Ю. П. Иванилов, Н. Н. Моисеев, Е. М. Столярова

Методы оптимизации

Н Н МОИСЕЕВ
Ю П. ИВАНИЛОВ
Е. М. СТОЛЯРОВА

МЕТОДЫ ОПТИМИЗАЦИИ

Допущено Министерством высшего и среднего специального образования СССР в качестве учебного пособия для студентов вузов, обучающихся по специальности «Прикладная математика»

МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1978

22 193
М 74
УДК 519.6

Методы оптимизации. Моисеев Н Н., Иванилов Ю П, Столярова Е М. «Наука», Главная редакция физико-математической литературы, М., 1978, 352 стр

Настоящая книга предназначена в качестве учебного пособия для студентов факультетов прикладной математики, факультетов по переподготовке специалистов в области использования вычислительной техники, а также для учащихся математических техникумов В ней излагается методика составления оптимизационных моделей в прикладных задачах, общие принципы линейного, нелинейного и динамического программирования Приводится обзор основных методов численного анализа для задач отыскания экстремумов функций.

М $\frac{20204-170}{053(02)-78}$ БЗ 30-89—78

ОГЛАВЛЕНИЕ

Предисловие 5

Глава I. **Задача отыскания экстремума функций многих переменных** 12

Введение 12
§ 1. Функция одной переменной. Условия экстремума . . . 13
§ 2. Функция многих переменных 22
§ 3. Относительный экстремум. Метод множителей Лагранжа 26

Глава II. **Численные методы отыскания безусловного экстремума** 40

Введение 40
§ 1. Градиентные методы 42
§ 2. Метод Ньютона 55
§ 3. Метод сопряженных градиентов 73
§ 4. Одномерный оптимальный поиск 84

Глава III. **Линейное программирование** 95

Введение 95
§ 1. О постановках задачи линейного программирования и ее приложениях 95
§ 2. Геометрическая интерпретация задач линейного программирования 102
§ 3. Некоторые свойства задач линейного программирования 109
§ 4. Симплекс-метод 115
§ 5. Двойственные задачи и методы 131

Глава IV. **Теория экстремума в нелинейных задачах с ограничениями** 151

Введение 151
§ 1 Выпуклые множества и конусы 152
§ 2. Выпуклые функции и опорные функционалы 165
§ 3. Условия экстремума в задачах нелинейного программирования 179
§ 4. Дискретный принцип максимума 204

Глава V. **Численные методы нелинейного программирования** . . . 217
Введение 217
§ 1. Методы спуска 217
§ 2. Методы штрафных функций 227

Глава VI. **Методы оптимизации, основанные на последовательном анализе вариантов** 254
Введение 254
§ 1. Аддитивные задачи 255
§ 2. Дискретные управляемые системы 271
§ 3. Задача о коммивояжере и ее обобщения 285

Приложение. **Диалоговая система оптимизации** 301
§ 1. Принципы построения диалоговых систем 301
§ 2. Библиотека программ решения задач безусловной минимизации 309
§ 3 Библиотека программ решения задач нелинейного программирования 313
§ 4. Примеры работы с ДИСО 324
§ 5. Некоторые подходы к проблеме создания управляющих программ 342

Литература 347
Предметный указатель 348

Глава I

ЗАДАЧА ОТЫСКАНИЯ ЭКСТРЕМУМА ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ

Введение

Одной из важных задач анализа является задача отыскания экстремума (наибольшего или наименьшего значения) скалярной функции $f(x)$ n-мерного векторного аргумента x при некоторых ограничениях. Эту задачу мы будем записывать следующим образом:

$$\min f(x), \tag{0.1}$$

$$x \in X. \tag{0.2}$$

Здесь X — некоторое подмножество n-мерного евклидова пространства E_n Впредь будем называть X *допустимым множеством* задачи (0.1)—(0.2), а точки, принадлежащие X, — ее *допустимыми точками*. Заметим, что задачу максимизации функции $f(x)$ тоже можно записать в виде (0.1) — (0.2), заменив $f(x)$ на $\tilde{f}(x) = -f(x)$.

В этой главе будут последовательно рассмотрены задача *безусловной минимизации* функции одной переменной ($X = E_1$), задача нахождения безусловного экстремума функции нескольких переменных ($X = E_n$) и, наконец, задача на *относительный экстремум*, т. е. задача минимизации функции нескольких переменных при наличии ограничений типа равенств, когда X — множество решений уравнения

$$g(x) = 0,$$

где $g(x)$ есть m-мерная вектор-функция, $m < n$.

Задача (0.1) — (0.2) является классической и рассматривается во всех курсах анализа. Теория решения таких задач развивалась еще в трудах Эйлера, Лагранжа, Бернулли, Лейбница. Она не потеряла своего значения и в настоящее время, несмотря на то, что с тех пор разработаны более общие методы, включающие классические, как частный случай. Классическая теория содержит значительную часть идей, лежащих в основе современных методов оптимизации. Поэтому изложение этих методов мы начнем с известных фактов анализа.

§ 1. Функция одной переменной. Условия экстремума

1. Предварительные рассмотрения. Возьмем задачу минимизации функции одной переменной $f(x)$ на множестве $X \subset E_1$:

$$\min f(x), \tag{1.1}$$

$$x \in X \subset E_1. \tag{1.2}$$

Ее формулировка нуждается в некоторых уточнениях. Поясним это на следующих примерах.

Пример 1.1. Пусть множество X состоит из четырех точек, значения функции $f(x)$ в которых заданы таблицей:

f	0,9	0,4	1,3	0,8
x	1,2	2,5	2,8	4,7

Перебрав эти точки и сравнив между собой значения функции в них, легко убедиться, что минимум достигается в точке $\hat{x} = 2{,}5$ и его величина есть

$$\hat{f} = f(\hat{x}) = 0{,}4.$$

Пример 1.2. Изменим несколько функцию примера 1.1 и зададим ее следующей таблицей:

f	0,9	0,4	1,3	0,4
x	1,2	2,5	2,8	4,7

Величина минимума в данном примере остается прежней, $\hat{f} = 0{,}4$, но достигается он теперь не в единственной точке, а в двух $\hat{x}' = 2{,}5$ и $\hat{x}'' = 4{,}7$. Поэтому целесообразно говорить о множестве точек минимума.

Приведенные примеры показывают, что необходимо четко сформулировать понятие решения задачи (1.1) — (1.2).

Определение 1.1. Точка $\hat{x}$ доставляет *глобальный минимум* функции $f(x)$ на множестве X, если $\hat{x} \in X$ и

$$f(\hat{x}) \leqslant f(x) \tag{1.3}$$

для всех $x \in X$.

Определение 1.2. Точка $\hat{x}$ называется точкой *строгого глобального минимума* $f(x)$ на множестве X, если $\hat{x} \in X$ и

$$f(\hat{x}) < f(x) \tag{1.4}$$

для всех $x \in X$, $x \neq \hat{x}$.

Под решением задачи (1.1) — (1.2) часто понимают любую точку, удовлетворяющую условию (1.3). Если

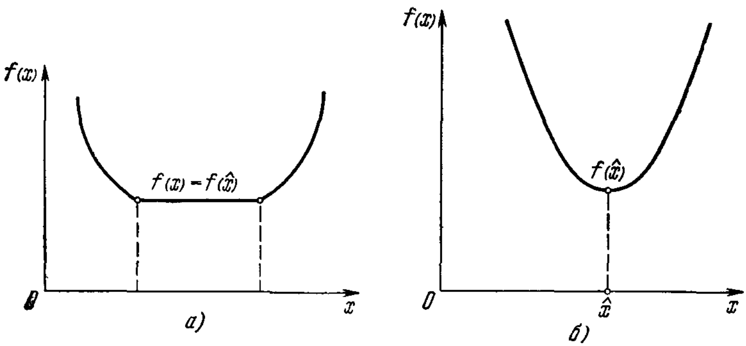

Рис. 1.1.

такая точка единственна, то она и является точкой строгого минимума. Если равенство в формуле (1.3) возможно при $x \neq \hat{x}$, то говорят, что реализуется нестрогий минимум, а под решением в этом случае понимают множество

$$\hat{X} = \{x \in X : f(x) = f(\hat{x})\}.$$

Пример 1.3. График функции, имеющей нестрогий минимум (определение (1.3)), может содержать горизонтальный участок в окрестности точки минимума (рис. 1.1*а*). Для функции со строгим минимумом это исключено (рис. 1.1*б*).

Наряду с задачей определения глобального минимума функции возникает задача поиска локального минимума. Дадим соответствующее определение

Определение 1.3. Точка $\hat{x} \in X$ доставляет *локальный минимум* функции $f(x)$ на множестве X, если при некотором достаточно малом $\varepsilon > 0$ для всех $x \neq \hat{x}$, $x \in X$, удовлетворяющих условию

$$|x - \hat{x}| \leqslant \varepsilon,$$

выполнено неравенство

$$f(\hat{x}) \leqslant f(x). \tag{1.5}$$

Если неравенство (1.5) — строгое, то точку $\hat{x}$ называют точкой *строгого локального минимума* функции $f(x)$. Понятно, что глобальный минимум является и локальным, но не наоборот.

Все определения для максимума функции получаются заменой в выражениях (1.3), (1.4), (1.5) знака неравенства на обратный.

Пример 1.4 Условию (1.5) могут удовлетворять одновременно как точки локального минимума, так и точки локального максимума функции $f(x)$ (рис. 1.2).

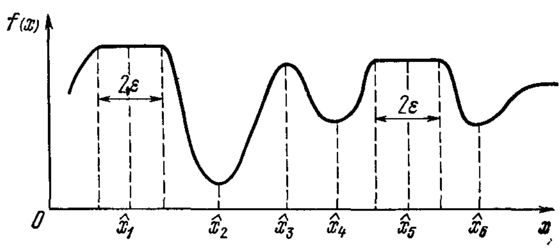

Рис. 1.2

Для функции, изображенной на рис. 1.2, точки $\hat{x}_1$, $\hat{x}_3$, $\hat{x}_5$ являются точками локального максимума, в точках $\hat{x}_4$, $\hat{x}_6$ реализуются локальные минимумы, а точка $\hat{x}_2$ — точка глобального минимума.

Обсудим теперь вопрос о существовании решения задачи (1.1) — (1.2). Рассмотрим некоторые варианты задания множества X. Естественно считать, что это множество непусто и не состоит из единственной точки. Только в этом случае задача минимизации содержательна, так как есть возможность выбора. Если множество X содержит конечное число точек, то решение задачи (1.1) — (1.2)

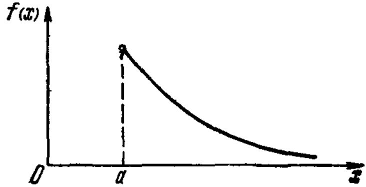

Рис. 1.3.

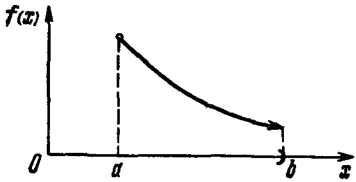

Рис. 1.4.

существует: можно перебрать все точки из X и выбрать среди них точку $\hat{x}$, удовлетворяющую условию (1.3). В случае, когда множество X содержит бесконечное число точек, задача минимизации $f(x)$ на X может не иметь решения.

Пример 1.5. Пусть

$$X = \{x : x \geqslant a\}$$

(рис. 1.3), а функция $f(x)$ монотонно убывает (например, $f(x) = e^{-x}$). Очевидно, что здесь точки $\hat{x}$, удовлетворяющей условию (1.3), не существует.

Пример 1.6. Пусть множество X задано в виде

$$X = \{x . a \leqslant x < b\},$$

т. е. не замкнуто: точка $x=b$ не содержится во множестве X (на рис. 1.4 это отмечено дужкой в точке b). Если функция $f(x)$ при $x \to b$ монотонно убывает, нижняя грань ее на множестве X не достигается (на рис. 1.4 это показано стрелкой) и точки $\hat{x}$, удовлетворяющей условию (1.3), не существует.

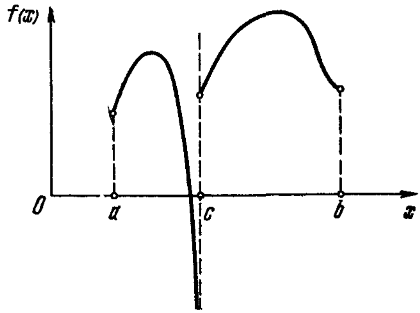

Рис. 1.5

Таким образом, рассмотренные примеры показывают, что в случаях, когда множество X не замкнуто, задача (1.1) — (1.2) может не иметь решения. Разумеется, это не означает, что в подобных случаях решения существовать не может. В самом деле, достаточно взять в примере 1.5 множество $X=\{x : x \leqslant a\}$, а в примере 1.6 $X=\{x : a < x \leqslant b\}$, $f(b) = \lim\limits_{x \to -b} f(x)$, чтобы соответствующие задачи были разрешимы.

Пример 1.7 Пусть

$$X=\{x : a \leqslant x \leqslant b\},$$

а $f(x)$ — неограниченная снизу функция, имеющая вертикальную асимптоту при $x=c$ (рис. 1.5). Здесь не существует точки, удовлетворяющей условию (1.3), т. е. неограниченная снизу на замкнутом ограниченном множестве X функция $f(x)$ глобального минимума на X не имеет.

Сформулируем теперь теорему Вейерштрасса, выделяющую широкий класс задач минимизации (максимизации), заведомо имеющих решение.

Теорема 1.1 *Задача минимизации непрерывной функции $f(x)$ на замкнутом ограниченном множестве X разрешима, т е. непрерывная функция $f(x)$ достигает на замкнутом ограниченном множестве своего минимума (во внутренней или граничной точке)*

Доказательство этой теоремы можно найти в любом курсе анализа

После того как мы сформулировали задачу минимизации функции и обсудили некоторые вопросы, связанные с существованием ее решения, необходимо указать признаки, с помощью которых можно найти точки $\hat{x}$, являющиеся решением задачи (1.1) — (1.2), или проверить, доставляет ли некоторая найденная точка $x \in X$ минимум функции $f(x)$. Эти признаки называются необходимыми условиями и достаточными условиями экстремума.

2. Необходимое условие первого порядка. Выясним условия, которые должны выполняться в точках локального экстремума функции. Рассмотрим те случаи, когда множество X представляет собой вещественную ось. Рассуждения сохраняют силу и для задач, в которых множество X не совпадает с E_1, но открыто, т. е. состоит только из внутренних точек, либо экстремум достигается в его внутренней точке. Изучение случаев, когда экстремум реализуется на границе множества X, требует, как мы увидим ниже, специальных методов.

При выводе условий экстремума будем предполагать, что функция $f(x)$ имеет в окрестности исследуемой точки $\hat{x}$ непрерывные производные до второго порядка включительно.

Теорема 1.2. *Для того чтобы функция $f(x)$, определенная на вещественной оси, имела безусловный локальный экстремум в точке $\hat{x}$, необходимо, чтобы выполнялось условие*

$$\left.\frac{df}{dx}\right|_{x=\hat{x}} = 0. \tag{1.6}$$

Доказательство. Пусть точка $\hat{x}$ доставляет локальный безусловный минимум функции $f(x)$ (случай максимума рассматривается аналогично). Тогда, согласно определению 1.3, найдется такая окрестность этой точки радиуса ε, что для всех ξ, удовлетворяющих неравенству $|\xi| \leqslant \varepsilon$,

$$f(\hat{x}+\xi) - f(\hat{x}) \geqslant 0. \tag{1.7}$$

По формуле Тейлора имеем

$$f(\hat{x}+\xi) = f(\hat{x}) + \xi f'(\hat{x}) + O(\xi^2).$$

Предположим, что $f'(\hat{x}) \neq 0$, и выберем $\xi = -f'(\hat{x})\rho$, где $\rho > 0$ — любое малое число такое, что $|f'(\hat{x})|\rho < \varepsilon$.

Тогда получим

$$\frac{f(\hat{x}+\xi)-f(\hat{x})}{\rho}=-(f'(\hat{x}))^2+\frac{O(\rho^2)}{\rho}.$$

Так как

$$\lim_{\rho\to 0}\frac{O(\rho^2)}{\rho}=0,$$

то найдется такое малое ρ^*, что второе слагаемое в правой части последнего выражения будет по абсолютной величине меньше первого, т. е. $f(\hat{x}+\xi)-f(\hat{x})<0$, что противоречит предположению (1.7). Итак,

$$f'(\hat{x})=0,$$

что и требовалось доказать.

Рассмотрим пример функции одной переменной, определенной на всей вещественной оси и удовлетворяющей

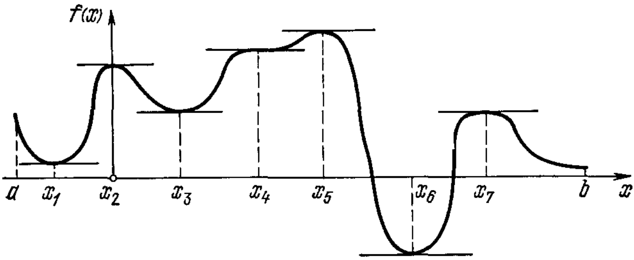

Рис. 1.6.

всем указанным выше условиям (рис. 1.6). Легко видеть, что наименьшего значения функция $f(x)$ достигает в точке x_6, и если, например, известно, что вне интервала $[a, b]$ функция $f(x)$ возрастает, то точка x_6 является точкой ее абсолютного глобального минимума на $X=E_1$. Заметим, что слева от этой точки функция $f(x)$ убывает, а справа от нее — возрастает. В самой же точке x_6 убывание функции приостанавливается. Поэтому она и называется *стационарной*. Вообще, стационарными называются все точки, удовлетворяющие условию (1.6). На рис. 1.6 — это точки, имеющие горизонтальную касательную. Однако, как видно из рис 1.6, не все из стационарных точек будут точками локального минимума (только точки x_1, x_3, x_6) Например, точки x_2, x_5, x_7 являются точками локального максимума, а точка x_4 — точкой перегиба.

Таким образом, условие (1.6) выделяет стационарные точки, но не определяет их характера: оно одинаково для точек максимума, минимума и перегиба. Чтобы избежать сравнения значений функции $f(x)$ во всех стационарных точках с целью определения глобального минимума (или локальных минимумов), желательно найти некоторые дополнительные условия, которые выполняются только в точках минимума.

3. Необходимые условия второго порядка. Пусть функция $f(x)$ удовлетворяет перечисленным выше условиям. Тогда справедлива

Теорема 1.3. *Для того чтобы функция $f(x)$ имела в стационарной точке $\hat{x}$ безусловный локальный минимум (максимум), необходимо, чтобы ее вторая производная была неотрицательна (неположительна), т. е.*

$$\left.\frac{d^2f}{dx^2}\right|_{x=\hat{x}} \geqslant 0 \qquad \left(\left.\frac{d^2f}{dx^2}\right|_{x=\hat{x}} \leqslant 0\right). \tag{1.8}$$

Доказательство. По теореме 1.2 первая производная в стационарной точке равна нулю. Соответственно, из формулы Тейлора при всех ξ получим

$$f(\hat{x}+\xi)-f(\hat{x})=\frac{1}{2}\xi^2 f''(\hat{x})+o(\xi^2). \tag{1.9}$$

Допустим теперь, что наша теорема неверна, т. е.

$$\left.\frac{d^2f}{dx^2}\right|_{x=\hat{x}} < 0. \tag{1.10}$$

Тогда для достаточно малых ξ второе слагаемое в правой части выражения (1.9) будет по абсолютной величине меньше первого и, следовательно, выполнится неравенство

$$f(\hat{x}+\xi)-f(\hat{x})<0.$$

Это противоречит определению точки $\hat{x}$ как точки локального минимума. Значит,

$$\left.\frac{d^2f}{dx^2}\right|_{x=\hat{x}} \geqslant 0.$$

Теорема доказана.

Условия (1.6), (1.8) называются необходимыми условиями минимума (максимума) второго порядка. Нетрудно показать, что совокупность условий (1.6), (1.8) не является достаточным условием минимума.

Пример 1.8. Пусть $f(x)=x^3$. Тогда точка $\hat{x}=0$ удовлетворяет необходимым условиям минимума второго порядка:

$$\left.\frac{df}{dx}\right|_{x=0}=3x^2|_{x=0}=0,$$
$$\left.\frac{d^2f}{dx^2}\right|_{x=0}=6x|_{x=0}=0.$$

При этом она является точкой перегиба, а не минимума функции $f(x)=x^3$.

4. Достаточные условия. Если функция $f(x)$ дифференцируема достаточное число раз, то можно построить аналогичные (1.6), (1.8) необходимые условия любого порядка. Налагая все более и более жесткие ограничения на выбор экстремальных точек, они, тем не менее, не дают окончательного ответа об их характере. Поэтому нужно иметь еще и достаточные условия экстремума.

Теорема 1.4. *Для того чтобы функция $f(x)$ имела в стационарной точке $\hat{x}$ безусловный локальный минимум (максимум), достаточно, чтобы ее вторая производная была в $\hat{x}$ положительна (отрицательна):*

$$\left.\frac{d^2f}{dx^2}\right|_{x=\hat{x}}>0 \qquad \left(\left.\frac{d^2f}{dx^2}\right|_{x=\hat{x}}<0\right). \tag{1.11}$$

Доказательство. Поскольку точка $\hat{x}$ — стационарная, разложение функции $f(x)$ в ряд Тейлора в окрестности $\hat{x}$ имеет вид

$$f(\hat{x}+\xi)-f(\hat{x})=\frac{1}{2}\xi^2 f''(\hat{x})+o(\xi^2).$$

При $f''(\hat{x})\neq 0$ для достаточно малых ξ знак правой части этого выражения определяется знаком второй производной $f''(\hat{x})$, т. е. из (1.11) вытекает неравенство

$$f(\hat{x}+\xi)-f(\hat{x})>0.$$

Последнее означает, что $\hat{x}$ — точка строгого локального минимума функции $f(x)$, что и требовалось доказать.

Обратимся теперь к случаю, когда в стационарной точке вторая производная функции $f(x)$ обращается в нуль, т. е.

$$\frac{df}{dx}(\hat{x})=0 \quad \text{и} \quad \frac{d^2f}{dx^2}(\hat{x})=0.$$

Как видно из примера 1.8, полученные нами условия экстремума не определяют в этом случае характера стационарной точки. Очевидно, необходимо рассмотреть производные более высоких порядков и снова воспользоваться разложением функции $f(x)$ в ряд Тейлора.

Пусть функция $f(x)$ имеет в окрестности точки $\hat{x}$ непрерывные производные до k-го порядка включительно, и пусть

$$f'(\hat{x}) = f''(\hat{x}) = \ldots = f^{(k-1)}(\hat{x}) = 0, \quad f^{(k)}(\hat{x}) \neq 0.$$

Тогда, согласно формуле Тейлора,

$$f(\hat{x} + \xi) - f(\hat{x}) = \frac{1}{k!}\xi^k f^{(k)}(\hat{x}) + o(\xi^k).$$

Анализируя это выражение, придем к следующей теореме (достаточному условию общего вида).

Теорема 1.5. *Пусть функция $f(x)$, определенная на множестве $X = E_1$, имеет непрерывные производные до k-го порядка включительно, причем в некоторой точке $\hat{x}$*

$$f'(\hat{x}) = f''(\hat{x}) = \ldots = f^{(k-1)}(\hat{x}) = 0, \quad f^{(k)}(\hat{x}) \neq 0.$$

Тогда, если k — четное число, то функция $f(x)$ имеет в точке $\hat{x}$ локальный максимум при $f^{(k)}(\hat{x}) < 0$ и локальный минимум при $f^{(k)}(\hat{x}) > 0$. Если k нечетно, то $f(x)$ не имеет в точке $\hat{x}$ ни максимума, ни минимума.

Заметим, что необходимые условия экстремума — это уравнения относительно неизвестных величин $\hat{x}$. Корни этих уравнений определяют некоторое множество «претендентов» на экстремум — значений переменной x, среди которых только и могут находиться интересующие нас точки $\hat{x}$, доставляющие максимум или минимум функции $f(x)$. Для того чтобы среди этих точек разыскать, например, точки минимума, мы должны еще в каждой точке множества «претендентов» проверить выполнение достаточных условий.

§ 2. Функция многих переменных

1. Необходимое условие экстремума. Снова рассмотрим задачу безусловной минимизации, но будем теперь считать, что $f(x)$ — скалярная функция векторного аргумента размерности n, т. е. $X = E_n$. Если $\hat{x}$ — точка ее безусловного локального экстремума, в $\hat{x}^j$ будет достигаться экстремум функции

$$f(\hat{x}^1, \hat{x}^2, \ldots, \hat{x}^{j-1}, x^j, \hat{x}^{j+1}, \ldots, \hat{x}^n)$$

одной переменной x^j, которая получается из функции $f(x)$, если зафиксировать все переменные, кроме x^j, положив $x^i = \hat{x}^i$ для $i \neq j$. Для функции же одной переменной

$$f(\hat{x}^1, \ldots, \hat{x}^{j-1}, x^j, \hat{x}^{j+1}, \ldots, \hat{x}^n)$$

получена теорема 1.2. Проведя это рассуждение для всех $j = 1, \ldots, n$, приходим к следующей теореме.

Теорема 2.1. *Для того чтобы в точке $\hat{x}$ функция $f(x^1, \ldots, x^n)$ имела безусловный локальный экстремум, необходимо, чтобы все ее частные производные обращались в $\hat{x}$ в нуль*:

$$\left.\frac{\partial f}{\partial x^i}\right|_{x=\hat{x}} = 0, \quad i = 1, 2, \ldots, n. \tag{2.1}$$

Условие стационарности (2.1) мы будем записывать еще в одной из следующих эквивалентных форм:

$$\operatorname{grad} f(\hat{x}) = \nabla f(\hat{x}) = 0, \tag{2.1'}$$

$$f'(\hat{x}) = 0, \tag{2.1''}$$

где $\nabla f(\hat{x}) = f'(\hat{x})$ — n-мерный вектор с компонентами $\frac{\partial f}{\partial x^i}(\hat{x})$, $i = 1, \ldots, n$, который принято называть градиентом функции $f(x)$ в точке $\hat{x}$.

Заметим, что необходимое условие экстремума (2.1) эквивалентно равенству нулю дифференциала функции $f(x)$ в точке $\hat{x}$:

$$df(\hat{x}) = 0.$$

В самом деле, если выполнено условие (2.1), то для любых dx^i, $i = 1, \ldots, n$, имеем

$$df(\hat{x}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(\hat{x})\, dx^i = 0.$$

Справедливо и обратное утверждение, так как из последнего равенства в силу произвольности независимых приращений dx^i, $i = 1, \ldots, n$, следует, что все частные производные в точке $\hat{x}$ равны нулю:

$$\frac{\partial f}{\partial x^i}(\hat{x})=0, \quad i=1, \ldots, n.$$

Условия (2.1) образуют систему n уравнений для определения n компонент вектора $\hat{x}$. Эти уравнения могут иметь различную природу и допускать любое количество решений, в частности, не иметь ни одного. Как и выше, точки $\hat{x}$, являющиеся решениями системы уравнений (2.1), будем называть стационарными, а условие (2.1) — необходимым условием экстремума первого порядка.

2. Необходимое условие второго порядка. Достаточные условия. После того как решение $\hat{x}$ системы уравнений (2.1) будет найдено, необходимо еще определить характер стационарной точки $\hat{x}$. Для этого нужно исследовать поведение функции $f(x)$ в окрестности стационарной точки $\hat{x}$. Снова воспользуемся разложением функции $f(x)$ в ряд Тейлора, предполагая ее дважды непрерывно дифференцируемой по всем переменным $x^1, \ldots, x^n$. Тогда получим

$$f(\hat{x}+\xi)=f(\hat{x})+\frac{1}{2}\sum_{k,i=1}^{n}\frac{\partial^2 f}{\partial x^i\,\partial x^k}(\hat{x})\,\xi^i\xi^k+o(\|\xi\|^2). \tag{2.2}$$

Здесь через $\frac{\partial^2 f}{\partial x^i\,\partial x^k}(\hat{x})$ мы обозначили элементы матрицы вторых производных функции $f(x)$ в стационарной точке $\hat{x}$, а через $\|\xi\|$ — какую-нибудь норму вектора ξ, например, $\|\xi\|=\sqrt{(\xi,\ \xi)}$. Далее матрицу вторых производных мы будем обозначать так:

$$f''(\hat{x})=\begin{bmatrix}\frac{\partial^2 f}{\partial x^1\,\partial x^1}(\hat{x}) & \ldots & \frac{\partial^2 f}{\partial x^1\,\partial x^n}(\hat{x})\\ \cdots & \cdots & \cdots\\ \frac{\partial^2 f}{\partial x^n\,\partial x^1}(\hat{x}) & \ldots & \frac{\partial^2 f}{\partial x^n\,\partial x^n}(\hat{x})\end{bmatrix}. \tag{2.3}$$

Характер стационарной точки $\hat{x}$ функции $f(x)$ связан со знакоопределенностью квадратичной формы

$$\sum_{i,k=1}^{n}\frac{\partial^2 f}{\partial x^i\,\partial x^k}(\hat{x})\,\xi^i\xi^k=(\xi,\ f''(\hat{x})\,\xi). \tag{2.4}$$

Напомним, что квадратичная форма называется неотрицательно определенной в точке $\hat{x}$, если

$$(\xi,\ f''(\hat{x})\,\xi) \geqslant 0, \tag{2.5}$$

и положительно определенной, если

$$(\xi,\ f''(\hat{x})\,\xi) > 0 \tag{2.6}$$

для любых векторов $\xi \neq 0$.

Соответственно, симметричная матрица вторых производных $f''(\hat{x})$ называется неотрицательно определенной в точке $\hat{x}$, если выполнено (2.5), и положительно определенной, если выполнено (2.6). Неположительно определенным и отрицательно определенным квадратичным формам и матрицам соответствуют противоположные знаки в неравенствах (2.5), (2.6).

Таким образом, с учетом разложения (2.2), приходим к следующей формулировке условий второго порядка экстремальности функции $f(x^1, \ldots, x^n)$.

Теорема 2.2. *Для того чтобы дважды непрерывно дифференцируемая функция n переменных $f(x)$ имела в стационарной точке $\hat{x}$ безусловный локальный минимум (максимум), необходимо, чтобы матрица ее вторых производных была неотрицательно (неположительно) определенной, и достаточно, чтобы она была положительно (отрицательно) определенной.*

Проверка знакоопределенности матриц может быть осуществлена, например, с помощью критерия Сильвестра. Согласно этому критерию, необходимым и достаточным условием положительной определенности квадратичной формы (x, Ax), где $A = \{a_{ij}\}$ — симметричная $n \times n$ матрица, является выполнение n неравенств:

$$a_{11} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \ldots, \quad \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix} > 0.$$

Необходимым и достаточным условием отрицательной определенности квадратичной формы (x, Ax) является выполнение цепочки следующих n неравенств:

$$(-1)^n a_{11} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \ldots$$

$$\ldots, (-1)^n \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} > 0.$$

Если квадратичная форма не меняет знака, но обращается в нуль при ненулевых значениях аргумента, то для определения характера стационарной точки $\hat{x}$ требуется исследование производных более высокого порядка.

3. Пример. Проиллюстрируем содержание настоящего параграфа на следующей задаче: определить экстремальные значения функции

$$f(x)=\frac{(x^1)^2}{a}+\frac{(x^2)^2}{b},$$
$$a\neq 0,\quad b\neq 0,\quad X=E_2.$$

Из необходимых условий (2.1) имеем

$$\frac{\partial f}{\partial x^1}=\frac{2x^1}{a}=0,\qquad \frac{\partial f}{\partial x^2}=\frac{2x^2}{b}=0.$$

Поэтому $\hat{x}^1=0$, $\hat{x}^2=0$ — стационарная точка. Коэффициенты квадратичной формы (2.4), вычисленные в ней, равны

$$\frac{\partial^2 f}{(\partial x^1)^2}=\frac{2}{a},\quad \frac{\partial^2 f}{\partial x^1\,\partial x^2}=\frac{\partial^2 f}{\partial x^2\,\partial x^1}=0,\quad \frac{\partial^2 f}{(\partial x^2)^2}=\frac{2}{b}.$$

Тогда, согласно теореме 2.2, имеем следующие случаи:

1) $a>0,\ b>0$ — функция $f(x)$ имеет в точке $\hat{x}=\{0,\ 0\}^T$ *) минимум;

2) $a>0,\ b<0$ \
3) $a<0,\ b>0$ } — экстремума нет;

4) $a<0,\ b<0$ — функция $f(x)$ имеет в точке $\hat{x}=\{0,\ 0\}^T$ максимум.

Отметим, что случаи 1) и 4) соответствуют поверхности, являющейся эллиптическим параболоидом, а случаи 2) и 3) — гиперболическому параболоиду, имеющему стационарную точку типа «седло».

§ 3. Относительный экстремум. Метод множителей Лагранжа

1. Метод исключения. Рассмотрим теперь задачу на относительный экстремум. Как мы видели в § 2, решение задачи об отыскании экстремумов функции n переменных $f(x)$ на всем пространстве E_n может быть сведено с помощью необходимых условий к решению системы урав-

*) Здесь и далее $\{x^1,\ \ldots,\ x^n\}^T$ — вектор-столбец.

нений (2.1), в результате чего определяются стационарные точки функции $f(x)$. Оказывается, что аналогичное сведение возможно и для задачи отыскания экстремумов функции $f(x)$ при наличии ограничений типа равенств

$$g_i(x) = 0, \quad i = 1, 2, \ldots, m. \tag{3.1}$$

Условия (3.1) принято еще называть уравнениями связи.

Уточним, что именно мы будем понимать под решением задачи на относительный экстремум. Напомним (см. введение), что точку x, удовлетворяющую условиям (3.1), мы договорились назвать допустимой.

Определение 3.1. Допустимая точка $\hat{x}$ доставляет *относительный локальный минимум* функции $f(x)$, если можно указать такое число $\varepsilon > 0$, что для всех x, удовлетворяющих уравнениям связи (3.1) и условию $\|x - \hat{x}\| < \varepsilon$, имеет место неравенство

$$f(x) \geqslant f(\hat{x}).$$

Определения строгого относительного локального минимума, а также все определения для относительного максимума получаются по аналогии с приведенными в § 1.

Рассмотрим случай, когда уравнения связи (3.1) могут быть разрешены относительно части переменных. Будем предполагать, что функции $g_i(x)$, $i = 1, \ldots, m$, имеют в окрестности рассматриваемой допустимой точки $\hat{x}$ непрерывные частные производные по всем аргументам до второго порядка включительно и, кроме того, ранг матрицы Якоби для функций $g_i(x)$, $i = 1, \ldots, m$, рассматриваемой в точке $\hat{x}$, равен m. Не нарушая общности, предположим, что отличен от нуля определитель (якобиан), составленный из частных производных по первым m аргументам, т. е.

$$\begin{vmatrix} \dfrac{\partial g_1}{\partial x^1} & \cdots & \dfrac{\partial g_1}{\partial x^m} \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ \dfrac{\partial g_m}{\partial x^1} & \cdots & \dfrac{\partial g_m}{\partial x^m} \end{vmatrix} \neq 0. \tag{3.2}$$

Тогда по теореме о неявных функциях в некоторой окрестности точки $\hat{x}$ система уравнений (3.1) разрешима относительно $x^1, \ldots, x^m$, т. е. представима в виде

$$x^j = \varphi_j(x^{m+1}, \ldots, x^n), \quad j = 1, 2, \ldots, m, \tag{3.3}$$

где $\varphi_j(x^{m+1}, \ldots, x^n)$ — непрерывно дифференцируемые в рассматриваемой окрестности функции. Переменные $x^{m+1}, \ldots, x^n$ естественно назвать «независимыми», в отличие от «зависимых» — $x^1, \ldots, x^m$. Подставляя выражения (3.3) в функцию $f(x)$, получим задачу отыскания безусловного экстремума функции $n-m$ переменных

$$f(\varphi_1(x^{m+1}, \ldots, x^n), \ldots, \varphi_m(x^{m+1}, \ldots, x^n), \quad x^{m+1}, \ldots, x^n) = \tilde{f}(x^{m+1}, \ldots, x^n).$$

Однако провести исключение части компонент вектора x обычно бывает трудно или даже невозможно. Поэтому мы используем другой путь определения точки $\hat{x}$, который не предполагает наличия явных выражений типа (3.3), хотя использует существенно условие (3.2).

2. Метод множителей Лагранжа. Как мы видели в замечании к теореме 2.1, в точке $\hat{x}$, доставляющей безусловный экстремум функции, ее полный дифференциал равен нулю, т. е.

$$df(\hat{x}) = \sum_{j=1}^{m} \frac{\partial f}{\partial x^j}(\hat{x})\, dx^j + \sum_{k=m+1}^{n} \frac{\partial f}{\partial x^k}(\hat{x})\, dx^k = 0, \tag{3.4}$$

где dx^j, $j=1, \ldots, m$, — дифференциалы «зависимых» переменных, связанные с дифференциалами «независимых» переменных dx^k, $k=m+1, \ldots, n$, следующим образом:

$$\sum_{j=1}^{m} \frac{\partial g_i}{\partial x^j}(\hat{x})\, dx^j + \sum_{k=m+1}^{n} \frac{\partial g_i}{\partial x^k}(\hat{x})\, dx^k = 0, \quad i=1, \ldots, m. \tag{3.5}$$

Уравнения (3.5) получены при дифференцировании полным образом уравнений связи (3.1). Исключим теперь дифференциалы «зависимых» переменных из уравнений (3.4), (3.5). Для этого умножим каждое из уравнений системы (3.5) на произвольные множители $\lambda_1, \ldots, \lambda_m$ и результаты сложим с уравнением (3.4), тогда получим следующее равенство:

$$\sum_{j=1}^{m} \left(\frac{\partial f}{\partial x^j}(\hat{x}) + \lambda_1 \frac{\partial g_1}{\partial x^j}(\hat{x}) + \ldots + \lambda_m \frac{\partial g_m}{\partial x^j}(\hat{x}) \right) dx^j + \sum_{k=m+1}^{n} \left(\frac{\partial f}{\partial x^k}(\hat{x}) + \lambda_1 \frac{\partial g_1}{\partial x^k}(\hat{x}) + \ldots + \lambda_m \frac{\partial g_m}{\partial x^k}(\hat{x}) \right) dx^k = 0. \tag{3.6}$$

Распорядимся множителями $\lambda_1, \ldots, \lambda_m$ таким образом, чтобы обратились в нуль коэффициенты при дифференциалах «зависимых» переменных, т. е.

$$\frac{\partial f}{\partial x^j}(\hat{x}) + \lambda_1 \frac{\partial g_1}{\partial x^j}(\hat{x}) + \ldots + \lambda_m \frac{\partial g_m}{\partial x^j}(\hat{x}) = 0, \quad j = 1, \ldots, m. \tag{3.7}$$

Это можно сделать, так как уравнения (3.7) являются системой линейных алгебраических уравнений относительно множителей $\lambda_1, \ldots, \lambda_m$, которая имеет единственное решение в силу того, что ее определитель (3.2) по условию отличен от нуля. При выбранных таким образом значениях множителей в равенстве (3.6) останутся только члены, содержащие дифференциалы «независимых» переменных. Поэтому коэффициенты при этих дифференциалах должны быть нулями, т. е.

$$\frac{\partial f}{\partial x^k}(\hat{x}) + \lambda_1 \frac{\partial g_1}{\partial x^k}(\hat{x}) + \ldots + \lambda_m \frac{\partial g_m}{\partial x^k}(\hat{x}) = 0, \quad k = m+1, \ldots, n. \tag{3.8}$$

Таким образом, мы получили систему $n+m$ уравнений (3.1), (3.7), (3.8) относительно $n+m$ неизвестных $\hat{x}^1, \ldots, \hat{x}^n, \hat{\lambda}_1, \ldots, \hat{\lambda}_m$. Этот результат представляет собой основное содержание *метода множителей Лагранжа* и позволяет определить множество «претендентов» на решение в задаче на относительный экстремум. Метод Лагранжа состоит из следующих этапов:

1) составляется функция $n+m$ переменных, которая называется функцией Лагранжа:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x); \tag{3.9}$$

2) вычисляются и приравниваются нулю ее частные производные по x и λ:

$$\begin{aligned} \frac{\partial L}{\partial x^j} &= \frac{\partial f}{\partial x^j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x^j} = 0, \quad && j = 1, \ldots, n, \\ \frac{\partial L}{\partial \lambda_i} &= g_i(x) = 0, && i = 1, \ldots, m; \end{aligned} \tag{3.10}$$

3) решается система (3.10) $n+m$ уравнений относительно $n+m$ неизвестных $x^1, \ldots, x^n, \lambda_1, \ldots, \lambda_m$.

Система уравнений (3.10) представляет собой необходимые условия первого порядка в задаче на относительный экстремум, а ее решения $\hat{x}^1, \ldots, \hat{x}^n$ принято называть *условно-стационарными точками*. Как и в случае задач на безусловный экстремум, необходимые условия первого порядка не определяют характера условно-стационарной точки. Для выяснения этого вопроса следует привлечь производные более высоких порядков функций $f(x)$ и $g(x)$.

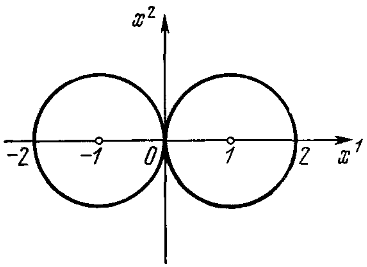

Рис. 3.1

Заметим, что требование неравенства нулю якобиана (3.2) является существенным. Только в этом случае система уравнений (3.7) разрешима, причем единственным образом, относительно множителей Лагранжа $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_m$.

Пример 3.1. Условие (3.2) может быть не выполнено, если решение задачи на относительный экстремум реализуется, например, в точке касания поверхностей ограничений (3 1) (начало координат на рис. 3.1). Пусть

$$n=2, \quad m=2, \quad f(x)=x^2,$$

$$g_1(x)=(x^1-1)^2+(x^2)^2-1,$$

$$g_2(x)=(x^1+1)^2+(x^2)^2-1.$$

Допустимая точка должна одновременно удовлетворять уравнениям $g_1(x)=0$, $g_2(x)=0$ и является единственной: $x^1=0$, $x^2=0$. Очевидно, что точка $\hat{x}^1=0$, $\hat{x}^2=0$ и будет решением задачи на относительный минимум функции $f(x)=x^2$ при ограничениях $g_1(x)=g_2(x)=0$. Составим для этой задачи функцию Лагранжа:

$$L(x, \lambda)=f(x)+\sum_{i=1}^{2}\lambda_i g_i(x)=$$

$$=x^2+\lambda_1[(x^1-1)^2+(x^2)^2-1]+\lambda_2[(x^1+1)^2+(x^2)^2-1].$$