Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тверской государственный технический университет»
(ТвГТУ)

Е.В. Борисова

ОПТИМАЛЬНЫЕ РЕШЕНИЯ В ИССЛЕДОВАНИИ УПРАВЛЯЕМЫХ СИСТЕМ: ТИПОВЫЕ МОДЕЛИ И ПРИКЛАДНЫЕ МЕТОДЫ

Учебное пособие

Тверь, 2017

УДК 551.5.001.57
ББК В6
В82

Борисова Е.В. Оптимальные решения в исследовании управляемых систем решения: типовые модели и прикладные методы: Учебное пособие. Тверь: ТвГТУ, 2017. 209 с.

В пособии изложены основные принципы построения и классификации моделей, отличительные черты задач анализа, синтеза, оптимизации. Детально разобраны методы решения задач линейного программирования; задач организации оптимальных перевозок, распределения и назначения; задач сокращения интервала неопределенности для унимодальных функций; основные формулы интерполяции; метод наименьших квадратов для нахождения параметров линейной и нелинейной регрессионных моделей. Приведены типовые алгоритмы и примеры прикладных задач.

Предназначено для магистров по направлению подготовки – 09.04.01 Информатика и вычислительная техника, профиль – информационное и программное обеспечение автоматизированных систем. Вид деятельности – научно-исследовательская. Пособие также может быть рекомендовано для магистрантов других технических направлений и профилей, занимающихся построением и исследованием моделей управляемых социально-экономических систем, а также для самообразования.

Рецензенты:

Хабаров А.Р. – заведующий кафедрой электронно-вычислительных машин Тверского государственного технического университета, кандидат технических наук, профессор;

Митин Н.А. – заведующий научно-образовательным центром «Прикладная математика», ведущий научный сотрудник Института прикладной математики им. М.В. Келдыша Российской академии наук, кандидат физико-математических наук, доцент.

ПРЕДИСЛОВИЕ

Магистратура – второй уровень высшего образования и в большей степени осознанный подход к выбранной сфере деятельности и другие карьерные перспективы. Как обозначено в методическом инструментарии по установлению квалификационных требований к претендентам на замещение должностей государственной гражданской службы и государственным гражданским служащим, это «руководители» высшей, главной и ведущей групп должностей гражданской службы; «советники» высшей, главной и ведущей групп должностей гражданской службы; «специалисты» высшей, главной и ведущей групп должностей.

Следовательно, студенты-магистранты – это специалисты, имеющие высшее профессиональное образование, определенный профессиональный опыт, в возрасте «периода достижений», когда личность использует интеллектуальные способности, чтобы сделать карьеру и избрать стиль жизни, уже имея за плечами определенный социальный, учебный, профессиональный опыт.

Пособие предназначено, прежде всего, для магистрантов, обучающихся по направлению «Информатика и вычислительная техника», изучающих дисциплину «Специальные главы высшей математики», содержание которой ориентировано на развитие общепрофессиональных и профессиональных компетенций, таких как: способность воспринимать математические, естественнонаучные и социально-экономические и профессиональные знания, умение приобретать, развивать и самостоятельно применять их для решения нестандартных задач, в том числе в новой или незнакомой среде и междисциплинарном контексте; владение существующими методами и алгоритмами решения задач распознавания и обработки данных.

Основной вид образовательной деятельности для магистранта – самостоятельная работа, это его организатор, технология и форма сбора материалов, анализа информации, инструмент самооценки и рефлексии. Для преподавателя – средство обратной связи и предмет оценочной деятельности. Консультирование в магистратуре, с одной стороны, выступает как условие обеспечения целостного индивидуального образовательного процесса, а с другой стороны, является самостоятельной структурной единицей взаимодействия преподавателя и магистранта.

Данное учебное пособие преследует две основные цели:

1. Дать студенту общее представление о принципах моделирования сложных систем (экономических, социальных, упарвленческих) и, более подробно, о методах решения базовых задач оптимального управления.

2. Предоставить студентам варианты классификаций моделей, методов для постановки и решения оптимизационных задач в различных прикладных областях.

Для достижения сформулированных целей:

- ✓ вводится однозначная терминология, используемая в процессе изложения материала;
- ✓ формируется представление о многообразии оптимизационных моделей;
- ✓ формулируются задачи моделирования как универсального инструмента исследования сложных систем;
- ✓ рассматриваются пошаговые методы расчета оптимальных значений;
- ✓ выполняется анализ алгоритмов, преимуществ и недостатков рассмотренных аналитических и численных методов оптимизации.

Учебное пособие подготовлено по результатам многолетнего опыта преподавания автором дисциплины «Специальные главы высшей математики» магистрантам, обучающимся по направлению «Информационное и программное обеспечение автоматизированных систем». При изложении аналитических методов расчета автор сознательно не приводит подробные выводы и сложные доказательства представленных математических зависимостей. Рассмотренные теоретические положения и примеры позволяют, с одной стороны, достаточно просто выполнить оценочные расчеты, с другой стороны, получить вполне адекватное представление о свойствах моделей реальных систем за счет их анализа.

Материал учебного пособия базируется на знаниях, умениях и компетенциях, приобретенных и сформированных в процессе освоения основной образовательной программы, в частности, дисциплин «Операционные системы», «Математика», «Информатика», «Экономика».

Для более эффективного усвоения материала фрагменты, представляющие наибольший интерес, выделены разными шрифтами, что позволяет акцентировать внимание читателя на аспектах, которые, по мнению автора, являются важными для понимания описанных моделей и методов. Курсив выделяет в тексте ключевые слова и фразы, термины и понятия, на которые следует обратить внимание и которые раскрывают смысл излагаемого материала.

Изложенный материал будет также полезен будущим специалистам, подготавливающим выпускные квалификационные работы, в которых требуется выполнить моделирование и исследование некоторой системы (объекта) с использованием методов поиска оптимальных решений.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Существуют справочники и существуют модели,
но соответствующие пары никогда не встречаются.
Качество вашей модели обратно пропорционально
количеству используемых справочников.

Законы расхождения Сэмюэля Клеменса

Математическое моделирование является мощным и эффективным инструментом исследования разнообразных объектов, систем и процессов в различных областях человеческой деятельности. Многообразие процессов, протекающих в исследуемых системах и объектах, обусловливает и многообразие математических методов и средств, используемых в теории моделирования. Одним из основных требований, предъявляемых к модели, является ее адекватность реальной системе, которая достигается за счет использования моделей с различным уровнем детализации, зависящим от особенностей структурно-функциональной организации системы и целей исследования. Процессы функционирования реальных систем практически невозможно описать полно и детально, что обусловлено их существенной сложностью. Основная проблема при разработке модели состоит в нахождении компромисса между простотой ее описания, что необходимо для исследования математическими методами, и необходимостью учета многочисленных особенностей, присущих реальной системе. Попытка построить единую универсальную модель сложной системы, несомненно, обречена на неудачу. Моделирование сложных систем в общем случае предполагает выполнение следующих основных этапов:

- ✓ формулировка целей моделирования;
- ✓ разработка концептуальной модели;
- ✓ разработка математической модели;
- ✓ параметризация модели;
- ✓ проверка адекватности модели (верификация модели);

✓ проведение экспериментов на модели (расчет характеристик);

✓ анализ результатов моделирования.

На этапе определения и формулирования целей моделирования определяется объект моделирования, формулируются задачи анализа и синтеза, выявляются наиболее важные характеристики, формулируется критерий эффективности, определяются требования к точности результатов моделирования и форме их представления.

Основное назначение концептуальной модели – выявление наиболее существенных аспектов структурно-функциональной организации, учет которых необходим для получения требуемых результатов. В концептуальной модели обычно в словесной форме приводятся сведения о природе и параметрах элементарных явлений исследуемой системы, о степени их взаимодействия, выявляются параметры, оказывающие наиболее существенное влияние на исследуемые характеристики системы. Одна и та же система может представляться различными концептуальными моделями, которые строятся в зависимости от целей исследования. Например, одна концептуальная модель может отображать временные аспекты функционирования системы, другая – показатели надежности, третья – физико-геометрические аспекты построения системы. Концептуальная модель служит основой для разработки математической модели в терминах конкретного математического аппарата.

Создание математической модели преследует две основные цели: 1) дать формализованное описание структуры и процесса функционирования системы для однозначности их понимания; 2) представить процесс функционирования системы в виде, допускающем аналитическое исследование системы с использованием методов, разработанных в рамках конкретного математического аппарата. Выбор метода моделирования зависит от многих факторов, в том числе от поставленных целей, сложности исследуемой системы, требований к номенклатуре исследуемых характеристик, требований к точности получаемых результатов.

Модели, заметим, не дают рекомендаций по решению проблем. Однако, они обеспечивают описательные результаты, то есть представляют моделируемую систему (например, дисперсия продаж некоторых товаров по месяцам в течение года). Специалист по моделям не применяет прямо полученный результат как решение, а сопоставляет его со своими оценками и прогнозами. Принимаемые решения основаны на сравнении путем обратной связи с первоначальной моделью, которая может модифицироваться при испытаниях в различных условиях. Результаты могут указывать, что проблема полностью не охвачена и требует изменений или реконструкции первоначальной модели. Иначе, моделирование представляет непрерывный процесс, а не одиночное решение одиночной проблемы. Существенную помощь в моделировании оказывают программные и компьютерные средства. Знание и понимание математических основ и методов построения моделей обеспечит их переложение на любой доступный программный продукт.

В практических задачах наибольшее значение придается:

- ✓ линейному программированию;
- ✓ оптимизационным моделям;
- ✓ регрессионному анализу;
- ✓ корреляционному и дисперсионному анализу.

Пособие содержит четыре основных раздела, введение, заключение, список литературы. Материал каждого раздела разбит на подразделы, которые имеют двойную нумерацию. Некоторые подразделы разбиты на пункты с заголовками, выделенными полужирным шрифтом.

Первый раздел содержит основные понятия и определения, касающиеся общих принципов моделирования, задачи, виды и классификацию моделей. Рассмотрены основы оценки погрешностей и компьютерных моделей. Определены отличительные черты задач анализа, синтеза, оптимизации.

Во втором разделе приводятся необходимые сведения из теории линейного программирования. Детально разобраны базовые методы решения задач линейного программирования: графический и симплексный. В отдельный подраздел выделена транспортная задача. Подробно представлены алгоритмы задач организации оптимальных перевозок, распределения и назначения.

В третьем разделе даются основные сведения теории одномерной оптимизации; методы сокращения интервала неопределенности для унимодальных функций. Вопросам аппроксимации и основным формулам интерполяции посвящена вторая часть этого раздела.

В четвертом разделе излагается теория нахождения параметров линейной и нелинейной регрессионных моделей, метод наименьших квадратов. В этом же разделе большое внимание уделяется элементам теории корреляции, результаты применения которой позволяют выявить и сформулировать ряд важных особенностей и закономерностей, присущих зависимым и независимым переменным.

Все разделы содержат значительное число типовых задач, решение которых подробно разобрано. Каждый раздел заканчивается перечнем вопросов для самоконтроля позволяющих читателю самостоятельно выполнить проверку степени усвоения изложенного материала.

Представленный список литературы не претендует на полноту и содержит ограниченный перечень литературных источников, которые в разной мере использовались при написании пособия. Этот перечень включает учебные издания и монографии, которые можно рекомендовать для дополнительного самостоятельного изучения.

Раздел 1. ОБЩИЕ ВОПРОСЫ МОДЕЛИРОВАНИЯ

1.1. Задачи и виды моделирования. Классификация моделей

Множество окружающих нас предметов и явлений обладают различными свойствами. Процесс познания этих свойств состоит в том, что мы создаем для себя некоторое представление об изучаемом объекте, помогающее лучше понять его внутреннее состояние, законы функционирования, основные характеристики. Такое представление, выраженное в той либо иной форме, называется моделью. Под *моделью* следует понимать любую систему, обладающую той же формальной структурой, при условии, что между системными характеристиками модели и оригиналом существует соответствие. Модель более доступна для изучения и исследования основных свойств объекта-оригинала. Замещение одного объекта другим с целью получения информации о важнейших свойствах объекта-оригинала можно назвать *моделированием*.

Модель (фр. modèle, от лат. modulus – «мера, аналог, образец») – это система, исследование которой служит средством для получения информации о другой системе, это упрощенное представление реального устройства и/или протекающих в нем процессов, явлений. Построение и исследование моделей, то есть моделирование, облегчает изучение имеющихся в реальном устройстве (процессе) свойств и закономерностей.

Моделирование – это и метод познания окружающего мира, дающий возможность управлять им. Он применялся еще в глубокой древности и постепенно охватывал все новые области знаний: технику, строительство, архитектуру, астрономию, физику, химию, биологию и, наконец, общественные, экономические науки, филологию, юриспруденцию. Признание практически во всех отраслях современной науки принес методу моделирования двадцатый век. Однако, методы моделирования долгое время развивалась независимо и отдельными науками. Отсутствовала единая система понятий, единая терминология. Например, еще в 1938 г. французский математик Курно в работе «Исследование математических принципов теории богатства» сформулировал «закон спроса». Ф. Кэнэ создал «экономическую таблицу», являющую собой попытку представить в форме математической модели процесс воспроизводства общественного продукта как единого целого. Количественный аспект анализа экономических явлений и процессов всегда занимал большое место в работах классиков отечественной и зарубежной экономики.

Создание электронно-вычислительных машин, обрабатывающих большие массивы информации, позволило решать задачи, требующие огромного объема вычислений, и разработать численные методы решения новых задач. Появились новые математические дисциплины, в частности *математическое программирование* – раздел математики, занимающийся решением задач нахождения экстремумов функции нескольких переменных с ограничениями на область допустимых значений этих переменных.

Математическое программирование и, его вариант линейное программирование, связано с целью решения задач по выбору оптимальной программы действий. Выделение класса экстремальных задач, определяемых линейным функционалом на множестве, задаваемом линейными ограничениями, следует отнести к 1930-м годам. Одними из первых, исследовавшими в общей форме задачи линейного программирования, были: Джон фон Нейман – математик и физик, доказавший основную теорему о матричных играх и изучивший экономическую модель, носящую его имя, и Н.В.Канторович – советский академик, лауреат Нобелевской премии (1975), сформулировавший ряд задач линейного программирования и предложивший в 1939 году метод их решения (метод разрешающих множителей), незначительно отличающийся от симплекс-метода.

Н.В. Канторович совместно с М.К. Гавуриным в 1949 году разработал метод потенциалов, который применяется при решении транспортных задач.

В последующих работах Н.В. Канторовича, В.С. Немчинова, В.В. Новожилова, А.Л. Лурье, А. Брудно, А.Г. Аганбегяна, Д.Б. Юдина, Е.Г. Гольштейна и других математиков и экономистов получили дальнейшее развитие математическая теория линейного и нелинейного программирования, так и приложение ее методов к исследованию различных экономических проблем.

Одновременно с развитием линейного программирования значительное внимание уделялось задачам нелинейного программирования, в которых либо целевая функция, либо ограничения, либо то и другое представлены нелинейными функциями. В 1951 году была опубликована работа Куна и Таккера, в которой приведены необходимые и достаточные условия оптимальности для решения задач нелинейного программирования. Эта работа послужила основой для последующих исследований в этой области. Начиная с 1955 года опубликовано много работ, посвященных квадратическому программированию (Баранкин Е.И., Dorfman R., Frank M., Wolfe P., Марковиц Г.). В работах Денниса Дж., Розена Дж. и Зонтендейка Г. разработаны градиентные методы решения задач нелинейного программирования. В настоящее время для эффективного применения методов математического программирования и решения задач на компьютерах разработаны алгебраические языки моделирования, такие как AMPL и LINGO.

Моделирование является обязательной частью исследований и разработок, неотъемлемой частью жизни, поскольку сложность любого материального объекта и окружающего его мира бесконечна вследствие неисчерпаемости материи и форм ее взаимодействия, – как внутри себя, так и с внешней средой. Преимущества моделирования состоят в возможности сравнительно простыми средствами изучать свойства системы, изменять ее параметры, вводить целевые и ресурсные характеристики внешней среды.

На идее моделирования базируется любой метод научного исследования, при этом, в теоретических методах используются различного рода знаковые, абстрактные модели, в экспериментальных – предметные модели. При исследовании сложное реальное явление заменяется некоторой упрощенной копией или схемой. Иногда построенная схема отражает какие-то существенные черты, позволяет разобраться в механизме явления, дает возможность предсказать его изменение. Одному и тому же явлению могут соответствовать разные модели. Задача исследователя – предсказывать характер явления и ход процесса. Иногда, бывает, что объект доступен, но эксперименты с ним дорогостоящи или привести к серьезным экологическим последствиям. Знания о таких процессах получают с помощью моделей. Как правило, моделирование используется для решения задач:

- ✓ исследования системы до того, как она спроектирована, с целью определения ее основных характеристик и правил взаимодействия элементов между собой и с внешней средой;
- ✓ проектирования системы для анализа и синтеза различных видов структур и выбора наилучшего варианта реализации с учетом сформулированных критериев оптимальности и ограничений;
- ✓ эксплуатации системы для получения оптимальных режимов функционирования и прогнозируемых оценок ее развития.

При этом одну и ту же систему можно описать различными типами моделей. Например, инженерные сети некоторого района можно промоделировать электрической схемой, гидравлической системой, математической моделью с использованием аппарата теории графов (Рис. 1.1).

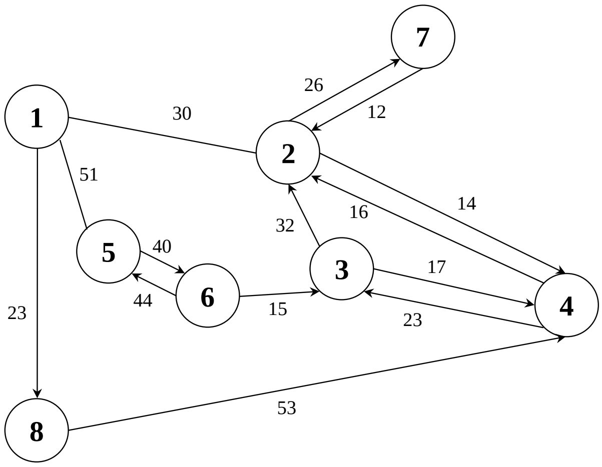


Рис. 1.1. Графовая модель участка транспортной сети

При наблюдении за объектом в голове у исследователя формируется некий мыслительный идеализированный образ объекта. При формировании когнитивной модели исследователи стремятся найти ответы на конкретные вопросы, поэтому от сложного устройства или объекта отсекается все второстепенное с целью конечного описания. Представление когнитивной модели на естественном языке называется содержательной моделью. Содержательная модель не равна когнитивной, так как не всегда исследователь может выразить в знаковой форме все элементы мысленного образа объекта. В естественнонаучных дисциплинах и в технике содержательную модель называют технической постановкой проблемы.

По функциональному признаку и цели содержательные модели подразделяют на:

✓ описательные – любое описание объекта;

✓ объяснительные – отвечают на вопрос: почему что-либо происходит?

✓ прогнозирующие – описывают будущее поведение объекта.

Содержательная модель, при формулировке которой используется понятие и представление предметных областей знаний занимающихся изучением объекта моделирования, называется концептуальной моделью. Если для описания системы используется естественный язык, то такое описание называется формальной содержательной моделью. Примерами содержательных моделей являются словесные постановки задач, программы и планы развития систем, деревья целей организации и др. Содержательные модели имеют самостоятельную ценность при решении задач исследования и управления системами, а также используются в смысле предварительного шага при разработке математических моделей. Поэтому качество математической модели зависит от качества соответствующей математической модели. В качестве языковых средств описания содержательных моделей используются естественный язык, диаграммы, таблицы, блок-схемы, графы.

Приведем классификацию моделей по форме представления.

Материальные – воспроизводят геометрические и физические свойства оригинала и всегда имеют реальное воплощение (детские игрушки, наглядные учебные пособия, макеты и др.):

✓ геометрически подобные масштабные, воспроизводящие пространственно-геометрические характеристики оригинала безотносительно его субстрату (макеты автомобилей, самолетов, зданий и сооружений, учебные муляжи и др.);

✓ воспроизводящие с масштабированием в пространстве и времени свойства и характеристики оригинала той же природы, что и модель, (гидродинамические модели судов, продувочные модели летательных аппаратов);

✓ аналоговые приборные, воспроизводящие исследуемые свойства и характеристики объекта оригинала в моделирующем объекте другой природы на основе некоторой системы прямых аналогий (разновидности электронного аналогового моделирования).

Информационные – совокупность информации, характеризующая свойства и состояния объекта, процесса, явления, а также их взаимосвязь с внешним миром).

Вербальные – описание свойств и характеристик оригинала на некотором естественном языке (текстовые материалы проектной документации, словесное описание результатов технического эксперимента).

Знаковые – информационная модель, выраженная специальными знаками (средствами любого формального языка).

Математические – математическое описание соотношений между количественными характеристиками объекта моделирования.

Графические – карты, чертежи, схемы, диаграммы, графы.

Табличные – таблицы: объект-свойство, объект-объект, двоичные матрицы и так далее.

Идеальные – материальная точка, абсолютно твердое тело, математический маятник, идеальный газ, геометрическая точка.

Неформализованные – системы представлений об объекте оригинале, сложившиеся в человеческом мозгу.

Графические иконические – черты, свойства и характеристики оригинала, реально или хотя бы теоретически доступные непосредственно зрительному восприятию (художественная графика, технологические карты).

Графические условные – данные наблюдений и экспериментальных исследований в виде графиков, диаграмм, схем.

Важный момент – сам характер науки предполагает изучение не одного конкретного явления, а широкого класса родственных явлений, предполагает необходимость формулировки общих категорических утверждений, которые называются законами. Естественно, что при такой формулировке многими подробностями пренебрегают. Чтобы более четко выявить закономерность сознательно идут на огрубление, идеализацию, схематичность, то есть изучают не само явление, а более или менее точную ее копию или модель. Всякий закон – это закон о моделях, а поэтому нет ничего удивительного в том, что с течением времени некоторые научные теории признаются непригодными. Это не приводит к краху науки, поскольку одна модель заменилась другой более современной.

Модель по определению всегда является лишь относительным, приближенным подобием объекта-оригинала и в информационном отношении принципиально беднее последнего. Это ее фундаментальное свойство. Произвольная природа объекта-оригинала означает, что этот объект может быть материально-вещественным, может носить чисто информационный характер и, наконец, может представлять собой комплекс разнородных материальных и информационных компонентов. Однако независимо от природы объекта, характера решаемой задачи и способа реализации математическая модель представляет собой информационное образование.

Пусть исследуется некоторая совокупность свойств $\{S\}$ реального объекта с помощь математики, тогда, например, объект – колебательная механическая система, а совокупность $\{S\}$ – характер и частота колебаний.

На основе содержательной модели переводим ее на формальный математический язык и тем самым переходим к математической модели; в этом заключается первый шаг – построение модели. Он существенно опирается на неформальное обсуждение постановки задачи и необходимую квалификацию исследователя в рассматриваемой области.

Второй шаг состоит в решении полученной математической задачи. Выбираем метод решения и реализуем его, причем сюда входит и проведение всех необходимых вычислений, в том числе и на ЭВМ. Это изучение проводится в рамках математики, но имеется и одна важная особенность: все элементы математической модели (в частности, все участвующие величины) являются как бы метками соответствующих реальных элементов. Это дает возможность в процессе решения математической задачи привлекать дополнительные сведения, которые могут упростить этот процесс, либо выделят из нескольких решений то, которое является наиболее приемлемым.

Полученное решение математической задачи нужно проанализировать, разобраться в его реальном смысле, сделать выводы, В этом состоит третий этап – этап интерпретации (истолкования) результата исследования модели. В него может входить и контроль правильности (как говорят, верификация) модели на основе сравнения результата с другими известными фактами, в частности с экспериментальными данными.

Основные этапы моделирования:

1. Постановка задачи. Определение цели анализа и пути ее достижения и выработки общего подхода к исследуемой проблеме. На этом этапе требуется глубокое понимание существа поставленной задачи. Иногда, правильно поставить задачу не менее сложно, чем ее решить. Постановка – процесс неформальный, общих правил нет.

2. Изучение теоретических основ и сбор информации об объекте оригинала. На этом этапе подбирается или разрабатывается подходящая теория. Если ее нет, устанавливаются причинно-следственные связи между переменными описывающими объект. Определяются входные и выходные данные, принимаются упрощающие предположения.

3. Формализация. Заключается в выборе системы условных обозначений и с их помощью записывать отношения между составляющими объекта в виде математических выражений. Устанавливается класс задач, к которым может быть отнесена полученная математическая модель объекта. Значения некоторых параметров на этом этапе еще могут быть не конкретизированы.

4. Выбор метода решения. На этом этапе устанавливаются окончательные параметры моделей с учетом условия функционирования объекта. Для полученной математической задачи выбирается какой-либо известный метод решения или разрабатывается специальный. При выборе метода учитываются знания исследователя, его предпочтения.

5. Реализация модели. Разрабатывается алгоритм, пишется программа, которая отлаживается, тестируется и, затем, находится решение поставленной задачи.

6. Анализ полученной информации. Сопоставляется полученное и предполагаемое решение, проводится контроль погрешности моделирования.

7. Проверка адекватности реальному объекту. Результаты, полученные по модели сопоставляются либо с имеющейся об объекте информацией или проводится эксперимент и его результаты сопоставляются с расчетными.

Процесс моделирования является итеративным. В случае неудовлетворительных результатов этапов 6 или 7 осуществляется возврат к одному из ранних этапов, который мог привести к разработке неудачной модели. Уточнение модели происходит до тех пор, пока не будут получены приемлемые результаты.

К *целям моделирования* можно отнести:

✓ понимание, как устроен конкретный объект, какова его структура, основные свойства, законы развития и взаимодействия с окружающим миром;

✓ проведение экспериментов;

✓ проектирование, управление и определение наилучших способов воздействия при заданных целях и критериях;

✓ прогнозирование прямых и косвенных последствий реализации заданных способов и форм воздействия на объект;

✓ тренировка и обучения специалистов;

✓ обработка информации.

Общие требования, предъявляемые к моделям:

1. Универсальность – характеризует полноту отображения моделью изучаемых свойств реального объекта.

2. Адекватность – способность отражать нужные свойства объекта с погрешностью не выше заданной.

3. Точность – оценивается степенью совпадения значений характеристик реального объекта и значения этих характеристик полученных с помощью моделей.

4. Экономичность – определяется затратами ресурсов ЭВМ памяти и времени на ее реализацию и эксплуатацию.

Пример. Рассмотрим модель в задаче о движении снаряда.

Снаряд пущен с Земли с начальной скоростью $V_0 = 30\, м/c$ под углом $\alpha = 45^0$ к ее поверхности; требуется найти траекторию движения снаряда и расстояние S между начальной и конечной точкой этой траектории.

Пренебрегая размерами снаряда, будем считать его материальной точкой. Введем систему координат *XOY*, совместив ее начало 0 с исходной точкой, из которой пущен снаряд, ось *OX* направим горизонтально, а ось *OY* – вертикально (Рис. 1.2).

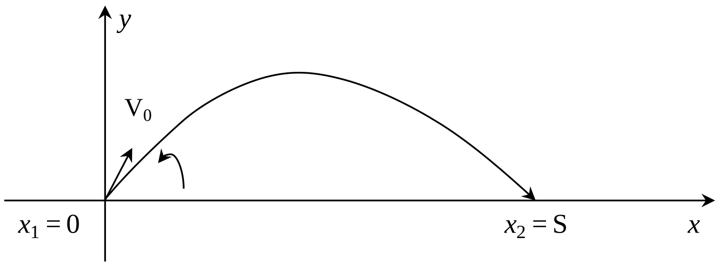


Рис. 1.2. Модельная траектория движения снаряда

Как это известно, из школьного курса физики, движение снаряда может быть описано формулами:

$$x = tV_0 \cos\alpha; \qquad y = tV_0 \sin\alpha - \frac{gt^2}{2},$$

где t – время; $g = 10\, м/c^2$ – ускорение свободного падения.

Эти формулы составляют математическую модель поставленной задачи. Выражая t через x из первого уравнения и подставляя во второе, получим уравнение траектории движения снаряда: $y = x tg\alpha - \dfrac{g x^2}{2V_0^2 \cos^2 \alpha}$.

Эта кривая (парабола) пересекает ось OX в двух точках: $x_1 = 0$ (начало траектории) и $x_2 = S$ (место падения снаряда). Подставляя в полученные формулы заданные значения V_0 и α, получим значение: $S = 90\ м$. Отметим, что при построении этой модели использован ряд предположений: например, считается, что Земля плоская, а воздух и вращение Земли не влияют на движение снаряда.

1.2. Математические модели

Математическое моделирование и связанный с ним компьютерный эксперимент незаменимы в случаях, когда натурный эксперимент невозможен или затруднен по тем или иным причинам. Например, нельзя поставить натурный эксперимент в истории, чтобы проверить, «что было бы, если бы...» Однако все можно сделать на компьютере, построив предварительно математические модели изучаемых явлений.

Математическое моделирование – это теоретико-экспериментальный метод познавательно-созидательной деятельности, метод исследования и объяснения явлений, процессов и систем (объектов-оригиналов) на основе создания новых объектов – математических моделей. Существующие математические методы и модели позволяют решать задачи большой размерности и учитывать большее число показателей и факторов влияния, а время решения задач значительно сокращается с применением компьютера.

Строительный материал и инструменты этих моделей – математические понятия. Они накапливались и совершенствовались в течение тысячелетий. Современная математика дает исключительно мощные и универсальные средства исследования. Практически каждое понятие в математике, каждый математический объект, начиная от понятия числа, является математической моделью.

При построении математической модели, изучаемого объекта или явления выделяют те его особенности, черты и детали, которые с одной стороны содержат более или менее полную информацию об объекте, а с другой допускают математическую формализацию. Математическая формализация означает, что особенностям и деталям объекта можно поставить в соответствие подходящие адекватные математические понятия: числа, функции, матрицы и так далее. Тогда связи и отношения, обнаруженные и предполагаемые в изучаемом объекте между отдельными его деталями и составными частями можно записать с помощью математических отношений: равенств, неравенств, уравнений. В результате получается математическое описание изучаемого процесса или явление, то есть его математическая модель.

Изучение математической модели всегда связанно с некоторыми правилами действия над изучаемыми объектами. Эти правила отражают связи между причинами и следствиями. Построение математической модели – это центральный этап исследования или проектирования любой системы. От качества модели зависит весь последующий анализ объекта.

Построение модели – это не формальная «механическая» процедура. Большое влияние на этот процесс оказывает опыт исследователя, его знания предметной области и всегда опирается на определенный эмпирический материал. Модель должна быть достаточно точной, адекватной и удобной для использования.

Математическая модель – это приближенное описание какого-либо класса явлений или объектов реального мира на языке математики. Такие модели появились вместе с математикой много веков назад. Связь с математической модели с реальностью осуществляется с помощью цепочки гипотез, идеализаций и упрощений. С помощью математических методов описывается, как правило, идеальный объект, построенный на этапе содержательного моделирования.

Требования, предъявляемые к математическим моделям созвучны с общими требованиями. Математическим моделям должен быть присущ ряд следующих свойств.

✓ *Универсальность* – то есть применимость модели к анализу ряда однотипных систем в одном или нескольких режимах функционирования. Это позволяет расширить область применимости модели для решения большего круга задач и характеризует полноту отображения изучаемых свойств реального объекта.

✓ *Адекватность* – способность отражать нужные свойства объекта с погрешностью не выше заданной. В общем смысле под адекватностью понимают правильное качественное и достаточно точное количественное описание именно тех характеристик системы, которые важны в данном конкретном случае, то есть соответствие модели исходной реальной системе и учет, прежде всего, наиболее важных качеств, связей. Оценить адекватность выбранной модели, особенно, например, на начальной стадии проектирования, когда вид создаваемой системы еще неизвестен, очень сложно. В такой ситуации часто полагаются на опыт предшествующих разработок или применяют определенные методы, например, метод последовательных приближений;

✓ *Точность* – оценивается степенью совпадения значений характеристик реального объекта и значения этих характеристик полученных с помощью моделей. Здесь важной задачей является оценка полученных результатов и имеющихся исходных данных, согласование их как между собой, так и с погрешностью используемой модели;

✓ *Экономичность* – определяется затратами ресурсов ЭВМ памяти и времени на ее реализацию и эксплуатацию. Свойство экономичности часто связывают с ее простотой. Экономичная модель, как показывает практика, – результат компромисса между отпущенными ресурсами и особенностями используемой модели.

✓ *Полнота* – позволяет отразить в достаточной мере именно те характеристики и особенности системы, которые интересуют исследователя с точки зрения поставленной цели проведения вычислительного эксперимента. Например, модель может достаточно полно описывать протекающие в системе процессы, но не отражать его габаритные, массовые или стоимостные показатели.

✓ *Робастность* – характеризует ее устойчивость по отношению к погрешностям исходных данных, способность нивелировать эти погрешности и не допускать их чрезмерного влияния на результат вычислительного эксперимента. Причинами низкой робастности модели могут быть необходимость при ее количественном анализе вычитания близких друг к другу приближенных значений величин или деления на малую по модулю величину, а также использование функций, быстро изменяющихся в промежутке, где значение аргумента известно с невысокой точностью.

✓ *Продуктивность* – связана с достоверностью исходных данных. Если они являются результатом измерений, то точность их измерения должна быть выше, чем для тех параметров, которые получаются при использовании моделирования. В противном случае модель будет непродуктивной и ее применение для анализа конкретной системы потеряет смысл. Ее можно будет использовать лишь для оценки и характеристик некоторого класса систем с гипотетическими исходными данными.

✓ *Наглядность* – является желательным, но необязательным свойством моделей. Однако использование модели и ее модификация упрощаются, если составляющие имеют ясный содержательный смысл. Это позволяет предвидеть результаты вычислительного эксперимента и облегчить контроль их правильности.

✓ *Простота* – отображение только существенных сторон объекта.

✓ *Конечность* – модель отображает оригинал лишь в конечном числе его отношений и, кроме того, ресурсы моделирования конечны.

✓ *Информативность* – содержание достаточной информации об объекте – в рамках гипотез, принятых при построении модели;

✓ *Потенциальность* – предсказуемость модели и ее свойств.

По форме представления математических моделей различают:

✓ *аналитические модели.* Их решения ищутся в замкнутом виде, в виде функциональных зависимостей. Удобны при анализе сущности описываемого явления или процесса и использовании в других математических моделях, но отыскание их решений бывает весьма затруднено. Примером успешного аналитического моделирования является открытие планеты Нептун на основании теоретического анализа движения планеты Уран. Французский ученый Урбан Леверье, на основании расчетов предсказал, что есть еще одна планета и указал ее координаты. Немецкий астроном Галле обнаружил данную планету в указанной точке;

✓ *численные модели*. Их решения – дискретный ряд чисел (таблицы). Эти модели универсальны, удобны для решения сложных задач, но не наглядны и трудоемки при анализе и установлении взаимосвязей между параметрами. Такие модели реализуют в виде программных комплексов – пакетов программ для расчета на компьютере. Программные комплексы бывают прикладные, привязанные к предметной области и конкретному объекту, явлению, процессу, и общие, реализующие универсальные математические соотношения (например, расчет системы алгебраических уравнений);

✓ *формально-логические информационные модели* – это модели, созданные на формальном языке. Например: модель формальной системы в математике и логике как любая совокупность объектов, свойства которых и отношения между которыми удовлетворяют аксиомам и правилам вывода формальной системы, служащей тем самым совместным (неявным) определением такой совокупности; модель в теории алгебраических систем как совокупность некоторого множества и заданных на его элементах свойств и отношений.

В основу классификации математических моделей могут быть положены различные принципы. Если исходить из общих задач математического моделирования в разных науках безотносительно к математическому аппарату, наиболее естественна такая классификация: дескриптивные (описательные) модели; оптимизационные; многокритериальные; игровые модели; детерминированные; вероятностные.

Дескриптивные (описательные) модели. Например, моделирование движения кометы, вторгшейся в Солнечную систему, производится с целью предсказания траектории ее полета, расстояния, на котором она пройдет от Земли, и т. д. В этом случае цели моделирования носят описательный характер, поскольку нет никаких возможностей повлиять на движение кометы, что-то в нем изменить.

Оптимизационные модели используются для описания процессов, на которые можно воздействовать, пытаясь добиться достижения заданной цели. В этом случае в модель входит один или несколько параметров, доступных влиянию. Например, меняя тепловой режим в зернохранилище, можно подобрать такой режим, чтобы достичь максимальной сохранности зерна, т.е. оптимизировать процесс хранения.

Многокритериальные модели. Нередко приходится оптимизировать процесс по нескольким параметрам одновременно, причем цели могут быть весьма противоречивыми. Например, зная цены на продукты и потребность человека в пище, нужно организовать питание больших групп людей физиологически правильно и, одновременно с этим, как можно дешевле. Ясно, что эти цели совсем не совпадают, т. е. при моделировании будет использоваться несколько критериев, между которыми нужно искать баланс.

Детерминированными называются модели, в которых отсутствуют, какие бы то ни было случайные изменения: внешних воздействий, внутренних параметров и самих переменных. В таких моделях все поведение объекта определяется конкретными значениями начальных условий и входных переменных. Иначе говоря, в них все точно определено (детерминировано).

Детерминированной моделью является график движения автомобиля с постоянным ускорением, так как скорость его меняется по определенному закону и легко прогнозируема (Рис. 1.3).

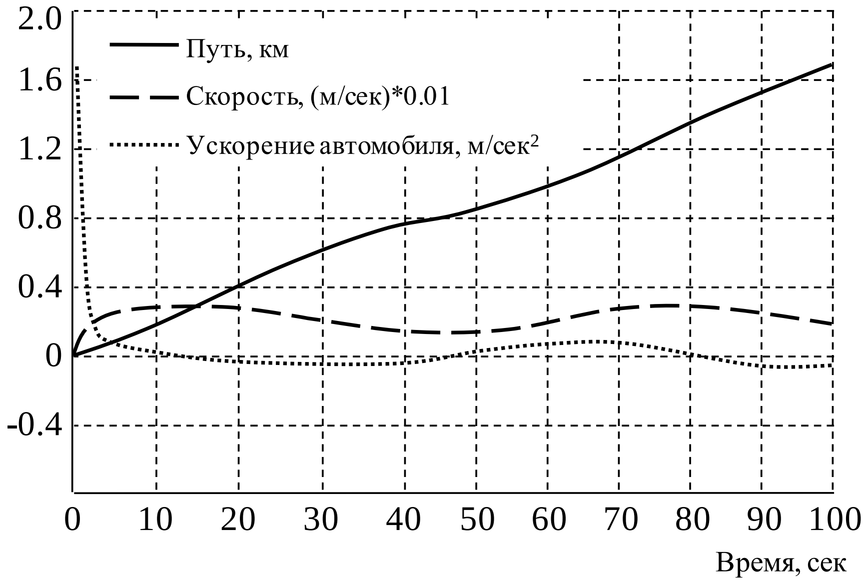


Рис. 1.3. Графическое представление характеристик движения автомобиля

Вероятностными являются модели, в которых учитывается случайный характер изменений значений входных, промежуточных и выходных переменных, а также параметров моделируемого объекта. Такие модели характеризуются функциями или плотностями распределения вероятностей и средними характеристиками смещения и рассеяния, например, математическим ожиданием и дисперсией. Вероятностную модель можно рассмотреть на примере графиков функций распределения вероятностей – постепенный переход от одних вероятностных моделей (1 – равномерное распределение) к другим вероятностным моделям (2 – нормальное распределение с разными значениями параметра), а также в пределе и к детерминированной модели 3 (Рис. 1.4).

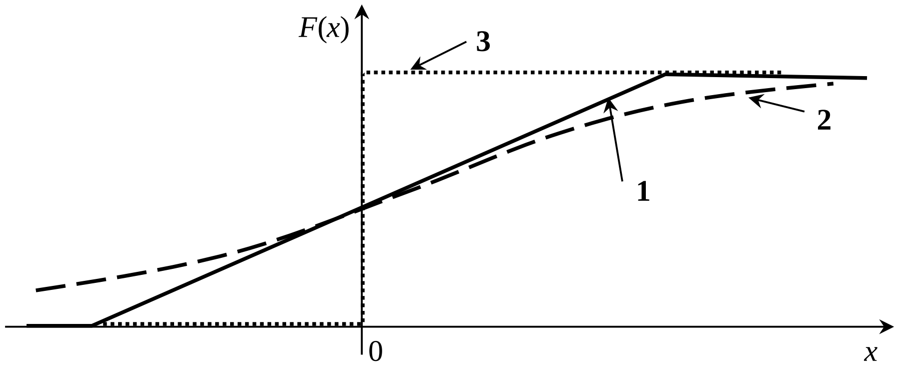


Рис. 1.4. Эволюция графиков вероятностных моделей

В дополнение к рассмотренным выше классам математических моделей приведем формальную классификацию, которая основывается на используемых математических средствах и часто строится в семантической форме, например:

- ✓ статические или динамические;
- ✓ дискретные или непрерывные;
- ✓ аналитические и численные;
- ✓ линейные или нелинейные модели.

Статистическая модель включает описание связей между основными переменными моделируемого объекта в установившемся режиме без учета изменения параметров во времени. В динамической модели описываются связи между основными переменными моделируемого объекта при переходе от одного режима к другому. В непрерывных моделях переменные принимают значения из некоторого промежутка, в дискретных моделях переменные принимают изолированные значения. Аналитические модели описывают объект в виде некоторых функциональных отношений и/или логических условий. Численные модели отражают элементарные этапы вычислений и последовательность их проведения.

Особое значение, вследствие важности, имеют линейные модели.

В математической теории линейность означает, что оператор преобразования «вход-выход» линеен. Иногда линейное свойство системы называют принципом суперпозиции. Динамические линейные стационарные системы с сосредоточенными параметрами описываются обыкновенными линейными дифференциальными уравнениями с постоянными коэффициентами. Для таких моделей гарантировано получение аналитического решения, согласно теореме Коши о существовании и единственности решения линейных дифференциальных уравнений n-го порядка с постоянными коэффициентами.

Системы, исходно нелинейные, но линеаризованные в окрестности опорных траекторий, также описываются с помощью линейных уравнений. Однако при этом не учитываются важные, интересные, тонкие эффекты, связанные с проявлением нелинейных свойств. Таким образом, линейные модели являются упрощенными по сравнению с нелинейными.

Нелинейные модели. Это динамическая система, в которой протекают процессы, описываемые нелинейными дифференциальными уравнениями. Нелинейные модели предполагаются параметрическими, описываемые нелинейными уравнениями. При использовании параметрической нелинейной регрессии, зависимая переменная (также называемая откликом) представляется в виде функции от комбинации нелинейных параметров и одной или нескольких независимых переменных (называемых предикторами). Модель может быть как одномерная (с одной выходной переменной), так и многомерной (с несколькими выходными переменными). Параметры могут быть в виде экспоненциальной, тригонометрической, степенной или любой другой нелинейной функции. Для нахождения нелинейных параметров обычно используются итерационный алгоритм.

Основу математических моделей многих процессов и явлений в физике, химии, биологии, экономике и других областях составляют уравнения различного вида: линейные уравнения, нелинейные уравнения, обыкновенные дифференциальные уравнения, дифференциальные уравнения в частных производных и т.д. Для решения подобных уравнений необходимо иметь возможность вычислять значения функций, входящих в описание математической модели рассматриваемого процесса или явления, при произвольном значении аргумента. Для сложных моделей подобные вычисления могут быть трудоемкими даже при использовании компьютера. Используемые в математических моделях функции могут быть заданы как аналитическим способом (в виде формулы), так и табличным, при котором функция известна только при определенных дискретных значениях аргумента.

Построение и исследование математических моделей возможно принципиально двумя способами. Аналитическим – то есть выводом из физических законов, математических аксиом или теорем. Аналитические методы направлены в основном на построение точных или асимптотических формул для изучения свойств решений с помощью этих формул. Точные формулы могут либо охватывать совокупность все решений заданного уравнения, либо представлять отдельные, частные решения, удовлетворяющие определенным свойствам. Экспериментальным путем, то есть посредством обработки результатов эксперимента и подбора аппроксимирующих (приближенно совпадающих) зависимостей.

В процессе работы с математическими моделями зачастую прибегают к процедурам линеаризации и идентификации с последующей проверкой на адекватность.

Линеаризация. Пусть дана математическая модель некоторой системы:

$$M = N(X, Y, A),$$

где X – множество входов; Y – множество выходов; A – множество состояний.

Изобразим ее схематически: $X \to A \to Y$. Отметим, что если X, Y, A – линейные пространства (множества), а $\varphi : X \to A$ и $\psi : A \to Y$ – линейные операторы (т. е. любые линейные комбинации $ax+by$ аргументов φ и ψ преобразуют в соответствующие линейные комбинации $a\varphi(X)+b\varphi(Y)$ и $a\psi(X)+b\psi(Y)$), то модель называется линейной.

Модели, не отвечающие указанным условиям, называются нелинейными. Они труднее поддаются исследованию, хотя и более актуальны. Нелинейные модели менее изучены, поэтому их часто линеаризуют – сводят к линейным моделям корректной процедурой линеаризации.

Пример. Применим операцию линеаризации к модели вида $y=\dfrac{at^2}{2}$, где $0 \le t \le 4$, которая является нелинейной (квадратичной). Для этого заменим один из множителей t на его среднее значение для рассматриваемого промежутка, т. е. на $t=2$. Такая линеаризация, хоть и очень наглядная, но достаточно грубая. Но выполненная замена дает линейную модель вида $y=at$. Более точную линеаризацию можно провести заменим множитель t не на среднее, а на значение в некоторой заранее не известной точке. Тогда, как следует из теоремы о среднем из курса высшей математики, замена будет достаточно точна. Но для нее необходимо оценить значение неизвестной точки. На практике используются и другие, более точные и тонкие процедуры линеаризации, использующие сложный математический аппарат.

Идентификация. Пусть существует некоторый набор моделей:

$$M=M(X,Y,A); \quad A=\{a_i\},$$

где $a_i=(a_{i1},a_{i2},\ldots,a_{ik})$ – вектор состояния объекта.

Если вектор a_i зависит от некоторых неизвестных параметров, то задача идентификации (модели, параметров модели) состоит в определении их допустимых значений по некоторым дополнительным условиям, например, экспериментальным данным, характеризующим состояние объекта.

Идентификация – задача построения по результатам наблюдений математических моделей различного типа, адекватно описывающих поведение объекта. Если $S = \{s_1, s_2, \ldots, s_n\}$ – некоторая последовательность сообщений, получаемых от источника информации об объекте, а $M = \{m_1, m_2, \ldots, m_z\}$ – последовательность моделей, описывающих S, среди которых, возможно, содержится оптимальная модель, то идентификация модели M означает, что последовательность S позволяет различать (по рассматриваемому критерию адекватности) две разные модели в наборе моделей M. Последовательность сообщений (данных) S называется информативной, если она позволяет различать разные модели в M. Цель идентификации – построение надежной, адекватной, эффективно функционирующей гибкой модели на основе минимального объема информативной последовательности сообщений. Наиболее часто используемые методы идентификации (параметров систем): метод наименьших квадратов, метод максимального правдоподобия, метод байесовских оценок, метод марковских цепных оценок, метод эвристик, экспертное оценивание.

Пример. Применим операцию идентификации параметра a в модели предыдущего примера. Для этого необходимо задать дополнительно значение y для некоторого t, например, $y = 6$ при $t = 3$. Тогда из модели получаем: $6 = 9a/2$, откуда найдем значение параметра $a = 12/9 = 4/3$. Идентифицированный параметр a определяет следующий вариант модели $y = \frac{2t^2}{3}$. Заметим, что методы идентификации моделей значительно сложнее, чем используемый в приведенном примере.

Пример. Оценим адекватность модели $y = at$, полученной в результате линеаризации выше, исходной модели $y = \frac{at^2}{2}$, при $0 \le t \le 4$. В качестве меры (критерия) адекватности рассмотрим меру – абсолютное значение разности между точным значением и модельным значением. Отклонение точной модели от линеаризованной будет в рамках этого критерия равно $|\frac{at^2}{2} - 2at|$. Если $a > 0$, то, как несложно оценить с помощью производной, эта погрешность будет экстремальна при $t = 2a$. Например, если $a = 1$, эта величина не превосходит 2. Полученное достаточно отклонение достаточно велико, и можно заключить, что линеаризованная модель в данном случае не является адекватной (как исходной системе, так и нелинеаризованной модели).

Дополнительно оценим чувствительность модели к изменениям входных параметров. Из рассматриваемого примера следует, что чувствительность модели $y = \frac{at^2}{2}$, при $0 \le t \le 4$ такова, что изменение входного параметра t всего на 1% приводит к изменению выходного параметра y более чем на 2%, то есть эта модель является чувствительной.

Математические модели универсальны и достаточно дешевы, они позволяют поставить «чистый» эксперимент (то есть в пределах точности модели исследовать влияние отдельного параметра при постоянстве других), прогнозировать развитие явления или процесса, отыскать способы управления ими. Результаты математического моделирования нуждаются в обязательном сопоставлении с данными физического моделирования – с целью проверки получаемых данных и для уточнения самой модели. С другой стороны, любая формула – это разновидность модели и, следовательно, не является абсолютной истиной, а всего лишь этап на пути ее познания.

1.3. Погрешности математических моделей. Компьютерное моделирование

В условиях большой размерности задач и нехватки компьютерных ресурсов исследователи вынуждены были ограничивать расчетную схему и представлять модели в упрощенном виде. С другой стороны, именно точность расчета позволяет получать прямую выгоду от использования моделирования. Погрешности имеют место, как при измерениях, так и при теоретическом моделировании. Для теоретических моделей, в соответствии с природой возникновения, различают:

- ✓ погрешности, возникающие при разработке физической модели;
- ✓ погрешности, возникающие при составлении математической модели;
- ✓ погрешности, возникающие при анализе математической модели;
- ✓ погрешности, связанные с конечным числом разрядов чисел при вычислениях. В последнем случае, например, число π в рамках символической записи как отношение длины окружности к диаметру представляет собой точное число, но попытка записать его в численном виде ($\pi = 3.14159265\ldots$) вызывает погрешность, связанную с конечным числом разрядов.

Перечисленные погрешности присутствуют всегда. Избежать их невозможно, и их называются методическими. При измерениях методические погрешности проявляют себя как систематические.

Пример. Рассмотрим погрешности физической и математической модели маятника, возникающие при измерении периода колебаний маятника в виде тела, подвешенного на нити. Физическая модель маятника:

- ✓ нить – невесома и нерастяжима;
- ✓ тело – материальная точка;
- ✓ трение отсутствует;
- ✓ тело совершает плоское движение;

✓ гравитационное поле – однородное (т. е. $g = const$ во всех точках пространства, в которых находится тело);

✓ влияние других тел и полей на движение тела отсутствует.

Уточним принятые допущения, очевидно, что реальное тело не может быть материальной точкой, оно имеет объем и форму, в процессе движения или со временем тело деформируется. Кроме того, нить имеет массу, она обладает упругостью и также деформируется. На движение маятника влияет движение точки подвеса, обусловленное действием вибраций, всегда имеющих место. Также на движение маятника влияет сопротивление воздуха, трение в нити и способ ее крепления, внешние магнитное и электрическое поля, неоднородность гравитационного поля Земли и даже влияние гравитационного поля Луны, Солнца и окружающих тел.

Перечисленные факторы, в принципе, могут быть учтены, однако сделать это достаточно трудно. Для этого потребуется привлечь почти все разделы физики. В конечном счете, учет этих факторов значительно усложнит физическую модель маятника и ее анализ. Не учет перечисленных, а также множества других, не упомянутых здесь факторов, существенно упрощает анализ, но приводит к погрешностям исследования.

Математическая модель маятника: в рамках выбранной простейшей физической модели математическая модель маятника – дифференциальное уравнение движения маятника – имеет следующий вид:

$$\frac{d^2\varphi}{dt^2} + \frac{g}{L} \cdot \sin\varphi = 0, \qquad (1)$$

где L – длина нити; φ – отклонение тела от положения равновесия.

При $\varphi << 1$ считают, что $\sin\varphi \approx \varphi$, и тогда уравнение движения записывается:

$$\frac{d^2\varphi}{dt^2} + \frac{g}{L} \cdot \varphi = 0. \qquad (2)$$

Получено линейное дифференциальное уравнение, которое может быть решено точно. Его решение имеет вид: $\varphi = \varphi_0 \cdot \sin(\omega_0 t + \varphi)$, где $\omega_0 = \sqrt{\frac{g}{L}}$. Отсюда следует, что период колебаний маятника $T_0 = 2\pi / \omega_0$ не зависит от амплитуды φ_0. Однако, это решение нельзя считать точным решением задачи о колебаниях маятника, представленного простейшей физической моделью, поскольку исходное уравнение (1) было другим.

Уточним решение. Если разложить $\sin\varphi$ в степенной ряд и учесть хотя бы первые два члена разложения, т. е. считать, что $\sin\varphi \approx \varphi + \varphi 3/6$, то решение дифференциального уравнения существенно усложнится. Приближенно его можно записать в виде $\varphi \approx \varphi_0(\sin\omega t + \varphi)$, где $\omega = \omega_0\left(1 - \frac{\varphi_0^2}{16}\right)$. Отсюда следует, что в данном приближении период колебаний маятника $T = 2\pi/\omega$ зависит от амплитуды колебаний по параболическому закону.

Таким образом, погрешность математической модели вида (2), связанная с заменой $\sin\varphi$ на φ, приводит к погрешности результата расчета периода колебаний маятника. Оценка этой погрешности может быть получена из решения задачи во втором приближении.

Проблема построения и анализа математической модели объекта исследования с заданной точностью, а также оценка погрешности расчетов в ряде случаев достаточно сложна. Требуется высокая математическая культура исследователя, необходим тщательный математический анализ и самой модели, и применяемых методов решения. Например, не имеет смысла требование решения уравнения (1) с точностью, существенно превышающей точность построения физической модели. В частности, в предыдущем примере нет смысла делать замену $\sin\varphi \approx \varphi + \varphi 3/6$ вместо

$\sin\varphi \approx \varphi$, если нить заметно деформируется или сопротивление воздуха велико.

Применение ЭВМ значительно увеличило возможности построения и исследования математических моделей, однако не следует думать, что совершенное знание математики, численных методов и языков программирования позволит решить любую физическую и прикладную задачу. Дело в том, что даже самые изящные и точные методы расчетов не могут исправить ошибки, допущенные при построении физической модели. Действительно, если длина нити маятника – L не постоянна, или если размеры тела сопоставимы с длиной нити, или трение велико и колебания маятника быстро затухают, то даже абсолютно точное решение уравнения (1) не позволит получить точное решение задачи о колебаниях маятника.

Модель адекватна объекту, если результаты теоретических исследований (расчет) совпадают с результатами опыта (измерений) в пределах погрешности последнего. Но с этой стороны также есть ряд сложностей.

Измерение – научно обоснованный опыт для получения количественной информации с требуемой или возможной точностью о параметрах объекта измерения. В метрологии определение понятия «измерение» дает ГОСТ 16.263-70. Измерение включает в себя следующие понятия: объект измерения; цель измерения; условия измерения (совокупность влияющих величин, описывающих состояние окружающей среды и объектов); метод измерения, т. е. совокупность приемов использования принципов и средств измерений. Принцип измерения – совокупность физических явлений, положенных в основу измерения; методика измерения, т. е. установленная совокупность операций и правил, выполнение которых обеспечивает получение необходимых результатов в соответствии с данным методом; средства измерения; измерительные приборы; измерительноинформационные системы; результаты измерений; погрешность измерений; понятия, характеризующие качество измерений: достоверность (характеризуется доверительной вероятностью, т. е. вероятностью того, что истинное значение измеряемой величины находится в указанных пределах); правильность (характеризуется значением систематической погрешности); сходимость (близость друг к другу результатов измерений одной и той же величины, выполняемых повторно одними и теми же методами и средствами и в одних и тех же условиях; отражает влияние случайных погрешностей на результат); воспроизводимость (близость друг к другу результатов измерений одной и той же величины, выполняемых в разных местах, разными методами и средствами, но приведенных к одним и тем же условиям).

Специфические операции математического моделирования, например, идентификация, линеаризация не сводятся в ЭВМ к преобразованию в ней программ. Отметим основные причины, несколько тормозящие выход математического моделирования на новые уровни информационных технологий:

✓ традиционное описание модели системами математических уравнений, соотношений; в то же время, большинство плохо структурированных и плохо формализуемых систем описываются с помощью экспертных данных, эвристических и имитационных процедур, интегрированных пакетов программ, графических образов и т. д.;

✓ существующие средства описания и представление моделей на ЭВМ не учитывают специфику моделирования, нет единого представления моделей, генерации новых моделей по банку моделей;

✓ недооценка возможностей компьютера, который может делать больше, чем простая реализация алгоритма, как правило, структурируемого и/или реализуемого хорошо, отсутствие доступа к опыту моделирования на ЭВМ.

В базовой пятерке: «система (исследуемая среда) – модель (описание среды) – алгоритм (программа) – компьютер (компьютерная технология) – пользователь (выработка решения)» при *компьютерном моделировании* главную роль играют уже алгоритм (программа), компьютер и технология, точнее, инструментальные системы для компьютера, компьютерные технологии. В имитационном моделировании (при отсутствии строгого и формально записанного алгоритма) главную роль играют технология и средства моделирования; аналогичная ситуация наблюдается в когнитивной графике. Модель не эквивалентна программе, а моделирование не сводится к программированию.

Основные функции компьютера при моделировании:

- ✓ исполнение роли вспомогательного средства для решения задач, доступных и для обычных вычислительных средств, алгоритмам, технологиям;
- ✓ исполнение роли средства постановки и решения новых задач, не решаемых традиционными средствами, алгоритмами, технологиями;
- ✓ исполнение роли средства конструирования компьютерных обучающих и моделирующих сред типа: «обучаемый – компьютер – обучающий», «обучающий – компьютер – обучаемый», «обучающий – компьютер – группа обучаемых», «группа обучаемых – компьютер – обучающий», «компьютер – обучаемый – компьютер»;
- ✓ исполнение роли средства моделирования для получения новых знаний;
- ✓ исполнение роли «обучения» новых моделей (самообучение модели).

Особенность компьютерных систем моделирования – их высокая интеграция и интерактивность. Часто компьютерные среды функционируют в режиме реального времени.

Вычислительный эксперимент – разновидность компьютерного моделирования. Можно говорить и о специальных пакетах прикладных программ, текстовых, графических и табличных процессоров, визуальных и когнитивных средах (особенно, работающих в режиме реального времени), позволяющих осуществлять компьютерное моделирование.

Компьютерное моделирование и вычислительный эксперимент становятся новым инструментом, новой технологией из-за возрастающей необходимости перехода от исследования линейных математических моделей (для которых достаточно хорошо известны или разработаны методы исследования) к исследованию сложных и нелинейных математических моделей (анализ которых гораздо сложнее). Образно, говоря: «наши знания об окружающем мире – линейны и детерминированы, а процессы в окружающем мире – нелинейны и стохастичны».

Компьютерное моделирование, от постановки задачи до получения результатов, состоит из *этапов*.

1. Постановка задачи.
   - ✓ Формулировка задачи.
   - ✓ Определение цели и приоритетов моделирования.
   - ✓ Сбор информации о системе, объекте моделирования.
   - ✓ Описание данных (их структуры, диапазона, источника и т. д.).
2. Предмодельный анализ.
   - ✓ Анализ существующих аналогов и подсистем.
   - ✓ Анализ технических средств моделирования (ЭВМ, периферийные устровйства).
   - ✓ Анализ программного обеспечения (языки программирования, пакеты прикладных программ, инструментальные среды).
   - ✓ Анализ математического обеспечения (модели, методы, алгоритмы).

3. Анализ задачи (модели).
   - ✓ Разработка структур данных.
   - ✓ Разработка входных и выходных спецификаций, форм представления данных.
   - ✓ Проектирование структуры и состава модели (подмоделей).
4. Исследование модели.
   - ✓ Выбор методов исследования подмоделей.
   - ✓ Выбор, адаптация или разработка алгоритмов, их псевдокодов.
   - ✓ Сборка модели в целом из подмоделей.
   - ✓ Идентификация модели, если в этом есть необходимость.
   - ✓ Формулировка критериев адекватности, устойчивости и чувствительности модели.
5. Программирование (проектирование программы).
   - ✓ Выбор метода тестирования и тестов (контрольных примеров).
   - ✓ Кодирование на языке программирования.
   - ✓ Комментирование программы.
6. Тестирование и отладка.
   - ✓ Синтаксическая отладка.
   - ✓ Семантическая отладка (отладка логической структуры).
   - ✓ Тестовые расчеты, анализ результатов тестирования.
   - ✓ Оптимизация программы.
7. Оценка моделирования.
   - ✓ Оценка средств моделирования.
   - ✓ Оценка адекватности моделирования.
   - ✓ Оценка чувствительности модели.
   - ✓ Оценка устойчивости модели.
8. Документирование.
   - ✓ Описание задачи, целей.
   - ✓ Описание модели, метода, алгоритма.

- ✓ Описание среды реализации.
- ✓ Описание возможностей и ограничений.
- ✓ Описание входных и выходных форматов, спецификаций.
- ✓ Описание тестирования.
- ✓ Создание инструкций для пользователя.

9. Сопровождение.
   - ✓ Анализ применения, периодичности использования, количества пользователей, типа использования (диалоговый, автономный и др.), анализ отказов во время использования модели.
   - ✓ Обслуживание модели, алгоритма, программы и их эксплуатация.
   - ✓ Расширение возможностей: включение новых функций или изменение режимов моделирования, в том числе и под модифицированную среду.
   - ✓ Нахождение, исправление скрытых ошибок в программе, если таковые найдутся.
10. Использование модели.

1.4. Задачи анализа, синтеза и оптимизации

Разработка и совершенствование современных инженерных методов и математических технологий проводится с целью повышения надежности, безопасности и экономической эффективности реальных объектов и процессов. Безусловно, это все очень вычислительно ресурсоемкие процессы, но их использование насущное требование современной практики.

Математические моделирование, решает задачи разработки и практического применения методов наиболее оптимального управления организационными системами. Как уже отмечалось, модель представляет собой «четырехместную конструкцию», компонентами которой являются субъект; задача, решаемая субъектом; объект-оригинал и язык описания или способ воспроизведения модели. Особую роль в структуре модели играет решаемая субъектом задача. Вне контекста задачи или класса задач понятие модели не имеет смысла. Каждому материальному объекту, вообще говоря, соответствует бесчисленное множество в равной мере адекватных, но различных по существу моделей, связанных с разными задачами. Паре задача-объект тоже соответствует множество моделей, содержащих в принципе одну и ту же информацию, но различающихся формами ее представления или воспроизведения.

Далеко не всегда вопрос о том, какого типа математическую задачу мы будем решать, даже какие величины мы будем искать, ясен с самого начала. Задача может быть поставлена не в конкретной форме («Найти частоту колебаний такой-то системы»), а в форме не столь определенной («Исследовать поведение такой-то системы», «Оптимизировать такое-то устройство путем подбора его параметров» и т. п.). Тогда требуется предварительное уточнение плана: какие величины было бы желательно найти, какие зависимости исследовать и откуда их можно было бы получить, по какому критерию проводить оптимизацию и т. д. Такой план, который впоследствии может видоизменяться и дополняться, желательно обдумать на возможно более ранней стадии исследования, поскольку он может существенно повлиять на формулировку математической модели: что считать исходными данными, какие величины искать, какого типа уравнения понадобятся для этого и т. д. При уточнении математической модели уточняется и план действий, в итоге четко формулируется математическая задача. Впрочем, бывает, что даже четко сформулированная задача видоизменяется в процессе дальнейшего исследования.

Прикладные математические задачи можно условно подразделить на два класса. В задачах одного класса речь идет об исследовании свойств заданного объекта – это *задачи анализа*. Задачи другого класса имеют целью выбор объекта из некоторой совокупности на основании каких-то требований – это *задачи синтеза*. Термин «задачи синтеза» применяется и в более специальном смысле; в частности, в теории систем управления он означает задачу о построении такой системы, имеющей предписанное функционирование на основе применения обратной связи. Конечно, это подразделение условно, так как многие задачи можно в равной мере отнести как к одному, так и к другому классу. Однако, из содержательной постановки задачи чаще всего бывает ясно, о задаче какого класса идет речь. Для задач анализа, математическая модель обычно сводится к уравнениям того или иного вида. Математическая модель задачи синтеза тоже может свестись к решению уравнений, если условия, на основании которых требуется выбрать объект, имеют вид некоторых равенств. Но часто условие выбора имеет другой характер: для выбираемого объекта некоторая заданная скалярная функция его параметров (целевая функция) должна принять наименьшее или наибольшее возможное значение. Тогда математическая модель сведется к задаче на экстремум (оптимум).

Оптимизация – в математике, информатике и исследовании операций задача нахождения экстремума (минимума или максимума) целевой функции в некоторой области конечномерного векторного пространства, ограниченной набором линейных и/или нелинейных равенств и/или неравенств. Теорию и методы решения задач оптимизации изучает *математическое программирование*. В процессе проектирования ставится обычно задача определения наилучших, в некотором смысле, структуры или значений параметров объектов. Такая задача называется оптимизационной.

Если оптимизация связана с расчетом оптимальных значений параметров при заданной структуре объекта, то она называется *параметрической оптимизацией*. Задача выбора оптимальной структуры является *структурной оптимизацией*.

Стандартная математическая задача оптимизации формулируется таким образом. Среди элементов χ, образующих множество X, найти такой элемент χ^*, который доставляет минимальное значение $f(\chi^*)$ заданной функции $f(\chi)$. Для того, чтобы корректно поставить задачу оптимизации, необходимо задать: допустимое множество; целевую функцию; критерий поиска.

Общая запись задач оптимизации задает большое разнообразие их классов. От класса задачи зависит подбор метода (эффективность ее решения). Классификацию задач определяют: целевая функция и допустимая область (задается системой неравенств и равенств или более сложным алгоритмом).

Локальные методы: сходятся к какому-нибудь локальному экстремуму целевой функции. В случае унимодальной целевой функции, этот экстремум единственен, и будет глобальным максимумом/минимумом.

Глобальные методы: имеют дело с многоэкстремальными целевыми функциями. При глобальном поиске основной задачей является выявление тенденций глобального поведения целевой функции.

Существующие в настоящее время методы поиска экстремума можно разбить на три большие группы:

✓ детерминированные;

✓ случайные (стохастические);

✓ комбинированные.

По критерию размерности допустимого множества методы оптимизации делят на методы *одномерной оптимизации* (золотого сечения, дихотомии, парабол, перебор по сетке, Фибоначчи и др.) и методы *многомерной оптимизации* (Нелдера-Мида, Хука-Дживса, конфигураций, Розенброка).

По виду целевой функции и допустимого множества задачи оптимизации и методы их решения можно разделить на следующие классы:

- ✓ если целевая функция и ограничения являются линейными функциями, такие задачи разрешаются методами *линейного программирования*;
- ✓ в противном случае имеют дело с задачей *нелинейного программирования* и применяют соответствующие методы.

По требованиям к гладкости и наличию у целевой функции частных производных их также можно разделить на:

- ✓ прямые методы, требующие только вычислений целевой функции в точках приближений;
- ✓ методы первого порядка, требующие вычисления первых частных производных функции;
- ✓ методы второго порядка, требующие вычисления вторых частных производных, то есть гессиана целевой функции.

Помимо того, оптимизационные методы делятся на следующие группы:

- ✓ аналитические методы (например, метод множителей Лагранжа и условия Куна-Таккера);
- ✓ численные методы;
- ✓ графические методы.

Как уже отмечено, способ нахождения экстремума полностью определяется классом задачи. Но перед тем, как оптимизировать математическую модель, нужно выполнить 4 шага.

Шаг 1. Определение границ системы оптимизации.

Шаг 2. Исключение тех связей объекта оптимизации с внешним миром, которые не могут сильно повлиять на результат оптимизации, а точнее тех, без которых решение упрощается.

Шаг 3. Выбор управляемых переменных. «Замораживаем» значения некоторых переменных (неуправляемые переменные). Другим «разрешаем» принимать любые значения из области допустимых решений (управляемые переменные). Определение ограничений на управляемые переменные (равенства и/или неравенства).

Шаг 4. Построение целевой функции. Выбор числового критерия оптимизации (например, показателя эффективности).

Характеризуя объект, сложно выбрать такой один критерий, который бы обеспечил всю полноту требований. А стремление к всеобъемлющему решению и назначение большого числа критериев сильно усложняет задачу. Поэтому в разных задачах количество критериев может быть различным. Задачи *однокритериальной оптимизации* иногда называют *скалярными*, а *многокритериальной – векторной оптимизацией*. Кроме того, количество параметров, характеризующих оптимизируемый объект (задачу), также может быть различным, причем параметры могут меняться непрерывно или дискретно.

В предельном случае решение практических задач можно свести к задаче двухкритериальной оптимизации, критериями в которой являются «цена» и «качество» (так называемые «цена-качество»). Это наглядно позволяет учесть и экономические (цена), и производственно-технические (качество продукции) требования. Сведение задачи к однокритериальной требует введения существенных допущений, но облегчает окончательный выбор.

Простейшим примером технико-экономической оптимизационной задачи может быть выбор диаметра трубопровода, по которому насосом перекачивается жидкость. При уменьшении диаметра трубы снижается ее стоимость, но увеличиваются затраты энергии на перекачку жидкости из-за возросшего гидравлического сопротивления. Примером задачи двухпараметрической оптимизации будет задача выбора диаметра трубопровода с горячей жидкостью или паром, так как одновременно выбирается диаметр трубопровода и толщина тепловой изоляции при постоянстве остальных. При этом оба параметра дискретны, так как существуют как сортамент труб, так и типовые параметры готовых теплоизоляционных сегментов. Оптимизации подлежат параметры многих технологических процессов, объемы производства предприятий, уровни надежности продукции и др.

Следует подчеркнуть, что оптимизация в отличие от обычного сравнения вариантов предполагает рассмотрение всех решений, попадающих в область допустимых значений параметров. Те решения, в процессе поиска которых не проводился полный просмотр возможных вариантов, обычно называют «рациональными».

Правильный выбор критериев играет существенную роль в выборе оптимального решения. В теории принятия решений не найдено общего метода выбора критериев оптимальности. В основном руководствуются опытом или рекомендациями. Наиболее изучен вопрос для финансово-экономических задач, в которых зачастую применяется единственный критерий – максимум показателя эффективности, прибыли, максимум рентабельности, либо минимум срока окупаемости и т. п. Применение для технических задач только одного критерия (например, максимум уровня безопасности, минимум потребления энергии, минимум экологического ущерба) часто приводит к абсурдным результатам, выходящим за область допустимых решений, поэтому обычно сочетается с экономическими критериями (например, минимум стоимости или максимум дохода). Большие сложности вызывают «неисчисляемые» критерии оптимальности, которые касаются, например, гуманитарных вопросов, художественного впечатления, изменения ландшафта и т. п. (например, максимум удобства, красоты). Для учета таких критериев могут применяться экспертные оценки.

Наиболее разработаны методы однокритериальной оптимизации, в большинстве случаев позволяющие получить однозначное решение. В задачах многокритериальной оптимизации абсолютное решение выбрать невозможно (за исключением частных случаев), так как при переходе от одного варианта к другому, как правило, улучшаются значения одних критериев, но ухудшаются значения других. Состав таких критериев называется противоречивым, и окончательно выбранное решение всегда будет компромиссным. Компромисс разрешается введением тех или иных дополнительных ограничений или субъективных предположений. Поэтому невозможно говорить об объективном единственном решении такой задачи.

Часто многокритериальную задачу сводят к однокритериальной применением «свертки» критериев в один комплексный, называемый функцией полезности. Например, в конкурсных процедурах выбора подрядчиков и поставщиков целевая функция рассчитывается на основе балльных критериев. В ряде случаев успешно применяются ранжирование и последовательное применение критериев оптимальности, метод анализа иерархий. Иногда общим методом для многокритериальных задач называют оптимальность по Парето, которое позволяет найти ряд «неулучшаемых» решений, однако этот метод не гарантирует глобальной оптимальности решений. Часто моделируемый объект сложен и расшифровка механизма его функционирования может оказаться трудоемкой и длинной во времени. В этом случае поступают следующим образом: на оригинале проводят эксперименты, обрабатывают полученные результаты и, не вникая в механизм и теорию моделируемого объекта с помощью методов математической статистики и теории вероятности, устанавливают связи между переменными, описывающими объект. Частным, но весьма важным для развитых в теоретическом отношении научных и технических дисциплин является случай, когда роль объекта-моделирования в исследовательской или прикладной задаче играет не фрагмент реального мира, рассматриваемый непосредственно, а некий идеальный конструкт, т. е. по сути дела другая модель, созданная ранее и практически достоверная. Подобное вторичное, а в общем случае n-кратное моделирование может осуществляться теоретическими методами с последующей проверкой получаемых результатов по экспериментальным данным, что характерно для фундаментальных естественных наук. В менее развитых в теоретическом отношении областях знания (биология, некоторые технические дисциплины) вторичная модель обычно включает в себя эмпирическую информацию, которую не охватывают существующие теории.

В заключении рассмотрим укрупненные этапы моделирования некоторого производства.

Этап 1. Содержательная постановка задачи.

Современное производство характерно тем, что часть производимой продукции (в стоимостном выражении) возвращается в виде инвестиций (т. е. части конечной продукции, используемой для создания основных фондов производства) в производство. При этом время возврата, ввода в оборот новых фондов может быть различным для различного рода производства. Необходимо промоделировать эту ситуацию и выявить динамику изменения величины основных фондов производства (капитала). Сложность и многообразие, слабая структурированность и плохая формализуемость основных экономических механизмов, определяющих работу предприятий, не позволяют преобразовать процедуры принятия решений в экономической системе в полностью эффективные математические модели и алгоритмы прогнозирования. Поэтому целесообразно использование простых, но гибких и надежных процедур принятия решения.

Этап 2. Формулировка гипотез, построение, исследование модели.

Динамика изменения величины капитала определяется в модели, в основном, простыми процессами производства и описывается так называемыми обобщенными коэффициентами амортизации (расхода фондов) и потока инвестиций (часть конечного продукта, используемого в единицу времени для создания основных фондов). Эти коэффициенты –

относительные величины (оцениваются за единицу времени). Необходимо разработать и исследовать модель динамики основных фондов. Считаем при этом допустимость определенных гипотез, определяющих систему производства. Пусть $x(t)$ – величина основных фондов (капитала) в момент времени t, где $0 \le t \le N$. Через промежуток времени Δt она будет равна $x(t+\Delta t)$. Абсолютный прирост составит $\Delta x = x(t+\Delta t) - x(t)$. Относительный прирост будет равен $\delta x = [x(t+\Delta t) - x(t)]/\Delta t$.

Сформулируем гипотезы:

1. Социально-экономические условия производства достаточно хорошие и способствуют росту производства, а поток инвестиций задается в виде известной функции $y(t)$.

2. Коэффициент амортизации фондов считается неизменным и равным m, и при достаточно малом значении Δt, изменение основных фондов прямо пропорционально текущей величине капитала, т. е. $dx = (y(t) - mx(t))\Delta t$.

Считая что $\Delta t \to 0$, а также учитывая определение производной, получим из предыдущего соотношения следующее математическое выражение закона изменения величины капитала – математическую модель (дифференциальное уравнение) динамики капитала: $x'(t) = y(t) - mx(t)$, при $x(0) = x_0$, где $x(0)$ – начальное значение капитала в момент времени $t = 0$.

Полученная простейшая модель не отражает следующего важного факта – социально-экономические ресурсы производства таковы, что между выделением инвестиций и их введением и использованием в выпуске новой продукции проходит некоторое время (T), называемое – лаг. Учитывая это, запишем модель в ином виде:

$$x'(t) = y(t-T) - mx(t); \quad x(0) = x_0.$$

Этой непрерывной, дифференциальной, динамической модели можно поставить в соответствие простую дискретную модель:

$$x_{i+1} = x_i + y_j - mx_i, \text{ при } x_0 = c,\ i = 0, \ldots, n,\ 0 < j < n.$$

где n – предельное значение момента времени при моделировании.

Дискретная модель следует из непрерывной для $\Delta t = 1$, при замене производной $x'(t)$ на относительное приращение при малых значениях Δt.

Этап 3. Построение алгоритма и программы моделирования.

Определим режим моделирования, когда значения m и c – известны и постоянны; y – увеличивается в каждый следующий момент времени на 1%, а также рассмотрим наиболее простой алгоритм моделирования в укрупненных шагах.

Шаг 1. Ввод начальных данных моделирования: $c = x(0)$ – начальный капитал; n – конечное время моделирования; m – коэффициент амортизации; s – единица измерения времени; y – объем инвестиций.

Шаг 2. Вычисление x_i от $i = 1$ до $i = n$ по рекуррентной формуле, приведенной выше (этап 2).

Шаг 3. Поиск стационарного состояния – такого момента времени j, $0 \le j \le n$, начиная с которого все $x_j, x_{j+1}, \ldots, x_n$ постоянны или изменяются на малую допустимую величину $\varepsilon > 0$.

Шаг 4. Выдача результатов моделирования и, по желанию пользователя, в графическом или табличном виде.

Этап 4. Постановка и проведение вычислительных экспериментов.

Эксперимент 1. Поток инвестиций – постоянный и в каждый момент времени равен 10 000. В начальный момент капитал – 1 000 000 руб. Коэффициент амортизации – 0.0025. Найти величину основных фондов через 20 суток, если лаг равен 5 суток.

Эксперимент 2. Основные фонды в момент времени $t = 0$ была равны 5 000. Через какое время общая их сумма превысит 120 000 руб., если поток инвестиций постоянный и равен 200, а $m = 0.02$, $T = 3$?

Эксперимент 3. Какую стратегию инвестиций лучше использовать, если величина инвестиций постоянная, в начальный момент капитал равен 100 000, величина амортизации постоянная?

Этап 5. Модификация (развитие) модели.

Модификация 1. Коэффициент амортизации можно взять в форме:

$$m = r - sx(t),$$

где r – коэффициент обновления фондов; s – коэффициент устаревания фондов, причем $r \geq 0, s \leq 1$. При этом модель примет вид:

$$x'(t) = y(t-T) - rx(t) + sx^2(t), x(0) = x_0.$$

Этой непрерывной, дифференциальной, динамической модели можно поставить в соответствие дискретную модель:

$$x_{i+1} = x_i + y_j - rx_i + sx_i^2, \text{ при } x_0 = c,\ i = 0,\dots,n,\ 0 < j < n,$$

где n – предельное значение момента времени при моделировании. Далее, следует исследовать непрерывную и дискретную модели.

Модификация 2. Одна из моделей математической экономики задается уравнением:

$$\frac{dz}{dt} = \left(\left(1-c\right)\cdot z(t) + k(t-T) + a\right)\cdot l,$$

где $z(t)$ – функция, которая характеризует выпуск продукции; k – коэффициент капиталовложений; a – независимые расходы производства; l – скорость реакции выпуска на капиталовложения; c – постоянная спроса; T – запаздывание (лаг).

Далее исследовать непрерывную и дискретную модели.

Модификация 3. Для модели динамики фондов с переменным законом потока инвестиций:

- ✓ построить гипотезы, модель и алгоритм моделирования;
- ✓ сформулировать планы вычислительных экспериментов;
- ✓ реализовать алгоритм и планы экспериментов на ЭВМ.

Полученные результаты передаются лицу, принимающему решения для окончательного выбора.

Математическое и компьютерное моделирование сегодня широко ставится на технологическую основу, в связи с этим необходимо отметить особую роль имитационного моделирования. Компьютерное моделирование (получение, накопление, переработка, хранение, использование, актуализация знаний с помощью ЭВМ), в отличие от математического, используется сравнительно недавно, хотя эти технологии моделирования тесно связаны. Компьютерное моделирование, как правило, применяется тогда, когда не удается построить математической аналитической модели или же такая модель трудоемка для исследования. Примером простейшей компьютерной модели может служить модель броуновского движения, получаемая генерацией компьютером нового случайного положения точки на экране и траектории ее движения; при этом отметим, что и сам «датчик случайных чисел компьютера» – это также компьютерная модель, соответствующая математической модели распределения случайной величины (обычно нормального распределения). Это распределение – псевдослучайное, получаемое по вполне детерминированному алгоритму.

Прогресс математического моделирования связан с разработкой систем компьютерного моделирования, которые поддерживает весь жизненный цикл модели, а прогресс в информационной технологии – с актуализацией опыта моделирования на компьютере, с созданием банков моделей, методов и программных систем, позволяющих собирать новые модели из моделей банка.

Вопросы самоконтроля:

1. Общее понятие модели.
2. Фундаментальное свойство моделей.
3. Задачи моделирования.
4. Виды моделей.
5. Что означает термин «моделирование».
6. Когнитивные модели и их подвиды.
7. Концептуальное моделирование.
8. Виды классификаций моделей.
9. Классификация по форме представления.
10. Этапы моделирования.
11. Цели моделирования.
12. Общие требования, предъявляемые к моделям.
13. Понятие математического моделирования.
14. Компьютерное моделирование.
15. Свойства математических моделей.
16. Дескриптивные модели.
17. Оптимизационные модели.
18. Многокритериальные модели.
19. Детерминированные модели.
20. Вероятностные модели.
21. Статические модели.
22. Отличие линейных и нелинейных моделей.
23. Способы построения и исследования математических моделей.
24. Понятие линеаризации.
25. Задачи идентификации моделей.
26. Методические погрешности, возникающие в процессе моделирования.
27. Измерение, как метод получения количественной информации об объекте.
28. Принципы измерений.
29. Задачи и принципы компьютерного моделирования.
30. Вычислительный эксперимент.
31. Этапы компьютерного моделирования.
32. Постановка задач анализа и синтеза.
33. Понятие оптимизации в моделировании.
34. Параметрическая и структурная оптимизация.
35. Экстремальные задачи математического программирования.
36. .Локальные и глобальные методы исследования целевой функции.
37. Понятие целевой функции.
38. Классификация оптимизационных задач.
39. Последовательность шагов при оптимизации математической модели.
40. Понятия однокритериальной и векторной оптимизации.
41. Критерий оптимальности.
42. Пример содержательных этапов моделирования реальных объектов.

Раздел 2. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

2.1. Основные виды задач линейного программирования

Математические методы и модели применяются также и для отыскания оптимальных решений в управлении. Всякое управление предполагает наличие цели, ради реализации которой оно осуществляется. Появление достаточно мощной вычислительной техники, способной за короткое время переработать значительные массивы информации, а также развитие теории математических моделей способствовало преобразованию искусства принятия решения в науку, доступную каждому, кто овладел ее принципами и методологией.

Среди методов математического программирования наибольшее распространение получили методы линейного программирования. Термин «программирование» показывает, что они применяются для составления плана (программы), который обеспечивал бы оптимальное использование материальных и трудовых ресурсов. Слово линейное определяет математическую природу этих моделей. Линейность связана с понятиями пропорциональности и возможностью суммирования результатов (аддитивностью). Она отражается в том, что условия задач выражаются системой линейных уравнений или неравенств, содержащих неизвестные только первой степени. Для любых задач линейного программирования характерны три следующих условия (по академику В.С. Немчинову):

- ✓ наличие системы взаимосвязанных факторов;
- ✓ строгое определение критерия оценки оптимальности;
- ✓ точная формулировка условий, ограничивающих использование наличных ресурсов.

Общая схема построения моделей задач принятия решений:

1. Выделяют параметры, которые описывают процесс принятия решения. Выбор тех или иных численных значений для параметров набора эквивалентен принятию того или иного решения. Эти параметры принято называть параметрами управления, число их может быть весьма значительным.

2. Выделяют и четко формулируют цель, ради достижения которой принимается то или иное решение. Как правило, целевую установку процесса принятия решения представляют в виде некоторой зависящей от параметров управления функции, значения которой дают оценку качества принятого решения. Эту функцию называют *целевой функцией* задачи или показателем качества ее решения. Целевая функция дает возможность сравнивать два различных решения: если при одном наборе параметров значение целевой функции меньше, чем при другом (или больше, в зависимости от задачи), то одно решение будет предпочтительнее другого.

3. Поскольку в задачах принятия решений не любой набор параметров управления может быть реализован практически, важно выделить среди всех решений множество возможных решений. Среди них существуют, достаточно типичные, группы:

✓ материальное обеспечение производства (построение оптимального плана поставок материалов);

✓ организация смежных работ:

✓ распределение поставок от m изготовителей к n потребителям, а при отсутствии равенства объемов – перемещение из резервов;

✓ комплектование специализированных подразделений из m видов работ и n видов техники с целью обеспечения максимальной производительности и темпа выполняемых работ.

Кроме того существует большой класс сетевых задач, позволяющих проектировать рациональные сети информационных коммуникаций.

При построении модели важно уменьшить «размерность» решаемой задачи. Поэтому основным элементом является выделение доминирующих переменных, параметров и ограничений.

Допустим, что предметом моделирования является процесс поставки изделий от нескольких промышленных предприятий. На предварительном этапе моделирования выявлен ряд факторов, влияющих на конечные экономические результаты:

- ✓ стоимость (отпускная цена);
- ✓ стоимость транспортировки, которая зависит от расстояния перевозки и принятой транспортной схемы;
- ✓ возможные потери в процессе доставки при переходе с одного вида транспорта на другой;
- ✓ степень экономической стабильности поставщиков и их дисциплинированность при выполнении договорных обязательств.

Назовем процесс поставки – конечной продукцией моделирования. Результат по конечной продукции осуществляется с помощью одного из выбранных способов поставок.

Сформулируем условие задачи. Необходимо составить план, при котором поставки в пределах имеющихся исходных запасов доставляли бы минимум функции затрат.

Формализуем условия задачи. Пусть n – число поставщиков; где j-й завод $(j=\overline{1,n})$ участник поставок на i-й объект производства $(i=\overline{1,m})$; m – перечень объектов, на которые необходимо выполнить поставки; a_j – запасы материалов на j-м заводе-производителе; c_{ij} – стоимость j-го ресурса при доставки его на i-й объект; b_i – потребность i-го объекта.

Составим математическую модель задачи:

1. Выбираем параметры управления (то, что можно менять в процессе производства в рамках поставленной задачи). Отметим, что в рамках поставленной задачи нельзя менять b_i, c_{ij}, a_j, m и n (см. выше). В процессе

производства можно менять только транспортную схему поставок на объекты, поэтому пусть x_j – объемы поставок на объекты. Следовательно, $x_1, x_2, ..., x_n$ – искомый набор параметров управления, а составленный план производства эквивалентен выбору конкретных значений параметров x_j.

2. Определяем целевую функцию. По условию задачи необходимо составить такую транспортную схему, чтобы прибыль компании была бы наибольшей. Следовательно, целевая функция должна отражать поставки и их стоимостные показатели:

$$f(x) = \sum_{j=1}^{n} c_j x_j = c_1 x_1 + c_2 x_2 + ... + c_n x_n .$$

Для нахождения оптимального решения необходимо максимизировать целевую функцию, то есть $f(x) \to max$.

3. Поскольку запасы поставок ограничены, то на параметры управления накладываются ограничения. Объемы поставок не должны превышать имеющихся запасов. Необходимо добавить еще ограничения $x_j \geq 0$, так как время поставок ресурсов не может быть отрицательным.

В результате получим математическую модель *общей задачи* линейного программирования задачи:

$$f(x) = \sum_{j=1}^{n} c_j x_j \to max;$$

$$\begin{cases} \sum_{j=1}^{n} a_{1j} x_j \leq b_1, \ \sum_{j=1}^{n} a_{2j} x_j \leq b_2, \ ..., \ \sum_{j=1}^{n} a_{mj} x_j \leq b_m; \\ x_j \geq 0, \ j = \overline{1, n}. \end{cases}$$

Из структуры модели видно, что целевая функция и ограничения линейно зависят от параметров управления. Подобные модели принято называть *задачами линейного программирования* (ЗЛП). Реально зависимости редко бывают строго линейными, однако предположение о линейности дает возможность разрабатывать эффективные методы, использование которых для ряда задач является вполне оправданным. Линейные модели могут содержать в качестве ограничений не только неравенства, но и уравнения.

Стандартная ЗЛП – определение максимального значения целевой функции: $f(x)=\sum_{j=1}^{n} c_j x_j \to max$, при ограничениях в виде неравенств:

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j \le b_i, & i=\overline{1,k}; \\ \sum_{j=1}^{n} a_{ij} x_j \ge b_i, & i=\overline{k+1,m}; \\ x_j \ge 0, & j=\overline{1,n}. \end{cases}$$

Каноническая ЗЛП – определение максимального значения целевой функции: $f(x)=\sum_{j=1}^{n} c_j x_j \to max$ при ограничениях в виде равенств:

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j = b_i, & i=\overline{1,m}; \\ x_j \ge 0, & j=\overline{1,n}. \end{cases}$$

Совокупность значений переменных $x_1, x_2, ..., x_n$, удовлетворяющих всем линейным ограничениям, называется допустимым (опорным) планом.

Переход от стандартной модели ЗЛП к канонической осуществляется добавлением новых неотрицательных переменных со знаком «+» для неравенств типа $\le$ и со знаком «–» для неравенств типа $\ge$. Само неравенство заменяется на равенство.

Если в ЗЛП на некоторую переменную x_k не накладывается условие неотрицательности, то добавляют две неотрицательные переменные и делают замену переменных $x_k = x_{k_2} - x_{k_1}$, $x_{k_2} \ge 0$, $x_{k_1} \ge 0$.

При переходе от задачи на минимум к задаче на максимум изменяется знак целевой функции.

Задача линейного программирования может принимать оптимальное значение в нескольких точках, такие решения называются *альтернативными* *оптимальными решениями* или альтернативным оптимумом, причем в каждой из этих точек целевая функция имеет одно и то же оптимальное значение.

Если в ЗЛП с двумя параметрами управления нормальный вектор целевой функции параллелен нормальному вектору прямой одного из ограничений, то ЗЛП может иметь множество оптимальных решений, например:

$$f(x) = 2x_1 + 3x_2 \to max;$$

$$\begin{cases} 4x_1 + 6x_2 \le 15; \\ 7x_1 + 5x_2 \le 20; \\ x_1 \ge 0, x_2 \ge 0. \end{cases}$$

Вектор целевой функции $\overrightarrow{N}(2;3)$ параллелен нормальному вектору $\overrightarrow{N_1}(4;6)$ прямой первого ограничения: $\overrightarrow{N_1} = 2\overrightarrow{N}$. В этом случае возможно множество оптимальных решений.

Если в ЗЛП с тремя параметрами управления нормальный вектор целевой функции параллелен нормальному вектору плоскости одного из ограничений или перпендикулярен к прямой пересечения двух плоскостей ограничений, то ЗЛП также может иметь множество оптимальных решений.

Для задач с большим количеством параметров управления сложно определить, имеется ли множество оптимумов по виду составленной математической модели.

Пример. Составить математическую модель задачи, привести ее к каноническому виду.

При проектировании региональной транспортной сети на автомобильной дороге требуется построить мостовые конструкции по одному из трех проектов (МКА). Расходы, связанные с каждой конструкцией, среди прочих, определяются количеством трудозатрат (человеко-часов), электроэнергии (киловатт-часов) и объемами потребляемого материала (т).

Определить расходы на производство конструкций мостов, обеспечивающий наибольший доход фирме-производителю.

Расходы ресурсов:	На единицу продукции			Имеется в наличии
	МКА1	МКА2	МКА3	
человеко-часов	10	15	10	320
киловатт-часов	18	16	20	500
материала, т	70	50	60	2 000
Доход с единицы продукции, млн. руб.	50	60	70	*max*

Решение.

1. В процессе производства можно менять количество выпускаемой продукции по определенному проекту МКА. Поэтому в качестве параметров управления x_1, x_2 и x_3 выберем количество выпускаемой продукции по соответствующему технологическому проекту МКА1, МКА2 и МКА3.

2. Целевая функция должна максимизировать доход от выпускаемой продукции: $f(x) = 50x_1 + 60x_2 + 70x_3 \to max$.

3. Запасы ресурсов ограничивают количество выпускаемой продукции. В процессе производства нельзя превысить количество имеющихся человеко-часов: $10x_1 + 15x_2 + 10x_3 \le 320$. Аналогично, составим условия для двух других ресурсов: $18x_1 + 16x_2 + 20x_3 \le 500$, $70x_1 + 50x_2 + 60x_3 \le 2000$. Необходимо добавить условие неотрицательности переменных: $x_1 \ge 0,\ \ x_2 \ge 0,\ \ x_3 \ge 0$.

В результате математическая модель принимает вид:

$$f(x) = 50x_1 + 60x_2 + 70x_3 \to max;$$

$$\begin{cases} 10x_1 + 15x_2 + 10x_3 \le 320; \\ 18x_1 + 16x_2 + 20x_3 \le 500; \\ 70x_1 + 50x_2 + 60x_3 \le 2000; \\ x_j \ge 0, \quad j = \overline{1,3}. \end{cases}$$

Это стандартная ЗЛП, ее решение: $f(x) = 1\,790$ руб., $x_1 = 0$, $x_2 = 10$, $x_3 = 17$.

Построим каноническую ЗЛП, добавив в первое, второе и третье неравенства неотрицательные переменные x_4, x_5, x_6:

$$f(x) = 50x_1 + 60x_2 + 80x_3 \to max;$$

$$\begin{cases} 10x_1 + 15x_2 + 10x_3 + x_4 = 320; \\ 18x_1 + 16x_2 + 20x_3 + x_5 = 500; \\ 70x_1 + 50x_2 + 60x_3 + x_6 = 2\,000; \\ x_j \ge 0, \quad j = \overline{1,6}. \end{cases}$$

Заметим, что стоимостные критерии не всегда являются доминирующими. При восстановлении дорог в районах чрезвычайных ситуаций на первый план может выйти критерий «времени восстановления прерванного движения». В этом случае, целевая функция примет вид:

$$L = \sum_{i}^{n} \sum_{j}^{m} t_{ij} x_{ij} \to min.$$

Математический аппарат линейного программирования позволяет не только получить оптимальный план, но и сделать ряд экономически значимых выводов, основанных на свойствах задачи, двойственной к исходной (прямой).

Двойственная задача по отношению к исходной задаче составляется по следующим правилам:

✓ в исходной задаче находится максимум целевой функции, а в двойственной – минимум;

✓ количество неизвестных двойственной задачи равно количеству ограничений исходной задачи;

✓ количество ограничений двойственной задачи равно количеству переменных исходной задачи;

✓ ограничению исходной задачи вида «$\ge$» соответствует неотрицательная переменная двойственной задачи, ограничению исходной задачи вида «$\le$» соответствует неположительная переменная двойственной задачи; ограничению исходной задачи вида «$=$» соответствует переменная двойственной задачи, которая может принимать как положительные, так и отрицательные значения;

✓ если на переменную исходной задачи наложено (не наложено) условие неотрицательности, то соответствующим ограничением двойственной задачи будет неравенство вида «$\geq$» (или уравнение);

✓ коэффициенты при неизвестных в целевой функции исходной задачи являются свободными членами в ограничениях двойственной задачи;

✓ свободные члены ограничений исходной задачи являются коэффициентами при неизвестных в целевой функции двойственной задачи;

✓ матрица, составленная из коэффициентов при неизвестных в системе ограничений исходной задачи, в двойственной задаче транспонируется.

Пример. Составить двойственную задачу для ЗЛП:

$$f(x) = x_1 + 2x_2 + 3x_3 \to max;$$

$$\begin{cases} x_1 + 5x_2 + 2x_3 = 70; \\ x_1 + x_2 - 4x_3 \leq 30; \\ x_1 - 2x_2 + x_3 \geq 40; \\ x_1 \geq 0, \quad x_2 \geq 0. \end{cases}$$

Решение. Исходная задача сформулирована на максимум, значит двойственная задача формулируется на минимум. Исходная задача имеет три ограничения, двойственная задача будет иметь три переменные y_1, y_2 и y_3. Второе ограничение исходной задачи имеет вид «$\leq$», тогда y_2 будет неотрицательным. Третье ограничение исходной задачи имеет вид «$\geq$», тогда y_3 будет неположительным. В исходной задаче $x_1 \geq 0$ и $x_2 \geq 0$, в двойственной задаче первые два ограничения будут записаны в виде неравенств типа «$\geq$». Коэффициенты при переменных целевой функции исходной задачи будут свободными членами ограничений двойственной задачи. Матрица коэффициентов при неизвестных в ограничениях исходной задачи транспонируется в двойственной задаче. Коэффициенты при переменных целевой функции двойственной задачи являются свободными членами ограничений исходной задачи.

Получаем двойственную задачу:

$$g(y) = 70y_1 + 30y_2 + 40y_3 \to min;$$

$$\begin{cases} y_1 + y_2 + y_3 \ge 1; \\ 5y_1 + y_2 - 2y_3 \ge 2; \\ 2y_1 - 4y_2 + y_3 = 3; \\ y_1 \ge 0, \quad y_2 \le 0. \end{cases}$$

Решение исходной задачи: $f(x) = 100$, $x_1 = 10$, $x_2 = 0$, $x_3 = 30$.

Решение двойственной задачи: $g(y) = 100$, $y_1 = 2$, $y_2 = 0$, $y_3 = -1$.

Экономическая интерпретация двойственной задачи заключается в определении оптимальных оценок y_i, называемых неявными или «теневыми» ценами каждого вида ресурса с целью минимизации общей ценности (стоимости) всех ресурсов. Значения переменных y_i в оптимальном решении двойственной задачи представляют собой оценки влияния свободных членов b_i системы ограничений исходной задачи на величину приращения целевой функции $\Delta f(x) = \Delta b_i y_i$.

Рассмотрим следующую производственную задачу.

Предприятие после выпуска основной продукции имеет излишки ресурсов двух типов: R_1 – 10 единиц, R_2 – 8 единиц. Существует два способа распорядиться этими ресурсами:

✓ организовать из них выпуск 3 новых видов продукции: P_1, P_2, P_3.

✓ продать их.

Рассмотрим оба способа и оценим целесообразность их реализации.

Исходные данные приведены в таблице:

Ресурсы	Расход ресурса на единицу продукции			Запас ресурсов
	P_1	P_2	P_3	
R_1	1	2	1	10
R_2	2	1	3	8
Удельная прибыль	\$6	\$4	\$4	

Согласно *первому способу*, надо составить такой план выпуска продукции, который максимизирует суммарную прибыль. Построим математическую модель этой задачи. Пусть x_j – план выпуска продукции P_j. Тогда целевая функция будет выглядеть следующим образом:

$$f(x) = 6x_1 + 4x_2 + 4x_3 \to max;$$

Ограничения по ресурсам:

$$x_1 + 2x_2 + 3x_3 \le 10;$$
$$2x_1 + x_2 + 3x_3 \le 8;$$
$$x_j \ge 0, \quad j = 1, 2, 3.$$

Получили стандартную задачу линейного программирования.

Рассмотрим *второй способ* использования ресурсов, а именно, их продажу. Интерес предприятия состоит в том, чтобы продать ресурсы по таким ценам, при которых доход от реализации ресурсов будет не меньше прибыли, которую можно получить от реализации продукции, изготовленной из этих ресурсов. В свою очередь покупатель заинтересован в приобретении ресурсов по таким ценам, при которых затраты на покупку будут минимальны.

Задача согласования цен на ресурсы, устраивающих обе стороны, может быть описана следующей математической моделью. Пусть y_1 – цена одной единицы ресурса R_1; y_2 – цена одной единицы ресурса R_2.

Интерес покупателя будет выражаться целевой функцией, равной суммарной стоимости приобретаемых ресурсов:

$$g(y) = 10y_1 + 8y_2 \to min.$$

Интерес продавца описываться ограничениями:

$$y_1 + 2y_2 \le 6;$$
$$2y_1 + y_2 \le 4;$$
$$y_1 + 3y_2 \le 4,$$

в которых левая часть означает стоимость ресурсов, затраченных на выпуск единицы соответствующей продукции, а правая – удельную прибыль от ее реализации. Дополняя естественные условия неотрицательности цен: y_1, y_2, y_3 $\ge$ 0, получаем *двойственную задачу* ЛП.

Таким образом, симметричной паре двойственных задач можно придать определенный экономический смысл.

Прямая задача. Определить такой план выпуска продукции $x = (x_1, x_2, \ldots, x_n)$, используя ограниченные запасы ресурсов, при котором прибыль от реализации продукции будет максимальной.

Двойственная задача. Установить такой набор цен ресурсов $y = (y_1, y_2, \ldots, y_m)$, при которых стоимость ресурсов, затраченных на выпуск единицы продукции будет не ниже прибыли от ее реализации, но при этом суммарная стоимость затрат будет минимальна.

Двойственные оценки выступают как инструмент балансировки и сопоставления затрат и результатов, они гарантируют рентабельность оптимального плана. Самую высокую ценность имеют те ресурсы, которые в наибольшей степени ограничивают выпуск продукции, прибыль предприятия и на увеличение которых предприятие согласно нести значительные расходы. Ресурс, который предприятие не использует полностью в оптимальном плане, получает нулевую оценку.

Если нужно использовать разнородные ресурсы, например, различные машины, материалы и т. д. для выполнения какой-либо работы, то применяется общий метод линейного программирования, который получил, в соответствии со своей математической основой, название симплекс-метода, предложенного американским ученым Дж. Данцигом.

2.2. Графический метод решения ЗЛП

Графические методы решения в линейном программировании более наглядны, но их применение ограниченно числом неизвестных переменных, которых должно быть не более трех. Рассмотрим методику графического решения основной ЗЛП.

Пример. Для изготовления изделий двух видов имеется 400 кг металла. На одно изделие первого вида расходуется 2 кг металла, а на изделие второго вида – 5 кг. Составить план производства, обеспечивающий наибольшую стоимость выпускаемых изделий, если отпускная цена одного изделия первого вида составляет 4 ед., а изделия второго вида – 3 ед., причем изделий каждого вида требуется изготовить не менее 50 и 20 штук соответственно.

Решение.

1. В качестве параметров управления x_1 и x_2 выберем соответственно количество изготовленных изделий первого и второго вида.

2. Целевая функция должна максимизировать прибыль, получаемую от реализации изделий, поэтому $f(x) = 4x_1 + 3x_2 \to max$.

3. Число выпускаемых изделий в первую очередь ограничено количеством металла (400 кг), которое не должно превысить в процессе производства: $2x_1 + 5x_2 \le 400$. Наличие плана также накладывает ограничение. Изделий первого и второго вида необходимо изготовить не менее 50 и 20 соответственно, т. е. $x_1 \ge 50$, $x_2 \ge 20$. Количество

изготовленных изделий принадлежит множеству неотрицательных целых чисел.

Необходимо добавить условие неотрицательности и целочисленности переменных: $x_{1,2} \in N_0$, где N_0 – множество целых неотрицательных чисел, хотя в данной задаче ограничение неотрицательности излишне, так как ограничения $x_1 \geq 50$, $x_2 \geq 20$ являются более строгими. В итоге получим математическую модель ЗЛП:

$$f(x) = 4x_1 + 3x_2 \to max;$$

$$\begin{cases} 2x_1 + 5x_2 \leq 400; \\ x_1 \geq 50; \\ x_2 \geq 20; \\ x_{1,2} \in N_0. \end{cases}$$

Аналитическое решение задачи: $f(x) = 660$ ед., $x_1 = 150$, $x_2 = 20$.

Получим решение этой задачи графическим методом.

Определим множество решения первого линейного неравенства $2x_1 + 5x_2 \leq 400$. На плоскости оно отображается полуплоскостью, ограниченной прямой $2x_1 + 5x_2 = 400$. Для построения прямой необходимо найти две точки, принадлежащие прямой. Обычно их находят путем последовательного обнуления каждой из переменных.

Если $x_1 = 0$, то $x_2 = 400/5 = 80$. Если $x_2 = 0$, то $x_1 = 400/2 = 200$. Часто именно точки пересечения прямой с координатными прямыми являются оптимальными решениями.

В системе координат Ox_1x_2 построим прямую по двум точкам (0; 80) и (200; 0). Построенная прямая делит координатную плоскость на две полуплоскости. Для определения искомой полуплоскости достаточно выбрать одну точку, не лежащую на прямой. Пусть это будет начало координат (0; 0). Подставив координаты выбранной точки в неравенство, получим: $0 \leq 400$. Неравенство верное, следовательно, нижняя полуплоскость, которой принадлежит начало координат, является искомой. На рисунке 2.1 *а* эта полуплоскость заштрихована.

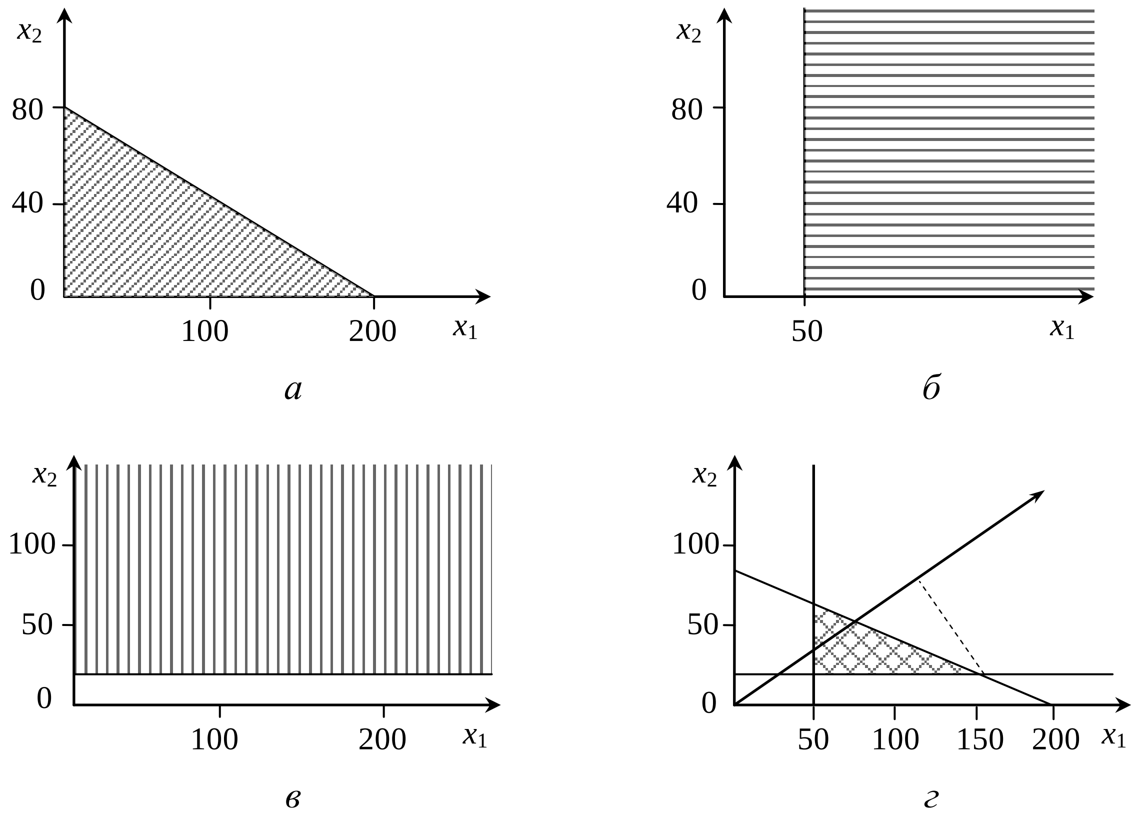


Рис. 2.1. Области допустимых решений ЗЛП

Аналогичная процедура выполняется для второго неравенства. Выберем точку (0; 0) и, подставив во второе неравенство, получим: $0 \geq 50$. Неравенство неверное, следовательно, выбираем правую полуплоскость, которой не принадлежит начало координат (Рис. 2.1 *б*).

Для третьего неравенства выбираем точку (0; 0) и, подставив в третье неравенство, получим: $0 \geq 20$. Неравенство неверное, следовательно, выбираем верхнюю полуплоскость, которой не принадлежит начало координат (Рис. 2.1 *в*).

В итоге выбираем общую заштрихованную область, точки которой удовлетворяют всем неравенствам задачи (Рис. 2.1 *г*). Область допустимых решений, или область Парето, изображается в виде замкнутого или незамкнутого многоугольника.

Оптимальное решение ЗЛП находится в одной из вершин многоугольника области Парето.

Возможны два пути нахождения оптимального решения графическим способом:

✓ находят координаты всех вершин многоугольника области Парето и подставляют их в целевую функцию, после чего выбирают оптимальное значение (наибольшее или наименьшее);

✓ строят нормальный вектор прямой целевой функции, проецируют область Парето на вектор или его продолжение (см. Рис. 2 1 *г*). Наиболее или наименее удаленная от начала координат проекция вершины области Парето на вектор или его продолжение будет оптимальным решением.

Для нахождения трех вершин необходимо решить три системы уравнений:

$$\begin{cases} x_1 = 50; \\ x_2 = 20. \end{cases} \quad \begin{cases} x_1 = 50; \\ 2x_1 + 5x_2 = 400. \end{cases} \quad \begin{cases} x_2 = 50; \\ 2x_1 + 5x_2 = 400. \end{cases}$$

В результате получим точки: $A(50; 20)$, $B(50; 60)$, $C(150; 20)$.

Находим значения целевой функции в вершинах многоугольника Парето:

$$F(50, 20) = 260,\ F(50, 60) = 380,\ F(150, 20) = 660.$$

Наибольшая стоимость выпускаемой продукции составит 660 ед. при изготовлении 150 изделий первого и 20 изделий второго вида.

При графическом решении *вторым способом* находим координаты только одной вершины $C(150; 20)$, поскольку проекция этой вершины на продолжение вектора целевой функции $\overline{N}(4; 3)$ будет наибольшей (см. Рис. 2.1 *г*). Второй способ применяется при использовании одинаковых масштабов на осях координат. В противном случае линия проекции на вектор целевой функции визуально не будет составлять 90^0.

Если не удается наглядно изобразить область Парето в системе координат с одинаковым масштабом осей координат, то поступают следующим способом. В системе координат с разным масштабом координатных осей рисуют линию уровня целевой функции $f(x)=0$ или $f(x)=const$ и перемещают ее параллельно в направлении вектора целевой функции до пересечения с областью Парето.

В процессе решения задачи использовано понятие «*область и оптимальность по Парето*». Рассмотрим его подробнее.

Пусть функции $f_1, f_2,..., f_K$ достигают максимум в одной и той же точке $X^* \in D$, в таком случае говорят, что задача имеет идеальное решение. Случаи существования идеального решения в многокритериальной задаче крайне редки. Поэтому основная проблема при рассмотрении таких задач – формализация *принципа оптимальности*, т. е. определение того, в каком смысле «оптимальное» решение лучше других. В случае отсутствия «идеального решения» в задаче *ищется компромиссное решение*.

Для всякой альтернативы $X \in D$ вектор из значений целевых функций $(f_1(X), f_2(X),..., f_K(X))$ является *векторной оценкой* альтернативы X. Векторная оценка альтернативы содержит полную информацию о ценности (полезности) этой альтернативы для главного конструктора системы, или, как принято говорить в системном анализе, лица, принимающего решение (ЛПР). Сравнение любых двух исходов заменяется сравнением их векторных оценок.

Пусть $X_1, X_2 \in D$. Если для всех критериев $f_1, f_2,..., f_K$ имеют место неравенства $f_k(X_2) \geq f_k(X_1)$, $k=1, 2,...,K$, причем хотя бы одно неравенство строгое, то говорят, что решение X_2 *предпочтительнее* решения X_1. Условие предпочтительности принято обозначать в виде $X_2 > X_1$.

В задаче точка $X_0 \in D$ называется *оптимальной по Парето*, если не существует другой точки $X \in D$, которая была бы предпочтительнее, чем X_0. Точки, оптимальные по Парето, образуют множество точек, оптимальных по Парето (множество неулучшаемых или эффективных точек) $D_p \subset D$.

Оптимальные решения многокритериальной задачи следует искать только среди элементов множества альтернатив D_p. В этой области ни один критерий не может быть улучшен без ухудшения хотя бы одного из других. Важным свойством множества Парето D_p является возможность «выбраковывать» из множества альтернатив D заведомо неудачные, уступающие другим по всем критериям. Обычно решение многокритериальной задачи должно начинаться с выделения множества D_p. При отсутствии дополнительной информации о системе предпочтений ЛПР должно принимать решение именно из множества Парето D_p.

В векторной оптимизации кроме множества Парето в общем случае нет общих правил, по которому варианту X_2 отдается предпочтение по сравнению с другим вариантом X_1. Часто решение многокритериальной задачи состоит в построении множества Парето-оптимальных точек и дальнейшем выборе одной из них на основе «здравого смысла» или с помощью какого-либо другого критерия.

Во всех случаях задача многокритериальной оптимизации одним из способом сводится к задаче с одним критерием. Существует много способов построения такого окончательного критерия, однако ни одному из них нельзя заранее отдать наибольшее предпочтение. Заметим, что целевые функции отображают множество точек, оптимальных по Парето $D_p \subset D \subset R^n$ в множество $F_p \subset F \subset R^K$, которое называется *множеством Парето.*

П р и м е р. Фирма «Папа Карло» для производства офисной мебели двух видов *A* и *B* использует блоки деревянных и металлических конструкций.

Затраты времени на производство каждого вида изделий, общий фонд рабочего времени, а также срок службы эксплуатации каждого вида изделия составляют:

Тип мебельной конструкции:	Затраты времени на производство мебели вида		Общий фонд рабочего времени, ч
	A	*B*	
деревянные блоки	2	5	200
металлические блоки	4	2	160
Срок службы, ед.	10	24	*max*

Требуется определить, какие мебельные конструкции каждого вида следует изготавливать, чтобы последующий их срок службы был максимальным.

Решение.

1. В качестве параметров управления x_1 и x_2 выберем количество производимой мебельной продукции первого и второго вида соответственно.

2. Целевая функция должна максимизировать срок службы изделий, поэтому $f(x) = 10x_1 + 24x_2 \to max$.

3. Количество используемых конструкций ограничено общим фондом рабочего времени. В процессе производства нельзя превысить общий фонд рабочего времени при изготовлении конструкций из дерева $2x_1 + 5x_2 \le 200$ и общий фонд рабочего времени для металлических конструкций $4x_1 + 2x_2 \le 160$. Необходимо добавить условие неотрицательности и целочисленности переменных: $x_{1,2} \in N_0$, где N_0 – множество целых неотрицательных чисел.

Получим математическую модель задачи:

$$f(x) = 10x_1 + 24x_2 \to max;$$

$$\begin{cases} 2x_1 + 5x_2 \le 200, & (1); \\ 4x_1 + 2x_2 \le 160, & (2); \\ x_{1,2} \in N_0. \end{cases}$$

Изобразим область допустимых решений в системе координат Ox_1x_2.

Найдем две точки, лежащие на прямой $2x_1 + 5x_2 = 200$. Пусть $x_1 = 0$, тогда из данного уравнения имеем: $x_2 = 200/5 = 40$. Пусть $x_2 = 0$, тогда из уравнения имеем $x_1 = 200/2 = 100$.

В итоге получим координаты двух точек прямой (1): *A*(0; 40), *B*(100; 0). В системе координат Ox_1x_2 построим прямую по точкам *A* и *B* (Рис. 2.2).

Аналогично для другой прямой $4x_1 + 2x_2 = 160$. Пусть $x_1 = 0$, тогда $x_2 = 160/2 = 80$. Пусть $x_2 = 0$, тогда $x_1 = 160/4 = 40$.

В итоге имеем координаты двух точек прямой (2): *C*(0; 80), *D*(40; 0). В системе координат Ox_1x_2 построим прямую по точкам *C* и *D*.

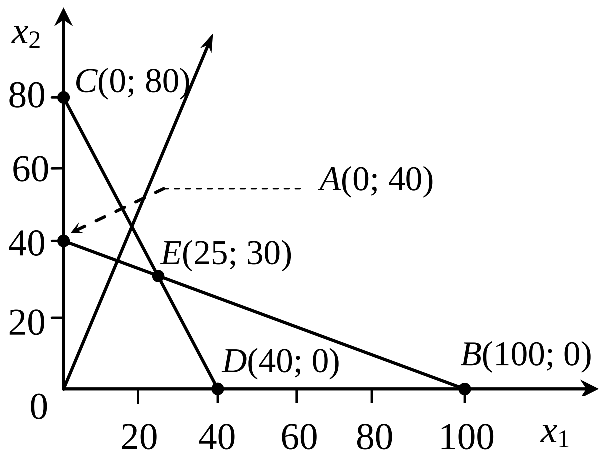


Рис. 2.2. Построение области Парето

В результате получили три замкнутые области и одну незамкнутую. Для определения области допустимых решений возьмем контрольную точку с координатами (0; 0), принадлежащую только четырехугольнику *OAED*. Координаты этой точки удовлетворяют всем ограничениям. Следовательно, выбранный четырехугольник *OAED* является искомой областью Парето. Контрольной точкой может быть не только вершина многоугольника, но и любая внутренняя точка полученных многоугольников.

Найдем координаты точки пересечения прямых (1) и (2) из системы уравнений:

$$\begin{cases} 2x_1 + 5x_2 = 200; \\ 4x_1 + 2x_2 = 160. \end{cases}$$

Разделим второе уравнение системы на два и вычтем из второго уравнения первое:

$$\begin{cases} 2x_1 + 5x_2 = 200; \\ -4x_2 = -120. \end{cases}$$

Разделим второе уравнение на (– 4):

$$\begin{cases} 2x_1 + 5x_2 = 200; \\ x_2 = 30. \end{cases}$$

Подставим в первое уравнение численное значение $x_2 = 30$. Получим координаты точки *E*(25; 30) пересечения прямых *AB* и *CD*. Координаты точек пересечения построенных прямых с осями координат найдены при построении прямых. Найдем значения целевой функции в вершинах четырехугольника *O*(0; 0), *A*(0; 40), *E*(25; 30) и *D*(40; 0):

F(0, 0) = 0; *F*(0, 40) = 960; *E*(25, 30) = 970; *F*(40, 0) = 400.

Наибольший срок службы составит 970 ед. при использовании 25 единиц блоков деревянных конструкций в мебели типа *A* и 30 блоков металлических конструкций для изделий типа *B*.

Рассмотрим пример нахождения нецелочисленного решения задачи линейного программирования графическим методом:

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

$$\begin{aligned} &x_1 - x_2 \geq -3; \\ &6x_1 + 7x_2 \leq 42; \\ &2x_1 - 3x_2 \leq 6; \\ &x_1 + x_2 \geq 4; \\ &x_1,\ x_2 \geq 0. \end{aligned}$$

Шаг 1. Построим область допустимых решений, т. е. решим графически систему неравенств. Для этого построим каждую прямую и определим полуплоскости, заданные неравенствами, знаки неравенств обозначены штрихом на прямых (Рис. 2.3).

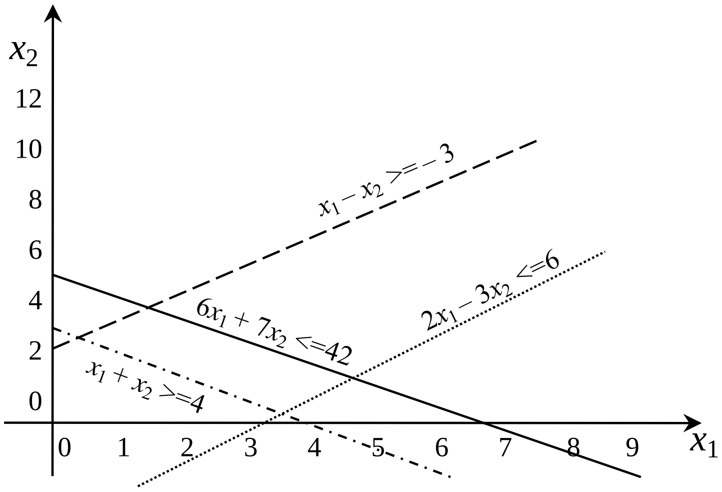


Рис. 2.3. Построение области ограничений

Шаг 2. Определение границы области допустимых решений. Пересечением полуплоскостей будет являться область, координаты точек которого удовлетворяют условию неравенствам системы ограничений задачи. Обозначим границы области многоугольника решений (Рис. 2.4).

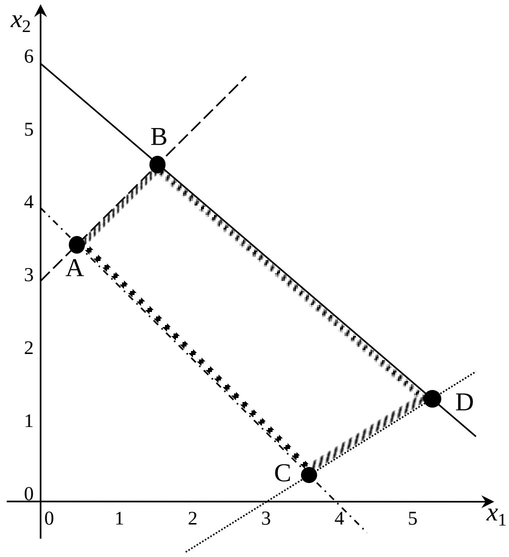


Рис. 2.4. Границы области Парето

Шаг 3. Рассмотрим целевую функцию задачи $F(x_1, x_2) = 2x_1 - x_2 \rightarrow max$. Построим прямую, отвечающую значению функции $F = 0$: $F = 2x_1 - x_2 = 0$. Вектор-градиент, составленный из коэффициентов целевой функции, указывает направление максимума функции $F(x_1, x_2)$. Начало вектора – точка (0; 0), конец – точка (2; – 1). Поскольку по условию задачи интересует максимальное решение, поэтому двигаем прямую до последней точки касания обозначенной области. На графике эта прямая обозначена пунктирной линией (Рис. 2.5).

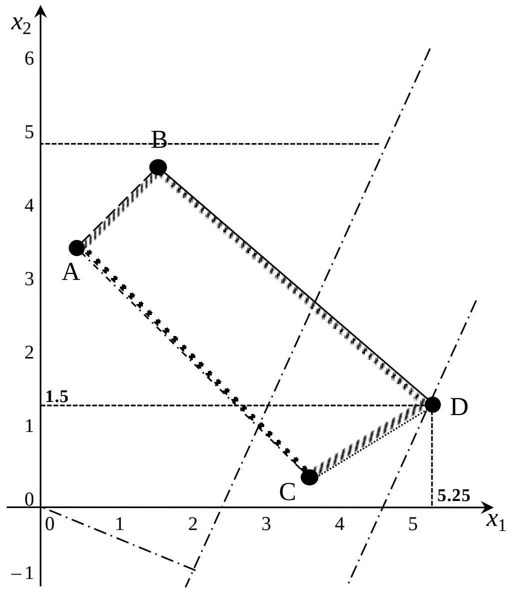


Рис. 2.5. Нахождение «крайней точки» на области решений

Прямая пересекает область в точке D. Так как точка D получена в результате пересечения прямых (2) и (3) из условий ограничений, то ее координаты удовлетворяют уравнениям этих прямых:

$$\begin{cases} 6x_1 + 7x_2 = 42; \\ 2x_1 - 3x_2 = 6. \end{cases}$$

Получим: $x_1 = 5.25$, $x_2 = 1.5$, значение целевой функции: $F(x_1, x_2) = 9$.

2.3. Симплексный метод

Симплексный метод относится к аналитическим методам решения задач линейного программирования. Он представляет собой последовательность шагов (итераций) перехода от одной вершины многогранника области Парето к другой с целью получения максимального значения целевой функции. ЗЛП любого типа могут быть решены симплекс-методом.

Возьмем в качестве примера математическую модель одну из рассмотренных задач без учета ограничений целочисленности переменных:

$$f(x) = 10x_1 + 24x_2 \to max;$$

$$\begin{cases} 2x_1 + 5x_2 \le 200, & (1); \\ 4x_1 + 2x_2 \le 160, & (2); \\ x_1 \ge 0, \quad x_2 \ge 0. \end{cases}$$

Добавим в 1-е и 2-е ограничения соответственно неотрицательные переменные x_3 и x_4, и запишем целевую функцию в неявном виде $F(x) = 0$:

$$f(x) - 10x_1 - 24x_2 = 0;$$

$$\begin{cases} 2x_1 + 5x_2 + x_3 = 200; \\ 4x_1 + 2x_2 + x_4 = 160; \\ x_1 \ge 0, \quad j = \overline{1, 4}. \end{cases}$$

Каждый коэффициент целевой функции $F(x)$, если он отрицателен, определяет величину положительного приращения целевой функции $f(x)$, а если положителен – отрицательного приращения целевой функции $f(x)$ при увеличении соответствующей переменной. Если x_1 увеличить на единицу, то целевая функция $f(x)$ увеличится на 10 единиц, а если x_2 увеличить на единицу, то целевая функция $f(x)$ увеличится на 24 единицы. Переменная x_2 сильнее влияет на целевую функцию, чем переменная x_1.

На первом шаге симплексного метода решения канонической ЗЛП необходимо попасть в одну из вершин многогранника области Парето. Выражаем из m ограничений канонической ЗЛП m переменных иразрешаем систему из m уравнений относительно выбранных m переменных. Исключаем эти m переменных из выражения целевой функции. Выбранные m переменные называют базисными или основными, а остальные – небазисными или свободными.

В рассматриваемой задаче в качестве базисных переменных удобно взять переменные x_3 и x_4. Они не входят в целевую функцию и относительно них легко разрешить систему двух уравнений:

$$f(x) - 10x_1 - 24x_2 = 0;$$

$$\begin{cases} x_3 = 200 - 2x_1 - 5x_2; \\ x_4 = 160 - 4x_1 - 2x_2; \\ x_j \geq 0, \quad j = \overline{1, 4}. \end{cases}$$

Первоначальное допустимое решение получается при нулевых значениях свободных переменных $x_1 = x_2 = 0$, $x_3 = 200$, $x_4 = 160$, $f(x) = 0$.

На втором шаге необходимо перейти в другую вершину многогранника области Парето, в которой целевая функция будет принимать большее значение. Аналитически это означает замену одной базисной переменной на другую, свободную переменную. Выбор новой базисной переменной осуществляют по следующим правилам.

Если в целевой функции имеются свободные переменные, коэффициенты при которых отрицательные, то желательно выбрать из них переменную с наибольшим по модулю значением коэффициента в качестве новой базисной переменной. Если все свободные переменные имеют неотрицательные коэффициенты, то оптимальное решение получено и улучшить его нельзя.

Свободная переменная x_2 имеет наибольший по модулю отрицательный коэффициент в целевой функции. Выберем в качестве новой базисной переменной x_2 и разрешим систему ограничений относительно базисных переменных x_2 и x_4, исключим новую базисную переменную x_2 из целевой функции. Эти преобразования удобно выполнять в виде симплекс-таблицы (табл. 2.1), которая составлена на основе канонической ЗЛП с целевой функцией $F(x)$, записанной в неявном виде.

Таблица 2.1

x_1	x_2	x_3	x_4	Свободные члены	Базис
2	5	1	0	200	x_3
4	2	0	1	160	x_4
– 10	– 24	0	0	0	$F(x)$

Столбцы при переменных x_3 и x_4 являются единичными. Единичный столбец – это столбец, все элементы которого равны нулю, кроме одного, равного единице. В симплекс-таблице количество единичных столбцов должно быть равно количеству базисных переменных. Все свободные члены должны быть неотрицательными. Неотрицательность свободных членов получают при составлении канонической ЗЛП.

Рассмотрим алгоритм ввода новой базисной переменной и исключения старой, реализованный симплекс-таблицей.

1. Выбираем наибольший по модулю отрицательный коэффициент целевой функции – это коэффициент при переменной x_2. Столбец $l = 2$ симплекс-таблицы, которому принадлежит новая базисная переменная x_2, называется ведущим столбцом.

2. Делим элементы столбца свободных членов на соответствующие элементы ведущего столбца, при этом учитываем только положительные элементы. Выбираем наименьшее частное $min\left\{\frac{200}{5}; \frac{160}{2}\right\} = min\{40; 80\}$. Строка $k = 1$, которой принадлежит наименьшее частное, называется ведущей строкой. На пересечении ведущего столбца и ведущей строки

находится ведущий элемент $a_{12} = 5$. Первая строка симплекс-таблицы не нумеруется.

3. Делим элементы ведущей строки на ведущий элемент:

$$a'_{kj} = \frac{a_{kj}}{a_{kl}},$$

где k – номер ведущей строки; l – номер ведущего столбца; a_{kl} – ведущий элемент.

Все остальные элементы симплекс таблицы пересчитываем по формуле:

$$a'_{ij} = a_{ij} - \frac{a_{kj}a_{il}}{a_{kl}} \quad \text{или} \quad a'_{ij} = a_{ij} - a'_{kj}a_{il}.$$

Элементы ведущего столбца, кроме ведущего элемента, обнуляются:

$$a'_{il} = a_{il} - \frac{a_{kl}a_{il}}{a_{kl}} = 0.$$

Первая строка (ведущая):

$$a'_{11} = \frac{a_{11}}{a_{12}} = \frac{2}{5} = 0.4; \quad a'_{12} = \frac{a_{12}}{a_{12}} = \frac{5}{5} = 1; \quad a'_{13} = \frac{a_{13}}{a_{12}} = \frac{1}{5} = 0.2;$$

$$a'_{14} = \frac{a_{14}}{a_{12}} = \frac{0}{5} = 0; \quad a'_{15} = \frac{a_{15}}{a_{12}} = \frac{200}{5} = 40.$$

Вторая строка:

$$a'_{21} = a_{21} - a'_{11}a_{22} = 4 - 0.4 \cdot 2 = 3.2;$$

$$a'_{22} = 0;$$

$$a'_{23} = a_{23} - a'_{13}a_{22} = 0 - 0.2 \cdot 2 = -0.4;$$

$$a'_{24} = a_{24} - a'_{14}a_{22} = 1 - 0 \cdot 2 = 1;$$

$$a'_{25} = a_{25} - a'_{15}a_{22} = 160 - 40 \cdot 2 = 80.$$

Третья строка:

$$a'_{31} = a_{31} - a'_{11}a_{32} = -10 - 0.4 \cdot (-24) = -4.4;$$

$$a'_{32} = 0;$$

$$a'_{33} = a_{33} - a'_{13}a_{32} = 0 - 0.2 \cdot (-24) = 4.8;$$

$$a'_{34} = a_{34} - a'_{14}a_{32} = 0;$$

$$a'_{35} = a_{35} - a'_{15}a_{32} = 0 - 40 \cdot (-24) = 480.$$

В столбце «Базис» старую базисную переменную ведущей строки x_3 заменяем на новую базисную переменную ведущего столбца x_2 (табл. 2.2).

Таблица 2.2

x_1	x_2	x_3	x_4	Свободные члены	Базис
0.4	1	0.2	0	40	x_2
3.2	0	– 0.4	1	80	x_4
– 4.4	0	4.8	0	480	$F(x)$

Первый итерационный шаг симплексного метода завершен. Получили новую симплекс-таблицу (см. табл.2.2) и новый опорный план (0; 40; 0; 80), $f(x) = 480$.

Первый шаг соответствует переходу от одной вершины O(0; 0) многоугольника области Парето к другой A(0; 40).

Коэффициент при переменной x_1 в целевой функции отрицателен, т. е. оптимальное значение еще не достигнуто. Повторяем шаги 1–3 еще раз.

1. Выбираем наибольший по модулю отрицательный коэффициент целевой функции – это коэффициент при переменной x_1. Столбец $l = 1$ симплекс-таблицы, которому принадлежит новая базисная переменная x_1, является ведущим столбцом.

2. Делим элементы столбца свободных членов на соответствующие элементы ведущего столбца, при этом учитываем только положительные элементы. Выбираем наименьшее частное $min\left\{\frac{40}{0.4}; \frac{160}{3.2}\right\} = min\{100; 25\}$.

Строка $k = 2$, которой принадлежит наименьшее частное, является ведущей строкой. На пересечении ведущего столбца и ведущей строки находится ведущий элемент $a_{21} = 3.2$.

3. Пересчитываем элементы симплекс-таблицы по формулам, приведенным на первом шаге. Начинаем с ведущей строки.

$$a'_{21} = \frac{a_{21}}{a_{21}} = 1; \qquad a'_{22} = \frac{a_{22}}{a_{21}} = \frac{0}{3.2} = 0; \qquad a'_{23} = \frac{a_{23}}{a_{21}} = \frac{-0.4}{3.2} = -0.125;$$

$$a'_{24} = \frac{a_{24}}{a_{21}} = \frac{1}{3.2} = 0.3125; \qquad a'_{25} = \frac{a_{25}}{a_{21}} = \frac{80}{3.2} = 25.$$

Первая строка:

$$a'_{11} = 0;$$

$$a'_{12} = a_{12} - a'_{22}a_{11} = 1 - 0 \cdot 0.4 = 1;$$

$$a'_{13} = a_{13} - a'_{23}a_{11} = 0.2 - (-0.125) \cdot 0.4 = 0.25;$$

$$a'_{14} = a_{14} - a'_{24}a_{11} = 0 - 0.3125 \cdot 0.4 = -0.125;$$

$$a'_{15} = a_{15} - a'_{25}a_{11} = 40 - 25 \cdot 0.4 = 30.$$

Третья строка:

$$a'_{31} = 0;$$

$$a'_{32} = a_{32} - a'_{22}a_{13} = 0 - 0 \cdot (-4.4) = 0;$$

$$a'_{33} = a_{33} - a'_{23}a_{13} = 4.8 - (-0.125) \cdot (-4.4) = 4.25;$$

$$a'_{34} = a_{34} - a'_{24}a_{13} = 0 - 0.3125 \cdot (-4.4) = 1.375;$$

$$a'_{35} = a_{35} - a'_{25}a_{13} = 480 - 25 \cdot (-4.4) = 590.$$

В столбце «Базис» прежнюю базисную переменную ведущей строки x_4 заменяем на новую базисную переменную ведущего столбца x_1 (табл. 2.3).

Таблица 2.3

x_1	x_2	x_3	x_4	Свободные члены	Базис
0	1	0.25	– 0.125	30	x_1
1	0	– 0.125	0.3125	25	x_2
0	0	4.25	1.375	590	$F(x)$

Второй шаг симплексного метода завершен. Получили новую симплекс-таблицу (см. табл. 2.3) и новый опорный план (25; 30; 0; 0), $f(x) = 590$.

Второй итерационный шаг соответствует переходу от вершины $A(0; 40)$ многоугольника к другой его вершине $B(25; 30)$. Коэффициенты целевой функции неотрицательны, поэтому новый опорный план (25; 30; 0; 0), $f(x) = 590$ является оптимальным.

Рассмотрим решение второго примера симплексным методом. Запишем математическую модель в каноническом виде без учета целочисленности переменных x_1 и x_2, добавив три неотрицательные переменные x_3, x_4 и x_5:

$$f(x) - 4x_1 - 3x_2 = 0;$$

$$\begin{cases} 2x_1 + 5x_2 + x_3 = 400; \\ x_1 - x_4 = 50; \\ x_2 - x_5 = 20; \\ x_j \geq 0, \quad j = \overline{1,5}. \end{cases}$$

Коэффициенты при переменных x_4 и x_5 отрицательны. При составлении симплекс-таблицы они не составят единичных столбцов, поэтому введем две искусственные неотрицательные переменные x_6 и x_7:

$$f(x) - 4x_1 - 3x_2 = 0;$$

$$\begin{cases} 2x_1 + 5x_2 + x_3 = 400; \\ x_1 - x_4 + x_6 = 50; \\ x_2 - x_5 + x_7 = 20; \\ x_j \geq 0, \ j = \overline{1,7}. \end{cases}$$

Искусственные переменные не входят в целевую функцию. Составим симплекс-таблицу (табл. 2.4).

Таблица 2.4

x_1	x_2	x_3	x_4	x_5	x_6	x_7	Свободные члены	Базис
2	5	1	0	0	0	0	400	x_3
1	0	0	– 1	0	1	0	50	x_6
0	1	0	0	– 1	0	1	20	x_7
– 4	– 3	0	0	0	–	–	0	$F(x)$

Выводим пошагово из базиса искусственные переменные x_6 и x_7. Геометрически это означает попадание в область Парето.

1. Выбираем вторую строку $k=2$ при переменной x_6 в качестве ведущей.

2. Выбираем первый столбец $l=1$ при переменной x_1 в качестве ведущего. На пересечении ведущего столбца и ведущей строки находится ведущий элемент $a_{21}=1$.

3. Пересчитываем элементы симплекс-таблицы по формулам, приведенным в предыдущей задаче. Начинаем с ведущей строки.

Вторая строка (ведущая) остается неизменной:

$$a'_{21}=1;\ a'_{22}=0;\ a'_{23}=0;\ a'_{24}=-1;\ a'_{25}=0;\ a'_{26}=1;\ a'_{27}=0;\ a'_{28}=50.$$

Первая строка (умножаем ведущую строку на два и результат вычитаем из первой строки):

$$a'_{11}=0;\ a'_{12}=5;\ a'_{13}=1;\ a'_{14}=2;\ a'_{15}=0;\ a'_{16}=-2;\ a'_{17}=0;\ a'_{18}=300.$$

Третья строка остается неизменной:

$$a'_{31}=0;\ a'_{32}=1;\ a'_{33}=0;\ a'_{34}=0;\ a'_{35}=-1;\ a'_{36}=0;\ a'_{37}=1;\ a'_{38}=20.$$

Четвертая строка (умножаем ведущую строку на четыре и результат прибавляем к четвертой строке):

$$a'_{41}=0;\ a'_{42}=-3;\ a'_{43}=0;\ a'_{44}=-4;\ a'_{45}=0;\ a'_{48}=200.$$

В столбце «Базис» прежнюю базисную переменную ведущей строки x_6 заменяем на новую базисную переменную ведущего столбца x_1. Вычеркиваем столбец выведенной искусственной переменной x_6 (табл. 2.5).

Таблица 2.5

x_1	x_2	x_3	x_4	x_5	x_7	Свободные члены	Базис
0	5	1	2	0	0	300	x_3
1	0	0	– 1	0	0	50	x_1
0	1	0	0	– 1	1	20	x_7
0	– 3	0	– 4	0	–	200	$F(x)$

Выводим из базиса искусственную переменную x_7.

1. Выбираем третью строку $k = 3$ при переменной x_7 в качестве ведущей.

2. Выбираем второй столбец $l = 2$ при переменной x_2 в качестве ведущего. На пересечении ведущего столбца и ведущей строки находится ведущий элемент $a_{32} = 1$.

3. Пересчитываем элементы симплекс-таблицы (см. табл. 2.5).

Третья строка (ведущая) остается неизменной:

$$a'_{31} = 0;\ a'_{32} = 1;\ a'_{33} = 0;\ a'_{34} = 0;\ a'_{35} = -1;\ a'_{37} = 1;\ a'_{38} = 20.$$

Первая строка (умножаем ведущую строку на пять и результат вычитаем из первой строки):

$$a'_{11} = 0;\ a'_{12} = 0;\ a'_{13} = 1;\ a'_{14} = 2;\ a'_{15} = 5;\ a'_{17} = -5;\ a'_{18} = 200.$$

Вторая строка остается неизменной:

$$a'_{21} = 1;\ a'_{22} = 0;\ a'_{23} = 0;\ a'_{24} = -1;\ a'_{25} = 0;\ a'_{27} = 0;\ a'_{28} = 50.$$

Четвертая строка (умножаем ведущую строку на три и результат прибавляем к четвертой строке):

$$a'_{41} = 0;\ a'_{42} = 0;\ a'_{43} = 0;\ a'_{44} = -4;\ a'_{45} = -3;\ a'_{48} = 260.$$

В столбце «Базис» прежнюю базисную переменную ведущей строки x_7 заменяем на новую базисную переменную ведущего столбца x_2. Вычеркиваем столбец выведенной искусственной переменной x_7 (табл. 2.6).

Таблица 2.6

x_1	x_2	x_3	x_4	x_5	Свободные члены	Базис
0	0	1	2	5	200	x_3
1	0	0	– 1	0	50	x_1
0	1	0	0	– 1	20	x_2
0	0	0	– 4	– 3	260	$F(x)$

Искусственные базисные переменные x_6 и x_7 выведены из базиса. Геометрически это означает попадание в одну из вершин многоугольника области Парето. Если не удается вывести искусственные переменные из базиса или получена неразрешимая задача, то исходная задача также неразрешима.

Далее используем стандартный алгоритм симплексного метода.

1. Выбираем наибольший по модулю отрицательный коэффициент целевой функции – это коэффициент при переменной x_4. Столбец $l = 4$ симплекс-таблицы, которому принадлежит новая базисная переменная x_4, является ведущим столбцом.

2. Делим элементы столбца свободных членов на соответствующие элементы ведущего столбца, при этом учитываем только положительные элементы. Выбираем наименьшее частное $min\left\{\frac{200}{2}\right\} = 100$. Строка $k = 1$, которой принадлежит наименьшее частное, является ведущей строкой. На пересечении ведущего столбца и ведущей строки находится ведущий элемент $a_{14} = 2$.

3. Пересчитываем элементы симплекс-таблицы (см. табл. 2.6). Начинаем с ведущей строки.

Первая строка (делим каждый элемент ведущей строки на ведущий элемент):

$$a'_{11} = 0;\ a'_{12} = 0;\ a'_{13} = 0.5;\ a'_{14} = 1;\ a'_{15} = 2.5;\ a'_{16} = 100.$$

Вторая строка (первую строку a'_{1j} прибавляем ко второй):

$$a'_{21} = 1;\ a'_{22} = 0;\ a'_{23} = 0.5;\ a'_{24} = 0;\ a'_{25} = 2.5;\ a'_{26} = 150.$$

Третья строка остается неизменной:

$$a'_{31} = 0;\ a'_{32} = 1;\ a'_{33} = 0;\ a'_{34} = 0;\ a'_{35} = -1;\ a'_{36} = 20.$$

Четвертая строка (умножаем первую строку a'_{1j} на четыре и результат прибавляем к четвертой строке):

$$a'_{41} = 0;\ a'_{42} = 0;\ a'_{43} = 2;\ a'_{44} = 0;\ a'_{45} = 7;\ a'_{46} = 660.$$

В столбце «Базис» прежнюю базисную переменную ведущей строки x_3 заменяем на новую базисную переменную ведущего столбца x_4 (табл. 2.7).

Таблица 2.7

x_1	x_2	x_3	x_4	x_5	Свободные члены	Базис
0	0	0.5	1	2.5	100	x_4
1	0	0.5	0	2.5	150	x_1
0	1	0	0	– 1	20	x_2
0	0	2	0	7	660	$F(x)$

Коэффициенты целевой функции неотрицательны, поэтому новый опорный план (150; 20; 0; 100; 0), $f(x) = 660$ является оптимальным.

Рассмотрим симплекс-метод получения нецелочисленного решения ЗЛП. Определим минимальное значение целевой функции $F(x_1, x_2) = 5x_1 - 3x_2$ следующих ограничениях:

$$\begin{aligned}
&3x_1 + 2x_2 \geq 6;\\
&-2x_1 + 3x_2 \leq 6;\\
&x_1 - x_2 \leq 4;\\
&4x_1 + 7x_2 \leq 28.
\end{aligned}$$

Для построения первого опорного плана систему неравенств приведем к системе уравнений путем введения дополнительных переменных (*переход к канонической форме*):

$$\begin{aligned}
&3x_1 + 2x_2 - 1x_3 + 0x_4 + 0x_5 + 0x_6 = 6;\\
&-2x_1 + 3x_2 - 0x_3 + 1x_4 + 0x_5 + 0x_6 = 6;\\
&1x_1 - 1x_2 + 0x_3 + 0x_4 + 1x_5 + 0x_6 = 4;\\
&4x_1 + 7x_2 + 0x_3 + 0x_4 + 0x_5 + 1x_6 = 28.
\end{aligned}$$

Введем *искусственные переменные* x: в 1-м равенстве вводим переменную x_7; а целевую функцию представим виде:

$$F(X) = Mx_7 \to min.$$

Из уравнений выражаем искусственную переменную: $x_7 = 6 - 3x_1 - 2x_2 + x_3$, которую подставим в целевую функцию: $F(X) = M(6 - 3x_1 - 2x_2 + x_3) \to min$, иначе:

$$F(X) = (-3M)x_1 + (-2M)x_2 + (M)x_3 + (6M) \to min.$$

Введем новую переменную $x_0 = -3x_1 - 2x_2$.

Выразим базисные переменные <7, 4, 5, 6> через небазисные:

$$\begin{aligned}
x_0 &= 6 - 3x_1 - 2x_2 + x_3;\\
x_7 &= 6 - 3x_1 - 2x_2 + x_3;\\
x_4 &= 6 + 2x_1 - 3x_2;\\
x_5 &= 4 - x_1 + x_2;\\
x_6 &= 28 - 4x_1 - 7x_2.
\end{aligned}$$

Переходим к основному алгоритму симплекс-метода.

Поскольку задача решается на минимум, то переменную для включения в текущий план выбирают по минимальному отрицательному числу в уравнении для x_0.

Проверка критерия оптимальности. В выражении для x_0 присутствуют положительные элементы. Следовательно, полученный план не оптимален.

Определение новой базисной переменной, как:

$$max\left(-3, -2, 1, 0, 0, 0, 0\right) = -3;$$

$$\begin{aligned}
x_0 &= 6 - 3x_1 - 2x_2 + x_3;\\
x_7 &= 6 - 3x_1 - 2x_2 + x_3;\\
x_4 &= 6 + 2x_1 - 3x_2;\\
x_5 &= 4 - x_1 + x_2;\\
x_6 &= 28 - 4x_1 - 7x_2.
\end{aligned}$$

В качестве новой переменной выбираем x_1. Вычислим значения D_i по всем уравнениям для этой переменной: b_i / a_{i1} из них выберем наименьшее:

$$min(6:3, -, 4:1, 28:4) = 2.$$

Вместо переменной x_7 в план войдет переменная x_1. Выразим переменную x_1 через x_7, то есть $x_1 = 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3 - \frac{1}{3}x_7$ и подставим во все выражения:

$$\begin{aligned}
x_0 &= 6 - 3\left(2 - \frac{2}{3}x_2 + \frac{1}{3}x_3 - \frac{1}{3}x_7\right) - 2x_2 + x_3;\\
x_4 &= 6 + 2\left(2 - \frac{2}{3}x_2 + \frac{1}{3}x_3 - \frac{1}{3}x_7\right) - 3x_2;\\
x_5 &= 4 - \left(2 - \frac{2}{3}x_2 + \frac{1}{3}x_3 - \frac{1}{3}x_7\right) + x_2;\\
x_6 &= 28 - 4\left(2 - \frac{2}{3}x_2 + \frac{1}{3}x_3 - \frac{1}{3}x_7\right) - 7x_2.
\end{aligned}$$

После приведения всех подобных, получаем новую систему, эквивалентную прежней:

$$\begin{aligned}
x_0 &= 0 + x_7;\\
x_1 &= 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3 - \frac{1}{3}x_7;\\
x_4 &= 10 - \frac{13}{3}x_2 + \frac{2}{3}x_3 - \frac{2}{3}x_7;\\
x_5 &= 2 + \frac{5}{3}x_2 - \frac{1}{3}x_3 + \frac{1}{3}x_7;\\
x_6 &= 20 - \frac{13}{3}x_2 - \frac{4}{3}x_3 + \frac{4}{3}x_7.
\end{aligned}$$

Полагая небазисные переменные $x = (1, 4, 5, 6)$ равными нулю, получим новый допустимый вектор и значение целевой функции:

$$x = (0, 0, 0, 0, 0, 0, -1),\ x_0 = 0.$$

Выражение для x_0 не содержит отрицательных элементов. Найден оптимальный план:

$$\begin{aligned}
x_0 &= 0 + x_7;\\
x_1 &= 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3 - \frac{1}{3}x_7;\\
x_4 &= 10 - \frac{13}{3}x_2 + \frac{2}{3}x_3 - \frac{2}{3}x_7;\\
x_5 &= 2 + \frac{5}{3}x_2 - \frac{1}{3}x_3 + \frac{1}{3}x_7;\\
x_6 &= 20 - \frac{13}{3}x_2 - \frac{4}{3}x_3 + \frac{4}{3}x_7.
\end{aligned}$$

На этом первый этап симплекс-метода завершен. Переходим ко *второму этапу*. Удаляем элементы с искусственными переменными.

$$\begin{aligned}
x_1 &= 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3;\\
x_4 &= 10 - \frac{13}{3}x_2 + \frac{2}{3}x_3;\\
x_5 &= 2 + \frac{5}{3}x_2 - \frac{1}{3}x_3;\\
x_6 &= 20 - \frac{13}{3}x_2 - \frac{4}{3}x_3.
\end{aligned}$$

Выразим базисные переменные: $x_1 = 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3$, которые подставим в целевую функцию:

$$F(X) = 5\left(2 - \frac{2}{3}x_2 + \frac{1}{3}x_3\right) - 3x_2 \text{ или } F(X) = 10 - \frac{19}{3}x_2 + \frac{5}{3}x_3.$$

Получаем новую систему переменных.

$$\begin{aligned}
x_0 &= 10 - \frac{19}{3}x_2 + \frac{5}{3}x_3;\\
x_1 &= 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3;\\
x_4 &= 10 - \frac{13}{3}x_2 + \frac{2}{3}x_3;\\
x_5 &= 2 + \frac{5}{3}x_2 - \frac{1}{3}x_3;\\
x_6 &= 20 - \frac{13}{3}x_2 - \frac{4}{3}x_3.
\end{aligned}$$

Проверка критерия оптимальности. В выражении для x_0 присутствуют положительные элементы. Следовательно, полученный план все еще не оптимален.

Определение новой базисной переменной:

$$max\left(0, -\frac{19}{3}, \frac{5}{3}, 0, 0, 0\right) = -\frac{19}{3};$$

$$x_0 = 10 - \frac{19}{3}x_2 + \frac{5}{3}x_3;$$
$$x_1 = 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3;$$
$$x_4 = 10 - \frac{13}{3}x_2 + \frac{2}{3}x_3;$$
$$x_5 = 2 + \frac{5}{3}x_2 - \frac{1}{3}x_3;$$
$$x_6 = 20 - \frac{13}{3}x_2 - \frac{4}{3}x_3.$$

В качестве новой переменной выбираем x_2. Вычислим значения D_i по всем уравнениям для этой переменной: b_i / a_{i2} из них выберем наименьшее:

$$min\left(2 : \frac{2}{3}, 10 : 4\frac{1}{3}, -, 20 : 4\frac{1}{3}\right) = 2\frac{4}{13}.$$

Вместо переменной x_4 в план войдет переменная x_2.

Выразим переменную x_2 через x_4: $x_2 = \frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4$ и подставим во все выражения:

$$x_0 = 10 - 6\frac{1}{3}\left(\frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4\right) + 1\frac{2}{3}x_3;$$
$$x_4 = 2 - \frac{2}{3}\left(\frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4\right) + \frac{1}{3}x_3;$$
$$x_5 = 2 + 1\frac{2}{3}\left(\frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4\right) - \frac{1}{3}x_3;$$
$$x_6 = 20 - 4\frac{1}{3}\left(\frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_{47}\right) - 1\frac{1}{3}x_3.$$

После приведения всех подобных, получаем новую систему, эквивалентную полученной ранее:

$$
\begin{aligned}
&x_0 = -\frac{60}{13} + \frac{9}{13}x_3 + \frac{19}{13}x_4;\\
&x_1 = \frac{6}{13} + \frac{3}{13}x_3 + \frac{2}{13}x_4;\\
&x_2 = \frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4;\\
&x_5 = \frac{76}{13} - \frac{1}{13}x_3 - \frac{5}{13}x_4;\\
&x_6 = 10 - 2x_3 + x_4.
\end{aligned}
$$

Полагая небазисные переменные $x = (1, 2, 5, 6)$ равными нулю, получим новый допустимый вектор и значение целевой функции:

$$
x = \left(0, 0, -\frac{9}{13}, -\frac{19}{13}, 0, 0\right),\ x_0 = -\frac{60}{13}.
$$

Выражение для x_0 не содержит положительных элементов, значит, найден оптимальный план. Окончательный вариант системы уравнений:

$$
\begin{aligned}
&x_0 = -\frac{60}{13} + \frac{9}{13}x_3 + \frac{19}{13}x_4;\\
&x_1 = \frac{6}{13} + \frac{3}{13}x_3 + \frac{2}{13}x_4;\\
&x_2 = \frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4;\\
&x_5 = \frac{76}{13} - \frac{1}{13}x_3 - \frac{5}{13}x_4;\\
&x_6 = 10 - 2x_3 + x_4.
\end{aligned}
$$

Так как в оптимальном решении отсутствуют искусственные переменные (они равны нулю), то данное решение является допустимым. Оптимальный план можно записать так: $x_1 = \frac{6}{13}$, $x_2 = 2\frac{4}{13}$.

Симплекс-метод является более громоздким по сравнению с графическим и достаточно трудоемким. При большом числе итераций ручной расчет симплекс-метода займет значительное количество времени, однако, его неоспоримое преимущество перед графическим методом –

точность расчетов. Разрешить противоречие между скоростью получения решений и их точностью можно путем использования специальных пакетов прикладных программ или широко распространенной компьютерной программы Excel. Она нужна для проведения расчетов, составления таблиц и диаграмм, вычисления простых и сложных функций и входит в состав пакета Microsoft Office.

2.4. Транспортная задача: открытая и закрытая

Транспортная задача – это задача о поиске оптимального распределения поставок однородного товара от поставщиков к потребителям при известных затратах на перевозку (тарифах) между пунктами отправления и назначения. Является задачей линейного программирования специального вида.

В модели транспортной задачи предполагается, что суммарные запасы поставщиков равны суммарным запросам потребителей, т. е.:

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

Такая задача называется задачей с *правильным балансом*, а модель задачи *закрытой*. Если же это равенство не выполняется, то задача называется задачей с *неправильным балансом*, а модель задачи – *открытой*.

Математическая формулировка транспортной задачи такова: найти переменные задачи $X = (x_{i\,j})$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, удовлетворяющие системе ограничений, условиям неотрицательности и обеспечивающие минимум целевой функции.

Учитывая неотрицательность объемов перевозок задача выглядит следующим образом: однородный груз сосредоточен у m поставщиков в объемах $a_1, a_2, \ldots, a_m$. Данный груз необходимо доставить n потребителям в объемах $b_1, b_2, \ldots, b_n$. Известны $C_{i\,j}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$ – стоимости перевозки единиц груза от каждого i-го поставщика каждому j-му потребителю. Требуется составить такой план перевозок, при котором запасы всех поставщиков вывозятся полностью, запросы всех потребителей удовлетворяются полностью, и суммарные затраты на перевозку всех грузов являются минимальными.

В транспортных задачах под поставщиками и потребителями понимаются различные предприятия, заводы, фабрики, склады, магазины и т. д. Однородными считаются грузы, которые могут быть перевезены одним видом транспорта. Под стоимостью перевозок понимаются тарифы, время, расход топлива и т. п. При решении транспортной задачи выбор критерия оптимальности имеет важное значение. Оценка экономической эффективности примерного плана может определяться по тому или иному показателю, положенного в основу расчета плана. Этот критерий является экономическим показателем, характеризующим качество плана. До настоящего времени нет общепринятого критерия, всесторонне учитывающего экономические факторы. При решении транспортной задачи, в качестве критерия оптимальности в различных случаях используют следующие показатели:

1. Объем работы транспорта (критерий – расстояние в т/км). Минимум пробега удобен для оценки планов перевозок, поскольку расстояние перевозки определяется легко и точно для любого направления. По этому критерию нельзя решать транспортные задачи с участием многих видов транспорта.

2. Тарифная плата за перевозку груза (критерий – тарифы провозных плат). Позволяет получить схему перевозок, наилучшую с точки зрения хозрасчетных показателей предприятия. Все надбавки, а также существующие льготные тарифы затрудняют его использование.

3. Эксплуатационные расходы на транспортировку грузов (критерий – себестоимость эксплуатационных расходов). Более верно отражает экономичность перевозок различными видами транспорта. Позволяет делать обоснованные выводы о целесообразности переключения с одного вида транспорта на другой.

4. Сроки доставки грузов (критерий – затраты времени).

5. Приведенные затраты (с учетом эксплуатационных расходов, зависящих от размеров движения и капиталовложения в подвижной состав).

6. Приведенные затраты (с учетом полных эксплуатационных расходов капиталовложений на строительство объектов в подвижной состав):

$$C_{приб} = C_{эи} + K_{эф}\left(K_{к} + \frac{T}{24}\cdot\frac{Ц}{365}\right) руб/т,$$

где $C_{эи}$ – эксплуатационные издержки; $K_{эф}$ – расчетный коэффициент эффективности капиталовложения; $K_{к}$ – капитальные вложения, приходящие на 1 т груза на протяжении участка; T – время следования; $Ц$ – цена одной тонны груза.

Результат решения ТЗ позволяет более полно производить оценку рационализации разных вариантов планов перевозок, с достаточно полной выраженностью количественно-одновременное влияние нескольких экономических факторов.

Сформулируем условие транспортной задачи (ТЗ) в общем виде.

Имеется несколько пунктов производства или сосредоточения какого-либо однородного продукта, а также некоторое количество пунктов, испытывающих потребность в этом продукте. Запасы в каждом пункте известны. Транспортные издержки, связанные с произвольной перевозкой, заданы. Необходимо составить такой план перевозок продукта из пунктов сосредоточения в пункты потребления, при котором весь объем материалов будет перевезен в пункты потребления, каждый пункт потребления будет полностью удовлетворен и суммарные транспортные затраты окажутся минимальными.

Переведем условие задачи на математический язык. Пусть A_i – пункты производства (их число m); B_j – пункты потребления (их число n); a_i – количество продукта в пункте A_i; а b_j – количество продукта в пункте B_j. Примем допущение, что стоимость перевозки продукции линейно зависит от количества перевозимой продукции. Пусть $c_{i\,j}(x)$ – стоимость перевозки x единиц продукта из A_i в B_j. С учетом наших допущений имеем: $c_{i\,j}(x) = c_{i\,j}x$. В качестве параметров управления выберем $x_{i\,j}$ – количество продукта, перевозимого из A_i в B_j. Тогда суммарные транспортные издержки (целевую функцию задачи) можно определить как:

$$f(x) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{i\,j}x_{i\,j}\ .$$

Составим ограничения. Необходимо вывести весь объем продукции из каждого пункта производства $\sum_{j=1}^{n} x_{i\,j} = a_i$ $(i = 1, 2, \ldots, m)$ и полностью удовлетворить потребность в продукте в пунктах спроса $\sum_{i=1}^{m} x_{i\,j} = b_j$ $(j = 1, 2, \ldots, n)$. Поскольку перевозки идут из пунктов производства в пункты потребления, то добавим ограничения $x_{i\,j} \geq 0$ для любых i и j $(\forall i, j)$.

В данной задаче производство полностью удовлетворяет потребление (выполняется условие баланса):

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

В результате получим математическую модель ТЗ:

$$f(x) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{i\,j}x_{i\,j} \to min;$$

$$\begin{cases} \sum_{j=1}^{n} x_{i\,j} = a_i,\ i = \overline{1, m}; \\ \sum_{i=1}^{m} x_{i\,j} = b_j,\ j = \overline{1, n}; \\ x_{i\,j} \ge 0,\ \ \forall i, j. \end{cases}$$

Решение *транспортной задачи* состоит из трех этапов:

✓ составление математической модели;

✓ нахождение опорного (первоначального) плана;

✓ нахождение оптимального плана.

Второй этап желателен, но необязателен. Для нахождения опорного плана обычно используют методы:

✓ «северо-западного угла»;

✓ метод наименьших стоимостей или минимального элемента.

П р и м е р. Имеются две базы с запасами $a_1 = 10$, $a_2 = 20$ и два магазина с потребностями: $b_1 = 5$, $b_2 = 25$. Стоимость перевозок единицы груза с баз в магазины задана: $c_{11} = 1$, $c_{12} = 2$, $c_{21} = 3$, $c_{22} = 1$. Необходимо составить математическую модель плана перевозок продукции с баз в пункты потребления, при котором все запасы на базах будут исчерпаны и каждый магазин будет полностью обеспечен необходимой продукцией, при этом суммарные транспортные затраты окажутся минимальными.

Р е ш е н и е. Поскольку условие баланса выполняется $(10 + 20 = 5 + 25)$, то ТЗ является закрытой и все ограничения записываются в виде равенств.

Пусть $x_{i\,j}$ – количество продукта, перевозимого из i-го пункта производства в j-й пункт потребления. Тогда:

$$F(x) = x_{11} + 2x_{12} + 3x_{21} + x_{22} \to min;$$

$$\begin{cases} x_{11} + x_{12} = 10; \\ x_{21} + x_{22} = 20; \\ x_{11} + x_{21} = 5; \\ x_{12} + x_{22} = 25; \\ x_{ij} \geq 0; \forall i, j. \end{cases}$$

Первое ограничение показывает, что с первой базы необходимо вывезти продукт в первый и второй магазины $x_{11} + x_{12}$ в количестве $a_1 = 10$. Аналогичен смысл и второго ограничения. Третье ограничение показывает, что в первый магазин необходимо привезти продукцию с первой и второй базы $x_{11} + x_{21}$ в количестве $b_1 = 5$. Аналогично и четвертое ограничение.

Составим первоначальный план перевозок методом наименьших стоимостей с двойным предпочтением. Исходные данные ТЗ представим в виде транспортной таблицы:

a_i / b_j	$b_1 = 5$	$b_2 = 25$
$a_1 = 10$	1**	2
$a_2 = 20$	3	1**

Отмечаем клетки с наименьшими стоимостями перевозок по каждой строке и каждому столбцу. Клетки, имеющие две отметки, заполняем в первую очередь. В клетке $a_1 b_1$ записываем число 5 и вычеркиваем полностью заполненный столбец b_1. В клетке $a_2 b_2$ записываем число 20 и вычеркиваем полностью заполненную строку a_2:

a_i / b_j	$b_1 = 5$	$b_2 = 25$
$a_1 = 10$	5	
$a_2 = 20$		20

Заполняем клетки с одной отметкой с учетом объемов баз и заявок магазинов, затем заполняем клетки без отметок. Осталась одна клетка $a_1 b_2$. В нее записываем число 5:

a_i / b_j	$b_1 = 5$	$b_2 = 25$
$a_1 = 10$	5	5
$a_2 = 20$		20

Получили опорный план методом минимизации, значение целевой функции рассчитываем по элементам опорного плана: $f(x) = 35$, $x_{11} = 5$, $x_{12} = 5$, $x_{22} = 20$. Заметим, что в данном примере опорный план совпал с оптимальным решением.

Рассмотрим открытую ТЗ, в которой не выполняется условие баланса.

Пример. Имеются два пункта производства товара $a_1 = 10$, $a_2 = 25$, и два пункта потребления $b_1 = 10$, $b_2 = 30$. Стоимость перевозок товара задана $c_{11} = 4$, $c_{12} = 2$, $c_{21} = 3$, $c_{22} = 1$. За недопоставку одной единицы товара первый пункт потребления терпит убыток, равный 1, а второй – 0.5. Необходимо составить математическую модель плана перевозок продукта из пунктов производства в пункты потребления, при котором весь продукт был бы перевезен в пункты потребления, а суммарные транспортные затраты и убытки оказались минимальными.

Решение. Введем обозначения. Пусть x_{ij} – количество товара, перевозимого из i-го пункта производства в j-й пункт потребления, $z_1 = b_1 - x_{11} - x_{21}$, $z_2 = b_2 - x_{12} - x_{22}$ – количество недопоставленного товара соответственно в первый и второй пункты потребления.

Составим математическую модель:

$$F(x) = 4x_{11} + 2x_{12} + 3x_{21} + x_{22} + z_1 + 0.5z_2 \to min;$$

$$\begin{cases} x_{11} + x_{12} = 10; \\ x_{21} + x_{22} = 25; \\ x_{11} + x_{21} \le 10; \\ x_{12} + x_{22} \le 30; \\ x_{ij} \ge 0; \forall i, j. \end{cases}$$

Прежде чем решать открытую ТЗ, ее необходимо сбалансировать. Вводим дополнительный фиктивный пункт производства a_{i+1} (в случае дефицита товара) или фиктивный пункт потребления b_{j+1} (в случае профицита товара). Мощность дополнительно введенного фиктивного пункта равна количеству недопоставленного или избыточного товара. Для нашего случая введем фиктивный пункт производства $a_3 = (30+10)-(25+10) = 5$.

Стоимость перевозок из фиктивного пункта производства равна нулю, поэтому целевая функция не меняется, а система ограничений принимает вид:

$$\begin{cases} x_{11} + x_{12} = 10; \\ x_{21} + x_{22} = 25; \\ x_{31} + x_{32} = 5; \\ x_{11} + x_{21} + x_{31} = 10; \\ x_{12} + x_{22} + x_{32} = 30; \\ x_{ij} \geq 0, \forall i, j. \end{cases}$$

Значение целевой функции $f(x) = 60$, при опорном плане:

$$x_{12} = 10; \qquad x_{21} = 5; \qquad x_{22} = 20.$$

Рассмотрим задачу большей размерности.

Пример. Завод имеет три цеха *A*, *B*, *C* и четыре склада №№ 1, 2, 3, 4. Цех *A* производит 30 тыс. изделий, цех *B* – 40 тыс. шт., цех *C* – 20 тыс. шт. за плановое время. Пропускная способность складов за то же время: склад № 1 – 20 тыс. шт., склад № 2 – 30 тыс. шт., склад № 3 – 30 тыс. шт., склад № 4 – 10 тыс. шт. Стоимость перевозок одной тысячи изделий в склады *№№ 1, 2, 3, 4 соответственно составляет: из цеха A – 2, 3, 3, 4 ед.; из цеха* *B* – 3, 2, 5, 1 ед.; из цеха *C* – 4, 3, 2, 6 ед. Составить план перевозок, при котором расходы на перевозку всех изделий минимальны.

Решение. Условие баланса выполняется $\sum_{i=1}^{3} a_i = \sum_{j=1}^{4} b_j = 90$, поэтому ТЗ является замкнутой, и все ограничения записываем в виде равенств.

Пусть $x_{i\,j}$ (тыс. шт.) – количество изделий, перевозимых из i-го цеха в j-й склад. Тогда математическая модель задачи будет иметь вид:

$$F(x) = 2x_{11} + 3x_{12} + 3x_{13} + 4x_{14} + 3x_{21} + 2x_{22} +$$
$$+ 5x_{23} + x_{24} + 4x_{31} + 3x_{32} + 2x_{33} + 6x_{34} \to min$$

$$\text{цеха:} \quad \begin{cases} x_{11} + x_{12} + x_{13} + x_{14} = 30; \\ x_{21} + x_{22} + x_{23} + x_{24} = 40; \\ x_{31} + x_{32} + x_{33} + x_{34} = 20, \end{cases}$$

$$\text{склады:} \quad \begin{cases} x_{11} + x_{21} + x_{31} = 20; \\ x_{12} + x_{22} + x_{32} = 30; \\ x_{13} + x_{23} + x_{33} = 30; \\ x_{14} + x_{24} + x_{34} = 10; \\ x_{ij} \ge 0. \end{cases}$$

Составим первоначальный план перевозок методом наименьших стоимостей с двойным предпочтением. Исходные данные ТЗ представим в виде транспортной таблицы. Отметим клетки с наименьшими стоимостями перевозок по каждой строке и каждому столбцу: Кажется в клетках b лишняя сточка – не получается убрать.

a_i / b_j	$b_1 = 20$	$b_2 = 30$	$b_3 = 30$	$b_4 = 10$
$a_1 = 30$	2**	3	3	4
$a_2 = 40$	3	2*	5	1**
$a_3 = 20$	4	3	2**	6

Клетки, имеющие две отметки, заполняем в первую очередь с учетом заявок производителей и потребителей: $x_{11} = 20$ (вычеркиваем первый полностью заполненный столбец), $x_{24} = 10$ (вычеркиваем четвертый столбец), $x_{33} = 20$ (вычеркиваем третью строку).

Затем заполняем клетки с одной отметкой с учетом заявок производителей и потребителей: $x_{22} = 30$ (вычеркиваем второй столбец и вторую строку).

Остается клетка $a_1 b_3 = x_{13} = 10$:

a_i / b_j	$b_1 = 20$	$b_2 = 30$	$b_3 = 30$	$b_4 = 10$
$a_1 = 30$	2** 20	3	3 10	4
$a_2 = 40$	3	2* 30	5	1** 0
$a_3 = 20$	4	3	2** 20	6

Получили начальный план перевозок:

$$x_{11} = 20;\ \ x_{13} = 10;\ \ x_{22} = 30;\ x_{24} = 10;\ x_{33} = 20;\ \ f(x) = 180 \text{ ед.}$$

2.5. Методы построения опорных планов

Существует несколько методов построения опорного плана при решении ТЗ:

✓ метод северо-западного угла;

✓ метод минимизации по строке (столбцу);

✓ метод Фогеля.

Рассмотрим основные положения каждого из них.

Метод северо-западного угла. В стандартных условиях ТЗ заданы возможности поставщиков a_{ij} и потребности потребителей b_j. Требуется найти допустимые объемы перевозки от каждого поставщика к каждому потребителю x_{ij}.

Шаг 1. Первая ячейка – с которой начинается распределение – будет «северо-западная» ячейка в левом верхнем углу таблицы (1-й поставщик, 1-й потребитель). Вписываем в эту ячейку максимальный объем, который позволяет запас поставщика и спрос потребителя (берем минимум между 20 и 30 *кг*, то есть 20 *кг*). Поскольку спрос 1-го потребителя полностью удовлетворен, ячейки соответствующего столбца заполняться больше не будут.

Шаг 2. Переходим в следующую «северо-западную» ячейку, не считая уже распределенной области. Этой ячейкой будет x_{12} (1-й поставщик, 2-й потребитель). Вписываем в эту ячейку максимальный объем, который позволяет запас поставщика и спрос потребителя (берем минимум между 30 и 10 *кг*, то есть 10 *кг*). Соответственно, уменьшаем оставшиеся не распределенными объемы поставки и потребления в строке и столбце на 10 *кг*. Запасы 1-го поставщика (в 1-й – верхней – строке) теперь исчерпаны, распределение по этой строке завершено.

Шаг 3. Переходим в следующую «северо-западную» ячейку, не считая уже распределенной области. Этой ячейкой будет x_{22} (2-й поставщик, 2-й потребитель). Вписываем в эту ячейку максимальный объем, который позволяет запас поставщика и спрос потребителя (берем минимум между 40 и 20 *кг*, то есть 20 *кг*). Соответственно, уменьшаем оставшиеся не распределенными объемы поставки и потребления в строке и столбце на 20 *кг*. Потребности 2-го потребителя теперь полностью удовлетворены, распределение по этому столбцу завершено.

Шаг 4. Переходим в следующую «северо-западную» ячейку, не считая уже распределенной области. Этой ячейкой будет x_{23} (2-й поставщик, 3-й потребитель). Вписываем в эту ячейку максимальный объем, который позволяет запас поставщика и спрос потребителя (берем минимум между 30 и 20 *кг*, то есть 20 *кг*). Соответственно, уменьшаем оставшиеся не распределенными объемы поставки и потребления в строке и столбце на 20 *кг*. Запасы 2-го поставщика (в 2-й сверху строке) теперь исчерпаны, распределение по этой строке завершено.

Шаг 5. Распределение оставшихся у последнего поставщика 20 *кг* груза по двум потребителям по 10 *кг*.

Результаты выполненных шагов, в условиях предыдущей задачи, то есть построенный опорный план – представлены в таблице.

	Потребитель B_1, потребность $20-20=0кг$	Потребитель B_2, Потребность $30-10-20=0кг$	Потребитель B_3, Потребность $30-10-20=0кг$	Потребитель B_4, потребность $10-10=0кг$
Поставщик A_1, запас $30-20-10=0кг$	$X_{11}=20кг$	$X_{12}=10кг$		
Поставщик A_2, запас $40-20-20=0кг$		$X_{22}=20кг$	$X_{23}=20кг$	
Поставщик A_3, запас $10-10=0кг$			$X_{33}=10кг$	$X_{34}=10кг$

При построении опорного плана по способу северо-западного угла совершенно не учитываются тарифы, потому план получается весьма далеким от оптимального. Для последующего нахождения оптимального решения ТЗ приходится делать много приближений (шагов).

М е т о д   м и н и м и з а ц и и   п о   с т р о к е . Этот способ учитывает тарифы перевозок и потому позволяет найти план, более близкий к оптимальному Суть метода заключается в том, что из всей таблицы стоимостей выбирают наименьшую, и в клетку, которая ей соответствует, помещают меньшее из чисел a_i, или b_j. Затем из рассмотрения исключают либо строку, соответствующую поставщику, запасы которого полностью израсходованы, либо столбец, соответствующий потребителю, потребности которого полностью удовлетворены, либо и строку и столбец, если израсходованы запасы поставщика и удовлетворены потребности потребителя.

Из оставшейся части таблицы стоимостей снова выбирают наименьшую стоимость, и процесс распределения запасов продолжают, пока все запасы не будут распределены, а потребности удовлетворены.

Запишем условие ТЗ в таблице и составим опорный план.

Поставщики	Потребители					Запасы
	b_1	b_2	b_3	b_4	b_5	
a_1	10	7	4	1 100	4	100
a_2	2 200	7 50	10	6	11	250
a_3	8	5	3	2	2 200	200
a_4	11	8 150	12 100	16	13 50	300
Потребности	200	200	100	100	250	850

Выбираем в таблице наименьшую стоимость (это стоимость, помещенная в клетке a_1, b_4) так как $a_1 = b_4 = 100$ ед. груза помещаем в этой клетке и исключаем из рассмотрения первую строку и четвертый столбец.

В оставшейся таблице стоимостей наименьшей является стоимость, расположенная в клетке a_2, b_1 и в клетке a_3, b_5. Заполняем любую из них, например, a_2, b_1. Имеем $200 < 250$, следовательно, записываем в нее 200 и исключаем из рассмотрения столбец b_1. В клетку a_3, b_5 записываем 200 ед. и исключаем из рассмотрения строку a_3. В оставшейся таблице стоимостей снова выбираем наименьшую стоимость и продолжаем процесс до тех пор, пока все запасы не будут распределены, а потребности удовлетворены. В результате получен опорный план: $X_{14} = 100$, $X_{21} = 200$, $X_{22} = 50$, $X_{35} = 200$, $X_{42} = 150$, $X_{43} = 100$, $X_{45} = 50$, остальные значения переменных равны нулю.

План не содержит циклов и состоит из семи положительных перевозок, следовательно, является вырожденным опорным планом. Определим его стоимость:

$$Z = 100 \cdot 1 + 200 \cdot 2 + 50 \cdot 7 + 200 \cdot 2 + 150 \cdot 8 + 100 \cdot 12 + 50 \cdot 13 = 4\ 300 \text{ (ед).}$$

Полученный план перевозок значительно ближе к оптимальному плану.

Недостатком метода является то, что при выборе минимального элемента в строке не учитывается стоимостной показатель по столбцу и наоборот, что не всегда экономически оправдано.

Метод наименьшего элемента. Суть метода в том, что в транспортной таблице сначала заполняются ячейки с наименьшими тарифами, а потом уже ячейки с большими тарифами. То есть выбираются перевозки с минимальной стоимостью доставки груза.

Шаг 1. Располагаем все клетки таблицы в очередь по мере возрастания тарифов, начиная с минимального.

Шаг 2. В клетку с минимальным тарифом записываем наибольшую возможную перевозку (исходя из запасов и потребностей), затем заполняем очередную по порядку клетку и т. д., пока не получим опорный план. При этом должен строго соблюдаться баланс по строкам и столбцам.

Пустые клетки прочеркиваем, а не заполняем нулями (чтобы было видно, что они не входят в план).

Если окажется, что есть несколько ячеек с одинаковыми и минимальными тарифами – выбираем любую из них. Он лучше плана, построенного по методу северо-западного угла, и для последующего нахождения оптимального плана потребуется меньше вычислений. Пример применения данного метода приведен в предыдущем подразделе.

Существенным недостатком рассмотренного метода является необходимость анализа всей таблицы исходных данных ТЗ, что очевидно неудобно при решении задач большой размерности.

Метод аппроксимации Фогеля. Первым делом добавляем к транспортной таблице дополнительные строку и столбец. Далее находим для каждой строки и каждого столбца абсолютные разности между двумя минимальными тарифами. Если в строке/столбце две клетки с одинаковыми и минимальными значениями тарифов, то берем именно их. Тогда разность будет равна 0. Найденные разности выписываем в добавочный столбец и добавочную строку. Среди вычисленных разностей (и по строкам, и по столбцам!) выбираем наибольшую. Затем в строке (или столбце), которой соответствует максимальная разность, ищем клетку с минимальным тарифом и заполняем ее. Если клеток с минимальным тарифом несколько, то заполняем ту из них, которой соответствует наибольшая разность. Затем повторяем все действия снова, уже не учитывая заполненные клетки.

Пример. Используя метод аппроксимации Фогеля, найти опорный план транспортной задачи, исходные данные которой приведены в таблице.

Пункты отправления	Пункты назначения				Запасы
	b_1	b_2	b_3	b_4	
a_1	7	8	1	2	160
a_2	4	5	9	8	140
a_3	9	2	3	6	170
Потребности	120	50	190	110	470

Для каждой строки и столбца таблицы условий найдем разности между двумя минимальными тарифами, записанными в данной строке или столбце, и поместим их в соответствующем дополнительном столбце или дополнительной строке таблица ниже.

Пункты отправления	Пункты назначения				Запасы	Разности по строкам					
	b_1	b_2	b_3	b_4							
a_1	7	8	1	2	160	1	6	–	–	–	–
a_2	4	5	9	8	140	1	1	1	1	1	0
a_3	9	2	3	6	170	1	1	1	–	–	–
Потребности	120	50	190	110	470						
Разности по столбцам	3	3	2	4							
	3	3	2	–							
	5	3	6	–							
	5	3	–	–							
	0	0	–	–							
	–	0	–	–							

Так, в строке a_2 минимальный тариф равен 4, а следующий за ним равен 5, разность между ними $5 - 4 = 1$. Точно так же разность между минимальными элементами в столбце b_4 равна $6 - 2 = 4$. Вычислив все эти разности, видим, что наибольшая из них соответствует столбцу b_4. В этом столбце минимальный тариф записан в клетке, находящейся на пересечении строки a_1 и столбца b_4. Таким образом, эту клетку следует заполнить. Заполнив ее, тем самым мы удовлетворим потребности пункта b_4. Поэтому исключим из рассмотрения столбец b_4 и будем считать запасы пункта a_1 равными $160 - 110 = 50$ ед. После этого определим следующую клетку для заполнения. Снова найдем разности между оставшимися двумя минимальными тарифами в каждой из строк и столбцов и запишем их во втором дополнительном столбце и во второй дополнительной строке таблицы. Как видно из этой таблицы, наибольшая указанная разность соответствует строке a_1. Минимальный тариф в этой строке записан в клетке, которая находится на пересечении ее со столбцом b_3. Следовательно, заполняем эту клетку. Поместив в нее число 50, тем самым предполагаем, что запасы в пункте a_1 полностью исчерпаны, а потребности в пункте b_3 стали равными 190 – 50 = 140 ед. Исключим из рассмотрения строку a_1 и определим новую клетку для заполнения. Продолжая итерационный процесс, последовательно заполняем клетки, находящиеся на пересечении строки a_3 и столбца b_3, строки a_3 и столбца b_2, строки a_2 и столбца b_1, строки a_2 и столбца b_2. В результате получим опорный план:

$$X = \begin{pmatrix} 0 & 0 & 50 & 110 \\ 120 & 20 & 0 & 0 \\ 0 & 30 & 140 & 0 \end{pmatrix}.$$

При этом плане общая стоимость перевозок такова:

$$Ц = 1 \cdot 50 + 2 \cdot 110 + 4 \cdot 120 + 5 \cdot 20 + 2 \cdot 30 + 2 \cdot 30 + 3 \cdot 140 = 1\,330.$$

Как правило, применение метода аппроксимации Фогеля позволяет получить либо опорный план, близкий к оптимальному, либо сам оптимальный план. Кстати, найденный выше опорный план транспортной задачи является и оптимальным. Во всех примерах опорный план содержит не более чем из $k + l - 1$ компонент, то есть не вырожденный. Это обусловлено тем, что с заполнением каждой клетки из рассмотрения исключается строка или столбец. Однако встречаются случаи, когда одновременное исключение и строки и столбца происходит не только в конце построения опорного плана, но и на промежуточных этапах. В результате число заполненных клеток становится меньше на одну. Подобные случаи называются вырождением плана. Вырожденные опорные планы практически не оптимизируются, так как методы оптимизации «зацикливаются» и происходит бесконечное число итераций, не приводящих к отысканию оптимума. Подсчитаем число занятых клеток таблицы в примере по методу Фогеля, их 6 и, согласно теории, должно быть $m + n - 1 = 6$. Следовательно, опорный план является невырожденным и ациклическим.

2.6. Построение оптимального плана перевозок

Рассмотрим методы оптимизации построенного опорного плана ТЗ.

С помощью рассмотренных выше методов построения начального плана можно получить вырожденный или невырожденный опорный план. Построенный план транспортной задачи как задачи линейного программирования можно было бы довести до оптимального с помощью симплексного метода. Однако из-за громоздкости симплексных таблиц, содержащих $m \times n$ неизвестных, и большого объема вычислительных работ для получения оптимального плана используют более простые методы. Одним из таких методов и является метод потенциалов.

Метод потенциалов – модификация симплекс-метода решения задачи линейного программирования применительно к транспортной задаче. Он позволяет, отправляясь от некоторого допустимого решения, получить оптимальное решение за конечное число итераций. Метод потенциалов, широко применяется на практике и используется при программировании.

Суть метода такова. На первом этапе составляют опорный план перевозок. На втором этапе строят систему потенциалов и проверяют начальный план на оптимальность. Каждому i-му поставщику устанавливается потенциал U_i, который можно интерпретировать как цену продукта в пункте поставщика, а каждому j-му потребителю устанавливается потенциал V_j, который характеризует цену продукта в пункте потребителя. Если опорный план является неоптимальным, то переходят к третьему этапу, суть которого заключается в корректировке плана прикрепления потребителей и поставщиков. Затем снова переходят ко второму этапу и т. д.

Алгоритм метода потенциалов:

Для улучшения уже имеющего опорного плана необходимо составить двойственную задачу:

$$u_1 \dots u_m \text{ и } v_1 \dots v_n \text{ любые значения;} \tag{1}$$

$$u_i + v_j \le c_{i\,j},\ j = 1,\ \overline{n};\, i = 1, \overline{m}; \tag{2}$$

$$T(u,v) = \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j \to max; \tag{3}$$

Пусть есть план: $X = \{x_{i\,j}\},\ \left(i = 1, \overline{m};\ j = 1, \overline{n}\right)$. (4)

Воспользуемся теоремой для дальнейшего построения алгоритма:

Теорема (критерий оптимальности): Для того чтобы допустимый план перевозок (4) в транспортной задаче был оптимальным, необходимо и достаточно, чтобы существовали такие числа $u_1 \dots u_m$ и $v_1 \dots v_n$, что:

$$u_i + v_j = c_{i\,j},\ \text{если } x_{i\,j} \ge 0; \tag{5}$$

$$u_i + v_j \le c_{i\,j},\ \text{если } x_{i\,j} = 0, \tag{6}$$

где u_i и v_j – потенциалы пункта назначения и отправления.

Сформулированная теорема позволяет построить алгоритм нахождения оптимального решения транспортной задачи:

Для опорного плана, в котором $m+n-1$ базисных клеток, можно определить потенциалы u_i и v_j так, чтобы выполнялось условие (5). Поскольку система содержит $m+n-1$ уравнений и $m+n$ неизвестных, то одну из них можно задать произвольно (например, приравнять к нулю). После этого из $m+n-1$ уравнений определяются остальные потенциалы и для каждой из свободных клеток вычисляются величины $u_i + v_j = c_{i\,j}$. Если оказалось, что $c'_{i\,j} \le c_{i\,j}$, то план оптимален. Если же хотя бы в одной свободной клетке $c'_{i\,j} > c_{i\,j}$, то план не является оптимальным и может быть улучшен путем переноса по циклу, соответствующему данной свободной клетке.

Процесс улучшения плана продолжается до тех пор, пока не будут выполнены условия (5) и (6). Рассмотрим вариант применения метода потенциалов.

Пример. Имеется сеть мостов, при построении которых требуются определенное количество металлоконструкций. Также имеется ряд баз поставщиков, где требуемый конструкции хранится, при этом на каждой базе различный объем хранимого товара. Кроме того, известны затраты на перевозку металлоконструкций от каждой базы к каждому мосту. Возникает потребность разработать такой план перевозок, чтобы мосты получили требуемое количество металлоконструкций с наименьшими затратами на транспортировку. Допустим, что опорный план перевозок уже составлен, но его необходимо улучшить. Именно в такой ситуации и применяется метод потенциалов.

Решение. Имеем задачу: улучшить опорный план перевозок, который является невырожденным и представлен в таблице.

<table>
<tr><td>Мосты/Базы</td><td colspan="2">b_1</td><td colspan="2">b_2</td><td colspan="2">b_3</td></tr>
<tr><td rowspan="2">a_1</td><td colspan="2">–</td><td colspan="2">–</td><td colspan="2">10</td></tr>
<tr><td></td><td>5</td><td></td><td>3</td><td></td><td>1</td></tr>
<tr><td rowspan="2">a_2</td><td colspan="2">15</td><td colspan="2">–</td><td colspan="2">5</td></tr>
<tr><td></td><td>3</td><td></td><td>2</td><td></td><td>4</td></tr>
<tr><td rowspan="2">a_3</td><td colspan="2">–</td><td colspan="2">20</td><td colspan="2">10</td></tr>
<tr><td></td><td>4</td><td></td><td>1</td><td></td><td>2</td></tr>
</table>

Для начала вычислим потенциалы для плана перевозок, представленного в таблице. Для этого сопоставим каждому поставщику a_i и каждому потребителю b_j величины u_i и v_j так, чтобы для всех базисных клеток плана было выполнено условие: $u_i + v_j = c_{ij}$. Результат показан в таблице 2.8.

Таблица 2.8. План перевозок с вычисленными потенциалами

<table>
<tr><td>Мосты/Базы</td><td colspan="2">b_1</td><td colspan="2">b_2</td><td colspan="2">b_3</td><td>U</td></tr>
<tr><td rowspan="2">a_1</td><td colspan="2">–</td><td colspan="2">–</td><td colspan="2">10</td><td rowspan="2">0</td></tr>
<tr><td>5</td><td>5</td><td>3</td><td>3</td><td>0</td><td>1</td></tr>
<tr><td rowspan="2">a_2</td><td colspan="2">15</td><td colspan="2">–</td><td colspan="2">5</td><td rowspan="2">3</td></tr>
<tr><td>0</td><td>3</td><td>– 1</td><td>2</td><td>0</td><td>4</td></tr>
<tr><td rowspan="2">a_3</td><td colspan="2">–</td><td colspan="2">20</td><td colspan="2">10</td><td rowspan="2">1</td></tr>
<tr><td>3</td><td>4</td><td>0</td><td>1</td><td>0</td><td>2</td></tr>
<tr><td>V</td><td colspan="2">0</td><td colspan="2">0</td><td colspan="2">1</td><td></td></tr>
</table>

Как видно из данных таблицы – план перевозок не удовлетворяет критерию оптимальности, поэтому не является оптимальным $c_{22} < 0$, поэтому его можно улучшить путем перераспределения поставок.

Для этого найдем ячейку с наибольшей по абсолютной величине отрицательной разностью c_{22} и построим цикл, в котором кроме этой ячейки все остальные являются базисными (такой цикл всегда существует и единственен). Отметим ячейку с отрицательной разностью c_{22} знаком «+», следующую знаком «–», и так далее, поочередно. Затем находим минимальное значение перевозки в ячейках цикла имеющих знак «–» и вписываем его в свободную ячейку со знаком «+». Затем последовательно обходим все ячейки цикла, поочередно вычитая и прибавляя к ним минимальное значение (в соответствии со знаками, которыми эти ячейки помечены: где минус – вычитаем, где плюс – прибавляем). Цикл повторяется до тех пор, пока не будет найдено оптимальное решение. Оптимальное решение, полученное с помощью метода потенциалов, для данной задачи представлено в таблице 2.9.

Таблица 2.9. Оптимальное решение

<table>
<tr><td>Мосты/Базы</td><td colspan="2">b_1</td><td colspan="2">b_2</td><td colspan="2">b_3</td><td>U</td></tr>
<tr><td rowspan="2">a_1</td><td colspan="2">–</td><td colspan="2">–</td><td colspan="2">10</td><td rowspan="2">0</td></tr>
<tr><td>4</td><td>5</td><td>3</td><td>3</td><td>0</td><td>1</td></tr>
<tr><td rowspan="2">a_2</td><td colspan="2">15</td><td colspan="2">5</td><td colspan="2">–</td><td rowspan="2">2</td></tr>
<tr><td>0</td><td>3</td><td>0</td><td>2</td><td>1</td><td>4</td></tr>
<tr><td rowspan="2">a_3</td><td colspan="2">–</td><td colspan="2">15</td><td colspan="2">15</td><td rowspan="2">1</td></tr>
<tr><td>2</td><td>4</td><td>0</td><td>1</td><td>0</td><td>2</td></tr>
<tr><td>V</td><td colspan="2">1</td><td colspan="2">0</td><td colspan="2">1</td><td></td></tr>
</table>

Поиск оптимального решения сводится к отысканию такой системы потенциалов, при которой выполняется условие оптимальности плана.

Рассмотрим еще один пример использования метода потенциалов.

П р и м е р . В условиях предыдущей задачи рассмотрим три базы и три моста, но с другими стоимостными показателями, представленными в таблице:

Мосты	Базы		
	b_1	b_2	b_3
a_1	6	3	4
a_2	3	7	2
a_3	2	8	6

Как следует спланировать перевозку, чтобы ее стоимость была минимальной?

Р е ш е н и е . Условие баланса выполняется: $\sum_{i=1}^{3} a_i = \sum_{i=1}^{3} b_j = 50$, поэтому ТЗ является закрытой и все ограничения записываем в виде равенств.

На первом этапе получаем опорный план перевозок методом наименьших стоимостей с двойным предпочтением.

a_1 / b_j	$b_1 = 20$	$b_2 = 12$	$b_3 = 18$	U_i
$a_1 = 15$	6 / 3	3** / 12	4	$U_1 = 0$
$a_2 = 25$	3 / 7	7	2** / 18	$U_2 = 3$
$a_3 = 10$	2** / 10	8	6	$U_3 = 4$
V_j	$V_1 = 6$	$V_2 = 3$	$V_3 = 5$	

Опорный план: $x_{11} = 3$, $x_{12} = 12$, $x_{21} = 7$, $x_{23} = 18$, $x_{31} = 10$, $f(x) = 131$.

На втором этапе построим систему потенциалов U_i и V_j для начального опорного плана.

Для занятых клеток составляем систему линейных уравнений вида $V_j = U_i + C_{ij}$:

$$V_1 = U_1 + 6;\ V_1 = U_2 + 3;\ V_1 = U_3 + 2;\ V_2 = U_1 + 3;\ V_3 = U_2 + 2.$$

Полагаем $U_1 = 0$, тогда $V_1 = 6$; $U_2 = 3$; $U_3 = 4$; $V_2 = 3$; $V_3 = 5$.

Найденные потенциалы указаны в вышеприведенной таблице.

Для оптимальности опорного плана необходимо и достаточно, чтобы система потенциалов U_i и V_j удовлетворяла условию:

$$V_j = U_i + C_{ij} \text{ для } x_{ij} > 0 \text{ (для занятой клетки);}$$

$$V_j \le U_i + C_{ij} \text{ для } x_{ij} = 0 \text{ (для незанятой клетки).}$$

Для незанятых клеток проверим условие оптимальности:

$$V_2 = 3 < U_2 + C_{22} = 3 + 7 = 10; \qquad V_2 = 3 < U_3 + C_{32} = 4 + 8 = 12;$$

$$V_3 = 5 > U_1 + C_{13} = 0 + 4 = 4; \qquad V_3 = 5 < U_3 + C_{33} = 4 + 6 = 10.$$

Третье неравенство, или клетка $a_1 b_3$, не удовлетворяет условию оптимальности, поэтому переходим к третьему этапу.

Для незанятых клеток, не удовлетворяющих условию оптимальности, находим наибольшую величину $\alpha_{ij} = V_j - U_i - C_{ij}$:

$$\alpha_{13} = V_3 - U_1 - C_{13} = 5 - 0 - 4 = 1.$$

Клетка a_1b_3 является положительной вершиной цикла. Строим замкнутый контур, начальная вершина которого находится в выбранной клетке, а остальные вершины контура – в занятых клетках. Линии контура могут быть горизонтальными либо вертикальными отрезками. Число отрезков и вершин будет четным. Первая вершина контура свободной клетки имеет знак «+», а в остальных вершинах контура расставляются поочередно знаки «–» и «+».

a_1/b_j	$b_1 = 20$	$b_2 = 12$	$b_3 = 18$	U_i
$a_1 = 15$	6 – 3	3 12	4 +	$U_1 = 0$
$a_2 = 25$	3 + 7	7	2 – 18	$U_3 = 3$
$a_3 = 10$	2 10	8	6	$U_3 = 4$
V_j	$V_1 = 6$	$V_2 = 3$	$V_3 = 5$	

Выбираем наименьшее из величин в вершинах контура с отрицательным знаком $min\{3{,}18\} = 3$. Уменьшаем на три единицы величину каждой вершины со знаком «–» и увеличиваем на три единицы величину каждой вершины со знаком «+». Получим:

a_1/b_j	$b_1 = 20$	$b_2 = 12$	$b_3 = 18$	U_i
$a_1 = 15$	6	3 12	4 3	$U_1 = 0$
$a_2 = 25$	3 10	7	2 15	$U_2 = 2$
$a_3 = 10$	2 10	8	6	$U_3 = 3$
V_j	$V_1 = 5$	$V_2 = 3$	$V_3 = 4$	

Новый опорный план:

$$x_{12} = 12;\ \ x_{13} = 3;\ \ x_{21} = 10;\ \ x_{23} = 15;\ \ x_{31} = 10;\ \ f(x) = 128.$$

Найдем потенциалы и проверим новый опорный план на оптимальность. Для занятых клеток составляем систему линейных уравнений вида $V_j = U_i + C_{ij}$:

$$V_1 = U_2 + 3; \quad V_1 = U_3 + 2; \quad V_2 = U_1 + 3, \quad V_3 = U_1 + 4, \quad V_3 = U_2 + 2.$$

Полагаем $U_1 = 0$, тогда $V_2 = 3$, $V_3 = 4$, $U_2 = 2$, $V_1 = 5$, $U_3 = 3$.

Найденные потенциалы указаны в вышеприведенной таблице.

Для незанятых клеток проверим условие оптимальности:

$$V_1 = 5 < U_1 + C_{11} = 0 + 6 = 6; \qquad V_2 = 3 < U_2 + C_{22} = 2 + 7 = 9;$$

$$V_2 = 3 < U_3 + C_{32} = 3 + 8 = 11; \qquad V_3 = 4 < U_3 + C_{33} = 3 + 6 = 9.$$

Полученный опорный план $x_{12} = 12$, $x_{13} = 3$, $x_{21} = 10$, $x_{23} = 15$, $x_{31} = 10$, удовлетворяет всем условиям оптимальности и целевая функция принимает значение $f(x) = 128$.

Распределительный метод состоит в последовательном улучшении опорного плана перевозок путем отыскания на каждом шаге выгодных циклов переноса грузов. Опорный план для данного метода можно сформировать, применяя метод «северо-западного» угла. Более подробно рассмотрим процесс формирования очередного цикла переноса на каждом новом шаге алгоритма. Очевидно, что при перемещении x единиц груза по некоторому циклу с ценой g стоимость перевозок изменяется на величину $x \times g$. Тогда, для улучшения текущего плана перевозок имеет смысл перемещать перевозки только по тем циклам, цена которых отрицательна. Если циклов с отрицательной ценой в таблице больше не осталось, это означает, что оптимальный план достигнут. При улучшении плана циклическими переносами пользуются приемом, заимствованным из симплекс-метода: на каждом шаге (цикле) заменяют одну свободную переменную на базисную, т. е. заполняют одну клетку и взамен того освобождают одну из базисных клеток.

Можно доказать, что для любой свободной клетки транспортной таблицы всегда существует цикл (и притом единственный), одна из вершин которого лежит в этой клетке, а все остальные в базисных клетках. Если цена такого цикла, с плюсом в свободной клетке, отрицательна, то план можно улучшить. Количество единиц груза (x), которые можно переместить, определяется минимальным значением перевозок, стоящих в отрицательных вершинах цикла.

П р и м е р. Распределительным методом найти оптимальный план перевозок транспортной задачи, имеющей следующую таблицу издержек:

Сток/ Исток	B_1	B_2	B_3	B_4	Запасы
A_1	10	7	6	8	31
A_2	5	6	5	4	48
A_3	8	7	6	7	38
Заявки	22	34	41	20	117

Р е ш е н и е.

1. Методом «северо-западного» угла найдем опорный план перевозок:

Сток/ Исток	B_1	B_2	B_3	B_4	Запасы
A_1	10 22	7 9	6	8	31
A_2	5	6 25	5 23	4	48
A_3	8	7	6 18	7 20	38
Заявки	22	34	41	20	117

Опорный план имеет шесть базисных клеток в соответствующей ему транспортной таблице, что позволяет его использовать без модификаций для дальнейшего решения задачи: $n+m-1=4+3-1=6$. Посчитаем стоимость найденного опорного плана: $L=22\cdot 10+9\cdot 7+\ldots=796$.

2. Попытаемся улучшить найденный опорный план перевозок методом циклических переносов. Вычислим цену цикла для каждой свободной клетки.

Количество свободных клеток в транспортной таблице данного опорного плана равно: $k = 3 \cdot 2 = 6$.

$$\gamma_{13} = 6 - 5 + 6 - 7 = 0;\ \gamma_{14} = 8 - 7 + 6 - 5 + 6 - 7 = 1;\ \gamma_{21} = 5 - 10 + 7 - 6 = -4;$$

$$\gamma_{24} = 4 - 7 + 6 - 5 = -2;\ \gamma_{31} = 8 - 10 + 7 - 6 + 5 - 6 = -2;\ \gamma_{32} = 7 - 6 + 5 - 6 = 0.$$

3. Для всех свободных переменных (клеток) с отрицательной ценой цикла вычислим максимальное количество груза, которое можно перенести по соответствующему циклу. Очевидно, что максимальное количество груза, которое можно переместить по некоторому выбранному циклу будет равно минимальному значению груза среди отрицательных клеток цикла:

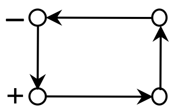
$$max\, x_{21} = min(22, 25) = 22; \qquad max\, x_{24} = min(20, 23) = 20;$$

$$max\, x_{31} = min(22, 25, 18) = 18.$$

4. Теперь для всех свободных переменных с отрицательной ценой циклов вычислим характеристику $y_{i\,j} = max\, x_{i\,j}$.

$i\,j$	2; 1	2; 4	3; 1
y	– 4	– 2	– 2
x	22	20	18
y_x	– 88	– 40	– 36

Выберем ту свободную переменную, которой соответствует наименьшее значение величины $y_{i\,j} = max\, x_{i\,j}$ и перенесем $max\, x_{i\,j}$ единиц груза по циклу, соответствующему выбранной переменной, получим:



Сток/ Исток	B_1	B_2	B_3	B_4	Запасы
A_1	10 22	7 9	6	8	31
A_2	5	6 25	5 23	4	48
A_3	8	7	6 18	7 20	38
Заявки	22	34	41	20	117

Таким образом, мы уменьшим значение целевой функции стоимости плана перевозок на 88 единиц.

Новому улучшенному плану перевозок будет соответствовать следующая таблица перевозок:

Сток/ Исток	B_1	B_2	B_3	B_4	Запасы
A_1	10	7 31	6	8	31
A_2	5 22	6 3	5 23	4	48
A_3	8	7	6 18	7 20	38
Заявки	22	34	41	20	117

Полученный на данном этапе первой итерации алгоритма новый план перевозок имеет шесть базисных клеток в соответствующей ему транспортной таблице, что позволяет его использовать без модификаций для дальнейшего решения задачи. Стоимость найденного плана перевозок равна: $L = 31 \cdot 7 + 22 \cdot 5 + ... = 708$, что несколько лучше результата опорного плана. Попытаемся опять улучшить найденный опорный план перевозок. Для этого перейдем к пункту 2 алгоритма.

Вычислим цену цикла для каждой свободной переменной:

$$\gamma_{13} = 6 - 5 + 6 - 7 = 0; \qquad \gamma_{14} = 8 - 7 + 6 - 5 + 6 - 7 = 1;$$

$$\gamma_{11} = 10 - 7 + 6 - 5 = 4; \qquad \gamma_{24} = 4 - 7 + 6 - 5 = -2;$$

$$\gamma_{31} = 8 - 5 + 5 - 6 = 2; \qquad \gamma_{32} = 7 - 6 + 5 - 6 = 0.$$

Найдем $max\, x_{24} = min(20, 23) = 20$.

Для единственной свободной переменной рассчитаем значение критерия $y_{i\,j} = max\, x_{i\,j} : y_{24} = (max\, x_{24}) = -40$.

Сток/ Исток	B_1	B_2	B_3	B_4	Запасы
A_1	10	7 31	6	8	31
A_2	5 22	6 3	5 23	4	48
A_3	8	7	6 18	7 20	38
Заявки	22	34	41	20	117

Перенесем $max\, x_{24} = 20$ единиц груза по циклу (2,4)+, (3,4)-, (3,3)+, (2,3)-, уменьшив этим значение целевой функции на 40 единиц. Новому улучшенному плану будет соответствовать следующая таблица перевозок:

Сток/ Исток	B_1	B_2	B_3	B_4	Запасы
A_1	10	7 31	6	8	31
A_2	5 22	6 3	5 3	4 20	48
A_3	8	7	6 38	7	38
Заявки	22	34	41	20	117

Полученный новый план перевозок имеет шесть базисных клеток в соответствующей ему транспортной таблице, что позволяет его использовать без модификаций для дальнейшего решения задачи.

Стоимость найденного плана перевозок равна:

$$L = 31 \cdot 7 + 22 \cdot 5 + \ldots = 668.$$

Попытаемся опять улучшить найденный опорный план перевозок. Для этого перейдем к пункту 2 алгоритма: вычислим цену цикла для каждой свободной переменной:

$\gamma_{13} = 6 - 5 + 6 - 7 = 0$; $\gamma_{14} = 8 - 4 + 6 - 7 = 3$; $\gamma_{11} = 10 - 7 + 6 - 5 = 4$;

$\gamma_{34} = 7 - 6 + 5 - 4 = 2$; $\gamma_{31} = 8 - 5 + 5 - 6 = 2$; $\gamma_{32} = 7 - 6 + 5 - 6 = 0$.

Так как не существует циклов свободных переменных с отрицательной ценой, полученный план перевозок улучшить нельзя, так как он является оптимальным. Стоимость этого плана перевозок, как было посчитано ранее, составляет 668 единиц.

2.7. Оптимальные назначения

При организации производственных работ важно место занимает распределение исполнителей по объектам и видам работ. Необходимо отметить, что задача о назначении есть полностью вырожденная транспортная задача. Действительно, при любом назначении исполнителя на работу автоматически «поставки» по строке совпадают со «спросом» по столбцу и вместо $2n-1$ получаем всего n ненулевых значений x_{ij}. Следовательно, нужно дополнить матрицу $(n-1)$ некоторыми, достаточно малыми, но неизвестными величинами для того, чтобы избавиться от вырожденности. Однако это требует громоздких вычислений и гораздо удобнее пользоваться специальными методами, разработанными для такого типа задач. Рассмотрим некоторые из них.

Решение стандартной задачи о назначении по алгоритму Флада. Отличительной особенностью таких задач является то, что ресурсы и объемы работ выражаются в различных единицах измерения (машино-сменами, человеко-днями и т. д.). Это не позволяет для их решения использовать рассмотренные выше методы.

Пусть имеются m исполнителей, которые могут выполнять различные m работ. Известна эффективность C_{ij} i-го исполнителя при выполнении j-й работы. Необходимо определить, какого исполнителя и на какую работу следует назначить, чтобы добиться минимального общего количества машино-смен, необходимых для выполнения всего комплекса работ при условии, что каждый исполнитель может быть назначен только на одну работу.

Для составления математической модели обозначим через $x_{i\,j}$ назначение i-го исполнителя на j-й работу. Так как количество исполнителей равно количеству работ, то каждый из них может быть назначен только на одну работу, поэтому $\sum_{i=1}^{m} x_{i\,j} = 1$ и $\sum_{j=1}^{m} x_{i\,j} = 1$. При назначении i-го исполнителя на j-ю работу эффективность равна $a_{i\,j}x_{i\,j}$.

Таким образом, приходим к следующей постановке вариации задачи линейного программирования:

$$x_{i,j} = x_{i,j}^2,\ \ i, j = 1,2,..m; \tag{1}$$

$$\sum_{i=1}^{m} x_{i\,j} = \sum_{j=1}^{m} x_{i\,j} = 1; \tag{2}$$

$$T = \sum_{i,j}^{m} a_{i\,j}^{(0)} x_{i\,j} = min. \tag{3}$$

Условие (1) – один означает, что величина, $x_{i\,j}$ принимает только два значения: либо 1, если i-й исполнитель назначается для выполнения j-й работы; либо 0, если i-й исполнитель не назначается для выполнения j-й работы.

Условие (2) – отражает вид итоговой матрицы назначений, то есть в каждой строке и в каждом столбце может быть только одна единица.

В условии (3) – индекс (0) означает, что значение стоимостной характеристики $a_{i\,j}$ взято из исходной матрицы.

Из постановки задачи понятно, что ее решением является квадратная матрица. Если в исходных данных количество исполнителей больше количества работ, то вводят либо фиктивную работу с нулевой производительностью, либо вторую часть ограничений записывают в виде неравенств.

Алгоритм М. Флада основан на двух теоремах. Первая утверждает, что решение задачи не изменится, если к любому столбцу (строке) исходной матрицы добавить (отнять) некую константу. Этим способом ряд величин $a_{i\,j}^{(0)}$ сводится к нулю, что обеспечивает минимизацию решения. Вторая теорема определяет условие оптимальности решений, получаемых на очередном шаге, состоящее в том, что минимальное число линий содержащих все нули, равно максимальному числу таких нулей. При этом никакие два из них не лежат на одной и той же прямой. Из комбинаторики известно, что для n^2 матрицы возможно получить $n!$ вариантов решения. Алгоритм Флада позволяет получить оптимальное решение задачи о назначении за сравнительно малое число шагов.

Пример. Рассмотрим оценочную таблицу (матрицу) отражающую затраты машино-смен на выполнение отдельных работ.

Исполнители, i	Работы, j				
	1	2	3	4	5
	Трудоемкость, машино-смен				
1	2	9	2	7	2
2	6	8	7	6	1
3	4	6	5	3	1
4	4	2	7	3	1
5	5	3	9	5	1

Анализируя данные таблицы, заметим, что есть возможность вычесть из всех строк минимальный элемент равный 1. В результате получим таблицу:

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

Проведем вычислительные действия с новой таблицей: вычтем минимальный элемент равный единице (1) из ненулевых элементов 1-го, 2-го, 3-го и 5-го столбцов, а из 4-го вычтем элемент равный двум (2). В результате получим новую таблицу:

0	7	0	4	|0
4	6	5	3	|0
2	4	3	0	|0
2	0	5	0	|0
3	1	7	2	|0

Проведем наименьшее из возможных число горизонтальных и вертикальных линий, пересекающих, по крайней мере, один раз все нули. Таких линий получилось четыре (4).

В теоретическом обосновании алгоритма Флада доказано, что если число таких линий меньше n, то полученная матрица не дает оптимального решения. Для дальнейших действий выбираем наименьший элемент, через который не проведена ни одна линия. В данном случае это единица во втором столбце. Следующим шагом вычитаем это значение из всех элементов, через которые не проведена ни одна линия, и прибавляем его ко всем остальным элементам, через которые проведены две линии. Таким образом, получим:

0	|7	0	|4	|1
3	|5	4	|2	|0
2	|4	3	|0	|1
2	|0	5	|0	|1
2	|0	6	|1	|0

Проверяем полученный вариант на оптимальность, легко заметить, что для этого опять достаточно четырех линий, то есть полученное решение не является оптимальным. Наименьшее из не вычеркнутых значений равно два.

Повторим действия, аналогичные предыдущему шагу и получим новый вариант назначений:

|0	|7	0	|6	|3
|1	|5	2	|2	|0
|0	|4	1	|0	|1
|0	|0	3	|0	|1
|0	|0	4	|1	|0

Вновь проверим план на оптимальность и установим, что для вычеркивания всех нулей в таблице требуется минимально 5 линий, то есть их число равно числу строк (столбцов) матрицы, что определяет условие оптимальности полученного варианта назначения машин на определенные виды работ (в таблице клетки назначений содержат выделенные курсивом элементы – нули). Заметим, что полученное решение не является единственным. Существуют еще три варианта выбора из имеющихся в таблице 11 нулей необходимых пяти так, чтобы в каждом столбце был только один ноль.

Полученное решение приведено в таблице:

Исполнители, *i*	Работы, *j*				
	1	2	3	4	5
	Трудоемкость, машино-смен				
1			1		
2					1
3				1	
4		1			
5	1				

Это решение дает оптимальное значение целевой функции: $T = 5 + 2 + 2 + 3 + 1 = 13$ (машино-смен). Полученный вариант назначений не является единственным. При отыскании других вариантов назначений заметим, что в эти решения должны войти нули, стоящие клетках последней таблицы как единственные в третьем столбце и во второй строке. Отдельно отметим, что значение целевой функции при каждом варианте сохраняется.

Если в подобных задачах требуется найти максимум целевой функции, например, наибольшую эффективность или выработку при использовании машин, то оценочная матрица $a_{i\,j}^{(0)}$ преобразуется по правилу:

$$a_{i\,j}^{(0+)} = a_{i\,jmax}^{(0)} - a_{i\,j}^{(0)},$$

где $a_{i\,jmax}^{(0)}$ – наибольшее из значений $a_{i\,j}$ стоящее в исходной оценочной матрице. Далее задача решается в соответствии с приведенным алгоритмом Флада и обеспечит поиск максимума целевой функции.

Решение задачи о назначении венгерским методом. Алгоритм был разработан и опубликован Гарольдом Куном в 1955 г. Сам Кун дал алгоритму название «венгерский», потому что он был в значительной степени основан на более ранних работах двух венгерских математиков: Денеша Кенига и Эйгена Эгервари. В 1957 г. Джеймс Манкрес показал, что этот алгоритм работает за (строго) полиномиальное время (т. е. за время порядка полинома, не зависящего от величины стоимостей). Поэтому в литературе данный алгоритм известен не только как «венгерский», но и как «алгоритм Куна-Манкреса». Впрочем, в 2006 г. выяснилось, что точно такой же алгоритм был изобретен за 100 лет до Куна немецким математиком Карлом Якоби. Дело в том, что его работа «About the research of the order of a system of arbitrary ordinary differential equations», напечатанная посмертно в 1890 г., содержавшая помимо прочих результатов и полиномиальный алгоритм решения задачи о назначениях, была написана на латыни, а ее публикация прошла незамеченной среди математиков.

Пример. По условиям предыдущей задачи, но с иной платежной матрицей решим задачу о назначении венгерским методом.

Решение. Исходная таблица (матрица):

2	4	1	3	3
1	5	4	1	2
3	5	2	2	4
1	4	3	1	4
3	2	5	3	5

Этап 1. В каждой строке ищем минимальный элемент (выделяем жирным в таблице) и отнимаем от всех элементов строки. Получим:

1	3	0	2	2
0	4	3	0	1
1	3	0	0	2
0	3	2	0	3
1	0	3	1	3

Теперь проводим аналогичную процедуру для всех столбцов: ищем наименьший элемент по столбцу (выделяем жирным в таблице) и отнимаем его из всех элементов столбца. Получим:

1	3	0	2	1
0	4	3	0	0
1	3	0	0	1
0	3	2	0	2
1	0	3	1	2

Целью является распределение всех подлежащих назначению единиц в клетки с нулевой стоимостью.

Этап 2. Выбираем строку с одним нулем (строка № 1), выделяем нуль жирным и зачеркиваем оставшиеся нулевые значения этого столбца (столбца № 3).

Далее выбираем строку с одним нулевым значением (строка № 5), выделяем нуль. Опять выбираем строку с одним нулем (строка № 3), выделяем нуль жирным и зачеркиваем (выделено серым) оставшиеся нулевые значения этого столбца (столбца № 4).

Аналогично, выбираем строку с одним нулем (строка № 4), выделяем нуль жирным и зачеркиваем (выделено серым) оставшиеся нулевые значения этого столбца (столбца № 1).

И, наконец, выбираем строку с одним нулевым значением (строка № 2), выделяем нуль, получим:

1	3	0	2	1
0	4	3	0	0
1	3	0	0	1
0	3	2	0	2
1	0	3	1	2

Получаем оптимальную матрицу назначений:

		1		
				2
			2	
1				
	2			

Значение целевой функции – стоимость (рациональность, время работ и т. д.) такого назначения составит: $1+1+2+2+2=8$. Как видно венгерский метод позволил получить оптимальное решение с меньшим количеством итераций и без дополнительных проверок на оптимальность.

Задача о нахождении оптимальной последовательности работ. Вопросы об определении последовательности выполняемых работ можно решить распределительными методами, однако более эффективным является специальный метод, носящий название «алгоритм Флада для распределительных задач».

Рассмотрим применение алгоритма на примере конкретной задачи. Имеется несколько объектов, на которых открыт фронт работ. Сначала выполняются подготовительные работы, затем основные. Время, требуемое на выполнение каждого вида работ на каждом из объектов, приведено в таблице.

Объект, i	Время подготовительных работ, t_i	Время основных работ, T_i
1	3	6
2	7	2
3	4	7
4	5	3

Предполагается, что работы можно развернуть на любом объекте. Для простоты положим, что время перемещения с объекта на объект несравнимо мало по отношению ко времени выполнения работ, поэтому им можно пренебречь. Заметим, что данное допущение не является обязательным для применения метода по алгоритму Флада.

Теоретически для распределительных задач установлено общее число возможных вариантов последовательностей строительства объектов. Оно равно числу перестановок $n!$. В рассматриваемом примере это значение составляет $4!=24$ варианта. Какой из них наиболее удачный или оптимальный?

На рисунке 2.2 показана последовательность работ в порядке возрастания номера объекта.

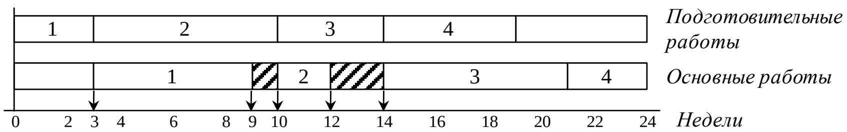


Рис. 2.2. Схема выполнения произвольной последовательности работ

Общее время выполнения всех работ составляет 24 недели и, очевидно, не является оптимальным. Применим алгоритм Флада.

Ш а г 1. Находим в таблице исходных данных объект с минимальным временем работы t_{imin} или T_{imin}. В пример это объект № 2 с $T_2=2$ недели. Так как это наименьшее время относится ко времени выполнения основной работы, то этот объект ставят на последнее место по подготовительным работам. Смысл такого действия вполне понятен – после завершения на объекте № 2 подготовительных работ основные будут закончены в кратчайший срок. Имеет место минимальный период свертывания работ. Если бы наименьшее время относилось ко времени выполнения подготовительных работ, то данный объект следовало бы поставить на первое место при разворачивании всех работ. Смысл такого действия также очевиден.

Шаг 2. Продолжаем анализ остальных объектов. Следующие t_{imin} или $T_{i\,min}$ со значениями 3 недели относятся к объекту № 1 – время подготовителных работ и объекту № 4 – время основных работ, можно расставлять любой из них в последовательность. Объект № 1 – на первое место, а объект № 4 на предпоследнее (так как последнее место уже отведено объекту № 2).

Шаг 3. Выполняя аналогичные действия, определяем всю последовательность работ: 1 – 3 – 4 – 2. Результатом является ликвидация простоев из-за отсутствия фронта работ, что приводит к оптимальному времени в 21 неделю. Экономия составила 3 недели, что существенно. На рисунке 2.3. показана оптимальная последовательность ввода объектов.

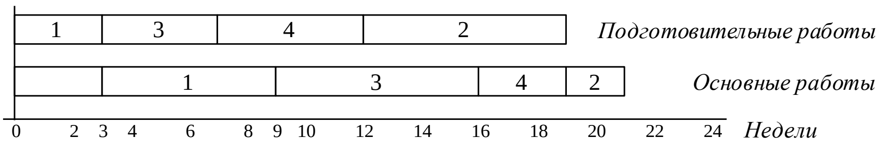


Рис. 2.3. Оптимальная последовательность выполнения работ на объектах

Если исключить из постановки задачи допущение о незначительности времени перемещения с объекта на объект, то следует этот параметр включить (равномерно) во временные показатели t_i и T_i. Алгоритм Флада может применяться и для более сложных случаев, так если видов работ больше двух, то возможно их различное чередование.

Вопросы самоконтроля:

1. Понятие «задачи математического программирования».
2. Условия для задач линейного программирования.
3. Модели задач принятия решений: общая схема.
4. Параметры управления.
5. Основные понятия и определения: задача оптимизации общего вида; целевая функция; ограничения; оптимальное решение; точность.
6. Классификация ЗО по виду ЦФ и ограничений.
7. Минимизация при ограничениях.
8. Общая задача линейного программирования (ЗЛП).
9. Стандартная задача линейного программирования.
10. Каноническая форма ЗЛП.
11. Оптимальные решения.
12. Двойственная задача линейного программирования.
13. Экономическая интерпретация двойственной задачи.
14. Сравнение прямой и двойственной ЗЛП.
15. Графический метод решения основной ЗЛП.
16. Область решений ЗЛП.
17. Оптимальность многокритериальных задач.
18. Способы графического отыскания оптимального решения.
19. Оптимальность по Парето.
20. Графический метод решения для целочисленных задач.
21. Прямые методы безусловной многомерной оптимизации. Симплекс-метод.
22. В чем суть перехода в симплекс-методе к канонической форме ЗЛП.
23. Базисные и искусственные переменные.
24. Алгоритм ввода новой базисной переменной.
25. Вывод из базиса искусственной переменной.
26. Геометрический смысл вывода из базиса искусственных переменных.
27. Критерий проверки плана на оптимальность.
28. Достоинство и недостатки симплекс-метода.
29. Транспортная задача (ТЗ). Математическая постановка.
30. Закрытая и открытая транспортные задачи.
31. Особенность поставки транспортных задач.
32. Показатели оптимальности при решении ТЗ.
33. Этапы решения ТЗ.
34. Приведение открытой ТЗ к закрытой, фиктивные переменные.
35. Методы построения опорного плана: северо-западного угла, минимизации по строке, наименьших стоимостей с двойным предпочтением, метод Фогеля.
36. Достоинства и недостатки методов построения опорного плана.
37. Оптимизация опорных планов ТЗ.
38. Метод потенциалов и алгоритм его реализации.
39. Основная теорема метода потенциалов.
40. Условие оптимальности в методе потенциалов.

41. Распределительный метод улучшения опорного плана.
42. Особенности постановки задач об оптимальных назначениях.
43. Алгоритм Флада для стандартной задачи о назначениях.
44. Смысл условий ЗЛП в вариации назначений.
45. Теоретические основания алгоритма Флада.
46. Условия оптимальности, используемые при реализации алгоритма Флада.
47. Преобразование оценочной матрицы при переходе к поиску максимума целевой функции.
48. Венгерский метод решения задачи о назначениях.
49. Нахождение оптимальной последовательности выполнения работ.
50. Алгоритм Флада для распределительных задач. Получение оптимального решения.

Раздел 3. МЕТОДЫ ОДНОМЕРНОЙ ОПТИМИЗАЦИИ

В этом разделе рассматривается простейшая математическая модель оптимизации, в которой целевая функция зависит от одной переменной, а допустимым множеством является отрезок вещественной оси:

$$f(x) \to min, \qquad x \in [a; b] \tag{1}$$

К математическим задачам вида (1) приводят прикладные задачи оптимизации с одной управляемой переменной. Кроме того, необходимость в минимизации функций одной переменной возникает при реализации некоторых методов решения более сложных задач. Как уже отмечалось, максимизация целевой функции ($f(x) \to max$) эквивалентна минимизации противоположной величины ($- f(x) \to min$), поэтому будем рассматривать только задачи минимизации.

3.1. Предварительные сведения

Пусть функция $f(x)$ определена на множестве X вещественной оси R.

1. Число $x^* \in X$ называется точкой глобального (абсолютного) минимума или просто точкой минимума функции $f(x)$ на множестве X, если $f(x^*) \le f(x)$ для всех $x \in X$. Значение $f^* = f(x^*) = \min\limits_X f(x)$ называют *глобальным* (абсолютным) *минимумом* или просто минимумом функции $f(x)$ на множестве X. Множество всех точек минимума $f(x)$ на X будем в дальнейшем обозначать через X^*.

2. Число $\tilde{x} \in X$ называется точкой локального минимума функции $f(x)$, если $f(\tilde{x}) \le f(x)$ для всех $x \in X$, достаточно близких к $\tilde{x}$, т. е. если

существует $\varepsilon > 0$ такое, что это неравенство выполняется для любого $x \in \{x | x \in X, |x - \tilde{x}| < \varepsilon\}$.

З а м е ч а н и я:

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

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

Например:

✓ если $f(x) = \ln x$, $X = (0; 1]$, то $X^* = \varnothing$;

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

✓ если $f(x) = \sin^2 \pi x$, $X = R$, то $X = Z$ – бесконечное множество.

Если множество точек минимума функции $f(x)$ на X – пусто, то задача минимизации $f(x)$ теряет смысл. В этом случае можно ограничиться поиском точки $\tilde{x}^* \in X$, в которой значение $f(x)$ с заданной погрешностью ε приближает точную нижнюю грань функции $f(x)$ на множестве X, т. е.:

$$f(\tilde{x}^*) - f_* < \varepsilon .$$

Широкий класс функций, для которых $X^* \neq \varnothing$, определяет известная из математического анализа теорема Вейерштрасса, согласно которой непрерывная на замкнутом ограниченном множестве функция достигает на этом множестве своих минимального и максимального значения.

Если функция $f(x)$ на множестве X имеет, кроме глобального еще и локальные минимумы, отличные от него, то минимизация $f(x)$, как правило, сильно затрудняется. В частности, многие методы поиска точки минимума $f(x)$ приспособлены только для 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, $a \le \alpha \le \beta \le b$, такие, что:

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

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

✓ при $x \in [\alpha;\beta]$ $\quad f(x) = f^* = \min\limits_{[a;b]} f(x)$.

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 рисунке 3.1.

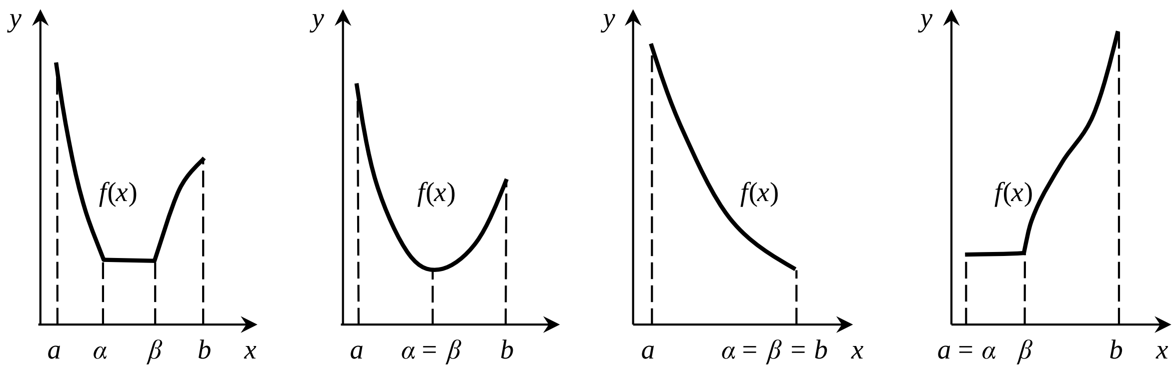


Рис. 3.1. Графики унимодальных функций на заданных интервалах

Из определения следуют основные 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 $[a;b]$, является унимодальной и на любом меньшем отрезке $[c;d] \subset [a;b]$.

3. Пусть $f(x) \in Q[a;b]$ и $a \le x_1 < x_2 \le b$. Тогда если:

✓ $f(x_1) \le f(x_2)$, то $x^* \in [a;x_2]$;

✓ $f(x_1) > f(x_2)$, то $x^* \in [x_1;b]$, где x^* – точка минимума $f(x)$ на отрезке $[a;b]$.

Определение. 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:

$$f[\alpha x' + (1-\alpha)x''] \le \alpha f(x') + (1-\alpha) f(x'') \tag{2}$$

Основные свойства выпуклых функций:

1. Если функция $f(x)$ *выпукла* на $[a;b]$, то на любом отрезке $[x';x''] \subset [a;b]$ ее график расположен не выше хорды, проведенной через точки графика с абсциссами x' и x'' (Рис. 3.2). Пусть x' и x'' – произвольные точки отрезка $[a;b]$, причем $x' < x''$. Легко проверить, что при любом $\alpha \in [0;1]$ точка $x_\alpha = \alpha x' + (1-\alpha)x''$ лежит на отрезке $[x';x'']$ и при непрерывном изменении α от 0 до 1 пробегает отрезок от точки x'' (при $\alpha = 0$) до точки x' (при $\alpha = 1$).

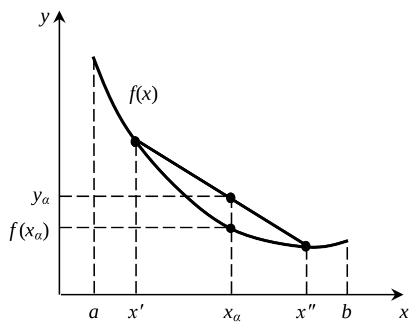


Рис. 3.2. Взаимное расположение $\alpha f(x') + (1-\alpha) f(x'')$.

2. Для того чтобы дифференцируемая на отрезке $[a;b]$ функция $f(x)$ была выпуклой на этом отрезке, необходимо и достаточно, чтобы производная $f'(x)$ не убывала на $[a;b]$;

3. 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 $f''(x) \ge 0$.

Условие выпуклости для дифференцируемой на отрезке $[a;b]$ функции $f(x)$ означает, что на этом отрезке любая касательная к графику $f(x)$ лежит не выше этого графика (Рис. 3.3).

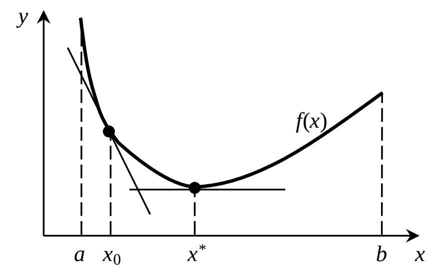


Рис. 3.3. Взаимные расположения графика функции и касательных к ней

Уравнение касательной к графику $f(x)$ в точке $(x_0, f(x_0))$, $x_0 \in [a;b]$ имеет вид:

$$y(x) = f(x_0) + f'(x_0)(x - x_0).$$

По формуле конечных приращений для любого $x \in [a;b]$ получим:

$f(x) = f(x_0) + f'(\xi)(x - x_0)$, где точка ξ лежит между x и x_0.

Поэтому $f(x) - y(x) = [f'(\xi) - f'(x_0)](x - x_0)$, $x \in [a;b]$, откуда с учетом того, что производная $f'(x)$ выпуклой функции не убывает, получаем:

$$f(x) - y(x) \geq 0 \text{ для всех } x \in [a;b].$$

4. Если $f(x)$ – выпуклая дифференцируемая на отрезке $[a;b]$ функция и в точке $x^* \in [a;b]$ выполняется равенство $f'(x^*) = 0$, то x^* является точкой глобального минимума $f(x)$ на $[a;b]$. Иначе, если касательная к графику $f(x)$ в точке с абсциссой x^* горизонтальна, а сам график расположен не ниже касательной, то x^* есть точка минимума функции $f(x)$.

Таким образом, равенство из пункта **4** для выпуклой дифференцируемой функции является не только необходимым условием глобального минимума (как для всякой дифференцируемой функции), но и его достаточным условием. Непосредственная проверка унимодальности функции с помощью определения в большинстве случаев вызывает затруднения, и для достаточно гладких функций часто используют те же критерии выпуклости. Если функция оказывается выпуклой, то можно утверждать, что она унимодальна. Разумеется, при отрицательном результате проверки функции на выпуклость нельзя сделать вывод о том, что она не унимодальна.

Классическая минимизация функции одной переменной

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

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

2. *Достаточные условия экстремума*. 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, $f^{(n)}(\tilde{x})\neq 0$. Тогда, если n – нечетно, то $\tilde{x}$ не является точкой локального экстремума функции $f(x)$. Если же n – четное число, то:

- ✓ при $f^{(n)}(\tilde{x})>0$ $\tilde{x}$ – точка локального минимума $f(x)$;
- ✓ при $f^{(n)}(\tilde{x})<0$ $\tilde{x}$ – точка локального максимума $f(x)$.

Перечисленные условия позволяют предложить следующий алгоритм решения задачи минимизации (1):

1. Находим все точки возможного экстремума функции $f(x)$ (стационарные точки), принадлежащие интервалу $(a;b)$; т. е. корни уравнения:

$$f'(\tilde{x})=0.$$

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

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 $[a;b]$.

З а м е ч а н и е. 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.

Пример. Решить задачу $f(x) = x^3 - 3x^2 + 1 \to min$, $x \in [-2; 2]$ классическим методом.

Решение.

Шаг 1. Находим корни уравнения $f'(x) = 3x^2 - 3 = 0$ из интервала $(-2; 2)$:

$$x_1 = -1,\ x_2 = 1.\ \text{Полагаем}\ x_0 = -2,\ x_3 = 2.$$

Шаг 2. Вычисляем значения $f(x)$ в точках x_i, $i = 0,\ldots,3$:

$$f(x_0) = -17,\ f(x_1) = 3,\ f(x_2) = -1,\ f(x_3) = 1.$$

Шаг 3. Находим f^* – наименьшее из $(-17, 3, -1, 1) = -17 = f(x_0)$.

Поэтому *координаты минимума* $x^* = x_0 = -2$, $f^* = -17$.

При решении практических задач оптимизации классический метод имеет ограниченное применение. Это объясняется тем, что, во–первых, во многих случаях значения целевой функции $f(x)$ находятся из измерений или экспериментов, а измерение производной $f'(x)$ затруднительно или невозможно и, во–вторых, если производная $f'(x)$ задана аналитически или поддается измерению, решение уравнения (необходимое условие 1) зачастую вызывает затруднения.

3.2. Прямые методы одномерной оптимизации унимодальных функций

Для решения задачи минимизации функции $f(x)$ на отрезке $[a; b]$ на практике, как правило, применяют приближенные методы. Они позволяют найти решение этой задачи с необходимой точностью в результате определения конечного числа значений функции $f(x)$ и ее производных в некоторых точках отрезка $[a;b]$. Методы, использующие только значения функции и не требующие вычисления ее производных, называются *прямыми методами* минимизации.

Большим достоинством прямых методов является то, что от целевой функции не требуется дифференцируемости и, более того, она может быть не задана в аналитическом виде. Единственное, на чем основаны алгоритмы прямых методов минимизации, это возможность определения значений $f(x)$ в заданных точках.

Рассмотрим наиболее распространенные на практике прямые методы поиска точки минимума. Самым слабым требованием к функции $f(x)$, позволяющим использовать эти методы, является ее унимодальность. Поэтому далее будем считать функцию $f(x)$ унимодальной на отрезке $[a;b]$

.

Метод равномерного поиска. Метод равномерного поиска или метод перебора относится к пассивным стратегиям поиска точки экстремума и является простейшим из прямых методов минимизации. Разобьем отрезок $[a;b]$ на n равных частей точками деления $x_i = a + i(b-a)/n, \quad i = 0,\ldots,n$. Вычислив значения $f(x)$ в точках x_i, путем сравнения найдем точку x_m, $0 \le m \le n$, для которой:

$$f(x_m) = \min_{0 \le i \le n} f(x_m). \tag{4}$$

Далее, положим $x^* \approx x_m$, $f^* \approx f(x_m)$.

Погрешность определения точки минимума x^* функции $f(x)$ методом перебора не превосходит величины $\varepsilon_m = (b-a)/n$.

Предположим, что x_m, из (4) является внутренней точкой разбиения отрезка $[a;b]$, т. е. $1 \le m \le n-1$ (случаи $m = 0$ и $m = n$ рассматриваются аналогично). Тогда с учетом свойства унимодальных функций следует:

✓ $f(x_{m-1}) \geq f(x_m)$, т. е. $x^* \in [x_m; b]$;

✓ $f(x_m) \leq f(x_{m+1})$, т. е. $x^* \in [a; x_{m+1}]$.

Отсюда получаем, что $x^* \in [x_{m-1}; b] \cap x^* \in [a; x_{m+1}] = [x_{m-1}; x_{m+1}]$. Длина последнего отрезка равна $2(b-a)/n$, а точка x_m является его серединой. Поэтому $\left|x_m - x^*\right| \leq (b-a)/n = \varepsilon_n$. Таким образом, чтобы обеспечить требуемую точность ε определения точки x^*, число отрезков разбиения n необходимо выбрать из условия $\varepsilon_n = (b-a)/n \leq \varepsilon$, т. е. $n \geq (b-a)/\varepsilon$.

Пусть реализация метода перебора потребовала N вычислений функции $f(x)$. Это означает, что отрезок $[a; b]$ был разбит на $n = N-1$ частей и достигнутая точность определения x^* составила $\varepsilon_n = \varepsilon_{N-1} = \dfrac{b-a}{N-1}$.

Поэтому метод равномерного поиска обеспечивает точность решения в результате N вычислений $f(x)$: $\varepsilon(N) = \dfrac{b-a}{N-1}$. Для оценки сходимости используется характеристика относительного уменьшения начального интервала неопределенности $R(N) = \dfrac{2}{N+1}$.

Пример. Методом равномерного поиска найти минимальное значение функции $f(x) = x^4 + e^{-x} \to min$, $x \in [0; 1]$ с точностью до $\varepsilon = 0.1$.

Решение. Функция $f(x)$ унимодальна на отрезке $[0; 1]$. Найдем число n отрезков разбиения: $n \geq \dfrac{1-0}{0.1} = 10$, можно взять $n = 10$. Вычислим значения $f(x_i)$, где $i = 0, \ldots, 10$ и запишем результаты в таблице 3.1.

Таблица 3.1

x_i	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$f(x_i)$	1.00	0.90	0.82	0.75	0.70	<u>0.67</u>	0.68	0.74	0.86	1.06	1.37

Минимальное из вычисленных значений $f(x)$ в таблице подчеркнуто. Таким образом, искомая точка: $x^* \approx 0.5$, $f^* \approx 0.67$.

3.3. Методы сокращения интервала неопределенности

В методе равномерного поиска, рассмотренном выше, точки x_i, в которых определяются значения $f(x)$, выбирают заранее. Если же для выбора очередной точки вычисления (измерения) $f(x)$ использовать информацию, содержащуюся в уже найденных значениях $f(x)$, то поиск точки минимума можно сделать более эффективным, т. е. сократить число определяемых для этого значений $f(x)$. На вариант более эффективного поиска точки x^* указывает одно из свойств унимодальных функций. Пусть $a < x_1 < x_2 < b$. Сравнив значения $f(x)$ в точках x_1 и x_2 (*пробных точках*), можно сократить отрезок поиска точки x^*, перейдя к отрезку $[a; x_2]$, если $f(x_1) \le f(x_2)$ или к отрезку m $[x_1; b]$, если $f(x_1) > f(x_2)$ (Рис. 3.4).

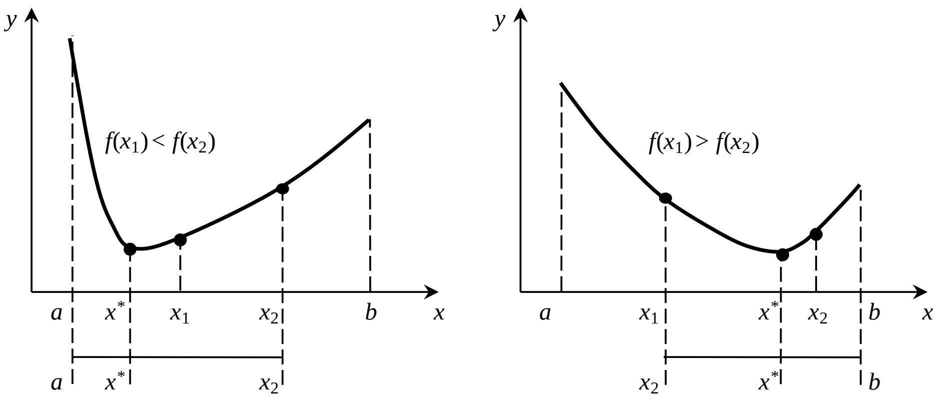


Рис. 3.4. Уменьшение отрезка поиска точки минимума

Описанную процедуру можно повторить необходимое число раз, последовательно уменьшая длину отрезка, содержащего точку минимума. Когда длина последнего из найденных отрезков станет достаточно малой, следует положить $x^* \approx \bar{x}$, где $\bar{x}$ – одна из точек этого отрезка, например, его середина. Методы минимизации, основанные на этом принципе, называются *методами сокращения интервалов неопределенности*. Чтобы относительное уменьшение отрезка на каждой итерации не зависело от того, какая из его частей исключается из дальнейшего рассмотрения, пробные точки следует располагать симметрично относительно середины исходного отрезка. В зависимости от способа выбора пробных точек получаются различные методы исключения отрезков.

Метод дихотомии. *Дихотомия* (от греч. «надвое»+«деление») – раздвоенность, последовательное деление на две части, не связанные между собой. Метод относится к последовательным стратегиям. Задается начальный интервал неопределенности и требуемая точность. Алгоритм опирается на анализ значений функции в двух точках. Для их нахождения текущий интервал неопределенности делится пополам и в обе стороны от середины откладывается $\frac{\delta}{2}$, где δ малое положительное число. Поиск заканчивается, когда длина текущего интервала неопределенности становится меньше заданной величины точности.

В методе дихотомии точки x_1 и x_2 располагаются близко к середине очередного отрезка $[a; b]$:

$$x_1 = \frac{b + a - \delta}{2}, \quad x_2 = \frac{b + a + \delta}{2}.$$

При этом отношение длин нового и исходного отрезков $\tau = \frac{b - x_1}{b - a} = \frac{x_2 - a}{b - a}$ близко к $1/2$, этим и объясняется название метода.

Отметим, что для любых точек x_1 и x_2 величина $\tau > 1/2$, поэтому указанный выбор пробных точек объясняется стремлением обеспечить максимально возможное относительное уменьшение отрезка на каждой итерации поиска x^*. В конце вычислений по методу дихотомии в качестве приближенного значения x^* берут середину последнего из найденных отрезков $[a;b]$, убедившись предварительно, что достигнуто неравенство по заданной точности: $\frac{b-a}{2} \le \varepsilon$.

С точки зрения компьютерной реализации этот метод наиболее прост и используется во многих стандартных программных средствах.

Алгоритм метода дихотомии:

Шаг 1. Определить x_1 и x_2. Вычислить $f(x_1)$ и $f(x_2)$.

Шаг 2. Сравнить $f(x_1)$ и $f(x_2)$. Если $f(x_1) \le f(x_2)$, то перейти к отрезку $[a; x_2]$, положив $b = x_2$, иначе – к отрезку $[x_1; b]$, положив $a = x_1$.

Шаг 3. Найти достигнутую точность $\varepsilon_n = \frac{b-a}{2}$. Если $\varepsilon_n > \varepsilon$, то перейти к следующей итерации, вернувшись к шагу 1. Если $\varepsilon_n \le \varepsilon$, то завершить поиск x^*, перейдя к шагу 4.

Шаг 4. Положить $x^* \approx \bar{x} = \frac{a+b}{2}$, $f^* \approx (\bar{x})$

Замечания.

1. Значение δ выбирают на интервале $(0; 2\varepsilon)$ из следующих соображений:

✓ чем меньше δ, тем больше относительное уменьшение длины отрезка на каждой итерации, т. е. при уменьшении δ достигается более высокая скорость сходимости метода дихотомии;

✓ при чрезмерно малом δ сравнение значений $f(x)$ в точках x_1 и x_2, отличающихся на величину δ, становится затруднительным. Поэтому выбор δ должен быть согласован с точностью определения функции $f(x)$ и с количеством верных десятичных знаков при задании аргумента x.

2. Число n итераций метода дихотомии, необходимое для определения точки x^* с точностью до ε, определяется неравенством:

$$n \geq \log_2 \frac{b-a-\delta}{2\varepsilon-\delta}.$$

Например, при исходном единичном интервале и точности $\varepsilon \sim 10^{-6}$ порядка 6 знаков после десятичной точки достаточно провести 20 итераций.

3. В результате n итераций длина отрезка поиска точки x^* станет:

$$\Delta_n = \frac{b-a}{2^n} + \left(\frac{1}{2^n} + \frac{1}{2^{n-1}} + \ldots + \frac{1}{2}\right) = \frac{b-a}{2^n} + \left(1 - \frac{1}{2^n}\right)\delta.$$

При этом будет достигнута точность определения точки минимума из условия:

$$\varepsilon_n = \frac{b-a}{2^n} + \left(1 - \frac{1}{2^n}\right)\frac{\delta}{2} \leq \varepsilon.$$

4. Величина δ может быть выбрана достаточно малой, поэтому, пренебрегая ею, получаем: $\varepsilon_n = \dfrac{b-a}{2^{n+1}}$. На каждой итерации метода дихотомии вычисляют два значения $f(x)$. Поэтому после N вычислений $f(x)$ производят $n = N/2$ итераций и достигают точность определения x^*:

$$\varepsilon(N/2) = \frac{b-a}{2^{N/2+1}}.$$

П р и м е р. Найти минимум функции предыдущего примера, методом дихотомии. Задана функция $f(x) = x^4 + e^{-x} \to min$, $x \in [0; 1]$, $\varepsilon = 0.1$.

Р е ш е н и е. Выберем $\delta = 0.02$. И т е р а ц и я 1.

Ш а г 1. $x_1 = 0.49$, $x_2 = 0.51$. $f(x_1) = 0.670$, $f(x_2) = 0.688$.

Ш а г 2. $f(x_1) > f(x_2)$, поэтому полагаем $a = x_1 = 0.49$.

Ш а г 3. $(b-a)/2 = 0.255 > 0.1$, значит переходим к следующей итерации. Результаты вычислений на остальных итерациях записаны в таблице 3.2.

Таблица 3.2

Номер итерации	a	b	$\frac{b-a}{2}$	x_1	x_2	$f(x_1)$	$f(x_2)$	Сравнение $f(x_1)$ и $f(x_2)$
2	0.49	1	0.26	0.735	0.755	0.771	0.792	$f(x_1)<f(x_2)$
3	0.49	0.755	0.13	0.613	0.633	0.683	0.691	$f(x_1)<f(x_2)$
4	0.49	0.633	0.07	0.07 < 0.1 – *точность достигнута*				

Таким образом, $x^* \approx \frac{0.49+0.633}{2} \approx 0.56$, $f^* \approx f(0.56) \approx 0.67$ (сравните с результатами решения другими методами).

Метод деления отрезка пополам является вариацией метода дихотомии. Этот метод, использует на каждой итерации три пробные точки и обеспечивает последовательное уменьшение длины отрезка, содержащего x^*, ровно вдвое.

Разделим отрезок $[a;b]$ на четыре равные части пробными точками $x_i = a + \frac{b-a}{4} i$, $i = 1,2,3$. Сравним значения $f(x_1)$ и $f(x_2)$.

Если $f(x_1) \le f(x_2)$, то уменьшенный вдвое отрезок поиска точки x^* найден – это $[a; x_2]$. Если $f(x_1) > f(x_2)$, то произведем еще одно сравнение значений $f(x)$: при $f(x_2) \le f(x_3)$, перейдем к отрезку $[x_1; x_3]$, а в противном случае – к отрезку $[x_2; b]$.

Отметим, что каким бы ни оказался новый отрезок, одна из уже использованных пробных точек переходит на его середину, становясь новой точкой x_2. Таким образом, для проведения следующей итерации на вновь полученном отрезке потребуется вычисление не более двух новых значений $f(x)$ (либо только в точке x_1, либо еще и в точке x_3).

Алгоритм метода деления отрезка пополам:

Шаг 1. Положить $x_1 = \dfrac{a+b}{2}$. Вычислить значение $f(x_2)$ и перейти к шагу 2.

Шаг 2. Положить $x_2 = \dfrac{a+b}{2}$. Вычислить значение $f(x_1)$ и перейти к шагу 3.

Шаг 3. Сравнить $f(x_1)$ и $f(x_2)$. Если $f(x_1) \le f(x_2)$, то продолжить поиск на отрезке $[a; x_2]$, положив $b = x_2$, $x_2 = x_1$, $f(x_1) = f(x_2)$, и перейти к шагу 5, иначе – положить $x_3 = \dfrac{x_2 + b}{2}$, вычислить значение $f(x_3)$ и перейти к шагу 4.

Шаг 4. Сравнить $f(x_2)$ и $f(x_3)$. Если $f(x_2) \le f(x_3)$, то перейти к отрезку $[x_1; x_3]$, положив $a = x_1$, $b = x_3$, иначе – продолжить поиск на отрезке $[x_2; b]$, положив $a = x_2$, $x_2 = x_3$, $f(x_2) = f(x_3)$. Перейти к шагу 5.

Шаг 5. Проверка окончания поиска. Вычислить $\varepsilon_n = \dfrac{b-a}{2}$ и сравнить с заданной точностью ε. Если $\varepsilon_n > \varepsilon$, то перейти к следующей итерации, вернувшись к шагу 2, иначе – завершить поиск, положив $x^* \approx x_2$, $f^* \approx f(x_2)$.

Пример. Найти минимум функции предыдущего примера, методом деления отрезка пополам. Задана функция: $f(x) = x^4 + e^{-x} \to min$, $x \in [0;1]$, $\varepsilon = 0.1$.

Решение. Итерация 1.

Шаг 1. Находим $x_2 = 0.5$, $f(x_2) = 0.669$. Переходим к шагу 2.

Шаг 2. Определяем $x_1 = 0.25$, $f(x_1) = 0.783$. Переходим к шагу 3.

Шаг 3. $f(x_1) > f(x_2)$, поэтому полагаем $x_3 = 0.75$, вычисляем $f(x_3) = 0.789$ и переходим к шагу 4.

Ш а г 4. $f(x_1) < f(x_2)$, поэтому полагаем $a = 0.25$, $b = 0.75$ и переходим к шагу 5.

Ш а г 5. Находим $\varepsilon_n = 0.25 > 0.1$, т. е. переходим к следующей итерации, начиная с шага 2.

Результаты вычислений остальных итераций записаны в таблице 3.3.

Таблица 3.3

Номер итерации	a	b	ε_n	x_1	x_2	x_3	$f(x_1)$	$f(x_2)$	$f(x_3)$	Сравнение $f(x_1)$ и $f(x_2)$
2	0.250	0.750	0.25	0.375	0.500	0.625	0.707	0.669	0.688	$f(x_1) > f(x_2)$ $f(x_2) < f(x_3)$
3	0.375	0.625	0.13	0.438	0.500	0.563	0.669	0.669	0.670	$f(x_1) = f(x_2)$ $f(x_2) < f(x_3)$
4	0.438	0.563	0.06	0.06 < 0.1 – *точность достигнута*						

Таким образом, получаем решение $x^* \approx x_2 = 0.5$, $f^* \approx f(x_2) = 0.67$. Сравните с результатами решения другими методами.

З а м е ч а н и е. В первой итерации метода деления отрезка пополам вычисляется не более трех значений $f(x)$, а на остальных – не более двух. Поэтому N вычислений $f(x)$ гарантируют осуществление $(N-1)/2$ итераций, и достигнутая точность определения x^* составляет:

$$\varepsilon(N) = \frac{b-a}{2^{N-1/2+1}}.$$

М е т о д з о л о т о г о с е ч е н и я. Название «золотое сечение» произошло от названия отношения, в котором некоторый отрезок делится на две части так, что отношение целого к большей части равно отношению большей части к меньшей, т. е. находится в так называемом «золотом отношении». С позиций математики золотое сечение (золотая пропорция) это такое соотношение двух величин b и a, при $a > b$, когда справедливо равенство: $\frac{a}{b} = \frac{a+b}{a} = \frac{1+\sqrt{5}}{2} \approx 1.62$. Число, равное отношению a/b, обычно обозначается прописной греческой буквой Φ в честь древнегреческого скульптора и архитектора Фидия. В процентном округленном значении «золотое сечение» – это деление какой-либо величины в отношении 62% и 38%.

Метод относится к последовательным стратегиям. Задается начальный интервал неопределенности и требуемая точность. Алгоритм опирается на анализ значений функции в двух точках. В качестве точек вычисления функции выбираются точки золотого сечения. На каждой итерации, кроме первой, требуется только одно новое вычисление функции. Поиск заканчивается, если длина текущего интервала неопределенности меньше заданной величины.

Рассмотрим такое симметричное расположение точек x_1 и x_2 на отрезке $[a; b]$, при котором одна из них становится пробной точкой и на новом отрезке, полученном после исключения части исходного отрезка. Использование таких точек позволяет на каждой итерации метода, кроме первой, ограничиться определением только одного значения $f(x)$, тогда как другое значение уже найдено на предыдущих итерациях.

Найдем точки x_1 и x_2, обладающие указанным свойством. Рассмотрим отрезок $[0; 1]$ и для определенности предположим, что при его уменьшении исключается правая часть этого отрезка. Пусть $x_2 = \tau$, тогда симметрично расположенная точка $x_1 = 1 - \tau$ (Рис. 3.5).

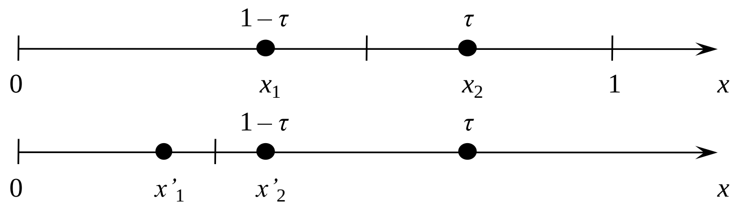


Рис. 3.5. Определение пробных точек в методе золотого сечения

Пробная точка x_1 отрезка $[0; 1]$ перейдет в пробную точку $x_2' = 1 - \tau$ нового отрезка $[0; \tau]$. Чтобы точки $x_2 = \tau$ и $x_2' = 1 - \tau$ делили отрезки $[0; 1]$ и

$[0;\tau]$ в одном и том же отношении, должно выполняться равенство $\frac{1}{\tau}=\frac{\tau}{1-\tau}$ или $\tau^2=1-\tau$, откуда находим положительное значение $\tau=\frac{\sqrt{5}-1}{2}=0.61803$

Таким образом, $x_1=1-\tau=\frac{3-\sqrt{5}}{2}$, $x_2=\tau=\frac{\sqrt{5}-1}{2}$.

Для произвольного отрезка $[a;b]$ выражения для пробных точек примут вид: $x_1=a+\frac{3-\sqrt{5}}{2}(b-a)$; $x_2=a+\frac{\sqrt{5}-1}{2}(b-a)$.

Замечания.

1. Точки x_1 и x_2 обладают свойством золотого сечения отрезка $[a;b]$. Это и объясняет название рассматриваемого метода.

2. На каждой итерации исключения отрезков с пробными точками одна из них $\bar{x}$ переходит на следующий отрезок и значение $f(x)$ в этой точке вычислять не следует. Если новым отрезком становится $[a;x_2]$, то на него переходит пробная точка $\bar{x}=x_1$ исходного отрезка, становясь его второй пробной точкой ($x_2'=x_1$) (Рис. 3.5). В случае перехода к отрезку $[x_1;b]$ пробная точка $\bar{x}=x_2$ исходного отрезка становится первой пробной точкой отрезка $[x_1;b]$.

3. Легко проверить, что $x_1=a+b-x_2$ и $x_2=a+b-x_1$. Поэтому на каждой итерации метода золотого сечения недостающую пробную точку нового отрезка можно найти по перешедшей на него пробной точке с помощью сложения и вычитания.

4. В конце вычислений по методу золотого сечения в качестве приближенного значения x^* можно взять середину последнего из полученных отрезков $\bar{x}=\frac{a+b}{2}$.

5. На каждой итерации отрезок поиска точки минимума уменьшается в одном и том же отношении $\tau = \frac{\sqrt{5}-1}{2}$, поэтому в результате n итераций его длина становится $\Delta_n = \tau^n(b-a)$. Таким образом, точность ε_n определения точки x^* после n итераций находят из равенства:

$$\varepsilon_n = \frac{\Delta_n}{2} = \frac{1}{2}\left(\frac{\sqrt{5}-1}{2}\right)^n (b-a),$$

а условием окончания поиска точки x^* служит неравенство $\varepsilon_n \le \varepsilon$.

Алгоритм метода золотого сечения:

Шаг 1. Найти x_1 и x_2. Вычислить $f(x_1)$ и $f(x_2)$. Положить $\tau = \frac{\sqrt{5}-1}{2}$, $\varepsilon_n = \frac{b-a}{2}$.

Шаг 2. Проверка на окончание поиска: если $\varepsilon_n > \varepsilon$, то перейти к шагу 3, иначе – к шагу 4.

Шаг 3. Переход к новому отрезку и новым пробным точкам. Если $f(x_1) \le f(x_2)$, то положить $b = x_2$, $x_2 = x_1$, $f(x_2) \le f(x_1)$, $x_1 = b - \tau(b-a)$ и вычислить $f(x_1)$, иначе – положить $a = x_1$, $x_1 = x_2$, $f(x_1) = f(x_2)$, $x_2 = b + \tau(b-a)$ и вычислить $f(x_2)$. Положить $\varepsilon_n = \tau\varepsilon_n$ и перейти к шагу 2.

Шаг 4. Окончание поиска: положить $x^* \approx \bar{x} = \frac{a+b}{2}$, $f^* \approx f(\bar{x})$.

Пример. Найти минимум функции предыдущего примера, методом золотого сечения. Задана функция: $f(x) = x^4 + e^{-x} \to min$, $x \in [0;1]$, $\varepsilon = 0.1$.

Решение. Итерация 1.

Шаг 1. Находим: $x_1 = 0.382$, $x_2 = 0.618$, $f(x_1) = 0.704$, $f(x_2) = 0.685$, $\varepsilon_n = 0.5$.

Шаг 2. $\varepsilon_n = 0.5 > \varepsilon = 0.1$, поэтому переходим к шагу 3.

Ш а г 3. $f(x_1) > f(x_2)$, поэтому полагаем $a = 0.382$, $x_1 = 0.618$, $f(x_1) = 0.685$, $x_2 = 0.764$, $\varepsilon_n = 0.309$ и вычисляем $f(x_2) = 0.807$. Переходим к следующей итерации, начиная с шага 2.

Результаты вычислений остальных итераций представлены в таблице 3.4.

Таблица 3.4

Номер итерации	a	b	ε_n	x_1	x_2	$f(x_1)$	$f(x_2)$	Сравнение $f(x_1)$ и $f(x_2)$
2	0.382	1.000	0.309	0.618	0.764	0.685	0.807	$f(x_1) < f(x_2)$
3	0.382	0.764	0.191	0.528	0.618	0.668	0.685	$f(x_1) < f(x_2)$
4	0.382	0.618	0.118	0.472	0.528	0.673	0.668	$f(x_1) > f(x_2)$
5	0.472	0.618	0.073	0.073 < 0.1 – точность достигнута				

Получили решение $x^* \approx \dfrac{0.472 + 0.618}{2} \approx 0.55$, $f^* \approx f(0.55) = 0.67$ (сравните с решением по другим методам).

З а м е ч а н и е. Число итераций, необходимое для достижения заданной точности ε, можно найти из условия $\varepsilon_n < \varepsilon$ с учетом соотношения:

$$n \geq -2.1 \ln\left(\frac{2\varepsilon}{b-a}\right).$$

Так как N вычислений $f(x)$ позволяют выполнить $N-1$ итераций метода золотого сечения, то достигнутая в результате этих вычислений точность определения x^* составляет: $\varepsilon(N) = \varepsilon_{N-1} = \dfrac{1}{2}\left(\dfrac{\sqrt{5}-1}{2}\right)^{N-1}(b-a)$.

Метод Фибоначчи. В методе Фибоначчи реализована последовательная стратегия, обеспечивающая максимально гарантированное сокращение интервала неопределенности при заданном количестве вычисления функции. Эта стратегия опирается на числа Фибоначчи, определяемые по формуле:

$$F_{i_0} = F_{i_1} = 1,\ F_{i_k} = F_{i_{k-1}} + F_{i_{k-2}},\ k = 2, 3, 4, \ldots .$$

Последовательность чисел Фибоначчи имеет вид $1, 1, 2, 3, 5, 8, 13, 23, 34, 55, 89, 144, 233, \ldots$. Эти числа названы в честь великого математика Фибоначчи, настоящее имя которого Леонардо Пизанский.

Леонардо Пизанский (около 1170 – около 1250) являлся первым крупным математиком средневековой Европы. Фибоначчи – это прозвище, которое переводится с итальянского как «Хороший сын родился». Отец его был торговец, поэтому часто разъезжал по разным странам. Чаще всего он бывал в Алжире, где юный Леонардо обучался математике у арабских учителей, по переводам ознакомился с трудами античных и индийских математиков. Самый известный труд ученого называется «Книга абака». В этой книге изложены все арифметические и алгебраические сведения того времени с неимоверной ясностью и глубиной. В отдельной главе, посвященной арифметической и геометрической прогрессии, ряда квадрата, впервые описана последовательность чисел Фибоначчи – как элементов числовой последовательности, в которой каждое последующие число равно сумме двух предыдущих. Помимо числовой последовательности «Фибоначчи» Леонардо Пизанский сделал еще ряд открытий, среди них:

- ✓ формулировка правила нахождения суммы членов арифметической прогрессии;
- ✓ ввел термин «частное» для обозначения результата деления;
- ✓ описал способ приведения дробей к общему знаменателю с помощью
- ✓ нахождения наименьшего общего кратного знаменателя.

Метод относится к последовательным стратегиям. Задается начальный интервал неопределенности и количество вычислений функции. Алгоритм опирается на анализ значений функции в двух точках. Точки вычисления функции находятся с использованием последовательности из $N+1$ чисел Фибоначчи. На каждой итерации, кроме первой, требуется только одно новое вычисление функции. Поиск заканчивается, если длина текущего интервала неопределенности меньше заданной величины точности. Длина отрезка неопределенности в методе Фибоначчи $L_N \approx {(b-a)}/{F_N}$.

Следовательно, стратегия ясна с самого начала. Нужно поместить следующую точку внутри интервала неопределенности симметрично относительно уже находящейся там точке. Парадоксально, но, чтобы понять, как следует начинать вычисления, необходимо разобраться в том, как его следует заканчивать. Разберемся с этим парадоксом.

На $n-m$ вычислении n-ю точку следует поместить симметрично по отношению к $n-1$ точке. Положение этой последней точки в принципе зависит от нас. Для того чтобы получить наибольшее уменьшение интервала на данном этапе, следует разделить пополам предыдущий интервал. Тогда точка x_n будет совпадать с точкой x_{n-1}. Однако при этом мы не получаем никакой новой информации. Обычно точки x_{n-1} и x_n отстоят друг от друга на достаточном расстоянии, чтобы определить, в какой половине, левой или правой, находится интервал неопределенности. Они помещаются на расстоянии $\frac{\delta}{2}$ по обе стороны от средины отрезка L_{n-1} можно самим задать величину δ или выбрать эту величину равной минимально возможному расстоянию между двумя точками. Используемое значение δ может определяться из практических соображений. Оно должно быть меньше $\frac{L_1}{F_{n+1}}$, в противном случае напрасно будет тратиться время на вычисление функции. После того как найдено положение первой точки, числа Фибоначчи больше не нужны.

Алгоритм поиска минимума функции по методу Фибоначчи:

Шаг 1. Задается начальный интервал неопределенности $L_0=[a_0,b_0]$, $\alpha>0$ допустимая длина конечного интервала.

Шаг 2. Найти количество вычислений функции как наименьшее целое число, при котором удовлетворяется условие $F_{i_N}\geq\frac{|L_0|}{\alpha}$ и числа Фибоначчи $F_{i_0},F_{i_1},\ldots,F_{i_N}$.

Шаг 3. Вычислить $y_k=a_0+\frac{F_{i_{N-2}}}{F_{i_N}}(b_0-a_0)$, $z_k=a_0+\frac{F_{i_{N-1}}}{F_{i_N}}(b_0-a_0)$, $k=0$

.

Шаг 4. Вычислить значения функций $F(y_k),F(z_k)$.

Шаг 5. Если $F(y_k)\leq F(z_k)$, то принять $a_{k+1}=a_k$, $b_{k+1}=z_k$, $z_{k+1}=y_k$ и вычислить $y_{k+1}=a_{k+1}+\frac{F_{i_{N-k-3}}}{F_{i_{N-k-1}}}(b_{k+1}-a_{k+1})$, затем перейти к шагу 6.

Если $F(y_k)>F(z_k)$, принять $a_{k+1}=y_k$, $b_{k+1}=b_k$, $y_{k+1}=z_k$ и вычислить $z_{k+1}=a_{k+1}+\frac{F_{i_{N-k-2}}}{F_{i_{N-k-1}}}(b_{k+1}-a_{k+1})$.

Шаг 6. Если $k\neq N-3$, то принять $k=k+1$ и перейти к шагу 4.

Если $k=N-3$, то всегда $y_{N-2}=z_{N-2}=\frac{(a_{N-2}+b_{N-2})}{2}$, то есть отсутствует точка нового вычисления функции. В этом случае следует принять $y_{N-1}=y_{N-2}=z_{N-2}$; $z_{N-1}=y_{N-1}+\varepsilon$.

В точках y_{N-1}, z_{N-1} вычисляются значения функции и находятся границы конечного интервала неопределенности:

- ✓ если $F(y_{N-1})\leq F(z_{N-1})$, то принять $a_{N-1}=a_{N-2}$, $b_{N-1}=z_{N-1}$;
- ✓ если $F(y_{N-1})>F(z_{N-1})$, то принять $a_{N-1}=y_{N-1}$, $b_{N-1}=b_{N-2}$.

Процесс поиска завершается и $x^* \in [a_{N-1}, b_{N-1}]$. В качестве приближенного решения можно принять любую точку интервала, рекомендуется $x^* = \dfrac{a_{N-1} + b_{N-1}}{2}$.

Характеристика относительного уменьшения начального интервала неопределенности $R(N) = \dfrac{1}{F_{i_N}}$, где N количество вычислений функции. Интересно, что $\dfrac{F_{k-1}}{F_k} \sim z$ zпри $k \geq 3$, причем, чем больше k, тем ближе отношение $\dfrac{F_{k-1}}{F_k}$ к числу z определяемому золотым сечением.

Сравнение методов сокращения интервалов неопределенности и равномерного поиска. При сравнении прямых методов минимизации обычно учитывают количество N значений $f(x)$, гарантирующее заданную точность определения точки x^* тем или иным методом. Чем меньше N, тем эффективнее считается метод. При этом вспомогательные операции такие, как выбор пробных точек, сравнение значений $f(x)$ и т. п., не учитываются.

Эффективность методов минимизации можно также сравнивать, на основании гарантированной точности $\varepsilon(N)$ нахождения точки x^*, которую они обеспечивают в результате определения N значений $f(x)$. В таблице 3.5 приведены значения достигнутой точности $\varepsilon(N)$ в зависимости от количества N найденных значений $f(x)$ на отрезке единичной длины для некоторых методов.

Таблица 3.5

Методы минимизации	Количество найденных значений $f(x)$			
	$N = 5$	$N = 11$	$N = 21$	$N = 51$
Метод золотого сечения	0.073	$4.1 \cdot 10^{-3}$	$3.3 \cdot 10^{-5}$	$1.8 \cdot 10^{-11}$
Методы деления отрезка пополам	0.125	$1.6 \cdot 10^{-2}$	$4.9 \cdot 10^{-4}$	$1.5 \cdot 10^{-8}$
Метод перебора	0.250	0.100	0.050	0.020

Из анализа рассмотренных методов следует, что наиболее эффективным из сравниваемых методов является метод золотого сечения и близкий к нему метод Фибоначчи, за ними идут методы деления отрезка пополам и наименее эффективен метод перебора.

3.4. Численные методы одномерной минимизации с использованием производной

Поиск точки минимума методами сокращения интервалов неопределенности основан на сравнении значений функции в двух точках. При таком сравнении разности значений $f(x)$ в этих точках не учитываются, важны только их знаки. Рассматриваемые в этом подразделе методы, используются при минимальных требованиях к целевой функции – она должна быть унимодальной. Также предполагается, что целевая функция $Y = F(x)$ является выпуклой и дифференцируемой (один раз или дважды). Причем, производная может быть вычислена в произвольно выбранных точках. Считается, что эффективность методов, использующих информацию о производных при поиске точки минимума можно существенно повысить. К основным численным методам одномерной минимизации с использованием производной относятся: метод средней точки; метод хорд; метод Ньютона.

М е т о д   с р е д н е й   т о ч к и направлен на повышение эффективности метода деления отрезка пополам. Сутью является использование технологии исключения отрезков за счет замены вычислений функции в трех точках на операцию вычисления производной в средней точке $\tilde{x} = \frac{a+b}{2}$.

Если $F'(\tilde{x}) > 0$, то точка $\tilde{x}$ лежит на участке монотонного возрастания $F(x)$, поэтому $x^* < \tilde{x}$ и точку минимума следует искать на отрезке $[a, \tilde{x}]$.

Если $F'(\tilde{x}) < 0$, то точка $\tilde{x}$ лежит на участке монотонного убывания $F(x)$, поэтому $x^* > \tilde{x}$ и точку минимума следует искать на отрезке $[\tilde{x}, b]$.

Равенство $F'(\tilde{x}) = 0$ означает, что точка минимума найдена $x^* = \tilde{x}$.

Такое исключение отрезков требует на каждой итерации только одного вычисления $F'(\tilde{x})$ и уменьшает отрезок поиска точки минимума ровно в два раза. Поиск заканчивается, если абсолютная величина производной меньше заданной погрешности.

Алгоритм поиска точки минимума методом средней точки:

Ш а г 1. Задается начальный интервал неопределенности $L_0 = [a_0, b_0]$ и $\varepsilon > 0$ – требуемая точность.

Ш а г 2. Задать $k = 0$.

Ш а г 3. Вычислить среднюю точку $\tilde{x} = \frac{a_k + b_k}{2}$, $F'(\tilde{x})$.

Ш а г 4. Проверить условие окончания:

✓ если $|F'(\tilde{x})| \le \varepsilon$, то процесс поиска завершается и $x^* = \tilde{x}$, $F^* = F(x^*)$;

✓ если $|F'(\tilde{x})| > \varepsilon$, то сравнить $F'(\tilde{x})$ с нулем.

Если $F'(\tilde{x}) > 0$, то продолжить поиск на отрезке $L_k = [a_k, b_k]$, положив $k = k+1$, $a_k = a_{k-1}$, $b_k = \tilde{x}_{k-1}$.

Если $F'(\tilde{x}) \le 0$, то продолжить поиск на отрезке $L_k = [a_k, b_k]$, положив $k = k+1$, $a_k = \tilde{x}_{k-1}$, $b_k = b_{k-1}$. Перейти к шагу 3.

Метод хорд опирается на равенство $F'(x)=0$, которое является необходимым и достаточным условием глобального минимума выпуклой дифференцируемой функции $F(x)$. Если на концах отрезка $L=[a,b]$ производная имеет разные знаки, то на интервале (a,b) найдется точка, в которой $F'(x)=0$ и поиск точки минимума $F(x)$ на отрезке $[a,b]$ эквивалентен решению уравнения $F'(x)=0$, $x\in[a,b]$. Таким образом, любой приближенный метод решения уравнения $F'(x)=0$, $x\in[a,b]$ можно рассматривать как метод минимизации выпуклой дифференцируемой функции $F(x)$ на отрезке $[a,b]$. Метод хорд основан на исключении отрезка путем определения точки пересечения с осью Ox хорды графика функции $F'(x)$ на очередном отрезке:

$$\tilde{x}=a-\frac{F'(a)}{F'(a)-F'(b)}(a-b).$$

Отрезок дальнейшего поиска определяется по правилу. Новыми точками отрезка $[a,b]$ для осуществления следующей итерации являются концы того из отрезков $[a,\tilde{x}]$ и $[\tilde{x},b]$, который содержит точку x^*. Его определяют по знаку производной $F'(\tilde{x})$.

✓ Если $F'(\tilde{x})>0$, то точка $\tilde{x}$ лежит на участке монотонного возрастания $F(x)$, поэтому $x^*<\tilde{x}$ и точку минимума следует искать на отрезке $[a,\tilde{x}]$, то есть $b=\tilde{x}$.

✓ Если $F'(\tilde{x})<0$, то точка $\tilde{x}$ лежит на участке монотонного убывания $F(x)$, поэтому $x^*>\tilde{x}$ и точку минимума следует искать на отрезке $[\tilde{x},b]$, то есть $a=\tilde{x}$.

✓ Равенство $F'(\tilde{x})=0$ означает, что точка минимума найдена точно и $x^*=\tilde{x}$. На каждой итерации, кроме первой, следует вычислять одно новое значение $F'(x)$. Поиск заканчивается, если абсолютная величина производной меньше заданной погрешности.

Алгоритм поиска точки минимума методом хорд:

Шаг 1. Задается начальный интервал неопределенности $L_0 = [a_0, b_0]$ и $\varepsilon > 0$ – требуемая точность, малое положительное число.

Шаг 2. Задать $k = 0$. Вычислить $F'(a_k)$, $F'(b_k)$.

Если $F'(a_k) * F'(b_k) < 0$, то перейти к шагу 3, иначе к шагу 5.

Шаг 3. Вычислить $x_k = a_k - \dfrac{F'(a_k)}{F'(a_k) - F'(b_k)}(a_k - b_k)$, $F'(x_k)$.

Шаг 4. Проверить условие окончания:

✓ если $|F'(x_k)| \le \varepsilon$, то процесс поиска завершается и $x^* = x_k$, $F^* = F(x^*)$;

✓ если $|F'(x_k)| > \varepsilon$, то сравнить $F'(x_k)$ с нулем.

Если $F'(x_k) > 0$, то продолжить поиск на отрезке $L_k = [a_k, b_k]$, положив $k = k+1$, $a_k = a_{k-1}$, $b_k = x_{k-1}$, $F'(b_k) = F'(x_{k-1})$.

Если $F'(x_k) \le 0$, то продолжить поиск на отрезке $L_k = [a_k, b_k]$, положив $k = k+1$, $a_k = x_{k-1}$, $b_k = b_{k-1}$, $F'(a_k) = F'(x_{k-1})$. Перейти на шаг 3.

Шаг 5. Если $F'(a_k) > 0$, $F'(b_k) > 0$, то $F(x)$ возрастает на отрезке $L_k = [a_k, b_k]$ и, следовательно, $x^* = a_k$.

Если $F'(a_k) < 0$, $F'(b_k) < 0$, то $F(x)$ убывает на отрезке $L_k = [a_k, b_k]$ и, следовательно, $x^* = b_k$.

Если $F'(a_k) * F'(b_k) = 0$, то $x^* = a_k$ или $x^* = b_k$, в зависимости от того, на каком из концов отрезка $L_k = [a_k, b_k]$ производная $F'(x) = 0$.

Метод Ньютона. Предполагается, что функция $F(x)$ дважды дифференцируема, причем $F''(x) > 0$. Тогда для поиска корня уравнения $F'(x) = 0$ используется метод касательных. Сущность метода заключается в том, что в очередной точке x_k строится линейная аппроксимация функции $F(x)$ (касательная к графику $F(x)$), а точка, в которой линейная аппроксимирующая функция обращается в нуль, используется в качестве следующего приближения x_{k+1}. Координата точки x_{k+1} находится по формуле:

$$x_{k+1} = x_k - \frac{F'(x_k)}{F''(x_k)}, \qquad k = 0,1,\ldots,$$

где x_0 начальная точка выбирается произвольно. Вычисления по приведенной формуле продолжаются до тех пор, пока не выполнится условие $|F'(x_k)| \le \varepsilon$, после чего полагают $x^* = x_k$, $F^* = F(x^*)$.

Алгоритм поиска минимума функции методом Ньютона:

Шаг 1. Задается начальный интервал неопределенности $L_0 = [a_0, b_0]$ и $\varepsilon > 0$ – требуемая точность.

Шаг 2. Задать $k = 0$ и начальную точку $x_k \in [a_k, b_k]$.

Шаг 3. Вычислить $F'(x_k)$. Проверить условие окончания:

✓ если $|F'(x_k)| \le \varepsilon$, то процесс поиска завершается и $x^* = x_k$, $F^* = F(x^*)$;

✓ если $|F'(x_k)| > \varepsilon$, то вычислить $F''(x_k)$ и если $F''(x_k) > 0$ перейти к шагу 4.

В противном случае закончить вычисление связи с нарушением обязательного условия $F''(x_k) > 0$.

Шаг 4. Вычислить $x_{k+1} = x_k - \dfrac{F'(x_k)}{F''(x_k)}$.

Шаг 5. Принять $k = k+1$ и перейти к шагу 3.

Замечание. В связи с выбором начального приближения x_0, удаленного достаточно далеко от искомого решения x^*, возможно, что последовательность $\{x_k\}$ будет расходиться. В этом случае рекомендуется найти лучшее начальное приближение x_0 другим методом (метод золотого сечения и т. д.).

3.5. Аппроксимация кривыми

Аппроксимировать – означает «приближенно заменять». Допустим, известны значения некоторой функции в заданных точках. Требуется найти промежуточные значения этой функции. Это так называемая задача о *восстановлении функции*. Кроме того, при проведении расчетов сложные функции удобно заменять алгебраическими многочленами или другими элементарными функциями, которые достаточно просто вычисляются (*задача о приближении функции*).

✓ Аппроксимация – определение в явном виде параметров функции, описывающей распределение точек.

✓ Интерполяция – определение промежуточных значений функции по известному дискретному набору значений функции.

✓ Экстраполяция – определение значений функции за пределами первоначально известного интервала.

Учесть информацию, содержащуюся в относительных изменениях значений $f(x)$ в пробных точках, позволяют *методы полиномиальной аппроксимации*, основная идея которых состоит в том, что для функции $f(x)$ строится аппроксимирующий многочлен, и его точка минимума служит приближением к x^*. Для эффективного использования этих методов на функцию $f(x)$, кроме унимодальности, налагается дополнительное требование достаточной гладкости (по крайней мере, непрерывности). Для практики весьма важен случай аппроксимации функции многочленом. При этом коэффициенты a_i будут подбираться так, чтобы достичь наименьшего отклонения многочлена от данной функции.

Теоретическим обоснованием указанных методов является известная из математического анализа теорема Вейерштрасса об аппроксимации, согласно которой непрерывную на отрезке функцию можно с любой точностью приблизить на этом отрезке некоторым полиномом. Для повышения точности аппроксимации можно, во–первых, увеличивать порядок полинома и, во-вторых, уменьшать длину отрезка аппроксимации. Первый путь приводит к усложнению вычислительных процедур, поэтому на практике используются аппроксимирующие полиномы не выше третьего порядка. В то же время уменьшить отрезок, содержащий точку минимума унимодальной функции, не представляет особого труда.

Постановка задачи интерполяции. На интервале $[a, b]$ заданы точки x_i, $i = 0,1,...,N$; $a \le x_i \le b$, и значения неизвестной функции в этих точках $f_i(x)$. Требуется найти функцию $F(x)$, принимающую в точках x_i те же значения $f_i(x)$. Точки x_i называются узлами интерполяции, а условия $F(x_i) = f_i(x)$ – условиями интерполяции. При этом $F(x)$ ищем только на отрезке $[a, b]$. Такая задача имеет много решений, т. к. через заданные точки (x_i, f_i), $i = 0,1,...,N$ можно провести бесконечно много кривых, каждая из которых будет графиком функции, для которой выполнены все условия интерполяции. Для практики важен случай аппроксимации функции многочленами.

Методы интерполяции можно разделить на *локальные* и *глобальные*. В случае *локальной интерполяции* на каждом интервале $[x_{i-1}, x_i]$ строится отдельный полином. В случае *глобальной интерполяции* отыскивается единый полином на всем интервале $[a, b]$. При этом искомый полином называется *интерполяционным полиномом*.

Локальная интерполяция методом парабол. В простейшем методе полиномиальной аппроксимации – методе парабол используются полиномы второго порядка. На каждой итерации этого метода строят квадратный трехчлен, график которого (парабола) проходит через три выбранные точки графика функции $f(x)$ (Рис. 3.6).

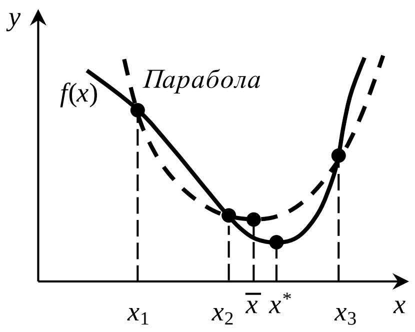


Рис. 3.6. Аппроксимация параболой

Рассмотрим унимодальную на отрезке $[a, b]$ функцию $f(x)$, достигающую минимума во внутренней точке этого отрезка. Выберем три точки x_1, x_2 и x_3 отрезка $[a, b]$, для которых выполняются неравенства:

$$x_1 < x_2 < x_3,\ f(x_1) \geq f(x_2) \leq f(x_3).$$

Из свойства унимодальности функции $f(x)$ следует, что $x^* \in [x_1, x_3]$. Построим квадратный трехчлен $q(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2)$, график которого проходит через точки $(x_1, f(x_1))$, $(x_2, f(x_2))$ и $(x_3, f(x_3))$ графика функции $f(x)$. Если $f(x_1) = f(x_2) = f(x_3)$, то поиск точки x^* на этом закончен, так как из свойства унимодальности функции $f(x)$ следует, что она достигает минимума в каждой точке отрезка $[x_1; x_3]$. Тогда из двойного неравенства вытекает, что ветви искомой параболы направлены вверх, а точка минимума $\bar{x}$ трехчлена $q(x)$ принадлежит отрезку $[x_1; x_3]$. Коэффициенты a_0, a_1 и a_2 определяем из системы уравнений:

$$q(x_1) = f(x_1) = f_1;$$

$$q(x_2) = f(x_2) = f_2;$$

$$q(x_3) = f(x_3) = f_3,$$

где $q(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2)$.

Выражения для коэффициентов имеют вид:

$$a_0 = f_1, \qquad a_1 = \frac{f_2 - f_1}{x_2 - x_1}, \qquad a_2 = \frac{1}{x_3 - x_2}\left(\frac{f_3 - f_1}{x_3 - x_1} \cdot \frac{f_2 - f_1}{x_2 - x_1}\right)$$

Значение точки минимума $\bar{x}$ квадратного трехчлена $q(x)$ вычислим, приравняв его производную к нулю:

$$\bar{x} = \frac{1}{2}\left(x_1 + x_2 - \frac{a_1}{a_2}\right) = \frac{1}{2}\left[x_1 + x_2 - \frac{(f_2 - f_1)(x_3 - x_2)}{x_2 - x_1} / \left(\frac{f_3 - f_1}{x_3 - x_1} - \frac{f_2 - f_1}{x_2 - x_1}\right)\right]$$

Значение $\bar{x}$ служит очередным приближением метода парабол к искомой точки минимума x^*. Далее описанная процедура повторяется для новых точек x_1, x_2, x_3, удовлетворяющих исходным неравенствам.

Выбрать точки среди x_1, x_2, x_3 и $\bar{x}$ можно с помощью перехода от исходного к новому отрезку $[x_1; x_3]$, содержащему точку x^*, методом исключения отрезков. Для этого перехода используют пробные точки x_2 и $\bar{x}$ и сравнивают значения $f(x)$ в этих точках . Начало и конец нового отрезка, а также пробная точка, попавшая на него, образуют новую тройку точек.

Заметим, что на каждой итерации метода парабол, кроме первой, определяется только одно новое значение $f(x)$. Условием окончания поиска служит близость к нулю разности Δ чисел $\bar{x}$, найденных на данной и предыдущей итерациях, т. е. неравенство $|\Delta| \le \varepsilon$, где ε – заданное число, характеризующее точность определения x^*.

Алгоритм метода парабол:

Шаг 1. Выбрать точки x_1, x_2, x_3, удовлетворяющие условиям метода.

Шаг 2. Найти $\bar{x}$ по формуле. На первой итерации перейти к шагу 4, на остальных – к шагу 3.

Шаг 3. Проверка на окончание поиска. Сравнить модуль разности значений $\bar{x}$ на двух сопредельных итерациях Δ с числом ε. Если $|\Delta| \le \varepsilon$, то поиск завершить, полагая $x^* \approx \bar{x}$, $f^* \approx f(x)$, иначе – перейти к шагу 4.

Шаг 4. Вычислить значение $f(\bar{x})$.

Ш а г 5. Определить новую тройку чисел x_1, x_2, x_3. Присвоить $f(x_1)$, $f(x_2)$ и $f(x_3)$ соответствующие значения $f(x)$ найденные ранее. Перейти к шагу 2.

П р и м е р . Методом парабол найти экстремум функции $f(x) = x^4 + e^{-x} \to min$, $x \in [0;1]$ с точностью $|\Delta| \le \varepsilon = 0.0025$.

Р е ш е н и е . И т е р а ц и я 1.

Ш а г 1. Выберем точки $x_1 = 0.25$, $x_2 = 0.5$, $x_3 = 0.75$. Функция принимает в этих точках значения, соответственно $f_1 = 0.7827$, $f_2 = 0.6690$, $f_3 = 0.7888$, удовлетворяющие неравенствам. Переходим к шагу 2.

Ш а г 2. По формуле находим: $\bar{x} = 0.4968$. Переходим к шагу 4.

Ш а г 4. Вычисляем $f(\bar{x}) = 0.6694$. Переходим к шагу 5.

Ш а г 5. На данной итерации имеем: $x_1 < \bar{x} < x_2 < x_3$, $f(\bar{x}) > f(x_2)$, следовательно, $x^* \in [\bar{x}; x_3]$. Поэтому полагаем: $x_1 = \bar{x} = 0.4968$, $f(x_1) = f(\bar{x}) = 0.6694$, а точки x_2, x_3 и значения $f(x)$ в них не изменяются. Переходим к следующей итерации, начиная с шага 2.

И т е р а ц и я 2.

Ш а г 2. Находим $\bar{x} = 0.5224$. Переходим к шагу 3.

Ш а г 3. $\Delta = |0.4968 - 0.5224| = 0.026 > 0.0025$, поэтому переходим к шагу 4.

Ш а г 4. Вычисляем $f(\bar{x}) = 0.6676$. Переходим к шагу 5.

Ш а г 5. На этой итерации $x_1 < x_2 < \bar{x} < x_3$, $f(x_2) > f(\bar{x})$, поэтому $x^* \in [x_2; x_3]$ и полагаем $x_1 = x_2 = 0.5$, $f(x_1) = f(x_2) = 0.6690$, $x_2 = \bar{x} = 0.5524$, $f(x_2) = f(\bar{x}) = 0.6676$, а точка x_3 и значение $f(x_3)$ остаются прежними. Переходим к следующей итерации.

И т е р а ц и я 3.

Ш а г 2. Находим $\bar{x} = 0.5248$. Переходим к шагу 3.

Шаг 3. Определяем $\Delta = \left| 0.5224 - 0.5248 \right| = 0.0024 < 0.0025$, т. е. требуемая точность достигнута. Поэтому полагаем $x^* = \bar{x} = 0.525$.

Отметим, что в результате трех итераций (пяти вычислений $f(x)$) точка x^* была найдена с весьма высокой точностью (сравните с точным до четвертого знака значением $x^* = 0.5283$).

Метод кубической интерполяции. В этом методе в качестве аппроксимирующего полинома используется многочлен третьей степени. Предполагается, что $F(x)$ – выпуклая дифференцируемая функция и найден интервал (x_1, x_2), содержащий ее точку минимума x^*. Это означает, что $F'(x_1) < 0$, а $F'(x_2) > 0$.

Аппроксимирующий кубический полином записывается в виде:

$$q(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2) + a_3(x - x_1)^2(x - x_2),$$

где неизвестными являются коэффициенты a_0, a_1, a_2, a_3.

Для определения неизвестных коэффициентов предполагается, что в точках x_1 и x_2 значения функции $F(x)$ и аппроксимирующего полинома $q(x)$ совпадают. Это позволяет найти два коэффициента. Для определения недостающих двух коэффициентов дополнительно полагают, что в точках x_1 и x_2 равны и их производные:

$$q'(x) = a_0 + a_1 + a_2(2x - x_1 - x_2) + 2a_3(x - x_1)(x - x_2) + a_3(x - x_1)^2 = F'(x).$$

Тогда уравнения для определения коэффициентов полинома примут вид:

$$q(x_1) = a_0 = F(x_1) \Rightarrow a_0;$$

$$q(x_2) = a_0 + a_1(x_2 - x_1) = F(x_2) \Rightarrow a_1;$$

$$q'(x_1) = a_1 + a_2(x_1 - x_2) = F'(x_1) \Rightarrow a_2;$$

$$q'(x_2) = a_1 + a_2(x_2 - x_1) + a_3(x_2 - x_1)^2 = F'(x_2) \Rightarrow a_3.$$

Для определения точки минимума производная $q'(x)$ приравнивается к нулю. В результате получается квадратное уравнение, имеющее два корня.

Из двух корней выбирается тот, который принадлежит интервалу (x_1, x_2). Этот корень $\widetilde{x}$ и является точкой приближения $q(x)$ на отрезке $[x_1, x_2]$ и используется в качестве приближения к точке минимума x^*.

Новыми точками x_2 и x_2 для осуществления следующей итерации являются концы того из отрезков $[a, \widetilde{x}]$ и $[\widetilde{x}, b]$, который содержит точку x^*. Его определяют по знаку производной $F'(\widetilde{x})$.

Если $F'(\widetilde{x}) > 0$, то точка $\widetilde{x}$ лежит на участке монотонного возрастания $F(x)$, поэтому $x^* < \widetilde{x}$ и точку минимума следует искать на отрезке $[a, \widetilde{x}]$.

Если $F'(\widetilde{x}) < 0$, то точка $\widetilde{x}$ лежит на участке монотонного убывания $F(x)$, поэтому $x^* > \widetilde{x}$ и точку минимума следует искать на отрезке $[\widetilde{x}, b]$.

Равенство $F'(\widetilde{x}) = 0$ означает, что точка минимума найдена точно и $x^* = \widetilde{x}$.

Алгоритм поиска минимума функции методом кубической интерполяции:

Шаг 1. Задать начальную точку x_0, величину шага $\Delta x > 0$, ε_1, ε_2 – малые положительные числа, характеризующие точность ($F'(\widetilde{x}) < \varepsilon_1$, $\left|\dfrac{\bar{x} - x_1}{\widetilde{x}}\right| \le \varepsilon_2$).

Шаг 2. Вычислить производную $F'(x_0)$.

Шаг 3. Проверить знак производной в точке x_0.

Если $F'(x_0) < 0$, то вычислить $x_{k+1} = x_k + 2^k \Delta x$, $k = 0,1,\ldots$ до точки x_m, в которой $F'(x_{m-1}) * F'(x_m) \le 0$.

Если $F'(x_0) > 0$, то вычислить $x_{k+1} = x_k - 2^k \Delta x$, $k = 0,1,\ldots$ до точки x_m, в которой $F'(x_{m-1}) * F'(x_m) \le 0$.

Шаг 4. Положить $x_1 = x_{m-1}$, $x_2 = x_m$ и вычислить $F(x_1) = F_1$, $F'(x_1) = F_1'$, $F(x_2) = F_2$, $F'(x_2) = F_2'$.

Шаг 5. Найти точку минимума кубического интерполяционного полинома по формуле:

$$\tilde{x} = \begin{cases} x^2, & \mu < 0; \\ x^2 - \mu(x_2 - x_1), & 0 \le \mu \le 1; \\ x_1, & \mu > 1, \end{cases}$$

где $\mu = \dfrac{F_2' + w - z}{F_2' - F_1' + 2w}$, $z = \dfrac{3(F_1 - F_2)}{x_2 - x_1} + F_1' + F_2'$,

$$w = \begin{cases} (z^2 - F_1' * F_2')^{1/2}, & x_1 < x_2; \\ -(z^2 - F_1' * F_2')^{1/2}, & x_1 > x_2. \end{cases}$$

Вычислить $F(\tilde{x})$.

Шаг 6. Проверить условие убывания.

Если $F(\tilde{x}) < F(x_1)$, перейти к этапу 7.

Если $F(\tilde{x}) \ge F(x_1)$, вычислить $\tilde{x}$ по формуле $\tilde{x} = \tilde{x} - \dfrac{1}{2}(\tilde{x} - x_1)$ до тех пор, пока не будет выполнено неравенство $F(\tilde{x}) \le F(x_1)$.

Шаг 7. Проверить выполнение условий окончания: $F'(\tilde{x}) < \varepsilon_1$, $\left|\dfrac{\tilde{x} - x_1}{\tilde{x}}\right| \le \varepsilon_2$.

Если оба условия выполнены, то процедура поиска точки минимума закончена и $x^* \approx \tilde{x}$.

Если хотя бы одно из условий не выполнено, то положить либо $x_1 = \tilde{x}$, $x_2 = x_1$ при $F'(\tilde{x}) * F'(x_1) < 0$, либо $x_1 = \tilde{x}$, $x_2 = x_2$ при $F'(\tilde{x}) * F'(x_2) < 0$. Перейти к шагу 5.

Замечание. На шагах 2-3 реализуется эвристическая процедура поиска границ интервала неопределенности, где изменение знака производной свидетельствует о переходе через точку минимума.

Интерполяционный метод Лагранжа. Принципиальные результаты в исследовании проблем интерполяции получил Жозеф Луи Лагранж. Исходя из выражений общих членов для рекуррентных рядов, найденных им в 1776 году, он нашел выражение интерполяционных многочленов, в последствие, получивших название по его имени. Интерполяционный многочлен Лагранжа – многочлен минимальной степени, принимающий данные значения в данном наборе точек.

Лагранж предложил строить интерполяционный полином в виде разложения: $L_n(x) = \sum_{i=0}^{n} l_i(x)\, y_i = y_0\, l_0(x) + y_1\, l_1(x) + \ldots + y_n\, l_n(x)$, где $l_i(x)$ – базисные функции.

Для того, чтобы такой полином, удовлетворял условиям Лагранжа, т. е. был бы интерполяционным, базисные функции $l_i(x)$ должны обладать следующими свойствами:

- ✓ быть полином степени n;
- ✓ удовлетворять условию: $l_i(x) = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$.

Приведем одну из форм записи интерполяционного многочлена Лагранжа:

$$L_n(x) = \sum_{i=0}^{n} y_i \frac{(x - x_0)(x - x_1)\cdots(x - x_{i-1})(x - x_{i+1})\cdots(x - x_n)}{(x_i - x_0)(x_i - x_1)\cdots(x_i - x_{i-1})(x_i - x_{i+1})\cdots(x_i - x_n)}.$$

Если ввести в рассмотрение многочлен специального вида $(n+1)$-й степени: $\omega_{n+1}(x) = (x - x_0)(x - x_1)\cdots(x - x_n)$, тогда многочлен Лагранжа можно представить в виде:

$$L_n(x) = \sum_{i=0}^{n} y_i \frac{\omega_{n+1}(x)}{(x - x_i)\,\omega_n'(x_i)}.$$

Если узлы упорядочены по величине, т. е. $x_{i+1} > x_i$ для всех i, то величины $\{h_i = x_{i+1} - x_i;\, 0 \le i \le n\}$ называют шагами интерполяции. Если

$h_i = h = const$, то есть $x_i = x_0 + h$ $(0 \le i \le n)$, то говорят об интерполяции по равноотстоящим узлам.

Интерполяционная формула Лагранжа заметно упрощается, если узлы интерполяции равноотстоящие. Вводя обозначение $t = \frac{x - x_0}{h}$, получим:

$$L_n(t) = \sum_{i=0}^{n} \frac{(-1)^{n-i} t(t-1)\cdots(t-n)}{i!(n-i)!(t-i)} f(x_i).$$

Рассмотрим частные случаи:

1. Пусть $n = 1$. Тогда $L_1(t) = (t-1)y_0 + ty_1$.

2. Пусть $n = 2$. Имеем $L_2(t) = \frac{1}{2}(t-1)(t-2)y_0 - t(t-2)y_1 + \frac{1}{2}t(t-1)y_2$.

В отличие от интерполяционного полинома в канонической форме для вычисления значений полинома Лагранжа не требуется предварительно определять коэффициенты полинома путем решения системы уравнений. Однако для каждого значения аргумента x полином Лагранжа приходится пересчитывать вновь, коэффициенты же канонического полинома вычисляются только один раз. Поэтому практическое применение полинома Лагранжа оправдано только в том случае, когда интерполяционная функция вычисляется в сравнительно небольшом количестве точек x.

З а м е ч а н и е. Интерполяционный полином Лагранжа оказывается очень удобным для приближенного вычисления определенных интегралов. Если, например, некоторую функцию заменить интерполяционным полином Лагранжа $f(x) \approx \sum_{i=0}^{n} f(x_i) l_i(x)$, то определенный интеграл от нее может быть вычислен следующим образом $\int_a^b f(x)dx \approx \sum_{i=0}^{n} f(x_i) \int_a^b l_i(x)dx$. Значения интегралов от $l_i(x)$ не зависят от $f(x)$ и могут быть легко вычислены аналитически.

На практике наиболее часто используется интерполяция первой $(n=1)$ *линейная*, второй $(n=2)$ *квадратичная* и третьей $(n=3)$ степени *кубическая*.

Пример. Положим $n=1$. В этом случае получаем две точки и значения функции в них:

x_0	x_1
$y_0=f(x_0)$	$y_1=f(x_1)$

Построенная интерполяционная формула Лагранжа дает уравнение прямой, проходящей через две заданные точки (Рис. 3.7). Соответствующая формула для записи многочлена Лагранжа первой степени получается в виде:

$$L_1(x)=\frac{x-x_1}{x_0-x_1}y_0+\frac{x-x_0}{x_1-x_0}.$$

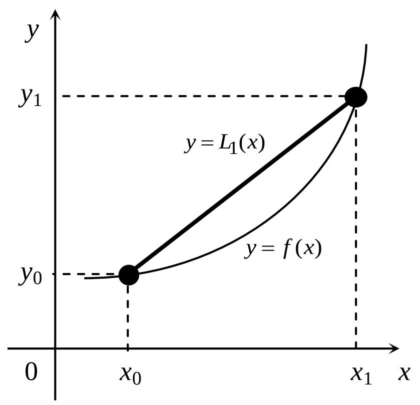


Рис. 3.7. Линейный полином

Пример. Положим $n=2$. Тогда мы получим уравнение параболы, проходящей через три точки.

x_0	x_1	x_2
$y_0=f(x_0)$	$y_1=f(x_1)$	$y_2=f(x_2)$

Соответствующая формула для записи многочлена Лагранжа второй степени:

$$L_2(x)=\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}y_0+\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}y_1+\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}y_2.$$

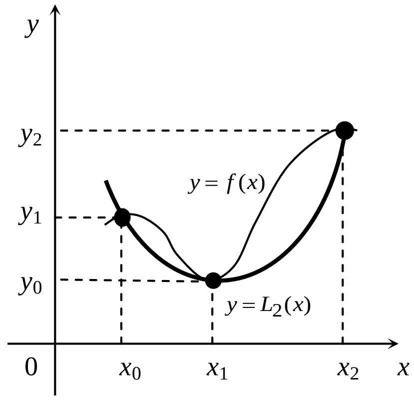


Рис. 3.8. Квадратичный полином

Вместе с тем интерполяционный метод Лагранжа имеет ряд существенных недостатков:

✓ поскольку степень многочлена Лагранжа определяется количеством узлов (минус 1), то любая попытка повысить точность аппроксимации путем увеличения количества узлов влечет за собой увеличению степени полинома;

✓ формула для расчета достаточно громоздка. Каждое слагаемое формулы является многочленом n-ой степени;

✓ если степень полинома выше 5, то на кривой появляется «волнистость», которая получила название эффекта Рунге–Мерей. Возможность улучшить эту ситуацию дает подбор расположения узлов в зависимости от конкретной функции в интерактивном режиме, но такая процедура достаточно неудобна.

Еще одним из путей уменьшения волнистости состоит в разбиении интервала интерполирования на несколько подынтервалов, для каждого из которых формируют многочлены Лагранжа, но уже более низких степеней. В дальнейшем многочлены Лагранжа «склеивают», однако объединенная аппроксимирующая функция может быть не дифференцируемой на границах подынтервалов.

Используются и другие записи (единственного) интерполяционного многочлена. Особенно часто используют запись в форме Ньютона.

Интерполяционный многочлен Ньютона. При равномерной дискретизации ($\Delta x = const$) удобно пользоваться интерполяционной формулой Ньютона. Как и другие интерполяционные формулы, она служит для построения многочлена n-й степени, который совпадает в $(n + 1)$ точке со значениями неизвестной искомой функции $y = f(x)$.

Пусть в точках $x_0, x_1, \ldots, x_{n+1}$ значения функции $y = f(x)$ равны соответственно $y_0 = f(x_0)$, $y_1 = f(x_1)$, ..., $y_{n+1} = f(x_{n+1})$. Построим интерполяционный многочлен Ньютона с помощью метода неопределенных коэффициентов. Для этого запишем искомый многочлен в виде:

$$P_m(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2)\ldots$$

Последовательно подставляя в эту формулу вместо x данные значения $x_0, x_1, \ldots, x_{n+1}$ и получим для нахождения неопределенных коэффициентов $b_0, b_1, \ldots, b_n$ «треугольную» систему уравнений:

$$\begin{cases} y_0 = b_0 \\ y_1 = b_0 + b_1(x_1 - x_0) \\ y_2 = b_0 + b_1(x_2 - x_0) + b_2(x_2 - x_0)(x_2 - x_1) \\ \ldots \\ y_{n+1} = b_0 + b_1(x_{n+1} - x_0) + \ldots + b_n(x_{n+1} - x_0)\ldots(x_{n+1} - x_n) \end{cases}$$

Система получается последовательной подстановкой в выражение многочлена вместо x значения x_0 в правой части равенства обратились в нуль все слагаемые, кроме первого: там везде был множитель $(x - x_0)$, обратившийся в нуль; при подстановке $x = x_1$ обратились в нуль все слагаемые, кроме первого и второго – они содержат множитель $(x - x_1)$ и т. д.). Полученную систему удобно решать также последовательно: из первого уравнения находим свободный член искомого многочлена b_0; подставив его во второе уравнение, находим коэффициент b_1 при первой степени x в искомом многочлене: $b_1 = \dfrac{y_1 - b_0}{x_1 - x_0}$ и т. д.

Важно, что для интерполяционного многочлена Ньютона можно выписать явные выражения коэффициентов через исходные данные задачи, а также и оценки точности замены неизвестной функции $f(x)$ этим многочленом.

Конечные разности. Пусть функция $y = f(x)$ задана таблицей своих значений. Величину $\Delta y_i = y_{i+1} - y_i$ принято называть *конечной разностью первого порядка* функции $y = f(x)$ в точке x_i (с шагом h). *Конечная разность второго порядка* определяется формулой $\Delta^2 y_i = \Delta y_{i+1} - \Delta y_i$. Аналогично определяются конечные разности третьего и более высоких порядков.

Общее определение *конечной разности порядка* k имеет вид:

$$\Delta^k y_i = \Delta^{k-1} y_{i+1} - \Delta^{k-1} y_i, \ \ k \ge 1 \text{ и } \Delta^0 y_i = y_i.$$

Если интерполируемая функция задана на таблице с постоянным шагом h (т. е. $x_i = x_0 + ih,\ 0 \le i \le n$), то многочлен Ньютона можно записать в следующем виде:

$$P_n(x) = P_n(x_0 + ht) = y_0 + \frac{\Delta y_0}{1!} t + \frac{\Delta^2 y_0}{2!} t(t-1) + \frac{\Delta^3 y_0}{3!} t(t-1)(t-2) + $$

$$+ \cdots + \frac{\Delta^n y_0}{n!} t(t-1)\ldots(t-n+1),$$

где $t = \dfrac{x - x_0}{h}$.

Полученный многочлен называется интерполяционным многочленом Ньютона с конечными разностями *для интерполяции вперед.* Если $t = \dfrac{x - x_n}{h}$, то можно записать многочлен в виде интерполяционного многочлена Ньютона с конечными разностями *для интерполяции назад*:

$$P_n(x) = P_n(x_n + ht) = y_n + \frac{\Delta y_{n-1}}{1!} t + \frac{\Delta^2 y_{n-2}}{2!} t(t+1) +$$

$$+ \frac{\Delta^3 y_{n-3}}{3!} t(t+1)(t+2) + \ldots + \frac{\Delta^n y_0}{n!} t(t+1)\cdots(t+n-1).$$

Частные случаи. Если в линейной интерполяции значение функции в точке, отличной от узлов интерполяции, определяется по двум известным значениям функции в двух узлах интерполяции, между которыми расположено интересующее нас значение аргумента x:

x_i	$x_{i+1} = x_i + h$
$y_i = f(x_i)$	$y_{i+1} = f(x_{i+1})$

Первая интерполяционная формула Ньютона $P_1(x) = y_i + t\Delta y_i$, где $t = \dfrac{x - x_i}{h}$, $\Delta y_i = y_{i+1} - y_i$ – первая конечная разность в точке x_i. Для получения приближенного значения функции $y = f(x)$ по формуле Ньютона достаточно к табличному значению y_i прибавить поправку, равную $\Delta y_i (x - x_i) h^{-1}$.

В случае квадратичной интерполяции используют три значения функции:

x_{i-1}	x_i	x_{i+1}
$y_{i-1} = f(x_{i-1})$	$y_i = f(x_i)$	$y_{i+1} = f(x_{i+1})$

Вторая интерполяционная формула Ньютона:

$$P_2(x) = y_{i-1} + t\Delta y_{i-1} + \frac{t(t-1)}{2!} \Delta^2 y_{i-1}.$$

Сравнение записей многочленов в форме Лагранжа и Ньютона. Для вычисления значения функции $f(x)$ в точке x, не являющейся узлом интерполяции, можно положить $f(x) \approx G_n(x)$, (где $G_n(x)$, либо $L_n(x)$, либо $P_n(x)$). Пусть многочлен $G_n(x)$ уже найден, но для уточнения необходимо привлечь еще один узел x_{n+1} и значение $f(x_{n+1})$ в нем. Тогда для вычисления $G_{n+1}(x)$ с помощью формулы Лагранжа нужно заново провести всю работу. Для вычисления по формуле Ньютона нужно рассчитать только поправку $\dfrac{\Delta^{n+1} y_0}{(n+1)!} t(t-1)\cdots(t-n)$ или $\dfrac{\Delta^{n+1} y_0}{(n+1)!} t(t+1)\cdots(t+n)$. Кстати, сразу будет видно, насколько она велика.

Отсюда видно различие между формулами Ньютона и Лагранжа. В формуле Лагранжа каждое из слагаемых представляет многочлен n-ой степени и все эти слагаемые равноправны. Поэтому невозможно заранее (то есть до производства вычислений) пренебрегать какими-либо из них. В формулу же Ньютона входят в качестве слагаемых многочлены *повышающихся степеней*, причем коэффициентами при них служат последовательные конечные разности, деленные на факториалы. Как уже отмечалось, последовательные разности обычно довольно быстро уменьшаются. Поэтому возникает возможность, не учитывать в формуле Ньютона тех слагаемых, коэффициенты при которых становятся пренебрежимо малыми. Сравнение форм Лагранжа и Ньютона для интерполяционного многочлена позволяет рекомендовать использование представления в форме Лагранжа:

- ✓ во-первых, в теоретических исследованиях, например при изучении вопроса о сходимости $G_n(x)$ к функции f при $n \to \infty$;
- ✓ во-вторых, при интерполировании нескольких функций на одной и той же сетке узлов, поскольку в этом случае можно один раз вычислить множители Лагранжа и использовать их для интерполяции всех функций.

Погрешность интерполяции. Ошибка приближения функции интерполяционным многочленом n-й степени в точке x – это разность $\rho_n(x) = f(x) - G_n(x)$. Оценить значение погрешности позволяет следующая теорема.

Теорема. Пусть функция $f(x)$ дифференцируема $(n+1)$ раз на отрезке $[a;\, b]$, содержащем узлы интерполяции x_i, $0 \le i \le n$. Тогда:

$$\rho_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega_{n+1}(x),\ \xi \in [a;\, b].$$

Откуда следует оценка погрешности интерполяции:

$$|\rho_n(x)| \le \frac{M_{n+1}}{(n+1)!}|\omega_{n+1}(x)|,\ M_{n+1} = \max_{[a;\, b]}\left|f^{n+1}(x)\right|.$$

Погрешность интерполяции в точке $x \in [a; b]$ относительно переменной t можно представить в виде:

$$\rho_n(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1}t(t-1)\ldots(t-n),$$

где ξ – некоторая точка, принадлежащая интервалу $(a;\, b)$.

Если $M_{n+1} = \max\limits_{[a;\, b]}\left|f^{n+1}(x)\right|$, то оценка погрешности интерполяции в точке $x \in [a; b]$, имеет вид: $|\rho| \le \dfrac{M_{n+1}}{(n+1)!}h^{n+1}|t(t-1)\ldots(t-n)|$.

Вопросы самоконтроля:

1. Глобальный и локальный минимум функции одной переменной.
2. Условие унимодальности функции и как это условие используется.
3. Понятие выпуклой функции. Ее основные свойства.
4. Как ведет себя производная в области точки экстремума.
5. Верно ли утверждение, что всякая выпуклая непрерывная на отрезке функция является на этом отрезке унимодальной?
6. Поведение касательной к выпуклой функции в области экстремума.
7. Можно ли считать, что глобальный минимум является локальным? А наоборот?
8. Необходимое и достаточное условие экстремума одномерной функции.
9. Классический алгоритм поиска точки экстремума функции одной переменной.
10. Различие между пассивным и последовательным поиском.
11. Прямые методы одномерной оптимизации.
12. Метод равномерного поиска.
13. Интервал неопределенности в задачах одномерной оптимизации.
14. Методы сокращения интервала неопределенности.
15. Условия «останова» применения методов.
16. Метод дихотомии?
17. Особенности применения метода дихотомии.
18. Метод деление отрезка пополам. Отличие от метода дихотомии.
19. Понятие золотого сечения.
20. Алгоритм метода «золотого сечения».
21. Преимущества и особенности метода «золотого сечения».
22. Последовательность Фибоначчи.
23. Принципиальное отличие метода Фибоначчи от других методов поиска минимума унимодальной функции.
24. Процесс сокращения интервала неопределенности по методу Фибоначчи.
25. Сравнение методов прямого поиска и сокращения интервалов неопределенности по скорости сходимости.
26. Методы одномерной оптимизации с использованием производной.
27. Суть и алгоритм метода средней точки.
28. Метод хорд.
29. Отличие метода Ньютона от методов средней точки и хорд.
30. Какие трудности возникают в методе квадратичной аппроксимации?
31. Каким образом сравнивают эффективность методов прямого поиска
32. Какой из интерполяционный полиномов имеет наименьшую степень?
33. В чем состоят недостатки метода Лагранжа?
34. В каких случаях удобно использовать метод Ньютона?
35. Понятия аппроксимации, интерполяции и экстраполяции.
36. Методы полиномиальной аппроксимации и их теоретическое обоснование.
37. Постановка задачи об аппроксимации функции.
38. Постановка задачи интерполяции.

39. Классификация методов интерполяции.
40. Что такое интерполяционный многочлен?
41. Локальная интерполяция методом парабол. Условия окончания поиска точки минимума.
42. Кубическая интерполяция. Преимущества и недостатки.
43. От чего зависит интерполяционный многочлен Лагранжа?
44. От чего зависит погрешность интерполяционного многочлена Лагранжа?
45. Что такое базисные функции многочлена Лагранжа.
46. Условие упрощения интерполяционной формулы Лагранжа.
47. Недостатки метода Лагранжа.
48. Условия выбора интерполяционной формулы Ньютона.
49. Достоинства и недостатки записей в форме Ньютона?
50. Что представляет собой графики интерполяционного многочлена при значениях $n=1$ и $n=2$?
51. Какими формулами (Лагранжа или Ньютона) удобнее пользоваться в случае равноотстоящих узлов интерполяции и почему?
52. Сколько интерполяционных многочленов можно построить для одной функции и одной системы узлов интерполяции?
53. Почему погрешность интерполяции для интерполяционных многочленов Лагранжа и Ньютона оценивается с помощью одной и той же формулы?
54. Понятие конечных разностей. Чем определяется их порядок?
55. Как составляются конечные разности?
56. Погрешность интерполяции.

Раздел 4. РЕГРЕССИОННЫЕ МОДЕЛИ

4.1. Основы регрессионного анализа

Регрессионный анализ – статистический метод исследования влияния одной или нескольких независимых переменных на зависимую переменную. Независимые переменные иначе называют регрессорами или предикторами, а зависимые переменные называются критериальными. Терминология *зависимых* и *независимых* переменных отражает лишь математическую связь переменных, а не их причинно-следственные отношения.

Цели регрессионного анализа:

- ✓ определение степени вариации зависимой переменной независимыми переменными;
- ✓ предсказание значения зависимой переменной;
- ✓ определение вклада отдельных независимых переменных в вариацию зависимой.

Регрессионный анализ нельзя использовать для определения наличия связи между переменными, поскольку наличие такой связи и есть предпосылка для применения анализа. Проблема регрессионного анализа характерна тем, что о распределениях изучаемых величин нет достаточной информации. Например, имеются основания предполагать, что случайная величина Y имеет некоторое распределение и нужно по результатам наблюдений определить значения параметров функции этого распределения. В зависимости от природы задачи и целей анализа результаты эксперимента по-разному интерпретируются в отношении переменной x. Довольно часто в практике исследований имеет место ситуация, когда важнейшие переменные, описывающие некоторый процесс, известны, а вот модель самого процесса неизвестна. В таких случаях возможны разные подходы.

Одним из них является построение эмпирических моделей в виде уравнений регрессии.

Для определения силы и направления связи между величинами в эксперименте используется модель, основанная на допущениях:

✓ величина x является контролируемой величиной, значения которой заранее задаются при планировании эксперимента;

✓ независимые переменные при различных измерениях одинаково распределены с нулевым средним и постоянной дисперсией.

В случае неконтролируемой переменной результаты наблюдений $(x_1, y_1), \ldots, (x_n, y_n)$ представляют собой выборку из некоторой двумерной совокупности. Методы регрессионного анализа одинаковы и в том, и в другом случае, однако интерпретация результатов различается (в последнем случае анализ существенно дополняется методами теории корреляции). Выбор вида функций иногда определяется по расположению экспериментальных значений (x, y) на диаграмме рассеяния, чаще из теоретических соображений. Функциональная зависимость между переменными x и y это правило, которое каждому элементу x из множества $\{X\}$ ставит в соответствие элемент y из множества $\{Y\}$.

Если график функции регрессии $y_x = f(x)$ или $x_y = \varphi(y)$ изображается прямой линией, то регрессию называют *линейной* ($y_x = ax + b$), в противном случае – *нелинейной*.

Для определения вида функции регрессии в системе координат наносят точки $(x; y_x)$, строят диаграмму рассеяния (Рис. 4.1) и по их расположению делают предположение о виде функции регрессии (выдвигают гипотезу).

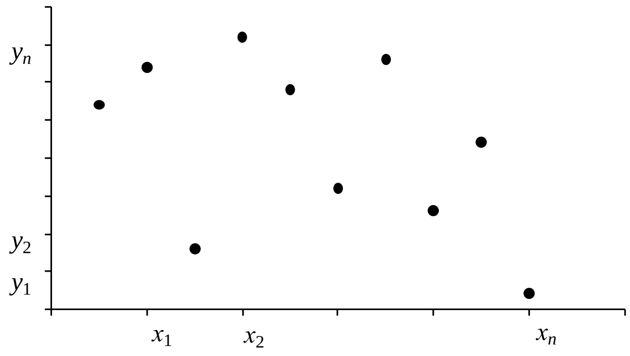


Рис. 4.1. Диаграмма рассеяния

Не существует общего правила для выбора подходящего вида функции. Можно лишь догадываться о форме уравнения регрессии. Однако существуют способы, с помощью которых можно проверить, является ли догадка удачной. Процесс нахождения теоретического уравнения линии регрессии заключается в выборе и обосновании типа прямой (кривой) и расчете его параметров. Результатом является возможность составить уравнение регрессии и получить количественную оценку влияния факторных признаков на результативный признак.

Регрессионный анализ естественным образом обобщается, когда зависимая переменная зависит не от одной, а от нескольких независимых переменных. Очевидно, что одновременный учет нескольких факторов, связанных с интересующей нас величиной, позволяет построить модель, точнее описывающую имеющиеся данные и лучше прогнозирующую зависимую переменную. Безусловно, процедура оценки наиболее эффективна при правильно спланированном эксперименте, но требуют рассмотрения и те случаи, когда в распоряжении имеются данные, полученные в заранее не спланированном эксперименте.

В таком случае следуют некоторым правилам, позволяющим выбрать наиболее подходящую модель с наименьшими возможными затратами:

✓ регрессионное уравнение должно содержать минимальное число коэффициентов;

✓ желательно, чтобы уравнение имело под собой содержательное обоснование. Например, демографические изменения, если нет ограничений на пищевые и другие ресурсы, осуществляются по экспоненте, поэтому модель этого процесса должна иметь соответствующую функциональную зависимость.

Лучшая процедура отбора наиболее подходящих моделей – пошаговый регрессионный анализ. Суть его в том, что отдельные переменные последовательно включаются в первоначальную модель и на каждом этапе анализируются, приводит ли добавление переменной к существенному или статистически значимому приближению предсказанных значений к эмпирическим данным.

В значительной мере достоверность полученных оценок зависит от некоторых предположений относительно поведения случайной ошибки:

✓ случайный характер – отдельные ошибки представляют собой случайные величины;

✓ нулевое среднее – каждое отклонение ошибки характеризуется нулевым математическим ожиданием и не зависит от значений x_i;

✓ дисперсия каждого отклонения равна и независима от x_i;

✓ отсутствие взаимосвязи (автокорреляции) ошибок;

✓ ошибка должна иметь нормальное распределение.

При выполнении этих условий можно оценить точность предсказания влияния x на y, так как именно предсказание является одной из главных целей регрессионного анализа.

4.2. Основной метод нахождения параметров уравнения регрессии

Классический подход к оцениванию параметров линейной регрессии основан на методе Ordinary Least Squares, OLS или наименьших квадратов (МНК), разработанный К. Гауссом и А. Лежандром. В его основу положена теория исследования на экстремум функции нескольких переменных.

Метод OLS (наименьших квадратов) позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений y фактических значений результативного признака от расчетных (теоретических) $\overline{y}_x$ минимальна:

$$S = \sum_i (y_i - \overline{y_{x_i}})^2 \to min .$$

Иными словами, из всего множества линий, линия регрессии на графике выбирается так, чтобы сумма квадратов расстояний по вертикали между экспериментальными точками и точками этой линией была минимальной:

$$S = \sum_i (y_i - \overline{y_{x_i}})^2 = \sum_i \left(y_i - ax_i - b\right)^2 = \sum_i \Delta_i^2 \to min .$$

Процесс выражения опытных данных функциональной зависимостью с помощью метода OLS состоит из двух этапов: сначала выбирают вид искомой формулы, а затем для данной формулы подбирают параметры.

Рассмотрим в качестве эмпирической формулы линейную зависимость. Если выборочное уравнение регрессии имеет вид $y_x = ax + b$, то параметры a и b находят из системы линейных уравнений, применяя метод Крамера:

$$\begin{cases} \left(\sum\limits_{i=1}^{n} x_i^2\right) \cdot a + \left(\sum\limits_{i=1}^{n} x_i\right) \cdot b = \left(\sum\limits_{i=1}^{n} x_i y_i\right); \\ \left(\sum\limits_{i=1}^{n} x_i\right) \cdot a + n \cdot b = \left(\sum\limits_{i=1}^{n} y_i\right); \end{cases}$$

где n – количество точек измерений.

$$\Delta = \begin{vmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{vmatrix} = n \cdot \sum x_i^2 - \left(\sum x_i^2\right)^2;$$

$$\Delta a = \begin{vmatrix} \sum y_i x_i & \sum x_i \\ \sum y_i & n \end{vmatrix} = n \cdot \sum y_i x_i - \sum x_i \cdot \sum y_i \, ;$$

$$\Delta b = \begin{vmatrix} \sum x_i^2 & \sum y_i x_i \\ \sum x_i & \sum y_i \end{vmatrix} = \sum x_i^2 \sum y_i - \sum x_i y_i \cdot \sum x_i \, .$$

Вычислим коэффициенты регрессии: $a = \dfrac{\Delta a}{\Delta}$; $b = \dfrac{\Delta b}{\Delta}$

или $a = \dfrac{n\sum y_i x_i - \sum y_i \sum x_i}{n\sum x_i^2 - \sum x_i^2}$; $\quad b = \dfrac{\sum x_i^2 \sum y_i - \sum x_i \sum y_i x_i}{n\sum x_i^2 - (\sum x_i)^2}$.

Критерием точности полученных коэффициентов уравнения регрессии является сумма квадратов отклонений значений расчетной и экспериментальной функций. В результате получим значение критерия, выраженное одним числом $\sum \Delta_i^2 = (y_{p_i} - y_{э_i})^2$.

Параметр a называется коэффициентом регрессии. Его величина показывает среднее изменение признака y на a единиц при изменении фактора x на одну единицу. Знак при коэффициенте a показывает направление связи: при $a > 0$ связь прямая, при $a < 0$ – обратная. Коэффициент b равен значению зависимой переменной y при $x = 0$, т. е. характеризует начальное состояние рассматриваемой зависимости – уровень отсчета. Параметр b может не иметь прикладного содержания. Интерпретировать можно знак: если $b > 0$, то относительное изменение результата y происходит медленнее, чем изменение фактора x. Иначе, вариация результата y меньше вариации фактора – коэффициент вариации по фактору x выше коэффициента вариации для результата y: $V_x > V_y$.

Например, если уравнение регрессии имеет вид $y_x = 2x + 5$, где x – уровень заработной платы персонала частной компьютерной фирмы, а y_x – полученная средняя покупательная способность работников внебюджетных организаций, то смысл коэффициентов $a = 2$ и $b = 5$ таков: при увеличении заработной платы на 1 условную единицу, покупательная способность (теоретически, в среднем) повышается на 2 две условных единицы. Если заработную плату «заморозить», покупательная способность составит 5 условных единиц.

Уравнение регрессии может быть найдено также в виде $x_y = cy + d$. Процесс нахождения значений коэффициентов уравнения и их суть при этом сохраняются, изменяются только «места» зависимых и независимых переменных.

Возможность достаточно понятной интерпретации коэффициентов линейной регрессии сделала этот вариант регрессионной модели распространенным в прикладных исследованиях. Зная уравнение регрессии, по точкам легко построить его наглядное представление – линию регрессии. А с помощью полученного уравнения можно, подставляя задаваемые значения фактора X, в прогнозируемом периоде получить планируемую величину показателя.

Целью анализа объекта исследования обычно является либо минимизация затрат ресурсов для достижения необходимого результата, либо максимизация эффективности использования имеющихся ограниченных ресурсов. Пусть проводится исследование по установлению необходимого типа и рационального номенклатурного перечня изделий предприятий (заводов, фирм, баз) по изготовлению комплектующих для систем автоматизированного контроля и управления усовершенствованного типа. По условиям использования различают крупные предприятия с номенклатурным перечнем изделий $R_б$ более 1 000 единиц, длительность работы (T) на рынке не менее 10 лет. Средние с перечнем $400 < R_б < 600$ и временем нахождения на рынке T от 3 до 5 лет и малые фирмы с перечнем изделий $50 < R_б < 100$.

Необходимый тип и рациональный перечень номенклатуры изделий могут быть определены из соотношения:

$$C = M + C_n + C_{mp},$$

где C – стоимость продукции, отнесенная на единицу устраиваемой системы автоматизированного контроля и управления; M – часть этой стоимости, не зависящая от искомого номенклатурного перечня изделий $R_б$; C_n – стоимость, обусловленная расходами по периодического изменения номенклатуры изделий, на каждые $R_б$ наименований; C_{mp} – часть стоимости, обусловленная затратами на обновление номенклатуры изделий предприятия при среднем изменении $R_б / 2$.

Если рассмотреть выявленное соотношение по составляющим, то в самом общем плане, когда требуется установить как рациональный перечень изделий $R_б$, так и тип предприятия изготовителя, величина M – отражает стоимость материалов $M_м$ и стоимость процессов производства – M_n, иначе $M = M_n + M_м$. Величину $M_м$ можно считать независимой от $R_б$. Характер функции $M_n = f(R_б)$ можно выявить на основе статистических данных стоимости изготовления систем автоматизированного контроля и управления на элементных базах, произведенных на предприятиях различных типов. В самом простейшем случае это может быть линейная зависимость вида: $M_n = a \cdot R_б + C_в$, где a – коэффициент связи, $C_в$ – стоимость изготовления. Для крупных и средних предприятий величина M_n должна быть меньше, чем для малых.

Пример. Найти коэффициенты линейной регрессии для функции $M_n = f(R_б)$ на основе экспериментальных данных. Обозначим, для удобства записи $y_э = M_n$ и $x_i = R_б$:

Аргумент, x_i	$x_1 = 1$	$x_2 = 2.2$	$x_3 = 4$	$x_4 = 4.5$	$x_5 = 6$	$x_6 = 9$
Функция, $y_э$	$y_1 = 1$	$y_2 = 1$	$y_3 = 2$	$y_4 = 4$	$y_5 = 4$	$y_6 = 0$

Решение. Выполним предварительные расчеты:

$$\sum_1^6 x_i = 26.7; \qquad \sum_1^6 y_i = 12; \qquad \sum_1^6 x^2{}_i = 159.09;$$

$$\sum_1^6 x_i y_i = 53.2; \qquad (\sum_1^6 x_i)^2 = 712.89.$$

Найдем значения коэффициентов уравнения регрессии:

$$a = \frac{6 \cdot 53.2 - 26.7 \cdot 12}{6 \cdot 159.09 - 712.89} = \frac{319.2 - 320.4}{954.54 - 712.89} = \frac{-1.2}{241.65} = -0.00497;$$

$$b = \frac{159.09 \cdot 12 - 26.7 \cdot 53.2}{6 \cdot 159.09 - 712.89} = \frac{1909.08 - 1420.44}{954.54 - 712.89} = \frac{488.64}{241.65} = 2.022.$$

Искомое уравнение линейной регрессии имеет вид: $y_p = -0.0045\,x + 2.022$. Вычислим «модельные» значения зависимой переменной и построим график линейной зависимости (Рис. 4.2.).

x_i	$x_1 = 1$	$x_2 = 2.2$	$x_3 = 4$	$x_4 = 4.5$	$x_5 = 6$	$x_6 = 9$
y_p	2.015	2.009	3.5	4.25	5	0

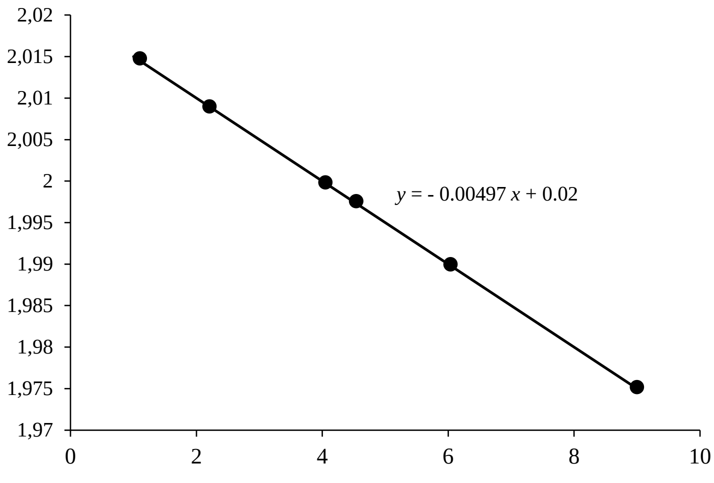


Рис. 4.2. График линейной регрессии

В практических исследованиях всегда имеет место некоторое рассеяние точек относительно линии регрессии. Оно обусловлено влиянием прочих, не учитываемых (не наблюдаемых или не измеряемых) в уравнении факторов. Имеют место отклонения фактических данных от теоретических результатов $y_э - y_p$. Величина этих отклонений и лежит в основе расчета остаточной дисперсии. Обозначим $\Delta = y_э - y_p$ и рассчитаем критерий оценки построенного уравнения регрессии:

$$\sum_1^6 \Delta_i^2 = \sum_1^6 \left(y_{эi} - y_{pi}\right)^2 = 1.03 + 1.02 + 0 + 4.01 + 3.9 = 14.$$

Чем меньше величина остаточной дисперсии, тем лучше уравнение регрессии подходит к исходным данным. Если остаточная дисперсия оказывается примерно одинаковой для нескольких функций, то на практике предпочтение отдается более простым видам функций, так как они в большей степени поддаются интерпретации и требуют меньшего объема наблюдений.

Результаты многих исследований применения метода OSL подтверждают, что число наблюдений должно в 6-7 раз превышать число рассчитываемых параметров при переменной x. Это означает, что искать линейную регрессию, имея менее всего лишь 6 наблюдений, бессмысленно. Для других видов функциональных зависимостей каждый параметр должен рассчитываться хотя бы по 7 наблюдениям.

4.3. Метод выравнивания нелинейных зависимостей

Много реальные процессы и явления предполагают нелинейный характер зависимости между переменными. Чаще всего функции регрессии, используемые при количественной оценке связей между переменными имеют вид:

$y_x = ax^2 + bx + c$ – квадратичная;

$y_x = bx^a$ – степенная;

$y_x = be^{ax}$ – экспоненциальная;

$y_x = ba^x$ – показательная;

$y_x = a\ln x + b$ – полулогарифмическая;

$y_x = b + \dfrac{a}{x}$; $\quad y_x = \dfrac{1}{ax+b}$ – гиперболические.

В случае полиномиального вида уравнение регрессии, например, когда уравнение имеет вид $y_x = ax^2 + bx + c$, то параметры a, b и c находят из системы линейных уравнений, тем же методом Крамера:

$$\begin{cases} \left(\sum\limits_{i=1}^{n} x_i^4\right)\cdot a + \left(\sum\limits_{i=1}^{n} x_i^3\right)\cdot b + \left(\sum\limits_{i=1}^{n} x_i^2\right)\cdot c = \left(\sum\limits_{i=1}^{n} x_i^2 y_i\right); \\ \left(\sum\limits_{i=1}^{n} x_i^3\right)\cdot a + \left(\sum\limits_{i=1}^{n} x_i^2\right)\cdot b + \left(\sum\limits_{i=1}^{n} x_i\right)\cdot c = \left(\sum\limits_{i=1}^{n} x_i y_i\right); \\ \left(\sum\limits_{i=1}^{n} x_i^2\right)\cdot a + \left(\sum\limits_{i=1}^{n} x_i\right)\cdot b + n\cdot c = \left(\sum\limits_{i=1}^{n} y_i\right). \end{cases} .$$

В более общих случаях нелинейности применяется аналогичная методика расчета коэффициентов уравнения, только нелинейные функции следует привести к линейному виду, применив метод выравнивания исходных данных.

Приведем пример линеаризации нелинейной зависимости вида $y = ab^x$.

Прологарифмируем обе части уравнения $\lg y = \lg a + x\lg b$ и введем новые переменные $Y = \lg y$, $A = \lg a$, $B = \lg b$. В итоге получим уравнение линейного вида $Y = A + Bx$, коэффициенты которого находим стандартным методом наименьших квадратов. После чего, выполнив обратную замену, получим искомую нелинейную модель.

Другие, вышеприведенные нелинейные функции так же легко приводятся к линейному виду путем введения новых переменных.

П р и м е р. Построим уравнение нелинейной регрессии заданного вида, применив метод выравнивания к закону гиперболического типа $y = \dfrac{1}{ax+b}$.

Аргумент, x_i	$x_1 = 1$	$x_2 = 2$	$x_3 = 3$	$x_4 = 4$	$x_5 = 5$
Значение, $y_э$	$y_1 = 2.3$	$y_2 = 1.8$	$y_3 = 3.8$	$y_4 = 5.3$	$y_5 = 4.3$

Решение. Введем новую переменную $Y = \frac{1}{y}$ и получим таблицу для расчета коэффициентов регрессионного уравнения по методу OSL:

x	1	2	3	4	5
y	2.3	1.8	3.8	5.3	4.3
$Y = 1/y$	0.4	0.5	0.3	0.2	0.2

$$\sum_1^5 x_i = 15; \quad \sum_1^5 Y_i = 1.6; \quad \sum_1^5 x^2{}_i = 55; \quad \sum_1^5 x_i Y_i = 4.1; \quad (\sum_1^5 x_i)^2 = 225.$$

Найдем коэффициент регрессии для выравненного закона $Y = A + Bx$.

$$a = \frac{5 \cdot 4.1 - 15 \cdot 1.6}{5 \cdot 55 - 225} = \frac{20.5 - 24}{50} = \frac{-3.5}{50} = -0.07;$$

$$b = \frac{55 \cdot 1.6 - 15 \cdot 4.1}{5 \cdot 55 - 225} = \frac{88 - 61.5}{50} = \frac{26.5}{50} = 0.53.$$

Тогда $Y = -0.07x + 0.53$. Вернемся к исходной переменной:

$$\frac{1}{y} = -0.07x + 0.53.$$

Окончательно уравнение нелинейной регрессии примет вид:

$$Y_p = \frac{1}{-0.07x + 0.53}.$$

Рассчитаем «модельные» значения, сумму квадратов отклонений и построим график нелинейной регрессии (Рис. 4.3.):

x	1	2	3	4	5
Y_p	2.2	2.6	3.1	4	5.5

$$\sum \Delta_i{}^2 = 0.01 + 0.64 + 0.49 + 1.69 + 1.11 = 4.27.$$

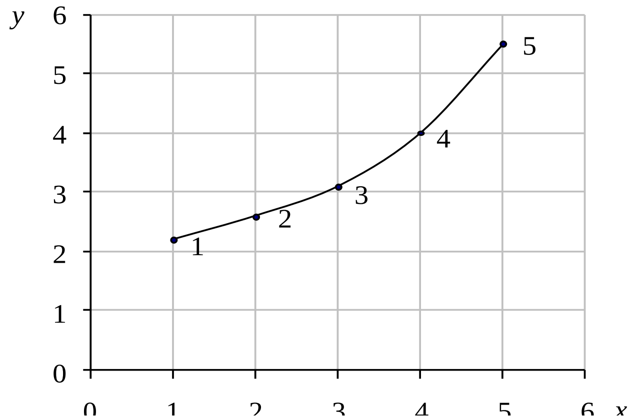


Рис. 4.3. График нелинейной регрессии

Для нелинейных видов функциональных зависимостей каждый параметр, как отмечалось, рассчитывается хотя бы по 6 наблюдениям. Следовательно, для параболы второй степени $y = ax^2 + bx + c$ необходимы не менее 12 данных, а для параболы третьей степени $y = ax^3 + bx^2 + cx + d$ – не менее 18 наблюдений.

4.4. Элементы теории корреляции

Интуитивно ясно, что о взаимозависимости между парой переменных можно говорить в тех случаях, когда уменьшению (увеличению) одной из них будет соответствовать уменьшение (увеличение) другой либо уменьшению (увеличению) первой будет соответствовать увеличение (уменьшение) второй переменной. В первом случае можно говорить о положительной корреляции между переменными (прямая зависимость), во втором – об отрицательной корреляции (обратная зависимость). Если рассмотреть разброс значений переменных относительно их средних, получим положительные и отрицательные разности, и знак их будет также различен.

При однонаправленных изменениях обеих переменных произведение их отклонений положительно, если же изменения переменных разнонаправлены, то произведение отрицательно.

Величина, полученная как отношение суммы произведений отклонений длине выборке без единицы, называется ковариацией. Ковариация вычисляется по формуле:

$$S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}.$$

Признаки x и y, по которым рассчитывается взаимосвязь, могут измеряться в разных единицах, иметь произвольные средние и дисперсии. Поэтому, вычитание соответствующих средних по каждой переменной делает ковариацию независимой от средних. Если разделить ковариацию на произведение стандартных отклонений, получим безразмерный коэффициент связи, который называется коэффициентом корреляции. *Коэффициент корреляции* (*Пирсона*) представляет собой численную меру степени взаимосвязи двух переменных, введенную в статистическую практику К. Пирсоном.

Коэффициент корреляции вычисляется по формуле:

$$r_{xy} = \frac{S_{xy}}{S_x S_y}.$$

Очевидно, что $S_{xy} = S_{yx}$, *следовательно*, $r_{xy} = r_{yx}$, поэтому с помощью коэффициента корреляции можно численно оценить величину и направленность взаимосвязи. Имеются разные модификации формулы линейного коэффициента корреляции, например:

$$r_{xy} = a\frac{\sigma_x}{\sigma_y} = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{\overline{yx} - \bar{y}\cdot\bar{x}}{\sigma_x \sigma_y},$$

где $\overline{xy} = \frac{\sum_{i=1}^{n} x_i\, y_i}{n}$; $\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - (\overline{x})^2}$; $\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} y_i^2}{n} - (\overline{y})^2}$;

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n};\ \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

По поводу интерпретации коэффициента корреляции сделаем два существенных замечания. При выводе расчетной формулы r_{xy} не делалось предположений о характере совместного двумерного распределения величин x и y. Очевиден вывод о пределах его изменения в диапазоне от –1 до +1. Выражения «сильная связь», «слабая связь», «умеренная связь» и т. д. справедливы только в рамках определенной статистической модели. Так, если частотные распределения величин x и y имеют разные значения асимметрии, то есть существенно скошены в разных направлениях, то даже при максимально возможной линейной связи между x и y величина коэффициента корреляции не будет по абсолютной величине превышать значения 0.6-0.7. Эта зависимость максимальной величины коэффициента корреляции от характера распределения x и y приводит к трудностям интерпретации получаемых его конкретных значений. Что означает $r_{xy} = 0.6$? Максимально возможную линейную связь при положительной и отрицательной асимметрии распределений x и y или умеренную связь этих переменных при совместном распределении, подобном двумерному нормальному распределению? Ответы на эти вопросы можно получить из качественного анализа диаграмм рассеяния и гистограмм распределения.

Второе замечание связано со значением коэффициентов, близких к нулю. Равенство нулю коэффициентов корреляции между переменными не всегда свидетельствует об отсутствии статистической связи между x и y. Так может проявляться, например, их нелинейная связь. Возможные варианты проявлений ложной корреляции могут быть еще связаны с появлением в совокупности исходных данных аномальных значений, или за счет неоднородности анализируемого материала, или за счет ошибок при регистрации данных.

Обобщим сказанное. Коэффициент корреляции находится в границах $-1 \le r_{xy} \le 1$. Если величина $r_{xy} \ge 0$, то зависимость между X и Y такова, что возрастание значений одной из переменных приводит к увеличению значений другой переменной. При значениях $r_{xy} < 0$ увеличение одной переменной приводит к уменьшению другой. Чем ближе значение r_{xy} к (± 1), тем сильнее (теснее) переменные X и Y связаны линейной функциональной зависимостью. Если $r_{xy} = \pm 1$, то линии регрессии $\overline{y}_x$ и $\overline{x}_y$ сливаются в одну, что соответствует строгой линейной зависимости между X и Y, в этом случае все наблюдаемые значения располагаются на общей прямой. При $r_{xy} = -1$ имеет место отрицательная линейная зависимость, при $r_{xy} = 1$ – положительная. Если $r_{xy} = 0$, то признаки X и Y некоррелированы. В этом случае линии регрессии $\overline{y}_x$ и $\overline{x}_y$ параллельны координатным осям.

Важно заметить, что величина линейного коэффициента корреляции оценивает тесноту связи рассматриваемых признаков в ее линейной форме. Поэтому близость абсолютной величины линейного коэффициента корреляции к нулю еще не означает, что X и Y независимы, если допускается отклонение этой зависимости от линейной. Следовательно, прямое утверждение $r_{xy} = 0$ не означает независимости исследуемой пары признаков.

В то же время, если X и Y независимы, то верно утверждение $r_{xy} = 0$. Таким образом, при отклонении парной статистической зависимости от линейной коэффициент корреляции теряет свой смысл как характеристика степени тесноты связи. В этом случае следует воспользоваться другим измерителем связи – корреляционным отношением.

Допустим, что выборочный коэффициент корреляции (найденный по выборке) оказался отличным от нуля. Так как выборка случайна, то еще нельзя заключить, что коэффициент корреляции генеральной совокупности также отличен от нуля. В этом случае проверяют гипотезу о значимости (существенности) выборочного коэффициента корреляции. Другая формулировка нулевой гипотезы: коэффициент корреляции генеральной совокупности r_{xy} равен нулю. Для оценки коэффициента корреляции r_{xy} нормально распределенной генеральной совокупности можно воспользоваться формулой:

$$r_{xy} - 3\cdot\frac{1-r_{xy}^2}{\sqrt{n}} \le r_{г.с.} \le r_{xy} + 3\cdot\frac{1+r_{xy}^2}{\sqrt{n}} \text{ (при } n \ge 50).$$

Выполним расчет коэффициента корреляции по формуле:

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n\sum x_i^2 - \left(\sum x_i\right)^2]\cdot[n\sum y_i^2 - \left(\sum y_i\right)^2]}} =$$

$$= \frac{6\cdot 53.2 - 26.7\cdot 12}{\sqrt{(6\cdot 159.09 - 712.89)\cdot(6\cdot 38 - 144)}} = \frac{-1.2}{142.47} = -0.00842.$$

Полученные результаты значений и коэффициента a в уравнении регрессии, и коэффициента корреляции r_{xy} согласуются с содержательным анализом условия задачи и подтверждают наличие обратной связи между изучаемыми величинами функции $M_n = f(R_б)$.

Для изучения степени неравномерности распределения определенного суммарного показателя между единицами отдельных групп вариационного ряда часто используют *кривую Лоренца* (или кривая концентрации). Кривая Лоренца была предложена американским экономистом Максом Отто Лоренцем в 1905 году как показатель неравенства в доходах населения. В прямоугольной системе координат кривая Лоренца является выпуклой вниз и проходит под диагональю единичного квадрата, расположенного в I координатной четверти. Каждая точка на кривой Лоренца соответствует утверждению вида: «20 самых бедных процентов населения получают всего 7% дохода». В случае равномерного распределения каждая группа населения имеет доход, пропорциональный своей численности. Такой случай описывается кривой равенства, являющейся прямой, соединяющей начало координат и точку (1; 1). В случае полного неравенства (когда лишь один член общества имеет доход) кривая сначала «прилипает» к оси абсцисс, а потом из точки (1; 0) «взмывает» к точке (1; 1).

П р и м е р. Пусть имеется распределение городов по числу жителей и распределение информационных потоков (Интернет-трафик) в этих городах (графы 1, 2, 3). Построить кривую Лоренца.

Города с числом жителей (тыс. чел.)	Число городов, в % к итогу W_i	Интернет-трафик % к итогу Y_i	Кумулятивные итоги	
			% городов $cumW_i$	% Интернет трафика $cumY_i$
До 3	4.2	0.2	4.2	0.2
3-5	4.6	0.2	8.8	0.5
5-10	13.1	1.7	21.9	2.2
10-20	28.3	6.8	50.2	9.0
20-50	28.7	14.8	78.9	23.8
50-100	9.7	10.3	88.6	34.1
100-500	9.7	33.8	98.3	67.9
Свыше 500	1.7	32.1	100	100
Итого	100.0	100.0	–	–

Р е ш е н и е. На координатной плоскости (Рис. 4.4.) наносим точки ($cumW_i$; $cumY_i$) и по ним строим кривую.

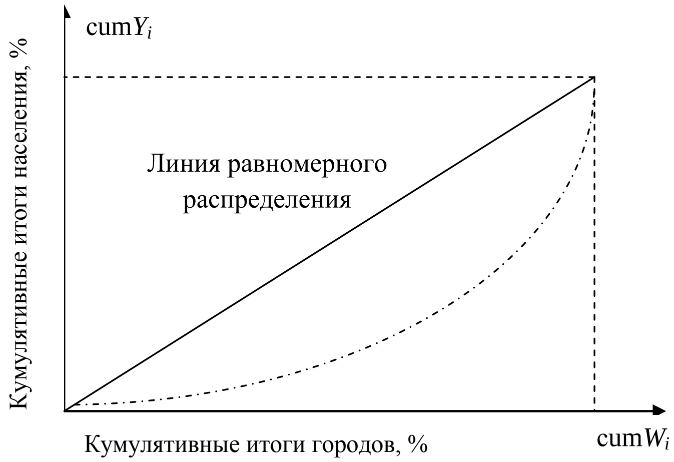


Рис. 4.4. Кривая Лоренца

Замечание, чем больше вогнутость (отличие кривой от линии равномерного распределения), тем выше концентрация (информационных потоков) в определенных группах единиц (крупных городах).

Кривые Лоренца применяют для распределений не только доходов, но и имущества домохозяйств, долей рынка для фирм в отрасли, природных ресурсов по государствам. *Концентрация* в производственной сфере представляет собой процесс сосредоточения изготовления продукции на ограниченном числе предприятий и в их производственных подразделениях. Материально-технической основой концентрации выступает научно-технический прогресс. Современное производство позволяет наиболее высоко использовать преимущества концентрации в интересах всего общества. Социальное значение концентрации состоит в том, что повышение технического предела производства создает климат для ликвидации тяжелого ручного труда, ускорения процесса ликвидации разницы между работниками умственного и физического труда.

Как форма организации производства концентрация имеет большую экономическую эффективность. Крупные предприятия обеспечивают, как правило, более высокий уровень производительности труда, лучшее качество продукции, более низкую себестоимость. Концентрация неразрывно связана со специализацией производства и комбинированием. Развитие специализированного производства выступает как прогрессивная форма концентрации однородного производства. Комбинирование, осуществляемое на крупных предприятиях, позволяет организовать производство на более высоком научно-техническом уровне. Концентрация производства может осуществляться путем увеличения размеров действующих предприятий, за счет их расширения, реконструкции или технического перевооружения; создания новых крупных предприятий и образования производственных объединений, что ведет к возникновению укрупненных производственных ячеек на базе централизации производства. Так как результатом централизации является расширение размеров предприятий, то ее также принято рассматривать как одну из форм осуществления концентрации. Уровень концентрации определяется системой показателей, основным из которых является объем производимой продукции.

В условиях научно-технического прогресса усиливается значение таких показателей концентрации, как стоимость основных производственных фондов, в том числе активного назначения, мощность энергоустановок и т. п. Для характеристики уровня концентрации применяется, и особенно широко в международных сопоставлениях, показатель численности работников, хотя этот показатель не всегда точно отражает динамику уровня концентрации, поскольку численность работников может сократиться при укрупнении производства в результате механизации и автоматизации трудовых процессов.

Вопросы самоконтроля:

1. Понятие регрессионной модели. Зависимые и независимые переменные.
2. Цели регрессионного анализа.
3. Линейная и нелинейная регрессия.
4. Диаграмма рассеяния. Гипотеза о виде функции регрессии.
5. Регрессионный анализ в случае нескольких независимых переменных.
6. От чего зависит достоверность полученных оценок?
7. Теоретические основы метода наименьших квадратов (МНК).
8. Смысл коэффициентов уравнения линейной регрессии.
9. Критерий оценки построенного уравнения регрессии.
10. Чем определяется необходимое число наблюдений для применения метода МНК.
11. Нелинейные зависимости. Полиномиальные уравнения.
12. Метод выравнивания для нелинейностей произвольного вила.
13. Понятие корреляции между переменными.
14. Ковариация. Коэффициент корреляции.
15. Интерпретация коэффициента корреляции.
16. Какие значения может принимать коэффициент корреляции. В чем их смысл?
17. Кривая Лоренца.
18. Оценка концентрации изучаемого признака по кривой Лоренца.

ЗАКЛЮЧЕНИЕ

В основе исследования сложных систем с использованием математического моделирования лежит системный подход, конечной целью которого является системное проектирование, направленное на построение системы с заданным качеством. В свою очередь системное проектирование базируется на результатах системного анализа, позволяющего выявить причинно-следственные связи между параметрами и характеристиками исследуемой системы и реализуемого с использованием математических моделей. При исследовании и проектировании сложных систем, наиболее эффективным оказывается использование комбинированного подхода, предполагающего совместное применение аналитических и численных методов. Аналитическими методами решаются задачи, связанные с формированием требований к структурным и функциональным параметрам, обеспечивающим заданное качество функционирования системы, однако получаемые при этом результаты могут иметь существенную погрешность. В процессе проверки адекватности модели необходимо определить область ее применения, то есть оценить диапазон изменения параметров, при котором точность результатов моделирования находится в допустимых пределах.

Численные и имитационные методы, основанные на использовании специализированных языков моделирования, например, GPSS, позволяют выполнять исследование систем с любой степенью детализации и должны применяться на заключительном этапе анализа изучаемой системы.

Синтез оптимальной системы направлен на построение системы, наилучшим образом соответствующей своему назначению. Решение задачи синтеза связано с определением зависимостей характеристик функционирования системы от параметров, которые представляются сложными математическими конструкциями. При этом возможность получения приемлемых результатов в процессе решения задач синтеза из-за их сложности и большой трудоемкости с учетом специфических особенностей реальных систем превосходит возможности математических методов оптимизации, и задача синтеза в общем виде оказывается математически неразрешимой. Для того, чтобы снизить сложность задачи синтеза, процесс проектирования разделяют на последовательность этапов, на каждом из которых решаются частные задачи синтеза – определяются параметры, связанные с отдельными аспектами структурно-функциональной организации системы, с использованием тех или иных моделей.

Исследования на моделях заключаются в проведении экспериментов, которые должны тщательно планироваться. Планирование, не в последнюю очередь, направлено на уменьшение количества и длительности экспериментов при условии обеспечения достоверности и полноты результатов моделирования. В случае большой размерности исследуемой системы, количество экспериментов и соответственно время, затраченное на моделирование, могут оказаться настолько большими, что полученные в конечном счете результаты потеряют свою актуальность. Особую значимость при планировании экспериментов приобретает использовании методов имитационного моделирования, характеризующихся большими компьютерными ресурсами. С увеличением числа проводимых экспериментов соответственно возрастает время моделирования. Все это обусловливает высокую стоимость компьютерного моделирования и требует тщательной подготовки эксперимента еще на этапе построения математической модели.

Приобретенные знания по мере изучения, представленного в учебном пособии материала, необходимы при изучении дисциплин, ориентированных на анализ данных и оценку качества управления системами, выполнению заданий в процессе научно-исследовательской и учебной практики, подготовки магистерской диссертации.

ЛИТЕРАТУРА

1. Аоки М. Введение в методы оптимизации.– М.: Наука, 1977.
2. Банди Б. Методы оптимизации. Вводный курс. Пер. с англ. М.: Радио и связь, 1988.
3. Батищев Д.И. Принятие оптимальных решений в экономических исследованиях. Горький: Горьковский госуниверситет, 1982.
4. Будак Б.М., Васильев Ф.П. Приближенные методы решения задач оптимального управления (тексты лекций). М.:МГУ, 1969. Вып. 2.
5. Васильев Ф.П. Численные методы решения экстремальных задач. М.:Наука. 1980.
6. Введение в математическое моделирование: Учебное пособие / Под ред. П.В. Трусова. М.: Логос, 2005.
7. Вентцель Е.С. Исследование операций. М.: Высшая школа, 2001.
8. Гилл Ф., Миррей У., Райт М. Практическая оптимизация. М.: Мир, 1985.
9. Гермейер Р. Введение в теорию исследования операций. М.: Наука, 1971.
10. Заварыкин В.М., Житомирский В.Г. Численные методы. М.: Просвещение, 1990.
11. Зарубин В.С. Математическое моделирование в технике. М.: Изд. МГТУ им. Н.Э. Баумана, 2003.
12. Иоффе А.Д., Тихомиров В.М. Теория экстремальных задач. М.: Наука, 1974.
13. Карманов В.Г. Математическое программирование. М.: Наука, 1980.
14. Красс М.С., Чупрынов Б.П. Математические методы и модели для магистрантов экономики: Питер, 2010.

15. Кудрявцев Л.Д. Математический анализ. Т. 1, 2. М.: Высшая школа, 1970.

16. Кузнецов Ю.Н., Кузубов В.И., Волощенко А.Б. Математическое программирование. М.: Высшая школа, 1976.

17. Лесин В.В., Лисовец Ю.П. Основы методов оптимизации. М.: Изд. МАИ, 1995.

18. Математические методы и модели исследования операций / Под ред. Колемаева В.А.. Изд.: Юнити-Дана, 2007, 592 с.

19. Моисеев Н.Н., Иваншов Ю.П., Столярова Е.М. Методы оптимизации. М.: Наука, 1988.

20. Муртаф Б. Современное линейное программирование. М.: Мир, 1984.

21. Ногин В.Д., Протодьяконов И.О. Основы теории оптимизации: Учебное пособие для студентов втузов. М.: Высшая школа, 1986.

22. Пантелеев А.В., Летова Т.А. Методы оптимизации в примерах и задачах. М.: Высшая школа, 2002.

23. Поляк Б.Т. Введение в оптимизацию. М.: Наука, 1983.

24. Розен В.В. Математические модели принятия решений в экономике. М: Высшая школа, 2002.

25. Советов Б.Я., Яковлев С.А. Моделирование систем: Учебник для вузов. 4-е изд., стер. М.: Высшая школа, 2005.

26. Сухарев А.Г. Оптимальный поиск экстремума. М.: МГУ, 1975.

27. Уайльд Д. Методы поиска экстремума, М.: Наука, 1967.

28. Хемминг Р.В. Численные методы. М.: Наука, 1968.

29. Юдин Д.Б., Гольдштейн Е.Г. Задачи и методы линейного программирования. М.: Сов. Радио, 1961.

30. Юрьева А.А. Математическое программирование: учеб. пособие / СПб.: Лань, 2014.