

И. В. Ремизова

**ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
И МЕТОДЫ ОПТИМИЗАЦИИ**

Учебное пособие

**Санкт-Петербург
2026**

Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«Санкт-Петербургский государственный университет
промышленных технологий и дизайна»
Высшая школа технологии и энергетики**

И. В. Ремизова

ИССЛЕДОВАНИЕ ОПЕРАЦИЙ И МЕТОДЫ ОПТИМИЗАЦИИ

Учебное пособие

Утверждено Редакционно-издательским советом ВШТЭ СПбГУПТД

Санкт-Петербург
2026

УДК 681.3(075)

ББК 30.10я7

Р 380

Рецензенты:

доктор технических наук, доцент, профессор кафедры аудиовизуальных систем
и технологий, ведущий научный сотрудник государственного института кино и телевидения

В. М. Пестриков;

кандидат технических наук, доцент, доцент кафедры ИИТСУ Высшей школы
технологии и энергетики Санкт-Петербургского государственного университета
промышленных технологий и дизайна

Е. П. Дятлова

Ремизова, И. В.

Р380 Исследование операций и методы оптимизации: учебное пособие /
И. В. Ремизова. — СПб.: ВШТЭ СПбГУПТД, 2026. — 106 с.

ISBN 978-5-91646-475-7

Учебное пособие соответствует программам и учебным планам дисциплины «Исследование операций и методы оптимизации» для студентов, обучающихся по направлению подготовки 09.03.03 «Прикладная информатика». Изложены классические методы решения экстремальных задач с критериями оптимальности в виде функции многих переменных. Дается классификация и приводятся примеры экстремальных задач, решаемых в различных функциональных подсистемах производства.

Пособие предназначено для подготовки бакалавров очной и заочной форм обучения. Отдельные разделы пособия могут быть полезны студентам других специальностей (например, 01.03.02 – «Прикладная математика и информатика»), магистрам, аспирантам и специалистам, занимающимся решением экстремальных задач.

УДК 681.3(075)

ББК 30.10я7

ISBN 978-5-91646-475-7

© ВШТЭ СПбГУПТД, 2026

© Ремизова И. В., 2026

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ. ОПТИМИЗАЦИЯ СИСТЕМ УПРАВЛЕНИЯ В ЭПОХУ ЦИФРОВЫХ ТЕХНОЛОГИЙ.....	4
ГЛАВА 1. ПРОБЛЕМА ОПТИМАЛЬНОГО УПРАВЛЕНИЯ НА ПРОИЗВОДСТВЕ.....	6
1.1. Общая постановка задачи оптимизации	6
1.2. Классификация экстремальных задач по функциональному признаку	7
1.3. Характеристика экстремальных задач, решаемых в АСУ ТП.....	7
1.4. Классификация экстремальных задач по виду их математической модели.....	10
ГЛАВА 2. ЗАДАЧИ ОТЫСКАНИЯ ЭКСТРЕМУМА ФУНКЦИЙ ОДНОЙ И МНОГИХ ПЕРЕМЕННЫХ. АНАЛИТИЧЕСКИЕ МЕТОДЫ РЕШЕНИЯ.....	11
2.1. Безусловный экстремум. Условия существования	11
2.2. Задача аналитической градуировки датчика	15
2.3. Задачи на условный экстремум с ограничениями-равенствами. Метод множителей Лагранжа. Условия существования экстремума в задаче с ограничениями	17
2.4. Примеры задач на условный экстремум с ограничениями-равенствами.....	22
2.5. Задачи с ограничениями смешанного типа. Общий алгоритм решения	27
2.6. Задача выбора оптимального технологического режима процесса сушки целлюлозы.....	31
ГЛАВА 3. ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ.....	35
3.1. Постановка задач линейного программирования	35
3.2. Геометрическая интерпретация задачи ЛП.....	36
3.3. Симплексный метод решения задачи линейного программирования с односторонними ограничениями	41
3.4. Модификация симплексного алгоритма для задачи ЛП с двусторонними ограничениями.....	63
3.5. Задача выбора оптимального технологического режима процесса обезвоживания осадка сточных вод	75
ГЛАВА 4. ЗАДАЧИ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ	78
4.1. Общая постановка задачи нелинейного программирования. Характеристика методов решения.....	78
4.2. Численные методы безусловной оптимизации	80
4.3. Решение задач на условный экстремум. Методы штрафных функций	93
4.4. Решение задач на условный экстремум. Методы допустимых направлений	100
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	106

ВВЕДЕНИЕ

ОПТИМИЗАЦИЯ СИСТЕМ УПРАВЛЕНИЯ В ЭПОХУ ЦИФРОВЫХ ТЕХНОЛОГИЙ

Проблема оптимальности остается одной из центральных в управлении, особенно при повсеместной цифровизации и внедрении интеллектуальных технологий. Сегодня практически любая задача, связанная с проектированием или эксплуатацией систем управления, формулируется как задача оптимизации, для решения которой применяются не только классические методы теории оптимальных систем, но и современные подходы – машинное обучение, предиктивная аналитика и адаптивное управление.

Особую актуальность проблема оптимальности приобрела в связи с переходом к концепциям Индустрии 4.0 и Индустрии 5.0, где ключевую роль играют киберфизические системы, цифровые двойники, облачные платформы и искусственный интеллект. Эти технологии позволяют не только автоматизировать, но и оптимизировать производственные процессы в режиме реального времени с учетом динамики рынка, энергоэффективности и экологических требований.

В современных условиях цифровые платформы управления (ЦПУ) и автоматизированные системы управления технологическими процессами (АСУ ТП) становятся стандартом для крупных промышленных предприятий. Например, в целлюлозно-бумажной промышленности (ЦБП) уже внедрены десятки таких систем, а в ближайшем времени планируется полный переход на адаптивные цифровые решения, интегрированные с системами мониторинга и управления на основе больших данных.

Внедрение ЦПУ и АСУ ТП требует значительных инвестиций – как временных, так и финансовых. Поэтому ключевым условием их эффективности становится окупаемость за счет повышения производительности, снижения издержек и улучшения качества продукции. Если система не обеспечивает оптимизацию процессов на всех уровнях – от операционного до стратегического, – ее внедрение теряет экономический смысл.

Современные исследования подтверждают: основной источник эффективности цифровых систем управления – решение задач оптимизации в режиме реального времени с учетом множества переменных и внешних факторов. Это стало возможным благодаря развитию вычислительных

мощностей, нейросетей и алгоритмов искусственного интеллекта, которые позволяют анализировать большие объемы данных и принимать оптимальные решения быстрее, чем традиционные методы.

В связи с этим настоящее учебное пособие посвящено постановке и методам решения задач оптимизации, характерных для современных АСУ ТП, с акцентом на цифровые технологии и практические кейсы из промышленности.

ГЛАВА 1. ПРОБЛЕМА ОПТИМАЛЬНОГО УПРАВЛЕНИЯ НА ПРОИЗВОДСТВЕ

1.1. Общая постановка задачи оптимизации

Исследование операций – это научный подход к решению задач организационного управления. Целью исследования операций является получение оптимальных решений задач, возникающих в процессе функционирования или создания сложных систем (включая коллективы людей, машины и средства вычислительной техники) [1]. Характерной чертой сложных систем¹ независимо от их физической природы – будь то биологическая система или сложная техническая система, – является целенаправленность поведения. При стремлении достигнуть цели сразу же возникают три задачи [2].

Первая задача – выбор и формулировка цели, которая зависит от очень многих условий и нередко сопряжена с большими трудностями. Очень часто мы знаем, что нам нужно, но не можем точно сформулировать. Как только цель выбрана, возникает вторая задача – согласование цели с имеющимися возможностями, т.е. учет ограничений. Даже ясная формулировка цели еще не является залогом ее достижения. И, наконец, третья задача – реализация способа достижения цели с учетом ограничений, которая сводится к выбору лучшего в некотором смысле варианта решения из допустимого множества вариантов.

Чтобы оценить качество того или иного варианта решения задачи, необходимо располагать некоторым показателем или численной характеристикой, определяющей степень достижения поставленной цели при выбранном варианте. Эта численная характеристика называется критерием оптимальности. Очевидно, что лучшему варианту соответствует экстремальное значение критерия оптимальности – минимум или максимум, в зависимости от конкретного смысла задачи. Поэтому такого рода задачи называют экстремальными.

Для решения экстремальной задачи необходимо перейти от ее содержательной постановки к математической формализации, т.е. записать в виде математических соотношений критерий оптимальности и все ограничивающие условия. Совокупность критерия оптимальности и системы дополнительных ограничений называется математической моделью задачи. Математическая модель должна быть адекватна содержательной формулировке задачи, с исчерпывающей полнотой отражая цель решения задачи и имеющиеся ограничения. Основная трудность при этом состоит в выборе критерия оптимальности, так как эта процедура не поддается формализации, а зависит от квалификации и интуиции разработчика задачи. В выбор критерия оптимальности всегда вносится значительный элемент субъективизма; в то же время удачный выбор критерия во многом определяет успех решения задачи, а

¹Сложной или большой системой принято называть систему, включающую большое число элементов различной природы со сложной структурой связи между ними.

если задача входит в состав задач АСУ, то и успех функционирования системы в целом. Поэтому всегда необходимо учитывать последствия, которые могут иметь место при оптимизации по выбранному критерию.

Система ограничений в экстремальной задаче должна определять то множество допустимых вариантов решений, на котором осуществляется поиск решения, оптимального в смысле заданного критерия. В поиске лучшего решения среди множества допустимых состоит сущность оптимизации. Нет возможности выбора – допустимое множество содержит одно или несколько решений – оптимизация теряет смысл.

Из сказанного выше ясно, насколько ответственен этап постановки экстремальной задачи и разработки ее математической модели.

1.2. Классификация экстремальных задач по функциональному признаку

Экстремальные задачи, решаемые в рамках АСУ, имеют самое различное смысловое содержание, однако по функциональному назначению среди этих столь разнообразных по содержанию задач может быть выделено не такое уж большое число классов.

Согласно предлагаемой классификации, будем различать задачи:

1. Оптимальной фильтрации, т.е. отделения полезного сигнала от помехи.
2. Идентификации, т.е. определения характеристик объектов управления по экспериментальным данным.
3. Оптимального управления, т.е. синтеза закона управления.
4. Адаптации, т.е. изменения параметров и структуры системы управления на основе текущей информации с целью достижения оптимального ее функционирования при изменяющихся характеристиках объекта управления или внешней среды.
5. Исследования операций или организационного управления производственным процессом.

1.3. Характеристика экстремальных задач, решаемых в АСУ ТП

Для установления связи экстремальных задач с уровнями управления АСУ ТП рассмотрим обобщенную функциональную структуру АСУ ТП.

Как любая система управления АСУ ТП включает две основные подсистемы: объект управления и управляющую часть. Объект управления в АСУ ТП представляет собой технологический процесс, протекающий в одном агрегате или в совокупности взаимосвязанных агрегатов, и охватывает участок производства, цех и даже завод. Например, объектом АСУ ТП могут быть: процесс варки целлюлозы, процесс промывки целлюлозы, варочно-промывной цех и, наконец, сульфат-целлюлозный завод. Таким образом, объект управления АСУ ТП является сложной системой, и, соответственно, управляющая часть АСУ ТП также является сложной системой и имеет иерархическую структуру,

т.е. строится с выделением нескольких функциональных уровней управления, функциональная структура АСУ ТП приведена на рисунке 1.1.

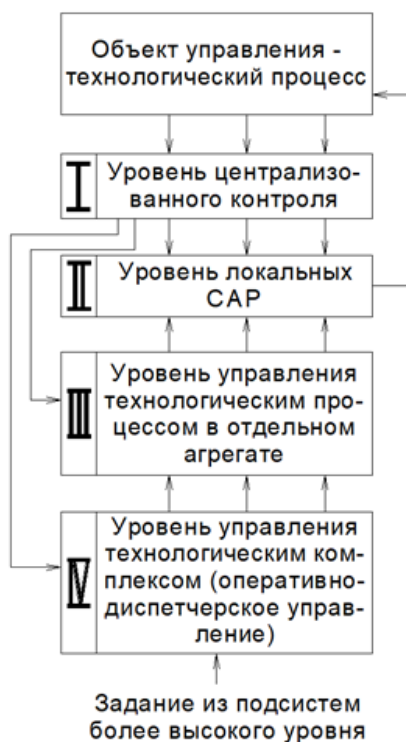


Рисунок 1.1. Иерархия уровней управления АСУ ТП

На рисунке 1.1 показано, что первый или нижний уровень управления – это уровень централизованного контроля. Учитывая, что в АСУ ТП состояние объекта может определяться десятками и сотнями параметров, целесообразно задачи контроля объединить в одну подсистему.

Информация подсистемы централизованного контроля используется для поддержания заданного значения отдельных параметров объекта с помощью локальных систем автоматического регулирования (САР). Задачи, решаемые в локальных САР, отнесены ко второму уровню управления АСУ ТП. Задание локальным САР поступает с верхнего третьего уровня, объединяющего задачи управления технологическим процессом в отдельном агрегате, и, наконец, на последнем уровне – оперативно-диспетчерского управления (ОДУ) – решаются задачи координации работы взаимосвязанных агрегатов.

Не останавливаясь на всех задачах, решаемых на каждом уровне АСУ ТП, дадим лишь краткую характеристику экстремальных задач.

Уровень централизованного контроля. Трудно переоценить значение подсистемы централизованного контроля для эффективности функционирования АСУ ТП: от полноты и достоверности информации, поступающей из этой подсистемы, зависит достоверность решения задач в узлах верхнего уровня. Поэтому выбор оптимальных параметров подсистемы централизованного контроля – это серьезная проблема, включающая следующие задачи из перечисленных выше классов:

- аналитической градуировки датчиков, позволяющие с заданной точностью получать математические модели датчиков;
- корректировки математических моделей датчиков на основании текущей информации;
- оптимальной фильтрации, позволяющие определить структуру и параметры сглаживающего фильтра, из условия, например, максимальной точности воспроизведения полезного сигнала;
- выбора оптимального периода опроса датчиков;
- выбора оптимальных алгоритмов расчета технико-экономических показателей (ТЭП) из условия минимальной погрешности.

На уровне локальных САР решаются следующие экстремальные задачи:

- идентификации объекта управления, т.е. определения статических и динамических характеристик каналов передачи возмущающих и управляющих воздействий;
- оптимального управления, к которым относятся задачи синтеза закона управления в САР и задачи параметрического синтеза, т.е. выбора оптимальных параметров настройки при заданной структуре регулятора;
- адаптации, т.е. корректировки статических и динамических характеристик объекта, а также закона управления или параметров регулятора при изменении свойств объекта или условий функционирования САР.

Уровень управления технологическим процессом в отдельном агрегате включает задачи:

- идентификации управления технологическим процессом в переходных режимах (при пуске, останове, изменении нагрузки и т.д.) и установившемся режиме (выбор параметров оптимального технологического режима);
- адаптации, т.е. коррекции математической модели объекта и алгоритмов управления в случае нестационарности характеристик процесса или окружающей среды. В подсистеме верхнего уровня (ОДУ) решаются главным образом задачи координации взаимосвязанных технологических процессов, которые относятся к классу организационного управления. Это задачи оптимального оперативно-календарного планирования выпуска и отгрузки продукции, распределения ограниченных ресурсов, управления запасами сырья и полуфабрикатов в буферных емкостях и т.д.

Из краткой характеристики экстремальных задач, которая дана выше, видно, что на всех уровнях управления, кроме уровня ОДУ, по своему функциональному назначению решаются задачи из одних и тех же классов, однако содержание и сложность задач различны. Если на нижнем уровне в качестве объекта рассматривается датчик, для которого решается задача идентификации, то на уровне локальных САР объектом является некоторый элемент технологического процесса, а на еще более высоком уровне –

технологический процесс в целом. Таким образом, с повышением уровня управления возрастает сложность экстремальных задач, в частности растет размерность, т.е. число переменных, от которых зависит критерий оптимальности и число ограничений.

1.4. Классификация экстремальных задач по виду их математической модели

В зависимости от вида математической модели можно рассмотреть два направления классификации экстремальных задач: по критерию оптимальности (рис. 1.2) и по типу дополнительных ограничений (рис. 1.3).

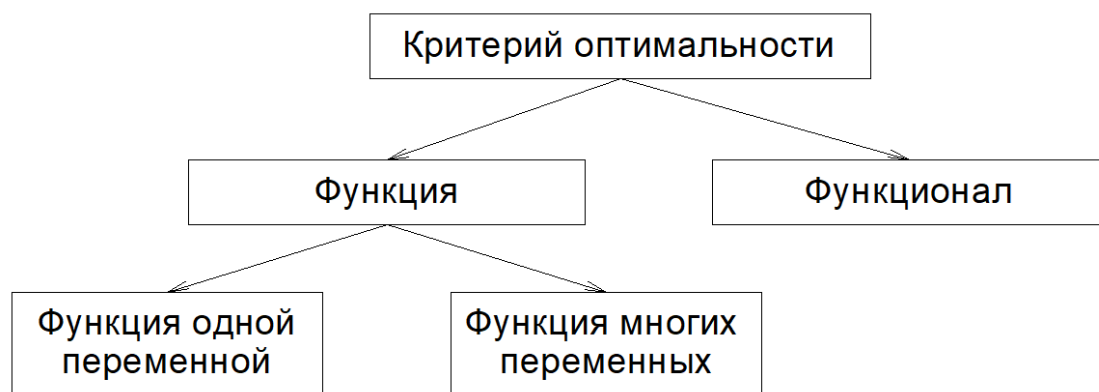


Рисунок 1.2. Классификация экстремальных задач по виду критерия оптимальности

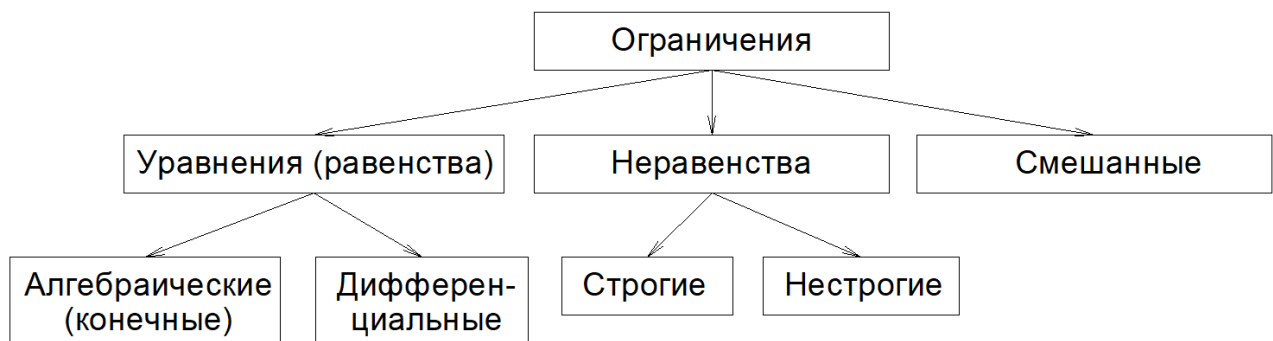


Рисунок 1.3. Классификация экстремальных задач по типу дополнительных ограничений

Если экстремальная задача не содержит никаких ограничений, то ее называют задачей на безусловный экстремум; при задаче с ограничениями – это условный экстремум критерия оптимальности. Таким образом, схема, представленная на рисунке 1.3, дает классификацию задач на условный экстремум.

ГЛАВА 2. ЗАДАЧИ ОТЫСКАНИЯ ЭКСТРЕМУМА ФУНКЦИЙ ОДНОЙ И МНОГИХ ПЕРЕМЕННЫХ. АНАЛИТИЧЕСКИЕ МЕТОДЫ РЕШЕНИЯ

2.1. Безусловный экстремум. Условия существования

Рассмотрим экстремальные задачи, в которых критерий оптимальности задан в виде скалярной функции $f(\bar{x})$ векторного аргумента $\bar{x}$: $\bar{x} = (x_1, x_2, \dots, x_n)$. В частном случае $n=1$ имеем функцию одной переменной. Пусть на аргумент $\bar{x}$ наложены ограничения, которые образуют множество x допустимых значений аргумента, тогда экстремальная задача формулируется следующим образом: найти условный экстремум функции $f(\bar{x})$ на множестве x .

$$\begin{aligned} & \text{extr } f(x) \\ & \bar{x} \in x \end{aligned} \quad (2.1)$$

Если на $\bar{x}$ никаких ограничений не наложено, то множество x совпадает с n -мерным евклидовым пространством $E_n (x = E_n)$, и задача (2.1) становится задачей на безусловный экстремум. Задача (2.1) является классической и рассматривается во всех курсах математического анализа [3, 4]. Сформулируем понятие решения задачи (2.1).

Определение 1. Функция $f(\bar{x})$ имеет в точке $\bar{x}^*$ глобальный экстремум, если для $\forall \bar{x} \in x$ выполняется неравенство:

$$f(\bar{x}) - f(\bar{x}^*) \geq 0 \text{ – в случае минимума;} \quad (2.2)$$

$$f(\bar{x}) - f(\bar{x}^*) \leq 0 \text{ – в случае максимума;} \quad (2.3)$$

Определение 2. Функция $f(\bar{x})$ имеет в точке $\bar{x}^*$ локальный экстремум, если при некотором достаточно малом $\varepsilon > 0$ для $\forall \bar{x} \in x$ удовлетворяющих условию

$$|\bar{x} - \bar{x}^*| \leq \varepsilon \quad (2.4)$$

выполняется неравенство (2.2) или (2.3).

Условие (2.4) означает, что точка $\bar{x}$ принадлежит достаточно малой ε - окрестности точки $\bar{x}^*$, которую можно представить в виде сферы радиуса ε . Иллюстрация определений 1 и 2 дана на рисунке 2.1 для функции одной переменной. На рисунке 2.1 функция $f(\bar{x})$ определена на отрезке $[a, b]$, который и представляет множество x допустимых значений аргумента x . Точки $\bar{x}_2^*$ и $\bar{x}_3^*$ – соответственно точки глобального минимума и максимума. В точке $\bar{x}_4^*$ достигается локальный минимум, в точках $\bar{x}_1^*$ и $\bar{x}_5^*$ – локальные максимумы – строгий и нестрогий.

Рисунок 2.2 иллюстрирует утверждение, известное под названием теоремы Вейерштрасса: непрерывная функция $f(x)$ достигает на замкнутом ограниченном множестве своего наибольшего и наименьшего значения (во внутренней или граничной точке).

Из курса математического анализа [3, 4] известно, что в точке безусловного экстремума функции $f(\bar{x})$ должны выполняться условия

- необходимые, позволяющие определить так называемые стационарные точки или точки, подозрительные на экстремум;
- достаточные, позволяющие дать окончательный ответ, какие из стационарных точек являются точками экстремума (минимума или максимума).

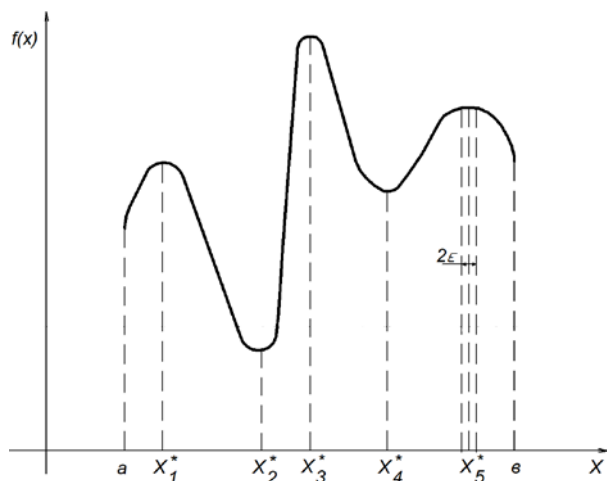


Рисунок 2.1. Экстремум функции $f(x)$

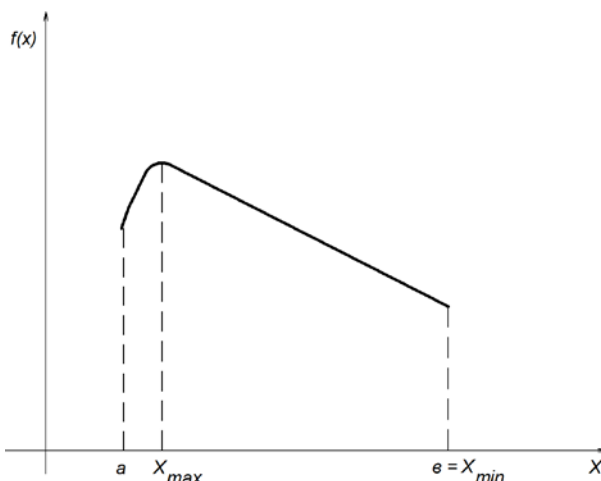


Рисунок 2.2. Иллюстрация теоремы Вейерштрасса

Необходимое условие (первого порядка) для функции одной переменной формулируется в виде теоремы Ферма: для того, чтобы непрерывная и имеющая непрерывные производные до второго порядка включительно функция $f(x)$ имела безусловный локальный экстремум в точке x^* , необходимо выполнение условия

$$\frac{df(x^*)}{dx} = 0. \quad (2.5)$$

Для функций многих переменных необходимое условие (2.5) записывается следующим образом: для того, чтобы функция $f(\bar{x})$ имела безусловный локальный экстремум в точке $\bar{x}^*$ необходимо, чтобы все ее частные производные первого порядка в этой точке обращались в нуль:

$$\frac{df(\bar{x}^*)}{dx_i} = 0, i = 1, \dots, n. \quad (2.6)$$

Для условия (2.6) часто используется и другая эквивалентная форма записи:

$$\text{grad}f(\bar{x}^*) = 0 \text{ или } f(\bar{x}) = 0, \quad (2.7)$$

где $\text{grad}f(\bar{x}^*)$ или $f(\bar{x})$ – градиент функции $f(\bar{x})$, представляющий n -мерный вектор с компонентами $\frac{df(\bar{x}^*)}{dx_i}$; геометрически градиент интерпретируется в виде нормали к поверхности функции $f(\bar{x})$ и указывает направление ее наискорейшего возрастания. По аналогии используется понятие антиградиента $(-\text{grad}f(\bar{x}))$, указывающего направление наискорейшего убывания $f(\bar{x})$.

Необходимое условие экстремума (2.7) эквивалентно равенству нулю дифференциала функции $f(\bar{x})$ в точке $\bar{x}$: $df(\bar{x}^*) = 0$.

Действительно, с учетом условия (2.7) имеем:

$$df(\bar{x}^*) = \sum_{i=1}^n \frac{df(\bar{x}^*)}{dx_i} \cdot dx_i = 0.$$

Таким образом, необходимые условия существования безусловного экстремума всегда приводят к уравнению (2.5) для функции одной переменной или к системе уравнений (2.7) для функции переменных, решение которых позволяет определить координаты стационарных точек. Если уравнения (2.5), (2.7) не могут быть решены аналитически, то координаты стационарных точек находят приближенно, пользуясь численными методами решения систем уравнений.

Чтобы среди стационарных точек найти точки экстремума, следует проверить, выполняются ли в этих точках достаточные условия экстремума.

Для функции одной переменной достаточными условиями, как известно, являются:

- изменения знака первой производной при прохождении через точку экстремума x^* (с плюса на минус в случае максимума и с минуса на плюс в случае минимума);
- выполнение условия (2.4) в ε -окрестности точки x^* ;
- знак второй производной $f''(x^*)$, (в точке минимума $f''(x^*) > 0$, в точке максимума $f''(x^*) < 0$).

Последнее условие рассмотрим более подробно. Пусть в окрестности стационарной точки x^* функция $f(x)$ имеет непрерывные производные до k порядка включительно. Разложим $f(x)$ в ряд Тейлора в окрестности точки x^* и определим знак приращения функции $\Delta f(x)$:

$$\Delta f(x) = f(x^* + \rho) - f(x^*) = \rho f'(x^*) + \frac{1}{2} \rho^2 f''(x^*) + \dots + \frac{1}{k!} \rho^k f^{(k)}(x^*) + o(\rho^k). \quad (2.8)$$

Согласно необходимому условию экстремума $f'(x^*) = 0$, и при любых достаточно малых ρ знак приращения $\Delta f(x)$ определяется знаком второй производной $f''(x^*)$, если $f''(x^*) \neq 0$. Если же $f''(x^*) = 0$, то знак приращения $\Delta f(x)$ при любых достаточно малых ρ определяется производной третьего порядка, и в случае $f'''(x^*) \neq 0$ стационарная точка x^* не является точкой экстремума $f(x)$. Действительно из (2.8) получим:

$$\Delta f(x) = f(x^* + \rho) - f(x^*) = \frac{1}{3!} \rho^3 f'''(x^*) + o(\rho^3),$$

откуда очевидно, что знак $\Delta f(x)$ будет меняться в зависимости от знака ρ и, следовательно, определение безусловного локального экстремума (2.4) нарушается. Обобщая сказанное на случай производных k порядка, можно сформулировать следующее достаточное условие экстремума функции одной переменной.

Пусть функция $f(x)$ имеет непрерывные производные до k порядка включительно и в некоторой точке x^*

$$f'(x^*) = f''(x^*) = \dots f^{k-1}(x^*) = 0, f^k(x^*) \neq 0,$$

тогда, если k – четное число, $f(x)$ имеет в точке x^* локальный экстремум – максимум или минимум в зависимости от знака $f^k(x^*)$, если же k – нечетное число, то в точке x^* экстремума нет.

Получили достаточное условие экстремума для функции многих переменных. Пусть $f(\bar{x})$ дважды непрерывно дифференцируемая функция и $\bar{x}^*$ – ее стационарная точка. Разложим $f(\bar{x})$ в ряд Тейлора в окрестности точки $\bar{x}^*$:

$$\Delta f(x) = f(\bar{x}^* + \bar{\rho}) - f(\bar{x}^*) = \sum_{i=1}^n \frac{\partial f(\bar{x}^*)}{\partial x_i} \bar{\rho}_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(\bar{x}^*)}{\partial x_i \partial x_j} \rho_i \rho_j + O(|\rho|^2). \quad (2.9)$$

Согласно необходимому условию оптимальности (2.7) линейные члены ряда $\frac{\partial f(\bar{x}^*)}{\partial x_i} = 0$ и знак приращения функции $\Delta f(x)$ при достаточно малом векторе ρ определяются суммой квадратических членов. Отсюда следует, что в точке $\bar{x}^*$ достигается безусловный локальный экстремум, если при любых достаточно малых $\bar{\rho}$ (кроме $\bar{\rho} = 0$) выполняется условие

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(\bar{x}^*)}{\partial x_i \partial x_j} \rho_i \rho_j &> 0 && \text{– для минимума;} \\ \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(\bar{x}^*)}{\partial x_i \partial x_j} \rho_i \rho_j &< 0 && \text{– для максимума.} \end{aligned} \quad (2.10)$$

Соотношение (2.10) представляет собой квадратичную форму, коэффициентами которой являются частные производные второго порядка функции $f(\bar{x})$, вычисленные в стационарной точке $\bar{x}^*$. Таким образом, чтобы в стационарной точке $\bar{x}^*$ функция $f(\bar{x})$ имела экстремум, квадратичная форма (2.10) должна быть положительно определена (в случае минимума) или отрицательно определена (в случае максимума). Проверка знакоопределенности квадратичной форма (2.10) может быть осуществлена с помощью критерия Сильвестра. Согласно этому критерию, проверяется знакоопределенность матрицы коэффициентов квадратичной формы (2.10), которая представляет собой матрицу вторых частных производных (матрица Гессе):

$$H = \begin{vmatrix} \frac{\partial^2 f(\bar{x}^*)}{\partial x_1^2} & \frac{\partial^2 f(\bar{x}^*)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\bar{x}^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\bar{x}^*)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\bar{x}^*)}{\partial x_2^2} & \dots & \frac{\partial^2 f(\bar{x}^*)}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f(\bar{x}^*)}{\partial x_n \partial x_1} & \frac{\partial^2 f(\bar{x}^*)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(\bar{x}^*)}{\partial x_n^2} \end{vmatrix} \quad (2.11)$$

Матрица Гессе, которую принято обозначать буквой H , положительно определена, если все ее главные миноры² строго положительны, т.е.

²

Минором k порядка называется определитель, элементы которого находятся на пересечении произвольных k строк и k столбцов матрицы H . Главные миноры расположены в верхнем левом углу матрицы H .

$$\Delta_1 = \frac{\partial^2 f(\bar{x}^*)}{\partial x_1^2} > 0; \Delta_2 = \begin{vmatrix} \frac{\partial^2 f(\bar{x}^*)}{\partial x_1^2} & \frac{\partial^2 f(\bar{x}^*)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(\bar{x}^*)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\bar{x}^*)}{\partial x_2^2} \end{vmatrix} > 0, \quad (2.12)$$

и тогда в точке $\bar{x}^*$ функция $f(\bar{x})$ имеет минимум.

Матрица Гессе отрицательно определена, если знаки ее главных миноров чередуются

$$\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \Delta_4 > 0, \quad (2.13)$$

и тогда в точке $\bar{x}^*$ функция $f(\bar{x})$ имеет максимум. Если условия (2.12) и (2.13) не выполняются, то в точке $\bar{x}^*$ экстремума нет, и имеется стационарная точка типа «седла». В случае равенства некоторых главных миноров нулю возникает неопределенность, требующая исследования знаков производных высших порядков, что неизмеримо труднее по сравнению с функцией одной переменной. Поэтому дополнительную информацию о характере экстремума стараются получить из физического смысла конкретной задачи.

2.2. Задача аналитической градуировки датчика

Среди прикладных задач оптимизации задачи на безусловный экстремум встречаются чрезвычайно редко. Как правило, они представляют собой лишь составляющую часть или этап решения более сложной экстремальной задачи.

В качестве примера рассмотрим задачу аналитической градуировки датчика, которая относится к задачам идентификации и обязательно решается в подсистеме централизованного контроля АСУТП. Для большей конкретности речь пойдет об аналитической градуировке датчика концентрации волокнистых суспензий. Как известно, эти датчики не имеют класса точности и служат главным образом индикаторами, т.е. указывают тенденцию изменения концентрации, не давая точной количественной оценки. В то же время информация о величине концентрации целлюлозной, бумажной, древесной массы необходима в АСУТП для решения задач управления технологическим процессом. Поэтому при внедрении АСУТП осуществляют градуировку датчиков концентрации по данным лабораторных анализов, т.е. находят зависимость

$$C = f(m), \quad (2.14)$$

где C – концентрация массы, %; m – сигнал на выходе датчика, мА.

Пусть для градуировки датчика проведена серия лабораторных измерений из 5 экспериментов (табл. 2.1). Из графика на рисунке 2.3 видно, что зависимость (2.14) носит линейный характер и, следовательно, математическая модель датчика может быть принята в виде:

$$C_p = b_0 + b_1 m_1, \quad (2.15)$$

где C_p – расчетное значение концентрации массы.

Таблица 2.1. Экспериментальные и расчетные данные C , m

№ п/п	m , мА	$C_{э}$, %	C_p , %
1	1,0	2,0	1,97
2	2,5	2,3	2,34
3	3,5	2,65	2,57
4	4,2	2,70	2,77
5	4,8	3,0	2,92

Требуется найти коэффициенты уравнения (2.15) так, чтобы обеспечить минимальное расхождение между расчетными (C_p) и экспериментальными ($C_{э}$) значениями концентрации.

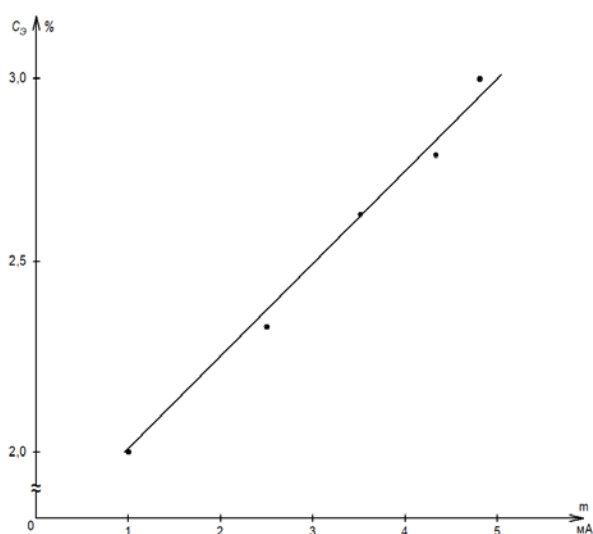


Рисунок 2.3. Градуировочная зависимость датчика концентрации

Определение коэффициентов математической модели (2.15) осуществляется обычно методом наименьших квадратов, обоснование и изложение которого не входит в нашу задачу. Рассмотрим лишь фрагмент этого метода, связанный с исследованием на экстремум функции многих переменных.

В качестве критерия оптимальности в задаче идентификации (2.15) принимают функцию вида

$$f = \sum_{i=1}^n (C_{pi} - C_{эi})^2. \quad (2.16)$$

С учетом (2.15) получим:

$$f = \sum_{i=1}^N (b_0 + b_1 m_i - C_{эi})^2. \quad (2.17)$$

Найдем коэффициенты b_0 и b_1 из условия минимума функции (2.17). Необходимые условия оптимальности функции $f(b_0, b_1)$ приводят к системе уравнений:

$$\begin{aligned} \frac{df}{db_0} &= 2 \sum_{i=1}^N (b_0 + b_1 m_i - C_{эi}) = 0 \\ \frac{df}{db_1} &= 2 \sum_{i=1}^N (b_0 + b_1 m_i - C_{эi}) m_i = 0 \end{aligned} \quad (2.18)$$

из решения которой получим:

$$b_0 = \frac{\sum_{i=1}^N C_{\text{э}i} \sum_{i=1}^N m_i^2 - \sum_{i=1}^N m_i \sum_{i=1}^N C_{\text{э}i} m_i}{N \sum_{i=1}^N m_i^2 - (\sum_{i=1}^N m_i)^2};$$

$$b_1 = \frac{\sum_{i=1}^N C_{\text{э}i} m_i - \sum_{i=1}^N C_{\text{э}i} \sum_{i=1}^N m_i}{N \sum_{i=1}^N m_i^2 - (\sum_{i=1}^N m_i)^2}.$$

Для данных в табл. 2.1 коэффициенты принимают значения $b_0 = 1,72$; $b_1 = 0,25$. Коэффициенты b_0 и b_1 , вычисленные из системы уравнений (2.18), определяют координаты стационарной точки $b^* = (b_0, b_1)$ функции $f(\bar{b})$. Для этого найдем частные производные второго порядка функции $f(\bar{b})$ и составим матрицу Гессе:

$$M = \begin{vmatrix} \frac{\partial^2 f}{\partial b_0^2} & \frac{\partial^2 f}{\partial b_0 \partial b_1} \\ \frac{\partial^2 f}{\partial b_1 \partial b_0} & \frac{\partial^2 f}{\partial b_1^2} \end{vmatrix} = \begin{vmatrix} N & \sum_{i=1}^N m_i \\ \sum_{i=1}^N m_i & \sum_{i=1}^N m_i^2 \end{vmatrix}$$

Определим знаки главных миноров:

$$\Delta_1 = N = 5 > 0, \Delta_2 = N \sum_{i=1}^N m_i^2 - \left(\sum_{i=1}^N m_i \right)^2 = 44,9 > 0$$

Проверка достаточных условий оптимальности подтверждает, что в точке $b^* = (b_0, b_1) = (1,72; 0,25)$ достигается минимум функции f .

2.3. Задачи на условный экстремум с ограничениями-равенствами. Метод множителей Лагранжа. Условия существования экстремума в задаче с ограничениями

Рассмотрим задачу на условный экстремум с ограничениями-равенствами. Математическая модель такой задачи в общем виде записывается следующим образом:

$$\text{найти } \text{extr } f(\bar{x}) \quad (2.19)$$

$$\text{при условии } g_i(\bar{x}) = 0, i = 1, \dots, m. \quad (2.20)$$

Ограничения $g_i(\bar{x}) = 0$ принято называть уравнениями связи. Чтобы задача оптимизации (2.19) имела смысл, должно выполняться условие $m < n$, где n – размерность вектора $\bar{x}$; в этом случае система уравнений (2.20) имеет бесчисленное множество решений. Дадим геометрическую интерпретацию задачи на примере функции двух переменных ($n=2$), тогда $m=1$.

Пример 2.1. Найти $\text{extr } f = (x_1 - 1)^2 + (x_2 - 3)^2$

при условии $g(\bar{x}) = x_1^2 + x_2^2 - 4 = 0$.

Ограничение $g(\bar{x})$ в рассматриваемом примере представляет собой окружность радиуса $R=2$ с центром в начале координат (рис. 2.4).

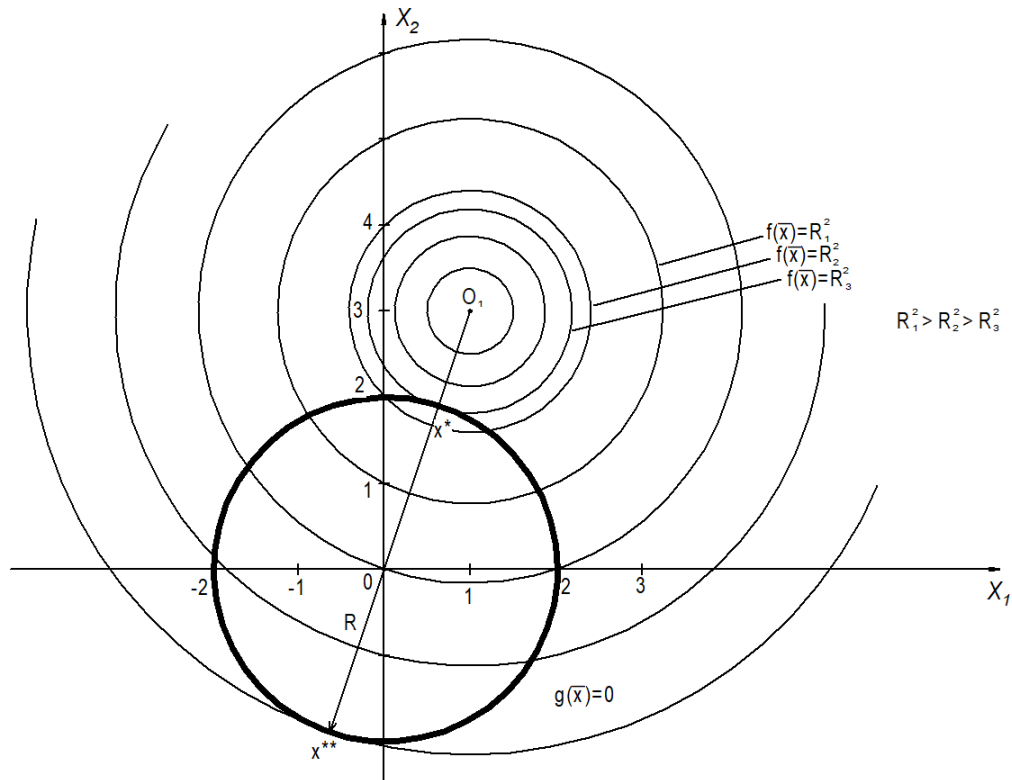


Рисунок 2.4. Геометрическая интерпретация задачи на условный экстремум

Таким образом, множество допустимых решений задачи – это множество точек, лежащих на окружности $g(\bar{x}) = 0$. Через любую из точек окружности $g(\bar{x}) = 0$ проходит линия уровня функции $f(\bar{x})$. На рисунке 2.4 показано, что линии уровня $f(\bar{x})$ – это семейство концентрических окружностей с центром в точке O_1 . Каждой точке множества допустимых решений соответствует вполне определенное значение функции $f(\bar{x})$, определяемое радиусом окружности $f(\bar{x}) = R_{\bar{x}}^2$, проходящей через эту точку. Сущность задачи на условный экстремум состоит в том, чтобы среди множества допустимых точек, т.е. точек, удовлетворяющих $g(\bar{x}) = 0$, найти такую точку или такие точки, в которых функция $f(\bar{x})$ достигала бы наибольшего значения. Из рисунка 2.4 видно, что это точки $x^*(0,63; 1,9)$ и $x^{**}(-0,63; -1,9)$; причем первая доставляет условный минимум $f(\bar{x})$ ($f(\bar{x}^*) = 1,35$), а вторая – условный максимум ($f(\bar{x}^{**}) = 26,7$). Для сравнения полезно отметить, что безусловный минимум $f(\bar{x})$ достигается в точке O_1 , в которой $f(O_1) = 0$, а безусловного максимума не существует.

Определение 3. Допустимая точка $\bar{x}^*$ доставляет условный глобальный экстремум функции $f(\bar{x})$, если для всех $\bar{x}$, удовлетворяющих уравнениям связи $g(\bar{x}) = 0, i = 1, \dots, t$, имеет место неравенство

$$f(\bar{x}) - f(\bar{x}^*) \geq 0 \text{ – для минимума,}$$

$$f(\bar{x}) - f(\bar{x}^*) \leq 0 \text{ – для максимума.}$$

Определение 4. Допустимая точка $\bar{x}^*$ доставляет условный локальный экстремум функции $f(\bar{x})$, если можно указать такое число $\varepsilon > 0$, что для всех $\bar{x}$, удовлетворяющих уравнениям связи $g(\bar{x}) = 0$, $i = 1, \dots, t$, и условию $|\bar{x} - \bar{x}^*| < \varepsilon$, имеет место неравенство

$$f(\bar{x}) - f(\bar{x}^*) \geq 0 \text{ или} \\ f(\bar{x}) - f(\bar{x}^*) \leq 0.$$

Получим необходимые условия оптимальности для задачи с ограничениями-равенствами. Строгое обоснование необходимых условий можно найти в литературе [3–5]. В настоящем учебном пособии для большей наглядности выведем необходимые условия из геометрической интерпретации задачи.

В общем виде геометрическая интерпретация задачи на условный экстремум дана на рисунке 2.5, где на плоскости чертежа нанесены линии уровня функции $f(\bar{x})$ и линия ограничения $g(\bar{x}) = 0$. Из рисунка 2.5 очевидно (это же следует и из рисунка 2.4), что условный локальный экстремум функции $f(\bar{x})$ реализуется только в точках, где линии уровня $f(\bar{x})$ касаются кривой $g(\bar{x}) = 0$ и, следовательно, в этих точках градиенты $f(\bar{x})$ и $g(\bar{x})$ направлены по одной прямой. Рассмотрим треугольники, образованные компонентами $\nabla f(\bar{x}^*)$ и $\nabla g(\bar{x}^*)$. Из подобия треугольников следует, что

$$\frac{\frac{\partial f(\bar{x}^*)}{\partial x_1}}{\frac{\partial g(\bar{x}^*)}{\partial x_1}} = \frac{\frac{\partial f(\bar{x}^*)}{\partial x_2}}{\frac{\partial g(\bar{x}^*)}{\partial x_2}}. \quad (2.21)$$

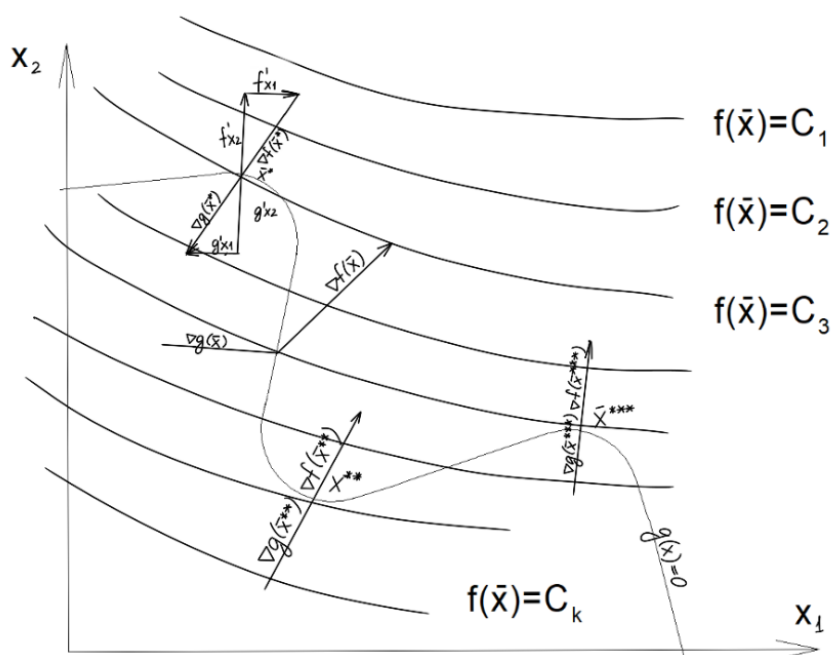


Рисунок 2.5. Геометрическая интерпретация необходимых условий оптимальности

Обозначим отношения (2.21) – λ , тогда получим следующие уравнения:

$$\begin{aligned}\frac{\partial f(\bar{x}^*)}{\partial x_1} + \lambda \frac{\partial g(\bar{x}^*)}{\partial x_1} &= 0, \\ \frac{\partial f(\bar{x}^*)}{\partial x_2} + \lambda \frac{\partial g(\bar{x}^*)}{\partial x_2} &= 0.\end{aligned}\quad (2.22)$$

Совместно с уравнением связи уравнения $g(\bar{x}) = 0$ (2.22) выражают необходимые условия оптимальности в задаче с ограничениями-равенствами.

Формально эти же условия могут быть получены, если ввести в рассмотрение функцию вида:

$$(\bar{x}, \bar{\lambda}) = f(\bar{x}) + \sum_{i=1}^m \lambda_i g_i(\bar{x}). \quad (2.23)$$

найти ее частные производные по переменным $\bar{x}$ и $\bar{\lambda}$ и приравнять их нулю:

$$\begin{aligned}\frac{\partial L(\bar{x}, \bar{\lambda})}{\partial x_j} &= \frac{\partial f(\bar{x})}{\partial x_j} + \sum_{i=1}^m \lambda_i \frac{\partial g_i(\bar{x})}{\partial x_j}, \quad j = 1, \dots, n, \\ \frac{\partial L(\bar{x}, \bar{\lambda})}{\partial x_i} &= g_i(\bar{x}) = 0, \quad i = 1, \dots, m.\end{aligned}\quad (2.24)$$

Функция $L(\bar{x}, \bar{\lambda})$ называется функцией Лагранжа и зависит от $n+m$ переменных; формальные переменные λ_i носят название неопределенных множителей Лагранжа – их число равно числу уравнений связи; метод, позволяющий вывести необходимые условия оптимальности с помощью функции Лагранжа, получил название метода множителей Лагранжа. По существу, этим методом задача (2.19), (2.20) на условный экстремум функции $f(\bar{x})$ сводится к задаче на безусловный экстремум функции Лагранжа (2.24), число переменных при этом увеличивается за счет введения m множителей Лагранжа.

Решим методом множителей Лагранжа задачу, геометрическая интерпретация которой дана на рисунке 2.4. Составим функцию Лагранжа:

$$L(\bar{x}, \bar{\lambda}) = f(\bar{x}) + \lambda g(\bar{x}) = (x_1 - 1)^2 + (x_2 - 3)^2 + \lambda(x_1^2 + x_2^2 - 4), \quad (2.25)$$

найдем ее частные производные, и приравняем их нулю:

$$\begin{aligned}\frac{\partial L}{\partial x_1} &= 2(x_1 - 1) + 2\lambda x_1 = 0, \\ \frac{\partial L}{\partial x_2} &= 2(x_2 - 3) + 2\lambda x_2 = 0, \\ \frac{\partial L}{\partial \lambda} &= x_1^2 + x_2^2 - 4 = 0.\end{aligned}$$

Решив систему уравнений, выражающих необходимые условия оптимальности, найдем координаты стационарных точек:

$$(\bar{x}^*, \bar{\lambda}^*) = (0,63; 1,9; 0,58); (\bar{x}^{**}, \bar{\lambda}^{**}) = (-0,63; -1,9; -2,58).$$

Чтобы определить, достигается ли экстремум в стационарных точках, следует проверить выполнение в этих точках достаточных условий оптимальности, что требует анализа вторых частных производных. Рассмотрим приращение функции $\Delta f(\bar{x}) = f(\bar{x}^* + \bar{\rho}) - f(\bar{x}^*)$, причем все точки $\bar{x} = \bar{x}^* + \bar{\rho}$, для которых вычисляется приращение, должны согласно определению 4 быть допустимыми, т.е. удовлетворять уравнениям связи, или, что то же самое, лежать на кривой $g(\bar{x}) = 0$ в ε -окрестности точки $\bar{x}^*$ (см. рис. 2.5).

Тогда приращение $\Delta f(\bar{x})$ можно заменить приращением функции Лагранжа $\Delta L(\bar{x}, \bar{\lambda}^*)$:

$$\Delta f(\bar{x}) = f(\bar{x}^* + \bar{\rho}) - f(\bar{x}^*) = L(\bar{x}^* + \bar{\rho}, \bar{\lambda}^*) - L(\bar{x}^*, \bar{\lambda}^*).$$

Разложим $L(\bar{x}^* + \bar{\rho}, \bar{\lambda}^*)$ в ряд Тейлора в окрестности точки $(\bar{x}^*, \bar{\lambda}^*)$

$$\begin{aligned} f(\bar{x}) = L(\bar{x}^* + \bar{\rho}, \bar{\lambda}^*) - L(\bar{x}^*, \bar{\lambda}^*) &= \sum_{j=1}^n \frac{\partial L(\bar{x}^*, \bar{\lambda}^*)}{\partial x_j} \rho_j + \\ &+ \frac{1}{2} \sum_{k=1}^n \sum_{j=1}^n \frac{\partial^2 L(\bar{x}^*, \bar{\lambda}^*)}{\partial x_k \partial x_j} \rho_k \rho_j + 0(|\bar{\rho}|^2). \end{aligned} \quad (2.26)$$

Очевидно, что первое слагаемое в правой части (2.26) равно нулю в силу необходимых условий оптимальности (2.24), и, следовательно, знак приращения $\Delta f(\bar{x})$ при достаточно малом $\bar{\rho}$ определяется знаком квадратичной формы

$$\Phi = \sum_{k=1}^n \sum_{j=1}^n \frac{\partial^2 L(\bar{x}^*, \bar{\lambda}^*)}{\partial x_k \partial x_j} \rho_k \rho_j, \quad (2.27)$$

а так как точка $(\bar{x}^* + \bar{\rho})$ должна быть допустимой, то координаты вектора смещения $\bar{\rho}$ не могут все задаваться произвольно, часть из них, а именно m должна вычисляться из условия

$$g_i(\bar{x}^* + \bar{\rho}) = 0, i = 1, \dots, m. \quad (2.28)$$

Разложим соотношения (2.28) в ряд Тейлора в окрестности точки $\bar{x}^*$:

$$g_i(\bar{x}^* + \bar{\rho}) = g_i(\bar{x}^*) + \sum_{j=1}^n \frac{\partial g_i(\bar{x}^*)}{\partial x_j} \rho_j + 0(|\bar{\rho}|) = 0.$$

Пренебрегая членами разложения более высокого порядка малости при достаточно малых $\bar{\rho}$, в линейном приближении получим:

$$\sum_{j=1}^n \frac{\partial g_i(\bar{x}^*)}{\partial x_j} \rho_j = 0, i = 1, \dots, m. \quad (2.29)$$

Каждое из линейных уравнений системы (2.29) геометрически определяет плоскость, касательную к поверхности $g_i(\bar{x})$ в точке $\bar{x}^*$, на рисунке 2.5 это прямая, касательная к кривой $g(\bar{x})$. Разрешив систему уравнений (2.29) относительно произвольных m зависимых переменных ρ_j , например, для

$j=1, \dots, m$, т.е. выразив зависимые переменные через независимые $n-m$ переменных, получим:

$$\rho_j = f_j(\rho_{n-m-1}, \dots, \rho_n), j = 1, \dots, m. \quad (2.30)$$

Подставив (2.30) в (2.27) и определив знак квадратичной формы Φ по правилу Сильвестра, установим, достигается ли в точке $\bar{x}^*$ экстремум. Проиллюстрируем изложенное выше на уже рассмотренном примере (см. рис. 2.4).

Найдем частные производные второго порядка функции Лагранжа (2.25):

$$\frac{\partial^2 L}{\partial x_1^2} = 2 + 2\lambda; \quad \frac{\partial^2 L}{\partial x_1 \partial x_2} = 0; \quad \frac{\partial^2 L}{\partial x_2^2} = 2 + 2\lambda$$

и составим квадратичную форму Φ для стационарной точки $(\bar{x}^*, \bar{\lambda}^*) = (0,63; 1,9; 0,58)$, вычислив в ней значения вторых частных производных:

$$\Phi(\bar{x}^*, \lambda^*) = \sum_{k=1}^n \sum_{j=1}^n \frac{\partial^2 L(\bar{x}^*, \lambda^*)}{\partial x_k \partial x_j} \rho_k \rho_j = 3,16\rho_1^2 + 3,16\rho_2^2. \quad (2.31)$$

Получим уравнение (2.29), разложив в ряд Тейлора уравнение связи $g(\bar{x}) = 0$:

$$\sum_{j=1}^2 \frac{\partial g_i(\bar{x}^*)}{\partial x_j} \rho_j = 2x_1^* \rho_1 + 2x_2^* \rho_2 = 1,26\rho_1 + 3,8\rho_2 = 0,$$

откуда найдем $\rho_2 = -0,33\rho_1$ и подставим в квадратичную форму Φ (2.31):

$$\Phi(\bar{x}^*, \lambda^*) = 3,16\rho_1^2 + 3,16(-0,33\rho_1)^2 = 3,48\rho_1^2.$$

Очевидно, что $(\bar{x}^*, \lambda^*) > 0$ при любых ρ_1 и, следовательно, в точке $\bar{x}^*$ функция $f(\bar{x})$ имеет минимум.

Убедиться, что в точке $\bar{x}^{**}$ достигается максимум $f(\bar{x})$, нетрудно.

Таким образом, сущность достаточных условий оптимальности в задачах с ограничениями-равенствами та же, что и в задачах на безусловный экстремум, с той лишь особенностью, что проверку экстремума следует проводить среди допустимых точек, т.е. с учетом уравнений связи $g(\bar{x}) = 0$.

2.4. Примеры задач на условный экстремум с ограничениями-равенствами

Задача оптимального проектирования очистных сооружений

Выбрать размеры N аэротенков для станции биологической очистки сточных вод, объемом V м³ каждый, соединенных между собой так, как показано на рисунке 2.6, из условия минимальных затрат материалов на их сооружение.

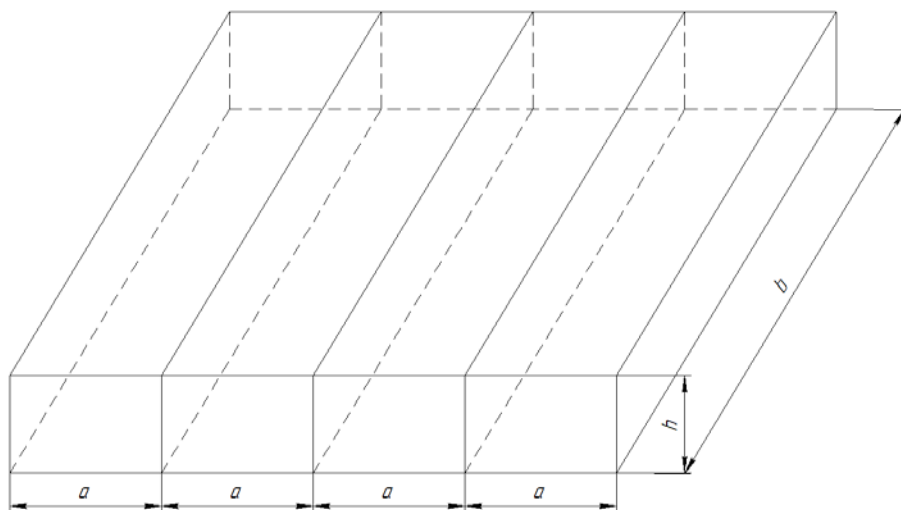


Рисунок 2.6. Схема очистных сооружений

Очевидно, что в качестве критерия оптимальности в этой задаче целесообразно выбрать общую поверхность аэротенков

$$S = Nab + 2Nah + (N + 1)bh.$$

и найти параметры a , b , h из условия минимума S при ограничении

$$g = abh - V = 0.$$

Решим эту задачу методом множителей Лагранжа. Составим функцию Лагранжа:

$$L(a, b, h, \lambda) = Nab + 2Nah + (N + 1)bh + \lambda(abh - V).$$

и получим необходимые условия оптимальности:

$$\begin{aligned} \frac{\partial L}{\partial a} &= Nb + 2Nh + 2bh = 0, \\ \frac{\partial L}{\partial b} &= Na + (N + 1)h + \lambda ah = 0, \\ \frac{\partial L}{\partial h} &= 2Na + (N + 1)b + \lambda ab = 0, \\ \frac{\partial L}{\partial \lambda} &= abh - V = 0. \end{aligned}$$

Решив систему уравнений, найдем:

$$a^* = \frac{N + 1}{2N} \sqrt[3]{\frac{4NV}{N + 1}}; \quad b^* = \sqrt[3]{\frac{4NV}{N + 1}}; \quad h^* = \frac{1}{2} \sqrt[3]{\frac{4NV}{N + 1}}; \quad \lambda = -4N \sqrt[3]{\frac{N + 1}{NV}}.$$

Убедимся, действительно ли параметры (a^*, b^*, h^*) доставляют минимум S . Для этого сформулируем достаточные условия оптимальности. Найдем частные

производные второго порядка функции Лагранжа, вычислим их в стационарной точке и составим квадратичную форму Φ :

$$\frac{\partial^2 L}{\partial^2 a} = 0; \frac{\partial^2 L}{\partial^2 b} = 0; \frac{\partial^2 L}{\partial^2 h} = 0; \frac{\partial^2 L}{\partial a \partial b} = N + \lambda h; \frac{\partial^2 L}{\partial a \partial h} = 2N + \lambda b; \frac{\partial^2 L}{\partial b \partial h} = (N + \lambda) + \lambda a;$$

$$\frac{\partial^2 L(a^*, b^*, h^*, \lambda^*)}{\partial a \partial b} = -N; \frac{\partial^2 L(a^*, b^*, h^*, \lambda^*)}{\partial a \partial h} = -2N; \frac{\partial^2 L(a^*, b^*, h^*, \lambda^*)}{\partial b \partial h} = -(N + 1);$$

$$\Phi = -N\Delta a\Delta b - 2N\Delta a\Delta h - (N + 1)\Delta b\Delta h,$$

где $\Delta a, \Delta b, \Delta h$ – малые приращения параметров a, b, h относительно значений a^*, b^*, h^* .

Рассмотрим линейную часть разложения в ряд Тейлора уравнения связи $g = 0$:

$$b^* h^* \Delta a + a^* h^* \Delta b + a^* b^* \Delta h = 0,$$

откуда

$$\Delta h = -\frac{b^* h^* \Delta a + a^* h^* \Delta b}{a^* b^*} = -\frac{N}{N + 1} \Delta a - \frac{1}{2} \Delta b.$$

Подставив выражение для Δh в квадратичную форму Φ , получим:

$$\Phi = \frac{2N^2}{N + 1} \Delta a^2 + N\Delta a\Delta b + \frac{N + 1}{2} \Delta b^2.$$

Составим матрицу коэффициентов квадратичной формы Φ и применим правило Сильвестра:

$$M = \begin{vmatrix} \frac{2N^2}{N + 1} & \frac{1}{2}N \\ \frac{1}{2}N & \frac{N + 1}{2} \end{vmatrix}; \quad \Delta_1 = \frac{2N^2}{N + 1} > 0$$

$$\Delta_2 = \frac{2N^2}{N + 1} \cdot \frac{N + 1}{2} - \frac{1}{4}N^2 = \frac{3}{4}N^2 > 0.$$

Следовательно, размеры аэротенков a^*, b^*, h^* обеспечивают минимальную поверхность при заданном объеме.

Из проделанного расчета видно, что оптимальные размеры аэротенков зависят от их числа.

Задача повышения достоверности измерительной информации

В общем виде эта задача сформулирована и решена в [6]. Сущность задачи состоит в том, что результаты измерений параметров на каком-либо участке производства корректируются на основе балансовых соотношений между ними и скорректированные величины принимаются за фактические значения параметров. Использование балансовых соотношений позволяет повысить достоверность информации, если только результаты измерений не содержат грубых ошибок. Алгоритм задачи входит в состав первичной обработки информации подсистемы централизованного контроля АСУТП.

Рассмотрим решение этой задачи на простейшем примере.

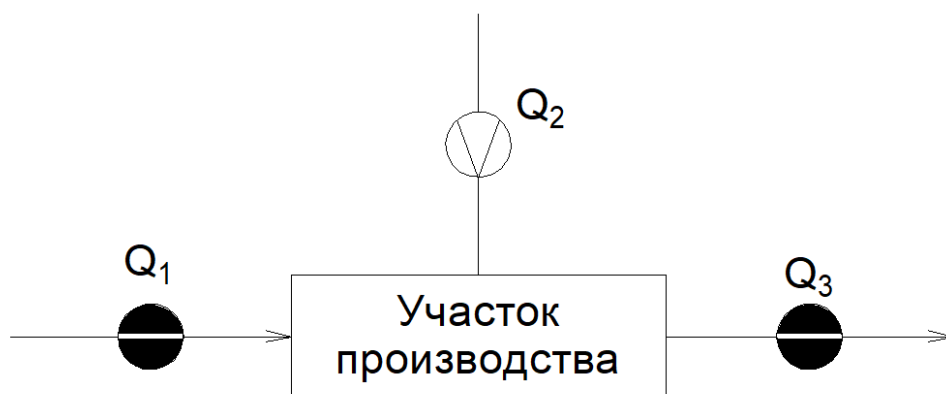


Рисунок 2.7. Схема материальных потоков

На рисунке 2.7 показана схема материальных потоков участка производства, на каждом потоке установлены расходомеры. Показания расходомеров обозначим соответственно Q_1^* , Q_2^* , Q_3^* и примем, что запас вещества на участке не меняется, тогда справедливо уравнение материального баланса для установившегося режима:

$$Q_1 + Q_2 - Q_3 = 0. \quad (2.32)$$

Однако из-за погрешностей измерения при подстановке значений Q_1^* , Q_2^* , Q_3^* в уравнение (2.32) получим

$$Q_1^* + Q_2^* - Q_3^* = \Delta,$$

где Δ – «невязка» уравнения материального баланса.

Пусть среднеквадратические погрешности измерений известны: обозначим их соответственно σ_1 , σ_2 , σ_3 .

Сформулируем следующую экстремальную задачу: найти скорректированные значения расходов Q_1 , Q_2 , Q_3 из условия минимума критерия оптимальности

$$f = \sum_{i=1}^3 \frac{1}{\sigma_i^2} (Q_i - Q_i^*)^2 \rightarrow \min \quad (2.33)$$

с учетом уравнения материального баланса (2.32). Очевидно, что (2.32), (2.33) – задача на полный экстремум. Решим ее методом множителей Лагранжа.

Запишем функцию Лагранжа:

$$L(\bar{Q}, \lambda) = \sum_{i=1}^3 \frac{1}{\sigma_i^2} (Q_i - Q_i^*)^2 + \lambda(Q_1 + Q_2 - Q_3).$$

Сформулируем необходимые условия оптимальности:

$$\frac{\partial L}{\partial Q_i} = \frac{2}{\sigma_i^2} (Q_i - Q_i^*) \pm \lambda = 0, i = 1, 2, 3;$$

$$\frac{\partial L}{\partial \lambda} = Q_1 + Q_2 - Q_3 = 0.$$

Решив полученную систему из четырех уравнений с четырьмя неизвестными, найдем координаты стационарной точки:

$$\lambda = \frac{2\Delta}{\sum_{i=1}^3 \frac{1}{\sigma_i^2}}; Q_i = Q_i^* \pm \frac{2\Delta}{\sum_{j=1}^3 \frac{\sigma_i^2}{\sigma_j^2}} \Delta. \quad (2.34)$$

В формуле (2.34) для расчета Q_i знак корректирующего члена противоположен знаку Q_i в уравнении материального баланса (2.32). Убедимся в том, что решение (2.34) действительно доставляет условный минимум функции (2.33).

Проверим для этого выполнение достаточных условий оптимальности.

Найдем частные производные второго порядка функции Лагранжа:

$$\frac{\partial^2 L}{\partial Q_i^2} = \frac{2}{\sigma_i^2}; \frac{\partial^2 L}{\partial Q_i \partial Q_j} = 0.$$

Запишем уравнение (2.32) в приращениях:

$$\Delta Q_1 + \Delta Q_2 - \Delta Q_3 = 0$$

и выразим из него ΔQ_3 :

$$\Delta Q_3 = \Delta Q_1 + \Delta Q_2. \quad (2.35)$$

Составим квадратичную форму:

$$\Phi = \sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial^2 L}{\partial Q_i \partial Q_j} \Delta Q_i \Delta Q_j = \frac{2}{\sigma_1^2} \Delta Q_1^2 + \frac{2}{\sigma_2^2} \Delta Q_2^2 + \frac{2}{\sigma_3^2} \Delta Q_3^2,$$

с учетом (2.35) получим:

$$\Phi = \left(\frac{2}{\sigma_1^2} + \frac{2}{\sigma_3^2} \right) \Delta Q_1^2 + \left(\frac{2}{\sigma_2^2} + \frac{2}{\sigma_3^2} \right) \Delta Q_2^2 + \frac{4}{\sigma_3^2} \Delta Q_1 \Delta Q_2.$$

Матрица Гессе для квадратичной формы Φ имеет вид:

$$H = \begin{pmatrix} \frac{2}{\sigma_1^2} + \frac{2}{\sigma_3^2} & \frac{2}{\sigma_3^2} \\ \frac{2}{\sigma_3^2} & \frac{2}{\sigma_2^2} + \frac{2}{\sigma_3^2} \end{pmatrix},$$

Очевидно, что главные миноры матрицы H

$$\Delta_1 = \frac{2}{\sigma_1^2} + \frac{2}{\sigma_3^2} > 0,$$

$$\Delta_2 = \frac{4}{\sigma_1^2 \sigma_2^2} + \frac{4}{\sigma_3^2} \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) > 0.$$

Следовательно, в стационарной точке (2.34) функция (2.33) имеет условный минимум. Из соотношений (2.34) видно, что корректировка каждого

параметра осуществляется с учетом относительного вклада его погрешности в суммарную погрешность всех измерений.

2.5. Задачи с ограничениями смешанного типа. Общий алгоритм решения

Прикладные задачи оптимизации, как правило, содержат ограничения смешанного типа, что значительно усложняет процедуру решения задачи. В общем виде задача со смешанной системой ограничений записывается следующим образом [7]:

Найти минимум (максимум)

$$f(\bar{x}), \bar{x} = (\bar{x}_1, \dots, \bar{x}_n); \quad (2.36)$$

при ограничениях

$$h_i(\bar{x}) = 0, i = 1, \dots, m, \quad (2.37)$$

$$g_i(\bar{x}) \geq 0, i = m + 1, \dots, r. \quad (2.38)$$

Число ограничений неравенств может быть любым.

Любая точка $\bar{x} \in E^n$, удовлетворяющая ограничениям в виде равенств (2.37) и в виде неравенств (2.38), будет допустимым решением задачи (2.36) – (2.38). Множество всех допустимых точек образует в E^n область допустимых решений (ОДР). Условный экстремум функции $f(\bar{x})$ – это точка $\bar{x}^*$ из ОДР, лежащая на пересечении всех гиперповерхностей, соответствующих $h_i(\bar{x}) = 0$. По отношению к ограничениям $g_i(\bar{x}) \geq 0$ точка $\bar{x}^*$ может быть либо внутренней, либо граничной. Если точка $\bar{x}^*$ оказывается граничной, т.е. какое-то неравенство выполняется как строгое равенство $g_i(\bar{x}^*) = 0$, то это неравенство называется активным.

Для иллюстрации изложенного рассмотрим несколько простейших примеров.

Пример 2.2. Минимизировать $f(\bar{x}) = x_1^2 + x_2$

при ограничениях $g_1(\bar{x}) = -(x_1^2 + x_2^2) + 9 \geq 0$,

$$g_2(\bar{x}) = -x_1 - x_2 + 1 \geq 0,$$

$$h_1(\bar{x}) = x_1 - x_2 - 2 = 0.$$

ОДР задачи – это множество точек, принадлежащих отрезку АВ, как показано на рисунке 2.8.

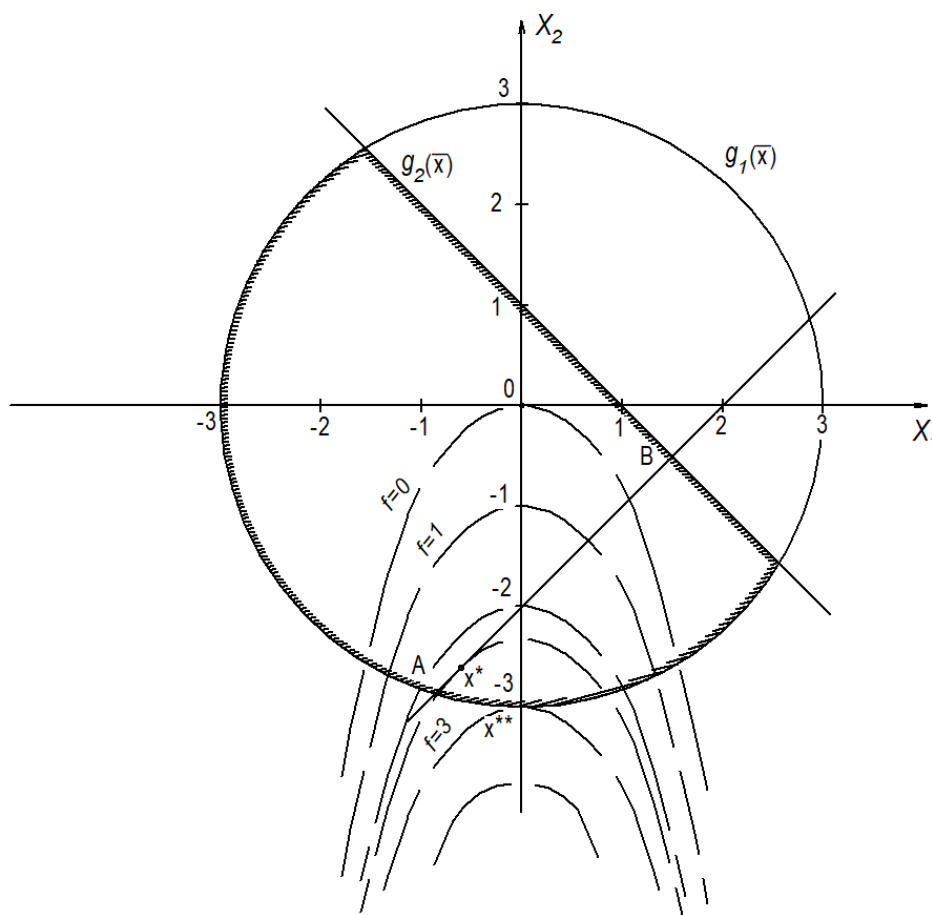


Рисунок 2.8. Геометрическая интерпретация задачи к примерам 2.2 и 2.3

Линии уровня функции $f(\bar{x})$ – семейство парабол (см. рисунок 2.8). Точка условного минимума $f(\bar{x})$ имеет координаты $\bar{x}^* = (-0,5; -2,5)$, при этом $f(\bar{x}^*) = -2,25$. По отношению к ограничениям $g_i(\bar{x})$ точка $\bar{x}^*$ является внутренней.

Пример 2.3. Изменим несколько условия задачи в примере 2.2. Пусть в систему ограничений входят только неравенства $g_1(\bar{x})$ и $g_2(\bar{x})$. ОДР в этом случае ограничена частью окружности $g_1(\bar{x}) = 0$ и отрезком прямой $g_2(\bar{x}) = 0$, на рисунке 2.8 ОДР заштрихована. Условный минимум $f(\bar{x})$ достигается на границе ОДР в точке с координатами $\bar{x}^{**} = (0; -3)$; $f(\bar{x}^{**}) = -3$. При этом ограничение $g_1(\bar{x}) = 0$ выполняется как строгое равенство, т.е. является активным.

Необходимые условия существования условного экстремума $f(\bar{x})$ в задаче (2.36) – (2.38) могут быть сформулированы в виде теоремы.

Теорема. Если вектор $\bar{x}^*$ доставляет условный минимум $f(\bar{x})$ и функции $f(\bar{x}), h_i(\bar{x}), g_i(\bar{x})$ дифференцируемы, то существуют векторы $\bar{u}^*, \bar{\lambda}^*$, благодаря которым выполняются соотношения:

$$\begin{aligned}
h_i(\bar{x}^*) &= 0, \quad i = 1, \dots, m, \\
g_i(\bar{x}^*) &\geq 0, \quad i = m + 1, \dots, r, \\
\bar{u}_i^* g_i(\bar{x}^*) &= 0, \quad i = m + 1, \dots, r, \\
\bar{u}_i^* &\geq 0, \quad i = m + 1, \dots, r, \\
\nabla L(\bar{x}^*, \bar{u}^*, \bar{\lambda}^*) &= 0.
\end{aligned} \tag{2.39}$$

В соотношении (2.39) функция L равна

$$f(\bar{x}) + \sum_{i=1}^m \lambda_i h_i(\bar{x}) - \sum_{i=m+1}^r u_i g_i(\bar{x}) \tag{2.40}$$

и рассматривается как обобщенная функция Лагранжа. Тогда можно записать

$$\nabla L(\bar{x}^*, \bar{u}^*, \bar{\lambda}^*) = \nabla f(\bar{x}^*) + \sum_{i=1}^m \lambda_i \nabla h_i(\bar{x}^*) - \sum_{i=m+1}^r u_i^* \nabla g_i(\bar{x}^*) = 0. \tag{2.41}$$

Условия (2.39) позволяют установить, какие из ограничений неравенств являются активными. Покажем это на рассмотренном выше примере 2.3. Функция Лагранжа для этого примера имеет вид:

$$\begin{aligned}
L(\bar{x}, \bar{u}) &= x_1^2 + x_2 - u_1(-x_1^2 - x_2^2 + 9) - u_2(-x_1 - x_2 + 1), \\
\frac{\partial L}{\partial x_1} &= 2x_1 + 2x_1 u_1 + u_2 = 0, \\
\frac{\partial L}{\partial x_2} &= 1 + 2x_2 u_1 + u_2 = 0.
\end{aligned} \tag{2.42}$$

Из условия $u_i^* g_i(\bar{x}^*) = 0$ получим:

$$\begin{aligned}
u_1(-x_1^2 - x_2^2 + 9) &= 0, \\
u_2(-x_1 - x_2 + 1) &= 0,
\end{aligned} \tag{2.43}$$

причем $u_1 \geq 0, u_2 \geq 0$.

Из условий (2.43) следует, что либо $u_i^* = 0$, либо $g_i(\bar{x}^*) = 0$. Однако если u_1^* и u_2^* одновременно положить равными нулю, возникает противоречие в соотношениях (2.42): во втором уравнении (2.42) получим, что $1 = 0$. Следовательно, не все u_i^* могут быть нулями. Предположим, что $u_1^* = 0$, тогда из второго уравнения (2.42) найдем $u_2^* = -1$, чего быть не может по условию (2.39). Положим теперь $u_2^* = 0$, тогда из первого уравнения (2.42) найдем:

$$x_1(1 + u_1) = 0,$$

а так как $u_1^* > 0$, то $x_1^* = 0$. Поскольку $u_1^* \neq 0$, то неравенство $g_1(\bar{x}^*)$ является активным, т.е. $-x_1^{*2} - x_2^{*2} + 9 = 0$, откуда $x_2^* = \pm 3$.

Однако в точке $(0; 3)$ нарушается ограничение $g_2(\bar{x})$, следовательно условный минимум достигается в точке $\bar{x}^* = (0; -3)$, при этом $u_1^* = 1/6$; $u_2^* = 0$ и все ограничивающие условия выполнены.

Если подобный анализ провести сложно для задач большой размерности, то можно воспользоваться алгоритмом, основанном на поэтапном исследовании функции $f(\bar{x})$ и границ области допустимых решений (ОДР).

1. При анализе функций $f(\bar{x})$, $h_i(\bar{x})$ и $g_i(\bar{x})$ полезно знать, являются ли они выпуклыми. Установить это можно с помощью условий выпуклости. Для любой

точки выпуклой функции одной переменной (см. рис. 2.7) справедливо неравенство:

$$f(x) \geq f(x^0) + f'(x^0)\Delta x;$$

соответственно для вогнутой функции $f(x) \leq f(x^0) + f'(x^0)\Delta x$. По аналогии записываются условия для функций многих переменных.

Для функций второго порядка выпуклость (вогнутость) легко проверяется по знаку матрицы Гессе. Для выпуклой функции матрица Гессе положительно определена (для вогнутой функции матрица Гессе отрицательно определена). Для функций более высокого порядка матрица Гессе становится функциональной и такая проверка затруднительна.

Если в задаче (2.36) – (2.38) функции $h_i(\bar{x})$ и $g_i(\bar{x})$ выпуклы, то они определяют выпуклую ОДР³, и в случае выпуклой задача имеет единственный экстремум. При отсутствии выпуклости задача может иметь несколько локальных экстремумов, что, естественно, немного усложняет решение.

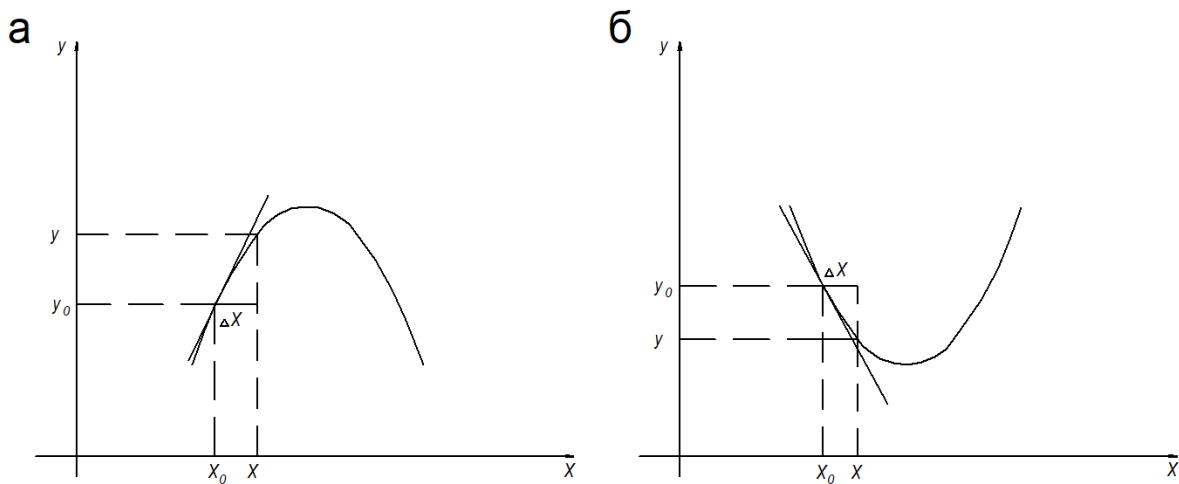


Рисунок 2.9. Функции:
а – вогнутая; б – выпуклая

2. После исследования свойств функций $f(\bar{x})$, $h_i(\bar{x})$, $g_i(\bar{x})$ решается задача на условный экстремум с ограничениями-равенствами, функция Лагранжа при этом имеет вид:

$$L(\bar{x}, \bar{\lambda}) = f(\bar{x}) + \sum_{i=1}^m \lambda_i h_i(\bar{x})$$

3. Из найденных в п. 2 стационарных точек для дальнейшего рассмотрения остаются только те, которые принадлежат ОДР, т.е. удовлетворяют ограничениям $g_i(\bar{x}) \geq 0$.

4. Проверяются границы ОДР: для этого ограничения $g_i(\bar{x})$ полагаются активными и включаются в функцию Лагранжа.

³ Область R называется выпуклой, если для $\forall \bar{x}^1 \in R$ и $\forall \bar{x}^2 \in R$ выполняется условие $\bar{x}_\alpha = \alpha \bar{x}^1 + (1 - \alpha) \bar{x}^2$.

5. В стационарных точках проверяются достаточные условия существования условного экстремума.

6. Сравнивая значения функции $f(\bar{x})$ в оставшихся после проверки достаточных условий точках, определяются точки абсолютного условного экстремума.

2.6. Задача выбора оптимального технологического режима процесса сушки целлюлозы

Найти параметры оптимального технологического режима работы пресспата (сушильной машины) из условия максимальной производительности. Процесс сушки целлюлозы определяется следующими параметрами:

- скоростью пресспата v , м/мин;
- давлением пара в сушильных цилиндрах p , МПа;
- массой 1м² полотна ρ , г/м²;
- влажностью полотна u , %.

В зависимости от параметров технологического режима производительность пресспата может быть рассчитана по формуле:

$$П = a \cdot \beta \cdot v \cdot \rho \cdot (1 - 0,01u), \quad (2.44)$$

где β – ширина полотна целлюлозы, м;

$a = 0,68 \cdot 10^{-4}$ – коэффициент, позволяющий выразить производительность в $\frac{\text{т} \cdot \text{в} \cdot \text{с} \cdot \text{ц}}{\text{ч}}$.

Параметры технологического режима связаны между собой; эта связь может быть описана эмпирически уравнением регрессии:

$$u = u_0 + b_1(\rho - \rho_0) + b_2(v - v_0) - b_3(p - p_0). \quad (2.45)$$

Параметры ρ_0, u_0, v_0, p_0 определяют некоторый номинальный режим, в окрестности которого справедливо линейное уравнение (2.45).

На диапазон измерений параметров технологическим регламентом или ГОСТом наложены ограничения:

$$0 \leq v \leq v_{\max}; \rho_{\min} \leq \rho \leq \rho_{\max}; u_{\min} \leq u \leq u_{\max}; p \leq p \leq p_{\max}. \quad (2.46)$$

С учетом соотношений (2.45) – (2.46) задача формулируется следующим образом:

Найти максимум критерия оптимальности (2.44) при ограничениях (2.45) и (2.46).

Исследуем задачу согласно изложенному выше алгоритму.

Подставим соотношение (2.45) в функцию цели (2.44):

$$П = a \cdot \beta \cdot v \cdot \rho \cdot (c_0 - c_1\rho - c_2v + c_3p), \quad (2.47)$$

где $c_0 = 1 - 0,01u_0 + 0,01b_1\rho_0 + 0,01b_2v_0 - 0,01b_3p_0$;

$c_1 = 0,01b_1, c_2 = 0,01b_2, c_3 = 0,01b_3$.

Сформулируем необходимые условия оптимальности для функции (2.47)

$$\begin{aligned}
\frac{\partial \Pi}{\partial v} &= ac_0 \rho - ac_1 \rho^2 - 2ac_2 v \rho + ac_3 \rho p = 0; \\
\frac{\partial \Pi}{\partial \rho} &= ac_0 v - 2ac_1 v \rho - ac_2 v^2 + ac_3 v \rho = 0; \\
\frac{\partial \Pi}{\partial p} &= ac_3 v \rho = 0.
\end{aligned} \tag{2.48}$$

Уравнения (2.48) обращаются в тождество только при $v = 0$ или $\rho = 0$, что не имеет смысла для рассматриваемой задачи. Тогда, согласно теореме Вейерштрасса, решение задачи (2.44) – (2.46) будет достигаться на границе области допустимых решений, определяемой системой ограничений (2.45), (2.46).

Исследуем границы ОДР.

Заметим, что параметр p входит в функцию цели линейно, причем $\frac{\partial \Pi}{\partial p} > 0$ (см. (2.47)), следовательно, для достижения максимума Π давление пара в сушильных цилиндрах следует поддерживать на верхней границе. Подставив в уравнение (2.45), сведем задачу к исследованию на условный экстремум функции цели Π , как функции двух переменных. Запишем уравнение (2.45) с учетом $p = p_{max}$ в виде

$$\Delta u = u - d = b_1(\rho - \rho_0) + b_2(v - v_0),$$

где $d = u_0 - b_3(p_{max} - p_0)$

и составим функцию Лагранжа, считая активным лишь одно из ограничений $u = u_{min}$, либо $u = u_{max}$:

$$L(\rho, v, \lambda) = a\beta v \rho (1 - 0,01\tilde{u}) + \lambda[b_1(\rho - \rho_0) + b_2(v - v_0) - \Delta\tilde{u}],$$

где $\tilde{u}$ – одно из граничных значений u ; $\Delta\tilde{u} = \tilde{u} - d$.

Необходимые условия оптимальности имеют вид:

$$\begin{aligned}
\frac{\partial L}{\partial v} &= a\beta \rho (1 - 0,01\tilde{u}) + \lambda b_2 = 0; \\
\frac{\partial L}{\partial \rho} &= a\beta v (1 - 0,01\tilde{u}) + \lambda b_1 = 0; \\
\frac{\partial L}{\partial \lambda} &= b_1(\rho - \rho_0) + b_2(v - v_0) - \Delta\tilde{u} = 0.
\end{aligned}$$

Решив систему уравнений, найдем координаты стационарной точки:

$$\begin{aligned}
\lambda^* &= \frac{(\Delta\tilde{u} + b_1\rho_0 + b_2v_0)a\beta(1 - 0,01\tilde{u})}{2b_1b_2}; \\
\rho^* &= \frac{\Delta\tilde{u} + b_1\rho_0 + b_2v_0}{2b_1}; \\
v^* &= \frac{\Delta\tilde{u} + b_1\rho_0 + b_2v_0}{2b_2}.
\end{aligned} \tag{2.49}$$

Решим эту задачу для конкретных исходных данных и сделаем геометрическую интерпретацию.

Пусть

$$\beta = 6\text{м}; b_1 = 0,1; b_2 = 0,5; b_3 = 150; u_{min} = 12\%; u_{max} = 18\%; u_0 = 15\%;$$

$$\rho_{min} = 500 \frac{\Gamma}{\text{М}^2}; \rho_{max} = 900 \frac{\Gamma}{\text{М}^2}; v_{max} = 120 \frac{\text{М}}{\text{МИН}}; p_{max} = 0,45 \text{ МПа};$$

$$\rho_0 = 700 \frac{\Gamma}{\text{М}^2}; v_0 = 90 \frac{\text{М}}{\text{МИН}}; p_0 = 0,42 \text{ МПа}.$$

На рисунке 2.10 в координатах ρ и v показана ОДР задачи (многоугольник ABCD), линии уровня функции цели Π представляют собой семейство гипербол. Считая активным ограничением $u = u_{min} = 12\%$, находим на этой границе точку условного максимума $\bar{x}^*$ согласно соотношениям (2.49):

$$\rho^* = 582,5 \frac{\Gamma}{\text{М}^2}; v^* = 116,5 \frac{\text{М}}{\text{МИН}};$$

$$\Pi(\rho^*, v^*, u_{min}) = 0,6 \cdot 10^{-4} \cdot 6 \cdot 116,5 \cdot 582,5 \cdot 0,88 = 24,36 \frac{\Gamma \cdot \text{В} \cdot \text{С} \cdot \text{Ц}}{\text{Ч}}.$$

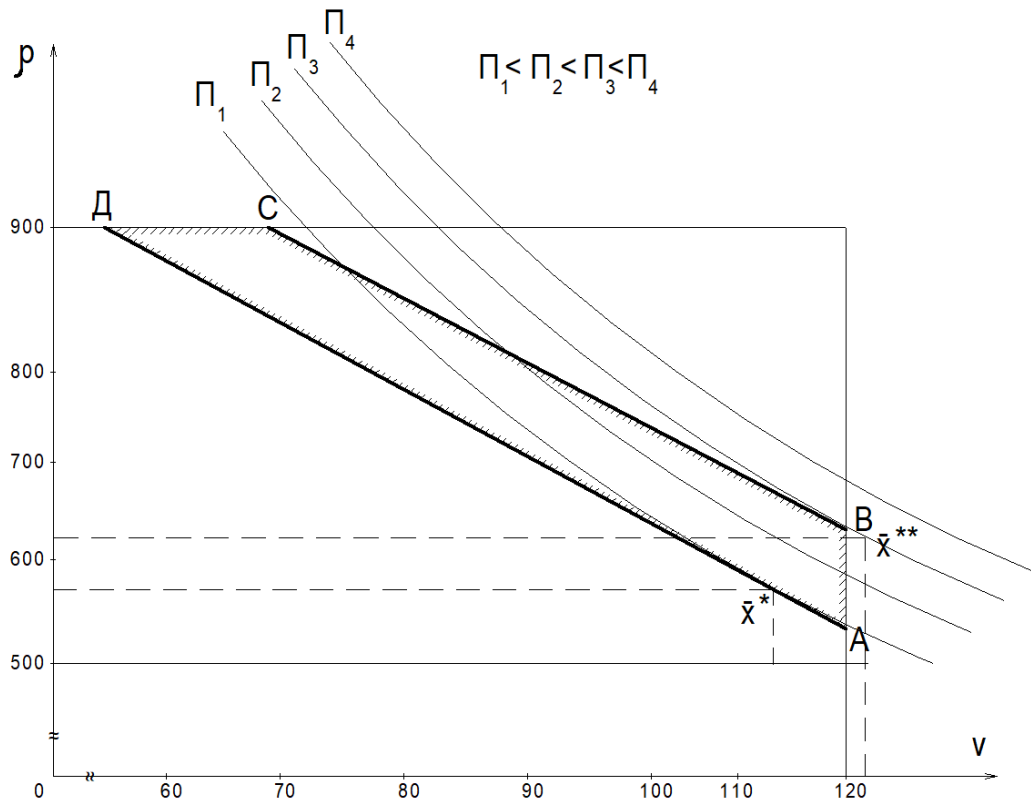


Рисунок 2.10. Геометрическая интерпретация задачи выбора оптимальных параметров процесса сушки целлюлозы

На верхней границе $u_{max} = 18\%$ точка условного максимума $\bar{x}^{**}$ имеет координаты:

$$\rho^{**} = 622,5 \frac{\Gamma}{\text{М}^2}; v^{**} = 122,5 \frac{\text{М}}{\text{МИН}}.$$

При определении точек условного максимума на нижней и верхней границе параметра u мы не учитывали ограничения на диапазон изменения параметров ρ и v . Проверив выполнение этих ограничений, замечаем, что точка $\bar{x}^{**}$ не является допустимой, так как $v^{**} > v_{max}$.

Скорректируем параметр v , положив $v = v_{max}$, и вычислим новое значение из условия, чтобы точка принадлежала границе ОДР $u = u_{max}$, т.е. из уравнения:

$u_{max} = 15 + 0,1(\rho - 700) + 0,5(v_{max} - 90) - 150(p_{max} - 0,42)$; получим, что $\rho = 625 \frac{\text{г}}{\text{м}^2}$.

Параметрам $v = v_{max} = 120 \frac{\text{м}}{\text{мин}}$ и $\rho = 625 \frac{\text{г}}{\text{м}^2}$ соответствует вершина в ОДР. Найдем значение функции цели Π в точке В:

$$\Pi(B) = 0,68 \cdot 10^{-4} \cdot 6 \cdot 120 \cdot 625 \cdot 0,82 = 25,1 \frac{\text{т} \cdot \text{в} \cdot \text{с} \cdot \text{ц}}{\text{ч}}.$$

Сравнивая значения $\Pi(\rho^*, v^*, u_{min})$ и $\Pi(B)$, видим, что в точке В функция цели Π имеет условный максимум.

ГЛАВА 3. ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

3.1. Постановка задач линейного программирования

Среди прикладных задач оптимизации достаточно часто встречаются задачи, математическая модель которых записывается с помощью линейных соотношений: линейного критерия оптимальности, называемого линейной формой, и системы линейных ограничений – уравнений и неравенств. Такие задачи относятся к классу задач линейного программирования (ЛП).

В АСУТП к задачам ЛП сводятся задачи выбора оптимального технологического режима агрегата (или совокупности агрегатов), оперативно-календарного планирования выпуска и отгрузки продукции, распределения ограниченных ресурсов, координации производительности взаимосвязанных агрегатов и т.д. Эти задачи решаются главным образом в подсистемах верхнего уровня АСУТП. В общем виде задача линейного программирования формулируется следующим образом [8]:

найти максимум (минимум) линейной формы

$$F = \sum_{j=1}^n c_j x_j \quad (3.1)$$

при ограничениях

$$\sum_{j=1}^n a_{ij} x_j \leq b_i, \quad i = 1, \dots, m. \quad (3.2)$$

Так как переменные x_j имеют физический смысл и, как правило, не могут быть отрицательными (например, расход, запасы, концентрация и т.д.), то в задаче ЛП выдвигается требование неотрицательности переменных x_j :

$$x_j \geq 0, \quad j = 1, \dots, n. \quad (3.3)$$

Ограничения (3.2) могут иметь вид уравнений, а также неравенств любого знака. Если ограничения (3.2) заданы в виде уравнений, то такая форма задачи ЛП называется канонической. В дальнейшем будет показано, что от системы неравенств легко перейти к равенствам и наоборот.

Для определенности впредь будем полагать, что требуется отыскать максимум линейной формы F ; задача на минимум всегда может быть сведена к задаче на максимум линейной формы, противоположной исходной по знаку.

Допустимым решением задачи ЛП называется любая совокупность неотрицательных переменных $x_j \geq 0, j = 1, \dots, n$, удовлетворяющая системе ограничений (3.2) [9].

Оптимальным решением задачи ЛП называется то из допустимых решений, при котором линейная форма F достигает максимума.

Задача ЛП необязательно имеет решение. Система ограничений (3.2) может оказаться противоречивой или иметь решения только в области отрицательных значений переменных. Может оказаться, что ОДР задачи ЛП

существует, но среди допустимых решений нет оптимального. Причиной всех этих случаев является некорректная постановка задачи.

Рассмотрим прежде всего вопрос о существовании допустимого решения. Пусть задача ЛП приведена к канонической форме – все ограничения заданы в виде уравнений:

$$\sum_{j=1}^n a_{ij}x_j = b_i, \quad i = 1, \dots, m. \quad (3.4)$$

Пусть также система уравнений (3.4) совместна, и все уравнения системы линейно-независимы, т.е. ни одно из них не может быть представлено как линейная комбинация остальных. Тогда очевидно, что задача ЛП будет иметь смысл, если выполнено условие $m < n$, так как в этом случае система уравнений (3.4) имеет бесчисленное множество решений. При этом $n-m$ переменным можно давать произвольные значения ($n-m$ переменные называются независимыми), остальные m зависимых переменных выразятся через независимые и свободные члены. Выберем в качестве зависимых произвольные m переменные (например, $x_1, \dots, x_m$) и разрешим относительно них систему уравнений (3.4), тогда получим:

$$x_i = \tilde{b}_i - \sum_{j=m+1}^n \tilde{a}_{ij}x_j, \quad i = 1, \dots, m, \quad (3.5)$$

где $\tilde{b}_i, \tilde{a}_{ij}$ – соответственно свободные члены и коэффициенты уравнений после преобразований.

Система уравнений (3.5) эквивалентна системе (3.4). Переменные x_i , стоящие в (3.5) слева, называют в задачах ЛП базисными. Вычисляя базисные переменные при произвольных значениях независимых переменных $x_{m+1}, \dots, x_n$, получим множество решений системы (3.4). Неотрицательные решения системы (3.4) образуют множество допустимых решений задачи ЛП. Решение системы (3.4), в котором все зависимые переменные равны нулю:

$$x_{m+1} = x_{m+2} = \dots = x_n = 0,$$

а базисные равны свободным членам $x_1 = \tilde{b}_1, x_2 = \tilde{b}_2, \dots, x_m = \tilde{b}_m$, называется базисным.

Базисное решение допустимо, если все

$$\tilde{b}_i \geq 0, \quad i = 1, \dots, m.$$

Таким образом для допустимости базисного решения все свободные члены в (3.5) должны быть неотрицательны. Если среди свободных членов b_i по крайней мере один равен нулю, то базисное решение называется вырожденным. Если в системе (3.5) все $\tilde{b}_i \geq 0$, то говорят, что задача ЛП задана в допустимой канонической форме.

3.2. Геометрическая интерпретация задачи ЛП

Задачу ЛП легко интерпретировать геометрически. Для наглядности интерпретации рассмотрим задачу ЛП в канонической форме, причем $n - m = 2$,

тогда любые две переменные могут быть независимыми. Дадим геометрическую интерпретацию задачи в плоскости независимых переменных. Для определенности примем в качестве независимых переменных x_{m+1}, x_{m+2} , тогда $x_1, x_2, \dots, x_m$ будут базисными. Запишем систему ограничений аналогично (3.5), разрешив все уравнения относительно независимых переменных:

$$\begin{aligned} x_1 &= \tilde{b}_1 - \tilde{a}_{1m+1}x_{m+1} - \tilde{a}_{1m+2}x_{m+2} \geq 0; \\ x_2 &= \tilde{b}_2 - \tilde{a}_{2m+1}x_{m+1} - \tilde{a}_{2m+2}x_{m+2} \geq 0; \\ x_m &= \tilde{b}_m - \tilde{a}_{mm+1}x_{m+1} - \tilde{a}_{mm+2}x_{m+2} \geq 0. \end{aligned} \quad (3.6)$$

Каждое неравенство (3.6) определяет в плоскости переменных x_{m+1}, x_{m+2} полуплоскость, все точки которой расположены с той стороны границы, определяемой уравнением

$$x_i = \tilde{b}_i - \tilde{a}_{im+1}x_{m+1} - \tilde{a}_{im+2}x_{m+2} = 0, \quad i = 1, \dots, m,$$

где выполняется условие $x_i > 0$.

На рисунке 3.1 полуплоскости, соответствующие каждому неравенству, заштрихованы. Точки, принадлежащие одновременно всем ограничениям (3.6), образуют множество допустимых решений задачи ЛП, представляющее собой выпуклый многоугольник (в общем случае многогранник).

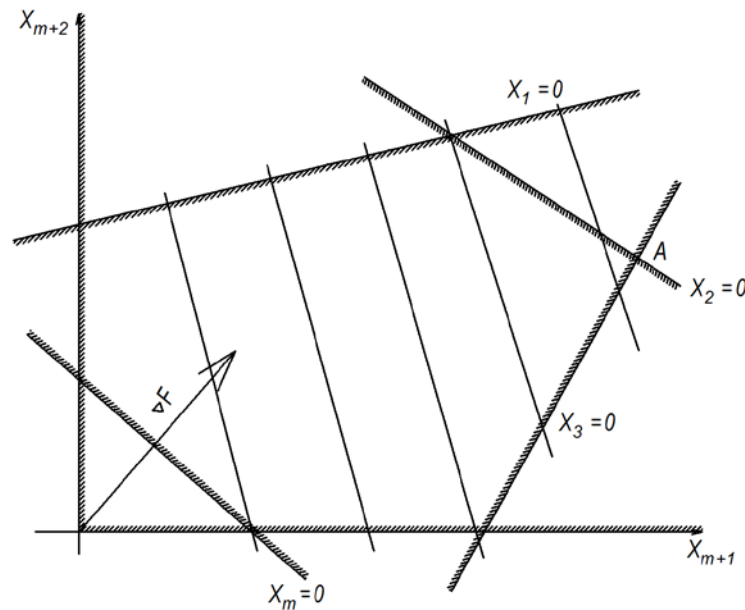


Рисунок 3.1. Геометрическая интерпретация задачи ЛП

Доказательство выпуклости ОДР задачи ЛП см., например, в [9]. Очевидно, что многоугольник допустимых решений лежит обязательно в первом квадранте плоскости, так как переменные x_{m+1}, x_{m+2} должны быть неотрицательны.

Чтобы выяснить, в какой из точек ОДР задача ЛП имеет решение, нанесем на плоскость чертежа линии уровня линейной формы F . Для этого, подставив в

линейную форму F соотношения (3.6), выразим F через независимые переменные x_{m+1}, x_{m+2} :

$$F = F_0 + \tilde{c}_{m+1}x_{m+1} + \tilde{c}_{m+2}x_{m+2}. \quad (3.7)$$

Задавая различные значения F , построим в плоскости независимых переменных x_{m+1}, x_{m+2} линии уровня F в виде семейства параллельных прямых, ортогональных $\nabla F = (\tilde{c}_{m+1}, \tilde{c}_{m+2})$. Из рисунка 3.1 следует, что решением задачи ЛП является вершина A ОДР, в которой линейная форма F достигает максимума.

Пример 3.1. На бумагоделательной машине (БДМ) выпускается два вида бумаги: упаковочная и обойная; при этом расходуется три вида сырья: сульфитная целлюлоза, древесная масса и макулатура. Требуется определить выпуск каждого вида бумаги на некоторый период времени при ограниченных запасах сырья, чтобы получить максимальную прибыль при реализации продукции. Исходные данные для решения задачи приведены в табл. 3.1.

Таблица 3.1. Исходные данные к задаче оптимального выпуска продукции

Сырье	Удельные нормы расхода сырья на бумагу		Запасы сырья, т
	упаковочную, т/т	обойную, т/т	
Сульфитная целлюлоза	$q_{11} = 0,5$	$q_{12} = 0,25$	45
Древесная масса	$q_{21} = 0,5$	$q_{22} = 0,46$	60
Макулатура	$q_{31} = 0$	$q_{32} = 0,19$	21
Прибыль на 1 т продукции, руб/т	50	35	

Обозначим x_1 и x_2 – количество упаковочной и обойной бумаги, которое следует произвести из условия максимума прибыли. Тогда математическая модель задачи оптимального выпуска бумаги будет иметь следующий вид: найти максимум прибыли

$$F = 50x_1 + 35x_2 \quad (3.8)$$

при ограничениях на запасы сырья

$$\begin{aligned} 0,5x_1 + 0,25x_2 &\leq 45, \\ 0,5x_1 + 0,46x_2 &\leq 60, \\ 0,19x_2 &\leq 21. \end{aligned} \quad (3.9)$$

Учитывая, что переменные x_1 и x_2 определяют выпуск продукции и не могут принимать отрицательные значения, потребуем выполнение условия

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

Обозначим количество каждого вида сырья, которое останется неиспользованным при выпуске продукции, y_1, y_2, y_3 . Очевидно, что эти переменные связаны с ограничениями на запасы (3.9) следующим образом:

$$\begin{aligned} y_1 &= 45 - 0,5x_1 - 0,25x_2 \geq 0, \\ y_2 &= 60 - 0,5x_1 - 0,46x_2 \geq 0, \\ y_3 &= 21 - 0,19x_2 \geq 0. \end{aligned} \quad (3.10)$$

С помощью переменных $y_i = 0, i = 1, 2, 3$, неравенства (3.9) сведены к эквивалентной им системе уравнений (3.10), в которой за независимые переменные удобно принять x_1, x_2 , а в качестве базисных $y_i, i = 1, 2, 3$. В плоскости независимых переменных x_1, x_2 (рис. 3.2) каждому ограничению $y_i = 0, i = 1, 2, 3$ системы (3.10) соответствует прямая линия, которая делит плоскость чертежа на две полуплоскости. Та полуплоскость, точки которой удовлетворяют условию $y_i > 0$, заштрихована. Множество точек, удовлетворяющих системе уравнений (3.10) и условиям неотрицательности x_1, x_2 , образует ОДР задачи. На рисунке 3.2 нанесены линии уровня функции прибыли F , направление возрастания которой указывает ∇F с координатами $\nabla F = (50, 35)$. Из рисунка 3.2 видно, что в точке А, координаты которой $x_1^A = 55, x_2^A = 70$, достигается максимальное значение функции прибыли F , которое найдем, подставив координаты точки А в выражение (3.8):

$$F_{max} = 5200.$$

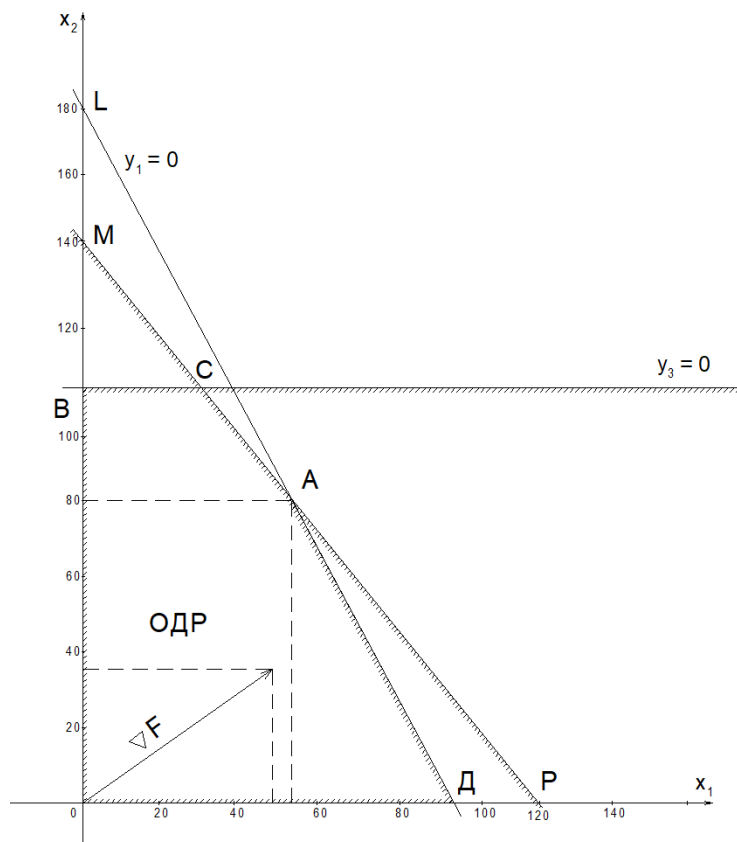


Рисунок 3.2. Геометрическая интерпретация задачи оптимального выпуска продукции

Покажем теперь, что решения задачи ЛП, принадлежащие вершинам ОДР, являются базисными. Действительно, при условии отсутствия вырожденности, в каждой вершине ОДР пересекается столько граней, сколько имеется независимых переменных, т.е. в общем случае $n-m$, а для задачи на рисунке 3.2 две границы. Следовательно, в каждой вершине ОДР независимые $n-m$ переменных равны нулю, остальные переменные являются базисными и могут быть определены из системы ограничений. Каждой вершине ОДР соответствует свой набор базисных переменных, причем при последовательном переходе из одной вершины ОДР в соседнюю одна из базисных переменных и одна из независимых меняются местами. Так, например, в вершинах ОДР на рисунке 3.2 находятся следующие решения:

$$\begin{aligned} \text{точка } O: x_1 = x_2 = 0; y_1 = 45; y_2 = 60; y_3 = 21; \\ \text{точка } B: x_1 = y_3 = 0; x_2 = 110; y_1 = 17,5; y_2 = 9,4; \\ \text{точка } C: y_2 = y_3 = 0; x_2 = 110; x_1 = 18,8; y_1 = 36,9; \\ \text{точка } A: y_1 = y_2 = 0; x_1 = 55; x_2 = 70; y_3 = 13,3; \\ \text{точка } D: y_1 = x_2 = 0; x_1 = 90; y_2 = 15; y_3 = 21. \end{aligned}$$

Вершины ОДР называют опорными вершинами, а принадлежащие им базисные решения – опорными. Помимо опорных вершин имеется ряд пересечений границ за пределами ОДР: базисные решения, соответствующие этим вершинам, являются недопустимыми (например, точки L, M, P на рисунке 3.2).

Из геометрической интерпретации задачи ЛП следует важный вывод: максимум линейной формы F достигается обязательно на границе ОДР, а именно: в одной из опорных вершин. Поэтому при решении задачи ЛП можно исключить из рассмотрения бесчисленное множество допустимых решений, оставив только опорные, число которых конечно.

Задача ЛП может иметь не единственное решение: если линии уровня линейной формы F параллельны одной из сторон ОДР, расположенной в направлении ∇F (см. рис. 3.3).

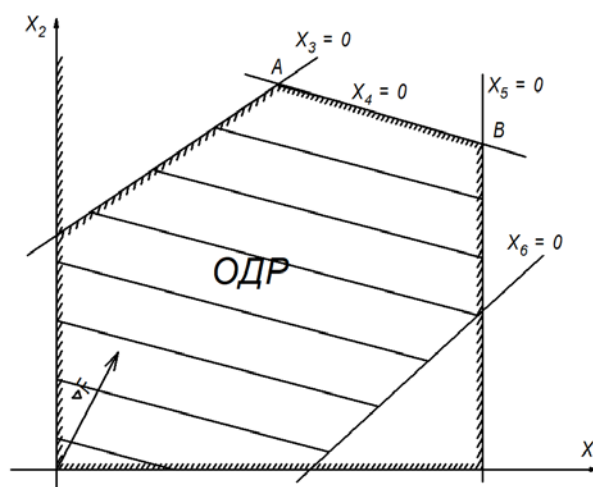


Рисунок 3.3. Геометрическая интерпретация задачи ЛП в случае не единственного решения

На рисунке 3.3 линии уровня F параллельны стороне АВ ОДР, с которой совпадает линия уровня, соответствующая $F = F_{max}$. Таким образом, любая из точек отрезка АВ является решением задачи ЛП.

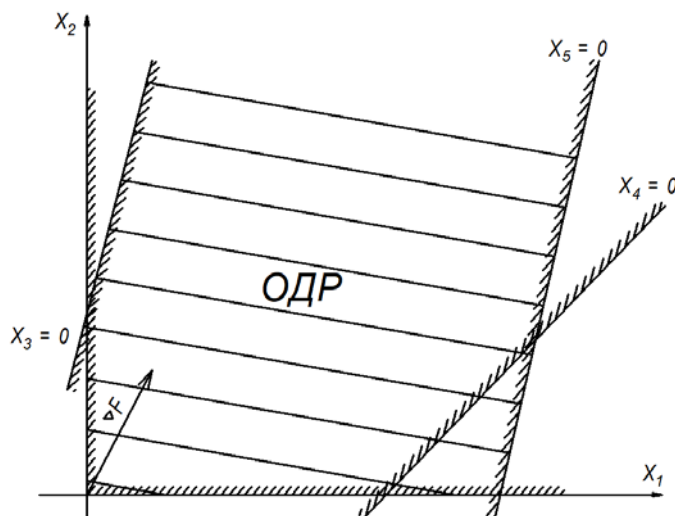


Рисунок 3.4. Геометрическая интерпретация задачи ЛП в случае неограниченности линейной формы F

Задача ЛП не имеет решения, если линейная форма F не ограничена сверху (рис. 3.4).

3.3. Симплексный метод решения задачи линейного программирования с односторонними ограничениями

Симплексный метод решения задач ЛП разработан американским математиком Дж. Данцигом в 1949 г., а основы отечественной теории линейного программирования заложены Л. В. Канторовичем в 1941 г.

Симплексный метод относится к численным конечным методам: он позволяет найти решение задачи ЛП, если оно существует, за конечное число шагов или итераций.

Основной аппарат симплексного метода – модифицированные жордановы исключения [8].

Модифицированные жордановы исключения

Для лучшего понимания рассмотрим вначале обыкновенные жордановы исключения, как это сделано в [8].

Пусть система из m линейно-независимых уравнений с n неизвестными разрешена относительно некоторых произвольно выбранных m базисных переменных $x_1, \dots, x_m$:

$$\begin{aligned} x_1 &= a_{1m+1}x_{m+1} + a_{1m+2}x_{m+2} + \dots + a_{1n}x_n + b_1, \\ x_2 &= a_{2m+1}x_{m+1} + a_{2m+2}x_{m+2} + \dots + a_{2n}x_n + b_2, \\ x_m &= a_{mm+1}x_{m+1} + a_{mm+2}x_{m+2} + \dots + a_{mn}x_n + b_m. \end{aligned} \quad (3.11)$$

Для единообразия записи введем переменную $x_{n+1} \equiv 1$ и новое обозначение для свободных членов $b_i = a_{in+1}$, тогда система (3.11) примет вид:

$$\begin{aligned} x_1 &= a_{1m+1}x_{m+1} + a_{1m+2}x_{m+2} + \dots + a_{1n}x_n + a_{1n+1}x_{n+1}, \\ x_2 &= a_{2m+1}x_{m+1} + a_{2m+2}x_{m+2} + \dots + a_{2n}x_n + a_{2n+1}x_{n+1}, \\ x_m &= a_{mm+1}x_{m+1} + a_{mm+2}x_{m+2} + \dots + a_{mn}x_n + a_{mn+1}x_{n+1}. \end{aligned}$$

Эта система может быть записана в виде таблицы:

	x_{m+1}	x_{m+2}	$\dots$	x_s	$\dots$	x_n	x_{n+1}	
x_1	a_{1m+1}	a_{1m+2}	$\dots$	a_{1s}	$\dots$	a_{1n}	a_{1n+1}	
x_2	a_{2m+1}	a_{2m+2}	$\dots$	a_{2s}	$\dots$	a_{2n}	a_{2n+1}	
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	
x_r	a_{rm+1}	a_{rm+2}	$\dots$	a_{rs}	$\dots$	a_{rn}	a_{rn+1}	
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	
x_m	a_{mm+1}	a_{mm+2}	$\dots$	a_{ms}	$\dots$	a_{mn}	a_{mn+1}	

(3.12)

Шагом обыкновенного жорданова исключения с разрешающим элементом $a_{rs} \neq 0$, с r разрешающей строкой и s разрешающим столбцом называется преобразование таблицы (3.12) для перемены мест x_r и x_s , т.е. операция решения уравнения

$$x_r = a_{rm+1}x_{m+1} + a_{rm+2}x_{m+2} + \dots + a_{rs}x_s + \dots + a_{rn+1}x_{n+1}$$

относительно x_s , подстановки его во все остальные уравнения системы и записи полученной системы в виде таблицы, аналогичной (3.12):

		s столбец							
		x_{m+1}	x_{m+2}	$\dots$	x_r	$\dots$	x_{n+1}		
г строка	x_1	d_{1m+1}	d_{1m+2}	$\dots$	a_{1s}	$\dots$	d_{1n+1}		
	x_2	d_{2m+1}	d_{2m+2}	$\dots$	a_{2s}	$\dots$	d_{2n+1}		
	$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$		
	x_s	$-a_{rm+1}$	$-a_{rm+2}$	$\dots$	1	$\dots$	$-a_{rn+1}$		
	$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$		
	x_m	d_{mm+1}	d_{mm+2}	$\dots$	a_{ms}	$\dots$	d_{mn+1}		

: a_{rs} (3.13)

В таблице (3.13) элементы d_{ij} подсчитываются следующим образом:

$$d_{ij} = a_{ij}a_{rs} - a_{is}a_{rj}, (i \neq r, j \neq s), \quad (3.14)$$

причем все элементы следует разделить на a_{rs} . Действительно, если $a_{rs} \neq 0$, то разрешив r уравнение относительно x_s , получим:

$$x_s = \frac{-\sum_{j=m+1}^{s-1} a_{rj}x_j + x_r - \sum_{j=s+1}^{n+1} a_{rj}x_j}{a_{rs}}.$$

Подставим x_s во все остальные уравнения системы, тогда i уравнение ($i \neq r$) преобразуется следующим образом:

$$\begin{aligned} x_i &= \sum_{j=m+1}^{s-1} a_{ij}x_j + a_{is} \frac{-\sum_{j=m+1}^{s-1} a_{rj}x_j + x_r - \sum_{j=s+1}^{n+1} a_{rj}x_j}{a_{rs}} + \sum_{j=s+1}^{n+1} a_{ij}x_j = \\ &= \sum_{j=m+1}^{s-1} \frac{a_{ij}a_{rs} - a_{is}a_{rj}}{a_{rs}} x_j + \frac{a_{is}}{a_{rs}} x_r - \sum_{j=s+1}^{n+1} \frac{a_{ij}a_{rs} - a_{is}a_{rj}}{a_{rs}} x_j. \end{aligned}$$

Таким образом один шаг жорданова исключения с разрешающим элементом a_{rs} переводит таблицу (3.12) в таблицу (3.13) по следующей схеме:

- 1) разрешающий элемент заменяется единицей;
- 2) остальные элементы разрешающего столбца остаются без изменения;
- 3) остальные элементы разрешающей строки меняют свои знаки;
- 4) элементы остальных строк и столбцов вычисляются по формуле прямоугольника (3.14);
- 5) все элементы новой таблицы делятся на разрешающий элемент.

Пример 3.2. Преобразовать систему уравнений, поменяв местами

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

базисную переменную x_5 и независимую переменную x_3 .

Запишем систему уравнений в виде таблицы:

	x_1	x_2	x_3	1
x_4	1	-2	3	2
x_5	-1	1	2	4
x_6	2	-1	-1	6

Сделаем один шаг обыкновенных жордановых исключений с разрешающим элементом 2:

	x_1	x_2	x_5	1				x_1	x_2	x_5	1
x_4	5	-7	3	-8	:2	$\rightarrow$	x_4	2,5	-3,5	1,5	-4
x_3	1	-1	1	-4			x_3	0,5	-0,5	0,5	-2
x_6	3	-1	1	16			x_6	1,5	-0,5	-0,5	4

В симплексном методе ЛП важно, чтобы элементы разрешающей строки при преобразованиях системы ограничений сохраняли знак, а элементы разрешающего столбца меняли его на противоположный. Поэтому для решения задач ЛП используются модифицированные жордановы исключения, согласно которым система уравнений (3.11) записывается в виде:

$$\begin{aligned}x_1 &= -a_{1m+1}(-x_{m+1}) - a_{1m+2}(-x_{m+2}) - \dots - a_{1n}(-x_n) + a_{1n+1}x_{n+1}, \\x_2 &= -a_{2m+1}(-x_{m+1}) - a_{2m+2}(-x_{m+2}) - \dots - a_{2n}(-x_n) + a_{2n+1}x_{n+1}, \\x_m &= -a_{mm+1}(-x_{m+1}) - a_{mm+2}(-x_{m+2}) - \dots - a_{mn}(-x_n) + a_{mn+1}x_{n+1}.\end{aligned}$$

Соответствующая системе таблица принимает вид:

	$-x_{m+1}$	$-x_{m+2}$	$\dots$	$-x_s$	$\dots$	$-x_n$	x_{n+1}
x_1	α_{1m+1}	α_{1m+2}	$\dots$	α_{1s}	$\dots$	α_{1n}	a_{1n+1}
x_2	α_{2m+1}	α_{2m+2}	$\dots$	α_{2s}	$\dots$	α_{2n}	a_{2n+1}
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_r	α_{rm+1}	α_{rm+2}	$\dots$	α_{rs}	$\dots$	α_{rn}	a_{rn+1}
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_m	α_{mm+1}	α_{mm+2}	$\dots$	α_{ms}	$\dots$	α_{mn}	a_{mn+1}

где для удобства обозначено $\alpha_{ij} = -a_{ij}$.

Один шаг модифицированных жордановых исключений с разрешающим элементом α_{rs} осуществляется по тем же правилам, что и шаг обыкновенных исключений с той лишь разницей, что несколько меняются пункты 2 и 3:

2) остальные элементы разрешающего столбца меняют свои знаки;

3) остальные элементы разрешающей строки остаются без изменения.

Пример 3.3. Преобразовать систему уравнений, поменяв местами базисную переменную x_5 и независимую переменную x_3 .

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

Табличная запись системы уравнений:

	$-x_1$	$-x_2$	$-x_3$	1
x_4	-2	1	-3	2
x_5	1	-4	2	3
x_6	-5	-2	4	-4

Выполним один шаг модифицированных жордановых исключений с разрешающим элементом 2:

	x_1	x_2	x_5	1				x_1	x_2	x_5	1
x_4	-1	-10	3	13	:2	$\rightarrow$	x_4	-0,5	-5	1,5	6,5
x_3	1	-4	1	3			x_3	0,5	-2	0,5	1,5
x_6	-14	12	-4	-20			x_6	-7	6	-2	-10

Из расположенной справа таблицы видим, что базисная переменная x_5 и независимая переменная x_3 поменялись местами.

Способы приведения задачи ЛП к канонической форме

Симплексный метод применяется к канонической форме задачи ЛП с упорядоченным множеством базисных переменных (см. (3.5)). Чтобы привести задачу к канонической форме, ограничения неравенства преобразовать в равенства. Для этого используются искусственные или дополнительные базисные переменные. Рассмотрим различные случаи ограничений в исходной задаче ЛП.

$$1. \quad \sum_{i=1}^n a_{ij}x_j \leq b_i, i = 1, \dots, m. \quad (3.15)$$

Для преобразования неравенства (3.15) в равенство к правой части каждого неравенства добавляется неотрицательная переменная y_i :

$$\sum_{i=1}^n a_{ij}x_j + y_i \leq b_i, i = 1, \dots, m,$$

$$y_i \geq 0.$$

Так как y_i входит только в одно уравнение, то легко разрешить систему относительно y_i , приняв их за базисные переменные:

$$2. \quad \sum_{i=1}^n a_{ij}x_j \geq b_i, i = 1, \dots, m. \quad (3.16)$$

Для перехода к уравнению из правой части неравенства (3.16) вычитается неотрицательная переменная y_i :

$$\sum_{i=1}^n a_{ij}x_j - y_i = b_i, i = 1, \dots, m.$$

Принимая y_i за базисные переменные и разрешая относительно y_i систему уравнений, получим

$$y_i = -b_i + \sum_{i=1}^n a_{ij}x_j \geq 0, i = 1, \dots, m.$$

$$3. \quad \sum_{i=1}^n a_{ij}x_j = b_i, i = 1, \dots, m. \quad (3.17)$$

Чтобы не делать сложных преобразований, разрешая систему уравнений (3.17) относительно некоторых m базисных переменных, в каждое из уравнений (3.17) вводится искусственная базисная переменная y_i

$$\begin{aligned} \sum_{i=1}^n a_{ij}x_j + y_i &= b_i, i = 1, \dots, n, \\ y_i &\geq 0. \end{aligned} \quad (3.18)$$

В отличие от рассмотренных случаев вводится и в линейную форму F с достаточно большим коэффициентом M со знаком плюс, если задача на минимум, и со знаком минус, если – на максимум:

$$F = \sum_{j=1}^n c_jx_j \pm M \sum_{i=1}^m y_i. \quad (3.19)$$

Задача (3.18), (3.19) называется расширенной задачей или задачей M . Если задача M имеет решение, в котором все искусственные переменные y_i равны нулю, то оно является и решением исходной задачи. Очевидно, что при этом сильно увеличивается размерность. В [8] приводится способ, позволяющий сократить размерность задачи путем последнего исключения уравнений (3.17). В настоящем пособии этот способ не рассматривается.

Алгоритм симплексного метода ЛП

Алгоритм симплексного метода состоит в направленном переборе опорных вершин ОДР, причем при переходе из одной опорной вершины в соседнюю значение линейной формы F увеличивается (при условии отсутствия вырожденности). Переход из одной вершины в другую осуществляется за одну итерацию алгоритма.

Алгоритм симплексного метода:

- позволяет найти опорное решение задачи ЛП или установить, что оно не существует;
- содержит признак, позволяющий определить, является ли полученное опорное решение оптимальным;
- указывает способ перехода от уже найденного опорного решения к другому, которому соответствует лучшее значение линейной формы F .

Алгоритм включает два этапа: на первом этапе отыскивается опорное решение задачи ЛП, на втором – оптимальное.

I этап. Нахождение опорного решения задачи ЛП

Математическая модель задачи ЛП в канонической форме с искусственными базисными переменными имеет вид:

$$F = \sum_{j=1}^n c_j x_j \rightarrow \max; \quad (3.20)$$

$$y_i = b_i + \sum_{j=1}^n a_{ij} x_j, i = 1, \dots, m, \quad (3.21)$$

где b_i – свободный член любого знака и, в том числе, равный нулю.

$$\begin{aligned} x_j &\geq 0; y_i \geq 0; \\ j &= 1, \dots, n; i = 1, \dots, m. \end{aligned} \quad (3.22)$$

Соотношения (3.20), (3.21) могут быть представлены в виде таблицы, называемой симплексной⁴.

В таблице введено обозначение $\alpha_{ij} = -a_{ij}$.

Найдем базисное решение:

	$-x_1$	$-x_2$	..	$-x_s$	...	$-x_n$	1
y_1	α_{11}	α_{12}	..	α_{1s}	...	α_{1n}	b_1
y_2	α_{21}	α_{22}	..	α_{2s}	...	α_{2n}	b_2
$\vdots$	...	...	..	...	...	...	...
y_k	α_{k1}	α_{k2}	..	α_{ks}	...	α_{kn}	b_k
$\vdots$	...	...	..	...	...	...	...
y_r	α_{r1}	α_{r2}	..	α_{rs}	...	α_{rn}	b_r
$\vdots$	...	...	..	...	...	...	...
y_m	α_{m1}	α_{m2}	..	α_{ms}	...	α_{mn}	b_m
F	$-C_1$	$-C_2$	..	$-C_s$	...	$-C_n$	0

положим все $x_j = 0$, тогда $y_i = b_i$.

Таким образом, исходной симплексной таблице с искусственным базисом соответствует вершина, образованная пересечением границ: $x_j = 0, j = 1, \dots, n$. Если все $b_i \geq 0$, то решение является опорным, и на этом 1 этап алгоритма заканчивается. Если же какие-либо свободные члены $b_k < 0$, необходимо

⁴ F-строка симплексной таблицы, участвует во всех преобразованиях, но разрешающий элемент в ней не выбирается.

преобразовать симплексную таблицу с помощью модифицированных жордановых исключений так, чтобы все b_k стали неотрицательными, либо установить факт отсутствия опорного решения.

При выполнении перехода от недопустимого к допустимому базисному решению необходимо выбирать разрешающий элемент по следующему правилу.

1. Анализируется столбец свободных членов и выбирается k строка, в которой $b_k < 0$. Если таких строк несколько, то выбирается первая по порядку строка. В k строке отыскивается элемент $\alpha_{ks} < 0$. Если такого элемента нет (все $\alpha_{ks} \geq 0$), то делается вывод: задача ЛП допустимых решений не имеет.

2. Пусть $\alpha_{ks} < 0$, тогда s столбец выбирается в качестве разрешающего. Если в k строке несколько отрицательных элементов, то за разрешающий принимается любой столбец, у которого в k строке находится отрицательный элемент.

3. Вычисляются все неотрицательные отношения свободных членов к элементам разрешающего столбца $\frac{b_i}{\alpha_{is}} \geq 0$ и среди них выбирается наименьшее:

$$\frac{b_r}{\alpha_{rs}} = \min_i \left\{ \frac{b_i}{\alpha_{is}} \geq 0 \right\}.$$

r строка принимается за разрешающую, а элемент α_{rs} – за разрешающий элемент.

Примечание: в случае вырождения, когда какие-либо $b_i = 0$, отношение $\frac{b_i}{\alpha_{is}}$ учитывается только в том случае, если $\alpha_{is} > 0$.

4. Выполняется один шаг модифицированных жордановых исключений над таблицей (3.23) с разрешающим элементом α_{rs} , при этом переменная x_3 вводится в базисное множество вместо переменной y_r , которая становится независимой, что соответствует переходу в соседнюю вершину, образованную пересечением границ:

$$x_j = 0, j = 1, \dots, s-1, s+1, \dots, n; y_r = 0.$$

Если разрешающим оказался элемент, то после шага исключения новый свободный член в этой строке станет положительным, в противном случае (если разрешающим оказался α_{rs} $r \neq k$) от отрицательности свободного члена в k строке за одну итерацию избавиться не удастся.

После завершения итерации все операции, описанные в п.п. 1–4, повторяются до тех пор, пока после конечного числа шагов либо будет установлен факт отсутствия опорного решения, либо в столбце свободных членов все элементы станут неотрицательными, т.е. опорное решение будет найдено.

Поясним смысл алгоритма поиска опорного решения на примерах.

Пример 3.4. Дана задача ЛП

$$F = 2x_1 + 3x_2 \rightarrow \max,$$

$$2x_1 - x_2 \leq 4,$$

$$2x_1 - 3x_2 \geq 6,$$

$$x_j \geq 0; j = 1, 2.$$

Приведем задачу к канонической форме с искусственными базисными переменными y_i и составим симплексную таблицу:

$$F = 2x_1 + 3x_2 \rightarrow \max,$$

$$y_1 = -2x_1 + x_2 + 4 \geq 0,$$

$$y_2 = 2x_1 - 3x_2 - 6 \geq 0,$$

$$x_j \geq 0; j = 1, 2.$$

	$-x_1$	$-x_2$	1	b_i/α_{i1}
y_1	2	-1	4	2
y_2	-2	3	-6	3
F	-2	-3	0	

Во второй строке симплексной таблицы имеется отрицательный свободный член, согласно алгоритму, находим в этой строке отрицательный элемент -2 , принимаем первый столбец за разрешающий и составляем неотрицательные отношения свободных членов к элементам первого столбца. Наименьшее из отношений находится в первой строке, следовательно, первая строка – разрешающая, на пересечении разрешающих столбца и строки стоит разрешающий элемент 2. После шага исключения с разрешающим элементом 2 получим новую симплексную таблицу, в которой базисными являются переменные x_1 и y_2 :

	$-y_1$	$-x_2$	1
x_1	0,5	-0,5	2
y_2	1	2	-2
F	1	-2	4

Во второй строке новой симплексной таблицы свободный член по-прежнему отрицателен, но больше ни одного отрицательного элемента в этой строке нет, отсюда следует вывод: задача ЛП допустимых решений не имеет. Действительно, отрицательность свободного члена означает, что одна из базисных переменных в нашем примере $y_2 < 0$. Чтобы сделать y_2 неотрицательной, следует увеличить ту из независимых переменных, которая входит в уравнение для y_2 с положительным коэффициентом (в симплексной таблице этот же коэффициент будет отрицательным).

Геометрически (рис. 3.5) увеличение x_1 означает движение из исходной точки с координатами $x_1 = 0, x_2 = 0$ к границам $y_1 = 0$ и $y_2 = 0$.

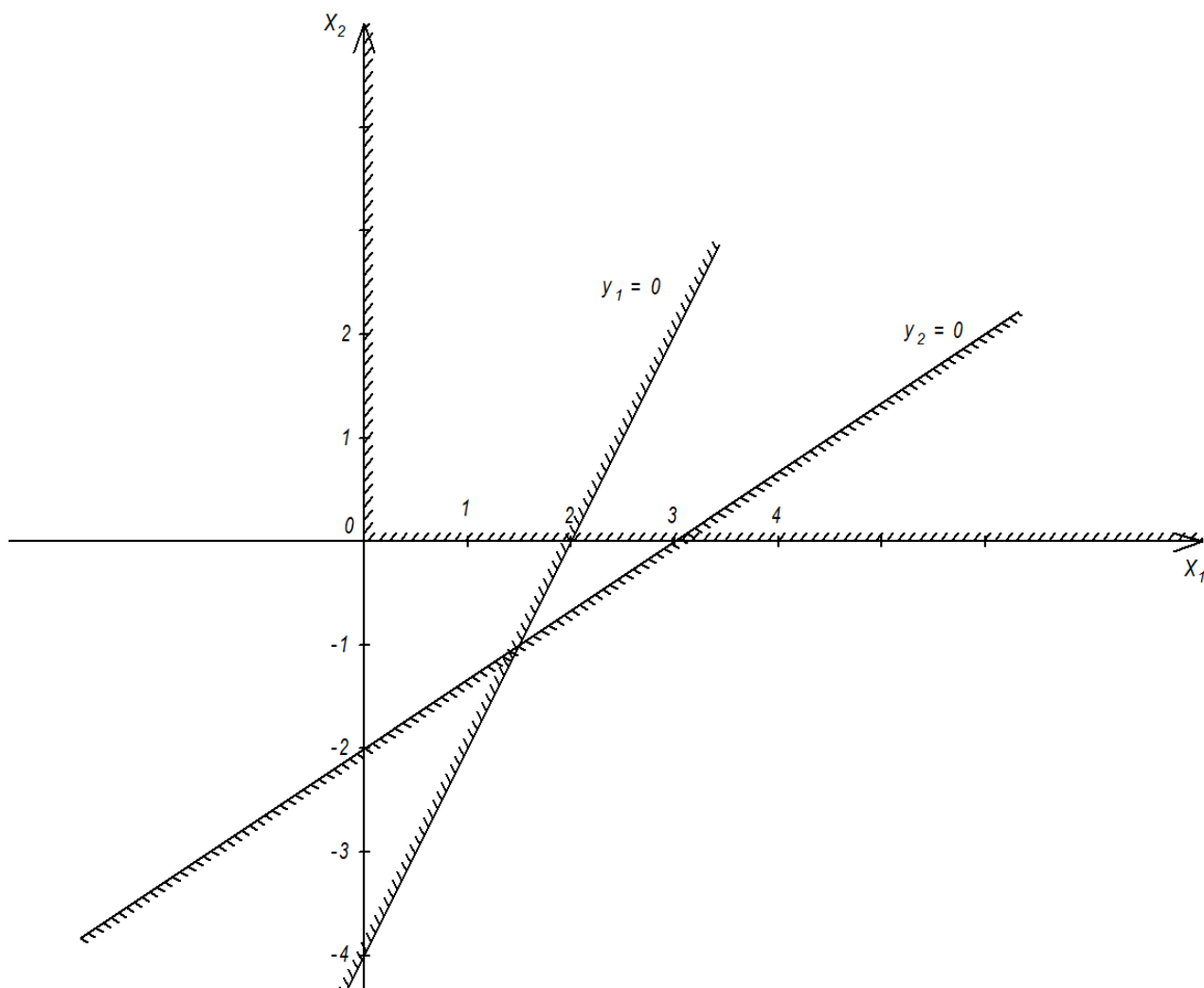


Рисунок 3.5. Геометрическая интерпретация случая несовместимости условий задачи ЛП

Первой будет достигнута ближайшая граница $y_1 = 0$, ей соответствует наименьшее из отношений $\frac{b_i}{\alpha_{i1}} = 2$. Симплексная таблица, полученная после одного шага исключений, определяет базисное решение в вершине с координатами $x_1 = 2, x_2 = 0$. Дальнейшее движение из этой вершины к границе $y_2 = 0$ лишено смысла, так как ОДР в этой задаче не существует, условия задачи несовместны. Признаком этого является положительность элементов во второй строке симплексной таблицы.

Пример 3.5. Задача ЛП задана в виде:

$$\begin{aligned}
 F &= 3x_1 + 5x_2 \rightarrow \max, \\
 -x_1 + 2x_2 &\geq 2, \\
 2x_1 + 2x_2 &\geq 6, \\
 -5x_1 + 3x_2 &\leq 15, \\
 9x_1 + 5x_2 &\leq 45, \\
 x_j &\geq 0; j = 1, 2.
 \end{aligned}$$

После приведения задачи к канонической форме составим первую симплексную таблицу:

$$F = 3x_1 + 5x_2 \rightarrow \max,$$

$$y_1 = -x_1 + 2x_2 - 2 \geq 0,$$

$$y_2 = 2x_1 + 2x_2 - 6 \geq 0,$$

$$y_3 = 5x_1 - 3x_2 + 15 \geq 0,$$

$$y_4 = -9x_1 - 5x_2 + 45 \geq 0,$$

$$x_j \geq 0; j = 1, 2.$$

	$-x_1$	$-x_2$	1	b_i/α_{i2}
y_1	1	-2	-2	1
y_2	-2	-2	-6	3
y_3	-5	3	15	5
y_4	9	5	45	9
F	-3	-5	0	

В первой и второй строках таблицы свободные члены отрицательны. Это означает, что базисное решение, соответствующее симплексной таблице, не принадлежит ОДР. Из рисунка 3.6 видно, что вершина, в которой находится это базисное решение, с координатами $x_1 = 0$, $x_2 = 0$ отделена от ОДР границами $y_1 = 0$ и $y_2 = 0$. Чтобы приблизиться к границе $y_1 = 0$, следует увеличивать x_2 . Для этого примем второй столбец симплексной таблицы, содержащий в первой строке отрицательный элемент -2 за разрешающий, вычислим отношения $\frac{b_i}{\alpha_{i2}} \geq 0$, выберем первую строку с наименьшим отношением $\frac{b_1}{\alpha_{12}} = 1$ за разрешающую. Выполнив один шаг модифицированных жордановых исключений с разрешающим элементом $\alpha_{12} = -2$, получим новую симплексную таблицу, в которой вместо переменной y_1 в базисное множество вошла переменная x_2 .

Новой таблице соответствует базисное решение, расположенное в вершине А с координатами $x_1 = 0$, $x_2 = 1$. Вершина А образована пересечением границ $x_1 = 0$ и $y_1 = 0$ и отделена от ОДР границей $y_2 = 0$. Чтобы из точки А попасть в опорную вершину, следует двигаться в направлении y_1 или x_1 . Для этого выберем в качестве разрешающего второй столбец симплексной таблицы (можно и первый, так как оба столбца содержат во второй строке отрицательные элементы). Вычислив наименьшее из отношений $\frac{b_2}{\alpha_{22}} = 4$, сделаем один шаг модифицированных жордановых исключений с разрешающим элементом $\alpha_{22} = -1$. Получим симплексную таблицу, соответствующую опорному решению в вершине В ОДР с координатами $x_1 = 0$; $x_2 = 3$:

	$-x_1$	$-y_2$	1
x_2	1	-0,5	3
y_1	3	-1	4
y_3	-8	1,5	6
y_4	4	2,5	30
F	2	-2,5	15

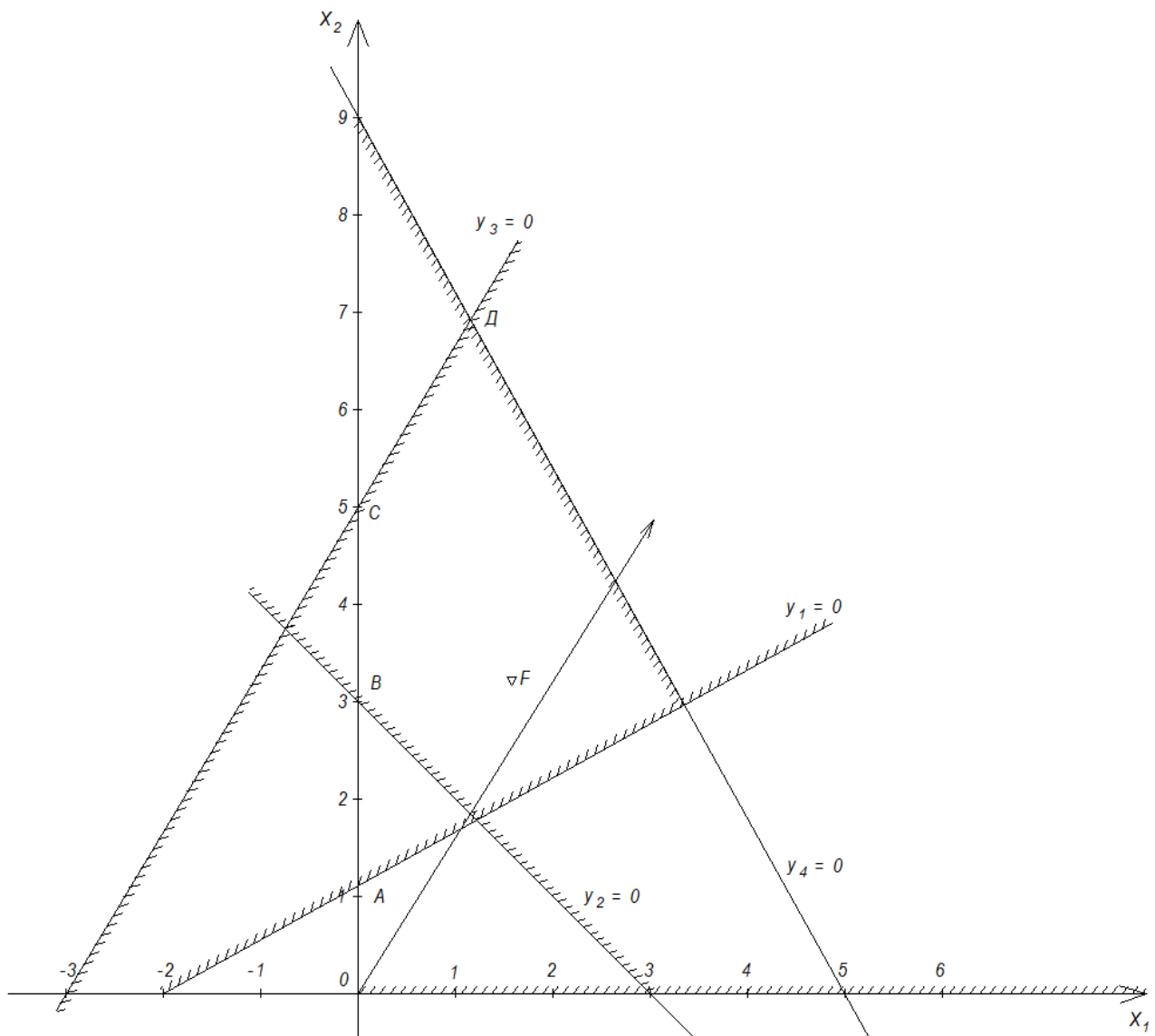


Рисунок 3.6. Геометрическая интерпретация задачи ЛП к примеру 3.5

Таким образом, чтобы попасть в опорную вершину, совершаются последовательно переходы из одной вершины в соседнюю с ней, расположенную ближе к ОДР.

Пример 3.6. Задача ЛП задана в виде

$$F = 3x_1 + x_2 \rightarrow \max,$$

$$x_1 - x_2 \geq 0,$$

$$y_2 = 2x_1 + 2x_2 - 6 \geq 0,$$

$$y_3 = -10x_1 - 5x_2 + 40 \geq 0,$$

$$x_j \geq 0; j = 1, 2.$$

Каноническая форма задачи и соответствующая ей симплексная таблица приведены ниже:

$$F = 3x_1 + x_2 \rightarrow \max,$$

$$y_1 = x_1 - x_2 \geq 0,$$

$$y_2 = 2x_1 + 2x_2 - 6 \geq 0,$$

$$y_3 = -10x_1 - 5x_2 + 40 \geq 0,$$

$$x_j \geq 0; j = 1, 2.$$

	$-x_1$	$-x_2$	1	$b_i/\alpha_{i1} \geq 0$	$b_i/\alpha_{i2} \geq 0$
y_1	-1	1	0		0
y_2	-2	-2	-6	3	3
y_3	10	5	40	4	8
F	-3	-1	0		

В этом примере мы впервые сталкиваемся со случаем вырождения, когда один из свободных членов $b_1 = 0$. Это означает, что в вершине 0 (рис. 3.7), в которой находится соответствующее исходной симплексной таблице базисное решение, число пересекающихся границ больше, чем число независимых переменных. Для получения опорного решения можно двигаться из точки 0 к определяющей ее от ОДР границе $y_2 = 0$ в направлении любой из осей x_1 и x_2 . Это видно и из симплексной таблицы, где любой из столбцов может быть разрешающим, так как во второй строке оба элемента отрицательны. Если за разрешающий взять первый столбец, т.е. движение по оси x_1 , то при этом граница $y_1 = 0$ не будет нарушена, так как увеличение x_1 приведет к увеличению y_1 . Поэтому в симплексной таблице отношение

$$\frac{b_1}{\alpha_{11}} = 0$$

не является лимитирующим при выборе разрешающего элемента, которым будет $\alpha_{21} = -2$. Если разрешающим будет второй столбец, то движение из точки 0 в направлении возрастания x_2 немедленно приведет к нарушению границы $y_1 = 0$, поэтому отношение

$$\frac{b_1}{\alpha_{12}} = 0$$

необходимо учесть, и разрешающий элемент в этом случае выбирается в первой строке – это $\alpha_{12} = 1$.

Один шаг модифицированных жордановых исключений с разрешающим элементом $\alpha_{21} = -2$ позволяет получить опорное решение (вершина В на рисунке 3.7); соответствующая этому решению таблица имеет вид

	$-y_2$	$-x_2$	1
y_1	-0,5	2	3
x_1	-0,5	1	3
y_3	5	-5	10
F	-4	1	0

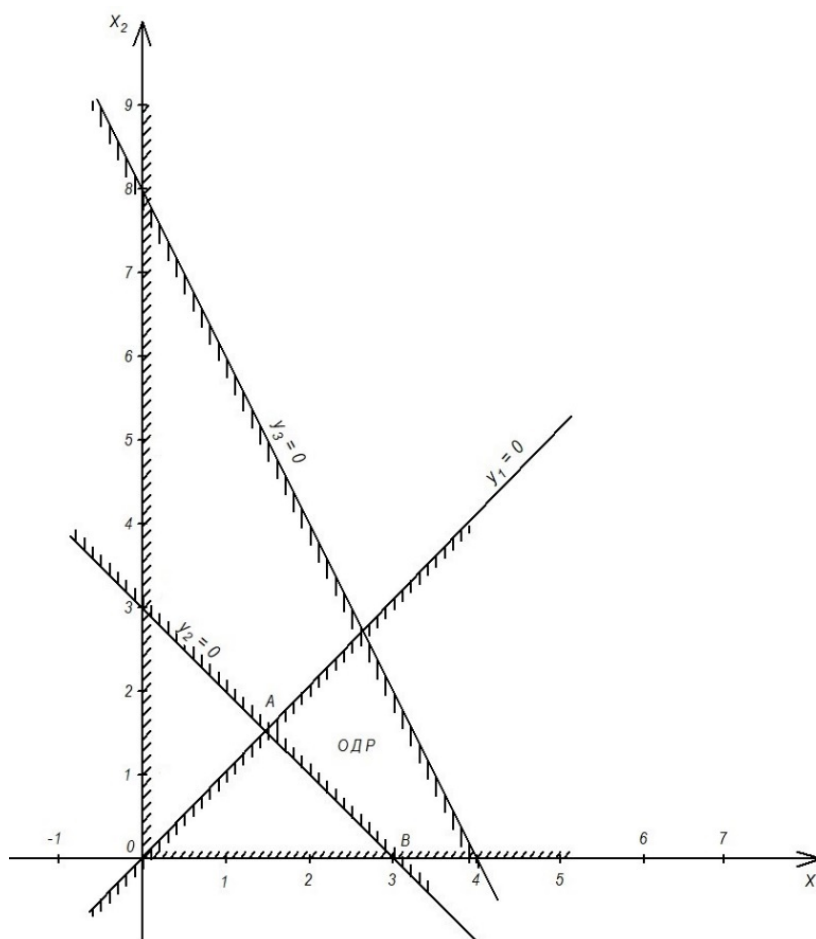


Рисунок 3.7. Геометрическая интерпретация задачи ЛП к примеру 3.6

Один шаг исключений с разрешающим элементом $\alpha_{12} = 1$ приводит к симплексной таблице с базисными переменными x_2, y_2, y_3 (вместо y_1, y_2, y_3 в исходной таблице), однако базисное решение принадлежит той же вершине 0, что и исходное.

	$-x_1$	$-x_2$	1	$b_i/\alpha_{i1} \geq 0$
y_1	-1	1	0	
y_2	-4	2	-6	1,5
y_3	15	-5	40	2,66
F	-4	-1	0	

Таким образом, в случае вырождения в одной вершине может находиться несколько базисных решений. Опорное решение получим, сделав еще один шаг модифицированных жордановых исключений с разрешающим элементом $\alpha_{21} = -4$:

	$-y_2$	$-y_1$	1
x_2	-0,25	0,25	1,5
x_1	-0,25	-0,5	1,5
y_3	3,75	2,5	17,5
F	-1	-1	6

Опорному решению соответствует вершина А ОДР.

II этап. Нахождение оптимального решения задачи ЛП

Пусть в результате 1 этапа получена симплексная таблица, которой соответствует опорное решение:

	$-x_1$	...	$-x_s$	...	$-x_n$	1
x_{n+1}	α_{11}	...	α_{1s}	...	α_{1n}	b_1
x_{n+2}	α_{21}	...	α_{2s}	...	α_{2n}	b_2
$\vdots$	...	...	...	...	...	...
x_{n+r}	α_{r1}	...	α_{rs}	...	α_{rn}	b_r
$\vdots$	...	...	...	...	...	...
x_{n+m}	α_{m1}	...	α_{ms}	...	α_{mn}	b_m
F	c_1	...	c_2	...	c_n	F_0

(3.24)

Для удобства записи не будем делать различий в обозначении искусственных и физических переменных.

Сформулируем признак оптимальности опорного решения. С этой целью обратимся к F-строке симплексной таблицы (3.24). Если все элементы F-строки $c_j \geq 0, j = 1, \dots, n$, то опорное решение

$$x_1 = x_2 = \dots = x_n = 0; x_{n+1} = b_1, \dots, x_{n+m} = b_m \quad (3.25)$$

является оптимальным, т.е. линейная форма F достигает на этом решении максимума. Действительно для этого опорного решения

$$F = -c_1x_1 - c_2x_2 - \dots - c_sx_s - \dots - c_nx_n + F_0 \leq F_0, \quad (3.26)$$

что выполняется при любых значениях независимых переменных $x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$.

В случае $x_1 > 0, x_2 > 0, \dots, x_n > 0, F < F_0$ при $x_1 = x_2 = \dots = x_n = 0, F = F_0$, следовательно, F_0 – максимальное значение линейной формы F.

Здесь могут представиться два случая:

- все элементы F-строки $c_j > 0, j = 1, \dots, n$, тогда решение задачи ЛП единственно;
- среди элементов F-строки имеются нулевые $c_j \geq 0$, тогда задача имеет бесчисленное множество решений, так как независимые переменные, для которых $c_j = 0$ могут принимать различные произвольные значения и при этом величина линейной формы F (3.26) не изменится.

Если среди элементов F-строки имеется по крайней мере один отрицательный, то опорное решение (3.25) неоптимально и может быть улучшено.

Пусть, например, $c_s < 0$, тогда

$$F = -c_1x_1 - c_2x_2 - \dots + c_sx_s - \dots - c_nx_n + F_0.$$

Очевидно, что с увеличением x_s линейная форма F будет возрастать, следовательно, существует лучшее опорное решение, которому соответствует большее значение линейной формы F.

Посмотрим, до какого значения можно увеличивать x_s . С ростом x_s базисные переменные $x_{n+1}, x_{n+2}, \dots, x_{n+m}$ начнут изменяться, увеличиваясь или уменьшаясь в зависимости от знака α_{is} :

$$x_i = -\alpha_{is}x_s + b_i \geq 0. \quad (3.27)$$

Если $\alpha_{is} < 0$, то при любом увеличении x_s $x_i > 0$; в случае, когда все $\alpha_{is} < 0, i = 1, \dots, m, x_s \rightarrow \infty$ и $F \rightarrow \infty$, т.е. линейная форма не ограничена сверху, и задача решения не имеет. Лимитирующими являются ограничения, в которых $\alpha_{is} > 0$, тогда с ростом x_s базисная переменная x_i будет уменьшаться. Очевидно, что x_s можно увеличивать до тех пор, пока какая-либо из базисных переменных не станет равна нулю, на основании соотношения (3.27) найдем, что

$$x_s = \min_{\alpha_{is} > 0} \left\{ \frac{b_i}{\alpha_{is}} \right\}.$$

Одна итерация алгоритма поиска оптимального решения включает следующие операции.

1. Анализируются элементы F-строки и за разрешающий принимается тот столбец, в котором $c_j < 0$. Обычно, если имеется несколько $c_j < 0$, среди них выбирается

$$c_s = \max_{c_j < 0} \{c_j\}$$

и S столбец становится разрешающим. Это позволяет найти максимум F за меньшее число итераций.

Более обоснованно выбор разрешающего столбца производится следующим образом:

- для всех $c_j < 0$ вычисляется наименьшее из отношений

$$\frac{b_r}{\alpha_{rj}} = \min_i \left\{ \frac{b_i}{\alpha_{ij}} \geq 0 \right\}; \quad (3.28)$$

- вычисляется приращение F

$$\Delta F = -c_j \frac{b_r}{b_{rj}};$$

- за разрешающий принимается S столбец, для которого

$$c_s \frac{b_r}{\alpha_{rs}} = \max_{c_j < 0} \left\{ c_j \frac{b_r}{\alpha_{rj}} \right\}.$$

При таком выборе разрешающего столбца гарантируется сходимость алгоритма за минимальное число итераций.

2. Вычисляются неотрицательные отношения свободных членов к элементам разрешающего столбца $\frac{b_i}{\alpha_{is}}$ ($\alpha_{is} > 0$) и выбирается наименьшее из отношений (см. (3.28)); r -строка принимается за разрешающую. Если все $\alpha_{is} < 0, i = 1, \dots, m$, то делается вывод, что оптимального решения не существует, так как линейная форма не ограничена сверху.

3. Выполняется шаг модифицированных жордановых исключений с разрешающим элементом α_{rs} .

Все операции алгоритма повторяются до тех пор, пока в F -строке не останется отрицательных элементов, либо не будет установлен факт отсутствия оптимального решения.

Конечность алгоритма симплексного метода

В предположении отсутствия вырожденности симплексный алгоритм сходится за конечное число итераций (либо устанавливается факт отсутствия оптимального решения). Конечность алгоритма симплексного метода вытекает из конечного числа опорных вершин ОДР. Способ выбора разрешающего столбца гарантирует переход в соседнюю опорную вершину с наибольшим значением линейной формы F . Таким образом, на каждой итерации линейная форма F неуклонно возрастает, а так как каждой опорной вершине соответствует единственное значение линейной формы F , то ни одна из опорных вершин не будет пройдена дважды.

Иначе обстоит дело в случае вырождения. Как показано на рисунке 3.7, в вершине O пересекаются три границы: $x_1 = 0, x_2 = 0, y_1 = 0$, хотя число независимых переменных равно двум. В результате одной вершине соответствует несколько базисных решений с одним и тем же значением линейной формы F . Поэтому после шага исключений может оказаться, что несмотря на замену одной из базисных переменных, значение F осталось прежним. Таким образом монотонность возрастания F нарушается. Из-за того, что одному и тому же значению F соответствует несколько опорных решений, возникает опасность закливания: повторения одних и тех же опорных решений. Закливание встречается довольно редко, поэтому способы избавления от него здесь не рассматриваются. Обычно после нескольких шагов удается выйти из вершины, которой соответствует несколько опорных решений.

Пример 3.7. В примере 3.5 за две итерации было получено опорное решение задачи ЛП, геометрическая интерпретация которой дана на рисунке 3.6. Найдем оптимальное решение этой задачи.

Перепишем соответствующую опорному решению симплексную таблицу:

	$-x_1$	$-x_2$	1	$b_i/\alpha_{i2} \geq 0$
x_2	1	-0,5	3	
y_1	3	-1	4	
y_3	-8	1,5	6	4
y_4	4	2,5	30	12
F	2	-2,5	15	

В F-строке таблицы имеется отрицательный элемент $c_2 = -2,5$, поэтому второй столбец принимаем за разрешающий. Вычислим отношения $b_i/\alpha_{i2} \geq 0$, наименьшее из отношений

$$\frac{b_3}{\alpha_{32}} = 4,$$

третья строка принимается за разрешающую. После шага исключений с разрешающим элементом $\alpha_{32} = 1,5$ получим новую симплексную таблицу, в которой вместо переменной y_3 в базис введена переменная y_2 :

	$-x_1$	$-y_3$	1	$b_i/\alpha_{i1} \geq 0$
x_2	-1,67	3,33	5	3
y_1	-2,33	0,67	8	
y_2	5,33	0,67	4	
y_4	17,33	-1,67	20	1,15
F	-11,33	1,67	25	

Симплексной таблице соответствует опорное решение, расположенное в вершине C с координатами $x_1 = 0, x_2 = 5$. В F-строке этой таблицы имеется отрицательный элемент $c = -11,33$, первый столбец делаем разрешающим, находим в этом столбце разрешающий элемент $\alpha_{41} = 17,33$ и выполняем еще один шаг исключений. Новая симплексная таблица с базисной переменной x_1 вместо y_4 не содержит в F-строке отрицательных элементов:

	$-y_4$	$-y_3$	1
x_2	0,96	3	7
y_1	0,13	0,4	10,7
y_2	0,3	0,15	10,15
x_1	0,05	-0,1	1,15
F	0,65	0,58	38

Следовательно, опорное решение

$$y_1 = y_3 = 0; x_2 = 7; x_1 = 1,15; y_1 = 10,7; y_2 = 10,15$$

является оптимальным, максимум $F=38$.

Оптимальному решению соответствует вершина D ОДР. Найденное решение единственно, так как все элементы F-строки строго положительны.

Пример 3.8. Решить задачу ЛП:

$$\begin{aligned} F &= 2x_1 + 4x_2 \rightarrow \max, \\ -2x_1 + x_2 &\leq 4, \\ x_1 - 2x_2 &\leq 4, \\ x_1 + 2x_2 &\leq 10, \\ x_j &\geq 0, j = 1, 2. \end{aligned}$$

Приводим задачу к канонической форме и составляем первую симплексную таблицу:

$$F = 2x_1 + 4x_2 \rightarrow \max,$$

$$y_1 = 2x_1 - x_2 + 4 \geq 0,$$

$$y_2 = -x_1 + 2x_2 + 4 \geq 0,$$

$$y_3 = -x_1 - 2x_2 + 10 \geq 0$$

$$x_j \geq 0; j = 1, 2.$$

	$-x_1$	$-x_2$	1	$b_i/\alpha_{i2} \geq 0$
y_1	-2	1	4	4
y_2	1	-2	4	
y_3	1	2	10	5
F	-2	-4	0	

Так как все элементы столбца свободных членов симплексной таблицы положительны, то приравняв независимые переменные нулю: $x_1 = x_2 = 0$, получим опорное решение:

$$y_1 = 4; y_2 = 4; y_3 = 10; F = 0,$$

расположенное в начале координат рисунка 3.8. Проанализировав элементы F- строки, убеждаемся, что все они отрицательны, следовательно, полученное опорное решение неоптимально. Находим

$$c_s = \max_{c_j < 0} |c_j| = c_2 = -4$$

и второй столбец принимаем за разрешающий. Находим наименьшее из соотношений

$$\frac{b_2}{\alpha_{r2}} = \min_{\alpha_{i2} > 0} \left\{ \frac{b_i}{\alpha_{i2}} \geq 0 \right\} = \frac{b_1}{\alpha_{12}} = 4$$

и выполняем шаг модифицированных жордановых исключений с разрешающим элементом $\alpha_{12} = 1$, получим симплексную таблицу с базисной x_2 переменной вместо y_1 :

	$-x_1$	$-y_1$	1	$b_i/\alpha_{i1} \geq 0$
x_2	-2	1	4	
y_2	-1	2	12	
y_3	5	-2	2	2/5
F	-10	4	16	

Приравняв независимые переменные нулю, получим опорное решение

$$x_1 = y_1 = 0; x_2 = 4; y_2 = 12; y_3 = 2; F = 16,$$

расположенное в вершине В ОДР (рисунок 3.8).

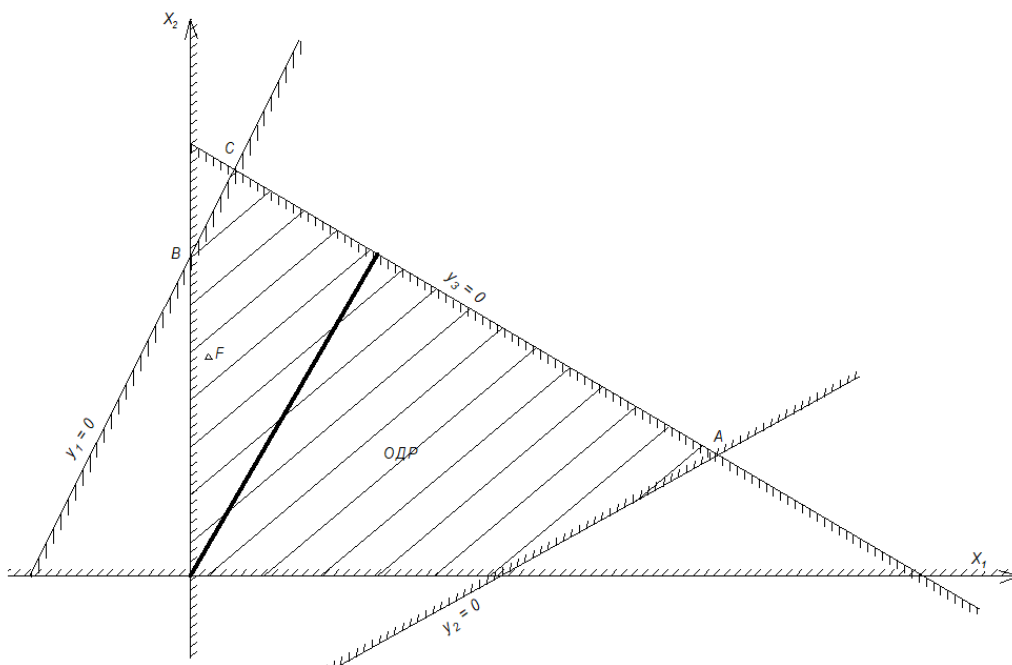


Рисунок 3.8. Геометрическая интерпретация задачи ЛП к примеру 3.8

Повторяем операции алгоритма отыскания оптимального решения. Находим в F-строке отрицательный элемент $c = -10$ и принимаем первый столбец за

разрешающий. Очевидно, что разрешающая строка третья, так как только в этой строке имеется положительный элемент $\alpha_{31} = 5$. После шага исключений с разрешающим элементом $\alpha_{31} = 5$ получим симплексную таблицу с базисной переменной x_1 вместо y_3 :

	$-y_3$	$-y_1$	1
x_2	0,4	0,2	4,8
y_2	0,2	1,6	12,4
x_1	0,2	-0,4	0,4
F	2	0	22

Приравняв независимые переменные нулю, найдем оптимальное опорное решение:

$$y_3 = y_1 = 0; x_2 = 4,8; y_2 = 12,4; x_1 = 0,4; F_{\max} = 22.$$

Так как элементы F-строки симплексной таблицы неотрицательны, то задача имеет бесчисленное множество решений, расположенных на отрезке АС ОДР.

Пример 3.9. Решить задачу ЛП:

$$\begin{aligned} F &= 3x_1 + 4x_2 \rightarrow \max, \\ -4x_1 + 5x_2 &\leq 20, \\ -3x_1 + 2x_2 &\leq 6, \\ 2x_1 - 4x_2 &\leq 8, \\ x_j &\geq 0, j = 1, 2. \end{aligned}$$

Приводим задачу к канонической форме и составляем первую симплексную таблицу:

$$F = 3x_1 + 4x_2 \rightarrow \max,$$

$$y_1 = 4x_1 - 5x_2 + 20 \geq 0,$$

$$y_2 = 3x_1 - 2x_2 + 6 \geq 0,$$

$$y_3 = -2x_1 + 4x_2 + 8 \geq 0$$

$$x_j \geq 0; j = 1, 2.$$

	$-x_1$	$-x_2$	1	$b_i/\alpha_{i2} \geq 0$
y_1	-4	5	20	4
y_2	-3	2	6	3
y_3	2	-4	8	
F	-3	-4	0	

Симплексной таблице соответствует опорное решение

$$x_1 = x_2 = 0; y_1 = 20; y_2 = 6; y_3 = 8,$$

расположенное в начале координат (рис. 3.9). Найдем оптимальное решение задачи ЛП. Выберем второй столбец, в котором находится $c_2 = \max_{c_j < 0} |c_j| = -4$, разрешающим, вычислим наименьшее из отношений

$$\frac{b_2}{\alpha_{22}} = \min \left\{ \frac{b_i}{\alpha_{i2}} \geq 0 \right\} = 3$$

и сделаем шаг модифицированных жордановых исключений с разрешающим элементом $\alpha_{22} = 2$. Получим симплексную таблицу с базисной переменной x_2 вместо y_2 :

	$-x_1$	$-y_2$	1	$b_i/\alpha_{i1} \geq 0$
y_1	3,5	-2,5	5	1,43
x_2	-1,5	0,5	3	
y_3	-5	2	20	
F	-9	2	12	

Опорное решение, которое дает симплексная таблица:

$$x_1 = y_2 = 0; y_1 = 5; x_2 = 3; y_3 = 20$$

находится в вершине В ОДР.

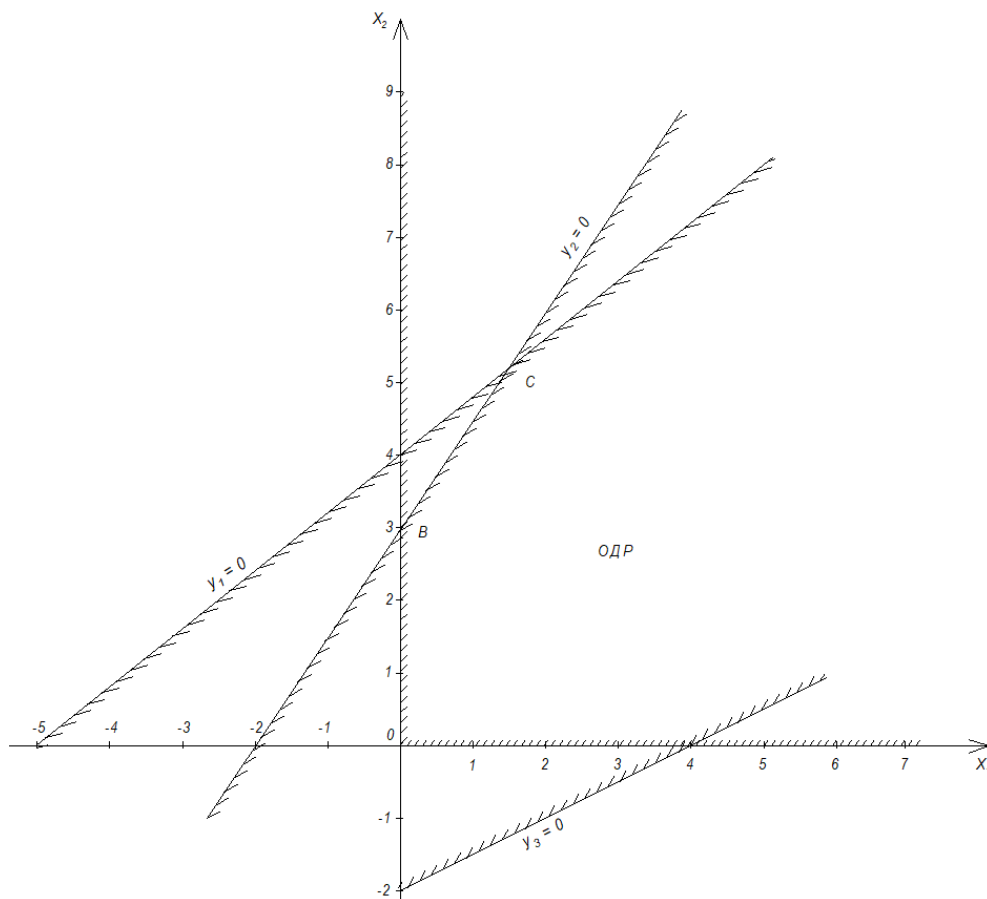


Рисунок 3.9. Геометрическая интерпретация задачи ЛП к примеру 3.9

Повторяем все операции алгоритма поиска оптимального решения: выбираем первый столбец разрешающим и, делая шаг модифицированных жордановых исключений с разрешающим элементом $\alpha_{11} = 3,5$, получим таблицу:

	$-y_1$	$-y_2$	1
x_1	0,28	-0,71	1,43
x_2	0,42	-0,57	5,1
y_3	1,43	-1,57	27,1
F	2,57	-4,43	24,8

Опорное решение, соответствующее таблице,

$$y_1 = y_2 = 0; x_1 = 1,43; x_2 = 5,1; y_3 = 27,1$$

находится в вершине С ОДР.

В F-строке симплексной таблицы имеется отрицательный элемент $c_2 = -4,43$, поэтому опорное решение оптимально, но так как в разрешающем столбце нет ни одного положительного элемента, делаем вывод об отсутствии оптимального решения из-за неограниченности сверху линейной формы F.

3.4. Модификация симплексного алгоритма для задачи ЛП с двусторонними ограничениями

Во многих прикладных задачах ЛП, связанных с выбором оптимальных технологических режимов, оптимальным планированием, распределением ограниченных ресурсов и т.д., допустимый диапазон изменения переменных задается либо ГОСТом, либо технологическим регламентом. В связи с этим приходится решать задачу ЛП с двусторонними ограничениями:

$$F = \sum_{j=1}^n c_j x_j \rightarrow \max, \quad (3.29)$$

$$\gamma_i^* \leq y_i = \sum_{j=1}^n \alpha_{ij}(-x_j) + b_i \leq \beta_i^*, \quad i = 1, \dots, m, \quad (3.30)$$

$$\gamma_j \leq x_j \leq \beta_j, \quad j = 1, \dots, n. \quad (3.31)$$

Ограничения (3.30), (3.31) определяют некоторый выпуклый многогранник – ОДР, где $\gamma_i^*, \gamma_j, \beta_i^*, \beta_j$ – заданные числа, некоторые из них могут быть и бесконечными. Если ограничения (3.30), (3.31) расчленим на $2m+2n$ неравенств, то получим обычную задачу ЛП, но размерность задачи при этом чрезвычайно увеличивается.

Поэтому для решения задач ЛП с двусторонними ограничениями разработан усложненный вариант симплекс-метода [8].

Для условий задачи (3.29), (3.30) составляется симплексная таблица:

	$-x_1 \quad \dots \quad -x_s \quad \dots \quad -x_n$	1
y_1	$\alpha_{11} \quad \dots \quad \alpha_{1s} \quad \dots \quad \alpha_{1n}$	b_1
$\vdots$	$\dots \quad \dots \quad \dots \quad \dots \quad \dots$	$\dots$
y_r	$\alpha_{r1} \quad \dots \quad \alpha_{rs} \quad \dots \quad \alpha_{rn}$	b_r
$\vdots$	$\dots \quad \dots \quad \dots \quad \dots \quad \dots$	$\dots$
y_m	$\alpha_{m1} \quad \dots \quad \alpha_{ms} \quad \dots \quad \alpha_{mn}$	b_m
F	$-C_1 \quad \dots \quad -C_s \quad \dots \quad -C_n$	

(3.32)

Будем считать, что в задаче нет свободных координат, т.е. для каждой переменной x_j , $j = 1, \dots, n$ по крайней мере одно из чисел γ_j или β_j конечно.

Выводим независимые переменные (записанные наверху таблицы (3.32)) на их верхние или нижние границы, т.е. заменяем x_j на x_j^+ или на x_j^- :

$$x_j^+ = \beta_j - x_j, \quad (3.33)$$

$$x_j^- = x_j - \gamma_j. \quad (3.34)$$

Из (3.31) получим, что

$$\begin{aligned} 0 &\leq x_j^+ \leq \beta_j - \gamma_j, \\ 0 &\leq x_j^- \leq \beta_j - \gamma_j. \end{aligned} \quad (3.35)$$

Таким образом x_j^+ (x_j^-) равняется нулю, когда x_j принимает значение на верхней (на нижней) границе. При этом симплексная таблица (3.32) преобразуется по следующему правилу:

Если независимая переменная x_s выводится на нижнюю границу, то из столбца свободных членов вычитается S столбец, все элементы которого предварительно умножаются на γ_s :

$$\tilde{b}_i = b_i - \alpha_{is}\gamma_s, i = 1, \dots, m.$$

Если же x_s выводится на верхнюю границу, то из столбца свободных членов вычитается S столбец, все элементы которого умножаются на β_s , и знаки элементов S столбца меняются на противоположные:

$$\begin{aligned} \tilde{b}_i &= b_i - \alpha_{is}\beta_s; \\ \tilde{\alpha}_{is} &= -\alpha_{is}, i = 1, \dots, m. \end{aligned}$$

Действительно, если переменная x_s выводится на нижнюю границу согласно (3.34), то очевидно, что ограничения y_i не должны быть нарушены, а для этого каждое i уравнение преобразуется следующим образом:

$$\begin{aligned}
y_i &= \sum_{j=1, j \neq s}^n \alpha_{ij}(-x_j) + \alpha_{is}(-x_s) + b_i = \sum_{j=1, j \neq s}^n \alpha_{ij}(-x_j) + \alpha_{is}(-(x_s - \gamma_s)) + b_i - \alpha_{is}\gamma_s = \\
&= \sum_{j=1, j \neq s}^n \alpha_{ij}(-x_j) + \alpha_{is}x_s^- + \tilde{b}_i.
\end{aligned}$$

Аналогично, если x_s выводится на верхнюю границу согласно (3.33), то каждое i уравнение также преобразуется:

$$\begin{aligned}
y_i &= \sum_{j=1, j \neq s}^n \alpha_{ij}(-x_j) + \alpha_{is}(-x_s) + b_i = \sum_{j=1, j \neq s}^n \alpha_{ij}(-x_j) + \alpha_{is}(\beta_s - x_s) + b_i - \alpha_{is}\beta_s = \\
&= \sum_{j=1, j \neq s}^n \alpha_{ij}(-x_j) - \alpha_{is}(-x_s^+) + \tilde{b}_i.
\end{aligned}$$

Выведение независимых переменных на их верхние или нижние границы означает замену случайно взятых координатных плоскостей $x_j = 0, j = 1, \dots, n$, граничными координатными плоскостями

$$\begin{aligned}
x_j^- &= x_j - \gamma_j = 0; \\
x_j^+ &= \beta_j - x_j = 0.
\end{aligned}$$

Пусть все независимые переменные x_j выведены на их верхние или нижние границы и получена симплексная таблица:

	$-\tilde{x}_1$	$\dots$	$-\tilde{x}_s$	$\dots$	$-\tilde{x}_n$	1
y_1	α_{11}	$\dots$	α_{1s}	$\dots$	α_{1n}	α_1
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_r	α_{r1}	$\dots$	α_{rs}	$\dots$	α_{rn}	α_r
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_m	α_{m1}	$\dots$	α_{ms}	$\dots$	α_{mn}	α_m
F	$-p_1$	$\dots$	$-p_s$	$\dots$	$-p_n$	F_0

(3.36)

где $\tilde{x}_j$ – независимая переменная, выведенная на верхнюю или нижнюю границу; $\alpha_{1j}, \alpha_i, p_j, F_0$ – коэффициенты и свободные члены уравнений и линейной формы F после приведения переменных x_j к нижним или верхним границам.

Преобразование симплексной таблицы (3.36) с целью замены какой-либо базисной переменной (например, y_r) на $\tilde{x}_s$ осуществляется с помощью усложненного шага модифицированных жордановых исключений, алгоритм которого дополняется следующими операциями:

1. Новая независимая переменная y_r выводится на одну из своих границ.
2. Новая базисная переменная $\tilde{x}_s$ заменяется переменной x_s , при этом, если $\tilde{x}_s = x_s^-$, то к новому свободному члену r разрешающей строки прибавляется γ_s ; если же $\tilde{x}_s = x_s^+$, то строка умножается на -1 и к новому свободному члену прибавляется β_s .

I этап модифицированного симплекс-алгоритма: отыскание опорного решения

1. В симплексной таблице (3.36) анализируется столбец свободных членов, каждый элемент столбца α_i сравнивается с нижней γ_i и верхней β_i^* границами. Возможны три случая:

$$1) \gamma_i \leq \alpha_i \leq \beta_i^*, i = 1, \dots, m, \quad (3.37)$$

что соответствует опорному решению, расположенному в начале координат;

$$2) \alpha_r < \gamma_r^*;$$

$$3) \alpha_r > \beta_r^*.$$

Последние два случая означают, что для r ограничения нарушена нижняя (верхняя) граница. Тогда в r строке отыскивается в случае нарушения нижней границы – отрицательный элемент; а при нарушении верхней границы – положительный элемент. Пусть, например, это элемент α_{rs} :

$$\alpha_{rs} < 0 \text{ в случае } \alpha_r < \gamma_r^*,$$

$$\alpha_{rs} > 0 \text{ в случае } \alpha_r > \beta_r^*.$$

Тогда S столбец принимается за разрешающий. Если в r строке нет ни одного отрицательного (положительного) элемента, то условия задачи несовместны.

2. Вычисляются неотрицательные отношения $\frac{\alpha_i - \gamma_i^*}{\alpha_{is}} \geq 0$ и выбирается наименьшее из отношений: $t_1 = \min_{\alpha_{is} > 0} \left\{ \frac{\alpha_i - \gamma_i^*}{\alpha_{is}} \geq 0 \right\}$.

Вычисляются неотрицательные отношения $\frac{\beta_i^* - \alpha_i}{-\alpha_{is}} \geq 0$ и выбирается

$$t_2 = \min_{\alpha_{is} > 0} \left\{ \frac{\beta_i^* - \alpha_i}{-\alpha_{is}} \geq 0 \right\}.$$

Вычисляется $t_3 = \beta_s - \gamma_s$.

3. Определяется $t = \min\{t_1, t_2, t_3\}$.

Если $t = t_1$ ($t = t_2$), то в качестве разрешающего берется тот из элементов S столбца, на котором достигается t_1 (t_2), например α_{ks} . В случае вырождения, когда $t = t_1 = 0$ ($t = t_2 = 0$), соответствующий коэффициент S столбца принимается за разрешающий лишь тогда, когда он положителен (отрицателен).

Если же $t = t_3$, то координата $\tilde{x}_s$ переводится на другую границу: с верхней на нижнюю или наоборот, и операции 1–3 повторяются.

4. Выполняется усложненный шаг модифицированных жордановых исключений с разрешающим элементом α_{ks} .

Все операции алгоритма поиска опорного решения повторяются до тех пор, пока не будет получена таблица, для всех свободных членов которой

выполняются неравенства (3.37). Тогда, приравняв все независимые переменные нулю, получим опорное решение.

Дадим обоснование алгоритму поиска опорного решения.

Пусть $\alpha_r - \gamma_r^* < 0$ – это значит, что вершину с базисным решением, соответствующим таблице (3.36), отделяет от ОДР граница $y_r - \gamma_r^* = 0$. Чтобы уменьшить расстояние до границы, следует перейти в соседнюю вершину, расположенную ближе к ОДР. Для этого в уравнении $y_r - \gamma_r^* = \sum_{j=1}^n \alpha_{rj}(-\tilde{x}_j) + \alpha_r - \gamma_r^* = 0$ будем увеличивать ту из независимых переменных $\tilde{x}_j$, у которой в r строке симплексной таблицы отрицательный коэффициент, например $\alpha_{rs} < 0$. Увеличение x_s равносильно движению по оси $\tilde{x}_s$ до первой ближайшей границы. Если нарушена граница $\beta_r^* - y_r = 0$, т.е. $\beta_r^* - \alpha_r < 0$, то очевидно, что в уравнении $\beta_r^* - y_r = -\sum_{j=1}^n -\alpha_{rj}\tilde{x}_j + \beta_r^* - \alpha_r = 0$ следует увеличивать ту переменную $\tilde{x}_s$, которой в r строке симплексной таблицы (3.36) соответствует положительный коэффициент, например $\alpha_{rs} > 0$. В этом случае движение также будет в сторону увеличения $\tilde{x}_s$ до первой ближайшей границы. Такой границей может быть:

- 1) $y_q - \gamma_q^* = 0$;
- 2) $\beta_q^* - y_q = 0$;
- 3) $\tilde{x}_s = \beta_s - \gamma_s$.

В первом случае координата t_1 точки пересечения $\tilde{x}_s$ с границей $y_q - \gamma_q^* = 0$ определится следующим образом

$$t_1 = \frac{\alpha_q - \gamma_q^*}{\alpha_{qs}} = \min_{\alpha_{is} > 0} \left\{ \frac{\alpha_i - \gamma_i^*}{\alpha_{is}} \geq 0 \right\}.$$

Во втором случае координата t_2 точки пересечения $\tilde{x}_s$ с границей $\beta_q^* - y_q = 0$ равна

$$t_2 = \frac{\beta_q^* - \alpha_q}{-\alpha_{qs}} = \min_{\alpha_{is} > 0} \left\{ \frac{\beta_i^* - \alpha_i}{-\alpha_{is}} \geq 0 \right\}.$$

В третьем случае ближайшей является верхняя граница переменной $\tilde{x}_s$ $t_3 = \beta_s - \gamma_s$. Таким образом, ближайшая вершина имеет по оси $\tilde{x}_s$ координату

$$t = \min\{t_1, t_2, t_3\}.$$

Выбрав t , мы определим ближайшую границу, которую встретим при движении по оси $\tilde{x}_s$ в направлении ОДР. При этом, если $t = t_1$, то этой границей является $y_q^- = y_q - \gamma_q^* = 0$ и переменная y_q заменяется в множестве базисных переменных на x_s , становится независимой и выводится на нижнюю границу. Если же $t = t_2$, то такой границей является $y_q^+ = \beta_q^* - y_q = 0$, тогда x_s вводится в базисное множество, а y_q становится независимой вместо x_s и выводится на верхнюю границу. Если же $t = t_3$, то ближайшей будет граница $\tilde{x}_s = \beta_s - \gamma_s$,

поэтому $\tilde{x}_s$ переводится на верхнюю границу, если ранее она была на нижней (и наоборот).

II этап модифицированного симплексного алгоритма: отыскание оптимального решения задачи ЛП

Пусть опорное решение получено и ему соответствует таблица

	$-\tilde{x}_1$	$\dots$	$-\tilde{x}_s$	$\dots$	$-\tilde{x}_n$	1
x_{n+1}	α_{11}	$\dots$	α_{1s}	$\dots$	α_{1n}	α_1
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_{n+r}	α_{r1}	$\dots$	α_{rs}	$\dots$	α_{rn}	α_r
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_{n+m}	α_{m1}	$\dots$	α_{ms}	$\dots$	α_{mn}	α_m
F	p_1	$\dots$	p_s	$\dots$	p_n	F

(3.38)

в которой для удобства обозначений не делается различий между физическими и искусственными переменными. Свободные члены таблицы (3.38) удовлетворяют ограничениям:

$$\gamma_{n+i} \leq \alpha_i \leq \beta_{n+i}, \quad i = 1, \dots, m,$$

где индексы ограничений соответствуют индексам базисных переменных. Аналогично обычному симплекс-алгоритму опорное решение задачи ЛП с двусторонними ограничениями оптимально, если все коэффициенты F-строки $p_j \geq 0$.

Поэтому правило выбора разрешающего столбца остается прежним: в F-строке отыскивается отрицательный элемент, например $p_s < 0$, и S столбец принимается за разрешающий. После этого вычисляется:

$$\begin{aligned} t_1 &= \min_{\alpha_{is} < 0} \left\{ \frac{\alpha_i - \gamma_{n+i}}{\alpha_{is}} \geq 0 \right\}, \\ t_2 &= \min_{\alpha_{is} > 0} \left\{ \frac{\beta_{n+i} - \alpha_i}{-\alpha_{is}} \geq 0 \right\}, \\ t_3 &= \beta_s - \gamma_s. \end{aligned}$$

При этом, если β_s или γ_s бесконечно и, кроме того, $\gamma_{n+i} = -\infty$ для всех $\alpha_{is} > 0$ и $\beta_{n+i} = \infty$ для всех $\alpha_{is} < 0$, то линейная форма F не ограничена сверху.

Определяется первая ближайшая вершина при движении по оси $\tilde{x}_s$ и координата t этой вершины на оси $\tilde{x}_s$:

$$t = \min\{t_1, t_2, t_3\}.$$

Если $t = t_1$ или t_2 , то выполняется усложненный шаг модифицированных жордановых исключений, в результате которого совершается переход в

соседнюю опорную вершину с большим значением F , причем новая независимая переменная x_{n+r} переводится на свою нижнюю ($t = t_1$) или верхнюю ($t = t_2$) границу, новая базисная переменная $\tilde{x}_s$ приводится к старому обозначению x_s с соответствующим преобразованием γ -строки новой симплексной таблицы.

Если $t = t_3$, то независимая переменная $\tilde{x}_s$ переводится на другую границу. После конечного числа шагов либо будет найдено максимальное значение линейной формы F , либо будет установлен факт ее неограниченности сверху.

Пример 3.10. Дана задача ЛП в канонической форме с двусторонними ограничениями

$$\begin{aligned} F &= 4x_1 + x_2 \rightarrow \max, \\ y_1 &= x_1 - x_2 + 1 \geq 0, \\ y_2 &= 5x_1 - x_2 - 5 \geq 0, \\ 4 &\geq y_3 = -x_1 - x_2 + 6 \geq 0, \\ 4 &\geq x_1 \geq 0, \\ 3 &\geq x_2 \geq -1. \end{aligned}$$

Составляем симплексную таблицу

	$-x_1$	$-x_2$	1
y_1	-1	1	1
y_2	-5	1	-5
y_3	1	1	6
F	-4	-1	0

и выводим переменные x_1 и x_2 на нижнюю границу. У x_1 нижняя граница равна нулю, поэтому дополнительных преобразований не требуется. Перевод x_2 на нижнюю границу выполним так, как это описано выше: умножим элементы второго столбца α_{i2} на $\gamma_2 = -1$ и вычтем из столбца свободных членов, получим таблицу:

	$-x_1^-$	$-x_2^-$	1	$\alpha_i - \gamma_i$	$b_i - \alpha_i$
y_1	-1	1	2	2	∞
y_2	-5	1	-4	-4	∞
y_3	1	1	7	7	-3
F	-4	-1	-1		

(3.39)

где $x_1^- = x_1 - \gamma_1 = x_1$;

$$x_2^- = x_2 - \gamma_2 = x_2 - 1;$$

$$0 \leq x_1^- \leq 4, \quad 0 \leq x_2^- \leq 4.$$

Приравняв x_1^- и x_2^- нулю, получим базисное решение, расположенное в точке O_1 (рисунок 3.10, а) и в начале координат (рисунок 3.10, б). Геометрическая интерпретация задачи дана для сравнения в плоскости независимых переменных x_1 и x_2 (рисунок 3.10, а) и в плоскости этих же переменных, выведенных на нижние границы. Из рисунка 3.10 видно, что вершины O_1 и O отделены от ОДР двумя границами $y_2 = 0$ и $y_3 = 4$. Этот же результат получим, проанализировав столбец свободных членов симплексной таблицы (3.39):

$$\alpha_2 = -4 < 0; \alpha_3 = 7 > 4.$$

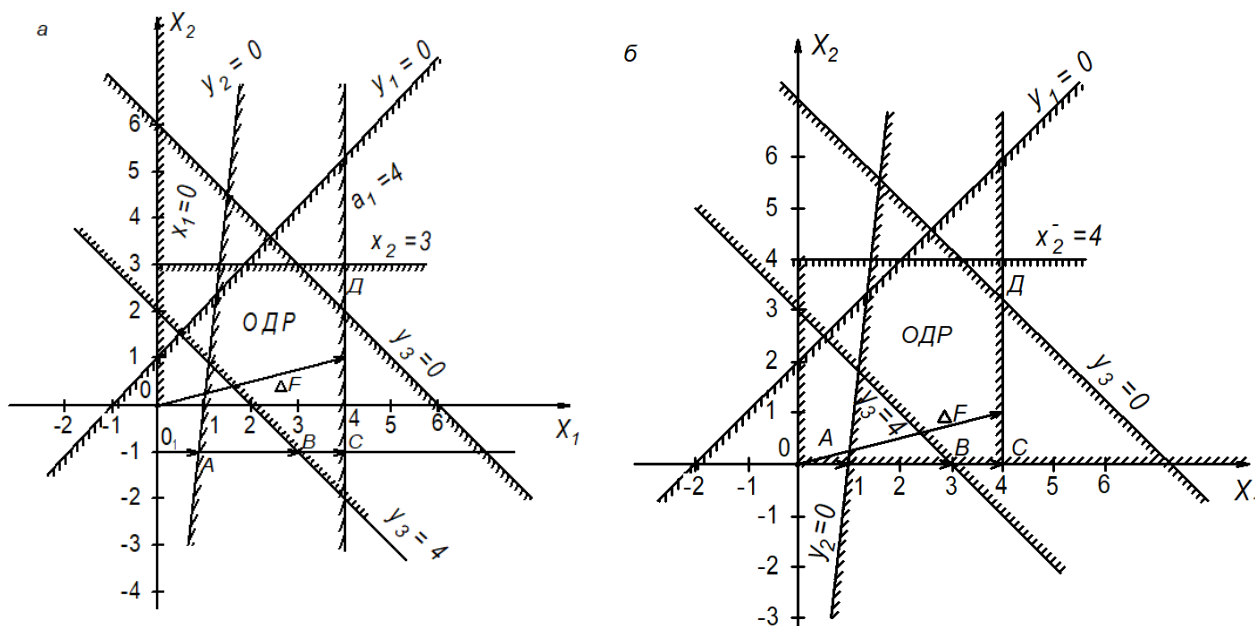


Рисунок 3.10. Геометрическая интерпретация задачи ЛП с двусторонними ограничениями к примеру 3.10:

а – в плоскости переменных x_1 и x_2 ; б – x_1^- и x_2^-

Приступим к поиску опорного решения.

Найдем во второй строке отрицательный элемент $\alpha_{12} = -5$ и первый столбец примем за разрешающий. Действительно, чтобы уменьшить величину $\alpha_2 - \gamma_2 = -4$, следует увеличить ту из независимых переменных, у которой в уравнении

$$y_2 = 5x_1^- - x_2^- - 4 = 0$$

положительный коэффициент (а в симплексной таблице отрицательный), т.е. x_1^- . Это равносильно движению из точки O по оси x_1^- в направлении $y_2 = 0$ до ближайшей границы. В нашем случае ближайшей является граница $y_2 = 0$, как следует из рисунка 3.10.

Убедимся в этом. Вычислим неотрицательные отношения $\frac{\alpha_i - \gamma_i}{\alpha_{i1}} \geq 0$ и выберем из них наименьшее

$$t_1 = \min \left\{ \frac{-4}{-5}; \frac{7}{1} \right\} = \frac{4}{5}.$$

Вычислим неотрицательные отношения $\frac{\beta_i - \alpha_i}{-\alpha_{i1}} \geq 0$ и выберем из них наименьшее

$$t_2 = \min \left\{ \frac{\infty}{-(-1)}; \frac{\infty}{-(-5)}; \frac{-3}{-1} \right\} = 3.$$

Вычислим $t_3 = 4$ и найдем $t = \min\{t_1, t_2, t_3\} = t_1 = \frac{4}{5}$.

Значение t_1 получено для элемента $\alpha_{12} = -5$, который принимаем за разрешающий и выполняем усложненный шаг модифицированных жордановых исключений, в результате которого x_1^- становится базисной переменной вместо y_2 и записывается в старых обозначениях, при этом дополнительных преобразований свободного члена разрешающей строки делать не надо, так как нижний предел x_1 равен нулю. Независимая переменная y_2 выводится на нижнюю границу. Новая симплексная таблица имеет вид:

	$-y_2^-$	$-x_2^-$	1	$\alpha_i - \gamma_i$	$b_i - \alpha_i$
y_1	-0,2	0,8	2,8	2,8	∞
x_1	-0,2	-0,2	0,8	0,8	3,2
y_3	0,2	1,2	6,2	6,2	-2,2
F	-0,8	-1,8	2,2		

(3.40)

Приравняв в (3.40) независимые переменные y_2^- и x_2^- нулю, найдем базисное решение:

$$y_1 = 2,8; x_1 = 0,8; y_3 = 6,2; y_2 = 0; x_2 = -1.$$

Ему соответствует вершина А, которая отделяется от ОДР границей $y_3 = 4$ (см. рис. 3.10).

Повторяем операции алгоритма поиска опорного решения. Проанализировав столбец свободных членов, убеждаемся, что нижние границы нигде не нарушены, а верхняя граница нарушена в строке y_3 . Выбираем в третьей строке положительный элемент, например, $\alpha_{31} = 0,2$ и первый столбец принимаем за разрешающий. Находим:

$$t_1 = \frac{\alpha_3 - \gamma_3}{\alpha_{31}} = \frac{6,2}{0,2} = 31;$$

$$t_2 = \min \left\{ \frac{\beta_i - \alpha_i}{-\alpha_{i1}} \geq 0 \right\} = \min \left\{ \frac{\infty}{-(-0,2)}; \frac{3,2}{-(-0,2)}; \frac{-2,2}{-0,2} \right\} = 11;$$

$$t_3 = \infty.$$

Выполняем усложненный шаг модифицированных жордановых исключений с разрешающим элементом $\alpha_{31} = 0,2$, в результате которого y_2 становится базисной переменной и записывается в старых обозначениях, а новая независимая переменная y_3 выводится на верхнюю границу. Для этого после

обычного шага исключений вычитаем из столбца свободных членов первый столбец, умноженный на $\beta_3 = 4$, и меняем знаки элементов первого столбца на противоположные. Получим симплексную таблицу:

	$-y_3^+$	$-x_2^-$	1	$\alpha_i - \gamma_i$	$b_i - \alpha_i$
y_1	-1	2	5	5	∞
x_1	-1	1	3	3	1
y_2	-5	6	11	11	∞
F	-4	3	11		

(3.41)

Приравняв в таблице (3.41) независимые переменные y_3^+ и x_2^- нулю, получим базисное решение:

$$y_1 = 5; x_1 = 3; y_3 = 4; y_2 = 11; x_2 = -1; F = 11.$$

Этому решению соответствует опорная вершина В ОДР (см. рис. 3.10).

В том, что опорное решение найдено, убеждаемся, проанализировав столбец свободных членов таблицы (3.41).

Найдем оптимальное решение задачи ЛП. Анализируем элементы F- строки таблицы (3.41) – среди них содержится один отрицательный $p_1 = -4$, следовательно, первый столбец должен быть разрешающим. Так как в разрешающем столбце все элементы отрицательны, то при движении из вершины В ОДР границы типа $y_1 = 0, x_1 = 0; y_2 = 0$ не встретятся.

Вычисляем поэтому только

$$t_2 = \min_{\alpha_{i1} > 0} \left\{ \frac{\infty}{-(-1)}; \frac{1}{-(-1)}; \frac{\infty}{-(-5)} \right\} = 1,$$

$$t_3 = \beta_3 - \gamma_3 = 4.$$

Находим $t = \min\{t_2, t_3\} = \min\{1, 4\} = 1$.

Таким образом, $t = t_2$ и, следовательно, $\alpha_{21} = -1$ разрешающий элемент. Выполняем усложненный шаг исключений, в результате которого x_1 становится независимой и выводится на верхнюю границу, а y_3^+ вводится в базис и записывается в обычном обозначении, при этом элементы разрешающей строки меняют свои знаки и к свободному члену разрешающей строки прибавляется $\beta_3 = 4$. Указанные преобразования приведены ниже.

Сначала выполняем обычный шаг модифицированных жордановых исключений над таблицей (3.41) с разрешающим элементом $\alpha_{21} = -1$:

	$-x_1$	$-x_2^-$	1
y_1	-1	1	2
y_3^+	-1	-1	-3
y_2	-5	1	-4
F	-4	-1	-1

Переводим x_1 на верхнюю границу $\beta_1 = 4$:

	$-x_1^+$	$-x_2^-$	1
y_1	1	1	6
y_3^+	1	-1	1
y_2	5	1	16
F	4	-1	15

Записываем y_3^+ в обычном обозначении:

	$-x_1^+$	$-x_2^-$	1	$\alpha_i - \gamma_i$	$b_i - \alpha_i$
y_1	1	1	6	6	∞
y_3	-1	1	3	3	1
y_2	5	1	16	16	∞
F	4	-1	15		

(3.42)

Приравняв в таблице (3.42) независимые переменные x_1^+ и x_2^- нулю, получим опорное решение

$$y_1 = 6; x_1 = 4; y_3 = 3; y_2 = 16; x_2 = -1; F = 15.$$

Это решение находится в опорной вершине С ОДР (рис. 3.10).

В F-строке симплексной таблицы (3.42) элемент $p_2 = -1$, следовательно, опорное решение неоптимально и может быть улучшено. Делаем второй столбец разрешающим и вычисляем:

$$t_1 = \min_{\alpha_{i2} > 0} \{6; 3; 16\} = 3,$$

$$t_3 = \beta_2 - \gamma_2 = 4.$$

Величина t_2 не вычисляется, так как все элементы разрешающего столбца положительны. Это означает, что при движении из вершины С в направлении увеличения линейной формы F границы типа $y_3 = 4$ (верхний предел) не встретятся.

Находим $t = \min\{t_1; t_3\} = \min\{3; 4\} = 3$.

Так как $t = t_1 = 3$, то выполняем усложненный шаг модифицированных жордановых исключений с разрешающим элементом $a_{22} = 1$. В результате этого шага y_3 становится независимой переменной и выводится на нижнюю границу, а x_2^- вводится в базис, записывается обычным образом, при этом к свободному члену разрешающей строки добавляется $\gamma_2 = -1$. Все преобразования приведены ниже. После обычного шага исключений получим таблицу:

	$-x_1^+$	$-y_3$	1
y_1	2	-1	3
x_2^-	-1	1	3
y_2	6	-1	13
F	3	1	18

Приведение y_3 к нижней границе не требует дополнительных преобразований таблицы, так как нижний предел y_3 $\gamma_3 = 0$.

Приведем x_2^- к обычному обозначению и получим окончательно симплексную таблицу в виде:

	$-x_1^+$	$-y_3^-$	1
y_1	2	-1	3
x_2	-1	1	2
y_2	6	-1	13
F	3	1	18

(3.43)

Приравняв независимые переменные x_1^+ и y_3^- нулю, найдем опорное решение:

$$y_1 = 3; x_1 = 4; y_3 = 0; y_2 = 13; x_2 = 2; F = 18.$$

В F-строке таблицы (3.43) все элементы положительны, следовательно, опорное решение оптимально и единственно; ему соответствует максимум линейной формы $F_{max} = 18$. Оптимальное решение расположено в вершине Д ОДР (рис. 3.10).

3.5. Задача выбора оптимального технологического режима процесса обезвоживания осадка сточных вод

Процесс обезвоживания осадка сточных вод осуществляется на барабанных вакуум-фильтрах. Одним из основных показателей процесса является производительность вакуум-фильтра по сухому веществу, которая зависит от таких переменных фильтров как уровень в ванне, скорость вращения барабана, концентрация ила в ванне. В задаче выбора оптимального режима работы вакуум-фильтра обеспечение максимума производительности является целевой установкой. При этом должны соблюдаться ограничения по толщине слоя ила и его влажности в месте съема барабана. Оба эти показателя также зависят от перечисленных выше параметров. Введем обозначения:

F – производительность вакуум-фильтра по сухому осадку, $\frac{\text{кг}}{\text{м}^3 \cdot \text{ч}}$;

y_1 – влажность осадка, %;

y_2 – толщина слоя ила, м;

x_1 – скорость вращения барабана, об/мин;

x_2 – уровень в ванне, м;

x_3 – концентрация ила в ванне, т/л;

Зависимости от переменных факторов x_1, x_2, x_3 были получены по данным пассивно-активного эксперимента в виде линейных уравнений регрессии.

Математическая модель задачи оптимизации процесса, составленная на основе линейных уравнений регрессии, приняла вид задачи ЛП с двусторонними ограничениями:

$$\begin{aligned} F &= 1,21x_1 + 2,52x_2 + 0,78x_3 + 6,49 \rightarrow \max; \\ 75 &\leq y_1 = 2,1x_1 + 1,45x_2 - 0,91x_3 + 79,8 \leq 85, \\ 0,05 &\leq y_2 = -0,0046x_1 + 0,029x_2 + 0,084 \leq 11, \\ 0,2 &\leq x_1 \leq 0,8, \\ 0,8 &\leq x_2 \leq 0,9, \\ 8 &\leq x_3 \leq 30. \end{aligned}$$

Для решения задачи воспользуемся модифицированным симплекс-алгоритмом. Составим исходную симплексную таблицу:

	$-x_1$	$-x_2$	$-x_3$	1
y_1	-2,1	-1,46	0,91	79,8
y_2	0,0046	-0,029	0	0,084
F	-1,21	-2,52	-0,78	6,49

Выведем переменные x_1 и x_2 на их верхние границы, а x_3 – на нижнюю. Для этого вычтем из столбца свободных членов первый и второй столбец, умноженные на величины верхних границ x_1 и x_2 соответственно, и третий столбец, умноженный на значение нижней границы x_3 , а, кроме того, поменяем знаки у элементов первого и второго столбцов:

	$-x_1^+$	$-x_2^+$	$-x_3^-$	1	$\alpha_i - \gamma_i^*$	$b_i^* - \alpha_i$
y_1	2,1	1,46	0,91	75,5	0,5	9,5
y_2	-0,0046	0,029	0	0,106	0,056	0,004
F	1,21	2,52	-0,78	15,92		

Проверим, все ли свободные члены симплексной таблицы удовлетворяют ограничениям, т.е. выполняется ли условие

$$\gamma_i^* \leq \alpha_i \leq \beta_i^*.$$

Из двух вспомогательных столбцов таблицы видно, что ограничения на y_1 и y_2 выполнены, следовательно, опорное решение найдено, в опорном решении параметры принимают следующие значения:

$$x_1 = 0,8; x_2 = 0,9; x_3 = 8; y_1 = 75,5; y_2 = 0,106; F = 15,92.$$

В F-строке симплексной таблицы имеется отрицательный элемент $p_3 = -0,78$; третий столбец принимаем за разрешающий. Вычисляем координату пересечения оси x_3^- с нижней границей y_n :

$$t_1 = \frac{\alpha_1 - \gamma_i}{\alpha_{13}} = \frac{0,5}{0,91} = 0,55.$$

Пересечения x_3^- с верхними границами y_1 и y_2 не будет, так как в разрешающем столбце нет отрицательных элементов, поэтому величину t_2 не вычисляем.

Находим $t_3 = \beta_3 - \gamma_3 = 30 - 8 = 22$,

$$t = \min\{t_1, t_3\} = 0,55.$$

Выполняем усложненный шаг модифицированных жордановых исключений с разрешающим элементом $\alpha_{13} = 0,91$. В результате получим симплексную таблицу, в которой новая базисная переменная x_3^- записывается в обычных обозначениях, а независимая y_1 выводится на нижнюю границу:

	$-x_1^+$	$-x_2^+$	$-y_1^-$	1
x_3	2,3	1,6	1,1	8,5
y_2	-0,0046	0,029	0	0,106
F	3,0	3,8	0,86	16,6

Симплексной таблице соответствует оптимальное решение

$$x_1 = 0,8; x_2 = 0,9; x_3 = 8,5; y_1 = 75; y_2 = 0,106; F = 16,6.$$

Так как элементы F-строки строго положительны, то решение единственно.

Геометрическая интерпретация задачи в пространстве трех переменных x_1^+, x_2^+, x_3^- дана на рисунке 3.11.

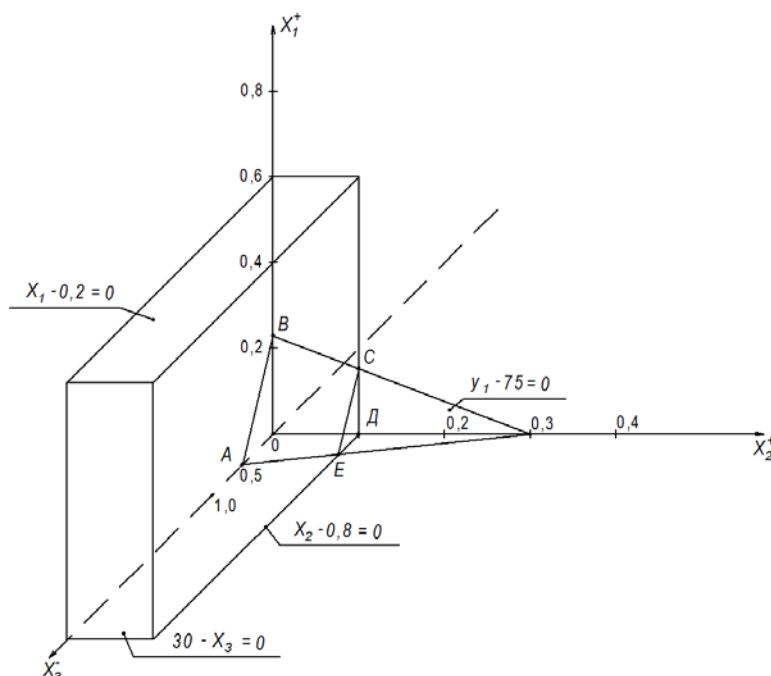


Рисунок 3.11. Геометрическая интерпретация ОДР в задаче обезвреживания осадков сточных вод

Область допустимых решений задачи – выпуклый многогранник ОАВСДЕ образован границами

$$\begin{aligned} x_1^+ &= 0,8 - x_1 = 0; & x_2^+ &= 0,9 - x_2 = 0; & x_2^- &= x_2 - 0,8 = 0; \\ x_3^- &= x_3 - 8 = 0; & y_1^- &= y_1 - 75 = 0. \end{aligned}$$

Решением задачи является опорная вершина А ОДР, координаты которой приведены выше.

ГЛАВА 4. ЗАДАЧИ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

4.1. Общая постановка задачи нелинейного программирования Характеристика методов решения

Задача нелинейного программирования состоит в отыскании экстремума нелинейной функции цели при заданных ограничениях в виде уравнений или неравенств. Ограничения могут быть линейными и нелинейными. Общая задача нелинейного программирования формулируется следующим образом [7, 13]: найти минимум

$$f(\bar{x}), \bar{x} \in E^n, \quad (4.1)$$

при m линейных или нелинейных ограничениях в виде равенств

$$h_i(\bar{x}) = 0, \quad i = 1, \dots, m \quad (4.2)$$

и $(p-m)$ линейных и нелинейных ограничениях в виде неравенств

$$g_i(\bar{x}) \leq 0, \quad i = m + 1, \dots, p. \quad (4.3)$$

Напомним, что отыскание минимума функции цели (4.1) введено для определенности, поскольку задача на максимум $f(\bar{x})$ легко сводится к задаче минимизации функции, противоположной исходной по знаку. Аналогично в условии (4.3) знак неравенства может быть изменен на обратный путем умножения на -1 , что не изменит математической модели задачи.

Общая постановка задачи (4.1) – (4.3) уже рассматривалась во 2 главе, где речь шла о применении аналитических методов решения. Там же формулировались (см. (2.39) – (2.41)) необходимые условия оптимальности для этой задачи, которые приводили к обобщенной функции Лагранжа (2.40).

Нелинейное программирование – наиболее обширный класс задач: подавляющее большинство самых различных по функциональному назначению экстремальных задач, решаемых в АСУТП, относится к классу нелинейного программирования. Однако лишь немногие из них могут быть решены аналитическими методами. Сложность математической модели задач – большая размерность, сложный характер функции цели и ограничений приводят к необходимости применения численных приближенных методов решения и, следовательно, средств вычислительной техники.

Среди задач нелинейного программирования принято выделять:

- задачи квадратичного программирования, в которых функция цели представляет собой квадратичную форму (положительно или отрицательно определенную), а ограничения линейны;
- задачи выпуклого программирования, в которых $f(\bar{x})$ – выпуклая функция и выпуклые функции в ограничениях.

В задачах квадратичного и выпуклого программирования область допустимых решений выпукла и решение задачи достигается либо внутри ОДР, либо на ее границе, а в силу выпуклости $f(\bar{x})$ гарантируется глобальность найденного экстремума. Для решения этих задач получены наиболее

эффективные алгоритмы [7, 9, 13]. Общая задача нелинейного программирования, не обладающая свойствами выпуклости, вызывает при решении большие трудности. Теория задач нелинейного программирования в полной мере еще не разработана. Применяемые в настоящее время методы нелинейного программирования реализуют итеративную (многошаговую) процедуру последовательного улучшения некоторого начального решения. Причем заранее нельзя сказать (кроме случаев, когда $f(\bar{x})$ – квадратичная функция), какое наибольшее число итераций гарантирует нахождение экстремума с заданной точностью. Разнообразие методов нелинейного программирования во многом объясняется стремлением достигнуть заданной точности решения за наименьшее число шагов. Первым этапом решения задачи нелинейного программирования является выбор начальной точки. Если в алгоритме метода не требуется, чтобы точка $\bar{x}^0$ была допустимой, то задание координат $\bar{x}^0$ не представляет труда. Если же точка $\bar{x}^0$ должна принадлежать ОДР, то выбор $\bar{x}^0$ является самостоятельной задачей, так как координаты точки $\bar{x}^0 = (\bar{x}_1^0, \bar{x}_2^0, \dots, \bar{x}_n^0)$ должны удовлетворять системе уравнений (4.2) и неравенств (4.3). Здесь могут быть применены различные подходы. Один из подходов следующий: зададим произвольно $n-m$ координат точки $\bar{x}^0$ и подставим их в уравнение (4.2). Получим систему из m уравнений с n неизвестными:

$$h_i(x_1, x_2, \dots, x_m, x_{m+1}^0, x_{m+2}^0, \dots, x_n^0) = 0, \quad i = 1, 2, \dots, m, \quad (4.4)$$

найдем решения системы уравнений (4.4) и выберем из них те, которые удовлетворяют неравенствам (4.3). При этом может возникнуть необходимость численного решения системы уравнений (4.4).

Другой подход состоит в решении вспомогательной экстремальной задачи минимизации функции вида:

$$\Phi(x) = \sum_{i=1}^m h_i^2(\bar{x}). \quad (4.5)$$

Действительно, функция $\Phi(x)$ квадратичная, минимум ее равен нулю и достигается в точках $\bar{x}$, которые являются корнями системы уравнений $h_i(\bar{x})$, поэтому любая стационарная точка функции $\Phi(x)$ может быть выбрана в качестве точки начального приближения $\bar{x}^0$, если она удовлетворяет неравенствам (4.3). Очевидно, что решение задачи (4.5) может повлечь за собой немалые трудности.

После выбора точки начального приближения начинается этап решения задачи нелинейного программирования. Методы нелинейного программирования, применяемые в настоящее время, можно объединить в две большие группы. Одна группа методов состоит в том, что задача с ограничениями сводится к последовательности задач безусловной оптимизации специальным образом сконструированных вспомогательных функций, называемых функциями штрафа. Эта группа методов получила название методов штрафных функций.

В методах другой группы реализуется шаговое движение к экстремуму функции цели, причем на каждом шаге контролируется соблюдение ограничений

задачи. Поэтому методы второй группы называют методами допустимых направлений.

Алгоритмы практически всех названных методов непременно включают как составляющий этап – решение задачи безусловной оптимизации. Поэтому прежде, чем изучить методы решения задач с ограничениями, целесообразно ознакомиться с численными методами отыскания безусловного экстремума.

4.2. Численные методы безусловной оптимизации

Рассмотрим задачу поиска безусловного минимума функции $f(\bar{x})$, заданной на всем пространстве E^n . Принято различать два подхода к решению этой задачи. Один подход связан с отысканием стационарных точек функции $f(\bar{x})$ из системы уравнений, выражающей необходимые условия оптимальности с последующей проверкой выполнения стационарных точек достаточных условий оптимальности. Этот подход подробно рассмотрен в главе 2. Другой подход к отысканию безусловного экстремума $f(\bar{x})$ связан с методами прямого поиска или методами спуска, в которых использование необходимых условий оптимальности необязательно. Методы спуска реализуют итеративную процедуру построения последовательности точек $\{\bar{x}^0, \bar{x}^1, \dots, \bar{x}^k\}$ в пространстве E^n . Значения функции цели $f(\bar{x})$ в точках $\{\bar{x}^k\}$ образуют убывающую сходящуюся последовательность

$$f(\bar{x}^0) \geq f(\bar{x}^1) \geq f(\bar{x}^2) \geq \dots \geq f(\bar{x}^k),$$

такую, что $\{f(\bar{x}^k)\} \rightarrow f(\bar{x}^*)$,

где $\bar{x}^*$ – точка минимума $f(\bar{x})$.

Различные методы спуска отличаются друг от друга способами выбора направления спуска и длины шага вдоль этого направления. На каждой итерации алгоритма спуска очередная точка приближения к экстремуму вычисляется по формуле

$$\bar{x}^{k+1} = \bar{x}^k + \alpha^{k+1} \bar{S}^k, \quad (4.6)$$

где k – номер итерации;

$\bar{x}^k, \bar{x}^{k+1}$ – точки приближения к экстремуму на k и $k+1$ итерациях;

$\bar{S}^k$ – направления спуска;

α^{k+1} – коэффициент, определяющий величину шага.

В зависимости от порядка производных функции $f(\bar{x})$, вычисляемых в процессе поиска экстремума, различают методы нулевого, первого и второго порядка. К методам нулевого порядка относятся метод покоординатного спуска, симплексный метод нелинейного программирования, метод сканирования и другие. Группу методов первого порядка образуют градиентные методы. К методам второго порядка относится метод Ньютона и его различные модификации. Особую группу составляют методы сопряженных направлений, которые могут иметь первый или второй порядок в зависимости от их модификации.

Одним из показателей эффективности методов спуска является скорость сходимости, т.е. скорость приближения к экстремуму. Методы первого порядка имеют линейную сходимость, которая оценивается по формуле

$$|\bar{x}^{k+1} - \bar{x}^*| \leq q |\bar{x}^k - \bar{x}^*|, \quad (4.7)$$

где $q = \frac{M-m}{M+m}$;

M, m – оценки максимального и минимального из собственных чисел матрицы Гессе.

У метода Ньютона скорость сходимости квадратична:

$$|\bar{x}^{k+1} - \bar{x}^*| \leq c |\bar{x}^k - \bar{x}^*|^2, \\ c \geq 0. \quad (4.8)$$

Среди методов спуска методы второго порядка требуют наименьшего числа итераций для нахождения точки экстремума с заданной точностью. Еще одним показателем эффективности являются затраты машинного времени на поиск, которые у методов второго порядка могут оказаться значительными из-за необходимости вычислять вторые частные производные функции цели.

Для оценки точности попадания в окрестность точки экстремума $\bar{x}^*$ применяются различные критерии. В методах нулевого порядка может оцениваться, например, расстояние, пройденное за последние g шагов

$$|\bar{x}^k - \bar{x}^{k+r}| \leq \delta,$$

где δ – точность вычисления экстремума, или приращение функции цели $f(\bar{x})$ за g шагов

$$|f(\bar{x}^k) - f(\bar{x}^{k+r})| \leq \delta.$$

Удобно для оценки попадания в заданную окрестность точки экстремума использовать величину градиента $f(x)$: этот показатель применяется в качестве признака окончания вычислений в методах первого и второго порядка:

$$|\nabla f(\bar{x}^k)| \leq \delta.$$

Ниже дается характеристика наиболее распространенных методов прямого поиска, кроме методов случайного поиска, которые в настоящем пособии не рассматриваются.

Методы нулевого порядка

Метод покоординатного спуска

Движение к экстремуму функции цели в этом методе осуществляется в осевом направлении. Движение начинается из некоторой точки $\bar{x}^0$, координаты которой задаются произвольно.

Рассмотрим k итерацию алгоритма метода. Пусть к началу k итерации найдена точка $\bar{x}^{k-1}$ с координатами $\bar{x}_1^{k-1}, \bar{x}_2^{k-1}, \dots, \bar{x}_n^{k-1}$ и вычислено значение функции цели $f(\bar{x}^{k-1})$, заданы величина шага α^k и точность определения экстремума δ .

1. Фиксируем все координаты точки $\bar{x}^{k-1}$, кроме $\bar{x}_i^{k-1}$.
2. В направлении x_i выполняем шаг:

$$\bar{x}_i^k = \bar{x}_i^{k-1} \pm \alpha_k.$$

Примечание: Направление движения вдоль оси x_i определяется в результате пробного шага так, чтобы функция $f(\bar{x})$ убывала (в случае поиска минимума).

3. Фиксируем координату $\bar{x}_i^k$ и выполняем шаг в направлении x_{i+1} и т.д. до тех пор, пока не переберем все основные направления. За одну итерацию выполняется n шагов, по одному в каждом из осевых направлений.

4. Оцениваем изменение функции цели за n шагов k итерации

$$\Delta f(\bar{x}^k) = f(\bar{x}^{k-1}) - f(\bar{x}^k).$$

5. Если $\Delta f(\bar{x}^k) < \delta$, то расчет окончен.

Алгоритм метода покоординатного спуска может быть реализован и несколько иначе. Например, после шага в направлении x_i вычисляем $f(x_1^{k-1}, \dots, x_{i-1}^{k-1}, x_i^k, x_{i+1}^{k-1}, \dots, x_n^{k-1})$ и, если выполняется условие $\Delta f = f(x^{k-1}) - f(x_1^{k-1}, \dots, x_{i-1}^{k-1}, x_i^k, x_{i+1}^{k-1}, \dots, x_n^{k-1}) > 0$, делаем в том же направлении еще один шаг, продолжая шаги до тех пор, пока функция цели на этом направлении убывает. Как только после некоторого шага в направлении x_i значение функции цели ухудшится, переходим к п. 3.

Геометрическая интерпретация алгоритма покоординатного спуска в плоскости двух переменных дана на рисунке 4.1.

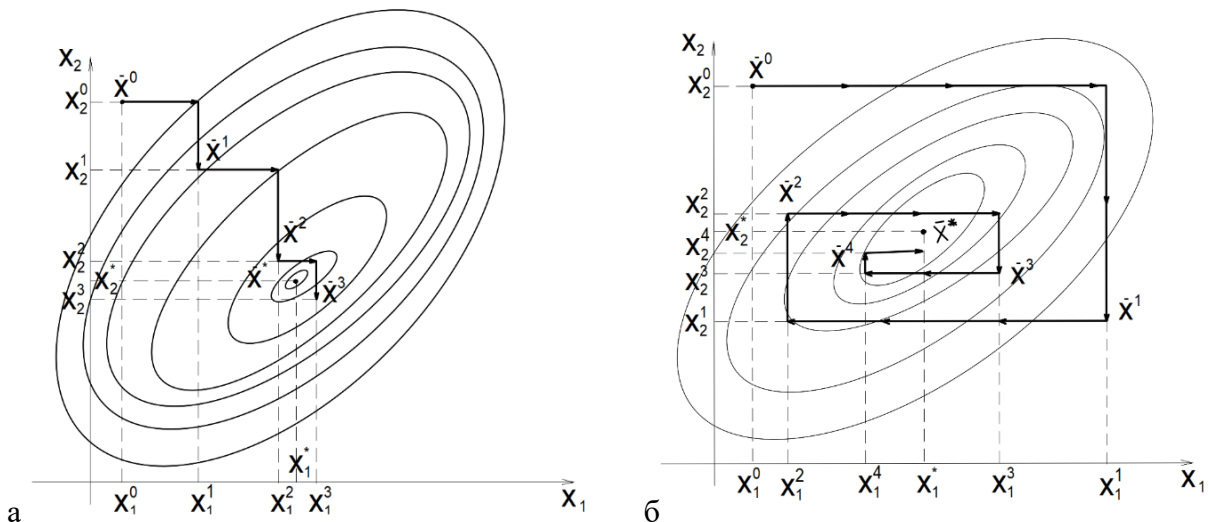


Рисунок 4.1. Геометрическая интерпретация алгоритма покоординатного спуска в случае:

a – одного шага в каждом осевом направлении; *б* – нескольких шагов в каждом из осевых направлений

Сходимость вычислительного процесса в методе покоординатного спуска существенно зависит от выбора величины шага. Очевидно, что по мере приближения к экстремуму шаг должен уменьшаться, поэтому в алгоритме обычно предусматривается корректировка шага на каждой итерации. Процедуры выбора и корректировки шага не поддаются строгой формализации и, как правило, являются эвристическими. Например, в рассматриваемом методе шаг может корректироваться в зависимости от приращения $f(\bar{x})$ на предыдущей итерации. Очевидно, что если $f(\bar{x})$ меняется мало, шаг следует увеличить, и наоборот, если $f(\bar{x})$ меняется сильно, надо уменьшить шаг; отсюда алгоритм корректировки шага:

$$\alpha^{k+1} = \begin{cases} \alpha^k, & \text{если } \Delta f_{\min} \leq f(\bar{x}^k) - f(\bar{x}^{k-1}) \leq \Delta f_{\max} \\ a\alpha^k, & \text{если } \Delta f(\bar{x}^k) < \Delta f_{\min} \\ b\alpha^k, & \text{если } \Delta f(\bar{x}^k) > \Delta f_{\max} \end{cases},$$

где $a > 1$, $b < 1$.

Коэффициенты a , b и граничные значения $\Delta f_{\min}$, $\Delta f_{\max}$ задаются эвристически.

Метод покоординатного спуска прост в реализации, так как на каждой итерации рассчитываются только значения $f(\bar{x})$.

Недостатки метода:

- сходимость вычислительной процедуры существенно зависит от ориентации поверхности функции цели относительно оси координат;
- в случае многоэкстремальной функции цели $f(\bar{x})$ алгоритм метода позволяет найти только локальный экстремум. Для определения глобального экстремума необходимо многократно решать задачу, начиная поиск из различных точек начальных приближений.

Симплексный метод нелинейного программирования

Симплексом в пространстве E^n называется выпуклый многогранник, имеющий $n+1$ вершину; против каждой вершины симплекса расположена только одна грань. В плоскости симплекс – это треугольник, в трехмерном пространстве – пирамида и т.д.

Основная идея симплексного метода заключается в том, что по известным значениям функции цели в вершинах симплекса, построенного в пространстве, в котором определена функция цели $f(\bar{x})$, находится направление движения к экстремуму. В симплексном методе обычно используются правильные симплексы, у которых все грани равны. На любой грани симплекса можно построить новый симплекс, отличающийся от прежнего только расположением одной вершины.

Например, на грани $\bar{x}^2\bar{x}^3$ симплекса $\bar{x}^1\bar{x}^2\bar{x}^3$ (рисунок 4.2) можно построить новый симплекс $\bar{x}^2\bar{x}^3\bar{x}^4$, вершина $\bar{x}^1$ которого является зеркальным отражением вершины $\bar{x}^1$ относительно грани $\bar{x}^2\bar{x}^3$.

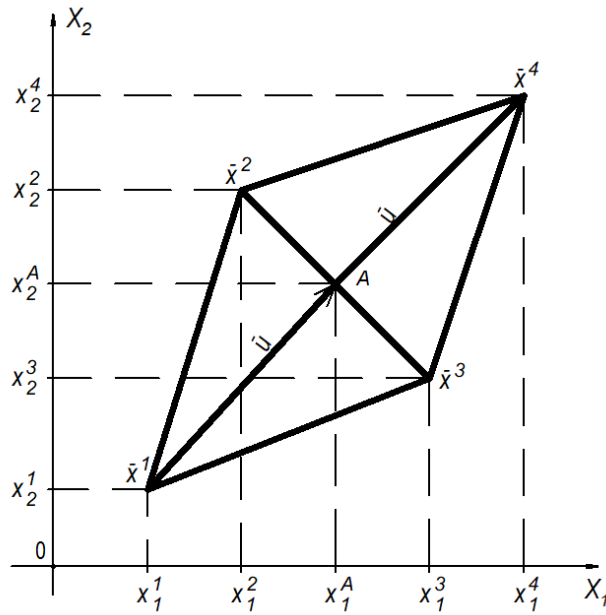


Рисунок 4.2. Геометрическая интерпретация построения симплексов

Чтобы построить новый симплекс, найдем координаты его вершины $\bar{x}^4$ по известным координатам вершины симплекса $\bar{x}^1 \bar{x}^2 \bar{x}^3$. Координаты точки А – центра грани $\bar{x}^2 \bar{x}^3$ могут быть вычислены по формуле:

$$x_j^A = \frac{1}{2}(x_j^2 + x_j^3), \quad j = 1, 2.$$

В общем случае n-мерного пространства формула приобретает вид:

$$x_j^A = \frac{1}{n} \sum_{i=1, i \neq r}^{n+1} x_j^i, \quad j = 1, \dots, n, \quad (4.9)$$

где i – номер вершины исходного симплекса;

r – номер вершины, расположенной против грани, на которой строится новый симплекс.

Обозначим $\bar{u}$ – высоту симплекса. Очевидно, что $\bar{u} = A - \bar{x}^1$, тогда $\bar{x}^4 = \bar{A} + \bar{u} = 2\bar{A} - \bar{x}^1$. В общем виде в координатной форме получим:

$$x_j^{n+1} = \frac{2}{n} \sum_{i=1, i \neq r}^{n+1} x_j^i - x_j^r, \quad (4.10)$$

где x_j^{n+1} – j координата вершины нового симплекса.

Движение к экстремуму функции цели в симплексном методе начинается из вершины некоторого исходного симплекса, расположение которого выбирается произвольно. В вершинах симплекса вычисляются значения $f(\bar{x})$ и выбирается худшее, т.е. наибольшее значение в случае поиска минимума $f(\bar{x})$. В представленной на рисунке 4.3 геометрической интерпретации алгоритма метода наибольшему значению $f(\bar{x})$ соответствует вершина $\bar{x}^1$. На грани $\bar{x}^2 \bar{x}^3$, расположенной против вершины $\bar{x}^1$, строится новый симплекс с вершиной $\bar{x}^4$, координаты которой вычисляются по формулам (4.9), (4.10). В вершине $\bar{x}^4$ вычисляется функция цели и выбирается та из вершин нового симплекса, в

которой $f(\bar{x})$ достигает наибольшего значения. На рисунке 4.3 это вершина $\bar{x}^3$. На грани $\bar{x}^2\bar{x}^4$ строится очередной симплекс с вершиной $\bar{x}^5$ и т.д. В результате процедуры построения симплексов получим последовательность точек, которая сходится к точке минимума $f(\bar{x})$. Вблизи точки $\bar{x}^*$ может возникнуть заикливание вычислительного процесса, которое состоит в том, что на некоторой итерации вместо построения новой вершины вернемся в вершину предыдущего симплекса. Случай заикливания иллюстрируется на рисунке 4.3: из вершины $\bar{x}^8$ мы приходим в вершину $\bar{x}^{11}$, которая, как это следует из расположения линий уровня $f(\bar{x})$, является худшей в новом симплексе, и из нее мы вновь попадем в вершину $\bar{x}^8$. Это объясняется тем, что точка минимума $\bar{x}^*$ находится внутри симплекса. Чтобы избавиться от заикливания, уменьшают размеры симплекса и продолжают движение к экстремуму. Вычисления прекращаются, когда размеры симплекса, например, высота, становятся меньше заданной точности попадания в окрестность экстремума: $|\bar{u}| \leq \delta$.

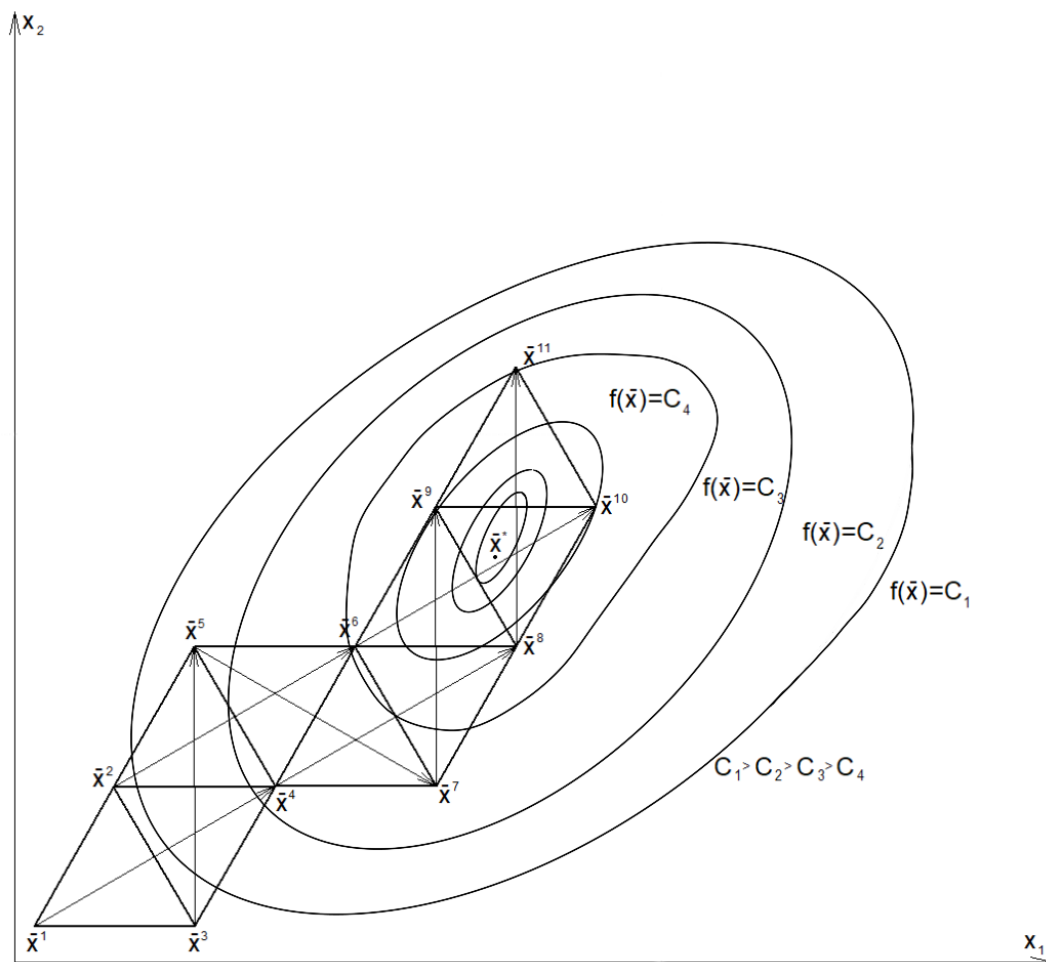


Рисунок 4.3. Геометрическая интерпретация симплексного метода

Достоинством симплексного метода является простота его реализации, так как на каждой итерации строится всего лишь одна вершина, и в ней и вычисляется значение функции цели. Только на первой итерации функция цели вычисляется во всех вершинах симплекса.

Недостатки метода:

Эффективность поиска зависит от размеров и ориентации начального симплекса. При неудачном расположении симплекса может возникнуть заикливание задолго до попадания в окрестность точки экстремума. Этот случай иллюстрируется на рисунке 4.4: зеркальным отражением «худшей» вершины $\bar{x}^1$ исходного симплекса является вершина $\bar{x}^4$, которая в своем симплексе также является «худшей».

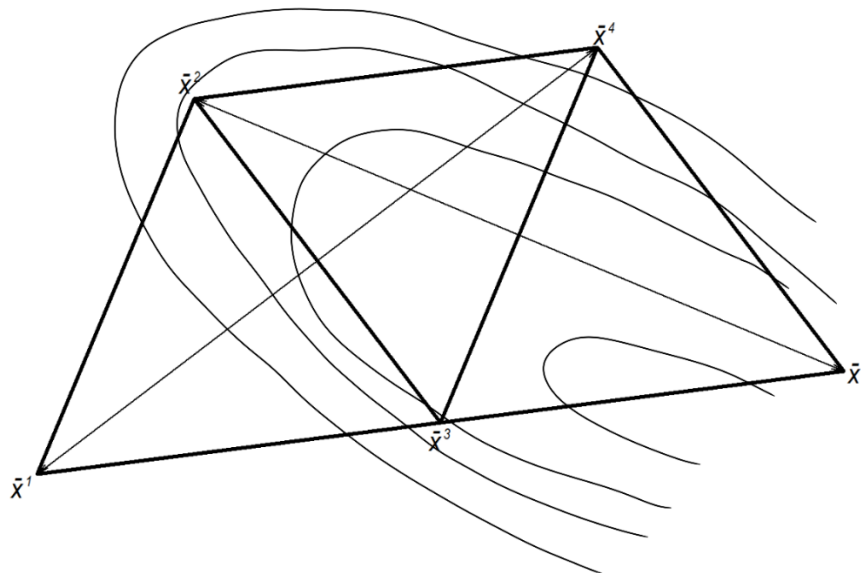


Рисунок 4.4. Иллюстрация заикливания при неудачной ориентации симплекса

Тогда, чтобы избавиться от заикливания, выбирают следующую «плохую» вершину $\bar{x}^2$ и строят новый симплекс на грани $\bar{x}^3\bar{x}^4$.

Методы первого порядка

Как уже указывалось, к методам первого порядка относятся градиентные методы, в которых движение к экстремуму функции цели осуществляется по градиенту; на некоторой k итерации очередная точка приближения к экстремуму вычисляется по формуле:

$$\bar{x}^k = \bar{x}^{k-1} \pm \alpha^k \nabla f(\bar{x}^{k-1}), \quad (4.11)$$

где α^k – коэффициент, определяющий величину шага в направлении $\nabla f(\bar{x}^{k-1})$.

Знак плюс или минус ставится в зависимости от того, в каком направлении осуществляется движение – по градиенту (к максимуму) или по антиградиенту (к минимуму).

Различные модификации градиентных методов отличаются способом выбора величины α^k на каждом шаге алгоритма. Если коэффициент α^k задается перед началом вычислений и не меняется в процессе поиска экстремума, то говорят о градиентном методе с фиксированным шагом, хотя это название не вполне точно, так как фактический шаг $\Delta \bar{x}^k = \alpha^k \nabla f(\bar{x}^{k-1})$ изменяется на каждой итерации за счет изменения – убывания $\nabla f(\bar{x})$ по мере приближения к экстремуму. Если величина α задается в начале вычислений, но на каждой

итерации корректируется по некоторому алгоритму, то такой метод называется градиентным методом с корректировкой шага. Один из способов корректировки шага описан в [12, 14]. Большую популярность завоевал метод, в котором коэффициент α вычисляется на каждой итерации из условия минимума (максимума) функции цели $f(\bar{x})$ на направлении антиградиента (градиента). На направлении градиента, вычисленного в точке $\bar{x}^{k-1}$, $f(\bar{x})$ является функцией только одного параметра α . Используя необходимые условия оптимальности функции одной переменной, получим:

$$\frac{df(\bar{x}^k)}{d\alpha} = \frac{df(\bar{x}^{k-1} \pm \alpha \nabla f(\bar{x}^{k-1}))}{d\alpha} = 0. \quad (4.12)$$

Уравнение (4.12) получено с учетом соотношения (4.11). Коэффициент α^k выбирается из корней уравнения (4.12)⁵. Поскольку на каждом шаге величина α^k выбирается оптимально, метод получил название метода наискорейшего спуска. Каждая точка $\bar{x}^k$ является точкой минимума (максимума) $f(\bar{x})$ на направлении $\nabla f(\bar{x}^{k-1})$, поэтому новое направление спуска $-\nabla f(\bar{x}^k)$, ортогонально предыдущему (см. рисунок 4.5). Из ортогональности направлений спуска можно получить еще одно правило для нахождения α :

$$(\nabla f(\bar{x}^{k-1}), \nabla f(\bar{x}^k)) = (\nabla f(\bar{x}^{k-1}), \nabla f(\bar{x}^{k-1} \pm \alpha \nabla f(\bar{x}^{k-1}))) = 0. \quad (4.13)$$

Очевидно, что уравнения (4.12) и (4.13) приводят к одним и тем же результатам.

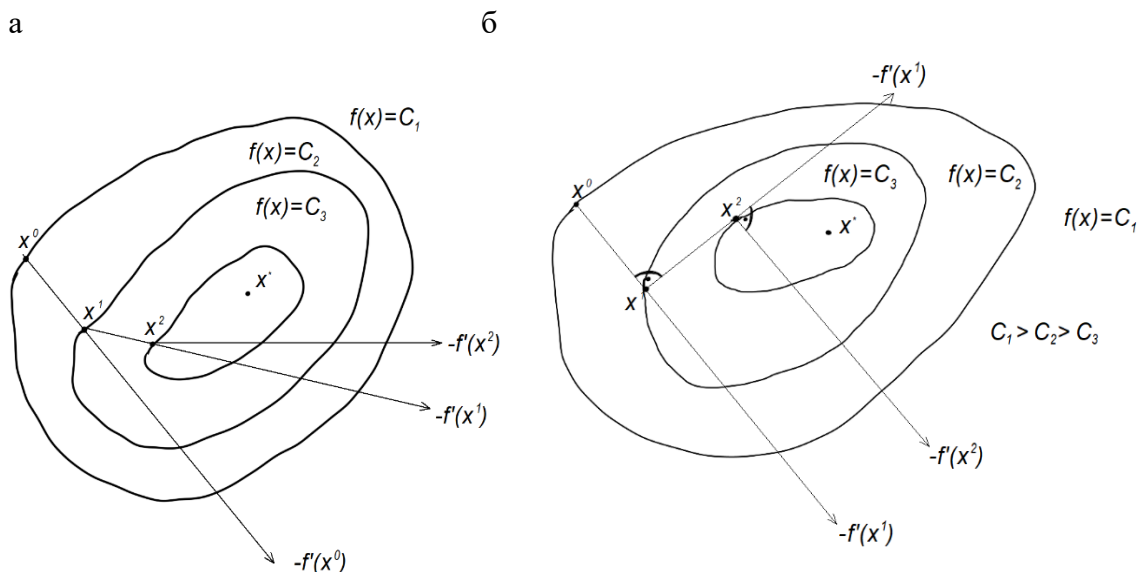


Рисунок 4.5. Геометрическая интерпретация градиентных методов:
а – с фиксированным шагом; б – наискорейшего спуска

Для функций высокого порядка могут возникнуть трудности при решении уравнений (4.12), (4.13). Для функций второго порядка коэффициент α вычисляется с помощью довольно простого соотношения:

⁵ Так как направление спуска в уравнении (4.12) учтено, то в качестве α^k выбирается наименьший положительный корень уравнения.

$$\alpha^k = - \frac{(\nabla f^T(\bar{x}^{k-1}), \nabla f(\bar{x}^{k-1}))}{(\nabla f^T(\bar{x}^{k-1})H\nabla f(\bar{x}^{k-1}))}, \quad (4.14)$$

где H – матрица Гессе.

Примечание. При использовании формулы (4.14) коэффициент α^k получается с учетом направления спуска, поэтому в алгоритме (4.11) направление спуска учитывать не следует.

Доказано, что градиентные методы сходятся из любой начальной точки, если функция $f(\bar{x})$ ограничена снизу и ее градиент удовлетворяет условию Липшица [5]:

$$|\nabla f(\bar{x}^1) - \nabla f(\bar{x}^2)| \leq R|\bar{x}^1 - \bar{x}^2|,$$

где $\bar{x}^1, \bar{x}^2$ – произвольные точки;

R – константа.

Доказано также, что если $f(\bar{x})$ выпуклая функция, то алгоритмы градиентных методов сходятся со скоростью геометрической прогрессии со знаменателем

$$q = \frac{M - m}{M + m},$$

(см. соотношение (4.7)).

Если функция $f(\bar{x})$ выпукла, то найденный экстремум будет единственным, в противном случае с помощью градиентных методов будет найден локальный экстремум, причем тот, в чьей области притяжения находится начальная точка. Сходимость градиентных методов не может считаться существенным преимуществом, так как скорость сходимости невелика и существенно зависит от характера поверхности функции цели. Если линия уровня $f(\bar{x})$ вытянута вдоль какого-либо направления и поверхность $f(\bar{x})$ имеет «овражный» характер, то градиентные методы работают плохо: поиск «застревает на дне оврага». «Овражному» характеру $f(\bar{x})$ соответствует плохая обусловленность матрицы Гессе: большое расхождение между ее собственными числами M и m приводит к тому, что коэффициент $q \approx 1$ и продвижение к экстремуму на каждой итерации будет незначительно. Проиллюстрируем это на примере [5].

Пример 4.1. Найти минимум

$$f(\bar{x}) = x_1^2 + 16x_2^2$$

из начальной точки $\bar{x}^0 = (2; 2)$ с точностью $\delta = 0,1$.

Очевидно, что минимум $f(\bar{x})$ достигается в точке $\bar{x}^* = (0; 0)$. Линии уровня $f(\bar{x})$ сильно вытянуты вдоль оси x_1 (см. рис. 4.6). Матрица Гессе в данном случае постоянна и равна

$$H = \begin{vmatrix} 2 & 0 \\ 0 & 32 \end{vmatrix}.$$

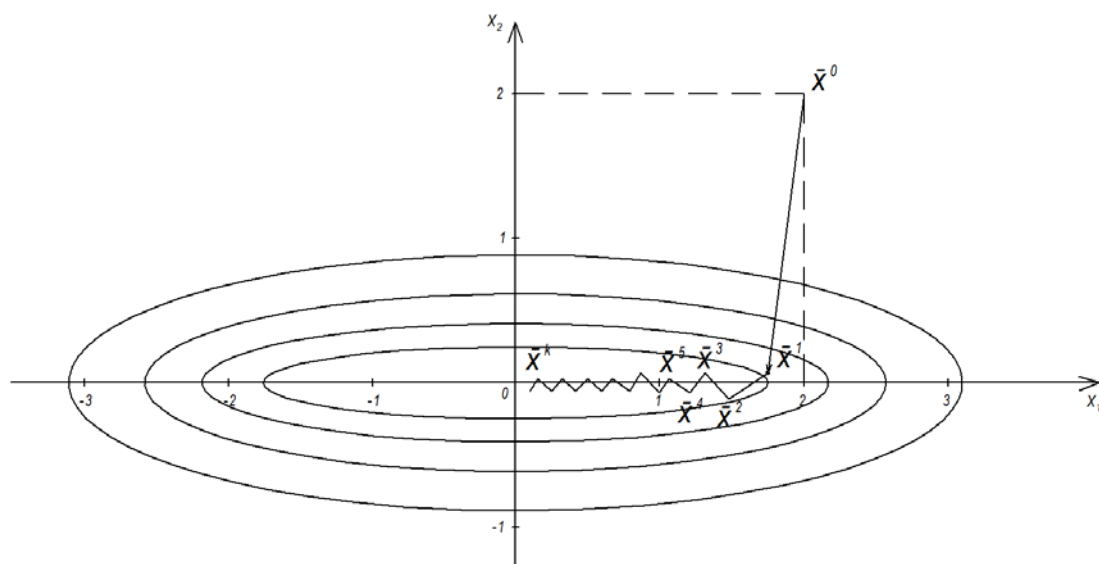


Рисунок 4.6. Геометрическая интерпретация метода наискорейшего спуска к примеру 4.1

Наименьшее и наибольшее собственные числа матрицы Гессе равны соответственно: $M = 32$, $m = 2$, т.е. сильно различаются между собой. Воспользуемся для отыскания минимума $f(\bar{x})$ методом наискорейшего спуска, результаты вычислений приведены в таблице 4.1, траектория спуска показана на рисунке 4.6; она характеризуется довольно быстрым спуском на «дно оврага» и медленным зигзагообразным движением в точку минимума.

Таблица 4.1. Результаты минимизации функции $f(\bar{x})$ методом наискорейшего спуска

Номер итерации	Координаты точки $\bar{x}^k$	$\nabla f(\bar{x}^k)$	α^k	$f(\bar{x}^k)$
0	(2; 2)	(4; 64)	-	68
1	(1,88; 0,08)	(3,76; 2,56)	-0,03	3,64
2	(1,55; -0,14)	(3,1; -4,48)	-0,087	2,72
3	(1,395; 0,084)	(2,79; 2,69)	-0,05	2,06
4	(1,23; -0,077)	(2,46; -2,47)	-0,06	1,61
5	(1,08; 0,071)	(2,16; 2,27)	-0,06	1,25

Изменить характер линий уровня функции цели можно, изменив масштаб независимых переменных. Проще всего это сделать, если $f(\bar{x})$ – квадратичная сепарабельная функция. Действительно, если $f(\bar{x})$ имеет вид:

$$f(\bar{x}) = \sum_{i=1}^n a_i x_i^2,$$

где коэффициенты a_i сильно различаются по величине, то поверхности уровней $f(\bar{x})$ вытянуты вдоль тех осей x_i , которым соответствуют малые a_i . Введем новые переменные:

$$x_i = \mu_i y_i \text{ или } y_i = a_i^{1/2} x_i,$$

где $\mu_i = \frac{1}{\sqrt{a_i}}$;

тогда в новых координатах y_i линии уровня $f(\bar{y})$ – концентрические окружности

$$f(\bar{y}) = \sum_{i=1}^n y_i^2.$$

В нашем примере мы начали движение из точки $\bar{x}^0 = (2; 2)$, и так как составляющие градиента в этой точке

$$\frac{\partial f(\bar{x}^0)}{\partial x_1} = 4; \quad \frac{\partial f(\bar{x}^0)}{\partial x_2} = 64$$

сильно различаются, то направление спуска не совпадает с направлением к точке минимума $\bar{x}^*$. Замена переменных для этой функции имеет вид:

$$y_1 = x_1; \quad y_2 = 4x_2.$$

Линии уровня $f(\bar{y})$ – концентрические окружности и $\nabla f(\bar{y}^0)$ теперь направлен в точку минимума, которую найдем за одну итерацию:

$$\alpha^1 = -\frac{(\nabla f^T(\bar{y}^0), \nabla f(\bar{y}^0))}{(\nabla f^T(\bar{y}^0)H, \nabla f(\bar{y}^0))} = -\frac{(4; 16) \begin{pmatrix} 4 \\ 16 \end{pmatrix}}{(4; 16) \begin{pmatrix} 2 & 0 \\ 0 & 32 \end{pmatrix} \begin{pmatrix} 4 \\ 16 \end{pmatrix}} = -0,5;$$

$\bar{y}^1 = \bar{y}^0 + \alpha^1 \nabla f(\bar{y}^0)$ или в координатной форме:

$$\begin{pmatrix} y_1^1 \\ y_2^1 \end{pmatrix} = \begin{pmatrix} y_1^0 \\ y_2^0 \end{pmatrix} + \alpha^1 \begin{pmatrix} \frac{\partial f(\bar{y}^0)}{\partial y_1} \\ \frac{\partial f(\bar{y}^0)}{\partial y_2} \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \end{pmatrix} - 0,5 \begin{pmatrix} 4 \\ 16 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Вернемся к переменным x_i и получим, что $\bar{x}^1 = (0; 0)$.

Если $f(\bar{x})$ не квадратичная функция, то выбирают

$$\mu_i = \left(\frac{\partial^2 f(\bar{x})}{\partial x_i^2} \right)^{-1/2}.$$

Такое преобразование не превратит поверхности уровня $f(\bar{x})$ в сферы, но уменьшит их вытянутость. Гарантированно исправить топографию функции $f(\bar{x})$ можно, если учесть все элементы матрицы Гессе, а не только ее диагональные члены.

Указанное преобразование приводит к методам второго порядка, в которых при выборе направления спуска учитывается кривизна поверхности функции цели.

Методы второго порядка

Метод Ньютона

В основу метода Ньютона положена аппроксимация функции $f(\bar{x})$ в окрестности точки $\bar{x}^{k-1}$ квадратичной формой:

$$\tilde{f}(\bar{x}) = f(\bar{x}^{k-1}) + (\nabla f^T(\bar{x}^{k-1}), (\bar{x} - \bar{x}^{k-1})) + \frac{1}{2}((\bar{x} - \bar{x}^{k-1})H(\bar{x}^{k-1}), (\bar{x} - \bar{x}^{k-1})). \quad (4.15)$$

Если матрица Гессе $H(\bar{x}^{k-1})$ положительно определена⁶, то квадратичная форма (4.15) имеет единственную точку минимума, которая является корнем уравнения:

$$\nabla \tilde{f}(\bar{x}) = \nabla f(\bar{x}^{k-1}) + H(\bar{x}^{k-1})(\bar{x} - \bar{x}^{k-1}) = 0. \quad (4.16)$$

Координаты точки минимума квадратичной формы $\tilde{f}(\bar{x})$ принимаются за очередное приближение $\bar{x}^k$ к экстремуму $f(\bar{x})$. Решив уравнение (4.16), получим:

$$\bar{x}^k = \bar{x}^{k-1} - H^{-1}(\bar{x}^{k-1})\nabla f(\bar{x}^{k-1}), \quad (4.17)$$

где $H^{-1}(\bar{x}^{k-1})$ – матрица, обратная матрице Гессе.

Способы вычисления обратных матриц см. [10].

Таким образом, метод Ньютона интерпретируется как последовательный поиск точек минимума квадратичных функций, аппроксимирующих функцию $f(\bar{x})$ в окрестности точек $\bar{x}^0, \bar{x}^1, \dots, \bar{x}^k$.

Геометрически итеративный процесс минимизации $f(\bar{x})$ выглядит так, как это показано на рисунке 4.7 на примере функции одной переменной.

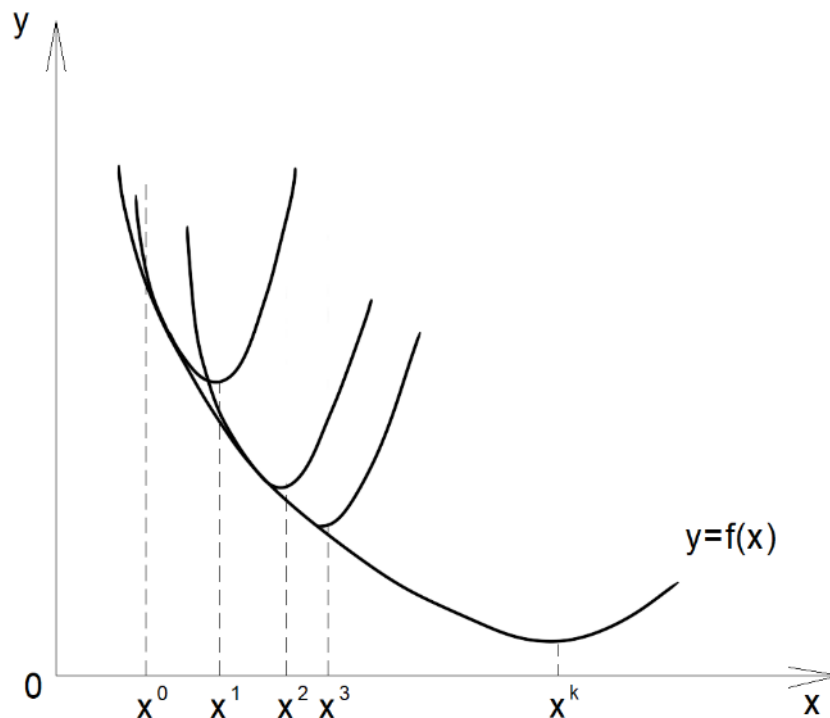


Рисунок 4.7. Геометрическая интерпретация метода Ньютона

⁶ В случае поиска максимума $f(\bar{x})$ матрица $H(\bar{x}^{k-1})$ должна быть отрицательно определена.

Из алгоритма (4.17) следует, что направление спуска на каждой итерации метода Ньютона определяется вектором, который получается при умножении градиента $\nabla f(\bar{x}^{k-1})$ на обратную матрицу Гессе. Это соответствует повороту $\nabla f(\bar{x}^{k-1})$ на некоторый угол, такой, что точка минимума аппроксимирующей функции $\tilde{f}(\bar{x})$ оказывается на направлении этого вектора. Поэтому шаг на направлении спуска полагается равным единице.

Вычислительный процесс в методе Ньютона, так же, как и в градиентных методах, заканчивается, когда на итерации k будет выполнено условие:

$$|\nabla f(\bar{x}^k)| \leq \delta.$$

Метод Ньютона сходится за меньшее число итераций, чем градиентные методы, так как обладает квадратичной скоростью сходимости (см. соотношение (4.8). Для квадратичных функций метод Ньютона сходится за одну итерацию, что очевидно из соотношения (4.15).

Продemonстрируем это свойство метода Ньютона на рассмотренном выше примере 4.1.

Матрица, обратная матрице Гессе для функции $f(\bar{x}) = x_1^2 + 16x_2^2$, имеет вид:

$$H^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/32 \end{pmatrix}.$$

Согласно (4.17) найдем:

$$\bar{x}^1 = \bar{x}^0 - H^{-1} \nabla f(\bar{x}^0);$$

$$\begin{pmatrix} x_1^1 \\ x_2^1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{32} \end{pmatrix} \begin{pmatrix} 4 \\ 64 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

таким образом координаты точки $\bar{x}^1 = (0; 0)$.

Недостатки метода Ньютона.

1. Несмотря на меньшее число итераций по сравнению с градиентными методами, общие затраты машинного времени в методе Ньютона могут оказаться значительными вследствие необходимости вычислять матрицу Гессе и обратную к ней матрицу.

2. Метод Ньютона сходится не из любой точки начального приближения. При неудачном выборе начального приближения последовательность $\{\bar{x}^k\}$ не будет стремиться к точке $\bar{x}^*$.

Условием сходимости вычислительного процесса в методе Ньютона является положительная определенность (а в случае максимума $f(\bar{x})$ отрицательная определенность) матрицы Гессе. На практике применяют сочетание метода градиента и метода Ньютона. Вдали от точки экстремума направление спуска выбирают по градиенту, а вблизи от точки экстремума, когда скорость сходимости градиентного метода падает, переходят к методу Ньютона.

Рассмотрим примеры решения некоторых задач [7] методом градиента и методом Ньютона.

4.3. Решение задач на условный экстремум. Методы штрафных функций

Методы штрафных функций применяются для решения задач с ограничениями в виде уравнений и неравенств, общая постановка которых имеет вид (см. также (4.1) + (4.3):

найти минимум

$$f(\bar{x}), \bar{x} = (x_1, x_2, \dots, x_n) \quad (4.18)$$

при ограничениях

$$h_i(\bar{x}) = 0, i = 1, \dots, m; \quad (4.19)$$

$$g_i(\bar{x}) \geq 0, i = m + 1, \dots, q. \quad (4.20)$$

В основу методов штрафных функций положена идея преобразования общей нелинейной задачи (4.18) – (4.20) на условный экстремум в последовательность задач без ограничений [7; 13]. В этом отношении методы штрафных функций схожи с методом множителей Лагранжа. Однако метод множителей Лагранжа оказывается малоэффективным в случае нелинейных задач большой размерности, так как приводит к решению системы уравнений, выражающей необходимые условия оптимальности для функции Лагранжа (см. соотношения (2.39) – (2.41)), что требует применения численных методов, реализация которых весьма трудоемка.

В методах штрафных функций вводится формальная функция цели, называемая расширенной функцией или просто штрафной функцией:

$$P(\bar{x}^k, r^k) = f(\bar{x}) + \sum_{i=1}^m r_i^k H(h_i(\bar{x}^k)) + \sum_{i=m+1}^q r_i^k G(g_i(\bar{x}^k)), \quad (4.21)$$

где $r_i^k, i = 1, \dots, q$ – весовые коэффициенты;
 $r_i^k \geq 0$;

$H(h_i(\bar{x}^k))$ и $G(g_i(\bar{x}^k))$ конструируются на основе функций $h_i(\bar{x})$ и $g_i(\bar{x})$;

k – число завершенных этапов вычислительного оптимизационного процесса.

Вид функций $H(h_i(\bar{x}))$ и $G(g_i(\bar{x}))$ выбирается таким образом, чтобы выполнялись условия:

$H(h_i(\bar{x})) \rightarrow 0$ при $h_i(\bar{x}) \rightarrow 0$, здесь обычно полагают

$$H(h_i(\bar{x})) = h_i^2(\bar{x});$$

$$G(g_i(\bar{x})) \rightarrow +\infty \text{ при } g_i(\bar{x}) \rightarrow 0^+, \quad (4.22)$$

для этого необходимо, чтобы точка $\bar{x}$ всегда была внутренней, т.е. удовлетворяла неравенству $g_i(\bar{x}) > 0, i = m + 1, \dots, q$; вместо требования (4.22) может быть выдвинуто требование

$$G(g_i(\bar{x})) \rightarrow 0 \text{ при } g_i(\bar{x}) \rightarrow 0^-; \quad (4.23)$$

при этом точка $\bar{x}$ должна быть внешней, т.е. $g_i(\bar{x}) < 0, i = m + 1, \dots, q$.

В зависимости от того, какое из требований принимается за основу при выборе $G(g_i(\bar{x}))$ (4.22) или (4.23), различают методы внутренней точки и методы внешней точки, а также комбинированные методы внутренней и внешней точки.

Методы внутренней точки

Рассмотрим задачу (4.18), (4.20), т.е. задачу с ограничениями только типа неравенств. Пусть в результате предварительного этапа исследования функции $f(\bar{x})$ (см. раздел 2.5) установлено, что решение задачи достигается на границе ОДР.

Для решения этой задачи методом внутренней точки штрафная функция $P(\bar{x}, r)$ формируется с учетом требования (4.22), причем наибольшее распространение получили штрафные функции двух видов:

$$P(\bar{x}, r) = f(\bar{x}) - r \sum_{i=m+1}^q \ln q_i(\bar{x}), \quad (4.24)$$

которая называется логарифмической штрафной функцией, и

$$P(\bar{x}, r) = f(\bar{x}) - r \sum_{i=m+1}^q \frac{1}{q_i(\bar{x})}, \quad (4.25)$$

называемая обратной штрафной функцией, где r – весовой коэффициент.

Нетрудно видеть, что функции (4.24), (4.25) удовлетворяют условию (4.22):

если $g_i(\bar{x}) \rightarrow 0^+$, $-\ln g_i(\bar{x}) \rightarrow +\infty$ и $\frac{1}{g_i(\bar{x})} \rightarrow +\infty$ (см. рис. 4.8).

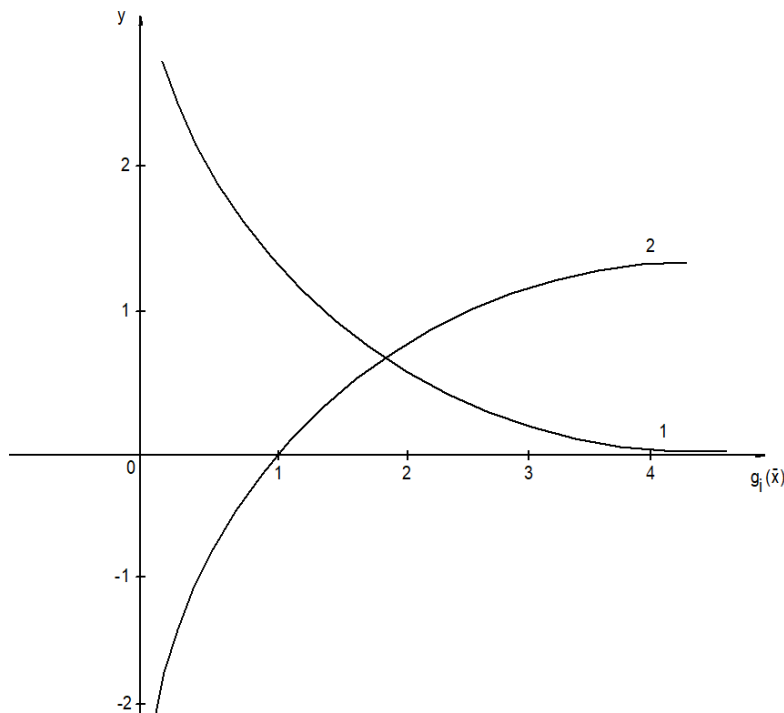


Рисунок 4.8. Иллюстрация выполнения условия (4.22) для штрафной функции

$$P(\bar{x}, r): 1 - y = \frac{1}{g_i(\bar{x})}; 2 - y = \ln q_i(\bar{x}).$$

Сформулировав для функции $P(\bar{x}, r)$ необходимые условия оптимальности, получим

$$\nabla P(\bar{x}, r) = \nabla f(\bar{x}) - r \sum_{i=m+1}^q \frac{1}{g_i(\bar{x})} \nabla g_i(\bar{x}) = 0 \quad (4.26)$$

или

$$\nabla P(\bar{x}, r) = \nabla f(\bar{x}) - r \sum_{i=m+1}^q \frac{1}{g_i^2(\bar{x})} \nabla g_i(\bar{x}) = 0. \quad (4.27)$$

Решив уравнения (4.26) и (4.27), найдем стационарные точки функции $P(\bar{x}, r)$ вида (4.24) и (4.25) соответственно. Очевидно, что положение стационарной точки будет зависеть от величины весового коэффициента r , т.е. $\bar{x} = \bar{x}(r)$.

В [7, 13] доказывается, что при определенных условиях, а именно: при существовании внутренних точек в ОДР задачи, т.е. таких точек $\bar{x}$, для которых $g_i(\bar{x}), i = m + 1, \dots, q$ выполняются как строгие неравенства;

- непрерывно дифференцируемых $f(\bar{x})$ и $g_i(\bar{x}), i = m + 1, \dots, q$;
- выпуклой $f(\bar{x})$ и вогнутых $g_i(\bar{x})$;
- матрица Гессе для функции $P(\bar{x}, r)$ относительно $\bar{x}$ не обращается в нуль ни в одной точке ОДР; функция $P(\bar{x}, r)$ будет выпуклой и последовательность стационарных точек

$$\{\bar{x}(r^k)\} \rightarrow \bar{x}^* \text{ при } r^k \rightarrow 0 \text{ и } k \rightarrow \infty, \quad (4.28)$$

где $\bar{x}^*$ – точка минимума исходной задачи.

Существенным требованием здесь является выпуклость штрафной функции $P(\bar{x}, r)$, при этом сходимость вычислительной процедуры (4.28) гарантируется. Однако, в [7, 13] показано, что сходимость может иметь место и при менее жестких условиях.

Алгоритм методов внутренней точки включает следующие операции:

1. Строится штрафная функция $P(\bar{x}, r)$ вида (4.24) или (4.25). В качестве начальной выбирается точка $\bar{x}^*$ – внутренняя точка ОДР. Задается весовой коэффициент r^1 .

2. Одним из методов безусловной оптимизации, например, методом наискорейшего спуска или методом Ньютона, осуществляется переход из точки $\bar{x}^*$ в точку $\bar{x}(r^1)$ локального минимума $P(\bar{x}, r^1)$. При этом $\bar{x}(r^1)$ обязательно будет внутренней точкой ОДР, так как в противном случае $G(g_i(\bar{x})) \rightarrow +\infty$.

3. Согласно принятому алгоритму, задается новый весовой коэффициент $r^2 < r^1$ и, начиная из точки $\bar{x}(r^1)$, определяется безусловный локальный минимум $P(\bar{x}, r^2)$.

4. Процедура повторяется, т.е. для каждого этапа k вычисляется локальный минимум $P(\bar{x}, r^k)$, начиная из точки $\bar{x}(r^{k-1})$ для строго убывающей последовательности $\{r^k\}$.

Сходимость вычислительного процесса существенно зависит от способа задания убывающей последовательности $\{r^k\}$ и, в частности, от величины r^1 .

При слишком больших начальных значениях r^1 точки $\bar{x}(r)$ располагаются слишком далеко от границы ОДР и на продвижение к границе затрачивается

значительное машинное время. При слишком малых r^1 точки $\bar{x}(r)$ оказываются близки к границе ОДР, тогда движение к $\bar{x}^*$ должно осуществляться вдоль границы, что в случае нелинейных ограничений затруднительно. В [7] предлагаются некоторые способы задания $\{r^k\}$, простейший из них состоит в том, что полагают $r^1 = 1$, а каждое последующее значение $r^k = \frac{1}{c} r^{k-1}$, где константа c принимает значения от 2 до 10.

Пример 4.2. Найти минимум $f(\bar{x}) = x_1 + x_2$

при ограничениях $g_1(\bar{x}) = -x_1^2 + x_2 \geq 0$;

$$g_2(\bar{x}) = x_1 \geq 0.$$

Примем в качестве $P(\bar{x}, r)$ логарифмическую штрафную функцию:

$$P(\bar{x}, r) = f(\bar{x}) - r \sum_{i=1}^2 \ln g_i(\bar{x}) = x_1 + x_2 - r[\ln(-x_1^2 + x_2) + \ln x_1].$$

Поскольку функция $P(\bar{x}, r)$ имеет простой вид, решим задачу аналитически. Сформулируем необходимые условия оптимальности для функции $P(\bar{x}, r)$:

$$\begin{aligned} \frac{\partial P(\bar{x}, r)}{\partial x_1} &= 1 + \frac{2rx_1}{-x_1^2 + x_2} - \frac{r}{x_1} = 0; \\ \frac{\partial P(\bar{x}, r)}{\partial x_2} &= 1 - \frac{r}{-x_1^2 + x_2} = 0. \end{aligned}$$

Решив систему уравнений, найдем:

$$x_1(r) = \frac{-1 \pm \sqrt{1+8r}}{4}; \quad x_2(r) = \frac{(-1 \pm \sqrt{1+8r})^2}{16} + r.$$

Учитываем только ту стационарную точку $\bar{x}(r)$, которая удовлетворяет ограничениям $g_i(\bar{x})$, т.е.:

$$x_1(r) = \frac{-1+\sqrt{1+8r}}{4}; \quad x_2(r) = \frac{(-1+\sqrt{1+8r})^2}{16} + r. \quad (4.29)$$

В табл. 4.2 приведены значения $x_1(r)$ и $x_2(r)$ для четырех различных значений r . Геометрическая интерпретация задачи приведена на рисунке 4.9. В пределе при $r^k \rightarrow 0$ минимизирующая последовательность $\{\bar{x}(r^k)\} = \{\bar{x}_1(r^k), \bar{x}_2(r^k)\}$ стремится к точке $x^* = (0; 0)$.

Таблица 4.2. Значения r и $\bar{x}(r)$ из примера 4.2

Номер итерации	r	$x_1(r)$	$x_2(r)$	$f(\bar{x}^r)$
1	1,00	0,500	1,250	1,75
2	0,50	0,309	0,595	0,904
3	0,25	0,183	0,283	0,466
4	0,10	0,085	0,107	0,192

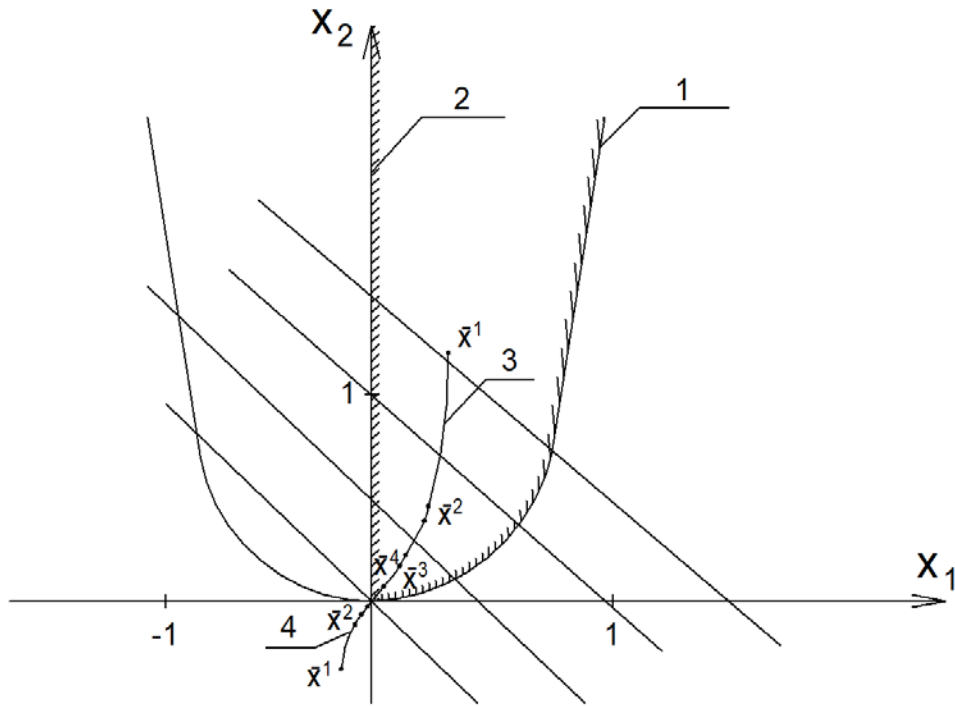


Рисунок 4.9. Геометрическая интерпретация задачи в примере 4.2:
1 – граница $g_1(\bar{x}) = 0$; 2 – граница $g_2(\bar{x}) = 0$; 3 – траектория спуска методом внутренней точки; 4 – траектория спуска методом внешней точки

В этой задаче при каждом значении r существует только один локальный экстремум. Это объясняется тем, что исходная задача имеет единственное решение. В задачах со многими локальными минимумами получают последовательность решений безусловных задач, сходящуюся к одному из локальных минимумов.

Методы внешней точки

В методах внешней точки штрафную функцию $P(\bar{x}, r)$ строят так, чтобы выполнялось условие (4.23). В отличие от методов внутренней точки, движение к экстремуму $\bar{x}^*$ может в этих методах начинаться из любой начальной точки, поэтому здесь не требуется существования внутренних точек у ОДР, вследствие чего, методы внешней точки применимы для решения задач с ограничениями в форме равенств. Для преобразования исходной задачи (4.18) – (4.20) в последовательность задач без ограничений наибольшее распространение получили штрафные функции двух видов:

$$P(\bar{x}, r) = f(\bar{x}) + r[\sum_{i=1}^m h_i^2(\bar{x}) + \sum_{i=m+1}^q G_i(\bar{x})], \quad (4.30)$$

называемая квадратичной функцией штрафа,

$$\text{где } G_i(\bar{x}) = \begin{cases} 0, & \text{если } g_i(\bar{x}) \geq 0 \\ g_i^2(\bar{x}), & \text{если } g_i(\bar{x}) < 0 \end{cases}$$

и функция

$$P(\bar{x}, r) = f(\bar{x}) + r[\sum_{i=1}^m |h_i(\bar{x})| + \sum_{i=m+1}^q G_i(\bar{x})], \quad (4.31)$$

где $G_i(\bar{x}) = \begin{cases} 0, & \text{если } g_i(\bar{x}) \geq 0 \\ g_i(\bar{x}), & \text{если } g_i(\bar{x}) < 0 \end{cases}$

Достоинство функции (4.30) по сравнению с (4.31) в том, что если $h_i(\bar{x}), g_i(\bar{x})$ непрерывно дифференцируемы, то и $P(\bar{x}, r)$ вида (4.29) будет обладать этим свойством, и в процедуре минимизации этой функции могут быть использованы методы первого и второго порядков. Поэтому чаще используется штрафная функция (4.30).

Сформулировав для функции (4.30) необходимые условия оптимальности аналогично (4.26), (4.27) и решив систему уравнений, найдем стационарные точки $P(\bar{x}, r)$, положение которых так же, как и в методах внутренней точки, будет зависеть от величины весового коэффициента r :

$$\bar{x} = \bar{x}(r).$$

В [11] доказано, что условия сходимости последовательности стационарных точек

$$\{\bar{x}(r^k)\} \rightarrow \bar{x}^* \text{ при } r^k \rightarrow \infty, \quad (4.32)$$

где $\bar{x}^*$ – точка условного минимума в исходной задаче (4.18) – (4.20), такие же, как в методах внутренней точки за исключением требования существования внутренних точек у ОДР.

Алгоритм методов внешних точек включает следующие операции.

1. Строится штрафная функция $P(\bar{x}, r)$, задается коэффициент r^1 и из произвольной начальной точки $\bar{x}^0$ осуществляется спуск в точку локального минимума $P(\bar{x}, r^1) - \bar{x}(r^1)$.

2. Согласно принятому алгоритму, задается весовой коэффициент $r^2 > r^1$ и, начиная из точки $\bar{x}(r^1)$, определяется безусловный локальный минимум $P(\bar{x}, r^2) - \bar{x}(r^2)$.

3. Процедура повторяется: на каждом этапе k решается задача безусловной минимизации функции $P(\bar{x}, r^k)$ из начальной точки $\bar{x}(r^{k-1})$ для строго возрастающей последовательности $\{r^k\}$.

Положение точек последовательности $\{\bar{x}(r^k)\}$ зависит от выбора значений $\{r^k\}$. Обычно полагают $r^1 = 1$, а каждое последующее значение r^k вычисляют: $r^k = cr^{k-1}$, где константа $c > 1$. С ростом r^k увеличивается вклад в функцию $P(\bar{x}, r)$ тех ограничений, от которых удалена точка $\bar{x}$. Поэтому при $r^k \rightarrow \infty$ безусловные локальные минимумы должны приближаться к ОДР, причем движение осуществляется из недопустимой области в допустимую.

Пример 4.3. Решим с помощью метода внешних точек задачу, рассмотренную в примере 4.2.

Составим квадратичную функцию штрафа

$$P(\bar{x}, r) = x_1 + x_2 + r[(-x_1^2 + x_2)^2 + x_1^2]$$

и найдем безусловный локальный минимум $P(\bar{x}, r)$ аналитическим методом:

$$\frac{\partial P(\bar{x}, r)}{\partial x_1} = 1 + r[-4x_1(-x_1^2 + x_2) + 2x_1] = 0,$$

$$\frac{\partial P(\bar{x}, r)}{\partial x_2} = 1 + 2r(-x_1^2 + x_2) = 0.$$

Решив систему уравнений, получим:

$$x_1(r) = -\frac{1}{2(r+1)}; \quad x_2(r) = \frac{1}{4(r+1)^2} - \frac{1}{2r}. \quad (4.33)$$

Очевидно, что точки $\bar{x}(r) = (x_1(r), x_2(r))$ – недопустимые, так как нарушается условие $g_1(\bar{x}) = x_1 \geq 0$.

Очевидно также, что при $r^k \rightarrow \infty \quad \bar{x}(r) \rightarrow \bar{x}^* = (0; 0)$.

Таблица 4.3. Результаты расчета $\bar{x}(r)$ и $f(\bar{x}(r))$

Номер итерации	r	$x_1(r)$	$x_2(r)$	$f(\bar{x})$
1	1	-0,250	-0,437	-0,687
2	2	-0,166	-0,222	-0,388
3	3	-0,125	-0,150	-0,275
4	10	-0,045	-0,048	-0,093

В таблице 4.3 приведены значения $x_1(r)$ и $x_2(r)$ для различных r^k , траекторию движения к точке $\bar{x}^*$ см. на рисунке 4.9.

Сравнительная оценка методов внутренней и внешней точки

1. Общим достоинством метода внутренней и внешней точки является возможность решения задач на безусловный экстремум с использованием методов прямого поиска.

2. При сопоставлении этих методов в качестве преимущества методов внутренней точки обычно указывают то обстоятельство, что при обращении к ним гарантируется соблюдение ограничений на всем протяжении поиска. Это существенно в том случае, когда функция цели $f(\bar{x})$ не определена за пределами ОДР и, кроме того, позволяет прекратить поиск на любой итерации, получив допустимое решение.

3. Методы внешних точек обладают тем достоинством, что не требуют существования внутренних точек у ОДР, соответственно не требуют определения начальной допустимой точки и пригодны для решения задач с ограниченными равенствами.

4. Важной особенностью методов внешней точки является также и то, что в штрафную функцию входят лишь ограничения, которые нарушаются в текущей точке $\bar{x}(r)$. По мере приближения к ОДР остается все меньше нарушенных ограничений и вычисление $P(\bar{x}, r^k)$ и ее производных упрощается. В методах внутренней точки в функциях штрафа $P(\bar{x}, r^k)$ участвуют все ограничения, а конструкция $P(\bar{x}, r^k)$ такова, что ограничения приходится вычислять и контролировать их соблюдение на каждом шаге поиска, поскольку

за пределами ОДР функция $P(\bar{x}, r^k)$ теряет смысл. Это увеличивает объем вычислений в методах внутренней точки.

5. Помимо указанных различий между методами внешней и внутренней точки у них имеются и общие свойства. Поверхность функции $P(\bar{x}, r^k)$ зависит от величины параметра r , причем с изменением r^k при $k \rightarrow \infty$ поверхность $P(\bar{x}, r)$ приобретает все более выраженный «овражный» характер.

4.4. Решение задач на условный экстремум. Методы допустимых направлений

В методах допустимых направлений реализуется идея движения к экстремуму функции цели путем последовательного перехода от точки к точке, которые являются допустимыми или почти допустимыми. В результате получают последовательность точек $\{\bar{x}^k\}$, удовлетворяющую ограничениям с точностью до ε , где ε – допустимая в задаче степень нарушения ограничений.

Последовательности точек $\{\bar{x}^k\}$ соответствует последовательность значений $\{f(\bar{x}^k)\}$, которая при $k \rightarrow \infty$ сходится к $f(\bar{x}^*)$.

В настоящее время существует множество различных методов допустимых направлений [5, 8]. Рассмотрим один из этих методов, называемый методом проектирования градиента. Метод проектирования градиента применяется для решения задач с ограничениями в виде уравнений и неравенств. Пусть задача нелинейного программирования задана в виде (4.18), (4.20).

Алгоритм метода проектирования градиента для этой задачи включает следующие операции.

1. Осуществляется допустимая начальная точка $\bar{x}^0$, и из точки $\bar{x}^0$ осуществляется движение методом наискорейшего спуска. На каждом шаге проверяется соблюдение ограничений $g_i(\bar{x}^k) \geq 0, i = m + 1, \dots, q$. Движение продолжается до тех пор, пока не будет достигнута точка, для которой становятся активными одно или несколько ограничений.

2. Пусть в точке $\bar{x}^k$ активными являются r ограничений, т.е. $g_i(\bar{x}^k) = 0, i = 1, \dots, r$. Тогда определяется допустимое направление движения к экстремуму $f(\bar{x})$. Поскольку точка $\bar{x}^k$ находится на границе ОДР, то движение к экстремуму должно осуществляться по поверхности ограничений, что в случае нелинейных функций $g_i(\bar{x})$ представляет значительные трудности. Поэтому в качестве допустимого принимается направление в плоскости, касательной к поверхности ограничений в точке $\bar{x}^k$. Векторы, расположенные в этой плоскости, находятся из условия ортогональности $\bar{u}(\bar{x}^k)$ и $\nabla g_i(\bar{x}^k)$

$$(\bar{u}, \nabla g_i(\bar{x}^k)) = 0, \quad i = 1, \dots, r, \quad (4.34)$$

где $\bar{u}$ – допустимый вектор.

Очевидно, что лучшим из множества допустимых направлений, задаваемых системой (4.34), будет такой вектор, движение вдоль которого

обеспечит наискорейшее убывание функции $f(\bar{x})$. Этот вектор должен иметь максимальную проекцию на направление антиградиента $-\nabla f(\bar{x}^k)$:

$$(\bar{u}, -\nabla f(\bar{x}^k)) \rightarrow \max. \quad (4.35)$$

В итоге, для выбора допустимого направления $\bar{u}(\bar{x}^k)$ необходимо решить задачу линейного программирования:

найти максимум линейной формы:

$$F(\bar{u}) = \sum_{i=1}^n \frac{\partial f(\bar{x}^k)}{\partial x_j} u_j \quad (4.36)$$

при ограничениях

$$\sum_{j=1}^n \frac{\partial g_i(\bar{x}^k)}{\partial x_j} u_j = 0, \quad i = 1, \dots, r. \quad (4.37)$$

Решив задачи (4.36), (4.37), найдем касательный вектор $\bar{u}^k$, обеспечивающий наискорейший спуск к экстремуму $f(\bar{x})$.

3. В направлении вектора $\bar{u}^k$ выполняется шаг и вычисляется точка $\bar{x}^{k+1}$:

$$\bar{x}^{k+1} = \bar{x}^k + \alpha^k \bar{u}^k. \quad (4.38)$$

Любой шаг из точки $\bar{x}^k$ в направлении вектора $\bar{u}^k$ приводит к нарушению ограничений (рис. 4.10) $g_i(\bar{x}) \geq 0$. Если степень нарушения ограничений не превышает допустимой

$$g_i(\bar{x}) > -\varepsilon, \quad i = 1, \dots, r, \quad (4.39)$$

то в направлении $\bar{u}^k$ может быть сделан еще шаг согласно алгоритму (4.38), в противном случае точка $\bar{x}^{k+1}$ проектируется на поверхность ограничений.

4. Проектирование точки $\bar{x}^{k+1}$ на поверхность ограничений может быть выполнено путем решения вспомогательной задачи минимизации функции вида

$$\Phi(\bar{x}) = \sum_{i=1}^{r_1} g_i^2(\bar{x}),$$

где r_1 – число нарушений в точке $\bar{x}^{k+1}$ ограничений;

тогда после серии шагов по антиградиенту $\Phi(\bar{x})$:

$$\bar{x}^s = \bar{x}^{s-1} - h^s \nabla \Phi(\bar{x}^{s-1})$$

найдем точку $\bar{x}^{k+1}$, в которой ограничения $g_i(\bar{x})$ нарушаются не больше, чем на ε (см. соотношение (4.39)).

5. После завершения п. 4 все операции алгоритма повторяются, начиная с п. 2.

Признаком конца вычислений является выполнение в некоторой точке $\bar{x}^k$ условия:

$$|\bar{u}^k, \nabla f(\bar{x}^k)| \leq \delta, \quad (4.40)$$

где δ – требуемая точность вычисления экстремума.

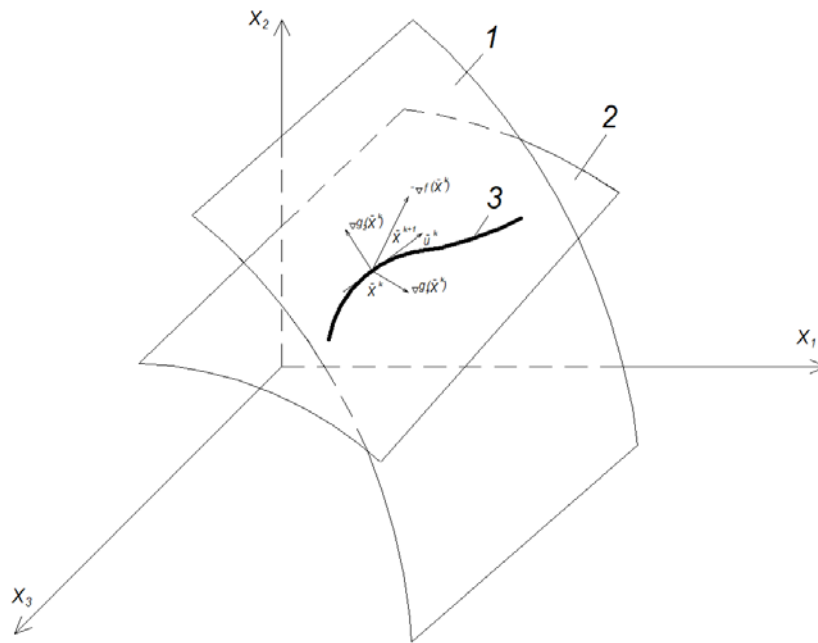


Рисунок 4.10. Определение допустимого направления движения к экстремуму $f(\bar{x})$ в пространстве переменных:
 1 – поверхность $g_1(\bar{x})$; 2 – поверхность $g_2(\bar{x})$; 3 – пересечение поверхностей $g_1(\bar{x})$ и $g_2(\bar{x})$, определяющее границу ОДР

Очевидно, что в точке $\bar{x}^*$ условного экстремума задачи (4.18), (4.20) проекция вектора $\bar{u}$ на направление $\nabla f(\bar{x}^*)$ должна быть равна нулю: $(\bar{u}, \nabla f(\bar{x}^*)) = 0$, что имеет место в случае ортогональности этих векторов (рис. 4.11), либо в случае равенства нулю одного из векторов (рис. 4.12). В окрестности точки $\bar{x}^*$ выполняется условие (4.40).

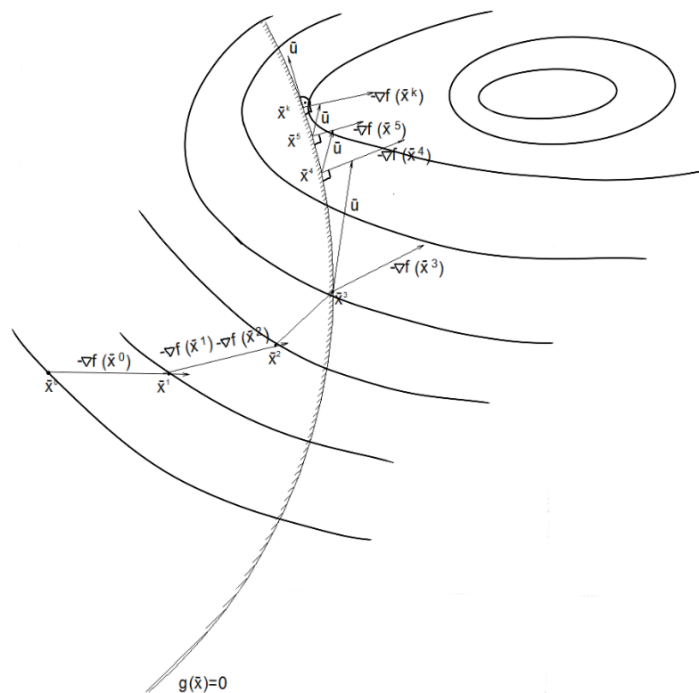


Рисунок 4.11. Геометрическая интерпретация метода проектирования градиента для случая ортогональности $\bar{u}$ и $\nabla f(\bar{x}^*)$

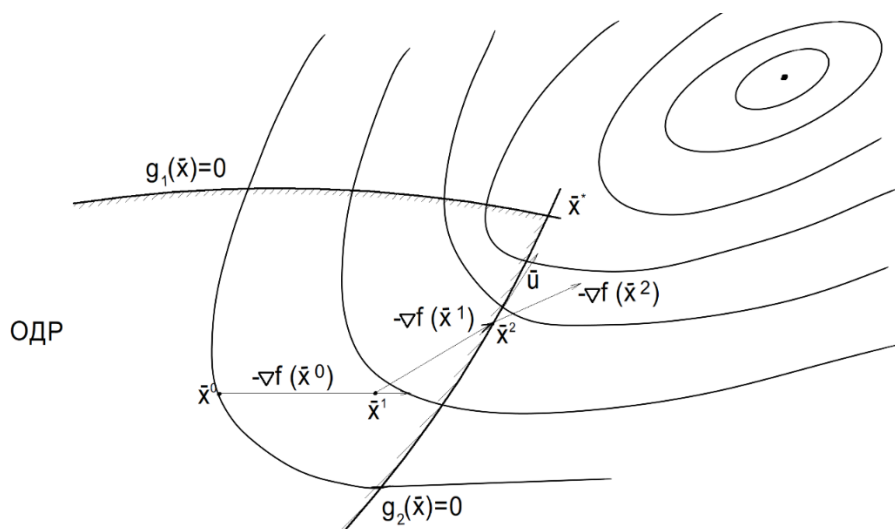


Рисунок 4.12. Геометрическая интерпретация метода проектирования градиента для случая, когда в точке $\bar{x}^*$ вектор допустимых направлений $\bar{u} = 0$

Пример 4.4. Найти минимум функции

$$f(\bar{x}) = x_1^2 + 1,5x_2^2 \quad (4.41)$$

при ограничении

$$g(\bar{x}) = x_1 + x_2 - 1 \geq 0 \quad (4.42)$$

с точностью

$$\delta = 0,15.$$

Геометрическая интерпретация задачи дана на рисунке 4.13.

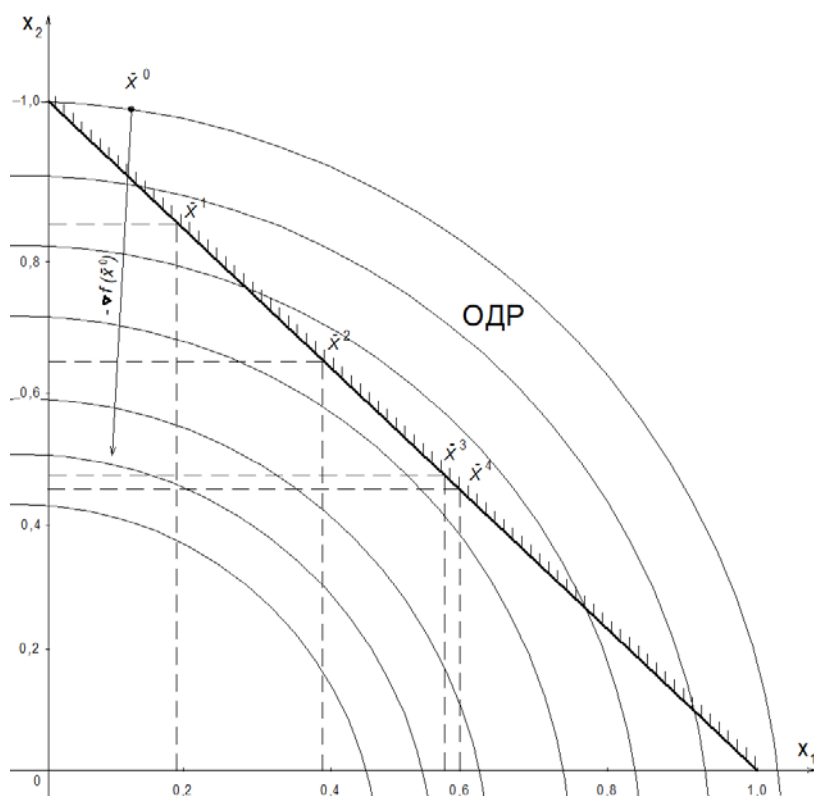


Рисунок 4.13. Геометрическая интерпретация решения задачи к примеру 4.4

Возьмем в качестве начальной точки внутреннюю точку ОДР с координатами $x_1^0 = 0,2$; $x_2^0 = 1,0$ и выполним операции алгоритма метода проектирования градиента.

Определим $\nabla f(\bar{x}) = \begin{pmatrix} 2x_1 \\ 3x_2 \end{pmatrix}$.

Вычислим $\nabla f(\bar{x}) = \begin{pmatrix} 0,4 \\ 3,0 \end{pmatrix}$; $|\nabla f(\bar{x}^0)| = 3,026 > \delta$.

Согласно алгоритму метода наискорейшего спуска

$$x_1^1 = x_1^0 - \alpha \frac{\partial f(\bar{x}^0)}{\partial x_1} = 0,2 - 0,4\alpha;$$

$$x_2^1 = x_2^0 - \alpha \frac{\partial f(\bar{x}^0)}{\partial x_2} = 1,0 - 3,0\alpha.$$

Найдем коэффициент α из условия, при котором точка $\bar{x}^1$ будет допустимой, т.е. не выйдет за границу ОДР; для этого подставим координаты x_1^1 и x_2^1 в неравенство (4.42):

$$x_1^1 + x_2^1 - 1 = 0,2 - 0,4\alpha + 1,0 - 3,0\alpha - 1 \geq 0.$$

Из условия $g(\bar{x}) = 0$ получим $\alpha = 0,0588$, при этом коэффициенте точка $\bar{x}^1$ с координатами $x_1^1 = 0,176$ и $x_2^1 = 0,824$ окажется на границе ОДР. Вычислим

$$\nabla f(\bar{x}^1) = \begin{pmatrix} 0,352 \\ 2,469 \end{pmatrix}; |\nabla f(\bar{x}^1)| = 2,494 > \delta.$$

Найдем допустимое направление $\bar{u}$, касательное к границе ОДР, решив задачу линейного программирования:

$$\begin{aligned} (-\nabla f(\bar{x}^1), \bar{u}) &\rightarrow \max_{\bar{u}} \\ (\nabla g(\bar{x}^1), u) &= 0 \end{aligned},$$

которая после подстановки численных значений $-\nabla f(\bar{x}^1)$ и $\nabla g(\bar{x}^1)$ примет вид:

$$\begin{aligned} F(\bar{u}) &= -0,352u_1 - 2,469u_2 \rightarrow \max \\ u_1 + u_2 &= 0. \end{aligned}$$

Из уравнения связи найдем, что $u_1 = -u_2$, а из условия максимума линейной формы F получим

$$u_1 > 0; u_2 < 0.$$

Так как нас интересует только направление вектора $\bar{u}$, примем, что $u_1 = 1$, тогда $u_2 = -1$. Очевидно, что направление вектора $\bar{u}$ совпадает с границей ОДР благодаря линейному ограничению (4.42), поэтому в данной задаче движение в допустимом направлении фактически будет движением по границе ОДР. Каждая очередная точка $\bar{x}^k$ на направлении $\bar{u}$ вычисляется по алгоритму:

$$\bar{x}^k = \bar{x}^{k-1} + \alpha \bar{u}.$$

Результаты вычислений для каждого шага в направлении $\bar{u}$ к минимуму $f(\bar{x})$ приведены в таблице 4.4.

Таблица 4.4. Результаты решения задачи методом проектирования градиента

Номер итерации k	Координаты		Координаты		$\nabla f(\bar{x}^k)$	$ (-\nabla f, \bar{u}) $	$f(\bar{x}^k)$	α
	x_1^k	x_2^k	$\frac{\partial f}{\partial x_1}$	$\frac{\partial f}{\partial x_2}$				
1	0,176	0,824	0,352	2,469	2,494	2,117	1,049	0,0558
2	0,376	0,624	0,752	1,872	2,017	1,12	0,725	0,2
3	0,576	0,424	1,152	1,272	1,716	0,12	0,601	0,2

На третьей итерации выполнилось условие конца вычислений (4.40):

$$|(-\nabla f(\bar{x}^3), \bar{u})| = 0,12 < \delta,$$

следовательно, точка $\bar{x}^3$ с координатами $x_1^3 = 0,576, x_2^3 = 0,424$ с точностью $\delta \leq 0,15$ является точкой условного минимума $f(\bar{x})$. Для сравнения приведем координаты точки $\bar{x}^*$: $x_1^* = 0,6; x_2^* = 0,4$. Чтобы найти $\bar{x}^*$ с большей точностью, следует уменьшить шаг α . Метод проектирования градиента при выпуклой функции $f(\bar{x})$ и вогнутой $g_i(\bar{x})$ позволяет определить глобальный минимум; если $f(\bar{x})$ невыпуклая функция, то находится один из ее локальных минимумов.

К недостаткам метода следует отнести большой объем вычислений, необходимость решения вспомогательных экстремальных задач при определении допустимых направлений $\bar{u}$ и проектировании точки $\bar{x}^k$ на поверхность ограничений.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Исследование операций: цикл статей // Хабр : [сайт]. – 2026. – URL: <https://habr.com/ru/articles/562774/> (дата обращения: 27.03.2026). – Текст: электронный.
2. Цыпкин, Я. З. Адаптация и обучение в автоматических системах / Я. З. Цыпкин. – М.: Наука, 1968. – 400 с. – Текст: непосредственный.
3. Смирнов, В. И. Курс высшей математики. Т.1 / В. И. Смирнов. – М. : Наука, 1974. – 479 с. – Текст: непосредственный.
4. Фихтенгольц, Г.М. Основы математического анализа. Часть 1 : учебник для вузов / Г. М. Фихтенгольц. – 18-е изд., стер. – СПб.: Лань, 2026. – 448 с. – Текст : непосредственный.
5. Васильев, Ф. П. Методы оптимизации. В 2-х ч. Часть I. Конечномерные задачи оптимизации. Принцип максимума. Динамическое программирование : учебник / Ф. П. Васильев. – М.: Изд-во МЦНМО, 2021. – 619 с. – Текст : непосредственный.
6. Ицкович, Э. Л. Контроль производства с помощью вычислительных машин – Э. Л. Ицкович. – М.: Энергия, 1975. – 417 с. – Текст: непосредственный.
7. Нелинейное программирование в современных задачах оптимизации : учебное пособие / Ю. В. Бородакий, А. М. Загребаев, Н. А. Крицына [и др.]. – М.: НИЯУ МИФИ, 2011. – 244 с. – Текст : непосредственный.
8. Зуховицкий, С. И. Линейное и выпуклое программирование / С. И. Зуховицкий, Л. И. Авдеева. – М.: Наука, 1967. – 460 с. – Текст: непосредственный.
9. Вентцель, Е. С. Исследование операций: задачи, принципы, методология / Е. С. Вентцель. – 5-е изд., стер. М.: КноРус, 2014. – 192 с. – Текст: непосредственный.
10. Кузнецов, А. В. Высшая математика. Математическое программирование : учебник / А. В. Кузнецов, В. А. Сакович, Н. И. Холод; под общ. ред. А.В. Кузнецова. – Изд. 4-е, стер.. – СПб. : Лань, 2013. – 351 с. – Текст: непосредственный.
11. Фиакко, А. Нелинейное программирование. Методы последовательной безусловной оптимизации / А. Фиакко, Г. Мак-Кормик. – М.: Мир, 1972. – 240 с. – Текст: непосредственный.
12. Акулич, И. Л. Математическое программирование в примерах и задачах : учебное пособие / И. Л. Акулич. – 3-е изд., стер. – СПб.: Лань, 2022. – 352 с. – Текст: непосредственный.

Учебное издание

Ремизова Ирина Викторовна

Исследование операций и методы оптимизации

Учебное пособие

Редактор и корректор Д. А. Романова
Техн. редактор М. Д. Баранова

Темплан 2025 г., поз. 5059

Подписано к печати 07.04.2026.	Формат 60x84/16.	Бумага тип № 1.
Печать офсетная.	Печ. л. 6,6.	Уч.-изд. л. 6,6.
Тираж 30 экз.	Изд. № 23.	Цена «С».
		Заказ №

Ризограф Высшей школы технологии и энергетики СПбГУПТД,
198095, Санкт-Петербург, ул. Ивана Черных, 4.