МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Белгородский государственный технологический университет
имени В. Г. Шухова

А. Г. Брусенцев, О. В. Осипов

МЕТОДЫ ОПТИМИЗАЦИИ

Учебное пособие

Белгород
2018

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Белгородский государственный технологический университет имени В. Г. Шухова

А. Г. Брусенцев, О. В. Осипов

МЕТОДЫ ОПТИМИЗАЦИИ

Утверждено учёным советом университета в качестве учебного
пособия для студентов направлений подготовки
09.04.01 - Информатика и вычислительная техника,
09.04.04 - Программная инженерия

Белгород
2018

УДК 519.8
ББК 22.1
Б89

Рецензенты:

Доктор технических наук, профессор Белгородского государственного технологического университета им. В. Г. Шухова ***Г. М. Редькин*** Кандидат физико-математических наук, доцент Белгородского государственного национального исследовательского университета (НИУ «БелГУ») ***В. В. Флоринский***

Брусенцев А. Г

Б89 Методы оптимизации: учебное пособие / А. Г. Брусенцев, О. В. Осипов. – Белгород: Изд-во БГТУ, 2018. – 263 с.

В пособии рассматриваются методы решения задач оптимизации. Изложены основы линейного программирования, теории двойственности и теории игр. Представлены различные методы дискретного, нелинейного и динамического программирования. Рассматриваются методы бесконечномерной оптимизации: классические и прямые методы вариационного исчисления. Изложение иллюстрируется примерами. В каждой главе приведены контрольные вопросы и задачи для самостоятельной работы.

Издание будет способствовать полноценному усвоению основных методов и алгоритмов оптимизации.

Учебное пособие предназначено для студентов высших учебных заведений, обучающихся по направлениям 09.04.01 – Информатика и вычислительная техника и 09.04.04 – Программная инженерия, а также может быть использовано и студентами других инженерно-экономических специальностей, изучающих дисциплину «Методы оптимизации».

Публикуется в авторской редакции.

УДК 519.8
ББК 22.1

Содержание

ПРЕДИСЛОВИЕ

Настоящая книга является учебным пособием по дисциплине «Методы оптимизации» и возникла в процессе чтения лекций магистрам специальностей «Информатика и вычислительная техника» и «Программная инженерия» Белгородского технологического университета им. В.Г. Шухова. Термин оптимизация можно понимать очень широко, включая в него методику всевозможных компьютерных переборов, лишенных какого-либо математического содержания. Предметом же настоящей книги являются математические методы оптимизации, т. е. математические методы решения экстремальных задач. Первые формулировки таких задач возникли в глубокой древности, а современные методы решения большинства из них стали известны сравнительно недавно. Возникновение и бурное развитие вычислительной техники привело к значительному росту популярности и практической значимости методов оптимизации. В настоящее время любое достаточно серьезное технологическое исследование содержит в себе ту или иную оптимизацию. Поэтому хорошо подготовленный специалист в области информационных технологий должен иметь представление о правильной постановке экстремальных задач и об эффективном выборе алгоритмов их решения. Авторы стремились к популярности изложения, которая, однако, не препятствовала бы полноценному усвоению основных методов и алгоритмов, а также всего курса в целом. Разумеется, охватить все содержание дисциплины в таком пособии невозможно. Здесь представлены лишь основы большинства разделов. Ряд вопросов излагается в обзорном порядке, и даются литературные ссылки для подробного углубленного изучения. Каждая глава снабжена контрольными вопросами и задачами для самостоятельного решения, которые позволяют проверить качество усвоения материала.

В книге затрагивается широкий круг вопросов, освещавшийся в большом количестве источников. Список литературы, приведенный в конце книги, содержит лишь те публикации, которые были использованы в ней или близко примыкают к ней, дополняя ее содержание.

Для чтения книги необходимы знания математического анализа, линейной алгебры, аналитической геометрии и теории вероятностей в объеме университетского курса бакалавриата. При рассмотрении вопросов бесконечномерной оптимизации необходимы также некоторые сведения из функционального анализа, которые, однако, кратко изложены в соответствующих местах настоящего пособия.

ВВЕДЕНИЕ

Под оптимизацией понимается выбор наилучшего варианта из некоторого множества альтернатив. Постоянное стремление осуществить такой выбор характерно для всех эпох развития человечества. В повседневной жизни задачи оптимизации, как правило, не требуют особых научных методов. Для их решения порой достаточно здравого смысла и накопленного ранее опыта. Однако в более сложных случаях не обойтись без расчетов, использующих математические модели исследуемых объектов. Под математической моделью явления или процесса понимают совокупность формул, уравнений, неравенств и т.д., отражающую существенные черты этого явления или процесса. При создании математических моделей для оптимизации первоочередной задачей является определение параметров, однозначно описывающих исследуемую ситуацию. Эти параметры обычно подразделяют на контролируемые, неконтролируемые и целевые. Контролируемые параметры x1, x2, ... , xn являются переменными, принимающими чаще всего числовые значения, причем исследователь может придавать им эти значения по своему усмотрению. Неконтролируемые параметры нельзя менять по своему усмотрению; более того, их значения во многих случаях исследователю неизвестны. Они могут быть случайными величинами с известными или неизвестными вероятностными характеристиками. Наконец, целевые параметры характеризуют эффективность альтернатив. Математическая модель должна связывать контролируемые и неконтролируемые параметры с целевыми параметрами. Задача нахождения по контролируемым и неконтролируемым параметрам целевых параметров называется прямой задачей оценки альтернатив. Наряду с прямой задачей часто решают обратную задачу, в которой требуется определить такие значения контролируемых параметров, при которых целевые параметры удовлетворяют тем или иным условиям оптимальности. Такие задачи часто еще называют задачами оптимизации. В зависимости от разновидности и сложности математической модели применяют те или иные методы решения обратной задачи, то есть те или иные методы оптимизации. Используемые в настоящее время математические модели подразделяются на детерминированные и стохастические. В детерминированных моделях неконтролируемые параметры, как правило, отсутствуют, а по значениям контролируемых параметров целевые параметры определяются математической моделью однозначно. К детерминированным моделям относятся модели линейного и нелинейного программирования, модели оптимального управления и т.д. Методы оптимизации здесь разработаны достаточно хорошо. В настоящем пособии, в основном, рассмотрены методы оптимизации для детерминированных моделей. В стохастических моделях связь между контролируемыми и целевыми параметрами носит вероятностный характер. Помимо указанного разделения моделей существенную роль играет их разделение по количеству переменных (контролируемых параметров). Говорят о задачах одномерной, многомерной и даже бесконечномерной оптимизации. Первые две главы настоящего пособия посвящены методам решения задач линейного программирования. В главах 3 и 4 кратко излагается теория двойственности линейного программирования и ее приложения в теории игр. Глава 5 посвящена методам решения задач дискретного линейного программирования. Основам нелинейного программирования посвящена глава 6, в которой изложен классический подход к теории экстремумов и общие численные методы решения задач. Здесь рассмотрены также задачи выпуклого и квадратичного выпуклого программирования. Задачам динамического программирования посвящена глава 7. Все перечисленные главы освещают вопросы конечномерной оптимизации. Бесконечномерной оптимизации посвящена последняя восьмая глава. Здесь мы кратко описываем основы вариационного исчисления и методы решения вариационных задач. Особое внимание уделяется прямым методам вариационного исчисления. Отметим, что в главе 8 рассмотрены лишь некоторые классические общие прямые методы вариационного исчисления. Существует большое число прямых методов, разработанных для решения тех или иных частных задач. В добавлении в качестве примера изложен прямой метод решения задачи об оптимальном распределении источников тепла.

1. ОСНОВНЫЕ ПОНЯТИЯ И МЕТОДЫ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

В 1939 году Леонид Витальевич Канторович опубликовал работу «Математические методы организации и планирования производства», в которой сформулировал новый класс экстремальных задач с ограничениями и разработал эффективный метод их решения. Тем самым были заложены основы линейного программирования. Джордж Данциг разработал симплекс метод и считается «отцом линейного программирования» на западе. Слово программирование здесь означает составление оптимального плана (программы) производства.

1.1. Формулировка задач линейного программирования. Основные формы линейных моделей

Сущность линейного программирования состоит в нахождении точек наибольшего или наименьшего значения некоторой линейной функции при определенном наборе линейных ограничений, налагаемых на аргументы. Ограничения образуют *систему ограничений*, которая имеет, как правило, бесконечное множество решений. Каждая совокупность значений переменных (аргументов функции z), которые удовлетворяют системе ограничений, называется *допустимым планом* задачи линейного программирования. Функция z, максимум или минимум которой определяется, называется *целевой функцией задачи.* Допустимый план, на котором достигается максимум или минимум функции z, называется *оптимальным планом задачи.* Система ограничений, определяющая множество планов, диктуется условиями производства. Задача линейного программирования состоит в выборе из множества допустимых планов наиболее выгодного (оптимального).

Пример 1.1. Мастерская имеет два станка S_1, S_2 на которых последовательно обрабатывается два вида продукции P_1 и P_2. Станок S_1 обрабатывает единицу продукции P_1 за 1 час, а единицу продукции P_2 — за 2 часа. Станок S_2 затрачивает на единицу продукции P_1 2 часа, а на единицу продукции P_2 — 1 час. Станок S_1 может работать в сутки не более 10 часов, а станок S_2 — не более 8 часов. Стоимость единицы продукции P_1 составляет с 1 рублей, а стоимость единицы продукции P_2 — с 2 рублей. Требуется определить такой суточный план выпуска двух видов продукции P_1 и P_2 мастерской, чтобы выручка от реализации произведенной продукции была максимальна.

Обозначим через x_1 и x_2 количество продукции P_1 и P_2 соответственно, которые планируется произвести на станках S_1 и S_2. Стоимость произведенной продукции $z = c_1x_1 + c_2x_2$. Мы должны назначить x_1 и x_2 так, чтобы величина z была максимальной. Переменные x_1 и x_2 не могут принимать произвольных значений. Их значения ограничены условиями производства, а именно, тем, что станки могут работать ограниченное время. На изготовление продукции P_1 станок S_1 тратит $1 \cdot x_1$ часов, а на изготовление продукции $P_2 - 2 \cdot x_2$ часов. Поскольку время работы станка S_1 не превосходит 10 часов, то величины x_1 и x_2 должны удовлетворять неравенству $x_1 + 2 \cdot x \leq 10$. Аналогично можно получить неравенство для станка S_2: $2 \cdot x_1 + x_2 \leq 8$. Кроме того, величины x_1 и x_2 не могут быть отрицательными: $x_1 0$, $x_2 0$.

Таким образом, задача заключается в нахождении точек максимума функции z среди точек с координатами x_1 и x_2, которые удовлетворяют указанным неравенствам. Такие задачи кратко записываются следующим образом:

$$z = c_1x_1 + c_2x_2 \to max$$
$$\begin{cases} x_1 + 2x_2 \leq 10, \\ 2x_2 + x_1 \leq 8; \end{cases}$$

$$x_1 \geq 0; x_2 \geq 0.$$

Рассмотрим несколько экономико-математических моделей, допускающих формулировки в виде задач линейного программирования.

Задача об оптимальном составе смеси. Эта задача является одной из первых задач, решенных методами линейного программирования. Изложим ее как задачу определения оптимального рациона кормления скота. Пусть нам известно содержание необходимых для кормления животного питательных веществ в различных применяемых кормах. Известна также стоимость единицы каждого вида корма. Требуется составить рацион — набор и количество кормов — так, чтобы каждое питательное вещество содержалось в нем в необходимом количестве и чтобы суммарные расходы на этот рацион были минимальны. Введем следующие обозначения:

m — число различных питательных веществ (витаминов, белков, углеводов и т. д.);

n — число применяемых видов кормов (силос, сено и т. д.);

a_{ik} — количество единиц i-го питательного вещества, содержащегося в единице k-го вида кормов;

b_i — минимальная суточная потребность животного в i-ом питательном веществе;

c_k — стоимость единицы k-го вида корма;

x_k — количество единиц k-го вида корма, используемое в суточном рационе и подлежащее определению.

Оптимальным планом здесь являются такие числа x_1, x_2,..., x_n которые удовлетворяют следующим ограничениям:

$x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0$ (количество корма не может быть отрицательным),

$\sum\limits_{k=1}^{n} a_{ik}x_k \geq b_i,\ i = 1, 2, \ldots, m$ (общее количество i-го питательного вещества в рационе не может быть ниже заданного b_i), и минимизируют суммарную стоимость составляемого рациона

$$z = \sum_{k=1}^{n} c_k x_k \to \min$$

Отметим, что такая же математическая задача возникает при отыскании оптимального *состава шихты* в металлургическом производстве. Известно, что для получения легированной стали нужно использовать шихту определенного химического состава. В состав шихты входят весьма дорогостоящие ингредиенты. В то же время можно использовать и малоценные материалы (чугун, лом, отходы с определенным содержанием присадок). Возникает задача отыскания такого состава шихты, который содержал бы необходимые количества химических веществ и имел бы минимальную стоимость. Математическая модель этой задачи совпадает с сформулированной выше. Еще один пример возникновения такой модели дает задача об оптимальном *смешивании нефтепродуктов.* Бензины разных сортов получают путем смешивания нефтепродуктов, имеющих различные технические характеристики. Заданные показатели качества бензина должны выдерживаться весьма точно. С другой стороны, от того, какие нефтепродукты смешиваются и в каком количестве, зависит рентабельность производства. Требуется построить такой план смешивания, который обеспечивал бы максимальную рентабельность производства и позволял, в то же время, получать бензины заданных сортов.

Задача об оптимальном использовании ресурсов. Эта задача возникает при составлении планов выпуска продукции предприятием и имеет важное практическое значение. Пусть номенклатура выпускаемой продукции состоит из n наименований. Производство продукции использует m видов ресурсов. Обозначим через a_{ik} затраты i-го вида ресурсов ($i = 1, 2, \ldots, m$) на производство единицы продукции k-го вида ($k = 1, 2, \ldots, n$). Пусть b_i — полный объем ресурсов i-го вида ($i = 1, 2, \ldots, m$), а c_k — прибыль, получаемая предприятием при изготовлении и реализации единицы k-го вида продукции ($k = 1, 2, \dots, n$). Требуется составить такой план выпуска продукции, который был бы технологически осуществим по имеющимся ресурсам и, в то же время, приносил наибольшую общую прибыль предприятию. Математическая модель задачи состоит в следующем. Обозначим через $X = \{x_1, x_2, \dots, x_n\}$ номенклатурный план выпуска продукции. Величины x_k должны удовлетворять следующим условиям:

$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$ (количество производимой продукции не может быть отрицательным),

$\sum\limits_{k=1}^{n} a_{ik} x_k \geq b_i$, $i = 1, 2, \dots, m$ (технологические ограничения). Кроме того, общая прибыль от производства и реализации продукции должна быть максимальной.

$$z = \sum_{k=1}^{n} c_k x_k \to \max$$

Приведение примеров математических моделей задач планирования и управления можно продолжить. Ниже мы еще вернемся к некоторым из таких моделей. Сейчас обсудим форму возникающих задач независимо от их содержания. Все задачи были задачами на минимум или максимум линейной целевой функции с ограничениями неравенствами. Переменные подчинялись условиям неотрицательности.

В других ситуациях могут возникать задачи, в систему ограничений которых, кроме неравенств, могут входить и равенства. Поэтому в наиболее общей форме задачу линейного программирования формулируют следующим образом:

$$z = c_0 + c_1 x_1 + c_2 x_2 + \dots + c_n x_n \to \max(\min)$$
$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \, R_1 b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \, R_2 b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \, R_m b_m \end{cases}$$

где $R_j (j = 1, 2, \dots, m)$— один из знаков сравнения ($\leq, \geq, =$).

Одна и та же задача линейного программирования может быть сформулирована в различных эквивалентных формах. Наиболее важными формами задачи линейного программирования являются *каноническая и стандартная*. В *канонической форме* задача является задачей на максимум некоторой линейной функции z, а ее система ограничений состоит только из равенств (уравнений); при этом переменные задачи $x_1, x_2, \dots x_n$ являются неотрицательными:

$$z = c_0 + c_1 x_1 + c_2 x_2 + \dots + c_n x_n \to \max$$
$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases}$$
$$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0.$$

К канонической форме можно привести любую задачу линейного программирования. Если в исходной задаче некоторое ограничение (например, первое) было неравенством, то оно преобразуется в равенство введением в левую часть некоторой неотрицательной переменной. Действительно

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1.$$

Вводим переменную $x_{n+1} = b_1 - a_{11}x_1 - a_{12}x_2 - \dots - a_{1n}x_n$. Тогда неравенство запишется в виде

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + x_{n+1} = b_1$$

В каждое из неравенств вводится своя “уравнивающая” переменная, после чего система ограничений становится системой уравнений. При этом “уравнивающая” переменная автоматически будет неотрицательной. Если в исходной задаче некоторая переменная x_i не подчинена условию неотрицательности, то ее заменяют разностью неотрицательных переменных x_i' и x_i'' :

$$x_i = x_i'' - x_i';\ x_i', x_i'' \leq 0.$$

Наконец, если исходная задача была задачей на минимум, то введением новой целевой функции $z_1 = -z$ мы преобразуем нашу задачу на минимум функции z в задачу на максимум функции z_1.

Таким образом, всякую задачу линейного программирования можно сформулировать в канонической форме.

В *стандартной форме* задача линейного программирования является задачей на максимум линейной целевой функции. Система ограничений ее состоит из линейных неравенств типа «$\geq$». Все переменные задачи неотрицательны.

$$z = c_0 + c_1x_1 + c_2x_2 + \cdots + c_nx_n \to \max(\min)$$
$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n\, b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n\, b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n\, b_m \end{cases}$$
$$x_1 \leq 0, x_2 \leq 0, \dots, x_n \leq 0.$$

Всякую задачу линейного программирования можно сформулировать в стандартной форме. Преобразование задачи на минимум в задачу на максимум, а также обеспечение неотрицательности переменных производится, как и раньше. Всякое равенство в системе ограничений равносильно системе взаимно противоположных неравенств:

$$a_{i1}x_i + a_{i2}x_2 + \cdots + a_{in}x_n = b_i \leftrightarrow \begin{cases} a_{i1}x_i + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i \\ -a_{i1}x_i - a_{i2}x_2 - \cdots - a_{in}x_n \leq -b_i \end{cases}$$

Существуют и другие способы преобразования системы равенств в систему неравенств, то есть всякую задачу линейного программирования можно сформулировать в стандартной форме.

Рассмотренные выше записи задач линейного программирования в канонической и стандартной формах обычно называют скалярными записями. Иногда более удобной является векторно-матричная запись. Введем векторы

$$\vec{x} = \{x_1, x_2, \dots, x_n\},\ \ \vec{c} = \{c_1, c_2, \dots, c_n\},\ \ \vec{b} = \{b_1, b_2, \dots, b_n\}$$

и матрицу $A = (a_{ij})$ коэффициентов в системе ограничений. Тогда задачи в канонической и стандартной формах можно записать соответственно в виде

$$\begin{aligned} & z = c_0 + (\vec{c}, \vec{x}) \to \max, \\ & A\vec{x} = \vec{b}, \vec{x} \geq \vec{0} \end{aligned}$$

$$\begin{aligned} & z = c_0 + (\vec{c}, \vec{x}) \to \max, \\ & A\vec{x} \leq \vec{b}, \vec{x} \geq \vec{0} \end{aligned}$$

Здесь и ниже неравенство $\vec{a} \leq \vec{e}$ ($\vec{a} < \vec{e}$) между векторами одинаковой размерности означает, что выполнены неравенства между всеми соответствующими координатами этих векторов. При этом условия неотрицательности (положительности) переменных часто записывают просто в виде $\vec{x} \geq 0$ ($\vec{x} > 0$).

В заключение параграфа отметим, что произвольная задача линейного программирования может не иметь решения. Эта ситуация может возникнуть в двух случаях: 1) система ограничений несовместна (область допустимых планов пуста); 2) целевая функция задачи неограничена на области допустимых планов. В первом случае задачу называют *недопустимой*.

1.2. Геометрическое истолкование задачи в стандартной форме в случае двух переменных

Задача линейного программирования в стандартной форме с двумя перемененными имеет вид:

$$z = c_0 + c_1x_1 + c_2x_2 \to max;$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \le b_1, \\ a_{21}x_1 + a_{22}x_2 \le b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 \le b_m; \end{cases}$$

$$x_1, x_2 \ge 0.$$

Эти задачи допускают простое геометрическое истолкование. Рассмотрим вначале геометрическое истолкование системы ограничений задачи. Каждую совокупность значений переменных $x1, x2$ можно изобразить точкой на плоскости, если ввести систему координат и по одной оси откладывать $x1$, а по другой — $x2$. Выясним, что геометрически означает совокупность решений одного отдельно взятого неравенства

$$a_1x_1 + a_2x_2 \le b.$$

Рассмотрим прямую линию на плоскости с уравнением

$$a_1x_1 + a_2x_2 = b.$$

Эта прямая делит плоскость на две полуплоскости, в одной из которых справедливо наше неравенство, а в другой — противоположное неравенство. Чтобы проверить, какая из полуплоскостей состоит из решений нашего неравенства, следует взять точку из какой-либо полуплоскости и проверить, выполняется ли наше неравенство в этой точке. Множество решений отдельно взятого линейного неравенства представляет собой полуплоскость. Для системы из нескольких таких неравенств точки, координаты которых удовлетворяют всем неравенствам одновременно, должны находиться во всех соответствующих полуплоскостях, то есть принадлежать теоретико-множественному пересечению этих полуплоскостей. Множество точек на плоскости, удовлетворяющих системе ограничений, составляет, таким образом, некоторую выпуклую многоугольную область. Условия неотрицательности переменных $x1, x2 \ge 0$ приводят к тому, что эта область находится в первой координатной четверти.

Пример 1.2. Построим на плоскости множество решений системы неравенств (рис.1.1)

$$\begin{cases} x_1 + 2x_2 \le 10, \\ 2x_1 + x_2 \le 8 \end{cases} \qquad x_1 \ge 0; x_2 \ge 0.$$

Множество точек на плоскости, удовлетворяющих системе ограничений задачи, называют часто допустимым многоугольником. Эта область может быть неограниченной или вовсе пустой.

Целевая функция задачи геометрически изображается с помощью прямой уровня, то есть прямой, на которой

$$z = c_0 + c_1x_1 + c_2x_2$$

принимает постоянное значение. Если C — произвольная константа, то уравнение прямой уровня имеет вид $c_0 + c_1x_1 + c_2x_2 = C$

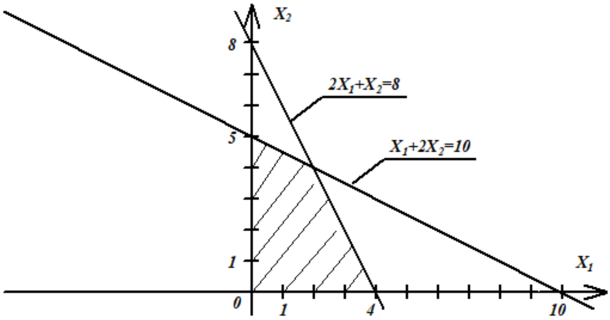

Рис. 1.1

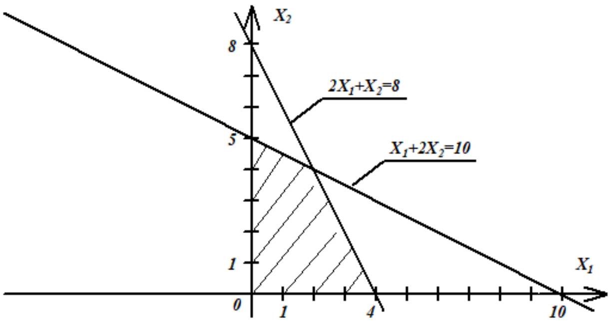

Рис. 1.2

При изменении константы C мы получаем различные прямые, параллельные друг другу. При увеличении C прямая уровня перемещается в направлении наискорейшего возрастания функции z, то есть, в направлении ее градиента. Вектор градиента

$$gradz = \left\{ \frac{\delta z}{\delta x_1}; \frac{\delta z}{\delta x_2} \right\} = \{c_1, c_2\}$$

Геометрический метод решения задачи состоит в следующем. Строится допустимый многоугольник и некоторое положение линии уровня целевой функции. Определяется направление перемещения прямой уровня. Точкой минимума z будет точка первого касания линии уровня с допустимым многоугольником. Точкой максимума — точка отрыва линии уровня от допустимого многоугольника. Эти точки чаще всего совпадают с некоторыми вершинами допустимого многоугольника. Таких точек может быть бесчисленное множество, если линия уровня z параллельна одной из сторон допустимого многоугольника.

Пример 1.3. Решим геометрически задачу примера 1.1 при $c_1 = 1, c_2 = 1$ (рис 1.2).

Точкой максимума здесь является точка A, координаты которой определяются из следующей системы уравнений:

$$\begin{cases} x_1 + 2x_2 = 10, \\ 2x_1 + x_2 = 8. \end{cases}$$

Решая эту систему, получаем точку максимума $A(2,4), z_{max} = 6$.

1.3. Общие системы линейных уравнений. Базисный вид системы. Метод Гаусса-Жордана

Ниже нам понадобятся некоторые сведения о системах уравнений, состоящих из m линейных уравнений с n неизвестными x_1,x_2,...,x_n

$$\begin{cases} a_{11}x_1 + a_{12}x_1 + \dots + a_{1n}x_1 = b_1 \\ a_{21}x_1 + a_{22}x_1 + \dots + a_{2n}x_1 = b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_1 + \dots + a_{mn}x_1 = b_m \end{cases}$$

Такая система может иметь множество решений, состоящее из единственного решения или бесконечного количества решений. Возможен также случай *несовместности системы*, когда множество решений является пустым. Две системы называютсяэ*квивалентными*, если они имеют одинаковые множества решений. Следующие операции над системой приводят к новой системе, эквивалентной исходной:

1) перестановка уравнений системы;
2) умножение обеих частей уравнения системы на одно и то же отличное от нуля число;
3) прибавление к заданному уравнению системы любого другого уравнения, умноженного на произвольное число.

Эти операции называют элементарными операциями над системой.

Основной задачей теории линейных систем является задача нахождения всего множества решений системы или установления ее несовместности. Для ее решения нужно привести систему к эквивалентной системе, имеющей особый *базисный вид*. При этом систему условно записывают в виде расширенной матрицы системы

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right)$$

Система называется имеющей базисный вид, если среди столбцов коэффициентов при неизвестных в ее расширенной матрице имеется столько различных единичных столбцов, сколько ненулевых строк в этой матрице. Единичным столбцом мы называем столбец, в котором на некотором месте стоит единица, а на всех остальных местах — нули. Единичные столбцы считаются различными, если единицы у них находятся на различных местах. Неизвестные, отвечающие этим различным единичным столбцам, называются *базисными неизвестными системы*. Остальные неизвестные называются *свободными неизвестными*. Базисные неизвестные входят по одному в каждое из уравнений и легко выражаются через свободные неизвестные. Если свободным неизвестным придавать произвольные значения, то базисные неизвестные по ним определятся однозначно, и мы получим все решения системы.

Пример 1.4. Рассмотрим систему

$$\begin{cases} x_1 + 3x_2 + 5x_3 = 6, \\ 7x_2 - 2x_3 + x_4 = -1. \end{cases}$$

Ее расширенная матрица

$$\left(\begin{array}{cccc|c} 1 & 3 & 5 & 0 & 6 \\ 0 & 7 & -2 & 1 & -1 \end{array}\right).$$

Система имеет базисный вид. Базисными неизвестными будут x_1 и x_4, а свободными — x_2 и x_3.

$$\begin{cases} x_1 = 6 - 3x_2 - 5x_3 \\ x_4 = -1 - 7x_2 + 2x_3. \end{cases}$$

Полагая, свободные неизвестные произвольными $x_2 = c_1$, $x_3 = c_2$, получаем двухпараметрическое семейство всех решений нашей системы $(6-3c_1-5c_2; c_1; c_2; -1-7c_1+2c_2)$.

Свободные неизвестные могут отсутствовать в базисном виде системы. Тогда, очевидно, система имеет только одно решение.

Метод Гаусса-Жордана представляет собой некоторый алгоритм, приводящий систему к базисному виду с помощью цепочки элементарных преобразований, которую удобно выполнять не над системой, а над ее расширенной матрицей. При этом элементарные операции над системой становятся следующими операциями над расширенной матрицей:

1) перестановка строк в матрице;
2) умножение всех элементов некоторой строки на одно и то же отличное от нуля число;
3) прибавление к данной строке любой другой, умноженной на произвольное число.

Метод Гаусса-Жордана состоит из ряда однотипных шагов. Опишем первый шаг алгоритма. Он состоит из трех этапов:

1) среди коэффициентов при неизвестных расширенной матрицы системы выбирается отличное от нуля число, которое в дальнейшем мы называем *разрешающим элементом* шага метода;
2) строка разрешающего элемента делится на разрешающий элемент и полученная строка, становясь основным инструментом для преобразования матрицы, называется нами в дальнейшем *ведущей строкой* шага алгоритма;
3) ведущая строка преобразует остальные строки матрицы путем прибавления ее к этим строкам после умножения на так подобранные числа, чтобы после преобразований в столбце бывшего разрешающего элемента стояли нули на всех местах, кроме места самого разрешающего элемента (на котором находится единица).

Описанные преобразования являются цепочкой элементарных операций над расширенной матрицей системы, и после завершения шага алгоритма метода Гаусса-Жордана в матрице появляется единичный столбец. Затем шаги повторяются, но на очередном шаге запрещается выбирать разрешающий элемент в строках, в которых он уже выбирался на предыдущих шагах. Шаги продолжаются до тех пор, пока количество единичных столбцов не сравняется с количеством ненулевых строк расширенной матрицы. Мы получаем систему в базисном виде.

При работе методом Гаусса-Жордана возможны следующие две особые ситуации. В результате выполнения очередного шага могут появиться либо нулевая строка $(0, 0, \dots, 0|0)$, либо строка вида $(0, 0, \dots, 0|b)$, где $b \neq 0$. В первом случае в новой системе будет уравнение вида $0 * x_1 + 0 * x_2 + \dots + 0 * x_n = 0$, которое является тождеством, справедливым при любых значениях неизвестных. Отбрасывание этого уравнения не меняет множества решений системы, поэтому обычно нулевая строка отбрасывается, и работа алгоритма продолжается. Во втором случае в новой системе появляется уравнение $0 * x_1 + 0 * x_2 + \dots + 0 * x_n \neq 0$, которое не может выполняться. Это свидетельствует о том, что новая и первоначальная системы несовместны. В этом случае работа алгоритма прекращается.

Пример 1.5. Реализуем метод Гаусса-Жордана для системы

$$\begin{cases} x_1 + 3x_2 + 5x_3 = 2, \\ 3x_1 - 2x_+ 16x_3 = -3. \end{cases}$$

В расширенной матрице системы выберем разрешающий элемент в первой строке и первом столбце. Получим сразу ведущую строку. Умножая первую строку на (-3) и прибавляя ко второй, получим

$$\left(\begin{array}{rrr|r} [1] & 3 & 5 & 2 \\ 3 & -2 & 16 & -3 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1 & 3 & 5 & 2 \\ 0 & -11 & 1 & -9 \end{array}\right).$$

Выбирая разрешающий элемент во второй строке и третьем столбце, умножая вторую строку на (-5) и прибавляя к первой, получим

$$\left(\begin{array}{rrr|r} 1 & 3 & 5 & 2 \\ 0 & -11 & [1] & -9 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1 & 58 & 0 & 47 \\ 0 & -11 & 1 & -9 \end{array}\right).$$

Последняя матрица соответствует системе уравнений

$$\begin{cases} x_1 + 58x_2 = 47, \\ -11x_2 + x_3 = -9. \end{cases}$$

с базисными неизвестными x_1, x_3 и свободной неизвестной x_2.

Отметим, что рассмотренный нами метод содержит большую долю произвола при выборе разрешающего элемента. Полученный в результате базисный вид системы тоже определяется неоднозначно. У совместной системы существует некоторая конечная совокупность базисных видов.

Переход от одного базисного вида к другому можно произвести с помощью, так называемой, *операции замещения*. Эта операция переводит заданную базисную неизвестную x_i в разряд свободных, а заданную свободную неизвестную x_j — в базисную. Операция замещения состоит в дополнительном шаге алгоритма Гаусса-Жордана с особым выбором разрешающего элемента. Этот элемент выбирается в строке, содержащей единицу при базисной неизвестной x_i и в столбце, отвечающем свободной неизвестной x_j. Выполним операцию замещения в базисном виде системы предыдущего примера, замещая свободной переменной x_2 базисную x_1

$$\left(\begin{array}{rrr|r} 1 & [58] & 0 & 47 \\ 0 & -11 & 1 & -9 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1/58 & [1] & 0 & 47/58 \\ 0 & -11 & 1 & -9 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1/58 & 1 & 0 & 47/58 \\ 11/58 & 0 & 1 & -5/58 \end{array}\right).$$

Тем самым получен новый базисный вид системы

$$\begin{cases} \frac{1}{58}x_1 + x_2 = \frac{47}{58}, \\ \frac{11}{58}x_1 + x_3 = -\frac{5}{58}. \end{cases}$$

В следующем параграфе при изучении симплекс-метода мы встретимся с базисным видом линейной системы уравнений и операцией замещения. При этом неизвестные будут называться переменными.

1.4. Симплекс-метод

Этот метод является универсальным, применимым к любой задаче линейного программирования в канонической форме. Система ограничений здесь — система линейных уравнений, в которой количество неизвестных обычно больше количества уравнений. Если ранг системы равен r, то мы можем выбрать r неизвестных, которые выразим через все остальные неизвестные. Для определенности предположим, что выбраны первые, идущие подряд, неизвестные $x_1, x_2, \ldots, x_r$. Тогда наша система уравнений может быть записана в виде

$$\begin{cases} x_1 = b'_1 - a'_{1,r+1}x_{r+1} - \cdots - a'_{1n}x_n, \\ x_2 = b'_2 - a'_{2,r+1}x_{r+1} - \cdots - a'_{2n}x_n, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ x_r = b'_r - a'_{r,r+1}x_{r+1} - \cdots - a'_{rn}x_n, \end{cases}$$

К такому виду можно привести любую совместную систему, например методом Гаусса-Жордана. Правда, не всегда можно выражать через остальные первые r неизвестных (мы это сделали для определенности записи). Однако такие r неизвестных обязательно найдутся. Эти неизвестные (переменные) называются *базисными*. Остальные переменные называются *свободными*. Придавая определенные значения свободным переменным и вычисляя значения базисных (выраженных через свободные), мы будем получать различные решения нашей системы ограничений. Таким образом, можно получить любое ее решение. Нас будут интересовать особые решения, которые получаются, когда свободные переменные равны нулю. Такие решения называются *базисными*. Базисных решений столько же, сколько различных базисных видов у данной системы ограничений. Базисное решение называется *допустимым базисным решением* или *опорным решением*, если в нем значения переменных неотрицательны. Если в качестве базисных взяты переменные $x_1,x_2,\dots,x_r$, то решение $\{b'_1,b'_2,\dots,b'_r,0,\dots,0\}$ будет опорным, если $b'_1 \le 0$, $b'_2 \le 0$, $\dots$, $b'_r \le 0$. Симплекс-метод основан на следующей теореме, которая называется *фундаментальной теоремой симплекс-метода*.

Теорема 1.1. Среди оптимальных планов задачи линейного программирования в канонической форме обязательно есть опорное решение ее системы ограничений. Если оптимальный план задачи единственен, то он совпадает с некоторым опорным решением.

Различных опорных решений системы ограничений конечное число. Поэтому решение задачи в канонической форме можно было бы искать перебором опорных решений и выбором среди них того решения, для которого значение z самое большое. Но, во-первых, все опорные решения неизвестны, и их нужно находить, а во-вторых, в реальных задачах этих решений очень много, и прямой перебор вряд ли возможен. Симплекс-метод представляет собой некоторую процедуру направленного перебора опорных решений. Исходя из некоторого, найденного заранее, опорного решения по определенному алгоритму симплекс-метода, мы подсчитываем новое опорное решение, на котором значение целевой функции z не меньше, чем на старом. После ряда шагов мы приходим к опорному решению, которое является оптимальным планом.

Процедура симплекс-метода на примере

Пусть требуется найти решение следующей задачи линейного программирования

$$z = -x_4 + x_5 \to max$$

$$\begin{cases} x_1 = 1 - x_4 + 2x_5, \\ x_2 = 2 + 2x_4 - x_5, \\ x_3 = 3 - 3x_4 - x_5; \end{cases}$$

$$x_1, x_2, \dots, x_5 \le 0.$$

Первое опорное решение имеет следующий вид B_1={1; 2; 3; 0; 0}. На этом базисном решении системы ограничений целевая функция $z(B_1) = 0$. Из вида целевой функции заключаем, что она может быть увеличена при выходе из решения B_1 путем увеличении переменной x_5 от нуля. При этом нужно следить, чтобы соблюдались равенства нашей системы и все переменные оставались неотрицательными. Из первого ограничения видно, что x1 останется неотрицательным при произвольном увеличении x5. Второе ограничение показывает, что x_2 становиться отрицательным при $x_5 > 2$. Из третьего ограничения заключаем, что x_3 остается неотрицательным при увеличении x_5 до трех. Таким образом, все переменные остаются неотрицательными при увеличении x_5 до 2. Предположим, что $x_5 = 2$, но, по-прежнему, $x_4 = 0$. Тогда остальные переменные примут значения $x_1 = 5$; $x_2 = 0$; $x_3 = 1$. Мы перешли к некоторому новому решению системы ограничений B_2={5; 0; 1; 0; 2}. Это решение будет опорным решением и соответствующий базисный вид можно получить с помощью операции замещения, при которой x2 выводится из числа базисных, а x_5 становится базисной переменной. Другими словами, нужно x_5 выразить из второго равенства системы ограничений и полученное выражение подставить вместо x_5 в первое и третье равенства. В результате получим базисный вид системы

$$\begin{cases} x_1 = 5 - 2x_2 + 3x_4, \\ x_5 = 2 - x_2 - 2x_4, \\ x_3 = 1 + x_2 - 5x_4; \end{cases}$$

которому отвечает базисное решение B_2={5; 0; 1; 0; 2}. С помощью нового вида системы исключим x_5 из целевой функции задачи

$$z = -x_4 + x_5 = 2 - x_2 + 2x_4 - x_4 = 2 - x_2 + x_4.$$

Значение целевой функции на новом базисном решении $z(B_2) = 2$. При этом целевую функцию можно еще увеличить, если выйти из B_2, увеличивая переменную x_4. В первом и во втором равенствах перестроенной системы ограничений x_4 можно увеличивать неограниченно. В третьем равенстве x_4 можно увеличивать лишь до 1/5. В противном случае переменная x_3 станет отрицательной. Положим $x_4 = 1/5$, $x_2 = 0$. Тогда $x_1 = 5 + 3/5 = 28/5$; $x_5 = 2 + 2/5 = 12/5$; $x_3 = 0$. Получаем следующее опорное решение $B_3 = \{28/5; 0; 0; 1/5; 12/5\}$. При этом переменная x_3 должна быть выведена из состава базисных переменных, а вместо нее базисной переменной должна стать переменная x4. Производя операцию замещения, получаем следующий базисный вид системы ограничений

$$\begin{cases} x_1 = (28/5) - (7/5)x_2 + (3/5)x_3, \\ x_5 = (12/5) - (3/5)x_2 - (2/5)x_3, \\ x_3 = (1/5) + (1/5)x_2 - (1/5)x_3; \end{cases}$$

с помощью которого можно исключить базисные переменные из выражения для целевой функции

$$z = 2\breve{\ } x_2 + x_4 = 2\breve{\ } x2 + ((1/5) + (1/5)x_2\breve{\ }(1/5)x_3) = (11/5)\breve{\ }(4/5)x_2\breve{\ }(1/5)x_3.$$

Из последнего выражения видно, что увеличить значение целевой функции переходом к новому опорному решению нельзя. Поэтому $z_{max} = z(B_3) = 11/5$. Оптимальный план совпадает с $B_3 = \{28/5; 0; 0; 1/5; 12/5\}$.

Процедура симплекс-метода в общем случае

В рассмотренном примере мы сделали два шага, переходя последовательно от базисного решения B_1 к B_2, а затем — к B_3. Вычисления по симплекс-методу обычно организуются в виде так называемых симплекс-таблиц. Чтобы разобраться в устройстве симплекс-таблицы рассмотрим один шаг симплекс-метода в общем случае. Предположим, что система ограничений задачи в канонической форме приведена к допустимому базисному виду и базисными переменными являются x_1, x_2, …, x_r. Целевая функция при этом выражена через свободные переменные x_{r+1}, …, x_n, то есть задача имеет вид

$$z = \gamma_0 + \gamma_{r+1}x_{r+1} + \cdots + \gamma_n x_n \to$$

$$\begin{cases} x_1 = b_1 - (a_{1r+1}x_{r+1} + \cdots + a_{1n}x_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_r = b_r - (a_{rr+1}x_{r+1} + \cdots + a_{rn}x_n, ; \end{cases}$$

$$x_i \geq 0, i = 1, 2, \ldots, n.$$

Поскольку базисный вид является допустимым, то b_1, b_2, ..., $b_r \leq 0$. Работа по симплекс-методу начинается с просмотра коэффициентов целевой функции, то есть величин γ_{r+1}, ..., γ_n. Здесь могут представиться два случая.

1) Все коэффициенты неположительные: γ_{r+1}, ..., γ_n. В этом случае исходное базисное решение будет оптимальным, так как переходом к другому базисному решению мы не можем увеличить целевую функцию.
2) Среди коэффициентов целевой функции есть положительные.

Пусть $\gamma_j > 0$. Это значит, что увеличение переменной x_j ведет к увеличению целевой функции. При этом увеличивать x_j нужно так, чтобы базисные переменные оставались неотрицательными. Для определения границы увеличения x_j просмотрим столбец коэффициентов при x_j в системе ограничений, то есть числа a_{1j}, a_{2j}, ..., a_{rj}. Если все эти коэффициенты отрицательны, то x_j можно увеличивать неограниченно, сохраняя неотрицательными базисные переменные. Это означает, что целевая функция неограниченна на области допустимых значений и задача не имеет решений. Если же среди этих коэффициентов есть положительные, то в соответствующих уравнениях x_j нельзя увеличивать неограниченно. Предположим, что мы придаем x_j значение $\rho(x_j = \rho)$, а остальные свободные переменные оставляем равными нулю. Система ограничений примет вид

$$\begin{cases} x_1 = b_1 - a_{1_j}\rho, \\ x_2 = b_2 - a_{2_j}\rho, \\ \ldots\ldots\ldots\ldots\ldots \\ x_r = b_r - a_{r_j}\rho, \end{cases}$$

Значение ρ не может быть произвольным в i-ом равенстве, если $a_{ij} > 0$. При этом ρ можно увеличивать в i-ом равенстве до величины b_i/a_{ij}. Таким образом, граница увеличения ρ может быть найдена выбором среди чисел a_{1j}, a_{2j}, ..., a_{rj} положительных, а затем для них минимального значения отношения b_i/a_{ij}. Это отношение и даст нам границу увеличения ρ. Если минимальным будет отношение b_{i_0}/a_{i_0j}, то граница увеличения ρ определяется i_0-ым равенством и можно положить $\rho = b_{i_0}/a_{i_0j}$. При этом $x_{i_0} = 0$. При перестройке системы ограничений переменная x_j должна быть введена в число базисных переменных, а переменная должна стать свободной. Перестройка производится с помощью операции замещения, которая выполняется с помощью разрешающего элемента a_{i_0j}. Этот элемент принято называть разрешающим или генеральным элементом шага симплекс-метода. После операции однократного замещения мы получим новый базисный вид системы ограничений, новое опорное решение и новое выражение для целевой функции через новые свободные переменные. На этом шаг симплекс-метода завершается. Следующий шаг начинается с просмотра коэффициентов нового выражения целевой функции. Шаги симплекс-метода повторяются до тех пор, пока на очередном шаге все коэффициенты выражения целевой функции через свободные переменные не станут неположительными. В этом случае очередное опорное решение системы ограничений будет точкой максимума.

Вычисления по симплекс-методу организуются в виде симплекс-таблиц, которые являются сокращенной записью задачи линейного программирования в канонической форме. *Перед составлением симплекс-таблицы задача должна быть преобразована. Система ограничений приводится к допустимому базисному виду, с помощью которого из целевой функции должны быть исключены базисные переменные.* Вопрос об этих преобразованиях мы рассмотрим ниже. Сейчас же будем считать, что они уже выполнены и задача имеет вид

$$z = \gamma_0 + \gamma_{r+1}x_{r+1} + \dots + \gamma_n x_n \to max;$$

$$\begin{cases} x_1 = b'_1 - a'_{1,r+1}x_{r+1} - \dots - a'_{1n}x_n, \\ x_2 = b'_2 - a'_{2,r+1}x_{r+1} - \dots - a'_{2n}x_n, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ x_r = b'_r - a'_{r,r+1}x_{r+1} - \dots - a'_{rn}x_n, \end{cases}$$

$$x_i \leq 0, i = 1, 2, \dots, n.$$

Здесь для определенности записи считается, что в качестве базисных можно взять переменные x_1,x_2,...,x_r и что при этом $b'_1 \leq 0$,$b'_2 \leq 0$,...,$b'_r \leq 0$. (соответствующее базисное решение является опорным).

Для составления симплекс-таблицы во всех равенствах в условии задачи члены, содержащие переменные, переносятся в левую часть, а свободные члены оставляются справа, т.е. задача записывается в виде следующей системы равенств:

$$\begin{cases} x_1 + a'_{1,r+1}x_{r+1} + \dots + a'_{1n}x_n = b'_1, \\ x_2 + a'_{2,r+1}x_{r+1} + \dots + a'_{2n}x_n = b'_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ x_r + a'_{r,r+1}x_{r+1} + \dots + a'_{rn}x_n = b'_r, \\ z - \gamma_{r+1}x_{r+1} - \dots - \gamma_n x_n = \gamma_0. \end{cases}$$

Далее эта система оформляется в виде таблицы:

Баз. пер.	Св. чл.	x_1	x_2	...	x_r	x_{r+1}	x_{r+2}	...	x_N
x_1	b_1	1	0	...	0	$a_{1,r+1}$	$a_{1,r+2}$	...	a_{1N}
x_2	b_2	0	1	...	0	$a_{2,r+1}$	$a_{2,r+2}$	...	a_{2N}
...	...	...	...	...	...	...	...	...	...
x_r	b_r	0	0	...	1	$a_{r,r+1}$	$a_{r,r+2}$	...	a_{rN}
z	γ_0	0	0	...	0	$-\gamma_{r+1}$	$-\gamma_{r+2}$	...	$-\gamma_N$

Еще раз напомним, что названия базисных переменных здесь взяты лишь для определенности записи и в реальной таблице могут оказаться другими.

Порядок работы с симплекс-таблицей

Первая симплекс-таблица подвергается преобразованию, суть которого заключается в переходе к новому опорному решению. Порядок перехода к следующей таблице такой.

1. Просматривается последняя строка таблицы и среди коэффициентов этой строки (исключая γ_o) выбирается отрицательное число. Если такового нет, то исходное базисное решение является оптимальным и данная таблица является последней.
2. Просматривается столбец таблицы, отвечающий выбранному отрицательному коэффициенту в последней строке, и в этом столбце выбираются положительные коэффициенты. Если таковых нет, то целевая функция неограниченна на области допустимых значений переменных, и задача решений не имеет.

3. Среди отобранных коэффициентов столбца выбирается тот, для которого отношение соответствующего свободного члена, находящегося в столбце свободных членов, к этому элементу, минимально. Этот коэффициент называется *разрешающим* или *генеральным элементом таблицы*. В дальнейшем базисная переменная, отвечающая строке разрешающего элемента, должна быть переведена в разряд свободных, а свободная переменная, отвечающая столбцу разрешающего элемента, вводится в число базисных.
4. Строится новая таблица, содержащая новые названия базисных переменных. Строка разрешающего элемента делится на этот элемент, и полученная строка записывается в новую таблицу на то же место. В остальные клетки новой таблицы записываются результаты преобразования элементов старой таблицы. Для этого умножают первую из заполненных строк (строку разрешающего элемента) на некоторые числа и складывают ее со строками старой таблицы. Числа эти подбираются так, чтобы в столбце разрешающего элемента получились нули, кроме клетки разрешающего элемента, в которой стоит единица. В результате получают новую симплекс-таблицу, которая отвечает новому базисному решению.
5. Теперь следует обратиться к пункту 2, т.е. просмотреть строку целевой функции и повторить все вышеперечисленное. Составление новых симплекс-таблиц производится до тех пор, пока все коэффициенты последней строки (кроме стоящего на месте γ_0) в очередной таблице не станут неотрицательными. После этого считается, что задача решена и по последней симплекс-таблице прочитывается ответ задачи. Максимальное значение z_{max} целевой функции стоит в первой клетке последней строки на месте y_o. Неотрицательные значения новых базисных переменных стоят в остальных клетках столбца свободных членов. Остальные переменные в точке максимума равны нулю.

Замечание 1.1. Замечание. Изложенный алгоритм содержит возможность неопределенности при выборе разрешающего элемента. Могут появиться несколько столбцов, пригодных для его выбора. Да и в заданном столбце может быть несколько чисел, каждое из которых можно назвать разрешающим элементом. Для однозначной организации вычислений приходится вводить добавочные условные правила. При выборе столбца разрешающего элемента в последней строке симплекс-таблицы выбирается максимальный по модулю отрицательный коэффициент. Если есть несколько таких коэффициентов с одинаковым максимальным модулем, выбирается тот, что отвечает переменной с минимальным номером и т.д. Отметим также, что если в столбце, пригодном для выбора разрешающего элемента, нет положительных чисел, то задача не имеет решений по причине неограниченности целевой функции на области допустимых планов.

Рассмотрим порядок решения задачи с помощью симплекс-таблиц на примерах.

Пример 1.6. Решить следующую задачу, уже рассмотренную в качестве примера:

$$z = x_5 - x_4 \to max;$$
$$\begin{cases} x_1 = 1 - x_4 + 2x_5, \\ x_2 = 2 + 2x_4 - x_5, \\ x_3 = 3 - 3x_4 - x_5; \end{cases}$$
$$x_i \leq 0, i = 1, \ldots, 5.$$

Запишем задачу в виде равенств:

$$\begin{cases} x_1 + x_4 - 2x_5 = 1, \\ x_2 - 2x_4 + x_5 = 2, \\ x_3 + 3x_4 + x_5 = 3, \\ z + x_4 - x_5 = 0. \end{cases}$$

Составляем первую симплекс-таблицу. Находим разрешающий элемент.

Базисные переменные	Свободные члены	x_1	x_2	x_3	x_4	$\downarrow x_5$
x_1	1	1	0	0	1	-2
$\leftarrow x_2$	2	0	1	0	-2	1
x_3	3	0	0	1	3	1
z	0	0	0	0	1	-1

Составляем новую симплекс-таблицу. Снова находим разрешающий элемент.

Базисные переменные	Свободные члены	x_1	x_2	x_3	$\downarrow x_4$	x_5
x_1	5	1	2	0	-3	0
x_2	2	0	1	0	-2	1
$\leftarrow x_3$	1	0	-1	1	5	0
z	2	0	1	0	-1	0

Переходим к следующей таблице: Эта таблица является последней, по ней читаем

Базисные переменные	Свободные члены	x_1	x_2	x_3	x_4	x_5
x_1	$\frac{28}{5}$	1	$\frac{7}{5}$	$\frac{3}{5}$	0	0
x_5	$\frac{12}{5}$	0	$\frac{3}{5}$	$\frac{2}{5}$	0	1
x_4	$\frac{1}{5}$	0	$\frac{-1}{5}$	$\frac{1}{5}$	1	0
z	$\frac{11}{5}$	0	$\frac{4}{5}$	$\frac{1}{5}$	0	0

ответ задачи: $z_{max} = \frac{11}{5}$ Координаты точки максимума: $x_1 = \frac{28}{5}$; $x_2 = 0$; $x_3 = 0$; $x_4 = \frac{1}{5}$; $x_5 = \frac{12}{5}$;

Пример 1.7. Решить задачу:

$$z = x_1 + x_2 \to max;$$

$$\begin{cases} x_3 = 1 + x_1 - x_2, \\ x_4 = 2 - x_1 + 2x_2; \end{cases}$$

$$x_i \leq 0 (i = 1, 2, 3, 4,).$$

Составляем первую симплекс-таблицу и находим разрешающий элемент.

Вторая таблица имеет вид:

Поскольку в последней таблице в столбце, пригодном для выбора разрешающего элемента, нет положительных чисел, то целевая функция неограниченна на области допустимых значений, то есть задача решения не имеет.

Базисные переменные	Свободные члены	$\downarrow x_1$	x_2	x_3	x_4
x_3	1	-1	1	1	0
$\leftarrow x_4$	2	1	-2	0	1
z	0	-1	-1	0	0

Базисные переменные	Свободные члены	x_1	x_2	x_3	x_4
x_3		0	-1	1	1
x_1	2	1	-2	0	1
z	2	0	-3	0	1

Зацикливание симплекс алгоритма

При работе симплекс методом мы переходим от одного опорного решения системы ограничений задачи к другому, причем так, чтобы значение целевой функции на следующем решении не уменьшалось. Поскольку опорных решений конечное число, мы должны прийти в оптимальное опорное решение, существование которого гарантируется фундаментальной теоремой. Однако мы можем не достигнуть оптимального решения, если в процессе перебора симплекс методом вернемся в опорное решение, которое уже встречалось ранее, и затем будем повторять цепочку опорных решений, пройденных ранее. Вычисления по симплекс методу войдут в бесконечно повторяющийся цикл и никогда не закончатся. Такое явление называется *зацикливанием симплекс алгоритма.*

Зацикливание — явление очень редкое. В литературе по линейному программированию имеются лишь несколько примеров задач, в которых возможно возникновение зацикливания.

Несмотря на возможность зацикливания, любую задачу линейного программирования надлежащей формы можно решить симплекс методом до конца. При рассмотрении симплекс алгоритма мы видели, что на очередном шаге разрешающий элемент может выбираться неоднозначно. Для однозначной организации вычислений приходится вводить добавочные правила. Можно показать, что зацикливание наступает лишь в случае возможности неоднозначного выбора разрешающего элемента. Если зацикливание наступило, следует изменить порядок вычислений, выбирая разрешающий элемент по-другому. Произойдет выход из цикла. Для борьбы с зацикливанием используют особые подпрограммы, гарантирующие выход из цикла в случае наступления зацикливания.

1.5. Метод искуственного базиса

Симплекс-метод применяется к задачам частного вида, у которых система ограничений имеет допустимый базисный вид. Непосредственное отыскание первого допустимого базисного вида системы ограничений обычно очень затруднительно. Существует два метода преодоления этой трудности. Первый из них называется *методом искусственного базиса*, а второй — *методом больших штрафов.*

Рассмотрим задачу в канонической форме

$$z = c_0 + c_1x_1 + c_2x_2 + ... + c_nx_n \to max,$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{1n}x_n = b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m; \end{cases}$$

$$x_i \geq 0, i = 1, 2, ..., n.$$

Непосредственно к такой задаче симплекс-метод не применим. Следует привести систему ограничений к допустимому базисному виду. Это можно сделать с помощью симплекс-метода. Не ограничивая общности, мы можем считать, что свободные члены $b_1, b_2, ..., b_m$ неотрицательны, так как в противном случае умножением некоторых уравнений на (-1) этого всегда можно добиться. Введем в нашу систему уравнений дополнительные переменные $y_1, y_2, ..., y_m$ следующим образом

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n + y_1 = b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{1n}x_n + y_2 = b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n + y_m = b_m. \end{cases}$$

Эта новая система ограничений имеет допустимый базисный вид с базисными переменными $y_1, y_2, ..., y_m$. Новая и исходная системы эквивалентны при условии, что $y_1, y_2, ..., y_m = 0$. Если путем эквивалентных преобразований переменные $y_1, y_2, ..., y_m$ вывести из числа базисных, заменяя их другими переменными, и в полученном базисном виде положить $y_1, y_2, ..., y_m = 0$, то мы получим базисный вид исходной системы. Преобразование новой системы к допустимому базисному виду, в котором $y_1, y_2, ..., y_m$ являются свободными переменными можно провести в процессе решения следующей вспомогательной задачи линейного программирования

$$f = -y_1 - y_2 - ... - y_m \to max,$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n + y_1 = b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{1n}x_n + y_2 = b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n + y_m = b_m. \end{cases}$$

$$x_i, y_j \geq 0, i = 1, 2, ..., n; j = 1, 2, ..., m.$$

При решении этой задачи могут представиться следующие два случая.

1) $\max f < 0$

В этом случае система ограничений задачи не имеет допустимого базисного вида, в котором искусственные переменные $y_1, y_2, ..., y_m$ являются свободными. Это означает, что исходная система не имеет допустимого базисного вида, т.е. любая задача линейного программирования в канонической форме с этой системой ограничений недопустима.

2) $\max f = 0$

В этом случае в точке максимума искусственные переменные $y_1, y_2, ..., y_m = 0$. При решении вспомогательной задачи симплекс-методом могут встретиться два случая:

a) В результате цепочки шагов все искусственные переменные становятся свободными. Полагая их равными нулю, мы получим допустимый базисный вид исходной системы ограничений, который можно использовать при решении исходной задачи симплекс-методом.

б) Не все искусственные переменные выводятся из состава базисных. Некоторыми простыми преобразованиями симплекс-таблицы всегда можно добиться вывода искусственных переменных из базиса, завершив тем самым подготовку к решению исходной задачи. Для этого следует произвести ряд прямых замещений базисных искусственных переменных еще оставшимися свободными переменными x_i. Поскольку в последнем опорном решении $y_1, y_2, ..., y_m = 0$, столбец свободных членов симплекс-таблицы при этом не изменяется.

Замечание 1.2. В исходной системе ограничений некоторые переменные могут входить по одному в соответствующие уравнения, т.е. быть «уединенными». При составлении вспомогательной задачи можно не вводить искусственные переменные в уравнения, содержащие «уединенные» переменные, знаки коэффициентов при которых не противоположны знакам свободных членов. После деления этих уравнений на коэффициенты при «уединенных» переменных, эти переменные будут играть роль базисных во вспомогательной задаче.

Пример 1.8. Решить следующую задачу:

$$z = -x_1 - 2x_2 \to max,$$

$$\begin{cases} 3x_1 - 5x_2 + x_3 + 2x_4 = 1, \\ 2x_1 - 2x_2 + x_4 - x_5 = -4, \\ x_1 - 3x_2 + 2x_4 - x_5 = -5; \end{cases}$$

$$x_i \geq 0, i = 1, 2, ..., 5.$$

В первое уравнение можно не вводить искусственную переменную. Роль базисной переменной может играть x_3. Сформулируем вспомогательную задачу

$$f = -y_1 - y_2 \to max,$$

$$\begin{cases} 3x_1 - 5x_2 + x_3 + 2x_4 = 1, \\ -2x_1 + 2x_2 - x_4 + x_5 + y_1 = 4, \\ -x_1 + 3x_2 - 2x_4 + x_5 + y_2 = 5; \end{cases}$$

$$y_i, x_i \geq 0 (j = 1, 2; i = 1, 2, ..., 5).$$

Исключим базисные переменные y_1, y_2, из целевой функции и составим первую симплекс-таблицу

Таблица 1

Баз.пер.	Св.чл.	x_1	x_2	x_3	x_4	$\downarrow x_5$	y_1	y_2
x_3	1	3	-5	1	2	0	0	0
$\leftarrow y_1$	4	-2	2	0	-1	1	1	0
y_2	5	-1	3	0	-2	1	0	1
f	-9	3	-5	0	3	-2	0	0

Подготовка системы ограничений завершена. Отбрасывая в последней таблице два последних столбца и последнюю строку, получим после исключения базисных переменных из целевой функции исходной задачи ее первую симплекс-таблицу

По последней симплекс-таблице читаем ответ исходной задачи: $z_{max} = -5/4$, точка максимума $X(3/4; 1/4; 0; 0; 5)$.

Таблица 2

Баз.пер.	Св.чл.	x_1	$\downarrow x_2$	x_3	x_4	x_5	y_1	y_2
x_3	1	3	-5	1	2	0	0	0
x_5	4	-2	2	0	-1	1	1	0
$\leftarrow y_2$	1	1	1	0	-1	0	-1	1
f	-1	-1	-1	0	1	0	2	0

Таблица 2

Баз.пер.	Св.чл.	x_1	x_2	x_3	x_4	x_5	y_1	y_2
x_3	6	8	0	1	-3	0	0	5
x_5	2	-4	0	0	1	1	1	-2
x_2	1	1	1	0	-1	0	-1	1
f	0	0	0	0	0	0	1	1

Иногда при решении вспомогательной задачи в последней симплекс-таблице не все искусственные переменные выводятся из состава базисных.

Пример 1.9. Привести к допустимому базисному виду следующую систему ограничений

$$\begin{cases} -x_1 - 2x_2 + 3x_3 - 3x_4 - x_5 - x_6 = -1, \\ x_1 + 4x_2 - 5x_3 - 5x_4 - 3x_5 - x_6 = 2, \\ -4x_1 + 4x_2 - 12x_3 - 2x_5 + 2x_6 = 2; \end{cases}$$

$$x_i \geq 0, i = 1, 2, ..., 6.$$

Сформулируем вспомогательную задачу

$$f = -y_1 - y_2 - y_3 \rightarrow max,$$

$$\begin{cases} x_1 + 2x_2 - 3x_3 + 3x_4 + x_5 + x_6 = 1, \\ x_1 + 4x_2 - 5x_3 - 5x_4 - 3x_5 - x_6 = 2, \\ -4x_1 + 4x_2 - 12x_3 - 2x_5 + 2x_6 = 2; \end{cases}$$

$$x_i, y_i \geq 0, i = 1, 2, ..., 6; j = 1, 2, 3.$$

Составим первую симплекс-таблицу, производя попутно исключение базисных переменных из целевой функции.

Таблица 2 является последней в решении вспомогательной задачи. Однако искусственные переменные еще не выведены из состава базисных. Мы выведем переменную y_2 из состава базисных, вводя вместо нее переменную x_3, несмотря на то, что правило выбора разрешающего элемента здесь нарушено. Строку для f можно при этом вовсе отбросить.

Произведем замещение базисной переменной y_3, например, на переменную x_1.

Теперь можно записать допустимый базисный вид исходной системы ограничений, меняя для удобства записи уравнения местами.

Таблица 1

Баз.пер.	Св.чл.	$\downarrow x_1$	x_2	x_3	x_4	x_5
$\leftarrow x_3$	6	8	0	1	−3	0
x_5	2	−4	0	0	1	1
x_2	1	1	1	0	−1	0
z	−2	−1	0	0	2	0

Таблица 1

Баз.пер.	Св.чл.	x_1	x_2	x_3	x_4	x_5
x_1	3/4	1	0	1/8	−3/8	0
x_5	5	0	0	1/2	−1/2	1
x_2	1/4	0	1	−1/8	−5/8	0
z	−5/4	0	0	1/8	13/8	0

$$\begin{cases} x_1 + 6x_4 + \dfrac{17}{6}x_5 + \dfrac{3}{2} = 0, \\ x_2 - 11x_4 - \dfrac{23}{6}x_5 - \dfrac{5}{2}x_6 = 2, \\ x_3 - 5x_4 - \dfrac{13}{6}x_5 - \dfrac{3}{2}x_6 = 0. \end{cases}$$

1.6. Метод больших штрафов

Метод искусственного базиса решает задачу линейного программирования в два этапа. Вначале при подготовке исходной задачи решается вспомогательная задача, а затем симплекс-метод применяется к решению исходной задачи. В методе больших штрафов ограничиваются одним этапом. По данной задаче строят некоторую вспомогательную задачу (M-задачу), имеющую нужную форму, и ее решают симплекс-методом. Затем по решению этой задачи находят и решение исходной задачи. Пусть дана произвольная задача в канонической форме:

$$z = c_0 + c_1x_1 + c_2x_2 + \dots + c_nx_n \to \max$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m \end{cases}$$

$$x_i \geq 0,\ i = 1, 2, \dots, n.$$

Как и раньше считаем, что свободные члены $b_1, b_2, \dots, b_m \geq 0$. Введем искусственные переменные $y_1, y_2, \dots, y_m \geq 0$ в систему ограничений и целевую функцию исходной задачи следующим образом:

$$z = c_0 + c_1x_1 + c_2x_2 + \dots + c_nx_n - M(y_1 + y_2 + \dots + y_m) \to \max;$$

Таблица 1

Баз.пер.	Св.чл.	x_1	$\downarrow x_2$	x_3	x_4	x_5	x_6	y_1	y_2	y_3
$\leftarrow y_1$	1	1	2	-3	3	1	1	1	0	0
y_2	2	1	4	-5	-5	-3	-1	0	1	0
y_3	2	-4	4	-12	0	-2	2	0	0	1
f	-5	2	-10	20	2	4	-2	0	0	0

Таблица 2

Баз.пер.	Св.чл.	x_1	x_2	$\downarrow x_3$	x_4	x_5	x_6	y_1	y_2	y_3
x_2	$1/2$	$1/2$	1	$-3/2$	$3/2$	$1/2$	$1/2$	$1/2$	0	0
$\leftarrow y_2$	0	-1	0	1	-11	-5	-3	-2	1	0
y_3	0	-6	0	-6	-6	-4	0	-2	0	1
f	0	7	0	5	17	9	3	5	0	0

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + y_1 = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + y_2 = b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + y_n = b_m \end{cases}$$
$$x_i \geq 0,\ i = 1, 2, \dots, n;\ y_j \geq 0,\ j = 1, 2, \dots, m.$$

Здесь $M > 0$ — достаточно большое число. Построенная задача называется M-задачей по отношению к исходной задаче. Система ограничений этой задачи имеет допустимый базисный вид. Правда, целевая функция содержит базисные переменные $y_1, y_2, \dots, y_m \geq 0$. Но они легко исключаются из выражения для z_M с помощью самих уравнений системы ограничений. Имеют место следующие утверждения:

1. Если для всех достаточно больших $M > 0$ M-задача имеет решение, то и исходная задача имеет решение, причем $z_{max} = z_{Mmax}$, а точка максимума для z может быть получена из точки максимума для z_M отбрасыванием значений искусственных переменных. При этом в точке максимума z_M значения искусственных переменных равны нулю.

2. Если при всех достаточно больших $M > 0$ M-задача не имеет решения, то и исходная задача не имеет решения.

Таким образом, для решения исходной задачи следует выбрать некоторое достаточно большое M и решить M-задачу. Если число M выбрано недостаточно большим, то точка максимума для z_M может иметь отличные от нуля значения искусственных переменных. В этом случае определить по решению M-задачи решение исходной задачи нельзя. Обычно число M берётся на порядок больше, чем коэффициенты в системе ограничений и целевой функции исходной задачи. Отметим также, что при введении искусственных переменных нет нужды вводить их во все ограничения задачи. Искусственные переменные можно не вводить в ограничения, содержащие уединённые переменные, знак коэффициентов которых совпадает со знаком соответствующих свободных членов.

Пример 1.10. Решить задачу:

$$z = -3x_1 - 2x_2 \to \max;$$

Таблица 3

Баз.пер.	Св.чл.	$\downarrow x_1$	x_2	x_3	x_4	x_5	x_6	y_1	y_2	y_3
x_2	$1/2$	-1	1	0	-15	-7	-4	$-5/2$	$3/2$	0
x_3	0	-1	0	1	-11	-5	-3	-2	1	0
$\leftarrow y_3$	0	-12	0	0	-72	-34	-18	-14	6	1

Таблица 4

Баз.пер.	Св.чл.	x_1	x_2	x_3	x_4	x_5	x_6	y_1	y_2	y_3
x_2	$1/2$	0	1	0	-11	$-23/6$	$-5/2$	$-8/6$	1	$-\dfrac{1}{12}$
x_3	0	0	0	1	-5	$-13/6$	$-3/2$	$-5/6$	$1/2$	$-\dfrac{1}{12}$
x_1	0	1	0	0	6	$17/6$	$3/2$	$7/6$	$-1/2$	$-\dfrac{1}{12}$

$$\begin{cases} x_1 + x_2 = 10, \\ x_1 - x_3 = 4; \end{cases}$$
$$x_1, x_2, x_3 \geq 0.$$

Система ограничений этой задачи имеет базисный вид, однако он не соответствует опорному решению. В первом уравнении роль базисной может играть переменная x_2. Во второе уравнение придется ввести искусственную переменную y. M-задачу построим при $M = 20$:

$$z_M = -3x_1 - 2x_2 - 20y \to \max;$$
$$\begin{cases} x_1 + x_2 = 10, \\ x_1 - x_3 + y = 4; \end{cases}$$
$$x_1, x_2, x_3, y \geq 0.$$

Выразим базисные переменные x_2, y из системы ограничений и подставим в целевую функцию z_M. Задача примет вид

$$z_M = -100 + 19x_1 - 20x_3 \to \max;$$
$$\begin{cases} x_1 + x_2 = 10, \\ x_1 - x_3 + y = 4; \end{cases}$$
$$x_1, x_2, x_3, y \geq 0.$$

Для этой задачи можно составить симплекс-таблицы:

Баз. пер.	Св. чл.	x_1	x_2	x_3	y
x_2	10	1	1	0	0
Y	4	1	0	-1	1
z_M	-100	-19	0	20	0

Баз. пер.	Св. чл.	x_1	x_2	x_3	y
x_2	6	0	1	1	-1
x_1	4	1	0	-1	1
z_M	-24	0	0	0	19

M-задача имеет следующее решение:

$$z_{Mmax} = -24;\ x_1 = 4;\ x_2 = 6;\ x_3 = 0;\ y = 0.$$

Решение исходной задачи:

$$z_{max} = -24;\ x_1 = 4;\ x_2 = 6;\ x_3 = 0.$$

1.7. Понятие о «трудоёмкости» симплекс-алгоритма и о методах внутренних точек

Полувековая практика решения задач линейного программирования показала высокую эффективность симплекс-метода и различных его модификаций. При решении задач с m ограничениями и n переменными, как правило, оказывалось достаточно m итераций. При этом количество элементарных арифметических операций имело порядок n^2m. Однако теоретические исследования по оценке «трудоемкости» решения задач линейного программирования симплекс-методом показала, что эта «трудоемкость» может быть значительно больше. В 1972 г. американские ученые В. Кли и Дж. Минти построили пример задачи линейного программирования с n переменными и $2n$ ограничениями, для решения которой требуется не менее $2^n - 1$ итераций симплекс-метода. Тем самым было показано, что симплекс-метод на классе всех линейных задач является алгоритмом «экспоненциальной трудоемкости» — количество необходимых вычислений оценивается экспоненциальной функцией параметров задачи. Этот факт означает, что существуют задачи не слишком большой размерности, решение которых симплекс-методом невозможно за обозримое время. Хотя все примеры такого рода искусственны и для задач, пришедших из практики, ничего подобного не происходит, возник естественный вопрос о существовании алгоритма, для которого необходимый объем вычислений при решении произвольной задачи линейного программирования оценивается полиномом от параметров задачи. Другими словами, обладает ли класс задач линейного программирования экспоненциальной сложностью или эта сложность полиномиальная? Вопрос этот нашел свое решение в теореме, доказанной советским математиком Л.Г. Хачияном [42]. Он применил к задачам линейного программирования новый алгоритм (метод эллипсоидов), построенный усилиями советских ученых А.С. Немировского, Н.З. Шоша и Д.Б. Юдина. Для формулировки теоремы Хачияна обозначим через h максимум модулей коэффициентов и свободных членов в целевой функции и системе ограничений задачи в стандартной форме. Для произвольных функций $f(t), g(t)$ будем писать $f = O(g)$, если существует такая константа C, что $f(t) \le Cg(t)$ для всех t.

Теорема 1.2. Для решения задачи линейного программирования достаточно $O(n^4(n+m)\ln hn)$ элементарных операций (сложение, вычитание, умножение, деление, извлечение квадратного корня, нахождение наибольшего из двух чисел). При этом достаточно использовать для вычислений число разрядов, равное $O(n \ln hn)$. Таким образом, *класс задач линейного программирования имеет полиномиальную сложность*. К сожалению, в вычислительном плане метод эллипсоидов оказался неперспективным. Однако факт полиномиальной сложности задач линейного программирования привёл в дальнейшем к созданию целого класса эффективных алгоритмов линейного программирования, которые получили название методов внутренней точки [18]. Первым из этих методов был алгоритм Н. Кармаркара, предложенный в 1984 г. Метод внутренней точки тоже переходит от точки к точке, улучшая значение целевой функции, но остаётся при этом во внутренности области допустимых значений задачи. Подобравшись достаточно близко к оптимальному опорному решению, он находит это решение. Алгоритм Кармаркара имеет время работы $O(n^4L)$, где L — длина битовой записи входных данных. В заключение отметим, что приведенные оценки указывают на более высокую эффективность методов внутренней точки по сравнению с симплекс-методом лишь для самых «плохих» задач линейного программирования, далеких от реальных. На практике же симплекс-метод по-прежнему является основным инструментом в линейном программировании.

Контрольные вопросы и задачи для самостоятельного решения

1. Как формулируется общая задача линейного программирования?
2. Когда задача линейного программирования называется имеющей каноническую форму?
3. Какая форма задачи линейного программирования называется стандартной?
4. В чем заключается геометрическое истолкование системы ограничений и целевой функции задачи в случае двух переменных?
5. Дайте определения базисного вида системы линейных уравнений, базисного и опорного решений такой системы.
6. Сформулируйте фундаментальную теорему симплекс-метода.
7. К какому виду должна быть приведена задача линейного программирования перед применением симплекс-метода?
8. Как составить первую симплекс-таблицу?
9. Опишите порядок работы с симплекс-таблицей. В чем заключается признак того, что симплекс-таблица является последней? Как прочесть решение задачи по последней симплекс-таблице? В каком случае по последней симплекс-таблице можно заключить, что задача не имеет решения по причине неограниченности целевой функции на области допустимых значений?
10. Для чего применяется метод искусственного базиса? Какие основные случаи могут представиться при работе этим методом?
11. Опишите метод больших штрафов. Как составить М-задачу для задачи линейного программирования в канонической форме?
12. Как избежать зацикливания симплекс алгоритма?
13. Что понимается под трудоемкостью симплекс метода? Что означает его экспоненциальная трудоемкость на классе всех задач линейного программирования?
14. Существуют ли алгоритмы решения задач линейного программирования полиномиальной трудоемкости? Обладает ли класс всех задач линейного программирования полиномиальной сложностью?

Построить математические модели в задачах 1.1-1.4

1.1. Для изготовления трех видов изделий А, В, С используется токарное, фрезерное, сварочное и шлифовальное оборудование. Затраты времени на обработку одного изделия для каждого из типов оборудования указаны в следующей ниже таблице. В ней же указан общий фонд рабочего времени каждого из типов оборудования, а также прибыль от реализации одного изделия каждого вида. Требуется определить, сколько изделий каждого вида следует изготовить предприятию, чтобы прибыль от их реализации была максимальной.

1.2. Кондитерская фабрика для производства трех видов карамели A, B и C использует три вида основного сырья: сахарный песок, патоку и фруктовое пюре. Нормы

Тип оборудования	Затраты времени (станко-ч) на обработку одного изделия вида			Общий фонд рабочего времени оборудования (ч)
	A	B	C	
Фрезерное	2	4	5	120
Токарное	1	8	6	280
Сварочное	7	4	5	240
Шлифовальное	4	6	7	360
Прибыль (руб.)	10	14	12	

расхода сырья каждого вида на производство 1 т. карамели данного вида приведены в таблице. В ней же указано общее количество сырья каждого вида, которое может быть использовано фабрикой, а также приведена прибыль от реализации 1 т. карамели каждого вида.

Вид сырья	Нормы расхода сырья (т) на 1 т карамели			Общее количество сырья (т)
	A	B	C	
Сахарный песок	2	4	5	120
Патока	1	8	6	280
Фруктовое пюре	7	4	5	240
Прибыль от реализации 1т карамели (руб.)	108	112	126	

Найти оптимальный план производства карамели, обеспечивающий максимальную прибыль от ее реализации.

1.3. При откорме животных каждое животное ежедневно должно получать не менее 60 ед. питательного вещества А, не менее 50 ед. вещества В и не менее 12 ед. вещества С. Указанные питательные вещества содержаться в трех видах корма. Содержание единиц питательных веществ в 1 кг каждого из видов корма приведено в следующей таблице: Составить дневной рацион, обеспечивающий получение необходимого количества питательных веществ при минимальных денежных затратах, если цена 1 кг корма I-го вида составляет 9 коп., корма II-го вида — 12 коп., а корма III-го вида — 10 коп.

1.4. При производстве чугунного литья используется n различных исходных шихтовых материалов (чугун различных марок, стальной лом, феррофосфор и др.) Химический состав чугунного литья определяется содержанием в нем химических элементов

Питательные вещества	Количество единиц питательных веществ в 1 кг корма вида		
	I	II	III
A	1	3	4
B	2	4	2
C	1	4	3

(кремния, марганца, фосфора и др.). Готовый чугун должен иметь строго определенный химический состав, который определяется величинами H_j, представляющими собой доли (в процентах) j-го химического элемента в готовом продукте ($j = 1, 2, \ldots, m$). При этом считаются известными величины h_{ij} — содержание (в процентах) j-го химического элемента в i-ом исходном шихтовом материале, а также величины c_i — цены единицы каждого шихтового материала ($i = 1, 2, \ldots, n$). Определить состав шихты, обеспечивающий получение литья заданного качества при минимальной общей стоимости используемых шихтовых материалов.

В задачах 1.5 — 1.8 *привести математическую модель линейного программирования к каноническому виду.*

1.5.

$$z = -2x_1 - x_2 + x_3 \to min,$$

$$\begin{cases} 2x_1 & -x_2 & +6x_3 & \leq & 12, \\ 3x_1 & +5x_2 & -12x_3 & = & 14, \\ -3x_1 & +6x_2 & -4x_3 & \leq & 18, \end{cases}$$

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

1.6.

$$z = -2x_1 + x_2 + 5x_3 \to min,$$

$$\begin{cases} 4x_1 & +2x_2 & +5x_3 & \leq & 12, \\ 6x_1 & -3x_2 & +4x_3 & = & 15, \\ 3x_1 & +3x_2 & -2x_3 & \leq & 16, \end{cases}$$

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

1.7.

$$z = 2x_1 - 5x_2 + 3x_3 \to min,$$

$$\begin{cases} -x_1 & +x_2 & +x_3 & \geq & 12, \\ x_1 & +5x_2 & -6x_3 & \leq & 16, \\ 3x_1 & +x_2 & +x_3 & \geq & 18, \end{cases}$$

$$x_1, x_2 \geq 0.$$

1.8.

$$z = -3x_1 + x_2 - 5x_3 \to min,$$

$$\begin{cases} 2x_1 & +5x_2 & -7x_3 & \leq & 4, \\ -4x_1 & -3x_2 & +8x_3 & \geq & 15, \\ 3x_1 & -2x_2 & +10x_3 & \leq & 11, \end{cases}$$

$$x_2, x_3 \geq 0.$$

Используя геометрическое истолкование задач линейного программирования, найти решения задач 1.9 — 1.13.

1.9.

$$z = x_1 + x_2 \to max;$$

$$\begin{cases} x_1 & + & 2x_2 & \leq & 14, \\ -5x_1 & + & 3x_2 & \leq & 15, \\ 4x_1 & + & 6x_2 & \geq & 24, \end{cases}$$

$$x_1, x_2 \geq 0.$$

Ответ: точка максимума (14; 0);

$$z_{max} = 14.$$

1.10.

$$z = x_1 + 2x_2 \to max;$$

$$\begin{cases} 4x_1 & - & 2x_2 & \leq & 12, \\ -x_1 & + & 3x_2 & \leq & 6, \\ 2x_1 & + & 4x_2 & \geq & 16, \end{cases}$$

$$x_2, x_3 \geq 0.$$

Ответ: точка максимума (4,8; 3,6);

$$z_{max} = 12.$$

1.11.

$$z = -2x_1 + x_2 \to min;$$

$$\begin{cases} 3x_1 & - & 2x_2 & \leq & 12, \\ -x_1 & + & 2x_2 & \leq & 8, \\ 2x_1 & + & 3x_2 & \geq & 6, \end{cases}$$

$$x_1, x_2 \geq 0.$$

Ответ: точка минимума (10;9);

$$z_{min} = -11.$$

1.12.

$$z = -x_1 + 4x_2 + 2x_4 - x_5 \to max;$$

$$\begin{cases} x_1 & - & 5x_2 + x_3 & = & 5, \\ -x_1 & + & x_2 + x_4 & = & 4, \\ x_1 & + & x_2 + x_5 & = & 8, \end{cases}$$

$$x_1, x_2, \ldots, x_5 \geq 0.$$

Ответ: точка максимума (2; 6; 33; 0; 0;);

$$z_{max} = 22.$$

1.13.

$$z = -5x_1 + x_2 - x_3 \to max;$$

$$\begin{cases} 3x_1 & - & x_2 - x_3 & = & 4, \\ x_1 & - & x_2 + x_3 - x_4 & = & 1, \\ 2x_1 & + & x_2 + 2x_3 + x_5 & = & 7, \end{cases}$$

$$x_1, x_2, \dots, x_5 \geq 0.$$

Ответ: точка максимума $(\frac{4}{3};\ 0;\ 0;\ \frac{1}{3};\ \frac{13}{3})$;

$$z_{max} = -\frac{20}{3}.$$

В задачах 1.14-1.17 *привести систему уравнений к какому-нибудь базисному виду.*

1.14.

$$\begin{cases} x_1 & -2x_2 & +3x_3 & +4x_4 & = 1, \\ 4x_1 & -7x_2 & +2x_3 & +x_4 & = 3, \\ 3x_1 & -5x_2 & -x_3 & -3x_4 & = 2, \end{cases}$$

1.15.

$$\begin{cases} x_1 & 4x_2 & -2x_3 & +3x_5 & = 2, \\ 2x_1 & 9x_2 & -x_3 & -4x_4 & = 5, \\ x_1 & 5x_2 & +x_3 & -4x_4 + 3x_5 & = 3, \end{cases}$$

1.16.

$$\begin{cases} x_1 & +3x_2 & -x_3 & -2x_4 & = 1, \\ 2x_1 & +7x_2 & -4x_3 & -3x_4 & = 3, \\ x_1 & +4x_2 & -3x_3 & -x_4 & = 2, \end{cases}$$

1.17.

$$\begin{cases} x_1 & -5x_2 & +3x_3 & +4x_4 & = 4, \\ 2x_1 & -9x_2 & +2x_3 & +x_5 & = 7, \\ x_1 & -4x_2 & -x_3 - 4x_4 & +x_5 & = 3, \end{cases}$$

С помощью симплекс-метода и его модификаций найти решение задач 1.18-1.27.

1.18.

$z = 3x_1 + 2x_3 - 6x_6 \to max;$

$$\begin{cases} 2x_1 & +x_2 & -2x_4 & +x_5 & = 16, \\ -3x_1 & +2x_2 & +x_3 & -3x_4 & = 18, \\ x_1 & +3x_2 & +4x_4 & +x_6 & = 24, \end{cases}$$

$x_i \geq 0(i = \overline{1,6})$.

Ответ: точка максимума (18; 0; 6; 66; 0; 0); $z_{max} = 66$.

1.19.

$z = 2x_1 + 3x_2 - x_4 \to max;$

$$\begin{cases} 2x_1 & -2x_2 & -2x_4 & +x_5 & = 16, \\ 3x_1 & +2x_2 & +x_3 & -3x_4 & = 18, \\ -x_1 & +3x_2 & +4x_4 & +x_6 & = 24, \end{cases}$$

$x_i \geq 0(i = \overline{1,6})$.

Ответ: точка максимума $(\frac{6}{11};\frac{90}{11}$ 0; 0; $\frac{254}{11}$;0); $z_{max} = \frac{282}{11}$.

1.20.

$z = 8x_2 + 7x_4 + x_6 \to max;$

$$\begin{cases} x_1 & -2x_2 & -3x_4 & -2x_6 & = 12, \\ 4x_2 & +x_3 & -4x_4 & -3x_6 & = 12, \\ 5x_2 & +5x_4 & +x_5 & +x_6 & = 25, \end{cases}$$

$x_i \geq 0(i = \overline{1,6})$.

Ответ: точка максимума (23; 4; 0; 1; 0; 0); $z_{max} = 39$.

1.21.

$z = x_1 + 3x_2 - 5x_4 \to max;$

$$\begin{cases} 2x_1 & +4x_2 & +x_3 & +2x_4 & = 28, \\ -3x_1 & +5x_2 & -3x_4 & +x_5 & = 30, \\ 4x_1 & -2x_2 & +8x_4 & +x_6 & = 32, \end{cases}$$

$x_i \geq 0 (i = \overline{1,6})$.

Ответ: точка максимума $(\frac{10}{11}; \frac{72}{11}$ 0; 0; 0; $\frac{456}{11})$; $z_{max} = \frac{226}{11}$.

1.22.

$z = 3x_1 + 2x_2 - 5x_6 \to max$;

$$\begin{cases} 2x_1 & +x_2 & -3x_5 & +5x_6 & = 34, \\ 4x_1 & +x_3 & +2x_5 & -4x_6 & = 28, \\ -3x_1 & +x_4 & -3x_5 & +6x_6 & = 24, \end{cases}$$

$x_i \geq 0 (i = \overline{1,6})$.

Ответ: точка максимума (0; 76; 0; 66; 14; 0); $z_{max} = 28$.

1.23.

$z = x_1 + 2x_2 - x_3 \to max$;

$$\begin{cases} -x_1 & +4x_2 & -2x_3 & \leq & 6, \\ x_1 & +x_2 & +2x_3 & \geq & 6, \\ 2x_1 & -x_2 & +2x_3 & = & 4, \end{cases}$$

$x_1, z_2, x_3 \geq 0$.

Ответ: точка максимума (2,8; 2,4; 0,4); $z_{max} = 7,2$.

1.24.

$z = 8x_1 - 3x_2 + x_3 + 6x_4 - 5x_5 \to max$;

$$\begin{cases} 2x_1 & +4x_2 & +x_3 & +x_4 & -2x_5 & = 28, \\ x_1 & -2x_2 & & +x_4 & +x_5 & = 31, \\ -x_1 & +3x_2 & +5x_3 & +4x_4 & -8x_5 & = 118, \end{cases}$$

$x_i \geq 0 (i = \overline{1,5})$.

Ответ: точка максимума (0; 0; 6; 28; 3); $z_{max} = 159$.

1.25.

$z = 2x_1 - 3x_2 + 4x_3 + 5x_4 - x_5 + 8x_6 \to max$;

$$\begin{cases} x_1 & +5x_2 & -3x_3 & -4x_4 & +2x_5 & +x_6 & = 120, \\ 2x_1 & +9x_2 & -5x_3 & -7x_4 & +4x_5 & +2x_6 & = 320, \end{cases}$$

$x_i \geq 0 (i = \overline{1,6})$.

Ответ: точка максимума (0; 0; 0; 80; 0; 440); $z_{max} = 3920$.

1.26.

$z = -3x_1 + 5x_2 - 3x_3 + x_4 + x_5 + 8x_6 \to max$;

$$\begin{cases} x_1 & -3x_2 & +4x_3 & +5x_4 & -6x_5 & +x_6 & = 60, \\ 7x_1 & -17x_2 & +26x_3 & +31x_4 & -35x_5 & +6x_6 & = 420, \end{cases}$$

$x_i \geq 0 (i = \overline{1,6})$.

Ответ: точка максимума (0; 0; 0; 0; 60; 420); $z_{max} = 3420$.

1.27.

$z = 8x_1 - 3x_2 + x_3 + 6x_4 - 5x_5 \to max$;

$$\begin{cases} 2x_1 & -x_2 & & +3x_4 & +x_5 & -x_6 & = 36, \\ -x_1 & +2x_2 & +x_3 & +2x_4 & & +2x_6 & = 20, \\ 3x_1 & -x_2 & +2x_3 & -x_4 & +3x_5 & +x_6 & = 30, \end{cases}$$

$x_i \geq 0 (i = \overline{1,6})$.

Ответ: точка максимума (6; 0; 10; 8; 0; 0); $z_{max} = 190$.

2. Транспортная задача

2.1. Закрытая транспортная задача. Нахождение опорных планов

Одной из задач линейного программирования является транспортная задача, состоящая, в общей постановке, в отыскании оптимального плана перевозок некоторого однородного груза с m баз $A_1, A_2, \ldots, A_m$ n потребителям $B_1, B_2, \ldots, B_n$. Пусть имеются определенные запасы груза на базах $A_1, A_2, \ldots, A_m$, которые мы обозначим $a_1, a_2, \ldots, a_m$ соответственно. Заказы каждого из потребителей (потребности) обозначим $b_1, b_2, \ldots, b_n$. Общее количество имеющегося груза обозначим $A(A = a_1 + a_2 + \ldots + a_m)$, а общие потребности – через $B(B = b_1 + b_2 + \ldots + b_n)$. При условии $A = B$ мы имеем закрытую модель, а при $A \neq B$ – открытую модель транспортной задачи.

В этом параграфе мы рассмотрим закрытую транспортную задачу. Предположим, что базы и потребители соединены коммуникациями, например железными дорогами. Обозначим стоимость перевозки единицы груза из пункта A_i в пункт B_j через c_{ij}. Числа c_{ij} можно расположить в виде матрицы

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}$$

Эта матрица называется *матрицей стоимостей*. Задача заключается в нахождении такого плана перевозок, чтобы общая стоимость их была минимальной. При этом необходимо выполнить следующие два условия:

1) все запасы груза должны быть вывезены;
2) заказы потребителей должны быть выполнены.

Такие требования можно соблюсти лишь при условии $A = B$, то есть в случае закрытой модели. Обозначим через x_{ij} количество груза, которое планируется перевезти из пункта A_i в пункт B_j. Эти величины будут представлять собой переменные в нашей задаче. Их также можно расположить в виде матрицы, которая называется матрицей перевозок. Общая стоимость перевозок (целевая функция нашей задачи) имеет вид

$$z = \sum_{i,j} c_{ij} x_{ij}$$

Эту функцию требуется минимизировать. При этом, однако, величины x_{ij} не могут принимать произвольные значения. Они *неотрицательны* и удовлетворяют следующим ограничениям, которые выражают требования 1) и 2), сформулированные выше:

$$\begin{cases} x_{11} + x_{12} + \ldots + x_{1n} = a_1; \\ x_{21} + x_{22} + \ldots + x_{2n} = a_2; \\ \ldots \\ x_{m1} + x_{m2} + \ldots + x_{mn} = a_m. \end{cases} \qquad \begin{cases} x_{11} + x_{12} + \ldots + x_{1n} = b_1; \\ x_{21} + x_{22} + \ldots + x_{2n} = b_2; \\ \ldots \\ x_{m1} + x_{m2} + \ldots + x_{mn} = b_m. \end{cases}$$

Приведенная система ограничений состоит из двух подсистем, которые называются системой горизонтальных и вертикальных уравнений. Таким образом, закрытая транспортная задача является задачей линейного программирования с ограничениями равенствами. Система ее ограничений содержит $n + m$ уравнений с $n \cdot m$ неизвестными. Для решения транспортной задачи также применяют симплекс-метод, но в силу специфики задачи здесь можно обойтись без симплекс-таблиц. План перевозок с указанием запасов и потребностей, а также стоимостей перевозок, удобно записывать в виде таблицы данных:

Запасы	Потребности			
	b_1	b_2	...	b_n
a_1	$\boxed{c_{11}}$ x_{11}	$\boxed{c_{12}}$ x_{12}	...	$\boxed{c_{1n}}$ x_{1n}
a_2	$\boxed{c_{12}}$ x_{21}	$\boxed{c_{22}}$ x_{22}	...	$\boxed{c_{2n}}$ x_{2n}
...	...	...	...	...
a_m	$\boxed{c_{m1}}$ x_{m1}	$\boxed{c_{m2}}$ x_{m2}	...	$\boxed{c_{mn}}$ x_{mn}

Решение задачи можно получить путем некоторых преобразований этой таблицы, которые соответствуют переходу от одного плана перевозок к другому. Как и в общем случае, оптимальное решение ищется среди опорных решений системы ограничений транспортной задачи. Ранг этой системы равен $m + n - 1$, поэтому среди $m \cdot n$ переменных x_{ij} выделяются $m + n - 1$ базисных переменных, а остальные $(m - 1)(n - 1)$ переменные являются свободными. В базисном решении свободные переменные равны нулю. Обычно эти нули в таблицу не вписывают, оставляя соответствующие клетки пустыми. Таким образом, при внесении в таблицу перевозок вместо x_{ij} чисел, представляющих опорный план, мы имеем $m + n - 1$ заполненных и $(m - 1)(n - 1)$ пустых клеток. Как и в общем случае, решение транспортной задачи начинается с отыскания первого опорного плана (исходного опорного решения системы ограничений). Мы рассмотрим два метода построения такого опорного плана. Суть обоих методов состоит в том, что опорный план составляется последовательно в несколько шагов (точнее, $m + n - 1$ шагов). На каждом из шагов заполняется одна клетка. При рассмотрении клетки с номерами (k, r) на первом шаге может представиться три случая:

а) $a_k > b_r$
б) $a_k = b_r$
в) $a_k < b_r$

В случае а) в клетку ставится число b_r, вычеркивается r-тый столбец, а запасы в пункте A_k полагаются равными $a_k - b_r$. В случае в) в клетку записывается число a_k, вычеркивается k-тая строка, а потребности в пункте b_r полагаются равными $b_r - a_k$. В случае б) в клетку ставится число $a_k = b_r$ и вычеркивается, по выбору, строка или столбец (вычеркивать и строку, и столбец нельзя). Оставшиеся потребности в пункте b_r или запасы в пункте a_k полагаются равными нулю. После первого шага наша таблица сократилась на одну строку или на один столбец, а потребности или запасы соответственно подправлены. В сокращенной таблице снова выбираем для заполнения клетку и повторяем все сначала. Начиная с первоначально данной таблицы и повторив $m + n - 2$ раз описанный шаг, мы придем к “таблице”, состоящей из одной клетки. Заполнив эту последнюю клетку и совершив $m + n - 1$ шаг, мы получим искомый опорный план. Правда, мы не указали, каким образом на каждом шаге мы выбираем клетку для заполнения. Различие методов отыскания первого опорного плана как раз и состоит в различии способов выбора заполняемой клетки.

Метод северо-западного угла. В этом методе на каждом шаге построения первого опорного плана заполняется верхняя левая клетка ("северо-западный угол") оставшейся таблицы.

Метод наименьшей стоимости. В этом методе на каждом шаге построения первого опорного плана заполняется та клетка оставшейся таблицы, которая имеет наименьший тариф (c_{kr}). Если такая клетка не единственная, то заполняется любая из них.

Пример 2.1. Найдем первый опорный план следующей транспортной задачи методом северо-западного угла.

		(10) (30)		
	Запасы	Потребности		
		40	30	10
	10	5 10	7	5
	20	2 20	1	4
(10) (40)	50	6 10	3 30	2 10

В клетку x_{11} ставим 10 и вычеркиваем строку. Потребности в B_1 станут равными 30. В клетку x_{21} ставим 20 и вычеркиваем строку. Потребности в B_1 равны 10. В клетку x_{31} ставим 10 и вычеркиваем столбец. В клетку x_{32} ставим 30 и вычеркиваем столбец. Запасы в A_3 станут равными 10. В клетку x_{33} ставим 10. Процесс нахождения опорного плана окончен. Базисными переменными здесь будут $x_{11}, x_{21}, x_{31}, x_{32}, x_{33}$. Остальные переменные свободные.

Метод наименьшей стоимости дает здесь другой опорный план:

			(10)	(0)
	Запасы	Потребности		
		40	30	10
	10	5 10	7	5
	20	2	1 20	4
(40)	50	6 30	3 10	2 10

Часто значение целевой функции на опорном плане, найденном методом наименьшей стоимости, меньше, чем на плане, полученном методом северо-западного угла. Этот опорный план как бы ближе к оптимальному. Хотя это и не всегда так, но часто нахождение оптимального решения, исходя из опорного плана, построенного методом наименьшей стоимости, требует меньше вычислений.

2.2. Комбинаторные свойства циклов в матрице

Под матрицей в этом параграфе понимается таблица, состоящая из клеток. *Циклом* в матрице будем называть ломаную линию (рис. 2.1), звенья которой располагаются по строкам и столбцам матрицы, и которая удовлетворяет следующим двум условиям:

1) эта ломаная является связной, т.е. из любой её вершины можно попасть в любую другую по звеньям ломаной;

2) в каждой вершине сходятся ровно два звена, причем одно из них располагается по строке, а другое по столбцу таблицы.

Циклы в таблице могут иметь *самопересечения*, т.е. звенья ломаной могут пересекаться в точке, не являющейся вершиной цикла (рис. 2.1, б).

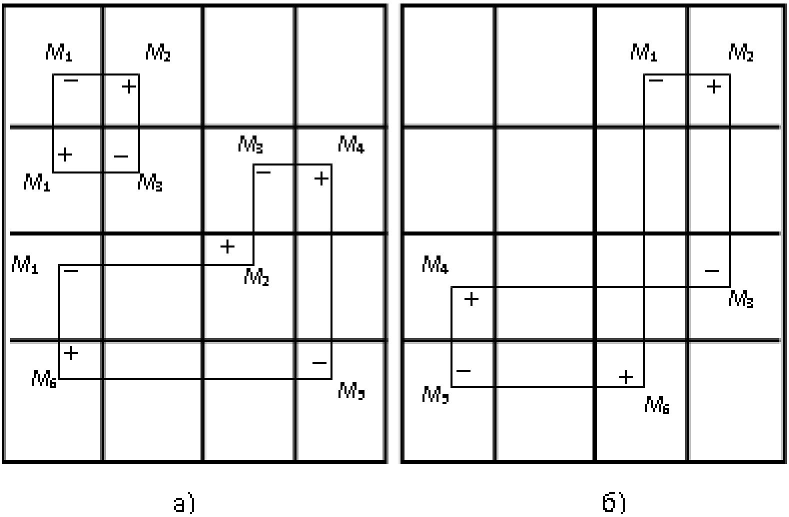

Рис. 2.1. Примеры циклов в матрице

Циклы в матрице обладают рядом свойств, которые описываются леммами 2.1 - 2.3.

Лемма 2.1. Пусть дана матрица, состоящая из n строк и m столбцов, для которой выполнено неравенство $n \cdot m \geq n + m$. Если в этой матрице произвольно отмечено $n+m$ клеток, то существует цикл с вершинами в отмеченных клетках (быть может не во всех).

Доказательство. Докажем лемму по индукции, которую будем вести по величине $n + m$. Проверим утверждение леммы для случая $n + m = 4$. Для меньших значений не может выполняться неравенство $n \cdot m \geq n + m$. Единственной возможностью здесь является $n = m = 2$. Легко увидеть цикл, лежащий в четырех отмеченных клетках. Предположим, что утверждение доказано для всех матриц, для которых $n+m < p$ при некотором натуральномр. Докажем, что наше утверждение справедливо и при $n + m = p$. Действительно, может представиться лишь две возможности

1) для каждой отмеченной клетки в строке и в столбце матрицы находится другая отмеченная клетка;
2) есть хотя бы одна отмеченная клетка, в столбце или строке которой других отмеченных клеток нет.

Рассмотрим вначале случай 2). Если есть отмеченная клетка, для которой нет других отмеченных клеток в строке (или столбце), то ее можно удалить, вычеркивая эту строку (или столбец). Получим матрицу, для которой $n + m < p$. По индуктивному предположению должен существовать цикл с вершинами в отмеченных клетках этой новой матрицы, и утверждение доказано. Если же имеет место случай 1), то из любой отмеченной клетки можно перейти в другую отмеченную клетку по строке (или столбцу). Из новой отмеченной клетки можно перейти в следующую и т. д. Поскольку клеток конечное число, то через ряд шагов мы вернемся в исходную клетку, то есть получим искомый цикл. Лемма 2.1 доказана.

Лемма 2.2. Колчество вершин любого цикла в матрице четно.

Доказательство. Поскольку в каждой вершине сходится ровно два взаимно перпендикулярных звена, то при переходе через вершину звено поворачивается на угол $\pm\pi/2$. В результате обхода цикла суммарный угол поворота должен быть равен $2\pi q$, где $q \in Z$. Обозначим через k количество вершин, в которых звено поворачивается на угол $\pi/2$, а через l — количество вершин, в которых оно поворачивается на угол $-\pi/2$. Тогда

$$\frac{\pi}{2}k - \frac{\pi}{2}l = 2\pi q l = l = 4q$$

Поэтому количество всех вершин $k+l = (k-l)+2l$ будет четным. Лемма 2.2 доказана.

Определение 2.1. Цикл в матрице называется означенным, если каждой вершине его сопоставлен знак «+» или «-», причем при обходе цикла знаки чередуются, т. е. если вершина имеет какой-то знак, то соседним вершинам сопоставляется противоположный знак (рис. 2.1, а), б)).

Из леммы 2.2 следует, что любой цикл в матрице можно сделать означенным, причем двумя различными способами.

Лемма 2.3. Во всяком означенном цикле число положительных вершин, лежащих в каждой строке или столбце, равняется количеству отрицательных вершин в этой строке или столбце.

Доказательство. Для каждой вершины в строке (столбце) имеется единственная соседняя вершина и для двух различных вершин в строке (столбце) соседние вершины различны. Рассмотрим положительные вершины в какой-нибудь строке (столбце). Для них в этой строке (столбце) имеется столько же соседних вершин. Они должны быть отрицательными по определению означенного цикла. Отсюда следует справедливость леммы 2.3.

2.3. Преобразование решения системы ограничений транспортной задачи сдвигом по означенному циклу

Пусть имеется некоторое решение системы ограничений транспортной задачи. Запишем это решение в виде матрицы перевозок. Пусть в матрице перевозок задан некоторый означенный цикл.

Определение 2.2. Сдвигом по означенному циклу на число x матрицы перевозок называется такое преобразование этой матрицы, при котором меняются лишь элементы в клетках, где находятся вершины цикла, причем элемент в клетке с положительной вершиной увеличивается на число x, а элемент в клетке с отрицательной вершиной уменьшается на это же число.

Теорема 2.1. При сдвиге по означенному циклу на число x решение системы ограничений транспортной задачи переходит снова в решение этой же системы ограничений.

Доказательство. Рассмотрим горизонтальное уравнение, в котором слева стоит сумма элементов некоторой строки матрицы. Некоторые из этих элементов увеличиваются на число x, а некоторые уменьшаются на это же число. В силу леммы 2.3 количество положительных вершин в строке равно количеству отрицательных. Поэтому, если уравнение удовлетворялось до сдвига, то оно будет удовлетворяться и после сдвига, так как его левая часть не изменится. То же можно сказать и о каждом вертикальном уравнении. Теорема доказана.

Лемма 2.4. Если матрица перевозок является базисным решением системы ограничений транспортной задачи, то не существует цикла с вершинами только в базисных клетках.

Д о к а з а т е л ь с т в о. Переменные, отвечающие базисным клеткам, выражаются через переменные свободных клеток. Поэтому, если свободные переменные равны нулю, то базисные определяются однозначно. Предположим, что утверждение леммы неверно, т. е. существует цикл с вершинами в базисных клетках. Означим произвольно этот цикл и произведем сдвиг по нему на произвольное число x. Матрица останется решением системы ограничений транспортной задачи. При этом базисные переменные меняются, а свободные нет. Этого не может быть, так как базисные переменные определяются свободными однозначно. Лемма 2.4 доказана.

Определение 2.3. Циклом пересчета для данной свободной клетки называется означенный цикл, одна из вершин которого находится в данной свободной клетке, а остальные — в базисных клетках. Цикл пересчета означивается так, что вершине в данной свободной клетке приписывается знак «+».

Теорема 2.2. Для каждой свободной клетки опорного решения системы ограничений транспортной задачи существует цикл пересчета и притом только один.

Д о к а з а т е л ь с т в о. Рассмотрим произвольную свободную клетку и добавим к ней $m + n - 1$ базисную клетку. Получим $m + n$ отмеченных клеток. Согласно лемме 2.1 существует цикл с вершинами в этих отмеченных клетках. Поскольку по лемме 2.4 нет цикла с вершинами в базисных клетках, то этот цикл должен иметь вершину в данной свободной клетке, т. е. существование цикла пересчета доказано. Если бы существовало два различных цикла пересчета для данной свободной клетки, то из них можно было бы скомбинировать цикл с вершинами только в базисных клетках, а это невозможно. Теорема доказана.

Цикл пересчета является инструментом с помощью, которого производится переход от одного опорного решения к другому. Действительно, произведем сдвиг по циклу пересчета для некоторой свободной клетки на число x равное наименьшей из перевозок, стоящих в клетках с вершинами, имеющими знак “-”. При этом, по крайней мере, одна из прежних базисных переменных имеет значение 0 и мы можем перевести ее в число свободных переменных, сделав вместо нее базисной ту переменную, которая была свободной. Решение, полученное в результате сдвига по циклу пересчета на выбранное число x, по-прежнему будет опорным, но отвечающим уже другому базисному виду системы ограничений.

Пример 2.2. Преобразуем первое опорное решение примера 2.1, вводя в качестве новой базисной переменной x_{12}.

Составляем цикл пересчета для клетки x_{12}. Минимальное из чисел, стоящих в клетках с отрицательными вершинами, равно 10. Производя сдвиг по циклу на число 10, получим новое опорное решение.

Запасы	Потребности		
	40	30	10
10	10 −	+	
20	20		
50	10 +	− 30	10

Запасы	Потребности		
	40	30	10
10		10	
20	20		
50	20	20	10

2.4. Нахождение коэффициентов при свободных переменных в базисном виде системы ограничений транспортной задачи

Пусть система ограничений транспортной задачи приведена к базисному виду. Базисная переменная x_{kl} выражается через свободные переменные, причем коэффициенты этого выражения определяет следующая

Теорема 2.3. Коэффициент при свободной переменной x_{ij} в выражении для базисной переменной x_{kl} может принимать лишь три возможных значения 0, 1 и -1. При этом

а) этот коэффициент равен 0, если в клетке переменной x_{kl} нет вершины цикла пересчета для переменной x_{ij};

б) этот коэффициент равен 1, если в клетке переменной x_{kl} положительная вершина цикла пересчета для переменной x_{ij};

в) этот коэффициент равен -1, если в клетке переменной x_{kl} отрицательная вершина цикла пересчета для переменной x_{ij};

Доказательство. Пусть дан некоторый базисный вид и соответствующее базисное решение системы ограничений транспортной задачи. Рассмотрим базисную переменную x_{kl} и свободную переменную x_{ij}. В базисном решении x_{ij}=0. Построим цикл пересчета для клетки x_{ij} и произведем сдвиг рассматриваемого решения на некоторое число x_0. Получим новое решение системы ограничений транспортной задачи, в котором изменилась лишь одна свободная переменная x_{ij}. Поэтому базисная переменная получает новое значение $x_{kl}^{(1)} = x_{kl} + qx_0$, где q — является коэффициентом, с которым переменная x_{ij} входит в выражение для x_{kl}. С другой стороны, это новое решение получено сдвигом по означенному циклу. Поэтому $q = 0$, если в клетке x_{kl} нет вершины цикла пересчета для клетки x_{ij}; $q = 1$, если в клетке x_{kl} — положительная вершина и $q = -1$, если в клетке x_{kl} — отрицательная вершина. Теорема доказана.

2.5. Выражение целевой функции транспортной задачи через свободные переменные

В выражении целевой функции транспортной задачи $z = \sum_{i,j} c_{ij}x_{ij}$ суммирование распространяется на все клетки. Если система ограничений приведена к базисному виду, то базисные переменные можно из этого выражения исключить. Получим выражение

$$z = \gamma_0 + \sum_{\text{св. пер.}} \gamma_{ij}x_{ij},$$

где суммирование производится по свободным клеткам. Для дальнейшего важно знать коэффициенты γ_{ij}. Назовем *алгебраической суммой стоимостей по означенному циклу* в таблице данных транспортной задачи разность между суммой стоимостей в клетках с положительными вершинами и суммой стоимостей в клетках с отрицательными вершинами этого означенного цикла.

Теорема 2.4. Для любого базисного вида системы ограничений транспортной задачи коэффициент γ_{ij} в выражении через свободные переменные целевой функции этой задачи равен алгебраической сумме стоимостей по циклу пересчета для свободной клетки x_{ij}.

Доказательство. Зафиксируем некоторую свободную переменную x_{ij} и подсчитаем коэффициент γ_{ij} для нее. В первоначальное выражение для z переменная x_{ij} входит непосредственно с коэффициентом c_{ij}, а также посредством базисных переменных. Если учесть теорему 2.3, то становиться очевидным, что каждая базисная переменная x_{kl}, в клетке которой есть вершина цикла пересчета для переменной x_{ij}, вносит в коэффициент γ_{ij} слагаемое $\pm c_{kl}$. Причем, если в клетке x_{kl} находится положительная вершина, то это слагаемое равно c_{kl}, а в случае отрицательной вершины —

$(-c_{kl})$. Теорема 2.4 доказана.

2.6. Распределительный метод

Этот метод, по сути, является симплекс-методом для транспортной задачи. При этом вычисления организуются без симплекс-таблиц, а используются циклы пересчета для всевозможных свободных клеток опорных решений. Пусть имеется первое опорное решение. В нем свободные переменные равны нулю. Мы переходим к другому опорному решению, увеличивая от нуля некоторую свободную переменную. Поскольку мы решаем задачу на минимум, то для уменьшения целевой функции нужно выбрать для увеличения от нуля ту свободную переменную, для которой $\gamma_{ij} < 0$. Увеличение выбранной переменной x_{ij} нельзя производить неограниченно, поскольку решение может стать недопустимым. Граница увеличения x_{ij} определяется способом перехода к следующему опорному решению. Этот переход мы производим с помощью цикла пересчета для клетки x_{ij}. Для того, чтобы сдвиг по циклу не привел к недопустимому решению, нужно следить за тем, чтобы базисные переменные не стали отрицательными. Для этого просматривают клетки цикла пересчета с отрицательными вершинами и выбирают среди них ту, в которой стоит минимальная величина перевозки. Эта клетка и определяет базисную переменную, которую мы выведем из числа базисных, вводя вместо нее переменную x_{ij}. Порядок работы по распределительному методу можно описать следующим образом:

1. Находится первое опорное решение одним из рассмотренных выше способов.
2. Для каждой свободной клетки строим цикл пересчета и определяем коэффициент γ_{ij} как алгебраическую сумму стоимостей по циклу пересчета. Если все коэффициенты $\gamma_{ij} \geq 0$, то задача решена и найденное опорное решение является точкой минимума задачи. В противном случае переходим к пункту 3.
3. Выбираем свободную клетку с отрицательным значением γ_{ij} и рассматриваем величины перевозок в клетках с отрицательными вершинами цикла пересчета для x_{ij}. Из этих перевозок выбираем наименьшую, которую обозначим через x.
4. Производим сдвиг по циклу пересчета для клетки, выбранной в пункте 3 на число x. Получаем новое опорное решение, на котором значение целевой функции будет меньше, чем на старом.
5. Переходим к пункту 2, то есть снова подсчитываем коэффициенты γ_{ij} для новых свободных клеток. Описанные шаги производятся до тех пор, пока на очередном шаге все коэффициенты γ_{ij} не станут неотрицательными.

2.7. Метод потенциалов

В распределительном методе наиболее трудоемкой частью является подсчет коэффициентов γ_{ij} для каждой свободной клетки. Существует метод, позволяющий подсчитывать эти коэффициенты без составления циклов пересчета. Этот метод называется методом потенциалов.

Каждому пункту отправления (базе) A_i сопоставим величину u_i (потенциал пункта отправления), каждому потребителю B_j сопоставим потенциал v_j. Будем находить эти потенциалы из того условия, что *для каждой базисной клетки* x_{kl} *сумма потенциалов равна стоимости в этой клетке*

$$u_k + v_l = c_{kl}.$$

Поскольку базисных клеток $m+n-1$, то мы имеем систему $m+n-1$ уравнений с $m+n$ неизвестными. Эта система имеет бесконечно много решений. Если положить один из потенциалов равным нулю, то мы легко найдем некоторое решение этой системы уравнений.

Теорема 2.5. Для любой свободной клетки x_{ij} алгебраическая сумма стоимостей по циклу пересчета γ_{ij} равна разности между стоимостью в этой клетке и суммой соответствующих потенциалов

$$\gamma_{ij} = c_{ij} - (u_i + v_j)$$

Д о к а з а т е л ь с т в о. Будем обходить цикл пересчета, выйдя из свободной клетки x_{ij}, например, по столбцу. Сначала мы попадем в базисную клетку x_{kj}, расположенную в k-ой строке и в том же j-ом столбце, что и клетка x_{ij}. Затем, двигаясь по k-ой строке, мы перейдем из клетки x_{kj} в базисную клетку x_{kl}, расположенную в l-ом столбце и в той же k-ой строке, что и клетка x_{kj}, и т. д. Наконец, выйдя из некоторой базисной клетки x_{iv}, лежащей в v-ом столбце, мы вернемся, и при том по i-ой строке, в свободную клетку x_{ij}, так как вышли из нее по столбцу. Получаем следующую последовательность клеток (стрелки указывают направление обхода)

$$x_{ij} \to x_{kj} \to x_{kl} \to x_{sl} \to \ldots \to x_{uv} \to x_{iv}.$$

Составим алгебраическую сумму стоимостей γ_{ij} по этому циклу пересчета.

$$\gamma_{ij} = c_{ij} - c_{kj} - c_{kl} - c_{sl} + \ldots + c_{uv} - c_{iv}.$$

Все клетки, кроме x_{ij}, в этом цикле являются базисными. Поэтому все стоимости, кроме c_{ij}, равны суммам соответствующих потенциалов и $\gamma_{ij} = c_{ij} - (u_k + v_j) + (u_k + v_l) - (u_s + v_l) + \ldots + (u_u + v_v) - (u_i + v_v)$.

После раскрытия скобок все потенциалы, кроме u_i и v_j уничтожаются и $\gamma_{ij} = c_{ij} - (u_i + v_j)$, что и требовалось доказать.

Опираясь на порядок работы по распределительному методу, сформулируем правила работы по методу потенциалов.

1. Нахождение первого опорного решения.
2. Нахождение потенциалов пунктов отправления и назначения.
3. Нахождение коэффициентов $\gamma_{ij} = c_{ij} - (u_i + v_j)$ для свободных клеток. Если все $\gamma_{ij} \geq 0$, то данное опорное решение является оптимальным и задача решена. В противном случае переходим к пункту 4).
4. Выбор свободной клетки с $\gamma_{ij} < 0$. Построение цикла пересчета для выбранной клетки. Сдвиг по этому циклу опорного решения на величину минимальной перевозки среди тех, что стоят в клетках с отрицательными вершинами цикла пересчета. Получение нового опорного решения. Переход к пункту 2.
5. Операции, указанные в пунктах 1) — 4) повторяются до тех пор, пока величины γ_{ij} для всех свободных клеток не станут неотрицательными. Соответствующее опорное решение будет точкой минимума транспортной задачи.

Пример 2.3. Решить следующую транспортную задачу, находя первое опорное решение методом наименьшей стоимости.

		v_1	v_2	v_3
	Запасы	Потребности		
		40	30	10
u_1	10	5	7	1 10
u_2	20	2 +	1 − 20	4
u_3	50	6 40 −	3 10 +	2 0

Составляем и решаем систему уравнений для потенциалов по базисным клеткам:

$$\begin{cases} u_1 + v_3 = 1, \\ u_2 + v_2 = 1, \\ u_3 + v_1 = 6, \\ u_3 + v_2 = 3, \\ u_3 + v_3 = 2; \end{cases} \quad \begin{cases} u_1 = -1, \\ u_2 = -2, \\ u_3 = 0; \end{cases} \quad \begin{cases} v_1 = 6, \\ v_2 = 3, \\ v_3 = 2; \end{cases}$$

Здесь мы положили $u_3 = 0$ и решили систему. Для каждой свободной клетки вычисляем γ_{ij}.

$$\gamma_{11} = 5 - (-1 + 6) = 0;$$
$$\gamma_{12} = 7 - (-1 + 3) = 5;$$
$$\gamma_{21} = 2 - (-2 + 6) = -2;$$
$$\gamma_{23} = 4 - (-2 + 2) = 4.$$

Имеется одна свободная клетка x_{21} с отрицательным $\gamma_{21} = -2$. Строим цикл пересчета для этой клетки. Наименьшее число, стоящее в клетках с отрицательными вершинами, равно 20. Производим сдвиг на 20 по этому циклу пересчета. Получаем новое опорное решение.

		v_1	v_2	v_3
		Потребности		
		40	30	10
u_1	10	$\boxed{5}$	$\boxed{7}$	$\boxed{1}$ 10
u_2	20	$\boxed{2}$ 20	$\boxed{1}$	$\boxed{4}$
u_3	50	$\boxed{6}$ 20	$\boxed{3}$ 30	$\boxed{2}$ 0

Снова составляем систему уравнений для потенциалов:

$$\begin{cases} u_1 + v_3 = 1, \\ u_2 + v_1 = 2, \\ u_3 + v_1 = 6, \\ u_3 + v_2 = 3, \\ u_3 + v_3 = 2; \end{cases} \quad \begin{cases} u_1 = -1, \\ u_2 = -4, \\ u_3 = 0; \end{cases} \quad \begin{cases} v_1 = 6, \\ v_2 = 3, \\ v_3 = 2; \end{cases}$$

Снова подсчитываем коэффициенты γ_{ij} для свободных клеток:

$$\gamma_{11} = 5 - (-1 + 6) = 0;$$
$$\gamma_{12} = 7 - (-1 + 3) = 5;$$
$$\gamma_{21} = 1 - (-4 + 3) = 2;$$
$$\gamma_{23} = 4 - (-4 + 2) = 6.$$

Здесь все коэффициенты неотрицательны. Это означает, что последнее опорное решение является оптимальным. При этом $z_{min} = 1 \cdot 10 + 2 \cdot 20 + 6 \cdot 20 + 3 \cdot 30 + 2 \cdot 0 = 260$. Координаты точки минимума: $_{11} = 0;_{12} = 0;_{13} = 10;_{21} = 20;_{22} = 0;_{23} = 0;_{31} = 20;_{32} = 30;_{33} = 0$.

2.8. Об одном свойстве закрытой транспортной задачи

Матрицу стоимостей транспортной задачи можно определенным образом изменить, не меняя оптимального плана перевозок. Справедлива следующая

Теорема 2.6. Если к элементам произвольного столбца или произвольной строки матрицы стоимостей транспортной задачи прибавить одно и то же произвольное число, то целевая функция задачи изменится на постоянное слагаемое. Таким образом, получаются две транспортные задачи с разными матрицами стоимостей, но одинаковыми оптимальными планами.

Д о к а з а т е л ь с т в о. Пусть дана транспортная задача с матрицей стоимостей $C = (c_{ij})$. Рассмотрим транспортную задачу с другой матрицей стоимостей $C_1 = (c_{ij} + \alpha_i + \beta_j)$, где $\alpha_i (i = 1, ..., n)$ и$\beta_j (j = 1, ..., m)$ - произвольные числа. Запишем целевую функцию второй задачи

$$z_1 {=} \sum_{i,j} (c_{ij} + \alpha_i + \beta_j) x_{ij} = \sum_{i,j} c_{ij} x_{ij} + \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{m} x_{ij} + \sum_{j=1}^{m} \beta_j \sum_{i=1}^{n} x_{ij}.$$

Учитывая горизонтальные и вертикальные уравнения системы ограничений, получим

$$z_1 {=} \sum_{i,j} c_{ij} x_{ij} + \sum_{i=1}^{n} \alpha_i a_i + \sum_{j=1}^{m} \beta_j b_j = z + \sum_{i=1}^{n} \alpha_i a_i + \sum_{j=1}^{m} \beta_j b_j,$$

что и доказывает утверждение теоремы. Очевидно, что оптимальные планы перевозок задач с матрицами перевозок C и C_1 совпадают.

Сформулированное свойство транспортной задачи позволяет использовать преобразование матрицы перевозок для нахождения оптимального опорного решения. Можно так изменить элементы матрицы стоимостей, чтобы они, оставаясь неотрицательными, стали равными нулю в базисных клетках некоторого опорного плана. На этом основан так называемый венгерский метод решения транспортной задачи (см. [15], с.180).

2.9. Открытые транспортные задачи

Выше мы предполагали, что общие запасы A равны общим потребностям B. В случае $A \neq B$ формулировка транспортной задачи различна в случаях $A < B$ и $A > B$. В случае$A < B$ мы можем обеспечить вывоз груза из пунктов отправления, однако удовлетворить все пункты назначения (потребителей) мы не можем. При математической формулировке горизонтальные уравнения остаются без изменений, а вертикальные уравнения превратятся в неравенства типа "$\leq$". Транспортная задача будет формулироваться так:

$$z = \sum_{(i,j)} c_{ij} x_{ij} \to \min;$$

$$\begin{cases} lx_{11} + x_{12} + ... + x_{1\,n} = a_1, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ x_{m\,1} + x_{m\,2} + ... + x_{m\,n} = a_m; \end{cases}$$

$$\begin{cases} lx_{11} + x_{21} + ... + x_{m\,1} \leq b_1, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ x_{1n} + x_{2\,n} + ... + x_{m\,n} \leq b_n; \end{cases}$$

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

В случае $A > B$ горизонтальные уравнения превращаются в неравенства, то есть система ограничений состоит из первой группы неравенств и второй группы равенств. Если в условиях открытой транспортной задачи все пункты отправления и назначения равноправны, то такая задача легко сводится к закрытой транспортной задаче. Равноправие пунктов понимается в том смысле, что нет потребителей, которых необходимо обязательно удовлетворить (при $A < B$), и нет баз, которые необходимо обязательно освободить от груза (при $A > B$).

Для сведения нашей задачи к закрытой транспортной задаче в случае $A < B$ введем фиктивный пункт отправления A_{m+1} с запасами

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

Рассмотрим закрытую задачу с пунктами отправления A_1, A_2, ... A_m, A_{m+1} и с теми же пунктами назначения B_1, B_2, ..., B_n. В клетках, отвечающих фиктивному пункту A_{m+1}, стоимости будем считать равными 0. Решение этой закрытой задачи даст решение нашей исходной задачи. Перевозки, отвечающие фиктивному пункту, имеют смысл недопоставок груза соответствующим потребителям. В случае $A > B$ следует ввести фиктивный пункт назначения B_{n+1} с потребностями $B_{n+1} = A - B$ и положить стоимости в клетках, связанных с этим пунктом, равными 0. Фиктивные перевозки будут означать количество груза, которое останется на соответствующих базах.

2.10. Открытые транспортные задачи с неравноправными пунктами. Блокирование клеток

В некоторых случаях в формулировку транспортной задачи входят также условия обязательного удовлетворения некоторых выделенных пунктов потребления или обязательного освобождения от груза некоторых баз. Первые условия возникают при $A < B$, а вторые — при $A > B$. В математической формулировке подобные условия означают, что в подсистеме неравенств некоторые неравенства должны быть на самом деле равенствами. Такие задачи с выделенными пунктами также могут быть сведены к закрытым задачам. Соответствующий прием сведения называется *блокированием клеток*.

Поскольку есть выделенные пункты, то в случае $A < B$ эти пункты надо удовлетворить за счет реальных баз, а в случае $A > B$ требуется вывозить груз из выделенных баз в реальные пункты потребления. Таким образом, в задачах с выделенными пунктами запрещается рассматривать решения, в которых отличны от нуля некоторые переменные, связанные с фиктивными пунктами. Исключение этих планов и осуществляется приемом, который называется блокированием клеток. При этом в закрытой транспортной задаче, построенной так, как описано выше, стоимости, отвечающие выделенным пунктам и фиктивным пунктам, полагаются равными достаточно большому числу $M > 0$. В остальных клетках, связанных с фиктивными пунктами, стоимости остаются равными нулю. Из-за большой стоимости выделенные клетки как бы блокируются, то есть в оптимальном решении закрытой задачи соответствующие перевозки получаются равными нулю. При этом в случае $A < B$ выделенные пункты назначения удовлетворяются за счет реальных баз, а в случае $A > B$ груз из выделенных баз вывозится целиком в реальные пункты потребления.

Пример 2.4. Свести к закрытой транспортной задаче следующую открытую задачу при дополнительном условии обязательного удовлетворения пункта B_1. Тот факт, что пункт B_1 выделен, условно записывается в виде звездочки над столбцом B_1 (рис. 2.2). Эта задача является открытой, так как общие потребности B=120 больше общих запасов A=100. Вводим фиктивный пункт отправления A_4 с запасами a_4=120 - 100 = 20. В клетке, соответствующей пунктам B_1 и A_4, стоимость полагаем M= 100. Остальные стоимости в фиктивных клетках остаются равными нулю (таких клеток всего одна). Получаем обычную закрытую транспортную задачу, которую можно решить методом потенциалов.

Запасы	Потребности	
	70	50
30	1	2
20	4	6
50	7	1

Запасы	Потребности	
	70	50
30	1	2
20	4	6
50	7	1
20	100	0

Рис. 2.2. Сведение открытой транспортной задачи к закрытой

Наконец, если строго выделенных пунктов нет, но недополучение груза пунктами назначения (при $A < B$) или остаток на базах груза (при $A > B$) приводит к определенным убыткам, то стоимости в фиктивных пунктах полагают равными этим убыткам.

Пример 2.5. Рассмотрим транспортную задачу предыдущего примера при другом дополнительном условии. Предположим, что недополучение единицы груза пунктом B_1 приводит к убытку в 8 единиц стоимости, а для пункта B_2 этот убыток равен 10 единицам стоимости. Как и раньше вводим фиктивный пункт отправления A_4 с запасами a_4=20. В клетках фиктивного пункта стоимости полагаем равными соответственно c_{41}=8, c_{42}=10. Получаем обычную закрытую транспортную задачу

Запасы	Потребности	
	70	50
30	1	2
20	4	6
50	7	1
20	8	10

Отметим, что прием *блокирования клеток* применяется для решения транспортных задач (закрытых или открытых) при дополнительном условии невозможности перевозок по некоторым коммуникациям. В некоторых случаях перевозки груза из пункта A_i в пункт B_j не могут быть осуществлены. При определении оптимальных планов таких задач полагают, что стоимость перевозки единицы груза, отвечающая клетке x_{ij} является достаточно большой величиной M. После этого известными методами находится решение новой транспортной задачи.

2.11. О других типах транспортных задач

В реальных условиях формулировка транспортной задачи может осложняться дополнительными ограничениями, которые меняют тип задачи и требуют значительной модификации метода ее решения.

Например, коммуникации, связывающие пункты отправления и назначения могут иметь ограниченные пропускные способности, т. е. по этим коммуникациям за рассматриваемое время можно перевезти ограниченное количество груза. Возникает транспортная задача с ограничениями по пропускной способности, в которой добавляются дополнительные ограничения вида

$$0 \le x_{ij} \le d_{ij},$$

где d_{ij} — пропускная способность коммуникации из A_i в B_j. Такая задача значительно отличается от обычной транспортной задачи. В частности, она не всегда имеет решение. Если сумма пропускных способностей коммуникаций, идущих в данный пункт назначения меньше потребности в этом пункте, то задача неразрешима. Существует модификация метода потенциалов, которая позволяет либо решить задачу с ограничениями по пропускной способности, либо установить ее неразрешимость.

Груз может обрабатываться на промежуточных станциях (подвергаться перевалке). Это ведет к дополнительным ограничениям, связанным с временем, которое необходимо для обработки груза. При этом транспортная задача значительно усложняется.

В приложениях часто возникает *транспортная задача по критерию времени*, в которой оптимизируется не стоимость перевозок, а время, необходимое на эти перевозки. Такая задача вообще не является задачей линейного программирования. В ней вместо матрицы стоимостей задается матрица $T{=}t_{ij})$, где $T{=}t_{ij}$ — время, необходимое для перевозки груза из пункта A_i в пункт B_j. Предполагается, что оно не зависит от объема перевозимой продукции, но в случае нулевой перевозки считается равной нулю. Целевая функция задачи имеет вид $z = \max\limits_{i,j} t_{ij}$. Очевидно, что за время z будут осуществлены все перевозки по заданному допустимому плану. Требуется выбрать такой допустимый план, при котором время осуществления всех перевозок будет минимальным. Решение этой задачи может быть сведено к последовательному решению нескольких задач линейного программирования.

К модели транспортной задачи иногда приводят задачи, по своему содержанию никак не связанные с транспортом, с планированием перевозок. В таких случаях говорят, что задача может быть сформулирована в терминах транспортной задачи. Примером такой задачи может служить задача о назначениях, о которой пойдет речь в главе о дискретном программировании. Поскольку для транспортной задачи имеются эффективные алгоритмы решения, всегда полезно, если возможно, сформулировать данную задачу в терминах транспортной задачи.

Наконец многие практические задачи приводят к модели, по отношению к которой обычная транспортная задача является частным случаем. Как и в транспортной задаче, здесь ограничения делятся на две группы, такие, что каждая переменная входит явно лишь в одно ограничение каждой группы, то есть сохраняется одна из основных особенностей транспортной задачи.

Однако в отличие от нее в одной из групп ограничений допускаются произвольные, не обязательно единичные, ненулевые коэффициенты при входящих в ограничения переменных. Такие задачи называются *распределительными задачами.* Для них разработаны различные алгоритмы решения, хотя и более сложные, чем для транспортной задачи, однако значительно более простые, чем общие методы линейного программирования.

Отметим, что нами рассматривалась транспортная задача в матричной постановке. Существуют сетевые постановки транспортной задачи. Подробно с задачами транспортного типа, в том числе, и в сетевой постановке можно ознакомиться по книге [15].

Контрольные вопросы для самостоятельного решения

1. Как формулируется транспортная задача? Что такое матрица перевозок? Как выглядит математическая модель закрытой транспортной задачи?
2. Как записать транспортную задачу в форме таблицы данных?
3. Нахождение первого опорного решения системы ограничений транспортной задачи. В чем заключаются метод северо-западного угла и метод наименьшей стоимости?
4. Что называют циклом в матрице? Какими комбинаторными свойствами обладают циклы?
5. Означенный цикл. Что называют сдвигом по означенному циклу в матрице перевозок? Каким основным свойством обладает этот сдвиг?
6. Что называется циклом пересчета для данной свободной клетки?

7. Как находятся коэффициенты при свободных переменных в базисном виде системы ограничений транспортной задачи?
8. Как находится выражение целевой функции транспортной задачи через свободные переменные для произвольного базисного вида системы ограничений?
9. В чем заключается распределительный метод решения закрытой транспортной задачи?
10. Опишите порядок работы по методу потенциалов.
11. При каких преобразованиях матрицы стоимостей транспортной задачи оптимальный план перевозок не меняется?
12. Открытые транспортные задачи. Как сводится открытая транспортная задача с равноправными пунктами к закрытой задаче?
13. В каких случаях при решении открытой транспортной задачи используется прием блокирования клеток?
14. Какие другие типы транспортных и подобных им задач Вы знаете?

Составить математические модели транспортных задач 2.1 — 2.4.

2.1. Четыре предприятия данного экономического района для производства продукции используют три вида сырья. Потребности в сырье каждого из предприятий соответственно равны 120, 50, 190 и 110 ед. Сырье сосредоточено в трех местах его получения, а запасы его соответственно равны 160, 140 и 170 ед. На каждое из предприятий сырье может завозиться из любого пункта его получения. Тарифы перевозок являются известными величинами и задаются матрицей стоимостей

$$\begin{pmatrix} 7 & 8 & 1 & 2 \\ 4 & 5 & 9 & 8 \\ 9 & 2 & 3 & 6 \end{pmatrix}$$

Требуется составить такой план перевозок, при котором общая стоимость перевозок является минимальной.

2.2. На трех складах оптовой базы сосредоточен однородный груз в количествах 180, 40 и 80 ед. Этот груз необходимо перевести в четыре магазина. Каждый из магазинов должен получить соответственно 120, 40, 60, и 80 ед. груза. Тарифы перевозок единицы груза из каждого из складов во все магазины задаются матрицей

$$\begin{pmatrix} 2 & 3 & 4 & 3 \\ 5 & 3 & 1 & 2 \\ 2 & 1 & 4 & 2 \end{pmatrix}$$

Требуется составить такой план перевозок, при котором общая стоимость перевозок является минимальной.

2.3. Производственное объединение имеет в своем составе три филиала, которые производят однородную продукцию соответственно в количествах 50, 30 и 10 ед. Эту продукцию получают четыре потребителя, расположенные в разных местах. Их потребности соответственно равны 30, 30, 10 и 20 ед. Тарифы перевозок единицы продукции от каждого из филиалов соответствующим потребителям задаются матрицей

$$\begin{pmatrix} 1 & 2 & 4 & 1 \\ 2 & 3 & 1 & 5 \\ 3 & 2 & 4 & 4 \end{pmatrix}$$

Требуется составить такой план прикрепления получателей продукции к ее поставщикам, чтобы общая стоимость перевозок была минимальной.

2.4. Три предприятия производственного объединения производят однородную продукцию в количествах равных соответственно 180, 350 и 20 ед. Эта продукция должна быть поставлена пяти потребителям в количествах равных соответственно

110, 90, 120, 80 и 150 ед. Затраты, связанные с производством и доставкой единицы продукции, задаются матрицей

$$\begin{pmatrix} 7 & 12 & 4 & 6 & 5 \\ 1 & 8 & 6 & 5 & 3 \\ 6 & 13 & 8 & 7 & 4 \end{pmatrix}$$

Требуется составить такой план прикрепления потребителей продукции к ее поставщикам, чтобы общие затраты были минимальными.

2.5. — 2.8 *Записать формулировки транспортных задач* 2.1 — 2.4 *с помощью таблицы данных и найти для каждой из задач первое опорное решение методами северо-западного угла и наименьшей стоимости. Результаты сравнить между собой.*

2.9. — 2.11 *Найти оптимальное решение задач* 2.1—2.3 *распределительным методом.*

Задачи

2.12. — 2.22 *решить методом потенциалов. В этих задачах задаются векторы запасов* $\vec{a} = (a_1, a_2, \ldots, a_m)$ *и потребностей* $\vec{b} = (b_1, b_2, \ldots, b_n)$, *а также матрица стоимостей* C.

2.13.
$\vec{a} = (4, 6, 10, 10)$;
$\vec{b} = (7, 7, 7, 7, 2)$;

$$C = \begin{pmatrix} 16 & 30 & 17 & 10 & 16 \\ 30 & 27 & 26 & 9 & 23 \\ 13 & 4 & 22 & 3 & 1 \\ 3 & 1 & 5 & 4 & 24 \end{pmatrix}$$

Ответ: $z_{min} = 191$.

2.14.
$\vec{a} = (20, 20, 20, 20)$;
$\vec{b} = (19, 19, 19, 19, 4)$;

$$C = \begin{pmatrix} 15 & 1 & 22 & 19 & 1 \\ 21 & 18 & 11 & 4 & 3 \\ 26 & 29 & 23 & 26 & 24 \\ 21 & 10 & 3 & 19 & 27 \end{pmatrix}$$

Ответ: $z_{min} = 684$.

2.15.
$\vec{a} = (13, 17, 17, 13)$;
$\vec{b} = (12, 12, 12, 12, 12)$;

$$C = \begin{pmatrix} 20 & 26 & 24 & 26 & 29 \\ 15 & 20 & 29 & 26 & 23 \\ 4 & 10 & 27 & 30 & 7 \\ 9 & 16 & 29 & 20 & 3 \end{pmatrix}$$

Ответ: $z_{min} = 868$.

2.16.
$\vec{a} = (18, 12, 17, 13)$;
$\vec{b} = (8, 8, 8, 8, 28)$;

$$C = \begin{pmatrix} 21 & 22 & 2 & 13 & 7 \\ 27 & 10 & 4 & 24 & 9 \\ 3 & 16 & 25 & 5 & 4 \\ 28 & 11 & 17 & 10 & 29 \end{pmatrix}$$

Ответ: $z_{min} = 392$.

2.17.
$\vec{a} = (15, 15, 15, 15)$;
$\vec{b} = (11, 11, 11, 11, 16)$;

$$C = \begin{pmatrix} 17 & 20 & 29 & 26 & 25 \\ 3 & 4 & 5 & 15 & 24 \\ 19 & 2 & 22 & 4 & 13 \\ 20 & 27 & 1 & 17 & 19 \end{pmatrix}$$

Ответ: $z_{min} = 542$.

2.18.
$\vec{a} = (15, 15, 19, 11)$;
$\vec{b} = (9, 24, 9, 9, 9)$;

$$C = \begin{pmatrix} 10 & 17 & 9 & 20 & 30 \\ 13 & 4 & 24 & 26 & 26 \\ 22 & 24 & 30 & 27 & 29 \\ 25 & 12 & 11 & 24 & 23 \end{pmatrix}$$

Ответ: $z_{min} = 859$.

2.19.
$\vec{a} = (21, 19, 15, 25)$;
$\vec{b} = (15, 15, 15, 15, 20)$;

$$C = \begin{pmatrix} 30 & 24 & 11 & 12 & 25 \\ 26 & 4 & 29 & 20 & 24 \\ 27 & 14 & 14 & 10 & 18 \\ 6 & 14 & 28 & 8 & 2 \end{pmatrix}$$

Ответ: $z_{min} = 693$.

2.20.
$\vec{a} = (9, 11, 14, 16)$;
$\vec{b} = (8, 9, 13, 8, 12)$;

$$C = \begin{pmatrix} 5 & 15 & 3 & 6 & 10 \\ 23 & 8 & 13 & 27 & 12 \\ 30 & 1 & 5 & 24 & 25 \\ 8 & 26 & 7 & 28 & 9 \end{pmatrix}$$

Ответ: $z_{min} = 339$.

2.21.
$\vec{a} = (22, 13, 17, 18)$;
$\vec{b} = (7, 7, 7, 7, 42)$;

$$C = \begin{pmatrix} 9 & 17 & 29 & 28 & 8 \\ 13 & 21 & 27 & 16 & 29 \\ 20 & 30 & 24 & 7 & 26 \\ 11 & 19 & 30 & 6 & 2 \end{pmatrix}$$

Ответ: $z_{min} = 726$.

2.22.
$\vec{a} = (16, 15, 14, 15)$;
$\vec{b} = (6, 6, 13, 20, 15)$;

$$C = \begin{pmatrix} 30 & 2 & 5 & 6 & 15 \\ 5 & 29 & 9 & 5 & 7 \\ 16 & 24 & 14 & 6 & 26 \\ 13 & 28 & 4 & 25 & 8 \end{pmatrix}$$

Ответ: $z_{min} = 329$.

3. ЭЛЕМЕНТЫ ТЕОРИИ ДВОЙСТВЕННОСТИ

Для любой задачи линейного программирования можно сформулировать некоторую задачу, называемую двойственной. Решение каждой из этой пары задач часто автоматически приводит к решению другой задачи, то есть количество решенных задач увеличивается вдвое. В некоторых случаях одна из двойственных задач проще и ее решать удобнее. Но важнее всего то, что во многих приложениях линейного программирования требуется решать обе двойственные задачи. Поэтому изучение взаимосвязи этих задач очень полезно. Мы кратко опишем эту взаимосвязь, не останавливаясь подробно на обоснованиях.

3.1. Симметрично-двойственные задачи

Рассмотрим задачу линейного программирования в стандартной форме.

$$z = c_0 + c_1x_1 + \ldots + c_nx_n \to max;$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leq b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \leq b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leq b_m; \end{cases}$$

$$x_1, x_2, \ldots, x_n \geq 0.$$

Назовем эту задачу задачей 1. Построим по задаче 1 задачу 2, которая имеет следующий вид:

$$f = c_0 + b_1y_1 + \ldots + b_my_m \to min;$$

$$\begin{cases} a_{11}y_1 + a_{21}y_2 + \ldots + a_{m1}y_m \geq c_1, \\ a_{12}y_1 + a_{22}y_2 + \ldots + a_{m2}y_m \geq c_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}y_1 + a_{2n}y_2 + \ldots + a_{mn}y_m \geq c_n; \end{cases}$$

$$y_1, y_2, \ldots, y_m \geq 0.$$

Задача 2 построена по задаче 1 с помощью следующих правил:

1. Число переменных задачи 2 равно числу ограничений задачи 1.
2. Матрица системы ограничений задачи 2 является транспонированной по отношению к соответствующей матрице задачи 1, то есть, получена из нее заменой строк столбцами с сохранением порядка.
3. Свободный член целевой функции f задачи 2 равен свободному члену в выражении целевой функции z задачи 1, а коэффициенты при переменных в выражении для f равны правым частям неравенств системы ограничений задачи 1.
4. Неравенства в системе ограничений задачи 2 имеют смысл «$\geq$ », а правые части этих неравенств равны коэффициентам при переменных в целевой функции задачи 1.
5. Задача 2 является задачей на минимум, в то время как задача 1 является задачей на максимум.

Отметим, что если рассматривать задачу 2 как исходную для построения двойственной задачи по сформулированным правилам, то двойственная задача будет равносильна задаче 1. Разумеется, при этом задачу 2 нужно сформулировать в стандартной форме.

Пример 3.1. Исходная задача 1:

$$z = 1 + 2x_1 - 3x_2 \to max;$$

$$y_1, y_2, y_3 \begin{cases} 2x_1 + 4x_2 \le 10, \\ 4x_1 - 2x_2 \le 1, \\ 5x_1 + 6x_2 \le 12; \end{cases}$$

$$x_1, x_2 \ge 0.$$

Двойственная задача 2:

$$f = 1 + 10y_1 + y_2 + 12y_3 \to min;$$

$$\begin{cases} 2y_1 + 4y_2 + 5y_3 \ge 2, \\ 4y_1 - 2y_2 + 6y_3 \ge -3; \end{cases}$$

Запишем задачу 2 в стандартной форме:

$$f_1 = -1 - 10y_1 - y_2 - 12y_3 \to max;$$

$$\begin{cases} -2y_1 - 4y_2 - 5y_3 \le -2, \\ -4y_1 + 2y_2 - 6y_3 \le 3; \end{cases}$$

$$y_1, y_2, y_3 \le 0.$$

Двойственная к ней задача:

$$z_1 = -1 - 2x_1 + 3x_2 \to min;$$

$$\begin{cases} -2x_1 - 4x_2 \ge -10, \\ -4x_1 + 2x_2 \ge -1, \\ -5x_1 - 6x_2 \ge -12; \end{cases}$$

$$x_1, x_2 \ge 0.$$

Полученная задача равносильна первоначально взятой задаче и совпадает с ней, если перейти к стандартной форме. Отмеченное свойство взаимности является причиной того, что пара задач (1;2) называется парой взаимно-двойственных задач.

3.2. Экономический смысл взаимно-двойственных задач

Рассмотрим *задачу оптимального использования ресурсов* предприятием, которое производит n видов продукции: $P_1, P_2, \ldots, P_n$ и использует для этого m видов ресурсов $R_1, R_2, \ldots, R_m$. Будем считать, что ресурсы при необходимости можно продать, не изготавливая продукции. Если $b_1, b_2, \ldots, b_m$ — наличные количества ресурсов; a_{ij} — расход ресурса R_i на изготовление единицы продукции P_j, а x_j — плановое количество продукции P_j , то переменные x_1, …, x_n должны удовлетворять системе неравенств:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le b_m; \end{cases}$$

$$x_1, x_2, \ldots, x_n \ge 0.$$

Для максимизации прибыли от реализации продукции предприятие должно решить задачу на максимум для функции $z = c_1x_1 + \ldots + c_nx_n$, где $c_1, c_2, \ldots, c_n$ — прибыли от реализации единицы соответствующего вида продукции.

Пусть наше предприятие имеет возможность продать ресурсы некоторому другому предприятию или лицу. Обозначим через $y_1, \ldots, y_m$ соответствующие цены ресурсов $R_1, R_2, \ldots, R_m$. Покупатель заинтересован в понижении цен, но продавец не будет продавать ресурсы, если продажа менее выгодна, чем производство продукции. Станем на точку зрения покупателя и определим наиболее выгодные цены ресурсов с его точки зрения. При полной покупке ресурсов он должен минимизировать общую сумму выплаты за ресурсы:

$$f = b_1 y_1 + \ldots + b_m y_m \to min.$$

При этом стоимость ресурсов, идущих на изготовление единицы каждого вида продукции, не должна быть меньше, чем прибыль от ее реализации, иначе продавцу невыгодно продавать ресурсы. Поэтому цены $y_1, y_2, \ldots, y_m$ должны удовлетворять неравенствам:

$$\begin{cases} a_{11}y_1 + a_{21}y_2 + \ldots + a_{m1}y_m \geq c_1, \\ a_{12}y_1 + a_{22}y_2 + \ldots + a_{m2}y_m \geq c_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}y_1 + a_{2n}y_2 + \ldots + a_{mn}y_m \geq c_n; \end{cases}$$
$$y_1, y_2, \ldots, y_m \geq 0.$$

Таким образом, предприятие-покупатель ресурсов при назначении цен решает задачу линейного программирования двойственную по отношению к задаче, которую решает предприятие-производитель при составлении оптимального плана выпуска продукции. Величины $y_1^0, y_2^0, \ldots, y_m^0$, составляющие точку минимума для задачи покупателя, *называют двойственными ценами (оценками) ресурсов*. Они имеют и более глубокий смысл, о котором речь пойдет ниже.

3.3. Несимметрично-двойственные задачи

Рассмотрим задачу линейного программирования в общем форме. Приведем эту задачу к стандартной форме с помощью рассмотренных выше приемов. Равенства в системе ограничений заменим парами противоположных неравенств, а переменные, принимающие значения произвольного знака, заменим разностями неотрицательных переменных. К полученной задаче в стандартной форме применим сформулированные выше правила построения двойственной задачи. Для простоты рассмотрим эту ситуацию на конкретном примере.

Исходная задача:

$$z = 5 - x_1 + 3x_2 - x_3 \to min;$$
$$\begin{cases} x_1 - 2x_2 + x_3 = 2, \\ x_1 + x_2 - x_3 \geq -3; \end{cases}$$
$$x_1, x_2 \geq 0.$$

Переменная x_3 может принимать значения произвольного знака, поэтому полагаем $x_3 = x_3^{'} - x_3^{''}$; $x_3^{'}, x_3^{''} \geq 0$. Запишем задачу в стандартной форме.

$$z_1 = -5 + x_1 - 3x_2 + x_3^{'} - x_3^{''} \to max;$$
$$y_1, y_2, y_3 \begin{cases} x_1 - 2x_2 + x_3^{'} - x_3^{''} \leq 2, \\ -x_1 + 2x_2 - x_3^{'} + x_3^{''} \leq -2, \\ -x_1 + x_2 - 3x_3^{'} + 3x_3^{''} \leq 3; \end{cases}$$
$$x_1, x_2, x_3^{'}, x_3^{''} \geq 0.$$

Построим двойственную задачу:

$$f = -5 + 2y_1 - 2y_2 + 3y_3 \to min;$$
$$\begin{cases} y_1 - y_2 - y_3 \geq 1, \\ -2y_1 + 2y_2 + y_3 \geq -3, \\ y_1 - y_2 - 3y_3 \geq 1, \\ -y_1 + y_2 + 3y_3 \geq -1; \end{cases}$$

$$y_1, y_2, y_3 \geq 0.$$

Отметим, что переменные y_1 и y_2 входят в эту задачу лишь в комбинации $y_1 - y_2$, а третье и четвертое неравенства системы ограничений двойственной задачи являются взаимно противоположными и могут быть заменены равенством $y_1 - y_2 - 3y_3 = 1$. Обозначая $y_1 - y_2$ через y, можем записать двойственную задачу в виде:

$$f = -5 - 2y + 3y_3 \to min;$$

$$\begin{cases} y - y_3 \geq 1, \\ -2y + y_3 \geq -3, \quad y_3 \geq 0. \\ y - 3y_3 = 1; \end{cases}$$

Отметим, что эта же задача может быть получена из первоначальной задачи, записанной в виде задачи на максимум с неравенством типа «$<$» :

$$z_1 = -5 + x_1 - 3x_2 + x_3 \to max;$$

$$y, y_3 \begin{cases} x_1 - 2x_2 + x_3 = 2, \\ -x_1 + x_2 - 3x_3 \leq 3; \end{cases}$$

$$x_1, x_2 \geq 0,$$

с помощью изложенного выше правила составления симметрично-двойственной задачи со следующими добавлениями:

а) если в исходной задаче в системе ограничений есть равенство, то в двойственной задаче соответствующая переменная не подчинена условию неотрицательности;

б) если в исходной задаче переменная не подчинена условию неотрицательности, то в двойственной задаче соответствующее ограничение является равенством.

В общем случае, если задачу общего вида сформулировать как задачу на максимум, все неравенства в системе ограничений которой имеют смысл «$\leq$ », то двойственную задачу можно сформулировать по правилам для симметрично двойственных задач с добавлениями а) и б).

3.4. Первая и вторая теоремы двойственности

Все рассматриваемые ниже утверждения относятся к парам двойственных задач общего вида, то есть *необязательно* симметрично-двойственных.

Теорема 3.1. Для всякой пары двойственных задач если исходная задача имеет решение, то двойственная задача также имеет решение, и оптимальные значения целевых функций этих задач совпадают, то есть $z_{max} = f_{min}$. Если исходная задача не имеет решения по причине неограниченности целевой функции на области допустимых значений, то двойственная задача недопустима.

Отметим, что первая теорема двойственности не позволяет по решению одной из задач найти точку оптимума другой. С ее помощью отыскивается лишь оптимальное значение целевой функции. Связь между точкой минимума и максимума пары двойственных задач определяет следующая:

Теорема 3.2. Для того, чтобы n-мерный вектор удовлетворяющий системе ограничений исходной задачи, и m-мерный вектор $\{y_1^0, y_2^0, \ldots, y_m^0\}$ удовлетворяющий системе ограничений двойственной задачи, были соответственно точкой максимума исходной и точкой минимума двойственной задач, необходимо и достаточно, чтобы выполнялись следующие условия:

$$\begin{cases} y_1^0\left(a_{11}x_1^0 + a12x_2^0 + \ldots + a_{1n}x_n^0 - b_1\right) = 0, \\ y_2^0\left(a_{21}x_1^0 + a_{22}x_2^0 + \ldots + a_{2n}x_n^0 - b_2\right) = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m^0\left(a_{m1}x_1^0 + a_{m2}x_2^0 + \ldots + a_{mn}x_n^0 - b_m\right) = 0, \\ x_1^0\left(a_{11}y_1^0 + a_{21}y_2^0 + \ldots + a_{m1}y_m^0 - c_1\right) = 0, \\ x_2^0\left(a_{12}y_1^0 + a_{22}y_2^0 + \ldots + a_{m2}y_0^m - c_2\right) = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n^0\left(a_{1n}y_1^0 + a_{2n}y_2^0 + \ldots + a_{mn}y_m^0 - c_n\right) = 0. \end{cases}$$

Вторая теорема двойственности содержит равенства, которые называются условиями дополнительной нежесткости. Она позволяет решать следующие задачи:

1) определять оптимальный план одной из двойственных задач, если известно решение другой;
2) проверять, не решая задачи, является ли некоторая совокупность чисел оптимальным планом одной из двойственных задач.

Пример 3.2. Пусть дана задача:

$$z = x_1 + x_2 + x_3 \to max;$$

$$y_1, y_2 \begin{cases} 2x_1 + 3x_2 - x_3 \le 2, \\ x_1 - x_2 + 2x_3 \le 3; \end{cases}$$

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

Двойственную задачу

$$f = 2y_1 + 3y_2 \to min;$$

$$\begin{cases} 2y_1 + y_2 \ge 1, \\ 3y_1 - y_2 \ge 1, \\ -y_1 + 2y_2 \ge 1; \end{cases}$$

$$y_1, y_2 \ge 0.$$

можно решить графически (рис.3.1.).

Точкой минимума двойственной задачи является точка $A\left(\frac{3}{5}; \frac{4}{5}\right)$. $z_{max} = f_{min} = f(A) = 2 \cdot \frac{3}{5} + 3 \cdot \frac{4}{5} = \frac{6}{6} + \frac{12}{5} = \frac{18}{5}$. Для определения точки максимума исходной задачи составим условия дополнительной нежесткости:

$$\begin{cases} y_1^0\left(2x_1^0 + 3x_2^0 - x_3^0 - 2\right) = 0, \\ y_1^0\left(x_1^0 - x_2^0 + 2x_n^0 - 3\right) = 0, \\ x_1^0\left(2y_1^0 + y_1^0 - 1\right) = 0, \\ x_2^0\left(3y_1^0 - y_2^0 - 1\right) = 0, \\ x_3^0\left(-y_1^0 + 2y_2^0 - 1\right) = 0. \end{cases}$$

Подставим в эти равенства $y_1^0 = \frac{3}{5}; y_2^0 = \frac{4}{5}$ Учитывая, что четвертое и пятое равенства при этом выполняются при всех x_3^0 и x_2^0 , для определения x_1^0, x_2^0, x_3^0 получаем систему уравнений

$$\begin{cases} 2x_1^0 + 3x_2^0 - x_3^0 = 3, \\ x_1^0 - x_2^0 + 2x_3^0 = 3, \\ x_1^0 = 0, \end{cases}$$

которая равносильна системе

$$\begin{cases} 3x_2^0 - x_3^0 = 2, \\ -x_2^0 + 2x_3^0 = 3, \\ x_1^0 = 0, \end{cases}$$

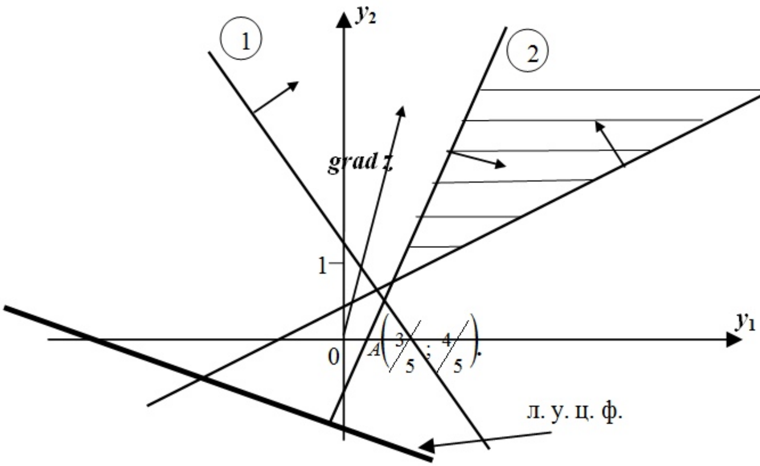

Рис. 3.1. Графическое решение двойственной задачи

Ее решение: $x_1^0 = 0; x_2^0 = \frac{7}{5}; x_3^0 = \frac{11}{5}$. Таким образом, точка максимума исходной задачи $\left(0; \frac{7}{5}; \frac{11}{5}\right)$.

Пример 3.3. Будет ли план $\overrightarrow{x} = \{1;\ 1\colon\ 1\}$ оптимальным планом задачи

$$z = x_1 + 2x_2 + 3x_3 \to max;$$

$$y_1, y_2 \begin{cases} x_1 - x_2 - x_3 \le 1, \\ x_1 + x_2 - x_3 \le 1; \end{cases}$$

$$x_1, x_2, x_3 \ge 0?$$

Составим двойственную задачу:

$$f = y_1 + y_2 \to min;$$

$$\begin{cases} y_1 + y_2 \ge 1, \\ -y_1 + y_2 \ge 2, \\ -y_1 - y_2 \ge 3; \end{cases}$$

$$y_1, y_2 \ge 0.$$

Составим условия дополнительной нежесткости:

$$\begin{cases} y_1^0 \left(x_1^0 - x_2^0 - x_3^0 - 1\right) = 0, \\ y_2^0 \left(x_1^0 + x_2^0 - x_n^0 - 1\right) = 0, \\ x_1^0 \left(y_1^0 + y_2^0 - 1\right) = 0, \\ x_2^0 \left(-y_1^0 + y_2^0 - 2\right) = 0, \\ x_3^0 \left(-y_1^0 - y_2^0 - 3\right) = 0. \end{cases}$$

Подставим в эти равенства $x_1^0 = x_2^0 = x_3^0 = 1$.
Получим систему уравнений

$$\begin{cases} -2y_1^0 = 0, \\ y_1^0 + y_2^0 = 1, \\ -y_1^0 + y_2^0 = 2, \\ -y_1^0 - y_2^0 = 3, \end{cases}$$

которая является несовместной. Следовательно, рассматриваемый план оптимальным планом не будет.

3.5. Третья теорема двойственности

Пусть некоторая задача линейного программирования решена и получено некоторое значение z_{max}. При изменении правых частей ограничений $b_1, b_2, \ldots, b_m$ задача меняется и, в конечном счёте, меняется значение z_{max}. Будем рассматривать zmax как функцию величин $b_1, b_2, \ldots, b_m$, то есть рассмотрим функцию $z_{max}(b_1, b_2, \ldots, b_m)$.

Теорема 3.3. В оптимальном плане двойственной задачи координата y_1^0 численно равна частной производной функции $z_{max}(b_1, b_2, \ldots, b_m)$ по переменной b_i, то есть

$$\frac{\partial z_{max}}{\partial b_i} = y_1^0; \quad i = 1, 2, \ldots, m.$$

Эта теорема может быть истолкована и следующим образом. Пусть величинам $b_1, b_2, \ldots, b_m$ придаются приращения $\Delta b_1, \Delta b_2, \ldots, \Delta b_m$. Функция $z_{max}(b_1, b_2, \ldots, b_m)$ получает приращение Δz_{max}. По сформулированной теореме при малых приращениях $\Delta b_1, \Delta b_2, \ldots, \Delta b_m$ справедливо следующее приближенное равенство

$$\Delta z_{max} \approx y_1^0 \Delta b_1 + y_2^0 \Delta b_2 + \ldots + y_m^0 \Delta b_m. \tag{3.1}$$

На самом деле функция $z_{max}(b_1, b_2, \ldots, b_m)$ такова, что при изменении Δb_i в некоторых пределах, равенство (3.1) является точным. При этом величины Δb_i могут быть и не малыми по сравнению с b_i.

Определить эти пределы изменения величин Δb_i можно с помощью второй теоремы двойственности. Равенство (3.1) будет точным, если при изменении b_i величины y_i^0 не изменяются. Пределы же изменения приращений Δb_i, при которых оптимальный план двойственной задачи не меняется, можно определить следующим образом. Придадим фиксированным значениям $b_1^0, b_2^0, \ldots, b_m^0$ неопределенные приращения $\Delta b_1, \Delta b_2, \ldots, \Delta b_m$ и рассмотрим условия дополнительной нежесткости с наращенными значениями $b_i = b_i^0 + \Delta b_i$, в которые подставим старые значения y_i^0. Относительно оптимального плана исходной задачи $x_1^0, x_2^0, \ldots, x_n^0$ получится система линейных уравнений, которую можно решить при определенных условиях с помощью, например, метода обратной матрицы. Получим зависимость

$$x_i^0 = x_i^0 \left(\Delta b_1, \Delta b_2, \ldots, \Delta b_m\right)$$

оптимального плана от приращений. Поскольку x_i^0 должны удовлетворять неравенствам $x_i^0 \geq 0, i = 1, 2, \ldots, n$, а также системе ограничений исходной задачи, то мы получим систему условий, которым должны удовлетворять Δb_i, чтобы оптимальный план двойственной задачи не менялся. Эта область изменения Δb_i называется областью устойчивости двойственных оценок.

Пример 3.4. Рассмотрим пару двойственных задач примера 3.2. Оптимальный план двойственной задачи здесь $y_1^0 = \frac{3}{5}; y_2^0 = \frac{4}{5}$. Наращенные значения правых частей ограничений исходной задачи равны: $b_1 = 2 + \Delta b_1, b_2 = 3 + \Delta b_2$. Условия дополнительной нежесткости имеют вид

$$\begin{cases} \frac{3}{5}\left(2x_1^0 + 3x_2^0 - x_3^0 - 2 - \Delta b_1\right) = 0, \\ \frac{4}{5}\left(x_1^0 - x_2^0 + 2x_3^0 - 3 - \Delta b_2\right) = 0, \\ x_1^0\left(2 \cdot \frac{3}{5} + \frac{4}{5} - 1\right) = 0, \\ x_2^0\left(\frac{9}{5} - \frac{4}{5} - 1\right) = 0, \\ x_3^0\left(-\frac{3}{5} + \frac{8}{5} - 1\right) = 0. \end{cases}$$

Четвёртое и пятое равенства — тождества, а из третьего следует, что $x_1^0 = 0$. Для x_2^0 и x_3^0 получаем систему равенств

$$\begin{cases} 3x_2^0 - x_3^0 = 2 + \Delta b_1, \\ -x_2^0 + 2x_3^0 3 + \Delta b_2. \end{cases}$$

Отсюда

$$x_2^0 = \frac{2}{5}(2 + \Delta b_1) + \frac{1}{5}(3 + \Delta b_2) = \frac{7}{5} + \frac{2}{5}\Delta b_1 + \frac{1}{5}\Delta b_2,$$
$$x_3^0 = \frac{1}{5}(2 + \Delta b_1) + \frac{3}{5}(3 + \Delta b_2) = \frac{11}{5} + \frac{1}{5}\Delta b_1 + \frac{3}{5}\Delta b_2.$$

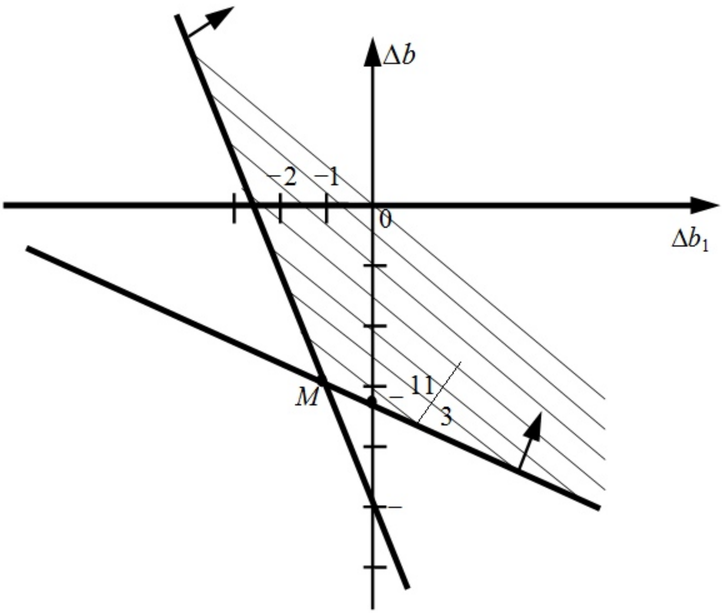

Рис. 3.2. Область устойчивости двойственных оценок примера 3.4

Приращения Δb_1 и Δb_2 должны удовлетворять неравенствам

$$\begin{cases} \frac{7}{5} + \frac{2}{5}\Delta b_1 + \frac{1}{5}\Delta b_2 \geq 0, \\ \frac{11}{5} + \frac{1}{5}\Delta b_1 + \frac{3}{5}\Delta b_2 \geq 0. \end{cases}$$

Система ограничений исходной задачи с правыми частями $2 + \Delta b_1$ и $3 + \Delta b_2$ удовлетворяется автоматически. Таким образом, можно изобразить на плоскости область изменения величин $\Delta b_1, \Delta b_2$, в которой оптимальный план двойственной задачи не меняется.

В области, изображенной на рис. 3.2, приращение Δz_{max} может быть найдено по точной формуле:

$$\Delta z_{max} = y_1^0 \Delta b_1 + y_2^0 \Delta b_2 = \frac{3}{5}\Delta b_1 + \frac{4}{5}\Delta b_2.$$

3.6. Послеоптимизационный анализ

Пусть рассматривается задача оптимального использования ресурсов. Пусть решена и сама эта задача, и двойственная к ней. Знание двойственных оценок $y_1^0, y_2^0, \ldots, y_m^0$ позволяет судить о том, насколько ценным в производстве является данный ресурс. Если величина $y_i^0 \neq 0$ то в силу условий дополнительной нежесткости ресурс при реализации оптимального плана выпуска продукции используется полностью. Не полностью используются ресурсы, для которых $y_i^0 = 0$. Таким образом, двойственные оценки ресурсов характеризуют их дефицитность. Различные варианты наращивания ресурсов (или их сокращения) приводят к разным величинам изменения оптимальной прибыли Δz_{max}. Чем больше величина y_i^0, тем больше приращение оптимальной прибыли мы получим при увеличении объема ресурса R_i на 1 единицу. В этом и заключается более глубокий смысл двойственных оценок ресурсов по сравнению с изложенным ранее.

При анализе возможностей расширения производства знание двойственных оценок ресурсов играет большую роль. Обычно такой анализ называют послеоптимизационным, поскольку он выполняется после решения прямой и двойственной задач. Здесь, однако, следует помнить, что для использования двойственных оценок в экономическом анализе нужно знать область их устойчивости, то есть область изменения ресурсов, в которой двойственные оценки ресурсов не меняются.

3.7. Двойственный симплекс-метод для пары симметрично-двойственных задач

Если размерности исходной и двойственной задач велики, то определение решения одной из задач по решению другой с помощью условий дополнительной нежесткости затруднительно. В этом случае применяют двойственный симплекс-метод, который позволяет по последней симплекс-таблице одной из задач прочесть решение двойственной задачи. Этот метод мы рассмотрим сначала в упрощенной форме для пары симметрично-двойственных задач, причем для такой пары, в которой одна из задач после уравнивания неравенств системы ограничений имеет эту систему ограничений в допустимом базисном виде, то есть, может решаться симплекс-методом непосредственно. В следующем параграфе будет рассмотрен подобный метод для произвольной задачи в канонической форме.

Пусть при уравнивании системы ограничений исходной задачи мы ввели переменные $x_{n+1}, x_{n+2}, \ldots, x_{n+m}$. Эти переменные играют роль базисных. При уравнивании системы ограничений двойственной задачи вводятся переменные $y_{m+1}, y_{m+2}, \ldots, y_{m+n}$, которые также играют роль базисных. Определим соответствие между переменными исходной и двойственной задач, сопоставляя каждой свободной переменной исходной задачи базисную переменную двойственной в соответствии с порядком следования номеров, а каждой базисной переменной исходной задачи — свободную переменную двойственной задачи. Запишем это соответствие в таком виде:

$$\begin{array}{cccccccc} x_1, & x_2, & \ldots, & x_n, & x_{n+1}, & x_{n+2}, & \ldots, & x_{n+m} \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ y_{m+1}, & y_{m+2}, & \ldots, & y_{m+n}, & y_1, & y_2, & \ldots, & y_m. \end{array} \qquad (F)$$

Отметим, что это соответствие переменных устанавливает взаимосвязь базисных решений данной и двойственной задач. Если составлена таблица, аналогичная симплекс-таблице, по базисному решению данной задачи, то соответствующее базисное решение двойственной задачи прочитывается по последней строке этой симплекс-таблицы с использованием построенного выше соответствия переменных. А именно, значение y_i в базисном решении двойственной задачи прочитывается в строке целевой функции симплекс-таблицы в столбце переменной x_j, соответствующей переменной y_i, по соответствию переменных (F). Эта взаимосвязь базисных решений сохраняется при всевозможных преобразованиях замещения данной таблицы в предположении, что в двойственной задаче происходит переход к соответствующему новому базисному решению.

Пусть одна из задач решена симплекс-методом и получена последняя симплекс-таблица. Решение (точка оптимума) другой задачи может быть прочитано по последней симплекс-таблице с использованием соответствия переменных (F).

Пример 3.5. Рассматривается пара двойственных задач:

1) $z = -x_1 + x_2 \to max;$

$$\begin{cases} x_1 - 2x_2 \le 1, \\ -2x_1 + x_2 \le 2, \\ 3x_1 + x_2 \le 3; \end{cases}$$

$$x_1 \ge 0, x_2 \ge 0.$$

2) $f = y_1 + 2y_2 + 3y_3 \to min;$

$$\begin{cases} y_1 - 2y_2 + 3y_3 \ge -1, \\ -2y_1 + y_2 + y_3 \ge 1; \end{cases}$$

$$y_1 \ge 0, y_2 \ge 0, y_3 \ge 0.$$

После уравнивания неравенств задачи имеют следующий вид:

1) $z = -x_1 + x_2 \to max;$

$$\begin{cases} x_1 - 2x_2 + x_3 = 1, \\ -2x_1 + x_2 + x_4 = 2, \\ 3x_1 + x_2 + x_5; \end{cases}$$

$$x_1, x_2, \ldots, x_5 \ge 0.$$

2) $f = y_1 + 2y_2 + 3y_3 \to min;$

$$\begin{cases} y_1 - 2y_2 + 3y_3 - y_4 = -1, \\ -2y_1 + y_2 + y_3 - y_5 = 1; \end{cases}$$

$$y_1, y_2, \ldots, y_5 \ge 0.$$

Составим соответствие переменных:

$$\begin{array}{ccccc} x_1, & x_2, & x_3, & x_4, & x_5 \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ y_1, & y_2, & y_3, & y_4, & y_5. \end{array}$$

Решая исходную задачу симплекс-методом, получаем следующую последнюю симплекс-таблицу:

Базисные переменные	Свободные члены	x_1	x_2	x_3	x_4	x_5
x_3	28/5	0	0	1	7/5	3/5
x_2	12/5	0	1	0	3/5	2/5
x_1	1/5	0	0	0	4/5	1/5
z	11/5	0	0	0	4/5	1/5

По первой теореме двойственности и в соответствии со сказанным выше $f_{min} = \frac{11}{5}$; точка максимума имеет координаты

$$y_1^0 = 0; y_2^0 = \frac{4}{5}; y_3^0 = \frac{1}{5}.$$

Двойственный симплекс-метод можно распространить и на произвольные задачи, однако, при этом он заметно усложняется.

3.8. Метод последовательного уточнения оценок. Обобщенный двойственный симплекс-метод

Метод последовательного уточнения оценок применяется к задаче линейного программирования в канонической форме

$$z = c_0 + \left(\overrightarrow{c}, \overrightarrow{x}\right) \to max,$$

$$A\overrightarrow{x} = \overrightarrow{b}, \overrightarrow{x} \geq \overrightarrow{0},$$

где А — $(m \times n)$ — матрица. Он так же, как и симплекс метод заключается в таком последовательном переходе от одного n-мерного вектора $\overrightarrow{x}$ к другому, при котором после конечного числа переходов либо получается оптимальное решение задачи, либо устанавливается ее неразрешимость. Однако в отличие от симплекс-метода эти последовательно получаемые векторы, вообще говоря, не являются допустимыми решениями системы ограничений (допустимыми планами), а являются в некотором смысле «почти» допустимыми, так называемыми *псевдопланами* данной задачи. Эти псевдопланы являются базисными решениями системы уравнений задачи, но не обязательно удовлетворяют условиям неотрицательности. Процесс перехода от одного псевдоплана к другому построен так, что как только очередной псевдоплан $\overrightarrow{x}$ оказывается неотрицательным, т.е. допустимым планом задачи, он является одновременно и оптимальным решением этой задачи. Каждому псевдоплану по его определению соответствует опорное решение системы ограничений двойственной задачи. При этом каждому следующему певдоплану соответствует лучшее, чем предыдущее, опорное решение двойственной задачи. Оптимальному решению данной задачи соответствует оптимальное решение двойственной. Таким образом, метод последовательного уточнения оценок — это применение обычного симплекс метода к двойственной задаче, дополненное на каждой итерации построением n-мерного вектора $\overrightarrow{x}$, являющегося псевдопланом данной задачи. Геометрически этот метод можно истолковать как построение последовательности псевдопланов $\overrightarrow{x_1}, \overrightarrow{x_2}, \ldots, \overrightarrow{x_5}$, являющихся точками n-мерного пространства R^n. находящимися вне допустимого многогранника. При этом каждый следующий вектор этой последовательности определяет гиперплоскость уровня целевой функции, которая находится ближе к области допустимых значений, чем предыдущая. Через конечное число шагов мы получим точку $\overrightarrow{x_5}$, принадлежащую допустимому многограннику, либо выяснится, что допустимый многогранник пуст. Перейдем к более детальному описанию метода. Предположим, что система ограничений задачи в канонической форме приведена к базисному виду, который, вообще говоря, не является допустимым и пусть из целевой функции исключены базисные переменные. Составим таблицу, аналогичную симплекс таблице. Отличие заключается лишь в том, что эта таблица составляется по, вообще говоря, недопустимому базисному виду системы уравнений. Эту таблицу будем называть обобщенной симплекс таблицей. Коэффициенты Δ_j при свободных переменных в последней строке этой таблицы назовем оценками свободных переменных для данного базисного решения.

Определение 3.1. Базисное решение системы ограничений задачи линейного программирования в канонической форме называется псеввдопланом этой задачи, если для него все оценки свободных переменных неотрицательны ($\Delta_j \geq 0$).

Утверждение 3.1. Если псевдоплан задачи линейного программирования в канонической форме является допустимым планом этой задачи, то он одновременно является оптимальным планом.

Д о к а з а т е л ь с в о. Если псевдоплан является допустимым, то составленная нами таблица будет обычной симплекс таблицей, у которой все коэффициенты $\Delta_j = -\gamma_j$ в последней строке неотрицательны. Значит, получена последняя симплекс таблица и соответствующий план будет оптимальным, что и требовалось доказать.

Утверждение 3.2. Если псевдоплан задачи линейного программирования в канонической форме имеет отрицательную компоненту $b_k < 0$, а все коэффициенты k-ой строки таблицы неотрицательны $a_{kj} \geq 0$, то задача не имеет решения по причине отсутствия допустимых решений системы ограничений.

Д о к а з а т е л ь с т в о. k-ая строка таблицы отвечает уравнению

$$x_k = b_k - \sum_j a_{kj} x_j,$$

которое при наших условиях не может иметь неотрицательного решения. Следовательно, система ограничений задачи допустимых решений не имеет. Утверждение доказано.

Рассмотрим некоторую процедуру перехода от одной обобщенной симплекс таблицы к другой, которая соответствует переходу от одного псевдоплана к другому. Если таблица соответствует псевдоплану, который не является допустимым планом задачи, то в столбце свободных членов среди отрицательных элементов выберем максимальное по модулю отрицательное число $b_k < 0$ и в k-ой строке отметим все отрицательные числа $a_{kj} < 0$. Если таковых не окажется, то в силу утверждения 2 задача является недопустимой и решения не имеет. Среди отмеченных элементов k-ой строки выберем элемент a_{kj_0}, на котором достигается $\underset{j}{min}\left(-\frac{\Delta_j}{a_{kj}}\right)$. Если таких элементов несколько, то выбираем любой из них. Этот выбранный элемент берем в качестве разрешающего и производим операцию замещения в нашей таблице, выводя переменную x_k из числа базисных и вводя в число базисных переменную x_{j_0}. Отметим, что новая таблица тоже будет отвечать псевдоплану задачи вследствие правила выбора разрешающего элемента. При этом в новой таблице значение целевой функции γ_0 будет меньше, чем в старой. Таким образом, справедливо

Утверждение 3.3. При описанном переходе от обобщенной симплекс таблицы, отвечающей некоторому псевдоплану, либо выяснится, что исходная задача не имеет решения, либо получится новая таблица, отвечающая некоторому новому псевдоплану. При этом значение целевой функции на новом псевдоплане будет меньше, чем на старом.

Описанные итерации повторяются до тех пор, пока очередной псевдоплан не станет допустимым планом задачи, который одновременно будет оптимальным планом. Важно заметить, что каждому псевдоплну однозначно сопоставляется некоторый опорный план двойственной задачи. Действительно, если система ограничений приведена к базисному виду, а из целевой функции исключены базисные переменные, то базисные переменные можно убрать, превращая уравнения в неравенства. Для полученной формы задачи построим двойственную задачу. Приведение данной и двойственной задач к каноническому виду дает пару задач с одинаковым числом переменных, которые можно привести во взаимно однозначное соответствие вида (F). С использованием этого соответствия и обобщенной симплекс таблицы мы можем, как и в предыдущем параграфе каждому псевдоплану сопоставить допустимый план двойственной задачи. Наши итерации, по сути, являются переходами от одного опорного решения двойственной задачи к другому с улучшением значения целевой функции. Поскольку координаты опорного решения двойственной задачи совпадают с оценками свободных переменных псевдоплана, то мы, по сути, последовательно улучшаем эти оценки. Отсюда происходит название рассмотренного метода.

Утверждение 3.4. Если исходная задача не имеет решения по причине неограниченности целевой функции на области допустимых значений, то она не имеет ни одного псевдоплана.

Д о к а з а т е л ь с т в о. Предположим противное, т. е. что существует хотя бы один псевдоплан. Ему должен соответствовать допустимый план двойственной задачи. Но по первой теореме двойственности при наших условиях двойственная задача должна быть недопустимой, т. е. ее область допустимых значений пуста. Пришли к противоречию. Утверждение 4 доказано.

Пример 3.6. Решить задачу

$$z = 16 - x_1 - x_2 \to max;$$

$$\begin{cases} x_1 + x_2 \leq 8, \\ x_1 - x_2 \geq 4, \\ x_1 + 2x_2 \geq 6; \end{cases}$$
$$x_1, x_2 \geq 0.$$

Построим двойственную задачу

$$f = 16 + 8y_1 - 4y_2 - 6y_3 \to min;$$

$$\begin{cases} y_1 - y_2 - y_3 \geq -1, \\ y_1 + y_2 - 2y_3 \geq -1; \end{cases}$$
$$y_1, y_2, y_3 \geq 0.$$

Приведем исходную и двойственную задачи к каноническому виду.

$$z = 16 - x_1 - x_2 \to max;$$

$$\begin{cases} x_1 + x_2 + x_3 = 8, \\ -x_1 + x_2 + x_4 = -4, \\ -x_1 - 2x_2 + x_5 = -6; \end{cases}$$
$$x_1, x_2, x_3, x_5 \geq 0.$$

$$f = 16 + 8y_1 - 4y_2 - 6y_3 \to min;$$

$$\begin{cases} -y_1 + y_2 + y_3 + y_4 = 1, \\ -y_1 - y_2 + 2y_3 + y_5 = 1; \end{cases}$$
$$y_1, y_2, y_3 \geq 0.$$

Составим обобщенную симплекс таблицу исходной задачи.

Таблица 1

Базисные переменные	Свободные члены	x_1	$\downarrow x_2$	x_3	x_4	x_5
x_3	8	1	1	1	0	0
x_4	-4	-1	1	0	1	0
$\leftarrow x_5$	-6	-1	[-2]	0	0	1
z	16	1	1	0	0	0

Таблица отвечает псевдоплану. Согласно соответствию переменных (F) этому псевдоплану соответствует допустимый план (0, 0, 0, 1, 1) двойственной задачи. По сформулированному выше правилу выбираем разрешающий элемент (-2) и перейдем к следующей таблице.

Эта таблица тоже соответствует псевдоплану, а также допустимому плану (0, 0, $\frac{1}{2}$, $\frac{1}{2}$, 0) двойственной задачи.

В последней симплекс таблице псевдоплан является допустимым планом, а, следовательно, оптимальным планом задачи. Получаем ответ в исходной задаче: $z_{max} = \frac{32}{3}; (\frac{14}{3}; \frac{2}{3})$. Решение двойственной задачи тоже прочитывается по последней симплекс таблице: $f_{min} = \frac{32}{3}$; точка минимума (0; $\frac{1}{3}$; $\frac{2}{3}$).

Отметим, что для получения решения исходной задачи методом последовательного уточнения оценок строить двойственную задачу, как в рассмотренном примере, совсем не обязательно. Рассмотренный пример показывает, по сути, равносильность

Таблица 2

Базисные переменные	Свободные члены	$\downarrow x_1$	x_2	x_3	x_4	x_5
x_3	5	1/2	0	1	0	1/2
$\leftarrow x_4$	-7	[-3/2]	0	0	1	1/2
x_5	3	1/2	1	0	0	-1/2
z	13	1/2	0	0	0	1/2

Таблица 3

Базисные переменные	Свободные члены	x_1	x_2	x_3	x_4	x_5
x_3	8/3	0	0	1	1/3	2/3
x_1	14/3	1	0	0	-2/3	-1/3
x_2	2/3	0	1	0	1/3	-1/3
z	32/3	0	0	0	1/3	2/3

нашего метода и двойственного симплекс метода, рассмотренного в предыдущем параграфе. Поэтому в дальнейшем метод последовательного уточнения оценок мы часто будем называть тоже *двойственным симплекс методом*. Этот метод удобен в случае, когда удалось выйти на псевдоплан. Такая ситуация встречается довольно часто (см., например, примеры 5.3 и 5.4). В общем же его применение может привести к сокращению количества итераций, необходимых в среднем для решения произвольной задачи линейного программирования в канонической форме по сравнению с применением простого симплекс метода или его модификаций типа метода искусственного базиса. Следующий составной алгоритм часто называют *обобщенным двойственным симплекс методом*.

1. Привести систему ограничений задачи к базисному виду и исключить базисные переменные из целевой функции.
2. Проверить, будет ли полученное базисное решение допустимым планом задачи. Если да, то решить задачу обычным симплекс методом и перейти к п. 5, иначе перейти к п. 3.
3. Проверить, будет ли полученное базисное решение псевдопланом задачи. Если да, то решить задачу двойственным симплекс методом и перейти к п. 5, иначе перейти к п. 4.
4. Перейти к следующему базисному виду системы ограничений и перейти к п. 2.
5. Конец.

3.9. Понятие об устойчивости задач линейного программирования. Регуляризация неустойчивых задач

Содержание этого параграфы важно для практического применения линейного программирования и существенно использует основные результаты теории двойственности. При численном решении задачи линейного программирования вида

$$z = (\overrightarrow{c}, \overrightarrow{x}) \to max \tag{3.1}$$

$$A\overrightarrow{x} \leq \overrightarrow{b}, \overrightarrow{x} \geq 0 \tag{3.2}$$

ЭВМ оперирует лишь приближенными значениями параметров задачи, округляемыми в процессе счета. По сути дела происходит замена задачи (3.2) некоторой задачей

$$z = (\overrightarrow{c}(\delta), \overrightarrow{x}) \to max \tag{3.3}$$

$$A(\delta)\overrightarrow{x} \leq \overrightarrow{b}(\delta), \overrightarrow{x} \geq 0, \tag{3.4}$$

где относительно матрицы $A(\delta)$ и векторов $\overrightarrow{b}(\delta)$, $\overrightarrow{c}(\delta)$ известно только, что они в определенной степени близки к истинным значениям матрицы А и векторов $\overrightarrow{b}, \overrightarrow{c}$. Кроме того, при исследовании математических моделей реальных явлений параметры модели получаются на основании экспериментальных данных и, чаще всего, известны приближенно с определенной степенью точности. Поэтому, если исследуемая модель имеет вид (3.2), то можно сказать, лишь, что истинная модель описывается одной из задач вида (3.3). В связи со сказанным вопросы о взаимосвязи этих задач представляют большой практический интерес.

Близость матриц и векторов будем выражать в обычной евклидовой метрике, т. е. расстоянием между двумя матрицами $A = (a_{ij})$ и $A^{(1)} = (a_{ij})$ размера $m \times n$ назовем число

$$\|A - A^{(1)}\| = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m}(a_{ij} - a_{ij}^{(1)})^2}.$$

Аналогично, если $\overrightarrow{b} = b_1, b_2, \ldots, b_m$, $\overrightarrow{b}^{(1)} = b_1^{(1)}, b_2^{(2)}, \ldots, b_m^{(1)}$, то

$$\|\overrightarrow{b} - \overrightarrow{b}^{(1)}\| = \sqrt{\sum_{i=1}^{m}(b_i - b_i^{(1)})^2}.$$

При фиксированных матрице А и векторах $\overrightarrow{b}, \overrightarrow{c}$ для любого $\delta > 0$ символами $A(\delta)$, $\overrightarrow{b}(\delta)$, $\overrightarrow{c}(\delta)$ будем обозначать любые элементы из соответствующих δ-окрестностей:

$$\|A - A(\delta)\| < \delta, \|\overrightarrow{b} - \overrightarrow{b}(\delta)\| < \delta, \|\overrightarrow{c} - \overrightarrow{c}(\delta)\| < \delta.$$

Задачу (3.3) будем называть *возмущенной задачей*, принадлежащей δ-окрестности задачи (3.2), если выполнены условия (U).

Обозначим через $X^{(0)}$ множество решений задачи (3.2), а через d —оптимальное значение ее целевой функции. Аналогично через$X^{(0)}(\delta), d(\delta)$ обозначим множество решений и оптимальное значение возмущенной задачи (3.3).

Определение 3.2. Задачу (3.2) назовем устойчивой, если существует такое число $\delta_0 > 0$, что для всех δ таких, что $0 \leq \delta \leq \delta_0$, задача (3.3) имеет решение. Другими словами, задача (3.2) устойчива, если она имеет решение, а также имеет решение любая задача, получающаяся из нее небольшими изменениями параметров.

Определение 3.3. Задачу (3.2) назовем устойчивой по функционалу, если

а) она устойчива;

б) для любого $\varepsilon > 0$ существует та кое $\delta > 0$, что как только выполнены условия (U), то выполнено неравенство $\|d - d(\delta)\| < \epsilon$. Иначе можно сказать оптимальное значение целевой функции задачи (3.2) как функция $d(A, \overrightarrow{b}, \overrightarrow{c})$ ее параметров в этом случае непрерывна.

Определение 3.4. Задачу (3.2) назовем устойчивой по решению, если

а) она устойчива;

б) для любого $\varepsilon > 0$ существует такое $\delta > 0$, что как только выполнены условия (U) для любого $(\overrightarrow{x}^0()\delta) \in X^{(0)}(\delta)$ найдется вектор $(\overrightarrow{x}^0(\delta) \in X^{(0)}(\delta)$, удовлетворяющий неравенству $\|\overrightarrow{x}^0 - \overrightarrow{x}^0(\delta)\| < \varepsilon$.

Ясно, что если задача (3.2) неустойчива в смысле хотя бы одного из определений 3.2 — 3.4, то ее решение на ЭВМ может привести к результатам, сильно отличающимся от истинных.

С помощью средств, выходящих за рамки настоящего пособия, можно установить, что все три определения устойчивости эквивалентны. Точнее, если задача (3.2) устойчива в смысле определения 3.2, то она устойчива и по функционалу и по решению. Приведем также без доказательства необходимые и достаточные условия устойчивости задачи (3.2).

Теорема 3.4. Стандартная задача линейного программирования (3.2) устойчива тогда и только тогда, когда существуют такие векторы $\overrightarrow{x}^0 \in R^n$ и $\overrightarrow{p}^0 \in R^m$, что $\overrightarrow{x}^0 > 0, \overrightarrow{p}^0 > 0, A\overrightarrow{x}^0 < \overrightarrow{b}, A^T\overrightarrow{p}^0 > \overrightarrow{c}$, где A^T — матрица, транспонированная по отношению к матрице А. Сформулированная теорема не позволяет устанавливать устойчивость или неустойчивость задачи (3.2) по ее внешнему виду, однако является весьма полезной при численном решении задач, которые имеют решение, но могут оказаться неустойчивыми.

В рамках общей методологии регуляризации некорректных задач, принадлежащей А.Н. Тихонову (см., например, [40]), построены методы, позволяющие решать произвольную задачу линейного программирования с любой степенью точности безотносительно к тому, устойчива она или нет. Однако при этом регуляризованная задача уже не является задачей линейного программирования. Здесь мы изложим метод регуляризации, не выводящий за рамки линейного программирования (см.[3] стр.271-286).

Предположим, что данная задача вида (3.2) не является устойчивой. Как следует из теоремы 3.4, это может происходить по двум причинам: а) не существует положительного вектора $\overrightarrow{x}^0 > 0, A\overrightarrow{x}^0 < \overrightarrow{b}$ б) не существует положительного вектора $\overrightarrow{p}^0 > 0$, для которого $A^T\overrightarrow{p}^0 > \overrightarrow{c}$. Будем предполагать, что задача (3.2) имеет решение. Это означает, в частности, что допустима как задача (3.2), так и двойственная к ней, т.е. существуют векторы $\overrightarrow{x}_1 \geq 0 : A\overrightarrow{x}_1 \leq \overrightarrow{b}$ и $\overrightarrow{p}_1 \geq 0 : A^T\overrightarrow{p}_1 \geq \overrightarrow{c}$. Сопоставим задаче (3.2) возмущенную задачу вида

$$z = (\overrightarrow{c} - \delta\overrightarrow{e}, \overrightarrow{x}) \to max, \tag{3.5}$$

$$A\overrightarrow{x} \leq \overrightarrow{b} + \delta\overrightarrow{\nu}, \overrightarrow{x} \geq 0, \tag{3.6}$$

где число $\delta > 0, \overrightarrow{e} = \{1, 1, \ldots, 1\} \in R^n, \overrightarrow{\nu} = \{1, 1, \ldots, 1\} \in R^m$.

Теорема 3.5. Если задача (3.2) имеет решение, то задача (3.4) устойчива при любом $\delta > 0$.

Д о к а з а т е л ь с т в о. Покажем, что для задачи (3.4) выполняются условия теоремы 3.4. Положим $\overrightarrow{x}^0 = \overrightarrow{x}_1 + \alpha\overrightarrow{e}, \overrightarrow{p}^0 = \overrightarrow{p}_1 + \alpha\overrightarrow{\nu}$. Поскольку $A\overrightarrow{x}_1 < \overrightarrow{b} + \delta\overrightarrow{\nu}, A^T\overrightarrow{p}_1 > \overrightarrow{c} - \delta\overrightarrow{e}$, то найдется такое число $\alpha > 0$, что

$$A\overrightarrow{x}^0 = A\overrightarrow{x}_1 + \alpha A\overrightarrow{e} < \overrightarrow{b} + \delta\overrightarrow{\nu},$$

$$A^T\overrightarrow{p}^0 = A^T\overrightarrow{p}_1 + \alpha A^T\overrightarrow{\nu} > \overrightarrow{c} - \delta\overrightarrow{e},$$

При этом очевидно, что $\overrightarrow{x}^0 > 0, \overrightarrow{p}^0 > 0$. Теорема доказана.

Задачу (3.4) назовем регуляризованной по отношению к задаче (3.2). Следующие две теоремы показывают, что с помощью решения регуляризованной задачи в принципе можно найти с любой степенью точности решение (если оно существует) задачи вида (3.2) безотносительно к ее устойчивости или неустойчивости.

Теорема 3.6. *Пусть $d, d(\beta)$ - оптимальные значения соответственно задачи вида (3.2) и регуляризованной по отношению к ней задачи (3.4), тогда справедливо равенство* $\lim_{\beta\to 0} d(\beta) = d$.

Д о к а з а т е л ь с т в о. Пусть $\overrightarrow{x}^0$ и $\overrightarrow{p}^0$ решения соответственно задачи (3.2) и двойственной к ней. Вектор $\overrightarrow{x}^0$ является допустимым для задачи (3.4) при любом $\delta > 0$. Следовательно, $(\overrightarrow{x} - \delta\overrightarrow{e}, \overrightarrow{x}^0) \leq d(\delta)$ или $(\overrightarrow{c}, \overrightarrow{x}^*) - \delta(\overrightarrow{e}, \overrightarrow{x}^*) \leq d(\delta)$. Учитывая, что $\overrightarrow{c}, \overrightarrow{x}^0) = d$, получим $d - \delta(\overrightarrow{e}, \overrightarrow{x}^0) \geq d(\delta)$. Проводя аналогичные рассуждения для задач двойственных к задачам (3.2), (3.4), получим $d(\delta) \leq d + \delta(\overrightarrow{v}, \overrightarrow{p}^0)$. Объединяя эти неравенства имеем

$$d - \delta(\overrightarrow{e}, \overrightarrow{x}^0) \leq d(\delta) \leq d + \delta(\overrightarrow{v}, \overrightarrow{p}^0).$$

Отсюда вытекает утверждение теоремы.

Приведем без доказательства теорему, показывающую что решение $\overrightarrow{x}(\delta)$ задачи (3.4) при малых $\delta > 0$ близко к множеству $X^{(0)}$ решений задачи (3.2).

Теорема 3.7. *Для любого $\epsilon > 0$ существует такое число $\delta > 0$, что для каждого решения $\overrightarrow{x}(\delta) \in X^{(0)}(\delta)$ задачи (3.4) найдется решение $\overrightarrow{x}^0 \in X^{(0)}$ задачи (3.2), удовлетворяющее неравенству $\|\overrightarrow{x}^0 - \overrightarrow{x}(\delta)\| < \epsilon$.*

Таким образом теоремы 3.6 и 3.7 обосновывают следующий способ приближенного решения задачи вида (3.2), имеющей решение, но, вообще говоря, неустойчивой. Выберем последовательность положительных чисел $\{\delta_k\}$ такую, что $\lim_{k\to\inf} \delta_k = 0$. Будем поочередно решать задачи (3.4) с $\delta = \delta_k$ при $k = l, 2, \ldots$. Решение каждой последующей задачи является более точным приближением к решению задачи (3.2), чем решение предыдущей.

Контрольные вопросы и задачи для самостоятельного решения

1. Сформулируйте правило составления задачи, двойственной по отношению к данной задаче линейного программирования в стандартной форме. Какие пары задач называют симметричными взаимно двойственными?
2. В чем заключается экономический смысл пары симметрично двойственных задач?
3. Несимметрично двойственные задачи. В чем состоит общее правило построения двойственных задач?
4. Сформулируйте первую теорему двойственности. Что позволяет сказать эта теорема о задаче линейного программирования, если известно решение двойственной задачи?
5. Сформулируйте вторую теорему двойственности. Какие задачи позволяет решать эта теорема?
6. Третья теорема двойственности. Область устойчивости двойственных оценок и ее отыскание с помощью второй теоремы двойственности?
7. Что такое послеоптимизационный анализ?
8. В чем заключается двойственный симплекс-метод для пары симметрично двойственных задач.
9. Что называется псевдопланом задачи линейного программирования в канонической форме?
10. Опишите алгоритм последовательного уточнения оценок.

Сформулируйте двойственные задачи по отношению к задачам 3.2-3.4

3.1.

$$z = 3x_1 + 3x_2 - 4x_3 \to max;$$

$$\begin{cases} 2x_1 & +x_2 & -3x_3 & \geq 18, \\ 4x_1 & & -5x_3 & \leq 12, \\ 3x_1 & -2x_2 & +x_3 & \geq 14, \end{cases}$$

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

3.2.

$$z = 6x_1 - x_2 + 3x_3 \to min;$$

$$\begin{cases} 4x_1 & -7x_2 & +5x_3 & \leq 15, \\ 2x_1 & +3x_2 & -4x_3 & = 16, \\ 6x_1 & +5x_2 & -8x_3 & \geq 12, \end{cases}$$

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

3.3.

$$z = -2x_1 + 5x_2 - 4x_3 \to max;$$

$$\begin{cases} 4x_1 & +2x_2 & -3x_3 & \geq 9, \\ 3x_1 & -2x_2 & +5x_3 & \leq 8, \\ x_1 & +3x_2 & +4x_3 & \geq 12, \end{cases}$$

$$x_1, x_2 \geq 0.$$

3.4.

$$z = -3x_1 + 4x_2 6x_3 \to min;$$

$$\begin{cases} 2x_1 & +3x_2 & -x_3 & \geq 8, \\ -3x_1 & +2x_2 & -2x_3 & = 10, \\ 5x_1 & -4x_2 & +x_3 & \geq 7, \end{cases}$$

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

Сформулируйте двойственные задачи по отношению к задачам 3.5, 3.6 и найдите их решения графически

3.5.

$$z = 6x_1 + 8x_2 + x_3 \to max,$$

$$\begin{cases} x_1 & +x_2 & +x_3 & \leq 3, \\ x_1 & +2x_2 & & \leq 4, \end{cases}$$

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

Отв.$f_{min} = 20$ при $y_1^0 = 4, y_2^0 = 2$.

3.6.

$$z = 27x_1 + 10x_2 + 15x_3 + 28x_4 \to max;$$

$$\begin{cases} 3x_1 & +2x_2 & +x_3 & +2x_4 & \leq 2, \\ 3x_1 & +x_2 & +3x_3 & +4x_4 & \leq 5, \end{cases}$$

$$x_1, \ldots, x_4 \geq 0.$$

Отв.$f_{min} = 29$ при $y_1^0 = 12, y_2^0 = 1$.

3.7. , **3.8** *В задачах 3.5, 3.6 найти решение исходной задачи, используя решение двойственной и вторую теорему двойственности.*

3.9. *Для модели задачи 1.2 составить двойственную модель. Кроме того: а) найти оптимальный план двойственной задачи; б) определить область устойчивости двойственных оценок по отношению к изменениям количества сырья каждого типа; в) определить увеличение максимальной стоимости изготавливаемой продукции при увеличении количества сырья каждого типа соответственно на 30, 40 и 50 т. Найдите решения задач 3.10 —3.15, используя двойственный симплекс-метод.*

3.10.

$$z = -4x_1 - 7x_2 - 8x_3 - 5x_4 \to max;$$

$$\begin{cases} x_1 & +x_2 & +2x_4 & \geq 4, \\ 2x_1 & +x_2 & +2x_3 & \geq 6, \end{cases}$$

$$x_1, \ldots, x_4 \geq 0.$$

Отв.$z_{max} = 29/2$, точка max $(3; 0; 0; 1/2)$.

3.11.

$$z = 5x_1 + 6x_2 + x_3 + x_4 \to min;$$

$$\begin{cases} 1,5x_1 & +6x_2 & +x_3 & +x_4 & \geq 18, \\ 3x_1 & & +2x_3 & -4x_4 & \geq 24, \end{cases}$$

$$x_1, \ldots, x_4 \geq 0.$$

Отв.$z_{min} = 52$, точка min $(8; 2; 0; 1/2)$.

3.12.

$$z = x_1 + 3x_2 + 4x_3 + 2x_4 \to min;$$

$$\begin{cases} x_1 & -x_2 & +4x_3 & +5x_4 & \geq 27, \\ 2x_1 & +3x_2 & -x_3 & +4x_4 & \geq 24, \end{cases}$$

$$x_1, \ldots, x_4 \geq 0.$$

Ответ.$z_{min} = 12$, точка min $(2; 0; 0; 5)$.

3.13.

$$z = -2x_1 - 8x_2 - x_3 + 5x_4 \to max;$$

$$\begin{cases} -2x_1 & +x_2 & -3x_3 & +x_4 & \geq 18, \\ x_1 & +2x_2 & +4x_3 & +2x_4 & \geq 24, \\ 3x_1 & +4x_2 & +2x_3 & -3x_4 & \geq 30, \end{cases}$$

$$x_1, \ldots, x_4 \geq 0.$$

Ответ.$z_{max} = 126$, точка max $(0; 12; 0; 6)$.

3.14.

$z=-x_1-7x_2+4x_3-9x_4-8x_5+3x_6 \to max;$

$$\begin{cases} 3x_1 & +2x_2 & +3x_3 & -2x_4 & +x_5 & +x_6 & =18, \\ 2x_1 & +x_2 & -x_3 & -3x_4 & +2x_5 & & \geq 24, \end{cases}$$

$$x_i \geq 0 (i=\overline{1,6}).$$

Ответ.$z_{max} = -75$, точка max $(3;0;0;0;9;0)$.

3.15.

$$z=2x_1+3x_2+5x_4 \to max;$$

$$\begin{cases} -2x_1 & +x_2 & -x_3 & =12, \\ x_1 & +2x_2 & +x_4 & =10, \\ 3x_1 & -x_2 & -x_5 & =18, \end{cases}$$

$$x_i \geq 0 (i=\overline{1,5}).$$

Ответ.:Задача не имеет решения.

4. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ ТЕОРИИ ИГР

Во многих видах человеческой деятельности, а особенно в экономике, часто встречаются ситуации, в которых интересы различных лиц, организаций и т.д. противоречат друг другу. Такие ситуации называют конфликтными. Например, при определении цен на товары интересы покупателя и продавца противоположны друг другу. Реальные конфликтные ситуации сложны для анализа, поэтому их нужно предварительно формализовать, отбросив все несущественные детали. Хорошей формой конфликтной ситуации, которая легко поддается формализации, является игра. Разрешение многих конфликтных ситуаций можно представить себе в виде игр, снабженных теми или иными правилами. Поэтому математическое моделирование игр является важной задачей.

Игры можно классифицировать по различным признакам, например, по количеству участников, по правилам распределения выигрышей и т.д. Мы рассмотрим простейшие игры с двумя участниками, в которых выигрыш одного из игроков является проигрышем другого. Такие игры называются играми двух игроков с нулевой суммой.

Во всех играх присутствует понятие стратегии игры, т.е. описание поведения игрока в зависимости от сложившейся ситуации. Игроки могут применять различные стратегии с тем или иным успехом.

4.1. Матричная игра двух игроков с нулевой суммой

Мы рассмотрим следующую формализацию игры, которая состоит из двух наборов стратегий первого и второго игроков. При этом конкретное содержание стратегий нас интересовать не будет. Обозначим возможные стратегии первого игрока через A_1, A_2, ..., A_n, а стратегии второго — B_1, B_2, ..., B_m. Игра является одноходовой: первый игрок применяет одну из своих возможных стратегий, а второй отвечает стратегией из своего набора. После этого происходит распределение выигрышей, которое задается числами — выигрышем первого игрока при условии, что он применяет стратегиюA_i, а второй игрок отвечает стратегией B_j. При этом выигрыш первого игрока является проигрышем второго. Числа образуют матрицу, которую называют матрицей выигрышей первого игрока или платежной матрицей. Величины a_{ij} могут быть как положительными, так и отрицательными или равными нулю. Если $a_{ij} < 0$, то первый игрок проигрывает, а второй выигрывает сумму, равную $|a_{ij}|$.

Пример 4.1. Игра заключается в том, что первый игрок накрывает монету гербом или решкой вверх, а второй отгадывает. Если второй игрок угадывает, то получает выигрыш, равный 1. Если второй игрок не угадывает, то он проигрывает 1. Возможные стратегии первого игрока: A_1 — накрыть монету гербом вверх, A_2 — накрыть монету решкой вверх. У второго игрока тоже две стратегии: B_1 — назвать орла, B_2 — назвать решку. Платежная матрица имеет следующий вид:

	$\boldsymbol{B}_1$	$\boldsymbol{B}_2$
$\boldsymbol{A}_1$	-1	1
$\boldsymbol{A}_2$	1	-1

4.2. Анализ игры в чистых стратегиях

Основной задачей теории игр является выбор для игроков оптимальной в том или ином смысле стратегии. Платежную матрицу анализируют, стремясь найти такую стратегию, которая при любом поведении противника гарантировала бы достаточно высокий выигрыш. При этом считается, что противник ведет себя наиболее неблагоприятным образом.

Пример 4.2. Пусть матричная игра задается матрицей:

	$\boldsymbol{B}_1$	$\boldsymbol{B}_2$	$\boldsymbol{B}_3$	$\boldsymbol{B}_4$	α_i
$\boldsymbol{A}_1$	10	7	4	1	$\boxed{1}$
$\boldsymbol{A}_2$	5	-1	6	8	-1
$\boldsymbol{A}_3$	2	-8	5	3	-8
β_j	10	7	$\boxed{6}$	8	1 \ 6

При наиболее неблагоприятном поведении второго игрока стратегия A_1 первого игрока доставляет ему минимальный гарантированный выигрыш 1. Стратегии A_2 и A_3 в этом смысле хуже, чем A_1, так как при неблагоприятных условиях ведут к проигрышу. С точки зрения второго игрока наилучшей является стратегия B_3, так как ее применение может привести к проигрышу 6, а при остальных стратегиях можно проиграть больше. В этом смысле A_1 наилучшая стратегия первого, а B_3 — второго игрока.

В общем случае игроки оценивают свои стратегии следующим образом. Первый игрок, рассматривая свои стратегии, ищет $\min\limits_j a_{ij}$, а затем выбирает такую стратегию A_i, при которой эта величина наибольшая. При этом он вычисляет величину

$$\alpha = \max_i \min_j a_{ij} \tag{4.1}$$

Величина α называется *нижней ценой игры*, а правило выбора наилучшей стратегии называется *правилом максимина.* В нашем примере α=1. Второй игрок находит $\max\limits_i a_{ij}$, затем выбирает стратегию, при которой эта величина наименьшая, то есть он вычисляет величину $\beta = \min\limits_i \max\limits_j a_{ij}$. β называется *верхней ценой игры*, а правило, которым пользуется второй игрок для выбора своей наилучшей стратегии, — *правилом минимакса.* В нашем примере $\beta = 6$. Можно доказать, что для любой платежной матрицы $\alpha \leqslant \beta$

В рассмотренном примере большую роль играет информация о ходе противника. Если второй игрок знает, что первый применит стратегию A_1, он может, выбирая стратегию B_4 уменьшить свой проигрыш по сравнению со стратегией B_3. Если первый знает, что второй применит стратегию B_3, он имеет возможность увеличить свой выигрыш, применяя стратегию A_2. То есть приведенная выше оценка качества стратегий теряет смысл при наличии информации о поведении противника. Такая ситуация характерна для большинства игр. Информация о поведении противника — залог успеха в игре. Однако для некоторых платежных матриц это не так.

Пример 4.3. Пусть игра задаётся матрицей

В этом примере верхняя и нижняя цены игры совпадают $\alpha = \beta = 6$. Если первый игрок придерживается стратегии A_2, второму ничего не остается, как выбрать стратегию B_2, иначе он проиграет больше. Если второй игрок придерживается стратегии B_2, то первому нужно придерживаться стратегии A_2, иначе он выиграет меньше.

	B_1	B_2	B_3	B_4	α_i
A_1	2	4	7	5	2
A_2	7	6	8	7	$\boxed{6}$
A_3	5	3	4	1	1
β_j	7	$\boxed{6}$	8	7	

Пара стратегий A_2, B_2 является парой оптимальных стратегий для обоих игроков, и информация о поведении противника здесь роли не играет.

В случае, когда верхняя и нижняя цены игры совпадают, говорят, что игра имеет седловую точку в *чистых стратегиях*. В этом случае всегда найдется такой элемент $a_{i_0 j_0}$ платежной матрицы, который удовлетворяет неравенствам

$$a_{ij_0} \leqslant a_{i_0 j_0} \leqslant a_{i_0 j} \tag{4.2}$$

при любых i, j. Этот элемент $a_{i_0 j_0}$ называется *седловой точкой игры,* а стратегии A_{i_0}, B_{j_0} называют *оптимальными стратегиями,* отвечающими седловой точке. Слова "чистая стратегия" означают просто стратегии первого и второго игроков. Существование седловой точки является исключением. Большинство игр не имеет седловой точки в чистых стратегиях. Если же игра имеет седловую точку, то она практически теряет смысл, так как ходы противников предрешены. Часто говорят, что в этом случае игра решается в чистых стратегиях, а пару оптимальных стратегий называют решением игры в чистых стратегиях.

4.3. Понятие смешанной стратегии. Седловая точка игры в смешанных стратегиях

Предположим, что игра повторяется много раз, то есть, проводится достаточно длинная серия партий, и каждый из игроков выбирает в каждой партии свою чистую стратегию случайным образом, в соответствии с некоторыми вероятностями выбора стратегий.

Определение 4.1. *Смешанной стратегией игрока называется совокупность вероятностей выбора им своих чистых стратегий.*

Обозначим через $p_1,\ p_2,\ \dots\ ,\ p_n$ вероятности выбора чистых стратегий первым игроком в серии партий. Эта совокупность чисел и составляет смешанную стратегию первого игрока. Смешанная стратегия — это n-мерный вектор, компоненты которого удовлетворяют условиям

$$p_i \geqslant 0 \ (i = 1, 2, ..., n);\ p_1 + p_2 + ... + p_n = 1. \tag{4.3}$$

Под смешанной стратегией второго игрока будем понимать любой m-мерный вектор $q_1, q_2, ..., q_m$, удовлетворяющий условиям

$$q_i \geqslant 0 \ (i = 1, 2, ..., m);\ q_1 + q_2 + ... + q_m = 1. \tag{4.4}$$

Определение 4.2. *Средним выигрышем первого игрока или платежной функцией игры называется функция $n+m$ переменных $M(\vec{p}, \vec{q}) = \sum\limits_{i,j} a_{ij} p_i q_j$ определенная на смешанных стратегиях первого и второго игроков.*

При конкретном выборе смешанных стратегий $M(\vec{p}, \vec{q})$ является математическим ожиданием выигрыша первого игрока.

Определение 4.3. *Седловой точкой игры в смешанных стратегиях называется такая пара смешанных стратегий первого и второго игроков*

$$\vec{p}^0 = \{p_1^0, p_2^0, ..., p_n^0\}; \ \vec{q}^0 = \{q_1^0, q_2^0, ..., q_m^0\}, \tag{4.5}$$

для которой при любых смешанных стратегиях $\vec{p}, \vec{q}$ выполняются неравенства

$$M(\vec{p}^0, \vec{q}^0) \leqslant M(\vec{p}^0, \vec{q}^0) \leqslant M(\vec{p}^0, \vec{q}) \tag{4.6}$$

Седловая точка представляет собой наиболее выгодную пару смешанных стратегий для каждого игрока, так как отступление от нее невыгодно им обоим.

Отметим, что чистые стратегии можно рассматривать как частный случай смешанных. Например, стратегию A1 первого игрока можно рассматривать как смешанную стратегию {1, 0, 0, ...,0}.

Пример 4.4. Составим платежную функцию игры в так называемое тюремное очко. Каждый из игроков выбрасывает определенное число пальцев (1, 2 или 3). Если сумма выброшенных пальцев четная, то выигрывает первый игрок и его выигрыш равен сумме выброшенных пальцев. Если сумма нечетная, выигрывает второй игрок и его выигрыш равен выброшенной сумме. Здесь каждый из игроков имеет по три стратегии и легко составить платежную матрицу

	B_1	B_2	B_3
A_1	2	-3	4
A_2	-3	4	-5
A_3	4	-5	6

Если игра повторяется много раз и чистые стратегии выбираются в соответствии с вероятностями $\{p_1, p_2, p_3\}$ для первого игрока и $\{q_1, q_2, q_3\}$ для второго, то платежная функция этой игры имеет вид

$$M(\vec{p}, \vec{q}) = 2p_1q_1 - 3p_1q_2 + 4p_1q_3 - 3p_2q_1 + 4p_2q_2 - 5p_2q_3 + 4p_3q_1 - 5p_3q_2 + 6p_3q_3. \tag{4.7}$$

Посчитаем значение платежной функции при следующих смешанных стратегиях $\vec{p}^0 = \left\{\frac{1}{4}; \frac{1}{2}; \frac{1}{4}\right\}$, $\vec{q}^0 = \left\{\frac{1}{4}; \frac{1}{2}; \frac{1}{4}\right\}$, $M(\vec{p}^0, \vec{q}^0) = 0$.

Можно показать, что пара стратегий $\vec{p}^0$, $\vec{q}^0$ в этом примере является седловой точкой.

4.4. Теорема Фон-Неймана. Нахождение седловой точки игры в смешанных стратегиях

Следующая теорема доказана американским математиком Фон-Нейманом (см., например, [34]).

Теорема 4.1. *Для каждой матричной игры двух игроков с нулевой суммой существует пара смешанных стратегий игроков, которая является седловой точкой игры в смешанных стратегиях.*

Теорема утверждает существование седловой точки, но вовсе не ее единственность. В некоторых случаях игра может иметь бесчисленное множество седловых точек в смешанных стратегиях.

Нахождение седловой точки в смешанных стратегиях называют *решением игры* в смешанных стратегиях. Решение игры сводится к решению пары двойственных задач линейного программирования. Сформулируем эту пару задач.

Для первого игрока смешанная стратегия в том и только в том случае отвечает седловой точке в смешанных стратегиях, когда она является точкой максимума задачи

$$A)\ z = \nu \to \max \tag{4.8}$$

$$\begin{cases} a_{11}p_1 + a_{21}p_2 + ... + a_{n1}p_n \leqslant \nu, \\ a_{12}p_1 + a_{22}p_2 + ... + a_{n2}p_n \leqslant \nu, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1m}p_1 + a_{2m}p_2 + ... + a_{nm}p_n \leqslant \nu; \\ p_1 + p_2 + ... + p_n = 1; \end{cases} \tag{4.9}$$

Здесь ν — независимая переменная, которая, вообще говоря, не подчиняется условию неотрицательности.

Для второго игрока смешанная стратегия $q_1^0, q_2^0, ..., q_m^0$ *в том и только в том случае отвечает седловой точке, когда она является точкой минимума задачи*

$$B)\ f = u \to \min \tag{4.10}$$

$$\begin{cases} a_{11}q_1 + a_{12}q_2 + ... + a_{1m}q_m \leqslant u, \\ a_{21}q_1 + a_{22}q_2 + ... + a_{2m}q_m \leqslant u, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \quad q_i \geqslant 0(i = \overline{1,m}) \\ a_{n1}q_1 + a_{n2}q_2 + ... + a_{nm}q_m \leqslant u, \\ q1 + q_2 + ... + q_m = 1; \end{cases} \tag{4.11}$$

Здесь переменная u не подчинена условию неотрицательности.

Форму этих задач линейного программирования нетрудно объяснить. Пусть ν — оптимальный выигрыш первого игрока. Оценивая свою смешанную стратегию $\{p_1,\ p_2,\ \dots\ ,\ p_n\}$, первый игрок стремится, чтобы при каждой чистой стратегии второго игрока его средний выигрыш был не меньше оптимального. Если второй игрок применяет свою стратегию B_j, то первый должен выбирать смешанную стратегию, удовлетворяющей неравенству

$$a_{1j}p_1 + a_{2j}p_2 + ... + a_{nj}p_n \geqslant \nu,\ j = 1, 2, ..., m. \tag{4.12}$$

Аналогично рассуждает и второй игрок. Если u — проигрыш второго игрока, который он стремится минимизировать, то оценивая свою смешанную стратегию $\{q_1,\ q_2,\ \dots\ ,\ q_m\}$, второй игрок должен стремиться к выполнению неравенств

$$a_{i1}q_1 + a_{i2}q_2 + ... + a_{im}q_m \leqslant u, \ i = 1, 2, ..., n. \tag{4.13}$$

Задачи А) и В) двойственны друг другу. Решение их в общем случае громоздко. Но для игр размера $2 \times m$ и $n \times 2$ одну из задач можно решить графически, а решение другой найти с помощью условий дополнительной нежесткости. Кроме того, в общем случае можно так преобразовать задачи А) и В), что они могут быть решены двойственным симплекс-методом в форме, которая рассматривалась выше.

4.5. Графическое решение игр размера $2 \times m$ и $n \times 2$

Рассмотрим игру размера $2 \times m$. Для оптимальной стратегии первого игрока имеем задачу

$$z = \nu \to \max; \tag{4.14}$$

$$\begin{cases} a_{11}p_1 + a_{21}p_2 \geqslant \nu, \\ a_{12}p_1 + a_{22}p_2 \geqslant \nu, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1m}p_1 + a_{2m}p_2 \geqslant \nu, \\ p1 + p_2 = 1; \ p_1, p_2 \geqslant 0 \end{cases} \tag{4.15}$$

Поскольку $p_2 = 1 - p_1$, то, подставляя это выражение вместо p_2 в остальные ограничения, получим задачу с двумя переменными p_1 и ν

$$z = \nu \to \max; \tag{4.16}$$

$$\begin{cases} (a_{11} - a_{21})p_1 + a_{21} - \nu \geqslant 0, \\ (a_{12} - a_{22})p_1 + a_{22} - \nu \geqslant 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ (a_{1m} - a_{2m})p_1 + a_{2m} - \nu \geqslant 0; \\ 0 \leqslant p_1 \leqslant 1. \end{cases} \tag{4.17}$$

Эту задачу можно решать графически. Для задач размером $n \times 2$ точно также преобразуется задача для второго игрока

$$f = u \to \min; \tag{4.18}$$

$$\begin{cases} a_{11}q_1 + a_{12}q_2 \leqslant u, \\ a_{21}q_1 + a_{22}q_2 \leqslant u, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{n1}q_1 + a_{n2}q_2 \leqslant u, \\ q_1 + q_2 = 1; \ q_1, q_2 \geqslant 0. \end{cases} \tag{4.19}$$

	B_1	B_2	B_3	B_4
A_1	2	1	5	3
A_2	1	3	4	0,5

Пример 4.5. Решить игру с платежной матрицей
Соответствующая задача линейного программирования имеет вид:

$$z = \nu \to \max; \tag{4.20}$$

$$\begin{array}{r} q_1 \\ q_2 \\ q_3 \\ q_4 \\ u \end{array} \begin{cases} 2p_1 + p_2 \geqslant \nu \\ p_1 + 3p_2 \geqslant \nu \\ 5p_1 + 4p_2 \geqslant \nu \\ 3p_1 + 0,5p_2 \geqslant \nu \\ p_1 + p_2 = 1 \end{cases} \tag{4.21}$$

$$p_1, p_2 \geqslant 0 \tag{4.22}$$

Исключим переменную p_2; p_2=1-p_1 :

$$z = \nu \to max; \tag{4.23}$$

$$\begin{cases} p_1 - \nu \geqslant -1, \\ -2p_1 - \nu \geqslant -3, \\ p_1 - \nu \geqslant -4, \\ 2,5p_1 - \nu \geqslant -0,5; \end{cases} \tag{4.24}$$

$$0 \leqslant p_1 \leqslant 1. \tag{4.25}$$

Полученную задачу решаем графически (рис. 4.1).

Точкой максимума является точка C, удовлетворяющая системе уравнений

$$\begin{cases} p_1 - \nu = -1, \\ 2p_1 + \nu = 3. \end{cases} \tag{4.26}$$

Отсюда $p_1^0 = \frac{2}{3}; p_2^0 = \frac{1}{3}$ — смешанная стратегия первого игрока в седловой точке. Для нахождения смешанной стратегии второго игрока применим условия дополнительной нежесткости. Двойственная задача имеет вид

$$f = u \to min; \tag{4.27}$$

$$\begin{array}{r} p_1 \\ p_2 \\ \nu \end{array} \begin{cases} 2q_1 + q_2 + 5q_3 + 3q_4 \leqslant u, \\ q_1 + 3q_2 + 4q_3 + 0,5q_4 \leqslant u, \\ q_1 + q_2 + q_3 + q_4 = 1; \end{cases} \tag{4.28}$$

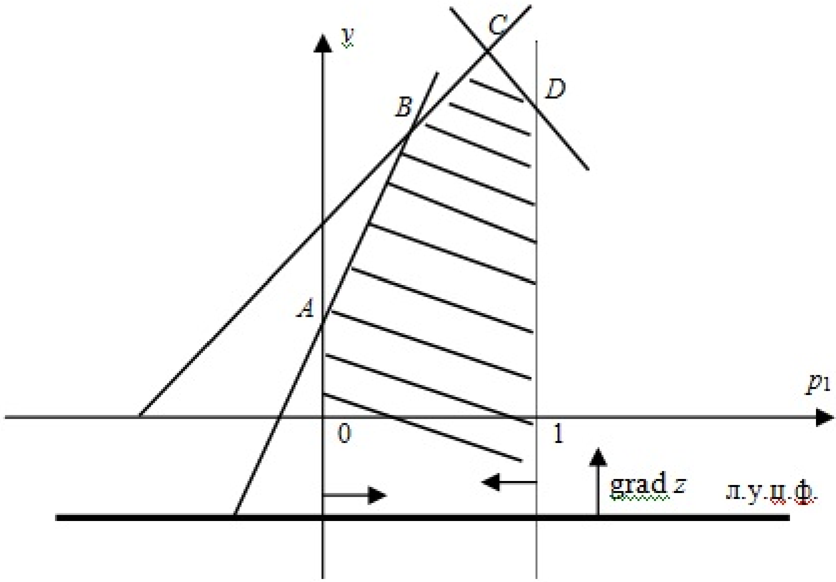

Рис. 4.1. Графическое решение задачи примера 4.5

$$q_1, q_2, q_3, q_4 \geqslant 0. \tag{4.29}$$

Условия дополнительной нежесткости запишутся так:

$$\begin{cases} p_1^0(2q_1^0 + q_2^0 + 5q_3^0 + 3q_4^0 - u^0) = 0, \\ p_2^0(q_1^0 + 3q_2^0 + 4q_3^0 + 0,5q_4^0 - u^0) = 0, \\ \nu^0(q_1^0 + q_2^0 + q_3^0 + q_4^0 - 1) = 0, \\ q_1^0(2p_1^0 + p_2^0 - \nu^0) = 0, \\ q_2^0(p_1^0 + 3p_2^0 - \nu^0) = 0, \\ q_3^0(5p_1^0 + 4p_2^0 - \nu^0) = 0, \\ q_4^0(3p_1^0 + 0,5p_2^0 - \nu^0) = 0, \\ u_1^0(p_1^0 + p_2^0 - 1) = 0. \end{cases} \tag{4.30}$$

Подставим найденные $p_1^0 = \frac{2}{3}$ и $p_2^0 = \frac{1}{3}$ в эти равенства, учитывая, что $\nu^0 = u^0 = \frac{5}{3}$. Четвертое, пятое и восьмое равенства будут тождествами. Из шестого и седьмого следует, что $q_3^0 q_4^0 = 0$. Для определения q_1^0 и q_2^0 имеем

$$\begin{cases} 2q_1^0 + q_2^0 = \frac{5}{3}, \\ q_1^0 + 3q_2^0 = \frac{5}{3}. \end{cases} \tag{4.31}$$

Откуда следует $q_1^0 = \frac{2}{3}; q_2^0 = \frac{1}{3}$. Таким образом, седловая точка игры

$$\vec{p}^0 = \{\frac{2}{3}; \frac{1}{3}\}; \vec{q}^0 = \{\frac{2}{3}; \frac{1}{3}; 0; 0\}. \qquad (4.32)$$

Отметим, что средний выигрыш в седловой точке называется обычно *ценой игры* в смешанных стратегиях. В нашей паре двойственных задач для первого и второго игроков цена игры равна $\nu^0 = u^0$ В примере 4.5 цена игры равна 5/3.

4.6. Решение игры двойственным симплекс-методом

Преобразуем пару двойственных задач *A)* и *B)* так, чтобы удобно было применять двойственный симплекс-метод. Прежде всего, отметим, что седловая точка игры не изменится, если ко всем элементам платежной матрицы прибавить одну и ту же константу. Действительно, если ко всем элементам прибавить константу C, то мы получим новую платежную функцию

$$M_1(\vec{p}, \vec{q}) = \sum_{i,j} (a_{i,j} + C) p_i q_j = \sum_{i,j} a_{i,j} p_i q_j + C \left(\sum_{i,j} p_i q_j \right) = M(\vec{p}, \vec{q}) + C. \qquad (4.33)$$

Таким образом, новая платежная функция отличается от старой постоянным слагаемым C. Если $\{\vec{p}^0, \vec{q}^0\}$ - седловая точка старой платежной функции, то

$$M(\vec{p}, \vec{q}^0) \leqslant M(\vec{p}^0, \vec{q}^0) \leqslant M(\vec{p}^0, \vec{q}). \qquad (4.34)$$

Прибавляя к обеим частям каждого этих неравенств константу C, получим

$$M_1(\vec{p}, \vec{q}^0) \leqslant M_1(\vec{p}^0, \vec{q}^0) \leqslant M_1(\vec{p}^0, \vec{q}). \qquad (4.35)$$

То есть $\{\vec{p}^0, \vec{q}^0\}$ является седловой точкой и новой платежной функции. Прибавляя константу C ко всем элементам матрицы мы всегда можем добиться того, чтобы цена игры была положительна, а седловая точка осталась бы прежней. При этом, поскольку в решениях задач *A)* и *B)* u^0 и v^0 положительны, можно считать переменные подчиненными дополнительным условиям $v > 0$; $u > 0$, которые на точки оптимума задач не повлияют.

Рассмотрим ограничение $p_1 + p_2 + ... + p_n = 1$ и разделим обе его части на v, обозначив $\frac{p_i}{v} = x_i (i = 1, 2, ..., n)$. Поскольку в нашей задаче $v \to max$ мы получим задачу на минимум для новой целевой функции

$$z_1 = \frac{1}{v} = x_1 + x_2 + ... + x_n \to min. \qquad (4.36)$$

Разделив все неравенства системы ограничений задачи *A)* на v, получим для переменных x_i условия

$$\begin{cases} a_{11}x_1 + a_{21}x_2 + ... + a_{n1}x_n \geqslant 1, \\ a_{12}x_1 + a_{22}x_2 + ... + a_{n2}x_n \geqslant 1, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1m}x_1 + a_{2m}x_2 + ... + a_{nm}x_n \geqslant 1; \end{cases} \tag{4.37}$$

$$x_i \geqslant 0, i = \overline{1, n}. \tag{4.38}$$

Производя аналогичные преобразования задачи *B)* и вводя новые переменные $y_j = \frac{q_j}{u}$, получаем новую задачу:

$$f_1 = \frac{1}{u} = y_1 + y_2 + ... + y_m \to max \tag{4.39}$$

$$\begin{cases} a_{11}y_1 + a_{12}y_2 + ... + a_{1m}y_m \leqslant 1, \\ a_{21}y_1 + a_{22}y_2 + ... + a_{2m}y_m \leqslant 1, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{n1}y_1 + a_{n2}y_2 + ... + a_{nm}y_m \leqslant 1; \end{cases} \tag{4.40}$$

$$y_i \geqslant 0, i = \overline{1, m}. \tag{4.41}$$

Мы получили пару симметрично-двойственных задач. Последнюю задачу можно решать симплекс-методом непосредственно после уравнивания неравенств, а решение другой можно будет прочесть по последней симплекс-таблице. Если $y_1^0, y_2^0, ..., y_m^0$ - точка максимума задачи для второго игрока, то вероятности чистых стратегий в седловой точке можно найти по формулам

$$q_1^0 = \frac{y_1^0}{f_{1max}}; q_2^0 = \frac{y_2^0}{f_{1max}}; ...; q_m^0 = \frac{y_m^0}{f_{1max}} \tag{4.42}$$

Если $x_1^0, x_2^0, ..., x_n^0$ — точка минимума задачи для первого игрока, то вероятности стратегий в седловой точке равны

$$p_1^0 = \frac{x_1^0}{z_{1min}}; p_2^0 = \frac{x_2^0}{z_{1min}}; ...; p_m^0 = \frac{x_m^0}{z_{1min}}. \tag{4.43}$$

Отметим, что $f_{min} = z_{max} = A$. Цена игры при этом равна $A - C$, где C — константа, которую мы прибавляли, гарантируя неравенства $v > 0; u > 0$.

Пример 4.6. Наметим нахождение седловой точки в смешанных стратегиях для игры примера 4.4. Прибавим ко всем элементам матрицы число 6. Получим матрицу

$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} 8 & 3 & 10 \\ 3 & 10 & 1 \\ 10 & 1 & 12 \end{pmatrix} \end{array}$$

Построим пару двойственных задач в переменных x_1, x_2, x_3 и y_1, y_2, y_3.

1)$z_1 = x_1 + x_2 + x_3 \to min$;

$$\begin{cases} 8x_1 + 3x_2 + 10x_3 \geqslant 1, \\ 3x_1 + 10x_2 + x_3 \geqslant 1, \\ 10x_1 + x_2 + 12x_3 \geqslant 1; \end{cases}$$

$x_1, x_2, x_3 \geqslant 0$.

2)$f_1 = y_1 + y_2 + y_3 \to max$;

$$\begin{cases} 8y_1 + 3y_2 + 10y_3 \leqslant 1, \\ 3y_1 + 10y_2 + y_3 \leqslant 1, \\ 10y_1 + y_2 + 12y_3 \leqslant 1; \end{cases}$$

$y_1, y_2, y_3 \geqslant 0$.

Вторую задачу можно решить симплекс-методом непосредственно после уравнивания неравенств системы ограничений.

4.7. Доминирование и дублирование стратегий. Упрощение игры

Если платежная матрица такова, что каждый элемент некоторой строки с номером i не меньше соответствующего элемента строки с номером k и, по меньшей мере, один ее элемент строго больше соответствующего элемента строки с номером k, то говорят, что стратегия A_i первого игрока доминирует над его стратегией A_k. Очевидно, что стратегия первого игрока, для которой есть доминирующая, не может быть оптимальной чистой стратегией для него, или входить в его оптимальную смешанную стратегию с ненулевой вероятностью. Таким образом, такую стратегию можно исключить из рассмотрения, вычеркнув из матрицы строку с номером k. Аналогично, если каждый элемент столбца с номером j платежной матрицы не больше соответствующего элемента столбца с номером r и, по меньшей мере, один его элемент строго меньше соответствующего элемента столбца с номером r, то говорят, что стратегия B_j второго игрока доминирует над его стратегией B_r. При поиске оптимальных стратегий столбец с номером r можно вычеркнуть.

Вообще говоря, в платежной матрице могут быть одинаковые строки или столбцы. Соответствующие стратегии называются дублирующими. Очевидно, что дублирующие стратегии целесообразно отбросить.

Выявление доминирующих и дублирующих стратегий позволяет произвести упрощение игры, то есть сократить размеры ее платежной матрицы, исходя из того очевидного факта, что стратегии, над которыми есть доминирующие, применяться в игре не будут, а избавиться от дублирования стратегий можно в виду их полной взаимозаменяемости. Упрощение игры обычно производят в следующем порядке. Вначале с точки зрения первого игрока выявляют его стратегии, над которыми есть доминирующие, а также его дублирующие стратегии. Соответствующие строки платежной матрицы вычеркивают. Для сокращенной матрицы проделывают то же самое, но с точки зрения второго игрока. Упрощение игры считается завершенным, если на очередном шаге в платежной матрице будут отсутствовать доминирующие и дублирующие стратегии для первого и второго игроков.

Замечание 4.1. *Если исходная платежная матрица имеет седловую точку в чистых стратегиях, то процесс упрощения игры приведет к матрице размера* 1×1*, состоящей из одного элемента (седловой точки).*

Пример 4.7. Рассмотрим игру с платежной матрицей

$$\begin{pmatrix} 2 & 4 & 7 & 5 \\ 7 & 6 & 8 & 7 \\ 5 & 3 & 4 & 1 \\ 7 & 6 & 0 & 9 \end{pmatrix}$$

Нетрудно видеть, что стратегия первого игрока A_2 доминирует над стратегиями A_1 и A_3, которые можно отбросить. В оставшейся матрице с точки зрения второго игрока стратегия B_2 доминирует над стратегиями B_1 и B_4. Отбрасывая последние две стратегии, получим упрощенную игру

	$\boldsymbol{B}_2$	$\boldsymbol{B}_3$
$\boldsymbol{A}_2$	6	8
$\boldsymbol{A}_4$	6	0

Возвращаясь на точку зрения первого игрока, мы видим, что в последней игре стратегия A_2 доминирует над стратегией A_4. Отбрасывая последнюю стратегию и сравнивая оставшиеся стратегии второго игрока, приходим к выводу, что стратегия B_3 должна быть отброшена. Получилась игра размера 1×1 и седловая точка (A_2, B_2) в чистых стратегиях.

4.8. Типичный пример применения теории игр в экономике

Пусть предприятие может выпускать три вида продукции: A, B и C. Прибыль от реализации продукции зависит от уровня спроса, о котором ничего неизвестно, но он принимает три возможных значения. Прибыль от реализации каждого изделия вида A, B или C в зависимости от уровня спроса представлена в таблице:

Продукция \ **Спрос**	**I**	**II**	**III**
A	8	3	10
B	3	10	1
C	10	1	12

Эту ситуацию можно рассматривать как игру двух игроков, если предприятие будет считать, что спрос ведет себя самым невыгодным для него образом. Рассматривая соответствующую игру в смешанных стратегиях, нетрудно найти седловую точку, в которой вероятности стратегий p_1^0, p_2^0, p_3^0 первого игрока (предприятия) укажут удельные веса каждого вида продукции, которые обеспечат предприятию оптимальную гарантированную прибыль при самом неблагоприятном поведении спроса. Игры, подобные рассмотренной, носят название *игр с природой.* Решение в рамках матричной игры с нулевой суммой позволяет принимать решения, являющиеся наилучшими при самых неблагоприятных условиях. Однако при более благоприятных условиях это решение может оказаться далеко не лучшим.

4.9. Другие разновидности игр

В настоящее время теория игр является довольно развитым разделом математики, содержащим различные обобщения матричных игр с нулевой суммой. Рассмотрим кратко основные направления таких обобщений. Отказ от абсолютной противоположности интересов игроков (от антагонистичности игры) приводит к неантагонистическим играм, простейшими из которых являются *биматричные игры.* Биматричной игрой называется игра двух игроков с ненулевой суммой, в которой выигрыши задаются двумя матрицами отдельно для каждого из игроков. В каждой матрице строка соответствует стратегии первого игрока, а столбец — стратегии второго игрока. При этом на пересечении строки и столбца в первой матрице находится выигрыш первого игрока, а во второй — второго. Если считать, что первый игрок имеет n стратегий, $i = 1, 2, ..., n$, а у второго игрока имеется m стратегий, $j = 1, 2, ..., m$, то выигрыши первого и второго игроков соответственно задаются матрицами

$$A = \begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{ij} & \dots & a_{im} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nm} \end{pmatrix}, B = \begin{pmatrix} b_{11} & \dots & b_{1j} & \dots & b_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ b_{i1} & \dots & b_{ij} & \dots & b_{im} \\ \dots & \dots & \dots & \dots & \dots \\ b_{n1} & \dots & b_{nj} & \dots & b_{nm} \end{pmatrix} \tag{4.44}$$

Обычная матричная игра является частным случаем биматричной при $a_{ij} = -b_{ij}$. Будем по-прежнему называть полный набор вероятностей $\vec{p} = \{p_1, p_2, ..., p_n\}$ применения первым игроком своих чистых стратегий смешанной стратегией первого игрока, и $\vec{q} = \{q_1, q_2, ..., q_m\}$ — смешанной стратегией второго игрока. Тогда средние выигрыши первого и второго игроков соответственно равны

$$M_1(\vec{p}, \vec{q}) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} p_i q_j, M_2(\vec{p}, \vec{q}) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij} p_i q_j. \tag{4.45}$$

Вместо понятия седловой точки здесь вводится понятие *точки равновесия биматричной игры.* Это пара смешанных стратегий первого и второго игроков $\vec{p}^0 = \{p_1^0, p_2^0, ..., p_n^0\}$ и $\vec{q}^0 = \{q_1^0, q_2^0, ..., q_n^0\}$, отступление от которых не выгодно обоим игрокам, то есть, для которых выполнены неравенства

$$M_1(\vec{p}^0, \vec{q}^0) \geqslant M_1(\vec{p}, \vec{q}^0); M_2(\vec{p}^0, \vec{q}^0) \geqslant M_1(\vec{p}^0, \vec{q}) \tag{4.46}$$

при любых смешанных стратегиях $\vec{p}$ и $\vec{q}$.

Теорема 4.2. *Всякая биматричная игра имеет, хотя бы одну, точку равновесия.*

Существуют способы нахождения точек равновесия, однако эти способы значительно сложнее, чем для матричных игр с нулевой суммой.

Другим естественным направлением обобщения матричных игр являются *бесконечные игры*, то есть игры, в которых хотя бы один из игроков имеет бесконечное количество возможных стратегий. При формализации таких игр будем обозначать их через $\Gamma(X,\ Y,\ a_1,\ a_2)$, где X, Y — множества возможных стратегий первого и второго игроков соответственно, а $a_i(x,y)$ — выигрыш i-го игрока. Часто каждую стратегию из множеств X и Y можно взаимно однозначно сопоставить определенному числу из единичного интервала (0,1). В этом случае функции $a_i(x,y)$ являются функциями двух аргументов, определенных на внутренности единичного квадрата, а игру называют игрой на единичном квадрате.

Бесконечная игра называется антагонистической, если $a_2(x,y) = -a_1(x,y) = -a(x,y)$. Такую игру обозначим Γ(X, Y, a). *Бесконечные антагонистические игры* изучены лучше остальных и во многом аналогичны матричным играм с нулевой суммой, хотя и имеют свои особенности. Их можно анализировать как в чистых, так и в смешанных стратегиях. Назовем чистой нижней ценой игры величину

$$\alpha = \max_{x\in X} \min_{y\in Y} a(x,y), \tag{4.47}$$

а чистой верхней ценой игры величину

$$\beta = \min_{y\in Y} \max_{x\in X} a(x,y). \tag{4.48}$$

Для матричных игр величины α и β всегда существуют, а для бесконечных игр они могут не существовать. Если $\alpha = \beta = V$, то такая игра имеет решение в чистых стратегиях, т. е. оптимальной стратегией первого игрока является выбор числа $x_0 \in X$ и второго игрока — числа $y_0 \in Y$, при которых $a(x_0,y_0) = V$. В этом случае V называется ценой игры, а (x_0,y_0) — седловой точкой игры в чистых стратегиях.

Если игра $\Gamma(X,\ Y,\ a)$ не имеет седловой точки в чистых стратегиях, то оптимальные стратегии можно искать среди смешанных стратегий. Рассмотрим игру на единичном квадрате. Пусть $F(x)$ — функция распределения вероятностей применения чистых стратегий первым игроком, т. е., если ξ — чистая стратегия первого игрока, то

$$F(x) = P\{\xi \leqslant x\}, \tag{4.49}$$

где P $\{\xi \leqslant x\}$ означает вероятность того, что случайно выбранная чистая стратегия ξ не будет превосходить x. Аналогично рассматривается функция распределения вероятностей применения чистых стратегий η вторым игроком

$$G(y) = P\{\eta \leqslant y\}. \tag{4.50}$$

Функции $F(x)$ и $G(y)$ называются смешанными стратегиями соответственно первого и второго игроков. Определим средний выигрыш первого игрока, как математическое ожидание случайной величины $a(\xi,\eta)$

$$E(F,G) = M(a(\xi,\eta)) = \int\limits_0^1 \int\limits_0^1 a(x,y)dF(x)dG(y). \tag{4.51}$$

В антагонистической бесконечной игре $\Gamma(X,\ Y,\ a)$ пара смешанных стратегий $F^*(x)$ и $G^*(y)$ соответственно для первого и второго игроков образует седловую точку в смешанных стратегиях, если для любых смешанных стратегий $F(x)$ и $G(y)$ справедливы соотношения

$$E(F, G^*) \leqslant E(F^*, G^*) \leqslant E(F^*, G). \tag{4.52}$$

В отличие от матричных игр с нулевой суммой, для бесконечных антагонистических игр седловая точка в смешанных стратегиях существует не всегда.

Теорема 4.3. *Всякая бесконечная антагонистическая игра двух игроков* Γ*(X, Y, a) с непрерывной функцией выигрышей a(x,y) на единичном квадрате имеет решение (игроки имеют оптимальные смешанные стратегии, отвечающие седловой точке в смешанных стратегиях).*

Коснемся еще одного направления обобщения понятия игры. Реальные конфликтные ситуации часто приводят к формализации в виде игры с количеством игроков больше двух, например, n игроков. Такие игры называются *играми n участников* (игроков). Поскольку в них участвуют не менее трех игроков, то возможны два варианта правил:

1) игрокам не разрешается вступать в соглашения,
2) игрокам разрешается вступать в соглашения.

В первом случае каждый игрок должен самостоятельно и независимо от желаний других выбирать свои стратегии с целью максимального увеличения своего выигрыша, т. е. игрокам не разрешается вступать к коалиции. Поэтому такая игра называется *бескоалиционной.* Во втором случае некоторые игроки могут по соглашению объединяться (кооперироваться) в действиях против других игроков, образовывая коалиции с целью максимизации выигрыша коалиции. Такие игры называются коалиционными или (в некоторых случаях) кооперативными. Методы исследования игр n участников существенно зависят от возможности образования или запрета коалиций.

Подробнее с математической теорией игр можно ознакомиться по книгам [25], [34].

4.10. Задачи теории статистических решений

Близкой по идеям и методам к теории игр является теория статистических решений. В отличие от теории игр второй игрок здесь вовсе не имеет никаких интересов, не пытается противодействовать, но выбор им своей стратегии заранее неизвестен. Как правило, в роли этого второго игрока выступает комплекс объективных условий, который обычно называют *природой.* С такой ситуацией мы уже встречались. Для принятия решения здесь нужно иметь какую-нибудь информацию о поведении природы. Если предположить, что она ведет себя самым неблагоприятным образом, то решать задачу можно так же, как в теории игр. Но обычно такое решение в итоге оказывается не самым выгодным. Анализ матрицы выигрышей здесь часто заменяют анализом матрицы рисков.

Под риском для стратегии A_i при данном комплексе условий Π_i понимают величину $r_{ij} = \max\limits_k (a_{kj})$ - a_{ij}. Например, при игре с природой, имеющей матрицу выигрышей

	Π_1	Π_2	Π_3	Π_4
A_1	1	4	5	9
A_2	3	8	4	3
A_3	4	6	6	2
$\max\limits_k a_{kj}$	4	8	6	9

соответствующая матрица рисков имеет такой вид:

	Π_1	Π_2	Π_3	Π_4
A_1	3	4	1	0
A_2	1	0	2	6
A_3	0	2	0	7

Риск—это плата за отсутствие информации. В нашем примере $r_{21} = 1$; $r_{24} = 6$, тогда как выигрыши a_{ij} в обоих случаях одинаковы. Если бы мы узнали, что природа находится в состоянии Π_1, то наш максимальный выигрыш получим, применив стратегию A_3, и этот выигрыш равен 4.

Если информации нет, и мы применяем стратегию A_2, то рискуем недополучить сумму $r_{21} = 1$. При состоянии природы Π_4 и применении стратегии A_2 риск больше ($r_{24} = 6$). При выборе стратегий в игре с природой естественно стремиться или к получению максимального выигрыша, или к минимизации риска, сопровождающего выбор решения.

Пусть о состоянии природы имеется информация в виде вероятностей q_1, q_2, ..., q_m соответствующих комплексов условий Π_1, Π_2, ..., Π_m. Если нам нужно выбрать решение в чистых стратегиях, то естественно выбирать ту стратегию A_i для которой среднее значение выигрыша максимально. Причём выбрать i так, чтобы

$$a_i = \sum_{j=1}^{m} a_{ij} q_j \to max. \tag{4.53}$$

В другом случае своя стратегия A_i выбирается из соображений минимальности среднего риска $r_i = \sum\limits_{j=1}^{m} r_{ij}\ q_j \to \min$. Можно показать, что оба способа выбора решения приводят к одинаковому результату.

Задача выбора решения ставится и в смешанных стратегиях. В этом случае приходится решать простые задачи линейного программирования. Если $\{p_1, p_2, ..., p_n\}$ — искомая смешанная стратегия, то она должна быть или точкой максимума задачи

$$z = \sum_{i=1}^{n} (\sum_{j=1}^{m} a_{ij} q_j) p_i \to max, \tag{4.54}$$

$$p_1 + p_2 + ... + p_n = 1, \tag{4.55}$$

$$p_i \geqslant 0, i = 1, 2, ..., n; \tag{4.56}$$

или точкой минимума другой задачи:

$$z_r = \sum_{i=1}^{n} (\sum_{j=1}^{m} r_{ij} q_j) p_i \to min, \tag{4.57}$$

$$p_1 + p_2 + ... + p_n = 1, \tag{4.58}$$

$$p_i \geqslant 0, i = 1, 2, ..., n; \tag{4.59}$$

Очень часто вероятности q_j в принципе существуют, но неизвестны. Иногда при этом пользуются принципом недостаточного основания Лапласа, который состоит в том, что $q_1 = q_2 = ...\ q_m = \dfrac{1}{m}$.Однако чаще всего предпринимают меры для определения хотя бы ориентировочных значений этих вероятностей. При этом можно воспользоваться, например, методом экспертных оценок. В некоторых случаях для определения статистических данных о состоянии природы производят специальные эксперименты, используют анализ уже имеющегося статистического материала.

В случае, когда вероятности состояний природы либо вообще не существуют, либо не поддаются оценке даже приближенно, объективный выбор решения становится невозможным. Здесь все зависит от точки зрения на ситуацию, от позиции принимающего решение лица. Опишем несколько возникающих при этом критериев при принятии решений в чистых стратегиях.

Максиминнный критерий Вальда. Природа здесь рассматривается как разумный агрессивный противник, стремящийся создать самые неблагоприятные условия. Выбор оптимальной стратегии производится так же, как и при анализе игры в чистых стратегиях, то есть выбирается стратегия, гарантирующая в любом случае выигрыш не меньший, чем $\alpha = \max_i \min_j a_{ij}$. Принимающее решение лицо здесь находится на точке зрения "крайнего пессимизма".

Критерий минимаксного риска Сэвиджа. Здесь в качестве оптимальной стратегии выбирается стратегия, при которой величина риска в наихудших условиях минимальна. При этом находится величина $S = \min_i \max_j r_{ij}$.

Оценка стратегий производится тоже с точки зрения "крайнего пессимизма но "пессимизм"понимается по-иному.

Критерий пессимизма-оптимизма Гурвица. Выбор стратегии производится в процессе вычисления величины

$$H = \max_i [\theta \min_j a_{ij} + (1-\theta) \max_j a_{ij}], \tag{4.60}$$

где θ —"коэффициент пессимизма выбираемый между нулем и единицей. При $\theta = 1$ критерий Гурвица превращается в критерий Вальда; при $\theta = 0$ — в критерий “крайнего оптимизма выбирающий стратегию, для которой самый большой выигрыш в строке максимален.

Контрольные вопросы и задачи для самостоятельного решения

1. Что обычно называют конфликтной ситуацией? Как строится простейшая модель конфликтной ситуации в виде матричной игры двух игроков с нулевой суммой?
2. Как игроки оценивают свои стратегии в процессе анализа игры в чистых стратегиях? Что такое нижняя и верхняя цены игры в чистых стратегиях?
3. Что такое седловая точка игры в чистых стратегиях?
4. Что такое смешанная стратегия игрока? Дайте определение платежной функции игры.
5. Что такое седловая точка игры в смешанных стратегиях? Сформулируйте теорему фон Неймана о существовании седловой точки игры в смешанных стратегиях.
6. Как строится пара двойственных задач для определения седловой точки в смешанных стратегиях?
7. В чем состоит графический метод решения игр размера $2 \times m$ и $n \times 2$?
8. Как решить игру в смешанных стратегиях двойственным симплекс-методом?
9. Дайте определения биматричной игры и ее точки равновесия.
10. Что такое бесконечная антагонистическая игра, ее чистые и смешанные стратегии?

11. Что вы знаете об играх n лиц? Когда игра называется бескоалиционной, а когда — коалиционной или кооперативной?
12. Что обычно называют игрой с природой? Постановка задачи различные подходы к ее решению. Что такое матрица рисков?
13. В чем состоит принцип недостаточного основания Лапласа?
14. Что представляют собой критерии Вальда, Сэвиджа и Гурвица?

В задачах 4.1 — 4.4 *для заданной платежной матрицы произвести анализ игры в чистых стратегиях.*

4.1.

$$\begin{pmatrix} 2 & -5 & 3 & -1 \\ 4 & 7 & -6 & 8 \\ -9 & 11 & 2 & 3 \\ -4 & 2 & -7 & 12 \end{pmatrix}.$$

4.2.

$$\begin{pmatrix} 2 & 4 & -9 & 4 \\ -5 & 7 & 11 & 2 \\ 3 & -6 & 2 & -7 \\ -1 & 8 & 3 & 12 \end{pmatrix}.$$

4.3.

$$\begin{pmatrix} 1 & -2 & 3 & -4 & 5 \\ 6 & -7 & 8 & -9 & 10 \\ 5 & -4 & 3 & -2 & 1 \end{pmatrix}.$$

4.4.

$$\begin{pmatrix} 4 & -2 & 3 \\ -5 & 6 & -7 \\ 8 & -9 & 12 \end{pmatrix}.$$

В задачах 4.5—4.8 *найти седловую точку игры в смешанных стратегиях, используя графический метод.*

4.5.

$$\begin{pmatrix} 7 & 10 \\ 9 & 6 \\ 8 & 9 \end{pmatrix}$$

Ответ:

$$\begin{cases} \vec{p}^{(0)} = (0; 1/4; 3/4), \\ \vec{q}^{(0)} = (3/4; 1/4). \end{cases}$$

4.6.

$$\begin{pmatrix} 6 & 5 \\ 4 & 6 \\ 2 & 7 \\ 1 & 8 \end{pmatrix}$$

Ответ:

$$\begin{cases} \vec{p}^{(0)} = (7/8; 0; 0; 1/8), \\ \vec{q}^{(0)} = (3/8; 5/8). \end{cases}$$

4.7.

$$\begin{pmatrix} 3 & 5 & 1 & 4 & 6 \\ 7 & 4 & 9 & 5 & 3 \end{pmatrix}$$

Ответ:

$$\begin{cases} \vec{p}^{(0)} = (1/2; 1/2), \\ \vec{q}^{(0)} = (0; 0; 0; 3/4; 1/4). \end{cases}$$

4.8.

$$\begin{pmatrix} 3 & 9 \\ 5 & 4 \\ 1 & 7 \\ 4 & 5 \\ 6 & 3 \end{pmatrix}$$

Ответ:

$$\begin{cases} \vec{p}^{(0)} = (1/3; 0; 0; 0; 2/3), \\ \vec{q}^{(0)} = (2/3; 1/3). \end{cases}$$

В задачах 4.9—4.12 *найти решения игр сведением к парам двойственных задач линейного программирования.*

4.9.

$$\begin{pmatrix} 3 & 4 & 5 \\ 7 & 6 & 4 \end{pmatrix}$$

Ответ:

$$\begin{cases} \vec{p}^{(0)} = (3/5; 2/5), \\ \vec{q}^{(0)} = (1/5; 0; 4/5). \end{cases}$$

4.10.

$$\begin{pmatrix} 4 & 7 \\ 9 & 5 \\ 5 & 9 \\ 6 & 9 \end{pmatrix}$$

Ответ:

$$\begin{cases} \vec{p}^{(0)} = (0; 3/7; 0; 4/7), \\ \vec{q}^{(0)} = (4/7; 3/7). \end{cases}$$

4.11.

$$\begin{pmatrix} 8 & 4 & 7 \\ 6 & 5 & 9 \\ 7 & 7 & 8 \end{pmatrix}$$

Ответ:

$$\begin{cases} \vec{p}^{(0)} = (0; 0; 1), \\ \vec{q}^{(0)} = (0; 1; 0). \end{cases}$$

4.12.

$$\begin{pmatrix} 7 & 6 & 7 & 5 \\ 6 & 7 & 9 & 8 \\ 5 & 8 & 4 & 6 \end{pmatrix}$$

Ответ:

$$\begin{cases} \vec{p}^{(0)} = (1/2; 1/2; 0), \\ \vec{q}^{(0)} = (3/4; 0; 0; 1/4). \end{cases}$$

5. ЭЛЕМЕНТЫ ДИСКРЕТНОГО ПРОГРАММИРОВАНИЯ

5.1. Некоторые задачи, приводящие к требованию целочисленности. Постановка задач дискретного программирования

Среди практически важных задач отыскания условного экстремума линейной функции особое место занимают задачи с требованием целочисленности всех или части переменных. Они получили название задач целочисленного (частично целочисленного) программирования. Исторически первой задачей целочисленного типа явилась, опубликованная венгерским математиком Е. Эгервари в 1932 году, задача о назначении персонала.

Задача о назначениях. Пусть требуется выполнить n видов различных работ и имеется n исполнителей (машин или людей) для их выполнения. Каждый исполнитель может использоваться на любой работе. Производительности исполнителей на различных работах, вообще говоря, различны. Обозначим через c_{ij} производительность i-го исполнителя на j-ой работе. Задача заключается в таком распределении исполнителей по работам, при котором суммарная производительность максимальна.

Построим математическую модель этой задачи. Для описания каждого варианта распределения исполнителей по работам введем переменные x_{ij}, относительно которых условимся, что $x_{ij} = 1$, если i-й исполнитель назначен на j-ю работу, и $x_{ij} = 0$ — в противном случае. Выбор варианта назначения исполнителей на работы равносилен приданию переменным x_{ij} значений 0 или 1. При этом должны выполняться условия

$$\sum_{j=1}^{n} x_{ij} = 1,\ i = 1, 2, ..., n,$$

которые означают, что каждый исполнитель назначается на работу, а также условия

$$\sum_{i=1}^{n} x_{ij} = 1,\ j = 1, 2, ..., n,$$

которые означают, что каждая работа обеспечивается исполнителем. Суммарная производительность при данном варианте назначений выразится суммой

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

Таким образом, получаем следующую математическую модель задачи

$$z = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} \to \max$$

$$\sum_{j=1}^{n} x_{ij} = 1,\ i = 1, 2, ..., n,\ \sum_{i=1}^{n} x_{ij} = 1,\ j = 1, 2, ..., n,$$

$$x_{ij} = \begin{cases} 1 \\ 0 \end{cases},\ \ i = 1, 2, ..., n,\ \ j = 1, 2, ..., n.$$

Последние из условий носят название условий булевости переменных x_{ij}, а задачи математического программирования с такими условиями называются задачами с булевыми переменными. Отметим, что замена условий булевости условиями неотрицательности переменных превращает нашу задачу в обычную закрытую транспортную задачу, если ее сформулировать как задачу на минимум. Если эту задачу решить методом потенциалов, то полученное решение автоматически будет удовлетворять условию булевости. Таким образом, в задаче о назначениях условия булевости переменных можно заменить условиями их неотрицательности. Отметим, что возможен открытый вариант задачи о назначениях, в котором количество исполнителей и видов работ не совпадают. С помощью введения фиктивных исполнителей или фиктивных видов работ эта задача сводится к такой же закрытой задаче, в которой производительности фиктивных пунктов полагаются равными нулю. При решении задачи некоторые исполнители могут оказаться без работы или некоторые виды работ — без исполнителей. При этом по-прежнему условия булевости можно заменить условиями неотрицательности.

Пример 5.1. Фирма объявила набор работников на две появившиеся вакансии. Были поданы заявления от трех претендентов, готовых занять каждую из двух вакансий. Предварительные испытания определили, что производительности претендентов на работе по каждой из двух вакансий задаются матрицей

$$\begin{pmatrix} 4 & 7 & 1 \\ 5 & 3 & 6 \end{pmatrix}.$$

Требуется так заполнить вакансии имеющимися претендентами, чтобы суммарная производительность принятых работников была максимальной. Предполагается, что каждая вакансия заполняется лишь одним претендентом, а каждый претендент может занять не более одной вакансии.

Для составления математической модели задачи введем такие переменные x_{ij}, что $x_{ij} = 1$, если i-й исполнитель назначен на j-ю работу, и

$x_{ij} = 0$ — в противном случае. Получаем следующую задачу на максимум

$$\begin{cases} x_{11} + x_{12} + x_{13} = 1, \\ x_{21} + x_{22} + x_{23} = 1; \end{cases} \quad \begin{cases} x_{11} + x_{21} \le 1, \\ x_{12} + x_{22} \le 1, \\ x_{13} + x_{23} \le 1; \end{cases}$$

$$x_{ij} = \begin{cases} 1 \\ 0 \end{cases}, \quad i = 1, 2, 3, \quad j = 1, 2, 3.$$

После введения фиктивной вакансии мы получаем задачу, в которой условия бул евости можно заменить условиями неотрицательности переменных. Задачу можно рассматривать как транспортную задачу на минимум целевой функции $z = -4x_{11} - 7x_{12} - x_{13} - 5x_{21} - x_{22} - 6x_{23} \to \max$, которую можно записать в виде следующей таблицы данных

		v_1	v_2	v_3
	Запасы	Потребности		
		1	1	1
u_1	1	$\boxed{-4}$	$\boxed{-7}$ 1	$\boxed{-1}$ 0
u_2	1	$\boxed{-5}$ 1	$\boxed{-1}$ 0	$\boxed{-6}$ 0
u_3	1	$\boxed{0}$	$\boxed{0}$	$\boxed{0}$ 1

Находим первое опорное решение методом наименьшей стоимости. Запишем и решим систему уравнений для потенциалов.

$$\begin{cases} u_1 + v_2 = -7, \\ u_1 + v_3 = -1, \\ u_2 + v_1 = -5, \\ u_2 + v_3 = -6, \\ u_3 + v_3 = 0. \end{cases} \quad \begin{cases} u_1 = 0, & v_1 = 0, \\ u_2 = -5, & v_2 = -7, \\ u_3 = 1, & v_3 = -1; \end{cases}$$

Вычислим коэффициенты γ_{ij} для свободных клеток. $\gamma_{11} = -4$, $\gamma_{22} = 11$, $\gamma_{31} = -1$, $\gamma_{32} = 6$. Составим цикл пересчета для клетки x_{31} и произведем сдвиг по этому циклу на единицу. Получим следующую таблицу

		v_1	v_2	v_3
	Запасы	Потребности		
		1	1	1
u_1	1	$\boxed{-4}$ 0	$\boxed{-7}$ 1	$\boxed{-1}$
u_2	1	$\boxed{-5}$ 1	$\boxed{-1}$	$\boxed{-6}$
u_3	1	$\boxed{0}$ 0	$\boxed{0}$	$\boxed{0}$ 1

Снова составим и решим систему уравнений для потенциалов.

$$\begin{cases} u_1 + v_1 = -4, \\ u_1 + v_2 = -7, \\ u_2 + v_1 = -5, \\ u_3 + v_1 = 0, \\ u_3 + v_3 = 0. \end{cases} \quad \begin{cases} u_1 = 0, & v_1 = -4, \\ u_2 = -1, & v_2 = -7, \\ u_3 = 4, & v_3 = -4; \end{cases}$$

Следовательно, $\gamma_{13} = 1$, $\gamma_{22} = 7$, $\gamma_{23} = -1$, $\gamma_{32} = 3$. Составим цикл пересчета для клетки x_{23} и произведем сдвиг по этому циклу. Получим таблицу

		v_1	v_2	v_3
	Запасы	Потребности		
		1	1	1
u_1	1	$\boxed{-4}$ 1	$\boxed{-7}$ 0	$\boxed{-1}$
u_2	1	$\boxed{-5}$	$\boxed{-1}$ 1	$\boxed{-6}$
u_3	1	$\boxed{0}$ 0	$\boxed{0}$	$\boxed{0}$ 1

$$\begin{cases} u_1 + v_1 = -4, \\ u_1 + v_2 = -7, \\ u_2 + v_2 = -1, \\ u_3 + v_1 = 0, \\ u_3 + v_3 = 0. \end{cases} \qquad \begin{cases} u_1 = 0, \quad v_1 = -4, \\ u_2 = 6, \quad v_2 = -7, \\ u_3 = 4, \quad v_3 = -4; \end{cases}$$

$\gamma_{13} = 3$, $\gamma_{21} = -7$, $\gamma_{23} = -8$, $\gamma_{32} = 3$. Производим сдвиг по циклу пересчета для клетки x_{21}.

		v_1	v_2	v_3
	Запасы	Потребности		
		1	1	1
u_1	1	$\boxed{-4}$ 1	$\boxed{-7}$ 0	$\boxed{-1}$
u_2	1	$\boxed{-5}$	$\boxed{-1}$ 1	$\boxed{-6}$
u_3	1	$\boxed{0}$ 0	$\boxed{0}$	$\boxed{0}$ 1

$$\begin{cases} u_1 + v_1 = -4, \\ u_1 + v_2 = -7, \\ u_2 + v_2 = -5, \\ u_3 + v_1 = 0, \\ u_3 + v_3 = 0. \end{cases} \qquad \begin{cases} u_1 = 0, \quad v_1 = -4, \\ u_2 = -1, \quad v_2 = -7, \\ u_3 = 4, \quad v_3 = -4; \end{cases}$$

$\gamma_{13} = 3$, $\gamma_{22} = 7$, $\gamma_{23} = -1$, $\gamma_{32} = 3$. Преобразуем таблицу с помощью цикла пересчета для клетки x_{23}.

		v_1	v_2	v_3
	Запасы	Потребности		
		1	1	1
u_1	1	$\boxed{-4}$ 0	$\boxed{-7}$ 1	$\boxed{-1}$
u_2	1	$\boxed{-5}$	$\boxed{-1}$	$\boxed{-6}$ 1
u_3	1	$\boxed{0}$ 1	$\boxed{0}$	$\boxed{0}$ 0

$$\begin{cases} u_1 + v_1 = -4, \\ u_1 + v_2 = -7, \\ u_2 + v_2 = -6, \\ u_3 + v_1 = 0, \\ u_3 + v_3 = 0. \end{cases} \quad \begin{cases} u_1 = 0, & v_1 = -4, \\ u_2 = -2, & v_2 = -7, \\ u_3 = 4, & v_3 = -4; \end{cases}$$

$\gamma_{13} = 3$, $\gamma_{21} = 1$, $\gamma_{22} = 8$, $\gamma_{32} = 3$. Получена последняя таблица, содержащая решение задачи. Второй претендент получает первую вакансию, третий — вторую, а первый остается безработным.

Не во всякой задаче с условиями целочисленности эти последние условия можно игнорировать. В большинстве случаев они играют очень существенную роль.

Задача о ранце. Имеется вектор ограниченных ресурсов $\vec{b} = \{b_1, b_2, ..., b_m\}$, которые можно использовать для перевозки различных по своим характеристикам грузов. Каждый из этих n видов грузов имеет следующие свойства:

1) неделимость, т.е. для транспортировки груз с номером j может выбираться в количестве кратном единице;
2) определена полезность (или стоимость) c_j единицы груза;
3) известным является расход i-го ресурса для перевозки единицы j-го груза a_{ij}, $\quad i = 1, 2, ..., m, \quad j = 1, 2, ..., n.$

Требуется определить такой набор груза различных видов, при котором общая полезность рейса максимальна. Под полезностью рейса понимается суммарная стоимость перевезенного за рейс груза.

Построим математическую модель задачи. Пусть x_j - количество выбранных для перевозки предметов j-го вида. Требованию неделимости соответствует условие

$$x_j \geq 0, \ \text{— целые числа}, \ \ j = 1, 2, ..., n.$$

Сопоставляя расход ресурсов при транспортировке с их наличным количеством, получаем ограничения

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

Общая полезность рейса определяется значением целевой функции

$$z = \sum_{j=1}^{n} c_j x_j.$$

Частным случаем этой задачи является классическая задача о ранце, в которой переменные x_j подчинены условию булевости. Смысл этих условий заключается в том, что любой из заданного набора предметов (грузов) может быть выбран для перевозки или нет, т.е. x_j может принимать значения 0 или 1.

Часто возникают и другие подобные задачи, которые можно сформулировать в виде следующей задачи на максимум

$$\begin{aligned} & z = \sum_{j=1}^{n} c_j x_j \to \max, \\ & \sum_{j=1}^{n} a_{ij} x_j \ \{\leq = \geq\} b_i, \quad i = 1, 2, ..., m, \\ & x_j \geq 0, \quad j = 1, 2, ..., n, \\ & x_j \ \text{— целые числа}, \quad j = 1, 2, ..., k. \end{aligned} \tag{P}$$

Задачу (P) при $k < n$ будем называть задачей частично целочисленного программирования. При $k = n$ сформулированная задача называется задачей целочисленного программирования. Более общими являются задачи дискретного или частично дискретного программирования. В этих задачах все или часть переменных принимают значения из заранее заданного дискретного множества. Ниже мы в основном будем рассматривать целочисленные задачи.

Основные понятия обычного линейного программирования переносятся и на дискретный случай. Вектор $X = \{x_1, x_2, ..., x_n\}$, удовлетворяющий системе ограничений задачи, называется допустимым решением (планом) задачи. Допустимое решение, доставляющее максимум целевой функции, называют оптимальным решением (планом). Процесс решения задачи дискретного программирования состоит в нахождении оптимального плана.

Казалось бы, естественный путь решения целочисленной задачи состоит в решении соответствующей обычной линейной задачи с последующим округлением компонент оптимального плана $X = \{x_1, x_2, ..., x_n\}$ до ближайших целых чисел. На самом деле такой путь в большинстве случаев не только уводит от оптимума, но даже приводит иногда к недопустимому решению задачи.

Пример 5.2. Рассмотрим задачу целочисленного программирования с двумя переменными

$$
\begin{aligned}
&z = x_1 + x_2 \to \max,\\
&\begin{cases} x_1 + x_2 \le 7,\\ 4x_1 - 5x_2 \le 5,\\ 2x_1 + 11x_2 \le 38; \end{cases}\\
&x_1 \ge 0,\ x_2 \ge 0,\ x_1,\ x_2\ \text{— целые числа.}
\end{aligned}
$$

Отбросим условия целочисленности и решим задачу графически. Экстремум достигается в точках $A\left(\frac{13}{3}, \frac{8}{3}\right)$ и $B\left(\frac{40}{9}, \frac{23}{9}\right)$, а также в любой точке отрезка AB, и равен 7. Округляя значения координат A, получим точку , которая не принадлежит области поиска. Можно показать, что максимум достигается в точках $N(3, 2)$ и $M(2, 3)$ и равен 5.

Рассмотренный пример показывает, что для задач дискретного программирования необходимо применять особые методы оптимизации.

5.2. Методы отсечения. Первый алгоритм Гомори

Методы отсечения сводятся к решению некоторой последовательности специально построенных задач линейного программирования без условий целочисленности. Каждая последующая задача получается из предыдущей добавлением дополнительного линейного ограничения (неравенства), называемого сечением. Если обозначить через З$_0$, З$_1$, З$_2$, ..., З$_l$, ... задачи указанной последовательности, то l-ым сечением называется линейное ограничение, вводимое в задачу З$_{l-1}$ для образования задачи З$_l$, и удовлетворяющее двум условиям:

a) любое целочисленное решение системы ограничений задачи З$_{l-1}$ ему удовлетворяет;
b) найденное нецелочисленное решение задачи З$_{l-1}$ ему не удовлетворяет (“отсекается”).

Рассмотрим задачу целочисленного программирования, которая имеет следующий вид

$$\begin{aligned} & z = c_0 + (\vec{c}, \vec{x}) \to \max, \\ & A\vec{x} = \vec{b}, \;\; \vec{x} \geq \vec{0}, \\ & x_j - \textit{целые}(j = 1, 2, ..., n), \end{aligned} \tag{P_1}$$

то есть является задачей с ограничениями равенствами. Частным случаем такой задачи можно считать задачу вида

$$\begin{aligned} & z = \sum_{j=1}^{n} c_j x_j \to \max, \\ & \sum_{j=1}^{n} a_{ij} x_j \leq b_i \quad i = 1, 2, ..., m, \\ & x_j \geq 0, \;\; x_j \text{ - целые числа}, \;\; j = 1, 2, ..., n, \end{aligned} \tag{P_2}$$

в которой предполагается, что величины b_i и a_{ij} также являются целочисленными. Задачу $(P2)$ легко свести к задаче вида $(P1)$ с помощью уравнивания неравенств. Задачу вида $(P1)$ (или $(P2)$) будем называть канонически целочисленной.

Существует несколько методов построения сечений для целочисленных задач. В этом параграфе мы рассмотрим наиболее простой первый алгоритм Р. Гомори, который применяется к задаче вида $(P1)$, т.е. к канонически целочисленной задаче.

Пусть задача $З_{l-1}$ уже решена симплекс методом и ее решение $\overline{X}_{l-1}$ не удовлетворяет условию целочисленности. Обозначим через a дробную часть числа a, $a = a - [a]$, где $[a]$-целая часть числа a. Пусть k — индекс свободных переменных в последней симплексной таблице. Обозначим через s — номер строки в этой таблице с *наибольшим* значением $\{b_s^*\}$ для свободного члена b_s^*, отвечающего базисной переменной. Тогда сечение Гомори запишется в виде

$$\{b_s^*\} - \sum_k \{a_{sk}^*\} x_k \leq 0. \tag{S_1}$$

Не удовлетворяющее условию целочисленности решение $\overline{X}_{l-1}$ задачи $З_{l-1}$ условию $(S1)$ не удовлетворяет. Действительно, в оптимальном опорном решении свободные переменные $x_k = 0$, а поскольку $\{b_s^*\}>0$, то $(S1)$ на $\overline{X}_{l-1}$ не удовлетворяется. Следовательно, ограничение $(S1)$ удовлетворяет условию b), упомянутому выше. Покажем, что условие $(S1)$ удовлетворяет и условию a). Сначала мы установим, что оно удовлетворяется для любого целочисленного решения системы ограничений задачи

$(P1)$. Рассмотрим i-ое ограничение этой задачи $\sum_{j=1}^{n} a_{ij}x_j = b_i$, и возьмем дробную часть от обеих частей равенства. Поскольку дробная часть суммы не превосходит суммы дробных частей слагаемых ($\{a+b\} \geq \{a\}+\{b\}$), то справедливо неравенство $\sum_{j=1}^{n}\{a_{ij}\}x_j \geq \{b_j\}$. Поскольку x_j — целое неотрицательное число, то $\{a_{ij}x_j\} \leq \{a_{ij}\}x_j$. Таким образом, для любого целочисленного решения выполнено неравенство $\sum_{j=1}^{n}\{a_{ij}\}x_j \geq \{b_i\}$. Это означает, что условие a) выполнено для сечения $(S1)$, построенного после решения задачи $З_0$, при построении задачи $З_1$. Для решения задачи $З_1$ симплекс методом неравенство, задающее сечение, нужно уровнять введением новой переменной. Можно показать, что эту новую переменную можно вывести из числа базисных переменных. В процессе решения задачи $З_1$ эта переменная в число базисных вернуться не может, так как она не входит в выражение для целевой функции. Поэтому мы можем рассматривать задачу $З_1$ в качестве задачи целочисленного программирования с увеличенным на единицу количеством переменных. В силу сказанного выше, построенное после решения задачи $З_1$ сечение будет обладать свойствами a), и b). То же можно сказать и о задаче $З_{l-1}$ при любом l.

Алгоритм решения канонически целочисленной задачи состоит из последовательности итераций, каждая из которых включает следующие пункты:

1) решается задача $З_{l-1}$ и находится ее оптимальное решение;
2) если решение этой задачи удовлетворяет условию целочисленности, то процесс заканчивается, и мы получаем оптимальный план. В противном случае переходим к пункту 3).
3) На основании последней симплекс-таблицы задачи $З_{l-1}$ записываем сечение Гомори $(S1)$.
4) Добавление ограничения из предыдущего пункта к условиям задачи $З_{l-1}$ приводит к задаче $З_l$, после чего снова возвращаемся к пункту 1) с увеличенным на единицу l.

Отметим, что при некотором l задача $З_l$ может оказаться не имеющей решения. Очевидно, что в случае недопустимости задачи $З_l$ и исходная задача целочисленного программирования тоже является недопустимой. Нетрудно показать, что если целевая функция задачи $З_l$ неограниченна на области допустимых решений, то это же справедливо и для исходной задачи целочисленного программирования.

Можно доказать, что в результате конечного числа итераций рассмотренного алгоритма мы придем к задаче $З_l$, которая либо имеет целочисленное решение, либо не имеет решений. В первом случае мы получаем решение исходной задачи целочисленного программирования, а во втором — устанавливаем, что эта задача не имеет решения.

Пример 5.3. Требуется найти оптимальный план задачи

$$z = x_1 + 4x_2 \to \max,$$
$$\begin{cases} -x_1 + 2x_2 \le 2, \\ 3x_1 + 2x_2 \le 6; \end{cases}$$
$$x_i \ge 0, \;\; x_i \text{ - целые числа} (i = 1, 2).$$

Отбрасывая условие целочисленности, решаем симплекс-методом задачу З$_0$:

$$z = x_1 + 4x_2 \to \max,$$
$$\begin{cases} -x_1 + 2x_2 + x_3 = 2, \\ 3x_1 + 2x_2 + x_4 = 6; \end{cases}$$
$$x_1 \ge 0, \;\; i = 1, 2.$$

Таблица 1

Базисные переменные	Свободные члены	x_1	$\downarrow x_2$	x_3	x_4
$\leftarrow x_3$	2	-1	2	1	0
x_4	6	3	2	0	1
z	0	-1	-4	0	0

Таблица 2

Базисные переменные	Свободные члены	$\downarrow x_1$	x_2	x_3	x_4
x_2	1	-1/2	1	1/2	0
$\leftarrow x_4$	4	4	0	-1	1
z	4	-3	0	2	0

Таблица 3

Базисные переменные	Свободные члены	x_1	x_2	x_3	x_4
x_2	3/2	0	1	3/8	1/8
x_1	1	1	0	-1/4	1/4
z	7	0	0	5/4	3/4

Из последней симплекс-таблицы получаем оптимальное решение задачи З$_0$: $\overline{X}_0 = \{1;\ 3/2;\ 0;\ 0\}$. Это решение не является целочисленным. Построим задачу З$_1$. Единственная строка с нецелочисленным значением b$_s$ — первая строка. Запишем сечение Гомори (S1)

$$\{3/2\} - (\{3/8\}x_3 + \{1/8\}x_4) \leq 0,$$

то есть.

$$1/2 - (3/8)x_3 - (1/8)x_4 \leq 0.$$

Преобразуя и уравнивая неравенство, получим

$$-(3/8)x_3 - (1/8)x_4 + u1 = -1/2.$$

Добавляя это ограничение к ограничениям задачи З$_0$, получим задачу З$_1$. Решение этой последней задачи удобно начинать, добавив к табл. 3 еще одну строку. Получена таблица, отвечающая псевдоплану. Применим двойственный симплекс метод (метод последовательного уточнения оценок).

Таблица 1

Базисн. перем.	Своб. члены	x_1	x_2	x_3	x_4	u_1
x_2	3/2	0	1	3/8	1/8	0
x_1	1	1	0	1/4	1/4	0
u$_1$	-1/2	0	0	[-3/8]	-1/8	1
z	7	0	0	5/4	3/4	0

По правилу выбора разрешающего элемента в двойственном симплекс методе выбираем элемент (-3/8) и переходим к следующей таблице

Таблица 2

Базисн. перем.	Своб. члены	x_1	x_2	x_3	x_4	u_1
x_2	1	0	1	0	0	1
x_1	4/3	1	0	0	1/3	-2/3
x_3	4/3	0	0	1	1/3	-8/3
z	16/3	0	0	0	1/3	10/3

Полученный псевдоплан оказывается одновременно оптимальным решением задачи З$_1$. Оно не является целочисленным. Строим соответствующее сечение Гомори

$$\{4/3\} - (\{1/3\}x_4 + \{-2/3\}u_1) \leq 0.$$

Поскольку {4/3}=1/3; {1/3}=1/3; {-2/3}=1/3, после преобразований и уравнивания неравенства получаем ограничение

$$-(1/3)x_4 - (1/3)u_1 + u_2 = -1/3.$$

Добавляя это ограничение, получаем задачу З$_2$, которую можно записать в виде следующей обобщенной симплекс таблицы, отвечающей псевдоплану.

Таблица 1

Базисн. перем.	Своб. члены	x_1	x_2	x_3	x_4	u_1	u_2
x_2	1	0	1	0	0	1	0
x_1	4/3	1	0	0	1/3	-2/3	0
x_3	4/3	0	0	1	1/3	-8/3	0
u$_2$	-1/3	0	0	0	[-1/3]	-1/3	1
z	16/3	0	0	0	1/3	10/3	0

Выбирая разрешающий элемент, переходим к следующей таблице.

Таблица 2

Базисн. перем.	Своб. члены	x_1	x_2	x_3	x_4	u_1	u_2
x_2	1	0	1	0	0	1	0
x_1	1	1	0	0	0	-1	1
x_3	1	0	0	1	0	-3	1
x_4	1	0	0	0	1	1	-3
z	5	0	0	0	0	3	1

Получили псевдоплан, который является допустимым и одновременно оптимальным и целочисленным. Таким образом, получаем решение исходной задачи целочисленного программирования: z_{max}=5; точка максимума: $x_1 = x_2 = 1$.

5.3. Второй алгоритм Гомори

Рассмотренный первый алгоритм Гомори применим не к любым задачам целочисленного программирования, а лишь к канонически целочисленным. Существуют методы отсечения применимые как для целочисленных (вообще говоря, не канонически целочисленных), так и для частично целочисленных задач. Самым распространенным из этих методов является второй алгоритм Гомори, который мы рассмотрим без обоснования. Как и ранее строим последовательность задач линейного программирования без условий целочисленности З$_0$, З$_1$, З$_2$, ..., З$_l$, ... Каждая последующая получается из последней симплекс таблицы предыдущей задачи с помощью добавления еще одного условия (сечения Гомори второго рода). Если в последней симплекс таблице задачи З$_{l-1}$ решение не удовлетворяет условиям целочисленности, т.е. в столбце свободных членов есть дробные числа b_i^*, которые по условию должны быть целыми, то выбрав строку с максимальным $\{b_i^*\}$ $(max\{b_i^*\} = \{b_s^*\})$, построим дополнительное ограничение вида

$$\{b_s^*\} - \sum_k \gamma_{sk} x_k \leq 0. \qquad (S_2)$$

где γ_{sk} определяются следующим образом:

1) для x_k, которые могут принимать нецелочисленные значения,

$$\gamma_{sk} = \begin{cases} a^*_{sk} & \text{при } a^*_{sk} \geq 0. \\ \dfrac{\{b^*_s\}}{1-\{b^*_s\}}[a^*_{sk}] & \text{при } a^*_{sk} < 0; \end{cases}$$

2) для x_k, которые могут принимать только целочисленные значения,

$$\gamma_{sk} = \begin{cases} \{a^*_{sk}\} & \text{при } \{a^*_{sk}\} \leq \{b^*_s\}. \\ \dfrac{\{b^*_s\}}{1-\{b^*_s\}}[1-\{a^*_{sk}\}] & \text{при } a^*_{sk} < 0; \end{cases}$$

После добавления ограничения (S2) к системе ограничений задачи $З_{l-1}$ получаем задачу $З_l$, которую снова решаем симплекс методом. Итерации повторяются до тех пор, пока решение очередной задачи не будет удовлетворять условиям целочисленности или не обнаружится, что очередная задача не имеет решения. Как и в случае первого алгоритма Гомори, можно показать, что при применении сечения Гомори второго рода задача решается за конечное число итераций.

Пример 5.4. Пусть требуется найти решение следующей задачи

$$z = x_1 + 4x_2 \to max,$$

$$\begin{cases} 2x_1 + x_2 \leq \dfrac{19}{3}, \\ x_1 + 3x_2 \leq 4; \end{cases}$$

$$x_i \geq 0, \quad x_i - \text{ целые числа } (i = 1, 2),$$

которая не является канонически целочисленной, т. к. после уравнивания неравенств получаем систему ограничений

$$\begin{cases} 2x_1 + x_2 + x_3 = \dfrac{19}{3}, \\ x_1 + 3x_2 = 4; \end{cases}$$

где переменная x_3 может принимать нецелочисленные значения.

Отбрасывая условия целочисленности, получаем следующую симплекс таблицу

Таблица 1

Базисные переменные	Свободные члены	x_1	$\downarrow x_2$	x_3	x_4
x_3	19/3	2	1	1	0
$\leftarrow x_4$	4	1	3	0	1
z	0	-1	-4	0	0

Таблица 2

Базисные переменные	Свободные члены	x_1	x_2	x_3	x_4
x_3	5	5/3	0	1	-1/3
x_2	4/3	1/3	1	0	1/3
z	2/3	1/3	0	0	4/3

Получили последнюю симплекс таблицу задачи $З_0$. Переменная x_2 должна быть целочисленной, но в точке максимума задачи $З_0$ принимает дробное значение. Построим сечение (S2)

$$1/3 \leq (1/3)x_1 + (1/3)x_4 \text{ или } -(1/3)x_1 - (1/3)x_4 + u_1 = -1/3.$$

Добавим в последнюю симплекс таблицу строку, отвечающую построенному сечению. Получим таблицу, отвечающую псевдоплану.

Таблица 1

Базисн. перем.	Своб. члены	x_1	x_2	x_3	x_4	u_1
x_3	5	5/3	0	1	-1/3	0
x_4	4/3	1/3	1	0	1/3	0
u_1	-1/3	[-1/3]	0	0	-1/3	1
z	16/3	1/3	0	0	4/3	0

Выбираем разрешающий элемент и переходим к новой таблице.

Таблица 2

Базисн. перем.	Своб. члены	x_1	x_2	x_3	x_4	u_1
x_3	10/3	0	0	1	-2	5
x_4	1	0	1	0	0	1
x_1	1	1	0	0	1	-3
z	5	0	0	0	1	1

Эта таблица является и последней для задачи $З_1$. Она дает решение и исходной задачи целочисленного программирования, т. к. в этой задаче допускается дробное значение переменной x_3. Таким образом, для исходной задачи точка максимума $x_1 = x_2 = 1$, $z_{max} = 5$.

5.4. Метод ветвей и границ

Этот метод часто называют методом последовательного анализа вариантов. В этом методе также решается ряд задач без условия целочисленности, но эти задачи не образуют единой последовательности.

Рассмотрим задачу вида (Р). Начнем опять с задачи $З_0$, которая получается из исходной отбрасыванием условий целочисленности. Пусть переменная x_k, которая должна принимать целочисленное значение, принимает в точке максимума задачи $З_0$ дробное значение x_{k_0} . Нетрудно найти такое целое n, что выполнены неравенства $n<x_{k_0}<n+1$. Рассмотрим две задачи $З_1$ и $З_2$, полученные из $З_0$ добавлением условий

$$1)x_k \geq n+1, \text{ и } 2)x_k \leq n$$

соответственно. Для каждой из этих задач найдем решение. Если решение одной из задач будет удовлетворять условиям целочисленности задачи (Р), то эту задачу будем называть прозондированной. Одна из задач может оказаться и не имеющей решения. В этом случае задачу тоже будем называть прозондированной. Если же задача имеет решение, не удовлетворяющее условиям целочисленности, то ее «разветвляют», как и раньше, на две задачи по какой-нибудь переменной, которая должна быть целочисленной, но принимает в полученном решении дробное значение. Прозондированные задачи далее уже не ветвятся. По ходу решения задач определяются оптимальные значения целевой функции. Все сказанное можно изобразить в виде схемы, приведённой на рис. 5.1.

Вычисления заканчиваются после того, когда все очередные задачи окажутся прозондированными. Если решается задача на максимум, то

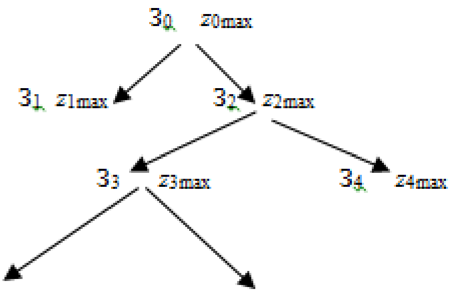

Рис. 5.1. Схема метода ветвей и границ

среди всех разрешимых прозондированных задач выбирается та, у которой самое большое значение z_{max}. Эта задача и дает точку максимума исходной задачи.

Метод ветвей и границ является универсальным методом, применимым как к полностью целочисленным задачам, так и к частично целочисленным, и даже к произвольным задачам дискретного программирования. Коммерческие задачи обычно решаются этим методом. В некоторых случаях, однако, этот метод трудно реализовать из-за большого объема машинного счета.

Следует отметить, что большинство типов задач дискретного программирования имеют экспоненциальную сложность (см. п. 1.7). Исключение составляют задача о назначениях, задача о ранце с булевыми переменными и некоторые другие, имеющие полиномиальную сложность. Таким образом, при больших размерностях большинство задач дискретного программирования за разумное время решить нельзя.

Контрольные вопросы и задачи для самостоятельного решения

1. Какие задачи называют задачами линейного целочисленного (частично целочисленного, дискретного, частично дискретного) программирования?
2. Сформулируйте задачу о назначениях. В чем заключается связь между задачей о назначениях и транспортной задачей?
3. Сформулируйте задачу о ранце.
4. В чем заключается основная идея методов отсечений? Опишите первый алгоритм Гомори для полностью целочисленных задач.
5. Как строится сечение Гомори второго рода?
6. Какова роль двойственного симплекс метода (метода последовательного уточнения оценок) при применении сечений Гомори первого и

второго рода?
7. В чем заключается метод ветвей и границ?
8. то можно сказать о сложности задач дискретного программирования?

Решить задачи о назначениях 5.1, 5.2.

5.1. Имеется пять механизмов и пять видов работ, которые этими механизмами могут выполняться. Матрица эффективностей имеет вид

$$\begin{pmatrix} 2 & 3 & 3 & 5 & 4 \\ 4 & 2 & 4 & 6 & 2 \\ 2 & 2 & 2 & 4 & 3 \\ 4 & 3 & 4 & 3 & 5 \\ 0 & 1 & 0 & 3 & 0 \end{pmatrix}.$$

Найти оптимальное распределение механизмов по видам работ.

Ответ: Первый механизм назначается на пятую работу, второй — на четвертую, третий — на первую, четвертый — на третью, пятый — на вторую.

5.2. Для выполнения четырех видов землеройных работ используются четыре экскаватора четырех различных типов. Производительность экскаватора i-ого типа при выполнении j-ой работы задается матрицей

$$\begin{pmatrix} 0,9 & 0,6 & 0,7 & 0,9 \\ 0,7 & 0,8 & 0,9 & 0,8 \\ 0,8 & 0,6 & 0,8 & 0,9 \\ 0,8 & 0,7 & 0,9 & 0,7 \end{pmatrix}.$$

Найти такое распределение экскаваторов по видам работ, которое обеспечивает максимальную производительность.

Ответ: Экскаватор первого типа следует назначить на первый вид работ, второго типа — на второй, третьего типа — на четвертый и четвертого типа — на третий вид работ.

С помощью графического истолкования найти решения задач целочисленного программирования 5.3 — 5.5.

5.3.

$$z = x_1 + x_2 \to max,$$

$$\begin{cases} x_1 + x_2 \leq 7, \\ 4x_1 - 5x_2 \leq 5, \\ 2x_1 + 11x_2 \leq 38; \end{cases}$$

$$x_1 \geq 0, x_2 \geq 0, \quad x_1, x_2 - \text{ целые}$$

Ответ: $z_{max} = 5$; точек максимума две: M(3,2) и N(2,3).

5.4.

$$z = 3x_1 + x_2 \to min,$$

$$\begin{cases} -4x_1 + x_2 \leq 29, \\ 3x_1 - x_2 \leq 15, \\ 5x_1 + 2x_2 \geq 38; \end{cases}$$

$$x_1, x_2 \geq 0, \ x_1, x_2 - \text{ целые}$$

Ответ: $z_{min} = 19$; точка минимума (0;19).

5.5.

$$z = 5x_1 + 7x_2 \to min,$$

$$\begin{cases} -3x_1 + 14x_2 \leq 78, \\ 5x_1 - 6x_2 \leq 26, \\ x_1 + 4x_2 \geq 25; \end{cases}$$

$$x_1, x_2 \geq 0, \ x_1, x_2 - \text{ целые}$$

Ответ: $z_{min} = 52$; точка минимума (2; 6).

Пользуясь алгоритмами Гомори найти решения задач 5.6 — 5.10.

5.6.

$$z = 2x_1 + x_2 \to max;$$

$$\begin{cases} 6x_1 + 4x_2 + x_3 = 24, \\ 3x_1 - 3x_2 + x_4 = 9, \\ -x_1 + 3x_2 + x_5 = 3; \end{cases}$$

$$x_i \geq 0, \quad x_i - \text{ целые} (i = \overline{1,5}).$$

Ответ: $z_{max} = 7$; точка максимума (3; 1; 2; 3; 3).

5.7.

$$z = 3x_1 + 2x_2 \to max;$$

$$\begin{cases} x_1 + x_2 + x_3 = 13, \\ x_1 - x_2 + x_4 = 6, \\ -3x_1 + x_2 + x_5 = 9; \end{cases}$$

$$x_i \geq 0, \quad x_i\text{– целые}(i = \overline{1,5}).$$

Ответ: $\text{z}_{max} = 35$; точка максимума (9; 4; 0; 1; 32).

5.8.

$$z = 7x_1 + x_2 \to max;$$

$$\begin{cases} 9x_1 + 4x_2 + x_3 = 110, \\ 11x_1 - 3x_2 + x_4 = 24, \\ 2x_1 - 7_2 + x_5 = 15; \end{cases}$$

$$x_i \geq 0, \quad x_i\text{– целые}(i = \overline{1,5}).$$

Ответ: $\text{z}_{max} = 84$; точка максимума (12; 0; 2; 108; 9).

5.9.

$z = 60x_1 + 70x_2 + 120,4x_3 + 130x_4 \to max;$

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 16, \\ x_1 + 1,85x_2 + x_3 + x_4 \leq 16, \\ 4x_1 + 6,9x_2 + 10x_3 + 13x_4 \leq 100, \\ 6,3x_1 + 5x_2 + 4x_3 + 3x_4 \leq 100, \\ x_1 + x_2 + x_3 + x_4 \geq 10; \end{cases}$$

$x_i \geq 0, \quad x_i$– целые$(i = 1,4)$.

Ответ: $\text{z}_{max} = 1322,4$; точка максимума (10; 0; 6; 0).

5.10.

$z = 5x_1 + 2x_2 - 3x_3 + 2x_4 + 3x_5 - 3x_6 \to max.$

$$\begin{cases} 5x_1 + 6x_2 + 4x_3 + 2x_4 - 3x_5 + 5x_6 = 11, \\ 5x_1 + 5x_2 + 7x_3 + 3x_5 + 5x_6 = 10, \\ 2x_1 + 2x_2 + 2x_3 + 3x_5 = 4; \end{cases}$$

$x_i \geq 0, \quad x_i$– целые$(i = \overline{1,6})$.

Ответ: $\text{z}_{max} = 7$; точка максимума (1; 1; 0; 0; 0; 0).

6. НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

6.1. Постановка задачи и общие понятия

1.Пример возникновения задачи нелинейного программирования в экономике. Пусть предприятие изготавливает два вида продукции P_1 и P_2. Если обозначить прибыль от реализации единицы продукции P_1 через c_1, а P_2 — через c_2, то суммарную прибыль предприятия можно записать в виде

$$z = c_1x_1 + c_2x_2,$$

где x_1 и x_2 - плановые количества продукции P_1 и P_2.

Предприятие стремится к максимуму функции z при ограничениях, определяемых условиями производства. Ранее мы считали величины c_1 и c_2 константами, независящими от x_1 и x_2. Однако такое предположение оправдано лишь при измененииx_1 иx_2 в некоторых пределах, да и то приближенно. Величины c_1 и c_2 зависят от x_1 и x_2. Чем больше продукции попадает на рынок, тем меньше цены на эту продукцию. Предположим, что истинная прибыль от реализации единицы продукции P_1 равна $c_1-\dfrac{x_1}{b_1}$, а продукции P2 — $c_2-\dfrac{x_2}{b_2}$, где c_1, b_1 и c_2, b_2 —некоторые константы. Общая прибыль определяется формулой

$$z = \left(c_1 - \frac{x_1}{b_1}\right)x_1 + \left(c_2 - \frac{x_2}{b_2}\right)x_2$$

Эта целевая функция уже не является линейной, т.к. в ее выражении присутствуют квадраты переменных. Задача на максимум для такой функции уже не может быть решена методами линейного программирования.

2.Общая формулировка задачи нелинейного программирования и связанные с ней общие понятия. Задачей нелинейного программирования мы будем называть задачу

$$z = f(x_1, x_2, \ldots, x_n) \to \max(\min);$$

$$\begin{cases} \varphi_1(x_1, x_2, \ldots, x_n) = 0, \\ \varphi_2(x_1, x_2, \ldots, x_n) = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \varphi_k(x_1, x_2, \ldots, x_n) = 0, \\ \varphi_{k+1}(x_1, x_2, \ldots, x_n) \le 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \varphi_m(x_1, x_2, \ldots, x_n) \le 0, \end{cases}$$

В этой задаче часть ограничений является равенствами, а часть– неравенствами. Все входящие в формулировку функции могут быть нелинейными. В дальнейшем часто совокупность значений переменных $x_1, x_2, \ldots, x_n$ мы будем рассматривать как точку $\vec{x}$. в n-мерном евклидовом пространстве $R^n \left(\vec{x} \in R^n \right)$. Множество точек R^n которые удовлетворяют системе ограничений задачи, обозначим через M. Сокращенно задачу можно записать так:

$$f(\vec{x}) \to \max(\min)$$

$$\vec{x} \in \mathrm{M} \subset R^n$$

Отметим, что конкретный вид системы ограничений можно изменять, не изменяя множества M, т.е. формулировать задачу в различных формах. Каждое равенство можно заменить системой неравенств

$$\begin{cases} \varphi_i(\vec{x}) \leqslant 0, \\ -\varphi_i(\vec{x}) \leqslant 0. \end{cases}$$

Поэтому можно считать, что система ограничений состоит только из неравенств. С другой стороны, каждое неравенство $\varphi_i(\vec{x}) \leq 0$ можно заменить эквивалентным равенством $\varphi_i(\vec{x}) + x_{n+1}^2 = 0$ вводя новую переменную x_{n+1}. Отметим, что в нелинейном программировании словие неотрицательности переменной не играет особой роли и включается (если оно есть) в систему ограничений "на общих основаниях". Количество ограничений тоже можно изменить, не меняя, по сути, задачу. Например, всегда можно систему ограничений записать в виде одного равенства. Для этого вначале преобразуем систему ограничений в систему равенств

$$\begin{cases} \varphi_1(x_1, x_2, \ldots, x_n) = 0, \\ \varphi_2(x_1, x_2, \ldots, x_n) = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \varphi_m(x_1, x_2 \ldots, x_n) = 0. \end{cases}$$

Эта же система эквивалентна одному уравнению

$$\Phi(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{m} {\varphi_i}^2(x_1, x_2, \ldots, x_n) = 0$$

Здесь, однако, следует заметить, что такое уменьшение количества ограничений часто не упрощает, а усложняет решение задачи.

Определение 6.1. Локальным минимумом задачи называется такая точка $\vec{x}^{(0)}$ удовлетворяющая системе ограничений ($\vec{x}^{(0)} \in M$), для которой выполнено неравенство $f(\vec{x}^0) \leqslant f(\vec{x})$. Здесь $\vec{x}$ – произвольная точка множества $M \cap K_\varepsilon(\vec{x}^{(0)})$, $K_\varepsilon(\vec{x}^{(0)})$ – ε - окрестность точки $\vec{x}^{(0)}$.

Под ε - окрестностью точки $\vec{x}^{(0)} = \left\{x_1^0, x_2^0, \ldots, x_n^0\right\}$ понимается множество точек $\vec{x} = \left\{x_1, x_2 \ldots, x_n\right\}$ для которых

$$\sqrt{\left(x_1 - x_1^0\right)^2 + \left(x_2 - x_2^0\right)^2 + \cdots + \left(x_n - x_n^0\right)^2} < \varepsilon.$$

Другими словами, в точке локального минимума значение целевой функции не больше, чем во всех точках множества M достаточно близких к точке $\vec{x}^{(0)}$.

Аналогично определяется понятие локального максимума, т.е. точки $\vec{x}^{(0)}$, для которой выполняется неравенство

$$f(\vec{x}^0) \geqslant f(\vec{x}) \text{ при } \vec{x} \in M \cap K_\varepsilon(\vec{x}^{(0)}).$$

Определение 6.2. Глобальным минимумом (максимумом) задачи называется точка $\vec{x} \in M$ для которой $f(\vec{x}^{(0)}) \leqslant f(\vec{x})(f(\vec{x}^{(0)}) \geqslant f(\vec{x}))$ для любой точки $\vec{x} \in M$.

Решением задачи нелинейного программирования является точка глобального минимума или максимума (экстремума). В настоящее время регулярных общих методов нахождения глобальных экстремумов не существует. Правда, существуют методы случайного и эволюционного поиска (генетические алгоритмы), которые дают, вообще говоря, приемлемое, а не оптимальное решение. Здесь мы рассматриваем методы нахождения точек локального экстремума. Эти методы особенно важны для так называемых *одноэкстремальных задач*.

Определение 6.3. Задача на минимум нелинейного программирования называется одноэкстремальной, если каждый ее локальный минимум одновременно является и глобальным. Аналогично понимается одноэкстремальность задачи на максимум.

Всякая задача линейного программирования является одноэкстремальной.

6.2. Метод множителей Лагранжая

1. Задача на условный экстремум. Необходимые условия экстремума. Рассмотрим задачу нелинейного программирования с ограничениями равенствами:

$$z = f(x_1, x_2, \ldots, x_n) \to \max(\min);$$

$$\begin{cases} \varphi_1(x_1, x_2, \ldots, x_n) = 0, \\ \varphi_2(x_1, x_2, \ldots, x_n) = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \varphi_m(x_1, x_2 \ldots, x_n) = 0. \end{cases}$$

Будем рассматривать в каждой точке $\vec{x} \in M$ $m+1$ вектор градиентов функций $f(x), \varphi_1(\vec{x}), \varphi_2(\vec{x}), \ldots, \varphi_m(\vec{x})$. Обозначим

$$\nabla f = gradf, \nabla\varphi_1 = grad\varphi_1, \ldots, \varphi_m = grad\varphi_m.$$

Справедливо следующее необходимое условие того, что точка $\vec{x}^{(0)} = \left\{x_1^0, x_2^0, \ldots, x_n^0\right\}$ является точкой локального экстремума задачи.

Теорема 6.1. Если точка $\vec{x}^{(0)} \in M$ является точкой локального экстремума, то в этой точке система из $m+1$ векторов градиентов $f(x), \varphi_1(\vec{x})$, $\varphi_2(\vec{x}), \ldots, \varphi_m(\vec{x})$ является линейно зависимой системой, то есть найдутся такие $m+1$ чисел, $\lambda_0, \lambda_1, \ldots, \lambda_m$, не все равные нулю, для которых в точке $\vec{x}^{(0)}$ выполнено равенство

$$\lambda_0 \nabla f(\vec{x}^{(0)}) + \sum_{i=1}^{m} \lambda_i \nabla\varphi_i(\vec{x}^{(0)}) = 0$$

Для большинства задач в каждой точке множества M система векторов $\nabla\varphi_1, \nabla\varphi_2, \ldots, \nabla\varphi_m$ линейно независима. Последнее условие назовем *"условие A"*.

Функцией Лагранжа задачи называется функция

$$L(\vec{x}, \vec{\lambda}) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i \varphi_i(\vec{x}),$$

которая является функцией $m+n$ переменных

$$\lambda_1 \lambda_2 \ldots \lambda_m, x_1, x_2 \ldots, x_n$$

Теорема 6.2. Пусть выполнено условие A. Для того чтобы точка $\vec{x}^{(0)}$ была точкой локального экстремума, необходимо, чтобы существовали такие числа $\left\{\lambda_1^0, \lambda_2^0, \ldots, \lambda_n^0\right\}$ при которых все частные производные функции Лагранжа по всем ее n + m переменным равнялись бы нулю:

$$\frac{\partial L}{\partial x_i}(\vec{x}^0, \vec{\lambda}^0) = 0 (i = 1, 2, \ldots, n); \frac{\partial L}{\partial \lambda_j}(\vec{x}^0, \vec{\lambda}^0) = 0 (j = 1, 2, \ldots, m)$$

Эта теорема дает необходимые условия локального экстремума. Если систему равенств теоремы 2 рассматривать, как систему уравнений относительно неизвестных $x_1, x_2 \ldots, x_n, \lambda_1 \lambda_2 \ldots \lambda_m$, то мы можем выделить точки, подозреваемые на экстремум.

Пример 6.1. Исходная задача:

$$z = 2x_1^2 + 3x_2^2 + x_3^2 \to \max,$$

$$\begin{cases} x_1 - x_2 + x_3 + 5 = 0, \\ x_1 + 4x_2 - x_3 - 1 = 0. \end{cases}$$

Составим функцию Лагранжа:

$$L(x_1, x_2, x_3, \lambda_1 \lambda_2) = 2x_1^2 + 3x_2^2 + x_3^2 +$$

$$+\lambda_1(x_1 - x_2 + x_3 + 5) + \lambda_2(x_1 + 4x_2 - x_3 - 1).$$

Необходимые условия имеют вид

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 4x_1 + \lambda_1 + \lambda_2 = 0, \\ \dfrac{\partial L}{\partial x_2} = 6x_2 - \lambda_1 + 4\lambda_2 = 0, \\ \dfrac{\partial L}{\partial x_3} = 2x_3 + \lambda_1 - \lambda_2 = 0, \\ \dfrac{\partial L}{\partial \lambda_1} = x_1 - x_2 + x_3 + 5, \\ \dfrac{\partial L}{\partial \lambda_2} = x_1 + 4x_2 - x_3 - 1 = 0. \end{cases}$$

Исключая λ_1 и λ_2 из первых трех уравнений, получим

$$\begin{cases} 6x_1 - 9x_2 - 9x_3 = 0 \\ x_1 - x_2 + x_3 + 5 = 0, \\ x_1 + 4x_2 - x_3 - 1 = 0. \end{cases}$$

Решим систему методом Гаусса:

$$\left(\begin{array}{ccc|c} 6 & -9 & -2 & 0 \\ 1 & -1 & 1 & -5 \\ 1 & 4 & -1 & 1 \end{array}\right) \Longleftrightarrow \left(\begin{array}{ccc|c} 0 & -9 & -2 & 30 \\ 1 & -1 & 1 & -5 \\ 0 & 5 & -2 & 6 \end{array}\right) \Longleftrightarrow \left(\begin{array}{ccc|c} 0 & 1 & \dfrac{8}{3} & -10 \\ 1 & -1 & 1 & -5 \\ 0 & 5 & -2 & 6 \end{array}\right)$$

$$\Longleftrightarrow \left(\begin{array}{ccc|c} 0 & 1 & \dfrac{8}{3} & -10 \\ 1 & 0 & \dfrac{11}{3} & -15 \\ 0 & 0 & -\dfrac{46}{3} & 1 \end{array}\right) \Longleftrightarrow \left(\begin{array}{ccc|c} 0 & 1 & \dfrac{8}{3} & -10 \\ 1 & 0 & \dfrac{11}{3} & -15 \\ 0 & 1 & 0 & \dfrac{66}{23} \end{array}\right)$$

$$\Longleftrightarrow \left(\begin{array}{ccc|c} 0 & 1 & 0 & -\dfrac{406}{23} \\ 1 & 0 & 0 & -\dfrac{587}{23} \\ 0 & 0 & 1 & \dfrac{66}{23} \end{array}\right)$$

Координаты точки, подозреваемой на экстремум:

$$x_1 = -\frac{587}{23}; x_2 = -\frac{406}{23}; x_3 = \frac{66}{23}$$

2. Достаточные условия условного экстремума. Пусть в некоторой задаче необходимые условия дают точку $(x_1^0, x_2^0 \ldots, x_n^0,\ \lambda_1^0 \lambda_2^0 \ldots \lambda_m^0)$. Построим матрицу вторых частных производных функции Лагранжа по переменным $x_1, x_2, \ldots, x_n$,

$$H = \begin{pmatrix} \dfrac{\partial^2 L}{\partial x_1^2} & \dfrac{\partial^2 L}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 L}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 L}{\partial x_2 \partial x_1} & \dfrac{\partial^2 L}{\partial x_2^2} & \cdots & \dfrac{\partial^2 L}{\partial x_2 \partial x_n} \\ \dfrac{\partial^2 L}{\partial x_n \partial x_1} & \dfrac{\partial^2 L}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 L}{\partial x_n^2} \end{pmatrix}$$

которая называется матрицей Гесса функции Лагранжа. В точке $(x_1^0, x_2^0, \ldots, x_n^0, \lambda_1^0 \lambda_2^0, \ldots, \lambda_m^0)$ матрица Гесса — это обычная числовая матрица размера $n \times n$. Перейдем из стационарной точки $M(x_1^0, x_2^0, \ldots, x_n^0, \lambda_1^0 \lambda_2^0 \ldots \lambda_m^0)$ в точку M_1 с координатами $x_1^0 + \nabla x_1, x_2^0 + \nabla x_2, \ldots, x_n^0 + \nabla x_n, \lambda_1^0 \lambda_2^0 \ldots \lambda_m^0$ По формуле Тейлора приращение функции Лагранжа

$$\nabla L = L_1(M_1) - L(M) = \frac{1}{2} \sum_{i,j} \frac{\partial^2 L}{\partial x_i \partial x_j}\bigg|_M \nabla x_i \nabla x_j + o(|\nabla \vec{x}|^2) =$$

$$= \frac{1}{2} d^2 L + o(|\nabla \vec{x}|^2)$$

имеет главную часть, которая является квадратичной формой с матрицей, совпадающей с матрицей Гесса функции Лагранжа H. Если ∇x_i $(i = 1, 2, \ldots, n)$ выбираются так, что точка M_1 тоже удовлетворяет системе условий задачи, то $\nabla L = \nabla f$ и знак величины ∇f (при достаточно малых ∇x_i) определяется знаком этой же квадратичной формы. Таким образом, справедлива следующая

Теорема 6.3. Если в исследуемой точке все собственные числа матрицы Гесса функции Лагранжа положительны, то в точке $\left\{x_1^0, x_2^0, \ldots, x_n^0\right\}$

находится локальный минимум. Если же эти собственные числа отрицательны, то $\vec{x}^{(0)}=\left\{x_1^0, x_2^0, \ldots, x_n^0\right\}$ являетсяточкой локального максимума.

Продолжим рассмотрение примера 1. Матрица Гесса здесь легко вычисляется, во всех точках:

$$H = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Поскольку собственные числа диагональной матрицы равны диагональным элементам, эти собственные числа положительны, т.е. в точке $\left\{-\dfrac{587}{23}; -\dfrac{406}{23}; \dfrac{66}{23}\right\}$ локальный минимум. Можно показать, что она является и точкой глобального минимума.

Приведенные достаточные условия являются довольно грубыми. Их можно уточнить следующим образом. При смещении из точки области допустимых значений в другую ее точку приращения ∇x_i не являются независимыми. Взаимосвязь между приращениями ∇x_i задается системой уравнений

$$\begin{cases} \dfrac{\partial \varphi_1}{\partial x_1}\nabla x_1 + \dfrac{\partial \varphi_1}{\partial x_2}\nabla x_2 + \cdots + \dfrac{\partial \varphi_1}{\partial x_n}\nabla x_n = 0, \\ \dfrac{\partial \varphi_2}{\partial x_1}\nabla x_1 + \dfrac{\partial \varphi_2}{\partial x_2}\nabla x_2 + \cdots + \dfrac{\partial \varphi_2}{\partial x_n}\nabla x_n = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial \varphi_m}{\partial x_1}\nabla x_1 + \dfrac{\partial \varphi_m}{\partial x_2}\nabla x_2 + \cdots + \dfrac{\partial \varphi_m}{\partial x_n}\nabla x_n = 0. \end{cases} \tag{6.1}$$

Поскольку условие считается выполненным, ранг матрицы этой системы равен m (обычно $m < n$) Поэтому из переменных ∇x_i $(i = 1, 2, \ldots, n)$ можно выбрать m базисных, которые выражаются через n–m свободных переменных. Если в выражении для квадратичной формы в формуле для ∇L базисные переменные выразить через свободные переменные, мы получим новую квадратичную форму с матрицей H' порядка n–m.

Теорема 6.4. Если в исследуемой точке все собственные числа матрицы H' положительны, то в точке $\left\{x_1^0, x_2^0, \ldots, x_n^0\right\}$ находится локальный минимум. Если же эти собственные числа отрицательны, то $\left\{x_1^0, x_2^0, \ldots, x_n^0\right\}$ является точкой локального максимума. Если у матрицы H' имеется хотя бы два собственных числа разных знаков, то точка $\left\{x_1^0, x_2^0, \ldots, x_n^0\right\}$ не является ни точкой минимума, ни точкой максимума.

Последняя теорема не охватывает случая, когда у матрицы H' нет собственных чисел разных знаков, но есть равные нулю собственные числа. В этом случае нужны дополнительные исследования с помощью дифференциалов высших порядков.

Пример 6.2. Рассмотрим простейшую задачу

$$f = x_1^2 - x_2^2 \to min$$

$$3x_1 + 4x_2 - 7 = 0$$

Построим функцию Лагранжа

$$L(x_1, x_2, \lambda) = x_1^2 - x_2^2 + \lambda(3x_1 + 4x_2 - 7)$$

Применим необходимые условия условного экстремума.

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 4x_1 + 3\lambda = 0, \\ \dfrac{\partial L}{\partial x_2} = -2x_2 + 4\lambda = 0, \\ \dfrac{\partial L}{\partial \lambda} = 3x_1 + 4x_2 - 7 = 0. \end{cases}$$

Получим стационарную точку $M(-3, 4, 2)$. Матрица Гесса функции Лагранжа в этой точке

$$H = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

не позволяет применить теорему 3. Переменные $\nabla x_1, \nabla x_2$ удовлетворяют системе, состоящей из одного уравнения $3\nabla x_1 + 4\nabla x_2 = 0$, или $\nabla x_2 = -3/4\nabla x_1$. Квадратичная форма $\frac{1}{2}d^2L = \nabla x_1^2 - \nabla x_2^2 = \nabla x_1^2 - (-3/4\nabla x_1)^2 = 7/16\nabla x_1^2$. Матрица H' имеет порядок, равный единице. $H' = (7/16)$, поэтому по теореме 4 заключаем, что точка $(-3, 4)$ является точкой минимума нашей задачи.

3. Критерий Сильвестра. Вычисление собственных чисел матрицы является довольно громоздкой задачей. Однако для применения теорем 3 и 4 сами эти числа не нужны. Матрица квадратичной формы является симметрической и все ее собственные числа вещественны. Для определения знаков собственных чисел симметрической матрицы можно использовать теорему, принадлежащую американскому математику Сильвестру. Рассмотрим симметрическую матрицу

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{31} & a_{32} & \dots & a_{3n} \end{pmatrix}$$

Назовем угловыми минорами этой матрицы следующие определители

$$\nabla_1 = a_{11}, \nabla_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \dots, \nabla_n = detA$$

Теорема 6.5. 1.Для положительности всех собственных чисел симметрической матрицы необходимо и достаточно, чтобы все угловые миноры этой матрицы были положительны.

2. Для отрицательности собственных чисел необходимо и достаточно, чтобы все угловые миноры матрицы были отличны от нуля, их знаки чередовались, и выполнялось условие $\nabla_1 < 0$.

Из этой теоремы и теоремы 6.4 следует, что из положительности угловых миноров матрицы H' вытекает, что в исследуемой стационарной точке локальный минимум, а из чередования знаков этих миноров при $\nabla_1 < 0$ вытекает, что в стационарной точке функции Лагранжа локальный максимум. Если учесть, *что произведение всех собственных чисел симметрической матрицы равняется ее определителю*, получаем, что при $\nabla_{n-m} \neq 0$ отсутствие положительности угловых миноров матрицы H' или указанного выше их знакочередования означает, что в исследуемой точке нет локального экстремума. Случай $\nabla_{n-m} = 0$ ледует рассматривать особо. В этом случае возможно существование пары собственных чисел разных знаков, что гарантирует отсутствие локального экстремума. Возможно также, что нет собственных чисел разных знаков. Поскольку при этом есть нулевые собственные числа, то для установления характера стационарной точки в этом случае следует привлекать дифференциалы функции Лагранжа более высокого порядка.

Пример 6.3. Найти точки условного экстремума функции $z = x_1 x_2^2 x_3^3$ при условии $x_1 + x_2 + x_3 = 12$, лежащие в области $\left\{\vec{x} \in R^3 : x_1, x_2, x_3 > 0\right\}$, и исследовать характер этих точек.

Построим функцию Лагранжа

$$L(\vec{x}, \lambda) = x_1, x_2^2, x_3^3 + \lambda(x_1 + x_2 + x_3 - 12)$$

и найдем ее частные производные

$$\frac{\partial L}{\partial x_1} = x_2^2 x_3^3 + \lambda; \frac{\partial L}{\partial x_2} = 2x_1 x_2 x_3^3 + \lambda; \frac{\partial L}{\partial x_3} = 3x_1 + x_2^2 x_3^3 + \lambda;$$

$$\frac{\partial L}{\partial \lambda} = x_1 + x_2 + x_3 - 12.$$

Приравнивая эти частные производные нулю, получаем стационарную точку $M_0(2,4,6)$; $\lambda = -3456$. Найдем матрицу Гесса в этой точке.

$$H|_{M_0} = \left.\begin{pmatrix} 0 & 2x_2x_3^3 & 3x_2^2x_3^2 \\ 2x_2x_3^3 & 2x_1x_3^3 & 6x_1x_2x_3^2 \\ 3x_2^2x_3^2 & 6x_1x_2x_3^2 & 6x_1x_2^2x_3 \end{pmatrix}\right|_{M_0} = 576\begin{pmatrix} 0 & 3 & 3 \\ 3 & 1 & 3 \\ 3 & 3 & 2 \end{pmatrix}$$

Теорема 6.3 здесь ничего не дает. Второй дифференциал является квадратичной формой

$$576(\nabla x_2^2 + 2\nabla x_3^2 + 6\nabla x_1\nabla x_2 + 6\nabla x_1\nabla x_3 + 6\nabla x_2\nabla x_3)$$

Приращения ∇x_i в силу соотношений (3.1) удовлетворяют равенству $\nabla x_1 = -\nabla x_2 - \nabla x_3$. Поэтому квадратичная форма преобразуется к виду

$$576(-5\nabla x_2^2 - 4\nabla x_3^2 - 6\nabla x_2\nabla x_3)$$

Матрица этой квадратичной формы имеет вид

$$H' = \begin{pmatrix} -5 & -3 \\ -3 & -4 \end{pmatrix}$$

Очевидно, что здесь $\nabla_1 < 0$, $\nabla_2 > 0$. Таким образом, в рассматриваемой области находится только одна точка экстремума, которая является точкой максимума.

6.3. Задачи выпуклого программирования

Определение 6.4. Множество $M \subset R^n$ называется выпуклым, если вместе с любыми двумя его точками $\vec{x}_1$ и $\vec{x}_2$ множеству M принадлежит всякий вектор, имеющий вид $(1-\alpha)\vec{x}_1 + \alpha\vec{x}_2$, $(0 \leqslant \alpha \leqslant 1)$.

При $n = 2$ или 3 векторы вида $(1-\alpha)\vec{x}_1 + \alpha\vec{x}_2$, $(0 \leqslant \alpha \leqslant 1)$ лежат на отрезке, соединяющем точки $\vec{x}_1$ и $\vec{x}_2$.

Определение 6.5. Функцию $f(\vec{x})$ будем называть выпуклой, если она определена на выпуклом множестве $\Omega \subseteq R^n$ и для любых $\vec{x}_1, \vec{x}_2 \in \Omega, \alpha \in (0,1)$ выполнено неравенство

$$f((1-\alpha)\vec{x}_1 + \alpha\vec{x}_2) \leqslant (1-\alpha)f(\vec{x}_1) + \alpha(\vec{x}_2).$$

Нетрудно видеть, что для выпуклой функции, определенной на всем R^n множество точек, удовлетворяющих неравенству $f(\vec{x}) \leqslant C$ с любой константой C, либо выпукло, либо пусто. То же самое можно сказать и относительно системы таких неравенств.

Всякую функцию вида $\mathrm{g}(\vec{x}) = -f(\vec{x})$, где $f(\vec{x})$ - выпуклая функция, обычно называют *вогнутой функцией*. Каждая линейная функция, очевидно, выпукла (и одновременно вогнута).

Задача выпуклого программирования формулируется следующим образом:

$$z = f(x_1, x_2, \ldots, x_n) \to min$$

$$\begin{cases} \varphi_1(x_1, x_2, \ldots, x_n) \leqslant 0, \\ \varphi_2(x_1, x_2, \ldots, x_n) \leqslant 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \varphi_m(x_1, x_2 \ldots, x_n) \leqslant 0. \end{cases}$$

$$\vec{x} = \Big\{x_1, x_2, \ldots, x_n\Big\} \in \Omega$$

где функции $f(\vec{x}), \varphi_1(\vec{x}), \ldots, \varphi_m(\vec{x})$ являются выпуклыми, а Ω – выпуклое множество, которое может в конкретном случае совпадать с R^n.

Задача линейного программирования является частным случаем задачи выпуклого программирования. Нетрудно показать, что *всякая задача выпуклого программирования одноэкстремальна*. Действительно, если $\vec{x}_0 \in M$ - точка локального минимума, то найдется такая окрестность $U(\vec{x}^0)$, что $f(\vec{x}) \geqslant f(\vec{x}^0)$ при всех $\vec{x} \in M \cap U(\vec{x}^0)$. Предположим, что $\vec{x}^0$ не является точкой глобального минимума. Тогда найдется такая точка $\vec{x}^1 \in M$, что $f(\vec{x}^1) < f(\vec{x}^0)$. При любом $\alpha \in (0, 1)$ получим

$$f((1-\alpha)\vec{x}^0 + \alpha x^1) \leqslant (1-\alpha)f(\vec{x}^0) + \alpha f(\vec{x}^1) < f(\vec{x}^0)$$

Однако при достаточно малом α выполнено включение $(1-\alpha)\vec{x}^0 + \alpha\vec{x}^1 \in M \cup U(\vec{x}^0)$. Получаем противоречие. Следовательно, в точке $\vec{x}^0$ достигается глобальный минимум, что и требовалось доказать.

Функцией Лагранжа задачи выпуклого программирования называется функция

$$L_B(\vec{x}, \vec{\lambda}) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i \varphi_i(\vec{x}).$$

Задача выпуклого программирования называется *удовлетворяющей условию Слэйтера*, если существует хотя бы одна точка $\vec{x} \in \Omega$ в которой все неравенства системы ограничений выполняются и являются строгими неравенствами.

Введем множество $E_m = \Big\{\vec{\lambda} \in R^m : \lambda_i \geqslant 0, (i = \overline{1, m})\Big\}$

Теорема 6.6. (теорема Куна-Таккера). Пусть задача выпуклого программирования удовлетворяет условию Слэйтера. Тогда точка $\vec{x}^0 \in \Omega$ в том и только в том случае является точкой минимума задачи, если существуют такие неотрицательные числа $\lambda_1^0, \lambda_2^0, \ldots, \lambda_m^0 \geqslant 0$, $(\vec{\lambda} \in E_m)$ что $(x_1^0, x_2^0 \ldots, x_n^0,\ \lambda_1^0 \lambda_2^0 \ldots \lambda_m^0) \in R^{n+m}$ является седловой точкой функции Лагранжа задачи, то есть для любых $\vec{x}^0 \in \Omega$ и $\vec{\lambda} \in E_m$ выполняются неравенства

$$L_B(\vec{x}^0, \vec{\lambda}) \leqslant L_B(\vec{x}^0, \vec{\lambda}^0) \leqslant L_B(\vec{x}, \vec{\lambda}^0)$$

Таким образом, решение задачи выпуклого программирования сводится к нахождению седловых точек функции Лагранжа. Правда, задача нахождения седловых точек далеко не всегда допускает непосредственное решение. Теорема Куна-Таккера является одной из основ теории двойственности выпуклого программирования, которая выходит за пределы настоящего пособия. Приведем лишь обобщение второй теоремы двойственности, которое будет полезно в дальнейшем.Рассмотрим задачу выпуклого программирования с множеством $\Omega = E_n = \Big\{\vec{x} \in R^n : x_i \geqslant 0, (i = \overline{1, n})\Big\}$

Теорема 6.7. Если в задаче выпуклого программирования целевая функция $f(\vec{x})$ и функции $\varphi_1(\vec{x}), \ldots, \varphi_m(\vec{x})$ являются непрерывно дифференцируемыми, а множество $\Omega = E_n$, то точка $(\vec{x}^0, \vec{\lambda}^0)$ тогда и только тогда является седловой точкой функции Лагранжа задачи, когда в этой точке выполнены следующие соотношения

$$\left.\frac{\partial L_B}{\partial x_j}\right|_{(\vec{x}^0, \vec{\lambda}^0)} \geqslant 0; \qquad x_j^0\ \left.\frac{\partial L_B}{\partial x_j}\right|_{(\vec{x}^0, \vec{\lambda}^0)} = 0; \tag{6.2}$$

$$\left.\frac{\partial L_B}{\partial \lambda_i}\right|_{(\vec{x}^0, \vec{\lambda}^0)} \leqslant 0; \qquad x_j^0\ \left.\frac{\partial L_B}{\partial \lambda_i}\right|_{(\vec{x}^0, \vec{\lambda}^0)} = 0; \tag{6.3}$$

$$x_j^0 \geqslant 0; \qquad j = \overline{1, n}; \qquad \lambda_i^0 \geqslant 0; \qquad i = \overline{1, m}. \tag{6.4}$$

В случае задачи линейного программирования в стандартной форме первые неравенства в (6.3) и (6.4) фактически совпадают с системами ограничений соответственно двойственной и заданной задач, а равенства представляют собой условия дополнительной нежесткости.

6.4. Задачи квадратичного программирования

Одним из наиболее простых и важных классов задач нелинейного программирования являются задачи квадратичного программирования, которые состоят в минимизации (или максимизации) квадратичной функции при линейных ограничениях. Поскольку задача максимизации изменением знака целевой функции сводится к задаче минимизации, мы будем считать, что общая задача квадратичного программирования в векторно-матричной форме записывается в следующем виде

$$f(\vec{x}) = (D\vec{x}, \vec{x}) + (\vec{c}, \vec{x}) \to min,$$

$$A\vec{x} \leqslant \vec{b}, \vec{x} \geq \vec{0}$$

где $\vec{x}$ — неизвестный n-мерный вектор, D — симметричная матрица квадратичной формы размера $n \times m$, A — матрица ограничений размера $m \times n$, $\vec{c}$ — n-мерный вектор, $\vec{b}$ — m-мерный вектор.

Способы решения такой задачи определяются свойствами матрицы D. Если все собственные числа матрицы D неотрицательны, то это задача выпуклого программирования и любой локальный минимум дает ее решение. Если все собственные числа матрицы D неположительны, то мы имеем задачу вогнутого программирования, у которой может быть большое число неэквивалентных локальных минимумов, но глобальный минимум (если он существует) достигается в одной из вершин многогранного множества допустимых значений. Еще более сложный случай, когда собственные числа матрицы D имеют разные знаки. Задача здесь также является многоэкстремальной, но глобальный минимум может достигаться в точке, которая не является вершиной множества допустимых значений. Из сказанного видно, что из всех задач квадратичного программирования простейшими являются задачи выпуклого квадратичного программирования. Для них существует ряд способов решения. Мы коснемся одного из них, который основан на теоремах 6.6, 6.7

Функция Лагранжа задачи выпуклого квадратичного программирования имеет вид

$$L_B = (\vec{x}, \vec{\lambda}) = (D\vec{x}, \vec{x}) + (\vec{c}, \vec{x}) + (\vec{\lambda}, A\vec{x} - \vec{b})$$

Если функция L_B имеет седловую точку $(x_1^0, x_2^0 \ldots, x_n^0,\ \lambda_1^0 \lambda_2^0 \ldots \lambda_m^0) \in R^{n+m}$, то в этой точке выполняются соотношения (6.3), (6.3) и (6.4) еоремы 6.7. Введем уравнивающие неотрицательные переменные u_i и v_i $(j = \overline{1,n}, i = \overline{1,m})$, бращающие неравенства в (6.2), (6.3) в равенства. Тогда соотношения (6.2) — (6.4) запишутся в виде

$$\left.\frac{\partial L_B}{\partial x_j}\right|_{(\vec{x}^0, \vec{\lambda}^0)} - u_j = 0; \qquad j = \overline{1,n} \tag{6.5}$$

$$\left.\frac{\partial L_B}{\partial \lambda_i}\right|_{(\vec{x}^0, \vec{\lambda}^0)} + v_i = 0; \qquad i = \overline{1,m} \tag{6.6}$$

$$x_j^0 u_j^0 = 0 (j = \overline{1,n}) \qquad \lambda_i^0 v_i = 0 (i = \overline{1,m}) \tag{6.7}$$

$$x_j^0 \geqslant 0, u_j^0 \geqslant 0 (j = \overline{1,n}) \qquad \lambda_i^0 \geqslant 0, v_i \geqslant 0 (i = \overline{1,m}) \tag{6.8}$$

Таким образом, для нахождения решения задачи выпуклого квадратичного программирования нужно определить неотрицательное решение системы линейных уравнений (6.5), (6.6), удовлетворяющее условиям (6.7). Система (6.5), (6.6) имеет базисный вид, который, однако, не дает неотрицательного базисного решения, поскольку является недопустимым. Переход к допустимому базисному виду можно произвести с помощью метода искусственного базиса. Вводя искусственные переменные y_k в систему (6.5), (6.6) и решая задачу на максимум функции $z = -\sum\limits_k y_k$ с системой ограничений (6.5), (6.6) получим допустимый базисный вид системы (6.5), (6.6) и соответствующее базисное решение будет неотрицательным решением этой системы, вообще говоря, не удовлетворяющим некоторым из условий (6.7). Можно показать, что в случае разрешимости исходной задачи с помощью конечного числа операций замещения можно прийти к допустимому базисному решению системы (6.5), (6.6) удовлетворяющему всем условиям (6.7) При этом отсутствие среди допустимых базисных решений системы (6.5), (6.6) такого, которое удовлетворяет условиям (6.7), означает неразрешимость исходной задачи.

Пример 6.4. Найти минимальное значение функции

$$f = x_1^2 + 2x_2^2 - 2x_1 - 4x_2$$

при условия

$$\begin{cases} x_1 + 2x_2 \leqslant 8, \\ 2x_1 - x_2 \leqslant 12 \end{cases}$$

$$x_1, x_2 \geqslant 0.$$

Функция f является выпуклой, так как собственные числа матрицы D положительны. Система ограничений состоит из линейных неравенств. Мы имеем задачу выпуклого квадратичного программирования. Составим функцию Лагранжа задачи

$$L_B - x_1^2 + 2x_2^2 - 2x_1 - 4_2 + \lambda_1(x_1 + 2x_2 - 8) + \lambda_2(2x_1 - x_2 - 12).$$

Запишем систему уравнений (6.5), (6.6), и условия (6.8), (6.7)

$$\begin{cases} 2x_1 + \lambda_1 + 2\lambda_2 - u_1 = 2 \\ 4x_2 + 2\lambda_1 - \lambda_2 - u_2 = 4 \\ x_1 + 2x_2 + \nu_1 = 8 \\ 2x_1 - x_2 + \nu_2 = 12 \end{cases}$$

$$x_1, x_2, \lambda_1, \lambda_2, u_1, u_2, \nu_1, \nu_2 \geqslant 0$$

$$u_1x_1 = u_2x_2 = \nu_1\lambda_1 = \nu_2\lambda_2 = 0 \qquad (6.9)$$

Для нахождения допустимого базисного решения построенной системы линейных уравнений применим метод искусственного базиса, т.е. решим следующую задачу линейного программирования

$$z = -y_1 - y_2 \to max$$

$$\begin{cases} 2x_1 + \lambda_1 + 2\lambda_2 - u_1 + y_1 = 2 \\ 4x_2 + 2\lambda_1 - \lambda_2 - u_2 + y_2 = 4 \\ x_1 + 2x_2 + \nu_1 = 8 \\ 2x_1 - x_2 + \nu_2 = 12 \end{cases}$$

$$x_1, x_2, \lambda_1, \lambda_2, u_1, u_2, \nu_1, \nu_2, y_1, y_2 \geqslant 0$$

Здесь y_1, y_2 — искусственные переменные. Выразив искусственные переменные через свободные и подставив полученные выражения в целевую функцию, получим следующую симплекс-таблицу

Таблица 1

Баз. пер.	Св. чл.	x_1	x_2	λ_1	λ_2	u_1	u_2	ν_1	ν_2	y_1	y_2
x_1	1	1	0	1/2	1	-1/2	0	0	0	1/2	0
x_2	1	0	1	1/2	-1/4	0	-1/4	0	0	0	1/4
ν_1	5	0	0	-5/2	-3/2	1/2	1/2	1	0	-1/2	-1/2
ν_2	11	0	0	-1/2	-9/4	1	-1/4	0	1	-1	1/4
z	0	0	0	0	0	0	0	0	0	1	1

Таблица 2

Баз. пер.	Св. чл.	x_1	$\downarrow x_2$	λ_1	λ_2	u_1	u_2	ν_1	ν_2	y_1	y_2
y_1	2	2	0	1	2	-1	0	0	0	1	0
$\leftarrow y_2$	4	0	4	2	-1	0	-1	0	0	0	1
ν_1	8	1	2	0	0	0	0	1	0	0	0
ν_2	12	2	-1	0	0	0	0	0	1	0	0
z	-6	-2	-4	-3	-1	1	1	0	0	0	0

Таблица 3

Баз. пер.	Св. чл.	$\downarrow x_1$	x_2	λ_1	λ_2	u_1	u_2	ν_1	ν_2	y_1	y_2
$\leftarrow y_1$	2	2	0	1	2	-1	0	0	0	1	0
x_2	1	0	1	1/2	-1/4	0	-1/4	0	0	0	1/4
ν_1	6	1	0	-2	-1/2	0	1/2	1	0	0	-1/2
ν_2	13	2	0	1/2	-1/4	0	-1/4	0	1	0	1/4
z	-2	-2	0	-1	-2	1	0	0	0	0	1

Последняя симплекс таблица доставляет допустимое базисное решение системы, удовлетворяющее условиям (6.9). Поэтому точка $(\vec{x}^0, \vec{\lambda}^0) = (1; 1; 0; 0)$ является седловой точкой функции Лагранжа исходной задачи, а $\vec{x}^0 = (1; 1)$ — оптимальным планом исходной задачи. При этом $f_{min} = -3$.

6.5. Задачи дробно-линейного программирования

Задача дробно-линейного программирования (ДЛП) может быть сформулирована в векторной форме в следующем виде

$$z = \frac{(\vec{c}, \vec{x} + \alpha)}{(\vec{d}, \vec{x} + \beta)} \to \max\ (\min),$$

где α и β - скалярные константы, постоянные векторы $\vec{c}$ и $\vec{d} \in R^n, \vec{x} \in R^n$ - вектор искомых переменных. Предполагается, что $\vec{x} \in M \subset R^n$. При этом

$$M = \left\{ \vec{x} \in R^n : A\vec{x} = \vec{b}, \vec{x} \geqslant 0, \vec{b} \in R^m \right\}$$

Таким образом, в качестве целевой функции используется отношение двух линейных функций, условия-ограничения задачи являются линейными равенствами и неравенствами.

Обычно предполагают также, что знаменатель целевой функции не обращается в нуль на области допустимых значений M. Не ограничивая общности, можно считать, что знаменатель положителен на множестве M.

Задачи дробно-линейного программирования решают в тех приложениях, когда оптимизируются относительные показатели, например, рентабельность, производительность и т. д.. Особенно часто такие задачи встречаются в области финансовой деятельности: планировании доходов корпораций, управлении статьями банковского баланса и т. п.

Задачи ДЛП имеют одну важную общую черту с задачами линейного программирования. Поверхности уровня целевой функции этих задач определяются линейными уравнениями, т. е. являются гиперплоскостями. Действительно, равенство z=C при произвольной константе C можно записать в виде $(\vec{c}, \vec{x}) + \alpha = C((\vec{d}, \vec{x}) + \beta)$, а это уравнение задает некоторую гиперплоскость. При неопределенном C мы получаем пучок гиперплоскостей, пересекающихся по линейному многообразию размерности n-2, часто называемому множеством вращения. Множество вращения является множеством точек, удовлетворяющих системе уравнений

$$\begin{cases} (\vec{c}, \vec{x}) = -\alpha, \\ (\vec{d}, \vec{x}) = -\beta. \end{cases}$$

Мы будем считать, что эта система совместна, т. к. в противном случае задачу ДЛП легко свести к обычной задаче линейного программирования. В случае n=2 гиперповерхности уровня будут прямыми линиями, а множество вращения состоит из одной точки. Вращая линию уровня целевой функции вокруг этой точки, мы будем увеличивать или уменьшать значение целевой функции. На этом основан графический метод решения задачи ДЛП. По крайней мере, одна из угловых точек допустимого многоугольника будет оптимальной.

В общем случае обычно применяют сведение задачи ДЛП к задаче линейного программирования с помощью приема преобразования переменных. Еще раз отметим, что знаменатель целевой функции можно считать положительным на множестве M, т.к. в противном случае можно умножить числитель и знаменатель на (-1). Введем следующие переменные

$y_0 = ((\vec{d}, \vec{x}) + \beta)^{-1}$ и $y_i{=}y_0 x_i$ при $i{=}1,\ 2,\ ...,\ n$. Задача ДЛП может быть записана в виде следующей задачи линейного программирования

$$z = (\vec{c}, \vec{y}) + \alpha y_0 \to max(min),$$

$$\begin{cases} A\vec{y} - y_0\vec{b} = 0, \\ (\vec{d}, \vec{y}) + \beta y_0 = 1, \end{cases}$$

$$0 \le \vec{y}\ \in\ R^n, \quad 0 \le y_0\ \in\ R^1.$$

Количество переменных в этой задаче равно $n{+}1$. После решения этой задачи линейного программирования решение исходной задачи легко найти по формулам $x_i = y_i/y_0$ при $i{=}1,\ 2,\ ...,\ n$.

Пример 6.5. Найти решение следующей задачи

$$z = \frac{2x_1 + x_2}{x_1 + x_2} \to min,$$

$$\begin{cases} 2x_1 + 8x_2 \le 26, \\ x_1 + x_2 \le 4, \\ 12x_1 + 3x_2 \le 39, \end{cases}$$

$$x_1,\ x_2 \ge 0.$$

Для нахождения решения вначале уравняем неравенства системы ограничений с помощью введения неотрицательных переменных.

$$z = \frac{2x_1 + x_2}{x_1 + x_2} \to min,$$

$$\begin{cases} 2x_1 + 8x_2 + x_3 = 26, \\ x_1 + x_2 + x_4 = 4, \\ 12x_1 + 3x_2 + x_5 = 39, \end{cases}$$

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

Введем новые переменные $y_0 = (x_1 + x_2)^{-1}, y_i = y_0 x_i\ (i = \overline{1, 2})$. Получим задачу линейного программирования

$$z = 2y_1 + y_2 \to min,$$

$$\begin{cases} 2y_1 + 8y_2 + y_3 - 26y_0 = 0, \\ y_1 + y_2 + y_4 - 4y_0 = 0, \\ 12y_1 + 3y_2 + y_5 - 39y_0 = 0, \\ y_1 + y_2 = 1, \end{cases}$$

$$y_i \geq 0(i = \overline{0,5}).$$

Воспользуемся модификацией симплекс-метода, изложенном в 1.6. Для этого перейдем к задаче на максимум, введем искусственную переменную u в последнее равенство системы ограничений, составим М-задачу при М=200 и исключим искусственную переменную из целевой функции. Получим задачу

$$z_1 = -200 + 198y_1 + 199y_2 \to max,$$

$$\begin{cases} 2y_1 + 8y_2 + y_3 - 26y_0 = 0, \\ y_1 + y_2 + y_4 - 4y_0 = 0, \\ 12y_1 + 3y_2 + y_5 - 39y_0 = 0, \\ y_1 + y_2 + u = 1, \end{cases}$$

$$y_i \geq 0(i = \overline{0,5}).$$

Составим симплекс-таблицу

Таблица 1

Баз. пер.	Св. чл.	y_1	$\downarrow y_2$	y_3	y_4	y_5	y_0	u
y_3	0	2	8	1	0	0	-26	0
$\leftarrow y_4$	0	1	1	0	1	0	-4	0
y_5	0	12	3	0	0	1	-39	0
u	1	1	1	0	0	0	0	1
z_1	-200	-198	-199	0	0	0	0	0

Таблица 2

Баз. пер.	Св. чл.	y_1	y_2	y_3	y_4	y_5	$\downarrow y_0$	u
$\leftarrow y_3$	0	-6	0	1	-8	0	6	0
y_2	0	1	1	0	1	0	-4	0
y_5	0	9	0	0	-3	1	-27	0
u	1	0	0	0	-1	0	4	1
z_1	-200	1	0	0	199	0	-796	0

Таблица 3

Баз. пер.	Св. чл.	y_1	y_2	y_3	$\downarrow y_4$	y_5	y_0	u
y_0	0	-1	0	1/6	-4/3	0	1	0
y_2	0	-3	1	2/3	-13/3	0	0	0
y_5	0	-18	0	9/2	-39	1	0	0
$\leftarrow$ u	1	4	0	-2/3	13/3	0	0	1
z_1	-200	-795	0	398/3	-2587/3	0	0	0

Таблица 4

Баз. пер.	Св. чл.	y_1	y_2	y_3	y_4	y_5	y_0	u
y_0	4/13				0			
y_2	1				0			
y_5	9				0			
y_4	3/13	12/13	0	-2/13	1	0	0	3/13
z_1	-1	1	0	0	0	0	0	3/13

Таким образом, ответ в исходной задаче: $z_{min} = -z_{1max} = 1$, точка минимума $x_1 = 0, x_2 = 13/4$

6.6. Численные методы решения задач нелинейного программирования

1. Методы безусловной оптимизации. Метод градиетного спуска и метод Ньютона-Рафсона. Рассматриваемые ниже методы являются способами нахождения точек локального экстремума для задач без ограничений и поэтому называются методами безусловной оптимизации. В приложениях чаще всего встречаются задачи с ограничениями, которые можно свести к задачам на безусловный экстремум с помощью метода *штрафных функций.*

Рассмотрим задачу без ограничений

$$z = f(x_1, x_2, ..., x_n) \to \min.$$

В любой точке $\vec{x} \in R^n$ можно построить вектор градиента

$$\left\{ \frac{\partial f}{\partial x_1}; \frac{\partial f}{\partial x_2}; ...; \frac{\partial f}{\partial x_n} \right\}$$

Этот вектор указывает направление наискорейшего возрастания функции $f(\vec{x})$. Противоположный вектор $(-\bigtriangledown f)$ называется вектором антиградиента и указывает направление наискорейшего убывания функции $f(\vec{x})$в данной точке.

Идея **метода градиентного спуска** заключается в том, что мы из произвольно взятой точки в R^n сдвигаемся по направлению антиградиента в некоторую близкую точку. При этом сдвиг выбирается таким, чтобы в новой точке значение функции было меньше, чем в старой. Затем в этой новой точке определяется антиградиент. В его направлении делается новый шаг в некоторую новую точку с уменьшением значения целевой функции $f(\vec{x})$. В результате ряда шагов мы можем сколь угодно близко подойти к точке локального минимума функции $f(\vec{x})$. Если функция $f(\vec{x})$ имеет лишь один экстремум, который является и глобальным минимумом, то мы получаем приближенное решение задачи. Если же у функции$f(\vec{x})$ несколько локальных минимумов, то мы "скатимся"в результате ряда шагов в окрестность одного из локальных минимумов. Причем в зависимости от начальной точки можно попасть в окрестность того или иного локального минимума.

Расчетные формулы метода градиентного спуска имеют следующий вид:

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - h_k \bigtriangledown f_k,$$

где $\vec{x}^{(k+1)}$ - вектор координат точки, получаемой на k+1 шаге вычислительного процесса; $\vec{x}^{(k)}$ - точка , полученная на к-том шаге; $\bigtriangledown f_k$ - градиент функции в точке $\vec{x}^{(k)}$; h_k - неотрицательное число, определяющее шаг в направлении антиградиента. В покоординатной записи расчетные формулы имеют вид

$$x_i^{(k+1)} = x_i^{(k)} - h_k \left(\frac{\partial f}{\partial x_i}\right)_k, i = 1, 2, K, n.$$

При организации вычислительного процесса требуется на каждом шаге определять величину h_k и сформулировать критерий окончания счета. Поскольку в самой точке минимума все частные производные равны нулю, то в достаточно малой ее окрестности они должны быть достаточно малы. Поэтому счет заканчивается, когда длина градиента $|\bigtriangledown f|$ в очередной точке становится меньше некоторого, наперед заданного числа.

Выбор величины h_k можно производить по-разному. Если функция $f(\vec{x})$ не очень сложна, то на каждом шаге h_k можно выбрать оптимальным образом. Для этого построим функцию

$$\varphi_k(h) = f(\vec{x}^{(k)} - h(\bigtriangledown f)_k).$$

Найдем минимум этой функции по переменной h. Для этого обычно решают уравнение

$$\varphi_k{}'(h) = 0.$$

Если это уравнение удается решить, то мы и получаем искомое число h_k. Вычислительный метод, включающий такое определение h_k, называют *методом наискорейшего градиентного спуска.* В некоторых случаях выбор оптимального шага затруднен. Иногда, решая задачу на ЭВМ, программируют эмпирический подбор шага, т.е. подбирают h_k так, чтобы выполнялось неравенство $f(\vec{x}^{(k+1)}) \leq f(\vec{x}^{(k)})$.

В ***методе Ньютона-Рафсона*** приближенно решается система уравнений

$$\frac{\partial f}{\partial x_i} = 0, i = 1, 2, K, n.$$

Так же, как и в методе градиентного спуска, этот метод можно применять в случае одноэкстремальных задач. Расчетные формулы метода Ньютона-Рафсона имеют вид:

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - H_k^{-1} \cdot (\bigtriangledown f)_k,$$

где $\vec{x}^{(k)}, (\bigtriangledown f)_k, \vec{x}^{(k+1)}$ рассматриваются как матрицы-столбцы размера $n \times 1$, а H_k^{-1} - матрица, обратная по отношению к матрице Гесса для функции $f(\vec{x})$ в точке $\vec{x}^{(k)}$,

$$H_k = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \wedge & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \wedge & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \wedge & \wedge & \wedge & \wedge \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \wedge & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

Вторые частные производные берутся в точке $\vec{x}^{(k)}$. Окончание счета, как и в методе градиентного спуска, производится по критерию малости $|\bigtriangledown f|$. Метод Ньютона-Рафсона сходится быстрее к точке минимума, чем метод градиентного спуска в случае задач выпуклого программирования ($f(\vec{x})$ - выпуклая функция). В частности, для квадратичной функции $f(\vec{x})$ этот метод приводит к точке минимума за один шаг. Применение метода Ньютона-Рафсона требует вычисления обратной матрицы, что при большом числе переменных является очень громоздкой задачей.

Оба описанных метода часто применяют для решения произвольных задач нелинейного программирования. Здесь, однако, отсутствуют теоретические обоснования сходимости методов и скорости их сходимости.

2. Метод штрафных функций. Рассмотрим задачу выпуклого программирования:

$$z = f(x_1, x_2, k, x_n) \leftarrow min;$$

$$\begin{cases}\varphi_1(x_1,x_2,k,x_n)\le 0,\\ \varphi_2(x_1,x_2,k,x_n)\le 0,\\ \wedge\wedge\wedge\wedge\wedge\wedge\wedge\wedge\wedge\wedge\\ \varphi_m(x_1,x_2,k,x_n)\le 0.\end{cases}$$

Здесь функции $f, \varphi_1, \varphi_2, K, \varphi_m$ - выпуклые, а множество $\Omega = R^n$.

Метод штрафных функций заключается в составлении по данной задаче некоторой задачи на безусловный экстремум, точка минимума которой совпадает с точкой минимума исходной задачи. Построенную задачу можно решать приближенно с помощью описанных выше методов.

Вспомогательная задача на безусловный экстремум строится с помощью введения в целевую функцию дополнительных штрафных слагаемых, которые очень велики в тех точках R^n, которые не принадлежат области допустимых значений . При минимизации функции точка минимума вынуждена находиться в области допустимых значений исходной задачи.

Рассмотрим функцию

$$\mu(y)=\begin{cases}y, y>0;\\ 0, y\le 0.\end{cases}$$

Построим по нашей задаче функцию

$$\Phi_R(x_1,x_2,K,x_n)=f(x_1,x_2,K,x_n)+R\sum_{j=1}^{m}\mu(\varphi_j(x_1,x_2,K,x_n)),$$

где $R>0$ - достаточно большая константа.

Теорема 6.8. *Если $\vec{x}^{(0)}$ является точкой минимума задачи выпуклого программирования, то существует такое число $R_0>0$, что при $R>R_0$ точка $\vec{x}^{(0)}$ будет точкой безусловного минимума функции $\Phi_R(\vec{x})$. Если для всех $R>R_0$ $\vec{x}^{(0)}$ является точкой безусловного экстремума функции $\Phi_R(\vec{x})$, то эта точка является и точкой минимума задачи выпуклого программирования.*

Отметим, что функция $\Phi_R(\vec{x})$ сама является выпуклой, поэтому задача $\Phi_R(\vec{x})\to\min$ одноэкстремальна. Для решения можно применить метод градиентного спуска. Компоненты градиента здесь являются разрывными функциями, поэтому используют обобщенный градиент, имеющий вид

$$\nabla\Phi_R=\nabla f+R\sum_{j=0}^{m}\mu(sign\varphi_i)\nabla\varphi_j,$$

$$\text{где sign}\varphi_j=\begin{cases}\text{1, при }\varphi_j>0,\\ \text{0, при }\varphi_j=0,\\ \text{-1, при }\varphi_j<0.\end{cases}$$

Организация наискорейшего спуска встречает трудности из-за сложности функции $\Phi_R(\vec{x})$ Поэтому применяют градиентный спуск с эмпирическим выбором шага. Подбор величины R часто тоже производится эмпирически. Величина R достаточно велика, если ее увеличение не приводит к изменению точки минимума $\Phi_R(\vec{x})$. Существуют точные оценки для R, при выполнении которых точка минимума функции $\Phi_R(\vec{x})$ совпадает с решением соответствующей задачи выпуклого программирования (см. [27], стр. 335).

Иногда, чтобы избежать трудностей, связанных с разрывностью градиента, вместо функции $\Phi_R(\vec{x})$ используют функцию

$$S_R(x_1, x_2, K, x_n) = f(x_1, x_2, K, x_n) + R\sum_{j=1}^{m} \mu^2(\varphi_j(x_1, x_2, K, x_n)),$$

с квадратичной функцией штрафа. При гладкости функций, входящих в формулировку исходной задачи, функция $S_R(\vec{x})$ гладкая, и для нахождения ее точки минимума можно использовать любой из градиентных методов. Недостатком функции $S_R(\vec{x})$ является то, что ее точка минимума удовлетворяет системе ограничений лишь приближенно, причем тем точнее, чем больше R. При этом скорость сходимости градиентных методов при больших R медленная. На практике метод штрафных функций применяется и при решении одноэкстремальных задач, не являющихся задачами выпуклого программирования.

6.7. Многокритериальные задачи

В рассмотренных до сих пор задачах имелся лишь один целевой параметр, который мы выражали через контролируемые параметры и получали одну целевую функцию. В экономике такие задачи возникают при рассмотрении небольших по масштабу и скромных по значению мероприятий. Когда речь идет о крупномасштабных операциях, затрагивающих разнообразные интересы, их эффективность не может быть выражена с помощью одного единственного показателя.

При организации работы промышленного предприятия естественно стремиться к максимуму валового объема продукции V, к максимуму чистого дохода D , к минимуму себестоимости продукции S, а также к максимуму производительности труда P. При более глубоком анализе может появиться еще ряд дополнительных критериев. Такая множественность показателей эффективности, из которых одни желательно обратить в максимум, а другие — в минимум, характерна для любой достаточно сложной задачи исследования операций. Такие задачи называются *многокритериальными*.

Если все целевые показатели удается выразить через контролируемые параметры $x_1, x_2, \ldots, x_n$, мы получаем задачу на отыскание точки $\vec{x}^{(0)} \in$ M, в которой достигался бы глобальный экстремум (можно считать минимум) сразу нескольких функций $f_1(\vec{x}), f_2(\vec{x}), K, f_k(\vec{x})$. В общем случае такая задача решения не имеет. При принятии решения приходится выбирать компромиссный вариант, приемлемый сразу по всем критериям. Для такого выбора, кроме самих функций $f_i(\vec{x})$ и ограничений, определяющих множество M, требуются дополнительные (часто субъективные) соображения. Объективно в множестве M можно выделить подмножество $\wp$ так называемых *паретовских решений*, из которых и производится выбор компромиссного варианта.

Определение 6.6. *Паретовским решением многокритериальной задачи на минимум называется такая точка $\vec{x}^{(0)} \in M$, для которой нельзя найти «лучшую» по любому данному критерию точку, «не ухудшающую» значения других критериев. Таким образом, $\vec{x}^{(0)}$ характеризуется следующим свойством. Если для некоторого $\vec{x}^{(0)} \in M$ выполнено неравенство $f_i(\vec{x}^{(0)}) > f_i(\vec{x})$, то найдется j $(1 \le j \le k)$, для которого выполняется неравенство $f_j(\vec{x}^{(0)}) < f_j(\vec{x})$.*

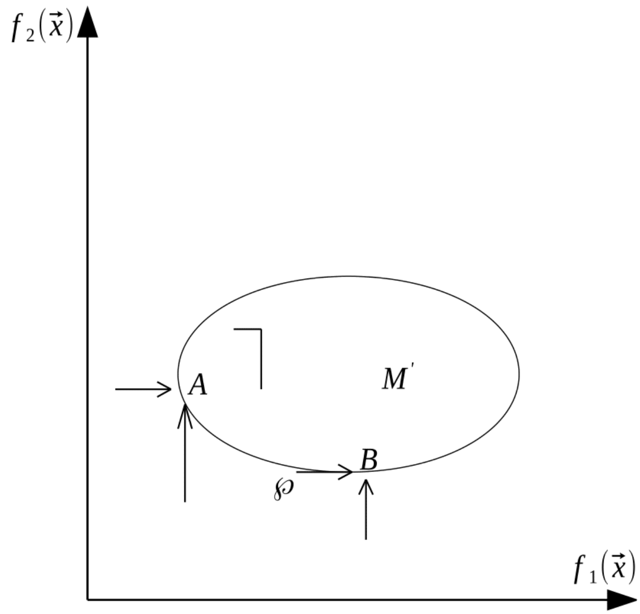

Рис. 6.1. Образ множества Парето в плоскости критериев f_1, f_2

Можно сказать, что паретовское решение $\vec{x}^{(0)}$ - это такое решение системы ограничений, для которого нет решений этой системы, доминирующих над $\vec{x}^{(0)}$ по всем критериям. Чаще всего множество $\wp$ гораздо уже множества М. Это видно на следующем примере. Пусть рассматривается задача с двумя критериями $f_1(\vec{x})$ и $f_2(\vec{x})$. Для каждого значения $\vec{x} \in M$ мы имеем точку на плоскости с координатами $(f_1(\vec{x}), f_2(\vec{x}))$. Множество всех таких точек обозначим M' (рис.6.1). Паретовским решениям отвечают точки дуги границы множества M'.

Непосредственное вычисление множества Парето $\wp$ довольно сложно. Часто применяют *эвристические приемы решения многокритериальных задач*. Рассмотрим несколько таких приемов.

1. Способ агрегированного критерия. Рассмотрим задачу

$$\sum_{i=1}^{k} a_i f_i(\vec{x}) \to min, \vec{x} \in M,$$

где a_i - некоторые неотрицательные числа, называемые весовыми коэффициентами. Они характеризуют относительную важность критериев. Чем больше число a_i, тем большую важность придают критерию $f_i(\vec{i})$. Если все функции $f_i(\vec{i})$ выпуклые и множество М — тоже, то у этой задачи есть точка минимума, и она обязательно принадлежит множеству $\wp$. При этом в качестве точки минимума можно получить любую точку множества $\wp$ надлежащим выбором весовых коэффициентов. Выбор этих коэффициентов обычно производится эвристически, например, группой специальных экспертов.

2. Способ главного показателя. Здесь эвристически выбирается некоторый главный показатель, например, $f_1(\vec{x})$. Его стремятся минимизировать при условии, что $\vec{x} \in M$; кроме того, остальные показатели остаются не больше некоторых приемлемых значений
b_2, b_3, K, b_k $(f_i(\vec{x}) \le b_i, i = 2, 3, K, k)$.При таком подходе все показатели, кроме главного, переводятся в разряд заданных ограничений.

3. Способ последовательных уступок. Этот способ построения компромиссного решения предполагает предварительное расположение целевых функций в порядке убывающей важности: $f_1(\vec{x}), f_2(\vec{x}), K, f_k(\vec{x})$. Сначала находится решение, обращающее в минимум важнейший показатель $f_1(\vec{x})$. Пусть $f_{1min} = b_1$. Затем назначается некоторая "уступка"$\triangle b_1$, которую мы согласны сделать для того, чтобы минимизировать второй показатель $f_2(\vec{x})$. К первоначальной системе ограничений добавляется условие $f_1(\vec{x}) \le b_1 + \triangle b_1$ и с новой системой ограничений находится $\min f_2(\vec{x}) = b_2$. Затем назначается новая уступка $\triangle b_2$ и к системе ограничений первого шага добавляется условие $f_2(\vec{x}) \le b_2 + \triangle b_2$. С новой системой ограничений решается задача на минимум функции $f_3(\vec{x})$ и т.д. На последнем шаге к системе ограничений добавляется условие $f_{k-1}(\vec{x}) \le b_{k-1} + \triangle b_{k-1}$ и находится $\min f_k(\vec{x})$. Последняя точка минимума $\vec{x}^{(0)}$ считается решением задачи. Такой способ построения компромиссного решения хорош тем, что здесь видно, ценой какой "уступки"в одном показателе приобретается выигрыш в другом, и какова величина этого выигрыша. Отметим, что последний способ требует эвристического упорядочивания показателей и такого же назначения последовательных уступок.

При любом способе постановки многокритериальной задачи для получения ее конкретного решения требуется вмешательство человеческого сознания для проведения неформализуемых оценок.

Контрольные вопросы и задачи для самостоятельного решения

1. Как формулируется общая задача нелинейного программирования? Как можно менять формулировку задачи, оставляя задачу эквивалентной исходной?
2. Дайте определение локального экстремума задачи нелинейного программирования. Что такое глобальный экстремум? Какие задачи называются одноэкстремальными?
3. Когда задача нелинейного программирования называется задачей на условный экстремум? Что такое функция Лагранжа задачи на условный экстремум? В чем заключаются необходимые условия условного экстремума?
4. Сформулируйте достаточные условия условного экстремума. Что такое критерий Сильвестра и как он используется при выяснении характера стационарной точки функции Лагранжа?
5. Дайте определение выпуклого множества в Rn и выпуклой функции. Что такое задача выпуклого программирования? Является ли такая задача одноэкстремальной?
6. Что такое функция Лагранжа задачи выпуклого программирования и ее седловая точка?
7. Сформулируйте теорему Куна-Таккера.
8. Как формулируются необходимые и достаточные условия Седловой точки?
9. Как формулируются необходимые и достаточные условия Седловой точки?
10. Как решить задачу выпуклого квадратичного программирования сведением к вспомогательной задаче линейного программирования?
11. Как формулируется задача дробно-линейного программирования? Как истолковать эту задачу геометрически в случае двух переменных?
12. Как сводится задача дробно-линейного программирования к задаче линейного программирования с помощью введения новых переменных?
13. Опишите метод наискорейшего градиентного спуска для задач на безусловный экстремум. В чем состоит расчетная схема метода Ньютона-Рафсона?

14. В чем состоит основная идея метода штрафных функций? Линейная и квадратичная функции штрафа. В чем заключаются преимущества и недостатки каждой из них?
15. Что такое паретовское решение многокритериальной задачи? Основные подходы, используемые для решения многокритериальных задач. Что такое способ агрегированного критерия?
16. Опишите суть способов главного показателя и последовательных уступок.

Используя геометрическое истолкование задачи нелинейного программирования решить задачи 6.1 — 6.4.

6.1.

$$z = x_1x_2 \to max,$$

$$\begin{cases} 6x_1 + 4x_2 \geq 12, \\ 2x_1 + 3x_2 \leq 24, \\ -x_1 + 4x_2 \leq 12, \\ x1, \quad x2 \geq 0. \end{cases}$$

Ответ: $z_{max} = 24$, точка max (6; 4).

6.2.

$$z = 9(x_1-5)^2 + 4(x_2-6)^2 \to min,$$

$$\begin{cases} 3x_1 + 2x_2 \geq 12, \\ x_1 - x_2 \leq 6, \\ -x_2 \leq 4, \\ x1, \quad x2 \geq 0. \end{cases}$$

Ответ: $z_{min} = 16$, точка min (5; 4).

6.3.

$$z = 4x_1 + 3x_2 \to max,$$

$$\begin{cases} x_1^2 - 2x_1 + x_2^2 - 2x_2 - 34 \leq 0, \\ x_1 \geq 1, \\ -x_2 \geq 1. \end{cases}$$

Ответ: $z_{max} = 37$, точка max (5,8;4,6).

6.4.

$$z = x_1x_2 \to max,$$

$$\begin{cases} x_1^2 + 2x_1 + x_2^2 - 2x_2 - 14 \geq 0, \\ 2x_1 + x_2 \leq 10, \\ -x_1,\ x_2 \geq 0. \end{cases}$$

Ответ: $z_{max} = 12{,}5$, точка max (2,5; 5).

В задачах 6.5 — 6.8 найти условные экстремумы указанных функций и определить их характер.

6.5.

$$z = x_1^2 + x_2^2 + x_3$$

при условиях

$$\begin{cases} x_1 + x_2 + x_3 = 4, \\ 2x_1 - 3x_2 = 12. \end{cases}$$

Ответ: $\begin{cases} z_{min} = 16\frac{53}{64} \textit{ в точке} \\ (27/8; -7/4; 19/8). \end{cases}$

6.6.

$$z = x_1x_2x_3$$

при условиях

$$\begin{cases} 2x_1x_2 + x_2x_3 = 4, \\ 2x_1 - 3x_2 = 12. \end{cases}$$

Ответ: $\begin{cases} z_{min} = -56/27 \textit{ в точке} \\ (-1,3; -26/3; -28/39), \\ z_{max} = 72 \textit{ в точке} \\ (3; -2; -12). \end{cases}$

6.7.

$$z = x_1x_2 + x_2x_3$$

при условиях

$$\begin{cases} x_1 + x_2 = 4, \\ x_2 + x_3 = 4. \end{cases}$$

Ответ: $\begin{cases} z_{max} = 8 \text{ *в точке*} \\ (2; 2; 2). \end{cases}$

6.8.

$$z = 3x_1^2 + 2x_1 + 2x_2^2 + 4x_2x_3$$

при условиях

$$\begin{cases} x_1^2 + 2x_2^2 = 19, \\ x_1 + 2x_2x_3 = 11. \end{cases}$$

Ответ: $\begin{cases} z_{min} = -43 \text{ *в точах*} \\ (-1; 3; 2)(-1; -3; -2). \end{cases}$

Решить задачи выпуклого квадратичного программирования 6.9 — 6.12.

6.9.

$$f = 2x_1^2 + 2x_2^2 - x_1x_2 - x1 - 4x_2 \to min;$$

$$\begin{cases} x_1 + 2x_2 \le 12, \\ 3x_1 + x_2 \le 15, \end{cases}$$

$$x_1, \quad x_2 \ge 0.$$

Ответ: $f_{min} = -\frac{35}{15} при x_1 = \frac{8}{15}, x_2 = \frac{17}{15}.$

6.10.

$$f = x_1^2 + x_2^2 - x1 - 8x_2 \to min;$$

$$\begin{cases} x_1 + 2x_2 \le 7, \\ x_2 \le 5, \end{cases}$$

$$x_1, \quad x_2 \ge 0.$$

Ответ: $f_{min} = -\frac{65}{4} при x_1 = \frac{1}{4}, x_2 = 4.$

6.11.

$$f = x_1^2 + x_2^2 - 2x1 - 8x_2 \to min;$$

$$\begin{cases} x_1 + 2x_2 \le 12, \\ -x_1 + x_2 \ge -8, \end{cases}$$

$$x_1, \quad x_2 \ge 0.$$

Ответ: $f_{min} = -16 \text{ при } x_1 = 0, x_2 = 4.$

6.12.

$$f = x_1^2 + x_2^2 + 2x_3^3 - 2x_2 - 3x3 \to min;$$

$$\begin{cases} x_1 + x_2 + x_3 \le 18, \\ x_2 \le 12, \\ x_1 + 2x_3 \le 14, \end{cases}$$

$$x_1, \quad x_2, \quad x_3 \ge 0.$$

Ответ:$f_{min} = -\dfrac{17}{8} при x_1 = 0, x_2 = 1, x_3 = \dfrac{3}{4}.$

Решить задачи дробно-линейного программирования 6.13 — 6.17.

6.13.

$$z = \frac{-5x_1 + 2x_2}{3x_1 + 4x_2} \to max,$$

$$\begin{cases} x_1 + 4x_2 + x_3 = 16, \\ 2x_1 + 3x_2 - 4x_4 = 12, \\ 3x_1 - 2x_2 + x_5 = 18, \end{cases}$$

$$x_i \ge 0(i = \overline{1,5}).$$

Ответ:$z_{max} = 1/2$ в точке $(0; 4; 0; 0; 26)$.

6.14.

$$z = \frac{3x_1 - x_2}{x_1 + x_2} \to max,$$

$$\begin{cases} x_1 + x_2 - x_3 = 5, \\ -x_1 + 3x_2 + x_4 = 7, \\ 3x_1 - x_2 + x_5 = 11, \end{cases}$$

$$x_i \ge 0(i = \overline{1,5}).$$

Ответ:$z_{max} = 2,2$ в точке $(4; 1; 0; 8; 0)$.

6.15.

$$z = \frac{8x_1 + 3x_2 + 2x_3 + x_4}{x_1 + x_2 + x_3 + x_4} \to max,$$

$$\begin{cases} 2x_1 + x_2 + x_3 + 3x_4 \le 300, \\ x_1 + 2x_3 + x_4 \le 70, \\ x_1 + 2x_2 + x_3 \le 340, \end{cases}$$

$$x_i \ge 0(i = \overline{1,4}).$$

Ответ:$z_{max} = 8$ в точке $(70; 0; 0; 0)$.

6.16.

$$z = \frac{5x_1 - x_2 + 8x_3 + 10x_4 - 5x_5 + x_6}{x_1 + x_2 + x_3 + x_4 + x_5 + x_6} \to max,$$

$$\begin{cases} 2x_1 - x_2 + 3x_4 + x_5 - x_6 = 40, \\ -x_1 + 2x_2 + x_3 + 2x_4 + 2x_6 = 20, \\ 3x_1 - x + 2x_3 - x_4 + 3x_5 + x_6 = 30, \end{cases}$$

$$x_i \geq 0 (i = \overline{1, 6}).$$

Ответ:$z_{max} = 489/62$ в точке $(6, 8; 0; 9, 2; 8, 8; 0; 0)$.

6.17.

$$z = \frac{2x_1 - 3x_2 + 4x_3 + 5x_4 5x_5 + 8x_6}{x_1 + x_2 + x_3 + x_4 + x_5 + x_6} \to max,$$

$$\begin{cases} x_1 + 5x_2 - 3x_3 - 4x_4 + 2x_5 + x_6 = 120, \\ 2x_1 + 9x_2 - 5x_3 - 7x_4 + 4x_5 + 2x_6 = 320, \end{cases}$$

$$x_i \geq 0 (i = \overline{1, 6}).$$

Ответ:$z_{max} = 98/13$ в точке $(0; 0; 0; 80; 0; 440)$.

7. ЭЛЕМЕНТЫ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Под динамическим программированием понимается метод оптимизации для операций, которые можно разбить на ряд шагов (этапов). Выбор решений на отдельных этапах можно рассматривать как управление реализацией данной операции. Для точной постановки задачи требуется введение некоторых новых понятий.

7.1. Прикладной пример и основные понятия

На ферме имеется стадо скота. Ежегодно часть стада отправляется на мясозаготовки, а остальная часть остается на ферме для воспроизводства. Доход от продажи скота выражается функцией $\varphi(\mathrm{u})$, где u — количество проданного скота. Функция $\varphi(\mathrm{u})$ может иметь, например, следующий вид Рис. 7.1

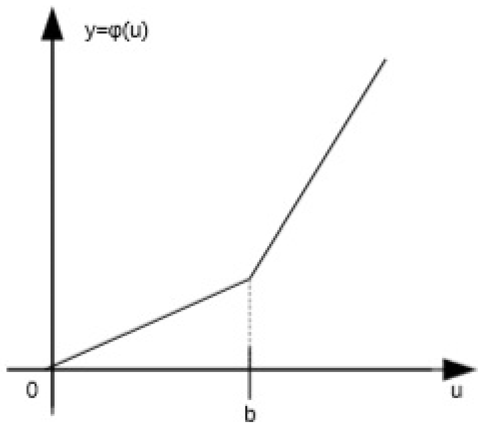

Рис. 7.1. Зависимость цены от количества проданного скота

Здесь поставки мяса сверх планового задания b оплачиваются по более высокой цене. Количество скота, оставленного на ферме для воспроизводства, в следующем году до начала мясозаготовок увеличивается в a раз, где $a \geq 1$. Требуется таким образом спланировать мясозаготовки на N лет, чтобы в итоге получить максимальный доход при условии, что минимальные ежегодные мясозаготовки составляют b .

Обозначим через $x(0)$ начальное количество скота на ферме, а через $x(t)$ — количество скота, оставленного на ферме к концу t-го года, $t = 1, 2, ..., N$.

Количество голов скота, проданного для мясозаготовок в t-м году, обозначим через $u(t), t = 1, 2, ..., N$. В $t - 1$ году, было оставлено для воспроизводства количество скота, равное $x(t-1)$. Следовательно, в t-м году перед мясозаготовками количество скота будет $ax(t-1)$. Из этого количества $u(t)$ будет продано для мясозаготовок, а остальная часть, т.е. $ax(t-1) - u(t)$, останется на ферме для воспроизводства. Доход фермы за N лет составит:

$$I = \varphi(u(1)) + \varphi(u(2)) + \varphi(u(N)) = \sum_{t=1}^{N} \varphi(u(t))$$

Учитывая обязательные мясозаготовки, мы имеем для управляющего параметра $u(t)$ ограничения :

$$u(t) \geq b, t = 1, 2, ..., N.$$

Кроме того, естественно считать выполненным условие

$$x(N) \geq d,$$

где d — плановое задание по разведению скота к концу N-летнего периода.

Итак, мы можем сформулировать следующую математическую задачу: требуется выбрать числа $u(1), u(2), ..., u(N)$ таким образом, чтобы максимизировать функцию

$$I = \sum_{k=1}^{N} \varphi(u(t)) \to max, \tag{7.1}$$

при условиях:

$$x(t) = ax(t-1) - u(t), (t = 1, 2, ...N); \tag{7.2}$$

$$u(t) \geq b, t = 1, 2, ...N; \tag{7.3}$$

$$x(0) = x_0, x(N) \geq d, x(t) \geq 0. \tag{7.4}$$

Полученная задача называется задачей оптимального управления. целочисленного аргумента $u(t)$ называется управлением, поскольку описывает управление деятельностью фермы. Функция $x(t)$ называется переменной состояния или фазовой переменной, поскольку описывает состояние управляемого объекта (фермы).Соотношения (7.2) обычно называют уравнениями состояния, поскольку они определяют эволюцию состояний объекта при известном управлении. Условия (7.3) являются ограничениями на выбор управления. Условие (7.4) — условие на начальное и конечное состояние объекта.

Обобщая рассматриваемый пример, укажем общее математическое описание дискретных управляемых объектов. Будем считать, что переменная t может принимать лишь значения $t = 0, 1, 2, ..., N$, где N — фиксированное натуральное число. В общем случае предполагается, что можно воздействовать на управляемый объект, выбирая r управляющих параметров $u_1(t), ..., u_r(t)$, или, что тоже самое, при каждом t выбирается управляющая точка имеющая r координат:

$$\vec{u}(t) = \{u_1(t), u_2(t), ..., u_r(t)\}$$

Управлением мы условимся называть последовательность точек $\vec{u}(0), \vec{u}$ $(1), \vec{u}(2), ..., \vec{u}(N)$ в пространстве переменных $u_1, u_2, ..., u_r$. Будем также считать, что в каждый момент времени t состояние объекта характеризуется n фазовыми координатами $x_1, x_2, ..., x_n$, т.е. в каждый момент времени t фазовое состояние $\tilde{x}(t)$ имеет n координат:

$$\vec{x}(t) = x_1(t), x_2(t), ..., x_r(t)$$

При этом последовательность $\vec{x}(0), \vec{x}(1), ..., \vec{x}(N)$ состояний объекта в моменты $t = 0, 1, ..., N$ будем называть траекторией движения. В рассмотренном нами примере управление одномерно $(r = 1)$, и фазовое состояние описывается одним параметром $(n = 1)$.

Начальное состояние обычно считается заданным: $\vec{x}(0) = \vec{x}_0$. Дальнейшее поведение объекта однозначно определяется, если выбрано некоторое управление, с помощью следующих уравнений состояния:

$$x_1(t) = f_1(t, \vec{x}(t-1), \vec{u}(t)), \tag{7.5}$$

где $i = 1, 2, ..., n; t = 1, 2, ..., N$.

Соотношение (7.5) часто называют законом движения дискретного управляемого объекта. Траекторию объекта, удовлетворяющую уравнениям (7.5) называют соответствующей начальному состоянию $\vec{x}(0)$ и управлению $\vec{u}(0)$

Во многих задачах управление $\vec{u}(0)$ не является произвольным. В нашем примере оно удовлетворяет условию (7.3). В общем же случае для каждого состояния $\tilde{x}$ и момента времени t задается в пространстве управлений некоторое непустое множество $U_t(\tilde{x})$— область управления. При этом рассматривают лишь управления, которые удовлетворяют условию

$$\vec{u}(t) \in U_t(\vec{x}(t-1)), t = 1, 2, ..., N,) \tag{7.6}$$

где траектория объекта исходит из начальной точки $\vec{x}(0)$. Управление, удовлетворяющее этому условию, называют допустимым (относительно начального состояния $\vec{x}(0)$). Соотношения (7.5) и (7.6) определяют дискретный управляемый объект. Процесс управления таким объектом осуществляется следующим образом. Поскольку задано начальное состояние $\vec{x}(0)$, нам известна соответствующая область управления $U_1(\vec{x}(0))$. Мы можем выбрать произвольную управляющую точку $\vec{u}(1) \in U_1(\vec{x}(0)$. После этого по формулам (7.5) мы определим состояние $x(1)$ при $t = 1$. Далее, зная, мы можем рассмотреть соответствующую область управления $U_2(\vec{x}(1))$. Выбрав произвольную управляющую точку $\vec{u}(2) \in U_2(\vec{x}(1))$, мы можем согласно (7.5) найти следующее состояние $\vec{x}(2)$ и т.д. Очевидно, что управление $\vec{u}(0), \vec{u}(1), ..., \vec{u}(N)$, получающееся в результате такого последовательного выбора, является допустимым (относительно исходного начального состояния $\vec{x}(0)$). При этом траектория объекта является соответствующей данному управлению.

Теперь можно поставить общую задачу оптимального управления для управляемого объекта. В качестве критерия эффективности, т.е. функции, показывающей насколько выгодным был выбранный процесс управления, рассмотрим выражение

$$I = \sum_{t=1}^{N} f_0(t, \vec{x}(t-1), \vec{u}(t)). \tag{7.7}$$

Задача оптимального управления заключается в том, чтобы, зная начальное состояние, выбрать такое допустимое управление для объекта (7.5),(7.6), которое придает функционалу (7.7) максимальное значение.

В некоторых случаях речь может идти о минимальном значении функционала типа (7.7).

Сформулированная задача, которую часто называют основной, может быть также названа задачей с закрепленным левым концом и свободным правым. Здесь начальное состояние является заданным, а состояние в правом конце отрезка времени, т.е. $\vec{x}(N)$ ничем не связано (лишь бы значение функционала (7.7) было максимальным). Кроме основной задачи можно также рассматривать задачу с подвижными концами. В этом случае в фазовом пространстве задаются два множества M_0 и M_N. Требуется определить такое начальное состояние $\vec{x}(0) \in M_0$ и такое допустимое (относительно) управление, чтобы было выполнено соотношение $\vec{x}(N) \in M_N$ и при этом функционал (7.7) принимал максимальное значение. Ясно, что если M_0 состоит из одной точки, а M_N совпадает со всем фазовым пространством, то задача с подвижными концами превращается в уже рассмотренную основную задачу.

Отметим, что в нашем примере управления фермой множество M_0 состоит из одной точки, а множество M_N определяется неравенством $x \geq d$, т.е. мы имеем здесь задачу с подвижными концами (точнее, с закрепленным левым и подвижным правым концом).

Наиболее общей задачей оптимального управления является задача с ограничениями на фазовые координаты. В этой задаче для каждого $t = 1, 2, ..., N$ задается в фазовом пространстве некоторое непустое множество M_t и требуется найти такое начальное состояние $\vec{x}(0)$ и такое допустимое (относительно $\vec{x}(0)$) управление, чтобы были выполнены соотношения $\vec{t} \in M_t(t = 0, 1, ..., N)$, и при этом функционал (7.7) принимал наибольшее возможное значение. Если множества $M_1, M_2, ..., M_{N-1}$ совпадают со всем фазовым пространством, то задача превращается в задачу с подвижными концами.

7.2. Дальнейшие примеры и принцип оптимальности. Метод динамического программирования

Выше мы сформулировали общую задачу оптимального управления объектом с дискретным временем. Приведем еще несколько различных примеров, чтобы убедиться в том, что подобные математические задачи встречаются часто и могут иметь весьма разнообразное содержание.

1.Нелинейная транспортная задача. В некоторых случаях можно рассматривать управляемый объект, не имея в виду его реальное существование во времени. Часто приходится вводить время искусственно, расчленяя решение некоторой задачи на условные шаги. Математическая формулировка задачи в этом случае формально совпадает с задачей оптимального управления и для ее решения можно применять аналогичные методы. Характерным примером могут служить распределительные задачи. Рассмотрим задачу о перевозке груза (сырья) от производителей (складов) к местам переработки (заводам).

Будем считать, что число складов равно трем, а число заводов обозначим через N. Предположим, что запас равен спросу, т.е. если обозначить через a_1, a_2, a_3 количество груза на складах, а через $b_1, b_2, ..., b_N$ потребности заводов, то $a_1 + a_2 + a_3 = b_1 + b_2 + ... + b_N$. Стоимость перевозки u единиц груза с i-го склада на t-й завод зависит от u (т.е., от того, сколько сырья нужно привезти), а также от i и t (этими числами характеризуется дальность перевозки). Обозначим эту стоимость через $\varphi_{it}(u)$. Функция $\varphi_{it}(u)$ может, например, иметь вид, изображенный на Рис. 7.2.

Задача заключается в том, чтобы перевезти сырье со складов на заводы с минимальными транспортными затратами. Распределение груза будем производить поэтапно. На каждом этапе удовлетворим некоторый завод. Таким образом, мы намечаем количество сырья, которое надо завести на 1-й завод с первого и второго складов, затем количество груза, перевозимого с этих же складов на 2-й завод, на 3-й завод и т.д. Количество же сырья, поставленного третьим складом, определится тогда однозначно: на каждый завод надо довезти с третьего склада недостающее количество сырья. Обозначим через $u_1(t)$ количество сырья, поставленного на t-й завод первым складом, а через $u_2(t)$ — вторым складом. При этом с третьего склада надо будет завести на t-й завод сырье в количестве $b_t(t) - u_1(t) - u_2(t)$.

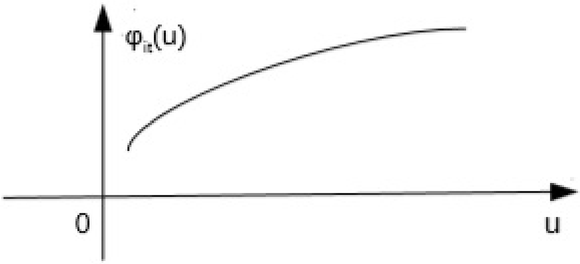

Рис. 7.2. Зависимость стоимости перевозки от количества перевезенного груза

Для составления плана нам достаточно выбрать числа $u_1(t), u_2(t)$ при $t = 1, 2, ..., N$. При этом по смыслу задачи

$$u_1(t) \geq 0, u_2 \geq 0, u_1(t) + u_2(t) \leq b_t t = 1, 2, ..., N. \tag{7.8}$$

Иными словами, "управляющая точка"$\vec{u}(t) = (u_1(t); u_2(t))$ должна находиться в треугольнике. Поскольку необходимо выполнить условия

$$u_1(t) \leq a_1, u_2(t) \leq a_2,$$

то множество U_t может быть пятиугольником Рис. 7.3.

Состояние складов (состояние объекта) при таком поэтапном распределении будем характеризовать величинами:

$\mathrm{x}_1(\mathrm{t})$ — количество сырья, вывозимого с первого склада на первые t заводов $\mathrm{x}_2(\mathrm{t})$ — количество сырья, вывозимого со второго склада на первые t заводов.

Ясно, что

$$\begin{cases} x_1(t) = x_1(t-1) + u_1(t), \\ x_2(t) = x_2(t-1) + u_2(t) \end{cases} \tag{7.9}$$

где $t = 1, 2, ..., N$.

Равенства (7.9) представляют собой уравнения состояния. Для удобства мы будем считать, что

$$x_1(0) = x_2(0) = 0 \tag{7.10}$$

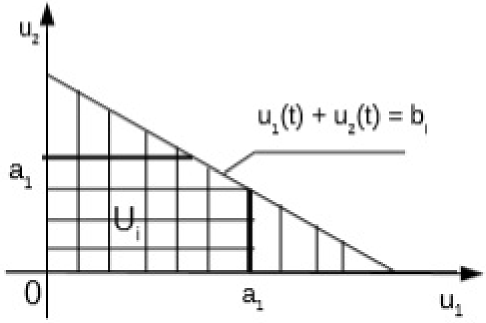

Рис. 7.3. Область допустимых управлений нелинейной транспортной задачи

Так как запас равен спросу, то после вывоза сырья на все N заводов склады должны остаться пустыми, т.е.

$$x_1(N) = a_1; x_2(N) = a_2 \tag{7.11}$$

Критерием эффективности плана здесь является стоимость всех перевозок:

$$I = \sum_{k=1}^{N} [\varphi_{1t}(u_1(t) + \varphi_{2t}(u_2(t)) + \varphi_{3t}(b_t - u_1(t) - u_2(t))].$$

Таким образом, транспортная задача приобретает следующую формулировку: для управляемого объекта (7.9) найти такое допустимое управление $u_1(t), u_2(t)$ (удовлетворяющее условию (7.8)), чтобы для соответствующей траектории $x_1(t), x_2(t)(t = 1, 2, ..., N)$ с начальным условием (7.10) удовлетворялось конечное условие (7.11) и при этом функционал I принимал наименьшее возможное значение. Очевидно, что это частный случай задачи с подвижными концами, в которой множества M_0 и M_1 состоят из одной точки каждое.

2. Задача о распределении ресурсов. Мы рассмотрим наиболее простой вариант этой задачи. В нашем распоряжении имеется какой-то запас средств (ресурсов) в количестве единиц. Эти средства нужно распределить между предприятиями Π_1, Π_2, ..., Π_N. Каждое предприятие $_t$ при вложении в него средств и приносит доход, зависящий отu, т.е. представляющий собой функцию $\varphi_t(u)$, которая считается известной. Функции $\varphi_t(u)(t = 1, 2, ..., N)$ являются неубывающими. Задача заключается в распределении средств между предприятиями с целью получения максимального суммарного дохода. Хотя в такой постановке задача не содержит упоминания о времени, можно все же операцию распределения средств между предприятиями мысленно развернуть в какой-то последовательности, считая за первый шаг вложение средств в предприятие Π_1, за второй — в Π_2 и т.д. Управляемым объектом здесь являются средства, а состояние объекта определяется параметром $x(t)$ — количеством средств, которые остаются после вложения в t-е предприятие. Управляющим параметром здесь является $u(t)$ — количество средств, вложенных в предприятие Π_t. Очевидно, что

$$x(t) = x(t-1) - u(t), (t = 1, 2, ..., N).$$

Эти равенства представляют собой уравнения состояния. Управление по своему смыслу неотрицательно и не больше общей суммы наличных средств:

$$u(t) \geq 0, u(t) \leq x(t-1)$$

Последние неравенства определяют множество $U_t(x(t-\mathit{1}))$ допустимости управления.Задача наша заключается в нахождении оптимального управления $u(\mathit{1}), u(\mathit{2}), ..., u(N)$, для которого суммарный доход максимален:

$$I = \sum_{k=1}^{N} \varphi_t(u(t)) \to max$$

При этом в начальный момент времени $x(\mathit{0}) = K$, кроме того, $x(N) = \mathit{0}$. Последнее условие означает, что все средства распределяются.

3. Принцип оптимальности Беллмана. Решение задач оптимального управления с дискретным временем основано на особом принципе, который называют принципом оптимальности. Метод решения этих задач называется методом динамического программирования. Сам принцип оптимальности очень прост. Его можно пояснить следующим образом.

Критерий оптимальности или целевой функционал (7.7) является суммой слагаемых, каждое из которых зависит от состояния объекта в начале t-го шага и от управления, выбираемого на этом шаге процесса. Можно пытаться так выбрать управление, чтобы оптимизировать именно это t-е слагаемое. Однако, уже выбрав управление, мы изменим состояние объекта и это, может быть, не позволит нам в будущем получить еще больший выигрыш на других слагаемых. Таким образом, шаговое управление должно выбираться дальновидно, с учетом всех его последствий в будущем. Значит, планируя многошаговую операцию, надо выбирать управление с учетом всех его будущих последствий на еще предстоящих шагах. Однако среди всех шагов есть один, который можно планировать без оглядки на будущее, чтобы он сам, как таковой, принес наибольшую выгоду. Это последний шаг. И если управление уже найдено, и в нем перед последним шагом объект находится в состоянии $\vec{x}(N-1)$, то управление $\vec{u}(N)$ должно быть точкой максимума функции

$$f_0(N, \vec{x}(N-1), \vec{u}(N)).$$

Правда, найти это управление нельзя, поскольку нам неизвестно состояние $\vec{x}(N-1)$ объекта перед последним шагом. Аналогично этому в оптимальном управлении перед T-м шагом, когда объект находится в состоянии $\vec{x}(T-1)$, управления $\vec{u}(T), \vec{u}(T+1), ..., \vec{u}(N)$, должны обеспечивать максимум функции

$$I(T) = \sum_{t=T}^{N} f_0(t, \vec{x}(t-1), \vec{u}(t)). \tag{7.12}$$

Теперь можно сформулировать сам принцип.

Принцип оптимальности. *Оптимальное управление обладает тем свойством, что каково бы ни было начальное состояние и начальное управление, последующее управление должно быть оптимальным по отношению к состоянию, получающемуся в результате действия начального управления.*

Использование этого принципа гарантирует, что управление, выбранное на любом шаге, является не локально лучшим, а лучшим с точки зрения процесса в целом.

Теперь изложим в общих чертах сам метод динамического программирования. Рассматривается задача с заданным . Уже отмечалось, что управление на последнем шаге $\vec{u}(N)$ должно быть точкой максимума функции

$$f_0(N, \vec{x}(N-1), \vec{u}(N)).$$

Этот максимум определить нельзя, т.к. нам неизвестно состояние $\vec{x}(N-1)$ после $N-1$-го шага. Предположим, что мы можем находить максимум функции $f_0(N, \vec{x}, \vec{u})$ по переменным $\vec{u}$ при любых допустимых значениях $\vec{x}$. Естественно, что точка максимума $\vec{u}$ может зависеть от значений $x_1, x_2, ..., x_N = \vec{x}$, т.е. $\vec{u}_{max} = \vec{u}_N(\vec{x}) = u_{1N}(\vec{x}, u_{2N})(\vec{x}), ..., u_{rN}(\vec{x})$. Будем считать функцию $\vec{u}_N(\vec{x})$ определённой. При её нахождении, разумеется, учитываются ограничения на $\vec{u}(\vec{u} \in U_N(\vec{x})$ Если подставить вместо $\vec{u}$ функцию $\vec{u}(\vec{x})$ в выражение для $f_0(N, \vec{x}, \vec{u})$, то мы получим некоторую функцию

$$\omega_N(\vec{x}) = f_0(N, \vec{x}, \vec{u}_N(\vec{x}))$$

Рассмотрим теперь $(N-1)$-й шаг. В его начале объект находится в состоянии $\vec{x}(N-2)$, которое нам неизвестно. При этом,$x_i(N-1) = f_i(N-$

$1, \vec{x}(N-2), \vec{u}(N-1)$. Согласно принципу максимума $\vec{u}(N-1)$ должна быть точкой максимума функции

$$f_0(N-1, \vec{x}(N-2), \vec{u}(N-1)) + \omega_N(\vec{f}(N-1, \vec{x}(N-2), \vec{u}(N-1)))$$

Этот максимум определить нет возможности, т.к. состояние $\vec{x}(N-2)$ нам неизвестно. Пусть нам удается найти максимум функции

$$f_0(N-1, \vec{x}, \vec{u}) + \omega_N(\vec{f}(N-1, \vec{x}, \vec{u})) \tag{7.13}$$

который в общем случае зависит от $\vec{x}$. Обозначим его через $\vec{u}_{N-1}(\vec{x})$ и, подставляя в выражение (13), получим некоторую функцию $\omega_{N-1}(\vec{x})$. Продолжая рассматривать шаги с меньшими номерами, мы на k-ом шаге находим функцию

$$\omega_{k-1}(\vec{x}) = \max_{\vec{u} \in U_{r-1}(\vec{x})} (f_0(k-1, \vec{x}, \vec{u}) - \omega_k(\vec{f}(k-1, \vec{x}, \vec{u}))$$

Последнее уравнение называется функциональным уравнением Беллмана. С помощью этого уравнения мы рекуррентно находим функции

$$\omega_N(\vec{x}), \omega_{N-1}(\vec{x}), ..., \omega_1(\vec{x})$$

и также точки максимума

$$\vec{u}_N(\vec{x}), \vec{u}_{N-1}(\vec{x}), ..., \vec{u}_1(\vec{x}).$$

Эти точки максимума позволяют определить оптимальное управление. Действительно, нам известно начальное состояние $\vec{x}(0)$.

Поэтому управление на первом шаге есть

$$\vec{u}(1) = \vec{u}_1(\vec{x}(0)).$$

Зная это управление, мы находим состояние в начале второго шага с помощью уравнений состояния. Затем находим управление на втором шаге:

$$\vec{u}(2) = \vec{u}_2(\vec{x}(1)).$$

Таким образом, определяется оптимальное управление шага t:

$$\vec{u}(t) = \vec{u}_t(\vec{x}(t-1)).$$

Соответствующая траектория находится с помощью уравнений состояния параллельно с нахождением оптимального управления.

В процессе оптимизации управления методом динамического программирования многошаговый процесс “проходится” дважды. Первый раз — от конца к началу, в результате чего находятся условные оптимальные управления $\vec{u}_t(\vec{x})$ и условные оптимальные выигрыши $\omega_k(\vec{x})$ за оставшийся "хвост" процесса. Второй раз процесс "проходится" от начала к концу, когда нам остается только "прочитать" уже готовые рекомендации и найти безусловное оптимальное управление $\vec{u}(t)$ состоящее из оптимальных шаговых управлений $\vec{u}(1), \vec{u}(2), ..., \vec{u}(N)$ Первый этап — этап условной оптимизации — несравненно сложнее и длительнее второго. Второй этап почти не требует дополнительных вычислений.

7.3. Пример решения задач динамического программирования

Пример 7.1. Пользуясь методом динамического программирования, решим конкретную задачу распределения ресурсов, размер которых K=200 млн.руб., между четырьмя предприятиями П1, П2, П3, П4 (N=4). Функции дохода на каждом из четырех предприятий задаются равенствами:

$$f_1(u) = 0,4u;$$
$$f_2(u) = 0,6u;$$
$$f_3(u) = 0,8u;$$
$$f_4(u) = 0,7u;$$

Таким образом, функции дохода считаются линейными, а общий целевой функционал имеет вид

$$I = 0,4u(1) + 0,6u(2) + 0,8(3) + 0,7u(4).$$

Управление u(t) (t=1,2,3,4) нужно выбрать так, чтобы максимизировать I. Уравнения состояния здесь имеют вид

$$x(t) = x(t-1) - u(t), t = 1, 2, 3, 4.$$

Управления выбираются из условий

$$u(t) \geq 0, u(t) \leq x(t-1)$$

Кроме того, мы считаем, что x(0) = K, x(4) = 0.

Согласно общей схеме на последнем шаге мы максимизируем функцию $f_4(u)$ на отрезке $0 \leq u \leq x(3)$ Очевидно максимум этой функции будет достигаться в точке $x(3)$ Таким образом, мы имеем

$$\omega(x) = 0,7(x); u_4(x) = x.$$

Рассмотрим теперь третий шаг. В его начале состояние есть $x(2)$, причем $x(3) = x(2) - u(3)$. Поэтому на третьем шаге мы должны максимизировать функцию

$$0,8u(3) + 0,7(x(2) - u(3))$$

на отрезке $0 \leq u(3) \leq x(2)$ Следовательно

$$\omega_3(x) = \max_{0 \leq u \leq x}(0,1u + 0,7x) = 0,8x; u_3(x) = x.$$

На втором шаге x(2)=x(1)-u(2), и максимизируется функция

$$0,6u(2) + 0,8(x(1) - u(2)),$$

то есть

$$\omega_2(x) = \max_{0 \leq u \leq x}(0,8x - 0,2u) = 0,8x; u_2(x) = 0.$$

Наконец, на первом шаге $x(1) = x(0) - u(1)$, и максимизируется функция

$$0,4u(1) + 0,8(x(0) - u(1));$$

$$\omega_1(x) = \max_{0 \leq u \leq x}(0,8x - 0,4u) = 0,8x; u_1(x) = 0.$$

Поскольку $x(0) = 200$, то мы получим

$$u(1) = u_1(200) = 0; \omega_1(200) = 160 (= 0,8 * 200);$$

$$x(1) = 200 - 0 = 200.$$

$$u(2) = u_2(x(1)) = u_2(200) = 0; \omega_2(200) = 160;$$

$$x(2) = 200 - 0 = 200.$$

$$u(3) = u_3(x(2)) = u_3(200) = 200; \omega_3(200) = 0;$$

$$x(3) = 200 - 200 = 0.$$

$$u(4) = u_4(x(3)) = u_4(0) = 0,7 * 0 = 0; \omega_4(0) = 0,7 * 0 = 0;$$

$$x(2) = 200 - 0 = 200.$$

Таким образом, мы имеем следующие управления:

$$u(1) = 0; u(2) = 0; u(3) = 200; u(4) = 0.$$

Общий доход при этом составит $\omega 1(200) = 160$ млн. руб.

Этот результат, однако, можно было бы предвидеть заранее, поскольку третье предприятие имеет наивысший коэффициент прибыли на вложенный рубль и выгоднее всего вкладывать все средства именно в это предприятие.

Пример 7.2. Решим теперь задачу об оптимальных мясозаготовках. В начале периода управления на ферме имелось 1200 голов скота. Период управления 4 года. Минимальные ежегодные мясозаготовки составляют 150 голов. После периода управления численность скота на ферме

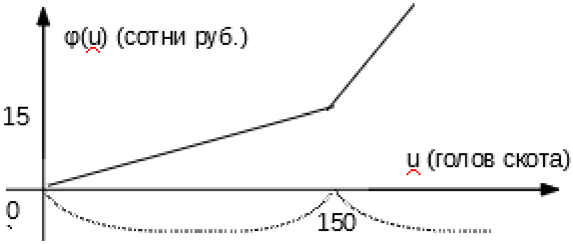

Рис. 7.4. Функция φ(u) примера 7.2

не должна быть меньше 800 голов. Численность скота, оставляемого на ферме для воспроизводства, в следующем году увеличивается в 1,4 раза. Доход от продажи скота выражается функцией, представленной на Рис. 7.4. Введем обозначения: $x(t)$ — количество скота, оставленного на ферме к концу t-го года; $u(t)$ — количество скота, проданное в t-ом году. Имеем уравнение состояния:

$$x(t) = 1,4x(t-1) - u(t), t = 1,2,3,4.$$

Требуется выбрать управление, чтобы максимизировать

$$I = \varphi(u(1)) + \varphi(u(2)) + \varphi(u(3)) + \varphi(u(4)).$$

При этом 150≤u(t)≤1,4x(t-1) и x(0)=1200; x(4)≥800. Последние неравенства означают, что u(4) ≤ 1,4x(3) - 800.

Начинаем оптимизацию с четвертого шага. Нужно определить максимум функции

$$\varphi(u(4)) = u - 135 \text{ при } 150 \leq u(4) \leq 1,4x(3) - 800$$

Поскольку функция φ монотонно растет, то $u_4(x) = 1,4x - 800$,

$$\omega_4(x) = 1,4x - 800 - 135 = 1,4x - 935.$$

На третьем шаге максимизируется функция (x(3) = 1,4x(2) – u(3)):

$$\varphi(u(3)) + \omega_4(1,4x(2) - u(3)) \ \ 150 \leq u(3) \leq 1,4x(2),$$

то есть нужно определить

$$\omega_3(x) = \max_{150 \leq u \leq 1,4x}(u - 135 + 1,4(1,4x - u - 935)) =$$

$$\max_{150 \leq u \leq 1,4x}(1,96 - 0,44u - 1070) = 1,96x - 1130;$$

$$u_3(x) = 150.$$

На втором шаге максимизируется функция

$$\varphi(u(2)) + \omega_3(1,4x(1) - u(2)) \text{ при условиях } 150 \leq u(2) \leq 1,4x(1).$$

$$\omega_1(x) = \max_{150 \leq u \leq 1,4x} (u - 135 + 1,96(1,4x - u) - 1130)$$

$$= \max_{150 \leq u \leq 1,4x} (2,704x - 0,96u - 1265) = 2,704 - 1409$$

$$u_2(x) = 150.$$

На первом шаге максимизируется функция

$$\varphi(u(1)) + \omega_2(1,4x(0) - u(1)) \text{ при условиях } 150 \leq u(1) \leq 1,4x(0).$$

$$\omega_1(x) = \max_{150 \leq u \leq 1,4x} (u - 135 + 2,704(1,4x - u) - 1409)$$

$$= \max_{150 \leq u \leq 1,4x} (3,7856x - 0,704u - 1544) =$$

$$= 3,7856x - 1649,6;$$

$$u_1(x) = 150.$$

Этап условной оптимизации закончен. Оптимальный выигрыш за четыре года составит

$$\omega_1(x(0)) = \omega_1(1200) = 4542,72 - 1649,6 = 2893,12;$$

$$u(1) = u_1(1200) = 150;$$

$$x(1) = 1,4x(0) - u(1) = 1,4 * 1200 - 150 = 1530;$$

$$u(2) = u_2(1530) = 150;$$

$$x(2) = 1,4x(1) - u(2) = 1,4 * 1530 - 150 = 1992;$$

$$u(3) = u_3(1992) = 150;$$

$$x(3) = 1,4x(2) - u(3) = 1,4 * 1992 - 150 = 2788,8 - 150 = 2638,8;$$

$$u(4) = u_4(2638,8) = 1,4 * 2638,8 - 800 = 2894,32;$$

$$x(4) = 1,4x(3) - u(4) = 800.$$

Таким образом, оптимальная стратегия в данном случае заключается в том, чтобы продавать в первые три года минимальное количество скота (150 голов), а в последнем году продать весь скот, оставив на ферме лишь 800 голов.

Пример 7.3. Попытаемся теперь решить ту же задачу с иной функцией дохода $\varphi(u)$. Пусть эта функция имеет вид Рис. 7.5

$$\varphi(u) = \begin{cases} 2u, u \leq 300 \\ 0,5u + 450, u > 300. \end{cases}$$

Все остальные условия предыдущей задачи остаются прежними.

На четвертом шаге

$$\omega_4 = \underset{150 \leq u \leq 1,4(x) - 800}{max} \begin{cases} 2(1,4x - 800)1,4x - 800 \leq 300 \\ 0,5(4x - 800) + 450, 1,4x - 800 > 300. \end{cases}$$

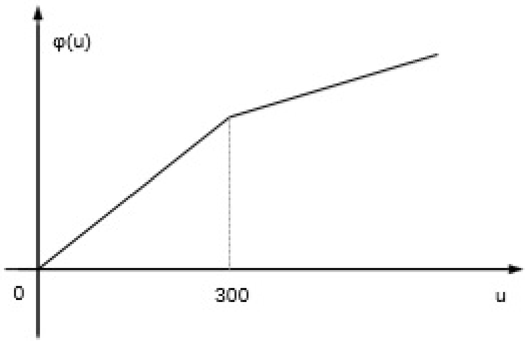

Рис. 7.5. функция $\varphi(u)$ примера 7.3

$$\omega_4 = \begin{cases} 2,8x - 1600, x \leq \frac{5500}{7}; \\ 0,7x - 50, x > \frac{5500}{7}; \end{cases}$$

$$u4(x) = 1,4x\breve{\ }800$$

На третьем шаге $(x(3) = 1,4x(2) - u(3))$:

$$\omega_3(x) = \max_{150 \leq u \leq 1,4(x)} \varphi(u) + \omega_4(1,4x - u)).$$

Вычисление функции ω_3(x) встречает большие трудности и требует большой изобретательности. Нахождение функций ω_2(x) и ω_1(x) встречает еще большие трудности. Таким образом, аналитическое решение задач динамического программирования часто приводит к сложным задачам на глобальный экстремум для функций, зависящих от параметра. Преодолеть эти трудности часто удается с помощью численной реализации алгоритма динамического программирования на ЭВМ.

При численной реализации управляющим параметрам придают лишь дискретное множество значений, которые, например, кратны некоторым величинам. В одних задачах это новое условие является естественным, в других — вводится искусственно. При этом решение, полученное с этим дополнительным условием, дает приближенное представление об истинном оптимальном управлении. Так, в задаче примера 3 предположение, что сдача скота на мясозаготовки может происходить партиями по 50 голов является искусственным. Однако при этом количество возможных значений параметра u будет конечным. Конечным будет и количество возможных состояний. Предположение же, чтоu принимает значения, кратные единице, является естественным. Однако оно приводит к очень большому числу (хотя и конечному) возможных управлений и состояний. Численная реализация задачи тем проще, чем меньше возможное количество управлений и состояний нужно рассматривать на каждом шаге управления. Поэтому предположение, что u в примере 7.3 принимает значения, кратные 50, приводит к более простой численной реализации задачи. Впрочем, проследить эту реализацию и в этом случае для нас сложно из-за большого числа возможных управлений.

Пример 7.4. Рассмотрим задачу о распределении ресурсов. Исходный запас средств K=10 (условных единиц). Требуется его оптимальным образом распределить между пятью (N=5) предприятиями. Предполагается, что вкладываются лишь целые количества средств. Функции дохода φ_t(u) заданы в следующей таблице:

u	φ_1(u)	φ_2(u)	φ_3(u)	φ_4(u)	φ_5(u)
1	0,5	0,1	0,6	0,3	1,0
2	1,0	0,5	1,1	0,6	1,2
3	1,4	1,2	1,2	1,3	1,3
4	2,0	1,8	1,4	1,4	1,3
5	2,5	2,5	1,6	1,5	1,3
6	2,8	2,9	1,7	1,5	1,3
7	3,0	3,5	1,8	1,5	1,3

В каждом столбце, начиная с какой-то суммы вложений, доходы перестают возрастать (остаются постоянными). Это связано с тем, что каждое предприятие способно "освоить"лишь ограниченное количество средств. Число возможных значений u на первом шаге равно 11 (с учетом значения u=0). Поскольку функция φ_t(u) стабилизируется при t$\geq$7, мы не приводим еще трех строк этой таблицы. Напомним, что уравнение состояния здесь имеет вид

$$x(t) = x(t-1) - u(t), t = 1, 2, 3, 4, 5$$

где $x(t)$ — количество еще не распределенных средств;$u(t)$ — количество средств, выделяемых предприятию с номером t,

$$0 \leq u(t) \leq x(t-1), t = 1, 2, 3, 4, 5.$$

Кроме того,$u(t)$ при каждом t может принимать целые значения.

$$x(0) = 10, x(5) = 0.$$

Мы должны максимизировать функцию $I = \sum\limits_{t=1}^{5} \varphi_t(u(t))$.

При t=5 мы находим $\omega_5(x) = \max\limits_{0 \leq u \leq x} \varphi_5(u)$

Поскольку функция φ_5(x) не убывающая, то $\omega_5(x) = \varphi_5(x), u_5(x) = x$.

Результаты условной оптимизации заносим в следующую таблицу:

	t=5		t=4		t=3		t=2		t=1	
x	u_5	ω_5	u_4	ω_4	u_3	ω_5	u_2	ω_2	u_1	ϖ_1
1	$\boxed{1}$	1,0	$\boxed{0}$	1,0	0	1,0	0	1,0		
2	2	1,2	1	1,3	1	1,6	0	1,6		
3	3	1,3	2	1,6	2	2,1	0	2,1		
4	4	1,3	3	2,3	2	2,4	0	2,4		
5	5	1,3	3	2,5	1	2,9	0	2,9		
6	6	1,3	4	2,6	2	3,4	5	3,5		
7	7	1,3	4	2,7	2	3,6	5	4,1		
8	8	1,3	5	2,8	4	3,7	$\boxed{5}$	4,6		
9	9	1,3	6	2,8	5	3,9	7	5,1		
10	10	1,3	7	2,8	5	4,1	7	5,6	$\boxed{2}$	$\boxed{5,6}$

Заполняя эту таблицу при $t = 4$, мы находим

$$\omega_4(x) = \max_{0 \leq u \leq x} (\varphi_4(u) + \omega_4(x-u)).$$

Нахождение максимума производится непосредственно при каждом x :

$x = 1$

u	$\varphi_4(u) + \omega_5(1-u)$	
0	$0 + 1,0$	$\boxed{1,0}$
1	$0,3 + 0$	0,3

$\omega_4(1)$=1,0

$u_4(1)$=0

$x = 2$

u	$\varphi_4(u) + \omega_5(2-u)$	
0	$0 + 1,2$	1,2
1	0,3+1,0	$\boxed{1,3}$
2	$0,6 + 0$	0,6

$\omega_4(2)$=1,3

$u_4(2)$=1

$x = 3$

u	$\varphi_4(u) + \omega_5(3-u)$	
0	$0 + 1,3$	1,3
1	0,3+1,2	$1,5$
2	$0,6 + 1,0$	$\boxed{1,6}$
3	$1,3 + 0$	1,3

$\omega_4(3)$=1,6

$u_4(3)$=2

$x = 4$

u	$\varphi_4(u) + \omega_5(4-u)$	
0	$0 + 1,3$	1,3
1	0,3+1,3	$1,6$
2	$0,6 + 1,2$	$1,8$
3	$1,3 + 1,0$	$\boxed{2,3}$
4	$1,4 + 0$	$1,4$

$\omega_4(4)$=2,3

$u_4(4)$=3

$x = 5$

u	$\varphi_4(u) + \omega_5(5-u)$	
0	$0 + 1,3$	1,3
1	0,3+1,3	$1,6$
2	$0,6 + 1,3$	$1,9$
3	$1,3 + 1,2$	$\boxed{2,5}$
4	$1,4 + 1,0$	$2,4$
5	$1,5 + 0$	$1,5$

$\omega_4(5)$=2,5

$u_4(5)$=3

$x = 6$

u	$\varphi_4(u) + \omega_5(6-u)$	
0	$0 + 1,3$	1,3
1	0,3+1,3	$1,6$
2	$0,6 + 1,3$	$1,9$
3	$1,3 + 1,3$	2,6
4	$1,4 + 1,2$	$\boxed{2,6}$
5	$1,5 + 1,0$	$2,5$
6	$1,5 + 0$	$1,5$

$\omega_4(6)$=2,6

$u_4(6)$=4

$x = 7$

u	$\varphi_4(u) + \omega_5(7-u)$	
0	$0 + 1,3$	1,3
1	0,3+1,3	$1,6$
2	$0,6 + 1,3$	$1,9$
3	$1,3 + 1,3$	2,6
4	$1,4 + 1,3$	$\boxed{2,7}$
5	$1,5 + 1,2$	$2,7$
6	$1,5 + 1,0$	$2,5$
7	$1,5 + 0$	$1,5$

$\omega_4(7)$=2,7

$u_4(7)$=4

$x = 8$

u	$\varphi_4(u) + \omega_5(8-u)$	
0	$0 + 1,3$	1,3
1	0,3+1,3	$1,6$
2	$0,6 + 1,3$	$1,9$
3	$1,3 + 1,3$	2,6
4	$1,4 + 1,3$	$2,7$
5	$1,5 + 1,3$	$\boxed{2,8}$
6	$1,5 + 1,2$	$2,7$
7	$1,5 + 1,0$	$2,5$
8	$1,5 + 0$	$1,5$

$\omega_4(8)$=2,8

$u_4(8)$=5

$x = 9$

u	$\varphi_4(u) + \omega_5(9-u)$	
0	$0 + 1,3$	1,3
1	0,3+1,3	$1,6$
2	$0,6 + 1,3$	$1,9$
3	$1,3 + 1,3$	2,6
4	$1,4 + 1,3$	$2,7$
5	$1,5 + 1,3$	$2,8$
6	$1,5 + 1,3$	$\boxed{2,8}$
7	$1,5 + 1,2$	$2,7$
8	$1,5 + 1,0$	$2,5$
9	$1,5 + 0$	$1,5$

$\omega_4(9)$=2,8

$u_4(9)$=6

$x=10$	u	$\varphi_4(u)+\omega_5(10-u)$		
	0	$0+1,3$	1,3	$\omega_4(10)$=2,8
	1	0,3+1,3	$1,6$	$u_4(10)$=7
	2	$0,6+1,3$	$1,9$	
	3	$1,3+1,3$	2,6	
	4	$1,4+1,3$	$2,7$	
	5	$1,5+1,3$	$2,8$	
	6	$1,5+1,3$	$2,8$	
	7	$1,5+1,3$	$\boxed{2,8}$	
	8	$1,5+1,2$	$2,7$	
	9	$1,5+1,0$	$2,5$	
	10	$1,5+0$	$1,5$	

При $t=3$ мы находим

$$\omega_3(x)=\max_{0\leq u\leq x}(\varphi_3(u)+\omega_4(x-u)).$$

Соответствующие столбцы для $t=3$ заполняются аналогично с использованием столбца для $\omega_4(x)$ и столбца $\varphi_3(x)$ таблицы функций дохода. Так же мы заполняем и столбцы, отвечающие $t=2$. При этом

$$\omega_2(x)=\max_{0\leq u\leq x}(\varphi_2(u)+\omega_4(x-u)).$$

При заполнении столбцов для $t=1$ можно ограничиться случаем $x=10$, поскольку начальное состояние $x(0)=10, u(1)=u1(10)=2$. Максимальная прибыль здесь равна $5,6$ условных единиц. Состояние при $t=1 x(1)=x(0)-u(1)=10-2=8$. По таблице значений $u2(x)$ находим $u2(8)=u(2)=5$. Потом $x(2)=x(1)-u(2)=8-5=3$. Далее $u(3)=u3(3)=2$, $x(3)=x(2)-u(3)=3-2=1$, $u(4)=u4(1)=0$, $x(4)=x(3)-u(4)=1-0=1$.

Таким образом, мы имеем следующее оптимальное управление:

$$u(1)=2, u(2)=5, u(3)=2, u(4)=0, u(5)=1,$$

т.е. первому предприятию следует выделить 2 ед. средств, второму — 5 ед., третьему —2 ед., четвертому — 0 ед. и пятому — 1 ед.

Контрольные вопросы и задачи для самостоятельного решения

1. Что входит в описание дискретного управляемого объекта? Приведите пример дискретного управляемого объекта, определив для него переменные состояния, управляющие параметры, уравнение состояния и целевой функционал.
2. В чем заключается основная задача оптимального управления дискретным управляемым объектом? Сформулируйте задачу с подвижными концами, а также задачу с ограничениями на фазовые переменные.
3. Дайте содержательную и математическую формулировки нелинейной транспортной задачи.
4. Сформулируйте задачу об оптимальном распределении ресурсов в виде задачи оптимального управления.
5. В чем заключается принцип оптимальности Беллмана?
6. Опишите метод динамического программирования?

Решение каждой содержательной задачи состоит из двух этапов: 1) этап формализации, то есть составление оптимизационной математической модели; 2) этап решения задачи с помощью того или иного метода. Для задач динамического программирования этап формализации требует выделения управляемой системы, переменных состояния, управляющих параметров (управлений), составления уравнений состояния и целевого функционала. Этап формализации часто вызывает значительные трудности.

По изложенным выше образцам формализовать задачи **7.1; 7.2.**

7.1. В склад емкостью W(м^3) требуется поместить N различных типов оборудования. Объем одной единицы оборудования i-го типа (i=1,2,3,...,N) равен V_i(м^3). Стоимость единицы оборудования i-го типа равна C_i(руб). Определить, сколько оборудования каждого типа следует поместить в склад, чтобы общая стоимость складированного оборудования была максимальной.

7.2. Детали N видов могут обрабатываться на двух станках. Время обработки детали i-го вида (i=1,2,3,...,N) на первом станке равно a_i минут, а на втором станке-b_i минут. Очередность обработки одна и та же: сначала деталь обрабатывается на первом станке, а затем на втором. Требуется выбрать такую последовательность обработки деталей, при которой время изготовления всех деталей будет минимальным.

Следующие конкретные задачи решить методом динамического программирования.

7.3. Найти оптимальный план загрузки склада в условиях задачи 7.1 при W=90м^3; N=3; v_1=24м^3; v_2= 19м^3; v_3=16м^3; c_1=960руб; c_2=500руб; c_3=250руб.

7.4. Найти оптимальную последовательность обработки деталей в условиях задачи 7.2 при N=4; a_1=20мин; a_2= 35мин; a_3=15мин; a_4=50мин; b_1=5мин; b_2=10 мин; b_3=20мин; b_4=7мин.

7.5. Найти оптимальное распределение средств K=700тыс.руб. между тремя предприятиями, если зависимость между капиталовложениями и приростом выпуска продукции задается следующей таблицей:

Объем капиталовложений	Прирост выпуска продукции (тыс. руб.)		
(тыс. руб.)	Предприятие 1	Предприятие 2	Предприятие 3
0	0	0	0
100	30	50	40
200	50	80	50
300	90	90	110
400	110	150	120
500	170	190	180
600	180	210	220
700	210	220	240

Оптимальное распределение средств должно обеспечить максимальный суммарный прирост выпуска продукции.

7.6. Решить методом динамического программирования линейную транспортную задачу со следующими условиями:

Запасы	Потребности			
	70	10	20	20
50	$\boxed{1}$	$\boxed{3}$	$\boxed{4}$	$\boxed{2}$
45	$\boxed{7}$	$\boxed{6}$	$\boxed{3}$	$\boxed{1}$
25	$\boxed{9}$	$\boxed{4}$	$\boxed{8}$	$\boxed{2}$

8. ВАРИАЦИОННОЕ ИСЧИСЛЕНИЕ

8.1. Примеры задач с бесконечным числом степеней свободы

Всюду выше рассматривались оптимизационные задачи в конечномерном пространстве. Это значит, что целевые функции этих задач являлись функциями конечного числа n переменных, которые удовлетворяли некоторым ограничениям. Однако сравнительно давно появились оптимизационные задачи, в которых разыскиваемую «точку оптимума» нельзя охарактеризовать конечным числом параметров (степеней свободы). Подобные задачи и составляют содержание вариационного исчисления, основанного в XVIII веке Л. Эйлером и Ж. Лагранжем, и значительно развитого в XIX веке в трудах ряда выдающихся математиков. В настоящее время вариационное исчисление превратилось в один из важнейших разделов теоретической и прикладной математики. Приведем ряд примеров задач, сыгравших важную роль в зарождении вариационного исчисления.

Пример 8.1. Исторически первой задачей, известной из глубокой древности и отнесенной впоследствии к задачам вариационного исчисления, явилась так называемая задача Дидо. Согласно легенде Дидо – царица одного из государств Древней Греции, преследуемая царем соседнего государства, бежала в Северную Африку и попросила у местного населения участок земли, который можно охватить шкурой вола. Получив согласие на столь ничтожную просьбу, она на глазах изумленных зрителей разрезала шкуру на тонкие ремешки и, связав их друг с другом, охватила полученной нитью довольно большой по тем временам участок. Развернув на нем строительство, она основала на этом участке знаменитый в древности город Карфаген.

Античные ученые заинтересовались математической стороной этой легенды: если нить уже связана, как расположить ее, чтобы охватить участок с площадью, наибольшей из возможных. В этой задаче требуется найти кривую заданной длины, которая охватывает наибольшую площадь. Она имеет ряд вариантов, и впоследствии получила название изопериметрической задачи. Уже в древности было обнаружено, что искомой формой нити здесь является дуга окружности (рис. 8.1). Один из вариантов задачи состоит в максимизации интеграла

$$S = \int\limits_a^b y(x)\,dx$$

($y = y(x)$–уравнение сухопутной границы участка), при заданном значении длины нити, то есть интеграла

$$L = \int\limits_a^b \sqrt{1+y'^2}\, dx$$

и заданных краевых условиях $y(a) = y(b) = 0$.

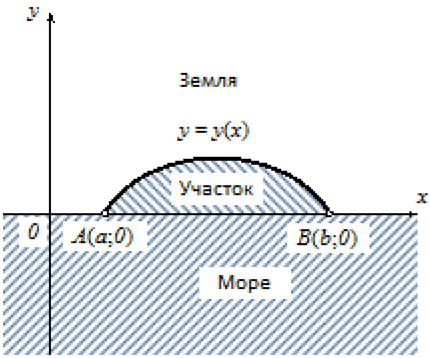

Рис. 8.1

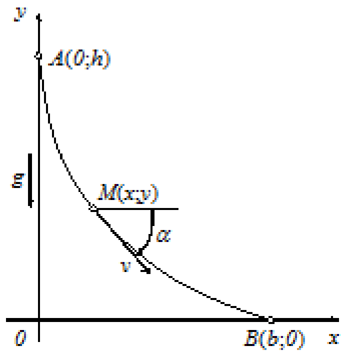

Рис. 8.2

Пример 8.2. Другой знаменитой задачей, приведшей к созданию методов вариационного исчисления, явилась задача о брахистохроне, предложенная И. Бернулли в 1696 году. Она состоит в следующем. Рассматривается материальная точка M, которая скатывается под действием силы тяжести без трения вдоль некоторой линии $y(x)$ из заданной точки A в заданную точку B (рис. 8.2), затрачивая на это время T. Требуется так выбрать путь скатывания $y(x)$, чтобы время T было минимальным из возможных. Перейти к чисто математической задаче здесь помогает закон сохранения энергии

$$gm(h-y) = \frac{mv^2}{2}, \text{ откуда } v = \sqrt{2g(h-y)}.$$

Далее

$$\frac{dx}{dt} = v\cos\alpha = \frac{\sqrt{2g(h-y)}}{\sqrt{1+y'^2}}, \text{ то есть } dt = \sqrt{\frac{1+y'^2}{2g(h-y)}}dx.$$

Поэтому

$$T = \int dt = \int\limits_0^b \sqrt{\tfrac{1+y'^2}{2g(h-y)}}\, dx.$$

Таким образом, речь идет о подборе функции $y(x)$, минимизирующей значение интеграла T. При этом функция $y(x)$ должна дополнительно удовлетворять граничным условиям $y(0) = h$,

$y(b) = 0$.

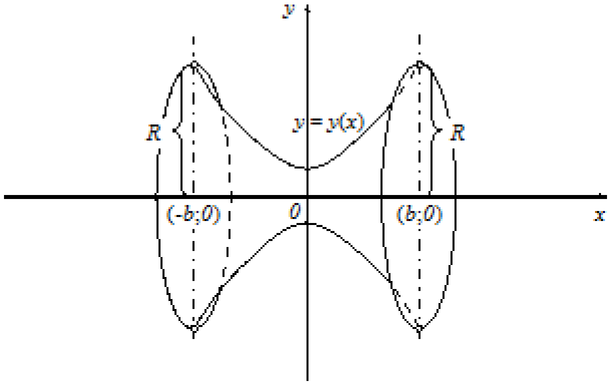

Рис. 8.3

Пример 8.3. Рассмотрим также задачу о форме равновесия мыльной пленки, натянутой на два кольца, насаженных на общую ось (рис. 8.3). Для простоты будем считать кольца одинаковыми, а ось, соединяющую их центры, перпендикулярной параллельным плоскостям колец. Примем также естественную гипотезу, что поверхность пленки имеет форму поверхности вращения, которая получается вращением вокруг оси центров колец кривой, заданной уравнением $y = y(x)$. Если пренебречь весом пленки, то из теории поверхностного натяжения будет следовать, что пленка должна иметь минимально возможную площадь. Математически задача сведется к минимизации интеграла

$$\sum = \int\limits_{-b}^{b} y\sqrt{1 + y'^2}\, dx,$$

равного площади поверхности вращения, при граничных условиях

$$y(-b) = y(b) = R.$$

Таким образом, и в этом примере речь идет о подборе функции $y(x)$, минимизирующей значение интеграла $\sum$.

Нетрудно увидеть в приведенных примерах общие черты. Прежде всего, в этих примерах появляются задачи на экстремум – минимум или максимум. В предыдущих главах мы имели дело с задачами на экстремум функций $z = f(x_1, x_2, \ldots, x_n)$ от n переменных. Искомым там был набор значений $x_1, x_2, \ldots, x_n$. Такую задачу обычно называют задачей с n степенями свободы. В задачах же этого параграфа искомой является линия, или, на аналитическом языке, функция, от которой требуется только, чтобы она удовлетворяла заданным граничным условиям. Но при произвольном выборе такой функции имеется бесконечное число степеней свободы, поскольку произвольную линию или функцию нельзя задать конечным набором параметров. Таким образом, рассмотренные примеры действительно можно отнести к оптимизационным задачам с бесконечным числом степеней свободы.

8.2. Понятие функционального пространства и функционала в нем. Локальный экстремум функционала. Общая схема исследования функционала на экстремум

Примеры функционалов. Основные понятия вариационного исчисления во многом аналогичны понятиям теории экстремумов функций конечного числа аргументов, где основную роль играет целевая функция со своей областью определения. Если, например, в интеграл $\sum$ примера 8.3 вместо $y(x)$ подставить любую конкретную функцию, заданную на сегменте $[-b, b]$ и удовлетворяющую краевым условиям $y(-b) = y(b) = R$, то он примет определенное числовое значение. *Такой закон соответствия, который каждой функции из некоторого класса функций ставит в соответствие число, называется функционалом.* Задача примера 8.3 состоит в подборе функции $y(x)$ для которой функционал $\sum$ принимает минимальное значение. Задача примера 8.1 является задачей на максимум функционала S. Вариационное исчисление изучает экстремумы функционалов.

Отметим, что функционал определяется как самим законом, сопоставляющим функции число, так и классом функций на котором он определен. Например, функционал $I = I\{y\} = \int\limits_0^2 y^2(x)\,dx$ можно считать определенным на функциях, заданных при $0 \le x \le 2$,принимающих там конечные значения (или даже бесконечные, если интеграл окажется сходящимся). Но его можно также определить на гладких функциях, удовлетворяющих граничным условиям $y(0) = y(2) = 0$. И это будет уже другой функционал. Можно сказать, что *функционал – это функция, у которой независимой переменной $y(x)$ служат обычные функции из некоторого класса функций, а значениями зависимой переменной $I\{y(x)\}$ служат числа.*

Нормированное векторное пространство. Общая точка зрения на функционал. . В теории экстремумов функций конечного числа переменных область определения функции является множеством в векторном пространстве R^n. Для функционалов область определения является множеством в *линейном векторном пространстве функций.*

Линейное векторное пространство H - это совокупность элементов $u, v, ...$, называемых векторами, для которых определены *линейные операции* (сложение $u + v$ двух векторов u, v и умножение αu вектора u на скаляр α), подчиняющиеся обычным правилам линейной алгебры. Ниже скаляры предполагаются *вещественными числами.* Обычно нулевой вектор обозначается символом $\mathbf{0}$, тем же, что и нулевой скаляр.

Векторы $u_1, u_2, ..., u_n$ называются *линейно независимыми*, если их линейная комбинация $\alpha_1 u_1 + ... + \alpha_n u_n$ равна нулю только в том случае, когда $\alpha_1 = ... = \alpha_n = 0$; в противном случае эта система векторов называется *линейно зависимой.* Под *размерностью* $dimH$ векторного пространства H понимают наибольшее число линейно независимых векторов в нем. Если такого наибольшего числа не существует, полагают, что $dimH = \infty$ и векторное пространство называют *бесконечномерным.* В противном случае говорят о *конечномерном* векторном пространстве.

Подмножество $M \subset H$ называется *линейным подпространством* или *однородным линейным многообразием* в H, если M само является векторным пространством относительно линейных операций в H. Размерность M не превосходит размерности пространства H.

Обычно для каждого вектора $u \in H$ определено число $\|u\|$, которое называется нормой вектора u. Норма обладает следующими свойствами:

1)$\|u\| \geq 0$
2)$\|u\| = 0$ тогда и только тогда, когда $u = \mathbf{0}$.
3)$\|\alpha u\| = |\alpha| \|u\|$
4)$\|u + v\| \leq \|u\| + \|v\|$

Линейное векторное пространство H, в котором определена норма, удовлетворяющая условиям 1)-4), называется *нормированным векторным пространством.* В вариационном исчислении наиболее употребительными являются следующие функциональные нормированные векторные пространства.

1. Пространство $C[a, b]$ функций $u(x)$, заданных и непрерывных на конечном отрезке $a \leq x \leq b$ с нормой

$$\|u\| = \max_{a \leq x \leq b} |u(x)|.$$

2. Пространство $C_1[a, b]$ функций $u(x)$, заданных и непрерывных при $a \leq x \leq b$ вместе со своей производной, с нормой

$$\|u\| = \max_{a \le x \le b} |u(x)| + \max_{a \le x \le b} |u'(x)|.$$

Аналогично вводится пространство $C_n[a,b]$ при $n = 2, 3, 4, \dots$

3. Гильбертово пространство $L_2[a,b]$ функций, заданных при $a \le x \le b$, не обязательно непрерывных, для которых норма

$$\|u\| = \sqrt{\int_a^b |u(x)|^2, dx}$$

принимает конечное значение. Это означает, что функция $u(x) \in L_2[a,b]$ должна быть *квадратично суммируемой.*

Легко проверить, что каждая из перечисленных выше норм удовлетворяет условиям 1)-4).

Всякая функция из $C_1[a,b]$ принадлежит $C[a,b]$, а всякая функция из $C[a,b]$ принадлежит $L_2[a,b]$. Тем не менее, $C_1[a,b]$ нельзя считать подпространством $C[a,b]$, поскольку эти пространства рассматриваются неотделимо от своих норм, а нормы в C_1 и C различны.

В нормированном пространстве вводится понятие расстояния ρ между любыми элементами u, v по формуле $\rho(u,v) = \|u - v\|$. Для пространств, состоящих из функций, это расстояние является уклонением функций друг от друга, причем каждая из норм дает свое измерение этого уклонения.

Элементы нормированного пространства не обязательно являются функциями. Поэтому на функционал можно смотреть с абстрактной точки зрения, *как на функцию $I\{u\}$, определенную на некотором множестве* $N(I) \subset H$. Дадим абстрактное определение экстремума функционала. При этом ϵ - окрестностью $U_\epsilon(u)$ точки $u \in H$ будем называть множество элементов $v \in H$, удовлетворяющих неравенству $\|u - v\| < \epsilon$ при некотором $\epsilon > 0$.

Определение 8.1. Элемент $U_0 \in N(I)$ называется точкой локального минимума (максимума) функционала $I\{u\}$, если найдется такая окрестность $U_\epsilon(u_0)$, что для всех $u \in U_\epsilon(u_0) \cap N(I)$ выполнено неравенство $I\{u\} \ge I\{u_0\}$ $(I\{u\} \le I\{u_0\})$. Элемент u_0 естественно назвать локальным экстремумом функционала $I\{u\}$.

Отметим, что в последнем определении, по сути, требуется, чтобы соответствующие неравенства выполнялись для всех элементов $u \in N(I)$ достаточно близких к u_0. Для функциональных пространств достаточная близость означает достаточно малое уклонение функции $u(x)$ от функции $u_0(x)$. Но эти уклонения могут измеряться различными способами в зависимости от того, какое функциональное пространство (норму) взять за основу. В соответствии с этим рассматривают различные типы экстремумов функционалов в функциональных пространствах. В дальнейшем мы будем чаще всего исследовать функционалы, которые естественно рассматривать либо в $C_1[a,b]$, либо в $C[a,b]$. В первом случае функцию $u_0(x)$ (элемент u_0) будем называть *слабым*, а во втором – *сильным экстремумом* функционала. При этом, сильный экстремум всегда будет также и слабым, но обратное не обязательно. Например, для функционала $I\{u\} = \int\limits_0^1 \sqrt[4]{1+u'^2}dx$, определенного на гладких функциях $u(x)$, удовлетворяющих краевым условиям $u(0) = 0$, $u(1) = 1$, функция $u_0(x) = x$ доставляет слабый минимум. В то же время сильного минимума эта функция не дает.

Для изложения абстрактной схемы исследования функционала на экстремум нам понадобиться следующее

Определение 8.2. Функционал $J\{u\}$ в линейном векторном пространстве H называется линейным, если его область определения $N(J) \subseteq H$ является линейным подпространством в H, и он обладает свойствами аддитивности: $J\{u+v\} = J\{u\} + J\{v\}$ и однородности: $J\{Cu\} = CJ\{u\}$, где $u, v \in N(J)$, $C = const$.

Вариация функционала. Необходимые условия экстремума. Предположим, что первоначально мы имели элемент u_0, а затем перешли к элементу $u = u_0 + \delta u$. Здесь элемент δu (символ нужно понимать как единый) принято называть вариацией элемента u_0 при условии, что вновь полученный элемент $u \in N(I)$. В остальном вариация δu считается произвольной и, обычно, малой по норме. Отметим, что вариация δu полностью аналогична конечномерному вектору приращений аргументов в теории экстремумов функций нескольких аргументов. Действуя по аналогии, рассмотрим приращение функционала $\Delta I = I\{u_0 + \delta u\} - I\{u_0\}$. Ниже мы будем предполагать, что функционал $I\{u\}$ обладает следующим свойством: для любого элемента $u_0 \in N(I)$ приращение функционала ΔI является суммой двух слагаемых

$$\Delta I = J\{u_0, \delta u\} + R_1\{u_0, \delta u\}, \tag{8.1}$$

первое из которых при фиксированном u_0 является линейным функционалом относительно δu, а второе удовлетворяет условию

$$|R_1\{u_0, \delta u\}| = o(\|\delta u\|), \text{ при } \|\delta u\| \to 0. \tag{8.2}$$

Отметим, что свойство (8.1), (8.2) аналогично свойству дифференцируемости функций одного или нескольких аргументов, и, если это свойство имеет место, совершенно аналогично можно установить необходимые условия экстремума функционала. Отметим, что слагаемое $\delta I\{u_0, \delta u\} = J\{u_0, \delta u\}$ в (8.1) называют вариацией функционала в данной точке u_0. Этот линейный по δu функционал можно считать определенным во всем пространстве H, хотя в конкретных случаях δu может принадлежать более узкому, чем H подпространству. Понятие вариации $\delta I\{u_0, \delta u\}$ полностью аналогично понятию дифференциала функции нескольких аргументов.

Теорема 8.1. *Если элемент u_0 является точкой экстремума функционала $I\{u\}$, то при любой возможной вариации δu аргумента вариация функционала при данном u_0 равна нулю*

$$\delta I\{u_0, \delta u\} = 0.$$

Доказательство. Предположим противное, то есть что найдется такая возможная вариация δu, при которой $\delta I\{u_0, \delta u\} > 0$. Для определенности будем считать, что $\delta I\{u_0, \delta u\} > 0$. При некоторой константе t рассмотрим приращение

$$\Delta I = I\{u_0 + t\delta u\} - I\{u_0\} = \delta I\{u_0, t\delta u\} + R_1\{u_0, t\delta u\} = t(\delta I\{u_0, \delta u\}+ +t^{-1} R_1\{u_0, t\delta u\}).$$

В силу условия (8.2) при $t \to 0$ второе слагаемое в последних скобках стремиться к нулю, и эти скобки положительны при достаточно малых $|t|$. Это значит, что знак приращения ΔI совпадает со знаком t и может принимать как положительные, так и отрицательные значения. Но поскольку u_0 по условию является экстремумом функционала приращение ΔI при достаточно малых $|t|$ не должно изменять знак. Пришли к противоречию. Теорема доказана.

Вариации функционала высших порядков. Достаточные условия экстремума. Еще раз подчеркнем глубокую аналогию между дифференциальным и вариационным исчислениями. Ее наличие позволяет ожидать, что при переходе к достаточным условиям экстремума функционала появиться понятие, играющее в этом вопросе ту же роль, что и дифференциал второго порядка при исследовании экстремума функций нескольких аргументов. Этим понятием является вторая вариация.

Для многих типов функционалов справедлива формула

$$\Delta I = I\{u_0 + \delta u\} - I\{u_0\} = \delta I\{u_0, \delta u\} + (\frac{1}{2!})\delta^2 I\{u_0, \delta u\} + \cdots + +(\frac{1}{n!})\delta^n I\{u_0, \delta u\} + R_n\{u_0, \delta u\},$$

совершенно аналогичная формуле Тейлора для функций конечного числа аргументов. Здесь $\delta^2 I\{u_0, \delta u\}$, $\delta^3 I\{u_0, \delta u\}$,...,$\delta^n I\{u_0, \delta u\}$ – *вариации функционала соответственно второго, третьего и т. д. порядков.* (*кратко — вторая, третья и т. д. вариации*). Они являются функционалами относительно δu уже не линейными, однако обладающими свойством однородности соответствующей степени

$$\delta^n I\{u_0, k\delta u\} = k^n \delta^n I\{u_0, \delta u\}, k = const.$$

Это означает, что члены $\delta^n I\{u_0, \delta u\}$ при $n = 2, 3, \ \ldots$ имеют последовательно второй, третий и т. д. порядок малости относительно $\|du\|$. Остаточный член $R_n\{u_0, \delta u\}$ обладает свойством

$$|R_n\{u_0, \delta u\}| \leq C\|\delta u\|^{n+1}. \tag{8.3}$$

Элемент $u_0 \in N(I)$, удовлетворяющий при всех возможных δu условию $(\delta I\{u_0, \delta u\} = 0)$ обычно называют *стационарным элементом функционала* $I\{u\}$. Всюду ниже мы предполагаем справедливость аналога формулы Тейлора при $n = 2$. Характер стационарного элемента можно выяснить с помощью *достаточных условий экстремума функционала*

Теорема 8.2. *Пусть элемент u_0 является стационарным элементом функционала $I\{u\}$. Тогда справедливы следующие утверждения.*

1^0. *Если при любой возможной вариации δu вторая вариация функционала удовлетворяет неравенству $\delta^2 I\{u_0, \delta u\} \geq m\|\delta u\|^2$ при некоторой константе $m > 0$, то стационарный элемент u0 является точкой локального минимума функционала.*

2^0. *Если при любой возможной вариации δu вторая вариация функционала удовлетворяет неравенству $\delta^2 I\{u_0, \delta u\} \leq -m\|\delta u\|^2$ при некоторой константе $m > 0$, то стационарный элемент u_0 является точкой локального максимума функционала.*

3^0. *Если при различных возможных вариациях δu вторая вариация функционала $\delta^2 I\{u_0, \delta u\}$ принимает значения разных знаков, то стационарный элемент u_0 не является точкой локального экстремума функционала (функционал имеет при $u = u_0$ минимакс).*

Д о к а з а т е л ь с т в о. Запишем приращение $\Delta I = I\{u_0 + \delta u\} - I\{u_0\}$ с помощью аналога формулы Тейлора с $n = 2$, учитывая, что u_0 - стационарный элемент функционала $I\{u\}$. Получим

$$\Delta I = I\{u_0 + \delta u\} - I\{u_0\} = \frac{1}{2}\delta^2 I\{u_0, \delta u\} + R_2\{u_0, \delta u\}$$

При условиях 1^0 с учетом свойства (8.3) получаем оценку $\Delta I \geq \frac{m}{2}\|\delta u\|^2 - C\|\delta u\|^3$, из которой следует положительность ΔI при $\|\delta u\| < \frac{m}{2C}$. То есть стационарный элемент u_0 является точкой локального минимума функционала, и пункт 1^0 доказан. При условиях 2^0 точно также получаем оценку $\Delta I \leq -\frac{m}{2}\|\delta u\|^2 + C\|\delta u\|^3$, из которой следует отрицательность ΔI при $\|\delta u\| < \frac{m}{2C}$. Тем самым и пункт 2^0 тоже доказан.

Для доказательства пункта 3^0 теоремы достаточно доказать, что при его условиях в любой окрестности стационарного элемента u_0 найдутся элементы u_1, u_2, для которых выполнены неравенства $I\{u_1\} > I\{u_0\}$ и $I\{u_2\} < I\{u_0\}$. По условию пункта 3^0 найдутся такие $(\delta u)_1$ и $(\delta u)_2$, что $\delta^2 I\{u_0, (\delta u)_1\} > 0$ и $\delta^2 I\{u_0, (\delta u)_2\} < 0$. Будем искать упомянутые u_1, u_2 в виде $u_1 = u_0 + t(\delta u)_1$, $u_2 = u_0 + t(\delta u)_2$, где t –некоторое число. Тогда справедливы следующие оценки.

$$I\{u_0 + t(\delta u)_1\} - I\{u_0\} = \tfrac{1}{2}\delta^2 I\{u_0, t(\delta u)_1\} + R_2\{u_0, t(\delta u)_1\} \geq \tfrac{t^2}{2}\delta^2 I\{u_0,$$
$$(\delta u)_1\} - C\|t(\delta u)_1\|^3 = t^2(\tfrac{1}{2}\delta^2 I\{u_0, (\delta u)_1\} - Ct\|(\delta u)_1\|^3);$$
$$I\{u_0 + t(\delta u)_2\} - I\{u_0\} = \tfrac{1}{2}\delta^2 I\{u_0, t(\delta u)_2\} + R_2\{u_0, t(\delta u)_2\} \leq \tfrac{t^2}{2}\delta^2 I\{u_0,$$
$$(\delta u)_1\} - C\|t(\delta u)_2\|^3 = t^2(\tfrac{1}{2}\delta^2 I\{u_0, (\delta u)_2\} - Ct\|(\delta u)_2\|^3);$$

Эти оценки показывают, что при всех достаточно малых $|t|$ выполнены неравенства $I\{u_1\} > I\{u_0\}$ и $I\{u_2\} < I\{u_0\}$ и теорема доказана.

Отметим, что теорема 8.2 не охватывает случай, когда $\delta^2 I\{u_0, \delta u\}$ знака не меняет, но обращается в нуль при $\delta u \neq 0$. Анализ этого сложного случая требует рассмотрения вариаций функционала порядка выше второго.

С помощью второй вариации можно сформулировать и необходимые условия экстремума.

Теорема 8.3. *Если элемент u_0 доставляет локальный минимум (максимум) функционалу $I\{u\}$, то для всех возможных δu выполняется неравенство $\delta^2 I\{u_0, \delta u\} \geq 0$ ($\delta^2 I\{u_0, \delta u\} \leq 0$).*

Доказательство. Докажем наше утверждение для случая, когда u_0 доставляет локальный минимум функционалу $I\{u\}$. Случай максимума рассматривается аналогично. Предположим противное, т. е., что найдется такая возможная вариация δu, для которой $\delta^2 I\{u_0, \delta u\} < 0$. Тогда, учитывая, что по теореме 8.1 u_0 является стационарным элементом, аналогично доказательству пункта 3^0 теоремы 8.2 можно написать, что для любого вещественного t справедливо неравенство

$$I\{u_0 + t\delta u\} - I\{u_0\} = \tfrac{1}{2}\delta^2 I\{u_0, t\delta u\} + R_2\{u_0, t\delta u\} \leq \tfrac{t^2}{2}\delta^2 I\{u_0,$$
$$\delta u\} + C\|t\delta u\|^3 = t^2(\tfrac{1}{2}\delta^2 I\{u_0, \delta u\} + Ct\|\delta u\|^3),$$

из которого следует, что при достаточно малом $|t|$ величина $I\{u_0 + t\delta u\} - I\{u_0\} < 0$. Последнее невозможно, поскольку u_0 является минимумом функционала $I\{u\}$. Теорема доказана.

До сих пор мы имели дело с локальными экстремумами функционала. *Элемент u_0 называют точкой глобального (абсолютного) минимума (максимума), если при всех $u \in N(I)$ выполнено неравенство $I\{u\} \geq I\{u_0\}$ ($I\{u\} \leq I\{u_0\}$).* Очевидно, что всякая точка глобального экстремума является и точкой локального экстремума, но, вообще говоря, не наоборот.

Задача на минимум функционала называется *одноэкстремальной*, если любой локальный минимум этого функционала является и глобальным минимумом. Аналогично определяется одноэкстремальность задачи на максимум. Точно также как и в конечномерном случае в любом линейном векторном пространстве можно ввести понятие выпуклого множества, а также *выпуклого функционала.*

Определение 8.3. Функционал $I\{u\}$ в линейном векторном пространстве H называется выпуклым, если его область определения $N(I)$ является *выпуклым* множеством и для любых u_1, $u_2 \in N(I)$ при всяком $\alpha \in [0,1]$ выполнено неравенство $I\{(1-\alpha)u_1+\alpha u_2\} \le (1-\alpha)I\{u_1\}+\alpha I\{u_2\}$. Функционал, для которого выполняется неравенство противоположное последнему, называют *вогнутым.*

Повторяя рассуждение параграфа 6.3, легко установить, что любая задача на минимум для выпуклого функционала является одноэкстремальной.

Пример 8.4. Функционал вида

$$I\{u\} = \int\limits_a^b (p(x)u'^2 + q(x)u^2)dx,$$

где функции $p(x)$, $q(x) \ge 0$ – непрерывны, заданный на гладких функциях, удовлетворяющих краевым условиям $u(a) = c_1$, $u(b) = c_2$ является выпуклым. Действительно, если функции $u_1(x)$ и $u_2(x)$ удовлетворяют краевым условиям, то тем же условиям удовлетворяет функция $(1-\alpha)u_1+\alpha u_2$ при любом $\alpha \in [0,1]$, т. е. область определения функционала является выпуклым множеством. Поскольку функция одной переменной $t2$ является выпуклой, то при любых вещественных $t1$ и $t2$ справедливо неравенство $((1-\alpha)t_1+\alpha t_2)^2 \le (1-\alpha)t_1^2+\alpha t_2$ $(a \in [0,1])$. Значит для любых гладких функций $u_1(x)$ и $u_2(x)$ при любом x выполнены неравенства: $((1-\alpha)u_1+\alpha u_2)^2 \le (1-\alpha)u_1^2+\alpha u_2^2$ и $((1-\alpha)u_1'+\alpha u_2')^2 \le (1-\alpha)u_1'^2+\alpha u_2'^2$. Из неотрицательности функций $p(x)$, $q(x)$, а также из возможности интегрирования неравенств получаем, выпуклость нашего функционала.

Можно продолжать развивать теорию экстремумов функционалов по аналогии с теорией функций конечного числа аргументов. Однако для решения конкретных задач нужно научиться реализовывать уже изложенную общую схему для конкретных типов функционалов. Этим мы и займемся в следующих параграфах.

8.3. Простейшая задача вариационного исчисления

В общей постановке простейшей задачи вариационного исчисления требуется найти гладкую функцию $u(x)$, удовлетворяющую заданным краевым условиям $u(a) = u_1$, $u(b) = u_2$ и доставляющую экстремум функционалу

$$I\{u\} = \int\limits_a^b f(x, u, u')dx. \qquad (8.4)$$

При этом функционал Iu рассматривается в одном из функциональных пространств $C[a, b]$ или $C_1[a, b]$. Иногда эту задачу рассматривают и в более широких функциональных пространствах. Функцию трех аргументов $f(\cdot, \cdot, \cdot)$ мы будем считать дифференцируемой.

Следует подчеркнуть, что сформулированная задача не всегда имеет решение. Вот классический пример, принадлежащий Вейерштрассу (1815–1897).

Пример 8.5. Пусть требуется найти минимум функционала $I\{u\} = \int\limits_{-1}^{1} x^2\, u'^2 dx$ в классе функций $C_1[-1, 1]$, при краевых условиях $u(-1) = -1$, $u(1) = 1$, которые задают область определения $N(I)$ функционала. Рассмотрим функцию $u_\alpha(x) = \operatorname{arctg}\dfrac{x}{\alpha}\Big/\operatorname{arctg}\dfrac{1}{\alpha}$ которая входит в область определения функционала при любом $a > 0$. Очевидно, что

$$I\{u_\alpha\}=\left(a^2\Big/\operatorname{arctg}^2\frac{1}{\alpha}\right)\int\limits_{-1}^{1}\frac{x^2dx}{(a^2+x^2)^2}<\left(a^2\Big/\operatorname{arctg}^2\frac{1}{\alpha}\right)\int\limits_{-1}^{1}\frac{dx}{a^2+x^2}=\frac{2a}{\operatorname{arctg}\dfrac{1}{\alpha}}$$

Следовательно, $\lim\limits_{a\to 0} I\{u_\alpha\} = 0$ Однако, на всякой допустимой функции $Iu > 0$. Поэтому число 0, являющееся точной нижней гранью функционала, минимумом не будет и рассматриваемая задача решения не имеет. Точнее у функционала нет точек глобального минимума. Поскольку функционал имеет вид функционала примера 8.4 , т. е. является выпуклым, задача на минимум для него является одноэкстремальной. Следовательно, отсутствуют и точки локального минимума.

Пример 8.6. Покажем, что функционал $I\{u\} = \int\limits_0^1 (x^2u'^2 + 12u^2)dx$ в классе функций $C_1[0, 1]$, при краевых условиях $u(0) = 0$, $u(1) = 1$ достигает глобального минимума на функции $u_0(x) = x^3$ и только на ней.

Представим любую функцию из области определения функционала $N(I)$ в виде $u(x) = u_0(x)+\delta u$, где δu–дифференцируемая функция, для которой $\delta u(0) = \delta u(1) = 0$. Последние краевые условия возникают из-за того, что должно быть $u(x) = u_0(x) + \delta u \in N(I)$. С помощью интегрирования по частям нетрудно установить, что

$$I\{u\} = I\{u_0 + \delta u\} =$$

$$\int_0^1 (x^2(x^3 + \delta u)'^2 + 12(x^3 + \delta u)^2)dx = \int_0^1 (21x^6 + (\delta u)'^2 + 12(\delta u)^2)dx$$

Поскольку $I\{u_0\} = 21\int_0^1 x^6\,dx$, то при $\delta u \neq 0$ справедливо неравенство $I\{u\} > I\{u_0\}$.Таким образом, на функции $u_0(x) = x^3$ достигается глобальный (абсолютный) минимум. Отметим, что и функционал этого примера также имеет вид функционала примера 8.4, т. е. является выпуклым.

В общем случае простейшей задачи вариационного исчисления можно найти вариацию $\delta I\{u_0, \delta u\}$ при любом $u_0 \in N(I)$. Затем, применяя теорему 8.1, можно определить все функции $u_0(x)$, которые могут быть точками экстремума. При этом будет полезной следующая

Лемма 8.1. *Пусть $M(x)$ функция, непрерывная на отрезке $[a, b]$ и удовлетворяющая условию*

$$\int_a^b M(x)\varphi(x)dx = 0$$

для всех функций $\varphi(x)$, непрерывных на отрезке $[a, b]$ вместе с производными до некоторого порядка $m \geq 0$ и обращающихся в нуль на концах этого отрезка вместе с указанными производными. Тогда $M(x) = 0$ при всех $x \in [a, b]$.

Доказательство. Предположим противное. Для определенности предположим, что найдется такое $c \in (a, b)$, что $M(c) > 0$. В силу непрерывности $M(x)$ найдется интервал $(c - \delta, c + \delta)$, в котором $M(x) > 0$. Рассмотрим функцию

$$\varphi_0(x) = \begin{cases} [\delta^2 - (x - c)^2]^\gamma, (x \in [c - \delta, c + \delta]), \\ 0, \ (x \notin [c - \delta, c + \delta]), \end{cases}$$

которая удовлетворяет всем наложенным на $\varphi(x)$ условиям при $\gamma > m$. Поэтому

$$0 = \int_a^b M(x)\varphi_0(x)dx = \int_{c-\delta}^{c+\delta} M(x)[\delta^2 - (x - c)^2]^\gamma dx$$

Последнее абсурдно, поскольку подынтегральное выражение в правой части положительно.

Уравнение Эйлера. Пользуясь дифференцируемостью функции $f(\cdot,\cdot,\cdot)$ в (8.4) определим вариацию $\delta I\{u_0,\delta u\}$ функционала для любого $u_0 \in N(I)$.При этом сам функционал будем рассматривать в пространстве $C_1[a,b]$. Отметим, что δu должна быть гладкой функцией, удовлетворяющей условиям $\delta u(a)=\delta u(b)=0$. Приращение функционала имеет вид

$$\Delta I=\int_a^b (f(x,u_0(x)+\delta u,u_0'(x)+\delta u')-f(x,u_0(x),u_0'(x)))dx$$

Поскольку справедливо равенство

$$f(x,u_0(x)+\delta u,u_0'(x)+\delta u')-f(x,u_0(x),u_0'(x))=$$
$$f_u'(x,u_0(x),u_0'(x))\delta u+f_{u'}'(x,u_0(x),u_0'(x))\delta u'+r_1(\delta u,\delta u')\text{'}$$

где $r_1(\delta u,\delta u')$ – член более высокого порядка малости, то, отбрасывая этот член, получим вариацию функционала

$$\delta I\{u_0,\delta u\}=\int_a^b (f_u'(x,u_0(x),u_0'(x))\delta u+f_{u'}'(x,u_0(x),u_0'(x))\delta u')dx.$$

Теорема 8.4. *Если функция $u_0(x)$ доставляет слабый экстремум функционалу вида* (8.4) *простейшей задачи вариационного исчисления, то она удовлетворяет дифференциальному уравнению*

$$f_u'(x,u(x),u'(x))-\frac{d}{dx}f_{u'}'(x,u(x),u'(x))=0. \tag{8.5}$$

и краевым условиям $u(a)=u_1$, $u(b)=u_2$.

Доказательство. Согласно теореме 8.1 справедливо равенство

$$\delta I\{u_0,\delta u\}=\int_a^b (f_u'(x,u_0(x),u_0'(x))\delta u+f_{u'}'(x,u_0(x),u_0'(x))\delta u')dx=0$$

С другой стороны δu является гладкой функцией, обращающейся в ноль на концах отрезка $[a,b]$, а $\delta u'$ - ее производная. Поэтому, интегрируя по частям, получим

$$\int_a^b f_{u'}'(x,u_0(x),u_0'(x))\delta u'dx=-\int_a^b \frac{d}{dx}(f_{u'}'(x,u_0(x),u_0'(x)))\delta u\,dx$$

Следовательно

$$\delta I\{u_0,\delta u\}=\int_a^b (f_u'(x,u_0(x),u_0'(x))-\frac{d}{dx}f_{u'}'(x,u_0(x),u_0'(x)))\delta u\,dx=0$$

Применяя лемму 8.1 с $M(x) = f'_u(x, u_0(x), u'_0(x)) - \dfrac{d}{dx} f'_{u'}(x, u_0(x), u'_0(x))$, $\varphi(x) = \delta u$, получаем, что функция $u_0(x)$ удовлетворяет дифференциальному уравнению (8.5). Краевые условия выполнены, поскольку $u_0(x) \in N(I)$. Теорема доказана.

Дифференциальное уравнение (8.5) называется *уравнением Эйлера простейшей задачи вариационного исчисления.* Здесь под производной $\dfrac{d}{dx}$ понимается полная производная, учитывающая зависимость u и u' от x. Уравнение (8.5), вообще говоря, является дифференциальным уравнением второго порядка. Любое решение этого уравнения называется *экстремалью* функционала (8.4). Если по некоторым соображениям известно, что простейшая задача для функционала (8.4) имеет решение, то это решение нужно искать среди экстремалей этого функционала. Экстремалями также часто называют интегральные кривые уравнения (8.5), т. е. графики его решений. Если о существовании экстремума функционала ничего не известно, то можно говорить лишь о нахождении экстремалей, удовлетворяющих краевым условиям, которые являются стационарными элементами функционала.

Совокупность экстремалей, как общее решение обыкновенного дифференциального уравнения второго порядка, является двухпараметрическим семейством функций. Двух граничных условий, вообще говоря, достаточно для нахождения требуемого частного решения. Однако уравнения Эйлера редко интегрируются в квадратурах. Поэтому для нахождения функций, доставляющих экстремум функционалу, часто приходится пользоваться численным интегрированием краевых задач.

Перечислим частные случаи, в которых уравнение Эйлера упрощается по сравнению с общим случаем.

1. Функция $f(\cdot,\cdot,\cdot)$ в (8.5) не зависит от u'. Уравнение Эйлера имеет вид $f'_u = (x, u(x)) = 0$. Это уравнение не является дифференциальным и его решение удовлетворяет краевым условиям лишь в исключительных случаях.

Пример 8.7. Найти экстремаль функционала $I\{u\} = \int\limits_a^b (e^{x+u} - u - \sin x)dx$, удовлетворяющую краевым условиям $u(0) = 0$, $u(1) = -1$. Здесь $f(x, u) = e^{x+u} - u - \sin x$. Имеем уравнение экстремали $f'_u(x,u) = e^{x+u} - 1 = 0$ или $u = -x$, которая удовлетворяет указанным краевым условиям.

2. Функция $f(\cdot,\cdot,\cdot)$ в (8.5) зависит только от u', т. е. $f(x,u,u') = f(u')$. Уравнение Эйлера имеет вид $f''_{u'u'}(u') \cdot u'' = 0$, а его общее решение – вид $u = C_1 x + C_2$. Таким образом, в этом случае экстремалями функционала являются всевозможные прямые.

Пример 8.8. Найти гладкую кривую на плоскости, соединяющую две заданные точки $M1(a,u_1)$ и $M2(b,u_2)$и имеющую минимальную длину. Эту задачу можно сформулировать как задачу нахождения функции $u(x)$, минимизирующей функционал $I\{u\} = \int\limits_a^b \sqrt{1+u'^2}dx$, и удовлетворяющей условиям $u(a) = u_1$, $u(b) = u_2$. Здесь $f(x,u,u') = \sqrt{1+u'^2} = f(u')$. Поэтому экстремалями будут всевозможные прямые, а искомой экстремалью будет прямая, соединяющая точки $M_1(a,u_1)$ и $M_2(b,u_2)$. Из геометрических соображений ясно, что среди гладких кривых, соединяющих две данные точки, кривая минимальной длины должна существовать, а кривая максимальной длины – нет. Поэтому найденная прямая будет решением нашей задачи.

3. Функция $f(\cdot,\cdot,\cdot)$ не зависит от u, т. е. $f(x,u,u') = f(x,u')$. Уравнение (8.5) в данном случае имеет вид $\dfrac{d}{dx} f'_{u'}(x,u'(x)) = 0$. Очевидно, это уравнение имеет первый интеграл $f'_{u'}(x,u'(x)) = C_1$, который является дифференциальным уравнением первого порядка. Решив это уравнение, мы получим экстремали функционала.

Пример 8.9. Найти экстремаль функционала $I\{u\} = \int\limits_{-1}^{1}(xu'+u'^2)dx$, удовлетворяющую краевым условиям $u(-1) = 1$, $u(1) = 0$. Здесь $f(x,u,u') = f(x,u') = \; = xu'+u'^2$. Уравнение Эйлера имеет первый интеграл $x+2u' = C_1$. Общее решение этого дифференциального уравнения первого порядка дает совокупность всех экстремалей функционала $u(x) = -(1/4)x^2 + (1/2)C_1 x + C_2$. Из них заданным краевым условиям удовлетворяет экстремаль $u(x) = -(1/4)x^2 - (1/2)x + 3/4$.

4. Функция $f(\cdot,\cdot,\cdot)$ не зависит от x, т. е. $f(x,u,u') = f(x,u')$. Уравнение Эйлера в данном случае имеет вид

$f'_u(u,u') - \dfrac{d}{dx} f'_{u'}(u,u') = 0$. По формуле дифференцирования сложной функции получим $f'_u(u,u') - f''_{uu'}(u,u')u' - f''_{u'u'}u'' = 0$.Если умножить обе части последнего уравнения на u', то оно запишется в виде $\dfrac{d}{dx}(f(u,u') - u'f'_{u'}(u,u')) = 0$. Отсюда получаем первый интеграл уравнения Эйлера $f(u,u') - u'f'_{u'}(u,u') = C_1$. Интегрируя это дифференциальное уравнение первого порядка, получим совокупность экстремалей функционала.

Пример 8.10. Найти экстремали функционала $I\{u\} = \int\limits_{-b}^{b} u\sqrt{1+u'^2}dx$, удовлетворяющие краевым условиям $u(-b) = u(b) = R$. Отметим, что в этом примере мы фактически демонстрируем решение задачи примера 8.3. Здесь $f(u,u') = u \cdot \sqrt{1+u'^2}$. Поэтому первый интеграл $f(u,u') - u'f'_{u'}(u,u') = C_1$ уравнения Эйлера имеет вид $\dfrac{u}{\sqrt{1+u'^2}} = C_1$. Разрешая последнее уравнение относительно u', разделяя переменные и интегрируя, получим общее решение $u = C_1 \operatorname{ch} \dfrac{x+C_2}{C_1}$. Из симметрии краевых условий следует, что $C_2 = 0$, а C_1 определиться из условия $C_1 \operatorname{ch} \dfrac{b}{C_1} = R$. Получаемая экстремаль называется *цепной линией*, а соответствующая поверхность вращения – *катеноидом.*

Поскольку задача примера 8.10 (или 8.3) имеет физический смысл, и поверхность пленки имеет определенную форму равновесия, задача на минимум имеет решение, которое совпадает с цепной линией.

Пример 8.11. Написать уравнение Эйлера для определенного на гладких функциях квадратичного функционала $I\{u\}=\int\limits_a^b(p(x)u'^2+q(x)u^2)dx$, с заданными функциями $p(x) \in C_1$, $q(x) \in C$ и краевыми условиями $u(a) = u_1$, $u(b) = u_2$. Такой функционал уже рассматривался в примере 8.4, однако здесь мы не предполагаем неотрицательности функций $p(x)$ и $q(x)$. Этот функционал не обязан быть выпуклым. Очевидно, что $f(x,u,u') = p(x)u'^2 + q(x)u^2$; $f'_u(x,u,u') = 2q(x)u$, $f'_{u'}(x,u,u') = 2p(x)u'$. Поэтому уравнение Эйлера имеет следующий вид: $-(p(x)u')' + q(x)u = 0$. Это уравнение является линейным однородным дифференциальным уравнением второго порядка и при $p(x) \neq 0$ имеет на отрезке $[a,b]$ двухпараметрическое семейство экстремалей. Такого семейства может не быть, если $p(x)$ обращается в нуль на $[a,b]$. Так для функционала примера 8.5 уравнение Эйлера $-(x^2u')' = 0$ имеет первый интеграл $x^2u' = C_1$, а также общее решение $u = -(C_1/x) + C_2$. При попытке удовлетворить граничным условиям $u(-1) = -1$, $u(1) = 1$ получим $C_1 = -1$, $C_2 = 0$. Но функция $u = \frac{1}{x}$ не является гладкой и даже непрерывной на отрезке $[-1,1]$. Из этого заключаем, что в примере 8.5 функционал не имеет не только минимума, но и максимума. Для функционала примера 8.6 уравнение Эйлера $-(x^2u')' + 12u = 0$ имеет решение $u = x^3$, которое удовлетворяет краевым условиям и доставляет минимум функционалу.

Условия экстремума, использующие вторую вариацию. Для функционала (8.4) простейшей задачи вариационного исчисления можно найти выражение для второй вариации, однако использование теоремы 8.2 здесь встречает трудности. Несмотря на это, используя вторую вариацию, можно получить некоторые условия экстремума, которые, в частности, вытекают и из этой теоремы. Предположим, что функция $f(\cdot,\cdot,\cdot)$ в (8.4) имеет непрерывные частные производные до третьего порядка включительно. Тогда справедлива обычная формула Тейлора для функции $f(\cdot,\cdot,\cdot)$

$$f(x,u_0(x)+\delta u, u_0'(x)+\delta u') - f(x,u_0(x),u_0'(x)) =$$
$$f'_u(x,u_0(x),u_0'(x))\delta u + f'_u(x,u_0(x),u_0'(x))\delta u' +$$
$$\frac{1}{2}(f''_{uu}(x,u_0(x),u_0'(x))\delta u^2 + 2f_{uu'}(x,u_0(x),u_0'(x))\delta u\delta u'$$
$$+f_{u'u'}(x,u_0(x),u_0'(x))\delta u'^2) + r_2(\delta u,\delta u')$$

где остаточный член имеет более высокий порядок малости. Интегрируя обе части последнего равенства по отрезку $[a,b]$, получим

$$\Delta I = I\{u_0+\delta u\} - I\{u_0\} = \delta I\{u_0,\delta u\} + \frac{1}{2}\delta^2 I\{u_0,\delta u\} + R_2\{u_0,\delta u\},$$

где

$$\delta^2 I\{u_0,\delta u\} =$$

$$\int\limits_a^b f''_{uu}(x,u_0(x),u_0'(x))\delta u^2 + 2f''_{uu'}(x,u_0(x),u_0'(x))\delta u\delta u' + f''_{u'u'}(x,u_0(x),u_0'(x))\delta u'^2)dx.$$

Отметим, что $u_0(x)$ здесь–фиксированная гладкая функция, удовлетворяющая краевым условиям $u(a) = u_1$, $u(b) = u_2$, а δu –

произвольная гладкая функция с нулевыми значениями на концах отрезка. Поскольку $2\delta u \delta u' = (\delta u^2)'$, то интегрируя по частям в выражении для второй вариации, получим

$$\delta^2 I\{u_0, \delta u\} =$$

$$\int_a^b (f''_{uu}(x, u_0(x), u'_0(x))\delta u^2 + \frac{d}{dx} f'_{uu'}(x, u_0(x), u'_0(x)))\delta u^2 dx + \int_a^b f'_{u'u'}(x, u_0(x), u'_0(x))\delta u'^2 dx.$$

Если в качестве δu выбрать функцию, не превосходящую единицы и отличную от нуля только в малой окрестности точки $x_0 \in (a, b)$, то первое слагаемое в выражении для $\delta^2 I\{u_0, \delta u\}$ будет сколь угодно малым за счет малости окрестности и знак всего выражения будет определяться значением $f''_{u'u'}(x, u_0(x), u'_0(x))$. Поэтому справедливы следующие утверждения:

если $f''_{u'u'}(x, u_0(x), u'_0(x))$ при некотором $x \in [a, b]$ принимает положительное значение, то существует допустимая вариация δu, при которой $\delta^2 I\{u_0, \delta u\} > 0$;

если $f''_{u'u'}(x, u_0(x), u'_0(x))$ при некотором $x \in [a, b]$ принимает отрицательное значение, то существует допустимая вариация δu, при которой $\delta^2 I\{u_0, \delta u\} < 0$.

Теорема 8.5. 1^0. *Если функционал $I\{u\}$ при некоторых краевых условиях имеет при $u = u_0(x)$ (хотя бы слабый) минимум (максимум), то при всех $x \in [a, b]$ выполнено неравенство $f''_{u'u'}(x, u_0(x), u'_0(x)) \geq 0$ ($f''_{u'u'}(x, u_0(x), u'_0(x)) \leq 0$).*

2^0. *Если функция $u = u_0(x)$ является стационарным элементом функционала $I\{u\}$ и выражение $f''_{u'u'}(x, u_0(x), u'_0(x))$ при $x \in [a, b]$ принимает значения обоих знаков, то функция $u = u_0(x)$ не является экстремумом функционала (функционал имеет при $u = u_0(x)$ минимакс).*

Доказательство. 1^0. Предположим противное, т. е., что $u_0(x)$ доставляет минимум функционалу и существует $x \in [a, b]$, в котором $f''_{u'u'}(x, u_0(x), u'_0(x)) < 0$. Тогда по первому утверждению перед теоремой 8.5 должна существовать такая допустимая вариация δu, что $\delta^2 I\{u_0, \delta u\} < 0$. Но по теореме 8.3 для всех возможных δu должно быть $\delta^2 I\{u_0, \delta u\} \geq 0$. Пришли к противоречию. Случай максимума рассматривается аналогично. Пункт 1^0 теоремы доказан. 2^0. Согласно утверждениям перед теоремой 8.5 существуют такие возможные вариации $(\delta u)_1$ и $(\delta u)_2$, что $\delta^2 I\{u_0, (\delta u)_1\} > 0$ и $\delta^2 I\{u_0, (\delta u)_2\} < 0$. Но согласно пункту 3^0 теоремы 8.2 функция $u_0(x)$ не доставляет экстремум функционалу и является элементом его минимакса. Теорема доказана.

Отметим, что пункт 1^0 последней теоремы дает еще одно *необходимое условие экстремума* функционала простейшей задачи вариационного исчисления. Это условие называется *необходимым условием Лежандра.* Пункт 2^0 представляет собой достаточное условие отсутствия экстремума в стационарном элементе функционала.

Пример 8.12. Доказать, что квадратичный функционал примера 8.11 с $p(x)$, меняющим знак на отрезке $[a, b]$ не имеет экстремумов.

Предположим противное, т. е. что существует функция $u = u_0(x)$, доставляющая экстремум этому функционалу. По теореме 8.1 эта функция является стационарным элементом функционала. Однако для квадратичного функционала $f''_{u'u'}(x, u_0(x), u_0'(x)) = 2p(x)$ и меняет знак на отрезке $[a, b]$. Согласно пункту 2^0 теоремы 8.5 эта функция не является экстремумом функционала.

Помимо простейшей задачи в вариационном исчислении рассматривается ряд других задач, методы решения которых во многом аналогичны вышеизложенным. Ниже мы формулируем ряд таких задач и опишем классические методы их решения, опуская, в основном, обоснования.

8.4. Обобщения простейшей задачи вариационного исчисления

Задача на экстремум функционала $I\{u(x)\}$, зависящего от производных высших порядков функции $u(x)$. На функциях, имеющих n непрерывных производных, рассмотрим функционал

$$I\{u(x)\} = \int_a^b f(x, u, u^{(1)}, u^{(2)}, \ldots, u^{(n)})dx, \tag{8.6}$$

где функция $f(x, u, u^{(1)}, u^{(2)}, \ldots, u^{(n)})$ имеет непрерывные частные производные до $n+1$-го порядка по всем аргументам, а $u(x)$ удовлетворяет граничным (краевым) условиям

$$u(a) = u_1, u^{(1)}(a) = u_1^{(1)}, \ldots, u^{(n-1)}(a) = u_1^{(n-1)},$$

$$u(b) = u_2, u^{(1)}(b) = u_2^{(1)}, \ldots, u^{(n-1)}(b) = u_2^{(n-1)}.$$

Задача на экстремум этого функционала может быть включена в общую схему исследования, если этот функционал рассматривать в пространстве $C_n[a, b]$. Совершенно аналогично теореме 8.4 доказывается

Теорема 8.6. *Если функционал (8.6) достигает на функции $u(x) \in C_n[a,b]$ локального экстремума, то эта функция удовлетворяет уравнению Эйлера-Пуассона*

$$f'_u - \frac{d}{dx} f'_{u^{(1)}} + \frac{d^2}{dx^2} f'_{u^{(2)}} - \ldots + (-1)^n \frac{d^n}{dx^n} f'_{u^{(n)}} = 0 \tag{8.7}$$

Пример 8.13. Найти экстремали функционала

$$I\{u(x)\} = \int\limits_0^1 (240xu - u''^2)dx,$$

удовлетворяющие краевым условиям $u(0) = u'(0) = 0$, $u(1) = 1$, $u'(1) = 6$. Здесь $f(x, u, u^{(1)}, u^{(2)}) = 240xu - u'^2$, $f'_u = 240x, \frac{d}{dx} f'_{u'} = 2u'', \frac{d^2}{dx^2} f'_{u'} = 2u^{(4)}$. Поэтому уравнение Эйлера-Пуассона имеет вид $u^{(4)} = 120x$. Его общее решение $u(x) = x^5 + C_1 x3 + C_2 x^2 + C_3 x + C_4$. С учетом краевых условий получаем для определения постоянных C_1, C_2, C_3, C_4: $C4_=0$, $C_3 = 0$, $C_1 + C_2 + C_3 + C_4 = 0$, $3C_1 + 2C_2 + C_3 = 1$. Отсюда заключаем, что $C_1 = 1$, $C_2 = -1$, $C_3 = C_4 = 0$. Экстремум функционала может достигаться только на кривой $u(x) = x^5 + x^3 - x^2$.

Задача на экстремум функционала, зависящего от нескольких функций. Еще одним обобщением простейшей задачи является задача на экстремум функционала, зависящего не от одной, а от нескольких функций $u_1(x)$, $u_2(x)$,..., $u_n(x)$ одного переменного x:

$$I\{u_1(x), u_2(x), \ldots, u_n(x)\} = \int\limits_a^b f(x, u_1, u_2, \ldots, u_n, u'_1, u'_2, \ldots, u'_n)dx, \tag{8.8}$$

где функция f(...) имеет непрерывные частные производные до второго порядка включительно, а краевые условия имеют вид $u_k(a) = u_{k1}$, $u_k(b) = u_{k2}$, $k = 1,\ldots, n$. Для включения этой задачи в общую схему нужно ввести линейное векторное пространство $C_1(n)[a,b]$ гладких n-компонентных вектор-функций с нормой $\|u\| = \max_{a \le x \le b} |\vec{u}(x)| + \max_{a \le x \le b} |\vec{u'}(x)|$, где $|\vec{v}|$ - длина вектора $\vec{v}$.

Теорема 8.7. *Если набор функций $u_1(x)$, $u_2(x)$,..., $u_n(x) \in C_1[a,b]$ доставляет слабый экстремум функционалу (8.8), то эти функции удовлетворяют системе дифференциальных уравнений Эйлера*

$$f_{u_k}' - \frac{d}{dx} f_{u_k}' = 0, k = 1, 2, \ldots, n.$$

Пример 8.14. Найти функции $u_1(x)$ и $u_2(x) \in C_1[0, \frac{\pi}{2}]$, на которых может достигаться экстремум функционала

$$I\{u_1(x),u_2(x)\}=\int\limits_0^{\frac{\pi}{2}}(u_1'^2+u_2'^2-2u_1u_2)dx$$

при краевых условиях $u_1(0)=u_2(0)=0$, $u_1(\frac{\pi}{2})=u_2(\frac{\pi}{2})=1$.

Система уравнений Эйлера имеет следующий вид

$$u_1''+u_2=0,$$

$$u_2''+u_1=0.$$

Исключая из второго уравнения функцию $u_2=-u_1''$, получим $u_1^{(4)}-u_1=0$. Общее решение последнего уравнения имеет вид $u_1(x)=C_1e^x+C_2e-x+C_3\cos x+C_4\sin x$. Отсюда находим, что $u_2(x)=-u_1''(x)=-C_1e^x-C_2e-x+C_3\cos x+C_4\sin x$. Из краевых условий следует, что $C_1=C_2=C_3=0$, $C_4=1$. Поэтому $u_1(x)=u_2(x)=\sin x$.

Задача на экстремум функционала, зависящего от функций нескольких аргументов. Рассмотрим случай, когда функционал зависит от функции нескольких аргументов. Для простоты рассмотрим функции $u(x,y)$ двух аргументов, причем точка (x,y) изменяется в некоторой области G плоскости. Тогда функционал, аналогичный (8.4), можно записать в виде

$$I\{u(x,y)\}=\iint\limits_{(G)}f(x,y,u,\frac{\partial u}{\partial x},\frac{\partial u}{\partial y})dxdy. \tag{8.9}$$

Этот функционал будем рассматривать на функциях, равных на границе ∂G области G некоторой заданной функции $\varphi(M)$:

$$\{M(x,y)\in\partial G\}\Rightarrow u(x,y)=\varphi(M). \tag{8.10}$$

Геометрически это означает, что сравниваются значения функционала $I\{u(x,y)\}$ на поверхностях с уравнением вида $z=u(x,y)$, «натянутых» на один и тот же контур, заданный в (8.10). Поскольку, как обычно, мы рассматриваем всевозможные функции вида $u(x,y)+\delta u$, то $\{M(x,y)\in\partial G\}\Rightarrow\delta u(x,y)=0$. Вводя для $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$ традиционные обозначения соответственно p и q, запишем выражение для первой вариации этого функционала:

$$\delta I\{u,\delta u\}=\iint\limits_{(G)}(f_u'\delta u+f_p'\frac{\partial(\delta u)}{\partial x}+f_q'\frac{\partial(\delta u)}{\partial y})dxdy.$$

Интегрируя по частям в последнем выражении и учитывая равенство нулю δu на границе области G, получим

$$\delta I\{u,\delta u\}=\iint\limits_{(G)}\Big(f_u'-\frac{\partial f_p'}{\partial x}-\frac{\partial f_q'}{\partial y}\Big)\delta udxdy.$$

Согласно теореме 8.1 для экстремальной функции $u(x,y)$ при всех допустимых δu справедливо равенство $\delta I\{u, \delta u\} = 0$. Отсюда получаем[1], что

$$f'_u + \frac{\partial f'_p}{\partial x} + \frac{\partial f'_q}{\partial y} = 0. \tag{8.11}$$

Последнее уравнение называют уравнением Эйлера для функционала (8.9), (8.10). Оно было получено М.В. Остроградским в 1834 году и потому часто называется также уравнением Остроградского. Обратим внимание на смысл частных производных $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$ в уравнении (8.11). Например, производная берется при фиксированном y, но с учетом того, что в в качестве подставлены их выражения через x и y. Поэтому развернутая форма записи этого уравнения имеет вид:

$$f''_{pp}\frac{\partial^2 u}{\partial x^2} + 2f'_{pq}\frac{\partial^2 u}{\partial x \partial y} + f'_{qq}\frac{\partial^2 u}{\partial y^2} + f''_{up}\frac{\partial u}{\partial x} + f''_{uq}\frac{\partial u}{\partial y} + f''_{xp} + f''_{yq} - f'_u = 0.$$

Последнее уравнение является дифференциальным уравнением в частных производных второго порядка и, тем самым, отыскание функций, подозреваемых на экстремум функционала (8.9), (8.10), сводится к отысканию решений этого уравнения, удовлетворяющих граничным условиям (8.10). Такие задачи называются краевыми и точные методы их решения в настоящем пособии не рассматриваются.

Пример 8.15. Выписать уравнение Эйлера-Остроградского для функционала

$$I\{u(x,y)\} = \iint\limits_{(G)} \left(\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 \right) dxdy,$$

определенного на функциях $u(x,y)$, удовлетворяющих граничным условиям (8.10) с некоторой функцией $\varphi(M)$.

[1] С использованием аналога леммы 8.1, справедливым и в многомерном случае

Очевидно, что $f'_u = 0$, $f'_p = 2\dfrac{\delta u}{\delta x}$, $f'_q = 2\dfrac{\partial u}{\partial y}$. Поэтому уравнение Эйлера-Остроградского имеет следующий вид: $\dfrac{\delta^2 u}{\delta x^2} + \dfrac{\partial^2 u}{\partial y^2} = 0$. Тем самым наша вариационная задача сводится к решению краевой задачи Дирихле для уравнения Лапласа.

8.5. Задача с подвижными концами

Вернемся опять к функционалу вида (8.4). В простейшей задаче вариационного исчисления считается, что этот функционал определен для кривых $u = u(x)$ с фиксированными концами. Однако в приложениях часто возникают задачи, в которых граничные условия вообще отсутствуют, т. е. значения функционала сравниваются для всех функций $u(x) \in C_1[a,b]$. Пусть функция $u_0(x)$ реализует экстремальное значение функционала вида (8.4) среди всех функций $u(x) \in C_1[a,b]$. Тогда этот экстремум достигается и среди более узкого множества функций с граничными значениями

$$u(a) = u_0(a),\ u'(a) = u'_0(a),\ u(b) = u_0(b),\ u'(b) = u'_0(b).$$

Отсюда следует, что функция $u0(x)$ удовлетворяет уравнению Эйлера. При произвольной ее вариации для вариации функционала с помощью интегрирования по частям получим

$$\delta I\{u_0, \delta u\} = \int\limits_a^b (f'_u(x, u_0(x), u'_0(x))\delta u + f'_{u'}(x, u_0(x), u'_0(x))\delta u')dx =$$

$$f'_{u'}(x, u_0(x), u'_0(x))\delta u\Big|_a^b + \int\limits_a^b (f'_u(x, u_0(x), u'_0(x)) - \frac{d}{dx} f'_u(x, u_0(x), u'_0(x)))\delta u\, dx$$

Поскольку последний интеграл равен нулю, то согласно теореме 8.1 получим

$$f'_{u'}(b, u_0(b), u'_0(b))\delta u(b) - f'_{u'}(a, u_0(a), u'_0(a))\delta u(a) = 0.$$

Поскольку $\delta u(a)$ и $\delta u(b)$ произвольны, то из последнего равенства вытекает, что экстремаль $u_0(x)$ должна удовлетворять следующим граничным условиям

$$f'_{u'}(a, u_0(a), u'_0(a)) = 0,\ f'_{u'}(b, u_0(b), u'_0(b)) = 0\,, \tag{8.12}$$

которые называются *естественными граничными условиями.* При решении вариационной задачи на фиксированном отрезке $[a, b]$ без краевых условий из двухпараметрического семейства экстремалей с помощью естественных граничных условий выбирается функция, подозреваемая в том, что она доставляет экстремум функционалу. При этом, если краевое условие снимается только на одном конце отрезка, то и естественное условие ставится только на этом конце.

Пример 8.16. Найти экстремали функционала в следующей задаче:

$$I\{u(x)\} = \int\limits_{a}^{\pi/4} (u'^2 - u^2)dx, \ u(0) = 1$$

В этой задаче отсутствует граничное условие при $x = \pi/4$, следовательно, на правом конце нужно использовать естественное граничное условие (8.12), которое имеет вид $2u'(\pi/4) = 0$. Уравнение Эйлера имеет вид $u'' + u = 0$, а его общее решение $u(x) = C_1 \cos x + C_2 \sin x$. Из условия на левом конце $u(0) = C_1 = 1$. Естественное условие на правом конце дает: $u'(\pi/4) = -\sin(\pi/4) + C_2 \cos(\pi/4) = 0$. Откуда $C_2 = 1$. Окончательно получаем $u(x) = \cos x + \sin x$.

Отметим, что в простейшей задаче вариационного исчисления для функционала (8.4) иногда ставят граничные условия, отличающиеся от $u(a) = u_1, u(b) = u_2$, например, такие

$$f_1(u(a), u'(a)) = 0, \ f_2(u(b), u'(b)) = 0.$$

В принципе на решении задачи это не сказывается, т. к. функция доставляющая экстремум функционалу при любых краевых условиях обязана удовлетворять уравнению Эйлера.

Приступим к рассмотрению более общей задачи с *подвижными концами.* Эта задача состоит в определении функции $u_0(x) \in C_1[a, b]$ и точек x_0 и $x_1 \in [a, b]$, $(x_0 < x_1)$, для которых функционал

$$I\{u\} = \int\limits_{x_0}^{x_1} f(x, u, u')dx \tag{8.13}$$

достигает слабого экстремума при условиях

$$u(x_0) = \varphi_0(x_0), \ u(x_1) = \varphi_1(x_1). \tag{8.14}$$

Здесь $\varphi_0(x), \varphi_1(x) \in C_1[a, b], f(\cdot, \cdot, \cdot)$ – заданные функции и $f(\cdot, \cdot, \cdot)$ имеет непрерывные частные производные второго порядка по всем аргументам.

Суть этой задачи состоит в том, что на отрезке $[a, b]$ задаются две кривые $\gamma_1 : u = \varphi_0(x)$ и $\gamma_2 : u = \varphi_1(x)$. Требуется найти такую гладкую кривую, которая соединяет какую-либо точку кривой γ_1 с какой-либо точкой кривой γ_2 и доставляет при этом слабый экстремум функционалу (8.13).

Определение 8.4. Функция $u_0(x) \in C_1[a,b]$ и точки ${x^0}_0$ и ${x^0}_1 \in [a,b]$, (${x^0}_0 < {x^0}_1$) называются доставляющими слабый экстремум функционалу (8.13) при условиях (8.14), если существует число $\varepsilon > 0$ такое, что для всех функций $u(x) \in C_1[a,b]$ и точек x_0 и $x_1 \in [a,b]$, ($x_0 < x_1$), удовлетворяющих (8.14) и неравенствам $\|u - u_0\|_1 < \varepsilon$, $|{x^0}_0 - x_0| < \varepsilon$, $|{x^0}_1 - x_1| < \varepsilon$, справедливо $I\{u_0(x)\} \leq I\{u(x)\}$ (локальный минимум) или $I\{u_0(x)\} \geq I\{u(x)\}$ (локальный максимум).

Теорема 8.8. *Для того, чтобы функционал* (8.13) *достигал на функции* $u(x) \in C_1[a,b]$ *слабого экстремума при условиях* (8.14) *необходимо, чтобы эта функция удовлетворяла уравнению Эйлера*

$$f'_u(x, u(x), u'(x)) - \frac{d}{dx} f'_{u'}(x, u(x), u'(x)) = 0$$

и условиям трансверсальности

$$\begin{aligned} &[f(x, u(x), u'(x)) + (\varphi_0(x) - u'(x)) f'_{u'}(x, u(x), u'(x))]_{x=x_0} = 0, \\ &[f(x, u(x), u'(x)) + (\varphi_1(x) - u'(x)) f'_{u'}(x, u(x), u'(x))]_{x=x_1} = 0 \end{aligned} \qquad (8.15)$$

Пример 8.17. Найти экстремали функционала в следующей задаче с подвижными концами

$$I\{u\} = \int\limits_{x_0}^{x_1} \sqrt{1 + u'^2} dx, \;\; u(x_0) = {x^2}_0 + 2, u(x_1) = x_1.$$

Поскольку $f(x, u, u') = \sqrt{1 + u'^2}$ зависит лишь от u', то общее решение уравнения Эйлера имеет вид $u(x) = C_1 x + C_2$. Запишем условия трансверсальности

$$\left[\sqrt{1 + u'^2} + (2x - u'(x)) \frac{u'}{\sqrt{1 + u'^2}}\right]_{x=x_0} = 0$$

$$\left[\sqrt{1 + u'^2} + (1 - u'(x)) \frac{u'}{\sqrt{1 + u'^2}}\right]_{x=x_1} = 0$$

Поскольку $u'(x) = C_1$, то с учетом (8.14) получим следующую систему уравнений относительно C_1, C_2, x_0и x_1

$$\begin{cases} 1 + 2x_0 C_1 = 0, \\ 1 + C_1 = 0, \\ C_1 x_0 + C_2 = x_0^2 + 2, \\ C_1 x_1 + C_2 = x_1, \end{cases}$$

решив которую получим $C_1 = -1$, $C_2 = 11/4$, $x_0 = 1/2$, $x_1 = 11/8$. Следовательно, уравнение экстремали имеет вид $u(x) = -x + 11/4$, $x_0 = 1/2$, $x_1 = 11/8$.

Задача Больца. К задачам вариационного исчисления с подвижными концами относится и задача Больца, состоящая в определении функции $u_0(x) \in C_1[a, b]$, доставляющей слабый экстремум функционалу

$$I\{u\} = \int\limits_a^b f(x, u, u')dx + \varphi(u(a), u(b)), \tag{8.16}$$

где $\varphi(s, t)$ – заданная функция, имеющая непрерывные частные производные по s и t. Необходимые условия экстремума функционала (8.16) дает следующая

Теорема 8.9. *Для того, чтобы функция $u(x) \in C_1[a, b]$ доставляла слабый экстремум функционалу* (8.16) *необходимо, чтобы она удовлетворяла уравнению Эйлера*

$$f'_u(x, u(x), u'(x)) - \frac{d}{dx} f'_{u'}(x, u(x), u'(x)) = 0$$

и условиям трансверсальности для задачи Больца

$$\left[f'_{u'}(x, u(x), u'(x)) - \frac{\partial \varphi}{\partial u(a)} \right]_{x=a} = 0, \ \left[f'_{u'}(x, u(x), u'(x)) - \frac{\partial \varphi}{\partial u(b)} \right]_{x=b} = 0. \tag{8.17}$$

Условия (8.17) используются для определения постоянных C_1 и C_2 в общем решении $u(x, C_1, C_2)$ уравнения Эйлера.

Пример 8.18. Найти экстремали следующего функционала:

$$I\{u(x)\} = \int\limits_0^1 u'^2 dx + u^2(0) - 2u^2(1).$$

Уравнение Эйлера $u'' = 0$ имеет общее решение $u(x) = C_1 x + C_2$. Очевидно также, что $\varphi(u(0), u(1)) = u^2(0) - 2u^2(1)$. Поэтому условия (8.17) примут вид:

$$[2u'(x) - 2u(0)]_{x=0} = 0, \ [2u'(x) - 4u(1)]_{x=1} = 0.$$

Получаем следующую систему уравнений:

$$\begin{cases} C_1 - C_2 = 0, \\ C_1 + C_2 = 0, \end{cases}$$

из которой следует, что $C_1 = C2 = 0$. Единственной экстремалью задачи является функция $u(x) \equiv 0$.

8.6. Задачи на условный экстремум

Под задачами на условный экстремум вариационного исчисления понимают задачи, в которых на искомые функции накладываются, помимо граничных условий, некоторые дополнительные ограничения. Различают задачи на условный экстремум *с интегральными связями* а также задачи *с конечными или дифференциальными связями.* В обоих случаях используется так называемая *функция Лагранжа задачи.*

Рассмотрим задачу на экстремум функционала (8.8), зависящего от нескольких функций, с краевыми условиями

$$u_k(a) = u_{k1}, u_k(b) = u_{k2}, k = 1, \dots, n \tag{8.18}$$

при дополнительных ограничениях

$$\int\limits_a^b g_i(x, u_1, u_2, \dots, u_n, u_1', u_2', \dots, u_n')dx = C_i, i = 1, 2, \dots, m \tag{8.19}$$

где m–произвольное натуральное число. Эту задачу с интегральными условиями часто называют *изопериметрической.* Построим функцию Лагранжа задачи следующим образом

$$L(x, u_1, u_2, \dots, u_n, u_1', u_2', \dots, u_n') = f(x, u_1, u_2, \dots, u_n, u_1', u_2', \dots, u_n') +$$

$$\sum_{i=1}^{m} \lambda_i g_i(x, u_1, u_2, \dots, u_n, u_1', u_2', \dots, u_n'), \tag{8.20}$$

где произвольные числа $\lambda i (i = 1, 2, \dots, m)$ называют множителями Лагранжа. При решении изопериметрической задачи используется следующее необходимое условие

Теорема 8.10. *Если функции $u_1(x), u_2(x), \dots, u_n(x) \in C_1[a, b]$ доставляют слабый экстремум функционалу* (8.8) *при условиях* (8.18), (8.19) *то существуют числа $\lambda i (i = 1, 2, \dots, m)$ (множители Лагранжа), при которых эти функции удовлетворяют следующей системе уравнений*

$$L_{u_k}' - \frac{d}{dx} L_{u_k}' = 0, k = 1, 2, \dots, n \tag{8.21}$$

где L–функция Лагранжа (8.20).

При применении теоремы 8.10 искомые функции $u_1(x), u_2(x), \dots, u_n(x)$ и множители Лагранжа $\lambda i (i = 1, 2, \dots, m)$ определяются из системы $n+m$ уравнений (8.21), (8.19).

Пример 8.19. Найти функцию $u(x)$, на которой может достигаться экстремум функционала $I\{u(x)\}$ в следующей изопериметрической задаче:

$$I\{u(x)\} = \int\limits_0^1 u'^2 dx;$$

$$u(0) = 0, u(1) = 1; \int\limits_0^1 xudx = 0.$$

Функция Лагранжа имеет вид: $L(x, u, u') = u'^2 + \lambda xu$. (8.21) запишется в виде уравнения $2u'' + \lambda x = 0$, общее решение которого $u = -\frac{\lambda x^3}{12} + C_1 x + C_2$. Из краевых условий следует, что $C_2 = 0$, $C_1 = 1 + (\frac{\lambda}{12})$, поэтому $u = -\frac{\lambda x^3}{12} + (\frac{\lambda}{12})x$. В силу интегрального условия получим $\int\limits_0^1 xudx = \int\limits_0^1 ((1 + \frac{\lambda}{12})x^2 - \frac{\lambda}{12}x^4)dx = 0$. Отсюда получаем, что $\lambda = -30$. Таким образом, экстремум может достигаться только на функции $u = \frac{5x^3}{2} - \frac{3}{2}x$.

Рассмотрим теперь задачу на условный экстремум для функционала вида (8.8) с краевыми условиями (8.18) при дополнительных ограничениях, заданных уравнениями связи

$$\varphi_i(x, u_1, u_2, \ldots, u_n, u_1', u_2', \ldots, u_n') = 0, i = 1, 2, \ldots, m, \tag{8.22}$$

где натуральное число $m < n$. Отметим, что в случае, когда левые части (8.22) не зависят от производных неизвестных функций, равенства (8.22) называют системой конечных связей, а в противном случае – системой дифференциальных связей. Саму же задачу (8.8), (8.18), (8.22) обычно называют *задачей Лагранжа.* Здесь также строится функция Лагранжа задачи в следующем виде

$$\begin{aligned} \tilde{L}(x, u_1, u_2, \ldots, u_n, u_1', u_2', \ldots, u_n') = f(x, u_1, u_2, \ldots, u_n, u_1', u_2', \ldots, u_n') + \\ \textstyle\sum_{i=1}^m \lambda_i(x)\varphi_i(x, u_1, u_2, \ldots, u_n, u_1', u_2', \ldots, u_n'), \end{aligned} \tag{8.23}$$

где $\lambda i(x) \in C_1[a, b]$ – произвольные функции *(множители Лагранжа).* При решении задачи Лагранжа используются следующие необходимые условия экстремума.

Теорема 8.11. *Если функции* $u_1(x), u_2(x), \ldots, u_n(x) \in C_1[a, b]$ *доставляют слабый экстремум функционалу* (8.8) *при условиях* (8.18), (8.22) *то существуют множители Лагранжа* $\lambda i(x) \in C_1[a, b](i = 1, 2, \ldots, m)$, *при которых эти функции удовлетворяют следующей системе уравнений*

$$\tilde{L}'_{uk} - \frac{d}{dx}\tilde{L}'_{uk} = 0, k = 1, 2, \ldots, n, \tag{8.24}$$

где $\tilde{L}$ *– функция Лагранжа* (8.23).

Так же, как и в задачах с интегральными условиями при решении задачи Лагранжа искомые функции $u_1(x), u_2(x), \ldots, u_n(x)$ и множители

Лагранжа $\lambda i(x)(i = 1, 2, \dots, m)$ определяются из системы $n+m$ уравнений (8.24), (8.22).

Пример 8.20. Найти функции $u_1(x), u_2(x)$, на которых может достигаться экстремум функционала в следующей задаче Лагранжа

$$I\{u_1, u_2\} = \int_0^1 (u_1'^2 + u_2'^2)dx;$$

$$u_1(0) = u_2(0) = 0, u_1(1) = 2\operatorname{ch} 1 - 2, u_2(1) = 2\operatorname{sh} 1;$$

$$u_1'(x) - u_2(x) = 0.$$

Запишем функцию Лагранжа данной задачи $\tilde{L} = u_1'^2 + u_2'^2 + \lambda(u_1'^2 - u_2)$. Для определения $u_1(x), u_2(x)$ и $\lambda(x)$ запишем систему равенств (8.24), (8.22):

$$\begin{cases} 2u_1'' + \lambda' = 0, \\ 2u_2' + \lambda = 0, \\ u_1' - u_2 = 0. \end{cases}$$

Из второго и третьего уравнений получаем: $\lambda = -2u_2'', u_2 = u_1'$. Производя подстановку в первое равенство, получим уравнение $u_1^{(4)} - u_1' = 0$. Его общее решение $u_1 = C_1e^x + C_2e^{-x} + C_3x + C_4$. Кроме того, $u_2 = u_1' = C_1e^x - C_2e^{-x} + C_3$. Для нахождения произвольных констант используем краевые условия. Получим следующую систему уравнений:

$$u_1(0) = C_1 + C_2 + C_4 = 0,$$

$$u_2(0) = C_1 - C_2 + C_3 = 0,$$

$$u_1(1) = C_1e + C_2e^{-1} + C_3 + C_4 = 2\operatorname{ch} 1 - 2,$$

$$u_2(1) = C_1e - C_2e^{-1} + C_3 = 2\operatorname{sh} 1.$$

Решая эту систему, получим: $C_1 = C_2 = 1, C_3 = 0, C_3 = \breve{}2$. Таким образом, в данной задаче функционал $I\{u_1, u_2\}$ может достигнуть экстремума лишь при $u_1(x) = 2\operatorname{ch} x - 2, u_2(x) = 2\operatorname{sh} x$.

8.7. Прямые методы вариационного исчисления

Рассмотренный выше метод решения вариационных задач, основанный на сведении их к уравнениям Эйлера, часто оказывается недостаточно эффективным, несмотря на его большое теоретическое значение. Рассмотренные выше примеры показывают, что в ряде случаев уравнения Эйлера удается решить в квадратурах и даже в элементарных функциях. Однако для более сложных задач, особенно для функционалов, определенных на функциях от нескольких переменных, такое решение найти не удается. Появляется необходимость численного решения линейной или нелинейной краевой задачи для дифференциального уравнения, что является очень непростой задачей и в настоящем пособии не рассматривается. Сравнительно давно были разработаны методы, позволяющие непосредственно (возможно, без перехода к уравнениям Эйлера) найти приближенное решение вариационной задачи. Эти методы называют *прямыми методами вариационного исчисления.* Ниже мы рассмотрим несколько таких методов, не останавливаясь подробно на их обосновании. Подробное теоретическое исследование этих методов и большое число примеров содержится в гл. IV книги [23], а также в книгах [29] – [31].

В математическом анализе имеется два основных подхода к решению задачи приближенного отыскания неизвестной функции. При первом подходе функция с самого начала заменяется набором значений в узлах некоторой сетки. Здесь типичными являются сеточные методы интегрирования дифференциальных уравнений. По сути, такой подход состоит в дискретизации аргументов функций и приближенной замене искомой функции соответствующей сеточной функцией. Несмотря на универсальность этого подхода, методы, разработанные в его рамках, требуют довольно большого объема вычислений, а результаты этих вычислений, представляющие собой числовые массивы большой размерности, часто неудобно использовать.

В математическом анализе имеется два основных подхода к решению задачи приближенного отыскания неизвестной функции. При первом подходе функция с самого начала заменяется набором значений в узлах некоторой сетки. Здесь типичными являются сеточные методы интегрирования дифференциальных уравнений. По сути, такой подход состоит в *дискретизации аргументов* функций и приближенной замене искомой функции соответствующей сеточной функцией. Несмотря на универсальность этого подхода, методы, разработанные в его рамках, требуют довольно большого объема вычислений, а результаты этих вычислений, представляющие собой числовые массивы большой размерности, часто неудобно использовать.

Второй подход объединяет так называемые *методы сужения числа степеней свободы.* В этих методах независимые переменные остаются непрерывными, но неизвестная функция ищется в некотором специальном виде, содержащем несколько параметров, которые потом подбираются из требования наилучшего удовлетворения условиям задачи. К этим методам и относятся прямые методы вариационного исчисления. Наиболее известными из них являются методы *Ритца, Канторовича и Галеркина.*

Метод Ритца предложен в 1908 году немецким физиком и математиком В. Ритцем и для простейшей задачи вариационного исчисления

$$I\{u\} = \int\limits_a^b f(x,u,u')dx, u(a) = c_1, u(b) = c_2 \qquad (8.25)$$

он состоит в том, что функционал (8.25) рассматривается не на произвольных допустимых кривых вариационной задачи, а на всевозможных линейных комбинациях вида

$$u_n(x) = \varphi_0(x) + \sum_{i=1}^n C_i\varphi_i(x), \qquad (8.26)$$

где $\varphi_0(a) = c_1, \varphi_0(b) = c_2$, а $\varphi_i(x)$ – элементы бесконечной последовательности функций, любая конечная часть которой линейно независима, причем $\varphi_i(a) = \varphi_i(b) = 0, i = 1, 2, \ldots$. Функции $\varphi_i(x)$ $(i = 1, 2, \ldots)$ называются *координатными.* При подстановке выражения (8.26) вместо u в (8.25) функционал превращается в функцию от n переменных: $I\{u_n\} = F(C_1, C_2, \ldots, C_n)$. Значения $C_1, C_2, \ldots, C_n$ выбираются так, чтобы функция $F(C_1, C_2, \ldots, C_n)$ достигала экстремума. Для этого решается система уравнений $\frac{\partial\Phi}{\partial C_i} = 0, i = 1, 2, \ldots, n$. По найденным из этой системы значениям $C_1, C_2, \ldots, C_n$ получаем приближенное решение вариационной задачи в виде (8.26).

Вопрос о сходимости минимизирующей последовательности $\{u_n(x), n = 1, 2, \ldots\}$ к искомой экстремали является довольно сложным и изучается в специальной литературе. Для оценки точности результатов, полученных методом Ритца или другими прямыми методами, обычно пользуются следующим практическим правилом. Вычисляют $u_n(x)$ и $u_{n+1}(x)$ сравнивают значения этих функций в нескольких точках отрезка $[a, b]$. Если в пределах требуемой точности эти значения совпадают, то считают, что в пределах требуемой точности решением вариационной задачи является $u_n(x)$. Если же такого совпадения нет, то находят $u_{n+2}(x)$ и сравнивают в этих же точках $u_{n+1}(x)$ и $u_{n+2}(x)$. Этот процесс продолжается до тех пор, пока значения $u_{n+k}(x)$ и $u_{n+k+1}(x)$ в выбранных точках не совпадут в пределах заданной точности.

Сравнительно простой является реализация метода Ритца для квадратичного функционала

$$I\{u\} = \int\limits_a^b (p(x)u'^2 + q(x)u^2 + r(x)u)dx, u(a) = c_1, u(b) = c_2. \qquad (8.27)$$

При подстановке выражения (8.26) в функционал (8.27) получится квадратичная функция $I\{u_n\} = F(C_1, C_2, \ldots, C_n)$. Поэтому система уравнений

$$\frac{\partial\Phi}{\partial C_i} = 0, i = 1, 2, \ldots, n$$

будет системой n линейных уравнений с n неизвестными, которая при надлежащем выборе координатных функций имеет единственное решение.

Последовательность координатных функций

$$\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x), \varphi_{n+1}(x), \dots$$

кроме сформулированных выше требований, должна быть *полной в пространстве функций* $f \in C_1[a,b]$, удовлетворяющих условиям $f(a) = f(b) = 0$.Полнота этой последовательности означает, что любую такую функцию $f(x)$ можно приблизить (по норме пространства $C_1[a,b]$) с любой степенью точности линейной комбинацией координатных функций. Чаще всего в качестве координатных функций берутся функции вида:

$$\varphi_k(x) = x^{k-1}(x-a)(x-b) \;\; (k = 1, 2, 3, \dots). \tag{8.28}$$

Иногда более удобной является последовательность функций вида:

$$\varphi_k(x) = \sin\left(k\pi(k-1)/(x-a)(x-b)\right) \;\; (k = 1, 2, 3, \dots). \tag{8.29}$$

Можно проверить, что каждая из этих последовательностей обладает свойством полноты, а любая часть каждой из них линейно независима. В качестве функции $\varphi_0(x)$ обычно выбирают

$$\varphi_0(x) = ((b-x)c_1 + (x-a)c_2)/(b-a). \tag{8.30}$$

Можно доказать, что если в функционале (8.27)$p(x) > 0$ и исходная задача имеет единственное решение, то приближенные решения $u_n(x)$, построенные по методу Ритца, при $n \to \infty$ сходятся в $C_1[a,b]$ к точному решению. Условие о единственности решения исходной задачи выполнено, например, при $q(x) \geq 0$. Однако в конкретных приложениях использовать теоремы о сходимости затруднительно. Для оценки точности результата приходится пользоваться приведенным выше практическим правилом.

Пример 8.21. С помощью метода Ритца найти приближенно экстремаль функционала

$$I\{u\} = \int\limits_0^1 (u'^2 + u^2 + 2ux)dx \tag{8.31}$$

при граничных условиях $u(0) = u(1) = 0$.

Согласно (8.30) $\varphi_0(x) \equiv 0\,(c_1 = c_2 = 0)$. Выберем систему координатных функций вида (8.28): $\varphi_k(x) = (x) = x^{k-1}x(x-1) = x^{k+1} - x^k$ $(k = 1, 2, 3, \dots)$. Возьмем $n = 2$. Тогда в соответствии с (8.26) получим

$$u_2(x) = C_1(x^2 - x) + C_2(x^3 - x^2),$$

$$u_2'(x) = C_1(2x-1) + C_2(3x^2 - 2x).$$

После подстановки в (8.31) $u_2(x)$ вместо $u(x)$ и интегрирования получим следующую квадратичную функцию от C_1, C_2:

$$\Phi(C_1, C_2) = I\{u\} = \frac{11}{30}C_1^2 + \frac{11}{30}C_1C_2 + \frac{1}{7}C_2^2 - \frac{1}{6}C_1 - \frac{1}{10}C_2.$$

Вследствие необходимых условий экстремума для функции $\Phi(C1, C2)$, находим

$$\begin{cases} \dfrac{\partial\Phi}{\partial C_1} = \dfrac{11}{15}C_1 + \dfrac{11}{30}C_2 - \dfrac{1}{6} = 0, \\ \dfrac{\partial\Phi}{\partial C_2} = \dfrac{11}{30}C_1 + \dfrac{2}{7}C_2 - \dfrac{1}{10} = 0. \end{cases}$$

Решая эту систему, получим $C_1 = 69/473$, $C_2 = 7/43$. Поэтому приближенное выражение для экстремали имеет вид

$$u_2(x) = \frac{69}{473}(x^2 - x) + \frac{7}{43}(x^3 - x^2) = \frac{1}{473}(77x^3 - 8x^2 - 69x).$$

В данном случае можно найти точное решение поставленной задачи:

$$u(x) = \frac{e}{e^2 - 1}(e^x - e^{-x}) - x.$$

Сравним приближенное решение $u_2(x)$ и точное $u(x)$ при некоторых значениях аргумента:

x	0,0	0,2	0,4	0,6	0,8	1,0
u(x)	0,0000	-0,0287	-0,0505	-0,0583	-0,0444	0,0000
u_2(x)	0,0000	-0,0285	-0,0506	-0,0585	-0,0442	0,0000

Сравнение показывает, точное и приближенное решение совпадают с точностью до 0,0002.

Отметим, что для функционалов содержащих производные высших порядков, для функционалов от нескольких функций и для задач на условный экстремум применение метода Ритца в принципе не отличается от рассмотренного выше и, в целом, является эффективным. Для функционалов, которые не являются квадратичными, для поиска экстремума функции $\Phi(C_1, C_2, \ldots, C_n)$ могут оказаться полезными градиентные методы, изложенные в параграфе 6.6. Трудности в применении метода Ритца возникают в задачах для функционалов на функциях от нескольких переменных.

Метод Канторовича предложен в 1933 году и является развитием метода Ритца для функционалов от функций нескольких переменных. Этот метод подробно изложен в монографии Л. В. Канторовича и В. И. Крылова [23]. Здесь мы ограничимся лишь изложением идеи метода.

Рассмотрим задачу на экстремум функционала $I\{u(x, y)\}$ (8.9), где область интегрирования G ограничена прямыми $x = a$ и $x = b$, а также графиками функций $y = \psi_1(x)$ и $y = \psi_2(x)$ ($\psi_1(x) \leq \psi_2(x)$ при

$x \in (a,b)$).Предположим, что искомая функция $u(x,y)$ удовлетворяет нулевому краевому условию $u(x,y)|_{\partial G} = 0$. В методе Канторовича приближенное выражение для экстремали ищется в виде

$$u_n(x,y) = \sum_{k=1}^{n} u_k(x)\varphi_k(x,y), \tag{8.32}$$

где известные координатные функции $\varphi_k(x,y)$ $(k = 1,2,3,\dots)$ считаются равными нулю на ∂G, кроме, быть может, прямых $x = a$ и $x = b$. Неизвестные функции одной переменной $u_k(x)$ определяются из требования, чтобы функционал $I\{u_n(x,y)\}$ достигал экстремального значения. Отметим, что функции $u_k(x)$ должны удовлетворять граничным условиям

$$u_k(a) = u_k(b) = 0,\ k = 1,2,3,\dots,n, \tag{8.33}$$

вытекающим из краевых условий задачи.

Записывая двойной интеграл в функционале (8.9) в виде повторного, получим:

$$I\{u(x,y)\} = \int_a^b \Big(\int_{\psi_1(x)}^{\psi_2(x)} f(x,y,u,\frac{\partial u}{\partial x},\frac{\partial u}{\partial y})\,dy\Big)dx.$$

Подставляя в последний интеграл вместо $u(x,y)$ выражение (8.32) и интегрируя по y во внутреннем интеграле, получим следующий функционал:

$$I\{u_n(x,y)\} = \int_a^b \Phi(x,u_1,u_2,K,u_n,u_1',u_2',K,u_n')dx.$$

Функции $u_k(x)$, доставляющие экстремум этому функционалу, должны удовлетворять системе уравнений Эйлера-Лагранжа

$$\Phi'_{u_k} - \frac{d}{dx}\Phi'_{u_k'} = 0,\ k = 1,2,K,n, \tag{8.34}$$

а также граничным условиям (8.33).

Таким образом, суть метода Канторовича состоит в том, что он приближенно сводит интегрирование уравнения в частных производных (8.11) к интегрированию системы обыкновенных дифференциальных уравнений (8.34).

Пример 8.22. Найти методом Канторовича экстремаль функционала

$$I\{u(x,y)\} = \iint_{(G)} \left(\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 - 2u\right)dxdy,$$

где $G = \{(x,y) :\ -a \le x \le a,\ -b \le y \le b\}$, $u(x,y)|_{\partial G} = 0$, $n = 1$.

Решение будем искать в виде

$$u_1(x,y) = u_1(x)(b^2 - y^2).$$

Граничные условия на прямых $y = \pm b$ выполняются. После подстановки $u_1(x,y)$ в функционал и интегрирования по y получаем функционал:

$$I\{u_1(x)\} = \int\limits_{-a}^{a} (\frac{16}{15} b^2 u_1'^2 + \frac{8}{3} b^3 u_1^2 - \frac{8}{3} b^3 u_1) dx.$$

Уравнение Эйлера для этого функционала имеет вид

$$u_1''(x) = -\frac{5}{2b^2} u_1(x) = -\frac{5}{4b^2}.$$

Находим общее решение этого уравнения:

$$u_1(x) = C_1 \operatorname{ch} \sqrt{2,5} \frac{x}{b} + C_2 \operatorname{sh} \sqrt{2,5} \frac{x}{b} + 0,5.$$

Постоянные C_1 и C_2 определяем из граничных условий $u_1(-a) = u_1(a) = 0$, из которых следует, что $C_1 = 0$, а $C_2 = -\left(2 \operatorname{ch}\left((2,5)^{1/2}(a/b)\right)\right)^{-1}$. Таким образом, приближенное решение имеет вид

$$u_1(x,y) = \frac{1}{2}(b^2 - y^2) \left(1 - \frac{ch\sqrt{2,5}\dfrac{x}{b}}{ch\sqrt{2,5}\dfrac{x}{b}} \right).$$

Метод Галеркина применяется при отыскании приближенных решений краевых задач для линейных обыкновенных дифференциальных уравнений и уравнений в частных производных. Эти краевые задачи могут возникать при решении задач вариационного исчисления. Поэтому имеет смысл этот метод называть приближенным методом вариационного исчисления. Отметим сразу, что метод Галеркина является типичным представителем методов сужения числа степеней свободы и не использует сеточные методы решения краевых задач, хотя и основан на переходе от вариационной задачи к уравнениям Эйлера-Лагранжа.

Пусть неизвестная функция $u(\overset{\rho}{x})$ в некоторой области D является решением следующей краевой задачи:

$$L(u(\overset{\rho}{x})) = f(\overset{\rho}{x}),\ \overset{\rho}{x} \in D \subset R^n,$$

$$\Gamma((u(\overset{\rho}{x})) = 0,\ \overset{\rho}{x} \in \partial D,$$

где L–некоторое линейное дифференциальное выражение, Γ–линейный оператор граничных условий. Приближенное решение этой краевой задачи ищется в виде суммы

$$u_n(\overset{\rho}{x}) = \sum_{k=1}^{n} c_k \varphi_k(\overset{\rho}{x}),$$

где c_k–неопределенные коэффициенты, а $\varphi_k(\overset{\rho}{x})$ – элементы бесконечной последовательности функций, любая конечная часть которой линейно независима, причем $\Gamma(\varphi_k(\overset{\rho}{x})) = 0$, $\overset{\rho}{x} \in \partial D$. Коэффициенты ck определяются из следующих условий ортогональности в области D невязки $L(u_n(\overset{\rho}{x}))-f(\overset{\rho}{x})$ в дифференциальном уравнении координатным функциям $\varphi_k(\overset{\rho}{x})$:

$$\int\limits_D [L(u_n(\overset{\rho}{x})) - f(\overset{\rho}{x})]\varphi_k(\overset{\rho}{x})dx = 0, \; k = 1, 2, K, n.$$

Если учесть линейность выражения L, то получим следующую систему для определения неопределенных коэффициентов c_i:

$$\sum_{i=1}^{n} c_i \int\limits_D f_i(\overset{\rho}{x})\varphi_k(\overset{\rho}{x})dx = \int\limits_D f(\overset{\rho}{x})\varphi_k(\overset{\rho}{x})dx, \; k = 1, 2, K, n,$$

где $f_i(\overset{\rho}{x}) \equiv L(\varphi_i(\overset{\rho}{x}))$. Отметим, что условие ортогональности невязки координатным функциям $\varphi_k(\overset{\rho}{x})$ можно заменить ее ортогональностью к функциям $f_k(\overset{\rho}{x}) \equiv L(\varphi_k(\overset{\rho}{x}))$, $k = 1, 2, \ldots, n$. Более того, при нахождении неопределенных коэффициентов c_k можно использовать ортогональность невязки к линейно независимой системе функций $\psi_1(\overset{\rho}{x}), \psi_2(\overset{\rho}{x}), \ldots, \psi_n(\overset{\rho}{x})$, линейная оболочка которой содержит функции $f_k(\overset{\rho}{x})(k = 1, 2, \ldots, n)$. Функции $\psi_k(\overset{\rho}{x})$ выбираются из соображений удобства вычислений соответствующих интегралов (см. решение нижеследующего примера 8.23).

Пример 8.23. Найти приближенно с помощью метода Галеркина экстремаль функционала

$$I\{u\} = \int\limits_0^1 (u'^2 - 2xu)dx$$

при граничных условиях $u(0) = u(1) = 0$. Составляя уравнение Эйлера, получим $u'' = -x$. Таким образом, $L_u = u''$, а $f(\overset{\rho}{x}) = -x$ и мы должны решать краевую задачу для уравнения $u'' = -x$ с граничными условиями $u(0) = u(1) = 0$. Выберем базисные функции в виде (8.28) с $a = 0$, $b = 1$, $k = 1, 2$, т. е. $\varphi_1(x) = x(x-1)$, $\varphi_2(x) = x^2(x-1)$. Приближенное решение будем искать в виде

$$u(x) = c_1 x(x-1) + c_2 x^2(x-1).$$

Поскольку $L(\varphi_1(x)) = 2$, а $L(\varphi_2(x)) = 6x - 2$, то невязку можно записать в виде

$$Lu - f(x) = u'' + x = 2c_1 + c_2(6x-2) + x.$$

Выберем функции $\psi_1(x) = 1$ и $\psi_2(x) = x$, линейная оболочка которых содержит $L(\varphi_1(x))$ и $L(\varphi_2(x))$. Запишем условия ортогональности невязки этим функциям. После интегрирования получим

$$\mathbf{2}c_1 + c_2 = -1/2,$$

$$c_1 + c_2 = -1/3.$$

Решая эту систему относительно c_1 и c_2, получаем $c_1 = -1/6$, $c_2 = -1/6$. Таким образом, мы имеем следующее приближение для экстремали:

$$u^*(x) = (-1/6)x(x-1) = (-1/6)x^2(x-1) = x(1-x^2)/6.$$

Отметим, что в данной задаче полученное приближенное решение совпадает с точным решением $u_{\text{т}}(x) = x(1-x^2)/6$.

Контрольные вопросы и задачи для самостоятельного решения

1. Примеры оптимизационных задач с бесконечным числом степеней свободы.
2. Понятие нормированного векторного пространства. Абстрактное определение функционала в нормированном векторном пространстве. Определение локального экстремума функционала.
3. Понятие линейного функционала. Вариация, вообще говоря, нелинейного функционала, необходимые условия локального экстремума. Стационарные элементы функционала.
4. Вариации функционалов второго и высших порядков. Достаточные условия экстремума. Необходимые условия, выраженные через вторую вариацию функционала.
5. Понятие глобального (абсолютного) экстремума функционала. Одноэкстремальность оптимизационной задачи. Выпуклые функционалы и одноэкстремальность соответствующих оптимизационных задач.
6. Простейшая задача вариационного исчисления. Уравнение Эйлера. Понятие экстремали.
7. Случаи простейших задач вариационного исчисления, для которых уравнение Эйлера упрощается по сравнению с общим случаем.
8. Условия экстремума простейшей задачи вариационного исчисления, использующие вторую вариацию.
9. Задача на экстремум функционала, зависящего от производных высших порядков неизвестной функции. Уравнение Эйлера-Пуассона.
10. Вариационная задача для функционала, зависящего от нескольких неизвестных функций. Система уравнений Эйлера.

11. Задача на экстремум функционала, зависящего от неизвестной функции нескольких аргументов. Уравнение Эйлера-Остроградского и краевая задача для него.
12. Задача с подвижными концами. Условия трансверсальности. Задача Больца.
13. Задача на условный экстремум с интегральными связями. Метод множителей Лагранжа.
14. Задача на условный экстремум с конечными или дифференциальными связями.
15. Понятие о прямых методах вариационного исчисления. Метод Ритца.
16. Метод Канторовича.
17. Метод Галеркина.

В задачах 8.1– 8.10 *найти экстремали функционала* $I\{u\}$, *удовлетворяющие указанным краевым условиям:*

8.1. $I\{u\} = \int\limits_0^1 (u'^2 + xu)dx;\;\; u(0) = u(1) = 0.$

Отв. $u(x) = (x^3 - x)/12.$

8.2. $I\{u\} = \int\limits_0^1 (u'^2 + uu' + 12xu)dx;\;\; u(0) = u(1) = 0.$

Отв. $u(x) = (x^3 - x).$

8.3. $I\{u\} = \int\limits_0^1 (u'^2 + u^2 + xu)dx;\;\; u(0) = u(1) = 0.$

Отв. $u(x) = \dfrac{\operatorname{sh} x}{\operatorname{sh} 1} - \dfrac{x}{2}.$

8.4. $I\{u\} = \int\limits_0^{\ln 2} (u'^2 + 3u^2)e^{2x}\,dx;\;\; u(0) = 0,\, u(\ln 2) = 15/8.$

Отв. $u(x) = e^x - e^{-3x}.$

8.5. $I\{u\} = \int\limits_0^1 (u'^2 + u^2 + 2ue^x)dx;\;\; u(0) = 0,\, u(1) = 1/(2e).$

Отв. $u(x) = 1/(2xe^x) - \operatorname{sh} x.$

8.6. $I\{u\} = \int\limits_0^1 u'^2\,dx;\;\; u(0) = 0,\, u(1) = 1.$

Отв. $u(x) = x.$

8.7. $I\{u\} = \int\limits_0^{3/2} (u'^3 + 2x)dx;\;\; u(0),\, u(3/2) = 1.$

Отв. $u(x) = (2x/3)^{3/2}.$

8.8. $I\{u\} = \int\limits_{-1}^1 (u'^2 + xu')dx;\;\; u(-1) = 1,\, u(1) = 0.$

Отв. $u(x) = (3/4) - (x^2 + 2x)/4.$

8.9. $I\{u\} = \int\limits_0^1 uu'\,dx;\;\; u(0) = 1,\, u(1) = \sqrt[3]{4}.$

Отв. $u(x) = (x+1)^{2/3}$ *или* $u(x) = (3x-1)^{2/3}.$

8.10. $I\{u\} = \int\limits_0^{\pi/2} (-u'^2 + u^2 + 2u)dx;\;\; u(0) = u(\pi/2) = 0.$

Отв. $u(x) = \cos x + \sin x - 1.$

В задачах 8.11– 8.15 *для функционала* $I\{u\}$*, зависящего от производных высших порядков от неизвестной функции, найти экстремали, удовлетворяющие указанным краевым условиям:*

8.11. $I\{u\} = \int\limits_0^1 u''\,dx;\;\; u(0) = u(1) = u'(1)0.$

Отв. $u(x) = x(x-1)^2.$

8.12. $I\{u\} = \int\limits_0^1 (u''^2 - 24xu)dx;\;\; u(0) = u'(0) = 0,\, u(1) = 1/5,\, u'(1) = 1.$

Отв. $u(x) = (x^5 + 3x^3 - 2x^2)/10.$

8.13. $I\{u\} = \int\limits_0^1 (-u''^2 + 48u)dx;\;\; u(0) = u'(0) = 0,\, u(1) = 1,\, u'(1) = 4.$

Отв. $u(x) = x^4.$

8.14. $I\{u\} = \int\limits_0^b (u''^2 + u'^2)dx;\;\; u(0) = u'(0) = u(b) = u'(b) = 0.$

Отв. $u(x) \equiv 0.$

8.15. $I\{u\} = \int\limits_0^\pi (-u'''^2 + u''^2)dx;\;\; u(0) = u'(0) = u''(0) = 0,\, u(\pi) = \pi,$ $u'(\pi) = 2,\, u''(\pi) = 0.$ *Отв.* $u(x) = x - \sin x.$

В задачах 8.16– 8.20 *для функционала* Iu_1, u_2*, зависящего от двух неизвестных функций, найти функции* $u_1(x)$*,* $u_2(x) \in C_1[a, b]$*, на которых может достигаться экстремум:*

8.16. $I\{u_1, u_2\} = \int\limits_0^{\frac{\pi}{2}} (u_1'^2 + u_2'^2 + 2u_1u_2)dx;$ $u_1(0) = u_2(0) = 0,$ $u_1(\frac{\pi}{2}) = 1, u_2(\frac{\pi}{2}) = -1.$ *Отв.* $u_1(x) = \sin x, u_2(x) = -\sin x.$

8.17. $I\{u_1, u_2\} = \int\limits_0^1 (u_1'u_2' + u_1u_2)dx; \quad u_1(0) = u_2(0) = 1, \quad u_1(1) = e,$ $u_2(1) = \frac{1}{e}$. *Отв.* $u_1(x) = e^x$, $u_2(x) = e^{-x}$.

8.18. $I\{u_1, u_2\} = \int\limits_0^1 (u_1'u_2' - u_1u_2)dx; \quad u_1(0) = u_2(0) = 0, \quad u_1(\frac{\pi}{2}) = 1,$ $u_2(\frac{\pi}{2}) = -1$. *Отв.* $u_1(x) = \sin x, u_2(x) = -\sin x$.

8.19. $I\{u_1, u_2\} = \int\limits_0^1 (u_1'^2 + u_2'^2 - 2u_1u_2)dx; \quad u_1(0) = u_2(0) = 0,$ $u_1(1) = \operatorname{sh} 1, u_2(1) = -\operatorname{sh} 1$. *Отв.* $u_1(x) = \operatorname{sh} x, u_2(x) = -\operatorname{sh} x$.

8.20. $I\{u_1, u_2\} = \int\limits_0^1 (u_1'^2 + u_2'^2 + 2u_1)dx; \quad u_1(0) = u_2(0) = 1, \quad u_1(1) = \frac{3}{2},$ $u_2(1) = 1$. *Отв.* $u_1(x) = (\frac{x^2}{2}) + 1, u_2(x) \equiv 1$

Найти экстремали функционала Iu в задачах 8.21–8.26 с подвижными концами:

8.21. $I\{u\} = \int\limits_0^{x_1} u'^2 dx; \quad u(x_1) = -x_1 - 1$. *Отв.* $u(x) = -2x, x_1 = 1$.

8.22. $I\{u\} = \int\limits_0^{x_1} (u'^2 + u^2)dx; \quad u(0) = 0, u(x_1) = 1$. *Отв. Нет решений.*

8.23. $I\{u\} = \int\limits_0^{x_1} u'^2 dx; \quad u(0) = 0, u(x_1) = \frac{2}{-x_1+1}$. *Отв.* $u(x) = \pm 4x$, $x_1 = \frac{1}{2}$.

8.24. $I\{u\} = \int\limits_0^1 (u'^2 + u)dx; \quad u(1) = 0$. *Отв.* $u(x) = \frac{x^2-1}{4}$.

8.25. $I\{u\} = \int\limits_0^{\frac{\pi}{2}} (u'^2 - u^2)dx; \quad u(0) = 0$. *Отв.* $u(x) = C\sin x, C \in R$.

8.26. $I\{u\} = \int\limits_0^{x_1} \frac{\sqrt{1+u'^2}}{u} dx; \quad u(0) = 1, u(x_1) = x1 - 1$. *Отв.* $u(x) = \sqrt{2 - (x-1)^2}, x_1 = 2$.

В задачах 8.27–8.30 найти экстремали следующих функционалов:

8.27. $I\{u\} = \int\limits_0^1 (u'^2 + u)dx - 2\operatorname{sh}1u(1)$. *Отв.* $u(x) = \operatorname{sh} x$.

8.28. $I\{u\} = \int\limits_0^1 u'^2 dx + u^2(0) - 2u^2(1)$. *Отв.* $u(x) \equiv 0$

8.29. $I\{u\} = \int\limits_0^3 4u^2 u'^2 dx + u^4(0) - 8u(3)$.

Отв. $u(x) = (x+1)^{\frac{1}{2}}$ *или* $u(x) = (\frac{4x^2}{3})^{\frac{1}{6}}$.

8.30. $I\{u\} = \int\limits_0^1 e^u u'^2 dx + 4e^{u(0)} + 32e^{-u(1)}$. *Отв.* $u(x) = 2\ln(x+1)$.

В изопериметрических задачах 8.31–8.35 найти функцию $u(x)$, для которой может достигаться экстремум функционала $I\{u\}$:

8.31. $I\{u\} = \int\limits_0^1 u'^2 dx$;

$u(0) = 0, u(1) = 1; \int\limits_0^1 xudx = 0$. *Отв.* $u(x) = \frac{5x^3}{2} - \frac{3x}{2}$.

8.32. $I\{u\} = \int\limits_0^1 u'^2 dx$;

$u(0) = u(1) = 0; \int\limits_0^1 xudx = 0, \int\limits_0^1 udx = 1$. *Отв.* $u(x) = 60x^3 - 96x^2 + 36x$.

8.33. $I\{u\} = \int\limits_0^\pi u'^2 dx$;

$u(0) = 0, u(\pi) = 1; \int\limits_0^\pi u \sin x dx = 0$. *Отв.* $u(x) = \frac{x - 2\sin x}{\pi}$.

8.34. $I\{u\} = \int\limits_0^\pi u \sin x dx$;

$u(0) = 0, u(\pi) = \pi; \int\limits_0^\pi u'^2 dx = \frac{3}{2}\pi$. *Отв.* $u(x) = (x + \sin x)$, *или* $u(x) = (x - \sin x)$.

8.35. $I\{u\} = \int\limits_0^1 (u'^2 + u^2)dx$;

$u(0) = 0, u(1) = e^{-1}; \int\limits_0^1 e^{-x} udx = \frac{1}{4}(1 - 3e^{-2})$. *Отв.* $u(x) = xe^{-x}$.

В задачах Лагранжа 8.36–8.40 найти функции $u_1(x), u_2(x)$, на которых может достигаться экстремум функционала $I\{u_1, u_2\}$:

8.36. $I\{u_1,u_2\}=\int\limits_0^1(u_1'^2+2u_2'^2+u_2^2)dx;$
$u_1(0)=-2, u_2(0)=1, u_1(1)=-e^{-1}, u_2(1)=0; u_1(x)-u_2'^2(x)=0.$
Отв. $u_1(x)=(x-2)e^{-x}, u_2(x)(1-x)e^{-x}.$

8.37. $I\{u_1,u_2\}=\int\limits_0^1(u_1'^2+u_2^2+x^3)dx;$
$u_1(0)=2, u_2(0)=1, u_1(1)=1, u_2(1)=2; u_1(x)-2u_2(x)+3x=0.$
Отв. $u_1(x)=(-x+2), u_2(x)=(x+1).$

8.38. $I\{u_1,u_2\}=\int\limits_0^1(u_1'^2+u_2'^2+1)dx;$
$u_1(0)=u_2(0)=u_2(1)=0, u_1(1)=2; u_1(x)+u_2(x)-2x^2=0.$
Отв. $u_1(x)=x^2+x, u_2(x)=x^2-x.$

8.39. $I\{u_1,u_2\}=\int\limits_0^\pi(u_1'^2-u_2'^2)dx;$
$u_1(0)=u_2(0)=u_1(\pi)=0, u_2(\pi)=\frac{\pi}{2}; u_1'(x)-u_2(x)+\cos x=0.$

Отв. $u_1(x)=\frac{x\sin x}{2}, u_2(x)=\frac{\sin x-x\cos x}{2}.$

8.40. $I\{u_1,u_2\}=\int\limits_0^{\frac{\pi}{2}}(u_1'^2+u_2'^2+2u_1u_2)dx;$
$u_1(0)=1, u_2(0)=-1, u_1(\frac{\pi}{2})=1+\frac{\pi^2}{4}, u_2(\frac{\pi}{2})=-1+\frac{\pi^2}{4};$
$u_1'(x)+u_2'(x)-4x=0.$
Отв. $u_1(x)=x^2+\cos x+\sin x, \quad u_2(x)=x^2-\cos x-\sin x.$

8.41. С помощью метода Ритца найти приближенно экстремаль функционала $I\{u\}=\int\limits_0^1(u'^2-u^2-2xu)dx$ при граничных условиях $u(0)=u(1)=0$, положив $n=2$. *Отв.* $u_2(x)=x(1-x)(0,192+0,171x).$

8.42. С помощью метода Канторовича найти приближенно экстремаль функционала $I\{u(x,y)\}=\iint\limits_{(G)}\left(\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2-2u\right)dxdy$, где $G=\{(x,y):$ $-2\le x\le 2, \quad -2\le y\le 2\}, \quad u(x,y)|_{\varepsilon G}=0, \quad n=1.$ *Отв.* $u_1(x,y)=\frac{1}{2}(4-y^2)\left(1-\frac{\operatorname{ch}\sqrt{2,5}\frac{x}{2}}{\operatorname{ch}\sqrt{2,5}}\right).$

8.43. С помощью метода Галеркина найти приближенно экстремаль функционала $I\{u\}=\int\limits_0^2(u'^2+xu')dx;$ при граничных условиях $u(0)=u(2)=0$, положив $n=1$. *Отв.* $u_1(x)=\frac{x(2-x)}{4}.$

Добавление.Оптимальное распределение плотности источников тепла

Рассмотренные выше прямые методы вариационного исчисления не исчерпывают всех возможных методов численного решения бесконечномерных задач. Существует ряд прикладных задач, при решении которых приходится комбинировать известные методы или изобретать новые.

Одной из инженерных задач, связанных с ресурсосберегающими технологиями, является задача об оптимальном расположении источников тепловых полей. Она всегда была актуальной задачей при проектировании в строительстве, металлургии и других областях техники. Эта задача допускает ряд постановок, которые не эквивалентны из-за различий в критериях оптимизации. По сути, здесь мы имеем целый ряд задач. Они различаются как по постановке, так и по методам решения. В настоящем разделе в качестве примера бесконечномерной задачи оптимизации рассматривается задача нахождения плотности источников тепла минимальной суммарной интенсивности, которая обеспечивает приемлемое распределение температур в области, находящейся в состоянии стационарного теплового баланса с окружающей средой. Умение решать такую задачу может быть использовано при оптимальной организации обогрева жилых и производственных помещений, для поддержания заданного температурного режима в однородных и неоднородных твердых телах и т.д. Перед тем, как дать точную формулировку задачи, сделаем несколько замечаний.

Существует три основных механизма распространения тепла: теплопроводность, конвекция и излучение. Последний играет существенную роль при высоких температурах (порядка $500-600^0$), поэтому в упомянутом выше круге задач этот механизм можно не учитывать. Чрезвычайно сложным является учет свободной конвекции, возникающей в гравитационном поле. Поскольку мы рассматриваем стационарные тепловые процессы, в которых свободная конвекция может отсутствовать[2], ее мы тоже не учитываем.

В итоге, как указано ниже, получается задача оптимального управления для эллиптических краевых задач. Вообще говоря, точного решения этой задачи может не существовать. Поэтому мы уточняем постановку задачи, вводя так называемое *квазирешение* или *ε-оптимальное решение* Далее кратко излагается методика численного нахождения этого решения. При этом для основных алгоритмов отсутствуют теоретические оценки необходимого количества операций. Поэтому эффективность этих алгоритмов далеко не очевидна и требует подтверждения в ходе вычисли-

[2] При определенных условиях могут возникать стационарные конвективные течения. В рассматриваемых здесь работах их влияние считается несущественным.

тельных экспериментов. В настоящем добавлении приводятся результаты таких экспериментов, выполненных с помощью разработанного авторами программного комплекса. Эти результаты свидетельствуют о достаточной эффективности этого комплекса и всей методики в целом.

Д.1.Задача нахождения оптимальной плотности источников тепла в неподвижной неоднородной среде

В работах [7]-[10], 35 рассматривалась следующая задача. В ограниченной связной области $D \subset R^m$ требуется определить функцию $f(\vec{x}) \geq 0$, доставляющую минимум линейному функционалу

$$J\{f\} = \int\limits_D f(\vec{x}) dV_m \to min, \tag{8.35}$$

при следующих условиях

$$\nabla * (\chi \nabla u) + f = 0,$$

$$(\frac{\partial u}{\partial n} + \beta u)|_{\partial D} = 0, \tag{8.36}$$

$$M(\vec{x}) - T_0 \geq u(\vec{x}) \geq m(\vec{x}) - T_0. \tag{8.37}$$

Здесь функция $u(\vec{x}) = T(\vec{x}) - T_0$, где $T(\vec{x})$ – установившаяся температура в точке области D, а T_0 константа, имеющая смысл температуры окружающей среды, $\chi(\vec{x}) > 0$ – температуропроводность в данной точке области D, а $\beta = \beta(\vec{x})$ –известная функция, определённая на границе ∂D, $m(\vec{x})$, $m(\vec{x})$ –задаваемые в области D минимальный и максимальный профили температуры, которые считаются непрерывными функциями. Плотность источников тепла $f(\vec{x})$ считается принадлежащей пространству $L_2(D)$ квадратично интегрируемых функций.

Обозначим через γ нижнюю границу значений функционала $J\{f\}$, когда пробегает множество неотрицательных функций из $L_2(D)$, удовлетворяющих условиям (8.36), (8.37).

Определение 1. Квазирешением оптимизационной задачи (8.35)–(8.37) при данном допуске $\varepsilon > 0$ назовём такую функцию $f_0(\vec{x}) \geq 0$ из $L_2(D)$, удовлетворяющую ограничениям (8.36), (8.37), для которой выполняется неравенство $J\{f_0\} \leq \gamma + \varepsilon$.

Приемлемость квазирешения определяется малостью ε. Ниже излагается метод численного нахождения квазирешения при любом заданном $\varepsilon > 0$. Предположим, что существуют такие положительные числа α_0 и δ_0 , для которых выполнены неравенства $\chi(\vec{x}) > \alpha_0$ $(\vec{x} \in D)$, $\beta(\vec{x}) \geq \delta_0$ $(\vec{x} \in \partial D)$. Тогда оператор $Lu = -\nabla \cdot (\chi \nabla u)$ с краевым условием $\left(\frac{\partial u}{\partial n} + \beta u\right)\Big|_{\partial D} = 0$ будет самосопряжённым, положительно определённым в $L_2(D)$, а значит, он имеет ограниченный обратный оператор $G{=}L^{-1}$. С его помощью можно переформулировать задачу (8.35)–(8.37), как задачу на минимум функционала $J\{f\}$ (8.35) при следующих условиях на плотность источников:

$$f(\vec{x}) \in L_2(D); \quad M(\vec{x}) - T_0 \geq (Gf)(\vec{x}) \geq m(\vec{x}) - T_0. \qquad (8.38)$$

Построим конечномерную аппроксимацию этой задачи в виде последовательности задач линейного программирования. Разобьём область D на n частей $\{D{=}\overset{[}{j}{=}\ 1]n\bigcup D_j\}$. Определим подпространство $S_n(D){\subset}s(D)$ кусочно-постоянных функций вида $f(\vec{x}){=}f_j$, $\vec{x}{\in}D_j(j{=}1,2,\ldots,n)$. Введём в $S_n(D)$ базис, состоящий из функций $e_j(\vec{x}){=}1$, $\vec{x}{\in}D_j$, и $e_j(\vec{x}){=}0$, $\vec{x}{\notin}D_j$. Тогда $f(\vec{x}){=}\overset{[}{i}{=}\ 1]n\sum f_je_j$. Введём обозначения $a_{ij}{=}(Ge_j,e_i)$, $(m(\vec{x}){-}T_0,e_i(\vec{x})){=}a_i$, $(M(\vec{x})-T_0,e_i(\vec{x})){=}b_i$, где $(\cdot{,}\cdot)$ скалярное произведение в $L_2(D)$. Подставляя выражение для $f(\vec{x})$ в (8.35) и умножая скалярно в $L_2(D)$ неравенства в (8.38) на $e_i(\vec{x})$, получаем задачу линейного программирования

$$\begin{aligned} &J_n(f) = \sum_{i=1}^{n}(mesD_j)f_j \to min, \\ &a_i \leq \sum_{i=1}^{n} a_{ij}f_i \leq b_i, f_i \geq 0 \quad (i = 1,2,\ldots,n) \end{aligned} \qquad (8.39)$$

Для выяснения связи между задачами (8.35)–(8.37) и (8.39), последнюю обозначим через $Z_0(n)$. Рассмотрим последовательность задач $Z_0(n)$, отвечающую такой последовательности разбиений области D, что $n{\to}\infty$. Назовем эту последовательность *конечномерной аппроксимацией задачи* (8.35)-(8.37). Обозначим через $(J_n)_{min}$ минимальное значение целевой функции задачи $Z_0(n)$.

Определение 2. Конечномерную аппроксимацию $Z_0(n)$ назовем *регулярной по функционалу*, если справедливо неравенство $\overset{[}{n}$ ${\to}\infty]\overline{\lim}(J_n)_{min} \leq \gamma$, где γ – число, фигурирующее в определении 1.

В работах [8],[9] показано, что при $m{\leq}3$ и условиях

$$1)m(\vec{x}) - T_0 \geq \delta_0 > 0; \quad 2) \lim_{n\to\infty} \max_j(diamD_j) = 0 \qquad (8.40)$$

конечномерная аппроксимация является регулярной по функционалу.

При наличии регулярности по функционалу и условии 2) решение $f(\vec{x}){=}\overset{[}{i}{=}\ 1]n\sum f_je_j$ конечномерной задачи $Z_0(n)$. при достаточно большом n можно считать приближенным квазирешением. Действительно, система ограничений в (8.39) означает, что неравенства (8.38) удовлетворяются в среднем по D_j, что при малых $diamD_j$ практически равносильно поточечным неравенствам. При этом значение $(J_n)_{min}$ при достаточно больших n не превосходит $\gamma + \varepsilon$.

Для построения задачи (8.39) требуется найти матрицу с элементами a_{ij}=(Ge_j, e_i), которую мы в дальнейшем называем *обменной матрицей.* Оператор G – интегральный оператор, ядро которого является функцией Грина краевой задачи (8.36). В одномерном случае (m=1) построить такую аппроксимацию можно без особого труда, так как в этом случае функцию Грина нетрудно найти в явном виде. При m>1 найти функцию Грина в явном виде затруднительно, поэтому обменная матрица строится численными методами. Находится функция u_i=Ge_j, которая является решением краевой задачи (8.36) с f=e_j. Эта краевая задача решается численно методом теплового баланса, а затем численным интегрированием определяются элементы aij. Построенную задачу (8.39) можно решить симплекс-методом, или одним из методов внутренних точек.

Замечание 1. При оценивании результатов приближенного решения задачи (8.35)–(8.37) весьма полезной является следующая оценка

$$\gamma \geq \overset{[}{\partial}\ D]\int \chi(\vec{x})\beta(\vec{x})(m(\vec{x}) - T_0)ds \tag{8.41}$$

В её справедливости нетрудно убедиться с помощью формулы Гаусса-Остроградского и условий (8.36), (8.37):

$$J\{f\}= \overset{[}{D}]\int f(\vec{x})dV_m=- \overset{[}{D}]\int \nabla\cdot(\chi \nabla u)dV_m=- \overset{[}{\partial}\ D]\int \chi \tfrac{\partial u}{\partial n}ds=$$
$$=\overset{[}{\partial}\ D]\int \chi\beta u ds \geq \overset{[}{\partial}\ D]\int \chi\beta(m(\vec{x}) - T_0)ds.$$

Д.2. Оптимальная плотность источников в движущейся среде

До сих пор предполагалось, что среда, заполняющая область не движется. В работе [8] поставлена оптимизационная задача без этого предположения. Полный учет конвекции приводит к очень сложной задаче, которую даже трудно точно сформулировать. Ниже поле скоростей среды $\vec{\nu}(\vec{x})$ в области предполагается фиксированным. Тем самым учитывается лишь искусственно создаваемая конвекция. Свободная конвекция в рассматриваемом стационарном процессе по-прежнему считается несущественной. Разобьём границу области D на три части $\partial D{=}_{+}\cup_{0}\cup_{-}$ где $_{+}$ часть границы, которая является входом среды в область D, $_{-}$ часть границы, являющейся выходом (стоком) среды, а $_{0}$ часть непроницаемой для среды границы. Справедливы следующие соотношения:

$$(\vec{n}, \vec{\nu})\big|_{0} = 0, \quad (\vec{n}, \vec{\nu})\big|_{-} > 0, \quad (\vec{n}, \vec{\nu})\big|_{+} < 0,$$

где $\vec{n}$ единичный вектор внешней нормали к границе ∂D. Последнее из этих трёх условий означает, что в область D поступает вещество внешней среды с температурой T_0. В дальнейшем мы предполагаем, что в нашем процессе присутствует поток тепла через непроницаемую для среды границу, равный $\alpha(\vec{x}) \cdot u(\vec{x}) \quad (\vec{x} \in_0)$, где $\alpha(\vec{x})>0$ коэффициент теплопередачи через $_0$. Рассмотрим задачу на минимум функционала $J\{f\}$ (8.35) при следующих условиях на плотность источников

$$\chi \Delta u - \nabla \cdot (\vec{\nu} u) + f = 0, \vec{x} \in D, \tag{8.42}$$

$$\begin{cases} \left(\chi \dfrac{\partial u}{\partial n} + \alpha u \right) \Big|_0 = 0, \\ \dfrac{\partial u}{\partial n} \Big|_- = 0, \\ \left(\chi \dfrac{\partial u}{\partial n} - (\vec{n}, \vec{\nu}) u \right) \Big|_+ = 0, \end{cases} \tag{8.43}$$

$$M(\vec{x}) - T_0 \geq u(\vec{x}) \geq m(\vec{x}) - T_0, \vec{x} \in D, \tag{8.44}$$

где χ коэффициент температуропроводности среды, который считается константой, $\vec{\nu}(\vec{x})$– поле скоростей среды, которое предполагается известным, подчиненным условию $div\vec{\nu}{=}0$ и потенциальным. Температурный режим (8.44) задается в некоторой подобласти $\tilde{D} \subseteq D$, которая в дальнейшем называется *областью контроля температуры.* В случае неподвижной среды считалось, что $\tilde{D}{=}D$. Вообще говоря, при постановке задачи (8.35)–(8.37) тоже можно требовать выполнение неравенств (Д.3) лишь в области , однако при теряется возможность оценки (8.37). С другой стороны, при в случае движущейся среды условие (8.44) нельзя удовлетворить при выполнении условия 1) в (8.40).

Для задачи (8.35), (8.42)–(8.44) имеет смысл понятие квазирешения, приближенное нахождение которого можно произвести, построив конечномерную аппроксимацию. Последнее требует преобразования краевой задачи (8.42)–(8.44), которое аналогично калибровочному преобразованию в электродинамике и использует потенциал $\varphi(\vec{x})$ поля скоростей $\vec{\nu}(\vec{x})$.

Отметим, что потенциал $\varphi(\vec{x})$ является решением следующей краевой задачи

$$\Delta \varphi = 0; \quad \frac{\partial \varphi}{\partial n} \Big|_0 = 0, \quad \frac{\partial \varphi}{\partial n} \Big|_- = s_1(\vec{x}), \quad \frac{\partial \varphi}{\partial n} \Big|_+ = -s_2(\vec{x}),$$

где положительные функции $s_2(\vec{x})$, $s_1(\vec{x})$ считаются известными и удовлетворяющими условию $\overset{[}{\partial}\ _+] \int s_1(\vec{x}) dS {=} \overset{[}{\partial}\ _-] \int s_2(\vec{x}) dS$ которое означает, что приток среды в область D равняется величине стока. Эта краевая задача имеет множество решений, отличающихся постоянным слагаемым. Для выделения единственного решения будем считать выполненным еще одно условие $\overset{[}{D}]\int\varphi(\vec{x})dV_m = 0$.

Преобразуем краевую задачу (8.42)–(8.44), вводя новую неизвестную функцию $w(\vec{x})$ следующим образом. Подставляя это выражение в (8.42) и учитывая, что, $\vec{\nu}(\vec{x})=\nabla\varphi(\vec{x})$, получим следующую краевую задачу

$$-\chi\Delta w + \left(|\nabla\varphi|^2/(4\chi)\right)w = f\cdot e^{-\varphi/(2\chi)}, \vec{x}\in D; \left(\frac{\partial w}{\partial n} + \sigma w\right)\bigg|_{\partial D} = 0 \quad (8.45)$$

где

$$\sigma(\vec{x}) = \begin{cases} \alpha(\vec{x})/\chi, & \vec{x}\in{}_0; \\ \dfrac{1}{2\chi}s_1(\vec{x}), & \vec{x}\in{}_-; \\ \dfrac{1}{2\chi}s_2(\vec{x}), & \vec{x}\in{}_+. \end{cases}$$

Оператор $Lw = -\chi\Delta w + \left(|\nabla\varphi|^2/(4\chi)\right)w$, действующий в пространстве $L_2(D)$ на достаточно гладкие функции $w(\vec{x})$, подчинённые краевым условиям (8.45), является самосопряжённым и положительно определённым, а поэтому имеет ограниченный обратный оператор $w = Gg$, определенный на $L_2(D)$. Поэтому мы можем переформулировать оптимизационную задачу (8.35),(8.42)–(8.44) следующим образом

$$J\{g\} = \overset{[}{D}]\int e^{\varphi(\vec{x})/(2\chi)}g(\vec{x})dV \to min,$$
$$\theta_2(\vec{x}) \geq Gg(\vec{x}) \geq \theta_1(\vec{x}), \vec{x}\in\tilde{D};$$
$$g(\vec{x})\in L_2(D), g(\vec{x})\geq 0, \vec{x}\in D,$$

где

$$g(\vec{x}) = f(\vec{x})e^{-(\varphi(\vec{x})/(2\chi))}; \theta_1(\vec{x}) = (m(\vec{x}) - T_0)e^{-(\varphi(\vec{x})/(2\chi))},$$
$$\theta_2(\vec{x}) = (M(\vec{x}) - T_0)e^{-(\varphi(\vec{x})/(2\chi))}.$$

Теперь можно построить конечномерную аппроксимацию этой задачи. Рассмотрим разбиение области D на части и введём кусочно-постоянные функции $g(\vec{x}) = \overset{[}{j} = 1]n\sum g_j e_j(\vec{x})$. Разбиение области D мы считаем и разбиением области $\tilde{D}\subseteq D$, т.е. при некотором натуральном p справедливо равенство $\tilde{D} = \overset{[}{i} = 1]p\bigcup D_i$. Как и выше, введём обозначения

$$\alpha_{ij} = (Ge_j, e_i), \alpha_i = (\theta_1, e_i), b_i = (\theta_2, e_i).$$

Заменяя класс функций $L_2(D)$ подпространством $S_n(D)$, умножая скалярно ограничения на базисные функции $e_i(\vec{x})$, получаем конечномерную аппроксимацию $Z_0(n)$ задачи

$$
\begin{aligned}
&J(g) = \overset{[}{j}= 1]n\sum c_j g_j \to min,\\
&\alpha_i \le \overset{[}{j}= 1]n\sum a_{ij} g_j \le b_i, \quad g_i \ge 0 \quad (i = 1, 2, \dots, p).
\end{aligned}
\tag{8.46}
$$

Здесь $c_j = c_j = \overset{[}{D}]\int e^{\varphi(\vec{x})/(2\chi)} dV$.

Можно показать, что конечномерная аппроксимация $Z_0(n)$ при выполнении условий (8.40) является регулярной по функционалу. А это означает, что при достаточно большом n решение этой задачи линейного программирования доставляет приближенное квазирешение задачи для движущейся среды. Таким образом, построение задачи (8.46) позволяет найти приближенное квазирешение. Численное построение этой задачи начинается с решения краевой задачи для определения потенциала $\varphi(\vec{x})$. Самым трудным при построении задачи (8.46) является нахождение обменной матрицы $a_{ij}{=}(Ge_j, e_i)$, поскольку оператор G явно не задан. Определение этой матрицы равносильно нахождению функций $w_j = Ge_j$, которые являются решениями уравнений $-\chi\Delta w + \left(\left|\nabla \varphi\right|^2/(4\chi)\right) w{=}e_j$ при краевых условиях (8.45). Эти краевые задачи так же, как и подобные им в случае неподвижной среды, решаются численно.

Д.3. Возможные модификации задачи

Выше рассмотрены основные формы задачи об оптимальном распределении источников тепла. Возможен ряд модификаций этой задачи как в случае неподвижной, так и для движущейся среды. Иногда ограничения сверху в (8.37), (8.44) отсутствуют или не существенны. Такую оптимизационную задачу назовём *односторонней*. В задачах (8.35)–(8.37), (8.42)–(8.44) возможно появление дополнительных условий, которые могут привести к некоторым новым модификациям. Одна из естественных модификаций состоит в требовании невозможности расположения источников тепла в некоторой части области D, т.е. возникает дополнительное требование равенства нулю плотности источников при $\vec{x}{\in}D_0{\subset}D$. Такую модификацию естественно назвать *задачей с ограничениями на локализацию источников*. Ещё одна модификация связана с присутствием некоторых фиксированных источников или стоков тепла до оптимизации. При этом плотность источников состоит из двух слагаемых, одно из которых известная функция, а второе подлежит определению. Такую задачу назовём *задачей с фиксированными источниками*. По форме она мало отличается от основной задачи. К функционалу $J\{\cdot\}$ добавляется постоянное слагаемое, а функции в ограничениях изменяются на однозначно определяемые слагаемые. Возможны различные комбинации рассмотренных выше модификаций. Наконец, можно ставить*задачу определения оптимального распределения стоков тепла* ($f(\vec{x}) \leq 0$), которую полезно решать при оптимальной организации охлаждения. Поскольку создание стоков тепла требует пропорциональных затрат энергии, эту задачу можно сформулировать как задачу на максимум отрицательного функционала (8.35).

Д.4. Численная реализация алгоритмов. Результаты вычислительных экспериментов

Для всех перечисленных модификаций задачи можно ввести понятие квазирешения, построить конечномерную аппроксимацию в виде последовательности задач линейного программирования, обладающую свойством регулярности по функционалу.

Для численного решения рассмотренных задач создан программный комплекс HeatCore, написанный на языке C# и не использующий сторонних математических библиотек. Предположим, что область D, имеющая форму параллелепипеда в R_m, разбита на ячейки D_j, которые считаются одинаковыми множествами: отрезками при m=1, прямоугольниками при m=2 и параллелепипедами при m=3.

В дальнейшем считается, что ячейки возникают в результате разрезания области D точками (m=1), двумя семействами параллельных сторонам D прямых (m=2) или тремя аналогичными семействами плоскостей (m=3). На рис. Д.1 приведена блок-схема алгоритма решения m-мерной задачи для движущейся среды с использованием численного метода нахождения обменной матрицы. Для неподвижной среды блок-схема аналогична, но не содержит блока нахождения потенциала поля скоростей. На рис. Д.2 и Д.3 приведены графики стабилизации значений $(J_n)_{min}$ с ростом n, полученные в результате численных экспериментов в трехмерном случае, соответственно для неподвижной и движущейся сред с различными значениями χ.

На рис. Д.4 приведён результат численного эксперимента для 3-мерной области D размером $5 \times 4 \times 3$ м с принудительным источником холода. В качестве вещества, заполняющего область, был выбран воздух с коэффициентом температуропроводности χ=$2,216 \times 10^{-52}$/. Границы области выполнены из кирпича с коэффициентом температуропроводности $\chi_{к}$=$5,2 \times 10^{-72}$/ толщиной 30 см. На дальней плоскости имеется окно с χ_c=$3,4 \times 10^{-72}$/ площадью 3 ^{2} (стеклянное окно). В правую грань области входит вещество со скоростью $s_2(x, y, z)$=10^{-4}/ ($_+$ имеет площадь 2 ^{2}). Симметрично с левой стороны имеется участок стока $_-$ такой же площади с такой же скоростью вещества $s_1(x, y, z)$=10^{-4}/. Минимальный

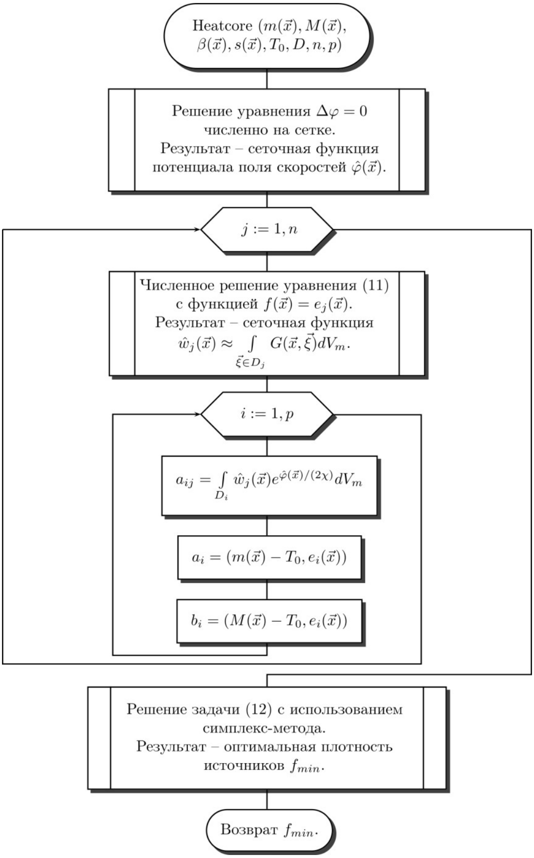

Рис. Д.1. Блок-схема алгоритма для решения m-мерной задачи

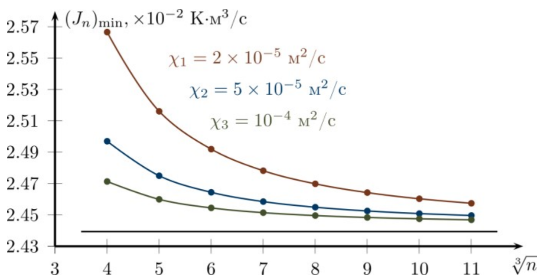

Рис. Д.2. Зависимость значения $(J_n)_{min}$ от n для неподвижной среды

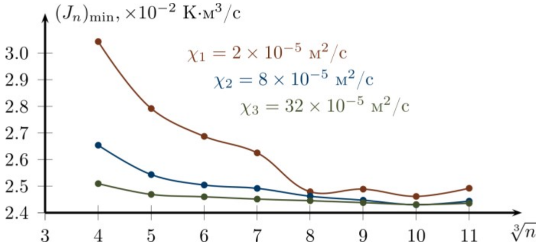

Рис. Д.3. Зависимость значения $(J_n)_{min}$ от n для движущейся среды

и максимальный профили температур заданы функциями $m(x, y, z)$=290, $M(x, y, z)$=320, температура внешней среды T_0=260. Область контроля занимает всю область за исключением пространства вблизи входа воздуха $_+$

$$\tilde{D}=D - \{x > 4,5 \ \ z > 0,5 \ \ z < 2,5\}.$$

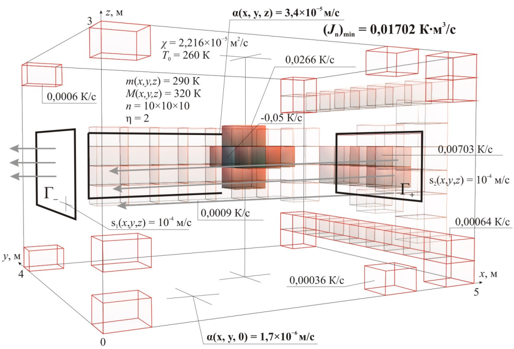

Рис. Д.4. Оптимальное расположение источников тепла внутри параллелепипеда

Разноцветными объёмами показана искомая оптимальная плотность источников тепла (красные) и сток (синий). Источники распределяются вблизи места входа вещества, вдоль окна, а также около стока, таким образом его нейтрализуя. Суммарная интенсивность источников имеет значение 0,01702 К $\cdot$ м^3/с. Для перевода в систему СИ необходимо вычислить потери тепла в единицу времени(Дж/с) через границу

$$J_1 = \int\limits_{\Gamma_0} \frac{\chi_1}{d} u(x, y, z) dS + \int\limits_{\Gamma_-} \frac{(\chi_1)_{\text{в}}}{\chi} s_1(x, y, z) u(x, y, z) dS,$$

где χ_1 – коэффициент теплопроводности через границу во внешнюю среду

[Вт/(м·К)], d – толщина стенки. Для стекла $(\chi_1)_c = 1$Вт/(м · К), кирпича $(\chi_1)_к = 0{,}5$ Вт/(м·К), воздуха $(\chi_1)_в = 0{,}0243$ Вт/(м·К). Результирующее значение J_1=13 131 Вт.

Для проверки эффективности был проведён следующий эксперимент. Было взято некоторое случайное распределение источников (рис. Д.5) и решена прямая задача теплопроводности для той же самой области (рис. Д.4) и тех же параметрах среды, но без стока тепла. В результате было получено, что случайное распределение источников нагревает область контроля в диапазоне температур 292,8−364,8 . Затем была решена прямая задача с $m(x, y, z)$=292,8 и $M(x, y, z)$=364,8 при $(x, y, z) \subseteq \tilde{D}$. В итоге получилось, что случайное распределение (рис. Д.5) имеет суммарную интенсивность $0,02687\ \cdot^3/$, а оптимальное (рис. Д.6) – $0,01476\ \cdot^3/$. Сэкономленная мощность при этом составляет около 45%.

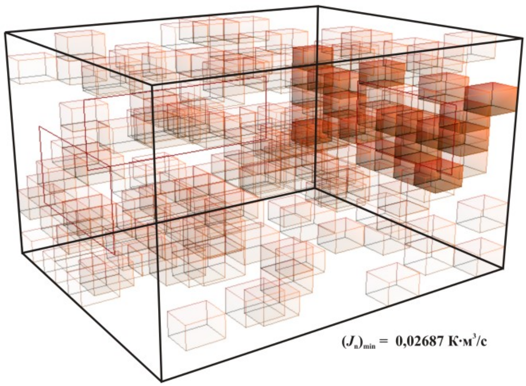

Рис. Д.5. Случайное расположение источников тепла внутри параллелепипеда

На рисунке Д.7 представлен результат расчёта оптимального распределения источников тепла внутри области сложной геометрической формы размером $5 \times 3,5 \times 10$ м, составленной из тетраэдров в программе Solidworks. Большая часть границы области выполнена из материала с коэффициентом теплопередачи α_1=4 ($^2\cdot$), за исключением небольшого окна с α_2=0, 5 ($^2\cdot$). Область заполнена воздухом с коэффициентом теплопроводности χ=0, 0267 ($\cdot$). До решения задачи оптимизации в область было добавлено n=544 источника тепла таким образом, чтобы они полностью заполнили область D (за исключением промежутков между источниками). Область контроля температуры $\tilde{D}$, температура внешней среды и

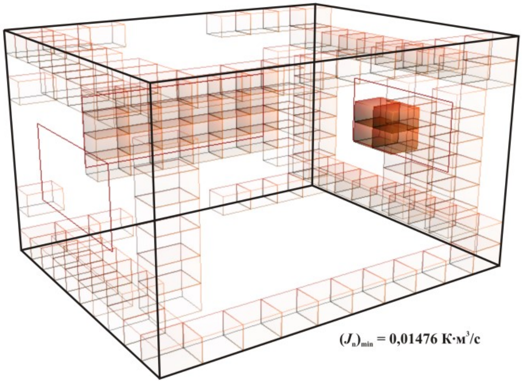

Рис. Д.6. Оптимальное расположение источников внутри параллелепипеда

температурный коридор были заданы следующими:

$$\tilde{D} = \left\{(x-2,5)^2 + (y-1)^2 < 2,2^2\right\} \cup \{z < 1\},$$

$$m(x,y,z) = \begin{cases} 5^oC, \ z > 1, \\ 25^oC, \ z \leq 1, \end{cases} \quad M(x,y,z) = 80^oC, \quad T_0 = 0^oC$$

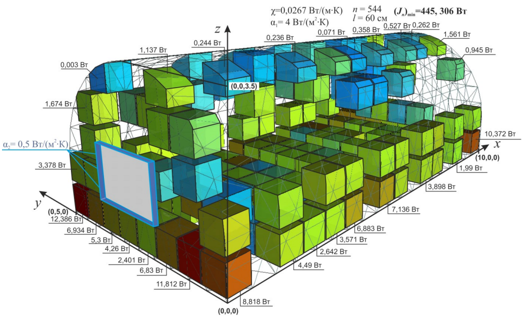

Рис. Д.7: Оптимальное расположение источников тепла при $D = D_f$

На рис. Д.7 оптимальное распределение источников показано в виде объёмов различного цвета. Источники красного оттенка являются самыми мощными, зелёного – средней мощности, синего – малой мощности. Некоторые источники имеют нулевую мощность, т.е. отсутствуют, и на рисунке не показаны. Источники стабилизируются в основном вдоль границы области. Это объясняется тем, что необходимо компенсировать рассеивание тепла через границу, которая имеет сравнительно большой коэффициент теплопередачи, и обеспечить около неё температурный коридор.

Самые мощные источники располагаются на углах, так как здесь самое большое рассеивание тепла. По условию эксперимента, верхнюю часть области ($z>1$) нужно нагреть только до температуры 5^0, поэтому алгоритм располагает в данной части только источники малой мощности. Суммарная мощность источников тепла составляет $(J_n)_{min}$=445, 306 . Исходная сетка содержала 6487 тетраэдров. После модификации сетки и добавления источников тепла количество многогранников увеличилось до 67912.

Выводы.Регулярность конечномерной аппроксимации $Z_0(n)$ по функционалу показывает лишь принципиальную возможность приближенного нахождения квазирешения. Остается открытым вопрос о том насколько большим нужно взять n, чтобы получить квазирешение с данным допуском ε. Из приведенных результатов вычислений видно, что последовательность $(J_n)_{min}$ довольно быстро стабилизируется с ростом n. Для случая неподвижной среды удается довольно близко подобраться к величине правой части оценки (8.41). Скорость стабилизации $(J_n)_{min}$ увеличивается с ростом температуропроводности среды χ. Значения $(J_n)_{min}$ при больших n слабо зависят от χ для неподвижной среды. Для движущейся среды эта зависимость значительно сильнее. Наконец, вычислительные эксперименты показывают возможность значительной экономии энергии в результате оптимизации по сравнению со случайной плотностью источников при таком же температурном коридоре.

Заключение

Завершая настоящее пособие, отметим, что в нем мы, в основном, рассматривали традиционные методы оптимизации. В настоящее время иногда успешно используют так называемые эволюционные методы и алгоритмы, в частности генетические алгоритмы. Все они моделируют базовые положения теории биологической эволюции – процессы отбора, мутации и воспроизводства. Здесь мы этих методов не касаемся, поскольку они относятся скорее к области искусственного интеллекта, а не к традиционным методам оптимизации. Подробно ознакомиться с эволюционными методами оптимизации можно по монографии [14].

Список литературы

[1] Акулич, И.Л. Математическое программирование в примерах и задачах / И.Л. Акулич. — М.: Высш. шк. ,1986. — 318 с.

[2] Ахиезер Н.И. Вариационное исчисление / Н.И. Ахиезер. – Харьков: Вища школа. – Изд-во при Харьк. ун-те, 1981.– 168с.

[3] Ашманов, С.А. Линейное программирование / С.А. Ашманов. — М.: Наука, Главная редакция физико-математической литературы, 1981.— 340 с.

[4] Ашманов, С.А. Теория оптимизации в задачах и упражнениях / С.А. Ашманов, А.В. Тихонов. — М.: Наука, 1991. — 447 с

[5] Беллман, Р. Динамическое программирование / Р Беллман. — М.: ИЛ, 1960. — 430 с.

[6] Болтянский, В.Г. Оптимальное управление дискретными системами / В.Г. Болтянский. — М.: Наука, 1973. — 446 с.

[7] Брусенцев А.Г., Брусенцева В.С. Задача об оптимальном выборе источников тепла. // Сб. трудов XXIII международной конференции «Математические методы в технике и технологиях». – Т.2. – 2010. – С. 43–46.

[8] Брусенцев А.Г., Осипов О.В. Оптимальный выбор источников тепла при наличии конвекции // Научные ведомости БелГУ. Математика. Физика. – № 26 (169). Выпуск 33. Белгород. – 2013. – С. 64–82.

[9] Брусенцев А.Г., Осипов О.В. Приближенное решение задачи об оптимальном выборе источников тепла // Научные ведомости Белгородского государственного университета. – №5 (124). Выпуск 26. Белгород. – 2012. – С. 60–69.

[10] Брусенцев А.Г., Осипов О.В. Численное нахождение обменной матрицы при решении задачи оптимизации распределения источников тепла // Вестник Белгородского государст-венного технологического университета имени В.Г. Шухова. №5. – 2016. – С.116-124.

[11] Вагнер, Г. Основы исследования операций / Г. Вагнер. — М.: Мир, 1972 — 1973. Т.1 — 3. — 987 с.

[12] Вентцель, Е.С. Исследование операций (Задачи, принципы, методология) / Е.С. Вентцель. — М: Наука, 1980. — 208 с.

[13] Волков, И.К. Исследование операций / И.К. Волков, Е.А. Загоруйко. — М.: Изд. МГТУ им. Н.Э. Баумана, 2004. — 440 с.

[14] Гладков Л. А., Курейчик В. В, Курейчик В. М. Биоинспирированные методы в оптимизации: монография / Л. А. Гладков, В. В Курейчик, В. М. Курейчик и др. —М.: Физматлит, 2009. — 384 с.

[15] Гольштейн, Е.Г. Задачи линейного программирования транспортного типа / Е.Г. Гольштейн, Д.Б. Юдин. — М.: Наука, 1969. — 382 с.

[16] Гольштейн, Е.Г. Линейное программирование / Е.Г. Гольштейн, Д.Б. Юдин. — М.: Наука, 1969. — 387 с.

[17] Данциг, Дж. Линейное программирование, его обобщения и приложения / Дж. Данциг. — М.: Прогресс, 1966. — 600 с.

[18] Дикин, И.И. Метод внутренних точек в линейном и нелинейном программировании / И.И. Дикин. — Изд. группа URSS, 2010. —120 с.

[19] Зайченко, Ю.П. Исследование операций / Ю.П. Зайченко. — Киев: Выща школа, 1988. — 550с.

[20] Заславский, Ю.Л. Сборник задач по линейному программированию / Ю.Л. Заславский. — М.: Наука, 1969. — 256с.

[21] Исследование операций в экономике / под редакцией профессора Н.Ш. Кремера. — М.: ЮНИТИ, 2003. — 407с.

[22] Калихман, И.Л. Сборник задач по математическому программированию / И.Л. Калихман . — М.: Высш. шк., 1975. —270 с.

[23] Канторович Л.В., Крылов В.И. Приближенные методы высшего анализа / Л.В. Канторович, В.И. Крылов. – М.: Физматгиз.–1962.– 709с.

[24] Карпелевич, Ф.И. Элементы линейной алгебры и линейного программирования / Ф.И. Карпелевич, Л.Е. Садовский. — М.: Наука, 1967. — 274 с.

[25] Крушевский, А.В. Теория игр / А.В. Крушевский. — Киев: Издательское объединение «Вища школа», 1977. — 216 с.

[26] Кузнецов, Ю.Н. Математическое программирование / Ю.Н. Кузнецов, В.И. Кузубов, А.Б. Волощенко. — М.: Высш. шк., 1980. — 371 с.

[27] Линейное и нелинейное программирование / под редакцией профессора И.Н. Ляшенко. — Киев: Издательское объединение «Вища школа», 1975. — 370с.

[28] Майника, Э. Алгоритмы оптимизации на сетях и графах: Пер. с англ. / Э. Майника. — М.: Мир, 1981. — 323 с.

[29] Михлин С.Г. Вариационные методы в математической физике / С.Г. Михлин.– М.: Наука, – 1970.– 512с.

[30] Михлин С.Г. Проблема минимума квадратичного функционала / С.Г. Михлин. – М.: Гостехиздат.– 1952.– 218с.

[31] Михлин С.Г. Численная реализация вариационных методов / С.Г. Михлин. – М.: Наука, – 1966.– 432с.

[32] Морозов, В.В. Исследование операций в задачах и упражнениях / В.В. Морозов, А.Г. Сухарев, В.В. Федоров. — М: Высш. шк., 1986. — 314 с.

[33] Мышкис А.Д. Математика. Специальные курсы для ВТУЗов / А.Д. Мышкис.– М.:Наука.– 1971.–632с.

[34] Нейман, Дж. Теория игр и экономическое поведение / Дж. Нейман, О. Моргерштерн. — М.: Наука, 1970. — 708 с.

[35] Осипов О.В. Оптимальное расположение источников тепла в неоднородной среде // Вестник Белгородского государственного технологического университета имени В.Г. Шухова. №1. – 2013. – С. 154–158.

[36] Саати, Т.Л. Математические методы исследования операций / Т.Л. Саати. — М.: Воениздат, 1963. — 353 с.

[37] Сборник задач по математике для ВТУЗов / под редакцией А.В. Ефимова и А.С. Поспелова. – М.: Изд-во Физико-матем. литер. – Т.3. – 2003.– 575с.

[38] Схрейвер, А. Теория линейного и целочисленного программирования: В 2-х т.: Пер с англ. / А. Схрейвер. — М.: Мир, 1991. — 360 с, 342 с.

[39] Таха, Х. Введение в исследование операций / Х. Таха. — Изд. Вильямс, 2005. — 903с.

[40] Тихонов, А.Н. Методы решения некорректных задач / А.Н. Тихонов, В.Я. Арсенин. — М.: Наука, 1974. — 222с.

[41] Форд, Л. Р. Потоки в сетях: Пер. с англ. / Л. Р. Форд, Д. Р. Фалкерсон. — М.: Мир, 1966. — 276 с.

[42] Хачиян, Л.Г. Полиномиальный алгоритм в линейном программировании / Л. Г. Хачиян. // ЖВМ и МФ. — 1980. — т. 20. — №1, с. 51 — 68.

[43] Хедли, Дж. Нелинейное и динамическое программирование / Дж. Хедли. — М.: Мир, 1967. — 560 с.

Учебное издание

Брусенцев Александр Григорьевич
Осипов Олег Васильевич

Методы оптимизации

Учебное пособие

Подписано в печать 12.09.17. Формат $60 \times 84/16$. Усл. печ. л. 15,3. Уч.-изд. л. 16,4. Тираж 75 экз. Заказ Цена Отпечатано в Белгородском государственном технологическом университете
им. В. Г. Шухова
308012, г. Белгород, ул. Костюкова, 46