

*Judson*  
Nov 1083 OKS

Б.Т.ПОЛЯК

ВВЕДЕНИЕ  
В ОПТИМИЗАЦИЮ



МОСКВА «НАУКА»  
ГЛАВНАЯ РЕДАКЦИЯ  
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ  
1983

§2.81  
П 54  
УДК 62-60

**Введение в оптимизацию.** Поляк Б. Т.—М.: Наука. Главная редакция физико-математической литературы, 1983.—384 с.

Книга является систематическим введением в современную теорию и методы оптимизации для конечномерных задач. Основное внимание уделяется идейным основам методов, их сравнительному анализу и примерам использования. Охвачен широкий круг задач — от линейного программирования, обсуждается методика постановки и решения прикладных проблем оптимизации. Приводятся условия экстремума, теоремы существования, единственности и устойчивости решения для основных классов задач. Исследуется влияние помех, неплавкости функций, вырожденности минимума. Книга предназначена для инженеров, экономистов, статистиков, вычислителей, сталкивающихся с задачами оптимизации. По своему математическому аппарату книга доступна студентам технических и экономических вузов.

Табл. 6, илл. 44, библ. 247.

**Борис Теодорович Поляк**

**ВВЕДЕНИЕ В ОПТИМИЗАЦИЮ**

Редактор Т. И. Кузнецова  
Технический редактор В. Н. Кондакова  
Корректоры О. А. Сидал, Л. С. Сомкова

ИБ № 11436

Сдано в набор 22.01.82. Подписано к печати 06.04.83. Т-08902. Формат 60×90<sup>1/6</sup>. Бумага № 2. Литография гарнитура. Высокая печать. Условия печ. л. 24. Уч.-изд. л. 26.32. Тираж 8500 экз. Заказ № 179. Цена 2 р. 40 к.

Издательство «Наука»  
Главная редакция физико-математической литературы  
117071, Москва, В-71, Ленинский проспект, 15

Ленинградская типография № 2 головное предприятие ордена Трудового Красного Знака Ленинградского объединения «Техническая книга» им. Евгении Соколовой Союза полиграфии при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52, Измайловский проспект, 29

## ОГЛАВЛЕНИЕ

Предисловие . . . . .	5
Список обозначений . . . . .	7
Введение . . . . .	9
ЧАСТЬ I. БЕЗУСЛОВНАЯ МИНИМИЗАЦИЯ . . . . .	15

**ГЛАВА 1. Основы теории и методов безусловной минимизации**

§ 1. Сведения из математического анализа . . . . .	15
§ 2. Условия экстремума . . . . .	22
§ 3. Существование, единственность, устойчивость минимума . . . . .	25
§ 4. Градиентный метод . . . . .	29
§ 5. Метод Ньютона . . . . .	36
§ 6. Роль теорем сходимости . . . . .	39

**ГЛАВА 2. Общие схемы исследования итеративных методов**

§ 1. Первый метод Ляпунова . . . . .	44
§ 2. Второй метод Ляпунова . . . . .	49
§ 3. Другие схемы . . . . .	59

**ГЛАВА 3. Методы минимизации**

§ 1. Модификации градиентного метода и метода Ньютона . . . . .	63
§ 2. Многошаговые методы . . . . .	68
§ 3. Другие методы первого порядка . . . . .	77
§ 4. Прямые методы . . . . .	87

**ГЛАВА 4. Влияние помех**

§ 1. Источники и типы помех . . . . .	94
§ 2. Градиентный метод при наличии помех . . . . .	97
§ 3. Другие методы минимизации при наличии помех . . . . .	100
§ 4. Прямые методы . . . . .	103
§ 5. Оптимальные методы при наличии помех . . . . .	107

**ГЛАВА 5. Минимизация недифференцируемых функций**

§ 1. Сведения из выпуклого анализа . . . . .	114
§ 2. Условия экстремума, существование, единственность и устойчивость решения . . . . .	124
§ 3. Субградиентный метод . . . . .	128
§ 4. Другие методы . . . . .	134
§ 5. Влияние помех . . . . .	144
§ 6. Поисковые методы . . . . .	146

**ГЛАВА 6. Вырожденность, многоэкстремальность, нестационарность**

§ 1. Вырожденный минимум . . . . .	150
§ 2. Многоэкстремальность . . . . .	166
§ 3. Нестационарность . . . . .	175

## ЧАСТЬ II. УСТОЙЧИВАЯ МИНИМИЗАЦИЯ

Г л а в а 7. Минимизация на простых множествах . . . . .	179
§ 1. Основы теории . . . . .	179
§ 2. Основные методы . . . . .	185
§ 3. Другие методы . . . . .	192
§ 4. Влияние помех . . . . .	196

Г л а в а 8. Задачи с ограничениями типа равенств . . . . .	199
§ 1. Основы теории . . . . .	199
§ 2. Методы минимизации . . . . .	210
§ 3. Учет возможных осложнений . . . . .	220

Г л а в а 9. Общая задача математического программирования . . . . .	225
§ 1. Выпуклое программирование (теория) . . . . .	225
§ 2. Нелинейное программирование (теория) . . . . .	240
§ 3. Методы выпуклого программирования . . . . .	247
§ 4. Методы нелинейного программирования . . . . .	263

Г л а в а 10. Линейное и квадратичное программирование . . . . .	268
§ 1. Линейное программирование (теория) . . . . .	268
§ 2. Конечные методы линейного программирования . . . . .	281
§ 3. Итерационные методы линейного программирования . . . . .	288
§ 4. Квадратичное программирование . . . . .	296

Часть III. ПРИКЛАДНОЙ АСПЕКТ . . . . .	298
Г л а в а 11. Примеры задач оптимизации . . . . .	301
§ 1. Задачи идентификации . . . . .	301
§ 2. Оптимизационные задачи в технике и экономике . . . . .	317
§ 3. Задачи оптимизации в математике и физике . . . . .	330

Г л а в а 12. Практическое решение задач оптимизации . . . . .	336
§ 1. Процесс решения . . . . .	336
§ 2. Программы оптимизации . . . . .	340
§ 3. Тестовые задачи и результаты вычислений . . . . .	343
Библиографические указания и комментарии . . . . .	361
Литература . . . . .	372
Предметный указатель . . . . .	383

## ПРЕДИСЛОВИЕ

Чрезвычайно широкое распространение задач оптимизации в технике, экономике, управлении привело к необходимости ознакомления широкого круга практиков с методами решения подобных задач. Однако инженеру или вычислителю трудно ориентироваться в литературе по оптимизации (большинство имеющихся книг написано «математиками для математиков»), несложно разобраться в многообразии задач и алгоритмов. В этой книге делается попытка систематического изложения основ современной теории и методов оптимизации в форме доступной инженеру. Используемый математический аппарат минималистичен — достаточно знания начал математического анализа, линейной алгебры и теории вероятностей. Изложение построено на последовательном усложнении рассматриваемых задач. Вначале описываются наиболее простые задачи безусловной минимизации гладких функций, затем исследуется влияние различных осложняющих факторов — помех, негладкости функций, вырожденности минимума, наличия ограничений. Анализ каждого класса задач проводится единообразно — вводится требуемый математический аппарат, затем обосновываются условия экстремума, результаты о существовании, единственности и устойчивости решения, и, наконец, описываются основные методы решения и исследуются их свойства. Главное внимание уделяется идейным основам методов, их сравнительному анализу; показано, как теоретические результаты служат фундаментом при построении и изучении методов. На примерах практической задач оптимизации обсуждается взаимоотношение общих и специальных методов решения. Данна обширная комментированная библиография, позволяющая читателю в случае необходимости обратиться к более подробным работам на интересующую его тему.

Включенный в книгу материал во многом отличается от традиционного. Нередко учебники по математическому программированию сводятся к описанию техники симплекс-метода линейного программирования. Мне этот круг вопросов не кажется центральным; ему посвящен лишь один параграф. В то же время большое внимание удалено задаче безусловной минимизации, которая дает богатый материал для обсуждения основных идей теории и методов оптимизации. Среди нестандартных разделов книги — задачи негладкой оптимизации, вырожденные

и нестационарные задачи, задачи с ограничениями типа равенств, условия устойчивости экстремума, влияние помех на методы оптимизации, анализ общих схем исследования сходимости итеративных методов и т. д. Систематически обсуждаются «наивные» вопросы, которые не принято задавать в математической литературе. Например: зачем нужны Условия экстремума? Какова польза от теоретических результатов о сходимости методов? Можно ли реально решать неустойчивые задачи оптимизации? и т. п.

Книга целиком посвящена конечномерным задачам. Это обусловлено как ограничениями на объем работы, так и предполагаемым уровнем математических знаний. Поэтому не рассматриваются такие важнейшие вопросы, как современная теория условий оптимальности в общих экстремальных задачах, задачи вариационного исчисления и оптимального управления и т. д. Вместе с тем мне кажется, что конечномерный случай очень богат идеями и результатами; он может служить прекрасной «моделью» более общих задач оптимизации. Знакомый с функциональным анализом читатель без труда заметит, что многие утверждения автоматически переносятся на задачи в гильбертовом или банаховом пространстве, однако в тексте подобные обобщения не приводятся. В книгу не включены также дискретные задачи оптимизации. Они требуют совсем иных методов исследования, чем непрерывные, и примыкают к комбинаторике и математической логике.

Мне неоднократно приходилось читать лекции по теории и методам оптимизации, в частности в Московском государственном университете и Институте проблем управления. На их примере я убедился, сколь различен подход к предмету у математиков, вычислителей и практиков. Предлагаемая книга представляет собой попытку некоторого компромиссного решения, рассчитанного на все эти категории читателей. Обращаясь к математикам, я хотел бы подчеркнуть, что данная книга — не учебник, она не связана с имеющимися вузовскими программами по курсу «методы оптимизации», не все приводимые формулы доказаны, много материала вынесено для самостоятельной работы в упражнения и т. д. С другой стороны, вычислительные программы показаны много материала вынесено для самостоятельной работы в упражнения и т. д. С другой стороны, вычислитель не найдет окончательных формул алгоритмов или готовых текстов программ оптимизации; в работе нет обсуждения важных деталей вычислительной «технологии»; ряд приводимых результатов представляет лишь теоретический интерес. Иными словами, книга не может служить сборником готовых рецептов решения конкретных задач. Наконец, инженеру или экономисту придется примирияться с несколько абстрактным характером изложения (примеры и приложения даны лишь в заключительных главах).

Большое значение для меня имела продолжительная совместная работа с Я. З. Цыпкиным, который является инициатором написания этой книги. Постоянная творческая активность Я. З. Цыпкина, его эрудиция и интерес к задачам оптимизации оказали серьезное влияние на формирование моих взглядов на предмет и методологию данной области науки. Я многому научился у своих учеников; в частности, вычислительный опыт Е. Н. Белова и В. А. Скокова существенно помог мне при написании книги. Большой труд по редактированию взял на себя Ю. Е. Нестеров. Неоценимую помощь в работе над книгой мне оказала Г. М. Корпелевич. Высокий профессионализм Г. Н. Архиповой позволил избежать многих трудностей при оформлении рукописи. Всем названным лицам я выражаю глубокую признательность.

Апрель, 1980 г.

Б. Т. Полак

## СПИСОК ОБОЗНАЧЕНИЙ

$\mathbf{R}^n$  —  $n$ -мерное вещественное евклидово пространство  
 $\{x_1, \dots, x_n\}$  — компоненты вектора  $x \in \mathbf{R}^n$ .

$\|\cdot\|$  — норма в  $\mathbf{R}^n$ ;  $\|x\|^2 = x_1^2 + \dots + x_n^2$   
 $\langle \cdot, \cdot \rangle$  — скалярное произведение в  $\mathbf{R}^n$ :  $(x, y) = x_1y_1 + \dots + x_ny_n$ .  
 $I$  — единичная матрица.

$A^T$  — матрица, транспонированная к  $A$ .  
 $A^+$  — псевдообратная матрица к  $A$  ( $\S$  1 гл. 6).  
 $A \geqslant B$  — матрицы  $A$  и  $B$  симметричны и  $A - B$  неотрицательно определены.

$\|A\|$  — норма матрицы  $A$ ;  $\|A\| = \max_{\|x\|=1} \|Ax\|$ .  
 $\rho(A)$  — спектральный радиус матрицы  $A$  ( $\S$  1 гл. 2).  
 $x \geqq y$  — все компоненты вектора  $x \in \mathbf{R}^n$  и  $y \in \mathbf{R}^n$  не меньше соответствующих компонент вектора  $y \in \mathbf{R}^n$ :  $x_i \geqq y_i$ ,  $i = 1, \dots, n$ .

$\mathbf{R}_+^n$  — неотрицательный ортант в  $\mathbf{R}^n$ ;  $\mathbf{R}_+^n = \{x \in \mathbf{R}^n; x \geqq 0\}$ .  
 $x_+$  — положительная часть вектора  $x \in \mathbf{R}^n$ :  $(x_+)_i = \max\{0, x_i\}$ ,  $i = 1, \dots, n$ .

$x^* = \arg \min_{x \in Q} f(x)$  — любая точка глобального минимума  $f(x)$  на  $Q$ :  
 $x^* \in Q$ ,  $f(x^*) = \min_{x \in Q} f(x)$ .  
 $X^* = \{\arg \min_{x \in Q} f(x)\}$  — множество точек глобального минимума  $f(x)$  на  $Q$ :

$\nabla f(x)$ ,  $f'(x)$  — градиент скалярной функции  $f(x)$  ( $\S$  1 гл. 1).  
 $\nabla g(x)$ ,  $g'(x)$  — производная векторной функции  $g(x)$ , матрица Якоби ( $\S$  1 гл. 1).  
 $\nabla^2 f(x)$ ,  $f''(x)$  — матрица вторых производных, гессиан ( $\S$  1 гл. 1).  
 $L'_x(x, y)$ ,  $L_{xx}(x, y)$  — градиент и матрица вторых производных  $L(x, y)$  по переменной  $x$ .  
 $\partial f(x)$  — субградиент выпуклой функции ( $\S$  1 гл. 5 и  $\S$  1 гл. 9).  
 $\partial_\varepsilon f(x)$  —  $\varepsilon$ -субградиент выпуклой функции ( $\S$  1 гл. 5).

по

$f'(x; y)$  — производная функции  $f(x)$  в точке  $x$  по направлению  $y$  (§ 1 гл. 1 и § 1 гл. 5).

$D(f)$  — область определения функции  $f(x)$  (§ 1 гл. 5).

$\text{Conv } Q$  — выпуклая оболочка множества  $Q$  (§ 1 гл. 5).

$Q^0$  — внутренность множества  $Q$ .

$\emptyset$  — пустое множество.

$P_Q(x)$  — проекция точки  $x$  на множество  $Q$  (§ 1 гл. 5).  $\rho(x, Q) = \inf_{y \in Q} \|x - y\|$ .

$o(h(x))$  — если  $g: \mathbf{R}^n \rightarrow \mathbf{R}^m$ ,  $h: \mathbf{R}^n \rightarrow \mathbf{R}^s$  и  $\|g(x)\|/\|h(x)\| \rightarrow 0$  при  $\|x\| \rightarrow 0$ , то  $g(x) = o(h(x))$ .

$O(h(x))$  — если  $g: \mathbf{R}^n \rightarrow \mathbf{R}^m$ ,  $h: \mathbf{R}^n \rightarrow \mathbf{R}^s$ ,  $\|g(x)\| \leqslant \alpha \|h(x)\|$  при  $\|x\| \leqslant \varepsilon$ , то  $g(x) = O(h(x))$ .  
 $O(u_k)$  — если последовательности  $u_k \in \mathbf{R}^n$ ,  $v_k \in \mathbf{R}^m$ ,  $k = 1, 2, \dots$ , такие, что  $\|v_k\|/\|u_k\| \rightarrow 0$  при  $k \rightarrow \infty$ , то  $v_k = o(u_k)$ .

$O(u_k)$  — если для последовательностей  $u_k \in \mathbf{R}^n$ ,  $v_k \in \mathbf{R}^m$ ,  $k = 1, 2, \dots$ , найдутся  $\alpha > 0$ ,  $k_0$  такие, что  $\|v_k\| \leqslant \alpha \|u_k\|$  при  $k \geqslant k_0$ , то  $v_k = O(u_k)$ .

$M_\xi$  — математическое ожидание случайной величины  $\xi$ .

$M(\xi|x)$  — условное математическое ожидание случайной величины  $\xi$ , зависящей от  $x$ , при фиксированном значении  $x$ .

$\forall$  — квантор общности:  $\forall x \in Q$  — «для всех  $x \in Q»$ .

$\blacktriangleleft$  — знак, ставящийся в конце доказательства (или в конце утверждения, если последнее приводится без доказательства).

Обычно буквы  $x, y, a, b, \dots$  используются для обозначения векторов,  $a, b, \dots$  — скаляров,  $A, B, \dots$  — матриц,  $i, j, k, \dots$  — целых чисел,  $Q, S, \dots$  — множеств. Иерархия последовательности векторов обозначается  $x^0, x^1, \dots, x^k, \dots, x_l$  — компоненты вектора  $x$ .

Нумерация лемм, теорем и формул в каждом параграфе независимая. При ссылках в пределах параграфа указывается только номер формулы, в пределах главы — номер формулы и параграфа, в остальных случаях приводится номер формулы, параграфа и главы.

## ВВЕДЕНИЕ

Обычно наши действия в условиях неоднозначности выбора определяются некоторой целью, которую мы стремимся достичь наилучшим образом. Тем самым человеческая деятельность связана с постоянным (сознательным или бессознательным) решением оптимизационных задач. Более того, многие законы природы носят вариационный характер, хотя здесь и неуместно говорить о наличии цели.

Можно было бы думать, что подобная распространенность задач оптимизации должна была найти свое отражение в математике. Однако в действительности до середины нынешнего столетия задачи на экстремум рассматривались в математике лишь эпизодически, развитая теория и методы решения подобных задач были созданы совсем недавно.

Наиболее простая задача безусловной минимизации функции многих переменных привлекла внимание математиков во времена, когда закладывались основы математического анализа. Она во многом стимулировала создание дифференциального исчисления, а необходимое условие экстремума (равенство градиента нулю), полученное Ферма в 1629, явилось одним из первых крупных результатов анализа. Позже в работах Ньютона и Лейбница были по существу сформулированы условия экстремума II порядка (т. е. в терминах вторых производных) для этой задачи.

Другой класс задач на экстремум, традиционно рассматривавшийся в математике, — это задачи вариационного исчисления. Следы интереса к ним можно найти и в античной математике (разного рода изотериметрические проблемы), однако подлинное рождение вариационного исчисления произошло в конце XVIII века, когда И. Бернулли сформулировал знаменитую задачу о брахистохроне. На современном языке классическая задача вариационного исчисления представляет собой бесконечномерную задачу безусловной оптимизации с минимизируемым функционалом специального (интегрального) вида. Условия экстремума I порядка в вариационном исчислении были получены Эйлером (уравнение Эйлера), а II порядка — Лежандром и Якоби. Важный вопрос о существовании решения в вариационном исчислении был впервые поставлен Вейерштрассом в второй половине XIX века.

Стоя задачи, о которых говорилось выше (конечномерная и бесконечномерная), являются примерами задач безусловной

минимизации. Задачи на условный экстремум рассматривались в классической математике лишь для ограниченных типов равенств. Знаменитое правило множителей Лагранжа (сформулированное в XVIII веке) представляет собой необходимое условие экстремума 1 порядка в подобных задачах (и в конечномерных, и в задачах вариационного исчисления). Поразительно, что такие же условия для задач с ограничениями типа неравенств были получены лишь недавно. Самы по себе системы неравенств (вне связи с задачами минимизации) изучали Фурье, Минковский, Бейль и другие ученые; созданный ими аппарат позволял без труда получить условия экстремума в задачах с ограничениями — неравенствами.

Первые работы по экстремальным задачам при наличии ограничений общего вида относятся к концу 30-х — началу 40-х годов нашего века. Корни этих работ были различны.

Специалисты по вариационному исчислению, принадлежавших к Чикагской школе (Блисс, Больц, Макшней, Грейвс, Хестенс и др.), стимулировал интерес возможно более широкой постановки вариационных задач. Здесь в 1937 г. появилась работа Валентайна, посвященная условиям экстремума для задач вариационного исчисления при наличии разного рода ограниченных типов неравенств. Позже были созданы (Макшней, Кокс) общие схемы анализа абстрактных экстремальных задач. Одному из аспирантов Чикагского Университета, Карушу, было поручено исследовать в качестве упражнения конечномерные задачи минимизации с общими ограничениями. Каруш получил в 1939 г. условий экстремума первого и второго порядков для гладкого случая, когда работе не отнеслись серьезно, и она не была опубликована. К тем же по существу условиям экстремума несколько позже пришел американский математик Фриц Джон, занимавшийся экстремальными проблемами в геометрии (типа отыскания эллипсоида наименьшего объема, описанного вокруг заданного выпуклого тела). Работа Джона была отвергнута одним серьезным математическим журналом и была напечатана лишь в 1949 г.

Независимо от американских исследований оптимизационная тематика развивалась и в СССР. Пионером в этой области был Л. В. Канторович, опубликовавший в 1939 г. книгу, содержащую математические постановки ряда экономических задач. Последние не укладывались в рамки стандартного математического аппарата, а именно, являлись задачами минимизации линейной функции на множестве, задаваемом линейными ограничениями типа равенств и неравенств. Л. В. Канторович разработал теорию подобных задач и предложил некоторые (не полностью алгоритмизированные) методы их решения. В 1940 г. появилась заметка того же автора, содержащая общую формулировку условий экстремума при наличии ограничений в беско-

нечномерном пространстве. Работы Л. В. Канторовича в то время не привлекли внимания математиков и остались, по сутиству, незамеченными. Как видит читатель, судьба не благоприятствовала первым исследованиям по неклассическим задачам оптимизации.

Время для них созрело несколько позже, в конце 40-х годов. Под влиянием прикладной тематики, которой ему приходилось заниматься в годы войны, Данциг в США стал изучать задачи минимизации линейной функции при линейных ограничениях, получившие название задач линейного программирования. Он сформулировал условия оптимальности решений в линейном программировании. Под влиянием работ фон Неймана по теории игр, Данциг, Гейл, Кун и Таккер создали теорию двойственности в линейном программировании — специфическую формулировку условий экстремума.

Вскоре после разработки теории линейного программирования возникает ее естественное обобщение на нелинейный случай. Задача минимизации нелинейной функции при нелинейных ограничениях была названа задачей математического программирования (что вряд ли можно признать удачным, учитывая перегруженность обоих терминов). Если и минимизируемая функция, и ограничения выпуклы, то говорят о задаче выпуклого программирования. Условия экстремума для задач математического программирования стали широко известны после работы Кунга и Таккера 1950 г.; по существу, это были те же условия Каруша — Джона. Для выпуклого случая Кун и Таккер сформулировали условия экстремума в, терминах седловой точки, эта формулировка пригодна и для негладких задач.

Существенный прогресс в теории оптимизации был достигнут при изучении так называемых задач оптимального управления, являющихся непосредственным обобщением классической задачи вариационного исчисления и заключающихся в оптимизации функционалов от решений обыкновенных дифференциальных уравнений, правые части которых включают голографические выборы функции («управления»). Необходимые условия оптимальности для этих задач были сформулированы и доказаны Л. С. Понtryaginom, В. Г. Болтянским и Р. В. Гамкрелидзе в 1956—1958 гг. в форме так называемого принципа максимума. В иной форме условия оптимальности для подобных задач были получены Беллманом на основе идей динамического программирования. Эти результаты были столь связаны со специфической формой задач оптимального управления, что не сразу было осознано их родство с условиями экстремума для задач математического программирования.

В 60-е годы появился цикл работ (А. Я. Дубовикова и А. А. Милотина, Б. Н. Пшеничного, Нейштадта, Халкина, Варги

и других авторов), в которых были предложены *общие схемы получения условий экстремума* для *абстрактных задач оптимизации* с ограничениями, позволившие охватить как теорему Куна — Таккера, так и принцип максимума. Это дало возможность по-новому взглянуть на известные результаты и, в частности, выделить в них стандартную часть, которую можно получать с помощью общих схем, и нестандартную, связанную со спецификой задачи. Удобным аппаратом для исследования экстремальных задач оказался *выпуклый анализ* — сравнительно новый раздел математики, получивший завершенную форму в работах Р. Рокафеллара. В настоящее время техника вывода условий оптимальности развита в совершенстве.

Выше в основном говорилось о той части теории оптимизации, которая связана с условиями экстремума. Однако найти с помощью условий экстремума явное решение задачи удается лишь в редких случаях. Сложность или невозможность отыскания аналитического решения обнаружилась и в других разделах математики; постепенно стало ясно, что любая задача может считаться решенной, если указан алгоритм, позволяющий численно построить приближенное решение с требуемой точностью. Этот принципиально новый подход, подкрепленный появлением ЭВМ и приведший к возникновению вычислительной математики, существенно затронул и проблематику оптимизации. Одним из центральных направлений здесь стала разработка и обоснование *численных методов* решения.

Математиков прошлого относительно мало интересовали вычислительные проблемы, и хотя некоторые методы решения линейных уравнений и безусловной минимизации связывают с именами Ньютона, Гаусса, Коши, эти результаты остались изолированными в творчестве упомянутых учёных и их последователей.

Пожалуй, первыми нужду в численных методах минимизации испытывали статистики. В задачах оценки параметров изменения метода максимального правдоподобия или метода наименьших квадратов приводило к необходимости отыскания экстремума функции многих переменных (вообще говоря, неквадратичной). Статистикам (Карри, Левенбергу, Крокету, Чернову и другим) принадлежат первые исследования по численным методам безусловной минимизации, выполненные в 40-х — 50-х годах. В связи с проблемами планирования эксперимента и решения уравнений регрессии в работах Бокса, Роббинса и Монро, Кифера и Вольфовича в начале 50-х годов были предложены методы минимизации функций при наличии случайных помех.

Другим разделом математики, где происходило зарождение методов оптимизации, была *линейная алгебра*. Необходимость

решения больших систем линейных уравнений, возникающих при конечно-разностной аппроксимации уравнений с частными производными, привела к развитию итеративных методов линейной алгебры. Но задача решения системы линейных уравнений эквивалентна минимизации квадратичной функции, и многие итеративные методы удобно строить и обосновывать, опираясь на этот факт. Таковы методы покоординатного спуска, наискосокового спуска, сопряженных градиентов и ряд других методов, которые были созданы в линейной алгебре к началу 50-х годов. Естественным шагом было перенесение подобных методов на неквадратичный случай.

С необходимостью решения задач оптимизации столкнулись специалисты по теории автоматического регулирования. Трудами В. В. Казакевича, А. А. Фельдбаума, А. А. Первозванского в 50-х годах была создана теория экстремального регулирования и предложены специальные методы оптимизации действующих объектов в реальном масштабе времени.

Первый численный метод нелинейного программирования — метод *итрафных функций* — был введен Курантом в 1943 г. из соображений, связанных с физической природой рассматривавшейся задачи.

Наконец,

метод дал предложенный Данлигом в конце 40-х годов *симплекс-метод* для решения задач линейного программирования. Обилие приложений и наличие хороших программ для ЭВМ привели к широкой популярности симплекс-метода прежде всего среди экономистов.

До какого-то времени такого рода исследования были спорадическими и не объединялись ни единим подходом, ни аппаратом. Однако к середине 60-х годов в рамках вычислительной математики сложилось самостоятельное направление, связанное с *численными методами оптимизации*. С тех пор непрерывно шло интенсивное развитие этого направления как вширь (разработка новых методов, исследование новых классов задач), так и вглубь (выработка единого аппарата для анализа сходимости и скорости сходимости, классификации и унификации методов). В настоящее время эта область вычислительной математики может считаться окончательно сформировавшейся. Разработано множество численных методов для всех основных классов задач оптимизации — безусловной минимизации гладких и негладких функций в конечно-мерных и бесконечно-мерных пространствах, условной минимизации при ограничениях типа равенств и (или) неравенств в выпуклом или невыпуклом случае и т. д. Для большинства методов имеется строгое обоснование, выяснена скорость сходимости, установлена область применимости. Конечно, многие проблемы еще не решены до конца (построение

## БЕЗУСЛОВНАЯ МИНИМИЗАЦИЯ

эффективных методов для некоторых специальных типов задач, проблема оптимальных методов, подробная численная проверка имеющихся алгоритмов, создание доступных и отработанных машинных программ и т. п.). Однако, по-видимому, период наибольшей активности в области численных методов оптимизации остался позади.

В предлагаемой вниманию читателя книге делается попытка систематического изложения современного состояния основ оптимизации.

Мы начинаем изучение проблем оптимизации с классической задачи безусловной минимизации гладкой функции:

$$\inf f(x), \quad x \in \mathbf{R}^n.$$

Этой задаче будет уделено большое внимание не только из-за ее важности, но и потому, что в силу ее простоты для нее наиболее четко видна схема математического исследования общих оптимизационных задач и идейные основы методов оптимизации.

## § 1. Сведения из математического анализа

**1. Дифференцирование скалярных функций.** Скалярная функция  $f(x)$   $n$ -мерного аргумента  $x$  (кратко это записывается :  $\mathbf{R}^n \rightarrow \mathbf{R}$ ) называется *дифференцируемой в точке*  $x$ , если найдется вектор  $a \in \mathbf{R}^n$  такой, что для всех  $y \in \mathbf{R}^n$

$$(x+y) = f(x) + (a, y) + o(y). \quad (1)$$

Вектор  $a$  в (1) называется *производной* или *градиентом* функции  $f(x)$  в точке  $x$  и обозначается  $f'(x)$  или  $\nabla f(x)$ . Итак, градиент определяется равенством

$$(x+y) = f(x) + (\nabla f(x), y) + o(y). \quad (2)$$

Иначе можно сказать, что функция дифференцируема в точке  $x$ , если она *допускает линейную аппроксимацию первого порядка* в этой точке, т. е. найдется линейная функция  $\tilde{f}(y) = \tilde{f}(x) + (\nabla \tilde{f}(x), y)$  такая, что  $|f(x+y) - \tilde{f}(y)| = o(y)$ . Ясно, что градиент определяется однозначно, при этом  $\nabla f(x)$  — вектор с компонентами  $(\partial f(x)/\partial x_1, \dots, \partial f(x)/\partial x_n)$ . Вычислять градиент можно, во-первых, непосредственно из определения, во-вторых, с помощью его координатной записи и, в-третьих, с помощью правила дифференцирования сложной функции (см. ниже (12)).

Пусть, например,  $f(x)$  — квадратичная функция

$$f(x) = (Ax, x)/2 - (b, x),$$

где  $A$  — симметричная  $n \times n$ -матрица,  $b \in \mathbf{R}^n$ . Тогда  $f(x+y) = (Ax+y, x+y)/2 - (b, x+y) = (Ax, x)/2 - (b, x) + (Ax-b, y) + (Ay, y)/2 = f(x) + (Ax-b, y) + (Ay, y)/2$ . Но  $|Ay, y| \leq \|A\| \|y\|^2$ , поэтому  $(Ay, y)/2 = o(y)$ . Итак,  $f(x)$  дифференцируема в любой точке  $x$  и

$$\nabla f(x) = Ax - b.$$

Функция  $f(x)$  называется *дифференцируемой на множестве*  $Q \subset \mathbf{R}^n$ , если она дифференцируема во всех точках  $Q$ . Если  $f(x)$  дифференцируема на всем пространстве  $\mathbf{R}^n$ , то говорят просто, что она *дифференцируема*.

Пусть  $f(x)$  дифференцируема на отрезке  $[x, x+ty]$  (т. е. для точек вида  $x+ty$ ,  $0 \leq t \leq 1$ ). Рассмотрим функцию одного переменного  $\Phi(\tau) = f(x+ty)$  и вычислим ее производную для  $0 \leq t \leq 1$ :

$$\frac{\Phi(\tau + \Delta\tau) - \Phi(\tau)}{\Delta\tau} = \frac{f(x + (\tau + \Delta\tau)y) - f(x + \tau y)}{\Delta\tau} =$$

$$= \frac{(\nabla f(x + \tau y), \Delta\tau y) + o(\Delta\tau y)}{\Delta\tau},$$

$$\Phi'(\tau) = \lim_{\Delta\tau \rightarrow 0} \frac{\Phi(\tau + \Delta\tau) - \Phi(\tau)}{\Delta\tau} = (\nabla f(x + \tau y), y).$$

Таким образом,  $\Phi(\tau)$  дифференцируема на  $[0, 1]$  и

$$\Phi'(\tau) = (\nabla f(x + \tau y), y). \quad (4)$$

Величина

$$f'(x; y) = \lim_{\varepsilon \rightarrow +0} \frac{f(x + \varepsilon y) - f(x)}{\varepsilon} \quad (5)$$

называется *производной по направлению* (или *вариацией*) функции  $f(x)$  в точке  $x$  по направлению  $y$ . Производная по направлению может существовать и для неплавких функций. Например, для  $f(x) = \|x\|$  имеем  $f'(0; y) = \|y\|$ . Если  $f(x)$  имеет в точке  $x$  производную по всем направлениям, линейную по  $y$ :  $f'(x; y) = (a, y)$ , то говорят, что  $f(x)$  *дифференцируема по Гато* в точке  $x$ . Такая функция имеет частные производные, причем  $f'(x; e_i) = \partial f(x)/\partial x_i$  ( $e_i$  — координатные орты),  $a = (\partial f/\partial x_1, \dots, \partial f/\partial x_n)$ . Из формулы (4) следует, что если  $f(x)$  дифференцируема в точке  $x$ , то она дифференцируема и по Гато, причем  $f'(x; y) = \Phi'(0) = (\nabla f(x), y)$ .

Обратное, вообще говоря, неверно. Например, функция  $f: \mathbf{R}^n \rightarrow \mathbf{R}^l$ ,  $n \geq 2$ , вида

$$f(x) = \begin{cases} 1, & \text{если } \|x - a\| = \|a\|, \\ 0 & \text{в остальных точках,} \end{cases} \quad (7)$$

где  $a \in \mathbf{R}^n$ ,  $a \neq 0$ , дифференцируема в точке  $0$  по любому направлению и  $f'(0; y) = 0$  для всех  $y$ , т. е. она дифференцируема по Гато в нуле, однако она не дифференцируема (и даже не непрерывна) в этой точке. Отметим еще, что иногда (чтобы подчеркнуть отличие от дифференцируемости по Гато) употребляют термин «*дифференцируемость по Фреше*» вместо «*дифференцируемость*».

Если функция  $f(x)$  дифференцируема на  $[x, x+ty]$ , то, пользуясь (4) и *формулой Ньютона—Лейбница*  $\Phi(1) = \Phi(0) + \int_0^1 \Phi'(\tau) d\tau$ , получаем запись остаточного члена в (2) в интегральной форме:

$$f(x+y) = f(x) + \int_0^1 (\nabla f(x+\tau y), y) d\tau =$$

$$= f(x) + (\nabla f(x), y) + \int_0^1 (\nabla f(y+\tau y) - \nabla f(x), y) d\tau. \quad (8)$$

Другой полезный результат — *теорема о среднем* — следует из *формулы конечных приращений*  $\Phi(1) = \Phi(0) + \Phi'(0)$ ,  $0 \leq \theta \leq 1$ , и (4):

$$f(x+y) = f(x) + (\nabla f(x+\theta y), y), \quad (9)$$

где  $0 \leq \theta \leq 1$  — некоторое число.

**Упражнение.** 1. Докажите, что: а)  $\nabla \|x\| = x/\|x\|$  при  $x \neq 0$ ; при  $x = 0$  функция  $\|x\|$  недифференцируема; б)  $\nabla \|x+y\|^2 = 2x+y$ .

2. Докажите, что из непрерывности по  $x$  производной Гато следует дифференцируемость.

**2. Дифференцирование векторных функций.** До сих пор речь шла о дифференцируемости скалярных функций. Совершенно аналогично определяется дифференцируемость векторных функций. Функция  $g: \mathbf{R}^n \rightarrow \mathbf{R}^m$  называется *дифференцируемой в точке*  $x$ , если найдется матрица  $A$  размерности  $m \times n$  такая, что для всех  $y \in \mathbf{R}^n$

$$g(x+y) = g(x) + Ay + o(y). \quad (10)$$

Матрица  $A$  называется *производной Якоби* отображения  $g(x)$ , и для нее применяется то же обозначение  $g'(x)$  или  $\nabla g(x)$ , что и в скалярном случае. Итак,

$$g(x+y) = g(x) + g'(x)y + o(y), \quad (11)$$

т. е. дифференцируемая в точке  $x$  функция допускает в этой точке линейную аппроксимацию первого порядка. Очевидно, что для дифференцируемой векторной функции  $g(x) = (g_1(x), \dots, g_m(x))$  элементы матрицы Якоби определяются формулой

$$g'(x)_{ij} = \partial g_i(x) / \partial x_j.$$

Пусть  $g: \mathbf{R}^n \rightarrow \mathbf{R}^m$  — дифференцируемая в точке  $x$  функция, а  $h: \mathbf{R}^m \rightarrow \mathbf{R}^s$  дифференцируема в точке  $g(x)$ . Тогда спрашивается правило дифференцирования сложных функций

$$[h(g(x))]' = h'(g(x))g'(x), \quad (12)$$

где в правой части стоит произведение матриц  $h'$  и  $g'$ .

**Теорема о среднем для векторных функций неверна, т. е., вообще говоря, не существует такого  $0 \leqslant \theta \leqslant 1$ , что  $\mathbf{g}(x+y) = \mathbf{g}(x) + \mathbf{g}'(x + \theta y)$**

для дифференцируемой на  $[x, x+y]$  функции  $\mathbf{g}: \mathbf{R}^n \rightarrow \mathbf{R}^m$ ,  $m > 1$ . Однако справедлива формула аналогичная (8): если  $\mathbf{g}'(x)$  дифференцируема на  $[x, x+y]$ , то

$$\mathbf{g}(x+y) = \mathbf{g}(x) + \int_0^1 \mathbf{g}'(x+\tau y) y d\tau =$$

$$= \mathbf{g}(x) + \mathbf{g}'(x)y + \int_0^1 (\mathbf{g}'(x+\tau y) - \mathbf{g}'(x))y d\tau. \quad (13)$$

Из формулы (13), в частности, получаем следующие полезные оценки. Если  $\|\mathbf{g}'(x+\tau y)\| \leqslant L$  при  $0 \leqslant \tau \leqslant 1$ , то

$$\|\mathbf{g}(x+y) - \mathbf{g}(x)\| \leqslant L\|y\|, \quad (14)$$

а если  $\mathbf{g}'(x)$  удовлетворяет условию Липшица на  $[x, x+y]$ :  $\|\mathbf{g}'(u) - \mathbf{g}'(v)\| \leqslant L\|u-v\|$ ,  $u, v \in [x, x+y]$ , то

$$\|\mathbf{g}(x+y) - \mathbf{g}(x) - \mathbf{g}'(x)y\| \leqslant L\|y\|^2/2. \quad (15)$$

Как и в скалярном случае, функция  $\mathbf{g}: \mathbf{R}^n \rightarrow \mathbf{R}^m$ , дифференцируемая во всех точках  $\mathbf{R}^n$ , называется *дифференцируемой*.

Упражнение. Используя (12) и результат упражнения 1, покажите, что  $\nabla(Ax - b)_+ \parallel A^T(Ax - b)_+$ , где  $A$  — матрица  $n \times n$ .

**3. Вторые производные.** Скалярная функция  $f(x)$  на  $\mathbf{R}^n$  называется *дважды дифференцируемой в точке*  $x$ , если она дифференцируема в этой точке и найдется симметричная  $n \times n$ -матрица  $H$  такая, что для всех  $y \in \mathbf{R}^n$

$$f(x+y) = f(x) + (\nabla f(x), y) + (\nabla^2 f(x+\theta y)y, y)/2. \quad (16)$$

Эта матрица называется *матрицей вторых производных*, *матрицей Гессе* или *гессианом* и обозначается  $f''(x)$  или  $\nabla^2 f(x)$ . Иначе говоря, функция дважды дифференцируема в точке  $x$ , если она допускает квадратичную аппроксимацию второго порядка в окрестности этой точки, т. е. существует квадратичная функция  $\tilde{f}(y) = f(x) + (\nabla f(x), y) + (\nabla^2 f(x)y, y)/2$  такая, что  $|\tilde{f}(x+y) - f(x+y)| = o(\|y\|^2)$ .

Уточним полученные ранее оценки для дважды дифференцируемых функций. Рассмотрим вновь скалярную функцию  $\Phi(\tau) = f(x + \tau y)$  в предположении, что  $f$  дважды дифференцируема

на  $[x, x+y]$ . Так же, как и выше, показывается, что эта функция дважды дифференцируема и

$$\Phi''(\tau) = (\nabla^2 f(x+\tau y)y, y). \quad (17)$$

Тогда из формулы Тейлора с остаточным членом в интегральной форме  $\Phi(1) = \Phi(0) + \Phi'(0) + \int_0^1 \int_0^1 \Phi''(\tau) d\tau dt$  следует

$$f(x+y) = f(x) + (\nabla f(x), y) + \int_0^1 \int_0^1 (\nabla^2 f(x+\tau y)y, y) d\tau dt. \quad (18)$$

В частности, если  $\|\nabla^2 f(x+\tau y)\| \leqslant L$  для  $0 \leqslant \tau \leqslant 1$ , то отсюда получаем

$$|\tilde{f}(x+y) - f(x) - (\nabla f(x), y)| \leqslant (L/2)\|y\|^2, \quad (19)$$

$$\text{а если } \|\nabla^2 f(x+\tau y) - \nabla^2 f(x)\| \leqslant L\tau\|y\|, \text{ то}$$

$$|\tilde{f}(x+y) - f(x) - (\nabla f(x), y) - (1/2)(\nabla^2 f(x)y, y)| \leqslant (L/6)\|y\|^3. \quad (20)$$

Если же пользоваться формулой Тейлора с остаточным членом в форме Лагранжа

$$f(x) = \varphi(0) + \varphi'(0) + \varphi''(\theta)/2, \quad 0 \leqslant \theta \leqslant 1,$$

то получаем, что найдется  $0 \leqslant \theta \leqslant 1$ , для которого

$$\tilde{f}(x+y) = f(x) + (\nabla f(x), y) + (\nabla^2 f(x+\theta y)y, y)/2. \quad (21)$$

#### Упражнения.

4. Покажите, что  $\nabla^2 f(x) — матрица с элементами  $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ .$

5. Докажите, что: а)  $\frac{\nabla^2}{\nabla^2}(Ax, x)/2 = (b, x) \equiv A$ , где  $A$  — симметричная

$n \times n$ -матрица,  $b \in \mathbf{R}^n$ ; б)  $\nabla^2 \|x\|^{-1} = I/\|x\| - \frac{xx^T}{\|x\|^3}$  для  $x \neq 0$ ; в)  $\nabla^2(c)x^2 = 2cc^T$ ,  $c \in \mathbf{R}^n$ .

6. Проверьте, что  $f''(x) = (f'(x))'$ , т. е. производная векторной функции  $f'(x)$  совпадает со второй производной  $f(x)$ .

**4. Выпуклые функции.** Понятие выпуклости играет огромную роль в теории экстремальных задач, и мы будем многократно обращаться к нему. Числовая функция  $f(x)$  на  $\mathbf{R}^n$  называется *выпуклой*, если

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y) \quad (22)$$

для любых  $x, y \in \mathbf{R}^n$ ,  $0 \leqslant \lambda \leqslant 1$ . Это определение имеет наглядный геометрический смысл — график функции на отрезке  $[x, y]$  лежит ниже хорды, соединяющей точки  $(x, f(x))$  и  $(y, f(y))$  (рис. 1). В определении выпуклости фигурируют две точки  $x, y$  и их выпуклые комбинации. Совершенно аналогичное неравенство справедливо для любого числа точек и их выпуклых комбинаций.

**Лемма 1** (неравенство Иенсена). Пусть  $f(x)$  — выпуклая функция на  $\mathbf{R}^n$ . Тогда для любых  $x^1, \dots, x^k \in \mathbf{R}^n$  и  $\lambda_i \geq 0$ ,

$$\sum_{i=1}^k \lambda_i = 1, \quad f(\lambda_1 x^1 + \dots + \lambda_k x^k) \leq \lambda_1 f(x^1) + \dots + \lambda_k f(x^k). \quad \blacktriangle \quad (23)$$

Функция  $f(x)$ , для которой  $-f(x)$  является выпуклой, называется *вогнутой*. Очевидно, что аффинная функция  $f(v) = (a, v) + \beta$  является и выпуклой и вогнутой.

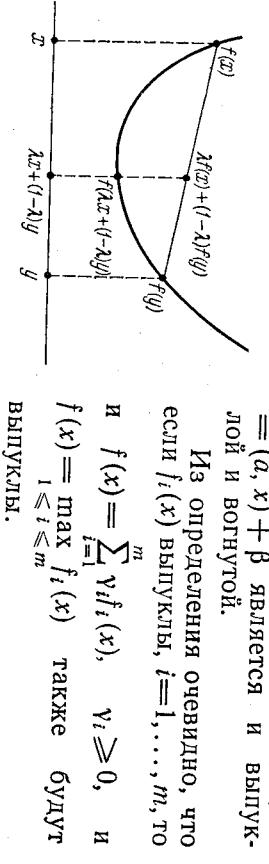


Рис. 1. Выпуклая функция.

Из определения очевидно, что если  $f_i(x)$  выпуклы,  $i=1, \dots, m$ , то и  $f(x) = \sum_{i=1}^m \nu_i f_i(x)$ ,  $\nu_i \geq 0$ , и  $f(x) = \max_{1 \leq i \leq m} f_i(x)$  также будут выпуклы.

Важным частным случаем выпуклых функций являются строго и сильно выпуклые функции. Функция  $f(x)$  на  $\mathbf{R}^n$  называется строго выпуклой, если для любых  $x \neq y$ ,  $0 < \lambda < 1$

$$(24)$$

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y),$$

и сильно выпуклой с константой  $l > 0$ , если при  $0 \leq \lambda \leq 1$

$$(25)$$

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - l\lambda(1 - \lambda)\|x - y\|^2/2.$$

Ясно, что сильно выпуклая функция строго выпукла.

Важно иметь аналитические критерии, по которым можно судить о выпуклости функций. Такие критерии существуют и очень просты для случая дифференцируемых функций. Они основываются на следующем элементарном результате.

**Лемма 2.** Пусть  $\psi(\tau)$  — дифференцируемая функция на  $\mathbf{R}$ . Тогда выпуклость  $\psi(\tau)$  эквивалентна монотонности производной  $(\psi'(\tau))'$  при  $\tau_1 \geq \tau_2$ , строгая выпуклость — строгой монотонности  $(\psi'(\tau_1) > \psi'(\tau_2))$  при  $\tau_1 > \tau_2$ , а сильная выпуклость — сильной монотонности  $(\psi(\tau_1) - \psi(\tau_2) \geq l(\tau_1 - \tau_2))$ ,  $\tau_1 > \tau_2$ .  $\blacktriangle$

**Лемма 3.** Для дифференцируемой функции  $f(x)$  на  $\mathbf{R}^n$  выпуклость эквивалента неравенству

$$f(x + y) \geq f(x) + (\nabla f(x), y), \quad (26)$$

$$f(x + y) > f(x) + (\nabla f(x), y), \quad y \neq 0. \quad (27)$$

$$\begin{aligned} \text{а) сильная выпуклость — неравенство} \\ f(x + y) \geq f(x) + (\nabla f(x), y) + l\|y\|^2/2 \end{aligned} \quad (28)$$

для любых  $x, y \in \mathbf{R}^n$ .  $\blacktriangle$  Иначе говоря, график (строго) выпуклой функции лежит (строго) выше касательной гиперплоскости, а для сильно выпуклой функции график лежит выше некоторого параболоида (рис. 2).

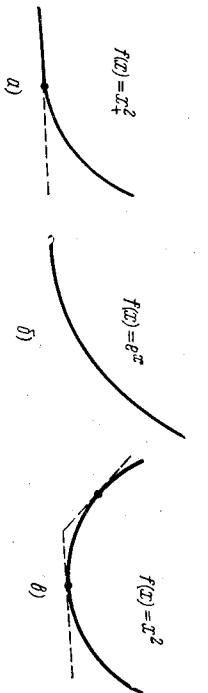


Рис. 2. Типы выпуклости: а) выпуклая функция; б) строго выпуклая функция.

Из (26) получаем полезное неравенство

$$(\nabla f(x) - \nabla f(y), x - y) \geq 0, \quad (29)$$

а для сильно выпуклой — условие сильной монотонности

$$(\nabla f(x) - \nabla f(y), x - y) > 0, \quad x \neq y. \quad (30)$$

Наиболее просто критерий выпуклости формулируется для дважды дифференцируемых функций  $f(x)$ : выпуклость эквивалента выполнению условия

$$\nabla^2 f(x) \geq 0, \quad (31)$$

а сильная выпуклость — выполнению условия

$$\nabla^2 f(x) \geq lI. \quad (32)$$

для всех  $x$ . Если же

$$\begin{aligned} \text{для всех } x, \text{ то } f(x) \text{ строго выпукла. Последнее условие является} \\ \text{лишь достаточным (например, для строго выпуклой функции} \\ f(x) = \|x\|^4 \text{ будет } \nabla^2 f(0) = 0). \\ \text{Пусть } x^* \text{ — точка минимума дифференцируемой сильно выпуклой (с константой } l) \text{ функции } f(x). \text{ Такая точка заведомо} \end{aligned}$$

существует, единственна и  $\nabla f(x^*) = 0$  (см. ниже §§ 2, 3). Поэтому из неравенств (28), (31) получаем

$$f(x) \geq f(x^*) + l\|x - x^*\|^2/2, \quad (35)$$

$$(\nabla f(x), x - x^*) \geq l\|x - x^*\|^2, \quad (36)$$

$$\|\nabla f(x)\| \geq l\|x - x^*\|. \quad (37)$$

### Упражнение.

7. Убедитесь в справедливости следующих утверждений:

a) функция  $(Ax, x)/2 - (b, x)$ ,  $A \geq 0$  сильно выпукла;

б) функция  $(Ax, x)/2 - (b, x)$  с вырожденной матрицей  $A \geq 0$  (в частности, линейная функция) выпукла, но не строго выпукла;

в) функция  $\|x\|^\alpha$  выпукла при  $\alpha \geq 1$ , строго выпукла при  $\alpha > 1$ , сильно выпукла только при  $\alpha = 2$ .

## § 2. Условия экстремума

Условия экстремума гладких функций на всем пространстве хорошо известны. Мы рассмотрим их, однако, достаточно подобно, так как они служат моделью, по которой строятся аналогичные условия в более сложных случаях.

### 1. Необходимое условие I порядка.

Точка  $x^*$  называется *локальным минимумом*  $f(x)$  на  $\mathbf{R}^n$ , если найдется  $\varepsilon > 0$  такое, что  $f(x) \geq f(x^*)$  для всех  $x$  из  $\varepsilon$ -окрестности  $x^*$  (т. е. при  $\|x - x^*\| \leq \varepsilon$ ). Иногда в таком случае говорят просто о *точке минимума*, отрасывая слово «локальный». Нужно, однако, помнить, что это различие между локальным и глобальным минимумом (т. е. точкой  $x^*$  такой, что  $f(x) \geq f(x^*)$  для всех  $x$ ). В необходимых условиях экстремума можно говорить просто о точке минимума, поскольку если некоторое условие выполнется для локального минимума, то оно же справедливо для глобального.

При формулировке достаточных условий нужно различать, какой из типов минимума подразумевается.

**Теорема 1** (Ферма). *Пусть  $x^*$  — точка минимума  $f(x)$  на  $\mathbf{R}^n$  и  $f'(x)$  дифференцируема в  $x^*$ . Тогда*

$$f'(x^*) = 0. \quad (1)$$

**Доказательство.** Пусть  $\nabla f(x^*) \neq 0$ . Тогда

$$\begin{aligned} &= f'(x^*) - \tau \|\nabla f(x^*)\|^2 + o(\tau \nabla f(x^*)) \\ &= f'(x^*) - \tau (\|\nabla f(x^*)\|^2 + \tau^{-1}o(\tau)) < f'(x^*) \end{aligned}$$

для достаточно малых  $\tau > 0$  по определению  $o(\tau)$ . Но это противоречит тому, что  $x^*$  — точка локального минимума.  $\blacktriangleleft$

Приведенное доказательство весьма поучительно. В предложении, что условие экстремума не выполняется, показано, как построить точку с меньшим значением  $f(x)$ . Таким образом, это доказательство указывает путь для построения метода



Рис. 3. Стационарные точки:  $a$  — точка минимума,  $b$  — точка перегиба,  $c$  — точка максимума.

Доказательство следует сразу из формулы (26) § 1, так как  $f(x) \geq f(x^*) + (\nabla f(x^*))^T(x - x^*) = f(x^*)$  для любого  $x \in \mathbf{R}^n$ .  $\blacktriangleleft$

Таким образом, для выпуклых функций необходимое условие экстремума является и достаточным. Впоследствии мы увидим, что эта ситуация является общей и для других типов выпуклых экстремальных задач.

**3. Необходимое условие II порядка.** Для невыпуклых задач можно продолжить исследование условий экстремума с помощью старших производных.

**Теорема 3.** *Пусть  $x^*$  — точка минимума  $f(x)$  на  $\mathbf{R}^n$  и  $f'(x)$  дифференцируема в  $x^*$ . Тогда*

$$\nabla^2 f(x^*) \geq 0.$$

**Доказательство.** По теореме 1  $\nabla f(x^*) = 0$ , поэтому для произвольного  $y$  и достаточно малых  $\tau$

$$\begin{aligned} f(x^*) &\leq f(x^* + \tau y) = f(x^*) + \tau^2 (\nabla^2 f(x^*) y, y)/2 + o(\tau^2), \\ (\nabla^2 f(x^*) y, y) &\geq o(\tau^2)/\tau^2. \end{aligned}$$

Переходя к пределу при  $\tau \rightarrow 0$ , получаем  $(\nabla^2 f(x^*) y, y) \geq 0$ .

В силу произвольности  $y$  это означает, что  $\nabla^2 f(x^*) \geq 0$ .  $\blacktriangleleft$

**4. Достаточное условие II порядка.**

**Теорема 4.** *Пусть в точке  $x^*$   $f(x^*)$  дважды дифференцируема, выполнено необходимое условие I порядка (т. е.  $\nabla f(x^*) = 0$ ) и*

$$\nabla^2 f(x^*) > 0. \quad (3)$$

*Тогда  $x^*$  — точка локального минимума.*

Доказательство. Пусть  $y$  — произвольный вектор с единичной нормой. Тогда

$$f(x^* + \tau y) = f(x^*) + \tau^2 (\nabla f(x^*) y, y)/2 + o(\tau^2 \|y\|^2) \geq$$

$$\geq f(x^*) + \tau^2/2 + o(\tau^2),$$

где  $\tau > 0$  — наименьшее собственное значение  $\nabla^2 f(x^*)$ , а функция  $o(\tau^2)$  не зависит от  $y$ . Поэтому найдется  $\tau_0$  такое, что при  $0 \leq \tau \leq \tau_0$  будет  $\tau^2/2 \geq o(\tau^2)$ , т. е.  $f(x^* + \tau y) \geq f(x^*)$ . ▲

Если в точке  $x^*$  выполняются необходимые условия I и II порядков (т. е.  $\nabla f(x^*) = 0$ ,  $\nabla^2 f(x^*) \geq 0$ ), но не выполняется положительно определенной), то  $x^*$  может и не являться минимума (например,  $f(x) = x^3$ ,  $x \in \mathbf{R}^1$ ) и в принципе анализ можно продолжить с помощью старших производных. Для однородного случая правило действий хорошо известно (нужно найти первую отличную от 0 производную), для многомерного случая техника вычислений сложна.

### 5. Зачем нужны условия экстремума?

Обычно в книгах по математическому анализу предлагается следующий рецепт для отыскания точек экстремума. Нужно найти все точки, удовлетворяющие необходимому условию I порядка, а затем использовать с помощью условий II порядка, отобрав из них точки минимума. Таким образом, создается впечатление, что условия экстремума — эффективный инструмент для решения задач оптимизации.

Нужно со всей определенностью подчеркнуть, что это способом не так. Отыскать в явной форме точку минимума с помощью условий экстремума удается лишь в редких случаях, для специально построенных примеров (они обычно и приводятся в учебниках). Дело в том, что решение системы уравнений  $\nabla f(x) = 0$  — задача ничуть не более простая, чем исходная, и явный вид ответа в ней найти, как правило, нельзя.

Зачем же в таком случае нужны условия экстремума и почему им уделяется столь большое внимание в теории экстремальных задач? Отчасти такое внимание является данью традиции, когда численные методы оптимизации не изучались, а решением задачи считалось лишь некоторое аналитическое выражение. Нередко при этом вывод условий экстремума для различных типов экстремальных задач превращается в чисто тематическую игру, где целью является получение изощренных формулировок для разного рода вырожденных ситуаций без всякой заботы о том, как пользоваться этими условиями экстремума. При чтении некоторых монографий создается впечатление, что формулировка условий оптимальности является главным (или даже единственным) объектом исследования в области экстремальных задач.

### § 3. СУЩЕСТВОВАНИЕ, ЕДИНСТВЕННОСТЬ, УСТОЙЧИВОСТЬ МИНИМУМА

Важной частью математической теории экстремальных задач (и в частности, задач безусловной оптимизации) являются проблемы существования, единственности и устойчивости решения. 1. Существование решения. Вопрос о существовании точки минимума обычно решается совсем просто с помощью следующей теоремы.

**Теорема 1** (Вейерштрасс). Пусть  $f(x)$  непрерывна на  $\mathbf{R}^n$  и множество  $Q_\alpha = \{x: f(x) \leq \alpha\}$  для некоторого  $\alpha$  непусто и ограничено. Тогда существует точка глобального минимума  $f(x)$  на  $\mathbf{R}^n$ .

Доказательство. Пусть  $f(x^k) \rightarrow \inf_{x \in \mathbf{R}^n} f(x) > \alpha$ , тогда  $x^k \in Q_\alpha$  для достаточно больших  $k$ . Множество  $Q_\alpha$  замкнуто (в силу непрерывности  $f(x)$ ) и ограничено, т. е. компактно, а потому у последовательности  $x_k$  существует предельная точка  $x^* \in Q_\alpha$ . Из непрерывности  $f(x)$  следует, что  $f(x^*) = \inf_{x \in \mathbf{R}^n} f(x)$ , т. е.  $x^* = \operatorname{argmin}_{x \in \mathbf{R}^n} f(x)$ . ▲

Предположение об ограниченности  $Q_\alpha$  существенно (например, функции  $x$  и  $1/(1+x^2)$  непрерывны на  $\mathbf{R}^1$ , но не имеют точки минимума). В некоторых случаях можно доказать существование решения и в ситуациях, не охватываемых теоремой 1 (см. ниже упр. 2).

#### Упражнения.

1. Докажите, что дифференцируемая сильно выпуклая функция на  $\mathbf{R}^n$  достигает минимума (вспомогайтесь неравенством (28) § 1 и теоремой 1).

На наш взгляд, условия экстремума являются той основой, на которой строятся методы решения оптимизационных задач, с этой точки зрения и нужно рассматривать вопрос о их полезности. В дальнейшем мы увидим, что, во-первых, в ряде случаев условия экстремума хотя и не дают возможности явно найти решение, но сообщают много информации о его свойствах. Во-вторых, доказательство условий экстремума или вид этих условий часто указывают путь построения методов оптимизации. Мы уже видели выше, что доказательство условия  $\nabla f(x^*) = 0$  естественно приводит к градиентному методу минимизации. В-третьих, при обосновании методов приходится делать ряд предположений. Обычно при этом требуется, чтобы в точке  $x^*$  выполнялось достаточное условие экстремума. Таким образом, условия экстремума фигурируют в теоремах о сходимости методов. Наконец, сами доказательства сходимости обычно строятся на том, что показывается, как «кнеязка» в условиях экстремума стремится к нулю. У читателя будет в дальнейшем много поводов убедиться в полезности условий экстремума.

**2.** Пусть  $f(x) = \langle Ax, x \rangle - (b, x)$ ,  $A \geqslant 0$  и  $f(x)$  ограничена снизу (например,  $f(x) = \|Bx - d\|^2$ ). Докажите, что  $f(x)$  достигает минимума на  $\mathbf{R}^n$ , хотя в теореме 1, вообще говоря, не выполняются (множество  $Q_\alpha$  не обязательно ограничено).

**2. Единственность решения.** Будем называть точку минимума локально единственно<sup>1</sup> если в некоторой ее окрестности нет других локальных минимумов. Будем говорить, что  $x^*$  — невырожденная точка минимума, если в ней выполнено достаточное условие экстремума II порядка, т. е.  $\nabla f(x^*) = 0$ ,  $\nabla^2 f(x^*) \geqslant 0$ .

**Теорема 2. Невырожденная точка минимума локально единственна.**

**Доказательство.** В соответствии с упражнением 6 § 1  $\nabla f(x^*) = \nabla f(x^*) + \nabla^2 f(x^*)(x - x^*) + o(x - x^*)$ , поэтому  $\|\nabla f(x)\| = \|\nabla^2 f(x^*)(x - x^*)\| + o(\|x - x^*\|) \geqslant l\|x - x^*\| + o(\|x - x^*\|) \geqslant 0$  для достаточно малых  $\|x - x^*\|$ , поскольку при  $\nabla^2 f(x^*) = A \geqslant 0$  имеем  $\|Ax\| \geqslant \|x\|$  для всех  $x$ , где  $l > 0$  — наименьшее собственное значение  $A$ . Таким образом, в некоторой окрестности  $x^*$  нет стационарных точек  $f(x)$ , значит, и точек минимума. Для выпуклых функций ответ на вопрос об единственности минимума часто может быть получен совсем просто.

**Теорема 3. Точка минимума строго выпуклой функции (глобально) единственна.**

Доказательство следует непосредственно из определения строгой выпуклости.  $\blacktriangleleft$

**3. Устойчивость решения.** При практическом решении задач оптимизации постоянно приходится сталкиваться со сложными проблемами. Пусть метод оптимизации приводит к построению минимизирующей последовательности, следует ли отсюда ее сходимость к решению? Если вместо исходной задачи минимизации решается задача, близкая к ней, можно ли утверждать близость их решений? Такого типа проблемы относятся к области теории экстремальных задач, связанной с понятиями устойчивости и корректности. Мы будем пользоваться термином «устойчивость» задач оптимизации, оставляя термин «корректность» для задач, не связанных с оптимизацией (речь идет о алгебраических, интегральных, операторных уравнениях и т. п.).

Точка  $x^*$  локального минимума  $f(x)$  называется локально устойчивой, если к ней сходится любая локальная минимизирующая последовательность, т. е. если найдется  $\delta > 0$  такое, что из  $f(x^k) \rightarrow f(x^*)$ ,  $\|x^k - x^*\| \leqslant \delta$  следует  $x^k \rightarrow x^*$ .

**Теорема 4. Точка локального минимума непрерывной функции  $f(x)$  локально устойчива тогда и только тогда, когда она локально единственна.**

**Доказательство.** Пусть  $x^*$  локально единственна. Возьмем произвольную локальную минимизирующую последовательность  $x^k$ ,  $\|x^k - x^*\| \leqslant \delta$ ,  $f(x^k) \rightarrow f(x^*)$ . В силу компактности

шара в  $\mathbf{R}^n$  из нее можно выбрать сходящуюся подпоследовательность  $x^{k_i} \rightarrow \bar{x}$ ,  $\|\bar{x} - x^*\| \leqslant \delta$ . Из непрерывности  $f(x)$  следует, что  $f(\bar{x}) = \lim f(x^{k_i}) = f(x^*)$ , но тогда  $\bar{x} = x^*$  (так как  $x^*$  локально единственная точка минимума). Поскольку это же верно для любой другой подпоследовательности, то и вся последовательность  $x^k$  сходится к  $x^*$ . Таким образом,  $x^*$  локально устойчива. Обратно, пусть  $x^*$  локально устойчива, но существует другая точка минимума  $x_1^* \neq x^*$ ,  $\|x_1^* - x^*\| \leqslant \delta$ . Тогда  $f(x_1^*) = f(x^*)$ . Возьмем последовательность точек  $x^1, x^2, \dots$ , поочередно совпадающих то с  $x^*$ , то с  $x_1^*$ . Она является минимизирующей, но не сходится, что противоречит локальной устойчивости  $x^*$ .  $\blacktriangleleft$

Аналогично доказывается следующий результат.

**Теорема 5. Пусть  $x^*$  — локально устойчива точка минимума непрерывной функции  $f(x)$ , а  $g(x)$  — непрерывная функция. Тогда для достаточно малых  $\varepsilon > 0$  функция  $f(x) + \varepsilon g(x)$  имеет локально единственную точку минимума  $x_\varepsilon$  в окрестности  $x^*$  и  $x_\varepsilon \rightarrow x^*$  при  $\varepsilon \rightarrow 0$ .  $\blacktriangleleft$**

Таким образом, из устойчивости следует близость точек минимума исходной и «возмущенной» функции.

Невырожденная точка минимума, как следует из теорем 2 и 4, является локально устойчивой. В этом случае результат теоремы 5 можно уточнить.

**Теорема 6. Пусть  $x^*$  — невырожденная точка минимума  $f(x)$ , а функция  $g(x)$  непрерывно дифференцируема в окрестности  $x^*$ . Тогда для достаточно малых  $\varepsilon > 0$  существует  $x_\varepsilon$  — локальная точка минимума функции  $f(x) + \varepsilon g(x)$  в окрестности  $x^*$ , причем**

$$x_\varepsilon = x^* - \varepsilon [\nabla^2 f(x^*)]^{-1} \nabla g(x^*) + o(\varepsilon). \quad (1)$$

Можно ввести и глобальное понятие устойчивости точек минимума. Для этого нужно в определении слова «локальный» заменить на «глобальный». Именно, точка глобального минимума называется глобально устойчивой, если к ней сходится любая минимизирующая последовательность. Будем в этом случае говорить о глобальной устойчивости задачи минимизации. Повторяя почти дословно доказательство теоремы 4, получаем, что если  $x^*$  — единственная точка глобального минимума непрерывной функции  $f(x)$  и множество  $Q_\alpha = \{x: f(x) \leqslant \alpha\}$  не пусто и ограничено для некоторого  $\alpha \geqslant f(x^*)$ , то  $x^*$  глобально устойчива. Требование ограниченности  $Q_\alpha$  существенно. Например, у функции  $f(x) = x^2/(1+x^4)$ ,  $x \in \mathbf{R}^1$ , точка глобального минимума  $x^* = 0$  единственна, но не глобально устойчива (так как минимизирующая последовательность  $x^k \rightarrow \infty$  не сходится к  $x^*$ ).

Можно было бы ввести следующее более широкое определение устойчивости, которое не предполагает единственности минимума. Множество  $X^*$  точек глобального минимума  $f(x)$ ,

назовем *слабо устойчивым*, если все предельные точки любой слабой устойчивости принадлежат  $X^*$ . Критерий качественности указан в упражнении 5.

Помимо качественной характеристики (устойчива или нет точка минимума), важно иметь количественные оценки устойчивости. Такие оценки, позволяющие судить о близости точки  $x$  к решению  $x^*$ , если  $f(x)$  близко к  $f(x^*)$ , уже были получены ранее для сильно выпуклых функций. Именно, из (35) § I получаем

$$\|x - x^*\|^2 \leq 2l^{-1} (f(x) - f(x^*)),$$

где  $l$  — константа сильной выпуклости. Аналогичная локальная оценка справедлива для невырожденной точки минимума:

$$\|x - x^*\|^2 \leq 2l^{-1} (f(x) - f(x^*)) + o(f(x) - f(x^*)), \quad (2)$$

где  $l$  — наименьшее собственное значение матрицы  $\nabla^2 f(x^*)$ .

Таким образом, число  $l$  характеризует «запас устойчивости» точки минимума. Оно, однако, не всегда удобно как мера устойчивости — например, оно меняется при умножении  $f(x)$  на константу. Поэтому часто используют следующий «нормированный» показатель. Назовем *обусловленностью* точки минимума  $x^*$  число

$$\mu = \overline{\lim}_{\delta \rightarrow 0} (\sup_{x \in L_\delta} \|x - x^*\|^2 / \inf_{x \in L_\delta} \|x - x^*\|^2),$$

$$L_\delta = \{x: f(x) = f(x^*) + \delta\}. \quad (4)$$

Иначе говоря,  $\mu$  характеризует степень вытянутости линий уровня  $f(x)$  в окрестности  $x^*$ . Ясно, что всегда  $\mu \geq 1$ . Если  $\mu$  велико, то линии уровня сильно вытянуты — функция имеет *обратный* характер, т. е. резко возрастает по одному направлению и слабо меняется по другим. В таких случаях говорят о *плохо обусловленных* задачах минимизации. Если же  $\mu$  близко к 1, то линии уровня  $f(x)$  близки к сферам; это соответствует хорошо обусловленным задачам. В дальнейшем мы увидим, что число обусловленности  $\mu$  возникает во многих проблемах, связанных с безусловной минимизацией, и может служить одним

Для квадратичной функции

$$f(x) = (Ax, x)/2 - (b, x), \quad A > 0$$

имеем  $L_0 = \{x: (A(x - x^*), x - x^*) = 2\delta\}$ , поэтому максимум

$l_1$  — нормированный собственный вектор, отвечающий наименьшему собственному значению  $\lambda_1$  матрицы  $A$ , а множитель  $\gamma_1$  определяется из условия  $x_1 \in L_\delta$ , т. е.  $\lambda_1 y_1^2 = 2\delta$ ,  $y_1 = (2\delta/\lambda_1)^{1/2}$ .

Аналогично минимум  $\|x - x^*\|$  по  $x \in L_\delta$  достигается на векторе

$$(5)$$

$x_n = x^* + y_n l_n$ ,  $l_n$  — собственный вектор, отвечающий наибольшему собственному значению  $\lambda_n$ ,  $y_n = (2\delta/\lambda_n)^{1/2}$  (рис. 4). Таким образом, отношение  $\mu(\delta) = \|x_1 - x^*\|^2 / \|x_n - x^*\|^2 = y_1^2 / y_n^2 = \lambda_n / \lambda_1$  фактически не зависит от  $\delta$  и

$$\mu = \frac{\lambda_n}{\lambda_1}. \quad (6)$$

Заметим, что отношение наибольшего к наименьшему собственному значению называется в линейной алгебре *числом обусловленности матрицы*.

Для случая неквадратичной функции обусловленность задачи ее минимизации равна числу обусловленности гессиана в точке минимума. Именно, если  $x^*$  — невырожденная точка минимума, то

$$\mu = \frac{L}{l}, \quad (7)$$

где  $L$  — наибольшее, а  $l$  — наименьшее собственное значение матрицы  $\nabla^2 f(x^*)$ .

Мы увидим в дальнейшем, что в практических приложениях часто возникают неустойчивые или плохо обусловленные задачи оптимизации. Обсуждение методов их решения будет дано в § 1 Гл. 6.

#### Упражнения.

3. Покажите, что точка минимума строго выпуклой непрерывной функции глобально устойчива.

4. Проверьте, что в условиях упражнения 2 множество точек минимума  $f(x)$  непрерывна и  $Q_\alpha = \{x: f(x) \leq \alpha\}$  непусто и слабо устойчиво.

5. Докажите, что если  $f(x)$  непрерывна и  $Q_\alpha = \{x: f(x) \leq \alpha\}$  ограничено некоторого  $\alpha > \inf f(x)$ , то множество точек минимума  $f(x)$  слабо устойчиво.

6. Покажите, что обусловленность задачи не меняется при монотонных преобразованиях функции и ортогональных преобразованиях переменных, т. е. обусловленность  $f(x)$  и  $f(Ux)$  одинакова, если  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  — ортогональная матрица.

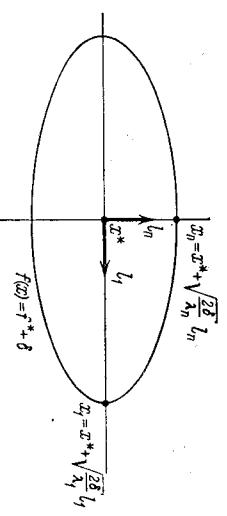


Рис. 4. Обусловленность квадратичной функции.

#### § 4. ГРАДИЕНТНЫЙ МЕТОД

##### 1. Эвристические соображения.

Мы переходим к анализу наиболее важных в идейном отношении методов безусловной минимизации — градиентного и Ньютона. Эти методы, редко

применяемые на практике в «чистом виде», служат моделью для построения более реалистических алгоритмов. На примере данных методов будет подробно разобран вопрос о сходимости — общая техника построения доказательства сходимости, описана разница между теоретическими результатами о сходимости и практическими результатами о сходимости и

Предположим, что в любой точке  $x$  можно вычислить градиент функции  $\nabla f(x)$ . В такой ситуации наиболее простым методом минимизации  $f(x)$  является градиентный, в котором, начиная с некоторого начального приближения  $x_0$ , строится итерационная последовательность

$$x^{k+1} = x^k - \gamma_k \nabla f(x^k), \quad (1)$$

где параметр  $\gamma_k \geq 0$  задает длину шага. К методу (1) можно прийти из разных соображений.

Во-первых, при доказательстве необходимых условий экстремума (теорема 1 § 2) мы использовали то обстоятельство, что если в точке  $x$  условие экстремума не выполняется ( $\nabla f(x) \neq 0$ ), то значение функции можно уменьшить, перейдя к точке  $x - \tau \nabla f(x)$  при достаточно малом  $\tau > 0$ . Итеративно применяя этот прием, приходим к методу (1).

Во-вторых, в точке  $x^k$  дифференцируемая функция  $f(x)$  приводится к линейной  $f_k(x) = f(x^k) + (\nabla f(x^k), x - x^k)$  с помощью аппроксимации  $f_k(x)$  в окрестности  $x^k$ . Поэтому можно искать минимум некоторым  $\varepsilon_k$  и решить вспомогательную задачу

$$\min \|x - x_*^k\| \leq \varepsilon_k. \quad (2)$$

Ее решение естественно принять за новое приближение  $x^{k+1}$ . Можно остаться в окрестности  $x^k$  и иначе, добавив к  $f_k(x)$  штраф за отклонение от  $x^k$ . Например, можно решить вспомогательную задачу

$$\min [f_k(x) + \alpha_k \|x - x^k\|^2] \quad (3)$$

и ее решение взять в качестве  $x^{k+1}$ . Читателю предстоит убедиться в том, что решение задач (2), (3) задается формулой (1).

В-третьих, можно в точке  $x^k$  выбрать направление локального наискорейшего спуска, т. е. то направление  $y^k$ ,  $\|y^k\| = 1$ , для которого достигается минимум  $f'(x^k; y)$ . Используя формулу (6) из § 1 для производной по направлению, получаем

$$y^k = \operatorname{argmin}_{\|y\|=1} (\nabla f(x^k), y) = -\nabla f(x^k)/\|\nabla f(x^k)\|. \quad (4)$$

Таким образом, направление наискорейшего спуска противоположно направлению градиента.

Мы привели здесь столь подробно эти соображения, поскольку они же будут использоваться при построении методов оптимизации в более сложных ситуациях (например, при наличии ограничений). Однако в этих ситуациях они могут привести к различным методам.

**2. Сходимость.** Рассмотрим простейший вариант градиентного метода, в котором  $\gamma_k \equiv \gamma$ :

$$x^{k+1} = x^k - \gamma \nabla f(x^k). \quad (5)$$

Нас будет интересовать поведение этого метода при различных предположениях относительно  $f(x)$  и  $\gamma$ .

**Теорема 1.** Пусть  $f(x)$  дифференцируема на  $\mathbb{R}^n$ , градиент  $f(x)$  удовлетворяет условию Липшица:

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|, \quad (6)$$

$f(x)$  ограничена сверху:

$$f(x) \geq f^* > -\infty \quad (7)$$

и  $\gamma$  удовлетворяет условию

$$\lim_{k \rightarrow \infty} \nabla f(x^k) = 0, \quad 0 < \gamma < 2/L. \quad (8)$$

Тогда в методе (5) градиент стремится к 0:

$$\lim_{k \rightarrow \infty} \nabla f(x^k) = 0,$$

а функция  $f(x)$  монотонно убывает:  $f(x^{k+1}) \leq f(x^k)$ .

Доказательство. Подставим в формулу (8) § 1  $x = x^k$ ,

$$y = -\gamma \nabla f(x^k) \text{ и воспользуемся (6):}$$

$$\begin{aligned} f(x^{k+1}) &= f(x^k) - \gamma \|\nabla f(x^k)\|^2 - \gamma \int_0^1 (\nabla f(x^k - \tau \gamma \nabla f(x^k)) - \\ &\quad - \nabla f(x^k), \nabla f(x^k)) d\tau \leq f(x^k) - \gamma \|\nabla f(x^k)\|^2 + \\ &\quad + L \gamma^2 \|\nabla f(x^k)\|^2 \int_0^1 \tau d\tau = f(x^k) - \gamma \left(1 - \frac{1}{2} L \gamma^2\right) \|\nabla f(x^k)\|^2. \end{aligned}$$

Суммируя неравенства

$$f(x^{k+1}) \leq f(x^k) - \alpha \|\nabla f(x^k)\|^2, \quad \alpha = \gamma(1 - L\gamma/2) \quad (9)$$

по  $k$  от 0 до  $s$ , получаем

$$f(x^{s+1}) \leq f(x^0) - \alpha \sum_{k=0}^s \|\nabla f(x^k)\|^2.$$

Поскольку  $\alpha > 0$  в силу (8), то

$$\sum_{k=0}^s \|\nabla f(x^k)\|^2 \leq \alpha^{-1} (f(x^0) - f(x^{s+1})) \leq \alpha^{-1} (f(x^0) - f^*)$$

при всех  $s, t$ , т. е.  $\sum_{k=0}^{\infty} \|\nabla f(x^k)\|^2 < \infty$ . Отсюда  $\|\nabla f(x^k)\| \rightarrow 0$ . ▲

Покажем, что все условия этой теоремы существенны. Нарушения условия (6) могут быть двух типов. Во-первых, функция  $f(x)$  может быть недостаточно гладкой в какой-либо точке. Пусть, например,  $f(x) = \|x\|^{1+\alpha}$ ,  $0 < \alpha < 1$ . Эта функция дифференцируема, но ее градиент не удовлетворяет условию Липшица, так как  $\|\nabla f(x) - \nabla f(0)\|/\|x - 0\| = (\alpha + 1)\|x\|^{\alpha-1} \rightarrow \infty$  при  $\|x\| \rightarrow 0$ . В этом случае будет  $\gamma \|\nabla f(x^k)\| \gg \|x^k - x^*\| = \|x^k\|$  при малых  $\|x^k\|$ , т. е. шаг в методе (5) получается слишком и монотонность убывания  $f(x)$  нарушается. Во-вторых, (6) не выполняется для функций, растущих быстрее квадратичной. Пусть, например,  $f(x) = \|x\|^{2+\alpha}$ ,  $\alpha > 0$ , тогда  $\|\nabla f(x) - \nabla f(0)\|/\|x - 0\| = (2 + \alpha)\|x\|^{\alpha} \rightarrow \infty$  при  $\|x\| \rightarrow \infty$ . При этом для всякого  $\gamma > 0$  можно указать такое  $x^0$ , что метод (5), примененный к функции  $\|x\|^{2+\alpha}$ ,  $\alpha > 0$ , с начальным приближением  $x^0$ , расходится, поскольку будет  $\|x^{k+1}\| > \|x^k\|$ ,  $k = 0, 1, \dots$ .

Если не выполнено условие (7), то функция  $f(x)$  не достигает минимума и градиент в методе (5) не обязан стремиться к 0 (например, если  $f(x)$  линейна:  $f(x) = (c, x)$ , то  $\|\nabla f(x)\| \equiv \|c\| > 0$ ).

Наконец, выбирать  $\gamma$ , нарушая условие (8), вообще говоря, также нельзя, что видно на примере функции  $f(x) = Lx^2/2$ ,  $x \in \mathbf{R}^1$ . Действительно, если  $\gamma \geq 2/L$ , то в методе (5) для этой функции будет  $f(x^{k+1}) \geq f(x^k)$ ,  $k = 0, 1, \dots$ , при любом  $x^0$ .

С другой стороны, при сделанных в теореме 1 предположениях нельзя доказать ничего большего, например, сходимость последовательности  $x^k$ . Примером может служить  $f(x) = 1/(1 + \|x\|^2)$ . Эта функция удовлетворяет условиям теоремы и при любом  $x^0 \neq 0$  будет  $\|x^k\| \rightarrow \infty$ .

Если потребовать, чтобы множество  $\{x : f(x) \leq f(x^0)\}$  было ограничено, то из  $x^k$  можно выбрать подпоследовательность, сходящуюся к некоторой стационарной точке  $x^*$ . Однако точка  $x^*$  не обязана быть точкой локального или глобального минимума. В частности, градиентный метод (5) (или даже (1) с произвольным выбором  $\gamma_k$ ), начатый из некоторой стационарной точки  $x^0$ , останется в этой точке:  $x^k = x^0$  для всех  $k$ . Иными словами, градиентный метод «застревает» в любой стационарной точке — точке максимума, минимума или седловой. Что же касается поиска глобального минимума, то градиентный метод «не отличает» точек локального минимума от глобального и

никакой гарантии сходимости к глобальному минимуму он не дает.

Наконец, в условиях теоремы 1 скорость сходимости  $\nabla f(x^k)$  к 0 может быть очень медленной. Например, для  $f(x) = 1/x$  при  $x \geq 1$  (вид  $f(x)$  при  $x < 1$  безразличен) метод (5) при  $\gamma = 1$ ,  $x^0 = 1$  принимает вид  $x^{k+1} = x^k + (x^k)^{-2}$ , при этом можно показать (используя лемму 6 § 2 гл. 2), что  $|f'(x^k)| = O(k^{-2/3})$ .

Рассмотрим поведение градиентного метода для более высокого класса функций — сильно выпуклых. Естественно, здесь удается доказать более сильные результаты, чем в теореме 1 — именно, сходимость итераций  $x^k$  к точке глобального минимума со скоростью геометрической прогрессии.

Нам понадобится несколько неравенств, относящихся к дифференцируемым, выпуклым и сильно выпуклым функциям. **Лемма 1.** Пусть  $f(x)$  дифференцируема,  $\nabla f(x)$  удовлетворяет условию Липшица с константой  $L$  и  $f(x) \geq f^*$  для всех  $x$ . Тогда

$$\|\nabla f(x)\|^p \leq 2L(f(x) - f^*). \quad (10)$$

**Доказательство.** Сделаем из точки  $x$  шаг градиентного метода с  $\gamma = 1/L$ . Тогда (см. (9))

$$f^* \leq f(x - L^{-1}\nabla f(x)) \leq f(x) - (2L)^{-1} \|\nabla f(x)\|^2. \quad \blacktriangleleft$$

**Лемма 2.** Пусть  $f(x)$  выпукла и дифференцируема, а  $\nabla f(x)$  удовлетворяет условию Липшица с константой  $L$ . Тогда  $(\nabla f(x) - \nabla f(y), x - y) \geq L^{-1} \|\nabla f(x) - \nabla f(y)\|^2$ . (11)

**Доказательство.** Покажем (11) лишь для дважды дифференцируемых функций. Тогда (см. (13) § 1)

$$\nabla f(y) = \nabla f(x) + \int_0^1 \nabla^2 f(x + \tau(y - x))(y - x) d\tau = \nabla f(x) + A(y - x),$$

где матрица  $A = \int_0^1 \nabla^2 f(x + \tau(y - x)) d\tau$  симметрична и неотрицательно определена в силу (32) § 1, т. е.  $A \geq 0$ . Кроме того,  $\|A\| \leq L$ , так как  $\|\nabla^2 f(x)\| \leq L$  для всех  $x$  в силу условия Липшица на градиент. Поэтому

$$(\nabla f(x) - \nabla f(y), x - y) = \\ = (A(x - y), x - y) \geq \|A\|^{-1} \|A(x - y)\|^2 \geq L^{-1} \|\nabla f(x) - \nabla f(y)\|^2.$$

## § 4. ГРАДИЕНТНЫЙ МЕТОД

*Лемма 3.* Пусть  $f(x)$  — дифференцируемая сильно выпуклая (с константой  $L$ ) функция,  $x^*$  — ее точка минимума (она существует, см. упр. 1 § 3). Тогда  $\|\nabla f(x)\| \geq 2l(f(x) - f(x^*))$ .  $\blacktriangle$

*Теорема 2.* Пусть  $f(x)$  дифференцируема на  $\mathbb{R}^n$ , ее гравиент удовлетворяет условию Липшица с константой  $L$  и  $f(x)$  является сильно выпуклой функцией с константой  $l$ . Тогда при  $0 < \gamma < 2/L$  метод (5) сходится к единственной точке глобального минимума  $x^*$  со скоростью геометрической прогрессии:

$$\|x^k - x^*\| \leq cq^k, \quad 0 \leq q < 1.$$

*Доказательство.* Выполнены все условия теоремы 1, поэтому справедливо неравенство (9):

$$f(x^{k+1}) \leq f(x^k) - \gamma(1 - Ly/2)\|\nabla f(x^k)\|^2.$$

Используем лемму 3:

$$f(x^{k+1}) \leq f(x^k) - ly(2 - Ly)(f(x^k) - f(x^*)).$$

Отсюда

$$\begin{aligned} f(x^{k+1}) - f(x^*) &\leq (1 - ly(2 - Ly))(f(x^k) - f(x^*)) = q_1(f(x^k) - f(x^*)), \\ f(x^k) - f(x^*) &\leq q_1^k(f(x^0) - f(x^*)), \quad q_1 = 1 - 2ly + Ly^2. \end{aligned}$$

Поскольку  $0 < \gamma < 2/L$ , то  $0 < q_1 < 1$ , и следовательно,  $f(x^k) \rightarrow f(x^*)$ . Из неравенства (35) § 1 следует

$$\|x^k - x^*\|^2 \leq (2/l)q_1^k(f(x^0) - f(x^*)). \quad \blacktriangle$$

Рассмотрим еще более узкий класс функций — сильно выпуклых дважды дифференцируемых.

*Теорема 3.* Пусть  $f(x)$  дважды дифференцируема и

$$U \leq \nabla^2 f(x) \leq LI, \quad l > 0,$$

для всех  $x$ . Тогда при  $0 < \gamma < 2/L$

$$\|x^k - x^*\| \leq \|x^0 - x^*\|q^k, \quad q = \max\{|1 - \gamma l|, |1 - \gamma L|\} < 1. \quad (14)$$

*Величина  $q$  минимальна и равна*

$$q^* = (L - l)(L + l) \text{ при } \gamma = \gamma^* = 2/(L + l). \quad (15)$$

*Доказательство.* По формуле (13) § 1

$$\nabla f(x^k) = \nabla f(x^*) + \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*)(x^k - x^*))d\tau = A_k(x^k - x^*),$$

где в силу (13)  $U \leq A_k \leq LI$ . Поэтому

$$\begin{aligned} \|x^{k+1} - x^*\| &= \|x^k - x^* - \gamma \nabla f(x^k)\| = \\ &= \|(I - \gamma A_k)(x^k - x^*)\| \leq \|I - \gamma A_k\| \|x^k - x^*\|. \end{aligned}$$

Для всякой симметричной матрицы  $A$  имеем  $\|I - A\| = \max\{|1 - \lambda_1|, |1 - \lambda_n|\}$ , где  $\lambda_1$  и  $\lambda_n$  — наименьшее и наибольшее собственные значения  $A$ . Поэтому  $\|x^{k+1} - x^*\| \leq \leq q\|x^{k+1} - x^*\|$ ,  $q = \max\{|1 - \gamma l|, |1 - \gamma L|\}$ . Поскольку  $0 < \gamma < 2/L$ ,  $0 < l \leq L$ , то  $|1 - \gamma l| < 1$ ,  $|1 - \gamma L| < 1$ , т. е.  $q < 1$ . Минимизируя  $q$  по  $\gamma$ , получаем (15).  $\blacktriangle$

Покажем, что оценка скорости сходимости, даваемая теоремой 3, точная, она достигается для любой квадратичной функции. Пусть  $f(x) = (Ax, x)/2 - (b, x)$ ,  $A \geq 0$ ,  $0 < l = \lambda_1 \leq \lambda_2 \dots \leq \lambda_n = L$ , где  $\lambda_i$  — собственные числа матрицы  $A$ . Возьмем произвольное  $0 < \gamma < 2/L$ . Предположим, что  $|1 - \gamma l| \geq |1 - \gamma L|$ . Выберем  $x^0 = x^* + e^1$ , где  $e^1$  — собственный вектор, отвечающий  $\lambda_1$ ,  $\|e^1\| = 1$ . Тогда  $x^k - x^* = (I - \gamma A)^k(x^0 - x^*) = = (1 - \gamma \lambda_1)^k e^1$ ,  $\|x^k - x^*\| = |(1 - \gamma l)|^k = q^k \|x^0 - x^*\|$ . Аналогичным образом, если  $|1 - \gamma L| \geq |1 - \gamma l|$ , то выберем  $x^0 = x^* + e^n$ ,  $e^n$  — собственный вектор, отвечающий  $\lambda_n$ ,  $\|e^n\| = 1$ , и получим также  $\|x^k - x^*\| = |(1 - \gamma L)|^k = q^k \|x^0 - x^*\|$ . Таким образом, для всякого  $0 < \gamma < 2/L$  найдется  $x^0$  такое, что  $\|x^k - x^*\| = q^k \|x^0 - x^*\|$ ,  $q = \max\{|1 - \gamma l|, |1 - \gamma L|\}$ .

Оценку  $\|x^k - x^*\| \leq (q^*)^k \|x^0 - x^*\|$ ,  $q^* = (L - l)/(L + l)$  нельзя улучшить, даже если выбирать  $\gamma$  оптимальным образом для каждого  $x^0$ . Действительно, возьмем  $x^0 = x^* + e^1 + e^n$  (обозначения те же, что и выше). Тогда при любом  $0 < \gamma < 2/L$

$$x^k - x^* = (I - \gamma A)^k(x^0 - x^*) = (1 - \gamma l)^k e^1 + (1 - \gamma L)^k e^n,$$

$$\|x^k - x^*\| = [(1 - \gamma l)^{2k} + (1 - \gamma L)^{2k}]^{1/2} \|x^0 - x^*\|/\sqrt{2}.$$

Поэтому, если либо  $|1 - \gamma l| > q^*$ , либо  $|1 - \gamma L| > q^*$ , то  $\|x^k - x^*\|$  убывает медленнее, чем  $(q^*)^k$ . Но  $q = \max\{|1 - \gamma l|, |1 - \gamma L|\} \leq |1 - \gamma^* L| = q^*$  лишь при  $\gamma = \gamma^*$ , при этом  $|1 - \gamma^* l| = \leq |1 - \gamma^* L| = q^*$  и  $\|x^k - x^*\| = (q^*)^k \|x^0 - x^*\|$ . Аналогичное рассуждение справедливо для любой точки  $x^0$  такой, что  $(x^0 - x^*, e^1) \neq 0$ ,  $(x^0 - x^*, e^n) \neq 0$ .

Локальный аналог теоремы 3 справедлив и для невыпуклых функций.

*Теорема 4.* Пусть  $x^*$  — невырожденная точка локального минимума  $f(x)$ . Тогда при  $0 < \gamma < 2/\|\nabla^2 f(x^*)\|$  метод (5) локально сходится к  $x^*$  со скоростью геометрической прогрессии, т. е. для всякого  $\delta > 0$  найдется  $\varepsilon > 0$  такое, что при  $\|x^0 - x^*\| \leq \varepsilon$  будет

$$\begin{aligned} \|x^k - x^*\| &\leq \|x^0 - x^*\|(q + \delta)^k, \\ q &= \max\{|1 - \gamma l|, |1 - \gamma L|\} < 1, \quad 0 < U \leq \nabla^2 f(x^*) \leq LI. \end{aligned} \quad (16)$$

Белличина  $q$  минимальна и равна  $q^* = (L - l)/(L + l)$  при  $\gamma^* = 2/(L + l)$ . ▲

Другие теоремы о сходимости градиентных методов при несколько иных предположениях будут приведены в следующих главах.

#### Упражнения.

- Подробно разберите поведение градиентного метода (5) для следующих функций на  $R^1$ : а)  $|x|^{1+\alpha}, 0 < \alpha < 1$ ; б)  $|x|^{2+\alpha}, \alpha > 0$ ; в)  $x^2, \gamma$  (1 +  $x^2$ ) $^{-1}$ . При каких  $x^0$  и  $\gamma$  метод сходится, при каких расходится? Ответы. а) Сходимости нет при любом  $\gamma > 0$  и  $x^0 \neq 0$ , при этом  $|x^k| \rightarrow 1/(1 + \alpha)$ ; б) знаки  $x^k$  и  $x^{k+1}$  чередуются при  $k \geq k_0$ . б) Метод сходится, если  $\gamma(2 + \alpha)|x^0|^\alpha < 2$ , и расходится в противоположном случае, причем  $|x^k| = |x^0|^\alpha$  при  $\gamma(2 + \alpha)|x^0|^\alpha = 2$  и  $|x^k| \rightarrow \infty$  при  $\gamma(2 + \alpha)|x^0|^\alpha > 2$ . в) Метод сходится при  $0 < \gamma < 2$  и расходится при  $\gamma \geq 2$  и любом  $x^0 \neq 0$ , при этом  $|x^k| = |x^0|$ , если  $\gamma = 2$  и  $|x^k| \rightarrow \infty$  при  $\gamma > 2$ . г)  $|x^k| \rightarrow \infty$  при любом  $x^0 \neq 0$ .

- Используя неравенство  $\sum_{k=0}^{\infty} \|\nabla f(x^k)\|^2 < \infty$ , полученное при доказательстве теоремы 1, покажите, что в ее условиях  $\lim_{k \rightarrow \infty} k \|\nabla f(x^k)\|^2 = 0$ .

#### § 5. Метод Ньютона

**1. Эвристические соображения.** В градиентном методе основой является идея локальной линейной аппроксимации минимизируемой функции  $f(x)$ . Если же функция дважды дифференцируема, то естественно попытаться использовать ее квадратичную аппроксимацию в точке  $x^k$ , т. е. функцию

$$f_k(x) = f(x^k) + (\nabla f(x^k), x - x^k) + (\nabla^2 f(x^k)(x - x^k), x - x^k)/2. \quad (1)$$

В градиентном методе следующее приближение  $x^{k+1}$  искалось из условия минимума линейной аппроксимации при дополнительных ограничениях на близость к  $x^k$  (так как линейная функция не достигает минимума на всем пространстве) — см. (2), (3) и (4) в § 4. Для квадратичной аппроксимации можно попытаться не накладывать таких ограничений, так как при  $\nabla^2 f(x^k) > 0$  функция  $f_k(x)$  достигает безусловного минимума. Выберем точку минимума  $f_k(x)$  в качестве нового приближения:  $x^{k+1} = \arg\min_{x \in R^n} f_k(x)$ .

Таким образом, мы получаем метод

$$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k). \quad (2)$$

К этому методу можно прийти и из несколько иных соображений. Точка минимума должна быть решением системы  $n$  уравнений с  $n$  переменными

$$\nabla f(x) = 0.$$

Одним из основных методов решения таких систем является *метод Ньютона*, заключающийся в линеаризации уравнений в точке  $x^k$  и решении линеаризованной системы (см. ниже п. 3). Эта линеаризованная система в данном случае имеет вид

$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0 \quad (4)$$

и ее решение  $x^{k+1}$ дается формулой (2).

#### 2. Сходимость.

Теорема 1. Пусть  $f(x)$  дважды дифференцируема,  $\nabla^2 f(x)$  удовлетворяет условию Липшица с константой  $L$ ,  $f(x)$  сильно выпукла с константой  $l$  и начальное приближение удовлетворяет условию

$$q = (Ll^{-2}/2) \|\nabla f(x^0)\| < 1. \quad (5)$$

Тогда метод (2) сходится к точке глобального минимума  $x^*$  в квадратичной скорости:

$$\|x^k - x^*\| \leq (2l/L) q^{2^k}. \quad (6)$$

Доказательство. Из условий Липшича на  $\nabla^2 f(x)$  следует (см. (15) § 1)

$$\|\nabla f(x + y) - \nabla f(x) - \nabla^2 f(x)y\| \leq (L/2) \|y\|^2.$$

Возьмем здесь  $x = x^k$ ,  $y = -[\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$ , тогда  $x + y = x^{k+1}$  и

$$\|\nabla f(x^{k+1})\| \leq (L/2) \|\nabla^2 f(x^k)\|^{-1} \|\nabla f(x^k)\|^2 \leq (L/2) \|\nabla^2 f(x^k)\|^{-1} \|\nabla^2 f(x^k)\|^2.$$

Но поскольку  $\nabla^2 f(x^k) \geq lI$  (условие сильной выпуклости, см. (33) § 1), то  $\|\nabla^2 f(x^k)\|^{-1} \leq l^{-1}I$  и  $\|\nabla^2 f(x^k)\|^{-1} \leq l^{-1}$ , т. е.  $\|\nabla^2 f(x^{k+1})\| \leq (Ll^{-2}/2) \|\nabla f(x^k)\|^2$ . Итерируя это неравенство, получаем

$$\|\nabla f(x^k)\| \leq \frac{2l^2}{L} \left( \frac{L}{2l^2} \|\nabla f(x^0)\| \right)^{2^k} = \frac{2l^2}{L} q^{2^k}.$$

Применение (37) § 1 завершает доказательство. ▲

Покажем, что все условия теоремы существенны, а усилить ее утверждение, вообще говоря, нельзя. Ясно, что существование второй производной требуется в самой формулировке метода, а условие сильной выпуклости гарантирует существование  $[\nabla^2 f(x^k)]^{-1}$ . Меньшие требования к гладкости (отказ от условия Липшица на  $\nabla^2 f(x)$ ) могут привести к уменьшению скорости сходимости метода. Пусть, например,  $f(x) = |x|^{5/2}, x \in R^1$ . Тогда при  $x \geq 0$   $f'(x) = (5/2)x^{3/2}$ ,  $f''(x) = (15/4)x^{1/2}$  и  $f'''(x)$  не удовлетворяет условию Липшица. Метод принимает вид (при  $x^0 \geq 0$ )  $x^{k+1} = x^k - (4/15)(x^k)^{-1/2}$ ,  $(5/2)(x^k)^{3/2} = (1/3)x^k$ , т. е.  $x^k = (1/3)x^0$  и метод сходится к  $x^* = 0$  со скоростью геометрической прогрессии (а не с квадратичной скоростью). Наконец,

$$(3)$$

нельзя утверждать сходимость метода при любом начальном приближении (не удовлетворяющем условию (5)). Пусть задача заключается в минимизации одномерной функции, произвольная которой изображена на рис. 5. Эта функция дважды дифференцируема, сильно выпукла (так как  $f''(x) \geq 1/2 > 0$  для всех  $x$ ),  $f''(x)$  удовлетворяет условию Липшица,  $x^* = 0$ . Однако если начать итерационный процесс из любой точки  $x^0$  с  $|x^0| > 1$ , то метод не сходится:  $|x^k| \equiv 1$  для всех  $k \geq 1$ .

### Условия теоремы 1

можно несколько ослабить лишь в одном направлении — можно глобальные требования на

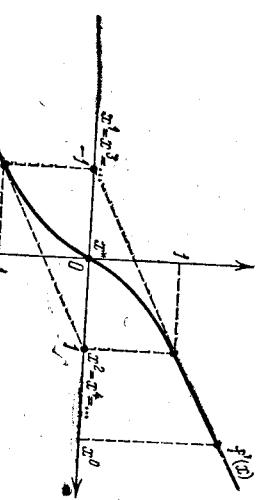


Рис. 5. Расходимость метода Ньютона.

находит  $\varepsilon > 0$  такое, что при  $\|x^0 - x^*\| \leq \varepsilon$  метод (2) сходится к  $x^*$  с квадратичной скоростью. ▲

Для квадратичной функции  $f(x) = (Ax, x)/2 - (b, x)$  с  $A > 0$  метод Ньютона сходится за 1 шаг, т. е.  $x^1 = x^*$  при любом  $x^0$ . Это очевидно, так как аппроксимирующая функция  $f_0(x)$  совпадает с  $f(x)$ . Чем ближе  $f(x)$  к квадратичной, тем быстрее сходится метод Ньютона. Формально — чем меньше  $L$ , тем в соответствии с теоремой больше область сходимости, определяемая (5), и тем быстрее скорость сходимости, определяемая величиной  $q$ .

**3. Метод Ньютона для уравнений.** Метод Ньютона может применяться не только для задач минимизации, но и для решения произвольных нелинейных уравнений

$$(7)$$

Он основан на той же идее линейной аппроксимации — на  $k$ -й итерации решается линеаризованное уравнение

$$g(x^k) + g'(x^k)(x - x^k) = 0,$$

$$x^{k+1} = x^k - g'(x^k)^{-1} g(x^k).$$

**Теорема 3.** Пусть уравнение (7) имеет решение  $x^*$ , функция  $g: \mathbf{R}^n \rightarrow \mathbf{R}^n$  дифференцируема в окрестности  $x^*$  и  $g'(x)$  удов-

летворяет условию Липшица в этой окрестности. Пусть матрица  $g'(x^*)$  невырождена. Тогда находится  $\varepsilon > 0$  такое, что при  $\|x^0 - x^*\| \leq \varepsilon$  метод (8) сходится к  $x^*$  с квадратичной скоростью.

Очевидно, что теорема 2 есть частный случай теоремы 3 при  $g(x) = \nabla f(x)$ ; доказательство остается прежним. ▲

Подчеркнем, что для сходимости (8) не нужно ни симметричности, ни положительной определенности  $g'(x)$ . В частности, метод Ньютона годится для отыскания стационарных точек функции  $f(x)$ , отличных от точек минимума.

### § 6. Роль теорем сходимости

1. Две крайние точки зрения. Возьмем какую-нибудь книгу по методам оптимизации, написанную «математиком для математиков» (типичным примером может служить монография Сеа [0.17]). Основную ее часть составляют теоремы о сходимости методов и их доказательства. Их формулировки максимально общи и абстрактны, используется аппарат современного функционального анализа. Критерии оценки результатов те же, что и в «чистой» математике — глубина, красота и простота утверждений и доказательств. Комментарии и примеры почти отсутствуют; сравнительный анализ методов не производится; важность или эффективность методов не обсуждается; численных примеров нет. Читателю, который интересуется использованием методов, приходится самому догадываться о связи математических результатов с практикой вычислений, и зачастую такую связь установить не просто. При этом нередко (особенно в научной литературе) такому же формальному исследованию подвергаются методы малоинтересные, а иногда и заведомо неэффективные. Это дало повод для появления остроумной пародии на «наукобразные» работы по методам оптимизации, написанной Вульфом [1.1]. Увы, эта пародия не исправила положения (более того, многие читатели восприняли статью всерьез, не поняв ее нарочитой нелепости).

Такая ситуация породила другой крайний взгляд, по существу отвергающий роль теории в разработке и изучении методов оптимизации. Его сторонники считают, что при создании тельство сходимости излишне, так как условия теорем труднопроверяемы в конкретных задачах, сам факт сходимости мало что дает, а оценки скорости сходимости непонятны и неэффективны. Кроме того, при реализации метода возникает масса обстоятельств, строгий учет которых невозможен (ошибки округления, приближенное решение различных вспомогательных задач и т. д.) и которые могут сильно повлиять на ход про-

песса. Поэтому единственным критерием оценки метода является практика его применения.

Не будем обсуждать эти точки зрения на абстрактном уровне, так как это потребовало бы решения общих проблем о предмете и стиле вычислительной математики. Вместо этого попытаемся на примере приведенных выше результатов о сходимости двух методов безусловной минимизации выяснить, в какой мере могут быть полезны теоремы о сходимости и почему они требуют к себе достаточно осторожного отношения.

**2. Зачем нужны теоремы о сходимости?** Ответ на этот «навивший» вопрос не так прост. Конечно, для математика, занимающегося теоретическим обоснованием методов, теоремы представляют самостоятельный интерес с точки зрения используемой в них техники, полноты исследования методов и т. д. Однако чем могут быть полезны такие теоремы человеку, собирающемуся решать практическую задачу?

Прежде всего, условия теорем выделяют класс задач, для которых можно рассчитывать на применимость метода. Эта информатия нередко носит отрицательный характер — если условия теоремы не выполняются, то метод может (но разумеется не обязан) оказаться неработоспособным. Так, наименее жесткие предположения, при которых можно обосновать сходимость градиентного метода в форме (5) § 4, заключаются в достаточной гладкости минимизируемой функции (теорема 1 § 4). При обсуждении теоремы мы видели, что нарушение этих предложений действительно может привести к расходимости процесса. Аналогичным образом более сильные условия гладкости функции для применимости метода Ньютона (теорема 1 § 5), как мы видели из примеров, также существенны. Удобно, когда подобные требования носят качественный характер (гладкость, выпуклость и т. п.) — это позволяет их проверять даже в сложных задачах. Важно также, чтобы требования в теоремах не были завышеными. Например, если судить по теореме 3 § 4, то для применимости градиентного метода нужно существование второй производной. Однако в действительности это требование излишне (см. теорему 1 § 4); оно нужно лишь для получения оценок скорости сходимости. Поэтому полезно иметь несколько теорем, в которых даются утверждения об одном методе при различных предположениях (таковы теоремы 1—4 § 4 для градиентного метода).

Теоремы о сходимости дают также важную информацию о качественном поведении метода: сходится ли он для любого начального приближения или только для достаточно хорошего, в каком смысле сходится (по функции, по аргументу или в пределе удовлетворяется условие экстремума и т. д.). Так, теорема 1 § 4 гарантирует применимость градиентного метода из любой начальной точки, в то же время утверждается лишь, что

#### § 6. РОЛЬ ТЕОРЕМ СХОДИМОСТИ

$\nabla f(x^k) \rightarrow 0$  (а сходимость по функции или аргументу может отсутствовать, что подтверждают рассмотренные там же примеры). В теореме 1 § 5 наоборот обосновывается сходимость метода Ньютона (по аргументу к глобальному минимуму) лишь для хорошего начального приближения, и как мы видели выше, это требование является существенным. Поэтому при практическом использовании метода Ньютона нужно иметь хорошее начальное приближение, в противном случае возможна расходимость метода.

Полезная информация нередко содержится и в самом доказательстве теорем о сходимости. Чаще всего они построены на той идеи, что некоторая скалярная величина монотонно убывает в процессе итераций (подробно этот вопрос будет обсуждаться в гл. 2). В теоремах 1, 2 § 4 такой величиной является сама минимизируемая функция, в теоремах 3, 4 § 4 — расстояние до точки минимума, в теореме 1 § 5 — норма градиента. Часто эта величина доступна ( $f(x)$ ,  $\|\nabla f(x)\|$ ) и по ее фактическому поведению в процессе вычислений можно судить о сходимости или расходимости метода — при нормальном течении процесса она должна убывать. Если же доказательство основано, например, на монотонном убывании  $\|x^k - x^*\|$ , то неразумно требовать монотонного убывания  $f(x)$  на каждом шаге. Особенно важные сведения о методе дают оценка скорости сходимости. Эти сведения могут быть как положительного, так и отрицательного характера. Например, оценка скорости сходимости метода Ньютона, содержащаяся в теореме 1 § 5, показывает, что метод сходится чрезвычайно быстро. Действительно, если начальное приближение достаточно близко к решению ( $q < 1$ ), то в соответствии с (6) § 5  $\|x^k - x^*\| \leq 2q^{2k}$  (так как  $l \leq L$ ). Поэтому для  $q = 0,5$  будет  $\|x^k - x^*\| \leq 2^{-2k+1}$ , так что  $\|x^5 - x^*\| < 10^{-9}$ , а для  $q = 0,1$  имеем  $\|x^k - x^*\| \leq 2 \cdot 10^{-2k}$ , так что  $\|x^4 - x^*\| < 10^{-16}$ . Иными словами, если метод Ньютона применим, то обычно требуется не более 4—5 итераций для получения решения с очень высокой точностью. С другой стороны, градиентный метод при оптимальном выборе  $\gamma$  в соответствии с теоремой 3 § 4 сходится со скоростью геометрической прогрессии со знаменателем  $q = (L-l)/(L+l)$ , причем мы видели, что эта оценка является точной для случая квадратичной функции. Для больших чисел обусловленности  $\mu = L/l$  знаменатель прогрессии  $q \approx 1 - 2/\mu$  близок к 1. Как мы видим в дальнейшем, нередко для самых простых задач среднеквадратичного приближения полиномами величина  $\mu$  достигает значений  $10^8$  и выше. Ясно, что при  $\mu = 10^8$  нужно сделать порядка 2·10<sup>8</sup> итераций, чтобы уменьшить  $\|x^0 - x^*\|$  в  $e$  раз. Иными словами, градиентный метод в такой ситуации неработоспособен. Этот отрицательный результат о поведении гра-

дентного метода удается получить чисто теоретически, не прибегая ни к каким численным экспериментам. Применительно к другим задачам минимизации он дает основания для настороженного отношения к градиентному методу — вряд ли можно рассчитывать на этот метод как эффективное средство решения сложных задач.

Теоретическая оценка скорости сходимости показывает также, от каких факторов зависит поведение метода. Так, для градиентного метода «трудные» задачи плохо обусловленные, а выбор начального приближения и близость функции к квадратической, но не обусловленностью задачи. Для метода соприженных градиентов, как мы увидим в дальнейшем, основную роль играет размерность задачи, тогда как в полученных выше оценках для методов градиентного и Ньютона размерность явно не входит. На основе этих соображений можно делать ориентировочные выводы о целесообразности применения различных методов в той или иной конкретной ситуации.

Наконец, при достаточно полной информации о задаче можно с помощью результатов о скорости сходимости заранее выбрать (или оценить) требуемое число итераций для достижения необходимой точности. Так, если мы находимся в условиях теоремы 3 § 4 и известны оценки для  $t$ ,  $L$  и  $\|x^0 - x^*\|$ , то можно указать число шагов  $k$ , гарантирующее точность  $\|x^k - x^*\| \leq \epsilon$  в градиентном методе с оптимальным  $\gamma = 2/(L+t)$ :

$$k \geq \frac{\epsilon}{\|x^0 - x^*\|} / \ln \frac{\mu - 1}{\mu + 1} \approx \frac{\mu}{2} \ln \frac{\|x^0 - x^*\|}{\epsilon}, \quad \mu = \frac{L}{t}.$$

**3. Необходима осторожность.** Прислушаемся к другой точке зрения, критикующей теоретическое исследование методов как излишнюю, а иногда и вредную роскошь.

Сторонники этой точки зрения указывают, что сам факт сходимости метода ровно ничего не говорит об эффективности последнего. Безусловно, это так. Ошибочно считать, что данный метод можно применять на практике, если его сходимость доказана — ведь скорость сходимости может быть безнадежно медленной. Однако мы уже отмечали, что теоремы сходимости (даже не содержащие оценок скорости сходимости) дают важную информацию об области применимости метода, его качественном поведении и т. п. Разумеется, вся эта информация недостаточна для окончательных выводов о целесообразности и возможности применения метода для решения конкретной задачи.

Далее, результаты о сходимости часто критикуют за неконструктивный характер. Их предположения трудно проверить,

входящие в них параметры неизвестны, оценки носят асимптотический характер и т. п. Такие обвинения во многом обоснованы. Нередко теоремы о сходимости чрезвычайно громоздки, и проверить их для какой-либо реальной задачи невозможно. Еще хуже, когда формулировки носят апостериорный характер («...пустя в процессе итераций выполняется такое-то соотношение...»). Поэтому бы тогда просто не предложил, что  $x^k \rightarrow x^*$ ? Однако не всегда ситуация столь мрачная. Как видно из теорем §§ 4 и 5, предположения в них прости и носят общий характер (требуются гладкость, выпуклость, сильная выпуклость, невырожденность и тому подобные естественные и часто легко проверяемые условия). Константы  $L$ ,  $t$  и  $q$ , входящие в формулировки теорем, обычно действительно неизвестны, поэтому конструктивный выбор  $\gamma$  в градиентном методе или явные оценки скорости сходимости невозможны. Однако существуют более сложные способы выбора  $\gamma$  в градиентном методе (гл. 3), для которых теоремы § 4 служат основой. Что же касается скорости сходимости, то хотя ее количественная оценка не всегда доступна, ее качественный характер — так, в теоремах 2, 3 § 4 и теореме 1 § 5 они верны для всех конечных  $k$ . Еще один упрек теоремам о сходимости заключается в том, что они рассматривают идеализированную ситуацию, отвлекаясь от наличия помех, ошибок округления, невозможности точного решения вспомогательных задач и т. п., а все эти факторы сильно влияют на поведение метода в реальных условиях. Действительно, в приведенных выше теоремах предполагалось, что градиент вычисляется точно, что обращение матрицы в методе Ньютона делается без погрешностей и т. д. В гл. 4 мы рассмотрим те же методы при наличии разного рода помех. Оказывается, их влияние заметно сказывается на эффективности методов. Поэтому оценки качества методов нужно делать с учетом более общих теорем о сходимости, рассчитанных на наличие помех.

Подводя итог, можно сказать, что теоретические исследования методов оптимизации могут дать много информации для читателя-практику. Нужно лишь при этом проявлять разумную осторожность и здравый смысл.

## Г л а з а 2

### ОБЩИЕ СХЕМЫ ИССЛЕДОВАНИЯ ИТЕРАТИВНЫХ МЕТОДОВ

Результаты о сходимости и скорости сходимости алгоритмов минимизации в гл. 1 были получены непосредственно, без привлечения каких-либо общих теорем. Такой подход был естествен, поскольку доказательства очень просты. Однако по мере усложнения задач и методов их обоснование становится более громоздким и трудоемким. Внимательный анализ применяемых доказательств показывает, что лежащие в их основе идеи просты и единообразны. Разумно выделить эти идеи «в явном виде», получить с их помощью ряд общих результатов о сходимости, а затем систематически использовать их при обосновании конкретных алгоритмов. Такого рода общие результаты и приводятся в данной главе.

### § 1. Первый метод Ляпунова

Идея этого подхода заключается в линеаризации итеративной процедуры, после чего вывод о сходимости удаётся сделать на основе анализа линеаризованного процесса. Предварительно приведем необходимые сведения из линейной алгебры.

**1. Сведения из линейной алгебры.** Пусть  $A$  — квадратная матрица  $n \times n$ ,  $\lambda_1, \dots, \lambda_n$  — ее собственные значения. Сспектральными радиусом  $A$  называется число

$$(1)$$

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|.$$

Другой важной характеристикой матрицы (не обязательно квадратной) является ее норма

$$(2)$$

$$\|A\| = \max_{\|x\|=1} \|Ax\|.$$

Используя тот факт, что у симметричной матрицы все собственные значения вещественны и существует полная ортогональная система собственных векторов, нетрудно доказать, что для симметричной матрицы  $\rho(A) = \|A\|$ . Для несимметричной матрицы  $\rho(A) \leq \|A\|$  и, вообще говоря,  $\rho(A) \neq \|A\|$ . Например,

$$\text{для матрицы } A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ оба собственные значения равны } 0, \text{ по-}$$

этому  $\rho(A) = 0$ , однако  $\|A\| = 1$ . Важная связь между  $\|A\|$  и  $\rho(A)$  устанавливается равенством

$$\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|^{1/k}. \quad (3)$$

Из (3) вытекает следующий фундаментальный факт.

Лемма 1. Чтобы  $\lim_{k \rightarrow \infty} A^k = 0$ , необходимо и достаточно выполнение условия  $\rho(A) < 1$ , при этом для всякого  $\varepsilon > 0$  найдется  $c = c(\varepsilon)$  такое, что  $\|A^k\| \leq c(\rho(A) + \varepsilon)^k$  для всех натуральных  $k$ .  $\blacktriangle$

Следствие. Для того чтобы итерационная последовательность векторов  $x_{k+1} = Ax_k$  сходилась к 0 при  $k \rightarrow \infty$  при любом  $x^0$ , необходимо и достаточно выполнение условия  $\rho(A) < 1$ .  $\blacktriangle$

Лемма 2. Пусть  $\rho(A) < 1$ . Тогда матричное уравнение  $ATUA = U - C$  имеет решение  $U$ , которое симметрично, если матрица  $C$  симметрична, и  $U \geq C$  при  $C \geq 0$ .

Доказательство. Поскольку  $\|A^k\| \leq cq^k$ ,  $q < 1$  (лемма 1), то ряд  $\sum_{k=0}^{\infty} (A^T)^k CA^k$  сходится к некоторой матрице  $U$ .

Эта матрица симметрична при симметричной  $C$ ,  $U \geq 0$  при  $C \geq 0$ ,  $ATUA = \sum_{k=1}^{\infty} (A^T)^k CA^k = U - C$ ,  $U = C + A^TUA \geq C$  при  $C \geq 0$ .  $\blacktriangle$

Назовем квадратную матрицу  $A$  с собственными значениями  $\lambda_1, \dots, \lambda_n$  устойчивой (урвицовой), если  $\operatorname{Re} \lambda_i < 0$ ,  $i = 1, \dots, n$ .  $\quad (5)$

Лемма 3. Для того чтобы  $\lim_{t \rightarrow \infty} e^{At} = 0$ , необходимо и достаточно, чтобы  $A$  была устойчива. При этом для всякого  $\varepsilon > 0$  найдется  $c = c(\varepsilon)$  такое, что  $\|e^{At}\| \leq c(\varepsilon)e^{(\gamma+\varepsilon)t}$  для всех  $t \geq 0$ ,  $\gamma = \max_i \operatorname{Re} \lambda_i$ .

Действительно, собственными значениями  $B = e^A$  являются  $e^{\lambda_i}$ , поэтому  $\rho(B) = \max_i e^{\operatorname{Re} \lambda_i} = e^\gamma$ . Поскольку  $e^\gamma < 1$  тогда и только тогда, когда  $\gamma < 0$ , то условие  $\rho(B) < 1$  эквивалентно условию  $\gamma < 0$ . Теперь остается воспользоваться леммой 1 (точнее, ее обобщением из упражнения 3).  $\blacktriangle$

Лемма 4 (Ляпунов). Пусть матрица  $A$  устойчива, а матрица  $C$  симметрична. Тогда уравнение  $AU + UAT = -C$   $\quad (6)$

имеет решение, причем  $U \geq 0$  ( $U \geq 0$ ), если  $C \geq 0$  ( $C \geq 0$ ).

**Доказательство.** В соответствии с леммой 3 матрица

$$U = \int_0^\infty e^{At} C e^{A^T t} dt \text{ определена. Матрица } Z(t) = e^{At} C e^{A^T t} \text{ является}$$

решением дифференциального уравнения  $\dot{Z}(t) = AZ + ZA^T$ ,

$$Z(0) = C, \text{ т. е. } U = \int_0^\infty Z(t) dt, \text{ поэтому } AU + UAT = \int_0^\infty (AZ +$$

$$+ ZA^T) dt = \int_0^\infty \dot{Z}(t) dt = -Z(0) = -C. \text{ Отсюда } U = \int_0^\infty e^{At} C e^{A^T t} dt$$

является искомым решением, и из этой же формулы следует, что  $U > 0$  ( $U \geq 0$ ) при  $C \geq 0$  ( $C \geq 0$ ).  $\blacktriangleleft$

Связь между устойчивыми матрицами и матрицами с  $\rho(A) < 1$  устанавливается следующей леммой.

**Лемма 5.** Пусть матрица  $A$  устойчива,  $B = I + \gamma A, 0 < \gamma <$

$$< \min_i (-2 \operatorname{Re} \lambda_i / |\lambda_i|^{-2}). Тогда \rho(B) < 1.$$

Действительно, если  $\lambda_i$  — собственные значения  $A$ ,  $\mu_i$  — собственные значения  $B$ , то  $\mu_i = 1 + \gamma \lambda_i$ ,  $|\mu_i|^2 = (1 + \gamma \operatorname{Re} \lambda_i)^2 + \gamma^2 (\operatorname{Im} \lambda_i)^2 = 1 + 2\gamma \operatorname{Re} \lambda_i + \gamma^2 |\lambda_i|^2 < 1$ , т. е.  $\rho(B) < 1$ .  $\blacktriangleleft$

**Упражнение.**

1. Покажите, что если матрица  $A$  симметрична или имеет попарно различные собственные значения, то в лемме 1 можно взять  $\varepsilon = 0$ ,  $c(\varepsilon) = 1$ .

2. Приведите пример матрицы  $A$  с  $\rho(A) \geq 1$  и некоторого  $x^0 \neq 0$ , для которых  $A x^k \rightarrow 0$  при  $k \rightarrow \infty$ .

3. Покажите, что лемма 1 справедлива и для неделых показателей, т. е.  $\|A^k\| \leq c(\varepsilon) (\rho(A) + \varepsilon)^k$  для всех вещественных  $t \geq 0$ .

## 2. Теоремы о линейной сходимости.

Мы будем часто употреблять термин *линейная сходимость* как синоним сходимости со скоростью геометрической прогрессии. Аналогично *сверхлинейная сходимость* означает сходимость более быструю, чем определяемую любой геометрической прогрессией. Наконец, термин *квадратичная сходимость* используется для процессов, в которых справедлива оценка вида  $u_{k+1} \leq c u_k^2$ , где  $u_k$  — некоторая мера близости к решению на  $k$ -й итерации.

Рассмотрим итерационный процесс вида

$$x^{k+1} = g(x^k), \quad (7)$$

где  $g$  — некоторое отображение из  $\mathbb{R}^n$  в  $\mathbb{R}^n$ . Точку  $x^*$  будем называть *неподвижной точкой* (7), если  $x^* = g(x^*)$ . В этом случае при  $x^k = x^*$  будет  $x^s \equiv x^*$  для всех  $s \geq k$ .

**Теорема 1.** Пусть  $x^*$  — неподвижная точка (7),  $g(x)$  дифференцируема в  $x^*$  и спектральный радиус матрицы Якоби

$\frac{g'(x^*)}{g(x)}$  удовлетворяет условию  $\rho = \rho(g'(x^*)) < 1$ . Тогда процесс (7) локально линейно сходится к  $x^*$ , а именно, для всякого  $0 < \varepsilon < 1 - \rho$  находится  $\delta > 0$  и с такие, что для всех  $k \geq 0$  будет

$$\|x^k - x^*\| \leq c(\rho + \varepsilon)^k \quad (8)$$

Дадим краткую схему доказательства. Обозначим  $A = g'(x^*)$ , тогда в соответствии с определением производной  $g(x) = g(x^*) + A(x - x^*) + o(x - x^*)$ . Поэтому процесс (7) может быть записан в виде

$$z^{k+1} = Az^k + y^k, \quad z^k = x^k - x^*, \quad y^k = o(z^k).$$

Отсюда

$$z^{k+1} = A^{k+1} z^0 + \sum_{i=0}^k A^{k-i} y^i,$$

$$\|z^{k+1}\| \leq \|A^{k+1}\| \|z^0\| + \sum_{i=0}^k \|A^{k-i}\| \|y^i\|. \quad (9)$$

Из леммы 1  $\|A^k\| \leq c(\varepsilon)(\rho + \varepsilon)^k$ , подставляя эту оценку в (9) и используя тот факт, что  $\|y^k\| = o(z^k)$ , можно получить утверждение теоремы.  $\blacktriangleleft$

Теорема 1 гарантирует локальную сходимость метода (7). В некоторых случаях можно утверждать и глобальную сходимость. Один из таких случаев очевиден — это случай линейной функции  $g(x)$ . Приведем результат о глобальной сходимости и для нелинейных функций. При этом нам удобнее будет рассматривать итерационный процесс, заданный в виде

$$x^{k+1} = x^k - \gamma(Ax^k + \Phi(x^k)). \quad (10)$$

Теорема 2. Пусть матрица  $-A$  устойчива, а  $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$  удовлетворяет условию  $\|\Phi(x)\| \leq L \|x\|$ .

Тогда, если

$$L < \frac{1}{2\|U\|}, \quad 0 < \gamma < \frac{\|U\|^{-1} - 2L}{(L + \|A\|)^2}, \quad (11)$$

где  $U$  — решение матричного уравнения

$$UA + A^T U = I, \quad (12)$$

то процесс (10) сходит к 0 со скоростью геометрической прогрессии при любом  $x^0$ :

$$\|x^k\| \leq \|x^0\|^2 \|U^{-1}\| \|U\| q^k,$$

$$q = 1 - (1/2)\gamma \|U\|^{-1} + \gamma L + (1/2)\gamma^2 (\|A\| + L)^2. \quad (13)$$

Для доказательства достаточно ввести  $u_k = (Ux_k, x_k)$  и по-

лучить соотношение  $u_{k+1} \leq q u_k$ .  $\blacktriangle$

Полученные выше результаты можно применить для иссле-

дования *уравнения в конечных разностях*

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + \dots + a_n y_{k-n} + \varphi(y_{k-1}, \dots, y_{k-n}), \quad (14)$$

где  $y_i \in \mathbf{R}^1$ . Для этого введем векторы  $x^k = (y_{k-1}, \dots, y_{k-n}) \in \mathbf{R}^n$ ,  $x^{k+1} = (y_k, y_{k-1}, \dots, y_{k-n+1}) \in \mathbf{R}^n$ , тогда  $x^{k+1} = Ax^k + h(x^k)$ , где

$$A = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 1 \end{pmatrix}, \quad h(x) = \begin{pmatrix} \varphi(x) \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (15)$$

Таким образом, итерационный процесс приведен к форме (7),

описанный прием типичен при исследовании многошаговых итеративных процессов, в которых каждое приближение зависит от нескольких предыдущих. Тогда увеличение размерности задачи позволяет свести ее к одноступенчатому процессу.

**Упражнение.** Докажите, что если все корни характеристического уравнения  $\lambda^n = a\lambda^{n-1} + \dots + a_n$  по модулю меньше 1, то для матрицы  $A$  вида (15) будет  $\rho(A) < 1$ .

**3. Теорема о сверхлинейной сходимости.** В случае, когда  $g'(x^*) = 0$ , из теоремы 1 следует, что метод (7) сходится быстрее любой геометрической прогрессии. Этот результат можно уточнить.

**Теорема 3.** Пусть  $x^*$  — неподвижная точка (7),  $g(x)$  дифференцируема в  $S = \{x: \|x - x^*\| \leq \|x^0 - x^*\|\}$ ,  $g'(x)$  непрерывна в  $S$  условие Липшица и  $g'(x^*) = 0$ . Тогда, если

$$q = (L/2) \|x^0 - x^*\| < 1, \quad (16)$$

то

$$\|x^k - x^*\| \leq (2/L) q^{2^k}. \quad (17)$$

**Доказательство.** Очевидно,  $x^0 \in S$ . В силу формулы (15) § 1 гл. 1  $\|x^1 - x^*\| = \|g(x^0) - g(x^*) - g'(x^*)(x^0 - x^*)\| \leq (L/2) \|x^0 - x^*\| \leq q \|x^0 - x^*\|$ , поэтому  $x^1 \in S$ . Аналогичным образом  $x^k \in S$  для всех  $k$ . Поэтому мы имеем право пользоваться той же оценкой:

$$\|x^{k+1} - x^*\| = \|g(x^k) - g(x^*) - g'(x^*)(x^k - x^*)\| \leq (L/2) \|x^k - x^*\|^2.$$

Отсюда следует требуемый результат.  $\blacktriangle$

**Упражнение.** Пусть  $x^*$  — невырожденная точка минимума  $f(x)$ ,  $\nabla f(x)$  удовлетворяет условию Липшица в окрестности  $x^*$ . Тогда метод

$$x^{k+1} = x^k - [\nabla^2 f(x^*)]^{-1} \nabla f(x^k) \quad (18)$$

локально сходится к  $x^*$  с квадратичной скоростью. Докажите этот результат с помощью теоремы 3.

## § 2. Второй метод Ляпунова

Этот метод является наиболее распространенным при обосновании сходимости итерационных процессов. Его идея заключается в том, что вводится некоторая скалярная неотрицательная функция  $V(x)$  (*функция Ляпунова*) и рассматривается ее значения на последовательных итерациях  $x^k$ . Если они monotonно убывают и ограничены снизу, то  $V(x^k) - V(x^{k+1}) \rightarrow 0$ . Отсюда при некоторых дополнительных предположениях следует сходимость метода.

Если посмотреть с этой точки зрения на приведенные выше результаты о сходимости, то окажется, что большинство из них получено именно по такой схеме. Так, при обосновании градиентного метода в гл. 1 в качестве функции Ляпунова в теоремах 1, 2 § 4 выступала сама минимизируемая функция  $f(x) - f^*$ . В теоремах 3, 4 § 4 такую роль играло расстояние до точки минимума. При обосновании метода Ньютона (теорема 1 § 5) использовалось монотонное убывание нормы градиента (т. е. неявные в выполнении условия экстремума). Наконец, в теореме 2 § 1 данной главы в доказательстве была построена специальная квадратичная функция Ляпунова. Эти же приемы выбора функции Ляпунова обычно применяются и для других, более сложных задач.

**1. Леммы о числовых последовательностях.** Для значений функции Ляпунова  $u_k = V(x^k)$  на  $k$ -м шаге процесса обычно получается итерационное соотношение вида

$$u_{k+1} \leq \Phi_k(u_k). \quad (1)$$

Отсюда делается вывод, что  $u_k \rightarrow 0$  и дается оценка скорости сходимости  $u_k$ . Поэтому важно исследовать поведение последовательностей вида (1) для нескольких «типовых» функций  $\Phi_k$ . С некоторыми простейшими соотношениями (1) мы уже столкнулись. Так, при доказательстве сходимости градиентного метода (§ 4 гл. 1) мы получали неравенство

$$u_{k+1} \leq q u_k, \quad 0 \leq q < 1 \quad (2)$$

(где  $u_k = f(x^k) - f^*$ , либо  $u_k = \|x^k - x^*\|^2$ , либо  $u_k = \|\nabla f(x^k)\|$ ). Из (2) следует оценка  $u_k \leq u_0 q^k$ . При обосновании метода

Ньютона (§ 5 гл. 1) было получено соотношение для  $u_k = \|\nabla f(x^k)\|$ :

$$u_{k+1} \leq c u_k^2, \quad c > 0. \quad (3)$$

Отсюда  $u_k \leq c^{-1} (c u_0)^{2^k}$  и, если  $c u_0 < 1$ , то  $u_k \rightarrow 0$ .

В других задачах, однако, соотношение (1) имеет более сложный вид, и его анализ не столь тривиален.

Начнем с линейных неравенств вида

$$u_{k+1} \leq q_k u_k + \alpha_k, \quad q_k \geq 0.$$

Отсюда

$$u_k \leq q_{k-1} q_{k-2} \dots q_0 u_0 + q_{k-1} \dots q_1 \alpha_0 + \dots + q_{k-1} \alpha_{k-2} + \alpha_{k-1}. \quad (5)$$

Рассмотрим несколько частных случаев.

Лемма 1. *Пусть*

$$u_{k+1} \leq q u_k + \alpha, \quad 0 \leq q < 1, \quad \alpha > 0.$$

Тогда

$$u_k \leq \alpha/(1-q) + (u_0 - \alpha/(1-q)) q^k. \quad (6)$$

Доказательство. Обозначая  $v_k = u_k - \alpha/(1-q)$ , из неравенства (6) получаем  $v_{k+1} \leq v_k q$ , что и дает (7).  $\blacktriangle$

Таким образом,  $u_k$  сходится в область  $u \leq \alpha/(1-q)$  со скоростью геометрической прогрессии со знаменателем  $q$ .

Лемма 2. Пусть  $u_k \geq 0$

$$\begin{aligned} u_{k+1} &\leq (1 + \alpha_k) u_k + \beta_k, \quad \alpha_k \geq 0, \quad \beta_k \geq 0, \\ \sum_{k=0}^{\infty} \alpha_k &< \infty, \quad \sum_{k=0}^{\infty} \beta_k < \infty. \end{aligned} \quad (8)$$

Тогда  $u_k \rightarrow u \geq 0$ .

Доказательство совпадает с приводимым ниже доказательством более общей леммы 9.  $\blacktriangle$

Лемма 3. *Пусть*

$$\begin{aligned} u_{k+1} &\leq q_k u_k + \alpha_k, \quad 0 \leq q_k < 1, \quad \alpha_k \geq 0, \\ \sum_{k=0}^{\infty} (1 - q_k) &= \infty, \quad \alpha_k/(1 - q_k) \rightarrow 0. \end{aligned} \quad (9)$$

Тогда  $\overline{\lim}_{k \rightarrow \infty} u_k \leq 0$ . В частности, если  $u_k \geq 0$ , то  $u_k \rightarrow 0$ .  $\blacktriangle$

Следствие. Если в (9)  $q_k = q < 1$ ,  $\alpha_k \rightarrow 0$ ,  $u_k \geq 0$ , то  $u_k \rightarrow 0$ .  $\blacktriangle$

В условиях леммы 3 можно для ряда случаев оценить скорость сходимости,

Лемма 4 (Чжун). Пусть  $u_k \geq 0$  и

$$u_{k+1} \leq \left(1 - \frac{c}{k}\right) u_k + \frac{d}{k^{p+1}}, \quad d > 0, \quad p > 0, \quad c > 0. \quad (10)$$

Тогда

$$u_k \leq d(c-p)^{-1} k^{-p} + o(k^{-p}) \quad при \quad c > p, \quad (11)$$

$$u_k = O(k^{-c} \ln k) \quad при \quad p = c, \quad (12)$$

$$u_k = O(k^{-c}) \quad при \quad p > c. \quad (13)$$

Доказательство. При любом соотношении  $c$  и  $p$  мы находимся в условиях применения леммы 3, так как  $1 - q_k = c/k$ ,  $\sum_{k=0}^{\infty} (1 - q_k) = \infty$ ,  $\alpha_k (1 - q_k)^{-1} = dc^{-1} k^{-p} \rightarrow 0$ , поэтому

$u_k \rightarrow 0$ . Пусть  $c > p$ . Введем  $v_k = k^p u_k - d(c-p)^{-1}$ . Тогда

$$v_{k+1} = (k+1)^p u_{k+1} - \frac{d}{c-p} \leq k^p \left(1 + \frac{1}{k}\right)^p \left(\left(1 - \frac{c}{k}\right) u_k + \frac{d}{k^{p+1}}\right) -$$

$$\begin{aligned} &- \frac{d}{c-p} = k^p u_k \left(1 - \frac{c-p}{k} + o\left(\frac{1}{k}\right)\right) + \\ &+ \frac{d}{k} \left(1 + \frac{p}{k} + o\left(\frac{1}{k}\right)\right) - \frac{d}{c-p} = \\ &= \left(v_k + \frac{d}{c-p}\right) \left(1 - \frac{c-p}{k} + o\left(\frac{1}{k}\right)\right) + \\ &+ \frac{d}{k} \left(1 + \frac{p}{k} + o\left(\frac{1}{k}\right)\right) - \frac{d}{c-p} = \\ &= v_k \left(1 - \frac{c-p}{k} + o\left(\frac{1}{k}\right)\right) + \frac{dp}{k^2} + o\left(\frac{1}{k^2}\right). \end{aligned}$$

Применяя лемму 3, получаем  $\overline{\lim}_{k \rightarrow \infty} v_k \leq 0$ , что и доказывает (11).

Пусть теперь  $p \geq c$ . Введем  $v_k = u_k k^c$ . Тогда

$$\begin{aligned} v_{k+1} &= u_{k+1} (k+1)^c \leq \\ &\leq \left[\left(1 - \frac{c}{k}\right) u_k + \frac{d}{k^{p+1}}\right] \cdot k^c \left(1 + \frac{c}{k} + \frac{c^2}{2k^2} + o\left(\frac{1}{k^2}\right)\right) = \\ &= \left(1 - \frac{c^2}{2k^2} + o\left(\frac{1}{k^2}\right)\right) v_k + \frac{d}{k^{p-c+1}} \left(1 + o\left(\frac{1}{k}\right)\right) \leq v_k + \frac{d'}{k^{p-c+1}} \end{aligned}$$

для достаточно больших  $k$ . Суммируя по  $k$ , получаем, что  $v_k$  ограничено при  $p > c$  (так как ряд  $\sum_{k=1}^{\infty} \frac{1}{k^a}$  сходится при  $a > 1$ )

и  $v_k = O(\ln k)$  при  $p = c$  (так как  $\sum_{i=1}^k \frac{1}{i^a} = O(\ln k)$ ). Это доказывает (12) и (13).  $\blacktriangle$

**Лемма 5** (Чжун). Пусть  $u_k \geq 0$  и

$$u_{k+1} \leq \left(1 - \frac{c}{k^s}\right) u_k + \frac{d}{k^t}, \quad 0 < s < 1, \quad s < t. \quad (14)$$

Тогда  $u_k \leq \frac{d}{c} \frac{1}{k^{t-s}} + o\left(\frac{1}{k^{t-s}}\right)$ .  $\blacktriangle$

Перейдем к исследованию рекуррентных неравенств, задаваемых нелинейными соотношениями.

**Лемма 6.** Пусть  $u_k > 0$  и

$$u_{k+1} \leq u_k - \alpha_k u_k^{1+p}, \quad \alpha_k \geq 0, \quad p > 0.$$

Тогда

$$u_k \leq u_0 \left(1 + pu_0^p \sum_{i=0}^{k-1} \alpha_i\right)^{-1/p}. \quad (15)$$

В частности, если  $\alpha_k \equiv \alpha$ ,  $p = 1$ , то

$$u_k \leq u_0 / (1 + \alpha k u_0). \quad (17)$$

Докажем (16) лишь для случая  $p \geq 1$ . Разделим обе части (15) на  $u_k^p u_{k+1}$ :

$$u_k^{-p} \leq u_k^{1-p} u_{k+1}^{-1} - \alpha_k u_k u_{k+1}^{-1}.$$

Поскольку  $u_{k+1} \leq u_k$ ,  $u_k^{1-p} \leq u_{k+1}^{1-p}$  при  $p \geq 1$ , то получаем

$u_k^{-p} \leq u_{k+1}^{-p} - \alpha_k$ . Суммируя неравенства, приходим к (16).  $\blacktriangle$

**2. Леммы о случайных последовательностях.** При исследовании итеративных методов, включающих элементы случайности (методы случайного поиска, задачи с помехами), обычно меняется та же техника, основанная на функциях Ляпунова. Однако здесь значения функции Ляпунова оказываются случайной величиной, поэтому нужно получить аналоги приведенных выше лемм для случайных последовательностей.

Напомним различные виды сходимости случайных величин. Пусть  $v^1, \dots, v^k, \dots$  — последовательность  $n$ -мерных случайных векторов. Мы обычно не будем выписывать то вероятностное пространство  $(\Omega, \mathfrak{F}, \mathbf{P})$ , на котором заданы эти величины (т. е. не будем писать  $v^1(\omega), \dots, v^k(\omega)$ ,  $\omega \in \Omega$ ,  $\Omega$  — пространство элементарных событий,  $\mathfrak{F}$  — заданная на нем  $\sigma$ -алгебра измеримых множеств,  $\mathbf{P}$  — вероятностная мера на  $\mathfrak{F}$ ). Говорят, что последовательность  $v^k$  сходится к случайному вектору  $v$ :

a) *почти наверное* (с вероятностью 1), если  $\mathbf{P}(\lim_{k \rightarrow \infty} v^k = v) = 1$  (здесь и далее  $\mathbf{P}(A)$  обозначает вероятность события  $A$ ), при этом пишут  $v^k \rightarrow v$  (п. н.).  
б) *по вероятности*, если для каждого  $\varepsilon > 0$   $\lim_{k \rightarrow \infty} \mathbf{P}(\|v^k - v\| > \varepsilon) = 0$ , что обозначается  $v^k \xrightarrow{P} v$ .

**в) в среднем квадратичном**, если  $\lim_{k \rightarrow \infty} \mathbf{M} \|v^k - v\|^2 = 0$  (здесь и далее  $\mathbf{M}x$  обозначает математическое ожидание случайной величины  $x$ ).

Основным инструментом при изучении сходимости случайных величин является теория полумартингалов. Последовательность скалярных случайных величин  $v_0, \dots, v_k, \dots$  называется *полумартингалом*, если  $\mathbf{M}(v_{k+1} | v_0, \dots, v_k) \leq v_k$ ,  $\mathbf{M}v_0 < \infty$ . Здесь  $\mathbf{M}(v_{k+1} | v_0, \dots, v_k)$  — условное математическое ожидание  $v_{k+1}$  при данных  $v_0, \dots, v_k$ . Часто в данном случае употребляют также термин *супермартингал*, для неравенства противоположного знака говорят о *субмартингале*, а для равенства — о *martингале*. Полумартингал является обобщением на стохастический случай понятия монотонно убывающей последовательности.

Ключевой результат о сходимости числовых последовательностей (ограниченная снизу монотонно убывающей последовательность имеет предел) для случайных величин приобретает следующий вид.

**Лемма 7.** Пусть  $v_0, \dots, v_k, \dots$  — полумартингал, причем  $v_k \geq 0$  для всех  $k$ . Тогда существует случайная величина  $v \geq 0$ ,  $v_k \rightarrow v$  (п. н.).  $\blacktriangle$

Известное неравенство Чебышева (если  $v \geq 0$ ,  $\varepsilon > 0$ ,  $\mathbf{M}v < \infty$ , то  $\mathbf{P}(v \geq \varepsilon) \leq \varepsilon^{-1} \mathbf{M}v$ ) для полумартингалов может быть усилено.

**Лемма 8** (неравенство Колмогорова). Пусть  $v_0, \dots, v_k, \dots$  — полумаргингал,  $v_k \geq 0$ ,  $\varepsilon > 0$ . Тогда

$$\mathbf{P}(v_k \geq \varepsilon \ \forall k) \leq \varepsilon^{-1} \mathbf{M}v_0. \quad \blacktriangle$$

Используя эти результаты, получим стохастические аналоги лемм 2 и 3.

**Лемма 9** (Гладышев). Пусть имеется последовательность случайных величин  $v_0, \dots, v_k \geq 0$ ,  $\mathbf{M}v_0 < \infty$  и

$$\mathbf{M}(v_{k+1} | v_0, \dots, v_k) \leq (1 + \alpha_k) v_k + \beta_k,$$

$$\sum_{k=0}^{\infty} \alpha_k < \infty, \quad \sum_{k=0}^{\infty} \beta_k < \infty, \quad \alpha_k \geq 0, \quad \beta_k \geq 0. \quad (19)$$

Тогда  $v_k \rightarrow v$  (п. н.), где  $v \geq 0$  — некоторая случайная величина.

Доказательство. Введем  $u_k = \prod_{i=k}^{\infty} (1 + \alpha_i) v_i + \sum_{i=k}^{\infty} \beta_i \times \prod_{j=i+1}^{\infty} (1 + \alpha_j)$ . Тогда  $u_k \geq 0$ ,  $\mathbf{M}u_0 < \infty$  (так как  $\prod_{i=0}^{\infty} (1 + \alpha_i) < \infty$ ,

$$\sum_{k=0}^{\infty} \beta_k < \infty, \quad \mathbf{M}v_0 < \infty$$

). При этом

$$\mathbf{M}(u_{k+1} | u_0, \dots, u_k) =$$

$$= \prod_{i=k+1}^{\infty} (1 + \alpha_i) \mathbf{M}(v_{k+1} | v_0, \dots, v_k) + \sum_{i=k+1}^{\infty} \beta_i \prod_{j=i+1}^{\infty} (1 + \alpha_j) \leq$$

$$\leq \prod_{i=k}^{\infty} (1 + \alpha_i) \sigma + \sum_{i=k}^{\infty} \beta_i \prod_{j=i+1}^{\infty} (1 + \alpha_j) = u_k,$$

т. е.  $u_k$  — полумартингал, и по лемме 7  $u_k \rightarrow v$  (п. н.),  $v \geq 0$ .

Поэтому и  $v_k = \left( u_k - \sum_{i=k}^{\infty} \beta_i \right) / \prod_{i=k}^{\infty} (1 + \alpha_i) \rightarrow v$  (п. н.).  $\blacktriangle$

**Лемма 10.** Пусть  $v_0, \dots, v_k$  — последовательность случайных величин,  $v_k \geq 0$ ,  $Mv_0 < \infty$  и

$$\mathbf{M}(v_{k+1} | v_0, \dots, v_k) \leq (1 - \alpha_k) v_k + \beta_k.$$

$$0 \leq a_k \leq 1, \quad \beta_k \geq 0, \quad \sum_{k=0}^{\infty} a_k = \infty, \quad \sum_{k=0}^{\infty} \beta_k < \infty, \quad \frac{\beta_k}{\alpha_k} \rightarrow 0. \quad (20)$$

Тогда  $v_k \rightarrow 0$  (п. н.),  $Mv_k \rightarrow 0$ , причем для всякого  $\varepsilon > 0$ ,  $k > 0$

$$\mathbf{P}(v_j \leq \varepsilon \text{ для всех } j \geq k) \geq 1 - \varepsilon^{-1} \left( Mv_k + \sum_{i=k}^{\infty} \beta_i \right). \quad (22)$$

**Доказательство.** Беря безусловное математическое ожидание от обеих частей (20), получаем  $Mv_{k+1} \leq (1 - \alpha_k) Mv_k + \beta_k$ . Отсюда по лемме 3  $Mv_k \rightarrow 0$ . С другой стороны,  $u_k = v_k + \sum_{i=k}^{\infty} \beta_i$  — полумартингал (ср. с доказательством леммы 9).

Используя леммы 8 и 9, получаем требуемый результат.  $\blacktriangle$

**3. Основные теоремы.** Рассматривается итеративный процесс

$$x^{k+1} = x^k - \gamma_k s^k, \quad (23)$$

где  $k$  — номер итерации,  $x^k, s^k$  — векторы в  $\mathbf{R}^n$ ,  $\gamma_k \geq 0$  — скалярный множитель, характеризующий длину шага. Мы объединим детерминированный и стохастический случаи — будет рассматриваться общая ситуация, когда  $x^k$  и  $s^k$  случайны, а детерминированный процесс включается в нее как частный случай. Основные предположения о процессе заключаются в следующем. Зависит только от  $x^k$  и  $k$ ,  $s^k = s^k(x^k)$ , величины  $s^k, s^{k-1}, \dots$  взаимно независимы.

Б. Существует скалярная функция (*функция Липшица*)

$$V(x) \geq 0, \quad \inf_{x \in \mathbf{R}^n} V(x) = 0, \quad V(x) \text{ дифференцируема и } \nabla V(x) \text{ удо-}$$

вливает условию Липшица с константой  $L$ .

В. Процесс (23) является псевдоградиентным по отношению к  $V(x)$ :

$$(\nabla V(x^k), \mathbf{M}(s^k | x^k)) \geq 0, \quad (24)$$

т. е.  $-s^k$  в среднем является направлением убывания  $V(x)$  в точке  $x^k$ .

Г. Выполняется следующее условие роста на  $s^k$ :

$$\mathbf{M}(\|s^k\|^2 | x^k) \leq \sigma^2 + \tau (\nabla V(x^k), \mathbf{M}(s^k | x^k)). \quad (25)$$

Величина  $\sigma^2$  обычно характеризует уровень аддитивных помех. Случай  $\sigma = 0$  типичен для детерминированных задач.

Д. Начальное приближение удовлетворяет условию

$$\mathbf{M}V(x^0) < \infty. \quad (26)$$

Разумеется, это условие выполняется, если  $x^0$  — детерминированный вектор.

Е. Длина шага такова, что

$$\gamma_k \geq 0, \quad \sum_{k=0}^{\infty} \gamma_k = \infty, \quad \lim_{k \rightarrow \infty} \gamma_k < \frac{2}{L\tau}. \quad (27)$$

Приведем основные теоремы о сходимости. При условиях А—Е нельзя, вообще говоря, утверждать, что  $V(x^k) \rightarrow 0$  для процесса (23) в каком-либо вероятностном смысле. Например, если  $s^k \equiv 0$ , то все условия выполняются, но  $x^k \equiv x^0$ . Однако некоторые утверждения о сходимости справедливы даже при этих минимальных предположениях.

Теорема 1. Пусть выполнены условия А—Е и либо  $\sigma^2 = 0$ ,

либо  $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$ . Тогда при любом  $x^0$  алгоритм (23)

$$V(x^k) \rightarrow V(n, n), \quad \lim_{k \rightarrow \infty} (\nabla V(x^k), \mathbf{M}(s^k | x^k)) = 0 \text{ (п. н.)}. \quad (28)$$

Доказательство. Используя условие Б и формулу (15), § 1 гл. 1, получаем

$$V(x^{k+1}) \leq V(x^k) - \gamma_k (\nabla V(x^k), s^k) + L \gamma_k^2 \|s^k\|^2/2.$$

Возьмем условное математическое ожидание обеих частей этого неравенства и применим условие Г:

$$\begin{aligned} \mathbf{M}(V(x^{k+1}) | x^k) &\leq \\ &\leq V(x^k) - \gamma_k (\nabla V(x^k), \mathbf{M}(s^k | x^k)) + L \gamma_k^2 \mathbf{M}(\|s^k\|^2 | x^k)/2 \leq \\ &\leq V(x^k) - \gamma_k (1 - (1/2)L\tau\gamma_k) (\nabla V(x^k), \mathbf{M}(s^k | x^k)) + L \gamma_k^2 \sigma^2/2. \end{aligned} \quad (29)$$

В силу условий В и Г

$$\mathbf{M}(V(x^{k+1})|x^k) \leq V(x^k) + L\gamma_k^2\sigma^2/2. \quad (30)$$

Применяя лемму 9, получаем, что  $V(x^k) \rightarrow V$  (п. н.). Перейдем в (29) к безусловным математическим ожиданиям:

$$\mathbf{M}V(x^{k+1}) \leq \mathbf{M}V(x^k) - \gamma_k(1 - (1/2)L\tau\gamma_k)u_k + L\gamma_k^2\sigma^2/2,$$

$$u_k = \mathbf{M}(\nabla V(x^k)), \mathbf{M}(s^k|x^k)).$$

Для достаточно больших  $k$ , в силу условия Е,  $1 - (1/2)L\tau\gamma_k \geq \varepsilon > 0$ , т. е.

$$\mathbf{M}V(x^{k+1}) \leq \mathbf{M}V(x^k) - \gamma_k\varepsilon u_k + L\gamma_k^2\sigma^2/2.$$

Поскольку  $\mathbf{M}V(x^0) < \infty$  (условие Д) и  $\sigma^2 \sum_{k=0}^{\infty} \gamma_k^2 < \infty$ , то отсюда

следует, что  $\sum_{k=0}^{\infty} \gamma_k u_k < \infty$ . Но так как  $\sum_{k=0}^{\infty} \gamma_k = \infty$ , то это означает, что  $\lim_{k \rightarrow \infty} u_k = 0$ . Из свойств сходимости в среднем следует, что если для случайных величин  $z^k \geq 0$ ,  $\mathbf{M}z^k \rightarrow 0$ , то найдется подпоследовательность  $z^{k_l} \rightarrow 0$  (п. н.). Поэтому

$$\lim_{k \rightarrow \infty} (\nabla V(x^k), \mathbf{M}(s^k|x^k)) = 0 \text{ (п. н.)}. \blacktriangle$$

Заменим условие В на условие В' сильной псевдоградиентности:

$$\mathbf{B}', (\nabla V(x^k), \mathbf{M}(s^k|x^k)) \geq lV(x^k), \quad l > 0.$$

**Теорема 2.** Пусть выполнены условия А — Е и В' и либо

$\sigma^2 = 0$ , либо  $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$ . Тогда при любом  $x^0$  в алгоритме (23)  $V(x^k) \rightarrow 0$  (п. н.),

$$\mathbf{P}(V(x^t) \leq \varepsilon \quad \forall t \geq k) \geq 1 - \varepsilon^{-1} \left( \mathbf{M}V(x^k) + \frac{1}{2}L\sigma^2 \sum_{i=k}^{\infty} \gamma_i^2 \right). \quad (31)$$

**Доказательство.** Из (29) и условия В' получаем

$$\mathbf{M}(V(x^{k+1})|x^k) \leq (1 - l\gamma_k(1 - (1/2)L\tau\gamma_k))V(x^k) + L\gamma_k^2\sigma^2/2. \quad (32)$$

Из леммы 10 и условия Е следует требуемый результат.  $\blacktriangle$

Перейдем к условиям сходимости в среднем.

**Теорема 3.** Пусть выполнены условия А — Е, В' и либо  $\sigma^2 = 0$ , либо  $\gamma_k \rightarrow 0$ . Тогда в алгоритме (23)

$$\mathbf{M}V(x^k) \rightarrow 0.$$

**Доказательство.** Беря безусловное математическое ожидание в (32), имеем

$$\mathbf{M}V(x^{k+1}) \leq (1 - l\gamma_k(1 - (1/2)L\tau\gamma_k))\mathbf{M}V(x^k) + L\gamma_k^2\sigma^2/2. \quad (34)$$

Поскольку  $1 - (1/2)L\tau\gamma_k \geq \varepsilon > 0$  для достаточно больших  $k$ , то

$$\mathbf{M}V(x^{k+1}) \leq (1 - l\gamma_k)\mathbf{M}V(x^k) + L\gamma_k^2\sigma^2/2.$$

По лемме 3  $\mathbf{M}V(x^k) \rightarrow 0$ .  $\blacktriangle$

Из неравенства (34) можно получать и другие результаты, в том числе оценки скорости сходимости. Приведем несколько примеров.

**Теорема 4.** Пусть выполнены условия А — Е, В' и  $\gamma_k \equiv \gamma$ ,  $0 < \gamma < 2/(L\tau)$ . Тогда

$$\mathbf{M}V(x^k) \leq \mathbf{M}V(x^0)q^k + \frac{Ly\sigma^2}{l(2 - L\tau)}(1 - q^k), \quad (35)$$

$$q = 1 - l\gamma(1 - (1/2)L\tau\gamma).$$

Этот результат следует из (34) и леммы 1.  $\blacktriangle$

Таким образом, если  $\sigma^2 > 0$ , то  $\lim_{k \rightarrow \infty} \mathbf{M}V(x^k) \leq Ly\sigma^2/l(2 - L\tau\gamma)$ , если же  $\sigma^2 = 0$ , то  $\mathbf{M}V(x^k)$  стремится к 0 со скоростью геометрической прогрессии.

**Теорема 5.** Пусть выполнены условия А — Е, В',  $\sigma^2 \geq 0$  и  $\gamma_k = \gamma/k$ . Тогда

$$\mathbf{M}V(x^k) = \begin{cases} O(1/k) & \text{при } l\gamma > 1, \\ O(1/k^2) & \text{при } l\gamma < 1. \end{cases} \quad (36)$$

Этот результат легко можно получить из (34) и леммы 4.  $\blacktriangle$

1. В качестве следствия теоремы 1 получите теорему 1 § 4 гл. 1, взяв  $V(x) = f(x) - f^*$ .

2. Примите теорему 4 для доказательства теорем 2, 3 § 4 гл. 1, беря

$$V(x) = f(x) - f^*, \text{ или } V(x) = \|x - x^*\|^2.$$

**4. Возможные модификации.** Приведенные теоремы о сходимости отнюдь не являются самыми общими и охватывающими все случаи. Они могут быть видоизменены в различных направлениях.

Во-первых, условия В, В' и Г могут быть обобщены следующим образом:

$$(\nabla V(x^k), \mathbf{M}(s^k|x^k)) \geq l_k V(x^k) - \beta_k, \quad (37)$$

$$\mathbf{M}(\|s^k\|^2|x^k) \leq \sigma_k^2 + \tau_k(\nabla V(x^k), \mathbf{M}(s^k|x^k)) + \mu_k V(x^k). \quad (38)$$

При определенных условиях на  $l_k$ ,  $\beta_k$ ,  $\sigma_k$ ,  $\tau_k$  и  $\mu_k$  можно с помощью лемм данного параграфа доказать аналоги теорем 1—3.

Такого рода ситуации, когда выполняются условия (37) и (38), встречаются нам далее при изучении конечно-разностных вариантов градиентного метода, методов регуляризации и т. д.

Во-вторых, все приведенные до сих пор результаты носили глобальный характер — предполагалось, что условия на  $V(x)$ ,  $s^k(x)$  и т. д. выполняются для всех  $x$ , а начальное приближение  $x^0$  могло быть любым. Однако нередко такого рода предположения выполняются лишь локально, в окрестности градиента. Естественно, что при этом и утверждения о сходимости должны носить локальный характер. Примерами могут служить теоремы 4 § 4 и 1 § 5 гл. 1 о локальной сходимости градиентного метода и метода Ньютона. Наличие случайных помех внеходит некоторые осложнения — возникает неизвестная вероятность выхода из области, в которой выполнены предположения. Помимо локальных утверждений о сходимости могут выполняться лишь с некоторой вероятностью  $1 - \delta$ ,  $\delta > 0$ . Приведем соответствующий аналог теоремы 2. Пусть

$$Q = \{x: V(x) \leq \varepsilon\},$$

где  $\varepsilon > 0$  — некоторое число.

**Теорема 6.** Пусть условия А — Е, В' выполнены для всех  $x$ ,  $x^k \in Q$ . Тогда для метода (23):

а) если  $x^0$  детерминировано,  $x^0 \in Q$ ,  $\sigma^2 = 0$ ,  $s^k$  детерминировано, то  $V(x^k) \rightarrow 0$ ;

б) если  $\sum_{k=0}^{\infty} v_k^2 < \infty$ , то

$$\mathbf{P}(x^k \in Q \forall k) \geq 1 - \delta, \quad \mathbf{P}(V(x^k) \rightarrow 0) \geq 1 - \delta,$$

$$\delta = \varepsilon^{-1} M V(x^0) + \frac{1}{2} L \sigma^2 \varepsilon^{-1} \sum_{k=0}^{\infty} v_k^2. \blacksquare \quad (39)$$

Далее, можно рассматривать *непрерывные аналоги итеративных методов* — процессы, описываемые обыкновенными дифференциальными уравнениями

$$dx/dt = s(x, t), \quad x(0) = x^0. \quad (40)$$

Для них можно применить ту же технику, основанную на функции Ляпунова. При этом формулировки многих теорем о сходимости упрощаются и приобретают более наглядный смысл.

Исторически метод функций Ляпунова и возник применительно к подобным задачам. Мы, однако, не будем приводить соответствующие результаты и рассматривать непрерывные методы. Дело в том, что развитие цифровой техники привело к тому, что теперь ЭВМ являются основными средствами решения вычислительных задач. Но при реализации процесса (40) на ЭВМ

### § 3. ДРУГИЕ СХЕМЫ

нужно переходить к его дискретной аппроксимации, т. е. вновь вернуться к итеративным методам. В то же время нужно иметь в виду, что переход к «пределной» форме дискретной траектории может быть целесообразен с методической точки зрения для упрощения формулировок и «угадывания» различных методов. Для обоснования сходимости подобный подход систематически используется в монографии В. З. Беленьского, В. А. Волконского и др. [2, 1].

Наконец, часто итерационный процесс рассматривается в форме

$$x^{k+1} = T(x^k), \quad T: \mathbf{R}^n \rightarrow \mathbf{R}^n, \quad (41)$$

а не в виде (23). Постулируется существование функции  $V(x)$ , обладающей свойством

$$V(T(x)) < V(x), \quad x \neq T(x), \quad (42)$$

при этом ни дифференцируемости, ни гладкости  $V(x)$  и  $T(x)$  не требуется. Достаточно предположить, например, полуунпредынность снизу функции  $\Phi(x) = V(T(x))$  и ограниченность множества  $\{x: V(x) \leq V(x^0)\}$ . При этих условиях удаётся доказать, что у последовательности (41) есть предельные точки, и каждая из них является неподвижной точкой  $T(x)$ . Схемы такого типа предложены и исследованы в [0,6, 0,13, 1,6, 2,9]. Перспективной в круге данных идей представляется схема Е. А. Нурминского [2, 9]; в ней не требуется монотонного убывания  $V(x)$  на каждом шаге и она пригодна для стохастического случая. К сожалению, подобные подходы не дают никакой информации о скорости сходимости процесса.

### § 3. Другие схемы

Не нужно думать, что первый и второй методы Ляпунова исчерпывают все многообразие схем исследования сходимости итерационных процедур. Иногда эти схемы основываются на несколько иных соображениях. Упомянем кратко некоторые из них.

**1. Принцип сжимающих отображений.** Пусть  $g: \mathbf{R}^n \rightarrow \mathbf{R}^n$  — некоторое отображение. Оно называется *сжимающим*, если

$$\|g(x) - g(y)\| \leq q \|x - y\|, \quad q < 1, \quad (1)$$

для всех  $x, y \in \mathbf{R}^n$ , т. е. если оно удовлетворяет условию Липшица с константой, меньшей 1. Рассмотрим итеративный процесс

$$x^{k+1} = g(x^k).$$

**Теорема 1** (принцип сжимающих отображений). Если  $g$  — сжимающее отображение, то оно имеет единственную

неподвижную точку  $x^*$ , к которой сходится процесс (2) при любом  $x^0$  со скоростью геометрической прогрессии

$$\|x^k - x^*\| \leq q^k (1 - q)^{-1} \|g(x^0) - x^0\|. \quad (3)$$

Доказательство.

$$\begin{aligned} \|x^{k+1} - x^k\| &= \|g(x^k) - g(x^{k-1})\| \leq q \|x^k - x^{k-1}\|, \\ \|x^{k+1} - x^k\| &\leq q^k \|x^1 - x^0\|, \\ \|x^{k+s} - x^k\| &\leq \sum_{i=k}^{k+s-1} \|x^{i+1} - x^i\| \leq \\ &\leq (q^{k+s-1} + q^{k+s-2} + \dots + q^k) \|x^1 - x^0\| \leq \frac{q^k}{1-q} \|x^1 - x^0\|. \end{aligned} \quad (4)$$

Следовательно,  $\|x^{k+s} - x^k\| \rightarrow 0$  при  $k \rightarrow \infty$  и любом  $s$ , т. е.  $x^k \rightarrow x^*$  последовательность Коши в  $\mathbf{R}^n$ . В силу полноты  $\mathbf{R}^n$   $x^k$  имеет предел  $x^*$ . Так как  $g(x)$  непрерывна в силу (1), то из  $x^k \rightarrow x^*$  следует  $g(x^k) \rightarrow g(x^*)$ , но  $g(x^k) = x^{k+1} \rightarrow x^*$ . Поэтому  $x^* = g(x^*)$ . Переходя в (4) к пределу при  $s \rightarrow \infty$ , получаем  $\|x^* - x^k\| \leq (q^k / (1 - q)) \|x^1 - x^0\|$ . Единственность неподвижной точки сразу следует из (1). ▲

Принцип сжимающих отображений у花开 тем, что он не только утверждает сходимость итеративного процесса, но и гарантирует существование неподвижной точки. Поэтому он традиционно применяется в математике для получения разнообразных теорем существования.

Принцип сжимающих отображений допускает различные обобщения и модификации. Однако, как показывают приводимые ниже упражнения 1–3, существенно расширить его нельзя.

Отметим еще, что попытка непосредственно применить принцип сжимающих отображений к задачам, рассмотренным в § 1, не дает результатов. В самом деле, там было показано, что если спектральный радиус  $\rho(A)$  матрицы  $A$  меньше 1, то итерации  $x^{k+1} = Ax^k$  сходятся. Однако в этих условиях линейное отображение  $g(x) = Ax$  не является, вообще говоря, сжимающим, так как не обязательно  $\|A\| < 1$ , см. § 1.

#### Упражнения.

1. Постройте пример отображения  $g(x)$ , обладающего свойством:  $\|g(x) - g(y)\| < \|x - y\|$  для любых  $x \neq y$ , но не имеющего неподвижной точки.
2. Постройте пример нерастягивающего отображения:  $\|g(x) - g(y)\| \leq \|x - y\|$ , имеющего неподвижную точку, для которой итерации  $x^{k+1} = g(x^k)$  не сходятся.
3. Постройте пример сжимающих отображений  $g_k$  с общей константой скажет  $q < 1$ , для которых итерации  $x^{k+1} = g_k(x^k)$  не сходятся.

**2. Теорема о неявной функции.** Удобным инструментом при исследовании итеративных методов, не разрешенных относительно  $x^{k+1}$ , является хорошо известная из анализа теорема о неявной функции. Пусть  $F(x, y)$  — отображение из  $\mathbf{R}^n \times \mathbf{R}^n$  в  $\mathbf{R}^n$ . Будем обозначать  $F_x(x, y)$ ,  $F_y(x, y)$  производные  $F$  по соответствующим переменным.

Теорема 2 (о неявной функции). Пусть  $F(x^*, y^*) = 0$ ,  $F_x(x, y)$  непрерывна по  $\{x, y\}$  в окрестности  $x^*, y^*$ ,  $F_x(x, y)$  дифференцируема по  $x$  в окрестности  $x^*, y^*$ ,  $F_x'(x^*, y^*)$  не вырождена. Тогда существует единственная непрерывная в окрестности  $y^*$  функция  $x = \Phi(y)$  такая, что  $x^* = \Phi(y^*)$ ,  $F(\Phi(y), y) = 0$ . Если, кроме того,  $F_y(x^*, y^*)$  существует, то  $\Phi(y)$  дифференцируема в  $y^*$  и

$$F'(y^*) = -[F'_x(x^*, y^*)]^{-1} F'_y(x^*, y^*). \quad (5)$$

Иными словами, уравнение  $F(x, y) = 0$  может быть решено относительно  $x$  в окрестности  $y^*$ . Применим этот результат прежде всего для исследования существования и устойчивости решения уравнений.

Теорема 3. Пусть уравнение  $g(x) = 0$ ,  $g: \mathbf{R}^n \rightarrow \mathbf{R}^n$ , имеет решение  $x^*$ , причем  $g'(x)$  дифференцируема в окрестности  $x^*$ ,  $g'(x)$  непрерывна в  $x^*$  и матрица  $g'(x^*)$  не вырождена. Тогда уравнение

$$g(x) = y \quad (6)$$

имеет решение  $x(y)$  при достаточно малых  $y$ , причем

$$x(y) = x^* - g'(x^*)^{-1} y + o(y). \quad (7)$$

Приведенные результаты позволяют исследовать итерационные процессы, в которых новое приближение  $x^{k+1}$  задается неявным выражением, например оно является решением некоторой вспомогательной задачи безусловной минимизации. Именно так обстоит дело в методе регуляризации и многих методах решения задач с ограничениями (например, методах штрафных функций).

**3. О роли общих схем исследования сходимости.** Общие теоремы типа приведенных в этой главе берут на себя стандартную, рутинную часть доказательств сходимости и тем самым упрощают процесс обоснования алгоритмов. Однако не нужно преувеличивать их роль и считать, что они делают анализ сходимости элементарным. Во-первых, во многих случаях проверка их условий представляет самостоятельную нетривиальную проблему. Во-вторых, для простых задач непосредственное, «в лоб», доказательство ничуть не сложнее обращения к общим

теоремам. Примеры этому мы видели в гл. 1. Конечно, можно было бы ее результаты доказать с помощью схем настоящей главы. Однако такой подход лишен наглядности и поучительности прямых доказательств. Наконец, в ряде задач целесообразно применить тот или иной специальный прием, использующий особенности задачи.

Таким образом, анализ сходимости остается творческим процессом, требующим искусства и здравого смысла. Попытки уложить его в прокрустово ложе некоторой единой схемы (что свойственно, в частности, некоторым монографиям), производят впечатление неестественности.

### Г л а в а 3 МЕТОДЫ МИНИМИЗАЦИИ

В гл. 1 были рассмотрены два алгоритма минимизации, наиболее простые в идейном отношении — градиентный метод и метод Ньютона. В настоящее время известны десятки (если не сотни) других методов безусловной минимизации дифференцируемых функций. Ниже будут описаны некоторые из них — наиболее интересные либо в теоретическом, либо в вычислительном отношении. На протяжении всей главы речь идет о задаче

$$\min f(x), \quad x \in \mathbf{R}^n,$$

где  $f(x)$  — дифференцируемая функция.

#### § 1. Модификации градиентного метода и метода Ньютона

**1. Недостатки и достоинства исходных методов.** В гл. 1 были подробно исследованы *градиентный метод*

$$x^{k+1} = x^k - \gamma \nabla f(x^k) \quad (1)$$

и *метод Ньютона*

$$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k). \quad (2)$$

Этот анализ позволяет составить следующую таблицу сравнительных достоинств и недостатков каждого из методов.

Таблица 1

Метод	Достоинства	Недостатки
Градиентный	Глобальная сходимость. Слабые требования к $f(x)$ . Простота вычислений	Медленная сходимость. Необходимость выбора $\gamma$
Ньютона	Быстрая сходимость	Локальная сходимость. Жесткие требования к $f(x)$ . Большой объем вычислений

Точный смысл этих терминов был разъяснен в гл. 1. Как видно из таблицы, достоинства и недостатки данных методов

являются взаимно дополнительными, и хотелось бы разработать метод, объединяющий достоинства каждого из них и в то же время свободный от их недостатков. Хотя такого идеального решения нет, в данной главе будут описаны возможные шаги по направлению к нему.

Прежде всего, оказывается, что некоторые из недостатков (необходимость выбора  $\gamma$  для градиентного метода, локальность метода Ньютона) могут быть устранены за счет простой модификации методов. К этому мы и перейдем.

**2. Модификация градиентного метода.** Рассмотрим общий градиентный метод

$$x^{k+1} = x^k - \gamma_k \nabla f(x^k) \quad (3)$$

при различных способах выбора длины шага  $\gamma_k$ . На первый взгляд кажется, что можно значительно повысить эффективность градиентного метода, если идти до минимума по направлению антиградиента:

$$\gamma_k = \arg\min_{\gamma \geq 0} \Phi_k(\gamma), \quad \Phi_k(\gamma) = f(x^k - \gamma \nabla f(x^k)). \quad (4)$$

При этом мы получаем так называемый *метод скорейшего спуска*.

**Теорема 1.** Пусть  $f(x)$  — непрерывно дифференцируемая функция и  $\{x: f(x) \leq f(x^0)\}$  ограничено. Тогда в методе (3), (4)  $\nabla f(x^k) \rightarrow 0$  и  $y$  последовательности  $x^k$  существуют предельные точки, каждая из которых стационарна, т. е. *найдется подследовательность*  $x^{k_i} \rightarrow x^*$ ,  $u \nabla f(x^*) = 0$ .

Этот результат нетрудно доказать с использованием техники гл. 2.  $\blacktriangle$

По сравнению с теоремой 1 § 4 гл. 1 здесь условие Липшица на градиент удастся заменить более слабым требованием непрерывности градиента. Это естественно, поскольку способ выбора длины шага (4) является более гибким, чем выбор  $\gamma_k \equiv \gamma$ . Метод (3), (4) сходится для тех примеров, на которых в § 4 гл. 1 демонстрировалась расходимость градиентного метода с постоянным шагом при невыполнении условия Липшица.

Выясним вопрос о скорости сходимости метода. Рассмотрим пример квадратичной функции

$$f(x) = (Ax, x)/2 - (b, x), \quad A > 0. \quad (5)$$

В данном случае  $\gamma_k$  из (4) может быть выписано явно:

$$\gamma_k = \frac{\|\nabla f(x^k)\|^2}{(A\nabla f(x^k), \nabla f(x^k))}. \quad (6)$$

Метод (3), (6) выгодно отличается от (1) тем, что в него не входит поллежащий выбору параметр  $\gamma$ .

#### § 1. ГРАДИЕНТНЫЕ МЕТОДЫ И МЕТОД НЬЮТОНА

**Теорема 2.** Для метода (3), (6) для функции (5) справедлива оценка

$$f(x^k) - f(x^*) \leq (f(x^0) - f(x^*)) \left( \frac{L-l}{L+l} \right)^{2k}, \quad (7)$$

где  $l$  — наименьшее, а  $L$  — наибольшее собственное значение матрицы  $A$ ,  $x^* = A^{-1}b$  — точка минимума  $f(x)$ .

Доказательство. Используя вид  $\Phi_k(\gamma)$  и  $\gamma_k$ , имеем

$$f(x^{k+1}) = f(x^k) - \gamma_k (\nabla f(x^k), \nabla f(x^k)) + \gamma_k^2 (A\nabla f(x^k), \nabla f(x^k))/2 =$$

$$= f(x^k) - \frac{1}{2} \frac{\|\nabla f(x^k)\|^4}{(A\nabla f(x^k), \nabla f(x^k))}.$$

Так как  $2(f(x^k) - f(x^*)) = (A(x^k - x^*), x^k - x^*) = (A^{-1}\nabla f(x^k), \nabla f(x^k))$ , то

$$\frac{f(x^{k+1}) - f(x^*)}{f(x^k) - f(x^*)} = 1 - \frac{\|\nabla f(x^k)\|^4}{(A^{-1}\nabla f(x^k), \nabla f(x^k))(A\nabla f(x^k), \nabla f(x^k))}.$$

Используя неравенство Канторовича

$$(Ax, x)(A^{-1}x, x) \leq (4Ll)^{-1}(L+l)^2 \|x\|^4 \quad \forall x \in \mathbb{R}^n, \quad (8)$$

получаем

$$\frac{f(x^{k+1}) - f(x^*)}{f(x^k) - f(x^*)} \leq \left( \frac{L-l}{L+l} \right)^2,$$

что и дает требуемую оценку (7).  $\blacktriangle$

Поскольку  $2(f(x) - f(x^*)) = (A(x-x^*), x-x^*) \geq l\|x-x^*\|^2$ , то из (7) следует

$$\|x^k - x^*\| \leq \sqrt{2l^{-1}(f(x^0) - f(x^*))} q^k, \quad q = (L-l)/(L+l). \quad (9)$$

Оценка (7) точная, так как нетрудно построить двумерный пример, для которого неравенство в (7) превращается в равенство. Сопоставляя (7) и (9) с теоремой 3 § 4 гл. 1, приходим к несколько неожиданному выводу — метод скорейшего спуска для квадратичной функции сходится, вообще говоря, не быстрее, чем простой градиентный метод (1) при соответствующем выборе  $\gamma$ . Этот же вывод справедлив и для общего неквадратичного случая. Итак, добиться выигрыша в скорости сходимости за счет более полной одномерной минимизации (т. е. выбора шага в соответствии с (4)) в градиентном методе нельзя.

Отсюда не следует делать вывод, что в принципе нельзя ускорить сходимость градиентного метода путем выбора длины шага. Например, если для минимизации квадратичной функции (5) применить градиентный метод (3) с  $\gamma_k = 1/\lambda_{k+1}$ ,  $k = 0, \dots, n-1$ , где  $\lambda_i$  — собственные значения  $A$ , то такой метод

будет конечен, т. е.  $x^n = x^*$  (проверьте!). Конечно, этот результат вряд ли представляет практический интерес, так как собственные значения  $A$  обычно неизвестны, а их нахождение — задача более трудная, чем решение системы  $Ax = b$ .

Перейдем к другому способу выбора  $\gamma_k$ . Простейший выбор  $\gamma_k \equiv \gamma$ ,  $0 < \gamma < 2/L$  (теорема 1 § 4 гл. 1), неконструктивен, так как константа  $L$  обычно неизвестна. Можно предложить следующую процедуру подбора  $\gamma$ . Задаются  $0 < \varepsilon < 1$ ,  $0 < \alpha < 1$  и некоторое  $\gamma$ . На каждой итерации вычисляется  $f(x^k) - \gamma Vf(x^k)$  и проверяется неравенство

$$f(x^k) - \gamma Vf(x^k) \leq f(x^k) - \varepsilon V\|Vf(x^k)\|^2. \quad (10)$$

Если оно выполняется, то  $x^{k+1} = x^k - \gamma Vf(x^k)$ , если же нет, то  $\gamma$  заменяется на  $\gamma\alpha$  и проверка повторяется.

Можно показать, что в условиях теорем 1 и 2 § 4 гл. 1 та-кая процедура требует конечного числа дроблений  $\gamma$  на каждой итерации и остается в силе утверждения этих теорем. Таким образом, правило выбора длины шага нетрудно сделать конструктивным. Однако главный недостаток градиентного метода — его медленную сходимость для плохо обусловленных за-дач — простыми средствами устранить не удается.

**3. Модификации метода Ньютона.** Придать методу Ньютона свойство глобальной сходимости можно различными способами. Один из них связан с регулировкой длины шага:

$$x^{k+1} = x^k - \gamma_k [\nabla^2 f(x^k)]^{-1} \nabla f(x^k). \quad (11)$$

Его часто называют *демпфированным методом Ньютона*. Параметр  $\gamma_k$  может выбираться по-разному, например

$$\gamma_k = \underset{\gamma > 0}{\operatorname{argmin}} f(x^k - \gamma [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)) \quad (12)$$

или  $\gamma$  дробится (умножается на  $0 < \alpha < 1$ ), начиная с  $\gamma = 1$ , до выполнения условия

$$f(x^{k+1}) \leq f(x^k) - \gamma q ([\nabla^2 f(x^k)]^{-1} \nabla f(x^k), \nabla f(x^k)), \quad 0 < q < 1, \quad (13)$$

или условия

$$\|\nabla f(x^{k+1})\|^2 \leq (1 - \gamma q) \|\nabla f(x^k)\|^2, \quad 0 < q < 1. \quad (14)$$

Для гладких сильно выпуклых функций демпфированный ме-тод Ньютона глобально сходится (упр. 1). Что касается скоро-сти сходимости, то на начальных итерациях можно утверждать лишь сходимость со скоростью геометрической прогрессии. При попадании же в окрестность  $x^*$ , в которой выполняются условия теоремы 1 § 5 гл. 1, будет иметь место квадратичная сходимость (упр. 2).

Возможна и другая модификация (называемая *методом Ленбеберга — Марквардта*), в которой само направление движения

отличается от задаваемого методом Ньютона. Постулим, так же, как при одном из обоснований градиентного метода (см. (3) § 4 гл. 1) — добавим к аппроксимирующей функции квадра-тичный штраф за отклонение от точки  $x^*$ , т. е. будем искать  $x^{k+1}$  из условия минимума

$$f_k(x) + (a_k/2) \|x - x^k\|^2, \quad f_k(x) = f(x^k) + (\nabla f(x^k), x - x^k) + (\nabla^2 f(x^k)(x - x^k), x - x^k)/2. \quad (15)$$

Тогда приходим к методу

$$x^{k+1} = x^k - (\nabla^2 f(x^k) + a_k I)^{-1} \nabla f(x^k). \quad (16)$$

При  $a_k = 0$  метод переходит в метод Ньютона, при  $a_k \rightarrow \infty$  направление движения стремится к антиградиенту. Таким обра-зом, (16) представляет собой компромисс между эми двумя методами. За счет выбора  $a_k$  можно добиться глобальной схо-димости метода (упр. 3).

Метод (16) обладает перед (11) тем преимуществом, что он (как и градиентный метод) пригоден не только для выпук-лых функций (см. упр. 3), тогда как в методе (11) требуется положительная определенность матрицы  $\nabla^2 f(x)$  (упр. 4).

Есть специальные модификации метода Ньютона, в которых

матрица  $\nabla^2 f(x^k)$  заменяется на некоторую положительно опре-деленную, если сама  $\nabla^2 f(x^k)$  таковой не является.

Однако во всех описанных модификациях метода Ньютона каждая итерация (как и в основном методе Ньютона) требует очень большой вычислительной работы (вычисление  $\nabla^2 f(x)$ , ре-решение систем линейных уравнений), а скорость сходимости вдали от минимума, вообще говоря, не высока.

Таким образом, попытки «слегка подправить» градиентный метод и метод Ньютона хотя и позволяют устранить некоторые их недостатки, но не меняют положение с наиболее серьезными их дефектами — медленной сходимостью градиентного метода и трудоемкостью метода Ньютона.

#### Упражнения.

1. Пусть  $f(x)$  — гладкая дифференцируемая сильно выпуклая функция,  $\|\nabla^2 f(x)\| \leq L$ . Тогда в процедуре (13), (14) число дроблений  $\gamma$  на каждой итерации конечно, а метод (11) с любым правилом (12) — (14) выбора  $\gamma_k$  при любом  $x^0$  сходится к точке минимума  $x^*$  со скоростью геометрической прогрессии. Докажите, воспользовавшись теоремами § 2 гл. 2  $\gamma_k = f(x) - f(x^*)$  или  $V(x) = \|\nabla f(x)\|^2$ .

2. Покажите, что в условиях теоремы 1 § 5 гл. 1 в достаточно малой окрестности  $x^*$  в методах (13) и (14) будет  $\gamma_k = 1$ . Для гладких сильно выпуклых функций,  $\|\nabla^2 f(x)\| \leq L$ , множество  $\{x : f(x) \leq f(x^*)\}$  ограничено, а точка  $x^*$ , в которой  $\nabla f(x^*) = 0$ , единственна. Покажите, что можно указать такие  $\underline{\gamma}$  и  $\bar{\gamma}$ , что при  $\underline{\gamma} \leq \gamma_k \leq \bar{\gamma}$  методе (16) будет  $x^k \rightarrow x^*$  (воспользовавшись теоремами § 2 гл. 2 с  $V(x) = f(x) - f(x^*)$ ).

4. Приведите примеры, показывающие, что если матрица  $\nabla^2 f(x^k)$  не является положительно определенной, то метод (11) может потерять смысл ( $[\nabla^2 f(x^k)]^{-1}$  не существует), а в методе (11), (12) может оказаться  $y_k = 0$  в точке, где  $\nabla f(x^k) \neq 0$ .

## § 2. Многошаговые методы

В градиентном методе на каждом шаге никак не используется информация, полученная на предыдущих итерациях. Естественно попытаться учесть «предысторию» процесса для ускорения сходимости. Такого рода методы, в которых новое приближение зависит от  $s$  предыдущих:

$$x^{k+1} = \Phi_k(x^k, \dots, x^{k-s+1}), \quad (1)$$

называются *s-шаговыми*. Градиентный метод и метод Ньютона были одношаговыми, теперь рассмотрим *многошаговые* ( $s > 1$ ) методы.

**1. Метод тяжелого шарика.** Одним из простейших многошаговых методов является двухшаговый *метод тяжелого шарика*

$$(2)$$

$$x^{k+1} = x^k - \alpha \nabla f(x^k) + \beta(x^k - x^{k-1}),$$

где  $\alpha > 0$ ,  $\beta \geq 0$  — некоторые параметры. Ясно, что при  $\beta = 0$  метод (2) переходит в градиентный. Свое название метод получил из-за следующей физической аналогии. Движение тела («тяжелого шарика») в потенциальном поле при наличии силы трения (или вязкости) описывается дифференциальным уравнением второго порядка

$$\mu \frac{d^2 x(t)}{dt^2} = -\nabla f(x(t)) - p \frac{dx(t)}{dt}. \quad (3)$$

Ясно, что из-за потери энергии на трение тело в конце концов окажется в точке минимума потенциала  $f(x)$ . Таким образом, тяжелый шарик «решает» соответствующую задачу минимизации. Если рассмотреть разностный аналог уравнения (3), то придем к итерационному методу (2).

Введение инерции движения (член  $\beta(x^k - x^{k-1})$ ) в итерационный процесс может привести к ускорению сходимости. Это видно, например, из рис. 6 — вместо зигзагообразного движения в градиентном методе в данном случае получается более плавная траектория по «дну оврага». Эти эвристические соображения подкрепляются следующей теоремой.

**Теорема 1.** Пусть  $x^*$  — невырожденная точка минимума  $f(x)$ ,  $x \in \mathbb{R}^n$ . Тогда при

$$0 \leq \beta < 1, \quad 0 < \alpha < 2(1 + \beta)L, \quad II \leq \nabla^2 f(x^*) \leq LI \quad (4)$$

находится  $\varepsilon > 0$  такое, что при любых  $x^0, x^1, \|x^0 - x^*\| \leq \varepsilon$ ,  $\|x^1 - x^*\| \leq \varepsilon$  метод (2) сходится к  $x^*$  со скоростью геометрической прогрессии:



Рис. 6. Метод тяжелого шарика (a) и градиентный метод (b).

размерности пространства, позволяющий свести многошаговый процесс к одношаговому (см. (15) § 1 гл. 2). Введем  $2n$ -мерный вектор  $z^k = \{x^k - x^*, x^{k-1} - x^*\}$ . Тогда итерационный процесс (2) может быть записан в форме

$$z^{k+1} = Az^k + o(z^k), \quad (7)$$

где квадратная матрица  $A$  размерности  $2n \times 2n$  имеет вид

$$A = \begin{pmatrix} (1 + \beta)I - \alpha B & -\beta I \\ I & 0 \end{pmatrix}, \quad B = \nabla^2 f(x^*). \quad (8)$$

Пусть  $I = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n = L$  — собственные значения матрицы  $B$ . Тогда собственные значения  $\rho_j$ ,  $j = 1, \dots, 2n$ , матрицы  $A$  совпадают с собственными значениями матриц  $2 \times 2$  вида

$$\begin{pmatrix} 1 + \beta - \alpha \lambda_i & -\beta \\ 1 & 0 \end{pmatrix}.$$

Следовательно, они являются корнями уравнений

$$\rho^2 - \rho(1 + \beta - \alpha \lambda_i) + \beta = 0, \quad i = 1, \dots, n. \quad (9)$$

Можно показать, что если  $0 < I \leq \lambda_i \leq L$ ,  $0 \leq \beta < 1$ ,  $0 < \alpha < 2(1 + \beta)/L$ , то  $|\rho| < 1$ , где  $\rho$  — любой корень уравнения (9).

Теперь мы можем воспользоваться теоремой 1 § 1 гл. 2 о логической сходимости итерационных процессов вида (7), что дает возможность получить оценку (5). Вычисля  $\min_{\alpha, \beta} \max_{1 \leq i \leq 2n} |\rho_i|$ , находим приведенные в теореме оптимальные значения  $\alpha^*$ ,  $\beta^*$  и соответствующее им  $q^*$ . ▲

Сравним скорость сходимости, даваемую одношаговым и двухшаговым методами при оптимальном выборе параметров. И в том, и в другом случаях имеем сходимость со скоростью геометрической прогрессии, но знаменатель прогрессии для одношагового метода равен

$$q_1 = (L - l)/(L + l), \quad (10)$$

а для двухшагового

$$q_2 = (\sqrt{L} - \sqrt{l})/(\sqrt{L} + \sqrt{l}). \quad (11)$$

Для больших значений числа обусловленности  $\mu = L/l$

$$q_1 \approx 1 - 2/\mu, \quad q_2 \approx 1 - 2/\sqrt{\mu}. \quad (12)$$

Поэтому, чтобы приблизиться к решению в  $\epsilon = 2,7 \dots$  раз, в одношаговом методе требуется порядка  $\mu/2$  итераций, в двухшаговом — порядка  $\sqrt{\mu}/2$ . Иными словами, для плохо обусловленных задач метод тяжелого шарика дает выигрыш в  $\sim \sqrt{\mu}$  раз по сравнению с градиентным. Для больших  $\mu$  эта разница весьма значительна. С вычислительной же точки зрения метод (2) немногим сложнее одношагового.

Правда, подбор оптимальных значений  $\alpha$  и  $\beta$  в (2) не прост — формулами (6) непосредственно воспользоваться не удается, так как границы спектра  $\nabla^2 f(x^*)$  (числа  $l$  и  $L$ ) обычно неизвестны.

#### Упражнение.

1. Докажите глобальную сходимость метода (2) для квадратичной  $f(x)$ .

**2. Метод сопряженных градиентов.** Рассмотрим другой вариант двухшагового метода — метод сопряженных градиентов, в котором параметры находятся из решения двумерной задачи оптимизации:

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k) + \beta_k (x^k - x^{k-1}), \quad (13)$$

$$\{\alpha_k, \beta_k\} = \operatorname{argmin}_{\alpha, \beta} f(x^k - \alpha \nabla f(x^k) + \beta (x^k - x^{k-1})). \quad (14)$$

Для случая квадратичной функции

$$f(x) = (Ax, x)/2 - (b, x), \quad A > 0,$$

эта задача может быть решена явно:

$$\alpha_k = \frac{\|r^k\|^2 (A\rho^k, p^k) - (r^k, p^k)(A_r^k, p^k)}{(A_r^k, r^k)(A\rho^k, p^k) - (A_r^k, p^k)^2}, \quad r^k = \nabla f(x^k) = Ax^k - b, \\ \beta_k = \frac{\|r^k\|^2 (A_r^k, p^k) - (r^k, p^k)(A_r^k, r^k)}{(A_r^k, r^k)(A\rho^k, p^k) - (A_r^k, p^k)^2}, \quad p^k = x^k - x^{k-1}. \quad (16)$$

Могло бы показаться, что соотношение методов (13), (14) и (2) такое же, как методов (3), (4) и (1) в § 1, — если метод скорейшего спуска не дает, как мы видели, выигрыша в сходимости по сравнению с градиентным методом с постоянным оптимальным  $\gamma$ , то и от двухшагового варианта скорейшего спуска (13), (14) трудно ждать существенного ускорения по сравнению с методом тяжелого шарика (2). Оказывается, ситуация здесь иная: так, в квадратичном случае метод (13), (14) (при специальном выборе  $p^1$ ) является конечным, т. е. дает точный минимум функции (15) за конечное число итераций.

Пусть начальное приближение  $x^0$  произвольно, а  $x^1$  получено из него методом скорейшего спуска:

$$x^1 = x^0 - \frac{\|r^0\|^2}{(A_r^0, r^0)} r^0, \quad r^0 = \nabla f(x^0) = Ax^0 - b. \quad (17)$$

**Лемма 1.** Градиенты  $r^0, r^1, \dots$  в методе (13), (16), (17) попарно ортогональны.

Доказательство. Воспользуемся индукцией по  $k$ . Пусть  $(r^i, r^k) = 0$  при  $0 \leq i < k$ ,  $k \geq 2$ , и  $r^i \neq 0$ ,  $i = 0, \dots, k$ . Ортогональность  $r^0, r^1, r^2$  следует непосредственно из определения метода. Тогда, умножая (13) слева на  $A$ , получаем

$$r^{k+1} = r^k - \alpha_k A r^k + \beta_k (r^k - r^{k-1}).$$

Из  $r^i \neq 0$  для  $i \leq k$  следует, что  $\alpha_k \neq 0$ . Поэтому  $A r^k$  есть линейная комбинация  $r^{k+1}, r^k$  и  $r^{k-1}$ , аналогично  $A r^i$ ,  $i < k$ , есть линейная комбинация  $r^{i+1}, r^i, r^{i-1}$  и в силу предположения индукции  $(A r^i, r^j) = 0$ ,  $|i - j| \geq 1$ ,  $i < k$ ,  $j \leq k$ . Следовательно,  $(r^{k+1}, r^i) = (r^k - \alpha_k A r^k + \beta_k (r^k - r^{k-1}), r^i) = 0$

при  $i = 0, \dots, k - 2$ .

Далее, непосредственно из формул (13), (16) следует, что  $(r^{k+1}, r^k) = 0$ ,  $(r^{k+1}, p^k) = 0$ .

Наконец, из (13), заменив  $k$  на  $k - 1$ , имеем  $r^k = -\alpha_{k-1} r^{k-1} + \beta_{k-1} p^{k-1}$ . Применяя это соотношение последовательно, получаем, что  $r^k$  есть линейная комбинация  $r^0, r^1, \dots, r^{k-1}$ , причем  $r^{k-1}$  входит с коэффициентом  $-\alpha_{k-1} \neq 0$ . Поэтому из  $(r^{k+1}, p^k) = 0$ ,  $(r^{k+1}, r^i) = 0$ ,  $i \leq k - 2$ , следует, что  $(r^{k+1}, r^{k-1}) = 0$ . Итак, для всех  $i \leq k$  будет  $(r^{k+1}, r^i) = 0$ . ▲

Если  $r^k$  обращается в 0, то  $x^k$  — точка минимума  $f(x)$ . Но в  $\mathbf{R}^n$  не может существовать более  $n$  ортогональных ненулевых векторов, поэтому для некоторого  $k \leq n$  будет  $r^k = 0$ . Итак, мы доказали следующий результат.

**Теорема 2.** *Метод (13), (16), (17) дает точку минимума квадратичной функции  $f(x)$  вида (15) за число итераций, не превосходящее  $n$ .* ▲

Мы установим в дальнейшем (см. гл. 7), что если  $L$  — некоторое подпространство в  $\mathbf{R}^n$ ,  $f(x)$  — выпуклая дифференцируемая функция, то условие

$$(\nabla f(x^*), a) = 0 \quad \text{для всех } a \in L$$

необходимо и достаточно для того, чтобы  $x^*$  было минимумом  $f(x)$  на  $L$ . Отсюда из леммы 1 следует, что  $x^k$  — точка минимума квадратичной функции  $f(x)$  вида (5) на подпространстве, проходящем через  $x^0$  и порожденном  $r^0, \dots, r^{k-1}$ . Этот несколько неожиданный факт (мы ищем минимум  $k$  раз последовательно на 2-мерных подпространствах, а он оказывается минимумом на всем  $k$ -мерном подпространстве) является важнейшей особенностью метода сопряженных градиентов и объясняет его коленность.

Последовательные направления движения  $r^k$  в методе сопряженных градиентов удовлетворяют соотношению

$$(Ap^i, p^j) = 0, \quad i \neq j. \quad (19)$$

Действительно,  $p^i = x^i - x^{i-1}$ , поэтому  $Ap^i = Ax^i - Ax^{i-1} = r^i - r^{i-1}$ . С другой стороны, мы уже отмечали, что  $p^k$  есть линейная комбинация  $r^0, \dots, r^{k-1}$ ,  $p^k = \sum_{j=0}^{k-1} \mu_j r^j$ . Поэтому для

$$i > k \text{ имеем } (Ap^i, p^k) = \left( r^i - r^{i-1}, \sum_{j=0}^{k-1} \mu_j r^j \right) = 0 \text{ в силу леммы 1.}$$

Векторы  $p^i$ , связанные соотношением (19), называются *сопряженными* или *д-ортогональными* (они ортогональны в метрике, задаваемой матрицей  $A$ ). Это объясняет название метода — в нем строятся линейные комбинации последовательных градиентов, являющиеся сопряженными.

Отметим, что знание произвольных сопряженных направлений  $s^i$ ,  $i = 1, \dots, n$ ,  $(As^i, s^i) = 0$ ,  $i \neq j$ , позволяет без труда решить систему

$$Ax = b, \quad A > 0.$$

Действительно, будем искать решение в виде  $x = \sum_{i=1}^n \alpha_i s^i$ . Тогда, подставляя это в (20), умножая скалярно на  $s^i$  и используя

$A$ -ортогональность, имеем

$$\alpha_i = (b, s^i)/(As^i, s^i). \quad (21)$$

Этому решению можно придать рекуррентную форму: зададимся произвольным  $x^0$  и построим  $x^{k+1} = x^k + \alpha_k s^k$ , где  $\alpha_k$  задаются (21). Тогда  $x^n = x^*$  — решение (20). Поскольку  $\alpha_k$  в (21) можно определить иначе:  $\alpha_k = \arg\min_a f(x^k + as^k)$ , то мы получаем,

что знание системы сопряженных направлений позволяет найти минимум квадратичной функции с помощью  $n$  одномерных минимизаций. Этот важный факт будет неоднократно использоваться в дальнейшем при построении других методов минимизации. В методе сопряженных градиентов сопряженные направления не выбираются заранее, а строятся по рекуррентным формулам.

Если применять метод (13), (14) для неквадратичных функций, то, сопоставляя его с методом скорейшего спуска, нетрудно доказать его глобальную сходимость, а сопоставляя с методом тяжелого шарика, — оценить скорость сходимости (упр. 3 и 4).

Методу сопряженных градиентов можно придать и иную форму. Рассмотрим итерационный процесс

$$x^{k+1} = x^k + \alpha_k p^k, \quad \alpha_k = \arg\min_a f(x^k + ap^k),$$

$$p^k = -r^k + \beta_k p^{k-1}, \quad \beta_k = \|r^k\|^2/\|r^{k-1}\|^2, \\ r^k = \nabla f(x^k), \quad \beta_0 = 0. \quad (22)$$

**Лемма 2.** *Для случая квадратичной функции (15) методы (13), (16), (17) и (22) при одинаковом  $x^0$  определяют одну и ту же последовательность точек  $x^k$ .* ▲

Поскольку  $p^k$  в (22) и (16) отличаются лишь скалярными (ненулевыми) множителями, а  $r^k$  в (22) и (16) совпадают, то процесс (22) обладает теми же свойствами, что и (13), (16): векторы  $p^i$  являются сопряженными, а градиенты  $r^i$  — взаимно ортогональны. Из леммы 2 и теоремы 1 следует, что метод (22) дает точку минимума квадратичной функции (15) в  $\mathbf{R}^n$  за число итераций, не превосходящее  $n$ . Для неквадратичных задач метод (22) проще, чем (13), (14), так как требует решения лишь одномерной (а не двумерной) вспомогательной задачи минимизации. Разумеется, в неквадратичном случае теряется свойство конечности метода и (22) превращается в, вообще говоря, бесконечный итерационный двухшаговый метод. Результат о его сходимости приведен в упражнении 5.

Обычно для неквадратичных задач метод сопряженных градиентов применяется в несколько иной форме. В него вводится процедура *обновления* — время от времени шаг делается не по формуле (22), а как в начальной точке, т. е. по градиенту.

Наиболее естественно производить обновление через число итераций, равное размерности пространства:

$$x^{k+1} = x^k + \alpha_k s^k, \quad \alpha_k = \underset{a \geq 0}{\operatorname{argmin}} f(x^k + as^k),$$

$$\begin{aligned} s^k &= -r^k + \beta_k s^{k-1}, \quad r^k = \nabla f(x^k), \\ \beta_k &= \begin{cases} 0, & k = 0, n, 2n, \dots \\ \frac{\|r^k\|^2 \|r^{k-1}\|^2}{\|r^k\|^2}, & k \neq 0, n, 2n, \dots \end{cases} \end{aligned} \quad (23)$$

Нетрудно доказать, что метод сопряженных градиентов с обновлением обладает свойством глобальной сходимости (упр. 6). Оказывается, что в то же время в окрестности минимума он сходится с квадратичной скоростью.

**Теорема 3.** Пусть  $x^*$  — невырожденная точка минимума, и в ее окрестности  $\nabla^2 f(x)$  удовлетворяет условию Липшица. Тогда для метода (23) в окрестности  $x^*$  справедлива оценка

$$\|x^{(m+1)n} - x^*\| \leq c \|x^m - x^*\|^2.$$

Иначе говоря, по скорости сходимости  $n$  шагов метода сопряженных градиентов эквивалентны одному шагу метода Ньютона. Мы не приводим доказательства теоремы, так как оно довольно промозгло. В его основе лежит идея квадратичной аппроксимации  $f(x)$  и факт конечности метода для квадратичных функций (см. теорему 2). ▲

Возможны иные вычислительные схемы метода сопряженных градиентов для неквадратичных функций. С одной из них, требующей решения двумерной задачи минимизации на каждом шаге, мы начали анализ этого метода — см. (13), (14). Другие, подобно (22), обычно включают лишь одномерные вспомогательные задачи, но отличаются от (22) правилом выбора  $\beta_k$ . Примером может служить схема

$$x^{k+1} = x^k + \alpha_k s^k, \quad \alpha_k = \underset{a \geq 0}{\operatorname{argmin}} f(x^k - as^k),$$

$$\begin{aligned} s^k &= -r^k + \beta_k s^{k-1}, \quad \beta_k = \frac{(r^k, r^k - r^{k-1})}{\|r^{k-1}\|^2}, \\ r^k &= \nabla f(x^k), \quad \beta_0 = 0. \end{aligned} \quad (24)$$

Как и для (22), здесь возможны варианты либо с обновлением, либо без него. Для квадратичной функции последовательности  $x^k$ , порождаемые методами (22) и (24), совпадают.

Как показывает опыт вычислений, для неквадратичного случая несколько более быструю сходимость обычно дает схема (24).

Представляет интерес поведение метода для задач большой размерности (когда число итераций меньше размерности). Оказывается, здесь можно гарантировать лишь сходимость со ско-

ростью геометрической прогрессии даже для квадратичного случая. Пусть  $A$  — матрица  $n \times n$ ,

$$U \leq A \leq L I, \quad l > 0, \quad (25)$$

и  $f(x)$  — соответствующая ей квадратичная функция на  $\mathbf{R}^n$ :

$$f(x) = (Ax, x)/2 - (b, x), \quad b \in \mathbf{R}^n. \quad (26)$$

Точка  $x^*$  может быть представлена в виде

$$x^k - x^* = P_k(A)(x^0 - x^*), \quad (27)$$

где  $P_k(A)$  — матричный полином  $k$ -й степени вида

$$P_k(A) = I + a_{1k}A + \dots + a_{kk}A^k. \quad (28)$$

Поэтому

$$\begin{aligned} \|x^k - x^*\|^2 &\leq 2(f(x^k) - f^*)/l = (AP_k(A))(x^0 - x^*), \quad x^0 - x^*/l \leq \\ &\leq (L/l)\|x^0 - x^*\|^2 \max_{l \leq \lambda \leq L} P_k^2(\lambda), \end{aligned}$$

где  $P_k(\lambda) = 1 + a_{1k}\lambda + \dots + a_{kk}\lambda^k$  — обычный полином. В силу свойств метода оценка для  $|f(x^k) - f^*|$  справедлива для всех  $P_k(\lambda)$ ,  $P_k(0) = 1$ , в частности, для

$$P_k(\lambda) = T_k\left(\frac{L+t-2\lambda}{L-t}\right)/T_k\left(\frac{L+t}{L-t}\right),$$

где

$$\begin{aligned} T_k(\lambda) &= [(t + \sqrt{t^2 - 1})^k + (t - \sqrt{t^2 - 1})^k]/2, \quad |\lambda| > 1; \\ T_k(\lambda) &= \cos(k \arccos \lambda), \quad |\lambda| \leq 1. \end{aligned}$$

Поэтому

$$\begin{aligned} \|x^k - x^*\| &\leq 2\left(\frac{L}{t}\right)^{\frac{1}{2}} \left[ \left( \frac{\sqrt{L} + \sqrt{t}}{\sqrt{L} - \sqrt{t}} \right)^k + \left( \frac{\sqrt{L} - \sqrt{t}}{\sqrt{L} + \sqrt{t}} \right)^k \right]^{-1} \|x^0 - x^*\| \leq \\ &\leq 2\left(\frac{L}{t}\right)^{\frac{1}{2}} q^k \|x^0 - x^*\|, \quad q = (\sqrt{L} - \sqrt{t})/(\sqrt{L} + \sqrt{t}). \quad (30) \end{aligned}$$

Можно показать на примерах, что оценка (30) неулучшаема.

Итак, при  $k < n$  для метода сопряженных градиентов, примененного для минимизации квадратичной функции, можно гарантировать сходимость со скоростью геометрической прогрессии со знаменателем  $q = (\sqrt{L} - \sqrt{t})/(\sqrt{L} + \sqrt{t}) \sim 1 - 2/\sqrt{\mu}$ ,  $\mu = L/t$ , т. е. такую же, как для метода тяжелого дзарика при оптимальном выборе его параметров. По сравнению с последним в методе сопряженных градиентов нет проблемы выбора

параметров — они определяются автоматически, хотя это и требует дополнительных вычислений для решения одномерной задачи минимизации.

Мы видим, что в методе сопряженных градиентов  $x^k$  является точкой минимума квадратичной функции  $f(x)$  на подпространстве, порожденном первыми  $k$  градиентами. Отсюда следует, что никакой метод, использующий только градиенты функции (точнее, в котором шаг делается по линейной комбинации предыдущих градиентов), не может сходиться быстрее. Иными словами, метод сопряженных градиентов является оптимальным по скорости сходимости в классе методов первого порядка. Но полученного выше результата вытекает, что для задач большой размерности с квадратичными функциями  $f(x)$ , удовлетворяющими условию (25), для всех методов первого порядка нельзя ждать сходимости более высокой, чем скорость геометрической прогрессии со знаменателем  $q = (\sqrt{L} - \sqrt{l}) / (\sqrt{L} + \sqrt{l})$ . Естественно, большая скорость сходимости не может достигаться и в более широком классе сильно выпуклых с константой  $l$  функций, градиент которых удовлетворяет условию Липшица с константой  $L$ . Факт квадратичной сходимости (теорема 3) имеет место только при числе итераций, существенно большем размерности пространства.

#### Упражнения.

2. Проверьте, что если  $x^1$  выбирается произвольно (а не по формуле (17)), то метод (13), (16) сходится к точке минимума (15) со скоростью геометрической прогрессии, но, вообще говоря, не является конечным. Для доказательства можно воспользоваться, например, тем фактом, что по определению метода (13), (14)  $f(x^{k+1}) \geq f(\tilde{x}^{k+1})$ , где  $\tilde{x}^{k+1}$  — точка, полученная из  $x^k$ ,  $x^{k-1}$  методом тяжелого шарика.

3. Пусть  $f(x)$  — непрерывная дифференцируемая функция, множество  $\{x: f(x) \leq f(x^0)\}$  ограничено. Докажите, что тогда при любых  $x^0, x^1$  в методе (13), (14) будет  $\nabla f(x^k) \rightarrow 0$  (используйте теорему 1 § 1).

4. Пусть  $x^*$  — невырожденная точка минимума  $f(x)$ . Используя то же соображение, что и в упражнении 2, докажите локальную сходимость метода (13), (14) со скоростью геометрической прогрессии.

5. Докажите следующий результат о сходимости метода сопряженных градиентов. Пусть  $f(x)$  — лифференцируемая сильно выпуклая функция, градиент которой удовлетворяет условию Липшица. Тогда метод (22) сходится при любом  $x^0$  к точке минимума  $f(x)$ . Используйте при этом следующие свойства метода:  $(r^*, p^*) = 0$ ,  $(r^*, p^*) = -\|r^*\|^2$  и лемму Абеля — Дэни (ряды

$\sum_{k=0}^{\infty} e_k, \sum_{k=0}^{\infty} e_k / (e_0 + \dots + e_k)$  сходятся или расходятся одновременно), применения ее к  $e_k = \|r^k\|^2 / \beta_1^2 \dots \beta_k^2$ . Попытайтесь оценить скорость сходимости.

6. Пусть  $f(x)$  непрерывно дифференцируема, а множество  $\{x: f(x) \leq f(x^0)\}$  ограничено. Докажите, что тогда в методе (23)  $\nabla f(x^k) \rightarrow 0$ . Это же справедливо для любого правила выбора моментов обновления, если их число бесконечно.

7. Докажите, что  $T_k(x)$ , определяемое (29), действительно является полиномом  $k$ -й степени.

### § 3. Другие методы первого порядка

В основе всех методов, описываемых в этом параграфе, лежит идея восстановления квадратичной аппроксимации функции по значениям ее градиентов в ряде точек. Тем самым методы объединяют достоинства градиентного метода (не требуется вычисление матрицы вторых производных) и метода Ньютона (быстрая сходимость вследствие использования квадратичной аппроксимации).

1. **Квазиньютоныкие методы.** Эти методы имеют общую структуру:

$$x^{k+1} = x^k - V_k H_k \nabla f(x^k), \quad (1)$$

где матрица  $H_k$  пересчитывается рекуррентным способом на основе информации, полученной на  $k$ -й итерации, так что  $H_k - [\nabla^2 f(x^k)]^{-1} \rightarrow 0$ . Таким образом, методы в пределе переходят в ньютоновский, что и объясняет их название. Отметим некоторые общие свойства методов такого типа. Доказательство приводимых ниже лемм может быть без труда получено с использованием описанной ранее техники.

Лемма 1. Пусть  $f(x) \geq f^*$ ,  $f(x)$  дифференцируема,  $\nabla f(x)$  удовлетворяет условию Липшица и

$$mI \leq H_k \leq MI, \quad m > 0. \quad (2)$$

Тогда в методе (1) с  $\gamma_k = 1$  достаточно мало, будет  $\nabla f(x^k) \rightarrow 0$ . ▲

Лемма 2. Пусть  $x^*$  — невырожденная точка минимума  $f(x)$ ,  $f(x)$   $\hat{\delta}$ -зажаты непрерывно дифференцируема в окрестности  $x^*$  и

$$\|H_k - [\nabla^2 f(x^*)]^{-1}\| \rightarrow 0. \quad (3)$$

Тогда метод (1) с  $\gamma_k = 1$  локально сходится к  $x^*$  быстрее любой геометрической прогрессии. ▲

Таким образом, при любых равномерно положительно определенных  $H_k$  метод (1) обладает глобальной сходимостью, а при условии (3) в окрестности минимума метод сходится со сверхлинейной скоростью.

Перейдем к вопросу о способах построения матриц  $H_k$ , аппроксимирующих  $[\nabla^2 f(x^k)]^{-1}$ . В принципе их можно формировать с помощью конечно-разностной аппроксимации. Именно, из точки  $x^k$  можно сделать  $n$  «пробных шагов» длины  $\alpha_k$  по координатным осям и вычислить в этих точках градиенты. Соответствующая разностная аппроксимация будет искомой, если  $\alpha_k \rightarrow 0$  (см. упр. 1).

Однако такой прямолинейный способ аппроксимации неизмен — в нем делается  $n$  пробных вычислений градиента на каждой итерации и никак не используются градиенты,

найденные на предыдущих итерациях. Кроме того, в нем требуется обращать матрицу. Основная идея квазиньютоновских методов заключается, во-первых, в том, чтобы не делать специальных пробных шагов, а использовать найденные градиенты в предыдущих точках (поскольку они близки к  $x^k$ ), а во-вторых, в том, чтобы строить аппроксимацию непосредственно для обратной матрицы  $[\nabla^2 f(x^k)]^{-1}$ . Обозначим

$$p^k = -H_k \nabla f(x^k), \quad y^k = \nabla f(x^{k+1}) - \nabla f(x^k). \quad (4)$$

Тогда для квадратичной функции  $f(x) = (Ax, x)/2 - (b, x)$ ,  $A > 0$ , имеем  $y^k = A(x^{k+1} - x^k) = \gamma_k A p^k$ , т. е.

$$\gamma_k p^k = A^{-1} y^k. \quad (5)$$

Поэтому для нового приближения  $H_{k+1}$  к  $[\nabla^2 f(x^{k+1})]^{-1}$  достаточно потребовать выполнения так называемого *квазиньютонского условия*

$$H_{k+1} y^k = \gamma_k p^k. \quad (6)$$

Кроме того, удобно получать  $H_{k+1}$  как поправку к  $H_k$  с помощью матриц первого или второго ранга. Наконец, эти поправки должны быть такими, чтобы для квадратичного случая оказалось  $H_n = A^{-1}$ .

Основным техническим инструментом анализа подобных методов является следующая лемма об обращении матриц.

**Лемма 3.** Пусть  $B$  — матрица  $n \times n$ , для которой  $B^{-1}$  существует, а,  $b$  — векторы из  $\mathbb{R}^n$ ,  $(B^{-1}a, b) \neq -1$ ,  $A = B + ab^T$ . Тогда

$$A^{-1} = B^{-1} - (1 + (B^{-1}a, b))^{-1} B^{-1}a(B^{-1}b)^T. \quad (7)$$

Лемма доказывается прямой проверкой. ▲

Таким образом, если известна матрица, обратная к  $B$ , а матрица  $A$  получена из  $B$  добавлением матрицы ранга 1, то обратная к  $A$  находится без труда.

Приведем примеры формул пересчета матриц  $H_k$ :

a) *метод Давидона — Флетчера — Паузла (ДФП)*:

$$H_{k+1} = H_k - \frac{H_k y^k (y^k)^T H_k}{(H_k y^k, y^k)} + \gamma_k \frac{\rho^k (\rho^k)^T}{(\rho^k, y^k)}, \quad H_0 > 0; \quad (8)$$

б) *метод Броидена:*

$$H_{k+1} = H_k - \frac{(\gamma_k p^k - H_k y^k)(\gamma_k p^k - H_k y^k)^T}{(\gamma_k p^k - H_k y^k, y^k)}, \quad H_0 > 0; \quad (9)$$

в) *метод Броидена — Флегнера — Шено (БФШ):*

$$H_{k+1} = H_k + \frac{\rho_k p^k (\rho^k)^T - p^k (y^k)^T H_k - H_k y^k (\rho^k)^T}{(\gamma_k, \rho^k)}, \quad (10)$$

Оказывается, для всех формул (8) — (10) выполнено квазиньютоновское условие (6). А если  $\gamma_k \geq 0$  — произвольные числа,  $p^k$  — произвольные, линейно независимые векторы,  $y^k$  удовлетворяют соотношению (5) с  $A^{-1} \geq 0$ , то при любом  $H_0 \geq 0$  будет  $H_n = A^{-1}$ . Отсюда следует

**Теорема 1.** При любых  $x^0$ ,  $H_0 \geq 0$  метод (1), (4) с любой из формулы пересчета (8), (9), (10) и  $\gamma_k = \arg \min_y f(x^k + y p^k)$

$$\text{для } f(x) = (Ax, x)/2 - (b, x), \quad A > 0, \quad \text{будет конечным: } x^* = \gamma^* = A^{-1}b. \quad \blacktriangle$$

Более того, можно показать, что, несмотря на различие формул пересчета, последовательности  $x^k$ , генерируемые каждым вариантом метода, для квадратичной функции  $f(x)$  совпадают.

Для неквадратичных функций квазиньютоновские методы в записанной выше форме применимы, но они, естественно, перестают быть конечными. В связи с этим при  $k > n$  можно либо продолжать счет по этим же формулам, либо ввести процедуру обновления (заменять матрицу  $H_k$  на  $H_0$  через каждые  $n$  итераций).

В настоящее время доказана сверхлинейная (или квадратичная) скорость сходимости многих вариантов квазиньютоновских методов в окрестности невырожденной точки минимума.

Эти результаты выглядят естественными в свете утверждений лемм 1 и 2 и теоремы 1, однако их полное доказательство весьма громоздко.

Квазиньютоновские методы чрезвычайно популярны, им посвящен огромный поток работ. Такое внимание объясняется утомившимися выше достоинствами методов — они требуют лишь одного вычисления градиента на каждом шаге, в них не нужно обращать матрицу или решать систему линейных уравнений, они обладают глобальной сходимостью, в окрестности решения скорость сходимости высока (часто квадратична) и т. п. Однако они имеют и дефекты по сравнению, например, с методом сопряженных градиентов. Главный из них заключается в необходимости хранить и пересчитывать матрицу  $H_k$  размерности  $n \times n$ , что для больших  $n$  требует значительного объема памяти ЭВМ.

При численной проверке методов обычно наилучшие результаты дает вариант (10).

**Упражнение.**

1. Пусть  $e_1, \dots, e_n$  — координатные орты в  $\mathbf{R}^n$ ,  $f(x)$  — дифференцируема в окрестности точки  $x$  и дважды дифференцируема в  $x$ . Пусть  $H(\alpha)$  — матрица  $t$ -й строкой которой является  $\alpha^{-1}(\nabla f)(x + \alpha e_i) - \nabla f(x)$ . Докажите, что  $H(\alpha) \rightarrow \nabla^2 f(x)$  при  $\alpha \rightarrow 0$ .

**2. Методы переменной метрики и методы сопряженных направлений.** Выше квазиньютоновские методы были получены как приближения к методу Ньютона. Однако на них можно посмотреть и с другой точки зрения.

Выясним прежде всего, как влияет выбор метрики на вид и свойства градиентного метода. Пусть в пространстве  $\mathbf{R}^n$  наряду с исходным скалярным произведением  $(x, y)$  задано с помощью матрицы  $A \geq 0$  другое скалярное произведение

$$(x, y)_1 = (Ax, y). \quad (11)$$

В этом случае  $A$  задает новую метрику в  $\mathbf{R}^n$ :

$$\|x - y\|_1^2 = (A(x - y), x - y). \quad (12)$$

Выпишем градиент дифференцируемой функции  $f(x)$  в новой метрике:

$$\begin{aligned} f(x + y) &= f(x) + (\nabla f(x), y) + o(\|y\|) = \\ &= f(x) + (AA^{-1}\nabla f(x), y) + o(\|y\|) = f(x) + (a, y)_1 + o(\|y\|_1), \\ a &= A^{-1}\nabla f(x). \end{aligned}$$

В соответствии с определением вектор  $a$  есть градиент  $f(x)$  в пространстве со скалярным произведением (11). Итак,

$$\nabla f(x) = A^{-1}\nabla f(x). \quad (13)$$

В новой метрике градиентный метод приобретает вид

$$x^{k+1} = x^k - \gamma_k \nabla f(x^k) = x^k - \gamma_k A^{-1}\nabla f(x^k) \quad (14)$$

и отличается от исходного градиентного метода наличием матрицы  $A^{-1}$ . Иными словами, градиентный метод не инвариантен к выбору метрики пространства. Естественно попытаться выбрать метрику так, чтобы ускорить сходимость метода. Для квадратичной функции

$$f(x) = (Bx, x)/2 - (b, x) = (\frac{1}{2})(A^{-1}Bx, x)_1 - (A^{-1}b, x)_1, \quad (15)$$

скорость сходимости (14) определяется знаменателем прогрессии  $q = (L - l)/(L + l)$ , где  $L$ ,  $l$  — наибольшее и наименьшее собственные значения матрицы  $A^{-1}B$ . Чем ближе эта матрица к единичной, тем меньше  $q$ . Наилучший способ — выбрать  $A = B$ , тогда  $A^{-1}B = I$ ,  $q = 0$ , т. е. если задать метрику с помощью матрицы  $B$ , то градиентный метод (с  $\gamma_k \equiv 1$ ) даст точное решение за 1 шаг. Это не удивительно, так как в этой метрике  $f(x) =$

$= (1/2)(x, x)_1 - (A^{-1}b, x)_1$ , т. е. линии уровня  $f(x)$  — сферы, а обусловленность  $\mu$  равна единице.

Для неквадратичной функции метод

$$x^{k+1} = x^k - \gamma_k H_k \nabla f(x^k), \quad H_k > 0, \quad (16)$$

может рассматриваться как градиентный в метрике

$$(x, y)_1 = (H_k^{-1}x, y), \quad (17)$$

и «оптимальным» выбором метрики является  $H_k = [\nabla^2 f(x^k)]^{-1}$ . Иными словами, квазиньютоновские методы могут трактоваться как градиентные, в которых на каждом шаге выбирается новая метрика, по возможности близкая к наилучшей. В связи с этим часто употребляют термин *методы переменной метрики* как синоним квазиньютоновских методов.

Такая интерпретация полезна и как эвристический способ построения новых вариантов квазиньютоновских методов. Например, можно получить новую метрику путем «растяжения» пространства в направлении последнего градиента или в направлении разности двух последовательных градиентов и т. п. Мы остановимся на таких методах подробнее в гл. 5.

Другой подход к построению эффективных методов первого порядка связан с использованием понятия *сопряженных направлений*. Мы уже отмечали в § 2, что, зная набор сопряженных направлений  $p^1, \dots, p^n$ :

$$(Ap^i, p^j) = 0, \quad i \neq j,$$

можно найти минимум квадратичной функции  $f(x) = (Ax, x)/2 - (b, x)$  за  $n$  одномерных минимизаций:

$$x^{k+1} = x^k - \alpha_k p^k, \quad \alpha_k = \underset{a}{\operatorname{argmin}} f(x^k - ap^k). \quad (18)$$

Тогда при любом  $x^0$  будет  $x^n = x^* = A^{-1}b$ . Один способ построения сопряженных направлений использовался в методе сопряженных градиентов — в нем процессу А-ортогонализации подвергались последовательно вычисляемые градиенты. Однако возможны и другие способы.

Пусть  $p^1, \dots, p^k$ ,  $k < n$  — уже построенные сопряженные векторы,

$$(Ap^i, p^j) = 0, \quad 1 \leq i, j \leq k, \quad i \neq j, \quad (20)$$

а  $x^k$  — соответствующие им точки в методе (19). Следующий вектор  $p^{k+1}$  должен удовлетворять соотношению

$$(p^{k+1}, Ap^i) = 0, \quad i = 1, \dots, k.$$

Поскольку  $p^i = \alpha_i^{-1}(x^{i+1} - x^i)$ ,  $Ap^i = \alpha_i^{-1}(\nabla f(x^{i+1}) - \nabla f(x^i)) = \alpha_i^{-1}y^i$ , то это эквивалентно условию

$$(p^{k+1}, y^i) = 0, \quad i = 1, \dots, k, \quad (21)$$

Итак, новое сопряженное направление  $p^{k+1}$  должно удовлетворять условиям ортогональности (21). Подвергая такому процессу ортогонализации любой набор линейно независимых векторов, получим различные наборы сопряженных направлений.

Этот же процесс может быть применен к неквадратичной функции:

$$x^{k+1} = x^k - \alpha_k p^k, \quad \alpha_k = \underset{\alpha \geq 0}{\operatorname{argmin}} f(x^k + \alpha p^k), \quad (22)$$

$$(p^{k+1}, y^i) = 0, \quad i = 1, \dots, k, \quad y^i = \nabla f(x^{i+1}) - \nabla f(x^i).$$

Обычно при этом ищут  $p^{k+1}$  в виде

$$p^{k+1} = H_{k+1} \nabla f(x^{k+1}), \quad H_{k+1} = H_k + \Delta H_k \quad (23)$$

и вместо непосредственного запоминания векторов  $y^i, i = 1, \dots, k$ , запоминают матрицу  $H_k$ . Таким образом, методы принимают ту же форму (1), что и квазиньютоны. Разница лишь в том, что при этом не обязательно  $H_k \rightarrow [\nabla^2 f(x^k)]^{-1}$ ; в некоторых вариантах метода оказывается (для квадратичной функции)  $H_n = 0$ . Поэтому в таких методах обязательно должно осуществляться обновление.

Выпишем алгоритм одного из простейших методов данного класса:

$$x^{k+1} = x^k + \alpha_k p^k, \quad \alpha_k = \underset{\alpha \geq 0}{\operatorname{argmin}} f(x^k + \alpha p^k), \quad (24)$$

$$p^k = -H_k \nabla f(x^k), \quad y^k = \nabla f(x^{k+1}) - \nabla f(x^k),$$

$$H_{k+1} = H_k - \frac{H_k y^k (y^k)^T H_k}{(H_k y^k, y^k)}, \quad k+1 \neq n, 2n, \dots$$

$$H_0 = H_n = H_{2n} = \dots = I.$$

Оказывается, что для квадратичной функции в методе (24)  $p^k$  являются сопряженными направлениями,  $H_k \geq 0$  для всех  $k \leq n$ ,  $H_n = 0$ . Для неквадратичных функций доказана квадратичная локальная сходимость методов данного класса в окрестности невырожденного минимума.

**3. Метод секущих.** Одним из простейших и наиболее распространенных методов решения одномерного уравнения

$$g(x) = 0 \quad (25)$$

является *метод секущих*, сущность которого видна из рис. 7. Его можно обобщить на многомерный случай — если  $g: \mathbf{R}^n \rightarrow \mathbf{R}^n$ , то можно вычислить  $g$  в  $n+1$  точках, построить линейную аппроксимацию и найти ее корень, который является очередным приближением к решению (25).

Применимально к задаче минимизации  $f(x)$  в  $\mathbf{R}^n$ , т. е. к задаче решения уравнения  $\nabla f(x) = 0$ , метод принимает следующий вид. Пусть  $x_k, x_{k-1}, \dots, x_{k-n} = n+1$  точек в  $\mathbf{R}^n$ ,  $\nabla f(x_k), \dots$

$\dots, \nabla f(x^{k-n})$  — вычисленные в них градиенты. Решим систему  $n+1$  линейных уравнений с  $n+1$  переменными  $\lambda_0, \lambda_1, \dots, \lambda_n$ :

$$\sum_{i=0}^n \lambda_i \nabla f(x^{k-i}) = 0, \quad \sum_{i=0}^n \lambda_i = 1 \quad (26)$$

и построим точку

$$x^{k+1} = \sum_{i=0}^n \lambda_i x^{k-i}. \quad (27)$$

Далее процесс повторяется для  $n+1$  последних точек  $x^{k+1}, \dots, x^{k-n+1}$  и т. д. Нетрудно проверить, что для  $n=1$  такой метод совпадает с методом секущих для решения уравнения  $\nabla f(x) = 0$ .

**Теорема 2.** Если векторы  $x^1 - x^0, \dots, x^n - x^0$  линейно независимы, а  $f(x)$  квадратична с  $\nabla^2 f(x) \equiv A > 0$ , то  $x^{n+1}$  — точка минимума  $f(x)$ . ▲

В системе линейных уравнений (26) на каждой итерации меняется лишь один столбец, поэтому нет необходимости решать ее каждый раз заново, а можно воспользоваться следующим результатом.

**Лемма 4.** Пусть  $B$  — квадратная

матрица  $n \times n$  со столбцами  $b^1, \dots, b^n$ , а  $B$  отличается от ее первым столбцом ( $b^1$  заменено на  $b^1$ ). Тогда

$$\tilde{c}^i = c^i - \frac{(b^1 - b^1, c^i)}{1 + (b^1 - b^1, c^i)} c^1, \quad (28)$$

где  $c^i$  — строки  $B^{-1}$ ,  $\tilde{c}^i$  — строки  $\tilde{B}^{-1}$ .

При доказательстве достаточно представить  $\tilde{B}$  в виде  $\tilde{B} = B + (\tilde{b}^1 - b^1)e^T$ , где  $e = (1, 0, \dots, 0)$ , и воспользоваться леммой 3. ▲

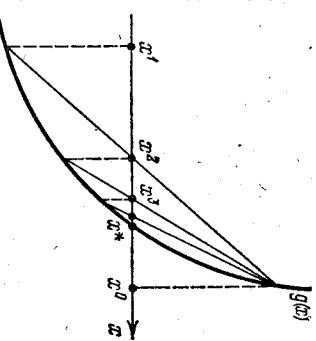


Рис. 7. Метод секущих.

Однако в описанной выше форме метод секущих не является удовлетворительным. Так, он не обладает свойством глобальной сходимости. Для устранения этого недостатка можно применять стандартные средства, например регулировку длины шага (из  $x^k$  делается шаг по направлению  $\sum_i \lambda_i x^{k-i}$ ). Вторым дефектом метода является его склонность к вырождению — в процессе счета последовательные приближения оказываются лежащими (приближенно) в подпространстве пространства  $\mathbf{R}^n$ . Соответствующая система линейных уравнений (26) плохо обусловлена и ее решение неустойчиво. Для преодоления этого недостатка

можно модифицировать метод с тем, чтобы система базисных точек была заведомо невырожденной. Например, можно добавлять на каждой итерации точку, делая шаг по координатным осям (в циклическом порядке). Для модифицированных подобным образом методов можно доказать сверхлинейную сходимость.

**4. Другие идеи построения методов первого порядка.** При всем разнообразии описанных выше алгоритмов первого порядка идея их оставалась одинаковой — использовать квадратичную аппроксимацию функции вблизи минимума. Как правило, эти алгоритмы конечны для квадратичных функций, а в общем случае их эффективность тем выше, чем ближе функции к квадратичной. Однако квадратичная модель может считаться естественной лишь в окрестности экстремума; вдали от него поведение минимизируемой функции может быть совсем иным. Поэтому для всех описанных выше методов отнюдь не гарантируется даже разумность стратегии оптимизации на начальных этапах поиска.

В связи с этим целесообразно использовать другие модели функций, отличные от квадратичной. На первый взгляд естественно попытаться строить полиномиальные модели на основе старших производных — следующих членов ряда Тейлора. Такие попытки делались, однако они вряд ли перспективны. Во-первых, прямое вычисление старших производных в многомерных задачах обычно требует слишком громоздких вычислений и большого объема памяти, а их восстановление по младшим производным предполагает вычисление последних в огромном числе точек. Во-вторых, решение вспомогательных задач минимизации полиномиальных функций, за редкими исключениями, не может быть осуществлено в аналитической форме.

Простой и важный класс представляют модели, основанные на аппроксимации функции однородной. Функция  $f(x)$ ,  $x \in \mathbb{R}^n$ , называется однородной относительно точки  $x^*$  с показателем  $\gamma > 0$ , если

$$f(x^* + \lambda(x - x^*)) - f(x^*) = \lambda^\gamma (f(x) - f(x^*)) \quad (29)$$

для всех  $x \in \mathbb{R}^n$  и  $\lambda \geq 0$ . Примеры однородных функций приведены в упражнениях 2—4 и 6.

Дифференцируемая однородная функция удовлетворяет важному соотношению

$$f(x) - f(x^*) = \gamma^{-1} (\nabla f(x), x - x^*). \quad (30)$$

Чтобы доказать (30), возьмем в (29)  $\lambda = 1 + \varepsilon$ :

$$f(x + \varepsilon(x - x^*)) - f(x^*) = (1 + \varepsilon)^\gamma (f(x) - f(x^*)),$$

$$\varepsilon^\gamma (f(x) - f(x^*)) = \varepsilon (\nabla f(x), x - x^*) + o(\varepsilon).$$

Устремляя  $\varepsilon$  к 0, получаем (30).

Точка  $x^*$  не обязательно является минимумом  $f(x)$  (см. примеры в упр. 2 и 3). Однако если  $f(x)$  достигает минимума, то  $x^*$  — точка глобального минимума  $f(x)$ . Действительно, пусть  $\tilde{x} = \tilde{x}^* = \min f(x)$ , тогда  $\nabla f(\tilde{x}) = 0$ . Подставляя  $\tilde{x}$  вместо  $x$  в (30), получаем, что  $f(x^*) = f(\tilde{x}) = \tilde{f}^*$ , т. е.  $x^*$  — точка глобального минимума. Именно этот случай и будет рассматриваться далее.

С помощью (30) можно найти точку минимума  $x^*$ , вычислив  $f(x)$  и  $\nabla f(x)$  в конечном числе точек. Действительно, если  $\gamma$  известно, то, взяв  $n+1$  точек  $x^0, \dots, x^n$ , мы получаем систему

$$\nabla f(x^i) - \alpha + (\nabla f(x^i), x^*) = (\nabla f(x^i), x^i), \quad i = 0, \dots, n, \quad (31)$$

линейную относительно  $n+1$  переменных  $x^*, \alpha$  ( $\alpha = \nabla f(x^*)$ ). Исключая переменную  $\alpha$ , получаем  $n$  линейных уравнений для определения  $x^* \in \mathbb{R}^n$ :

$$\begin{aligned} (\nabla f(x^i) - \nabla f(x^0), x^*) &= \\ &= (\nabla f(x^i), x^i) - (\nabla f(x^0), x^0) - \gamma (f(x^i) - f(x^0)), \quad i = 1, \dots, n. \end{aligned} \quad (32)$$

Если же  $\gamma$  неизвестно, то можно взять  $n+2$  точек  $x^0, \dots, x^{n+1}$  и определить  $n+1$  переменных  $\gamma, x^*$  из линейной системы (32), в которой следует взять  $n+1$  уравнений.

Аналогичный подход можно применить для минимизации функций общего вида подобно тому, как это делалось в методе секущих. В самом деле, пусть уже построены приближения  $x^0, \dots, x^k, k > n$ . Взяв последние  $n+1$  из них (или  $n+2$ , если  $\gamma$  неизвестно), решим систему (относительно  $x, \alpha, \gamma$ , либо  $x, \alpha$ )

$$(\nabla f(x^i), x) - \alpha + \gamma f(x^i) = (\nabla f(x^i), x^i), \quad i = k, k-1, \dots, \quad (33)$$

а решение  $x$  выберем в качестве  $x^{k+1}$ . Для  $\gamma = 2$  получаем метод, близкий к методу секущих, но отличающийся от него (в нем, в отличие от метода секущих, используются не только  $\nabla f(x^i)$ , но и значения функции  $f(x^i)$ ).

Такой процесс следует модифицировать с помощью тех же приемов, что и метод секущих (бороться с вырождением точек  $x^k$  путем добавления новых точек, линейно независимых от предыдущих, регулировать длину шага и т. д.). Полезно также сравнивать фактическое значение  $f(x^{k+1})$  с «предсказанным» (равным  $\alpha/\gamma$ ). Это может служить проверкой предположения о близости функции к однородной. При решении систем линейных уравнений целесообразно использовать близость этих уравнений на соседних итерациях (см. лемму 4).

Для минимизации однородных и близких к ним функций можно применять и другие методы. Так, в *градиентном методе* можно применять специальные способы выбора длины шага. Пусть функция  $f(x)$  удовлетворяет условию (30), причем вели-

чины  $f^* = f(x^*)$  и  $\gamma$  известны. Рассмотрим градиентный метод

$$x^{k+1} = x^k - \frac{\gamma(f(x^k) - f^*)}{\|\nabla f(x^k)\|^2} \nabla f(x^k). \quad (34)$$

Выбор шага  $\gamma_k = \frac{\gamma(f(x^k) - f^*)}{\|\nabla f(x^k)\|^2}$  здесь сделан так, чтобы для  $x^{k+1} = x^k - \gamma_k \nabla f(x^k)$  удовлетворялось равенство  $f(x^k) - f^* = \gamma^{-1}(\nabla f(x^k), x^k - x^{k+1})$  — ср. с (30). Тогда

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|x^k - x^*\|^2 - \frac{2\gamma(f(x^k) - f^*)}{\|\nabla f(x^k)\|^2} (\nabla f(x^k), x^k - x^*) + \\ &\quad + \frac{\gamma^2(f(x^k) - f^*)^2}{\|\nabla f(x^k)\|^2} = \|x^k - x^*\|^2 - \frac{\gamma^2(f(x^k) - f^*)^2}{\|\nabla f(x^k)\|^2}. \end{aligned}$$

Отсюда следует, что если  $\|\nabla f(x)\|$  ограничена на множестве  $\{x: \|x - x^*\| \leq \|x^0 - x^*\|\}$ , то  $f(x^k) \rightarrow f^*$ . Нетрудно видеть, что этот же результат остается справедливым, если в (30) равенство заменить на неравенство

$$f(x) - f^* \leq \gamma^{-1}(\nabla f(x), x - x^*). \quad (35)$$

Несколько иной класс (по сравнению с однородными) задается формулой

$$f(x) = F(\Phi(x)), \quad \Phi(x) = (Ax, x)/2 - (b, x), \quad A > 0, \quad (36)$$

где  $F: \mathbf{R}^l \rightarrow \mathbf{R}^l$  — монотонная на  $[f^*, \infty)$  функция,  $\Phi^* = \Phi(x^*) = \min \Phi(x)$ . Очевидно, что  $x^*$  является точкой минимума  $f(x)$ .

Если задан явный вид  $F$  и  $\Phi$ , то в соответствии с последним замечанием вместо минимизации  $f(x)$  можно решать более простую задачу минимизации  $\Phi(x)$ . Однако часто доступна меньшая информация о задаче. Тогда можно применить следующий *вариант метода сопряженных градиентов*:

$$\begin{aligned} x^{k+1} &= x^k + \alpha_k p^k, \quad \alpha_k = \arg \min_{\alpha \geq 0} f(x^k + \alpha p^k), \\ p^k &= -\nabla f(x^k) + \beta_k p^{k-1}, \\ \beta_k &= \frac{F'(\Phi(x^{k-1})) \|\nabla f(x^k)\|^2}{F'(\Phi(x^{k-1})) \|\nabla f(x^k)\|^2}, \quad \beta_0 = 0. \end{aligned} \quad (37)$$

Нетрудно проверить, что метод (37) порождает ту же последовательность точек, что и метод сопряженных градиентов для минимизации  $\Phi(x)$ , а потому является конечным.

Величину  $\rho_k = F'(\Phi(x^k))/F'(\Phi(x^{k-1}))$ , входящую в формулу для  $\beta_k$ , можно оценивать приближенно, аппроксимируя  $F(z)$  квадратичной или степенной функцией. При этом метод (37) можно применять и для минимизации функций, не обязательно имеющих вид (36).

В целом методы, основанные на однородных и близких к ним аппроксимациях функций, пока мало исследованы.

#### Упражнения.

2. Покажите, что аффинная функция  $f(x) = (a, x) - \beta$  однородна с  $\gamma = 1$  для любого  $x^*$ .

3. Проверьте, что квадратичная функция  $f(x) = (Ax, x)/2 - (b, x)$ , где  $A^{-1}$  существует, однородна относительно  $x^* = A^{-1}b$  с  $\gamma = 2$ .

4. Пусть существует решение  $x^*$  системы  $(a_i^t, x) = \beta_i$ ,  $i = 1, \dots, m$ ,  $x \in \mathbb{R}^n$ . Докажите, что функция  $f(x) = \sum_{i=1}^m |(a_i^t, x) - \beta_i|^\gamma$ ,  $\gamma > 0$ , однородна относительно  $x^*$  с показателем  $\gamma$ .

5. Докажите, что для дважды дифференцируемой однородной функции справедливо соотношение  $\nabla^2 f(x)(x - x^*) = (\gamma - 1) \nabla f(x)$ .

6. Покажите, что если  $\Phi^* = 0$ ,  $F(z) = |z|^\alpha$ ,  $\alpha > 0$ , то  $f(x)$  вида (36) — однородная относительно  $x^*$  с показателем  $2\alpha$ .

## § 4. Прямые методы

**1. Общая характеристика.** Во многих задачах минимизируемая функция задается с помощью некоторого алгоритма вычисления ее значений в произвольной точке. Вид алгоритма может быть неизвестен (например, вычисление значений функции производится либо с помощью модели, либо на реальном объекте) или он может быть столь сложен, что аналитическое вычисление градиента слишком громоздко. Во всех этих случаях единственная информация, которой мы располагаем, — значения  $f(x)$ .

Методы, использующие только эту информацию, называются *методами нулевого порядка* (часто говорят также о *прямых методах, методах поиска* или о *методах без вычисления производных*). Наиболее прямолинейная стратегия в такой ситуации заключается в использовании значений функции для *конечно-разностной аппроксимации производных* — градиента или гессiana. Более экономный способ связан с учетом значений функции в предыдущих точках. Наконец, известен ряд специфических методов нулевого порядка, не имеющих аналогов среди методов первого или второго порядков.

**2. Метод линейной аппроксимации.** Для оценки градиента функции  $f: \mathbf{R}^n \rightarrow \mathbf{R}^l$  в точке  $x$  составим конечно-разностные отношения

$$\Delta_1 = \alpha^{-1}[f(x + \alpha y) - f(x)], \quad \Delta_2 = (2\alpha)^{-1}[f(x + \alpha y) - f(x - \alpha y)], \quad (1)$$

где  $y \in \mathbf{R}^n$  — произвольный вектор.

**Лемма 1. а)** Если  $f$  дифференцируема в  $x$ , то

$$|\Delta_1 - (\nabla f(x), y)| \rightarrow 0 \text{ при } \alpha \rightarrow 0. \quad (2)$$

б) Если  $\nabla f$  удовлетворяет условию Липшица с константой  $L$  в окрестности  $x$ , то при достаточно малых  $\alpha$

$$|\Delta_1 - (\nabla f(x), y)| \leq L\alpha \|y\|^2/2. \quad (3)$$

в) Если  $f$  дифференцируема и  $\nabla^2 f$  удовлетворяет условию Липшица в окрестности  $x$ , то при достаточно малых  $\alpha$

$$|\Delta_2 - (\nabla f(x), y)| \leq L\alpha^2 \|y\|^3/6. \quad (4)$$

г) Если  $f(x)$  квадратична, то при любом  $\alpha$

$$\Delta_2 = (\nabla f(x), y). \quad (5)$$

Лемма 1 легко доказывается с использованием формул (2), (15), (20) § 1 гл. 1.  $\blacktriangleleft$

Таким образом, разностные отношения  $\Delta_1$  и  $\Delta_2$  могут служить приближением для линейной аппроксимации  $f(x)$ . Рассмотрим методы вида

$$x^{k+1} = x^k - \gamma_k s^k, \quad (6)$$

где  $\gamma_k \geq 0$  — длина шага, а  $s^k$  вычисляется по одной из двух формул

$$s^k = \sum_{i=1}^m \alpha_k^{-1} [f(x^k + \alpha_k h^i) - f(x^k)] h^i, \quad (7)$$

$$s^k = \sum_{i=1}^m (2\alpha_k)^{-1} [f(x^k + \alpha_k h^i) - f(x^k - \alpha_k h^i)] h^i. \quad (8)$$

Здесь  $h^i$ ,  $i = 1, \dots, m$ , — векторы, задающие направления промежуточных шагов,  $\alpha_k$  — длина пробного шага. Выбирая различные  $h^i$  и  $m$ , получим те или иные алгоритмы.

а) *Разностный аналог градиентного метода:*  $m = n$ ,  $h^i = e_i$ ,  $i = 1, \dots, n$ , где  $e_i$  — координатные орты. Иначе говоря, промежуточные шаги делаются по координатным осям, так что метод (6), (7) в координатной записи имеет вид

$$x_i^{k+1} = x_i^k - (\gamma_k/\alpha_k) [f(x^k + \alpha_k e_i) - f(x^k)]. \quad (9)$$

В соответствии с леммой 1

$$s^k = \sum_{i=1}^n (\nabla f(x^k), e_i) e_i + \varepsilon^k = \nabla f(x^k) + \varepsilon^k, \quad (10)$$

где остаточный член  $\varepsilon^k$  может быть оценен для каждой из формул (7), (8) в зависимости от гладкости  $f(x)$ .

б) *Метод покоординатного спуска:*  $m = 1$ ,  $h = e_i$ ,  $j = k(\text{mod } n)$ . Шаги делаются по координатным осям, выбираемым в циклическом порядке:

$$x_i^{k+1} = \begin{cases} x_i^k - (\gamma_k/\alpha_k) [f(x^k + \alpha_k e_i) - f(x^k)], & i = k(\text{mod } n), \\ x_i^k & \text{в противном случае.} \end{cases} \quad (11)$$

При этом  $s^k = \nabla f(x^k)_j e_j + \varepsilon^k$ .

в) *Метод случайного покоординатного спуска:*  $m = 1$ ,  $h = e_j$ , где  $j$  принимает значения  $1, \dots, n$  с равной вероятностью. Шаг делается, как и выше, по координатным осям, но они выбираются в случайном порядке.

г) *Метод случайного поиска:*  $m = 1$ ,  $h$  — случайный вектор, равномерно распределенный на единичной сфере. Здесь движение производится по случайному направлению, а знак и величина шага определяются разностным отношением:

$$x^{k+1} = x^k - (\gamma_k/\alpha_k) [f(x^k + \alpha_k h) - f(x^k)] h. \quad (12)$$

Сходимость всех методов гарантируется условием  $\alpha_k \rightarrow 0$  (см. упр. 1).

Скорость сходимости зависит от гладкости  $f(x)$  и способа выбора  $\alpha_k$ . С точки зрения погрешностей вычисления выгодно брать  $\alpha_k$  большим. Так как чем меньше  $\alpha_k$ , тем больше влияние ошибок округления при вычислении разностных отношений (в (1) приходится вычислять разность двух близких чисел и делить на малое число; это всегда связано с потерей точности). Однако для больших  $\alpha_k$  ухудшается связь с потерей точности (лемма 1). Можно показать, что в условиях теоремы 3 § 4 гл. 1 можно обеспечить в описанных выше методах сходимость со скоростью геометрической прогрессии, если  $\alpha_k \leq cq^k$ , где  $q < 1$  — некоторое число.

Вопрос о соотношении скоростей сходимости различных вариантов метода довольно сложен. Рассмотрим важный частный случай, который может служить моделью более реалистических ситуаций. Пусть  $f(x)$  квадратична:

$$f(x) = (Ax, x)/2 - (b, x), \quad A > 0, \quad (13)$$

а  $\gamma_k$  выбирается из условия скорейшего спуска:

$$x^{k+1} = x^k - \gamma_k s^k, \quad \gamma_k = \arg \min_{\gamma \geq 0} f(x^k - \gamma s^k). \quad (14)$$

Сравним три способа выбора  $s^k$ : *симметричная разностная аппроксимация градиента*

$$s^k = \sum_{i=1}^n (2\alpha)^{-1} [f(x^k + \alpha e_i) - f(x^k - \alpha e_i)] e_i = \nabla f(x^k) \quad (15)$$

(последнее равенство в силу (5)); *покоординатный спуск*

$$s^k = (2\alpha)^{-1} [f(x^k + \alpha h^k) - f(x^k - \alpha h^k)] h^k = (\nabla f(x^k), h^k) h^k, \quad (16)$$

где  $h^k$  — равномерно распределенный на единичной сфере вектор. Таким образом, (14), (15) совпадают с методом наискорей-

шего спуска ((4), § 1), а (14), (16) хорошо известен в линейной алгебре как *метод Гаусса—Зейделя*.

Соотношение скоростей сходимости методов зависит от различных причин: приведем несколько крайних случаев. Если  $A = I$ , то (14), (15) и (14), (16) приводят к решению за 1 шаг, тогда как метод случайного поиска сходится в средневквадратичном не быстрее некоторой геометрической прогрессии. Если

$$(Ax, x) = \sum_{i=1}^n \lambda_i x_i^2, \quad \lambda_i > 0, \quad \text{то метод (14), (16) конечен, тогда как}$$

$$(14), (15) — нет. Наконец, если задача плохо сбусловлена ( $\mu \gg 1$ ), то можно показать, что метод случайного поиска сходится быстрее градиентного (с учетом разницы в числе вычислений  $f(x)$  на одной итерации методов). Грубо говоря, для таких задач случайное направление в среднем лучше указывает на решение, чем антиградиент. Метод Гаусса—Зейделя имеет еще один резерв ускорения сходимости — если заменить в нем  $y_k$  на  $x_k$ ,  $1 < \alpha < 2$  (так называемая *сверхрелаксация*), то оказывается, что в ряде случаев сходимость резко улучшается.$$

В целом можно рекомендовать в классе поисковых методов описанного типа метод покоординатного спуска как по его простоте, так и по скорости сходимости.

#### Упражнения.

1. Докажите, что в условиях теоремы 1 § 4 гл. 1 при  $\alpha_k \rightarrow 0$ ,  $u_k = v$ , где  $\nabla f(x^*) \rightarrow 0$  п. н. Используйте технику доказательства теоремы 1 § 2 гл. 2.
2. Предложите по аналогии с (10) § 1 конструктивный способ регулировки  $\alpha_k$ , обеспечивающий линейную скорость сходимости.

#### 3. Нелокальная линейная аппроксимация.

В конечно-разностном градиентном методе (9) *пробные и рабочие шаги* были разделены — точки  $x^k + \alpha_k e_j$  служили только для оценки градиента в  $x^k$ , в  $x^{k+1}$  вся работа проводится заново. Можно поступить и иначе, и строить линейную аппроксимацию по набору точек, расположенных достаточно далеко.

Типичным примером служит так называемый *симплексный метод* (не путать с симплекс-методом в линейном программировании!). Пусть выбраны  $n+1$  точек  $x^0, x^1, \dots, x^n$ , образующие вершины правильного симплекса. Вычислим значения  $f(x)$  в вершинах и найдем ту, для которой  $f(x)$  максимальна:  $j = \arg\max_{0 \leq i \leq n} f(x^i)$ . Построим новый симплекс, отличающийся от старого лишь одной вершиной;  $x^j$  заменяется на  $x^{n+1}$ :

$$x^{n+1} = 2x^1(x^0 + \dots + x^{j-1} + x^{j+1} \dots + x^n) - x^j \quad (18)$$

(т. е.  $x^{n+1}$  симметрично с  $x^j$  относительно грани, противолежащей  $x^j$ ). Если окажется, что в новом симплексе максимум достигается в  $x^{n+1}$ , то возвращаемся к исходному симплексу, заменив  $x^j$  на вершину, в которой значение  $f(x)$  максимально среди

оставшихся вершин и т. д. Если какая-либо точка сохраняется в  $n+1$  последовательном симплексе, то последний симплекс сокращается вдвое подобным преобразованием с центром в этой вершине (рис. 8).

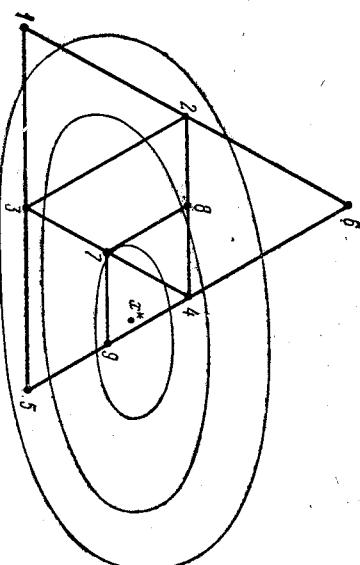


Рис. 8. Симплексный метод.

Мы описали лишь простейший вариант метода. Существует много его модификаций, в которых симплекс не обязательно правильный, а величина шага и условия дробления могут быть иными. С теоретической точки зрения подобные методы слабо исследованы. Практика показывает их работоспособность для не слишком плохо обусловленных задач.

**4. Квадратичная аппроксимация.** Вычислив значения  $f(x)$  в достаточном числе точек, можно построить квадратичную аппроксимацию  $f(x)$ . Удобно это сделать, например, следующим образом (*метод барицентрических координат*). Выбирается (как и в симплексном методе)  $n+1$  базисных точек  $x^0, \dots, x^n$ . Вычисляются значения функции во всех этих точках и серединах соединяющих их отрезков (обозначим  $f((x^i+x^j)/2) = f_{ij}$ ,  $f(x^i) = f_{ii}$ ,  $i, j = 0, \dots, n$ ). После этого решается система линейных (относительно  $\lambda_0, \lambda_1, \dots, \lambda_n$ ) уравнений

$$4 \sum_{j=0}^n f_{ij} \lambda_j + \lambda = f_{ii}, \quad i = 0, \dots, n, \quad (19)$$

$$\sum_{i=0}^n \lambda_i = 1$$

и строится точка

$$x^{n+1} = \sum_{i=0}^n \lambda_i x^i, \quad (20)$$

Нетрудно проверить, что если  $f$  квадратична, то  $x^{n+1} = x^* = A^{-1}b$  при любых  $x^0, \dots, x^n$  таких, что  $x^n - x^0, \dots, x^1 - x^0$  линейно независимы.

Далее (для неквадратичной  $f(x)$ ) точка  $x^{n+1}$  включается в число базисных, а одна из прежних базисных точек (точка  $x^0$  или та, в которой  $f(x)$  максимальна) удаляется. На следующей итерации достаточно вычислить  $f(x)$  в  $n+1$  точках (в  $x^{n+1}$  и серединах отрезков, соединяющих  $x^{n+1}$  с остальными базисными точками). Новая система уравнений для  $\lambda_i$  будет отличаться от (19) лишь одной строкой, так что можно использовать результат леммы 4 § 3 для построения решения. Аналогичным образом процесс продолжается дальше.

Удобство метода в том, что сама квадратичная аппроксимация функции не выписывается явно, строится лишь точка минимума этой аппроксимации. По сравнению с конечно-разностным аналогом метода Ньютона здесь существенно меньше вычислений  $f(x)$  на каждом шаге ( $n+1$  вместо  $n(n+1)/2$ ). Для признания устойчивости процессу в нем нужно ввести регулировку длины шага, принять меры для предотвращения вырождения системы базисных точек, проверять условие выпуклости  $f_{ii} \leq (f_{ii} + f_{jj})/2$  и т. п.

Другая группа методов прямого поиска использует идеи метода сопряженных направлений и сводит исходную задачу к

*последовательности одномерных минимаций*. В отличие от метода покоординатного спуска, где система направлений спуска жестко фиксируется (этой системой являются координатные орты), в данных

методах направления спуска строятся в процессе минимизации. Принцип их построения — сделать их (для задачи минимизации квадратичной функции) сопряженными; тогда, как мы знаем (см. § 2) процесс минимизации конечен в квадратичном случае. Основная идея методов этой группы иллюстрируется рис. 9 — три последовательные одномерные минимизации приводят в точку минимума. В многомерном пространстве верен аналогичный результат.

**Лемма 2.** Пусть  $f(x) = (Ax, x)/2 - (b, x)$ ,  $A > 0$ ,  $x \in \mathbf{R}^n$ ,  $p^1, \dots, p^k$  — сопряженные векторы:  $(Ap^i, p^j) = 0$ ,  $i \neq j$ ,  $k < n$ ,  $L^0 = \left\{ x: x = x^0 + \sum_{i=1}^k \lambda_i p^i \right\}$ ,  $x^1 \equiv L^0$ ,  $L^1 = \left\{ x: x = x^1 + \sum_{i=1}^k \lambda_i p^i \right\}$ ,  $y^0 = \underset{x \in L^0}{\operatorname{argmin}} f(x)$ ,  $y^1 = \underset{x \in L^1}{\operatorname{argmin}} f(x)$ . Тогда вектор  $p^{k+1} = y^1 - y^0$  является сопряженным с  $p^1, \dots, p^k$ .

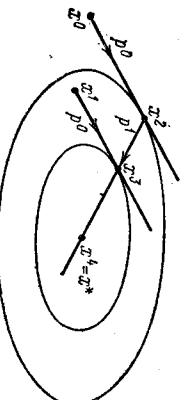


Рис. 9. Метод сопряженных направлений.

Этот результат следует из условия минимума  $f(x)$  на подпространстве (см. замечание после теоремы 2 § 2).  $\blacktriangle$

На этой основе можно построить метод минимизации, например, следующим образом. Пусть  $x_k$  — полученное на  $k$ -й итерации приближение к решению,  $p^0, \dots, p^k$  — найденные направления ( $x^0$  и  $p^0$  произвольны). Построим  $\bar{x}^k = x^k + h_k$ , где  $h_k$  — произвольный вектор, не являющийся линейной комбинацией  $p^0, \dots, p^k$ . Проведем цикл последовательных одномерных минимизаций по направлениям  $p^0, \dots, p^k$ , начиная из точки  $\bar{x}^k$ ; обозначим полученную в результате точку  $\hat{x}^{k+1}$ . В качестве  $x^{k+1}$  возьмем минимум  $f(x)$  на прямой, соединяющей  $\hat{x}^{k+1}$  с  $x^k$ , а в качестве  $p^{k+1}$  — вектор  $\hat{x}^{k+1} - x^k$ . Для квадратичной функции в  $\mathbf{R}^n$  такой метод Пауэлла приводит к минимуму не более чем за  $n$  шагов.

Существует и много других модификаций, основанных на подобной идее. Всего для отыскания минимума в квадратичном случае требуется  $n(n+1)/2$  одномерных минимизаций. Если считать, что каждая из них включает три вычисления функции, то видно, что метод менее экономен, чем (19), (20) (где нужно  $n(n+1)/2$  вычислений для той же цели). Однако в неквадратичном случае метод работоспособен даже для плохого начального приближения (если принять меры против вырождения системы  $p^i$ ), тогда как метод барицентрических координат подобно методу Ньютона требует хорошего начального приближения.

типа среднего риска:

$$f(x) = MQ(x, \omega) = \int Q(x, \omega) dP(\omega), \quad (1)$$

## Г л а в а 4

### ВЛИЯНИЕ ПОМЕХ

Цель этой главы — выяснить поведение методов безусловной минимизации дифференцируемых функций при наличии помех. Оказывается, что чувствительность методов к помехам различна.

Грубо говоря, чем эффективнее метод в идеальном случае (без помех), тем более чувствителен он к разного рода ошибкам. Можно модифицировать методы, сделав их работоспособными в условиях помех. При этом априорная информация о помехах (их уровнях, законах распределения и т. д.) может быть эффективно использована.

### § 1. ИСТОЧНИКИ И ТИПЫ ПОМЕХ

**1. Источники помех.** В реальных задачах применить методы гл. 1 и 3 «в чистом виде» нельзя — ситуация неизбежно осложняется наличием разного рода ошибок и погрешностей. Перечислим некоторые из причин их возникновения.

В простейшем случае, когда минимизируемая функция и ее градиент заданы формулами, ошибки возникают вследствие *погрешности вычисления*, связанных с округлением при выполнении арифметических действий на ЭВМ. В результате  $f(x^k)$ ,  $\nabla f(x^k)$  и т. д. вычисляются с некоторой ошибкой, т. е. вместо вектора  $\nabla f(x^k)$  мы получаем вектор  $s^k = \nabla f(x^k) + r^k$ . Здесь по меха  $r^k$  является детерминированной (ошибки округления в ЭВМ не носят случайного характера) и можно оценить ее уровень  $\|r^k\| \leq \varepsilon$ , так как законы образования погрешностей округления хорошо изучены. Величину  $\varepsilon$  обычно можно считать постоянной (не зависящей от  $x^k$ ) и, как правило, не слишком большой. В случае необходимости  $\varepsilon$  можно уменьшить, производя вычисления с двумя точностями.

В ряде задач значения  $f(x^k)$  и  $\nabla f(x^k)$  получаются не с помощью вычислений, а в *результате измерений*. Такова ситуация при оптимизации на реальном объекте (экстремальное регулирование, планирование эксперимента). Тогда помехи носят случайный характер, свойственный погрешностям измерений. При этом обычно бывает доступна информация об уровне и статистической природе помех.

Нередко (особенно в задачах адаптации, обучения, распознавания и т. д.) проблема оптимизации ставится следующим образом. Нужно минимизировать детерминированную функцию  $f(x)$ ,

где функция  $Q(x, \omega)$  известна, однако распределение  $P(\omega)$  не задано. Дана лишь выборка  $\omega_1, \dots, \omega_k$  из этого распределения. Тогда точное вычисление  $f(x)$  и  $\nabla f(x)$  в принципе невозможno. В качестве приближенного значения этих величин можно взять

$$\frac{1}{k} \sum_{i=1}^k Q(x, \omega_i) \quad \text{и} \quad \frac{1}{k} \sum_{i=1}^k \nabla_x Q(x, \omega_i), \quad (2)$$

или более просто

$$Q(x, \omega_k) \quad \text{и} \quad \nabla_x Q(x, \omega_k). \quad (3)$$

В этом случае значения функции и градиента содержат случайную помеху. Если брать в качестве приближений для  $f(x^k)$  и  $\nabla f(x^k)$  величины  $Q(x^k, \omega_k)$  и  $\nabla_x Q(x^k, \omega_k)$ , то помехи будут не-зависимы в различных точках.

Аналогичная ситуация возникает в *методе Монте-Карло*, когда задача заключается в минимизации  $f(x)$  вида (1) и распределение  $P(\omega)$  известно, однако вычисление интеграла (1) слишком трудоемко. Тогда можно точные значения  $f(x)$  и  $\nabla f(x)$  заменить выборочными значениями, как и выше.

В ряде задач ошибки возникают из-за того, что значения функции и градиента вычисляются по упрощенным или приближенным формулам. Нередко точное вычисление требует громоздкого расчета функций влияния, решения сложных вспомогательных задач, учета взаимодействия всех параметров и т. д. Все эти вычисления нецелесообразно (а иногда и невозможно) проводить полностью. Их упрощение и огрубление приводят к погрешностям в определении функции и градиента. Это так называемые *неустранимые погрешности*.

Наконец, во многих методах ошибки возникают не из-за приближенного вычисления функции или градиента, а из-за необходимости решения вспомогательных задач, которое не может быть осуществлено точно. Например, в методе Ньютона на каждом шаге нужно решать систему линейных уравнений, что необходимо сопряжено с ошибками; в методе сопряженных градиентов требуется проводить одномерную минимизацию, что также может быть сделано лишь приближенно и т. д. В таком случае говорят о *погрешностях метода*.

**2. Типы помех.** Как мы видели выше, ошибки при вычислении функции и градиента могут иметь различное происхождение и различную природу. Несколько упрощая реальную ситуацию, можно выделить следующие основные типы помех. Всюду ниже

речь идет о вычислении градиента, когда вместо точного значения  $\nabla f(x^k)$  нам доступен вектор

$$s^k = \nabla f(x^k) + r^k, \quad (4)$$

где  $r^k$  — помехи. Случай приближенного вычисления  $f(x)$  исследуется аналогично (см. § 4).

а) *Абсолютные детерминированные помехи* удовлетворяют условию

$$\|r^k\| \leq \varepsilon, \quad (5)$$

т. е. градиент вычисляется с заданной абсолютной ошибкой. Предполагается, что про помехи не известно ничего, кроме этого условия. В частности, вектор  $r^k$  может не являться случайным, либо он может быть коррелирован с предыдущими помехами и т. д. Такая ситуация характерна для погрешностей вычислений и систематических ошибок измерений.

б) *Относительные детерминированные помехи* удовлетворяют условию

$$\|r^k\| \leq \varepsilon / \|\nabla f(x^k)\|. \quad (6)$$

Иначе говоря, градиент вычисляется с относительной ошибкой. В остальном, как и выше, о природе  $r^k$  ничего не известно. Такие помехи возникают, например, при использовании приближенных формул, дающих фиксированную относительную ошибку.

в) *Абсолютные случайные помехи*. Предположим, что помехи  $r^k$  случайны, независимы при различных  $x$ , центрированы и имеют ограниченную дисперсию:

$$(7)$$

Помехи такого типа характерны для задач, в которых градиент отыскивается в результате измерений на реальном объекте (экстремальное регулирование, планирование эксперимента), а также для задач с функцией среднего риска (1).

1) *Относительные случайные помехи* обладают теми же свойствами, что и в п. в), однако их дисперсия убывает по мере приближения к точке минимума:

$$Mr^k = 0, \quad M\|r^k\|^2 \leq \sigma^2. \quad (8)$$

Разумеется, на практике часто встречаются и другие типы помех, например случайные помехи с *систематической ошибкой* ( $\|Mr^k\| \leq \varepsilon$ ) или случайные ограниченные помехи ( $Mr^k = 0$ ,  $\|r^k\| \leq \varepsilon$ ). Однако их можно рассматривать как комбинацию основных типов, описанных выше. Поэтому мы ограничимся этими наиболее важными классами помех. Иногда (особенно в теоретических работах) предполагают, что уровень помех  $\varepsilon_k$  зависит от номера итерации и  $\varepsilon_k \rightarrow 0$  при  $k \rightarrow \infty$ . Такое предпо-

ложение представляется не очень реалистичным. Впрочем, в некоторых случаях можно добиться его выполнения путем повышения точности вычислений и уменьшения погрешности метода.

## § 2. ГРАДИЕНТНЫЙ МЕТОД ПРИ НАЛИЧИИ ПОМЕХ

1. *Постановка задачи*. Рассмотрим градиентный метод минимизации дифференцируемой функции  $f(x)$  на  $\mathbf{R}^n$  в ситуации, когда градиент вычисляется с ошибкой:

$$x^{k+1} = x^k - \gamma_k s^k, \quad s^k = \nabla f(x^k) + r^k. \quad (1)$$

Относительно помех  $r^k$  будут делаться предположения об их принадлежности одному из классов, описанных в § 1. Функция  $f(x)$  будет предполагаться сильно выпуклой (с константой  $L$ ) и с градиентом, удовлетворяющим условию Липшица (с константой  $L$ ) — этот класс функций наиболее важен (см. гл. 1 и 3). Нас будет интересовать поведение обычного градиентного метода с  $\gamma_k = \gamma$  при наличии помех, а также вопрос о целесообразном выборе длины шага в условиях помех. Обоснование метода будет вестись с помощью общих теорем § 2 гл. 2.

**2. Абсолютные детерминированные помехи.** Теорема 1. Пусть  $\|r^k\| \leq \varepsilon$ ,  $\gamma_k = \gamma$ . Тогда найдется  $\gamma > 0$  такое, что при  $0 < \gamma < \gamma_\beta$  методе (1) будет

$$\|x^k - x^*\| \leq \rho + q^k \|x^0 - x^*\|, \quad (2)$$

где  $0 \leq q < 1$ ,  $\rho = \rho(\varepsilon) \rightarrow 0$  при  $\varepsilon \rightarrow 0$ ,  $x^*$  — точка минимума  $f(x)$ .

Доказательство. Введем функцию Ляпунова

$$V(x) = \frac{1}{2} (\|x - x^*\| - \frac{1}{l} \varepsilon)_+^2. \quad (3)$$

Используя результат упражнения 1, получаем

$$(\nabla V(x^k), s^k) = \left( \|x^k - x^*\| - \frac{1}{l} \varepsilon \right)_+ \frac{(\nabla f(x^k) + r^k, x^k - x^*)}{\|x^k - x^*\|} \geq \left( \|x^k - x^*\| - \frac{1}{l} \varepsilon \right)_+ (l \|x^k - x^*\| - \varepsilon) = 2lV(x^k),$$

$$\|s^k\|^2 = \|\nabla f(x^k) + r^k\|^2 \leq (L \|x^k - x^*\| + \varepsilon)^2 \leq a + bV(x^k) \leq a + (b/(2l)) (\nabla V(x^k), s^k),$$

где  $a, b$  — некоторые константы, причем  $a \rightarrow 0$  при  $\varepsilon \rightarrow 0$ . Применим теорему 4 § 2 гл. 2, получаем требуемый результат.  $\blacktriangle$

Как нетрудно проверить на примерах (упр. 2), оценка (2) не является завышенной. Таким образом, наличие аддитивных помех приводит к тому, что градиентный метод с постоянным  $\gamma$  перестает сходиться к точке минимума. Он дает лишь

возможность попасть в некоторую окрестность минимума, размежевой которой тем меньше, чем меньше уровень помех. Сходимость к этой окрестности происходит со скоростью геометрической прогрессии.

Мы не выписывали выше точных значений констант (величин  $\rho$ ,  $\gamma$ ,  $q$ ), интересуясь лишь качественной картиной процесса. В упражнении 2 эти значения указаны для случая квадратичной функции.

### Упражнение.

1. Докажите, что  $V(x)$  вида (3) дифференцируема,  $\nabla V(x) = (\|x - x^*\| - \varepsilon/l)_+ \|x - x^*\|^{-1} (x - x^*)$ ,  $\nabla V(x)$  удовлетворяет условию Липшица с константой 1. Напишите график  $V(x)$  для  $x \in \mathbb{R}^l$ .

2. Пусть  $f(x) = (Ax, x)/2 - (b, x)$ ,  $II \leq A \leq LI$ ,  $l > 0$ ,  $\|r^k\| \leq \varepsilon$ ,  $0 < \gamma < 2/l$ . Покажите, что тогда в методе (1) будет  $\|x^{k+1} - x^*\| \leq q \|x^k - x^*\| + \gamma\varepsilon$ ,  $q = \max\{|1 - \gamma L|, |1 - \gamma L|\}$ . Используя лемму 1 § 2 гл. 2, получите оценку  $\|x^k - x^*\| \leq \gamma\varepsilon/(1 - q) + q^k (\|x^0 - x^*\| - \gamma\varepsilon/(1 - q))$ . В частности, при  $\gamma = 2/(L + l)$  отсюда следует  $\|x^k - x^*\| \leq \frac{\varepsilon}{l} + \left(\|x^0 - x^*\| - \frac{\varepsilon}{l}\right) \left(\frac{L - l}{L + l}\right)^k$ . Проверьте на примере, что эта оценка не является завышенной. Исследуйте предельный случай  $\varepsilon = 0$ .

### 3. Относительные детерминированные помехи.

Теорема 2. Пусть  $\|r^k\| \leq \alpha \|\nabla f(x^k)\|$ ,  $\alpha < 1$ ,  $\gamma_k \equiv \gamma$ . Тогда найдется  $\bar{\gamma} > 0$  такое, что при  $0 < \gamma < \bar{\gamma}$  метод (1) сходится к  $x^*$  со скоростью геометрической прогрессии.

Доказательство. Возьмем в качестве функции Ляпунова  $V(x) = f(x) - f(x^*)$ . Тогда (см. леммы 1 и 3 § 4 гл. 1)

$$\begin{aligned} (\nabla V(x^k), s^k) &= (\nabla f(x^k), \nabla f(x^k) + r^k) \geq (1 - \alpha) \|\nabla f(x^k)\|^2 \geq \\ &\geq (1 - \alpha) 2lV(x^k), \end{aligned}$$

$$\|s^k\|^2 \leq \|\nabla f(x^k)\|^2 (1 + \alpha)^2 \leq 2(1 + \alpha)^2 LV(x^k).$$

Составляется применить теорему 4 § 2 гл. 2. ▲

Таким образом, градиентный метод устойчив к относительным ошибкам, если их уровень менее 100%. Причина этого очевидна — всякое направление, составляющее с антиградиентом острый угол, является направлением убывания  $f(x)$  и может быть использовано в качестве направления движения вместо градиента.

### 4. Абсолютные случайные помехи. Пусть помехи $r^k$ случайны, независимы, $\mathbf{M}r^k = 0$ и $\mathbf{M}\|r^k\|^2 \leq \sigma^2$ .

Теорема 3. Найдется  $\bar{\gamma} > 0$  такое, что при  $\gamma_k \equiv \gamma$ ,

то  $\mathbf{M}\|x^k - x^*\|^2 \rightarrow 0$ . Если же

$$\sum_{k=0}^{\infty} \gamma_k^2 < \infty, \quad \sum_{k=0}^{\infty} \gamma_k = \infty, \quad (6)$$

то  $x^k \rightarrow x^*$  п. н. Наконец, если  $\gamma_k = \gamma/k$ ,  $\gamma > (2l)^{-1}$ , то

$$\mathbf{M}(f(x^k) - f^*) \leq \frac{L\sigma^2\gamma^2}{2(2l\gamma - 1)k} + o\left(\frac{1}{k}\right). \quad (7)$$

Доказательство. Возьмем  $V(x) = f(x) - f^*$ . Тогда  $(\nabla V(x^k), \mathbf{M}s^k) = (\nabla f(x^k), \nabla f(x^k)) \geq 2lV(x^k)$ ,  $\mathbf{M}\|s^k\|^2 = \|\nabla f(x^k)\|^2 + \mathbf{M}\|r^k\|^2 \leq \sigma^2 + (\nabla V(x^k), \mathbf{M}s^k)$ .

Теперь остается воспользоваться теоремами 2—5 § 2 гл. 2. ▲

Мы увидим далее (теорема 4), что вышеприведенные оценки не завышены, поэтому теорема 3 дает основания для следующих выводов. Во-первых, обычный вариант градиентного метода ( $\gamma_k \equiv \gamma$ ) при наличии аддитивных случайных помех не сходится к точке минимума, а приводит лишь в окрестность минимума. Размеры этой области тем меньше, чем меньше  $\gamma$ . Во-вторых, выбирая убывающие  $\gamma_k$ , можно сделать метод сходящимся в том или ином вероятностном смысле (в среднем при  $\gamma_k \rightarrow 0$  и почти наверное при  $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$ ). В-третьих, скорость сходимости при этом довольно мелкана (порядка  $O(1/k)$ ). Как мы увидим в дальнейшем, более высокой скорости сходимости нельзя добиться ни при каком выборе  $\gamma_k$ .

Уточним теорему 3 для квадратичной функции и помех постоянного уровня. Итак, пусть

$$f(x) = (Ax, x)/2 - (b, x), \quad II \leq A \leq LI, \quad l > 0,$$

$$\mathbf{M}r^k = 0, \quad \mathbf{M}r^k(r^k)^T = \sigma^2 I. \quad (8)$$

Будем считать, что начальное приближение  $x^0$  случайно и симметрично распределено вокруг  $x^*$ :  $\mathbf{M}(x^0 - x^*)(x^0 - x^*)^T = \alpha I$ .

Теорема 4. При любом  $0 < \gamma < 2/l$ ,  $\gamma_k \equiv \gamma$  в методе (1) при условиях (8) для величины

$$U_k = \mathbf{M}(x^k - x^*)(x^k - x^*)^T \quad (9)$$

справедливы соотношения

$$U_k \rightarrow U_{\infty} = \gamma\sigma^2 A^{-1}(2I - \gamma A)^{-1}, \quad (10)$$

$$\|U_k - U_{\infty}\| \leq \|U_0 - U_{\infty}\| q^k, \quad q = \max\{(1 - \gamma l^2, (1 - \gamma L)^2) < 1. \quad (11)$$

$$\text{Если } \gamma_k = \gamma/k, \quad \gamma > (2l)^{-1}, \text{ то}$$

$$U_k = \frac{1}{k} B(\gamma) + o\left(\frac{1}{k}\right), \quad B(\gamma) = \gamma\sigma^2 \left(2A - \frac{1}{\gamma} I\right)^{-1}. \quad (12)$$

**Величина  $\|B(\gamma)\|$  минимальна при  $\gamma = 1/l$ ,**

$$\|U_k\| = \frac{1}{k} \frac{\sigma^2}{l^2} + o\left(\frac{1}{k}\right). \quad \blacktriangle \quad (13)$$

**5. Относительные случайные помехи.** Пусть помехи  $r^k$  такие же, как в предыдущем пункте, но их дисперсия удовлетворяет условию

$$\mathbf{M} \|r^k\|^2 \leq \alpha \|\nabla f(x)\|^2. \quad (14)$$

**Теорема 5.** *При любом  $\alpha$  существует  $\bar{\gamma}$  такое, что при  $\gamma_k \equiv \gamma$ ,  $0 < \gamma < \bar{\gamma}$  в методе (1) будет*

$$\mathbf{M} \|x^k - x^*\|^2 \leq cq^k, \quad q < 1. \quad \blacktriangle \quad (15)$$

Мы видим, что наличие случайных относительных помех любого уровня не приводит к нарушению сходимости.

Итак, в зависимости от типа помех их присутствие может либо сохранять, либо нарушать сходимость градиентного метода. Иногда сходимость можно восстановить за счет регулировки длины шага.

### § 3. Другие методы минимизации при наличии помех

**1. Метод Ньютона.** Вопрос о поведении метода Ньютона при наличии помех значительно более сложен, чем тот же вопрос для градиентного метода. Дело в том, что в этом методе может быть несколько источников помех (вычисление  $\nabla f(x)$ ,  $\nabla^2 f(x)$ , обращение  $\nabla^2 f(x)$ ) и их природа может быть различна (например, случайные ошибки в вычислении градиента и систематические в обращении матрицы). Мы не будем стараться рассмотреть все возможные ситуации, а остановимся на нескольких характерных примерах, интересуясь лишь качественным анализом процесса.

Пусть в результате всех вычислений (градиента, гессiana, решения системы линейных уравнений) получается вектор, отличающийся от истинного:

$$s^k = [\nabla^2 f(x^k)]^{-1} \nabla f(x^k) + r^k, \quad (1)$$

где  $r^k$  — помеха, и делается шаг

$$x^{k+1} = x^k - s^k. \quad (2)$$

Предположим, что помеха может содержать систематическую ошибку:

$$\|r^k\| \leq \varepsilon. \quad (3)$$

Как мы знаем, метод Ньютона сходится локально в некоторой области  $U$ . Ясно, что если  $\varepsilon$  больше диаметра  $U$ , то сходимости заведомо нет — при любом  $x^0$ , сколь угодно близком к  $x^*$ , про-

цесс выходит из  $U$ . Таким образом, возникает ситуация, которой не было в градиентном методе: при достаточно высоком уровне абсолютных помех метод Ньютона может вести себя бессмыслицальным образом (например,  $\|x^k - x^*\|$  может возрастать) при любом  $x^0$ .

Возникновение систематических ошибок в методе Ньютона неизбежно, даже если число обусловленности  $\mu$  точки минимума (§ 3 гл. 1) велико (а именно тогда применение метода Ньютона наиболее целесообразно), то матрица  $\nabla^2 f(x^k)$  оказывается плохо обусловленной. Поэтому результат решения системы линейных уравнений  $\nabla^2 f(x^k)z = \nabla f(x^k)$  для определения шага метода отличается от точного решения вследствие ошибок округления в ЭВМ. Это отличие (для плохо обусловленных систем) может быть значительным и приводит к развалу метода Ньютона.

Присутствие случайных или относительных ошибок не столь катастрофично, но может повлечь существенное замедление метода Ньютона. Пусть, например, требуется минимизировать квадратичную функцию

$$f(x) = (Ax, x)/2 - (b, x), \quad A > 0, \quad (4)$$

причем матрицы  $A$  и  $A^{-1}$  вычисляются точно, а градиент содержит случайную ошибку:

$$s^k = \nabla f(x^k) + r^k = Ax^k - b + r^k, \quad \mathbf{M} r^k = 0, \quad \mathbf{M} \|r^k\|^2 = \sigma^2. \quad (5)$$

Рассмотрим метод

$$x^{k+1} = x^k - \gamma_k A^{-1} s^k, \quad (6)$$

являющийся обобщением метода Ньютона за счет введения параметра  $\gamma_k$ . Как мы увидим в дальнейшем (теорема 1 § 5), этот метод ни при каком способе выбора  $\gamma_k$  не может сходиться быстрее, чем  $O(1/k)$ . Но скорость сходимости такого же порядка может обеспечить гораздо более простой градиентный метод. Таким образом, здесь теряется основное преимущество метода Ньютона — его высокая скорость сходимости. Аналогичная ситуация возникает при наличии относительной ошибки. Если, например, градиент вычисляется с относительной ошибкой, то метод Ньютона может сходиться лишь со скоростью геометрической прогрессии.

Лишь при высокой точности вычислений метод Ньютона сохраняет свои преимущества (см. упр. 1).

#### Упражнение.

1. Докажите следующий результат. Пусть  $r^k$  в (1) удовлетворяет условию

$$\|r^k\| \leq c \|\nabla f(x^k)\|^2, \quad (7)$$

а к  $f'(x)$  применима теорема 1 § 5 гл. 1. Тогда при достаточно малом  $c$  метод (2) локально сходится с квадратичной скоростью.

**2. Многошаговые методы.** Ограничимся вновь анализом некоторых характерных частных случаев. Начнем с *метода тяжелого шарика*. Можно показать, что при наличии абсолютных детерминированных помех в определении градиента он сходится в область вокруг минимума. Громоздкая выкладка показывает, что для квадратичной функции размер этой области, вообще говоря, больше, чем для градиентного метода. Приведем аналогичный результат, относящийся к абсолютным случайным помехам. Пусть

$$\mathbf{f}(\mathbf{x}) = (\mathbf{A}\mathbf{x}, \mathbf{x})/2 - (\mathbf{b}, \mathbf{x}), \quad U \leq A \leq L, \quad l > 0, \quad (8)$$

$\mathbf{s}^k = \nabla f(\mathbf{x}^k) + \mathbf{r}^k = \mathbf{A}\mathbf{x}^k - \mathbf{b} + \mathbf{r}^k, \quad \mathbf{M}\mathbf{r}^k = 0, \quad \mathbf{M}\mathbf{r}^k(\mathbf{r}^k)^T = \sigma I,$  причем помехи  $\mathbf{r}^k$  взаимно независимы. Как можно показать, метод тяжелого шарика с постоянными коэффициентами

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha \mathbf{s}^k + \beta (\mathbf{x}^k - \mathbf{x}^{k-1}) \quad (9)$$

в такой ситуации не сходится к  $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$ , а приводит лишь в область вокруг  $\mathbf{x}^*$ . Поэтому рассмотрим метод с переменными коэффициентами, который удобно записать в форме

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha_k \mathbf{y}^k, \quad \mathbf{y}^{k+1} = \mathbf{y}^k - \beta_k (\mathbf{y}^k - \mathbf{s}^k). \quad (10)$$

Наряду с ним рассмотрим градиентный метод

$$(11)$$

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \mathbf{V}_k \mathbf{s}^k.$$

Ограничимся коэффициентами вида

$$\alpha_k = \frac{1}{k} \alpha, \quad \beta_k = \frac{1}{k} \beta, \quad \mathbf{V}_k = \frac{1}{k} \mathbf{V}. \quad (12)$$

**Теорема 1.** При любом выборе  $\alpha, \beta$  метод (10), (12) сходится асимптотически к быстрее (в смысле величины  $\|\mathbf{M}(\mathbf{x}^k - \mathbf{x}^*)(\mathbf{x}^k - \mathbf{x}^*)^T\|$ ), чем метод (11) с  $\mathbf{V}_k = 1/(k)$ . ▲

Таким образом, метод тяжелого шарика, превосходящий градиентный метод по скорости сходимости для задач без помех, является относительно менее эффективным при наличии помех. Этот вывод относится только к асимптотическому поведению метода. На начальных итерациях, когда относительная величина помех мала, двухшаговый метод может превосходить одношаговый, как и для задач без помех.

Примерно такова же ситуация с методом сопряженных градиентов. Полный анализ его поведения при наличии помех очень сложен. При этом разные его варианты по-разному реагируют на ошибки. По-видимому, наиболее устойчивы к попрещностям формулы (13), (14) из § 2 гл. 3, несколько менее — формулы (23) и (24) § 2 гл. 3. Можно показать, что при абсолютных и относительных помехах метод сопряженных градиентов вблизи минимума теряет преимущества перед градиентным. Лишь если

помехи удовлетворяют условию типа (7), то метод сопряженных градиентов сохраняет свои достоинства.

**3. Другие методы.** Квазиньютоновские методы очень чувствительны к ошибкам вычисления градиента. Действительно, в них восстанавливается матрица  $A = \nabla^2 f(\mathbf{x})$  по измерениям гравиента:

$$\Delta p^i \approx y^i, \quad p^i = x^{i+1} - x^i, \quad y^i = \nabla f(x^{i+1}) - \nabla f(x^i), \quad i = 0, \dots, k-1. \quad (13)$$

Если шаги малы ( $x^{i+1}$  близко к  $x^i$ ), а измерения  $\nabla f(x^i)$  содержат ошибки, то матрица восстанавливается плохо. Для задач со случайными аддитивными помехами с этим эффектом можно бороться путем увеличения числа измерений — нужно восстанавливать не по  $n$  значениям  $\nabla f(x)$ , как в детерминированном случае, а по  $N > n$  замерам. При этом можно выписать рекуррентные формулы, аналогичные приведенным в § 3 гл. 3. Для неподчиненных помех такой прием, вообще говоря, не приводит к повышению точности.

Совершенно аналогичные замечания относятся и к методу секущих — чтобы сделать его работоспособным при наличии случайных помех, нужно брать число базисных точек заметно большее, чем размерность пространства.

Однако нужно помнить, что возможности всех методов, основанных на квадратичной аппроксимации, весьма ограничены в задачах с помехами — даже знание точной матрицы вторых производных не спасает положения (см. анализ для метода Ньютона в § 2).

## § 4. ПРЯМЫЕ МЕТОДЫ

**1. Постановка задачи.** Пусть в произвольной точке  $\mathbf{x}^k$  имеется значение  $f(\mathbf{x}^k)$  с ошибкой  $\eta_k$ . По-прежнему будем говорить об абсолютной (относительной) детерминированной ошибке, если  $|\eta_k| \leq \varepsilon$  ( $|\eta_k| \leq \alpha (f(\mathbf{x}^k) - f(\mathbf{x}^*))$ ), и об абсолютной (относительной) случайной ошибке, если  $\eta_k$  случаи, независимы,  $\mathbf{M}\eta_k = 0$  и  $\mathbf{M}\eta_k^2 \leq \sigma^2$  ( $\mathbf{M}\eta_k^2 \leq \tau (f(\mathbf{x}^k) - f(\mathbf{x}))$ ). Задача заключается в изучении влияния разного рода ошибок на прямые методы минимизации (§ 4 гл. 3) и в модификации этих методов для преодоления влияния помех.

**2. Разностные методы при случайных помехах.** Рассмотрим методы типа приведенных в § 4 гл. 3 в ситуации со случайными помехами. Начнем с наиболее типичного примера — *метода Кифера — Вольфовича* (метода разностной аппроксимации гравиента):

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \mathbf{V}_k \mathbf{s}^k, \quad \mathbf{s}^k = \sum_{i=1}^n \frac{1}{2\alpha_k} (\tilde{f}(\mathbf{x}^k + \alpha_k e_i) - \tilde{f}(\mathbf{x}^k - \alpha_k e_i)) e_i, \quad (1)$$

$e_i$  — координатные орты. Здесь и далее

$$\begin{aligned} \tilde{f}(x) &= f(x) + \eta, \\ \text{причем случайные ошибки } \eta \text{ независимы в различных точках и} \end{aligned} \quad (2)$$

$$\begin{aligned} M\eta = 0, \quad M\eta^2 \leq \sigma^2. \\ \text{Обсудим вопрос о выборе пробных и рабочих шагов } \alpha_k, \gamma_k. \end{aligned} \quad (3)$$

где помехи  $\eta$  независимы в различных точках. Составим метод Кифера — Вольфовича (градиентный)

$$\begin{aligned} x^{k+1} &= x^k - \gamma_k s^k, \\ s^k &= \sum_{i=1}^n \frac{1}{2\alpha_k} [\tilde{f}(x^k + \sigma_k e_i) - \tilde{f}(x^k - \sigma_k e_i)] e_i \end{aligned} \quad (8)$$

и метод случайного поиска

$$x^{k+1} = x^k - \gamma_k s^k, \quad (9)$$

где  $h^k$  — случайный вектор, равномерно распределенный на единичной сфере (и не зависящий от  $\eta$ ). Поскольку для квадратичной функции систематическая ошибка в разностной аппроксимации градиента равна 0 при любом  $\alpha_k$  (лемма 1 § 4 гл. 3), здесь не нужно стремить  $\alpha_k$  к 0. Будем считать, что в (8) и (9)  $\alpha_k \equiv \alpha > 0$ . Используя теорему 4 § 2, нетрудно доказать, что в методе (8) при  $\gamma_k = \gamma/k$ ,  $\gamma \geq 1/(2L)$

$$M(x^k - x^*)^T (x^k - x^*)^T = \frac{1}{k} \cdot \frac{\gamma \sigma^2}{2\alpha^2} \left( 2A - \frac{1}{\gamma} I \right)^{-1} + o\left(\frac{1}{k}\right), \quad (10)$$

а в методе (9) при  $\gamma_k = \gamma/k$ ,  $\gamma > n/(2L)$

$$M(x^k - x^*)^T (x^k - x^*)^T = \frac{1}{k} \cdot \frac{\gamma \sigma^2}{2\alpha^2} \left( 2A - \frac{n}{\gamma} I \right)^{-1} + o\left(\frac{1}{k}\right). \quad (11)$$

Отсюда следует, что если брать  $\gamma_k$  в (8) в  $n$  раз больший, чем в (9), то  $n$  шагов метода (9) будут асимптотически эквивалентны одному шагу метода (8). Учитывая, что трудоемкость метода (8) в  $n$  раз больше, чем метода (9), получаем, что в данной ситуации методы (8) и (9) эквивалентны по их асимптотической эффективности. Любопытно, что этот вывод не зависит от обусловленности или каких-либо других свойств  $A$  (ср. с иной ситуацией в задачах без помех в § 4 гл. 3).

Отметим в заключение, что к асимптотическим оценкам типа приведенных в теореме 1, следует относиться с большой осторожностью. Например, выбор  $\alpha_k = \alpha k^{-1/6}$  означает, что нужно сделать миллион итераций, чтобы уменьшить пробный шаг в 10 раз. Поэтому практически счет будет происходить при постоянном  $\alpha_k$ .

**Упражнения.**

1. Покажите, что среди  $\alpha_k, \gamma_k$  вида  $\alpha_k = \alpha k^p$ ,  $\gamma_k = \gamma k^r$  в условиях теоремы 1 наилучшим в смысле асимптотических оценок скорости сходимости выбором 1 является приведенный в теореме:  $r = -1$ ,  $p = -1/6$ .
2. Сформулируйте аналог теоремы 1 для несимметричной разностной аппроксимации градиента и в предположении, что  $\nabla f(x)$  удовлетворяет условию Липшица. Покажите, что в этом случае наилучший выбор параметров таков:  $\gamma_k = \gamma/k$ ,  $\alpha_k = \alpha/k^{1/4}$ , при этом  $M\|x^k - x^*\|^2 = O(1/k^{1/2})$ .

$$\begin{aligned} \tilde{f}(x) &= f(x) + \eta, \\ \text{причем случайные ошибки } \eta \text{ независимы в различных точках и} \end{aligned} \quad (2)$$

$$\begin{aligned} M\eta = 0, \quad M\eta^2 \leq \sigma^2. \\ \text{Обсудим вопрос о выборе пробных и рабочих шагов } \alpha_k, \gamma_k. \end{aligned} \quad (3)$$

где помехи  $\eta$  независимы в различных точках. Составим метод Кифера — Вольфовича (градиентный)

$$\begin{aligned} x^{k+1} &= x^k - \gamma_k s^k, \\ s^k &= \sum_{i=1}^n \frac{1}{2\alpha_k} [\tilde{f}(x^k + \sigma_k e_i) - \tilde{f}(x^k - \sigma_k e_i)] e_i \end{aligned} \quad (8)$$

и метод случайного поиска

$$x^{k+1} = x^k - \gamma_k s^k, \quad (9)$$

где  $h^k$  — случайный вектор, равномерно распределенный на единичной сфере (и не зависящий от  $\eta$ ). Поскольку для квадратичной функции систематическая ошибка в разностной аппроксимации градиента равна 0 при любом  $\alpha_k$  (лемма 1 § 4 гл. 3), здесь не нужно стремить  $\alpha_k$  к 0. Будем считать, что в (8) и (9)  $\alpha_k \equiv \alpha > 0$ . Используя теорему 4 § 2, нетрудно доказать, что в методе (8) при  $\gamma_k = \gamma/k$ ,  $\gamma \geq 1/(2L)$

$$M(x^k - x^*)^T (x^k - x^*)^T = \frac{1}{k} \cdot \frac{\gamma \sigma^2}{2\alpha^2} \left( 2A - \frac{1}{\gamma} I \right)^{-1} + o\left(\frac{1}{k}\right), \quad (10)$$

а в методе (9) при  $\gamma_k = \gamma/k$ ,  $\gamma > n/(2L)$

$$M(x^k - x^*)^T (x^k - x^*)^T = \frac{1}{k} \cdot \frac{\gamma \sigma^2}{2\alpha^2} \left( 2A - \frac{n}{\gamma} I \right)^{-1} + o\left(\frac{1}{k}\right). \quad (11)$$

Отсюда следует, что если брать  $\gamma_k$  в (8) в  $n$  раз больший, чем в (9), то  $n$  шагов метода (9) будут асимптотически эквивалентны одному шагу метода (8). Учитывая, что трудоемкость метода (8) в  $n$  раз больше, чем метода (9), получаем, что в данной ситуации методы (8) и (9) эквивалентны по их асимптотической эффективности. Любопытно, что этот вывод не зависит от обусловленности или каких-либо других свойств  $A$  (ср. с иной ситуацией в задачах без помех в § 4 гл. 3).

Отметим в заключение, что к асимптотическим оценкам типа приведенных в теореме 1, следует относиться с большой осторожностью. Например, выбор  $\alpha_k = \alpha k^{-1/6}$  означает, что нужно сделать миллион итераций, чтобы уменьшить пробный шаг в 10 раз. Поэтому практически счет будет происходить при постоянном  $\alpha_k$ .

**Упражнения.**

1. Покажите, что среди  $\alpha_k, \gamma_k$  вида  $\alpha_k = \alpha k^p$ ,  $\gamma_k = \gamma k^r$  в условиях теоремы 1 наилучшим в смысле асимптотических оценок скорости сходимости выбором 1 является приведенный в теореме:  $r = -1$ ,  $p = -1/6$ .
2. Сформулируйте аналог теоремы 1 для несимметричной разностной аппроксимации градиента и в предположении, что  $\nabla f(x)$  удовлетворяет условию Липшица. Покажите, что в этом случае наилучший выбор параметров таков:  $\gamma_k = \gamma/k$ ,  $\alpha_k = \alpha/k^{1/4}$ , при этом  $M\|x^k - x^*\|^2 = O(1/k^{1/2})$ .

**3. Другие методы.** Для задач с помехами перестают быть работоспособными все методы, построенные на одномерных минимизациях (например, методы сопряженных направлений § 4 гл. 3), поскольку такую минимизацию нельзя осуществить. Более перспективными являются методы, в которых строится локальная аппроксимация функции по ее значениям в ряде точек (типа симплексного поиска или метода барицентрических координат, см. § 4 гл. 3). Влияние помех оказывается в том, что эти методы перестают работать в окрестности минимума, где уровень помех сравним с приращениями функции. Если помехи случайны и центрированы, то методы можно модифицировать так, что они останутся работоспособными и в указанной области. Общая идея такой модификации — использовать большее число точек для построения аппроксимации функции, чем в детерминированном случае. Это позволяет усреднять помехи и получать все более точную аппроксимацию. Например, в симплексном методе можно многократно проводить вычисления функции в каждой вершине симплекса, сопоставляя точность оценки значений функции с их разностью в различных вершинах.

Более экономный способ заключается в пересчете аппроксимации после каждого нового измерения. Опишем лишь схему подобных методов на упрощенной модели. Пусть можно предполагать, что функция  $f(x)$ ,  $x \in \mathbb{R}^n$ , аффинна в некоторой области:  $f(x) \approx (a, x) + \beta$ , и уже вычислены ее значения с помехой в  $k$  ( $k \geq n+1$ ) точках:  $y_i = (a, x^i) + \beta + \eta_i$ ,  $i = 1, \dots, k$ , где  $\eta_i$  — случайные независимые помехи,  $M_{\eta_i} = 0$ ,  $M_{\eta_i^2} = \sigma^2$ . Рассмотрим  $(n+1)$ -мерные векторы  $z^i = \{x^i, 1\}$ ,  $c^* = \{a, \beta\}$  и запишем измерения в виде  $y_i = (c^*, z^i) + \eta_i$ . Найдем оценку для  $c^*$  методом наименьших квадратов, т. е.

$$c^k = \underset{c}{\operatorname{argmin}} \sum_{i=1}^k (y_i - (c, z^i))^2 = \left( \sum_{i=1}^k z^i (z^i)^T \right)^{-1} \left( \sum_{i=1}^k z^i y_i \right) = \\ = \Gamma_k \sum_{i=1}^k z^i y_i, \quad \Gamma_k = \left( \sum_{i=1}^k z^i (z^i)^T \right)^{-1}. \quad (12)$$

Этому методу можно придать рекуррентную форму — новое измерение в точке  $x^{k+1}$ :  $y_{k+1} = (c^*, z^{k+1}) + \eta_{k+1}$ ,  $z^{k+1} = \{x^{k+1}, 1\}$ , может быть учтено с помощью следующей формулы:

$$c^{k+1} = c^k - \Gamma_{k+1} z^{k+1} ((c^k, z^{k+1}) - y_{k+1}), \quad (13)$$

$$\Gamma_{k+1} = \Gamma_k - \frac{\Gamma_k z^{k+1} (\Gamma_k z^{k+1})^T}{1 + (\Gamma_k z^{k+1}, z^{k+1})}, \quad k \geq n+1,$$

$$\Gamma_{n+1} = \left( \sum_{i=1}^{n+1} z^i (z^i)^T \right)^{-1}.$$

Таким образом, на каждом шаге не нужно заново вычислять оценку для аппроксимирующей функции, решая систему линейных уравнений (12), а достаточно использовать простую рекуррентную формулу (13). Оценка  $c^k$  может быть использована для реализации шага спуска:  $x^{k+1} = x^k - \gamma_k a^k$ ,  $c^k = \{a^k, \beta_k\}$ , и проверки согласованности линейной модели функции с измерениями. Разумеется, в реальных задачах линейная модель функции правомерна лишь локально, и метод минимизации должен включать «забывание» информации, полученной на ранних итерациях.

Совершенно аналогичные способы могут быть применены для восстановления квадратичной аппроксимации функции по результатам измерений, содержащих случайную ошибку.

### § 5. Оптимальные методы при наличии помех

**1. Потенциальные возможности итеративных методов при наличии помех.** Для детерминированных «невозмущенных» задач, как мы видели, существует множество методов, каждому из которых присуща своя скорость сходимости. Так, для гладких сильно выпуклых функций метод тяжелого шарика сходится быстрее градиентного, метод сопряженных градиентов — еще более быстро и т. д. Вопрос об оптимальном в смысле скорости сходимости методе здесь весьма сложен. Оказывается, наличие помех в определенном смысле упрощает ситуацию — оно ограничивает возможности любых методов минимизации. В этом случае существует некая предельная скорость сходимости, которая не может быть превзойдена. Тот метод, для которого эта предельная скорость достигается, естественно считать оптимальным.

Начнем с результатов, устанавливающих *потенциальные возможности* по скорости сходимости произвольных итеративных алгоритмов (не обязательно связанных с минимизацией) при наличии случайных помех. Рассмотрим итерационный процесс в  $\mathbb{R}^n$ :

$$x^{k+1} = x^k - \gamma_k s^k, \quad s^k = R(x^k) + g^k, \quad (1)$$

где  $\gamma_k \geq 0$  — детерминированные скалярные множители,  $R(x)$  — некоторая функция, а  $g^k$  — случайные помехи, предполагающиеся независимыми и центрированными ( $M_{g^k} = 0$ ). Начальное приближение  $x^0$  может быть либо детерминированным, либо случайнym, в последнем случае предполагается, что  $M_{|x|^2} < \infty$  и  $x^0, g^k$  независимы. Предположим, что существует единственная точка  $x^*$  такая, что  $R(x^*) = 0$  и  $R(x)$  удовлетворяет условию линейного роста:

$$\|R(x)\| \leq L \|x - x^*\|.$$

**Теорема 1.** Пусть для всех  $k$

$$\mathbf{M} \|\xi^k\|^2 \geq \sigma^2. \quad (3)$$

Тогда при сделанных выше предположениях для любого метода (1)

$$\mathbf{M} \|x^k - x^*\|^2 \geq 1/(a + kb), \quad a = 1/\mathbf{M} \|x^0 - x^*\|^2, \quad b = L^2/\sigma^2. \quad (4)$$

Подчеркнем, что в этой теореме, в отличие от любых теорем сходимости, приводившихся ранее, даются оценки скорости сходимости не сверху, а снизу. Теорема относится к любому способу выбора  $\gamma_k$  — в частности, и такому, для которого сходимость не имеет места.

**Доказательство.** Оценим условное математическое ожидание

$$\mathbf{M} (\|x^{k+1} - x^*\|^2 | x^k);$$

$$\mathbf{M} (\|x^{k+1} - x^*\|^2 | x^k) = \|x^k - x^* - \gamma_k R(x^k)\|^2 + \gamma_k^2 \mathbf{M} \|\xi^k\|^2,$$

$$\begin{aligned} \|x^k - x^* - \gamma_k R(x^k)\|^2 &\geq (\|x^k - x^*\| - \gamma_k \|R(x^k)\|)_+^2 \\ &\geq (\|x^k - x^*\| - \gamma_k L \|x^k - x^*\|)_+, \end{aligned}$$

$$\mathbf{M} (\|x^{k+1} - x^*\|^2 | x^k) \geq (1 - \gamma_k L)_+^2 \|x^k - x^*\|^2 + \gamma_k^2 \sigma^2.$$

Отсюда

$$\mathbf{M} \|x^{k+1} - x^*\|^2 \geq (1 - \gamma_k L)_+^2 \mathbf{M} \|x^k - x^*\|^2 + \gamma_k^2 \sigma^2. \quad (5)$$

Стоящая справа кусочно-квадратичная функция достигает минимума по  $\gamma_k$  при  $\gamma_k^* = L \mathbf{M} \|x^k - x^*\|^2 / (L^2 \mathbf{M} \|x^k - x^*\|^2 + \sigma^2)$ . Отсюда получаем

$$\begin{aligned} \mathbf{M} \|x^{k+1} - x^*\|^2 &\geq (1 - \gamma_k^* L)_+^2 \mathbf{M} \|x^k - x^*\|^2 + (\gamma_k^*)^2 \sigma^2 = \\ &= \sigma^2 \mathbf{M} \|x^k - x^*\|^2 / (L^2 \mathbf{M} \|x^k - x^*\|^2 + \sigma^2), \end{aligned}$$

или, обозначая  $u_k = 1/(\mathbf{M} \|x^k - x^*\|^2)$ ,  $u_{k+1} \leq L^2/\sigma^2 + u_k$ . Таким образом,  $u_k \leq u_0 + k L^2/\sigma^2$ , т. е.

$$\mathbf{M} \|x^k - x^*\|^2 \geq [1/\mathbf{M} \|x^0 - x^*\|^2 + k L^2/\sigma^2]^{-1}. \blacksquare$$

Из теоремы 1 следует, что любой метод вида (1) при сделанных выше предположениях не может сходиться быстрее  $O(1/k)$ . Приведем некоторые примеры использования этого результата. Вновь, как и в § 2, рассмотрим градиентный метод минимизации  $f(x)$ :

$$x^{k+1} = x^k - \gamma_k s^k, \quad s^k = \nabla f(x^k) + \xi^k \quad (5)$$

при абсолютных случайных помехах:

$$\mathbf{M} \xi^k = 0, \quad \mathbf{M} \|\xi^k\|^2 \geq \sigma^2$$

(обратите внимание, что здесь знак неравенства для дисперсии помех изменен на обратный по сравнению с § 2). Предположим, что  $f(x)$  имеет точку минимума  $x^*$ , а градиент  $\nabla f(x)$  удовлетворяет условию Липшица с константой  $L$ . Тогда мы находимся в условиях применимости теоремы 1, и из нее следует, что при любом выборе  $\gamma_k$  для метода (5) справедлива оценка

$$\mathbf{M} \|x^k - x^*\|^2 \geq (1/\mathbf{M} \|x^0 - x^*\|^2 + k L^2/\sigma^2)^{-1}. \quad (7)$$

Иначе говоря, никакой вариант градиентного метода при наличии абсолютных случайных помех не может сходиться быстрее  $O(1/k)$  (точнее,  $\mathbf{M} \|x^k - x^*\|^2 \geq \sigma^2/(L^2 k) + o(1/k)$ ). Заметим, что для градиентного метода с  $\gamma_k = \gamma/k$  было  $\mathbf{M} \|x^k - x^*\|^2 = O(1/k)$ , т. е. он асимптотически оптимален по порядку сходимости сходимости. Более точно вопрос об оптимальности градиентного метода будет исследован далее.

Рассмотрим теперь метод Ньютона при наличии помех. Будем считать, что матрица  $[\nabla^2 f(x^k)]^{-1}$  вычисляется точно, а градиент содержит аддитивную случайную помеху  $\xi^k$ . В этом случае метод Ньютона (модифицированный за счет введения параметра, задающего длину шага) принимает вид

$$x^{k+1} = x^k - \gamma_k [\nabla^2 f(x^k)]^{-1} (\nabla f(x^k) + \xi^k). \quad (8)$$

Относительно помех  $\xi^k$  будем считать, что они независимы и

$$\mathbf{M} \xi^k = 0, \quad \mathbf{M} \|\xi^k\|^2 \geq \sigma^2. \quad (9)$$

Можно показать, что в условиях теоремы 1 § 5 гл. 1 о сходимости «невозмущенного» метода Ньютона детерминированная часть процесса (8) (т. е.  $R(x^k) = [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$ ) в окрестности решения удовлетворяет условию Липшица, а случайная часть имеет дисперсию, ограниченную снизу. Таким образом, метод (8) также не может сходиться быстрее, чем со скоростью  $O(1/k)$ . Иначе говоря, наличие случайных помех уничтожает преимущества быстро сходящихся методов минимизации.

Приведем результат, аналогичный теореме 1, но применительно к относительным помехам.

**Теорема 2.** Пусть выполнены предположения, сформулированные в начале параграфа, и для всех  $k$

$$\mathbf{M} \|\xi^k\|^p \geq \tau \|x^k - x^*\|^p.$$

Тогда для любого метода (1)

$$\mathbf{M} \|x^k - x^*\|^p \geq \mathbf{M} \|x^0 - x^*\|^p q^k, \quad q = \tau / (L^2 + \tau). \blacksquare \quad (11)$$

В качестве первого примера использования теоремы 2 рассматриваем градиентный метод при случайных относительных помехах. Пусть  $f(x)$  дифференцируем, существует точка минимума  $x^*$ ,  $\nabla f(x)$  удовлетворяет условию Липшица с константой  $L$ , а помеха в определении градиента независима при различных  $k$

и удовлетворяет условиям  $\mathbf{M}\xi^k = 0$ ,  $\mathbf{M}\|\xi^k\|^2 \geq \tau\|x^k - x^*\|^2$ . Тогда в методе (5) при любых  $\gamma_k$  выполняется неравенство (11). Иными словами, градиентный метод при случайных относительных помехах не может сходиться быстрее, чем со скоростью геометрической прогрессии.

Вторым примером может служить метод *случайного поиска*. Пусть  $f(x) = (Ax, x)/2 - (b, x)$ ,  $A \leq I \leq L$ ,  $l > 0$ .

Рассмотрим метод

$$x^{k+1} = x^k - (\gamma_k/(2a))(f(x^k + \alpha h^k) - f(x^k - \alpha h^k))h^k, \quad (12)$$

где  $h^k$  — случайный равномерно распределенный на единичной сфере вектор,  $\alpha > 0$  — фиксированная длина пробного шага.

Метод может быть записан в виде (см. § 4 гл. 3)

$$x^{k+1} = x^k - \gamma_k h^k (h^k)^T \nabla f(x^k) = x^k - \gamma_k s^k,$$

$$s^k = h^k (h^k)^T \nabla f(x^k).$$

Используя результат упражнения 1, получаем

$$R(x^k) = \mathbf{M}s^k = \frac{1}{n} \nabla f(x^k),$$

$$\mathbf{M}\|\xi^k\|^2 = \mathbf{M}\|s^k - R(x^k)\|^2 = \frac{n-1}{n^2} \|\nabla f(x^k)\|^2 \geq \frac{n-1}{n^2} l^2 \|x^k - x^*\|^2.$$

Из теоремы 2 следует, что при любом способе выбора  $\gamma_k$  метод случайного поиска не может сходиться быстрее, чем геометрическая прогрессия со знаменателем

$$q = (n-1)l^2/(L^2 + (n-1)l^2). \quad (14)$$

В частности, для  $f(x) = \|x\|^2/2$ ,  $x \in \mathbf{R}^n$ , метод случайного поиска сходится не быстрее прогрессии со знаменателем  $(n-1)/n$ .

Теорему 2 можно несколько уточнить для случая, когда  $R(x)$  линейна, а для помех известна оценка снизу не только для дисперсии, но и для матрицы ковариаций. Итак, рассматривается метод

$$x^{k+1} = x^k - \Gamma_k(A(x^k - x^*) + \xi^k), \quad (15)$$

где  $\xi^k$  независимы,  $x^0$  случайный вектор,  $A^{-1}$  существует и

$$\mathbf{M}\xi^k = 0, \quad \mathbf{M}\xi^k(\xi^k)^T \geq B > 0, \quad \mathbf{M}(x^0 - x^*)(x^0 - x^*)^T > 0, \quad (16)$$

а  $\Gamma_k$  — детерминированные матрицы  $n \times n$ .

**Теорема 3.** В методе (15) при любых  $\Gamma_k$  справедлива оценка

$$\mathbf{M}(x^k - x^*)(x^k - x^*)^T \geq [(\mathbf{M}(x^0 - x^*)(x^0 - x^*)^T)^{-1} + kA^T B^{-1} A]^{-1} = \frac{1}{k} A^{-1} B(A^T)^{-1} + o\left(\frac{1}{k}\right). \quad \blacktriangle \quad (17)$$

В качестве приложения рассмотрим *обобщение градиентного метода минимизации квадратичной функции*

$$f(x) = (Ax, x)/2 - (b, x), \quad A \geq I > 0$$

при наличии помех:

$$x^{k+1} = x^k - \Gamma_k(\nabla f(x^k) + \xi^k), \quad \mathbf{M}\xi^k = 0, \quad \mathbf{M}\xi^k(\xi^k)^T = \sigma^2 I. \quad (18)$$

$$\mathbf{M}(x^k - x^*)(x^k - x^*)^T \geq \left(U_0^{-1} + \frac{k}{\sigma^2} A^2\right)^{-1} = \frac{\sigma^2}{k} A^{-2} + o\left(\frac{1}{k}\right), \quad (19)$$

$$U_0 = \mathbf{M}(x^0 - x^*)(x^0 - x^*)^T,$$

$$\|\mathbf{M}(x^k - x^*)(x^k - x^*)^T\| \geq \frac{\sigma^2}{k^2} + o\left(\frac{1}{k}\right), \quad (20)$$

$$\Gamma_k = (kA + \sigma^2 A^{-1} U_0^{-1})^{-1} = k^{-1} A^{-1} + o(1/k). \quad (21)$$

Сопоставляя (20) с оценкой (13) § 2 для градиентного метода, получаем, что при данных условиях выбор  $\gamma_k = 1/(kl)$  в градиентном методе является асимптотически оптимальным.

#### Упражнения.

1. Пусть  $h$  — случайный вектор, равномерно распределенный на единичной сфере в  $\mathbf{R}^n$ . Докажите, что  $\mathbf{M}hh^T = n^{-1}I_n$ , и если  $a$  — произвольный вектор, то  $\mathbf{M}\|ah^T a - n^{-1}a\|^2 = (n^{-1} - n^{-2})\|a\|^2$ .

2. Покажите, что если  $\mathbf{M}\xi^k(\xi^k)^T = B$ , то неравенство в (17) переходит в равенство при  $\Gamma_k$ , задаваемом (21).

**2. Оптимальные алгоритмы.** До сих пор мы ограничивались весьма узким классом алгоритмов — линейными рекуррентными. Однако вопрос об оптимальности можно решать для гораздо более общего класса процедур. Оказывается, что в ряде случаев можно установить поленицальные возможности любых (не обязательно рекуррентных или линейных) методов минимизации при наличии случайных помех. Основным инструментом здесь является известное в статистике неравенство Крамера — Рао (информационное неравенство).

Пусть функция  $f(x)$  квадратична:

$$f(x) = (Ax, x)/2 - (b, x), \quad A > 0, \quad (22)$$

а ее градиент вычисляется со случайной помехой  $\xi$ . Предположим, что помехи  $\xi$  независимы и одинаково распределены (раньше мы такого предположения не делали). Пусть уже вычислены значения  $r^1 = \nabla f(x^1) + \xi^1, \dots, r^k = \nabla f(x^k) + \xi^k$  в некоторых точках  $x^1, \dots, x^k$ . Наконец, пусть матрицы  $A$  и  $A^{-1}$  известны. Тогда  $x^i - x^* = A^{-1}r^i - A^{-1}\xi^i$ ,  $i = 1, \dots, k$ . Обозначим  $z^i = x^i - A^{-1}r^i$ ,  $\eta^i = -A^{-1}\xi^i$ . Тогда  $z^i = x^* + \eta^i$ . Величины  $z^i$  известны (так как

$x^i, r^i$  и  $A^{-1}$  известны), а величины  $\eta^i$  независимы и одинаково распределены (ибо такими являются  $\xi^i$ ). Таким образом, задача свелась к следующей. Заданы векторы  $z^i = x^* + \eta^i$ , где  $\eta^i$  — реализации независимой, однаково распределенной случайной величины. Требуется по ним оценить  $x^*$ .

Это — классическая задача оценки параметров, рассматриваемая в статистике. Для нее справедливо *неравенство Крамера — Рао*, утверждающее, что если  $\eta^i$  имеют плотность  $p_\eta(z)$ , эта плотность регулярна (т. е. справедливо равенство  $\int \nabla p_\eta(z) dz = 0$ ) и существует *фишеровская информационная матрица*

$$J = \int \frac{\nabla p_\eta(z) \nabla^T p_\eta(z)}{p_\eta(z)} dz, \quad 0 < J < \infty, \quad (23)$$

то для любой несмешанной оценки  $\hat{x}^k$  вектора  $x^*$  по измерениям  $z^i, i = 1, \dots, k$ , имеет место неравенство

$$\mathbf{M}(\hat{x}^k - x^*)(\hat{x}^k - x^*)^T \geq k^{-1} J^{-1}. \quad (24)$$

Иными словами, существует нижняя граница точности произвольных несмешанных оценок. Используя (24) и результат упражнения 4, приходим к следующему результату.

**Теорема 4.** *Пусть помехи  $\xi^i$  имеют плотность  $p(z)$ , причем  $p(z)$  регулярна и  $J = \int \frac{\nabla p \nabla^T p}{p} dz$  существует,  $0 < J < \infty$ . Тогда для любой несмешанной оценки  $\hat{x}^k$  точки минимума  $x^*$  функции (22), построенной по измерениям  $r^i = \nabla f(x^*) + \xi^i, i = 1, \dots, k$ , в  $k$  точках, справедливо неравенство*

$$\mathbf{M}(\hat{x}^k - x^*)(\hat{x}^k - x^*)^T \geq k^{-1} A^{-1} J A^{-1}. \quad (25)$$

Важно, что сюда не входят точки измерения  $x^1, \dots, x^k$ . Таким образом, при любом способе выбора  $k$  точек измерения градиента нельзя найти минимум с точностью, большей чем задаваемая неравенством (25).

Остается построить метод, для которого указанная нижняя граница достигается. Если ограничиться линейными алгоритмами

$$x^{k+1} = x^k - \Psi_k H (\nabla f(x^k) + \xi^k), \quad (26)$$

где  $H > 0$  — некоторая матрица, то получаем, что асимптотически оптимальный выбор  $\Psi_k$  и  $H$  таков:

$$\Psi_k = 1/k, \quad H = A^{-1}, \quad (27)$$

при этом

$$\mathbf{M}(x^k - x^*)(x^k - x^*)^T \leq k^{-1} A^{-1} B A^{-1} + o(k^{-1}), \quad B = M_{\xi\xi}^{-1}, \quad (28)$$

С учетом упражнения 3 отсюда получаем, что если  $\xi^i$  распределены нормально, то правая часть (25) совпадает с правой частью (28). Таким образом, для случая нормальных помех алгоритм (26), (27) является *асимптотически оптимальным* (не только среди линейных или рекуррентных алгоритмов). Для других распределений помехи алгоритм (26), (27), вообще говоря, не оптимален. Более того, можно показать, что правая часть (25) строго меньше правой части (28) для любого распределения, отличного от нормального. В этом случае оптимальный алгоритм можно получить, введя нелинейность в итерационный процесс

$$x^{k+1} = x^k - \Psi_k \Phi(\nabla f(x^k) + \xi^k), \quad (29)$$

где функция  $\Phi: \mathbf{R}^n \rightarrow \mathbf{R}^n$  и  $\Psi_k = 1/k$ .

$$\Phi(z) = J^{-1} A^{-1} \nabla \ln p(z), \quad \Psi_k = 1/k. \quad (30)$$

Для нормальных помех метод (29), (30) переходит в (26), (27).

Можно показать, что при определенных условиях на  $p(z)$  распределение величины  $\sqrt{k}(\hat{x}^k - x^*)$  для метода (29), (30) стремится к нормальному со средним 0 и матрицей ковариаций  $A^{-1} J A^{-1}$ . Сопоставляя это с правой частью (25), получаем, что метод (29), (30) является *асимптотически оптимальным*.

Практическая реализация метода (29), (30) затруднительна, так как в нем нужно знать матрицу  $A^{-1}$ , а также плотность распределения помех. Мы не будем останавливаться на способах преодоления этих трудностей. Здесь более важен принципиальный факт — возможность построения асимптотически оптимального алгоритма решения задачи минимизации при наличии случайных помех, причем этот алгоритм оказывается рекуррентным.

Подчеркнем еще, что все выводы здесь носят асимптотический характер. Оптимальный алгоритм для конечных  $k$  в случае нормальных помех дается выражением (21). Видно, что на начальных шагах ( $k \ll \sigma^2 A^{-2} U_0^{-1}$ )  $\Gamma_k$  примерно постоянно:  $\Gamma_k \approx \sigma^{-2} U_0 A$ , а для больших  $k$   $\Gamma_k$  убывает как  $k^{-1}$ :  $\Gamma_k = k^{-1} A^{-1} + o(k^{-1})$ .

Отметим также, что оптимальные алгоритмы предполагают точное знание закона распределения помех и неустойчивы к отклонению истинного распределения от предполагаемого. Существуют способы преодоления этой трудности (так называемые *робастные алгоритмы минимизации*).

3. Пусть случайный вектор  $\eta$  распределен нормально со средним 0 и ковариационной матрицей  $S$ . Покажите, что в этом случае информационная матрица (23) определяется формулой  $J = S^{-1}$ .

4. Пусть случайные векторы  $\xi$  и  $\eta$  связаны зависимостью  $\eta = B\xi$ , где  $B$  — некоторая матрица. Докажите, что для соответствующих информационных матриц справедливо соотношение  $J_\eta = B J_\xi B^T$ .

выпуклая комбинация его точек, т. е.

$$x^i \in Q, \sum_{i=1}^m \lambda_i = 1, \lambda_i \geq 0, i = 1, \dots, m \Rightarrow \sum_{i=1}^m \lambda_i x^i \in Q. \quad (2)$$

## Г л а в а 5 МИНИМАЗАЦИЯ НЕДИФФЕРЕНЦИРУЕМЫХ ФУНКЦИЙ

Во многих случаях функции, которые нужно минимизировать, оказываются недифференцируемыми. Примеры такого типа, связанные с декомпозицией, штрафными функциями, используянием теорем двойственности и т. д., будут неоднократно встречаться в дальнейшем. Негладкие функции возникают также в задачах наилучшего приближения, в статистике при оценке параметров по методу наименьших модулей, в задаче Штейнера и родственных ей. В приложениях типичны ситуации, когда оптимизируемый показатель — технический или экономический — недифференцируемым образом зависит от параметров (например, эта зависимость часто бывает кусочно-линейной). По указанным причинам при анализе задач оптимизации нельзя ограничиваться случаем гладких функций.

Разумеется, в общем виде задача минимизации недифференцируемых функций чрезвычайно сложна. Эти функции могут быть столь «плохо устроены», что значения функции в любом конечном множестве точек не несут информации об ее поведении в других точках. Ясно, что такие функции минимизировать безнадежно. Поэтому мы в основном ограничимся важным частным случаем негладких функций — выпуклыми функциями.

### § 1. Сведения из выпуклого анализа

Сравнительно недавно (в основном в 60-е годы нашего века) был создан простой и удобный аппарат для работы с выпуклыми функциями и множествами, получивший название *выпуклого анализа*. Мы неоднократно будем обращаться к технике выпуклого анализа. Пока потребуются лишь простейшие результаты из этой области.

**1. Выпуклые множества и проектирование.** Напомним, что множество  $Q$  в  $\mathbf{R}^n$  называется *выпуклым*, если оно содержит всякий отрезок, концы которого принадлежат  $Q$ , т. е. если для любых

$$x, y \in Q, 0 \leq \lambda \leq 1$$

$$\lambda x + (1 - \lambda)y \in Q.$$

Отсюда следует (по индукции), что в  $Q$  содержится и любая

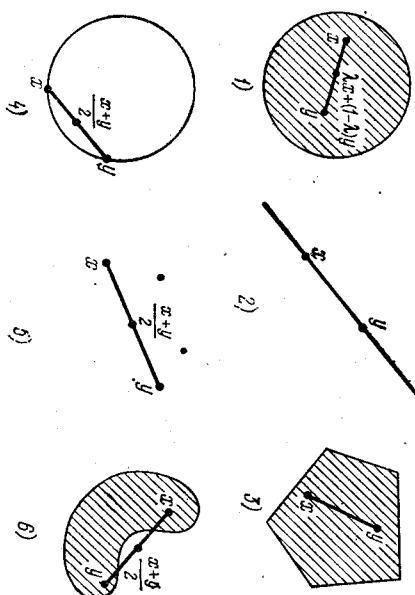


Рис. 10. Примеры выпуклых (1—3) и невыпуклых (4—6) множеств.

Для выпуклой функции  $f(x)$  множество  $Q = \{x: f(x) \leq \alpha\}$ , очевидно, выпукло при любом  $\alpha$ . Обратное, вообще говоря, неверно — функция  $f(x) = \sqrt{\|x\|}$  не выпукла, но множество  $\{x: f(x) \leq \alpha\}$  выпуклы (такие функции называются *квазивыпуклыми*).

Если множество  $Q$  невыпукло, то его можно «овыпуклить». В *выпуклой оболочке*  $\text{Conv } Q$  множества  $Q$  называется наименьшее выпуклое множество, содержащее  $Q$ , т. е. пересечение всех выпуклых множеств, содержащих  $Q$ . Такое множество существует и непусто для всех непустых  $Q$ . Например, выпуклой оболочкой сферы является шар, выпуклой оболочкой двух точек — соединяющий их отрезок. Нетрудно проверить, что выпуклую оболочку можно определить и иначе — как выпуклую комбинацию конечного числа точек из  $Q$ , т. е.

$$\text{Conv } Q =$$

$$= \left\{ x = \sum_{i=1}^m \lambda_i x^i, x^i \in Q, \lambda_i \geq 0, i = 1, \dots, m, \sum_{i=1}^m \lambda_i = 1 \right\}. \quad (3)$$

Непосредственно из определения видно, что такие множества, как шар, параллелепипед, линейное многообразие, многогранилое выпуклы, а сфера или конечный набор точек невыпуклы (рис. 10).

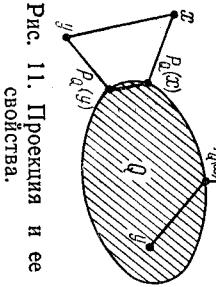
Лемма 1 (Каратеодори). Для  $Q \subset \mathbf{R}^n$  в (3) можно взять  $m \leq n+1$ . ▲

Для замкнутого  $Q$  множество  $\text{Conv } Q$  не обязательно замкнуто (например, для  $Q = \{x \in \mathbf{R}^2 : x_2 = x_1^{1/2}, x_1 > 0\}$  будет  $\text{Conv } Q = \{x \in \mathbf{R}^2 : 0 < x_2 \leq x_1^{1/2}, x_1 > 0\} \cup \{0, 0\}$ ).

**Лемма 2.** Если  $Q$  замкнуто и ограничено, то таковыми же будет и  $\text{Conv } Q$ .  $\blacktriangleleft$

В дальнейшем нам часто будет нужна операция проектирования. Проекцией точки  $x \in \mathbf{R}^n$  на множество  $Q \subset \mathbf{R}^n$  называется точка из  $Q$  (обозначаемая  $P_Q(x)$ ), находящаяся на кратчайшем расстоянии от  $x$ , т. е.

$$P_Q(x) = \underset{y \in Q}{\operatorname{argmin}} \|x - y\|. \quad (4)$$



Ясно, что если  $x \in Q$ , то  $P_Q(x) = x$ . Используя теорему Вейерштрасса (§ 3 гл. 1) получаем, что для замкнутого  $Q$  проекция существует. Если  $Q$  выпукло, то проекция единственна, так как  $P_Q(x) = \underset{y \in Q}{\operatorname{argmin}} \varphi(y)$ ,  $\varphi(y) = \|x - y\|^2$  — строго выпуклая функция (теорема 3 § 3 гл. 1).

Наконец, для выпуклого замкнутого  $Q$  проекция обладает следующими свойствами (рис. 11):

$$(x - P_Q(x), y - P_Q(x)) \leq 0 \quad \text{для всех } y \in Q, \quad (5)$$

$$\|P_Q(x) - P_Q(y)\| \leq \|x - y\| \quad \text{для любых } x, y. \quad (6)$$

#### Упражнение.

1. Докажите, что если  $Q$  выпукло, то множества  $\alpha Q = \{x = \alpha y, y \in Q\}$ ,  $AQ = \{x = Ay, y \in Q\}$  выпуклы (здесь  $\alpha \in \mathbf{R}^1$ ,  $A$  — матрица  $m \times n$ ), а если  $Q_1$  и  $Q_2$  выпуклы, то и  $Q_1 \cap Q_2$  и  $Q_1 + Q_2 = \{x = x_1 + x_2, x_1 \in Q_1, x_2 \in Q_2\}$  выпуклы.

2. Докажите, что функция  $\rho_Q(x) = \|x - P_Q(x)\|$  непрерывна для замкнутого  $Q$  и выпукла для выпуклого  $Q$ , а функция  $\varphi(x) = \rho_Q^2(x)/2$  выпукла и дифференцируема для выпуклого замкнутого  $Q$ , причем  $\nabla \varphi(x) = (x - P_Q(x))P_Q(x)$ .

3. Пусть  $x$  — внутренняя точка выпуклого множества  $Q$ ,  $y$  — его границная точка. Докажите, что точка  $(1 - \lambda)x + \lambda y$  является внутренней точкой  $Q$  при  $0 \leq \lambda < 1$  и не принадлежит  $Q$  при  $\lambda \geq 1$ .

**2. Теоремы отделимости.** В основе выпуклого анализа лежат так называемые теоремы отделимости (теоремы Хана — Бана-ха). Два множества  $Q_1$  и  $Q_2$  в  $\mathbf{R}^n$  называются *отделимыми*, если найдется разделяющая их гиперплоскость (рис. 12), иначе говоря, если найдется число  $\alpha$  и вектор  $a \in \mathbf{R}^n$ ,  $a \neq 0$ , такие, что  $(a, x) \geq \alpha$  для всех  $x \in Q_1$  и  $(a, x) \leq \alpha$  для всех  $x \in Q_2$ . Они называются *строго отделимыми*, если найдутся  $a \in \mathbf{R}^n$  и  $\alpha_1 > \alpha_2$  такие, что  $(a, x) \geq \alpha_1$  для  $x \in Q_1$  и  $(a, x) \leq \alpha_2$  для  $x \in Q_2$ .

**Теорема 1** (теорема отделимости). Пусть  $Q_1, Q_2$  — выпуклые замкнутые непресекающиеся множества в  $\mathbf{R}^n$ , причем  $Q_2$  ограничено. Тогда  $Q_1$  и  $Q_2$  строго отделимы.

**Доказательство.** Функция  $\rho_1(x) = \rho_{Q_1}(x) = \|x - P_{Q_1}(x)\|$  в соответствии с упражнением 2 непрерывна. Следовательно, на ограниченном замкнутом множестве  $Q_2$  она достигает минимума. Пусть  $a_1 = P_{Q_1}(a_2)$ ,  $a_2 = \underset{x \in Q_1}{\operatorname{argmin}} \rho_1(x)$ . Тогда  $a_1 \neq a_2$  (так как  $Q_1$  и  $Q_2$  не пересекаются),  $\|a_1 - a_2\| = \rho(Q_1, Q_2) = \min \{\|x - y\|, x \in Q_1, y \in Q_2\}$  и  $a_2 = P_{Q_1}(a_1)$ . Из (5) следует, что  $(a_1 - a_2, x) \geq (a_1 - a_2, a_1) = \alpha_1$  для  $x \in Q_1$ ,  $(a_1 - a_2, x) \leq (a_1 - a_2, a_2) = \alpha_2$  для  $x \in Q_2$ ,  $\alpha_1 - \alpha_2 = \|a_1 - a_2\|^2 > 0$ . Таким образом, множества  $Q_1$  и  $Q_2$  строго отделимы.  $\blacktriangleleft$

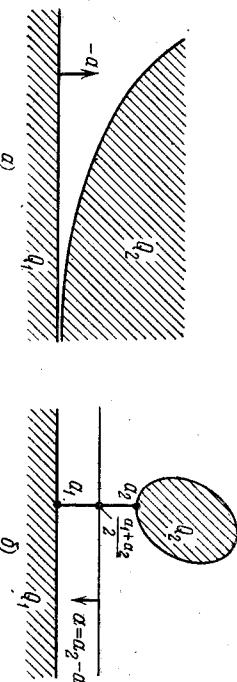


Рис. 12. Теорема отделимости: а) множества отделимы, б) множества строго отделимы.

Геометрически это доказательство совершенно наглядно (рис. 12, б). Требование ограниченности  $Q_2$  в теореме отделимости нельзя отбросить: множества  $Q_1 = \{x \in \mathbf{R}^2, x_2 \leq 0\}$ ,  $Q_2 = \{x \in \mathbf{R}^2, x_2 \geq x_1^{-1}, x_1 > 0\}$  не являются строго отделимыми.

Теорема отделимости позволяет доказать приводимый ниже результат об опорной гиперплоскости. Гиперплоскостью  $L = \{x : (a, x) = \alpha\}$  называется *опорной* к множеству  $Q$  в точке  $x^0$ , если  $x^0 \in L$ , а все множество  $Q$  лежит в полупространстве, задаваемом  $L$ , т. е.  $(a, x) \leq \alpha$  для  $x \in Q$  (рис. 13).

**Теорема 2** (об опорной гиперплоскости). Пусть  $Q$  — выпуклая границчная точка  $Q$ . Тогда существует гиперплоскость, опорная к  $Q$  в  $x^0$ .  $\blacktriangleleft$

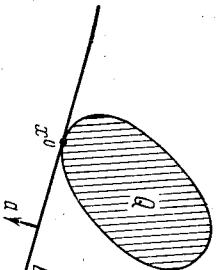


Рис. 13. Опорная гиперплоскость.

#### Упражнения.

4. Докажите следующие варианты теоремы отделимости:

- а) Пусть  $Q_1, Q_2$  — выпуклые множества, причем  $Q_1$  и  $Q_2$  имеют внутренние точки, ни одна из которых не является общей для обоих множеств. Тогда  $Q_1$  и  $Q_2$  отделимы.

б) Пусть  $Q_1, Q_2$  — выпуклые непересекающиеся множества. Тогда они отделены.

5. Убедитесь, что множества  $Q_1 = \{x \in \mathbf{R}^2 : |x_1| \leq 1, x_2 = 0\}$  и  $Q_2 = \{x \in \mathbf{R}^2 : x_1 = 0, |x_2| \leq 1\}$  выпуклы, не имеют общих внутренних точек, но неотделимы (ср. с упр. 4а)).

6. Докажите, что если  $x$  — граничная точка  $\text{Соп} Q$ , то в лемме 1 можно заменить  $n+1$  на  $n$ .

**3. Выпуклые недифференцируемые функции.** Определение выпуклой функции, данное в § 1 гл. 1, остается в силе и для недифференцируемых функций. Именно, скалярную функцию  $f(x)$ , заданную на всем пространстве  $\mathbf{R}^n$ , назовем выпуклой, если для любых  $x, y \in \mathbf{R}^n$  и любых  $0 \leq \lambda \leq 1$  выполняется неравенство

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad (7)$$

(см. рис. 1). Подчеркнем, что вследу в этой главе мы рассматриваем лишь функции, областью определения которых является все пространство. В дальнейшем (гл. 9) мы введем более общий класс выпуклых функций, заданных на некотором множестве. Для них многие утверждения данного параграфа (например, лемма 3) будут неверны.

Нетрудно проверить, что операции сложения, умножения на неотрицательное число и взятия максимума не выводят из класса выпуклых функций. Выпуклые функции обладают и рядом других удобных свойств. В частности, оказывается, что выпуклые функции на  $\mathbf{R}^n$  устроены достаточно просто.

**Лемма 3. Всякая выпуклая функция на  $\mathbf{R}^n$  непрерывна.**

Доказательство. Возьмем произвольные  $x \in \mathbf{R}^n$ ,  $\delta > 0$  и рассмотрим точки  $a^1 = x + \delta e_1$ ,  $a^2 = x - \delta e_1$ , ...,  $a^{2n-1} = x + \delta e_n$ ,  $a^{2n} = x - \delta e_n$ , где  $e_1, \dots, e_n$  — координатные орты. Пусть  $\Delta(\delta) = \max_{1 \leq i \leq 2n} |f(a^i) - f(x)|$ . Образуем многогранник  $Q(\delta)$  с вершинами в этих точках:

$$\begin{aligned} Q(\delta) &= \left\{ \sum_{i=1}^{2n} \mu_i a^i, \mu_i \geq 0, \sum_{i=1}^{2n} \mu_i = 1 \right\} = \\ &= \left\{ x + \delta \sum_{i=1}^n \mu_i e_i, |\mu_i| \leq 1 \right\}. \end{aligned}$$

Докажем, что  $\sup_{y \in Q(\delta)} |f(y) - f(x)| \leq \Delta(\delta)$ . Действительно, пусть

$y = \sum_{i=1}^{2n} \mu_i a^i$ ,  $\mu_i \geq 0$ ,  $\sum_{i=1}^{2n} \mu_i = 1$ , тогда по неравенству Иенсена (лемма 1 § 1 гл. 1)  $f(y) \leq \sum_{i=1}^{2n} \mu_i f(a^i) \leq \max_i f(a^i) \leq f(x) + \Delta(\delta)$ .

С другой стороны,  $f(y) \geq f(x) - f(y')$ , где  $y' \in Q(\delta)$  — точка,

симметричная с  $y$  относительно  $x$ , т. е. если  $y = x + \delta \sum_{i=1}^n \nu_i e_i$ , то  $y' = x - \delta \sum_{i=1}^n \nu_i e_i$ . Поэтому  $f(y) \geq 2f(x) - f(y') \geq f(x) - \Delta(\delta)$ , так как по доказанному  $f(y') \leq f(x) + \Delta(\delta)$ . Итак, действительно  $|f(y) - f(x)| \leq \Delta(\delta)$  для всех  $y \in Q(\delta)$ .

Заметим теперь, что верен следующий факт: всякая одномерная выпуклая функция  $\varphi(\tau)$  непрерывна. Действительно, при  $\varepsilon > 0$   $\varphi(\tau + \varepsilon) = \varphi((1 - \varepsilon)\tau + \varepsilon(\tau + 1)) \leq (1 - \varepsilon)\varphi(\tau) + \varepsilon\varphi(\tau + 1) = \varphi(\tau) + \varepsilon(\varphi(\tau + 1) - \varphi(\tau))$ ; с другой стороны,  $\varphi(\tau) = \varphi\left(\frac{\varepsilon}{1+\varepsilon}(\tau - 1) + \frac{1}{1+\varepsilon}(\tau + \varepsilon)\right) \leq \frac{\varepsilon}{1+\varepsilon}\varphi(\tau - 1) + \frac{1}{1+\varepsilon} \times \times \varphi(\tau + \varepsilon)$ , т. е.  $\varphi(\tau + \varepsilon) \geq \varphi(\tau) + \varepsilon(\varphi(\tau) - \varphi(\tau - 1))$ . Отсюда  $\varphi(\tau + \varepsilon) \rightarrow \varphi(\tau)$  при  $\varepsilon \rightarrow +0$ . Совершенно аналогично рассматривается случай  $\varepsilon < 0$ . Отметим попутно, что отсюда же следует существование левой и правой производных  $\varphi'_-(\tau) = \varphi'(\tau; -1)$  и  $\varphi'_+(\tau) = \varphi'(\tau; +1)$ , причем

$$\varphi(\tau) - \varphi(\tau - 1) \leq \varphi'(\tau; 1) \leq \varphi(\tau + 1) - \varphi(\tau). \quad (8)$$

В силу непрерывности  $\Delta_i(\delta) = |f(a^i) - f(x)| = |f(x \pm \delta e_i) - f(x)|$  стремится к 0 при  $\delta \rightarrow 0$ . Поэтому и  $\Delta(\delta) = \max_i \Delta_i(\delta)$  стремится к 0 при  $\delta \rightarrow 0$ . Итак,  $\sup_{y \in Q(\delta)} |f(y) - f(x)| \rightarrow 0$  при  $\delta \rightarrow 0$ , откуда и следует искомое утверждение.  $\blacktriangle$

**Следствие. Если  $f(x)$  — выпуклая функция, то множество  $Q(\alpha) = \{x : f(x) \leq \alpha\}$  выпукло и замкнуто. В частности, множество  $X^* = \underset{x \in \mathbf{R}^n}{\text{Argmin}} f(x)$  выпукло и замкнуто.  $\blacktriangle$**

**Лемма 4. Выпуклая функция  $f(x)$  имеет в произвольной точке  $x$  одностороннюю производную по любому направлению, равномерно ограниченную по направлениям:**

$$\begin{aligned} f'(x; y) &= \lim_{\alpha \rightarrow +0} \frac{f(x + \alpha y) - f(x)}{\alpha} \leq \\ &\leq f(x + y) - f(x) \leq \max_{\|z\|=\|y\|} (f(x + z) - f(x)). \quad (9) \end{aligned}$$

**4. Субградиент.** Разумеется, выпуклая функция не обязательно является дифференцируемой (рис. 14). Однако для нее можно ввести понятие, во многих отношениях аналогичное градиенту. Пусть  $f(x)$  — выпуклая функция на  $\mathbf{R}^n$ . Вектор  $a \in \mathbf{R}^n$ , для которого выполняется неравенство

$$f(x + y) \geq f(x) + (a, y)$$

для всех  $y \in \mathbf{R}^n$ , называется *субградиентом* функции  $f(x)$  в точке  $x$  и обозначается  $\partial f(x)$ . Как видно из простейших примеров (рис. 14), субградиент определяется, вообще говоря, неоднозначно. Мы будем применять обозначение  $\partial f(x)$  как для всего множества субградиентов, так и для его произвольного представителя (обычно по смыслу ясно, что именно имеется в виду).

Перейдем к исследованию свойств субградиента.

**Лемма 5.** Если  $f(x)$  дифференцируема в точке  $x$ , то субградиент определен однозначно и совпадает с градиентом:  $\partial f(x) = \nabla f(x)$ .

Так как градиент удовлетворяет неравенству (26) § 1 гл. 1:  $f(x+y) \geq f(x) + \nabla f(x), y$ , то он является субградиентом. Вычитая из неравенства (10) равенство  $f(x+y) = f(x) + (\nabla f(x), y) + o(y)$ , получаем  $(\partial f(x) - \nabla f(x), y) \geq o(y)$ , что возможно при всех  $y$  лишь если  $\partial f(x) = \nabla f(x)$ .  $\blacktriangleleft$

Можно показать, что всякая выпуклая функция почти всюду (т. е. за исключением множества меры нуль) дифференцируема. Это — известная теорема Радемахера.

**Лемма 6. Множество субградиентов в любой точке не-**

*пусто, выпукло, замкнуто и ограничено.*

Приведем схему доказательства. Рассмотрим множество  $Q = \{x, \alpha: \alpha \geq f(x)\}$  в пространстве  $\mathbf{R}^{n+1}$  (это множество называется *надграфиком* или *энграфом* функции  $f(x)$ ) (рис. 15). Множество  $Q$ , очевидно, выпукло, а из леммы 3 следует, что оно имеет внутренние точки. Точка  $\{x, f(x)\}$  является граничной точкой  $Q$ . По теореме 2 существует опорная гиперплоскость к  $Q$  в этой точке, задаваемая вектором  $\{a, -1\}$ . Тогда  $a$  является субградиентом  $f(x)$  в точке  $x$ . Выпуклость и замкнутость множества субградиентов следуют непосредственно из определения, ограниченность следует из леммы 4.  $\blacktriangleleft$

Для малых функций справедливо неравенство, аналогичное (29) § 1 гл. 1: для любых  $x, y$

$$(\partial f(x) - \partial f(y), x - y) \geq 0, \quad (11)$$

т. е. субградиент является *монотонным* оператором.

Знание субмножества позволяет вычислять производную по направлению (9) с помощью формулы, являющейся обобщением (6) § 1 гл. 1.

**Лемма 7. Для любых  $x, y$**

$$f'(x; y) = \max_{a \in \partial f(x)} (a, y). \quad (12)$$

Наметим доказательство (12). Поскольку  $f(x + \varepsilon y) - f(x) \geq \varepsilon(a, y)$  для всех  $a \in \partial f(x)$ , то  $f'(x; y) \geq \max_{a \in \partial f(x)} (a, y)$ . Пусть найдется  $y^0$  такое, что  $f'(x; y^0) > \max_{a \in \partial f(x)} (a, y^0)$ . Рассмотрим

в  $\mathbf{R}^{n+1}$  луч  $L = \{a, z: a = f(x) + \lambda f'(x; y^0), z = x + \lambda y^0, \lambda > 0\}$  и надграфик  $A = \{a, z: a > f(z)\}$ . Поскольку  $f(z) \geq f(x) + \lambda f'(x; y^0)$ , то множества  $A$  и  $L$  не пересекаются. Применив теорему отделимости, получим противоречие.  $\blacktriangleleft$

Отсюда и из (6) § 1 гл. 1 следует результат, обратный лемме 5, — если  $\partial f(x)$  состоит из одного элемента, то  $f(x)$  дифференцируема в этой точке.

Леммы 3, 4 и 7 приводят к следующему результату.

**Лемма 8. Субградиенты выпуклой функции  $f(x)$  ограничены на всяком ограниченном множестве или множестве вида  $\{x: f(x) \leq \alpha\}$ .  $\blacktriangleleft$**

В дальнейшем нам придется иногда оперировать с суммами множеств (например, с выражением  $\alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$  в предположении ниже леммы 10). Напомним, что если  $A, B, C$  — множества в  $\mathbf{R}^n$ ,  $\beta, \gamma \in \mathbf{R}^1$ , то  $A = \beta B + \gamma C$  означает, что  $A = \{a = \beta b + \gamma c, b \in B, c \in C\}$ . Мы уже знаем (упр. 1), что сумма выпуклых множеств выпукла;  $B + C = \emptyset$ , если  $B = \emptyset$ . **Лемма 9. Если  $B$  и  $C$  замкнуты и ограничены, то  $B + C$  замкнуто и ограничено.**  $\blacktriangleleft$

Предположение об ограниченности здесь существенно: например, если  $B = \{x \in \mathbf{R}^2: x_2 \geq x_1^{-1}, x_1 > 0\}$ ,  $C = \{x \in \mathbf{R}^2: x_1 = 0\}$ , то  $B$  и  $C$  замкнуты, но  $B + C = \{x \in \mathbf{R}^2: x_1 > 0\}$  незамкнуто.

Приведем три леммы, дающие возможность вычислять субградиенты сложных функций.

**Лемма 10. Если  $f_1(x), f_2(x)$  выпуклы,  $f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x)$  и  $\alpha_1, \alpha_2 \geq 0$ , то**

$$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x). \quad (13)$$

Доказательство. Операция дифференцирования по направлению очевидным образом линейна:  $f'(x; y) = \alpha_1 f'_1(x; y) + \alpha_2 f'_2(x; y)$  для всех  $x, y$ . Воспользуемся формулой (12):  $\max_{a \in \partial f(x)} (a, y) = \max_{a \in \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)} (a, y)$ .

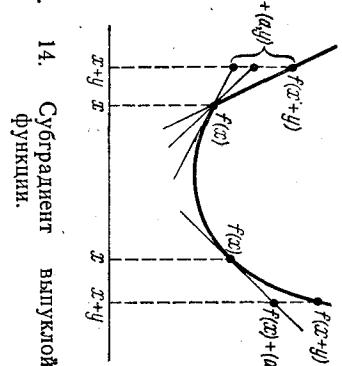


Рис. 14. Субградиент выпуклой функции.

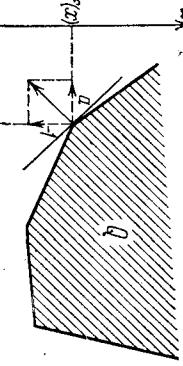


Рис. 15. К доказательству существования субградиента.

Множества  $\partial f(x)$  и  $\alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$  в силу лемм 6, 9 и упражнения 1 выпуклы, замкнуты и ограничены. Но если для всех  $y \in \mathbf{R}^n$   $\max_{a \in A} (a, y) = \max_{b \in B} (b, y)$ , для выпуклых, замкнутых ограниченных множеств  $A$  и  $B$ , то  $A$  и  $B$  совпадают (это легко показать с помощью теоремы 1). Поэтому справедливо (13).  $\blacktriangleleft$

Разумеется, формула (13) распространяется на сумму нескольких выпуклых функций:

$$\partial \left( \sum_{i=1}^m \alpha_i f_i(x) \right) = \sum_{i=1}^m \alpha_i \partial f_i(x), \quad \alpha_i \geq 0. \quad (14)$$

Следующая лемма дает правило вычисления субградиента максимума нескольких функций.

**Лемма 11.** Пусть  $f(x) = \max_{1 \leq i \leq m} f_i(x)$ , где  $f_i(x)$  выпуклы.

Тогда

$$\partial f(x) = \text{Conv} \bigcup_{i \in I(x)} \partial f_i(x), \quad I(x) = \{i : f_i(x) = f(x)\}. \quad (15)$$

**Доказательство.** В силу лемм 6, 2 множество  $A = \text{Conv} \bigcup_{i \in I(x)} \partial f_i(x)$  выпукло, замкнуто и ограничено; таковым же является и множество  $\partial f(x)$ . Нетрудно видеть, что  $f'(x; y) = \max_{i \in I(x)} f'_i(x; y)$  для всех  $y$ . Но по лемме 7 и в силу определения  $\text{Conv} Q \max_{i \in I(x)} f'_i(x; y) = \max_{\substack{i \in I(x) \\ \lambda_i \geq 0}} \sum_{i \in I(x)} \lambda_i f'_i(x; y) = \max_{a \in A} (a, y)$ . С другой стороны,  $f'(x; y) = \max_{a \in \partial f(x)} (a, y)$  (лемма 7).

Если  $\max_{a \in A} (a, y) = \max_{a \in \partial f(x)} (a, y)$  для всех  $y$ , то (ср. с доказательством леммы 10)  $A = \partial f(x)$ .  $\blacktriangleleft$

**Лемма 12.** Пусть  $A$  — матрица  $m \times n$ ,  $\Phi(y)$  — выпуклая функция на  $\mathbf{R}^m$ ,  $f(x) = \Phi(Ax)$ ,  $x \in \mathbf{R}^n$ . Тогда

$$\partial f(x) = A^T \partial \Phi(Ax). \quad \blacktriangleleft$$

С помощью лемм 10—12 можно вычислять субградиенты различных функций столь же просто, как и градиенты гладких функций по обычным правилам дифференцирования.

**Упражнение.**

7. Вычислите субградиенты следующих функций: а)  $f(x) = \|x\|$ ; б)  $f(x) = \sum_{i=1}^k |(a^i, x) - b_i|$ ; в)  $f(x) = \max_{1 \leq i \leq k} ((a^i, x) - b_i)$ .

**Ответы.** а)  $\partial f(x) = \begin{cases} \frac{x}{\|x\|}, & x \neq 0, \\ 0, & \|x\| \leq 1, x = 0; \end{cases}$  б)  $\partial f(x) = \sum_{i=1}^k \text{sign}((a^i, x) - b_i)$ .

в)  $\partial f(x) = \sum_{i=1}^k a_i a^i$ ,  $a_i = 0$  при  $(a^i, x) - b_i < f(x)$ ,  $a_i \geq 0$ ,  $\sum_{i=1}^k a_i = 1$ .

**5.  $\varepsilon$ -субградиент.** Понятие субградиента может быть обобщено следующим образом. Вектор  $a \in \mathbf{R}^n$  называется  $\varepsilon$ -субградиентом выпуклой функции  $f(x)$  в точке  $x$ , если

$$f(x + y) \geq f(x) + (a, y) - \varepsilon \quad (17)$$

для всех  $y \in \mathbf{R}^n$ . Здесь  $\varepsilon \geq 0$  — некоторое фиксированное число.

Множество  $\varepsilon$ -субградиентов и произвольный  $\varepsilon$ -субградиент будем обозначать  $\partial_\varepsilon f(x)$ . Из определения следует, что  $\partial f(x) = \partial_0 f(x)$ ,  $\partial f(x) \subset \partial_\varepsilon f(x)$  для всех  $\varepsilon > 0$  и, более того,  $\partial f(x) = \bigcap_{\varepsilon > 0} \partial_\varepsilon f(x)$ . Геометрически  $\varepsilon$ -субградиент соответствует гиперплоскостям в  $\mathbf{R}^{n+1}$ , разделяющим надграфик  $f(x)$  и точку  $\{f(x) - \varepsilon, x\}$  (рис. 16). В отличие от субградиента,  $\varepsilon$ -субградиент

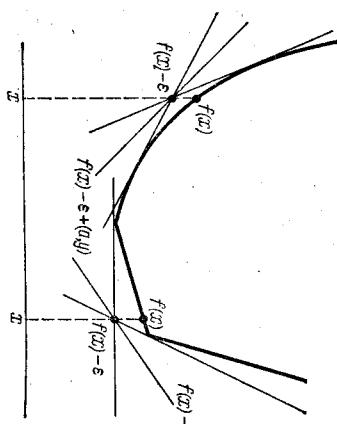


Рис. 16.  $\varepsilon$ -субградиент.

при  $\varepsilon > 0$  не определяется локальными свойствами  $f(x)$ . Ясно, что  $\varepsilon$ -субградиент не единственный даже для дифференцируемых функций; исключением является аффинная функция  $f(x) = (c, x) + \alpha$ , тогда  $\partial_\varepsilon f(x) = c$  для всех  $\varepsilon, x$ .

Правила вычисления  $\varepsilon$ -субградиентов не столь просты, как субградиентов. Мы приведем один важный частный случай, когда отыскание  $\varepsilon$ -субградиента требует меньших вычислений, чем субградиента. Пусть

$$f(x) = \max_{y \in Q} \Phi(x, y), \quad (18)$$

где  $x \in \mathbf{R}^n$ ,  $Q$  — компактное множество,  $\Phi(x, y)$  непрерывна по  $y$  и выпукла по  $x$ . В частности,  $Q$  может состоять из конечного числа элементов (тогда получаем функцию из леммы 11). Очевидно, что  $f(x)$  определена на  $\mathbf{R}^n$  и выпукла. Пусть  $\tilde{y} = \tilde{y}(x) —$  произвольная точка из  $Q$  такая, что  $\Phi(x, \tilde{y}) \geq f(x) - \varepsilon$ . Иначе говоря,  $\tilde{y}$  — произвольная точка, в которой приближенно (с точностью до  $\varepsilon$ ) достигается максимум  $\Phi(x, y)$  по  $y \in Q$ .

$$(19)$$

**Лемма 13.**

$$\partial_x \Phi(x, \tilde{y}) \subseteq \partial_\varepsilon f(x). \quad (20)$$

**Доказательство.** Для любого  $z$  из определения субградиента и (18), (19) следует

$$f(x+z) = \max_{y \in Q} \Phi(x+z, y) \geq \Phi(x+z, \tilde{y}) \geq \Phi(x, \tilde{y}) +$$

$$+ (\partial_x \Phi(x, \tilde{y}), z) \geq f(x) + (\partial_x \Phi(x, \tilde{y}), z) - \varepsilon. \quad \blacktriangle$$

Таким образом, чтобы найти один из  $\varepsilon$ -субградиентов  $f(x)$  вида (18), достаточно приближенно отыскать максимум по  $y$  и взять субградиент соответствующей функции  $\Phi$ . Вычисление же субградиента  $f(x)$  требует точной максимизации  $\Phi$  по  $y$ .

## § 2. Условия экстремума, существование, единственность и устойчивость решения

Для задачи

$$\min f(x), \quad x \in \mathbf{R}^n, \quad (1)$$

где  $f(x)$  — выпуклая недифференцируемая функция на  $\mathbf{R}^n$ , проведем анализ, аналогичный осуществленному в §§ 2 и 3 гл. 1 для гладких функций.

**1. Условия экстремума.** Необходимые и достаточные условия минимума формулируются в терминах субградиентов чрезвычайно просто.

**Теорема 1. Условие**

$$0 \in \partial f(x^*) \quad (2)$$

необходимо и достаточно для того, чтобы точка  $x^*$  являлась решением (1).

**Доказательство.** Необходимость. Пусть  $x^*$  — точка минимума  $f(x)$ , тогда  $f(x^* + y) \geq f(x^*) + (0, y)$  для всех  $y$ . Это означает (10) § 1, что 0 является субградиентом  $f(x)$  в точке  $x^*$ .

Достаточность. Если 0 является субградиентом в точке  $x^*$ , то  $f(x^* + y) \geq f(x^*) + (0, y) = f(x^*)$  при всех  $y$ , т. е.  $x^*$  — решение (1).  $\blacktriangle$

**2. Существование и единственность минимума.**

**Теорема 3.** Пусть функция  $f(x)$  выпукла на  $\mathbf{R}^n$ , а множество  $Q_\alpha = \{x: f(x) \leq \alpha\}$  непусто и ограничено для некоторого  $\alpha$ . Тогда  $f(x)$  достигает минимума на  $\mathbf{R}^n$ .

Действительно, по лемме 3 § 1  $f(x)$  непрерывна, в силу чего применима теорема Вейерштрасса (§ 3 гл. 1).  $\blacktriangle$

Разумеется, в точке минимума могут быть и ненулевые субградиенты (например, для  $f(x) = \|x\|$ :  $\partial f(0) = \{a: \|a\| \leq 1\}$ , см. упр. 7 § 1), и это условие (2) отличается от условия  $\nabla f(x) = 0$  для гладких функций. Иначе говоря, условия экстремума в не-гладком случае не сводятся к решению системы уравнений.

В данном случае еще более очевидным является утверждение, которое мы приводили в § 3 гл. 1, — условия экстремума вовсе не предназначены для конструктивного отыскания минимума,

С помощью понятия  $\varepsilon$ -субградиента можно сформулировать необходимые и достаточные условия того, что точка  $x_\varepsilon$  является приближенным решением задачи (1).

**Теорема 2. Условие**

$$0 \in \partial_\varepsilon f(x_\varepsilon) \quad (3)$$

выполняется тогда и только тогда, когда  $f(x_\varepsilon) \leq \inf_x f(x) + \varepsilon$ .  $\blacktriangle$

**Упражнения.**

1. Убедитесь в справедливости следующих условий экстремума:

a)  $f(x) = \sum_{i=1}^m \alpha_i \|x - a^i\|$ ,  $\alpha_i > 0$ ,  $x, a^i \in \mathbf{R}^n$ . Тогда  $\nabla f(x^*) = 0$ , если

$$x^* \neq a^i, \text{ и } \alpha_i \geq \left\| \sum_{j \neq i} \frac{\alpha_j(a^j - a^i)}{\|a^j - a^i\|} \right\|, \text{ если } x^* = a^i.$$

b)  $f(x) = \sum_{i=1}^m |(a^i, x) - b_i|$ . Существуют  $|\lambda_i^*| \leq 1$ ,  $i \in I^* = \{i: (a^i, x^*) = b_i\}$  такие, что  $\sum_{i \in I^*} \lambda_i^* a^i + \sum_{i \in I_+} a^i - \sum_{i \in I_-} a^i = 0$ ,  $I_+ = \{i: (a^i, x^*) > b_i\}$ ,

$I_- = \{i: (a^i, x^*) < b_i\}$ .  
в)  $f(x) = \max_{1 \leq i \leq m} f_i(x)$ , где  $f_i(x)$  — выпуклые дифференцируемые функции. Тогда существуют  $\lambda_i^* \geq 0$ ,  $i \in I^* = \{i: f_i(x^*) = f(x^*)\}$ ,  $\sum_{i \in I^*} \lambda_i^* = 1$  такие, что  $\sum_{i \in I^*} \lambda_i^* \nabla f_i(x^*) = 0$ .

2. Пусть  $f(x)$  та же, что и в упражнении 1в). Пусть для точки  $x_\varepsilon$  существуют  $\lambda_i \geq 0$ ,  $i \in I_\varepsilon = \{i: f_i(x_\varepsilon) \geq f(x_\varepsilon) - \varepsilon\}$ ,  $\sum_{i \in I_\varepsilon} \lambda_i = 1$  такие, что  $\sum_{i \in I_\varepsilon} \lambda_i \nabla f_i(x_\varepsilon) = 0$ . Тогда  $f(x_\varepsilon) \leq \inf_x f(x) + \varepsilon$ . Докажите, используя лемму 13 § 1 и теорему 2.

**3. Существование и единственность минимума.**

**Теорема 3.** Пусть функция  $f(x)$  выпукла на  $\mathbf{R}^n$ , а множество  $Q_\alpha = \{x: f(x) \leq \alpha\}$  непусто и ограничено для некоторого  $\alpha$ . Тогда  $f(x)$  достигает минимума на  $\mathbf{R}^n$ .  
Действительно, по лемме 3 § 1  $f(x)$  непрерывна, в силу чего применима теорема Вейерштрасса (§ 3 гл. 1).  $\blacktriangle$

Вопрос об единственности минимума просто решается для строго выпуклых функций. Напомним (§ 1 гл. 1), что функция называется строго выпуклой, если для любых  $x \neq y$ ,  $0 < \lambda < 1$  для гладких функций. Иначе говоря, условия экстремума в не-гладком случае не сводятся к решению системы уравнений.

В данном случае еще более очевидным является утверждение, которое мы приводили в § 3 гл. 1, — условия экстремума вовсе не предназначены для конструктивного отыскания минимума,

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y). \quad (4)$$

**Теорема 4. Точка минимума строго выпуклой функции единственна.** Доказательство очевидно.  $\blacktriangle$

**Упражнения.**

3. Докажите, что для строго выпуклой функции выполняется при всех  $y \neq 0$  неравенство

$$f(x+y) > f(x) + (\partial f(x), y). \quad (5)$$

4. Убедитесь, что функция  $f(x) = \|x\|$  не является строго выпуклой, а функция  $\sum_{i=1}^m a_i \|x - a^i\|$ ,  $a_i > 0$ , строго выпукла, если только точки  $a^i$  не лежат на одной прямой.

5. Докажите, что функция  $f(x) = \sum_{i=1}^m a_i \|x - a^i\|$ ,  $a_i > 0$ , достигает минимума на  $\mathbf{R}^n$ , причем единственного, если  $a^i$  не лежат на одной прямой.

**3. Устойчивость минимума.**

**Теорема 5. Единственная точка минимума выпуклой функции глобально устойчива, т. е. к ней сходится любая минимизирующая последовательность. Ограниченнное множество точек минимума  $X^*$  слабо устойчиво, т. е. всякая минимизирующая последовательность имеет предельные точки и все они принадлежат  $X^*$ . ▲**

Эти утверждения немедленно следуют из непрерывности  $f(x)$  (лемма 3 § 1) и следующего легко проверяемого факта.

**Лемма 1.** Если  $Q_\alpha = \{x: f(x) \leq \alpha\}$  ограничено и не пусто при некотором  $\alpha$  для выпуклой функции  $f(x)$ , то  $Q_\alpha$  ограничено при всех  $\alpha$ . ▲

Количественные оценки устойчивости легко получить для класса сильно выпуклых функций. Напомним определение *сильно выпуклости*, данное в § 1 гл. 1 и относящееся как к гладким, так и негладким функциям  $f(x)$ : найдется  $l > 0$  такое, что

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - l\lambda(1-\lambda)\|x-y\|^2/2 \quad (6)$$

для любых  $x, y$  и  $0 \leq \lambda \leq 1$ . Такие функции обладают следующими свойствами.

**Лемма 2.** Для сильно выпуклой функции  $f(x)$  справедливо неравенство

$$f(y) \geq f(x) + (\partial f(x), y-x) + l\|y-x\|^2/2 \quad (7)$$

для всех  $x, y$ , при этом  $f(x)$  достигает единственного минимума  $x^*$  и для всех  $x$

$$f(x) \geq f(x^*) + l\|x-x^*\|^2/2. \quad (8)$$

**Доказательство.** Из определения субградиента имеем

$$f(\lambda x + (1-\lambda)y) = f(x + (1-\lambda)(y-x)) \geq f(x) + (1-\lambda)(\partial f(x), y-x).$$

Подставляя это неравенство в (6) и сокращая на  $1-\lambda$ , получаем

$$f(y) \geq f(x) + (\partial f(x), y-x) + l\lambda\|x-y\|^2/2.$$

Это верно для всех  $\lambda < 1$ ; переходя к пределу при  $\lambda \rightarrow 1$ , получаем (7). Из (7) следует, что  $Q = \{y: f(y) \leq f(x)\}$  ограничено, и из теорем 3 и 4 вытекает существование и единственность  $x^*$ .

Используя теорему 1 и (7), приходим к (8). ▲

Неравенство (8) позволяет оценивать близость  $x$  к  $x^*$  по близости  $f(x)$  к  $f(x^*)$ . Частный случай оценки (8) для гладких функций мы уже приводили в § 3 гл. 1.

Однако нужно отметить, что для негладких задач свойство сильной выпуклости, вообще говоря, не характерно.

Есть другой важный класс функций, для которого можно гарантировать устойчивость, причем этот класс включает только негладкие функции. Будем говорить, что  $x^*$  — точка острого минимума  $f(x)$ , если для всех  $x$  (рис. 17)

$$f(x) \geq f(x^*) + \alpha\|x-x^*\|, \quad (9)$$

$\alpha > 0$ .

Рис. 17. Острый минимум.

Такое условие заведомо не может выполняться для гладких функций (упр. 8 § 3 гл. 1).

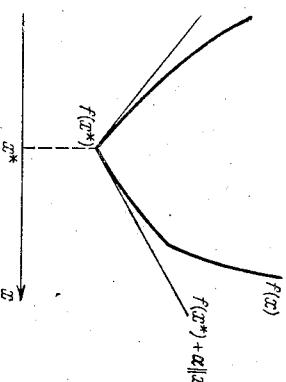
**Лемма 3. Следующие условия эквивалентны (9) для выпуклой функции  $f(x)$ :**

- а)  $f'(x^*, y) \geq \alpha \geq 0$  для всех  $y$ ;
- б)  $0$  является внутренней точкой  $\partial f(x^*)$ . ▲

С помощью (9) можно оценивать близость  $x$  к  $x^*$ , зная близость  $f(x)$  к  $f(x^*)$ . Однако более интересно свойство «сверхустойчивости» острого минимума, которого нет для задач с сильно выпуклыми функциями. Оказывается, точка острого минимума не изменяется при малом возмущении функции.

**Теорема 6.** Пусть  $f(x)$  — выпуклая функция на  $\mathbf{R}^n$ ,  $x^*$  — точка ее острого минимума,  $g(x)$  — выпуклая функция. Тогда найдется  $\varepsilon_0 > 0$  такое, что при  $0 \leq \varepsilon < \varepsilon_0$  точка минимума функции  $f(x) + \varepsilon g(x)$  единственна и совпадает с  $x^*$ .

Доказательство. По лемме 10 § 1 для  $\Phi_\varepsilon(x) = f(x) + \varepsilon g(x)$  имеем  $\partial \Phi_\varepsilon(x) = \partial f(x) + \varepsilon \partial g(x)$ . Поскольку  $0$  — внутренняя точка  $\partial f(x^*)$  (лемма 3), а  $\partial g(x^*)$  ограничено (лемма 6 § 1), то при достаточно малом  $\varepsilon$  будет  $\varepsilon \partial g(x^*) \subset -\partial f(x^*)$ , т. е.  $0 \in \partial \Phi_\varepsilon(x^*)$ . По теореме 1  $x^*$  — точка минимума  $\Phi_\varepsilon(x)$ . ▲



### Упражнения.

6. Докажите следующее обобщение теоремы 6. Пусть  $f(x) — выпуклая функция,  $X^* = \operatorname{Argmin}_{x \in \mathbb{R}^n} f(x) \neq \emptyset$  и выполняется условие  $f(x) \geq f^* + \varphi(x, X^*)$ ,  $\alpha > 0$ , где  $f^* = f(x^*)$ ,  $x^* \in X^*$ ,  $\varphi(x, X^*) = \|x - P_{X^*}(x)\|$ . Пусть  $g(x) — выпуклая функция, множество  $X_g^* = \operatorname{Argmin}_{x \in X^*} g(x) \neq \emptyset$  и ограничено. Тогда$$

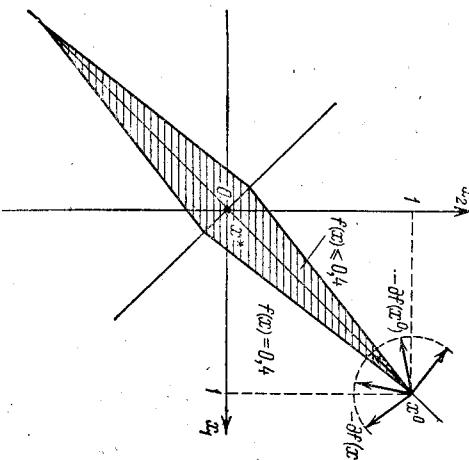
$$X_g^* = \operatorname{Argmin}_{x \in \mathbb{R}^n} [f(x) + g(x)] \text{ при достаточно малых } \varepsilon > 0.$$

7. Рассмотрите понятие обусловленности точки минимума (§ 3 гл. 1) применительно к негладким  $f(x)$ . Чему равно  $\mu$  для  $f(x) = \sum_{i=1}^n \lambda_i |x_i|$ ,  $\lambda_i > 0$ ?

$$\text{Минимально к негладким } f(x). \text{ Чему равно } \mu \text{ для } f(x) = \sum_{i=1}^n \lambda_i |x_i|, \lambda_i > 0?$$

### § 3. Субградиентный метод

1. Идея метода. Основные алгоритмы минимизации гладких функций — градиентный и Ньютона — были построены на использовании линейной или квадратичной аппроксимации функции, задаваемой первыми членами ряда Тейлора. Однако



для нелиффеирируемой функции эта идея неприменима — такая функция не может быть хорошо аппроксимирована ни линейной, ни квадратичной функциями. Другие методы минимизации гладких функций, описанные в гл. 3, также теряют работоспособность при переходе к нелиффеирируемым функциям. Приведем

несколько примеров.

Пусть  $f(x) = |x_1 - x_2| + 0,2|x_1 + x_2|$ . Тогда вектор  $\{1, 2; -0,8\}$  является ее субградиентом в точке  $\{1; 1\}$ , однако движение по нему приводит к возрастанию функции при любом выборе длины шага  $y_k$  (рис. 18). Таким образом, значение функции в методе (3) не могут убывать монотонно. Оказывается, однако, что при этом монотонно убывает другая функция — расстояние до точки минимума, и в этом и заключается основная идея субградиентного метода (3). Вторая особенность метода заключается в правиле выбора длины шага. Ясно, что в (3) нельзя брать  $y_k \equiv \gamma$ , как в градиентном методе. Например, для функции  $f(x) = \|x\|$  имеем  $\|\partial f(x)\| = 1$  для всех  $x \neq 0$ , и тем самым будет  $\|x_{k+1} - x_k\| \equiv \gamma$ , поэтому метод не будет сходиться. С другой стороны, нельзя выбирать  $y_k$ , как в методе скорейшего спуска — ведь  $f(x)$  не обязательно убывает по направлению  $-\partial f(x^k)$ . В субградиентном методе можно уменьшать длину шага, либо используя близость значения функции в текущей точке к минимальному, либо выбирая некоторую априори стремящуюся к 0 последовательность. Рассмотрим ниже оба эти способа.

2. Основные результаты. Пусть  $f(x) — выпуклая функция и точкой минимума (рис. 18). Поэтому метод покординатного спуска неприменим для минимизации подобных нелиффеирируемых функций.$

Можно было бы попытаться построить аналог метода скользящего спуска. Назовем вектор  $s = s(x) \in \mathbb{R}^n$ ,  $\|s\| = 1$ , направлением скорейшего спуска в точке  $x$ , если по этому направлению

$$s(x) = \operatorname{argmin}_{\|y\|=1} f'(x; y). \quad (1)$$

В силу формулы (12) § 1 для выпуклой функции направление наискорейшего спуска существует и определяется формулой  $s = -P_{\partial f(x)}(0) / \|P_{\partial f(x)}(0)\|$ . (2)

т. е.  $s$  задается субградиентом с минимальной нормой. Однако можно построить пример выпуклой функции, для которой метод скорейшего спуска

$$x^{k+1} = x^k + y_k s(x^k), \quad y_k = \operatorname{argmin}_{y>0} f(x^k + ys(x^k))$$

«застрекает», не дойдя до точки минимума.

Поэтому разработка методов минимизации негладких функций требует привлечения новых идей. Одна из них, принадлежащая Н. З. Шору, выглядит несколько неожиданно. Пишется прямой аналог градиентного метода с заменой градиента на произвольный субградиент функции:

$$x^{k+1} = x^k - y_k \partial f(x^k). \quad (3)$$

Рассмотрим снова функцию  $f(x) = |x_1 - x_2| + 0,2|x_1 + x_2|$ . Тогда вектор  $\{1, 2; -0,8\}$  является ее субградиентом в точке  $\{1; 1\}$ , однако движение по нему приводит к возрастанию функции при любом выборе длины шага  $y_k$  (рис. 18). Таким образом, значение функции в методе (3) не могут убывать монотонно. Оказывается, однако, что при этом монотонно убывает другая функция — расстояние до точки минимума, и в этом и заключается основная идея субградиентного метода (3). Вторая особенность метода заключается в правиле выбора длины шага. Ясно, что в (3) нельзя брать  $y_k \equiv \gamma$ , как в градиентном методе. Например, для функции  $f(x) = \|x\|$  имеем  $\|\partial f(x)\| = 1$  для всех  $x \neq 0$ , и тем самым будет  $\|x_{k+1} - x_k\| \equiv \gamma$ , поэтому метод не будет сходиться. С другой стороны, нельзя выбирать  $y_k$ , как в методе скорейшего спуска — ведь  $f(x)$  не обязательно убывает по направлению  $-\partial f(x^k)$ . В субградиентном методе можно уменьшать длину шага, либо используя близость значения функции в текущей точке к минимальному, либо выбирая некоторую априори стремящуюся к 0 последовательность. Рассмотрим ниже оба эти способа.

2. Основные результаты. Пусть  $f(x) — выпуклая функция и точкой минимума (рис. 18). Поэтому метод покординатного спуска неприменим для минимизации подобных нелиффеирируемых функций.$

Можно было бы попытаться построить аналог метода скользящего спуска. Назовем вектор  $s = s(x) \in \mathbb{R}^n$ ,  $\|s\| = 1$ , направлением скорейшего спуска в точке  $x$ , если по этому направлению

$$x^{k+1} = x^k - \frac{\partial f(x^k)}{\|\partial f(x^k)\|}, \quad y_k \rightarrow 0, \quad \sum_{k=0}^{\infty} y_k = \infty. \quad (4)$$

Иначе говоря, из точки  $x^k$  делается шаг фиксированной длины  $\gamma_k$  по направлению, противоположному субградиенту. Длина шага стремится к 0, а суммарная длина шагов бесконечна. При мерами последовательностей  $\gamma_k$ , удовлетворяющих условиям (4), могут служить

$$\gamma_k = \frac{\gamma}{k+c}, \quad \gamma_k = \frac{\gamma}{k^\rho}, \quad 0 < \rho \leq 1, \quad \gamma_k = \frac{\gamma}{k \ln k}. \quad (5)$$

Утверждение о сходимости в теореме 1 (и во многих случаях в дальнейшем) относится к величине

$$\Phi_k = \min_{0 \leq i \leq k} f(x^i) \quad (6)$$

— «рекордному» значению  $f(x)$  за  $k$  итераций.

**Теорема 1.** В методе (4) для выпуклой  $f(x)$ :  $\Phi_k \rightarrow f^* = \inf_{x \in \mathbb{R}^n} f(x)$ .

Подчеркнем, что здесь не требуется ни существования точки минимума, ни даже ограниченности  $f(x)$  снизу (возможно, что  $f^* = -\infty$ ).

**Доказательство.** Предположим, что  $f(x^k) \geq \tilde{f}$  для всех  $k$  и некоторого  $\tilde{f} > f^*$ . Выберем точку  $\tilde{x}$  такую, что  $f(\tilde{x}) < \tilde{f}$ . В силу непрерывности  $f(x)$  (лемма 3 § 1) найдется  $\rho > 0$  такое, что  $f(x) \leq \tilde{f}$  при  $\|x - \tilde{x}\| \leq \rho$ . В частности, для  $x_\rho = \tilde{x} + \rho \frac{\partial f(x^k)}{\partial f(x^k)}$  будет  $f(x_\rho) \leq \tilde{f}$ . С другой стороны,  $f(x_\rho) \geq f(x^k) + (\partial f(x^k), x_\rho - x^k) \geq \tilde{f} + (\partial f(x^k), \tilde{x} - x^k) + (\partial f(x^k), x_\rho - \tilde{x}) \doteq \tilde{f} + (\partial f(x^k), \tilde{x} - x^k) + \rho \|\partial f(x^k)\|$ , т. е.  $(\partial f(x^k), x_\rho - \tilde{x}) \|\partial f(x^k)\| \geq \rho$ .

Оценим теперь расстояние до  $\tilde{x}$  в процессе итераций:

$$\|x^{k+1} - \tilde{x}\|^2 = \|x^k - \tilde{x}\|^2 - 2\gamma_k \left( \frac{\partial f(x^k)}{\|\partial f(x^k)\|}, x^k - \tilde{x} \right) + \gamma_k^2 \leq \|x^k - \tilde{x}\|^2 - 2\gamma_k \rho + \gamma_k^2.$$

Поскольку  $\gamma_k \rightarrow 0$ , то найдется  $k_0$  такое, что  $\gamma_k \leq \rho$  при  $k \geq k_0$ . Поэтому при  $k \geq k_0$

$$\|x^{k+1} - \tilde{x}\|^2 \leq \|x^k - \tilde{x}\|^2 - \gamma_k \rho.$$

Суммируя эти неравенства по  $k$ , получаем  $\rho \sum_{k=k_0}^{\infty} \gamma_k \leq \|x^{k_0} - \tilde{x}\|^2$ ,

что противоречит условию  $\sum_{k=0}^{\infty} \gamma_k = \infty$ . Итак, неравенство  $f(x^k) \geq$

$\geq \tilde{f} > f^*$  при всех  $k$  невозможно, что эквивалентно условию  $\Phi_k \rightarrow f^*$ .  $\blacktriangle$

Можно получить и утверждения относительно сходимости  $x^k$  в случае непустоты множества точек минимума  $X^*$  (упр. 1).

Ясно, что метод (4) не может сходиться быстро — грубо говоря, расстояние до точки минимума не может быть меньше длины шага  $\gamma_k$ , а эта величина убывает медленно, так как должно выполняться условие  $\sum_{k=0}^{\infty} \gamma_k = \infty$ . В частности, можно показать, что в методе (4) заведомо не может быть сходимости со скоростью геометрической прогрессии. Кроме того, выбор  $\gamma_k$  из условий  $\gamma_k \rightarrow 0$ ,  $\sum_{k=0}^{\infty} \gamma_k = \infty$  неудовлетворителен, так как существует много таких последовательностей и совершенно не ясно, какую из них целесообразно выбрать. Поэтому описем другие способы регулировки длины шага.

В некоторых задачах бывает известно минимальное значение функции (обозначим его  $f^*$ ). Так, например, если систему совместных линейных уравнений

$$(a^t, x) = b_t, \quad t = 1, \dots, n, \quad x \in \mathbb{R}^n,$$

свести к минимизации функции

$$f(x) = \sum_{t=1}^n |(a^t, x) - b_t|$$

или функции  $f(x) = \max_{1 \leq t \leq n} |(a^t, x) - b_t|$ , то в обоих случаях

$f^* = 0$ . Значение  $f^*$  позволяет построить следующий вариант субградиентного метода, не содержащий произвольных параметров:

$$x^{k+1} = x^k - \frac{f(x^k) - f^*}{\|\partial f(x^k)\|^2} \partial f(x^k). \quad (7)$$

Геометрический смысл такого выбора длины шага показан на рис. 19.

**Теорема 2.** Пусть  $f(x)$  — выпуклая функция на  $\mathbb{R}^n$ , множество точек минимума  $X^*$  которой непусто. Тогда в методе (7)  $x^k \rightarrow x^* \in X^*$ . Оценка скорости сходимости следующая: для произвольной функции  $f$

$$\lim_{k \rightarrow \infty} \sqrt{k} (f(x^k) - f^*) = 0; \quad (8)$$

для функции, имеющей острый минимум, можно утверждать сходимость со скоростью геометрической прогрессии.

**Доказательство.** Пусть  $\tilde{x}$  — произвольная точка минимума. Тогда

$$\begin{aligned} \|x^{k+1} - \tilde{x}\|^2 &= \|x^k - \tilde{x}\|^2 - 2 \frac{\langle \partial f(x^k), x^k - \tilde{x} \rangle (f(x^k) - f^*)}{\|\partial f(x^k)\|^2} + \\ &+ \frac{(f(x^k) - f^*)^2}{\|\partial f(x^k)\|^2} \leq \|x^k - \tilde{x}\|^2 - \frac{(f(x^k) - f^*)^2}{\|\partial f(x^k)\|^2}. \end{aligned} \quad (9)$$

Отсюда следует, что  $(f(x^k) - f^*)/\|\partial f(x^k)\| \rightarrow 0$ . Так как последовательность  $x^k$  ограничена:  $\|x^k - \tilde{x}\| \leq \|x^0 - \tilde{x}\|$ , то (лемма 8 § 1)  $\|\partial f(x^k)\| \leq c$ . Поэтому  $f(x^k) \rightarrow f^*$ . Следовательно, найдется подпоследовательность  $x_{k_i} \rightarrow x^*$ , где  $x^*$  — некоторая точка минимума. Заменяя в полученной выше оценке  $\tilde{x}$  на  $x^*$ , получаем,

$$\text{что } \|x^k - x^*\| \text{ монотонно убывает, а } \|x_{k_i} - x^*\| \rightarrow 0. \text{ Отсюда}$$

$$\sum_{k=0}^{\infty} \frac{(f(x^k) - f^*)^2}{\|\partial f(x^k)\|^2} < \infty,$$

Перейдем к оценке скорости сходимости. Из (9)

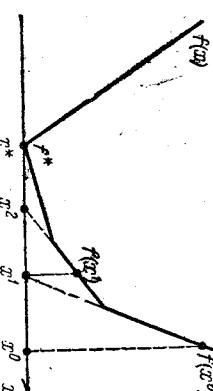


Рис. 19. Способ выбора длины  $a$  из ограниченности  $\|\partial f(x^k)\|$  шага в субградиентном методе.

следует  $\sum_{k=0}^{\infty} (f(x^k) - f^*)^2 < \infty$ . Если предположить, что  $\lim_{k \rightarrow \infty} \sqrt{k}(f(x^k) - f^*) > 0$ , то  $f(x^k) - f^* > a/\sqrt{k}$  для достаточно больших  $k$ , что противоречит условию  $\sum_{k=0}^{\infty} (f(x^k) - f^*)^2 < \infty$ . Итак,  $\lim_{k \rightarrow \infty} \sqrt{k}(f(x^k) - f^*) = 0$ .

Пусть теперь  $f(x)$  имеет острый минимум, т. е.  $f(x) - f^* \geq \alpha \|x - x^*\|$ . Тогда  $\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - (\alpha^2/c^2)\|x^k - x^*\|^2 = q\|x^k - x^*\|^2$ ,  $q = 1 - \alpha^2/c^2$ ,

что и означает сходимость со скоростью геометрической прогрессии. ▲

Знаменатель этой прогрессии может быть, однако, очень большим при  $k \geq 1$ , если линии уровня функции  $f(x)$  сильно вытянуты (т. е. если задача минимизации плохо обусловлена).

В случае, когда величина  $f^*$  неизвестна, метод можно модифицировать. Так, можно применять итерационный процесс

$$x^{k+1} = x^k - \frac{f(x^k) - f^*}{\|\partial f(x^k)\|^2} \partial f(x^k), \quad (10)$$

где  $\tilde{f}$  — некоторая оценка для  $f^*$ , и на основе поведения  $x^k$  пересчитывать  $\tilde{f}$ .

Ранее уже отмечалось, что итерационный процесс (7) можно применять и для минимизации гладких выпуклых функций, причем его скорость сходимости того же порядка, что и других «хороших» вариантов градиентного метода (см. (34) § 3 гл. 3).

### Упражнения.

1. Докажите следующий вариант теоремы 1. Если  $f(x)$  выпукла,  $\lambda^* \neq \emptyset$  и  $\sum_{k=0}^{\infty} \psi_k^2 < \infty$ , то в методе (4)  $x^k \rightarrow x^* \in X^*$ .
2. Что можно сказать о поведении следующих методов: а)  $x^{k+1} = x^k - \nu \partial f(x^k)/\|\partial f(x^k)\|$ ,  $\nu > 0$ ; б)  $x^{k+1} = x^k - \nu_k \partial f(x^k)$ ,  $\nu_k \rightarrow 0$ ,  $\sum_{k=0}^{\infty} \nu_k = \infty$ ; в)  $x^{k+1} = x^k - \nu_k \partial f(x^k)/\|\partial f(x^k)\|$ ,  $\nu_k = \nu_0 q^k$ ,  $q < 1$ ; г)  $x^{k+1} = x^k - \nu (f(x^k) - f^*) \partial f(x^k)/\|\partial f(x^k)\|^{-2}$ . Ответы. а) Метод «сходится с точностью до  $\Psi$ », т. е. существует функция  $\Psi(\nu) > 0$ ,  $\Psi(\nu) \rightarrow 0$  при  $\nu \rightarrow 0$ , такая, что  $\lim_{k \rightarrow \infty} \Phi_k \leq f^* + \Psi(\nu)$ ,  $\Phi_k = \min_{1 \leq i \leq k} f(x^i)$ ; б) если  $\|\partial f(x)\| \leq c$  для всех  $x$ , то справедливы утверждения теоремы 1; в) для случая острого минимума можно для данного  $x^0$  так выбрать  $\nu_0$  и  $q$ , что метод будет сходиться со скоростью геометрической прогрессии; г) при  $0 < \nu < 2$  справедливы утверждения теоремы 2.

### 3. $\varepsilon$ -субградиентный метод. Рассмотрим возможность замены субградиента на $\varepsilon$ -субградиент в методах типа (3). Такая замена может быть целесообразна, так как в ряде задач вычисление $\varepsilon$ -субградиента проще, чем градиента (см. лемму 13 в § 1).

Наиболее прямолинейный подход связан с заменой  $\partial f(x)$  в методе (4) на произвольный  $\varepsilon$ -субградиент. Однако если  $\varepsilon$  фиксировано, то такой метод может не сходиться — например, в соответствии с теоремой 2 § 2  $\varepsilon$ -субградиент может обращаться в 0 в любой точке, где значение  $f(x)$  отличается от оптимума меньше, чем на  $\varepsilon$ , поэтому метод остановится в любой такой точке. Значит, для сходимости нужно изменять  $\varepsilon$ , устремляя его к 0 в процессе итераций. Таким образом, приходим к методу

$$x^{k+1} = x^k - \frac{\partial_{\varepsilon_k} f(x^k)}{\left\| \partial_{\varepsilon_k} f(x^k) \right\|},$$

где  $\varepsilon_k \rightarrow 0$ ,  $\sum_{k=0}^{\infty} \varepsilon_k = \infty$ ,  $\varepsilon_k \rightarrow 0$  при  $k \rightarrow \infty$ . (11)

**Теорема 3.** В методе (11) для выпуклой функции  $f(x)$  имеем  $\Phi_k = \min_{1 \leq i \leq k} f(x^i) \rightarrow \inf f(x)$ .

Доказательство строится по той же схеме, что и в теореме 1. ▲

#### § 4. Другие методы

**1. Вводные замечания.** Как мы видели выше, субградиентный метод чрезвычайно прост по форме и сходится при минимальных предположениях о функции. Однако скорость его сходимости может быть мала. Во-первых, для гладких функций он переходит в градиентный метод, отличающийся от него обычных вариантов лишь способами выбора длины шага. Но мы уже знаем, что градиентный метод неэффективен для плохо обусловленных функций. Во-вторых, субградиентный метод в форме (4) § 3, как уже отмечалось, не может сходиться быстро (даже со скоростью геометрической прогрессии) ни для каких функций. Наконец, вариант (7) субградиентного метода, как следует из доказательства теоремы 2, также сходится медленно (как геометрическая прогрессия со знаменателем, близким к 1) для плохо обусловленных негладких функций. Итак, субградиентный метод не может считаться эффективным средством решения для выпуклых недифференцируемых задач, и возникает потребность в более мощных методах оптимизации.

В гладком случае такие методы строились на базе метода Ньютона, т. е. на основе квадратичной аппроксимации минимизируемой функции. Для негладких задач требуются иные идеи. Один возможный подход связан с кусочно-линейной аппроксимацией, являющейся естественной для недифференцируемых функций. Однако множество негладких выпуклых функций слишком «разнообразно», чтобы его можно было хорошо приблизить довольно «бедным» классом кусочно-линейных функций. Поэтому возможности такого подхода ограничены. Впрочем, оказывается, что задача минимизации произвольной выпуклой функции вообще слишком сложна, и в принципе невозможен метод, использующий только субградиенты и быстро сходящийся для всех функций данного класса.

**2. Многошаговые методы.** Простейший путь ускорения сходимости связан с использованием информаций, полученной на предыдущих итерациях. Пусть уже построены точки  $x^0, \dots, x^k$  и в них вычислены субградиенты  $\partial f(x^0), \dots, \partial f(x^k)$ . В силу соглашений

$$f(x^*) \geq f(x^i) + (\partial f(x^i), x^* - x^i)$$

можно утверждать, что точка минимума  $x^*$  лежит в области, задаваемой линейными неравенствами

$$Q_k = \{x: (\partial f(x^i), x - x^i) \leq f^* - f(x^i), i = 0, \dots, k\}, \quad (1)$$

а в случае, если величина  $f^* = f(x^*)$  неизвестна, в более широкой области

$$Q_k = \{x: (\partial f(x^i), x - x^i) \leq 0, i = 0, \dots, k\}. \quad (2)$$

Естественно добавить новую точку  $x^{k+1}$  так, чтобы в том или ином смысле максимально сократить эту область (рис. 20). Делается это различными способами. Ниже приводятся различные варианты подобных методов и результаты об их сходимости. Они, как правило, утверждают сходимость величины  $\Phi_k - f^*$  к 0 с той или иной скоростью, где

$$\Phi_k = \min_{x \in \mathbb{R}^n} f(x), \quad \Phi_k = \min_{0 \leq i \leq k} f(x^i).$$

Во всех методах предполагается, что задан многогранник  $Q_0$ , содержащий  $x^*$  — область априорной локализации минимума. Обычно в практических задачах легко указать, например, возможный диапазон изменения каждой переменной, и этот параллелепипед можно принять за  $Q_0$ .

В методе отсекающей гиперплоскости в качестве точки  $x^{k+1}$  выбирается точка минимума кусочно-линейной аппроксимации  $f(x)$ , определяемой значениями  $f(x^i)$  и  $\partial f(x^i)$ ,  $i = 0, \dots, k$ , на множестве  $Q_0$ . Иначе говоря,  $x^{k+1}$  — решение задачи линейного программирования:

$$\min z,$$

$$f(x^i) + (\partial f(x^i), x - x^i) \leq z, \quad i = 0, \dots, k, \\ x \in Q_0. \quad (3)$$

Здесь  $z \in \mathbb{R}^1$  — вспомогательная переменная — ордината аппроксимирующей функции (рис. 21). В этом методе мы впервые сталкиваемся с ситуацией, когда в задаче безусловной минимизации на каждой итерации нужно решать вспомогательную задачу линейного

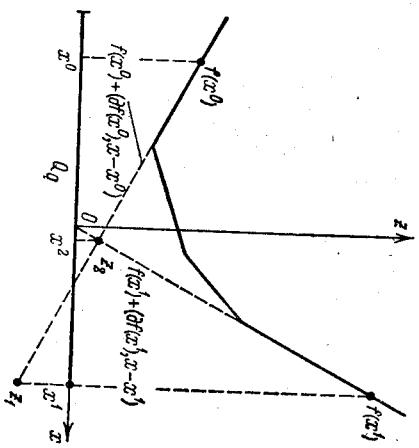


Рис. 21. Метод отсекающей гиперплоскости.

Программирования, т. е. задачу условной минимизации. Такая ситуация типична именно для негладких задач, требующих кусочно-линейную аппроксимацию. Наоборот, для задач с ограничениями, как мы увидим далее, нередко применяются методы, основанные на сведении задачи к безусловной минимизации. В этом, вообще говоря, нет противоречия, если

получающиеся вспомогательные задачи проще исходных. Для оценки эффективности подобных подходов нужно аккуратно оценивать сложность решения вспомогательных задач.

**Теорема 1.** Пусть  $f(x)$  — выпуклая функция на  $\mathbf{R}^n$ , а множество  $Q_0$  ограничено и содержит точку минимума  $x^*$ . Тогда в методе (3)  $\Phi_k \rightarrow f^*$ .

Доказательство. Пусть  $z_{k+1}$ ,  $x^{k+1}$  — решение задачи (3), тогда  $z_{k+1} \leq \max_{0 \leq i \leq k} [f(x^i) + (\partial f(x^i), x - x^i)]$  для всех  $x \in Q_0$  (см. рис. 21) и, в частности,  $z_{k+1} \leq \max_{0 \leq i \leq k} [f(x^i) + (\partial f(x^i), x^* - x^i)]$ .

В силу выпуклости  $f(x)$  имеем  $f(x^*) \geq f(x^i) + (\partial f(x^i), x^* - x^i)$ ,  $0 \leq i \leq k$ , т. е.  $f^* \geq \max_{0 \leq i \leq k} [f(x^i) + (\partial f(x^i), x^* - x^i)]$ . Сопоставляя полученные неравенства, получаем  $z_{k+1} \leq f^*$ . С другой стороны,  $f(x^{k+1}) \geq f^*$ , т. е.  $z_{k+1} \leq f^* \leq f(x^{k+1})$ . Предположим, что  $f(x^{k+1}) - z_{k+1} \geq \varepsilon > 0$  для всех  $k \geq k_0$ . Тогда

$$\begin{aligned} f(x^i) &\geq f(x^{k+1}) + (\partial f(x^{k+1}), x^i - x^{k+1}) \geq z_{k+1} + \varepsilon + \\ &+ (\partial f(x^{k+1}), x^i - x^{k+1}) \geq f(x^i) + (\partial f(x^i), x^{k+1} - x^i) + \varepsilon + \\ &+ (\partial f(x^{k+1}), x^i - x^{k+1}) \geq f(x^i) + \varepsilon - 2L \|x^i - x^{k+1}\|, \end{aligned}$$

где  $L = \max \|\partial f(x)\|$  (эта величина ограничена по лемме 8 § 1). Отсюда  $\|x^{k+1} - x^i\| \geq \varepsilon/2L$  для всех  $i = 0, \dots, k$  и всех  $k \geq k_0$ . Это противоречит компактности  $Q_0$ . Поэтому  $\lim_{k \rightarrow \infty} (f(x^k) - z_k) = 0$ ,

а так как  $0 \leq f(x^k) - f^* \leq f(x^k) - z_k$ , то  $\lim_{k \rightarrow \infty} f(x^k) = f^*$ .  $\blacktriangle$

Вопрос о скорости сходимости метода исследован мало. Для некоторых классов задач (например, задач с острым минимумом) он, по-видимому, сходится быстро. Так, для кусочно-линейных задач этот метод конечен. В общем же случае он может сходиться очень медленно. Рассмотрим одномерную задачу

$$\min p^{-1} x^p, \quad 0 \leq x \leq 1, \quad x_0 = 0, \quad x_1 = 1.$$

Каждая вспомогательная задача (3) имеет неединственное решение; предполагается, что в качестве  $x_{k+1}$  берется наибольшее из решений. Тогда  $x_{k+1} = x_k - p^{-1} x_k = q x_k$ ,  $x_k = q^{k-1}$ ,  $q = 1 - p^{-1}$ , и для больших  $p$  знаменатель прогрессии  $q$  близок к 1. Для многомерных задач, по-видимому, нельзя обеспечить линейной скорости сходимости даже для гладких сильно выпуклых функций.

Недостатком метода является необходимость решать задачу линейного программирования, в которой число ограничений возрастает. Можно модифицировать метод, устранив этот дефект — грубо говоря, нужно оставлять лишь те ограничения, которые удовлетворяются как равенства. Можно также при решении каждой новой задачи использовать решение предыдущей как

начальное приближение — для этого следует перейти к двойственной задаче.

Другой способ выбора точки  $x^{k+1}$  применяется в методе чебышевских центров. Здесь предлагается в качестве  $x^{k+1}$  брать чебышевский центр многогранника  $Q_k$  вила (1) или (2), т. е. точку, максимум расстояния которой от граней многогранника минимален. Иными словами,  $x^{k+1}$  — решение либо задачи

$$\max z,$$

$$\left( \frac{\partial f(x^i)}{\|\partial f(x^i)\|}, x - x^* \right) + z \leq 0, \quad i = 0, \dots, k, \quad x \in Q_0, \quad (4)$$

либо, если  $f^*$  известно, задачи

$$\max z, \quad \left( \frac{\partial f(x^i)}{\|\partial f(x^i)\|}, x - x^* \right) + z \leq f^* - f(x^i), \quad i = 0, \dots, k, \quad x \in Q_0. \quad (5)$$

Можно показать, что для (4) и (5) справедлив аналог теоремы 1. Относительно скорости сходимости метода отметим следующее. Нетрудно видеть, что для одномерного случая метод

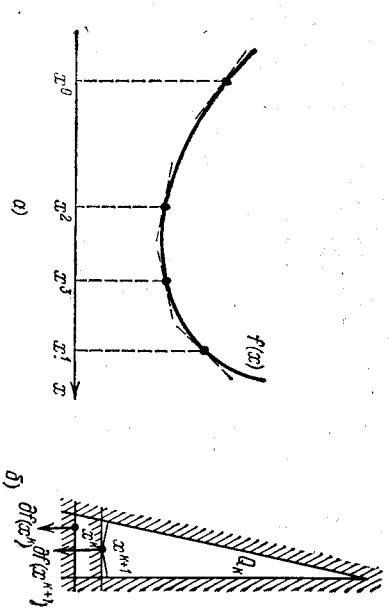


Рис. 22. Метод чебышевских центров: а) одномерный случай; б) двумерный случай.

(4) превращается в *метод дихотомии*: в качестве  $x^{k+1}$  берется середина наименьшего отрезка с концами в точках  $x^i$ , для которых  $\partial f(x^i)$  имеет разные знаки на концах отрезка (рис. 22, а). Отсюда следует, что метод (4), в отличие от (3), не является конечным для кусочно-линейных  $f(x)$ . Для многомерного случая, как видно из рис. 22, б, скорость сходимости может быть достаточно медленной.

Можно точку  $x^{k+1}$  строить иначе. Выберем произвольное множество индексов  $I$  из набора  $0, \dots, k$ . Например, может быть

$I = \{0, \dots, k\}$  или  $I = \{k\}$  или  $I = \{k, k-1\}$  и т. п. В качестве  $x^{k+1}$  возьмем точку, ближайшую к  $x^k$  и удовлетворяющую ограничениям вида (1) для набора  $I$ . Иначе говоря,  $x^{k+1}$  — решение задачи

$$\min \|x - x^k\|^2,$$

(предполагается, что величина  $f^*$  известна). Вспомогательная задача (6) является задачей квадратичного программирования с минимизируемой функцией вида  $\|x - a\|^2$ . Таким образом, она сводится к проектированию точки  $x^k$  на многогранник, задаваемый линейными ограничениями (6). От этой задачи удобно перейти к двойственной — именно, искать решение в виде (см. § 4 гл. 10):

$$x^{k+1} = x^k - \sum_{i \in I} \lambda_i^k \partial f(x^i), \quad (7)$$

где  $\lambda_i^k$  — решение задачи

$$\min_{\lambda_i \geq 0} \left[ \sum_{i \in I} \lambda_i \partial f(x^i) \right]^2 - 4 \sum_{i \in I} \lambda_i (\partial f(x^i), x^k - x^i) -$$

$$- 4 \sum_{i \in I} \lambda_i (f(x^i) - f^*)]. \quad (8)$$

Эта задача минимизации квадратичной функции на неограниченном ортантне решается достаточно просто (см. § 3 гл. 7). Ясно, что если  $I = \{k\}$ , то метод совпадает с субградиентным методом (7) § 2.

Достоинством метода (6) по сравнению с ранее рассматривавшимися (3), (4) и (5) является то, что в нем множество  $I$  может содержать не все предыдущие индексы, так что вспомогательные задачи, решаемые на каждом шаге, могут быть небольшой размерности. К недостаткам его относится необходимость знания  $f^*$ .

Наконец, оригинальный подход к выбору  $x^{k+1}$  предлагается в методе центров тяжести. Пусть

$$Q_k = \{x \in Q_0: (\partial f(x^i), x - x^i) \leq 0, \quad i = 1, \dots, k\}, \quad (9)$$

$x^{k+1}$  — центр тяжести  $Q_k$ .

Такой выбор объясняется следующим результатом, относящимся к теории выпуклых тел.

**Лемма 1.** Пусть  $Q$  — выпуклое тело (т. е. множество с неустойчивой внутренностью) в  $\mathbb{R}^n$ ,  $a$  — его центр тяжести,  $L$  — гиперплоскость, проходящая через  $a$ ,  $v_1$  и  $v_2$  — объемы частей, на которые  $L$  делит  $Q$  (рис. 23). Тогда

$$\frac{v_i}{v} \leq 1 - \left( \frac{n}{n+1} \right)^n \leq 1 - \frac{1}{e}, \quad i = 1, 2, \quad v = v_1 + v_2. \quad (10)$$

Для точек, отличных от  $a$ , правая часть в (10) может лишь уменьшаться. ▲

Иными словами, объем части, «отрезаемой» от  $Q$  гиперплоскостью, проходящей через центр тяжести, составляет всегда не менее чем  $e^{-1}$  часть от объема  $Q$ . Для других точек объем «отрезаемой» части может оказаться меньше. Это и объясняет причину выбора центра тяжести  $Q_k$  в качестве  $x^{k+1}$ .

**Теорема 2.** Пусть  $f(x)$  — выпуклая функция на  $\mathbb{R}^n$ ,  $Q_0$  — ограниченное, замкнутое выпуклое множество. Тогда в методе (9)

$$f_k - f^* \leq c q^k, \quad q = \left(1 - \left(\frac{n}{n+1}\right)^n\right)^{1/n} \leq \left(1 - \frac{1}{e}\right)^{1/n} = 1 - \frac{1}{ne} + o\left(\frac{1}{n}\right),$$

$$c = \max_{x \in Q_0} (f(x) - f^*). \quad (11)$$

**Доказательство.** В соответствии с леммой 1 объем  $v_k$  множества  $Q_k$  удовлетворяет неравенству  $v_{k+1} \leq v_k \beta$ ,  $\beta = 1 - \left(\frac{n}{n+1}\right)^n$ , т. е.  $v_k \leq v_0 \beta^k$ . Возьмем произвольную точку minimuma  $x^* \in Q_k$  и построим множество  $S$ , получающееся из  $Q_k$  подобным преобразованием с центром в  $x^*$  и коэффициентом растяжения  $\alpha = \beta^{-k/n}$ , т. е.  $S = \{x: x^* + \alpha y, y \in Q_k\}$ . Тогда его объем  $v(S) = \alpha^n v_k \leq \alpha^n v_0 \beta^k = v_0$ . Поэтому множество  $Q_0$  не может помешаться строго внутри  $S$  и, следовательно, найдется  $z \in Q_0, z \in S^0$ . Отсюда следует, что  $u = (1 - \alpha^{-1})x^* + \alpha^{-1}z \in Q_0^0$  (так как  $z$  получается из  $u$  указанным выше растяжением). Но если  $u \in Q_k^0$ , то

(в силу определения  $Q_k$ ) найдется  $i, 1 \leq i \leq k$ , такое, что  $(\partial f(x^i), u - x^i) \geq 0$ . Поэтому  $f(u) \geq f(x^i) + (\partial f(x^i), u - x^i) \geq f(x^i) \geq \Phi_k$ . Используя выпуклость  $f(x)$ , находим  $f(u) = f((1 - \alpha^{-1})x^* + \alpha^{-1}z) \leq (1 - \alpha^{-1})f^* + \alpha^{-1}f(z) \leq f^* + c/\alpha$ ,  $c = \max_x (f(x) - f^*)$ , причем  $c < \infty$  в силу непрерывности  $f(x)$  и ограниченности  $Q_0$ . Итак,

$$\Phi_k - f^* \leq c \alpha^{-1} = c \left(1 - \left(\frac{n}{n+1}\right)^n\right)^{k/n}. \quad \blacktriangle$$

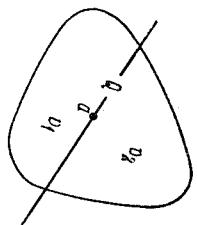


Рис. 23. Лемма о центре тяжести.

Для  $n = 1$  множество  $Q_k$  — отрезок, и  $x^{k+1}$  — его середина.

Поэтому метод центра тяжести переходит в метод дихотомии. Для  $n = 2$  можно построить способ отыскания центра тяжести, основываясь на том, что центром тяжести треугольника является точка пересечения его медиан, а центр тяжести объединения двух фигур вычисляется по формуле  $\tilde{x} = \alpha \tilde{x}_1 + (1 - \alpha) \tilde{x}_2$ , где  $\tilde{x}_1, \tilde{x}_2$  — центры тяжести  $A_1, A_1, A_2$  (причем  $A = A_1 \cup A_2$ ),  $\alpha = s_2/(s_1+s_2)$ ,  $s_1, s_2$  — площади  $A_1, A_2$ . Производя триангуляцию

$Q_k$  для  $n = 2$ , можно таким образом найти  $x^{k+1}$ . Для  $n \geq 2$  отыскание центра тяжести многогранника превращается в чрезвычайно громоздкую задачу, и метод вряд ли может быть реализован.

Однако метод центров тяжести представляет большой теоретический интерес. Во-первых, для него удается получить оценку скорости сходимости, зависящую только от размерности пространства и « начальной неопределенности » — величины  $\max_{x \in Q_0} f(x) - \min_{x \in Q_0} f(x)$ , но не от индивидуальных характеристик функции типа ее обусловленности. Все ранее приводившиеся оценки такими свойствами не обладали. При этом для задач небольшой размерности скорость сходимости достаточно велика. Действительно, из (11) видно, что за  $n$  итераций можно увеличить точность решения приблизительно в  $e$  раз. Таким образом, при  $n = 10$ , чтобы получить решение с точностью до 0,1% (т. е. добиться соотношения  $\Phi_k - f^* \leq \max_{x \in Q_0} (f(x) - f^*) \cdot 10^{-3}$ ) нужно сделать порядка  $11 \ln 10^3 \sim 190$  итераций — сравнительно не большое число. Во-вторых, как мы увидим далее, этот метод является в определенном смысле оптимальным.

#### Упражнения.

1. Покажите, что если  $f(x)$  — кусочно-линейная функция,  $I = \{k, \dots, m\}$ , а  $t$  достаточно велико, то метод (6) конечен.
2. Покажите, что для любой функции  $f(x)$  метод центров тяжести не может сходиться слишком быстро — именно,  $v_k \geq e^{-kv_0}$ , где  $v_0$  — объем  $Q_k$ .

**3. Оптимальные методы.** Для задачи безусловной минимизации выпуклой функции можно установить потенциальные возможности любого метода, использующего лишь субградиенты и значения функции. Имеет место следующий результат, который мы приведем в не совсем четкой, но наглядной формулировке.

**Теорема 3** (Немировский — Юдин). Для любого метода минимизации функции  $f(x)$ ,  $x \in \mathbb{R}^n$ , использующего значения  $f(x)$  и  $\partial f(x)$ , найдется такая выпуклая функция, для которой данный метод сходится (по функции) не быстрее, чем геометрическая прогрессия со знаменателем  $1 - c/n$ , или не быстрее чем  $O(1/\sqrt{k})$  равномерно по размерности (здесь  $c$  — некоторая абсолютная константа). ▲

Мы не приводим доказательства этого утверждения, так как оно потребовало бы строгого (и достаточно громоздкого) определения понятия «любой метод, использующий значения  $f(x)$  и  $\partial f(x)$ », уточнения имеющейся априорной информации о функции (начальное приближение, область локализации минимума, граничные для  $f(x)$  и  $\partial f(x)$  и т. п.). Идея доказательства состоит в том, что при заданных  $x^0, \dots, x^k, f(x^0), \dots, f(x^k), \partial f(x^0), \dots, \partial f(x^k)$  строится кусочно-линейная выпуклая функция с указанными значениями функции и субградиента в данных точках,

но у которой минимум по возможности сильно отличается от величин  $f(x^i)$ .

Сопоставляя этот результат с полученными выше оценками скорости сходимости, приходим к важному выводу: скорость сходимости метода центров тяжести не может быть по порядку превзойдена ни для какого метода оптимизации, использующего ту же информацию (т. е. значения  $f(x)$  и  $\partial f(x)$ ). Иными словами, этот метод является в определенном смысле оптимальным, и попытки создать более быстро сходящиеся методы заведомо безнадежны.

Однако к этому выводу нужно относиться с осторожностью. Во-первых, он относится к широкому классу «всех выпуклых функций». В действительности редко приходится иметь дело с «произвольными» выпуклыми функциями. Обычно минимизируемая функция принадлежит к более узкому классу (например, сильно выпукла, или имеет острый минимум, или имеет вид  $\max_{1 \leq i \leq k} f_i(x)$ , где  $f_i(x)$  — гладкие функции и т. д.). Для более узких классов могут существовать и более эффективные методы. Во-вторых, вывод имеет минимаксный характер — найдется функция, «плохая» для данного метода. Однако при минимизации конкретной функции метод может сходиться и значительно быстрее, чем для «наихудшего» случая. В то же время метод центра тяжести сходится одинаково как для «хороших», так и для «плохих» функций. Наконец, при оценке скорости сходимости в теореме Немировского — Юдина учитывается лишь число вычислений  $f(x')$  и  $\partial f(x')$ . Объем же вычислительной работы по решению возникающих вспомогательных задач не принимается во внимание. Поэтому, скажем, трудность отыскания центра тяжести многогранника не учитывается, так как она не связана с дополнительными вычислениями функции и субградиента. В действительности, конечно, метод центра тяжести не может быть рекомендован не только как оптимальный, но и как разумный метод оптимизации при  $n > 2$ . Все это показывает, что вопрос о выборе метода минимизации, даже при наличии теоретических выводов об «оптимальном» в определенном смысле методе, остается весьма сложным.

**4. Методы с растяжением пространства.** Естественна попытка модифицировать метод центров тяжести, устранив его основные недостатки — трудоемкую операцию отыскания центра тяжести и необходимость запоминания  $\partial f(x')$ , полученных на всех предыдущих итерациях, но сохранив скорость сходимости. Это можно сделать следующим образом (рис. 24). Если поместить многогранник  $Q_k$  внутрь шара, то центр тяжести последнего, конечно, найти легко — он совпадает с центром шара. Выбирая эту точку в качестве  $x^{k+1}$  и вычислив  $\partial f(x^{k+1})$ , «отсекаем» поло-

эллипсоид минимального объема. Делая линейное преобразование пространства, превращаем этот эллипсоид в шар, после чего процедура повторяется. При таком подходе нет нужды запоминать сам многогранник  $Q_k$  и задающие его ограничения, т. е. величины  $\partial f(x^i)$ ,  $i = 0, \dots, k$ . Достаточно на  $k$ -м шаге помнить точку  $x^k$  и линейное преобразование пространства, задаваемое матрицей  $H_k$ . При этом приходим к методу эллипсоидов:

$$x^{k+1} = x^k - \gamma_k H_k \partial f(x^k), \quad \gamma_k = \frac{\rho}{n+1} \left( \frac{n}{\sqrt{n^2 - 1}} \right)^k, \quad (12)$$

$$H_{k+1} = H_k - \frac{2}{n+1} \frac{H_k \partial f(x^k) \partial f(x^k)^T H_k}{(H_k \partial f(x^k), \partial f(x^k)^T H_k)}, \quad H_0 = I.$$

Здесь  $\rho$  — радиус исходного шара с центром в  $x^0$ , в котором локализуется точка минимума.

**Теорема 4.** Для метода (12) в пространстве  $R^n$  справедлива оценка

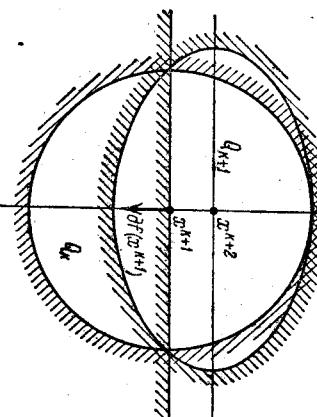
$$\Phi_k - f^* \leq c q^k, \quad c = \max_{\|x-x^*\| \leq \rho} (f(x) - f^*), \quad (13)$$

$$q = n(n-1)^{-\frac{n-1}{2n}} (n+1)^{-\frac{n+1}{2n}}.$$

Оно основано на том легко проверяемом факте, что объем эллипса, описанного вокруг полушиара (рис. 24), в  $2q^{1/n}$  раз больше объема этого полушара. Поэтому на каждом шаге объем области локализации минимума уменьшается в  $q^{1/n}$  раз. Дальнейшая часть доказательства строится так же, как и в теореме 1. ▲

Таким образом, поведение метода (12) сохраняет черты метода центра тяжести (сходимость со скоростью геометрической прогрессии со знаменателем, не зависящим от свойств минимизируемой функции, а лишь от размерности пространства). Однако знаменатель прогрессии здесь ближе к 1, именно,  $q \sim 1 - 1/(2n^2)$  вместо  $q \sim 1 - 1/(en)$  в методе (9). Для больших размерностей пригрешив скрытости значителен, и метод (12) становится малоэффективным. Например, при  $n = 10$  нужно порядка 200 итераций, чтобы увеличить точность (по функции) в  $e$  раз, при  $n = 30$  — порядка 2000 итераций.

Рис. 24. Метод эллипсоидов.



В своих работах Н. З. Шор пришел к методам вида (12) из других соображений. Он предложил объединять субградиентный метод с процедурой *растяжения пространства*. Последняя проводится либо в направлении последнего субградиента, либо в направлении разности двух последних субградиентов. Величина растяжения задается некоторым параметром, выбираемым эвристически. Таким образом (см. упр. 3 и 4) получаются методы вида

$$x^{k+1} = x^k - \gamma_k H_k \partial f(x^k), \quad H_{k+1} = H_k - \left( 1 - \frac{1}{\alpha_k^2} \right) \frac{H_k s^k (s^k)^T H_k}{(H_k s^k, s^k)}, \quad H_0 = I, \quad (14)$$

где  $\alpha_k$  — коэффициент растяжения пространства на  $k$ -й итерации,  $\gamma_k$  определяет длину шага,  $s^k$  — направление растяжения. Все эти величины могут выбираться различными способами. Например:

$$s^k = \partial f(x^k), \quad \gamma_k = \frac{2(f(x^k) - f^*)}{(H_k \partial f(x^k), \partial f(x^k))}, \quad \alpha_k = \infty, \quad (15)$$

$$s^k = \partial f(x^k), \quad \gamma_k = \lambda \frac{f(x^k) - f^*}{(H_k \partial f(x^k), \partial f(x^k))}, \quad \alpha_k = \alpha, \quad (16)$$

$$s^k = \partial f(x^k) - \partial f(x^{k-1}), \quad \gamma_k = \arg\min_{\gamma} f(x^k - \gamma H_k \partial f(x^k)), \quad (17)$$

$$\alpha_k = \alpha,$$

где величина  $f^* = \min f(x)$  предполагается известной.

Очевидна их связь с методами переменной метрики для минимизации гладких функций, описанными в § 3 гл. 3. Методы пригодны как для гладкой, так и для негладкой оптимизации. Их сходимость для квадратичного случая утверждается следующей теоремой.

**Теорема 5.** Пусть  $f(x) = (Ax, x)/2 - (b, x)$ ,  $A > 0$ . Тогда методы (14), (15) и (17) с  $\alpha = \infty$  конечны:  $x^n = x^*$   $\Rightarrow A^{-1}b$ . ▲

Относительную сходимость и скорость сходимости методов (14) в общем случае известно немного. На основе теоремы Н. Мирского — Юдина можно утверждать, что для произвольной выпуклой функции эти методы не могут сходиться быстрее, чем геометрическая прогрессия со знаменателем  $1 - 1/(cn)$ . Н. З. Шор рассматривает иной класс функций, удовлетворяющих условию

$$N(f(x) - f^*) \leq (\partial f(x), x - x^*) \leq M(f(x) - f^*). \quad (18)$$

Такие функции естественно называть *приближенно-однородными* (ср. с (30) § 3 гл. 3). Для них можно доказать, что если  $\alpha_k = \alpha = (M + N)/(M - N)$ ,  $\lambda = 2MN/(M + N)$ ,

$$(19)$$

то метод (14), (16) сходится со скоростью геометрической прогрессии со знаменателем  $\alpha^{1/n}$ .

$$\Phi_k - f^* \leq c\sqrt{k}\alpha^{-k/n}. \quad (20)$$

Поэтому, чем ближе  $M$  к  $N$  (т. е. чем ближе функция к однородной), тем больше  $\alpha$  и тем быстрее сходится метод. В пределе для однородной функции ( $M = N$ ) можно взять  $\alpha = \infty$ , и метод оказывается конечным — этот факт для квадратичной функции ( $M = N = 2$ ) уже отмечался в теореме 5.

#### Упражнения.

3. Пусть для некоторого  $\alpha > 0$  и  $s \in \mathbf{R}^n$ ,  $\|s\| = 1$ ,  $R_\alpha(s)$  — линейный оператор в  $\mathbf{R}^n$ , действующий по формуле  $R_\alpha(s)x = x + (\alpha - 1)ss^T x$ . Приверите, что  $R_\alpha^T(s) = R_\alpha(s)$  — оператор растяжения в  $\alpha$  раз в направлении  $s$ , т. е.  $R_\alpha(s)s = \alpha s$ ,  $R_\alpha(s)x = x$  для  $(x, s) = 0$ .

4. Покажите, что  $H_k$  в (14) — результат последовательного применения операторов растяжения, именно,  $H_k = P_k P_{k'}^T P_0 = I$ ,  $P_{i+1} = P_i R_{\alpha_i}^{-1}(s^i)$ .

5. Убедитесь, что при  $\alpha = 1$  (т. е. когда растяжение не производится) метод (14), (16) с  $\lambda = 1$  переходит в субградиентный метод (7) из § 3.

6. Докажите теорему 5 и сопоставьте ее с результатами § 3 гл. 3, что означает выбор  $\alpha_k = \infty$ ?

## § 5. Влияние помех

**1. Постановка задачи.** Рассмотрим поведение субградиентного метода минимизации выпуклой функции  $f(x)$  на  $\mathbf{R}^n$  при наличии помех.

Пусть

$$x^{k+1} = x^k - y_k s^k, \quad s^k = \partial f(x^k) + r^k. \quad (1)$$

Здесь  $r^k$  — помехи, наложенные на субградиент. Их природа может быть различна — погрешности вычисления, ошибки измерений, приближенные формулы и т. д. С формальной точки зрения помехи могут быть абсолютными или относительными, детерминированными или случайными. Проанализируем различные типичные случаи. Нас будут интересовать вопросы сходимости, оценки скорости сходимости, а также разумные способы выбора  $y_k$ , т. е. те же задачи, которые в гл. 4 решались для гладкого случая.

**2. Абсолютные детерминированные помехи.** Пусть ошибки вычисления субградиента удовлетворяют условию

$$\|r^k\| \leq \varepsilon, \quad (2)$$

где  $\varepsilon$  — абсолютный уровень помех. Как мы видели, в гладких задачах такие помехи приводили к нарушению сходимости — градиентный метод склонился лишь в некоторую окрестность минимума, размеры которой зависели от  $\varepsilon$  и от обусловленности задачи. Для негладких задач ситуация может оказаться иной —

при низком уровне помех в случае острого минимума сходимость

может сохраняться при специальном способе выбора  $y_k$ . Это является следствием того, что  $\partial f(x)$  не стремится к 0 при приближении к острому минимуму.

**Теорема 1.** Пусть  $f(x)$  — выпуклая функция на  $\mathbf{R}^n$ ,  $x^* - ee$  точка острого минимума, т. е.  $f(x) - f(x^*) \geq \alpha \|x - x^*\|$ ,  $\alpha > 0$ . Пусть  $\vartheta < \alpha$ . Тогда для всякого  $x^0$  найдется числа  $\gamma_0 > 0$ ,  $q < 1$ , такие, что в методе (1) при  $y_k = \gamma_0 q^k$  будет

$$\|x^k - x^*\| \leq \|x^0 - x^*\| q^k. \quad (3)$$

Чтобы воспользоваться приведенным в теореме 1 способом выбора длины шага, нужно располагать подробной информацией о задаче (иметь оценки для  $L$ ,  $\alpha$ ,  $\varepsilon$ ,  $\|x^0 - x^*\|$ ). При отсутствии такой правильное назначение  $\gamma_0$  и  $q$  может привести к остановке метода не в точке минимума. Мы не будем обсуждать более реалистические способы регулировки длины шага — для нас важнее принципиальная возможность сходимости субградиентного метода для негладких задач при наличии абсолютных помех со скоростью геометрической прогрессии.

**3. Относительные детерминированные помехи.** Пусть задан относительный уровень помех:

$$\|r^k\| \leq \alpha \|\partial f(x^k)\|. \quad (4)$$

В гладких задачах сходимость имеет место при любом  $\alpha < 1$  (теорема 2 § 2 гл. 4). Негладкие функции вновь вносят свои особенности. Приведем краткое анализ сходимости. Условие псевдоградиентности алгоритма (1) относительно функции Ляпунова

$$V(x) = \|x - x^*\|^2/2 \quad (5)$$

имеет вид  $(s^k, x^k - x^*) \geq 0$ . Но

$$(s^k, x^k - x^*) = (\partial f(x^k) + r^k, x^k - x^*) \geq (\cos \varphi_k - \alpha) \|\partial f(x^k)\| \|x^k - x^*\|, \quad (6)$$

где  $\varphi_k$  — угол между  $\partial f(x^k)$  и  $x^k - x^*$ ,  $0 \leq \varphi_k \leq \pi/2$ . Поэтому, если

$$0 \leq \varphi_k \leq \Phi < \pi/2, \quad \alpha \leq \cos \Phi,$$

то условие псевдоградиентности выполняется. Условие (6) существенно более ограничительно, чем условие  $\alpha < 1$ . Чем хуже обусловленность функции, тем меньше  $\cos \Phi$  и тем чувствительнее метод к относительным помехам. На рис. 18 видно, что даже малая ошибка в определении направления субградиента приводит к тому, что это направление не приближает к точке минимума. По этой причине кажущаяся естественным обобщение субградиентного метода

$$x^{k+1} = x^k - y_k H \partial f(x^k), \quad (7)$$

где  $H > 0$  — некоторая матрица, вообще говоря, может и не сходитьсь,

**4. Абсолютные стохастические помехи.** Пусть помехи  $r^k$  случайны, взаимно независимы, центрированы и имеют ограниченную дисперсию:

$$\mathbf{M}r^k = 0, \quad \mathbf{M}\|r^k\|^2 \leq \sigma^2. \quad (8)$$

**Теорема 2.** Пусть  $f(x)$  — выпуклая функция,  $\|\partial f(x)\| \leq c$  для всех  $x$ , существует точка минимума  $x^*$ , выполняется условие (8) и

$$\sum_{k=0}^{\infty} \gamma_k = \infty, \quad \sum_{k=0}^{\infty} \gamma_k^2 < \infty. \quad (9)$$

Тогда в методе (1)  $\min_{0 \leq i \leq k} f(x^i) \rightarrow f(x^*)$  ( $n, n$ ). ▲

Таким образом, здесь, как и для гладкого случая, при аддитивных случайных помехах любого уровня имеет место сходимость, если  $\gamma_k$  удовлетворяет условию (9). Различие гладкого и негладкого случаев заключается в том, что в первом наличие помех приводит к необходимости изменения способа регулировки длины шага (приходится выбирать  $\gamma_k \rightarrow 0$  вместо  $\gamma_k \equiv \gamma$ ), во втором же нет столь сильной разницы между ситуациями с помехой и без помех (и в той, и в другой ситуациях следует брать  $\gamma_k \rightarrow 0$ ). Относительную скорость сходимости в условиях (8) не все ясно. Если  $f(x)$  сильно выпукла, то, выбирая  $\gamma_k = \gamma/k$  при достаточно большом  $\gamma$ , можно добиться сходимости порядка  $O(1/k)$ ; это доказывается по стандартной схеме. Однако в более типичной для негладких задач ситуации с острым минимумом скорость сходимости недостаточно исследована.

#### Упражнение.

1. Покажите, что если  $f(x)$  имеет острый минимум с константой  $b_*$ , то для всех  $x$  из области  $S = \{x : \|x - x^*\| \leq p\}$  справедливо неравенство  $(\partial f(x))_x \geq (l/L) \|\partial f(x)\| \|x - x^*\|$ , где  $L = \max_{x \in S} \|\partial f(x)\|$ , т. е. выполняется (6) с

$$\cos \varphi = l/L.$$

#### § 6. Поисковые методы

Рассмотрим задачу минимизации выпуклой функции  $f(x)$  в сущности, когда единственной доступной информацией о  $f(x)$  являются ее значения в произвольной точке.

**1. Одномерный случай.** Поиск минимума одномерной выпуклой функции  $f(x)$  на отрезке  $[a, b] \subset \mathbb{R}$  легко организовать на основе следующего геометрически очевидного соображения

$$(1 + \alpha)/2 = \beta, \quad \beta^2 = 1 - \beta, \quad \beta = (\sqrt{5} - 1)/2 \quad (2)$$

(рис. 25, а). Если вычислены значения  $f(x)$  в двух точках  $x_1, x_2$ ,  $a < x_1 < x_2 < b$ , то точка минимума  $x^*$  не может находиться на отрезке  $[a, x_1]$ , если  $f(x_1) > f(x_2)$ , и на отрезке  $[x_2, b]$ , если  $f(x_2) > f(x_1)$  (если  $f(x_1) = f(x_2)$ , то одна из точек минимума принадлежит  $[x_1, x_2]$ ). Поэтому после вычисления двух значе-

ний функции область локализации минимума может быть сокращена. Простейший алгоритм, реализующий эту идею, рассказывает точки на каждом отрезке симметрично относительно его центра (рис. 25, б)):

$$a_0 = a, \quad b_0 = b, \quad e_k = \alpha(b_k - a_k)/2, \quad 0 < \alpha < 1, \quad (1)$$

$$a_{k+1} = \begin{cases} a_k, & \text{если } f((a_k + b_k)/2 - e_k) < f((a_k + b_k)/2 + e_k), \\ (a_k + b_k)/2 - e_k & \text{в противном случае,} \end{cases}$$

$$b_{k+1} = \begin{cases} b_k, & \text{если } f((a_k + b_k)/2 + e_k) > f((a_k + b_k)/2 - e_k), \\ (a_k + b_k)/2 + e_k & \text{в противном случае,} \end{cases}$$

$$a_{k+1} = (a_k + b_k)/2 - e_k, \quad b_{k+1} = (a_k + b_k)/2 + e_k, \quad \Rightarrow f((a_k + b_k)/2 + e_k) = f((a_k + b_k)/2 - e_k).$$

Очевидно,  $0 \leq b_{k+1} - a_{k+1} \leq (1 + \alpha)(b_k - a_k)/2$ , так что длина отрезка, на котором локализуется минимум, сокращается на

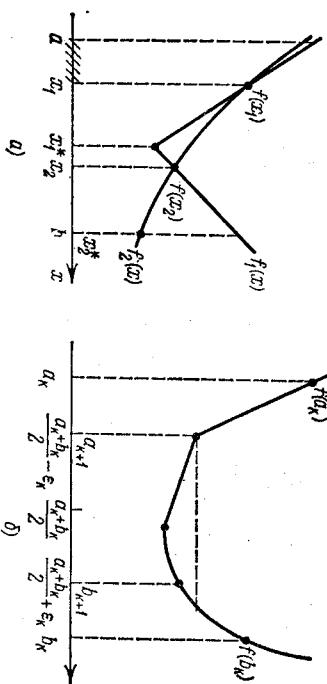


Рис. 25. Одномерный поиск.

Каждой итерации примерно в 2 раза, если  $\alpha$  мало. Ясно, что (1) при  $\alpha \ll 1$  есть просто разностный аналог метода дихотомии (§ 3).

Можно поступить более экономно, если использовать вычисленные ранее значения функции (одно из них на  $[a_{k+1}, b_{k+1}]$  за- ведомо найдено на предыдущей итерации). При этом, если выбрать  $\alpha$  из соотношения

$$(1 + \alpha)/2 = \beta, \quad \beta^2 = 1 - \beta, \quad \beta = (\sqrt{5} - 1)/2 \quad (2)$$

(уравнение «золотого сечения» отрезка), то одна из точек  $(a_{k+1} + b_{k+1})/2 \pm \epsilon_{k+1}$  будет совпадать со старой  $(a_k + b_k)/2 \mp \epsilon_k$ , т. е. каждая итерация будет требовать лишь одного вычисления функции. При этом в методе деления пополам (1) с  $\alpha \ll 1$  за одно вычисление функции отрезок сокращается в

$\sqrt{2} \approx 1,41$  раза, а в методе золотого сечения (1), (2) в  $2/(1+\alpha) = (\sqrt{5}+1)/2 \approx 1,62$  раза, что несколько лучше.

Можно выиграть и чуть больше, если сделать  $\alpha$  зависящим от  $k$ . Именно так поступают в методе Фибоначчи, подробно описанном во многих книгах (например, [0,2, 0,8, 0,18]). Нетрудно видеть, что все описанные выше методы отыскивают минимум не только выпуклой, но и любой *унимодальной* функции (т. е. такой, для которой локальный минимум совпадает с глобальным). Про метод Фибоначчи можно показать, что он скрашает длину отрезка локализации в расчете на одно вычисление функции наиболее быстро, именно, он является оптимальным в минимаксном смысле в классе унимодальных функций. Тем не менее методом Фибоначчи редко пользуются на практике по следующим причинам. Во-первых, он дает совсем небольшой выигрыш по сравнению с методом золотого сечения. В то же время он связан с дополнительными вычислениями для построения новых точек. Во-вторых, в нем требуется заранее выбрать число итераций, которые будут осуществляться. Поскольку естественным критерием окончания процесса одномерной минимизации является близость полученного значения функции к минимальному, а не размер области локализации минимума, то заранее определить нужное число шагов затруднительно. В третьих, он оптимален лишь в минимаксном смысле, т. е. в расчете на «наихудшую» унимодальную функцию. Для конкретных функций более быструю сходимость могут дать другие методы.

Все это лишний раз показывает, сколь осторожно нужно относиться к теоретическим выводам об оптимальности методов (ср. выше п. 3 § 4).

**2. Многомерный случай.** Большинство идей, лежащих в основе поисковых методов минимизации гладких функций (§ 4 гл. 3), не переносятся на негладкий случай. Так, методы последовательной одномерной минимизации типа покоординатного спуска, как мы уже видели (см. рис. 18), могут не сходиться для недифференцируемых функций. Идея локальной линейной или квадратичной аппроксимации минимизируемой функции также не работает. С другой стороны, субградиентный метод (§ 2) и его обобщения (§ 3) нельзя применять, если заменить субградиент на его конечно-разностную аппроксимацию — мы уже отмечали (§ 4), что субградиентный метод, вообще говоря, неустойчив по отношению к детерминированным погрешностям. Наконец, описанный выше одномерный поисковый метод не просто перенести на многомерный случай. Дело в том, что, вычислив функцию в нескольких точках, трудно локализовать область минимума в многомерном случае. В силу указанных выше трудностей в настоящее время известно сравнительно мало теоретически исследованных и обоснованных поисковых методов минимизации негладких функций.

Опишем один из них, идея которого весьма проста и перспективна. Для задачи минимизации выпуклой функции  $f(x)$  на  $\mathbf{R}^n$  он имеет вид

$$\begin{aligned} x^{k+1} &= x^k - \gamma_k s^k, \\ s^k &= \delta_k^{-1} [f(x^k + \alpha_k g^k + \delta_k h^k) - f(x^k + \alpha_k g^k)] h^k, \end{aligned} \quad (3)$$

где  $g^k$ ,  $h^k$  — независимые случайные векторы, равномерно распределенные на кубе  $Q = \{x: |x_i| \leq 1, i = 1, \dots, n\}$ ,  $\alpha_k, \delta_k, \gamma_k$  — некоторые числовые последовательности. Иначе говоря, делается шаг случайного поиска (по направлению  $h^k$ ), но не из точки  $x^k$ , а из «рандомизированной» точки  $x^k + \alpha_k g^k$ . Благодаря введению такой *рандомизации* происходит как бы стягивание исходной функции. Можно показать, что

$$M(S^k | x^k) = c \nabla f(x^k, \alpha_k) + \beta_k, \quad \|\beta_k\| \leq c_1 \delta_k / \alpha_k, \quad (4)$$

где  $f(x, \alpha)$  — слаженная функция,

$$f(x, \alpha) = \frac{1}{(2\alpha)^n} \int f(x + ay) dy, \quad (5)$$

причем  $f(x, \alpha)$  является выпуклой лифференцируемой функцией, градиент которой удовлетворяет условию Липшица с константой  $c \sqrt{n}/\alpha_k$ . Таким образом, (3) может рассматриваться как градиентный метод минимизации слаженной функции (5) при различии помех. Регулируя коэффициент слаживания  $\alpha_k$ , длину пробного шага  $\delta_k$  и рабочего шага  $\gamma_k$ , можно добиться сходимости метода. Именно, если

$$\sum_{k=0}^{\infty} \gamma_k = \infty, \quad \sum_{k=0}^{\infty} \gamma_k^2 < \infty, \quad \gamma_k/\alpha_k \rightarrow 0, \\ \delta_k/\alpha_k \rightarrow 0, \quad \alpha_k \rightarrow 0, \quad |\alpha_k - \alpha_{k+1}|/\gamma_k \rightarrow 0, \quad (6)$$

то метод сходится с вероятностью 1 к множеству точек минимума (если последнее не пусто). Аналогичным образом процедура стягивания путем рандомизации может быть применена для построения других методов.

Разумеется, скорость сходимости метода (3) очень мала. Задача построения эффективных поисковых методов минимизации негладких выпуклых функций в многомерном случае остается открытой.

**Г л а в а 6**  
**ВЫРОЖДЕННОСТЬ, МНОГОЭКСТРЕМАЛЬНОСТЬ,  
 НЕСТАЦИОНАРНОСТЬ**

В практических задачах редко приходится иметь дело с той идеализированной ситуацией, которая рассматривалась в гл. 1 и 3. С двумя возможными осложнениями (наличие помех и отсутствие дифференцируемости) мы уже познакомились в предыдущих главах. Перейдем к анализу других причин, затрудняющих решение задач безусловной оптимизации — вырожденности минимума, многоэкстремальности и нестационарности. Нас будет интересовать как поведение стандартных методов в подобных ситуациях, так и специальные приемы преодоления упомянутых трудностей.

**§ 1. Вырожденный минимум**

В гл. 1 и 3 мы исследовали методы оптимизации, в основном для случая невырожденного минимума (т. е. в предположении, что в точке минимума  $x^* \nabla f(x^*) > 0$ ). Откажемся от этого предположения.

**1. Поведение стандартных методов.** Изучим поведение простейшего градиентного метода безусловной минимизации дифференцируемой функции  $f(x)$ .

(1)

$$x^{k+1} = x^k - \gamma \nabla f(x^k)$$

В ситуации, когда невырожденность точки минимума не предполагается, но зато  $f(x)$  выпуклая. Мы уже знаем (теорема 1 § 4 гл. 1), что при минимальных предположениях  $\nabla f(x^k) \rightarrow 0$  для (1). Оказывается, для выпуклых функций справедлив более сильный результат.

**Теорема 1.** Пусть  $f(x)$  — выпуклая дифференцируемая функция в  $\mathbb{R}^n$ , градиент которой удовлетворяет условию Липшица с константой  $L$ , а множество точек минимума  $X^* = \arg \min_{x \in \mathbb{R}^n} f(x)$  непусто. Тогда метод (1) с  $0 < \gamma < 2/L$  сходится к некоторой точке  $\tilde{x} \in X^*$ ,  $f(\tilde{x}) = f^*$ , причем

(2)

$$f(x^k) - f^* = o(1/k).$$

**Доказательство.** Используем неравенство (лемма 2 § 4 гл. 1)  $(\nabla f(x) - \nabla f(y), x - y) \geq L^{-1} \|\nabla f(x) - \nabla f(y)\|^2$ , справедливо для выпуклых функций, у которых градиент удовлетворяет условию Липшица с константой  $L$ . Отсюда  $(\nabla f(x), x - \tilde{x}) \geq$

$$\sum_{k=0}^{\infty} \|\nabla f(x^k)\|^2 < \infty, \quad (4)$$

т. е.  $\nabla f(x^k) \rightarrow 0$ . Последовательность  $x^k$  ограничена, так как  $\|x^k - \tilde{x}\| \leq \|x^0 - \tilde{x}\|$ , поэтому можно выбрать сходящуюся подпоследовательность  $x^{k_i} \rightarrow \tilde{x}$ . В силу непрерывности  $\nabla f(x)$  при этом  $\nabla f(\tilde{x}) = 0$ , т. е.  $\tilde{x} \in X^*$ . Заменяя в (3)  $\tilde{x}$  на  $\tilde{x}$ , получаем, что

$$(9) \quad \begin{aligned} \text{§ 4 гл. 1 } f(x^{k+1}) &\leq f(x^k) - \alpha \|\nabla f(x^k)\|^2, \quad \alpha = \gamma(1 - L\gamma/2) > 0. \\ \text{в силу выпуклости } f(x): \quad f(x^k) - f^* &\leq (\nabla f(x^k), x^k - \tilde{x}) \leq \|\nabla f(x^k)\| \|x^k - \tilde{x}\|. \end{aligned} \quad \text{Поэтому для } u_k = f(x^k) - f^* \text{ получаем } u_{k+1} \leq u_k - \alpha \|x^k - \tilde{x}\|^2 u_k, \text{ и применение леммы 6 § 2 гл. 2 дает}$$

$$u_k \leq \left( \frac{1}{u_0} + \alpha \sum_{i=0}^{k-1} \|x^i - \tilde{x}\|^{-2} \right)^{-1}$$

$$ku_k \leq \left( \frac{1}{u_0 k} + \alpha/k \sum_{i=0}^{k-1} \|x^i - \tilde{x}\|^{-2} \right)^{-1}$$

Поскольку по доказанному  $\|x^i - \tilde{x}\| \rightarrow 0$  при  $i \rightarrow \infty$ , то  $\|x^i - \tilde{x}\|^{-2} \rightarrow \infty$  и  $k^{-1} \sum_{i=0}^{k-1} \|x^i - \tilde{x}\|^{-2} \rightarrow \infty$  при  $k \rightarrow \infty$ , поэтому

правая часть последнего неравенства стремится к 0 при  $k \rightarrow \infty$ . Это и означает, что  $u_k = o(1/k)$ .  $\blacktriangle$

Отметим, что точка  $\tilde{x} \in X^*$ , к которой сходится  $x^k$ , не может быть указана заранее. Так, она может быть различна для различных  $\gamma$  (при фиксированных  $x^0$ ) и не обязательно совпадает с  $x^*$  — точкой из  $X^*$ , ближайшей к  $x^0$ . Однако  $\tilde{x}$  не может быть слишком далека от  $x^*$ . Действительно, заменяя в (3)  $\tilde{x}$  на  $x^*$ , получаем  $\|x^k - x^*\| \leq \|x^0 - x^*\|$ , т. е.

$$\|\tilde{x} - x^*\| \leq \|x^0 - x^*\| = \rho(x^0, X^*).$$

Из теоремы 1 следует, что градиентный метод сходится (в выпуклом случае) без всяких предположений о невырожденности минимума. При этом по функции гарантируется скорость сходимости порядка  $o(1/k)$ . Однако скорость сходимости по переменным может быть существенно меньше. Пусть, например,  $f(x) = p^{-1} |x|^\rho$  при  $|x| \leq 1$ ,  $f(x) = |x|$  при  $|x| > 1$ ,  $p > 2$ ,  $x \in \mathbb{R}$ . Тогда  $f(x)$  удовлетворяет условиям теоремы 1,  $x^* = 0$  и

из (1) имеем  $|x^{k+1}| = |x^k - \gamma(x^k)^{p-1}|$  при  $|x^0| \leq 1$ . Используя результат упражнения 3 § 2 гл. 2, находим, что при  $0 < \gamma < 2$  будет  $|x^k| = O(k^{-1/(p-2)})$ . Таким образом, выбирая достаточно большое  $p$ , получаем, что для любого  $\alpha > 0$  существует такая функция  $f(x)$ , для которой градиентный метод сходится медленнее, чем  $k^{-\alpha}$ . Отметим, что для этого же примера  $f(x^k) = O(k^{-p/(p-2)})$ , что согласуется с оценкой (2) и демонстрирует невозможность ее существенного улучшения.

Рассмотрим более подробно вопрос о поведении градиентного метода для квадратичной функции:

$$f(x) = (Ax, x)/2 - (b, x), \quad A > 0. \quad (6)$$

Хотя задача минимизации  $f(x)$  невырожденная (так как  $A > 0$ , то точка минимума  $x^*$  существует, единственна, глобально устойчива, а  $f(x)$  сильно выпукла), нас будет интересовать случай плохо обусловленной задачи, который в определенном смысле близок к вырожденному. Пусть  $L$  и  $l$  — наибольшее и наименьшее собственные значения  $A$ ,  $\mu = L/l \gg 1$ . Как мы знаем (теорема 3 § 4 гл. 1), при выборе  $\gamma = 2/(L+l)$  (такой выбор является наилучшим) для градиентного метода (1) справедлива оценка  $\|x^k - x^*\| \leq \|x^0 - x^*\|^{q^k}$ ,  $q = (L+l)/(L+l) = (\mu-1)/(\mu+1)$ , и эта оценка неулучшаема (см. примеры после упомянутой теоремы). Поскольку  $2(f(x^k) - f^*) = (A(x^k - x^*), x^k - x^*) \leq \|A\| \|x^k - x^*\|^2 \leq L \|x^0 - x^*\|^{2q^k}$ , то можно гарантировать сходимость по функции со скоростью геометрической прогрессии со знаменателем  $q_1 = q^2$ . Однако для плохо обусловленных задач  $\mu \gg 1$ , и  $q_1 \approx 1 - 4/\mu$  очень близко к 1. Оказывается, можно получить оценку скорости сходимости по функции, не зависящую от обусловленности.

**Теорема 2.** *Метод (1) для минимизации (6) при  $0 < \gamma < 2/L$  сходится к  $x^*$ , причем для достаточно больших  $k$*

$$f(x^k) - f^* \leq \frac{\|x^0 - x^*\|^2}{2\gamma(2k+1)} \left(1 - \frac{1}{2k+1}\right)^{2k} \leq \frac{\|x^0 - x^*\|^2}{4\gamma\mu k}. \quad (7)$$

Доказательство.

$$\begin{aligned} x^k - x^* &= (I - \gamma A)^k (x^0 - x^*), \\ 2(f(x^k) - f^*) &= (A(I - \gamma A)^{2k} (x^0 - x^*), x^0 - x^*) \leq \\ &\leq \|x^0 - x^*\|^2 \|A(I - \gamma A)^{2k}\| \leq \|x^0 - x^*\|^2 \max_{l \leq \lambda \leq L} |\lambda(1 - \gamma\lambda)^{2k}| \leq \\ &\leq \|x^0 - x^*\|^2 \max_{0 \leq \lambda \leq L} \Phi(\lambda), \end{aligned}$$

где  $\Phi(\lambda) = \lambda(1 - \gamma\lambda)^{2k}$ . Поскольку корнями  $\Phi'(\lambda)$  являются  $\lambda_1 = 1/\gamma$  и  $\lambda_2 = 1/(\gamma(2k+1))$  и  $\Phi(\lambda_1) = 0$ ,  $\Phi(0) = 0$ , то максимум  $\Phi(\lambda)$  на  $[0, L]$  может достигаться либо при  $\lambda = \lambda_2$ , либо

при  $\lambda = L$ . Так как

$$\Phi(\lambda_2) = \frac{1}{\gamma(2k+1)} \left(1 - \frac{1}{2k+1}\right)^{2k} \leq \frac{1}{2\gamma\mu k},$$

$$\Phi(L) = L(1 - \gamma L)^{2k}, \quad \text{а} \quad |1 - \gamma L| < 1,$$

то для достаточно больших  $k$  будет  $\max_{0 \leq \lambda \leq L} \Phi(\lambda) = \Phi(\lambda_2)$ , что

и дает (7).  $\blacktriangleleft$

Таким образом, можно гарантировать оценку типа  $f(x^k) - f^* \leq c/k$ , где константа  $c$  не зависит от обусловленности.

Что же касается скорости сходимости по аргументу, то здесь нельзя получить никакой «равномерной по обусловленности» оценки. Именно, для любого  $0 < \alpha < 1$  и любого  $k$  можно построить такую квадратичную функцию вида (6) и такое начальное приближение  $x^0$ , что будет  $\|x^k - x^*\| > \alpha \|x^0 - x^*\|$  для метода (1) при любом  $\gamma$ . Более того, здесь достаточно брать  $n=2$ , а множество таких  $x^0$  достаточно «обширно».

Перейдем к анализу другого стандартного метода минимизации — метода сопряженных градиентов (§ 2 гл. 3). Вопрос о его поведении для случая вырожденного минимума в общей ситуации не исследован; по-видимому, основное достоинство метода — его быстрая сходимость — здесь теряется. Рассмотрим лишь случай квадратичной функции (6), причем будем предполагать, что размерность задачи велика (так что мы не можем воспользоваться результатом о конечности метода). В (30) § 2 гл. 3 была найдена оценка скорости сходимости метода:

$$\|x^k - x^*\| \leq 2(L/l)^{1/2} \|x^0 - x^*\|^{q^k}, \quad q = (\sqrt{\mu} - 1)/(\sqrt{\mu} + 1).$$

Здесь знаменатель прогрессии  $q$  зависит от обусловленности и близок к 1 для плохо обусловленных задач. Как и выше, получим оценку скорости сходимости по функции, не зависящую от обусловленности.

**Теорема 3.** *В методе сопряженных градиентов для функции (6) справедлива оценка*

$$f(x^k) - f^* \leq \frac{L \|x^0 - x^*\|^2}{2(2k+1)^2}. \quad (8)$$

Доказательство. В соответствии с (27) § 2 гл. 3

$$\begin{aligned} x^k - x^* &= P_k(A)(x^0 - x^*), \\ \text{где } P_k(\lambda) &— полином } k-\text{й степени, обладающий свойством} \\ 2(f(x^k) - f^*) &= (AP_k(A)^2(x^0 - x^*), x^0 - x^*) = \\ &= \min_{R \in \mathcal{R}} (AR(A)^2(x^0 - x^*), x^0 - x^*), \quad (9) \end{aligned}$$

где  $\mathcal{R}$  — множество полиномов  $R(\lambda)$ ,  $k$ -й степени, удовлетворяющих условию  $R(0) = 1$ . Возьмем

$$R^*(\lambda) = \frac{T_{2k+1}(\sqrt{\lambda}/\sqrt{L})}{(2k+1)(\sqrt{\lambda}/\sqrt{L})}, \quad (10)$$

где  $T_k(x) = \cos(k \arccos x)$  — полином Чебышева. Поскольку  $T_{2k+1}(x)$  содержит только нечетные степени  $x$ , то  $R_0(x) = T_{2k+1}(\sqrt{x})/\sqrt{x}$  является полиномом  $k$ -й степени от  $x$ ,  $R_0(0) = T_{2k+1}(0) = 2k+1$ . В соответствии с этим  $R^*(\lambda) \in \mathcal{R}$ . Итак,

$$\begin{aligned} 2(f(x^k) - f^*) &\leq (AR^*(A)^2(x^0 - x^*), x^0 - x^*) \leq \\ &\leq \|x^0 - x^*\|^2 \max_{0 \leq \lambda \leq L} |\lambda R^*(\lambda)| = \end{aligned}$$

$$= \frac{L \|x^0 - x^*\|^2}{(2k+1)^2} \max_{0 \leq \lambda \leq L} \left| T_{2k+1}\left(\frac{\sqrt{\lambda}}{\sqrt{L}}\right) \right| = \frac{L \|x^0 - x^*\|^2}{(2k+1)^2},$$

поскольку  $\max_{0 \leq x \leq 1} |T_k(x)| = 1$ . ▲

Мы видим, что независимо от обусловленности задачи метод сопряженных градиентов гарантирует достаточно высокую скорость сходимости по функции типа  $O(k^{-2})$  вместо  $O(k^{-1})$ , как в градиентном методе. Оценку (8) усилить нельзя. Так, для всякого  $k$  можно построить такую квадратичную функцию в пространстве  $R^n$ ,  $n = k+1$ , и указать такое  $x^0$ , что

$$f(x^k) - f^* = L \|x^0 - x^*\|^2/(2(2k+1)^2).$$

Более того, можно показать, что любой метод минимизации квадратичных функций, использующий лишь информацию о градиентах, не может дать скорости сходимости более высокой, чем  $(8)$ , равномерно по размерности  $n$  и по всему классу квадратичных функций  $f(x)$ .

Что же касается скорости сходимости по аргументу метода сопряженных градиентов, то здесь нельзя получить никаких оценок, не зависящих от обусловленности и размерности.

Любопытно сопоставить с полученными результатами те оценки скорости сходимости по функции, которые мы имели ранее для негладкого случая. Для субградиентного метода в формуле (7) § 3 гл. 5 справедлив результат о сходимости (теорема 2 § 3 гл. 5), аналогичной теореме 1. Скорость сходимости, однако, более медленна, чем в гладком случае — в соответствии с (8) § 3 гл. 5  $f(x^k) - f^*$  убывает как  $O(k^{-1/2})$ . Для метода эллипсоидов (12) § 4 гл. 5 была доказана линейная сходимость по функции (теорема 4 § 4 гл. 5). Знаменатель прогрессии при этом зависел от размерности, но не зависел от каких-либо

характеристик функции типа обусловленности. Поскольку для гладкого вырожденного случая мы нигде не имели линейной скорости сходимости, то можно ожидать, что в задачах небольшой размерности целесообразно применять метод эллипсоидов и при минимизации гладких плохо обусловленных функций.

В заключение этого пункта — несколько слов о поведении метода Ньютона в вырожденном случае. Прежде всего, этот метод не всегда будет корректно определен, так как матрица  $\nabla^2 f(x^k)$  может оказаться вырожденной в сколь угодно малой окрестности  $x^*$ . Поэтому метод неприменим для решения вырожденных задач. Существует более узкий класс задач, в которых эта трудность отсутствует. Именно, пусть  $\nabla^2 f(x) > 0$  для всех точек  $x \neq x^*$  из окрестности  $x^*$ , а в самой точке  $x^*$  матрица  $\nabla^2 f(x^*) \geq 0$  не имеет обратной. Тогда при некоторых дополнительных предположениях метод Ньютона будет сходиться. Однако скорость сходимости будет существенно ниже, чем в невырожденном случае. Пусть, например,  $f(x) = |x|^p$ ,  $p > 2$ ,  $x \in \mathbb{R}^1$ . Тогда  $f'(x) = p|x|^{p-1} \operatorname{sign} x$ ,  $f''(x) = p(p-1)|x|^{p-2}$ ,  $f''(x) > 0$  при  $x \neq 0$  и  $f''(x^*) = f''(0) = 0$ . Метод Ньютона при  $x_0 > 0$  принимает вид  $x_{k+1} = x_k - (p-1)^{-1}x_k = qx_k$ ,  $q = (\rho-2)/(\rho-1) < 1$ . Отсюда  $x_k = q^k x_0$ , т. е. в данном случае метод Ньютона сходится со скоростью геометрической прогрессии, знаменатель которой близок к 1 при больших  $\rho$ . Конечно, это гораздо хуже, чем квадратичная сходимость, которая имела место в невырожденном случае. В других примерах (см. упр. 4) скорость сходимости может быть еще меньше.

Подводя итог, можно сказать, что в основном стандартные методы минимизации остаются сходящимися при поиске вырожденного минимума гладкой выпуклой функции. Однако скорость сходимости падает, иногда существенно.

#### Упражнения.

1. Для фиксированного числа шагов  $k$  укажите наилучший способ выбора параметра  $\gamma$  в методе (1) для минимизации (6), исходя из оценок, полученных при доказательстве теоремы 2. Указание. Выберите  $\gamma$  так, чтобы минимизировать  $\max_{0 \leq \lambda \leq L} \lambda(1-\gamma\lambda)^{2k}$  при известных  $k$ ,  $L$ .

2. Рассмотрите градиентный метод вида  $x^{k+1} = x^k - \gamma_k \nabla f(x^k)$  для минимизации (6) и выберите (при фиксированном числе шагов  $k$  и известной константе  $\frac{L}{\gamma_k}$ )  $\gamma_k$ ,  $0 \leq \gamma_k \leq k-1$ , чтобы выполнить те же оценки для негладкого случая. Для субградиентного метода в формуле (7) § 3 гл. 5 справедлив результат о сходимости (теорема 2 § 3 гл. 5), аналогичной теореме 1. Скорость сходимости, однако, более медленна, чем в гладком случае — в соответствии с (8) § 3 гл. 5  $f(x^k) - f^*$  убывает как  $O(k^{-1/2})$ . Для метода эллипсоидов (12) § 4 гл. 5 была доказана линейная сходимость по функции (теорема 4 § 4 гл. 5). Знаменатель прогрессии при этом зависел от размерности, но не зависел от каких-либо

3. Рассмотрите случай квадратичной  $f(x) \in A \geq 0$  и непустым множеством точек минимума. Покажите, что для него справедливы все результаты о сходимости и скорости сходимости, полученные в этом параграфе для  $A > 0$ .

4. Исследуйте скорость сходимости метода Ньютона для функции  $f(x) = \exp(-x^2)$ ,  $x \in \mathbb{R}$ , в окрестности точки минимума  $x^* = 0$ .

## 2. Специальные методы решения вырожденных задач.

1°. Метод регуляризации. Пусть задача минимизации выпуклой гладкой функции  $f(x)$  — «плохая», например имеет вырожденный минимум. Тогда можно немножко изменить задачу, добавив к  $f(x)$  «хорошую» функцию  $\varepsilon g(x)$  с малым «весом». Найдя точку минимума «улучшенной» функции  $f(x) + \varepsilon g(x)$ , можно устремить параметр  $\varepsilon$  к 0. Естественно на-деляясь, что последовательность найденных при этом точек минимума будет сходиться к решению исходной задачи. В этом заключается идея метода регуляризации. Величина  $\varepsilon$  называется параметром регуляризации, функция  $g(x)$  — регуляризую-щей функцией.

Сначала рассмотрим метод регуляризации в «идеализированном» варианте, когда минимум вспомогательной задачи отыскивается точно.

Теорема 4. Пусть  $f(x)$  — выпуклая непрерывная функция в  $\mathbb{R}^n$ , имеющая непустое множество точек минимума  $X^*$ , а  $g(x)$  — сильно выпуклая непрерывная функция. Пусть

$$x_\varepsilon = \arg\min_{x \in \mathbb{R}^n} \Phi_\varepsilon(x), \quad \Phi_\varepsilon(x) = f(x) + \varepsilon g(x), \quad \varepsilon > 0. \quad (11)$$

Тогда  $x_\varepsilon \rightarrow x^*$  при  $\varepsilon \rightarrow +0$ , где  $x^*$  — та из точек минимума  $f(x)$ , для которой  $g(x)$  минимальна, т. е.  $x^* = \arg\min_{x \in X^*} g(x)$ .

Доказательство. Функция  $f + \varepsilon g$  сильно выпуклая, поэтому точка  $x_\varepsilon$  существует и единственна. Далее, из определения  $x_\varepsilon$  для произвольной точки  $\hat{x} \in X^*$  получаем  $f(x_\varepsilon) + \varepsilon g(x_\varepsilon) \leq f(\hat{x}) + \varepsilon g(\hat{x})$ ,  $f(\hat{x}) \leq f(x_\varepsilon)$ , т. е.  $g(x_\varepsilon) \leq g(\hat{x})$ , а потому и  $g(x_\varepsilon) \leq g(x^*)$ . Так как  $g(x)$  сильно выпукла, то множество  $\{x: g(x) \leq \alpha\}$  ограничено, т. е. совокупность  $x_\varepsilon$  ограничена. Выберем подпоследовательность  $x_{\varepsilon_i}$ , сходящуюся к некоторой точке  $\tilde{x}$ . Так как  $f(x)$  и  $g(x)$  непрерывны, то  $\lim_{i \rightarrow \infty} g(x_{\varepsilon_i}) = g(\tilde{x})$ , т. е.  $g(\tilde{x}) \leq g(x^*)$ , а переходя к пределу в неравенстве  $f(x_{\varepsilon_i}) + \varepsilon_i g(x_{\varepsilon_i}) \leq f(x^*) + \varepsilon g(x^*)$ , получаем  $f(\tilde{x}) \leq f(x^*)$ . Таким образом,  $\tilde{x} \in X^*$ , и из неравенства  $g(\tilde{x}) \leq g(x^*)$  и определения  $x^*$  следует, что  $\tilde{x} = x^*$ . Итак,  $x_{\varepsilon_i} \rightarrow x^*$ . Но  $\|x_{\varepsilon_i} - x^*\|^2 \leq \varepsilon(g(x^*) - g(x_{\varepsilon_i}))$ , т. е. вся последовательность  $x_\varepsilon$  сходится к  $x^*$ . ▲

Конечно, воспользоваться методом регуляризации в описанном выше виде, как правило, не удается из-за невозможности точного решения вспомогательной задачи. Один из немногих случаев, когда такое решение в принципе можно найти, связан с квадратичными функциями. Пусть  $f(x) = (Ax, x)/2 - (b, x)$ , где  $A \geq 0$ , и пусть  $f(x)$  достигает минимума на  $\mathbb{R}^n$  на непустом множестве  $X^*$ . Пусть  $g(x) = (B(x - a), x - a)/2$ , где  $B > 0$ .

Тогда в методе регуляризации на каждом шаге минимизируется квадратичная функция, и поэтому

$$x_\varepsilon = (A + \varepsilon B)^{-1}(b + \varepsilon Ba). \quad (13)$$

В силу (12) матрица  $A + \varepsilon B$  имеет обратную при любом  $\varepsilon > 0$ . Из теоремы 4 следует, что  $x_\varepsilon \rightarrow x^* \in X^*$ , где  $x^* = \arg\min_{x \in X^*} g(x)$ . В частности, когда  $B = I$ ,  $a = 0$  (т. е. когда регуляризатор имеет вид  $g(x) = \|x\|^2/2$ ), то  $x^*$  — точка минимума  $f(x)$  с наименьшей нормой (она называется *нормальным решением* задачи). В этом случае

$$x_\varepsilon = (A + \varepsilon I)^{-1}b. \quad (14)$$

Метод регуляризации для квадратичной задачи тесно связан с понятием так называемой псевдообратной матрицы. Пусть  $C$  — произвольная матрица размерности  $m \times n$  (не обязательно квадратная). Тогда функция

$$f(x) = \|Cx - d\|^2, \quad x \in \mathbb{R}^n, \quad (15)$$

достигает минимума на  $\mathbb{R}^n$  (см. упр. 2 § 3 гл. 1). Точка минимума  $f(x)$  с наименьшей нормой (нормальное решение, обозначим его  $x^*$ ) единственна. Можно показать, что  $x^*$  линейно зависит от  $d$ :

$$x^* = C^+d, \quad (16)$$

где  $C^+$  — некоторая матрица размерности  $n \times m$ , называемая псевдообратной к  $C$ . Из теоремы 4 и равенства (14) следует, что

$$C^+ = \lim_{\varepsilon \rightarrow +0} (C^T C + \varepsilon I)^{-1} C^T. \quad (17)$$

Другие свойства операции псевдообращения приведены в упр. 6.

Вернемся к методу регуляризации. Ясно, что чем меньше  $\varepsilon$ , тем ближе  $x_\varepsilon$  к решению, так что, казалось бы, выгодно брать очень малые  $\varepsilon$ . Однако мы увидим в дальнейшем, что из-за влияния погрешностей в вычислении функции и градиента, а также из-за ошибок округления при решении вспомогательной задачи выбирать слишком малые  $\varepsilon$  нельзя. Поэтому возникает вопрос о точности решения, даваемого методом регуляризации, при конечных  $\varepsilon$ . Приведем примеры, показывающие, что  $|x_\varepsilon - x^*|$  может быть велика даже для малых  $\varepsilon$ .

Пусть  $f(x) = p^{-1}x^p$ ,  $x \in \mathbb{R}^1$ ,  $p > 2$ ,  $g(x) = (x - 1)^2/2$ . Тогда  $x^* = 0$ , и нетрудно оценить, что  $|x_\varepsilon - x^*| \approx |x_\varepsilon| \approx \varepsilon^{1/(p-1)}$ . Поэтому, если  $p$  велик, то  $|x_\varepsilon - x^*|$  сравнительно велико даже для маленьких  $\varepsilon$ . Так, при  $p = 7$ ,  $\varepsilon = 10^{-6}$  получаем  $|x_\varepsilon - x^*| \approx 10^{-1}$ ,

**2°. Прокс-метод.** Метод регуляризации для регуляризующей функции  $g(x) = \|x - a\|^2/2$  записывается в форме

$$x^k = x_{\varepsilon_k} = \operatorname{argmin}_{x \in \mathbb{R}^n} (f(x) + (\varepsilon_k/2)\|x - a\|^2), \quad \varepsilon_k \rightarrow 0.$$

Можно попытаться поступить иначе — на каждом шаге менять не параметр регуляризации  $\varepsilon_k$ , а точку  $a$ , заменяя ее на  $x^k$ . Таким образом приходим к методу

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left( f(x) + \frac{1}{2} \varepsilon \|x - x^k\|^2 \right), \quad \varepsilon > 0. \quad (18)$$

Этот метод называется *проксимационным* (или прокс-методом), поскольку он тесно связан с так называемым проксимационным отображением. Пусть  $f(x)$  — выпуклая функция на  $\mathbb{R}^n$ ,  $\varepsilon > 0$  — некоторый параметр. Тогда оператор

$$\operatorname{Prox}_x a = \operatorname{argmin}_{x \in \mathbb{R}^n} \left( f(x) + \frac{1}{2} \varepsilon \|x - a\|^2 \right)$$

называется *проксимационным*. Его свойства и явный вид для ряда примеров приведены в упражнениях 7 и 8.

Во введенных обозначениях метод записывается так:

(19)

*Теорема 5.* Пусть  $f(x)$  — выпуклая функция на  $\mathbb{R}^n$ , имеющая непустое множество точек минимума  $X^*$ ,  $\varepsilon > 0$ . Тогда метод (13) сходится к некоторой точке  $x^* \in X^*$ .

*Доказательство.* В соответствии с упр. 7 функция  $\Psi(a) = \min_x [f(x) + 1/2 \varepsilon \|x - a\|^2]$  выпукла, дифференцируема,  $\nabla \Psi(a) = \varepsilon(a - \operatorname{Prox}_a a)$  удовлетворяет условию Липшица с константой  $\varepsilon$  и  $X^* = \operatorname{Argmin}_a \Psi(a) \neq \emptyset$ . Применим для минимизации  $\Psi(a)$  градиентный метод с  $\gamma = 1/\varepsilon$ :

$$a^{k+1} = a^k - \varepsilon^{-1} \nabla \Psi(a^k) = a^k - \varepsilon^{-1} \varepsilon(a^k - \operatorname{Prox}_a a^k) = \operatorname{Prox}_a a^k.$$

Иначе говоря, прокс-метод (19) может рассматриваться как градиентный метод минимизации  $\Psi(a)$ . Применив теорему 1 (все ее условия выполнены), получаем требуемый результат.  $\blacktriangle$

Преимуществом прокс-метода перед методом регуляризации является то, что обусловленность вспомогательных задач минимизации в нем не ухудшается (параметр  $\varepsilon$  остается постоянным). Однако прокс-метод (так же как и градиентный метод) не приводит, вообще говоря, к нормальному решению.

Для квадратичной функции вида (6) прокс-метод может быть записан в явной форме:

$$x^{k+1} = (A + \varepsilon I)^{-1}(b + \varepsilon x^k).$$

**3°. Итеративная регуляризация.** Во всех описанных выше методах предполагалось, что на каждом шаге решается (точно или приближенно) некоторая вспомогательная задача безусловной минимизации. При этом мы никак не фиксировали метод ее решения. Можно поступить иначе — задавая некоторым методом безусловной минимизации и делать несколько итераций этого метода для очередной вспомогательной задачи (число итераций может быть либо заранее определено, либо регулироваться в процессе вычислений). В простейшем варианте методов такого типа делается один шаг градиентного спуска для минимизации регуляризованной функции, после чего изменяется параметр регуляризации. Таким образом, приходим к методу *итеративной регуляризации*:

$$x^{k+1} = x^k - \gamma_k (\nabla f(x^k) + \varepsilon_k \nabla g(x^k)), \quad (21)$$

где  $g(x)$  — регуляризующая функция,  $\varepsilon_k$  — меняющийся на каждой итерации параметр регуляризации.

*Теорема 6.* Пусть  $f(x)$ ,  $g(x)$  — однажды дифференцируемые выпуклые функции на  $\mathbb{R}^n$ , причем  $\|\nabla f(x)\| \leq L$ ,  $\|\nabla g(x)\| \leq L$ ,  $L > 0$ , для всех  $x$ ,  $X^* = \operatorname{Argmin}_x f(x) \neq \emptyset$  и

$$0 \leq \frac{\varepsilon_{k-1} - \varepsilon_k}{\varepsilon_k^2} \rightarrow 0, \quad 0 \leq \varepsilon_k \rightarrow 0, \quad \sum_{k=0}^{\infty} \varepsilon_k = \infty, \quad (22)$$

$$\gamma_k = \gamma, \quad 0 < \gamma < \frac{2}{(1 + \varepsilon_0)L}. \quad (23)$$

Тогда в методе (21)  $x^k \rightarrow x^*$ , где  $x^* \in X^*$ ,  $x^* = \operatorname{argmin}_{x \in X^*} g(x)$ .

*Доказательство.* Пусть  $y^k = \operatorname{argmin}_{x \in X^*} \Phi_k(x)$ ,  $\Phi_k(x) = f(x) + \varepsilon_k g(x)$ , при сделанных предположениях  $y^k$  существует однозначно определена и  $y^k \rightarrow x^*$  (см. теорему 4). Функция  $\Phi_k(x)$  сильно выпукла с константой  $b_{\varepsilon_k}$ , поэтому (см. (35) § 1 гл. 1)  $\Phi_k(y^{k-1}) \geq \Phi_k(y^k) + ((\varepsilon_k/2)\|y^k - y^{k-1}\|^2)$ . Аналогично, из сильной выпуклости  $\Phi_{k-1}(x)$  получаем  $\Phi_{k-1}(y^k) \geq \Phi_{k-1}(y^{k-1}) + ((\varepsilon_{k-1}/2)\|y^k - y^{k-1}\|^2)$ . Складывая эти неравенства, имеем  $(\varepsilon_k - \varepsilon_{k-1})(g(y^k) - g(y^{k-1})) + l(\varepsilon_{k-1} + \varepsilon_k)\|y^k - y^{k-1}\|^2/2 \leq 0$ . Поскольку  $\{y^k\}$  ограничено, то найдется такая константа  $M$ , что  $\|g(y^k) - g(y^{k-1})\| \leq M\|y^k - y^{k-1}\|$ . Поэтому

$$\|y^k - y^{k-1}\| \leq \frac{2M(\varepsilon_{k-1} - \varepsilon_k)}{l(\varepsilon_{k-1} + \varepsilon_k)} \leq N \frac{\varepsilon_{k-1} - \varepsilon_k}{\varepsilon_k}, \quad N = \frac{M}{l}. \quad (24)$$

Оценим теперь в методе (21) расстояние от  $x^{k+1}$  до  $y^k$ :

$$\|x^{k+1} - y^k\| = \|x^k - y^k - \gamma \nabla \Phi_k(x^k)\| = \|x^k - y^k - \gamma A(x^k - y^k)\|.$$

Здесь в соответствии с (13), § 1 гл. 1 и условием  $\nabla \Phi_k(y^k) = 0$

$$A = \int_0^1 \nabla^2 \Phi_k(y^k + \tau(x^k - y^k)) d\tau.$$

В силу сделанных предположений

$$Ie_k I \leq \nabla^2 \Phi_k(x) \leq L(1 + \varepsilon_k) I \leq L(1 + \varepsilon_0) I,$$

поэтому  $Ie_k I \leq A \leq L(1 + \varepsilon_0) I$  и

$$\|x^{k+1} - y^k\| \leq \|I - \gamma A\| \|x^k - y^k\| \leq \\ \leq \max_{Ie_k \leq \lambda \leq L(1 + \varepsilon_0)} \|I - \gamma \lambda\| \|x^k - y^k\| = (1 - \gamma \varepsilon_k) \|x^k - y^k\| \quad (25)$$

для достаточно больших  $k$  в силу условия  $\varepsilon_k \rightarrow 0$ . Используя (24) и (25), получаем

$$\|x^{k+1} - y^k\| \leq (1 - \gamma \varepsilon_k) \|x^k - y^k\| \leq (1 - \gamma \varepsilon_k) \|x^k - y^{k-1}\| + \\ + (1 - \gamma \varepsilon_k) \|y^k - y^{k-1}\| \leq (1 - \gamma \varepsilon_k) \|x^k - y^{k-1}\| + \mu_k,$$

$$\mu_k = (1 - \gamma \varepsilon_k) N(e_{k-1} - e_k) \varepsilon_k^{-1}.$$

Применение леммы 3 § 2 гл. 2 для  $u_* = \|x^k - y^{k-1}\|$  с учетом условий (22) дает  $u_k \rightarrow 0$ . Но  $\|x^k - x^*\| \leq \|x^k - y^{k-1}\| + \|y^{k-1} - x^*\| \rightarrow 0$ , так как  $\|x^k - y^{k-1}\| \rightarrow 0$  по доказанному выше, а  $\|y^k - x^*\| \rightarrow 0$  в силу теоремы 4. ▲

Что касается скорости сходимости, то в силу условия  $\sum_{k=0}^{\infty} \varepsilon_k = \infty$  параметр  $\varepsilon_k$  нельзя устремлять к 0 слишком быстро. С другой стороны, метод сходится не быстрее, чем метод регуляризации, а последний, как мы видели выше, может сходиться медленно.

Упражнения.

5. Пусть  $f(x)$  — выпуклая функция в  $\mathbf{R}^n$ ,  $X^* = \operatorname{Argmin}_{x \in \mathbf{R}^n} f(x) \neq \emptyset$ , функция  $g(x)$  строго выпукла и множество  $\{x; g(x) \leq a\}$  ограничено и непусто для некоторого  $a$ . Докажите (по той же схеме, что и теорему 4) сходимость метода регуляризации в этом случае.

6. Используя определение  $C^+$  и формулу (7), докажите следующие свойства псевдобротных матриц:

- а) если  $m = n$  и  $C^{-1}$  существует, то  $C^+ = C^{-1}$ ;
- б)  $AA^+A = A$ ,  $A^+AA^+ = A^+$ ;
- в)  $(A^+)^+ = A$ ;
- г)  $(A^T)^+ = (A^+)^T$ .

7. Докажите следующие свойства оператора Prox:

- а) он однозначно определен;
- б) он является нерастягивающим, т. е.  $\|\operatorname{Prox} a - \operatorname{Prox} b\| \leq \|a - b\|$ ;

в) функция  $\Psi(a) = \min_x (f(x) + (\varepsilon/2) \|x - a\|^2)$  выпукла, дифференцируема, ее градиент удовлетворяет условию Липшица с константой  $\varepsilon$  и равен  $\nabla \Psi(a) = \varepsilon(a - \operatorname{Prox} a)$ ;

г) если  $X^* = \operatorname{Argmin}_x f(x) \neq \emptyset$ , то  $X^* = \operatorname{Argmin}_a \Psi(a)$ .

8. Вычислите  $\operatorname{Prox} a$  и  $\Psi(a)$  (упр. 7) для следующих примеров:

- а)  $f(x) = (Ax, x)/2 - (b, x)$ ,  $A \geq 0$ ;
- б)  $f(x) \equiv 0$ ;
- в)  $f(x) = \|x\|$ .

Ответы. а)  $\operatorname{Prox} a = (A + \varepsilon I)^{-1}(b + \varepsilon a)$ ,  $\Psi(a) = (1/2)[\varepsilon \|a\|^2 - ((A + \varepsilon I)^{-1}(b + \varepsilon a), (b + \varepsilon a))]$ ; б)  $\operatorname{Prox} a = a$ ,  $\Psi(a) = a$ ; в)  $\operatorname{Prox} a = [1 - 2/(\varepsilon \|a\|)]a$ ,  $\Psi(a) = \|a\|^2/2\varepsilon$ ,  $\Psi(a) = \|a\|$  при  $\|a\| > 2/\varepsilon$ .

3. Методы при наличии помех. Анализ методов отыскания вырожденного минимума производился выше в идеализированной ситуации, когда значения градиента минимизируемой функции известны точно (в методах: градиентном, сопряженных градиентов, итеративной регуляризации) или когда вспомогательная задача минимизации на каждой итерации решается точно (в методах регуляризации и проксимационном). Рассмотрим влияние помех на поведение упомянутых методов, ограниченных наибольшими характерными случаями.

1°. Градиентный метод. Пусть имеются абсолютные детерминированные погрешности в определении градиента, т. е. в точке  $x^k$  доступен вектор  $x^k = \nabla f(x^k) + r^k$ ,  $\|r^k\| \leq \varepsilon$ . В этой ситуации градиентный метод (1) принимает вид

$$s^{k+1} = x^k - \gamma s^k. \quad (26)$$

Как мы знаем (из теоремы 1 § 2 гл. 4), для невырожденного минимума можно гарантировать сходимость в некоторую область вокруг точки минимума  $x^*$ . Радиус этой области (см. упр. 2 § 2 гл. 4) зависит от константы сильной выпуклости  $l$  и стремится к бесконечности при  $l \rightarrow 0$ . Поэтому из этих результатов нельзя сделать выводы о поведении метода в вырожденном случае (кроме предположения о неработоспособности метода). Действительно, при попадании в область малых значений градиента метод (27) начинает вести себя бессмысленным образом — направление движения становится практически произвольным. Поэтому метод (27) следует видоизменить — нужно прекращать итерации, как только величина  $\|s^k\|$  станет достаточно малой. В такой форме метод оказывается в определенном смысле работоспособным.

Теорема 7. Пусть  $f(x)$  — выпуклая дифференцируемая функция в  $\mathbf{R}^n$ , градиент которой удовлетворяет условию Липшица

с константой  $L$ , а  $X^* = \operatorname{Argmin}_{x \in \mathbb{R}^n} f(x) \neq \emptyset$ . Пусть известны величины  $L$ ,  $\varepsilon$  (см. (26)) и  $\rho \geq \|x^0 - x^*\|$ , где  $x^* = P_{X^*}(x^0)$  — точка минимума, ближайшая к  $x^0$ . Пусть итерации (27) с  $0 < \gamma < 2/L$  продолжаются до выполнения условия

$$\|s^k\| \leq \varepsilon + 2 \sqrt{\frac{\varepsilon L \rho}{2 - L \gamma}} \quad (28)$$

и  $x_\varepsilon$  — та точка  $x^k$ , в которой это условие впервые выполняется. Тогда процесс закончится не более чем за  $\rho/(\gamma\varepsilon) + 1$  итераций, при этом  $\|\nabla f(x_\varepsilon)\| \leq \left( \varepsilon + \sqrt{\frac{\varepsilon L \rho}{2 - L \gamma}} \right) \rho$ .

Доказательство. Из (26) и (27)

$$\|x^{k+1} - x^*\| = \|x^k - x^* - \gamma \nabla f(x^k) - \gamma r^k\| \leq \|x^k - x^* - \gamma \nabla f(x^k)\| + \gamma \varepsilon.$$

Из неравенства (3) (с заменой  $\hat{x}$  на  $x^*$ )  $\|x^k - x^* - \gamma \nabla f(x^k)\|^2 \leq \|x^k - x^*\|^2 - \gamma(2L - \gamma)\|\nabla f(x^k)\|^2$ .

Поскольку для произвольных  $a \geq b > 0$  справедливо неравенство  $\sqrt{a^2 - b^2} \leq a - b^2/(2a)$ , то  $\|x^k - x^* - \gamma \nabla f(x^k)\| \leq \|x^k - x^*\| - \gamma(2 - L\rho)\|\nabla f(x^k)\|^2(2L\|x^k - x^*\|)^{-1}$ . Таким образом,

$$\|x^{k+1} - x^*\| \leq \|x^k - x^*\| - \frac{\gamma(2 - L\rho)\|\nabla f(x^k)\|^2}{2L\|x^k - x^*\|} + \gamma \varepsilon. \quad (29)$$

Предположим, что  $x^k$  не является точкой остановки. Тогда

$$\varepsilon + 2 \sqrt{\frac{\varepsilon L \rho}{2 - L \gamma}} \leq \|s^k\| \leq \|\nabla f(x^k)\| + \varepsilon. \quad (30)$$

Отсюда  $\|\nabla f(x^k)\|^2 \geq (4\varepsilon L \rho)/(2 - L \gamma)$ . Подставляя это в (29), получаем

$$\|x^{k+1} - x^*\| \leq \|x^k - x^*\| - \gamma \varepsilon \left( \frac{2\rho}{\|x^k - x^*\|} - 1 \right). \quad (31)$$

Поскольку  $\|x^0 - x^*\| \leq \rho$ , то отсюда  $\|x^1 - x^*\| \leq \rho - \gamma \varepsilon$  и вообще  $\|x^k - x^*\| \leq \rho - k \gamma \varepsilon$  для всех  $k$  до остановки процесса. Поэтому число итераций до остановки процесса не превосходит  $\rho/(\gamma \varepsilon) + 1$ . Поскольку в точке остановки

$$\|\nabla f(x^k)\| - \varepsilon \leq \|s^k\| \leq \varepsilon + 2 \sqrt{\varepsilon L \rho / (2 - L \gamma)},$$

то  $\|\nabla f(x_\varepsilon)\| \leq (\varepsilon + \sqrt{\varepsilon L \rho / (2 - L \gamma)})$ . ▲

Обсудим полученный результат. Для выбранной модификации градиентного метода гарантируется, во-первых, что будет получена точка с достаточно малой нормой градиента  $\|\nabla f(x_\varepsilon)\| \leq \Phi(\varepsilon)$ , где  $\Phi(\varepsilon) = O(\sqrt{\varepsilon}) \rightarrow 0$  при  $\varepsilon \rightarrow 0$ , во-вторых что эта точка не слишком далека от точки минимума, ближайшей к начальному приближению. В силу неравенства  $f(x_\varepsilon) -$

$-f(x^*) \leq \|\nabla f(x_\varepsilon)\| \|x_\varepsilon - x^*\|$  можно гарантировать, что в точке  $x_\varepsilon$  значение функции также близко к минимальному:

$$f(x_\varepsilon) - f(x^*) = O(\sqrt{\varepsilon}). \quad (31)$$

В этом смысле можно считать, что точка  $x_\varepsilon$  дает приближенное решение задачи минимизации. Разумеется, дать какую-либо явную оценку близости  $x_\varepsilon$  к множеству  $X^*$  нельзя.

Для задач со случайными абсолютными помехами можно доказать результат о сходимости почти наверное градиентного метода, в котором шаг  $\gamma_k$  стремится к 0 (см. упр. 9).

Если интенсивность помех зависит от номера итерации:

$$\|r^k\| \leq \varepsilon_k, \text{ при } \sum_{k=0}^{\infty} \varepsilon_k < \infty \text{ градиентный метод сходится в обычном смысле.}$$

2°. Метод регуляризации. Вследствие неизбежных погрешностей при вычислении  $f(x)$ , а также из-за невозможности отыскания точного минимума неквадратичной функции, вспомогательная задача безусловной минимизации (11) в Методе регуляризации может быть решена лишь приближенно, с точностью до некоторой величины  $\delta$ . Пусть

$$\Phi_\varepsilon(x_\varepsilon^\delta) \leq \Phi_\varepsilon^* + \delta,$$

где  $\Phi_\varepsilon(x) = f(x) + \varepsilon g(x)$ ,  $\Phi_\varepsilon^* = \min_{x \in \mathbb{R}^n} \Phi_\varepsilon(x)$ .

Теорема 8. Пусть выполнены условия теоремы 4. Тогда при

$$\varepsilon \rightarrow 0, \quad \delta/\varepsilon \rightarrow 0 \quad (33)$$

будет  $x_\varepsilon^\delta \rightarrow x^*$ .

Доказательство проводится точно так же, как и в теореме 4. ▲

Оценку близости  $\|x_\varepsilon^\delta - x^*\|$  в явной форме для произвольного функционала  $f(x)$  дать нельзя (см. примеры, относящиеся к теореме 4).

3°. Другие методы. Аналогичному исследованию могут быть подвергнуты другие методы, описанные ранее, — в частности, прокс-метод и метод итеративной регуляризации. Не будем на этом останавливаться подробнее, так как и техника исследования, и получающиеся результаты подобны теоремам 7 и 8.

#### Упражнение.

9. Пусть  $f(x)$  — выпуклая дифференцируемая функция в  $\mathbb{R}^n$ ,  $\nabla f(x)$  удовлетворяет условию Липшица,  $X^* = \operatorname{Argmin}_{x \in \mathbb{R}^n} f(x) \neq \emptyset$ . Пусть  $s^k = \nabla f(x^k) + \xi^k$ , где случайные помехи  $\xi^k$  независимы и  $M\xi^k = 0$ ,  $M\|\xi^k\|^2 \leq \sigma^2$ . Рассматривается градиентный метод  $x^{k+1} = x^k - \gamma_k s^k$  при условиях  $\sum_{k=0}^{\infty} \gamma_k = \infty$ ,  $\sum_{k=0}^{\infty} \gamma_k^2 <$

<  $\infty$ . Докажите, используя метод доказательства теоремы 1 этого параграфа и теоремы I § 2 гл. 2, что  $x \xrightarrow{k \text{ п. н.}} x^* \in X^*$ , где точка  $x$  может быть разной для различных реализаций процесса.

**4. Заключительное обсуждение.** Теперь настало время ответить на основной вопрос — можно ли практически решать задачи оптимизации с вырожденным минимумом? Ответ на этот вопрос не столь прост, как кажется, и он заставляет еще раз осмыслять соотношение теоретических результатов о сходимости и практики вычислений, которое мы уже обсуждали в § 6 гл. 1. Нужно сказать, что сама постановка такого вопроса в математических работах нередко считается неуместной и ее избегают, ограничиваясь результатами о сходимости типа вышеупомянутых. Однако отнюдь не очевидно, какие практические выводы можно сделать, например, из теоремы 8 — предлагаем читателю самому обдумать эту проблему в качестве упражнения.

Прежде всего, нужно ясно понимать, что именно требуется от приближенного решения той или иной оптимизационной задачи. Ответ здесь зависит от того, как будет использоваться дальнейшее найденное решение. В некоторых случаях для нас основной интерес представляет сама точка минимума (будем говорить тогда об *аргументных задачах* минимизации). Например, оценка физический констант по результатам прямых или косвенных измерений сводится (путем применения метода максимального правдоподобия или наименьших квадратов, см. гл. 11) к минимизации некоторой функции. Здесь прямой физический смысл имеют сами аргументы минимума этой функции, а найденные их значения — оценки искомых параметров — будут использоваться в разнообразных задачах, не связанных непосредственно с исходной задачей минимизации. Поэтому в данном случае важно возможно точнее найти точку минимума, т. е. мы имеем аргументную задачу. Аналогичным образом обстоит дело в некоторых других задачах оценивания и идентификации. Однако в большинстве случаев сами координаты точки минимума не представляют интереса, а важно лишь обеспечить минимально возможное значение критерия оптимальности. Это соответствует *критериальными задачами* оптимизации. Например, в задачах наилучшего приближения требуется аппроксимировать заданную функцию  $a(t)$  некоторым более простым выражением, например полиномом  $n$ -й степени  $\sum_{i=1}^{n+1} x_i t^{i-1}$ . После выбора соответствующей нормы ( $L_1, L_2, L_\infty$  и т. п.) задача сводится к минимизации функции  $f(x) = \left\| a(t) - \sum_{i=1}^{n+1} x_i t^{i-1} \right\|$ . Однако при этом величина коэффициентов  $x_i^*$ , минимизирующих  $f(x)$ , не представляют никакого самостоятельного интереса — важна лишь малость  $f(x)$ . Более того, вместо алгебраических полиномов мы

могли бы выбрать тригонометрические или искать приближение по какой-либо другой системе функций. Подобная же ситуация имеет место и во многих других задачах, в которых нужно наилучшим образом описать объект с помощью некоторой модели, выбор которой в достаточной мере произволен, а цель заключается в минимизации «крайсогласования» между выходами модели и объекта. Другими примерами критериальных задач могут служить оптимизационные экономические задачи, задачи оптимального проектирования и т. п.

Для критериальных задач вырожденность минимума не представляет особой трудности, поскольку нам достаточно почасть в область малых значений минимизируемой функции  $f(x)$ . Формальным подтверждением этому служат оценки (2), (10), (31) точности по функции приближенных решений, даваемых различными методами. Так, оценка (10) показывает, что при минимизации без помех производственной квадратичной функции (возможно, с вырожденным минимумом) независимо от размерности пространства метод сопряженных градиентов гарантирует оценку  $f(x_k) - f^* = O(k^{-2})$ . Это означает, что за 100 итераций можно уменьшить значение функции приблизительно в 10 000 раз, что обычно достаточно для практических целей. При наличии помех оценка (31) дает гарантию, что если уровень помех мал, то градиентный метод с правилом остановки (28) дает возможность найти достаточно хорошее приближение по функции независимо от вырожденности минимума и от размерности задач с вырожденным минимумом удается построить удовлетворительно работающие алгоритмы.

Значительно сложнее положение с аргументными задачами. Заметим, что даже при отсутствии помех мы имели лишь результаты о сходимости методов (теоремы 1–6), однако ни в одном случае не была получена оценка скорости сходимости. Как уже отмечалось, теоремы о сходимости без оценок скорости сходимости не могут считаться достаточным обоснованием рабочеспособности метода. Более того, мы видели на примерах, что скорость сходимости всех рассматривавшихся методов могла быть очень мала. Поэтому ни один из методов не может гарантировать отыскание вырожденного минимума (с заданной точностью по аргументам) за априори определенное число итераций. В практических задачах вычисления осложняются неизбежными погрешностями. Результаты о поведении методов при наличии помех (теоремы 7, 8) не содержат никаких оценок близости получаемого приближенного решения к точному (по аргументу). Теорема 8 дает некоторое асимптотическое утверждение — если уровень помех стремится к 0, то приближенные решения сходятся к точному. Однако на практике мы решаем задачу при одном фиксированном уровне помех, и этот асимpto-

тический результат (не сопровождающийся оценками точности), по существу, не несет никакой информации о гарантированной точности решения.

Приведенные выше пессимистические соображения не означают, что аргументные вырожденные задачи во всех случаях нельзя решить. Часто мы обладаем обширной априорной информацией о решении, и она может быть эффективно использована. Так, если известна близость решения к некоторой точке  $a$ , то последняя может быть выбрана в качестве начального приложения для итеративных методов (например, градиентного).

В соответствии с теоремой 7 в градиентном методе гарантируется отыскание приближенного решения  $x_e$  такого, что  $\|x_e - x^*\| \leq \|a - x^*\|$ , а  $f(x_e)$  тем меньше, чем меньше уровень помех. Другой способ учета априорной информации в данном случае связан с выбором регуляризующей функции вида  $\|x - a\|^2$ . Нередко имеется априорная информация о каких-либо свойствах решения. Такая информация может быть учтена в итеративных методах путем выбора подходящей нормы в методе регуляризации путем выбора  $g(x)$ . Далее, в статистических задачах типа оценки параметров информации о решении обычно формулируется в терминах априорного распределения. Используя байесовский подход, удается включить эту информацию в минимизируемую функцию и тем самым облегчить отыскание решения.

Подводя итог, можно сказать, что возможность решения аргументных задач с вырожденным минимумом обычно определяется наличием априорной информации о решении. При отсутствии такой информации трудно рассчитывать на получение сколько-нибудь точного решения.

## § 2. Многоэкстремальность

До сих пор мы в основном занимались задачами минимизации выпуклых функций, для которых всякий локальный минимум совпадает с глобальным (теорема 2 § 2 гл. 1). В случае, когда функция *многоэкстремальная* (т. е. имеет много локальных минимумов), проблема отыскания *глобального минимума* очень сложна. К ее анализу мы и переходим. Всюду в этом параграфе речь будет идти о задаче

$$\min f(x), \quad x \in \mathbf{R}^n, \quad (1)$$

где функция  $f(x)$  гладкая, но не выпуклая.

**1. Предварительные замечания.** Мы знаем (теорема 1 § 2 гл. 1), что всякая точка локального минимума  $x^*$  в задаче (1) является стационарной, т. е.  $\nabla f(x^*) = 0$ . Обратно, если в стационарной точке  $\nabla^2 f(x^*) > 0$ , то  $x^*$  является точкой локального (или глобального) минимума (теорема 4 § 2 гл. 1). Совершенно анало-

гично можно показать, что если  $\nabla f(x^*) = 0$  и  $\nabla^2 f(x^*) < 0$ , то  $x^*$  — точка локального максимума. Наконец, если матрица  $\nabla^2 f(x^*)$  в стационарной точке  $x^*$  не является знакопределенной, то найдутся как векторы  $y$ , для которых  $f(x^* + \varepsilon y) \geq f(x^*)$  для достаточно малых  $\varepsilon > 0$ , так и векторы, для которых  $f(x^* + \varepsilon y) < f(x^*)$ . Первые называются *направлениями возрастания*, вторые — *направлениями убывания*, а точка  $x^*$  называется *седловой*. Сведем эти результаты в следующую теорему.

**Теорема 1.** Пусть  $\nabla f(x^*) = 0$ , матрица  $\nabla^2 f(x^*)$  невырождена,  $\lambda_1 \leq \dots \leq \lambda_n$  — ее собственные значения,  $e_1^1, \dots, e^n$  — соответствующие им ортонормированные собственные векторы. Тогда, если  $\lambda_1 > 0$ , то  $x^*$  — точка минимума, если  $\lambda_n < 0$ , то  $x^*$  — точка максимума, а если  $\lambda_1 < 0 < \lambda_n$ , то  $x^*$  — седловая точка, причем векторы  $y \in L_- = \left\{ \sum_{i: \lambda_i < 0} \gamma_i e^i \right\}$ ,  $y \neq 0$ , являются

направлениями убывания, а  $y \in L_+ = \left\{ \sum_{i: \lambda_i > 0} \gamma_i e^i \right\}$ ,  $y \neq 0$ , являются направлениями возрастания. При этом  $\mathbf{R}^n = L_- \oplus L_+$ ,

т. е.  $\mathbf{R}^n$  — прямая сумма подпространств  $L_-$  и  $L_+$ .  $\blacktriangle$  Точка  $x^*$  с  $\nabla f(x^*) = 0$  и невырожденным гессианом называется *невырожденной стационарной точкой*, размерность подпространства  $L_-$  называется *индексом стационарной точки*, так что индекс равен 0 тогда и только тогда, когда  $x^*$  — точка минимума.

Перейдем к анализу поведения основных методов минимизации к окрестности различных стационарных точек. Начнем с градиентного метода вида  $x^{k+1} = x^k - \gamma \nabla f(x^k)$ . Мы знаем (теорема 4 § 4 гл. 1), что в окрестности невырожденного минимума градиентный метод при  $0 < \gamma < 2/\|\nabla^2 f(x^*)\|$  сходится к  $x^*$ , независимо от того, каким минимумом является  $x^*$ , локальным или глобальным. Пусть теперь  $x^*$  — невырожденная стационарная точка с индексом, отличным от 0. Тогда

$$x^{k+1} - x^* = x^k - x^* - \gamma \nabla f(x^k) = -(I - \gamma \nabla^2 f(x^*)) (x^k - x^*) + o(x^k - x^*). \quad (2)$$

Если  $x^*$  — точка максимума, то все собственные значения матрицы  $I - \gamma \nabla^2 f(x^*)$  больше 1 при любом  $\gamma > 0$  (они равны  $1 - \gamma \lambda_i$ ,  $i = 1, \dots, n$ , но все  $\lambda_i < 0$ , см. теорему 1). Поэтому  $\|(I - \gamma \nabla^2 f(x^*))z\| \geq q\|z\|$ ,  $q > 1$ , для всех  $z$ . Отсюда следует, что для достаточно малых  $\|x^k - x^*\| \neq 0$  будет  $\|x^{k+1} - x^*\| > \|x^k - x^*\|$ . Итак, если точка  $x^0$  близка к  $x^*$ , но не совпадает с  $x^*$ , то итерации в градиентном методе будут удаляться от  $x^*$ . Иначе говоря, точка максимума — *точка отталкивания* для градиентного процесса, и траектория, попавшая в окрестности такой точки, заведомо выйдет из нее (за исключением особого случая, когда  $x^0 = x^*$ ).

Для случая, когда  $x^*$  — седловая точка, анализ прост, если  $f(x) — квадратичная функция$ . Тогда  $x^{k+1} — x^* = (I - \gamma A)(x^k - x^*)$ ,  $x^k - x^* = (I - \gamma A)^k(x^0 - x^*)$ , (3)

где  $A = \nabla^2 f(x)$ . В обозначениях теоремы 1  $(x^k - x^*, e^i) = (I - \gamma \lambda_i)(x^0 - x^*, e^i)$ . Если  $\lambda_i > 0$ ,  $0 < \gamma < 2\|A\|$ , то  $|1 - \gamma \lambda_i| < 1$ , и потому  $(x^k - x^*, e^i) \rightarrow 0$ . Если же  $\lambda_i < 0$ , то  $(x^k - x^*, e^i) = q_i^k(x^0 - x^*, e^i)$ , где  $q_i = 1 - \gamma \lambda_i > 1$ , и поэтому  $(x^k - x^*, e^i) \rightarrow \infty$  при  $(x^0 - x^*, e^i) \neq 0$ . Поскольку  $\|x^k - x^*\|^2 = \sum_{i=1}^n (x^k - x^*)^i$ , то получаем  $\|x^k - x^*\| \rightarrow \infty$ , если  $x^0 - x^* \in L_+$ . Итак, если начальное приближение не принадлежит подпространству  $L_+$ , то траектория градиентного метода будет удаляться от седловой точки. Для неквадратичного случая анализ более сложен, но приводит к аналогичным выводам, — лишь для исключительного множества начальных точек градиентные итерации приводят к седловой точке.

Грубо говоря, градиентный метод «почти никогда» не сходится к точке максимума или седловой точке. В то же время он не различает точки локального и глобального минимума и сходится к произвольной из них.

Несколько иначе ведет себя метод Ньютона. Из теоремы 3 § 5 гл. 1 следует, что для его сходимости не требуется положительной определенности  $\nabla^2 f(x^*)$  — достаточно невырожденность этой матрицы. Поэтому метод Ньютона может сходиться к любой стационарной точке, так как он не отличает максимумов от минимумов или от седловых точек.

Не будем подвергать исследованию другие методы минимизации, рассмотренные в предыдущих главах. Некоторые из них существенно опираются на предположение о выпуклости функции и при невыполнении этого предположения теряют работоспособность (таковы почти все методы гл. 5). Другие могут сходить к любой стационарной точке (некоторые варианты квазиньютоновских методов). Наконец, большинство методов, как правило, сходится к произвольной точке локального минимума. Важно подчеркнуть, что при этом ни один метод не дает гарантии попадания в глобальный минимум.

**2. Точные методы.** Все методы многоэкстремальной оптимизации можно условно разделить на две группы. Для первых существуют какие-либо точные утверждения об их сходимости к глобальному минимуму, для вторых приходится ограничиваться некоторыми правдоподобными рассуждениями об их разумном поведении в многоэкстремальной ситуации. Будем говорить в первом случае о *точных* методах, во втором — об *эвристических*. Придумать точные методы нетрудно, однако их ценность, как правило, невелика. Приведем характерный пример.

**Теорема 2.** Пусть  $f(x) — непрерывная функция на множестве  $Q = \{a \leq x \leq b\} \subset \mathbb{R}^n$ ,  $x^k — последовательность независимых равномерно распределенных на  $Q$  случайных векторов$ . Тогда  $\min_{1 \leq i \leq k} f(x^i) \xrightarrow{P} \min f(x)$ .$

Доказательство. По теореме Вейерштрасса (§ 3 гл. 1) существует точка  $x^*$  глобального минимума  $f(x)$  на  $Q$ . Пусть  $\varepsilon > 0$  произвольно, в силу непрерывности  $f(x)$  найдется окрестность  $U$  точки  $x^*$ , для которой  $f(x) \leq f(x^*) + \varepsilon$  при  $x \in U$ . Пусть  $v — объем  $U \cap Q$ ,  $V — объем  $Q$ , тогда  $v \geq 0$  в силу открытости  $U$ . Вероятность попадания точки  $x^i$  в  $U \cap Q$  равна  $v/V$ ; вероятность, что хотя бы одна из точек  $x^1, \dots, x^k$  попадет в  $U \cap Q$ , равна  $p_k = 1 - (1 - v/V)^k$ . Очевидно, что  $p_k \rightarrow 1$  при  $k \rightarrow \infty$ , т. е.  $P\{\min_{1 \leq i \leq k} f(x^i) > f(x^*) + \varepsilon\} \rightarrow 0$  при  $k \rightarrow \infty$ . Это и означает сходимость по вероятности. ▲$$

Теорема 2 столь же проста и универсальна, сколь и бескордажательна. Оценим число вычислений функции, требуемое для отыскания решения с небольшой точностью. Для одного примера. Пусть  $x = (x_1, \dots, x_{10}) \in \mathbb{R}^{10}$ ,  $f(x) = \max_{1 \leq j \leq 10} x_j$ ,  $Q = \{x\}$   $0 \leq x_j \leq 1$ ,  $j = 1, \dots, 10\}$ , и зададимся точностью  $\varepsilon = 10^{-2}$ . Тогда  $x^* = 0$ ,  $f(x^*) = 0$ ,  $v = (10^{-2})^{10} = 10^{-20}$ ,  $V = 1$ ,  $p_k \approx k \cdot 10^{-20}$ , т. е., чтобы вероятность отыскания  $x^*$  с точностью до  $1\%$  была равна хотя бы  $10\%$ , потребуется  $\sim 10^{19}$  итераций. Иначе говоря, метод случайного поиска (в описанной в теореме 2 форме) абсолютно непригоден для нахождения глобального минимума уже для размерностей порядка 10. Мы вновь столкнулись с тем фактом, что сама по себе теорема о сходимости отнюдь не гарантирует работоспособности метода. Тем не менее регулярно появляются работы, содержащие результаты типа теоремы 2 (это дало повод для появления уже упоминавшейся пародийной статьи Бульфа [1.11], в которой с полной серьезностью обсуждается детерминированный вариант теоремы 2). Конечно, авторы подобных работ обычно не столь откровенны, и предлагаемые ими методы выглядят более солидно, однако математическое обоснование бывает столь же «глубоким», как теорема 2. В то же время читателю должно быть ясно, что для произвольных непрерывных или даже гладких функций невозможен метод существенно лучший, чем в теореме 2. На рис. 26 приведены примеры функций, для которых глобальный минимум нельзя найти иначе, чем путем перебора ее значений на достаточно мелкой сетке. Поэтому нужно сузить класс рассматриваемых функций. Рассмотрим функции, удовлетворяющие условию Липшица

$$|f(x) - f(y)| \leq L \|x - y\|, \quad (4)$$

причем будем предполагать, что константа  $L$  известна. При

минимизации такого рода функций можно руководствоваться следующим соображением. Пусть уже найдено рекордное значение  $f(x)$  по  $k-1$  предыдущим итерациям:  $\Phi_{k-1} = \min_{\mathbf{x} \in \mathcal{X}_{k-1}} f(\mathbf{x})$  — и вычислено  $f(x^k)$ . Тогда, если  $f(x^k) < \Phi_{k-1}$ , то улучшается значение рекорда:  $\Phi_k = f(x^k)$ , если же  $f(x^k) > \Phi_{k-1}$ , то в шаре  $\{\mathbf{x}: \|\mathbf{x} - \mathbf{x}^k\| < L^{-1}(f(x^k) - \Phi_{k-1})\}$  заведомо не может быть точки глобального минимума, что приводит к сокращению области возможной локализации минимума. Эту идею нетрудно реализовать в вычислительный алгоритм для одномерного случая.

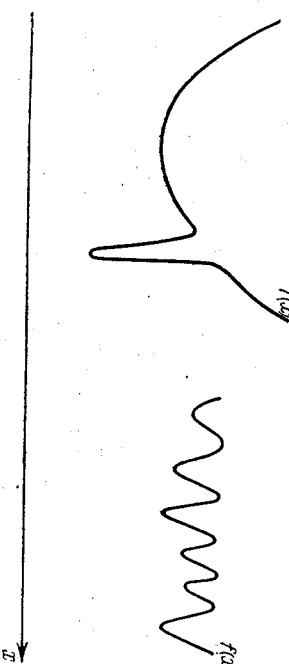


Рис. 26. Функции, для которых трудно найти глобальный минимум.

В многомерном случае возникают серьезные трудности, связанные со сложностью описания области локализации минимума и реализации правила выбора новой точки. Эффективность такого рода методов зависит от вида функции и расстановки точек.

Например, если разница между  $f(x^1)$  и  $f(x^2)$  велика, то удается сразу отсечь область большого объема. Если же функция имеет вид, изображенный на рис. 26, то при любых правильных выборах  $x^k$  метод будет не лучше полного перебора. Кроме того, в реальных задачах константа  $L$  в (4) редко бывает известна, а ее неправильное назначение может либо резко замедлить метод, либо привести к потере глобального минимума.

Аналогичная ситуация возникает, если известны оценки производных минимизируемой функции. Например, если  $\nabla f(x)$  удовлетворяет условию Липшица

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \quad (5)$$

и  $L$  известно, то сокращение области можно осуществлять, используя неравенство

$$|f(x) - f(x^k) - (\nabla f(x^k), x - x^k)| \leq (L/2) \|x - x^k\|^2. \quad (6)$$

Мы не будем на этом подробнее останавливаться, так как этому методу присущи те же недостатки, что и описанные выше,

другие классы многоэкстремальных функций, которые были бы естественны и допускали бы простое описание, пока не придуманы. В целом можно сказать, что существующие точные методы поиска глобального экстремума не могут рассматриваться как эффективные для решения многомерных задач.

### 3. Детерминированные эвристические методы.

Один из возможных подходов к решению многоэкстремальных задач заключается в сочетании методов локальной оптимизации с той или иной процедурой перебора начальных точек. Например, можно произвести спуск методом сопряженных градиентов из вершин грубой равномерной сетки, покрывающей область априорной локализации минимума. Исходные «пробные точки» можно расставлять и иначе. Так, существуют способы «более равномерно» распределять точки в многомерном параллелепипеде, чем в вершинах прямоугольной сетки, примером могут служить так называемые *ЛП-последовательности* [6.14]. При этом число проенных точек может быть невелико (несколько десятков). Процесс очередной локальной минимизации следует прекратить, если мы либо попадаем в уже обследованную зону локального минимума, либо если значение функции в грубо найденном локальном минимуме заметно больше уже достигнутого рекорда.

Более интересны методы, в которых глобальный поиск представим как единий итеративный процесс. Для этого алгоритм должен обладать способностью «выходить» из локальных минимумов. Простейшим примером является *метод тяжелого шарика* (§ 2 гл. 3), в котором приближения  $x^k$  связаны формулами

$$x^{k+1} = x^k - \alpha \nabla f(x^k) + \beta (x^k - x^{k-1}). \quad (7)$$

Ясно, что если  $\nabla f(x^k) = 0$ , но  $x^k \neq x^{k-1}$ , то будет  $x^{k+1} \neq x^k$ , т. е. метод не «застревает» в стационарной точке. Из механической аналогии (7) (движение тяжелого шарика по неровной поверхности) следует, что если скорость шарика достаточно велика, то он «проскаивает» неглубокие ямки. Можно проверить на примерах, что метод действительно обладает свойством выходить из небольших локальных минимумов. Однако он может «скатиться» в более глубокий минимум и уже не выйти из него. Поэтому метод тяжелого шарика (7) — ненадежный способ отыскания глобального минимума.

Более перспективным является *обратный метод* И. М. Гельфандова и М. Л. Цетлина. Авторы исходят из представления об овражном строении минимизируемой функции, т. е. считается, что функция слабо меняется по некоторым направлениям (обратным дно оврага) и резко — по другим (направления склонов оврага). Примером однозадачной овражной функции может служить квадратичная функция с плохо обусловленной матрицей. Вообще, в окрестности локального минимума овражные функции характеризуются большим числом обусловленно-

сти  $\mu$  (см. § 3 гл. 1). Овражный метод состоит из шагов спуска, произволящихся с помощью какого-либо локального метода (обычно градиентного) и осуществляющих спуск на дно оврага, и *овражных шагов* вдоль дна оврага. Структура метода видна из рис. 27, на нем  $x^0$  и  $x^1$  — два начальных приближения, тонкими линиями обозначены шаги спуска, жирными — овражные шаги. Траектория в овражном методе проходит в основном по дну оврага, не задерживаясь в локальных минимумах (важно подчеркнуть, что овражные шаги делаются определенной длины).

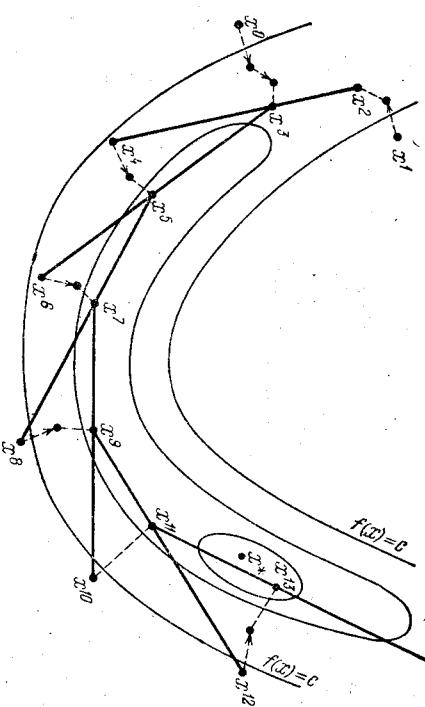


Рис. 27. Овражный метод.

независимо от возрастания или убывания функции в данном направлении). Овражный метод предназначен для грубого просмотра всей области определения функции. Найденные в результате точки с малыми значениями  $f(x)$  следует далее уточнить с помощью более мелких локальных методов. Овражный метод не свободен от ряда недостатков. Во-первых, довольно трудной проблемой является подбор длины овражного шага, так как если она велика, то метод «перепрыгивает» многие минимумы, если же мала, то метод не отслеживает направления дна оврага и движение становится хаотическим. Во-вторых, направление овражного шага не определено однозначно и зависит от многих обстоятельств (точность локальных спусков, расположение предыдущей точки и т. д.). Вообще, наличие многих «свободных параметров» в овражном методе приводит к тому, что его использование требует большого опыта и обстоятельной предварительной «отладки».

Идеи овражного метода используются в *методе СПП* (спуск-подъем-перевал). В нем весь процесс поиска глобального ми-

нимума разделен на три этапа, которые повторяются циклически. На этапе спуска с помощью метода сопряженных градиентов отыскивается локальный минимум. На этапе подъема происходит выход из зоны минимума. Движение осуществляется по направлению «к наимедленнейшему подъему», которое находится следующим образом. В точке  $x^k$  строится функция  $f_k(v) = f(x) - (\nabla f(x^k), v)$ . Очевидно, что  $\nabla f_k(x^k) = \nabla^2 f(x^k) > 0$ , то  $x^k$  — точка локального минимума  $f_k(x)$ , если же  $\nabla^2 f(x^k)$  не знакоопределенна, то  $x^k$  — седловая точка  $f_k(x)$ . Из точки  $z^0 = x^k + \varepsilon d_{k-1}$  (где  $d_{k-1}$  — направление предыдущего движения,  $\varepsilon > 0$  — параметр) делается несколько шагов градиентного метода для  $f_k(x)$ :  $z^{i+1} = z^i - \gamma \nabla f_k(z^i)$ . Если точки  $z^i$  стремятся к  $x^k$ , то это признак знакоопределенности  $\nabla^2 f_k(x^k) = \nabla^2 f(x^k)$  (см. исследование поведения градиентного минимума в п. 1 в окрестности локального минимума и седловой точки), и направление  $d^k = (z^i - x^k) / \|z^i - x^k\|$  берется в качестве направления подъема. Легко проверить, что оно близко к собственному вектору  $\nabla^2 f(x^k)$ , отвечающему наименьшему собственному значению этой матрицы. Делается шаг  $\tilde{x}^{k+1} = x^k + \lambda_k d^k$ , а градиентный шаг из  $\tilde{x}^{k+1}$  приводит к новой точке  $x^{k+1}$ , где повторяется этап подъема. Если же точки  $z^i$  удаляются от  $x^k$ , то делается вывод о невыпуклости  $f(x)$  в окрестности  $x^k$  и происходит переход к этапу перевала. При этом вектор  $d^k = (z^i - x^k) / \|z^i - x^k\|$  задает направление движения на перевал, которое сочетается с градиентным спуском после каждого шага. При прохождении перевала (критерием этого является перемена знака величины  $(\nabla f(x^{k+1}), d^k)$ ) начинается спуск в новый локальный минимум. Типичная траектория метода СПП показана на рис. 28. Разделение поиска на этапы представляется более целесообразным по сравнению с едицобразной процедурой движения в овражном методе.

Известны и многие другие эвристические методы глобальной оптимизации. К сожалению, какие-либо строгие результаты об их эффективности отсутствуют, а проверка на тестовых задачах недостаточно убедительна, да пока и не произволовилась достаточно обстоятельно.

**4. Стохастические эвристические методы.** Здесь можно выделить два подхода. В первом случайность вносится в процесс минимизации (*методы случайного поиска*), во втором строится та или иная *стохастическая модель минимизируемой функции*.

Методы случайного поиска для локальной оптимизации уже описывались в § 4 гл. 3. Для приятия им глобального характера нужно, чтобы в этих методах были разрешены большие шаги, выводящие из окрестности локального минимума. Описанная простейший варант такого метода. Пусть ищется глобальный минимум  $f(x)$ ,  $x \in \mathbf{R}^n$ , на единичном кубе  $Q = \{x: 0 \leq x_i \leq 1\}$ . В точке  $x^k$  выбирается вектор  $h_k$  с независимыми компо-

нентами, равномерно распределенными на  $[-1, 1]$ , и если  $x^k + h^k \in Q$  и  $f(x^k - h^k) < f(x^k)$ , то берется  $x^{k+1} = x^k + h^k$ . В приведенном случае выбирается новая реализация  $h^k$ . Метод выглядит вполне разумно, для него можно доказать теорему о сходимости к глобальному минимуму и т. д. Однако легко убедиться, что он (с точностью до обозначений) совпадает с методом перебора из теоремы 2, т. е. является, как мы видели, совершенно неэффективным. К сожалению, эта же опасность подстерегает

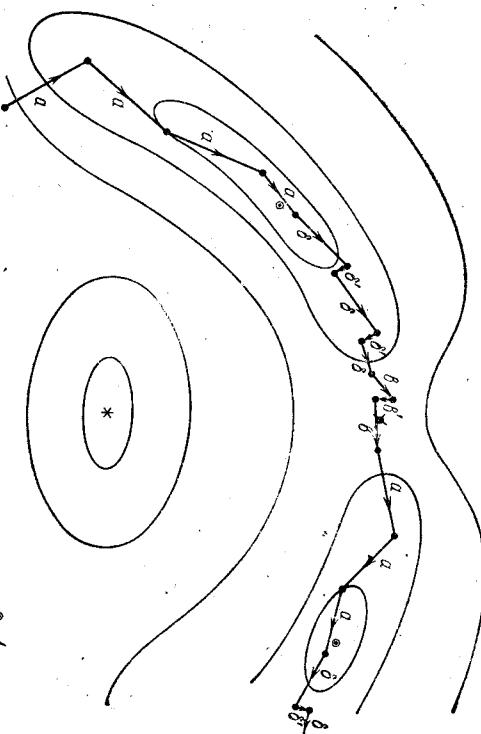


Рис. 28. Метод СПП для глобальной минимизации:  $a$  — этап спуска,  $b$  — этап подъема,  $\delta$  — этап перевала;  $f$  — точка минимума,  $2$  — седловая точка,  $3$  — точка максимума.

нас и в других методах случайногопоиска, хотя она может быть и не столь очевидна, как в приведенном «книжном» варианте метода. Поэтому трудно разделить оптимизм энтузиастов случайногопоиска, считающих, что они обладают эффективным средством глобальной оптимизации. Читатели, желающие ознакомиться с разнообразными модификациями методов случайногопоиска, мы отсылаем к обширной литературе по этой тематике. Упоминавшийся выше второй подход использования случайности в глобальной оптимизации основывается на идеи, что после вычисления минимизируемой функции в  $k$  точках  $x^1, \dots, x^k$  можно говорить о вероятностях тех или иных ее значений в других точках. Понятию «вероятность» при этом иногда придается точный смысл — считается, что имеется ансамбль функций с заданной на нем вероятностной мерой, а минимизируемая функция  $f(x)$  есть элемент этого ансамбля; тогда можно говорить об

условных вероятностях тех или иных событий при реализации значений  $f(x^1), \dots, f(x^k)$ . Чаще же ограничиваются нестрогой вероятностной моделью. Обычно считают, что реализация значения  $f(x^k)$  в точке  $x^k$  «повышает вероятность» близких к  $f(x^k)$  значений  $f(x)$  для точек из окрестности  $x^k$  и не меняет их вдали от  $x^k$ . Задаваясь тем или иным априорным распределением и достаточно произвольным правилом пересчета, получают априорное распределение значений  $f(x)$  для всех  $x$ . В качестве  $x^{k+1}$  выбирается точка, где «математическое ожидание»  $f(x)$  минимально, и, после вычисления  $f(x^{k+1})$  пересчет вероятностей производится заново. Известно много конкретных реализаций подобной идеи. Для всех методов такого типа трудности представляют способы записи апостериорных вероятностей и процедуры поиска «наиболее перспективной» точки. Кроме того, произвол в исходных предположениях при построении подобных методов оставляет ощущение неудовлетворенности.

Мы постарались дать лишь общую картину положения в области глобальной оптимизации. Как видят читатель, положение здесь далеко от благополучного. Требуются интенсивные дальнейшие теоретические и численные исследования, имеющихся методов. Однако больше всего ощущается потребность в новых идеях, прежде всего в классификации многоэкстремальных задач и выделении сравнительно узких классов задач, допускающих специальные и достаточно эффективные методы их решения.

### § 3. Нестационарность

В некоторых практических задачах, связанных с управлением объектами в реальном масштабе времени, критерий оптимальности не остается постоянным, а изменяется с течением времени (например, вследствие дрейфа характеристик объекта). Четкая постановка подобных *нестационарных задач* оптимизации может быть различной в зависимости от целей управления и доступной информации. Опишем кратко некоторые возможные ситуации.

**1. Известен вид  $f(x, t)$ .** Пусть минимизируемая функция зависит от скалярного параметра  $t$  (не обязательно имеющего смысла реального времени), т. е. имеет вид  $f(x, t)$ ,  $x \in \mathbf{R}^n$ ,  $t \in \mathbf{R}$ . Обозначим точку локального или глобального минимума  $f(x, t)$  при фиксированном  $t = t_0$  через  $x_0^*$ :

$$x_0^* = \underset{x \in \mathbf{R}^n}{\operatorname{argmin}} f(x, t_0). \quad (1)$$

Тогда, в соответствии с необходимыми условиями минимума,  $x_0^*$  является решением уравнения

$$\nabla_x f(x_0^*, t_0) = 0.$$

Если предположить, что выполнено достаточное условие минимума  $\nabla_{xx}^2 f(x^*, t_0) > 0$ , матрица  $\nabla_{xx}^2 f(x, t)$  непрерывна в  $\{x^*, t_0\}$ ,  $\nabla_x f(x, t)$  дифференцируем по  $t$  в  $\{x_0, t_0\}$  и  $\nabla_x f(x, t)$  непрерывен в окрестности  $\{x_0, t_0\}$ , то выполнены условия теоремы о неявной функции (теорема 2 § 3 гл. 2) и потому в окрестности  $t_0$  существует дифференцируемая функция  $x^*(t)$ , для которой  $\nabla_x f(x^*(t), t) = 0$ , и она определяется уравнением

$$\frac{dx^*}{dt} = -[\nabla_{xx}^2 f(x^*(t), t)]^{-1} \nabla_{xt} f(x^*(t), t), \quad x^*(t_0) = x_0. \quad (2)$$

В силу непрерывности  $\nabla_{xx}^2 f(x, t)$  в окрестности  $t_0$  выполняется условие  $\nabla_{xx}^2 f(x^*(t), t) > 0$ , являющееся достаточным условием экстремума, т. е.

$$x^*(t) = \operatorname{argmin}_{x \in \mathbb{R}^n} f(x, t). \quad (3)$$

Иными словами, зная точку минимума при одном  $t = t_0$ , можно из (2) найти точку минимума и для близких значений  $t$ . Если  $f(x, t)$  сильно выпукла по  $x$  при каждом  $t$ , то глобальный минимум  $x^*(t) = \operatorname{argmin}_x f(x, t)$  существует и единственный для всех  $t$

и описывается уравнением (2), которое имеет решение, продолжаемое на всей оси. Итак, если вид функции  $f(x, t)$  известен (достаточно, чтобы были доступны ее производные  $\nabla_{xx}^2 f(x, t)$  и  $\nabla_{xt}^2 f(x, t)$ ), то траектория точек минимума  $x^*(t)$  можно отследить, решая дифференциальное уравнение (2), если точка минимума известна в какой-либо момент  $t_0$ .

Конечно, такой подход имеет в основном принципиальное значение, так как, во-первых, дифференциальное уравнение (2) нельзя решить точно, во-вторых, точка минимума  $x_0^*$  в момент  $t_0$  может быть найдена лишь приближенно, и, в-третьих, вид зависимости  $f$  от  $t$  обычно неизвестен. Чтобы преодолеть первые две трудности, можно перейти к дискретному времени, т. е. заменить дифференциальное уравнение конечно-разностным, а в качестве начального приближения выбирать не обязательно точку минимума.

**2. Вид  $f(x, t)$  неизвестен.** Рассмотрим несколько иную ситуацию, в которой отсутствует какая-либо информация о законе изменения минимизируемой функции во времени. Пусть в  $k$ -й момент времени (речь идет о дискретном варианте задачи) имеется функция  $f_k(x)$ , причем можно вычислять ее значения и значения производных в произвольной точке. Тогда можно следить за несколько итераций какого-либо метода минимизации  $f_k(x)$ , а полученную точку взять в качестве начального приближения для минимизации  $f_{k+1}(x)$ . В простейшем варианте можно делать лишь один шаг градиентного метода

$$x^{k+1} = x^k - \gamma \nabla f_k(x^k) \quad (4)$$

или метода Ньютона  $x^{k+1} = x^k - [\nabla^2 f_k(x^k)]^{-1} \nabla f_k(x^k)$ . (5)

Нас будет интересовать вопрос о поведении подобных итераций. С другой точки зрения этот вопрос заключается в исследовании градиентного метода или метода Ньютона в нестационарной ситуации.

В действительности мы уже занимались близкими задачами, когда изучали влияние помех на методы оптимизации. Например, если существует предельная функция  $f(x)$  — такая, что  $f_k(x) \rightarrow f(x)$ ,  $\nabla f_k(x) \rightarrow \nabla f(x)$ , то  $\nabla f_k(x^k)$  можно записать в виде  $\nabla f_k(x^k) = \nabla f(x^k) + (\nabla f_k(x^k) - \nabla f(x^k))$  и последний член рассматривать как «помеху». Тогда (4) есть не что иное, как градиентный метод минимизации  $f(x)$  при наличии помех, и можно применять результаты § 2 гл. 4. Если предельной функции не существует, то методы (4), (5) нужно исследовать непосредственно. Ограничимся одним характерным примером. Пусть все функции  $f_k(x)$  дважды дифференцируемы, причем для всех  $x$  и  $k$

$$L \leq \nabla^2 f_k(x) \leq L, \quad l > 0. \quad (6)$$

Тогда каждая функция  $f_k(x)$  имеет единственную точку минимума  $x_k^*$ . Предположим, что эти точки минимума дрейфуют с ограниченной скоростью

$$\|x_k^* - x_{k+1}^*\| \leq a. \quad (7)$$

**Теорема 1.** При сделанных выше предположениях для метода (4) с  $0 < \gamma < 2/L$  справедлива оценка

$$\lim_{k \rightarrow \infty} \|x^k - x^*\| \leq \frac{a}{1-q}, \quad q = \max \{ |1 - \gamma l|, |1 - \gamma L| \} < 1. \quad (8)$$

**Доказательство.** Так же как при доказательстве теоремы 3 § 4 гл. 1, имеем

$$\|x^{k+1} - x_k^*\| = \|x^k - \gamma \nabla f_k(x^k) - x_k^*\| \leq q \|x^k - x_k^*\|.$$

Отсюда  $\|x^{k+1} - x_{k+1}^*\| \leq \|x^{k+1} - x_k^*\| + \|x_{k+1}^* - x_k^*\| \leq q \|x^k - x_k^*\| + a$ .

Используя лемму 1 § 2 гл. 2 для  $u_k = \|x^k - x_k^*\|$ , получаем (8). ▲

Таким образом, градиентный метод (4) отслеживает нестационарный минимум с точностью до величин порядка  $a$ . Учитывая полное отсутствие информации о законе движения минимума, трудно рассчитывать на большее.

В ряде случаев упомянутая информация может быть доступна. Например, может быть известно, что траектория оптимумов описывается разностным уравнением

$$x_{k+1}^* = g_k(x_k^*), \quad (9)$$

где неизвестно начальное условие  $x_0^*$  (ср. с описанием непрерывной траектории  $x^*(t)$  с помощью уравнения (2)). В этом случае целесообразно ввести прогнозирование, даваемое (9), в методы минимизации. В частности, градиентный метод (4) принимает вид

$$x^{k+1} = g_k(x^k) - \gamma \nabla f_k(g_k(x^k)). \quad (10)$$

**3. Заключительные замечания.** При анализе методов оптимизации мы начали с наиболее простого случая — невырожденного безусловного минимума гладкой функции при полной информации о задаче — и постепенно учитывали возможные усложнения этой ситуации (недоступность производных, наличие помех, негладкость функции, вырожденность минимума, многоэкстремальность, нестационарность). Однако не нужно думать, что тем самым мы полностью исчерпали тематику безусловной минимизации. Многообразие практических задач оптимизации столь велико, что они не укладываются в рамки даже наиболее общих схем. В частности, мы совершенно не рассматривали методы, пригодные для минимизации функций специального вида. Некоторые из таких методов будут описаны позже, когда в части III мы познакомимся с конкретными примерами задач оптимизации.

## ЧАСТЬ II УСЛОВНАЯ МИНИМИЗАЦИЯ

### Глава 7

#### МИНИМИЗАЦИЯ НА ПРОСТЫХ МНОЖЕСТВАХ

Изучение задач условной минимизации начнем с наиболее простых, имеющих вид

$$\min_{x \in Q \subset \mathbb{R}^n} f(x), \quad (A)$$

где  $Q$  — множество «простой» структуры. В принципе условия на это множество в приводимых ниже теоремах являются самыми общими (выпуклость, замкнутость и т. д.). Однако сoderжательный характер эти результаты приобретают, лишь если для множества  $Q$  можно достаточно просто найти фигурирующие в теоремах объекты (опорную гиперплоскость, проекцию и т. п.). В этом смысле и понимается термин *простое множество*. В качестве примеров можно привести параллелепипед  $Q = \{x: a \leq x \leq b\}$ , шар  $Q = \{x: \|x\| \leq \alpha\}$ , линейное многообразие  $Q = \{x: Ax = b\}$  и т. д. Ограничения, задаваемые такими множествами, часто обуславливаются либо физической природой переменных (например, требованием неотрицательности), либо априорными знаниями о решении.

#### § 1. Основы теории

**1. Условия экстремума в гладком случае.** Точка  $x^* \in Q$  называется *локальной точкой минимума* (или просто *точкой минимума*) в задаче (A), если  $f(x) \geq f(x^*)$  для всех  $x \in Q$ ,  $\|x - x^*\| \leq \varepsilon$  при некотором  $\varepsilon > 0$ . Если  $f(x) \geq f(x^*)$  для всех  $x \in Q$ , то будем говорить о *глобальном минимуме*.

**Теорема 1** (необходимое условие минимума I порядка). *Пусть  $f(x)$  дифференцируема в точке минимума  $x^*$ , а  $Q$  — выпуклое множество. Тогда*

$$(\nabla f(x^*), x - x^*) \geq 0 \quad \text{для всех } x \in Q. \quad (1)$$

**Доказательство.** Пусть  $(\nabla f(x^*), x^0 - x^*) < 0$  для некоторого  $x^0 \in Q$ . Тогда  $x(\alpha) = x^* + \alpha(x^0 - x^*) \in Q$  при  $0 \leq \alpha \leq 1$  в силу выпуклости  $Q$  и

$$f(x(\alpha)) = f(x^*) + \alpha(\nabla f(x^*), x^0 - x^*) + o(\alpha) < f(x^*)$$

при достаточно малых  $\alpha > 0$ , что противоречит локальной оптимальности  $x^*$ .  $\blacktriangle$

Вектор  $a \in \mathbb{R}^n$ , удовлетворяющий условию  $(a, x - x^*) \leq 0$  для всех  $x \in Q$ , называется *опорным* к  $Q$  в точке  $x^* \in Q$  (если

$a \neq 0$ , то он задает опорную гиперплоскость  $(a, x - x^*) = 0$ , см. § 1 гл. 5). Поэтому условие (1) может быть сформулировано иначе: вектор  $-\nabla f(x^*)$  является опорным к  $Q$  в точке локального минимума  $x^*$ . Далее, всякий вектор вида  $s = x - x^*$ ,  $x \in Q$ , называется допустимым направлением в точке  $x^*$  относительно выпуклого множества  $Q$ . Это название объясняется тем, что  $x^* + \alpha s \in Q$  при всех  $0 \leq \alpha \leq 1$ . Вспомнив формулу (6) § 1 гл. 1 для производной по направлению  $f'(x; s) = (\nabla f(x), s)$ , можем условие экстремума сформулировать и так: производная по любому допустимому направлению в точке минимума негативна.

Геометрический смысл (1) очень прост (рис. 29) — множество  $Q$  и множество  $S = \{x: (\nabla f(x^*), x - x^*) < 0\}$ , образованное направлениями локального минимума  $x^*$  в точке  $x^*$ , не должны пересекаться.

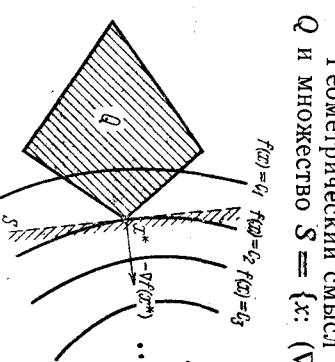


Рис. 29. Условия экстремума на множестве  $Q$ .

Теорема 2 (достаточное условие минимума I порядка). Пусть  $f(x)$  дифференцируема в точке  $x^* \in Q$ ,  $Q$  выпукло и выполняется условие (2) для всех  $x \in Q$ ,  $\|x - x^*\| \leq \varepsilon$ ,  $\varepsilon > 0$ . Тогда  $x^*$  — точка локального минимума  $f(x)$  на  $Q$ .

Доказательство. Выберем  $\varepsilon_1 > 0$ ,  $\varepsilon_1 \leq \varepsilon$ , так что

$$|f(x^* + y) - f(x^*) - (\nabla f(x^*), y)| \leq \alpha \|y\|^2$$

при  $\|y\| \leq \varepsilon_1$ . Тогда для  $x \in Q$ ,  $\|x - x^*\| \leq \varepsilon_1$ , имеем

$$f(x) \geq f(x^*) + (\nabla f(x^*), x - x^*) - \alpha \|x - x^*\|^2 \geq f(x^*) + \alpha \|x - x^*\|^2,$$

т. е.  $x^*$  — точка локального минимума.  $\blacktriangleleft$

Заметим, что (2) заведомо не выполняется, если  $x^*$  — внутренняя точка  $Q$ , поэтому в условиях теоремы 2 минимум достигается обязательно в граничной точке  $Q$  (рис. 30).

В терминах производной по направлению (2) может быть записано так:

$$f'(x^*; s) \geq \alpha \|s\|, \quad \alpha > 0 \quad (3)$$

для всех допустимых  $s$ . Заметим, что каждое естественное достаточное условие экстремума вида  $\langle \nabla f(x^*), s \rangle > 0$  для всех допустимых  $s$  в действительности неверно (упр. 1).

Конкретизируем полученные условия экстремума для нескольких важнейших примеров множеств  $Q$ .

Пусть

$$Q = \{x \in \mathbf{R}^n: a \leq x \leq b\}. \quad (4)$$

Тогда из теорем 1 и 2 для  $f(x)$ , дифференцируемой в  $x^* \in Q$ , получаем, что если  $x^*$  — точка минимума  $f(x)$  на  $Q$ , то

$$\begin{cases} \frac{\partial f(x^*)}{\partial x_i} \leq 0, & a_i < x_i^* < b_i, \\ \frac{\partial f(x^*)}{\partial x_i} \geq 0, & x_i^* = a_i, \\ \frac{\partial f(x^*)}{\partial x_i} \leq 0, & x_i^* = b_i, \end{cases} \quad (5)$$

а если  $x_i^* = a_i$  или,  $x_i^* = b_i$  для всех  $1 \leq i \leq n$  и

$$\frac{\partial f(x^*)}{\partial x_i} \begin{cases} > 0, & x_i^* = a_i, \\ < 0, & x_i^* = b_i, \end{cases} \quad (6)$$

то  $x^*$  — точка минимума  $f(x)$  на  $Q$ . В частности, если ищется минимум  $f(x)$ ,  $x \in \mathbf{R}^1$ , при ограничении  $x \geq 0$ , то условие  $f'(0) \geq 0$  является необходимым, а  $f'(0) > 0$  — достаточным для минимума в 0. С помощью (5) в принципе можно искать точки

минимума на параллелепипеде  $Q$  путем перебора: разбиваем множество индексов  $I = \{1, \dots, n\}$  на три подмножества  $I = I_0 \cup I_+ \cup I_-$ , полагаем  $x_i = a_i$  при  $i \in I_+$ ,  $x_i = b_i$  при  $i \in I_-$  и решаем систему уравнений относительно  $x_i$ ,  $i \in I_0$ :  $\frac{\partial f(x)}{\partial x_i} = 0$ ,  $i \in I_0$ . Если

в найденной точке  $x^*$  окажется, что  $a_i < x_i^* < b_i$ ,  $i \in I_0$  и  $\frac{\partial f(x^*)}{\partial x_i} \geq 0$ ,  $i \in I_+$ ,  $\frac{\partial f(x^*)}{\partial x_i} \leq 0$ ,  $i \in I_-$ , то в  $x^*$  выполнены необходимые условия экстремума. Рассуждаем, так как путь не может рассматриваться как реалистический метод решения задачи. В дальнейшем мы опишем гораздо более эффективные способы минимизации, опирающиеся на условия экстремума.

В качестве второго примера рассмотрим минимизацию на линейном многообразии:

$$Q = \{x \in \mathbf{R}^n: Ax = b\},$$

где  $b \in \mathbf{R}^m$ ,  $A$  — матрица  $m \times n$ . Из теоремы 1 следует, что  $(\nabla f(x^*), x - x^*) \geq 0$  для всех  $x \in Q$ , т. е.  $(\nabla f(x^*), s) \geq 0$  для всех  $s \in L = \{s: As = 0\}$ . Если бы нашлось  $s^0 \in L$  такое, что  $(\nabla f(x^*), s^0) > 0$ , то  $(\nabla f(x^*), -s^0) < 0$ , что невозможно, так как  $-s^0 \in L$ . Поэтому  $(\nabla f(x^*), s) = 0$  для всех  $s \in L$ . Отсюда

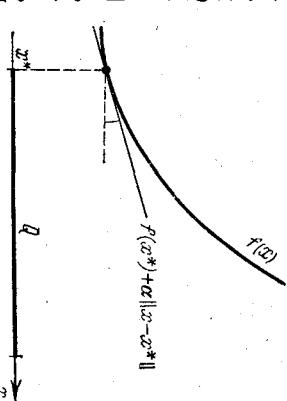


Рис. 30. Острый минимум в задаче с ограничениями.

следует (лемма 1 § 1 гл. 8), что находится  $y^* \in \mathbf{R}^m$  такое, что

$$\nabla f(x^*) = A^T y^*. \quad (8)$$

Итак, (8) является необходимым условием минимума  $f(x)$  на  $Q$  вида (7).

Упражнение.

1. Рассмотрите пример в  $\mathbf{R}^2: \min(y - y^2), y \geq 0, y = x_2 - x_1^2$ . Убедитесь в том, что для  $x^* = 0$   $f'(x^*; s) \geq 0$  для любого допустимого  $s$ , однако  $x^*$  не является точкой локального минимума.

**2. Условия экстремума в выпуклом случае.** Мы будем использовать сведения из теории выпуклых функций, приведенные в § 1 гл. 5.

**Теорема 3.** Пусть  $f(x)$  — выпуклая функция на  $\mathbf{R}^n$ ,  $Q$  — выпуклое множество в  $\mathbf{R}^n$ ,  $x^* \in Q$ . Тогда условие: найдется субградиент  $\partial f(x^*)$  такой, что для всех  $x \in Q$

$$(\partial f(x^*), x - x^*) \geq 0, \quad (6)$$

является необходимым и достаточным для того, чтобы  $x^*$  был глобальным минимумом  $f(x)$  на  $Q$ .

**Доказательство.** Необходимость. Пусть такого субградиента нет. Тогда множества  $S = \partial f(x^*)$  и  $K = \{y: (y, x - x^*) \geq 0, x \in Q\}$  не пересекаются. По лемме 6 § 1 гл. 5  $S$  выпукло, замкнуто и ограничено. Без труда проверяется, что множество  $K$  выпукло и замкнуто. Поэтому применима теорема отделимости (теорема 1 § 1 гл. 5), т. е. найдутся  $c \in \mathbf{R}^n$ ,  $c \neq 0$  и  $\alpha > 0$  такие, что  $(a, c) \leq -\alpha$  для всех  $a \in S$  и  $(c, y) \geq 0$  для всех  $y \in K$ . Пусть  $\Gamma$  — замыкание конуса, порожденного допустимыми направлениями, т. е.  $\Gamma = \{x: x = \lim_{k \rightarrow \infty} \lambda_k(x^k - x^*)\}$ ,  $\lambda_k > 0, x^k \in Q\}$ . Если  $c \in \Gamma$ , то вновь применим теорему отделимости (это возможно, так как множество  $\Gamma$  выпукло и замкнуто) и найдем  $b$  такое, что  $(b, x) \geq 0, x \in \Gamma$ , и  $(b, c) < 0$ . Тогда из определения  $K$  и  $\Gamma$  следует, что  $b \in K$ , поэтому неравенство  $(b, c) < 0$  противоречит условию  $(c, y) \geq 0$  для всех  $y \in K$ .

Итак,  $c \in \Gamma$ , следовательно, найдутся последовательности  $\lambda_k > 0$  и  $x^k \in Q$  такие, что  $\lambda_k(x^k - x^*) \rightarrow c$ . Выберем  $k$  так, чтобы было  $\|\lambda_k(x^k - x^*) - c\| \leq q/(2L)$ ,  $L = \max_{a \in S} \|a\|$ . Тогда в соответствии с леммой 6 § 1 гл. 5

$$f'(x^*; \lambda_k(x^k - x^*)) = \max_{a \in S} (a, \lambda_k(x^k - x^*)) =$$

$$= \max_{a \in S} (a, c) + \max_{a \in S} (\lambda_k(x^k - x^*) - c, a) \leq -\alpha + \frac{1}{2}\alpha = -\frac{1}{2}\alpha.$$

Поэтому  $f'(x^*; x^k - x^*) < 0$  и для достаточно малых  $\gamma > 0$  будет  $f(x^* + \gamma(x^k - x^*)) < f(x^*)$ , что невозможно, если  $x^*$  — точка минимума.

Упражнение.

2. С помощью теоремы 3 покажите, что  $b = P_Q(a)$  тогда и только тогда, когда  $(b - a, x - b) \geq 0$  для всех  $x \in Q$  (ср. с (5) § 1 гл. 5). (Указание.  $P_Q(a)$  есть решение задачи минимизации  $\|x - a\|^2, x \in Q$ ).

### 3. Существование, единственность и устойчивость минимума.

Теорема существования мало меняется по сравнению с теоремой 1 § 3 гл. 1 — условие ограниченности множества  $\{x: f(x) \leq \alpha\}$  заменяется на условие ограниченности множества  $\{x \in Q: f(x) \leq \alpha\}$ , а доказательство остается прежним.

**Теорема 4** (Вейерштрасс). Пусть  $f(x)$  — непрерывная функция на  $Q \subset \mathbf{R}^n$ , множество  $Q$  замкнуто, а множество  $\{x \in Q: f(x) \leq \alpha\}$  ограничено и непусто для некоторого  $\alpha$ . Тогда задача (A) имеет решение. ▲

Если выполняется достаточное условие минимума (2), то минимум единственный.

**Теорема 5.** В условиях теоремы 2  $x^*$  — локально единственная точка минимума.

**Доказательство** следует из неравенства

$$f(x) \geq f(x^*) + \alpha \|x - x^*\|^2, \quad \|x - x^*\| \leq \varepsilon, \quad (10)$$

полученного при доказательстве теоремы 2. ▲

Единственность решения по-прежнему можно гарантировать для строго выпуклой  $f(x)$ . Однако можно наложить и некоторые условия на  $Q$ , приводящие к единственности минимума. Назовем множество  $Q$  строго выпуклым, если для любых  $x_1 \in Q, x_2 \in Q, x_1 \neq x_2, 0 < \lambda < 1$  точка  $\lambda x_1 + (1 - \lambda)x_2$  является внутренней точкой  $Q$ .

**Теорема 6.** Пусть  $f(x)$  — выпуклая функция на  $\mathbf{R}^n$ , множество  $Q$  строго выпукло и  $\|\partial f(x)\| \geq \varepsilon > 0$  для всех субградиентов и всех  $x \in Q$ . Тогда точка минимума  $f(x)$  на  $Q$  единственна. ▲

Понятие устойчивости для задачи (A) можно ввести различным образом. Как и ранее, будем называть задачу минимизации (глобально) устойчивой, если всякая минимизирующая последовательность сходится, т. е. если из  $x^k \in Q, f(x^k) \rightarrow f^* = \inf_{x \in Q} f(x)$  следует, что  $x^k \rightarrow x^*, f(x^*) = f^*$ . Можно определить понятие обобщенно-минимизирующей последовательности:  $f(x^k) \rightarrow f^* = \inf_{x \in Q} f(x)$ ,

$\rho(x^k, Q) \rightarrow 0$ , где  $\rho(x^k, Q) = \inf_{x \in Q} \|x^k - x\|$  и называть задачу

обобщено-устойчивой, если всякая обобщенно-минимизируемая последовательность сходится к точке минимума.

**Теорема 7.** Если  $f(x)$  непрерывна на  $\mathbf{R}^n$ ,  $Q$  замкнуто и подмножество  $\{x \in Q : f(x) \leq \alpha\}$  ограничено и непусто для некоторого  $\alpha$ , а точка глобального минимума  $x^*$  единственна, то задача минимизации устойчива и обобщено-устойчива. ▲

Можно получить количественные оценки устойчивости для сильно выпуклых функций — эти оценки совершенно аналогичны результатам леммы 2 § 2 гл. 5 для безусловного минимума (см. упр. 6). Более интересен случай острого минимума.

Назовем  $x^*$  точкой (глобального) острого минимума  $f(x)$  на  $Q$ , если для всех  $x \in Q$

$$f(x) \geq f(x^*) + \alpha \|x - x^*\|, \quad \alpha > 0, \quad (11)$$

(ср. с (9) § 2 гл. 5). Можно дать аналогичное определение локального острого минимума, а также рассмотреть более общий случай неединственного минимума, но мы ограничимся наиболее простой ситуацией. Для задач с ограничениями острый минимум может достигаться и при гладких  $f(x)$  (см. рис. 30).

**Лемма 1.** Следующие условия эквивалентны (11) для выпуклой  $f(x)$  и выпуклого  $Q$ :

- а)  $f'(x^*, s) \geq \alpha \|s\|$  для всех допустимых направлений  $s$ ;
- б) множество  $-\partial f(x^*)$  и множество опорных к  $Q$  в точке  $x^*$  векторов имеют общую внутреннюю точку. ▲

Из а) следует, что условия (2) и (11) эквивалентны для гладких выпуклых функций. Для невыпуклых функций можно показать, что в условиях теоремы 2  $x^*$  — точка локального острого минимума (см. неравенство (10)).

Основное свойство «сверхустойчивости» острого минимума (см. теорему 6 § 2 гл. 5) сохраняется и для задач с ограничениями.

**Теорема 8.** Пусть  $f(x)$  — выпуклая функция на  $\mathbf{R}^n$ ,  $Q$  — выпуклое замкнутое множество,  $x^*$  — точка острого минимума  $f(x)$  на  $Q$ ,  $g(x)$  — выпуклая функция. Тогда находится  $\varepsilon_0 > 0$  такое, что при  $0 < \varepsilon < \varepsilon_0$  точка минимума функции  $f(x) + \varepsilon g(x)$  на  $Q$  единственна и совпадает с  $x^*$ . ▲

#### Упражнения.

3. Докажите, что шар — строго выпуклое множество, а параллелепипед и полупространство — нет.

4. Докажите, что если  $f(x)$  — строго выпуклая функция на  $\mathbf{R}^n$ , то множество  $\{x : f(x) \leq \alpha\}$  строго выпукло при любом  $\alpha$ .

5. Приведите пример выпуклой (но не строго) функции  $f(x)$ , для которой множество  $\{x : f(x) \leq \alpha\}$  строго выпукло.

6. Пусть  $f(x)$  — сильно выпуклая функция на  $\mathbf{R}^n$ , а множество  $Q$  выпукло и замкнуто. Докажите, что решение  $x^*$  задачи (A) существует и единствено, и  $f(x) \geq f(x^*) + l\|x - x^*\|^2$  для всех  $x \in Q$ , где  $l$  — константа сильной выпуклости.

7. Исследуйте, при каких  $c \in \mathbf{R}^n$  в задаче  $\min_{x \in Q} \|x - c\|^2$ ,  $a \leq x \leq b$ , достигается первый минимум. Ответ. Если  $c_i \geq b_i$  или  $c_i \leq a_i$  для всех  $i \leq i \leq n$ . Покажите, что при условии (6)  $x^*$  — точка острого минимума.

## § 2. ОСНОВНЫЕ МЕТОДЫ

Переходим к исследованию основных методов решения задачи (A). Они аналогичны градиентному методу и методу Ньютона для безусловной минимизации.

**1. Метод проекции градиента.** Этот метод является непосредственным обобщением градиентного метода. Поскольку последний, вообще говоря, выходит за пределы множества, можно добавить операцию проектирования на  $Q$ . Таким путем получаем метод (рис. 31):

$$x^{k+1} = P_Q(x^k - \gamma \nabla f(x^k)), \quad (1)$$

где  $P_Q$  — проектор на  $Q$ . (см. § 1 гл. 5).

**Теорема 1.** Пусть  $f(x)$  — выпуклая дифференцируемая функция в  $\mathbf{R}^n$ , градиент которой удовлетворяет на  $Q$  условию Липшица с константой  $L$ . Пусть  $Q$  выпуклое и замкнутое,  $X^* = \operatorname{Argmin}_{x \in Q} f(x) \neq \emptyset$  и  $0 < \gamma < 2/L$ .

Тогда:

- а)  $x^k \rightarrow x^* \in X^*$ ;
- б) если  $f(x)$  сильно выпукла, то  $x^k \rightarrow x^*$  со скоростью геометрической прогрессии;
- в) если  $f(x)$  дифференцируема и  $lL \leq \nabla^2 f(x) \leq Ll$ ,  $x \in Q$ ,  $l > 0$ , то знаменатель прогрессии равен  $q = \max\{|1 - \gamma l|, |1 - \gamma L|\}$ ;
- г) если  $x^*$  — точка острого минимума, то метод конечен:  $x^k = x^*$  для некоторого  $k$ .

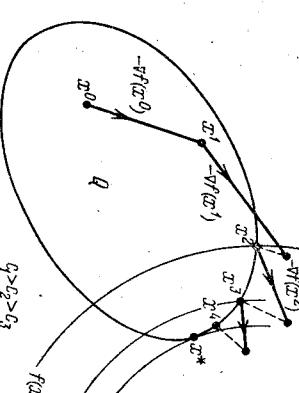


Рис. 31. Метод проекции градиента.

Показательно, что для произвольной точки  $\hat{x}$  и для произвольной точки  $\hat{x}^k$  на  $Q$  имеет место равенство  $\|\hat{x} - \hat{x}^k\|^2 = \|P_Q(\hat{x} - \gamma \nabla f(\hat{x})) - P_Q(\hat{x} - \gamma \nabla f(\hat{x}^k))\|^2$

$\leq \|\hat{x} - \hat{x}^k - \gamma(\nabla f(\hat{x}^k) - \nabla f(\hat{x}))\|^2 = \|\hat{x}^k - \hat{x}\|^2 - 2\gamma(\nabla f(\hat{x}^k) - \nabla f(\hat{x}), \hat{x}^k - \hat{x}) + \gamma^2 \|\nabla f(\hat{x}^k) - \nabla f(\hat{x})\|^2$

Используя лемму 2 § 4 гл. 1, получаем

$$\begin{aligned} \|x^{k+1} - \hat{x}\|^2 &\leq \|x^k - \hat{x}\|^2 - \gamma(2 - \gamma L)(\nabla f(x^k) - \nabla f(\hat{x}), x^k - \hat{x}) \leq \\ &\leq \|x^k - \hat{x}\|^2 - \gamma(2 - \gamma L)(f(x^k) - f(\hat{x})) \end{aligned}$$

Здесь мы применили условие экстремума (1) § 1. Суммируя полученные неравенства по  $k$  и учитывая, что  $\gamma(2 - \gamma L) > 0$ , находим  $\sum_{k=0}^{\infty} (f(x^k) - f(\hat{x})) < \infty$ . Отсюда следует, что  $\lim_{k \rightarrow \infty} f(x^k) = f^* = f(\hat{x})$ . Последовательность  $x^k$  ограничена, так как  $\|x^k - \hat{x}\| \leq \|x^0 - \hat{x}\|$ , поэтому можно выбрать сходящуюся подпоследовательность  $x^{k_l} \rightarrow x^*$ . В силу непрерывности  $f(x)$  при этом  $f(x^{k_l}) \rightarrow f(x^*)$ , т. е.  $f(x^*) = f^*$ . Заменяя  $\hat{x}$  на  $x^*$  и учитывая монотонность  $\|x^k - x^*\|$ , получаем, что  $x^k \rightarrow x^* \in X^*$ .

Пусть  $f(x)$  сильно выпукла, тогда из неравенства  $\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \gamma(2 - \gamma L)(\nabla f(x^k) - \nabla f(x^*), x^k - x^*)$  с учетом (31) § 1 гл. 1 имеем  $\|x^{k+1} - x^*\|^2 \leq (1 - \gamma/(2 - \gamma L))\|x^k - x^*\|^2$ , т. е.  $x^k \rightarrow x^*$  со скоростью геометрической пропрессии.

Если  $f(x)$  дважды дифференцируема, то из полученного выше неравенства  $\|x^{k+1} - x^*\| \leq \|x^k - x^*\| - \gamma(\nabla f(x^k) - \nabla f(x^*))$  оценка скорости сходимости выводится так же, как и при доказательстве теоремы 3 § 4 гл. 1.

Остается доказать конечность метода для острого минимума. Для произвольного  $x \in Q$ , используя (2) § 1 (с учетом леммы 1 § 1), имеем

$$\begin{aligned} (x^k - \nabla f(x^k) - x^*, x - x^*) &= \\ &= (x^k - x^* - \gamma(\nabla f(x^k) - \nabla f(x^*)), x - x^*) - \gamma(\nabla f(x^*), x - x^*) \leq \\ &\leq (1 + \gamma L)\|x^k - x^*\|\|x - x^*\| - \gamma\alpha\|x - x^*\| = \\ &= ((1 + \gamma L)\|x^k - x^*\| - \gamma\alpha)\|x - x^*\| \leq 0 \end{aligned}$$

**Упражнения.**

- Докажите, что условие экстремума (1) § 1 можно записать в виде:  $x^* = P_Q(x^* - \gamma \nabla f(x^*))$  для любого  $\gamma > 0$ .
- Пусть условие острого минимума в теореме 1 заменено на более обобщенное:  $f(x) - f(x^*) \geq \alpha \|x - P_{X^*}(x)\|$ ,  $\alpha > 0$ . Докажите, что метод (1) остается конечным.
- Покажите, что если не требовать выпуклости  $f(x)$ , то сходимости  $x^k$  к точке локального или глобального минимума может и не быть, однако можно доказать, что  $f(x^{k+1}) \leq f(x^k)$  и  $\|x^{k+1} - x^k\| \rightarrow 0$ .
- Предложите конструктивное правило выбора длины шага в методе проекции градиента, аналогичное (10) § 1 гл. 3.
- Приведите пример, показывающий, что метод  $x^{k+1} = P_Q(x^* - \gamma H \nabla f(x^*))$  не сходится при  $H \neq I$ ,  $H > 0$ .

Рассмотрим несколько примеров. Пусть  $Q = \{x: x \geq 0\}$ ,  $x \in \mathbf{R}^n$ . Тогда  $P_Q(x) = x_+$  и метод проекции градиента принимает вид

$$x^{k+1} = (x^k - \nabla f(x^k))_+$$

(2)

Пусть  $Q = \{x: a \leq x \leq b\}$ ,  $x \in \mathbf{R}^n$ . Будем обозначать для скалярных величин  $\tau$ ,  $\alpha \leq \beta$

$$(\tau)_a^\beta = \begin{cases} \tau, & \alpha \leq \tau \leq \beta, \\ a, & \tau < a, \end{cases} \quad (3)$$

Аналогичный смысл имеет обозначение  $(x)_a^\beta$  для вектора  $x$ ,  $a \leq b$  — это вектор,  $i$ -я компонента которого равна  $(x_i)_a^\beta$ . Тогда метод проекции градиента для данного  $Q$  выглядит так:  $x^{k+1} = (x^k - \nabla f(x^k))_a^\beta$ .

Далее, пусть  $Q$  — шар:  $Q = \{x: \|x\| \leq \rho\}$ . Тогда  $x^{k+1} = \begin{cases} x^k - \gamma \nabla f(x^k), & \text{если } \|x^k - \gamma \nabla f(x^k)\| \leq \rho, \\ \rho \frac{x^k - \gamma \nabla f(x^k)}{\|x^k - \gamma \nabla f(x^k)\|}, & \text{если } \|x^k - \gamma \nabla f(x^k)\| > \rho. \end{cases}$

Наконец, пусть  $Q$  — линейное многообразие,  $Q = \{x \in \mathbf{R}^n: Cx = d\}$ , где  $C$  — матрица  $m \times n$ ,  $d \in \mathbf{R}^m$ .

Тогда

$$x^{k+1} = (I - C^+ C)(x^k - \gamma \nabla f(x^k)) + C^+ d. \quad (6)$$

Здесь  $C^+$  — псевдообратная матрица к  $C$  (§ 1 гл. 6). Если при этом  $x^0 \in Q$ , то  $x^{k+1} - x^0 = T(x^k - x^0 - \gamma \nabla f(x^k))$ ,

$$\text{где } T = I - C^+ C, \text{ а если } C \text{ — матрица ранга } m < n, \text{ то } T = I - C^T (C C^T)^{-1} C.$$

**2. Метод проекции субградиента.** Аналогом субградиентного метода безусловной минимизации негладких функций является метод проекции субградиента

$$x^{k+1} = P_Q(x^k - \gamma \partial f(x^k)),$$

(8)

где, как и ранее,  $\partial f(x^k)$  — любой из субградиентов выпуклой функции  $f(x)$  в точке  $x^k$ . Правила выбора  $y_k$  аналогичны рассмотренным в § 3 гл. 5, и мы упомянем лишь два важнейших:

$$\psi_k \rightarrow 0, \quad \sum_{k=0}^{\infty} \psi_k = \infty, \quad (9)$$

$$\psi_k = \frac{f(x^k) - f^*}{\|\partial f(x^k)\|^2}, \quad f^* = \min_{x \in Q} f(x). \quad (10)$$

**Теорема 2.** Пусть  $f(x)$  — выпуклая функция на  $\mathbf{R}^n$ ,  $Q$  — выпуклое замкнутое множество, множество  $X^*$  точек минимума  $f(x)$  на  $Q$  непусто. Тогда метод (8), (10) сходится к  $x^* \in X^*$ , если  $\|\partial f(x)\| \leq c$  для всех  $x \in Q$ , то и метод (8), (9) сходится к  $x^* \in X^*$ .  $\blacktriangle$

**3. Метод условного градиента.** Вспомним, что в основе градиентного метода лежала идея линеаризации функции. Можно попытаться применить эту же идею для задачи с ограничениями: в очередной точке  $x^k$  линеаризуем функцию  $f(x)$ , затем решим задачу минимизации линейной функции на  $Q$  и найденную точку используем для выбора направления движения. Из этих соображений приходим к *методу условного градиента*:

$$\tilde{x}_k = \operatorname{argmin}_{x \in Q} (\nabla f(x^k), x), \quad (11)$$

$$x^{k+1} = x^k + \psi_k (\tilde{x}_k - x^k).$$

При этом, во-первых, предполагается, что задача минимизации линейной функции на  $Q$  имеет решение (для чего естественно потребовать ограниченность  $Q$ ), во-вторых, что это решение может быть найдено достаточно просто, лучше всего — в явной форме (см. примеры в упражнении 7) и, в-третьих, нужно указать правило выбора  $\psi_k$ ,  $0 \leq \psi_k \leq 1$ .

**Теорема 3.** Пусть  $f(x)$  — дифференцируемая функция, гравиент которой на  $Q$  удовлетворяет условию Липшица с константой  $L$ , а  $Q$  выпукло, замкнуто и ограничено. Пусть  $\psi_k$  определяется из условия скорейшего спуска:

$$\psi_k = \operatorname{argmin}_{0 \leq \psi \leq 1} f(x^k + \psi (\tilde{x}_k - x^k)). \quad (12)$$

Тогда

а)  $(\nabla f(x^k), x^k - \tilde{x}_k) \rightarrow 0$  и для всякой предельной точки последовательности  $x^k$  удовлетворяется необходимое условие экстремума (1) § 1;

б) если  $f(x)$  выпукла, то предельные точки — точки минимума  $f(x)$  на  $Q$  и справедлива оценка

$$f(x^k) - f^* = O(1/k), \quad f^* = \min_{x \in Q} f(x),$$

$$f(x^k) \geq f^* \geq f(x^k) + (\nabla f(x^k), \tilde{x}_k - x^k),$$

$$(13)$$

в) если задача имеет острый минимум, то метод (11), (12) конечен.

**Доказательство.** Прежде всего, метод определен, так как при сделанных предположениях точка  $\tilde{x}_k$  существует. Введем  $V(x) = f(x) - f^*, s^k = x^k - \tilde{x}_k$ . Тогда в силу условия Липшица на  $\nabla V(x)$  (см. (15) § 1 гл. 1)

$$V(x^{k+1}) = \min_{0 \leq \psi \leq 1} V(x^k - \psi s^k) \leq \min_{0 \leq \psi \leq 1} \Phi(\psi),$$

$$\Phi(\psi) = V(x^k) - \psi (\nabla f(x^k), s^k) + \frac{\psi^2 L \|s^k\|^2}{2}.$$

Обозначим  $\psi_k^* = \frac{(\nabla f(x^k), s^k)}{L \|s^k\|^2}$ . По определению  $\tilde{x}_k$ :  $(\nabla f(x^k), s^k) \geq 0$ , т. е.  $\psi_k^* \geq 0$ . Возможны два случая: 1)  $\psi_k^* \leq 1$ , 2)  $\psi_k^* > 1$ . В первом

$$V(x^{k+1}) \leq \Phi(\psi_k^*) = V(x^k) - \frac{(\nabla f(x^k), s^k)^2}{2L \|s^k\|^2} \leq V(x^k) - \frac{(\nabla f(x^k), s^k)^2}{2LR^2}, \quad (14)$$

$$\text{где } R \text{ — диаметр множества } Q. \text{ Во втором } L \|s^k\|^2 < (\nabla f(x^k), s^k) \text{ и}$$

$$V(x^{k+1}) \leq \Phi(1) \leq V(x^k) - \frac{(\nabla f(x^k), s^k)^2}{2L \|s^k\|^2} \leq V(x^k) - (\nabla f(x^k), s^k)/2. \quad (15)$$

Итак, в обоих случаях  $V(x^k)$  монотонно убывает, а так как  $V(x) \geq 0$ , то  $V(x^{k+1}) \rightarrow 0$ . В соответствии с (14) и (15) это влечет сходимость  $(\nabla f(x^k), s^k) \rightarrow 0$ .

Пусть теперь  $x^*$  — произвольная предельная точка последовательности  $x^k$  (она заведомо существует, так как  $Q$  ограничено),  $x^k \rightarrow x^*$ . Тогда для любого  $x \in Q$

$$(\nabla f(x^*), x - x^*) = (\nabla f(x^*) - \nabla f(x^k), x - x^*) + (\nabla f(x^k), x - \tilde{x}_k) + (\nabla f(x^k), \tilde{x}_k - x^k) + (\nabla f(x^k), x^k - x^*).$$

Первый и четвертый члены правой части стремятся к 0 при  $i \rightarrow \infty$ , так как  $x^k \rightarrow x^*$ , второй неотрицателен в силу определения  $\tilde{x}_k$ , а третий стремится к 0 по доказанному. Поэтому  $(\nabla f(x^*), x - x^*) \geq 0$ , т. е. выполнено условие (1) § 1.

Пусть  $f(x)$  выпукла, тогда  $x^*$  — точка минимума,  $V(x^*) \rightarrow 0$  и  $V(x_k) \leq (\nabla f(x_k), x^k - x^*) \leq (\nabla f(x_k), x_k - \tilde{x}_k)$ , т. е.  $V(x_k) \leq (\nabla f(x_k), s^k)$ . С другой стороны, из (14) и (15) получаем  $(\nabla f(x_k), s^k) \leq \max \{(2LR^2(V(x^k) - V(x^{k+1}))^{1/2}, 2(V(x^k) - V(x^{k+1}))\} \leq 2LR^2(V(x^k) - V(x^{k+1}))^{1/2}$  для достаточно больших  $k$ , так как  $V(x^k) \rightarrow 0$ . Отсюда  $V(x^{k+1}) \leq V(x^k) - (2LR^2)^{-1} V(x^k)^2$ . Используя лемму 6 § 2 гл. 2, получаем оценку (13).

Наконец, пусть  $f(x)$  имеет острый минимум на  $Q$  в точке  $x^*$ . Тогда для произвольного  $x \in Q$

$$(\nabla f(x^*), x^* - x) = (\nabla f(x^*), x^* - x) + (\nabla f(x^*), x^* - x) \leq -\alpha \|x - x^*\| + L \|x^k - x^*\| \|x - x^*\| \leq 0$$

для  $x^k$ , достаточно близких к  $x^*$ . Поэтому для таких  $x^k$  будет  $\tilde{x}^k = x^*$  (в соответствии с определением  $\tilde{x}^k$ ). Поскольку  $x^* - \tilde{x}^{k+1} = x^*$ .  $\blacktriangle$

Покажем на примере, что оценка (13) не может быть улучшена даже для сильно выпуклых  $f(x)$ . Пусть  $x \in \mathbb{R}^2$ ,

$$f(x) = x_1^2 + (1 + x_2)^2, \quad Q = \{x: |x_1| \leq 1, 0 \leq x_2 \leq 1\} \quad (16)$$

(рис. 32). Тогда  $x^* = \{0, 0\}$ , если  $x^k < 0$  и  $\tilde{x}^k = \{-1, 0\}$ , если  $x^k > 0$ . При этом для всех  $k$  (14) обращается в равенство, и мы получаем  $v_{k+1} = v_k - (1/4) \|s_k\|^2 v_k^2, \|s_k\|^2 \rightarrow 1$ , т. е.

$$v_k = 4/k + o(1/k), \quad \text{где } v_k = f(x^k) - f^* = \|x^k - x^*\|^2.$$

Эта ситуация является типичной: если  $Q$  — многоугольник, а минимум гладкой функции  $f(x)$  достигается не в вершине  $Q$ , то скорость сходимости столь же низка. Это не удивительно, так как в качестве  $\tilde{x}^k$  могут выступать лишь вершины  $Q$ , поэтому направления движения  $\tilde{x}^k - x^k$  сильно отличаются от направления на минимум  $x^* - x^k$ .

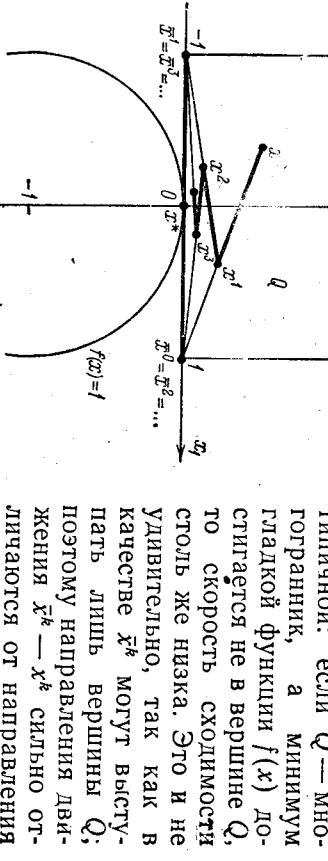


Рис. 32. Медленная сходимость метода условного градиента.

Мы видели, если в задаче имеется острый минимум, то метод условного градиента сходится не только от свойств  $f(x)$  и  $Q$  (гладкость, выпуклость, сильная выпуклость и т. п.), но и от структуры решения.

Параметр  $\gamma_k$  в методе условного градиента можно выбирать иначе, чем в (12) (упр. 9). Однако наиболее простой способ  $\gamma_{k+1} = \tilde{x}^k$  (т. е.  $\gamma_k = 1$ ) не годится.

Далее, метод условного градиента не обобщается на негладкие задачи. Дело заключается в том, что точка минимума  $f(x)$  на  $Q$  не является неподвижной точкой метода вида (11), в котором градиент заменен на произвольный субградиент.

#### Упражнения.

6. Когда существует решение задачи а)  $\min(c, x), x \geq 0, x \in \mathbb{R}^n$ ; б)  $\min(c, x), Ax = b, x \in \mathbb{R}^n$ ; в)  $\min(c, x), (Ax, x) \leq \beta, \beta > 0, A \geq 0$ ? Ответ. а) Если  $c \geq 0$ ; б) если  $A^T c = 0$ ; в) если  $(c, e_i) = 0$  для всех собственных векторов  $e_i$  матрицы  $A$ , отвечающих нулевым собственным значениям.

7. Убедитесь в правильности решения следующих элементарных задач минимизации:

$$\text{а)} \quad x^* = \arg \min_{a \leq x \leq b} (c, x), \quad x_i^* = \begin{cases} a_i, & \text{если } c_i > 0, \\ b_i, & \text{если } c_i < 0, \end{cases} \quad \text{любому между } a_i \text{ и } b_i, \text{ если } c_i = 0;$$

$$\text{б)} \quad x^* = \arg \min_{\|x\| \leq \rho} (c, x) = -\frac{c\rho}{\|c\|};$$

в)  $x^* = \arg \min_{(1/2)(Ax, x) - (b, x) \leq a} (c, x), \quad A > 0, \quad a > 0, \quad x^* = A^{-1}(b - Ac)$ , где  $\lambda$  находится из условия  $(Ax^*, x^*)/2 - (b, x^*) = a$ .

8. Пусть  $x^*$  — решение задачи  $\min(c, x), x \in Q$ . Докажите, что  $x^* = \lim P_Q(\lambda c)$  при  $\lambda \rightarrow \infty$ .

9. Докажите, что все утверждения теоремы 3 остаются в силе, если шаг выбирает из условия  $y_k = \min\{1, (\nabla f(x^k), x^k - \tilde{x}^k)/L\} \|x^k - \tilde{x}^k\|_2^2$ .

10. Назовем множество  $Q$  сильно выпуклым, если найдется  $\beta > 0$  такое, что если  $x \in Q, y \in Q$ , то  $z \in Q$  при  $\|z - (x + y)/2\| \leq \beta \|x - y\|^2$ . Покажите, что если  $f(x)$  — сильно выпуклая функция на  $\mathbb{R}^n$ , то множество  $Q_\alpha = \{x: f(x) \leq \alpha\}$  сильно выпукло. Докажите, что сильно выпуклое множество, отличное от  $\mathbb{R}^n$ , ограничено.

11. Докажите, что если  $f(x)$  выпуклая и  $\|\nabla f(x)\| \geq \epsilon > 0$  для  $x \in Q$ , а  $Q$  сильно выпукло, то метод условного градиента в условиях теоремы 3 сходится со скоростью геометрической прогрессии. Докажите, что если  $f(x)$  выпуклая, то  $\Psi(x) \leq \frac{y \in Q}{f(x)}$  для всех  $x \in Q$ , причем равенство достигается тогда и только тогда, когда  $x = \arg \min_{x \in Q} f(x)$ . Попробуйте исключить, что если  $f(x)$  выпуклая, дифференцируемость и т. п.).

Положите, как можно интерпретировать метод условного градиента в терминах функции  $\Psi(x)$ .

4. Метод Ньютона. Для построения метода Ньютона в задаче (A) можно использовать ту же идею квадратичной аппроксимации  $f(x)$ , что и для безусловного минимума. Разница лишь в том, что нужно отыскивать минимум аппроксимации не на всем пространстве, а на множестве  $Q$ . Эти соображения приводят к методу

$$x^{k+1} = \arg \min_{x \in Q} f_k(x),$$

$$f_k(x) = f(x^k) + (\nabla f(x^k), x - x^k) + (\nabla^2 f(x^k)(x - x^k), x - x^k)/2. \quad (17)$$

Теорема 4. Пусть  $f(x)$  достигает минимума на выпуклом замкнутом множестве  $Q$  в точке  $x^*$ , причем в окрестности  $x^*$   $f(x)$  однажды дифференцируема на  $Q$ ,  $\nabla^2 f(x)$  удовлетворяет условию Липшица и

$$\nabla^2 f(x^*) > 0.$$

Тогда метод (17) локально сходится к  $x^*$  с квадратичной скоростью.

Доказательство. В точке  $x^{k+1}$  удовлетворяется необходимое условие минимума  $f_k(x)$  на  $Q$ , т. е.

$$(\nabla f_k(x^{k+1}), x - x^{k+1}) = (\nabla f(x^k) + \nabla^2 f(x^k)(x^{k+1} - x^k), x - x^{k+1}) \geq 0$$

для всех  $x \in Q$ , в частности для  $x = x^*$ . Поэтому

$$0 \leq (\nabla f(x^k) + \nabla^2 f(x^k)(x^{k+1} - x^k), x^* - x^{k+1}) = (\nabla f(x^*), x^* - x^{k+1}) +$$

$$+ (\nabla f(x^k) - \nabla f(x^*), x^{k+1} - x^k), x^* - x^{k+1}).$$

Первый член неположителен в силу § 1. Для  $\nabla f(x^k) - \nabla f(x^*) \leq L(x^k - x^*)$  имеем оценку  $\nabla f(x^k) - \nabla f(x^*) = \nabla^2 f(x^k)(x^k - x^*) + r$ ,  $\|r\| \leq (L/2)\|x^k - x^*\|^2$  (см. (15) § 1 гл. 1). Поэтому

$$0 \leq (\nabla^2 f(x^k)(x^k - x^*), x^* - x^{k+1}) \leq$$

$$\leq -l\|x^{k+1} - x^*\|^2 + (L/2)\|x^k - x^*\|^2 \|x^{k+1} - x^*\|.$$

Здесь мы воспользовались тем, что  $\nabla^2 f(x^k) \geq II$ ,  $l > 0$ , для всех  $x^k$ , достаточно близких к  $x^*$ , в силу (18) и условия Липшица на гессиан. Отсюда либо  $x^{k+1} = x^*$ , либо

$$(19)$$

$\|x^{k+1} - x^*\| \leq L\|x^k - x^*\|^2/(2l)$ .

Если  $L\|x^0 - x^*\|/(2l) < 1$ , то из (19) следует, что и все  $x^k$  оста-

нутся в той же окрестности  $x^*$ , а оценка (19) означает квадра-

тическую скорость сходимости. ▲

Для случая острого минимума нетрудно доказать конечность метода. Однако в этой ситуации вряд ли имеет смысл использовать метод Ньютона, так как гораздо более простые методы (проекции градиента и условного градиента) также обладают свойством конечности.

Метод Ньютона можно применить лишь в случае, если задача минимизации квадратичной функции на  $Q$  легко решается. Если  $Q$  — многогранник, то (17) является общей задачей квадратичного программирования. Как мы увидим в гл. 10, для нее существуют конечные алгоритмы решения. Для частного случая, когда  $Q$  — параллелепипед, задача (17) может быть решена с помощью модификации метода сопряженных градиентов, описанного в следующем параграфе. Наконец, в простейших случаях, когда  $Q$  — шар или линейное многообразие, (17) имеет достаточно простое решение.

#### Упражнения.

13. Покажите, что решением задачи  $\min[(Ax, x)/2 - (b, x)]$ ,  $A > 0$ ,  $\|x\| \leq \rho$  является точка  $(A + \lambda I)^{-1}b$ , где  $\lambda = 0$ , если  $\|A^{-1}b\| \leq \rho$ , и  $\lambda$  на-ходитя из уравнения  $\|(A + \lambda I)^{-1}b\| = \rho$  в противном случае.

14. Рассмотрите модификацию метода Ньютона, аналогичные описанным в § 1 гл. 3 для безусловной минимизации, и докажите их глобальную сходимость.

### § 3. Другие методы

1. **Квазиньютоновские методы.** Отметим, что все методы расширения гладких задач, описанные в предыдущем параграфе, могут быть получены с помощью общей схемы. Пусть

$$x^{k+1} = \underset{x \in Q}{\operatorname{argmin}} \left( (\nabla f(x^k), x - x^k) + \frac{1}{2}(H_k(x - x^k), x - x^k) \right), \quad (1)$$

где  $H_k \geq 0$  — некоторая матрица.

Очевидно, что при  $H_k = \nabla^2 f(x^k)$  метод (1) переходит в метод Ньютона, а при  $H_k = \gamma^{-1}I$  — в метод проекции градиента (так как последний может быть записан в виде  $x^{k+1} = \arg \min_{x \in Q} \|x - (x^k - \gamma \nabla f(x^k))\|^2$ ). Можно еще несколько расширить класс методов (1), введя в них процедуру одномерной ми-

$$\begin{aligned} \bar{x}^k &= \underset{x \in Q}{\operatorname{argmin}} \left( (\nabla f(x^k), x - x^k) + \frac{1}{2}(H_k(x - x^k), x - x^k) \right), \\ x^{k+1} &= x^k + \gamma_k s^k, \quad s^k = \bar{x}^k - x^k, \quad \gamma_k = \underset{0 \leq y \leq 1}{\operatorname{argmin}} f(x^k + ys^k). \end{aligned} \quad (2)$$

В частности, при  $H_k = 0$  получаем из (2) метод условного градиента. Сходимость подобных методов требует специального анализа. Нельзя, например, воспользоваться результатами о сходимости методов безусловной минимизации типа  $x^{k+1} = x^k - \gamma_k H_k \nabla f(x^k)$  с произвольной  $H_k \geq 0$  (лемма 1 § 3 гл. 3) по-добно тому, как это было сделано в § 2 при доказательстве теоремы 1. Дело в том, что в лемме 1 § 3 гл. 3 в качестве функции Липунова выбирается не расстояние до минимума, а  $f(x) - f(x^*)$ , поэтому нельзя утверждать релаксационность оператора проектирования (упр. 1). Однако можно вести доказательство по той же схеме, что и для теоремы 4 § 2. Приведем характерный результат.

Теорема 1. Пусть  $f(x)$  дважды дифференцируема и  $II \leq \nabla^2 f(x) \leq LI$ ,  $l > 0$ , для всех  $x \in Q$ ,  $Q$  выпукло, замкнуто и  $\|H_k - \nabla^2 f(x^k)\| \leq \varepsilon < l/2$ .

Тогда в методе (1)  $x^k$  локально сходится к  $x^* = \underset{x \in Q}{\operatorname{argmin}} f(x)$  со скоростью геометрической прогрессии, а если

$$\|H_k - \nabla^2 f(x^k)\| \rightarrow 0, \quad (4)$$

то скорость сходимости сверхлинейна.

Доказательство. Из определения  $x^{k+1}$

$$0 \leq (\nabla f(x^k) + H_k(x^{k+1} - x^k), x^* - x^{k+1}) \leq$$

$$\leq (\nabla f(x^k) - \nabla f(x^*), x^{k+1} - x^k) - (l - \varepsilon)\|x^{k+1} - x^*\|^2 +$$

$$+ (l/2)\|x^k - x^*\|^2 \|x^{k+1} - x^*\|,$$

$$\text{Но } \nabla f(x^k) - \nabla f(x^*) = \nabla^2 f(x^k)(x^k - x^*) + r, \quad \|r\| \leq \frac{1}{2}L\|x^k - x^*\|^2,$$

поэтому

$$0 \leq ((\nabla^2 f(x^k) - H_k)(x^k - x^*) + H_k(x^{k+1} - x^*) + r, x^* - x^{k+1}) \leq$$

$$\leq \varepsilon \|x^k - x^*\| \|x^{k+1} - x^*\| - (l - \varepsilon)\|x^{k+1} - x^*\|^2 +$$

$$+ (l/2)\|x^k - x^*\|^2 \|x^{k+1} - x^*\|,$$

$$\|x^{k+1} - x^*\| \leq \frac{\varepsilon + (l/2)\|x^k - x^*\|}{l - \varepsilon} \|x^k - x^*\|.$$

Отсюда следует, что если  $(L/2)\|x^0 - x^*\| < l - 2\epsilon$ , то  $x^k \rightarrow x^*$  со скоростью геометрической прогрессии, знаменатель которой тем меньше, чем меньше  $\epsilon$ . Аналогичным образом при условии (4) получаем  $\|x_{k+1} - x^*\| \leq q_k \|x_k - x^*\|$ ,  $q_k \rightarrow 0$  при  $k \rightarrow \infty$ , что и означает сверхлинейную сходимость. ▲

Теорема 1 показывает, что в задачах, где вычисление матрицы  $\nabla^2 f(x^k)$  невозможно или слишком трудоемко, желательно в качестве  $H_k$  брать аппроксимацию к гессиану. Это можно сделать так же, как в квазиньютоновских методах безусловной минимизации, используя информацию, полученную на предыдущих итерациях. Именно, если доступны градиенты в предыдущих точках, то можно восстановить  $H$  из приближенных равенств

$$\nabla f(x^{i+1}) - \nabla f(x^i) \approx H(x^{i+1} - x^i), \quad i = k, \dots, k-n+1, \quad (5)$$

если только  $x^i$  не лежат в одном подпространстве.

Мы не будем подробно останавливаться на подобных методах, так как они в основном используют ту же технику, что и для безусловной минимизации. Разница здесь лишь в том, что в задаче с ограничениями вектор  $\nabla f(x^{i+1})$ , вообще говоря, не ортогонален направлению движения  $x^{i+1} - x^i$ .

#### Упражнения.

1. Приведите пример функции  $f(x)$  и матрицы  $H \geq 0$ , для которых метод  $x^{k+1} = x^k - H \nabla f(x^k)$  сходится, но  $\|x^k - x^*\|$  не убывает монотонно ( $x^* -$  точка минимума  $f(x)$ ). С его помощью постройте пример расходимости метода (1) с  $H_k \equiv H$ .

2. Докажите результат о глобальной сходимости метода (1) для случая, когда матрица  $H_k$  достаточно близка к  $Y^{-1}$ ,  $0 < Y < 2L$ .

3. Покажите, что метод  $x^{k+1} = P_Q(x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k))$ , вообще говоря, не сходится. В частности, если  $f(x)$  квадратична, то при любом  $x^0$  метод останавливается в точке  $x^1$  (одинаковой для всех  $x^0$ ), вообще говоря, не являющейся решением.

**2. Метод сопряженных градиентов.** Начнем со случая, когда  $f(x) — квадратичная функция, а  $Q — подпространство в  $\mathbf{R}^n$ ,  $Q = \{x: Cx = 0\}$ ,  $C — матрица  $m \times n$  ранга  $m$ . Как мы знаем, проекция вектора на это подпространство задается формулой  $P_Q(x) = (I - C^T C)x = (I - C^T(CC^T)^{-1}C)x$ .$$$

Запишем метод сопряженных градиентов, в котором вектор  $\nabla f(x)$  заменен на его проекцию на  $Q$ :

$$x^{k+1} = x^k + \alpha_k p^k, \quad \alpha_k = \underset{a}{\operatorname{argmin}}_f(x^k + ap^k), \quad x^0 \in Q,$$

$$p^k = -P_Q \nabla f(x^k) + \beta_k p^{k-1}, \quad \beta_k = \underset{x^k + ap^k \geq 0}{\operatorname{argmin}}_f(x^k + ap^k). \quad (6)$$

Можно доказать (упр. 4), что если  $f(x) — квадратичная функция  $f(x) = (Ax, x)/2 — (b, x)$ , причем  $(Ax, x) \geq \alpha \|x\|^2$ ,  $\alpha > 0$ , при всех  $x \in Q$ , то метод (6) заканчивается не более чем за  $n-m$  шагов.$

Таким образом, метод сопряженных градиентов остается конечным при минимизации квадратичной функции на подпространстве, причем число шагов тем меньше, чем больше ограничений. Правда, каждая итерация метода связана с дополнительными вычислениями — проектированием на подпространство.

Пусть теперь  $Q — положительный ортант в  $\mathbf{R}^n$ , т. е.  $Q = \{x: x \geq 0\}$ , а  $f(x) — по-прежнему квадратичная функция$ . Тогда ее минимизацию на  $Q$  можно свести к последовательной минимизации на граниях  $Q$ . Эти грани имеют вид  $\{x_i = 0, i \in I, x_i > 0, i \in J\}$ , где  $I — некоторый набор индексов из множества  $1, \dots, n$ . Минимизация на подпространстве  $L = \{x: x_i = 0, i \in J\}$  проводится просто — надо проводить вычисления, как в методе сопряженных градиентов, заменяя нулевыми компоненты из множества  $J$  как у векторов  $x^k$ , так и у градиентов  $\nabla f(x^k)$  (см. (6) и упр. 5). С учетом этих соображений приходим к методу минимизации  $f(x)$  на  $Q$ , который мы запишем в координатной форме:$$

$$x^{k+1} = x^k + \alpha_k p^k, \quad \alpha_k = \underset{\substack{\alpha \geq 0 \\ x^k + \alpha p^k \geq 0}}{\operatorname{argmin}}_f(x^k + \alpha p^k), \quad (7)$$

$$\beta_k = \begin{cases} \sum_{i \in I_k} (\nabla f(x^k)_i)^2 / \sum_{i \in I_k} (\nabla f(x^{k-1})_i)^2, & \text{если } I_k = I_{k-1}, \\ 0, & \text{если } k = 0 \text{ или } I_k \neq I_{k-1}, \\ \{i: x_i^k = 0, \nabla f(x^k)_i > 0\}, & \text{для всех } i \in I_{k-1}, \\ I_{k-1} \cup \{i: x_i^k = 0\} & \text{в остальных случаях.} \end{cases}$$

Иными словами, ведется процесс минимизации методом сопряженных градиентов  $f(x)$  на множестве  $L_k = \{x: x_i = 0, i \in I_k, x_i > 0, i \in J_k\}$ . Он заканчивается либо когда одна из компонент (не принадлежащих  $I_k$ ) вектора  $x^k$  обращается в 0 (тогда в множестве  $I_k$  добавляется номер этой компоненты), либо когда найден минимум на  $L_k$  (в этом случае «обновляется» множество  $I_k$ ). Можно показать, что если  $f(x) = (Ax, x)/2 — (b, x)$ ,  $A \geq 0$ , то описанный метод конечен.

Таким образом, мы получили конечный метод решения задачи минимизации квадратичной функции при ограничениях  $x \geq 0$ . Можно придумать и другие конечные варианты метода сопряженных градиентов, а также обобщить его на случай, когда ограничения имеют вид  $a \leq x \leq b$ .

Для минимизации неквадратичной  $f(x)$  на ортанте или параллелепипеде можно использовать те же идеи. При этом нужно

регулировать точность решения задачи минимизации на грани. Разумеется, такие методы не будут конечными.

#### Упражнения.

4. Пусть  $Q$  — подпространство в  $\mathbf{R}^n$ ,  $f(x)$  — дифференцируемая функция на  $\mathbf{R}^n$ . Рассмотрим  $f_Q(x)$  — ее сужение на  $Q$ . Тогда градиент  $\nabla f_Q(x)$  в точке  $x \in Q$  определяется равенством  $\hat{f}_Q(x+y) = \hat{f}_Q(x) + (\nabla \hat{f}_Q(x)y) + o(y)$  для всех  $y \in Q$ , где  $\nabla \hat{f}_Q(x) \in Q$ . Докажите, что (6) представляет собой метод сопряженных градиентов для безусловной минимизации  $f_Q(x)$ . Отсюда следует, что если градиента квадратична, то метод, конечно.

5. Покажите, что если  $Q = \{x: x_i = 0, i \in I\}$ , то  $P_Q(x)_i = 0$ , если  $i \in I$ , и  $P_Q(x)_i = x_i$ , если  $i \notin I$ .

**3. Минимизация негладких функций.** При описании методов безусловной минимизации выпуклых негладких функций в § 4 гл. 5 мы предполагали, что задана область локализации минимума. Если в качестве такой области взять множество  $Q$ , то окажется, что все эти методы пригодны и для задач с ограничениями. Таким образом, методы отсекающей гиперплоскости, чебышевских центров, центра тяжести и т. д. дословно переносятся на задачи (A). При этом в них на каждом шаге решается задача минимизации линейной или квадратичной функции на множестве  $Q_k$ , которое задается с помощью условия  $x \in Q$  и некоторых дополнительных линейных ограничений. Если  $Q$  — многоугольник, то получаем задачу линейного или квадратичного программирования, которая может быть решена стандартными методами. Все результаты о сходимости и скорости сходимости, приведенные в § 4 гл. 5, остаются справедливыми и для задач с ограничениями. Отметим, что наличие острого минимума в негладком случае не приводит, вообще говоря, к конечности методов.

#### § 4. Влияние помех

Не будем рассматривать здесь все ситуации столь же подробно, как в задачах безусловной минимизации (гл. 4). Нас будут интересовать главным образом новые эффекты, связанные с наличием ограничений.

**1. Абсолютные детерминированные помехи.** Пусть вместо градиента  $\nabla f(x^k)$  (или субградиента  $\partial f(x^k)$ ) известны лишь их приближенные значения  $\tilde{\nabla}f(x^k)$  ( $\tilde{\partial}f(x^k)$ ), причем

$$\|\tilde{\nabla}f(x^k) - \nabla f(x^k)\| \leq \varepsilon \quad (\|\tilde{\partial}f(x^k) - \tilde{\partial}f(x^k)\| \leq \varepsilon).$$

Пусть мы применяем методы § 2 в этой ситуации, т. е. в методах  $\nabla f(x^k)$  и  $\tilde{\nabla}f(x^k)$  заменены на  $\nabla f(x^k)$  и  $\tilde{\nabla}f(x^k)$ . При этом, вообще говоря, методы проекции градиента и проекции субградиента перестают сходиться, а приводят к попаданию в некоторую окрестность минимума, размеры которой зависят от  $\varepsilon$ . Несколько иначе обстоит дело с методом условного градиента.

Прежде всего, он включает операцию одномерной минимизации, которая не может быть выполнена точно. Кроме того, точка  $\tilde{x}^k$  может сильно измениться при замене  $\nabla f(x^k)$  на  $\tilde{\nabla}f(x^k)$ . Поэтому метод условного градиента вряд ли целесообразен для задач с помехами.

Новая ситуация возникает в случае острого минимума. Теорема 1. Пусть  $x^*$  — точка острого минимума дифференцируемой выпуклой функции  $f(x)$  на выпуклом множестве  $Q$ . Предположим, что проектирование на  $Q$  осуществляется точно. Тогда при условиях теоремы 1 § 2 метод проекции градиента остается конечным, если в нем  $\nabla f(x^k)$  заменен на  $\tilde{\nabla}f(x^k)$ , а  $\varepsilon > 0$  достаточно мало.

Доказательство проводится так же, как в теореме 1 § 2. Таким образом, задачи с гладкой  $f(x)$  и острым минимумом обладают свойством сверхустойчивости некоторых методов — последние дают точное решение даже при наличии абсолютных (но достаточно малых) помех.

**2. Абсолютные случайные помехи.** Пусть помехи  $\xi^k = \nabla f(x^k) - \tilde{\nabla}f(x^k)$  ( $\xi^k = \partial f(x^k) - \tilde{\partial}f(x^k)$ ) случайны, независимы, центрированы и имеют ограниченную дисперсию:

$$\mathbf{M}\xi^k = 0, \quad \mathbf{M}\|\xi^k\|^2 \leq \sigma^2. \quad (3)$$

**Теорема 2.** Пусть  $f(x)$  — выпуклая функция на  $\mathbf{R}^n$ ,  $Q$  — ограниченное замкнутое выпуклое множество. Тогда в методе

$$x^{k+1} = P_Q(x^k - y_k \tilde{\nabla}f(x^k)), \quad \sum_{k=0}^{\infty} y_k^2 < \infty, \quad \sum_{k=0}^{\infty} y_k = \infty, \quad (4)$$

при выполнении условий (2), (3) будет  $x^k \rightarrow x^*$  (п. н.), где  $x^*$  — некоторая точка минимума  $f(x)$  на  $Q$ . Если  $f(x)$  сильно выпукла, то  $\mathbf{M}\|x^k - x^*\|^2 \rightarrow 0$  (условие  $\sum_{k=0}^{\infty} y_k^2 < \infty$  можно при этом заменить на  $y_k \rightarrow 0$ ), а если  $y_k = y/k$  и  $y$  достаточно велико, то

$$\mathbf{M}\|x^k - x^*\|^2 = O(1/k). \blacksquare$$

Для задач с острым минимумом, по-видимому, нет необходимости устремлять  $y_k$  к 0. Можно предположить, что при привильной регулировке длины шага метод проекции градиента при наличии случайных помех будет почти наверное конечным в случае острого минимума. Однако вопрос этот не исследован.

Для метода условного градиента на первый взгляд кажется естественным поступить так же, как в методе (4), т. е. заменить

точное значение градиента на приближенное и устремить шаг к 0:

$$\bar{x}^k = \underset{x \in Q}{\operatorname{argmin}} (\tilde{\nabla} f(x^k), x),$$

$$x^{k+1} = x^k + v_k(\bar{x}^k - x^k), \quad v_k \rightarrow 0, \quad \sum_{k=0}^{\infty} v_k = \infty. \quad (5)$$

Однако такой метод, как правило, не сходится. Пусть, например, ищется минимум гладкой функции  $f(x)$ ,  $x \in \mathbf{R}^1$ , на отрезке  $Q = [-\alpha, \beta]$ ,  $\alpha > 0$ ,  $\beta > 0$ , а минимум достигается в  $x^* = 0 \in Q$ . Тогда при  $x^k = x^*$  будет  $\tilde{\nabla} f(x^k) = 0$  и  $\bar{x}^k = -\alpha$ , если  $\xi_k > 0$ ,  $\bar{x}^k = \beta$ , если  $\xi_k > 0$ . Для симметрично распределенной помехи  $\mathbf{M}(\bar{x}^k - x^k) = (\beta - \alpha)/2 \neq 0$  при  $\beta \neq \alpha$ . Итак, в точке минимума  $f(x)$  среднее значение направления движения отлично от 0, поэтому метод не может сходиться к этой точке.

Сходимости в методе условного градиента можно добиться, если ввести процедуру усреднения градиентов:

$$\bar{x}^k = \underset{x \in Q}{\operatorname{argmin}} (y^k, x),$$

$$y^k = y^{k-1} + \mu_k (\tilde{\nabla} f(x^k) - y^{k-1}), \quad \mu_k \rightarrow 0, \quad \sum_{k=0}^{\infty} \mu_k = \infty, \quad (6)$$

$$x^{k+1} = x^k + v_k(\bar{x}^k - x^k), \quad v_k \rightarrow 0, \quad \sum_{k=0}^{\infty} v_k = \infty.$$

Здесь  $y^k$  — усредненное по предыдущим итерациям значение градиента.

**3. Относительные помехи.** Пусть помехи удовлетворяют условию

$$\|\nabla f(x) - \tilde{\nabla} f(x)\| \leq \alpha \|\nabla f(x)\|. \quad (7)$$

Мы видели (теорема 2 § 2 гл. 4), что градиентный метод устойчив к таким помехам, если их уровень меньше 100% (т. е.  $\alpha < 1$ ). В задачах с ограничениями дело обстоит не так: поскольку в точке минимума  $x^*$ , вообще говоря,  $\nabla f(x^*) \neq 0$ , то величина  $\|\tilde{\nabla} f(x) - \nabla f(x)\|$  не обязана стремиться к 0 при приближении  $x$  к  $x^*$ . Поэтому здесь ситуация с абсолютными и относительными помехами мало отличается, и, например, нельзя гарантировать сходимость метода проекции градиента при деградированных относительных помехах любого уровня.

Действительным аналогом относительных ошибок для задач с ограничениями являются условия типа

$$\|\tilde{\nabla} f(x^k) - \tilde{\nabla} f(x^k)\| \leq \alpha \|x^k - x^*\|, \quad (8)$$

$$\|\tilde{\nabla} f(x^k) - \tilde{\nabla} f(x^k)\| \leq \alpha \|x^{k+1} - x^k\|. \quad (9)$$

Однако такие условия выглядят довольно искусственно, и мы не будем заниматься их анализом.

## Г л а в а 8 ЗАДАЧИ С ОГРАНИЧЕНИЯМИ ТИПА РАВЕНСТВ

Данная глава посвящена задаче вида

$$\min f(x), \quad x \in \mathbf{R}^n, \quad g_i(x) = 0, \quad i = 1, \dots, m, \quad (A)$$

где  $f$ ,  $g_i$  — гладкие функции. Эта задача, являющаяся частным случаем общей задачи математического программирования (см. гл. 9), будет рассмотрена отдельно и достаточно подробно, так как идеи основных подходов и методов в данном случае наиболее наглядны.

### § 1. Основы теории

**1. Правило множителей Лагранжа.** Пусть  $Q = \{x: g_i(x) = 0, i = 1, \dots, m\}$ . Точки  $x \in Q$  называются *допустимыми*. Точка  $x^*$  называется (*локальным*) *минимумом* в задаче (A), если она допустима и  $f(x^*) \leq f(x)$  для всех допустимых  $x$ , достаточно близких к  $x^*$ .

**Теорема 1** (необходимое условие минимума I порядка). *Пусть  $x^*$  — точка минимума в задаче (A), функции  $f(x)$ ,  $g_i(x)$  непрерывно дифференцируемы в окрестности  $x^*$ . Тогда найдутся числа  $y_0^*, y_1^*, \dots, y_m^*$ , не все равные 0, такие, что*

$$y_0^* \nabla f(x^*) + \sum_{i=1}^m y_i^* \nabla g_i(x^*) = 0. \quad (1)$$

Будем называть  $x^*$  *регулярной точкой минимума*, если  $f'(x^*)$ ,  $g_i'(x^*)$  непрерывно дифференцируемы в ее окрестности и  $\nabla g_i(x^*)$ ,  $i = 1, \dots, m$ , линейно независимы.

**Теорема 2** (правило множителей Лагранжа). *Если  $x^*$  — регулярная точка минимума, то найдутся  $y_1^*, \dots, y_m^*$  такие, что*

$$\nabla f(x^*) + \sum_{i=1}^m y_i^* \nabla g_i(x^*) = 0. \quad (2)$$

Числа  $y_1^*, \dots, y_m^*$  в (2) называются *множителями Лагранжа*.

То, что правило множителей Лагранжа справедливо, вообще говоря, лишь при условии регулярности, можно увидеть из приведенных примеров. Так, в задаче тип  $x$ ,  $x^2 = 0$ ,  $x \in \mathbf{R}^1$  точка  $x^* = 0$  является точкой минимума (но не регулярной), и равен-

ство (2) не выполняется ни при каком  $y^*$ , так как  $\gamma(x^*) = 1$ ,  $g'(x^*) = 0$  (см. также упр. 1).

Теорема 2 немедленно следует из теоремы 1. Действительно, в регулярном случае  $y_0^* \neq 0$  (иначе  $\sum_{i=1}^m y_i^* \nabla g_i(x^*) = 0$ ,  $y_1^*, \dots, y_m^*$  не все равны 0, что противоречит линейной независимости  $\nabla g_i(x^*)$ ). Разделив равенство (1) на  $y_0^*$ , получаем (с точностью

линейно зависимы:  $\sum_{i=1}^m \mu_i \nabla g_i(x^*) = 0$ ,  $\sum_{i=1}^m \mu_i^2 \neq 0$ , то равенство (1) справедливо при  $y_0^* = 0$ ,  $y_i^* = \mu_i$ ,  $i = 1, \dots, m$ . Следовательно, достаточно доказать теорему 2. Ниже будут рассмотрены три различных доказательства. Это вызвано как важностью при построении, так и использованием идей этих доказательств при

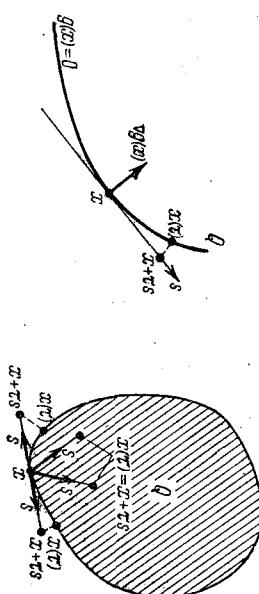


Рис. 33. Касательные векторы.

Составим так называемую *функцию Лагранжа*

$$L(x, y) = f(x) + (y, g(x)) = f(x) + \sum_{i=1}^m y_i g_i(x), \quad (3)$$

определенную на  $\mathbf{R}^n \times \mathbf{R}^m$ . Здесь и далее мы используем векторную запись  $y = (y_1, \dots, y_m)$ ,  $g(x) = (g_1(x), \dots, g_m(x))$ . Тогда правило множителей Лагранжа формулируется так:

$$L'_x(x^*, y^*) = 0, \quad L'_y(x^*, y^*) = 0, \quad (4)$$

где  $L'_x$ ,  $L'_y$  означают производные по соответствующим переменным. Запись в виде (4) удобна симметрией по переменным  $x$  и  $y$ , называемым соответственно *прямым* и *двойственным*.

*Доказательство*, основанное на исключении и непрерывности. Пусть  $Q \subset \mathbf{R}^n$  — некоторое множество,  $x \in Q$ . Вектор  $s \in \mathbf{R}^n$  называется *касательным* к  $Q$  в точке  $x$ , если для всех достаточно малых  $t \geq 0$  найдутся точки  $x(\tau) \in Q$  такие, что  $\|x(\tau) - (x + ts)\| = o(\tau)$  (рис. 33). Если  $Q$  выпукло, то всяко

допустимое направление (<§ 1 гл. 7>) является и касательным, но не наоборот (см. упр. 3). Очевидно, что касательные векторы образуют конус  $S_Q(x)$  (т. е. если  $s \in S$ , то и  $\lambda s \in S$  при  $\lambda \geq 0$ ).

Обратите внимание на то, что если  $x$  — граничная точка шара, то конус  $S$  является полупространством, а не гиперплоскостью.

Так что здесь термин «касательный вектор» имеет другой смысл, чем в геометрии.

**Теорема 3 (Люстерник).** Пусть  $Q = \{x \in \mathbf{R}^n: g_i(x) = 0, i = 1, \dots, m\}$ , где  $g_i(x)$  — непрерывно дифференцируемые  $\sigma$ -окрестности  $x^* \in Q$  функции, причем  $\nabla g_i(x^*)$ ,  $i = 1, \dots, m$ , линейно независимы. Тогда

$$S_Q(x^*) = \{s \in \mathbf{R}^n: (s, \nabla g_i(x^*)) = 0, i = 1, \dots, m\}, \quad (5)$$

т. е. касательные к множеству  $Q$  в точке  $x^*$  векторы образуют подпространство, ортогональное векторам  $\nabla g_1(x^*)$ ,  $\dots$ ,  $\nabla g_m(x^*)$ . ▲

Для доказательства правила множителей Лагранжа нам понадобится следующий результат.

**Лемма 1.** Пусть  $A$  — матрица  $m \times n$ ,  $L = \{x \in \mathbf{R}^n: Ax = 0\}$  и  $(c, x) \geq 0$  для всех  $x \in L$ . Тогда  $c = A^T y$ ,  $y \in \mathbf{R}^m$  и  $(c, x) = 0$  при  $x \in L$ .

**Доказательство.** Множество  $L_1 = \{x \in \mathbf{R}^n: x = A^T y$ ,  $y \in \mathbf{R}^m\}$  выпукло и замкнуто, как полупространство в  $\mathbf{R}^n$ . Если  $c \in L_1$ , то по теореме отделимости точка  $c$  может быть строго отделена от  $L_1$ , т. е. найдется  $a \in \mathbf{R}^n$ , такое, что  $(a, c) < 0$  и  $(a, x) \geq 0$ ,  $x \in L_1$ . Но тогда  $0 \leq (a, x) = (Aa, y) = (Aa, y)$  для всех  $y \in \mathbf{R}^m$ . Это возможно лишь, если  $Aa = 0$ ,  $a \in L$ , что противоречит условию  $(a, c) < 0$ . Итак,  $c \in L_1$ . ▲

Доказательство теоремы 2. Пусть  $s$  — касательный вектор к множеству  $Q = \{x: g_i(x) = 0, i = 1, \dots, m\}$  в точке  $x^*$ . Тогда найдутся  $x(\tau)$  такие, что  $g_i(x(\tau)) = 0$ ,  $i = 1, \dots, m$ ,  $\|x^* + \tau s - x(\tau)\| = o(\tau)$ . Поэтому

$$f(x(\tau)) = f(x^* + \tau s + o(\tau)) = f(x^*) + \tau (\nabla f(x^*), s) + o(\tau).$$

Поскольку  $f(x(\tau)) \geq f(x^*)$  для достаточно малых  $\tau$ , то отсюда  $(\nabla f(x^*), s) \geq 0$ . По теореме Люстерника  $(s, \sum_{i=1}^m \mu_i \nabla g_i(x^*)) = 0$ ,

$i = 1, \dots, m$ . Используя лемму 1, получаем  $\nabla f(x^*) = \sum_{i=1}^m \mu_i \nabla g_i(x^*)$ ,

где  $\mu_i$  — некоторые числа. Положив  $y_i^* = -\mu_i$ ,  $i = 1, \dots, m$ , приходим к (2). ▲

**2°.** Доказательство, основанное на исключении и непрерывности. Если  $\nabla g_i(x^*)$  линейно независимы, то матрица  $g'(x^*)$ , строками которой являются  $\nabla g_1(x^*)$ ,  $\dots$ ,  $\nabla g_m(x^*)$ , имеет ранг  $m$ . Поэтому найдутся компоненты вектора  $x$ , число которых равно  $m$  (их множество обозначим  $I$ ), такие, что ма-трица с элементами  $\partial g_i(x^*)/\partial x_i$ ,  $i = 1, \dots, m$ ,  $i \in I$  имеет

обратную. Запишем вектор  $x \in \mathbf{R}^n$  в виде  $x = \{u, v\}$ , где  $u \in \mathbf{R}^m$  — компоненты  $x$  с индексами из  $I$ ,  $v \in \mathbf{R}^{n-m}$  — остальные компоненты. Тогда матрица  $g'_u(u^*, v^*)$  (где  $g(u, v) = g(x)$ ,  $x^* = \{u^*, v^*\}$ ) имеет обратную. Рассмотрим равенство  $g'(u, v) = 0$ . Поскольку  $g'(u^*, v^*) = 0$ ,  $g$  непрерывно дифференцируема в окрестности  $\{u^*, v^*\}$  и матрица  $g'_u(u^*, v^*)$  невырождена, то по теореме о неявной функции (теорема 2 § 3 гл. 2) найдется дифференцируемая в окрестности  $v^*$  функция  $u(v)$ , такая, что  $u(v^*) = u^*$ ,  $g(u(v), v) = 0$  и  $u'(v) = -[g''(u(v), v)]^{-1} g'_v(u(v), v)$ . Рассмотрим теперь функцию  $\Phi(v) = f(u(v), v)$ , где  $f(u, v) = f(x)$ . Функция  $\Phi(v)$  достигает локального безусловного минимума в  $v^*$ . Действительно, при любом  $v$ , близком к  $v^*$ ,  $g(u(v), v) = 0$ , т. е. точка  $x = (u(v), v)$  является допустимой, а потому  $\Phi(v^*) = f(u(v^*), v^*) = f(x^*) \leq f(x) = f(u(v), v) = \Phi(v)$ . Следовательно,  $\nabla \Phi(v^*) = 0$ . Но по правилу дифференцирования сложной функции  $\nabla \Phi(v) = u'(v)^T f'_u(u(v), v)^T + f'_v(u(v), v)^T$ . Поэтому

$$0 = \nabla \Phi(v^*) = -g'_v(u^*, v^*)^T [g'_u(u^*, v^*)^T]^{-1} f'_u(u^*, v^*)^T + f'_v(u^*, v^*)^T.$$

Обозначим

$$[g'_u(u^*, v^*)^T]^{-1} f'_u(u^*, v^*)^T = -y^*. \quad (6)$$

Тогда  $f'_u(u^*, v^*)^T + g'_u(u^*, v^*)^T y^* = 0$ ,  $f'_v(u^*, v^*)^T + g'_v(u^*, v^*)^T y^* = 0$ , что эквивалентно равенству (2).  $\blacktriangle$

Это доказательство основывается на идее сведения задачи с ограничениями к задаче на безусловный минимум путем исключения переменных. Именно, переменные  $x \in \mathbf{R}^n$  разбиваются на две группы  $u \in \mathbf{R}^m$ ,  $v \in \mathbf{R}^{n-m}$ , из равенств  $g(x) = 0$  выражаем одну группу через другую:  $u = u(v)$  и рассматриваем задачу безусловной минимизации для  $\Phi(v) = f(u(v), v)$ . Необходимое условие минимума для нее ( $\nabla \Phi(v^*) = 0$ ) порождает условие экстремума для исходной задачи. При этом формула (6) дает явное выражение для множителей Лагранжа.

З<sup>о</sup>. Доказательство, основанное на штрафных функциях. Пусть  $U = \{x: \|x - x^*\| \leq \varepsilon\}$ , где величина  $\varepsilon > 0$  такова, что  $f, g_i$  непрерывно дифференцируемы в  $U$  и  $x^*$  — точка глобального минимума  $f$  на  $Q \cap U$ .

Рассмотрим задачу

$$\min_{x \in U} f_k(x), \quad f_k(x) = f(x) + \frac{1}{2} K \sum_{i=1}^m g_i^2(x) + \frac{1}{2} \|x - x^*\|^2, \quad (7)$$

где  $K$  — некоторый параметр. В силу непрерывности  $f_k(x)$  за-

дача (7) имеет решение  $x^k$ . Тогда

$$f_k(x^k) \leq f_k(x^*),$$

$$f(x^k) + \frac{1}{2} K \sum_{i=1}^m g_i^2(x^k) + \frac{1}{2} \|x^k - x^*\|^2 \leq f(x^*),$$

$$\sum_{i=1}^m g_i^2(x^k) \leq \frac{2}{K} (f(x^*) - f(x^k) - \frac{1}{2} \|x^k - x^*\|^2).$$

Величина в правой части стремится к 0 при  $K \rightarrow \infty$  (так как  $\|x^k - x^*\| \leq \varepsilon$ ), поэтому  $g(x^k) \rightarrow 0$ . Выберем подпоследовательность  $x^{k_i} \rightarrow \tilde{x} \in U$ , тогда  $g(\tilde{x}) = 0$ ,  $f(\tilde{x}) + \|\tilde{x} - x^*\|^2/2 \leq f(x^*)$ , но с другой стороны, поскольку  $x^*$  — точка минимума на  $Q$ , то  $f(x^*) \leq f(\tilde{x})$ . Поэтому  $\tilde{x} = x^*$ . Так как всякая предельная точка для  $x^k$  совпадает с  $x^*$ , то  $x^k \rightarrow x^*$  при  $K \rightarrow \infty$ . Поэтому для достаточно больших  $K > 0$  точка  $x^k$  лежит внутри  $U$ . Следовательно, условие минимума для нее приобретает вид  $\nabla f_k(x^k) = 0$ , т. е.

$$\nabla f(x^k) + K \sum_{i=1}^m g_i(x^k) \nabla g_i(x^k) + x^k - x^* = 0. \quad (8)$$

Введем величины

$$y_0^k = \frac{1}{\sqrt{1 + K^2 \sum_{i=1}^m g_i^2(x^k)}}, \quad y_i^k = \frac{K g_i(x^k)}{\sqrt{1 + K^2 \sum_{i=1}^m g_i^2(x^k)}}, \quad i=1, \dots, m.$$

Равенство (8) можно записать в виде

$$y_0^k \nabla f(x^k) + \sum_{i=1}^m y_i^k \nabla g_i(x^k) + (x^k - x^*) y_0^k = 0. \quad (9)$$

Имеем  $\sum_{i=0}^m (y_i^k)^2 = 1$  для всех  $k$ , поэтому существует последовательность  $k_i \rightarrow \infty$  такая, что  $y_i^k \rightarrow y_i^*, i=0, \dots, m$ ,  $\sum_{i=0}^m (y_i^*)^2 = 1$ . Переходя к пределу в (9), получаем (1).  $\blacktriangle$

В данном доказательстве эксплуатируется та же идея, что и в предыдущем, — использование необходимого условия экстремума в задаче без ограничений для получения необходимого условия в задаче с ограничениями. Однако способ сведения одноточной задачи к другой здесь совсем иной — строится последовательность ( $K \rightarrow \infty$ ) задач безусловной минимизации, отличаю-

щихся все большим «штрафом» за нарушение ограничений (член

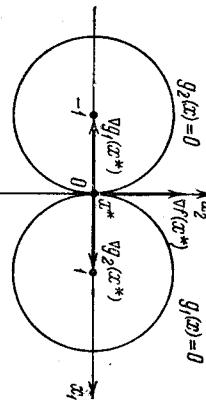
$$\frac{1}{2} K \sum_{i=1}^m g_i^2(x) \text{ в } f_k(x)), \text{ решения которых в пределе стремятся к решением исходной задачи условной минимизации.}$$

Приведенное в данном пункте доказательство наиболее простое по используемым в нем средствам (не применяется ни теорема Листерника, ни теорема о неявной функции, ни аналогичные им утверждения).

- Упражнения.**
1. Рассмотрите задачу (A) в  $\mathbf{R}^2$ :  $f(x) = x_2$ ,  $g_1(x) = (x_1 - 1)^2 + x_2^2 - 1$ ,  $g_2(x) = (x_1 + 1)^2 + x_2^2 - 1$  (рис. 34). Покажите, что  $x^* = \{0, 0\}$  — не регулярная точка минимума, и в ней не выполняется (2).
  2. Убедитесь, что если  $g_i(x) = (a_i^T x - b_i)$ ,  $i = 1, \dots, m$ , то (2) совпадает с (8) § 1 гл. 7.
  3. Покажите, что если  $Q$  выпукло, то касательный конус  $S$  является выпуклым и совпадает с  $\Gamma$  — замыканием конуса, порожденного допустимыми направлениями (см. доказательство теоремы 3 § 1 гл. 7).

Рис. 34. Задача с нерегулярным минимумом.

- 5.** Покажите, что если точка  $x^*$  — локально единственный минимум, то в функцию  $f_k(x)$  (см. доказательство в п. 3°) можно не включать член  $\|x - x^*\|^2/2$ .
- 6.** Убедитесь, что если минимум регуляризаций, то для величин  $K g_i(x^k)$  (см. доказательство в п. 3°) существует предел при  $K \rightarrow \infty$ :  $K g_i(x^k) \rightarrow y_i^*/b_i^*$ .



## 2. Условия минимума II порядка.

Теорема 4 (необходимое условие II порядка). Пусть  $x^*$  — регулярная точка минимума в задаче (A),  $f(x)$  и  $g_i(x)$  — дифференцируемы в окрестности  $x^*$ , а  $y_i^*$ ,  $i = 1, \dots, m$  — множители Лагранжа. Тогда

$$(L''_{xx}(x^*, y^*)s, s) = \left( (\nabla^2 f(x^*) + \sum_{i=1}^m y_i^* \nabla^2 g_i(x^*))s, s \right) \geq 0 \quad (10)$$

для всех  $s \in S = \{s: (\nabla g_i(x^*), s) = 0, i = 1, \dots, m\}$ .

Иными словами, матрица  $L''_{xx}(x^*, y^*)$  неотрицательно определена на касательном подпространстве  $S$  (см. (5)).

Доказательство. Пусть  $s \in S$ . В соответствии с теоремой Листерника найдутся допустимые  $x(\tau)$  такие, что

$$\begin{aligned} \|x^* + \tau s - x(\tau)\| &= o(\tau). \text{ Тогда, используя (4), получаем} \\ f(x^*) &\leq f(x(\tau)) = L(x(\tau), y^*) = L(x^*, y^*) + (L'_x(x^*, y^*), x(\tau) - x^*) + \\ &+ (L''_{xx}(x^*, y^*)(x(\tau) - x^*)/2 + o(\tau^2)) = \\ &= f(x^*) + (\tau^2/2)(L''_{xx}(x^*, y^*)s, s) + o(\tau^2), \end{aligned}$$

откуда  $(L''_{xx}(x^*, y^*)s, s) \geq 0$ . ▲

Как уже на первый взгляд естественным более обще необходимое условие экстремума  $L''_{xx}(x^*, y^*) \geq 0$  на самом деле неверно (см. упр. 8).

Прежде чем переходить к достаточным условиям экстремума, приведем некоторые вспомогательные результаты, относящиеся к матрицам специального вида. Эти результаты будут систематически использоваться в дальнейшем.

Лемма 2. Пусть  $A$  — симметричная матрица  $n \times n$ ,  $C$  — матрица  $m \times n$  ранга  $m$  и  $(Ax, x) \geq 0$  для всех  $x \neq 0$  таких, что  $Cx = 0$ . Тогда блочная матрица

$$B = \begin{pmatrix} A & C^T \\ C & 0 \end{pmatrix} \quad (11)$$

размерности  $(m+n) \times (m+n)$  имеет обратную. ▲

Лемма 3. В условиях леммы 2 найдутся такие  $K_0 > 0$ ,  $\alpha > 0$ , что  $A + KCTC \leq \alpha I$  при  $K \geq K_0$ . ▲

Лемма 4. В условиях леммы 2 для достаточно больших  $K$   $\|(A + KCTC)^{-1}Cr\| \leq \alpha_1/K$ ,  $\|C(A + KCTC)^{-1}\| \leq \alpha_2/K$ ,

$$\|I - KC(A + KCTC)^{-1}Cr\| \leq \alpha_3/K,$$

где  $\alpha_i$  — некоторые константы. ▲

Лемма 5. В условиях леммы 2 для матрицы

$$B_k = \begin{pmatrix} A & C^T \\ C & -\frac{1}{K} I \end{pmatrix} \quad (12)$$

при достаточно больших  $K$  существует  $B_k^{-1}$ , причем  $\|B_k^{-1}\| \leq \gamma$ . ▲

Вернемся к формулировке условий экстремума.

Теорема 5 (достаточное условие II порядка). Пусть  $g_i(x^*) = 0$ ,  $i = 1, \dots, m$ , функции  $f(x)$  и  $g_i(x)$  — дифференцируемы в окрестности  $x^*$ ,  $\nabla g_i(x^*)$ ,  $i = 1, \dots, m$ , линейно независимы, выполнено необходимое условие минимума (4) и

$$L''_{xx}(x^*, y^*)s, s > 0 \quad (13)$$

при всех  $s$  таких, что  $(\nabla g_i(x^*), s) = 0$ ,  $i = 1, \dots, m$ . Тогда  $x^*$  — точка локального минимума в задаче (A).

Иначе говоря, если в  $x^*$  выполнено необходимое условие экстремума II порядка и матрица  $L''_{xx}(x^*, y^*)$  положительно

определенна на касательном подпространстве  $S$ , то  $x^*$  — точка минимума. Точку  $x^*$ , в которой выполняются условия теоремы 5, будем называть *невырожденным минимумом*.

Доказательство. Введем функцию

$$\begin{aligned} M(x, y, K) &= f(x) + (y, g(x)) + (K/2) \|g(x)\|^2 = \\ &= L(x, y) + (K/2) \|g(x)\|^2, \quad (14) \end{aligned}$$

где  $K > 0$  — некоторый параметр. Тогда  $M'_x(x^*, y^*, K) = L'_x(x^*, y^*) = 0$ ,  $M''_{xx}(x^*, y^*) = L''_{xx}(x^*, y^*) + Kg'(x^*)^T g'(x^*)$ . Для матриц  $A = L''_{xx}(x^*, y^*)$  и  $C = g'(x^*)$  применима лемма 3, поэтому для достаточно больших  $K > 0$  будет

$$(15)$$

$$M''_{xx}(x^*, y^*, K) > 0.$$

Таким образом, выполнено достаточное условие локального минимума  $M(x, y^*, K)$  (теорема 4 § 2 гл. 1), т. е.  $M(x, y^*, K) \geq M(x^*, y^*, K)$  для всех  $x$ , достаточно близких к  $x^*$ . Но для  $x \in Q$  (т. е. для допустимых  $x$ ) имеем  $M(x, y^*, K) = f(x)$ , т. е.  $f(x) \geq f(x^*)$  для  $x \in Q$  из окрестности  $x^*$ . ▲

Функция  $M(x, y, K)$  (14) называется *модифицированной функцией Лагранжа*. Она играет важную роль в теории условной оптимизации. Остановимся на некоторых ее свойствах. Прежде всего, она отличается от обычной функции Лагранжа (3) наличием «штрафного члена»  $(K/2) \|g(x)\|^2$  и совпадает с (3) при  $K = 0$ :  $M(x, y, 0) = L(x, y)$ . Далее, если  $x \in Q$ , то  $M(x, y, K) = L(x, y) = f(x)$  и  $M'_x(x, y, K) = L'_x(x, y)$ , а  $M'_y(x, y, K) = L'_y(x, y) = g(x)$ . Поэтому необходимо условие минимума I порядка имеет вид, аналогичный (4):

$$(16)$$

$$M'_x(x^*, y^*, K) = 0, \quad M'_y(x^*, y^*, K) = 0,$$

где множители Лагранжа  $y^*$  те же, что и в (4).

Однако на уровне условий II порядка начинаются различия между  $M(x, y, K)$  и  $L(x, y)$ . Мы убедились при доказательстве теоремы 5, что если  $x^*$  — невырожденная точка минимума в задаче с ограничениями (A), то  $x^*$  — невырожденная точка безусловного минимума  $M(x, y^*, K)$  при достаточно больших  $K$ . Для обычной функции Лагранжа аналогичное утверждение неверно — точка  $x^*$  является стационарной точкой  $L(x, y^*)$ , однако не обязательно точкой минимума (см. упр. 8). Указанное свойство модифицированной функции Лагранжа позволяет строить эффективные методы оптимизации с ее помощью (§ 2).

**Упражнения.**

7. Покажите, что в задаче  $\min f(x)$ ,  $x \in \mathbb{R}^n$ ,  $g(x) = 0$ ,  $f(x) = x_1^2 - x_2^2$ , решением является  $x^* = \{0, 0\}$ , при этом  $y^* = 0$ . Проверьте, что матрица  $L''_{xx}(x^*, y^*)$  неанализирована.

8. В задаче  $\min f(x)$ ,  $x \in \mathbb{R}^2$ ,  $g(x) = 0$ ,  $f(x) = x_1^2 - x_2^2$ , решением является  $x^* = \{0, 0\}$ , при этом  $y^* = 0$ . Проверьте, что матрица  $m \times n$ ,  $B = \begin{pmatrix} A & C^T \\ C & D \end{pmatrix}$  — матрица  $(n+m) \times (n+m)$ . Докажите, что условие  $A \geq 0$ ,  $CA + C^T - D \geq 0$  (обобщение критерия Сильвестра на матричный случай).

10. Докажите, что в условиях теоремы 5  $x^* = \arg \min_{x \in S} L(x, y^*)$ .

**3. Использование условий экстремума.** В обычных курсах математического анализа изучение задач условной минимизации заканчивается выводом условий экстремума. Считается, что они дают возможность найти решение. Разумеется, это не так. Правило множителей Лагранжа задает систему уравнений (4) относительно  $x^*, y^*$ . Эти уравнения нелинейны (за исключением случая квадратичной  $f(x)$  и линейных  $g_i(x)$ ), и их решение найти в явной форме обычно невозможно. Все приводимые в учебниках примеры, демонстрирующие возможность решения задач с помощью множителей Лагранжа, являются специально подобранными исключениями из правил типа приводимых в упражнении 11.

Подлинное значение условий экстремума заключается в другом (ср. аналогичные замечания в § 2 гл. 1). Во-первых, с их помощью строятся численные методы отыскания решения. Во-вторых, они дают возможность после того, как решение найдено, судить о его единственности, устойчивости и т. д. (см. ниже). Наконец, они дают те естественные требования к задаче, при которых ее удобно анализировать (например, исследовать сходимость методов).

Многочисленные примеры подобного использования условий экстремума будут постоянно встречаться нам далее.

**Упражнение.**

11. Найдите решения следующих задач с помощью множителей Лагранжа и докажите оптимальность ответов с помощью достаточных условий экстремума:

$$\text{a) } \min \sum_{i=1}^n x_i, \quad \sum_{i=1}^n x_i = 1;$$

$$\text{б) } \min \sum_{i=1}^n x_i, \quad \sum_{i=1}^n x_i^2 = 1;$$

$$\text{в) } \min (Ax, x), \quad \|x\| = 1;$$

$$\text{г) } \min \|x\|^2, \quad (Ax, x) = 1.$$

Ответы. а)  $x_i^* = 1/n$ ,  $i = 1, \dots, n$ ; б)  $x_i^* = -1/\sqrt{n}$ ,  $i = 1, \dots, n$ ; в)  $x^* = e^1$  — нормированный собственный вектор, отвечающий наименьшему

собственному значению матрицы  $A$ ; г)  $x^* = \lambda_n^{-1/2} e^n$ ,  $e^n$  — нормированный собственный вектор, отвечающий наибольшему собственному значению  $\lambda_n$  матрицы  $A$ , решение существует при  $\lambda_n > 0$ .

#### 4. Существование, единственность и устойчивость решения.

Вопрос о существовании решения вновь решается с помощью теоремы 4 § 1 гл. 7, и здесь специфика задачи (A) не играет роли.

Что касается единственности решения, то воспользоваться теоремой о единственности минимума строго выпуклой функции на выпуклом множестве для задачи (A) обычно не удается, так как множество  $Q$ , задаваемое нелинейными ограничениями типа равенств, невыпукло (за исключением вырожденных случаев) (см. упр. 12). Однако здесь можно привести априорные условия единственности.

**Теорема 6. Невырожденная точка минимума локально единственна.**

Действительно, при доказательстве теоремы 5 мы получили, что невырожденное решение  $x^*$  задачи (A) является невырожденной точкой безусловного минимума  $M(x, y^*, K)$ , поэтому найдется  $l > 0$  такое, что  $M(x, y^*, K) - M(x^*, y^*, K) \geq l \|x - x^*\|_B^2$  некоторой окрестности  $x^*$  (см. (2) § 3 гл. 1). Поскольку  $f(x) = M(x, y^*, K)$  для  $x \in Q$ , то

$$f(x) - f(x^*) \geq l \|x - x^*\|^2$$

для  $x \in Q$  из окрестности  $x^*$ . ▲

Совершенно очевиден следующий результат о единственности множителей Лагранжа, вытекающий из определения регулярной точки.

**Теорема 7. Для регулярного минимума множители Лагранжа определяются однозначно.** ▲

Переходя к анализу устойчивости, начнем с вопроса об устойчивости решения по отношению к возмущениям в ограничениях. Наряду с исходной задачей (A) введем «возмущенную» задачу

$$\min f(x), \quad (17)$$

$$g_i(x) = \varepsilon_i, \quad i = 1, \dots, m,$$

где  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_m) \in \mathbf{R}^m$  — некоторый вектор. Обозначим ее ре-

шение (если такое существует) через  $x_\varepsilon$ , и пусть  $\Phi(\varepsilon) = f(x_\varepsilon)$ . Нас будет интересовать, когда  $x_\varepsilon \rightarrow x^*$  при  $\varepsilon \rightarrow 0$  ( $x^*$  — решение (A)), а также оценки близости  $x_\varepsilon$  к  $x^*$  и поведение  $\Phi(\varepsilon)$  для малых  $\varepsilon$ .

**Теорема 8. Пусть  $x^*$  — невырожденное решение задачи (A). Тогда при достаточно малых  $\|\varepsilon\|$  существует  $x_\varepsilon$ , причем**

$$\|x_\varepsilon - x^*\| = O(\varepsilon),$$

$$\nabla \Phi(0) = -y^*.$$

**Доказательство.** Пусть  $z = \{x, y\} \in \mathbf{R}^{n+m}$ ,  $x \in \mathbf{R}^n$ ,  $y \in \mathbf{R}^m$ ,  $R(z)$  — оператор в  $\mathbf{R}^{n+m}$ , определяемый следующим образом:  $R(z) = \{L_x(x, y), L_y(x, y)\}$ . Тогда система уравнений (4) может быть записана в виде

(20)

$$R(z) = 0.$$

Очевидно, что  $R(z^*) = 0$ , где  $z^* = \{x^*, y^*\}$ ,  $x^*$  — решение задачи (A),  $y^*$  — соответствующие множители Лагранжа. Вычислим  $R'(z^*)$ . Имеем

$$R'(z^*) = \begin{pmatrix} A & C^T \\ C & 0 \end{pmatrix}, \quad A = L_{xx}(x^*, y^*), \quad C = L_{yx}(x^*, y^*) = g'(x^*). \quad (21)$$

Из леммы 2 следует, что матрица  $R'(z^*)$  невырождена. В соответствии с теоремой 3 § 3 гл. 2 система

$$R(z) = a \quad (22)$$

имеет решение  $z_a$  при достаточно малых  $\|a\|$ , причем

$$z_a = z^* - [R'(z^*)]^{-1} a + o(a). \quad (23)$$

Возьмем  $a = \{0, \varepsilon\}$ ,  $a \in \mathbf{R}^{n+m}$ ,  $\varepsilon \in \mathbf{R}^m$ . При этом система (22) эквивалентна следующей:

$$\nabla f(x) + g'(x)^T y = 0, \quad g(x) = \varepsilon, \quad (24)$$

и у нее существует решение  $z_a = \{x_\varepsilon, y_\varepsilon\}$  при достаточно малых  $\varepsilon$ . Тогда точка  $x_\varepsilon$ , во-первых, удовлетворяет ограничениям задачи (18). Во-вторых, в силу непрерывности  $Vg_i(x)$  и регулярности  $x^*$  градиенты  $Vg_i(x_\varepsilon)$  также линейно независимы при достаточно малых  $\|\varepsilon\|$ . В-третьих, в  $x_\varepsilon$  выполнено (в силу (24)) необходимое условие минимума в задаче (18) с множителями Лагранжа  $y_\varepsilon$ . Наконец, в силу непрерывности первых и вторых производных и линейной независимости  $Vg_i(x^*)$  выполняется условие  $L''_{xx}(x_\varepsilon, y_\varepsilon) > 0$  на подпространстве  $S_\varepsilon = \{s: (Vg_i(x_\varepsilon), s) = 0, i = 1, \dots, m\}$ . Итак, в точке  $x_\varepsilon$  выполняется достаточное условие экстремума II порядка, т. е.  $x_\varepsilon$  — решение задачи (18).

Из соотношения (23) следует  $\|z_a - z^*\| \leq \alpha \|a\|$ ,  $\alpha$  — некоторая константа, поэтому  $\|x_\varepsilon - x^*\| \leq \alpha \|\varepsilon\|$ . Наконец, из соотношения (24) получаем

$$\Phi(\varepsilon) = f(x_\varepsilon) = f(x^*) + (\nabla f(x^*), x_\varepsilon - x^*) + o(\|x_\varepsilon - x^*\|) = f(x^*) - (g'(x^*)^T y^*, x_\varepsilon - x^*) + o(\varepsilon) = f(x^*) - (g(x_\varepsilon) - g(x^*), y^*) + o(\varepsilon) = f(x^*) - (y^*, \varepsilon) + o(\varepsilon).$$

Значит,  $\nabla \Phi(0) = -y^*$ . ▲

Из теоремы 8 следует устойчивость невырожденного минимума по отношению к возмущениям в ограничениях. В частности, если для последовательности  $x^k$  из окрестности невырож-

денного минимума  $x^*$  будет  $\lim_{k \rightarrow \infty} g_i(x^k) = 0$ ,  $i = 1, \dots, m$ ,

$$\lim_{k \rightarrow \infty} f(x^k) = f(x^*), \text{ то эта последовательность сходится к } x^*.$$

Устойчивость по отношению к возмущениям минимизируемой функции исследуется точно так же. Приведем характерный результат.

**Теорема 9.** Пусть  $x_\varepsilon$  — решение задачи  $\min [f(x) + \varepsilon h(x)]$ ,  $g_i(x) = 0$ ,  $i = 1, \dots, m$ , где  $\varepsilon \in \mathbf{R}^1$ ,  $h(x)$  — однократно непрерывно дифференцируемая в окрестности  $x^*$  функция,  $x^*$  — невырожденное решение задачи (A). Тогда при малых  $|\varepsilon|$   $x_\varepsilon$  существует и  $x_\varepsilon \rightarrow x^*$  при  $\varepsilon \rightarrow 0$ . ▲

Предположение о невырожденности минимума в теоремах 8, 9 существенно. Например, если решается задача  $\min x$ ,  $g(x) = x^2 = 0$ ,  $x \in \mathbf{R}^1$ , то ее решение  $x^* = 0$  нерегулярно. Для возмущенной задачи с ограничением  $g(x) = \varepsilon$  при  $\varepsilon < 0$  решения не существует (допустимое множество пусто), а при  $\varepsilon > 0$   $x_\varepsilon = \sqrt{\varepsilon}$  и оценка (19) нарушается.

### Упражнение.

12. Пусть  $g : \mathbf{R}^n \rightarrow \mathbf{R}^1$  — строго выпуклая функция. Докажите, что если множество  $\{x: g(x) = 0\}$  содержит более чем одну точку, то оно невыпукло.

## § 2. Методы минимизации

**1. Классификация методов.** Методы решения задач условной оптимизации многочисленны и разнообразны. Их можно классифицировать как по формальным признакам, так и по содержательным.

Как и ранее, можно выделить методы нулевого, первого и второго порядков в зависимости от порядка используемых производных. Мы в основном будем иметь дело с методами первого порядка (в которых вычисляются градиенты  $\nabla f(x)$  и  $\nabla g_i(x)$ ) и второго порядка (в которых требуется знание  $\nabla^2 f(x)$  и  $\nabla^2 g_i(x)$ ). Далее, методы делятся на прямые (в которых итерации ведутся в пространстве прямых переменных  $x$ ) и двойственные (которые существенно используют двойственные переменные  $y$ ). Во многих методах на каждом шаге решается некоторая вспомогательная задача, и с вычислительной точки зрения удобно вести классификацию по ее типу. Это может быть задача безусловной минимизации, задача минимизации линейной или квадратичной функции при линейных ограничениях и т. д. Наконец, сами идеи, лежащие в основе методов, разнообразны. Это идейно исключенные переменных, линеаризация, штрафных функций, обычной и модифицированной функций Лагранжа и т. д. Ниже будут рассмотрены наиболее важные в идейном или вычислительном отношении методы.

**2. Метод линеаризации.** В этом методе на каждой итерации минимизируемая функция и ограничения линеаризуются. Поскольку задача минимизации линейной функции при линейных ограничениях может не иметь решения, в функцию добавляется квадратичный член (ср. с аналогичным приемом при построении градиентного метода (3) § 4 гл. 1). Таким путем приходим к методу, где очередное приближение  $x^{k+1}$  является решением следующей вспомогательной задачи:

$$\begin{aligned} \min & [(\nabla f(x^k), x - x^k) + (2\gamma)^{-1} \|x - x^k\|^2], \\ g_i(x^k) + (\nabla g_i(x^k), x - x^k) &= 0, \quad i = 1, \dots, m, \end{aligned} \quad (1)$$

где  $\gamma > 0$  — некоторый параметр. С такого рода задачей мы уже встречались неоднократно (гл. 7). С одной стороны, метод (1) может быть записан как метод проекции градиента на линеаризованные ограничения:

$$\begin{aligned} x^{k+1} &= P_{Q_k}(x^k - \gamma \nabla f(x^k)), \\ Q_k &= \{x: g(x^k) + g'(x^k)(x - x^k) = 0\}. \end{aligned} \quad (2)$$

С другой стороны, решением системы линейных уравнений

$$\begin{aligned} (1/\gamma)(x - x^k) + g'(x^k)^T y &= -\nabla f(x^k), \\ g'(x^k)(x - x^k) &= -g(x^k) \end{aligned} \quad (3)$$

является вектор  $\{x^{k+1}, y^{k+1}\} \in \mathbf{R}^{n+m}$ , первые компоненты которого совпадают с  $x^{k+1}$ . Таким образом, для нахождения  $x^{k+1}$  достаточно решить систему линейных уравнений (3) (размерности  $n+m$ ). Получаемый при этом вектор  $y^{k+1}$  является, как мы увидим далее, оценкой для множителей Лагранжа  $y^*$ .

**Теорема 1.** Пусть  $x^*$  — точка невырожденного минимума, а  $\nabla f(x)$ ,  $\nabla^2 g_i(x)$  двойствуют условию Липшица в окрестности  $x^*$ . Тогда существует  $\gamma > 0$  такое, что при  $0 < \gamma < \gamma$  метод (1) корректно определен и локально сходится к  $x^*$  со скоростью геометрической прогрессии.

Доказательство. Прежде всего, в силу линейной независимости  $\nabla g_i(x^*)$ ,  $i = 1, \dots, m$ , векторы  $\nabla g_i(x^k)$ ,  $i = 1, \dots, m$ , также линейно независимы для  $x^k$ , достаточно близких к  $x^*$ , и потому  $Q_k \neq \emptyset$  и метод (2) корректно определен, т. е. точка  $x^{k+1}$  существует. В соответствии с (2) метод может быть записан в виде (см. (6) § 2 гл. 7)

$$x^{k+1} = (I - C_k^+ C_k)(x^k - \gamma \nabla f(x^k)) + C_k^+(C_k x^k - g(x^k)), \quad (4)$$

Поскольку  $C_k = C + g''(x^*)(x^k - x) + o(x^k - x^*)$ , где  $C = g'(x^*)$ ,  $\nabla f(x^k) = \nabla f(x^*) + \nabla^2 f(x^*)(x^k - x^*) + o(x^k - x^*)$ ,  $\nabla f(x^*) + C^T y^* = 0$ , оценки методы.

а  $C^+ = CT(CCT)^{-1}$ , то для  $x^k$ , достаточно близких к  $x^*$ , получаем  
 $x^{k+1} - x^* = D(x^k - x^*) + o(x^k - x^*)$ ,

$$D = (I - CT(CCT)^{-1}C)(I - \gamma A), \quad A = L''_{xx}(x^*, y^*). \quad (5)$$

Покажем, что при достаточно малых  $\gamma > 0$  будет  $\rho(D) < 1$ , где  $\rho(D)$  — спектральный радиус  $D$ . Действительно, рассмотрим итерации  $u^{k+1} = D u^k$  при произвольном  $u^0 \in \mathbf{R}^n$ . Поскольку  $D u = P_3(I - \gamma A) u$ ,  $S = \{x: Cx = 0\}$ , то все  $u^k$  принадлежат  $S$ ,  $k \geq 1$ . Поэтому для  $k \geq 1$

$$\begin{aligned} \| (I - \gamma A) u^k \| ^2 &= \| u^k \| ^2 - 2\gamma \| Au^k, u^k \| + \gamma^2 \| Au^k \| ^2 \leq \\ &\leq \| u^k \| ^2 - 2\gamma \| u^k \| ^2 + \gamma^2 \| A \| \| u^k \| ^2, \end{aligned}$$

так как  $(Au, u) \geq l \| u \| ^2$ ,  $l > 0$ , для  $u \in S$  в силу невырожденности  $x^*$ . Значит,  $\| (I - \gamma A) u^k \| \leq q \| u^k \|$ ,  $q < 1$ , для малых  $\gamma > 0$ . Но  $\| I - CT(CCT)^{-1}C \| \leq 1$ , так как  $I - CT(CCT)^{-1}C$  — оператор проектирования, см. (6) § 1 гл. 5. Итак,  $\| u^{k+1} \| \leq q \| u^k \|$ ,  $q < 1$ ,  $k \geq 1$ , т. е.  $u^k \rightarrow 0$ , что эквивалентно условию  $\rho(D) < 1$  (следствие из леммы 1 § 1 гл. 2). Применяя теперь теорему 1 § 1 гл. 2 к (5), получаем требуемый результат. ▲

Теорема 1 является типичной в ряде отношений. Во-первых, в ней доказывается лишь локальная сходимость метода. Это естественно, так как в задачах с нелинейными ограничениями типа равенств допустимое множество, как правило, является невыпуклым, что уже отмечалось выше. Поэтому для таких задач нельзя надеяться ни на какой глобальный результат. Во-вторых, основным аппаратом при доказательстве является теорема 1 § 1 гл. 2. Впрочем, чаще итерационный процесс рассматривается одновременно для прямых и двойственных переменных ( $x$  и  $y$ ) и эта теорема применяется в пространстве  $\mathbf{R}^{n+m}$ . В-третьих, дается предположение о невырожденности минимума. В случае вырождения можно иногда доказать факт сходимости того или иного метода первого порядка, но нельзя гарантировать сходимость со скоростью геометрической прогрессии.

Доказательство. В обозначениях  $z^k = \{x^k - x^*, y^k - y^*\}$  метод может быть записан в виде

$$z^{k+1} = D z^k + o(z^k),$$

$$D = I - \gamma B, \quad B = \begin{pmatrix} A & C^T \\ -C & 0 \end{pmatrix}, \quad A = L''_{xx}(x^*, y^*), \quad C = g'(x^*). \quad (7)$$

Покажем, что матрица  $-B$  устойчива. Это эквивалентно тому, что для системы  $\dot{z} = -Bz$  будет  $z(t) \rightarrow 0$  при  $t \rightarrow \infty$  и любом  $z(0)$  (см. лемму 3 § 1 гл. 2). В переменных  $x, y$  система принимает вид  $\dot{x} = -Ax - C^T y$ ,  $\dot{y} = Cx$ . Возьмем  $\rho(t) = (\|x(t)\|^2 + \|y(t)\|^2)/2$ , тогда  $\dot{\rho} = (\dot{x}, x) + (\dot{y}, y) = -(Ax, x) - (C^T y, x) + (Cx, y) = - (Ax, x) \leq -\alpha \|x\|^2$ , так как  $A > 0$ . Поэтому  $\rho$  монотонно убывает, отсюда  $\dot{\rho} \rightarrow 0$ , и потому  $x(t) \rightarrow 0$  при  $t \rightarrow \infty$ . Поскольку  $x(t)$  — решение линейного дифференциального уравнения, то из  $x(t) \rightarrow 0$  следует и  $\dot{x}(t) \rightarrow 0$ . Таким образом,  $Cy = -\dot{x} - Ax \rightarrow 0$  при  $t \rightarrow \infty$ . Поскольку ранг  $C$  равен  $m$ , то это влечет  $y \rightarrow 0$ . Итак,  $z(t) \rightarrow 0$ . В силу леммы 5 § 1 гл. 2 спектральный радиус  $D \equiv I - \gamma B$  меньше 1 при достаточно малых  $\gamma > 0$ . Применяя теорему 1 § 1 гл. 2 для (7), получаем требуемый результат. ▲

Можно модифицировать метод (6), осуществив полную ми-

нимизацию функции Лагранжа по  $x$  вместо одного шага градиентного метода:

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbf{R}^n} L(x, y^k), \quad y^{k+1} = y^k + \gamma g(x^{k+1}). \quad (8)$$

Оказывается, что в предположениях теоремы 2 метод корректно определен при  $y^0$ , достаточно близких к  $y^*$  (т. е.  $x^{k+1}$  существует), и для (8) справедливы утверждения теоремы 2. Таким образом, переход к более трудоемкому методу (8) (в нем требуется на каждом шаге решать задачу безусловной минимизации) не меняет качественной картины поведения метода (6), хотя и может привести к уменьшению знаменателя прогрессии.

**4. Методы модифицированной функции Лагранжа.** Значительные лучшими свойствами обладают методы типа (6), (8), в которых обычная функция Лагранжа заменена на модифицированную. Начнем с аналого метода (6):

$$\begin{aligned} x^{k+1} &= x^k - \gamma M'_x(x^k, y^k, K) = x^k - \gamma (\nabla f(x^k) + \\ &\quad + g'(x^k)^T y^k + K g'(x^k)^T g(x^k)), \end{aligned} \quad (9)$$

$$y^{k+1} = y^k + \gamma M'_y(x^k, y^k, K) = y^k + \gamma g(x^k).$$

**Теорема 3.** Пусть  $x^*$  — невырожденная точка минимума,  $\nabla^2 f(x)$  и  $\nabla^2 g_i(x)$  удовлетворяют условию Липшица в окрестности  $x^*$ . Тогда метод (9) при достаточно больших  $K$  и  $0 < \gamma < \bar{\gamma}$  локально сходится к  $x^*$ ,  $y^*$  со скоростью геометрической прогрессии.

**Доказательство.** Задача (A) эквивалентна задаче

$$\min [f(x) + (K/2) \|g(x)\|^2], \quad (10)$$

$$g(x) = 0.$$

Для последней функция Лагранжа совпадает с  $M(x, y, K)$ , а метод (6) переходит в (9). Но в соответствии с (15), § 1  $M''_{xx}(x^*, y^*, K) > 0$ , поэтому применима теорема 2 о сходимости (6), а тем самым и (9). ▲

Таким образом, метод (9) сходится для любого невырожденного минимума без дополнительных предположений, требовавшихся для методов (6) и (8), основанных на функции Лагранжа. Оказывается при этом, что за счет использования модифицированной функции Лагранжа можно добиться и достаточно высокой скорости сходимости. Это относится к аналогу метода (8):

$$x^{k+1} = \underset{x}{\operatorname{argmin}} M(x, y^k, K), \quad y^{k+1} = y^k + K g(x^{k+1}). \quad (11)$$

**Теорема 4.** Пусть выполнены условия теоремы 3. Тогда для всякого  $y^0$ , достаточно близкого к  $y^*$ , найдется  $K_0$  такое, что при  $K > K_0$  метод (11) сходится к  $x^*$ ,  $y^*$  со скоростью геометрической прогрессии, знаменатель которой  $q = O(1/K)$ .

Доказательство. Обозначим  $\Phi(x) = M(x, y^*, K)$ ,  $\Psi_k(x) = (y^k - y^*, g(x))$ , тогда  $M(x, y^k, K) = \Phi(x) + \Psi_k(x)$ . В силу (15) и (16), § 1  $x^*$  — точка невырожденного безусловного минимума  $\Phi(x)$ . В силу теоремы 6 § 3 гл. 1, если  $\|y^k - y^*\|$  доста-

точно мало, то существует  $x^{k+1}$  — точка локального минимума  $\Phi(x) + \Psi_k(x)$  в окрестности  $x^*$ , причем  $x^{k+1} - x^* = -[\nabla^2 \Phi(x^*)]^{-1} \nabla \Psi_k(x^*) + o(y^k - y^*)$ .

Поскольку  $\nabla^2 \Phi(x^*) = A + K C^T C$ ,  $g'(x^*) = C$ ,  $A = L''_{xx}(x^*, y^*)$ ,  $\nabla \Psi_k(x^*) = C^T (y^k - y^*)$ , то

$$\begin{aligned} \|x^{k+1} - x^*\| &\leqslant \|(A + K C^T C)^{-1} C^T\| \|y^k - y^*\| + o(y^k - y^*) \leqslant \\ &\leqslant (\alpha_1/K) \|y^k - y^*\| + o(y^k - y^*) \end{aligned} \quad (12)$$

В силу леммы 4 § 1. Далее вновь применим эту лемму:

$$\begin{aligned} \|y^{k+1} - y^*\| &\leqslant \|I - KC(A + K C^T C)^{-1} C^T\| \|y^k - y^*\| + o(y^k - y^*) \leqslant \\ &\leqslant (\alpha_2/K) \|y^k - y^*\| + o(y^k - y^*). \end{aligned}$$

Таким образом, при достаточно малых  $y^k - y^*$  будет  $y^k \rightarrow y^*$  со скоростью геометрической прогрессии со знаменателем  $q = O(1/K)$ , а из (12) следует, что  $x^k - x^*$  с той же скоростью.

Более аккуратный учет остаточных членов в доказательстве теоремы 4 позволяет получить более сильное утверждение.

**Теорема 5.** Результаты теоремы 4 справедливы для любого  $y^0$  (при этом  $K_0$  зависит от  $\|y^0 - y^*\|$ ). ▲

Таким образом, метод (11) обладает рядом преимуществ перед методом (8). Во-первых, он не требует хорошего начального приближения по  $y$ , что важно, так как такое приближение обычно неизвестно. Во-вторых, он сходится при минимальных предположениях — достаточно невырожденности минимума и гладкости функций. При этом он не сложнее метода (8) с вычислительной точки зрения. Он не требует подбора длины шага  $\gamma$  (впрочем, в нем возникает проблема выбора  $K$ ). Наконец, что очень важно, его скорость сходимости может быть сделана высокой за счет выбора  $K$ .

Конечно, не нужно переопределять достоинства метода (11). В формулировках теорем 4 и 5 нет слов «локально сходится», однако это не значит, что метод дает возможность находить глобальное решение задачи, просто трудность перенесена на вспомогательную задачу (11). В определении  $x^{k+1}$  неявно предполагается, что это точка безусловного минимума функции  $M(x, y^k, K)$ , близкая к  $x^*$  (такая существует); однако отыскание такой точки остается проблемой. Далее, если сделать  $K$  очень большим, то скорость сходимости итерации (11) увеличивается, однако каждая итерация становится более трудоемкой. Дело в том, что задача минимизации  $M(x, y^k, K)$  становится плохо обусловленной.

Вопрос о компромиссном выборе  $K$  с учетом указанных условий довольно сложен и не решен до конца. В практических вычислениях можно менять  $K$  на каждой итерации, подбирая его в зависимости от результатов счета.

**5. Метод штрафных функций.** Применим для решения (A) ту же идею сведения к последовательности задач безусловной минимизации, что и при доказательстве правила множителей Лагранжа с помощью штрафных функций (см. (7) § 1):

$$x^k = \underset{x \in Q_0}{\operatorname{argmin}} f_k(x), \quad f_k(x) = f(x) + \frac{1}{2} K_k \|g(x)\|^2, \quad K_k \rightarrow \infty. \quad (13)$$

Здесь  $Q_0$  — некоторое ограниченное множество локализации минимума, введенное для того, чтобы решение задачи безусловной минимизации существовало. Докажем сходимость метода при минимальных предположениях (не будем требовать даже дифференцируемости  $f$  и  $g_i$ ).

**Теорема 6.** Пусть задача (A) имеет решения, их множество обозначим  $X^*$ . Пусть  $f$  и  $g_i$  непрерывны,  $Q_0$  ограничено и замкнуто,  $Q_0 \cap X^* \neq \emptyset$ . Тогда всякая предельная точка метода (13) (под  $x^k$  понимается глобальный минимум  $f_k(x)$  на  $Q_0$ ) является глобальным минимумом для задачи (A).

Доказательство. Метод (13) корректно определен, так как  $f_k(x)$  непрерывна и  $Q_0$  ограничено и замкнуто, поэтому по теореме 4 § 1 гл. 7 точки  $x^k$  существуют. Поскольку  $x^k \in Q_0$ , то существует хотя бы одна предельная точка  $\tilde{x}$  для последовательности  $x^k$ . Пусть  $x^* \in Q_0 \cap X^*$ , тогда по определению (13)

$$f_k(x^k) \leq f_k(x^*) = f(x^*). \quad \text{Отсюда } \|g(x^k)\|^2 \leq (2/K_k)(f(x^*) - f(x^k)) \text{ и, переходя к пределу при } k \rightarrow \infty, \text{ получаем, что } g(\tilde{x}) = 0, \text{ т. е. } \tilde{x} \text{ — допустимая точка. С другой стороны, } f'_k(x^k) \leq f'(x^*) - (1/2)K_k \|g(x^k)\|^2 \leq f'(x^*), \text{ следовательно } f'(\tilde{x}) \leq f'(x^*). \text{ Поэтому } \tilde{x} \text{ является решением (A).} \blacksquare$$

Для невырожденного минимума можно оценить и скорость сходимости метода (13).

**Теорема 7.** Пусть  $x^*$  — точка невырожденного минимума и  $\nabla^2 f(x)$ ,  $\nabla^2 g_i(x)$  удовлетворяют условию Липшица в окрестности  $x^*$ . Тогда  $x^k \rightarrow x^*$  при  $K_k \rightarrow \infty$ , причем

$$\|x^k - x^*\| = O(1/K_k), \quad K_k g(x^k) \rightarrow y^*. \blacksquare \quad (14)$$

Разумеется, из теоремы 7 не следует, что за счет выбора  $K_k$  можно добиться сколь угодно быстрой сходимости — как уже отмечалось выше, с ростом  $K_k$  ухудшается обусловленность вспомогательных задач и тем самым растет трудность их решения. Однако в методе штрафных функций неизбежно нужно брать  $K_k \rightarrow \infty$  (иначе он не сходится), и в этом главный недостаток метода по сравнению с методом модифицированной функции Лагранжа, в котором можно не увеличивать штрафной

коэффициент  $K$ . При практических расчетах  $K_k$  увеличивают таким образом, чтобы  $x^{k-1}$  могло служить начальным приближением при поиске  $x^k$  (при быстром росте  $K_k$  область сходимости методов безусловной минимизации для  $f_k(x)$  сокращается, и  $x^{k-1}$  может не попасть в эту область). Отметим, что метод штрафных функций совпадает с вариантом метода модифицированной функции Лагранжа, в котором не производится пересчет двойственных переменных ( $y^k \equiv 0$ ) (см. упр. 2). Вообще, в методе штрафных функций двойственные переменные никак не используются, хотя их можно было бы найти в соответствии с (14). По существу, метод модифицированной функции Лагранжа может рассматриваться как такое видоизменение метода штрафных функций, в котором систематически используется информация о множителях Лагранжа. Это видоизменение оказывается существенно более эффективным, чем исходный метод. Единственным возможным преимуществом метода штрафных функций может служить его большая универсальность — он сходится при весьма слабых требованиях (см. теорему 6).

#### Упражнения.

1. Докажите теорему 7 по той же схеме, что и теорему 4, записав (13) в виде  $x^k = \underset{x}{\operatorname{argmin}} M(x, 0, K_k)$ .

2. Рассмотрите метод  $x^k = \underset{x}{\operatorname{argmin}} M(x, y^0, K_k)$  при некотором постоянном  $y^0$  и покажите, что для него справедливы все результаты, относящиеся к методу штрафных функций (в котором  $y^0 = 0$ ). Докажите, что скорость сходимости тем выше, чем ближе  $y^0$  к  $y^*$ .

**6. Метод приведенного градиента.** Использование идей исключении переменных, примененной при втором способе доказательства правила множителей Лагранжа (§ 1) приводит к следующему методу, называемому методом *приведенного (или раздупированного) градиента*. Пусть  $x = \{u, v\}$ ,  $u \in \mathbf{R}^m$ ,  $v \in \mathbf{R}^{n-m}$  — *дупированного градиента*. Пусть  $x = \{u, v\}$ ,  $u \in \mathbf{R}^m$ ,  $v \in \mathbf{R}^{n-m}$  — разбиение переменных на две группы, причем и можно найти через  $v$  из уравнения  $g(x) = g(u, v) = 0$ . Построим градиентный метод безусловной минимизации функции  $\Phi(v) = f(u(v), v)$ , где  $f(u, v) = f(x)$ , т. е.

$$v^{k+1} = v^k - \gamma \nabla \Phi(v^k),$$

$$\nabla \Phi(v^k) = -g'_v(u^k, v^k)^T [g'_u(u^k, v^k)^T]^{-1} f'_u(u^k, v^k)^T + f'_v(u^k, v^k)^T, \quad (15)$$

где  $u^k$  — решение уравнения  $g(u, v^k) = 0$ . Таким образом, в методе (15) не нужно находить зависимость  $u(v)$  в явном виде, достаточно решать уравнение  $g(u, v^k) = 0$  при фиксированном  $v^k$ .

**Теорема 8.** Пусть  $x^*$  — точка невырожденного минимума и  $\nabla^2 f$ ,  $\nabla^2 g_i$  удовлетворяют условию Липшица в окрестности  $x^*$ . Тогда найдется  $\gamma > 0$  такое, что при  $0 < \gamma < \bar{\gamma}$  метод (15)

локально сходится к  $\{u^*, v^*\} = x^*$  со скоростью геометрической прогрессии.

Доказательство сводится к проверке невырожденности точки  $v^*$  безусловного минимума  $\Phi(v)$  и применению теоремы 4 § 4 гл. I. ▲

Остановимся отдельно на частном случае линейных ограничий. Пусть ограничения имеют вид

$$A_1 u + A_2 v = b,$$

где  $A_1$  — невырожденная матрица  $m \times m$ ,  $A_2$  — матрица  $m \times (n-m)$ ,  $b \in \mathbf{R}^m$ , а минимизируемая функция *сепаралльна* по  $u$  и  $v$ :

$$f(x) = f_1(u) + f_2(v). \quad (16)$$

Тогда метод (15) принимает вид

$$\begin{aligned} v^{k+1} &= v^k - y(\nabla f_2(v^k) - A_2^T (A_1^T)^{-1} \nabla f_1(u^k)), \\ u^{k+1} &= A_1^{-1}(b - A_2 v^{k+1}). \end{aligned} \quad (17)$$

Он сходится глобально, если функции  $f_1(u)$  и  $f_2(v)$  сильно выпуклые и гладкие. Можно также гарантировать глобальную сходимость, если только  $f_2(v)$  сильно выпукла, а  $f_1(u)$  имеет достаточно малую вторую производную.

**7. Метод Ньютона.** Для решения системы уравнений (4) § 1 можно применить метод Ньютона, т. е. новое приближение  $x^{k+1}$ ,  $y^{k+1}$  искать как решение системы линеаризованных уравнений

$$L''_{xx}(x^k, y^k)(x - x^k) + L''_{yx}(x^k, y^k)(y - y^k) = -L'_x(x^k, y^k), \quad (19)$$

(так как  $L''_{yy}(x, y) \equiv 0$ ), или в более подробной записи:

$$\begin{aligned} \left[ \nabla^2 f(x^k) + \sum_{i=1}^m y_i^k \nabla^2 g_i(x^k) \right] (x - x^k) + g'(x^k)^T (y - y^k) &= \\ &= -\nabla f(x^k) - g'(x^k)^T y^k \end{aligned} \quad (20)$$

$$g'(x^k)(x - x^k) = -g(x^k).$$

Теорема 9. Пусть  $x^*$  — невырожденная точка минимума и  $\nabla^2 f$  и  $\nabla^2 g_i$  удовлетворяют условию Липшица в окрестности  $x^*$ . Тогда метод (20) локально квадратично сходится к  $x^*, y^*$ .

Доказательство. Как и ранее (см. доказательство теоремы 8 § 1), введем  $z = \{x, y\} \in \mathbf{R}^{n+m}$ ,  $z^k = \{x^k, y^k\}$ ,  $z^* = \{x^*, y^*\}$ . Тогда (4) § 1 можно записать в виде

$$R(z) = 0, \quad R(z) = \{L'_x(x, y), L'_y(x, y)\},$$

$$R: \mathbf{R}^{n+m} \rightarrow \mathbf{R}^{n+m},$$

а (20) превращается в метод Ньютона для решения этого уравнения, т. е.  $z^{k+1}$  является решением системы линеаризованных уравнений

$$R'(z^k)(z - z^k) = -R(z^k). \quad (22)$$

Поскольку  $R(z^*) = 0$ ,  $R'(z)$  удовлетворяет условию Липшича в окрестности  $z^*$  и матрица  $R'(z^*)$  невырождена (см. (21) § 1 и лемму 2 § 1), то применим общий результат о сходимости метода Ньютона (теорема 3 § 5 гл. 1), что и дает требуемое утверждение. ▲

Метод Ньютона (20) обладает теми же достоинствами и недостатками, что и для безусловной минимизации: он быстро сходится, но требует трудоемкого вычисления вторых производных и хорошего начального приближения. Такое приближение особенно трудно найти для двойственных переменных.

Метод Ньютона можно представить в другой форме. Составим квадратичную аппроксимацию (по  $x$ ) функции Лагранжа при фиксированном значении  $y^k$  и будем искать минимум этой аппроксимации при линеаризованных ограничениях. В качестве  $y^{k+1}$  выберем  $y^k + u^k$ , где  $u^k$  — множители Лагранжа для вспомогательной задачи

$$\min \{(L'_x(x^k, y^k), x - x^k) + (L''_{xx}(x^k, y^k)(x - x^k), x - x^k)/2\}, \quad (23)$$

где  $L'_x(x^k, y^k) + g'(x^k)(x - x^k) = 0$ .

Нетрудно показать, что методы (20) и (23) эквивалентны. (Ср. различные формы записи метода линеаризации (1) — (3).)

**8. Другие квадратично сходящиеся методы.** Мы уже отмечали (см. упр. 10 § 1), что для невырожденного минимума  $x^*$

$$x^* = \underset{x \in S}{\operatorname{argmin}} L(x, y^*), \quad S = \{x: g'(x^*)(x - x^*) = 0\}. \quad (24)$$

Поэтому естественно построить метод, в котором на  $k$ -м шаге ищется минимум  $L(x, y^k)$  на подпространстве, образованном линеаризацией ограничения  $g(x) = 0$  в точке  $x^k$ :

$$x^{k+1} = \underset{x \in Q_k}{\operatorname{argmin}} L(x, y^k),$$

$$Q_k = \{x: g(x^k) + g'(x^k)(x - x^k) = 0\},$$

$$y^{k+1} = y^k + u^k,$$

где  $u^k$  — множители Лагранжа в задаче (25). Этот метод очень близок к (23) с той лишь разницей, что на  $Q_k$  ищется минимум функции  $L(x, y^k)$ , а не ее квадратичной аппроксимации. Поэтому в этом методе не нужно вычислять вторые производные — минимизация на подпространстве может быть осуществлена каким-либо эффективным методом первого порядка, например методом сопряженных градиентов (6) § 3 гл. 7.

Естественно, что свойства метода (25) те же, что и для метода (23).

Можно предложить и другие методы, идейно близкие к (23) и (25). Таковы, например, квазиньютоновские методы, в которых решается вспомогательная задача

$$\min_{x \in Q_k} [L'_x(x^k, y^k)(x - x^k) + \frac{1}{2}(H_k(x - x^k), x - x^k)], \quad (26)$$

где матрица  $H_k$  — аппроксимация для  $L''_{xx}(x^k, y^k)$ , построенная по предыдущим значениям градиентов.

### § 3. Учет возможных осложнений

В предыдущем параграфе мы рассматривали идеализированную ситуацию — пренебрегали наличием помех, ограничивались локальными результатами и случаем невырожденного минимума и т. д. Попытаемся учесть, в какой мере такая идеализация опасна.

**1. Глобальный минимум.** Все утверждения §§ 1—2 носили локальный характер, что, как уже отмечалось, связано с невыпуклостью задачи (A) (ср. с упр. 12 § 1). Даже если исходная задача (A) не имеет локальных минимумов, то система уравнений (4) § 1 (т. е. запись правила множителей Лагранжа) обычно имеет не единственное решение. Например, в задаче

$$(1)$$

$$\min_{\|x\|^2=1} (Ax, x),$$

где  $A$  — симметричная матрица  $n \times n$ , нет локальных минимумов, однако всякий собственный вектор  $A$  удовлетворяет необходимым условиям экстремума, которые в данном случае принимают вид  $Ax + yx = 0, y \in \mathbf{R}^1$ . Неудивительно, что и результаты о сходимости методов минимизации в основном были локальными, т. е. требовали хорошего начального приближения к решению. Исключением является метод штрафных функций — его сходимость носит глобальный характер (теорема 6 § 2). Конечно, в этом методе возникает проблема поиска глобального минимума во вспомогательных задачах (см. § 2 гл. 6). Можно при этом залиться одним разумным коэффициентом штрафа, и из полученных в процессе глобальной минимизации «полозрительных» точек произвести спуск локальными методами § 2.

В целом можно считать, что проблема поиска глобального решения задачи (A) очень сложна, и удовлетворительных универсальных способов ее решения нет.

**2. Влияние помех.** Теоремы об устойчивости, приведенные в

§ 2, дают основание надеяться, что малые помехи в вычислении функций и градиентов не приведут к каким-либо катастрофическим последствиям в невырожденном случае. Действительно,

можно показать, что для невырожденного минимума рассмотренные в § 2 методы при достаточно малых абсолютных помехах приводят в окрестность решения, причем размер этой окрестности тем меньше, чем ниже уровень помех.

Приведем один типичный результат. Рассмотрим «возмущенный» метод линеаризации:

$$x^{k+1} = \operatorname{argmin}_{x \in \tilde{Q}_k} ((\nabla f(x^k), x - x^k) + \frac{1}{2\nu} \|x - x^k\|^2), \quad (2)$$

$$\tilde{Q}_k = \{x: \tilde{g}(x^k) + \tilde{\nabla}g(x^k)(x - x^k) = 0\},$$

причем для всех  $x$  из некоторой окрестности  $U$  точки  $x^*$

$$\|\tilde{\nabla}f(x) - \nabla f(x)\| \leq \varepsilon_1, \quad \|\tilde{g}(x) - g(x)\| \leq \varepsilon_2, \quad \|\tilde{\nabla}g(x) - \nabla g(x)\| \leq \varepsilon_3.$$

**Теорема 1.** В условиях теоремы 1 § 2 существует такое  $\varepsilon_0 > 0$ , что для всякого  $\varepsilon > 0$  найдутся  $\delta_i > 0$ ,  $i = 1, 2, 3$ , для которых в методе (2) будет  $\|x^k - x^*\| \leq \varepsilon$  для всех достаточно больших  $k$ , если  $\|x^0 - x^*\| \leq \varepsilon_0$ ,  $\varepsilon_i < \delta_i$ ,  $i = 1, 2, 3$ .

Приведем схему доказательства. Так же как в § 2, можно показать, что  $x^{k+1} - x^* = D(x^k - x^*) + o(x^k - x^*) + r^k$ , где  $\|r^k\| = O(\varepsilon_1 + \varepsilon_2 + \varepsilon_3)$ ,  $\rho(D) < 1$ . Пусть  $U > I$  — решение матричного уравнения  $D^T U D = U - I$  (лемма 2 § 1 гл. 2),  $v_k = (U(x^k - x^*), x^k - x^*)$ . Тогда (ср. с доказательством теоремы 2 § 1 гл. 2) для достаточно малых  $v_0$  будет  $v_{k+1} \leq q v_k + \alpha_k$ , где  $q < 1$ ,  $\alpha_k = O(\varepsilon_1 + \varepsilon_2 + \varepsilon_3)$ , что и приводит к требуемому результату.  $\blacktriangleleft$

Совершенно аналогичные утверждения относятся и к другим методам § 2.

Однако если помехи не слишком малы, то может произойти полный «развал» методов. Пусть, например, при вычислении  $g_i(x)$  допускается систематическая ошибка  $\varepsilon_i$ ,  $i = 1, \dots, m$ . Это приводит к тому, что решается задача с ограничениями вида  $g_i(x) = \varepsilon_i$ ,  $i = 1, \dots, m$ . При этом может случиться, что такая система не имеет решения, даже если в исходной задаче существует невырожденное решение  $x^*$ . В такой ситуации любые методы могут привести к бессмысличному ответу (см. ниже п. 4).

Подчеркнем, что этот эффект отсутствовал в задачах безусловной минимизации (см. гл. 4).

Не будем подробно останавливаться на задачах с относительными помехами. Здесь многое зависит от того, какой смысл вкладывается в этот термин. Если предполагать, что

$$\|r(x)\| \leq a \|x - x^*\|, \quad (3)$$

где  $r(x)$  — всевозможные ошибки в вычислении градиентов и функций, то можно показать, что для невырожденного минимума при достаточно малых  $a$  сохранится сходимость методов § 2.

Если же считать, что ошибки удовлетворяют условиям типа

$$\|\tilde{V}f(x) - Vf(x)\| \leq \alpha \|Vf(x)\|, \quad (4)$$

то эта ситуация эквивалента по существу ситуации с абсолютными помехами, так как  $\tilde{V}f(x^*) \neq 0$ .

Наконец, своеобразное положение возникает при наличии случайных помех. Рассмотрим, например, метод типа Эрроу — Гурвица (см. (6) § 2):

$$\begin{aligned} x^{k+1} &= x^k - \psi_k (\tilde{V}f(x^k) + g'(x^k)^T y^k) \\ y^{k+1} &= y^k + \psi_k g(x^k), \end{aligned} \quad (5)$$

где  $\tilde{V}f(x^k) = Vf(x^k) + \xi^k$ , а  $\xi^k$  — независимые центрированные случайные помехи (для простоты предполагается, что  $g_i(x)$  и  $Vg_i(x)$  известны точно). Так как  $\xi^k$ , вообще говоря, не ограничена, то существует ненулевая вероятность «выброса» точки из области сходимости. В результате, используя теорему 6 § 2 гл. 2, удается лишь доказать следующее: если

$$\sum_{k=0}^{\infty} \psi_k = \infty,$$

$\sum_{k=0}^{\infty} \psi_k^2 < \infty$ , то в условиях теоремы 2 § 2 метод (5) сходится к решению с вероятностью  $1 - \delta$ , где  $\delta$  тем меньше, чем меньше дисперсия помех и чем точнее начальное приближение. Аналогичные результаты справедливы и при наличии помех в вычислении  $g_i(x)$  и  $Vg_i(x)$ , а также для ряда других методов из § 2.

**3. Вырожденный минимум.** Большинство результатов § 2 было доказано в предположении невырожденности минимума. Отклонения от этого предположения могут вызываться двумя причинами. Во-первых, в точке  $x^*$  может нарушиться условие линейной независимости градиентов ограничений. Для таких задач, вообще говоря, неверно правило множителей Лагранжа и неприменима теорема о касательном подпространстве. Поэтому все методы, основанные на обычной и модифицированной функции Лагранжа, а также на линеаризации ограничений, могут потерять работоспособность. Рассмотрим пример в  $\mathbf{R}^2$ , с которым мы уже сталкивались (упр. 1 § 1 и рис. 34):

$$\min x_2,$$

$$\begin{aligned} g_1(x) &= (x_1 - 1)^2 + x_2^2 - 1 = 0, \\ g_2(x) &= (x_1 + 1)^2 + x_2^2 - 1 = 0. \end{aligned} \quad (6)$$

Решением является  $x^* = \{0, 0\}$ , при этом  $Vg_1(x^*) = \{-2, 0\}$  и  $Vg_2(x^*) = \{2, 0\}$  линейно зависимы. Тогда в точке  $x^0 = \{\varepsilon, 0\}$  линеаризация ограничений приводит к множеству  $Q_0 = \{x; g_1(x^0) + (Vg_1(x^0), x - x^0) = 0, g_2(x^0) + (Vg_2(x^0), x - x^0) = 0\}$

$= \{0\}$ , которое пусто при любом  $\varepsilon \neq 0$ . Таким образом, метод линеаризации неприменим для сколь угодно близкого к  $x^*$  начального приближения. По этой же причине теряет смысл метод Ньютона (см. его запись (23) § 2). Наконец, все двойственные методы для данной задачи не могут сходиться, так как для них не существует множителей Лагранжа. По-видимому, можно построить методы, основанные на условиях экстремума для нерегулярного случая (теорема 1 § 1), однако этот подход пока не исследован.

Во-вторых, вырожденная точка минимума может быть регулярной, но вместо достаточного условия экстремума (13) § 1 в ней может быть выполнено лишь более слабое необходимое условие (10) § 1. Эта ситуация во многом близка вырожденному минимуму для задач без ограничений (см. § 1 гл. 6). Так, для задач с линейными ограничениями и выпуклой  $f(x)$  можно доказать сходимость метода линеаризации и методов, основанных на модифицированной функции Лагранжа, однако сходимости со скоростью геометрической прогрессии при этом утверждать нельзя. В то же время метод, использующий обычную функцию Лагранжа, может и не сходиться. Рассмотрим простейший пример в  $\mathbf{R}^1$ :  $\min f(x), g(x) = 0$ , где  $f(x) = x, g(x) = x$ . Тогда  $x^* = 0$  — решение,  $y^* = -1, L(x, y^*) \equiv 0$ , так что  $L''_{xx}(x, y^*) \equiv 0$ , и условия теоремы 2 § 2 не выполнены. Метод (6) § 2 принимает в данном случае вид  $x^{k+1} = x^k - y(y^k + 1)$ ,  $y^{k+1} = y^k + \psi_k$ , поэтому для  $\rho_k = (x^k - x^*)^2 + (y^k - y^*)^2$  получаем  $\rho_{k+1} = \rho_k(1 + \psi_k^2)$ , т. е.  $\rho_k \rightarrow \infty$  при любом  $\psi \neq 0$ .

Наименее чувствителен ко всем формам вырождения метод штрафных функций. В условиях его сходимости (теорема 6 § 2) нет требований невырожденности минимума (или даже его регулярности). Однако вырождение приводит к замедлению сходимости метода штрафных функций. Чтобы в этом убедиться, нет нужды строить специальные примеры — достаточно обратить внимание на его связь с методом регуляризации. Действительно, задача  $\min [f(x) + K\|g(x)\|^2]$  эквивалентна задаче  $\min [\varepsilon f(x) + \|g(x)\|_1^2]$  с  $\varepsilon = 1/K$ , а последняя может рассматриваться как задача тип  $\|g(x)\|^2$ , регуляризованная с помощью функции  $f(x)$ . Метод же регуляризации, как мы знаем (из § 1 гл. 6) может сходиться очень медленно в вырожденном случае.

**4. Противоречивость ограничений.** Может случиться, что задача (A), которую требуется решить, плохо поставлена — множество допустимых точек в ней пусто. Причины подобной ситуации различны. Нередко в технических и экономических задачах исходные требования к объекту являются завышенными, а потому противоречивыми. Иногда дело заключается в противоречиях имеющейся информации, ошибках в характеристиках объектов и т. д. Наконец, если задача (A) возникает как вспомогательная при решении более сложных задач, то

несовместность ограничений может возникнуть вследствие необходимости использования аппроксимации. Например, если при менять метод линеаризации (§ 2) для плохих начальных приближений, то получающиеся вспомогательные задачи могут не иметь решений.

Возникает вопрос — как поведут себя различные методы минимизации в таких условиях? Часть из методов попросту потеряет смысл. Так, метод линеаризации и метод Ньютона не будут определены корректно, так как задачи, решаемые в них на каждой итерации, будут иметь противоречивые ограничения. Другие методы формально будут применимы, но заведомо не будут сходиться. Так, двойственные методы (6), (8), (9), (11) § 2 расходятся, так как не существует точки, в которой  $g(x) = 0$ , а поэтому шаг по двойственным переменным не стремится к 0. Наиболее разумным образом будет вести себя метод штрафных функций. Пусть, например, ограничения линейны, а функция  $f(x)$  выпукла. Тогда, используя отмечавшуюся ранее связь между методом регуляризации и методом штрафных функций, можно доказать, что последний сходится к *псевдобрешено* задачи — той из точек минимума  $\|g(x)\|^2$ , для которой  $f(x)$  минимальна.

### Г л а в а 9.

#### ОБЩАЯ ЗАДАЧА МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ

В данной главе будет исследоваться *общая задача математического программирования*:

$$\begin{aligned} \min f(x), \quad x \in \mathbf{R}^n, \\ g_i(x) \leqslant 0, \quad i = 1, \dots, r, \\ g_i(x) = 0, \quad i = r + 1, \dots, m, \end{aligned} \quad (A)$$

$x \in Q,$

где  $Q \subset \mathbf{R}^n$  — «простое» множество (ср. с гл. 7), а  $g_i: Q \rightarrow \mathbf{R}$ ,  $i = 1, \dots, m$ . Точки, в которых удовлетворяются все ограничения, будем называть *допустимыми*. Частные случаи задачи (A) уже исследовались в гл. 7 ( $r = m = 0$ ) и гл. 8 ( $r = 0$ ,  $Q = \mathbf{R}^n$ ). Анализ будет производиться для двух основных классов задач (A) — *нелинейного программирования* ( $f(x)$ ,  $g_i(x)$  дифференцируемы,  $Q = \mathbf{R}^n$ ) и *выпуклого программирования* ( $f(x)$ ,  $g_i(x)$ ,  $i = 1, \dots, r$  выпуклы,  $r = m$ ,  $Q$  выпукло).

### § 1. Выпуклое программирование (теория)

**1. Сведения из выпуклого анализа.** Нам потребуются некоторые новые сведения сверх приведенных в § 1 гл. 5.

Ранее рассматривались лишь выпуклые функции, определенные на всем пространстве  $\mathbf{R}^n$ . Целесообразно расширить класс выпуклых функций, отказавшись от последнего условия. Пусть  $Q \subset \mathbf{R}^n$  — некоторое множество, а скалярная функция  $f(x)$  определена на  $Q$  и не определена вне  $Q$ . Будем называть  $Q$  *областью определения*  $f(x)$  и обозначать  $D(f)$  (от английского domain — область). Иногда употребляют обозначение  $\text{dom } f$  и термин *эффективная область*. Функция  $f(x)$  называется *выпуклой на*  $D(f)$ , если для любых  $x \in D(f)$ ,  $y \in D(f)$   $0 \leqslant \lambda \leqslant 1$  будет  $\lambda x + (1 - \lambda)y \in D(f)$  и

$$f(\lambda x + (1 - \lambda)y) \leqslant \lambda f(x) + (1 - \lambda)f(y). \quad (1)$$

Из определения следует, что  $D(f)$  для выпуклой функции  $f(x)$  — выпуклое множество.

Удобно несколько иначе подойти и к самому понятию функции. Будем считать, что функция может принимать не только конечные значения, но и значение  $+\infty$ , подчиняющееся

естественным правилам для арифметических действий и неравенств:

$$\alpha < \infty, \quad \infty + \alpha = \infty,$$

$$\alpha \cdot \infty = \infty \quad (\alpha > 0), \quad \infty \leqslant \infty,$$

$$\infty + \infty = \infty, \quad \max\{\alpha, \infty\} = \infty$$

для всех  $\alpha \in \mathbf{R}^1$ . Выражения  $\infty - \infty$ ,  $\infty / \infty$  не определены, но положим  $0 \cdot \infty = 0$ . С учетом этого соглашения доопределим выпуклую функцию  $f(x)$ , заданную на  $D(f) \subset \mathbf{R}^n$ , на все  $\mathbf{R}^n$ , положив

$$f(x) = +\infty \quad \text{для } x \in D(f). \quad (3)$$

Тогда неравенство (1) остается справедливым для всех  $x, y \in \mathbf{R}^n$ . Будем теперь под *выпуклой* понимать функцию  $f(x)$ , принимающую значения из  $\mathbf{R}^1 \cup \{+\infty\}$ , заданную на всем  $\mathbf{R}^n$  и

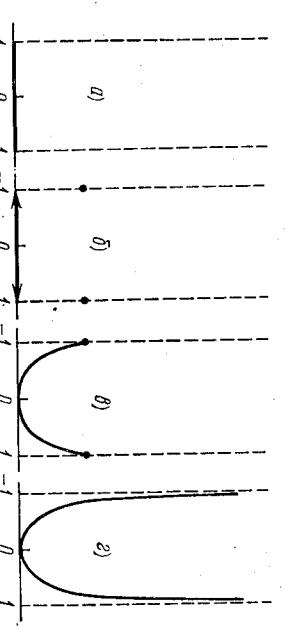


Рис. 35. Выпуклые функции с ограниченной областью определения.

удовлетворяющую неравенству (1) для всех  $x, y \in \mathbf{R}^n$ ,  $0 \leqslant \lambda \leqslant 1$ . При этом

$$D(f) = \{x: f(x) < \infty\}, \quad (4)$$

причем всегда будем предполагать, не оговаривая этого особо, что  $D(f)$  не пусто (иногда в этом случае говорят о *собственных выпуклых функциях*). Аналогично, функцию  $f(x)$  со значениями в  $\mathbf{R}^1 \cup \{-\infty\}$  будем называть *возногтой*, если  $-f(x)$  выпукла, при этом будем обозначать  $D(f) = \{x: f(x) > -\infty\}$ . В качестве примеров рассмотрим следующие четыре скалярные функции (рис. 35а—г):

$$f(x+y) \geqslant f(x) + f(y) \quad (9)$$

для всех  $y \in \mathbf{R}^n$ . Таким образом, если  $D(f) = \mathbf{R}^n$ , то это определение совпадает с (10) § 1 гл. 5. Однако при  $D(f) \neq \mathbf{R}^n$  возникают некоторые новые ситуации. Прежде всего очевидно, что  $\partial f(x)$  не существует в любой точке  $x \in D(f)$  (неравенство (9) не выполняется ни при каком  $a$  для  $x+a \in D(f)$ ). В границах точек  $D(f)$  субградиент может как существовать (пример (5)), так и не существовать (примеры (6)–(8)). Для внутренних же точек  $D(f)$  справедлив результат, аналогичный леммам 6 и 8 § 1 гл. 5.

**Лемма 3.** Для  $x \in D(f)^0$  множество  $\partial f(x)$  не пусто, выпукло, замкнуто и ограничено, причем для любого ограниченного множества из  $D(f)^0$  субградиенты равномерно ограничены.  $\blacktriangleleft$

Основные леммы о правилах вычисления субградиентов из § 1 гл. 5 остаются справедливыми при некоторых дополнительных предположениях.

$$f(x) = \begin{cases} 1 - \sqrt{1-x^2}, & |x| \leqslant 1, \\ \infty, & |x| > 1; \end{cases} \quad (7)$$

$$f(x) = \begin{cases} x^2(1-x^2)^{-1}, & |x| < 1, \\ \infty, & |x| \geqslant 1. \end{cases} \quad (8)$$

$$f(x) = \begin{cases} 0, & |x| \leqslant 1, \\ \infty, & |x| > 1; \end{cases} \quad (5)$$

$$f(x) = \begin{cases} 0, & |x| < 1, \\ 1, & |x| = 1, \\ \infty, & |x| > 1; \end{cases} \quad (6)$$

**Лемма 4** (Моро — Рокафеллар). *Пусть  $f_1(x), f_2(x)$  — выпуклые функции,  $\hat{f}(x) = f_1(x) + f_2(x)$  и  $D(f_1)^0 \cap D(f_2)^0 \neq \emptyset$ .*

Тогда

$$\partial\hat{f}(x) = \partial f_1(x) + \partial f_2(x). \blacksquare$$

По индукции получаем обобщение (11) на случай  $m$  функций: если  $\hat{f}_1(x), \dots, \hat{f}_m(x)$  выпуклы и

$$D(\hat{f}_1)^0 \cap \dots \cap D(\hat{f}_{m-1})^0 \cap D(\hat{f}_m)^0 \neq \emptyset,$$

то

$$\partial(\hat{f}_1(x) + \dots + \hat{f}_m(x)) = \partial f_1(x) + \dots + \partial f_m(x).$$

Лемма Моро — Рокафеллара является мощным инструментом при доказательстве разнообразных результатов выпуклого

$$(13)$$

анализа и теории экстремальных задач. Ограничимся пока одним важным примером. Напомним, что множество  $K \subset \mathbf{R}^n$ , называемое *конусом*, если для всех  $x \in K$  будет  $\lambda x \in K$  при любом  $\lambda > 0$ . Примерами конусов могут служить подпространство, полупространство, *прямой круговой конус*

$$\delta_Q(x) = \begin{cases} 0, & x \in Q, \\ \infty, & x \notin Q \end{cases}$$

(см. упр. 5). Тогда  $f(x) = f_1(x) + \dots + f_m(x)$ ,  $\partial f_i(0) = K_i^*$ ,  $\partial f(0) = K^*$  (см. упр. 6),  $D(f_i) = K_i$ ,  $D(f) = K$ , поэтому из (15) следует (12), а из (13) получаем (16).  $\blacksquare$

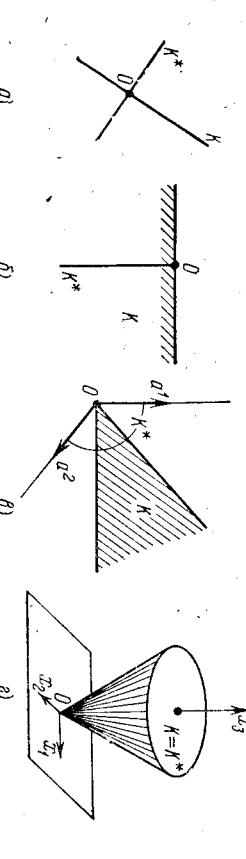
Условие (15) не может быть, вообще говоря, отброшено (упр. 4). Однако в важнейшем частном случае (когда все  $K_i$  — замкнутые полупространства) от него удается освободиться.

**Лемма 6** (Фаркаш). *Пусть  $K_i = \{x \in \mathbf{R}^n : (a^i, x) \geq 0\}$ ,*

$i = 1, \dots, m$ ,  $K = K_1 \cap \dots \cap K_m \neq \emptyset$ . Тогда

$$K^* = K_1^* + \dots + K_m^* = \left\{ \sum_{i=1}^m y_i a^i, y_i \geq 0 \right\}.$$

Рис. 36. Конусы и сопряженные конусы: а) подпространство, б) полуправило, в) многогранный конус, г) прямой круговой конус.



Пусть  $A$  — матрица  $m \times n$ , строками которой являются  $a^i$ . Тогда требуемый результат может быть записан в виде:

если  $K = \{x \in \mathbf{R}^n : Ax \geq 0\}$ ,  
то  $K^* = \{x \in \mathbf{R}^n : x = A^T y, y \geq 0, y \in \mathbf{R}^m\}$ . (18)

В этих же терминах лемма 1 § 1 гл. 8 может быть записана так: если  $K = \{x \in \mathbf{R}^n : Ax = 0\}$ ,  
то  $K^* = \{x \in \mathbf{R}^n : x = A^T y, y \in \mathbf{R}^m\}$ . (19)

Доказательство (18) может быть проведено по той же схеме, что и (19) (см. доказательство леммы 1 § 1 гл. 8). При этом, однако, приходится использовать замкнутость многогранного конуса вида  $\{x = A^T y, y \geq 0\}$  — этот факт требует специального доказательства.  $\blacksquare$

Конус  $K^*$  называется *сопряженным* к конусу  $K \subset \mathbf{R}^n$ , если (рис. 36)

$$K^* = \{a \in \mathbf{R}^n : (a, x) \geq 0 \forall x \in K\}. \quad (14)$$

Конус  $-K^*$  иногда называют *полярой* конуса  $K$ .

**Лемма 5** (Лубовицкий — Милотин). *Пусть  $K_1, \dots, K_m$  — выпуклые конусы в  $\mathbf{R}^n$ ,  $K = K_1 \cap \dots \cap K_m$  и*

$$K_1^0 \cap \dots \cap K_{m-1}^0 \cap K_m \neq \emptyset.$$

Тогда

$$K^* = K_1^* + \dots + K_m^*.$$

Доказательство. Введем  $f_i(x) = \delta_{K_i}(x)$ ,  $f(x) = \delta_K(x)$ , где  $\delta_Q(x) — индикаторная функция$  множества  $Q$ :

$$(16)$$

Пусть  $K_i = \{x \in Q : (a^i, x) \geq 0\}$ ,  $i = 1, \dots, m$ ,  $K = K_1 \cap \dots \cap K_m \neq \emptyset$ . Тогда

$$(17)$$

Доказательство леммы 6. Пусть  $f(x) = f_1(x) + \dots + f_m(x)$ ,  $\partial f_i(0) = K_i^*$ ,  $\partial f(0) = K^*$  (см. упр. 6),  $D(f_i) = K_i$ ,  $D(f) = K$ , поэтому из (15) следует (12), а из (13) получаем (16).  $\blacksquare$

Условие (15) не может быть, вообще говоря, отброшено (упр. 4). Однако в важнейшем частном случае (когда все  $K_i$  — замкнутые полупространства) от него удается освободиться.

**Лемма 6** (Фаркаш). *Пусть  $K_i = \{x \in \mathbf{R}^n : (a^i, x) \geq 0\}$ ,  $i = 1, \dots, m$ ,  $K = K_1 \cap \dots \cap K_m \neq \emptyset$ . Тогда*

$$(18)$$

Доказательство (18) может быть проведено по той же схеме, что и (19) (см. доказательство леммы 1 § 1 гл. 8). При этом, однако, приходится использовать замкнутость многогранного конуса вида  $\{x = A^T y, y \geq 0\}$  — этот факт требует специального доказательства.  $\blacksquare$

Из леммы Фаркаша следует, что если все конусы  $K_i$  многоугольные, то в лемме Дубовицкого — Милютина можно отбросить требование (15). Аналогичным образом может быть усиlena лемма Моро — Рокафеллара. Лемма Фаркаша дает возможность выписать вид опорных векторов к многогранному множеству.

**Л е м м а 7.** *Пусть  $Q = \{x \in \mathbf{R}^n : (a^i, x) \leq b_i, i = 1, \dots, m\}$ ,  $x^* \in Q$ ,  $I^* = \{i : (a^i, x^*) = b_i\}$  — множество активных ограничений в  $x^*$ ,  $K = \{c : (c, x - x^*) \geq 0 \forall x \in Q\}$  — конус опорных векторов к  $Q$  в точке  $x^*$ . Тогда*

$$K = \left\{ \sum_{i \in I^*} y_i a^i, y_i \geq 0, i \in I^* \right\}. \quad (20)$$

**Доказательство.** Пусть  $\Gamma = \{z : z = \lambda(x - x^*), \lambda \geq 0, x \in Q\}$  — конус, порожденный возможными направлениями (см. § 1 гл. 7). Тогда  $\Gamma = \{z : (a^i, z) \leq 0, i \in I^*\}$ . На основании леммы Фаркаша  $\Gamma^* = \left\{ \sum_{i \in I^*} y_i a^i, y_i \geq 0 \right\}$ . Если  $c \in \Gamma^*$ , то по определению сопряженного конуса  $\lambda(c, x - x^*) \geq 0$  для всех  $x \in Q$ ,  $\lambda \geq 0$ , это эквивалентно тому, что  $(c, x - x^*) \geq 0$  для всех  $x \in Q$ , т. е.  $\Gamma^* \subset K$ . Обратное включение очевидно.  $\blacktriangleleft$

В заключение выпишем вид опорного конуса для множества, задаваемого с помощью выпуклой функции.

**Л е м м а 8.** *Пусть  $f(x)$  — выпуклая функция,  $x^* \in D(f)^0$ ,  $Q = \{x : f(x) \leq 0\}$  и  $\inf f(x) < 0$ . Тогда  $Q^* = \{c : (c, x - x^*) \geq 0 \forall x \in Q\} = \{\partial f(x^*), \lambda \geq 0\}$ , если  $f(x^*) = 0$ , и  $Q^* = \{0\}$ , если  $f(x^*) < 0$ .*

Доказательство немедленно следует из определения субradienta и лемм 2 и 6 (при  $m = 1$ ).  $\blacktriangleleft$

#### Упражнения.

1. Докажите, что если  $Q_\alpha = \{x : f(x) \leq \alpha\}$  непусто и ограничено для некоторого  $\alpha \in \mathbf{R}^1$  и выпуклой  $f(x) \in D(f)^0 \supseteq Q_\alpha$ , то  $Q_\alpha$  ограничено для всех  $\alpha < \infty$  (ср. с леммой 1 § 2 гл. 5). Указание. Воспользуйтесь леммой 2.

2. Докажите лемму 4. (Указание. Включение  $\partial f(x) \supseteq \partial f_1(x) + \partial f_2(x)$  очевидно; для доказательства противоположного включения рассмотрите два множества в  $\mathbf{R}^{n+1}$ :  $Q_1 = \{\alpha \in \mathbf{R}^1, z \in \mathbf{R}^n : \alpha \geq f_1(z) + f_2(z)\}$ ,  $Q_2 = \{\alpha \in \mathbf{R}^1, z \in \mathbf{R}^n : \alpha < f_1(z) + f_2(z)\}$  и примените к ним теорему открытости).

3. Докажите, что ортант и прямой круговой конус являются самосопряженными, т. е.  $K = K^*$ .

4. Пусть  $K_1 = \{x \in \mathbf{R}^n : (a, x) \geq 0\} \cup \{0\}$ ,  $K_2 = \{x : (a, x) = 0\}$ ,  $a \neq 0$ ,  $a \in \mathbf{R}^n$ . Покажите, что равенство  $(K_1 \cap K_2)^* = K_1^* + K_2^*$  в этом случае неверно.

5. Покажите, что индикаторная функция выпуклого множества  $Q$  (17) является выпуклой и имеет субградиент в любой точке из  $Q$ , причем  $\partial \delta_Q(x)$  совпадает с конусом векторов, опорных к  $Q$  в точке  $x$ .

6. Пусть  $K$  — выпуклый конус. Покажите, что  $\partial \delta_K(0) = K^*$ .

**2. Теорема Куна — Таккер.** Прежде всего приведем признак безусловного минимума выпуклой функции, частным случаем которого является теорема 1 § 2 гл. 5.

**Л е м м а 9.** *Пусть  $f(x)$  — выпуклая функция. Точка  $x^*$  — точка глобального минимума  $f(x)$  на  $\mathbf{R}^n$  тогда и только тогда, когда  $0 \in \partial f(x^*)$ .*

Доказательство следует непосредственно из определения субградиента. Обратите внимание, что из (21) вытекает существование субградиента в точке минимума.  $\blacktriangleleft$

Как и ранее, для выпуклых функций локальный минимум совпадает с глобальным.

**Л е м м а 10.** *Пусть  $f(x)$  — выпуклая функция,  $\varepsilon \geq 0$ ,  $x^* \in D(f)^0$  и  $f(x) \geq f(x^*)$  для всех  $x$  таких, что  $\|x - x^*\| \leq \varepsilon$ . Тогда*

$$\Delta \text{оказательство. Пусть } \|x - x^*\| > \varepsilon, \lambda = \varepsilon/\|x - x^*\| < 1, x_\varepsilon = \lambda x + (1 - \lambda)x^*. \text{ Тогда } \|x_\varepsilon - x^*\| = \varepsilon, \text{ так что } f(x_\varepsilon) \geq f(x^*), \text{ но из (1)} f(x_\varepsilon) \leq \lambda f(x) + (1 - \lambda)f(x^*), \text{ т. е. } f(x) \geq \lambda^{-1}f(x_\varepsilon) - f(x^*) + f(x^*) \leq f(x^*). \blacktriangleleft$$

Предположение  $x^* \in D(f)^0$  существенно (иначе в приведенном выше выкладке выражение  $f(x_\varepsilon) - f(x^*)$  может быть равно  $-\infty$  — и потерять смысл). Например, для функции (5) точка  $x = 2$  не является глобальным минимумом, хотя в ее окрестности  $f(x) = f(2) = \infty$ . Полезно в связи с этим обратить внимание на необходимость осторожности в выкладках, содержащих выражения, которые могут обратиться в  $\infty$ .

Рассмотрим общую задачу выпуклого программирования вида

$$\min f(x), \quad x \in \mathbf{R}^n, \\ g_i(x) \leq 0, \quad i = 1, \dots, m, \quad (22)$$

$$x \in Q$$

и сформулируем для нее необходимые и достаточные условия экстремума.

**Т е о р е м а 1** (Кун — Таккер). *Пусть  $f(x)$ ,  $g_i(x)$ ,  $i = 1, \dots, m$  — выпуклые функции,  $Q \subset D(g_i)^0$ ,  $i = 1, \dots, m$ , и выполняется условие Слейтера: найдется  $x_0^* \in Q$  такое, что*

$$g_i(x^0) < 0, \quad i = 1, \dots, m. \quad (23)$$

Тогда допустимая точка  $x^*$  является глобальным решением (22), если и только если найдутся  $y_i^* \geq 0$ ,  $i = 1, \dots, m$ , такие, что

$$\dot{y}_i g_i(x^*) = 0, \quad i = 1, \dots, m \quad \text{и} \\ L(x, y^*) \geq L(x^*, y^*) \quad \forall x \in Q, \quad (24)$$

$$\text{где } y^* = (y_1^*, \dots, y_m^*) \in \mathbf{R}^m, \quad g(x) = (g_1(x), \dots, g_m(x)) \text{ и} \\ L(x, y) = f(x) + (y, g(x)). \quad (25)$$

Функцию  $L(x, y)$ , как и для задач с равенствами (см. § 1 гл. 8), будем называть *функцией Лагранжа*. Вектор  $y^*$  — *множеством переменных*, условие  $y_i^*g_i(x^*) = 0$ ,  $i = 1, \dots, m$  — *условием дополняющей нежесткости*, набор-индексов  $I^* = \{i; g_i(x^*) = 0\}$  — *множеством активных ограничений*. Очевидно, что  $y_i^* = 0$  при  $i \in I^*$ . Если для задачи (22) выполняются условия теоремы 1, то будем говорить о *регулярной точке минимума* или *регулярной задаче*. Теорема Куна — Таккера утверждает, что в случае регулярной задачи минимума  $x^*$  найдутся неотрицательные множители Лагранжа  $y^*$ , удовлетворяющие условию дополняющей нежесткости, такие что функция Лагранжа при  $y = y^*$  достигает минимума на  $Q$  в точке  $x^*$ . Таким образом, появляется возможность свести задачу с неравенствами  $L_i(x) \leq 0$  к задаче минимизации без этих ограничений.

Доказательство. Достаточность. Пусть  $x$  — произвольная допустимая точка и выполняется (24), тогда

$$f(x) \geq f(x) + (y^*, g(x)) = L(x, y^*) \geq L(x^*, y^*) =$$

$$= f(x^*) + (y^*, g(x^*)) = f(x^*),$$

т. е.  $x^*$  — точка глобального минимума в (22). Отметим, что при этом не используется условие Слейтера (23).

Небходимость. Введем функции

$$f_0(x) = \delta_Q(x), \quad f_i(x) = \delta_{Q_i}(x),$$

$$Q_i = \{x; g_i(x) \leq 0\}, \quad i = 1, \dots, m,$$

$$F(x) = f(x) + \sum_{i=0}^m f_i(x),$$

где  $\delta_Q(x)$  — индикаторная функция (17). Тогда  $F(x) = f(x)$ , если  $x$  допустимая точка, и  $F(x) = \infty$  в противном случае. Поэтому задача (22) эквивалентна безусловной минимизации  $F(x)$ . Функция  $F(x)$  выпукла, поэтому по лемме 9  $0 \in \partial F(x^*)$ . Точка  $x^0$ , фиксирующаяся в (23), такова, что  $x^0 \in D(f_0^0)$ ,  $x^0 \in D(g_i^0)$ ,  $x^0 \in Q_i^0$  в силу леммы 2, и поэтому  $x^0 \in D(f_i^0)$ ,  $i = 1, \dots, m$ . Следовательно,  $D(f)^0 \cap D(f_0) \cap D(f_1)^0 \cap \dots \cap D(f_m)^0 \neq \emptyset$ , и поэтому к  $F(x)$  применима лемма Моро — Рокафеллара:  $\partial F(x^*) = \partial f(x^*) + \partial f_0(x^*) + \dots + \partial f_m(x^*)$ .

Но  $\partial f_0(x^*) = \partial \delta_Q(x^*)$  — конус опорных к  $Q$  в точке  $x^*$  (см. упр. 5),  $\partial f_i(x^*) = \{y_i^* \partial g_i(x^*), y_i \geq 0\}$ ,  $i \in I^*$ ;  $\partial f_i(x^*) = 0$ ,  $i \in I^*$  (лемма 8).

Поэтому найдутся такие числа  $y_i^* \geq 0$ ,  $i \in I^*$ , что

$$0 \in \partial f(x^*) + \partial \delta_Q(x^*) + \sum_{i \in I^*} y_i^* \partial g_i(x^*).$$

Введем вектор  $y^* = (y_1^*, \dots, y_m^*)$ , где  $y_i^*$  определены выше при  $i \in I^*$ ,  $y_i^* = 0$  при  $i \in I^*$ , и функцию

$$\Phi(x) = L(x, y^*) + \delta_Q(x) = f(x) + \delta_Q(x) + \sum_{i \in I^*} y_i^* g_i(x). \quad (28)$$

Лемму Моро — Рокафеллара вновь можно применить, так как  $x^* \in D(f)^0$ ,  $x^* \in D(g_i)^0$ ,  $i \in I^*$ ,  $x^* \in D(\delta_Q)$ , и мы получаем, что  $\partial \Phi(x^*) = \partial f(x^*) + \partial \delta_Q(x^*) + \sum_{i \in I^*} y_i^* \partial g_i(x^*)$ . Таким образом, (27) имеет вид  $0 \in \partial \Phi(x^*)$ , и по лемме 9  $x^*$  — точка безусловного минимума  $\Phi(x)$ . Это эквивалентно тому (см. (28)), что  $x^*$  — точка минимума  $L(x, y^*)$  на  $Q$ .  $\blacktriangle$

Отметим, что из (27) следует условие экстремума в субградиентной форме:

$$(\partial_x L(x^*, y^*), x - x^*) \geq 0 \quad \forall x \in Q,$$

$$\partial_x L(x^*, y^*) = \partial f(x^*) + \sum_{i \in I^*} y_i^* \partial g_i(x^*), \quad (29)$$

что является обобщением теоремы 3 § 1 гл. 7, в которой рассматривалась задача без ограничений типа неравенств. Если же  $Q = \mathbf{R}^n$ , то (29) принимает вид

$$0 \in \partial_x L(x^*, y^*). \quad (30)$$

Для теоремы Куна — Таккера известно и много других доказательств (например, непосредственно на основе теорем отдельности). Приведенное простое доказательство наглядно демонстрирует эффективность техники выпуклого анализа. Если попытаться воспользоваться условиями экстремума из гл. 7, рассматривая множество, задаваемое всеми ограничениями, то потребуется выписать опорный вектор к этому множеству. Именно это и составляет нетривиальную часть теоремы Куна — Таккера.

Условие Слейтера играет ту же роль, что и условие регулярности в задаче с равенствами (§ 1 гл. 8). При его невыполнении допустимое множество может оказаться слишком «толстым» и теорема Куна — Таккера не будет верна. Например, если задача (22) имеет вид (см. рис. 34)

$$\min x_2, \quad x \in \mathbf{R}^2,$$

$$(x_1 - 1)^2 + x_2^2 - 1 \leq 0,$$

$$(x_1 + 1)^2 + x_2^2 - 1 \leq 0,$$

то  $x^* = 0$ ,  $I^* = \{1, 2\}$  и, как нетрудно проверить, не существует таких  $y_1^*, y_2^*$ , чтобы выполнялось условие (30).

Если в задаче имеются линейные ограничения, то их удобно отнести к множеству  $Q$ . В частности, если других ограничений нет, т. е. задача имеет вид

$$\begin{aligned} \min f(x), \\ (a^i, x) \leq b_i, \quad i = 1, \dots, m, \end{aligned} \tag{32}$$

то все эти ограничения целесообразно рассматривать как за-

дающие многогранное множество  $Q$  и использовать лемму 7 о виде опорных к такому множеству.

**Теорема 2.** Если  $f(x)$  выпукла и  $D(f)^0$  содержит допустимое множество, то необходимое и достаточное условие экстремума в задаче (32) имеет вид: находится  $y_i^* \geq 0$ ,  $i \in I^* = \{i : (a^i, x^*) = b_i\}$  такие, что

$$\sum_{i \in I^*} y_i^* a^i \equiv -\partial f(x^*). \blacktriangle \tag{33}$$

Подчеркнем, что в этом случае не требуется условие Слейтера, т. е. допустимое множество в (32) может и не иметь внутренней точки.

Теорему Куна — Таккера часто записывают в несколько ином виде, в терминах седловой точки. Введем требуемые для этого понятия. Пусть  $Q \subset \mathbf{R}^n$ ,  $S \subset \mathbf{R}^m$  — два множества,  $\Phi: Q \times S \rightarrow \mathbf{R}^1$ . Пара  $x^* \in Q$ ,  $y^* \in S$  называется седловой точкой функции  $\Phi(x, y)$  на  $Q \times S$ , если

$$\Phi(x^*, y) \leq \Phi(x^*, y^*) \leq \Phi(x, y^*) \quad \forall x \in Q \text{ и } \forall y \in S. \tag{34}$$

Иначе говоря,  $x^*$  является точкой минимума  $\Phi(x, y^*)$  по  $x$  на  $Q$ , а  $y^*$  — точкой максимума  $\Phi(x^*, y)$  по  $y$  на  $S$ . Если ниже приведенные выражения определены, то равенство

$$\min_{x \in Q} \max_{y \in S} \Phi(x, y) = \max_{y \in S} \min_{x \in Q} \Phi(x, y) = \Phi(x^*, y^*) \tag{35}$$

эквивалентно (34), т. е. наличие седловой точки означает, что операции минимизации и максимизации можно переставлять. **Теорема 3** (Кун — Таккер). В условиях теоремы 1  $x^*$  является решением задачи (22) тогда и только тогда, когда пара  $x^*, y^*$  при некотором  $y^* \geq 0$  является седловой точкой  $L(x, y)$  на  $Q \times \mathbf{R}_+^m$ , т. е.

$$L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*) \quad \forall x \in Q \text{ и } \forall y \geq 0. \tag{36}$$

Доказательство. Пусть  $x^*$  — решение (22), по теореме 1 найдется  $y^* \geq 0$  такое, что  $(y^*, g(x^*)) \leq (y^*, g(x^*))$  для всех  $y \geq 0$  для всех  $x \in Q$ . Но тогда  $L(x^*, y^*) = f(x^*) \geq f(x^*) + (y, g(x^*)) = L(x^*, y)$  для любых  $y \geq 0$ , поскольку  $g(x^*) \leq 0$ . Итак, пара  $x^*, y^*$ , где  $y^*$  — решение (22), а  $y^*$  — множители Лагранжа, является седловой точкой  $L(x, y)$  на  $Q \times \mathbf{R}_+^m$ .

Обратно, пусть  $x^*, y^*$  — седловая точка. Тогда  $L(x^*, y^*) \geq L(x^*, y)$  означает, что  $(y, g(x^*)) \leq (y^*, g(x^*))$  для всех  $y \geq 0$ . Это возможно лишь тогда, когда  $g(x^*) \leq 0$ ,  $(y^*, g(x^*)) = 0$ . Поэтому для любого допустимого  $x$

$$L(x^*, y^*) = f(x^*) \leq L(x, y^*) = f(x) + (y^*, g(x)) \leq f(x),$$

что и означает, что  $x^*$  — решение (22).  $\blacktriangle$

### Упражнения.

7. Получите из теоремы 2 условия (8) § 1 гл. 7. (Указание. Запишите ограничения  $(a^i, x) = b_i$  в виде  $(a^i, x) \leq b_i$  —  $(a^i, x) \geq b_i$ .)

8. Проверьте, что функции  $\Phi(x, y) = xy$  и  $\Phi(x, y) = -xy$ ,  $x \in \mathbf{R}^1$ ,  $y \in \mathbf{R}^1$ , обе имеют единственную седловую точку  $(0, 0)$  на  $\mathbf{R}^1 \times \mathbf{R}^1$ . Убедитесь на примере из упражнения 8, что из (34) не следуют равенства  $X^* = \operatorname{Argmin}_{x \in Q} \Phi(x, y^*)$ ,  $Y^* = \operatorname{Argmax}_{y \in S} \Phi(x^*, y)$ , где  $X^* \times Y^*$  — множество седловых точек.

### 3. Двойственность.

В формулировку теоремы 3 прямые и двойственные переменные входят симметричным образом. Поэтому можно ожидать, что аналогичная симметрия существует и для задач оптимизации, т. е. что (36) является условием экстремума не только для исходной задачи (22), но и для другой задачи оптимизации относительно двойственных переменных. Такую задачу можно получить из следующих соображений. Введем

$$\Phi(x) = \sup_{y \geq 0} L(x, y), \tag{37}$$

тогда очевидно, что

$$\Phi(x) = \begin{cases} f(x), & \text{если } g_i(x) \leq 0, \quad i = 1, \dots, m, \\ \infty, & \text{в противном случае.} \end{cases}$$

Поэтому исходная задача может быть записана в виде

$$\min_{x \in Q} \Phi(x). \tag{38}$$

Поступим аналогичным образом, поменяв роль переменных и операций максимизации и минимизации. Именно, введем

$$\Psi(y) = \inf_{x \in Q} L(x, y) \tag{39}$$

(возможно, что  $\Psi(y) = -\infty$  для некоторых  $y$ ) и рассмотрим задачу

$$\max_{y \geq 0} \Psi(y). \tag{40}$$

Задача (40) называется *двойственной*, а (38) или (22) — *прямой*. **Теорема 4** (теорема двойственности). *Справедливы следующие соотношения двойственности:*

а) Для любых допустимых  $x$  и  $y$  (т. е. для  $x \in Q$ ,  $g_i(x) \leq 0$ ,  $i = 1, \dots, m$ ,  $y \geq 0$ )

$$\begin{aligned} & f(x) \geq \Psi(y). \\ & \text{б) Если прямая задача регулярна, } x^* \text{ — ее решение, } y^* \text{ — множители Лагранжа, то } y^* \text{ — решение (40) и} \\ & f(x^*) = \Psi(y^*). \end{aligned} \quad (41)$$

в) Если для допустимых  $x^*, y^*$  имеет место (42), то  $x^*$  — решение прямой, а  $y^*$  — решение двойственной задачи.

Доказательство. а) Если  $x \in Q$ ,  $g(x) \leq 0$ ,  $y \geq 0$ , то

$$f(x) \geq f(x) + (y, g(x)) = L(x, y) \geq \inf_{x' \in Q} L(x', y) = \Psi(y).$$

б) Пусть  $x^*$  — решение (22),  $y^*$  — множители Лагранжа, тогда по теореме 2

$$\Psi(y^*) = \inf_{x \in Q} L(x, y^*) = L(x^*, y^*) \geq L(x^*, y) \geq \inf_{x \in Q} L(x, y) = \Psi(y)$$

для всех  $y \geq 0$ , т. е.  $y^*$  — решение (40), при этом, поскольку  $L(x^*, y^*) = f(x^*)$ , то  $\Psi(y^*) = f(x^*)$ .

в) Пусть  $g(x^*) \geq 0$ ,  $x^* \in Q$ ,  $y^* \geq 0$  и  $f(x^*) = \Psi(y^*)$ , тогда для произвольных допустимых  $x$ ,  $y$  в силу (41)  $f(x) \geq \Psi(y^*) = f(x^*) \geq \Psi(y)$ , т. е.  $x^*, y^*$  — решения (22) и (40).  $\blacktriangleleft$

Рассмотрим несколько примеров. Для задачи в  $\mathbf{R}^2$

$$\min x_1, \quad x_2 \leq 0, \quad (43)$$

имеем  $L(x, y) = x_1 + x_2 y$ ,  $\Psi(y) = \inf_{x \in \mathbf{R}^2} L(x, y) = -\infty$ . Здесь ни

прямая, ни двойственная задачи не имеют решения. Пусть

$$f(x) = 1/x, \quad x \in \mathbf{R}^1, \quad g(x) = -x \leq 0. \quad (44)$$

Тогда  $L(x, y) = 1/x - xy$ ,  $\Psi(y) = -\infty$  для  $y > 0$ ,  $\Psi(0) = 0$ . Здесь двойственная задача имеет решение  $y^* = 0$ , а прямая — нет. Наконец, если задача имеет вид

$$\min x, \quad x^2 \leq 0, \quad x \in \mathbf{R}^1, \quad (45)$$

то  $L(x, y) = x + yx^2$ ,  $\Psi(y) = -(4y)^{-1}$ , и прямая задача имеет решение  $x^* = 0$ , а двойственная — нет. Все эти «патологические» примеры показывают, что в общем случае соотношение между прямой и двойственной задачами может быть достаточно произвольным. Однако в регулярном случае, в соответствии с теоремой 4б), обе задачи одновременно разрешимы и их оптимальные значения равны.

Полезность теоремы двойственности по сравнению с обычными условиями экстремума типа теорем 1—3 определяется следующими факторами. Во-первых, удается свести исходную за-

дачу к другой, которая может оказаться проще. Так, если  $m \ll n$ , то размерность двойственной задачи (равная  $m$ ) существенно меньше размерности прямой. Во-вторых, неравенство (41) позволяет получить оценку снизу для минимума в (22) и тем самым оценить точность приближенного решения. Конечно, плодовитость двойственного подхода в большей мере зависит от того, насколько просто вычисляется функция  $\Psi(y)$ . В ряде случаев (в частности, для задач линейного, квадратичного, сепарабельного и геометрического программирования, см. гл. 10, 11), двойственный подход оказывается очень эффективным.

Отметим, что двойственную задачу (40) будем иногда записывать в иной форме. Во-первых, максимум  $\Psi(y)$  может достигаться лишь в тех точках, где  $\Psi(y) \neq -\infty$ , поэтому (40) равносильна задаче

$$\max \Psi(y), \quad y \geq 0, \quad y \in D(\Psi), \quad (46)$$

где  $D(\Psi) = \{y: \Psi(y) > -\infty\}$ . Во-вторых, привычнее иметь дело с задачами минимизации, а не максимизации. Если ввести

$$\theta(y) = -\Psi(y), \quad (47)$$

то вместо (40) получим выпуклую (упр. 10) задачу

$$\min \theta(y), \quad y \geq 0, \quad (48)$$

либо вместо (46) — задачу

$$\min \theta(y), \quad y \geq 0, \quad y \in D(\theta); \quad (49)$$

при этом соотношение двойственности (41) приобретает вид

$$f(x) + \theta(y) \geq 0. \quad (50)$$

Наконец, обратим внимание на то, что в зависимости от того, какие ограничения записаны в виде  $g_i(x) \leq 0$ , а какие отнесены к множеству  $Q$ , мы получим различные двойственные задачи. Вообще, для каждой задачи оптимизации существует много различных двойственных задач, но мы не будем останавливаться на общей теории их формирования.

#### Упражнения.

10. Докажите следующие свойства  $\Psi(y)$ :

- а) множество  $D(\Psi) = \{y: \Psi(y) > -\infty\}$  выпукло, функция  $\Psi(y)$  вогнута на  $D(\Psi)$ ;
- б) если  $f(x), g_i(x)$  непрерывны,  $Q$  замкнуто и ограничено, то  $D(\Psi) = \mathbf{R}^n$  и  $\Psi(y)$  непрерывна;
- в) если  $f(x), g_i(x)$  выпуклы,  $Q$  выпукло,  $Q \subset D(f)^0, Q \subset D(g_i)^0, f(x)/\|x\| \rightarrow \infty$  при  $x \in Q$ ,  $\|x\| \rightarrow \infty$ , то  $\mathbf{R}_+^m \subset D(\Psi)$ ;
- г) если  $f(x)$  строго выпукла,  $g_i(x)$  выпуклы,  $Q$  выпукло, замкнуто и ограничено, то  $\Psi(y)$  дифференцируема при  $y \geq 0$  и  $\nabla \Psi(y) = g(x(y))$ , где  $x(y) = \arg \min_{x \in Q} L(x, y)$ ;
- д) если, кроме того,  $f(x)$  сильно выпукла, то  $\nabla \Psi(y)$  удовлетворяет условию Липшица.

- 11.** Напишите двойственные к задачам:
- $\min_{x} \{c, x\}, \|x\| \leq 1;$
  - $\min_{x} \{c, x\}, \|x\|^2 \leq 1$ . Убедитесь, что получаются разные задачи, хотя исходные задачи эквивалентны.

#### 4. Существование, единственность и устойчивость решения.

Лемма 2 и результат упражнения 1 позволяют сформулировать следующий вариант теоремы Вейерштрасса для задачи (22).

**Теорема 5.** Пусть  $f(x), g_i(x), i = 1, \dots, m$ , — выпуклые функции,  $Q$  выпукло и замкнуто, для  $S = \{x \in Q : g_i(x) \leq 0, i = 1, \dots, m\}$  имеет  $S \subset D(f)^0$ ,  $S \subset D(g_i)^0, i = 1, \dots, m$ , и множество  $\{x \in S : f(x) \leq \alpha\}$  непусто и ограничено для некоторого  $\alpha$ . Тогда решение (22) существует.  $\blacktriangleleft$

Единственность решения, как обычно, можно гарантировать для строго выпуклой функции  $f(x)$ . Кроме того, если множество  $Q$  строго выпукло, а  $\|\partial_x L(x, y^*)\| \geq \varepsilon > 0$  для всех  $x \in Q$  и выполняются условия теоремы 1, то нетрудно получить, что решение единственно. Наконец, если какая-либо из функций  $g_i(x)$  строго выпукла, а отвечающий ей множитель Лагранжа положителен  $y_i^* > 0$ , то решение также единственно.

Что касается двойственной задачи, то предположений регулярности недостаточно для единственности ее решения. Например, для задачи

$$\min x, x \in \mathbf{R}^1, g_1(x) = -x \leq 0, g_2(x) = x^2 - 2x \leq 0 \quad (51)$$

множители Лагранжа определяются неоднозначно. Грубо говоря, это связано с тем, что первое ограничение является линейным — оно не меняет допустимой области. В то же время при условиях регулярности можно утверждать ограниченность решения единственной задачи.

Перейдем к анализу устойчивости. Нам понадобится следующий результат о непрерывной зависимости множества решений системы выпуклых неравенств от их правых частей.

**Лемма 11.** Пусть  $g_i(x), i = 1, \dots, m$ , — выпуклые функции, множество

$$S = \{x : g_i(x) \leq 0, i = 1, \dots, m\} \quad (52)$$

непусто и ограничено,  $D(g_i)^0 \supseteq S$ . Тогда:

$$S_\varepsilon = \{x : g_i(x) \leq \varepsilon, i = 1, \dots, m\} \quad (53)$$

непусто и ограничено для любых  $\varepsilon > 0$ ,  $\varepsilon = \{\varepsilon_1, \dots, \varepsilon_m\}$  и  $\Delta(S_\varepsilon, S) \rightarrow 0$  при  $\varepsilon \rightarrow +0$ ;  $\delta$ ) если при этом существует точка  $x^0$  такая, что  $g_i(x^0) \leq -\delta < 0$ , то  $S_\varepsilon$  непусто и ограничено при любых  $\varepsilon_i \geq g_i(x^0), i = 1, \dots, m$ , и

$$\rho(x_e, S) \leq c(\max \varepsilon_i)_+ \quad \forall x_e \in S_e, c = \frac{1}{\delta} \max_{x \in S} \|x - x^0\|. \quad \blacktriangleleft \quad (54)$$

Здесь  $\Delta(S_\varepsilon, S) — хаусдорфово расстояние между множествами  $S_\varepsilon$  и  $S$ , т. е.$

$$\Delta(S_\varepsilon, S) = \max \left\{ \max_{x \in S_\varepsilon} \rho(x, S), \max_{x \in S} \rho(x_\varepsilon, S) \right\}$$

Лемма 11 дает возможность немедленно получить результат о слабой устойчивости (см. § 3 гл. 1) задачи выпуклого программирования при возмущении ограничений.

**Теорема 6.** Пусть  $X^* = \operatorname{Argmin}_{x \in S} f(x)$ ,  $X_\varepsilon = \operatorname{Argmin}_{x \in S_\varepsilon} f(x)$ , где  $S$  и  $S_\varepsilon$  имеют вид (52) и (53),  $f(x), g_i(x)$  выпуклы и непрерывны на  $D(f)^0$  и  $D(f)^0 \supseteq S, D(g_i)^0 \supseteq S, i = 1, \dots, m$ , причем

а) при всех  $\varepsilon > 0$  множество  $X_\varepsilon^*$  непусто и ограничено и  $\Delta(X_\varepsilon, X^*) \rightarrow 0$ ,  $f_\varepsilon^* \rightarrow f^*$  при  $\varepsilon \rightarrow +0$ , где  $f^* = f(x^*)$ ,  $x^* \in X^*$ ,

б) если существует  $x^0 : g_i(x^0) < 0, i = 1, \dots, m$ , то  $X_\varepsilon^*$  непусто и ограничено при любых  $\varepsilon_i \geq g_i(x^0), i = 1, \dots, m$ , и  $\Delta(X_\varepsilon^*, X^*) \rightarrow 0$ ,  $f_\varepsilon^* \rightarrow f^*$  при  $\varepsilon \rightarrow 0$ .

Доказательство. Множество  $X^*$  можно записать в виде  $X^* = \{x : g_i(x) \leq 0, i = 0, 1, \dots, m\}$ , где  $g_0(x) = f(x) - f^*$ . В соответствии с леммой 11 множество  $X_\varepsilon = \{x : g_i(x) \leq \varepsilon_i, i = 0, 1, \dots, m\}$  (где  $\varepsilon_0 = 0$  в случае а) и  $\varepsilon_0 = f(x^0) - f^*$  в случае б)) является ограниченным и непустым. Но минимизация  $f(x)$  на  $S_\varepsilon$  эквивалентна минимизации  $f(x)$  на  $X_\varepsilon$ . Множество  $X_\varepsilon$  по доказанному ограничено и непусто, в силу выпуклости и непрерывности  $f(x)$  и  $g_i(x)$  оно выпукло и замкнуто. Поэтому минимум  $f(x)$  на  $X_\varepsilon$  достигается, т. е.  $X_\varepsilon^*$  непусто и ограничено. Поскольку  $f(x)$  непрерывна в окрестности  $X^*$ , то получаем  $f_\varepsilon^* \rightarrow f^*$ . Записав  $X_\varepsilon^*$  в виде  $X_\varepsilon^* = \{x : g_i(x) \leq \varepsilon_i, i = 0, 1, \dots, m\}$ , где  $\varepsilon_0 = f^* - f_\varepsilon^*$  и вновь применяя лемму 11, получаем, что  $\Delta(X_\varepsilon^*, X^*) \rightarrow 0$ .  $\blacktriangleleft$

Мы не останавливаемся на других результатах об устойчивости (сильная устойчивость решений, устойчивость по отношению к возмущениям минимизируемой функции, устойчивость множителей Лагранжа, более общие виды возмущений и т. д.). Отметим лишь, что при исследовании подобных проблем очень полезна бывает теорема двойственности. Пусть, например, вместо задачи (22) рассматривается возмущенная задача

$$\min f(x), g_i(x) \leq \varepsilon_i, i = 1, \dots, m, x \in Q. \quad (55)$$

Составим для нее функцию Лагранжа

$$L_\varepsilon(x, y) = f(x) + y, g(x) - \varepsilon = L(x, y) - (\varepsilon, y) \quad (56)$$

и двойственную задачу

$$\max_{y \geq 0} \Psi_\varepsilon(y), \Psi_\varepsilon(y) = \inf_{x \in Q} L_\varepsilon(x, y) = \Psi(y) - (\varepsilon, y). \quad (57)$$

Таким образом, проблема свелась к исследованию устойчивости двойственной задачи при возмущении целевой функции.

#### Упражнение.

12. Рассмотрите пример в  $\mathbf{R}^2$ , где исходная задача:  $\min_{\mathbf{x}} \mathbf{x}_1 - \mathbf{x}_2 \leq 0$ , возмущенная:  $\min_{\mathbf{x}} (\mathbf{x}_1 + \varepsilon_1 \mathbf{x}_2, -\mathbf{x}_1 - \varepsilon_2 \mathbf{x}_2 \leq 0$ . Убедитесь, что в исходной задаче при решении существует  $(\mathbf{x}_1^*, \mathbf{x}_2^*$  произвольно), а в возмущенной задаче при  $\varepsilon_1 \neq \varepsilon_2$  решения нет.

## § 2. Нелинейное программирование (теория)

Ниже рассматривается общая задача нелинейного программирования

$$\begin{aligned} & \min f(\mathbf{x}), \\ & g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, r, \\ & g_i(\mathbf{x}) = 0, \quad i = r+1, \dots, m, \end{aligned} \quad (1)$$

где все функции предполагаются дифференцируемыми, но не обязательно выпуклыми.

**1. Необходимые условия минимума.** Для произвольной допустимой точки  $\mathbf{x}^*$  введем множества индексов

$$\begin{aligned} I^* &= \{i: g_i(\mathbf{x}^*) = 0, i = 1, \dots, r\}, \\ I &= \{i: g_i(\mathbf{x}^*) = 0, i = 1, \dots, m\}, \end{aligned} \quad (2)$$

характеризующих *активные ограничения* ( $I^*$  относится к неравенствам,  $I$  — ко всем ограничениям).

**Теорема 1** (Каруш — Джон). *Пусть  $\mathbf{x}^*$  — точка локального минимума в (1), а функции  $f(\mathbf{x})$ ,  $g_i(\mathbf{x})$ ,  $i = 1, \dots, m$ , непрерывно дифференцируемы в окрестности  $\mathbf{x}^*$ . Тогда найдутся числа  $y_0^*$ ,  $y_i^*$ ,  $i \in I$ , не все равные 0, такие, что  $y_0^* \geq 0$ ,  $y_i^* \geq 0$ ,  $i \in I^*$ , и*

$$y_0^* \nabla f(\mathbf{x}^*) + \sum_{i \in I} y_i^* \nabla g_i(\mathbf{x}^*) + \sum_{i \in I^*} y_i^* z_i \geq 0. \quad (3)$$

**Доказательство.** Можно предположить, что  $\nabla g_i(\mathbf{x}^*)$ ,  $i = r+1, \dots, m$ , линейно независимы, иначе утверждение теоремы тривиально (можно взять  $y_0^* = 0$ ,  $y_i^* = 0$ ,  $i \in I^*$ ). Построим два множества в  $\mathbf{R}^{m+1}$ , причем будем обозначать компоненты  $\mathbf{z} \in \mathbf{R}^{m+1}$  через  $\{z_0, z_1, \dots, z_m\}$ :

$$\begin{aligned} A &= \{\mathbf{z} \in \mathbf{R}^{m+1}: z_0 = (\nabla f(\mathbf{x}^*), s), z_i = (\nabla g_i(\mathbf{x}^*), s), i \in I, s \in \mathbf{R}^n\}, \\ B &= \{\mathbf{z} \in \mathbf{R}^{m+1}: z_0 < 0, z_i < 0, i \in I^*; z_i = 0, i = r+1, \dots, m\}. \end{aligned}$$

Покажем, что эти множества не пересекаются. Пусть это не так, тогда найдется  $s \in \mathbf{R}^n$  такое, что

$$\begin{aligned} & (\nabla f(\mathbf{x}^*), s) < 0, \\ & (\nabla g_i(\mathbf{x}^*), s) < 0, \quad i \in I^*, \\ & (\nabla g_i(\mathbf{x}^*), s) = 0, \quad i = r+1, \dots, m. \end{aligned} \quad (4)$$

По теореме Люстерника (теорема 3 § 1 гл. 8) найдутся точки  $\mathbf{x}_\lambda$  такие, что  $g_i(\mathbf{x}_\lambda) = 0$ ,  $i = r+1, \dots, m$ ,  $\mathbf{x}_\lambda = \mathbf{x}^* + \lambda s + o(\lambda)$ . Тогда в силу (4)

$$\begin{aligned} f(\mathbf{x}_\lambda) &= f(\mathbf{x}^*) + \lambda (\nabla f(\mathbf{x}^*), s) + o(\lambda) < f(\mathbf{x}^*) \\ &\text{при достаточно малых } \lambda > 0. \end{aligned}$$

С другой стороны, по той же причине  $g_i(\mathbf{x}_\lambda) < 0$  при  $i \in I^*$  и достаточно малых  $\lambda > 0$ . Наконец,  $g_i(\mathbf{x}_\lambda) < 0$  при  $i \in I$  и малых  $\lambda$  в силу непрерывности  $g_i(\mathbf{x})$  и того факта, что  $g_i(\mathbf{x}^*) < 0$ ,  $i \in I$ . Итак, точка  $\mathbf{x}_\lambda$  при малых  $\lambda > 0$  удовлетворяет всем ограничениям, и для нее  $f(\mathbf{x}_\lambda) < f(\mathbf{x}^*)$ , что противоречит определению локального минимума. Следовательно,  $A$  и  $B$  не пересекаются.

Множества  $A$  и  $B$  очевидным образом выпуклы и непусты. По теореме отделимости найдется  $y^* \in \mathbf{R}^{m+1}$ ,  $y^* \neq 0$ , такое, что  $(y^*, z) \geq 0$ ,  $z \in A$ ,  $(y^*, z) \leq 0$ ,  $z \in B$  (в правой части неравенств можно взять 0, так как  $A$  и  $B$  конусы). Обозначим компоненты  $y^*$  через  $y_0^*$ ,  $y_1^*$ ,  $\dots$ ,  $y_m^*$ . Тогда неравенство  $(y^*, z) \geq 0$  для всех  $z \in A$  означает, что

$$y_0^* \nabla f(\mathbf{x}^*) + \sum_{i \in I} y_i^* \nabla g_i(\mathbf{x}^*) + \sum_{i \in I^*} y_i^* z_i \geq 0$$

для всех  $s \in \mathbf{R}^n$  и всех  $z_i \in \mathbf{R}$ ,  $i \in I$ . Это возможно лишь тогда, когда  $y_0^* \nabla f(\mathbf{x}^*) + \sum_{i \in I} y_i^* \nabla g_i(\mathbf{x}^*) = 0$  и  $y_i^* = 0$  при  $i \in I$ .

Наконец, поскольку для всех  $z \in B$

$$(y^*, z) = y_0^* z_0 + \sum_{i \in I^*} y_i^* z_i \leq 0$$

при любых  $z_0 < 0$ ,  $z_i < 0$ ,  $i \in I^*$ , то  $y_0^* \geq 0$ ,  $y_i^* \geq 0$ ,  $i \in I^*$ .  $\blacktriangleleft$

Другой путь доказательства теоремы указан в упражнении 2.

Нас будут интересовать условия (они называются *условиями регулярности*), при которых можно гарантировать, что  $y_0^* \neq 0$ . Так как  $y_i^*$  определены с точностью до положительного множителя, то без ограничения общности можно тогда считать  $y_0^* = 1$ . Иначе говоря, необходимое условие экстремума приобретает вид

$$\begin{aligned} L'_x(\mathbf{x}^*, y^*) &= 0, \quad y_i^* \geq 0, \quad i \in I^*, \quad (y^*, g(\mathbf{x}^*)) = 0, \\ L(x, y) &= f(\mathbf{x}) + (y, g(\mathbf{x})), \quad y^* = (y_1^*, \dots, y_m^*), \end{aligned} \quad (5)$$

Функцию  $L(x, y)$  мы по-прежнему будем называть *функцией Лагранжа*, числа  $y_i^*$ ,  $i = 1, \dots, m$  — *множителями Лагранжа*.

Простейшее из условий регулярности имеет следующий вид: Условие регулярности А: векторы  $\nabla g_i(\mathbf{x}^*)$ ,  $i \in I$ , либо независимы.

**Теорема 2.** Пусть выполнены условия теоремы 1 и условие регулярности А. Тогда находится  $y^* \in \mathbf{R}^m$ , для которого выполняется (5).

Доказательство немедленно следует из (3), так как предположение  $y^* = 0$  противоречит условию регулярности А.  $\blacktriangleleft$

Условие А является довольно ограничительным и не всегда выполняется.

**Условие регулярности Б:** векторы  $\nabla g_i(x^*)$ ,  $i = r+1, \dots, m$ , линейно независимы, и находится вектор  $s^0 \in \mathbf{R}^n$  такой, что

$$(\nabla g_i(x^*), s^0) = 0, \quad i = r+1, \dots, m; \quad (\nabla g_i(x^*), s^0) < 0, \quad i \in I^*.$$

Иначе говоря, находится элемент из касательного подпространства к ограничениям типа равенств в точке  $x^*$ , который ведет строго внутрь каждого из множеств  $g_i(x) \leq 0$ ,  $i \in I^*$  (рис. 37). В частности, если функции  $g_i(x)$ ,  $i = 1, \dots, m$ , выпуклы, а ограничения типа равенств отсутствуют и выполняется условие Слейтера (23) § 1, то, взяв  $s^0 = x^0 - x^*$ , мы получаем, что выполнено и условие Б. Условие Слейтера удобно тем, что не требует знания решения  $x^*$ .

**Теорема 3.** При замене условия А на условие Б утверждение теоремы 2 остается справедливым.

Доказательство. Предположим, что  $y_0^* = 0$  и умножим (3) скалярно на  $s^0$ . Тогда получаем

$$\sum_{i \in I^*} y_i^* (\nabla g_i(x^*), s^0) = 0,$$

что в сочетании с условиями  $(\nabla g_i(x^*), s^0) > 0$ ,  $y_i^* \geq 0$ ,  $i \in I^*$  дает

$$y_i^* = 0, \quad i \in I^*. \quad \text{Поэтому (3) переходит в } \sum_{i=r+1}^m y_i^* \nabla g_i(x^*) = 0, \quad y_i^* \text{ не}$$

все равны 0, что противоречит линейной независимости  $\nabla g_i(x^*)$ ,  $i = r+1, \dots, m$ .  $\blacktriangleleft$

#### Упражнение.

1. Сопоставьте доказательство теоремы 1 с первым доказательством приведенным Лагранжа (см. § 1 гл. 8). Какое утверждение использовалось там вместо теоремы отделимости?

2. Докажите теорему 1 по схеме, использующей штрафные функции (ср. с третьим доказательством теоремы 2 § 1 гл. 8). (Указание. Введите

$$f_k(x) = f(x) + \frac{1}{2} K \left( \sum_{i=1}^r g_i(x)_+^2 + \sum_{i=r+1}^m g_i(x)^2 \right) + \|x - x^*\|^2,$$

т. е., что  $x^*$  — точка минимума  $f_k(x)$  на  $Q = \{x: \|x - x^*\| \leq \varepsilon\}$  — для больших  $K$  лежит внутри  $Q$ , воспользовавшись необходимыми условиями минимума без

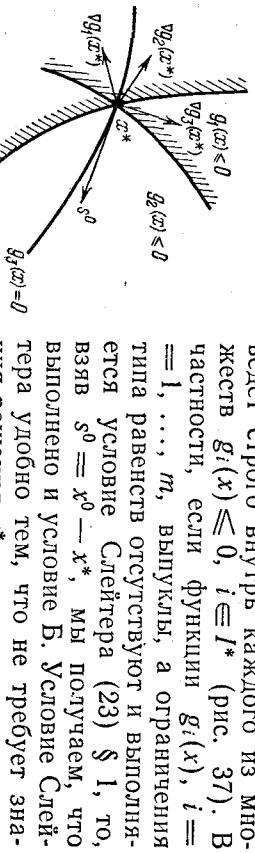


Рис. 37. Условие регулярности Б.

ограничений, т. е.  $\nabla f_k(x^*) = 0$ , и перейдите к пределу при  $K \rightarrow \infty$ . Покажите, что если точка минимума  $x^*$  локально единственна, то член  $\|x - x^*\|^2$  в  $f_k(x)$  можно отбросить.

**2. Достаточные условия минимума.** Оказывается, в некоторых случаях можно утверждать, что  $x^*$  является решением, пользуясь только первыми производными и не предполагая выпуклости задачи. У этого результата нет аналога ни в задачах безусловной минимизации, ни в задачах с ограничениями типа равенств, но он близок к условию острого минимума (см. § 1 гл. 7).

**Теорема 4** (достаточные условия I порядка). Пусть  $x^*$  допустима точка, функции  $f(x)$ ,  $g_i(x)$ ,  $i = 1, \dots, m$ , дифференцируемы в  $x^*$ , число активных ограничений равно  $n$ , а их градиенты (*t. e.*  $\nabla g_i(x^*)$ ,  $i \in I$ ) линейно независимы. Пусть находится числа  $y_i^*$ ,  $i \in I$ , такие, что  $y_i^* > 0$ ,  $i \in I^*$ , и

$$\nabla f(x^*) + \sum_{i \in I} y_i^* \nabla g_i(x^*) = 0. \quad (7)$$

Тогда  $x^*$  — точка локального минимума в задаче (1).

Доказательство. Возьмем произвольную последовательность допустимых точек  $x^k \rightarrow x^*$ , и пусть  $s \in \mathbf{R}^n$ ,  $\|s\| = 1$ , — предельная точка последовательности векторов  $(x^k - x^*)/\|x^k - x^*\|$ . Тогда из дифференцируемости  $g_i(x)$ ,  $i = r+1, \dots, m$ , в точке  $x^*$  следует:

$$0 = g_i(x^k) = g_i(x^*) + (\nabla g_i(x^*), x^k - x^*) + o(x^k - x^*) =$$

$$= (\nabla g_i(x^*), x^k - x^*) + o(x^k - x^*).$$

Разделив на  $\|x^k - x^*\|$  и переходя к пределу, получаем

$$(\nabla g_i(x^*), s) = 0, \quad i = r+1, \dots, m.$$

Совершенно аналогично

$$(\nabla g_i(x^*), s) \leq 0, \quad i \in I^*.$$

Умножим равенство (7) скалярно на  $s$  и используем (8):

$$0 = (\nabla f(x^*), s) + \sum_{i \in I^*} y_i^* (\nabla g_i(x^*), s). \quad (10)$$

Величина  $(\nabla g_i(x^*), s)$  не может быть равна 0 для всех  $i \in I^*$ , так как тогда вектор  $s \neq 0$  был бы ортогонален к  $n$  линейно независимым векторам в  $\mathbf{R}^n$ . Поэтому  $\Phi(s) = \min_{i \in I^*} (\nabla g_i(x^*), s) < 0$  в силу (9). Введем множество  $S = \{s \in \mathbf{R}^n: \|s\| = 1, (\nabla g_i(x^*), s) = 0, i = r+1, \dots, m; (\nabla g_i(x^*), s) \leq 0, i \in I^*\}$ . Тогда  $\max_{s \in S} \Phi(s) = -\varepsilon < 0$  в силу непрерывности и отрицательности  $\Phi(s)$  на  $S$  и компактности  $S$ .

Учитывая, что  $y_i^* > 0$ ,  $i \in I^*$ , из (10) получаем

$$(\nabla f(x^*), s) = - \sum_{i \in I^*} y_i^* (\nabla g_i(x^*), s) \geq \varepsilon \min_{i \in I^*} y_i^* = 2a > 0$$

для любых  $s$ , являющихся предельными точками  $(x^k - x^*)/\|x^k - x^*\|$  при допустимых  $x^k$ ,  $x^k \rightarrow x^*$ . Поскольку

$$f(x^k) = f(x^*) + \|x^k - x^*\|(\nabla f(x^*), (x^k - x^*)/\|x^k - x^*\|) + o(x^k - x^*),$$

то справедливо неравенство

$$f(x^k) \geq f(x^*) + \alpha \|x^k - x^*\| \quad (11)$$

для всех допустимых  $x^k$ , достаточно близких к  $x^*$ . Таким образом,  $x^*$  — точка локального минимума. ▲

Обратим внимание, что в действительности мы доказали (11), т. е. что  $x^*$  — точка острого минимума (ср. с § 1 гл. 7).

Для простейшей одномерной задачи

$$\min f(x), \quad x \in \mathbf{R}^1, \quad x \geq 0 \quad (12)$$

теорема 4 утверждает, что если  $f''(0) > 0$ , то  $x^* = 0$  — точка минимума. Этот факт геометрически очевиден (см. рис. 30).

Вспомогателься достаточными условиями первого порядка не всегда удается, так как теорема 4 применима лишь тогда, когда число активных ограничений равно размерности пространства. Например, она заведомо не работает, если общее число ограничений меньше  $n$ .

**Теорема 5** (достаточные условия II порядка). *Пусть  $x^*$  — допустимая точка, функции  $f(x)$ ,  $g_i(x)$ ,  $i = 1, \dots, m$ , дважды дифференцируемы в  $x^*$ . Пусть для некоторого  $y^* \in \mathbf{R}^m$  выполнены (5) и для всякого  $s \neq 0$ , для которого*

$$\begin{aligned} (\nabla g_i(x^*), s) &= 0, \quad i \in I^*, \quad y_i^* > 0, \\ (\nabla g_i(x^*), s) &= 0, \quad i \in I^*, \quad y_i^* = 0, \\ (\nabla g_i(x^*), s) &\geq 0, \quad i \in I^*, \quad y_i^* = 0, \end{aligned} \quad (13)$$

выполняется неравенство

$$(L''_{xx}(x^*, y^*)s, s) > 0.$$

Тогда  $x^*$  — точка локального минимума в задаче (11). ▲

**3. Единственность и устойчивость решения.**

**Теорема 6.** *Пусть выполнены условия теоремы 4 или теоремы 5. Тогда  $x^*$  — локально единственное решение.*

Действительно, в условиях теоремы 4 было доказано первенство (11), из которого и следует единственность решения. Аналогичным образом можно показать, что в условиях теоремы 5 справедливо либо условие  $(\nabla f(x^*), s) > 0$ , либо условия  $(\nabla f(x^*), s) = 0$ ,  $(\nabla^2 f(x^*)s, s) > 0$ , где  $s$  имеет тот же смысл, что и в теореме 4. Отсюда нетрудно получить нужное утверждение. ▲

Представляет интерес также вопрос об единственности множителей Лагранжа.

**Теорема 7.** *Пусть выполнены условия теоремы 2. Тогда множители Лагранжа определяются однозначно. ▲*

Этот результат очевиден — предположение о неединственности противоречит условию регулярности А. Как мы уже отмечали в § 1 в связи с условием Слейтера, из условия регулярности Б не следует единственность множителей Лагранжа.

Исследуем устойчивость решения по отношению к разного рода возмущениям. Первый результат демонстрирует инвариантность решения при возмущениях целевой функции, если выполнено достаточное условие первого порядка.

**Теорема 8.** *Пусть выполнены условия теоремы 4. Тогда точка минимума в задаче*

$$\min_{x \in Q} [f(x) + \varepsilon f_1(x)], \quad (15)$$

$$Q = \{x: g_i(x) \leq 0, \quad i = 1, \dots, r; \quad g_i(x) = 0, \quad i = r+1, \dots, m\}$$

(где  $f_1(x)$  дифференцируема в  $x^*$ ) при достаточно малых  $\varepsilon > 0$  совпадает с  $x^*$ .

Доказательство. В силу неравенства (11)  $x^*$  является точкой острого минимума в задаче (1). Аналогично теореме 8 § 1 гл. 7 она не меняется при малых возмущениях минимизируемой функции. ▲

Перейдем к изучению более общей проблемы устойчивости. Здесь и в дальнейшем понадобится следующее понятие. Назовем точку  $x^* \in \mathbf{R}^n$  невырожденным решением задачи (1), если:

- а) функции  $f(x)$ ,  $g_i(x)$ ,  $i = 1, \dots, m$ , дважды дифференцируемы в окрестности  $x^*$ ;
- б) точка  $x^*$  является допустимой, для нее множества  $I^*$  и  $I$  определены в (2);
- в) выполнено условие регулярности А, т. е.  $\nabla g_i(x^*)$ ,  $i \in I$ , линейно независимы;

1) выполнено достаточное условие второго порядка и условие строгой дополняющей нежесткости, т. е. найдется вектор  $y^*$  такой, что

$$L'_x(x^*, y^*) = 0, \quad y_i^* > 0, \quad i \in I^*; \quad y_i^* = 0, \quad i \in I^*, \quad 1 \leq i \leq r, \quad (16)$$

а для всякого  $s \neq 0$ , для которого

$$(\nabla g_i(x^*), s) = 0, \quad i \in I, \quad (17)$$

выполняется неравенство

$$(L''_{xx}(x^*, y^*)s, s) > 0. \quad (18)$$

Для такой точки выполнены условия теоремы 5, так что она действительно является точкой минимума. Более того, она линейно единственна (см. теорему 6), а множители Лагранжа  $y^*$  определены однозначно (см. теорему 7).

Для невырожденной точки минимума вопрос об устойчивости может быть решен в наиболее общей формулировке. Рассмотрим *возмущенную задачу*

$$\begin{aligned} g_i(x) &\leq \varepsilon_i, & i = 1, \dots, r, \\ g_i(x) &= \varepsilon_i, & i = r+1, \dots, m, \end{aligned} \quad (19)$$

где  $f(x)$  — дважды дифференцируемая в окрестности  $x^*$  функция. Обозначим через  $\varepsilon$  вектор в  $\mathbf{R}^{m+1}$  с компонентами  $\varepsilon_0, \dots, \varepsilon_m$ .

**Теорема 9.** *Пусть  $x^*$  — невырожденное решение (1). Тогда для достаточно малых  $\varepsilon$  существуют  $x_\varepsilon, y_\varepsilon$  — решение задачи (19) и соответствующие ему множители Лагранжа. При этом  $x_\varepsilon \rightarrow x^*$ ,  $y_\varepsilon \rightarrow y^*$  при  $\varepsilon \rightarrow 0$ .*

Причем  $x_\varepsilon, y_\varepsilon$  дифференцируемы в 0 и для  $\Phi(\varepsilon) = f(x_\varepsilon)$

$$\partial\Phi(0)/\partial\varepsilon_i = 0, \quad \partial\Phi(0)/\partial\varepsilon_i = y_i^*, \quad 1 \leq i \leq m. \quad (21)$$

Наметим кратко доказательство. В силу необходимых условий экстремума в точке  $x_\varepsilon$  (если она существует и множество активных ограничений в ней то же, что в  $x^*$ ) должны выполняться равенства

$$\nabla f(x_\varepsilon) + \sum_{i \in I} y_{ei} \nabla g_i(x_\varepsilon) + \varepsilon_0 \nabla f_1(x_\varepsilon) = 0,$$

$$g_i(x_\varepsilon) = \varepsilon_i, \quad i \in I. \quad (22)$$

Эту систему уравнений относительно  $z = \{x, y_i, i \in I\}$  можно записать в виде

$$R(z) = T(z)\varepsilon. \quad (23)$$

При этом  $z^* = \{x^*, y_i^*, i \in I\}$  удовлетворяет уравнению  $R(z^*) = 0$ . Матрица Якоби  $R'(z^*)$ , как легко проверяется непосредственным подсчетом, имеет вид

$$R'(z^*) = \begin{pmatrix} L''_{xx}(x^*, y^*) & \nabla^T g(x^*) \\ \hat{V}_g(x^*) & 0 \end{pmatrix}, \quad (24)$$

где  $\hat{V}_g(x^*)$  — матрица, строками которой являются  $\nabla g_i(x^*)$ ,  $i \in I$ .

В силу леммы 2 § 1 гл. 8 и условия г) в определении невырожденной точки минимума матрица  $R'(z^*)$  невырождена. В соответствии с теоремой 3 § 3 гл. 2 система (23) имеет решение  $z_\varepsilon$  и  $z_\varepsilon = z^* - R'(z^*)^{-1} T(z^*) \varepsilon + o(\varepsilon)$ .

Отсюда следует, что для достаточно малых  $\varepsilon$  найдется решение  $x_\varepsilon, y_\varepsilon$ ,  $i \in I$ , системы (23), дифференцируемое по  $\varepsilon$ . При этом будет  $y_\varepsilon > 0$ ,  $i \in I^*$ , так как  $y_i^* > 0$ ,  $i \in I^*$ . Можно пока-

зать, что при этом выполняются и достаточные условия второго порядка (в силу их выполнения в  $x^*$  и соображений непрерывности). Точка  $x_\varepsilon$  — допустимая, поскольку она удовлетворяет всем ограничениям (ограничениям  $g_i(x) \leq \varepsilon_i$ ,  $i \in I$ , — по определению  $x_\varepsilon$ , ограничениям  $g_i(x) \leq \varepsilon_i$ ,  $i \in I$ , — по определению  $x_\varepsilon$ , ограничениям  $g_i(x^*) < 0$  для таких  $i$  и из соображений непрерывности). Итак,  $x_\varepsilon$  — решение задачи (19), а  $y_\varepsilon$  — соответствующие множители Лагранжа.

Далее, в силу дифференцируемости  $x_\varepsilon$ :  $\Phi(\varepsilon) - \Phi(0) = f(x_\varepsilon) - f(x^*) = (\nabla f(x^*), x_\varepsilon - x^*) + o(\varepsilon) = - \sum_{i=1}^m (y_i^* \nabla g_i(x^*), x_\varepsilon - x^*) + o(\varepsilon)$ . Используя формулу (25) для  $z_\varepsilon - z^*$  и проделав требуемые выкладки, получаем формулу (21) для  $\nabla \Phi(0)$ .  $\blacktriangle$

Выражения (21) позволяют трактовать множители Лагранжа как коэффициенты влияния ограничений. В самом деле, они показывают, что скорость изменения целевой функции при возмущении какого-либо ограничения равна соответствующему множителю Лагранжа. Чем больше последний, тем больше чувствительность по отношению к данному уравнению. Наоборот, для неактивных ограничений  $y_i^* = 0$ , что соответствует факту нечувствительности решения к возмущению этих ограничений.

**Упражнение.**  
3. Докажите теорему 8 иначе — покажите, что уравнение  $\nabla_{L_\varepsilon}(x^*, y) = 0$  (где  $L_\varepsilon(x, y)$  — функция Лагранжа для (15)) имеет решение  $y_\varepsilon$  при достаточно малых  $\varepsilon$ , причем  $y_{ei}^* > 0$ ,  $i \in I^*$ . Тем самым в точке  $x^*$  выполняется достаточное условие экстремума для задачи (15).

4. Докажите, что если выполнены все требования в определении невырожденного минимума, за исключением (18), а число активных ограничений равно  $n$ , то выполняется и (18), так как из (17) следует  $s = 0$ . При этом минимум является острым, т. е.  $f(x) \geq f(x^*) + \alpha \|x - x^*\|$ ,  $\alpha > 0$ , для всех допустимых  $x$ , достаточно близких к  $x^*$ .

### § 3. Методы выпуклого программирования

Рассмотрим сначала задачу выпуклого программирования

$$\begin{aligned} \min f(x), \quad &x \in \mathbf{R}^n, \\ g_i(x) &\leq 0, \quad i = 1, \dots, m, \end{aligned} \quad (1)$$

где  $f(x)$ ,  $g_i(x)$  — выпуклые дифференцируемые на  $\mathbf{R}^n$  функции, и остановимся лишь на методах, специфических для данного класса задач. При этом нужно иметь в виду, что к задаче (1) можно применять как методы, рассчитанные на более общие

задачи (§ 4), так и методы, рассчитанные на «простые» ограничения (гл. 7).

Относительно способов классификации методов можно повторить сказанное в § 2 гл. 8. В данном случае методы можно

делить и по тому принципу, дают ли они последовательность допустимых точек (т. е. удовлетворяющих ограничениям  $g_i(x) \leq 0$ ,  $i = 1, \dots, m$ ), или нет. Первые часто называют *методами возможных направлений*.

**1. Методы возможных направлений.** Пусть  $x^k$  — некоторая допустимая точка. Определим для нее « $\varepsilon$ -активные» ограничения, т. е. такие, которые с точностью до параметра  $\varepsilon_k \geq 0$  обрашаются в равенства:

$$I_k = \{i: g_i(x^k) \geq -\varepsilon_k\}. \quad (2)$$

Линеаризуем минимизируемую функцию и  $\varepsilon$ -активные ограничия в точке  $x^k$  и найдем направление  $s^k$ , которое является допустимым для линеаризованной задачи и одновременно приводит к наиболее быстрому убыванию функции. Иными словами, выберем в качестве  $s^k$  решение вспомогательной задачи

$$\max z, \quad (\nabla f(x^k), s) \leq -z, \quad i \in I_k, \quad s \in S_k, \quad (3)$$

где  $z \in \mathbf{R}^1$  — дополнительная переменная, а  $S_k$  — некоторое «простое» ограниченное множество, которое вводится для того, чтобы данная задача заведомо имела решение. Например, можно взять в качестве  $S_k$  некоторый шар

$$S_k = \{s: \|s\|^2 \leq \rho_k^2\}$$

или куб

$$S_k = \{s: |s_i| \leq \rho_k, \quad i = 1, \dots, n\}. \quad (5)$$

Эти способы называются *нормализации N1* и *N2* соответственно. Далее делается шаг

$$x^{k+1} = x^k + v_k s^k, \quad (6)$$

где длина шага  $v_k > 0$  выбирается так, чтобы в точке  $x^{k+1}$  все ограничения не нарушались, а  $f(x^{k+1})$  принимало бы наименьшее значение. В новой точке  $x^{k+1}$  процедура повторяется.

Параметр  $\varepsilon_k$  можно регулировать по-разному, нужно лишь, чтобы было  $\varepsilon_k > 0$ ,  $\varepsilon_k \rightarrow 0$ . Если взять  $\varepsilon_k \equiv 0$  (т. е. учитывать лишь активные в точке  $x^k$  ограничения), то метод «не чувствует» ограничений, которые в точке  $x^k$  выполняются «почти» как равенства (т. е.  $|g_i(x^k)|$  мало), а их присутствие резко ограничивает длину шага из  $x^k$ . Поэтому может произойти «застревание» метода в окрестности точки, не являющейся решением.

Одним из первых конкретных алгоритмов типа возможных направлений был *метод проекции градиента Розена*. В нем делается шаг по проекции градиента  $\nabla f(x^k)$  на многообразие, задаваемое линеаризацией активных ограничений. Следует отли-

чать такой метод от описанного выше (§ 2 гл. 7) метод проекции градиента. В последнем производится движение по градиенту, а затем проектирование на допустимое множество — это не то же, что движение по проекции градиента на грань допустимого множества (см. рис. 38, где показан шаг каждого из методов для случая линейных ограничений).

Не будем приводить теоретические результаты, связанные с обоснованием метода довольно громоздко (хотя идеально и не сложно). Во-первых, скорость сходимости метода может быть мала даже для «хороших» задач, что связано с введением параметра  $\varepsilon_k$  и необходимостью оставаться внутри допустимого множества. Во-вторых, трудности определенные выбором национальной допустимой точки, предстают в виде известных методов, свободные от указанных выше недостатков и в то же время не более сложные. К описанию одного из таких методов мы и переходим.

**Упражнения.**

1. Рассмотрите метод проекции градиента и условного градиента (§ 2 гл. 8) для задач с линейными ограничениями как конкретизацию метода возможных направлений.
2. Покажите, что задача (3), (4) эквивалентна задаче квадратичного программирования

$$\min [-z + \lambda_k \|s\|^2], \quad (\nabla f(x^k), s) \leq -z, \quad (\nabla g_i(x^k), s) \leq -z, \quad i \in I_k,$$

при некотором  $\lambda_k \geq 0$ .

**2. Метод линеаризации.** Линеаризуем все ограничения в точке  $x^k$ , решим вспомогательную задачу квадратичного программирования

$$\min [(\nabla f(x^k), x - x^k) + (2V)^{-1} \|x - x^k\|^2], \quad (7)$$

$$g_i(x^k) + (\nabla g_i(x^k), x - x^k) \leq 0, \quad i = 1, \dots, m,$$

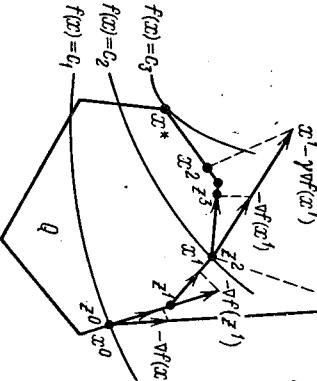


Рис. 38. Метод проекции градиента и условного градиента (точки  $x^0, x^1, x^2, \dots, z^0, z^1, z^2, \dots$ ).

Теорема 1. Пусть множество  $X^*$  решений задачи (1) не пусто, функции  $f(x)$ ,  $g_i(x)$  выпуклы и дифференцируемы, а их градиенты удовлетворяют условию Липшица, и выполняется

*Условие Слейтера.* Тогда находится  $\bar{\gamma} > 0$  такое, что при  $0 < \gamma < \bar{\gamma}$  метод (7) сходится к точке  $x^* \in X^*$ . Если при этом  $f(x)$  сильно выпукла, то  $\|x^k - x^*\| \leq c q^k$ ,  $0 \leq q < 1$ . ▲

Опустим доказательство этого результата, так как оно довольно громоздко, а локальный вариант теоремы 1 (при нескольких иных предположениях) будет доказан позже для более общей задачи (см. теорему 1 § 4). Отметим, что если ограничения отсутствуют, то метод (7) переходит в градиентный метод безусловной минимизации  $f(x)$ , а теорема 1 совпадает с теоремами 1 § 1 гл. 6 и 2 § 4 гл. 1. Метод (7) довольно близок к методу возможных направлений с нормализацией  $N_1$ , но, вообще говоря, не совпадает с ним (ср. с упр. 2). В частности, получающиеся в (7) точки не обязаны быть допустимыми.

В описанном выше варианте метода производилась линеаризация всех ограничений. В действительности можно учитывать некоторым  $\varepsilon > 0$ , выбираем

$$I_k = \{i: g_i(x^k) \geq \max_{1 \leq i \leq m} g_i(x^k)_+ - \varepsilon\} \quad (8)$$

и в качестве  $x^{k+1}$  возьмем решение задачи

$$\min [(\nabla f(x^k), x - x^k) + (2y)^{-1} \|x - x^k\|^2], \\ g_i(x^k) + (\nabla g_i(x^k), x - x^k) \leq 0, \quad i \in I_k. \quad (9)$$

Для метода (8), (9) справедливы утверждения, аналогичные теореме 1. В то же время эта модификация метода более экономна — на каждом шаге решается задача меньшей размерности, чем в варианте (7).

Способ выбора постоянного параметра  $y$  в (7) и (9), конечно, не является единственным возможным. Существуют конструктивные приемы подбора  $y_k$ , аналогичные алгоритмам регулировки длины шага в градиентном методе безусловной минимизации (10) § 1 гл. 3.

Наконец, можно использовать другие вычислительные схемы метода (7) или (9). Так, преобразуя метод (7) в виде

$$x^{k+1} = P_{Q_k}(x^k - y \nabla f(x^k)),$$

$$Q_k = \{x: g_i(x^k) + (\nabla g_i(x^k), x - x^k) \leq 0, \quad i = 1, \dots, m\}. \quad (10)$$

Отсюда, в частности, видно, что для задач с линейными ограничениями метод линеаризации совпадает с методом проекции гравиента (см. § 2 гл. 7). С другой стороны, если записать двойственную задачу к (9), то (см. (12) § 4 гл. 10) она будет иметь вид

$$\min_{i \in I_k} \left[ \frac{y}{2} \|\nabla f(x^k) + \sum_{i \in I_k} y_i \nabla g_i(x^k)\|^2 - \sum_{i \in I_k} y_i g_i(x^k) \right]. \quad (11)$$

Обозначим ее решение  $y_i^k$ ,  $i \in I_k$ , и построим

$$x^{k+1} = x^k - y \left( \nabla f(x^k) + \sum_{i \in I_k} y_i^k \nabla g_i(x^k) \right). \quad (12)$$

В соответствии с (13) § 4 гл. 10 точка  $x^{k+1}$  будет та же, что и в методе (9).

Преимущество такого подхода в том, что вспомогательная задача (11) сводится к отысканию минимума квадратичной функции на  $\mathbf{R}_+^m$ , который может быть найден, например, методом сопряженных градиентов за конечное число шагов (см. § 3 гл. 7).

**3. Двойственные методы.** В схеме (11), (12) метода линеаризации фигурируют как прямые, так и двойственные переменные. Известно много методов такого типа. По-видимому, наиболее простой в идейном смысле схемой пересчета прямых и двойственных переменных является градиентный метод Эрроу — Гравина — Узбади (ср. с (6) § 2 гл. 8):

$$x^{k+1} = x^k - y L'_x(x^k, y^k) = x^k - y \left( \nabla f(x^k) + \sum_{i=1}^m y_i^k \nabla g_i(x^k) \right), \\ y^{k+1} = [y^k + y L'_y(x^k, y^k)]_+ = [y^k + y g(x^k)]_+. \quad (13)$$

В нем для отыскания седловой точки функции  $L(x, y)$  на  $\mathbf{R}^n \times \mathbf{R}^m_+$  (а этому эквивалентна задача (1), см. теорему 3 § 1) делается градиентный шаг минимизации по  $x$  и шаг метода проекции гравиента по  $y$ . Однако метод (13) может и не сходиться. Рассмотрим простейший пример — одномерную задачу линейного программирования

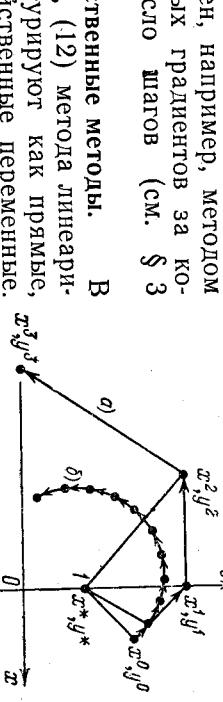
$$\min x, \quad x \in \mathbf{R}^1, \quad -x \leq 0. \quad (14)$$

В ней  $x^* = 0$ ,  $y^* = 1$ , а метод (13) принимает вид

$$x^{k+1} = x^k - y(1 - y^k), \quad y^{k+1} = (y^k - y x^k)_+. \quad (15)$$

Траектория такого процесса показана на рис. 39. Видно, что при малых  $y$  нет сходимости (даже в окрестность  $x^*, y^*$ ).

Если предположить, что  $f(x)$  сильно выпукла, то такие эффекты не наблюдаются. Однако вопрос о строгом обосновании метода и в этом случае не прост. Стандартная техника доказательства, основанная на введении функции Ляпунова вида  $\|x - x^*\|^2 + \|y - y^*\|^2$ , здесь не работает (см. упр. 4). Этим путем удается доказать лишь более слабое утверждение — о том,



## § 3. МЕТОДЫ ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ

что найдется подпоследовательность  $x^k$ , сходящаяся к  $x^*$ . Однако ценность такого результата невелика, так как остается неясным, как выделять упомянутую подпоследовательность. Дело в том, что соответствующая подпоследовательность  $y^k$  не обязательно сходится к  $y^*$ , так что величина  $L'_x(x^k, y^k)$  не стремится к 0 и не может служить характеристикой точности решения.

Другой вариант метода предполагает, что вместо одного градиентного шага проводится полная минимизация  $L(x, y^k)$  по  $x$  (ср. с (8) § 2 гл. 8):

$$x^{k+1} = \operatorname{argmin}_x L(x, y^k), \quad y^{k+1} = [y^k + \gamma g(x^{k+1})]_+. \quad (16)$$

Этот метод может интерпретироваться как метод проекции градиента для двойственной задачи. В самом деле, как мы видели (в § 1), двойственная задача имеет вид

$$\max_{y \geq 0} \Psi(y), \quad \Psi(y) = \min_{x \in \mathbb{R}^n} L(x, y),$$

при этом, если  $f(x)$  сильно выпукла, то  $\Psi(y)$  вогнута, вследствие определена и дифференцируема, причем градиент  $\Psi(y)$  удовлетворяет условию Липшица и равен  $L'_y(x(y), y)$ , где  $x(y) = \operatorname{argmin}_x L(x, y)$  (упр. 10 § 1). Поэтому (16) можно записать в виде

$$y^{k+1} = [y^k + \gamma \nabla \Psi(y^k)]_+$$

$$(17)$$

и применить теорему 1 § 2 гл. 7 о сходимости метода проекции градиента.

Область применимости методов множителей Лагранжа может расширяться, а скорость сходимости повысить, если от обычной функции Лагранжа перейти к модифицированной. Составим

$$M(x, y, K) = f(x) + \frac{1}{2K} \| (y + Kg(x))_+ \|^2 - \frac{1}{2K} \| y \|^2, \quad (18)$$

где  $K$  — некоторый параметр (ср. с (14) § 1 гл. 8). Приведем без доказательства основные ее свойства.

**П е м а 1.** Пусть  $f(x)$ ,  $g_i(x)$  выпуклы,  $K > 0$ . Тогда:

а)  $M(x, y, K)$  выпукла по  $x$  и вогнута по  $y$ ;

б)  $\lim_{K \rightarrow 0} M(x, y, K) = \begin{cases} L(x, y), & \text{если } y \geq 0, \\ -\infty & \text{в противном случае;} \end{cases}$

в) множества  $X^* \times Y^*$  седловых точек  $M(x, y, K)$  на  $\mathbf{R}^n \times \mathbf{R}^m$  и  $L(x, y)$  на  $\mathbf{R}^n \times \mathbf{R}_+^m$  совпадают;

г) если  $f(x)$ ,  $g_i(x)$  дифференцируемы, а  $X^* \times Y^*$  непусто, то  $(M'_x(x, y, K), x - x^*) - (M'_y(x, y, K), y - y^*) \leq -K \| M'_y(x, y, K) \|^2$  при всех  $x \in \mathbf{R}^n$ ,  $y \in \mathbf{R}^m$ ,  $x^* \in X^*$ ,  $y^* \in Y^*$ . ▲

Рассмотрим аналог метода (13), в котором  $L(x, y)$  заменена на  $M(x, y, K)$  (ср. с (9) § 2 (гл. 8)):

$$x^{k+1} = x^k - \gamma M'_x(x^k, y^k, K), \quad y^{k+1} = y^k + \gamma M'_y(x^k, y^k, K). \quad (20)$$

**Теорема 2.** Пусть  $f(x)$ ,  $g_i(x)$  выпуклы (причем  $f(x)$  сильно выпукла),  $\partial g_i(x)$  дифференцируемы, их градиенты и вторые производные удовлетворяют условию Липшица. Пусть для решения  $x^*$  и множителей Лагранжа  $y^*$  справедливы условия:  $\nabla g_i(x^*)$  линейно независимы при  $i \in I^*$  и  $y_i^* > 0$  при  $i \in I^*$ , где  $I^* = \{i : g_i(x^*) = 0\}$ . Тогда для любых  $x^0, y^0$  найдется  $\bar{\gamma} > 0$  такое, что при  $0 < \gamma < \bar{\gamma}$  метод (20) сходит к  $x^*, y^*$  со скоростью геометрической прогрессии.

Наметим краткую схему доказательства. Введем функцию Ляпунова

$$V(x, y) = \|x - x^*\|^2 + \|y - y^*\|^2 - K(M(x, y, K) - M(x^*, y^*, K)). \quad (21)$$

Используя неравенство (19), можно показать, что

$$\begin{aligned} V(x^{k+1}, y^{k+1}) &\leq V(x^k, y^k) - \lambda (\|M'_x(x^k, y^k, K)\|^2 + \|M'_y(x^k, y^k, K)\|^2), \quad \lambda > 0. \end{aligned}$$

Отсюда следует сходимость метода. Поэтому для достаточно больших  $k$  будет  $y_i^k \equiv 0$  для  $i \in I^*$ . При этом итерационный процесс совпадает с методом модифицированной функции Лагранжа для задачи с ограничениями в форме равенств

$$\min f(x), \quad g_i(x) = 0, \quad i \in I^*.$$

Но для последнего в теореме 3 § 2 гл. 8 была доказана локальная сходимость со скоростью геометрической прогрессии. ▲

Аналогом (16) является метод, в котором на каждом шаге имеется безусловный минимум модифицированной функции Лагранжа (ср. с (11) § 2 гл. 8):

$$x^{k+1} = \operatorname{argmin}_x M(x, y^k, K), \quad y^{k+1} = y^k + Kg(x^{k+1}). \quad (23)$$

**Теорема 3.** Пусть выполнены условия теоремы 2. Тогда для всякого  $y^0 \geq 0$  найдется  $K$  такое, что при  $K > K$  метод (23) сходится к решению со скоростью геометрической прогрессии со знаменателем  $O(1/K)$ .

Вновь ограничимся лишь наброском доказательства. В качестве функции Ляпунова целесообразно взять  $V(y) = \|y - y^*\|^2$ . Тогда

$$V(y^{k+1}) \leq V(y^k) - \|y^{k+1} - y^k\|^2. \quad (24)$$

Отсюда следует сходимость метода. Локальное поведение метода такое же, как для соответствующего алгоритма решения задачи

(22). Используя результат о скорости сходимости последнего (теорема 4 § 2 гл. 8), получаем оценку скорости сходимости, приведенную в теореме 3. ▲

Метод (23) может рассматриваться как один из наиболее эффективных подходов к решению задач выпуклого программирования. Он позволяет использовать мощные алгоритмы безусловной минимизации гладких функций (гл. 3). В то же время в нем нет недостатков, свойственных методу штрафных функций (см. ниже п. 4) — штрафной коэффициент  $K$  не возрастает, поэтому обусловленность вспомогательных задач не ухудшается от итерации к итерации. Наконец, по сравнению с методом (16), основанным на обычных функциях Лагранжа, метод (23) обладает существенно более высокой скоростью сходимости, так как знаменатель геометрической прогрессии может быть сделан малым за счет выбора  $K$ . Впрочем, здесь нужно соблюдать некоторый компромисс — ведь при слишком большом  $K$  усложняются задачи безусловной минимизации (они становятся плохо обусловленными).

Наконец, упомянем еще об одном методе, в котором итерации по прямым и двойственным переменным ведутся следующим образом:

$$y^k = \operatorname{argmin}_{\substack{y_i \geq 0, i \in I_k}} \left\| \nabla f(x^k) + \sum_{i \in I_k} y_i \nabla g_i(x^k) \right\|^2, \\ I_k = \{i : g_i(x^k) \geq -\varepsilon\}, \\ x^{k+1} = x^k - y_k \left( \nabla f(x^k) + \sum_{i \in I_k} y_i \nabla g_i(x^k) \right). \quad (25)$$

Он называется методом *одновременного решения прямой и двойственной задач*. Его смысл достаточно прозрачен: для точки  $x^k$  находятся такие приближения  $y^k$  двойственных переменных, которые минимизируют невязку в условиях экстремума. Это приближение используется для уточнения прямых переменных (именно, делается шаг градиентного метода минимизации по  $x$  функции  $L(x, y^k)$ ). Метод (25) близок схеме (11), (12) метода линеаризации. С другой стороны, можно проверить, что мы получим (25), если в методе возможных направлений с нормализацией  $N1$  (3), (4), (6) перейдем к двойственной задаче.

**Упражнение.**

3. Покажите, что в методе (15) для примера (14) будет  $(x^{k+1} - x^*)^2 + (y^{k+1} - y^*)^2 \geq (x^k - x^*)^2 + (y^k - y^*)^2$  при любых  $x^k, y^k \geq 0$  и любом  $y \geq 0$ .

4. Выпишите метод (13) для задачи в  $\mathbf{R}^n$ :  $\min x^2, x \leq 0$ , и убедитесь, что  $v_k = (x^k - x^*)^2 + (y^k - y^*)^2$  не для всех  $x^k, y^k$  убывает монотонно при сколь угодно малых  $y \geq 0$ .

5. Докажите теорему 3, рассматривая (23) как метод градиента для решения следующей двойственной задачи:  $\max_y \Psi(y)$ , где  $\Psi(y) = \inf_x M(x, y, K)$ .

**4. Методы штрафов и ролевые им.** Идея метода штрафных функций, рассматривавшегося в § 2 гл. 8 для задач с ограничениями типа равенств, применима и для случая ограничений в виде неравенств. При этом появляется значительно больше возможностей.

1°. **Метод штрафов** (ср. с (13) § 2 гл. 8):

$$x^k = \operatorname{argmin}_x f_k(x), \quad (26)$$

$$f_k(x) = f(x) + \frac{1}{2} K_k \sum_{i=1}^m g_i(x)_+^2, \quad K_k > 0, \quad K_k \rightarrow \infty.$$

**Теорема 4.** Пусть  $f(x), g_i(x)$  выпуклы и конечны на  $\mathbf{R}^n$ , множество решений  $X^*$  задачи (1) непусто и ограничено. Тогда  $f_k(x)$  выпукла,  $X_k^* = \operatorname{Argmin}_{x \in \mathbf{R}^n} f_k(x) \neq \emptyset$ , последовательность  $x_k$  ограничена и все ее предельные точки принадлежат  $X^*$ , причем  $f(x^k) \leq f^* = f(x^*)$ ,  $x^* \in X^*$ . ▲

Условие ограниченности  $X^*$  существенно (см. упр. 6). При дополнительных предположениях метод (26) дает возможность получить и множители Лагранжа. Действительно,

$$0 = \nabla f_k(x^k) = \nabla f(x^k) + K \sum_{i=1}^m g_i(x^k)_+ \nabla g_i(x^k).$$

Поскольку  $x^k \rightarrow x^*$ , а в точке  $x^*$  выполняется условие

$$\nabla f(x^*) + \sum_{i=1}^m y_i^* \nabla g_i(x^*) = 0,$$

то должно быть  $K g_i(x^k)_+ \rightarrow y_i^*$ .

2°. **Метод барьеров:**

$$x^k = \operatorname{argmin}_x f_k(x), \quad (27)$$

$$f_k(x) = f(x) - \varepsilon_k \sum_{i=1}^m \frac{1}{g_i(x)}, \quad \varepsilon_k > 0, \quad \varepsilon_k \rightarrow 0.$$

Здесь минимум берется по  $x$ , для которых  $g_i(x) < 0, i = 1, \dots, m$ . Этот метод называется также *методом внутренних штрафов*, так как, в отличие от (26) (метода внешних штрафов), штрафный член в (27) не равен 0 даже для допустимых точек и возрастает при приближении к границе изнутри.

**Теорема 5.** Пусть  $f(x), g_i(x)$  выпуклы и конечны, множество решений  $X^*$  задачи (1) непусто и ограничено и выпукло, *меньше* условие Слейтера. Тогда в методе (27)  $X_k^* = \operatorname{Argmin}_x f_k(x) \neq \emptyset$ , функции  $f_k(x)$  выпуклы, последовательность  $x^k$  ограничена

и все ее предельные точки принадлежат  $X^*$ , причем  $f(x^k) \geq f^* = f(x^*)$ ,  $x^* \in X^*$ . ▲

В методе (27) можно получить приближения и для множества Лагранжа (упр. 7). Функция  $1/g_i(x)$  в (27) может быть заменена на другие функции, например на  $\log(-g_i(x))$ .

З<sup>о</sup>. Метод сдвиг штрафов. Регулировать штраф можно не только путем выбора штрафных коэффициентов, но и путем сдвига уровня штрафования:

$$x^k = \operatorname{argmin}_x f_k(x), \quad f_k(x) = f(x) + K \sum_{i=1}^m (g_i(x) + \lambda_i^k)^2, \quad (28)$$

где уровни  $\lambda_i^k$  нужно увеличивать, если  $x^k$  недопустимо, и уменьшать в противном случае. Например, можно взять

$$\lambda_i^{k+1} = \lambda_i^k + g_i(x^k), \quad i = 1, \dots, m. \quad (29)$$

Если сделать замену  $y_i^k = K\lambda_i^k$ , то обнаруживается, что метод (28), (29) переходит в (23). Итак, метод модифицированной функции Лагранжа (23) можно интерпретировать и как метод сдвиг штрафов.

4<sup>о</sup>. Метод подбора  $f^*$ . Если бы  $f^* = \min\{f(x) : g_i(x) \leq 0, i = 1, \dots, m\}$  было известно, то задача (1) была бы эквивалентна минимизации функции  $(f(x) - f^*)^2 + \|g(x)\|_+^2$ . Величину  $f^*$  можно подбирать рекуррентно, решая задачи

$$x^k = \operatorname{argmin}_x f_k(x), \quad f_k(x) = (f(x) - f_k)_+^2 + \sum_{i=1}^m g_i(x)_+^2. \quad (30)$$

Ясно, что если  $f_k(x^k) > 0$ , то значение  $f_k$  следует повысить. Одно из возможных правил пересчета  $f_k$  таково:

$$f_{k+1} = f_k + f_k(x^k)/(f(x^k) - f_k). \quad (31)$$

**Теорема 6.** Пусть  $f(x), g_i(x)$  выпуклы и конечны,  $x^* —$  регулярная точка минимума задачи (1) и  $f_0 < f^*$ . Тогда в методе (30), (31)  $x^k \rightarrow x^*, f_k \leq f^*$  и  $g_i(x^k)_+/f(x^k) - f_k \rightarrow y_i^*$ ,  $i = 1, \dots, m$ . ▲

Сравним описанные выше методы. Вид функций  $f_k(x)$  для каждого из них для простейшей задачи  $\min_x x, x \in \mathbf{R}^1, -x \leq 0$  приведен на рис. 40.

Метод штрафных функций (26) сходится при минимальных предположениях. Он сводится к безусловной минимизации дифференцируемой выпуклой функции на всем  $\mathbf{R}^n$ , в нем не требуется находить начальную допустимую точку. Некоторым его недостатком является то, что точки  $x^k$  не удовлетворяют ограничениям, однако этот дефект легко устранить — если известна точка  $x^0$  такая, что  $g_i(x^0) < 0$ ,  $i = 1, \dots, m$ , то можно брать  $\bar{x}^k = \lambda_k x^0 + (1 - \lambda_k) x^k$ ,  $\lambda_k = \min\{\lambda : g_i(\lambda x^0 + (1 - \lambda)x^k) \leq 0, i =$

$= 1, \dots, m\}$ , тогда  $\bar{x}^k$  являются допустимыми и сходятся к решению. При этом  $f(\bar{x}^k) \geq f^* \geq f(x^k)$ , т. е. получаются двусторонние оценки точности решения. Более серьезен недостаток, присущий всем вариантам метода штрафов (ср. с. § 2 гл. 8), — плохая обусловленность функции  $f_k(x)$  при больших  $K$ , что серьезно усложняет решение вспомогательной задачи безусловной минимизации в (26).

Метод барьеров имеет, пожалуй, единственное преимущество перед методом штрафов — функция  $f_k(x)$  в (27) более гладкая,

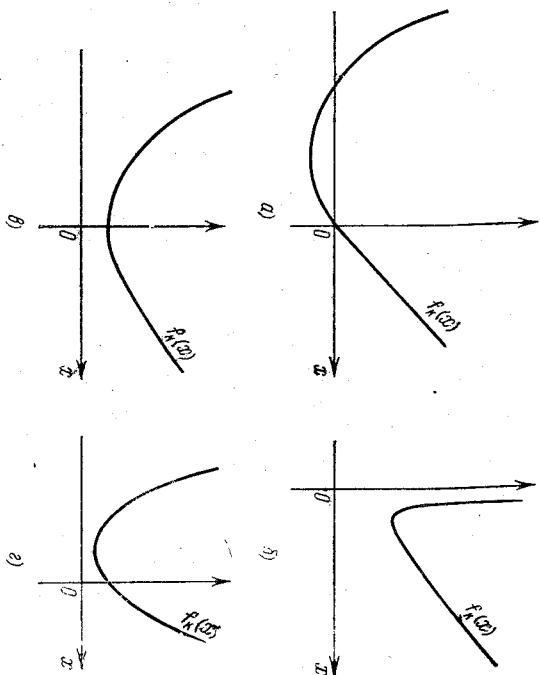


Рис. 40. Вид вспомогательных задач безусловной минимизации: а) метод подбора штрафов, б) метод барьеров, в) метод сдвига штрафов

чем в (26) ( $f_k(x)$  в (26) лишь однократно дифференцируема). В то же время для сходимости (27) требуется условие Слейтера (иначе метод просто теряет смысл); нужно знать начальную точку  $x^0$  такую, что  $g_i(x^0) < 0$ ,  $i = 1, \dots, m$ ; при проведении процесса минимизации нужно следить за тем, чтобы точка не вышла из допустимой области. Последнее обстоятельство особенно неприятно и не позволяет применять стандартные программы безусловной минимизации. Впрочем, их можно модифицировать, изменив соответствующим образом процедуру определения длины шага. Наконец, методу (27) присущи те же эффекты плохой обусловленности — функция  $f_k(x)$  существенно по-разному ведет себя по разным направлениям (резко растет при приближении к границе и мало меняется при движении вдоль границы).

Метод сдвига штрафов (28), (29), как мы видели, совпадает с методом модифицированных функций Лагранжа и обладает всеми его достоинствами: не требует роста штрафного коэффициента  $K$ , и потому не приводит к плохо обусловленным задачам; позволяет увеличивать скорость сходимости за счет выбора  $K$  и т. д.  $K$  недостаткам можно отнести более жесткие условия, нужные для сходимости метода.

Наконец, метод подбора  $f^*$  (30), (31) обладает рядом привлекательных черт — в нем нет каких-либо коэффициентов, стремящихся к 0 или  $\infty$ ,  $f_k(x)$  всюду определена и т. д. Однако ему также присущ эффект плохой обусловленности даже для регулярных задач (см. упр. 8). Кроме того, он требует решения вспомогательных задач минимизации с большой точностью, поскольку в нем приходится оперировать с малыми величинами  $f(x^k) - f_k$ . Например, если  $f_k > f^*$ , то  $\min_x f_k(x) = 0$ , однако из-за неизбежных ошибок точка  $x^k$  будет найдена приближенно, и может оказаться, что  $f(x^k) > f_k$  и значение  $f_k$  будет увеличено (тогда как в действительности его нужно уменьшить).

#### Упражнения.

6. Рассмотрите задачу минимизации в  $\mathbb{R}^2$ :  $\min_{x_1, x_2} x_1 - x_1 + 1 \leq 0, x_1 + 1 \leq 0, \rho^2(x, Q) \leq 0$ , где  $Q = \{x: x_2 \geq x_1^{-1}, x_1 > 0\}$ . Для нее  $X^* = \{x: x_1 = 1, x_2 \geq 1\}$ , функция  $g_i(x) = \rho^2(x, Q)$  выпукла (как квадрат расстояния до выпуклого множества  $Q$ , см. упр. 2 § 1 гл. 5). Покажите, что при  $0 < K \leq 1$  функция  $f_k(x)$  (26) для этой задачи не имеет точки минимума.

7. Докажите, что в (27) для невырожденной точки минимума  $s_k/g_i(x^k)^2 \rightarrow y_i^*$ .

8. Рассмотрите задачу в  $\mathbb{R}^2$ :  $\min_{x_1, x_2} x_1 - x_1 \leq 0, (x_1 + 1)^2 + x_2^2 - 1 \leq 0$ , для нее  $x^* = 0, f^* = 0$ . Покажите, что функция  $\Phi(x) = (f(x) - f^*)^2 + g_1(x)_+^2 + g_2(x)_+^2$  (возникающая в методе (30), (31) при  $f_k = f^*$ ) имеет в точке  $x^*$  вырожденный минимум (в частности,  $\partial^2 \Phi(0)/\partial x_2^2 = 0$ ).

5. Методы для негладких задач. Ниже рассматривается задача

$$\min f(x), \quad g_i(x) \leq 0, \quad i = 1, \dots, m, \quad x \in Q, \quad (33)$$

где  $Q$  — выпуклое замкнутое множество в  $\mathbb{R}^n$ , функции  $f(x)$ ,  $g_i(x)$  выпуклы и конечны на  $\mathbb{R}^n$ . При этом предполагается, что множество  $Q$  является «простым» в смысле гл. 7, а  $f(x)$  и  $g_i(x)$ , вообще говоря, недифференцируемы, и в произвольной точке  $x \in Q$  можно вычислить их субградиенты  $\partial f(x)$ ,  $\partial g_i(x)$ . Ясно, что  $m$  ограничений  $g_i(x) \leq 0$  эквивалентны одному

$$g(x) \leq 0, \quad g(x) = \max_{1 \leq i \leq m} g_i(x), \quad (34)$$

где  $g(x)$  выпукла и в соответствии с леммой 11 § 1 гл. 5

$$\partial g(x) = \left\{ \sum_{i \in I(x)} \lambda_i \partial g_i(x), \quad \lambda \geq 0, \quad \sum_{i \in I(x)} \lambda_i = 1 \right\},$$

$$I(x) = \{i: g_i(x) = g(x)\}.$$

Рассмотрим следующее обобщение метода проекции субградиента:

$$x^{k+1} = P_Q(x^k - \gamma_k s^k), \quad s^k = \begin{cases} \partial f(x^k), & \text{если } g(x^k) \leq 0, \\ \partial g(x^k), & \text{если } g(x^k) > 0. \end{cases} \quad (35)$$

Иначе говоря, делается шаг по проекции субградиента целевой функции, если точка  $x^k$  допустима, и по проекции субградиента нарушенных ограничений в противном случае.

Теорема 7. Пусть множество  $Q$  ограничено и существует  $x^0 \in Q$ ,  $g_i(x^0) < 0$ ,  $i = 1, \dots, m$ . Тогда все предельные точки последовательности  $x^k$  в методе (35) являются решениями задачи (33).

Доказательство. В силу предположений  $f(x)$ ,  $g_i(x)$  непрерывны на  $Q$ , допустимое множество замкнуто и ограничено, а потому решение существует. Пусть  $x^* — любое из решений. Выберем  $\varepsilon > 0$  и рассмотрим  $S_\varepsilon = \{x \in Q: f(x) \leq f(x^*) + \varepsilon, g(x) \leq 0\}$ . Тогда на отрезке  $[x^0, x^*]$  находится точка  $\bar{x} \in S_\varepsilon$  такая, что  $f(\bar{x}) \leq f(x^*) + \delta$ ,  $g(\bar{x}) \leq -\delta$  для некоторого  $0 < \delta < \varepsilon$ . Оценим расстояние от  $x^{k+1}$  до  $\bar{x}$ . Имеем$

$$\begin{aligned} \|x^{k+1} - \bar{x}\|^2 &= \|P_Q(x^k - \gamma_k s^k) - \bar{x}\|^2 \leq \|x^k - \gamma_k s^k - \bar{x}\|^2 \\ &= \|x^k - \bar{x}\|^2 - 2\gamma_k(s^k, x^k - \bar{x}) + \gamma_k^2 \|s^k\|^2. \end{aligned}$$

Поскольку субградиент непрерывной функции  $f(x)$  или  $g(x)$  ограничен на ограниченном множестве  $Q$  (лемма 8 § 1 гл. 5), то  $\|s^k\| \leq c$ . Если  $x^k \in S_\varepsilon$ , то либо  $g(x^k) \leq 0$ ,  $f(x^k) \geq f(x^*) + \varepsilon$ , либо  $g(x^k) > 0$ ,  $f(x^k) \leq f(x^*) + \varepsilon$ . В первом случае  $s^k = \partial f(x^k)$ ,  $(s^k, x^k - \bar{x}) = (\partial f(x^k), x^k - \bar{x}) \geq f(x^k) - f(\bar{x}) \geq f(x^*) + \varepsilon - f(x^*) - \delta = \varepsilon - \delta$ . Во втором случае  $s^k = \partial g(x^k)$ ,  $(s^k, x^k - \bar{x}) = (\partial g(x^k), x^k - \bar{x}) \geq g(x^k) - g(\bar{x}) \geq -g(\bar{x}) \geq \delta$ .

Итак, если  $x^k \in S_\varepsilon$ , то

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - 2\alpha\gamma_k + \gamma_k^2 c, \quad \alpha = \min\{\delta, \varepsilon - \delta\} > 0.$$

Стандартным образом (см., например, доказательство теоремы 1 § 3 гл. 5) отсюда следует, что это неравенство не может выполняться при всех  $k$ . Итак, для всякого  $\varepsilon > 0$  найдется  $x^k \in S_\varepsilon$ . Поскольку  $\bigcap_{\varepsilon > 0} S_\varepsilon = X^*$ , где  $X^*$  — множество решений (42), то все предельные точки  $x^k$  принадлежат  $X^*$ . ▲

<sup>9\*</sup>

Другой путь применения субградиентного метода к задаче (33) заключается в ее сведении к задаче безусловной минимизации на основе следующей леммы.

**Лемма 2.** *Пусть выполняется условие Слейтера, а множество точек минимума функции*

$$\Phi(x) = f(x) + K \sum_{i=1}^m g_i(x)_+, \quad (36)$$

на  $Q$  при всех  $K > \bar{K} = \max_{1 \leq i \leq m} y_i^*$ , где  $y_i^*$  — множители Лагранжа.

Этот результат немедленно следует из теоремы Куна — Таккера и формулы для  $\partial\Phi(x)$  (основываящейся на леммах 10 и 11 § 1 гл. 5). ▲

Таким образом, можно выбрать достаточно большое  $K$  и применить метод проекции субградиента для минимизации  $\Phi(x)$ :

$$x^{k+1} = P_Q(x^k - v_k \partial\Phi(x^k)) = \\ = P_Q \left[ x^k - v_k \left( \partial f(x^k) + K \sum_{i=1}^m \mu_{ik} \partial g_i(x^k) \right) \right], \quad (37)$$

$$\mu_{ik} = \begin{cases} 0, & g_i(x^k) \leq 0, \\ 1, & g_i(x^k) > 0 \end{cases}$$

Поскольку нижняя граница  $K$  для  $K$  обычно неизвестна, можно сделать этот параметр переменным, увеличивая его на каждом шаге

$$x^{k+1} = P_Q \left[ x^k - v_k \left( \partial f(x^k) + K_k \sum_{i=1}^m \mu_{ik} \partial g_i(x^k) \right) \right], \quad K_k \rightarrow \infty. \quad (38)$$

Описанные выше методы сходятся, естественно, не быстрее, чем субградиентные методы безусловной минимизации. При этом для больших  $K$  функция  $\Phi(x)$  (36) имеет сильно вытянутые линии уровня, и ее минимизация затруднена. Поэтому трудно рассчитывать на эффективность подобных методов.

Способы ускорения сходимости субградиентных методов, такие как и в задаче безусловной минимизации, связаны с использованием на  $k$ -й итерации информации, полученной на предыдущих итерациях.

Наиболее прямолинейный подход заключается в построении кусочно-линейной аппроксимации минимизируемой функции и ограничений. Пусть уже вычислены субградиенты в точках  $x^1, \dots, x^k$ . Тогда функция  $f(x)$  может быть заменена выражением  $\max_{1 \leq j \leq k} [f(x^j) + (\partial f(x^j), x - x^j)]$ , аналогично можно поступить с  $g_i(x)$ . Тогда исходная задача аппроксимируется сле-

дующей:

$$\min z,$$

$$f(x') + (\partial f(x'), x - x') \leq z, \quad j = 1, \dots, k, \\ g_i(x') + (\partial g_i(x'), x - x') \leq 0, \quad i = 1, \dots, m, j = 1, \dots, k, x \in Q, \quad (39)$$

которая является задачей линейного программирования (если  $Q$  — многогранник). Ее решение может быть взято в качестве  $x^{k+1}$ . Нетрудно убедиться, что в случае отсутствия ограничений  $g_i(x) \leq 0$  этот метод совпадает с методом отсекающей гиперплоскости, описанном в § 4 гл. 5. Можно сократить число ограничений во вспомогательной задаче (39), если все ограничения заменить одним  $g(x) \leq 0$ . Удобно взять  $g(x) = \max_{1 \leq i \leq m} g_i(x)$ , так как вычисление субградиента  $g(x)$  требует вычисления лишь одного субградиента  $g_i(x)$  (именно,  $\partial g(x) = \partial g_i(x)$ ,  $j = \arg\max_{1 \leq i \leq m} g_i(x)$ ).

Есть и другие способы построения кусочно-линейных аппроксимаций задачи выпуклого программирования. Кроме того, многие методы, приведенные в § 4 гл. 5 для задач безусловной минимизации (методы чебышевских центров, центра тяжести, методы с растяжением пространства и т. д.) переносятся и на задачи с ограничениями.

Наконец, описанные выше в данном параграфе методы, использующие двойственные переменные, естественным образом обобщаются на негладкий случай — нужно лишь градиентный метод заменить на метод проекции субградиента. В частности, простейший алгоритм (13) принимает вид

$$x^{k+1} = P_Q[x^k - v_k \partial_x L(x^k, y^k)] =$$

$$= P_Q \left[ x^k - v_k \left( \partial f(x^k) + \sum_{i=1}^m y_i^k \partial g_i(x^k) \right) \right],$$

$$y^{k+1} = [y^k + v_k \partial_y L(x^k, y^k)]_+ = [y^k + v_k g(x^k)]_+, \quad (40)$$

$$v_k \rightarrow 0, \quad \sum_{k=0}^{\infty} v_k = \infty.$$

Используя тот же пример (14), нетрудно проверить, что метод (49) без дополнительных предположений типа строгой выпуклости  $f(x)$  не сходится. Вопрос о сходимости метода до конца не решен.

Если минимум  $L(x, y)$  по  $x \in Q$  при всяком  $y \in \mathbf{R}_+^m$  существует и его легко найти, то можно применить аналог метода (16):

$$x^{k+1} = \arg\min_{x \in Q} L(x, y^k),$$

$$y^{k+1} = [y^k + v_k g(y^k)]_+, \quad v_k \rightarrow 0, \quad \sum_{k=0}^{\infty} v_k = \infty. \quad (41)$$

Его интерпретация как субградиентного метода решения двойственной задачи

$$\max_{y \geq 0} \psi(y), \quad \psi(y) = \min_{x \in Q} L(x, y) \quad (42)$$

позволяет получить результаты о сходимости.

К сожалению, в негладких задачах редко удается найти минимум  $L(x^*, y)$  по  $x$  в явном виде, что затрудняет применение метода (41) (а также его аналогов с заменой  $L(x, y)$  на  $M(x, y, K)$ ).

#### Упражнения.

9. Исследуйте модификацию метода (35), в котором в качестве  $g(x)$  взято

$$(34) \quad g(x) = \sum_{i=1}^m g_i(x)_+$$

10. Рассмотрите другие правила выбора длины шага в методе (35), аналогичные описанному в § 3 гл. 5. Исследуйте их сходимость и проверьте, в каких случаях можно отбросить требование ограниченности  $Q$ .

11. Проанализируйте метод типа (35), в котором  $s^k = \partial f(x^k)$  при  $g(x^k) \leq -\varepsilon_k$ ,  $s^k = \partial g(x^k)$  при  $g(x^k) > -\varepsilon_k$ ,  $\varepsilon_k \geq 0$ ,  $\varepsilon_k \rightarrow 0$ .

**6. Обсуждение.** В § 3 гл. 8 были указаны трудности, которые могут возникнуть при практическом решении задач с ограничениями типа равенств. Посмотрим, в какой мере эти трудности присущи методам выпуклого программирования.

Прежде всего, проблема многоэкстремальности, представляемая основную сложность для невыпуклого случая, не возникает в выпуклых задачах. Действительно, все вспомогательные задачи минимизации, рассматривавшиеся в данном параграфе (для метода линеаризации, двойственных методов, методов штрафов и т. д.), являются выпуклыми, так что локальные минимумы в них совпадают с глобальными.

Далее, утверждения о сходимости методов выпуклого программирования носят глобальный характер. Поэтому нет необходимости иметь хорошие начальные приближения или опасаться расходимости методов. В связи с этим не представляет особой опасности наличие помех — если в задачах гл. 8 помех могли вывестись из области сходимости методов, то в выпуклом случае этого не происходит. Мы не будем останавливаться на подробном анализе влияния помех (он во многом аналогичен приведенному в гл. 4, § 5 гл. 5 и § 4 гл. 7), ограничившись этим кратким замечанием.

В невыпуклых задачах может случиться, что вспомогательные задачи минимизации в том или ином методе не имеют решения (например, ограничения в них противоречивы), даже если в исходной задаче минимум достигается. В выпуклом программировании эта ситуация невозможна. Например, задача квадратичного программирования (7) в методе линеаризации всегда имеет решение, поскольку ограничениям в (7) удовлетворяет

любая допустимая точка (1). Если окажется, что во вспомогательной задаче ограничения противоречивы, то это же имеет место в исходной задаче.

Таким образом, в задачах выпуклого программирования отсутствуют многие осложнения, свойственные более общим задачам.

## § 4. Методы нелинейного программирования

Рассмотрим ту же задачу, что и в § 2:

$$\min f(x), \quad x \in \mathbf{R}^n, \\ g_i(x) \leq 0, \quad i = 1, \dots, r, \\ g_i(x) = 0, \quad i = r + 1, \dots, m, \quad (1)$$

где  $f(x)$ ,  $g_i(x)$  — дифференцируемые функции. Методы ее решения и результаты о сходимости во многом близки к описанным в § 2 гл. 8 и в § 3 этой главы, где исследовались частные случаи задачи (1). Поэтому изложение будет сравнительно кратким.

### 1. Метод линеаризации.

Идея этого метода, заключающаяся в линеаризации минимизируемой функции и ограничений, хорошо знакома нам. Для общей задачи (1) метод принимает вид

$$\begin{aligned} \min & [(\nabla f(x^k), x - x^k) + (2y)^{-1} \|x - x^k\|^2], \\ & g_i(x^k) + (\nabla g_i(x^k), x - x^k) \leq 0, \quad i \in I^*, \\ & g_i(x^k) + (\nabla g_i(x^k), x - x^k) = 0, \quad i = r + 1, \dots, m, \\ & I_k^* = \{i: 1 \leq i \leq r, \quad g_i(x^k) \geq -\varepsilon\}, \quad \varepsilon > 0. \end{aligned} \quad (2)$$

В качестве  $x^{k+1}$  берется решение этой задачи, и процесс повторяется. Обратите внимание, что линеаризуются не все ограничения, а лишь « $\varepsilon$ -активные», т. е. для  $i \in I_k^*$ . Вместо задачи квадратичного программирования (2) можно ввести двойственную ей (см. (12), § 4 гл. 10):

$$\min_{y \in S_k} \left[ y \left\| \nabla f(x^k) + \sum_{i=1}^m y_i \nabla g_i(x^k) \right\|^2 - \sum_{i=1}^m y_i g_i(x^k) \right], \quad (3)$$

$$S_k = \{y \in \mathbf{R}^m: y_i \geq 0, \quad i \in I_k^*; \quad y_i = 0, \quad i \in I_k^*, \quad 1 \leq i \leq r\}.$$

Эта задача минимизации квадратичной функции при условиях неотрицательности части переменных может быть решена методом сопряженных градиентов (§ 3 гл. 7) за конечное число шагов. Найдя ее решение  $y_k^*$ , можно вычислить новое приближение для прямых переменных (см. (13) § 4 гл. 10):

$$x^{k+1} = x^k - y \left( \nabla f(x^k) + \sum_{i=1}^m y_i \nabla g_i(x^k) \right). \quad (4)$$

Поскольку мы не делаем никаких предположений о выпуклости, то естественно, что результаты о сходимости метода носят локальный характер. Кроме того, требуется невырожденность решения (см. § 2).

**Теорема 1.** *Пусть  $x^*$  — невырожденная точка минимума (1). Тогда найдутся  $\bar{\gamma} > 0$ ,  $\bar{\varepsilon} > 0$  такие, что при  $0 < \gamma < \bar{\gamma}$ ,  $0 < \varepsilon < \bar{\varepsilon}$  метод (2) локально сходится к решению со скоростью геометрической прогрессии.*

Наметим схему доказательства. Во-первых, ясно, что для достаточно малых  $\varepsilon > 0$  и  $x^k$ , достаточно близких к  $x^*$ , будет  $I_k^* = I^*$ . Во-вторых, при этих же предположениях доказывается, что в задаче (2) будет  $g_i(x^k) + (\nabla g_i(x^k), x^{k+1} - x^k) = 0$ ,  $i \in I_k^* = I^*$ , т. е. ограничения, активные для  $x^*$  в (1), будут активными и для решения (2). Таким образом, локально метод будет совпадать с методом линеаризации для задачи с ограничениями типа равенств для активных в  $x^*$  ограничений:

$$\min f(x), \quad g_i(x) = 0, \quad i \in I. \quad (5)$$

Но сходимость последнего со скоростью геометрической прогрессии доказана (теорема 1 § 2 гл. 8). ▲

Оказывается, что для задач с острым минимумом метод линеаризации локально сходится к квадратичной скорости. **Теорема 2.** *Пусть  $x^*$  — невырожденная точка минимума, а число активных ограничений равно  $n$ . Тогда метод (2) при достаточно малых  $\gamma > 0$ ,  $\varepsilon > 0$  локально сходится к решению с квадратичной скоростью.*

Доказательство основывается на том, что при сделанных предположениях (2) локально совпадает с методом Ньютона для решения системы  $g_i(x) = 0$ ,  $i \in I$ , а последний сходится локально с квадратичной скоростью. ▲

**2. Ньютоновский и квазиньютоновский методы.** Если доступны вторые производные всех функций  $f(x)$ ,  $g_i(x)$  и их вычисление не слишком трудоемко, то можно применить следующий аналог метода Ньютона. В точке  $x^k$  ограничения линеаризуются, функция Лагранжа аппроксимируется квадратичной, и решается вспомогательная задача квадратичного программирования

$$\min [(L'_x(x^k, y^k), x - x^k) + (H_k(x - x^k), x - x^k)/2],$$

$$\begin{aligned} g_i(x^k) + (\nabla g_i(x^k), x - x^k) &\leq 0, \quad i \in I^*, \\ g_i(x^k) + (\nabla g_i(x^k), x - x^k) &= 0, \quad i = r+1, \dots, m, \\ I_k^* &= \{i: 1 \leq i \leq r, g_i(x^k) \geq -\varepsilon\}, \quad \varepsilon > 0. \end{aligned} \quad (6)$$

$F_{k*}$ ,  $G_i^k$  — аппроксимации для  $\nabla^2 f(x^k)$ ,  $\nabla^2 g_i(x^k)$ , построенные по значениям градиентов этих функций в предыдущих точках с помощью рекуррентных формул (см. § 3 гл. 3).

Близкий вариант метода заключается в решении на  $k$ -й итерации вспомогательной задачи минимизации при линейных ограничениях (ср. с (24) § 2 гл. 8):

$$\min_{i \in I_k} L(x, y^k), \quad (8)$$

$$\begin{aligned} g_i(x^k) + (\nabla g_i(x^k), x - x^k) &\leq 0, \quad i \in I_k, \\ g_i(x^k) + (\nabla g_i(x^k), x - x^k) &= 0, \quad i = r+1, \dots, m. \end{aligned} \quad (9)$$

Правило пересчета  $y^k$  для метода (9) остается прежним. Для (9) справедливы те же утверждения о сходимости, что и для метода Ньютона (6).

Серьезный недостаток методов типа (6) заключается в возможной невыпуклости вспомогательных задач. Дело в том, что матрица  $L''_{xx}(x^k, y^k)$ , вообще говоря, не является положительно определенной даже для  $x^k, y^k$ , сколь угодно

а вторые производные функций  $f(x)$ ,  $g_i(x)$  удовлетворяют условию Липшица в окрестности  $x^*$ . Тогда для достаточно малых  $\varepsilon > 0$  метод (6) локально сходится к  $x^*$ ,  $y^*$  с квадратичной скоростью.

Можно доказать эту теорему, показав сначала, что метод (6) локально совпадает с методом Ньютона для решения задачи (5), и используя результат о сходимости последнего (теорема 9 § 2 гл. 8). ▲

Метод (6), вообще говоря, очень трудоемок — он требует вычисления вторых производных минимизируемой функции и ограничений и решения задачи квадратичного программирования на каждом шаге. От первого недостатка можно избавиться, перейдя к квазиньютоновским методам подобно тому, как это делалось для безусловной минимизации (§ 3 гл. 3). Именно, вместо (6) на каждом шаге нужно решать вспомогательную задачу

$$\min [(L'_x(x^k, y^k), x - x^k) + (H_k(x - x^k), x - x^k)/2],$$

$$\begin{aligned} g_i(x^k) + (\nabla g_i(x^k), x - x^k) &\leq 0, \quad i \in I^*, \\ g_i(x^k) + (\nabla g_i(x^k), x - x^k) &= 0, \quad i = r+1, \dots, m, \\ I_k^* &= \{i: 1 \leq i \leq r, g_i(x^k) \geq -\varepsilon\}, \quad \varepsilon > 0, \end{aligned} \quad (7)$$

где матрица  $H_k$  является аппроксимацией для  $L''_{xx}(x^k, y^k)$  и имеет вид

$$H_k = F_k + \sum_{i \in I_k} y_i^k G_i^k, \quad I_k = \{i: i \in I_k^* \cup i = r+1, \dots, m\}, \quad (8)$$

$F_k$ ,  $G_i^k$  — аппроксимации для  $\nabla^2 f(x^k)$ ,  $\nabla^2 g_i(x^k)$ , построенные по значениям градиентов этих функций в предыдущих точках с по-

мощью рекуррентных формул (см. § 3 гл. 3).

Близкий вариант метода заключается в решении на  $k$ -й итерации вспомогательной задачи минимизации при линейных ограничениях (ср. с (24) § 2 гл. 8):

$$\min_{i \in I_k} L(x, y^k), \quad (8)$$

$$\begin{aligned} g_i(x^k) + (\nabla g_i(x^k), x - x^k) &\leq 0, \quad i \in I_k, \\ g_i(x^k) + (\nabla g_i(x^k), x - x^k) &= 0, \quad i = r+1, \dots, m. \end{aligned} \quad (9)$$

Правило пересчета  $y^k$  для метода (9) остается прежним. Для (9) справедливы те же утверждения о сходимости, что и для метода Ньютона (6).

Серьезный недостаток методов типа (6) заключается в возможной невыпуклости вспомогательных задач. Дело в том, что матрица  $L''_{xx}(x^k, y^k)$ , вообще говоря, не является положительно определенной даже для  $x^k, y^k$ , сколь угодно

ближких к  $x^*, y^*$ . Поэтому целевая функция может оказаться невыпуклой, а вся вспомогательная задача (6) в целом — многоэкстремальной, что существенно затрудняет решение. В связи с этим лучше использовать методы типа (7) с положительно определенными матрицами  $H_k$ . Например, можно начать с  $H_0 = I$  (что соответствует методу линеаризации) и при пересчете  $H_k$  следить за тем, чтобы знакопределеннность не нарушалась.

Наконец, сходимость метода Ньютона существенно локальная. С этим можно бороться, применяя «гибриды» метода линеаризации и Ньютона. Они имеют вид (7), где  $H_k$  — некоторая «промежуточная» между  $L''_{xx}(x^k, y^k)$  и  $\frac{1}{\gamma} I$  матрица — например, их сумма — как в методе Левенберга — Марквардта (см. (16) § 1 гл. 3) для безусловной минимизации. Однако и метод линеаризации не обладает, вообще говоря, глобальной сходимостью, поэтому такой компромисс не всегда гарантирует преодоление указанной трудности.

**3. Другие методы.** Поскольку наиболее мощные методы решения задачи нелинейного программирования (типа метода Ньютона) обладают, как мы видели, серьезными недостатками, то это стимулирует интерес к более простым и «грубым» методам. Ограничимся качественным обсуждением наиболее важных из них.

1°. **Методы штрафных функций.** Один из простейших вариантов метода основан на учете всех ограничений с помощью внешних квадратичных штрафов. Иначе говоря, решается последовательность вспомогательных задач безусловной минимизации

$$\min_{x \in \mathbb{R}^n} f_k(x), \quad f_k(x) = f(x) +$$

$$+ \frac{1}{2} K_k \left( \sum_{i=1}^m g_i(x)_+^2 + \sum_{i=r+1}^m g_i(x)^2 \right), \quad K_k \rightarrow \infty. \quad (10)$$

**Теорема 4.** Пусть  $x^*$  — локально единственное решение задачи (1), а функции  $f(x)$ ,  $g_i(x)$  непрерывно дифференцируемы в окрестности  $x^*$ . Тогда для достаточно больших  $K_k$  найдется точка  $x^k$  локального минимума  $f_k(x)$  в окрестности  $x^*$ , и  $x^k \rightarrow x^*$ .  $\blacktriangleleft$

Подчеркнем некоторые особенности метода (10) и приведенного результат о его сходимости. Во-первых, сходимость доказана при минимальных предположениях — не требуется ни условий регулярности (а тем более невырожденности) минимума, ни слишком сильных предположений о гладкости. Несколько отличительным является требование единственности, однако и при его отсутствии можно получить некоторые результаты о сходимости.

Разумеется, методу штрафных функций для задачи (1) присущ общий недостаток этой группы методов — при больших значениях коэффициента штрафа задача безусловной минимизации становится плохо обусловленной, и потому сложность решения вспомогательных задач возрастает. Кроме того, эти задачи, вообще говоря, многоэкстремальны (теорема 4 утверждает лишь, что среди локальных минимумов  $f_k(x)$  найдется близкий к  $x^*$ ). В основном все сказанное выше о конкретном методе (10) относится и к иным близким подходам, описанным в § 3.

2°. **Двойственные методы.** Нетрудно скомбинировать методы, описанные ранее для частных задач (только с равенствами или неравенствами) так, чтобы они были пригодны для общей задачи (1). Мы предоставляем это читателю в качестве упражнения. Однако эти методы вряд ли можно отнести к ряду «грубых и надежных» по причинам, обсуждавшимся выше.

3°. **Метод приведенного градиента.** Довольно общий подход к задаче (1) заключается в следующем. Пусть уже получено приближение  $x^k$ , тогда можно выделить  $\varepsilon$ -активные ограничения (т. е. равенства и те неравенства, для которых  $|g_i(x^k)| \leq \varepsilon$ ,  $\varepsilon > 0$ ), отбросить остальные и применить для задачи с равенствами одну итерацию какого-либо из методов § 2 гл. 8. (Ранее в методах линеаризации и Ньютона мы поступали несколько иначе: активные ограничения — неравенства включали во вспомогательные задачи в качестве неравенств.) Часто удобно использовать в этой роли метод приведенного градиента (§ 2 гл. 8). Естественно, при этом возникает множество сложных проблем (возможность расходимости для плохих начальных приближений, выбор правил регулирования  $\varepsilon$  и точности решения вспомогательных задач и т. д.). Тем не менее метод приведенного градиента является одним из наиболее употребимых в программах для решений общей задачи математического программирования.

Наконец, наиболее общими являются *задачи со смешанными ограничениями*:

$$\begin{aligned} \min(c, x), \quad x \in \mathbf{R}^n, \\ A_1x = b^1, \quad A_2x \leqslant b^2, \quad x \geqslant 0. \end{aligned} \quad (4)$$

## Г л а в а 10

### ЛИНЕЙНОЕ И КВАДРАТИЧНОЕ ПРОГРАММИРОВАНИЕ

Мы уже многократно сталкивались с задачами линейного или квадратичного программирования. Они возникали как вспомогательные при решении более общих задач выпуклого или нелинейного программирования и при безусловной минимизации негладких функций. Задачи линейного и квадратичного программирования представляют большой интерес и сами по себе вследствие их фундаментальной роли в экономических, технических и других прикладных проблемах. Математическая теория таких задач проста, красива и завершена. Общие методы решения применительно к задачам линейного и квадратичного программирования приобретают особенно наглядный вид и обладают многими замечательными свойствами. Эти задачи составят объект рассмотрения в данной главе.

#### § 1. Линейное программирование (теория)

**1. Типы задач.** В задаче линейного программирования минимизируемая функция и ограничения линейны:

$$\begin{aligned} \min(c, x), \quad x \in \mathbf{R}^n, \\ (a^i, x) \leqslant b_i, \quad i = 1, \dots, m, \end{aligned}$$

или в матричной форме

$$\min(c, x), \quad x \in \mathbf{R}^n,$$

$$Ax \leqslant b,$$

где  $A$  — матрица  $m \times n$ ,  $m < n$ , то, представив ее в виде  $A = (A_1 | A_2)$ , где  $A_1$  — матрица  $m \times m$ ,  $A_2$  — матрица  $m \times (n-m)$ , и разбив переменные  $x$  на две группы  $x = \{u, v\}$ ,  $u \in \mathbf{R}^m$ ,  $v \in \mathbf{R}^{n-m}$ , можем записать (9) в виде

$$A_1u + A_2v = b, \quad u \geqslant 0, \quad v \geqslant 0. \quad (9)$$

Исключив  $v$  из системы равенств (что возможно, если  $A_1$  невырождена), получаем ограничения типа неравенств для  $v$ :

$$A_1^{-1}A_2v \leqslant A_1^{-1}b, \quad v \geqslant 0. \quad (10)$$

Наконец, всегда можно сделать так, чтобы минимизируемая функция зависела только от одной переменной. Действительно, имеется условие неотрицательности переменных, то его удобно не включать в матрицу  $A$  и выделить отдельно:

$$\begin{aligned} \min(c, x), \quad x \in \mathbf{R}^n, \\ Ax \leqslant b, \quad x \geqslant 0. \end{aligned} \quad (2)$$

Допустимы ограничения типа равенств, например задача может иметь вид

$$\min(c, x), \quad x \in \mathbf{R}^n,$$

$$Ax = b, \quad x \geqslant 0. \quad (3)$$

Каждый указанный тип задач может быть сведен к любому другому с помощью одного из следующих преобразований. Ограничение типа равенств

$$(a^i, x) = b_i \quad (5)$$

эквивалентно двум неравенствам

$$(a^i, x) \leqslant b_i, \quad -(a^i, x) \leqslant -b_i. \quad (6)$$

Неравенство

$$(a^i, x) \leqslant b_i \quad (7)$$

может быть сведено к равенству и условию неотрицательности дополнительного переменного  $z_i$ :

$$(a^i, x) + z_i = b_i, \quad z_i \geqslant 0. \quad (8)$$

Наконец, в системе равенств можно выразить одни переменные через другие и исключить их. Так, если условия имеют вид

$$Ax = b \quad x \geqslant 0, \quad (9)$$

где  $A$  — матрица  $m \times n$ ,  $m < n$ , то, представив ее в виде  $A = (A_1 | A_2)$ , где  $A_1$  — матрица  $m \times m$ ,  $A_2$  — матрица  $m \times (n-m)$ , и разбив переменные  $x$  на две группы  $x = \{u, v\}$ ,  $u \in \mathbf{R}^m$ ,  $v \in \mathbf{R}^{n-m}$ , можем записать (9) в виде

$$A_1u + A_2v = b, \quad u \geqslant 0, \quad v \geqslant 0.$$

Исключив  $v$  из системы равенств (что возможно, если  $A_1$  невырождена), получаем ограничения типа неравенств для  $v$ :

$$A_1^{-1}A_2v \leqslant A_1^{-1}b, \quad v \geqslant 0. \quad (11)$$

Описанные преобразования дают возможность распространить любой результат или метод решения на задачи, записанные в произвольной форме.

$$\begin{aligned} \min t, \quad (c, x) = t = 0, \quad x \in Q. \\ \min t, \quad (c, x) = t = 0, \quad x \in Q. \end{aligned} \quad (12)$$

Несколько слов о геометрической интерпретации задачи линейного программирования. Линейные ограничения выделяют многогранное множество (не обязательно ограниченное) в пространстве переменных, и задача заключается в минимизации линейной функции на этом множестве (рис. 41, а). Для формы (3) допустимое множество образуется пересечением линейного (12) представляет собой задачу об отыскании «найшер многообразия с неотрицательным ортантом» (рис. 41, б). Наконец, «точки» многогранного множества (рис. 41, в).

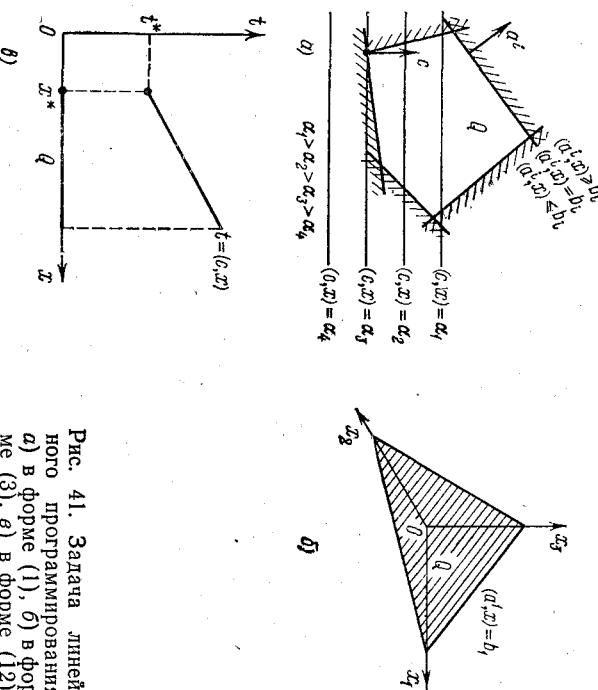


Рис. 41. Задача линейного программирования:  
а) в форме (1), б) в форме (3), в) в форме (12).

**2. Структура многогранных множеств.** Нам понадобится вспомогательный материал, относящийся к теории линейных неравенств (а на геометрическом языке — к теории многогранных множеств). До сих пор мы определяли **многогранное множество**  $Q$  как множество решений системы линейных неравенств:

$$Q = \{x \in \mathbb{R}^n : (a^i, x) \leq b_i, i = 1, \dots, m\}, \quad (13)$$

задающих грани  $Q$  (рис. 41, а). Ниже будет показано, что многогранное множество допускает и иное представление — как выпуклая оболочка своих вершин (если  $Q$  не ограничено, то  $Q$  — выпуклая оболочка вершин и конечного числа лучей) (рис. 42, а).

Точка  $x \in Q$  называется **крайней точкой** множества  $Q \subset \mathbb{R}^n$ , если она не является внутренней точкой какого-либо отрезка, лежащего в  $Q$  (рис. 42, б). Крайние точки многогранного множества легко описать алгебраически.

**Лемма 1.** *Крайними являются те и только те точки множества (13), для которых имеется  $n$  линейно независимых активных ограничений.*

**Доказательство.** Пусть  $x^0 \in Q$ , т. е.  $(a^i, x^0) \leq b_i, i = 1, \dots, m$ ,  $I_0 = \{i : (a^i, x^0) = b_i\}$  — множество активных ограничений. Если среди векторов  $a^i, i \in I_0$ , менее чем  $n$  линейно независимых, то система однородных уравнений  $(a^i, s) = 0$ ,  $i \in I_0$ , имеет ненулевое решение  $s^0$ . Тогда векторы  $x^1 = x^0 + \gamma s^0$ ,  $x^2 = x^0 - \gamma s^0$  при достаточно малом  $\gamma > 0$  также принадлежат  $Q$ , а потому точка  $x^0 = (x^1 + x^2)/2$  не может быть крайней.

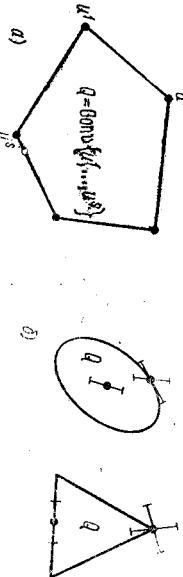


Рис. 42. К теореме о представлении а) выпуклой оболочки конечного числа точек, б) крайние точки множества.

Обратно, пусть среди  $a^i, i \in I_0$ , есть  $n$  линейно независимых. Предположим, что  $x^0 = (x^1 + x^2)/2$ , где  $x^1, x^2 \in Q$ . Тогда для  $i \in I_0$

$$b_i = (a^i, x^0) = [(a^i, x^1) + (a^i, x^2)]/2 \leq (b_i + b_i)/2 = b_i.$$

Отсюда  $(a^i, x^1) = (a^i, x^2) = b_i$ . Но система уравнений  $(a^i, x) = b_i, i \in I_0$  не может иметь два различных решения (ее ранг равен  $n$ ). Поэтому  $x^1 = x^2$ , т. е.  $x^0$  — крайняя точка. ▲

**Следствие.** Число крайних точек многогранного множества **конечно**. ▲

Крайние точки многогранного множества называют также **вершинами**. Лемма 1 показывает, что эта терминология согласуется с геометрическим определением вершины как 0-мерной грани.

Многогранное множество может и не иметь крайних точек (например, если оно является подпространством, см. упр. 1). Однако для ограниченного  $Q$  крайние точки существуют, более того, они «порождают» все множество  $Q$ .

**Лемма 2.** *Непустое ограниченное многогранное множество является выпуклой оболочкой своих вершин.*

**Доказательство.** Пусть  $x^0 \in Q$ ,  $I_0 = \{i : (a^i, x^0) = b_i\}$ . Если  $x^0$  — крайняя точка, то найдутся  $z^1, z^2 \in Q$ ,  $z^1 \neq z^2$  такие, что  $x^0 = (z^1 + z^2)/2$ . Рассмотрим прямую  $x^0 + \lambda(z^2 - z^1)$ ,  $\lambda \in \mathbb{R}^1$ . В силу ограниченности  $Q$  она не может целиком содержаться в  $Q$ , поэтому в некоторой точке  $x^1 = x^0 + \lambda_1(z^2 - z^1)$  она

выходит из  $Q$ . При этом  $I_1 = \{i: (a^i, x^1) = b_i\} \supseteq I_0$ , так как если  $i \in I_0$ , то  $(a^i, (z^1 + z^2)/2) = b_i$ ,  $(a^i, z^1) \leq b_i$ ,  $(a^i, z^2) \leq b_i$ , что означает  $(a^i, z^1) = (a^i, z^2) = b_i$ , поэтому  $(a^i, x^1) = (a^i, x^0) + \lambda_{11}(a^i, z^2) - (a^i, z^1) = b_i$ . Более того, включение  $I_1 \supseteq I_0$  — стягое, так как в точке  $x^1$  добавляется по меньшей мере одно активное ограничение. С точкой  $x^1$  повторим ту же процедуру и придем к точке  $x^2$ ,  $I_2 \supseteq I_1$  и т. д. Поскольку процесс увеличения числа активных ограничений не может продолжаться бесконечно, то в результате придем к крайней точке. Итак, крайние точки существуют; при этом, как следует из построения, для любой точки  $x^0$  найдется крайняя точка  $x^*$ ,  $I_0 \subset I^* = \{i: (a^i, x^*) = b_i\}$ .

Покажем теперь, что любая точка  $x^0 \in Q$  может быть представлена в виде  $x^0 = \sum_{i=1}^s \lambda_i u^i$ ,  $\lambda_i \geq 0$ ,  $\sum_{i=1}^s \lambda_i = 1$ , где  $u^1, \dots, u^s$  — крайние точки. Если  $x^0$  — не крайняя, то найдем вершину  $u^1 \neq x^0$  (по доказанному такая найдется), возьмем соединяющий  $u^1$  и  $x^0$  отрезок и продолжим его в сторону  $x^0$  максимально возможно (т. е. найдем  $x^1 = u^1 + \lambda^*(x^0 - u^1)$ ,  $\lambda^* = \max\{\lambda: u^1 + \lambda(x^0 - u^1) \in Q\} < \infty$ ). В точке  $x^1$  множество активных ограничений  $I_1 \neq \emptyset$ . Найдем крайнюю точку  $u^2$ , для которой  $I_1 \subset \{i: (a^i, u^2) = b_i\}$ . Соединим  $u^2$  с  $x^1$  и продолжим этот отрезок в сторону  $x^1$  максимально возможно, получим  $x^2$ ,  $I_2 \supseteq I_1$ ,  $I_2 \neq I_1$ . Продолжая этот процесс далее, найдем точки  $x^3, x^4, \dots, x^s$  и крайние точки  $u^3, \dots, u^s$ . При этом для некоторого  $s$  окажется  $x^s = u^s$  (так как множество активных ограничений все время расширяется). По индукции легко доказывается, что  $x^0$  — выпуклая комбинация  $u^1, \dots, u^s$ .  $\blacktriangle$

Переходя к неограниченным многогранным множествам, начнем с самого простого из них — *многогранного конуса*

$$Q = \{x \in \mathbf{R}^n: (a^i, x) \leq 0, \quad i = 1, \dots, m\}, \quad (14)$$

и пусть  $A$  — матрица со строками  $a^1, \dots, a^m$ .

**Лемма 3.** *Конус  $Q$  (14) можно представить в виде выпуклой оболочки конечного числа лучей и прямых:*

$$Q = \left\{ x: x = \sum_{i=1}^s \lambda_i u^i + \sum_{i=1}^r \gamma_i w^i, \quad \lambda_i \geq 0, \quad \gamma_i \in \mathbf{R}^1 \right\}, \quad (15)$$

$$u^i, w^i \in Q. \quad \blacktriangle$$

Введем некоторые понятия, связанные с произвольным конусом  $K$ . Луч  $\{\lambda v, \lambda \geq 0\}$ ,  $v \in K$ , называется *крайним* для  $K$ , если он не может быть представлен как полусумма двух различных лучей из  $K$ . Крайние луки (они называются также *направляющими*) играют для конусов ту же роль, что и крайние точки для ограниченных множеств. Конус  $K$  называется *острым*, если он не содержит прямых, т. е. из  $v \in K$ ,  $-v \in K$  следует  $v = 0$ .

В этой терминологии из леммы 3 следует аналог леммы 2 для конусов.

**Лемма 4.** *Острый многогранный конус является выпуклой оболочкой своих крайних лучей, число которых конечно.*  $\blacktriangle$

Теперь мы можем получить основной результат данного раздела.

**Теорема 1** (о представлении многогранного множества).

Пусть

$$Q = \{x \in \mathbf{R}^n: Ax \leq b\} \neq \emptyset. \quad (16)$$

Тогда  $Q$  представимо как выпуклая оболочка конечного числа точек, лучей и прямых.

$$Q = \left\{ x: x = \sum_{i=1}^s \lambda_i u^i + \sum_{i=1}^r \mu_i v^i + \sum_{i=1}^p \nu_i w^i, \quad \begin{aligned} \lambda_i \geq 0, \quad \sum_{i=1}^s \lambda_i = 1, \quad \mu_i \geq 0, \quad \nu_i, \lambda_i, \mu_i \in \mathbf{R}^1 \end{aligned} \right\}, \quad (17)$$

где  $Au^i \leq 0$ ,  $i = 1, \dots, r$ ;  $Aw^i = 0$ ,  $i = 1, \dots, p$ ;  $(w^i, w^j) = 0$ ,  $i \neq j$ . Если  $Q$  не содержит прямых ( $p = 0$ ), то в качестве  $w^i$  можно взять вершины  $Q$ , в качестве  $v^i$  — крайние лучи  $K = \{x: Ax \leq 0\}$ .

Доказательство является комбинацией рассуждений, проводившихся выше. Сначала выделяются  $w^1, \dots, w^p$  такие, что  $Aw^i = 0$ ,  $i = 1, \dots, p$ ;  $(w^i, w^j) = 0$ ,  $i \neq j$  (если такие имеются). Затем рассматривается многогранное множество

$$Q_1 = \{x: A(x^0 + x) \leq b, \quad (x, w^i) = 0, \quad i = 1, \dots, p\},$$

где  $x^0$  — любая точка из  $Q$ , тогда  $0 \in Q_1$ , т. е.  $Q_1 \neq \emptyset$ , и показывается, что  $Q_1$  представима как выпуклая оболочка своих крайних точек и направляющих конуса  $K = \{x: Ax \leq 0, (x, w^i) = 0, i = 1, \dots, p\}$  — это делается так же, как при доказательстве леммы 2, но с учетом того, что  $Q_1$  может содержать лучи. Последние представляются как выпуклая оболочка направляющих  $K$  с помощью леммы 3.  $\blacktriangle$

Вектор  $z \in \mathbf{R}^n$  такой, что луч  $x + \lambda z$ ,  $\lambda \geq 0$ ,  $x \in Q$ , целиком принадлежит  $Q$ , называется *направлением рёбер* выпуклого множества  $Q$ . Из теоремы 1 следует, что для многогранного множества все направления рёбер порождаются конечным числом векторов.

Справедлив и результат, двойственный к теореме 1 — выпуклая оболочка конечного числа точек, лучей и прямых является многогранным множеством, т. е. представима в виде (13).

**Упражнения:**

1. Докажите следующие утверждения:
  - а) всякая крайняя точка  $Q$  является граничной точкой  $Q$ ; в частности, открытое множество не имеет крайних точек;

- б) подпространство не имеет крайних точек;  
в) всякая точка сферы  $\{x: \|x\| = 1\}$  является крайней для шара  $Q = \{x: \|x\| \leq 1\}$ .

2. Покажите с помощью теоремы 1, что для  $Q = \{x: Ax \leq b\} \neq \emptyset$  следующие условия эквивалентны: а)  $Q$  ограничено; б) неравенства  $Ax \leq 0$  имеют только нулевое решение; в) уравнения  $A^T y = c$  имеют решение  $y \geq 0$  для любого  $c$  (используйте лемму Фаркаша).

**3. Условия экстремума.** Теорема 1 в принципе дает возможность найти решение задачи линейного программирования и полностью исследовать его.

**Теорема 2. Решение  $X^*$  задачи**

$$\begin{aligned} &\min(c, x), \quad x \in \mathbf{R}^n, \\ &x \in Q = \{x: Ax \leq b\}, \quad b \in \mathbf{R}^m, \end{aligned} \tag{18}$$

существует в том и только в том случае, когда  $Q \neq \emptyset$  и  $(c, v^i) \geq 0$ ,  $i = 1, \dots, r$ ;  $(c, w^i) = 0$ ,  $i = 1, \dots, p$ , при этом

$$X^* = \left\{ x^*: x^* = \sum_{i \in I_1} \lambda_i u^i + \sum_{i \in I_2} \mu_i v^i + \sum_{i=1}^p \nu_i w^i, \right.$$

$$\left. \lambda_i \geq 0, \quad \sum_{i \in I_1} \lambda_i = 1, \quad \mu_i \geq 0, \quad \nu_i \in \mathbf{R}^1 \right\}, \tag{19}$$

$\partial \partial e I_1 = \{i: (u^i, c) = f^*\}$ ,  $I_2 = \{i: (c, v^i) = 0\}$ ,  $f^* = (c, x^*)$ ,  $x^* \in X^*$ ,  $a u^i, v^i, w^i$  — векторы, фигурирующие в теореме 1.

Действительно, в силу (17) для произвольного  $x \in Q$

$$(c, x) = \sum_{i=1}^s \lambda_i (c, u^i) + \sum_{i=1}^r \mu_i (c, v^i) + \sum_{i=1}^p \nu_i (c, w^i). \tag{20}$$

Минимум этого выражения по  $\lambda_i \geq 0$ ,  $\sum_{i=1}^s \lambda_i = 1$ ,  $\mu_i \geq 0$ ,  $\nu_i \in \mathbf{R}^1$ , очевидно, достигается лишь тогда, когда  $(c, v^i) \geq 0$ ,  $i = 1, \dots, r$ ,  $(c, w^i) = 0$ ,  $i = 1, \dots, p$ , и решением являются любые  $\lambda_i^*$ ,  $\mu_i^*$ ,  $\nu_i^*$  такие, что  $\lambda_i^* = 0$  при  $i \in I_1$ ,  $\lambda_i^* \geq 0$  при  $i \in I_2$ ,  $\sum_{i \in I_1} \lambda_i^* = 1$ ,  $\mu_i^* \geq 0$  при  $i \in I_2$ ,  $\mu_i^* = 0$  при  $i \in I_2$ .  $\blacktriangleleft$

Следствие. Если допустимое множество не содержит промежуточных, а решение задачи (18) существует, то среди решений найдется единична  $Q$ . Если решение единственное, то оно достигается в вершине.  $\blacktriangleleft$

Разумеется, этот результат не может рассматриваться как конструктивный способ отыскания решения, так как число вершин огромно даже для задач небольшой размерности и полный перебор вершин — дело совершенно нереальное. Таким образом, теорема 2 отнюдь не делает тривиальной проблему поиска решения задач линейного программирования, поэтому имеет смысл провести обычный анализ задачи, связанный с условиями экстремума.

Задача (18) является весьма частным случаем задачи выпуклого программирования (§ 1 гл. 9) — минимизируемая функция и ограничения определены для всех  $x \in \mathbf{R}^n$ , выпуклы, дифференцируемы, а ограничения имеют специальный вид, позволяющий отказаться от требования выполнения условия Слейтера. Используя теоремы 2 и 3 § 1 гл. 9, приходим к следующим условиям экстремума.

**Теорема 3.** Для того чтобы допустимая точка  $x^*$  была решением задачи (18), необходимо и достаточно существование множества Лагранжа  $y^* \in \mathbf{R}^m$  таких, что

$$y^* \geq 0, \quad (y^*, Ax^* - b) = 0, \quad c + A^T y^* = 0. \tag{21}$$

Теорема 4. Для того чтобы  $x^* \in \mathbf{R}^n$  являлось решением задачи (18), необходимо и достаточно существование  $y^* \in \mathbf{R}^m$ ,  $y^* \geq 0$  такого, что

$$L(x, y^*) \geq L(x^*, y^*) \geq L(x^*, y) \quad \forall x \in \mathbf{R}^n, \quad \forall y \in \mathbf{R}_+^m, \tag{22}$$

$$\begin{aligned} &\partial \partial e \\ &L(x, y) = (c, x) + (y, Ax - b). \blacktriangleleft \end{aligned} \tag{23}$$

Выпишем задачу, двойственную к (18), и сформулируем теорему двойственности. Поскольку

$$\Psi(y) = \inf_{x \in \mathbf{R}^n} L(x, y) = \inf_{x \in \mathbf{R}^n} [(c + A^T y, x) - (b, y)],$$

то

$$\Psi(y) = \begin{cases} -\infty, & \text{если } c + A^T y \neq 0, \\ -(b, y), & \text{если } c + A^T y = 0. \end{cases}$$

Записывая двойственную задачу в форме (49) § 1 гл. 9, получаем

$$\min(b, y), \quad c + A^T y = 0, \quad y \geq 0. \tag{24}$$

Таким образом, задачей, двойственной к (18), также оказалась задача линейного программирования.

**Теорема 5** (теорема двойственности). Решения  $x^*, y^*$  пары одновременно, при этом  $y^*$  (соответственно  $x^*$ ) — множители Лагранжа для (18) (соответственно для (24)). Для любых допустимых  $x, y$  справедливо неравенство

$$(c, x) + (b, y) \geq 0, \tag{25}$$

причем равенство в (25) достигается тогда и только тогда, когда  $x, y$  — решения задач (18) и (24).  $\blacktriangleleft$

Рекомендуем читателю доказать эту теорему непосредственно на основе теорем 3 и 4, а не обращаясь к общей теореме двой-

ственности из § 1 гл. 9. Удобно при этом использовать приводимые ниже условия экстремума для задачи вида (24) (см. теорему 6).

Результаты теорем 3—5 относились к задаче в форме (18). Если исходная задача задана в иной форме, то, используя предобразования, описанные в п. 1, ее можно привести к виду (18), записать условия экстремума, а затем вернуться к исходным переменным. Если проделать это для задачи

$$\begin{aligned} \min(c, x), \quad x \in \mathbf{R}^n, \\ Ax = b, \quad b \in \mathbf{R}^m, \quad x \geq 0 \end{aligned} \quad (26)$$

(она называется *канонической формой* задачи линейного программирования), то получим следующие результаты.

**Теорема 6.** *Точка  $x^* \in \mathbf{R}^n$  является решением задачи линейного программирования (26) тогда и только тогда, когда найдется  $y^* \in \mathbf{R}^m$  такие, что выполняется любое из условий:*

- a)  $Ax^* = b, \quad x^* \geq 0, \quad A^T y^* \leq c, \quad (A^T y^* - c, x^*) = 0,$
- b)  $x^* \geq 0, \quad L(x, y^*) \geq L(x^*, y^*) \geq L(x^*, y)$

$$\forall x \in \mathbf{R}_+^n, \quad \forall y \in \mathbf{R}^m. \quad \blacksquare \quad (28)$$

Задача, двойственная к (26), имеет вид

$$\max(b, y), \quad A^T y \leq c. \quad (29)$$

**Теорема 7** (теорема двойственности). *Решения  $x^*, y^*$  пары одновременно, при этом  $-y^*$  (соответственно  $x^*$ ) — множества Лагранжа для (26) (соответственно для (29)). Для любых допустимых  $x, y$*

$$(c, x) \geq (b, y), \quad (30)$$

*причем равенство в (30) достигается тогда и только тогда, когда  $x, y$  — решения задачи (26) и (29).*  $\blacktriangle$

Теперь, когда мы умеем выписывать двойственную задачу для различных постановок исходной задачи, можно установить следующий факт, опровергающий термин «двойственность».

**Теорема 8.** *Правильная задача (18) является двойственной к одновременной задаче (24). Это же справедливо для задач (26), (29).*  $\blacktriangle$

Таким образом, можно говорить просто о паре двойственных задач, не уточняя, какая из них является прямой. Отметим, что тот способ построения двойственной задачи, который приведен в § 1 гл. 9, не позволяет сделать такое же заключение для общей задачи выпуклого программирования, так как мы не знаем,

при этом  $(c, x) \geq (b, y)$  для допустимых  $x, y$ , а равенство достигается только для решений.

**Упражнение.** 3. Получите из теоремы 2, что если  $Q$  содержит прямые, то решение задачи (18) не может быть ограниченным. Получите необходимые и достаточные условия ограниченности решения.

**4. Существование, единственность и устойчивость решения.** Как уже отмечалось, теорема 2 в принципе дает полное решение проблем существования и единственности, но носит неконструктивный характер. Нас будут интересовать более просто проверяемые условия существования и единственности решения, которые можно получить с ее помощью.

**Теорема 9.** *Пусть в задаче (18) допустимое множество не пусто, а целевая функция на нем ограничена снизу:*

$$Q = \{x: Ax \leq b\} \neq \emptyset, \quad \inf_{x \in Q} (c, x) > -\infty.$$

Тогда решение задачи (18) существует.

**Доказательство.** Функция  $(c, x)$  на  $Q$  задается (20). Но это выражение ограничено снизу при любых  $\mu_i \geq 0, v_i \in \mathbf{R}^n$ . Лишь тогда, когда  $(c, v^i) \geq 0, i = 1, \dots, r; (c, w^i) = 0, i = 1, \dots, s$ , при этом минимум  $(c, x)$  на  $Q$  достигается (теорема 2).  $\blacktriangle$

Подчеркнем, что в теореме 9 не предполагается ограниченность множества вида  $\{x \in Q: (c, x) \leq \alpha\}$ .

Как следствие из теоремы 9 получаем, что задача безусловной минимизации любой из следующих кусочно-линейных функций

$$f(x) = \sum_{i=1}^m ((a^i, x) - b_i)_+, \quad f(x) = \sum_{i=1}^m |(a^i, x) - b_i|, \quad (32)$$

$$f(x) = \max_{1 \leq i \leq m} ((a^i, x) - b_i)_+, \quad f(x) = \max_{1 \leq i \leq m} |(a^i, x) - b_i|$$

всегда имеет решение. Действительно, эти задачи сводятся к задачам линейного программирования, в которых целевая функция неограничительна.

Другим полезным следствием является следующее уточнение теоремы двойственности.

**Теорема 10.** *Если обе двойственные задачи линейного программирования имеют допустимые точки, то они имеют решение.*

Действительно, это сразу следует из соотношения (25) (для задач (18), (24)) и теоремы 9.  $\blacksquare$

Выяснить, является ли данное решение единственным, можно совсем просто.

**Теорема 11.** Пусть  $x^*$  — решение задачи (1),  $I^* = \{i: (a^i, x^*) = b_i\}$  — множество активных ограничений. Тогда если среди векторов  $a^i, i \in I^*$ , не найдется  $n$  линейно независимых, то  $x^*$  — не единствено, а если  $n$  таких векторов найдется и им отвечают  $y_i^* > 0$ , то решение единствено.

Доказательство. По следствию из теоремы 2 единственный минимум достигается в вершине, что вместе с леммой 1 дает первое утверждение теоремы. Пусть теперь  $a^i, i \in I \subset I^*$ ,  $i$  линейно независимых векторов из  $I^*$ , тогда уравнение  $(a^i, x) = b_i, i \in I$ , имеет единственное решение  $x^*$ . Поэтому для любого  $x \in Q$ ,  $x \neq x^*$ , найдется  $j \in I$  такое, что  $(a^j, x) < b_j$ . Тогда в соответствии с теоремой 3

$$(c, x) = -(A^T y^*, x) = -\sum_{i \in I^*} y_i^* (a^i, x) > -\sum_{i \in I^*} y_i^* b_i =$$

$$= -\sum_{i \in I} y_i^* (a^i, x^*) = -(A^T y^*, x^*) = (c, x^*). \blacksquare$$

Что касается устойчивости решения, то оказывается, что для задач линейного программирования всегда выполняется соотношение, обобщающее условие острого минимума (11) § 1 гл. 7. **Лемма 5.** Пусть  $X^*$  — множество решений задачи (18),  $X^* \neq \emptyset$ . Тогда найдется  $\alpha > 0$  такое, что  $(c, x) - f^* \geq \alpha \rho(x, X^*)$  для всех допустимых  $x$ . Здесь  $f^* = (c, x^*)$ ,  $x^* \in X^*$ ,  $\rho(x, X^*) = \min_{x^* \in X^*} \|x - x^*\|$ .

Приведем доказательство для случая, когда  $Q$  ограничено. Тогда  $Q = \left\{ x: x = \sum_{i=1}^s \lambda_i u^i, \lambda_i \geq 0, \sum_{i=1}^s \lambda_i = 1 \right\}$ ,  $X^* = \left\{ x: x = \sum_{i \in I^*} \lambda_i^* u^i, \lambda_i^* \geq 0, \sum_{i \in I^*} \lambda_i^* = 1 \right\}$ , где  $I^* = \{i: (c, u^i) = f^*\}$  (см. теоремы 1, 2).

Введем  $\alpha = \min_{i \in I^*} \frac{(c, u^i) - f^*}{\rho(u^i, X^*)}$  и покажем, что это  $\alpha$  — искомое. Действительно,  $\alpha > 0$ , так как  $(c, u^i) - f^* > 0$  для всех  $i \in I^*$ . Пусть  $x \in Q$  произвольно,  $x = \sum_{i=1}^s \lambda_i u^i$ ,  $\lambda_i \geq 0$ ,  $\sum_{i=1}^s \lambda_i = 1$ . При меняя к выпуклой функции  $\rho(x, X^*)$  (см. упр. 2 § 1 гл. 5) неравенство Иенсена (лемма 1 § 1 гл. 1), получаем

$$\begin{aligned} \rho(x, X^*) &= \rho\left(\sum_{i=1}^s \lambda_i u^i, X^*\right) \leq \sum_{i=1}^s \lambda_i \rho(u^i, X^*) = \sum_{i \in I^*} \lambda_i \rho(u^i, X^*) \leq \\ &\leq \alpha^{-1} \sum_{i \in I^*} \lambda_i ((c, u^i) - f^*) = \alpha^{-1} \sum_{i=1}^s \lambda_i ((c, u^i) - f^*) = \\ &= \alpha^{-1} ((c, x) - f^*). \blacksquare \end{aligned}$$

Из леммы 5, в частности, вытекает, что единственное решение  $x^*$  задачи линейного программирования удовлетворяет условию острого минимума:

$$(c, x) - f^* \geq \alpha \|x - x^*\| \quad \forall x \in Q. \quad (34)$$

**Лемма 6** (Хоффман). Пусть  $Q = \{x \in \mathbf{R}^n: (a^i, x) \leq b_i, i = 1, \dots, m\} \neq \emptyset$ . Тогда найдется  $\alpha > 0$  такое, что

$$\sum_{i=1}^m ((a^i, x) - b_i)_+ \geq \alpha \rho(x, Q) \quad \forall x \in \mathbf{R}^n. \quad (35)$$

Доказательство. Рассмотрим задачу линейного программирования

$$\min \sum_{i=1}^m t_i, \quad (36)$$

$$(a^i, x) - b_i = t_i - z_i, \quad i = 1, \dots, m,$$

$$t_i \geq 0, \quad z_i \geq 0, \quad i = 1, \dots, m,$$

где  $t_i, z_i$  — дополнительные переменные. Для любого  $x \in \mathbf{R}^n$  вектор  $\{x, t, z\}$ , где  $t_i = ((a^i, x) - b_i)_+$ ,  $z_i = (b_i - (a^i, x))_+$ ,  $i = 1, \dots, m$ , является допустимой точкой в этой задаче, а решения (36) имеют вид  $\{x, 0, (b - Ax)_+\}$ , где  $x \in Q$ . Применяя лемму 5 к этой задаче, получаем (35).  $\blacksquare$

Лемма Коффмана утверждает, что если в какой-либо точке невязка в нарушении линейных неравенств мала, то эта точка близка к множеству допустимых точек. Отметим, что аналогичный результат для задач с выпуклыми неравенствами (лемма 11 § 1 гл. 9) был получен в предположении, что выполняется условие Слейтера, а множество  $Q$  ограничено.

Для задачи линейного программирования из леммы 6 вытекает следующее утверждение об устойчивости.

**Лемма 7.** Пусть  $X^*$  — непустое множество решений задачи (1),  $f^* = (c, x^*)$ ,  $x^* \in X^*$ . Тогда найдется  $\alpha > 0$  такое, что для любого  $x$

$$\rho(x, X^*) \leq \sum_{i=1}^m ((a^i, x) - b_i)_+ + ((c, x) - f^*)_+. \quad (37)$$

Для доказательства достаточно записать  $X^*$  в виде  $X^* = \{x: (a^i, x) \leq b_i, i = 1, \dots, m, (c, x) \leq f^*\}$  и воспользоваться леммой 6.  $\blacksquare$

Из (37) следует, что всякая обобщенно-минимизирующая последовательность  $x_k$  (т. е. такая, что  $(c, x_k) \rightarrow f^*$ ,  $((a^i, x_k) - b_i)_+ \rightarrow 0$ ) сходится к множеству решений задачи линейного програм-

Тот факт, что в задаче линейного программирования всегда выполняется обобщенное условие острого минимума (33), приводит к инвариантности решения при малом возмущении минимизируемой функции.

**Теорема 12.** Пусть множество  $X^*$  решений (18) непусто, функция  $g(x)$  выпукла,  $D(g)^0 \supseteq X^*$ ,  $\|\partial g(x)\| \leq L$  при  $x \in X^*$  и  $g(x)$  достигает минимума на  $X^*$ :  $X_g^* = \underset{x \in X^*}{\operatorname{Argmin}} g(x) \neq \emptyset$ . Тогда для достаточно малых  $\varepsilon > 0$  минимум в задаче

$$\min_{x \in Q} [(c, x) + \varepsilon g(x)] \quad (38)$$

достигается на множестве  $X_g^*$ .

Доказательство. Пусть  $x$  — произвольная точка из  $Q$ ,  $x^*$  — проекция  $x$  на  $X_g^*$ ,  $x_g^*$  — точка из  $X_g^*$ . Тогда в силу леммы 5

$$(c, x) + \varepsilon g(x) \geq f^* + \alpha \|x - x^*\| + \varepsilon (g(x_g^*) + (\partial g(x_g^*), x - x_g^*)).$$

Но  $(\partial g(x_g^*), x - x_g^*) = (\partial g(x_g^*), x - x^*) + (\partial g(x_g^*), x^* - x_g^*)$ , где  $(\partial g(x_g^*), x_g^* - x_g^*) \geq 0$  в силу условий экстремума  $g(x)$  на  $X^*$ . Поэтому

$$(c, x) + \varepsilon g(x) \geq f^* + \varepsilon g(x_g^*) + \alpha \|x - x^*\| + \varepsilon (\partial g(x_g^*), x - x^*) \geq$$

$$\geq (c, x_g^*) + \varepsilon g(x_g^*) + (\alpha - \varepsilon L) \|x - x^*\| \geq (c, x_g^*) + \varepsilon g(x_g^*)$$

при  $\varepsilon < \alpha/L$ .  $\blacktriangleleft$

Из этого важного результата, как мы увидим в дальнейшем, вытекает конечность многих итеративных методов линейного программирования. Применяя теорему 12 к двойственной задаче, получаем, что при малых возмущениях ограничений множители Лагранжа остаются прежними (упр. 4). Отсюда следует, что теорема 9 § 1 гл. 9 для задач линейного программирования может быть уточнена (упр. 5).

#### Упражнение.

4. Покажите, что если решения задач  $\min(c, x)$ ,  $Ax \leq b$  и  $\min(c, x)$ ,  $Ax \leq b + \varepsilon$  существуют, то для достаточно малых  $\varepsilon \in \mathbf{R}^m$  множители Лагранжа второй задачи являются и множителями Лагранжа для первой задачи.

5. Пусть решения  $x^*$  задачи  $\min(c, x)$ ,  $Ax \leq b$  и соответствующие ему множители Лагранжа  $y^*$  единственны. Тогда функция  $\varphi(\varepsilon) = \inf_{Ax \leq b+\varepsilon} (c, x)$ ,  $\varepsilon \in \mathbf{R}^m$ , дифференцируема в 0 и  $\nabla \varphi(0) = y^*$ .

6. Пусть множества  $X^* = \operatorname{Argmin}(c, x)$ ,  $X_1^* = \operatorname{Argmin}_{Ax \leq b} (c^1, x)$ , ...,  $X_k^* = \operatorname{Argmin}_{Ax \leq b} (c^k, x)$  непусты. Докажите, что тогда решением задачи  $\min_{x \in X_{k-1}^*} [(c, x) + \varepsilon (c^1, x) + \dots + \varepsilon^k (c^k, x)]$  при достаточно малых  $\varepsilon > 0$  является  $x \in X_k^*$ .

7. Используя предыдущий результат, покажите, что неединственность решения устраняется малым возмущением, т. е. если решение задачи  $Ax \leq b$  не единствено, то найдутся такие  $\varepsilon > 0$  и  $c^1 \in \mathbf{R}^n$ , что решение задачи  $\min(c + \varepsilon c^1, x)$ ,  $Ax \leq b$  будет единствено.

## § 2. КОНЕЧНЫЕ МЕТОДЫ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

### 1. Симплекс-метод

Как мы знаем (следствие из теоремы 2 § 1), решение задачи линейного программирования достигается в вершине многогранного множества. С другой стороны, лемма 1 § 1 дает простое описание вершин как решений систем линейных уравнений. Поскольку число вершин конечно, в принципе можно найти решение за конечное число шагов (каждый из которых включает решение системы линейных уравнений, т. е. требует конечного числа арифметических действий) путем перехода вершин.

Симплекс-метод использует, по существу, ту же идею, но реализует ее не столь прямолинейно. Во-первых, просмотр вершин ведется таким образом, что значения целевой функции монотонно убывают. При этом перебор значительно сокращается — отбрасываются те вершины, в которых значения  $(c, x)$  большие уже найденных. Во-вторых, перебор идет по соседним вершинам, поэтому система линейных уравнений, которую приходится решать на очередном шаге, мало отличается от предыдущей. Используются специальные экономные приемы решения таких систем.

Симплекс-метод применяют для задач, записанных в канонической форме:

$$\min(c, x), \quad Ax = b, \quad x \geq 0. \quad (1)$$

Здесь  $x \in \mathbf{R}^n$ ,  $A$  — матрица  $m \times n$ ,  $b \in \mathbf{R}^m$ ,  $c \in \mathbf{R}^n$ . Поскольку в § 1 мы в основном имели дело с иной формой записи допустимого множества, переформулируем полученные там результаты.

#### Лемма 1. Пусть

$$Q = \{x \in \mathbf{R}^n: Ax = b, x \geq 0\} \neq \emptyset, \quad b \in \mathbf{R}^m. \quad (2)$$

Тогда

- а)  $Q$  — многогранное множество, не содержащее прямых;
- б) вершины  $Q$  являются точки  $x$  из  $Q$ , для которых некоторые  $A_i^t$ ,  $i \in I = \{i: x_i > 0\}$ ;
- в)  $Q$  представимо в виде выпуклой оболочки своих вершин и крайних лучей конуса  $K = \{x \geq 0, Ax = 0\}$ .  $\blacktriangleleft$

В соответствии с леммой 1 в вершине  $Q$  не более  $m$  компонент вектора  $x$  могут быть положительны. Назовем  $x$  невырожденной вершиной, если число положительных компонент в ней равно  $m$ . В дальнейшем предполагается, что допустимое множество в (1) непусто, а все его вершины невырождены. В симплекс-методе используется специфическая терминология, впрочем не всегда одинаковая в различных руководствах. Допустимые точки называют планом, вершину — опорным планом или допустимым базисным решением, решение задачи — оптимальным

*планом*, столбцы  $A$ , отвечающие положительным компонентам опорного плана, — *базисом*. Мы не будем злоупотреблять этой терминологией, а сохраним прежнюю (допустимая точка, вершина, решение и т. д.).

Перейдем к описанию симплекс-метода для решения задачи (1). Пусть на  $k$ -м шаге получена точка  $x^k$ , являющаяся вершиной (2), и

$$I_k = \{i: x_i^k > 0\}.$$

В силу предположения невырожденности  $I_k$  содержит  $m$  элементов. Разобьем вектор  $x \in \mathbb{R}^n$  на две группы  $x = \{u, v\}$ , где  $u \in \mathbb{R}^m$  отвечает компонентам из  $I_k$ ,  $v \in \mathbb{R}^{n-m}$  — компонентам, не принадлежащим  $I_k$ . Тогда система  $Ax = b$  может быть записана в виде

$$A_1 u + A_2 v = b, \quad (4)$$

где  $A_1$  — матрица  $m \times m$ , столбцы которой являются столбцами  $A$  с индексами из  $I_k$ ,  $A_2$  — матрица  $m \times (n-m)$ , составленная из остальных столбцов  $A$ . В силу утверждения б) леммы 1 и условия невырожденности матрицы  $A_1$  имеет обратную, поэтому

$$u = A_1^{-1}(b - A_2 v). \quad (5)$$

Целевая функция приобретает вид

$$(c, x) = (c^1, u) + (c^2, v), \quad (6)$$

где  $c^1 \in \mathbb{R}^m$  — вектор с компонентами  $c_i$ ,  $i \in I_k$ ,  $c^2 \in \mathbb{R}^{n-m}$  — вектор с компонентами  $c_j$ ,  $j \in I_k$ . С учетом (5) целевая функция может быть выражена только через  $v$  и приобретает вид

$$(c^1, A_1^{-1}(b - A_2 v)) + (c^2, v) = (c^2 - A_2^T (A_1^{-1})^T c^1, v) + (c^1, A_1^{-1} b).$$

Таким образом, исходная задача эквивалентна следующей:

$$\min(c^2 - A_2^T (A_1^{-1})^T c^1, v), \quad A_1^{-1} (b - A_2 v) \geq 0, \quad v \geq 0. \quad (7)$$

При этом точке  $x^k$  соответствует вектор  $\{u^k, v^k\}$ , где  $u^k > 0$ ,  $v^k = 0$ . Точка  $v^k = 0$  является допустимой для задачи (7), при этом ограничение  $A_1^{-1} (b - A_2 v) \geq 0$  удовлетворяется как строгое неравенство, так как  $A_1^{-1} (b - A_2 v^k) = A_1^{-1} b = u^k > 0$ . Поэтому оно может быть отброшено (как неактивное) при анализе точки  $v^k$  на оптимальность. Но в задаче

$$\min(d, v), \quad v \geq 0 \quad (8)$$

минимум достигается в 0 тогда и только тогда, когда  $d \geq 0$ .

Итак, если

$$d = c^2 - A_2^T (A_1^{-1})^T c^1 \geq 0, \quad (9)$$

то точка  $v = 0$  является решением (8) и одновременно (7), а тем самым  $x^k$  является решением (1). Если же среди компонент  $d$  есть отрицательные (например,  $d_j < 0$ ), то  $v = 0$  не является решением (8), и, увеличивая  $v_j$ , можно уменьшить значение левой функции в (8). Итак, сделаем шаг

$$v^{k+1} = v^k + \gamma_k e_j, \quad j: d_j < 0, \quad e_j — j\text{-я орт.} \quad (10)$$

Выбор  $\gamma_k$  производится из следующих соображений. С увеличением  $\gamma_k$  целевая функция уменьшается, однако может нарушиться ограничение  $A_1^{-1} (b - A_2 v) \geq 0$ , которое в точке  $v^k = 0$  было неактивным. Итак,  $\gamma_k$  нужно выбрать из условия

$$\gamma_k = \max \{ \gamma \geq 0: A_1^{-1} (b - \gamma A_2 e_j) \geq 0 \}. \quad (11)$$

Если при этом  $\gamma_k = \infty$ , то задача не имеет решения, так как  $\inf_{x \in Q} (c, x) = -\infty$ . Если же  $\gamma_k < \infty$ , то получаем новую точку  $x^{k+1} = \{u^{k+1}, v^{k+1}\}$ , где  $u^{k+1} = A_1^{-1} (b - \gamma_k A_2 e_j)$ ,  $v^{k+1} = v^k + \gamma_k e_j$ .

Эта точка вновь является вершиной (она допустима и имеет  $m$  положительных компонент, так как  $j$ -я компонента стала положительной, а одна из компонент  $u^{k+1}$  обратилась в 0). Поэтому в ней можно повторить всю процедуру заново. Поскольку  $(c, x^{k+1}) < (c, x^k)$  (так как  $(c, x^{k+1}) = (d, v^{k+1}) + \gamma_k (d, e_j) = (c, x^k) + \gamma_k d_j$ ,  $\gamma_k > 0, d_j < 0$ ), то метод обладает

свойством монотонности целевой функции. Следовательно, возвращение в какую-либо из ранее пройденных вершин невозможно, и в силу конечности общего числа вершин метод конечен.

С геометрической точки зрения симплексный метод реализует последовательный переход от одной вершины допустимого множества к соседней (см. упр. 1) с меньшим значением целевой функции. Ясно, что такой процесс конечен.

Идея исключения переменных, используемая в симплекс-методе, уже встречалась нам в гл. 8 при анализе задач с ограничениями типа равенств. Так, второе доказательство правила множителей Лагранжа (§ 1 гл. 8) основывалось на сведении исходной задачи к задаче безусловной минимизации путем (невинного) выражения одних переменных через другие и использования условий экстремума в задаче без ограничений. Аналогично применялся в методе приведенного градиента для построения метода оптимизации (см. формулу (18) § 2 гл. 8 с формулами симплекс-метода). В симплекс-методе используется, по существу, тот же подход, разница лишь в том, что, кроме равенств  $Ax = b$ , имеются неравенства  $x \geq 0$ , а каждая задача, получающаяся после исключения переменных, является задачей линейного программирования простой структуры (7). Кстати, условие оптимальности (9) в симплекс-методе позволяет

другим способом получить условие экстремума в исходной задаче, правда, при более жестких предположениях (см. упр. 2).

#### Упражнения.

1. Назовем *ребром* (1-мерной гранью) многоугольного множества  $Q$  отрезок  $L = [a, b] \subset Q$  такой, что любая его точка не является внутренней точкой другого отрезка из  $Q$  (т. е. если  $x \in L, y \in Q, z \in Q$  и  $x = (y+z)/2$ , то это возможно лишь тогда, когда  $y \in L, z \in L$ ). Докажите, что концы ребра являются вершинами (такие вершины называются *соседними*) и что для любой вершины существует соседняя (если чисто вершине больше 1). Дайте описание ребер аналогичное описанию вершин в лемме 1 § 1 и в п. 6 леммы 1. Покажите, что точки  $x^*$  и  $x^{*-1}$  в симплекс-методе являются соседними вершинами.

2. Получите из (9) условия экстремума в задаче (1). Покажите, что они совпадают с условиями теоремы 6 § 1. (Указание. Докажите, что если  $x^*$  — решение, то  $(A_1^{-1})^T c^1 = y^*$ .)

**2. Реализация симплекс-метода.** Конечно, приведенное выше описание симплекс-метода еще далеко от полного. Чтобы перейти от него к четкому алгоритму, нужно ответить на ряд вопросов.

Первый связан с выбором начального приближения. Точку  $x^0$ , являющуюся вершиной  $Q$ , можно найти с помощью следующего приема (*метода искусственного базиса*). Введем дополнительные переменные  $z_i$ , играющие роль невязок в ограничениях, и рассмотрим задачу их минимизации:

$$\min \sum_{i=1}^m z_i, \quad (a^t, x) + z_i = b_i, \quad i = 1, \dots, m, \quad x \geq 0, \quad z \geq 0. \quad (12)$$

В этой задаче искомым является вектор  $\{x, z\}$  размерности  $n+m$ , а точка  $\{0, b\} \in \mathbb{R}^{n+m}$  является вершиной (всегда можно считать, изменив в случае необходимости знак ограничения  $(a^t, x) = b_i$ , что  $b \geq 0$ ; предполагем, что среди компонент  $b$  нет нулевых). Поэтому для (12) можно применить симплекс-метод с данным начальным приближением. В результате получим либо точку  $\{x^0, 0\}$ , где  $x^0$  — вершина в исходной задаче, либо точку  $\{z^0, 0\}$ , где  $z^0$  — вершина в исходной задаче, например, исходная задача имеет вид

Введения подобного дополнительного этапа в решение получения линейного программирования нередко можно избежать. Пусть, например, исходная задача имеет вид

$$\min(c, x), \quad Ax \leq b, \quad x \geq 0,$$

причем  $b > 0$ . Приведем ее к канонической форме, как в (8),

§ 1:

$$\min(c, x), \quad Ax + z = b, \quad x \geq 0, \quad z \geq 0. \quad (14)$$

Тогда точка  $\{0, b\}$  является вершиной, так что применять метод искусственного базиса излишне.

Второй вопрос связан с численной реализацией метода. Симплекс-метод имеет ряд вычислительных схем, связанных со способами вычисления, хранения и преобразования физических в нем матриц и векторов. В основном варианте (алгоритм с обратной матрицей) вычисляется и хранится матрица  $A_1^{-1}$ , а выражения типа  $A_1^T (A_1^{-1})^T c^1$  (см. (9)) вычисляются путем умножения этой матрицы на соответствующие векторы. При вычислении же  $A_1^{-1}$  используется тот факт, что матрицы  $A_1$  на последовательных итерациях получаются заменой лишь одного столбца. Поэтому можно обращаться эти матрицы рекуррентно, используя соотношения типа лемм 3 и 4 § 3 гл. 3. Начальное приближение для  $A_1^{-1} b$  в методе искусственного базиса (12) и в задаче (14) не требует обращения матрицы, так как в этом случае  $A_1 = I$ .

Мы не приводим подробных формул симплекс-метода. В настоящее время программы симплекс-метода широко доступны и достигли высокого уровня совершенства (с вычислительной и программистской точки зрения). Поэтому вряд ли многим читателям придется решать задачи линейного программирования вручную или самостоятельно писать программу симплекс-метода (в этом случае можно обратиться к многочисленным руководствам по линейному программированию).

Последний вопрос, о котором стоит упомянуть, — это проблема вырождения в симплекс-методе. При изложении метода делалось предположение о невырожденности всех вершин. Отказ от этого предположения может привести к невозможности обращения матрицы  $A_1$ . Однако многочисленные эксперименты с симплекс-методом показали, что в практических задачах вырождение встречается редко, и его опасность можно пренебречь.

**3. Другие конечные методы.** Мы уже упоминали, что существуют различные вычислительные схемы симплекс-метода и кратко описали одну из них (*алгоритм с обратной матрицей*). В модифицированном симплекс-методе хранится и пересчитываются матрицы не хранятся в готовом виде, а вычисляются из начальной с помощью элементарных преобразований, соответствующих предыдущим итерациям.

Целесообразность выбора той или иной модификации определяется соотношением  $m$  и  $n$ , степенью заполненности матрицы ограничений (в большинстве задач значительной размерности в  $A$  много нулевых элементов), объемом памяти ЭВМ, требованиями к точности решения и т. д. Эти вопросы разработаны весьма подробно. Однако в принципе (если отвлечься от ошибок

округления и т. д.) все модификации симплекс-метода дают одну и ту же последовательность приближений  $x^k$ .

В то же время существуют методы, основанные на идеях симплекс-метода, но не совпадающие с ним. Мы уже отмечали (см. упр. 2), что симплекс-метод одновременно дает решения и для двойственной задачи. Именно, величина  $y^k = (A_1^{-1})^T c^1$  может рассматриваться как приближение к решению двойственной задачи (не являющееся допустимым для нее, причем  $(b, y^k) \geq (b, y^*)$  до последней итерации, где  $y^k = y^*$  (см. упр. 3)). Поэтому возможен следующий путь решения (1). Применим симплекс-метод к двойственной задаче. Тогда одновременно получаются приближения для прямых переменных  $x^k$ , причем они дают оценку снизу для целевой функции:  $(c, x^k) < (c, x^*)$ . Можно организовать вычисление таким образом, что при решении двойственной задачи основным является итерационный процесс именно для прямых переменных. Такой способ решения называется *двойственным симплекс-методом*. Применять его удобно в следующих ситуациях. Во-первых, если прямая задача имеет вид (1) § 1, тогда двойственная записывается в канонической форме, что позволяет применить к ней симплекс-метод без дополнительных преобразований. Во-вторых, часто приходится решать последовательность задач линейного программирования, каждая из которых получается из предыдущей добавлением нового ограничения типа неравенств (см., например, метод отсекающей гиперплоскости и другие методы решения негладких задач из § 4 гл. 5). Тогда решение предыдущей задачи не является допустимой точкой для новой задачи и потому не может быть использовано в симплекс-методе. Однако решение двойственной задачи оказывается вершиной для новой двойственной задачи и представляет собой хорошее начальное приближение для двойственного симплекс-метода (см. упр. 4).

Известны и некоторые другие конечные методы, построенные на основе симплекс-метода. Таков, например, метод *одновременного решения прямой и двойственной задач*. Он удобен тем, что в нем получаются оценки для целевой функции сверху и снизу, что позволяет не доводить итерации до конца, ограничиваясь получением приближения с требуемой точностью. Однако в целом эти методы не получили широкого распространения. Парadoxальным образом это объясняется успехами программной реализации симплекс-метода; в результате создавалось ощущение ненужности каких-либо других методов линейного программирования.

#### Упражнения.

3. Покажите, что для величины  $y^k = (A_1^{-1})^T c^1$  в симплекс-методе спра- ведливо соотношение  $(b, y^k) \geq (b, y^*)$ , если только  $x^k$  не является реше- нием задачи (1).

4. Пусть  $x^*$  — решение задачи тип  $(c, x)$ ,  $Ax \leq b$ ,  $y^*$  — решение двой- ственной задачи. Запишите задачу, двойственную к типу  $(c, x)$ ,  $Ax \leq b$ ,  $(a_{m+1}^m, x) \leq b_{m+1}$ , и покажите, что  $\{y^*, 0\} \in R^{m+1}$  — вершина для этой задачи.

4. **Почему симплекс-метод позволяет найти решение?** Все приведенные выше рассуждения не дают никаких оснований для вывода об эффективности симплекс-метода как средства решений задач линейного программирования сколько-нибудь значительной размерности. Более того, могло бы оказаться, что этот метод просто неработоспособен для подобных задач. Действительно, единственное, что мы знаем о сходимости метода, — это факт его конечности. Каково же требуемое при этом число шагов? Априори мы можем гарантировать лишь то, что это число не превосходит числа вершин допустимого множества (обозначим его  $N$ ). Из леммы 1 § 1 следует, что справедлива оценка  $N \leq C_m^n$  (речь идет об ограничениях вида (13) § 1). Если  $m = 2n$ , то по формуле Стирлинга  $C_m^n \sim 2^{2n} (n/e)^n$  — астрономическое число даже для  $n$  порядка нескольких десятков. Конечно, это лишь оценка сверху для  $N$ , причем сильно завышенная, однако она производит мрачное впечатление. Как показывают простые примеры, число вершин многогранника действительно велико (хотя и существенно меньше, чем  $C_m^n$ ). Пусть, например,  $Q = \{x \in R^n : 0 \leq x \leq a\}$ ,  $a > 0$ , тогда очевидно, что  $N = 2^n$ . Для  $n = 100$  (это задача небольших размеров по масштабам линейного программирования) получаем  $N = 2^{100} \approx 10^{30}$  — число, пре- восходящее возможности любых существующих или потенци- ально возможных вычислительных машин. Эти соображения относятся лишь к числу вершин многогранника, а не непосред- ственно к числу шагов симплекс-метода. Однако построены специальные примеры, в которых траектория симплекс-метода проходит по всем вершинам многогранника условий [10.18]. Для этих примеров число шагов экспоненциально растет с ростом размерности.

Казалось бы, эти рассуждения приводят к пессимистическому выводу о возможностях симплекс-метода. Именно таково было и первоначальное мнение Дж. Даннига, создателя симплекс-ме- тода. Как он пишет в [105], «симплекс-метод интуитивно был отвергнут как неэффективный и лишь благодаря счастливой случайности был проверен и принят». Численная проверка показала поразительную эффективность метода для практических задач вопреки всем ожиданиям. Оказалось, что для подавляющего большинства задач число итераций бывает порядка  $m \div 2m$  (речь идет о задаче в канонической форме (1)), и отнюдь не растет экспоненциально с ростом размерности. Лишь в очень редких случаях (а в настоящее время симплекс-методом решены многие тысячи задач) число итераций было существенно больше

чем  $t$ . Причина такого удивительного «выведения» до сих пор неясна.

#### Упражнение.

5. Выпишите симплекс-метод для задачи  $\min \{c, x\}, 0 \leqslant x_i \leqslant 1, i = 1, \dots, n$ . Покажите, что он всегда заканчивается не более чем за  $n$  шагов.

### § 3. Итерационные методы линейного программирования

**1. Актуальность итерационных методов.** В § 2 говорилось о достоинствах симплекс-метода. Однако его возможности не безграничны. Так, он чувствителен к погрешностям вычислений, и для плохо обусловленных задач их влияние может привести к значительной потере точности (впрочем, этот вопрос недостаточно исследован ни теоретически, ни численно). Далее, в симплекс-методе необходимо хранить элементы обратной матрицы  $A_1^{-1}$  или какие-либо эквивалентные величины. Хотя в развитых программах симплекс-метода приняты специальные меры по экономии записи, при больших размерах решаемых задач требуемый объем памяти велик. Симплекс-метод не позволяет использовать обычно имеющуюся априорную информацию о задаче типа хорошего начального приближения для прямых и (или) двойственных переменных. Наконец, большинство программ симплекс-метода не рассчитано на их использование в качестве подпрограмм, тогда как часто приходится многократно решать однотипные задачи — например, в методах нелинейного программирования и негладкой оптимизации, где на каждой итерации делают актуальной разработку новых методов линейного программирования.

**2. Итерационно-конечные методы.** Противопоставление итерационных и конечных методов в известной мере условно. Так, рассмотрим еще несколько итерационных методов, описанных ранее применительно к нелинейным задачам. Оказывается, для случая линейного программирования они заканчиваются за конечное число шагов.

Начнем с двух методов, представляющих скорее теоретический, а не вычислительный интерес (решаемые в них на каждом шаге вспомогательные задачи не проще исходной).

1°. Метод регуляризации. Вместо исходной задачи

$$\min \{c, x\}, Ax \leqslant b \quad (1)$$
будем решать последовательность регуляризованных задач
$$\min \{[c, x] + e_k \|x\|^2\}, Ax \leqslant b. \quad (2)$$

Решение такой задачи существует (если допустимое множество

непустого) и единствено, обозначим его  $x^k$ . Из теоремы 12 § 1 непосредственно следует конечность метода.

**Теорема 1.** Пусть множество решений (1)  $X^*$  непусто. Тогда существует  $\bar{\varepsilon} > 0$  такое, что при всех  $0 < \varepsilon_k \leqslant \bar{\varepsilon}$  будет  $x^k = x^*$ , где  $x^*$  — нормальное решение исходной задачи (т. е. элемент  $X^*$  с минимальной нормой). ▲

Таким образом, в методе регуляризации для линейного программирования, в отличие от общего случая (см. теорему 4 § 1 гл. 6), нет нужды устремлять параметр регуляризации к 0 — решение перестает меняться при всех достаточно малых значениях  $e_k$ .

2°. Метод проекции градиента. Построим, как в прокс-методе (см. § 1 гл. 6), вспомогательную задачу на  $k$ -й итерации вида

$$\min \{[c, x] + (2y)^{-1} \|x - x^k\|^2\}, Ax \leqslant b \quad (3)$$

и ее решение возьмем в качестве  $x^{k+1}$ . Очевидно, что этот метод может быть записан в форме метода проекции градиента

$$x^{k+1} = P_Q(x^k - yc), Q = \{x: Ax \leqslant b\}. \quad (4)$$

**Теорема 2.** Если решение  $x^*$  задачи (1) существует и единственno, то при любом  $y > 0$  метод (3) конечен, т. е.  $x^k = x^*$  для некоторого  $k$ .

Действительно, в (1) достигается острый минимум (см. лемму 5 § 1), а в соответствии с общей теоремой 1 § 2 гл. 7 метод проекции градиента для острого минимума конечен. ▲

Метод (3), конечно, не эффективен — в нем вместо исходной задачи решается последовательность более трудных вспомогательных задач. Его можно сделать более разумным, если на каждом шаге учитывать лишь  $\varepsilon$ -активные ограничения, т. е. решать задачу

$$\min \{[c, x] + (2y)^{-1} \|x - x^k\|^2\},$$

$$(a_i^t, x) \leqslant b_i, i \in I_k, I_k = \{i: (a_i^t, x^k) \geqslant b_i - \varepsilon\}. \quad (5)$$

Однако более перспективным является переход к двойственным методам решения вспомогательных задач, который и описывается ниже.

3°. Метод штрафных функций. Вместо квадратичной задачи (2), фигурирующей в методе регуляризации, будем решать двойственный ей. В соответствии с (12) и (13) § 4 точка  $x^k$  — решение (2) — может быть найдена по формуле

$$x^k = -(2e_k)^{-1} (c + A^T y^k),$$

где  $y^k$  — решение задачи

$$\min_{y \geqslant 0} \left[ (b, y) + \frac{1}{4e_k} \|c + A^T y\|^2 \right]. \quad (6)$$

Задача (7) сводится к минимизации квадратичной функции на неограниченном ортанте; минимум может быть найден методом сопряженных градиентов (§ 3 гл. 7). С другой стороны, (7) есть не что иное, как метод штрафных функций для решения задачи, двойственной к (1):

$$\min (b, y), \quad A^T y = -c, \quad y \geqslant 0, \quad (8)$$

где в качестве штрафного коэффициента берется  $1/(4\varepsilon_k)$ ,  $\varepsilon_k \rightarrow 0$ .

Таким образом, метод регуляризации эквивалентен методу штрафных функций для двойственной задачи.

Можно, наоборот, использовать метод штрафных функций для исходной задачи

$$\min (c, x), \quad Ax = b, \quad x \geqslant 0, \quad (9)$$

решая последовательность квадратичных задач минимизации на  $\mathbf{R}_+^n$ :

$$\min_{x \geqslant 0} \left[ (c, x) + \frac{1}{2} K_k \|Ax - b\|^2 \right], \quad K_k \rightarrow \infty. \quad (10)$$

Обозначая решение этой задачи через  $x^k$  и используя связь с методом регуляризации и теорему 1, получаем следующий результат.

**Теорема 3.** Если решение задачи (9) существует, то для достаточно больших  $K_k$  метод (10) дает точное решение двойственной задачи:

$$y^* = K_k(Ax^k - b). \quad \blacktriangle \quad (11)$$

Таким образом, метод штрафных функций для задач линейного программирования конечен для двойственной задачи (хотя при этом не обязательно  $x^k$  — решение прямой задачи).

4°. Метод модифицированной функции Лагранжа. Выписывая задачу, двойственную к (3), мы приходим к методу

$$y^k = \arg \min_{y \geqslant 0} \left[ (b, y) - (A^T y, x^k) + \frac{1}{2} y \|A^T y + c\|^2 \right], \quad (12)$$

$$x^{k+1} = x^k - y(A^T y^k + c).$$

Но это не что иное, как метод модифицированной функции Лагранжа для двойственной задачи (8) с  $K = y$ . Обратно, если исходная задача имеет каноническую форму (9), то метод модифицированной функции Лагранжа

$$x^k = \arg \min_{x \geqslant 0} M(x, y^k, K),$$

$$M(x, y, K) = (c, x) + (y, Ax - b) + \frac{1}{2} K \|Ax - b\|^2,$$

$$y^{k+1} = y^k + K(Ax^k - b)$$
(13)

эквивалентен методу проекции градиента для двойственной к (9) задачи с  $y = K$ . С учетом теоремы 2 и упражнения 1 получаем следующее утверждение.

**Теорема 4.** Если решение задачи (9) существует, то при любом  $K \geq 0$  метод (13) за конечное число шагов дает решение прямой и двойственной задач.  $\blacktriangle$

Метод модифицированной функции Лагранжа (13) имеет много привлекательных черт. Поскольку в нем  $K$  не обязательно должно быть слишком большим, то вспомогательные задачи в (13) не являются неизбежно плохо обусловленными. С другой стороны, при увеличении  $K$  число итераций сокращается и в принципе может быть доведено до 1 (см. упр. 4). Решение (13) может быть найдено с помощью хорошо отработанного метода сопряженных градиентов за конечное число шагов. Впрочем, нет нужды решать вспомогательные задачи (13) точно — можно получить результаты о сходимости и для случая, когда минимизация в (13) осуществляется приближенно, с определенным итерационным методом, в (13) приходится иметь дело лишь с исходной матрицей ограничений  $A$ , что позволяет использовать свойство ее слабой заполненности. Благодаря этим достоинствам метод (13) нашел широкое применение в вычислительной практике и в некоторых ситуациях успешно конкурирует с симплекс-методом.

### Упражнения.

1. Докажите теорему 2 без предположения об единственности решения.
2. Пусть исходная задача имеет вид (1) и применяется метод штрафных функций:  $x^* = \arg \min [(c, x) + (K_k/2) \|Ax - b\|^2]$ . Докажите, что для достаточно больших  $K_k$  вектор  $K_k(Ax^* - b)_+$  является решением двойственной задачи (8).

3. Докажите, что если допустимое множество в (1) ограничено, то найдется  $\hat{y} > 0$  такое, что при всех  $x^0 \in Q$  и всех  $y \geq \hat{y}$  метод (3) дает решение за 1 шаг, т. е.  $x^1 = x^*$ .

4. Используя предыдущий результат, покажите, что при достаточно большом  $K$  метод (13) дает решение за 1 шаг.

**3. Сведение к негладкой минимизации.** Описанные выше методы, несмотря на свои достоинства, не могут считаться универсальным средством решения любых задач линейного программирования. Дело прежде всего в том, что на каждом шаге в них требуется решать довольно сложную вспомогательную задачу и при очень большой размерности вычисления становятся трудоемкими или даже нереализуемыми. Поэтому представляют интерес и другие итерационные методы, в которых вычисления на отдельной итерации максимально просты (возможно, за счет более медленной сходимости). Эти методы дают возможность находить (хотя бы и грубо), приближенное решение очень больших задач.

Первый из такого рода подходов связан с использованием простых методов негладкой оптимизации, описанных в гл. 5. Свести же исходную задачу линейного программирования к безусловной минимизации кусочно-линейной функции можно различными способами. Так, можно ввести

$$f(x) = (c, x) + K \max_{1 \leq i \leq m} [(a^i, x) - b_i]_+. \quad (14)$$

Тогда для достаточно больших  $K \geq K$  точки минимума  $f(x)$  на  $\mathbf{R}^n$  совпадают с множеством решений  $X^*$  задачи

$$\min(c, x), \quad (a^i, x) \leq b_i, \quad i = 1, \dots, m. \quad (15)$$

Отметим, что аналогичный метод негладких штрафов уже обсуждался применительно к задачам выпуклого программирования (лемма 2 § 3 гл. 9). Один из субградиентов (14) вычисляется очень просто:

$$\partial f(x) = c + K a^j, \quad j = \operatorname{argmax}_{1 \leq i \leq m} [(a^i, x) - b_i]_+, \quad (16)$$

т. е. в точке  $x$  достаточно найти «наиболее нарушенное» ограничение. Недостатком такого подхода является то, что число  $K$  обычно неизвестно.

Весьма простой способ трансформации задачи связан с использованием двойственной функции. Пусть в исходной задаче имеются двусторонние ограничения:

$$\min(c, x), \quad x \in \mathbf{R}^n, \quad Ax = b, \quad b \in \mathbf{R}^m, \quad a \leq x \leq d. \quad (17)$$

Введем двойственную функцию (см. (47) § 1 гл. 9):

$$\theta(y) = -\min_{a \leq x \leq d} [(c, x) + (y, Ax - b)]. \quad (18)$$

Для всякого  $y$  решение  $x(y)$  задачи минимизации по  $x$  отыскивается чрезвычайно просто:

$$x(y)_i = \begin{cases} a_i, & \text{если } (c + A^T y)_i > 0, \\ d_i, & \text{если } (c + A^T y)_i < 0. \end{cases} \quad (19)$$

Функция  $\theta(y)$  выпукла, кусочно-линейна, всюду определена, и в силу теоремы двойственности исходная задача эквивалента безусловной минимизации  $\theta(y)$ . Субградиент  $\theta(y)$  имеет вид

$$\partial \theta(y^k) = -(Ax^k - b), \quad x^k = x(y^k). \quad (20)$$

Такой подход особенно выгоден, если размерность двойственных переменных мала:  $m \ll n$ .

Наконец, записав для задачи (15) условия экстремума (см. теорему 3 § 1) в виде равенств и неравенств:

$$Ax \leq b, \quad (c, x) = (b, y), \quad y \geq 0, \quad A^T y - c = 0, \quad (21)$$

можно свести задачу линейного программирования к минимизации невязок в системе (21):

$$\Phi(x, y) = \|Ax - b\|_+ + |(c, x) - (b, y)| + \|y\|_+ + \|A^T y - c\|. \quad (22)$$

Этот способ не экономен, так как в нем размерность задачи равна  $n + m$ . Единственное его преимущество заключается в том, что минимальное значение  $\Phi$  известно:  $\Phi^* = 0$ .

После того как задача сведена к безусловной минимизации тем или иным способом (для определенности будем говорить о (14)), можно применить субградиентный метод:

$$x^{k+1} = x^k - \gamma_k \partial f(x^k). \quad (23)$$

При обычных условиях на  $y_k$  результаты § 3 гл. 5 гарантируют сходимость. Если решение  $x^*$  единственно, то  $f(x)$  имеет острый минимум и тем самым справедливы оценки скорости сходимости, полученные в § 3 гл. 5 для острого минимума.

Вместо субградиентного метода (23) можно использовать более молниевые методы ускорения сходимости из § 4 гл. 5, например, метод эллипсоидов (12) § 4 гл. 5 или метод с растяжением пространства (14) § 4 гл. 5. Возможности таких способов ограничены, так как в них требуется запоминать матрицу  $H_k$  высокой размерности, а скорость сходимости не слишком велика (см. оценки в теореме 4 § 4 гл. 5). Однако на этом пути Хачияном [10.15] получен интересный теоретический результат. Оказывается, что можно найти точное решение за конечное число арифметических действий, которое полиномиально зависит от размерности задачи (мы отсылаем к [10.15] за точной формулировкой, см. также упр. 5 и 6). Основную роль в доказательстве этого факта играет оценка из теоремы 4 § 4 гл. 5 — для достижения некоторой фиксированной точности по функции нужно сделать порядка  $O(n^2)$  итераций метода (12) § 4 гл. 5. С другой стороны, для симплекс-метода и других известных конечных методов получаются лишь экспоненциальные оценки числа действий (мы уже указывали в § 2, что существуют примеры, в которых симплекс-метод требует числа итераций порядка  $\exp(O(n))$ ). Не стоит делать из этих оценок вывод о том, что итерационный метод, основанный на сведении задачи линейного программирования к безусловной минимизации и применения метода эллипсоидов, превосходит симплекс-метод для больших размерностей. Во-первых, для конечного  $n$  нельзя сказать, что больше:  $O(n^c)$  или  $\exp(O(n))$ , — все зависит от входящих в эти формулы неизвестных констант и величины  $n$ . Во-вторых (и это главное), оценки симплекс-метода рассчитаны на «наихудшие» задачи. В подавляющем же большинстве реальных задач, как уже отмечалось, для получения решения в действительности требуется порядка  $m \div 2m$  итераций. В то же время оценка скорости сходимости метода эллипсоидов реализуется для

всех задач (см. § 4 гл. 5). Наконец, кроме числа действий, работоспособность методов зависит от многих факторов (устойчивость к погрешностям вычислений, требуемый объем памяти и т. д.).

#### Упражнения.

5. Пусть задана система неравенств  $\sum_{l=1}^n a_{il}x_l \leq b_i, i = 1, \dots, m$ , где  $a_{il}$ ,  $b_i$  — целые числа;  $|a_{il}| \leq h$ ,  $|b_i| \leq h$ , и пусть  $L = n \log h \sqrt{n} + \log_2(n+1)$ . Покажите, что если система имеет решение, то найдется решение  $x^*$  такое, что  $\|x^*\| \leq 2^L$ , а если система не имеет решения, то  $\min f(x) \geq 2^{-L}$ , где  $f(x) = \max_{1 \leq i \leq m} \left( \sum_{l=1}^n a_{il}x_l - b_i \right)$ .

6. Используя предыдущий результат, покажите, что метод (12) § 4 гл. 5, примененный для минимизации  $f(x)$ , за число итераций  $O(n^2 L)$  дает возможность выяснить, имеет ли систему неравенств решение или нет.

**4. Использование функций Лагранжа.** В § 3 гл. 9 описывались двойственные методы выпуклого программирования, являющиеся методами отыскания седловой точки функции Лагранжа. Простейший из них — метод Эрроу — Гурвица — Улзавы — для задач линейного программирования не сходится (см. пример (14) в § 3 гл. 9). Однако если функцию Лагранжа заменить на модифицированную, то соответствующий метод (см. (20) § 3 гл. 9) является сходящимся. Именно, рассмотрим для задачи (1) метод

$$\begin{aligned} x^{k+1} &= x^k - \gamma M'_x(x^k, y^k, K) = x^k - \gamma(c + \\ &\quad + A^T[y^k + K(Ax^k - b)]_+), \end{aligned} \quad (24)$$

$$y^{k+1} = y^k + \gamma M'_y(x^k, y^k, K) = y^k + \gamma \max\{-y^k/K, Ax^k - b\}$$

и соответственно для задачи (9) метод

$$\begin{aligned} x^{k+1} &= [x^k - \gamma M'_x(x^k, y^k, K)]_+ = \\ &= [x^k - \gamma(c + A^T y^k + K A^T(Ax^k - b))]_+, \quad (25) \\ y^{k+1} &= y^k + \gamma M'_y(x^k, y^k, K) = y^k + \gamma(Ax^k - b). \end{aligned}$$

Теорема 5. Если решение задач (1) или (9) существует, то для всякого  $K > 0$  найдется такое  $\tilde{\gamma} > 0$ , что при  $0 < \gamma < \tilde{\gamma}$  методы (24) и (25) сходятся к решению соответствующих прямой и двойственной задач со скоростью геометрическойvergence. Следует обратить внимание на то, что знаменатель прогрессии нельзя сделать малым за счет выбора  $K$ , так что скорость сходимости может быть невелика.

В методе Эрроу — Гурвица — Улзавы делается градиентный шаг минимизации по  $x$  и максимизации по  $y$  функции Лагранжа.

Вместо этого можно поочередно находить минимум по  $x$  и максимум по  $y$  этой функции и сдвигаться в полученном направлении. Именно, пусть решается задача

$$Ax \leq b, \quad b \in \mathbf{R}^m, \quad x \geq 0, \quad c > 0, \quad b > 0,$$

$$Ax \leq b, \quad x \in \mathbf{R}^n,$$

и пусть  $x^0$  некоторая допустимая точка,  $(c, x^0) = \alpha$ . Тогда решение задачи (26) не изменится, если добавить ограничение  $(c, x) \leq \alpha$ ; аналогично в силу теоремы двойственности всякое решение двойственной задачи должно удовлетворять условию  $(b, y) \leq \alpha$ . Поэтому (26) можно свести к такой задаче о седловой точке:

$$\begin{aligned} \min_{x \in Q} \max_{y \in S} L(x, y) &= \max_{y \in S} \min_{x \in Q} L(x, y), \\ L(x, y) &= (c, x) + (y, Ax - b), \\ Q &= \{x \geq 0, (c, x) \leq \alpha\}, \quad S = \{y \geq 0, (b, y) \leq \alpha\}. \end{aligned} \quad (27)$$

Будем отыскивать седловую точку рекуррентно, минимизируя  $L(x, y^k)$  по  $x \in Q$  и максимизируя  $L(x^k, y)$  по  $y \in S$ :

$$x^{k+1} = x^k + \gamma_k(\bar{x}^k - x^k), \quad \bar{x}^k = \operatorname{Argmin}_{x \in Q} L(x, y^k),$$

$$y^{k+1} = y^k + \gamma_k(\bar{y}^k - y^k), \quad \bar{y}^k = \operatorname{Argmax}_{y \in S} L(x^k, y). \quad (28)$$

Если  $b > 0$ ,  $c > 0$ ,  $\alpha > 0$ , то точки  $\bar{x}^k$ ,  $\bar{y}^k$  находятся элементарно:

$$\bar{x}^k = \begin{cases} \frac{\alpha}{(c + A^T y^k)_j} e_j, & j = \operatorname{argmin}_{1 \leq i \leq m} \frac{(c + A^T y^k)_i}{\alpha} \text{ в противном случае,} \\ 0, & \text{если } c + A^T y^k \geq 0, \end{cases} \quad (29)$$

$$\bar{y}^k = \begin{cases} \frac{\alpha}{(A x^k - b)_j} e_j, & j = \operatorname{argmax}_{1 \leq i \leq m} \frac{(A x^k - b)_i}{\alpha} \text{ в противном случае.} \\ 0, & \text{если } A x^k - b \leq 0, \end{cases}$$

Теорема 6. Если в задаче (26)  $c > 0$ ,  $b > 0$ ,  $\alpha > 0$ , то методе (28), (29) при  $\gamma$  — это так называемый метод фиктивной игры или метод Брауна — Робинсон. Его гравитательной чертой является все предельные точки являются решением прямой и двойственной задач. Метод (28), (29) первоначально был предложен в теории игр

все предельные точки являются решением прямой и двойственной задач. Метод (28), (29) первоначально был предложен в теории игр — это так называемый метод фиктивной игры или метод Брауна — Робинсон. Его гравитательной чертой является

чрезвычайная простота. К сожалению, и теоретические оценки и практика вычислений свидетельствуют о его медленной сходимости. Многочисленные эвристические попытки ускорения сходимости пока не дали однозначных результатов.

**4. Заключительные замечания.** До недавнего времени считалось, что симплекс-метод является универсальным эффективным аппаратом для решения любых задач линейного программирования, и нужна в каких-либо иных методах (в частности, итеративных) практически не возникала. Однако в настоящее время такая точка зрения не является общепринятой. Можно надеяться, что вскоре будут созданы вычислительные схемы итеративных методов, конкурентоспособные с симплексным алгоритмом.

#### § 4. Квадратичное программирование

Задачу минимизации квадратичной формы при линейных ограничениях

$$(1)$$

$$\min \{ (Cx, x)/2 - (d, x) \}, \quad x \in Q, \quad Q = \{x: Ax \leqslant b\}$$

будем называть задачей квадратичного программирования.

Здесь  $x \in \mathbf{R}^n$ ,  $C$  — симметричная матрица  $n \times n$ ,  $b \in \mathbf{R}^m$ ,  $A$  — матрица  $m \times n$ . Мы знаем, что квадратичная форма  $(Cx, x)/2 - (d, x)$  является выпуклой в том и только в том случае, когда  $C \geqslant 0$ . Отказ от этого условия приводит, вообще говоря, к многоэкстремальности задачи. В дальнейшем будем всегда предполагать, что

$$(2)$$

Отметим еще, что линейные ограничения в (1) можно преобразовать теми же способами, что и для задач линейного программирования, так что мы будем пользоваться любой удобной формой их записи.

**1. Условия экстремума.** Конкретизируя общие условия экстремума для выпуклых задач (теорема 2 § 1 гл. 9), немедленно получаем следующие результаты.

**Теорема 1.** *Необходимым и достаточным условием экстремума для задачи (1), (2) в точке  $x^*$ ,  $Ax^* \leqslant b$ , является существование  $y^* \in \mathbf{R}^m$  таких, что*

$$(3)$$

$$Cx^* - d + A^T y^* = 0, \quad y^* \geqslant 0, \quad y^*(Ax^* - b) = 0. \quad \blacktriangle$$

Чтобы выписать двойственную к (1) задачу, составим функцию Лагранжа

$$(4)$$

$$L(x, y) = (Cx, x)/2 - (d, x) + (y, Ax - b)$$

и двойственную функцию

$$(5)$$

$$\Psi(y) = \inf_x L(x, y).$$

Выясним вид  $D(\Psi)$ , т. е. области, где  $\Psi(y) \neq -\infty$ . Если квадратичная функция ограничена снизу на  $\mathbf{R}^m$ , то она достигает минимума (см. упр. 2 § 3 гл. 1). Поэтому для всякого  $y \in D(\Psi)$  найдется  $x = x(y)$  такой, что  $L'_x(x, y) = 0$ , т. е.

$$Cx - d + A^T y = 0.$$

При этом

$$\begin{aligned} \Psi(y) &= L(x(y), y) = (Cx, x)/2 + (A^T y - d, x) - (b, y) = \\ &= -(Cx, x)/2 - (b, y). \end{aligned}$$

Поскольку двойственная задача имеет вид (49) § 1 гл. 8:

$$\min \theta(y), \quad \theta(y) = -\Psi(y), \quad y \geqslant 0, \quad y \in D(\theta),$$

то ее решение тем самым эквивалентно решению задачи

$$(6)$$

$$\min \{ (Cx, x)/2 + (b, y) \}, \quad Cx + A^T y = d.$$

В двух важных частных случаях из (6) можно исключить  $x$ : при  $C = 0$  (задача линейного программирования) и при  $C \geqslant 0$ :

$$x = C^{-1}(d - A^T y), \quad (7)$$

при этом  $D(\Psi) = \mathbf{R}^m$ , и мы получаем двойственную задачу

$$\min_{y \geqslant 0} \Gamma/2(C^{-1}(d - A^T y), (d - A^T y)) + (b, y). \quad (8)$$

Итак, справедливо следующее утверждение о паре двойственных задач.

**Теорема 2** (теорема двойственности). *При  $C > 0$  задача (8) двойственная задаче (1): их решения  $x^*, y^*$  существуют или не существуют одновременно, они связаны соотношением*

$$x^* = C^{-1}(d - A^T y^*), \quad (9)$$

*и выполняются неравенства*

$$(Cx, x)/2 - (d, x) \geqslant (C^{-1}(d - A^T y), d - A^T y)/2 + (b, y) \quad (10)$$

*для всех  $Ax \leqslant b$ ,  $y \geqslant 0$ , причем равенство возможно лишь при  $x = x^*, y = y^*$ .*  $\blacktriangle$

Таким образом, при  $C > 0$  задача квадратичного программирования сводится к минимизации квадратичной функции на  $\mathbf{R}_+^m$ . Мы неоднократно пользовались этим сведением ранее, при изучении различных итеративных методов, на каждом шаге которых решалась задача квадратичного программирования. Особенно просто выглядит двойственная задача, если  $C = y^{-1}I$ , т. е. если исходная задача имеет вид

$$\min \{ (d, x) + (2y)^{-1} \|x - a\|^2 \}, \quad Ax \leqslant b.$$

$$(11)$$

Построение двойственной к ней не требует обращения матриц:

$$(12)$$

$$\min \{[(y/2) \|d + A^T y\|^2 - (y, Aa - b)], y \geq 0,$$

а из теоремы 2 получаем, что решения  $x^*, y^*$  задач (11), (12) связаны соотношением

$$x^* = a - y(A^T y^* + d).$$

#### Упражнения.

1. Проверьте, что любая вершина многогранника  $Q = \{x \in \mathbb{R}^n: |x_i| \leq 1, i = 1, \dots, n\}$  является точкой локального минимума функции  $f(x) = -\|x\|^2$ .
2. Убедитесь, что для задач: а)  $\min \{[(Cx, x)/2 - (d, x)], Ax = b; x \geq 0\}$ ; б)  $\min \{[(Cx, x)/2 - (d, x)], Ax = b; x \geq 0\}$  условия экстремума имеют соответственно вид: а)  $Cx^* - d \geq 0, Cx^* - d, x^* = 0$ ; б)  $Cx^* - d + A^T y^* = 0$  (ср. с § 1 гл. 7); в)  $x^* - a + A^T y^* = 0$ ,  $Cx^* - d + A^T y^* \geq 0, (Cx^* - d + A^T y^*, x^*) = 0$ .

**2. Существование, единственность и устойчивость решения.** Случай положительно определенной матрицы  $C$  исследуется особенно просто.

**Теорема 3.** Пусть  $Q$  непусто,  $a, C > 0$ . Тогда решение  $x^*$  задачи (1) существует, единственно и выполняется оценка

$$f(x) - f(x^*) \geq a \|x - x^*\|^2, \quad a > 0 \quad (14)$$

для всех  $x \in Q$ .

Действительно, в этом случае минимизируемая функция сильно выпукла, и можно применить результат упражнения 6 § 1 гл. 7.  $\blacktriangle$

Случай  $C = 0$  уже анализировался ранее — задача квадратичного программирования переходит в задачу линейного программирования. Ситуация при  $C \geq 0$  описывается близкими результатами.

**Теорема 4.** Пусть  $C \geq 0$ ,  $Q$  непусто,  $(Cx, x)/2 - (d, x)$  ограничена сверху на  $Q$ . Тогда решение задачи (1) существует. Этот результат доказывается по той же схеме, что и теорема 9 § 1.  $\blacktriangle$

#### Следствие. Задача

$$\min \|Cx - d\|^2, \quad x \in Q = \{x: Ax \leq b\} \quad (15)$$

имеет решение, если  $Q$  непусто.  $\blacktriangle$

Основные результаты об устойчивости решения следуют из общих утверждений, относящихся к задаче выпуклого программирования (см. § 1 гл. 9). Приведем лишь один специфический для квадратичного программирования факт.

**Теорема 5.** Пусть  $C \geq 0$  и множество решений  $X^*$  задачи (1) непусто,  $f^* = \{x^*\}, x^* \in X^*$ . Тогда найдутся  $\alpha > 0, \beta > 0$  такие, что для любого  $x \in Q$  либо  $f(x) - f^* \geq \alpha^2 (x, X^*),$  либо  $f(x) - f^* \geq \beta \rho(x, X^*),$  либо  $f(x) - f^* \geq \beta \rho^2(x, X^*).$   $\blacktriangle$  (16)

**3. Конечные методы.** Принципиальная возможность решения задачи квадратичного программирования за конечное число арифметических действий связана с тем, что минимизация квадратичной функции на линейном многообразии сводится к решению системы линейных уравнений, а число граней многогранника конечно. Поэтому, последовательно решая задачи на различных гранях многогранника ограничений, можно найти решение задачи.

Один из возможных путей реализации такого подхода заключается в следующем. Пусть решается задача (1) при  $C > 0$  и на  $k$ -м шаге найдена допустимая точка  $x^k$  и выделено множество ограничений  $I_k$ , причем все  $i \in I_k$  являются активными. Найдем решение задачи

$$\min \{[(Cx, x)/2 - (d, x)], (a^i, x) = b_i, i \in I_k\}. \quad (17)$$

Эта задача имеет решение в силу теоремы 3 и условия  $(a^i, x^k) = b_i, i \in I_k$ ; оно может быть найдено из системы линейных уравнений (см. упр. 26))

$$Cx - d + \sum_{i \in I_k} y_i a^i = 0, \quad (a^i, x) = b_i, \quad i \in I_k. \quad (18)$$

Если окажется, что  $\bar{x}^k, y^k$  (решения этой задачи) таковы, что  $\bar{x}^k$  удовлетворяет всем ограничениям  $(a^i, x) \leq b_i, i = 1, \dots, m$ , а  $y_i^k \geq 0, i \in I_k$ , то  $\bar{x}^k$  — решение, см. теорему 1. Если точка  $\bar{x}^k$  не удовлетворяет ограничениям, то возьмем

$$x^{k+1} = x^k + \lambda_k (\bar{x}^k - x^k), \quad (19)$$

где  $0 < \lambda_k < 1$  выбирается из условия  $\lambda_k = \max \{\lambda: x^k + \lambda (\bar{x}^k - x^k) \in Q\}$ , включим новые активные ограничения в  $I_{k+1}$  и повторим вычисления. Наконец, если точка  $\bar{x}^k$  допустимая, но среди  $y_i^k$  есть отрицательные, то возьмем  $x^{k+1} = \bar{x}^k, I_{k+1} = \{i \in I_k, y_i^k > 0\}$ . Пусть описанный выше процесс начинается с точки  $x^0 \in Q, I_0 = \{i: (a^i, x^0) = b_i\}$ . Тогда за конечное число итераций будет найдено решение.

Можно использовать идеи исключения переменных подобно тому, как это делается в симплекс-методе. При этом целесообразно задавать задачу в канонической форме

$$\min \{[(Cx, x)/2 - (d, x)], Ax = b, x \geq 0\}. \quad (20)$$

Здесь достаточно требовать выполнения условия  $C \geq 0$  (а не  $C > 0$ ). Наконец, путем последовательного введения искусственных переменных можно сделать вычислительную схему совершенно аналогичной схемам симплекс-метода. Мы не будем на этом подробнее останавливаться, так как с идеальной точки зрения эти вопросы мало интересны.

**4. Итеративные методы.** Двойственная задача (8) сводится к минимизации квадратичной функции на  $\mathbf{R}^m_+$ , что может быть сделано методом сопряженных градиентов (см. § 3 гл. 7). Таким образом, получаем итеративный метод, приводящий к решению за конечное число итераций. Впрочем, этот подход требует для вычислений градиента либо построения  $C^{-1}$ , либо решения на каждом шаге системы линейных уравнений с матрицей  $C$ . Это целесообразно лишь тогда, когда  $C$  достаточно проста (например, единичная).

Описанные в § 3 итерационно-конечные методы линейного программирования перестают быть конечными для квадратичных задач. Сходимость и оценки ее скорости для этих методов можно получить с помощью общих результатов, относящихся к гладким выпуклым задачам (см. § 3 гл. 9).

В целом для задач квадратичного программирования общие итеративные методы выпуклого программирования не обладают какой-либо повышенной эффективностью. С другой стороны, специфические итеративные алгоритмы для таких задач также мало разработаны. В частности, не очень ясно, как наиболее целесообразно использовать идеи метода сопряженных градиентов для общей задачи квадратичного программирования. Поэтому проблема решения больших квадратичных задач не может считаться полностью решенной.

### ЧАСТЬ III ПРИКЛАДНОЙ АСПЕКТ

#### Г л а в а 11

##### ПРИМЕРЫ ЗАДАЧ ОПТИМИЗАЦИИ

Первые две части книги носили несколько абстрактный характер. Каждая глава начиналась фразой типа «рассмотрим задачу минимизации функции  $f(x)$  на  $\mathbf{R}^n$  в предположении, что  $f(x)$  выпукла и дифференцируема...». Однако, где возникает такого рода задачи? Насколько они распространены? Каков конкретный вид минимизируемой функции? Приводимые ниже примеры — попытка дать ответ на подобные вопросы. После формулировки задачи будет обсуждаться возможность ее анализа и решения методами, описанными выше. При этом нас особенно будет интересовать проблема сочетания общих и специфических методов решения. Оказывается, что для многих конкретных задач можно предложить специальные методы, использующие особенности данной задачи, и более эффективные, чем стандартные. Некоторые из таких приемов будут описаны ниже; разумеется, они отнюдь не исчерпывают всех возможностей. Как увидит читатель, решение всякой конкретной задачи не является рутинным процессом и требует изобретательности и опыта.

Отметим еще, что в каждой из областей науки или практики, где возникают задачи оптимизации, сложилась своя терминология и система обозначений, иногда резко отличающиеся от примененных в теории экстремальных задач. Так, в работах по идентификации принято обозначать неизвестные параметры через  $a, b, c, \dots$ , результаты измерений — через  $x, y, \dots$ , так что типичная задача записывается в виде  $\min J(c), J(c) = \sum_{i=1}^m F(x_i - c)$ .

Мы не будем использовать подобные «локальные» обозначения, а будем придерживаться традиций, принятых в задачах оптимизации.

#### § 1. Задачи идентификации

Пожалуй, наиболее часто необходимость оптимизации возникает в связи с проблемами идентификации, т. е. при построении модели по имеющимся данным об объекте. Задачи идентификации отличаются большим разнообразием в зависимости от вида модели (статическая, динамическая, линейная, нелинейная) и предположений о данных (прямые, косвенные, статистические, детерминированные).

**1. Статистические задачи оценки параметров.** Пусть имеется последовательность измерений  $z^1, \dots, z^m, z^i \in \mathbf{R}^s$ , которая рассматривается как независимая выборка с плотностью распределения  $p(z, x^*)$ . При этом вид плотности  $p(z, x)$  известен, а неизвестное значение параметра  $x^* \in \mathbf{R}^n$  требуется определить.

Например, пусть заданы  $m$  измерений скалярной величины  $x^*$ , содержащие случайную ошибку:

$$z^i = x^* + \xi_i, \quad i = 1, \dots, m, \quad (1)$$

где  $\xi_i$  — независимые случайные величины, нормально распределенные со средним 0 и дисперсией 1. Тогда

$$p(z, x) = (2\pi)^{-1/2} \exp(-(z - x)^2/2), \quad (2)$$

и задача заключается в отыскании оценки для  $x^*$  по  $z^1, \dots, z^m$  и плотности (2). С аналогичной задачей мы уже сталкивались в § 5 гл. 4.

Наиболее распространенным способом оценивания в такой ситуации является *метод максимального правдоподобия*, в котором рекомендуется выбрать оценку для  $x^*$ , максимизирующую вероятность реализации  $z^1, \dots, z^m$ . Поскольку последняя пропорциональна  $p(z^1, x)p(z^2, x)\dots p(z^m, x)$  в силу независимости измерений, то метод сводится к максимизации  $p(z^1, x)\dots p(z^m, x)$  по  $x$ . Если взять логарифм этой функции (что не меняет решения) и изменить знак, то мы приходим к задаче безусловной минимизации:

$$\min f_m(x), \quad f_m(x) = -\sum_{i=1}^m \ln p(z^i, x). \quad (3)$$

Вектор

$$x_m^* = \underset{x \in \mathbf{R}^n}{\operatorname{argmin}} f_m(x) \quad (4)$$

будем называть *оценкой максимального правдоподобия*. В част-

ности, для задачи (1), (2)  $f_m(x) = \frac{1}{2} \sum_{i=1}^m (z^i - x)^2 + \frac{1}{2} \ln 2\pi$  и минимум  $f_m(x)$  отыскивается в явном виде

$$x_m^* = \frac{1}{m} \sum_{i=1}^m z^i, \quad (5)$$

т. е. оценкой максимального правдоподобия в данном случае является среднее арифметическое измерений.

Метод максимального правдоподобия обладает важным свойством оптимальности, которое и оправдывает его широкое распространение. Оказывается, при некоторых предположениях

регулярности относительно  $p(z, x)$  величина  $\sqrt{m}(x_m^* - x^*)$  в пределе при  $m \rightarrow \infty$  распределена нормально со средним 0 и матрицей ковариаций  $S$ :

$$S = J^{-1}, \quad J = \int \frac{\nabla_x p(z, x^*) \nabla_x^T p(z, x^*)}{p(z, x^*)} dz. \quad (6)$$

Здесь  $J$  — *фишеровская информационная матрица* (ср. с (23) § 5 гл. 4). Но величина  $J^{-1}$  в силу неравенства Крамера — Рао (см. (24) § 5 гл. 4) — это нижняя граница для матрицы ковариаций величин  $\sqrt{m}(\hat{x}_m - x^*)$ , где  $\hat{x}_m$  — произвольная несмещенная оценка. Таким образом, оценка максимального правдоподобия имеет асимптотически наилучшую характеристику точности. На языке статистики: оценка  $x_m^*$  является *асимптотически нормальной, состоятельной и асимптотически эффективной*.

Итак, для построения оценки максимального правдоподобия нужно найти безусловный минимум  $f_m(x)$  вида (3). Конечно, явно отыскать  $x_m^*$  удается лишь в редких случаях (подобных задаче (1), (2)). Обычно для минимизации  $f_m(x)$  приходится прибегать к численным методам. Отметим особенности  $f_m(x)$ , позволяющие выбрать определенный метод минимизации. Во-первых, размерность задачи обычно невелика (число искомых параметров редко превосходит 10). Во-вторых, оптимальность оценки максимального правдоподобия доказывается при предложении регулярности, которые включают гладкость  $p(z, x)$  по  $x$ . Поэтому для тех задач, где метод максимального правдоподобия обоснован, функция  $f_m(x)$  является гладкой. В-третьих,  $f_m(x)$  имеет специфический вид — она является суммой одинаковых (отличающихся лишь значениями  $z^i$ ) функций. Поэтому обычно нетрудно выписать ее производные, значит, вычисление  $\nabla f_m(x)$  и  $\nabla^2 f_m(x)$  не представляет труда. Наконец, для большинства задач (хотя и не для всех)  $f_m(x)$  является выпуклой, и проблема многоэкстремальности не возникает. В то же время, как показывает практика,  $f_m(x)$  часто бывает плохо обусловленной.

На основании сказанного и анализа методов в гл. 1 и 3 можно сделать вывод, что для минимизации  $f_m(x)$  вида (3) целесообразно применять метод Ньютона в сочетании с приемами, гарантирующими его глобальную сходимость (§ 1 гл. 3).

1. Убедитесь, что метод максимального правдоподобия для оценки среднего и дисперсии нормального распределения приводит к минимизации в  $\mathbf{R}^2$  функции  $f_m(x) = m \ln x_2 + \sum_{i=1}^m \frac{(z^i - x_1)^2}{2x_2^2}$ . Исследуйте эту функцию (выпуклость, существование и единственность минимума, область определения); выпишите ее производные.

**2.** Запишите метод максимального правдоподобия для задачи отыскания минимума квадратичной функции по измерениям ее градиентов (см. п. 2 § 5 гл. 4).

**2. Задачи регрессии.** В рассмотренной выше задаче все измерения имели одинаковое распределение. Более общей является ситуация, когда измеряемая величина меняется в зависимости от значений некоторых входных переменных, предполагаемых известными. Простейшей здесь является *регрессионная задача*. Пусть

$$y_i = \Phi(u^i, x^*) + \xi_i, \quad i = 1, \dots, m,$$

где  $y_i$  — результаты измерений,  $\Phi: \mathbf{R}^s \times \mathbf{R}^n \rightarrow \mathbf{R}^1$  — известная функция,  $u^i \in \mathbf{R}^s$  — входные величины,  $x^* \in \mathbf{R}^n$  — искомые параметры,  $\xi_i \in \mathbf{R}^1$  — случайные ошибки. Требуется, зная  $y_i$ ,  $u^i$ ,  $i = 1, \dots, m$ , оценить  $x^*$ . Такие задачи часто возникают в эконометрике (восстановление параметров экономических моделей по статистическим данным за прошлые годы), при моделировании химических реакторов (вычисление параметров кинетических уравнений по экспериментальным данным), в кристаллографии (оценка параметров кристаллической решетки по данным рентгеноструктурного анализа) и т. д.

Предположим, что  $\xi_i$  случайны, независимы и одинаково распределены с известной плотностью  $p(z)$ . Тогда оценка максимального правдоподобия имеет вид

$$x_m^* = \operatorname{argmin}_{x \in \mathbf{R}^n} f_m(x), \quad f_m(x) = -\sum_{i=1}^m \ln p(y_i - \Phi(u^i, x)). \quad (8)$$

Можно показать, что при некоторых предположениях оценка (8) по-прежнему является асимптотически нормальной, состоятельной и асимптотически эффективной.

Наряду с (8) для решения регрессионных задач применяют и другие методы, в частности *метод наименьших квадратов*, в котором

$$x_m^* = \operatorname{argmin}_{x \in \mathbf{R}^n} f_m(x), \quad f_m(x) = \sum_{i=1}^m (y_i - \Phi(u^i, x))^2. \quad (9)$$

Этот метод, во-первых, не требует знания закона распределения помехи, а во-вторых, более прост. Для нормально распределенных помех метод максимального правдоподобия совпадает с методом наименьших квадратов. Для других законов распределения оценка наименьших квадратов не является, вообще говоря, эффективной или асимптотически эффективной. Она, однако, обладает одним важным свойством. Пусть модель линейна:

$$y_i = (u^i, x^*) + \xi_i, \quad i = 1, \dots, m, \quad x^* \in \mathbf{R}^n, \quad u^i \in \mathbf{R}^n. \quad (10)$$

Тогда оценка наименьших квадратов линейно зависит от измерений:

$$x_m^* = U^+ y, \quad (11)$$

где  $y = \{y_1, \dots, y_m\} \in \mathbf{R}^m$ ,  $U$  — матрица  $m \times n$  со строками  $u^i$ , а  $U^+$  — псевдообратная к ней (§ I гл. 6). Известная в статистике теорема Гаусса — Маркова утверждает, что (11) обладает наименьшей среди всех линейных несмещанных оценок матрицы ковариаций независимо от закона распределения помехи.

Итак, наиболее распространенные методы оценивания для

задач регрессии вновь приводят к задаче безусловной минимизации (8) или (9). Функции  $f_m(x)$  при этом обычно являются гладкими, однако задачи (8) и (9), как правило, невыпуклы (для нелинейных по  $x$  функций  $\Phi(u, x)$ ) и могут иметь достаточно большую размерность (порядка нескольких десятков).

Функции (8) и (9) являются составными — они имеют вид

$$f(x) = F(z(x)), \quad (12)$$

где  $z(x): \mathbf{R}^n \rightarrow \mathbf{R}^m$ ,  $F(z): \mathbf{R}^m \rightarrow \mathbf{R}^1$ . Действительно, если взять  $z(x)_i = y_i - \Phi(u^i, x)$  и  $F(z) = -\sum_{i=1}^m \ln p(z_i)$ , то получим (8), а

для  $F(z) = \sum_{i=1}^m z_i^2$  получим (9). Для минимизации функций вида

(12) удобно применять методы первого порядка (такие, как метод сопряженных градиентов и квазиньютоновские методы, см. гл. 3), поскольку градиент  $f(x)$  нетрудно выписать:  $\nabla f(x) = z'(x)^T F(z(x))$ . Однако вычисление  $\nabla^2 f(x)$  весьма громоздко, так как оно требует знания вторых производных  $z(x)$  (т. е. знания  $m$  матрицы размерности  $n \times n$ ). Единственное исключение — случай линейной модели (10) и неквадратичной  $F(z)$ ; тогда

$$f(x) = \sum_{i=1}^m F_i(y_i - (u^i, x)), \quad \nabla f(x) = -\sum_{i=1}^m u^i \nabla F_i(y_i - (u^i, x)), \quad (13)$$

$$\nabla^2 f(x) = \sum_{i=1}^m u^i (u^i)^T \nabla^2 F_i(y_i - (u^i, x)),$$

и можно применить метод Ньютона. Для нелинейных моделей существует специальный метод, использующий вид функции (12). В нем строится итерационная последовательность  $x^k$ , причем в каждой точке  $z(x)$  линеаризуется, а  $F(x)$  аппроксимируется квадратично:

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbf{R}^n} f_k(x), \quad f_k(x) = F_k(z^k + z'(x^k)(x - x^k)),$$

$$z^k = z(x^k), \quad F_k(z) = F(z^k) + (\nabla F(z^k), z - z^k) + + (\nabla^2 F(z^k)(z - z^k)/2, z - z^k)/2, \quad (14)$$

В частности, в методе наименьших квадратов  $F_k(z) = F(z) = \|z\|^2$ , а для линейной модели (14) совпадает с методом Ньютона. Итерационный процесс (14) называется *методом Гаусса — Ньютона*; на каждом шаге в нем ищется оценка наименьших квадратов для линеаризованной модели. Можно показать локальную сходимость метода (14) со скоростью геометрической прогрессии; знаменатель прогрессии тем меньше, чем меньше  $z(x_m^*)$  (т. е. невязки в точке минимума). Методу (14) удается придать и глобальную сходимость подобно тому, как это делается в методе Ньютона (см. § 1 гл. 3).

Итак, для решения задачи вида (8), (9) разумно применять либо общие методы типа сопряженных градиентов, либо специальные методы типа Гаусса — Ньютона. Первые более просты; во вторых на каждом шаге нужно обращать матрицу, однако они могут сходиться быстрее. Численные эксперименты показали примерно равную эффективность обеих групп методов. Отметим еще, что задачи (8) и (9) не имеет смысла решать слишком точно, так как  $x_m^*$  все равно не совпадает с истинным значением  $x^*$ .

**3. Робастное оценивание.** Оценки максимального правдоподобия, будучи асимптотически эффективными, не являются устойчивыми к отклонениям закона распределения помехи от предполагаемого. Покажем это на простейшем примере. Оценка (5) обладает наименьшим вторым моментом среди любых несмещенных оценок для случая нормально распределенных данных. Пусть, однако, одно измерение распределено иначе — у него очень большая (или бесконечная) дисперсия. Тогда и у оценки (5) дисперсия будет велика или бесконечна. В то же время описанная ситуация часто возникает на практике — она соответствует, например, поломке измерительного прибора, грубой ошибке при записи данных, сбое вычислительной машины. Каждое из этих событий редко, однако их влияние на оценку (5) будет катастрофичным.

Путь борьбы с такого рода осложнениями был предложен швейцарским статистиком Хубером в 1964 г. Он исходил из предположения, что истинная плотность помехи  $\rho$  неизвестна, а известен лишь некоторый класс  $\mathcal{P}$ , которому она принадлежит. Хубер указал оценки, которые работоспособны для всех  $\rho \in \mathcal{P}$ . Примитивно к регрессионной задаче (10) эти оценки имеют вид

$$x_m^* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f_m(x), \quad f_m(x) = \sum_{i=1}^m |y_i - (u_i^t, x)|,$$

$$F(z) = -\ln \rho^*(z), \quad \rho^* = \underset{p \in \mathcal{P}}{\operatorname{argmin}} J(p), \quad J(p) = \int \frac{\rho'(z)^2}{\rho(z)} dz. \quad (15)$$

Иначе говоря, нужно найти *наименее благоприятное распределение*  $\rho^* \in \mathcal{P}$  (обладающее наименьшей фишеровской информацией), а затем использовать отвечающую ему оценку максимального правдоподобия. Оказывается, при естественных предположениях такие оценки являются *асимптотически минимаксными* на  $\mathcal{P}$ , т. е. в известном смысле оптимальными. Приведем хуберовские оценки для двух наиболее важных классов  $\mathcal{P}$ . Если  $\mathcal{P}$  — *класс всех невырожденных распределений*, то

$$x_m^* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f_m(x),$$

$$f_m(x) = \sum_{i=1}^m |y_i - (u_i^t, x)|, \quad (16)$$

т. е. если о распределении ничего не известно, то нужно пользоваться методом наименьших модулей. Если  $\mathcal{P}$  — *класс «приближенно нормальных» распределений*, то

$$x_m^* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f_m(x),$$

$$f_m(x) = \sum_{i=1}^m F(y_i - (u_i^t, x)),$$

$$F(z) = \begin{cases} z^2/2, & |z| \leq d, \\ |z|d - d^2/2, & |z| > d, \end{cases} \quad (17)$$

где параметр  $d$  зависит от «уровня загрязнения» основного распределения. Иначе говоря, если распределение близко к нормальному, то нужно пользоваться методом, «промежуточным» между методом наименьших квадратов и методом наименьших модулей.

Как мы видим, робастное оценивание вновь приводит к необходимости решать задачи безусловной минимизации функций.

Последние являются, однако, недостаточно гладкими.

При минимизации (17) метод Ньютона применять нельзя.

Например, если точка  $x$  такова, что  $|y_i - (u_i^t, x)| > d$ ,  $i = 1, \dots, m$ , то  $\nabla f_m(x) = 0$ , и метод Ньютона теряет смысл. В этой ситуации можно пользоваться либо методом Левенберга — Марквардта (§ 1 гл. 3), либо методами первого порядка типа сопряженных градиентов (см. § 2 гл. 3). Поскольку  $f_m(x)$  кусочно-квадратична, то последний метод можно сделать конечным. Существуют и специальные методы минимизации функций вида (17). Например, в методе *Мудрова — Кукко* функции  $F(z)$  аппроксимируются квадратичными, вписанными в график  $F(z)$ ,

т. е. на каждой итерации решается задача методом наименьших

квадратов:

$$\begin{aligned} x^{k+1} &= \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^m \mu_i^k (y_i - (u^i, x))^2, \\ z^k &= y_t - (u^t, x^k), \\ \mu_i^k &= \begin{cases} 1, & |z_i^k| \leq d, \\ 1/|z_i^k|, & |z_i^k| > d. \end{cases} \end{aligned} \quad (18)$$

Задача (16) — это задача безусловной минимизации кусочно-линейной функции. Она сводится к задаче линейного программирования при введении дополнительных переменных:

$$\min \sum_{i=1}^m (\ell_i + s_i),$$

$$(u^i, x) - y_i = \ell_i - s_i, \quad i = 1, \dots, m, \quad (19)$$

$$\ell_i \geq 0, \quad s_i \geq 0.$$

Однако при этом резко возрастает число переменных (обычно в задачах регрессии  $m \gg n$ ). Удобно применять итеративные методы минимизации негладких функций из гл. 5, в частности метод с растяжением пространства (14) § 4 гл. 5. Приведем еще один способ решения. Запишем (16) для линейного случая в виде

$$\min \sum_{i=1}^m |\ell_i|, \quad Ux - y = t$$

и составим модифицированную функцию Лагранжа:

$$M(x, t, v, K) = \sum_{i=1}^m |\ell_i| + (v, Ux - y - t) + K \|Ux - y - t\|^2/2, \text{ где}$$

$v \in \mathbb{R}^m$  играет роль двойственных переменных. Переменную  $t$  можно исключить, так как минимум  $M$  по  $t$  находится в явном виде, и метод модифицированной функции Лагранжа (23) § 3 приобретает вид

$$x^{k+1} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^m F((u^i, x) - y_i + \frac{1}{K} v^k),$$

$$v^{k+1} = K F'((u^i, x^{k+1}) - y_i + \frac{1}{K} v^k), \quad (20)$$

$$F(z) = \begin{cases} z^2/2, & |z| \leq 1/K, \\ 1/K|z| - 1/2K^2, & |z| > 1/K. \end{cases}$$

В результате получаем *метод Бергекаса*, в котором для решения негладкой задачи (16), на каждой итерации нужно искать

минимум дифференцируемой функции вида (17). Иначе говоря, производится итеративное сглаживание функции (16). В силу конечности метода модифицированной функции Лагранжа для задач линейного программирования (см. теорему 4 § 3 гл. 10) метод (20) конечен.

Упражнение.

3. Найдите решение задачи (16) для простейшей задачи с  $u^i = 1, x \in \mathbb{R}^1$ . Ответ.  $x_m^* — выборочная медиана$ , т. е. точка, для которой среди чисел  $y_i$  найдется равное число больших и меньших ее.

**4. Рекуррентное оценивание.** Все описанные выше методы оценивания сводятся к довольно сложной задаче безусловной минимизации. Более того, при поступлении каждого нового измерения эту задачу приходится решать заново. Поэтому при обработке последовательно приходящих данных в реальном масштабе времени методы типа максимального правдоподобия не-рациональны.

Опишем иной подход, удобный для оценивания в такой ситуации. Для задачи оценки параметров из п. 1 введем функцию  $f(x) = MQ(z, x) = \int Q(z, x) p(z, x^*) dz$ ,  $Q(z, x) = -\ln p(z, x)$ . (21)

Легко видеть, что  $x^*$  — точка минимума  $f(x)$ :

$$\begin{aligned} f(x) &= - \int (\ln p(z, x^*)) p(z, x^*) dz - \\ &\quad - \int \ln \left[ 1 + \frac{p(z, x) - p(z, x^*)}{p(z, x^*)} \right] p(z, x^*) dz \geqslant \\ &\geqslant f(x^*) - \int [p(z, x) - p(z, x^*)] dz = f(x^*). \end{aligned}$$

Здесь мы воспользовались тем, что  $-\ln(1 + \alpha) \geqslant -\alpha$  для всех  $\alpha$ . Итак, исходная задача заключается в минимизации функции (21). Последняя нам неизвестна (так как в (21) входит неизвестная плотность  $p(z, x^*)$ ). В методе максимального правдоподобия  $f(x)$  аппроксимировалась функцией (3):

$$f_m(x) = \frac{1}{m} \sum_{i=1}^m Q(z^i, x),$$

и в качестве оценки  $x^*$  бралась точка минимума этой аппроксимации. Однако можно поступить иначе. Пусть измерения  $z^1, \dots, z^k, \dots$  поступают последовательно, и  $x^k$  — найденное после обработки  $k$  измерений приближение для  $x^*$ . Тогда можно счи-тать  $\nabla Q(z^{k+1}, x^k)$  приближением для  $\nabla f(x^k) = M \nabla_x Q(z, x^k)$  (ср.

с (1) — (3) в § 1 гл. 4) и применить метод градиентного типа для минимизации  $f(x)$ :

$$x^{k+1} = x^k - \gamma_k \nabla_x Q(z^{k+1}, x^k) = x^k - \gamma_k \frac{\nabla_x p(z^{k+1}, x^k)}{p(z^{k+1}, x^k)}. \quad (22)$$

Мы получили метод иного типа, нежели рассматривавшиеся ранее. Ранее обработке подвергались одновременно *т* измерений — составлялась функция вида (3), и для ее минимизации применялся какой-либо итеративный метод. В процедуре (22) каждое измерение используется лишь однократно, не требуется ни запоминания всех предыдущих измерений, ни оперирования со всеми членами суммы в (3). В подобных случаях говорят об *аддитивных* или *рекуррентных оценках* (часто употребляют также термин *стохастическая аппроксимация*).

Используя результаты § 2 гл. 4, нетрудно получить утверждения о сходимости (22) для  $k \rightarrow \infty$  при естественных предположениях. Иными словами, оценки (22) являются состоятельными. Могло бы оказаться, что эти оценки сильно уступают оценкам максимального правдоподобия (4) по точности. Однако это не так. Если взять в (22)  $\gamma_k = \gamma/k$ , то при соответствующем подборе  $\gamma$  скорость сходимости оказывается по порядку такой же, как для (4), т. е. порядка  $O(1/k)$ . Если же рассмотреть несколько более общий метод, чем (22):

$$x^{k+1} = x^k - \frac{1}{k+1} J^{-1} \frac{\nabla_x p(z^{k+1}, x^k)}{p(z^{k+1}, x^k)}, \quad (23)$$

где  $J$  — фишеровская информационная матрица (6), то оценки (23) оказываются асимптотически эффективными. Иначе говоря, рекуррентные оценки, будучи гораздо проще оценок максимального правдоподобия, не уступают им по асимптотическим свойствам.

Аналогичный подход возможен и для задачи регрессии (10). Рекуррентные оценки вида

$$x^{k+1} = x^k - H_k u^{k+1} \Psi((\mu^{k+1}, x^k) - y_{k+1}) \quad (24)$$

используют измерения последовательно. При достаточно общих предположениях о матрицах  $H_k$  и функциях  $\Psi: \mathbf{R}^l \rightarrow \mathbf{R}^l$  они оказываются состоятельными, а если выбрать

$$H_k = \frac{1}{k+1} J^{-1} B^{-1}, \quad J = \int \frac{p'(z)^2}{p(z)} dz, \quad (25)$$

то и асимптотически эффективными. Таким образом, здесь нет нужды решать сложные задачи типа (8), а результаты (с асимп-

тической точки зрения) оказываются эквивалентными оценкам (8).

В простейшем случае задачи (1), (2) оценки (23) — (25) принимают вид  $x^{k+1} = x^k - (x^k - z^{k+1})/(k+1)$  и в точности совпадают с нерекуррентной оценкой (5). В общем случае рекуррентные и нерекуррентные оценки, вообще говоря, различны.

В ситуации, когда плотность распределения помехи  $p$  не полностью известна, можно применить *рекуррентный вариант рабочей оценки* (15). Для этого нужно в (24), (25) заменить  $\Psi(z)$  на  $-\rho^*(z)'/\rho^*(z)$ , где  $\rho^*$  — «наименее благоприятное» расположение (15). При этом сохраняются те же асимптотические свойства, что и для оценки (15).

Таким образом, в случае большого объема выборки рекуррентные процедуры оценивания обладают заметными преимуществами перед нерекуррентными.

Отметим еще, что для метода наименьших квадратов можно выписать рекуррентную схему (см. (13) § 4 гл. 4), не требующую решения систем линейных уравнений и дающих точно те же оценки, что и (11), для любого числа измерений.

**Упражнение.** 4. Покажите, что  $J = \nabla^2 f(x^*)$  для  $f(x)$  вида (21). Таким образом, (23) есть метод типа Ньютона (вида (18) § 1 гл. 2) для минимизации  $f(x)$  при наличии помех, однако в нем не используются явно вторые производные.

**5. Анализ данных.** Выведено предполагалось, что имеется некоторая модель объекта, известная с точностью до параметров. Требовалось по «выходам» этой модели, осложненным случайными ошибками, восстановить неизвестные параметры. Однако такая схема нередко является слишком идеализированной. Во первых, мы обычно не знаем точного вида модели. В некоторых ситуациях он известен лишь приближенно. Например, при математическом описании химического реактора пренебрегают побочными реакциями, неоднородностью процесса по объему, стационарным катализаторов и т. д. В других случаях адекватная модель объекта известна, но она слишком громоздка, чтобы ее можно было воспользоваться. Наконец, во многих задачах анализируемый объект вообще не имеет точного математического описания. Таковы задачи, связанные с деятельностью человека, с описанием социальных и экономических процессов, с разного рода геологическими и космическими явлениями (против землетрясений или солнечной активности). Во-вторых, делавшиеся выше вероятностные предположения о природе помех в реальных задачах обычно также не выполняются. Так, при обосновании метода максимального правдоподобия для задач регрессии (8) приходится предполагать, что помехи  $\xi_i$  независимы, центрированы, одинаково распределены с известной плотностью. Между тем на практике ошибки бывают коррелированы, содержат систематическую составляющую, их распределение меняется и т. д.

Как мы видели на примере робастных оценок, все эти нарушения исходных предположений могут быть отнюдь не безобидны.

В связи с этим возможности и области применения статистических методов идентификации ограничены. Задачи описания объектов при отсутствии каких-либо предположений об их вероятностной природе относят к области *анализа данных*. В таких задачах требуется описать имеющиеся данные с помощью какой-либо модели так, чтобы рассогласование между моделью и объектом было достаточно мало. При этом выбор как класса моделей, так и меры рассогласования в достаточной мере произволен и определяется такими факторами, как простота вычислений оценок, простота получаемых формул и т. п. Конечно, такой эмпирический подход имеет и оборотную сторону — неясно, какова достоверность полученного описания объекта, насколько им можно пользоваться для целей экстраполяции и прогноза и т. п. Вопрос о выборе типа модели весьма сложен, и мы остановимся лишь на задачах подгонки параметров модели после того, как выбор последней уже сделан. Типичная задача имеет следующий вид. Пусть вектор  $u \in \mathbf{R}^s$  — «вход» объекта,  $y \in \mathbf{R}^1$  — «выход» объекта,  $x \in \mathbf{R}^n$  — подбираемые параметры,  $\Phi(u, x)$  — модель объекта (т. е. прогноз выхода при фиксированных значениях входа и параметров),  $F(z)$  — выбранная мера для учета рассогласования модели и объекта. Имеются данные  $u_i^i, y_i$ ,  $i = 1, \dots, m$ , — результаты  $m$  измерений выходов при значениях входных переменных, равных  $u^i$ . Требуется подобрать параметры так, чтобы минимизировать отклонение предсказываемых значений выхода от фактических, т. е. нужно найти

$$\min_{x \in \mathbf{R}^n} f(x), \quad f(x) = \sum_{i=1}^m F(y_i - \Phi(u^i, x)). \quad (26)$$

В качестве меры рассогласования обычно берут

$$F(z) = |z|, \quad (27)$$

$$F(z) = z^2. \quad (28)$$

Обратите внимание на то, что мы вновь пришли к задаче вида (12), однако из соображений, никак не связанных с математической статистикой.

Критерий рассогласования не обязательно должен носить аддитивный характер, например часто выбирают минимаксный критерий

$$f(x) = \max_{1 \leq i \leq n} |y_i - \Phi(u^i, x)|. \quad (29)$$

Нужно сказать, что ранее в практике идентификации в качестве  $F(z)$  почти исключительно бралась квадратичная функция (28), т. е. применялся метод наименьших квадратов. При-

чины этого были связаны как с традициями (опиравшимися на слишком расширенное применение теорем математической статистики об оптимальности метода наименьших квадратов), так и с соображениями простоты вычислений. Ниже оба эти мотива потеряли свое значение в связи с неробастностью метода наименьших квадратов при отходе от лежащих в его основе предположений и развитием численных методов минимизации, позволяющих практически одинаково просто решать задачи с различными  $F(z)$ .

В ряде случаев об объекте имеются в виде *непрерывных*, а не дискретных измерений. Пусть, например,  $t \in \mathbf{R}^1$  играет роль времени, и задан выход объекта  $y(t)$  при известном в ходе  $u(t)$ ,  $0 \leq t \leq T$ . Тогда аналоги задач (26), (27); (26), (28); (29) принимают вид

$$f(x) = \int_0^T |y(t) - \Phi(u(t), x)| dt, \quad (30)$$

$$f(x) = \int_0^T (y(t) - \Phi(u(t), x))^2 dt, \quad (31)$$

$$f(x) = \max_{0 \leq t \leq T} |y(t) - \Phi(u(t), x)|. \quad (32)$$

Таким образом, мы вновь приходим к задаче конечномерной безусловной минимизации, однако каждое вычисление функции  $f(x)$  включает операцию интегрирования или взятия максимума. В некоторых случаях эти операции могут быть выполнены в явном виде. Пусть, например, решается задача (31) для  $\Phi(u(t), x) = \sum_{i=1}^n x_i u_i(t)$ . Она сводится к минимизации функции

$$f(x) = \sum_{i,j=1}^n x_i x_j \int_0^T u_i(t) u_j(t) dt - 2 \sum_{i=1}^n x_i \int_0^T y(t) u_i(t) dt. \quad (33)$$

Если, например,  $u_i(t) = t^{i-1}$  (т. е. ищется среднеквадратичное приближение  $y(t)$  полиномами),  $T = 1$ , то

$$\int_0^T u_i(t) u_j(t) dt = \frac{1}{i+j-1}, \quad \int_0^T y(t) u_i(t) dt = \int_0^1 y(t) t^{i-1} dt.$$

Иначе говоря, для решения требуется вычислить моменты функции  $y(t)$  и найти минимум квадратичной формы (33) с матрицей Гильберта

$$\left( \left( \frac{1}{i+j-1} \right) \right)_{i,j=1}^n. \quad (34)$$

Эта задача весьма поучительна. Оказывается, что обусловленность матрицы Гильберта равна  $\mu = 1,5 \cdot 10^7$  для  $n = 6$  и  $\mu = 1,6 \cdot 10^{13}$  для  $n = 10$ . Иначе говоря, даже для задач малой размерности возникают очень плох обусловленные задачи минимизации. Им соответствует пизкая скорость сходимости градиентных методов — в соответствии с теоремой 3 § 4 гл. I и теоремой 2 § 1 гл. 3 при оптимальном выборе шага градиентный метод сходится со скоростью геометрической прогрессии со знаменателем  $q \approx 1 - 2\mu^{-1}$ , т. е.  $q \approx 1 - 10^{-7}$  для  $n = 6$  и  $q \approx 1 - 10^{-13}$  для  $n = 10$ , так что нужно сделать  $\sim 10^7$  шагов градиентного метода при  $n = 6$ , чтобы увеличить точность приближения всего в  $e$  раз. Иными словами, уже для  $n = 6$  градиентный метод сходится (по  $x$ ) безнадежно медленно. Кроме того, задача минимизации становится практически неустойчивой (см. § 3 гл. I). Действительно, ошибки порядка  $\varepsilon$  при определении моментов приводят к ошибкам порядка  $\varepsilon$  в определении точки минимума. Поэтому, даже если точка минимума ищется точно, но допускается малая ошибка ( $\sim 10^{-7}$ ) при вычислении моментов, то она приведет к ошибке  $\sim 1$  в определении коэффициентов полинома для  $n = 6$  и  $\sim 10^6$  для  $n = 10$ , т. е. для  $n = 10$  найденный полином не будет иметь никакого отношения к истинному полиному наилучшего среднеквадратичного приближения.

Однако не стоит драматизировать ситуацию. Во-первых, данная задача является критериальной (см. § 1 гл. 6) — нам не важны коэффициенты теоретического полинома наилучшего приближения; требуется указать любой полином, хорошо приближающий  $y(t)$  в среднеквадратичном смысле. Поэтому производный вектор  $x$ , соответствующий значениям  $f(x)$ , близким к минимальному, является приемлемым. При этом, как уже отмечалось в § 1 гл. 6, итеративные методы типа сопряженных градиентов сходятся достаточно быстро даже для вырожденных задач, позволяют учесть априорную информацию о решении и устойчивы к погрешностям. Многочисленные эксперименты [11, 22] показали, что хороший вариант метода сопряженных градиентов дает практически точный минимум по функции, а для гладких  $y(t)$  получается и хорошее приближение к решению (для  $n \leq 10$ ). Во-вторых, все трудности при решении задачи могут быть сняты, если искать разложение  $y(t)$  не по обычным полиномам, а по *ортогональным полиномам*. Тогда

$$\int_0^T u_i(t) u_j(t) dt = \begin{cases} 0, & i \neq j, \\ 1, & i = j, \end{cases}$$

поэтому  $f(x)$  (33) приобретает вид

$$f(x) = \sum_{i=1}^n (x_i^2 - 2a_i x_i), \quad a_i = \int_0^T y(t) u_i(t) dt$$

и ее минимум может быть найден явно:  
 $x_i^* = a_i, \quad i = 1, \dots, n.$

Таким образом, здесь не возникает никаких проблем, связанных с плохой обусловленностью. Этот пример показывает, какую большую роль играет правильный выбор формы записи модели. Разного рода замены переменных, масштабные преобразования, и т. п. могут резко улучшить свойства минимизируемой функции, уменьшить число обусловленности и существенно упростить процесс минимизации.

#### Упражнения.

5. Решите явно задачи в  $\mathbf{R}^1$  для  $y_1 = \dots = y_{m-1} = 0, \quad y_m = \alpha$ :

- a)  $\min \sum_{i=1}^m (y_i - x)^2$ ; b)  $\min \sum_{i=1}^m |y_i - x|$ ; в)  $\min \max_{1 \leq i \leq m} |y_i - x|$ . О тв.
- в. т. а)  $x^* = \alpha/m$ ; б)  $x^* = 0$ ; в)  $x^* = \alpha/2$ . Сделайте отсюда выводы о влиянии единичных выбросов при различных критериях оптимальности.
6. Требуется аппроксимировать давние  $y_i$  в точках  $t_i, t = 1, \dots, m$ , за-  
висимостью вида  $x_1 \exp x_2 t$ . Составьте следующие способы: а) найти  $\min \sum_{i=1}^m (y_i - x_1 \exp x_2 t_i)^2$  по  $x \in \mathbf{R}^2$ ; б) найти  $\min \sum_{i=1}^m |y_i - x_1 \exp x_2 t_i|$  по  $x \in \mathbf{R}^2$ ; в) сделать замену  $a_i = \ln y_i, \quad z_1 = \ln x_1, z_2 = x_2$  и найти  $\min \sum_{i=1}^m (a_i - z_1 - z_2 t_i)^2$  по  $z \in \mathbf{R}^2$ ; г) составить разности  $\Delta y_i = y_{i+1} - y_i$ .

$$\Delta t_i = t_{i+1} - t_i > 0 \text{ и взять } x_2^m = \frac{1}{m} \sum_{i=1}^m \frac{1}{t_i} \frac{\Delta y_i}{\Delta t_i}, \quad x_1^m = \frac{1}{m} \sum_{i=1}^m y_i \exp(-x_2^m t_i).$$

Какой из способов наиболее прост? Какой дает точное решение, если  $y_i = x_1^* \exp x_2^* t_i$ ? Какой более надежен (на эвристическом уровне)?

6. Другие задачи идентификации. Выше речь шла лишь о статических моделях объекта типа регрессионной (7). При идентификации динамических объектов приходится рассматривать модели, описываемые дифференциальными или разносточными уравнениями и подбирать входящие в них параметры или начальные условия. Здесь возможны те же подходы, что и для статических задач, однако возникающие при этом задачи безусловной минимизации имеют свои особенности. Во-первых, вычисление градиента функции требует решения вспомогательных линейных дифференциальных уравнений (уравнений чувствительности) высокой размерности, а потому возможно лишь в простейших задачах. Во-вторых, даже для линейных моделей и квадратичных критерии минимизируемые функции могут оказаться неквадратичными относительно искомых параметров. В-третьих, обычно по физическому смыслу задачи параметры не могут принимать произвольные значения (например, описываемый объект должен быть устойчивым). Это приводит к необходимости

минимизации при наличии ограничений. Наконец, рекуррентные оценки (типа (24)) требуют специального (и весьма сложного) обоснования. Все это показывает, что задачи идентификации динамических объектов являются существенно более сложными, чем в статическом случае.

Все рассмотренные выше задачи были связаны с обработкой результатов «пассивных» экспериментов. В ряде случаев имеется возможность активного вмешательства в формирование данных (выбор входных переменных в регрессионной модели (7), выбор моментов измерения и т. п.). Подобными проблемами занимается теория *планирования эксперимента*. Здесь возникают задачи оптимизации трех типов. Первый связан с выбором оптимального в том или ином смысле плана эксперимента (*A*-оптимальные, *D*-оптимальные и т. д. планы). Здесь характерны детерминированные задачи на экстремум при наличии ограничений (последние связаны с ограничениями на входные воздействия). Возникающие при этом задачи многоэкстремальны и редко допускают явное решение. Положение существенно упрощается при переходе к так называемым непрерывным планам. В планировании эксперимента существует большой набор решений подобных оптимизационных задач, и соответствующие оптимальные (или приближенно оптимальные) планы собраны в специальные каталоги. Второй тип задач связан с обработкой результатов эксперимента при фиксированном плане, т. е. с решением рассмотренной выше проблемы идентификации.

Наконец, во многих случаях целью исследования является не локальное описание зависимости некоторого показателя от других переменных, а отыскание экстремума этого показателя. Например, пусть ищется состав многокомпонентного сплава, обладающего максимальной прочностью. Тогда вначале с помощью оптимального плана эксперимента строится зависимость прочности от состава в окрестности некоторого исходного сплава. Затем на основе этой зависимости выбирается новая базисная точка, и процесс повторяется. Таким образом, здесь приходится решать задачу безусловной минимизации функции, значения которой в каждой точке вычисляются со случайной помехой (см. § 4 гл. 4). Следует отметить, однако, что теория планирования экспериментов развивается в отрыве от теории оптимизации, и в ней практически не используются разработанные в последней методы. Этому во многом способствует специфическая терминология, сложившаяся в планировании эксперимента и затрудняющая взаимопонимание.

Последний класс задач идентификации, который мы лишь упомянем, связан с восстановлением функций, входящих в описание системы. Задачи такого типа возникают в геофизике (интерпретация данных электрических и сейсморазведки, обратные задачи магнито- и гравиметрии), медицине (обработка электро-

кардиограмм и энцефалограмм), радиотехнике и т. д. Данный тип проблем идентификации приводит к бесконечномерным задачам минимизации, как правило, неустойчивым.

Все рассмотренные выше задачи были связаны с обработкой результатов «пассивных» экспериментов. В ряде случаев имеется возможность активного вмешательства в формирование данных (выбор входных переменных в регрессионной модели (7), выбор моментов измерения и т. п.). Подобными проблемами занимаются теория *планирования эксперимента*. Здесь возникают задачи оптимизации трех типов. Первый связан с выбором оптимального в том или ином смысле плана эксперимента (*A*-оптимальные, *D*-оптимальные и т. д. планы). Здесь характерны детерминированные задачи на экстремум при наличии ограничений (последние связаны с ограничениями на входные воздействия). Возникающие при этом задачи многоэкстремальны и редко допускают явное решение. Положение существенно упрощается при переходе к так называемым непрерывным планам. В планировании эксперимента существует большой набор решений подобных оптимизационных задач, и соответствующие оптимальные (или приближенно оптимальные) планы собраны в специальные каталоги. Второй тип задач связан с обработкой результатов эксперимента при фиксированном плане, т. е. с решением рассмотренной выше проблемы идентификации.

Наконец, во многих случаях целью исследования является не локальное описание зависимости некоторого показателя от других переменных, а отыскание экстремума этого показателя. Например, пусть ищется состав многокомпонентного сплава, обладающего максимальной прочностью. Тогда вначале с помощью оптимального плана эксперимента строится зависимость прочности от состава в окрестности некоторого исходного сплава. Затем на основе этой зависимости выбирается новая базисная точка, и процесс повторяется. Таким образом, здесь приходится решать задачу безусловной минимизации функции, значения которой в каждой точке вычисляются со случайной помехой (см. § 4 гл. 4). Следует отметить, однако, что теория планирования экспериментов развивается в отрыве от теории оптимизации, и в ней практически не используются разработанные в последней методы. Этому во многом способствует специфическая терминология, сложившаяся в планировании эксперимента и затрудняющая взаимопонимание.

Последний класс задач идентификации, который мы лишь упомянем, связан с восстановлением функций, входящих в описание системы. Задачи такого типа возникают в геофизике (интерпретация данных электрических и сейсморазведки, обратные задачи магнито- и гравиметрии), медицине (обработка электро-

### Упражнения.

7. Рассмотрите модель  $y_{t+1} = x^*y_t + \xi_t$ ,  $y_1, \dots, y_k \in \mathbf{R}^1$  — заданные числа,  $\xi_t$  — случайные независимые помехи. Выпишите метод наименьших квадратов и метод максимального правдоподобия для оценки параметра  $x^* \in \mathbf{R}^1$ . Напишите рекуррентные версии этих методов.

8. Заданная функция  $y^0(t)$  приближена уравнением  $dy/dt = -x^*y(t)$  с неизвестным начальным условием. Придумайте различные постановки задачи оценки  $x^*$ . В каких случаях решение выписывается явно? Сравните с упражнением 6.

## § 2. ОПТИМИЗАЦИОННЫЕ ЗАДАЧИ В ТЕХНИКЕ И ЭКОНОМИКЕ

1. Оптимальное проектирование. После того как составлена математическая модель проектируемого объекта, можно решать задачу об оптимальном выборе параметров. Предполагается, что задан некоторый показатель (целевая функция), который подлежит оптимизации, и сформулированы ограничения на допустимые значения параметров и других характеристик объекта. Здесь возможны задачи, связанные с наилучшим выбором параметров отдельного изделия (например, оптимизация характеристик электрических машин, электронных приборов, строительных конструкций), технологического процесса (например, оптимизация химического реактора), предприятия или отрасли (например, оптимизация задачи оптимального размещения производства, оптимизация трассы газопровода или электрической сети). При этом типичен случай нелинейной целевой функции и нелинейных ограничений при не слишком большой размерности (редко приходится подбирать больше чем несколько десятков параметров).

Поскольку характеристики объекта обычно описываются достаточно сложными зависимостями, их часто аппроксимируют полиномиальными выражениями. При этом возникают задачи вида

$$\min f_0(x), \quad x \in \mathbf{R}^n,$$

$$f_l(x) \leqslant 1, \quad l = 1, \dots, m, \quad (1)$$

$$0 < a_i \leqslant x_i \leqslant b_i, \quad i = 1, \dots, n,$$

где

$$f_l(x) = \sum_{j=1}^N a_{jl} x_1^{b_{1jl}} x_2^{b_{2jl}} \dots x_n^{b_{njl}}, \quad (2)$$

Если  $a_{jl} \geqslant 0$ , а  $b_{ijl}$  — произвольные числа, то функции (2) называют *позициями*, а (1), (2) — *задачей геометрического программирования*. Позиции, вообще говоря, не являются выпуклыми функциями. Поэтому общие методы нелинейного

программирования неэффективны для задач геометрического программирования. Однако существуют специальные приемы, позволяющие свести задачу (1), (2) к выпуклой. Можно сделать преобразование

$$z_i = \ln x_i,$$

тогда (1) принимает вид

$$\min g_0(z), \quad g_i(z) = \sum_{j=1}^N a_{ij} \exp\left(\sum_{l=1}^n b_{il} z_l\right), \quad (3)$$

$$g_i(z) \leqslant 1, \quad l = 1, \dots, m,$$

$c_i \leqslant z_i \leqslant d_i, \quad c_i = \ln a_i, \quad d_i = \ln b_i, \quad i = 1, \dots, n.$  Поскольку  $g(z) = \exp(b, z)$  — выпуклая функция, то (4) — задача выпуклого программирования (с гладкими  $g_i(z)$ ). Можно свести (1) и к более простой задаче; покажем это на примере задачи без ограничений. Исходная задача

$$\min \sum_{j=1}^N a_j x^{b_j} \dots x_n^{b_n}, \quad x > 0, \quad (5)$$

заменой (3) приводится к задаче

$$\min \sum_{j=1}^N a_j \exp u_j, \quad (6)$$

$$(b^j, z) = u_j, \quad j = 1, \dots, N, \quad b^j = \{b_{1j}, \dots, b_{nj}\},$$

где  $u_j$  — дополнительные переменные. Составим задачу, двойственную к (6):

$$L(z, u, y) = \sum_{j=1}^N (a_j \exp u_j + y_j((b^j, z) - u_j)),$$

$$\Psi(y) = \inf_{z, u} L(z, u, y) =$$

$$= \begin{cases} \sum_{j=1}^N y_j \left(1 - \ln \frac{y_j}{a_j}\right), & \text{если } \sum_{j=1}^N y_j b^j = 0, \quad y > 0, \\ -\infty & \text{в противном случае,} \end{cases}$$

т. е. двойственная задача имеет вид

$$\min \sum_{j=1}^N (y_j \ln y_j - y_j(1 + \ln a_j)), \quad \sum_{j=1}^N y_j b^j = 0, \quad y > 0 \quad (7)$$

и оказывается задачей с выпуклой гладкой минимизируемой функцией и линейными ограничениями. Для ее решения можно

применять методы гл. 7; найдя  $y^*$ , следует вычислить  $u_j^* = \ln(y_j^*/a_j)$ , решить систему  $(b^j, z) = u_j^*, j = 1, \dots, N$ , и отыскать оптимальные значения исходных переменных  $x_i^* = \exp z_i^*$ . Полезно обратить внимание на следующие обстоятельства. Во-первых, мы еще раз убеждаемся, как важно правильно выбрать независимые переменные. Исходная задача (1) невыпукла и многоэкстремальна, однако замена (3) приводит к выпуклой задаче. Во-вторых, соображения двойственности оказались работоспособными в ситуации, где на первый взгляд не имеет смысла говорить о двойственности — именно, для задачи безусловной минимизации (5). Впрочем, мы уже сталкивались с подобным приемом (ср. с методом (20) § 1 для задачи безусловной минимизации (16) § 1).

Конечно, задачи оптимального проектирования не исчерпываются схемой геометрического программирования. Трудности, связанные с многоэкстремальностью, не всегда обходятся столь легко. Еще более сложную проблему представляет типичное для задач проектирования требование целочисленности некоторых переменных (параметры изделий принимают лишь значения, задаваемые стандартами; количество станков должно быть целым, объем производства проектируемого завода должен соответствовать типовым проектам; существует конечное число конструкций соединительной сети или размещений электронных устройств на схеме и т. д.). В этой книге, однако, мы не касаемся подобных комбинаторных оптимизационных задач и методов их решения.

**2. Оптимальное распределение ресурсов.** Ряд задач поиска неисправностей, обнаружения цели, планирования эксперимента укладываются в следующую простую схему. Имеется некоторый ресурс. Как наилучшим образом распределить его по  $n$  объектам, если эффективность использования на  $i$ -м объекте задается функцией  $\Phi_i(x)$ ? Математически:

$$\min [\Phi_1(x_1) + \dots + \Phi_n(x_n)], \quad x_1 + \dots + x_n = 1, \quad x_i \geqslant 0, \quad i = 1, \dots, n. \quad (8)$$

Вид минимизируемой функции и ограничений здесь так прост, что решение часто удается получить в явном или «полуявном» виде.

Пусть  $\Phi_i(x_i)$  — вогнутые функции (в частности, линейные), тогда минимум достигается в вершине симплекса, задающего ограничения. Но этих вершин всего  $n$  и они имеют вид  $\{1, 0, \dots, 0\}, \{0, 1, 0, \dots, 0\}, \dots, \{0, 0, \dots, 1\}$ . Поэтому достаточно найти  $j = \arg\min_{1 \leqslant i \leqslant n} \Phi_i(1)$  и взять в качестве решения  $x_j^* = 1, x_i^* = 0, i \neq j$ . Иначе говоря, весь ресурс следует сосредоточить на одном объекте.

Пусть теперь  $\Phi(x_i)$  — выпуклые функции. Составим функцию  $L(x, y) = \Phi_1(x_1) + \dots + \Phi_n(x_n) + y(x_1 + \dots + x_n - 1)$ ,  $y \in \mathbf{R}^1$ , (9) и введем одномерные функции

$$\Psi_i(y) = \inf_{x_i \geq 0} [\Phi_i(x_i) + yx_i]. \quad (10)$$

Тогда

$$\Psi(y) = \inf_{x \geq 0} L(x, y) = \Psi_1(y) + \dots + \Psi_n(y) - y, \quad (11)$$

а теорема двойственности (см. § 1 гл. 9) гарантирует, что точка максимума  $\Psi(y)$  является множителем Лагранжа для исходной задачи. Итак, для решения задачи нужно построить функции  $\Psi_i(y)$  (10), найти точку максимума  $y^*$  одномерной вогнутой функции  $\Psi_i(y)$  и затем найти  $x_i^* = \arg \min_{x_i \geq 0} [\Phi_i(x_i) + y^* x_i]$ . Такую процедуру часто можно осуществить в аналитическом виде. Так, если  $\Phi_i(x_i) = (\lambda_i/2)(x_i - a_i)^2$ ,  $\lambda_i > 0$ , то

$$\Psi_i(y) = -\frac{1}{2\lambda_i} (\lambda_i a_i - y)^2 + \frac{\lambda_i a_i^2}{2},$$

$$\Psi'(y) = \sum_{i=1}^n \frac{1}{\lambda_i} (\lambda_i a_i - y)_+ - 1, \quad (12)$$

т. е.  $\Psi'(y)$  — кусочно-линейная функция. Ее корень можно найти так: упорядочить числа  $\lambda_i a_i$  и последовательно вычислять  $\Psi'(\lambda_i a_i)$ , пока это выражение не переменит знак, после чего  $y^*$  находится линейной интерполяцией.

Наконец, еще один подход к задаче (8) связан с идеями *динамического программирования*. Обозначим

$$f_k(\alpha) = \min_{\substack{x_1 + \dots + x_k = \alpha \\ x_i \geq 0}} [\Phi_1(x_1) + \dots + \Phi_k(x_k)]. \quad (13)$$

Тогда функции  $f_k(\alpha)$  связаны рекуррентным соотношением

$$\hat{f}_{k+1}(\alpha) = \min_{0 \leq x_{k+1} \leq \alpha} [f_k(\alpha - x_{k+1}) + \Phi_{k+1}(x_{k+1})], \quad (14)$$

причем  $\hat{f}_1(\alpha) = \Phi_1(\alpha)$ , а задача заключается в вычислении  $f_n(1)$ .

Таким образом, вместо одной задачи минимизации функции  $n$  переменных мы получили последовательность из  $n-1$  одномерных задач (14). Правда, каждая из этих задач требует решения одномерных задач минимизации при всех значениях параметра  $\alpha \in [0, 1]$ . В некоторых случаях построение функций  $f_k(\alpha)$  удастся в явном виде, в других это нужно делать приближенно,

вычисляя  $f_k(\alpha)$  на некоторой сетке. Существенно, что при таком подходе не требуется никаких предположений типа выпуклости. Существует и много других, более общих постановок задач оптимального распределения ресурсов (например, относящихся к многомерным ресурсам). Однако основные идеи специальных методов их решения (использование теоремы двойственности или динамического программирования) сохраняются.

**3. Оптимальное планирование.** Исторически именно экономические задачи привели к возникновению и развитию линейного программирования. Это естественно, так как в экономике производство экономических моделей являются центральной, а большому программированию сводятся многие задачи оперативного и долгосрочного планирования в рамках предприятий или организаций, задачи оптимизации снабжения и перевозок, задачи управления запасами и т. д. В экономических задачах двойственные переменные могут интерпретироваться как цены на ресурсы и продукцию, а условия экстремума гарантируют, что существуют цены такие, что оптимальный в смысле этих цен план является оптимальным и для исходной задачи. Поэтому линейное программирование является не только инструментом отыскания решения, но и лежит в основе теории математической экономики.

Характерной чертой экономических задач является очень большая размерность (нередко задачи с сотнями и тысячами переменных и ограничений). Матрица ограничений обычно слабо заполнена (большинство ее элементов — нули), хотя и может содержать тысячи ненулевых элементов. В такой ситуации наиболее остро при численном решении стоит проблема памяти ЭВМ, а также вопрос о времени вычислений и их устойчивости к погрешностям. Существующие хорошо отработанные программы симплекс-метода позволяютправляться с этими проблемами, но возможности таких программ не безграничны. Весьма перспективными являются итеративные методы (§ 3 гл. 10). Наконец, очень важным ресурсом при решении больших задач является максимально возможное использование их специфики. Остановимся на этом подробнее.

Важным классом задач линейного программирования являются *транспортные*:

$$\min \sum_{i,j} c_{ij} x_{ij},$$

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

$$\sum_{j=1}^m x_{ij} = b_j, \quad j = 1, \dots, n, \quad x_{ij} \geq 0. \quad (15)$$

В содержащихся терминах:  $a_i$  — объем производства в  $i$ -м пункте,  $b_i$  — объем потребления в  $j$ -м пункте,  $c_{ij}$  — стоимость перевозки единицы груза из  $i$ -го в  $j$ -й пункт,  $x_{ij}$  — искомые объемы перевозок из  $i$ -го в  $j$ -й пункт. Для многогранника ограниченный в транспортной задаче вершины находятся очень просто (см. упр. 1). Поэтому нет нужды во введении дополнительных переменных для отыскания начальной вершины (см. (12) § 2 гл. 10), существенно упрощаются и вычисления на каждой итерации симплекс-метода. Поэтому, хотя в (15) число переменных равно  $m \times n$ , удается решать задачи с большими  $m$  и  $n$  (порядка  $10^3$ ).

Многие задачи линейного программирования имеют *блочную структуру*:

$$\min \sum_{i=1}^s (c^i, x_i), \quad (16)$$

$$x_i \in Q_i, \quad \sum_{i=1}^s A_i x_i = b.$$

Здесь  $x_i \in \mathbf{R}^{n_i}$ ,  $Q_i = \{x_i; B_i x_i \leq d_i\}$ ,  $d_i \in \mathbf{R}^{m_i}$ ,  $A_i$  — матрицы  $m \times n_i$ ,  $i = 1, \dots, s$ ,  $b \in \mathbf{R}^m$ . Всего в этой задаче  $M = m_1 + \dots + m_s + m$  ограничений и  $N = n_1 + \dots + n_s$  переменных. Задачи такого типа описываются объектами с относительно слабо связанными между собой блоками. Например,  $x_i$  могут относиться к  $i$ -му заводу в отрасли, а условие  $\sum_{i=1}^s A_i x_i = b$  задает глобальные ограничения на ресурсы и производимую продукцию всех заводов. Для решения подобных задач целесообразно применять *методы декомпозиции*, в которых последовательно решаются вспомогательные задачи для отдельных блоков. Существует много подобных методов. В одном из них строится функция Лагранжа (только для глобальных ограничений)

$$L(x, y) = \sum_{i=1}^s (c^i, x_i) + \left( y, \sum_{i=1}^s A_i x_i - b \right), \quad x \in \mathbf{R}^N, \quad y \in \mathbf{R}^m, \quad (17)$$

*и. двойственная функция*

$$\Psi(y) = \min_{\substack{x_i \in Q_i \\ i=1, \dots, s}} L(x, y) = \sum_{i=1}^s \min_{x_i \in Q_i} (c^i + A_i^T y, x_i) - (y, b). \quad (18)$$

На основании теоремы двойственности (см. § 1 гл. 10) задача (16) сводится к отысканию  $\max_{y \in \mathbf{R}^m} \Psi(y)$ ,

т. е. к  $m$ -мерной задаче безусловной максимизации кусочно-линейной вогнутой функции. Для ее решения можно применять методы гл. 5 (например, субградиентный метод или методы с растяжением пространства), при этом для вычисления  $\Phi(y^k)$ ,  $\partial\Phi(y^k)$  нужно решить  $s$  задач линейного программирования для блоков:

$$x_i^k = \arg \min_{x_i \in Q_i} (c^i + A_i^T y^k, x_i), \quad (20)$$

тогда (ср. с (20) § 3 гл. 10)

$$\Psi(y^k) = \sum_{i=1}^s (c^i + A_i^T y^k, x_i^k) - (y^k, b), \quad (21)$$

$$\partial\Psi(y^k) = \sum_{i=1}^s A_i x_i^k - b. \quad (22)$$

Ряд других итерационных методов линейного программирования (см. § 3 гл. 10) применительно к задаче (16) также обладает декомпозиционной структурой и потому позволяет решать задачи большой размерности. Отметим еще, что методы декомпозиции могут рассматриваться не только как численные процедуры отыскания оптимума, но и как механизмы итеративного планирования. В частности, описанный выше метод допускает интерпретацию как способ достижения оптимального плана с помощью подбора цен.

#### Упражнение.

1. Для задачи (15) берется произвольная пара  $(i_0, j_0)$ . Если  $a_{i_0} < b_{j_0}$ , то строится  $x_{i_0 j_0} = a_{i_0}$ , ограничение  $\sum_{j=1}^n x_{i_0 j} = a_{i_0}$  вычеркивается, а  $b_{j_0}$  заменяется на  $b_{j_0} - a_{i_0}$ . Если же  $b_{j_0} < a_{i_0}$ , то  $x_{i_0 j_0} = b_{j_0}$ , ограничение с  $b_{j_0}$  вычеркивается, а  $a_{i_0}$  заменяется на  $a_{i_0} - b_{j_0}$ . Этот процесс продолжается, пока не будут вычеркнуты все ограничения. Докажите, что полученный в результате вектор  $x = \{x_{ij}\}$  является вершиной (15) и все вершины могут быть заданы подобным образом.

#### 4. Оптимизация в условиях неопределенности.

Во всех рассмотренных выше примерах предполагалось, что имеется полное описание объекта и условий его деятельности. Однако в ряде случаев мы не обладаем столь исчерпывающей информацией, да она и невозможна в силу наличия неизбежной неопределенности и случайных факторов. Так, при планировании сельскохозяйственного производства мы сталкиваемся с неопределенностью метеорологических условий, при оптимизации режима электростанций — со случайным характером спроса на электроэнергию, при планировании запасов — с неопределенностью требований и т. д. Такого рода задачами, включающими случайность, называется *стochasticкое программирование*. В простейшей

постановке (называемой иногда задачей *перспективного стохастического программирования*) решение должно приниматься заранее, до производства испытаний; этим решением является  $n$ -мерный вектор; вся неопределенность носит вероятностный характер, а оптимум и ограничения должны выполняться в среднем. С математической точки зрения задача имеет вид

$$\min f(x), \quad f(x) = MQ_0(x, \omega) = \int Q_0(x, \omega) d\mathbf{P}(\omega),$$

$$g_i(x) \leqslant 0,$$

$$g_i(x) = MQ_i(x, \omega) = \int Q_i(x, \omega) d\mathbf{P}(\omega), \quad i = 1, \dots, m, \quad (23)$$

$$x \in S \subset \mathbf{R}^n.$$

Иначе говоря, задача сводится к детерминированной, в которой целевая функция и ограничения имеют вид математических ожиданий. Степень информированности о задаче может быть различной. Если функции  $Q_i(x, \omega)$  и распределение  $\mathbf{P}(\omega)$  известны, то в принципе задача эквивалентна обычной детерминированной задаче, но каждое вычисление  $f(x)$  и  $g_i(x)$  и их градиентов требует вычисления интегралов и потому весьма трудоемко. Поэтому даже в этом случае нередко удобно поступать так же, как при неизвестном распределении  $\mathbf{P}(\omega)$ . Именно, если задана лишь выборка  $\omega^1, \dots, \omega^k$  из этого распределения, то можно (как в методе Монте-Карло) аппроксимировать  $MQ(x, \omega)$  выражением

$$\frac{1}{k} \sum_{i=1}^k Q(x, \omega^i) \quad \text{и решать получающуюся детерми-}$$

нированную задачу. Другой способ (который особенно естествен

в ситуации, где реализации  $\omega^i$  поступают последовательно) основан на более простой замене  $f(x)$  и  $\nabla f(x)$  (аналогично и  $g_i(x)$ ,  $\nabla g_i(x)$ ) в точке  $x^k$  на  $Q_0(x^k, \omega^k)$  и  $\nabla Q_0(x^k, \omega^k)$ . Такие методы можно рассматривать как методы оптимизации при наличии случайных помех (см. гл. 4), они называются методами *стохастической аппроксимации* или *аддитивными методами стохи-*

*стической оптимизации*. Мы уже рассматривали подобный подход для задач оценки параметров (см. § 1).

В стохастическом программировании имеется ряд других постановок задач, отличных от (23). Так, решение может выбираться не априори, а уточняться в процессе наблюдений (*многоэтапные задачи стохастического программирования*). В других ситуациях само решение должно носить случайный характер, т. е. искомым является не конечномерный вектор, а закон распределения случайной величины. Такие постановки приводят к бесконечномерным задачам оптимизации.

Наконец, неопределенность в задачах оптимизации не обязательно сводится к случайным факторам. Так, оптимизируемая

функция  $F(x, u)$  может зависеть от некоторого параметра  $u$ , который не является случайнym, но не известен нам (известно лишь, что  $u \in U$ , где  $U$  — некоторое множество в  $\mathbf{R}^m$ ). Задача оптимизации в такой ситуации допускает различные постановки. Наиболее распространенным является *минимаксный подход*, т. е. расчет на наихудшее значение параметра. Тогда задача сводится к следующей:

$$\min f(x), \quad f(x) = \max_{u \in U} F(x, u) \quad (24)$$

(для простоты предполагаем, что ограничения на  $x \in \mathbf{R}^n$  отсутствуют). В частности, если  $u$  может принимать лишь конечное число значений, получаем *простейшую задачу на минимакс*:

$$\min f(x), \quad f(x) = \max_{1 \leqslant i \leqslant m} F_i(x). \quad (25)$$

К решению задач на минимакс можно подойти двояко.

Во-первых, (24) можно рассматривать как задачу безусловной минимизации негладкой функции ( $f(x)$  вида (25), как правило, нелинейно-дифференцируема, даже если  $F_i(x)$  гладкие; так, если  $F_i(x)$  аффинны:  $F_i(x) = (a^i, x) - b_i$ , то  $f(x)$  — кусочно-линейная, т. е. заранее негладкая функция). При этом, если  $F_i(x)$  выпуклы по  $x$ , то, используя правило вычисления субградиента (лемма 11 § 1 гл. 5), легко найти субградиент  $f(x)$  вида (25). Тем самым для минимизации  $f(x)$  можно применять методы, описанные в гл. 5. Для функций вида (24) вычисляется  $\varepsilon$ -субградиент (лемма 13 § 1 гл. 5), при этом нет нужды отыскивать максимум по  $u \in U$  точно. Поэтому применим метод  $\varepsilon$ -субградиента (11) § 3 гл. 5.

Во-вторых, задачу (25) можно свести к задаче нелинейного программирования, введя дополнительную переменную  $t \in \mathbf{R}^1$ :  $\min t, \quad F_i(x) \leqslant t, \quad i = 1, \dots, m$ .

Задача (26) имеет ряд особенностей — в ней всегда удовлетворяется условие Слейтера; нет проблем с выбором допустимой точки (достаточно взять произвольное  $x^0$  и  $t^0 \geqslant \max_{1 \leqslant i \leqslant m} F_i(x^0)$ ); переменная  $t$  исключается из вспомогательных задач для большинства методов. Поясним последнее замечание на примере методов, основанных на функции Лагранжа. Введем

$$L(x, t, y) = t + \sum_{i=1}^m y_i (F_i(x) - t). \quad (27)$$

Очевидно, что  $\inf_{t \in \mathbf{R}^1} L(x, t, y) = -\infty$ , если  $\sum_{i=1}^m y_i \neq 1$ . Поэтому

Метод (16) § 3 гл. 9 приобретает вид

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \hat{L}(x, y^k), \quad \hat{L}(x, y) = \sum_{i=1}^m y_i F_i(x),$$

$$y^{k+1} = P_S(y^k + \gamma F(x^{k+1})), \quad S = \left\{ y: y \geq 0, \sum_{i=1}^m y_i = 1 \right\}, \quad (28)$$

$$F(x) = \{F_1(x), \dots, F_m(x)\}.$$

Аналогичным образом модифицируются и другие методы нелинейного программирования применительно к задаче (26).

#### Упражнения

2. Напишите условия экстремума для задачи (25) при выпуклых  $F_i(x)$  двумя способами: а) из условия  $0 \equiv \partial f(x^*)$  (см. § 2 гл. 5); б) из теоремы Куна — Таккера для задачи (26).

3. Сформулируйте условия, при которых в задаче (25) достигается острый минимум (см. (9) § 2 гл. 5 и (11) § 2 гл. 9).

**5. Экстремальное регулирование.** Теория экстремальных систем выросла из классической теории автоматического регулирования, связанной с проблемами установления равновесия. Рассматривается система, действие которой описывается некоторым показателем качества. Таковым может быть коэффициент полезного действия (для энергетических установок), производительность совокупности машин (для технологических процессов), выход полезного продукта (для химических реакторов) и т. д. Значение показателя качества зависит от ряда параметров системы. Целью управления является выбор этих настраиваемых параметров, дающих экстремум показателя качества. Особенность задачи в том, что аналитическое описание зависимости показателя качества от параметров отсутствует. Таким образом, теория экстремального регулирования связана с оптимизацией действующих объектов, адекватная модель которых неизвестна в реальном масштабе времени. В силу специфики реальных задач (наличие шумов) выход объекта является случайной величиной.

С математической точки зрения эти задачи могут рассматриваться как задачи минимизации функции, единственной информацией о которой являются ее значения, вычисленные со случайными ошибками. Методы решения подобных задач были описаны в § 4 гл. 4. Следует однако отметить, что задачи экстремального регулирования имеют особенности. Во-первых, поскольку оптимизация проводится на реальном объекте, важно получать малые значения функции не только в конце вычислений, но и в ходе всего итерационного процесса. Поэтому, в частности, нельзя делать большие шаги, которые могут привести к возрастанию функции. Во-вторых, регулируемый объект, как правило, бывает нестационарным. В силу неизбежных колебаний неконтролируемых параметров и дрейфа характеристик оптимизируемая функция меняется во времени, т. е. мы нахо-

димся в условиях, описанных в § 3 гл. 6. Наконец, оптимизация ведется в непрерывном (а не дискретном) времени, а регулируемый объект обладает инерционностью. Все это приводит к тому, что применяемые на практике способы экстремального регулирования заметно отличаются от стандартных алгоритмов безусловной минимизации.

**6. Оптимальное управление.** Другой важный класс экстремальных задач, возникших из проблем автоматического регулирования, — это задачи оптимального управления. В них требуется найти закон изменения управляющих воздействий для динамической системы так, чтобы минимизировать некоторый критерий оптимальности. При этом предполагается, что система описывается обыкновенными дифференциальными уравнениями, которые известны (равно как и критерий оптимальности). Оптимальное управление нашло широкое применение в задачах выбора траекторий летательных и космических аппаратов, выбора режимов работы электрических машин и химических реакторов и т. д. Основные достижения в теории оптимального управления относятся к условиям оптимальности, формулируемым в виде принципа максимума Понтиягина или принципа оптимальности Белмана. Большие успехи достигнуты и в развитии численных методов оптимального управления. Мы, однако, не имеем возможности останавливаться на этих важных результатах, так как данная книга посвящена лишь конечномерным задачам, и отсылаем читателя к обширной литературе по оптимальному управлению.

Опишем лишь *дискретный вариант задачи оптимального управления*:

$$\min \left[ F(x_N) + \sum_{i=0}^{N-1} f_i(x_i, u_i) \right],$$

$$x_{i+1} = \Phi_i(x_i, u_i), \quad i = 0, 1, \dots, N-1, \quad x_0 = a,$$

$$u_i \in U_i, \quad i = 0, \dots, N-1, \quad x_i \in X_i, \quad i = 1, \dots, N.$$

Здесь  $x_i \in \mathbf{R}^n$ ,  $i = 0, \dots, N$ , — состояния процесса или фазовые переменные,  $u_i \in \mathbf{R}^m$ ,  $i = 0, \dots, N-1$ , — управление или управляемые переменные, соотношения  $x_{i+1} = \Phi_i(x_i, u_i)$  называются уравнениями состояния и задают траекторию процесса  $\{x_0, x_1, \dots, x_N\}$ ,  $u_i \in U_i$  и  $x_i \in X_i$  — ограничения на управления и состояния соответственно. Частными случаями задачи (29) являются: задача оптимизации конечного состояния (терминальная задача) без ограничений на состояния:

$$\min F(x_N),$$

$$x_{i+1} = \Phi_i(x_i, u_i), \quad i = 0, 1, \dots, N-1, \quad x_0 = a, \quad (30)$$

$$u_i \in U_i, \quad i = 0, \dots, N-1,$$

*дискретная аппроксимация классической задачи вариационного исчисления:*

$$\min \sum_{i=0}^{N-1} F_i(x_i, u_i),$$

$$x_{i+1} = x_i + \varepsilon u_i, \quad x_i \in \mathbf{R}^1, \quad u_i \in \mathbf{R}^1, \quad i = 0, \dots, N-1, \quad (31)$$

$$x_0 = a, \quad x_N = b,$$

$$\text{где } p_i^* - \text{решение сопряженной системы}$$

и ряд других. К анализу задачи (29) можно подходить с различными точек зрения. Во-первых, можно рассматривать векторы  $x_i, u_i$  как независимые переменные, а уравнения состояния — как связывающие их ограничения типа равенства. Такой прямолинейный подход обычно не очень эффективен, так как получается задача с большим числом переменных и ограничений, а специфическая структура задачи (29) практически не используется. Во-вторых, можно считать независимыми переменными лишь  $u_i$ , а  $x_i$  выражать через них с помощью уравнений состояния и начального условия для  $x_0$ . Так, задача (30) записывается в виде

$$\begin{aligned} \min f(u), \quad u = \{u_0, \dots, u_{N-1}\} \in \mathbf{R}^{Nm}, \\ u \in U, \quad U = U_0 \times \dots \times U_{N-1}, \end{aligned} \quad (32)$$

где  $f(u) = F(x_N)$ , а  $x_N = x_N(u)$  находится рекуррентно из соотношений  $x_{i+1} = \Phi_i(x_i, u_i)$ ,  $i = 0, \dots, N-1$ ,  $x_0 = a$ .

Пусть функции  $F(x_N)$ ,  $\Phi_i(x_i, u_i)$  дифференцируемы по своим аргументам; вычислим градиент  $f(u)$ . Пусть  $\tilde{u} = \{\tilde{u}_0, \dots, \tilde{u}_{N-1}\}$  — приращение управления, тогда  $\tilde{x}$  — линейная часть приращения состояний — описывается уравнениями

$$\tilde{x}_{i+1} = \Phi'_x(x_i, u_i)\tilde{x}_i + \Phi'_u(x_i, u_i)\tilde{u}_i, \quad i = 0, \dots, N-1, \quad \tilde{x}_0 = 0.$$

*Введем сопряженную систему*

$$p_i = \Phi'_x(x_i, u_i)^T p_{i+1}, \quad p_i \in \mathbf{R}^n, \quad i = 0, \dots, N-1, \quad p_N = \nabla F(x_N).$$

Тогда

$$\begin{aligned} (p_{i+1}, \tilde{x}_{i+1}) &= (p_{i+1}, \Phi'_x(x_i, u_i)\tilde{x}_i + \Phi'_u(x_i, u_i)\tilde{u}_i) = \\ &= (p_i, \tilde{x}_i) + (\Phi'_u(x_i, u_i)^T p_{i+1}, \tilde{u}_i). \end{aligned}$$

Суммируя эти равенства, получаем

$$(VF(x_N), \tilde{x}_N) = \sum_{i=0}^{N-1} (\Phi'_u(x_i, u_i)^T p_{i+1}, \tilde{u}_i).$$

Отсюда следует, что

$$\nabla f(u) = \{\Phi'_u(x_0, u_0)^T p_1, \dots, \Phi'_u(x_{N-1}, u_{N-1})^T p_N\}. \quad (33)$$

Используя необходимое условие минимума  $f(u)$  на  $U$  (теорема 1 § 1 гл. 7), получаем, что если  $U_0, \dots, U_{N-1}$  выпуклы, управление  $u^*$  является оптимальным для (30), а  $x^{*1}$  — соответствующая ему траектория, то должно выполняться условие (*локальный принцип максимума*):

$$(\Phi'_u(x_i^*, u_i^*)^T p_{i+1}^*, u_i - u_i^*) \geq 0 \quad \forall u_i \in U_i, \quad (34)$$

$$p_i^* = \Phi'_x(x_i^*, u_i^*)^T p_{i+1}^*, \quad i = 0, \dots, N-1, \quad p_N^* = \nabla F(x_N^*). \quad (35)$$

Используя вид  $\nabla f(u)$  (33), нетрудно выписать итеративные методы решения задачи (32), описанные в § 2 гл. 7. Так, в методе проекции градиента последовательность приближений  $u^k$  из строится следующим образом. Определяется траектория  $x^k$  из уравнений состояния

$$x_{i+1}^k = \Phi_i(x_i^k, u_i^k), \quad i = 0, \dots, N-1, \quad x_0^k = a,$$

рекуррентно (от  $N$  до 0) вычисляются

$$p_i^k = \Phi'_x(x_i^k, u_i^k)^T p_{i+1}^k, \quad i = N-1, \dots, 0, \quad p_N^k = \nabla F(x_N^k)$$

и затем определяются  $u^{k+1}$ :

$$u_i^{k+1} = P_{U_i}(p_i^k - \gamma \Phi'_u(x_i^k, u_i^k)^T p_{i+1}^k).$$

Описанный подход к задаче (30) прост и эффективно использует ее особенности. Однако существенные трудности для него представляют фазовые ограничения. При их наличии удобен следующий прием. Предположим, что в задаче (29) отсутствуют ограничения на управление, а уравнение  $x_{i+1} = \Phi_i(x_i, u_i)$  однозначно разрешимо относительно  $u_i \in \mathbf{R}^m$  при любых  $x_i, x_{i+1} \in \mathbf{R}^n$ . Отметим, что эта ситуация нетипична, так как обычно  $m < n$ , однако в задаче (31) мы имеем именно такой случай. Тогда, исключив управления с помощью уравнений состояния, приходим к задаче вида

$$\min \sum_{i=0}^{N-1} \Phi_i(x_i, x_{i+1}), \quad (36)$$

$$x_i \in X_i, \quad i = 1, \dots, N, \quad x_0 = a.$$

Для решения последней задачи можно применять как общие методы типа проекции градиента (они упрощаются благодаря виду ограничений в (36)), так и специальные. В частности, в методе динамического программирования вводятся функции  $V_i(x_i)$ , связанные рекуррентным соотношением

$$V_{i+1}(x_{i+1}) = \min_{x_i \in X_i} [\Phi_i(x_i, x_{i+1}) + V_i(x_i)], \quad i = 1, \dots, N-1,$$

$$V_1(x_1) = \Phi_0(a, x_1), \quad (37)$$

при этом минимум  $V_N(x_N)$  по  $x_N \in X_N$  дает решение (ср. с (14)). С другой стороны, для минимизации (36) удобно применять методы типа покоординатного спуска (т. е. последовательно проводить минимизацию по  $x_i \in X_i$ ,  $i = 1, \dots, N$ , при фиксированых значениях остальных переменных). Действительно, при этом вычисляются значения лишь двух функций  $\Phi_{i-1}(x_{i-1}, x_i)$  и  $\Phi_i(x_i, x_{i+1})$  в (36), а ограничения на каждую переменную  $x_i \in X_i$  задаются независимо. Такой метод называется *методом локальных вариаций*.

#### Упражнения.

4. Запишите задачу распределения ресурсов (8) в форме задачи оптимального управления.

5. Покажите, что если  $\Phi_i(x_i, u_i) = Ax_i + Bu_i$ ,  $F(x_N)$  выпукла и  $U_i$  выпуклы, то (34) является достаточным условием экстремума.

6. Найдите явный вид решения задачи (31) с  $F_i(x_i, u_i) = \alpha_i x_i^2 + \beta_i u_i^2$ ,  $\alpha_i \geq 0$ ,  $\beta_i \geq 0$ : а) из условий экстремума, б) с помощью метода динамического программирования. мин  $(c_w, x_N)$ ,  $x_{i+1} = Ax_i + Bu_i$ ,  $x_i \in X_i$ ,  $u_i \in U_i$ ,  $x_0 = a$ , где  $X_i$ ,  $U_i$  — многогранные множества.

### § 3. Задачи оптимизации в математике и физике

Задачи на экстремум издавна рассматривались в математике, задолго до развития общей теории оптимизации. Аналогичным образом в механике, оптике и других разделах физики традиционно исследовались задачи, связанные с вариационными принципами. Ниже приводится несколько примеров.

**1. Задачи наилучшего приближения.** Пусть имеется определенная система линейных уравнений

$$(a^i, x) = b_i, \quad i = 1, \dots, m, \quad x \in \mathbf{R}^n.$$

Такая система может и не иметь решения (особенно при  $m > n$ ), поэтому естественно ставить задачу о минимизации какой-либо нормы от невязки. В частности,

$$\min \sum_{i=1}^m |(a^i, x) - b_i|, \quad (2)$$

$$\min \sum_{i=1}^m ((a^i, x) - b_i)^2, \quad (3)$$

$$\min \max_{1 \leq i \leq m} |(a^i, x) - b_i|. \quad (4)$$

являются соответственно задачами *наилучшего в норме  $l_1$* , *наилучшего среднеквадратичного* и *наилучшего равномерного* (или *чебышевского*) приближения. Отметим, что такого рода задачи возникают и в прикладных проблемах идентификации (см. § 1).

Покажем, как общая теория может быть применена для анализа задач (2)–(4). Прежде всего, используя теорему 9 § 1 гл. 10 и теорему 4 § 4 гл. 10, получаем, что задачи (2)–(4) всегда имеют решение. Нормальное решение (см. § 1 гл. 6) задачи (3) выписывается явно:

$$x^* = A^+ b,$$

где  $A$  — матрица  $m \times n$  со строками  $a^i$ ,  $A^+$  — псевдообратная к ней,  $b = (b_1, \dots, b_m)$ . Получим условия экстремума в задаче (4). Используя лемму 11 § 1 гл. 5 и теорему 1 § 2 гл. 5, убеждаемся, что точка  $x^*$  с  $f(x^*) \geq 0$ ,  $f(x) = \max_{1 \leq i \leq m} |(a^i, x) - b_i|$ , является решением (4) тогда и только тогда, когда 0 принадлежит выпуклой оболочке векторов  $a_i a^i$ ,  $i \in I^*$ , где  $I^* = \{i: |(a^i, x^*) - b_i| = f(x^*)\}$ ,  $a_i = \text{sign}((a^i, x^*) - b_i)$ . Тогда по лемме Каратеодори (лемма 1 § 1 гл. 5) 0 является выпуклой комбинацией  $n+1$  из этих векторов. Итак, необходимое и достаточное условие экстремума принимает вид: найдутся  $n+1$  чисел  $\lambda_i \geq 0$ ,  $i \in I^*$ , таких, что

$$\sum_{i \in I^*} \lambda_i a_i a^i = 0, \quad \sum_{i \in I^*} \lambda_i = 1. \quad (6)$$

Наряду с дискретными задачами (2)–(4) можно рассматривать и их *непрерывные аналоги*. Именно, пусть на отрезке  $[0, 1]$  задана функция  $b(t)$  и система функций  $a_1(t), \dots, a_m(t)$ . Задача наилучшего приближения  $b(t)$  по данной системе соответственно в нормах  $L_1$ ,  $L_2$ ,  $L_\infty$  имеет вид

$$\min \int_0^1 \left| b(t) - \sum_{i=1}^m x_i a_i(t) \right| dt, \quad (7)$$

$$\min \int_0^1 \left( b(t) - \sum_{i=1}^m x_i a_i(t) \right)^2 dt, \quad (8)$$

$$\min_{0 \leq t \leq 1} \left| b(t) - \sum_{i=1}^m x_i a_i(t) \right|. \quad (9)$$

Здесь уже нельзя столь просто и без всяких дополнительных предположений о  $b(t)$  и  $a_i(t)$  получить результаты о существовании решения и условия экстремума. Однако такой анализ может быть проведен. В частности, используя обобщение леммы 11 § 1 гл. 5 на случай бесконечного числа функций (вил  $f(x) = \max_{0 \leq t \leq 1} f(x, t)$ ) можно получить классическую теорему Чебышева о наилучшем равномерном приближении для случая, когда  $a_i(t)$  — полином степени  $i-1$ .

Мы упомянули лишь о простейших задачах наилучшего приближения. Целью было показать связь между теорией приближений и теорией оптимизации. Разумеется, задачи приближения имеют специфику, и для их исследования создан весьма тонкий математический аппарат. Тем не менее упомянутая связь полезна как при анализе условий наилучшего приближения, так и особенно при разработке численных методов для задач аппроксимации.

#### Упражнение.

1. Рассмотрите дискретную задачу наилучшего равномерного приближения полиномами, т. е. задачу (4) с  $a^i = \{1, t_i, \dots, t_i^{n-1}\}$ ,  $0 \leq t_i \leq \dots \leq t_m \leq 1$ . Получите дискретный аналог теоремы Чебышева: чтобы величина  $\|x^*\|$  было решением, необходимо и достаточно, чтобы величины  $|e_i|$  (где  $e_i =$

$$= \sum_{j=1}^n x_j^* t_j^{i-1} - b_i$$
) достигали максимума не менее чем в  $n+1$  точках, а знаки  $e_i$  в этих точках чередовались. (Указание. Рассмотрите (6) как систему уравнений относительно  $n+1$  переменных  $x_i = \lambda_i e_i$ ; найдите решение и воспользуйтесь свойством определителя Вандермонда менять знак при перестановке двух столбцов.)

**2. Геометрические задачи на экстремум.** В геометрии возникает множество задач на максимум и минимум, отличающихся разнообразием формулировок и сложностью решения. Одна группа задач связана с нахождением тел (заданных с точностью до параметров), имеющих минимальный объем, поверхность или другие характеристики. С одной такой задачей мы столкнулись при построении метода эллипсоидов (см. § 4 гл. 5). Требовалось найти эллипсоид наименьшего объема, описанный вокруг полушара. Задача без труда сводится к минимизации функции одного переменного и может быть решена явно.

Другой тип задач возникает, когда вид тела, обладающего тем или иным экстремальным свойством, не фиксирован. Например: найти радиус минимального шара, внутри которого может быть помещено любое тело диаметра 1 (ответ дается теоремой Юнга), или найти радиус максимального шара, который может быть помещен внутри произвольного выпуклого тела шириной 1 (ответ дается теоремой Банаха) и т. д. Примером решения экстремальной задачи такого типа может служить лемма 1 § 4 гл. 5, понадобившаяся нам в связи с обоснованием метода центра тяжести.

Сюда же примыкают так называемые изоллерметрические задачи, в которых требуется оптимизировать одну из геометрических характеристик тела (например, объем) при фиксированных значениях других характеристик (например, поверхности). Такова задача Диодона об отыскании фигуры наибольшей пло-

щади в  $\mathbf{R}^2$ , образованной дугой длины 1 с концами на прямой. Подобные задачи рассматриваются в вариационном исчислении и являются бесконечномерными.

Наконец, существует множество задач типа наименее плотной упаковки тел, нахождения минимальной  $\varepsilon$ -сети и т. п. Им присущ комбинаторный характер, и они обычно очень трудны.

Не останавливаясь на всех этих разнообразных задачах, приведем два примера, демонстрирующих плодотворность использования теории оптимизации для анализа геометрических проблем.

Первая задача в исходной постановке никак не связана с оптимизацией. Именно, попробуем доказать *теорему Хелли*: если  $A_i$ ,  $i = 1, \dots, m$ , — выпуклые множества в  $\mathbf{R}^n$ ,  $A_1$  ограничено и пересечение любых  $n+1$  множеств из  $\{A_i\}$  непусто, то все множества имеют общую точку. Введем функцию  $f(x) = \max_{1 \leq i \leq m} \rho(x, A_i)$ , где  $\rho(x, A_i)$  — расстояние от  $x$  до  $A_i$ . Функция  $\rho(x, A_i)$  и  $f(x)$  выпуклы (см. упр. 2 § 1 гл. 5); множество  $\{x: f(x) \leq \alpha\} \subset \{x: \rho(x, A_i) \leq \alpha\}$ , а потому ограничено для любого  $\alpha \geq 0$ . Следовательно (теорема 3 § 2 гл. 5),  $f(x)$  достигает минимума на  $\mathbf{R}^n$  в некоторой точке  $x^*$ , поэтому (теорема 1 § 2 гл. 5)  $0 \in \partial f(x^*)$ . Используя вид субградиента  $f(x)$  (лемма 11 § 1 гл. 5) и лемму Каратеодори (лемма 1 § 1 гл. 5), получаем, что найдутся  $n+1$  индексов (для определенности первые  $n+1$ ), для которых  $\rho(x^*, A_i) = f(x^*)$ , и чисел  $\lambda_i \geq 0$ ,  $i = 1, \dots, n+1$ ,

и лемму Каратеодори (лемма 1 § 2 гл. 5), получаем, что найдутся  $n+1$  индексов (для определенности первые  $n+1$ ), для которых  $\rho(x^*, A_i) = f(x^*)$ , и чисел  $\lambda_i \geq 0$ ,  $i = 1, \dots, n+1$ ,  $\sum_{i=1}^{n+1} \lambda_i = 1$ , таких, что  $\sum_{i=1}^{n+1} \lambda_i \partial \rho(x^*, A_i) = 0$ . Но это — достаточное условие безусловного минимума (теорема 1 § 2 гл. 5) функции  $\Phi(x) = \max_{1 \leq i \leq n+1} \rho(x, A_i)$ . Поскольку  $\bigcap_{i=1}^{n+1} A_i \neq \emptyset$  (по предположению теоремы), то  $\min_{x \in \mathbf{R}^n} \Phi(x) = 0$ , т. е.  $\Phi(x^*) = 0$ ,  $\rho(x^*, A_i) = 0$ ,  $i = 1, \dots, n+1$ . Но  $\rho(x^*, A_i) = f(x^*)$ ,  $i = 1, \dots, n+1$ , поэтому  $f(x^*) = 0$ , что и означает, что  $x^* \in A_i$ ,  $i = 1, \dots, n+1$ .

Конечно, теорема Хелли может быть доказана чисто геометрическими средствами, однако приведенное выше доказательство поучительно — оно демонстрирует возможность применения стандартной техники для решения нестандартных задач.

Рассмотрим теперь задачу Штейнера: найти точку в  $\mathbf{R}^n$ , для которой сумма расстояний до  $m$  заданных точек  $a^1, \dots, a^m$  минимальна. Иными словами, ищется

$$\min f(x), \quad f(x) = \sum_{i=1}^m \|x - a^i\|. \quad (10)$$

То, что здесь минимум достигается и единствен (если только не все  $a^i$  лежат на одной прямой), уже отмечалось (в упр. 4, 5 § 2 гл. 5). Если точка минимума  $x^*$  не совпадает ни с одной из

точек  $a^i$ , то  $f(x)$  дифференцируема в  $x^*$  (упр. 1 § 1 гл. 1) и (теорема 1 § 2 гл. 1)

$$\nabla f(x^*) = \sum_{i=1}^m e^i = 0, \quad e^i = \frac{x^* - a^i}{\|x^* - a^i\|}. \quad (11)$$

В некоторых случаях это уравнение можно решить. Пусть, например,  $n = 2$ ,  $m = 3$ , т. е. ищется точка, наименее удаленная от вершин треугольника. Из (11) следует, что  $e^1 + e^2 + e^3 = 0$ ,  $\|e^i\| = 1$ , это возможно лишь тогда, когда векторы  $e^i$  образуют между собой равные углы. Итак, точка  $x^*$  такова, что отрезки, соединяющие ее с вершинами, образуют угол  $120^\circ$  (такая точка называется *точкой Торричелли*, рис. 43, а). Если такой точки не

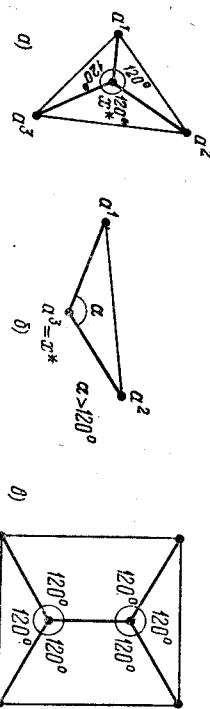


Рис. 43. Задача Штейнера.

найдется (последнее произойдет, если один из углов в треугольнике не менее  $120^\circ$ ), то минимум  $f(x)$  достигается в соответствующей вершине (рис. 43, б). В общем случае для минимизации (10) нужно применять численные методы. При этом в окрестности минимума  $f(x)$  может быть как гладкой (если  $x^*$  не совпадает с  $a^1, \dots, a^m$ ), так и негладкой (если  $x^* = a^i$ ). Поэтому здесь следует разумно комбинировать методы гладкой и негладкой минимизации.

**Упражнение.**  
2. Пусть  $A_1, \dots, A_m$  — выпуклые замкнутые множества в  $\mathbb{R}^n$  и  $A = A_1 \cap \dots \cap A_m \neq \emptyset$ . Докажите, что метод последовательного проектирования вида  $x_{k+1} = P_k(x^*)$ , где  $P_k$  — проектор на наиболее удаленную от  $x^*$  множества, сходится к некоторой точке из  $A$ . (Указание. Рассмотрите метод как субградиентный метод вида (7) § 3 гл. 5 для минимизации  $f(x) = -\max_{1 \leq i \leq m} \rho(x, A_i)$ .)

**3. Вариационные принципы физики.** Во всех задачах § 2 задачи оптимизации возникали в связи с возможностью выбора того или иного решения и наличием цели управления. Достойно удивления, что многие процессы и явления в природе, не будучи управляемыми, описываются с помощью вариационных принципов. Таковы, например, вариационные принципы механики

(принцип наименьшего действия, принцип минимума потенциальной энергии), оптики (принцип наименьшего времени Ферма), гидравлики и т. д. В силу этого можно рассматривать процессы установления равновесия в подобных системах как своеобразной природой. На первых шагах развития математического программирования делались попытки непосредственного использования электрических, гидравлических и тому подобных устройств для решения задач оптимизации. Ныне, с развитием цифровых вычислительных машин, подобные физические модели потеряли актуальность. Однако анализ физических аналогий может быть полезен при разработке численных методов оптимизации. С такой ситуацией мы уже сталкивались при описании метода тяжелого шарика в § 2 гл. 3. Приведенная там модель (движения тела в потенциальном поле при наличии силы трения) позволяла

сделать качественные выводы о поведении итеративного процесса (возможность ускорения за счет введения инерции, способность «проскакивать» небольшие локальные минимумы и т. д.). Конечно, эти выводы нуждаются в строгом математическом обосновании (типа теоремы 1 § 2 гл. 3), однако их эвристическая польза несомненна.

С другой стороны, ряд результатов математической теории оптимизации оказался полезен при изучении физических проблем. Так, например, некоторые варианты теоремы двойственности (в содержащих терминах) были известны в аналитической механике и теории электрических цепей; выяснение математической природы этих утверждений способствовало большей методологической ясности и позволило перенести эти утверждения на более общий класс объектов.

Особенно широкое применение для решения физических задач нашли численные методы оптимизации. Упомянем лишь о задачах расчета статически неопределенных упругих систем, о вариационных задачах равновесия жидкости, о расчетах упруго-пластических и вязко-пластических деформаций (особенностю последних задач является их негладкость) и т. д.

**Упражнение.**  
4. Свет распространяется в одной среде со скоростью  $v_1$ , в другой —  $v_2$ , граница раздела между средами прямолинейна. Выведите из принципа Ферма (свет между двумя точками распространяется по пути, претерпевшему наименьшего времени) закон преломления Снеллиуса:  $\sin \alpha_1 / \sin \alpha_2 = v_1 / v_2$ , где  $\alpha_1$  — угол падения,  $\alpha_2$  — угол преломления.

## Г л а в а 12

### ПРАКТИЧЕСКОЕ РЕШЕНИЕ ЗАДАЧ ОПТИМИЗАЦИИ

В этой главе мы остановимся на таких вопросах, связанных с практическим решением задач оптимизации, как методика формализации задач, критерии выбора метода, процесс численного решения и анализ его результатов. Здесь же приводятся сведения о тестовых задачах оптимизации, имеющихся программах и некоторых результатах вычислений.

#### § 1. Процесс решения

Несколько методических замечаний о постановке и решении задач оптимизации, содержащихся в этом параграфе, не претендуют на общность и весьма конспективны. Во-первых, практические задачи и ситуации слишком разнообразны, чтобы можно было охватить даже основные их типы. Во-вторых, процесс решения всегда остается искусством, которому можно научиться лишь путем проб и ошибок. Наконец, собственный опыт автора в этой области ограничен, и высказываемые ниже мнения во многом субъективны.

**1. Формализация задач.** Лишь в редких случаях имеется готовая математическая запись задачи оптимизации, предназначенной для решения. Обычно первым шагом является *четкая постановка* и формализация практической задачи. Этот этап должен осуществляться совместно специалистом в данной конкретной области («заказчиком») и математиком, который будет решать задачу («вычислителем»). Хорошо, когда оба эти лица совмещаются в одном; в противном случае процесс выработки общего языка может оказаться длительным. Вычислителю рекомендуется обязательно разобраться в задаче «по существу», даже если она принесена в готовом виде, так как без этого вероятность получения бессмысленного ответа очень велика.

При постановке задачи заказчик должен прежде всего четко понять сам, какова цель исследования, для чего предполагается использовать решение, доступна ли исходная информация и какова ее точность, какие предположения являются существенными, а такие нет, и т. д. Часто уже на этом этапе оказывается, что задачу, собственно, решать нельзя (из-за отсутствия данных) или незачем (так как ожидаемый выигрыш от оптимизации будет очень незначителен по сравнению с другими факторами).

Далее следует попытаться сформулировать *критерий оптимальности и ограничения*, на содержательном уровне. В практических задачах обычно возникает проблема многокритериальности — скажем, при проектировании нового устройства желательно, чтобы все его характеристики (производительность, цена, вес, надежность и т. п.) были оптимальными. Ясно, что одновременно это сделать невозможно. Поэтому следует либо выработать единый обобщающий критерий, либо задаться допустимыми значениями всех характеристик, кроме одной, а последнюю оптимизировать.

Следующим шагом является *математическая формализация задачи*. Здесь очень важно участие вычислителя — как с точки зрения привнесения общей математической культуры, так и для решения конкретных вопросов, связанных с возможностями численной реализации. На этом этапе следует выбрать независимые переменные и записать минимизируемую функцию и ограничения. При этом заказчик поясняет, какие из параметров и условий являются основными, а какими можно пренебречь; в какой мере правомерна линеаризация функций; какие содержательные ограничения должны быть учтены, а вычислитель пытается построить в этих условиях наиболее простую модель (по возможности линейную и с наименьшим числом переменных и ограничений). В результате вырабатывается некоторая предварительная математическая формализация задачи.

Эта формализация нуждается в проверке по существу (все ли в ней учтено, не слишком ли она груба) и *первоначальном математическом изучении*. Здесь важно обдумать, к какому классу относится задача, решались ли (и насколько успешно) подобные задачи ранее, какие вычислительные и программные средства имеются и т. д. Следует также прикинуть общие характеристики задачи (выпуклая она или нет, насколько гладки входящие в формулировку функции, есть ли опасность многоэкстремальности, насколько велики ошибки при вычислениях, сильно ли вырождена задача и т. п.). В связи с этим нередко приходится несколько изменять постановку задачи — скажем, выбирать новые переменные, в которых минимизируемая функция лучше обусловлена или выпукла, записывать одно негладкое ограничение как несколько гладких, переходить к двойственной задаче с меньшим числом переменных и т. п.

В построенной модели, как правило, не все величины бывают известны. Проблема получения необходимой *исходной информации* или нередко превращается в одну из сложнейших. Это особенно относится к экономическим задачам, где трудности извлечения достоверного и достаточного статистического материала очень велики. В технических задачах приходится решать вспомогательные задачи по идентификации искомых параметров; нужно обязательно представлять (хотя бы качественно) точность

получаемых при этом результатов. В результате всего этого сложного и «многошагового» процесса вырабатывается окончательная математическая формулировка задачи, подлежащей оптимизации.

**2. Выбор метода и программы.** Лишь в исключительных случаях (примеры которых приводились выше) решение задачи оптимизации может быть найдено в явной форме. Обычно экстремум приходится отыскивать численно с помощью ЭВМ. При выборе метода и его программной реализации следует учитывать в первую очередь имеющиеся возможности вычислительных средств, наличие соответствующих программ и количество однотипных задач, которые требуется решить.

Если решается единственная задача оптимизации и она не очень сложна, то вряд ли стоит тратить силы на выбор наилучшего метода и программы. Достаточно, чтобы в имеющемся математическом обеспечении ЭВМ нашлась программа, пригодная для решения задач подобного типа и объема. То, что время решения окажется несколько большим, чем для специально написанной программы (например, 1 час вместо 1 минуты), не так существенно. Аналогичным образом при отсутствии готовых программ для решения уникальной задачи разумно программировать простейший метод ее решения, пренебрегая его скоростью сходимости и прочими характеристиками.

Лишь при *массовом решении* однотипных задач (а такая ситуация обычно для задач оптимизации) вопрос о выборе способа численного решения стоит остро, поскольку проигрыш во времени счета или точности решения увеличивается многократно. Здесь имеет смысл опробовать различные имеющиеся программы на нескольких примерах, и если результаты окажутся неудовлетворительными (слишком медленная сходимость либо невозможность получить решение с заданной точностью), то следует составить специальную программу для решения данного класса задач. При выборе метода нужно учитывать *скорость сходимости, объем вычислений на каждой итерации, устойчивость к различного рода помехам, требуемый объем памяти ЭВМ*, т. е. именно те факторы, которые мы исследовали при теоретическом изучении методов в первых частях книги. Когда из подобных априорных соображений метод выбран, встает вопрос о его *алгоритмической и программной реализации*. Переход от методов на том уровне подробности, который принят в данной книге, к четкому алгоритму, отнюдь не прост. Нужно задать полпрограммы решения вспомогательных задач, параметры, определяющие точность решения этих задач, критерии прерывания счета и т. п. От этих малозначащих на идеином уровне подробностей в очень большой мере зависят эффективность алгоритма. Аналогичные проблемы возникают при написании программы. Выбор языка, на котором она составляется, организация массивов информа-

ции, способы реализации арифметических выражений — все это очень важные детали. Мы уже отмечали, в частности, что высокий уровень отработки подобных деталей в программах символьского метода послужил одной из основных причин успешной работы конечных методов линейного программирования при практическом решении задач.

После составления первого варианта программы и ее отладки начинается процесс *подгонки алгоритма*. Берется несколько характерных задач и делается попытка их решения. Она, как правило, оказывается неудовлетворительной; следует выяснить причины этого и попытаться устраниить их. С этой целью можно изменять параметры алгоритма, заменять отдельные подпрограммы (например, процедуру одномерной минимизации), проводить некоторые вычисления с двойной точностью и т. п. В лучшем случае эти попытки оказываются успешными и приводят к окончательному варианту программы, в худшем — выясняется, что выбранный метод не дает возможности решить задачу и следует обратиться к другому методу, учитывая неудачный опыт предыдущей попытки. Иногда следует остановиться на некоторой последовательной комбинации методов.

Разумеется, не всегда процесс подготовки к счету происходит подобным образом. Так, при решении задач очень большой размерности главная проблема связана с размещением информации в памяти ЭВМ, и этим в первую очередь определяется выбор метода и программы. При использовании *микропроцессоров* для целей оптимизации основным является вопрос о простоте метода и его устойчивости к погрешностям вычислений. Особые проблемы возникают при решении задач в *реальном масштабе времени*, здесь важно успеть обработать поступающую информацию и рекуррентно использовать найденные приближения. Новые возможности появляются при работе в *диалоговом режиме* — в частности, резко упрощается настройка параметров алгоритма, удается исследовать влияние изменений условий задачи на ход решения, можно осуществлять последовательную проверку различных методов и т. п.

**3. Анализ результатов.** После выбора программы и проведения счета процесс решения не является завершенным. Во-первых, полезно оценить *точность найденного решения*. В ряде случаев это можно сделать теоретически — используя теорему двойственности, проверяя невязку в выполнении условия оптимальности, применяя оценки типа теоремы 2 § 5 гл. 1 и т. д. Возможна и экспериментальная проверка точности. Так, пециально необходимо осуществить спуск из разных начальных точек и посмотреть, насколько разные результаты получаются. Во-вторых, обязательно следует провести *содержательный анализ найденного решения*, учитывая, что математическая формулировка была получена в результате ряда упрощений и огрублений

исходной задачи. В итоге может оказаться, что хотя найденная точка и является решением математической модели, она не удовлетворяет заказчика, так как не все факторы были учтены в исходной постановке. Нередко так и случается с экономическими задачами, что приводит к сравнительно низкому уровню использования оптимационных моделей в экономической практике. В такой ситуации следует пересмотреть исходную модель и внести в нее изменения, после чего надо заново решить задачу оптимизации.

## § 2. Программы оптимизации

**1. Общие требования к программам.** Разумеется, требования к программам оптимизации зависят прежде всего от назначения и характера использования этих программ. Здесь можно выделить несколько типичных ситуаций.

Если программа предназначена для однократного решения конкретной задачи, то она может быть произвольной, лишь бы позволяла достичь требуемой цели. В случае, когда предполагается многократное решение однотипных задач одним пользователем, следует обратить особое внимание на экономию времени счета. В частности, заметного выигрыша можно добиться, если писать программу на автокоде, а не на алгоритмическом языке. Однако в остальном подобные программы «для внутреннего пользования» не обязаны подчиняться каким-либо жестким ограничениям.

Серьезные требования предъявляются к стандартным программам, предназначенным для решения типовых задач многими пользователями. Эти программы должны допускать использование на различных ЭВМ (и потому пишутся на алгоритмических языках, чаще всего на Фортране); предоставлять определенные удобства пользователю (особенно это важно в задачах линейного программирования, связанных с большими объемами информации, где требуется иметь возможность формировать и изменять эту информацию различными способами); допускать использование в качестве подпрограмм (например, безусловная минимизация часто возникает как вспомогательная задача при решении более общей задачи оптимизации); представлять результаты вычислений в удобном и достаточно подробном виде и т. д. Стандартные программы должны сопровождаться описаниями и тестовыми примерами. Пользователь не обязан быть хорошо знакомым с применяемым методом минимизации; не следует требовать от него задания многих параметров алгоритма. Вообще, задаваемую им информацию нужно сделать по возможностям минимальной (например, в задачах безусловной минимизации эта информация может сводиться к процедурам вычисления функции и градиента, указанию начальной точки и

критерия точности вычислений). Должно быть предусмотрено, что в случае невозможности найти решение выдается на печать причина отказа программ. Разумеется, здесь упомянута лишь часть требований к стандартным программам оптимизации.

В настоящее время широкое распространение получили *пакеты программ оптимизации*. Они предназначены для решения больших классов задач, входящие в них программы взаимосвязаны и используют обшире модули (типа одномерной минимизации); предъявляются общие требования к способу задания исходной информации и выводу результатов на печать и т. д. Пакеты программ более экономны, чем набор отдельных программ и предоставают большие удобства пользователю. Ими целесообразно пользоваться, если приходится регулярно решать разнообразные оптимизационные задачи.

**2. Где найти нужную программу.** Некоторые программы оптимизации включены в *математическое обеспечение* наиболее распространенных машин. В основном это относится к программам безусловной минимизации гладких функций. Наиболее часто при этом используется метод сопряженных градиентов и квазиньютоновские методы (см. §§ 2, 3 гл. 3). Относительно доступны пакеты программ для решения задач линейного программирования. Таков ПМП (пакет математического программирования) на базе операционной системы ОСЕС для ЭВМ ЕС; его прототип является пакет MPS фирмы IBM. На ряде машин (например, фирмы ICL) имеется LP1 — пакет программ линейного программирования, обладающий большими возможностями.

В крупнейших вычислительных центрах страны созданы собственные пакеты программ оптимизации. Так, в ВЦ АН СССР разработана диалоговая система для ЭВМ БЭСМ-6. Она предназначена для решения задач безусловной минимизации, нелинейного программирования и оптимального управления. Там же эксплуатируется пакет ЛП (линейного программирования) для БЭСМ-6, позволяющий решать задачи, содержащие до тысячи ограничений и нескольких тысяч переменных. В ЦЭМИ АН СССР разрабатывается пакет, предназначенный для решения основных типов задач оптимизации и включающий наиболее мощные из известных алгоритмов. В ВЦ МГУ создан пакет программ на языке Фортран, содержащий программы для безусловной минимизации и для задач с простыми ограничениями. Ряд пакетов оптимизации имеется или проектируется в Институте кибернетики АН УССР, в ВЦ ЛГУ, в Институте математики и механики УНЦ АН СССР и других организациях. Они в основном носят *экспериментальный* характер. В целом пакеты оптимизации еще не стали обязательной частью математического обеспечения любой ЭВМ. Использовать же «чужой» пакет на «своей» машине довольно сложно по ряду технических и организационных причин. Поэтому проблема получения нужной

программы оптимизации стоит весьма остро. Отчасти она может быть решена за счет стандартных программ, опубликованных в различных изданиях, либо включенных в Государственный фонд алгоритмов и программ. Упомянем некоторые из опубликованных программ.

В книге Химмельблау [0.21] приведены тексты на Фортране ряда алгоритмов безусловной и условной минимизации. Для поиска безусловного минимума гладкой функции даны несколько квазиньютоновских алгоритмов (в частности, реализующие метод Давидона — Флэтчера — Паузэлла (8) § 3 гл. 3 и Брайдена (9) § 3 гл. 3) и алгоритм сопряженных градиентов (в варианте (22) § 2 гл. 3), а также прямые методы (не требующие вычисления производных) — симплексный алгоритм (18) § 4 гл. 3 и алгоритм Паузэлла, в котором строятся сопряженные направления (п. 4 § 4 гл. 3). Для решения общей задачи нелинейного программирования приводится так называемый алгоритм скользящего допуска, идущий близкий к методу штрафных функций (10) § 4 гл. 9; для решения вспомогательных задач безусловной минимизации применяется метод прямого программирования.

В книгах Кустегра и Майза [12.12], Юнки, Шаха и Зендера [12.13] собрано довольно много программ (на Фортране и Алголе), относящихся в основном к линейному программированию и безусловной минимизации. Программы линейного программирования содержатся также в работах И. В. Романовского [10.14], Орчард-Хайса [12.15], Лэнд и Паузэлла [12.14]; в последней имеются и программы для квадратичных задач. В приложении к сборнику [11.6] приведен текст программы на Фортране, реализующий метод безусловной минимизации негладких функций вида (14), (17) § 4 гл. 5 (метод с растяжением пространства в направлении разности последовательных субградиентов). В учебнике [0.10] опубликовано описание диалоговой системы оптимизации ВЦ АН СССР, упоминавшейся выше.

Пожалуй, этим кратким списком почти исчерпываются публикации программ оптимизации в доступных советскому читателю книгах. Упомянем еще издаваемые различными организациями «*ротапринтные сборники* с текстами программ. Серия «Программы и алгоритмы», издаваемая ЦЭМИ АН СССР, включает ряд выпусков, составляющих в целом комплекс программ оптимизации. Описание этого комплекса содержится в статьях [12.1, 12.2]. В него входят программы безусловной минимизации — метод сопряженных градиентов в модификации (24) § 2 гл. 3 и его комбинации с методом переменной метрики; программы минимизации на параллелепипеде — метод сопряженных градиентов типа (7) § 3 гл. 7; программы для общей задачи нелинейного программирования, реализующие метод модифицированной функции Лагранжа (23) § 3 гл. 9; аналогичные программы для задач линейного и квадратичного программирования и некото-

рые другие программы, составленные на Алголе и Фортране. Этот комплекс — один из немногих доступных наборов универсальных и достаточно эффективных программ. В сборниках «Практическая оптимизация» (написанный на языках Фортран и PL-1), «Вычислительный центр МГУ» издает серию «Стандартные программы решения задач математического программирования» и «Численный анализ на Фортране» (в последней дано описание пакета программ оптимизации). Вычислительный центр ЛГУ — серию «Алгол-процедуры» (содержащую в основном программы линейного программирования), Институт математики и механики УНЦ АН СССР — серию «Программы оптимизации». Ряд программ можно найти в сборниках «Алгоритмы и алгоритмические языки», издаваемых ВЦ АН СССР.

Наконец, неопубликованные программы, включенные в Государственный фонд алгоритмов и программ, можно заказать в этом фонде. Списки поступающих в него материалов публикуются в регулярно издаваемых ВНТИ Центром информационных бюллетенях «Алгоритмы и программы».

Следует представлять себе *трудности*, связанные с отысканием, получением и использованием готовых программ. Как мы видели, опубликовано лишь небольшое число программ оптимизации; они рассеяны по многочисленным, нередко труднодоступным изданиям. Многие из программ при этом недостаточно проверены и могут содержать ошибки; реализуемые в них методы не всегда относятся к числу самых эффективных. Нередко полученная с большим трудом программа оказывается непригодной для использования на конкретной ЭВМ с конкретным транслятором и потому нуждается в переработке; однако производить изменения в чужой программе довольно сложно. По указанным причинам часто бывает проще самостоятельно написать программу для своей задачи, чем искать готовую.

### § 3. Тестовые задачи и результаты вычислений

В данном параграфе приводится несколько популярных тестовых задач оптимизации и полученные для них с помощью ряда основных методов результаты. Эти данные полезны как ориентировочные при оценке возможностей существующих и разработке новых алгоритмов. Полутно обсуждается вопрос о критериях численного сравнения методов и способах описания результатов вычислений.

**1. Критерии сравнения алгоритмов и описание результатов экспериментов.** Сравнить между собой несколько методов оптимизации по результатам численных экспериментов и сделать выводы о преимуществах одного из них отнюдь не просто,

Во-первых, сравнению подвергаются не методы, а *машинные реализации* соответствующих алгоритмов. Хороший метод можно «загубить» плохим программированием, неудачным выбором параметров алгоритма, проведением вычислений на машине малой разрядности и т. д.; мы уже упоминали об этом выше.

Во-вторых, неясно, каким образом соизмерять трудоемкость различных методов. Естественный на первый взгляд критерий — затраченное машинное время — в действительности не очень удобен. Трудно сравнивать быстродействие различных машин, не всегда имеются данные о затратах времени (при работе в мультипроцессорном режиме), время существенно зависит от языка программирования и особенностей транслятора. Более надежным показателем является число вычислений минимизируемой функции или другая «внутренняя» характеристика метода. Однако и здесь возникает ряд проблем. Неясно, как соизмерять трудоемкость вычисления функции и решения различных вспомогательных задач (вспомним, например, что метод центров тяжести (9) § 4 гл. 5 оптимальен по требуемому числу вычислений субградиента, но приводит к очень сложной вспомогательной задаче на каждой итерации). Затруднительно сопоставлять вычисления функции и ее производных. В тех задачах, где применяются конечно-разностные аппроксимации производных, дело обстоит просто — одно вычисление градиента эквивалентно  $n(n+1)/2$  вычислениям функции и т. д. Однако для дискретной «дороже» функции (см. (33) § 2 гл. 11); для квадратичной функции градиент вычисляется даже проще, чем ее значения; так же дело обстоит и с субградиентом в минимаксной задаче и т. д.

В-третьих, методы могут по-разному вести себя на разных этапах процесса минимизации. Пусть, например, два метода одинаковы по трудоемкости на каждой итерации, но для одного из них величина  $\|x^k - x^*\|$  сначала быстро убывает, а потом почти не меняется, а для другого эта же величина изменяется медленно, но с постоянной скоростью. Какой из двух методов следует считать лучшим? Более того, поведение разных величин, характеризующих точность решения ( $\|x^k - x^*\|$ ,  $\|\nabla f(x^k)\|$ ,  $f(x^k) - f^*$  и т. п.), может быть также различным.

Никакого удовлетворительного способа преодоления указанных трудностей не существует. Единственное, что можно сделать в подобной ситуации — приводить данные о результатах вычислений в развернутой форме, чтобы иметь возможность сравнить методы по разным критериям.

В общем, при *публикации результатов* проверки методов следует придерживаться следующих правил:

а) приводить точную формулировку задачи, для которой проводился счет, включая все ее параметры и начальное прижение;

б) указывать тип ЭВМ, длину машинного слова, язык программирования, транслятор, сведения о программе;

в) давать подробное описание применяемого алгоритма или отсылать к его публикации, если она имеется;

г) выводить не только окончательный результат, но и промежуточные (например, через каждые 100 итераций или при каждом увеличении точности на порядок);

д) сообщать различные характеристики точности приближения ( $\|x^k - x^*\|$ ,  $f(x^k) - f^*$ , невязку в ограничениях и выполнении условий экстремума); в задачах малой размерности приводить и сами приближения  $x^k$ ,

е) указывать подробно сведения о трудоемкости вычислений (число итераций, число вычислений  $f(x)$ ,  $\nabla f(x)$ , машинное время).

Лишь при выполнении этих условий можно воспроизводить полученные данные и сравнивать их с другими по разным показателям. Разумеется, столь подробное описание нужно лишь в работах, специально посвященных численной проверке методов. В других ситуациях скрупулезное выполнение упомянутых правил невозможно да и нелесообразно (пример тому — нижеприводимые краткие сведения о результатах счета).

**2. Общие требования к тестовым задачам.** Нередко в статьях, где предлагается новый метод минимизации, его проверка производится на 1—2 примерах, построенных автором. При этом сравнивают с другими алгоритмами либо не дается вообще, либо приводятся результаты счета для других, заведомо плохих методов. Такое положение типично, например, для многих работ по случайному поиску. Конечно, сделать какие-либо выводы о сравнимых достоинствах методов при подобной информации невозможно.

Сравнение методов целесообразно проводить на стандартных специально подобранных задачах-тестах. Желательно, чтобы тесты удовлетворяли следующим требованиям:

а) тесты должны быть унифицированы и быть общепринятыми; обычно такие задачи формируются в результате свободного естественного отбора, основным достоинством теста является его популярность;

б) тесты должны моделировать типовые трудности для данного класса задач (например, для задач безусловной минимизации нужны тесты с разной обусловленностью, разной размерностью, разной кривизной линий уровня, одно- и многоэкстремальные и т. д.);

в) решение в тестовой задаче должно быть известно;

г) задачи должны быть достаточно компактными (не требовать больших массивов информации или сложных правил вычисления функций);

д) в качестве тестовых не годятся задачи, обладающие специфическими особенностями, дающими преимущества тому или иному методу; так, если взять  $f(x) = f_1(x_1) + \dots + f_n(x_n)$  в качестве тестовой функции для безусловной минимизации, то метод покоординатного спуска будет для нее чрезвычайно эффективен; для функции вида  $f(x) = F(\varphi(x))$ , где  $\varphi(x)$  — квадратичная, а  $F$  — скалярная функция, линии уровня являются эллипсами, а потому ряд методов (скажем, метод Пауэлла, см. лемму 2 § 4 гл. 3) будут для них конечными, и т. д.

К сожалению, единых общепринятых тестовых задач пока нет. Более того, нет классификации тестовых задач по их сложности и характеру присущих им трудностей. Ниже делается попытка собрать несколько тестов для основных классов задач на оптимизацию, представляющих посильное приближение к сформулированным выше требованиям.

**3. Безусловная минимизация гладких функций.** Всюду далее  $x = \{x_1, \dots, x_n\} \in \mathbf{R}^n$ ,  $f(x)$  — минимизируемая функция,  $x^*$  — точка глобального минимума  $f(x)$  на  $\mathbf{R}^n$ ,  $f^* = f(x^*)$ ,  $x^0$  — начальное приближение,  $x^k$  — полученная в результате вычислений точка,  $\delta f = f(x^k) - f^*$  — точность решения по функции,  $\delta x = \max_{1 \leq i \leq n} |x_i^* - x_i^k|$  — точность решения по аргументам,  $\mu$  — число обусловленности  $x^*$  (см. (4) § 3 гл. 1),  $k$  — число итераций применявшегося метода,  $k_0$  — число вычислений функций,  $k_1$  — число вычислений ее градиента.

Проще всего тестовые задачи с известным решением можно построить следующим образом. Выбирается точка  $x^*$  и функции  $\Phi_i(x)$ ,  $i = 1, \dots, m$ . Тогда очевидно, что

$$f(x) = \sum_{i=1}^m (\Phi_i(x) - \Phi_i(x^*))^2 \quad (1)$$

достигает минимума в точке  $x^*$ . При этом  $f^* = 0$ ,  $\nabla^2 f(x^*) = -\sum_{i=1}^m \nabla \Phi_i(x^*) \nabla^T \Phi_i(x^*)$ , так что можно регулировать обусловленность задачи за счет выбора  $\Phi_i(x)$ . В частности, если  $m < n$  или  $m = n$ , но  $\nabla \Phi_i(x^*) = 0$  для некоторых  $i$ , то получаем вырожденную точку минимума. Функции вида (1), вообще говоря, невыпуклы и могут иметь локальные (и даже глобальные) минимумы, отличные от  $x^*$ .

Другой способ связан с выбором функций вида

$$f(x) = \sum_{i=1}^{n+1} f_i(x_i, x_{i-1}), \quad x_0 = a, \quad x_{n+1} = b. \quad (2)$$

Тогда, задавшись  $x_0^* = a$  и решив последовательно одномерные уравнения  $\partial f(x)/\partial x_i = \partial/\partial x_i [f_i(x_i^*, x_{i-1}^*) + f_{i+1}(x_{i+1}, x_i^*)] = 0$  (решение обозначим  $x_{i+1}^*$ ) для  $i = 0, \dots, n$ , выберем  $b = x_{n+1}^*$ . Очевидно, что тогда  $\nabla f(x^*) = 0$ . Если  $f_i(x_i, x_{i-1})$  выпуклы по  $x_i, x_{i-1}$ , то  $f(x)$  выпукла и  $x^*$  — точка минимума  $f(x)$ . Наконец, можно взять произвольную точку  $x^*$  и построить функцию  $\Phi(x)$  и произвольную точку  $x^*$  и построить  $f(x) = \Phi(x) - (\nabla \Phi(x^*), x)$ .

Тогда  $\nabla f(x^*) = 0$ ,  $f(x)$  выпукла и потому  $x^*$  — точка глобального минимума  $f(x)$ . Существует и много других приемов построения функций с известной точкой минимума.

Приводимые ниже примеры построены по одному из описанных способов. Следует помнить, что для функций вида (1), (2) существуют специальные методы, обладающие повышенной эффективностью. Так, для (1) эффективен метод Гаусса — Ньютона (14) § 1 гл. 11, для (2) — метод динамического программирования и покоординатного спуска (36), (37) § 2 гл. 11. Аналогичным образом специфика задач (1), (2) может оказаться на поведении общих методов минимизации.

Перейдем к описанию конкретных тестовых задач.

**Задача 1** (функция Розенброка [0.21, задача 2]),  $n = 2$ :

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2, \\ x^0 = (-1, 2; 1), \quad x^* = (1; 1), \quad f^* = 0.$$

Функция плохо обусловленная ( $\mu = 2500$ ), невыпуклая, с парabolическим оврагом, точка  $x^0$  далека от  $x^*$  (рис. 44). Возможное многомерное обобщение задачи 1 предложено в [12.3]:

$$f(x) = 100 \sum_{i=2}^n (x_i - x_{i-1}^2)^2 + (1 - x_1)^2,$$

$$x^* = (1; \dots; 1), \quad f^* = 0, \quad x_1^0 = -1, \quad x_i^0 = (x_{i-1}^0)^2 - 0, \quad i = 2, \dots, n.$$

**Задача 2** (функция Пауэлла [0.21, задача 26]),  $n = 4$ :

$$f(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4,$$

$$x^0 = (3; -1; 0; 1), \quad x^* = (0; 0; 0; 0), \quad f^* = 0.$$

Функция невыпуклая, точка минимума — вырожденная ( $\mu = \infty$ ).

**Задача 3** (разностный аналог задачи о брахистохроне [3.9]),  $n > 1$  — любое:

$$f(x) = \sum_{i=1}^{n+1} \left[ \frac{0.0016 + (x_i - x_{i-1})^2}{0.04i} \right]^{1/2}, \quad x_0 = x_0^* = 0, \quad x_{n+1} = x_{n+1}^*,$$

$$x_{i+1}^* = x_i^* + 0.04 \left[ \frac{0.0099099(i+1)}{1 - 0.0099099(i+1)} \right]^{1/2}, \quad i = 0, \dots, n,$$

$$x^0 = (0; \dots; 0).$$

Эта функция выпуклая (и даже сильно выпуклая на любом ограниченном множестве, хотя и не на всем пространстве), обусловленность растет с ростом  $n$ , но не слишком быстро ( $\mu \sim 10^4$  для  $n = 50$ ).

Задача 4 (среднеквадратичная аппроксимация экспонента [3.9]),  $n = 4$ :

$$f(x) = \alpha^{-2} \sum_{j=1}^{10} [\alpha \exp(-0.2jx_2) + 2\alpha \exp(-0.4j)] - x_1 \exp(-0.2jx_2) - x_3 \exp(-0.2jx_4)]^2, \quad f^* = 0,$$

4 а)  $x^0 = (0,5; 0; 2,5; 3)$ ,  $x^* = (1; 1; 2; 2)$ ,  $\alpha = 1$ ,

4 б)  $x^0 = (500; 0; 2500; 3)$ ,  $x^* = (1000; 1; 2000; 2)$ ,  $\alpha = 1000$ .

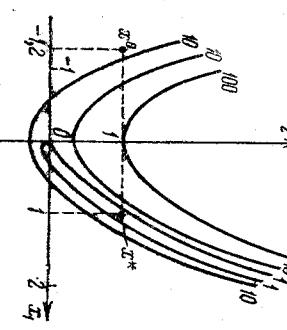
Функция  $f(x)$  невыпуклая, с искривленным оврагом, обусловленность велика (особенно для варианта 4 б)).

Задача 5 (среднеквадратичная

аппроксимация полиномами [11.22]),

$$f(x) = \sum_{i=1}^{101} \left( \sum_{j=1}^n x_i t_j^{i-1} - \sum_{j=1}^n x_i^* t_j^{i-1} \right)^2, \quad t_j = 0,01(j-1), \quad j = 1, \dots, 101,$$

$$x^0 = (2; \dots; 2), \quad x^* = (1; \dots; 1), \quad f^* = 0.$$



Функция  $f(x)$  квадратична; ее обусловленность чрезвычайно велика (в п. 5 § 1 гл. 11 приводились числа обусловленности для непрерывного варианта задачи:  $\mu = 1,5 \cdot 10^7$  для  $n = 6$ ,  $\mu = 1,6 \cdot 10^{13}$  для  $n = 10$ ; можно думать, что для дискретной задачи обусловленность того же порядка).

Численная проверка методов на тестовых задачах 1—5 осуществлялась В. А. Скоковым и Ю. Е. Нестеровым; остановимся кратко на полученных ими результатах.

В таблице 1 приведены данные для трех первых задач и четырех прямых методов (т. е. не требующих вычисления производной, см. § 4 гл. 3), именно, разностного аналога градиентного метода (9) § 4 гл. 3 («Градиент»), комбинации методов покордантного спуска (11) § 4 гл. 3, разностного варианта сопряженных градиентов (24) § 2 гл. 3, и барицентрических координат (19) § 4 гл. 3 («Барицентр», см. [122]), метода Паузэлла п. 4 § 4 гл. 3 («Паузэлл») и симплексного метода (18) § 4 гл. 3 («Симплекс»). Знакочерк «н» означает, что метод не привел к решению, прочек — что вычисления не проводились.

Как видно из табл. 1, градиентный метод не работоспособен даже для задач малой ( $n = 2,4$ ) размерности. Остальные методы работают, грубо говоря, примерно одинаково.

Таблица 1. Сравнение прямых методов для задач 1—3

Метод	Задача 1		
	$k_0$	$\delta x$	$\delta f$
«Градиент»	7657	$0,2 \cdot 10^{-5}$	$10^{-9}$
«Барицентр»	674	$10^{-9}$	$10^{-17}$
«Паузэлл»	151	—	$10^{-10}$
«Симплекс»	200	—	$10^{-8}$

Метод	Задача 2		
	$k_0$	$\delta x$	$\delta f$
«Градиент»	924	$10^{-4}$	$10^{-18}$
«Барицентр»	433	—	$10^{-13}$
«Паузэлл»	209	—	$0,7 \cdot 10^{-7}$
«Симплекс»	—	—	—

Метод	Задача 3, $n = 50$		
	$k_0$	$\delta x$	$\delta f$
«Градиент»	21	$10^{-4}$	$10^{-7}$
«Барицентр»	721	—	—
«Паузэлл»	—	—	—
«Симплекс»	—	—	—

В следующей таблице 2 показаны результаты вычислений (для тех же задач 1—3) с помощью методов первого порядка (т. е. использующих  $\nabla f(x)$ ). Вычисления проводились для методов: градиентного в варианте наискорейшего спуска (3), (4) § 1 (гл. 3 («Градиент»); двух версий метода сопряженных градиентов (22) и (24) § 2 гл. 3 («Конград-1» и «Конград-2») соответственно); метода Давидона — Флетчера — Паузэлла (8) § 3 гл. 3 («ДФП»); метода Брайдена — Флетчера — Шенно (10) § 3 гл. 3 («БФШ») и метода Шора (14), (17) § 4 гл. 5 («Шор-1») — по данным [5.15]. Результаты в таблице вновь показывают, что градиентный метод непригоден для решения (даже сравнительно простых задач). Квазиньютоновские методы (особенно «БФШ»)

Таблица 2. Сравнение методов первого порядка для задач 1–3

Метод	Задача 1				Задача 2			
	$k$	$k_1$	$\delta f$	$\delta x$	$k$	$k_1$	$\delta f$	$\delta x$
«Градиент»	217	476	$10^{-4}$	$10^{-2}$	1000	2003	$10^{-3}$	$10^{-1}$
«Конграл-1»	32	65	$10^{-9}$	$10^{-5}$	39	82	$10^{-5}$	$10^{-2}$
«Конграл-2»	30	63	$10^{-9}$	$10^{-5}$	27	59	$10^{-4}$	$10^{-2}$
«ДФП»	20	87	$10^{-9}$	$10^{-5}$	16	54	$10^{-9}$	$10^{-3}$
«БФП»	22	72	$10^{-11}$	$10^{-6}$	16	54	$10^{-9}$	$10^{-3}$
«Шор-1»	39	$k_0 = 286$	$10^{-15}$	$10^{-7}$	50	$k_0 = 695$	$10^{-13}$	$10^{-6}$

Метод	Задача 3, $n=20$				Задача 3, $n=50$			
	$k$	$k_1$	$\delta f$	$\delta x$	$k$	$k_1$	$\delta f$	$\delta x$
«Градиент»	288	701	$10^{-4}$	$10^{-1}$	1000	2369	0,1	1
«Конграл-1»	66	136	$10^{-6}$	$10^{-2}$	104	222	$10^{-5}$	$10^{-1}$
«Конграл-2»	67	138	$10^{-6}$	$10^{-2}$	129	265	$10^{-5}$	$10^{-1}$
«ДФП»	61	148	$10^{-8}$	$10^{-3}$	109	267	$10^{-7}$	$10^{-2}$
«БФП»	36	103	$10^{-8}$	$10^{-3}$	96	253	$10^{-8}$	$10^{-3}$
«Шор-1»	—	—	—	—	—	—	—	—

работают несколько лучше метода сопряженных градиентов; следует при этом помнить, что они требуют большей памяти ЭВМ и большего числа действий на каждой итерации. Из двух схем метода сопряженных градиентов более эффективна (хотя и ненамного) схема «Конграл-2». Нужно отметить высокую эффективность метода Шора — он успешно конкурирует с лучшими квазиньютоновскими методами. Учитывая, что метод Шора приведен и для минимизации неплавких функций (для которых, собственно, он и был разработан, см. § 4 гл. 5), его можно признать и эффективным, и универсальным.

Задача 3 подверглась также обстоятельному численному исследованию в [0.19, § 31]. Автор ставил, однако, цель получить по возможности точное решение непрерывной задачи о брахистохроне. Разумеется, с этой точки зрения задача 3 является плохой дискретной аппроксимацией для непрерывной задачи, но это никак не порочит ее как самостоятельную тестовую задачу конечномерной минимизации. В [0.19] показано также, как построить быстрый сходящийся метод минимизации, использующий специфический вид задачи 3. При оценке подобной ситуации нужно иметь в виду, что для каждой конкретной тестовой задачи можно указать специальные методы высокой эффективности, однако назначение тестовых задач — служить инструментом для проверки общих методов минимизации.

Для задачи 4 сравнивались следующие методы: метод «Конграл-2», сочетание метода сопряженных градиентов с Методом переменных метрики, предложенное в [3.9] («Конграл-3»); метод «Шор-1»; метод типа Гаусса–Ньютона (14) § 1 гл. 11, описанный в [11.22] («Гаусс — Ньютон»). Результаты приведены в таблице 3. Видно, что метод сопряженных градиентов для задачи 4 сходится довольно медленно, а для более трудной задачи 46 вообще не приводит к решению. Модификация этого метода («Конграл-3») значительно более эффективна. Метод Шора вновь, как и для предыдущих задач, дает хорошие результаты. Специализированный метод типа Гаусса — Ньютона работает гораздо лучше, чем общие методы. Такая высокая эффективность этого метода во многом объясняется особенностями задачи (метод Гаусса — Ньютона работает тем лучше, чем меньше  $f^*$ , см. § 1 гл. 11; в данном случае  $f^* = 0$ ). Следует также иметь в виду, что объем вычислений на каждую итерацию для этого метода существенно больше, чем для других.

Таблица 3. Сравнение методов для задачи 4

Метод	Задача 4а				Задача 4б			
	$k=k_1$	$k_0$	$\delta f$	$\delta x$	$k=k_1$	$k_0$	$\delta f$	$\delta x$
«Конграл-2»	179	781	$10^{-13}$	$10^{-4}$	н	110	$10^{-8}$	$10^{-4}$
«Конграл-3»	47	205	$10^{-18}$	$10^{-8}$	н	622	$10^{-18}$	$4 \cdot 10^{-4}$
«Шор-1»	83	400	$6 \cdot 10^{-5}$	$10^{-7}$	72	410	$4 \cdot 10^{-14}$	$10^{-3}$
«Гаусс — Ньютон»	6	—	$10^{-21}$	$10^{-11}$	11	—	$10^{-22}$	$10^{-7}$

Для задачи 5 сравнивались два метода — упомянутый выше вариант метода Гаусса — Ньютона [11.22] и обычный метод наименьших квадратов («МНК») — система линейных уравнений  $\nabla f(x) = 0$  решалась методом исключения Гаусса; результаты даны в таблице 4. Видно, что оба метода при  $n \leq 10$  дают практически точный минимум по функции; что касается точности по аргументу, то она гораздо выше для итеративного метода. Впрочем, время вычислений для метода Гаусса — Ньютона примерно в 10 раз больше, чем для «МНК».

#### Упражнения.

1. Как построить тестовую задачу вида  $f(x) = \sum_{i=1}^m ((a^i, x) - b_i)^2$ , в ко-

торой  $x^*$  известно, но  $f^* > 0$ ?

2. Полумайте, велики или малы величины  $\delta x$  в табл. 4 по сравнению с теоретическими для данных чисел обусловленности  $\epsilon$ .

Таблица 4. Сравнение методов для задачи 5

Метод	$n$	2	3	4	5
«Гаусс — Ньютон»	$\delta_f$	0	0	0	0
«МНК»	$\delta_f$	0	0	$2 \cdot 10^{-17}$	$3 \cdot 10^{-18}$
«Гаусс — Ньютон»	$\delta_x$	$10^{-8}$	$10^{-8}$	$10^{-8}$	$10^{-8}$
«МНК»	$\delta_x$	$10^{-8}$	$10^{-8}$	$6 \cdot 10^{-8}$	$2 \cdot 10^{-6}$

#### 4. Безусловная минимизация негладких функций. Негладкие тестовые функции вида

$$f(x) = \max_{1 \leq i \leq m} f_i(x)$$

с известным минимумом можно построить следующим образом. Зададимся выпуклыми гладкими функциями  $\varphi_i(x)$ ,  $i = 1, \dots, m$  (например, квадратичными или линейными) и точкой  $x^* \in \mathbb{R}^n$ . Вычислим градиенты  $\nabla \varphi_i(x^*)$  и найдем такое множество индексов  $I$  и числа  $\lambda_i > 0$ ,  $i \in I$ , что  $\sum_{i \in I} \lambda_i \nabla \varphi_i(x^*) = 0$  (этого всегда можно добиться, добавляя к набору  $\varphi_i(x)$  линейные функции вида  $(x, e_j)$ , где  $e_j$  — координатные орты). После этого возьмем  $f_i(x) = \varphi_i(x) - \alpha_i$ , где  $\alpha_i = \varphi_i(x^*)$  для  $i \in I$  и  $\alpha_i > \varphi_i(x^*)$  для  $i \notin I$ ; тогда в соответствии с условиями экстремума (см. лемму 11 § 1 гл. 5) функция (4) будет иметь глобальный минимум в  $x^*$ . Существуют и другие приемы построения тестовых функций (см., например, ниже задачу 8).

**Задача 6** (функция Шора [5.15, с. 176]),  $n = 5$ :

$$f(x) = \max_{1 \leq i \leq 10} b_i \|x - a^i\|^2, \quad b = (1; 5; 10; 2; 4; 3; 1,7; 2,5; 6; 3,5),$$

$$A = \begin{pmatrix} 0 & 2 & 1 & 1 & 3 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 4 & 2 & 2 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & \\ 0 & 1 & 1 & 2 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 3 & 2 & 2 & 1 & 1 & 1 & 0 & 0 \end{pmatrix},$$

$$\begin{aligned} x^* &= (1,12434; 0,97945; 1,47770; 0,92023; 1,12429), \\ f^* &= 22,60016, \quad I = (2, 4, 5, 9), \quad x^0 = (0; 0; 0; 0; 1). \end{aligned}$$

Здесь  $A$  — матрица, столбцами которой являются  $a^i$ ,  $I$  — множество активных индексов:  $I = \{i: b_i \|x^* - a^i\|^2 = f(x^*)\}$ .

В работе [5.15] использовалась одна из модификаций метода «Шор-1» (т. е. метода (14), (17) § 4 гл. 5), отличавшаяся способом выбора длины шага  $y_k$ , при этом за 51 итерацию (т. е. за 51 вычисление  $\partial f(x)$ ) было получено  $\delta_f = 7 \cdot 10^{-5}$ ,  $\delta_x = 4 \cdot 10^{-4}$ . Для этой же задачи в [0.19, с. 415—417] применялся метод типа метода линеаризации; потребовалось 14 раз решать вспомогательную задачу линейного программирования для достижения точности  $\delta_f = 5 \cdot 10^{-5}$ .

**Задача 7** [5.20, с. 151],  $n = 10$ :

$$f(x) = \max_{1 \leq k \leq 5} \{(A^{(k)} x, x) - (b^k, x)\},$$

$$A_{ij}^{(k)} = \exp(i/j) \cos j \sin k, \quad i \neq j,$$

$$A_{ii}^{(k)} = 0,1 \sin k + \sum_{j \neq i} A_{ij}^{(k)},$$

$$b_i^k = \exp(i/k) \sin ik, \quad i, j = 1, \dots, 10, \quad k = 1, \dots, 5,$$

$$x^* = (-0,1263; -0,0346; -0,0067; 0,2668; 0,0673; 0,2786; 0,0744; 0,1387; 0,0839; 0,0385),$$

$$f^* = -0,8414, \quad x^0 = (1; \dots; 1), \quad f^0 = 5337.$$

Вычисления, проведенные К. Лемарепалем, дали следующие результаты. Метод «Шор-1» дал  $\delta_f < 10^{-4}$  за  $k_1 = 60$  (т. е. за 60 вычислений субградиента). Два других метода негладкой минимизации, предложенные Ф. Бульфом и К. Лемарепалем [5.20], потребовали для достижения той же точности соответственно  $k_1 = 186$  и  $k_1 = 96$ . Попытка применить к задаче один из лучших квазиньютоновских методов гладкой минимизации «БФШ» не привела к точному решению — за  $k_1 = 86$  была достигнута точка с  $\delta_f = 0,08$ , после чего метод остановился.

**Задача 8** (апроксимация полиномами в  $l_1$ ),  $n$  — любое:

$$f(x) = \sum_{j=1}^{101} \left| \sum_{i=1}^n x_i t_j^{i-1} - \sum_{i=1}^n x_i^* t_j^{i-1} \right|,$$

$$t_j = 0,01(j-1), \quad j = 1, \dots, 101,$$

$$x^* = (1/n; \dots; 1/n), \quad f^* = 0, \quad x^0 = (0; \dots; 0).$$

Эта задача является негладким аналогом задачи 5. В ряде методов требуется задать  $Q_0$  — область локализации минимума, в этом случае бралось  $Q_0 = \{x: \|x\| \leq 1\}$ . Вычисления проводились (для различных алгоритмов) В. И. Венцом, О. В. Гулинским и А. С. Немировским. Метод эллипсоидов (12) § 4 гл. 5 скончился (по функции) для задач большой размерности, вплоть до  $n = 50$ . Интересно, что при этом достигалась и сравнительно

хорошая точность по аргументам. Так, для  $n = 5$  за 1200 итераций было получено  $\delta_f = 5 \cdot 10^{-11}$ ,  $\delta_x = 10^{-9}$ , для  $n = 20$  за 9500 итераций было найдено  $\delta_f = 10^{-11}$ ,  $\delta_x = 10^{-3}$ . Метод «Шор-1» ([14], (17) § 4 гл. 5) сходился значительно быстрее; так, для  $n = 5$  потребовалось лишь 47 итераций, чтобы получить  $\delta_f = 4 \cdot 10^{-5}$ ,  $\delta_x = 10^{-4}$ , а для  $n = 20$  за 180 итераций была найдена точка с  $\delta_f = 10^{-7}$ . Правда, в последнем случае точность по аргументу была очень плохой ( $\delta_x = 1$ ). Скорость сходимости существенно зависела от выбора  $\alpha$  (см. (17), (19) § 4 гл. 5); поскольку функция  $f(x)$  однородна с показателем 1, то для нее можно взять  $M = N = 1$ ,  $\alpha = \infty$  и добиться (в принципе) конечной сходимости метода. Проверялся также метод (14), (16) § 4 гл. 5 («Шор-2»). Его применение для данной задачи облегчается тем, что известно значение  $f^* = 0$ . Если не предполагать величину  $f^*$  известной, то регулировка длины шага в методе «Шор-2» усложняется, и его скорость сходимости существенно замедляется.

**5. Нелинейное программирование.** Приведем один из способов построения тестовых задач выпуклого программирования с ограничениями типа неравенств. Выбирается точка  $x^* \in \mathbf{R}^n$ , набор выпуклых дифференцируемых функций  $g_i(x)$ ,  $i = 0, \dots, m$ , и множество индексов  $I^*$  (число элементов в котором не превосходит  $n$ ). Далее конструируется задача

$$\begin{aligned} & \min [g_0(x) + (c - \nabla g_0(x^*), x)], \\ & g_i(x) - g_i(x^*) \leq 0, \quad i \in I^*, \\ & g_i(x) - a_i \leq 0, \quad i \in I^*, \end{aligned} \quad (5)$$

где  $a_i > g_i(x^*)$  — произвольные числа, а вектор  $c$  имеет вид

$$c = - \sum_{i \in I^*} \lambda_i \nabla g_i(x^*), \quad (6)$$

$\lambda_i \geq 0$ ,  $i \in I^*$ , — произвольные числа. Тогда в соответствии с теоремой Куна — Таккера  $x^*$  — решение задачи (5). Выбирая число элементов в  $I^*$ , можно получить различную размерность многообразия, на котором достигается решение. В частности, если в  $I^*$  включить  $n$  элементов так, что градиенты  $\nabla g_i(x^*)$ ,  $i \in I^*$ , линейно независимы,  $\lambda_i > 0$ ,  $i \in I^*$ , то  $x^*$  — острый минимум (см. теорему 4 § 2 гл. 9).

Ниже в дополнение к введенным выше будем применять следующие обозначения:  $m$  — общее число ограничений (не считая ограничений вида  $x \geq 0$ ),  $r$  — число ограничений типа неравенств,  $\delta g$  — невязка по ограничениям полученного решения:  $\delta g = \sum_{i=1}^r g_i(x^k)_+^2 + \sum_{i=r+1}^m g_i(x^k)^2$ ,  $k_0$  и  $k_1$  — число вычислений всех функций и их градиентов (например, число вычислений модифицированной функции Лагранжа и ее градиента).

**Задача 9** ([0.21, задача 18]),  $n = 15$ ,  $m = 5$ :

$$f(x) = - \sum_{i=1}^{10} b_i x_i + \sum_{i,j=1}^5 c_{ij} x_{10+i} x_{10+j} + 2 \sum_{j=1}^5 d_j x_{10+j}^3,$$

$$g_i(x) = \sum_{j=1}^{10} a_{ij} x_j - 2 \sum_{j=1}^5 c_{ij} x_{10+j} - 3d_j x_{10+j}^2 - e_i \leq 0, \quad i = 1, \dots, 5,$$

$$x_i \geq 0, \quad i = 1, \dots, 15,$$

$$A = \begin{pmatrix} -16 & 0 & -3,5 & 0 & 0 & 2 & -1 & -1 & 1 & 1 \\ 2 & -2 & 0 & -2 & -9 & 0 & -1 & -2 & 2 & 1 \\ 0 & 0 & 2 & 0 & -2 & -4 & -1 & -3 & 3 & 1 \\ 1 & 0,4 & 0 & -4 & 1 & 0 & -1 & -2 & 4 & 1 \\ 0 & 2 & 0 & -1 & -2,8 & 0 & -1 & -1 & 5 & 1 \end{pmatrix},$$

$$b = (-40; -2; -0,25; -4; -4; -1; -40; -60; 5; 1),$$

$$d = (4; 8; 10; 6; 2),$$

$$e = (-15; -27; -36; -18; -12),$$

$$x_i^0 = 0,0001, \quad i = 1, \dots, 15, \quad i \neq 7, \quad x_7^0 = 60, \quad f^0 = 2400,01.$$

И минимизируемая функция, и ограничения в этой задаче являются нелинейными и невыпуклыми. Точное решение неизвестно. В [0.21] приведено следующее приближенное решение:  $x^* = (0; 0; 5,174; 0; 3,0611; 11,8395; 0; 0; 0,1039; 0; 0,3; 0,3335; 0,4; 0,4283; 0,224)$ ,  $f^* = 32,386$ . Однако в [12.5] получено несколько иное решение с меньшим значением функции:  $x^* = (0; 0; 5,174; 0; 3,061117; 11,839466; 0; 0; 0,103877; 0; 0,300002; 0,333466; 0,400003; 0,428306, 0,223964)$ ,  $f^* = 32,348679$ .

Задача 9 является задачей средней трудности. На ней в ряде работ (ссылки можно найти в [0.21, 12.5]) проводились сравнение различных методов. По данным [0.21] (табл. 9.3.5, 9.3.6)] такие методы, как скользящего допуска, линеаризации (программа ПОП), Розенброка, обобщенного градиентного (ОГМОП), (по поводу терминологии см. [0.21]) не привели к решению задачи. Различные варианты метода барьерных функций и штрафов для  $k_1$  порядка нескольких тысяч дают лишь достаточно грубое решение  $\delta x \sim 0,1 \div 0,3$  и  $\delta f \sim 10^{-2} \div 10^{-3}$ . В таблице 5 приведены результаты вычислений из [12.5] для метода штрафных функций (10) § 4 гл. 9 и метода модифицированной функции Лагранжа (23) § 3 гл. 9 при трех начальных точках  $x^0$ :

- a)  $x_i^0 = 0, i = 1, \dots, 10, i \neq 7; x_7^0 = 60;$   
 $x_i^0 = 0,0001, i = 11, \dots, 15; x_i^0 = 0, i = 16, \dots, 20,$   
b)  $x_i^0 = 1, i = 1, \dots, 15; x_i^0 = 0, i = 16, \dots, 20,$   
b)  $x_i^0 = 1, i = 1, \dots, 15; x_i^0 = 10, i = 16, \dots, 20.$

Таблица 5. Сравнение методов для задачи 9

Начальная точка	$\delta x$	$\delta f$	$\delta g$	$k_0$	$k_1$
Метод модифицированной функции Лагранжа					
a)	$10^{-6}$	$10^{-6}$	$0,5 \cdot 10^{-12}$	1199	3921
b)	$10^{-4}$	$10^{-6}$	$0,4 \cdot 10^{-11}$	1248	4256
b)	$10^{-6}$	$10^{-6}$	$0,5 \cdot 10^{-10}$	1488	4911
Метод штрафных функций					
a)	$10^{-2}$	$0,4 \cdot 10^{-3}$	$0,2 \cdot 10^{-6}$	1776	6151
b)	$10^{-2}$	$0,6 \cdot 10^{-3}$	$0,3 \cdot 10^{-6}$	1188	4135
b)	$2 \cdot 10^{-2}$	$0,3 \cdot 10^{-3}$	$0,2 \cdot 10^{-6}$	1078	3895

Здесь у вектора  $x$  20 координат, так как задача записывается в форме

$$\min f(x),$$

$$g_i(x) + x_{15+i} = 0, i = 1, \dots, 5, x_i \geq 0, i = 1, \dots, 20,$$

являющейся канонической для программ, используемых в [12.5].

Данные в таблице 5 показывают, что спуск из существенно различных начальных точек проходит приблизительно одинаково, а метод модифицированной функции Лагранжа заметно эффективнее метода штрафных функций. Отметим еще, что коэффициент штрафа  $K$  регулируется и в первом методе становится равным  $\sim 100$ , во втором  $\sim 1000$  к концу вычислений. Расчет для одного варианта требовал порядка 3 мин. на машине БЭСМ-6 в системе БЭСМ-АЛГОЛ.

Задача 10 ([0.21, задача 20]),  $n = 24, r = 14, m = 20$ .

$$f(x) = \sum_{i=1}^{12} a_i(x_i + x_{i+12}),$$

$$g_i(x) = \frac{x_{i+12}}{24} - \frac{c_i x_i}{12} = 0, i = 1, \dots, 12,$$

$$b_i \sum_{j=13}^b \frac{x_j}{b_{i-12}} - 40b_i \sum_{j=1}^{24} \frac{x_j}{b_i} = 0, i = 1, \dots, 12,$$

$$g_{13}(x) = \sum_{i=1}^{24} x_i - 1 = 0,$$

$$g_{14}(x) = \sum_{i=1}^{12} \frac{x_i}{d_i} + 142,224705 \sum_{i=13}^{24} \frac{x_i}{b_{i-12}} - 1,671 = 0,$$

$$g_{15}(x) = \frac{x_1 + x_{13}}{24} - 0,1 \leqslant 0,$$

$$g_{16}(x) = \frac{x_2 + x_{14}}{24} - 0,3 \leqslant 0,$$

$$g_{17}(x) = \frac{x_3 + x_{15}}{24} - 0,4 \leqslant 0,$$

$$g_{18}(x) = \frac{x_7 + x_{19}}{24} - 0,3 \leqslant 0,$$

$$g_{19}(x) = \frac{x_8 + x_{20}}{24} - 0,6 \leqslant 0,$$

$$g_{20}(x) = \frac{x_9 + x_{21}}{24} - 0,3 \leqslant 0,$$

$$x_i \geq 0, i = 1, \dots, 24,$$

$$a = (0,0693; 0,0577; 0,05; 0,2; 0,26; 0,55; 0,03; 0,1;$$

$$0,12; 0,18; 0,1; 0,09),$$

$$b = (44,094; 58,12; 58,12; 137,4; 120,9; 170,9; 62,501; 84,94;$$

$$133,425; 82,507; 46,07; 60,097),$$

$$c = (123,7; 31,7; 45,7; 14,7; 84,7; 27,7; 49,7; 7,1; 2,1;$$

$$17,7; 0,85; 0,64),$$

$$d = (31,244; 36,12; 34,784; 92,7; 82,7; 91,6; 56,708; 82,7;$$

$$80,8; 64,517; 49,4; 49,1),$$

$$x_i^0 = 0,04, i = 1, \dots, 24.$$

Задача имеет невыпуклые и нелинейные ограничения. Ее можно было бы несколько упростить, заменив  $\sum_{i=1}^{24} x_i$  на 1 в огра-

ничениях 15–20, однако такие приемы считаются «запрещенными» при испытании общих методов. Решения, приведенные в [0.21], являются довольно грубыми; наиболее точное решение получено в [12.5] и имеет вид  $x^* = (0; 0,107248; 0,111390; 0; 0; 0; 0,0755407; 0; 0; 0; 0,0111949; 0; 0,192752; 0,288611; 0; 0; 0; 0,212858; 0; 0; 0; 0)$ ,  $f^* = 0,0556580$ .

Эта задача является наиболее сложной среди приведенных в [0.21]; большинство стандартных программ не привело к решению. Работоспособными по данным [0.21] оказались лишь методы приведенного градиента и скользящего допуска, причем последний для достижения невысокой точности  $\delta_f \sim 10^{-3}$ ,  $\delta_x \sim 0,4 \cdot 10^{-1}$  потребовал 8,5 мин. счета на СДС-6600. В [12.5] задача решалась методом модифицированной функции Лагранжа и методом штрафных функций. Для первого метода было получено  $\delta_x = 0,5 \cdot 10^{-6}$ ,  $\delta_f = 10^{-8}$ ,  $\delta_g = 10^{-13}$  при  $k_0 = 1828$ ,  $k_1 = 5436$  и за 5 мин. на БЭСМ-6, для второго  $\delta_x = 0,3 \cdot 10^{-3}$ ,  $\delta_f = 0,3 \cdot 10^{-4}$ ,  $\delta_g = 0,3 \cdot 10^{-4}$ ,  $k_0 = 204$ ,  $k_1 = 600$ , но более точных результатов при продолжении вычислений в методе штрафных функций получить не удается. Таким образом, для грубого решения годится метод штрафных функций, а дальнейшее уточнение можно производить методом модифицированной функции Лагранжа.

**Упражнение.** Постройте двойственную к задаче 9; покажите, что она может быть записана в виде  $\min \left[ (e, y) + (C_y, y) + \sum_{i=1}^5 d_i y_i^3 \right]$ ,  $y \in \mathbb{R}^5$ ,  $A^T y \geq b$ ,

$y \geq 0$  и является выпуклой, а ее решением является  $y^* = (0,3; 0,3335; 0,4; 0,4285; 0,2224)$ .

4. Как подобно (5) построить тестовую задачу, в которой были бы ограничения типа равенств?

**4. Линейное программирование.** Как это ни странно, не существует общепринятых тестовых задач линейного программирования различной размерности и сложности. Как правило, новые программы проверяют сначала на «игрушечных» задачах малой размерности (чтобы убедиться в правильности их работы), а затем применяют к реальному задачам. При этом обычно не делается попыток систематического сравнения эффективности программ на одинаковом числовом материале. Это связано с тем, что содержательную проверку методов и программ имеет смысл проводить на задачах, где  $m \geq 50$ ,  $n \geq 100$  (речь идет о канонической форме записи). Задачи меньшей размерности, как правило, решаются хорошими программами симплекс-метода на современных ЭВМ за пренебрежимо малое время (порядка секунд). Запись же информации в задачах утомительной размерности весьма трудоемка, и тратить силы на подобную работу без крайней необходимости никому не хочется.

По-видимому, целесообразно иметь два типа тестовых задач линейного программирования: специальных трудных задач малой размерности и больших задач, в которых матрица ограничений легко генерируется. К задачам первого типа можно отнести, во-первых, сконструированные примеры, в которых симплекс-метод требует большого числа шагов (экспоненциально зависящего от размерности), см. ниже задачу 11, а во-вторых, неустойчивые задачи, в которых матрицы линейных уравнений, решаемые на шаге симплекс-метода, плохо обусловлены. Последние можно получить с помощью дискретных задач равномерного или в норме  $l_1$  приближения функций полиномами, записанных в форме линейного программирования (см. (19) § 1 гл. 11).

Второй класс тестовых задач можно построить, например, следующим образом. Возьмем какую-либо простую выпуклую нелинейную функцию  $f(x)$  в  $\mathbb{R}^n$  и набор точек  $x^1, \dots, x^m$ , среди которых есть  $x^*$  — точка минимума  $f(x)$ . Тогда задача линейного программирования

$$\min t, \quad (Vf(x^i), x - x^i) - t \leq -f(x^i), \quad i = 1, \dots, m, \quad (7)$$

будет иметь решение  $\{x^*, 0\}$  (ср. с методами минимизации  $f(x)$ , описанными в § 4 гл. 5). Так, если взять  $f(x) = \|x\|^2$ , а в качестве  $x^i$  — точки вида  $\{x_i, \dots, x_n\}$ , где  $x_j, j = 1, \dots, n$ , принимают значения  $0, \pm 1$ , то получим задачу линейного программирования

$$\min t, \quad 2(x^i, x) - t \leq \|x^i\|^2, \quad (8)$$

в которой  $3^n$  ограничений, а решением является  $x^* = 0, t^* = 0$ . Насколько нам известно, численное исследование подобных многомерных и легко задаваемых задач (матрица ограничений в них легко генерируется) не проводилось.

Задача 11,  $n$  любое:

$$-1 \leq x_1 \leq 1, \\ -2x_{i-1} - 2^{2i-2} \leq x_i \leq 2x_{i-1} + 2^{2i-2}, \quad i = 2, \dots, n, \quad (9)$$

$$x_1^* = 1, \quad x_i^* = 2x_{i-1}^* + 2^{2i-2}, \quad i = 2, \dots, n.$$

В этой задаче многогранник ограничений имеет  $2^n$  вершин; если начать с вершины

$$x_1^0 = 1, \quad x_i^0 = 2x_{i-1}^0 + 2^{2i-2}, \quad i = 2, \dots, n-1, \\ x_n^0 = -2x_{n-1}^0 - 2^{2n-2},$$

то итерации симплекс-метода при одном из способов перехода к соседней вершине пройдут по всем вершинам многогранника.

Задачу можно задать и в канонической форме:

$$\begin{aligned} \max & \sum_{j=1}^n c_j x_j, \\ \sum_{j=1}^t a_{ij} x_j + x_{n+i} &= b_i, \quad i = 1, \dots, n, \\ x_j &\geq 0, \quad j = 1, \dots, 2n, \\ a_{ii} &= 1, \quad a_{ij} = (-1)^{i+j} 2^{i-j+1}, \quad j < i, \\ b_i &= (2^{2i} - (-2)^i)/3, \quad i = 1, \dots, n, \\ c_j &= (-2)^{n-j}, \quad j = 1, \dots, n, \\ x_i^0 &= 0, \quad i = 1, \dots, n, \quad x_{n+i} = b_i, \quad i = 1, \dots, n. \end{aligned} \tag{10}$$

Здесь  $x_i^*$ ,  $i = 1, \dots, n$  — те же, что и ранее;  $x_i^* = 0$ ,  $i = n+1, \dots, 2n$ ; решение двойственной задачи имеет вид  $y_i^* = 2^{n-i}$ ,  $i = 1, \dots, n$ . Численное исследование этой задачи было выполнено Н. А. Соколовым. При фактическом применении симплекс-метода для  $n = 15$  число шагов оказалось равным 21402 (вместо теоретических  $2^{15} = 32\,768$ ); по-видимому, это связано с применявшимся в конкретном алгоритме способом выбора соседней вершины. Однако и такое число шагов нужно признать чрезвычайно большим (для задач подобной размерности обычно число итераций не превосходит 100). Решение с помощью итерационного метода, основанного на модифицированной функции Лагранжа (типа (12) § 3 гл. 10) дало точность порядка 2—3% (и по аргументам, и по критерию) за 556 итераций, при этом каждая итерация включала порядка 7 циклов метода покординатного спуска для минимизации модифицированной функции Лагранжа.

#### Упражнение.

5. Покажите, что следующая задача [10.18] эквивалентна (9):

$$\max x_n,$$

$$\begin{aligned} \varepsilon x_{i-1} \leq x_i \leq 1 - \varepsilon x_{i-1}, \quad i = 2, \dots, n, \quad 0 < \varepsilon < 1/2, \\ x^0 = (0; \dots; 0), \quad x^* = (0; \dots; 0; 1). \end{aligned}$$

## БИБЛИОГРАФИЧЕСКИЕ УКАЗАНИЯ И КОММЕНТАРИИ

### К ГЛАВЕ 1

**§ 1.** Основные факты, связанные с дифференцированием функций многих переменных, можно найти в любом курсе математического анализа, например в учебнике Г. Е. Шилова [1.10]. При переходе к абстрактным пространствам эти факты становятся только более наглядными. О дифференцировании функционалов и операторов можно прочесть в монографиях Л. В. Kantorovicha и Г. П. Akilova [1.5], М. М. Bainbridge [1.2].

Свойства выпуклых и сильно выпуклых функций приводятся в большинстве монографий по математическому программированию, в частности в книгах Ф. П. Васильева [0.2], В. Г. Карманова [0.8], Ортеги и Рейнболда [0.12]. Впервые понятие сильно выпуклой функции было введено в [1.7].

**§ 3.** С общей проблемой корректности математических задач читатель может ознакомиться по монографиям [1.4, 1.8]. Применительно к задачам оптимизации мы приводим в учебниках Ф. П. Васильева [0.2] и В. Г. Карманова [0.8]. Отметим, что употребляемая там терминология не всегда совпадает с вышеприведенной.

**§ 4.** Идея градиентного метода восходит к Коши (1830 г.). Первые доказательства сходимости при различных предположениях получены Л. В. Kantorovichem, M. M. Vajbergom, Karri, Krokem и Chernovym в 1940—1950-х гг. В изложении материала этого параграфа мы следуем в основном статье [1.6]. Лемма 2 доказана Е. Г. Гольштейном и Н. В. Третьяковым [1.3]. Много материала о поведении градиентного метода содержится в монографии [0.12].

**§ 5.** Линеаризацию как способ решения уравнений применял еще Ньютона, а первые результаты о сходимости принадлежат Фурье и Коши. Систематическое изучение метода Ньютона осуществлено Л. В. Kantorovichem в конце 40-х годов, его результаты приведены в книге [1.5]. Различные обобщения метода и относящиеся к нему теоремы сходимости описаны в [0.12]. Отметим также, что передко метод называют методом Ньютона — Рафсона или Ньютона — Канторовича.

**§ 6.** Серьезное обсуждение проблем о соотношении теории и практики в вычислительной математике содержит, например, в известных книгах Н. С. Бахвалова и Хемминга по численным методам [1.1, 1.9]. Мы рекомендуем также читателю ознакомиться с точкой зрения Р. П. Федоренко [0.19].

### К ГЛАВЕ 2

**§ 1.** Идея использования линеаризации для исследования асимптотического поведения траекторий, описываемых дифференциальными уравнениями, была высказана А. М. Ляпуновым (первый метод Ляпунова) в конце прошлого века и развита в работах Пуанкаре, Бендикиона и многих других авторов. Первые результаты, относящиеся к дискретным процессам (описываемым разностными уравнениями), были получены Пуанкаре и Перроном. Так, основная теорема 1 по существу доказана Перроном в 1929 г. (хотя ее часто называют теоремой Островского, получившего этот результат в 1957 г.). Систематическое исследование сходимости итеративных процедур на основе первого метода Ляпунова начато в 50-х годах в работах Беллмана, П. В. Бромберга и Островского. Наиболее подробно эта теория изложена в монографии Ортеги и Рейнболда [0.12], где приведена и обширная библиография.

**§ 2.** Различные леммы о чистовых последовательностях рассеяны в многочисленных работах, нередко не связанных со сходимостью итеративных методов, и установление их авторства представляет трудности.

Применение скалярных функций, монотонно убывающих на траекториях процесса, для исследования асимптотической сходимости и устойчивости обыкновенных дифференциальных уравнений восходит к Ляпунову (прямой или второй метод Ляпунова). Этот метод является наиболее употребительным инструментом анализа в данной области. Для анализа дискретных процессов метод функций Ляпунова применялся, начиная с работ Беллмана и П. В. Бромберга. Наиболее полные результаты были получены Ю. И. Любичем и Г. Д. Майстрским [2.7]. Впрочем, в работах по вычислительной математике доказательства, основанные на идеи использования функции Ляпунова, являются традиционными (см., например, [2.4]). В стохастических задачах техника второго метода Ляпунова впервые была применена Бломом [2.12]. С тех пор она многократно использовалась и обобщалась [2.1, 2.6, 2.8, 2.9, 2.11].

В формулировках и доказательствах лемм и теорем § 2 мы следовали изложению [2.11].

**§ 3.** Принцип сжимающих отображений традиционно использовался в математике не столько для исследования сходимости, сколько для доказательства теорем существования. Примером может служить классическое доказательство существования и единственности решения дифференциальных уравнений, проходящего через данную точку, предложенное Пикаром. Абстрактная формулировка принципа дана Банахом. Различные обобщения можно найти в [0.12].

## К ГЛАВЕ 3

Методам безусловной минимизации удалено много внимания в монографиях [0.2, 0.8, 0.10, 0.12, 0.13, 0.15, 0.20, 0.21, 0.22, 1.2]; им посвящены также обзоры [3.8, 3.10].

**§ 1.** Теорема 1 о сходимости метода наискорейшего спуска получена Карри в 1944 г. Скорость сходимости этого метода для квадратичных функций (теорема 2) исследовалась Л. В. Кантовичем [1.5]. Процедура регулировки длины шага (10) предлагалась различными авторами, в частности Армико, С. А. Смоляком, Б. Н. Пшеничным и Ю. М. Данининым [0.15]. Метод Ньютона с регулировкой длины шага (11) предложил Л. В. Кантович в 1948 г. Алгоритм типа (16) для специальных типов задач (минимизации сумм квадратов нелинейных функций) впервые использовал Легенберг в 1944 г., а широкую известность он приобрел благодаря работе Марквардта [3.20]. Различные детали и более подробные библиографические указания по материалам данного параграфа можно найти в монографии Ортеги и Рейнboldта [0.12].

**§ 2.** Двухшаговые методы для ускорения сходимости итерационных процедур традиционно применялись в линейной алгебре [3.6, 3.11]. На неквадратичные задачи минимизации метод (2) был перенесен в [3.7], там же получены утверждения о сходимости типа теоремы 1.

Метод сопряженных градиентов, играющий очень большую роль в теории и практике оптимизации, появился в 1952 г. в работах Хестенса и Штифеля [3.17] как способ решения систем линейных уравнений с положительно определенной матрицей. В [3.16] содержится подробное исследование свойств метода и его модификаций, сравнение различных вычислительных схем. В 1964 г. Флэтчер и Ривз [3.15] предложили применять метод в форме (22) для неквадратичных задач и высказали эвристические соображения об эффективности метода. Первое доказательство сходимости дано Дэниелом (см. [3.12]). Вычислительная схема (24) предложена в [0.13] и [3.9], там же даны доказательства сходимости для вариантов с обновлением.

**§ 3.** Первый квазиньютоновский алгоритм был предложен Давидом в 1959 г. и стал популярным благодаря работе Флегчера и Паулэла [3.14]. Этот метод Давида — Флегчера — Паулэла описывается формулами (1), (8). С тех пор появилось множество работ, в которых либо формулируются новые варианты квазиньютоновских методов, либо делается попытка классифицировать эти методы и получить их с помощью единого подхода [3.8], либо дается обоснование известным методам [3.2]. Подробно эти вопросы изложены в монографиях [0.15, 0.22, 3.16] и обзоре [3.13].

По поводу обоснования различных вариантов методов переменной метрики и сопряженных направлений можно обратиться к упомянутымся работам [0.15, 0.22, 3.2]. Метод секущих для решения одномерных уравнений известен очень давно, для двумерного случая, как указывает Островский [2.10], он был предложен еще Гауссом. В общей форме метод описан Вульфом [3.22]. Обоснование метода дано Ю. М. Данининым [3.1]. В обзоре [3.8] приведены некоторые целеособоразные с вычислительной точки зрения модификации метода. Идея использования однородной модели принадлежит Джекобсону и Оксману [3.19].

**§ 4.** Рангообразные методы конечно-разностной аппроксимации и оценки их точности приведены в [3.4]. Методы вида (6)–(8) исследовались в [4.2]. Метод покоординатного спуска для квадратичных функций хорошо известен в линейной алгебре как метод Гаусса — Зейделя [3.6, 3.11]. Инициатором применения и активным пропагандистом методов случаного поиска является Л. А. Растигин [0.16]. Симплексный метод был предложен в [3.21] и обобщался многими авторами (см. посвященную ему монографию [3.3]). Описанный в п. 4 метод барицентрических координат построен С. С. Лавровым [3.5].

## К ГЛАВЕ 4

Вопрос о влиянии помех на метод оптимизации в целом мало исследован в литературе. В основном изучалась случай аддитивных случайнных помех. Первые работы в этой области появились в статистике и были связаны с алгоритмами стохастической аппроксимации [4.14, 4.12]. Специально для задач минимизации эти исследования подытожены в монографии Ю. М. Ермолева [4.2], Кушнира и Кларка [4.13].

**§ 1.** Источники и типы помех в вычислительных задачах вообще освещены в книгах по численным методам [1.1, 1.9]. Многочисленные примеры задач адаптации, обучения, распознавания и т. д., пригодные к минимизации функций типа среднего риска, приведены в монографиях [4.2, 4.9, 4.11]. Классификация помех, приводимая в этом параграфе, не является общепринятой.

**§ 2.** Случай детерминированных помех затрагивалась в ранних работах по экстремальному регулированию [4.4] и в литературе по методам оптимизации не исследовалась. Задачи со случайнными помехами, как уже отмечалось, изучены лучше [4.1, 4.2, 4.12, 4.14, 0.16, 0.18, 2.3, 2.8].

**§ 3.** Сравнение скорости сходимости одно- и двухшаговых методов в условиях постепенного регулирования [4.6].

**§ 4.** Первые результаты типа теоремы 1 были получены еще в [4.12], см. также [2.3, 2.8]. При построении разностного варианта градиентного метода и метода случайного поиска мы следуем [4.7]. Рекуррентная версия метода наименьших квадратов (13) является частным случаем фильтра Калмана.

**§ 5.** Результаты п. 1 получены в [4.10]. Асимптотически оптимальные алгоритмы оптимизации при наличии случайнных помех указаны и обоснованы в [4.8]. Требуемые сведения из математической статистики (например, неизвестно Крамера — Рао) можно найти в [4.3].

## К ГЛАВЕ 5

В настоящее время существует несколько монографий, специально посвященных задаче минимизации нелинейных функций. Это, прежде всего, книга Н. З. Шора [5.15], являющаяся пионером в области методов неглажкой оптимизации. Монография В. Ф. Деминова [5.2, 5.3] посвящены минимизации функций типа максимума. В книге Ю. М. Ермолова [4.2] много внимания уделено неглажкой оптимизации при случайных помехах. Вопрос об оптимальных методах выпуклой минимизации подробно исследован в монографии А. С. Немировского и Д. Б. Юдина [0.11]. Обобщения на случай некоторых невыпуклых неглажких задач можно найти в книге Е. А. Нурминского [2.9] и А. М. Гулаша [5.1]. Наконец, имеются сборники статей по этой тематике [5.19, 5.20].

**§ 1.** Использованный материал по выпуклому анализу содержится в фундаментальной монографии Рокафеллера [5.13], см. также [5.1]. Первые результаты о субградиентах и их свойствах (типа лемм 6—12) принадлежат Б. Н. Пшеничному [5.12]. Различные обобщения леммы 7 можно найти в книге В. Ф. Деминова [5.2]. Свойства  $\varepsilon$ -субградиентов описаны в [2.9].

**§ 2.** Задачи с острым минимумом подробно изучены в статье [5.21].

**§ 3.** Субградиентный метод (с постоянной длиной шага), был впервые предложен Н. З. Шором в 1962 г. (см. библиографию в [5.15, 5.19, 5.20]). Способ регулировки длины шага (4) указан Ю. М. Ермоловым [0.5] и Б. Т. Поляком [5.8]. Субградиентный метод в форме (7) для частных задач был введен И. И. Ереминым [5.4]. Подробное исследование такого метода осуществлено в [5.9]. Варианты метода из упражнения 2 описаны в [5.5, 5.15]. Детальный анализ скорости сходимости различных модификаций субградиентного метода и их численная проверка даны в [5.16]. Результаты, относящиеся к  $\varepsilon$ -субградиентному методу, имеются в [2.9, 5.3]. Различные его обобщения предложены Лемарешалем, Бульфом и другими [5.19, 5.20, 2.9].

**§ 4.** Метод отсекающей гиперплоскости разработан Келли [5.17] еще до появления более простого субградиентного метода. Результаты о сходимости метода Келли имеются в [0.9]. Методы чебышевских центров посвящены работы [5.6, 5.10]. Вариант (6) предложен в [5.9]. Оригинальный метод центра тяжести разработан одновременно и независимо А. Ю. Левиным [5.7] и Ньюменом [5.18]. Теорема 2 в приведенной выше форме доказана А. С. Немировским и Д. Б. Юдиным [0.11].

Вопрос об оптимальных методах глубоко исследован эими же авторами [0.11]. Ими же предложен и обоснован метод (12) [0.11]. С другой точки зрения к этому методу пришел Н. З. Шор. Он также разработал многочисленные варианты методов с растяжением пространства (эти работы подтверждены в [5.15]). Форма (14) для этих методов предложена В. А. Скоковым [5.14].

**§ 5.** Субрадиентный метод при налении помех изучал Ю. М. Ермолов [4.2], см. также [2.9, 4.1, 5.21].

**§ 6.** Метод (3) принадлежит А. М. Гулашу [5.1]; им же исследована сходимость метода.

## БИБЛИОГРАФИЧЕСКИЕ УКАЗАНИЯ И КОММЕНТАРИИ

Результат о сходимости из теоремы 1 был приведен без доказательства в [6.12] и обоснован Е. Г. Гольштейном и Н. В. Третьяковым [1.3]. Метод регуляризации для некорректных задач решения операторных уравнений предложен А. Н. Тихоновым в 1963 г. (см. библиографию в [1.8]). Теорема 4 дополнена Е. С. Левитиным и автором [6.9]. Многочисленные результаты, относящиеся к псевдообратным матрицам, можно найти в книге Альберта [6.1].

Понятие проекционного отображения введено Моро, его результаты описаны в § 31 монографии Рокафеллера [5.13]. Метод (18) был предложен Мартине [6.19], его обобщения и подробный анализ даны Рокафеллером [6.20]. Метод итеративной регуляризации (21) предложен А. Б. Бакушинским и автором [6.2]. Идея «регуляризации» с помощью выбора момента остановки итеративного процесса при решении линейных операторных уравнений высказывалась А. Б. Бакушинским, М. М. Лаврентьевым и другими авторами. Наиболее полное исследование этого подхода с указанием конструктивных правил остановки содержится в [6.6]. Метод регуляризации в задачах минимизации при наличии помех впервые исследовался В. А. Морозовым [6.10]. Дальнейшие результаты можно найти в [0.2, 0.8, 1.8].

**§ 2.** Задаче глобальной оптимизации посвящены монографии [6.15, 6.16], сборник [6.21], ей удалено большое внимание в книгах [6.3, 6.4, 6.11]. Длительные градиентные методы липшицевых функций предлагаются рядом авторов (см. [6.3, 6.16]). Овражный метод предложен И. Л. Гельфандом и М. Л. Петлинским [6.5]. Идея разделения процесса поиска на этапы спуска, подъема и пересечения и описанный способ определения направления «клиамида-леннейшего подъема» принадлежит И. Е. Федоровой [6.17]. Другие детерминированные методы глобальной оптимизации приведены в [6.21].

Методом случайного поиска посвящена общирная литература. С этими методами можно ознакомиться по монографиям Л. А. Растринина [0.16, 6.13]. Последние результаты в этой области отражены в работе изательством «Эзинате». Статистические модели глобальной оптимизации исследованы в работах И. Б. Молкуса [6.11], Р. Г. Строгина [6.15], Д. И. Батиева [6.3], А. Жильинского [6.22].

**§ 3.** Нестационарные задачи оптимизации при наличии ограничений рассматриваются в монографии [6.7]. Первоначально вопросы зависимости решения от параметра изучались в теории, посвященной методам продолжения по параметру для решения нелинейных уравнений (см. [0.12], §§ 7.5, 10.4). Совсем с другой точки зрения нестационарные задачи оптимизации исследованы в работах по экстремальному регулированию [0.16, 4.1] и по динамическим методам [6.23].

## К ГЛАВЕ 7

Значительное внимание задаче минимизации на простом множестве уделено в монографиях [0.2, 0.4, 0.15, 0.17, 3.4, 4.2] и работах [0.9, 0.14].

**§ 1.** Впервые условие экстремума типа теоремы 1 получил еще в 1940 г. Л. В. Канторовичем в выпущенном виде теорема 1 доказана В. Ф. Деминовым и А. М. Рубиновым [0.4]. Результат для неглажкого случая (теорема 3) принадлежит Б. Н. Пшеничному [5.12]. Задачи с условиями (2) и более общими такого типа исследовались в [5.21]. Обобщение минимизирующие постепенности и соответствующее условие устойчивости введены в [0.9]. Результаты о «сверхустойчивости» острого минимума получены в [5.21].

**§ 2.** Метод проекции градиента независимо предложен в работах [7.6, 0.4, 0.9]. Первая часть теоремы 1 (п. а — в) получена в [0.9]. Результат о количестве итераций метода в случае острого минимума содержится в [5.21]. Дальнейшие исследования метода проекции градиента можно найти в [7.2].

## К ГЛАВЕ 6

**§ 1.** Некорректным задачам (частным случаем которых являются задачи безусловной минимизации с вырожденным минимумом) уделяется очень большое внимание в исследованиях последних лет. Укажем в этой связи монографии А. Н. Тихонова и В. Я. Арсеньева [1.8] и В. К. Иванова, В. В. Васина и В. П. Тананы [1.4], М. М. Лаврентьева [6.8]. В этих работах основное внимание уделяется решению линейного уравнения  $Ax = b$  в гильбертовом пространстве в ситуации, когда оператор  $A$  не имеет обратного.

Для негладких задач метод проекции субградиента (8), (9) введен в [5.8], а в форме (8), (10) — в [5.9].

Франк и Вульф [7.5] предложили метод (11) для задач квадратичного программирования (т. е. с квадратичной  $f(x)$  и многогранником  $Q$ ). Поэтому передко метод условного градиента называют методом Франк и Вульфа. Для общей задачи (A) метод условного градиента применял В. Ф. Демянтов [10.4]. Оценки скорости сходимости и различные правила выбора длины шага предложены в [0.9, 7.4]. Поведение метода для острого минимума изучено в [7.4, 5.21]. Пример типа (16), демонстрирующий медленную сходимость метода, построен в [7.3].

Метод Ньютона для задач с ограничениями введен в [0.9].

**§ 3.** Некоторые результаты, близкие к материалу п. 1 данного параграфа, можно найти в [0.15]. Метод сопряженных градиентов с ограничениями вида  $a \leq x \leq b$  введен и обоснован в [3.9]. Относительность многогранниковых методов минимизации негладких функций с ограничениями  $x \in Q$  см. упоминавшиеся в гл. 5 работы.

**§ 4.** Сходимость метода проекции субградиента при случайных помехах (теорема 2) доказана Ю. М. Ермолевым [4.1, 4.2]. Метод стохастического усиленного градиента в форме (6) предложен в [7.1], см. также [5.1]. Использование методов при ошибках, удовлетворяющих условию (9), проведено в [4.5]. Ряд результатов, относящихся к методам решения задачи (A) при на-личии помех, имеется в [3.4].

## К ГЛАВЕ 8

**§ 1.** Условие экстремума в форме (2) имеется в большинстве курсов математического анализа, например в [1.10].

Теорема Люстерника (для бесконечномерного случая) была доказана в 1934 г., с этим результатом можно ознакомиться по [8.4, 1.5, 9.1]. Доказательство, основанное на теореме о неявной функции, является стандартным в анализе (см., например, [1.10]). Идея применения штрафных функций для вывода правила множителей Лагранжа высказывалась различными авторами. Приводимое здесь доказательство следует [8.2].

Основная лемма 3 часто называется леммой Фишлера или леммой Дебре (см., например, [2.2] гл. 5). Еще обобщения на бесконечномерный случай и другие леммы этого параграфа получены в [8.9]. Модифицированная функция Лагранжа (14) впервые была введена в вариационном исчислении в 30-х — 40-х годах нашего века для доказательства достаточных условий оптимальности.

Утверждения об устойчивости типа теоремы 8 содержатся в [0.20]. Результаты для более общих задач (типа  $\min f(x, \varepsilon)$ ) при ограничениях  $g_i(x, \varepsilon) = 0, i = 1, \dots, m$ ) можно найти в [8.3, 8.5].

**§ 2.** Методы решения задач с ограничениями типа равенств рассматриваются в монографиях [0.15, 0.20, 0.22] и обзорах [0.14, 8.11]. В большинстве же руководств, посвященных математическому программированию, анализируется лишь выпуклый случай, не включающий задачи такого вида.

Метод линеаризации на уровне идеи предлагается различными авторами. Как четко сформулированный алгоритм его ввел Б. Н. Шеневич [8.10]. Эти результаты подробно описаны в [0.15]. Приводимая версия метода отличается от указанной в [0.15] более простым правилом выбора длины шага.

Методы множителей Лагранжа (6), (8) систематически изучались Эрроу, Гурвицем и Узлавой [0.23] в основном для выпуклых задач. Один раздел в монографии [0.23] (гл. 11) посвящен задачам с ограничениями типа равенств. Теорема 2 и результат о сходимости метода (8) доказаны автором [8.6].

Метод модифицированной функции Лагранжа в форме (11) был независимо предложен в работах [8.13—8.15]. Теоремы 4 и 5 о его сходимости

и скорости сходимости доказаны Н. В. Третьяковым и автором [8.9], см. также [8.11]. Вариант метода (9) для линейных ограничений предложен А. С. Анициным [8.1], а для общего случая в [8.8].

Метод штрафных функций исторически был первым методом условной минимизации, его ввел Кунт в 1943 г., а впервые обосновали Батлер и Мартин в 1962 г. Оценка скорости сходимости метода (теорема 7) получена в [8.7]. Упомянем также работы [0.9, 0.20, 0.22], где можно найти и дальнейшую библиографию.

Метод привеленного градиента издавна применялся практиками. Его исследовал Бульф [8.17] (см. также [0.22]).

Метод Ньютона (19) является наиболее прямолинейным подходом к задаче, использующим правило множителей Лагранжа. Условия его сходимости (теорема 9) получены в [8.6].

Робинсон [8.16] спел задачу условной минимизации к последовательности задач с линейными ограничениями вида (25). Ему принадлежит и результат о квадратичной сходимости метода. Методы типа (26) исследовались в [0.22].

**§ 3.** Поведение методов оптимизации с ограничениями типа равенств при наличии случайных помех подробно изучено в [8.8].

## К ГЛАВЕ 9

**§ 1.** Теория выпуклого программирования рассматривается в ряде учебников и монографий [0.7, 0.8, 0.20, 0.23, 2.5, 5.11, 5.12, 9.9, 9.26].

Как уже отмечалось, наиболее полным курсом выпуклого анализа является монография Рокафеллера [5.13], где можно найти доказательства лемм 1—4. Фундаментальная лемма 4 (бесконечномерном случае) см. ссылки в [5.13]. Лемма 5 получена (в бесконечномерном случае) А. Я. Дубовицким и А. А. Милютиным [9.7]. Этот результат лежит в основе нашей широкое применение техники Дубовицкого — Милютина для анализа общих экстремальных задач (см. в этой связи [9.3, 9.7, 9.12, 9.28]). Лемма Фаркаша или эквивалентные ей утверждения известны с 30-х годов нашего века, она является основным инструментом в конечномерной теории линейного и выпуклого программирования. Основные теоремы 1 и 3 доказаны Куном и Такером в 1950 г.; эти же авторы подчеркнули существенную роль условия регулярности, введенного ранее Слейтером.

Двойственная задача первоначально была построена для задачи линейного программирования (см. ссылки в гл. 10). Позже это было проделано для квадратичного программирования, хотя здесь разные авторы по-разному вводили понятие двойственной задачи. Еще больший разнобой царил в теории двойственности для общих задач выпуклого программирования. В ряде монографий и учебников можно найти, например, определение двойственной задачи, в котором фигурируют  $y$  и  $x$ . Излагаемая в п. 3 теория двойственности разработана Е. Г. Гольштейном, в монографии которого [9.4] можно найти много более общих результатов (см. также [9.9]). Современный подход к двойственности, связанный с сопряженными функциями и возмущениями общего вида, изложен в монографиях Рокафеллера [5.13] и А. Д. Иоффе и В. М. Тихомирова [9.12].

Общая постановка проблемы устойчивости можно найти в [9.4, 9.9, 8.3, 8.5].

**§ 2.** Теорема 1 — один из основных результатов в теории нелинейного программирования — является непосредственным обобщением правила множителей Лагранжа. Долгое время она наряду с теоремой 1 § 1 называлась теоремой Куна — Таккера. В действительности ранее этот результат был получен Джоном Однаком. Однако, как выяснилось позже, условия экстремума для гладких задач (и первого, и второго порядков) были впервые найдены Карушем еще в 1939 г. Полная драматизма и поучительная история открытия «теоремы Куна — Джона — Куна — Таккера» описана в статье Куна [9.25], там же

В качестве приложения опубликована рукопись Каруша. Наиболее подробно различные условия экстремума и условия регуляризации между ними исследованы в книге Манасаряна [9.26] (см. также [0.20]). Современные схемы получения общих условий экстремума как для гладких, так и для неоднадцатых задач в бесконечномерном случае развиваются весьма интенсивно и достигли высокого уровня абстракции (см. [9.1, 9.3, 9.7, 9.12, 9.28]).

Условия устойчивости типа теоремы 9 изучались Фиакко и Мак-Кормиком [0.20].

**§ 3.** Работы Зойтейделя, относящиеся к концу 50-х — началу 60-х годов, поглощены в его монографии [0.7]. В этой книге приводится общая конструкция методов возможных направлений, даны многочисленные модификации и конкретизация общих алгоритмов для частных задач. Независимо от Зойтейделя близкие схемы были предложены и исследованы С. И. Зуховским, Р. А. Поляком и М. Е. Примаком [0.14]. Одной из первых реализаций метода возможных направлений был метод проекции градиента Розена [9.31]. В работах [0.7, 9.31, 9.11] даны и доказательства сходимости методов.

Метод линеаризации вместе с его обоснованием был впервые предложен Б. Н. Пшеничным [8.10]. Его результаты подробно изложены в монографии [0.15]. Теорема 1, относящаяся к сходимости метода с постоянным  $Y$ , принадлежит А. Антипину [9.2].

Метод множителей Лангренка был одним из первых численных методов выпуклого программирования. Его исследование посвящена монография Эрроу, Гурвица и Узлави [0.23]. Авторы рассматривали в основном непрерывные варианты метода (траектория движения описывается дифференциальными разностными уравнениями), и лишь одна глава в [0.23], написанная Узлавой, относится к дискретному методу. В ней исследуется метод (13) и утверждается, что если  $f(x)$  строго выпукла, то для всякой начальной точки и любого  $\varepsilon > 0$  найдется достаточно малое  $\gamma \geq 0$ , такое, что метод приведет в  $\varepsilon$ -окрестность решения. К сожалению, в доказательстве этого утверждения (теорема 1, гл. 10 [0.23]) допущена принципиальная ошибка. На этом пути удается лишь доказать [0.6], что найдется последовательность  $x^*$ , попадающая в окрестность  $x^*$ . Сходимость метода типа (13) доказана в [9.14] иным способом. Метод (16) также предложен в [0.23], там же доказана его сходимость в некоторых случаях.

Модифицированные функции Лагранжа для задач выпуклого программирования введены и исследованы в работах Бергесекса, Вержникского, Е. Г. Гольштейна, Рокафеллера, Н. В. Третьякова [9.6, 9.18, 9.20, 9.30, 9.32]. Итерационные методы типа (20) для линейных ограничений изучал А. С. Антипин [8.1]. Результат, частным случаем которого является теорема 2, получен Г. Д. Маистровским [9.15], см. также [9.5]. Теорема 3 доказана Н. В. Третьяковым [9.18] и Рокафеллером [9.30].

Метод одновременного решения прямой и двойственной задачи (25) и близкая к нему двойственная схема метода линеаризации изучалась в [9.16, 9.17].

Методам штрафов, барьера и близким к ним посвящены монографии [0.20, 9.22] и обзор [9.19]. Оценки скорости сходимости для метода (26) даны в [9.8]. Результаты типа теорем 4, 5 можно найти в [0.20]. Метод спиги штрафов был введен Вержбником [9.32], позже оказалось, что он совпадает с методом модифицированной функции Лагранжа. Метод подбора  $f_k^*$  был предложен Моррисоном [9.27]. Другие способы пересчета  $f_k$  и результаты о сходимости можно найти в [9.13, 9.24].

Обобщение субградиентного метода на задачи выпуклого программирования однной задаче безусловной минимизации с негладкими штрафами (36) отмечалась рядом авторов [9.8, 9.29]. С обобщением метода отскакивающих гиперплоскостей и с другими методами, упомянутыми в п. 5, можно ознакомиться по работам, которые цитировались в § 4 гл. 5,

**§ 4.** Методы решения общих задач нелинейного программирования (1) рассматриваются в монографиях [0.15, 0.20, 0.22] и обзоре [0.14]. Результаты типа теорем 1, 2 о сходимости метода линеаризации (для нескольких иного правила выбора  $V$ ) получены в [0.15]. Метод Бьютона в форме (9) предложен Робинсоном [8.16]. Вычислительные схемы квазиньютоновских методов приведены в [0.22, 9.21, 9.23].

## К ГЛАВЕ 10

Существует ряд учебников и монографий, посвященных в основном задаче линейного программирования [0.7, 9.10, 10.1—10.5, 10.7, 10.11, 10.14, 10.16, 10.17]. Основное внимание в них уделяется алгоритмической реализации симплекс-метода и его модификаций.

**§ 1.** Заслуга постановки задачи линейного программирования принадлежит Л. В. Канторовичу [10.6]. С тех пор этой задаче было посвящено огромное количество исследований (см., например, библиографию в [10.2, 10.5]).

Результаты типа теоремы 1 получены Карапедори и Вейлем. В разработку условий оптимальности внесли вклад Л. В. Канторович, Данциг, фон Нейман, Гейль, Кун, Таккер (см. работы в сборнике [10.10]). Теорема 12 для частных случаев сформулирована в [0.23, 10.12], для общего — в [5.21].

**§ 2.** Симплекс-метод, как уже отмечалось, предложен Данцигом. Подробное изложение вычислительной схемы и модификаций можно найти в [0.7, 9.10, 10.1—10.3, 10.5, 10.7, 10.11, 10.14, 10.16, 10.17].

**§ 3.** Краткость метода модифицированной функции Лагранжа показана автором и Н. В. Третьяковым [10.12]. Решать задачи линейного программирования, сводя их к безусловной минимизации кусочно-линейных функций, первым предложил Н. З. Шор в начале 60-х годов (см. [5.15]). Работа Л. Г. Хачина [10.15], о которой говорится в § 3, опубликована недавно (в 1979 г.), но уже успела оказать большое влияние на взгляды специалистов по линейному программированию. Результаты типа теоремы 5 получены А. С. Антипиным [8.1] и Г. Д. Маистровским [9.15]. Метод «фиктивной игры» для решения матричных игр предложен Брауном (см. [10.4]). Общая формулировка и обоснование этого метода даны В. А. Волконским с соавторами [2.1]. Помимо описанных в тексте, существует и много других итеративных методов линейного программирования [0.7, 0.23, 10.4, 10.8, 10.13]. Большой опыт численного исследования одного из таких методов отражен в книге Р. П. Федоренко [0.19].

**§ 4.** Разработка алгоритмы квадратичного программирования (разработанные к началу 60-х годов) описаны в книгах [0.7, 10.9].

## К ГЛАВЕ 11

Прикладным задачам оптимизации посвящена монография [11.33] и обзор [11.34], много примеров можно найти в книгах [0.1, 0.16, 0.21].

**§ 1.** Постановка и методы решения классической задачи оценки параметров описываются в любом курсе математической статистики, например, [4.3].

Обзор методов для регрессионных задач дан в [11.3, 11.18]. Результаты вычислений для нелинейного метода наименьших квадратов приведены в [11.22]. Идеология рабочего оценивания изложена Хубером (см. обзор [11.13]). Метод (18) описан в [11.8], метод (20) — в [11.32]. Общие статистические подходы к проблеме восстановления зависимостей по экспериментальным данным исследовал В. Н. Вапник [11.5].

Широкую популярность методы адаптации завоевали благодаря работам Я. З. Цыпкина [4.9]. Рекуррентный вариант метода максимума правдоподобия (22), (23) предложен Сакрисоном (см. [2.8]). Рекуррентные оценки для регрессионных задач исследованы Я. З. Цыпкиным и автором [11.23]. Задачи подбора моделей по экспериментальным данным рассматривались в [11.3, 11.29]. С оценкой параметров в динамических моделях можно ознакомиться по работам, которые цитировались в § 4 гл. 5,

комиться по книге [11], а с проблемами оптимального планирования эксперимента — по [11.19, 11.27].

**§ 2.** Примеры прикладных задач оптимального проектирования содержатся в работах [11.6, 11.7, 11.20, 11.30, 6.3, 6.11]. Задачам геометрического программирования посвящена книга [11.10].

Подробное исследование разнотообразных задач распределения ресурсов содержится в [11.9]. Динамическое программирование, разработанное Беллманом [11.4] (см. также [11.2]), представляет собой специальный метод решения задач оптимизации, имеющих динамическую структуру. Идеи метода могут быть использованы и для некоторых других задач, например (8).

Задачи оптимизации в экономике впервые исследовал Л. В. Канторович [10.6]. Многочисленные примеры постановок экономических задач содержатся в [11.15, 11.21, 10.4—10.6, 10.16]. Методом декомпозиции посвящены работы [11.17, 11.26].

Различные постановки задач стохастического программирования и методы их решения рассматриваются в монографиях Ю. М. Ермольева [4.2, 11.12] и Д. Б. Юдина [4.11].

Задачи на минимакс интенсивно исследовалась В. Ф. Демьяновым, В. В. Федоровым и другими авторами [5.2, 5.3, 11.28, 2.9]. Постановки задач оптимизации в условиях неопределенности, отличные от минимаксных, рассматривал Ю. Б. Гермейер [11.8].

С проблемами экстремального регулирования можно ознакомиться по книгам [0.16, 11.14]. Теории и численным методам решения дискретных задач оптимального управления посвящена обширная литература, отметим лишь [11.2, 11.4, 11.24, 0.2, 0.10, 0.13, 0.19].

**§ 3.** Связь теории приближений с теорией оптимизации обсуждается в работах [11.16, 5.12]. Некоторые геометрические задачи на экстремум приведены в [9.1, 9.12]. Систематическое изучение взаимосвязей между теорией электрических цепей и математическим программированием предпринято Денисом [11.1]. Книга Б. С. Разумихина [0.13] посвящена использованию вариационных принципов физики для построения численных методов оптимизации. Ряд примеров вариационных задач механики приведен в монографии Ф. Л. Черноуско и Н. В. Баничука [11.31].

## К ГЛАВЕ 12

**§ 1.** Методические вопросы постановки и решения оптимизационных задач редко затрагиваются в литературе по оптимизации. Общая идеология взаимоотношений математика-вычислителя и практика-заказчика хорошо объясняется в книге Н. С. Бахвалова по численным методам [1.1]. Процесс решения реальных задач оптимизации (преимущественно задач оптимального управления) на большом материале иллюстрируется в монографии Р. П. Федоренко [0.19]. В этой связи полезны также книги по исследованию операций [11.8, 12.6].

**§ 2.** Требования к программам нелинейного программирования сформулированы в [12.16]. Тексты программ безусловной минимизации гладких функций имеются в [0.21, 12.12, 12.13], негладких функций — в [11.6], линейного программирования — в [10.14, 12.12—12.15], квадратичного программирования — в [12.14], нелинейного программирования — в [0.21, 12.12], а также в изданиях, упоминаемых в тексте. Обзор имеющихся программ линейного программирования дан в [12.4] (по состоянию на 1969 г.) и в гл. 16 [12.6] (современное состояние) (см. также [12.7, 12.8, 12.15]). Существующие диагностические системы оптимизации приведены в гл. 15 [12.6], конкретная диалоговая система описана в [0.10]. Комплекс созданных под руководством автора программ оптимизации охарактеризован в [12.1, 12.2]. Полное перечисление имеющихся на Западе программ нелинейного программирования содержится в статье [12.16].

**§ 3.** Обсуждение критериев сравнения алгоритмов имеется в [0.21]; там же приведена специальная таймер-программа для сопоставления быстродействия различных ЭВМ. Унифицированные требования к публикациям по сравнению численных алгоритмов оптимизации, разработанные специальной комиссией, сформулированы в [12.10].

Первым систематическим сборником тестовых задач и исследованием разнотообразных методов на этих задачах явилась работа Колвиля [12.9]. С тех пор многие из введенных в эту работу тестов стали классическими (например, функция Розенброка) и с их помощью анализируются новые программы и методы оптимизации. Большой набор тестовых задач на безусловный и условный экстремум приведен в книге Химмельбала [0.21]; одновременно там описаны результаты численных экспериментов по проверке многих методов с помощью этих тестов. Тесты по безусловной минимизации неплавких функций даны в [5.20]. Большой опыт решения тестовых задач неплавких функций программирования описан в статье В. А. Скокова [12.5]. Тесты, относящиеся к задачам геометрического программирования, собраны в работе [12.11]. Важный пример трудной для симплекс-метода задачи линейного программирования приведен в [10.18], на его основе построена задача 11. В целом нужно сказать, что актуальная работа по разработке тестов различной сложности для различных классов задач, их систематизации и проверке алгоритмов на этих тестах еще далека от завершения.

## К ГЛАВЕ 1 (см. также 0.1—0.3, 0.8, 0.10, 0.12, 0.13, 0.15, 0.17, 0.21)

- ЛИТЕРАТУРА**
- ОСНОВНЫЕ МОНОГРАФИИ, УЧЕБНИКИ И ОБЗОРЫ**
- 0.1. Аоки М. Введение в методы оптимизации. — М.: Наука, 1977.
  - 0.2. Васильев Ф. П. Лекции по методам решения экстремальных задач. — М.: МГУ, 1974.
  - 0.3. Габасов Р., Кириллова Ф. М. Методы оптимизации. — Минск: БГУ, 1975.
  - 0.4. Демьянов В. Ф., Рубинов А. М. Приближенные методы решения экстремальных задач. — Л.: ЛГУ, 1968.
  - 0.5. Ермолович Ю. М. Методы решения нелинейных экстремальных задач. — Кибернетика, 1966, № 4, с. 1—17.
  - 0.6. Зантиль У. Нелинейное программирование. Единый подход. — М.: Радио, 1973.
  - 0.7. Зейтендик Г. Методы возможных направлений. — М.: ИЛ, 1963.
  - 0.8. Карманов В. Г. Математическое программирование. — М.: Наука, 1975.
  - 0.9. Левитин Е. С., Поляк Б. Т. Методы минимизации при наличии ограничений. — Журн. вычисл. матем. и матем. физ., 1966, т. 6, № 5, с. 787—823.
  - 0.10. Моисеев Н. Н., Иванилов Ю. П., Столюрова Е. М. Методы оптимизации. — М.: Наука, 1978.
  - 0.11. Немировский А. С., Юдин Д. Б. Сложность задач и эффективность методов оптимизации. — М.: Наука, 1980.
  - 0.12. Оргела Д.Ж., Рейнболдт В. Итерационные методы решения нелинейных систем уравнений со многими неизвестными. — М.: Мир, 1975.
  - 0.13. Поляк Э. Численные методы оптимизации. Единый подход. — М.: Мир, 1974.
  - 0.14. Поляк Б. Т. Методы минимизации при наличии ограничений. — В кн.: Итоги науки и техники. Матем. анализ. Т. 12. М.: ВИНИТИ, 1974, с. 147—197.
  - 0.15. Пшеничный Б. Н., Даилин Ю. М. Численные методы в экстремальных задачах. — М.: Наука, 1975.
  - 0.16. Растигин Л. А. Системы экстремального управления. — М.: Наука, 1974.
  - 0.17. Сеа Ж. Оптимизация. Теория и алгоритмы. — М.: Мир, 1973.
  - 0.18. Уайлд Д. Дж. Методы поиска оптимума. — М.: Наука, 1967.
  - 0.19. Федоренко Р. П. Приближенное решение задач оптимального управления. — М.: Наука, 1978.
  - 0.20. Фиакко А., Мак-Кормик Дж. Нелинейное программирование: методы последовательной безусловной минимизации. — М.: Мир, 1972.
  - 0.21. Химельбаум Д. Практическое нелинейное программирование. — М.: Мир, 1975.
  - 0.22. Чистенные методы условной оптимизации/Под ред. Ф. Гилла, У. Моррея. — М.: Мир, 1977.
  - 0.23. Эрроу К. Дж., Гурвиц Л., Узаза Х. Исследования по линейному и нелинейному программированию. — М.: ИЛ, 1962.
- К ГЛАВЕ 2  
(см. также 0.6, 0.12, 0.13)**
- 2.1. Беленький В. З., Волконский В. А. и др. Итеративные методы в теории игр и программирования. — М.: Наука, 1974.
  - 2.2. Беллман Р. Введение в теорию матриц. — М.: Наука, 1969.
  - 2.3. Вазан М. Стохастическая аппроксимация. — М.: Мир, 1972.
  - 2.4. Евтушенко Ю. Г., Жалан В. Г. Применение метода функций Липунова для исследования сходимости численных методов. — Журн. вычисл. матем. и матем. физ., 1975, т. 15, № 1, с. 101—112.
  - 2.5. Карапин С. Математические методы в теории игр, программировании и экономике. — М.: ИЛ, 1964.
  - 2.6. Кушнер Г. Дж. Стохастическая устойчивость и управление. — М.: Мир, 1969.
  - 2.7. Любич Ю. И., Майстрюковский Г. Д. Общая теория динамических процессов для выпуклых функционалов. — УМН, 1970, т. 25, № 1, с. 57—112.
  - 2.8. Невельсон М. Б., Хасьминский Р. З. Стохастическая аппроксимация и рекуррентное определение. — М.: Наука, 1972.
  - 2.9. Нурмиский Е. А. Чистые методы решения детерминированных и стохастических минимаксных задач. — Киев: Наукова Думка, 1979.
  - 2.10. Островский А. М. Решение уравнений и систем уравнений. — М.: ИЛ, 1963.
  - 2.11. Поляк Б. Т. Сходимость и скорость сходимости итеративных стохастических алгоритмов. I. Общий случай. — Автом. и телемех., 1976, № 12, с. 83—94.
  - 2.12. Blum J. B. Multidimensional stochastic approximation method. — Ann. Math. Stat., 1954, v. 25, № 4, p. 737—742.

## К ГЛАВЕ 3

(см. также 0.1—0.3, 0.8, 0.10, 0.11—0.13, 0.15—0.17, 0.20, 0.22)

- 3.1. Данилин Ю. М. Об одном классе алгоритмов минимизации со сперхлинейной сходимостью. — Журн. вычисл. матем. и матем. физ., 1974, т. 14, № 3, с. 598—609.
- 3.2. Данилин Ю. М. Скорость сходимости методов сопряженных направлений. — Кибернетика, 1977, № 6, с. 97—105.
- 3.3. Дамбровская А. П. Симплекский поиск. — М.: Энергия, 1979.
- 3.4. Катковник В. Я. Линейные оценки и стохастические задачи оптимизации. — М.: Наука, 1976.
- 3.5. Лавров С. С. Применение барицентрических координат для решения некоторых вычислительных задач. — Журн. вычисл. матем. и матем. физ., 1964, т. 4, № 5, с. 905—911.
- 3.6. Марчук Г. И. Методы вычислительной математики. — 2-е изд. — М.: Наука, 1980.
- 3.7. Поляк Б. Т. Об одном способе ускорения сходимости итерационных методов. — Журн. вычисл. матем. и матем. физ., 1964, т. 4, № 5, с. 791—803.
- 3.8. Поляк Б. Т. Методы минимизации функций многих переменных: обзор. — Эконом. и матем. методы, 1967, т. 3, № 6, с. 881—902.
- 3.9. Поляк Б. Т. Метод сопряженных градиентов в задачах на экстремум. — Журн. вычисл. матем. и матем. физ., 1969, т. 9, № 4, с. 807—821.
- 3.10. Саульев В. К., Самойлова И. И. Приближенные методы безусловной оптимизации функций многих переменных. — В кн.: (Итоги науки и техники) Матем. анализ Т. 11. М.: ВИНИТИ, 1973, с. 91—128.
- 3.11. Фаддеев Д. К., Фаддеева В. Н. Вычислительные методы линейной алгебры. — 2-е изд. — М.: Физматгиз, 1963.
- 3.12. Daniel J. W. The approximate minimization of functionals. — Englewood Cliffs: Prentice Hall, 1971.
- 3.13. Dennis J. E., Moré J. J. Quasi-Newton methods, motivation and theory. — SIAM Review, 1977, v. 19, № 1, p. 46—89.
- 3.14. Fletcher R., Powell M. J. D. A rapidly convergent descent method for minimization. — Comput. J., 1963, v. 6, № 2, p. 163—168.
- 3.15. Fletcher R., Reeves C. M. Function minimization by conjugate gradients. — Comput. J., 1964, v. 7, № 2, p. 149—154.
- 3.16. Hestenes M. R. Conjugate direction methods in optimization. — N. Y., Heidelberg; Berlin: Springer, 1980.
- 3.17. Hestenes M. R., Stiefel E. Methods of conjugate gradients for solving linear systems. — J. Res. Nat. Bur. Stand. USA, 1952, v. 49, № 6, p. 409—436.
- 3.18. Huang H. G. Unified approach to quadratically convergent algorithms for function minimization. — J. Optim. Theory Appl., 1970, v. 5, № 6, p. 405—423.
- 3.19. Jacobson D., Oksman W. An algorithm that minimizes homogeneous function of  $n$  variables in  $n+2$  iterations and rapidly minimizes general functions. — J. Math. Anal. Appl., 1972, v. 38, № 3, p. 535—552.
- 3.20. Marquardt D. W. An algorithm for least-squares estimation of nonlinear parameters. — J. SIAM, 1963, v. 11, № 2, p. 431—441.
- 3.21. Spendley W., Hext G. R., Hinsworth F. R. Sequential application of simplex design in optimization and evolutionary operation. — Technometrics, 1962, v. 4, № 4, p. 441—461.
- 3.22. Wolfe P. The secant method for simultaneous nonlinear equations. — Comm. ACM, 1959, v. 2, № 1, p. 12—13.

## К ГЛАВЕ 4

(см. также 0.18, 1.1, 1.3, 2.1, 2.7—2.9, 3.4)

- 4.1. Ермольев Ю. М. О методе обобщенных стохастических градиентов и стохастических квазифайеровских последовательностях. — Кибернетика, 1969, № 2, с. 73—83.
- 4.2. Ермольев Ю. М. Методы стохастического программирования. — М.: Наука, 1976.
- 4.3. Крамер Г. Математические методы статистики. — 2-е изд. — Мир, 1975.
- 4.4. Первозванский А. А. Случайные процессы в нелинейных авторегressive системах. — М.: Физматгиз, 1962.
- 4.5. Поляк Б. Т. Сходимость методов возможных направлений в экспрессивных задачах. — Журн. вычисл. матем. и матем. физ., 1971, т. 11, № 4, с. 855—869.
- 4.6. Поляк Б. Т. Сравнение скорости сходимости одностошаговых и многошаговых алгоритмов оптимизации при наличии помех. — Технич. кибернет., 1977, № 1, с. 9—12.
- 4.7. Поляк Б. Т. К вопросу о сравнении градиентного метода и метода случайного поиска. — Автом. и вычисл. техн., 1977, № 3, с. 194—197.
- 4.8. Поляк Б. Т., Цыпкин Я. З. Оптимальные псевдоградиентные алгоритмы адаптации. — Докл. АН СССР, 1980, т. 250, № 5, с. 1084—1087.
- 4.9. Цыпкин Я. З. Адаптация и обучение в автоматических системах. — М.: Наука, 1968.
- 4.10. Цыпкин Я. З., Поляк Б. Т. Достигнутая точность алгоритмов адаптации. — Докл. АН СССР, 1974, т. 218, № 3, с. 532—535.
- 4.11. Юдин Д. Б. Задачи и методы стохастического программирования. — М.: Сов. радио, 1979.
- 4.12. Kiefer J., Wolfowitz J. Stochastic estimation of the maximum of a regression function. — Ann. Math. Stat., 1952, v. 23, № 3, p. 462—466.
- 4.13. Kushner H. J., Clark D. S. Stochastic approximation methods for constrained and unconstrained systems. — N. Y., Heidelberg; Berlin: Springer, 1978.
- 4.14. Robbins H., Monro S. A stochastic approximation method. — Ann. Math. Stat., 1951, v. 22, № 3, p. 400—407.
- К ГЛАВЕ 5
- (см. также 0.2, 0.5, 0.8, 0.9, 0.11, 0.18, 2.9, 3.4, 4.1, 4.2, 4.11)
- 5.1. Гупал А. М. Стохастические методы решения негладких экстремальных задач. — Киев: Наукова думка, 1979.

5.2. Демянов В. Ф. Минимакс: дифференцируемость по направлениям. — Л.: ЛГУ, 1974.

5.3. Демянов В. Ф., Малоземов В. Н. Введение в минимакс. — М.: Наука, 1972.

5.4. Еремин И. И. Релаксационный метод решения систем неравенств с выпуклыми функциями в левых частях. — Докл. АН СССР, 1965, т. 160, № 5, с. 994—996.

5.5. Ермольев Ю. М., Шор Н. З. О минимизации нелинейных функций. — Кибернетика, 1967, № 1, с. 101—102.

5.6. Зуховицкий С. И., Примак М. Е. О сходимости метода центрированных сечений для решения задачи выпуклых программирования. — Докл. АН СССР, 1975, т. 222, № 2, с. 273—276.

5.7. Левин А. Ю. Об одном алгоритме минимизации выпуклых функций. — Докл. АН СССР, 1965, т. 160, № 6, 1244—1247.

- 5.8. Поляк Б. Т. Один общий метод решения экстремальных задач. — Докл. АН СССР, 1967, т. 174, № 1, с. 33—36.
- 5.9. Поляк Б. Т. Минимизация неявных функционалов. — Журн. вычисл. матем. и матем. физ., 1969, т. 9, № 3, с. 509—521.
- 5.10. Примак М. Е. О сходимости модифицированного метода чебышевских центров решения задач выпуклого программирования. — Кibernetika, 1977, № 5, с. 100—102.
- 5.11. Пшеничный Б. Н. Выпуклый анализ и экстремальные задачи. — М.: Наука, 1980.
- 5.12. Пшеничный Б. Н. Необходимые условия экстремума. — М.: Наука, 1969.
- 5.13. Рокафеллер Р. Т. Выпуклый анализ. — М.: Мир, 1973.
- 5.14. Соколов В. А. Замечание к методам минимизации, использующим операцию растяжения пространства. — Кibernetika, 1974, № 4, с. 115—117.
- 5.15. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения. — Киев: Наукова думка, 1979.
- 5.16. Goffin J. L. On the convergence rates of subgradient optimization methods. — Math. Progr., 1977, v. 13, № 3, p. 329—347.
- 5.17. Kelley J. E. The cutting-plane method for solving convex programs. — J. SIAM, 1960, v. 8, № 4, p. 703—712.
- 5.18. Newmann D. J. Location of the maximum on unimodal surfaces. — J. ACM, 1965, v. 12, № 3, p. 395—398.
- 5.19. Nondifferentiable optimization/Math. Progr. Study 3. Eds. M. Bialiszki, P. Wolfe. — Amsterdam: North Holland, 1975.
- 5.20. Nonsmooth optimization/Proc. NASA Workshop, March 28—April 8, 1977. Eds. C. Lemaréchal, R. Mifflin. — Oxford: Pergamon Press, 1978.
- 5.21. Poljak B. T. Sharp minimum. — In: Generalized Lagrangians and applications. Proc. NASA Workshop, Dec. 17—22, 1979/Ed. A. Wierzbicki. — Oxford: Pergamon Press, 1982.
- К ГЛАВЕ 6**
- (см. также 0.2, 0.8, 0.11, 0.12, 0.16, 1.3, 1.4, 1.8)
- 6.1. Альберт А. Регрессия, псевдоинверсия и рекуррентное оценивание. — М.: Наука, 1977.
- 6.2. Бакушинский А. Б., Поляк Б. Т. О решении вариационных неравенств. — Докл. АН СССР, 1974, т. 219, № 5, с. 1038—1041.
- 6.3. Батышев Д. И. Поисковые методы оптимального проектирования. — М.: Сов. радио, 1975.
- 6.4. Булатов В. П. Методы погружения в задачах оптимизации. — Новосибирск: Наука, 1977.
- 6.5. Гельфанд И. М., Цеглин М. Л. Принцип нелокального поиска в системах автоматической оптимизации. — Докл. АН СССР, 1961, т. 137, № 2, с. 295—298.
- 6.6. Емелин И. В., Красносельский М. А. К теории некорректных задач. — Докл. АН СССР, 1979, 244, № 4, с. 805—808.
- 6.7. Еремин И. И., Мазуров В. Д. Нестационарные процессы математического программирования. — М.: Наука, 1979.
- 6.8. Лаврентьев М. М. О некоторых некорректных задачах математической физики. — Новосибирск: СО АН СССР, 1962.
- 6.9. Левитин Е. С., Поляк Б. Т. Сходимость минимизирующих последовательностей в задачах на условий экстремум. — Докл. АН СССР, 1966, т. 168, № 5, с. 997—1000.
- 6.10. Морозов В. А. О регуляризации некоторых классов экстремальных задач. — В кн.: Вычисл. методы и прогр. Т. 12. М.: МГУ, 1969, с. 24—37.
- 6.11. Монкус И. Б. Многокстремальные задачи в проектировании. — М.: Наука, 1967.
- К ГЛАВЕ 7**
- (см. также 0.2, 0.4, 0.8, 0.9, 0.13—0.15, 3.4, 3.9, 4.1, 4.2, 4.5, 5.1, 5.8—5.12, 5.15, 5.21)
- 7.1. Гупал А. М., Баженов Л. Г. Стохастический метод линеаризации. — Кibernetika, 1972, № 3, с. 116—117.
- 7.2. Bertsekas D. P. On the Goldstein—Levitin—Polyak gradient projection method. — IEEE Trans. Autom. Contr., 1976, v. 21, № 2, p. 174—184.
- 7.3. Cannon M. D., Cullum C. D. A tight upper bound on the rate of convergence of the Frank—Wolfe algorithm. — SIAM J. Contr., 1968, v. 6, № 4, p. 509—516.
- 7.4. Dunn J. C. Rates of convergence for conditional gradient algorithms near singular and nonsingular extrema. — SIAM J. Contr. Optim., 1979, v. 17, № 2, p. 187—211.
- 7.5. Frank M., Wolfe Ph. Algorithm for quadratic programming. — Naval Res. Log. Quart., 1956, v. 3, № 1—2, p. 95—110.
- 7.6. Goldstein A. A. Convex programming in Hilbert space. — Bull. Amer. Math. Soc., 1964, v. 70, № 5, p. 709—710.
- К ГЛАВЕ 8**
- (см. также 0.13, 0.15, 0.20—0.23)
- 8.1. Антипов А. С. Метод градиентного типа для отыскания седловой точки модифицированной функции Лагранжа. — Эконом. и матем. методы, 1977, т. 13, № 3, с. 560—565.
- 8.2. Волин Ю. М., Остроуский Г. М. Метод штрафных функций и необходимые условия оптимальности. — В кн.: Управляемые системы. Вып. 9. Новосибирск: СО АН СССР, 1971, с. 43—51.
- 8.3. Левитин Е. С. О дифференцируемости по параметру оптимального значения параметрических задач математического программирования. — Кibernetika, 1976, № 1, с. 44—59.
- 8.4. Люстерник Л. А., Соболев В. И. Элементы функционального анализа. — 2-е изд. — М.: Наука, 1965.

- 8.5. Первозванский А. А., Гайгори В. Г. Декомпозиция, агрегирование и приближенная оптимизация. — М.: Наука, 1979.
- 8.6. Поляк Б. Т. Итерационные методы, использующие множители Лагранжа, для решения экстремальных задач с ограничениями типа равенств. — Журн. вычисл. матем. и матем. физ., 1970, т. 10, № 5, с. 1098—1106.
- 8.7. Поляк Б. Т. О скорости сходимости метода штрафных функций. — Журн. вычисл. матем. и матем. физ., 1971, т. 11, № 1, с. 3—11.
- 8.8. Поляк Б. Т. Методы решения задач на условный экстремум при наличии случайных помех. — Журн. вычисл. матем. и матем. физ., 1979, т. 19, № 1, с. 70—78.
- 8.9. Поляк Б. Т., Третьяков Н. В. Метод штрафных оценок для задач на условный экстремум. — Журн. вычисл. матем. и матем. физ., 1973, т. 13, № 1, с. 34—46.
- 8.10. Пшеничный Б. Н. Алгоритмы для общей задачи математического программирования. — Кибернетика, 1970, № 5, с. 120—125.
- 8.11. Bertsekas D. P. Multiplier methods: a survey. — Automatica, 1976, № 12, р. 133—145.
- 8.12. Bertsekas D. P. Constrained optimization and Lagrange multiplier methods. — N. Y.: London: Acad. Press, 1982.
- 8.13. Harker P. C., Byles J. D. A new method for the optimization of a nonlinear function subject to nonlinear constraints. — Comput. J., 1970, v. 13, № 2, р. 178—184.
- 8.14. Hestenes M. R. Multiplier and gradient methods. — J. Optim. Theory Appl., 1969, v. 4, № 5, р. 303—320.
- 8.15. Powell M. J. D. A method for nonlinear constraints in minimization problems. — In: Optimization/Ed. R. Fletcher. London: Acad. Press, 1969, р. 283—298.
- 8.16. Robinson S. M. A quadratically convergent algorithm for general nonlinear programming problems. — Math. Progr., 1972, v. 3, № 2, р. 145—156.
- 8.17. Wolfe P. Methods for nonlinear constraints. — In: Nonlin. Progr./Ed. J. Abadie. Amsterdam: North-Holland, 1967, р. 120—131.
- К ГЛАВЕ 9  
(см. также 0.1—0.3, 0.5—0.11, 0.13—0.15, 0.17—0.23)
- 9.1. Алексеев В. М., Тихомиров В. М., Фомин С. В. Оптимальное управление. — М.: Наука, 1979.
- 9.2. Антилип А. С. Методы нелинейного программирования, основанные на прямой и двойственной модификации функции Лагранжа. Принт. — М.: ВНИИСИ, 1979.
- 9.3. Гирсанов И. В. Лекции по математической теории экстремальных задач. — М.: МГУ, 1970.
- 9.4. Гольштейн Е. Г. Теория двойственности в математическом программировании. — М.: Наука, 1971.
- 9.5. Гольштейн Е. Г. О сходимости градиентного метода отыскания седловых точек модифицированных функций Лагранжа. — Эконом. и матем. методы, 1977, т. 13, № 2, с. 322—329.
- 9.6. Гольштейн Е. Г., Третьяков Н. В. Модифицированные функции Лагранжа. — Эконом. и матем. методы, 1974, т. 10, № 3, с. 568—591.
- 9.7. Дубовик А. Я., Милютин А. А. Задачи на экстремум при наличии ограничений. — Докт. АН СССР, 1963, т. 149, № 4, с. 759—762.
- 9.8. Еремин И. И. О методе штрафов в выпуклом программировании. — Кибернетика, 1967, № 4, с. 63—67.
- 9.9. Еремин И. И., Астафьев Н. Н. Введение в теорию линейного и выпуклого программирования. — М.: Наука, 1976.
- 9.10. Зуховицкий С. И., Адельева Л. И. Линейное и выпуклое программирование. — М.: Наука, 1967.
- 9.11. Зуховицкий С. И., Поляк Р. А., Примак М. Е. Алгоритм для решения задачи выпуклого программирования. — Докл. АН СССР, 1963, т. 153, № 5 с. 991—994.
- 9.12. Иоффе А. Д., Тихомиров В. М. Теория экстремальных задач. — М.: Наука, 1974.
- 9.13. Лебедев В. Ю. О сходимости метода нагруженного функционала в задаче выпуклого программирования. — Журн. вычисл. матем. и матем. физ., 1977, т. 17, № 3, с. 765—768.
- 9.14. Майстронский Г. Д. О градиентных методах отыскания седловых точек. — Эконом. и матем. методы, 1976, т. 12, № 5, с. 917—929.
- 9.15. Мастровский Г. Д. О скорости сходимости градиентного метода для модифицированной функции Лагранжа. — Эконом. и матем. методы, 1979, т. 15, № 2, с. 380—386.
- 9.16. Поляк Р. А. Алгоритм для одновременного решения прямой и двойственной задачи выпуклого программирования. — В кн.: Экон. кибернет. и исслед. опер. в эконом. М.: Экономика, 1966, с. 53—64.
- 9.17. Пшеничный Б. Н. Двойственный метод в экстремальных задачах. I, II. — Кибернетика, 1965, № 3, с. 89—95; № 4, с. 64—69.
- 9.18. Третьяков Н. В. Метод штрафных оценок для задач выпуклого программирования. — Труды семинара. Киев: ИК АН УССР, 1973, № 526—540.
- 9.19. Эльстер К.Х., Россман Х. Решение нелинейных оптимизационных задач с помощью штрафных и барьерных функций. — В кн.: Примен. исслед. опер. в эконом. М.: Экономика, 1977, с. 95—161.
- 9.20. Bertsekas D. P. On the method of multipliers for convex programming. — IEEE Trans. Autom. Contr., 1975, v. 20, p. 385—388.
- 9.21. Garcia-Palomares U. M., Mangasarian O. L. Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems. — Math. Progr., 1976, v. 11, № 1, р. 1—13.
- 9.22. Grossmann Ch., Kaplan A. A. Strafmethoden und modifizierte Lagrangepunktionen in der nichtlinearen Optimierung. — Leipzig: BSB B. G. Teubner Verlag, 1979.
- 9.23. Han S. P. Superlinearly convergent variable metric algorithm for general nonlinear programming problem. — Math. Progr., 1976, v. 11, № 3, р. 263—282.
- 9.24. Kowalik J., Osborne M. R., Ryan D. M. A new method for constrained optimization problems. — Oper. Res., 1969, v. 7, № 6, р. 973—983.
- 9.25. Kuhn H. W. Nonlinear programming: a historical view. — In: Nonlinear programming. SIAM-AMS Proc. Vol. 9. Providence: Amer. Math. Soc., 1976, р. 1—26.
- 9.26. Mangasarian O. L. Nonlinear programming. — N. Y.: McGraw-Hill, 1969.
- 9.27. Morrison D. D. Optimization by least squares. — SIAM J. Numer. Anal., 1968, v. 5, № 1, р. 83—88.
- 9.28. Neustadt L. W. Optimization. A theory of necessary conditions. — Princeton: Princeton Univ. Press, 1976.
- 9.29. Pietrzykowski T. An exact potential method for constrained programming. — SIAM J. Numer. Anal., 1969, v. 6, № 2, р. 299—304.
- 9.30. Rockafellar R. T. The multiplier method of Hestenes and Powell applied to convex programming. — J. Optim. Theory Appl., 1973, v. 12, р. 555—562.
- 9.31. Rosen J. B. The gradient projection method for nonlinear programming. I. — SIAM J., 1960, v. 8, № 1, р. 180—217.
- 9.32. Wierzbicki A. P. A penalty function shifting method in constrained static optimization and its convergence properties. — Archiw. Autom. i Techn., 1971, v. 16, № 4, р. 395—416.

## К ГЛАВЕ 10

(см. также 0.7, 0.19, 0.22, 2.1, 5.15, 5.21, 6.2, 8.1, 8.12, 9.9, 9.10, 9.15)

ные методы линейного программирования. — М.: Наука, 1977.

10.2. Гасс С. Линейное программирование. — М.: Физматиз, 1961.

10.3. Габасов Р., Кирilloва Ф. М. Методы линейного програм-

мирования. — Минск: БГУ, 1977, т. 2; 1980, т. 3.

10.4. Гольштейн Е. Г., Юдин Д. Б. Новые направления в линей-

ном программировании. — М.: Сов. радио, 1966.

10.5. Данциг Дж. Линейное программирование, его применения и об-

общения. — М.: Прогресс, 1966.

10.6. Канторович Л. В. Математические методы в организации и

планировании производства. — Л: ЛГУ, 1939.

10.7. Карпелевич Ф. И., Садовский Л. Е. Элементы линейной

алгебры и линейного программирования. — М.: Физматиз, 1963.

10.8. Корпелевич Г. М. Экстраградиентный метод для отыскания

седловых точек и других задач. — Эконом. и матем. методы, 1976, т. 12, № 4,

с. 47—756.

10.9. Конди Г. П., Крелле В. Нелинейное программирование. — М:

Сов. радио, 1963.

10.10. Линейные неравенства и смежные вопросы/Под ред. Г. Куна,

А. Таккера. — М.: ИЛ, 1959.

10.11. Ниг И. В. Линейное программирование (с обсуждением некото-

рых нелинейных задач). — М.: МГУ, 1978.

10.12. Поляк Б. Т., Третьяков Н. В. Об одном итерационном ме-

тоде линейного программирования и его экономической интерпретации. — Эко-

ном. и матем. методы, 1972, т. 8, № 5, с. 740—751.

10.13. Разумихин Б. С. Физические модели и методы теории равновесия в программировании и экономике. — М.: Наука, 1975.

10.14. Романовский И. В. Алгоритмы решения экстремальных за-

дач. — М.: Наука, 1977.

10.15. Хачян Л. Г. Полиномиальный алгоритм в линейном програм-

мировании. — Докт. АН СССР, 1979, т. 244, № 5, с. 1003—1096.

10.16. Юдин Д. Б., Гольштейн Е. Г. Задачи и методы линейного

программирования. — М.: Сов. радио, 1961.

10.17. Юдин Д. Б., Гольштейн Е. Г. Линейное программирование.

Теория и конечные методы. — М.: Физматиз, 1963.

10.18. Klee V., Minty G. J. How good is simplex algorithm? — In: Inequalities-III/Ed. O. Shisha. N. Y., London: Acad. Press, 1972, p. 159—175.

## К ГЛАВЕ 11

(см. также 0.1, 0.19, 4.2—4.4, 4.9, 4.11, 6.3, 6.11, 6.13, 10.5, 10.6, 10.13)

11.1. Айдерсон Т. Статистический анализ временных рядов. — М.: Мир, 1976.

11.2. Арис Р. Дискретное динамическое программирование. — М.: Мир, 1969.

11.3. Бард И. Нелинейное оценивание параметров. — М.: Статистика, 1979.

11.4. Беллман Р. Динамическое программирование. — М.: ИЛ, 1960.

11.5. Бапник В. Н. Восстановление зависимостей по эмпирическим

данным. — М.: Наука, 1979.

11.6. Вычислительные методы выбора оптимальных проектных реше-

ний/Под ред. В. С. Михалевича. — Киев: Наукова думка, 1977.

11.7. Гемин В. И., Каган Б. М. Методы оптимального проекти-

рования. — М.: Энергия, 1980.

11.8. Гермейер Ю. Б. Введение в теорию исследования операций. — М.: Наука, 1971.

11.9. Гурин Л. С., Дымарский Я. С., Меркулов А. Д. Задачи и методы оптимального распределения ресурсов. — М.: Сов. радио, 1968.

11.10. Даффин Р., Петерсон Э., Зенер К. Геометрическое програм-

мирование. — М.: Мир, 1972.

11.11. Денисис Д. Б. Математическое программирование и электрические цепи. — М.: ИЛ, 1961.

11.12. Ермольев Ю. М., Ястребецкий А. И. Стохастические модели и методы в экономическом планировании. — М.: Наука, 1979.

11.13. Ершов А. А. Стабильные методы оценки параметров: обзор. — Автом. и телемех., 1978, № 8, с. 66—100.

11.14. Казакевич В. В., Родов А. Б. Системы автоматической опти-

мизации. — М.: Энергия, 1977.

11.15. Канторович Л. В. Экономический расчет наилучшего использо-

вания ресурсов. — М.: АН СССР, 1960.

11.16. Лоран П. Ж. Аппроксимация и оптимизация. — М.: Мир, 1976.

11.17. Лэслон Л. С. Оптимизация больших систем. — М.: Наука, 1975.

11.18. Мудров В. И., Кушко В. Л. Методы обработки измерений. — М.: Сов. радио, 1976.

11.19. Налимов В. В., Чернова Н. А. Статистические методы пла-

нирования экстремальных экспериментов. — М.: Наука, 1965.

11.20. Островский Г. М., Волин В. М. Методы оптимизации хими-

ческих реакторов. — М.: Химия, 1967.

11.21. Первозванский А. А. Математические модели в управлении производством. — М.: Наука, 1975.

11.22. Поляк Б. Т., Скоков В. А. Решение задач на минимум

суммы квадратов. — Эконом. и матем. методы, 1978, т. 14, № 6,

с. 1173—1180.

11.23. Поляк Б. Т., Цыпкин Я. З. Адаптивные алгоритмы оценивания

(сходимость, оптимальность, стабильность). — Автом. и телемех., 1979, № 3,

с. 71—84.

11.24. Пропой А. И. Элементы теории оптимальных дискретных процес-

сов. — М.: Наука, 1973.

11.25. Табак Д. Кую Б. Оптимальное управление и математическое

программирование. — М.: Наука, 1975.

11.26. Ульман С. Методы декомпозиции для решения задач оптимизации. — Таллин: Валгус, 1979.

11.27. Федоров В. В. Теория оптимального эксперимента. — М.: Наука, 1971.

11.28. Федоров В. В. Численные методы максимина. — М.: Наука, 1979.

11.29. Химмельбау Д. Анализ процессов статистическими метода-

ми. — М.: Мир, 1973.

11.30. Цирлин А. М. Вариационные методы выбора оптимальных пара-

метров аппаратов химической технологии. — М.: Машиностроение, 1978.

11.31. Чироуский Ф. Л., Баничук Н. В. Вариационные задачи ме-

ханики и управления (численные методы). — М.: Наука, 1973.

11.32. Bertsekas D. P. Approximation procedures based on the

method of multipliers. — J. Optim. Theory Appr., 1977, v. 23, № 4, p. 487—

510.

11.33. Blacken J., McCormick G. P. Selected applications of non-

linear programming. — N. Y.: Wiley, 1968.

11.34. Lasdon L. S., Waren A. D. Survey of nonlinear programming

applications. — Oper. Res., 1980, v. 28, № 5, p. 1029—1073.

К ГЛАВЕ 12  
(см. также 0.21, 10.14)

- 12.1. Б е л о в Е. Н. Алгоритмы и программы решения задач квадратичного и линейного программирования. — Эконом. и матем. методы, 1980, т. 16, № 1, 198—201.
- 12.2. Б е л о в Е. Н., П о л я к Б. Т., С к о к о в В. А. Комплекс программ оптимизации. — Эконом. и матем. методы, 1978, т. 14, № 4, см. 792—796.
- 12.3. Н е с т е р о в Ю. Е., С к о к о в В. А. К вопросу о тестировании алгоритмов безусловной минимизации. — В кн.: Численные методы решения задач линейной оптимизации. — М.: ЦЭМИ, 1980, с. 77—91.
- 12.4. М а л к о в У. Х. Обзор программ решения общей задачи линейного программирования. — Эконом. и матем. методы, 1969, т. 5, № 4, с. 594—597.
- 12.5. С к о к о в В. А. Некоторый вычислительный опыт решения задач нелинейного программирования. — В кн.: Матем. методы решения экономических задач, Сб. 7. М.: Наука, 1977, с. 51—69.
- 12.6. Современное состояние теории исследования операций/Под ред. Н. Н. Моисеева. — М.: Наука, 1979.
- 12.7. У и л к и н с о н Дж. Х., Р а й н ш К. Справочник алгоритмов на языке Алгол. Линейная алгебра. — М.: Машиностроение, 1976.
- 12.8. Computational practice in mathematical programming. Math. Progr. Study. Vol. 4.— Amsterdam: North Holland, 1975.
- 12.9. C o l v i l l e A. R. A comparative study of nonlinear programming codes. — In: Proc. Princeton Symp. Math. Progr./Ed. Kuhn H. W. Princeton: Princeton Univ. Press, 1970.
- 12.10. C r o w d e r H. P., D e m b o R. S., M u l v e y J. M. Reporting computational experiments in mathematical programming. — Math. Progr., 1978, v. 15, № 3, p. 316—329.
- 12.11. D e m b o R. S. A set of geometric programming test problems and their solutions. — Math. Progr., 1976, v. 10, № 2, p. 192—213.
- 12.12. K u e s t e r J. L., M i z e J. H. Optimization techniques with FORTRAN. — N. Y.: McGraw-Hill, 1973.
- 12.13. K u n z i H. P., T z s c h a c h H. G., Z e n d e r C. A. Numerical methods of mathematical optimization with ALGOL and FORTRAN programs. — N. Y.: Acad. Press, 1971.
- 12.14. L a n d A. H., P o w e l l S. Fortran codes for mathematical programming: linear, quadratic and discrete. — London; N. Y.: Wiley, 1973.
- 12.15. O r c h a r d - H a v y s W. Advanced linear programming computing techniques. — N. Y.: McGraw-Hill, 1968.
- 12.16. W a r r e n A. D., L a s d o n L. S. The status of nonlinear programming. — Oper. Res., 1979, v. 27, № 3, p. 431—456.

12.1. Б е л о в Е. Н. Алгоритмы и программы решения задач квадратичного и линейного программирования. — Эконом. и матем. методы, 1980, т. 16, № 1, 198—201.

12.2. Б е л о в Е. Н., П о л я к Б. Т., С к о к о в В. А. Комплекс программ оптимизации. — Эконом. и матем. методы, 1978, т. 14, № 4, см. 792—796.

12.3. Н е с т е р о в Ю. Е., С к о к о в В. А. К вопросу о тестировании алгоритмов безусловной минимизации. — В кн.: Численные методы решения задач линейной оптимизации. — М.: ЦЭМИ, 1980, с. 77—91.

12.4. М а л к о в У. Х. Обзор программ решения общей задачи линейного программирования. — Эконом. и матем. методы, 1969, т. 5, № 4, с. 594—597.

12.5. С к о к о в В. А. Некоторый вычислительный опыт решения задач нелинейного программирования. — В кн.: Матем. методы решения экономических задач, Сб. 7. М.: Наука, 1977, с. 51—69.

12.6. Современное состояние теории исследования операций/Под ред. Н. Н. Моисеева. — М.: Наука, 1979.

12.7. У и л к и н с о н Дж. Х., Р а й н ш К. Справочник алгоритмов на языке Алгол. Линейная алгебра. — М.: Машиностроение, 1976.

12.8. Computational practice in mathematical programming. Math. Progr. Study. Vol. 4.— Amsterdam: North Holland, 1975.

12.9. C o l v i l l e A. R. A comparative study of nonlinear programming codes. — In: Proc. Princeton Symp. Math. Progr./Ed. Kuhn H. W. Princeton: Princeton Univ. Press, 1970.

12.10. C r o w d e r H. P., D e m b o R. S., M u l v e y J. M. Reporting computational experiments in mathematical programming. — Math. Progr., 1978, v. 15, № 3, p. 316—329.

12.11. D e m b o R. S. A set of geometric programming test problems and their solutions. — Math. Progr., 1976, v. 10, № 2, p. 192—213.

12.12. K u e s t e r J. L., M i z e J. H. Optimization techniques with FORTRAN. — N. Y.: McGraw-Hill, 1973.

12.13. K u n z i H. P., T z s c h a c h H. G., Z e n d e r C. A. Numerical methods of mathematical optimization with ALGOL and FORTRAN programs. — N. Y.: Acad. Press, 1971.

12.14. L a n d A. H., P o w e l l S. Fortran codes for mathematical programming: linear, quadratic and discrete. — London; N. Y.: Wiley, 1973.

12.15. O r c h a r d - H a v y s W. Advanced linear programming computing techniques. — N. Y.: McGraw-Hill, 1968.

12.16. W a r r e n A. D., L a s d o n L. S. The status of nonlinear programming. — Oper. Res., 1979, v. 27, № 3, p. 431—456.

## ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Абсолютные ограничения** 96, 103, 144, 196, 221  
Активные ограничения 232, 240  
Аргументные задачи 164
- Безусловная минимизация** 9, 15, 346  
Блочные задачи 322
- Вариационное исчисление** 9, 328
- Вариационные принципы** 334
- Вонкругая функция 20, 226
- Выпуклая оболочка 115, 271
- Функция 19, 118, 226
- Выпуклое множество 114
- программирование 11, 225
- Выпуклый анализ 12, 114, 225
- Геометрическое программирование 317
- Глобальный минимум 166
- Градиент 15
- Градиентный метод 30, 63, 85, 97, 108, 150, 278
- Двойственность 235, 275, 297
- Двоястственные методы 210, 251, 267
- переменные 200, 232
- Динамическое программирование 11, 320, 329
- Допустимые направления 180
- точки 199, 225
- Единственность решения 26, 125, 208, 238,
- Идентификация 301
- Каноническая форма 276
- Касательный вектор 200
- Квадратичное программирование 206
- Квадратичновыпуклые методы 77, 103, 192, 289, 355
- эллипсоидов 142, 353
- Конус 228
- Крайняя точка 270
- Критериальные задачи 164
- Лемма Дубовикского — Милютина 229
- Карагеодори 116
- Ляпунова 45
- Моро — Рокафеллара 228
- об обращении матриц 78, 83
- Фаркаша 229
- Линейная складываемость 46
- Линейное программирование 11, 268, 358
- Найлучшее приближение 330
- Направление скорейшего спуска 30, 123
- Невырожденный минимум 26, 206, 245
- Нелинейное программирование 225, 354
- Мартингал 53
- Математическое программирование 11,
- Метод барциентрических координат 91, 348
- Метод барьера 255
- Метод градиента 78
- Метод Гаусса — Зейделя 90
- Иностранец 306, 351
- Лагранж 322
- ДЛФП 78, 349
- итеративной регуляризации 159
- Кифера — Вольфович 103
- Левенберга — Марквардта 66
- линейизаций 211, 249, 263
- Нестаджик оптимизации 114, 299, 352
- Ньютона 37, 38, 66, 100, 109, 191, 218, 294
- отсеивания 135
- Палмела 93, 348
- плавления 93
- поиска 87, 146
- полокомплектного спуска 88
- присоединенного градиента 217, 267
- проекции градиента 185, 248, 289
- субградиента 187, 259
- регуляризации 156, 163, 288
- с расщепленным пространства 143, 291
- скользящими 82
- скорейшего спуска 64, 349
- спрятанных градиентов 70, 86, 153, 194, 348
- тяжелого шарика 68, 102
- устойчивого градиента 188
- центров тяжести 138
- чебышевских центров 137
- эллиптических функций 13, 216, 255, 266,
- Минимакс 325
- Минимизирующая последовательность 26, 183
- Многогранное множество 270
- Многогранные методы 68, 102, 134
- Множества стрельбы 166, 220
- Множественный лагранжики 199, 232, 241
- Монотонированная функция Лагранжа 206, 252
- Монотонность 21, 120

- Неравенство Ценсена 20
  - Канторовича 65
  - Колмогорова 53
  - Крамера — Рао 112
- Нестационарные задачи 175
- Нормализация 248
- Нормальное решение 157
- Обусловленность 28, 314
- Овражная функция 28
- Овражный метод 171
- Ограничения общего вида 10, 225
  - типа равенств 10, 199
  - Однородные функции 84
- Опорный гиперболоскость 117
- Опорный вектор 179
- Оптимальное планирование 321
  - проектирование 317
  - распределение ресурсов 319
  - управление 11, 327
- Оптимальные методы 107
- Острый пикум 127, 184, 278
- Однородные множества 116
- Отделенные множества 96, 103, 145, 198, 221
- Погрешность вычислений 94
- Позиномы 317
- Правило множителей Лагранжа 199
  - принцип максимума 11, 327
  - сжимающих отображений 59
- Проекция 116
- Производная по направлению 16, 119
- Прокс-метод 158
- Прямые методы 87, 103, 210
- Псевдоградиентность 55
- Псевдообратная матрица 157
- Фишеровская информация 112, 303
  - Функция Лагранжа 200, 232, 241
    - Лапунова 49
    - среднего риска 95
- Разностные аппроксимации 87
- Рандомизация 149
- Репрессии 304
- Регуляризация 166
- Регуляризация минимума 199, 232
- Рекуррентные оценки 309
- Робастность 306, 311
- Сверхлинейная склонность 46
- Седловая точка 167, 234
- Сильная выпуклость 20, 126, 191
- Симплекс-метод 13, 281, 360
- Симплексный метод 90, 99
- Сопряженные векторы 72
  - конусы 229
- Стационарная точка 23, 166
- Стochasticеская аппроксимация 310
- Стochasticеское программирование 323
- Строгое выпуклость 20, 125, 183
- Субградиент 120, 123
- Существование решения 25, 125, 183, 238, 277, 298
- Сходимость случайных величин 52
- Теорема Вейтерштрасса 25, 183
  - Каруша — Джона 240
  - Куна — Танкера 231, 234
  - Листерника 201
  - Некирского — Юдина 140
  - о невыводной функции 61
  - Ферма 22
  - Хейли 333
- Тестовые задачи 345
- Транспортная задача 321
- Условие дополнительной неизвестности 232, 245
  - регулярности 241
  - Слейтера 231
  - экстремума 22, 124, 179, 199, 231, 240, 274, 296
- Устойчивость решения 26, 126, 183, 238, 245, 278

sakofen



SD