институт
математики
экономики и
информатики
ИМЭИ

ОПТИМИЗАЦИЯ, ИССЛЕДОВАНИЕ ОПЕРАЦИЙ И УПРАВЛЕНИЕ

ВЫПУСК 3

УНИВЕРСИТЕТСКИЙ УЧЕБНИК

А. В. Аргучинцев, А. И. Беников

ВВЕДЕНИЕ В ОПТИМИЗАЦИЮ

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Иркутский государственный университет»

Университетский учебник

А. В. Аргучинцев, А. И. Беников

Введение в оптимизацию

Учебное пособие

УДК 519.85
ББК 22.161.8я73
А79

Печатается по решению редакционно-издательского совета
Иркутского государственного университета

Рецензенты: д-р физ.-мат. наук, проф. В.А. Дыхта;
д-р физ.-мат. наук, проф. В.А. Срочко

Научные редакторы серии:
д-р физ.-мат. наук, проф. В.А. Дыхта;
д-р физ.-мат. наук, проф. В.А. Срочко

Аргучинцев А.В.

А79 Введение в оптимизацию: учеб. пособие / А.В. Аргучинцев, А.И. Беников. – Иркутск : Изд-во ИГУ, 2011. – Вып. 3. – 105 с. (Университетский учебник)

ISBN 978-5-9624-0537-7

Учебное пособие является вводным курсом в математическую теорию оптимизации. В сжатом виде дается обзор необходимых при изучении экстремальных задач знаний по векторам, матрицам, выпуклым множествам и функциям, дифференцируемости функций многих переменных. Изложены классические методы решения задач на безусловный и условный экстремум, простейшие численные методы решения задач одномерной оптимизации. Необходимый теоретический материал сопровождается иллюстративными примерами и задачами для самостоятельного решения.

Пособие предназначено для студентов математических, экономических и технических направлений и специальностей.

Библиогр. 19 назв. Ил. 25

УДК 519.85
ББК 22.161.8я73

ISBN 978-5-9624-0537-7

Оглавление

Предисловие

В современные учебные планы обучения студентов по математическим, экономическим и техническим направлениям и специальностям в обязательном порядке входят оптимизационные курсы, в которых изучаются теория и методы решения задач линейного и нелинейного программирования, вариационного исчисления и оптимального управления, динамического программирования и ряд других весьма важных для грамотного специалиста разделов. Освоение этих курсов невозможно без предварительного знания элементов математического анализа, линейной алгебры, обыкновенных дифференциальных уравнений.

Данное учебное пособие можно рассматривать как введение, содержащее минимально необходимый материал для изучения теории и методов решения экстремальных задач. Материал разбит на три главы.

В первой главе в сжатом виде дается обзор обязательных при изучении экстремальных задач знаний по векторам, матрицам, выпуклым множествам и функциям, дифференцируемости функций многих переменных. Рассмотрены определения выпуклых и вогнутых функций, а также соответствующие критерии.

Во второй главе приводятся определения и свойства максимума и минимума, супремума и инфимума, дается изложение методов решения задач на безусловный и условный экстремум, основанных на правиле множителей Лагранжа. Рассматривается как классическая задача с ограничениями типа равенств, так и общая задача математического программирования с ограничениями типа равенств, неравенств и включения.

В третьей, заключительной, главе дан краткий обзор численных методов поиска экстремума функций одной вещественной переменной. Данные методы весьма специфичны и отличаются от методов решения общих задач нелинейного программирования. С другой стороны, именно одномерная оптимизация очень часто используется в качестве вспомогательного инструмента при численном решении не только задач математического программирования, но и оптимального управления. Указанным обстоятельством объясняется включение данного раздела в вводный курс по оптимизации.

В каждом параграфе каждой главы дается сжатое теоретическое изложение сведений, содержащее необходимые определения, утверждения и теоремы. После этого разбираются примеры и приводится определенное количество задач и упражнений для самостоятельного решения. Задачи имеют различный уровень сложности.

Необходимо отметить, что краткое изложение основ теории ни в коем случае не освобождает студентов-математиков от подробного изучения материала, предусматривающего необходимость понимания доказательств соответствующих теорем и утверждений. Рекомендуемый список литературы приведен в конце пособия.

В книге используется автономная для каждой главы нумерация формул, определений, теорем, примеров, рисунков и задач. Символ ■ в тексте означает конец примера.

Глава 1
Множества и функции в n-мерном евклидовом пространстве

§ 1. Исходные понятия

1. Пусть n – натуральное число. Совокупность всевозможных упорядоченных систем n действительных чисел $(x_1, x_2, \ldots, x_n)$ называется *n–мерным вещественным пространством* и обозначается через R^n. Элементы $x = (x_1, x_2, \ldots, x_n)$ n–мерного вещественного пространства называются *векторами* или *точками*, а числа $x_1, x_2, \ldots, x_n$ – их *компонентами* или *координатами*.

Равенство векторов в n–мерном вещественном пространстве понимается как равенство их соответствующих компонент, т. е. вектор $x = (x_1, x_2, \ldots, x_n)$ равен вектору $y = (y_1, y_2, \ldots, y_n)$ тогда и только тогда, когда $x_1 = y_1$, $x_2 = y_2$, $\ldots$, $x_n = y_n$. Нулевой вектор, сложение векторов и умножение вектора на число определяются следующим образом:

$$0 = (0, 0, \ldots, 0),$$
$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),$$
$$\lambda(x_1, x_2, \ldots, x_n) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n).$$

Пусть $\alpha, \beta, \ldots, \gamma$ - некоторые числа. Вектор w называется *линейной комбинацией* векторов $x, y, \ldots, z$, если он представим в виде

$$w = \alpha x + \beta y + \ldots + \gamma z. \tag{1.1}$$

Числа $\alpha, \beta, \ldots, \gamma$ называются *коэффициентами линейной комбинации* (1.1). Линейная комбинация называется *тривиальной*, если все ее коэффициенты равны нулю, и *нетривиальной* в противном случае.

Векторы $x, y, \ldots, z$ называются *линейно зависимыми*, если существует нетривиальная линейная комбинация этих векторов, равная нулю, т. е., если существуют числа $\alpha, \beta, \ldots, \gamma$, среди которых есть отличные от нуля, такие, что

$$\alpha x + \beta y + \ldots + \gamma z = 0. \tag{1.2}$$

Если равенство (1.2) возможно только при $\alpha = \beta = \ldots = \gamma = 0$, то векторы $x, y, \ldots, z$ называются *линейно независимыми.*

Число

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i \tag{1.3}$$

называется *скалярным произведением* векторов $x = (x_1, x_2, \ldots, x_n)$ и $y = (y_1, y_2, \ldots, y_n)$. n-мерное вещественное пространство со скалярным произведением (1.3) называется *n-мерным евклидовым пространством.* Всюду ниже пространство R^n мы будем считать евклидовым.

Векторы $x, y \in R^n$ называются *ортогональными*, если $\langle x, y \rangle = 0$.

Длиной или *нормой* (*евклидовой нормой*) вектора $x \in R^n$ называется число

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Расстояние $\rho(x, y)$ между векторами x и y из R^n определяется как длина вектора $x - y$, т. е.

$$\rho(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

Непосредственно из определения скалярного произведения, нормы и расстояния вытекает, что

$$\langle x, y \rangle = \langle y, x \rangle,$$
$$\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle,$$

$$\langle \lambda x, y\rangle = \lambda\langle x, y\rangle,$$

$$\|x\| \geq 0 \ \ \forall x \in R^n \ \text{ и } \ \|x\| = 0 \Leftrightarrow x = 0,$$

$$\|\lambda x\| = |\lambda| \cdot \|x\|,$$

$$\rho(x, y) \geq 0 \ \ \forall x, y \in R^n \ \text{ и } \ \rho(x, y) = 0 \Leftrightarrow x = y,$$

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

Кроме того, для любых $x, y, z \in R^n$ справедливы неравенства

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|, \tag{1.4}$$

$$\|x + y\| \leq \|x\| + \|y\|, \tag{1.5}$$

$$\|x - z\| \leq \|x - y\| + \|y - z\|. \tag{1.6}$$

Неравенство (1.4) называется *неравенством Коши-Буняковского*, неравенства (1.5) и (1.6) – *неравенствами треугольника.*

Углом между векторами $x, y \in R^n$ называется угол φ, косинус которого определяется соотношением

$$\cos\varphi = \frac{\langle x, y\rangle}{\|x\| \cdot \|y\|}.$$

Из неравенства Коши-Буняковского следует, что $|\cos\varphi| \leq 1$.

Вектор-столбцом или просто *столбцом* размерности n называется матрица вида

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}. \tag{1.7}$$

Поскольку сложение матриц и умножение их на число определяются точно так же, как соответствующие операции над векторами, векторы $x \in R^n$ можно отождествлять со столбцами (1.7). Всюду ниже вектор $x \in R^n$ будет считаться вектор-столбцом, но для экономии места, если необходимо указать его компоненты, мы будем по-прежнему записывать его в строчку как $x = (x_1, x_2, \ldots, x_n)$. Запись же $[x_1\ x_2 \ldots x_n]$ означает, что это – матрица, состоящая из одной строки (*вектор-строка*, или просто *строка*), так что, если $(x_1, x_2, \ldots, x_n) = x$, то $[x_1\ x_2 \ldots x_n] = x^T$, где символ "^{T}"означает транспонирование матрицы (вектора).

Матрицу

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

мы будем кратко записывать как $[a_{ij}]_{i,j=1}^{m,n}$. Множество матриц размера $m \times n$ будем обозначать через $M_{m,n}$, а множество квадратных матриц порядка n – через M_n.

Пусть $A = [a_{ij}]_{i,j=1}^{n}$ – симметричная матрица порядка n, $x \in R^n$. Функция n переменных $x_1, \dots, x_n$

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = \langle Ax, x \rangle$$

называется *квадратичной формой* от переменных $x_1, \dots, x_n$, а числа a_{ij} – ее *коэффициентами*. Матрица A называется *матрицей квадратичной формы*. Матрица A (квадратичная форма $\langle Ax, x\rangle$) называется *положительно* (*неотрицательно*) *определенной*, если для всех $x \neq 0$ выполняется условие $\langle Ax, x\rangle > 0$ (≥ 0). Если $\langle Ax, x\rangle < 0$ (≤ 0) для всех $x \neq 0$, то матрица A (квадратичная форма $\langle Ax, x\rangle$) называется *отрицательно* (*неположительно*) *определенной*. Положительно и отрицательно определенные матрицы (квадратичные формы) называются *знакоопределенными*, неотрицательно и неположительно определенные – *полуопределенными*, а квадратичные формы, принимающие как положительные, так и отрицательные значения (и соответствующие им матрицы)– *знаконеопределенными*. Ясно, что если матрица A положительно определена (полуопределена), то матрица $[-A]$ определена отрицательно (неположительно).

Справедливы следующие утверждения:

i. Матрица A положительно определена тогда и только тогда, когда

$$\langle Ax, x\rangle \;\geq\; \mu \|x\|^2 \quad \forall x \in R^n ,$$

где $\mu > 0$ – наименьшее собственное число матрицы A.

ii. (*Критерий Сильвестра.*) Матрица A положительно определена тогда и только тогда, когда все ее главные угловые миноры положительны.

iii. Матрица A положительно полуопределена тогда и только тогда, когда все ее главные (не только угловые) миноры неотрицательны (напомним, что главными минорами матрицы A называются определители, соответствующие главным подматрицам матрицы A – таким ее подматрицам, которые образуются отбрасыванием строк и столбцов матрицы A с одинаковыми номерами).

В качестве следствия из критерия Сильвестра отметим тот факт, что положительно (а, значит, и отрицательно) определенная матрица не вырождена.

П р и м е р 1.1. Квадратичная форма

$$f(x) = 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3$$

положительно определена - главные угловые миноры матрицы

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

(матрицы квадратичной формы) положительны:

$$\det_1 = 2, \quad \det_2 = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3, \quad \det_3 = \det A = 4.$$

П р и м е р 1.2. Рассмотрим квадратичную форму

$$f(x) = 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3. \tag{1.8}$$

Главные угловые миноры первого, второго и третьего порядков матрицы

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

равны, соответственно, 2, 3 и 0. Критерий Сильвестра не выполняется. Вычислим все главные миноры матрицы A. Главные миноры порядка один

$$a_{11} = a_{22} = a_{33} = 2,$$

главные миноры порядка два

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3,$$

Главный минор порядка три, $\det A = 0$. Следовательно, квадратичная форма (1.8) – положительно полуопределенная.

Пусть $x, y \in R^n$. По правилу умножения матриц

$$x^T y = y^T x = \sum_{i=1}^{n} x_i y_i.$$

То есть в терминах матриц скалярное произведение $\langle x, y\rangle = x^T y$, $\|x\| = \sqrt{x^T x}$, квадратичная форма $\langle Ax, x\rangle = x^T Ax$, а неравенство Коши-Буняковского принимает вид:

$$|x^T y| \leq \|x\| \cdot \|y\|.$$

Произведение же вектор-столбца x на вектор-строку y^T, где $x, y \in R^n$, представляет собой уже матрицу

$$xy^T = [x_i y_j]_{i,j=1}^{n} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \dots & x_2 y_n \\ \dots & \dots & \dots & \dots \\ x_n y_1 & x_n y_2 & \dots & x_n y_n \end{bmatrix}.$$

Матрица xy^T называется *диадой* или *внешним произведением векторов* x и y и обозначается $x\rangle\langle y$ (в противоположность внутреннему произведению, как иногда называют скалярное произведение векторов). В отличие от внутреннего произведения, которое имеет смысл только для векторов одинаковой размерности, внешнее произведение определено для любых векторов x, y. А именно, если $x \in R^n$, $y \in R^m$, то xy^T – матрица размерности $n \times m$, а yx^T – матрица размерности $m \times n$ и $(yx^T)^T = xy^T$.

2. *Открытым шаром* радиуса r с центром в точке $a \in R^n$ называется множество

$$B(a, r) = \{x \in R^n : \|x - a\| < r\}.$$

Множество

$$B[a, r] = \{x \in R^n : \|x - a\| \leq r\}$$

называется *замкнутым шаром.* Открытый шар радиуса ε с центром в точке a называется также *ε-окрестностью* точки a и обозначается $O(a, \varepsilon)$.

Последовательность $\{x^k\} \in R^n$ называется *сходящейся* к точке $x \in \in R^n (x^k \to x)$, если для любого $\varepsilon > 0$ существует номер $N = N(\varepsilon)$ такой, что $\|x^k - x\| < \varepsilon$ для всех $k \geq N$, т. е., если $x^k \in O(x, \varepsilon)$ для всех $k \geq N$. Другими словами, последовательность $\{x^k\}$ сходится к точке x, если $\lim\limits_{k\to\infty} \|x^k - x\| = 0$. Точка x называется *пределом последовательности* $\{x^k\}$, что обозначается как $x = \lim\limits_{k\to\infty} x^k$. Ясно, что если последовательность $\{x^k\}$ сходится к точке x, то и всякая ее подпоследовательность сходится к той же самой точке.

Точка x называется *предельной точкой* последовательности $\{x^k\}$, если у этой последовательности имеется подпоследовательность, сходящаяся к x. Сходящаяся последовательность $x^k \to x$ имеет, очевидно, точку x своей предельной точкой, а других предельных точек уже не имеет. Несходящаяся последовательность может не иметь предельных точек, а может их иметь в любом количестве.

Последовательность $\{x^k\}$ называется *ограниченной*, если существует постоянная $M > 0$, такая, что $\|x^k\| \leq M$ для всех $k = 1, 2, \ldots$. Из любой ограниченной последовательности можно извлечь сходящуюся подпоследовательность (теорема Больцано – Вейерштрасса).

3. Множество $X \in R^n$ называется *ограниченным*, если существует постоянная $M > 0$ такая, что $\|x\| \leq M$ для всех $x \in X$. Нетрудно видеть, что множество X ограничено тогда и только тогда, когда для любых $x, y \in X$ расстояние $\rho(x, y) = \|x - y\|$ ограничено некоторой фиксированной постоянной. Величина $D = \sup\limits_{x,y\in X} \|x - y\|$ называется *диаметром* (ограниченного) множества X и обозначается $\operatorname{diam} X$.

Точка $x \in R^n$ называется *предельной точкой* множества $X \subseteq R^n$ если любая ее ε-окрестность содержит точки из X, отличные от x. Для того чтобы точка x была предельной для множества X, необходимо и достаточно чтобы в X существовала последовательность попарно различных точек, сходящаяся к x. Предельные точки множества X могут принадлежать, а могут не принадлежать множеству X. Множество X называется *замкнутым*, если оно содержит все свои предельные точки. Пустое множество, множество, состоящее из конечного числа точек и все пространство R^n – замкнутые множества. Множество X называется *компактным*, если любая последовательность $\{x^k\} \in X$ имеет хотя бы одну предельную точку x, причем $x \in X$.

Точка x называется *внутренней точкой* множества X, если она входит в X вместе с некоторой своей ε-окрестностью. Совокупность всех внутренних точек множества X называется его *внутренностью* и обозначается $\text{int}X$. Множество, все точки которого являются внутренними, называется *открытым.* Пустое множество и все пространство R^n – открытые множества.

Точка x называется *граничной точкой* множества X, если в любой ее ε-окрестности содержатся точки, как принадлежащие, так и не принадлежащие множеству X. Граничные точки сами могут принадлежать, а могут и не принадлежать множеству X. Совокупность всех граничных точек множества X называется его *границей* и обозначается через ∂X.

Всякая внутренняя точка множества, очевидно, является его предельной точкой. Однако не всякая граничная точка множества будет его предельной точкой – исключение здесь составляют изолированные точки множества. (Точка $x \in X$ называется *изолированной точкой* этого множества, если существует ε – окрестность этой точки, не содержащая ни одной точки множества X, отличной от x.)

Задачи и упражнения

1.1. Доказать, что

i. векторы $x, y, \ldots, z$ линейно зависимы тогда и только тогда, когда хотя бы один из них является линейной комбинацией остальных;

ii. если векторы $x, y, \ldots, z$ линейно независимы и вектор w есть их линейная комбинация,

$$w = \alpha x + \beta y + \ldots + \gamma z, \tag{1.9}$$

то представление (1.9) единственно (указание: предположить, что есть другое представление $w = \alpha_1 x + \beta_1 y + \ldots + \gamma_1 z$ и вычесть это равенство из равенства (1.9));

iii. если среди векторов $x, y, \ldots, z$ имеется нулевой вектор, то эти векторы линейно зависимы;

iv. система векторов, состоящая из одного ненулевого вектора, линейно независима.

1.2. Доказать, что любая система попарно ортогональных ненулевых векторов линейно независима.

1.3. Доказать неравенство Коши – Буняковского (указание: исследовать $\|x-\lambda y\|^2$ как квадратный относительно λ трехчлен). Показать, что при $x, y \neq 0$ равенство $|\langle x, y\rangle| = \|x\| \cdot \|y\|$ возможно в том и только в том случае, когда векторы x и y линейно зависимы: $x = \lambda y,\ \lambda = const.$

1.4. Доказать неравенства треугольника (1.5) и (1.6).

1.5. Доказать неравенство

$$\|x-y\| \geq |\, \|x\| - \|y\| \,|.$$

1.6. Доказать тождество параллелограмма:

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

1.7. Доказать, что векторы x и y ортогональны тогда и только тогда, когда $\|x+y\|^2 = \|x\|^2 + \|y\|^2$ (теорема Пифагора).

1.8. Доказать, что если векторы $x, y, z, \ldots, w$ попарно ортогональны, то

$$\|x+y+z+\ldots+w\|^2 = \|x\|^2 + \|y\|^2 + \|z\|^2 + \ldots + \|w\|^2.$$

Верно ли обратное утверждение?

1.9. Доказать, что при любых $x, y, z, \ldots, v, w \in R^n$

$$\|x-w\| \leq \|x-y\| + \|y-z\| + \ldots + \|v-w\|.$$

1.10. Доказать неравенство четырехугольника

$$\|x-y\| + \|z-v\| \geq |\, \|x-z\| - \|y-v\| \,|$$

(указание: воспользоваться результатом упр. 1.8).

1.11. Пусть $x, y \in R^n, x, y \neq 0$. Доказать, что

i. $x = \lambda y, \lambda > 0$, тогда и только тогда, когда угол между x и y равен нулю;

ii. $x = \lambda y, \lambda < 0$, тогда и только тогда, когда угол между x и y равен π.

1.12. В следующих примерах выяснить – является ли треугольник с вершинами x, y, z прямоугольным?, равнобедренным?, равносторонним?:

a) $x = (1, 2, -3, 1)$, $y = (0, 0, -4, 2)$, $z = (2, 0, -6, -5)$;

b) $x = (2, 1, 0, 1)$, $y = (1, 0, -1, 2)$, $z = (3, 0, 1, 0)$;

c) $x = (2, 4, 2, 4, 2)$, $y = (6, 4, 4, 4, 6)$, $z = (5, 7, 5, 7, 2)$;

d) $x = (3, 0, 4, 2)$, $y = (4, 2, 5, 2)$, $z = (2, 0, 5, 4)$.

1.13. Доказать, что если матрица A положительно определена, то положительно определены и матрицы $A^2 = AA$ и A^{-1}.

1.14. Доказать, что из положительной определенности матриц A и B следует положительная определенность матрицы $A + B$.

1.15. Пусть A и C – матрицы порядка n, матрица C – невырожденная. Доказать, что матрица $B = C^T AC$ положительно определена тогда и только тогда, когда положительно определена матрица A.

1.16. Пусть матрица $A \in M_{m,n}$. Тогда матрицы AA^T и $A^T A$ – неотрицательно определенные. Если $\mathrm{rank} A = n$, то матрица $A^T A$ положительно определена. Если $\mathrm{rank} A = m$, то положительно определена матрица AA^T. (Доказать.)

1.17. Вывести критерии отрицательной определенности и полуопределенности матриц. (Замечание: если A – матрица порядка k, то $\det\,[-A] = (-1)^k \det A$.)

1.18. Исследовать на знакоопределенность следующие квадратичные формы:

a) $x_1^2 + 2x_2^2 + 3x_3^2 + 2x_1x_2 + 2x_1x_3 + 4x_2x_3$;

b) $x_1^2 + x_2^2 + x_3^2 + 2x_1x_3$;

c) $-3x_1^2 - x_2^2 - 5x_3^2 + 2x_1x_2 + 4x_1x_3 - 4x_2x_3$;

d) $2x_1^2 + x_2^2 + 3x_3^2 + 6x_1x_2 + 2x_1x_3 + 6x_2x_3$;

e) $-x_1^2 - x_2^2 - x_3^2 + 2x_1x_2 + x_1x_3 + 2x_2x_3$.

1.19. Найти все значения параметра λ, при которых положительно определены следующие квадратичные формы:

a) $5x_1^2 + x_2^2 + \lambda x_3^2 + 4x_1x_2 - 2x_1x_3 - 2x_2x_3$;

b) $2x_1^2 + x_2^2 + 3x_3^2 + 2\lambda x_1x_2 + 2x_1x_3$;

c) $x_1^2 + x_2^2 + 5x_3^2 + 2\lambda x_1x_2 - 2x_1x_3 + 4x_2x_3$.

1.20. Пусть $G \in M_{m,n}$. Показать, что $\langle Gx, y\rangle = \langle G^T y, x\rangle$.

1.21. Доказать, что $rank\, xy^T = 1$ для любых $x, y \in R^n$, $x, y \neq 0$.

1.22. Пусть A – матрица размерности $m \times n$, B – матрица размерности $n \times r$. Обозначим через $a^1, \ldots, a^n$ столбцы матрицы A, через $b^1, \ldots, b^n$ столбцы матрицы B^T (т. е. $(b^1)^T, \ldots, (b^n)^T$ – строки матрицы B). Доказать, что $AB = \sum\limits_{k=1}^{n} a^k (b^k)^T$.

1.23. Доказать, что если последовательность $\{x^k\}$ имеет предел, то этот предел – единственный.

1.24. Доказать, что если $x = \lim\limits_{k\to\infty} x^k$, $y = \lim\limits_{k\to\infty} y^k$, то

$$\lim_{k\to\infty} \|x^k - y^k\| = \|x - y\|$$

(лемма о непрерывности расстояния).

1.25. Доказать, что последовательность $\{x^k\} \in R^n$ сходится к точке $x \in R^n$ тогда и только тогда, когда $x_i^k \to x_i$ для всех $i = 1, \ldots, n$.

1.26. Доказать, что сходящаяся последовательность ограничена.

1.27. Доказать, что открытый шар – открытое множество, замкнутый шар – замкнутое множество.

1.28. Доказать, что если множество X не имеет предельных точек, то оно замкнуто.

1.29. Доказать, что множество $X \in R^n$ компактно тогда и только тогда, когда оно замкнуто и ограничено (указание: воспользоваться теоремой Больцано – Вейерштрасса).

§ 2. Выпуклые множества

О п р е д е л е н и е 2.1. Пусть x, y – две различные точки в R^n. Совокупность точек вида

$$\lambda x + (1-\lambda)y = y + \lambda(x-y), \quad -\infty < \lambda < +\infty \tag{2.1}$$

называется *прямой*, проходящей через точки x и y.

О п р е д е л е н и е 2.2. *Прямой*, проходящей через точку $a \in R^n$ с *направляющим вектором* $p \in R^n$, называется совокупность точек вида

$$a + \lambda p, \quad -\infty < \lambda < +\infty \; . \tag{2.2}$$

Очевидно, что прямые (2.1) и (2.2) совпадают при $a = y$ и $p = x - y$ (рис. 2.1).

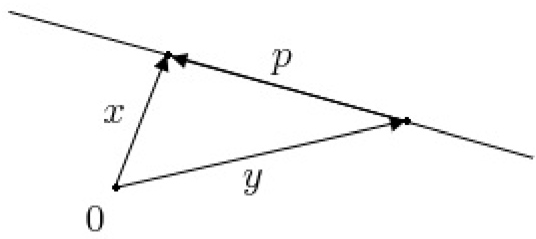

Рис. 2.1

О п р е д е л е н и е 2.3. *Лучом*, исходящим из точки $a \in R^n$ в направлении $p \in R^n$, называется множество точек вида

$$a + \lambda p, \quad \lambda \geq 0.$$

О п р е д е л е н и е 2.4. *Отрезком*, соединяющим точки $x, y \in R^n$, называется множество

$$[x, y] = \{z \in R^n : z = \lambda x + (1-\lambda)y = y + \lambda(x-y), \quad 0 \leq \lambda \leq 1\}.$$

О п р е д е л е н и е 2.5. Множество $X \subseteq R^n$ называется *выпуклым*, если вместе с любыми своими точками x и y оно содержит и весь отрезок их соединяющий, т. е., если точка $z = \lambda x + (1-\lambda)y$ принадлежит X при любых $x, y \in X$ и $\lambda \in [0, 1]$.

На рис. 2.2 и 2.3 изображены выпуклые множества, на рис. 2.4 – невыпуклое множество.

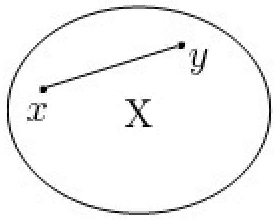

Рис. 2.2

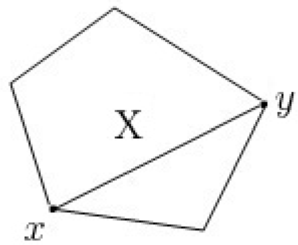

Рис. 2.3

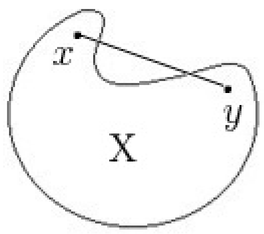

Рис. 2.4

Простейшими примерами выпуклых множеств являются прямая, луч, все пространство R^n. Пустое множество и множество, состоящее из одной точки, также считаются выпуклыми.

Рассмотрим другие примеры выпуклых множеств.

П р и м е р 2.1. *n-мерный прямоугольный параллелепипед*

$$\Pi = \{x \in R^n : \alpha_i \leq x_i \leq \beta_i,\ i = 1, \ldots, n\}$$

является выпуклым множеством.

В самом деле, для любых $x, y \in \Pi$ и $\lambda \in [0, 1]$

$$\alpha_i = \lambda\alpha_i + (1-\lambda)\alpha_i \leq \lambda x_i + (1-\lambda)y_i \leq$$

$$\leq \lambda\beta_i + (1-\lambda)\beta_i = \beta_i, \;\; i = 1, \ldots, n.$$

П р и м е р 2.2. Шар $B[a, r]$ радиуса r с центром в точке a – выпуклое множество.

Действительно, если $z = \lambda x + (1-\lambda)y$, где точки $x, y \in B[a, r]$ и $\lambda \in [0, 1]$, то

$$\|z - a\| = \|\lambda(x-a) + (1-\lambda)(y-a)\| \leq$$

$$\leq \lambda\|x-a\| + (1-\lambda)\|y-a\| \leq \lambda r + (1-\lambda)r = r.$$

О п р е д е л е н и е 2.6. Пусть X и x^0 – произвольные множество и вектор из R^n, соответственно. *Сдвигом* множества X на вектор x^0 называется множество

$$X + x^0 = \{z = x + x^0 :\ x \in X\}.$$

Нетрудно видеть, что сдвиг выпуклого множества на произвольный вектор x^0 является выпуклым множеством.

О п р е д е л е н и е 2.7. Пусть a – ненулевой вектор из R^n, β – число. Множество

$$H = H(a, \beta) = \{x \in R^n : \langle a, x \rangle = \beta\}$$

называется *гиперплоскостью*. Вектор a называется *нормальным вектором* (*вектором нормали*) гиперплоскости H.

Очевидно, что вектор a и число β определяют гиперплоскость с точностью до общего ненулевого множителя. При $n = 1$ гиперплоскость – это точка, при $n = 2$ – прямая, при $n = 3$ – плоскость.

Возьмем произвольный вектор $x^0 \in H(a, \beta)$. Любой вектор $x \in H(a, \beta)$ можно однозначно представить в виде $x = y + x^0$, где y принадлежит гиперплоскости $\langle a, x \rangle = 0$, проходящей через начало координат:

$$\langle a, x \rangle = \langle a, y \rangle + \langle a, x^0 \rangle = \beta \, .$$

Следовательно, гиперплоскость $H(a, \beta) = H(a, 0) + x^0$, т. е. представляет собой сдвиг на вектор x^0 гиперплоскости $H(a, 0)$ (множества векторов, ортогональных вектору a) (рис. 2.5). В этом смысле можно говорить, что гиперплоскость ортогональна своему нормальному вектору.

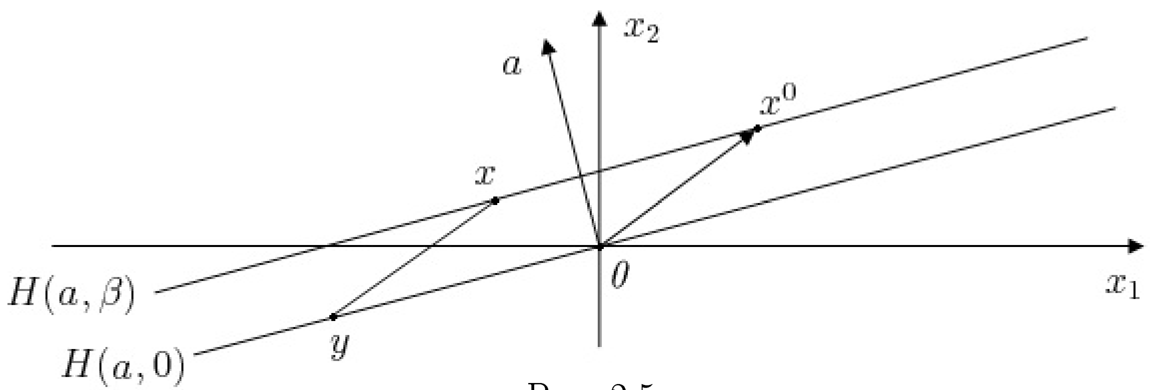

Рис. 2.5

Пусть точки $x, y \in H(a, \beta)$. Тогда и вся прямая, проходящая через эти точки, принадлежит гиперплоскости H:

$$\langle a, y + \lambda(x - y) \rangle = \langle a, y \rangle + \lambda[\langle a, x \rangle - \langle a, y \rangle] = \beta + \lambda[\beta - \beta] = \beta.$$

Отсюда также следует, что вектор нормали гиперплоскости ортогонален направляющему вектору $p = x - y$ любой прямой, проходящей через любые точки $x, y \in H$ (рис. 2.6).

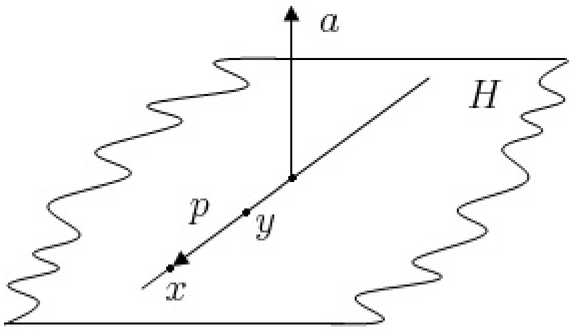

Рис. 2.6

О п р е д е л е н и е 2.8. Пересечение конечного числа гиперплоскостей называется *линейным многообразием* (если это пересечение не пусто).

Таким образом, линейное многообразие – это множество решений системы линейных алгебраических уравнений

$$\langle a^j, x\rangle = a_{j1}x_1 + \ldots + a_{jn}x_n = b_j \ , \ \ j = 1, \ldots, m,$$

или, в матричной форме,

$$Ax = b \ ,$$

где $A = [a_{ji}]_{j,i=1}^{m,n}$, $b = [b_j]_{j=1}^{m}$.

Пусть $H(a, \beta)$ – некоторая гиперплоскость. В пространстве R^n гиперплоскость $H(a, \beta)$ порождает два *полупространства* (рис. 2.7)

$$H_+ = \{x \in R^n : \ \langle a, x\rangle \geq \beta\} \ ,$$
$$H_- = \{x \in R^n : \ \langle a, x\rangle \leq \beta\} \ .$$

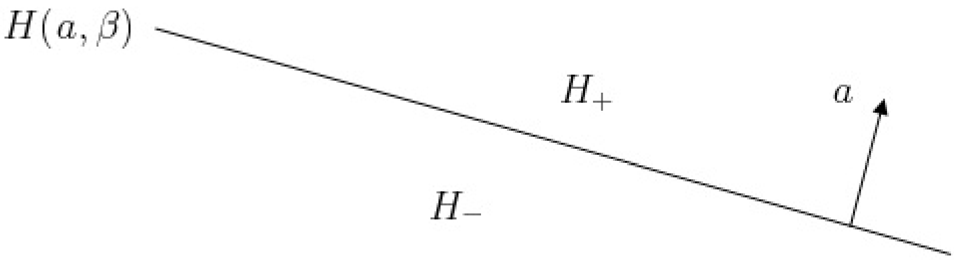

Рис. 2.7

О п р е д е л е н и е 2.9. Пересечение конечного числа гиперплоскостей и полупространств (если оно не пусто) называется *выпуклым многогранным множеством* или *выпуклым полиэдром*. Ограниченное выпуклое многогранное множество называется *выпуклым многогранником* или *политопом*.

То есть выпуклое многогранное множество – это множество решений системы линейных равенств и неравенств

$$\langle a^j, x\rangle = a_{j1}x_1 + \ldots + a_{jn}x_n = b_j\ , j = 1, \ldots, r, \tag{2.3}$$

$$\langle a^j, x\rangle = a_{j1}x_1 + \ldots + a_{jn}x_n \le b_j\ , j = r+1, \ldots, m \tag{2.4}$$

(случаи $r = 0$ и $r = m$ здесь не исключаются; в последнем случае выпуклое многогранное множество вырождается в линейное многообразие).

О п р е д е л е н и е 2.10. Система линейных условий вида (2.3), (2.4) называется *линейно независимой*, если линейно независима соответствующая система векторов $\{a^j\}$. *Рангом системы линейных условий* называется ранг матрицы, составленной из векторов a^j, отвечающих данной системе.

О п р е д е л е н и е 2.11. Пусть X – выпуклое многогранное множество, определяемое системой (2.3), (2.4). Точка $x \in X$ называется *вершиной* или *угловой точкой* множества X, если среди условий (2.3), (2.4) найдутся n линейно независимых условий, которым эта точка удовлетворяет как точным равенствам.

Таким образом, угловая точка является точкой пересечения не менее чем n из m гиперплоскостей $\langle a^j, x\rangle = b_j$. Если число гиперплоскостей, пересекающихся в угловой точке, в точности равно n, то угловая точка называется *невырожденной*. Если же в угловой точке пересекается более n гиперплоскостей, т. е. число соотношений (2.3), (2.4), которые эта точка обращает в равенства, больше n, то угловая точка – *вырожденная*. На рис. 2.8 многогранник X имеет пять угловых точек – a, b, c, d, e, одна из которых, точка b, – вырожденная.

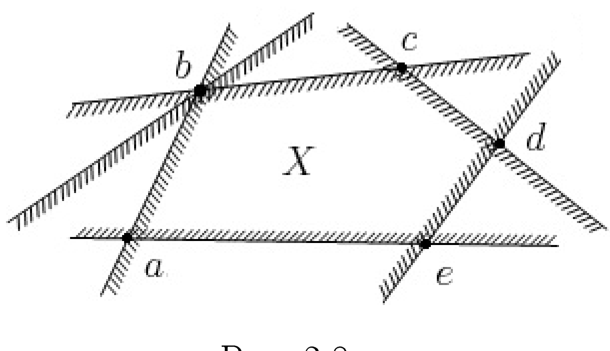

Рис. 2.8

Понятно, что если ранг системы условий (2.3), (2.4) меньше n, то выпуклое многогранное множество не имеет угловых точек. Если же выпуклое многогранное множество имеет угловые точки, то их число конечно – из системы (2.3), (2.4) можно составить лишь конечное число линейно независимых подсистем по n уравнений.

О п р е д е л е н и е 2.12. Пусть X и Y – непустые множества из R^n. Говорят, что гиперплоскость $H(a, \beta)$ разделяет множества X и Y (или отделяет X от Y), если для любых $x \in X$ и $y \in Y$ выполняется неравенство

$$\langle a, x \rangle \ \leq \ \beta \ \leq \ \langle a, y \rangle \, . \tag{2.5}$$

Если для всех $x \in X$, $y \in Y$ и некоторого числа $\delta > 0$ выполняется неравенство

$$\langle a, x \rangle \ + \ \delta \ \leq \ \beta \ \leq \ \langle a, y \rangle - \delta, \tag{2.6}$$

то говорят, что гиперплоскость $H(a, \beta)$ сильно отделяет множество X от множества Y.

Геометрически неравенства (2.5) и (2.6) означают, что множества X и Y лежат в разных полупространствах, определяемых гиперплоскостью H (рис. 2.9 – 2.11). На рис. 2.9 гиперплоскость H сильно разделяет множества X и Y. Рис. 2.12 и 2.13 иллюстрируют случаи, когда множества X и Y не могут быть разделены никакой гиперплоскостью.

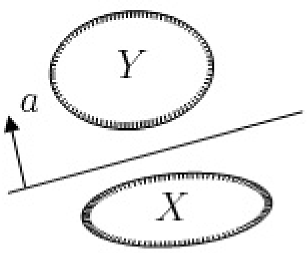

Рис. 2.9

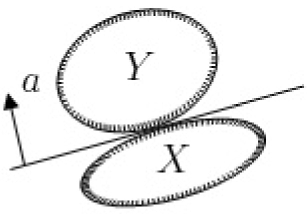

Рис. 2.10

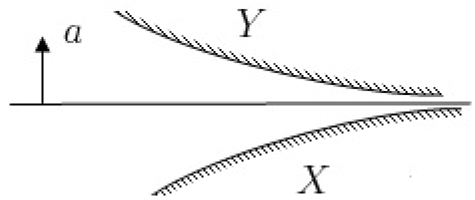

Рис. 2.11

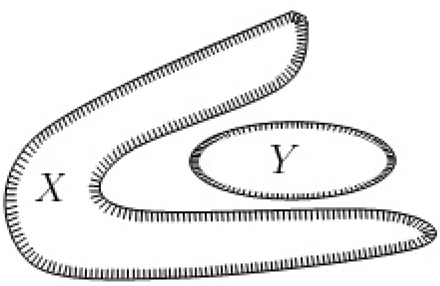

Рис. 2.12

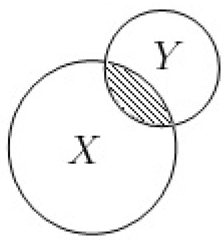

Рис. 2.13

Т е о р е м а 2.1. Пусть X и Y – непустые множества из R^n. Эти множества можно разделить некоторой гиперплоскостью тогда и только тогда, когда существует вектор $a \neq 0$ такой, что

$$\langle a, x\rangle \ \leq \ \langle a, y\rangle \quad \forall x \in X, y \in Y.$$

Их можно сильно отделить друг от друга некоторой гиперплоскостью тогда и только тогда, когда существуют вектор $a \neq 0$ и число $\varepsilon > 0$, такие, что

$$\langle a, x\rangle \ \leq \ \langle a, y\rangle - \varepsilon \quad \forall x \in X, y \in Y.$$

Т е о р е м а 2.2. Пусть X и Y – непустые выпуклые множества из R^n, причем хотя бы одно из них ограничено. Если замыкания множеств X и Y не пересекаются, то существует гиперплоскость, сильно отделяющая X от Y.

Ограниченность хотя бы одного из множеств в теореме 2.2 существенна: рис. 2.11 иллюстрирует случай, когда множества X и Y выпуклы, замкнуты и не имеют общих точек, но не могут быть сильно отделены друг от друга.

Т е о р е м а 2.3. Пусть X и Y – непустые выпуклые множества из R^n, не имеющие общих точек. Тогда существует гиперплоскость, разделяющая эти множества.

Т е о р е м а 2.4. Пусть X и Y – выпуклые множества из R^n и внутренность одного из них, скажем Y, не пуста. Если $X \cap \operatorname{int} Y = \emptyset$, то существует гиперплоскость, разделяющая множества X и Y.

Т е о р е м а 2.5. Пусть X и Y – выпуклые множества из R^n, причем $\operatorname{int} X \neq \emptyset$, $\operatorname{int} Y \neq \emptyset$ и $\operatorname{int} X \cap \operatorname{int} Y = \emptyset$. Тогда существует гиперплоскость, разделяющая множества X и Y.

О п р е д е л е н и е 2.13. Пусть X – непустое множество из R^n, точка $x^0 \in \partial X$. Гиперплоскость $H = \{x \in R^n : \langle a, x - x^0\rangle = 0\}$ называется опорной ко множеству X в точке x^0, если для любых $x \in X$ выполняется неравенство $\langle a, x\rangle \ \leq \ \langle a, x^0\rangle$.

Т е о р е м а 2.6. Во всякой граничной точке непустого выпуклого множества X существует опорная к нему гиперплоскость.

Графическая иллюстрация определения 2.13 и теоремы 2.6 приведена на рис.2.14 и 2.15.

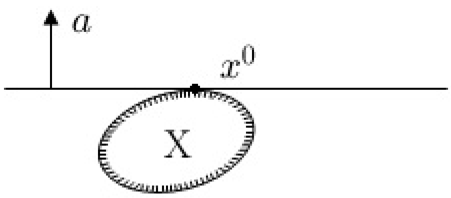

Рис. 2.14

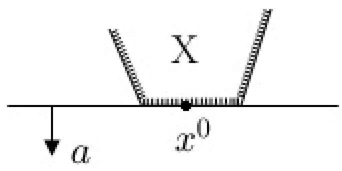

Рис. 2.15

Задачи и упражнения

2.1. Проверить, принадлежит ли точка $\overline{x}$ прямой, проходящей через точки x и y? Если принадлежит, то найти расстояние от точки $\overline{x}$ до ближайшей точки отрезка $[x, y]$:

a) $x = (3, -4, 2, -1),\ y = (1, -2, 4, 0),\ \overline{x} = (-3, 2, 8, 2);$

b) $x = (0, -4, 2, 3),\ y = (-4, -7, 5, 4),\ \overline{x} = (8, 2, -4, 1);$

c) $x = (-4, 0, 4, 1, -1),\ y = (-3, 1, 3, 0, 0),\ \overline{x} = (1, 5, -1, -4, 4);$

d) $x = (2, 1, -1, 0, 1),\ y = (-1, 4, -3, 10, 1),\ \overline{x} = (-4, 7, -5, 3, 1);$

e) $x = (-3, 10, 4, 0, -8),\ y = (7, 0, -6, 10, 2),\ \overline{x} = (5, 2, -4, 8, 0).$

2.2. Найти точку пересечения прямых, проходящих через точки x^1, y^1 и x^2, y^2 :

a) $x^1 = (-3, 0, 7),\ y^1 = (-2, 3, 11);$

$x^2 = (-3, 4, 14),\ y^2 = (1, 8, 16);$

b) $x^1 = (7, 12, 4, 5),\ y^1 = (2, 4, 2, 1);$

$x^2 = (5, 4, 0, 4),\ y^2 = (8, 4, -2, 7);$

c) $x^1 = (-6, -2, 2, 6, 0),\ y^1 = (-4, 4, 6, -2, 6);$

$x^2 = (8, 6, -5, 5, 6),\ y^2 = (9, 9, -3, 1, 9);$

d) $x^1 = (3, 9, -5, 3, 1),\ y^1 = (5, 6, -5, 3, 0);$

$x^2 = (4, 0, 1, 12, 2),\ y^2 = (5, 1, -1, 9, 1).$

2.3. Доказать, что следующее определение эквивалентно определению 2.5: «Множество X называется выпуклым, если $z = \frac{1}{2}x + \frac{1}{2}y \in X$ для любых $x, y \in X$».

2.4. В следующих задачах найти все значения параметра λ, при которых множество X является выпуклым:

a) $X = \{x \in R^2 : \lambda(x_1 - x_2^2) = 0,\ x_1 + x_2 = 1\}$;

b) $X = \{x \in R^2 : \lambda(x_1 - x_2^2) = 0,\ x_1 + x_2 = \lambda\}$;

c) $X = \{x \in R^2 : x_1^2(\lambda^2 + 3\lambda + 2) - x_2 \geq 0\}$;

d) $X = \{x \in R^2 : e^{x_1}(\lambda^2 - 5\lambda + 6) - x_2(\lambda^2 + 2) \leq 0\}$.

2.5. Доказать, что внутренность выпуклого множества выпукла.

2.6. Доказать, что замыкание выпуклого множества выпукло.

2.7. Пусть X – некоторое множество из R^n, $\overline{X}$ – замыкание множества X. Можно ли утверждать, что если $\overline{X}$ выпукло, то X также выпукло?

2.8. Пусть X – выпуклое множество с непустой внутренностью. Доказать, что $\overline{X} = \overline{(\mathrm{int}X)}$. Привести пример невыпуклого множества, для которого это равенство не имеет места.

2.9. Доказать, что пересечение любого числа выпуклых множеств – выпуклое множество. Верно ли это утверждение для объединения выпуклых множеств?

2.10. Доказать, что декартово произведение выпуклых множеств X и Y

$$X \times Y = \{(x, y) : x \in X, y \in Y\}$$

является выпуклым множеством.

2.11. *Линейной комбинацией множеств* $X_1, X_2, \ldots, X_k$ (с коэффициентами $\lambda_1, \lambda_2, \ldots, \lambda_k$) называется множество $Z = \lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_k X_k$, состоящее из тех и только тех точек z, которые представимы в виде $z = \lambda_1 x^1 + \lambda_2 x^2 + \ldots + \lambda_k x^k$, $x^i \in X_i$, $i = 1, 2, \ldots, k$. При $\lambda_1 = \lambda_2 = \ldots = \lambda_k = 1$ множество Z – *сумма множеств* $X_1, X_2, \ldots, X_k$, при $k = 1$ – *произведение множества на число*, при $k = 2$ и $\lambda_1 = -\lambda_2 = 1$ – *разность множеств* X_1 и X_2. Доказать, что множество $Z =$ выпукло, если выпуклы множества $X_1, X_2, \ldots, X_k$.

2.12. Пусть $X_1 = \{x \in R^2 : x_1 = 0, 0 \leq x_2 \leq 1\}$,

$$X_2 = \{x \in R^2 : 0 \leq x_1 \leq 1, x_2 = 2\}.$$

Опишите и изобразите на плоскости множества $X_1 + X_2$ и $X_1 - X_2$.

2.13. Пусть точки $x^1, \ldots, x^k \in R^n$. Точка $x = \sum\limits_{i=1}^{k} \lambda_i x^i$, где $\sum\limits_{i=1}^{k} \lambda_i = 1$, $\lambda_i \geq 0$ для всех $i = 1, \ldots, k$, называется *выпуклой комбинацией* точек $x^1, \ldots, x^k$. Множество всех выпуклых комбинаций точек $x^1, \ldots, x^k$ называется их *выпуклой оболочкой* и обозначается $\mathrm{conv}\{x^1, \ldots, x^k\}$. То есть

$$\mathrm{conv}\{x^1, \ldots, x^k\} =$$

$$= \{x \in R^n : \ x = \sum_{i=1}^{k} \lambda_i x^i, \ \sum_{i=1}^{k} \lambda_i = 1, \ \lambda_i \geq 0, \ i = 1, \ldots, k\}.$$

Выпуклая оболочка двух точек – это отрезок, соединяющий эти точки. Доказать, что множество $X \subset R^n$ выпукло тогда и только тогда, когда оно содержит все выпуклые комбинации любого конечного числа своих точек. (Указание: использовать метод математической индукции.)

2.14. Изобразите на плоскости выпуклую оболочку точек $a = (0,0)$, $b = (-1,-1)$, $c = (-2,1)$, $d = (-2,3)$, $e = (1,2)$, $f = (3,-3)$, $g = (-4,-2)$.

2.15. Пусть X – выпуклое множество. Точка $x \in X$ называется *крайней* или *экстремальной* точкой множества X, если представление $x = \lambda x^1 + (1 - \lambda) x^2$ невозможно ни для каких $x^1, x^2 \in X, x^1 \neq x^2$, и $\lambda \in (0,1)$. То есть крайняя точка не может быть внутренней точкой никакого отрезка, имеющего своими концами точки из X.

Показать, что крайняя точка выпуклого множества является его граничной точкой. Всякая ли граничная точка множества является его крайней точкой?

2.16. Доказать, что угловые точки выпуклого многогранного множества являются его крайними точками, и, наоборот, крайние точки выпуклого многогранного множества являются его угловыми точками.

2.17. Найти все угловые точки *фундаментального симплекса*

$$S = \{x \in R^n : \ x_1 + x_2 + \ldots + x_n = 1 \ , \ x_i \geq 0 \ , \ i = 1, \ldots, n\}.$$

2.18. Найти все угловые точки множества

$$X = \{x \in R^n : \ x_1 + x_2 + \ldots + x_n \leq 1 \ , \ x_i \geq 0 \ , \ i = 1, \ldots, n\}.$$

2.19. Для множеств, заданных следующими системами линейных равенств и неравенств, найти все угловые точки. Указать вырожденные и невырожденные угловые точки:

a) $\begin{aligned} 2x_1 + \ \ x_2 &\le 10, \\ x_1 + 2x_2 &\le 14, \\ 4x_1 + \ \ x_2 &\le 16; \end{aligned}$

b) $\begin{aligned} x_1 + \ \ x_2 &\le 1, \\ 3x_1 + 5x_2 &\le 4, \\ x_1 + 3x_2 &\le 7, \\ x_1 - \ \ x_2 &\le 0; \end{aligned}$

c) $\begin{aligned} x_1 + 2x_2 + \ \ x_3 &= 5, \\ -x_1 \qquad\quad - 2x_3 &\le 3, \\ x_i \ge 0, \ i &= 1, 2, 3; \end{aligned}$

d) $\begin{aligned} 4x_1 + 3x_2 + \ \ x_3 &= 4, \\ -x_1 + 6x_2 - \ \ x_3 &\ge 2, \\ 7x_1 - \ \ x_2 + 2x_3 &\le 5, \\ x_i \ge 0, \ i &= 1, 2, 3; \end{aligned}$

e) $\begin{aligned} x_1 - x_2 + x_3 &\le 1, \\ x_1 + x_2 + x_3 &\le 4, \\ x_i \ge 0, \ i &= 1, 2, 3; \end{aligned}$

f) $\begin{aligned} x_1 - x_2 + x_3 &\ge 1, \\ x_1 + x_2 + x_3 &\ge 4, \\ x_i \ge 0, \ i &= 1, 2, 3; \end{aligned}$

g) $\begin{aligned} x_1 + x_2 + 3x_3 + 4x_4 &= 13, \\ x_1 - x_2 + \ \ x_3 - \ \ x_4 &= 2, \\ x_1 \ge 0, \ x_2 \ge 0, \ x_3 &\ge 0; \end{aligned}$

h) $\begin{aligned} 2x_1 - \ \ x_2 + 3_3 + 5x_4 &= 3, \\ 3x_1 - 2x_2 + x_3 - \ \ x_4 &\le 4, \\ x_1 \ge 0, \ x_2 \ge 0, \ x_4 &\ge 0. \end{aligned}$

2.20. Вывести уравнение гиперплоскости, разделяющей выпуклые множества X_1 и X_2:

a) $X_1 = \{x \in R^2 : \dfrac{x_1^2}{4} + \dfrac{x_2^2}{9} \le 1\}, \ \ X_2 = \{x \in R^2 : x_2 \ge \dfrac{3}{x_1}\ , x_1 > 0\}$;

b) $X_1 = \{x \in R^2 : x_2(x_1 - 1) \ge 3\ , x_1 < 1\}$;

$X_2 = \{x \in R^2 : (x_2 + 4)(x_1 + 2) \ge 3\ , x_1 > -2\}$;

c) $X_1 = \{x \in R^n : x_1^2 + \ldots + x_n^2 \le 1\}$;

$X_2 = \{x \in R^n : x_1^2 + \ldots + x_{n-1}^2 + 1 \le x_n\}$.

2.21. Вывести уравнение гиперплоскости, опорной ко множеству X в точке x^0 :

a) $X = \{x \in R^2 : x_2 \ge e^{x_1}\}, \ \ x^0 = (0, 1)$;

b) $X = \{x \in R^2 : x_1 x_2 \ge 1\}, \ \ x^0 = (1, 1)$.

§ 3. Дифференцирование функций многих переменных

Пусть функция $f(x) : R^n \to R^1$.

О п р е д е л е н и е 3.1. Функция $f(x)$, определенная на множестве $X \subseteq R^n$, называется *непрерывной в точке* $a \in X$, если для любого числа $\varepsilon > 0$ существует число $\delta > 0$ такое, что для любого $x \in X$, удовлетворяющего условию $\|x - a\| < \delta$, будет $|f(x) - f(a)| < \varepsilon$.

Приведем также другое, эквивалентное, определение непрерывности.

О п р е д е л е н и е 3.2. Функция $f(x)$ непрерывна при $x = a$, если для любой последовательности $\{x^k\}$ из множества X, сходящейся к точке a, будет $f(x) \to f(a)$.

Функция, непрерывная в каждой точке множества X, называется *непрерывной на множестве* X. Класс таких функций обозначается через $C(X)$.

О п р е д е л е н и е 3.3. Функция $f(x)$, определенная на множестве $X \subseteq R^n$, называется *равномерно непрерывной* на этом множестве, если для любого числа $\varepsilon > 0$ можно указать такое число $\delta > 0$, зависящее только от ε, что для любых $x, y \in X$, удовлетворяющих условию $\|x - y\| < \delta$, выполняется неравенство $|f(x) - f(y)| < \varepsilon$.

Т е о р е м а 3.1 (о равномерной непрерывности). Функция, непрерывная на замкнутом ограниченном множестве X, равномерно непрерывна на этом множестве.

О п р е д е л е н и е 3.4. Пусть функция $f(x)$ определена в некоторой ε-окрестности $O(x, \varepsilon)$ точки x. Функция $f(x)$ называется *дифференцируемой в точке* x, если существует вектор $\nabla f(x) \in R^n$ (читается *«набла* $f(x)$*»*) такой, что приращение функции $\triangle f(x) = f(y) - f(x)$, соответствующее приращению аргумента $\triangle x = y - x$, $y \in O(x, \varepsilon)$ можно представить в виде

$$f(y) - f(x) = \langle \nabla f(x), y - x \rangle + o(\|y - x\|), \tag{3.1}$$

где $o(\|y - x\|)$ – бесконечно малая более высокого порядка, чем $\|y - x\|$:

$o(\|y-x\|)/\|y-x\| \to 0$ при $\|y-x\| \to 0$. Величина $\mathrm{d}f(x) = \langle \nabla f(x), y - x \rangle$ представляет главную линейную относительно $\triangle x$ часть приращения (3.1) и называется *дифференциалом* функции $f(x)$ в точке x, а вектор $\nabla f(x)$ – *градиентом* функции $f(x)$ в точке x.

О п р е д е л е н и е 3.5. Функция $f(x)$ называется *дифференцируемой на множестве* $X \subseteq R^n$, если она дифференцируема в каждой точке этого множества, и *непрерывно дифференцируемой* или *гладкой* на множестве X, ($f(x) \in C_1(X)$) если $\|\nabla f(y) - \nabla f(x)\| \to 0$ при $\|y - x\| \to 0$, $y, x \in X$.

З а м е ч а н и е. По определению, для дифференцируемости функции в точке требуется, чтобы функция была определена в некоторой окрестности этой точки. Следовательно, для дифференцируемости функции на множестве X необходимо, чтобы это множество было открыто, либо чтобы функция была определена на некотором открытом множестве, содержащем множество X.

Условие (3.1) однозначно определяет градиент функции $f(x)$ в точке x как вектор, составленный из ее частных производных в этой точке: $\nabla f(x) = (\nabla f(x)_1, \dots, \nabla f(x)_n) = (\partial f(x)/\partial x_1, \dots, \partial f(x)/\partial x_n)$. В самом деле, положим в (3.1) $y = x + \lambda e^i$, где $e^i = (0, \dots, 0, 1, 0, \dots, 0)$ – i-й орт, $|\lambda| < \varepsilon$. Тогда

$$f(x + \lambda e^i) - f(x) = \nabla f(x)_i \cdot \lambda + o(\lambda).$$

Деля обе части этого равенства на $\lambda \neq 0$ и устремляя затем $\lambda \to 0$, получим, что

$$\nabla f(x)_i \ = \lim_{\lambda \to 0} \frac{f(x + \lambda e^i) - f(x)}{\lambda} \ \stackrel{\text{def}}{=} \ \frac{\partial f(x)}{\partial x_i} \ .$$

Таким образом, функция, дифференцируемая в точке, имеет в этой точке частные производные по всем аргументам. Обратное, вообще говоря, не верно – функция может иметь в точке частные производные по всем аргументам, но не быть дифференцируемой в этой точке. Однако справедлива следующая

Т е о р е м а 3.2. Если функция $f(x)$ имеет частные производные по всем аргументам в некоторой окрестности точки x, причем все эти частные производные непрерывны в самой точке x, то функция $f(x)$ дифференцируема в точке x.

Пусть функция $f(x)$ дифференцируема в точке x. Положим $y = x + \lambda p$, где p – произвольный вектор (направление) из R^n, $\|p\| = 1$, $0 < \lambda < \varepsilon$. Тогда из (3.1) получим, что

$$f(y) - f(x) = \lambda\langle\nabla f(x), p\rangle + o(\lambda) = \lambda\left[\langle\nabla f(x), p\rangle + \frac{o(\lambda)}{\lambda}\right] .$$

Отсюда следует, что величина и знак приращения $f(y) - f(x)$ определяются (по крайней мере, в малой окрестности точки x) величиной и знаком скалярного произведения $\langle\nabla f(x), p\rangle$. Это скалярное произведение, в силу неравенства Коши – Буняковского, максимально при $p = \nabla f(x)/\|\nabla f(x)\|$. Следовательно, градиент функции $f(x)$ в точке x определяет направление наискорейшего (локального) возрастания функции $f(x)$, или, как принято говорить, является *направлением наискорейшего подъема* функции $f(x)$ в точке x. Понятно, что *антиградиент* $[-\nabla f(x)]$ является *направлением наискорейшего спуска.*

О п р е д е л е н и е 3.6. Пусть функция $f(x)$ дифференцируема в некоторой ε-окрестности $O(x, \varepsilon)$ точки x. Функция $f(x)$ называется *дважды дифференцируемой в точке* x, если существует симметричная матрица $\nabla^2 f(x)$ (читается «*набла два* $f(x)$») порядка n, такая, что при любом $y \in O(x, \varepsilon)$ справедливо представление

$$\nabla f(y) - \nabla f(x) = \nabla^2 f(x)(y - x) + o(\|y - x\|) , \tag{3.2}$$

где $o(\|y-x\|) = (o_1(\|y-x\|), \ldots, o_n(\|y-x\|))$, $o_i(\|y-x\|)/\|y-x\| \to 0$ при $\|y - x\| \to 0$, $i = 1, \ldots, n$.

О п р е д е л е н и е 3.7. Функция $f(x)$, дважды дифференцируемая в каждой точке множества $X \subseteq R^n$, называется *дважды дифференцируемой на множестве* X, а если, кроме того, $\|\nabla^2 f(y) - \nabla^2 f(x)\| \to 0$ при $\|y - x\| \to 0$, $y, x \in X$, то функция $f(x)$ называется *дважды непрерывно дифференцируемой* или *дважды гладкой* на множестве X.

Класс дважды гладких на множестве X функций обозначается через $C_2(X)$. Ясно, что если функция $f(x)$ дважды дифференцируема в точке x, то ее градиент непрерывен в этой точке (т. е. $\|\nabla f(y) - \nabla f(x)\| \to 0$ при $\|y - x\| \to 0$, $y \in O(x, \varepsilon)$).

Пусть функция $f(x)$ дважды дифференцируема в точке x (и, значит, дифференцируема в некоторой ε-окрестности $O(x, \varepsilon)$ этой точки). Полагая в (3.2) $y = x + \lambda e^j$, $|\lambda| < \varepsilon$, получим

$$\nabla f(x + \lambda e^j)_i - \nabla f(x)_i = \nabla^2 f(x)_{ij} \cdot \lambda + o_i(\lambda), \;\; i = 1, \ldots, n, \qquad (3.3)$$

где $\nabla^2 f(x)_{ij}$ – (i, j)-й элемент матрицы $\nabla^2 f(x)$. Деля обе части равенства (3.3) на $\lambda \neq 0$ и устремляя затем $\lambda \to 0$, получим, что

$$\nabla^2 f(x)_{ij} \; = \; \lim_{\lambda \to 0} \frac{\nabla f(x + \lambda e^j)_i - \nabla f(x)_i}{\lambda} \;\; \stackrel{\text{def}}{=} \;\; \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \; .$$

То есть матрица $\nabla^2 f(x)$ представляет собой матрицу вторых частных производных функции $f(x)$ в точке x. Она называется *матрицей Гессе* или *гессианом* функции $f(x)$ в точке x. Таким образом, дважды дифференцируемость функции $f(x)$ в точке x влечет существование и независимость от порядка дифференцирования ее вторых частных производных в этой точке. Обратное, вообще говоря, верно лишь в том случае, если все частные производные 2-го порядка функции $f(x)$ существуют в некоторой окрестности точки x и непрерывны в самой точке x.

Если функция $f(x)$ дифференцируема в некоторой ε-окрестности $O(x, \varepsilon)$ точки x и дважды дифференцируема в самой точке X, то для любого $y \in O(x, \varepsilon)$ справедлива *формула Тейлора*

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle + o(\|y - x\|^2) \; ,$$

где $o(\|y-x\|^2)/\|y-x\|^2 \to 0$ при $\|y-x\| \to 0$. Выражение $\langle \nabla^2 f(x)(y-x), y - x \rangle$ называется *вторым дифференциалом* функции $f(x)$ в точке x и обозначается $\mathrm{d}^2 f(x)$.

Задачи и упражнения

3.1. Доказать непрерывность линейной функции $f(x) = \langle c, x \rangle$ и расстояния $f(x) = \rho(x, a) = \|x - a\|$.

3.2. *Множеством Лебега* функции $f(x)$, определенной на множестве $X \subseteq R^n$, называется множество

$$X_k = \{x \in X : \; f(x) \leq k = const\}.$$

Доказать, что множество Лебега функции $f(x)$, непрерывной на замкнутом множестве X, замкнуто.

3.3. Пусть функции $f(x)$ и $g(x)$ непрерывны в точке $a \in R^n$. Доказать, что функции $f(x) \pm g(x)$, $f(x) \cdot g(x)$ и $f(x)/g(x)$ непрерывны в точке a (частное при условии $g(a) \neq 0$).

3.4. Доказать, что если функция $f(x)$ дифференцируема в точке x, то она и непрерывна в этой точке.

3.5. Пусть функции $f(x)$ и $g(x)$ дифференцируемы (дважды дифференцируемы) в точке x. Доказать, что функции $f(x) \pm g(x)$, $f(x) \cdot g(x)$ и $f(x)/g(x)$(при $g(x) \neq 0$) также дифференцируемы в точке x, а их «производные» имеют вид:

$$\nabla[f(x) \pm g(x)] = \nabla f(x) \pm \nabla g(x) \; ;$$

$$\nabla^2[f(x) \pm g(x)] = \nabla^2 f(x) \pm \nabla^2 g(x) \; ;$$

$$\nabla[f(x) \cdot g(x)] = f(x)\nabla g(x) + g(x)\nabla f(x) \; ;$$

$$\nabla^2[f(x) \cdot g(x)] = f(x)\nabla^2 g(x) + g(x)\nabla^2 f(x) + \\ + [\nabla g(x)\nabla f(x)^T + \nabla f(x)\nabla g(x)^T] \; ;$$

$$\nabla \left[\frac{f(x)}{g(x)}\right] = \frac{1}{g^2(x)}[\, g(x)\nabla f(x) - f(x)\nabla g(x)] \; ;$$

$$\nabla^2 \left[\frac{f(x)}{g(x)}\right] = \frac{1}{g^3(x)}\{g(x)[\, g(x)\nabla^2 f(x) - f(x)\nabla^2 g(x)] - \\ - [\nabla g(x)\nabla f(x)^T + \nabla f(x)\nabla g(x)^T] + 2f(x)\nabla g(x)\nabla g(x)^T\} \; .$$

3.6. Пусть функция $h(x) : R^n \to R^1$ дифференцируема (дважды дифференцируема) в точке x, а функция $\varphi(h) : R^1 \to R^1$ дифференцируема (дважды дифференцируема) в точке $h = h(x)$. Доказать, что тогда функция $f(x) = \varphi(h(x))$ дифференцируема (дважды дифференцируема) в точке x и

$$\nabla f(x) = \varphi'(h(x))\nabla h(x),$$

$$\nabla^2 f(x) = \varphi''(h(x))\nabla h(x)\nabla h(x)^T + \varphi'(h(x))\nabla^2 h(x) \; .$$

3.7. Пусть $\varphi(h) = h|h|$ – функция одной переменной h. Доказать, что $\varphi'(h) = 2|h|$.

3.8. Пусть функция $h(x) : R^n \to R^1$. Функция $h_+(x) = \max\{\,0,\ h(x)\,\}$ называется *срезкой* функции $h(x)$. Положим $p(x) = [h_+(x)]^2$. Тогда, если функция $h(x)$ дифференцируема в точке x, то функция $p(x)$ также дифференцируема в точке x и

$$\nabla p(x) = 2h_+(x)\nabla h(x)\ .$$

(Указание: представить $h_+(x)$ в виде $h_+(x) = \frac{1}{2}(h(x) + |h(x)|)$ и воспользоваться результатами упражнений 3.6 и 3.7.)

3.9. Вычислить градиент и гессиан функций

$$f(x) = \langle c, x\rangle\ ,$$

$$f(x) = \tfrac{1}{2}\|x\|^2\ ,$$

$$f(x) = \tfrac{1}{2}\langle Gx, x\rangle + \langle c, x\rangle + d\ ,$$

$$f(x) = \tfrac{1}{2}\|Ax - b\|^2\ ,\ \ A \in M_{m,n}\ ,$$

$$f(x) = \|x\|\ ,\ \ x \neq 0\ ,$$

$$f(x) = \|Ax - b\|\ ,\ \ Ax \neq b\ ,\ \ A \in M_{m,n}\ .$$

3.10. Пусть функция $f(x)$ дифференцируема в точке x. *Производной* функции $f(x)$ в точке x *по направлению* p , $\|p\,\| = 1$, называется предел

$$\frac{\partial f(x)}{\partial p} = \lim_{\lambda \to 0} \frac{f(x + \lambda p) - f(x)}{\lambda}$$

(если этот предел существует). В частности, если $p = e^i$, то производная по направлению – это частная производная функции $f(x)$ по переменной x_i. Доказать, что

$$\frac{\partial f(x)}{\partial p} = \langle \nabla f(x), p\rangle.$$

3.11. Пусть функция $f(x)$ определена на выпуклом множестве X из R^n. Для произвольных (фиксированных) $x, y \in X$ и $\lambda \in [0, 1]$ положим

$g(\lambda) = f(x+\lambda[y-x])$. Доказать, что если функция $f(x) \in C_r(X), r = 1, 2$, то функция одной переменной $g(\lambda) \in C_r([0, 1])$ и

$$g'(\lambda) = \langle \nabla f(x + \lambda[y - x]), y - x \rangle \,,$$

$$g''(\lambda) = \langle \nabla^2 f(x + \lambda[y - x])(y - x), y - x \rangle.$$

3.12. Пусть функция $f(x)$ определена на выпуклом множестве X. Доказать, что при соответствующих предположениях о дифференцируемости справедливы следующие формулы:

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + \tau[y - x]), y - x \rangle \mathrm{d}\tau,$$

$$f(y) - f(x) = \langle \nabla f(x + \theta[y - x]), y - x \rangle \,, \ \ 0 < \theta < 1,$$

$$f(y) - f(x) = \langle \nabla f(x), y - x \rangle + \frac{1}{2} \left\langle \nabla^2 f(x + \theta[y - x])(y - x), y - x \right\rangle \,,$$
$$0 < \theta < 1,$$

$$\nabla f(y) - \nabla f(x) = \int_0^1 \nabla^2 f(x + \tau[y - x])(y - x)\mathrm{d}\tau,$$

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle = \langle \nabla^2 f(x + \theta[y - x])(y - x), y - x \rangle,$$
$$0 < \theta < 1.$$

(Указание: положить $g(\lambda) = f(x+\lambda[y-x]), \ 0 \le \lambda \le 1$, и воспользоваться результатами упражнения 3.11.)

3.13. Пусть функция $f(x) \in C_1(X)$. Говорят, что градиент этой функции, $\nabla f(x)$, удовлетворяет на X *условию Липшица* с константой $L > 0$, если для любых точек $x, y \in X$

$$\|\nabla f(y) - \nabla f(x)\| \le L\|y - x\|.$$

Класс таких функций обозначается $C_{1,1}(X)$.

Доказать, что если функция $f(x) \in C_{1,1}(X)$, где X – выпуклое множество, то

$$f(y) \le f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} L\|y - x\|^2 \quad \forall x, y \in X \ .$$

§ 4. Выпуклые функции

О п р е д е л е н и е 4.1. Функция $f(x)$, определенная на выпуклом множестве $X \subseteq R^n$, называется *выпуклой* на этом множестве, если для любых $x, y \in X$ и числа $\lambda \in [0,1]$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y). \tag{4.1}$$

Если при любых $x, y \in X,\ x \neq y$ и $\lambda \in (0,1)$ неравенство (4.1) строгое, то функция $f(x)$ называется *строго выпуклой* на множестве X.

Графически выпуклость функции $f(x)$ означает следующее: при любых $x, y \in X$ график функции $f(x)$ на отрезке $[x, y]$ лежит не выше хорды, соединяющей точки $(x, f(x))$ и $(y, f(y))$ (рис. 4.1).

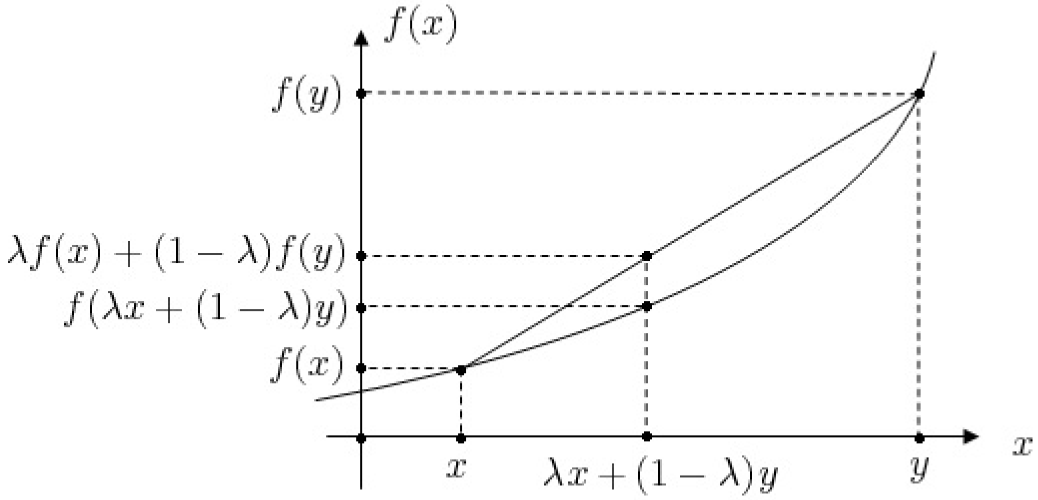

Рис. 4.1

З а м е ч а н и е. Функция $f(x)$, определенная на выпуклом множестве X, называется *вогнутой* на этом множестве, если

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y) \quad \forall\, x, y \in X, \lambda \in [0,1]\,.$$

Ясно, что если функция $f(x)$ выпукла на X, то функция $[-f(x)]$ вогнута на этом множестве, так что свойства вогнутых функций могут быть легко получены из соответствующих свойств выпуклых функций.

Простейшими примерами выпуклых функций (на R^n) являются линейная функция $f(x) = \langle c, x\rangle$ и расстояние $f(x) = \|x - a\|$. В самом деле,

$$\langle c, \lambda x + (1-\lambda) y\rangle = \lambda \langle c, x\rangle + (1-\lambda)\langle c, y\rangle \ ,$$

$$\|\lambda x + (1-\lambda) y - a\| = \|\lambda (x-a) + (1-\lambda)(y-a)\| \le$$

$$\le \lambda \|x-a\| + (1-\lambda)\|y-a\|$$

при $0 \le \lambda \le 1$. Линейная функция является одновременно и вогнутой на всем пространстве R^n.

Т е о р е м а 4.1. Функция $f(x)$, выпуклая на выпуклом множестве X, непрерывна во всех внутренних точках этого множества. В частности, если $X = R^n$, то функция $f(x)$ непрерывна на всем пространстве.

Т е о р е м а 4.2. Пусть функция $f(x) \in C_1(X)$. Тогда следующие утверждения эквивалентны:

1) $f(x)$ выпукла на X;

2) для любых $x, y \in X$

$$f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle;$$

3) для любых $x, y \in X$

$$\langle \nabla f(y) - \nabla f(x), y - x\rangle \ \ge 0.$$

Т е о р е м а 4.3. Функция $f(x) \in C_2(X)$ выпукла на X тогда и только тогда, когда для любых $x, y \in X$

$$\langle \nabla^2 f(x)(y-x), y - x\rangle \ \ge 0. \tag{4.2}$$

Если $X = R^n$, то условие (4.2) эквивалентно условию

$$\langle \nabla^2 f(x) p, p\rangle \ \ge 0 \quad \forall\, x, p \in R^n,$$

т. е. гессиан функции $f(x)$ неотрицательно определен на всем пространстве.

О п р е д е л е н и е 4.2. Функция $f(x)$, определенная на выпуклом множестве $X \subseteq R^n$, называется *сильно выпуклой* на этом множестве, если существует число $\mu > 0$ (*константа сильной выпуклости*) такое, что для любых $x, y \in X$ и $\lambda \in [0, 1]$

$$f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y) - \frac{1}{2}\mu\lambda(1-\lambda)\|x-y\|^2 \, .$$

Примером сильно выпуклой функции может служить функция $f(x) = \|x-a\|^2$. Действительно,

$$\|\lambda x + (1-\lambda)y - a\|^2 =$$

$$= \lambda\|x-a\|^2 + (1-\lambda)\|y-a\|^2 - \frac{1}{2}\cdot 2\lambda(1-\lambda)\|x-y\|^2 ,$$

так что $\|x-a\|^2$ сильно выпукла на R^n с константой сильной выпуклости $\mu = 2$. Тогда, очевидно, сильно выпуклой является и любая функция $f(x)$ вида $f(x) = h(x) + K\|x-a\|^2$, где $K > 0$, а функция $h(x)$ – выпуклая. Вообще, сумма выпуклой и сильно выпуклой функций является сильно выпуклой функцией. Ясно также, что сильно выпуклая на X функция будет выпуклой и даже строго выпуклой на этом множестве.

Графически сильная выпуклость функции $f(x)$ означает, что при любых $x, y \in X$ график функции $f(x)$ на отрезке $[x, y]$ лежит не выше параболы $\phi(\lambda) = \lambda f(x) + (1-\lambda)f(y) - \frac{1}{2}\mu\lambda(1-\lambda)\|x-y\|^2$, проходящей через точки $(x, f(x))$ и $(y, f(y))$ (рис. 4.2).

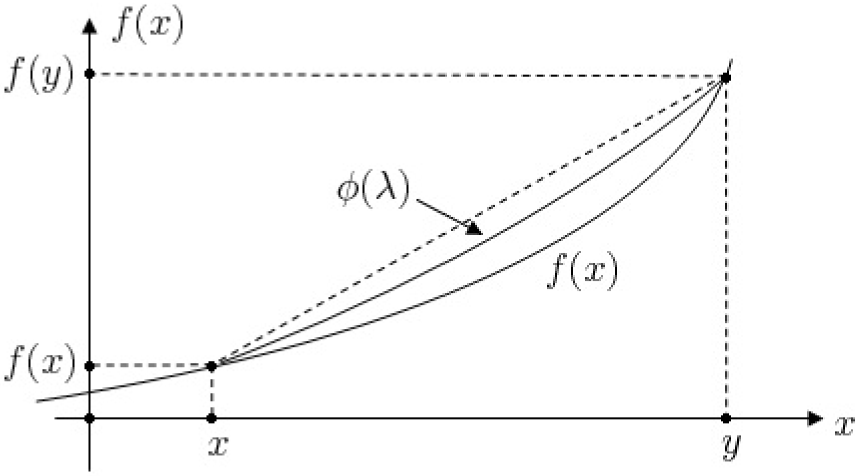

Рис. 4.2

Т е о р е м а 4.4. Пусть функция $f(x)$ сильно выпукла и непрерывна на выпуклом замкнутом множестве $X \subseteq R^n$ (в частности, может быть $X = R^n$). Тогда ее множество Лебега

$$X_k = \{x \in X : f(x) \le k\}$$

ограничено при любом k.

Т е о р е м а 4.5. Пусть функция $f(x) \in C_1(X)$. Тогда следующие утверждения эквивалентны:

1) $f(x)$ сильно выпукла на X с константой сильной выпуклости μ;

2) для любых $x, y \in X$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\mu \|y - x\|^2 \ ;$$

3) для любых $x, y \in X$

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \ \geq \mu \|y - x\|^2 \ .$$

Т е о р е м а 4.6. Функция $f(x) \in C_2(X)$ сильно выпукла на X с константой сильной выпуклости μ тогда и только тогда, когда для любых $x, y \in X$

$$\langle \nabla^2 f(x)(y - x), y - x \rangle \ \geq \mu \|y - x\|^2. \tag{4.3}$$

Если $X = R^n$, то условие (4.3) эквивалентно условию

$$\langle \nabla^2 f(x)p, p \rangle \ \geq \mu \|p\|^2 \quad \forall\, x, p \in R^n.$$

П р и м е р 4.1. Рассмотрим квадратичную функцию

$$f(x) = ax_1^2 + +bx_1x_2 + cx_2^2 \ .$$

Градиент функции $f(x)$, $\nabla f(x) = (2ax_1 + bx_2, bx_1 + 2cx_2)$, гессиан

$$\nabla^2 f(x) = \begin{bmatrix} 2a & b \\ b & 2c \end{bmatrix} .$$

Главные угловые миноры $\triangle_1 = 2a$, $\triangle_2 = 4ac - b^2$. При $a > 0$ и $4ac > b^2$ имеем $\triangle_1 > 0$, $\triangle_2 > 0$ и, следовательно, по критерию Сильвестра, матрица $\nabla^2 f(x)$ положительно определена для всех $x = (x_1, x_2)$. Более того, так как матрица $\nabla^2 f(x)$ не зависит от x, то

$$\langle \nabla^2 f(x)p, p \rangle \geq \mu \|p\|^2 \quad \text{для любых } x = (x_1, x_2),\ p = (p_1, p_2)\,,$$

где μ – наименьшее собственное число матрицы $\nabla^2 f(x)$. Следовательно, функция $f(x)$ сильно выпукла на R^n с константой сильной выпуклости μ. Если $a \geq 0$, $c \geq 0$ и $4ac \geq b^2$, то все главные миноры матрицы $\nabla^2 f(x)$ неотрицательны (главные миноры первого порядка равны $2c$ и $2a$). Следовательно, матрица $\nabla^2 f(x)$ неотрицательно определена для всех $x = (x_1, x_2)$ и функция $f(x)$ выпукла на R^n.

Задачи и упражнения

4.1. Доказать, что если функции $f_i(x)$, $i = 1, \dots, m$ выпуклы на выпуклом множестве X, то их неотрицательная линейная комбинация $f(x) = \sum\limits_{i=1}^{m} \alpha_i f_i(x)$, $\alpha_i \geq 0$, $i = 1, \dots, m$ также выпукла на X. В частности, выпуклыми являются сумма любого числа выпуклых функций и произведение выпуклой функции на положительное число. Если для некоторого $i = k$ функция $f_k(x)$ сильно выпукла и $\alpha_k > 0$, то функция $f(x)$ сильно выпукла.

4.2. Доказать *неравенство Иенсена*: Если функция $f(x)$ выпукла на выпуклом множестве X, то

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \leq \sum_{i=1}^{m} \lambda_i f(x_i) ,$$

где $x_i \in X$, $\lambda_i \geq 0$, $i = 1, \dots, m$, $\sum\limits_{i=1}^{m} \lambda_i = 1$.

4.3. Доказать, что если функция $f(x)$ выпукла на R^n, то функция одной переменной $g(\lambda) = f(x + \lambda p)$ выпукла при любых $x, p \in R^n$.

4.4. Пусть функция $f(x)$ выпукла и неотрицательна на выпуклом множестве X. Доказать, что тогда функция $\varphi(x) = f^2(x)$ также выпукла на X. Если функция $f(x)$ строго выпукла, то строго выпукла и функция $\varphi(x)$.

4.5. Пусть функция $g(x)$ вогнута и положительна на выпуклом множестве X. Доказать, что функция $f(x) = [g(x)]^{-1}$ выпукла на X.

4.6. Пусть функция $h(x)$ выпукла на выпуклом множестве X. Тогда её срезка $h_+(x) = \max\{0, h(x)\}$ также выпукла на X (доказать).

4.7. Пусть функции $f_i(x)$, $i = 1, \dots, m$, выпуклы на множестве X. Доказать, что тогда функция $\varphi(x) = \max\limits_{i=1,\dots,m} f_i(x)$ также выпукла на X.

4.8. Доказать, что если функция $f(x)$ выпукла на выпуклом множестве X, то её множество Лебега $X_k = \{x \in X : f(x) \leq k\}$ выпукло при любом k.

4.9. Пусть функция $f(x)$ выпукла и непрерывно дифференцируема на выпуклом множестве X. Показать, что гиперплоскость

$$H_k = \{x \in R^n : \langle \nabla f(x^k), x - x^k \rangle = 0\}$$

является опорной ко множеству Лебега

$$X_k = \{x \in X : f(x) \leq f(x^k)\}$$

в точке x^k.

4.10. Пусть функции $h_i(x), i = 1, \ldots, m$ выпуклы на выпуклом множестве X. Доказать, что тогда множество $X_0 = \{x \in X : \; h_i(x) \leq 0, \; i = 1, \ldots, m\}$ является выпуклым.

4.11. Пусть функции $g_j(x), \; j = 1, \ldots, r$ вогнуты на выпуклом множестве X. Доказать, что тогда множество $X_0 = \{x \in X : g_j(x) \geq 0 \;, j = 1, \ldots, r\}$ является выпуклым.

4.12. Вывести критерии строгой выпуклости гладких и дважды гладких функций.

4.13. Доказать теорему 4.4 в предположении, что функция $f(x)$ непрерывно дифференцируема на X.

4.14. Пусть функция $f(x)$ сильно выпукла и дифференцируема на выпуклом множестве X. Доказать, что

$$\nabla f(x) \neq \nabla f(y) \;\; \text{при} \quad x \neq y \;, \quad x, y \in X \;.$$

4.15. Показать, что квадратичная функция

$$f(x) = \frac{1}{2}\langle Gx, x \rangle + \langle c, x \rangle + d$$

выпукла тогда и только тогда, когда матрица G неотрицательно определена и сильно выпукла тогда и только тогда, когда матрица G определена положительно. Может ли она быть строго выпуклой?

4.16. Сформулировать критерии выпуклости, строгой и сильной выпуклости функций одной переменной.

4.17. Функция $f(x)$ $(R^n \to R^1)$ называется *сепарабельной*, если она представима в виде

$$f(x) = f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n).$$

Сепарабельными являются, например, функции $f(x) =< c, x >$ и $f(x) = = \|x - a\|^2$. Доказать, что сепарабельная функция $f(x)$ выпукла тогда и только тогда, когда выпуклы функции $f_1(x_1), \ldots, f_n(x_n)$.

4.18. В следующих примерах исследовать функции $f(x)$ на выпуклость (вогнутость) на заданных множествах X. Для сильно выпуклых (вогнутых) функций вычислить константу сильной выпуклости (вогнутости).

a) $f(x) = x_1^6 + x_2^2 + x_3^2 + x_4^2 + 10x_1 + 5x_2 - 3x_4 - 20, X = R^4$;

b) $f(x) = \exp(2x_1 + x_2), X = R^2$;

c) $f(x) = -x_1^3 - x_2^3 - x_3^3 + 10x_1 - x_2 + 15x_3 + 6,$

$$X = R^3_+ = \{x \in R^3 : x \geq 0\};$$

d) $f(x) = x_1^2 + x_2^2 + \frac{1}{2}x_3^2 + x_1x_2 - x_3 + 10, X = R^3$;

e) $f(x) = -x_1^2 - x_2^2 - 2x_3^2 + x_1x_2 + x_1x_3 + x_2x_3 + 5x_2,\ X = R^3$;

f) $f(x) = x_1^3 + 2x_3^3 + 10x_1 + x_2 - 5x_3 + 6,\ X = R^3_- = \{x \in R^3 : x \leq 0\}$;

g) $f(x) = 5x_1^4 + x_2^6 + x_3^2 - 13x_1 + 7x_3 - 8,\ R = E^3$;

h) $f(x) = -\frac{1}{2}x_2^7 + \frac{1}{2}x_3^4 + 2x_2x_3 + 11x_1 + 6,\ X = R^3_-$;

i) $f(x) = (x_1^2 - x_2)^2,\ X = R^2$;

j) $f(x) = x_1 \exp(-x_1 - x_2),\ X = R^2$;

k) $f(x) = 2x_1^2 + x_2^2 + 2x_3^2 + x_1x_2 - 6x_1x_3,\ X = R^3$;

l) $f(x) = 3x_1^2 + x_2^2 + 2x_3^2 + x_1x_2 + 3x_1x_3 + x_2x_3 + 3x_2,\ X = R^3$;

m) $f(x) = 5x_1^2 + \frac{1}{2}x_2^2 + 4x_3^2 + x_1x_2 + 2x_1x_3 + 2x_2x_3 + x_3,\ X = R^3$;

n) $f(x) = -2x_1^2 - \frac{1}{2}x_2^2 - 5x_3^2 + \frac{1}{2}x_1x_2 + 2x_1x_3 + x_2x_3 + 3x_1 - 2x_2,\ X = R^3$;

o) $f(x) = \sqrt{1 + x_1^2 + x_2^2},\ X = R^2$;

p) $f(x) = \ln \sum\limits_{i=1}^{n} \exp(x_i),\ X = R^n$.

Глава 2
Экстремумы функций многих переменных

§ 1. Минимумы и максимумы

Пусть функция $f(x)$ определена на множестве $X \subseteq R^n$.

О п р е д е л е н и е 1.1. Функция $f(x)$ называется *ограниченной снизу* на множестве X, если существует число α такое, что $\alpha \leq f(x)$ для всех $x \in X$. Число α называется *нижней границей* функции $f(x)$ на множестве X.

Ясно, что если α является нижней границей функции $f(x)$ на множестве X, то любое число $\alpha' < \alpha$ также является нижней границей функции $f(x)$ на множестве X.

О п р е д е л е н и е 1.2. Пусть функция $f(x)$ ограничена снизу на множестве X. Число α называется *точной нижней границей* или *нижней гранью* функции $f(x)$ на множестве X и обозначается $\inf\limits_{x \in X} f(x)$, если оно является нижней границей, а любое число $\alpha' > \alpha$ нижней границей уже не является. То есть число α – точная нижняя граница функции $f(x)$ на множестве X, если $\alpha \leq f(x)$ для всех $x \in X$ и для любого числа $\alpha' > \alpha$ найдется элемент $x \in X$ такой, что $f(x) < \alpha'$.

Т е о р е м а 1.1. Пусть функция $f(x)$ ограничена снизу на множестве X. Тогда ее нижняя грань на этом множестве существует.

Единственность нижней грани вытекает из ее определения.

О п р е д е л е н и е 1.3. Функция $f(x)$ называется *не ограниченной снизу* на множестве X, если для любого числа α найдется элемент $x \in X$ такой, что $f(x) < \alpha$.

Если функция $f(x)$ не ограничена снизу на множестве X, то ее нижняя грань на этом множестве принимается равной $-\infty: \inf\limits_{x \in X} f(x) = -\infty$.

О п р е д е л е н и е 1.4. Точка $x^* \in X$ называется *точкой минимума*, а величина $f(x^*)$ – минимальным значением, или *минимумом* функции $f(x)$ на множестве X, если $f(x^*) \le f(x)$ для всех $x \in X$.

Минимум функции $f(x)$ на множестве X обозначается как $\min\limits_{x \in X} f(x)$ или просто $\min\limits_{X} f(x)$. Точка минимума обозначается как $\operatorname{argmin}\limits_{x \in X} f(x)$, а множество точек минимума функции $f(x)$ на множестве X – $\operatorname{Argmin}\limits_{x \in X} f(x)$. Очевидно, что если минимум функции $f(x)$ на множестве X существует, то $\min\limits_{x \in X} f(x) = \inf\limits_{x \in X} f(x)$, и в этом случае говорят, что функция $f(x)$ достигает своей нижней грани на множестве X (в точке $x^* = \operatorname{argmin}\limits_{x \in X} f(x)$).

П р и м е р 1.1. Пусть функция $f(x) = e^x, \;\; x \in R^1$. Тогда

$$\inf_{x \ge 0} f(x) = \min_{x \ge 0} f(x) = 1, \; \operatorname{argmin}_{x \ge 0} f(x) = 0, \; \operatorname{Argmin}_{x \ge 0} f(x) = \{0\}.$$

В то же время $\inf\limits_{R^1} f(x) = 0 < f(x)$ для всех $x \in R^1$, т. е. минимум функции $f(x)$ на R^1 не существует: $\operatorname{Argmin}\limits_{R^1} f(x) = \emptyset$.

Функция $f(x)$ называется *ограниченной сверху* на множестве X, если существует число β (*верхняя граница* функции $f(x)$ на множестве X) такое, что $f(x) \le \beta$ для всех $x \in X$. Если, кроме того, для любого $\beta' < \beta$ найдется $x \in X$ такое, что $f(x) > \beta'$, то β называется *точной верхней границей* или *верхней гранью* функции $f(x)$ на множестве X. Верхняя грань функции $f(x)$, ограниченной сверху на множестве X, существует и единственна. Если функция $f(x)$ не ограничена сверху на множестве X, т. е. для любого числа β найдется $x \in X$ такой, что $f(x) > \beta$, то функция $f(x)$ называется *не ограниченной сверху* на множестве X, а ее верхняя грань принимается равной $+\infty$. Функция, ограниченная снизу и сверху на некотором множестве, называется *ограниченной* на этом множестве.

Точка $x^* \in X$ называется точкой максимума функции $f(x)$ на множестве X, а величина $f(x^*)$ – максимальным значением или *максимумом* функции $f(x)$ на множестве X, если $f(x^*) \geq f(x)$ для всех $x \in X$.

Верхняя грань функции $f(x)$ на множестве X обозначается $\sup\limits_{x \in X} f(x)$, а ее максимум на этом множестве – $\max\limits_{x \in X} f(x)$.

Т е о р е м а 1.2 (*теорема Вейерштрасса*). Пусть функция $f(x)$ непрерывна на замкнутом и ограниченном множестве X. Тогда она достигает на этом множестве своих точных верхней и нижней граней.

То есть максимум и минимум функции, непрерывной на компактном множестве, существуют.

О п р е д е л е н и е 1.5. Точка $x^* \in X$ называется *точкой локального минимума*, а величина $f(x^*)$ – *локальным минимумом* функции $f(x)$ на множестве X, если существует число $\varepsilon > 0$ такое, что

$$f(x^*) \leq f(x) \quad \forall\, x \in X \cap O(x^*, \varepsilon) \ . \tag{1.1}$$

Если при некотором $\varepsilon > 0$ равенство в (1.1) возможно лишь при $x = x^*$, то x^* – *точка строгого локального минимума*, а $f(x^*)$ – *строгий локальный минимум* функции $f(x)$ на множестве X.

Локальный минимум функции $f(x)$ на множестве X обозначают как $\operatorname*{locmin}\limits_{x \in X} f(x)$. В отличие от локального минимума, минимум в смысле определения 1.4 называют также *глобальным* или *абсолютным* минимумом функции $f(x)$ на множестве X. Чтобы подчеркнуть, что речь идет именно о глобальном минимуме, его обозначают также как $\operatorname*{absmin}\limits_{x \in X} f(x)$. Функция $f(x)$ может иметь на множестве X различные локальные минимумы. Глобальный минимум, если он существует, – единственный. Очевидно также, что всякий глобальный минимум является одновременно и локальным, обратное же, вообще говоря, не верно. Однако, если функция $f(x)$ выпукла на выпуклом множестве X, то всякая точка локального минимума является одновременно и точкой глобального минимума, причем множество точек минимума $X_* = \operatorname*{argmin}\limits_{X} f(x)$ выпукло. Если $f(x)$ строго выпукла на X, то множество X_* может содержать не более одной точки. Наконец, если $f(x)$ непрерывна и сильно выпукла, а множество X – выпукло и замкнуто (в частности, может быть $X = R^n$), то точка минимума существует и единственна.

Аналогично определяются локальный и глобальный максимумы функции $f(x)$ на множестве X. Локальные и глобальные минимумы и максимумы называются локальными и глобальными, соответственно, *экстремумами* функции $f(x)$ на множестве X. На рис. 1.1 точки x_1, x_3 и x_5 – точки локального минимума, точки x_2, x_4 и x_6 – точки локального максимума функции $f(x)$ на отрезке $[a, b]$, причем точки x_3 и x_6 являются одновременно и точками глобального минимума и максимума, соответственно.

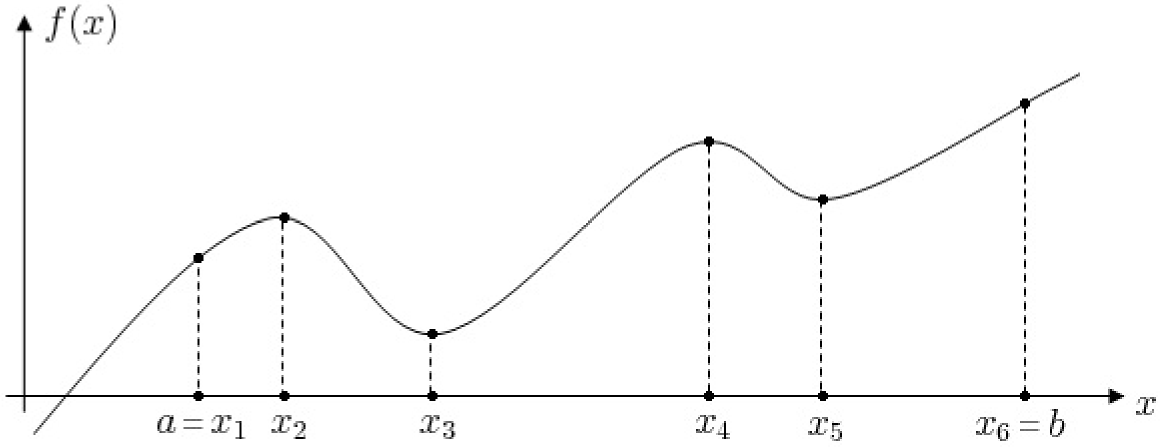

Рис. 1.1

Задачи и упражнения

1.1. Пусть функции $f(x)$ и $g(x)$ ограничены на множестве X. Доказать справедливость следующих соотношений:

a) $\inf\limits_X f(x) = -\sup\limits_X[-f(x)]$,

$\sup\limits_X f(x) = -\inf\limits_X[-f(x)]$;

b) $\inf\limits_X \lambda f(x) = \lambda \inf\limits_X f(x)$, $\lambda > 0$,

$\sup\limits_X \lambda f(x) = \lambda \sup\limits_X f(x)$, $\lambda > 0$;

c) $\inf\limits_X \lambda f(x) = \lambda \sup\limits_X f(x)$, $\lambda < 0$,

$\sup\limits_X \lambda f(x) = \lambda \inf\limits_X f(x)$, $\lambda < 0$;

d) $\inf\limits_X [f(x)+c] = \inf\limits_X f(x) + c$,

$\sup\limits_X [f(x)+c] = \sup\limits_X f(x) + c$;

e) если $f(x) \le g(x)$ для всех $x \in X$, то

$\inf\limits_X f(x) \le \inf\limits_X g(x)$,

$\sup\limits_X f(x) \le \sup\limits_X g(x)$;

f) $\inf\limits_X [f(x)+g(x)] \ge \inf\limits_X f(x) + \inf\limits_X g(x)$,

$\sup\limits_X [f(x)+g(x)] \le \sup\limits_X f(x) + \sup\limits_X g(x)$;

g) $\inf\limits_X [f(x)-g(x)] \le \inf\limits_X f(x) - \inf\limits_X g(x)$,

$\sup\limits_X [f(x)-g(x)] \ge \sup\limits_X f(x) - \sup\limits_X g(x)$;

h) если $f(x) \ge 0$ и $g(x) \ge 0$ для всех $x \in X$, то

$\inf\limits_X [f(x)g(x)] \ge \inf\limits_X f(x) \cdot \inf\limits_X g(x)$,

$\sup\limits_X [f(x)g(x)] \le \sup\limits_X f(x) \cdot \sup\limits_X g(x)$;

i) $\inf\limits_X \min\{f(x), g(x)\} = \min\{\inf\limits_X f(x), \inf\limits_X g(x)\}$,

$\sup\limits_X \max\{f(x), g(x)\} = \max\{\sup\limits_X f(x), \sup\limits_X g(x)\}$;

j) $\inf\limits_X \max\{f(x), g(x)\} \ge \max\{\inf\limits_X f(x), \inf\limits_X g(x)\}$,

$\sup\limits_X \min\{f(x), g(x)\} \le \min\{\sup\limits_X f(x), \sup\limits_X g(x)\}$.

Проверить выполнение соотношений a) – j) для случая, когда функции $f(x)$ и (или) $g(x)$ не ограничены снизу (сверху) на множестве X. Напомним, что

$$(+\infty) + a = +\infty, \text{ если } a \ne -\infty\,; \quad (+\infty)\cdot a = +\infty, \text{ если } a > 0\,;$$

$$(-\infty) + a = -\infty, \text{ если } a \ne +\infty\,; \quad (+\infty)\cdot a = -\infty, \text{ если } a < 0\,;$$

$$(-\infty)\cdot a = -\infty, \text{ если } a > 0\,; \quad (-\infty)\cdot a = +\infty, \text{ если } a < 0\,,$$

а действия $(+\infty) + (-\infty), (+\infty)\cdot 0$ и $(-\infty)\cdot 0$ лишены смысла.

1.2. Пусть функции $f(x)$ и $g(x)$ определены и непрерывны на замкнутом ограниченном множестве $X \in R^n$, $X_1 = \operatorname{argmin}_X f(x)$, $X_2 = \operatorname{argmax}_X f(x)$. Доказать, что

i. множества X_1 и X_2 замкнуты;

ii. $\min_X [f(x) + g(x)] \le \min_X f(x) + \min_{X_1} g(x)$;

iii. $\max_X [f(x) + g(x)] \ge \max_X f(x) + \max_{X_2} g(x)$.

1.3. Пусть $X \subset R^n$ – произвольное множество, $\overline{X}$ – замыкание X, функция $f(x)$ непрерывна на $\overline{X}$. Доказать, что

$$\inf_X f(x) = \inf_{\overline{X}} f(x)\,, \quad \sup_X f(x) = \sup_{\overline{X}} f(x)\,.$$

1.4. Пусть функция $F(x, y)$ определена на декартовом произведении множеств X и Y. Доказать *неравенство минимаксов*:

$$\sup_X \inf_Y F(X, Y) \ \le\ \inf_Y \sup_X F(X, Y)\,.$$

1.5. Пусть функция $F(x, y)$ определена на декартовом произведении множеств X и Y. Доказать, что

$$\inf_{X \times Y} F(x, y) = \inf_X \inf_Y F(x, y) = \inf_Y \inf_X F(x, y)\,.$$

1.6. Пусть функция $f(x)$ – сепарабельная: $f(x) = f_1(x_1) + \ldots + f_n(x_n)$, $x = (x_1, \ldots, x_n)$, $x_i \in X_i$, $i = 1, \ldots, n$, множество $X = X_1 \times \ldots \times X_n$. Доказать, что $\min_X f(x)$ существует тогда и только тогда, когда существуют экстремумы $\min_{X_i} f_i(x_i)$, $i = 1, \ldots, n$. При этом

$$\min_X f(x) = \sum_{i=1}^{n} \min_{X_i} f_i(x_i)\,.$$

1.7. Пусть $h(x) : X \to R^1$, $f(x)$ – возрастающая числовая функция, определенная на области значений функции $h(x)$. Доказать, что множества точек глобальных и локальных, строгих и нестрогих экстремумов функции $f(h(x))$ совпадают с соответствующими множествами для функции $h(x)$, причем

$$\min_X f(h(x)) = f\left(\min_X h(x)\right)\,, \quad \max_X f(h(x)) = f\left(\max_X h(x)\right).$$

Сформулировать и доказать аналоги данного утверждения для случая, когда f – убывающая функция (невозрастающая функция, неубывающая функция).

1.8. Показать, что все предположения теоремы 1.2 существенны. (Указание: привести примеры, когда невыполнение хотя бы одного из этих предположений приводит к отсутствию экстремума функции $f(x)$ на множестве X.)

1.9. Пусть X – замкнутое множество из R^n, функция $f(x)$ определена и непрерывна на X и существует точка $x^k \in X$ такая, что множество Лебега $X_k = \{x \in X : \ f(x) \leq f(x^k\}$ ограничено. Тогда множество $X_* = \operatorname*{argmin}_X f(x)$ не пусто. (Доказать.)

1.10. Пусть X – непустое замкнутое множество из R^n, функция $f(x)$ определена и непрерывна на X и для любой последовательности $x^k \in X$ такой, что $\|x^k\| \to \infty$ имеет место соотношение

$$\lim_{k\to\infty} f(x^k) = +\infty \qquad (\lim_{k\to\infty} f(x^k) = -\infty).$$

Тогда функция $f(x)$ достигает своего минимума (максимума) на множестве X. (Доказать.)

1.11. Пусть функция $f(x)$ непрерывна и сильно выпукла на выпуклом и замкнутом множестве $X \subset R^n$ (в частности, может быть $X = R^n$). Тогда точка $x^* = \arg\min_X f(x)$ существует и единственна. (Доказать.)

1.12. Пусть X – непустое множество из R^n. *Расстоянием от точки $v \in R^n$ до множества X* называется число $\rho(v, X) = \inf_{x\in X} \|x - v\|$. Доказать, что

$$|\rho(u, X) - \rho(v, X)| \leq \rho(u, v) \ \ \forall u, v \in R^n.$$

1.13. Пусть X – непустое выпуклое и замкнутое множество из R^n. Доказать, что для любого $v \in R^n$ существует единственная точка $\pi \in X$ такая, что $\rho(v, X) = \|\pi - v\|$. Точка π называется *проекцией точки v на множество X* и обозначается как $\pi_X(v)$ или просто $\pi(v)$.

1.14. Пусть $X \subset R^n$ – замкнутое и ограниченное множество, $f(x) = \langle c, x\rangle, c \neq 0$. Доказать, что $\operatorname*{argmin}_X f(x) \subset \partial X$, т. е. минимум линейной функции достигается на границе множества X. Если X – открытое множество, то $\operatorname*{argmin}_X f(x) = \emptyset$.

1.15. Пусть $X \subset R^n$ – выпуклое компактное множество, функция $f(x)$ вогнута и непрерывна на X, $f(x) \neq \mathrm{const}$ на X. Доказать, что минимум функции $f(x)$ на множестве X достигается на границе этого множества.

1.16. Пусть функция $f(x)$ определена и непрерывна на замкнутом ограниченном множестве $X \subset R^n$. Доказать, что если $f(x) > 0$ для всех $x \in X$, то

$$\min_X f(x) = [\max_X f^{-1}(x)]^{-1}\ ; \qquad \max_X f(x) = [\min_X f^{-1}(x)]^{-1}\ .$$

1.17. Доказать, что если функция $f(x)$ выпукла на выпуклом множестве $X \subset R^n$, то всякая точка ее локального минимума (если она существует) является одновременно и точкой глобального минимума, причем множество точек минимума $X_* = \operatorname{argmin}_X f(x)$ выпукло. Если $f(x)$ строго выпукла на X, то множество X_* может содержать не более одной точки.

1.18. Пусть функция $f(x)$ определена на множестве X. Последовательность $\{x^k\} \in X$ называется *минимизирующей* для функции $f(x)$ на множестве X, если

$$\lim_{k\to\infty} f(x^k) = \inf_{x\in X} f(x).$$

Доказать, что минимизирующая последовательность всегда существует.

1.19. Говорят, что *последовательность* $\{x^k\}$ *сходится к множеству* X, если $\rho(x^k, X) \to 0$ при $k \to \infty$. Доказать, что если функция $f(x)$ непрерывна на замкнутом ограниченном множестве $X \subset R^n$, то любая минимизирующая последовательность функции $f(x)$ на множестве X сходится к множеству $X_* = \operatorname{argmin}_{x\in X} f(x)$.

1.20. Последовательность $\{x^k\} \in X$ называется *релаксационной* для функции $f(x)$ на множестве X, если $f(x^1) \geq f(x^2) \geq \ldots$. Всякая ли релаксационная последовательность является минимизирующей? Минимизирующая – релаксационной?

1.21. Показать, что последовательность $\{x^k\}$, сходящаяся к множеству $X_* = \operatorname{argmin}_{x\in X} f(x)$, не обязательно сходится к какой-либо точке этого множества.

Указание: рассмотрите функцию одной переменной $f(x) = \cos(x)$ и последовательности

a. $x^{2k} = -\pi - \dfrac{1}{k}, \quad x^{2k+1} = \pi - \dfrac{1}{k}, \quad k = 1, 2, \ldots;$

b. $x^k = \pi(2k+1) - \dfrac{1}{k}, \quad k = 1, 2, \ldots .$

Выясните, можно ли из этих последовательностей выделить подпоследовательности, сходящиеся к конкретным точкам из X_*?

1.22. Пусть функция

$$f(x) = \frac{x^2}{x^6 + 2} \; .$$

Является ли последовательность $x^k = 2k\,(k = 1, 2, \ldots)$ минимизирующей для функции $f(x)$? Сходящейся к множеству $X_* = \operatorname*{argmin}_{x \in X} f(x)$?

1.23. Найти все точки экстремумов, дать их характеристики и вычислить верхние и нижние грани функции $f(x)$ на множестве X:

$$\text{a)} \quad f(x) = \begin{cases} x^4, & x \le -2, \\ 1, & -2 < x \le 0, \\ \cos x, & 0 < x \le \pi, \\ 2, & \pi < x \le 5, \\ 2e^{(x-5)}, & 5 < x < 10, \end{cases} \qquad X = \{x \in R^1 : \ x < 10\};$$

$$\text{b)} \quad f(x) = \begin{cases} x^2, & -1 \le x \le 1, \\ (x-3)^2 - 3, & 1 \le x \le 4, \\ 0, & 4 < x \le 5, \end{cases} \qquad X = [-1, 5] \in R^1;$$

$$\text{c)} \quad f(x) = \begin{cases} |\cos x|, & x \le 0, \\ |\sin x|, & x > 0, \end{cases} \qquad X = R^1.$$

Является ли последовательность $x^k = \dfrac{(-1)^k}{k}$, $k = 1, 2, \ldots$ минимизирующей для функции $f(x)$ из c)? Сходящейся к множеству точек глобального минимума (максимума)?

1.24. Исследовать на экстремум функцию $f(x) = |||x^2 - 1| - 1| - 1|$ на отрезке $[a, b]$ при различных a и b.

§ 2. Задача на безусловный экстремум

Задачей на безусловный экстремум называется задача нахождения минимума или максимума функции $f(x) : R^n \to R^1$ (и точек, в которых эти экстремумы достигаются) на всем пространстве R^n. Коротко эту задачу будем записывать следующим образом:

$$f(x) \to \min\,(\max)\,; \quad x \in R^n\,. \tag{2.1}$$

Здесь «$\to \min(\max)$» представляет собой единый символ и читается как «*минимизировать, найти минимум* (*максимизировать, найти максимум*)», а запись $x \in R^n$ означает, что на переменные задачи не наложено никаких ограничений. Поэтому задачу (2.1) называют также *задачей минимизации* (*максимизации*) *без ограничений* или просто *задачей безусловной минимизации* (*максимизации*). Функция $f(x)$ в задаче (2.1) называется *целевой функцией*.

Поскольку

$$\max_X f(x) = -\min_X[-f(x)]\,, \quad \min_X f(x) = -\max_X[-f(x)]\,,$$

то при изучении задач на минимум и максимум можно ограничиться лишь одной из них: задача на максимум сводится к задаче на минимум, а задача на минимум – к задаче на максимум изменением знака целевой функции. Следуя традиции, сложившейся в литературе по оптимизации, мы будем рассматривать, как правило, задачу на минимум.

Классический метод решения задачи (2.1) основан на следующих утверждениях.

Т е о р е м а 2.1 (*теорема Ферма*). Пусть x^* – точка локального минимума функции $f(x)$. Если функция $f(x)$ дифференцируема в точке x^*, то

$$\nabla f(x^*) = 0\,. \tag{2.2}$$

О б с у ж д е н и е

1. Так как всякая точка глобального минимума является одновременно и точкой локального минимума, то условие (2.2) является также необходимым условием глобального минимума.

2. Если функция $f(x)$ выпукла и непрерывно дифференцируема на всем пространстве, то условие (2.2) является не только необходимым,

но и достаточным условием глобального минимума функции $f(x)$. Это следует непосредственно из критерия выпуклости дифференцируемых функций (теорема 1.4.2):

$$f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle \quad \forall\, x \in R^n \ .$$

3. Условие локального минимума (2.2) называют *необходимым условием первого порядка* (по порядку используемых в нем производных). Следующая теорема дает *необходимое условие локального минимума второго порядка.*

Т е о р е м а 2.2. Пусть $x^* = \operatorname{arglocmin}_{R^n} f(x)$. Если функция $f(x)$ дважды дифференцируема в точке x^*, то ее матрица Гессе неотрицательно определена в этой точке:

$$\langle \nabla^2 f(x^*) p, p \rangle \geq 0 \quad \forall\, p \in R^n \ . \tag{2.3}$$

Достаточное условие локального минимума второго порядка представляет собой усиление условия (2.3).

Т е о р е м а 2.3. Пусть функция $f(x)$ дважды дифференцируема в точке x^*. Если в этой точке выполняется условие (2.2) и матрица вторых производных функции $f(x)$ положительно определена, то x^* – точка строгого локального минимума функции $f(x)$.

Условия максимума функции $f(x)$ легко получить из условий минимума, если воспользоваться отмеченной выше связью между минимумами и максимумами. А именно, необходимыми условиями локального максимума функции $f(x)$ в точке x^* являются равенство нулю градиента и неположительная определенность матрицы вторых производных в этой точке. Таким образом, условие (2.2) является просто необходимым условием локального экстремума (как минимума, так и максимума). Если функция $f(x)$ вогнута и непрерывно дифференцируема на всем пространстве, то условие (2.2) является необходимым и достаточным условием глобального максимума функции $f(x)$. Достаточным условием локального максимума функции $f(x)$ в точке x^* является равенство нулю градиента и отрицательная определенность матрицы Гессе в этой точке.

Точки x^*, удовлетворяющие условию (2.2), называются *стационарными* точками. Поскольку точки экстремума дифференцируемой функции

$f(x)$ могут находиться только среди ее стационарных точек, то отсюда, в силу теорем 2.1–2.3, вытекает следующая схема исследования функций на экстремум.

1. Находятся частные производные функции $f(x)$ по всем аргументам и решается система алгебраических уравнений

$$\begin{aligned} &\frac{\partial f(x_1,\ldots,x_n)}{\partial x_1} = 0\,, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &\frac{\partial f(x_1,\ldots,x_n)}{\partial x_n} = 0\,. \end{aligned} \tag{2.4}$$

Решения системы (2.4) являются стационарными точками. (Если система (2.4) не имеет решений, то функция $f(x)$ не имеет стационарных точек и, стало быть, не имеет экстремумов.)

2. В стационарных точках исследуется на знакоопределенность матрица вторых производных. Точки, в которых матрица $\nabla^2 f(x)$ положительно определена, являются точками строгого локального минимума, а точки, в которых матрица $\nabla^2 f(x)$ отрицательно определена, – точками строгого локального максимума.

3. Анализируются стационарные точки, в которых матрица вторых производных не является строго знакоопределенной. Если матрица неотрицательно (неположительно) определена, то соответствующая точка подозрительна на локальный минимум (максимум). Иногда, непосредственно изучая поведение функции в окрестности подозрительной точки, удается выяснить, является ли она точкой минимума или максимума. Наконец, если в стационарной точке матрица вторых производных не является знакоопределенной, то такая точка заведомо не может быть точкой экстремума.

Если известно, что глобальный экстремум исследуемой функции существует, то точкой глобального минимума (максимума) является та стационарная точка, в которой функция принимает наименьшее (наибольшее) значение. Для установления факта существования глобального экстремума может оказаться полезным утверждение из упражнения 1.10: если функция $f(x)$ непрерывна на всем пространстве и $\lim\limits_{k\to\infty} f(x^k) = +\infty$

($\lim\limits_{k\to\infty} f(x^k) = -\infty$) для любой последовательности $\{x^k\}$ такой, что $\|x^k\| \to \infty$, то функция $f(x)$ достигает своего минимума (максимума) на R^n. Напомним также, что если матрица $\nabla^2 f(x)$ неотрицательно (неположительно) определена на всем пространстве, то функция $f(x)$ является выпуклой (вогнутой) и любая стационарная точка будет точкой минимума (максимума).

З а м е ч а н и е. В некоторых случаях наглядное представление о поведении функции, ее минимумах и максимумах можно получить из рассмотрения ее линий уровня. Напомним, что *линией уровня* функции $f(x)$ называется множество точек, в которых функция $f(x)$ сохраняет какое-либо постоянное значение $f(x) = k$. Например, линии уровня линейной функции $f(x) = c_1x_1 + \ldots + c_nx_n$ суть параллельные гиперплоскости (для $n = 3$ – плоскости, для $n = 2$ – прямые) $c_1x_1 + \ldots + c_nx_n = k$. Линии уровня функции $f(x) = \lambda x_1^2 + \mu x_2^2$, $\lambda, \mu > 0$, для $k > 0$ – концентрическое семейство эллипсов $x_1^2/a^2 + x_2^2/b^2 = 1$, где $a^2 = k/\lambda$, $b^2 = k/\mu$, для $k = 0$ – начало координат, а для $k < 0$ – пустое множество. Линии уровня функции $f(x) = \max\{0, x_1 + x_2\}$ для $k > 0$ – параллельные прямые $x_1 + x_2 = k$, для $k = 0$ – полуплоскость $x_1 + x_2 \leq 0$, для $k < 0$ – пустое множество. Линии уровня используются для визуализации функций во многих дисциплинах. Например, на географических картах с помощью линий уровня (изолиний) изображают рельеф местности (земная поверхность в этом случае есть функция от географических координат), в теории потребления это кривые безразличия – множество наборов продуктов, имеющих для потребителя одинаковую полезность, в химии – изобары и изотермы и т. д.

П р и м е р 2.1. Рассмотрим квадратичную функцию

$$f(x) = x_1^2 + 2x_2^2 - x_1x_2 + x_1 - x_2 \ .$$

Вычислим частные производные функции $f(x)$ и приравняем их нулю:

$$\frac{\partial f(x)}{\partial x_1} = 2x_1 - x_2 + 1 = 0\,,$$

$$\frac{\partial f(x)}{\partial x_2} = 4x_2 - x_1 - 1 = 0\,.$$

Решив эту систему, найдем стационарную точку $x^* = (-3/7, 1/7)$. Матрица вторых производных

$$\nabla^2 f(x) = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}$$

положительно определена на всем пространстве R^2. Следовательно, функция $f(x)$ – сильно выпуклая и точка x^* – единственная точка глобального минимума. Очевидно, что $\sup\limits_{R^2} f(x) = +\infty$. Линии уровня функции $f(x)$ представлены на рис. 2.1. Для наглядности на них проставлены соответствующие значения функции $f(x)$.

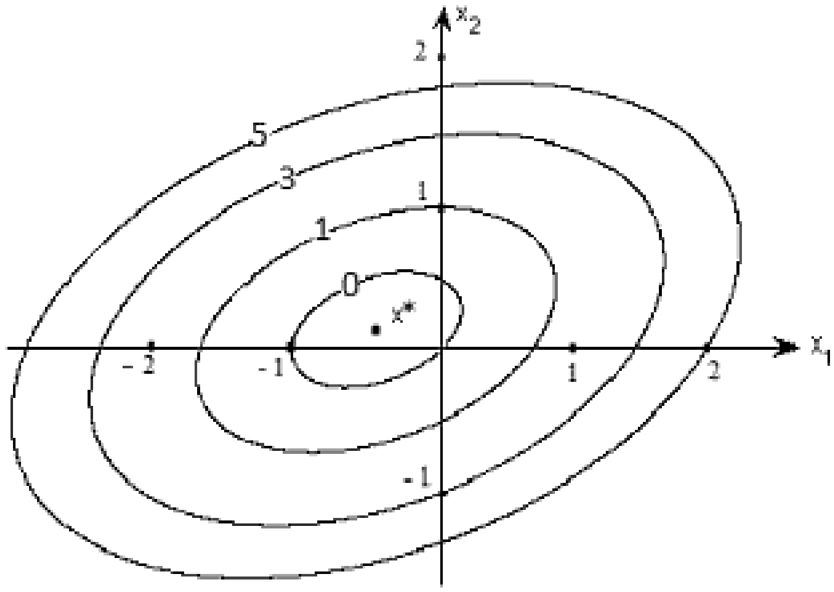

Рис. 2.1

П р и м е р 2.2. Исследуем на экстремум функцию

$$f(x) = x_1^2 + x_2^2 - 2x_1x_2 + x_1 - x_2 \ .$$

Вычисляя градиент функции $f(x)$ и приравнивая его нулю, получим систему линейных алгебраических уравнений

$$\begin{array}{rcrcrl} 2x_1 & - & 2x_2 & + & 1 & = 0\,, \\ -2x_1 & + & 2x_2 & - & 1 & = 0\,. \end{array}$$

Решение этой системы – любая точка прямой $2x_1 - 2x_2 = -1$. Матрица вторых производных

$$\nabla^2 f(x) = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}$$

положительно полуоределена на всем пространстве R^2. Следовательно, функция $f(x)$ – выпуклая и достигает своего минимума в любой точке прямой $2x_1 - 2x_2 = -1$. Очевидно также, что $\sup\limits_{R^2} f(x) = +\infty$. Линии уровня функции $f(x)$ представлены на рис. 2.2. Прямая X_* на рис. 2.2 – это множество точек минимума $2x_1^* - 2x_2^* = -1$.

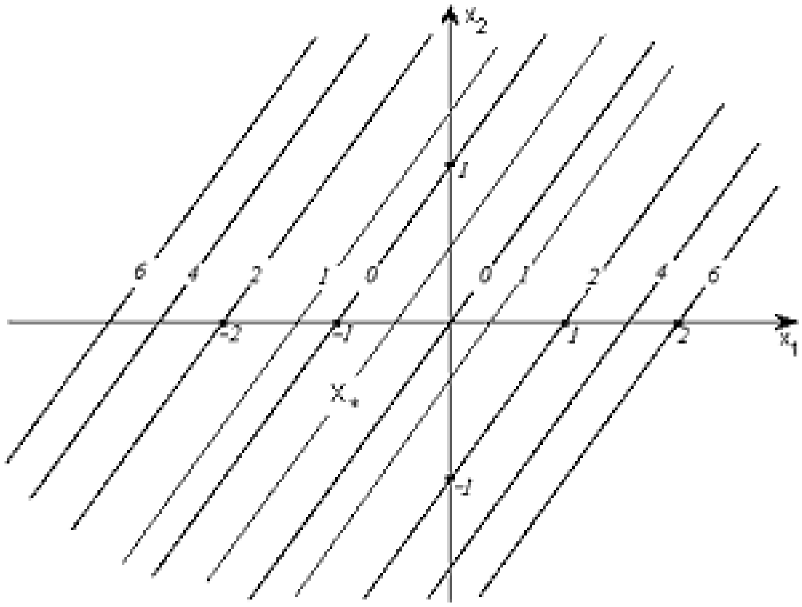

Рис. 2.2

П р и м е р 2.3. Рассмотрим функцию

$$f(x) = x_1^4 + x_2^4 - (x_1 + x_2)^2.$$

Стационарные точки функции $f(x) : x^1 = (0,0), x^2 = (-1,-1), x^3 = (1,1)$. Матрица вторых производных в точке x^1,

$$\nabla^2 f(x^1) = \begin{bmatrix} -2 & -2 \\ -2 & -2 \end{bmatrix},$$

является отрицательно полуопределенной. Следовательно, точка x^1 подозрительна на локальный максимум. Однако, поскольку $f(\lambda, -\lambda) = 2\lambda^4 > 0 = f(x^1)$ для любых сколь угодно малых $\lambda \neq 0$, точка x^1 не является точкой локального максимума. Матрицы

$$\nabla^2 f(x^2) = \nabla^2 f(x^3) = \begin{bmatrix} 10 & -2 \\ -2 & 10 \end{bmatrix}$$

положительно определены. Следовательно, точки x^2 и x^3 являются точками строгого локального минимума, причем $f(x^2) = f(x^3) = -2$. Далее, $\lim\limits_{k\to\infty} f(x^k) = +\infty$ для любой последовательности $\{x^k\}$ такой, что $\|x^k\| \to +\infty$. Следовательно, с одной стороны, $\sup\limits_{R^2} f(x) = +\infty$ и, с другой стороны, функция $f(x)$ достигает своего минимума на R^2. Последнее означает, что точки x^2 и x^3 – точки глобального минимума, поскольку других претендентов нет. Линии уровня функции $f(x)$ изображены на рис. 2.3.

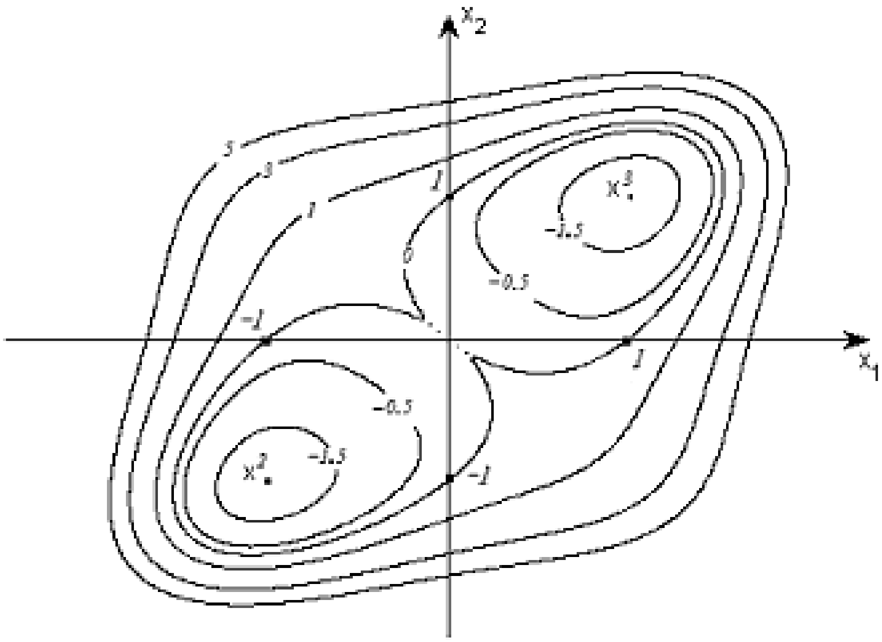

Рис. 2.3

П р и м е р 2.4. Этот пример интересен тем, что, хотя у рассматриваемой ниже функции имеется единственная точка локального экстремума, точек глобального экстремума не существует.

Рассмотрим функцию

$$f(x) = x_1^2 + 4x_2^3 - 2x_1x_2 \ .$$

Стационарные точки: $x^1 = (0,0)$, $x^2 = (1/6, 1/6)$. Матрица вторых производных в первой из них,

$$\nabla^2 f(x^1) = \begin{bmatrix} 2 & -2 \\ -2 & 0 \end{bmatrix} ,$$

является знаконеопределенной. Следовательно, точка x^1 не является точкой экстремума. Матрица вторых производных в точке x^2,

$$\nabla^2 f(x^2) = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}.$$

Эта матрица положительно определена и, следовательно, x^2 – точка строгого локального минимума. В то же время, например, $f(-2\lambda, -\lambda) = -4\lambda^3 \to -\infty$ при $\lambda \to +\infty$, $f(\lambda, 0) = \lambda^2 \to +\infty$ при $\lambda \to +\infty$. Линии уровня функции $f(x)$ представлены на рис. 2.4.

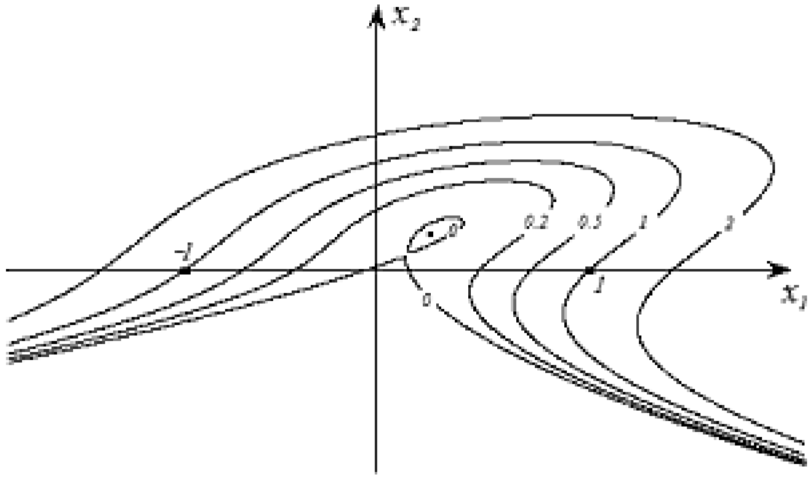

Рис. 2.4

П р и м е р 2.5. На рис. 2.5 изображены линии уровня функции $f(x) = x_1^2 - x_2^2$. Ее единственная стационарная точка, $x^* = (0, 0)$, не является точкой экстремума – матрица вторых производных $\nabla^2 f(x)$ – знаконеопределенная, $\sup\limits_{R^2} f(x) = +\infty$, $\inf\limits_{R^2} f(x) = -\infty$.

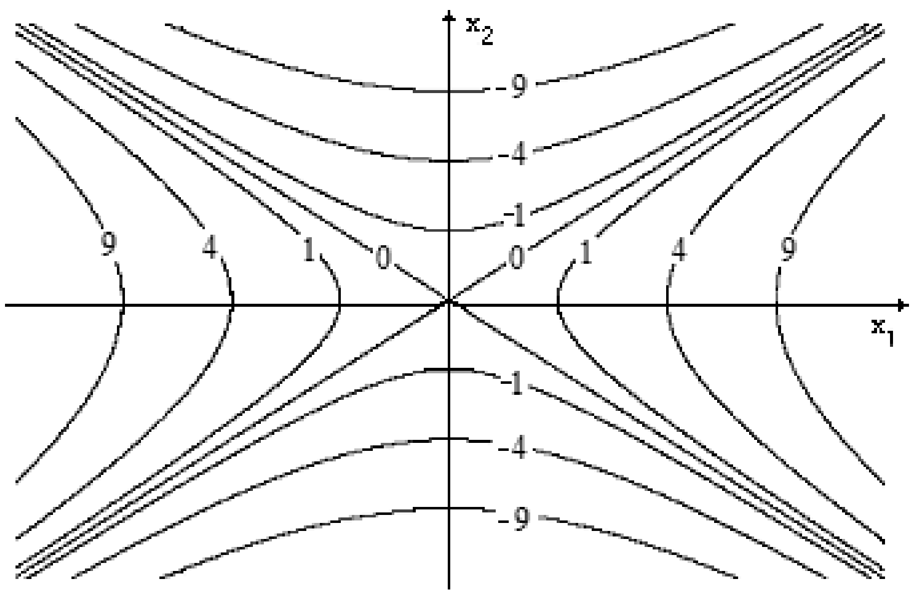

Рис. 2.5

График функции $f(x)$ (рис. 2.6) представляет собой поверхность, по форме подобную седлу, центр (седловина) которого находится в точке (в R^3) $(0,0,0) = (f(x^*), x_1^*, x_2^*)$. Точка x^* является *седловой точкой* (см. упражнение 2.14) функции $f(x)$. Заметим, что в этой точке функция $f(x)$ имеет минимум по переменной x_1 и максимум по переменной x_2 : $\min\limits_{x_1} f(x_1, 0) = \max\limits_{x_2} f(0, x_2) = f(x^*)$.

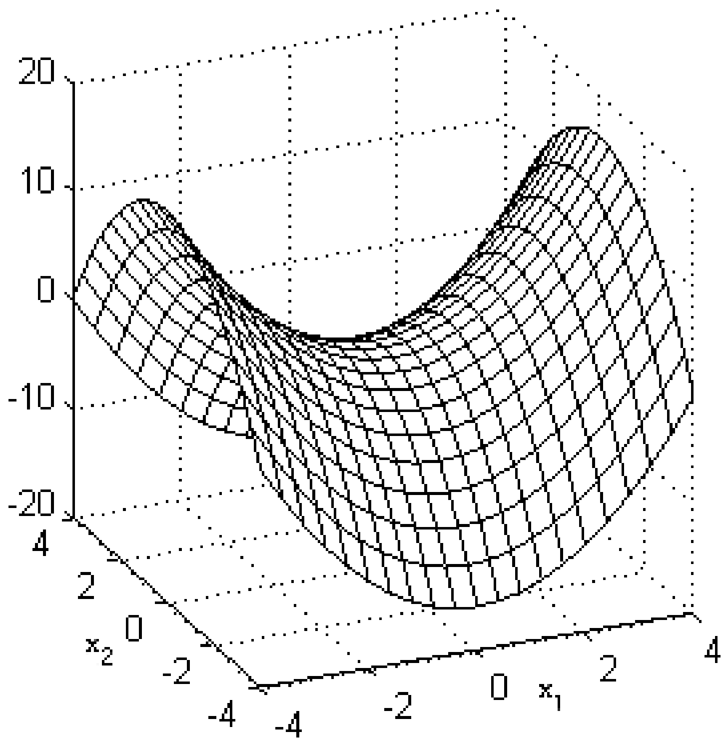

Рис. 2.6

Задачи и упражнения

2.1. Пусть функция $f(x)$ определена на множестве $X \subset R^n$. Доказать, что утверждения теорем 2.1–2.3 (при соответствующих предположениях о дифференцируемости) сохраняют силу в любой внутренней точке множества X.

2.2. Исследовать на экстремум следующие функции:

a) $\quad f(x) = x_1^2 + x_2^2 + x_3^2 - x_1x_2 + x_1 - 2x_3$;

b) $\quad f(x) = -x_1^2 - 4x_2^2 + 2x_1x_2 + x_1$;

c) $f(x) = -x_1^2 + x_1\sqrt{x_2} - x_2 + 6x_1 + 10$;

d) $f(x) = 2x_1^2 + 2x_2^2 + x_3^2 - 4x_1x_2 + x_1 + \cos x_3$;

e) $f(x) = x_1^3 + x_2^2 + 2x_3^2 - x_2x_3 - x_2$;

f) $f(x) = x_1^2 - 2x_2^2 - 2x_1x_2 + x_1$;

g) $f(x) = 2x_1^2 + x_2^3 + x_3^2 - x_1x_2 + 2x_1x_3 - x_2$;

h) $f(x) = x_1^4 + x_2^4 - 2(x_1 - x_2)^2$;

i) $f(x) = x_1x_2 + 20/x_1 + 50/x_2$;

j) $f(x) = e^{-x_1^2+x_2^2+2x_1x_2-x_2}$;

k) $f(x) = e^{-2x_1^2-5x_2^2+x_1x_2}$;

l) $f(x) = (4 - x_1)^2 + (x_1 - x_2^2)^2$;

m) $f(x) = x_1^2 - x_2^2 + 2e^{-x_1^2}$;

n) $f(x) = e^{-x_1^2-x_2^2}$;

o) $f(x) = e^{x_1^2+x_2^2+2x_1x_2+2}$;

p) $f(x) = (x_1^3 - 1)^4 + (x_2 - x_1)^2 - 2$;

q) $f(x) = x_1x_2^2(1 - x_1 - x_2)$;

r) $f(x) = (x_1 + x_2 - 1)e^{-x_1^2-x_1x_2+x_2^2}$;

s) $f(x) = x_1x_2^2x_3^2(1 - x_1 - 2x_2 - 3x_3)$;

t) $f(x) = 2x_1^{2/3} + x_2^{2/3} + 4x_3^{2/3}$;

u) $f(x) = (x_1 - 1)^3 + (x_2^3 - x_1)^2$;

v) $f(x) = x_1x_2\ln(x_1^2 + x_2^2)$;

w) $f(x) = ae^{-x_1} + be^{-x_2} + \ln(e^{x_1} + e^{x_2})\,, \quad a, b > 0$;

x) $f(x) = (x_1 + x_2)(x_1 - a)(x_2 - b)$;

y) $f(x) = x_1^2 - 2x_1x_2^2 + x_2^4 - x_2^5$;

z) $f(x) = x_1e^{x_1} - (1 + e^{x_1})\cos x_2$.

2.3. Исследовать на экстремум функцию $f(x_1, x_2)$, заданную неявно:

a) $x_1^2 + x_2^2 + f^2 - 2x_1 + 2x_2 - 4f - 10 = 0$;

b) $x_1^2 + x_2^2 + f^2 - x_1 f - x_2 f + 2x_1 + 2x_2 + 2f = 2$.

2.4. Найти все значения параметров a и b, при которых точка $(0, 1)$ является точкой экстремума для функции

$$f(x) = 2x_1^2 + a^3 e^{x_1} + bx_1x_2 + \frac{1}{2}b^3x_2^2 + ax_2 .$$

2.5. Найти все значения параметров a, b и c, при которых точка $(1, -4)$ является точкой глобального минимума для функции

$$f(x) = ax_1^2 + bx_1x_2 + x_2^2 + 6x_1 + cx_2 .$$

2.6. Доказать, что функция двух переменных

$$f(x_1, x_2) = x_1 e^{x_1} - (1 + e^{x_1}) \cos x_2$$

имеет бесконечно много точек локального минимума и ни одной точки локального максимума.

2.7. Доказать, что функция $f(x) = (x_1 - x_2^2)(2x_1 - x_2^2)$ достигает минимума в точке $x^* = (0, 0)$ вдоль каждой прямой, проходящей через x^*, но x^* не является точкой локального минимума этой функции.

2.8. В пространстве R^n найти точку, сумма квадратов расстояний от которой до k заданных точек $v^i, i = 1, 2, \ldots, k$ минимальна.

2.9. Исследовать на экстремум функцию $f(x) = \frac{1}{2}\|Ax - b\|^2$, где $A \in M_{m,n}$.

2.10. Исследовать на экстремум квадратичную функцию

$$f(x) = \frac{1}{2} < Gx, x > + < c, x > +d ,$$

где G – симметричная матрица порядка n.

2.11. Найти все решения задачи

$$f(x) = \|x\| - \langle c, x\rangle \to \min; \ \ x \in R^n.$$

2.12. Найти все точки экстремума, дать их характеристики и вычислить верхние и нижние грани функции

$$f(x) = \exp[\ (x_1^2 - 1)^2 + |x_2| + x_3^4\ ].$$

2.13. Построить линии уровня следующих функций $(R^2 \to R^1)$:

a) $f(x) = \|x - a\|$;

b) $f(x) = \|x - a\|^2$;

c) $f(x) = [1 + \|x - a\|^2]^{-1}$;

d) $f(x) = \exp(\|x - a\|)$;

e) $f(x) = \ln(1 + \|x - a\|)$;

f) $f(x) = w_1(x_1 - a_1)^2 + w_2(x_2 - a_2)^2, \ \ w_1,\, w_2 > 0$;

g) $f(x) = w_1|x_1 - a_1| + w_2|x_2 - a_2|, \ \ w_1,\, w_2 > 0$;

h) $f(x) = \max\{w_1|x_1 - a_1|, w_2|x_2 - a_2|\}, \ \ w_1,\, w_2 > 0$;

i) $f(x) = \exp(c_1x_1 + c_2x_2)$;

j) $f(x) = x_1x_2$;

k) $f(x) = x_1^2 + x_2^2 - 2x_1x_2$;

l) $f(x) = x_1^4 + x_2^2 - 2x_1^2x_2$;

m) $f(x) = \max\{x_1, x_2\}$;

n) $f(x) = \max\{0, -\|x\|^2\}$;

o) $f(x) = \min\{0, x_1\}$;

p) $f(x) = x_1^2 + 2x_2^2 - 4x_1 + 2x_2$;

q) $f(x) = x_1^2 - 2x_2^2 - 4x_1 + 2x_2$.

2.14. Точка (x^*, y^*) называется *седловой точкой* функции $f(x, y)$, если $f(x^*, y^*) = \min\limits_x f(x, y^*) = \max\limits_y f(x^*, y)$, или $f(x^*, y^*) = \max\limits_x f(x, y^*) = \min\limits_y f(x^*, y)$. Проверить, является ли точка (x^0, y^0) седловой точкой (или точкой экстремума) функции $f(x, y)$:

$$f(x, y) = -1 + 4(e^x - x) - 5x\sin y + 6y^2, \ \ (x^0, y^0) = (0, 0);$$

$$f(x, y) = (x^2 - 2x)\cos y, \ \ (x^0, y^0) = (1, \pi).$$

§ 3. Классическая задача на условный экстремум

Классической задачей на условный экстремум называется задача нахождения минимума (или максимума) функции $f(x) : R^n \to R^1$ на множестве решений системы алгебраических уравнений

$$g_1(x_1, \dots, x_n) = 0 \ ,$$

$$g_2(x_1, \dots, x_n) = 0 \ ,$$

$$\dots\dots\dots\dots\dots\dots$$

$$g_m(x_1, \dots, x_n) = 0 \ .$$

Коротко эту задачу будем записывать следующим образом:

$$f(x) \to \min\,(\max); \quad g_j(x) = 0, \ \ j = 1, \dots, m \tag{3.1}$$

или

$$f(x) \to \min\,(\max); \quad x \in X \ ,$$

где $X = \{x \in R^n : g_j(x) = 0, \ \ j = 1, \dots, m\}$. Множество X называется *допустимым множеством*, а точки $x \in X$ – *допустимыми точками*.

Задачу на условный экстремум называют также задачей (минимизации или максимизации) с ограничениями-равенствами, а уравнения $g_j(x) = 0$, задающие ограничения на переменные, – *уравнениями связей*, или, просто, *связями* задачи.

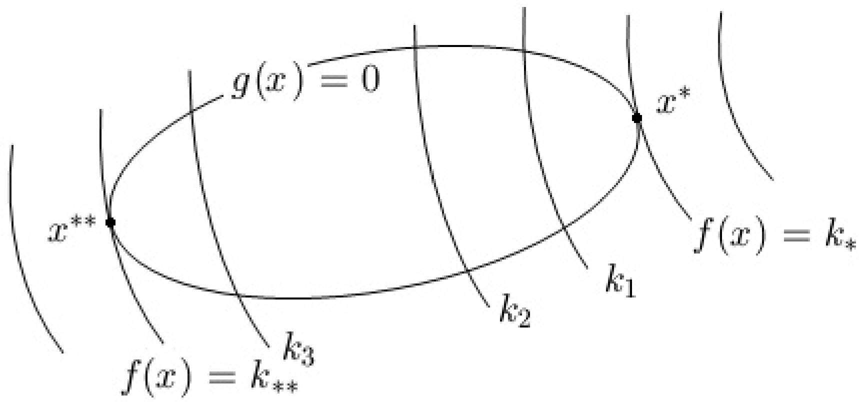

Рис. 3.1. $k_* < k_1 < k_2 < k_3 < k_{**}$

На рис. 3.1 приведена графическая иллюстрация задачи (3.1): точка минимума функции $f(x)$ на кривой $g(x) = 0$ – это точка x^*, принадлежащая линии уровня $f(x) = k_*$, где k_* – наименьшее среди всех k, при которых кривые $f(x) = k$ и $g(x) = 0$ имеют общие точки. Аналогично, точка максимума функции $f(x)$ на кривой $g(x) = 0$ – это точка x^{**}, принадлежащая линии уровня $f(x) = k_{**}$, где k_{**} – наибольшее среди всех k, при которых кривые $f(x) = k$ и $g(x) = 0$ имеют общие точки.

В некоторых случаях для решения задачи (3.1) удается применить *метод исключения переменных*. Допустим, что $m < n$ и из ограничений задачи можно выразить m каких-либо компонент вектора x через остальные $n - m$ компонент. Например, пусть, для определенности,

$$x_1 = \varphi_1(x_{m+1}, x_{m+2}, \dots, x_n),$$

$$x_2 = \varphi_2(x_{m+1}, x_{m+2}, \dots, x_n),$$

$$\dots\dots\dots\dots\dots\dots\dots\dots$$

$$x_m = \varphi_m(x_{m+1}, x_{m+2}, \dots, x_n).$$

Подставляя эти выражения в целевую функцию, приходим к задаче на безусловный экстремум функции $n - m$ переменных $\psi(x_{m+1}, \dots, x_n) = f(\varphi_1(x_{m+1}, \dots, x_n), \dots, \varphi_m(x_{m+1}, \dots, x_n), x_{m+1}, \dots, x_n)$.

П р и м е р 3.1. Рассмотрим задачу

$$f(x) = (x_1 - 1)^2 + 2x_2^3 + x_3^3 + (x_4 + 1)^2 \to \text{extr}; \tag{3.2}$$

$$x_1 + 2x_2 + x_3 = 1; \tag{3.3}$$

$$x_2 - 2x_3 + x_4 = 1. \tag{3.4}$$

Запись «$\to$ extr» означает, что требуется исследовать функцию на минимум и на максимум. Из уравнений связей имеем

$$x_1 = 1 - 2x_2 - x_3, \quad x_4 = 1 - x_2 + 2x_3. \tag{3.5}$$

Подставляя эти выражения в целевую функцию, приходим к задаче на безусловный экстремум функции двух переменных:

$$\psi(x_2, x_3) = (2x_2 + x_3)^2 + 2x_2^3 + x_3^3 + (2 - x_2 + 2x_3)^2 \to \text{extr}. \tag{3.6}$$

Частные производные функции $\psi(x_2, x_3)$ имеют вид:

$$\frac{\partial\psi(x_2, x_3)}{\partial x_2} = 6x_2^2 + 10x_2 - 4; \quad \frac{\partial\psi(x_2, x_3)}{\partial x_3} = 3x_3^2 + 10x_3 + 8.$$

Приравнивая их нулю и решая полученные уравнения, находим четыре стационарные точки в задаче (3.6): $(1/3, -2)$, $(1/3, -4/3)$, $(-2, -2)$, $(-2, -4/3)$. Матрица вторых производных

$$\nabla^2\psi(x_2, x_3) = \begin{bmatrix} 12x_2 + 10 & 0 \\ 0 & 6x_3 + 10 \end{bmatrix}$$

положительно определена в точке$(1/3, -4/3)$, отрицательно определена в точке $(-2, -2)$ и является знаконеопределенной в точках $(1/3, -2)$ и $(-2, -4/3)$. Следовательно, точка $(1/3, -4/3)$ является точкой строгого локального минимума, а точка $(-2, -2)$ – точкой строгого локального максимума функции $\psi(x_2, x_3)$. Точки $(1/3, -2)$ и $(-2, -4/3)$ не являются точками экстремума функции $\psi(x_2, x_3)$. Подставляя координаты точек экстремума в (3.5), находим две точки:

$$x^1 = (5/3, 1/3, -4/3, -2), \quad x^2 = (7, -2, -2, -1).$$

Первая из них является точкой строгого локального минимума, вторая – точкой строгого локального максимума в задаче (3.2)–(3.4). Нетрудно видеть, что $\sup\limits_X f(x) = \sup\limits_{R^2} \psi(x_2, x_3) = +\infty$, $\inf\limits_X f(x) = \inf\limits_{R^2} \psi(x_2, x_3) = -\infty$. ■

Рассмотрим классические условия экстремума в задаче (3.1). Для определенности будем считать, что задача (3.1) – задача на минимум.

О п р е д е л е н и е 3.1. Пусть функции $g_1(x), \ldots, g_m(x)$ дифференцируемы в точке $x \in X$. Говорят, что в точке x выполнено *условие регулярности*, если градиенты $\nabla g_1(x), \ldots, \nabla g_m(x)$ линейно независимы.

Т е о р е м а 3.1 (правило множителей Лагранжа). Пусть x^* – точка локального минимума в задаче (3.1). Если функции $f(x), g_1(x), \ldots, g_m(x)$ непрерывно дифференцируемы в некоторой окрестности точки x^*, а в самой точке x^* выполняется условие регулярности, то существуют числа $\lambda_1^*, \ldots, \lambda_m^*$, такие, что

$$\nabla f(x^*) + \sum_{j=1}^{m} \lambda_j^* \nabla g_j(x^*) = 0. \tag{3.7}$$

Числа $\lambda_1^*, \ldots, \lambda_m^*$ в (3.7) называются *множителями Лагранжа*, отвечающими точке x^*. Условие регулярности обеспечивает не только существование, но и единственность множителей Лагранжа. В самом деле, если предположить, что существуют числа $\mu_1^*, \ldots, \mu_m^*$ такие, что $\mu_j^* \neq \lambda_j^*$ хотя бы для одного номера j и

$$\nabla f(x^*) + \sum_{j=1}^{m} \mu_j^* \nabla g_j(x^*) = 0, \tag{3.8}$$

то, вычитая (3.7) из (3.8), получим, что

$$\sum_{j=1}^{m} (\mu_j^* - \lambda_j^*) \nabla g_j(x^*) = 0.$$

Но это противоречит условию регулярности.

Необходимо отметить, что условие регулярности является наиболее общим, но не единственным условием, при выполнении которого существуют множители Лагранжа. Например, если $\nabla f(x^*) = 0$, то условие (3.7) выполняется тривиально: $\lambda_1^* = \ldots = \lambda_m^* = 0$. Если при этом матрица $\nabla^2 f(x^*)$ положительно определена, то x^* – точка безусловного, а, значит, и условного, локального минимума функции $f(x)$. Другой, нетривиальный, пример – это задачи с линейными связями

$$g_j(x) = \langle a^j, x \rangle - b_j = 0, \ j = 1, \ldots, m. \tag{3.9}$$

Если условия (3.9) линейно независимы, т. е.

$$\mathrm{rank}[\nabla g_1(x) \ldots g_m(x)] = \mathrm{rank}[a^1 \ldots a^m] = m,$$

то справедлива теорема 3.1. Если же среди условий (3.9) имеются линейно зависимые условия (но система (3.9) совместна), то эти линейно зависимые условия можно исключить из системы без ущерба для множества ее решений. Не умаляя общности, можно считать, что после исключения в системе остались первые $r < m$ уравнений. Тогда, если x^* – точка локального минимума, то, в силу теоремы 3.1, существуют числа $\lambda_1^*, \ldots, \lambda_r^*$, такие, что

$$\nabla f(x^*) + \sum_{j=1}^{r} \lambda_j^* \nabla g_j(x^*) = 0.$$

Полагая $\lambda_{r+1}^* = \ldots = \lambda_m^* = 0$, отсюда получим (3.7). Очевидно, что в этом случае множители Лаграгжа могут определяться неоднозначно.

Заметим также, что если ограничения задачи имеют вид (3.9), то множество $X = \{x \in R^n : g_j(x) = 0, j = 1, \dots, m\}$ является выпуклым. Если, кроме того, функция $f(x)$ выпукла на множестве X, то задачу (3.1) будем называть *линейно-выпуклой задачей на условный экстремум.* В линейно-выпуклой задаче всякая точка локального минимума является точкой глобального минимума, а правило множителей Лагранжа – не только необходимым, но и достаточным условием экстремума.

Введем *функцию Лагранжа*

$$L(x, \lambda) = f(x) + \langle \lambda, g(x) \rangle = f(x) + \sum_{j=1}^{m} \lambda_j g_j(x),$$

где $\lambda = (\lambda_1, \dots, \lambda_m)$, $g(x) = (g_1(x), \dots, g_m(x))$. Тогда условие (3.7) можно записать в виде:

$$L_x(x^*, \lambda^*) = 0, \tag{3.10}$$

где $L_x(x, \lambda)$ – вектор, составленный из частных производных функции Лагранжа по переменным $x_1, \dots, x_n$:

$$L_x(x, \lambda) = \nabla f(x) + \sum_{j=1}^{m} \lambda_j \nabla g_j(x).$$

Матрицу вторых частных производных функции Лагранжа по переменным $x_1, \dots, x_n$ обозначим через $L_{xx}(x, \lambda)$:

$$L_{xx}(x, \lambda) = \nabla^2 f(x) + \sum_{j=1}^{m} \lambda_j \nabla^2 g_j(x).$$

Т е о р е м а 3.2 (необходимое условие второго порядка). Пусть x^* – точка локального минимума в задаче (3.1). Если функции $f(x), g_1(x), \dots,$ $g_m(x)$ дважды непрерывно дифференцируемы в некоторой окрестности точки x^*, а в самой точке x^* выполняются условия регулярности, то

$$\langle L_{xx}(x^*, \lambda^*)p, p \rangle \geq 0 \tag{3.11}$$

для всех $p \neq 0$ таких, что

$$\langle \nabla g_j(x^*), p \rangle = 0, \ j = 1, \dots, m. \tag{3.12}$$

Здесь λ^* – вектор множителей Лагранжа.

Т е о р е м а 3.3 (достаточное условие второго порядка). Пусть функции $f(x), g_1(x), \ldots, g_m(x)$ дважды непрерывно дифференцируемы в некоторой окрестности точки $x^* \in X$ и существуют числа $\lambda_1^*, \ldots, \lambda_m^*$ такие, что выполняются условие (3.10) и условие

$$\langle L_{xx}(x^*, \lambda^*)p, p \rangle > 0 \qquad (3.13)$$

для всех $p \neq 0$, удовлетворяющих (3.12). Тогда x^* – точка строгого локального минимума в задаче (3.1).

Обратим внимание, что теорема 3.3 не требует выполнения условия регулярности в точке x^* – достаточно лишь выполнения условий (3.10) и (3.13) для некоторых $\lambda_1^*, \ldots, \lambda_m^*$.

Условия локального максимума функции $f(x)$ на множестве X немедленно следуют из условий (3.10), (3.11) и (3.13), если задачу на максимум записать в виде

$$-f(x) \to \min; \; -g(x) = 0.$$

А именно, необходимое условие первого порядка имеет вид (3.10) и, таким образом, условие (3.10) является просто необходимым условием экстремума (как минимума, так и максимума), а в условиях второго порядка нужно просто поменять знаки неравенств на противоположные.

Допустимая точка x^*, удовлетворяющая при некотором наборе множителей $\lambda_1^*, \ldots, \lambda_m^*$ условию (3.10), называется *стационарной точкой* в задаче (3.1). Стационарные точки являются точками, подозрительными на экстремум. Для их нахождения нужно составить функцию Лагранжа $L(x, \lambda)$, вычислить ее частные производные по $x_1, \ldots, x_n$ и решить систему $n + m$ уравнений

$$L_x(x, \lambda) = 0 \qquad (n \text{ уравнений}), \qquad (3.14)$$

$$g(x) = 0 \qquad (m \text{ уравнений}) \qquad (3.15)$$

с $n+m$ неизвестными $x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m$. Уравнения (3.14) – это необходимое условие экстремума (3.10), уравнения (3.15) – ограничения задачи («условия допустимости» точки x). Найденные из решения системы (3.14), (3.15) стационарные точки исследуются на экстремум с помощью условий второго порядка. Описанная процедура называется *методом неопределенных множителей Лагранжа.*

З а м е ч а н и е. Поскольку $\partial L(x,\lambda)/\partial\lambda_j = g_j(x)$, то градиент функции Лагранжа $\nabla L(x,\lambda) = (L_x(x,\lambda), g(x))$. Следовательно, система (3.14), (3.15) представляет собой условие стационарности функции Лагранжа:

$$\nabla L(x,\lambda) = 0. \tag{3.16}$$

Решение этой системы, точка (x^*,λ^*), является стационарной точкой функции Лагранжа. Условия (3.14), (3.15), или, что – то же самое, условие (3.16) называются *условиями стационарности.*

П р и м е р 3.2. Рассмотрим задачу из примера 3.1:

$$f(x) = (x_1-1)^2 + 2x_2^3 + x_3^3 + (x_4+1)^2 \to \text{extr};$$

$$x_1 + 2x_2 + x_3 = 1,$$

$$x_2 - 2x_3 + x_4 = 1.$$

Ограничения задачи – линейные, линейно-независимые. Следовательно, во всех допустимых точках выполняется условие регулярности.

Функция Лагранжа

$$L(x,\lambda) = (x_1-1)^2 + 2x_2^3 + x_3^3 + (x_4+1)^2 + \\ + \lambda_1(x_1 + 2x_2 + x_3 - 1) + \lambda_2(x_2 - 2x_3 + x_4 - 1).$$

Условия стационарности:

$$\begin{aligned} L_{x_1}(x,\lambda) &= 2x_1 - 2 + \lambda_1 = 0,\\ L_{x_2}(x,\lambda) &= 6x_2^2 + 2\lambda_1 + \lambda_2 = 0,\\ L_{x_3}(x,\lambda) &= 3x_3^2 + \lambda_1 - 2\lambda_2 = 0,\\ L_{x_4}(x,\lambda) &= 2x_4 + 2 + \lambda_2 = 0,\\ g_1(x) &= x_1 + 2x_2 + x_3 = 1,\\ g_2(x) &= x_2 - 2x_3 + x_4 = 1. \end{aligned}$$

Данная система имеет четыре решения:

$$\begin{aligned} (x^1,\lambda^1) &= (7/3, 1/3, -2, -10/3, -8/3, 14/3),\\ (x^2,\lambda^2) &= (5/3, 1/3, -4/3, -2, -4/3, 2),\\ (x^3,\lambda^3) &= (7, -2, -2, -1, -12, 0),\\ (x^4,\lambda^4) &= (19/3, -2, -4/3, 1/3, -32/3, -8/3). \end{aligned}$$

Следовательно, в исходной задаче имеется четыре стационарные точки:

$$x^1 = (7/3, 1/3, -2, -10/3), \quad x^2 = (5/3, 1/3, -4/3, -2),$$
$$x^3 = (7, -2, -2, -1), \quad x^4 = (19/3, -2, -4/3, 1/3).$$

Матрица вторых производных

$$L_{xx}(x, \lambda) = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 12x_2 & 0 & 0 \\ 0 & 0 & 6x_3 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} .$$

Без учета условий (3.12) эта матрица не является знакоопределенной ни в одной из стационарных точек. Условия (3.12) (одни и те же для всех точек) имеют вид:

$$\langle \nabla g_1(x), p \rangle = p_1 + 2p_2 + p_3 = 0,$$

$$\langle \nabla g_2(x), p \rangle = p_2 - 2p_3 + p_4 = 0.$$

Отсюда

$$p_1 = -2p_2 - p_3, \quad p_4 = -p_2 + 2p_3. \tag{3.17}$$

С учетом этого, квадратичная форма

$$\langle L_{xx}(x^1, \lambda^1)p, p \rangle = 2p_1^2 + 4p_2^2 - 12p_3^2 + 2p_4^2 =$$

$$= 2(2p_2 + p_3)^2 + 4p_2^2 - 12p_3^2 + 2(p_2 - 2p_3)^2 = 14p_2^2 - 2p_3^2. \tag{3.18}$$

Аналогично,

$$\langle L_{xx}(x^2, \lambda^2)p, p \rangle = 2p_1^2 + 4p_2^2 - 8p_3^2 + 2p_4^2 = 14p_2^2 + 2p_3^2, \tag{3.19}$$

$$\langle L_{xx}(x^3, \lambda^3)p, p \rangle = 2p_1^2 - 24p_2^2 - 12p_3^2 + 2p_4^2 = -14p_2^2 - 2p_3^2, \tag{3.20}$$

$$\langle L_{xx}(x^4, \lambda^4)p, p \rangle = 2p_1^2 - 24p_2^2 - 8p_3^2 + 2p_4^2 = -14p_2^2 + 2p_3^2. \tag{3.21}$$

Так как $p \neq 0$, то, в силу (3.17), компоненты p_2 и p_3 вектора p не могут одновременно равняться нулю. Следовательно, квадратичная форма (3.19) определена положительно, квадратичная форма (3.20) определена отрицательно, а квадратичные формы (3.18) и (3.21) являются знаконеопределенными для всех p, удовлетворяющих условиям (3.12). Отсюда, в свою очередь, следует, что точка x^2 является точкой строгого локального минимума, точка x^3 – точкой строгого локального максимума, а стационарные точки x^1 и x^4 не являются точками экстремумов в рассматриваемой задаче.

П р и м е р 3.3. Рассмотрим задачу

$$f(x) = (x_1 - 1)^2 + (x_2 - 1)^2 + x_3^2 \to \min; \tag{3.22}$$

$$x_1 + x_2 - x_3 = 1, \tag{3.23}$$

$$2x_1 - x_2 + 2x_3 = 3, \tag{3.24}$$

$$3x_1 + x_3 = 4. \tag{3.25}$$

Ограничения задачи – линейные, функция $f(x)$ – сильно выпуклая. Следовательно, решение задачи существует и единственно, а условия стационарности являются необходимыми и достаточными условиями экстремума.

Функция Лагранжа задачи (3.22)–(3.25) имеет вид

$$L(x, \lambda) = (x_1 - 1)^2 + (x_2 - 1)^2 + x_3^2 + \lambda_1(x_1 + x_2 - x_3 - 1) +$$
$$+ \lambda_2(2x_1 - x_2 + 2x_3 - 3) + \lambda_3(3x_1 + x_3 - 4).$$

Условия стационарности функции Лагранжа:

$$L_{x_1}(x, \lambda) = 2(x_1 - 1) + \lambda_1 + 2\lambda_2 + 3\lambda_3 = 0, \tag{3.26}$$

$$L_{x_2}(x, \lambda) = 2(x_2 - 1) + \lambda_1 - \lambda_2 = 0, \tag{3.27}$$

$$L_{x_3}(x, \lambda) = 2x_3 - \lambda_1 + 2\lambda_2 + \lambda_3 = 0, \tag{3.28}$$

$$g_1(x) = x_1 + x_2 - x_3 - 1 = 0, \tag{3.29}$$

$$g_2(x) = 2x_1 - x_2 + 2x_3 - 3 = 0, \tag{3.30}$$

$$g_3(x) = 3x_1 + x_3 - 4 = 0. \tag{3.31}$$

Решение системы (3.26)–(3.31):

$$x^* = (29/26, 14/26, 17/26) \quad \text{(решение исходной задачи)},$$

$$\lambda^* = (-\alpha + 7/13, -\alpha - 5/13, \alpha), \quad -\infty < \alpha < +\infty.$$

Заметим, что ограничения (3.23)–(3.25) – линейно-зависимые. Ранг системы ограничений равен двум. ■

В гладких задачах на безусловный экстремум решение задачи находится среди ее стационарных точек. В задаче на условный экстремум это, к сожалению, не всегда так. Например, пусть множество $X \subset R^2$

определяется условием $g(x) = x_1^3 - x_2^2 = 0$, функция $f(x) = x_1$. Очевидно, что минимум функции $f(x)$ на множестве X достигается в точке $x^* = (0, 0)$. Однако

$$\nabla f(x^*) + \lambda \nabla g(x^*) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 0 \end{bmatrix} \neq 0$$

ни при каком λ. Причина – в точке x^* не выполняется условие регулярности: градиент ограничения $\nabla g(x^*) = 0$.

Точки $x \in X$, в которых не выполняется условие регулярности, называются *анормальными точками* множества X. Множители Лагранжа в анормальной точке могут существовать, как, скажем, в примере 3.3, где все допустимые точки – анормальные, а могут не существовать, как в только что рассмотренном примере. Для выяснения того, существует ли в анормальной точке какой-либо максимум или минимум, нужно провести дополнительное исследование поведения функции в окрестности этой точки.

Анормальные точки, если они существуют, являются, по определению, решениями системы алгебраических уравнений

$$\sum_{j=1}^{n} \lambda_j \nabla g_j(x) = 0, \tag{3.32}$$

$$g(x) = 0 \tag{3.33}$$

при дополнительном условии $\lambda \neq 0$. Очевидно, что решение системы (3.32), (3.33) неединственно – если (λ^*, x^*) – решение системы (3.32), (3.33), то (μ^*, x^*), где $\mu^* = \alpha\lambda^*$, также является решением этой системы при любом $\alpha \neq 0$.

Формально разыскание анормальных нестационарных точек можно «присоединить» к методу неопределенных множителей Лагранжа, введя так называемую *обобщенную функцию Лагранжа*

$$L(x, \bar{\lambda}) = \lambda_0 f(x) + \sum_{j=1}^{m} \lambda_j g_j(x),$$

где $\bar{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, а множитель λ_0 принимает два значения – 0 или 1. При $\lambda_0 = 1$ обобщенная функция Лагранжа совпадает с классичеческой: $L(x, \bar{\lambda}) = L(x, \lambda)$. Так как

$$L_x(x, \bar{\lambda}) = \lambda_0 \nabla f(x) + \sum_{j=1}^{m} \lambda_j \nabla g_j(x),$$

то условия (3.14), (3.15) и (3.33), (3.34) вместе можно записать в виде

$$L_x(x, \bar{\lambda}) = 0, \tag{3.35}$$

$$g(x) = 0, \tag{3.36}$$

где $\lambda_0 = 1$ или 0, соответственно. Для нахождения всех точек, подозрительных на экстремум, нужно решить систему (3.35), (3.36) сначала с $\lambda_0 = 1$, а затем с $\lambda_0 = 0$ (или наоборот). Разумеется, если априори известно, что анормальных нестационарных точек не существует (например, в случае линейных связей), то решение системы (3.35), (3.36) с $\lambda_0 = 0$ становится излишним.

П р и м е р 3.4. Исследуем на экстремум функцию $f(x) = x_2$ на множестве $X = \{x \in R^2 : x_1^3 + x_2^3 - 3ax_1x_2 = 0,\ a > 0\}$.

Составим обобщенную функцию Лагранжа

$$L(x, \bar{\lambda}) = \lambda_0 x_2 + \lambda(x_1^3 + x_2^3 - 3ax_1x_2). \tag{3.37}$$

Вычислим частные производные функции (3.37) по x_1, x_2, приравняем их нулю и добавим к полученной системе уравнение связи:

$$L_{x_1}(x, \bar{\lambda}) = 3\lambda(x_1^2 - ax_2) = 0, \tag{3.38}$$

$$L_{x_2}(x, \bar{\lambda}) = \lambda_0 + 3\lambda(x_2^2 - ax_1) = 0, \tag{3.39}$$

$$g_(x) \quad = x_1^3 + x_2^3 - 3ax_1x_2 = 0. \tag{3.40}$$

Положим $\lambda_0 = 1$. Тогда решением системы (3.38)–(3.40) будут стационарная точка $x^* = (a\sqrt[3]{2}, a\sqrt[3]{4})$ и множитель Лагранжа $\lambda^* = -(3a^2\sqrt[3]{2})^{-1}$. Матрица вторых производных функции Лагранжа в точке (x^*, λ^*)

$$L_{xx}(x^*, \lambda^*) = \begin{bmatrix} 6\lambda^* x_1^* & -3a\lambda^* \\ -3a\lambda^* & 6\lambda^* x_2^* \end{bmatrix} = 3a\lambda^* \begin{bmatrix} 2\sqrt[3]{2} & -1 \\ -1 & 2\sqrt[3]{4} \end{bmatrix}$$

является отрицательно определенной. Следовательно, x^* – точка строгого локального максимума, $f(x^*) = a\sqrt[3]{4}$.

Положим теперь $\lambda_0 = 0$. Тогда решение системы (3.38)–(3.40) – точка $x = (0,0)$ и произвольное $\lambda \neq 0$. Точка $(0,0)$ – анормальная. Покажем, что она не является точкой экстремума. Для этого положим $x_2 = tx_1$ и подставим в уравнение (3.40). В результате получим параметрическое представление кривой $x_1^3 + x_2^3 - 3ax_1x_2 = 0$:

$$x_1 = \frac{3at}{1+t^3}, \quad x_2 = \frac{3at^2}{1+t^3}, \ -\infty < t < +\infty, \ t \neq -1. \tag{3.41}$$

Отсюда ясно, что точка $(0,0)$ не является точкой экстремума – в любой сколь угодно малой окрестности этой точки функция $f(x) = x_2$ принимает (при $t \to \pm\infty$) значения разных знаков. Из (3.41) следует также, что $x_2 \to +\infty$ при $t \to -1+0$ и $x_2 \to -\infty$ при $t \to -1-0$. То есть функция $f(x)$ не ограничена на множестве X. Кривая (3.41) изображена на рис. 3.2. Она называется *декартов лист*. Точка $(0,0)$ – особая точка этой кривой.

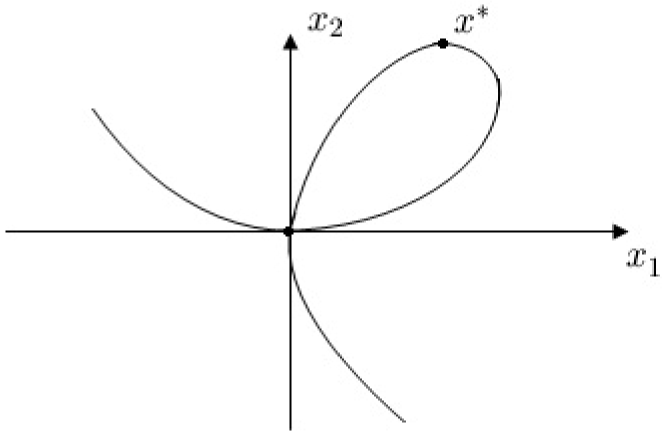

Рис. 3.2

В заключение заметим, что задачи, в которых среди ограничений на переменные имеются ограничения вида $g_j(x) = g_j(x_1, \ldots, x_n) \leq 0$ (*ограничения типа неравенств*) могут быть сведены к задачам на условный экстремум введением дополнительных переменных: $g_j(x) + x_{n+j}^2 = 0$. Рассмотрим следующий простой

П р и м е р 3.5. Найти минимум функции

$$f(x) = 3x_2^2 - 11x_1 - 3x_2 - x_3$$

на множестве

$$X = \{x \in R^3 : x_1 - 7x_2 + 3x_3 \leq -7, \quad 5x_1 + 2x_2 - x_3 \leq 2\}.$$

Прибавляя к левой части первого и второго ограничений x_4^2 и x_5^2, соответственно, приходим к задаче

$$f(x) = 3x_2^2 - 11x_1 - 3x_2 - x_3 \to \min; \tag{3.42}$$

$$x_1 - 7x_2 + 3x_3 + x_4^2 \qquad = -7, \tag{3.43}$$

$$5x_1 + 2x_2 - x_3 + \qquad x_5^2 = \quad 2. \tag{3.44}$$

Задачу (3.42)–(3.44) можно решить, например, методом исключения. Из (3.43) и (3.44) имеем

$$x_1 = \frac{1}{16}(x_2 - x_4^2 - 3x_5^2 - 1), \tag{3.45}$$

$$x_3 = \frac{1}{16}(37x_2 - 5x_4^2 + x_5^2 - 37). \tag{3.46}$$

Подставляя эти выражения в целевую функцию, получим задачу на безусловный минимум

$$3x_2^2 - 6x_2 + x_4^2 + 2x_5^2 + 3 \to \min .$$

Решение этой задачи очевидно: $x_2^* = 1, x_4^* = x_5^* = 0$. Отсюда, в силу (3.45) и (3.46), получим решение исходной задачи $x^* = (x_1^*, x_2^*, x_3^*) = (0, 1, 0)$ (точка глобального минимума).

Задачи и упражнения

3.1. Покажите, что в задаче

$$f(x) \to \min; \quad Ax = b, \quad A \in M_{m,n}$$

необходимые условия минимума имеют вид:

$$\nabla f(x^*) + A^T\lambda^* = 0, \quad \langle \nabla^2 f(x^*)p, p\rangle \geq 0 \ \ \forall p : Ap = 0.$$

3.2. Доказать, что правило множителей Лагранжа является необходимым и достаточным условием экстремума в линейно-выпуклой задаче на условный минимум.

3.3. Следующие задачи решить методом исключения переменных и с помощью правила множителей Лагранжа.

a) $f(x) = x_1 x_2 \to \text{extr}; \quad 3\mathrm{x}_1 + \mathrm{x}_2 = 6;$

b) $f(x) = e^{x_1 x_2} \to \text{extr}; \quad \mathrm{x}_1 + \mathrm{x}_2 = 1;$

c) $f(x) = 2x_1^2 + x_2^2 \to \text{extr}; \quad 2\mathrm{x}_1 + \mathrm{x}_2 = 1;$

d) $f(x) = x_1 + x_2 + x_3^2 \to \text{extr}; \quad -\mathrm{x}_1 + \mathrm{x}_3 = 1, \quad -\mathrm{x}_1\mathrm{x}_3 + \mathrm{x}_2 = 1;$

e) $f(x) = x_1 x_2 + x_1 x_3 + x_2 x_3 \to \text{extr}; \quad \mathrm{x}_1 + \mathrm{x}_2 + \mathrm{x}_3 = 3;$

f) $f(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 \to \text{extr}; \quad \mathrm{x}_1 + \mathrm{x}_2 - \mathrm{x}_4 = 6, \quad \mathrm{x}_1 + \mathrm{x}_3 + \mathrm{x}_4 = 9;$

g) $f(x) = (x_1 - 3)^2 + (x_2 - 2)^2 \to \text{extr}; \quad \mathrm{x}_1 + \mathrm{x}_2 = \mathrm{a},$

где a – произвольное число;

h) $f(x) = x_1^2 - x_2^2 \to \text{extr}; \quad \mathrm{x}_1 - \mathrm{x}_2 = 10;$

i) $f(x) = x_1^2 + 2x_2^2 + 3x_3^2 \to \min; \quad x_1 + x_2 = 1, \quad x_1 - x_3 = 2;$

j) $f(x) = 2x_1^2 + x_2^2 + x_3^2 \to \min; \quad x_1 + x_2 = 4, \quad x_2 + x_3 = 3;$

k) $f(x) = 3x_1^2 + 2x_2^2 + x_3^2 \to \min; \quad x_1 + x_2 + x_3 = 3, \quad x_1 - 2x_2 = -1;$

l) $f(x) = 4x_1^2 + x_2^2 + 2x_3^2 \to \min; \quad x_1 - 2x_2 + x_3 = 1, \quad x_1 - x_3 = -1;$

m) $f(x) = x_1^2 + x_2^2 + x_3^2 \to \min; \ x_1 + 2x_2 + 3x_3 = 7, \ 2x_1 + 2x_2 + x_3 = 9/2;$

n) $f(x) = x_1^2 + 2x_2^2 + (x_3 - 2)^2 \to \min; \ x_1 + x_2 + x_3 = 1, \ x_1 - x_2 + x_3 = 1;$

o) $f(x) = x_1^2 x_2^3 x_3^4 \to \text{extr}; \quad \mathrm{x}_1 + \mathrm{x}_2 + \mathrm{x}_3 = 18;$

p) $f(x) = x_1 x_2 + x_2 x_3 \to \text{extr}; \quad \mathrm{x}_1 + \mathrm{x}_2 = 4, \quad \mathrm{x}_2 + \mathrm{x}_3 = 4;$

q) $f(x) = x_1^2 + x_2^2 + \ldots + x_n^2 \to \min; \quad x_1 + 2x_2 + \ldots + nx_n = 1;$

r) $f(x) = x_1 + x_2 + x_3 + x_4 \to \text{extr}; \quad \mathrm{x}_1\mathrm{x}_2\mathrm{x}_3\mathrm{x}_4 = \mathrm{c}^4;$

s) $f(x) = a^2 x_1^2 + b^2 x_2^2 + c^2 x_3^2 - (a x_1^2 + b x_2^2 + c x_3^2)^2 \to \text{extr};$

$$x_1^2 + x_2^2 + x_3^2 = 1 \quad (a > b > c > 0).$$

3.4. Решить задачи:

a) $f(x) = x_1^2 + 3x_2^2 + 3 \to \text{extr}; \quad \mathrm{x}_1^3 + \mathrm{x}_2^3 = 1;$

b) $f(x) = 2x_1 \to \text{extr}; \quad \mathrm{x}_1^3 + \mathrm{x}_2^3 - 3\mathrm{x}_1\mathrm{x}_2 = 0;$

c) $f(x) = 4x_1x_2 \to \text{extr}; \quad \mathrm{x}_1^2 + \mathrm{x}_2^2 = 9;$

d) $f(x) = 2x_1^3 + 2x_2^3 + 3x_1 + 3x_2 \to \text{extr}; \quad \mathrm{x}_1^2 + \mathrm{x}_2^2 = 1;$

e) $f(x) = x_1 + x_2 + x_3 \to \min; \quad \dfrac{1}{x_1} + \dfrac{1}{x_2} + \dfrac{1}{x_3} = 1;$

f) $f(x) = x_1 + x_2 + x_3 \to \text{extr}; \quad \mathrm{x}_1^2 + \mathrm{x}_2^2 + \mathrm{x}_3^2 = 1;$

g) $f(x) = x_1x_2x_3 \to \min; \quad x_1^2 + x_2^2 + x_3^2 = 1, \quad x_1 + x_2 + x_3 = 1;$

h) $f(x) = x_1x_2 + x_2x_3 \to \max; \quad x_1^2 + x_2^2 = 2, \quad x_2 + x_3 = 2;$

i) $f(x) = x_1^2 + 12x_1x_2 + 2x_2^2 \to \text{extr}; \quad 4\mathrm{x}_1^2 + \mathrm{x}_2^2 = 25;$

j) $f(x) = x_1 \to \min; \quad 9x_1^3 - x_2^3 = 0;$

k) $f(x) = \sum\limits_{i=1}^{n} x_i \to \min; \quad \|x\| = 1;$

l) $f(x) = \|x\| \to \min; \quad \sum\limits_{i=1}^{n} x_i = 1;$

m) $f(x) = \sum\limits_{i=1}^{n} x_i^2 \to \min; \quad \sum\limits_{i=1}^{n} x_i^4 = 1;$

n) $f(x) = \langle Gx, x\rangle \to \min; \quad \|x\| = 1,$

где G – симметричная матрица порядка n с известными собственными значениями $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n > 0;$

o) $f(x) = \|x\| \to \min; \quad \langle Gx, x\rangle = 1;$

p) $f(x) = \dfrac{1}{2} < Gx, x > + < c, x > + d \to \min; \quad Ax = b,$

G – положительно определенная матрица порядка n,

$A \in M_{m,n}, \ \operatorname{rank} A = m;$

q) $f(x) = \|Gx - c\|^2 \to \min; \quad Ax = b,$

$G \in M_{k,n}, \ \operatorname{rank} G = n, \ A \in M_{m,n}, \ \operatorname{rank} A = m;$

r) $f(x) = \langle c, x\rangle \to \text{extr}; \quad \|\mathrm{x} - \mathrm{a}\| \leq \mathrm{r};$

s) $f(x) = x_1 \to \text{extr};\ \mathrm{x}_1^{\mathrm{p}} + \mathrm{x}_2^{\mathrm{q}} = 0,$ p и q – натуральные числа.

3.5. Проверить, является ли точка $x = (1, 1, 1)$ точкой экстремума в задаче

$$x_1x_2 + x_1x_3 + x_2x_3 \to \mathrm{extr}; \quad 2\mathrm{x}_1^3\mathrm{x}_2^2\mathrm{x}_3 + 4\mathrm{x}_1^2 + 5\mathrm{x}_2^2 + 6\mathrm{x}_3^2 = 17.$$

3.6. Исследовать на экстремум функцию $f = f(x_1, x_2, x_3)$, заданную неявно:

$$x_1^2 + x_2^2 + f^2 - x_1 f - x_2 f + 2x_1 + 4x_2 + 2x_3 + 2f - 1 = 0$$

при условии $x_2 + x_3 = 0$.

3.7. Решить задачи:

a) $2x_1^2 + 2x_1 + 4x_2 - 3x_3 \to \min;$

$8x_1 - 5x_2 + 4x_3 \leq 40, \quad -2x_1 + x_2 - x_3 = 0;$

b) $-x_1^2 - x_2^2 - x_3^2 \to \max;$

$x_1 + x_2 + x_3 = 3, \quad x_1 - x_2 + 2x_3 \leq 5;$

c) $x_1 \to \min; \; x_1^2 + x_2^2 \leq 1, \quad x_1^3 + x_2^3 = 1;$

d) $x_1x_2x_3 \to \min; \; x_1^2 + x_2^2 + x_3^2 = 1, \quad x_1 + x_2 + x_3 \leq 0;$

e) $\sum\limits_{i=1}^{n} x_i^4 \to \max; \quad \sum\limits_{i=1}^{n} x_i^2 \leq 1.$

3.8. Найти расстояние от точки $a = (1, 2)$ на плоскости до прямой $2x+3y = 1$ (определение расстояния от точки до множества см. в упражнении 2.1.12).

3.9. Найти расстояние между началом координат и кривой $y = x^{-2}$.

3.10. *Расстоянием между множествами* X и Y называется число $\rho(X, Y) = \inf\limits_{x \in X, y \in Y} \|x - y\|$. Найти расстояние между параболой $y = x^2$ и прямой $x - y = 2$.

3.11. Найти проекцию точки $v \in R^n$ на следующие множества:

a) гиперплоскость $\langle a, x\rangle = \beta$; b) полупространство $\langle a, x\rangle \leq \beta$;

c) шар $\|x - a\| \leq r$; d) отрезок $[x, y]$;

e) прямую $x = a + tp, \quad -\infty < t < +\infty$

(определение проекции точки на множество см. в упражнении 2.1.13).

3.12. Найти проекцию точки $v \in R^n$ на множество

$$X = \{x \in R^n : \langle a^1, x\rangle \le b_1, \quad \langle a^2, x\rangle \le b_2\}.$$

3.13. Пусть A – $(m \times n)$ – матрица, $\mathrm{rank} A = m$. Доказать,что проекция точки $v \in R^n$ на линейное многообразие $X = \{x \in R^n : \quad Ax = b\}$ вычисляется по формуле

$$\pi_X(v) = v - A^T(AA^T)^{-1}(Av - b).$$

3.14. Среди всех прямоугольных параллелепипедов, сумма ребер которых имеет заданную длину l, найти параллелепипед наибольшего объема.

3.15. Данное положительное число a представить в виде суммы n положительных слагаемых так, чтобы сумма их квадратов была наименьшей.

3.16. Определить размеры прямоугольного бассейна данного объема V, при которых на облицовку боковой поверхности и дна потребуется наименьшее количество материала.

3.17. Среди всех вписанных в данный круг радиуса R треугольников найти тот, площадь которого наибольшая. Дать геометрическую иллюстрацию.

3.18. Среди треугольных пирамид с заданными основанием и высотой найти ту, которая имеет наименьшую площадь боковой поверхности.

3.19. Рассмотрим задачу

$$\langle c, x\rangle \to \min; \;\; Ax = b,$$

где $A \in M_{m,n}$. Доказать, что либо эта задача не имеет решения, либо любая допустимая точка является решением, причем последнее имеет место в том и только том случае, если вектор c есть линейная комбинация строк матрицы A, т. е. существует вектор λ, такой, что $c = A^T\lambda$.

§ 4. Задача с ограничениями типа равенств и неравенств

Рассмотрим схему применения правила множителей Лагранжа для более общих задач на условный экстремум:

$$f(x) \to \min, \tag{4.1}$$

$$\begin{gathered} x \in X = \{x \in \Omega : g_i(x) \le 0, i = 1, 2, \ldots, r; \\ g_i(x) = 0, i = r+1, r+2, \ldots, m,\ m - r < n\}, \Omega \subset R^n. \end{gathered} \tag{4.2}$$

Т е о р е м а 4.1. Пусть x^* – точка локального минимума в задаче математического программирования (4.1)–(4.2), функции f, $g_1, \ldots, g_r$ дифференцируемы в точке x^*, функции g_{r+1}, g_{r+2}, $\ldots$, g_m непрерывно дифференцируемы в некоторой окрестности точки x^*, Ω – выпуклое множество. Тогда существует число λ_0^* и вектор $\lambda^* \in R^m$ такие, что выполняются:

1) условие нетривиальности: $\lambda_0^{*2} + \|\lambda^*\|^2 > 0$, т. е. хотя бы один из множителей Лагранжа $\lambda_0, \lambda_1, \ldots, \lambda_m$ отличен от нуля;

2) условие неотрицательности: $\lambda_0^* \ge 0, \lambda_i^* \ge 0, i = 1, 2, \ldots, r$, т. е. множители Лагранжа, отвечающие целевой функции и ограничениям-неравенствам, неотрицательны;

3) условия дополняющей нежесткости:

$$\lambda_i^* g_i(x^*) = 0, i = 1, 2, \ldots, r;$$

4) условие стационарности

$$< \frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x}, x - x^* > \ge 0 \tag{4.3}$$

для всех $x \in \Omega$. Здесь $L(x, \lambda_0, \lambda) = \lambda_0 f(x) + \sum_{i=1}^n \lambda_i g(x)$ – функция Лагранжа, $\frac{\partial L(x,\lambda_0,\lambda)}{\partial x}$ составлена из частных производных функции Лагранжа по компонентам вектора x.

Наиболее простыми для проверки достаточными условиями регулярности рассматриваемой задачи, гарантирующими существование набора множителей Лагранжа с $\lambda_0^* = 1$, является (в дополнение к условиям теоремы 4.1):

1) выпуклость функций $g_i, i = 1, 2, \ldots, r$, отсутствие ограничений-равенств ($m = r$) и существование точки $\bar{x} \in \Omega$ такой, что $g_i(\bar{x}) < 0$ для всех $i = 1, 2, \ldots, r$ (условие Слейтера)
или

2) линейность всех функций g_i, $i = r+1, r+2, \ldots, m$ при дополнительном предположении, что Ω – полиэдр, т. е. множество, являющееся пересечением конечного числа полупространств вида

$$< a^j, x > \leq \gamma_j, a^j \in R^n.$$

Условия оптимальности второго порядка достаточно громоздки при проверке. Отметим, в частности, что если функции f, $g_1, \ldots, g_m$ дважды дифференцируемы в точке x^*, существует набор $\{\lambda_0^*, \lambda^*\}$, для которого выполняются утверждения 1)–4) теоремы 4.1, а квадратичная форма

$$< \frac{\partial^2 L(x^*, \lambda_0^*, \lambda^*)}{\partial x^2} h, h >$$

положительно определена на R^n, тогда x^* – точка строгого локального минимума функции f на множестве X.

Укажем одно достаточное условие глобального минимума [16].

Т е о р е м а 4.2. Предположим, что в задаче (4.1)–(4.2) множество Ω выпукло, функции f, $g_1, \ldots, g_r$ выпуклы на Ω и дифференцируемы в точке $x^* \in X$, а функции $g_{r+1}, g_{r+2}, \ldots, g_m$ линейны. Если для некоторого набора множителей Лагранжа $\{\lambda_0^*, \lambda^*\}$, $\lambda_0^* = 1$ справедливы утверждения 1)–4) теоремы 4.1, то x^* – точка глобального минимума рассматриваемой целевой функции на множестве (4.2).

Итак, общая схема решения рассматриваемой задачи выглядит следующим образом.

1. Составляется функция Лагранжа.

2. Выписываются необходимые условия экстремума, сформулированные в утверждениях 1)–4) теоремы 4.1. К этим условиям добавляются ограничения, задающие множество (4.2). Полученная система алгебраических уравнений и неравенств используется для поиска подозрительных на экстремум точек. Так же как и для задачи с ограничениями-равенствами целесообразно проанализировать отдельно случай $\lambda_0 = 0$ и $\lambda_0 = 1$ (или λ_0 – любое положительное число). Если выполнено одно из условий регулярности, то вариант $\lambda_0 = 0$ рассматривать не надо.

3. Среди найденных стационарных точек отыскиваются точки минимума и проводится анализ на глобальный экстремум (если это возможно).

Заметим, что множество Ω обычно также задается с помощью ограничений типа равенства и неравенства. Однако для некоторых специальных типов множеств иногда имеет смысл не включать эти ограничения в основные ограничения – равенства и неравенства (4.2).

Рассмотрим два часто встречающихся случая.

а) $\Omega = R^n$.

Неравенство (4.3) выполняется для любых $x \in R^n$ в том и только том случае, когда (проверьте!)

$$\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_j} = 0, j = 1, 2, \ldots, n. \tag{4.4}$$

П р и м е р 4.1.

Решить задачу математического программирования:

$$f(x) = 3x_2^2 - 11x_1 - 3x_2 - x_3 \to \min,$$

$$x \in X = \{x \in R^3 : x_1 - 7x_2 + 3x_3 \leq -1, x_3 + 2 \geq 5x_1 + 2x_2\}.$$

В рассматриваемом примере выполняются условия регулярности, поскольку ограничения-равенства отсутствуют, а ограничения–неравенства задаются линейными функциями. Поэтому функцию Лагранжа рассмотрим сразу в нормальном виде ($\lambda_0 = 1$):

$$L(x, \lambda) = 3x_2^2 - 11x_1 - 3x_2 - x_3 + \lambda_1(x_1 - 7x_2 + 3x_3 + 7) +$$

$$+ \lambda_2(5x_1 + 2x_2 - x_3 - 2).$$

Уравнения (4.4):

$$\lambda_1 + 5\lambda_2 = 11,$$

$$6x_2 - 7\lambda_1 + 2\lambda_2 = 3,$$

$$3\lambda_1 - \lambda_2 = 1.$$

Условия неотрицательности множителей Лагранжа и дополняющей нежесткости:

$$\lambda_1 \geq 0, \lambda_2 \geq 0,$$

$$\lambda_1(x_1 - 7x_2 + 3x_3 + 7) = 0,$$

$$\lambda_2(5x_1 + 2x_2 - x_3 - 2) = 0.$$

Отсюда $\lambda_1^* = 1, \lambda_2^* = 2, x^* = (0, 1, 0)$. Поскольку целевая функция и функции, задающие ограничения, удовлетворяют условиям теоремы 4.2, то x^* – точка глобального минимума, а $\min\limits_{x \in X} f(x) = f(x^*) = 0$.

б) $\Omega = \{x \in R^n : \alpha_j \leq x_j \leq \beta_j, j = 1, 2, \ldots, n\}$.

Неравенство (4.3) выполняется для всех $x \in \Omega$ тогда и только тогда, когда (проверьте!)

$$\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_j}(x_j - x_j^*) \geq 0 \quad \forall x_j : \alpha_j \leq x_j \leq \beta_j, j = 1, 2, \ldots, n.$$

Следовательно,

$$\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_j} \begin{cases} = 0, & \text{если } \ \alpha_j < x_j^* < \beta_j, \\ \geq 0, & \text{если } \ x_j^* = \alpha_j, \\ \leq 0, & \text{если } \ x_j^* = \beta_j. \end{cases} \tag{4.5}$$

П р и м е р 4.2.

Решить с помощью правила множителей Лагранжа

$$f(x) = x_1^2 - x_2 \to \min,$$

$$x_1 + x_2 = 2, \ \Omega = \{x \in R^2 : 0 \leq x_1 \leq 5, 0 \leq x_2 \leq 1\}.$$

Функция Лагранжа (с учетом выполнения условия регулярности):

$$L(x, \lambda) = x_1^2 - x_2 + \lambda(x_1 + x_2 - 2).$$

Условия (4.5):

$$2x_1 + \lambda \begin{cases} = 0, & \text{если } \ 0 < x_1 < 5, \\ \geq 0, & \text{если } \ x_1 = 0, \\ \leq 0, & \text{если } \ x_1 = 5; \end{cases} \qquad -1 + \lambda \begin{cases} = 0, & \text{если } \ 0 < x_2 < 1, \\ \geq 0, & \text{если } \ x_2 = 0, \\ \leq 0, & \text{если } \ x_2 = 1. \end{cases}$$

Варианты $x_1 = 0$ и $x_1 = 5$ невозможны, так как при этом точки, удовлетворяющие ограничению-равенству, не принадлежат Ω. Следовательно, $x_1 = \frac{-\lambda}{2}$. Отсюда получаем единственную подозрительную на экстремум точку $x^* = (1,1)$, $\lambda^* = -2$. В силу теоремы 4.2 x^* – точка глобального минимума, причем $\min\limits_X f(x) = 0$.

В заключение укажем, что рассмотренная схема позволяет одновременно решать также и задачу максимизации, если исключить условие $\lambda_0 \geq 0$. При этом, случаю $\lambda_0 \leq 0$ будут соответствовать точки, подозрительные на максимум (переформулируйте соответствующим образом теорему 4.2). При решении некоторых задач удается применить метод исключения (см. пример 4.2; выразить $x_1 = 2 - x_2$). Наконец, некоторые задачи в пространстве R^2 можно решить геометрически, построив линии уровня целевой функции и множеств ограничений.

П р и м е р 4.3.

Рассмотрим модель поведения потребителя на рынке товаров и услуг. Пусть вектор $x = (x_1, x_2, \ldots, x_n)$ характеризует объем потребляемых товаров или услуг, т. е. x_i – количество единиц потребляемого товара или услуги i-го вида, $x_i \geq 0$.

О п р е д е л е н и е 4.1. Функция $u(x) = u(x_1, x_2, \ldots, x_n)$, характеризующая ценность для потребителя приобретаемого набора товаров и услуг называется *индивидуальной функцией полезности потребителя.*

Очевидно, что функция полезности есть функция, определенная для неотрицательных значений компонент вектора x. Если набор $\bar{x}$ предпочтительнее для потребителя, чем набор $\bar{\bar{x}}$, то $u(\bar{x}) > u(\bar{\bar{x}})$.

О п р е д е л е н и е 4.2. *Предельной полезностью товара* i называется частная производная функции полезности по переменной x_i :

$$M_i u(x) = \frac{\partial u(x)}{\partial x_i}.$$

Относительно функции полезности обычно принимаются следующие основные гипотезы:

1) Функция $u(x)$ строго монотонно возрастает по каждому из своих аргументов, т. е., если $\bar{x} = (\bar{x}_1, \dots, \bar{x}_i, \dots, \bar{x}_n)$,
$\bar{\bar{x}} = (\bar{x}_1, \dots, \bar{x}_{i-1}, \bar{\bar{x}}_i, \bar{x}_{i+1}, \dots, \bar{x}_n)$, $\bar{x}_i > \bar{\bar{x}}_i$, то $u(\bar{x}) > u(\bar{\bar{x}})$.

2) Функция полезности является строго вогнутой при $x > 0$.

Гипотеза строгой вогнутости, в частности, означает отрицательность вторых производных:

$$\frac{\partial^2 u(x)}{\partial x_i^2} = \frac{\partial}{\partial x_i}\Big[Mu_i(x)\Big] < 0,$$

при $x_i > 0,\ i = 1, 2, \dots, n$. Это означает убывание предельных полезностей с ростом объема потребления.

В прикладных задачах допускается замена условий 1)–2) более слабыми требованиями неубывания и вогнутости.

Пусть D – бюджет потребителя. Тогда соответствующая задача математического программирования может быть записана в виде:

$$u(x) \to \max,$$

$$\sum_{j=1}^{n} p_j x_j \le D, \quad x_j \ge 0, \quad j = 1, 2, \dots, n.$$

Перейдя к задаче минимизации функции $-u(x)$, запишем функцию Лагранжа в нормальном виде:

$$L(x, \lambda) = -u(x) + \lambda(< p, x > -D).$$

$$\frac{\partial L}{\partial x_j} = -M_j(x) + \lambda p_j,$$

где $M_j(x) = \dfrac{\partial u}{\partial x_j}$ – предельная полезность j-го товара. Условия (4.5) в данном случае имеют вид:

$$-M_j(x) + \lambda p_j \begin{cases} \ge 0, & x_j = 0, \\ = 0, & x_j > 0. \end{cases}$$

Таким образом, если в оптимальном наборе $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ приобретаемых товаров или услуг $x_j^* > 0$ (товар покупается), то

$\lambda^* = \dfrac{M_j(x^*)}{p_j}$. То есть отношение предельной полезности к цене каждого приобретаемого товара при оптимальном плане покупок является одной и той же константой, равной множителю Лагранжа.

Задачи и упражнения

4.1. Показать, что неравенство (4.3) в правиле множителей Лагранжа выполняется тогда и только тогда, когда справедливы равенства (4.4), если x^* – внутренняя точка множества Ω.

4.2. Получить условия, аналогичные (4.5), если

$$\Omega = \{x \in R^n : x_j \geq 0, j \in J \subset \{1, 2, \ldots, n\}\}.$$

4.3. Доказать, что если выполнены условия правила множителей Лагранжа, $\Omega = \{x \in R^n : \|x - r\|^2 \leq a^2\}$, а точка локального минимума x^* лежит на границе Ω, т. е. $\|x^* - r\|^2 = a^2$, то

$$x^* = r - a\frac{L_x(x^*, \lambda_0^*, \lambda^*)}{\|L_x(x^*, \lambda_0^*, \lambda^*)\|}.$$

4.4. Решить с помощью правила множителей Лагранжа. Результат проверить теоретически, построив множества ограничений и линии уровня целевых функций:

a) $f(x) = x_1^2 - 6x_1 - x_2 \to extr,$

$x_1 + 2x_2 \leq 15,$

$2x_1 + 3x_2 \leq 24,$

$3x_1 + 2x_2 \leq 24,\ 0 \leq x_2 \leq 2;$

b) $f(x) = (x_1 - 7)(x_2 - 1) \to extr,$

$x_1 + 2x_2 \leq 19,$

$x_1 + x_2 \leq 9,$

$x_1 \geq 0,\ x_2 \geq 0;$

c) $f(x) = (x_1 - 2)^2 + (x_2 - 3)^2 \to extr,$

$x_1 + x_2 \leq 9,$

$x_1 + 2x_2 \leq 12,$

$x_1 \geq 0,\ x_2 \geq 0;$

d) $f(x) = x_1^2 - x_1 - x_2 \to \min,$

$x_1^2 + 3x_2^2 + 2x_1 \leq 3,$

$x_1 \geq 0,\ x_2 \geq 0;$

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

$x_1^2 + x_2^2 \leq 25,$

$x_1 x_2 \geq 4,$

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

4.5. Применить правило множителей Лагранжа к решению следующих задач:

a) $2x_1^2 + 2x_1 + 4x_2 - 3x_3 \to \min,$

$8x_1 - 5x_2 + 4x_3 \leq 40,$

$-2x_1 + x_2 - x_3 = 0;$

b) $2x_1^2 + 2x_2^2 - x_1 x_2 - x_1 - 4x_2 \to \min,$

$x_1 + 2x_2 \leq 12,$

$x_1 \geq 0;$

c) $x_1^2 + x_2^2 + x_3^2 \to \max,$

$x_1 + x_2 + x_3 \leq 15,$

$x_i \geq 0, i = 1, 2, 3;$

d) $-x_1^2 - x_2^2 - x_3^2 \to \max,$

$x_1 + x_2 + x_3 = 3,$

$x_1 - x_2 + 2x_3 \leq 5;$

e) $x_1^2 + x_2^2 + x_3^2 - 2x_1 - 2x_2 - 2x_3 + 10 \to \min,$

$x_1^2 + x_2^2 + x_3^2 \leq 4,$

$x_3 \geq 0;$

f) $x_1 \to \min,$

$x_1^2 + x_2^2 \leq 1,$

$x_1^3 + x_2^3 = 1;$

g) $x_1 x_2 x_3 \to \min,$

$x_1^2 + x_2^2 + x_3^2 = 1,$

$x_1 + x_2 + x_3 \leq 0;$

h) $\sum\limits_{i=1}^{n} x_i^4 \to \max,$

$\sum\limits_{i=1}^{n} x_i^2 \leq 1.$

Примечание: во всех примерах достаточно просто удается найти точки глобального экстремума.

4.6. Найти расстояние от точки y до множества X:

a) $y = (2, 1, 0, 1),\ \ X = \{x \in R^4 : 2x_1 + x_2 + x_3 + x_4 \leq 1\};$

b) $y = (0, 1, 0, 1),\ \ X = \{x \in R^4 : 2x_1 + x_2 + x_3 - x_4 \geq 2, x_4 \geq 0\}.$

4.7. Пусть функция полезности $u(x) = \sqrt{x_1 x_2}$, вектор цен $p = (3, 1)$, бюджет потребителя равен 6 ед. Определить оптимальный набор приобретаемых товаров. Вычислить предельные полезности каждого товара при оптимальном плане покупок.

§ 5. Экстремум линейных и квадратичных функций на простых множествах

Поиск точек экстремума путем решения систем уравнений, используемых в правиле множителей Лагранжа, представляет весьма серьезную задачу. Однако для линейных и квадратичных целевых функций можно легко получить аналитические решения экстремальных задач на множествах специального вида. В дальнейшем такие множества будем называть *простыми*. Рассмотрим два типа подобных задач.

5.1. Минимизация линейной функции на выпуклом, замкнутом и ограниченном множестве. Рассмотрим задачу

$$f(x) =< c, x >= \sum_{i=1}^{n} c_i x_i \to \min, \quad x \in X. \tag{5.1}$$

Здесь c – заданный вектор, $c \in R^n$, $c \neq 0$; множество X – выпукло, замкнуто и ограничено.

Остановимся на важнейших свойствах задачи (5.1).

У т в е р ж д е н и е 5.1. Решение задачи (5.1) существует. Любая точка минимума или максимума является точкой глобального экстремума.

Справедливость данного утверждения следует из теоремы Вейерштрасса и того факта, что линейная функция является одновременно выпуклой и вогнутой.

У т в е р ж д е н и е 5.2. Экстремум линейной функции на некотором множестве может достигаться лишь на границе этого множества.

Достаточно просто удается найти экстремум линейной функции на многомерном параллелепипеде, шаре, эллипсоиде, а также некоторых других простейших множествах (см. упражнения).

5.2. Задача проецирования на выпуклое и замкнутое множество. Напомним, что *проекцией точки* $y \in R^n$ на множество $X \subset R^n$ называется такая точка $p \in X$, что

$$\|p - y\| = \min_{x \in X} \|x - y\|.$$

Будем применять обозначение $p = P_X(y)$. Таким образом, проекция точки y – это точка p, расстояние до которой от точки y равно расстоянию от точки y до множества X.

Перечислим основные свойства проекции [5, 7].

У т в е р ж д е н и е 5.3. Если $X \subset R^n$ – непустое замкнутое множество, то проекция на него любой точки $y \in R^n$ существует. Проекция принадлежит границе множества X, если $y \notin intX$.

У т в е р ж д е н и е 5.4. Если множество $X \subset R^n$ – непустое, замкнутое и выпуклое, то проекция на нее любой точки существует и единственна.

У т в е р ж д е н и е 5.5. Необходимым и достаточным условием того, чтобы точка p являлась проекцией y на выпуклое множество X является выполнение неравенства

$$< p - y, x - p > \geq 0 \quad \forall x \in X. \tag{5.2}$$

Необходимость решения задачи проецирования точки на множество возникает, в частности, при использовании метода проекций градиента для численного решения задач математического программирования.

Таким образом, для поиска проекции точки y на множество X необходимо решить задачу

$$\|x - y\| \to min, \ \ x \in X. \tag{5.3}$$

При решении задачи проецирования обычно (5.3) заменяется на эквивалентную задачу

$$\frac{1}{2}\|x-y\|^2 \to min, \;\; x \in X. \tag{5.4}$$

целевая функция в (5.4) является сепарабельной и всюду дифференцируемой.

Задачи и упражнения

Найти решения задач минимизации линейной функции $f(x) =< c, x >$, $c \in R^n$, $c \neq 0$ для следующих видов множества X :

5.1. X – n-мерный куб

$$X = \{x \in R^n : \;\; |x_i| \leq l, \;\; i = 1, 2, \ldots, n, \;\; l > 0\};$$

5.2. X – n-мерный прямоугольный параллелепипед

$$X = \{x \in R^n : \;\; \alpha_i \leq x_i \leq \beta_i, \;\; i = 1, 2, \ldots, n\};$$

5.3. X – n-мерный шар

$$X = \{x \in R^n : \;\; \|x - r\| \leq l, \;\; l > 0\},$$

r – заданный вектор из R^n;

5.4. X – n-мерный эллипсоид

$$X = \{x \in R^n : \;\; \sum_{i=1}^{n} \alpha_i^2 x_i^2 \leq l, \;\; l > 0\};$$

Числа α_i заданы, некоторые из них могут равняться нулю.

5.5. X – n-мерный фундаментальный симплекс

$$X = \{x \in R^n : \;\; \sum_{i=1}^{n} x_i = 1, \;\; x_i \geq 0, \;\; i = 1, 2, \ldots, n\};$$

5.6. X – простейший полиэдр

$$X = \{x \in R^n : \; < a, x > \leq b, \;\; x \geq 0\},$$

вектор $a \in R^n,\ a > 0$ и число $b > 0$ заданы.

5.7. X – множество Лебега квадратичной функции

$$X = \{x \in R^n :\ \frac{1}{2} < Ax, x > + < b, x > \leq r^2\},$$

A – симметрическая положительно определенная матрица.

5.8. Предположим, что экстремум линейной функции достигается во внутренней точке множества X. Показать, что этот случай невозможен, т. е. значение целевой функции можно улучшить. (Тем самым будет доказано утверждение 5.2.)

Решить задачи проецирования произвольной точки $y \in R^n$ на следующие множества X.

5.9. X – n-мерный куб

$$X = \{x \in R^n :\ \ |x_i| \leq l,\ \ i = 1, 2, \ldots, n,\ \ l > 0\};$$

5.10. X – n-мерный прямоугольный параллелепипед

$$X = \{x \in R^n :\ \ \alpha_i \leq x_i \leq \beta_i,\ \ i = 1, 2, \ldots, n\};$$

5.11. X – неотрицательный ортант

$$X = \{x \in R^n :\ \ x_i \geq 0,\ \ i = 1, 2, \ldots n\};$$

5.12. X – n-мерный шар

$$X = \{x \in R^n :\ \ \|x - r\| \leq l,\ \ l > 0\},$$

r – заданный вектор из R^n;

5.13. X – n-мерное пространство с «выколотым» шаром

$$X = \{x \in R^n :\ \ \|x - r\| \geq l,\ \ l > 0\},$$

r – заданный вектор из R^n;

5.14. X – гиперплоскость.

$$X = \{x \in R^n :\ < b, x >= \gamma\}$$

вектор $b \in R^n$ и число γ заданы;

5.15. X – полупространство

$$X = \{x \in R^n : \ < b, x > \leq \gamma\};$$

5.16. X – линейное многообразие

$$X = \{x \in R^n : \quad Ax = b\},$$

A – прямоугольная матрица размерности $m \times n$ $(m < n)$ с линейно независимыми строками, $b \in R^n$ – заданный вектор.

5.17. Множество $X \in R^n$ называется аффинным, если оно вместе с любыми двумя своими элементами целиком содержит соединяющую эти элементы прямую, т. е., если $x, y \in X$, то $\alpha x + (1 - \alpha) y \in X$ для всех чисел $\alpha \in R^1$. Показать, что для аффинного множества условие (5.2) равносильно следующему: $< p-y, x-p >= 0$ для всех $x \in X$. Проверить с помощью этого критерия результат решения упражнения 5.14.

Глава 3
Численные методы решения задач одномерной оптимизации

Пусть функция $f(x)$ непрерывна на отрезке $[a,b]$. Рассмотрим задачу

$$f(x) \to \min; \quad x \in [a,b]. \tag{1.1}$$

Решение задачи (1.1) обозначим через x^*. Через f_* будем обозначать минимальное значение функции $f(x)$ на отрезке $[a,b]$:

$$f_* = f(x^*) = \min_{[a,b]} f(x).$$

Разумеется, задача (1.1) в ряде случаев может быть решена классическим методом, известным еще из курса средней школы. Необходимо вычислить первую производную, приравнять ее нулю, найти корни соответствующего алгебраического уравнения, исследовать в критических точках вторые производные или посмотреть на изменение знака первой производной. При этом нужно не забыть также про концы отрезка и точки, в которых производная не существует. Сразу же видны возможные трудности применения классической схемы. Во-первых, целевая функция может быть вообще недифференцируемой. Во-вторых, нужно не просто определить точку, в которой равна нулю первая производная, а найти все такие точки. Это означает, что необходимо найти все корни, вообще говоря, нелинейного алгебраического уравнения. А ведь это самостоятельная задача, которая часто по сложности является не менее сложной, чем исходная задача на экстремум. И даже найдя все критические точки, мы еще не решим оптимизационную задачу - нужны дополнительные исследования. Наконец, очень часто задачи одномерной оптимизации возникают в качестве вспомогательных на итерациях решения более сложных экстремальных задач. При этом функции одной переменной нельзя записать в виде аналитического выражения в явном виде. Перечисленные выше обстоятельства указывают на важность эффективных численных методов решения задач одномерного поиска. Краткий обзор наиболее распространенных методов приведен в этом разделе.

1. Методы перебора. Простейший способ приближенного решения задачи (1.1) очевиден и состоит в следующем. Отрезок $[a, b]$ разбивается на n частей точками $a = x_0 < x_1 < \ldots < x_N = b$; в этих точках вычисляется значение функции $f(x)$ и среди них выбирается минимальное. Рассмотрим два варианта этого метода.

Метод равномерного перебора. Так называется метод, использующий равноотстоящие узлы $x_0, x_1, \ldots, x_N$. Формально его можно описать следующим образом:

$$x_{k+1} = x_k + h, \quad k = 0, \ldots, N-1,$$

$$x_0 = a, \quad h = \frac{b-a}{N}; \tag{1.2}$$

$$\widehat{x} = \arg \min_{k=\overline{0,N}} f(x_k). \tag{1.3}$$

Если функция $f(x)$ удовлетворяет на отрезке $[a, b]$ условию Липшица с константой L, т. е.

$$|f(x) - f(y)| \leq L|x - y|, \quad \forall x, y \in [a, b], \tag{1.4}$$

то при $h \leq 2\varepsilon/L$

$$f(\widehat{x}) - f_* \leq \varepsilon. \tag{1.5}$$

Метод адаптивного перебора. Пусть функция $f(x)$ удовлетворяет условию (1.4), $\varepsilon > 0$ – задано. Положим $h = 2\varepsilon/L$ и сетку $x_0, \ldots, x_N$ построим следующим образом:

$$x_{k+1} = x_k + h + h_k, \quad k = 0, \ldots, N-2,$$

$$x_0 = a, \quad x_N = b.$$

Здесь число узлов N не фиксируется заранее, а определяется в процессе вычислений из условия $x_{N-1} < b \le x_{N-1} + h + h_k$; шаг h_k вычисляется по формуле

$$h_k = \frac{f(x_k) - F_k}{L},$$

где $F_k = \min\limits_{i=\overline{0,k}} f(x_i)$. Приближенное решение задачи (1.1), точка $\widehat{x}$, определяется так же, как и в методе равномерного перебора, $\widehat{x} = \arg\min\limits_{k=\overline{0,N}} f(x_k)$, и удовлетворяет условию (1.5).

2. Метод ломаных. Пусть функция $f(x)$ удовлетворяет условию (1.4). Метод ломаных для решения задачи (1.1) состоит в построении двух последовательностей точек – вспомогательной $\{\bar{x}_0, \bar{x}_1, \ldots\}$ и основной $\{x_0, x_1, \ldots\}$ по следующим правилам:

$$x_0 \in [a, b] - \text{произвольна}; \tag{1.6}$$

$$\bar{x}_k = \arg\min_{[a,b]} f_k(x), \quad k = 1, 2, \ldots\ ; \tag{1.7}$$

$$f_k(x) = \max_{i=0,\ldots,k-1} [f(\bar{x}_i) - L|x - \bar{x}_i|]; \tag{1.8}$$

$$x_k = \arg\min_{i=\overline{0,k}} f(\bar{x}_i), \quad k = 0, 1, \ldots\ . \tag{1.9}$$

Таким образом, на каждой итерации метода ломаных строится функция $f_k(x)$, находится точка ее минимума $\bar{x}_k$ и затем по правилу (1.9) вычисляется очередное приближение к решению задачи (1.1). Условием остановки метода (1.6)–(1.9) является условие

$$f(x_k) - f_k(\bar{x}_k) \le \varepsilon, \tag{1.10}$$

где $\varepsilon > 0$ – заданная точность решения задачи (1.1). При выполнении условия (1.10) $f(x_k)$ отличается от f_* не более чем на ε. Название метода связано с его геометрической интерпретацией – функция $f_k(x)$ является кусочно-линейной, а ее график представляет собой непрерывную ломаную линию, состоящую из отрезков прямых с угловыми коэффициентами $+L$ или $-L$. Точка $\bar{x}_k$ является одной из нижних вершин ломаной $f_k(x)$ и может быть легко вычислена простым перебором этих вершин.

3. Методы минимизации выпуклых функций. В основе этих методов лежит следующее свойство выпуклых функций: если точки $x, y \in$

$[a,b],\ a < x < y < b$, то

$$x^* \in \begin{cases} [a,y]\,, & \text{если} \ \ f(x) \le f(y) \\ [x,b]\,, & \text{если} \ \ f(x) \ge f(y). \end{cases} \tag{1.11}$$

(Строго говоря, из (1.11) следует, что если $f(x) = f(y)$, то $x^* \in [x,y]$, однако при численных расчетах точное равенство выполняется крайне редко, поэтому мы не будем выделять данный случай и ограничимся включением (1.11).)

Отрезок, содержащий точку минимума, называется ее отрезком локализации. Из (1.11) следует, что для уменьшения отрезка локализации выпуклой функции достаточно сравнить между собой ее значения в двух внутренних точках этого отрезка. Отсюда вытекает следующий алгоритм последовательного сжатия отрезка локализации: положим $[a_0,b_0] = [a,b]$ и для $k = 0,1,\ldots$ построим отрезки $[a_{k+1},b_{k+1}]$ по правилу

$$[a_{k+1},b_{k+1}] = \begin{cases} [a_k,y_k]\,, & \text{если} \ \ f(x_k) < f(y_k) \\ [x_k,b_k]\,, & \text{если} \ \ f(x_k) \ge f(y_k). \end{cases}$$

Здесь $x_k < y_k$ – внутренние, вообще говоря, произвольные точки отрезка $[a_k,b_k]$. Процедуру сжатия отрезка локализации можно оборвать при выполнении условия $b_k - a_k \le \varepsilon$, где $\varepsilon > 0$ – заданная точность решения задачи (1.1), а в качестве приближенного решения задачи взять произвольную точку $\widehat{x}$ последнего отрезка локализации. Рассмотрим два варианта описанного алгоритма.

Метод деления отрезка пополам. В этом методе

$$x_k = \frac{1}{2}[a_k + b_k - \delta], \quad y_k = x_k + \delta,$$

где $\delta > 0$ – достаточно малое число, ограниченное снизу погрешностью представления чисел в компьютере, а сверху - заданной точностью решения задачи (1.1). Таким образом, точки x_k и y_k отстоят от середины отрезка на величину $\delta/2$, а сам отрезок локализации на каждой итерации уменьшается почти в два раза:

$$b_{k+1} - a_{k+1} = \frac{1}{2}(b_k - a_k) + \frac{\delta}{2}.$$

Метод золотого сечения. Здесь в качестве точек x_k и y_k на каждой итерации выбираются точки, задающие золотое сечение отрезка $[a_k,b_k]$.

Напомним, что *золотым сечением отрезка* называется деление его на две неравные части так, что отношение длины всего отрезка к длине большей части равно отношению длины большей части к длине меньшей:

$$x_k = a_k + \frac{3-\sqrt{5}}{2}(b_k - a_k) \simeq a_k + 0,382(b_k - a_k), \tag{1.12}$$

$$y_k = a_k + \frac{\sqrt{5}-1}{2}(b_k - a_k) \simeq a_k + 0,618(b_k - a_k). \tag{1.13}$$

Замечательным свойством золотого сечения является то, что точка x_k является одной из точек золотого сечения отрезка $[a_k, y_k]$, а точка y_k – одной из точек золотого сечения отрезка $[x_k, b_k]$. Это означает, что в отрезке $[a_{k+1}, b_{k+1}]$ уже имеется одна из точек золотого сечения с вычисленным в ней значением функции $f(x)$. Следовательно, на всех итерациях, кроме начальной, для сжатия отрезка локализации достаточно только одного вычисления значения функции.

4. Минимизация выпуклых дифференцируемых функций. Пусть функция $f(x)$ выпукла и непрерывно дифференцируема на $[a, b]$. Тогда (см. упражнение 1.14) для любой точки $x \in [a, b]$

$$x^* \in \begin{cases} [a, x], & \text{если } f'(x) \geq 0 \\ [x, b], & \text{если } f' \leq 0. \end{cases}$$

Случаи $x = a$ и $x = b$ здесь не исключаются. То есть, если $f'(a) \geq 0$, то $x^* = a$, а если $f'(b) \leq 0$, то $x^* = b$. Ну и, конечно, если $f'(x) = 0$, то $x^* = x$. Отсюда с очевидностью вытекает следующий алгоритм сжатия отрезка локализации функции $f(x)$:

1) если $f'(a) \geq 0$, то остановиться: $x^* = a$;

2) если $f'(b) \leq 0$, то остановиться: $x^* = b$;

3) если $f'(a) < 0$ и $f'(b) > 0$, то положить $[a_0, b_0] = [a, b]$;

4) для $k = 0, 1, \dots$ выбрать точку $x_k \in \operatorname{int}[a_k, b_k]$ и проверить условие $f'(x_k) = 0$; если это условие выполняется, то остановится: $x^* = x_k$, иначе положить

$$[a_{k+1}, b_{k+1}] = \begin{cases} [a_k, x_k], & \text{если } f'(x_k) > 0 \\ [x_k, b_k], & \text{если } f'(x_k) < 0. \end{cases}$$

На каждой итерации этого алгоритма необходимо также проверять условие $b_k - a_k \le \varepsilon$, где $\varepsilon > 0$ – заданная точность решения задачи (1.1). Впрочем, можно ограничиться и одним условием остановки: $|f'(x_k)| \le eps$ (точности ε и eps не обязаны быть равны).

В качестве точки x_k на каждой итерации удобно выбирать середину отрезка $[a_k, b_k]$ (метод деления отрезка пополам с вычислением производной) или точку пересечения касательных к функции $f(x)$ в точках a_k и b_k (метод касательных). Нетрудно убедиться, что в методе касательных

$$x_k = a_k + \frac{f(a_k) - f(b_k) + f'(b_k)[b_k - a_k]}{f'(b_k) - f'(a_k)}. \tag{1.14}$$

Практика показывает, что метод касательных во многих случаях сходится значительно быстрей метода деления отрезка пополам, но весьма чувствителен к погрешностям в вычислении производной $f'(x)$. Поэтому метод касательных можно применять только в том случае, если производные вычисляются точно, иначе лучше использовать метод деления отрезка пополам, для которого точность вычисления производных не существенна – важен лишь их знак.

5. Методы полиномиальной аппроксимации. В основе методов, рассматриваемых в этом пункте, лежит простая идея – аппроксимировать функцию $f(x)$ некоторым многочленом, найти его минимум на отрезке $[a, b]$ и точку минимума взять в качестве приближения к решению задачи (1.1). Так как нахождение минимума многочлена не составляет труда в том случае, когда его степень не превосходит трех, то в качестве интерполяционного полинома выбирают полином второй или третьей степени. Точность аппроксимации существенно зависит от близости точек, по которым строится аппроксимирующий многочлен, поэтому методы полиномиальной аппроксимации обычно применяются на заключительном этапе поиска минимума – после того, как каким-нибудь другим методом найден достаточно малый отрезок локализации.

Метод квадратичной аппроксимации (метод парабол). Формально этот метод состоит в следующем. Пусть значения функции $f(x)$ в точках $x_1 < x_2 < x_3$ из отрезка $[a, b]$ равны, соответственно, f_1, f_2, f_3. Через точки плоскости (x_1, f_1), (x_2, f_2) и (x_3, f_3) проведем параболу $\psi(x) = c_0(x - x_2)^2 + c_1(x - x_2)^2 + c_2$. Коэффициенты c_0, c_1 и c_2 легко определяются из соотношений $\psi(x_1) = f_1$, $\psi(x_2) = f_2$, $\psi(x_3) = f_3$. Точку минимума параболы $\psi(x)$ обозначим через $\bar{x}$. Эту точку можно использовать либо в качестве приближения к решению задачи (1.1), либо для построения следующего приближения, например, в соответствии со следующим правилом.

Пусть

$$f_{\min} = \min\{f_1, f_2, f_3\},$$

$$x_{\min} = \arg\min\{f_1, f_2, f_3\}.$$

Если разности $f_{\min} - f(\bar{x})$ и $x_{\min} - \bar{x}$ достаточно малы, то в качестве приближенного решения задачи (1.1) выбирается та из точек $x_{\min}$ или $\bar{x}$, в которой меньше значение функции $f(x)$. В противном случае берется наилучшая из точек $x_{\min}$ или $\bar{x}$ и две точки по обе стороны от нее. Через эти три точки проводится новая парабола, находится ее минимум и т. д.

Метод кубической аппроксимации. Предположим, что функция $f(x)$ выпукла и непрерывно дифференцируема на отрезке $[a, b]$, $f'(a) < 0$, $f'(b) > 0$. Аппроксимируем функцию $f(x)$ многочленом третьей степени

$$\psi(x) = c_0(x-a)^3 + c_1(x-a)^2 + c_2(x-a) + c_3,$$

где коэффициенты c_0, c_1, c_2 и c_3 определяются из условий

$$\psi(a) = f(a),\ \psi(b) = f(b),\ \psi'(a) = f'(a),\ \psi'(b) = f'(b).$$

Найдем точку $\bar{x} = \arg\min\limits_{[a,b]} \psi(x)$. Легко проверить, что

$$\bar{x} = a + \alpha(b-a),$$

где

$$\alpha = \frac{z + \omega - f'(a)}{f'(b) - f'(a) + 2\omega}, \qquad z = 3\,\frac{f(a) - f(b)}{b-a} + f'(a) + f'(b),$$

$$\omega = \sqrt{z^2 - f'(a)\cdot f'(b)}.$$

Точка $\bar{x}$ используется для сжатия отрезка локализации, также как и в методе деления отрезка пополам с вычислением производной: если $f'(\bar{x}) < 0$, то новым отрезком локализации является отрезок $[x, b]$, если $f'(\bar{x}) > 0$ – отрезок $[a, x]$.

Задачи и упражнения

1.1. Пусть функция $f(x) \in C_1[a,b]$. Доказать, что в этом случае она удовлетворяет условию Липшица на $[a,b]$ с константой $L = \max_{t\in[a,b]} |f'(x)|$.

1.2. Пусть функция $f(x)$ выпукла и непрерывно дифференцируема на отрезке $[a,b]$. Доказать, что тогда она удовлетворяет условию Липшица на $[a,b]$ с константой $L = \max\{|f'(a)|, |f'(b)|\}$.

1.3. Пусть функция $f(x) = h(x) + g(x)$, где функции $h(x)$ и $g(x)$ удовлетворяют на отрезке $[a,b]$ условию Липшица с константами L_h и L_g, соответственно. Доказать, что тогда $f(x)$ удовлетворяет на $[a,b]$ условию Липшица с константой $L \le L_h + L_g$.

1.4. Пусть функция $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$ — многочлен n-ой степени. Обозначим

$$g(x) = \sum_{\substack{0\le i\le n:\\ a_i>0}} a_i x^i, \quad h(x) = \sum_{\substack{0\le i\le n:\\ a_i<0}} |a_i| x^i.$$

Очевидно, что $f(x) = g(x) - h(x)$. Доказать, что если левая граница отрезка $[a,b]$, $a \ge 0$, то $f(x)$ удовлетворяет на $[a,b]$ условию Липшица с константой

$$L \le \max\{|g'(b) - h'(a)|, |g'(a) - h'(b)|\}.$$

1.5. Доказать, что если в методе равномерного перебора шаг $h \le 2\varepsilon/L$, где L — константа Липшица из (1.14), то справедлива оценка (1.5).

1.6. Доказать оценку (1.5) для метода адаптивного перебора.

1.7. С помощью методов равномерного и адаптивного перебора решить следующие задачи:

$$f(x) = e^{-x} \to \min, \quad x \in [0,3];$$

$$f(x) = -x^4 + \frac{1}{3}x^3 + x^2 + x \to \min, \quad x \in [0,2];$$

$$f(x) = \max\{\sin x, \cos x\} \to \min, \quad x \in [0, 2\pi].$$

Точность ε взять равной 10^{-2}. В каждом случае подсчитать потребовавшееся для решения задачи количество вычислений значений функции (узлов сетки). Объяснить полученные результаты.

1.8. Сделайте графически 5–6 итераций метода ломаных в задаче

$$f(x) = \min\{\sin x, \cos x\} \to \min, \quad x \in [0, 2\pi].$$

1.9. Докажите, что для выпуклых на отрезке $[a, b]$ функций выполняется включение (1.11).

1.10. Выведите формулы (1.12) и (1.13). Покажите, что в методе золотого сечения длина отрезка локализации $b_k - a_k = (b - a)/p^k$, где $p = \dfrac{1 + \sqrt{5}}{2}$.

1.11. Решите задачу

$$f(x) = x^3 - 2x - 5 \to \min, \quad x \in [0, 1]$$

с помощью методов деления отрезка пополам и золотого сечения. Точность решения задачи ε взять равной $10^{-2}, 10^{-3}, 10^{-4}$. В каждом случае подсчитайте потребовавшееся для решения задачи количество вычислений значений функции.

1.12. Функция $f(x)$ называется унимодальной на отрезке $[a, b]$, если она непрерывна на этом отрезке и существуют числа α, β,
$a \le \alpha \le \beta \le b$ такие, что:

a) $f(x)$ строго монотонно убывает при $a \le x \le \alpha$ (если $a < \alpha$);

b) $f(x)$ строго монотонно возрастает при $\beta \le x \le b$ (если $\beta < b$);

c) $f(x) = f_* = \min\limits_{[a,b]} f(x)$ при $\alpha \le x \le \beta$, так что

$$[\alpha, \beta] = \arg\min_{[a,b]} f(x).$$

Случаи, когда один или два из отрезков $[a, \alpha]$, $[\alpha, \beta]$, $[\beta, b]$ вырождаются в точку, здесь не исключаются. В частности, если $\alpha = \beta$, то $f(x)$ называется строго унимодальной на отрезке $[a, b]$. Докажите, что если функция $f(x)$ унимодальна на отрезке $[a, b]$, точки $x, y \in [a, b]$, $a < x < y < b$, то

$$x^* \in \begin{cases} [a, y], & \text{если } f(x) \le f(y), \\ [x, b], & \text{если } f(x) \ge f(y). \end{cases}$$

1.13. Докажите, что выпуклая непрерывная на отрезке $[a, b]$ функция $f(x)$ унимодальна на этом отрезке. Строго выпуклая на $[a, b]$ функция строго унимодальна на $[a, b]$.

1.14. Пусть функция $f(x)$ выпукла и непрерывно дифференцируема на отрезке $[a,b]$. Тогда при любом $x \in [a,b]$

$$x^* \in \begin{cases} [a,x]\,, & \text{если } f'(x) \geq 0, \\ [x,b]\,, & \text{если } f'(x) \leq 0. \end{cases}$$

Докажите.

1.15. Выведите формулу (1.14). Докажите, что точка
$x_k \in \operatorname{int}[a_k, b_k]$.

1.16. Рассмотрите метод парабол. Сформулируйте условия, при которых точка минимума параболы $\psi(x)$ принадлежит отрезку $[x_1, x_3]$.

1.17. Решите следующие задачи:

a) $f(x) = \min\{|x^2 - 2|,\ 1 - x/4\} \to \min, \quad x \in [-2, 2]$;

b) $f(x) = \max\{|x^2 - 2|,\ 1 - x/4\} \to \min, \quad x \in [-2, 2]$;

c) $f(x) = 3x^4 + (x-1)^2 \to \min, \quad x \in [0, 4]$;

d) $f(x) = 4x \sin x \to \min, \quad x \in [0, 2\pi]$;

e) $f(x) = 2(x-3)^2 + e^{0,5x^2} \to \min, \quad x \in [0, 3]$.

1.18. Одним из способов решения уравнения $f(x) = 0$ является сведение этой задачи к задаче минимизации функции $\varphi(x) = f^2(x)$. Ясно, что если x^* — точка минимума функции $\varphi(x)$, то, если $\varphi(x^*) = 0$, x^* — корень уравнения $f(x)$. В противном случае уравнение $f(x) = 0$ корней не имеет. Используя этот прием, решите уравнения

$$4x^3 - 3x^2 + 2x + e^{x/2} = 0,$$

$$3000 - 100x^2 - 4x^5 - 6x^6 = 0,$$

$$\sin x - e^x + 1 = 0.$$

Список литературы

1. А к у л и ч И. Л. Математическое программирование в примерах и задачах / И. Л. Акулич. – СПб. : Лань, 2009. – 347 с.

2. А л е к с е е в В. М. Сборник задач по оптимизации. Теория, примеры, задачи / В. М. Алексеев, Э. М. Галеев, В. М. Тихомиров. – М. : Физматлит, 2007. – 256 с.

3. А ш м а н о в С. А. Теория оптимизации в задачах и упражнениях / С. А. Ашманов, А. В. Тимохов. – М. : Наука, 1991. – 448 с.

4. Б а з а р а М. Нелинейное программирование. Теория и алгоритмы / М. Базара, К. Шетти. – М. : Мир, 1982. – 583 с.

5. В а с и л ь е в О. В. Лекции по методам оптимизации / О. В. Васильев. – Иркутск: Изд-во Иркут. ун-та, 1994. – 344 с.

6. В а с и л ь е в О. В. Методы оптимизации в задачах и упражнениях / О. В. Васильев, А. В. Аргучинцев. – М. : Физматлит, 1999. – 208 с.

7. В а с и л ь е в Ф. П. Методы оптимизации / Ф. П. Васильев. – М. : Факториал Пресс, 2002. – 824 с.

8. В а с и л ь е в Ф. П. Численные методы решения экстремальных задач / Ф. П. Васильев. – М. : Наука, 1988. – 552 с.

9. Г а л е е в Э. М. Оптимизация: теория, примеры, задачи / Э. М. Галеев. – М. : Книжный дом «ЛИБРОКОМ», 2010. – 336 с.

10. Г а л е е в Э. М. Краткий курс теории экстремальных задач / Э. М. Галеев, В. М. Тихомиров. – М. : Изд-во Моск. ун-та, 1989. – 204 с.

11. Г а л е е в Э. М. Оптимизация: теория, примеры, задачи / Э. М. Галеев, В. М. Тихомиров. – М. : Эдиториал УРСС, 2000. – 320 с.

12. Д е м и д о в и ч Б. П. Сборник задач и упражнений по математическому анализу / Б. П. Демидович. – М. : Наука, 1990. – 624 с.

13. К а р м а н о в В. Г. Математическое программирование / В. Г. Карманов. – М. : Физматлит, 2001. – 264 с.

14. К о р н е е н к о В. П. Методы оптимизации: учебник / В. П. Корнеенко. – М. : Высш. шк., 2007. – 664 с.

15. Н и н у л А. С. Оптимизация целевых функций: Аналитика. Численные методы. Планирование эксперимента / А. С. Нинул. – М. : Физматлит, 2009. – 336 с.

16. С у х а р е в А. Г. Курс методов оптимизации / А. Г. Сухарев,

А. В. Тимохов, В. В. Федоров. – М. : Физматлит, 2008. – 368 с.

17. Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления: Т. 1 / Г. М. Фихтенгольц. – М. : Наука, 1969. – 608 с.

18. Шилов Г. Е. Математический анализ (функции одного переменного): Ч. 1–2 / Г. Е. Шилов. – М. : Наука, 1969. – 528 с.

19. Шилов Г. Е. Математический анализ (функции нескольких вещественных переменных): Ч. 1–2 / Г. Е. Шилов. – М. : Наука, 1972. – 624 с.