

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Нижегородский государственный университет  
им. Н.И. Лобачевского

С.Ю. Городецкий

# Лекции по нелинейному математическому программированию

Учебно-методическое пособие

Рекомендовано методической комиссией  
института Информационных технологий, математики и механики  
для студентов ННГУ, обучающихся по направлениям подготовки  
01.03.02 «Прикладная математика и информатика»  
и 02.03.02 «Фундаментальная информатика и информационные технологии»

Нижний Новгород  
2020

УДК 519.85  
ББК В18  
Г-70

Г-70 Городецкий С.Ю. ЛЕКЦИИ ПО НЕЛИНЕЙНОМУ МАТЕМАТИЧЕСКОМУ ПРОГРАММИРОВАНИЮ: Учебно-методическое пособие. – Нижний Новгород: Нижегородский госуниверситет, 2020. – 173с.

*Рецензенты:* кандидат физ.-мат. наук, доцент **А.Г. Коротченко**.

Учебно-методическое пособие написано на основе текстов лекций, читаемых автором в институте Информационных технологий, математики и механики в рамках дисциплин «Методы оптимизации» для направления подготовки бакалавриата «Прикладная математика и информатика (общий профиль)» (ПМИ), а также «Методы оптимизации» и «Модели и методы вычислительной оптимизации» для направления подготовки бакалавриата «Фундаментальная информатика и информационные технологии» (ФИИТ).

Пособие содержит сведения преимущественно теоретического характера по динамическому программированию, элементам выпуклого анализа, условиям оптимальности в задачах математического программирования, численными методами локальной и многоэкстремальной оптимизации. Этот материал читается во всех указанных дисциплинах.

Пособие предназначено для студентов бакалавриата направлений ПМИ (общий профиль) и ФИИТ, а также может быть использовано в качестве дополнительного материала студентами направления «Программная инженерия» при изучении дисциплины «Методы нелинейной оптимизации».

УДК 519.85  
ББК В18

© Нижегородский государственный  
университет им. Н.И. Лобачевского, 2020

# Предисловие

Учебно-методическое пособие написано на основе курсов лекций по методам оптимизации, читаемых автором на нескольких направлениях подготовки бакалавров в институте Информационных технологий, математики и механики. Основные направления: «Прикладная математика и информатика (общий профиль)» (ПМИ) и «Фундаментальная информатика и информационные технологии» (ФИИТ). В качестве дополнительного материала учебное пособие может быть использовано на направлении «Программная инженерия», а также на очно-заочной форме обучения. Один из читаемых автором лекционных курсов включает в себя задачи и методы как конечномерной, так и бесконечномерной оптимизации (на направлении ПМИ), а остальные — только конечномерной (*математическое программирование*).

Следует отметить, что весь совокупный материал читаемых автором курсов по методам оптимизации издан небольшим тиражом в 2018 году в виде печатного учебного пособия «Лекции по методам нелинейной оптимизации» [1]. Поскольку теория и методы решения задач линейного программирования традиционно читаются на указанных направлениях подготовки в виде небольшого отдельного курса «Линейное программирование», этот материал не был включен в изданное учебное пособие, что и определило его название.

Наиболее востребованным студентами является лекционный материал по нелинейным конечномерным задачам и методам оптимизации, в виде учебно-методического пособия в электронной форме. Это побудило автора к подготовке такого электронного издания. По сравнению с печатным изданием 2018 года в данном пособии исправлены замеченные ошибки и сделаны некоторые дополнения. Изложение и иллюстративный материал пособия достаточно полно соответствуют тому, как они совокупно представлены на лекциях по нелинейной конечномерной оптимизации (*нелинейному математическому программированию*) на ПМИ и ФИИТ. Каждый из тематических разделов представлен в наиболее полном из двух вариантов изложения. По этой причине материал сгруппирован не по лекциям, а по главам.

Подробность изложения соответствует тому относительно небольшому объему часов, которое отведено для изучения дисциплин учебными планами: лекционные курсы по методам оптимизации на ПМИ и ФИИТ читаются в семестровом варианте. При лекционных курсах рабочими программами дисциплин предусмотрены практические занятия и лабораторные практикумы. Комплекты задач для практических занятий с примерами их решения [10], а также описания теории и заданий для лабораторных практикумов [4], [21], [31], представлены отдельными методическими изданиями. Значительная часть этого материала в читаемых лекциях не отражена и в настоящее пособие не включена. Материал лабораторных практикумов вынесен для са-

мостоятельного изучения студентами. Ссылки на необходимые методические издания [4], [21], [31] приведены в соответствующих разделах пособия. Материал этих изданий следует рассматривать как теоретическое расширение основного содержания лекционных курсов, а также содержания данного учебного пособия. Этот материал является неотъемлемой частью изучаемых дисциплин и включается, наряду с материалом лекций, в теоретическую часть программ экзаменов.

Порядок изложения в пособии соответствует лекционному и во многом продиктован потребностями практических занятий, а также лабораторного практикума. Он согласован со сложившейся методикой проведения всех этих видов занятий. Это определило, в частности, то, что раздел «Динамическое программирование» включен в начальную часть пособия как отдельная глава, хотя формально является частью математического программирования.

Структура распределения материала по конечномерной оптимизации в лекционных курсах и в пособии представлена на рис. 1.

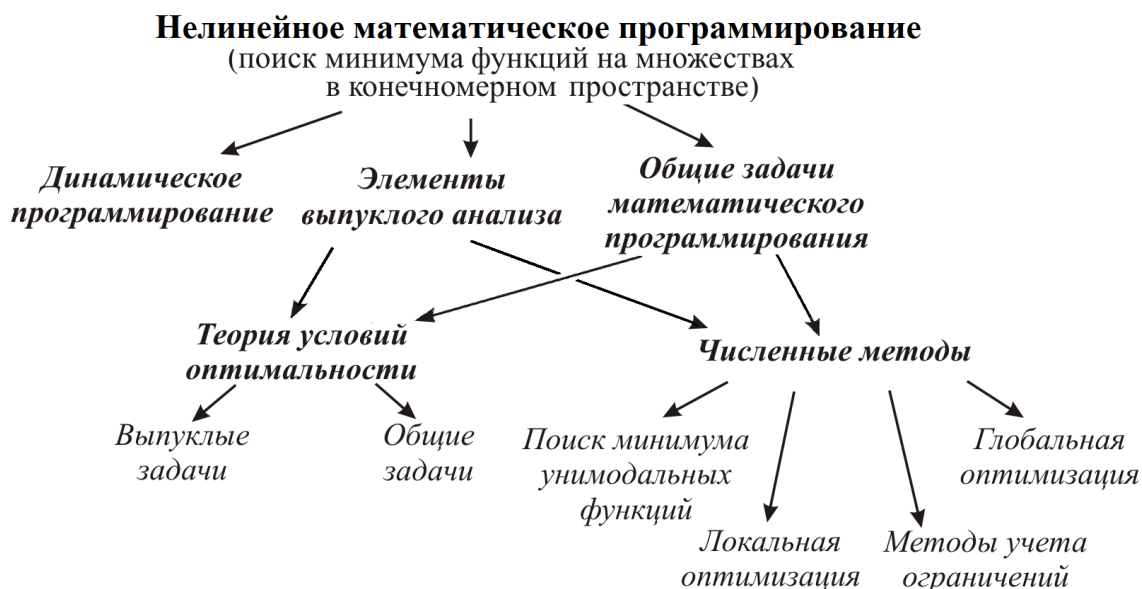


Рис. 1. Основные разделы нелинейного математического программирования в структуре лекционных курсов

Приведем краткий обзор содержания.

В главе 1 кратко рассмотрены общие постановки задач нелинейной оптимизации, вводятся понятия глобального и локального решений, рассматривается обобщение этих постановок и понятий решения на случай задач многокритериальной оптимизации, вводятся концепции оптимальности по Парето и Слейтеру. Студенты, более детально интересующиеся проблематикой и постановками конечномерных однокритериальных задач нелинейной оптимизации могут найти дополнительную информацию, например, в первой вводной главе книги [5].

Глава 2 посвящена теории дискретного (по времени) динамического программирования, методу рекуррентных уравнений Р. Беллмана, их обобщениям на задачи с нефиксированной длительностью процесса, принципу Беллмана. Расширенный материал по тематике первых двух глав приведен в описании лабораторного практикума по первой лабораторной работе [4], посвященной задачам оптимизации на графах с векторными весами ребер, методам сверток, применению метода рекуррентных уравнений Беллмана для задач на графах, методу Дейкстры, а также предложенному автором данного пособия безсверточному методу, позволяющему строить сразу все множество оптимальных по Парето путей на графах с неотрицательными векторными весами ребер. Эта лабораторная работа выполняется с использованием специализированной компьютерной программы учебно-исследовательского характера *OptWay*, разработанной в лаборатории «Динамика и оптимизация» кафедры теории управления и динамики систем ИИТММ, автор — Городецкий С.Ю.

Глава 3 предоставляет необходимую теоретическую основу в области теории условий оптимальности для задач нелинейного математического программирования. Условия оптимальности позволяют в ряде случаев аналитически находить решения задач на экстремум при ограничениях. Однако основное их назначение в другом, — они являются основой построения эффективных численных методов решения задач с ограничениями.

В начале главы приводятся необходимые сведения из математического анализа, включая теоремы об условиях оптимальности Ферма и методе множителей Лагранжа. Затем изучаются необходимые элементы выпуклого анализа. После этого последовательно строится теория условий оптимальности для задач со смешанными ограничениями в виде равенств и неравенств. Результаты по условиям оптимальности в недифференциальной форме (записанные через условие минимума или в терминах седловой точки функции Лагранжа), а также условия в градиентной форме (т.е. условия первого порядка) аккуратно обосновываются. Условия оптимальности второго порядка в задачах со смешанными ограничениями приводятся без доказательства.

Материал по условиям оптимальности, записанным в градиентной форме, подкрепляется второй лабораторной работой по анализу и исследованию двумерных невыпуклых задач с ограничениями на основе геометрической интерпретации условий оптимальности и их численного решения. В этой работе используются возможности одного из общих математических пакетов *Mathcad*, *MatLab* или *Mathematica*<sup>1</sup>, а именно, графические средства визуализации предложенных задач, возможности численного решения систем

---

<sup>1</sup>Лицензионные версии пакетов *Mathcad* производителя Mathsoft Engineering & Education, Inc. (США), *MatLab* производителя MathWorks® (американская компания с центральным офисом в США) и *Mathematica* производителя Wolfram Research, Inc. (США) установлены в компьютерных классах кафедры теории управления и динамики систем института ИТММ.

нелинейных уравнений и неравенств, численные методы непосредственного решения экстремальных задач с ограничениями.

Глава 4 целиком посвящена численным методам решения задач математического программирования. Эти методы являются практическим инструментом решения прикладных задач, и, как неотъемлемая часть, включаются в математические пакеты и системы автоматизированного проектирования. Без их использования сегодня не решается ни одна серьезная задача в области технического проектирования. Численные методы оптимизации применяются также в области финансов, в экономике и других сферах деятельности.

Материал главы разбит на несколько блоков. Вначале формализуется понятие численного метода математического программирования и рассматривается задача построения численных методов как оптимальных (в смысле гарантированного результата) или последовательно-оптимальных решающих правил. Техника получения таких методов демонстрируется на примере класса унимодальных функций. Рассмотрены также специальные методы поиска глобального минимума выпуклых функций на выпуклом многограннике.

Отдельный блок посвящен численным методам поиска локального минимума функций достаточно общего вида в  $R^N$ . Приведены алгоритмы выбора шагового множителя. Подробно рассмотрены простые методы первого и второго порядков (метод наискорейшего градиентного спуска и метод Ньютона), а также метод прямого поиска Хука-Дживса. Для первых двух анализируются условия и порядки скорости сходимости, а также другие свойства, обсуждаются как недостатки, так и положительные стороны этих простых методов.

Более изощренные методы (сопряженных градиентов, Ньютона с регулировкой шага и модификацией матрицы Гессе на положительную определенность, квазиньютоновские и некоторые другие) представлены в лекционных курсах и данном пособии, в основном, лишь в виде краткого обзора. Из них подробно (включая вывод всех соотношений) рассмотрен только метод сопряженных градиентов Флетчера-Ривса. Освоение теории по остальным указанным методам вынесено на самостоятельное изучение студентами в рамках существующих при лекционных курсах лабораторных практикумов. Соответствующая (третья) лабораторная работа выполняется в специализированной программной среде *LocOpt* (разработка лаборатории «Динамика и оптимизация» кафедры теории управления и динамики систем ИИТММ, автор — Городецкий С.Ю.). Описание среды *LocOpt*, а также необходимой теории по изучаемым методам содержится в методическом пособии [21].

Следующий блок материала связан с общими численными методами учета ограничений-равенств и неравенств в задачах локальной оптимизации. Такие задачи являются более сложными для численного решения по сравнению со случаем отсутствия ограничений. Рассмотрены методы внешнего штрафа, а также методы модифицированных функций Лагранжа. Построение методов последнего типа существенно опирается на теорию условий оптимальности,

рассмотренную в главе 3. Изучение методов учета ограничений поддержано четвертой лабораторной работой, выполняемой в указанной выше программной среде *LocOpt*, которая, в частности, позволяет самостоятельно конструировать тестовые задачи с ограничениями, имеет развитые графические средства визуализации и исследования структуры построенных задач, а также позволяет в наглядной форме изучать работу методов локальной оптимизации с учетом ограничений.

Последний блок четвертой главы посвящен методам многоэкстремальной оптимизации. Следует заметить, что многие прикладные задачи могут быть поставлены именно как задачи глобальной (многоэкстремальной) оптимизации, поскольку сводятся к отысканию допустимых значений параметров с наименьшими значениями минимизируемой функции в целом по допустимому множеству. Проблема, однако, заключается в том, что решение таких задач требует значительного количества вычислений, быстро возрастающего с увеличением размерности. В ситуациях, когда вычисления значений функций дорогостоящи, требуется разработка специальных решающих правил, позволяющих эффективно планировать размещение новых точек измерений с учетом ранее накопленной информации о проведенных вычислениях функций.

Исследования в ННГУ им. Н.И. Лобачевского по этой тематике связаны с именами Ю.И. Неймарка и Р.Г. Стронгина [27], [28], а также и их учеников (среди многочисленных публикаций см., например, [29], [30], а также [2]). Следует отметить, что Нижегородская школа многоэкстремальной оптимизации, руководимая Р.Г. Стронгиным, является признанным лидером в области разработки методов глобальной оптимизации указанного типа. В основе многих из разработанных методов лежат вероятностные модели поведения многоэкстремальных липшицевых функций.

В лекционном курсе, а также учебном пособии приводится обзор основных подходов и методов Нижегородской школы многоэкстремальной оптимизации, однако большая часть материала посвящена классическим вопросам липшицевой многоэкстремальной оптимизации и более простым геометрическим методам, на которых иллюстрируются основные концепции. Этот материал подкрепляется пятой лабораторной работой [31] по методам одномерной глобальной оптимизации.

В заключение автор выражает глубокую благодарность магистру направления «Прикладная математика и информатика» Кандаловой Татьяне за неоценимую помощь в подготовке рукописи.

*С.Ю. Городецкий*

# Глава 1. Постановки задач оптимизации, понятия решения

## 1.1. Общий вид рассматриваемых задач, локальные и глобальные решения

Пусть задано множество элементов  $D \subseteq B$ , где  $B$  — некоторое линейное нормированное пространство с метрикой, порождаемой нормой. Если такое пространство обладает еще и свойством полноты (т.е. любая фундаментальная последовательность его элементов сходится к некоторому элементу из этого же пространства), то оно является банаховым пространством. Банаховы пространства позволяют рассматривать бесконечномерные задачи, элементами которых могут быть, например, функции с определенными свойствами. В данном пособии ограничимся конечномерными задачами, когда в качестве  $B$  будет принято пространство  $R^N$ , где  $R^N$  — действительное нормированное пространство размерности  $N$ .

Пусть на множестве  $D$  определена функция

$$Q : D \rightarrow R^1,$$

которая описывает «качество» элемента  $z$  из  $D$ . Функцию  $Q(z)$  будем называть *целевой функцией*,  $D$  — *допустимым множеством*, любой элемент  $z \in D$  — *допустимым элементом*. Для определенности будем считать, что «лучшим» является тот допустимый элемент  $z$ , который соответствует меньшему значению  $Q(z)$ .

Далее всегда будем предполагать, что  $D$  и  $Q$  обладают свойствами, обеспечивающими достижимость точной нижней грани значений  $Q(z)$  для  $z \in D$ . Это будет выполнено, например, при компактности  $D$  из  $R^N$  и непрерывности  $Q$  на множестве  $D$ .

Задачи оптимизации вида

$$Q(z) \rightarrow \min, \quad z \in D \tag{1.1}$$

можно ставить различным образом: как задачи определения глобального или локального минимумов.

**Определение.** *Глобальным минимумом* целевой функции  $Q$  на  $D$  называют элемент  $z^* \in D$ , для которого  $\forall z \in D : Q(z^*) \leq Q(z)$ .

Будем использовать следующую форму записи:

$$z^* = \arg \min \{Q(z) : z \in D\},$$



$$Q^* = Q(z^*) = \min \{Q(z) : z \in D\}.$$

В ряде случаев бывает достаточно определить не глобально-оптимальное, а лишь локально-оптимальное решение.

**Определение.** Элемент  $z^o \in D$  называют *локальным минимумом* задачи (1.1), если  $\exists \varepsilon > 0$ , что  $\forall z \in D \cap O_\varepsilon(z^o) : Q(z^o) \leq Q(z)$ , где  $O_\varepsilon(z^o) = \{z \in B : \|z - z^o\| < \varepsilon\}$  — открытый шар радиуса  $\varepsilon$  с центром в  $z^o$ . Если в этом определении для рассматриваемых  $z \neq z^o$  выполняется строгое неравенство  $Q(z^o) < Q(z)$ , то *локальный минимум называют строгим*.

Понятие локального минимума зависит от способа задания нормы. Это наиболее существенно для случая функциональных пространств  $B$ .

**Замечание.** Если рассматривать глобальный или локальный минимумы функции  $Q$  в задаче без ограничений, т.е. когда в (1.1)  $D = B$ , то их называют безусловными минимумами (глобальным и локальным). Для того, чтобы подчеркнуть влияние ограничений, минимумы функции  $Q(x)$  (глобальные и локальные) при  $D \neq B$  иногда называют условными.

На рис. 1.1 для случая  $B = R^N$  с  $N = 1$  и  $D = [a, b]$  показана ситуация, когда функция имеет континуум локальных минимумов.

Далее будем предполагать, что множество  $D$  неявно задано с помощью требований, накладываемых на значения функций-ограничений

$$g = (g_1, \dots, g_m)^T : E \rightarrow R^m \quad \text{и} \quad h = (h_1, \dots, h_p)^T : E \rightarrow R^p,$$

$$D = \{z \in E \subseteq B : g(z) \leq 0, h(z) = 0\}, \quad (1.2)$$

где  $E$  — либо все пространство  $B$ , т.е.  $E = B$ , либо множество простой геометрической структуры, например, для случая  $B = R^N$

$$E = \Pi = \{z \in R^N : a \leq z \leq b\}.$$

Если  $B$  — функциональное пространство, т.е. его элементами  $z$  являются функции с определенными свойствами, то функции  $Q, g, h$  из (1.1), (1.2) принято называть *функционалами*.

Если же задача (1.1), (1.2) рассматривается в конечномерном пространстве  $B = R^N$ , ее называют *задачей математического программирования*.

Может показаться, что постановки задач (1.1), (1.2) чрезвычайно просты и здесь нет предмета для изучения. Однако это далеко не так. Дело в том, что функции  $(Q, g, h)$ , входящие в постановки задач, даже в более простом случае  $B = R^N$  могут обладать существенно разными свойствами (линейные, квадратичные, выпуклые, гладкие, липшицевы и т.д.) и могут быть заданы в различной форме (аналитически, алгоритмически). Принадлежность функций

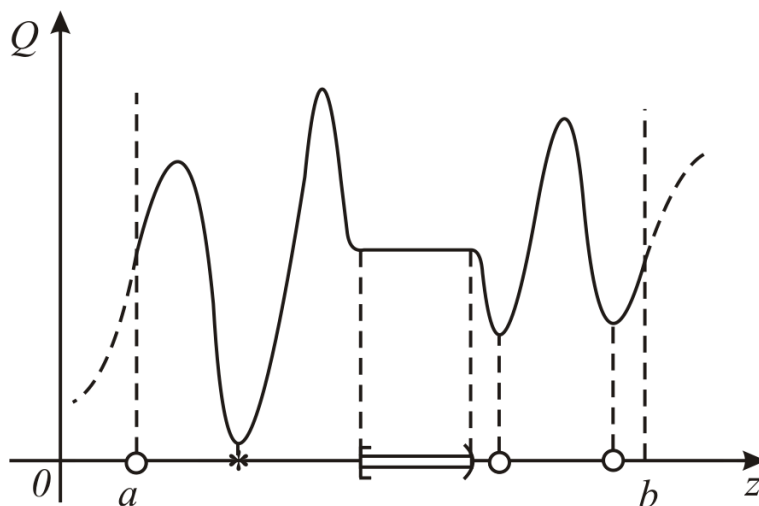


Рис. 1.1. Локальный и глобальный минимумы:  $\circ, *, \text{E}$  — точки локальных минимумов;  $*$  — точка глобального минимума

задачи разным классам требует различных методов определения решения. Например, алгоритмическая форма задания функций в задачах математического программирования неизбежно приводит к необходимости построения численных методов, основанных на проведении измерений локальных характеристик функций задачи (их значений, производных и т.п.) в отдельных точках множества  $E$  из (1.2).

## 1.2. Другие постановки задач оптимизации

В реальных задачах целевая функция  $Q$  может зависеть не только от аргумента  $z$  из множества  $D$ , но еще и от некоторого неконтролируемого фактора  $\xi$ :  $Q = Q(z, \xi)$ . В зависимости от того, какой смысл имеет указанная величина  $\xi$ , существуют различные постановки задач оптимизации. Укажем на некоторые.

1. *Задачи оптимизации в условиях противодействия.* Неопределенность  $\xi$  вызвана противодействием разумного противника. Например, мы минимизируем по  $z$  функцию  $Q$ , а противник максимизирует ее по  $\xi$ . Принятие решений в условиях разумного противодействия является объектом исследования теории игр.
2. *Задача оптимизации в условиях риска.* В этом случае  $\xi$  — стохастический фактор, т.е. случайная величина с известным законом распределения. Принятие решения в этих условиях изучается теорией оптимальных статистических решений (игры с «природой»).
3. *Задачи в условиях неопределенности.* Неопределенность вызвана недостаточной осведомленностью об условиях, в которых осуществляется выбор решения. Для неопределённого фактора  $\xi$  может быть известно только

множество, в котором он может принимать свои значения,  $\xi \in \Sigma$ . Величина  $\xi$  неизвестна в момент выбора  $z$ .

4. *Многокритериальные задачи.* Они возникают в том случае, когда множество  $\Sigma = \{\xi_1, \xi_2, \dots, \xi_n\}$  и все возможные значения  $\xi$  одновременно реализуются. Это эквивалентно наличию конечного числа целевых функций, образующих векторный критерий. Многокритериальная задача запишется следующим образом:

$$Q(z) = \{Q_1(z), Q_2(z), \dots, Q_n(z)\} \rightarrow \min, \quad z \in D. \quad (1.3)$$

### 1.3. Обобщения понятия решения в задачах векторной оптимизации

Обобщим введенное в пункте 1.1 понятие глобально-оптимального решения на ситуацию, когда нужно оптимизировать не одну функцию, а сразу несколько

$$Q(z) = \{Q_1(z), Q_2(z), \dots, Q_n(z)\}.$$

*Многокритериальные задачи* вида (1.3) называют также задачами *векторной оптимизации*. В этом случае задача обычно становится противоречивой в том смысле, что минимум различных компонент векторной функции  $Q$  может достигаться на разных элементах  $z \in D$  (рис. 1.2).

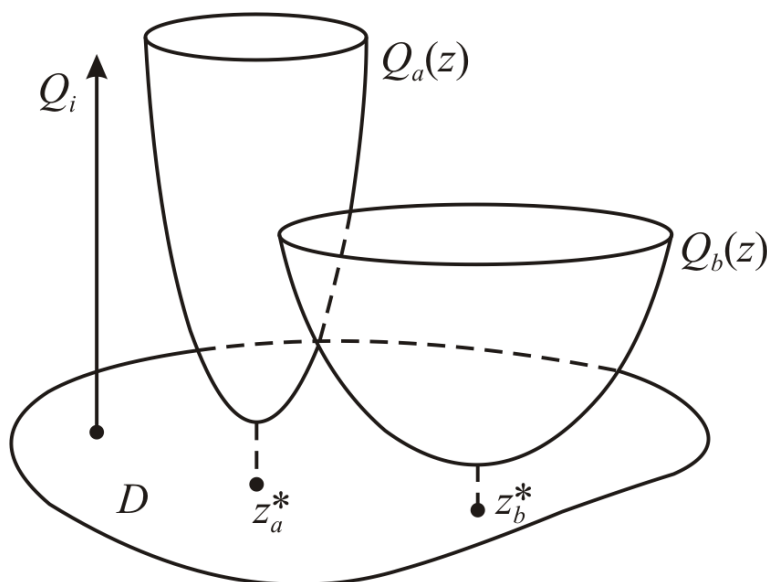


Рис. 1.2. Противоречивость задачи векторной оптимизации с  $Q(z) = (Q_a(z), Q_b(z))$

Очевидно, что требуется ввести новое понятие оптимальности. Пусть  $W$  — образ множества  $D$  при векторном отображении  $Q : D \rightarrow W$ , и элемент  $\tilde{z}$  отображается в  $\tilde{w} = Q(\tilde{z})$ . В каком случае  $\tilde{w}$  не может быть принят в качестве оптимального значения векторной функции  $Q$ ?

Рассмотрим множество  $R^-(\tilde{w})$  векторов  $w$ , покомпонентно строго лучших вектора  $\tilde{w}$ :

$$R^-(\tilde{w}) = \{w \in R^n : \forall i = \overline{1, n}, w_i < \tilde{w}_i\}.$$

Интуитивно понятно, что  $\tilde{w}$  из  $W$  не может быть признан оптимальным, если пересечение  $R^-(\tilde{w}) \cap W \neq \emptyset$ , т.е. если  $\exists z \in D$ , что  $Q(z) \in R^-(\tilde{w})$  (см. рис. 1.3). Таким образом, для оптимальных  $\tilde{w}$  пересечение должно быть пусто. Геометрически это означает, что открытое множество  $R^-(\tilde{w})$ , имеющее на рис. 1.3 вид угла с вершиной в точке  $\tilde{w}$ , не должно иметь общих точек с множеством  $W$ , чтобы значение  $\tilde{w}$  могло быть признано оптимальным. При этом значение  $\tilde{w}$  нельзя улучшить одновременно по всем компонентам векторного критерия  $Q(z)$  за счет выбора  $z$  из  $D$ .

Такие значения на рис. 1.3 образуют участки границы множества  $W$ , отмеченные двойной или жирной линией. Их называют *слабо эффективными (оптимальными по Слейтеру)*. Подмножество значений, отмеченных жирной линией соответствует *эффективным (оптимальным по Парето)* решениям. Дадим формальные определения.

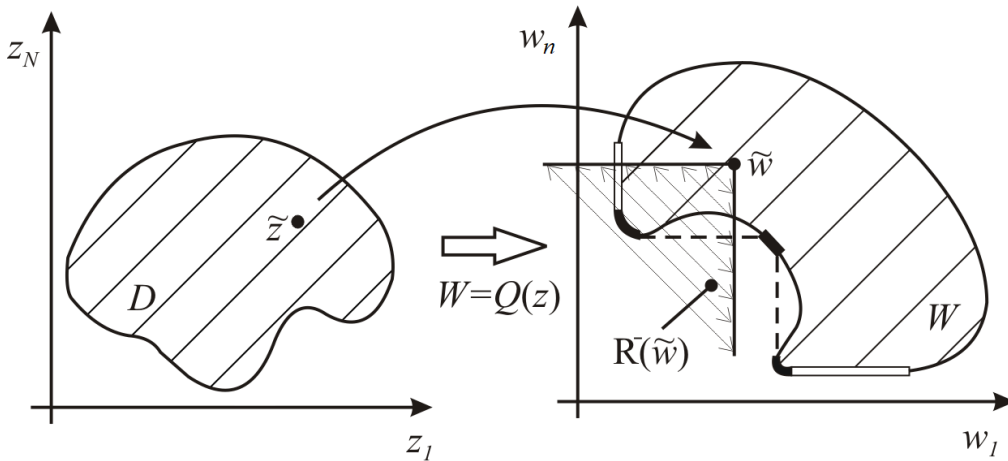


Рис. 1.3. Оптимальность в многокритериальных задачах:  
 — слабоэффективные значения;  — эффективные значения

**Определение.** Элемент  $z^o \in D$  называется *полуэффективным* или *оптимальным по Слейтеру*, если не существует другого  $z \in D$ , что  $\forall i = 1, \dots, n : Q_i(z) < Q_i(z^o)$ .

Множество таких элементов обозначим  $Z^o$ .

**Определение.** Элемент  $z^* \in D$  называется *эффективным* или *оптимальным по Парето*, если не существует другого  $z \in D$ , что  $\forall i = 1, \dots, n : Q_i(z) \leq Q_i(z^*)$  и хотя бы одно неравенство строгое, т.е.  $Q(z) \neq Q(z^*)$ .

Множество таких элементов обозначим  $Z^*$  (см. рис. 1.3).

**Замечание.** *Образы множеств элементов, оптимальных по Слейтеру (Парето) при отображении  $Q$ , называют множествами значений оптимальных по Слейтеру (Парето). Операторы выделения этих множеств будем обозначать как  $S(W)$  и  $P(W)$ . Таким образом,*

$$Q(Z^o) = S(W), \quad Q(Z^*) = P(W).$$

Для определения множеств решений  $Z^o$  или  $Z^*$  могут быть использованы различные подходы. Наиболее употребительны различные способы сведения задачи векторной оптимизации к серии вспомогательных задач

$$\Psi_\lambda(z) \rightarrow \min, \quad z \in D \quad (1.4)$$

со скалярным критерием  $\Psi_\lambda(z)$ , зависящим от набора параметров  $\lambda$ . Вид критерия  $\Psi_\lambda$  выбирают так, чтобы при любых значениях  $\lambda \in \Lambda$  решение  $z_\lambda^*$  задачи (1.4) принадлежало  $Z^o$  или  $Z^*$ . Подходы, основанные на решении задач вида (1.4) называют *методами сверток*.

Существуют методы построения оценок множеств  $Z^o$  и  $Z^*$  без использования параметрических сверток.

Более подробно методы решения многокритериальных задач, в частности, методы сверток рассмотрены в книге [2] (разделы 1.1-1.3), а в [3] изложены основные понятия и теория условий оптимальности по Парето и Слейтеру. Расширением материала данного пособия является теория, представленная в описании лабораторного практикума [4], посвященного вопросам поиска оптимальных путей на графах с векторными весами ребер. В частности, в данном лабораторном практикуме рассмотрены методы линейной свертки, свертки Ю.В. Гермейера в сочетании с применением метода рекуррентных уравнений Р. Беллмана, а также метода Дейкстры. Кроме того, изложен метод поиска множества оптимальных по Парето путей на графах с векторными весами ребер, не использующий свертку.

Этот материал разбирается студентами самостоятельно, отрабатывается на занятиях лабораторного практикума и включается в программу экзамена по соответствующей дисциплине.

# Глава 2. Дискретное динамическое программирование

## 2.1. Задачи с фиксированным временем начала и окончания процесса

Пусть задана *управляемая динамическая система*, состояние которой в момент времени  $t$  определяется вектором  $x(t) \in R^n$ . Пусть  $u(t)$  — управление, которое действует на систему в момент  $t$ . Время изменяется дискретно:

$$t \in \Theta = \{t_s, t_s + 1, \dots, t_f - 1, t_f\},$$

где  $t_s$  — время старта,  $t_f$  — время финиша. Будем считать, что  $t_f - t_s = N$ , т.е. процесс  $N$ -шаговый.

Понятие управляемой динамической системы предполагает возможность однозначного определения значения состояния  $x(t+1)$  в следующий момент времени по известным текущим значениям  $x(t)$  и  $u(t)$  в момент времени  $t$ .

Обозначим оператор динамической системы, описывающий закон изменения ее состояния как

$$x(t+1) = f(x(t), u(t), t). \quad (2.1)$$

Для текущего состояния  $x(t)$  управления  $u(t)$  выбираются из допустимого множества, которое может зависеть только от текущего состояния  $x(t)$  и момента времени:

$$u(t) \in U(x(t), t) \in R^m. \quad (2.2)$$

Кроме того задано множество начальных состояний

$$x(t_s) = x^s \in X^s \quad (2.3)$$

и множество конечных состояний

$$x(t_f) = x^f \in X^f. \quad (2.4)$$

Принимается, что переход динамической системы из состояния  $x(t)$  под действием управления  $u(t)$  связан с затратами, которые определяются *функцией затрат* вида  $q(x(t), u(t), t)$  (см. рис. 2.1).

Необходимо найти допустимое (с точки зрения ограничений (2.2)–(2.4)) оптимальное управление  $u^*[t_s, t_f - 1]$  и связанную с ним оптимальную траекторию  $x^*[t_s, t_f]$ , обеспечивающие с учетом всех связей и ограничений оптимальное значение целевой функции  $Q$ :

$$Q^* = Q(x^*[t_s, t_f], u^*[t_s, t_f - 1]) = \min_{\substack{x[t_s, t_f]; \\ u[t_s, t_f - 1]}} Q(x[t_s, t_f], u[t_s, t_f - 1]),$$

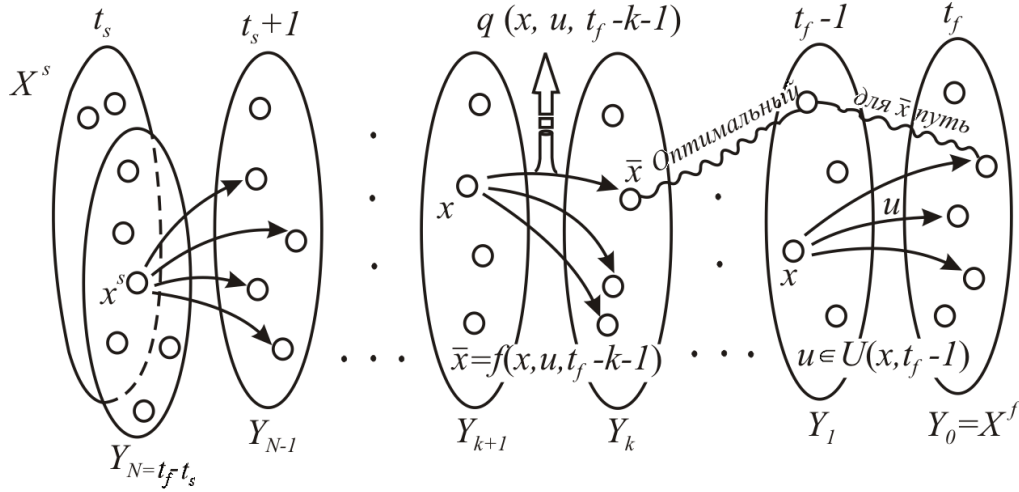


Рис. 2.1. Управляемый динамический процесс с функцией перехода  $f$  и затратами  $q$  на шаге

где функция  $Q$  определяется некоторым образом через значения затрат  $q(x(t), u(t), t)$  на шагах в зависимости от типа задачи.

1. *Задачи с аддитивным критерием:*

$$Q(x[t_s, t_f], u[t_s, t_f - 1]) = \sum_{t=t_s}^{t_f-1} q(x(t), u(t), t). \quad (2.5)$$

2. *Задачи с критерием типа максимума:*

$$Q(x[t_s, t_f], u[t_s, t_f - 1]) = \max_{t \in [t_s, t_f-1]} q(x(t), u(t), t).$$

3. *Задачи с мультипликативным критерием:*

$$Q(x[t_s, t_f], u[t_s, t_f - 1]) = \prod_{t=t_s}^{t_f-1} q(x(t), u(t), t)$$

(при дополнительном условии положительности  $q(x, u, t) > 0$ ).

Заметим, что в контексте применяемых обозначений на указанных интервалах времени, например, вида  $[t_s, t_f]$  используются только допустимые дискретные значения из множества  $\Theta$ .

Далее изложение будет проведено на примере критерия (2.5). Задачи с критериями других типов рассматриваются на практических и лабораторных занятиях (см. [4], [10]).

Сделаем еще одно важное замечание. При решении методами динамического программирования содержательных прикладных задач (см., например, [7], [8]) часто приходится сталкиваться с ситуацией, когда исходная постановка задачи внешне сильно отличается от (2.1) – (2.5). Трудности связаны с

выбором такого описания  $\varkappa(t)$ , которое могло бы являться состоянием  $x(t)$ . При этом следует помнить, что всегда можно проверить правильно ли введено понятие состояния  $x(t)$ . Критерий проверки следующий: должна быть возможна запись в форме (2.1), выбор  $u(t)$  должен подчиняться ограничениям вида (2.2), а функция затрат на шаге должна зависеть только от  $x(t)$ ,  $u(t)$ ,  $t$ .

## 2.2. Метод рекуррентных соотношений Р. Беллмана

Метод рекуррентных соотношений Ричарда Беллмана [6], [7], [8] состоит в том, что исходная  $N$ -шаговая задача высокой размерности погружается в семейство более простых рекурсивно связанных друг с другом задач более низкой размерности. Оптимальное управление строится поэтапно. На каждом этапе строятся оптимальные правила выбора управления для укороченных задач, включающих только один последний шаг, два шага, три шага и т.д. Вместе с тем на каждом шаге управление выбирается с учетом последствий, так как управление, оптимизирующее целевую функцию только для каждого данного шага, может привести к неоптимальному результату всего процесса.

Управление на каждом шаге должно быть оптимальным с точки зрения процесса в целом от текущего момента времени до конца.

Каково бы ни было начальное состояние системы перед очередным шагом, управление на этом этапе выбирается так, чтобы затраты на данном шаге плюс оптимальные затраты на всех последующих шагах были минимальными.

Обозначим через  $Y_k$  множество состояний, из которых ровно за  $k$  шагов динамическая система может попасть в множество конечных состояний  $X^f$ . При этом  $Y_0 = X^f$ .

**Определение.** Пусть система находится в состоянии  $x \in Y_k$ . *Функцией Беллмана*  $S_k(x)$  называют функцию оптимальных (!) затрат при переводе динамической системы из  $x$  в множество финальных состояний за  $k$  шагов:

$$S_k(x) = \min \sum_{t=t_f-k}^{t_f-1} q(x(t), u(t), t),$$

где минимум берется при условии, что  $x(t_f - k) = x$ ,  $x(t+1) = f(x(t), u(t), t)$ ,  $u(t) \in U(x(t), t)$  для  $t \in \{t_f - k, \dots, t_f - 1\}$ , и  $x(t_f) \in X^f$ .

Рассмотрим *прямой ход* вычислительного процесса. Его идея состоит в том, что процесс вычислений, двигаясь от множества конечных состояний к множеству начальных состояний, на каждом этапе для текущего состояния  $x \in Y_k$  выбирает такое управление  $u^*(x, t_f - k)$ , под действием которого затраты на текущем шаге вместе с оптимально возможными затратами на всех



последующих шагах оказались бы минимальными. Рассмотрим этот процесс более детально.

1. Пусть  $x \in Y_1$ , тогда осталось выполнить один последний шаг. Находим функцию Беллмана, равную оптимальным затратам при переходе из  $x$  в  $X^f$  (см. правую часть рис. 2.1)

$$S_1(x) = \min_{\substack{u \in U(x, t_f - 1) \\ f(x, u, t_f - 1) \in X^f}} \{q(x, u, t_f - 1)\}. \quad (2.6)$$

Пусть для состояния  $x$  минимум в (2.6) достигается при использовании управления  $u^*(x, t_f - 1)$ .

2. Допустим, что  $S_1(x), S_2(x), \dots, S_k(x)$ , а также  $u^*(x, t_f - 1), \dots, u^*(x, t_f - k)$  уже найдены.

3. Выведем рекуррентное соотношение, связывающее  $S_{k+1}(x)$  и  $S_k(x)$ . Предположим, что  $x \in Y_{k+1}$ . Найдем величину минимальных затрат при переходе из данного состояния  $x$  в финальное множество  $X^f$  (см. среднюю часть рис. 2.1).

По определению:

$$S_{k+1}(x) = \min_{\substack{u \in U(x, t_f - k - 1) \\ x(t_f - k) = f(x, u, t_f - k - 1) \\ x(t+1) = f(x(t), u(t), t) \\ u(t) \in U(x(t), t) \\ t \in \{t_f - k, \dots, t_f - 1\} \\ x(t_f) \in X^f}} \left\{ q(x, u, t_f - k - 1) + \sum_{t=t_f - k}^{t_f - 1} q(x(t), u(t), t) \right\}.$$

Поскольку первый элемент  $q(\dots)$  в сумме, стоящей в фигурных скобках, не зависит от группы переменных  $x(t), u(t)$  для  $t \geq t_f - k$ , то минимум по обеим группам переменных можно брать только от оставшейся суммы.

$$S_{k+1}(x) = \min_{u \in U(x, t_f - k - 1)} \left\{ q(x, u, t_f - k - 1) + \right. \\ \left. + \min_{\substack{x(t_f - k) = f(x, u, t_f - k - 1) \\ x(t+1) = f(x(t), u(t), t) \\ u(t) \in U(x(t), t) \\ t \in \{t_f - k, \dots, t_f - 1\}; x(t_f) \in X^f}} \sum_{t=t_f - k}^{t_f - 1} q(x(t), u(t), t) \right\}. \quad (2.7)$$

Формальным обоснованием законности этого перехода является доказанная ниже лемма о расщеплении операции взятия инфинума. Используя определение функции Беллмана, из (2.7) получим, что для  $x \in Y_{k+1}$ :

$$S_{k+1}(x) = \min_{\substack{u \in U(x, t_f - k - 1) \\ f(x, u, t_f - k - 1) \in Y_k}} \{q(x, u, t_f - k - 1) + S_k(f(x, u, t_f - k - 1))\}. \quad (2.8)$$

Из этого уравнения для состояния  $x$  определяется, кроме того, значение на данном шаге оптимального управления:  $u^*(x, t_f - k - 1)$ . Заметим, что дополнительное ограничение  $f(x, u, t_f - k - 1) \in Y_k$  появилось в (2.8) в силу того, что в нем использована функция  $S_k(\bar{x})$ , областью определения которой является множество  $Y_k$ .

4. Таким образом, продолжая процесс до  $x \in Y_N$  найдем функцию  $S_N(x)$ .

Уравнения (2.6), (2.8) описывают последовательность *рекуррентных соотношений* для определения функций Беллмана (т.е. функций оптимальных затрат) для семейства укороченных задач, а также оптимальных правил поведения  $u^*(x, t)$ . Используя найденные функции, а также оптимальное начальное состояние  $x_s^*$  из  $X^s$ , можно найти решение рассматриваемой задачи динамического программирования.

Оптимальное начальное состояние  $x_s^* \in X^s$ , являющееся началом оптимального пути до  $X^f$  определим из решения задачи:

$$x_s^* = \arg \min_{x \in Y_N \cap X^s} S_N(x). \quad (2.9)$$

Итак, из (2.6), (2.8) находим оптимальные правила поведения, т.е. управления  $u^*(x, t_f - k - 1)$ , под действием которых динамическая система должна переходить из состояния  $x \in Y_{k+1}$  в  $Y_k$  ( $k = 0, N - 1$ ), а также оптимальное начальное состояние  $x_s^* \in Y_N$ . Найденное описание оптимального поведения в форме  $u^*(x, t)$  можно трактовать как *оптимальный регулятор*, действующий по принципу *обратной связи* (см. рис. 2.2).

Построим теперь оптимальное управление как функцию времени  $u^*(t)$ . Для этого выполним *обратный ход* вычислений метода Беллмана. При этом найдем в явной форме оптимальную программу управления  $u^*(t)$  и оптимальную траекторию  $x^*(t)$  из  $x_s^*$  в  $X^f$ .

Для состояния  $x^*(t_s) = x_s^*$  выбираем оптимальное управление  $u^*(t_s) = u^*(x^*(t_s), t_s)$ , используя оптимальное правило, которое найдено на прямом ходе метода. Под действием этого управления переходим в состояние

$$x^*(t_s + 1) = f(x^*(t_s), u^*(t_s), t_s).$$

Теперь, определим  $u^*(t_s + 1) = u^*(x^*(t_s + 1), t_s + 1)$  и находим следующую

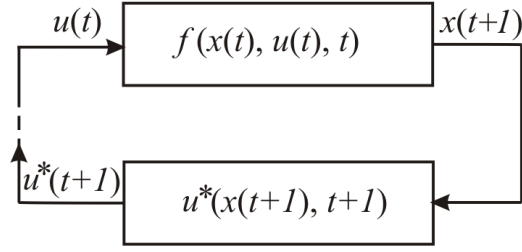


Рис. 2.2. Оптимальные правила  $u^*(x, t)$  определяют оптимальный регулятор

точку пути —  $x^*(t_s + 2)$ . Продолжая этот процесс до момента времени  $t_f$ , получим оптимальный набор состояний, т.е. оптимальную траекторию

$$x^*(t), \quad t \in \{t_s, \dots, t_f\}$$

и оптимальные управления

$$u^*(t), \quad t \in \{t_s, \dots, t_f - 1\}.$$

Для строгого обоснования проведенных выкладок при выводе рекуррентных уравнений Р.Беллмана докажем возможность расщепления в операции взятия минимума, которая использовалась при получении формулы (2.8).

**Лемма** (о расщеплении операции взятия инфимума). *Всегда верно следующее соотношение*

$$\inf_{u \in U, w \in W(u)} \{V_1(u) + V_2(u, w)\} = \inf_{u \in U} \left\{ V_1(u) + \inf_{w \in W(u)} \{V_2(u, w)\} \right\}.$$

*Доказательство.* Обозначим значение правой части равенства через  $R$ , а левой части — через  $L$ . Пусть также  $C(u) = \inf_{w \in W(u)} \{V_2(u, w)\}$ .

Пусть пара  $(u_\varepsilon^*; w_\varepsilon^*)$  дает  $\varepsilon$ -приближение левой фигурной скобки к  $L$ , а  $w_\varepsilon^*(u)$  дает  $\varepsilon$ -приближение к  $C(u)$ . Тогда верны неравенства:

$$\varepsilon + L \geq V_1(u_\varepsilon^*) + V_2(u_\varepsilon^*, w_\varepsilon^*) \geq V_1(u_\varepsilon^*) + \inf_{w \in W(u_\varepsilon^*)} V_2(u_\varepsilon^*, w) \geq R.$$

Таким образом, для  $\forall \varepsilon > 0 : \varepsilon + L \geq R$ . В пределе при  $\varepsilon \rightarrow 0 : L \geq R$ .

С другой стороны, для  $u \in U$ :

$$L \leq V_1(u) + V_2(u, w_\varepsilon^*(u)) \leq V_1(u) + \inf_{w \in W(u)} V_2(u, w) + \varepsilon.$$

Устремляя  $\varepsilon$  к нулю и беря от крайних частей полученного неравенства точную нижнюю грань по  $u \in U$ , получим:

$$L \leq \inf_{u \in U} \left\{ V_1(u) + \inf_{w \in W(u)} V_2(u, w) \right\} + \varepsilon = R + \varepsilon.$$

Таким образом, при  $\varepsilon \rightarrow 0 L \leq R$ . Откуда следует, что  $L = R$ . □

Важно уяснить роль рекуррентных уравнений Беллмана в задачах динамического программирования.

Исходная задача (2.1)–(2.5) имеет размерность  $n \cdot N + m \cdot N$ , где  $N$  обычно велико. С использованием уравнений Беллмана они сводятся к решению  $N$  задач размерности  $m$  и одной задачи размерности  $n$ , хотя первые  $N$  задач приходится решать многократно для различных значений аргументов  $x$  функций Беллмана  $S_k(x)$ . Однако, общие затраты при решении этим методом будут линейно зависеть от  $N$ , тогда как при прямом решении исходной задачи зависимость от  $N$ , в общем случае, экспоненциальная.

Рассмотрим пример задачи, в которой применение метода рекуррентных уравнений, позволяет получить аналитическое решение.

**Пример** (задача о распределении сырья). Пусть требуется распределить (не обязательно полностью) начальный запас сырья  $a$  между  $N$  предприятиями, каждое из которых приносит доход  $q_i(u_i) = \alpha_i \sqrt{u_i}$ , где  $u_i$  — количество сырья, которое выделяется  $i$ -ому предприятию. Целью распределения является получение максимального суммарного дохода

$$\sum_{k=1}^N \alpha_k \sqrt{u_k}.$$

Заметим, что пропорциональность дохода не  $u_k$ , а корню из  $u_k$  можно объяснить, например, тем, что сырье является скоропортящимся, поэтому скорость увеличения дохода с ростом  $u_k$  должна уменьшаться (см. рис. 2.3).

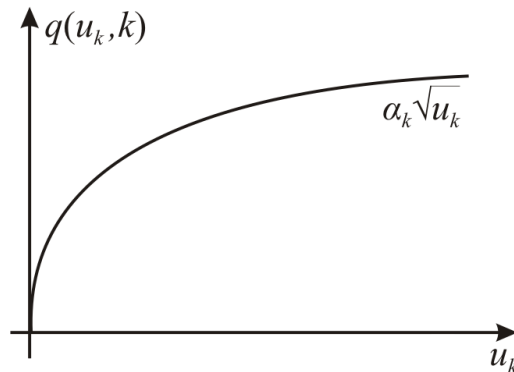


Рис. 2.3. Вид функции дохода от переработки скоропортящегося сырья

Чтобы представить задачу в форме задачи динамического программирования, будем рассматривать распределение сырья как процесс последовательного дележа. Будем выделять сырье вначале 1-му предприятию, затем 2-му и т.д. Примем, что момент времени  $t$  совпадает с номером предприятия, т.е.  $t \in \{1, 2, \dots, N\}$ . Объемы сырья, выделяемые каждому предприятию,

$$u_1, u_2, \dots, u_N, (u_i \geq 0)$$

примем за управления, а в качестве состояния  $x_k$  выберем количество сырья, оставшееся после распределения  $k$  предприятиям. Тогда начальное состояние процесса  $x_0 = a$ , а закон изменения состояния процесса  $x_{k+1} = f(x_k, u_{k+1}) = x_k - u_{k+1}$ . Следовательно,  $u_{k+1} \in [0, x_k]$ . Таким образом,  $U(x_k, k) = [0, x_k]$ . Кроме того,  $X_s = \{a\}$ , а  $X_f = [0, a]$ .

Составим уравнения Беллмана. Пусть сырье распределено по всем предприятиям, кроме  $N$ -ого, и управляемый процесс находится в состоянии  $x_{N-1}$ , тогда максимальный доход (за счет выбора  $u_N$ ) в оставшемся одношаговом процессе составит

$$S_1(x_{N-1}) = \max_{0 \leq u_N \leq x_{N-1}} \alpha_N \sqrt{u_N} = \beta_N \sqrt{x_{N-1}},$$

где принято  $\beta_N = \alpha_N$ ;  $u_N^*(x_{N-1}) = x_{N-1}$ .

Пусть теперь сырье распределено по всем предприятиям, кроме двух последних:  $N-1$ -го и  $N$ -го, тогда

$$\begin{aligned} S_2(x_{N-2}) &= \max_{0 \leq u_{N-1} \leq x_{N-2}} \{\alpha_{N-1} \sqrt{u_{N-1}} + S_1(x_{N-2} - u_{N-1})\} = \\ &= \max_{0 \leq u_{N-1} \leq x_{N-2}} \{\alpha_{N-1} \sqrt{u_{N-1}} + \beta_N \sqrt{x_{N-2} - u_{N-1}}\}. \end{aligned}$$

Значение  $u_{N-1}^*(x)$ , определяющее максимум, ищем из условия обращения в ноль производной функции

$$\begin{aligned} \rho(x, u) &= \alpha \sqrt{u} + \beta \sqrt{x - u}. \\ \rho'(x, u) &= \frac{1}{2} \left( \frac{\alpha}{\sqrt{u}} - \frac{\beta}{\sqrt{x - u}} \right) = 0, \\ \alpha \sqrt{x - u} - \beta \sqrt{u} &= 0. \end{aligned}$$

Отсюда

$$u = \frac{\alpha^2 x}{\alpha^2 + \beta^2},$$

лежит между 0 и  $x$ . Так как  $\rho'(x, 0) = +\infty$ ,  $\rho'(x, x) = -\infty$ , то значение

$$u_{N-1}^*(x_{N-2}) = \frac{\alpha_{N-1}^2 x_{N-2}}{\alpha_{N-1}^2 + \beta_N^2}$$

обеспечивает именно максимум выражения в фигурной скобке. Подставив это значение, получим:

$$S_2(x_{N-2}) = \sqrt{\alpha_{N-1}^2 + \beta_N^2} \sqrt{x_{N-2}} = \beta_{N-1} \sqrt{x_{N-2}},$$

где принято  $\beta_{N-1}^2 = \alpha_{N-1}^2 + \beta_N^2$ .

Методом математической индукции нетрудно доказать, что на любом шаге  $S_k(x_{N-k}) = \beta_{N-k+1}\sqrt{x_{N-k}}$ , где  $\beta_{N-k+1}^2 = \alpha_{N-k+1}^2 + \beta_{N-k+2}^2$ , и  $u_{N-k+1}^*(x_{N-k}) = \frac{\alpha_{N-k+1}^2 x_{N-k}}{\beta_{N-k+1}^2}$ . Продолжая этот процесс, придем к начальному состоянию, когда сырье не распределено между предприятиями. Тогда общий максимальный доход определится как значение  $S_N(a) = \beta_1^2 \sqrt{a}$ , где  $\beta_1^2 = \sum_{s=1}^N \alpha_s^2$ , а начальное значение оптимального управления составит:

$$u_1^* = u_1^*(a) = \frac{\alpha_1^2}{\alpha_1^2 + \beta_2^2} a = \frac{\alpha_1^2}{\beta_1^2} a.$$

Найдем состояние  $x_1^*$ :

$$x_1^* = a - \frac{\alpha_1^2}{\beta_1^2} a = \frac{\beta_2^2}{\beta_1^2} a.$$

Отсюда найдем значение оптимального управления во 2-й момент времени:

$$u_2^* = u_2^*(x_1^*) = \frac{\alpha_2^2 \beta_2^2}{\beta_2^2 \beta_1^2} a = \frac{\alpha_2^2}{\beta_1^2} a.$$

Продолжая этот процесс, получим, что каждому предприятию оптимально выделить сырье в размере:

$$u_k^* = \frac{\alpha_k^2}{\sum_{s=1}^N \alpha_s^2} a.$$

Другие примеры решения задач динамического программирования можно найти в [7], [8], [10].

## 2.3. Принцип Беллмана

Рассмотрим принцип Беллмана применительно к задачам динамического программирования с дискретным временем.

**Утверждение** (принцип Беллмана в форме необходимого условия). *Оптимальное управление обладает тем свойством, что если под его воздействием в некоторый промежуточный момент времени управляемая система попала в состояние  $x^*(\tau)$ , то оставшиеся такты этого управления будут оптимальны на оставшемся промежутке времени по отношению к состоянию  $x^*(\tau)$ , в которое попала управляемая система.*

*Доказательство.* (Для аддитивного критерия (2.5)).

Покажем, что выбор любого другого управления  $v[\tau, t_f]$  вместо  $u^*[\tau, t_f]$  не может привести к уменьшению затрат на второй части пути, если процесс находился в начальном состоянии  $x^*(\tau)$ .

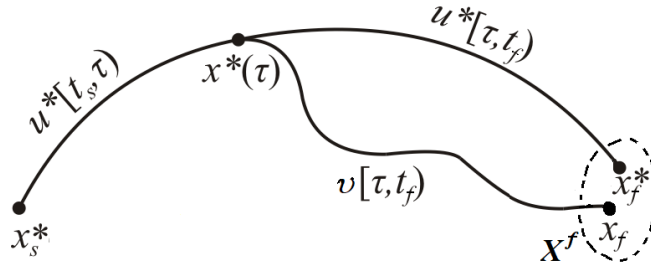


Рис. 2.4. К доказательству принципа Беллмана как необходимого условия

Доказательство проведем методом от противного (см. рис. 2.4). Пусть существует такое управление  $v[\tau, t_f]$ , что

$$Q(x^*(\tau), v[\tau, t_f]) < Q(x^*(\tau), u^*[\tau, t_f]).$$

Рассмотрим следующее управление:

$$\bar{u}(t) = \begin{cases} u^*(t), & t \in [t_s, \tau - 1] \\ v(t), & t \in [\tau, t_f - 1] \end{cases}.$$

Из аддитивности минимизируемой функции следует, что

$$\begin{aligned} Q(x^*(t_s), \bar{u}[t_s, t_f - 1]) &= Q(x^*(t_s), u^*[t_s, \tau - 1]) + Q(x^*(\tau), v[\tau, t_f - 1]) < \\ &< Q(x^*(t_s), u^*[t_s, \tau - 1]) + Q(x^*(\tau), u^*[\tau, t_s - 1]) = Q(x^*(t_s), u^*[t_s, t_f - 1]). \end{aligned}$$

Пришли к противоречию, так как  $u^*$  — оптимальное управление.  $\square$

**Задание.** Покажите, что рассмотренное выше утверждение неверно для критерия типа максимума.

Заметим, что для построения эффективных вычислительных методов важно выполнение принципа Беллмана в форме достаточного условия [11]. Ряд методов, отличных от метода рекуррентных уравнений, явно или неявно опирается на принцип Беллмана.

**Утверждение** (принцип Беллмана в форме достаточного условия). *Если в каждом текущем состоянии использовать управление, являющееся первым тактом управления, оптимального на оставшемся промежутке движения по отношению к состоянию, в которое попал динамический процесс, то такое управление будет в целом оптимально.*

**Определение.** Если для управления  $u_I[t_s, \tau - 1]$  существует его продолжение  $u_{II}[\tau, t_f - 1]$  такое, что управление  $u_I + u_{II}$  будет оптимально в целом, то управление  $u_I[t_s, \tau - 1]$  назовем *оптимальным с учетом последствий*.

*Доказательство.* Пусть  $u^*(x(\tau), [\tau, t_f - 1])$  является оптимальным управлением на промежутке  $[\tau, t_f)$  по отношению к состоянию  $x(\tau)$ ,  $u^*(x(\tau), \tau)$  — первым тактом этого управления, т.е. значением этого управления в момент  $\tau$ , а  $u_{II}^*[\tau + 1, t_f - 1]$  — оставшимися тактами этого управления на оставшемся промежутке времени.

Покажем, что если начиная с  $x^*(t_s)$  использовать значения управления  $u(\tau) = u^*(x(\tau), \tau)$ , где  $x(\tau)$  — текущее состояние, в котором при этом оказалась управляемая система, то полученное управление будет оптимальным в целом.

Докажем по индукции:

1. На первом шаге очевидно, что управление  $u^*(x^*(t_s), t_s)$  оптимально с учетом последствия, так как это первый такт оптимального в целом управления.
2. Предположим (см. рис. 2.5(a)), что на первых  $k - 1$  шагах указанное правило дает управление  $u_I[t_s, \tau - 1]$ , оптимальное с учетом последствия. То есть существует продолжение  $u_{II}[\tau, t_f - 1]$  (с затратами  $Q_{II}$ ), что  $Q_I + Q_{II} = Q^*$ , где  $Q_I$  — затраты, порожденные управлением  $u_I[t_s, \tau - 1]$ , а  $Q^*$  — оптимальные затраты. Обозначим состояние, в которое приводит управление  $u_I[t_s, \tau - 1]$  через  $x_I(\tau)$ .

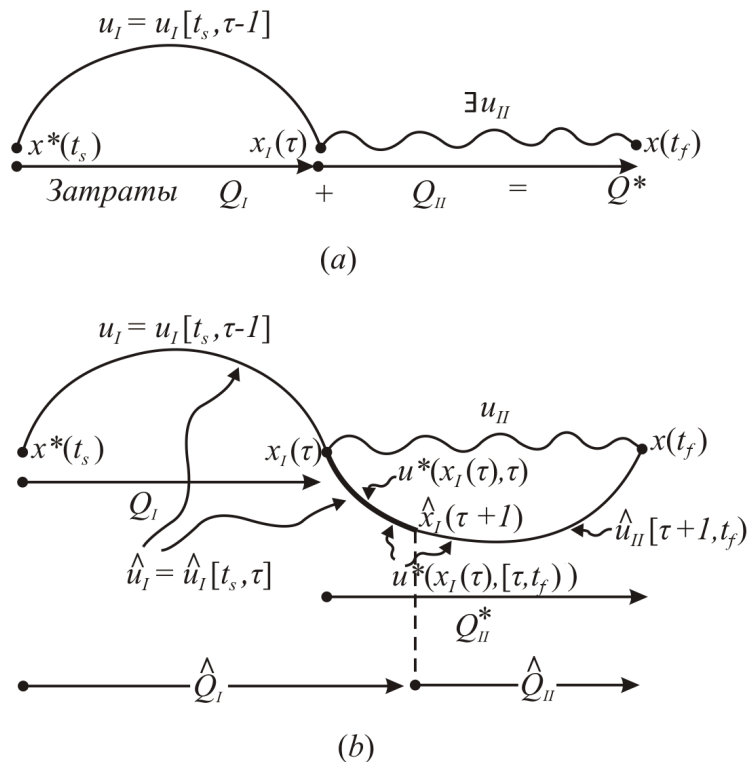


Рис. 2.5. К доказательству принципа Беллмана как достаточного условия

3. Докажем, что это свойство выполняется и для  $k$ -го шага. Обозначим управление, построенное по принципу Беллмана за  $k$  шагов через  $\hat{u}_I[t_s, \tau]$ .



- а) Из способа построения управления  $\hat{u}_I$  следует, что оно включает  $u_I[t_s, \tau-1]$  и еще первый такт  $u^*(x_I(\tau), \tau)$  управления  $u^*(x_I(\tau), [\tau, t_f])$ , оптимального для состояния  $x_I(\tau)$  на  $[\tau, t_f - 1]$  (см. рис. 2.5(b)).

Обозначим затраты, связанные с этим управлением, через  $Q_{II}^*$ . Очевидно, что  $Q_{II} \geq Q_{II}^*$ . Но управление  $(u_I[t_s, \tau-1]; u^*(x_I(\tau), [\tau, t_f]))$  не обязано быть оптимальным, поэтому  $Q_I + Q_{II}^* \geq Q^*$ . Следовательно

$$Q^* = Q_I + Q_{II} \geq Q_I + Q_{II}^* \geq Q^*.$$

Откуда следует, что  $Q_{II} = Q_{II}^*$ .

- б) согласно пункту (а):  $\hat{u}_I = \hat{u}_I[t_s, \tau] = (u_I[t_s, \tau-1]; u^*(x_I(\tau), \tau))$ . Выберем в качестве продолжения этого управления оставшиеся (после исключения  $u^*(x_I(\tau), \tau)$ ) такты управления  $u^*(x_I(\tau), [\tau, t_f])$ . Обозначим их  $\hat{u}_{II}$ . Соответствующие  $\hat{u}_I$  и  $\hat{u}_{II}$  затраты обозначим через  $\hat{Q}_I$  и  $\hat{Q}_{II}$  соответственно. Затраты при использовании управления  $(\hat{u}_I, \hat{u}_{II})$ :

$$\begin{aligned} \hat{Q}_I + \hat{Q}_{II} &= \underbrace{Q_I + q(x_I(\tau), u^*(x_I(\tau), \tau), \tau)}_{\hat{Q}_I} + \\ &+ \underbrace{Q_{II}^* - q(x_I(\tau), u^*(x_I(\tau), \tau), \tau)}_{\hat{Q}_{II}} = Q_I + Q_{II}^* = Q_I + Q_{II} = Q^*. \end{aligned}$$

Следовательно,  $\hat{u}_I[t_s, \tau]$  является оптимальным с учетом последействия. По индукции получаем оптимальность управления в целом.  $\square$

## 2.4. Запись рекуррентных соотношений Беллмана относительно начала процесса

Опишем без подробного вывода запись уравнений Беллмана от начала процесса (см. рис. 2.6).

Обозначим через  $X_k$  множество состояний, в которое можно попасть из  $X^s$  за  $k$  шагов.

**Определение.** *Функцией Беллмана  $Z_k(x)$  относительно начала процесса* называют функцию оптимальных затрат по переходу за  $k$  шагов из множества начальных состояний  $X^s$  в  $x \in X_k$ .

Пусть  $x \in X_1$ . Тогда, по определению (см. левую часть рис. 2.6):

$$Z_1(x) = \min_{\substack{\bar{x} \in X^s \\ u \in U(\bar{x}, t_s) \\ f(\bar{x}, u, t_s) = x}} q(\bar{x}, u, t_s). \quad (2.10)$$

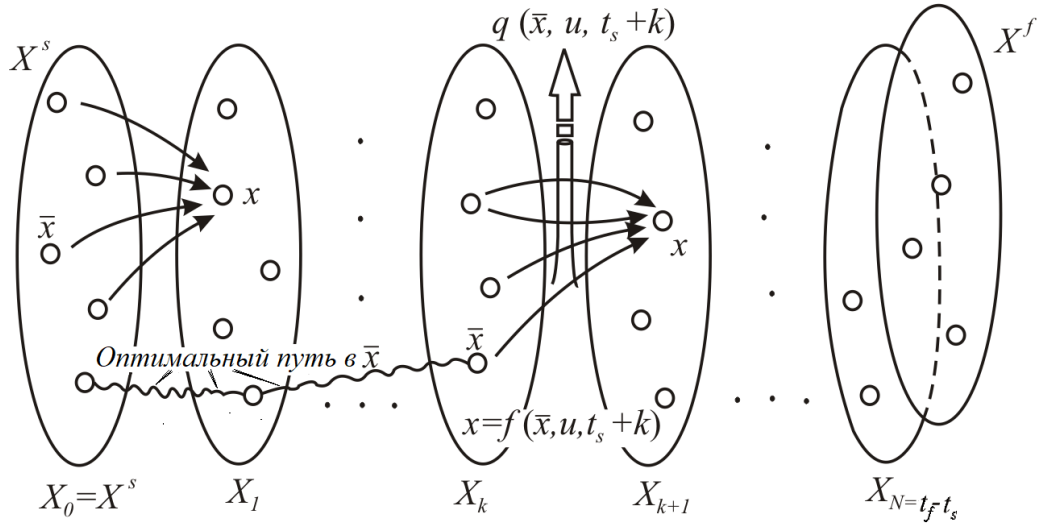


Рис. 2.6. Иллюстрация к записи уравнений Беллмана от начала процесса

Таким образом, в общем случае минимум берется по совокупности переменных  $\bar{x}$  и  $u$ . Их значения, обеспечивающие минимум в (2.10) обозначим через  $\bar{x}^*(x, t_s)$  и  $\bar{u}^*(x, t_s)$ . Пусть теперь  $x \in X_{k+1}$ . Функция Беллмана запишется в виде (см. центральную часть рис. 2.6):

$$Z_{k+1}(x) = \min_{\substack{\bar{x} \in X_k \\ u \in U(\bar{x}, t_s + k) \\ f(\bar{x}, u, t_s + k) = x}} \{Z_k(\bar{x}) + q(\bar{x}, u, t_s + k)\}. \quad (2.11)$$

Значения  $\bar{x}$  и  $u$ , обеспечивающие минимум в (2.11), обозначим как

$$\begin{aligned} \bar{x}^*(x, t_s + k), \\ \bar{u}^*(x, t_s + k). \end{aligned}$$

Эти пары определяют оптимальные правила поведения. Продолжая этот процесс найдем  $Z_N(x)$  для  $x \in X_N$ .

Получим величину оптимальных затрат при движении из множества начальных состояний в доступные состояния множества  $X^f$ :

$$Z_N^* = \min_{x \in X^f \cap X_N} Z_N(x),$$

а также оптимальное конечное состояние

$$x_f^* = x^*(t_f) = \arg \min_{x \in X^f \cap X_N} Z_N(x). \quad (2.12)$$

Отсюда, выполняя обратный ход вычислительного процесса (исходя из (2.12)),

определим предшествующие оптимальные значения:

$$\begin{aligned}x^*(t_f - 1) &= \bar{x}^*(x^*(t_f), t_f - 1), \\u^*(t_f - 1) &= \bar{u}^*(x^*(t_f), t_f - 1), \\&\dots \\x^*(t_s) &= \bar{x}^*(x^*(t_s + 1), t_s), \\u^*(t_s) &= \bar{u}^*(x^*(t_s + 1), t_s).\end{aligned}$$

**Замечание.** Запись уравнений Беллмана от начала удобна тогда, когда уравнение динамики  $f(\bar{x}, \bar{u}, t) = x$  однозначно разрешимо относительно  $\bar{x}$ . При этом в (2.10), (2.11) минимум можно брать только по независимой переменной  $\bar{u}$ .

## 2.5. Обобщение рекуррентных соотношений Беллмана на задачи с не фиксированной длительностью процесса

Введем несколько изменений в постановку задачи [11].

**А.** Будем считать, что  $t_s$  задано, а  $t_f$  заранее неизвестно и определяется следующим образом:

$$\forall t < t_f \quad x(t) \notin X^f, \text{ а } x(t_f) \in X^f.$$

При этом для заданного начального состояния  $x$  при разных управлениях  $u(t)$  время  $t_f$  достижения финального множества  $X^f$  может быть различным.

**В.** Будем рассматривать только те задачи, в которых  $U$ ,  $f$  и  $q$  автономны, т.е. не зависят от времени явно:

$$\begin{aligned}x(t + 1) &= f(x(t), u(t)), \\u(t) &\in U(x(t)), \\q(x(t), u(t)) &\text{ — затраты на шаге.}\end{aligned}$$

**С.** Также будем считать, что в пространстве состояний динамической системы нет циклических траекторий.

Вместо функции Беллмана введем некоторую оценку сверху этой функции —  $\bar{S}(x)$  (верхняя оценка функции оптимальных затрат по попаданию в  $X^f$  из  $x$ ).

В построенном ниже вычислительном процессе после  $k$  итераций функция  $\bar{S}(x)$  будет описывать величину оптимальных затрат при переходе из  $x$  в  $X^f$  не более, чем за  $k$  шагов.

Запишем аналог рекуррентных соотношений. Пусть  $Y_0 = X^f$ ,  $Y_1$  — множество состояний, из которых можно за один шаг попасть в  $Y_0$ . Далее, пусть  $Y_{k+1}$  — множество состояний, из которых можно за один шаг попасть в  $Y_k$ . Причем,  $Y_i$  и  $Y_j$  могут пересекаться при  $i \neq j$ .

На первом шаге ( $k = 0$ )

$$\bar{S}(x) = \begin{cases} 0, & x \in Y_0 = X^f, \\ \infty, & x \notin Y_0 = X^f. \end{cases} \quad (2.13)$$

Пусть уже выполнено  $k$  итераций. Опишем  $(k + 1)$ -ю. Рассматриваются  $x \in Y_{k+1}$ . Значения  $\bar{S}(x)$  в правой части равенства (2.14) определяют оптимальные затраты по переходу из  $x$  в  $X^f$  не более, чем за  $k$  шагов.

$$\bar{S}(x) := \min \left\{ \bar{S}(x); \min_{\substack{u \in U(x) \\ f(x,u) \in Y_k}} \{q(x, u) + \bar{S}(f(x, u))\} \right\}. \quad (2.14)$$

Знак  $:=$  подчеркивает, что слева от него подразумевается новый экземпляр функции  $\bar{S}$ , который должен заменить «старый» экземпляр этой функции только после завершения формирования нового экземпляра. В процессе вычислений запоминаются значения  $u^*(x)$ , приведшие к улучшению  $\bar{S}(x)$ . Процесс закончен, когда  $Y_{k+1} = \emptyset$ . В этот момент  $\bar{S}(x)$  определяет оптимальные затраты по достижению  $X^f$  из  $x$ , т.е. становится функцией Беллмана. После этого определяется оптимальное начальное состояние  $x_s^* = \arg \min_{x \in X^s} \bar{S}(x)$ .

**Пример.** Приведем пример использования обобщенных уравнений Беллмана в задаче поиска кратчайшего пути на ориентированном графе. Вершины графа примем за состояния  $x$ , множество ребер, исходящих из вершины будем трактовать как множество управлений  $U(x)$ . Правила перехода  $\bar{x} = f(x, u)$  задаются структурой графа. Шаг перехода соответствует переходу по одному ребру. Вес ребра трактуется как значение функции затрат  $q(x, u)$ .

Рис. 2.7 иллюстрирует последовательные стадии процесса решения, выполняемого обобщенным методом рекуррентных уравнений Беллмана. Текущие значения функции  $\bar{S}(x)$  показаны у вершин в прямоугольных рамках, просматриваемые для вершин  $x$  ребра из  $U(x)$  отмечены волнистой линией. Ребра, соответствующие текущим значениям  $u^*(x)$  отмечены короткими жирными стрелками.

Построенный оптимальный путь выделен двойной линией.

Заметим, что применительно к специальным задачам поиска оптимальных путей на графах существуют методы, более удобные, чем метод Беллмана, например, метод Дейкстры [9]. Он применим в случае неотрицательности

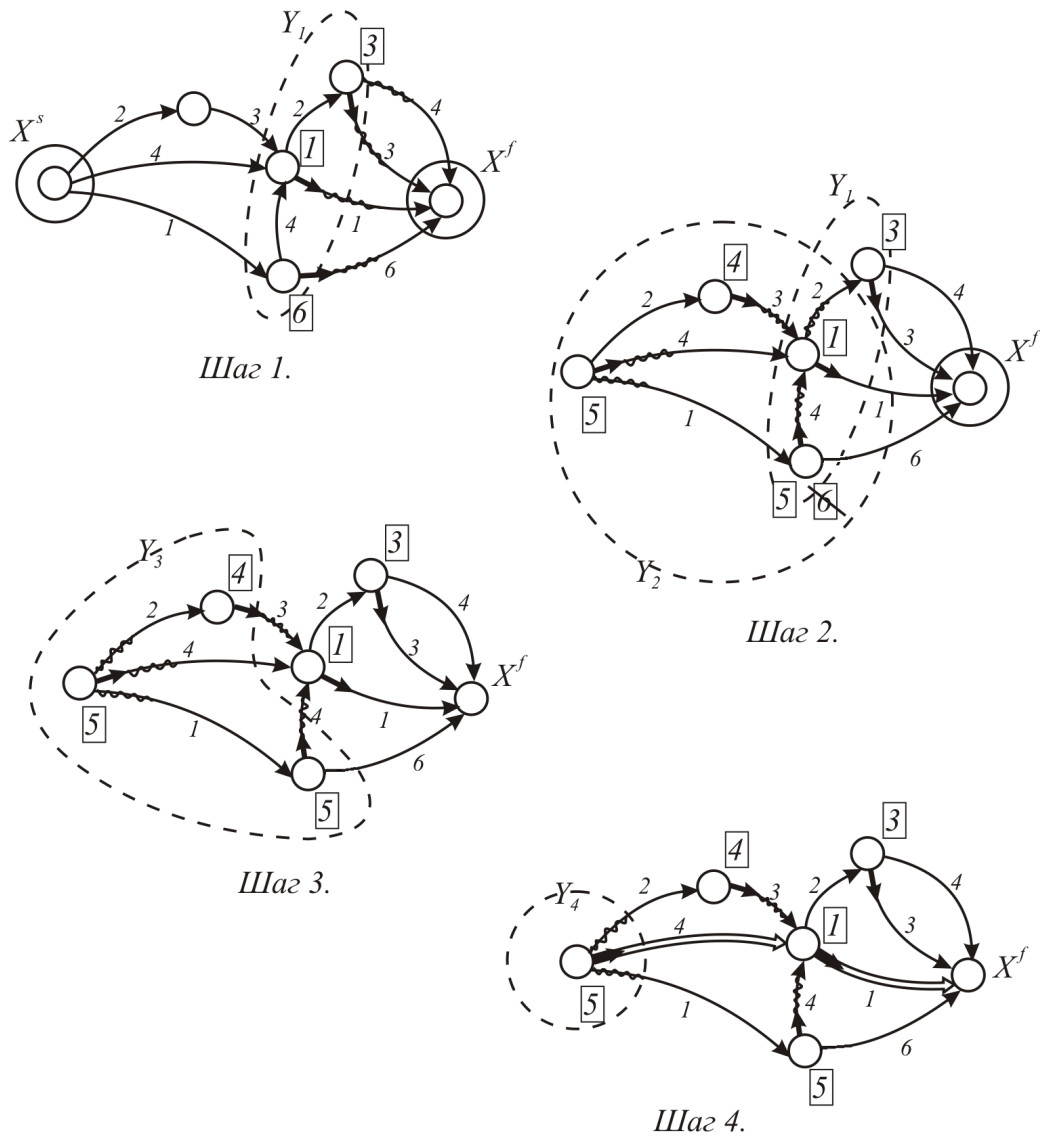


Рис. 2.7. Применение обобщенных уравнений Беллмана

весов ребер. Описание метода Дейкстры, обоснование и связь с принципом Беллмана приведены в описании теоретической части к лабораторной работе «Методы оптимизации на графах с векторными весами ребер» [4], включенной в программу лекционных курсов, соответствующих данному пособию. Материал этой лабораторной работы расширяет лекционный материал и материал данного пособия и должен быть изучен студентами самостоятельно.

# Глава 3. Условия оптимальности в задачах математического программирования

## 3.1. Необходимые сведения из математического анализа

Целью раздела является получение условий, определяющих точку экстремума в задачах с ограничениями

$$\min_{x \in D} Q(x), \quad (3.1)$$

$$D = \{x \in R^N : g(x) \leq 0, h(x) = 0\}, \quad (3.2)$$

где  $g = (g_1, g_2, \dots, g_m)$  и  $h = (h_1, h_2, \dots, h_p)$ . Относительно функций  $Q, g, h$  будем пока предполагать их достаточную гладкость.

Начнем с более простых вопросов и задач, вспоминая известные факты из математического анализа.

1. *Вектор градиента и гиперплоскость, касательная к поверхности равного уровня функции*

Будем обозначать вектор градиента следующим образом

$$\nabla Q(x) = \begin{pmatrix} \frac{\partial Q}{\partial x_1}(x) \\ \dots \\ \frac{\partial Q}{\partial x_n}(x) \end{pmatrix}.$$

Рассмотрим геометрическое место точек  $x$ , для которых  $Q(x) = C$ . В общем случае (а именно, при выполнении условия, когда  $\nabla Q(x) \neq 0$  во всех точках указанного множества), это будет поверхность, которую называют *поверхностью (гиперповерхностью) равного уровня*. Пусть она проходит через точку  $x^0$ . Тогда  $C = Q(x^0)$  и на поверхности равного уровня выполняется равенство

$$Q(x) = Q(x^0). \quad (3.3)$$

Если функция  $Q(x)$  дифференцируема в точке  $x^0$ , то

$$Q(x) = Q(x^0) + \sum_{i=1}^N \frac{\partial Q(x^0)}{\partial x_i} (x_i - x_i^0) + o(\|x - x^0\|). \quad (3.4)$$

При этом с использованием векторной формы записи и эвклидова скалярного произведения:

$$\sum_{i=1}^N \frac{\partial Q(x^0)}{\partial x_i} (x_i - x_i^0) = (\nabla Q(x^0), x - x^0).$$

Подставляя эти выражения в (3.3) и сокращая на  $Q(x^0)$ , получим следующее представление поверхности равного уровня:

$$(\nabla Q(x^0), x - x^0) + o(\|x - x^0\|) = 0.$$

Отбрасывая последний нелинейный член суммы, будем иметь *линейную аппроксимацию поверхности* в окрестности  $x^0$ :

$$(\nabla Q(x^0), x - x^0) = 0.$$

Мы получили *уравнение касательной в точке  $x^0$  гиперплоскости* к поверхности равного уровня. Очевидно, вектор  $\nabla Q(x^0)$  ортогонален к ней. Следовательно, *вектор градиента локально ортогонален также поверхности равного уровня функции в точке  $x^0$* . Это иллюстрирует рис. 3.1.

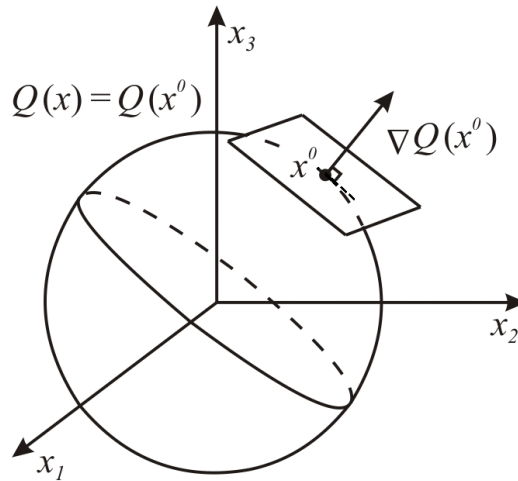


Рис. 3.1. Поверхность равного уровня и касательная гиперплоскость

2. *Производная по направлению, множество направлений строго локального убывания*

Пусть  $v$  — вектор направления. Считаем его нормированным ( $\|v\| = 1$ ). По определению

$$\frac{\partial Q}{\partial v}(x^0) = \left. \frac{dQ(x^0 + \tau v)}{d\tau} \right|_{\tau=+0} = \lim_{\tau \rightarrow +0} \frac{Q(x^0 + \tau v) - Q(x^0)}{\tau}.$$

**Замечание.** *Производная по направлению может существовать и при отсутствии дифференцируемости  $Q$  в точке  $x^0$ . В качестве примера можно указать на недифференцируемую в  $x = 0$  функцию скалярного аргумента  $Q(x) = |x|$ , у которой производные по двум направлениям  $v = \pm 1$  в точке 0 существуют и обе равны 1.*

Получим выражение для производной по направлению в случае дифференцируемости функции в точке  $x^0$ , используя представление (3.4).

$$\frac{\partial Q(x^0)}{\partial v} = \lim_{\tau \rightarrow +0} \frac{Q(x^0) + (\nabla Q(x^0), \tau v) + o(\tau \cdot \|v\|) - Q(x^0)}{\tau} = (\nabla Q(x^0), v).$$

**Замечание.** Множество направлений строгого локального убывания функции, дифференцируемой в точке  $x^0$ , есть открытое полупространство вида:

$$\{v : (\nabla Q(x^0), v) < 0\}$$

с границей, ортогональной  $\nabla Q(x^0)$  (см. рис. 3.2).

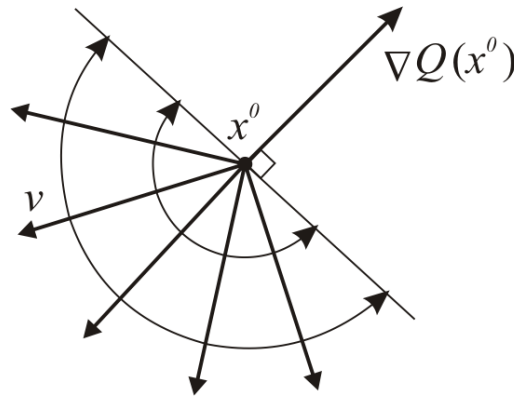


Рис. 3.2. Полупространство направлений  $v$  строгого локального убывания в точке  $x^0$  функции  $Q$ , дифференцируемой в  $x^0$

### 3. Геометрический смысл вектора градиента

Поскольку

$$\frac{\partial Q(x^0)}{\partial v} = (\nabla Q(x^0), v) = \|\nabla Q(x^0)\| \cdot \|v\| \cos \varphi,$$

где  $\varphi$  — угол между векторами  $\nabla Q(x^0)$  и  $v$ ,  $\|v\| = 1$ ,  $\varphi \in [0, 2\pi]$ , то при  $\nabla Q(x^0) \neq 0$  производная по направлению будет максимальна при  $\varphi = 0$ , что соответствует

$$v = \nabla Q(x^0) / \|\nabla Q(x^0)\|,$$

т.е.  $\nabla Q(x^0)$  — направление скорейшего локального возрастания функции, аналогично при  $v = -\nabla Q(x^0) / \|\nabla Q(x^0)\|$  производная по направлению достигает своего минимума (при значении меньшем нуля). Этому направлению соответствует  $\varphi = \pi$ . Таким образом, вектор  $-\nabla Q(x^0)$  (антиградиент) является направлением скорейшего локального убывания.

### 4. Условие экстремума первого порядка при отсутствии ограничений

Рассмотрим задачу поиска минимума функции (см. [12], гл. 4-5):

$$\min_{x \in R^N} Q(x).$$



**Теорема (Ферма).** Если  $x^o$  — локальный экстремум функции  $Q$  в  $R^N$  (минимум или максимум), и  $Q$  дифференцируема в точке  $x^o$ , то  $\nabla Q(x^o) = 0$ .

*Доказательство.* Обоснуем теорему на примере задачи на минимум. Пусть утверждение не верно и  $\nabla Q(x^o) \neq 0$ . Рассмотрим точки, смещенные от  $x^o$  в направлении  $v = -\nabla Q(x^o)/\|\nabla Q(x^o)\|$ . Из дифференцируемости следует, что

$$\begin{aligned} Q(x^o + \tau v) &= Q(x^o) + \left. \frac{dQ(x^o + \tau v)}{d\tau} \right|_{\tau=+0} \cdot \tau + o(\tau) = \\ &= Q(x^o) + (\nabla Q(x^o), v)\tau + o(\tau) = Q(x^o) - \|\nabla Q(x^o)\|\tau + o(\tau) < Q(x^o) \end{aligned}$$

при любых достаточно малых  $\tau$ . Это противоречит локальной оптимальности точки  $x^o$ . □

Доказательство теоремы Ферма опирается на возможность смещения из точки  $x^o$  в любом направлении в силу отсутствия ограничений. При наличии ограничений, когда  $x \in D \neq R^N$ , градиент  $\nabla Q(x^o)$  в точке локального минимума может быть не равен нулю, если все направления локального убывания будут выводить из допустимого множества. То есть, при наличии ограничений для точек локального минимума ориентация вектора градиента  $\nabla Q(x^o)$  должна быть согласована с направлением локальных нормалей к гладким фрагментам границы допустимого множества в точке  $x^o$ .

### 5. Функция Лагранжа

В общем случае задачи (3.1)–(3.2) условия экстремума удобно записывать через функцию Лагранжа. Для этой задачи она имеет вид:

$$L(x, \lambda_0, \lambda, \mu) = \lambda_0 Q(x) + \sum_{j=1}^m \lambda_j g_j(x) + \sum_{i=1}^p \mu_i h_i(x), \quad (3.5)$$

где  $\lambda_0 \in R^1$ ,  $\lambda \in R^m$ ,  $\mu \in R^p$  — множители Лагранжа.

Если ограничений–неравенств в задаче нет, то сумма  $\sum_{j=1}^m \lambda_j g_j(x)$  исчезает.

Для задачи с ограничениями–равенствами

$$\min_{h(x)=0} Q(x) \quad (3.6)$$

функция Лагранжа имеет вид:

$$L(x, \lambda_0, \mu) = \lambda_0 Q(x) + \sum_{i=1}^p \mu_i h_i(x). \quad (3.7)$$

**Замечание.** Сам Лагранж рассматривал именно задачу (3.6) и функцию вида (3.7) при  $\lambda_0 = 1$ .

### 3.2. Необходимые условия оптимальности в задаче с равенствами в форме теоремы Лагранжа, понятие регулярности допустимого множества

**Теорема** (Лагранжа). Пусть в задаче с равенствами (3.6)  $Q$  и  $h$  дифференцируемы в окрестности точки  $x^o$ , и их производные непрерывны в точке  $x^o$ . Тогда, чтобы  $x^o$  была локальным экстремумом (минимумом или максимумом), необходимо существование нетривиального набора множителей  $(\lambda_0^*, \mu^*) \neq 0$ , при котором для функции Лагранжа (3.7) выполняется условие стационарности по совокупности переменных  $x, \mu$  в точке  $(x^o, \mu^*)$ :

$$\nabla_{x,\mu} L(x^o, \lambda_0^*, \mu^*) = 0. \quad (3.8)$$

**Замечание.** Поскольку

$$\nabla_x L(x, \lambda_0, \mu) = \lambda_0 \nabla Q(x) + \sum_{i=1}^p \mu_i \nabla h_i(x),$$

$$\nabla_\mu L(x, \lambda_0, \mu) = h(x),$$

равенство (3.8) эквивалентно записывается в виде системы условий:

$$\begin{cases} \lambda_0^* \nabla Q(x^o) + \sum_{i=1}^p \mu_i^* \nabla h_i(x^o) = 0, & (3.9) \\ h(x^o) = 0. & (3.10) \end{cases}$$

Заметим, что условие (3.8) выглядит так же, как условие  $\nabla Q(x^o) = 0$  в теореме Ферма, но оно записывается уже для функции Лагранжа, и рассматривается в расширенном пространстве  $x, \mu$ . Система (3.9)-(3.10) — полная система уравнений относительно  $x^o$  и  $\mu^*$ . При этом  $\lambda_0^*$  выполняет роль параметра.

*Доказательство.* Условие (3.10) означает всего лишь допустимость точки  $x^o$ . То есть, оно очевидно. А условие (3.9) при  $(\lambda_0^*, \mu^*) \neq 0$  означает линейную зависимость входящих в него векторов (их размерность равна  $N$ )

$$\nabla Q(x^o), \nabla h_1(x^o), \dots, \nabla h_p(x^o). \quad (3.11)$$

Пусть  $p+1 > N$ . Тогда эти векторы, очевидно, зависимы и теорема верна.

Пусть  $p+1 \leq N$ . Предположим, что утверждение неверно и система векторов (3.11) линейно независима. Рассматривая все указанные векторы как вектор-столбцы, образуем из них матрицу (см. рис.3.3). Ее столбцовый

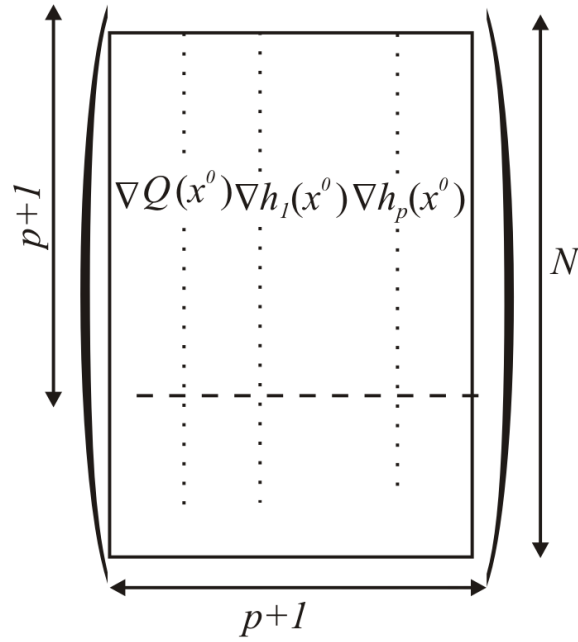


Рис. 3.3. Выделение невырожденного квадратного блока

ранг равен  $(p + 1)$  и равен строковому. Поэтому в матрице найдется ровно  $(p + 1)$  независимая строка. За счет перенумерации переменных переставим эти строки в верхнюю часть матрицы, образовав невырожденный квадратный блок  $J$  размера  $(p + 1) \times (p + 1)$ .

Разобьем векторы  $x$  и  $x^o$  (после перенумерации) на соответствующие части, где первая часть включает  $(p + 1)$  компоненту:

$$x = (x_I, x_{II}), \quad x^o = (x_I^o, x_{II}^o).$$

Построим вектор-функцию

$$F(x_I) = (Q(x_I, x_{II}^o) - Q(x^o), h_1(x_I, x_{II}^o), \dots, h_p(x_I, x_{II}^o)),$$

в которой переменными являются только компоненты  $x_I$ .  $F(x_I)$  непрерывно отображает окрестность точки  $x_I^o \in R^{p+1}$  в окрестность точки  $0 \in R^{p+1}$  (так как  $F(x_I^o) = 0$  и функции  $Q$  и  $h_i$  непрерывны по  $x_I$  в окрестности  $x_I^o$ ).

Заметим, что подматрица  $J$  является матрицей Якоби для отображения  $F$  в точке  $x_I = x_I^o$ . Так как определитель  $\det J \neq 0$ , то существует обратное отображение  $F^{-1}$ , которое отображает окрестность точки  $0 \in R^{p+1}$  в окрестность точки  $x_I^o \in R^{p+1}$  (обобщение теоремы об обратной функции).

Если производные функции  $Q$  и  $h$  непрерывны, то элементы матрицы Якоби  $J(x)$  непрерывны в точке  $x^o$ . Этого достаточно для того, чтобы обратное отображение  $F^{-1}$  было непрерывным. Тогда  $F^{-1}$  будет переводить точки близкие к точке  $0$  в точки близкие к  $x_I^o$ . Теорема о непрерывности обратного отображения доказана в анализе [12].

Дальнейшие рассуждения проведем применительно к задаче на минимум. В задаче на максимум рассуждения аналогичны.

Выберем точку  $(-\varepsilon, 0, \dots, 0)$ , где  $\varepsilon > 0$ . Эта точка по  $F^{-1}$  отобразится в некоторую точку  $x_I^\varepsilon$ . Из непрерывности  $F^{-1}$  следует, что

$$x_I^\varepsilon \xrightarrow{\varepsilon \rightarrow 0} x_I^0.$$

Но  $F(x_I^\varepsilon) = (-\varepsilon, 0, \dots, 0)$ , то есть  $Q(x_I^\varepsilon, x_{II}^0) - Q(x^0) = -\varepsilon < 0$ ,  $h(x_I^\varepsilon, x_{II}^0) = 0$ . Это противоречит тому, что точка  $x^0$  — локальный минимум.  $\square$

Заметим, что выполнение (3.9) возможно и при  $\lambda_0^* = 0$ . Но при этом получается, что свойства функции  $Q(x)$  вообще не учитываются в условиях оптимальности. Ясно, что это связано с вырождением структуры допустимого множества в точке  $x^0$ . Такие случаи называют случаями *нерегулярности*.

Дадим определение более точно. Пусть рассматривается класс задач вида

$$\min_{x \in D} Q(x),$$

где предполагается, что  $Q$  и  $D$  удовлетворяют определенным требованиям. Пусть также для этого класса задач существуют условия оптимальности, записываемые через функцию Лагранжа.

**Определение.** Говорят, что допустимое множество  $D$  *регулярно в точке*  $x^0 \in D$ , если для любой функции  $Q$  (такой, что полученная задача принадлежит рассматриваемому классу), имеющей в точке  $x^0$  условный локальный минимум на  $D$ , условия оптимальности могли выполняться при  $\lambda_0^* \neq 0$ .

Допустимое множество  $D$  *регулярно в целом*, если оно регулярно в каждой своей точке.

Получим простые достаточные условия регулярности применительно к множествам, описанным гладкими ограничениями–равенствами.

**Теорема** (достаточное условие регулярности). *Для класса задач, рассматриваемого в теореме Лагранжа, для регулярности допустимого множества  $D = \{x : h(x) = 0\}$  в точке  $x^0$  достаточно, чтобы набор градиентов*

$$\nabla h_1(x^0), \dots, \nabla h_p(x^0)$$

*был линейно независим.*

*Доказательство.* Пусть регулярности нет. Тогда существует такая функция  $Q$  с нужными свойствами, имеющая на  $D$  в точке  $x^0$  условный локальный

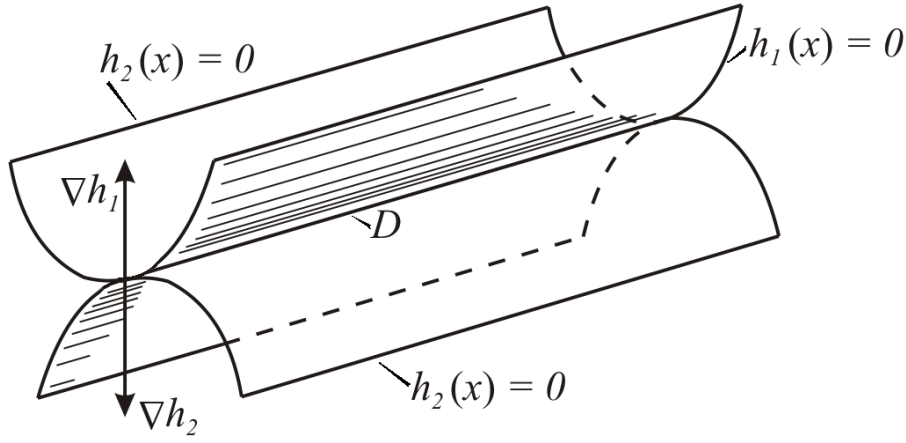


Рис. 3.4. Нарушение достаточного условия регулярности, вызванное касанием поверхностей вдоль линии их пересечения

минимум, что условия оптимальности теоремы Лагранжа выполняются при  $\lambda_0^* = 0$ . Но так как  $(\lambda_0^*, \mu^*) \neq 0$ , то  $\mu^* \neq 0$ , а из (3.9) будет вытекать, что

$$\sum_{i=1}^p \mu_i^* \nabla h_i(x^0) = 0.$$

То есть получили линейную зависимость, что противоречит условиям.  $\square$

1. *Геометрическая иллюстрация достаточного условия регулярности на примере*

Пусть  $p = 2, N = 3$ . Нарушение достаточного условия в каждой из точек допустимого множества будет в случае касания вдоль этого множества поверхностей  $h_1(x) = 0$  и  $h_2(x) = 0$  (рис.3.4). Градиенты  $\nabla h_1(x), \nabla h_2(x)$  локально ортогональны этим поверхностям, а в точках множества  $D$ , вдоль которого происходит касание поверхностей, градиенты будут направлены по одной прямой, т.е. окажутся линейно зависимы.

2. *Геометрическая интерпретация условий Лагранжа при регулярности допустимого множества*

При выполнении условия регулярности равенство (3.9) переписется в следующем виде:

$$-\nabla Q(x^0) = \sum_{i=1}^p \mu_i^* \nabla h_i(x^0). \quad (3.12)$$

Допустимое множество  $D$  представлено в виде нелинейного локально гладкого многообразия, порождаемого пересечением поверхностей  $h_i(x) = 0$ , где  $(i = \overline{1, p})$ . Нормальными к ним в точке  $x^0$  являются  $\nabla h_1(x^0), \dots, \nabla h_p(x^0)$  (см. рис. 3.5). При всевозможных  $\mu_i$  множество векторов  $u = \sum_{i=1}^p \mu_i \nabla h_i(x^0)$  образуют линейное многообразие  $M$ , локально ортогональное к  $D$ .

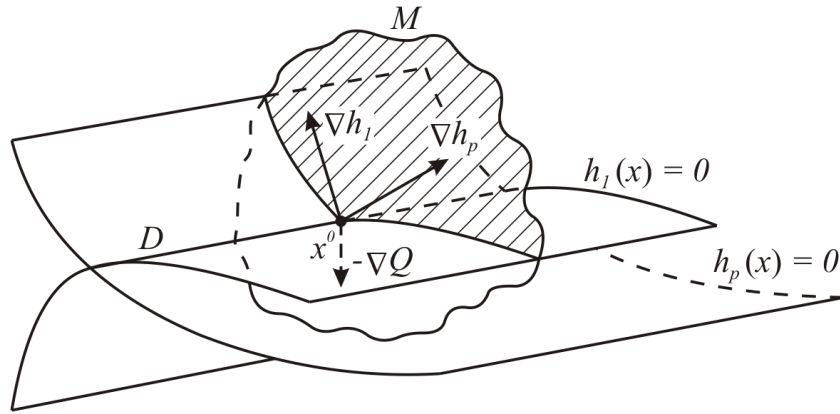


Рис. 3.5. Многообразие  $M$ , локально ортогональное  $D$  в точке  $x^0$ , образовано всевозможными комбинациями нормалей  $\nabla h_1(x^0), \dots, \nabla h_p(x^0)$

Если  $x^0$  — локальный минимум, то  $-\nabla Q(x^0)$  в силу равенства (3.12) принадлежит  $M$ . Это означает, что  $-\nabla Q(x^0)$  локально ортогонален многообразию  $D$  в точке  $x^0$ . Следовательно, поверхность равного уровня  $Q(x) = Q(x^0)$  касается множества  $D = \{x : h_1(x) = 0, \dots, h_p(x) = 0\}$  в точке  $x^0$ . Это показано на рис. 3.6.

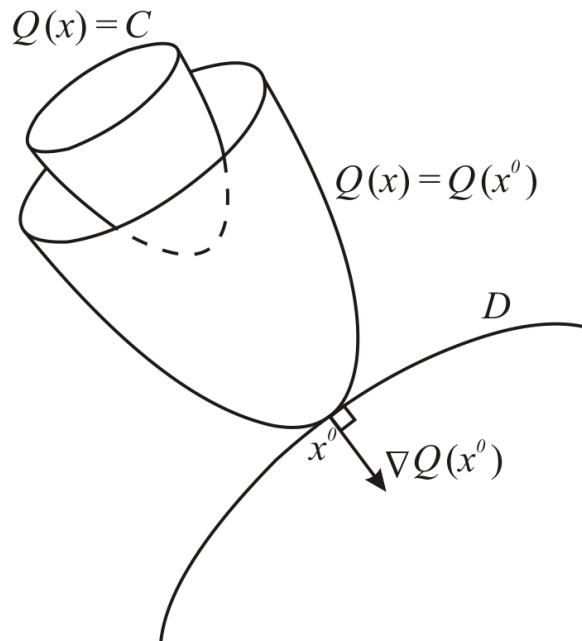


Рис. 3.6. Выполнение условий оптимальности Лагранжа означает касание в  $x^0$  поверхности равного уровня  $Q(x)$  с допустимым множеством  $D$

**Пример** (иллюстрация к теореме Лагранжа).

Найти экстремумы функции  $\frac{1}{3}x_1^3 + x_2$  при ограничении-равенстве вида  $x_1^2 + x_2^2 = 1$ :

$$\min_{x_1^2 + x_2^2 = 1} \left( \frac{1}{3}x_1^3 + x_2 \right).$$

Градиент  $\nabla Q(x)$  направлен по нормали к линии уровня  $\frac{1}{3}x_1^3 + x_2 = C$ . Градиент  $\nabla h(x)$  направлен по нормали к окружности  $x_1^2 + x_2^2 = 1$  в направлении роста квадрата расстояния от точки  $(0, 0)$ .

$$\nabla Q(x) = \begin{pmatrix} x_1^2 \\ 1 \end{pmatrix}, \nabla h(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}.$$

Так как  $\nabla h(x) \neq 0$  на этой окружности, значит допустимое множество всюду регулярно, то есть  $\lambda_0^* \neq 0$ . Положим  $\lambda_0^* = 1$ .

Из теоремы Лагранжа:  $-\nabla Q(x^o) = \mu_1^* \nabla h(x^o)$ , то есть  $x^o$  — такая точка, где  $\nabla Q(x^o)$  и  $\nabla h(x^o)$  направлены по одной прямой.

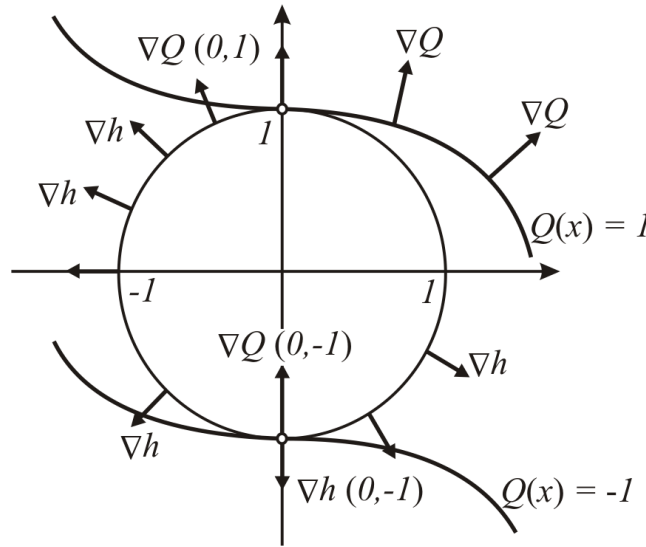


Рис. 3.7. Допустимое множество и линии уровня функции  $Q$ ; ориентации градиентов

Из рис. 3.7 очевидно, что это выполняется для точек  $(0, \pm 1)$ . Действительно, запишем систему уравнений Лагранжа

$$\begin{cases} \begin{pmatrix} -x_1^2 \\ -1 \end{pmatrix} = \mu \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}, \\ x_1^2 + x_2^2 = 1. \end{cases}$$

Из первого уравнения  $-x_1 \cdot x_1 = \mu \cdot 2x_1$ . Возможно два варианта:  $x_1 = 0$  и  $x_1 \neq 0$ .

Если  $x_1 = 0$ , тогда из последнего уравнения системы  $x_2 = \pm 1$ , а из второго  $\mu = \mp \frac{1}{2}$ . Если  $x_1 \neq 0$ , тогда  $x_1 = \gamma = -2\mu$  и из второго уравнения системы  $1 = \gamma x_2$ . Отсюда, используя третье уравнение, имеем:

$$\gamma^2 + \frac{1}{\gamma^2} = 1;$$

$$\gamma^4 - \gamma^2 + 1 = 0.$$

Полученное уравнение не имеет действительных решений. Следовательно остается два варианта:  $x_1 = 0, x_2 = \pm 1$ .

Заметим, что одна из этих точек — минимум, другая — максимум. Это точки касания линий равного уровня функции  $Q(x)$  и допустимого множества, имеющего вид окружности.

Нас будут интересовать дальнейшие обобщения условий оптимальности на задачи с неравенствами. Для этого понадобятся некоторые факты из теории выпуклых множеств и функций.

### 3.3. Элементы теории выпуклых множеств и функций

#### 3.3.1. Выпуклые множества, проекция, теоремы об отделимости

Начала теории выпуклых множеств были заложены немецким математиком Германом Минковским (1864–1909). Необходимую информацию по данной теме можно найти в книгах [2], [5], дополнительную — в [15], [18].

Далее будем предполагать, что работаем в пространстве  $R^N$ , в котором введено эвклидово скалярное произведение

$$(x, y) = \sum_{i=1}^N x_i y_i,$$

норма

$$\|x\| = \sqrt{(x, x)}$$

и метрика

$$\rho(x, y) = \|x - y\|.$$

**Определение.** Отрезком  $[x, y]$  в  $R^N$  называют геометрическое место точек

$$\{x^\alpha : x^\alpha = \alpha x + (1 - \alpha)y, \alpha \in [0, 1]\}.$$

**Определение.** Множество  $D$  называют *выпуклым*, если для любых  $x$  и  $y$  из  $D$  отрезок  $[x, y] \subseteq D$ , то есть если  $\forall x, y \in D$  и  $\forall \alpha \in [0, 1]$  :

$$\alpha x + (1 - \alpha)y \in D.$$

Определение иллюстрирует рис. 3.8.

**Замечание.** Пустое множество  $\emptyset$  выпукло по определению.

*Задания на самостоятельную проработку*

Пусть  $G_i$  — выпуклые множества. Доказать выпуклость множеств:



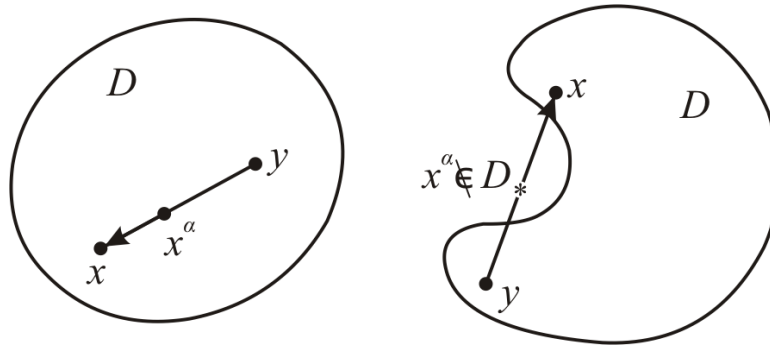


Рис. 3.8. Примеры выпуклого и невыпуклого множеств

1.  $\bigcap_{i=1}^s G_i$ ;
2.  $\sum_{i=1}^s G_i = \left\{ x : x = \sum_{i=1}^s u^i, u^i \in G_i \right\}$ ;
3.  $\alpha G_1 = \{x : x = \alpha u^1, u^1 \in G_1\}$  для  $\alpha \geq 0$  и  $\alpha < 0$ ;
4.  $G_1 - G_2 = G_1 + (-1)G_2$ ;
5. Множества

$$\Lambda(x^1, \dots, x^k) = \left\{ x : x = \sum_{i=1}^k x^i \lambda_i, \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1 \right\},$$

где  $k > 1$ , называемого *выпуклой линейной оболочкой*.

**Определение.** *Проекцией* точки  $y$  на непустое множество  $D$  называют точку  $x_y \in D$  (если таковая существует), что

$$\forall x \in D : \|x_y - y\| \leq \|x - y\|.$$

То есть проекция — это *точка из  $D$ , ближайшая к точке  $y$*  (см. рис. 3.9).

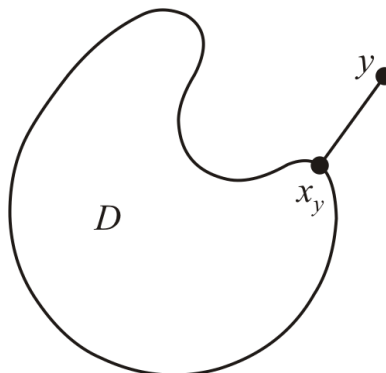


Рис. 3.9. Точка  $y$  и ее проекция  $x_y$  на множество  $D$

Рассмотрим несколько простых лемм о проекции.

## Леммы о свойствах проекций

**Лемма 1.** Существует единственная проекция точки  $y$  на непустое, выпуклое, замкнутое множество  $D$ .

*Доказательство.*

1. Если  $y \in D$ , то  $x_y = y$ , так как тогда

$$\|x_y - y\| = 0 \leq \|x - y\|.$$

Это неравенство выполняется для любого  $x \in D$ . Кроме того, в этом случае не найдется иной точки  $\tilde{x}_y$  из  $D$ , чтобы  $\|\tilde{x}_y - y\| = 0$ . То есть имеет место единственность.

2. Пусть  $y \notin D$ . Так как  $D \neq \emptyset$ , то существует  $\tilde{x} \in D$ . Пусть  $R = \|\tilde{x} - y\|$  и  $\overline{O}_R(y) = \{x : \|x - y\| \leq R\}$  — замкнутый шар.

Построим  $\tilde{D} = D \cap \overline{O}_R(y)$ . Множество  $\tilde{D}$  непустое, ограниченное, замкнутое. Следовательно, по соответствующей теореме математического анализа, оно будет компактно.

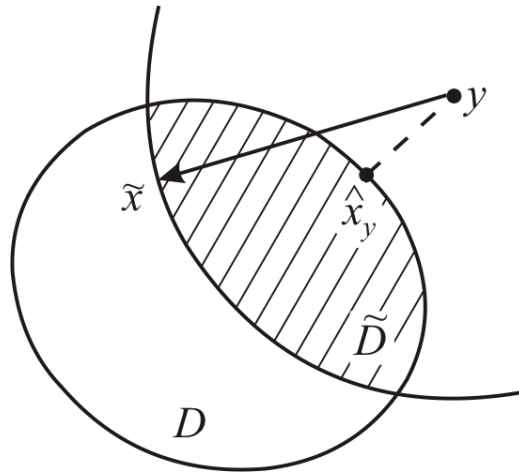


Рис. 3.10. Иллюстрация к доказательству леммы 1 при  $y \notin D$

Введем функцию расстояния  $f(x) = \|x - y\|$ . Она непрерывна на  $\tilde{D}$ . Так как  $\tilde{D}$  — компакт, то по теореме Вейерштрасса она достигает на  $\tilde{D}$  своей точной нижней грани  $\inf_{x \in \tilde{D}} f(x)$ . То есть существует  $\hat{x}_y \in \tilde{D}$ , что для любого  $x \in \tilde{D}$ :

$$\|\hat{x}_y - y\| \leq \|x - y\| \leq R.$$

Но для точек  $x \in D \setminus \tilde{D} = \{x : x \in D \text{ \& } x \notin \tilde{D}\}$ , очевидно, выполняется  $\|x - y\| > R$ , поэтому

$$\|\hat{x}_y - y\| \leq \|x - y\|$$

для любых  $x \in D$ . Следовательно,  $\hat{x}_y$  и есть проекция.

Докажем её единственность.

Пусть существуют по крайней мере две проекции  $x'_y$  и  $x''_y$ , что норма разности  $\|x'_y - x''_y\| \neq 0$ . Построим линейное подпространство  $L$  размерности 2, натянутое на вектора  $x'_y - y$  и  $x''_y - y$ .

Точка  $\bar{x} = (x'_y + x''_y)/2$  будет принадлежать  $D$  в силу его выпуклости. Она также будет принадлежать  $L$ . В  $L$  можно опираться на геометрические представления планиметрии. Треугольник  $(y, x'_y, x''_y)$  — равнобедренный,  $(y; \bar{x})$  — его медиана. Она же будет и его высотой. Получаем два равных прямоугольных треугольника:  $(y, x'_y, \bar{x})$  и  $(y, x''_y, \bar{x})$  на двумерном многообразии  $L$ . Век-

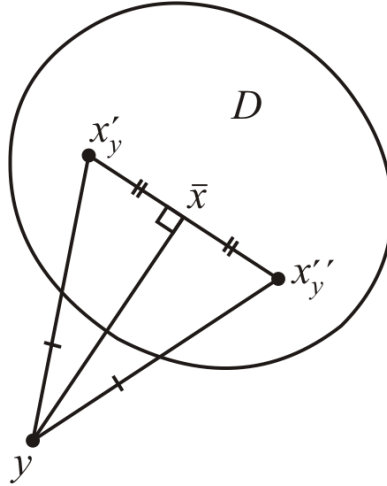


Рис. 3.11.  $x'_y \neq x''_y$  не могут быть проекциями на выпуклое  $D$

тор  $\bar{x} - y$  будет ортогонален вектору  $x'_y - x''_y$ . Поскольку длины катетов  $[x'_y, \bar{x}]$  и  $[\bar{x}, x''_y]$  положительны, то высота  $[y, \bar{x}]$  полученного треугольника  $(x'_y, \bar{x}, y)$  меньше гипотенузы, то есть

$$\|\bar{x} - y\| < \|x'_y - y\| = \|x''_y - y\|.$$

Следовательно,  $\bar{x}$  находится от  $y$  на меньшем расстоянии, чем  $x'_y$  и  $x''_y$ . Однако, точка  $\bar{x} \in D$ . Следовательно,  $x'_y$  и  $x''_y$  не являются проекциями. Полученное противоречие завершает доказательство леммы 1. □

**Лемма 2.** Пусть  $D$  — непустое, выпуклое множество. Точка  $x_y \in D$  является проекцией точки  $y$  на  $D$  тогда и только тогда, когда

$$\forall x \in D : (x - x_y, y - x_y) \leq 0. \quad (3.13)$$

Геометрически это означает, что признаком проекции является то, что угол между векторами  $x - y$  и  $y - x_y$  всегда неострый (рис.3.12).

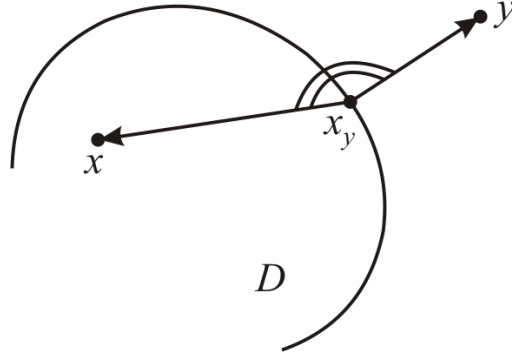


Рис. 3.12. Угол при  $x_y$  не может быть острым — критерий проекции

*Доказательство.* Возьмем произвольную точку  $x \in D$ :

$$\|x - y\|^2 = \|(x - x_y) - (y - x_y)\|^2 = \|x - x_y\|^2 - 2(x - x_y, y - x_y) + \|x_y - y\|^2.$$

Обозначим  $A = \|x - x_y\|^2 - 2(x - x_y, y - x_y)$ ,  $B = (x - x_y, y - x_y)$ .

*Достаточность.*

Пусть  $x_y \in D$  и для любого  $x \in D$  верно (3.13). Тогда  $B \leq 0$ , откуда следует, что  $A \geq 0$ . Отсюда

$$A + \|x_y - y\|^2 \geq \|x_y - y\|^2.$$

Но  $A + \|x_y - y\|^2 = \|x - y\|^2$ , значит  $x_y$  — проекция, в силу произвольности выбора  $x$  из  $D$ .

*Необходимость.*

Пусть теперь  $x_y$  — проекция. Тогда для любого  $x \in D$  выполняется

$$\|x_y - y\|^2 \leq \|x - y\|^2 = A + \|x_y - y\|^2.$$

Отсюда  $A \geq 0$ .

Итак, для любого  $x \in D$

$$\|x - x_y\|^2 - 2(x - x_y, y - x_y) \geq 0. \quad (3.14)$$

Нам хотелось бы избавиться от первого элемента этой суммы. Используем произвол в  $x$ . Вначале выберем  $x$  произвольно из  $D$ . Затем введем вектор

$$x_\gamma = x_y + \gamma(x - x_y).$$

При  $\gamma \in [0, 1]$   $x_\gamma \in [x, x_y] \subseteq D$ . Тогда для  $x = x_\gamma$  верно (3.14).

После подстановки  $x_\gamma$  вместо  $x$  в (3.14) имеем для любого  $\gamma \in [0, 1]$ :

$$\gamma^2 \|x - x_y\|^2 - 2\gamma(x - x_y, y - x_y) \geq 0.$$

Если предположить, что скалярное произведение  $B > 0$ , то при  $\gamma \rightarrow 0$  получим отрицательный знак у данной разности, чего быть не может. Следовательно,  $B \leq 0$ . Лемма доказана.  $\square$

## Понятие отделимости множества и точки. Теоремы отделимости

Теоремы отделимости понадобятся при выводе условий оптимальности при ограничениях–неравенствах.

**Определение.** Множество  $D$  и точка  $y$  *отделимы*, если существует вектор  $a : \|a\| = 1$ , что

$$\forall x \in D : (a, x) \leq (a, y).$$

**Определение.** Множество  $D$  и точка  $y$  *строго отделимы*, если существует вектор  $a : \|a\| = 1$ , что

$$\forall x \in D, (x \neq y) : (a, x) < (a, y).$$

**Определение.** Множество  $D$  и точка  $y$  *сильно отделимы*, если существует вектор  $a : \|a\| = 1$  и существует число  $\varepsilon > 0$ , что

$$\forall x \in D : (a, x) \leq (a, y) - \varepsilon.$$

### Геометрическая интерпретация отделимости

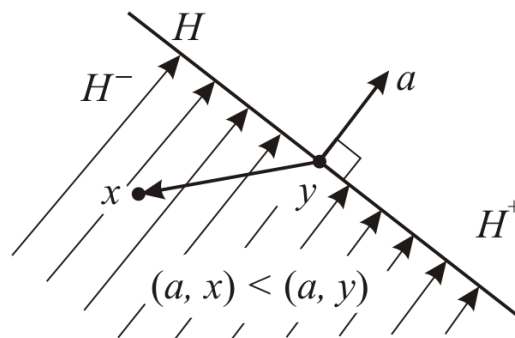


Рис. 3.13. Два полупространства, порождаемые гиперплоскостью  $(a, x) = (a, y)$

Равенство

$$(a, x) = (a, y)$$

определяет гиперплоскость  $H = \{x : a_1x_1 + a_2x_2 + \dots + a_Nx_N = (a, y)\}$ , проходящую через точку  $y$ . Вектор  $a$  является ее нормалью (рис. 3.13).

При отделимости  $y$  от  $D$  множество  $D$  лежит целиком в замыкании отрицательного полупространства  $H^-$ , порождаемого этой гиперплоскостью:

$$H^- = \{x : (x, a) < (y, a)\}.$$

На рис. 3.14 показаны разные виды отделимости. Поскольку под отделимостью понимается отделимость гиперплоскостью, то ситуация, показанная на рис.3.14 (d), — это неотделимость.

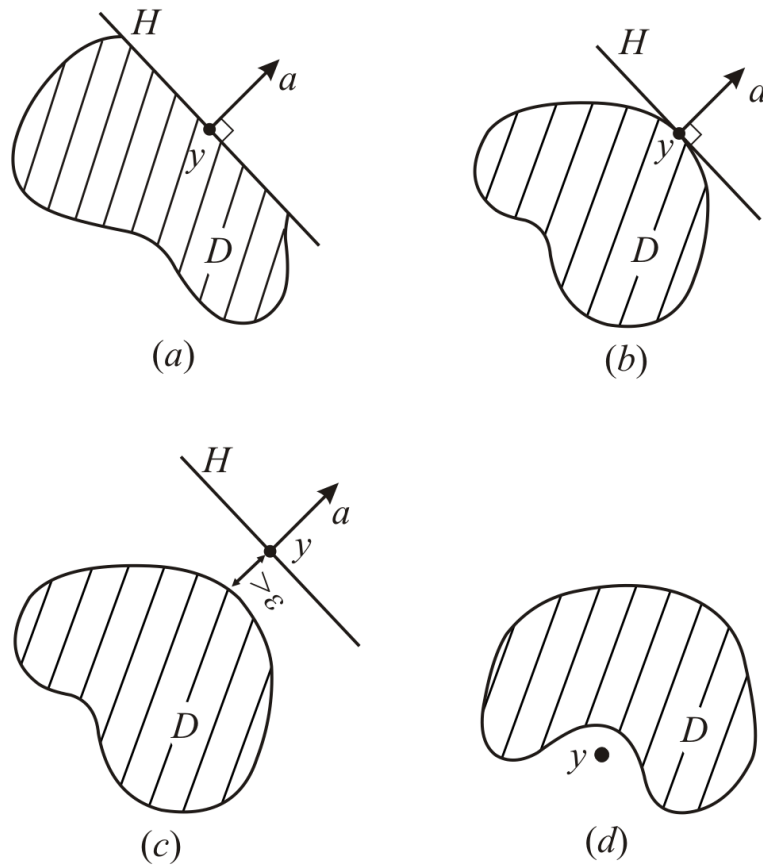


Рис. 3.14. Случаи обычной, строгой и сильной отделимости — а, b, с. Случай неотделимости — d

**Теорема 1** (о сильной отделимости). *Непустое, замкнутое, выпуклое множество  $D$  и не принадлежащая ему точка  $y$  сильно отделимы.*

*Доказательство.* По лемме 1 существует единственная точка  $x_y \in D$  являющаяся проекцией точки  $y$  на  $D$ . Возьмем вектор  $d = y - x_y$ .  $d \neq 0$ , так как  $y \notin D$ . В качестве  $a$  будем выбирать  $d/\|d\|$  (рис.3.15).

По второй лемме о свойствах проекции

$$\forall x \in D : (x - x_y, y - x_y) = (x - x_y + y - y, y - x_y) = (x - y + d, d) \leq 0.$$

Отсюда

$$(x, a) \leq (y, a) - \|d\|.$$

Т.е. выполнено условие сильной отделимости при  $\varepsilon = \|d\|$ .  $\square$

**Теорема 2** (об отделимости). *Всякое непустое, выпуклое множество и любая его граничная точка отделимы.*

*Доказательство.* Рассмотрим упрощенное доказательство.

Обозначим через  $\bar{D}$  замыкание множества, о котором говорится в теореме. Точка  $y$  — граничная точка множества  $D$ . Можно показать, что в силу

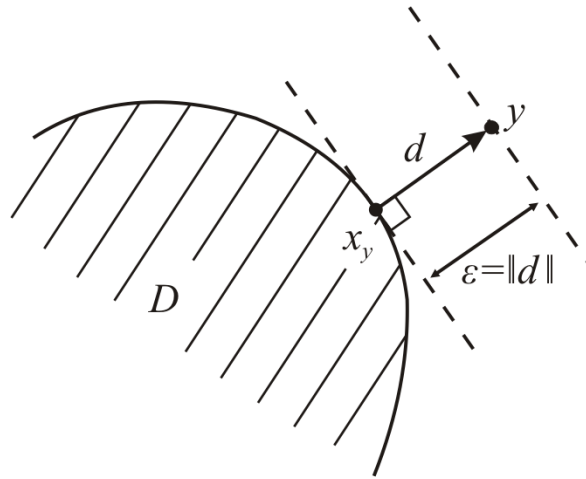


Рис. 3.15. К доказательству теоремы о сильной отделимости

выпуклости  $D$  точка  $y$  будет одновременно являться граничной точкой замыкания  $\overline{D}$ , которое также является выпуклым (это предлагается доказать самостоятельно). Тогда в любой окрестности  $y$  существуют точки как принадлежащие  $\overline{D}$ , так и не принадлежащие ему. Построим последовательность точек  $y_k \notin \overline{D}$ , чтобы:

$$\lim_{k \rightarrow \infty} y_k = y.$$

Для пары  $\overline{D}, y_k$  применима предыдущая теорема 1 (рис. 3.16). Значит

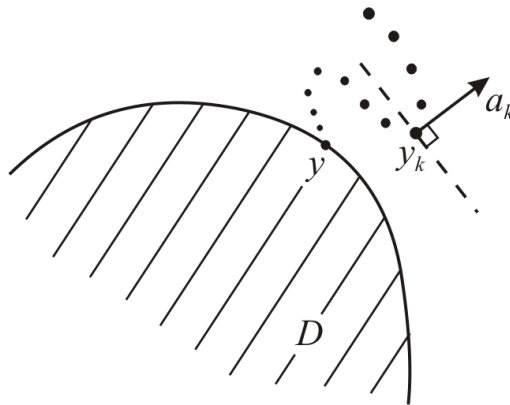


Рис. 3.16. К доказательству отделимости граничной точки

существуют векторы  $a_k$ , ( $\|a_k\| = 1$ ) и числа  $\varepsilon_k > 0$ , что выполняется:

$$\forall x \in \overline{D} : (a_k, x) \leq (a_k, y_k) - \varepsilon_k,$$

следовательно

$$(a_k, x) < (a_k, y_k). \quad (3.15)$$

Так как  $\{a : \|a\| = 1\}$  — компакт, то из последовательности  $\{a_k\}$  можно выделить сходящуюся подпоследовательность  $a_{k_s} \xrightarrow{s \rightarrow \infty} a^*$ . Переходя к пределу

в (3.15) на подпоследовательности  $k = k_s$  получим:

$$\forall x \in D : (a^*, x) \leq (a^*, y),$$

то есть отделимость есть. □

На основе доказанной теоремы позднее будут получены условия оптимальности в выпуклой задаче математического программирования.

### 3.3.2. Выпуклые функции и их свойства

В задачах математического программирования оказывается, что если их элементы обладают свойствами выпуклости, то эти задачи значительно проще решать. Для них существуют более сильные аналитические результаты, а также специальные численные методы. Поэтому понятие выпуклости важно.

**Определение.** Функция  $Q$ , определенная на выпуклом множестве  $D$ , называется *выпуклой*, если для любых  $x, y \in D$  и всех  $\alpha \in [0, 1]$  выполняется

$$Q(\alpha x + (1 - \alpha)y) \leq \alpha Q(x) + (1 - \alpha)Q(y). \quad (3.16)$$

Если для любых  $x, y \in D$  при  $x \neq y$  и всех  $\alpha \in (0, 1)$  в (3.16) выполняется строгое неравенство, то  $Q$  *строго выпукла*. При обратных знаках неравенств  $Q$  называется *вогнутой* и *строго вогнутой* или *выпуклой вверх* и *строго выпуклой вверх*.

Заметим, что выпуклость  $D$  требуется для того, чтобы точка  $\alpha x + (1 - \alpha)y$  не выходила из области определения функции  $Q$ .

Эквивалентная форма для (3.16):

$$Q(\underbrace{y + (x - y)\alpha}_{x^\alpha}) \leq \underbrace{Q(y) + (Q(x) - Q(y))\alpha}_{z^\alpha}.$$

В пространстве точка  $(x^\alpha, z^\alpha)$  при изменении  $\alpha$  от 0 до 1 пробегает от  $(y, Q(y))$  до  $(x, Q(x))$ . За счет линейного вхождения  $\alpha$  возникает отрезок, соединяющий две эти точки.

Таким образом, (3.16) означает, что в любом сечении функции  $Q$  вдоль  $[x, y]$  (т.е. вдоль любого отрезка в  $D$ ), график функции в этом сечении лежит всегда нестрого ниже отрезка прямой, соединяющей точки  $(x, Q(x))$  и  $(y, Q(y))$  (см. рис. 3.17).

На рисунке 3.18 приведен пример функции одной переменной, которая изменяет характер выпуклости при переходе от одного участка к другому.



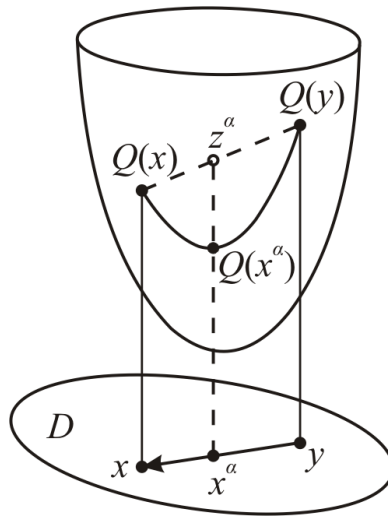


Рис. 3.17. Структура выпуклой функции

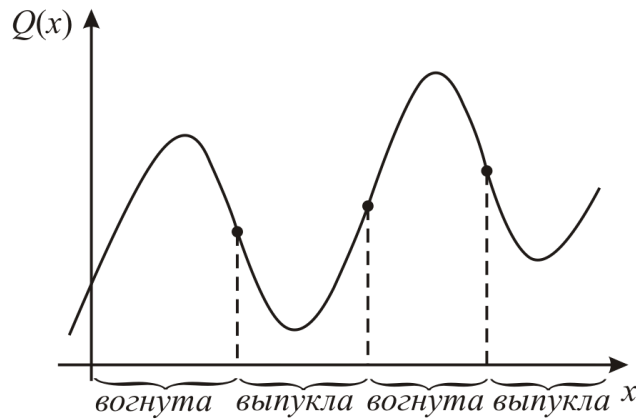


Рис. 3.18. Пример чередования характера выпуклости

### Некоторые свойства выпуклых функций

**Свойство 1.** Если  $Q(x)$  выпукла на выпуклом  $D$ , то  $Q(x)$  непрерывна в любой внутренней точке.

Приводится без доказательства.

**Замечание.** На границе  $D$  выпуклая функция может иметь разрывы.

На рис. 3.19 в качестве примера приведена выпуклая на  $D = \{x : x \geq 0\}$  функция

$$Q(x) = \begin{cases} x, & x > 0 \\ 1, & x = 0. \end{cases}$$

Она терпит разрыв в граничной точке 0 множества  $D$ .

**Свойство 2.** Если  $Q(x)$  — выпукла на выпуклом  $D$ , то в любой внутренней точке  $x$  множества  $D$  функция  $Q(x)$  имеет производные по любому направлению  $v$ .

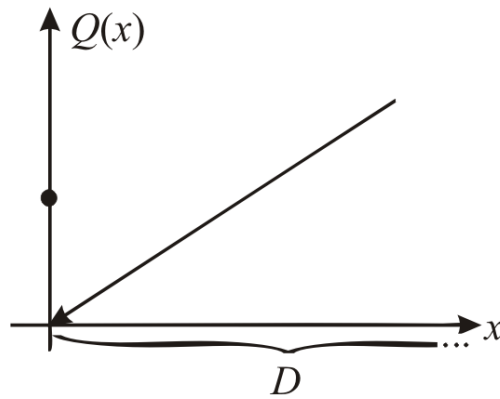


Рис. 3.19. Пример разрыва в граничной точке  $D$

Заметим, что существование производной в точке  $x$  по любому направлению не означает дифференцируемости функции в точке. На рис. 3.20 приведен пример выпуклой функции  $|x|$ . Ее производные в точке 0 по направлениям  $v_1, v_2$  равны 1, но функция в точке 0 не дифференцируема.

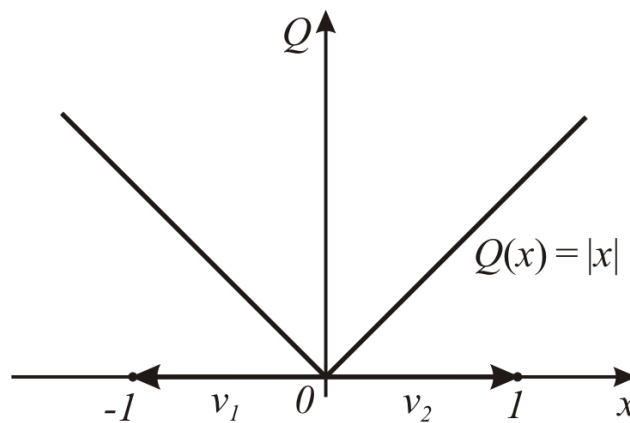


Рис. 3.20. В точке 0 существуют производные по обоим направлениям

*Доказательство.* Приведем его лишь схематично. Пусть  $\|v\| = 1$ . Введем функцию  $\varphi(\alpha)$ :

$$\varphi(\alpha) = \frac{Q(x + \alpha v) - Q(x)}{\alpha}.$$

Можно доказать, что при любых  $0 < \alpha_1 < \alpha_2$  верно  $\varphi(\alpha_1) \leq \varphi(\alpha_2)$ . Далее показываем, что для функций с таким свойством

$$\lim_{\alpha \rightarrow +0} \varphi(\alpha) = \inf_{\alpha \geq 0} \varphi(\alpha),$$

поскольку их определения в нашем случае совпадут. Но точная нижняя грань всегда существует, следовательно существует и нужный нам предел, а это и есть  $\frac{\partial Q(x)}{\partial v}$ . Полное доказательство следует провести самостоятельно.  $\square$

**Свойство 3.** Критерий выпуклости дифференцируемой функции

**Теорема.** Пусть  $Q(x)$  определена и дифференцируема на открытом, выпуклом множестве  $D$ . Функция  $Q$  выпукла на  $D$  тогда и только тогда, когда

$$\forall x, y \in D : \quad Q(y) + (\nabla Q(y), x - y) \leq Q(x). \quad (3.17)$$

Приведем геометрическую интерпретацию свойства. Заметим, что в случае дифференцируемости  $Q$  в точке  $y$  возможно представление:

$$Q(x) = Q(y) + (\nabla Q(y), x - y) + o(\|x - y\|).$$

Аффинная часть этого выражения является линейной аппроксимацией функции  $Q$  и совпадает с левой частью неравенства (3.17).

Возможный геометрический образ, соответствующий неравенству (3.17): поверхность  $z = Q(x)$ , порожденная функцией  $Q(x)$ , располагается не ниже

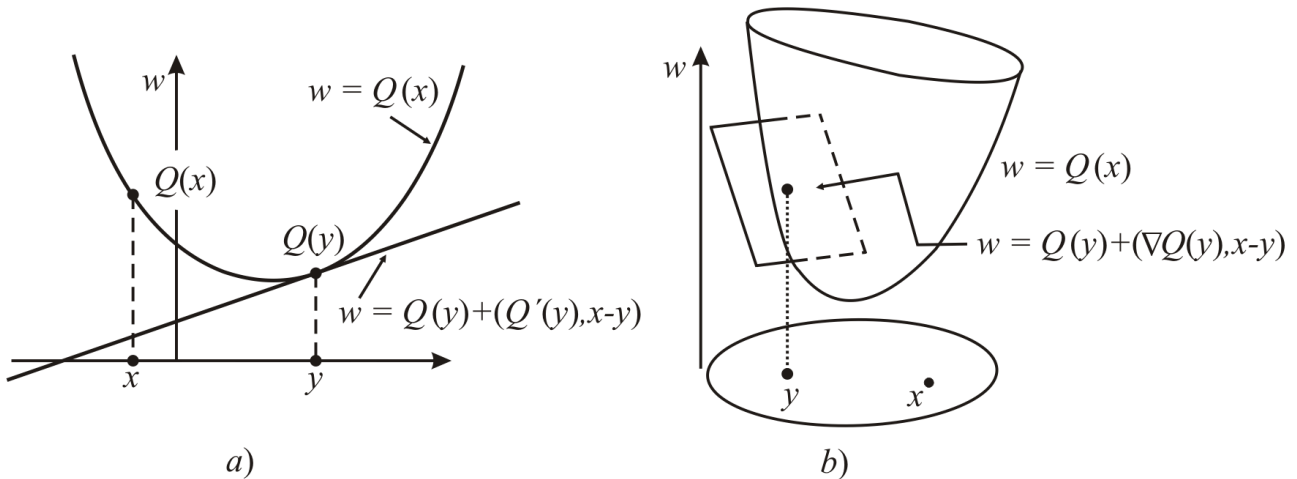


Рис. 3.21. Геометрическая интерпретация первого критерия выпуклости: а) — для одной переменной; б) — для  $N = 2$

касательной гиперплоскости, проведенной к поверхности  $W = Q(x)$  в произвольной точке  $(y; Q(y))$ . Касательная гиперплоскость порождена линейной аппроксимацией этой функции, построенной по локальным характеристикам функции  $Q$  в точке  $y$ :  $Q(y)$  и  $\nabla Q(y)$ .

*Доказательство.*

*Достаточность.* Пусть выполняется неравенство (3.17). Так как область  $D$  выпукла, то  $\forall x, y \in D$  и  $\forall \alpha \in [0, 1]$  вектор  $z = \alpha x + (1 - \alpha)y \in D$ .

Дважды запишем неравенство (3.17), используя вместо  $y, x$  пары  $z, x$  и  $z, y$ :

$$+ \begin{cases} Q(z) + (\nabla Q(z), x - z) \leq Q(x) & \times \alpha, \\ Q(z) + (\nabla Q(z), y - z) \leq Q(y) & \times (1 - \alpha). \end{cases}$$

Домножая первое неравенство на  $\alpha$ , второе на  $1 - \alpha$  и складывая их, получим:

$$Q(z) + \underbrace{(\nabla Q(z), z - z)}_{=0} \stackrel{=0}{\leq} \alpha Q(x) + (1 - \alpha)Q(y),$$

где  $z = \alpha x + (1 - \alpha)y$ . Это неравенство совпадает с определением выпуклой функции.

*Необходимость.* Пусть  $Q$  — выпуклая функция. По определению  $\forall x, y \in D, \forall \alpha \in [0, 1]$ :

$$Q(y + \alpha(x - y)) \leq Q(y) + \alpha(Q(x) - Q(y)).$$

Рассмотрим случай  $x \neq y$  (для  $x = y$  неравенство (3.17), очевидно, верно). Преобразуем так:

$$Q\left(y + \alpha\|x - y\| \frac{(x - y)}{\|x - y\|}\right) \leq Q(y) + \alpha\|x - y\| \frac{(Q(x) - Q(y))}{\|x - y\|}.$$

Введем обозначения:

$$v = \frac{(x - y)}{\|x - y\|}$$

и

$$\beta = \alpha\|x - y\|.$$

Тогда после несложных перестановок получим

$$\frac{Q(y + \beta v) - Q(y)}{\beta} \leq \frac{Q(x) - Q(y)}{\|x - y\|}.$$

Устремляя  $\beta$  к нулю, из последнего неравенства получим:

$$\frac{\partial Q(y)}{\partial v} \leq \frac{Q(x) - Q(y)}{\|x - y\|}. \quad (3.18)$$

Для дифференцируемых функций неравенство (3.18) эквивалентно запишется в виде:

$$\left( \nabla Q(y), \frac{x - y}{\|x - y\|} \right) \leq \frac{Q(x) - Q(y)}{\|x - y\|}$$

или

$$(\nabla Q(y), x - y) \leq Q(x) - Q(y).$$

□

Это свойство чрезвычайно важно с точки зрения построения численных методов. Частично это ясно из следствия.

**Следствие.** Пусть функция  $Q$  выпукла и дифференцируема в  $R^N$ , а градиент  $\nabla Q(y) \neq 0$ . Тогда глобальный минимум  $x^*$  функции  $Q$  не принадлежит множеству

$$G = \{x : (\nabla Q(y), x - y) > 0\}.$$

*Доказательство.* Заметим, что глобальный минимум может находиться в точке  $x$  только при выполнении неравенства  $Q(x) \leq Q(y)$ . Так как  $Q$  — выпуклая функция, то из неравенства (3.17):

$$Q(y) \leq Q(x) - (\nabla Q(y), x - y).$$

Учитывая оба неравенства и сокращая справа и слева на  $Q(x)$ , получим, что для точек  $x$ , которые могут быть глобально оптимальными:

$$0 \leq -(\nabla Q(y), x - y),$$

то есть  $(\nabla Q(y), x - y) \leq 0$ . Следовательно, все глобально оптимальные точки содержатся в этом замкнутом полупространстве. Тогда в дополняющем открытом полупространстве  $G$  глобальных минимумов нет, что и требовалось.  $\square$

Приведем геометрическую иллюстрацию. Пусть  $D$  — выпукло,  $Q$  — выпукла и дифференцируема в  $D$ . Будем проводить измерение градиента.

Измерение  $\nabla Q(y_1)$  позволяет отсечь заштрихованную часть (см. рис.3.22). Далее — измеряем  $\nabla Q(y_2)$  в геометрическом центре оставшегося множества  $D_1$ . Отброшенная часть помечена сеткой. Процесс продолжается. В итоге глобальный минимум — в сужающейся остающейся части.

На этой идее основан метод эллипсоидов А. С. Нимеровского и Д. Б. Юдина [26].

Если выпуклая функция не является дифференцируемой, то для нее можно ввести обобщение понятия градиента, используя аналог неравенства (3.17).

**Определение.** Пусть  $Q(x)$  — выпукла в  $R^N$ ,  $y \in R^N$ . Говорят, что  $v(y)$  является субградиентом  $Q$  в точке  $y$ , если  $\forall x \in R^N : Q(x) \geq Q(y) + (v(y), x - y)$ .

**Определение.** Субдифференциалом выпуклой функции  $Q$  в точке  $y$  называют множество  $\partial Q(y)$ , состоящее из всех субградиентов функции  $Q$  в этой точке.

**Пример.** Рассмотрим функцию:

$$Q(x_1, x_2) = 5|x_1 - x_2| + |x_2 - 1|.$$

Качественный вид линий уровня показан на рис. 3.23:

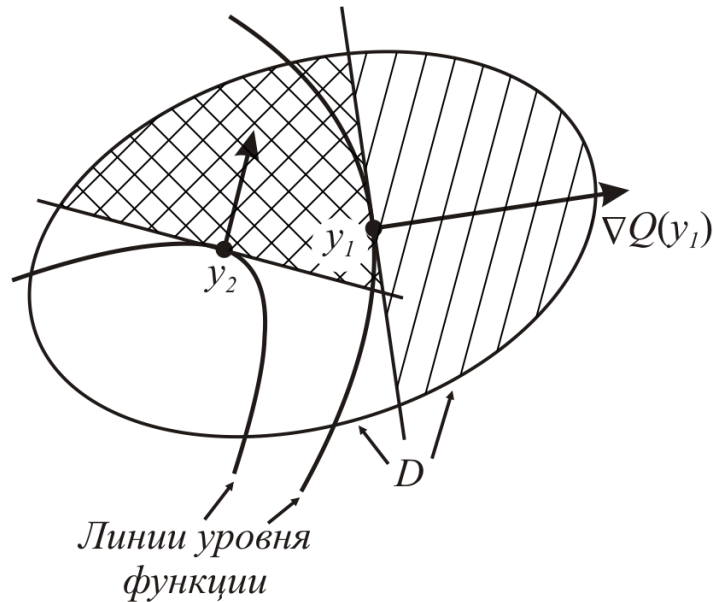


Рис. 3.22. Измерения градиента позволяют отсекал части области, не содержащие глобального минимума

Можно показать, что в точке 0 субдифференциал функции  $\partial Q(0)$  является конусом, натянутым на  $v_1$  и  $v_2$ . Эти вектора являются нормальными к гладким фрагментам границы области  $Q(x) < Q(0)$  в точке 0. Заметим, что субградиент — всегда направление возрастания, но при этом вектор обратный по знаку («антисубградиент») не обязан быть направлением убывания, что хорошо видно из рис. 3.23.

И еще известно, что в случае дифференцируемости функции её субдифференциал содержит единственный вектор, равный градиенту.

**Свойство 4.** Критерий выпуклости дважды непрерывно дифференцируемой функции

**Теорема.** Пусть  $Q$  определена и дважды непрерывно дифференцируема на некотором непустом, выпуклом, открытом множестве  $D$ . В этом случае  $Q(x)$  выпукла на  $D$  тогда и только тогда, когда  $\forall x \in D$  матрица вторых частных производных функции  $Q$  неотрицательно определена.

Матрицу вторых частных производных будем обозначать как

$$\nabla_{xx}^2 Q(x) = \begin{bmatrix} \frac{\partial^2 Q(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 Q(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 Q(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial^2 Q(x)}{\partial x_n \partial x_1} & \frac{\partial^2 Q(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 Q(x)}{\partial x_n \partial x_n} \end{bmatrix}.$$

Ее называют *матрицей Гессе* или *гессианом*.

Неотрицательную определенность обозначают с помощью неравенства:

$$\nabla_{xx}^2 Q(x) \geq 0.$$

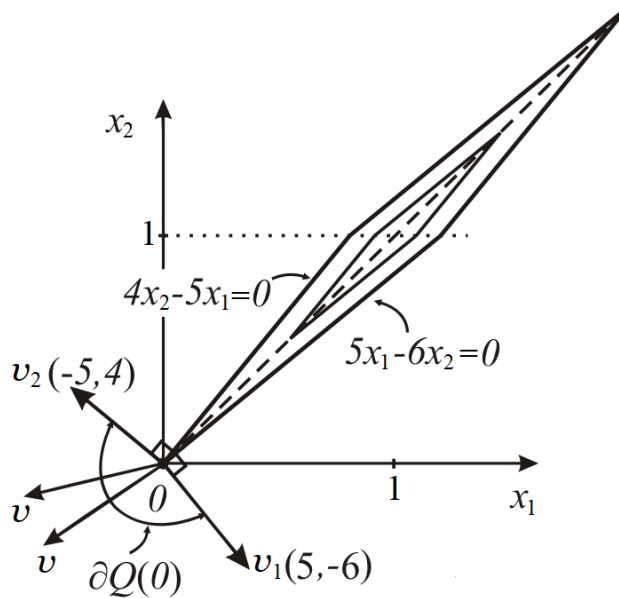


Рис. 3.23. Направления субградиентов для недифференцируемой функции

Заметим, что из непрерывности вторых производных вытекает симметрия гессиана:  $\nabla_{xx}^2 Q(x) = (\nabla_{xx}^2 Q(x))^T$ .

**Определение.** Пусть матрица  $A$  — симметричная (симметрическая). Матрицу  $A$  называют *неотрицательно определенной* ( $A \geq 0$ ), если  $\forall d : d^T A d \geq 0$ .

Говорят, что матрица  $A$  *положительно определена* ( $A > 0$ ), если  $\forall d \neq 0 : d^T A d > 0$ .

Говорят, что  $A$  *знаконеопределена*, если  $\exists d' : d'^T A d' > 0$ , а также  $\exists d'' : d''^T A d'' < 0$ .

Геометрические иллюстрации для перечисленных случаев ( $f(d) = d^T A d$  — *квадрика*) приведены на рис.3.24, где показаны виды поверхности  $w = f(d)$ .

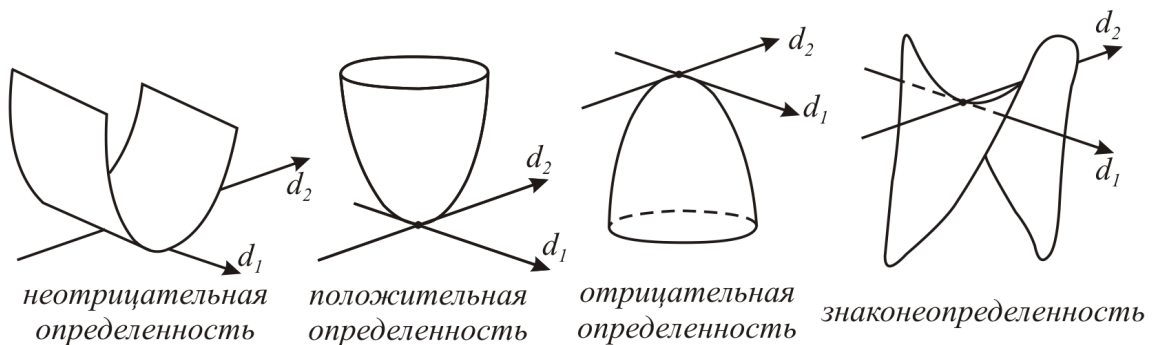


Рис. 3.24. Геометрическая иллюстрация знакоопределенности матрицы

Напомним способы анализа характера знакоопределенности:

- *Критерий Сильвестра*

*Симметричная матрица  $A$  положительно определена тогда и только тогда, когда все её главные миноры положительны:  $\Delta_1 > 0, \Delta_2 > 0,$*

$\dots, \Delta_N > 0$ . Симметричная матрица отрицательно определена тогда и только тогда, когда  $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \dots$ .

Заметим, что при  $\Delta_1 \geq 0, \dots, \Delta_N \geq 0$  про характер знакоопределенности сказать ничего нельзя, т.к. критерий неприменим. Пример этого приведен в учебно-методическом пособии [10].

- *По собственным числам*

У симметричной матрицы собственные числа  $\lambda_1, \lambda_2, \dots, \lambda_N$  — действительные. При неотрицательной определенности все  $\lambda_i \geq 0$ , при положительной — все  $\lambda_i > 0$ . При знакоотрицательности — все  $\lambda_i < 0$ . При знаконеопределенности:  $\exists \lambda_i < 0$  и  $\exists \lambda_i > 0$ .

*Доказательство критерия выпуклости функций  $Q \in C^2$ .*

*Достаточность.* Пусть матрица вторых частных производных  $\nabla_{xx}^2 Q(x)$  неотрицательно определена для любых  $x$ . Разложим функцию  $Q(x)$  в ряд Тейлора с остаточным членом:

$$Q(x) = Q(y) + (\nabla Q(y), x - y) + (x - y)^T \frac{\nabla_{xx}^2 Q(\xi)}{2} (x - y),$$

где  $\xi \in [x, y] \subseteq D$ .

Матрица  $\nabla_{xx}^2 Q(\xi)$  неотрицательно определена, поэтому

$$(x - y)^T \nabla_{xx}^2 Q(\xi) (x - y) \geq 0.$$

Следовательно, верно неравенство

$$Q(x) \geq Q(y) + (\nabla Q(y), x - y)$$

для  $\forall x \in D, \forall y \in D$ . По свойству 3 отсюда следует выпуклость.

*Необходимость.* Предположим, что  $Q$  — выпукла, но матрица  $\nabla_{xx}^2 Q$  не является неотрицательно определенной. Значит,  $\exists \tilde{y}, \exists d$ , что  $d^T \nabla_{xx}^2 Q(\tilde{y}) d < 0$ .

Так как элементы матрицы Гессе непрерывны, то знак неравенства сохранится и в некоторой окрестности  $O_\varepsilon(\tilde{y})$  точки  $\tilde{y}$ . Возьмем произвольное  $\tilde{x} = \tilde{y} + \gamma d$  из этой окрестности, где  $\gamma$  достаточно малое (см. рис. 3.25). Тогда, разложив  $Q(\tilde{x})$  в ряд Тейлора, имеем:

$$Q(\tilde{x}) = Q(\tilde{y}) + (\nabla Q(\tilde{y}), \tilde{x} - \tilde{y}) + (\tilde{x} - \tilde{y})^T \frac{\nabla_{xx}^2 Q(\eta)}{2} (\tilde{x} - \tilde{y}),$$

где  $\eta \in [\tilde{x}, \tilde{y}] \subseteq O_\varepsilon(\tilde{y})$ . Или

$$Q(\tilde{x}) = Q(\tilde{y}) + (\nabla Q(\tilde{y}), \tilde{x} - \tilde{y}) + \underbrace{\gamma^2 d^T \frac{\nabla_{xx}^2 Q(\eta)}{2} d}_{<0}.$$



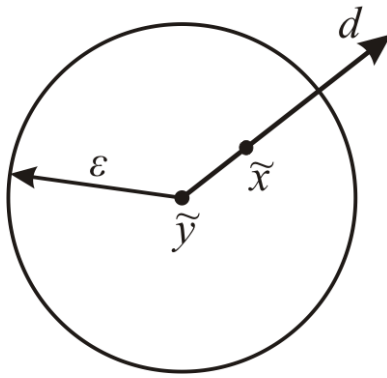


Рис. 3.25. Выбор точки  $\tilde{x} = \tilde{y} + \gamma d$

Откуда ясно, что

$$Q(\tilde{x}) < Q(\tilde{y}) + (\nabla Q(\tilde{y}), \tilde{x} - \tilde{y}).$$

Но это противоречит свойству 3 выпуклой функции. Таким образом гессиан всюду в  $D$  неотрицательно определен.  $\square$

**Свойство 5.** Пусть  $g(x)$  выпукла в  $R^N$ , а  $h(x)$  аффинна в  $R^N$  (т.е.  $h(x) = (a, x) + b$ ). Тогда множества  $G = \{x : g(x) \leq C\}$ ,  $H = \{x : h(x) = C\}$  выпуклы (доказать самостоятельно, см. рис.3.26).

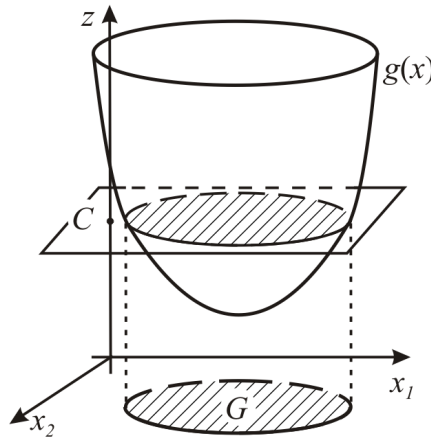


Рис. 3.26. Выпуклое ограничение–неравенство порождает выпуклое множество

**Определение.** Если функция  $Q(x)$  выпукла на выпуклом множестве  $D$ , то задача математического программирования вида

$$\min_{x \in D} Q(x)$$

называется *выпуклой задачей математического программирования* или *задачей выпуклого математического программирования*.

**Свойство 6.** О минимумах в выпуклых задачах математического программирования

**Теорема.** В любой выпуклой задаче математического программирования:

1. Любой локальный минимум является глобальным;
2. Множество глобальных минимумов выпукло (см. рис. 3.27);

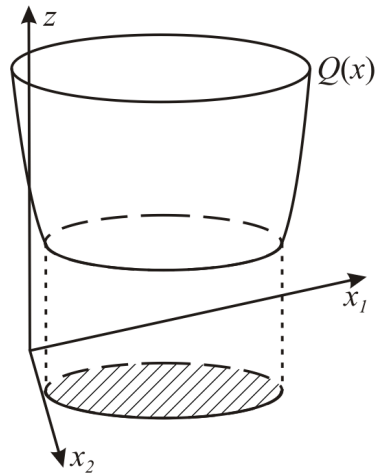


Рис. 3.27. Множество глобальных минимумов функции  $Q(x)$

3. Если  $Q$  строго выпукла, то глобальный минимум будет единственным.

**Задание.** Самостоятельно доказать утверждения теоремы.

### 3.4. Условия оптимальности для задач с ограничениями–равенствами и неравенствами

#### 3.4.1. Условия оптимальности в недифференциальной форме

Общую задачу математического программирования запишем в следующем виде:

$$\begin{aligned}
 & \min Q(x) \\
 & g_i(x) \leq 0 \quad (i = 1, \dots, m) \\
 & h_j(x) = 0 \quad (j = 1, \dots, p) \\
 & x \in E \subseteq R^N.
 \end{aligned}
 \tag{3.19}$$

Введем обозначение для допустимого множества:

$$D = \{x \in E : g(x) \leq 0, h(x) = 0\}.$$

#### Условия Каруша–Куна–Таккера в недифференциальной форме

1.  $x^o \in D$  — допустимость;
2.  $\exists (\lambda_0^*, \lambda^* \in R^m, \mu^* \in R^p) \neq 0$  — существование и нетривиальность;
3.  $\lambda_0^* \geq 0, \lambda^* \geq 0$  — неотрицательность;

4. Для функции Лагранжа вида

$$L(x, \lambda_0, \lambda, \mu) = \lambda_0 Q(x) + (g(x), \lambda) + (h(x), \mu)$$

выполняется:

$$L(x^o, \lambda_0^*, \lambda^*, \mu^*) = \min_{x \in E} L(x, \lambda_0^*, \lambda^*, \mu^*) \text{ — принцип минимума;}$$

5.  $\forall i = 1, \dots, m : \lambda_i^* g_i(x^o) = 0$  — условия дополняющей нежесткости.

Ниже будет показано, что эти условия при определенных требованиях к задаче будут являться необходимыми или же достаточными условиями глобального минимума в точке  $x^o$ . Заметим, что задача с функциональными ограничениями (3.19) свелась к условиям, в которых операция взятия минимума (условие из п.4) выполняется только по множеству  $E$ , а все функциональные ограничения оказались «спрятаны» в функцию Лагранжа.

Поясним смысл условия дополняющей нежесткости. Введем понятие *активного ограничения*.

**Определение.** Если ограничение–неравенство выполняется в точке  $x$  как равенство, то такое ограничение называют *активным* в точке  $x$ .

Множество номеров активных в точке  $x$  ограничений обозначим как  $I(x)$ :

$$I(x) = \{i \in \{1, \dots, m\} : g_i(x) = 0\}.$$

Условие из п.5 означает, что, если ограничение  $g_i$  неактивно в точке  $x^o$ , то оно не входит в выражение для функции  $L(x, \lambda_0^*, \lambda^*, \mu^*)$  в принципе минимума из п.4, поскольку из  $g_i(x^o) < 0$  в п.5 следует, что  $\lambda_i^* = 0$ .

Прежде чем формулировать теоремы об условиях оптимальности, приведем другую эквивалентную (при некоторых дополнительных требованиях) форму записи условий из п.п.1–5.

Договоримся далее обозначить эти условия из п.п.1–5 как ККТ(1–5).

### Условия ККТ в терминах седловой точки функции Лагранжа

1.  $x^o \in D$  — допустимость;
2.  $\exists (\lambda^*, \mu^*), \lambda_0^* = 1$  — существование и регулярность<sup>1</sup>;
3.  $\lambda^* \geq 0$  — неотрицательность;
4.  $\forall \lambda \geq 0, \forall \mu, \forall x \in E$ :

$$L(x^o, 1, \lambda, \mu) \leq L(x^o, 1, \lambda^*, \mu^*) \leq L(x, 1, \lambda^*, \mu^*) \quad (3.20)$$

— условие седловой точки  $(x^o, \lambda^*, \mu^*)$  для функции Лагранжа.

Договоримся обозначать условия в терминах седловой точки как ККТ–СТ(1–4).

<sup>1</sup>Общее определение регулярности допустимого множества дано в подразделе 3.2

**Теорема** (об эквивалентности). В основной задаче математического программирования при регулярности допустимого множества  $D$  условия ККТ (1–5) выполняются тогда и только тогда, когда выполняются условия ККТ–СТ(1–4).

*Доказательство.* При регулярности  $D$  коэффициент  $\lambda_0^* \neq 0$  и его можно принять равным 1. Тогда первые три условия эквивалентны. Условие ККТ 4 (принцип минимума) равносильно выполнению правой части неравенства (3.20)  $\forall x \in E$ . Остается показать, что условие ККТ 5 выполняется тогда и только тогда, когда  $\forall \lambda \geq 0, \forall \mu$  выполняется левая часть неравенства (3.20).

Вначале перепишем его явно:

$$Q(x^o) + (\lambda, g(x^o)) + \underbrace{(\mu, h(x^o))}_{=0} \leq Q(x^o) + (\lambda^*, g(x^o)) + \underbrace{(\mu^*, h(x^o))}_{=0}.$$

Следовательно,  $\forall \lambda \geq 0$  выполняется:

$$(g(x^o), \lambda) \leq (g(x^o), \lambda^*). \quad (3.21)$$

Пусть условие ККТ 5 верно. Это значит, что  $(g(x^o), \lambda^*) = 0$  и следовательно справедливо неравенство

$$(g(x^o), \lambda) \stackrel{\leq 0}{\leq} 0 \stackrel{\geq 0}{=} (g(x^o), \lambda^*),$$

которое совпадает с (3.21). Следовательно, левое неравенство в (3.20) верно. Пусть теперь верна левая часть неравенства (3.20), т.е. выполняется (3.21).

Пусть  $\lambda \rightarrow 0$ , тогда из (3.21) получаем:

$$0 \leq (g(x^o), \lambda^*) \stackrel{\geq 0}{\leq} 0.$$

Отсюда

$$\sum_{i=1}^m g_i(x^o) \lambda_i^* = 0,$$

где каждое слагаемое неположительное. Следовательно,  $\forall i = 1, \dots, m : \lambda_i^* g_i(x^o) = 0$ . □

Теперь докажем несколько важных теорем.

**Теорема** (о достаточности условий ККТ). В случае, когда в задаче математического программирования (3.19) допустимое множество регулярно в точке  $x^o$  (т.е.  $\lambda_0^* \neq 0$ ), выполнение условий ККТ(1–5) или эквивалентных им условий ККТ–СТ(1–4) достаточно для глобальной оптимальности точки  $x^o$ .

*Доказательство.* Пусть выполнено ККТ(1–5), но утверждение теоремы неверно. Тогда  $\exists \bar{x} \in D : Q(\bar{x}) < Q(x^o)$ . Оценим:

$$L(\bar{x}, \lambda_0^*, \lambda^*, \mu^*) = \lambda_0^* Q(\bar{x}) + \overset{\leq 0}{(g(\bar{x}), \lambda^*)} + \overset{\geq 0}{(h(\bar{x}), \mu^*)} \overset{=0}{\leq} \lambda_0^* Q(\bar{x}) \underset{\text{т.к. } \lambda_0^* > 0}{<} \\ \underset{\text{т.к. } \lambda_0^* > 0}{<} \lambda_0^* Q(x^o) = \lambda_0^* Q(x^o) + \underbrace{(g(x^o), \lambda^*)}_{=0 \text{ по усл. 5}} + \underset{=0}{(h(x^o), \mu^*)} = L(x^o, \lambda_0^*, \lambda^*, \mu^*).$$

Таким образом,  $L(\bar{x}, \lambda_0^*, \lambda^*, \mu^*) < L(x^o, \lambda_0^*, \lambda^*, \mu^*)$ , что противоречит принципу минимума 4. Значит, утверждение теоремы нарушаться не может и  $x^o$  — глобальный минимум в задаче.  $\square$

Итак, при регулярности допустимого множества условия ККТ и ККТ–СТ эквивалентны и достаточны для глобальной оптимальности точки  $x^o$ . Беда заключается в том, что для многих задач они могут и не выполняться, т.к. не являются необходимыми.

Рассмотрим теперь класс задач, где они необходимы! Это будут выпуклые задачи.

**Теорема** (Каруша–Куна–Таккера в недифференциальной форме). *Если в задаче (3.19) множество  $E$  выпукло,  $Q(x)$ ,  $g_i(x)$  ( $i = \overline{1, m}$ ) — выпуклы на множестве  $E$ , а  $h_i(x)$  ( $i = \overline{1, p}$ ) — аффинные, то:*

1. *Условия Каруша–Куна–Таккера (1–5) являются необходимыми условиями глобальной оптимальности точки  $x^o$  в задаче (3.19);*
2. *В случае регулярности (при  $\lambda_0^* \neq 0$ ) эти условия являются также достаточными для глобальной оптимальности точки  $x^o$  в задаче (3.19).*

**Замечание.** *В условиях рассмотренной теоремы при регулярности допустимого множества глобальный минимум функции  $Q$  на  $D$  можно определить путем отыскания седловой точки функции Лагранжа по переменным  $x \in E$ ,  $\lambda \geq 0$ ,  $\mu \in R^p$ . Это вытекает из теоремы об эквивалентности.*

*Доказательство теоремы.*

Достаточность (пункт 2) уже была показана.

Докажем пункт 1. Его обоснование сведется к теореме об отделимости. Вначале построим множество и точку, которые будем разделять.

1. Возьмем произвольное  $x \in E$ . Для него построим множество

$$S(x) = \{(z_Q, z_g, z_h) : Q(x) \leq z_Q, g(x) \leq z_g, h(x) = z_h\}.$$

Наконец введем множество  $S = \bigcup_{x \in E} S(x)$ .

Теперь построим точку  $P = (Q^*, 0, 0) \in R^{1+m+p}$ , где  $Q^* = Q(x^o)$  — глобально оптимальное значение.

Структуру множеств  $S$  и  $S(x)$  можно проиллюстрировать на примере, но только для упрощенной ситуации, когда ограничений–равенств нет. Тогда в  $S(x)$  будут входить пары  $(z_Q, z_g)$ . Пусть  $N = 1$ ,  $m = 1$ ,  $p = 0$ . На рис. приведен пример конкретного вида функций  $Q(x)$  и  $g_1(x)$  и  $E$ , а также вид соответствующего множества  $S$ .

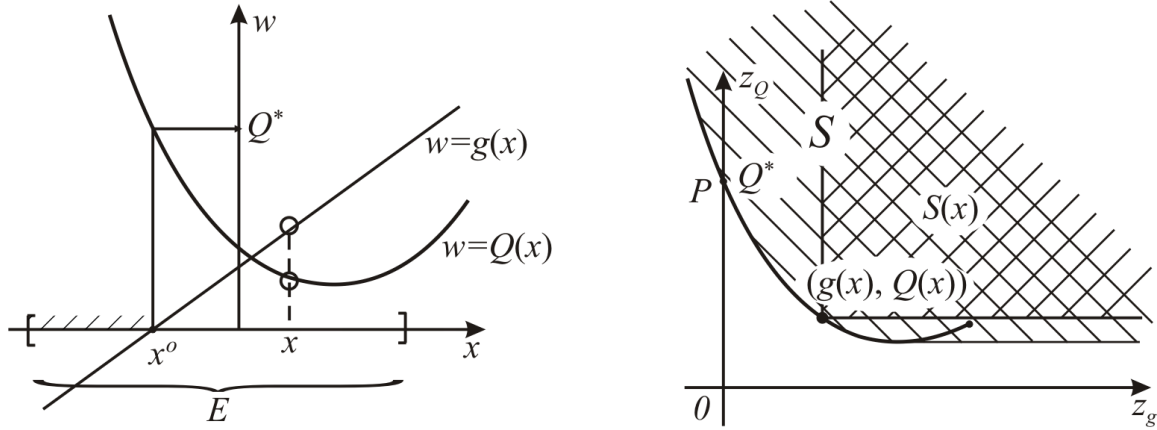


Рис. 3.28. Структура множеств  $S(x)$  и  $S$  для конкретной задачи

2. Докажем выпуклость  $S$ . Пусть  $z' = (z'_Q, z'_g, z'_h) \in S$ . Тогда существует точка  $x' \in E$ :  $z' \in S(x')$  и, следовательно,  $Q(x') \leq z'_Q$ ,  $g(x') \leq z'_g$ ,  $h(x') = z'_h$ . Пусть есть еще одна точка  $z'' = (z''_Q, z''_g, z''_h) \in S$ . Тогда существует точка  $x'' \in E$ :  $z'' \in S(x'')$  и, следовательно,  $Q(x'') \leq z''_Q$ ,  $g(x'') \leq z''_g$ ,  $h(x'') = z''_h$ .

Возьмем произвольное число  $\alpha \in [0, 1]$  и построим точку

$$z^\alpha = \alpha z' + (1 - \alpha) z''.$$

Если эта точка принадлежит множеству  $S$ , то оно выпукло по определению. Докажем это.

Так как множество  $E$  выпукло, то точка  $x^\alpha = \alpha x' + (1 - \alpha) x'' \in E$ . В силу выпуклости функции  $Q$  верно неравенство:

$$Q(x^\alpha) \leq \alpha Q(x') + (1 - \alpha) Q(x'') \leq \alpha z'_Q + (1 - \alpha) z''_Q = z^\alpha_Q.$$

Точно также можно получить, что  $g(x^\alpha) \leq z^\alpha_g$  и  $h(x^\alpha) = z^\alpha_h$ . Значит точка  $z^\alpha \in S(x^\alpha) \subseteq S$  и  $S$  — выпуклое множество.

3. Покажем, что  $P$  — граничная точка множества  $S$ .

Сначала покажем, что точка  $P \in S$ , а именно, что  $P \in S(x^0)$ .

Так как  $x^0$  — глобальный минимум, то  $x^0 \in E$  и является допустимой точкой. Следовательно выполняются условия

$$\begin{cases} Q(x^0) = Q^*, \\ g(x^0) \leq 0, \\ h(x^0) = 0. \end{cases}$$

Откуда следует, что  $P \in S(x^o) \subseteq S$ .

Теперь покажем, что  $P$  является граничной точкой. Сдвинем  $P$  на сколь угодно малое  $\varepsilon > 0$ . Покажем, что точка

$$P_\varepsilon = (Q^* - \varepsilon, 0, 0)$$

не принадлежит  $S$ .

Пусть это не так. Тогда существует  $\tilde{x} \in E : P_\varepsilon \in S(\tilde{x})$ . Но в этом случае  $g(\tilde{x}) \leq 0$ ,  $h(\tilde{x}) = 0$  (следовательно,  $\tilde{x}$  — допустима) и кроме того:

$$Q(\tilde{x}) \leq Q^* - \varepsilon < Q^*.$$

Поскольку значение  $Q(x)$  в допустимой точке не может быть меньше глобально-оптимального, то получаем противоречие.

Таким образом  $\forall \varepsilon > 0 : P_\varepsilon \notin S$ , т.е.  $P$  — граничная точка  $S$  (в любой окрестности  $P$  есть точки принадлежащие  $S$  (сама точка  $P$ ) и не принадлежащие  $S$ ).

4. Из второй теоремы об отделимости следует, что множество  $S$  и точка  $P$  отделимы. По определению отделимости найдется ненулевой вектор  $(\lambda_0^*, \lambda^*, \mu^*)$ , что для любой точки  $(z_Q, z_g, z_h) \in S$ :

$$\lambda_0^* z_Q + (\lambda^*, z_g) + (\mu^*, z_h) \geq \lambda_0^* Q^* + (\lambda^*, 0) + (\mu^*, 0).$$

Заменяя  $z_Q, z_g, z_h$  на их явные выражения, запишем полученное условие иначе:

$$\exists (\lambda_0^*, \lambda^*, \mu^*) \neq 0 : \forall x \in E, \forall \varepsilon \geq 0, \forall \delta \geq 0$$

выполняется неравенство

$$\lambda_0^*(Q(x) + \varepsilon) + (\lambda^*, g(x) + \delta) + (\mu^*, h(x)) \geq \lambda_0^* Q^*. \quad (3.22)$$

5. Докажем выполнение условий Каруша–Куна–Таккера.

Условия ККТ 1 и 2 уже выполняются.

А). Проверим условие  $\lambda_0^* \geq 0$ .

В силу того, что неравенство (3.22) верно для любых  $x \in E$ ,  $\varepsilon \geq 0$  и  $\delta \geq 0$ , то возьмем  $x = x^o$ ,  $\delta = -g(x^o)$ ,  $\varepsilon > 0$ ; получим:

$$\lambda_0^* Q(x^o) + \lambda_0^* \varepsilon + 0 + 0 \geq \lambda_0^* Q(x^o),$$

откуда:

$$\lambda_0^* \varepsilon \geq 0,$$

следовательно  $\lambda_0^* \geq 0$ .

В). Покажем, что верно условие  $\lambda^* \geq 0$ .

Возьмем  $x = x^o, \varepsilon = 0$ ,

$$\delta_j = \begin{cases} -g_j(x^o), & j \neq i, \\ -g_i(x^o) + 1, & j = i. \end{cases}$$

и подставим эти переменные в неравенство (3.22):

$$\lambda_0^* Q(x^o) + \lambda_i^* + 0 \geq \lambda_0^* Q(x^o).$$

Отсюда видно, что  $\lambda_i^* \geq 0$ . При этом  $i$  было выбрано произвольно. Таким образом, условие ККТ 3 Каруша–Куна–Таккера выполнено.

С). Проверим, что верно также условие ККТ 5.

Снова воспользуемся неравенством (3.22). Пусть  $x = x^o, \delta = 0, \varepsilon = 0$ :

$$\lambda_0^* Q(x^o) + (\lambda^*, g(x^o)) + 0 \geq \lambda_0^* Q(x^o),$$

откуда

$$(\lambda^*, g(x^o)) = \sum_{i=1}^m \lambda_i^* g_i(x^o) \geq 0.$$

Но, так как  $\forall i = \overline{1, m} : g_i(x^o) \leq 0$ , а  $\lambda_i^* \geq 0$ , то скалярное произведение  $\sum_{i=1}^m \lambda_i^* g_i(x^o) \leq 0$ . Следовательно,

$$\sum_{i=1}^m \underbrace{\lambda_i^* g_i(x^o)}_{\substack{\geq 0 \\ \leq 0}} = 0.$$

Каждое слагаемое суммы неположительно, поэтому из равенства нулю суммы следует, что каждое слагаемое нулевое, т.е.  $\lambda_i^* g_i(x^o) = 0$  ( $\forall i = 1, m$ ).

D). Остается проверить условие ККТ 4.

Для произвольного  $x$  из  $E, \varepsilon = 0, \delta = 0$  выполняется неравенство (3.22):

$$\lambda_0^* Q(x) + (\lambda^*, g(x)) + (\mu^*, h(x)) \geq \lambda_0^* Q(x^*) + \underbrace{(\lambda^*, g(x^*))}_{=0} + \underbrace{(\mu^*, h(x^*))}_{=0},$$

следовательно  $\forall x \in E$ :

$$L(x, \lambda_0^*, \lambda^*, \mu^*) \geq L(x^*, \lambda_0^*, \lambda^*, \mu^*).$$

Значит, условие ККТ 4 тоже выполнено.

□



**Теорема** (достаточное условие Слейтера регулярности допустимого множества). Пусть рассматривается класс задач

$$\min_{x \in D} Q(x),$$

где  $D = \{x \in E : g(x) \leq 0, h(x) = 0\}$ . Причем множество  $E$  выпукло,  $Q(x)$ ,  $g_i(x)$  ( $i = 1, \dots, m$ ) — выпуклы на  $E$ ,  $h_j(x)$  ( $\forall j = 1, \dots, p$ ) — аффинны. Тогда, если существует  $\bar{x} \in E$ , что  $h(\bar{x}) = 0$ ,  $g(\bar{x}) < 0$  и нулевой вектор размерности  $p$  — внутренняя точка множества  $h(E)$ , т.е.  $0 \in \text{int } h(E)$ , то  $D$  — регулярно.

*Доказательство.* Предположим, что допустимое множество  $D$  не регулярно. Тогда найдется такая выпуклая на  $D$  функция  $Q(x)$ , которая имеет глобальный минимум в точке  $\tilde{x}$  из  $D$ , но условия Каруша–Куна–Таккера для нее выполняются в  $\tilde{x}$  при  $\lambda_0^* = 0$ . В этом случае справедливы оценки:

$$\begin{aligned} L(\tilde{x}, \lambda_0^*, \lambda^*, \mu^*) &= 0 \cdot Q(\tilde{x}) + (\lambda^*, g(\tilde{x})) + (\mu^*, h(\tilde{x})) = 0 \leq \\ &\leq L(\bar{x}, \lambda_0^*, \lambda^*, \mu^*) = 0 \cdot Q(\bar{x}) + (\lambda^*, g(\bar{x})) + (\mu^*, h(\bar{x})). \end{aligned} \quad (3.23)$$

Неравенство здесь вытекает из принципа минимума. Отсюда

$$0 \leq (\lambda^*, g(\bar{x})) + (\mu^*, h(\bar{x})). \quad (3.24)$$

Так как  $\lambda_0^* = 0$ , но по условиям  $(\lambda_0^*, \lambda^*, \mu^*) \neq 0$ , то  $(\lambda^*, \mu^*) \neq 0$ . Рассмотрим два случая.

1. Существует номер  $i = 1, \dots, m$ , что  $\lambda_i^* > 0$ . Тогда из (3.24) следует, что

$$0 \leq \underbrace{(\lambda^*, g(\bar{x}))}_{<0} + \underbrace{(\mu^*, h(\bar{x}))}_{=0}.$$

Откуда получаем неверное неравенство  $0 < 0$ .

2. Пусть  $\lambda^* = 0$ . Тогда  $\mu^* \neq 0$ . Поскольку в (3.23) неравенство будет выполнено не только для  $\bar{x}$ , но и для любых  $x \in E$ , то

$$0 \leq (\lambda^*, g(x)) + (\mu^*, h(x)),$$

а поскольку  $\lambda^* = 0$ , то при  $\mu^* \neq 0$

$$0 \leq (\mu^*, h(x)).$$

Но всегда (за счет того, что  $0$  внутренняя точка множества  $h(E)$ ) можно подобрать такое  $x' \in E$ , что знаки  $h_i(x')$  будут обратны знакам ненулевых элементов  $\mu_i^*$ , поэтому получим  $(\mu^*, h(x)) < 0$  при  $x = x'$  и придем к противоречию вида  $0 < 0$ .

□

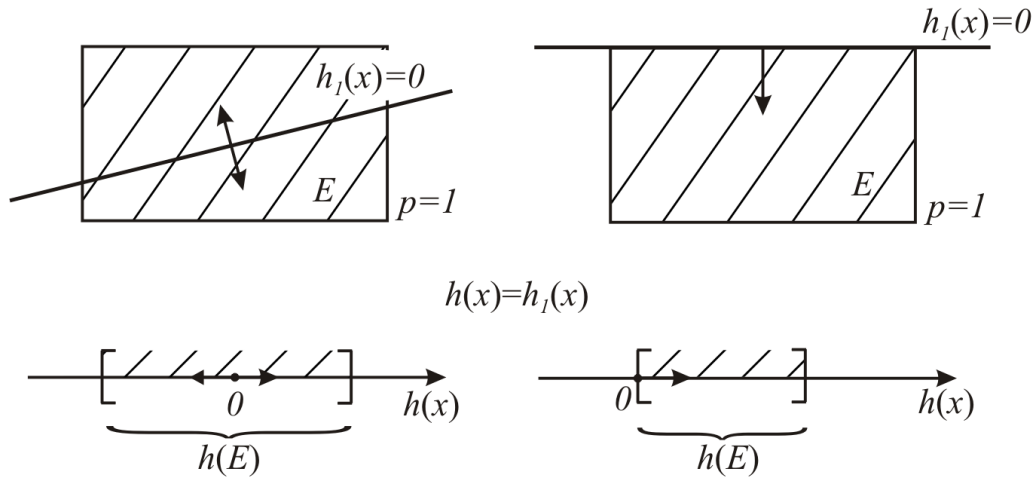


Рис. 3.29. Ситуации выполнения (слева) и нарушения (справа) условия  $0 \in \text{int } h(E)$  и виды множеств  $h(E)$  для этих случаев

На рис. 3.29 приведены примеры ситуаций, когда условие  $0 \in \text{int } h(E)$  выполнено и когда оно нарушено. Нарушение связано с тем, что фрагмент границы множества  $E$  для случая, показанного в правой части рис. 3.29, лежит на линейном многообразии, порожденном аффинными ограничениями. В последнем случае линейное многообразие, порожденное ограничением  $h_1(x) = 0$ , может быть смещено в пределах  $E$  только в одном направлении. Это значит, что, перебирая элементы  $x$  из  $E$ , мы будем получать значения  $h(x)$  одного знака, например, положительные, а также значение 0. Тем самым 0 окажется на границе  $h(E)$  (см. правую часть рис. 3.29).

### 3.4.2. Условия оптимальности в дифференциальной форме

#### Условия первого порядка

**Теорема** (Каруша–Куна–Таккера для дифференцируемого выпуклого случая). *Рассмотрим задачу*

$$\min_{x \in D} Q(x),$$

где  $D = \{x \in R^N : g(x) \leq 0, h(x) = 0\}$ . При этом  $Q, g_i, (i = 1, \dots, m)$  — выпуклы в  $R^N$  и дифференцируемы. Ограничения  $h_j(x) (j = 1, \dots, p)$  — аффинны, множество  $D$  — регулярно (например, выполняются условия Слейтера:  $\exists \bar{x} \in R^N : g(\bar{x}) < 0, h(\bar{x}) = 0$ ). Тогда для глобальной оптимальности точки  $x^*$  необходимо и достаточно выполнение следующих условий:

1.  $g(x^*) \leq 0, h(x^*) = 0$  — допустимость;
2.  $\exists (1, \lambda^*, \mu^*)$  — существование множителей Лагранжа при  $\lambda_0^* = 1$ ;
3.  $\lambda^* \geq 0$  — неотрицательность;
4.  $\nabla_x L(x^*, 1, \lambda^*, \mu^*) = 0$  — стационарность;
5.  $\forall i = 1, \dots, m : \lambda_i^* g_i(x^*) = 0$  — условие дополняющей нежесткости.

Заметим, что вместо условий из п.4 и п.5 достаточно потребовать выполнения одного условия 4':

$$4'. \quad -\nabla Q(x^*) = \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^*),$$

где  $I(x^*) = \{i \in \{1, \dots, m\} : g_i(x^*) = 0\}$ .

Множество  $I(x^*)$  — множество номеров активных в точке  $x^*$  ограничений. Напомним, что ограничение  $g_i(x) \leq 0$  активно в некоторой точке  $\bar{x}$ , если оно принимает в этой точке нулевое значение, т.е.  $g_i(\bar{x}) = 0$ .

*Доказательство.* Покажем, что условие 4 равносильно условию 4 Каруша–Куна–Таккера в недифференциальной форме (принципу минимума):

$$L(x^*, \lambda_0^*, \lambda^*, \mu^*) = \min_{x \in E} L(x, \lambda_0^*, \lambda^*, \mu^*)$$

при  $E = R^N$ .

А). Пусть выполняется принцип минимума функции Лагранжа. В рассматриваемой задаче  $\lambda_0^* = 1$ ,  $L(x, 1, \lambda^*, \mu^*)$  — дифференцируема по  $x$  и  $E = R^N$ . Поэтому в точке минимума функции Лагранжа по  $x$  ее градиент по этим переменным обращается в ноль:  $\nabla_x L(x^*, 1, \lambda^*, \mu^*) = 0$ .

В). Докажем теперь обратное. Пусть верно 4. Функция Лагранжа  $L(x, 1, \lambda^*, \mu^*)$  — выпуклая и дифференцируемая, как сумма выпуклых, дифференцируемых функций с неотрицательными коэффициентами. В этом случае по первому критерию выпуклости:

$$\forall x \in R^N : L(x, 1, \lambda^*, \mu^*) \geq L(x^*, 1, \lambda^*, \mu^*) + (\nabla_x L(x^*, 1, \lambda^*, \mu^*), x - x^*).$$

Однако градиент функции Лагранжа в точке  $x^*$  равен 0. Из оставшегося неравенства следует выполнение принципа минимума.  $\square$

Мы получили необходимые и достаточные условия глобального минимума для выпуклой задачи с регулярным допустимым множеством в случае дифференцируемости. Можно свести вышперечисленные условия в систему уравнений и неравенств для нахождения этого минимума. Существуют численные методы решения задач с ограничениями, основанные на численном решении таких систем. Методы этой группы называют *методами Лагранжа*. В некоторых численных методах невязку в системе условий оптимальности 1–5 используют для определения близости оценки решения  $\hat{x}^*$  к истинному условному минимуму  $x^*$ .

Есть два варианта записи системы.

Рассмотрим первый вариант. Составим систему из  $N + p + m$  уравнений и  $2m$  неравенств:

$$\begin{cases} \nabla_x L(x^*, 1, \lambda^*, \mu^*) = 0, \\ h(x^*) = 0, \\ \lambda_i^* g_i(x^*) = 0 \quad (i = 1, \dots, m), \\ \lambda^* \geq 0, \\ g(x^*) \leq 0. \end{cases}$$

Неизвестные  $\lambda^*$ ,  $\mu^*$  и  $x^*$  находят численным решением этой системы.

Для относительно простых задач можно найти глобальный минимум и аналитически. Для этого желательно избавиться от существенно нелинейных условий дополняющей нежесткости. Составим систему из  $N + p + r$  уравнений, где неизвестное нам множество номеров активных в точке  $x^*$  ограничений–неравенств  $I(x^*) = J$ . Получим:

$$\begin{cases} -\nabla Q(x^*) = \sum_{i \in J} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^*), \\ h(x^*) = 0, \\ g_i(x^*) = 0, \quad i \in J = \{i_1, i_2, \dots, i_r\}. \end{cases} \quad (3.25)$$

Получена замкнутая система с  $N + p + r$  неизвестными  $(x^*, \lambda_{i_1}^*, \dots, \lambda_{i_r}^*, \mu^*)$ . Если она нелинейная, то может иметь несколько решений. Если решений нет, то множество  $J$  выбрано неверно. Если решение найдено, для проверки оставшихся условий оптимальности необходимо выяснить выполнение следующих неравенств:

$$\begin{cases} g_i(x^*) \leq 0, \quad i \notin J, \\ \lambda_{i_s}^* \geq 0, \quad i_s \in J. \end{cases}$$

Если оказалось, что  $\exists i : g_i(x^*) > 0$ , то необходимо скорректировать гипотезу  $J$ , приняв  $J := J \cup \{i\}$ , и повторно составить систему (3.25). Если  $\exists \lambda_{i_s}^* < 0$ , то можно показать, что следует исключить  $i_s$ -ое ограничение из списка  $J$ , приняв  $J := J \setminus \{i_s\}$ , и решить новую систему вида (3.25) еще раз.

### Геометрическая интерпретация условий оптимальности

Приведем геометрическую интерпретацию выполнения основного градиентного соотношения в системе условий (3.25), а также ситуацию появления  $\lambda_{i_s}^* < 0$ . Рассмотрим упрощенный случай, когда ограничения–равенства отсутствуют, т.е.  $p = 0$ ,  $N = 2$ , и число  $r$  активных в точке  $x^*$  ограничений равно двум ( $J = I(x^*) = \{i_1, i_2\}$ ). Основное градиентное соотношение в (3.25) при этом приобретает вид

$$-\nabla Q(x^*) = \lambda_{i_1}^* \nabla g_{i_1}(x^*) + \lambda_{i_2}^* \nabla g_{i_2}(x^*), \quad (3.26)$$

где  $\nabla g_{i_1}(x^*)$  и  $\nabla g_{i_2}(x^*)$  являются внешними нормальными к активным в  $x^*$  фрагментам границы допустимого множества. При выполнении условия  $\lambda_{i_1}^* \geq 0$ ,  $\lambda_{i_2}^* \geq 0$  вектор  $-\nabla Q(x^*)$  размещается в конусе, натянутом на векторы  $\nabla g_{i_1}(x^*)$  и  $\nabla g_{i_2}(x^*)$  (рис. 3.30). Именно только при такой его ориентации точ-

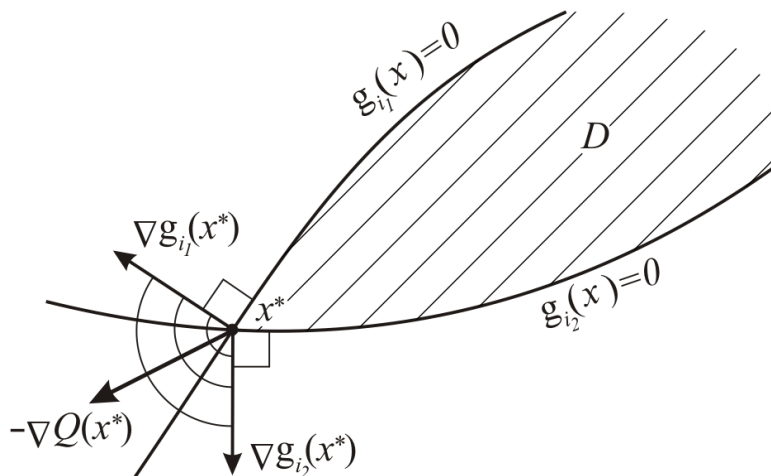


Рис. 3.30. Ориентация антиградиента при выполнении условий оптимальности

ка  $x^*$  может быть минимумом в задаче.

Действительно, направления  $\nu$ , в которых  $Q(x)$  строго локально убывает в точке  $x^*$ , образуют открытое полупространство, граница которого ортогональна вектору  $-\nabla Q(x^*)$  (рис. 3.31). Очевидно, пока антиградиент остается в указанном конусе, ни один из векторов  $\nu$  строгого локального убывания не входит в допустимое множество. Если же антиградиент выходит, напри-

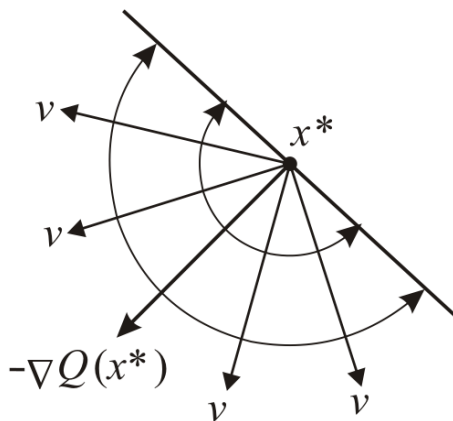


Рис. 3.31. Конус направлений строгого локального убывания функции  $Q$  в точке  $x^*$

мер, за вектор  $\nabla g_{i_1}(x^*)$  (см. рис. 3.32), то в соотношении (3.26) нарушится правильный знак у  $\lambda_{i_2}^*$  (окажется  $\lambda_{i_2}^* < 0$ ).

При этом возникнет пересечение полупространства направлений  $\nu$  строгого локального убывания с допустимым множеством. В частности, малое смещение из точки  $x^*$  вдоль границы  $g_{i_1}(x) = 0$  ограничения  $g_{i_1}(x) \leq 0$  (с

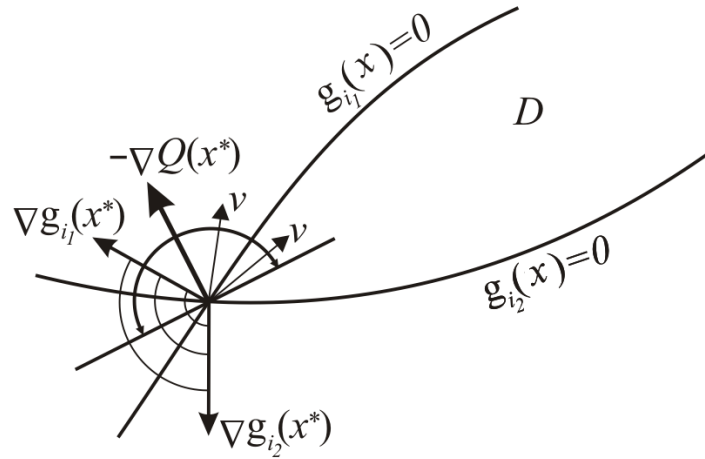


Рис. 3.32. Нарушение условий оптимальности:  $\lambda_{i_2}^* < 0$  — смещение вдоль  $g_{i_1}(x) = 0$  с уходом в область  $g_{i_2}(x) < 0$  уменьшает  $Q$

уходом с границы  $g_{i_2}(x) = 0$ ) приведет к уменьшению значения функции  $Q(x)$  по сравнению с точкой  $x^*$ . Именно этим обусловлено правило исключения из активного набора  $J$  того активного неравенства, которое породило  $\lambda_i^* < 0$ .

**Теорема** (Каруша–Куна–Таккера для невыпуклого случая в дифференциальной форме). Пусть в задаче

$$\min_{\substack{g(x) \leq 0 \\ h(x) = 0}} Q(x),$$

функции  $Q$  и  $g$  дифференцируемы в точке  $x^o$ , а  $h(x)$  непрерывно дифференцируемы в некоторой окрестности точки  $x^o$ . Тогда для того, чтобы  $x^o$  была локальным минимумом необходимо выполнение условий:

- 1)  $g(x^o) \leq 0, h(x^o) = 0$  — допустимость;
- 2)  $\exists (\lambda_0^*, \lambda^*, \mu^*) \neq 0$  — существование и нетривиальность;
- 3)  $\lambda_0^* \geq 0, \lambda^* \geq 0$  — неотрицательность;
- 4)  $\nabla_x L(x^o, \lambda_0^*, \lambda^*, \mu^*) = 0$  — стационарность;
- 5)  $\forall i = 1, \dots, m : \lambda_i^* g_i(x^o) = 0$  — условия дополняющей нежесткости.

*Доказательство.* Допустимость точки  $x^o$  очевидна. Докажем остальное.

А) Вначале исключим тривиальный случай, когда набор градиентов

$$\nabla h_1(x^o), \nabla h_2(x^o), \dots, \nabla h_p(x^o)$$

линейно зависимый.

Тогда существует  $\mu^* \neq 0 : \sum_{j=1}^p \mu_j^* \nabla h_j(x^o) = 0$ . Если теперь величины  $\lambda_0^*, \lambda^*$  принять равными нулю, то пункты 2–5 утверждения теоремы будут выполнены.

В) Пусть теперь этот набор линейно независим.

Доказательство будет основано на построении линейной аппроксимации задачи. Введем следующие множества:

$$\begin{aligned} H &= \{x : h(x) = 0\}, \\ G_i &= \{x : g_i(x) \leq 0\}, \\ G_Q &= \{x : Q(x) < Q(x^o)\}. \end{aligned}$$

Если точка  $x^o$  — локальный минимум, то найдется такая ее окрестность, что в пересечении этой окрестности с допустимым множеством не найдется точек  $x$  со значениями  $Q(x) < Q(x^o)$ , т.е.

$$\exists \varepsilon > 0 : G_Q \cap O_\varepsilon(x^o) \cap H \cap \underbrace{\bigcap_{i=1}^m G_i}_{=D} = \emptyset. \quad (3.27)$$

Это иллюстрирует рис. 3.33.

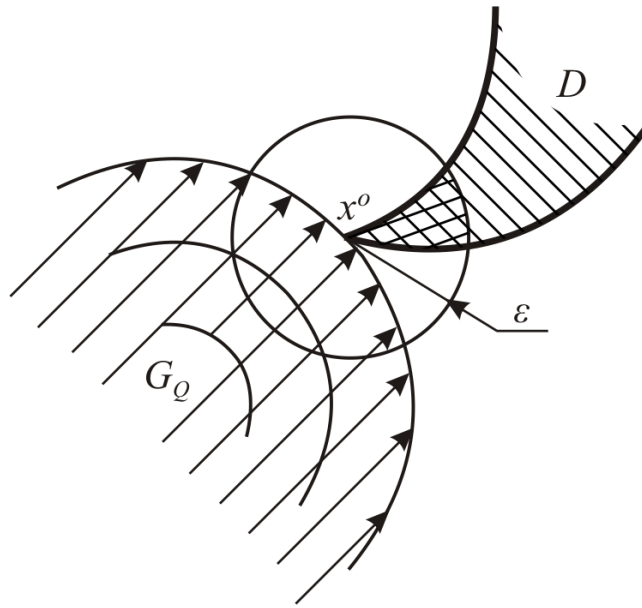


Рис. 3.33. Определение локального минимума  $x^o$  в терминах пустоты пересечения множеств

Введем конические аппроксимации этих множеств. Выполним замену переменных, перейдя от  $x$  к  $d$ :

$$x = x^o + d.$$

Через  $K_H(x^o)$  обозначим множество векторов  $d$ , лежащих на пересечении касательных гиперплоскостей к ограничениям–равенствам  $h_j(x)$  в точке  $x^o$ .  $K_H(x^o)$  является линейным касательным многообразием ко множеству  $H$  в точке  $x^o$ :

$$K_H(x^o) = \{d : (\nabla h_j(x^o), d) = 0, \forall j = 1, \dots, p\}.$$

Аналогично введем конус в виде открытого полупространства, содержащий направления строгого локального убывания  $Q$  в точке  $x^o$ , аппроксимирующий множество  $G_Q$  в окрестности точки  $x^o$ :

$$K_Q(x^o) = \{d : (\nabla Q(x^o), d) < 0\},$$

а также  $r$  конических множеств, аппроксимирующих в окрестности  $x^o$  множества  $G_i$  для активных ограничений:

$$K_{g_i}(x^o) = \{d : (\nabla g_i(x^o), d) < 0\}, \quad \forall i \in I(x^o) = \{i_1, \dots, i_r\}.$$

Виды конических аппроксимаций показаны на рис. 3.34 (a) – (c).

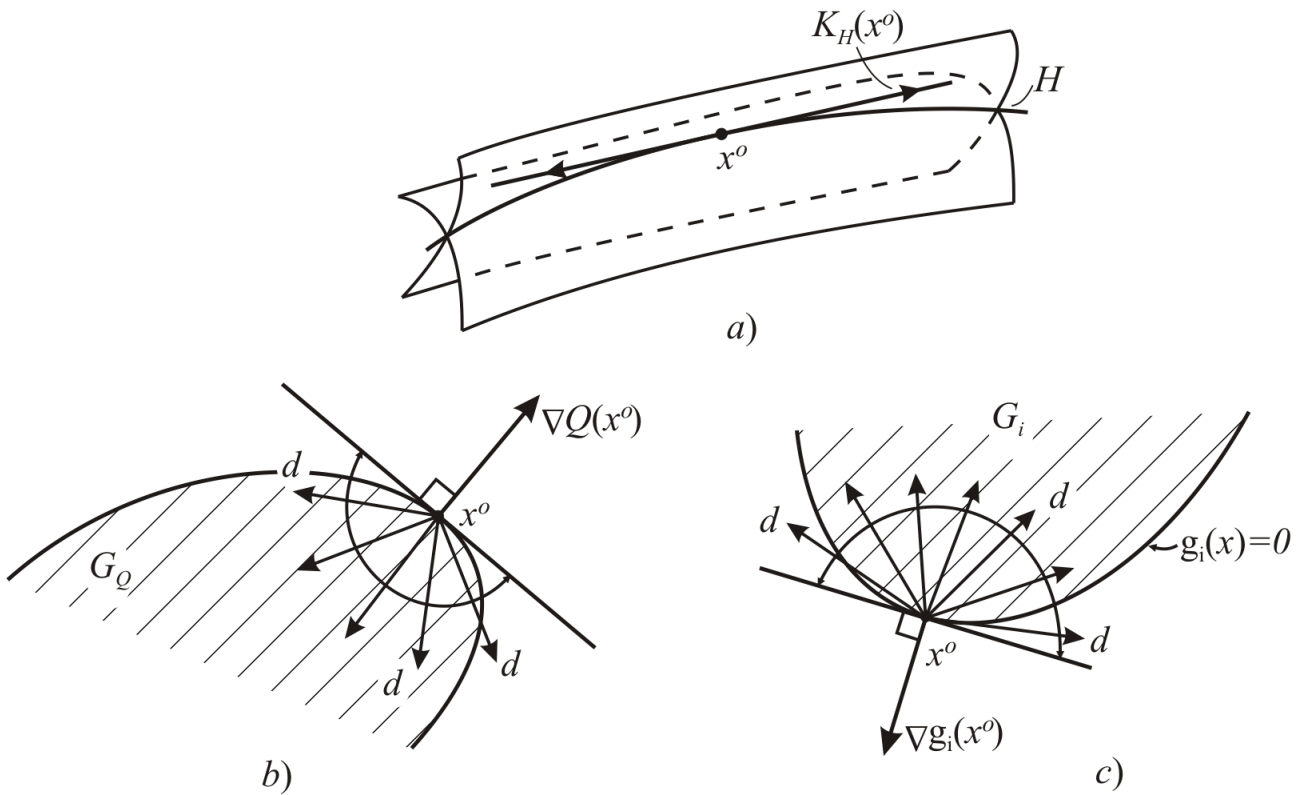


Рис. 3.34. Виды конических аппроксимаций: а)  $K_H(x^o)$ ; б)  $K_Q(x^o)$ ; в)  $K_i(x^o)$

Следующая лемма показывает, что пустота множества (3.27) сохраняется после замены множеств, входящих в (3.27), их коническими аппроксимациями.

**Лемма (основная).** *Если  $x^o$  – локальный минимум, то*

$$K_Q(x^o) \cap \underbrace{\left( \bigcap_{i \in I(x^o)} K_{g_i}(x^o) \cap K_H(x^o) \right)}_{K_D(x^o)} = \emptyset. \quad (3.28)$$



Заметим, что конус  $K_D(x^0)$  следует рассматривать как коническую аппроксимацию допустимого множества  $D$  в окрестности  $x^0$ .

*Доказательство.* Пусть множество из (3.28) непустое. Тогда найдется хотя бы один вектор  $\bar{d} \in K_Q(x^0)$ ,  $\bar{d} \in K_D(x^0)$ . Для доказательства леммы понадобится следующая известная теорема Люстерника.

**Теорема (Люстерника).** Пусть функции  $h(x)$  непрерывно дифференцируемы в окрестности точки  $x^0$  и набор градиентов

$$\nabla h_1(x^0), \dots, \nabla h_p(x^0)$$

линейно независим. Тогда для любого вектора  $d$ , удовлетворяющего условиям:  $(d, \nabla h_i(x^0)) = 0$ ,  $\forall i = 1, \dots, p$  и любого достаточно малого  $\varepsilon$  существует  $o(\varepsilon)$ , при котором выполняется равенство

$$h(x^0 + d\varepsilon + o(\varepsilon)) = 0.$$

Теорема приводится без доказательства.

Продолжим доказательство леммы. Поскольку  $\bar{d} \in K_D(x^0)$ , то вектор  $\bar{d} \in K_H(x^0)$ . Кроме того, для случая (В) набор градиентов функций  $h_i(x)$  в точке  $x^0$  линейно независим. Тогда по теореме Люстерника найдется  $o(\varepsilon)$ , что  $x^\varepsilon = x^0 + \varepsilon\bar{d} + o(\varepsilon)$  будет удовлетворять ограничениям–равенствам (см. рис. 3.35), т.е.  $x^\varepsilon \in H : h(x^0 + \bar{d}\varepsilon + o(\varepsilon)) = h(x^\varepsilon) = 0$ . Выясним, какое значение при-

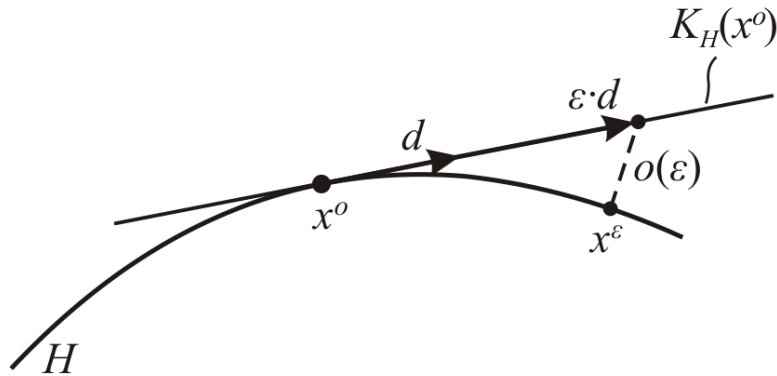


Рис. 3.35. Добавка порядка  $o(\varepsilon)$  возвращает на нелинейное многообразие  $H$  при сдвиге на  $\varepsilon d$  вдоль касательного многообразия

нимает целевая функция в этой точке. По определению дифференцируемости  $Q$  в точке  $x^0$ :

$$Q(x^\varepsilon) = Q(x^0) + (\nabla Q(x^0), \bar{d}\varepsilon + o(\varepsilon)) + \tilde{o}(\varepsilon) = Q(x^0) + \underbrace{(\nabla Q(x^0), \bar{d})}_{<0} \varepsilon + \hat{o}(\varepsilon),$$

где  $\tilde{o}(a)$ ,  $\hat{o}(\varepsilon)$  — некоторые бесконечно малые более высокого порядка, чем  $\varepsilon$ . При достаточно малых  $\varepsilon > 0$ :  $(\nabla Q(x^o), \bar{d})\varepsilon + \hat{o}(\varepsilon) < 0$ . Отсюда  $Q(x^\varepsilon) < Q(x^o)$ . Проверим, удовлетворяет ли точка  $x^\varepsilon$  ограничениям–неравенствам. Рассмотрим произвольное  $i \in I(x^o)$ . Значение функции  $g_i(x^\varepsilon)$  разложим в ряд Тейлора в точке  $x^o$ , и после очевидных преобразований получим:

$$g_i(x^\varepsilon) = g_i(x^o) + (\nabla g_i(x^o), \bar{d})\varepsilon + o(\varepsilon) = (\nabla g_i(x^o), \bar{d})\varepsilon + o(\varepsilon).$$

Поскольку  $\bar{d} \in K_D(x^o)$ , то  $\bar{d} \in K_{g_i}(x^o)$ . Поэтому скалярное произведение меньше нуля. Значение  $g_i(x^o) = 0$ , следовательно при малом  $\varepsilon$  выражение справа от равенства отрицательное. Теперь аналогично рассмотрим любое  $i \notin I(x^o)$ . Для него выполняется равенство

$$g_i(x^\varepsilon) = g_i(x^o) + (\nabla g_i(x^o), \bar{d})\varepsilon + \hat{o}(\varepsilon).$$

Поскольку  $g_i(x^o) < 0$ , то при достаточно малом  $\varepsilon$  сумма справа от равенства будет отрицательна. Мы показали, что при любом достаточно малом  $\varepsilon > 0$  точка  $x^\varepsilon$  является допустимой и значение функции  $Q$  в этой точке меньше  $Q(x^o)$ . Таким образом,  $x^o$  — не локальный минимум. Это противоречие, показывающее, что лемма верна.  $\square$

Продолжим доказательство основной теоремы. Нужно рассмотреть все логически возможные случаи выполнения условия (3.28).

1. Рассмотрим случай, когда  $K_D(x^o) \neq \emptyset$ . Докажем вспомогательную лемму.

**Лемма** (для случая 1). *Если  $K_D(x^o) \neq \emptyset$  и  $K_Q(x^o) \cap K_D(x^o) = \emptyset$ , то значение  $d^* = 0$  является глобальным минимумом задачи*

$$\min_{\substack{(\nabla g_i(x^o), d) \leq 0, \forall i \in I(x^o) \\ (\nabla h_j(x^o), d) = 0, \forall j = 1, \dots, p}} (\nabla Q(x^o), d) \quad (3.29)$$

по переменным  $d$ .

Заметим, что допустимое множество элементов  $d$  в приведенной задаче совпадает с  $\bar{K}_D(x^o)$  — замыканием конуса  $K_D(x^o)$ .

*Доказательство.* Точка  $d^* = 0$  допустима и целевая функция в ней равна нулю. Предположим, что она не является глобальным минимумом. Тогда найдется другая допустимая точка  $\eta \in \bar{K}_D(x^o)$ , что  $(\nabla Q(x^o), \eta) < 0$ .

Множество  $\bar{K}_D(x^o)$  регулярно по Слейтеру, т.к. ограничения–неравенства в (3.29) выпуклы, равенства линейны и  $K_D(x^o) \neq \emptyset$ . Поскольку  $K_D(x^o) \neq \emptyset$ ,  $\exists \xi \in K_D(x^o) : (\nabla h_j(x^o), \xi) = 0$  для всех  $j$ , и  $(\nabla g_i(x^o), \xi) < 0 \forall i \in I(x^o)$ .

Сконструируем дополнительную точку  $d_\varepsilon = \eta + \varepsilon\xi$ . Тогда очевидно, что

$$(\nabla Q(x^o), d_\varepsilon) = \underbrace{(\nabla Q(x^o), \eta)}_{<0} + \varepsilon(\nabla Q(x^o), \xi) < 0$$

при малых  $\varepsilon$ , т.е.  $d_\varepsilon \in K_Q(x^o)$ . Также  $(\nabla h_j(x^o), d_\varepsilon) = 0$ .

Для  $i \in I(x^o)$  :

$$(\nabla g_i(x^o), d_\varepsilon) = \underbrace{(\nabla g_i(x^o), \eta)}_{\leq 0} + \varepsilon \underbrace{(\nabla g_i(x^o), \xi)}_{< 0} < 0.$$

Значит  $d_\varepsilon \in K_D(x^o)$ . Но с другой стороны  $d_\varepsilon \in K_Q(x^o)$ . А это противоречит тому, что множество  $K_D(x^o) \cap K_Q(x^o) = \emptyset$ . Таким образом  $d^* = 0$  — глобальный минимум в задаче (3.29). Лемма доказана.  $\square$

Задача (3.29) является выпуклой гладкой задачей с регулярным множеством. Запишем условия Каруша–Куна–Таккера для задачи (3.29), эти условия являются для нее необходимыми и достаточными:

1.  $d^* = 0$  — допустимая точка;
2.  $\exists(1, \lambda^*, \mu^*) \neq 0$ ;
3.  $\lambda^* \geq 0$ ;
4.  $\nabla_d L(d^*, 1, \lambda^*, \mu^*) = \nabla Q(x^o) + \sum_{i \in I(x^o)} \lambda_i^* \nabla g_i(x^o) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^o) = 0$ ;
5.  $\forall i \in I(x^o) \quad \lambda_i^* (\nabla g_i(x^o), 0) = 0$ .

Заметим, что последнее условие 5 заведомо выполнено и никаких дополнительных требований не налагает. Таким образом, для случая, когда конус  $K_D(x^o)$  не является пустым, основная теорема доказана, поскольку показана справедливость всех ее утверждений.

2. Остается рассмотреть случай пустой области  $K_D(x^o)$ . Здесь возможно два подслучая. Подслучай 2а:  $\exists l : 1 \leq l < r$ , что  $\bigcap_{i=i_{l+1}}^{i_r} K_{g_i}(x^o) \cap K_H \neq \emptyset$ . Подслучай 2б:  $K_{g_{i_r}}(x^o) \cap K_H = \emptyset$ . Рассмотрим первый из них.

2а) Для этого случая нужно доказать еще одну лемму.

**Лемма** (для случая 2а). *Если найдется такое число  $l$ , что  $\bigcap_{i=i_{l+1}}^{i_r} K_{g_i}(x^o) \cap K_H(x^o) \neq \emptyset$ , а  $\bigcap_{i=i_l}^{i_r} K_{g_i}(x^o) \cap K_H(x^o) = \emptyset$ , то точка  $d^* = 0$  является глобальным минимумом в задаче:*

$$\min_{\substack{(\nabla g_{i_s}(x^o), d) \leq 0, \forall s \in l+1, \dots, r \\ (\nabla h_j(x^o), d) = 0, \forall j = 1, \dots, p}} (\nabla g_{i_l}(x^o), d). \quad (3.30)$$

*Доказательство* леммы 2а полностью аналогично доказательству предыдущей леммы. Таким образом, имеем выпуклую, гладкую задачу (3.30) с регулярным допустимым множеством, для которой справедливы следующие условия Каруша–Куна–Таккера:

1.  $d^* = 0$  — допустимая точка;
2.  $\exists(\lambda_{i_l}^* = 1, \lambda_{i_{l+1}}^*, \dots, \lambda_{i_r}^*, \mu^*) \neq 0$ ;
3.  $\lambda_{i_{l+1}}^* \geq 0, \dots, \lambda_{i_r}^* \geq 0$ ;
4.  $\nabla_d L(d^*, 1, \lambda^*, \mu^*) = 1 \cdot \nabla g_{i_l}(x^o) + \sum_{s=l+1}^r \lambda_{i_s}^* \nabla g_{i_s}(x^o) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^o) = 0$ ;
5.  $\forall s = l+1, \dots, r : \lambda_{i_s}^*(\nabla g_{i_s}(x^o), 0) = 0$ .

Здесь условие 5 тривиально выполнено. Как видим, основная теорема в этом случае выполняется при  $\lambda_0^* = 0$ ,  $\lambda_j^* = 0$  для  $j \notin \{i_{l+1}, \dots, i_r\}$ , т.е. это случай нерегулярности.

2b) Рассмотрим последний случай

$$K_{g_{i_r}}(x^o) \cap K_H(x^o) = \emptyset.$$

Здесь  $K_{g_{i_r}}(x^o)$  — открытое полупространство с точкой 0 на границе, а  $K_H(x^o)$  — линейное подпространство. Пересечение может быть пусто только в том случае, когда градиент  $\nabla g_{i_r}(x^o)$  ортогонален множеству  $K_H(x^o)$ , т.е. лежит в ортогональном ему подпространстве (рис. 3.36). Следовательно, как этот

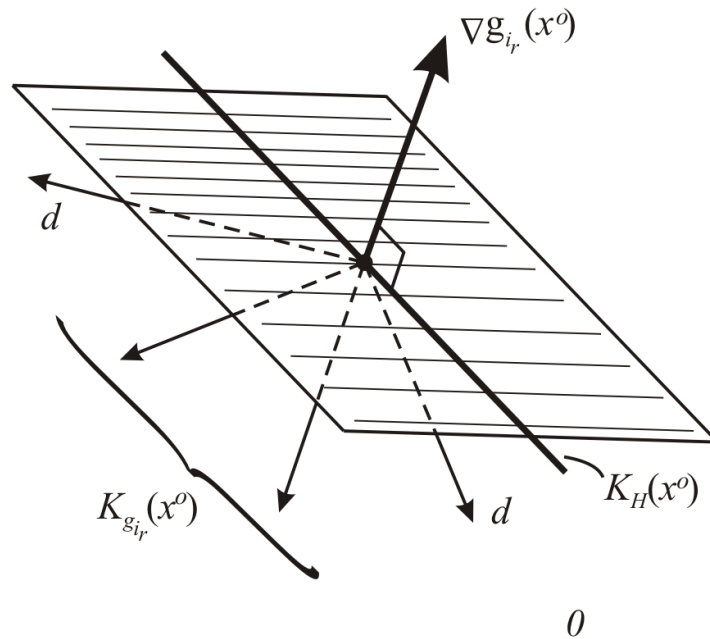


Рис. 3.36. Ортогональность вектора  $\nabla g_{i_r}(x^o)$  к  $K_H(x^o)$

вектор, так и обратный к нему разложимы по системе нормалей  $\nabla h_j(x^o)$  к многообразиям, порожденным равенствами  $h_j(x) = 0$ , ( $j = 1, \dots, p$ ).

Таким образом, существует  $\mu^*$ , что

$$\nabla g_{i_r}(x^o) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^o) = 0.$$

Это означает выполнение утверждений основной теоремы при  $\lambda_0^* = 0$  и  $\lambda_j^* = 0$ , ( $j \neq i_r$ ),  $\lambda_{i_r}^* = 1$ .

Теорема доказана.  $\square$

**Теорема** (достаточное условие регулярности в форме независимости градиентов). *Для класса задач, указанных в предыдущей теореме, для регулярности в точке  $x^o$  допустимого множества вида*

$$D = \{x : g(x) \leq 0, h(x) = 0\}$$

*достаточно, чтобы набор градиентов*

$$\nabla g_{i_1}(x^o), \dots, \nabla g_{i_r}(x^o), \nabla h_1(x^o), \dots, \nabla h_p(x^o)$$

*был линейно независим, где  $\{i_1, \dots, i_r\} = I(x^o)$  — множество номеров активных в точке  $x^o$  ограничений–неравенств.*

*Доказательство.* Предположим, что набор градиентов линейно независим, но множество  $D$  нерегулярно в точке  $x^o$ . Тогда найдется функция  $Q(x)$  из заданного класса задач, имеющая локальный минимум на  $D$  в точке  $x^o$ , и такая, что в условиях 1) – 5) теоремы Каруша–Куна–Таккера  $\lambda_0^* = 0$ . Следовательно верно выражение

$$0 = \sum_{i \in I(x^o)} \lambda_i^* \nabla g_i(x^o) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^o).$$

Но так как набор  $(\lambda_0^*, \lambda^*, \mu^*) \neq 0$ ,  $\lambda_0^* = 0$ , и  $\lambda_i^* = 0$  для  $i \notin I(x^o)$ , то  $(\lambda_{i_1}^*, \lambda_{i_2}^*, \dots, \lambda_{i_r}^*, \mu_1^*, \dots, \mu_p^*) \neq 0$ .

Нашлась нетривиальная линейная комбинация градиентов, равная нулю, значит набор градиентов линейно зависим. Полученное противоречие доказывает теорему.  $\square$

### Достаточные условия локального минимума первого порядка

В общем случае для невыпуклой гладкой задачи общего вида

$$\begin{aligned} & \min_{x \in D} Q(x), \\ & D = \{x \in R^N : g_i(x) \leq 0, (i = 1, \dots, m); h_j(x) = 0, (j = 1, \dots, p)\} \end{aligned} \quad (3.31)$$

рассмотренные выше условия оптимальности первого порядка не могут быть достаточными, поскольку никак не учитывают информацию ни о локальной искривленности поверхностей равного уровня минимизируемой функции, ни о локальной искривленности фрагментов границ допустимого множества в окрестности исследуемой точки  $x^o$ .

Легко привести примеры задач от двух переменных, наглядно подтверждающих данное суждение.

Однако, все же существуют особые случаи, когда условия оптимальности первого порядка могут быть достаточны для определения условного локального минимума. Такие ситуации возможны только тогда, когда совокупность ограничений–равенств и активных в допустимой точке  $x^o$  ограничений–неравенств образуют полную локально невырожденную систему из  $N$  уравнений. При этом та часть антиградиента  $-\nabla Q(x^o)$  целевой функции, которая остается после разложения по нормальям к ограничениям–равенствам, должна находиться строго внутри конуса, натянутого на внешние нормали к активным ограничениям–неравенствам. Эта ситуация подобна той, которая показана на рис. 3.30.

Приведем точную формулировку.

**Теорема** (достаточные условия оптимальности первого порядка). Пусть в задаче (3.31) функции  $Q$ ,  $g$  и  $h(x)$  дифференцируемы в точке  $x^o$ . Тогда выполнение следующих условий:

- 1)  $g(x^o) \leq 0$ ,  $h(x^o) = 0$ , т.е.  $x^o$  допустима;
- 2) существует непустое множество номеров  $I(x^o) = \{i_1, \dots, i_r\}$ , ( $r > 0$ ) активных в точке  $x^o$  ограничений–неравенств;
- 3) в точке  $x^o$  выполнено достаточное условие регулярности в форме линейной независимости градиентов

$$\nabla g_{i_1}(x^o), \dots, \nabla g_{i_r}(x^o), \nabla h_1(x^o), \dots, \nabla h_p(x^o)$$

- 4)  $\exists (\lambda_0^* = 1, \lambda_{i_1}^*, \lambda_{i_r}^*, \mu^*)$ , что справедливо разложение

$$-\nabla Q(x^o) = \sum_{s=1}^r \lambda_{i_s}^* \nabla g_{i_s}(x^o) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^o)$$

- 5)  $p + r = N$  и  $\forall s = 1, \dots, r : \lambda_{i_s}^* > 0$ ;

достаточно, чтобы точка  $x^o$  являлась острым локальным минимумом в задаче (3.31).

Теорема приводится без доказательства. Его можно найти в [25] (гл.9, §2.2, теорема 4).

Поясним не встречавшийся ранее термин острого локального минимума.

**Определение.** Допустимая точка  $x^o$  называется острым локальным минимумом (условным), если  $\exists \varepsilon > 0$  и  $\alpha > 0$ , что  $\forall x \in D \cap O_\varepsilon(x^o) : Q(x) \geq Q(x^o) + \alpha \|x - x^o\|$ .

## Необходимые, а также достаточные условия оптимальности второго порядка

Рассматривается основная задача математического программирования вида (3.31).

Пусть все функции дважды непрерывно дифференцируемы в некоторой окрестности точки  $x^o$ . Введем несколько условий.

A) Условия оптимальности первого порядка:

1. точка  $x^o$  — допустима;
2.  $\exists (\lambda_0^*, \lambda^*, \mu^*) \neq 0$ , что  $\lambda_0^* \geq 0$ ,  $\lambda^* \geq 0$ ;
3.  $\nabla_x L(x^o, \lambda_0^*, \lambda^*, \mu^*) = 0$  — выполнено условие стационарности;
4.  $\forall i = 1, \dots, m : \lambda_i^* g_i(x^o) = 0$  — условие дополняющей нежесткости.

B) Выполнено достаточное условие регулярности в форме линейной независимости системы векторов:

$$\nabla g_{i_1}(x^o), \dots, \nabla g_{i_r}(x^o), \nabla h_1(x^o), \dots, \nabla h_p(x^o),$$

где  $\{i_1, \dots, i_r\} = I(x^o)$ .

C) Для любого вектора  $d \neq 0$ , удовлетворяющего условиям:

$$\begin{aligned} (d, \nabla g_{i_s}(x^o)) &= 0, & s &= 1, \dots, r; \\ (d, \nabla h_j(x^o)) &= 0, & j &= 1, \dots, p \end{aligned} \quad (3.32)$$

выполняется неравенство

$$d^T \nabla_{xx}^2 L(x^o, \lambda_0^*, \lambda^*, \mu^*) d > 0. \quad (3.33)$$

**Теорема** (необходимые и достаточные условия оптимальности второго порядка). Пусть в задаче математического программирования (3.31) функции  $Q$ ,  $g$ ,  $h$  дважды непрерывно дифференцируемы в окрестности точки  $x^o$ , т.е.  $\{Q, g, h \in C^2(O_\varepsilon(x^o))\}$ . Тогда справедливо следующее.

1. *Необходимость:*

если выполняется достаточное условие регулярности B и  $x^o$  — локальный минимум, то выполняются условия оптимальности первого порядка A при  $\lambda_0^* \neq 0$ , и кроме того  $\forall d \neq 0$ , удовлетворяющего (3.32), выполнено неравенство

$$d^T \nabla_{xx}^2 L(x^o, \lambda_0^*, \lambda^*, \mu^*) d \geq 0$$

2. *Достаточность:*

выполнение условий A, B, C достаточно для того, чтобы  $x^o$  являлся строгим локальным минимумом в задаче (3.31).

Теорема приводится без доказательства, с ним можно самостоятельно познакомиться в [12], гл. XX.

Заметим, что достаточными условиями приведенной теоремы можно пользоваться лишь в том случае, когда суммарное число ограничений–равенств и активных ограничений–неравенств строго меньше размерности задачи.

Остановимся на геометрическом смысле условия  $C$ . Его следует рассматривать как обобщение на случай задач с ограничениями известного результата из математического анализа для гладких до второго порядка задач без ограничений. В таких задачах необходимое условие второго порядка для локального минимума заключается в неотрицательной определенности матрицы Гессе  $\nabla_{xx}^2 Q(x^o)$ , а достаточное условие строгого локального минимума — в выполнении условия стационарности  $\nabla Q(x^o) = 0$  и положительной определенности матрицы Гессе  $\nabla_{xx}^2 Q(x^o)$ , что означает локальную строгую выпуклость  $Q(x)$  (локальную положительную кривизну поверхности  $Q(x)$  в любых направлениях).

Легко заметить, что при отсутствии ограничений в условии  $C$  следует рассматривать произвольные  $d \neq 0$ , и неравенство (3.33) или его нестрогий аналог превратятся в приведенные выше требования к знакоопределенности гессиана функции  $Q$ .

Непосредственный геометрический смысл самого условия  $C$  заключается в локальной положительной кривизне функции Лагранжа по переменным  $x$  при значениях  $\lambda_o^*, \lambda^*, \mu^*$ , но не во всем пространстве, а только на линейном многообразии направлений, касательных в точке  $x^o$  к ограничениям–равенствам и активным в  $x^o$  неравенствам.



# Глава 4. Численные методы решения задач математического программирования

## 4.1. Понятие численного метода математического программирования

Прежде чем ввести понятие «численные методы математического программирования» заметим, что их называют также *поисковыми методами*. Пусть задано множество элементов  $E$ . Оно может быть различным, однако чаще всего встречаются случаи, когда либо  $E = R^N$ , либо

$$E = \{x \in R^N : a_i \leq x_i \leq b_i, i = 1, \dots, N\}.$$

Далее множество  $E$  будем называть *областью поиска*.

Рассмотрим задачу математического программирования:

$$\min_{x \in D} Q(x) \tag{4.1}$$

$$D = \{x \in E : g_i(x) \leq 0, (i = 1, \dots, m); h_s(x) = 0, (s = 1, \dots, p)\},$$

где, как и раньше, ограничения могут быть записаны в векторной форме:  $g(x) \leq 0, h(x) = 0$ .

Введем вектор-функцию  $f(x) = (Q, g, h)$ , описывающую весь набор функций задачи. Тогда при заданной области поиска  $E$  задача (4.1) полностью определяется функцией  $f$ . Будем предполагать, что относительно  $f$ , имеется некоторая начальная информация. Её называют *априорной информацией*. Она всегда присутствует в той или иной форме, может быть по своей природе весьма различной. Априорная информация необходима для построения численных методов. Наиболее простым и, в тоже время, весьма общим способом её описания является требование принадлежности функции  $f$  некоторому классу функций  $\Phi$  с заданными свойствами. Например, это может быть класс липшицевых функций  $\Phi_{lip}$ , класс выпуклых функций  $\Phi_{conv}$ , унимодальных -  $\Phi_{un}$  и т.д. (см.рис. 4.1). В более сложном случае на классе функций может быть задана вероятностная мера, описывающая относительные частоты встречаемости в данном классе тех или иных функций (см. [2] — глава 4, [28] — раздел 1.1.).

В отличие от аналитических методов решения задач математического программирования (применимых только к достаточно простым задачам) при использовании численных методов единственной возможностью получения дополнительной информации о решаемой задаче является проведение её испытаний. Испытания могут выполняться в произвольных точках  $x^1, x^2, \dots, x^k, \dots$  заданной области поиска  $E$ .

Испытанием задачи в точке  $x^k \in E$  называют вычисление набора  $J_k = J(f, x^k)$  локальных характеристик функции  $f$  задачи в точке  $x^k$ . В понятие *локальной характеристики* могут включаться значения самой функции  $f$ , её градиента, вторых частных производных и т.д. при этом, наивысший порядок вычисляемых производных называется *порядком испытания*.

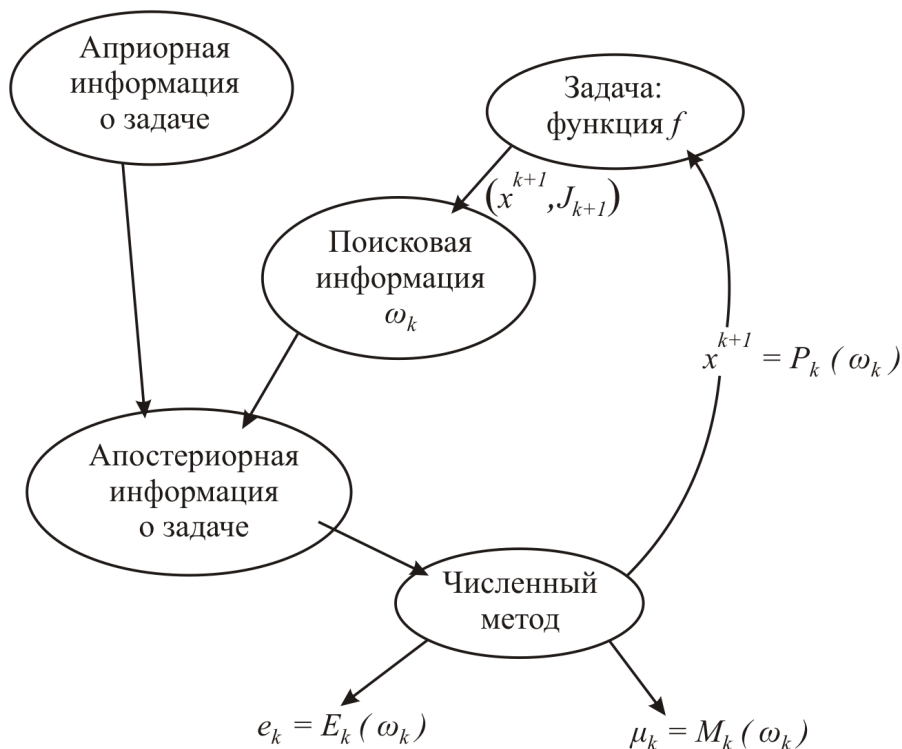


Рис. 4.1. Структура и схема функционирования численного метода оптимизации

Например, при проведении испытаний нулевого порядка

$$J_k = (Q(x^k), g(x^k), h(x^k)) = (Q_k, g^k, h^k),$$

если проводятся испытания первого порядка, то

$$J_k = (Q_k, g^k, h^k; \nabla Q_k, \nabla g^k, \nabla h^k).$$

Допустим, проведена серия испытаний в точках  $x^1, \dots, x^k$  области поиска  $E$  и получены результаты испытаний  $J_1, \dots, J_k$ . Множество  $\omega_k$ , включающее все эти данные, называют *поисковой информацией*.

$$\omega_k = \{(x^s, J_s) : s = 1, \dots, k\}.$$

При накоплении поисковой информации происходит её интерпретация с использованием априорной информации. В результате образуется *апостериорная информация*, которую использует численный метод (рис. 4.1).

Основной составляющей численного метода является решающее правило  $P_k(\omega_k)$ , определяющее выбор точки  $x^{k+1}$  для проведения нового испытания.

Метод обращается к задаче, передавая ей значение  $x^{k+1}$  и получает в ответ результат испытания  $J_{k+1}$ . Пара  $(x^{k+1}, J_{k+1})$  пополняет поисковую информацию  $\omega_k$ .

Кроме того, в структуру метода должны включаться два дополнительных правила  $E_k(\omega_k)$  и  $M_k(\omega_k)$ . Первое из них определяет текущую *оценку решения*. Она может быть разного вида (точечная оценка по координате, оценка по значению целевой функции, интервальная оценка и т.п.). Второе правило определяет, должен ли остановиться метод. Например,  $M_k(\omega_k) = 1$  может означать останов, а  $M_k(\omega_k) = 0$  — необходимость продолжения поиска (см. [2], [28]).

**Определение.** Численным методом математического программирования  $\alpha$  называют набор правил  $\alpha = \langle P_k, E_k, M_k \rangle$  ( $k = 1, 2, \dots$ ), где  $P_k$  — правило выбора новой точки,  $E_k$  — правило, определяющее текущую оценку решения,  $M_k$  — правило, определяющее остановку метода.

**Определение.** Если метод использует испытания  $k$ -го порядка, то его называют *методом  $k$ -го порядка*.

**Определение.** Если правило  $P_k$  в методе  $\alpha$  не зависит от  $\omega_k$ , то метод  $\alpha$  называют *пассивным*. Иначе метод называют *последовательным*.

Пассивный метод строит последовательность точек  $\{x^k\}$ , которая не зависит от результатов испытаний функции  $f$ .

**Определение.** Если правило останова  $M_k$  в составе метода завершает его ровно после  $n$  испытаний, то такой метод называется  *$n$ -шаговым*.

Введенное понятие численного метода активно используется при их конструировании. В некоторых случаях (при достаточно богатой априорной информации о классе функций) удастся построить методы, являющиеся наилучшими по отношению к некоторым введенным критериям качества. Этот подход будет использован в следующем подразделе. Если же построение метода из условий оптимальности невозможно, используются иные критерии: обеспечение сходимости, высокий порядок скорости сходимости, глобализация сходимости.

## 4.2. Понятие оптимального и $\varepsilon$ -оптимального метода

Пусть необходимо найти решение задачи (4.1) в области поиска  $E$ . Будем предполагать, что априорная информация о вектор-функции  $f$  задана в терминах принадлежности  $f$  множеству задач  $\Phi$  с определенными свойствами. Кроме того, будем считать, что имеется некоторый класс методов  $A$ , применимых к задаче (4.1).

Введем  $d(\alpha, f)$  — функцию эффективности метода  $\alpha$ , примененного к задаче  $f$ . Примем, что чем меньше  $d(\alpha, f)$ , тем эффективность выше. Например,  $d(\alpha, f)$  можно интерпретировать как значение остаточной погрешности.

Может показаться, что оптимальным на классе  $A$  следует назвать тот метод  $\alpha^*$  из  $A$ , для которого

$$d(\alpha^*, f) = \inf_{\alpha \in A} d(\alpha, f).$$

Однако такое определение некорректно, т.к.  $\alpha^*$  зависело бы от решаемой задачи  $f$ , которая при выборе метода неизвестна! Нужно избавиться от влияния  $f$ . Для этого введем понятие гарантированной эффективности.

**Определение.** Гарантированной эффективностью метода  $\alpha$  на классе задач  $\Phi$  называют функцию

$$d^*(\alpha) = \sup_{f \in \Phi} d(\alpha, f). \quad (4.2)$$

Гарантированная эффективность определяет наихудшее значение эффективности, которую может иметь метод  $\alpha$  на задачах из  $\Phi$ .

**Определение.** Метод  $\alpha^* \in A$  называют оптимальным в смысле гарантированной эффективности, если

$$d^*(\alpha^*) = \inf_{\alpha \in A} d^*(\alpha). \quad (4.3)$$

Так как точная нижняя грань не всегда достижима, то оптимального алгоритма может и не быть. В этом случае используют понятие  $\varepsilon$ -оптимальности.

**Определение.** Метод  $\alpha^o \in A$  называют  $\varepsilon$ -оптимальным в смысле гарантированной эффективности, если для  $\varepsilon > 0$

$$d^*(\alpha^o) \leq \inf_{\alpha \in A} d^*(\alpha) + \varepsilon. \quad (4.4)$$

Это лишь один из способов определения оптимального алгоритма. Кроме принципа наилучшего гарантированного результата используется принцип оптимальности в среднем, понятие асимптотической оптимальности, а также более сложный принцип последовательной оптимальности применяемый как в гарантированном смысле, так и в среднем. Широко используется также принцип одношаговой оптимальности, являющийся упрощением последовательной оптимальности [2], [28].

Приведем один из вариантов определения последовательно-оптимального метода. Необходимость введения этого понятия связана с тем, что в приведенных ранее определениях (4.2)–(4.4) речь фактически шла об оптимальности по отношению лишь к априорному классу задач  $\Phi$ . Ниже в п. 1.3. будет

показана эффективность этих определений при построении методов поиска решения унимодальных задач. В то же время, Ю. Б. Гермейером было показано, что на классе липшицевых функций эти же определения приводят к парадоксу. Исправить ситуацию можно за счет введения более сложного *принципа последовательной оптимальности*. Этот принцип учитывает, что информация о классе функций изменяется по мере накопления поисковой информации, чего не принимают во внимание оптимальные алгоритмы. Опишем соответствующее понятие в предположении, что рассматривается класс  $A_n$  —  $n$ -шаговых алгоритмов, а указанные ниже точные нижние и верхние грани достигаются.

Пусть метод  $\alpha$  уже выполнил  $k$  испытаний, результаты которых запомнены в виде поисковой информации  $\omega_k$ . Обозначим через  $\Phi_{\omega_k}$  апостериорный класс функций:

$$\Phi_{\omega_k} = \{f \in \Phi : \forall (x^s, J_s) \in \omega_k : J(x^s, f) = J_s\}.$$

Таким образом, в классе  $\Phi_{\omega_k}$  содержатся только те функции из исходного класса  $\Phi$ , результаты испытаний которых во всех точках  $x^s$  уже проведенных испытаний совпадают с имеющимися результатами  $J_s$  этих испытаний.

Функции из  $\Phi_{\omega_k}$  не различимы по результатам испытаний в точках  $x^s$  ( $s = 1, \dots, k$ ).

**Определение.** Говорят, что  $n$ -шаговый метод  $\alpha_n \in A_n$  *последовательно оптимален на классе  $\Phi$*  (в гарантирующем смысле), если его правила  $P_k$  порождают выбор точек испытаний  $x^{k+1}$  исходя из следующего условия:

$$\inf_{x^1 \in E} \sup_{\substack{J_1 = J(x^1, f) \\ f \in \Phi}} \inf_{x^2 \in E} \sup_{\substack{J_2 = J(x^2, f) \\ f \in \Phi_{\omega_1}}} \dots \inf_{x^n \in E} \sup_{\substack{J_n = J(x^n, f) \\ f \in \Phi_{\omega_{n-1}}}} d((x^1, J_1); \dots; (x^n, J_n)). \quad (4.5)$$

**Определение.** Метод  $\alpha$  называют *последовательно оптимальным на один шаг* (или *одношагово-оптимальным*), если для любого  $k$ :

$$x^{k+1} = \arg \inf_{x \in E} \sup_{\substack{J = J(x, f) \\ f \in \Phi_{\omega_k}}} d(\omega_k \cup \{(x, J)\}). \quad (4.6)$$

### 4.3. Построение методов поиска минимума унимодальной функции на отрезке

Материал данного раздела с различных точек зрения представлен в разных источниках. Кроме ранее приведенных [2], [28], укажем на [5], [10], [19].

Рассмотрим задачу поиска минимума функции

$$\min_{x \in [a, b] \subseteq \mathbb{R}^1} Q(x). \quad (4.7)$$

**Определение.** Функцию  $Q(x)$ , имеющую на  $[a, b]$  глобальный минимум, назовем *униmodalьной* на отрезке  $[a, b]$ , если существует точка  $x^* \in [a, b]$ , что на отрезке  $[a, x^*]$  функция  $Q(x)$  строго убывает, а на  $[x^*, b]$  — строго возрастает.

Для класса униmodalьных на  $[a, b]$  функций введем обозначение —  $\Phi_{un}[a, b]$ . Изучим вопрос о построении методов нулевого порядка, когда  $J_k = J(x^k, Q) = Q(x^k)$ .

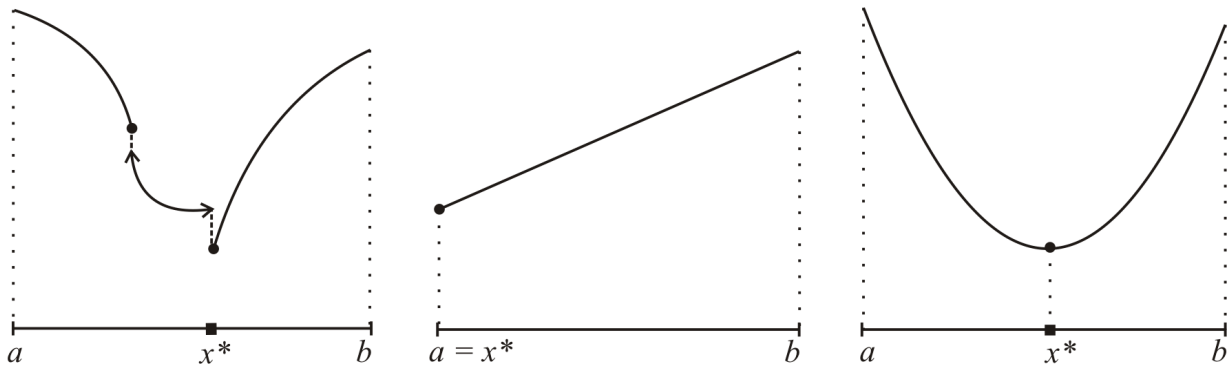


Рис. 4.2. Несколько примеров поведения униmodalьных функций

**Утверждение.** Пусть  $Q(x)$  — униmodalьна на  $[a, b]$ , и  $a < x_1 < x_2 < b$ . Тогда:

а) если  $Q(x_1) \leq Q(x_2)$ , то  $x^* \notin (x_2, b]$ ; б) если  $Q(x_1) \geq Q(x_2)$ , то  $x^* \notin [a, x_1)$ .

*Доказательство.* Рассмотрим пункт (а) (см. рис. 4.3). Предположим, что утверждение (а) неверно, т.е.  $x^* \in (x_2, b]$  и, следовательно, функция  $Q$  строго убывает на  $[a, x^*]$ :

$$Q(x_1) > Q(x_2) > Q(x^*).$$

Однако, левое неравенство противоречит условию пункта (а). Аналогично доказывается утверждение для пункта (б).  $\square$

**Следствие.** Пусть функция  $Q(x)$  униmodalьна на отрезке  $[a, b]$  и в его внутренних точках  $x_1 < x_2 < \dots < x_n$  проведены измерения с результатами  $Q_k = Q(x_k)$ , ( $k = 1, \dots, n$ ). Если формально обозначить  $x_0 = a$  и  $x_{n+1} = b$ , получим

$$a = x_0 < x_1 < x_2 < \dots < x_n < x_{n+1} = b,$$

причем  $a$  и  $b$  не являются точками измерений. Пусть  $k^*$  — номер точки измерения с наименьшим значением функции, т.е.

$$k^* = \arg \min_{k=1, \dots, n} Q_k, \quad (4.8)$$

тогда

$$x^* \in [x_{k^*-1}, x_{k^*+1}]. \quad (4.9)$$

Заметим, что  $k^*$  зависит от  $Q$ , т.е.  $k^* = k^*(Q)$ .

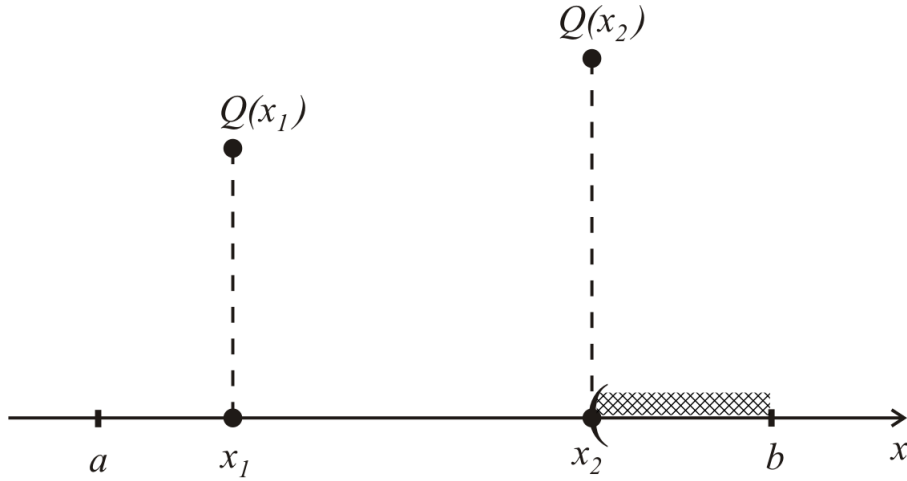


Рис. 4.3. Сокращение области поиска по двум измерениям (будет удалено множество, отмеченное штриховкой)

*Доказательство* получается в результате многократного применения к имеющимся парам результатов измерений обоснованного выше утверждения (см. рис. 4.4).

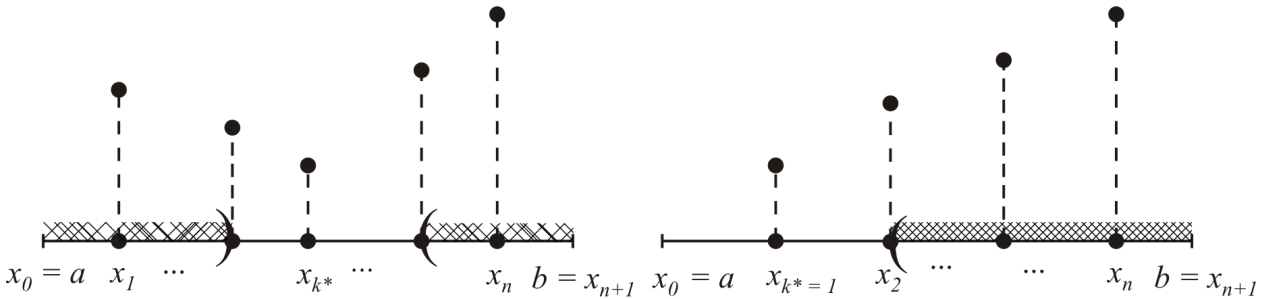


Рис. 4.4. Примеры сокращения области поиска по  $n$  измерениям

### 4.3.1. Построение оптимальных и $\varepsilon$ -оптимальных методов в классе пассивных $n$ -шаговых

Напомним, что метод  $\alpha$  называется *пассивным*, если правило  $P_k$  размещения очередного измерения в этом методе не зависит от поисковой информации  $\omega_k$ .

Рассмотрим класс  $A_n^{\text{пас}}$  пассивных  $n$ -шаговых методов. Пусть метод  $\alpha_n^{\text{пас}} \in A_n^{\text{пас}}$ , функция  $Q$  принадлежит классу унимодальных функций на отрезке  $[a, b]$ , а функция эффективности определяет длину наименьшего интервала, содержащего решение  $x^*$ . Из (1.7), (1.8) следует, что для пассивного

метода выражение для функции эффективности будет выглядеть следующим образом:

$$d(\alpha_n^{\text{пас}}, Q) = x_{k^*(Q)+1} - x_{k^*(Q)-1},$$

где

$$k^*(Q) = \arg \min_{k=1, \dots, n} Q(x_k).$$

В пассивных методах точки измерений  $a < x_1 < x_2 < \dots < x_n < b$  выбираются до проведения испытаний. После вычисления значений  $Q_k = Q(x_k)$ , ( $k = 1, \dots, n$ ) применяется стандартное правило отбрасывания (4.8), (4.9). Таким образом, пассивный  $n$ -шаговый алгоритм полностью определяется выбором  $n$  точек измерений.

Гарантированная эффективность на классе унимодальных функций для  $n$ -шаговых, пассивных алгоритмов имеет вид:

$$d^*(\alpha_n^{\text{пас}}) = \sup_{Q \in \Phi_{un}[a, b]} (x_{k^*(Q)+1} - x_{k^*(Q)-1}) = \max_{k=1, \dots, n} (x_{k+1} - x_{k-1}). \quad (4.10)$$

Рассмотрим  $n$ -шаговые пассивные алгоритмы и попробуем среди них найти оптимальный. Отдельно исследуем методы, выполняющие нечетное и четное число измерений.

### Случай нечетного числа измерений $n=2r+1$

Построим оптимальный метод  $\alpha_n^{\text{пас}}$ . Отрезки вида  $x_{k+1} - x_{k-1}$ , участвующие в сравнении длин в выражении (4.10), распадаются на две группы. На рис. 4.5 они отмечены верхними и нижними дугами. «Верхних» отрезков будет  $L = (n + 1)/2$  штук. Уменьшение длины наибольшего из них может происходить при выравнивании длин всех этих отрезков (рис. 4.5-b).

Таким образом,  $\inf_{a < x_1 < \dots < x_n < b} \max_{k=1, \dots, n} (x_{k+1} - x_{k-1})$  достигается, если отрезок  $[a, b]$  разбить дугами равной длины. Оптимальная гарантированная эффективность такого метода будет определяться выражением

$$d^*(\alpha_n^{\text{пас}}) = \inf_{a < x_1 < \dots < x_n < b} \max_{k=1, \dots, n} (x_{k+1} - x_{k-1}) = (b - a) / \left(\frac{n + 1}{2}\right). \quad (4.11)$$

Заметим, что полученный оптимальный метод — не единственный. Существует континуум оптимальных методов, так как точки измерений, связанных нижними дугами можно произвольно смещать без перехода через остальные точки. Более того, точки нижнего ряда можно сближать друг с другом не допуская переходов через другие точки. Гарантированная эффективность метода от этого не изменится.



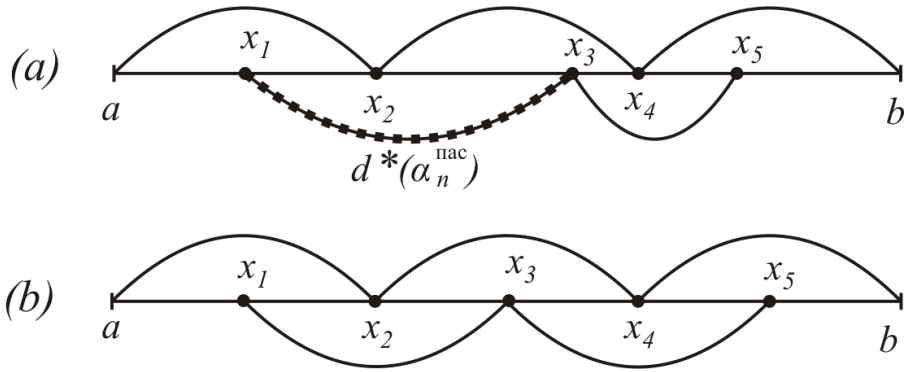


Рис. 4.5. (а) — выделен интервал, определяющий гарантированную эффективность; (в) — одно из оптимальных размещений для нечетного  $n$

### Случай четного числа измерений $n = 2r$

Повторяя предыдущие рассуждения (см. рис. 4.6-а) легко увидеть, что при четном  $n$  оптимального метода не существует, т.к. предельное уменьшение длин отрезков, показанных на рис. 4.6-а, приведёт к попарному слиянию точек. Но можно построить  $\varepsilon$ -оптимальные методы, если разбить отрезок  $[a, b]$  на  $L = (n/2) + 1$  равных частей и разместить точки измерений попарно слева и справа вблизи от разделяющих линий. Тогда для полученного  $\varepsilon$ -оптимального алгоритма  $\alpha_n^{o\text{пас}}$  гарантированная эффективность:

$$d^*(\alpha_n^{o\text{пас}}) = \inf_{a < x_1 < \dots < x_n < b} \max_{k=1, \dots, n} (x_{k+1} - x_{k-1}) + \varepsilon = \frac{b-a}{\frac{n}{2} + 1} + \varepsilon. \quad (4.12)$$

Один из вариантов  $\varepsilon$ -оптимального размещения приведен на рис. 4.6-в.

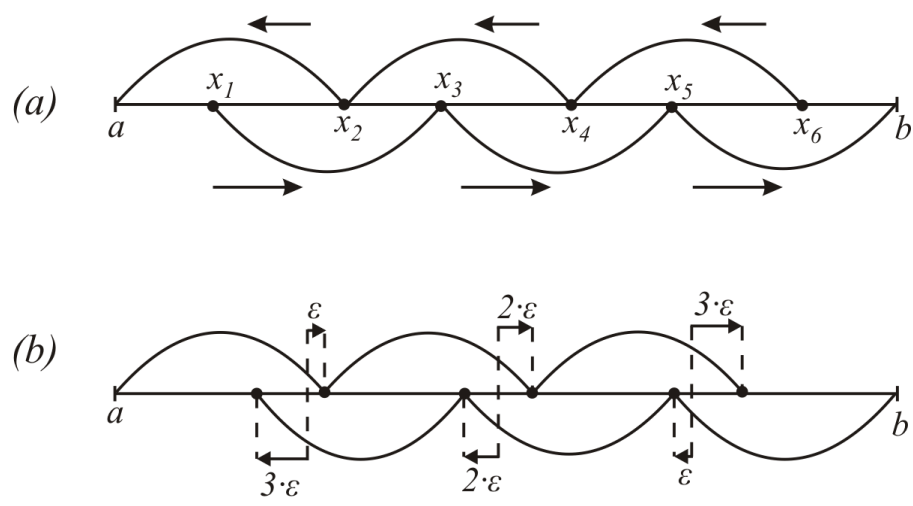


Рис. 4.6. (а) — оптимизация размещения точек; (в) — одно из  $\varepsilon$ -оптимальных размещений при четном  $n$

Заметим, что эффективность построенных оптимальных и  $\varepsilon$ -оптимальных методов невелика. Это связано с тем, что они выбирались в классе пассивных. Далее покажем, что расширение этого класса за счет включения

последовательных методов позволит повысить гарантированную эффективность.

### 4.3.2. Методы дихотомии — неоптимальные последовательные методы

Методы дихотомии используются для нахождения безусловного минимума унимодальных функций  $Q(x)$ . Они являются одними из самых простых методов последовательного поиска. Название связано с тем, что каждые два испытания в два раза (или примерно в два раза) сокращают область поиска.

#### Первый вариант метода дихотомии

Введем параметр точности  $\delta > 0$  и параметр  $\varepsilon > 0$  (причем  $\varepsilon \ll \delta$ ).

*Алгоритм метода «дихотомия-1»*

**Шаг 0.** Задать  $\delta$ ;  $0 < \varepsilon \ll \delta$ .

**Шаг 1.** Пока  $b - a > \delta$  выполнять шаги 1.1 и 1.2:

**Шаг 1.1.** Присвоить:  $x_1 = \frac{b+a}{2} - \varepsilon$ ;  $x_2 = \frac{b+a}{2} + \varepsilon$ .

**Шаг 1.2.** Если  $Q(x_1) < Q(x_2)$ , то  $b := x_2$ , иначе  $a := x_1$

**Шаг 2.** Завершить выполнение метода с интервальной оценкой решения  $x^* \in [a, b]$ , где  $b - a \leq \delta$ .

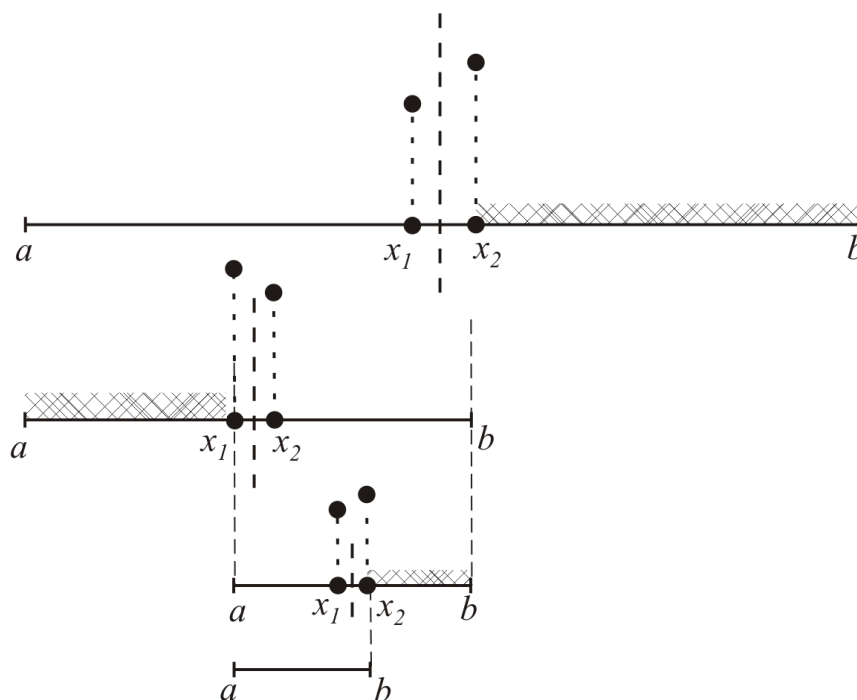


Рис. 4.7. Сжатие области поиска при использовании метода дихотомии-1

Гарантированная эффективность метода почти экспоненциально улучша-

ется с ростом числа измерений. При  $\varepsilon \rightarrow 0$

$$d^*(\alpha_n^{\text{дих-1}}) \approx \frac{b-a}{2^{n/2}}.$$

В этом методе  $n$  всегда четно.

### Второй вариант метода дихотомии

В этом варианте метода первое измерение выполняется в точке  $c$  — середине исходного интервала. Оно не приводит к его сокращению. Далее каждые два новых измерения сокращают область поиска ровно в два раза. Гарантированная эффективность этого метода

$$d^*(\alpha_n^{\text{дих-2}}) = \frac{b-a}{2^{(n-1)/2}}. \quad (4.13)$$

В этом методе  $n$  всегда нечетно.

*Алгоритм метода «дихотомия-2»*

**Шаг 0.** Задать  $\delta > 0$  — точность решения.

**Шаг 1.** Положить  $c = (b+a)/2$  и  $Q_c = Q(c)$ .

**Шаг 2.** Пока  $b-a > \delta$  выполнять шаги 2.1 и 2.2:

**Шаг 2.1.** Вычислить:  $x = (a+c)/2$ ,  $Q_x = Q(x)$ ;  $y = (c+b)/2$ ,  $Q_y = Q(y)$ ;

**Шаг 2.2.** Если  $Q_x \leq Q_c < Q_y$ , положить  $b := c$ ,  $c := x$ ,  $Q_c := Q_x$ ; иначе, если  $Q_x > Q_c \leq Q_y$ , то положить  $a := x$ ,  $b := y$ ; иначе положить  $a := c$ ,  $c := y$ ,  $Q_c := Q_y$ .

**Шаг 3.** Завершить выполнение метода с интервальной оценкой решения  $x^* \in [a, b]$ , где  $b-a \leq \delta$ .

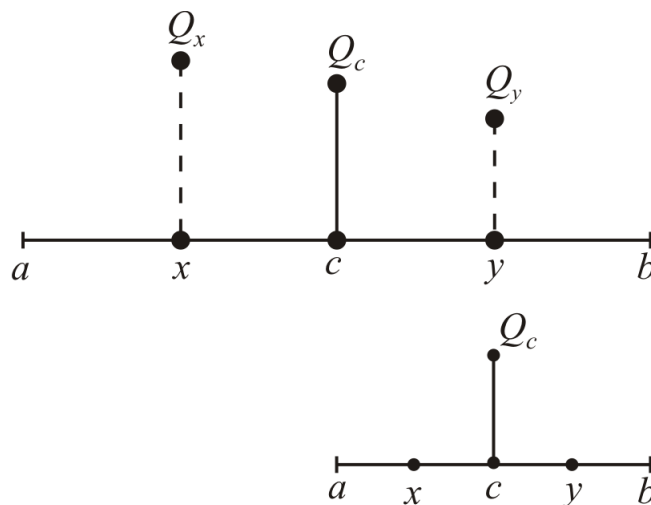


Рис. 4.8. Один из трех возможных случаев сжатия интервала поиска в методе дихотомии-2 после двух проведенных измерений в точках  $x$  и  $y$

Таким образом, гарантированная эффективность простых последовательных методов оказалась выше гарантированной эффективности оптимальных и  $\varepsilon$ -оптимальных пассивных.

### 4.3.3. Построение $\varepsilon$ -оптимального последовательного метода для класса унимодальных задач, метод Фибоначчи

Построим теперь оптимальный или  $\varepsilon$ -оптимальный последовательный метод. По-прежнему будем рассматривать задачу математического программирования

$$\min_{x \in [a, b]} Q(x), \quad (4.14)$$

где  $Q$  принадлежит классу унимодальных на отрезке  $[a, b]$  функций, т.е.  $Q \in \Phi_{un}[a, b]$ . Ограничимся изучением некоторого подкласса последовательных методов, а именно класса  $A_n^{сим}$  —  $n$ -шаговых симметричных алгоритмов.

Приведем описание симметричного метода. Пусть вначале имеем отрезок  $[a, b]$ , длины  $l = b - a$ . Будем последовательно размещать очередную точку испытания симметрично относительно середины оставшегося отрезка по отношению к имеющейся на нем точке прежнего испытания. При этом примем, что расстояние от этих точек до наиболее удаленных от них концов отрезка составляет долю в общей длине, равную  $\lambda_k$  после размещения  $k$ -го испытания.

Для точек первых двух измерений  $x$  и  $y$ , симметрично размещенных на отрезке так, что  $x < y$ , имеем:

$$\lambda_2 = \frac{y - a}{b - a} = \frac{b - x}{b - a}.$$

Проведя испытания, по их результатам исключим из исходного отрезка ту его часть, где минимума заведомо быть не может. Останется отрезок длины  $(b - a)\lambda_2 = l\lambda_2$ , содержащий внутри одну из точек ранее проведенного испытания. Обозначим ее через  $v$ . На новом отрезке достаточно разместить всего одну новую точку, симметрично к уже имеющейся точке  $v$ . Долю, в которой эти точки делят текущий отрезок обозначим через  $\lambda_3$ . Продолжим процесс до тех пор, пока не проведем все  $n$  измерений.

Для симметричных методов можно определить связь между  $\lambda_k$  и  $\lambda_{k+1}$ . Пусть на  $k$ -ом шаге рассматривался текущий отрезок  $[\bar{a}, \bar{b}]$  длины  $(\bar{b} - \bar{a})$ . Согласно методу, на  $[\bar{a}, \bar{b}]$  испытания проводятся после измерений функции в двух симметричных точках  $x, y$  ( $x < y$ ). Пусть (для определенности) оказался отброшен правый подынтервал  $(y, \bar{b}]$ . Оставшийся обозначим  $[\tilde{a}, \tilde{b}]$ , при этом прежнюю точку  $x$  обозначим через  $v$ . Вычислим двумя различными способами (по рис. 4.9-а и 4.9-б) длину интервала  $[\bar{a}, x] = [\tilde{a}, v]$  и приравняем эти выражения. Получим:

$$(\bar{b} - \bar{a}) \cdot \lambda_k \cdot \lambda_{k+1} = (1 - \lambda_k)(\bar{b} - \bar{a}).$$

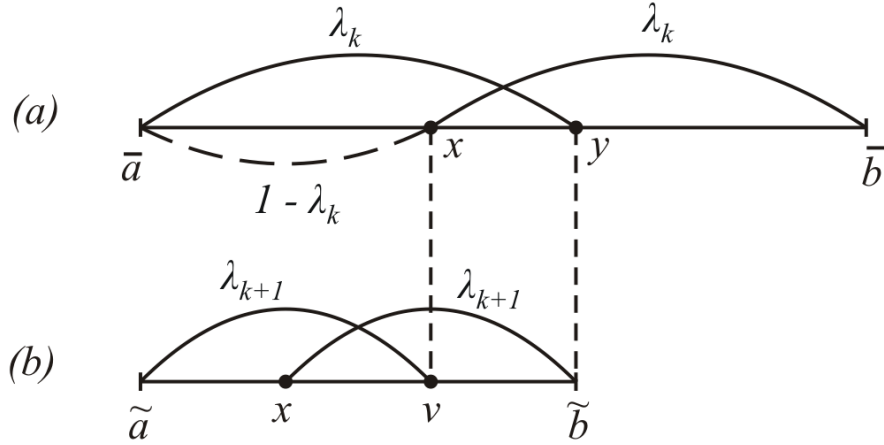


Рис. 4.9. Две последовательные стадии (а) и (в) применения симметричного метода, в котором левая точка становится правой, а правая — левой

Отсюда находим зависимость:

$$\lambda_k = \frac{1}{1 + \lambda_{k+1}}, \quad (4.15)$$

справедливую для всех симметричных методов, у которых точка, которая была правой в паре  $x, y$ , становится на следующей итерации левой и наоборот.

Заметим, что можно рассмотреть методы с обратным поведением точек испытаний. Однако, как легко показать, лучшую эффективность можно получить именно для рассматриваемого нами варианта.

Определим гарантированную эффективность симметричных методов, основанных на соотношении (4.15)

$$\begin{aligned} d^*(\alpha_n^{\text{сим}}) &= \sup_{Q \in \Phi_{un}[a,b]} d(\alpha_n^{\text{сим}}, Q) = (b-a)\lambda_2\lambda_3\dots\lambda_n = \\ &= (b-a)\lambda_n \cdot \frac{1}{1+\lambda_n} \cdot \frac{1+\lambda_n}{2+\lambda_n} \cdot \frac{2+\lambda_n}{3+2\lambda_n} \cdot \dots \cdot \frac{F_{n-3}+F_{n-4}\lambda_n}{F_{n-2}+F_{n-3}\lambda_n} = (b-a) \frac{\lambda_n}{F_{n-2}+F_{n-3}\lambda_n}. \end{aligned}$$

Нетрудно видеть, что существует закономерность в значениях коэффициентов  $F_i$ , причем эта закономерность отвечает правилам формирования последовательности Фибоначчи:

$$F_0 = F_1 = 1, F_2 = F_1 + F_0 = 2, \dots, F_{k+1} = F_k + F_{k-1}.$$

Из предыдущего очевидно, что

$$\lambda_k = \frac{F_{n-k-1} + F_{n-k-2} \cdot \lambda_n}{F_{n-k} + F_{n-k-1} \cdot \lambda_n}. \quad (4.16)$$

Значение гарантированной эффективности перепишем в виде

$$d^*(\alpha_n^{\text{сим}}) = \frac{(b-a)}{F_{n-3} + \frac{F_{n-2}}{\lambda_n}}, \quad (4.17)$$

где  $\frac{1}{2} < \lambda_n < 1$ .

Для определения метода, оптимального в классе  $n$ -шаговых симметричных, необходимо найти значение параметра  $\lambda_n = \lambda_n^* \in (\frac{1}{2}; 1)$ , обеспечивающее точную нижнюю грань выражения (4.17). Очевидно, что

$$\inf_{\frac{1}{2} < \lambda_n < 1} d^*(\alpha_n^{\text{сим}}) = \lim_{\lambda_n \rightarrow 1/2} \frac{(b-a)}{F_{n-3} + \frac{F_{n-2}}{\lambda_n}} = \frac{b-a}{2F_{n-2} + F_{n-3}} = \frac{b-a}{F_{n-2} + F_{n-1}} = \frac{b-a}{F_n}.$$

Поскольку значение инфимума не достигается ни при каком  $\lambda_n \in (\frac{1}{2}, 1)$ , приходим к выводу, что *оптимального алгоритма не существует*. Построим  $\varepsilon$ -оптимальный. Для этого достаточно выбрать  $\lambda_n^o = \frac{1}{2} + \delta$ , где  $\delta = \delta(\varepsilon)$  — достаточно малое положительное число. Однако, это приведет к тому, что  $\delta$  войдет в выражения для всех  $\lambda_k^o$ , что неудобно в использовании. Поступим иначе, а именно, примем формально  $\lambda_n^* = \frac{1}{2}$ . Тогда для остальных  $\lambda_k^*$ , ( $k = 2, 3, \dots, n-1$ ) из (4.16) получим компактные соотношения вида:

$$\lambda_k^* = \frac{F_{n-k+1}}{F_{n-k+2}}. \quad (4.18)$$

Причем для  $k = 2, \dots, n-1$  будет выполнено:  $\lambda_k^* \in (\frac{1}{2}, 1)$ .

Однако использовать  $\lambda_n^* = \frac{1}{2}$  на последней итерации для размещения последнего измерения не удастся, т.к. оно попало бы в центр оставшегося интервала и при этом совпало бы с уже имеющейся там точкой прежнего измерения. Поэтому на последнем шаге придется нарушить симметрию метода и разместить точку последнего испытания с  $\varepsilon$ -сдвигом от центра оставшегося интервала.

Таким образом, мы построили метод, называемый методом Фибоначчи. Его гарантированная эффективность равна

$$d^*(\lambda_n^{\text{Фиб}}) = \frac{b-a}{F_n} + \varepsilon. \quad (4.19)$$

*Алгоритм метода Фибоначчи для решения задач с заданной точностью*

**Шаг 0.** Задать параметры  $\varepsilon$  и  $\delta$ , причем  $0 < \varepsilon \ll \delta$ , где  $\delta$  — требуемая точность решения исходной задачи.

**Шаг 1.** Определить необходимое число измерений  $n$ , используя (4.19):

$$n = \min \{s : (b-a)/F_s + \varepsilon \leq \delta\}.$$

**Шаг 2.** Применить симметричный метод, размещающий точки испытаний с использованием пропорций  $\lambda_k = \lambda_k^*$  из (4.18) при  $k = 2, 3, \dots, n-1$ .

**Шаг 3.** Последнее  $n$ -е измерение провести с  $\varepsilon$ -смещением от середины интервала, оставшегося после выполнения шага 2 (рис. 4.10).

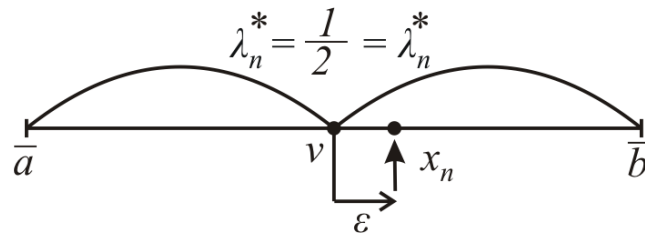


Рис. 4.10. Смещение точки последнего испытания  $x_n$  в методе Фибоначчи от ранее проведенного, находящегося в центре предпоследнего интервала  $[\bar{a}, \bar{b}]$

**Замечание.** Метод Фибоначчи был получен как  $\varepsilon$ -оптимальный в классе симметричных  $n$ -шаговых алгоритмов. В действительности можно доказать, что он является  $\varepsilon$ -оптимальным в классе всех последовательных  $n$ -шаговых алгоритмов на множестве унимодальных задач.

#### 4.3.4. Метод золотого сечения

Метод золотого сечения является более простым, чем метод Фибоначчи, но близким к нему по эффективности последовательным симметричным методом. Пропорции деления отрезка  $\lambda_k$  в методе золотого сечения не зависят ни от числа  $n$  предполагаемых измерений, ни от желаемой точности  $\delta$  определения решения  $x^*$ . Поэтому его часто используют во встроенных процедурах одномерной оптимизации, применяемых для определения шагового множителя в ряде методов многомерной локальной оптимизации (см. п.4.5.1). Это связано с тем, что в контексте такого применения требуется, с одной стороны, экономия числа измерений, а с другой — останов поиска по критериям, отличающимся от заданной точности по длине интервала, содержащего точку минимума функции одномерного сечения. В этих условиях метод Фибоначчи, очевидно, не может быть применен.

Ранее в (4.15) для класса симметричных методов было получено соотношение, связывающее пропорции деления интервала на двух последовательных шагах:

$$\lambda_k = \frac{1}{1 + \lambda_{k+1}}. \quad (4.20)$$

Это соотношение является нелинейным разностным уравнением, решениями которого являются последовательности  $\{\lambda_k\}$  с  $\lambda_k \in (\frac{1}{2}, 1)$ . Метод Фибоначчи использует решение в виде конечной последовательности  $\lambda_2^*, \lambda_3^*, \dots, \lambda_n^*$ , минимизирующей произведение  $\lambda_2 \cdot \lambda_3 \cdot \dots \cdot \lambda_n$ . Этому решению соответствует краевое условие  $\lambda_n^* = \frac{1}{2}$ , приводящее значение  $\lambda_n^*$  на границу допустимого интервала.

Теперь найдем другое, наиболее простое (стационарное) решение того же разностного уравнения, положив  $\lambda_k = \tau$ , ( $k = 1, 2, \dots$ ). Для этого подставим

$\lambda_k = \tau$  в уравнение (4.20), и получим квадратное уравнение

$$\tau^2 + \tau - 1 = 0.$$

Единственным корнем этого уравнения, удовлетворяющим условию  $\tau > \frac{1}{2}$  является число  $\tau = \frac{\sqrt{5}-1}{2} \approx 0,62$ .

Выберем  $\lambda_k \equiv \tau, \forall k = 1, 2, \dots$  и будем пользоваться симметричным методом с этим значением пропорций деления до момента, когда исчерпается заданное количество шагов  $n$ , либо до момента когда текущий интервал станет меньше заданного  $\delta$ , если решение требуется найти с точностью  $\delta$ .

Построенный метод получил название *метода золотого сечения*. Гарантированная эффективность этого метода при  $n$  измерениях определяется соотношением

$$d^*(\alpha_n^{\text{Зол.сеч}}) = \frac{b-a}{(1/\tau)^{n-1}}. \quad (4.21)$$

Заметим, что по построению (при  $\varepsilon \rightarrow 0$ ) всегда  $d^*(\alpha_n^{\text{Зол.сеч}}) > d^*(\alpha_n^{\text{Фиб.}})$ , т.е. метод золотого сечения уступает по эффективности методу Фибоначчи.

Чтобы выяснить, насколько близки по эффективности и поведению эти два метода, исследуем связь между пропорциями  $\lambda_k^*$  метода Фибоначчи и пропорциями  $\lambda_k^{\text{Зол.сеч}} = \tau$ . Построим диаграмму Кёнигса–Ламерея для соотношения (4.20), рассматривая его как точечное отображение отрезка  $(0, 1]$  в себя при преобразовании  $\lambda_{k+1}$  в  $\lambda_k$ .

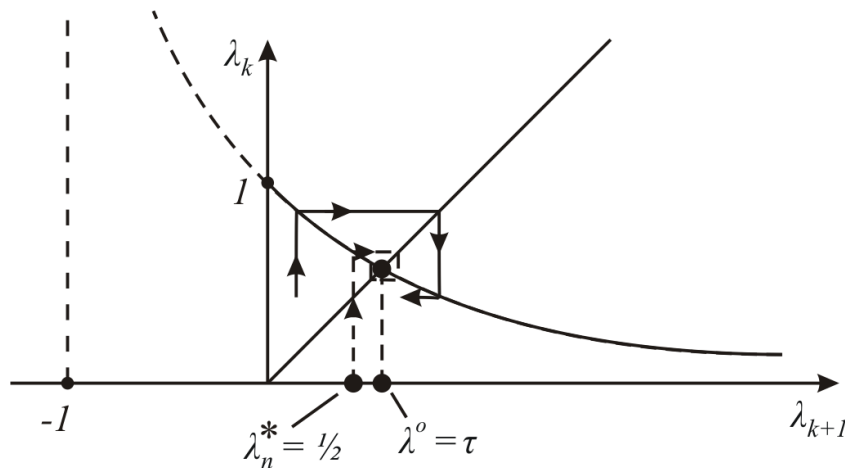


Рис. 4.11. Приближение начальных пропорций  $\lambda_k^*, \lambda_{k-1}^*, \dots$  метода Фибоначчи к пропорции золотого сечения  $\lambda^{\text{Зол.сеч}} = \tau = \lambda^o$  при больших  $n$

Поскольку модуль производной от правой части выражения (4.20) по  $\lambda_{k+1}$  всюду на  $(0, 1]$  меньше 1, отображение (4.20) на  $(0, 1]$  является сжимающим. Поэтому с увеличением числа итерации, связанных с пересчетами  $\lambda_{k+1}$  в  $\lambda_k$ , начальные (при малых  $k$ ) пропорции  $\lambda_k$  будут асимптотически приближаться при  $n \rightarrow \infty$  к неподвижной точке точечного отображения  $\lambda^o = \tau = \lambda^{\text{Зол.сеч}}$ .



Поэтому в методе Фибоначчи при достаточно больших значениях  $n$  на первых итерациях, т.е. при  $k = 2, 3, \dots$  пропорции  $\lambda_k^*$  берут близки к пропорциям метода золотого сечения  $\tau$  (см. рис. 4.11).

**Заключение.** Таким образом, если метод Фибоначчи собирается выполнить достаточно большое число итераций  $n$ , то его поведение на начальной стадии поиска будет близко к поведению метода золотого сечения.

Заметим дополнительно, что имеет место следующее асимптотическое соотношение при  $n \rightarrow \infty$ :  $F_n \approx 1,17 \cdot (\frac{1}{\tau})^{n-1}$ . Таким образом, учитывая (4.19) и (4.21), асимптотически при  $\varepsilon \rightarrow 0, n \rightarrow \infty$  имеем:

$$d^*(\alpha_n^{\text{Фиб.}}) \approx 0,85 \cdot d^*(\alpha_n^{\text{Зол.сеч}}).$$

#### 4.3.5. Сравнительная таблица эффективности различных методов

Таблица 4.1. Коэффициенты сокращения исходного интервала после  $n$  измерений при  $\varepsilon \rightarrow 0$

$n$	Оптимальный пассивный	Дихотомии-1 $\varepsilon \rightarrow 0$	Дихотомии-2	Фибоначчи $\varepsilon \rightarrow 0$	Золотого сечения
$2k + 1$	$\frac{n+1}{2}$ для $n = 2k + 1$	—	$2^{\frac{n-1}{2}}$	$F_n$	$(1/\tau)^{n-1}$
$2k$	$\frac{n}{2} + 1$ для $n = 2k$ при $\varepsilon \rightarrow 0$	$2^{\frac{n}{2}}$	—	$F_n$	$(1/\tau)^{n-1}$
2	2	2	—	2	1,6
3	2	—	2	3	2,6
4	3	4	—	5	4,2
5	3	—	4	8	6,8
6	4	8	—	13	10,9
7	4	—	8	21	17,6
8	5	16	—	34	28,4
9	5	—	16	55	45,8
10	6	32	—	89	73,9

Табл. 4.1 заполнена числами, показывающими во сколько раз сократилась длина исходного интервала после проведения нескольких измерений. Видно, что оптимальные пассивные методы уступают последовательным неоптимальным методам дихотомии. Лучшим оказывается последовательный метод

Фибоначчи, поскольку он  $\varepsilon$ -оптимален в классе последовательных. Метод золотого сечения близок к нему по эффективности, хотя несколько уступает.

В таблице сокращение интервала для методов дихотомии-1, дихотомии-2 и Фибоначчи приведено для случая  $\varepsilon \rightarrow 0$ .

#### 4.4. Построение методов поиска минимума дифференцируемой выпуклой функции на выпуклом многограннике

В этом разделе кратко, в обзорном плане, представлены некоторые методы поиска минимума строго выпуклых дифференцируемых функций.

Рассмотрим задачу:

$$\min_{x \in M} Q(x), \quad (4.22)$$

где  $Q(x)$  — дифференцируемая строго выпуклая функция,  $M$  — выпуклый многогранник

$$M = \{x \in R^N : Ax \leq b\}. \quad (4.23)$$

Пусть результатами испытаний в точках  $x$  являются значения функции  $Q_i = Q(x^i)$  и значения градиента этой функции  $\nabla Q_i = \nabla Q(x^i)$ .

##### Построение последовательно-оптимального одношагового метода

В силу критерия выпуклости для дифференцируемой функции:

$$\forall x \in R^N : Q(x) \geq Q_i + (\nabla Q_i, x - x^i). \quad (4.24)$$

Рассмотрим множество

$$D_i = \{x \in R^N : (\nabla Q_i, x - x^i) > 0\}.$$

Для  $x \in D_i$ :

$$Q(x) \geq Q_i + (\nabla Q_i, x - x^i) > Q_i \geq Q(x^*).$$

Следовательно, глобальный минимум  $x^* \notin D_i$ .

Из этого наблюдения вытекает следующее правило сокращения исходного допустимого множества путем отбрасывания подобластей, не содержащих глобального минимума  $x^*$ . Пусть  $M_0 = M$ , а после проведения  $(k - 1)$  испытаний множество  $M_0$  сократилось до  $M_{k-1}$ . Если проведено испытание в точке  $x^k$ , то

$$M_k = M_{k-1} \cap \{x \in R^N : (\nabla Q_k, x - x^k) \leq 0\}. \quad (4.25)$$

Для получения правила размещения новых испытаний  $x^{k+1}$  воспользуемся принципом последовательной одношаговой оптимальности (4.6). Введем

функцию эффективности, описывающую меру  $V(M_{k+1})$  оставшегося множества  $M_{k+1}$ , являющегося текущей оценкой решения. Пусть  $\bar{x}$  — предполагаемая точка нового испытания, а  $z \in R^N$  — гипотетическое значения вектора градиента функции  $Q$  в точке  $\bar{x}$ .

Тогда при наличии пары  $(\bar{x}; z)$  эффективность такого испытания на  $(k + 1)$ -м шаге будет определяться значением

$$d(\omega_k, \bar{x}, z) = V(M_k \cap \{x \in R^N : (z, x - \bar{x}) \leq 0\}).$$

Поскольку значение  $z$  до проведения испытания неизвестно и может быть любым, а на величину  $d(\cdot)$  влияет лишь ориентация этого вектора, то значение гарантированной эффективности определяется соотношением

$$d^*(\omega_k, \bar{x}) = \sup_{\|z\|=1} V(M_k \cap \{x \in R^N : (z, x - \bar{x}) \leq 0\}). \quad (4.26)$$

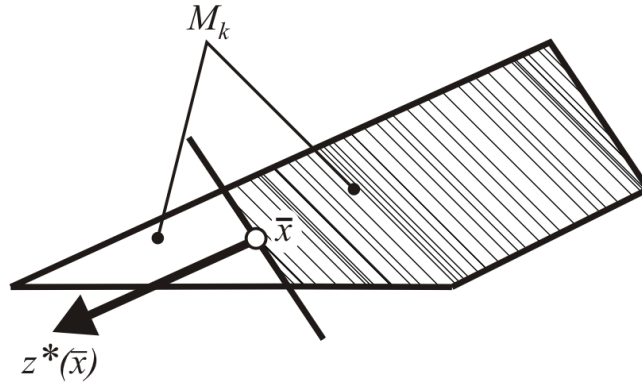


Рис. 4.12. Штриховкой отмечено множество меры  $d^*(\omega_k, \bar{x})$ , содержащее решение  $x^*$  для наихудшей ориентации  $z^*(\bar{x})$  — градиента функции при испытании в точке  $\bar{x}$

Значение  $z = z^*(\bar{x})$ , определяющее точную верхнюю грань в (4.26) приближенно показано на рис. 4.12.

Согласно принципам последовательной одношаговой оптимальности (4.6) приходим к следующему правилу

$$x^{k+1} = \arg \min \{d^*(\omega_k, \bar{x}) : \bar{x} \in M_k\}. \quad (4.27)$$

Точная реализация правила (4.27) в общем случае невозможна, однако в качестве эвристического правила, близкого к (4.27), может быть использован *метод центров тяжести*, согласно которому очередное испытание  $x^{k+1}$  следует проводить в центре тяжести оставшегося многогранника  $M_k$ . После выполнения испытания и вычисления  $\nabla Q_{k+1}$  выполняется сокращение многогранника  $M_k$  по правилу (4.25).

Поскольку определение центра тяжести произвольного многогранника вычислительно трудоёмко, А. С. Немировским и Д. Б. Юдиным [13] разработан *метод эллипсоидов*, приближенно реализующий идею описанного выше метода.

### Метод, основанный на сведениях к серии задач линейного программирования

Для поиска решений выпуклых задач могут быть использованы другие подходы, основанные на правиле выбора точек очередных испытаний  $x^{k+1}$  как оптимальных решений вспомогательных оценочных задач. Эти задачи строятся на основе минорант целевой функции, т. е. нижних оценок ее значений.

Минорантой (нижней оценкой) функции  $Q(x)$  будет кусочно-линейная функция  $Q_k^-(x)$ , получаемая на основе (4.24):

$$Q_k^-(x) = \max_{i=1, \dots, k} \{Q_i + (\nabla Q_i, x - x^i)\}. \quad (4.28)$$

Тогда для всякого  $x$  из множества, в котором задана функция  $Q$ , будем иметь  $Q(x) \geq Q_k^-(x)$ .

Миноранта  $Q_k^-(x)$  определяет нижнюю оценку возможных значений функции  $Q$  в точке  $x$ . Нижняя оценка построена по результатам предыдущих испытаний функции  $Q$ . Будем выбирать точку очередного испытания в допустимом множестве из условия достижения минимума нижней оценки функции:

$$x^{k+1} = \arg \min_{Ax \leq b} Q_k^-(x). \quad (4.29)$$

Построенную вспомогательную задачу (4.28)-(4.29) выбора  $x^{k+1}$  можно свести к задаче линейного программирования в расширенном пространстве переменных  $(x; z)$ , где  $z \in R^1$ :

$$(x^{k+1}; z^{k+1}) = \arg \min_{(x; z) \in D_k} z, \quad (4.30)$$

где  $D_k = \{x : Q_i + (\nabla Q_i, x - x^i) \leq z, (i = 1, \dots, k); Ax \leq b\}$ .

За текущую оценку решения примем наименьшее вычисленное значение функции  $Q_k^* = \min \{Q_i : i = 1, \dots, k\}$  и соответствующую ему точку испытания  $x_k^*$ . Геометрический смысл правила (4.30) иллюстрирует рис. 4.13.

Критерием останова метода является достаточная малость гарантированной погрешности текущей оценки  $Q_k^*$  минимального значения функции:

$$\min_{i=1, \dots, k} Q_i - z^{k+1} \leq \varepsilon. \quad (4.31)$$

Таким образом, при этом подходе задача выпуклого программирования (4.22), (4.23) сводится к решению последовательности задач линейного программирования.

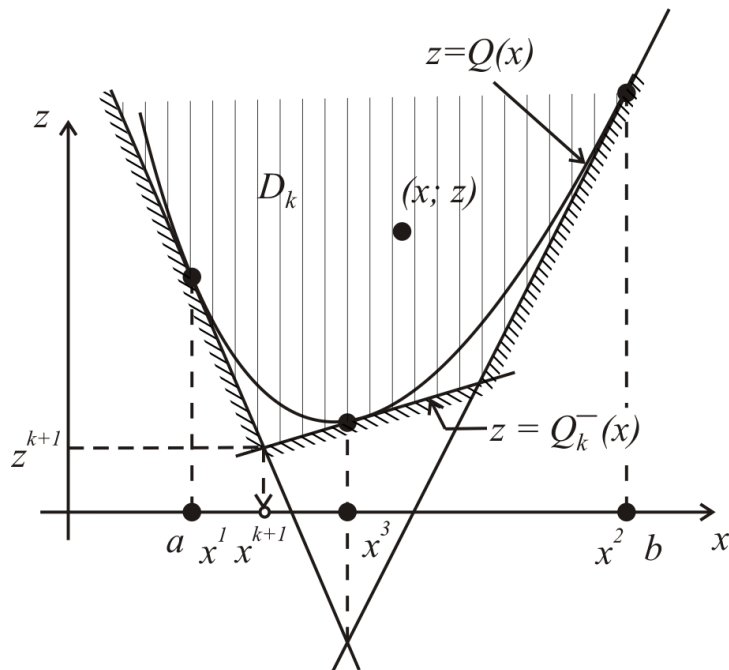


Рис. 4.13. Вид вспомогательной задачи (4.30) при числе переменных  $N = 1$  и количестве проведенных испытаний  $k = 3$

**Теорема (сходимости).** Для задачи поиска глобального минимума строго выпуклой дифференцируемой функции на выпуклом компактном многограннике (4.22), (4.23) метод (4.30), (4.31) при  $\varepsilon > 0$  остановится через конечное число шагов, и полученная к моменту останова оценка  $Q_k^*$  минимального значения функции на многограннике (4.23) будет иметь погрешность не более  $\varepsilon$  по значению функции  $Q$ .

*Доказательство* не приводится, но оно непосредственно вытекает из более общей теоремы 3 в [14] — § 6.2.

**Замечание.** Метод (4.30), (4.31) применим также для выпуклых недифференцируемых функций  $Q$ , если в правиле выбора точки  $x^{k+1}$  очередного испытания (4.30) в выражении нижней оценки функции заменить значения градиента функции на значения ее субградиента.

## 4.5. Методы поиска локального минимума в задачах без функциональных ограничений

В этом разделе рассмотрены методы поиска локально-оптимальных решений для достаточно общих классов задач, характеризующихся, как правило, лишь гладкостью некоторого порядка. Также приведены некоторые методы решения задач с кусочно непрерывными функциями. Для задач обоих типов невозможно определить погрешность текущей оценки решения по конечному числу испытаний, что не позволяет ввести функцию эффективности. Это, в

свою очередь, не дает возможности использовать при построении численных методов принципы оптимальности, которые успешно применялись нами до этого.

Необходимы иные подходы к построению методов (см. [2], [5], [15], [17], [20]-[24]). Обычно в общих задачах поиска локального минимума используется *принцип локального спуска*, заключающийся в том, что для последовательности основных испытаний  $x^k$  обеспечивается убывание значений целевой функции  $Q(x^k) > Q(x^{k+1})$ , при этом убывание должно быть «существенным».

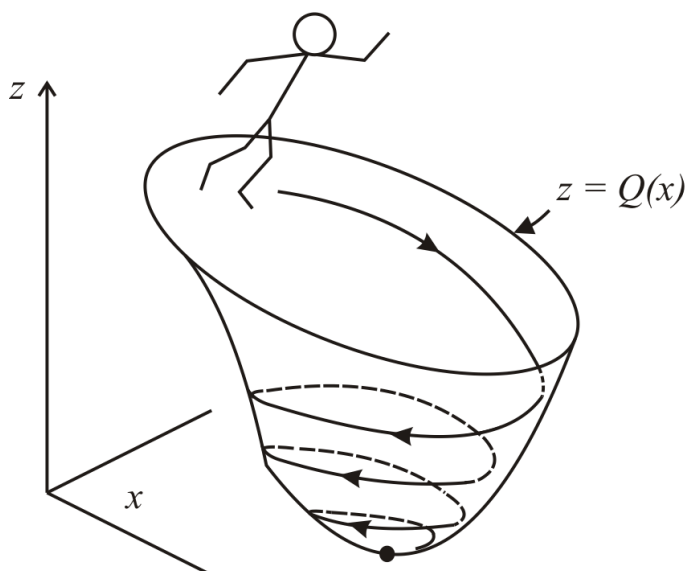


Рис. 4.14. Метод локальной оптимизации подобен слепому (доступна только локальная информация о свойствах поверхности: высота, наклон, кривизна), спускающемуся в низину

Требование «существенности» убывания должно обеспечивать сходимость к локально-оптимальному решению, поскольку простое строгое убывание, очевидно, недостаточно для обеспечения сходимости (см. рис.4.14). Важной характеристикой процесса сходимости является *порядок скорости сходимости* (понятие будет введено ниже).

Итак, рассмотрим задачу математического программирования без ограничений

$$\min_{x \in R^N} Q(x),$$

где  $Q$  принадлежит некоторому классу функций  $\Phi$ . Обычно  $\Phi = C^1$  или  $\Phi = C^2$ . Зададим также начальную точку поиска  $x^0$ . Несмотря на неограниченность области поиска всегда будем предполагать, что локальный минимум  $x^o$  в задаче существует, и начальная точка  $x^0$  (для применяемого численного метода) находится в *области притяжения*  $x^o$  для этого метода.

Заметим, что хоть область притяжения локального минимума зависит от применяемого метода оптимизации, это понятие создает полезный геометрический образ.

Использование принципа локального спуска порождает методы сходной структуры. Их можно называть *траекторными*. Почти все методы такого типа могут быть представлены в виде итерационного соотношения:

$$x^{k+1} = x^k + t_k d^k,$$

где  $x^k$  — точки *основных испытаний*, состоящих в вычислении  $J_k = J(x^k)$  — набора тех или иных *локальных характеристик* целевой функции в точке  $x^k$ ,  $d^k$  — направление смещения из точек  $x^k$ , а  $t_k$  — коэффициенты, определяющие величины смещений вдоль выбранных направлений (*шаговые множители*). Для нахождения величин смещений  $t_k$  вдоль направлений  $d^k$  методы могут выполнять *вспомогательные (рабочие) шаги*, приводящие к дополнительным испытаниям — измерениям локальных характеристик целевой функции вдоль направлений  $d^k$  (см. рис. 4.15). Заметим, что порядок испытаний на рабочих шагах может быть ниже порядка основных испытаний  $J_k$ .

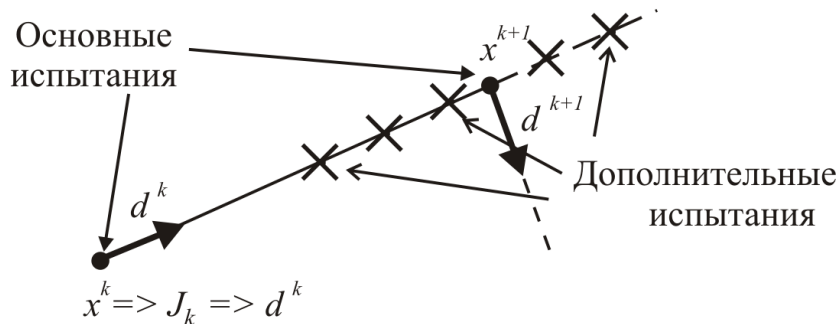


Рис. 4.15. Структура траекторного метода локальной оптимизации

Основные испытания  $J_k$ , как правило, определяют выбор направлений  $d^k$ , а рабочие испытания — выбор величины шагового множителя  $t_k$ .

**Определение.** *Порядком численного метода оптимизации* называют порядок его основных испытаний, т.е. наивысший порядок вычисляемых при этом производных.

Введем несколько определений, касающихся характера сходимости методов. Пусть  $x^*$  — решение задачи.

**Определение.** *Линейной* называют сходимость, при которой мера близости к решению  $\|\Delta x^k\| = \|x^k - x^*\|$  убывает по закону геометрической прогрессии с некоторым знаменателем  $0 < q < 1$ . Таким образом,  $\|\Delta x^k\|$  и  $\|\Delta x^{k+1}\|$  являются бесконечно малыми одного порядка при  $k \rightarrow \infty$ . Начиная с некоторого шага  $k$ :  $\|\Delta x^{k+1}\| \leq q \|\Delta x^k\|$ .

**Определение.** *Сверхлинейной* называют сходимость более высокого порядка, чем у любой геометрической прогрессии. При сверхлинейной сходимости  $\|\Delta x^k\| \rightarrow 0$  при  $k \rightarrow \infty$  и  $\|\Delta x^{k+1}\|$  имеет более высокий порядок малости по отношению к  $\|\Delta x^k\|$  при  $\|\Delta x^k\| \rightarrow 0$ , т.е.  $\frac{\|\Delta x^{k+1}\|}{\|\Delta x^k\|} \rightarrow 0$  при  $k \rightarrow \infty$ . Таким образом, можно сказать, что метод *сходится к  $x^*$  сверхлинейно*, если при достаточной близости точки  $x^0$  к  $x^*$  найдется такое  $K$  и последовательность чисел  $q_1, q_2, \dots, q_s$  из интервала  $[0, 1)$ , стремящаяся к 0 при  $s \rightarrow \infty$ , что для всякого  $k \geq K$  и любого  $s > 0$  будет выполнено неравенство

$$\|\Delta x^{k+s}\| \leq q_1 q_2 \dots q_s \|\Delta x^k\|.$$

**Определение.** *Квадратичной* называют сверхлинейную сходимость, при которой мера погрешности на следующем шаге  $\|\Delta x^{k+1}\|$  является величиной второго порядка малости по отношению к  $\|\Delta x^k\|$ , при  $\|\Delta x^k\| \rightarrow 0$  с  $k \rightarrow \infty$  т.е.  $\exists C > 0$ , что начиная с некоторого шага  $K$  при достаточной малости  $\|\Delta x^0\|$  для всякого  $k \geq K$ :

$$\|\Delta x^{k+1}\| \leq C \|\Delta x^k\|^2,$$

а при более высоком показателе степени в правой части неравенства подобное соотношение уже не выполняется.

#### 4.5.1. Критерии и алгоритмы выбора шагового множителя

В общем случае величина шагового множителя  $t_k$  выбирается алгоритмически, исходя из некоторых заданных критериев. Они подбираются так, чтобы удовлетворяющий им шаговый множитель не мог стать слишком большим, но, с другой стороны, не был бы слишком относительно малым. Это нужно для обеспечения сходимости методов.

**Критерий 1.** *Критерий существенности убывания функции.*

Он играет роль сдерживающего фактора, препятствующего неограниченному увеличению  $t_k$ . В основе критерия лежит требование

$$t_k \in \Pi_1(\mu), \quad 0 < \mu < 1, \quad (4.32)$$

$$\Pi_1(\mu) = \{t \geq 0 : Q(x^k + td^k) \leq Q_k + \mu(\nabla Q_k, d^k)t\}. \quad (4.33)$$

**Критерий 2.** *Критерий близости к минимуму по направлению.*

В его основе лежит требование, чтобы в точке  $x^k + t_k d^k$  модуль скорости изменения функции  $Q$  в направлении  $d^k$  был в заданное число раз меньше модуля скорости ее изменения в точке  $x^k$ . Это требование формализуется следующим образом. Задается малый положительный коэффициент  $\eta$ , и величина  $t^k$  определяется условием



$$t_k \in \Pi_2(\eta), \quad 0 < \eta < 1, \quad (4.34)$$

$$\Pi_2(\eta) = \{t \geq 0 : |(\nabla Q(x^k + td^k), d^k)| \leq \eta |(\nabla Q_k, d^k)|\}. \quad (4.35)$$

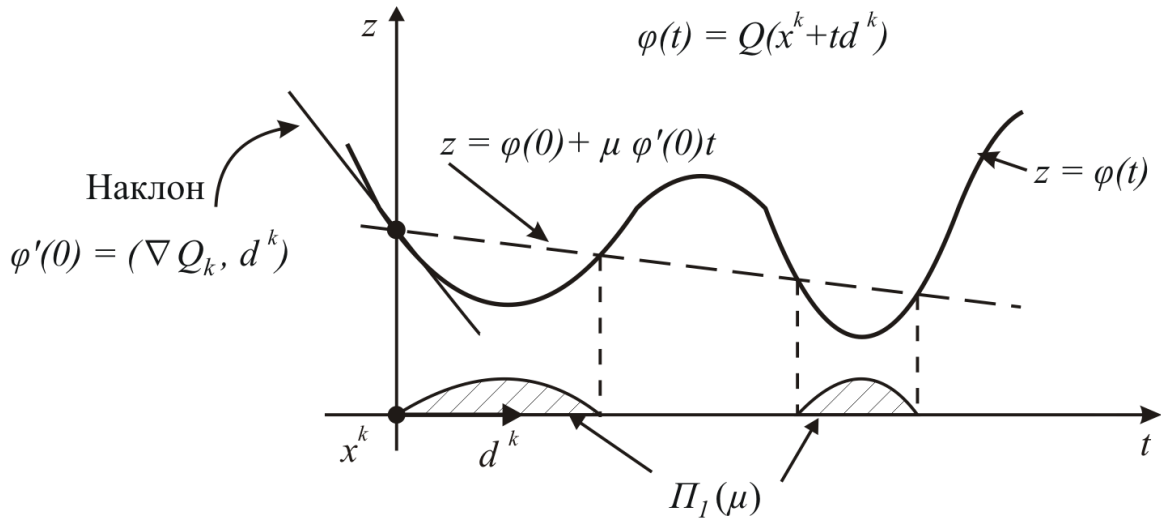


Рис. 4.16. Множество значений  $t$ , удовлетворяющих критерию 1

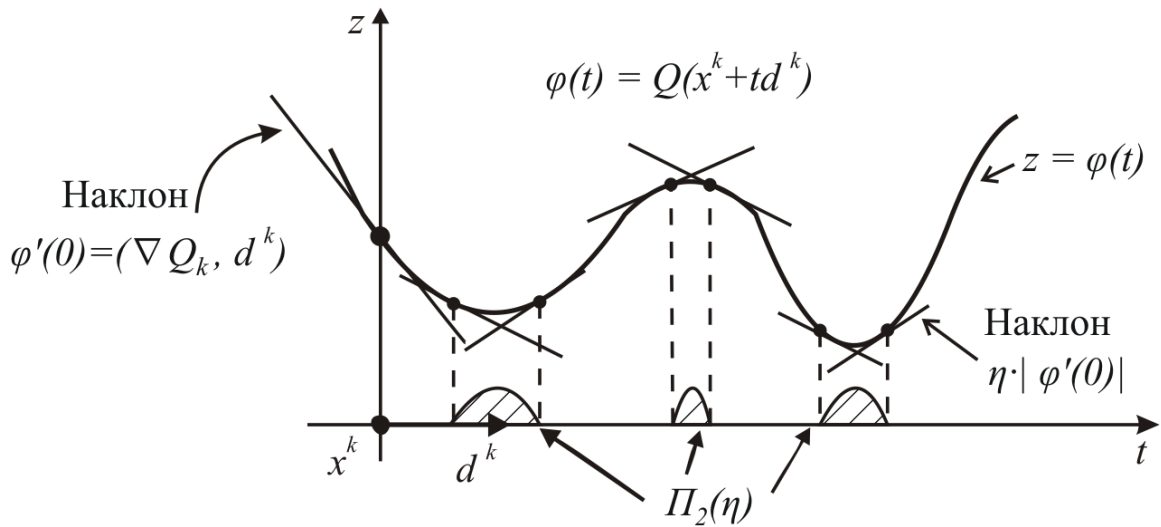


Рис. 4.17. Множество значений  $t$ , удовлетворяющих критерию 2

### Критерий 3. Модифицированный критерий 2

$$t_k \in \Pi_3(\eta), \quad 0 < \eta < 1, \quad (4.36)$$

$$\Pi_3(\eta) = \{t \geq 0 : (\nabla Q(x^k + td^k), d^k) \geq \eta \cdot (\nabla Q_k, d^k)\}. \quad (4.37)$$

Рассмотрим несколько правил выбора шагового множителя и реализующие их алгоритмы.

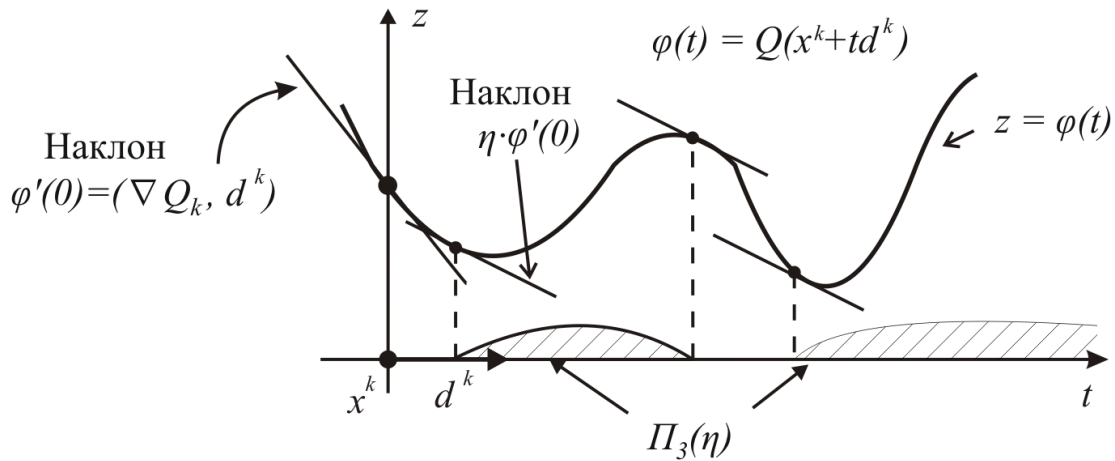


Рис. 4.18. Множество значений  $t$ , удовлетворяющих критерию 3

### Правило Армихо

Если вычисления градиента функции дорогостоящие, то при проведении дополнительных испытаний может оказаться выгоднее использовать простые алгоритмы выбора шагового множителя, в которых градиент измеряется только единожды. Самым простым является *правило Армихо*. Оно может быть описано алгоритмически.

*Алгоритм правила Армихо*

**Шаг 0.** Задаем параметры  $\alpha > 0$ ,  $0 < \mu < 1/2$  и  $0 < \gamma < 1$ .

**Шаг 1.** Полагаем  $t = \alpha$ .

**Шаг 2.** Пока  $t \notin \Pi_1(\mu)$ , полагаем  $t := \gamma t$ .

**Шаг 3.** Принимаем  $t_k = t$ .

**Замечание.** Требование  $0 < \mu < 1/2$  нужно для того, чтобы для функций  $\varphi(t)$ , имеющих вид параболы с  $\varphi''(t) > 0$  (такой вид они будут иметь при достаточной близости  $x^k$  к локальному минимуму дважды непрерывно дифференцируемой функции  $Q(x)$ ), значение  $t_k$ , соответствующее точке её минимума, удовлетворяло критерию 1.

*Доказательство* проведите самостоятельно.

Это правило может сочетаться со способами выбора  $d^k$  в ряде численных методов, обеспечивая их сходимость. Например, правило Армихо можно использовать в методах градиентного поиска, в методах Ньютона с регулировкой шага (желательно при  $\alpha = 1$ ). Однако его нельзя применять в квазиньютоновских методах и методе сопряженных градиентов Флетчера–Ривса.

### Правило одномерной минимизации

При использовании этого правила значение шагового множителя выбирается согласно (4.38) с использованием методов поиска минимума функции

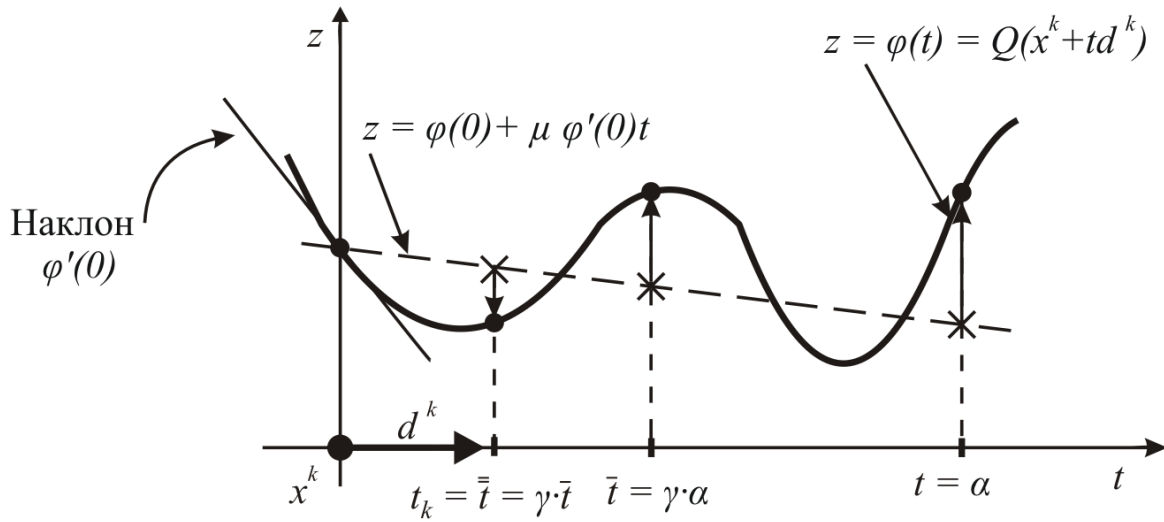


Рис. 4.19. Пример выбора шагового множителя  $t_k$  по правилу Армихо при  $\gamma = 1/2$

одной переменной,

$$t_k \in \Pi_1(\mu) \cap \Pi_2(\eta), \quad 0 < \mu < \eta < 1. \quad (4.38)$$

При этом требование  $\mu < \eta$  обеспечивает непустоту пересечения  $\Pi_1(\mu)$  и  $\Pi_2(\eta)$ .

Если правило (4.38) применяется с  $\eta \ll 1$ , то говорят, что шаговый множитель выбирается из условия «аккуратного» одномерного поиска.

При исследовании сходимости методов обычно принимается, что выбор шагового множителя из условия «аккуратного» одномерного поиска соответствует определению  $t_k$  из условия точной одномерной минимизации  $Q(x)$  на луче  $x^k + td^k$  при  $t \geq 0$ , т.е. выбору

$$t_k = \arg \min_{t \geq 0} \varphi(t) = \arg \min_{t \geq 0} Q(x^k + td^k). \quad (4.39)$$

Заметим, что правило (4.39) не может быть реализовано абсолютно точно. «Аккуратный» одномерный поиск на основе (4.33), (4.35), (4.38) следует рассматривать как его приближенную численную реализацию.

Определение  $t_k$  из условия (4.38) выполняется в два этапа.

**Этап 1.** Определяется отрезок  $[0, T]$ , содержащий точку локального минимума  $\varphi(t)$ , ближайшую к точке 0.

**Шаг 0.** Задаем малое  $0 < \delta \ll 1$ , полагаем  $\bar{t} = 0$ ,  $\bar{\bar{t}} = \delta$ .

**Шаг 1.** Пока  $\varphi(\bar{t}) > \varphi(\bar{\bar{t}})$  и  $\varphi'(\bar{\bar{t}}) \leq 0$ , полагаем  $\bar{t} = \bar{\bar{t}}$ ,  $\delta := 2 \cdot \delta$ ,  $\bar{\bar{t}} := \bar{\bar{t}} + \delta$ .

**Шаг 2.** После завершения циклической процедуры шага 1 принимаем  $T = \bar{\bar{t}}$ .

**Этап 2.** При  $0 < \mu < \eta < 1$  выполняется поиск минимума  $\varphi(t) = Q(x^k + td^k)$  на отрезке  $t \in [0, T]$  с нестандартным критерием останова. Процесс останавливается, как только первый раз при работе метода поиска минимума встретится значение  $t$ , удовлетворяющее условию (4.38). Это значение  $t$  принимается в качестве величины шагового множителя  $t_k$ .

Как правило, поиск минимума на этапе 2 выполняется в предположении, что  $\varphi(t)$  — унимодальная. Например, может быть применен метод золотого сечения и его модификации.

Правило одномерной минимизации сочетается практически с любым способом выбора  $d^k$ . Однако, это правило вычислительно наиболее дорогостоящее, поскольку требует многократного измерения градиента функции. Существуют методы локальной оптимизации, в которых определение минимума в сечениях на шаге не является обязательным, но простое правило Армихо применять нельзя. При этом, как правило, можно использовать *правило Вулфа*, которое вычислительно более экономично, чем одномерная минимизация.

### Правило Вулфа

По этому правилу выбор шагового множителя должен подчиняться условию  $t_k \in \Pi_1(\mu) \cap \Pi_3(\eta)$ ,  $0 < \mu < \eta < 1$ .

Сравнение рис. 4.17 и 4.18 наглядно показывает, что область возможных значений  $t_k$  в правиле Вулфа значительно шире, чем в правиле одномерной минимизации. Это позволяет определять значение шагового множителя  $t_k$  с меньшими затратами. Реализация требует специального алгоритма (см. [14]).

#### Алгоритм реализации правила Вулфа

**Шаг 0.** Задаем начальное  $\alpha > 0$ ,  $0 < \mu < \eta < 1$ , коэффициент «интерполяции»  $0 < \gamma < 1$  и коэффициент «экстраполяции»  $r > 1$

**Шаг 1.** Полагаем  $[\check{\alpha}; \hat{\alpha}] = [0; 0]$ ,  $t = \alpha$ .

**Шаг 2.**

Пока $t \notin \Pi_1(\mu) \cap \Pi_3(\eta)$	выполняем	Пока $t \notin \Pi_1(\mu)$	
		выполняем	$\hat{\alpha} = t$
			$t := \check{\alpha} \cdot \gamma + \hat{\alpha} \cdot (1 - \gamma)$
		Пока $t \notin \Pi_3(\eta)$	
	выполняем		$\check{\alpha} = t$
			Если $\hat{\alpha} = 0$ то $\{t := \check{\alpha} \cdot r; break; \}$
			иначе $t := \check{\alpha} \cdot \gamma + \hat{\alpha} \cdot (1 - \gamma)$

**Шаг 3.** Полагаем  $t_k = t$ .

На рис. 4.20-4.21 представлены иллюстрации работы данного алгоритма для случая малого начального  $\alpha$  и большого начального  $\alpha$ .

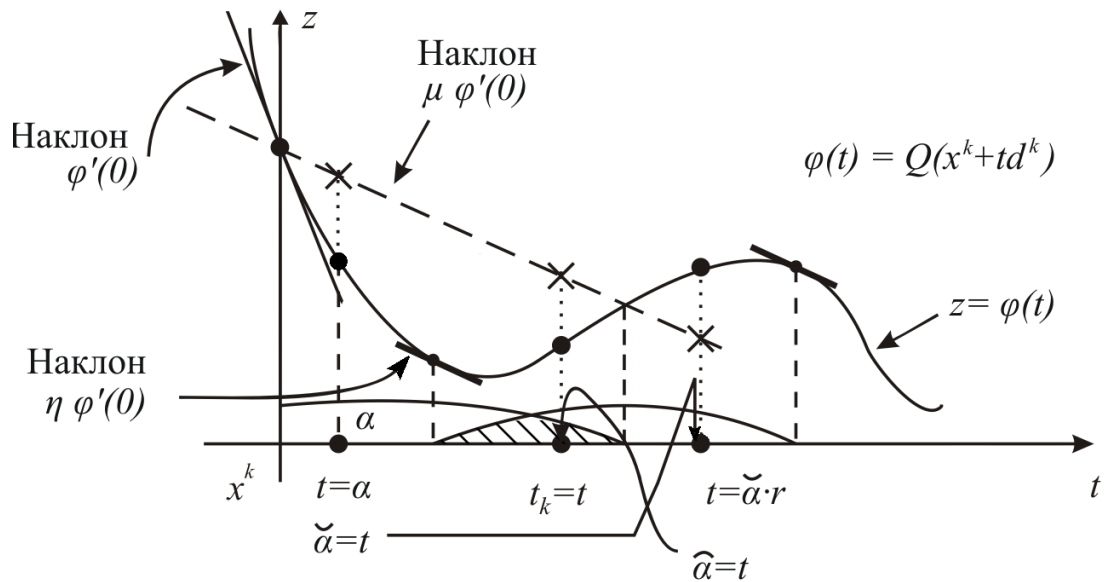


Рис. 4.20. Выполнение алгоритма правила Вульфа при малом  $\alpha$  и  $r = 5$ ,  $\gamma = 1/3$

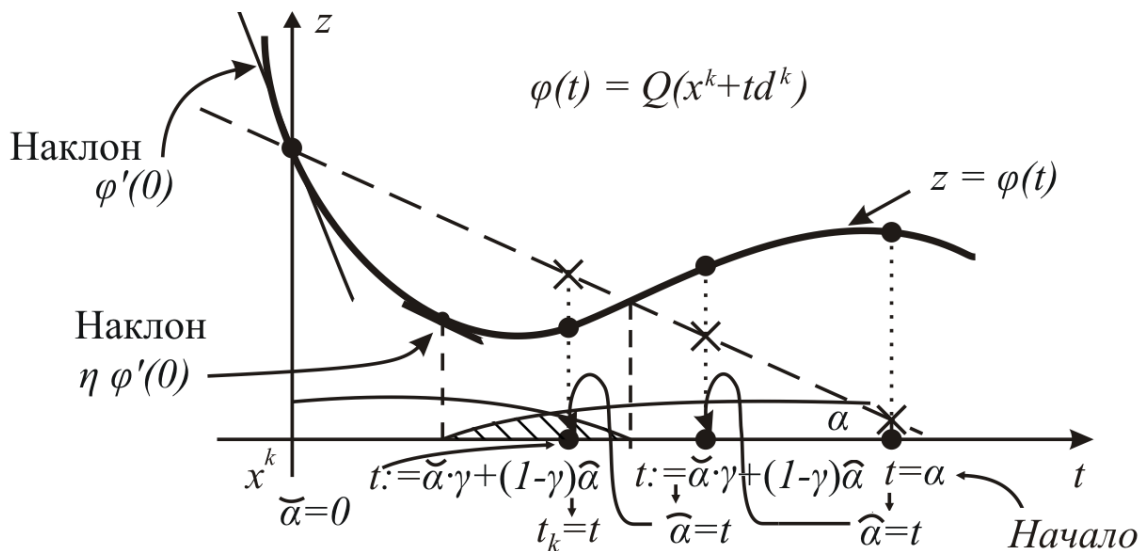


Рис. 4.21. Выполнение алгоритма правила Вульфа при большом  $\alpha$  и  $\gamma = 1/3$

Правило Вульфа можно применять в методе градиентного поиска, в методе Ньютона с регулировкой шага, в некоторых квазиньютоновских методах. Нельзя использовать в методе сопряженных градиентов Флетчера–Ривса. Метод сопряженных градиентов Флетчера–Ривса совместим только с правилом одномерной минимизации в варианте «аккуратного» одномерного поиска.

**Замечание.** Для корректной определенности алгоритмов выбора шагового множителя достаточно требовать от функции одномерного сечения непрерывной дифференцируемости и ограниченности снизу.

### 4.5.2. Простые методы градиентного поиска

Простые градиентные методы являются «идеологически» важными, хотя никогда не применяются в «чистом» виде. Они служат основой построения более совершенных методов. В данном разделе рассматривается задача минимизации функции  $Q$  в пространстве  $R^N$ . Пусть целевая функция  $Q(x) \in C^1$ , т.е. принадлежит классу непрерывно дифференцируемых функций. Дополнительные требования будут наложены позднее.

Градиентные методы (*называемые также методами градиентного поиска*) относятся к *методам первого порядка*, т.к. используют испытания вида  $J_k = (Q_k; \nabla Q_k)$ , включающие вычисление первых производных. Их можно описать общим итерационным соотношением

$$x^{k+1} = x^k + t_k d^k,$$

где  $d^k$  — направление смещения из точки  $x^k$ , совпадающее с направлением антиградиента, то есть

$$d^k = -\nabla Q_k.$$

Тем самым, при выборе направления  $d^k$  фактически используется локальная линейная модель поведения функции. Шаговые множители  $t_k$  могут выбираться различным образом при возможности обеспечения сходимости.

**Определение.** При выборе шагового множителя  $t_k$ , из условия достижения минимума в сечении, т.е. из условия

$$t_k = \arg \min_{t \geq 0} Q(x^k + t d^k), \quad (4.40)$$

градиентный метод называют *методом наискорейшего градиентного поиска*.

Очевидно, такой способ выбора  $t_k$  удовлетворяет требованию (4.38) при достаточной малости  $\mu$ .

Подчеркнем, что правило (4.40) точно вычислительно не реализуемо. Его приближенная реализация обеспечивается алгоритмом «аккуратного» одномерного поиска, исходя из требования (4.38), но при условии  $0 < \mu < \eta \ll 1$ , т.е. когда значение  $\mu$  и  $\eta$  выбираются весьма близкими к нулю.

В качестве текущей оценки решения в методе используется точка последнего испытания. В качестве критерия останова может быть использовано условие  $\|\nabla Q(x^k)\| \leq \varepsilon$ , определяющее малость невязки в необходимом условии оптимальности из теоремы Ферма.

**Теорема (сходимости).** Пусть функция  $Q(x)$  непрерывно дифференцируема, ограничена снизу и ее градиент удовлетворяет условию Липшица, т.е.  $\exists L > 0$ , что:

$$\forall x', x'' : \|\nabla Q(x') - \nabla Q(x'')\| \leq L \|x' - x''\|.$$

Тогда метод градиентного поиска из любой начальной точки  $x^0$  строит последовательность  $x^k$  такую, что  $\|\nabla Q(x^k)\| \rightarrow 0$  для  $k \rightarrow \infty$  при следующих вариантах выбора шагового множителя:

- a) при постоянном шаговом множителе  $t^k = \alpha$ , где  $0 < \alpha < 1/L$ ;
- b) по правилу Армихо при  $0 < \mu < \frac{1}{2}$ ;
- c) по правилу Вулфа при  $0 < \mu < \eta < 1$ ;
- d) по правилу одномерной минимизации при  $0 < \mu < \eta < 1$ ;
- e) по правилу наискорейшего градиентного поиска (4.40) (реализуются методом «аккуратного» одномерного поиска при  $0 < \mu < \eta \ll 1$ ).

*Доказательство.* Проведем доказательство для случая (e). Остальные случаи предлагается доказать самостоятельно.

Рассмотрим точки, лежащие на луче, порожденном направлением антиградиента:  $x = x(t) = x^k - t\nabla Q(x^k)$ ,  $t \geq 0$ . Построим верхнюю оценку для приращения функции.

Используя теорему о среднем из математического анализа, можно утверждать, что  $\exists \xi \in [x, x^k]$ , для которого выполнено первое из указанных ниже равенств. Полученное выражение можно оценить сверху, учитывая, что  $x = x^k - t\nabla Q_k$ :

$$\begin{aligned} Q(x) - Q_k &= (\nabla Q(\xi), x - x^k) = (\nabla Q(\xi) \pm \nabla Q_k, x - x^k) = \\ &= (\nabla Q(\xi) - \nabla Q_k, x - x^k) - t(\nabla Q_k, \nabla Q_k) \leq \\ &\leq \|\nabla Q(\xi) - \nabla Q_k\| \cdot \|x - x^k\| - t\|\nabla Q_k\|^2 \leq \\ &\leq L\|x - x^k\|^2 - t\|\nabla Q_k\|^2 \leq \underbrace{Lt(t - \frac{1}{L})}_{\psi(t)} \|\nabla Q_k\|^2. \end{aligned}$$

Значение  $x$  в левой части полученного неравенства зависит от  $t$ . Неравенство сохранится, если в обеих его частях перейти к минимально возможным при  $t \geq 0$  значениям. В результате этой операции и с учетом правила выбора  $t_k$  в наискорейшем градиентном поиске получим:

$$Q_{k+1} - Q_k \leq L\psi(1/2L)\|\nabla Q_k\|^2 = -\frac{1}{4L}\|\nabla Q_k\|^2 \leq 0.$$

Возможно два варианта. Если на некотором шаге  $k$  выполнится  $\|\nabla Q_k\| = 0$ , то утверждение теоремы уже верно. Если же всегда  $\|\nabla Q_k\| > 0$ , то будет иметь место строго монотонное убывание последовательности значений функции на траектории поиска:  $Q_{k+1} < Q_k$ . В силу ограниченности снизу этих значений, получаем, что  $Q_{k+1} - Q_k \rightarrow 0$  при  $k \rightarrow \infty$ , но тогда из оценки следует, что  $\|\nabla Q(x^k)\| \rightarrow 0$  при  $k \rightarrow \infty$ .  $\square$

## Дополнительные свойства метода наискорейшего градиентного поиска

**Свойство 1.** *Всякие два последовательных направления поиска  $d^{k+1}$  и  $d^k$  взаимно ортогональны (см. рис. 4.22), т.е.  $\forall k : (d^{k+1}, d^k) = 0$ .*

*Доказательство.* Это вытекает из правила выбора шагового множителя (4.40). Действительно, для функции сечения  $\varphi(t) = Q(x^k + td^k)$  в точке  $t_k$  выполняется:  $\varphi'(t_k) = 0$ . Однако,

$$\varphi'(t_k) = (\nabla Q(x^{k+1}), d^k) = (\nabla Q_{k+1}, d^k) = (-d^{k+1}, d^k) = 0.$$

□

**Свойство 2.** *Для квадратичной функции*

$$Q(x) = \frac{x^T A x}{2} + (b, x) + c \quad (4.41)$$

*с симметричной, положительно определенной матрицей  $A$  метод наискорейшего градиентного поиска сходится из любой начальной точки  $x^0$  линейно, т.е. со скоростью геометрической прогрессии со знаменателем  $a = a(x^0)$ , не превосходящим значения  $0 \leq a < 1$ . При этом справедливы следующие оценки:*

$$0 \leq a(x^0) \leq q = \left( \frac{1 - \lambda_{\min}/\lambda_{\max}}{1 + \lambda_{\min}/\lambda_{\max}} \right)^2 \quad - \text{соотношение Канторовича,}$$

$$Q(x^k) - Q(x^*) \leq (a(x^0))^k (Q(x^0) - Q(x^*)),$$

$$\|x^k - x^*\| \leq T \cdot (a(x^0))^{k/2} \|x^0 - x^*\|,$$

где  $T$  — некоторая положительная константа, а  $\lambda_{\min}$  и  $\lambda_{\max}$  — минимальное и максимальное собственные числа матрицы вторых производных  $\nabla_{xx}^2 Q(x) = A$ .

Предлагается самостоятельно доказать это свойство.

Заметим, что приведенные оценки скорости сходимости обобщаются и на неквадратичный случай.

Соотношение Канторовича связывает оценку скорости сходимости с разбросом собственных чисел матрицы Гессе.

Если  $\lambda_{\min} = \lambda_{\max}$ , то поверхности равного уровня квадратичной функции будут концентрически вложенными сферами. Для таких функций направление антиградиента точно указывает на локальный минимум и сходимость к нему метода наискорейшего градиентного поиска из любой начальной точки будет за один шаг (см. рис. 4.23).



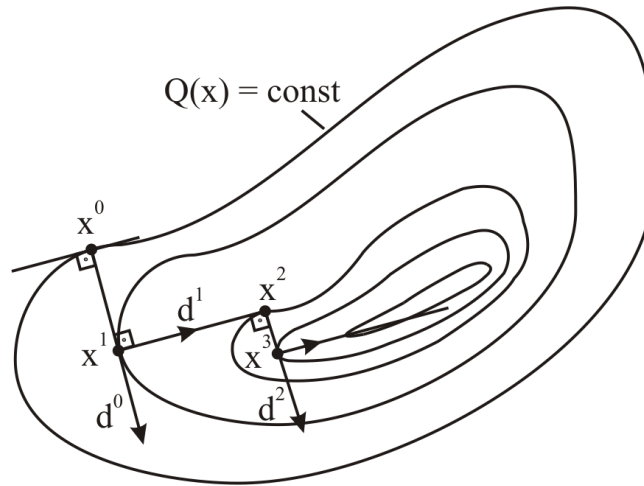


Рис. 4.22. Взаимная ортогональность направлений наискорейшего градиентного поиска

Действительно, из соотношения Канторовича вытекает, что при равенстве собственных чисел ( $\lambda_{min} = \lambda_{max}$ ) значение  $q = 0$  и, следовательно,  $Q(x^1) = Q(x^*)$ .

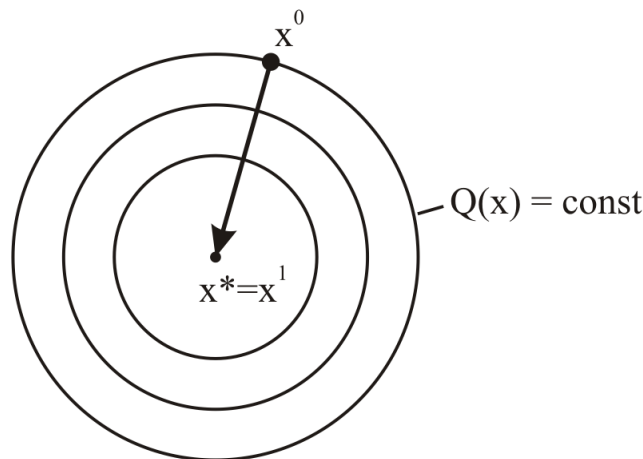


Рис. 4.23. Сходимость за один шаг при  $\lambda_{min} = \lambda_{max}$

Если же  $0 < \lambda_{min} \ll \lambda_{max}$ , то значение  $q$  близко к 1, что, в общем случае, соответствует чрезвычайно медленной сходимости (рис. 4.24). Сходимость за один шаг сохраняется только для начальных точек, расположенных на главных осях эллипсоидов  $Q(x) = const$  функции (4.41). При удалении начальной точки  $x^0$  от главных осей обеспечивается лишь медленная сходимость в пределе (за бесконечное число шагов). Это связано с тем, что свойство антиградиента указывать на локальный минимум, как это показано на рис. 4.23 для случая  $\lambda_{min} = \lambda_{max}$ , не инвариантно к линейным преобразованиям пространства. Линейные преобразования сжатия–растяжения, переводящие картину на рис.4.23 в вид, показанный на рис. 4.24 нарушают это свойство.

Проведённый анализ показывает, что классическая схема градиентного

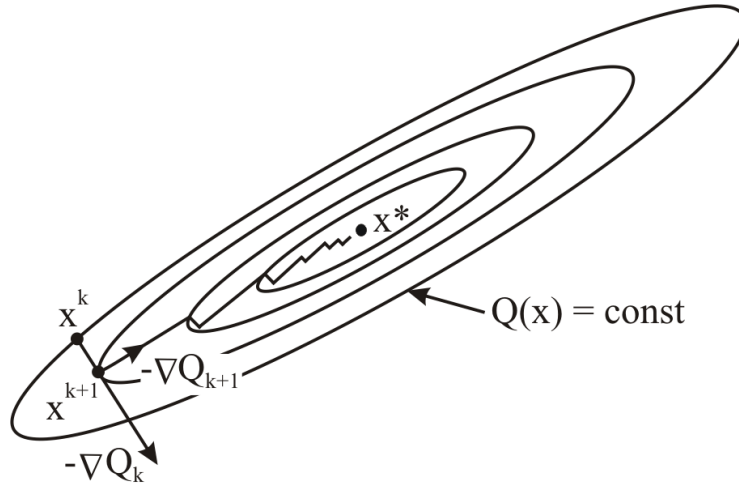


Рис. 4.24. Медленная сходимость при  $\lambda_{\min} \ll \lambda_{\max}$

поиска, когда  $d^k = -\nabla Q_k$ , обладает медленной сходимостью и не может быть рекомендована для практического применения. Однако, использование специальным образом скорректированных антиградиентных направлений позволяет строить эффективные методы (см., например, п.4.5.4).

### 4.5.3. Метод Ньютона

Данный метод является классическим *методом второго порядка*. Он использует локальные квадратичные аппроксимации функций. Концептуально важен как основа построения высоко эффективных методов.

Рассмотрим произвольную дважды непрерывно дифференцируемую функцию. Пусть в точке  $x^k$  для  $Q(x)$  измерено значение  $Q_k = Q(x^k)$ , градиент  $\nabla Q^k = \nabla Q(x^k)$  и матрица вторых производных  $\nabla_{xx}^2 Q_k = \nabla_{xx}^2 Q(x^k)$ , т.е.  $J(x^k) = (Q_k, \nabla Q_k, \nabla_{xx}^2 Q_k)$ . Построим локальную квадратичную аппроксимацию целевой функции (рис. 4.25):

$$P_k(x) = Q_k + (\nabla Q_k, x - x^k) + (x - x^k)^T \nabla_{xx}^2 Q_k (x - x^k) / 2.$$

Условие, определяющее стационарную точку  $\bar{x}$  полинома  $P_k(x)$ , имеет вид:

$$\nabla P_k(\bar{x}) = \nabla Q_k + \nabla_{xx}^2 Q_k (\bar{x} - x^k) = 0.$$

Если принять точку  $\bar{x}$  за точку очередного испытания  $x^{k+1}$ , получим *метод Ньютона*:

$$x^{k+1} = x^k + 1 \cdot d_{\text{НЬЮТ}}^k, \quad (4.42)$$

где  $d_{\text{НЬЮТ}}^k$  определяется из решения линейной системы вида

$$\nabla_{xx}^2 Q_k \cdot d_{\text{НЬЮТ}}^k = -\nabla Q_k. \quad (4.43)$$

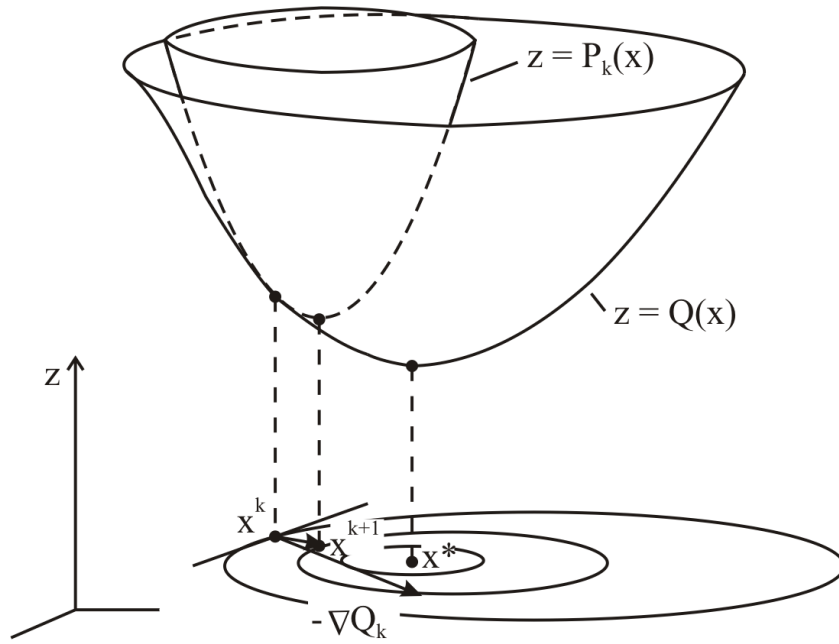


Рис. 4.25. Метод Ньютона размещает точку нового испытания  $x^{k+1}$  в стационарной точке текущей квадратичной аппроксимации

Итерационную форму метода Ньютона можно представить и в явном виде

$$x^{k+1} = x^k + (\nabla_{xx}^2 Q_k)^{-1}(-\nabla Q(x^k)). \quad (4.44)$$

Соотношение (4.44) показывает, что направление  $d_{\text{Ньюто}}^k$  получается из антиградиентного за счет домножения на него обращенной матрицы Гессе. Это равносильно линейному преобразованию пространства: повороту и растяжению–сжатию.

При численной реализации метода явная форма записи (4.44) не используется, поскольку обращение матрицы в (4.44) более трудоемко, чем решение линейной системы (4.43) относительно  $d_{\text{Ньюто}}^k$ .

### Свойства метода Ньютона

**Свойство 1.** Для квадратичной функции  $Q(x)$  при невырожденности  $\nabla_{xx}^2 Q$  метод Ньютона из любой начальной точки  $x^0$  переходит за один шаг в  $x^*$  — стационарную точку функции  $Q(x)$ , которая, в зависимости от характера знакоопределенности матрицы  $\nabla_{xx}^2 Q$ , может быть точкой минимума, максимума, а также — седловой точкой.

*Доказательство.* Вытекает из по способа построения метода. □

**Свойство 2.** Если матрица  $\nabla_{xx}^2 Q_k$  положительно определена и градиент  $\nabla Q(x^k) \neq 0$ , то  $d_{\text{Ньюто}}^k = (\nabla_{xx}^2 Q_k)^{-1}(-\nabla Q(x^k))$  является направлением строгого локального убывания функции, в точке  $x^k$ .

*Доказательство.* В указанных условиях  $\partial Q(x^k)/\partial d_{\text{НьюТ}}^k = (\nabla Q_k, d_{\text{НьюТ}}^k) < 0$  в силу положительной определенности матрицы Гессе, а, следовательно, и обратной к ней.  $\square$

**Свойство 3.** *Можно построить примеры достаточно гладких функций, имеющих локальный минимум в точке  $x^*$ , для которых при выборе начальной точки вне некоторой окрестности  $x^*$  будет наблюдаться расходимость.*

*Доказательство.* Предлагается самостоятельно построить такие примеры для  $N = 1$ . Чтобы принцип их построения стал понятен, обратим внимание на простую интерпретацию метода Ньютона в случае  $R^1$ . Для функций одного переменного соотношение (4.44) принимает вид

$$x^{k+1} = x^k - \frac{Q'_k}{Q''_k}. \quad (4.45)$$

Представим итерационный процесс на плоскости  $(x, z)$ , где  $z = Q'$ . Стационарной точке  $x^*$  функции  $Q$  будет соответствовать корень производной  $Q'(x^*) = 0$ , а соотношение (4.45) можно интерпретировать как метод касательных для поиска корня уравнения  $Q'(x) = 0$  (рис. 4.26).  $\square$

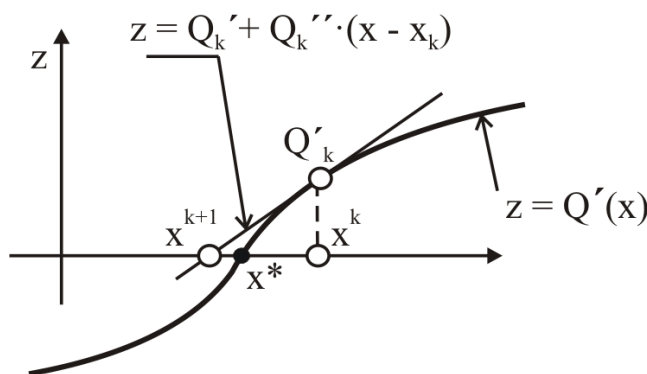


Рис. 4.26. В  $R^1$  метод Ньютона совпадает с методом касательных при поиске корня производной функции

В качестве обоснования приведите примеры ситуаций, в которых будет наблюдаться расходимость.

**Свойство 4.** *Сходимость.*

**Теорема** (об условиях и характере сходимости). *Для дважды непрерывно дифференцируемых функций  $Q \in C^2(R^N)$  с невырожденной матрицей  $\nabla_{xx}^2 Q(x^*)$  существует такая  $\varepsilon$ -окрестность стационарной точки  $x^*$  функции  $Q(x)$ , что для любой начальной точки  $x^0$  из этой окрестности метод*

Ньютона будет сходиться сверхлинейно. Если  $Q(x)$  — трижды непрерывно дифференцируемая функция, то сходимость квадратичная, т.е.  $\exists C > 0$ , что начиная с некоторого шага  $\|x^{k+1} - x^*\| \leq C \cdot \|x^k - x^*\|^2$ .

*Доказательство.* Обоснуем первое утверждение.

Предположим, что  $Q \in C^2(R^N)$ . Тогда возможно следующее разложение в ряд Тейлора для градиента:

$$\nabla Q(x^k) = \underbrace{\nabla Q(x^*)}_{=0} + \nabla_{xx}^2 Q(x^*)(x^k - x^*) + o(\|x^k - x^*\|),$$

и также представление

$$\nabla_{xx}^2 Q(x^k) = \nabla_{xx}^2 Q(x^*) + \beta(x^k, x^*),$$

где  $\beta(x^k, x^*)$  — бесконечно малая при  $\|x^k - x^*\| \rightarrow 0$ .

Для  $x^k$ , взятых в достаточно малой окрестности  $x^*$ , матрицы  $\nabla_{xx}^2 Q(x^k)$  будут невырождены. Подставляя второе выражение в первое имеем:

$$\nabla Q(x^k) = \nabla_{xx}^2 Q(x^k)(x^k - x^*) + \tilde{o}(\|x^k - x^*\|),$$

затем умножая слева на матрицу  $(\nabla_{xx}^2 Q(x^k))^{-1}$ , ограниченную при  $x^k \rightarrow x^*$ , получим:

$$x^* - (x^k + (\nabla_{xx}^2 Q(x^k))^{-1}(-\nabla Q_k)) = \tilde{o}(\|x^k - x^*\|).$$

Используя (4.44) придем к соотношению

$$x^* - x^{k+1} = \tilde{o}(\|x^k - x^*\|),$$

где через  $\tilde{o}(\Delta)$  обозначена некоторая бесконечно малая (при  $\Delta \rightarrow 0$ ), более высокого порядка, чем  $\Delta$ .

Возьмем норму от левой и правой части последнего равенства. Тогда при  $x^k \rightarrow x^*$

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \rightarrow 0.$$

Следовательно, при достаточной близости начальной точки  $x^0$  к  $x^*$ , найдется  $0 < q < 1$ , что для всех  $k$  будет выполнено:  $\|x^{k+1} - x^*\|/\|x^k - x^*\| \leq q < 1$ . Это означает сходимость к  $x^*$  из любой начальной точки  $x^0$  некоторой  $\varepsilon$ -окрестности точки  $x^*$  при  $k \rightarrow \infty$ . Таким образом, в этой окрестности  $\|x^k - x^*\| \rightarrow 0$  при  $k \rightarrow \infty$ , что означает сходимость. Но ранее было показано, что  $\|x^{k+1} - x^*\| = o(\|x^k - x^*\|)$  при  $\|x^k - x^*\| \rightarrow 0$ . Следовательно, сходимость будет сверхлинейной.

Докажем второе утверждение. Для трижды непрерывно дифференцируемой функции все рассуждения повторяются, только теперь можно утверждать, что в разложении градиента остаточный член будет не  $o(\|x^k - x^*\|)$ ,

а  $O(\|x^k - x^*\|^2)$ , а в разложении матрицы остаточный член будет иметь порядок  $O(\|x^k - x^*\|)$ :

$$\nabla Q(x^k) = \underbrace{\nabla Q(x^*)}_{=0} + \nabla_{xx}^2 Q(x^*)(x^k - x^*) + O(\|x^k - x^*\|^2),$$

откуда

$$\nabla_{xx}^2 Q(x^*) = \nabla_{xx}^2 Q(x^k) + O(\|x^k - x^*\|),$$

поэтому

$$x^* - (x^k + (\nabla_{xx}^2 Q(x^k))^{-1}(-\nabla Q_k)) = \tilde{O}(\|x^k - x^*\|^2).$$

Отсюда следует, что для некоторого  $C$  выполнится следующее неравенство

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^2} \leq C,$$

начиная с достаточно больших  $k$ . Это означает, что сходимость квадратичная.  $\square$

### Замечание.

1. Теорема в свойстве 4 утверждает сходимость к стационарной точке  $x^*$  функции  $Q$ . В зависимости от характера знакоопределенности матрицы  $\nabla_{xx}^2 Q(x^*)$  эта точка может быть точкой минимума, седловой точкой или точкой максимума.
2. Квадратичная сходимость может быть доказана при более мягких требованиях к  $Q(x)$ . Вместо гладкости третьего порядка достаточно потребовать липшицевости матриц Гессе. Однако доказательство в этом случае усложняется.

Квадратичная сходимость, характерная для метода Ньютона, означает чрезвычайно быстрое убывание погрешности. Например, если константа  $C \approx 1$ , а ошибка на шаге  $k$  составляла 0.1, то на последующих трех шагах она составит 0.01, 0.0001 и 0.00000001. Вместе с тем, сходимость обеспечивается лишь в некоторой неизвестной окрестности стационарной точки  $x^*$ . Вне этой окрестности метод может расходиться. Кроме того, если начальная точка оказывается близкой к точке локального максимума или стационарной точке седлового типа, то метод Ньютона будет сходиться к этим стационарным точкам.

Такое поведение делает невозможным применение классической схемы метода Ньютона для решения прикладных задач поиска минимума. Практически применяются его модификации, лишенные указанных недостатков, но сохраняющие высокий порядок сходимости. Сама идея использования квадратичных аппроксимаций, примененная в методе Ньютона, является весьма плодотворной и находит многочисленные применения в современных методах локальной оптимизации, включая задачи с ограничениями (см. [14]).

#### 4.5.4. Метод сопряженных градиентов Флетчера–Ривса

В этом разделе описано построение эффективного метода первого порядка, основанного на использовании специальным образом скорректированных антиградиентных направлений.

Метод был предложен в 1964 году Флетчером и Ривсом. При его построении используется квадратичная модель поведения функции. Для квадратичных функций вида (4.41) с симметричными положительно определенными матрицами данный метод является конечно сходящимся. Для определения локального минимума  $x^*$  ему требуется не более  $N$  итераций (сравните с бесконечным числом итераций для простого градиентного метода). Для достаточно широких классов неквадратичных функций метод обладает глобальной областью сходимости и сверхлинейным порядком скорости сходимости. В основе построения метода лежит понятие сопряженных направлений.

##### Сопряженные направления и их свойства

**Определение.** Система линейно-независимых векторов  $p^0, p^1, \dots, p^{N-1}$  для симметричной матрицы  $A$  называется  $A$ -сопряженной, если

$$\forall i = 1, \dots, N; j = 1, \dots, N; i \neq j : (p^i, Ap^j) = 0. \quad (4.46)$$

**Определение.** Пусть  $M$  — линейное многообразие,  $A$  — симметричная матрица,  $v \neq 0, v \notin M$  и

$$\forall z \in M : (v, Az) = 0, \quad (4.47)$$

тогда вектор  $v$  называют  $A$ -сопряженным с многообразием  $M$ .

Докажите самостоятельно следующую лемму.

**Лемма.** Если  $p^0, p^1, \dots, p^{N-1}$  — все отличны от нуля,  $A$  — не только симметрична, но еще и положительно определена, тогда из (4.46) следует линейная независимость векторов  $p^0, p^1, \dots, p^{N-1}$ .

Заметим, что в условиях леммы, т.е. положительной определенности матрицы  $A$ ,  $A$ -сопряженность означает ортогональность в смысле нового скалярного произведения  $\langle x, y \rangle = (x, Ay)$ .

Для построения методов, использующих сопряженные направления, чрезвычайно важным является свойство, определяемое следующей леммой.

**Лемма.** Пусть  $Q(x)$  — квадратичная функция вида (4.41) с симметричной положительно определенной матрицей  $A$ , а  $M$  и  $L$  — линейные многообразия, причем  $M \subset L$ . Тогда, если  $z$  — точка минимума  $Q(x)$  на  $M$ , а  $u$  — точка минимума  $Q(x)$  на  $L$ , причем  $u \neq z$ , то вектор  $(u - z)$  будет  $A$ -сопряжен с многообразием  $M$ , включенным в  $L$ .

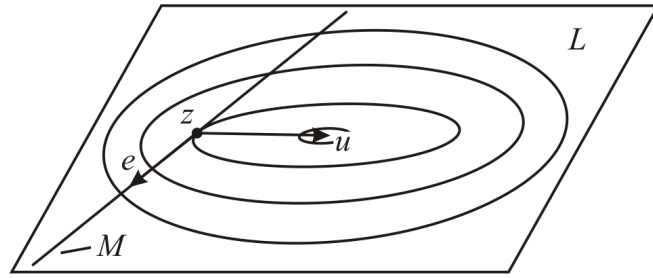


Рис. 4.27. Вектор  $u - z$   $A$  – сопряжен с  $M$

Это утверждение иллюстрирует рис. 4.27.

*Доказательство.* Рассмотрим произвольный вектор  $e \in M$ . Поскольку  $\nabla Q(x) = Ax + b$ , а матрица  $A$  симметрична, то  $(u - z, Ae) = (Au - Az, e) = (\nabla Q(u) - \nabla Q(z), e)$ . Последнее скалярное произведение равно нулю, т.к. по теореме Лагранжа в точках минимума  $u$  и  $z$  на линейных многообразиях  $L$  и  $M$  градиенты функции ортогональны к этим многообразиям, а поскольку  $e \in M \subset L$ , то  $\forall e \in M$ :

$$(\nabla Q(u), e) = 0, (\nabla Q(z), e) = 0.$$

Таким образом, для  $v = u - z$  выполнено (4.47), следовательно,  $(u - z)$  будет  $A$ -сопряжен с многообразием  $M$ .  $\square$

Построим теперь вычислительные процедуры поиска минимума квадратичной функции  $Q(x)$  из (4.41), использующие  $A$ -сопряженные направления.

**Определение.** Поисковые процедуры вида (4.48)-(4.49) называются *методами сопряженных направлений*:

$$x^{k+1} = x^k + t_k \cdot p^k, \quad (4.48)$$

$$Q(x^k + t_k p^k) = \min \{Q(x^k + t p^k) : -\infty < t < +\infty\}. \quad (4.49)$$

Применение сопряженных направлений при построении методов оптимизации связано с их замечательным свойством приводить в минимум строго выпуклой квадратичной функции не более чем за  $N$  шагов.

**Теорема.** Пусть  $Q(x)$  – квадратичная функция вида (4.37) с симметричной положительно определенной матрицей  $A$ , а  $p^0, p^1, \dots, p^{N-1}$  – система  $A$ -сопряженных векторов. Тогда для любой начальной точки  $x^0$  процедура поиска вида (4.44) – (4.45) приводит в минимум функции не более, чем за  $N$  шагов.



*Доказательство.* При поиске вдоль направления  $p^0$  метод определит точку  $x^1$  — минимум на одномерном многообразии  $L(p^0)$ , натянутом на  $p^0$ . На втором шаге при поиске вдоль направления  $p^1$  метод определит точку  $x^2$ . По построению вектор  $x^2 - x^1$  будет сопряжен с  $L(p^0)$ , т.е. ортогонален к  $p^0$  в смысле нового скалярного произведения  $\langle, \rangle$ . Если теперь рассмотреть линейное многообразие  $L(p^0, p^1)$ , натянутое на  $p^0, p^1$  и предположить, что минимум функции  $Q(x)$  достигается на нем в некоторой точке  $\bar{x}^2 \neq x^2$  то возникает противоречие. Действительно, по доказанной выше лемме мы получим еще один вектор  $\bar{x}^2 - x^1$ , не принадлежащий прямой, проходящей через  $x^2$  и  $x^1$ , лежащий в том же двумерном многообразии  $L(p^0, p^1)$  и ортогональный (в смысле нового скалярного произведения) к  $p^0$ , что невозможно. Значит  $\bar{x}^2 = x^2$ , и на втором шаге метод сопряженных направлений найдет минимум на двумерном многообразии  $L(p^0, p^1)$ .

Продолжая аналогичные рассуждения, приходим к выводу, что не более чем за  $N$  шагов метод найдет минимум на линейном многообразии  $L(p^0, \dots, p^{N-1})$  размерности  $N$ , т.е. во всем пространстве (рис 4.28). Следовательно, определение точки минимума  $x^*$  произойдет не более, чем за  $N$  итераций.  $\square$

Для того, чтобы можно было воспользоваться методом сопряженных направлений, необходим алгоритм вычисления  $A$ -сопряженных векторов  $p^0, \dots, p^{N-1}$ . Проблема, которая на первый взгляд кажется непреодолимой, заключается в том, что нужно построить  $A$ -сопряженные векторы не зная самой матрицы  $A$ . Однако эта задача может быть решена с использованием результатов испытаний первого порядка функции  $Q(x)$  из (4.41), выполняемых в основных точках траектории поиска.

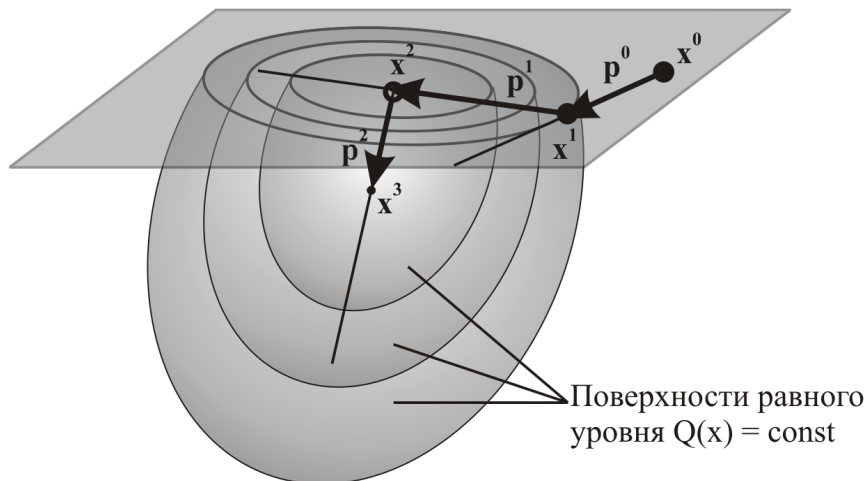


Рис. 4.28. Характер поиска в методе сопряженных направлений для строго выпуклой квадратичной функции

## Метод сопряженных градиентов Флетчера–Ривса для строго выпуклых квадратичных функций

Этот метод использует испытания первого порядка, когда у функции  $Q$  в точке  $x^k$  определяются значения  $Q_k$  и  $\nabla Q_k$ . Для построения метода сопряженных направлений для квадратичной функции  $Q(x)$  из (4.41) необходимо по результатам испытаний получить систему  $A$ -сопряженных векторов  $p^0, \dots, p^{N-1}$  при условии, что сама матрица  $A$  является неизвестной.

Выберем начальный вектор:

$$p^0 = -\nabla Q_0, \quad \nabla Q_0 = \nabla Q(x^0). \quad (4.50)$$

Пусть векторы  $p^0, \dots, p^{k-1}$  построены. Положим, согласно (4.48),

$$x^k = x^{k-1} + t_{k-1} \cdot p^{k-1}, \quad (4.51)$$

где

$$t_{k-1} = \arg \min \{ Q(x^{k-1} + tp^{k-1}) : t \geq 0 \}, \quad (4.52)$$

Примем

$$p^k = -\nabla Q_k + \beta_{k-1} p^{k-1}; \quad \nabla Q_k = \nabla Q(x^k). \quad (4.53)$$

Подберем  $\beta_{k-1}$  из условия  $A$ -сопряженности  $p^k$  и  $p^{k-1}$ :

$$(p^k, Ap^{k-1}) = 0.$$

Получим

$$\beta_{k-1} = ((\nabla Q_k)^T Ap^{k-1}) / ((p^{k-1})^T Ap^{k-1}), \quad (4.54)$$

но здесь значение матрицы  $A$  неизвестно, ее нужно исключить. Окончательный вид  $\beta_{k-1}$  приведен далее в (4.59), (4.60).

Значение  $t^k$ , удовлетворяющее (4.52) для функции  $Q$ , можно получить из условия оптимальности по направлению  $p^k$ :

$$(\nabla Q(x^k + t_k p^k), p^k) = 0,$$

если его переписать в виде

$$t^k = -((p^k)^T \nabla Q_k) / ((p^k)^T Ap^k) = -((\nabla Q_k)^T \nabla Q_k) / ((\nabla Q_k)^T Ap^k). \quad (4.55)$$

Этот результат может быть получен, если в числителе и знаменателе первой дроби выразить  $p^k$  из (4.53) и воспользоваться тем, что  $((p^{k-1})^T \nabla Q_k) = 0$  по теореме Лагранжа (условие оптимальности в (4.52), и  $(p^{k-1}, Ap^k) = 0$  по построению.

Кроме того, умножая (4.51) на  $A$  слева и добавляя в обе части равенства вектор  $b$  из (4.41), получим дополнительное соотношение

$$\nabla Q_{k+1} = \nabla Q_k + t_k Ap^k. \quad (4.56)$$

**Лемма.** При использовании правил (4.50) и (4.51)–(4.53) последовательность векторов градиентов  $\nabla Q_0, \nabla Q_1, \dots, \nabla Q_{N-1}$  образует взаимно ортогональную систему, а направления  $p^0, p^1, \dots, p^{N-1}$   $A$ -сопряжены.

*Доказательство.* Пользуясь соотношениями (4.53), (4.56), (4.55), приведем известное доказательство леммы методом математической индукции.

Действительно, по построению,  $p^1$  сопряжен с  $p^0$ . Кроме того,  $p^0 = -\nabla Q_0$ , а по теореме Лагранжа  $\nabla Q_1$  ортогонален  $p^0$ , следовательно,  $\nabla Q_1$  ортогонально  $\nabla Q_0$ . Таким образом, для систем из двух векторов лемма верна.

Предположим, что при  $k < (N - 1)$  векторы в системе  $p^0, p^1, \dots, p^k$  взаимно сопряжены, а векторы  $\nabla Q_0, \nabla Q_1, \dots, \nabla Q_k$  — взаимно ортогональны. Покажем, что эти свойства сохранятся у данных систем векторов при включении в них  $p^{k+1}$  и  $\nabla Q_{k+1}$ .

Рассмотрим значения  $i < k$ , тогда из (4.56) и (4.53):

$$((\nabla Q_{k+1})^T \nabla Q_i) = ((\nabla Q_k + t_k A p^k)^T \nabla Q_i) = t_k (A p^k)^T (-p^i + \beta_{i-1} p^{i-1}) = 0.$$

Рассмотрим теперь  $i = k$ . Применяя (4.56) и используя выражение из (4.55) для величины  $t_k$ , получим:

$$((\nabla Q_{k+1})^T \nabla Q_k) = ((\nabla Q_k + t_k A p^k)^T \nabla Q_k) = 0.$$

Теперь докажем сохранение сопряженности системы векторов  $p^i$  для  $i = 1, \dots, k + 1$ . Сопряженность двух последних векторов следует из способа их построения. Осталось рассмотреть только  $i < k$ . Применяя (4.53), учитывая сопряженность по предположению индукции векторов  $p^0, \dots, p^k$ , а также используя (4.56), получим

$$\begin{aligned} ((p^{k+1})^T A p^i) &= (-\nabla Q_{k+1} + \beta_k p^k)^T A p^i = (-\nabla Q_{k+1})^T A p^i = \\ &= (-\nabla Q_{k+1})^T (\nabla Q_{i+1} - \nabla Q_i) / t_i = 0. \end{aligned} \quad (4.57)$$

Последнее равенство нулю вытекает из уже доказанной ортогональности градиентов.  $\square$

Метод сопряженных направлений для строго выпуклых квадратичных функций в целом, построен. Однако, в формулу (4.54) для вычисления коэффициента  $\beta^{k-1}$  входит неизвестная матрица  $A$ . Но это не является существенным, поскольку выражение (4.54) можно переписать в другом виде. Чтобы показать это, выразим  $A p^{k-1}$  в числителе (4.54) из (4.56) а  $p^{k-1}$  в знаменателе (4.54) из (4.53). Тогда

$$\begin{aligned} \beta_{k-1} &= (\nabla Q_k, (\nabla Q_k - \nabla Q_{k-1})) / (t_{k-1} (-\nabla Q_{k-1} + \beta_{k-2} p^{k-2})^T A p^{k-1}) = \\ &= (\nabla Q_k, \nabla Q_k) / (-\nabla Q_{k-1})^T t_{k-1} A p^{k-1}. \end{aligned} \quad (4.58)$$

Выражая  $t_{k-1}Ap^{k-1}$  из (4.56) и пользуясь ортогональностью  $\nabla Q_k$  и  $\nabla Q_{k-1}$ , окончательно получим

$$\beta_{k-1} = \|\nabla Q_k\|^2 / \|\nabla Q_{k-1}\|^2, \quad (4.59)$$

Если не преобразовывать числитель в предшествующем соотношении, получим другую форму записи коэффициента

$$\beta_{k-1} = (\nabla Q_k, \nabla Q_k - \nabla Q_{k-1}) / \|\nabla Q_{k-1}\|^2, \quad (4.60)$$

которая рекомендуется для использования в неквадратичном случае.

**Замечание.** Построенный метод находит минимум любой строго выпуклой квадратичной функции не более чем за  $N$  шагов. Отметим, что для численного определения  $t^k$  должен быть использован «аккуратный» одномерный поиск (т.е. параметр  $\eta$  одномерного поиска должен быть выбран близким к нулю). Правила Армихо и Вульфа неприменимы.

### Метод Флетчера-Ривса для неквадратичных функций

Применение метода сопряженных градиентов к неквадратичной функции  $Q(x)$ , естественно, не может обеспечить конечность процедуры поиска минимума. В этом случае генерируемые направления  $p^k$  уже не являются  $A$ -сопряженными для какой-либо матрицы  $A$ . После выполнения серии из  $N$  шагов метод, как правило, повторно запускается из последней найденной точки, что приводит к встраиванию в процедуру поиска градиентных шагов. Это положительно влияет на глобализацию области сходимости метода. Кроме того, две приведенные в (4.59), (4.60) формулы для определения  $\beta_{k-1}$  перестают быть эквивалентными. Вычислительная практика показывает, что для неквадратичного случая предпочтительнее использовать формулу (4.60).

Свойство строгого локального убывания функции  $Q$  в точке  $x^{k+1}$  в направлении  $p^{k+1}$  определяется отрицательностью соответствующего скалярного произведения, которое в силу (4.53) представимо в виде

$$(\nabla Q_{k+1}, p^{k+1}) = -(\nabla Q_{k+1}, \nabla Q_{k+1}) + \beta_k (\nabla Q_{k+1}, p^k). \quad (4.61)$$

Если  $\nabla Q_{k+1} \neq 0$  и  $x^{k+1}$ , как это должно быть по теории, соответствует минимуму вдоль направления  $p^k$ , то  $(\nabla Q_{k+1}, p^k) = 0$  и знак всего выражения (4.61) — отрицательный. Ошибки в определении минимума по направлению могут нарушать это свойство. Поэтому в метод встраивается проверка отрицательности выражения (4.61) и выполняются рестарты метода при нарушении этого условия. Если минимум вдоль направления  $p^k$  будет определяться с большой погрешностью, то это будет приводить к частым рестартам. Возникающие при этом частые замены  $p^k$  на антиградиент могут существенно снизить эффективность метода. Именно поэтому метод Флетчера-Ривса

несовместим с правилом выбора  $t_k$  по правилам Армихо и Вулфа. Он требует выполнения «аккуратного» одномерного поиска, что вычислительно достаточно трудоемко.

С учетом сделанных замечаний получаем следующий алгоритм метода Флетчера–Ривса для неквадратичных функций

**Шаг 0.** Задаём  $\varepsilon > 0$  — параметр останова,  $0 < \mu < \eta \ll 1$  — параметры одномерного поиска,  $x^0$  — начальная точка.

**Шаг 1.** Вычисляем  $Q_0 = Q(x_0)$ ,  $\nabla Q_0 = \nabla Q(x^0)$ , полагаем  $p^0 = -\nabla Q_0$ ,  $k = 0$ .

**Шаг 2.** Если  $(\nabla Q_k, p^k) \geq 0$ , то направление  $p^k$  не является направлением локального убывания функции, поэтому полагаем  $p^0 = -\nabla Q_k$ ,  $x^0 = x^k$  и устанавливаем  $k = 0$ . Иначе направление  $p^k$  сохраняем. Переходим на шаг 3.

**Шаг 3.** Вычисляем величину шагового множителя  $t_k$  методом «аккуратного» одномерного поиска с параметрами  $\mu, \eta$ . Определяем

$$x^{k+1} = x^k + t_k p^k,$$

$$Q_{k+1} = Q(x^{k+1}), \nabla Q_{k+1} = \nabla Q(x^{k+1}).$$

Полагаем  $k = k + 1$ .

**Шаг 4.** Проверяем критерий останова: при выполнении условия  $\|\nabla Q_k\| \leq \varepsilon$  поиск прекращаем и  $x^k, Q_k$  выдаем как оценку решения; при  $\|\nabla Q_k\| > \varepsilon$  переходим на шаг 5.

**Шаг 5.** Если  $k = N$ , полагаем  $x^0 = x^N$  и выполняем возврат на шаг 1. Если  $k < N$ , то переходим на шаг 6.

**Шаг 6.** Вычисляем  $\beta^{k-1}$  по формуле  $\beta_{k-1} = (\nabla Q_k, \nabla Q_k - \nabla Q_{k-1}) / \|\nabla Q_{k-1}\|^2$  и  $p^k$  по следующим формулам:  $p^k = -\nabla Q_k + \beta_{k-1} p^{k-1}$ . Переходим на шаг номер 2 алгоритма.

Приведем без доказательства несколько фактов, касающихся сходимости метода Флетчера–Ривса (см. [16], [17]). В одной из теорем нам понадобится условие сильной выпуклости, которое ранее не использовалось.

**Определение.** Функция  $Q$  *сильно выпукла* в  $R^N$ , если существует константа  $\rho > 0$ , что для любых  $x'$  и  $x''$  и любого  $\alpha \in [0, 1]$

$$Q(x'\alpha + (1 - \alpha)x'') \leq \alpha Q(x') + (1 - \alpha)Q(x'') - \rho\alpha(1 - \alpha)\frac{\|x' - x''\|^2}{2}$$

**Теорема (сходимости).** I. Пусть  $Q(x) \in C^1(R^N)$ , т.е. непрерывно дифференцируема, а множество  $\{x : Q(x) \leq Q(x^0)\}$  ограничено. Тогда в методе Флетчера–Ривса  $\|\nabla Q(x^k)\| \rightarrow 0$  при  $k \rightarrow \infty$ .

II. Пусть  $Q(x)$  — дифференцируемая сильно выпуклая функция, градиент

которой удовлетворяет условию Липшица, тогда из любой начальной точки метод Флетчера–Ривса сходится к единственной точке локального минимума функции  $Q$ .

**Теорема** (об оценке скорости сходимости). Пусть  $Q(x) \in C^2(R^N)$ ,  $x^*$  — локальный минимум, причем  $\nabla_{xx}^2 Q(x^*)$  — положительно определена, тогда в окрестности  $x^*$  при наблюдении через  $N$  итераций метод сходится сверлинейно; если же дополнительно матрица Гессе  $\nabla_{xx}^2 Q(x)$  липшицева в окрестности  $x^*$ . Тогда для метода Флетчера–Ривса в окрестности точки  $x^*$  справедлива оценка

$$\|x^{(s+1) \cdot N} - x^*\| \leq c \|x^{sN} - x^*\|^2.$$

Таким образом, при наблюдении за методом через  $N$  итераций, в указанных выше условиях он сходится квадратично и его  $N$  итераций по порядку скорости сходимости равносильны одной итерации метода Ньютона.

#### 4.5.5. Обзор других эффективных методов локальной оптимизации первого и второго порядков

Рассмотрим задачи безусловной оптимизации, обладающие достаточной гладкостью. В этом разделе будут в обзорном плане указаны некоторые известные методы, являющиеся улучшенными модификациями базовых методов первого и второго порядков. Заметим, что подробно рассмотренный в п. 4.5.4. метод сопряженных градиентов Флетчера–Ривса, являющийся методом первого порядка, можно воспринимать как эффективное развитие простых градиентных методов.

Получить наглядные представления о поведении и сравнительной эффективности метода Флетчера–Ривса, а также других методов, рассматриваемых в данном разделе и разделе 4.5.6., можно с помощью учебно-демонстрационной программной системы *LocOpt*, разработанной на кафедре ТУиДС института ИТММ ННГУ (см. описание системы в [21]).

Рассмотрим теперь улучшенные модификации метода Ньютона — метода второго порядка.

#### Метод Ньютона с регулировкой шага

Высокий порядок скорости сходимости в некоторой окрестности стационарной точки функции, характерный для метода Ньютона, может сочетаться с отсутствием сходимости вне этой окрестности. Для расширения области сходимости этого метода вводят регулировку шагового множителя, который у исходного метода Ньютона тождественно равен единице:  $t_k \equiv 1$ .

Для возможности регулировки шага обеспечим, чтобы направление

$$d_{\text{ньют}}^k = (\nabla_{xx}^2 Q_k)^{-1}(-\nabla Q_k)$$

было направлением убывания. Потребуем, чтобы

$$\forall x \in R^N : mE \leq \nabla_{xx}^2 Q(x) \leq ME,$$

где  $E$  — единичная матрица, а  $0 < m \leq M < +\infty$ . Обозначим этот подкласс строго выпуклых функций через  $\Phi_{m,M}$ . Далее считаем, что  $Q \in \Phi_{m,M}$

Выберем шаговый множитель из принципа одномерной минимизации

$$\nabla_{xx}^2 Q_k \cdot d_{\text{НЬЮТ}}^k = -\nabla Q_k, \quad (4.62)$$

$$x^{k+1} = x^k + t_k d_{\text{НЬЮТ}}^k, \quad (4.63)$$

$$t_k = \arg \min_{t \geq 0} Q(x^k + t d_{\text{НЬЮТ}}^k). \quad (4.64)$$

Модифицированный таким образом метод называют *методом Ньютона–Рафсона*. Имеет место следующая теорема.

**Теорема** (сходимости). *На классе функций  $\Phi_{m,M}$  метод Ньютона–Рафсона сходится из любой начальной точки к единственной точке минимума. Локально сходимость является сверхлинейной.*

В методе Ньютона с регулировкой шага (4.62)–(4.64) вместо правила (4.64) можно использовать более дешевое правило Армихо или Вулфа. При этом желательно применять его с параметром  $\alpha = 1$ , т.к. при попадании  $x^k$  в малую окрестность решения начальное значение  $t_k = \alpha = 1$  начнет сразу приниматься алгоритмом и метод превратится в обычный метод Ньютона, обеспечивая высокий порядок локальной скорости сходимости. Заметим, что при этом на значение параметра  $\mu$  из (4.33) следует наложить ограничение  $0 < \mu < 1/2$ , при котором значение  $t = 1$ , соответствующее минимуму функции одномерного сечения  $\varphi(t) = Q(x^k + t d_{\text{НЬЮТ}}^k)$  в виде параболы, будет принадлежать множеству  $\Pi_1(\mu)$ .

### Метод Ньютона с регулировкой шага и модификацией матрицы

Если функция  $Q$  не принадлежит классу  $\Phi_{m,M}$ , то она может иметь участки знаконеопределенной или отрицательной кривизны. Поскольку при этом  $d_{\text{НЬЮТ}}^k$  может перестать быть направлением локального убывания, то предыдущие модификации метода окажутся неработоспособными.

В этом случае прибегают к замене матриц  $\nabla_{xx}^2 Q_k$  на возможно близкие к ним по малости нормы разности положительно определенные матрицы  $\tilde{\Gamma}_k$ , которые строятся в виде модифицированного разложения Холесского  $\nabla_{xx}^2 Q_k \Rightarrow \tilde{\Gamma}_k = \tilde{L}_k^T \tilde{D}_k \tilde{L}_k$ , где  $\tilde{D}_k$  — диагональная матрица с диагональю  $d_{ii} > \delta > 0$ , а  $\tilde{L}_k$  — верхняя треугольная матрица с единичной диагональю. Параметр преобразования  $\delta$  задается.

Методы выглядят следующим образом. Выполняется модифицированное преобразование Холецкого (подробнее см., например, в [2]):

$$\nabla_{xx}^2 Q_k \rightarrow \tilde{L}_k; \tilde{D}_k,$$

решаются две линейные системы с треугольными матрицами:

$$\tilde{L}_k^T \cdot \nu = -\nabla Q_k; (\tilde{D}_k \tilde{L}_k) \tilde{d}_{\text{НЬЮТ}}^k = \nu; \quad (4.65)$$

$$x^{k+1} = x^k + t_k \tilde{d}_{\text{НЬЮТ}}^k, \quad (4.66)$$

где шаговый множитель определяют либо из условия одномерной минимизации (что достаточно дорогостояще, но уменьшает число основных итераций):

$$t_k = \arg \min_{t \geq 0} Q(x^k + t \cdot \tilde{d}_{\text{НЬЮТ}}^k), \quad (4.67)$$

либо определяют  $t_k$  по правилу Армихо или Вулфа с параметрами  $\alpha = 1$ ,  $0 < \mu < 1/2$ .

На широких классах функций метод сходится к точкам локального минимума из любой нестационарной точки и сохраняет локально сверхлинейную или квадратичную скорость сходимости.

### Квазиньютоновские методы переменной матрики

Если матрица вторых частных производных недоступна измерению, её в модифицированном методе Ньютона можно заменить некоторыми оценочными матрицами  $G_k$  для коррекции которых используется информация об изменениях вектора градиента при переходе между точками основных испытаний:

$$G_k \cdot d_{\text{КВ.НЬЮТ}}^k = -\nabla Q_k; \quad (4.68)$$

$$x^{k+1} = x^k + t_k d_{\text{КВ.НЬЮТ}}^k; \quad (4.69)$$

$$t_k = \arg \min_{t \geq 0} Q(x^k + t d_{\text{КВ.НЬЮТ}}^k) \quad (4.70)$$

Пусть  $\Delta x^k = x^{k+1} - x^k$ , а  $\Delta y^k = \nabla Q_{k+1} - \nabla Q_k$ . Заметим, что если функция  $Q$  квадратична, то справедливо очевидное равенство

$$\Delta y^k = \nabla_{xx}^2 Q \cdot \Delta x^k,$$

где матрица  $\nabla_{xx}^2 Q$  — постоянна. Для неквадратичных функций оно выполняется лишь приближенно

$$\Delta y^k \approx \nabla_{xx}^2 Q_{k+1} \cdot \Delta x^k.$$

Для коррекций матриц  $G_k$  используется близкое условие, называемое *квазиньютоновским условием*:

$$\Delta y^k = G_{k+1} \cdot \Delta x^k. \quad (4.71)$$



Здесь

$$G_{k+1} = G_k + U_k, \quad (4.72)$$

при дополнительном требовании

$$G_{k+1} = G_{k+1}^T. \quad (4.73)$$

Поправка  $U_k$  обычно выбирается в виде симметричной матрицы ранга 1 или 2.

На первом шаге  $G_k = E$  — единичной матрице. Обычно через каждые  $N$  шагов вновь полагают  $G_k = E$ . Известно несколько конкретных способов определения поправочных матриц, например, Давидана–Флетчера–Пауэлла (DFP) или Бройдена–Флетчера–Гольдфарба–Шанно (BFGS), последняя имеет вид:

$$G_{k+1} = G_k - \frac{G_k \Delta x^k (\Delta x^k)^T G_k}{(\Delta x^k)^T G_k \Delta x^k} + \frac{\Delta y^k (\Delta y^k)^T}{(\Delta y^k)^T \Delta x^k}. \quad (4.74)$$

При их использовании шаговый множитель  $t_k$  можно определять не только по процедуре одномерной минимизации (4.70), но и с использованием более экономичного правила Вулфа с параметром  $\alpha = 1$ . При этом обеспечивается положительная определенность оценочных матриц из (4.70). Заметим, что правило Армихо в данных методах использовать нельзя.

Матрицы  $G_k$  в (4.68) целесообразно предварительно подвергать разложению Холесского или модифицированному разложению Холесского, что позволяет более экономично решать системы вида (4.68). Применение модифицированного разложения Холесского для  $G_k$  перед его использованием в (4.68) позволяет, кроме того, исправить случай потери положительной определенности матрицы  $G_k$ . Такие ситуации могут возникать из-за ошибок округления, хотя по теории (без учета влияния этих ошибок) матрицы  $G_k$  должны всегда оставаться положительно определенными.

Заметим, что вместо итераций по построению  $G_k$ , являющихся заменителем гессиана  $\nabla_{xx}^2 Q(x^k)$ , можно проводить итерационное построение матриц  $H_k$ , являющихся заменителем обращенного гессиана  $(\nabla_{xx}^2 Q(x^k))^{-1}$ . Для этого существуют формулы, аналогичные (4.74). Однако, с учетом указанных выше факторов целесообразно применять итерационное соотношение (4.74).

Квазиньютоновские методы обладают свойствами сходимости, аналогичными методу Флетчера–Ривса.

Квазиньютоновские методы и метод сопряженных градиентов являются наиболее часто используемыми при оптимизации достаточно гладких функций. При этом квазиньютоновские методы расходуют дополнительные операции и память на вычисление, хранение и использование матричных оценок, зато могут применять более экономичное правило Вулфа для определения шагового множителя. В то же время, метод сопряженных градиентов не тратит вычислительные ресурсы и память на построение матричных оценок, но

требует значительных затрат на дополнительные вычисления градиента для реализации «аккуратного» одномерного поиска. Поэтому можно отдать предпочтение методам сопряженных градиентов, если вычисления  $\nabla Q(x)$  относительно «дешевы», а размерность  $N$  велика. В обратной ситуации более экономичными могут оказаться квазиньютоновские методы (но только при хорошо продуманной вычислительной реализации).

#### 4.5.6. Методы прямого поиска

**Определение.** *Методы прямого поиска* — это методы нулевого порядка, не предполагающие гладкости функции.

Функция может не только не иметь производных, но может содержать разрывы. При поиске минимума эти методы измеряют только значения функций. Наиболее популярными в практике методами этого класса являются методы Хука–Дживса и Нелдера–Мида [2]. Ниже рассмотрен первый.

#### Метод Хука–Дживса

В этом методе поиск минимума на каждом шаге происходит в результате смещения вдоль некоторого направления — образца, которое строится, а затем корректируется в результате специальных пробных покоординатных перемещений, называемых построением конфигурации. *Построение конфигурации* из точки  $y$  выполняет отображение  $F : y \rightarrow z$ , где  $F$  — оператор построения конфигурации:  $z = F(y)$ . Он устроен так, что направление  $z - y$  является направлением убывания функции  $Q$  в точке  $y$ .

Для построения оператора  $F$  введем следующие обозначения:  $e^i$  —  $i$ -й координатный орт,  $h > 0$  — параметр, определяющий величину координатного перемещения. Тогда переход от  $y$  к  $z$  осуществляется согласно следующему алгоритму построения конфигурации:

**Шаг 0.** Полагаем  $\bar{x} = y$ .

**Шаг 1.** Для  $i$  от 1 до  $N$  выполняем действия:

- если  $Q(\bar{x} + he^i) < Q(\bar{x})$ , то  $\bar{x} := \bar{x} + he^i$ ;
- иначе, если  $Q(\bar{x} + he^i) \geq Q(\bar{x}) > Q(\bar{x} - he^i)$ , то  $\bar{x} := \bar{x} - he^i$ ;

**Шаг 2.**  $z := \bar{x}$ .

Описание алгоритма метода Хука–Дживса представлено на рис. 4.29.

Метод Хука–Дживса является эвристическим, был предложен для поиска минимума функций с «овражной» структурой. Обладает способностью достаточно быстро (по числу основных итераций) приходиться в окрестности решения, но не может быстро его уточнять. На рис. 4.30. представлен пример его применения при  $\alpha = 3$ .

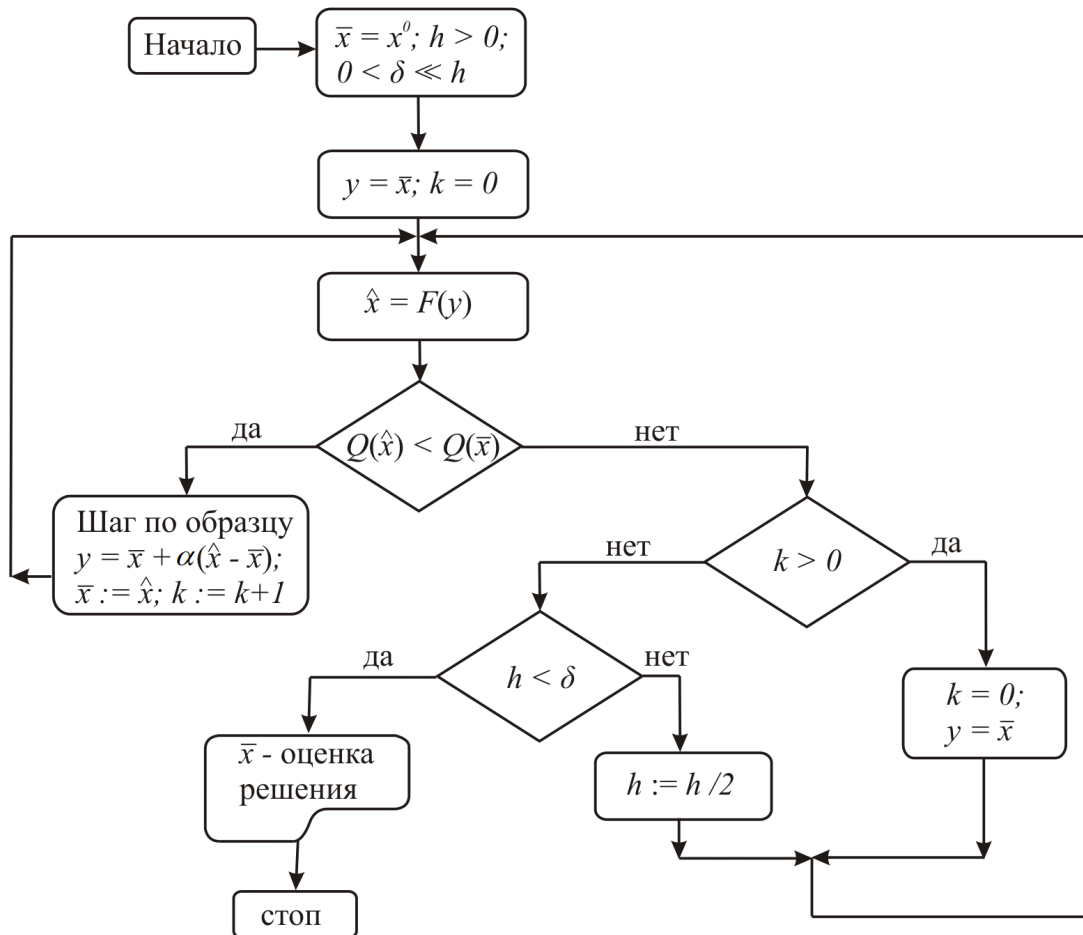


Рис. 4.29. Блок-схема метода Хука–Дживса; в авторской версии  $\alpha = 2$

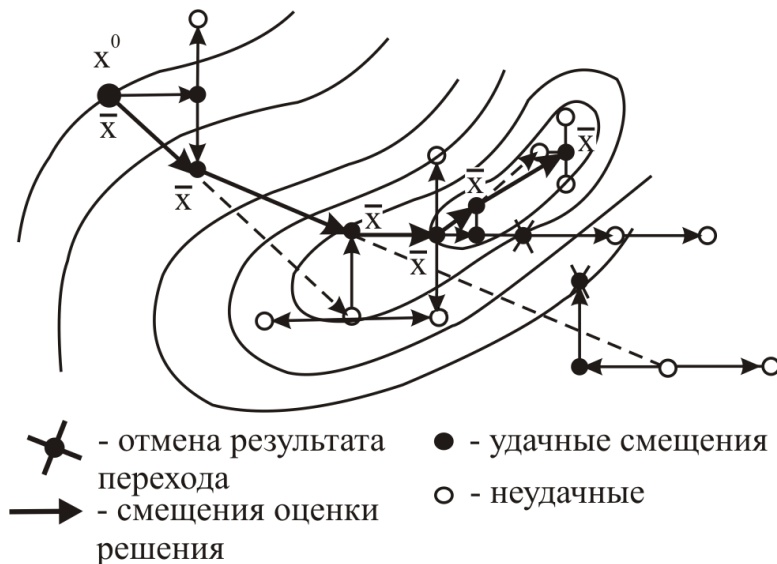


Рис. 4.30. Пример построения траектории методом Хука–Дживса при значении  $\alpha = 3$

После знакомства с методами безусловной локальной оптимизации, рассмотрим численные методы в задачах с ограничениями.

## 4.6. Численные методы решения задач с ограничениями

Рассмотрим задачу математического программирования:

$$\min_{x \in D} Q(x), \quad (4.75)$$

$$D = \{x \in E : g_i(x) \leq 0, i = 1, \dots, m; h_j(x) = 0, j = 1, \dots, p\}, \quad (4.76)$$

где  $E$  — некоторое множество в  $R^N$  простой структуры. Например,  $E = R^N$  или  $E = \{x \in R^N : a_i \leq x_i \leq b_i, i = 1, \dots, N\}$ .  $E$  является областью поиска решения, на  $E$  определены все функции задачи:  $Q, g, h$ .  $D$  — допустимое множество. В курсе лекций представлены только общие методы учета ограничений. Основная идея *общих методов учета ограничений* состоит в том, чтобы для задачи (4.75), (4.76) с функциональными ограничениями построить специальное семейство вспомогательных задач без функциональных ограничений:

$$S_{\lambda_k}(x) \rightarrow \min, x \in E, \quad (4.77)$$

в которых параметры  $\lambda_k$  должны подстраиваться или выбираться таким образом, чтобы обеспечивалась в том или ином смысле сходимость решений  $x_{\lambda_k}^*$  вспомогательных задач (4.77) к решениям исходной задачи. При этом предполагается, что для определения решений  $x_{\lambda_k}^*$  вспомогательных задач могут быть применены ранее рассмотренные методы поиска минимума без функциональных ограничений (для случая  $E = R^N$ ) или их модификации (для  $E \neq R^N$ ).

Известны несколько подходов к реализации этой общей идеи. Они представлены на схеме (рис. 4.31).

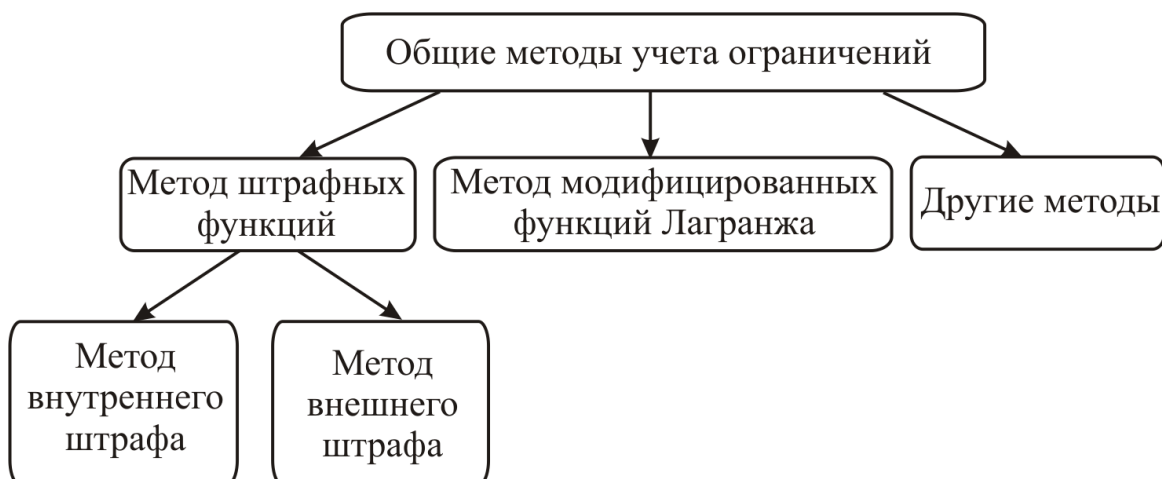


Рис. 4.31. Некоторые общие методы учета ограничений

Построению и исследованию общих и специальных методов учета ограничений посвящена обширная литература (см., например, [5], [14], [15], [19],

[23], [24], [25]). В силу ограниченности объема читаемых лекционных курсов, в данном учебном пособии рассмотрено только два общих метода. Первый, один из самых простых и наглядных, — метод внешнего штрафа. Его описание можно найти почти в каждой из указанных книг, а также в электронном материале для лабораторного практикума [21]. Метод штрафов весьма неприхотлив по требованиям к свойствам функций решаемой задачи, однако обладает весьма медленной сходимостью, обусловленной необходимостью неограниченного увеличения коэффициента штрафа.

За последние три–четыре десятилетия были разработаны более сложные, но зато более совершенные общие методы учета ограничений для локальной оптимизации, активно использующие модификации функций Лагранжа, приводящие к появлению у них локальной седловой точки в прямо–двойственном пространстве для невыпуклых задач.

В первую очередь следует указать на методы модифицированных функций Лагранжа (см. [5], [23], [24] и [14]), которые также рассмотрены в данном курсе лекций. В этих методах сходимость может быть обеспечена за счет настройки дополнительных параметров, роль которых выполняют двойственные переменные — множители Лагранжа. Отметим также еще более изощренный метод последовательного квадратичного программирования, основанный на решении последовательности задач квадратичного программирования, в которых целевая функция выбирается как квадратичная аппроксимация функции Лагранжа, а ограничения являются линейными аппроксимациями ограничений исходной задачи в окрестности текущего приближения (см. [14], а также краткое изложение в [2]). Также следует указать на метод точных гладких штрафных функций [14].

Дополнительно назовем также несколько специальных методов: проекции градиента, возможных направлений, условного градиента, условный метод Ньютона ([5], [14], [15], [22]). Специальные методы учета ограничений в лекционных курсах, связанных с данным пособием, не представлены.

#### 4.6.1. Метод внешнего штрафа

Общая идея метода внешнего штрафа чрезвычайно проста: добавить к целевой функции достаточно большую положительную штрафную добавку в области нарушения ограничений, чтобы эти точки оказались отбракованы процедурами поиска минимума.

Введем понятие функции штрафа применительно к методу внешнего штрафа для задачи (4.75), (4.76).

**Определение.** Заданную на множестве  $E$  функцию  $H(x)$ , удовлетворяющую условию (4.78) называют *функцией внешнего штрафа*.

$$H(x) = \begin{cases} 0, & \text{если } x \in D, \\ > 0, & \text{если } x \notin D, \end{cases} \quad (4.78)$$

Целевая функция задачи со штрафом строится в виде

$$S_\gamma(x) = Q(x) + \gamma H(x). \quad (4.79)$$

Параметр  $\gamma > 0$  называют *коэффициентом штрафа*. Метод заключается в решении последовательности задач со штрафом для возрастающей (обычно — неограниченно возрастающей) последовательности чисел  $\gamma = \gamma_k$

$$\min_{x \in E} S_{\gamma_k}(x). \quad (4.80)$$

Заметим, что минимум в (4.80) берется на множестве  $E$ , а все функциональные ограничения учтены в  $S_{\gamma_k}(x)$  из (4.79).

Пусть  $x_{\gamma_k}^*$  — решения задач со штрафом (4.79), а  $x_{\gamma_k}^*(\varepsilon_k)$  — приближенные решения этих задач, полученные численным методом. Погрешность приближенных решений понимается в следующем смысле:

$$S_{\gamma_k}(x_{\gamma_k}^*(\varepsilon_k)) \leq S_{\gamma_k}(x_{\gamma_k}^*) + \varepsilon_k.$$

Поскольку геометрическая структура допустимого множества  $D$  неизвестна и неявно определяется функциями ограничений  $g(x), h(x)$ , то штраф  $H(x)$  в (4.78) можно определить в виде функции, зависящей от значений  $g(x)$  и  $h(x)$ , т.е. в виде  $H(x) = \tilde{H}(g(x), h(x))$ . Существуют различные способы задания этой функции. Наиболее простой и достаточно часто используемой является *степенная функция штрафа* вида

$$H(x) = \tilde{H}(g(x), h(x)) = \sum_{i=1}^m (\max\{0; c_i g_i(x)\})^r + \sum_{j=1}^p |C_j h_j(x)|^r, \quad (4.81)$$

где  $c_i, C_j > 0, r > 0$ .

### Свойства функции внешнего степенного штрафа

- 1). При  $0 < r \leq 1$  функция  $\tilde{H}(g, h)$  из (4.81) не является гладкой по переменным  $g$  и  $h$ .
- 2). При  $1 < r \leq 2$  функция  $\tilde{H}(g, h)$  из (4.81) гладкая до первого порядка по  $g$  и  $h$ .
- 3.) При  $2 < r \leq 3$  функция  $\tilde{H}(g, h)$  из (4.81) гладкая до второго порядка по  $g$  и  $h$ .

**Замечание.** На практике лучше всего брать  $r$  чуть больше двух или равным двум.

## Геометрическая интерпретация метода

На рис. 4.32 в качестве геометрической иллюстрации метода приведен вид функции штрафа  $H(x)$  и функции вспомогательной штрафной задачи  $S_{\gamma_k}(x)$  в одномерном случае ( $N = 1$ ) при двух ограничениях–неравенствах (т.е.  $m = 2, p = 0$ ).

Обратите внимание, что пример подобран так, что безусловный минимум функции  $Q(x)$  (имеется ввиду минимум на  $E$ ) не принадлежит допустимому множеству  $D$ , а условный минимум  $x^*$  (т.е. минимум с учетом всех ограничений) располагается на границе  $D$ , причем так, что производная  $Q'(x^*)$  отлична от нуля. За счет этого при удалении от множества  $D$  влево  $Q(x)$  строго локально убывает. За счет того, что  $r = 2$ , штрафная функция обладает гладкостью, поэтому  $H'(x^*) = 0$ . Значит  $S'_{\gamma_k}(x^*) = Q'(x^*) \neq 0$ . Следовательно, при сколь угодно большом  $\gamma_k$  точка  $x_{\gamma_k}^* \notin D$ , а может лишь приближаться к  $x^*$  с увеличением  $\gamma_k$ .

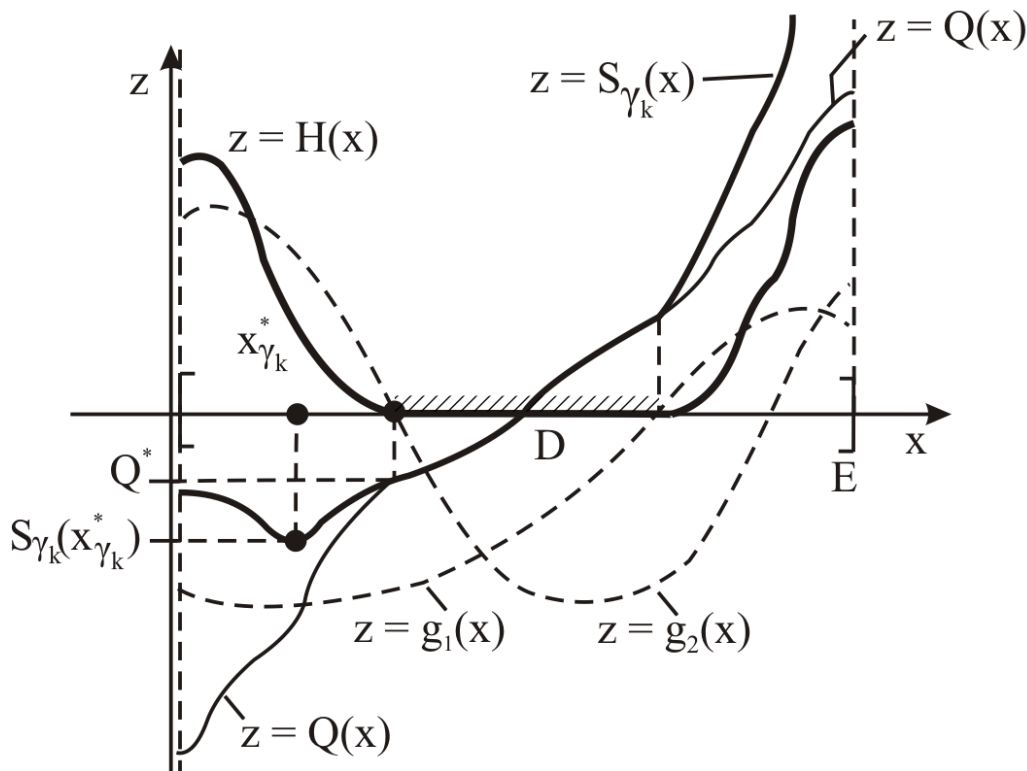


Рис. 4.32. Вид функции штрафа  $H(x)$  при  $r = 2$  и поведение функции штрафной задачи  $S_{\gamma_k}(x)$  при  $\gamma_k = 1$  в конкретном примере

## Свойства метода внешнего штрафа

**Свойство 1.** Достаточные условия сходимости

**Теорема** (сходимости). Пусть выполнены следующие условия:

1) функции  $Q(x)$ ,  $H(x)$  непрерывны на  $E$  и существует множество глобаль-

но-оптимальных решений исходной задачи  $X^* = \{x^*\}$ ;

2) для всяких достаточно больших  $\gamma_k > \bar{\gamma}$  и достаточно малых  $\varepsilon_k \geq 0$  существуют приближенные решения вспомогательных задач  $x_{\gamma_k}^*(\varepsilon_k)$ , не выходящие за пределы некоторого компакта  $K \subseteq E$ ;

3) при  $k \rightarrow \infty$  имеют место пределы:  $\gamma_k \rightarrow \infty$ ,  $\varepsilon_k \rightarrow 0$ . Тогда:

I) любая предельная точка  $x_\infty^*$  последовательности  $x_{\gamma_k}^*(\varepsilon_k)$  является решением исходной задачи, т.е.  $x_\infty^* \in X^*$  — множеству глобальных минимумов;

II) существует предел

$$\lim_{k \rightarrow \infty} Q(x_{\gamma_k}^*(\varepsilon_k)) = Q^*. \quad (4.82)$$

*Доказательство.* Докажем утверждение I. Так как  $x_{\gamma_k}^*$  — глобальный минимум  $S_{\gamma_k}$ , верна следующая оценка:

$$\varepsilon_k + S_{\gamma_k}(x^*) \geq S_{\gamma_k}(x_{\gamma_k}^*) + \varepsilon_k \geq S_{\gamma_k}(x_{\gamma_k}^*(\varepsilon_k)) = Q(x_{\gamma_k}^*(\varepsilon_k)) + \gamma_k H(x_{\gamma_k}^*(\varepsilon_k)). \quad (4.83)$$

Поскольку  $x^* \in D$ , то  $S_{\gamma_k}(x^*) = Q(x^*)$ . Разделим неравенство (4.83) на  $\gamma_k$ :

$$\frac{\varepsilon_k}{\gamma_k} + \frac{Q(x^*)}{\gamma_k} \geq \frac{Q(x_{\gamma_k}^*(\varepsilon_k))}{\gamma_k} + H(x_{\gamma_k}^*(\varepsilon_k)).$$

Поскольку последовательность  $\{x_{\gamma_k}^*(\varepsilon_k)\}$  принадлежит компакту  $K$ , у неё существуют сходящиеся подпоследовательности. Пусть  $x_\infty^*$  — предельная точка одной из них. Пусть этой подпоследовательности соответствуют значения  $k = k_s$ ,  $s = 1, 2, \dots$

В силу предположений теоремы значение  $Q(x^*)$  существует и ограничено, а поскольку  $Q$  и  $H$  непрерывны на  $E$ , и  $\gamma_k \rightarrow \infty$ , то в пределе при  $k = k_s$ ,  $s \rightarrow \infty$  получим:

$$0 \geq H(x_\infty^*).$$

Однако, из (4.78) всегда  $H(x) \geq 0$ , значит  $H(x_\infty^*) = 0$ . Из этого следует, что любая предельная точка  $x_\infty^*$  последовательности  $x_{\gamma_k}^*(\varepsilon_k)$  является допустимой.

Вернемся к (4.83). Так как

$$Q(x_{\gamma_k}^*(\varepsilon_k)) + \underbrace{\gamma_k H(x_{\gamma_k}^*(\varepsilon_k))}_{\geq 0} \geq Q(x_{\gamma_k}^*(\varepsilon_k)),$$

то из (4.83) следует:

$$\varepsilon_k + Q(x^*) \geq Q(x_{\gamma_k}^*(\varepsilon_k)).$$

Рассмотрим ту же, что и ранее, подпоследовательность с  $k = k_s$  и перейдем к пределу при  $s \rightarrow \infty$ , получим:

$$Q(x^*) \geq Q(x_\infty^*).$$



Поскольку  $x^*$  — глобальный минимум  $Q$  на  $D$  и  $x_\infty^* \in D$ , т.е.  $x_\infty^*$  — допустима, то  $Q(x^*) = Q(x_\infty^*)$  и, следовательно,  $x_\infty^* \in X^*$ , т.е. принадлежит множеству глобальных минимумов задачи.

Докажем утверждение II. Пусть (4.82) неверно. Тогда найдется такое  $\delta > 0$  и подпоследовательность  $x_{\gamma_k}^*(\varepsilon_k)$  с  $k = k_s$ , что  $\forall s = 1, 2, \dots$  выполнится:

$$\|Q(x_{\gamma_{k_s}}^*(\varepsilon_{k_s})) - Q^*\| \geq \delta.$$

Однако, если из  $\{x_{\gamma_{k_s}}^*(\varepsilon_{k_s})\}$  выделить сходящуюся подпоследовательность с номерами элементов  $k_{s_t}, t = 1, 2, \dots$ , что всегда возможно, то  $Q(x_{\gamma_{k_{s_t}}}^*(\varepsilon_{k_{s_t}}))$ , в силу доказанного выше, будет сходиться к  $Q^*$ , т.е. на элементах этой подпоследовательности, начиная с некоторого  $T$  для всех  $t \geq T$ :

$$\|Q(x_{\gamma_{k_{s_t}}}^*(\varepsilon_{k_{s_t}})) - Q(x^*)\| < \delta,$$

что противоречит предыдущему. Таким образом, утверждение (II) верно.  $\square$

**Свойство 2.** Плохая обусловленность задач. Если  $x_{\gamma_k}^* \notin D$ , то функции  $S_{\gamma_k}(x)$  вспомогательных задач при достаточно больших значениях параметра  $\gamma_k$  плохо обусловлены в окрестности  $x_{\gamma_k}^*$ , а именно, для граничных собственных чисел матриц вторых производных  $\nabla_{xx}^2 S_{\gamma_k}(x_{\gamma_k}^*)$ , вычисляемых в точках минимумов вспомогательных задач (4.80), выполнено:

$$0 < \lambda_{\min} \ll \lambda_{\max}.$$

Приводится без доказательства.

Свойство 2 означает, что при больших коэффициентах штрафа функции задач со штрафом будут сильно овражны.

**Свойство 3.** Характер приближения оценок к решению

**Теорема.** Если в методе штрафов коэффициенты штрафа образуют строго возрастающую последовательность  $\gamma_{k+1} > \gamma_k$ , а вспомогательные задачи (4.80) решаются точно (т.е.  $\varepsilon_k = 0$ ), то последовательность значений  $Q(x_{\gamma_k}^*)$  будет неубывающей, а последовательность  $H(x_{\gamma_k}^*)$  — невозрастающей:

$$Q(x_{\gamma_{k+1}}^*) \geq Q(x_{\gamma_k}^*), \quad H(x_{\gamma_{k+1}}^*) \leq H(x_{\gamma_k}^*).$$

*Доказательство.* Рассмотрим два значения коэффициента штрафа  $\gamma_{k+1} > \gamma_k$ . Тогда будут справедливы два очевидных неравенства:

$$\begin{aligned} Q(x_{\gamma_k}^*) + \gamma_k H(x_{\gamma_k}^*) &\leq Q(x_{\gamma_{k+1}}^*) + \gamma_k H(x_{\gamma_{k+1}}^*), \\ Q(x_{\gamma_{k+1}}^*) + \gamma_{k+1} H(x_{\gamma_{k+1}}^*) &\leq Q(x_{\gamma_k}^*) + \gamma_{k+1} H(x_{\gamma_k}^*). \end{aligned}$$

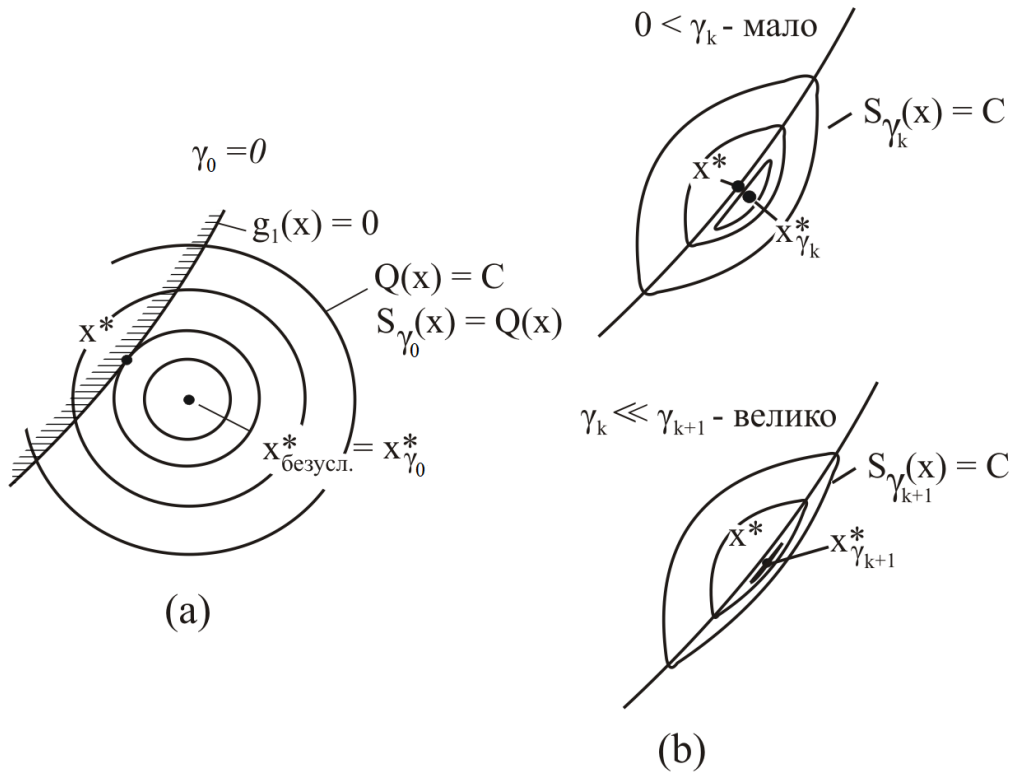


Рис. 4.33. (a) — вид линий равного уровня исходной задачи; (b) — возрастание овражности функции  $S_{\gamma_k}(x)$  с ростом  $\gamma_k$

После преобразования каждого из них получим, что

$$0 \leq Q(x^*_{\gamma_{k+1}}) - Q(x^*_{\gamma_k}) + \gamma_k(H(x^*_{\gamma_{k+1}}) - H(x^*_{\gamma_k})),$$

$$Q(x^*_{\gamma_{k+1}}) - Q(x^*_{\gamma_k}) + \gamma_{k+1}(H(x^*_{\gamma_{k+1}}) - H(x^*_{\gamma_k})) \leq 0.$$

Заметим, что эти два выражения совпадают с точностью до значения  $\gamma$ . Мы видим, что при увеличении коэффициента штрафа приведенное выше выражение из неотрицательного становится неположительным. Это возможно только при  $Q(x^*_{\gamma_{k+1}}) - Q(x^*_{\gamma_k}) \geq 0$  и  $H(x^*_{\gamma_{k+1}}) - H(x^*_{\gamma_k}) \leq 0$ , что и требовалось.  $\square$

Заметим, что свойство 3 может быть проинтерпретировано следующим образом: *если безусловный минимум функции  $Q$  выходит за границы множества  $D$  (тогда условный минимум  $x^*$  лежит на его границе), и  $\nabla Q(x^*) \neq 0$ , то подпоследовательность  $x^*_{\gamma_{k_s}}$ , сходящаяся к  $x^*$ , будет подходить к  $x^*$  извне множества  $D$ .*

#### Свойство 4. Оценки скорости сходимости

Прежде чем сформулировать само свойство, опишем дополнительные требования к задаче и процессу решения.

### Дополнительные предположения

- A.** Вспомогательные задачи со штрафом (4.80) решаются точно (т.е.  $\varepsilon_k \equiv 0$ ).
- B.** Используется функция степенного штрафа (4.81) с показателем степени  $r > 0$ .
- C.** Множество  $E$  — замкнуто, а  $Q$  — липшицева на  $E$  с константой  $L$  по отношению к норме  $\|x' - x''\|$ .
- D.**  $\exists \delta > 0$  и  $\alpha > 0$ , что  $\forall x \in (O_\delta(D) \setminus D) \cap E$  для функции невязки вида

$$G(x) = \max \{ \max \{ c_j g_j(x) : j = 1, \dots, m \} ; \max \{ |C_s h_s(x)| : s = 1, \dots, p \} ; 0 \}$$

и функции расстояния  $\rho(x, D) = \min \{ \|x - y\| : y \in D \}$  до множества  $D$  выполняется неравенство  $G(x) \geq \alpha \rho(x, D)$ .

Погрешность решения задачи с ограничениями оценим величиной

$$\Delta(\gamma) = Q(x^*) - S_\gamma(x_\gamma^*). \quad (4.84)$$

**Теорема** (об оценках скорости сходимости). Пусть выполнены условия теоремы сходимости и, кроме того, дополнительные предположения  $A - -D$ , тогда:

- 1) при  $r \leq 1$  существует достаточно большое  $\bar{\gamma}$ , что  $\forall \gamma \geq \bar{\gamma}$  оценка погрешности решения становится равной нулю, т.е.  $\Delta(\gamma) = 0$ ;
- 2) при  $r > 1$  и достаточно больших  $\gamma$

$$0 \leq \Delta(\gamma) \leq \frac{c}{\gamma^{\frac{1}{r-1}}}, \quad (4.85)$$

где  $c = (r - 1)(L/(r \cdot \alpha))^{\frac{r}{r-1}}$ .

Приводится без доказательства.

Заметим, что в случае  $r \leq 1$  возможно точное решение задачи за счет нарушения гладкости. Однако нарушение гладкости препятствует использованию локальных методов с высоким порядком сходимости.

Следует обратить внимание, что при обычно используемом  $r = 2$  оценка (4.85) погрешности (4.84) обратно пропорциональна коэффициенту штрафа  $\gamma$ . Для обеспечения высокой точности значение  $\gamma$  должно стать очень большим, но это приведет к плохой обусловленности задачи со штрафом, что вызовет потерю точности при численном решении задач (4.80).

**Замечание.** Метод внешнего штрафа многократно решает задачи вида (4.80), увеличивая  $\gamma_k$  и повышая точность решения задач (т.е. уменьшая  $\varepsilon_k$ ). При этом в качестве критерия останова можно использовать условие

$$G(x_{\gamma_k}^*(\varepsilon_k)) \leq \delta,$$

т.е. условие малости невязки в ограничениях.  $\delta > 0$  — параметр останова.

### 4.6.2. Метод модифицированных функций Лагранжа

В предыдущем подразделе было показано, что метод штрафных функций имеет ряд существенных недостатков. Многие из них удается преодолеть, если штрафную добавку присоединять не к целевой функции, а к функции Лагранжа. Множители Лагранжа выполняют при этом роль дополнительных настроечных коэффициентов, за счет изменения которых можно добиться совмещения минимума вспомогательной функции с минимумом исходной задачи. Возникающий при этом метод называют *методом модифицированных функций Лагранжа* [14], [24].

#### Построение метода для задач с равенствами

Пусть поставлена задача математического программирования с ограничениями–равенствами:

$$\min_{x \in D} Q(x), \quad D = \{x \in R^N : h(x) = 0\}. \quad (4.86)$$

В случае регулярности допустимого множества функция Лагранжа для такой задачи имеет следующий вид:

$$L(x, \mu) = Q(x) + (\mu, h(x)).$$

В гладком регулярном случае для (4.86) пара  $(x^*, \mu^*)$ , соответствующая ее локальному минимуму в точке  $x^*$ , является стационарной точкой функции Лагранжа. Как было показано в разделе 2, при выпуклости  $Q$  и аффинности  $h(x)$  пара  $(x^*, \mu^*)$  становится седловой точкой функции Лагранжа. Но в общем случае седловой точки у функции Лагранжа в  $(x^*, \mu^*)$  может не существовать, поскольку это условие в невыпуклом случае не является необходимым для оптимальности.

**Пример.** Рассмотрим задачу на минимум с невыпуклой функцией  $Q(x) = x^3$  и одним ограничением  $x + 1 = 0$ .

В этой задаче существует единственная допустимая точка  $x = -1$ . Условие экстремума дает  $-3(-1)^2 = \mu^*$ , следовательно  $\mu^* = -3$ , и  $L(x, \mu^*) = x^3 - 3(x+1)$ . В точке  $x = x^* = -1$ :  $\nabla_x L(x^*, \mu^*) = 0$ ,  $\nabla_{xx}^2 L(x^*, \mu^*) = -6 < 0$ , следовательно, при  $x = x^*$  по переменной  $x$  у функции Лагранжа достигается максимум, а не минимум.

Чтобы исправить рассмотренную ситуацию, следует так изменить функцию Лагранжа, чтобы модифицированная функция сохранила все основные свойства обычной функции Лагранжа, но для нее точка  $(x^*, \mu^*)$  стала хотя бы локально седловой.

Характер изменений может быть понят исходя из следующего наблюдения. В задаче (4.86) на допустимом множестве, т.е. при  $x \in D$ :

$$L(x, \mu) = Q(x).$$

Поэтому на множестве  $D$  функция Лагранжа заведомо имеет локальный минимум в точке  $x^*$ . Если же  $x$  покидает допустимое множество, то при используемых конкретных значениях  $\mu \neq \mu^*$ , в зависимости от направления смещения по переменным  $x$ , функция Лагранжа будет либо возрастать, либо убывать и даже при  $\mu = \mu^*$  станет иметь в  $x^*$  лишь стационарную точку, которая в общем случае может не являться локальным минимумом  $L(x, \mu^*)$  по  $x$ . Таким образом, следует добавить к ней дополнительное слагаемое, локально возрастающее при удалении от множества  $D$ .

Таким образом, приходим к *модифицированной функции Лагранжа*

$$L_\gamma(x, \mu) = Q(x) + (\mu, h(x)) + \frac{\gamma}{2} \|h(x)\|^2, \quad (4.87)$$

где  $\gamma > 0$ .

Наложим на задачу (4.86) некоторые ограничения:

- А) Пусть в задаче (4.86), функции  $Q$ ,  $h$  дважды дифференцируемы по крайней мере в некоторой окрестности допустимой точки  $x^*$  и их вторые производные непрерывны в этой точке.
- В) В точке  $x^*$  выполнено достаточное условие регулярности допустимого множества в форме линейной независимости градиентов  $\nabla h_1(x^*)$ ,  $\dots$ ,  $\nabla h_p(x^*)$ .
- С) Существуют значения  $\mu^*$  множителей Лагранжа, однозначно определяемые из условия стационарности

$$\nabla_x L(x^*, \mu^*) = 0,$$

- Д) Для пары  $x^*, \mu^*$  дополнительно выполняется условие второго порядка:

$$\forall d \neq 0 : (d, \nabla h_j(x^*)) = 0 \quad (j = 1, \dots, p)$$

$$d^T \nabla_{xx}^2 L(x^*, \mu^*) d > 0.$$

Заметим, что дополнительные требования  $A - -D$  совпадают с достаточными условиями второго порядка для строгого локального минимума. Их выполнение обеспечивает наличие в точке  $x^*$  строгого локального минимума исходной задачи (4.86).

Концептуально важным моментом является то, что если в условиях оптимальности первого и второго порядков заменить обычную функцию Лагранжа  $L(x, \mu)$  на модифицированную  $L_\gamma(x, \mu)$ , то все утверждения останутся справедливыми (предлагается самостоятельно проверить). Это и определяет название функции  $L_\gamma(x, \mu)$ .

**Лемма.** Пусть задача (4.86) удовлетворяет свойствам  $A--D$ . Тогда найдется такое  $\bar{\gamma} > 0$ , что модифицированная функция Лагранжа вида (4.88)

$$L_\gamma(x, \mu) = Q(x) + (\mu, h(x)) + \frac{\gamma}{2} \|h(x)\|^2 \quad (4.88)$$

при любых  $\gamma > \bar{\gamma} > 0$  будет иметь для  $x$  из некоторой окрестности  $U(x^*)$  точки  $x^*$  локальную седловую точку  $(x^*, \mu^*)$ , т.е.  $\forall x \in U(x^*), \forall \mu \in R^p$ .

$$L_\gamma(x^*, \mu) \leq L_\gamma(x^*, \mu^*) \leq L_\gamma(x, \mu^*).$$

Приводится без доказательства.

Лемма показывает, что даже для невыпуклых задач точке  $x^*$  строгого локального минимума в (4.86) соответствует локальная седловая точка  $(x^*, \mu^*)$  модифицированной функции Лагранжа. Будем отыскивать её, минимизируя  $L_\gamma(x, \mu)$  по  $x$  и, одновременно, максимизируя по  $\mu$ .

Для получения конкретной формы метода необходимо построить правила коррекции оценок множителей Лагранжа. С этой целью запишем условие оптимальности для вспомогательной задачи (4.89), учитывая, что окрестность  $U(x^*)$  является открытым множеством,

$$x^{k+1} = \arg \min_{x \in U(x^*)} L_{\gamma_k}(x, \mu^k); \quad (4.89)$$

в точке  $x^{k+1}$ .

$$\nabla Q(x^{k+1}) + \sum_{s=1}^p \nabla h_s(x^{k+1}) \mu_s^k + \gamma_k \sum_{s=1}^p \nabla h_s(x^{k+1}) h_s(x^{k+1}) = 0.$$

Если бы точка  $x^{k+1}$  была решением исходной задачи, то соответствующие множители Лагранжа  $\tilde{\mu}_s$  можно было бы определить из решения системы

$$-\nabla Q(x^{k+1}) = \sum_{s=1}^p \tilde{\mu}_s \nabla h_s(x^{k+1}). \quad (4.90)$$

Если  $x^{k+1}$  близка к  $x^*$ , то по введенным в пункте (B) требованиям  $\tilde{\mu}_s$  определяются из этой системы единственным образом. Сравнивая (4.90) с предыдущим равенством, видим, что в векторной форме записи

$$\tilde{\mu} = \mu^k + \gamma_k h(x^{k+1}).$$

В силу предполагаемой близости  $x^{k+1}$  к  $x^*$  следует принять  $\mu^{k+1} = \tilde{\mu}$ .

Таким образом, метод модифицированных функций Лагранжа может быть представлен в виде следующего итерационного соотношения:

$$x^{k+1} = \arg \min_{x \in R^N} L_{\gamma_k}(x, \mu^k), \quad (4.91)$$

$$\mu^{k+1} = \mu^k + \gamma_k h(x^{k+1}) \quad (4.92)$$

Заметим, что  $h(x^{k+1}) = \nabla_{\mu} L_{\gamma_k}(x^{k+1}, \mu^k)$ . Таким образом, формула (4.92) описывает шаг метода градиентного поиска в задаче максимизации функции  $L_{\gamma}(x^{k+1}, \mu)$  по  $\mu$ . При этом величина шагового множителя  $t_k = \gamma_k$ . Оказывается, в этом методе, в отличие от метода штрафов, последовательность  $\gamma_k$  не обязана стремиться к бесконечности для обеспечения сходимости. Сходимость может быть получена за счет настройки множителей  $\mu_k$  по правилу (4.92).

**Теорема** (об условиях и характере сходимости [14]). *Пусть для задачи с ограничениями–равенствами (4.86) выполнены условия A–D. Тогда  $\exists \bar{\gamma} > 0$ ,  $\delta > 0$ ,  $T > 0$ ,  $\bar{\delta} \in (0, \delta]$  и окрестность  $U(x^*)$  точки  $x^*$ , что при выполнении для начальных условий (см. рис. 4.34) требований:*

$$(\mu^0, \gamma_0) \in \Theta(\bar{\delta}, \bar{\gamma}) = \{(\mu, \gamma) : \|\mu - \mu^*\| < \bar{\delta}\gamma, \gamma \geq \bar{\gamma}\},$$

*а также требований  $\bar{\gamma} \leq \gamma_k \leq \gamma_{k+1}$ , на всех последующих шагах метода модифицированных функций Лагранжа (4.91)–(4.92) – будет выполнено:*

- 1) вспомогательные задачи (4.91) будут иметь в окрестности  $U(x^*)$  единственную стационарную точку, соответствующую  $x^{k+1}$ ;
- 2) всегда будет выполнено

$$(\mu^{k+1}, \gamma_{k+1}) \in \Theta(\delta, \bar{\gamma}) = \{(\mu, \gamma) : \|\mu - \mu^*\| < \delta\gamma, \gamma \geq \bar{\gamma}\}.$$

- 3) кроме того,  $\mu^k \rightarrow \mu^*$  и  $x^k \rightarrow x^*$  при  $k \rightarrow \infty$ , причем правило коррекции (4.92) определит оценку  $\mu^{k+1}$  так, что выполняются неравенства:

$$\|x^{k+1} - x^*\| < \frac{T}{\gamma_k} \|\mu^k - \mu^*\|, \quad (4.93)$$

$$\|\mu^{k+1} - \mu^*\| < \frac{T}{\gamma_k} \|\mu^k - \mu^*\|, \quad (4.94)$$

*что обеспечивает линейную скорость сходимости по множителям Лагранжа, если  $\gamma_k$  остается конечным, а при  $\gamma_k \rightarrow \infty$  сверхлинейную сходимость. При обеспечении линейной сходимости коэффициент сжатия  $T/\gamma_k$  может быть сделан сколь угодно малым за счет выбора значения коэффициента штрафа  $\gamma_k$ .*

Несмотря на то, что утверждение теоремы носит локальный характер, реальное поведение метода в конкретных задачах показывает, что сходимость к решению обычно наблюдается и при значительных отклонениях по  $x$  и  $\mu$  от  $(x^*, \mu^*)$ .

Теорема приводится без доказательства.

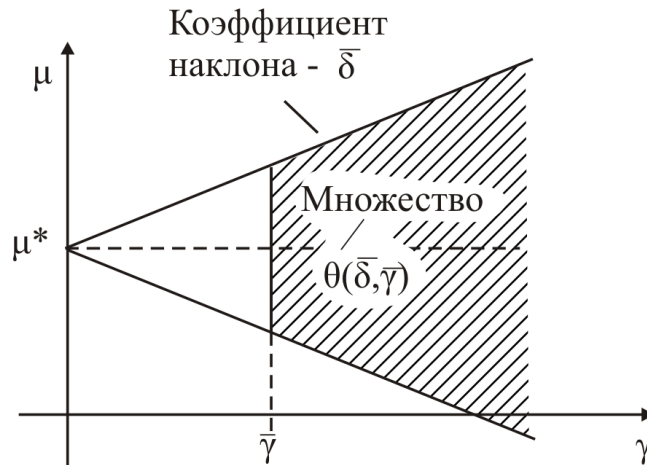


Рис. 4.34. Вид множества для возможного выбора начальных значений пары  $(\mu^0, \gamma_0)$

Укажем также, что в качестве критерия останова в данном методе нельзя принимать малость невязки в ограничениях, поскольку ее величина (в отличие от метода штрафов) ведет себя не монотонно. Останов можно выполнить при малости невязки в системе условий оптимальности:

$$\|h(x^{k+1})\| + \|\nabla Q(x^{k+1}) + \sum \mu_s^{k+1} \nabla h_s(x^{k+1})\| \leq \varepsilon.$$

Это требование можно дополнить условием неубывания невязки в течение заданного числа шагов. Если одно из этих двух условий выполнено, метод останавливается.

### Обобщение метода модифицированных функций Лагранжа на задачи с ограничениями смешанного типа

Рассмотрим задачу поиска минимума при ограничениях — равенствах и неравенствах

$$\min_{x \in D} Q(x), \quad D = \{x \in R^N : g(x) \leq 0, h(x) = 0\}. \quad (4.95)$$

Аналогично предыдущему случаю, наложим ограничения на задачу (4.95):

- A') Функции  $Q, g, h$  дважды дифференцируемы по крайней мере в некоторой окрестности допустимой точки  $x^*$  и их вторые производные непрерывны в этой точке.
- B') В  $x^*$  выполнено достаточное условие регулярности допустимого множества в форме линейной независимости градиентов

$$\nabla h_1(x^*), \dots, \nabla h_p(x^*), \nabla g_{i_1}(x^*), \dots, \nabla g_{i_r}(x^*),$$



где  $\{i_1, \dots, i_r\} = I(x^*)$  — набор номеров ограничений–неравенств, активных в точке  $x^*$ .

C') Существуют значения множителей Лагранжа  $\lambda_{i_1}^* \geq 0, \dots, \lambda_{i_r}^* \geq 0, \mu^*$ , однозначно определяемые из условия стационарности

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0$$

при  $\lambda_j^* = 0$  для  $j \notin I(x^*)$ .

D') Выполняется условие второго порядка:

$$\forall d \neq 0 : (d, \nabla h_j(x^*)) = 0 \quad (j = 1, \dots, p), \quad (d, \nabla g_s(x^*)) = 0 \quad (s = 1, \dots, r) :$$

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) d > 0.$$

Заметим, что при выполнении  $A' - -D'$  задача (4.95) имеет в точке  $x^*$  строгий локальный минимум.

Чтобы воспользоваться методом, ранее полученным для задач с ограничениями–равенствами, избавимся от ограничений–неравенств. Рассмотрим неравенства  $g_i(x) \leq 0$ . Введем дополнительную переменную  $z_i$ , чтобы  $g_i(x) = -z_i^2$ . Очевидно, что это равенство в пространстве переменных  $x, z$  равносильно предыдущему неравенству. Отсюда, эквивалентной формой записи неравенств будет следующая форма их записи в виде равенств:

$$g_i(x) + z_i^2 = 0.$$

Построим модифицированную функцию Лагранжа для задачи с этими дополнительными равенствами:

$$L_\gamma(x, z, \lambda, \mu) = Q(x) + (\mu, h(x)) + \frac{\gamma}{2} \sum_{i=1}^p h_i^2(x) + \underbrace{\sum_{i=1}^m \lambda_i (g_i(x) + z_i^2) + \frac{\gamma}{2} \sum_{i=1}^m (g_i(x) + z_i^2)^2}_{F(x, z)}. \quad (4.96)$$

Найдем минимум по  $z$  аналитически. Продифференцируем (4.96) по  $z_i^2$  и приравняем частную производную к нулю:

$$\lambda_i + \gamma(g_i(x) + z_i^2) = 0.$$

В силу выпуклости  $F(x, z)$  по  $z_i^2$  полученное условие определяет точку минимума по  $z_i^2$ , но без учета неотрицательности.

С учетом ограничения на знак  $z_i^2 \geq 0$ , получим

$$(z_i^*)^2 = \max \left\{ 0; -\frac{\lambda_i}{\gamma} - g_i(x) \right\}.$$

Теперь подставим это значение в (4.96):

$$\begin{aligned} F(x, z)|_{z=z^*} &= \sum_{i=1}^m \lambda_i \max \left\{ -\frac{\lambda_i}{\gamma}; g_i(x) \right\} + \frac{\gamma}{2} \sum_{i=1}^m \left( \max \left\{ -\frac{\lambda_i}{\gamma}; g_i(x) \right\} \right)^2 = \\ &= \frac{2\gamma}{2\gamma} \sum_{i=1}^m \lambda_i \max \{ \dots \} + \frac{\gamma^2}{2\gamma} \sum_{i=1}^m \left( \max \{ \dots \} \right)^2 = \\ &= \frac{1}{2\gamma} \left[ \sum_{i=1}^m \left( \max \{ -\lambda_i; \gamma g_i(x) \} \right)^2 + 2\lambda_i \max \{ -\lambda_i; \gamma g_i(x) \} + \lambda_i^2 - \lambda_i^2 \right] = \\ &= \frac{1}{2\gamma} \sum_{i=1}^m \left( \left( \max \{ 0; \lambda_i + \gamma g_i(x) \} \right)^2 - \lambda_i^2 \right); \end{aligned}$$

$$\begin{aligned} \tilde{L}_\gamma(x, \lambda, \mu) &= Q(x) + (\mu, h(x)) + \frac{\gamma}{2} \sum_{i=1}^p h_i^2(x) + \\ &+ \frac{1}{2\gamma} \sum_{i=1}^m \left( \left( \max \{ 0; \lambda_i + \gamma g_i(x) \} \right)^2 - \lambda_i^2 \right). \quad (4.97) \end{aligned}$$

Поскольку при  $z = z^*(\lambda_k, x^{k+1})$

$$\frac{\partial L_{\gamma_k}}{\partial \lambda_i^k}(x^{k+1}, z, \lambda^k, \mu^k) = \max \left\{ -\frac{\lambda_i^k}{\gamma_k}; g_i(x^{k+1}) \right\},$$

то соотношение для коррекции множителей  $\lambda_i$ , аналогичное (4.92), примет вид

$$\lambda_i^{k+1} = \lambda_i^k + \gamma_k \frac{\partial L_{\gamma_k}}{\partial \lambda_i^k}(x^{k+1}, z^*(\lambda_k, x^{k+1}), \lambda^k, \mu^k).$$

После подстановки выражения для частной производной получим:

$$\lambda_i^{k+1} = \lambda_i^k + \gamma_k \left( \max \left\{ -\frac{\lambda_i^k}{\gamma_k}; g_i(x^{k+1}) \right\} \right).$$

Окончательный вид итерационных формул для метода модифицированных функций Лагранжа при смешанных ограничениях следующий:

$$x^{k+1} = \arg \min_{x \in R^N} \tilde{L}_{\gamma_k}(x, \lambda^k, \mu^k), \quad (4.98)$$

$$\lambda_i^{k+1} = \max \{0; \lambda_i^k + \gamma_k g_i(x^{k+1})\}, \quad (4.99)$$

$$\mu^{k+1} = \mu^k + \gamma_k h(x^{k+1}). \quad (4.100)$$

Для построенного метода (4.98)–(4.100) при выполнении условий  $A' - -D'$  справедлива теорема, полностью аналогичная случаю задач с равенствами [14]. Опуская её полную формулировку, заметим лишь, что структура множества для выбора начальных значений  $(\mu^0, \lambda^0, \gamma^0)$  следующая

$$(\mu^0, \lambda^0, \gamma^0) \in \Theta'(\bar{\delta}, \bar{\gamma}),$$

$$\Theta'(\bar{\delta}, \bar{\gamma}) = \{(\mu, \lambda, \gamma) : \lambda \geq 0, \gamma \geq \bar{\gamma}; \|\mu - \mu^*\| \leq \gamma \bar{\delta}; \|\lambda - \lambda^*\| \leq \gamma \bar{\delta}\},$$

а оценки, аналогичные (4.93)–(4.94) имеют вид (4.101)–(4.103):

$$\|\mu^{k+1} - \mu^*\| \leq \frac{T}{\gamma_k} (\|\lambda^k - \lambda^*\| + \|\mu^k - \mu^*\|), \quad (4.101)$$

$$\|\lambda^{k+1} - \lambda^*\| \leq \frac{T}{\gamma_k} (\|\lambda^k - \lambda^*\| + \|\mu^k - \mu^*\|), \quad (4.102)$$

$$\|x^{k+1} - x^*\| \leq \frac{T}{\gamma_k} (\|\lambda^k - \lambda^*\| + \|\mu^k - \mu^*\|). \quad (4.103)$$

Заметим, что данные оценки говорят о высокой скорости сходимости данного метода, что выгодно отличает его от метода штрафов. Критерий остановки может быть построен аналогично случаю задач с ограничениями–равенствами на основе проверки малости невязки в системе условий оптимальности задачи со смешанными ограничениями.

## 4.7. Методы многоэкстремальной оптимизации

Пусть задано компактное множество элементов  $E \subseteq R^N$ . В большинстве случаев будем предполагать, что

$$E = \{x \in R^N : a_i \leq x_i \leq b_i, i = 1, \dots, N\}, \quad (4.104)$$

где  $a_i$  и  $b_i$  — конечны.

Вначале рассмотрим следующую постановку задачи без функциональных ограничений:

$$\min_{x \in E \subseteq R^N} Q(x). \quad (4.105)$$

Если мы не обладаем информацией о единственности локального минимума, то нужно рассматривать задачу (4.105) как *многоэкстремальную*.

Цель решения задач многоэкстремальной оптимизации заключается в получении (желательно, с наперед заданной точностью) оценки глобально минимального значения  $Q^* = Q(x^*)$  в (4.105), а также, возможно, точки глобального минимума  $x^*$  или множества  $\{X^*\}$  таких минимумов.

Пусть проведено конечное число испытаний задачи (4.105) в точках  $x^1, x^2, \dots, x^k$  с вычислением значений  $Q_1, Q_2, \dots, Q_k$  (см. рис. 4.35a). Что можно сказать о положении глобального минимума? Если о свойствах функции  $Q$  ничего априори неизвестно, то о положении  $x^*$  ничего сказать нельзя (см. рис. 4.35b).

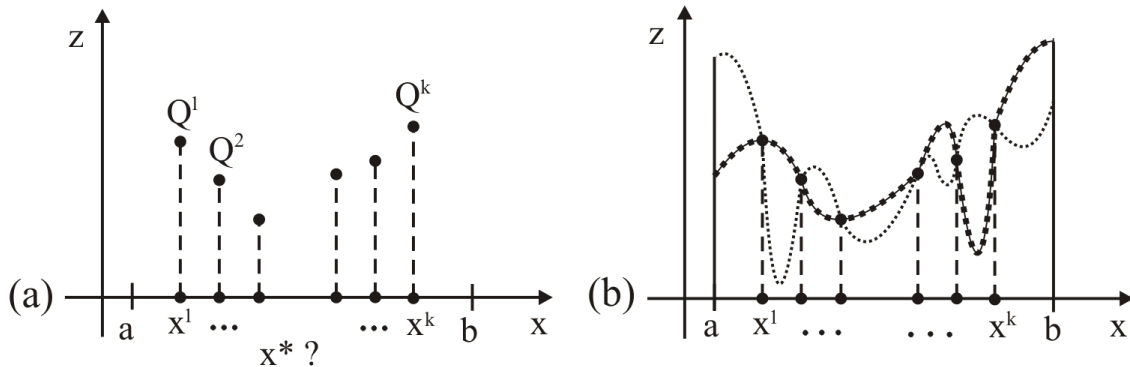


Рис. 4.35. (a) — результаты испытаний многоэкстремальной функции; (b) — несколько вариантов ее возможного поведения

Для возможности построения оценок для  $Q^*$  и  $x^*$  по конечному числу испытаний необходимо наложить специальные требования на рассматриваемый класс задач. Классическим примером таких требований является липшицевость функций [2], [28], [29].

### Липшицевы функции и их свойства

Будем рассматривать класс функций  $Q \in \Phi_{Lip}(E)$ , удовлетворяющих условию Липшица на  $E$  с константой  $L$ . Их называют также *непрерывными по Липшицу*.

Напомним, что функция  $Q$  *липшицева на  $E$  с константой  $L$* , если  $\forall x', x'' \in E: |Q(x') - Q(x'')| \leq L \cdot \|x' - x''\|$ .

**Определение.** Функцию  $Q$  называют *липшицевой на  $E$* , если  $\exists L \geq 0$ , что  $Q$  окажется липшицевой на  $E$  с этим значением константы.

Пусть проведено измерение в точке  $x^i$  и, как результат, получено  $Q_i$  — значение функции в этой точке. Тогда, в силу определения липшицевых функций, для  $x \in E$  верна оценка:

$$Q_i - L\|x - x^i\| \leq Q(x) \leq Q_i + L\|x - x^i\|. \quad (4.106)$$

Пусть теперь проведены измерения на наборе точек  $x^1, x^2, \dots, x^k$ . Тогда неравенства вида (4.106) верны для всех  $i = 1, 2, \dots, k$ . Рассматривая их в совокупности, получим более точную оценку для  $x \in E$ :

$$\underbrace{\max_{i=1, \dots, k} \{Q_i - L\|x^i - x\|\}}_{Q_k^-(x)} \leq Q(x) \leq \underbrace{\min_{i=1, \dots, k} \{Q_i + L\|x^i - x\|\}}_{Q_k^+(x)}. \quad (4.107)$$

Пусть  $Q_k^*$  — наименьшее вычисленное значение функции, а  $Q_k^-$  — наименьшее на  $E$  значение миноранты:

$$Q_k^* = \min_{i=1, \dots, k} Q_i,$$

$$Q_k^- = \min_{x \in E} Q_k^-(x).$$

Справедливы следующие свойства.

### Оценки глобального экстремума

**Свойство 1.** Оценка по значению функции:

$$Q_k^- \leq Q(x^*) \leq Q_k^*. \quad (4.108)$$

**Свойство 2.** Оценка по координате:

$$x^* \in D_k = \{x \in E : Q_k^-(x) \leq Q_k^*\}. \quad (4.109)$$

Поведение верхних и нижних оценок  $Q_k^+(x)$  и  $Q_k^-(x)$  показано на рис. 4.36 на примере одномерной задачи. Здесь  $E = [a, b]$ , количество измерений функции  $k = 3$ .

Оценки значения  $Q(x^*)$ , а также возможного положения  $x^*$ , отмечены на рис. 4.36 штриховкой. Заметим, что поскольку  $x^*$  может содержаться только в подмножестве  $D_k$ , то новые измерения функции  $Q(x)$  следует проводить только в пределах этого подмножества.

Одним из методов поиска глобального минимума липшицевой функции, заданной на компакте, является метод С. А. Пиявского [2], [10], [19], [31]. Он размещает новые испытания в точке глобального минимума текущей нижней оценки  $Q_k^-(x)$ . Заметим, что эти точки всегда принадлежат множеству  $D_k$  из (4.109).

#### 4.7.1. Метод С. А. Пиявского

Метод прост в описании и имеет достаточно «мягкие» условия применимости. Геометрически идея метода заключается в следующем: предлагается

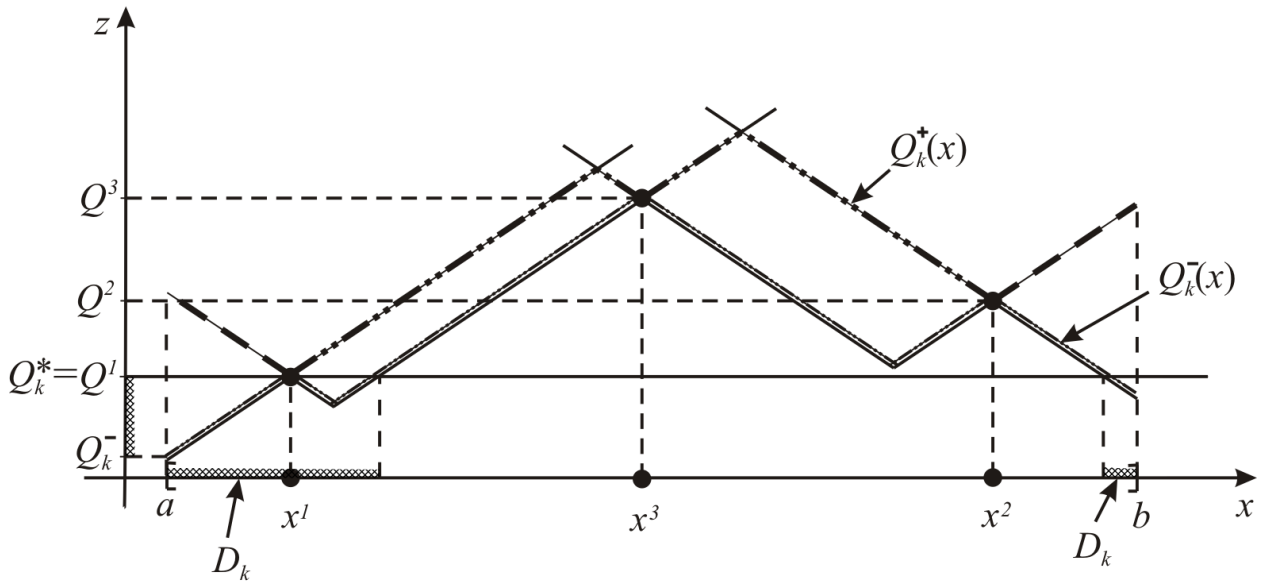


Рис. 4.36. Оценки значений липшицевой функции  $Q(x)$  и её глобального минимума при количестве испытаний  $k = 3$

проводить очередное испытание в точке минимума миноранты  $Q_k^-(x)$ . Можно доказать, что в случае  $R^1$  это правило реализует принцип одношаговой оптимальности при специальном образом введенной функции эффективности.

*Алгоритм метода С.А.Пиявского*

**Шаг 0.** Задается константа Липшица  $L > 0$  и точность по значению функции  $\varepsilon > 0$ .

**Шаг 1.** В  $E$  произвольно размещается набор начальных точек  $x^1, x^2, \dots, x^{k_0}$ . В них проводятся измерения  $Q_1, Q_2, \dots, Q_{k_0}$ , полагается  $k = k_0$ .

**Шаг 2.** Вычисляется

$$x^{k+1} = \arg \min_{x \in E} Q_k^-(x). \quad (4.110)$$

**Шаг 3.** Определяется

$$Q_k^* = \min_{i=1, \dots, k} Q_i. \quad (4.111)$$

**Шаг 4.** Проверяется критерий останова:

$$Q_k^* - Q_k^-(x^{k+1}) \leq \varepsilon. \quad (4.112)$$

Если он выполнен, то поиск останавливается,  $Q_k^*$  и соответствующая этому значению точка испытания  $x_*^k$  выдаются в качестве оценки  $Q^*$  и  $x^*$ , иначе выполняется переход на следующий шаг 5.

**Шаг 5.** Вычисляется значение функции в точке  $x^{k+1}$ :  $Q_{k+1} = Q(x^{k+1})$ , полагается  $k := k + 1$  и производится переход на шаг 2.

Работу алгоритма в случае  $R^1$  иллюстрирует рис. 4.37. Останов процесса поиска происходит после того, как длина интервала  $[Q_k^-, Q_k^*]$  становится меньше  $\varepsilon$ .

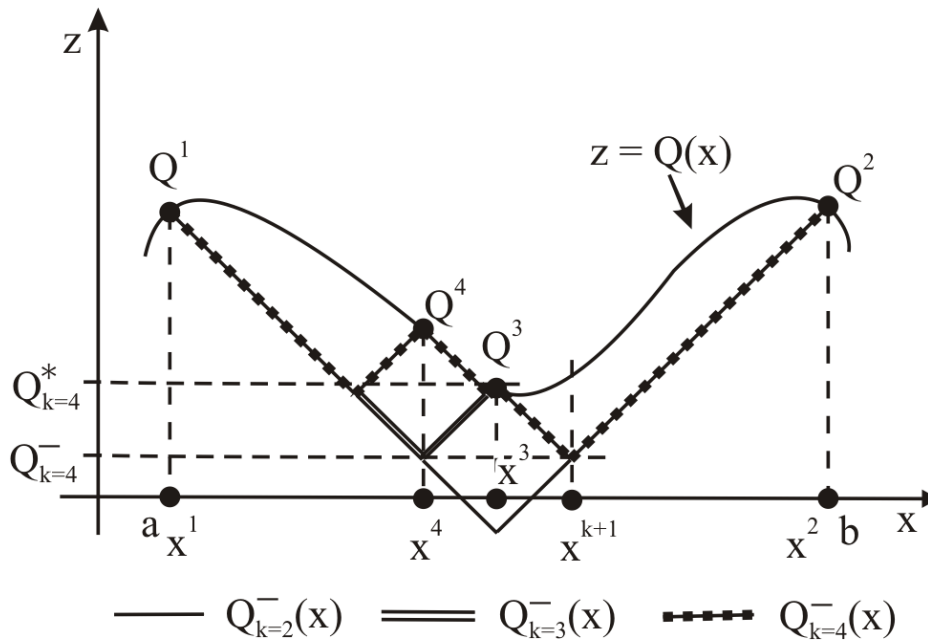


Рис. 4.37. Размещение первых пяти испытаний в методе Пиявского при  $N = 1$ ; измерение функции в точке  $x^{k+1}$  пятого испытания еще не проведено

Интуитивно понятно, что в областях, не содержащих окрестностей глобальных минимумов, метод разместит не более, чем конечное число испытаний. Таким образом, процесс будет сходиться в окрестность решения.

**Теорема.** Пусть множество  $E$ , на котором проводится поиск, является компактом, а целевая функция  $Q$  — липшицева на  $E$  с константой  $L > 0$ . Тогда для любого набора начальных точек  $x^1, \dots, x^{k_0} \in E$  метод Пиявского остановится за конечное число итераций и при останове выполнится:

$$0 \leq Q_k^* - Q(x^*) \leq \varepsilon,$$

т.е. задача будет решена с заданной точностью по значению функции.

*Доказательство.* Из свойства (4.108) вытекает, что:

$$0 \leq Q_k^* - Q(x^*) \leq Q_k^* - Q_k^-(x^{k+1}).$$

Значит, если удастся показать, что  $Q_k^* - Q_k^-(x^{k+1}) \rightarrow 0$  при  $k \rightarrow \infty$ , то для всякого  $\varepsilon > 0$  на некотором шаге  $k$  выполнится критерий останова (4.112), а следовательно, и оценка из утверждения теоремы. Таким образом, теорема будет доказана.

Докажем существование этого предела. Предположим, что предела не существует. Тогда найдется  $\varepsilon > 0$ , а также бесконечная подпоследовательность  $k = k_s$ , что для всех  $s = 1, 2, \dots$  окажется  $Q_{k_s}^* - Q_{k_s}^-(x^{k_s+1}) > \varepsilon$ . Тогда для всякого  $j \leq k_s$ :

$$\varepsilon < Q_{k_s}^* - Q_{k_s}^-(x^{k_s+1}) \leq Q_j - Q_{k_s}^-(x^{k_s+1}) = Q_{k_s}^-(x^j) - Q_{k_s}^-(x^{k_s+1}) \leq L \|x^j - x^{k_s+1}\|.$$

Последнее неравенство следует из того, что миноранта  $Q_k^-(x)$  сама является липшицевой с константой  $L$ . Таким образом, мы получили, что для любого  $j = 1, \dots, k_s$ :  $\|x^j - x^{k_s+1}\| \geq \varepsilon/L$ , и это справедливо для  $s = 1, 2, \dots$

Рассмотрим последовательность  $\{x^{k_s+1}\}_{s=1}^{\infty}$  точек из  $E$ , которая по свойству компакта должна иметь сходящуюся подпоследовательность. Но в силу последнего неравенства у этой последовательности не может быть точек сгущения, поскольку все её элементы отдалены друг от друга на расстояние не меньше, чем  $\varepsilon/L$ . Это противоречие доказывает теорему.  $\square$

### Применение метода при неизвестном $L$ (метод с оцениванием константы)

Оценим константу Липшица снизу величиной

$$l_k = \max_{1 \leq i < j \leq k} \frac{|Q_j - Q_i|}{\|x^j - x^i\|}. \quad (4.113)$$

Примем в качестве константы Липшица в выражении для миноранты из (4.107) следующую оценку:

$$L_k = \begin{cases} rl_k, & \text{если } l_k > 0; \\ 1, & \text{если } l_k = 0, \end{cases} \quad (4.114)$$

где параметр надежности  $r > 1$ .

**Замечание.** Для метода с оценкой константы Липшица вида (4.114) приведенная выше теорема останется справедлива в той своей части, которая утверждает конечность выполнения метода. Оценка точности полученного приближения будет верна только при дополнительном предположении, что оценка  $L_k$  на момент останова окажется не меньше значения  $L$ .

### Реализация метода С. А. Пиявского для $R^1$ . Метод ломаных

Рассмотрим задачу поиска глобального минимума липшицевой с константой  $L$  функции скалярного аргумента  $x$ , изменяющегося на отрезке

$$\min_{x \in [a, b] \subset R^1} Q(x). \quad (4.115)$$

Пусть точки измерений упорядочены по возрастанию координаты нижним индексом, причем количество начальных испытаний  $k_0 = 2$  и они размещены



на концах отрезка, т.е.  $x^1 = a, x^2 = b$ . Таким образом, после проведения  $k$  измерений будем иметь разбиение отрезка  $[a, b]$  на  $(k - 1)$  подинтервал точками  $a = x_1 < x_2 < \dots < x_k = b$ .

Введем обозначения:

$$\Delta Q_i = Q_{i+1} - Q_i,$$

$$\Delta x_i = x_{i+1} - x_i.$$

Нетрудно видеть, что на каждом из подинтервалов  $[x_i, x_{i+1}]$  миноранта  $Q_k^-(x)$  из (4.107) функции  $Q$ , используемая в (4.110) на шаге 2 метода С. А. Пиявского, будет иметь форму «зубца», вид которого показан на рис. 4.38. Из картинки очевидно, что при  $x \in [x_i, x_{i+1}]$  миноранта имеет единственный локальный минимум в точке  $\hat{x}_i$  со значением  $R_i$ . Величину  $R_i$  называют *характеристикой интервала*.

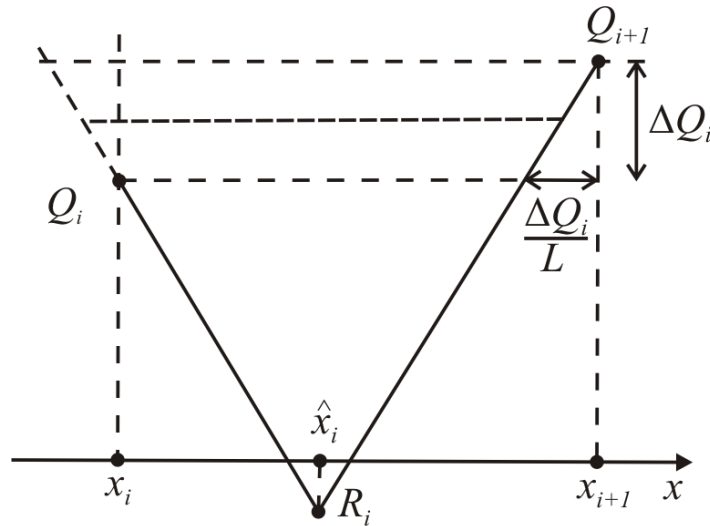


Рис. 4.38. Вид миноранты на подинтервале  $[x_i, x_{i+1}]$  и его характеристика  $R_i$

Из рис. 4.38 очевидно, что

$$\hat{x}_i = \frac{x_i + x_{i+1}}{2} - \frac{\Delta Q_i}{2L}, \quad (4.116)$$

$$R_i = \frac{Q_i + Q_{i+1}}{2} - L \frac{\Delta x_i}{2}. \quad (4.117)$$

Используя характеристики  $R_i$  можно компактно записать полученный алгоритм.

*Алгоритм метода ломаных*

**Шаг 0.** Задать константу  $L > 0$  и  $\varepsilon > 0$  — точность решения задачи по значению функции.

**Шаг 1.** Положить  $x^1 = a, x^2 = b$ , вычислить  $Q^1 = Q(x^1), Q^2 = Q(x^2), Q_2^* = \min \{Q^1; Q^2\}, x_2^* = \arg \min \{Q(x^1); Q(x^2)\}, k = 2$ .

**Шаг 2.** Упорядочить точки нижним индексом по возрастанию координаты

$$a = x_1 < \dots < x_k = b.$$

**Шаг 3.** Определить подынтервал  $[x_t; x_{t+1}]$  с минимальным значением характеристики (4.117):

$$R_t = \min \{R_i : i = 1, \dots, k - 1\}. \quad (4.118)$$

**Шаг 4.** Если выполнен критерий останова

$$Q_k^* - R_t \leq \varepsilon, \quad (4.119)$$

прекратить вычисления и принять  $Q_k^*$  и соответствующую точку  $x_k^*$  за оценку решения. Если же (4.119) не выполнено, перейти на шаг 5.

**Шаг 5.** Провести новое измерение в точке

$$x^{k+1} = \hat{x}_t, \quad Q^{k+1} = Q(x^{k+1}). \quad (4.120)$$

Положить

$$Q_{k+1}^* = \min \{Q_k^*; Q^{k+1}\}, \quad x_{k+1}^* = \arg \min \{Q(x^i) : i = 1, \dots, k + 1\},$$

принять  $k := k + 1$ , перейти на шаг 2.

Заметим, что данное правило вычислительно легко реализуемо. На практике обычно в алгоритм встраивают процедуру оценивания константы  $L$ , например, используя (4.113)–(4.114).

**Определение.** Методы одномерного поиска, в которых выбор и проведение очередного испытания, выполняется по правилам (4.118), (4.120) называют *характеристически представимыми*.

#### 4.7.2. Информационно–статистический метод Р. Г. Стронгина

Существуют другие характеристически представимые методы многоэкстремальной оптимизации для класса липшицевых функций. Одним из наиболее эффективных является информационно–статистический метод Р. Г. Стронгина [28]. Он использует первоначально предложенную Ю. И. Неймарком идею о введении вероятностного описания на классах многоэкстремальных функций. Опубликованная в 1963 году в их совместной статье [27] первоначальная концепция синтеза методов глобальной оптимизации исходила из принципа извлечения на каждой итерации наибольшего количества информации (по К. Шеннону) о положении глобального минимума целевой функции, при этом для вычисления энтропии предполагалось использовать апостериорные вероятностные распределения на классе минимизируемых функций.

Информационно–статистический метод был получен как одношагово–оптимальное в среднем решающее правило в рамках специально предложенной Р. Г. Стронгиным вероятностной модели поведения многоэкстремальной липшицевой функции. При этом при выводе правил метода использовались более удобные функции эффективности, отличные от количества извлекаемой информации по Шеннону. Процесс получения соотношений в этом методе является сложным и здесь не приводится (краткое изложение см. в [31]). В то же время, алгоритмическое описание информационно–статистического метода мало отличается от описания метода ломаных. Укажем на эти отличия.

### Отличия информационно–статистического алгоритма от метода ломаных

Пусть, как и прежде, точки измерений пронумерованы нижним индексом в порядке возрастания координаты, и два первых измерения выполнены на концах отрезка, т.е. в точках  $a$  и  $b$ . Остановимся на отличиях.

- 1) На шаге 3 при использовании правила выбора делимого интервала (4.118) применяется другое выражение для характеристики подинтервала. А именно, вместо (4.117) используется новая характеристика (4.121)

$$R_i = \frac{Q_i + Q_{i+1}}{2} - \frac{1 + ((\Delta Q_i / \Delta x_i) / L_k)^2}{2} L_k \frac{\Delta x_i}{2}. \quad (4.121)$$

При этом на шаге 5 в информационно–статистическом алгоритме применяется выражение для точки  $\hat{x}_t$  такое же, как в методе ломаных, т.е.

$$\hat{x}_i = \frac{x_i + x_{i+1}}{2} - \frac{\Delta Q_i}{2L_k}.$$

- 2) Метод всегда использует в качестве  $L$  оценку  $L_k$ , вычисляемую перед шагом 3 по правилу (4.114), где принимается

$$l_k = \max \{ |\Delta Q_i| / \Delta x_i : i = 1, \dots, k - 1 \}.$$

- 3) На шаге 4 применяется другой критерий останова — по малости делимого интервала:

$$x_{t+1} - x_t \leq \varepsilon.$$

Известно, что данный метод размещает точки испытаний, в среднем, более экономично, чем метод ломаных. Это связано с тем, что за счет наличия дополнительного множителя в виде дроби перед оценкой константы Липшица  $L_k$  в формуле (4.121), на разных подинтервалах метод использует различные значения оценок константы Липшица. А именно, на подинтервалах, где величина  $|\Delta Q_i / \Delta x_i|$  относительно мала, значение оценки константы может уменьшаться почти вдвое.

### 4.7.3. Многомерные методы липшицевой оптимизации: основные подходы

Заметим, что описанный в п. 4.7.1 метод С. А. Пияковского в своей исходной форме является многомерным. Однако, вид основного решающего правила (4.110) для выбора точек следующих испытаний таков, что не позволяет построить алгоритм, позволяющий точно реализовать это правило в многомерном случае. Это связано с тем, что миноранты  $Q_k^-(x)$  в (4.110) сами являются сложными многоэкстремальными функциями, и в (4.110) на каждой итерации нужно решать задачу глобальной оптимизации. В то же время, в  $R^1$  метод легко алгоритмизируется.

Информационно–статический метод изначально построен как одномерный.

Возникает естественный вопрос: «Как обобщить описанные подходы на многомерные задачи, чтобы построенные методы были численно эффективно реализуемы?»

Кратко ответим на этот вопрос. Используется несколько подходов.

- 1) Приближенная реализация многомерных правил (например, вида (4.110)) выбора точки нового испытания. Здесь могут использоваться простые методы типа Монте–Карло с локальным уточнением (см. п.4.7.7). Этот способ применяется редко.
- 2) Методы редукции размерности (т.е. понижения размерности) в нескольких вариантах реализации:
  - а) использование фрактальных кривых — разверток [28], [30] (также см. краткое изложение в [2]);
  - б) многошаговая схема редукции ([28], а также [2]);
- 3) Компонентные методы [2], [29]. Эти методы, в определенном смысле, также приводят к редукции размерности.

Остановимся на методах редукции размерности в рамках обзорного изложения подхода 2.

#### Редукция размерности с использованием кривых Пеано

Известно, что отрезок вещественной оси, например  $[0, 1]$  можно однозначно и непрерывно отобразить на многомерный гиперинтервал  $E$  из (4.104) таким образом, что каждая точка  $x$  из  $E$  будет иметь прообраз в  $[0, 1]$ . Отображения такого типа называют *кривыми Пеано* или *развертками Пеано*. Пусть  $x = \varphi(s)$ ,  $s \in [0, 1]$ , есть кривая Пеано, отображающая  $[0, 1]$  на  $E$ :

$$\{\varphi(s) : s \in [0, 1]\} = E = [0, 1]^N \subset R^N. \quad (4.122)$$

То есть,  $\forall x \in E \exists s \in [0, 1]$ , что  $\varphi(s) = x$ .

Тогда из непрерывности функций  $\varphi$  и  $Q$  вытекает существование минимума на  $[0, 1]$  одномерной функции  $\tilde{Q}(s) = Q(\varphi(s))$ , причем из (4.122) следует

$$\min_{s \in [0,1]} \tilde{Q}(s) = \min_{x \in E} Q(x) \quad (4.123)$$

Таким образом, решение задачи одномерной глобальной оптимизации на отрезке, будет определять решение исходной задачи на  $E \subset R^N$ .

К сожалению, точное построение кривых Пеано невозможно. Используются их приближения, например, по правилу Гильберта. Процесс получения приближений  $\varphi_m(s)$  для различного уровня  $m$  кривой Пеано–Гильберта иллюстрирует рис. 4.39

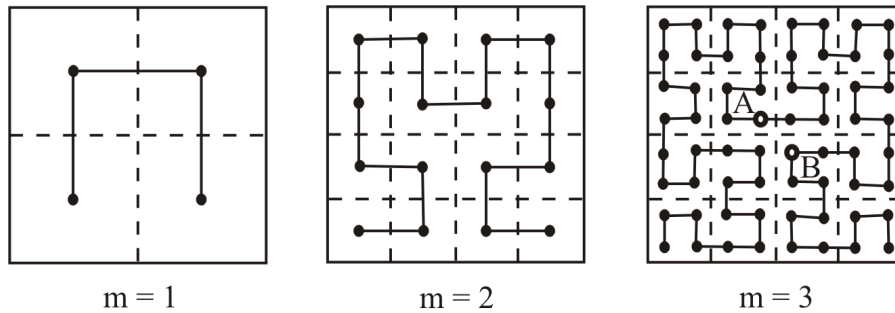


Рис. 4.39. Первые три уровня приближения в  $R^2$  кривыми Пеано-Гильберта  $\varphi_m(s)$  кривой Пеано  $\varphi(s)$

Использование разверток в многомерной глобальной оптимизации наиболее полно представлено в работах Р. Г. Стронгина и его учеников [28], [30]. Показано, что при липшицевости функций  $Q(x)$  с константой  $L$  на  $E = [0, 1]^N \subset R^N$  одномерная функция  $\tilde{Q}(s)$  из (4.123) будет удовлетворять на  $[0, 1]$  условию Гёльдера с константой  $K$ , т.е.  $\forall s', s''$  из  $[0, 1]$

$$|\tilde{Q}(s') - \tilde{Q}(s'')| \leq K \sqrt[N]{|s' - s''|}, \quad (4.124)$$

где  $K = 4 \cdot L \cdot \sqrt{N}$ .

Таким образом, с учетом свойства (4.124), для решения одномерных задач

$$\min_{s \in [0,1]} Q(\varphi_m(s))$$

следует применять не исходный информационно–статистический алгоритм или метод ломанных, а их обобщения на случай гёльдеровых функций.

Следует ещё обратить внимание на нежелательный эффект, связанный с использованием разверток. Можно заметить, что некоторые точки, например,  $A$  и  $B$  на рис. 4.38, близкие в  $R^N$ , будут иметь прообразы  $s_A, s_B$ , далёкие на отрезке  $[0, 1]$ . Для уменьшения фактора потери информации о близости

в многомерном пространстве используют одновременно несколько разверток  $\varphi_m^1, \dots, \varphi_m^p$ , получаемых, например, вращениями базовой развертки  $\varphi_m$  вокруг координатных осей на углы, кратные  $\pi/2$ . В результате одновременно решается не одна задача, а серия взаимосвязанных задач

$$\min_{s \in [0,1]} Q(\varphi_m^1(s)); \dots; \min_{s \in [0,1]} Q(\varphi_m^p(s)), \quad (4.125)$$

в которых испытание функции  $\tilde{Q}^j(s) = Q(\varphi_m^j(s))$  в точке  $s_j^{k+1}$  на  $k+1$ -м шаге в одной из задач приводит к определению значения  $Q_{k+1}$  функции  $Q(x)$  в точке  $x^{k+1} = \varphi_m^j(s_j^{k+1}) \in E$ . Однако точка  $x^{k+1}$  будет иметь прообразы  $s_1^{k+1}, \dots, s_p^{k+1}$  в каждой из  $p$  задач (4.125). Таким образом, одно проведенное измерение  $Q_{k+1}$  в  $E$  породит результаты измерений  $\tilde{Q}_{k+1}^i = Q(\varphi_m^i(s_i^{k+1})) = Q_{k+1}$  в каждой из подзадач ( $i = 1, \dots, p$ ) в точках  $s_i^{k+1}$ , размещаемых в копиях отрезка  $[0, 1]$ . Как правило, задачи из группы (4.125) решаются параллельно с обменов информацией через общую память, хранящую результаты испытаний исходной задачи.

### Многошаговая схема редукции размерности

Указанный способ понижения размерности основан на следующем общем соотношении (см., например, [2], [28]), в котором нижним индексом пронумерованы координаты вектора  $x$ :

$$\min_{x \in [a,b] \subseteq R^N} Q(x) = \min_{x_1 \in [a_1, b_1]} f_1(x_1), \quad (4.126)$$

где  $x = (x_1, \dots, x_N)$ , и

$$f_1(x_1) = \min_{x_2 \in [a_2, b_2]} f_2(x_1, x_2), \quad (4.127)$$

$$f_i(x_1, \dots, x_i) = \min_{x_{i+1} \in [a_{i+1}, b_{i+1}]} f_{i+1}(x_1, \dots, x_i, x_{i+1}) \quad (4.128)$$

для  $i = 2, \dots, N-1$ . Причем  $f_N(x_1, \dots, x_N) = Q(x)$ .

В монографии Р. Г. Стронгина 1978 года [28] доказано, что если функция  $Q(x)$  липшицева на  $[a, b] \in R^N$  с константой  $L$ , то каждая из функций  $f_i(x_1, \dots, x_i)$  липшицева на  $[a_i, b_i]$  с той же константой по переменной  $x_i$ . Таким образом, исходная липшицева многомерная задача сводится к решению  $N$  рекурсивно вложенных одномерных подзадач, к которым применяются методы одномерной многоэкстремальной липшицевой оптимизации.

### Компонентные методы

Идея компонентных методов состоит в том, что всю исходную область поиска  $E$  можно разбить на непересекающиеся по внутренним точкам подобласти-компоненты  $P_i$  определенной структуры (рис. 4.40).

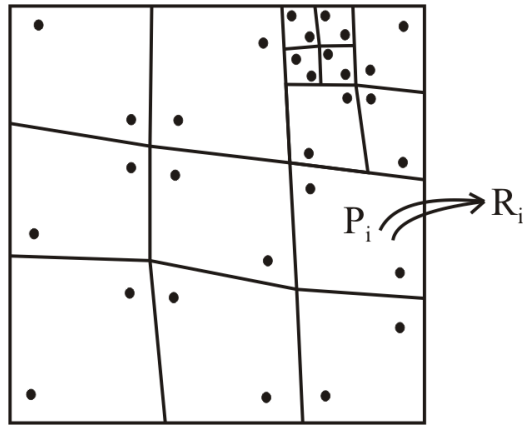


Рис. 4.40. Последовательное разделение области поиска: на шаге делится компонента  $P_t$ , имеющая наилучшую характеристику  $R_t$

В каждой компоненте испытания проводятся в строго определенных точках. Приоритетность компонент  $P_i$  при их выборе для следующего деления определяется значением *характеристики*  $R_i = R(P_i)$  данной компоненты. Характеристика  $R_i$  вычисляется с использованием результатов испытаний, проведенных только в компоненте  $P_i$ , а также с использованием построенных оценок константы Липшица.

Далее на каждой итерации определяется компонента с наилучшим приоритетом

$$R_t = \min_{i=1, \dots, n_k} R_i, \quad (4.129)$$

где  $n_k$  — число компоненты на шаге  $k$ .

После этого соответствующая компонента  $P_i$  по строго определенным правилам разделяется на более мелкие части с проведением в них новых испытаний. При этом постоянно поддерживается определенный тип разбиения. Поскольку правило (4.129) сводится к линейному перебору компонент, то этот подход также можно отнести к методам редукции.

В основе любого компонентного метода лежит выбор специальной формы компонент  $P_i$ , правил их деления, а также правил размещения в них испытаний. При этом преследуется возможность быстрого (желательно — аналитического) вычисления характеристик  $R_i$  по результатам испытаний, проведенных в компоненте. Как правило, в качестве  $R_i$  принимают простые оценки величины

$$Q_{P_i}^- = \min_{x \in P_i} Q_{P_i}^-(x),$$

где под  $Q_{P_i}^-(x)$  понимается поточечная миноранта значения функции  $Q(x)$ , построенная по результатам испытаний, размещенных в компоненте  $P_i$ .

В разных компонентных методах применяются компоненты различного вида с различными правилами размещения в них испытаний и различными правилами деления. Наиболее просты методы, использующие компоненты—

параллелепипеды в  $R^N$ . Испытания проводят либо только в геометрическом центре (так называемая *центральная схема*) либо в двух точках на концах главной диагонали (*диагональная схема*). В лекционных курсах рассмотрена только центральная схема, первоначально предложенная Ю. Г. Евтушенко в методе деления на два и методе деления на три.

#### 4.7.4. Метод деления на три

Рассмотрим многомерный параллелепипед (гиперинтервал)  $P_y$  с геометрическим центром в точке  $y$  и ребрами, параллельными осям системы координат. Введем обозначение:  $Q_y = Q(y)$ . Очевидно:

$$\forall x \in P_y : Q(x) \geq Q_{P_y}^-(x) = Q_y - L\|x - y\|,$$

где  $Q_{P_y}^-(x)$  — *поточечная нижняя оценка* значений функции  $Q$ . Тогда нижняя оценка функции  $Q$  на всем  $P_y$  определяется согласно соотношению

$$Q_{P_y}^- = \min_{x \in P_y} Q_{P_y}^-(x). \quad (4.130)$$

Минимум поточечной оценки в (4.130) достигается на  $P_y$  в наиболее удаленной от геометрического центра точке. Для обычно используемых норм — это одна из вершин  $P_y$ . Расстояние от неё до  $y$  равно половине диаметра  $P_y$ . Таким образом, получаем следующую оценку (см. рис. 4.41), принимаемую в качестве характеристики  $R(P_y)$  гиперинтервала  $P_y$ :

$$R(P_y) = Q_{P_y}^- = Q_y - L \cdot \frac{\text{diam}(P_y)}{2}. \quad (4.131)$$

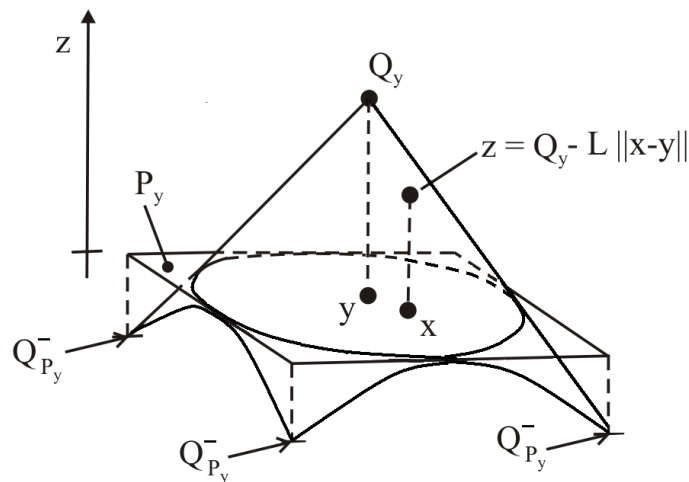


Рис. 4.41. Поточечная нижняя оценка  $Q_{P_y}^-(x)$  функции  $Q(x)$  и общая нижняя оценка  $Q_{P_y}^-$  по параллелепипеду  $P_y$  в целом



Метод деления на три применяется к задачам вида (4.104), (4.105), где  $Q$  — липшицева на  $E$  с константой  $L > 0$ . Он использует деление  $E$  на многомерные параллелепипеды  $P_w$ . Испытания, состоящие в вычислении целевой функции  $Q$ , проводятся в их геометрических центрах. Нижние оценки для  $Q(x)$  при  $x \in P_w$ , принимаемые в качестве характеристик  $R(P_w)$ , строятся только по результату измерения  $Q$  в точке  $w$  — геометрическом центре  $P_w$ , тогда как в методе Пиявского нижняя оценка для любых  $x$  строится с учетом всех проведенных испытаний. За счет этого упрощения (огрубления) нижней оценки метод деления на три легко реализуем для любой размерности  $N$ .

*Алгоритм метода деления на три*

**Шаг 0.** Задать константу Липшица  $L > 0$ , параметр останова  $\varepsilon > 0$ .

**Шаг 1.** Разделить исходный параллелепипед  $E$  на три равные части по большему ребру с получением параллелепипедов  $P_x, P_y, P_z$  с геометрическими центрами в точках  $x, y, z$ . Подсчитать значения функции в этих точках:  $Q_x, Q_y, Q_z$ . Вычислить нижние оценки  $Q_{P_x}^-, Q_{P_y}^-, Q_{P_z}^-$ , используя формулу (4.131). Сформировать начальный список компонент

$$\mathfrak{P} = \{(P_w, Q_w, Q_{P_w}^-) : w \in \{x, y, z\}\}.$$

**Шаг 2.** Найти компоненту  $P_{w^*}$  с наименьшим значением нижней оценки — характеристики:

$$R(P_{w^*}) = Q_{P_{w^*}}^- = \min \{Q_{P_w}^- : (P_w, Q_w, Q_{P_w}^-) \in \mathfrak{P}\}.$$

Определить наименьшее из вычисленных значений функции  $Q$ :

$$\tilde{Q}^* = \min \{Q_w : (P_w, Q_w, Q_{P_w}^-) \in \mathfrak{P}\}.$$

**Шаг 3.** Если  $\tilde{Q}^* - Q_{P_{w^*}}^- \leq \varepsilon$ , выполнить останов метода, приняв  $\tilde{Q}^*$  и соответствующую точку  $\tilde{x}^*$  за оценку глобального минимума, иначе — перейти на следующий шаг.

**Шаг 4.** Параллелепипед  $P_{w^*}$  разделить по большему ребру на три равные части  $\tilde{P}_{x'}, \tilde{P}_{y'}, \tilde{P}_{z'}$ , где  $y' = w^*$ . Вычислить  $Q_{x'}, Q_{z'}$ , положить  $Q_{y'} = Q_{w^*}$ . Вычислить нижние оценки  $Q_{P_w}^-$  для  $P_w \in \{\tilde{P}_{x'}, \tilde{P}_{y'}, \tilde{P}_{z'}\}$ .

**Шаг 5.** Исключить из списка  $\mathfrak{P}$  разделенный параллелепипед  $P_{w^*}$  и включить вместо него три новых:

$$\mathfrak{P} = \mathfrak{P} \setminus \{(P_{w^*}, Q_{w^*}, Q_{P_{w^*}}^-)\};$$

$$\mathfrak{P} = \mathfrak{P} \cup \{(P_w, Q_w, Q_{P_w}^-) : P_w \in \{\tilde{P}_{x'}, \tilde{P}_{y'}, \tilde{P}_{z'}\}\}.$$

Перейти на Шаг 2.

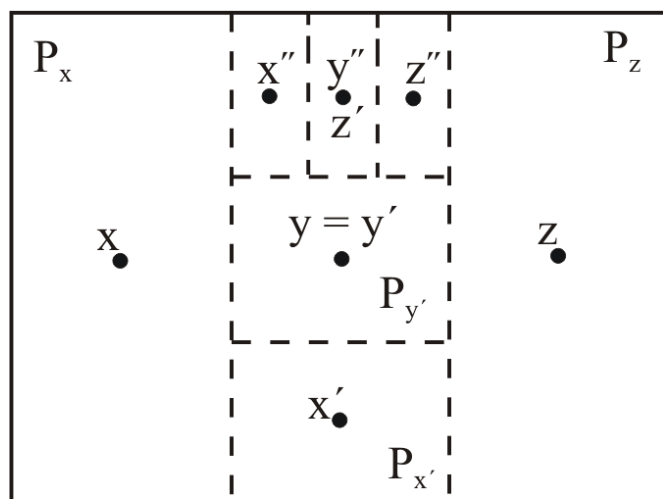


Рис. 4.42. Вариант разбиения области поиска в методе деления на три после трех итераций

Иллюстрация процесса деления компонент в данном методе приведена на рис.4.42.

Метод деления на три порождает неравномерное покрытие области поиска  $E$  точками испытаний, уплотняющееся в областях с меньшими значениями целевой функции. Для него справедлива теорема, почти полностью повторяющая теорему для метода С. А. Пиявского.

**Теорема.** Пусть функция  $Q$  — липшицева с константой  $L$  на многомерном параллелепипеде  $E$ , а параметр останова  $\varepsilon > 0$ , тогда метод деления на три остановится за конечное число шагов и в момент останова найденная оценка глобального минимума по значению функции будет отличаться от истинного значения не более, чем на  $\varepsilon$ , т.е. разность  $\tilde{Q}^* - Q(x^*) \leq \varepsilon$ .

*Доказательство* предлагается провести самостоятельно по аналогии с доказательством теоремы для метода С. А. Пиявского.

**Замечание.** Можно легко построить модификацию метода деления на три с оцениванием константы Липшица  $L$ .

#### 4.7.5. Учет ограничений–неравенств в методе деления на три

Заметим, что для методов липшицевой многоэкстремальной оптимизации нельзя применять способы учета ограничений, рассмотренные применительно к задачам локальной оптимизации. Причина проста — штрафные добавки «портят» значения констант Липшица минимизируемых функций. В многоэкстремальной оптимизации используются иные подходы (см. например, [2]).

Рассмотрим один из них на примере метода деления на три для задачи

условной многоэкстремальной оптимизации вида

$$\min_{x \in D} Q(x), \quad (4.132)$$

$$D = \{x \in E : g_1(x) \leq 0, \dots, g_m(x) \leq 0\}, \quad (4.133)$$

$$E = \{x \in R^N : a_i \leq x_i \leq b_i, i = 1, \dots, N\}, \quad (4.134)$$

где функции  $Q, g_1, \dots, g_m$  являются липшицевыми со значениями констант  $L, L_1, \dots, L_m$ .

Аналогично предыдущему, для каждого многомерного параллелепипеда  $P_y$  текущего разбиения в его геометрическом центре  $y$  проводятся измерения всех функций задачи с получением значений  $Q_y, g_y^1, \dots, g_y^m$ . Далее для всего  $P_y$  в целом строятся нижние оценки каждой из функций:

$$Q_{P_y}^- = Q_y - L \cdot \text{diam} P_y / 2, \quad (g^i)_{P_y}^- = g_y^i - L_i \cdot \text{diam} P_y / 2, \quad (i = 1, \dots, m). \quad (4.135)$$

В качестве характеристики компоненты  $P_y$  введем величину, использующую оценки (4.135),

$$R_{P_y} = \begin{cases} Q_{P_y}^-, & \text{при } (g^i)_{P_y}^- \leq 0, \forall i = 1, \dots, m; \\ +\infty, & \text{в остальных случаях.} \end{cases} \quad (4.136)$$

Эта характеристика определяет возможное наименьшее значение целевой функции в компоненте  $P_y$ , где хотя бы в некоторых точках возможно выполнение всех ограничений–неравенств. Для компонент, где это невозможно, характеристике присваивается бесконечно большое значение, что в дальнейшем сделает невозможным деление компонент, в которых ограничения не могут быть выполнены.

Изменим также оценку глобального минимума  $\tilde{Q}^*$ . Пусть  $\mathfrak{P}$  — текущий список компонент разбиения, включающий описания  $(P_w, Q_w, g_w^1, \dots, g_w^m, R_{P_w})$  для всех имеющихся компонент  $P_w$ . Тогда

$$\tilde{Q}^* = \begin{cases} +\infty, & \text{если } \forall (P_w, Q_w, g_w^1, \dots, g_w^m, R_{P_w}) \in \mathfrak{P} : \max_{i=1, \dots, m} \{g_w^i\} > 0 \\ \min \{Q_w : (P_w, Q_w, g_w^1, \dots, g_w^m, R_{P_w}) \in \mathfrak{P} \text{ и } \max_{i=1, \dots, m} \{g_w^i\} \leq 0\}, & \text{иначе.} \end{cases} \quad (4.137)$$

*Алгоритм метода деления на три для задач с ограничениями–неравенствами*

**Шаг 0.** Задать константы Липшица  $L, L_1, \dots, L_m$  и параметр останова  $\varepsilon > 0$ .

**Шаг 1.** Разбить исходный параллелепипед  $E$  на три части по большему ребру с получением параллелепипедов  $P_x, P_y, P_z$  с геометрическими центрами в точках  $x, y, z$ . Вычислить значения функций  $Q_w, g_w^1, \dots, g_w^m$  для  $w \in \{x, y, z\}$ . Определить нижние оценки вида (4.135) и характеристики (4.136) для  $P_x, P_y, P_z$ . Сформировать список описаний компонент

$$\mathfrak{P} = \{(P_w, Q_w, g_w^1, \dots, g_w^m, R_{P_w}) : w \in \{x, y, z\}\}$$

**Шаг 2.** Найти наиболее приоритетную компоненту  $P_{w^*}$  из условия

$$R_{P_{w^*}} = \min \{ R_{P_w} : (P_w, Q_w, g_1^w, \dots, g_m^w, R_{P_w}) \in \mathfrak{P} \}.$$

Вычислить оценку  $\tilde{Q}^*$  из (4.137).

**Шаг 3.** Если  $R_{P_{w^*}} = +\infty$ , остановить вычисления и выдать сообщение о пустоте допустимой области. Если  $R_{P_{w^*}} < +\infty$  и  $\tilde{Q}^* - R_{P_{w^*}} \leq \varepsilon$ , то остановить вычисления и выдать оценку решения  $\tilde{Q}^*$  и соответствующую ей точку  $\tilde{x}^*$ , иначе — перейти на шаг 4.

**Шаг 4.** Параллелепипед  $P_{w^*}$  разделить по большему ребру на три равные части  $\tilde{P}_{x'}, \tilde{P}_{y'}, \tilde{P}_{z'}$ , где  $y' = w^*$ . Вычислить  $Q_w, g_w^1, \dots, g_w^m$  для  $w \in \{x'; z'\}$ ; положить  $(Q_{y'}, g_{y'}^1, \dots, g_{y'}^m) = (Q_{w^*}, g_{w^*}^1, \dots, g_{w^*}^m)$ . Вычислить характеристики  $R_{P_w}$  для  $P_w \in \{ \tilde{P}_{x'}, \tilde{P}_{y'}, \tilde{P}_{z'} \}$

**Шаг 5.** Исключить из списка  $\mathfrak{P}$  разделенный параллелепипед  $P_{w^*}$  и включить вместо него три новых

$$\mathfrak{P} := \mathfrak{P} \setminus \{ (P_{w^*}, Q_{w^*}, g_{w^*}^1, \dots, g_{w^*}^m, R_{P_{w^*}}) \};$$

$$\mathfrak{P} := \mathfrak{P} \cup \left\{ (P_w, Q_w, g_w^1, \dots, g_w^m, R_{P_w}) : P_w \in \{ \tilde{P}_{x'}, \tilde{P}_{y'}, \tilde{P}_{z'} \} \right\}.$$

Перейти на шаг 2.

Описанный метод строит неравномерное покрытие области поиска  $E$  точками испытаний, уплотняющееся на допустимом множестве  $D$  в окрестностях глобально-оптимальных решений.

Можно построить варианты методов деления на три с оцениванием неизвестных значений констант липшица.

#### 4.7.6. Метод DIRECT с множественными значениями константы Липшица

Метод DIRECT был предложен сравнительно недавно, в 1993 году (авторы: Джонес Д. Р., Пертунен С. Д., Стукман Б. Е.) для задачи (4.104), (4.105) без функциональных ограничений. Эта работа открыла новое направление в разработке методов липшицевой оптимизации. Здесь приведем краткую информацию об основной идее метода.

При построении метода предполагается, что  $Q \in \Phi_{Lip}(E)$ , т.е. функция является липшицевой. Однако константа Липшица  $L$  предполагается неизвестной и методом не оценивается. Метод DIRECT строится в предположении, что  $L$  — может принимать любые значения из  $(0, +\infty)$ . Как же в этом случае выбирать точки новых измерений?

Метод относится к компонентным и использует компоненты-параллелепипеды  $P_w$  с испытаниями в центральных точках  $w$ . Деление выбранных приоритетных компонент происходит по схеме деления на три, как и в двух методах, описанных в п. 4.7.4–4.7.5.

Отличия от метода деления на три возникают в правиле отбора приоритетных компонент  $P_{w^*}$ . Отбор основан на выделении на каждом шаге множества так называемых *потенциально оптимальных* компонент  $P_{w^*}$ .

**Определение.** Компонента  $P_{w^*}$  называется *потенциально оптимальной*, если существует  $L > 0$  при котором

$$\forall P_w : Q_{w^*} - L \frac{\text{diam}(P_{w^*})}{2} \leq Q_w - L \frac{\text{diam}(P_w)}{2} \quad (4.138)$$

и, кроме того,

$$Q_{w^*} - L \frac{\text{diam}(P_{w^*})}{2} \leq \tilde{Q}^* - |\tilde{Q}^*| \varepsilon. \quad (4.139)$$

Здесь параметр  $\varepsilon > 0$  определяет желаемое текущее относительное улучшение достигнутого к данному шагу метода минимального значения  $\tilde{Q}^*$  целевой функции.

Чтобы правило (4.138), (4.139) стало геометрически понятным, перенумеруем все компоненты из  $\{P_w\}$  и сопоставим компоненте  $P_j$  точку с координатами  $(d_j; Q_j)$ , где  $d_j = \text{diam}(P_j)$ , а  $Q_j$  — значение  $Q$  в центре  $P_j$ .

Если множество компонент изобразить точками  $(d_j; Q_j)$  на плоскости  $d, Q$ , то искомые компоненты  $P_{w^*}$  будут соответствовать набору точек  $(d_{j^*}; Q_{j^*})$ , лежащих на право-нижнем участке границы выпуклой линейной оболочки множества  $\{(d_j; Q_j)\}$ , показанного на рис. 4.43.

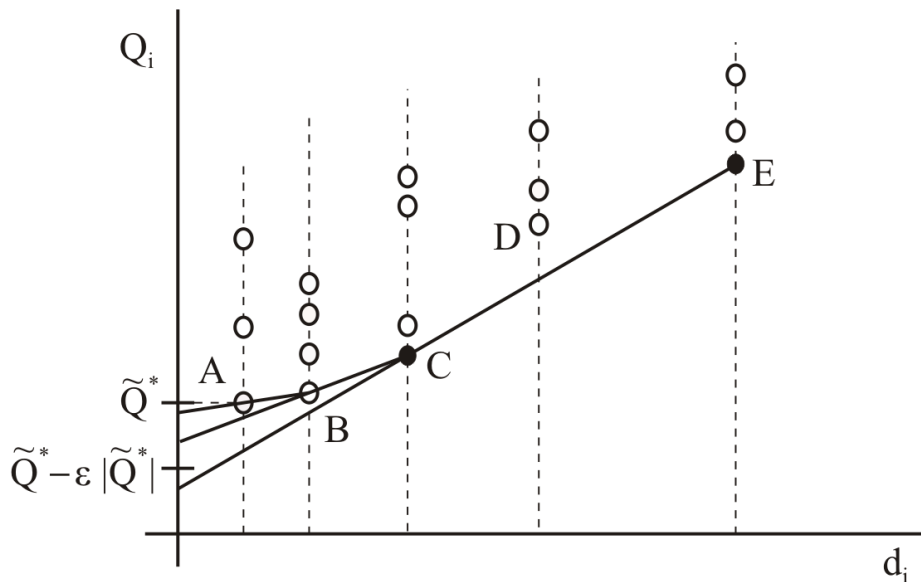


Рис. 4.43. Множество потенциально оптимальных компонент  $P_{w^*}$  изображено четными точками

Следует обратить внимание на то, что все точки сгруппированы в вертикальные ряды-столбцы, соответствующие компонентам одинакового диаметра. Право-нижнюю «границу» множества образуют точки A, B, C и E.

Очевидно, что для точек, расположенных на этой границе, первое условие (4.138) потенциальной оптимальности выполняется. Однако, на рис. 4.43 выделены не все точки этой «границы», а только точки  $C$  и  $E$ . Действительно, если для  $A$  и  $B$  при некотором  $L$  выполнено (4.138), то нарушается (4.139). Точка  $D$  — не лежит на границе выпуклой линейной оболочки множества точек  $(d_j; Q_j)$  и поэтому для нее не найдется такого  $L$ , что выполняется условия (4.138). Таким образом, в приведенной на рис. 4.43 ситуации делению на данном шаге подвергнутся два параллелепипеда  $P_{w^*}$ , изображенные жирными точками  $C$  и  $E$ .

Поскольку очевидно, что на каждом шаге в число делимых будет обязательно попадать одна из компонент с наибольшим диаметром (на рис. 4.43 это точка  $E$  — нижняя точка из самого правого столбца), то, за счет конечности числа элементов в самом правом столбце из точек-компонент, при бесконечном продолжении вычислений каждая компонента обязательно рано или поздно будет еще раз поделена. Это означает, что *любая точка области поиска будет предельной для последовательности испытаний*. Однако размещение точек будет неравномерным, что и определяет эффективность метода.

Заметим, что метод DIRECT не имеет правила останова по точности.

На рис.4.44 представлен пример размещения первых 300 испытаний функции

$$Q(x_1, x_2) = (4 + 2.1x_1^2 + x_1^4/3) \cdot x_1^2 + x_1 \cdot x_2 + 4(x_2^2 - 1)x_2^2$$

при поиске ее глобального минимума на множестве  $x_1 \in [-2.5, 2.5]$ ,  $x_2 \in [-1.5, 1.5]$  методом DIRECT с параметром  $\varepsilon = 10^{-4}$ .

#### 4.7.7. Эвристические методы глобального поиска

Методы липшицевой оптимизации, представленные в п.4.7.1–4.7.6 основаны на специальном планировании размещения испытаний. Целью планирования является достижение адаптивного поведения методов на решаемых задачах, приводящего к неравномерному размещению точек испытаний, существенно более плотному в окрестностях решений.

Однако сам процесс планирования требует вычислительных затрат, порой значительных. Это оправдано в задачах, где время  $\Delta t$  вычисления функции относительно велико.

Если же  $\Delta t$  мало, то применение сложных методов может сать нецелесообразным. В этом случае можно использовать так называемые мультистартовые эвристические методы, основанные на простых процедурах выбора точек запуска локальных методов с последующим определением лучшей найденной точки в качестве оценки глобального минимума. Выбор точек запуска может выполняться случайно. Известно достаточно много реализаций этого эвристического подхода.

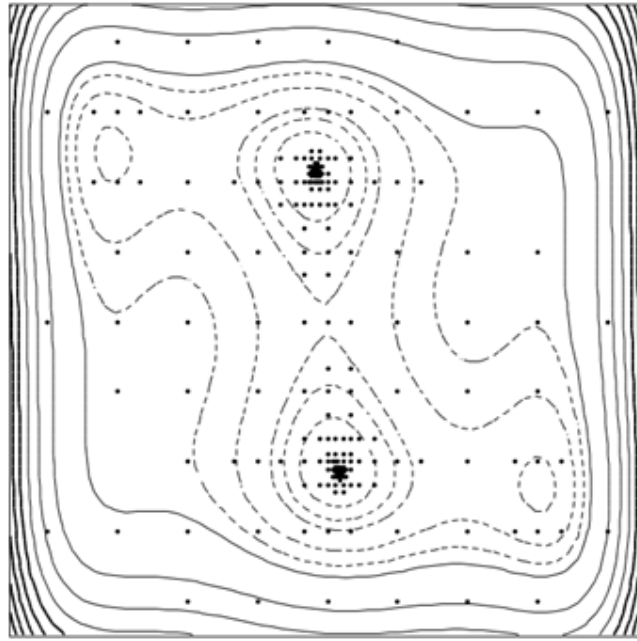


Рис. 4.44. Размещение первых 300 испытаний методом DIRECT на функции с двумя глобальными минимумами

В качестве примера проанализируем простейшую из таких процедур.

### Метод Монте–Карло с локальным уточнением

Рассмотрим задачу (4.104), (4.105) и предположим, что известна величина  $v$  — нижняя оценка меры  $V$  множества  $D^*$  точек из  $E$ , в которых функция  $Q$  принимает значение, меньшее, чем во всех локальных минимумах, не совпадающих с глобальными:

$$0 < v \leq V(D^*), D^* = \{x \in E : Q(x) \leq \min Q(x_{\text{лок}}^*), x_{\text{лок}}^* \notin \{x^*\}\}.$$

Пусть  $q^* = V(D^*)/V(E)$  — относительная мера множества  $D^*$ , а  $q$  — её нижняя оценка:  $q = v/V(E) \leq q^*$ .

Заметим, что при выборе начальной точки  $x^0 \in D^*$  метод локального спуска должен сойтись к одному из глобальных минимумов.

Определим  $P^*(n)$  вероятность попадания в  $D^*$  хотябы одной из  $n$  случайно и равномерно размещенных в  $E$  точек  $x^1, x^2, \dots, x^n$ . По формуле Бернулли имеем:

$$P^*(n) = 1 - (1 - q^*)^n \geq 1 - (1 - q)^n. \quad (4.140)$$

Если одна из точек  $x^1, x^2, \dots, x^n$  попала в  $D^*$ , то её можно определить как точку с наименьшим значением  $Q(x^i)$ , ( $i = 1, 2, \dots, n$ ). Локальный спуск из такой точки приведет нас к глобальному минимуму.

Если задать требуемую надежность поиска  $0 < \beta < 1$  и потребовать

выполнения неравенства  $P^*(n) \geq \beta$ , то из (4.140) получим:

$$n \geq n^* = \frac{\ln(1 - \beta)}{\ln(1 - q)}. \quad (4.141)$$

*Алгоритм метода*

**Шаг 0.** Задать  $0 < q < 1$  — нижнюю оценку относительной меры множества  $D^*$  в  $E$ ;  $0 < \beta < 1$  — требуемую надежность решения;  $\varepsilon > 0$  требуемую точность.

**Шаг 1.** Определить по формуле (4.141) необходимое число  $n^*$  случайно размещаемых точек.

**Шаг 2.** Случайно и равномерно разместить в  $E$   $n = n^*$  точек  $x^1, \dots, x^n$  и вычислить значения  $Q^i = Q(x^i)$  ( $i = 1, \dots, n$ ). Определить  $\tilde{x}^* = \arg \min \{Q(x^i) : i = 1, \dots, n\}$ .

**Шаг 3.** Выполнить локальный спуск с точностью  $\varepsilon$  из построенной начальной точки  $x^0 = \tilde{x}^*$ .

Приведем в качестве примера таблицу необходимого числа испытаний  $n^*$  на шаге 2 для нескольких значений оценок относительной меры  $q$  множества  $D^*$ . Во всех случаях уровень надежности, т.е. вероятности правильного определения глобального минимума, был выбран равным  $\beta = 0.999$ .

Таблица 4.2. Необходимое число измерений в методе Монте–Карло при  $\beta = 0.999$

q	0.001	0.0001	$10^{-5}$	$10^{-10}$	$10^{-15}$
$n^*$	$\approx 69 \cdot 10^2$	$\approx 69 \cdot 10^3$	$\approx 69 \cdot 10^4$	$\approx 69 \cdot 10^9$	$\approx 69 \cdot 10^{14}$

Заметим, что при фиксированных линейных размерах множеств  $D^*$  и  $E$  в многомерных задачах с ростом размерности  $N$  значение  $q$  будет экспоненциально убывать, что приведет к быстрому нарастанию числа необходимых испытаний.



## Список литературы

1. Городецкий С.Ю. Лекции по методам нелинейной оптимизации. — Нижний Новгород: Нижегородский государственный университет, 2018. — 219 с. — [Электронная версия: <http://www.itmm.unn.ru/tuds/obuchenie/materialy/>].
2. Городецкий С.Ю., Гришагин В.А. Нелинейное программирование и многоэкстремальная оптимизация. — Нижний Новгород: Изд-во ННГУ, 2007. — 489 с.
3. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. — М.: Наука, 1982. — 254 с.
4. Городецкий С.Ю. Методы оптимизации на графах с векторными весами ребер. — Нижний Новгород: Нижегородский государственный университет, 2004. — 33 с.
5. Карманов В.Г. Математическое программирование М.: Физматлит, 2008. — 263 с.
6. Арис Р. Дискретное динамическое программирование. — М.: Мир, 1969.
7. Беллман Р., Дрейфус С. Прикладные задачи динамического программирования. — М.: Наука, 1965.
8. Калихман И.Л., Войтенко М.А. Динамическое программирование в примерах и задачах. — М.: Высшая школа, 1979.
9. Кристофидес Н. Теория графов. Алгоритмический подход. — М.: Мир, 1978.
10. Методы оптимизации в примерах и задачах / Бирюков Р.С., Григорьева С.А., Городецкий С.Ю., Павлючонок З.Г., Савельев В.П. — Нижний Новгород: Нижегородский государственный университет, 2010. — 100 с. — [Электронная версия: <http://www.itmm.unn.ru/tuds/obuchenie/materialy/>].
11. Габасов Р., Кириллова Ф.М. Основы динамического программирования. — Минск: Из-во БГУ. 1975.
12. Фихтенгольц Г.М. Курс дифференциального и интегрального исчисления. Т.1. — Изд. 7-е стереотипное. — М.: Наука, 1969.
13. Немировский А.С., Юдин Д.Б. Сложность задач и эффективность методов оптимизации. — М.: Наука, 1979.
14. Измаилов А.Ф., Солодов М.В. Численные методы оптимизации. — М.: Физматлит, 2003.
15. Базара М., Шетти К. Нелинейное программирование. Теория и алгоритмы. — М.: Мир, 1982.
16. Полак Э. Численные методы оптимизации. Единый подход. — М.: Мир, 1974.

17. Мину М. Математическое программирование. — М.: Наука, 1990.
18. Пшеничный Б.Н. Выпуклый анализ и экстремальные задачи. — М.: Наука, 1980.
19. Васильев Ф.П. Численные методы решения экстремальных задач. — М.: Наука, 1982.
20. Васильев Ф.П. Методы оптимизации. — М.: Факториал Пресс, 2002.
21. Городецкий С.Ю. Лабораторный практикум по методам локальной оптимизации в программной системе LocOpt. [Электронный ресурс] — Нижний Новгород: ННГУ, 2007. URL: <http://www.unn.ru/e-library/aids.html?pscience=6posdate=2007>.
22. Сухарев А.Г., Тимохов А.В., Федоров В.В. Курс методов оптимизации. — М.: Наука, 1986.
23. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. — М.: Мир, 1985.
24. Бертсекас Д. Условная оптимизация и методы множителей Лагранжа. — М.: Радио и связь, 1987.
25. Поляк Б.Т. Введение в оптимизацию. — М.: Наука, 1983.
26. Немировский А.С., Юдин Д.Б. Сложность задач и эффективность методов оптимизации. — М.: Наука, 1979.
27. Неймарк Ю.И., Стронгин Р.Г. Информационный подход к задаче поиска экстремума функций // Известия АН СССР. Техническая кибернетика, 1966. № 1. С. 17–26.
28. Стронгин Р.Г. Численные методы в многоэкстремальных задачах (информационно-статистические алгоритмы). — М.: Наука, 1978.
29. Сергеев Я.Д., Квасов Д.Е. Диагональные методы глобальной оптимизации. — М.: Физматлит, 2008.
30. Стронгин Р.Г., Гергель В.П., Гришагин В.А., Баркалов К.А. Параллельные вычисления в задачах глобальной оптимизации / Предисл.: В.А. Садовничий. — М.: Изд-во Московского университета, 2013.
31. Городецкий С.Ю. Методы поиска глобального экстремума. — Горький: Горьковский государственный университет, 1990. — [Электронная версия: <http://www.itmm.unn.ru/tuds/obuchenie/materialy/>].

# Оглавление

<b>Предисловие</b>	<b>3</b>
<b>Глава 1. Постановки задач оптимизации, понятия решения</b>	<b>8</b>
1.1. Общий вид рассматриваемых задач, локальные и глобальные решения . . . . .	8
1.2. Другие постановки задач оптимизации . . . . .	10
1.3. Обобщения понятия решения в задачах векторной оптимизации	11
<b>Глава 2. Дискретное динамическое программирование</b>	<b>14</b>
2.1. Задачи с фиксированным временем начала и окончания процесса	14
2.2. Метод рекуррентных соотношений Р. Беллмана . . . . .	16
2.3. Принцип Беллмана . . . . .	22
2.4. Запись рекуррентных соотношений Беллмана относительно начала процесса . . . . .	25
2.5. Обобщение рекуррентных соотношений Беллмана на задачи с не фиксированной длительностью процесса . . . . .	27
<b>Глава 3. Условия оптимальности в задачах математического программирования</b>	<b>30</b>
3.1. Необходимые сведения из математического анализа . . . . .	30
3.2. Необходимые условия оптимальности в задаче с равенствами в форме теоремы Лагранжа, понятие регулярности допустимого множества . . . . .	34
3.3. Элементы теории выпуклых множеств и функций . . . . .	40
3.3.1. Выпуклые множества, проекция, теоремы об отделимости	40
3.3.2. Выпуклые функции и их свойства . . . . .	48
3.4. Условия оптимальности для задач с ограничениями–равенствами и неравенствами . . . . .	58
3.4.1. Условия оптимальности в недифференциальной форме .	58
3.4.2. Условия оптимальности в дифференциальной форме . .	66
<b>Глава 4. Численные методы решения задач математического программирования</b>	<b>81</b>
4.1. Понятие численного метода математического программирования	81
4.2. Понятие оптимального и $\varepsilon$ -оптимального метода . . . . .	83
4.3. Построение методов поиска минимума унимодальной функции на отрезке . . . . .	85
4.3.1. Построение оптимальных и $\varepsilon$ -оптимальных методов в классе пассивных $n$ -шаговых . . . . .	87
4.3.2. Методы дихотомии — неоптимальные последовательные методы . . . . .	90
4.3.3. Построение $\varepsilon$ -оптимального последовательного метода для класса унимодальных задач, метод Фибоначчи . . .	92

4.3.4.	Метод золотого сечения . . . . .	95
4.3.5.	Сравнительная таблица эффективности различных методов . . . . .	97
4.4.	Построение методов поиска минимума дифференцируемой выпуклой функции на выпуклом многограннике . . . . .	98
4.5.	Методы поиска локального минимума в задачах без функциональных ограничений . . . . .	101
4.5.1.	Критерии и алгоритмы выбора шагового множителя . . . . .	104
4.5.2.	Простые методы градиентного поиска . . . . .	110
4.5.3.	Метод Ньютона . . . . .	114
4.5.4.	Метод сопряженных градиентов Флетчера–Ривса . . . . .	119
4.5.5.	Обзор других эффективных методов локальной оптимизации первого и второго порядков . . . . .	126
4.5.6.	Методы прямого поиска . . . . .	130
4.6.	Численные методы решения задач с ограничениями . . . . .	132
4.6.1.	Метод внешнего штрафа . . . . .	133
4.6.2.	Метод модифицированных функций Лагранжа . . . . .	140
4.7.	Методы многоэкстремальной оптимизации . . . . .	147
4.7.1.	Метод С. А. Пиявского . . . . .	149
4.7.2.	Информационно–статистический метод Р. Г. Стронгина . . . . .	154
4.7.3.	Многомерные методы липшицевой оптимизации: основные подходы . . . . .	156
4.7.4.	Метод деления на три . . . . .	160
4.7.5.	Учет ограничений–неравенств в методе деления на три . . . . .	162
4.7.6.	Метод DIRECT с множественными значениями константы Липшица . . . . .	164
4.7.7.	Эвристические методы глобального поиска . . . . .	166

Станислав Юрьевич Городецкий

**ЛЕКЦИИ ПО НЕЛИНЕЙНОМУ  
МАТЕМАТИЧЕСКОМУ  
ПРОГРАММИРОВАНИЮ**

*Учебно-методическое пособие*

Федеральное государственное автономное образовательное учреждение  
высшего образования

«Национальный исследовательский Нижегородский государственный  
университет им. Н.И. Лобачевского»

603950, Нижний Новгород, пр. Гагарина, 23