Санкт-Петербургский политехнический университет

Петра Великого

Боревич А.З.

ДИФФЕРЕНЦИАЛЬНОЕ ИСЧИСЛЕНИЕ ФУНКЦИЙ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

Учебное пособие

Санкт-Петербург

2015

Оглавление

Глава 1. Предел. Непрерывность

В этой главе рассматриваются вопросы, связанные с непрерывностью функций нескольких переменных. §1 посвящен введению n-мерного пространства (пространства $\mathbf{R}^n$) – множества упорядоченных числовых наборов, состоящих из n компонент. Для элементов этого пространства вводятся операции сложения, умножения на число и скалярное умножение, определяются понятия нормы и сходимости, а также некоторые специальные подмножества этого пространства. Во втором параграфе содержатся основные определения и факты, относящиеся к понятиям предела и непрерывности функции нескольких переменных. Разбираются некоторые свойства предела и непрерывных функций.

§ 1. Пространство $\mathbf{R}^n$

1°. Скалярное произведение и норма. Рассмотрим множество всех упорядоченных пар $\left\{(x,y) \,\middle|\, x,y \ \text{ – вещественные числа}\right\}$ и введем на этом множестве две арифметические операции:

1) *сложение* по правилу

$$(x_1,y_1)+(x_2,y_2)=(x_1+x_2,\, y_1+y_2);$$

2) *умножение на вещественное число* α с помощью равенства

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

Это множество всех упорядоченных пар с введенными выше операциями будем называть *двумерным векторным пространством* и обозначать $\mathbf{R}^2$.

Пространство $\mathbf{R}^2$ нетрудно интерпретировать геометрически. Для этого рассмотрим плоскость и фиксированную прямоугольную декартову систему координат на ней. Между точками этой плоскости и элементами $\mathbf{R}^2$ можно установить взаимно однозначное соответствие. А именно, сопоставим каждой точке M плоскости упорядоченную пару вещественных чисел (x,y), состоящую из прямоугольных координат точки M в выбранной системе координат: $M(x,y)\to(x,y)\in\mathbf{R}^2$. Обратно, всякой упорядоченной паре чисел (x,y) отвечает единственная точка плоскости, имеющая число x своей абсциссой и число y – ординатой (рис.1.1): $(x,y)\in\mathbf{R}^2\to M(x,y)$.

Другая геометрическая интерпретация пространства $\mathbf{R}^2$ – это множество всех радиус-векторов той же плоскости. Каждой точке M на рассматриваемой плоскости соответствует единственный радиус-вектор $\boldsymbol{r}$, начало которого совпадает с началом координат – точкой O, а конец – с точкой M. Обратно, каждому радиус-вектору $\boldsymbol{r}$ сопоставляется единственная точка M на плоскости, а именно – его конец. Таким образом, возникает еще одно взаимно однозначное соответствие – между элементами пространства $\mathbf{R}^2$ и множеством всех радиус-векторов на плоскости (рис.1.1): $(x,y) \leftrightarrow M(x,y) \leftrightarrow \boldsymbol{r}(x,y)$

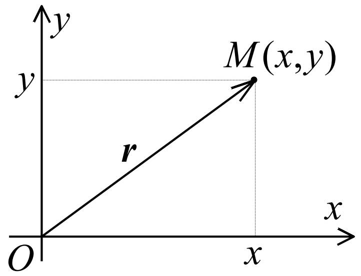

Рис. 1.1. Геометрическая интерпретация пространства $\mathbf{R}^2$.

Заметим, что в этой геометрической интерпретации введенным на множестве $\mathbf{R}^2$ операциям соответствуют известные правила сложения векторов (“правило параллелограмма”) и умножения вектора на вещественное число.

Указанные взаимно однозначные соответствия позволяют отождествить элементы пространства $\mathbf{R}^2$ с геометрическими объектами (точками на плоскости или соответствующими им радиус-векторами), поэтому сами элементы двумерного пространства также называют *точками* или *векторами*.

Аналогично множество $\left\{(x,y,z) \,\middle|\, x,y,z \text{ – вещественные числа}\right\}$ всех упорядоченных троек вещественных чисел можно превратить в трехмерное пространство $\mathbf{R}^3$.

Перейдем к рассмотрению общего случая.

Определение 1. Пусть n – натуральное число. Множество упорядоченных наборов, состоящих из n вещественных чисел, обозначаемых через $\boldsymbol{x} = (x_1, \ldots, x_n)$, где $x_1, \ldots, x_n$ – произвольные вещественные числа, для которых введены операции сложения и умножения на вещественные числа по следующим правилам:

$$\boldsymbol{x} + \boldsymbol{y} = (x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$

$$\alpha \boldsymbol{x} = \alpha (x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n),$$

где α – произвольное вещественное число, называется *n-мерным пространством* и обозначается $\mathbf{R}^n$. Элементы пространства $\mathbf{R}^n$ называются *точками* или *векторами* n-мерного пространства.

Отметим, что одномерное пространство $\mathbf{R}^1$ представляет собой не что иное, как множество всех вещественных чисел с обычными операциями сложения и умножения (без операции деления).

Вектор $\mathbf{0} = (0, \ldots, 0)$ называется *нулевым вектором*, вектор $(-1) \cdot \boldsymbol{x}$ обозначают также через $(-\boldsymbol{x})$ и называют *противоположным* вектору $\boldsymbol{x}$, а *разностью* векторов $\boldsymbol{x}$ и $\boldsymbol{y}$

называют вектор $\boldsymbol{x}+(-\boldsymbol{y})$ и обозначают его $\boldsymbol{x}-\boldsymbol{y}$.

Легко убедиться в справедливости следующих соотношений между произвольными элементами $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ пространства $\mathbf{R}^n$ (ниже λ, μ – произвольные вещественные числа):

1) $(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}=\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})$;
2) $\boldsymbol{x}+\mathbf{0}=\boldsymbol{x}$;
3) $\boldsymbol{x}+(-\boldsymbol{x})=\mathbf{0}$;
4) $\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$;
5) $1\cdot\boldsymbol{x}=\boldsymbol{x}$;
6) $\lambda(\boldsymbol{x}+\boldsymbol{y})=\lambda\boldsymbol{x}+\lambda\boldsymbol{y}$;
7) $(\lambda+\mu)\boldsymbol{x}=\lambda\boldsymbol{x}+\mu\boldsymbol{x}$;
8) $\lambda(\mu\boldsymbol{x})=(\lambda\mu)\boldsymbol{x}$.

На множестве элементов пространства $\mathbf{R}^n$ можно ввести операцию умножения, сопоставляющую двум любым элементам этого пространства определенное вещественное число по следующему правилу.

Определение 2. *Скалярным произведением* двух векторов $\boldsymbol{x}=(x_1,\ldots,x_n)$ и $\boldsymbol{y}=(y_1,\ldots,y_n)$ пространства $\mathbf{R}^n$ называется число, определяемое следующим образом

$$x_1y_1+x_2y_2+\ldots+x_ny_n,$$

которое обозначается через $\langle\boldsymbol{x},\boldsymbol{y}\rangle$.

Иногда для обозначения скалярного произведения используют символику, заимствованную из линейной алгебры. Очевидно, векторы $\boldsymbol{x}=(x_1,\ldots,x_n)$ и $\boldsymbol{y}=(y_1,\ldots,y_n)$ представляют собой матрицы размера $1\times n$, т.е. векторы-строки, а транспонированная матрица $\boldsymbol{y}^T$ будет в таком случае матрицей размера $n\times 1$, т.е. вектором-столбцом. Поэтому согласно определению операции умножения матриц имеем

$$\boldsymbol{x}\cdot\boldsymbol{y}^T=(x_1,\ldots,x_n)\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}=x_1y_1+\ldots+x_ny_n=\langle\boldsymbol{x},\boldsymbol{y}\rangle,$$

т.е. скалярное произведение векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ может быть записано в виде $\boldsymbol{x}\cdot\boldsymbol{y}^T$.

Непосредственно из определения 2 вытекают следующие *свойства скалярного произведения*:

Свойство 1. $\langle\boldsymbol{x},\boldsymbol{x}\rangle\geq 0$ для любого $\boldsymbol{x}\in\mathbf{R}^n$, причем $\langle\boldsymbol{x},\boldsymbol{x}\rangle=0 \iff \boldsymbol{x}=\mathbf{0}$;

Свойство 2. $\langle\boldsymbol{x},\boldsymbol{y}\rangle=\langle\boldsymbol{y},\boldsymbol{x}\rangle$ для любых $\boldsymbol{x},\boldsymbol{y}\in\mathbf{R}^n$;

Свойство 3. $\langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z}\rangle = \langle \boldsymbol{x}, \boldsymbol{z}\rangle + \langle \boldsymbol{y}, \boldsymbol{z}\rangle$; $\langle \boldsymbol{x}, \boldsymbol{y}+\boldsymbol{z}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{z}\rangle$ для любых $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbf{R}^n$;

Свойство 4. $\langle \lambda\boldsymbol{x}, \boldsymbol{y}\rangle = \lambda\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, \lambda\boldsymbol{y}\rangle$, где λ – произвольное вещественное число.

Определение 3. *Нормой* произвольного вектора $\boldsymbol{x} \in \mathbf{R}^n$ называется неотрицательное число, обозначаемое $\|\boldsymbol{x}\|$ и определяемое равенством:

$$\|\boldsymbol{x}\| = \sqrt{<\boldsymbol{x}, \boldsymbol{x}>} = \sqrt{x_1^2 + \ldots + x_n^2}\ .$$

Норму вектора называют также *длиной* этого вектора, так как в частном случае пространства $\mathbf{R}^2$ (и $\mathbf{R}^3$) определение 3 приводит к понятию длины геометрического вектора в традиционном смысле: если $\boldsymbol{x} = (x_1, x_2) \in \mathbf{R}^2$, то $\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2} = |\boldsymbol{x}|$.

Отметим следующие *свойства нормы*:

Свойство 1. $\|\boldsymbol{x}\| \ge 0$ для любого $\boldsymbol{x} \in \mathbf{R}^n$, причем $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \mathbf{0}$;

Свойство 2. $\|\lambda\boldsymbol{x}\| = |\lambda| \cdot \|\boldsymbol{x}\|$ для любого вещественного числа λ.

▶ В самом деле, имеем $\|\lambda\boldsymbol{x}\| = \sqrt{\langle \lambda\boldsymbol{x}, \lambda\boldsymbol{x}\rangle} = \sqrt{\lambda^2\langle \boldsymbol{x}, \boldsymbol{x}\rangle} = |\lambda|\sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle} = |\lambda| \cdot \|\boldsymbol{x}\|$. ◀

Свойство 3. $\|\boldsymbol{x}+\boldsymbol{y}\| \le \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

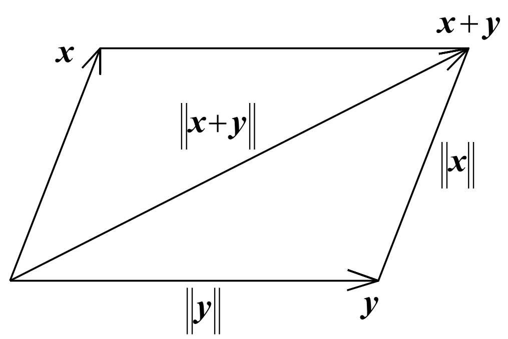

Рис. 1.2. Геометрическая интерпретация неравенства треугольника

Последнее свойство носит название “неравенство треугольника”, так как в случае пространства $\mathbf{R}^2$ (или $\mathbf{R}^3$) выражает известный геометрический факт: длина любой стороны треугольника не превосходит суммы длин двух других его сторон (рис.1.2).

Доказательство этого свойства проведем с использованием следующего классического неравенства.

Теорема 1 (неравенство Коши–Буняковского). Для любых двух наборов вещественных чисел $x_1, \ldots, x_n$ и $y_1, \ldots, y_n$ справедливо неравенство

$$\left|\sum_{i=1}^{n} x_i y_i\right| \le \sqrt{\sum_{i=1}^{n} x_i^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2}\ . \qquad (1)$$

В терминах скалярного произведения и нормы неравенство (1) можно переписать так: для любых $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{R}^n$ верно

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \le \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\| \qquad (2)$$

▶ В случае $\boldsymbol{x} = \mathbf{0}$ неравенство (2) выполняется тривиальным образом. Пусть теперь $\boldsymbol{x} \ne \mathbf{0}$. Рассмотрим вектор $t\boldsymbol{x}+\boldsymbol{y}$ при $t \in \mathbf{R}$. В силу первого свойства скалярного произведения при любом вещественном t. справедливо неравенство $\langle tx+y,\ tx+y\rangle \ge 0$. Раскрывая левую часть этого неравенства, на основании свойств 2), 3) и 4) скалярного произведения получим

$$\langle x, x\rangle t^2 + 2\langle x, y\rangle t + \langle y, y\rangle \ge 0 .$$

То, что последнее неравенство выполняется при любом вещественном t означает, что квадратный трехчлен с положительным коэффициентом при t^2, находящийся в левой части неравенства, имеет неположительный дискриминант, т.е. справедливо неравенство

$$\langle x, y\rangle^2 - \langle x, x\rangle\langle y, y\rangle \le 0 ,$$

а это равносильно неравенству (2). ◀

Замечание 1. Элементы пространства $\mathbf{R}^n$ – это упорядоченные наборы из n вещественных чисел. Как отмечалось в начале этого пункта, в частном случае пространства $\mathbf{R}^2$ (или $\mathbf{R}^3$), если на плоскости выбрана фиксированная система координат, упорядоченные пары (тройки) чисел можно рассматривать как геометрические радиус-векторы.

Введенное определением 2 скалярное произведение, как известно из векторной алгебры, совпадает с геометрическим определением скалярного произведения:

$$\langle \boldsymbol{a}, \boldsymbol{b}\rangle = |\boldsymbol{a}|\cdot|\boldsymbol{b}|\cdot\cos\varphi ,$$

где $|\boldsymbol{a}|$, $|\boldsymbol{b}|$ – длины векторов $\boldsymbol{a}$ и $\boldsymbol{b}$, а φ – угол между ними. Так как понятие длины геометрического вектора и понятие нормы элемента из $\mathbf{R}^2$ (см. определение 3) просто совпадают, неравенство (2) в этом случае очевидно:

$$|\langle \boldsymbol{a}, \boldsymbol{b}\rangle| = |\boldsymbol{a}|\cdot|\boldsymbol{b}|\cdot|\cos\varphi| \le |\boldsymbol{a}|\cdot|\boldsymbol{b}| = \|\boldsymbol{a}\|\cdot\|\boldsymbol{b}\| .$$

Обратимся теперь к доказательству свойства 3) (неравенство треугольника).

▶ Из неравенства (2) вытекает неравенство $\langle x, y\rangle \le \|x\|\cdot\|y\|$, к обеим частям которого после умножения на 2 прибавим сумму квадратов норм векторов x и y:

$$\|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \le \|x\|^2 + 2\|x\|\cdot\|y\| + \|y\|^2 .$$

На основании полученного неравенства имеем

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle =$$

$$= \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \le \|x\|^2 + 2\|x\|\cdot\|y\| + \|y\|^2 = \left(\|x\| + \|y\|\right)^2 .$$

Извлечение квадратного корня из крайних членов полученного неравенства приводит к неравенству треугольника. ◀

Замечание 2. Используя соответствие между элементами пространства $\mathbf{R}^2$ и геометрическими векторами на плоскости, можно, по аналогии, ввести понятие угла между векторами пространства $\mathbf{R}^n$. Назовем *углом* между векторами $x, y \in \mathbf{R}^n$ ($x \ne \mathbf{0}$, $y \ne \mathbf{0}$) такой угол $\varphi \in [0, \pi]$, что

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

В частности, если $\cos\varphi = 0$, то векторы $\boldsymbol{x}$ и $\boldsymbol{y}$ называют *ортогональными* и пишут: $\boldsymbol{x} \perp \boldsymbol{y}$. Нулевой вектор считается по определению ортогональным к любому вектору: $\mathbf{0} \perp \boldsymbol{x}$ для любого $\boldsymbol{x} \in \mathbf{R}^n$.

2°. Множества в пространстве $\mathbf{R}^n$. В этом пункте, почти сплошь состоящем из определений и пояснений к ним, рассматриваются специальные, важные для дальнейшего изложения, подмножества пространства $\mathbf{R}^n$ и вводится соответствующая терминология.

Определение 4. Пусть ε – некоторое положительное число, $\boldsymbol{x}_0$ – точка пространства $\mathbf{R}^n$. Подмножество пространства $\mathbf{R}^n$ вида

$$\left\{ \boldsymbol{x} \in \mathbf{R}^n \;\middle|\; \|\boldsymbol{x} - \boldsymbol{x}_0\| < \varepsilon \right\}$$

называется *открытым шаром с центром в точке* $\boldsymbol{x}_0$ *и радиусом* ε и обозначается: $u_\varepsilon(\boldsymbol{x}_0)$. Часто это же подмножество называют ε-*окрестностью точки* $\boldsymbol{x}_0$.

Иногда, если радиус упомянутого открытого шара не существенен в данном контексте и специально не фиксируется, говорят просто об окрестности точки $\boldsymbol{x}_0$ и пишут $u(\boldsymbol{x}_0)$.

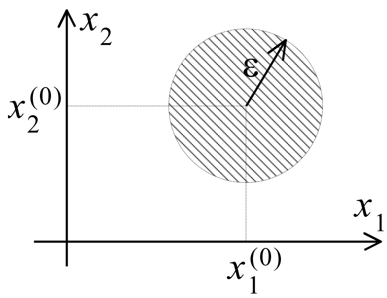

Рис. 1.3. ε-окрестность в пространстве $\mathbf{R}^2$

Для случая $\mathbf{R}^3$ множество $u_\varepsilon(\boldsymbol{x}_0)$ – это обычный шар с центром в точке $\boldsymbol{x}_0 = (x_1^{(0)}, x_2^{(0)}, x_3^{(0)})$ и радиусом ε, причем точки сферы, его ограничивающей, этому множеству не принадлежат. Для пространства $\mathbf{R}^2$ множество $u_\varepsilon(\boldsymbol{x}_0)$ составляют все точки, лежащие внутри круга с центром в точке $\boldsymbol{x}_0 = (x_1^{(0)}, x_2^{(0)})$ и радиусом ε (точки, лежащие на окружности, задающей этот круг, множеству $u_\varepsilon(\boldsymbol{x}_0)$ не принадлежат) (см. рис.1.3).

Рассмотрим произвольное непустое подмножество X пространства $\mathbf{R}^n$.Пусть $\boldsymbol{x} \in X$. Точка $\boldsymbol{x}$ называется *внутренней точкой множества* X, если она принадлежит X вместе с некоторой своей окрестностью; более точно: если существует окрестность $u(\boldsymbol{x}_0)$ точки $\boldsymbol{x}_0$ такая, что $u(\boldsymbol{x}_0) \subset X$. Иными словами, точка $\boldsymbol{x}_0$ является внутренней для множества X в том случае, если множеству X принадлежит не только она сама, но и все ее “ближайшее окружение”.

Пусть снова X – непустое подмножество пространства $\mathbf{R}^n$. Точка $\boldsymbol{x} \in \mathbf{R}^n$ (не обязательно принадлежащая X) называется *граничной точкой множества* X, если в любой ее окрестности найдется точка $\boldsymbol{x}^{(1)} \in X$ (возможно, сама точка $\boldsymbol{x}$, если она принадлежит X) и, одновременно, точка $\boldsymbol{x}^{(2)} \notin X$ (подчеркнем еще раз, что в определении речь идет о любой окрестности точки $\boldsymbol{x}$). Таким образом, если точка $\boldsymbol{x} \in X$ не является внутренней, то она граничная, т.е. *все точки множества X относятся либо к категории внутренних, либо к категории граничных*. В то же время точка может являться граничной для данного множества X, но ему не принадлежать.

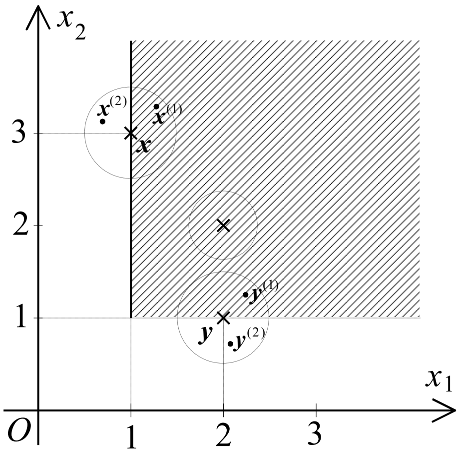

Рис. 1.4. Множество $D = \left\{ (x_1, x_2) \in \mathbf{R}^2 \;\middle|\; x_1 \ge 1,\ x_2 > 1 \right\}$.

Пример 1. Для плоского множества

$$D = \left\{ (x_1, x_2) \in \mathbf{R}^2 \;\middle|\; x_1 \ge 1,\ x_2 > 1 \right\}$$

точка $(2, 2)$ – внутренняя; точка $\boldsymbol{x} = (1, 3)$ – граничная, принадлежащая D; точка $\boldsymbol{y} = (2, 1)$ – граничная, но множеству D не принадлежит (см. рис.1.4).

Определение 5. Множество $X, X \subset \mathbf{R}^n$, называется *открытым*, если каждая его точка внутренняя.

Определение 6. Множество $X, X \subset \mathbf{R}^n$, называется *замкнутым*, если оно содержит все свои граничные точки.

Существует договоренность, согласно которой пустое множество, равно как и все пространство $\mathbf{R}^n$, относят одновременно и к категории открытых, и к категории замкнутых множеств. По некоторым причинам это удобно, а к недоразумениям не приводит.

Покажем, что открытый шар из определения 4 – открытое множество в смысле определения 5. Пусть $u_\varepsilon(\boldsymbol{x}_0)$ – открытый шар в пространстве $\mathbf{R}^n$ и $\boldsymbol{x}^* \in u_\varepsilon(\boldsymbol{x}_0)$ – любая точка из этого множества. Требуется проверить, что $\boldsymbol{x}^*$ – внутренняя точка множества $u_\varepsilon(\boldsymbol{x}_0)$. Так как $\left\| \boldsymbol{x}^* - \boldsymbol{x}_0 \right\| < \varepsilon$, число $\delta = \varepsilon - \left\| \boldsymbol{x}^* - \boldsymbol{x}_0 \right\|$ – положительно. Тогда окрестность $u_\delta(\boldsymbol{x}^*)$ точки $\boldsymbol{x}^*$ содержится в открытом шаре $u_\varepsilon(\boldsymbol{x}_0)$. Действительно, для любой точки $\boldsymbol{x} \in u_\delta(\boldsymbol{x}^*)$, используя неравенство треугольника пункта 1, имеем

$$\left\| \boldsymbol{x} - \boldsymbol{x}_0 \right\| = \left\| \boldsymbol{x} - \boldsymbol{x}^* + \boldsymbol{x}^* - \boldsymbol{x}_0 \right\| \le \left\| \boldsymbol{x} - \boldsymbol{x}^* \right\| + \left\| \boldsymbol{x}^* - \boldsymbol{x}_0 \right\| < \delta + \left\| \boldsymbol{x}^* - \boldsymbol{x}_0 \right\| = \varepsilon - \left\| \boldsymbol{x}^* - \boldsymbol{x}_0 \right\| + \left\| \boldsymbol{x}^* - \boldsymbol{x}_0 \right\| = \varepsilon,$$

т.е. $\boldsymbol{x} \in u_\varepsilon(\boldsymbol{x}_0)$.

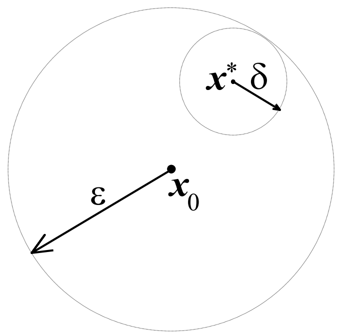

Рис. 1.5. Открытый шар является открытым множеством

Доказанное включение $u_\delta(\boldsymbol{x}^*) \subset u_\varepsilon(\boldsymbol{x}_0)$ и означает, что точка $\boldsymbol{x}^*$, произвольно выбранная из открытого шара $u_\varepsilon(\boldsymbol{x}_0)$, является его внутренней точкой (см. рис.1.5).

В качестве примера замкнутого подмножества пространства $\mathbf{R}^2$ приведем $G = \left\{(x_1, x_2) \in \mathbf{R}^2 \;\middle|\; 1 \le x_1 \le 2;\ 0 \le x_2 \le 1\right\}$. Встречавшееся выше множество $D = \left\{(x_1, x_2) \in \mathbf{R}^2 \;\middle|\; x_1 \ge 1,\ x_2 > 1\right\}$ нельзя отнести ни к открытым, ни к замкнутым, так как, например, точка $\boldsymbol{x} = (1, 3)$ принадлежит множеству D, но внутренней точкой этого множества не является, в то время как точка $\boldsymbol{y} = (2, 1)$, будучи граничной точкой множества D, ему не принадлежит.

Сферой радиуса r *с центром в точке* $\boldsymbol{x}_0$ называется множество

$$S = \left\{\boldsymbol{x} \in \mathbf{R}^n \;\middle|\; \|\boldsymbol{x} - \boldsymbol{x}_0\| = r\right\},$$

а *отрезком, соединяющим точки* $\boldsymbol{x}^{(1)}$ *и* $\boldsymbol{x}^{(2)}$, – множество

$$T = \left\{\boldsymbol{x} \in \mathbf{R}^n \;\middle|\; \boldsymbol{x} = \lambda(\boldsymbol{x}^{(2)} - \boldsymbol{x}^{(1)}) + \boldsymbol{x}^{(1)};\ \lambda \in [0, 1]\right\},$$

или, что то же самое,

$$T = \left\{\boldsymbol{x} \in \mathbf{R}^n \;\middle|\; \boldsymbol{x} = (1 - \lambda)\boldsymbol{x}^{(1)} + \lambda \boldsymbol{x}^{(2)};\ \lambda \in [0, 1]\right\}.$$

В случае $n = 2$ множество S представляет собой окружность, а в случае $n = 3$ – обычную геометрическую сферу. Множество T в обоих случаях – геометрический отрезок с концами в точках $\boldsymbol{x}^{(1)}$ и $\boldsymbol{x}^{(2)}$ (на плоскости и в пространстве соответственно). В этих частных случаях замкнутость S и T очевидна, так как каждое из этих двух множеств состоит из своих граничных точек и только из них. Замкнутость сферы и отрезка в случае $n > 3$ примем без доказательства.

Пусть снова $X, X \subset \mathbf{R}^n$, $X \ne \varnothing$. Точка $\boldsymbol{x}^{(0)} \in \mathbf{R}^n$ называется *предельной точкой* множества X, если в любой окрестности этой точки содержится, по крайней мере, одна точка $\boldsymbol{x} \in X$, такая, что $\boldsymbol{x} \ne \boldsymbol{x}^{(0)}$. Ясно, что каждая внутренняя точка множества X является предельной. В то же время граничная точка множества X может быть предельной точкой, а может и не быть таковой. Замкнутое множество, очевидно, содержит все свои предельные точки.

Наконец, множество X, $X \subset \mathbf{R}^n$, называется *ограниченным*, если существует такое число $M > 0$, что неравенство $\|\boldsymbol{x}\| \le M$ выполняется для каждой точки $\boldsymbol{x} \in X$

Так, например, непосредственно из определения отрезка T, соединяющего точки $\boldsymbol{x}^{(1)}$ и $\boldsymbol{x}^{(2)}$ следует его ограниченность. Действительно, для любого вектора $\boldsymbol{x} \in T$ выполняется требуемое условие, если в качестве числа M использовать константу $\left\|\boldsymbol{x}^{(2)}-\boldsymbol{x}^{(1)}\right\|+\left\|\boldsymbol{x}^{(1)}\right\|$. (В случае, когда точки $\boldsymbol{x}^{(1)}$ и $\boldsymbol{x}^{(2)}$ совпадают, отрезок T вырождается в точку, и ограниченность такого множества очевидна). Применяя неравенство треугольника, имеем

$$\|\boldsymbol{x}\|=\left\|\lambda(\boldsymbol{x}^{(2)}-\boldsymbol{x}^{(1)})+\boldsymbol{x}^{(1)}\right\| \leq \lambda\left\|\boldsymbol{x}^{(2)}-\boldsymbol{x}^{(1)}\right\|+\left\|\boldsymbol{x}^{(1)}\right\| \leq\left\|\boldsymbol{x}^{(2)}-\boldsymbol{x}^{(1)}\right\|+\left\|\boldsymbol{x}^{(1)}\right\|=M .$$

Аналогично проверяется ограниченность открытого шара $u_{\varepsilon}(\boldsymbol{x}_0)$ и сферы S. Множество D из примера 1 и все пространство $\mathbf{R}^n$ ограниченными не являются (такие множества называются *неограниченными*).

В заключение отметим еще один тип множеств, использование которых связано с понятием предела функции нескольких переменных (см. §2 гл.1). *Проколотой окрестностью* (ε-*окрест-*
ностью) *точки* $\boldsymbol{x}_0$ называют подмножество в $\mathbf{R}^n$ следующего вида:

$$\left\{\boldsymbol{x} \in \mathbf{R}^n \mid 0<\left\|\boldsymbol{x}-\boldsymbol{x}_0\right\|<\varepsilon\right\}.$$

Ясно, что проколотая ε-окрестность – это ε-окрестность, из которой удалена точка $\boldsymbol{x}_0$. Проколотая окрестность, так же как и собственно окрестность, представляет собой открытое ограниченное множество. Обозначать проколотую окрестность будем символом $\dot{u}(\boldsymbol{x}_0)$ или $\dot{u}_{\varepsilon}(\boldsymbol{x}_0)$.

3°. Сходимость в $\mathbf{R}^n$. По аналогии с обычной числовой последовательностью введем понятие последовательности точек в $\mathbf{R}^n$, которые для удобства будем обозначать большими латинскими буквами: M, N, P и т. д. Если каждому номеру $k \in \mathbf{N}$ поставлена в соответствие точка $M_k \in \mathbf{R}^n$, то говорят, что *задана последовательность точек*, которую обозначают $\{M_k\}_{k=1}^{\infty}$ или просто $\{M_k\}$.

Определение 6. Точка $M_0 \in \mathbf{R}^n$ называется *пределом последовательности* $\{M_k\}$, если $\lim\limits_{k \to \infty}\left\|M_k-M_0\right\|=0$.

Здесь символ lim обозначает предел обычной числовой последовательности $\left\{\left\|M_k-M_0\right\|\right\}$ (с неотрицательными членами). Тот факт, что точка M_0 является пределом последовательности $\{M_k\}$, записывается в виде $\lim\limits_{k \to \infty} M_k=M_0$ или $M_k \xrightarrow[k \to \infty]{} M_0$.

Таким образом, вспоминая определение предела числовой последовательности, можно сказать, что равенство $\lim\limits_{k\to\infty} M_k = M_0$ имеет место тогда и только тогда, когда для любого $\varepsilon > 0$ существует такой номер K, что для всех номеров $k > K$ выполняется неравенство $\|M_k - M_0\| < \varepsilon$ или, что то же, $M_k \in u_\varepsilon(M_0)$. Иными словами, точку M_0 называют пределом последовательности $\{M_k\}$, если в любой (сколь угодно малой) окрестности точки M_0 содержатся все точки данной последовательности, за исключением, быть может, некоторого конечного числа точек этой последовательности, не попавших в упомянутую окрестность.

Последовательности, имеющие предел, также как числовые, называются *сходящимися*.

Пример 2. Пусть $M_k = \left(\dfrac{k-1}{k};\ \dfrac{k+1}{k^2};\ \dfrac{1-k}{k+1}\right)$ для любого натурального k, и пусть $M_0 = (1, 0, -1)$. Тогда $\lim\limits_{k\to\infty} M_k = M_0$. Действительно, для квадрата нормы разности $M_k - M_0$ имеем

$$\|M_k - M_0\|^2 = \left(\frac{k-1}{k} - 1\right)^2 + \left(\frac{k+1}{k^2} - 0\right)^2 + \left(\frac{1-k}{k+1} + 1\right)^2 = \frac{1}{k^2} + \frac{(k+1)^2}{k^4} + \frac{4}{(k+1)^2},$$

и, следовательно, $\lim\limits_{k\to\infty} \|M_k - M_0\|^2 = 0$, а значит, и $\lim\limits_{k\to\infty} \|M_k - M_0\| = 0$.

Следующая ниже теорема, которую называют *теоремой о покоординатной сходимости*, позволяет несколько упростить процесс вычисления предела последовательности точек в $\mathbf{R}^n$.

Теорема 2. Точка $M_0 = (x_1^{(0)}, \dots, x_n^{(0)})$ является пределом последовательности $\{M_k\}_{k=1}^{\infty}$, где $M_k = (x_1^{(k)}, \dots, x_n^{(k)}) \in \mathbf{R}^n$, тогда и только тогда, когда имеет место покоординатная сходимость, т.е. равенства $\lim\limits_{k\to\infty} x_i^{(k)} = x_i^{(0)}$ выполнены для всех $i = 1, 2, \dots, n$.

▶ *Необходимость.* Очевидно, для всякого $i = 1, 2, \dots, n$ справедливы соотношения

$$0 \le \left|x_i^{(k)} - x_i^{(0)}\right| \le \sqrt{\sum_{j=1}^{n} (x_j^{(k)} - x_j^{(0)})^2} = \|M_k - M_0\|.$$

По условию выполнено равенство $\lim\limits_{k\to\infty} \|M_k - M_0\| = 0$ и, следовательно, из написанных выше соотношений по теореме о сжатой переменной для числовых последовательностей получаем $\lim\limits_{k\to\infty} \left|x_i^{(k)} - x_i^{(0)}\right| = 0$, а значит, и $\lim\limits_{k\to\infty} (x_i^{(k)} - x_i^{(0)}) = 0$ для всех $i = 1, 2, \dots, n$. Поэтому $\lim\limits_{k\to\infty} x_i^{(k)} = x_i^{(0)}$.

Достаточность. По условию, $\lim\limits_{k\to\infty} (x_i^{(k)} - x_i^{(0)}) = 0$ для всякого $i = 1, 2, \dots, n$. Используя обычные правила вычисления пределов числовых последовательностей, получаем, что $\lim\limits_{k\to\infty} \sum\limits_{i=1}^{n} (x_i^{(k)} - x_i^{(0)})^2 = 0$. Из непрерывности квадратного корня (степенной функции) вытекает равенство $\lim\limits_{k\to\infty}\sqrt{\sum\limits_{i=1}^{n}(x_i^{(k)}-x_i^{(0)})^2}=0$. Таким образом, доказано, что $\lim\limits_{k\to\infty}\|M_k-M_0\|=0$, а это и означает справедливость равенства $\lim\limits_{k\to\infty}M_k=M_0$. ◀

Возвращаясь к последовательности $M_k=\left(\dfrac{k-1}{k};\ \dfrac{k+1}{k^2};\ \dfrac{1-k}{k+1}\right)$ из примера 2, найдем ее предел с помощью доказанной теоремы. Действительно, в силу равенств

$$\lim_{k\to\infty}x_1^{(k)}=\lim_{k\to\infty}\frac{k-1}{k}=1=x_1^{(0)},\quad \lim_{k\to\infty}x_2^{(k)}=\lim_{k\to\infty}\frac{k+1}{k^2}=0=x_2^{(0)},\quad \lim_{k\to\infty}x_3^{(k)}=\lim_{k\to\infty}\frac{1-k}{1+k}=-1=x_3^{(0)},$$

из теоремы о покоординатной сходимости следует, что пределом последовательности $\{M_k\}$ является точка $M_0=(1,0,-1)$.

Теорема о покоординатной сходимости позволяет многие факты, установленные в свое время для сходящихся числовых последовательностей, перенести на последовательности точек в $\mathbf{R}^n$. Отметим лишь некоторые из них.

1. Если предел последовательности $\{M_k\}$ существует, то он единственный.

2. Всякая подпоследовательность сходящейся последовательности сходится, причем имеет тот же предел, что и исходная последовательность.

3. $\lim\limits_{k\to\infty}(M_k+N_k)=\lim\limits_{k\to\infty}M_k+\lim\limits_{k\to\infty}N_k$, при условии, что пределы в правой части равенства существуют.

4. **Теорема Больцано–Вейерштрасса.** Из всякой ограниченной последовательности точек пространства $\mathbf{R}^n$ можно выделить сходящуюся подпоследовательность.

Ограниченной последовательностью в пространстве $\mathbf{R}^n$ называют последовательность, ограниченную как множество точек в $\mathbf{R}^n$, т.е. ту, для которой найдется такое число $M>0$, что нормы всех члены этой последовательности его не превосходят.

Разумеется, утверждения, относящиеся к сходящимся числовым последовательностям, в которых члены последовательностей сравниваются по величине (так называемые теоремы сравнения: о сжатой переменной, о предельном переходе в неравенствах и т. п.) перенести на последовательности точек в $\mathbf{R}^n$ невозможно, поскольку отношение “>“ для элементов из $\mathbf{R}^n$ при $n>1$ не определено.

§ 2. Непрерывность функций нескольких переменных

1°. Понятие функции нескольких переменных.

Определение 1. Пусть на непустом множестве $X, X \subset \mathbf{R}^n$, задано правило *f*, согласно которому каждому элементу $\boldsymbol{x} \in X$ ставится в соответствие определенное вещественное число *y*. В таком случае говорят, что на множестве X задана *функция n переменных* $f\colon X \to \mathbf{R}$, которую обозначают $f(\boldsymbol{x})$, $f(x_1, \ldots, x_n)$, или $f(M)$, где M – точка пространства $\mathbf{R}^n$. Переменные $x_1, \ldots, x_n$ называют *независимыми переменными* или *аргументами*, а переменную y – *зависимой переменной*. Желая в обозначении функции отразить зависимую переменную, пишут: $y = f(\boldsymbol{x})$, $y = f(M)$ и т. п. Множество X называется *областью определения* функции $y = f(M)$.

Для функции двух переменных ее аргументы часто обозначают буквами *x*, *y*, а зависимую переменную буквой *z*: $z = f(x, y)$. Для таких функций возможна наглядная геометрическая интерпретация – аналог графика функции одной переменной. Если (x, y) – точка из области определения функции, а $z = f(x, y)$ – ее значение для этих аргументов, то в пространстве *Oxyz* отмечается точка (x, y, z). Перебрав таким образом все точки из области определения, получим, вообще говоря, некоторую поверхность, соответствующую области определения, лежащей в координатной плоскости *Oxy* (рис.1.6).

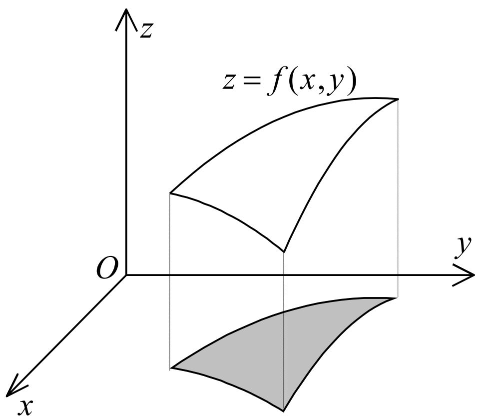

Рис.1.6. Геометрическая интерпретация функции двух переменных

Подобной наглядной интерпретации не имеют функции от трех и большего числа аргументов. В какой-то степени представить характер изменения функции позволяет рассмотрение так называемых множеств уровня, а именно, если $f(\boldsymbol{x})$ – функция от n переменных, c – вещественное число, то множество $\Gamma_c = \left\{ \boldsymbol{x} \in \mathbf{R}^n \;\middle|\; f(\boldsymbol{x}) = c \right\}$ называется *множеством уровня функции f, соответствующим значению c*. В случае $n = 2$ множество Γ_c называют *линией уровня*.

Пример 1. Как известно из аналитической геометрии, уравнение $z = x^2 - y^2$ определяет поверхность, носящую название *гиперболический параболоид*. Линии уровня (при $c \ge 0$) этой функции изображены на рис.1.7.

Если рассмотреть линии уровня с различными значениями *c*, составляющими арифметическую прогрессию, то по тому, как меняется расстояние между этими линиями можно судить о форме поверхности: “густо” расположенные линии свидетельствуют о быстром изменении z, т.е. о значительной крутизне поверхности; если линии уровня достаточно разрежены – поверхность в этой своей части сравнительно пологая. Заметим, что эту идею давно и плодотворно используют в картографии, где линия уровня – это множество точек на поверхности Земли с фиксированной высотой над уровнем моря.

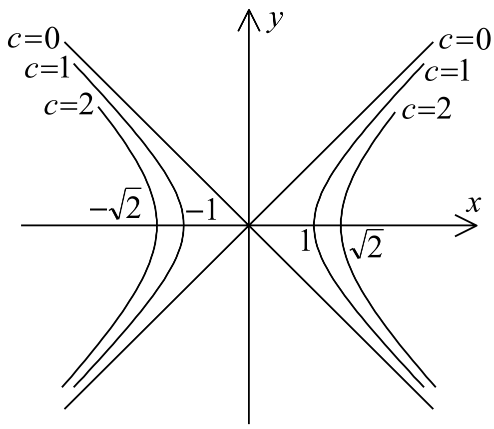

Рис. 1.7. Линии уровня для функции двух переменных $f(x,y)=x^2-y^2$.

Рассмотрим линейную и квадратичную функции от n переменных, как наиболее простые.

Линейной называется функция, определяемая равенством

$$y=\langle \boldsymbol{c},\boldsymbol{x}\rangle = c_1x_1+c_2x_2+\ldots+c_nx_n,$$

где $\boldsymbol{x}=(x_1,\ldots,x_n)$, а $\boldsymbol{c}=(c_1,\ldots,c_n)$ – фиксированный вектор из $\mathbf{R}^n$.

Функция вида $y=\langle \boldsymbol{c},\boldsymbol{x}\rangle+\alpha$, где α – вещественное число, называется *аффинной* (хотя часто и ее называют линейной).

Пусть A – вещественная симметричная матрица размера $n\times n$ вида

$$A=\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & . & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix},$$

где $a_{ij}=a_{ji}$ $(i,\ j=1,\ldots,n)$. *Квадратичной функцией* называется функция

$$y=\frac{1}{2}\langle \boldsymbol{x}A,\boldsymbol{x}\rangle+\langle \boldsymbol{c},\boldsymbol{x}\rangle+\alpha.$$

В частном случае $n=2$ имеем

$$\boldsymbol{x}A=(x_1,x_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}=\left(x_1a_{11}+x_2a_{21},\ x_1a_{12}+x_2a_{22}\right)$$

и, следовательно, учитывая, что $a_{12}=a_{21}$, получаем следующий общий вид *квадратичной функции двух переменных*:

$$y=\frac{1}{2}(a_{11}x_1^2+a_{21}x_2x_1+a_{12}x_1x_2+a_{22}x_2^2)+(c_1x_1+c_2x_2)+\alpha=$$

$$=\frac{1}{2}a_{11}x_1^2+a_{21}x_2x_1+\frac{1}{2}a_{22}x_2^2+c_1x_1+c_2x_2+\alpha.$$

Нетрудно понять, что для линейной функции $y = c_1 x_1 + c_2 x_2$ двух переменных линии уровня суть семейство параллельных прямых с вектором нормали $\boldsymbol{c} = (c_1, c_2)$, а для квадратичной функции вида $y = x_1^2 + x_2^2$ – это семейство концентрических окружностей с центром в начале координат.

2°. Предел функции нескольких переменных.

Определение 2. Пусть $f(M) = f(x_1, \dots, x_n)$ – функция n переменных, заданная на множестве $X, X \subset \mathbf{R}^n$, и M_0 – предельная точка множества X (п.2° §1 гл.1). Число A называется *пределом функции* $f(M)$ *в точке* M_0, если для любой последовательности $\{M_k\}_{k=1}^{\infty}$, такой, что $M_k \neq M_0$, $M_k \in X$ и $\lim\limits_{k\to\infty} M_k = M_0$, выполняется $\lim\limits_{k\to\infty} f(M_k) = A$. При этом пишут $\lim\limits_{M\to M_0} f(M) = A$ или $f(M) \xrightarrow[M\to M_0]{} A$.

Это определение предела функции *на языке последовательностей*. Так же, как и в случае функции одной переменной, можно ввести еще одно эквивалентное определение предела функции нескольких переменных *на языке* "ε–δ".

Определение 3. В условиях определения 2 число A называется *пределом функции* $f(M)$ *в точке* M_0, если для любого $\varepsilon > 0$ существует $\delta > 0$ такое, что для всех точек $M \in X \cap \dot{u}_\delta(M_0)$ выполняется неравенство $|f(M) - A| < \varepsilon$.

Напомним, что символ $\dot{u}_\delta(M_0)$, введенный в п.2° §1 гл.1, означает проколотую окрестность точки M_0, а именно, множество вида $\left\{ \boldsymbol{x} \in \mathbf{R}^n \;\middle|\; 0 < \|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta \right\}$.

Определение 4. Если $\lim\limits_{M\to M_0} f(M) = 0$, то функцию f называют *бесконечно малой функцией* в точке M_0.

Используя координатную запись точек $M(x_1, \dots, x_n)$, $M_0(x_1^{(0)}, \dots, x_n^{(0)})$, условие $M \in \dot{u}_\delta(M_0)$ из приведенных выше определений можно несколько видоизменить. Отметим, что если некоторое свойство имеет место для всех элементов из некоторого множества, то оно же (это свойство) и подавно выполняется для всех элементов из любого его подмножества. Рассмотрим множество

$$D_\delta(M_0) = \left\{ M(x_1, \dots, x_n) \;\middle|\; 0 < \left|x_i - x_i^{(0)}\right| < \delta;\; i = 1, \dots, n \right\}.$$

Для точек этого множества справедливо неравенство

$$\left\|\boldsymbol{x} - \boldsymbol{x}^{(0)}\right\| = \sqrt{\sum_{i=1}^{n} (x_i - x_i^{(0)})^2} < \sqrt{n} \cdot \delta,$$

и, следовательно, для положительного числа $\delta' = \dfrac{\delta}{\sqrt{n}}$ имеет место включение $D_{\delta'}(M_0) \subset \dot{u}_\delta(M_0)$.

С другой стороны, если точка M принадлежит множеству $\dot{u}_\delta(M_0)$, то, поскольку для каждого $i = 1,2,...,n$ верно неравенство $\left|x_i - x_i^{(0)}\right| \le \|\boldsymbol{x} - \boldsymbol{x}_0\|$, эта точка тем более принадлежит множеству $D_\delta(M_0)$, т.е. проколотая окрестность $\dot{u}_\delta(M_0)$ является подмножеством множества $D_\delta(M_0)$. Геометрически полученное означает, что в любой круг можно вписать квадрат и, обратно, в любой квадрат можно вписать круг.

Проиллюстрируем установленный факт на примере функции двух переменных (см. рис.1.8):

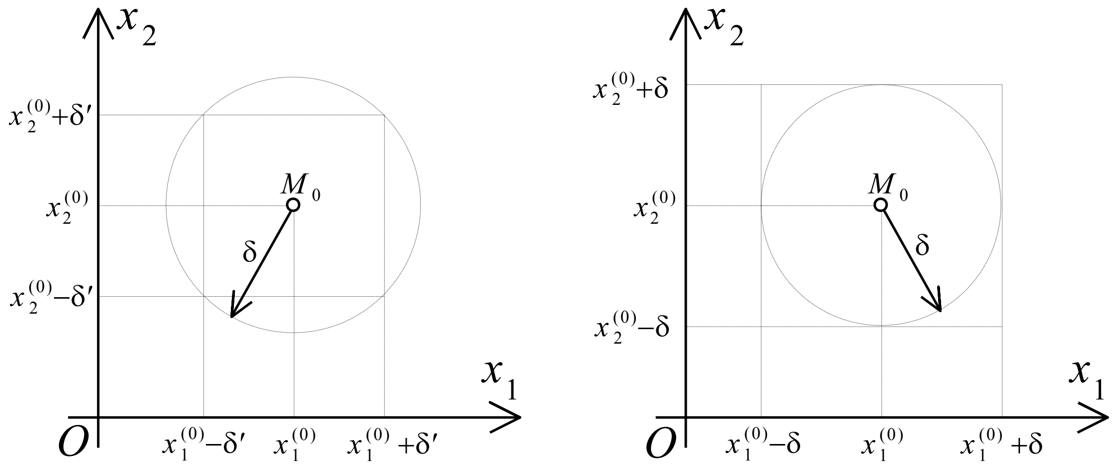

Рис. 1.8. Проколотая окрестность и ее модификация

Из всего сказанного следует, что *условие* $M \in \dot{u}_\delta(M_0)$ *в определениях 3-4 можно при желании заменить неравенствами* $0 < \left|x_i - x_i^{(0)}\right| < \delta$, $i = 1, \ldots, n$; *при этом получающееся определение будет эквивалентно исходному.*

Рассмотрим понятие предела для функции $f(x, y)$ двух переменных более детально. В этом случае для предела в точке $M_0(x_0, y_0)$ используют еще обозначение $\lim\limits_{\substack{x\to x_0\\ y\to y_0}} f(x, y)$ и называют его *двойным пределом*.

Наряду с двойным можно рассматривать так называемый *повторный предел* $\lim\limits_{x\to x_0}\left(\lim\limits_{y\to y_0} f(x, y)\right)$, где сначала вычисляется “внутренний” предел $\lim\limits_{y\to y_0} f(x, y) = \varphi(x)$ функции одной переменной y, при этом переменная x предполагается фиксированной, в результате чего образуется функция $\varphi(x)$ лишь одной переменной x, а затем вычисляется “внешний” предел, т.е. $\lim\limits_{x\to x_0} \varphi(x)$. Аналогично, рассматривая сначала $\lim\limits_{x\to x_0} f(x, y) = \psi(y)$, а затем $\lim\limits_{y\to y_0}\psi(y)$, можно говорить о еще одном повторном пределе для $f(x, y)$: $\lim\limits_{y\to y_0}\left(\lim\limits_{x\to x_0} f(x,y)\right)$.

Следующий пример показывает, что существование (и даже равенство между собой) двух повторных пределов не может гарантировать существования в этой точке двойного предела.

Пример 2. Функция $f(x,y)=\dfrac{xy}{x^2+y^2}$ определена на множестве $X=\mathbf{R}^2\setminus\{(0,0)\}$. Ее повторные пределы в точке $M_0(0,0)$ существуют и равны между собой:

$$\lim_{x\to 0}\left(\lim_{y\to 0}\frac{xy}{x^2+y^2}\right)=\lim_{x\to 0}0=0,\quad \lim_{y\to 0}\left(\lim_{x\to 0}\frac{xy}{x^2+y^2}\right)=\lim_{y\to 0}0=0,$$

в то время как двойной предел $\lim\limits_{\substack{x\to 0\\ y\to 0}}\dfrac{xy}{x^2+y^2}$ не существует.

▶ Действительно, для последовательности точек $M_k\left(\dfrac{1}{k},\dfrac{1}{k}\right)$, сходящейся к точке $M_0(0,0)$, соответствующая последовательность значений функции сходится к $1/2$:

$$f(M_k)=\frac{1/k^2}{1/k^2+1/k^2}=\frac{1}{2}\xrightarrow[k\to\infty]{}\frac{1}{2},$$

а для последовательности точек $M'_k\left(\dfrac{1}{k},\dfrac{2}{k}\right)\xrightarrow[k\to\infty]{}M_0(0,0)$ соответствующая последовательность значений функции сходится к $2/5$:

$$f(M_k)=\frac{2/k^2}{1/k^2+4/k^2}=\frac{2}{5}\xrightarrow[k\to\infty]{}\frac{2}{5}.$$

Неравенство $\lim\limits_{k\to\infty}f(M_k)\neq\lim\limits_{k\to\infty}f(M'_k)$ в соответствии с определением 2 означает, что в точке $M_0(0,0)$ данная функция предела (двойного) не имеет. ◀

Используя определения 2 и 3, можно перенести основные теоремы о пределах функции одной переменной на случай функций нескольких переменных. Приведем формулировки некоторых из них.

Теорема 1. Пусть функции $f(M)$ и $g(M)$ определены на множестве X, $X\subset\mathbf{R}^n$, $M_0\in X$ – предельная точка множества X, и существуют пределы $\lim\limits_{M\to M_0}f(M)=A$, $\lim\limits_{M\to M_0}g(M)=B$. Тогда функции $f(M)+g(M)$, $f(M)-g(M)$, $f(M)\cdot g(M)$ имеют в точке M_0 пределы $A+B$, $A-B$, AB соответственно. Если дополнительно известно, что $B\neq 0$, то и функция $f(M)/g(M)$ имеет в точке M_0 предел, равный A/B.

Из теоремы 1 немедленно следует, что сумма, разность и произведение двух бесконечно малых функций является бесконечно малой функцией.

Теорема 2 (о сохранении знака). Если функция нескольких переменных $f(M)$ имеет в точке M_0 предел, равный $A>0$, то существует такая окрестность $\dot{u}(M_0)$, что для всех точек M из пересечения этой окрестности с множеством X, на котором определена эта функция, выполняется неравенство $f(M)>0$. Аналогично, случай $A<0$.

Теорема 3 (о сжатой переменной). Если существуют пределы $\lim\limits_{M\to M_0} f(M)=A$ и $\lim\limits_{M\to M_0} g(M)=A$, причем в некоторой окрестности $\dot{u}(M_0)$ для всех точек $M\in\dot{u}(M_0)$ выполняется двойное неравенство: $f(M)\le h(M)\le g(M)$, то и у функции $h(M)$ в точке M_0 существует предел, равный тому же самому числу A.

Отметим, наконец, еще две ситуации, имеющие отношение к понятию предела. Говорят, что функция $f(M)$ стремится к бесконечности при $M\to M_0$, и пишут $\lim\limits_{M\to M_0} f(M)=\infty$, если для любого $\mathrm{E}>0$ существует $\delta>0$ такое, что для любой точки $M\in X\bigcap\dot{u}_\delta(M_0)$ выполняется неравенство $|f(M)|>\mathrm{E}$ (здесь, как в определении 2, M_0 – предельная точка области определения X функции $f(M)$). В этой случае функцию f называют также *бесконечно большой функцией* в окрестности точки M_0.

Можно еще говорить о пределе функции $f(\boldsymbol{x})=f(x_1,\dots,x_n)$.при $\|x\|\to+\infty$, а именно: число A называется пределом в этом случае, если для любого $\varepsilon>0$ существует $\Delta>0$ такое, что для всех точек $\boldsymbol{x}$, удовлетворяющих неравенству $\|\boldsymbol{x}\|>\Delta$, выполняется условие $|f(\boldsymbol{x})-A|<\varepsilon$. Этот факт записывается следующим образом: $\lim\limits_{\|\boldsymbol{x}\|\to+\infty} f(\boldsymbol{x})=A$.

3°. Непрерывность в точке и на множестве.

Определение 5. Пусть $f(M)$ – функция n переменных, заданная на множестве X, $X\subset\mathbf{R}^n$, и M_0 – предельная точка множества X. Говорят, что функция $f(M)$ *непрерывна в точке* M_0, если $\lim\limits_{M\to M_0} f(M)=f(M_0)$.

Таким образом, непрерывность функции в точке означает совпадение предела этой функции с ее значением в этой точке.

Из определения 5 и из теоремы об арифметических операциях над функциями, имеющими предел, непосредственно следует, что если $f(M)$ и $g(M)$ непрерывны в точке M_0, то и функции $f(M)+g(M)$, $f(M)-g(M)$, $f(M)\cdot g(M)$, $f(M)/g(M)$ непрерывны в этой точке (последняя при условии, что $g(M_0) \neq 0$).

Может случиться, что функция от n переменных, заданная на множестве X, $X \subset \mathbf{R}^n$, устроена так, что реально зависит не от всех переменных $x_1, \dots, x_n$, а лишь от некоторых из них:
например, функция $f(x_1, \dots, x_n) = px_k$ (где p – некоторое вещественное число) зависит лишь от одной переменной x_k (k – фиксированный номер переменной), а остальные переменные не участвуют в выражении для ее задания. Функцию px_k можно считать определенной на всем пространстве $\mathbf{R}^n$, причем из ее непрерывности как линейной функции одной переменной следует ее непрерывность в любой точке $\mathbf{R}^n$ как функции n переменных.

Это замечание и приведенный выше факт непрерывности суммы устанавливает непрерывность аффинной функции (см. п.1°) $y = \langle \boldsymbol{c}, \boldsymbol{x} \rangle + \alpha = c_1 x_1 + \dots + c_n x_n + \alpha$.

Аналогично, функция $\alpha_{ij} x_i x_j$ (α_{ij} – вещественное число), рассматриваемая как функция n переменных, непрерывна в любой точке пространства $\mathbf{R}^n$ (произведение двух непрерывных функций: $\alpha_{ij} x_i$ и x_j) и, следовательно, тем же свойством обладает сумма таких функций $f(x_1, \dots, x_n) = \sum\limits_{i \le j} \alpha_{ij} x_i x_j$ $(i,\ j = 1, \dots, n)$.

Отсюда в силу непрерывности суммы непрерывных функций вытекает непрерывность квадратичной функции (см. п.1°) $y = \frac{1}{2} \langle \boldsymbol{x}A, \boldsymbol{x} \rangle + \langle \boldsymbol{c}, \boldsymbol{x} \rangle + \alpha$.

Для функций нескольких переменных можно ввести понятие сложной функции. Пусть функция n переменных $f(\boldsymbol{x}) = f(x_1, \dots, x_n)$ определена на множестве X, $X \subset \mathbf{R}^n$, и пусть на множестве D, $D \subset \mathbf{R}^m$, заданы n функций x_i от m переменных: $x_i = g_i(\boldsymbol{t}) = g_i(t_1, \dots, t_m)$, $\boldsymbol{t} \in D$, причем для всех $\boldsymbol{t} \in D$ выполняется включение $\left(g_1(\boldsymbol{t}), \dots, g_n(\boldsymbol{t})\right) \in X$. В таком случае можно говорить о заданной на множестве D, $D \subset \mathbf{R}^m$, *сложной функции* (или иначе: *суперпозиции*)

$$f\left(g_1(\boldsymbol{t}), \dots, g_n(\boldsymbol{t})\right),$$

от m переменных $t_1, \dots, t_m$.

Данное определение является прямым обобщением известного понятия суперпозиции для функций одного аргумента.

Теорема 4. Пусть функции $g_1(\boldsymbol{t})$, …, $g_n(\boldsymbol{t})$, описанные выше, непрерывны в точке $\boldsymbol{t}_0 \in D \subset \mathbf{R}^m$, $\boldsymbol{x}_0 = \left(g_1(\boldsymbol{t}_0), \dots, g_n(\boldsymbol{t}_0)\right) \in X \subset \mathbf{R}^n$, и функция $f(\boldsymbol{x})$ непрерывна в точке $\boldsymbol{x}_0$.

Тогда сложная функция $F(\boldsymbol{t}) = f\big(g_1(\boldsymbol{t}), \dots, g_n(\boldsymbol{t})\big)$ непрерывна в точке $\boldsymbol{t}_0$.

▶ Рассмотрим произвольную последовательность точек $\boldsymbol{t}^{(k)} = \left(t_1^{(k)}, \dots, t_m^{(k)}\right)$ такую, что $\lim\limits_{k\to\infty} \boldsymbol{t}^{(k)} = \boldsymbol{t}_0$. Тогда, поскольку $\lim\limits_{k\to\infty} g_i(\boldsymbol{t}^{(k)}) = g_i(\boldsymbol{t}_0)$ для каждого $i = 1, \dots, n$ (это равенство означает непрерывность функции $g_i(t)$ в точке $\boldsymbol{t}_0$), на основании теоремы о покоординатной сходимости справедливо следующее равенство:

$$\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \lim_{k\to\infty}\left(g_1(\boldsymbol{t}^{(k)}), \dots, g_n(\boldsymbol{t}^{(k)})\right) = \left(g_1(\boldsymbol{t}_0), \dots, g_n(\boldsymbol{t}_0)\right) = \boldsymbol{x}_0 .$$

Используя определение 2 и непрерывность функции $f(\boldsymbol{x})$ в точке $\boldsymbol{x}_0$, получаем $\lim\limits_{k\to\infty} f(\boldsymbol{x}^{(k)}) = f(\boldsymbol{x}_0)$. Таким образом, для любой последовательности $\left\{\boldsymbol{t}^{(k)}\right\}_{k=1}^{\infty}$ с условием $\lim\limits_{k\to\infty} \boldsymbol{t}^{(k)} = \boldsymbol{t}_0$ выполнено:

$$\lim_{k\to\infty} F(\boldsymbol{t}^{(k)}) = \lim_{k\to\infty} f\left(g_1(\boldsymbol{t}^{(k)}), \dots, g_n(\boldsymbol{t}^{(k)})\right) = \lim_{k\to\infty} f(\boldsymbol{x}^{(k)}) = f(\boldsymbol{x}_0) = f\left(g_1(\boldsymbol{t}_0), \dots, g_n(\boldsymbol{t}_0)\right) = F(\boldsymbol{t}_0),$$

что в соответствии с определением 4 и определением предела функции на языке последовательностей (определение 2) доказывает теорему. ◀

Полученная теорема позволяет заключить, что *всякая функция нескольких переменных (как и функция одной переменной), полученная из основных элементарных функций при помощи конечного числа арифметических операций и конечного числа операций суперпозиции, непрерывна в любой предельной точке из области определения.*

Как уже отмечалось, для функций нескольких переменных, имеющих предел, справедлива теорема о сохранении знака. Этот же факт применительно к функциям нескольких переменных, непрерывным в точке, сформулируем в следующем виде.

Теорема 5. Пусть функция $f(\boldsymbol{x})$ – непрерывна в точке $\boldsymbol{x}_0$, являющейся внутренней точкой области определения X, $X \subset \mathbf{R}^n$ этой функции, и пусть $f(\boldsymbol{x}_0) > 0$ (или $f(\boldsymbol{x}_0) < 0$). Тогда существует такая окрестность $u_\delta(\boldsymbol{x}_0)$, что $f(\boldsymbol{x}) > 0$ (соответственно $f(\boldsymbol{x}) < 0$) для всех точек $\boldsymbol{x} \in u_\delta(\boldsymbol{x}_0)$.

▶ Ясно, что каким бы ни было положительное число $f(\boldsymbol{x})$, найдется такое число $\varepsilon > 0$, что разность $f(\boldsymbol{x}) - \varepsilon$ все еще положительна. Для этого числа ε согласно определению 3 существует такое число $\delta > 0$, что для всех $\boldsymbol{x} \in u_\delta(\boldsymbol{x}_0)$ выполняется неравенство $\left|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)\right| < \varepsilon$, или, что то же самое, $-\varepsilon < f(\boldsymbol{x}) - f(\boldsymbol{x}_0) < \varepsilon$. Рассматривая левую часть этого двойного неравенства, имеем $f(\boldsymbol{x}) > f(\boldsymbol{x}_0) - \varepsilon > 0$ для всех $\boldsymbol{x}$ из упомянутой окрестности.

Второй случай, когда $f(\boldsymbol{x}_0) < 0$, разбирается аналогично ◀

Будем говорить, что функция $f(\boldsymbol{x})$ *непрерывна на множестве* X, $X \subset \mathbf{R}^n$, если она непрерывна в каждой точке этого множества. Следующая теорема устанавливает важное свойство функций, непрерывных на множестве.

Теорема 6 (Вейерштрасс). Пусть X – непустое, замкнутое и ограниченное подмножество пространства $\mathbf{R}^n$ (п.2 §1 гл.1), и пусть функция $f(\boldsymbol{x})$ непрерывна на X. Тогда эта функция ограничена на множестве X (т.е. существует такое число $L > 0$, что $|f(\boldsymbol{x})| \le L$ для всех $\boldsymbol{x} \in X$) и, более того, она достигает на множестве X своих наибольшего и наименьшего значений, точнее говоря, существуют точки $\boldsymbol{x}^* \in X$ и $\boldsymbol{x}_* \in X$ такие, что выполнены равенства

$$f(\boldsymbol{x}^*) = \sup_{\boldsymbol{x}\in X} f(\boldsymbol{x}) = \max_{\boldsymbol{x}\in X} f(\boldsymbol{x}), \quad f(\boldsymbol{x}_*) = \inf_{\boldsymbol{x}\in X} f(\boldsymbol{x}) = \min_{\boldsymbol{x}\in X} f(\boldsymbol{x}).$$

Эта теорема обобщает известные для функций одной переменной теоремы Вейерштрасса на случай функций нескольких переменных. Доказательство теоремы 6 не приводится.

Резюме к главе 1

В §1 этой главы рассматривается *n*-мерное пространство и некоторые его свойства, вводятся понятия скалярного произведения и нормы, доказывается классическое неравенство Коши–Буняковского, изучаются некоторые специфические подмножества пространства $\mathbf{R}^n$. Вторая часть параграфа посвящена свойствам сходящихся последовательностей точек пространства $\mathbf{R}^n$, в частности, доказывается теорема о покоординатной сходимости.

В §2 изучаются функции нескольких переменных, рассматриваются множества уровня таких функций, вводится понятие предела функции в точке, приводятся основные факты, касающиеся функций, имеющих предел. В заключительной части параграфа рассматриваются свойства функций, непрерывных в точке и на множестве: непрерывность суперпозиции, теорема о сохранении знака, теорема Вейерштрасса и некоторые другие.

Задание на самостоятельную работу

1. Опишите граничные точки проколотой окрестности $\dot{u}_\varepsilon(\boldsymbol{x}_0)$. Докажите, что проколотая окрестность $\dot{u}_\varepsilon(\boldsymbol{x}_0)$ – открытое множество.

2. Сравните определения граничной и предельной точек множества $X, X \subset \mathbf{R}^n$. В каком случае граничная точка не является предельной? Приведите пример.

3. Приведите примеры замкнутых неограниченных множеств из пространства $\mathbf{R}^2$.

4. Приведите пример ограниченного множества из пространства $\mathbf{R}^4$, не являющегося ни открытым, ни замкнутым.

5. Может ли ограниченная последовательность $\{M_k\}_{k=1}^{\infty}$, $M_k \in \mathbf{R}^n$, $k = 1,2,...$, не иметь предела?

6. Докажите, что всякая подпоследовательность сходящейся последовательности $\{M_k\}_{k=1}^{\infty}$, $M_k \in \mathbf{R}^n$, $k = 1,2,...$, сходится, причем имеет тот же предел, что и сама последовательность (свойство 2 п.3° §1).

7. Изобразите на координатной плоскости *Oxy* линии уровня функций $z = \sqrt{x^2 + y^2}$, $z = x^2 + y^2$, $z = x^2 + |y|$.

8. Используя определение предела функции нескольких переменных на языке “ε–δ” (определение 3 п.2° §2),

а) докажите теорему о сохранении знака (теорема 2 п.2° §2);

б) докажите, что функция $f(M)$, имеющая предел в точке M_0, ограниченна в некоторой окрестности этой точки.

9. Приведите пример функции $f(\boldsymbol{x})$, непрерывной на ограниченном множестве X, $X \subset \mathbf{R}^n$, которая не является ограниченной на этом множестве.

10. Множества X_1 и X_2 определены следующим образом:

$$X_1 = \left\{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 \le 2(x + y)\right\}, \quad X_2 = \left\{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 \ge 1\right\}.$$

Являются ли множества X_1, X_2, $X_3 = X_1 \cap X_2$, $X_4 = X_1 \setminus X_2$ компактами?

Решите следующие задачи из сборника задач [3]:
7.4 – 7.16, 7.19 – 7.21, 7.24 – 7.26, 7.29 – 7.54.

Глава 2. Дифференцируемость функций нескольких переменных

В этой главе вводится понятие дифференцируемой функции нескольких переменных. В первых двух параграфах рассматриваются основные свойства дифференцируемых функций, определяются касательная плоскость, градиент, производная по направлению, дифференциал и доказывается формула Тейлора. Третий параграф посвящен функциям, заданным неявно.

§ 1. Частные производные и дифференцируемость

1°. Частные производные. Рассмотрим функцию двух переменных $f(x, y)$, определенную в окрестности $u(M_0)$ фиксированной точки $M_0(x_0, y_0)$. Пусть, далее, Δx – вещественное число (которое будем называть приращением переменной x в точке M_0) такое, что точка $M(x_0+\Delta x, y_0)$ также содержится в окрестности $u(M_0)$. Разность $f(x_0+\Delta x, y_0)-f(x_0, y_0)$ называется *частным приращением функции* $f(x, y)$ *по переменной* x *в точке* M_0 и обозначается $\Delta_x f(x_0, y_0)$. В этом обозначении не отражено то приращение Δx переменной x, которое, собственно, и породило данное частное приращение функции, но к недоразумениям это обстоятельство обычно не приводит. Отметим, что для фиксированной точки $M_0(x_0, y_0)$ частное приращение $\Delta_x f(x_0, y_0)$ представляет собой функцию одной переменной Δx. Аналогично определяется частное приращение функции $f(x, y)$ по переменной y. Таким образом, имеем равенства

$$\Delta_x f(x_0, y_0) = f(x_0+\Delta x, y_0) - f(x_0, y_0), \quad \Delta_y f(x_0, y_0) = f(x_0, y_0+\Delta y) - f(x_0, y_0).$$

Для $\Delta x \neq 0$ ($\Delta y \neq 0$) отношение $\dfrac{\Delta_x f(x_0, y_0)}{\Delta x}$ (соответственно $\dfrac{\Delta_y f(x_0, y_0)}{\Delta y}$) является функцией одной переменной Δx (соответственно Δy), определенной в проколотой окрестности $\dot{u}(0)$. После этих предварительных замечаний перейдем к определению частных производных.

Определение 1. Пусть функция $f(x, y)$ определена в окрестности $u(M_0)$ точки $M_0(x_0, y_0)$. Если существует предел $\lim\limits_{\Delta x\to 0}\dfrac{\Delta_x f(x_0, y_0)}{\Delta x}$, то он называется *частной производной функции* $f(x, y)$ *в точке* $M_0(x_0, y_0)$ *по переменной* x и обозначается одним из следующих способов: $f'_x(M_0)$, $f'_x(x_0, y_0)$, $\dfrac{\partial f(M_0)}{\partial x}$, $\dfrac{\partial f(x_0, y_0)}{\partial x}$.

Аналогично определяется частная производная по переменной y:

$$f'_y(M_0) = f'_y(x_0, y_0) = \frac{\partial f(M_0)}{\partial y} = \frac{\partial f(x_0, y_0)}{\partial y} = \lim_{\Delta y \to 0} \frac{\Delta_y f(x_0, y_0)}{\Delta y}.$$

Нетрудно заметить, что частная производная по переменной x (соответственно по y) представляет собой производную функции одной переменной $f(x, y_0)$ в точке x_0 (соответственно функции $f(x_0, y)$ одной переменной y, вычисленную в точке y_0). Это обстоятельство позволяет находить частную производную по какой-либо из переменных при помощи известных правил дифференцирования функций одной переменной, рассматривая оставшуюся переменную как величину постоянную в данном контексте.

Например, вычислим частные производные функции $f(x, y) = \cos\dfrac{x}{y} + e^{x^2 + y^3}$ в точке (x, y):

$$f'_x(x, y) = -\sin\left(\frac{x}{y}\right) \cdot \frac{1}{y} + e^{x^2+y^3} \cdot 2x, \quad f'_y(x, y) = -\sin\left(\frac{x}{y}\right) \cdot \left(-\frac{x}{y^2}\right) + e^{x^2+y^3} \cdot 3y^2.$$

В предыдущей главе (п.1 §2) отмечалось, что график функции $z = f(x, y)$ двух переменных в пространстве $Oxyz$ допускает геометрическую интерпретацию в виде некоторой поверхности, образованной точками пространства с координатами $\big(x, y, f(x, y)\big)$, где точки (x, y) принадлежат области определения функции $f(x, y)$. На основе такой интерпретации можно дать геометрическое истолкование введенному понятию частной производной. А именно, рассмотрим сечение указанной поверхности плоскостью $y = y_0$ (рис.2.1). Заметим, что график функции $z = f(x, y_0)$ одной переменной x представляет собой линию пересечения этой поверхности с плоскостью $y = y_0$. Поскольку частная производная $f'_x(x_0, y_0)$ совпадает с производной функции $f(x, y_0)$ в точке x_0, исходя из геометрического смысла производной функции одной переменной, можно сделать вывод: число $f'_x(x_0, y_0)$ есть тангенс угла θ наклона к положительному направлению оси Ox касательной, проходящей через точку $P\big(x_0, y_0, f(x_0, y_0)\big)$ поверхности, к сечению этой поверхности плоскостью $y = y_0$ (рис.2.1)

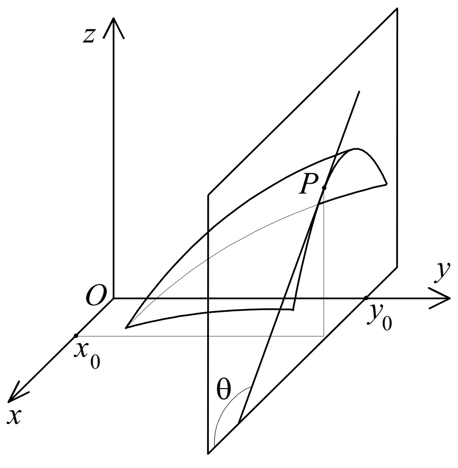

Рис. 2.1. Геометрический смысл частной производной

Перейдем к рассмотрению общего случая. Пусть функция n переменных

$f(\boldsymbol{x}) = f(x_1, \dots, x_n)$ определена в некоторой окрестности $u(\boldsymbol{x}^{(0)})$ точки $\boldsymbol{x}^{(0)} = (x_1^{(0)}, \dots, x_n^{(0)})$. Пусть, далее, h_i – вещественное число, называемое *приращением переменной* x_i, такое, что точка $\boldsymbol{x}(x_1^{(0)}, \dots, x_i^{(0)} + h_i, \dots, x_n^{(0)})$ принадлежит окрестности $u(\boldsymbol{x}^{(0)})$. Назовем *частным приращением функции* $f(x)$ *по переменной* x_i *в точке* $\boldsymbol{x}^{(0)}$, вызванным приращением h_i, разность

$$\Delta_i f(\boldsymbol{x}^{(0)}) = f(x_1^{(0)}, \dots, x_i^{(0)} + h_i, \dots, x_n^{(0)}) - f(x_1^{(0)}, \dots, x_n^{(0)}).$$

Для $h_i \neq 0$ отношение $\dfrac{\Delta_i f(\boldsymbol{x}^{(0)})}{h_i}$ представляет собой функцию одной переменной h_i, определенную в проколотой окрестности $\dot{u}(0)$.

Определение 2. Пусть функция $f(x)$ определена в окрестности $u(\boldsymbol{x}^{(0)})$ точки $\boldsymbol{x}^{(0)} \in \mathbf{R}^n$. Если существует предел $\lim\limits_{h_i \to 0} \dfrac{\Delta_i f(\boldsymbol{x}^{(0)})}{h_i}$, то он называется *частной производной функции* $f(\boldsymbol{x}) = f(x_1, \dots, x_n)$ *по переменной* x_i *в точке* $\boldsymbol{x}^{(0)}$ и обозначается $\dfrac{\partial f(\boldsymbol{x}^{(0)})}{\partial x_i}$ или $f'_{x_i}(\boldsymbol{x}^{(0)})$, $i = 1, 2, \dots, n$.

Таким образом, функция n переменных может иметь ровно n частных производных (по числу своих аргументов). Вычисление частной производной $\dfrac{\partial f}{\partial x_i}$ сводится, как и в случае функции двух переменных, к формальному нахождению производной от функции $f(x_1, \dots, x_n)$ по переменной x_i (остальные переменные $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ при этом предполагаются фиксированными, т.е. рассматриваются как константы).

Пример 1. Для функции четырех переменных $f(x_1, x_2, x_3, x_4) = e^{x_1^2 x_2} \sin(2x_3 - x_4)$ согласно сформулированному принципу имеем

$$f'_{x_1}(x_1, x_2, x_3, x_4) = 2x_1 x_2 e^{x_1^2 x_2} \sin(2x_3 - x_4), \qquad f'_{x_2}(x_1, x_2, x_3, x_4) = x_1^2 e^{x_1^2 x_2} \sin(2x_3 - x_4)$$

,

$$f'_{x_3}(x_1, x_2, x_3, x_4) = 2e^{x_1^2 x_2} \cos(2x_3 - x_4), \qquad f'_{x_4}(x_1, x_2, x_3, x_4) = -e^{x_1^2 x_2} \cos(2x_3 - x_4).$$

2°. Производная по направлению. Обобщением понятия частной производной является так называемая производная по направлению. Рассмотрим сначала функцию $f(x, y)$ двух переменных, определенную в некоторой окрестности фиксированной точки (x, y).

Из векторной алгебры известно, что произвольный двумерный вектор $\boldsymbol{e}$ единичной длины ($\|\boldsymbol{e}\|=1$) записывается в следующем виде: $\boldsymbol{e}=(\cos\alpha, \cos\beta)$. Проекции (координаты) этого вектора называются *направляющими косинусами* и связаны равенством (см. рис.2.2):

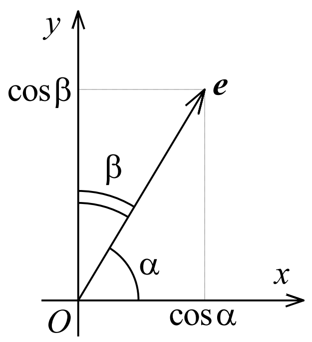

Рис. 2.2. Направляющие косинусы двумерного вектора.

$$\cos^2\alpha+\cos^2\beta=1=\|\boldsymbol{e}\|.$$

Определение 3. Если существует предел

$$\lim_{h\to+0}\frac{f(x+h\cos\alpha,\ y+h\cos\beta)-f(x,y)}{h},$$

то он называется *производной функции* $f(x,y)$ *в точке* (x,y) *по направлению* $\boldsymbol{e}$ и обозначается: $\dfrac{\partial f(x,y)}{\partial\boldsymbol{e}}$.

Непосредственно из этого определения видно, что производная по направлению представляет собой правостороннюю производную в точке $h=0$ от функции $f(x+h\cos\alpha,\ y+h\cos\beta)$, если ее рассматривать как функцию одной переменной h.

Поскольку дробь $\dfrac{f(x+h\cos\alpha,\ y+h\cos\beta)-f(x,y)}{h}$ есть отношение приращения функции $f(x,y)$ в точке (x,y) в направлении $\boldsymbol{e}$ к величине h, вызвавшей это приращение, предел этого отношения (т.е. производная по направлению $\boldsymbol{e}$), характеризует скорость изменения функции $f(x,y)$ в указанной точке в данном направлении.

Рассмотрим производную функции $f(x,y)$ в точке (x,y) по направлению $\boldsymbol{e}_1=(1,0)$ (вдоль положительной части оси Ox). Имеем:

$$\frac{\partial f(x,y)}{\partial\boldsymbol{e}_1}=\lim_{h\to+0}\frac{f(x+h,\ y)-f(x,y)}{h}=\frac{\partial f(x,y)}{\partial x},$$

т.е. если в точке (x,y) существует частная производная $\dfrac{\partial f(x,y)}{\partial x}$, то она совпадает с производной по направлению $\boldsymbol{e}_1$ (т.е. вдоль положительной части оси Ox). Однако производная по противоположному направлению $\boldsymbol{e}_2=(-1,0)$, т.е. в направлении отрицательной части оси Ox, отличается от частной производной $\dfrac{\partial f(x,y)}{\partial x}$ знаком:

$$\frac{\partial f(x,y)}{\partial\boldsymbol{e}_2}=\lim_{h\to+0}\frac{f(x-h,\ y)-f(x,y)}{h}=-\lim_{h\to+0}\frac{f(x-h,\ y)-f(x,y)}{(-h)}=$$

$$=-\lim_{t\to-0}\frac{f(x+t,\ y)-f(x,y)}{t}=-\frac{\partial f(x,y)}{\partial x}.$$

Точно так же частная производная по переменной y связана с производными по направлениям $\boldsymbol{e}_3=(0,1)$ и $\boldsymbol{e}_4=(0,-1)$:

$$\frac{\partial f(x,y)}{\partial \boldsymbol{e}_3}=\frac{\partial f(x,y)}{\partial y}, \quad \frac{\partial f(x,y)}{\partial \boldsymbol{e}_4}=-\frac{\partial f(x,y)}{\partial y}.$$

Перенесем понятие производной по направлению на случай функции $f(\boldsymbol{x})=f(x_1,\dots,x_n)$ n переменных. В приводимом ниже определении мы, кроме того, откажемся от требования к вектору, определяющему направление, иметь единичную длину.

Определение 4. Пусть функция $f(x)$ определена в некоторой окрестности точки $\boldsymbol{x}\in\mathbf{R}^n$, и пусть $\boldsymbol{y}$ – произвольный ненулевой вектор пространства $\mathbf{R}^n$. Если существует предел $\lim\limits_{h\to+0}\dfrac{f(\boldsymbol{x}+h\boldsymbol{y})-f(\boldsymbol{x})}{h}$, то он называется *производной функции* $f(x)$ *в точке* $\boldsymbol{x}$ *по направлению* $\boldsymbol{y}$ и обозначается $\dfrac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{y}}$ или $f'(\boldsymbol{x},\boldsymbol{y})$.

Как видно из этого определения, производная по направлению $\boldsymbol{y}=(y_1,\dots,y_n)$ представляет собой правостороннюю производную в точке $h=0$ от функции $f(x_1+hy_1,\dots,x_n+hy_n)$, если ее рассматривать как функцию одной переменной h.

Если вектор $\boldsymbol{y}$, определяющий направление, имеет единичную длину, то производная по направлению функции n переменных характеризует скорость ее изменения в заданной точке в направлении $\boldsymbol{y}$.

Пример 2. Пусть функция $f(x)$ имеет в точке $\boldsymbol{x}$ частную производную по переменной x_i, и пусть $\boldsymbol{e}_i=(0,\dots,\underset{i}{1},\dots,0)$. Вектор $\boldsymbol{e}_i$ определяет направление положительной части оси Ox_i. Тогда производная по направлению $\boldsymbol{e}_i$ совпадает с частной производной $\dfrac{\partial f(\boldsymbol{x})}{\partial x_i}$. Действительно,

$$f'(\boldsymbol{x},\boldsymbol{e}_i)=\lim_{h\to+0}\frac{f(x_1,\dots,x_i+h,\dots,x_n)-f(x_1,\dots,x_n)}{h}=\frac{\partial f(\boldsymbol{x})}{\partial x_i}.$$

3°. Дифференцируемые функции двух переменных. Рассмотрим функцию двух переменных $f(x,y)$, определенную в некоторой окрестности $u(M_0)$ точки $M_0(x_0,y_0)$. Пусть вещественные числа Δx и Δy (приращения переменных x и y соответственно) таковы, что точка $M(x_0+\Delta x, y_0+\Delta y)$ также содержится в окрестности $u(M_0)$. Разность

$$\Delta f(x_0,y_0)=f(x_0+\Delta x, y_0+\Delta y)-f(x_0,y_0),$$

рассматриваемая как функция двух переменных Δx и Δy, называется *полным приращением функции* $f(x,y)$ *в точке* $M_0(x_0,y_0)$.

Определение 5. Функция $f(x,y)$, определенная в окрестности $u(M_0)$ точки $M_0(x_0, y_0)$, называется *дифференцируемой* в этой точке, если существуют два таких числа A и B, что разность $\Delta f(x_0,y_0) - A\Delta x - B\Delta y$, представляющая собой функцию двух переменных Δx и Δy, является бесконечно малой более высокого порядка, чем $\rho = \sqrt{\Delta x^2 + \Delta y^2}$ при $\Delta x \to 0$, $\Delta y \to 0$.

Из неравенств $0 \le |\Delta x| \le \rho$, $0 \le |\Delta y| \le \rho$ следует, что одновременное стремление к нулю приращений Δx и Δy равносильно условию $\rho \to 0$. Используя для обозначения бесконечно малой функции более высокого порядка символ $o(\rho)$ (“o-малое от ρ”), условие дифференцируемости из определения 5 может быть записано в виде равенства $\Delta f(x_0,y_0) - A\Delta x - B\Delta y = o(\rho)$, или

$$\Delta f(x_0, y_0) = A\Delta x + B\Delta y + o(\rho), \tag{1}$$

причем $\lim\limits_{\rho\to 0}\dfrac{o(\rho)}{\rho} = 0$.

Равенству (1) можно придать несколько иную, эквивалентную форму. С этой целью введем бесконечно малую функцию $\lambda(\Delta x, \Delta y) = \Delta f(x_0,y_0) - A\Delta x - B\Delta y$. Если приращения аргументов Δx и Δy равны нулю, то равно нулю и полное приращение функции $\Delta f(x_0, y_0)$. Следовательно, $\lambda(0, 0) = 0$.

Рассмотрим окрестность точки $(0, 0)$ и определим в этой окрестности функцию $\varepsilon(\Delta x, \Delta y)$ двух переменных Δx и Δy следующим равенством:

$$\varepsilon(\Delta x, \Delta y) = \begin{cases} \dfrac{\lambda(\Delta x, \Delta y)}{\rho}, & \text{если} \quad \Delta x^2 + \Delta y^2 \ne 0, \\ 0, & \text{если} \quad \Delta x = \Delta y = 0. \end{cases}$$

Эта функция, очевидно, непрерывна в точке $(0, 0)$:

$$\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}} \varepsilon(\Delta x, \Delta y) = \lim_{\rho\to 0}\frac{\lambda(\Delta x, \Delta y)}{\rho} = 0 = \varepsilon(0, 0).$$

Таким образом, для любых значений приращений Δx и Δy из указанной окрестности, в том числе и равных нулю, справедливо равенство $\lambda(\Delta x, \Delta y) = \varepsilon(\Delta x, \Delta y)\cdot\rho$. Теперь условие (1) дифференцируемости функции $f(x,y)$ в точке (x_0, y_0) можно переписать в виде равенства

$$\Delta f(x_0, y_0) = A\Delta x + B\Delta y + \varepsilon(\Delta x, \Delta y)\cdot\rho, \tag{2}$$

где $\varepsilon(\Delta x, \Delta y)$ – непрерывная в точке $(0, 0)$ бесконечно малая функция, $\varepsilon(0,0) = 0$, и

$\rho = \sqrt{\Delta x^2 + \Delta y^2}$. В дальнейшем будут использоваться обе формы записи условия дифференцируемости (формулы (1) и (2)).

Переходя к рассмотрению некоторых свойств дифференцируемых функций двух переменных, отметим прежде всего, что так же, как это имело место для функций одной переменной, дифференцируемость функции двух переменных влечет за собой ее непрерывность в данной точке.

Теорема 1. Если функция двух переменных $f(x,y)$ дифференцируема в точке (x_0, y_0), то она непрерывна в этой точке.

▶ Рассмотрим предел функции $f(x,y)$ в точке (x_0, y_0) и обозначим через $\Delta x = x - x_0$, $\Delta y = y - y_0$ приращения ее аргументов. Используя определение непрерывности и условие дифференцируемости (2), имеем

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} f(x,y) = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} f(x_0 + \Delta x, y_0 + \Delta y) = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \big(\Delta f(x_0, y_0) + f(x_0, y_0)\big) =$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \big(A\Delta x + B\Delta y + \varepsilon(\Delta x, \Delta y)\cdot\rho\big) + f(x_0, y_0) = f(x_0, y_0).$$ ◀

Следующее утверждение проясняет смысл коэффициентов A и B в определении 5: эти числа (если они существуют) определяются однозначно и представляют собой частные производные функции $f(x,y)$ в рассматриваемой точке.

Теорема 2. Пусть функция двух переменных $f(x,y)$ дифференцируема в точке (x_0, y_0), т.е. выполнено условие (1) или равносильное ему условие (2). Тогда функция $f(x,y)$ имеет в этой точке частные производные, причем $\dfrac{\partial f(x_0, y_0)}{\partial x} = A$, $\dfrac{\partial f(x_0, y_0)}{\partial y} = B$, где A и B – коэффициенты из условия дифференцируемости.

▶ Положив $\Delta y = 0$ в формуле (2), получим

$$f(x_0 + \Delta x, y_0) - f(x_0, y_0) = A\,\Delta x + \varepsilon(\Delta x, 0)\left|\Delta x\right|.$$

В этом равенстве левая часть представляет собой частное приращение $\Delta_x f(x_0, y_0)$ функции $f(x,y)$ по переменной x в точке (x_0, y_0), а функция одной переменной $\varepsilon(\Delta x, 0)$ – бесконечно малая при $\Delta x \to 0$, поскольку

$$\lim_{\Delta x \to 0} \varepsilon(\Delta x, 0) = \varepsilon(0, 0) = 0.$$

Применяя известную из теории пределов теорему о произведении бесконечно малой функции на ограниченную, по определению частной производной (определение 1) получим

$$\frac{\partial f(x_0, y_0)}{\partial x} = \lim_{\Delta x \to 0} \frac{\Delta_x f(x_0, y_0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{A\,\Delta x + \varepsilon(\Delta x, 0)\,|\Delta x|}{\Delta x} = A + \lim_{\Delta x \to 0} \varepsilon(\Delta x, 0) \cdot \frac{|\Delta x|}{\Delta x} = A,$$

так как функция $\varphi(\Delta x) = \frac{|\Delta x|}{\Delta x}$ принимает всего два значения (± 1), а значит, является ограниченной. Таким образом, доказано существование производной $\frac{\partial f(x_0, y_0)}{\partial x}$ и ее совпадение с коэффициентом A формулы (2). Аналогично доказывается существование частной производной $\frac{\partial f(x_0, y_0)}{\partial y}$ и равенство $\frac{\partial f(x_0, y_0)}{\partial y} = B$. ◀

С учетом доказанной теоремы условия дифференцируемости (1) и (2) можно переписать соответственно в следующем виде:

$$\Delta f(x_0, y_0) = f'_x(x_0, y_0)\,\Delta x + f'_y(x_0, y_0)\,\Delta y + o(\rho) \tag{3}$$

и

$$\Delta f(x_0, y_0) = f'_x(x_0, y_0)\,\Delta x + f'_y(x_0, y_0)\,\Delta y + \varepsilon(\Delta x, \Delta y)\,\rho\,. \tag{4}$$

Существование частных производных согласно теореме 2 является необходимым условием дифференцируемости. Однако оно не является достаточным условием. Приведем пример функции двух переменных, не являющейся дифференцируемой в точке, но имеющей в ней обе частные производные.

Пример 3.

$$f(x, y) = \begin{cases} 0, & \text{если} \quad x = y = 0, \\ \dfrac{xy}{x^2 + y^2}, & \text{если} \quad x^2 + y^2 \neq 0. \end{cases}$$

▶ Нетрудно проверить, что частные производные этой функции в точке $(0, 0)$ равны нулю:

$$f'_x(0, 0) = \lim_{\Delta x \to 0} \frac{\Delta_x f(0, 0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(\Delta x, 0) - f(0, 0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{0 - 0}{\Delta x} = 0.$$

Аналогично, $f'_y(0, 0) = 0$. С другой стороны, в предыдущей главе (см. пример 2 §2,) было показано, что рассматриваемая функция не имеет предела в точке $(0, 0)$, а значит, не является непрерывной в этой точке. В таком случае, в силу теоремы 1, эта функция не может быть дифференцируемой в точке $(0, 0)$. ◀

Заметим, что существуют функции, которые имеют производные по любому направлению, но не являются дифференцируемыми (например, $f(x, y) = \sqrt{x^2 + y^2}$ в точке $(0, 0)$).

Поскольку проверить дифференцируемость той или иной конкретной функции, используя только определение дифференцируемости, достаточно сложно, в следующей теореме приводятся достаточные условия дифференцируемости функции двух переменных в точке.

Теорема 3. Пусть функция $f(x,y)$ определена в некоторой окрестности $u(M_0)$ точки $M_0(x_0, y_0)$ и имеет частные производные $f'_x(x,y)$ и $f'_y(x,y)$ в каждой точке (x,y) этой окрестности. В таком случае эти частные производные также представляют собой функции двух переменных, определенные в окрестности $u(M_0)$.

Пусть $f'_x(x,y)$ и $f'_y(x,y)$, рассматриваемые как функции переменных x и y, непрерывны в точке (x_0, y_0). Тогда функция $f(x,y)$ дифференцируема в этой точке.

▶ Окрестность $u(M_0)$ представляет собой внутренность круга с центром в точке (x_0, y_0) (определение 4 §1 гл.1). Обозначим радиус этого круга через r ($r>0$) и заметим, что если $\rho = \sqrt{\Delta x^2 + \Delta y^2} < r$, то точки $M_0(x_0, y_0)$, $M_1(x_0, y_0+\Delta y)$, $M(x_0+\Delta x, y_0+\Delta y)$, а также соединяющие их отрезки находятся в рассматриваемой окрестности $u(M_0)$ (рис.2.3).

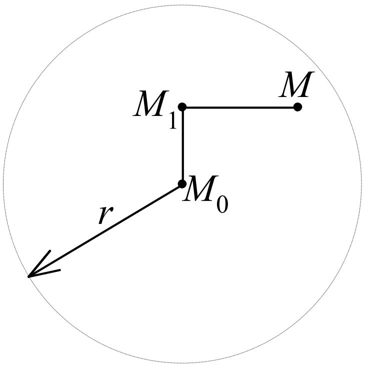

Рис. 2.3. К доказательству теоремы 3

Введем функцию одной переменной, определенную равенством

$$\varphi(y) = f(x_0, y)$$

для всех таких значений y, для которых $(x_0, y) \in u(M_0)$, т.е. для всех y из промежутка $(y_0 - r, y_0 + r)$. Функция $\varphi(y)$ дифференцируема на этом промежутке, причем $\varphi'(y) = f'_y(x_0, y)$, ибо

$$\varphi'(y) = \lim_{h\to 0}\frac{\varphi(y+h)-\varphi(y)}{h} = \lim_{h\to 0}\frac{f(x_0, y+h)-f(x_0, y)}{h} = f'_y(x_0, y)$$

на основании определения производной функции одной переменной и определения частной производной. Применим к функции $\varphi(y)$, рассматриваемой на отрезке с концами в точках y_0 и $y_0+\Delta y$, формулу конечных приращений Лагранжа. Согласно этой формуле существует такое число $\theta_1 \in (0,1)$, что выполняется равенство

$$\varphi(y_0+\Delta y) - \varphi(y_0) = \varphi'(y_0+\theta_1\Delta y)\Delta y,$$

откуда, переходя к функции $f(x,y)$, получаем

$$f(x_0, y_0+\Delta y) - f(x_0, y_0) = f'_y(x_0, y_0+\theta_1\Delta y)\,\Delta y.$$

Аналогичными рассуждениями устанавливается справедливость равенства

$$f(x_0+\Delta x, y_0+\Delta y) - f(x_0, y_0+\Delta y) = f'_x(x_0+\theta_2\Delta x, y_0+\Delta y)\,\Delta x,$$

где число θ_2 также из интервала $(0,1)$.

Перейдем теперь к рассмотрению полного приращения $\Delta f(x_0, y_0)$ функции $f(x, y)$. Имеем

$$\Delta f(x_0, y_0) = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) =$$
$$= f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0 + \Delta y) + f(x_0, y_0 + \Delta y) - f(x_0, y_0) =$$
$$= f'_x(x_0 + \theta_2 \Delta x, y_0 + \Delta y)\, \Delta x + f'_y(x_0, y_0 + \theta_1 \Delta y)\, \Delta y .$$

Введем следующие обозначения:

$$\varepsilon_1(\Delta x, \Delta y) = f'_y(x_0, y_0 + \theta_1 \Delta y) - f'_y(x_0, y_0)$$

(ввиду независимости от Δx правой части допустимо также обозначение $\varepsilon_1(0, \Delta y)$) и

$$\varepsilon_2(\Delta x, \Delta y) = f'_x(x_0 + \theta_2 \Delta x, y_0 + \Delta y) - f'_y(x_0, y_0) .$$

В силу ограниченности θ_1 $(0 < \theta_1 < 1)$, на основании непрерывности функции $f'_y(x, y)$ в точке (x_0, y_0) можно записать

$$\lim_{\rho \to 0} \varepsilon_1(\Delta x, \Delta y) = \lim_{\Delta y \to 0} f'_y(x_0, y_0 + \theta_1 \Delta y) - f'_y(x_0, y_0) = f'_y(x_0, y_0) - f'_y(x_0, y_0) = 0,$$

и, аналогично,

$$\lim_{\rho \to 0} \varepsilon_2(\Delta x, \Delta y) = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} f'_x(x_0 + \theta_2 \Delta x, y_0 + \Delta y) - f'_x(x_0, y_0) = f'_x(x_0, y_0) - f'_x(x_0, y_0) = 0$$

Таким образом, приращение $\Delta f(x_0, y_0)$ представлено в виде

$$\Delta f(x_0, y_0) = f'_x(x_0, y_0)\Delta x + f'_y(x_0, y_0)\Delta y + \varepsilon_2(\Delta x, \Delta y)\Delta x + \varepsilon_1(\Delta x, \Delta y)\Delta y, \tag{5}$$

где $\varepsilon_1(\Delta x, \Delta y)$ и $\varepsilon_2(\Delta x, \Delta y)$ – бесконечно малые при $\rho \to 0$.

Для завершения доказательства осталось проверить, что сумма двух последних слагаемых в формуле (5) есть бесконечно малая более высокого порядка, чем ρ, при $\rho \to 0$ (см. формулу 3).

Заметим, что функции $\dfrac{\Delta x}{\sqrt{\Delta x^2 + \Delta y^2}}$ и $\dfrac{\Delta y}{\sqrt{\Delta x^2 + \Delta y^2}}$ при условии $\rho = \sqrt{\Delta x^2 + \Delta y^2} \neq 0$ ограничены по абсолютной величине единицей, и обратимся к определению предела функции двух переменных (см. определение 3 §2 гл.1).

Для любого $\varepsilon > 0$ существует окрестность $\overset{\circ}{u}_{\delta_1}(0, 0)$, такая, что для всех точек $(\Delta x, \Delta y)$ этой окрестности выполняется неравенство

$$\left|\varepsilon_1(\Delta x, \Delta y)\right| < \frac{\varepsilon}{2} .$$

Для того же ε существует окрестность $\overset{\circ}{u}_{\delta_2}(0, 0)$, такая, что для всех точек $(\Delta x, \Delta y)$ из $\overset{\circ}{u}_{\delta_2}(0, 0)$ справедливо неравенство

$$\left|\varepsilon_2(\Delta x, \Delta y)\right| < \frac{\varepsilon}{2}.$$

В таком случае, для всех точек $(\Delta x, \Delta y) \in \overset{\circ}{u}_{\delta_1}(0,0) \bigcap \overset{\circ}{u}_{\delta_2}(0,0)$ выполнено

$$\left|\frac{\varepsilon_1(\Delta x, \Delta y)\Delta y + \varepsilon_2(\Delta x, \Delta y)\Delta x}{\rho}\right| \le \left|\varepsilon_1(\Delta x, \Delta y)\right| \frac{|\Delta y|}{\rho} + \left|\varepsilon_2(\Delta x, \Delta y)\right| \frac{|\Delta x|}{\rho} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

а это, согласно тому же определению предела функции двух переменных, означает, что

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{\varepsilon_1(\Delta x, \Delta y)\Delta y + \varepsilon_2(\Delta x, \Delta y)\Delta x}{\rho} = 0,$$

или, другими словами, $\varepsilon_1(\Delta x, \Delta y)\Delta y + \varepsilon_2(\Delta x, \Delta y)\Delta x = o(\rho)$. ◀

Пример 4. Рассмотрим функцию $f(x, y) = y^2 \ln x$, определенную на открытом множестве $D = \{(x, y) \mid x > 0\}$. В каждой точке множества D эта функция имеет частные производные

$$f'_x(x, y) = \frac{y^2}{x}, \quad f'_y(x, y) = 2y \ln x,$$

которые, очевидно, представляют собой непрерывные функции на множестве D. Согласно теореме 3, функция $f(x, y) = y^2 \ln x$ дифференцируема в каждой точке своей области определения.

4°. Дифференцирование сложной функции. В предыдущей главе (п.3° §2 гл.1) было введено понятие сложной функции (суперпозиции) и была доказана теорема, устанавливающая "наследственный" характер непрерывности функции в точке. Дифференцируемость функции нескольких переменных в точке по отношению к операции суперпозиции носит тот же характер.

Рассмотрим сначала функцию $f(x, y)$ двух переменных x и y, каждая из которых, в свою очередь, является функцией независимой переменной t: $x = x(t)$, $y = y(t)$. Тогда функция $f\big(x(t), y(t)\big)$ является сложной функцией независимой переменной t, а переменные x и y – промежуточными переменными.

Теорема 4. Пусть функции $x(t)$ и $y(t)$ дифференцируемы в точке t_0, а функция $f(x, y)$ дифференцируема в точке (x_0, y_0), где $x_0 = x(t_0)$, $y_0 = y(t_0)$. Тогда сложная функция $f\big(x(t), y(t)\big)$ дифференцируема в точке t_0, причем для ее производной справедливо равенство

$$\left.\frac{df\big(x(t),\,y(t)\big)}{dt}\right|_{t=t_0} = f'_x(x_0,\,y_0)\cdot x'(t_0)+f'_y(x_0,\,y_0)\cdot y'(t_0)\,. \qquad (6)$$

▶ Поскольку функция $f(x,y)$ дифференцируема в точке $(x_0,\,y_0)$, ее полное приращение в этой точке, согласно формуле (4), может быть записано в следующем виде:

$$\Delta f(x_0,\,y_0)=f'_x(x_0,\,y_0)\Delta x+f'_y(x_0,\,y_0)\Delta y+\varepsilon\,(\Delta x,\,\Delta y)\,\rho\,, \qquad (7)$$

где $\rho=\sqrt{\Delta x^2+\Delta y^2}$, а $\lim\limits_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\varepsilon(\Delta x,\Delta y)=0$.

Для произвольного (достаточно малого) $\Delta t\neq 0$ соответствующие приращения x и y в точке t_0 представляют собой разности:

$$\begin{aligned}\Delta x &= x\,(t_0+\Delta t)-x\,(t_0),\\ \Delta y &= y\,(t_0+\Delta t)-y\,(t_0).\end{aligned} \qquad (8)$$

Тогда функция $\varepsilon(\Delta x,\Delta y)$, рассматриваемая с учетом (8) как функция одной переменной Δt, есть бесконечно малая функция при $\Delta t\to 0$. В самом деле,

$$\lim_{\Delta t\to 0}\varepsilon(\Delta x,\Delta y)=\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\varepsilon(\Delta x,\Delta y)=\lim_{\rho\to 0}\varepsilon(\Delta x,\Delta y)=0\,.$$

Пусть теперь в равенстве (7) приращения аргументов Δx и Δy определяются формулами (8), т.е. зависят от приращения Δt.

Перейдем к вычислению производной сложной функции $f\big(x(t),\,y(t)\big)$ в точке t_0. Для этого используем определение производной и формулу (4):

$$\left.\frac{d\,f\big(x(t),y(t)\big)}{dt}\right|_{t=t_0}=\lim_{\Delta t\to 0}\frac{f\big(x(t_0+\Delta t),y(t_0+\Delta t)\big)-f\big(x(t_0),y(t_0)\big)}{\Delta t}=$$

$$=\lim_{\Delta t\to 0}\frac{f\,(x_0+\Delta x,\,y_0+\Delta y)-f\,(x_0,\,y_0)}{\Delta t}=\lim_{\Delta t\to 0}\frac{\Delta f\,(x_0,\,y_0)}{\Delta t}=$$

$$=\lim_{\Delta t\to 0}\frac{f'_x(x_0\,,\,y_0)\Delta x+f'_y(x_0\,,\,y_0)\Delta y+\varepsilon\,(\Delta x,\Delta y)\,\rho}{\Delta t}=$$

$$=f'_x(x_0,\,y_0)\lim_{\Delta t\to 0}\frac{x(t_0+\Delta t)-x(t_0)}{\Delta t}+f'_y(x_0,\,y_0)\lim_{\Delta t\to 0}\frac{y(t_0+\Delta t)-y(t_0)}{\Delta t}+\lim_{\Delta t\to 0}\frac{\varepsilon(\Delta x,\Delta y)\rho}{\Delta t}=$$

$$=f'_x(x_0,\,y_0)\,x'(t_0)+f'_y(x_0,\,y_0)\,y'(t_0)+\lim_{\Delta t\to 0}\left[\varepsilon\,(\Delta x,\Delta y)\cdot\frac{\rho}{\Delta t}\right].$$

Для завершения доказательства осталось показать, что последнее слагаемое равно нулю. Воспользуемся теоремой о произведении бесконечно малой функции на ограниченную функцию: произведение двух функций одной переменной, из которых одна – бесконечно малая в рассматриваемой точке, другая ограничена, само является бесконечно малой функцией в этой точке. Выше отмечалось, что функция $\varepsilon(\Delta x, \Delta y)$, рассматриваемая как функция одной переменной Δt есть бесконечно малая при $\Delta t \to 0$ и, следовательно, достаточно установить ограниченность функции $\frac{\rho}{\Delta t}$. Действительно, функция $\frac{\rho}{|\Delta t|}$ как функция от переменной Δt имеет предел в точке 0:

$$\lim_{\Delta t\to 0}\frac{\rho}{|\Delta t|} = \lim_{\Delta t\to 0}\frac{\sqrt{\Delta x^2+\Delta y^2}}{|\Delta t|} = \lim_{\Delta t\to 0}\sqrt{\left(\frac{x(t_0+\Delta t)-x(t_0)}{\Delta t}\right)^2+\left(\frac{y(t_0+\Delta t)-y(t_0)}{\Delta t}\right)^2} =$$

$$= \sqrt{\left(x'(t_0)\right)^2+\left(y'(t_0)\right)^2},$$

а, значит, является ограниченной в окрестности точки 0. В таком случае ограничена в той же окрестности и функция $\frac{\rho}{\Delta t}$. ◄

Введем обозначение для зависимой переменной функции двух переменных $f(x,y)$, положив

$$z = f(x,y).$$

Тогда суперпозиция функции $z = f(x,y)$ и функций $x = x(t)$, $y = y(t)$ примет вид

$$z = f\big(x(t),\, y(t)\big).$$

В условиях теоремы 4 формулу (6) часто записывают следующим образом:

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\cdot\frac{dx}{dt} + \frac{\partial z}{\partial y}\cdot\frac{dy}{dt}, \tag{9}$$

причем в левой части равенства переменная z рассматривается как зависимая переменная аргумента t, а в правой части она же – как зависимая переменная двух переменных x и y.

Перейдем к рассмотрению более общего случая. Пусть заданы две функции $x = x(u,v)$ и $y = y(u,v)$ двух переменных u и v, дифференцируемые в точке (u_0,v_0), а функция $z = f(x,y)$ двух переменных x и y дифференцируема в точке (x_0,y_0), где $x_0 = x(u_0,v_0)$, $y_0 = y(u_0,v_0)$. При сделанных предположениях суперпозиция $z = f\big(x(u,v),\, y(u,v)\big)$, рассматриваемая как функция двух переменных u и v, имеет частные производные в точке (u_0,v_0) по каждой из них. Эти частные производные могут быть вычислены по формулам, аналогичным формуле (9):

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\cdot\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial u}, \tag{10}$$

$$\frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\cdot\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial v}. \tag{11}$$

▶ Действительно, если зафиксировать одну из переменных, положив, например, $v = v_0$, функция $z = f\big(x(u,v_0), y(u,v_0)\big)$ оказывается функцией одной переменной u. Применяя формулу (6) из теоремы 4 к этой ситуации, для производной сложной функции в точке u_0 имеем

$$\left.\frac{dz}{du}\right|_{u=u_0} = f'_x(x_0, y_0)\left.\frac{dx}{du}\right|_{u=u_0} + f'_y(x_0, y_0)\left.\frac{dy}{du}\right|_{u=u_0}.$$

Из этого равенства непосредственно вытекает формула (10). В самом деле, частная производная функции нескольких переменных по одной из переменных, как отмечалось в п.1°, представляет собой обычную производную по этой переменной при фиксированных значениях остальных переменных, и, следовательно, справедливы следующие равенства:

$$\left.\frac{dx}{du}\right|_{u=u_0} = \left.\frac{dx(u,v_0)}{du}\right|_{u=u_0} = \frac{\partial x(u_0,v_0)}{\partial u},$$

$$\left.\frac{dy}{du}\right|_{u=u_0} = \left.\frac{dy(u,v_0)}{du}\right|_{u=u_0} = \frac{\partial y(u_0,v_0)}{\partial u},$$

$$\left.\frac{dz}{du}\right|_{u=u_0} = \left.\frac{dz\big(x(u,v_0), y(u,v_0)\big)}{du}\right|_{u=u_0} = \frac{\partial z(u_0,v_0)}{\partial u}.$$

Таким образом, полученные выше выражения для производной $\left.\dfrac{dz}{du}\right|_{u=u_0}$ можно переписать в виде:

$$\frac{\partial z(u_0,v_0)}{\partial u} = f'_x(x_0, y_0)\cdot\frac{\partial x(u_0,v_0)}{\partial u} + f'_y(x_0, y_0)\cdot\frac{\partial y(u_0,v_0)}{\partial u},$$

или, в других обозначениях,

$$\frac{\partial z(u_0,v_0)}{\partial u} = \frac{\partial z(x_0, y_0)}{\partial x}\cdot\frac{\partial x(u_0,v_0)}{\partial u} + \frac{\partial z(x_0, y_0)}{\partial y}\cdot\frac{\partial y(u_0,v_0)}{\partial u}.$$

Формула (11) устанавливается аналогично. ◀

5°. Дифференцируемые функции нескольких переменных. Дифференциал. В этом пункте понятия дифференцируемой функции, а также результаты двух предыдущих пунктов, переносятся на общий случай функций нескольких переменных.

Рассмотрим функцию n переменных $f(\boldsymbol{x}) = f(x_1, \dots, x_n)$, определенную в некоторой окрестности $u(\boldsymbol{x}_0) \subset \mathbf{R}^n$ точки $\boldsymbol{x}_0 = (x_1^{(0)}, \dots, x_n^{(0)})$, и пусть вектор $\boldsymbol{h} = (h_1, \dots, h_n)$ выбирается так, что точка $\boldsymbol{x}_0 + \boldsymbol{h}$ также содержится в окрестности $u(\boldsymbol{x}_0)$. *Полным приращением функции* $f(\boldsymbol{x})$ *в точке* $\boldsymbol{x}_0$ называется разность

$$\Delta f(\boldsymbol{x}_0) = f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0) = f(x_1^{(0)} + h_1, \dots, x_n^{(0)} + h_n) - f(x_1^{(0)}, \dots, x_n^{(0)}) .$$

Следующее определение является обобщением определения 5 из п.3°.

Определение 6. Функция $f(\boldsymbol{x})$, определенная в некоторой окрестности $u(\boldsymbol{x}_0)$ точки $\boldsymbol{x}_0$ называется *дифференцируемой в точке* $\boldsymbol{x}_0$, если существуют такие числа $A_1, \dots, A_n$, что разность $\Delta f(\boldsymbol{x}_0) - \sum\limits_{i=1}^{n} A_i h_i$, рассматриваемая как функция переменных $h_1, \dots, h_n$, представляет собой бесконечно малую более высокого порядка, чем $\|\boldsymbol{h}\| = \sqrt{\sum\limits_{i=1}^{n} h_i^2}$, при $\|\boldsymbol{h}\| \to 0$, т.е.

$$\Delta f(\boldsymbol{x}_0) - \sum_{i=1}^{n} A_i h_i = o\left(\|\boldsymbol{h}\|\right),$$

или

$$\Delta f(\boldsymbol{x}_0) = \sum_{i=1}^{n} A_i h_i + o\left(\|\boldsymbol{h}\|\right). \tag{12}$$

Формула (12) является обобщением формулы (1) из п.3° и допускает, подобно ей, еще одну, эквивалентную, форму записи:

$$\Delta f(\boldsymbol{x}_0) = \sum_{i=1}^{n} A_i h_i + \varepsilon(h_1, \dots, h_n) \cdot \|\boldsymbol{h}\|, \tag{13}$$

где функция $\varepsilon(h_1, \dots, h_n)$ такова, что

$$\lim_{\|\boldsymbol{h}\| \to 0} \varepsilon(h_1, \dots, h_n) = 0 .$$

Так же, как для функций двух переменных, устанавливается, что дифференцируемость функции $f(\boldsymbol{x})$ в точке $\boldsymbol{x}_0$ влечет ее непрерывность в этой точке, а также выясняется, что коэффициенты $A_1, \dots, A_n$ из определения 6 представляют собой частные производные по соответствующим аргументам функции $f(\boldsymbol{x})$:

$$A_1 = \frac{\partial f(\boldsymbol{x}_0)}{\partial x_1}; \quad \dots \quad ; \; A_n = \frac{\partial f(\boldsymbol{x}_0)}{\partial x_n} .$$

При этом формулы (12) и (13) могут быть переписаны в следующем виде:

$$\Delta f(\boldsymbol{x}_0) = \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x}_0)}{\partial x_i} h_i + \mathrm{o}\left(\|\boldsymbol{h}\|\right) \tag{14}$$

и

$$\Delta f(\boldsymbol{x}_0) = \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x}_0)}{\partial x_i} h_i + \varepsilon(h_1, \dots, h_n) \cdot \|\boldsymbol{h}\|, \tag{15}$$

где, как и выше, функция $\varepsilon(h_1, \dots, h_n)$ является бесконечно малой в точке $\mathbf{0}$, т.е. $\lim\limits_{\|\boldsymbol{h}\|\to 0} \varepsilon(h_1, \dots, h_n) = 0$. Два последних равенства являются уточнением формул (12) и (13) определения 6.

По аналогии с теоремой 3 п.3° доказывается, что существование частных производных $\dfrac{\partial f(\boldsymbol{x})}{\partial x_i}$, $i = 1, \dots, n$ функции $f(\boldsymbol{x})$ в каждой точке некоторой окрестности точки $\boldsymbol{x}_0$ и их непрерывность в самой точке $\boldsymbol{x}_0$ является достаточным условием дифференцируемости функции $f(\boldsymbol{x})$ в точке $\boldsymbol{x}_0$. Приведем, наконец, соответствующие формулы, относящиеся к вычислению частных производных сложной функции (обобщение формул (9) – (11) п.4°).

Рассмотрим n функций $x_1 = x_1(\boldsymbol{t}) = x_1(t_1, \dots, t_m)$, $\dots$, $x_n = x_n(\boldsymbol{t}) = x_n(t_1, \dots, t_m)$, каждая из которых является дифференцируемой в точке $\boldsymbol{t}_0 = (t_1^{(0)}, \dots, t_m^{(0)})$, и пусть $\boldsymbol{x}_0 = \left(x_1(\boldsymbol{t}_0), \dots, x_n(\boldsymbol{t}_0)\right)$. Пусть далее функция $z = f(\boldsymbol{x}) = f(x_1, \dots, x_n)$ дифференцируема в точке $\boldsymbol{x}_0$. В таком случае сложная функция $z = f\left(x_1(\boldsymbol{t}), \dots, x_n(\boldsymbol{t})\right)$, рассматриваемая как функция нескольких переменных $t_1, \dots, t_m$, имеет частные производные по каждой из них, причем справедливы следующие равенства:

$$\frac{\partial x}{\partial t_i} = \sum_{k=1}^{n} \frac{\partial z}{\partial x_k} \cdot \frac{\partial x_k}{\partial t_i}, \quad i = 1, \dots, m. \tag{16}$$

Перейдем к обсуждению понятия дифференциала – еще одного понятия, связанного с функциями, дифференцируемыми в точке.

Пусть функция $f(\boldsymbol{x}) = f(x_1, \dots, x_n)$ дифференцируема в точке $\boldsymbol{x}_0$ и, следовательно, для ее полного приращения в этой точке справедливо равенство (14):

$$\Delta f(\boldsymbol{x}_0) = \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x}_0)}{\partial x_i} h_i + o\left(\|\boldsymbol{h}\|\right).$$

Определение 7. *Дифференциалом функции* $f(\boldsymbol{x})$ *в точке* $\boldsymbol{x}_0$, *соответствующим приращению аргументов* $\boldsymbol{h} = (h_1, \dots, h_n)$ называется сумма

$$\sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x}_0)}{\partial x_i} h_i,$$

которая обозначается символом $df(\boldsymbol{x}_0)$, или, если в записи функции участвовала зависимая переменная y ($y = f(\boldsymbol{x})$), то просто dy.

Таким образом, дифференциал функции нескольких переменных имеет вид

$$dy = df(\boldsymbol{x}_0) = \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x}_0)}{\partial x_i} h_i. \tag{17}$$

Отметим, что в обозначении $df(\boldsymbol{x}_0)$ отсутствует упоминание о тех приращениях аргументов $\boldsymbol{h} = (h_1, \dots, h_n)$, которые, собственно, и породили этот дифференциал, а в обозначении dy, кроме того, и сама точка $\boldsymbol{x}_0$, в которой он вычисляется, но, естественно, подразумевается наличие и того, и другого. Формула (17) показывает, что дифференциал представляет собой линейную функцию от n переменных $h_1, \dots, h_n$, в которой частные производные $\dfrac{\partial f(\boldsymbol{x}_0)}{\partial x_i}$ играют роль коэффициентов.

Обратимся теперь к сравнению определения дифференциала (определение 7) с определением дифференцируемости функции в точке (определение 6). Из сопоставления формулы (14) с формулой (17) имеем

$$\Delta f(\boldsymbol{x}_0) = df(\boldsymbol{x}_0) + o\left(\|\boldsymbol{h}\|\right). \tag{18}$$

Кроме того, очевидно,

$$\lim_{\|\boldsymbol{h}\|\to 0} df(\boldsymbol{x}_0) = \lim_{\boldsymbol{h}\to\boldsymbol{0}} df(\boldsymbol{x}_0) = \lim_{\substack{h_1\to\boldsymbol{0}\\ \dots\\ h_n\to\boldsymbol{0}}} \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x}_0)}{\partial x_i} h_i = 0,$$

т.е. $df(\boldsymbol{x}_0)$ является бесконечно малой функцией при $\|\boldsymbol{h}\| \to 0$.

Таким образом, в равенстве (18) участвуют три бесконечно малые функции, и его при $\|\boldsymbol{h}\| \to 0$ можно воспринимать как совпадение полного приращения функции с ее дифференциалом с точностью до бесконечно малой функции $o\left(\|\boldsymbol{h}\|\right)$ более высокого порядка, чем $\|\boldsymbol{h}\|$.

6°. Касательная плоскость. Градиент. В п.1° отмечался геометрический смысл частных производных функции двух переменных $z = f(x, y)$. В этом случае понятию дифференцируемости функции в точке и понятию дифференциала также можно придать наглядную геометрическую интерпретацию.

Пусть функция $z = f(x, y)$ дифференцируема в точке (x_0, y_0) и, следовательно, имеет место формула (3):

$$\Delta f(x_0, y_0) = f'_x(x_0, y_0)\,\Delta x + f'_y(x_0, y_0)\,\Delta y + o(\rho),$$

где $\rho = \sqrt{\Delta x^2 + \Delta y^2} \to 0$ при $\Delta x \to 0$, $\Delta y \to 0$. Введем следующие обозначения:

$$x = x_0 + \Delta x, \quad y = y_0 + \Delta y,$$

и отметим, что условие $\rho \to 0$, равносильное одновременному стремлению к нулю приращений аргументов Δx и Δy, можно заменить на условия $x \to x_0$, $y \to y_0$.

Перепишем выражение для приращения функции несколько иначе:

$$z = z_0 + f_x'(x_0, y_0)(x - x_0) + f_y'(x_0, y_0)(y - y_0) + o(\rho), \tag{19}$$

где z и z_0 – значения функции $f(x, y)$ в точках (x, y) и (x_0, y_0) соответственно, т.е. $z = f(x, y)$, $z_0 = f(x_0, y_0)$.

Рассмотрим плоскость (α), определяемую в пространстве с системой координат *Oxyz* уравнением

$$z = z_0 + f_x'(x_0, y_0)(x - x_0) + f_y'(x_0, y_0)(y - y_0). \tag{20}$$

Будем переменную z по оси аппликат в этом уравнении обозначать символом $z_{\text{кас}}$, чтобы отличать ее от аналогичной координаты в уравнении (19).

$$z_{\text{кас}} = z_0 + f_x'(x_0, y_0)(x - x_0) + f_y'(x_0, y_0)(y - y_0). \tag{21}$$

Вычитанием левых и правых частей формул (19) и (21) получим равенство

$$z - z_{\text{кас}} = o(\rho), \tag{22}$$

означающее, что разность аппликат графика функции $z = f(x, y)$ и плоскости (α) является бесконечно малой функцией более высокого порядка, чем ρ, при $\rho \to 0$ (иначе говоря, при $x \to x_0$, $y \to y_0$).

Обратимся к геометрической интерпретации графика функции двух переменных $z = f(x, y)$ в виде некоторой поверхности (S) в трехмерном пространстве (п.1° §2 гл.1). Равенство (22) позволяет однозначно выделить плоскость (α) среди всех тех плоскостей, которые содержат точку $M_0(x_0, y_0, z_0)$, принадлежащую поверхности (S). Действительно, пусть для некоторой плоскости, задаваемой уравнением

$$z_{\text{кас}} = z_0 + A(x - x_0) + B(y - y_0), \tag{23}$$

выполнено условие (22). (Рассматривать плоскости, проходящие через точку M_0, которые не могут быть описаны уравнениями такого вида, т.е. параллельные оси аппликат, нет необходимости, ибо они не могут являться геометрическими образами никаких функций двух переменных x, y). Складывая левые и правые части равенств (22) и (23) и возвращаясь к первоначальным обозначениям, получим

$$\Delta f(x_0, y_0) = A\,\Delta x + B\Delta y + o(\rho),$$

откуда в силу теоремы 2 $A = f_x'(x_0, y_0)$, $B = f_y'(x_0, y_0)$.

Таким образом, среди всех плоскостей, проходящих через точку M_0 (т.е. описываемых уравнением (23)), единственной, для которой выполнено условие (22), является плоскость (α) (описываемая уравнением (21)). Иначе говоря, при стремлении точки (x, y) к точке (x_0, y_0)

сближение аппликат графика функции $z = f(x, y)$ и плоскости (α) происходит "существенно быстрее", чем сближение аппликаты графика функции и аппликаты любой другой плоскости. Плоскость (α) называется *касательной плоскостью к графику функции* $z = f(x, y)$ *в точке* (x_0, y_0, z_0), а функцию, заданную равенством (20), называют *линейной аппроксимацией функции* $f(x, y)$ *в окрестности точки* (x_0, y_0).

Заметим попутно, что из определения дифференциала (формула (17)) и формулы (21) следует

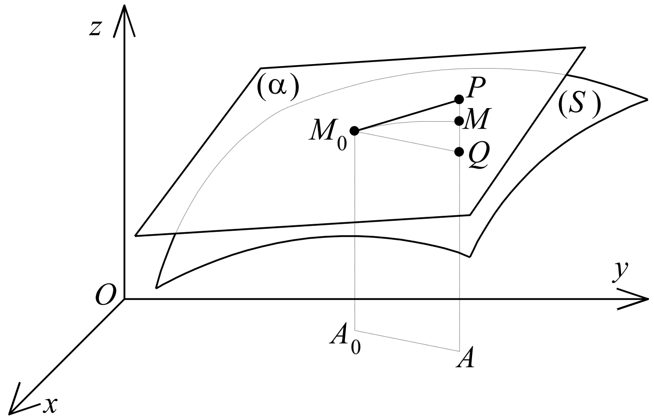

Рис. 2.4. Геометрическая иллюстрация к понятию касательной плоскости

$$z_{\text{кас}} - z_0 = df(x_0, y_0) = dz,$$

т.е. дифференциал равен приращению аппликаты касательной плоскости, вызванному приращениями аргументов $\Delta x = x - x_0$, $\Delta y = y - y_0$.

Геометрическая иллюстрация всего вышеизложенного представлена на рис.2.4. Сделаем необходимые пояснения к рисунку. Точки $A_0(x_0, y_0, 0)$ и $A(x, y, 0)$ лежат в плоскости xOy и принадлежат области определения функции $f(x, y)$. (S) – поверхность графика функции $z = f(x, y)$, точки $M_0(x_0, y_0, z_0)$ и $M(x, y, z)$ принадлежат поверхности (S). Плоскость (α) – касательная плоскость, определенная уравнением (21), точки M_0 и $P(x, y, z_{\text{кас}})$ принадлежат плоскости (α). Точка $Q(x, y, z_0)$ имеет те же абсциссу и ординату, что и точки A, M, P, а аппликату, равную $z_0 = f(x_0, y_0)$ и совпадающую с аппликатой точки M_0. В этих обозначениях $dz = z_{\text{кас}} - z_0 = |AP| - |AQ| = |PQ|$, а учитывая формулу (22), можно усмотреть и геометрический смысл величины $o(\rho)$:

$$o(\rho) = z - z_{\text{кас}} = |AM| - |AP| = -|MP|.$$

(Отрицательность величины $o(\rho)$ обусловлена конкретностью рассматриваемой геометрической ситуации).

Замечание. Обратимся еще раз к уравнению касательной плоскости (α):

$$z = z_0 + f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0).$$

Из аналитической геометрии известно, что вектор $\vec{n} = \left(f'_x(x_0, y_0),\ f'_y(x_0, y_0),\ -1\right)$ является нормалью к касательной плоскости (α). Этот вектор называют также *нормалью к* *поверхности* (S), *определяемой функцией* $z = f(x,y)$, *в точке* (x_0, y_0, z_0).

В п.2° была введена производная по направлению, с которой связано понятие так называемого градиента. Пусть функция $f(\boldsymbol{x})$ определена в окрестности точки $\boldsymbol{x}_0 \in \mathbf{R}^n$ и дифференцируема в этой точке. В таком случае, как отмечалось в п.5°, функция $f(\boldsymbol{x})$ имеет частные производные в этой точке по всем аргументам: $\dfrac{\partial f(\boldsymbol{x}_0)}{\partial x_1}, \ldots, \dfrac{\partial f(\boldsymbol{x}_0)}{\partial x_n}$. n-мерный вектор, составленный из этих частных производных, называется *градиентом функции* $f(\boldsymbol{x})$ в точке $\boldsymbol{x}_0$ и обозначается одним из следующих способов:

$$\nabla f(\boldsymbol{x}_0) = \operatorname{grad} f(\boldsymbol{x}_0) = \left(\frac{\partial f(\boldsymbol{x}_0)}{\partial x_1}, \ldots, \frac{\partial f(\boldsymbol{x}_0)}{\partial x_n} \right) \tag{24}$$

(обозначение ∇ читается "набла"). Используя понятия градиента и скалярного произведения векторов (определение 2 п.2° §1 гл.1), условие дифференцируемости функции (14) может быть переписано в виде

$$\Delta f(\boldsymbol{x}_0) = \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{h} \rangle + o\left(\|\boldsymbol{h}\|\right),$$

где вектор $\boldsymbol{h} = (h_1, \ldots, h_n)$ состоит из приращений независимых аргументов.

Связь между градиентом и производной по направлению устанавливает следующая теорема.

Теорема 5. Пусть функция $f(\boldsymbol{x})$ дифференцируема в точке $\boldsymbol{x}_0$, и $\boldsymbol{y}$ – произвольный ненулевой вектор пространства $\mathbf{R}^n$. Тогда функция $f(\boldsymbol{x})$ имеет в точке $\boldsymbol{x}_0$ производную по направлению $\boldsymbol{y}$ (в смысле определения 4), причем

$$\frac{\partial f(\boldsymbol{x}_0)}{\partial \boldsymbol{y}} = \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{y} \rangle. \tag{25}$$

▶ Заметим, что какой бы ни была окрестность $u(\boldsymbol{x}_0)$, точки вида $\boldsymbol{x}_0 + h\boldsymbol{y}$ при достаточно малых положительных h принадлежат этой окрестности. Действительно,

$$\|(\boldsymbol{x}_0 + h\boldsymbol{y}) - \boldsymbol{x}_0\| = \|h\boldsymbol{y}\| = h\|\boldsymbol{y}\| \xrightarrow[h\to+0]{} 0.$$

Выписывая для таких точек условие дифференцируемости функции $f(\boldsymbol{x})$ в точке $\boldsymbol{x}_0$, имеем

$$\Delta f(\boldsymbol{x}_0) = f(\boldsymbol{x}_0 + h\boldsymbol{y}) - f(\boldsymbol{x}_0) = \langle \nabla f(\boldsymbol{x}_0), h\boldsymbol{y} \rangle + o\left(\|h\boldsymbol{y}\|\right) = h\langle \nabla f(\boldsymbol{x}_0), \boldsymbol{y} \rangle + \|\boldsymbol{y}\| \cdot o(h).$$

Теперь по определению 4 производной по направлению можно записать

$$\frac{\partial f(\boldsymbol{x}_0)}{\partial \boldsymbol{y}} = \lim_{h\to+0} \frac{f(\boldsymbol{x}_0 + h\boldsymbol{y}) - f(\boldsymbol{x}_0)}{h} = \lim_{h\to+0} \frac{\langle \nabla f(\boldsymbol{x}_0), \boldsymbol{y} \rangle + \|\boldsymbol{y}\| \cdot o(h)}{h} =$$

$$= \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{y} \rangle + \|\boldsymbol{y}\| \lim_{h\to+0} \frac{o(h)}{h} = \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{y} \rangle.$$ ◀

Формула (25) позволяет, зная градиент функции, легко вычислить ее производную по любому направлению $\boldsymbol{y}$.

Рассмотрим подробнее случай функции трех переменных. Пусть функция $f(x,y,z)$ дифференцируема в точке $M_0(x_0, y_0, z_0)$, и пусть $\boldsymbol{e}_0$ – произвольный единичный вектор. Как отмечалось в п.1° §1 гл.1, скалярное произведение элементов из $\mathbf{R}^3$ совпадает с геометрическим определением скалярного произведения векторов $\langle \vec{a}, \vec{b} \rangle = |\vec{a}||\vec{b}| \cos\varphi$, где $|\vec{a}|$ и $|\vec{b}|$ длины векторов $\vec{a}$ и $\vec{b}$ соответственно, а φ – угол между ними. Теперь снова вернемся к формуле (25)

$$\frac{\partial f(M_0)}{\partial \boldsymbol{e}_0} = \langle \nabla f(M_0), \boldsymbol{e}_0 \rangle = |\nabla f(M_0)|\,|\boldsymbol{e}_0| \cos\theta = |\nabla f(M_0)| \cos\theta, \tag{26}$$

где θ – угол между векторами $\nabla f(M_0)$ и $\boldsymbol{e}_0$. Формула (26) раскрывает геометрический смысл направления вектора $\nabla f(M_0)$ (если градиент отличен от $\mathbf{0}$), а также смысл величины его модуля. Действительно, если векторы $\boldsymbol{e}_0$ и $\nabla f(M_0)$ сонаправлены, то производная по этому направлению имеет наибольшее значение и равна длине градиента $|\nabla f(M_0)|$ (так как в этом случае $\cos\theta = 1$). Для любого другого вектора $\boldsymbol{e}_0$ угол θ, который он составляет с вектором $\nabla f(M_0)$, отличен от нуля (а значит, $\cos\theta < 1$), и, следовательно, производная по направлению $\boldsymbol{e}_0$ в этом случае строго меньше длины градиента. Таким образом, градиент функции в точке, если он не равен нулевому вектору, определяет направление наибольшей скорости возрастания функции, а длина градиента и есть значение этой наибольшей скорости. Если же $\nabla f(M_0) = \mathbf{0}$, то в точке M_0 производные функции $f(x,y,z)$ по всем направлениям равны нулю. Эти свойства градиента, вытекающие из формулы (25), имеют место и в общем случае функции n переменных $f(\boldsymbol{x}) = f(x_1, \dots, x_n)$.

Пусть $\boldsymbol{c} = (c_1, \dots, c_n)$ – фиксированный вектор из $\mathbf{R}^n$, α – вещественное число и

$$l(\boldsymbol{x}) = l(x_1, \dots, x_n) = \langle \boldsymbol{c}, \boldsymbol{x} \rangle + \alpha = c_1 x_1 + \dots + c_n x_n + \alpha$$

– аффинная функция n переменных (п.1° §2 гл.1). Поскольку для такой функции частные производные имеют вид

$$\frac{\partial l(\boldsymbol{x})}{\partial x_i} = c_i, \quad i = 1, \dots, n,$$

для градиента аффинной функции получаем

$$\nabla l(\boldsymbol{x}) = (c_1, \dots, c_n) = \boldsymbol{c},$$

т.е. градиент представляет собой постоянный вектор $\boldsymbol{c}$ независимо от того, в какой точке $\boldsymbol{x}$

он вычислен.

Пусть $Q = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ – вещественная симметричная матрица $(a_{21} = a_{12})$ размера 2×2, $\boldsymbol{c} = (c_1, c_2)$ – фиксированный вектор из пространства $\mathbf{R}^2$, и α – вещественное число. Рассмотрим квадратичную функцию двух переменных $q(x_1, x_2)$ (п.1° §2 гл.1), определяемую равенством

$$q(\boldsymbol{x}) = q(x_1, x_2) = \frac{1}{2}\langle \boldsymbol{x}Q, \boldsymbol{x}\rangle + \langle \boldsymbol{c}, \boldsymbol{x}\rangle + \alpha = \frac{1}{2}a_{11}x_1^2 + a_{21}x_2x_1 + \frac{1}{2}a_{22}x_2^2 + c_1x_1 + c_2x_2 + \alpha .$$

Ее частные производные имеют вид

$$\frac{\partial q(\boldsymbol{x})}{\partial x_1} = a_{11}x_1 + a_{21}x_2 + c_1, \quad \frac{\partial q(\boldsymbol{x})}{\partial x_2} = a_{12}x_1 + a_{22}x_2 + c_2,$$

и градиент квадратичной функции можно записать в следующей форме

$$\nabla q(\boldsymbol{x}) = \left(\frac{\partial q(\boldsymbol{x})}{\partial x_1}, \frac{\partial q(\boldsymbol{x})}{\partial x_2}\right) = \left(a_{11}x_1 + a_{21}x_2 + c_1,\ a_{12}x_1 + a_{22}x_2 + c_2\right) =$$

$$= \left(a_{11}x_1 + a_{21}x_2,\ a_{12}x_1 + a_{22}x_2\right) + (c_1, c_2) = \boldsymbol{x}Q + \boldsymbol{c} .$$

Для квадратичной функции n переменных $q(\boldsymbol{x}) = \frac{1}{2}\langle \boldsymbol{x}Q, \boldsymbol{x}\rangle + \langle \boldsymbol{c}, \boldsymbol{x}\rangle + \alpha$ градиент вычисляется точно так же:

$$\nabla q(\boldsymbol{x}) = \boldsymbol{x}Q + \boldsymbol{c} .$$

Здесь Q – вещественная симметричная матрица размера $n \times n$, $\boldsymbol{c}$ – фиксированный вектор из $\mathbf{R}^n$, $\boldsymbol{x}$ – произвольный вектор из пространства $\mathbf{R}^n$.

В заключение этого параграфа отметим связь между линиями уровня функции двух переменных и ее градиентом. Рассмотрим функцию $f(M) = f(x, y)$, точку $M_0(x_0, y_0)$ из области ее определения и линию уровня Γ функции $f(M)$, соответствующую значению $f(M_0)$:

$$\Gamma = \left\{(x, y) \mid f(x, y) = f(x_0, y_0)\right\}. \tag{27}$$

Предположим, что линия уровня Γ представляет собой кривую, допускающую параметрическое задание при помощи уравнений

$$x = x(t), \quad y = y(t), \quad t \in [\alpha, \beta], \tag{28}$$

и пусть $t_0 \in [\alpha, \beta]$ – такое значение параметра t, для которого $x_0 = x(t_0)$, $y_0 = y(t_0)$. Определение кривой Γ (27) и уравнения (28), описывающие эту кривую, позволяют записать следующее равенство:

$$f\big(x(t),\, y(t)\big) = f\big(x(t_0),\, y(t_0)\big), \tag{29}$$

справедливое для всех $t \in [\alpha, \beta]$. Дифференцируя равенство (29) по t и подставляя в полученный результат $t = t_0$, имеем

$$\frac{\partial f(x_0, y_0)}{\partial x} \cdot x'(t_0) + \frac{\partial f(x_0, y_0)}{\partial y} \cdot y'(t_0) = 0. \tag{30}$$

Если интерпретировать переменную t как время, а уравнения (28) как уравнения движения материальной точки в зависимости от времени t, то вектор $\boldsymbol{v}(t_0) = \big(x'(t_0),\, y'(t_0)\big)$ представляет собой вектор скорости в момент времени $t = t_0$. Тогда равенство (30) можно переписать в виде

$$\langle \nabla f(x_0, y_0),\, \boldsymbol{v}(t_0) \rangle = 0, \tag{31}$$

что означает перпендикулярность векторов $\nabla f(x_0, y_0)$ и $\boldsymbol{v}(t_0)$. Как известно из физики, если $\boldsymbol{v}(t_0) \neq 0$, то вектор скорости $\boldsymbol{v}(t_0)$ направлен по касательной, проведенной в точке $(x_0, y_0) = \big(x(t_0),\, y(t_0)\big)$ к кривой Γ. Учитывая это обстоятельство, равенство (31) геометрически означает, что градиент $\nabla f(x_0, y_0)$ функции $f(x, y)$, вычисленный в точке (x_0, y_0), перпендикулярен касательной, проведенной в этой точке к линии уровня Γ (на рис.2.5 начало вектора $\nabla f(x_0, y_0)$ совмещено с точкой $M_0(x_0, y_0)$).

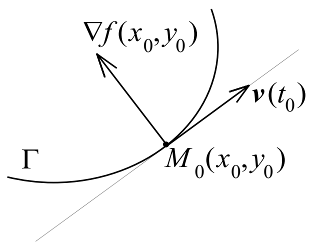

Рис. 2.5. Линии уровня функции $f(x, y)$ и ее градиент $\nabla f(x_0, y_0)$.

§ 2. Формула Тейлора для функций нескольких переменных

В этом параграфе выводится и обсуждается формула Тейлора для функций нескольких переменных, которую вполне обоснованно считают вершиной дифференциального исчисления. Эта формула активно используется в следующей главе, посвященной изучению экстремальных задач.

1°. Частные производные высших порядков. Рассмотрим функцию двух переменных $z = f(x, y)$, определенную в окрестности точки (x_0, y_0), и пусть эта функция имеет частные производные $f'_x(x, y)$, $f'_y(x, y)$ в каждой точке этой окрестности. В таком случае эти производные, в свою очередь, представляют собой функции двух переменных x и y, определенные в окрестности точки (x_0, y_0). Поэтому можно ставить вопрос о наличии у них частных производных по тому или иному аргументу. Если у функции $f'_x(x,y)$ в точке (x_0, y_0) существует производная по переменной x, то ее называют *частной производной второго порядка функции* $f(x,y)$ *по переменной* x в точке (x_0, y_0) и обозначают

$$f''_{xx}(x_0, y_0) = z''_{xx}(x_0, y_0) = \frac{\partial^2 z(x_0, y_0)}{\partial x^2} = \frac{\partial^2 f(x_0, y_0)}{\partial x^2}.$$

Если существует частная производная от функции $f'_x(x,y)$ по переменной y, то ее называют *смешанной частной производной второго порядка от функции* $f(x,y)$:

$$f''_{xy}(x_0,y_0) = z''_{xy}(x_0,y_0) = \frac{\partial^2 z(x_0,y_0)}{\partial y \partial x} = \frac{\partial^2 f(x_0,y_0)}{\partial y \partial x} = \frac{\partial}{\partial y}\left\{\frac{\partial f}{\partial x}\right\}_{(x_0,y_0)}.$$

Таким образом, функция двух переменных может иметь четыре частные производные второго порядка $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y \partial x}$, $\frac{\partial^2 f}{\partial x \partial y}$, $\frac{\partial^2 f}{\partial y^2}$, из которых две являются смешанными: $\frac{\partial^2 f}{\partial y \partial x}$ и $\frac{\partial^2 f}{\partial x \partial y}$.

Вообще, *частной производной порядка* k *от функции* $f(x,y)$ называют частную производную по какому-либо из аргументов от некоторой производной $(k-1)$-го порядка. Например,

$$\frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial x \partial y}\right) = \frac{\partial^3 f}{\partial x^2 \partial y} \quad \text{или} \quad \frac{\partial}{\partial y}\left(\frac{\partial^3 z}{\partial x \partial y^2}\right) = \frac{\partial^4 z}{\partial y \partial x \partial y^2}.$$

Символика, которая при этом используется, естественным образом переносится с частных производных второго порядка на частные производные более высокого порядка.

Пример. Пусть $z = x^3 y + 3x^2 y^2 - 2x^2 + y$.

Имеем: $z'_x = 3x^2 y + 6xy^2 - 4x$; $z'_y = x^3 + 6x^2 y + 1$;

$z''_{xx} = 6xy + 6y^2 - 4$; $z''_{yy} = 6x^2$;

$z''_{yx} = (z'_y)'_x = 3x^2 + 12xy$; $z''_{xy} = (z'_x)'_y = 3x^2 + 12xy$.

В общем случае значения смешанных производных зависят от порядка, в котором производится дифференцирование. Однако, при определенных, не слишком жестких условиях, имеет место их совпадение, как это и произошло в рассмотренном примере.

Теорема 1. Пусть в некоторой окрестности $u(M_0)$ точки $M_0(x_0, y_0)$ определены смешанные производные второго порядка $f''_{xy}(x,y)$ и $f''_{yx}(x,y)$ функции $f(x,y)$, и пусть эти производные непрерывны в самой точке M_0. Тогда справедливо равенство

$$f''_{xy}(x_0, y_0) = f''_{yx}(x_0, y_0). \tag{1}$$

▶ Пусть $\Delta x \neq 0$, $\Delta y \neq 0$ – приращения переменных x и y в точке $M_0(x_0, y_0)$ такие, что $M_1(x_0 + \Delta x,\ y_0 + \Delta y) \in u(M_0)$. Точки $M_2(x_0 + \Delta x, y_0)$ и $M_3(x_0, y_0 + \Delta y)$ в таком случае также принадлежат окрестности $u(M_0)$. Рассмотрим следующее выражение:

$$w = \left[f(x_0 + \Delta x,\ y_0 + \Delta y) - f(x_0 + \Delta x, y_0)\right] - \left[f(x_0, y_0 + \Delta y) - f(x_0, y_0)\right], \tag{2}$$

и определим на отрезке $[x_0, x_0 + \Delta x]$ функцию одной переменной t следующим равенством:

$$\varphi(t) = f(t, y_0 + \Delta y) - f(t, y_0).$$

В силу существования частной производной $f'_x(x, y)$ в окрестности $u(M_0)$, функция $\varphi(t)$ дифференцируема на отрезке $[x_0, x_0 + \Delta x]$, причем

$$\varphi'(t) = f'_x(t, y_0 + \Delta y) - f'_x(t, y_0). \tag{3}$$

Выражение w можно записать как приращение функции $\varphi(t)$ в точке x_0, вызванное приращением аргумента Δx:

$$w = \varphi(x_0 + \Delta x) - \varphi(x_0).$$

Применим к функции $\varphi(t)$, рассматриваемой на отрезке $[x_0, x_0 + \Delta x]$ формулу конечных приращений Лагранжа. Согласно этой формуле, существует такое число $\theta_1 \in (0, 1)$, что выполняется равенство

$$\varphi(x_0 + \Delta x) - \varphi(x_0) = \varphi'(x_0 + \theta_1 \Delta x)\,\Delta x.$$

Таким образом, с учетом формулы (3) получаем

$$w = \left[f'_x(x_0 + \theta_1 \Delta x,\ y_0 + \Delta y) - f'_x(x_0 + \theta_1 \Delta x, y_0)\right]\Delta x. \tag{4}$$

Еще раз применим формулу конечных приращений Лагранжа, на этот раз к функции $f'_x(x_0 + \theta_1 \Delta x, t)$, рассматриваемой на отрезке $[y_0, y_0 + \Delta y]$. Согласно этой формуле, существует такое число $\theta_2 \in (0, 1)$, что выполняется равенство

$$f'_x(x_0 + \theta_1 \Delta x, y_0 + \Delta y) - f'_x(x_0 + \theta_1 \Delta x, y_0) = f''_{xy}(x_0 + \theta_1 \Delta x, y_0 + \theta_2 \Delta y)\Delta y.$$

Подставляя полученный результат в формулу (4), окончательно находим

$$w = f''_{xy}(x_0 + \theta_1 \Delta x, y_0 + \theta_2 \Delta y)\Delta y \Delta x. \tag{5}$$

Аналогично, используя вспомогательную функцию

$$\psi(t) = f(x_0 + \Delta x, t) - f(x_0, t),$$

после двукратного применения формулы конечных приращений Лагранжа, для выражения (2) получим

$$w = f''_{yx}(x_0 + \theta_4 \Delta x, y_0 + \theta_3 \Delta y)\Delta x \Delta y,$$

где θ_3 и θ_4 – некоторые числа из промежутка $(0, 1)$. Сравнивая последнее равенство с равенством (5), можно записать

$$f''_{xy}(x_0+\theta_1\Delta x, y_0+\theta_2\Delta y) = f''_{yx}(x_0+\theta_4\Delta x, y_0+\theta_3\Delta y).$$

Переходя здесь к пределу при $\Delta x \to 0$, $\Delta y \to 0$, и учитывая непрерывность функций $f''_{xy}(x,y)$ и $f''_{yx}(x,y)$ в точке M_0 (данную по условию), приходим к требуемому равенству (1). ◀

Установленная теорема о равенстве смешанных производных 2-го порядка допускает очевидное обобщение: результат дифференцирования по различным переменным не зависит от порядка следования переменных, участвующих в процессе дифференцирования. Например, если частные производные $\dfrac{\partial^3 f}{\partial x^2 \partial y}$, $\dfrac{\partial^3 f}{\partial x \partial y \partial x}$, $\dfrac{\partial^3 f}{\partial y \partial x^2}$ непрерывны в некоторой точке, то в этой точке справедливо равенство

$$\frac{\partial^3 f}{\partial x^2 \partial y} = \frac{\partial^3 f}{\partial x \partial y \partial x} = \frac{\partial^3 f}{\partial y \partial x^2},$$

и т. п. Более того, этот результат справедлив и для функций с любым конечным числом аргументов. С теми же оговорками по поводу непрерывности, две смешанные производные одного и того же порядка равны, если только количество дифференцирований по соответствующим переменным одинаково. Например, $\dfrac{\partial^4 f}{\partial x \partial y \partial z^2} = \dfrac{\partial^4 f}{\partial z \partial x \partial z \partial y}$ и т. п.

2°. Дифференциалы высших порядков. В этом пункте будем предполагать, что все встречающиеся в нем функции имеют непрерывные частные производные в соответствующих точках вплоть до тех порядков, которые будут рассматриваться. Напомним, что в предыдущем параграфе для функции $y = f(\boldsymbol{x}) = f(x_1, \ldots, x_n)$ было введено понятие дифференциала в точке $\boldsymbol{x} = (x_1, \ldots, x_n)$ формулой (17):

$$d f(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x})}{\partial x_i} h_i. \tag{1}$$

В дальнейшем, чтобы не усложнять запись, будем опускать упоминание о самой точке $\boldsymbol{x}$, а также изменим, следуя традиции, обозначения h_i для приращений аргументов на dx_i ($i = 1, \ldots, n$). Будем называть величины dx_i не только приращениями независимых переменных x_i, но и *дифференциалами*. Иногда их называют даже *независимыми дифференциалами* в знак того, что они не зависят от точки $\boldsymbol{x} = (x_1, \ldots, x_n)$, в которой дифференциал вычисляется. Формально их "независимость" будет проявляться в том, что при дифференцировании по переменным $x_1, \ldots, x_n$ величины dx_i будут рассматриваться как постоянные. С учетом сделанных выше замечаний формула (1) принимает следующий вид

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i . \qquad (2)$$

Ясно, что дифференциал df зависит, вообще говоря, от переменных $x_1, \ldots, x_n$ и дифференциалов $dx_1, \ldots, dx_n$. Дифференциал функции $f(\boldsymbol{x})$, заданный формулой (2), будем называть *дифференциалом первого порядка* или *дифференциалом*. *Второй дифференциал функции* $f(\boldsymbol{x})$ (*дифференциал второго порядка*), соответствующий тем же дифференциалам $dx_1, \ldots, dx_n$, определяется равенством

$$d^2 f = d(df). \qquad (3)$$

Раскроем содержание правой части:

$$d^2 f = d\left(\sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i\right) = \sum_{j=1}^{n} \frac{\partial}{\partial x_j}\left(\sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i\right) dx_j = \sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i} dx_j dx_i . \qquad (4)$$

Из линейной алгебры известно, что функция n переменных $t_1, \ldots, t_n$ вида

$$\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij} t_i t_j ,$$

где $a_{ij} = a_{ji}$, называется *квадратичной формой*. Учитывая равенство смешанных производных

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j} ,$$

равенство (4) можно интерпретировать следующим образом: второй дифференциал функции $f(\boldsymbol{x})$ представляет собой квадратичную форму от независимых дифференциалов $dx_1, \ldots, dx_n$.

Дифференциал порядка k *функции* $f(\boldsymbol{x})$, соответствующий независимым дифференциалам $dx_1, \ldots, dx_n$, определяется по индукции с помощью рекуррентного соотношения

$$d^k f = d(d^{k-1} f), \quad k = 2, 3, \ldots .$$

Для третьего дифференциала (*дифференциала третьего порядка*), рассуждая так же, как при выводе формулы (4), получим

$$d^3 f = \sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i} dx_k dx_j dx_i . \qquad (5)$$

Рассмотрим функцию двух переменных $f(x, y)$ и, учитывая равенство соответствующих смешанных производных 2-го и 3-го порядка, вычислим по формулам (4) и (5) второй и третий дифференциалы функции $f(x,y)$:

$$d^2 f = \frac{\partial^2 f}{\partial x^2} dx^2 + 2\frac{\partial^2 f}{\partial x \partial y} dx\,dy + \frac{\partial^2 f}{\partial y^2} dy^2,$$

$$d^3 f = \frac{\partial^3 f}{\partial x^3} dx^3 + 3\frac{\partial^3 f}{\partial x^2 \partial y} dx^2 dy + 3\frac{\partial^3 f}{\partial x \partial y^2} dx\,dy^2 + \frac{\partial^3 f}{\partial y^3} dy^3,$$

где dx и dy – независимые дифференциалы переменных x и y соответственно. С помощью метода математической индукции можно получить общую формулу для дифференциала порядка k ($k = 1, 2, \dots$):

$$d^k f = \sum_{p=0}^{k} C_k^p \frac{\partial^k f}{\partial x^{k-p} \partial y^p} dx^{k-p} dy^p, \tag{6}$$

где C_k^p – соответствующие биномиальные коэффициенты, т.е. $C_k^p = \dfrac{k!}{p!(k-p)!}$.

Вместо формулы (6) нередко используют символическую формулу

$$d^k f = \left(\frac{\partial}{\partial x} dx + \frac{\partial}{\partial y} dy \right)^k f, \tag{7}$$

которую следует понимать следующим образом: если выражение $\left(\dfrac{\partial}{\partial x} dx + \dfrac{\partial}{\partial y} dy \right)^k$ рассматривать как алгебраическое и раскрыть его при помощи бинома Ньютона, а затем формально умножить на f, правая часть обретет свой первоначальный смысл, т.е. превратится в формулу (6). В случае функции $f(\boldsymbol{x}) = f(x_1, \dots, x_n)$ имеет место аналогичная символическая формула

$$d^k f = \left(\sum_{i=1}^{n} \frac{\partial}{\partial x_i} dx_i \right)^k f, \quad k = 1, 2, \dots .$$

В заключение приведем еще один способ описания второго дифференциала функции $f(\boldsymbol{x}) = f(x_1, \dots, x_n)$, который понадобится нам в дальнейшем. Матрицу размера $n \times n$, составленную из частных производных функции $f(\boldsymbol{x})$ в точке $\boldsymbol{x}_0$ следующим образом:

$$H(\boldsymbol{x}_0) = \left(\frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_i \partial x_j} \right)_{n \times n} = \begin{pmatrix} \dfrac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_1^2} & \dfrac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_1 \partial x_2} & \dots & \dfrac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_2^2} & \dots & \dfrac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \dfrac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_n \partial x_2} & \dots & \dfrac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_n^2} \end{pmatrix}, \tag{8}$$

называют *матрицей Гессе* или *гессианом*. Обозначим через $\boldsymbol{dx} = (dx_1, \dots, dx_n)$ вектор-строку, составленную из независимых дифференциалов $dx_1, \dots, dx_n$, и пусть $\boldsymbol{dx}^T$ – вектор-столбец. Будем рассматривать второй дифференциал функции $f(\boldsymbol{x})$ в точке $\boldsymbol{x}_0$, и пусть $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{dx}$ – некоторая точка из окрестности точки $\boldsymbol{x}_0$. Пользуясь операцией умножения матриц, формулу (4) теперь можно записать так:

$$d^2 f(\boldsymbol{x}_0) = \boldsymbol{dx} H(\boldsymbol{x}_0) \boldsymbol{dx}^T = (\boldsymbol{x} - \boldsymbol{x}_0) H(\boldsymbol{x}_0) (\boldsymbol{x} - \boldsymbol{x}_0)^T . \tag{9}$$

По существу, здесь применен стандартный способ записи квадратичной формы в матричном виде.

Заметим попутно, что для квадратичной функции $q(\boldsymbol{x}) = \frac{1}{2}\langle \boldsymbol{x}Q, \boldsymbol{x}\rangle + \langle \boldsymbol{c}, \boldsymbol{x}\rangle + \alpha$, где Q – вещественная симметричная матрица размера $n \times n$, $\boldsymbol{c}$ – фиксированный вектор из $\mathbf{R}^n$, α – вещественное число (см. п.1° §2 гл.1), гессиан совпадает с матрицей Q (и, следовательно, не зависит от выбора точки $\boldsymbol{x}_0$): $H(\boldsymbol{x}_0) = \left(\dfrac{\partial^2 q(\boldsymbol{x}_0)}{\partial x_i \partial x_j} \right)_{n \times n} = Q$.

3°. Формула Тейлора. Как известно, для функции одной переменной $F(t)$, удовлетворяющей определенным условиям, формула Тейлора порядка m с остаточным членом в форме Лагранжа при $t_0 = 0$ записывается следующим образом:

$$F(t) = F(0) + \frac{F'(0)}{1!} t + \dots + \frac{F^{(m)}(0)}{m!} t^m + \frac{F^{(m+1)}(\theta)}{(m+1)!} t^{m+1}, \tag{10}$$

где θ - некоторое число, расположенное между 0 и t.

Рассмотрим теперь функцию двух переменных $f(x, y) = f(M)$, имеющую непрерывные частные производные до $(m+1)$-го порядка включительно в каждой точке окрестности $u(M_0)$ точки $M_0(x_0, y_0)$. Пусть Δx и Δy такие произвольные приращения переменных x и y соответственно, что точка $M_1(x_0 + \Delta x,\ y_0 + \Delta y)$ принадлежит окрестности $u(M_0)$. Точки $M_t(x_0 + t\Delta x,\ y_0 + t\Delta y)$ при изменении переменной t от 0 до 1 пробегают отрезок на плоскости, соединяющий точки M_0 и M_1, и, следовательно, также принадлежат окрестности $u(M_0)$. Более того, если t меняется в пределах отрезка $[-\sigma, 1]$, где σ - достаточно малое положительное число, то точки M_t также принадлежат указанной окрестности. Введем на отрезке $[-\sigma, 1]$ вспомогательную функцию F одной переменной t равенством

$$F(t) = f(M_t) = f(x_0 + t\Delta x,\ y_0 + t\Delta y) . \tag{11}$$

Заметим, прежде всего, что

$$\Delta f(M_0) = f(M_1) - f(M_0) = f(x_0 + \Delta x,\ y_0 + \Delta y) - f(x_0, y_0) = F(1) - F(0) . \tag{12}$$

Функция $F(t)$ на отрезке $[-\sigma, 1]$ (в том числе и в нуле) имеет производные до $(m+1)$- го порядка включительно, причем $F^{(m+1)}(t)$ – непрерывна на данном отрезке. Этот факт проверяется непосредственно последовательным дифференцированием по t равенства (11) с использованием правила дифференцирования сложной функции (см. теорему 4 §1 гл.2). Действительно,

$$F'(t) = \frac{\partial f(M_t)}{\partial x}\Delta x + \frac{\partial f(M_t)}{\partial y}\Delta y,$$

$$F''(t) = \frac{\partial^2 f(M_t)}{\partial x^2}\Delta x^2 + 2\frac{\partial^2 f(M_t)}{\partial x \partial y}\Delta x \Delta y + \frac{\partial^2 f(M_t)}{\partial y^2}\Delta y^2,$$

. . . .

$$F^{(m+1)}(t) = \left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^{m+1} f(M_t).$$

Здесь правая часть последнего равенства истолковывается так, как это было указано в предыдущем пункте.

Установленные свойства функции $F(t)$ дают возможность записать для нее формулу Тейлора (10). Положив в этой формуле $t = 1$, при некотором $\theta \in (0,1)$ получаем

$$F(1) = F(0) + \frac{F'(0)}{1!} + \ldots + \frac{F^{(m)}(0)}{m!} + \frac{F^{(m+1)}(\theta)}{(m+1)!}.$$

Перенося $F(0)$ в левую часть и учитывая формулу (12), приходим к равенству

$$\Delta f(M_0) = \sum_{k=1}^{m}\frac{1}{k!}\left(\frac{\partial}{\partial x}\Delta x + \frac{\partial}{\partial y}\Delta y\right)^k f(M_0) + R_m(M_\theta), \tag{13}$$

где $R_m(M_\theta) = \dfrac{1}{(m+1)!}\left(\dfrac{\partial}{\partial x}\Delta x + \dfrac{\partial}{\partial y}\Delta y\right)^{m+1} f(M_\theta)$. Формула (13) носит название *формулы Тейлора для функции двух переменных* $f(M) = f(x,y)$ *с остаточным членом в форме Лагранжа*. Нередко ей придают несколько иную форму записи, внешне более напоминающую формулу Тейлора (10) для функции одной переменной:

$$\boxed{f(x,y) = f(x_0,y_0) + \sum_{k=1}^{m}\frac{1}{k!}\left(\frac{\partial}{\partial x}(x-x_0) + \frac{\partial}{\partial y}(y-y_0)\right)^k f(x_0,y_0) + R_m(x,y)}, \tag{14}$$

где $R_m(x,y) = \dfrac{1}{(m+1)!}\left(\dfrac{\partial}{\partial x}(x-x_0) + \dfrac{\partial}{\partial y}(y-y_0)\right)^{m+1} f(x_0 + \theta(x-x_0), y_0 + \theta(y-y_0))$ при некотором $0 < \theta < 1$.

Сформулируем более подробно полученный результат.

Теорема 2. Пусть функция двух переменных $f(x,y)$ имеет производные до порядка $m+1$ включительно в каждой точке некоторой окрестности точки (x_0, y_0). Тогда для любой точки (x,y) указанной окрестности найдется такое число $\theta \in (0,1)$, при котором имеет место формула Тейлора (14).

Аналогичным образом выводится *формула Тейлора для функции* n *переменных* $f(\boldsymbol{x}) = f(x_1,\ldots,x_n)$ и произвольной точки $\boldsymbol{x}$ из некоторой окрестности точки $\boldsymbol{x}_0 = (x_1^{(0)}, x_2^{(0)},\ldots,x_n^{(0)})$:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \sum_{k=1}^{m} \frac{1}{k!}\left(\frac{\partial}{\partial x_1}(x_1 - x_1^{(0)}) + \ldots + \frac{\partial}{\partial x_n}(x_n - x_n^{(0)})\right)^k f(\boldsymbol{x}_0) + R_m(\boldsymbol{x}), \tag{15}$$

где $R_m(\boldsymbol{x}) = \dfrac{1}{(m+1)!}\left(\dfrac{\partial}{\partial x_1}(x_1 - x_1^{(0)}) + \ldots + \dfrac{\partial}{\partial x_n}(x_n - x_n^{(0)})\right)^{m+1} f\big(\boldsymbol{x}_0 + \theta(\boldsymbol{x} - \boldsymbol{x}_0)\big)$ при некотором $\theta \in (0,1)$.

Подчеркнем, что второе слагаемое в правой части формулы (15) представляет собой многочлен степени m от n независимых переменных $x_1, \ldots, x_n$.

Можно доказать (см. [1], [2]), что остаточный член в форме Лагранжа $R_m(\boldsymbol{x})$ является бесконечно малой функцией более высокого порядка, чем $\|\boldsymbol{x} - \boldsymbol{x}_0\|^m$ при $\|\boldsymbol{x} - \boldsymbol{x}_0\| \to 0$, т.е.

$$R_m(\boldsymbol{x}) = o\left(\|\boldsymbol{x} - \boldsymbol{x}_0\|^m\right).$$

Такую форму записи остаточного члена называют *формой Пеано*. С учетом этого обстоятельства формулу (15) можно переписать в следующем виде

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \sum_{k=1}^{m} \frac{1}{k!}\left(\frac{\partial}{\partial x_1}(x_1 - x_1^{(0)}) + \ldots + \frac{\partial}{\partial x_n}(x_n - x_n^{(0)})\right)^k f(\boldsymbol{x}_0) + o\left(\| \boldsymbol{x} - \boldsymbol{x}_0 \|^m\right) \tag{16}$$

Равенство (16) после отбрасывания последнего слагаемого в правой части приводит к приближенной формуле:

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_0) + \sum_{k=1}^{m} \frac{1}{k!}\left(\frac{\partial}{\partial x_1}(x_1 - x_1^{(0)}) + \ldots + \frac{\partial}{\partial x_n}(x_n - x_n^{(0)})\right)^k f(\boldsymbol{x}_0) \tag{17}$$

Чем “ближе” точка $\boldsymbol{x}$ к $\boldsymbol{x}_0$ и чем больше m, тем более точным является равенство (17).

Отметим несколько важных частных случаев формул (15) – (17).

Если функция $f(\boldsymbol{x})$ имеет лишь непрерывные частные производные первого порядка (т.е. $m = 0$), то в правой части формулы (15) второе слагаемое отсутствует, и она принимает вид

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \left\langle \nabla f(\boldsymbol{x}_0 + \theta(\boldsymbol{x} - \boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0 \right\rangle \tag{18}$$

где $\theta \in (0,1)$ (см. формулу (24) §1 гл.2). Формула (18) представляет собой обобщение известной формулы Лагранжа для функции одной переменной.

Используя матрицу Гессе (8), равенство (15) при $m=1$ может быть переписано следующим образом:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{x}-\boldsymbol{x}_0 \rangle + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0) H\big(\boldsymbol{x}_0 + \theta(\boldsymbol{x}-\boldsymbol{x}_0)\big)(\boldsymbol{x}-\boldsymbol{x}_0)^T, \tag{19}$$

где θ – некоторое число из интервала $(0,1)$.

Формулы (16) и (17) при $m=1$ записываются следующим образом

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{x}-\boldsymbol{x}_0 \rangle + o\big(\|\boldsymbol{x}-\boldsymbol{x}_0\|\big), \tag{20}$$

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_0) + \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{x}-\boldsymbol{x}_0 \rangle. \tag{21}$$

Первое равенство по существу лежит в основе определения дифференцируемой функции (см. формулу (14) §1 гл.2), а второе приближенное равенство дает линейную аппроксимацию функции $f(\boldsymbol{x})$ в окрестности точки $\boldsymbol{x}_0$.

Наконец, при $m=2$ формулы (16) и (17) принимают следующий вид:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x}_0)}{\partial x_i}(x_i - x_i^{(0)}) + \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_j \partial x_i}(x_j - x_j^{(0)})(x_i - x_i^{(0)}) + o\big(\| \boldsymbol{x}-\boldsymbol{x}_0 \|^2\big) =$$

$$= f(\boldsymbol{x}_0) + \langle \nabla f(\boldsymbol{x}_0), (\boldsymbol{x}-\boldsymbol{x}_0) \rangle + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)\, H(\boldsymbol{x}_0)\, (\boldsymbol{x}-\boldsymbol{x}_0)^T + o\big(\|\boldsymbol{x}-\boldsymbol{x}_0\|^2\big), \tag{22}$$

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_0) + \langle f(\boldsymbol{x}_0), \boldsymbol{x}-\boldsymbol{x}_0 \rangle + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)\, H(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)^T. \tag{23}$$

Правая часть формулы (23) представляет собой квадратичную функцию от n переменных x_1, … , x_n, и потому ее называют *квадратичной аппроксимацией функции* $f(\boldsymbol{x}) = f(x_1, \ldots, x_n)$ *в окрестности точки* $\boldsymbol{x}_0$.

§ 3. Неявные функции

Определение. Пусть функция $f(x,y)$ определена на некотором множестве $D \subset \mathbf{R}^2$. Функция $\varphi(x)$ одной переменной, определенная на некотором множестве $X \subset \mathbf{R}$, такая, что для любого $x \in X$ выполняется

1) $\big(x, \varphi(x)\big) \in D$

2) $f\big(x, \varphi(x)\big) = 0$,

называется *неявной функцией, определенной уравнением*

$$f(x, y) = 0. \qquad (1)$$

Функцию $\varphi(x)$ при этом называют *решением уравнения* (1) и говорят, что функция $\varphi(x)$ *задана неявно уравнением* (*1*).

Пример. Рассмотрим функцию двух переменных и $f(x, y) = x^2 + y^2 - 1$ и соответствующее ей уравнение

$$x^2 + y^2 - 1 = 0. \qquad (2)$$

Согласно данному определению решениями этого уравнения являются, например, функции:

$$\varphi_1(x) = \sqrt{1 - x^2}, \qquad x \in [-1, 1];$$

$$\varphi_2(x) = -\sqrt{1 - x^2}, \qquad x \in [-1, 1];$$

$$\varphi_3(x) = \begin{cases} \sqrt{1 - x^2}, & x \in [0, 1], \\ -\sqrt{1 - x^2}, & x \in [-1, 0); \end{cases}$$

$$\varphi_4(x) = \begin{cases} \sqrt{1 - x^2}, & x \in \left[-1, -1/2\right] \cup \left[0, 1/2\right], \\ -\sqrt{1 - x^2}, & x \in \left(-1/2, 0\right) \cup \left(1/2, 1\right]; \end{cases}$$

$$\varphi_5(x) = \sqrt{1 - x^2}, \qquad x \in \left(1/3, 2/3\right);$$

$$\varphi_6(x) = \begin{cases} \sqrt{1 - x^2}, & \text{если } x \text{ – рациональное число, } |x| \le 1, \\ -\sqrt{1 - x^2}, & \text{если } x \text{ – иррациональное число, } |x| \le 1. \end{cases}$$

На рис.2.6 показаны графики первых пяти функций. График функции $\varphi_6(x)$ изобразить не представляется возможным.

Очевидно, уравнение (2) задает неявно бесконечное множество различных функций, даже определенных на одном и том же множестве X.

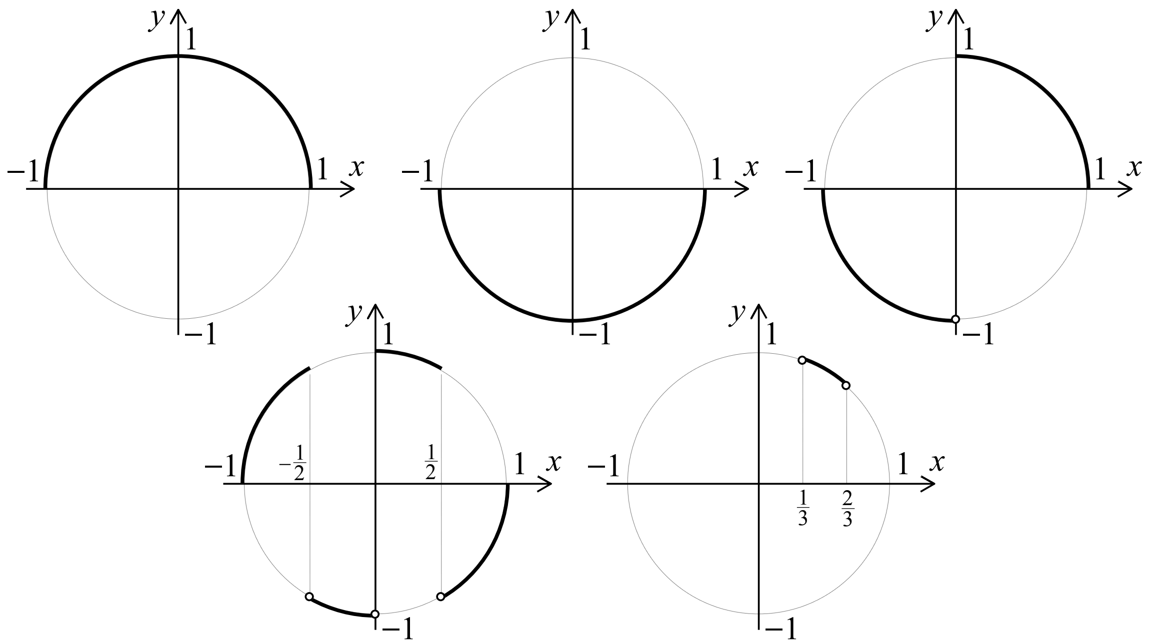

Рис. 2.6. Графики неявных функций $\varphi_1(x)$, $\varphi_2(x)$, $\varphi_3(x)$, $\varphi_4(x)$, $\varphi_5(x)$, заданных уравнением (2).

Конечно, уравнение (1) может вообще не определять ни одной функции (например, уравнение $x^2 + y^2 + 1 = 0$) или задавать неявную функцию, область определения которой состоит из единственной точки (например, уравнение $x^2 + y^2 = 0$). Однако часто имеют место случаи, когда уравнение (1), рассматриваемое в окрестности точки (x_0, y_0), такой, что $f(x_0, y_0) = 0$, однозначно определяет неявную функцию с определенными свойствами (например, непрерывную) по крайней мере в достаточно малой окрестности точки x_0. То же уравнение (2), рассматриваемое в окрестности точки (x_0, y_0), такой, что $x_0^2 + y_0^2 = 1$, $y_0 \neq 0$, определяет однозначно непрерывную функцию в достаточно малой окрестности точки x_0 (рис.2.7). Ниже приводятся достаточные условия существования неявной функции, определяемой уравнением (1), имеющей непрерывную производную, в зависимости от свойств функции $f(x, y)$.

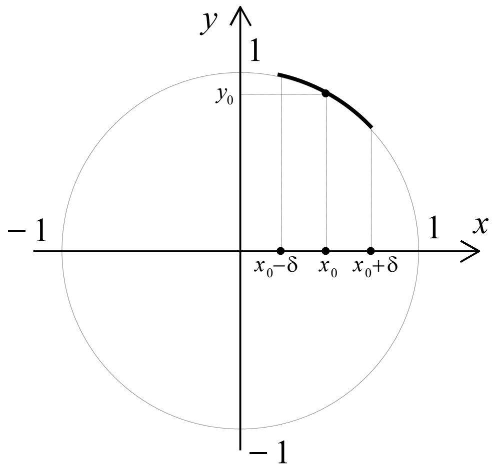

Рис. 2.7. График непрерывной неявной функции, определяемой уравнением (2).

Теорема 1. Пусть функция двух переменных $f(x, y)$ в окрестности $u(x_0, y_0)$ точки (x_0, y_0) имеет непрерывные частные производные $f_x'(x, y)$, $f_y'(x, y)$, причем $f_y'(x_0, y_0) \neq 0$, и пусть $f(x_0, y_0) = 0$. Тогда найдутся такие одномерные окрестности $u(x_0)$ и $u(y_0)$ точек x_0 и y_0 соответственно, что для любого $x \in u(x_0)$ существует единственное решение $y \in u(y_0)$

уравнения

$$f(x,y)=0.$$

Это решение, обозначаемое $y=\varphi(x)$, представляет собой функцию одной переменной, определенную на $u(x_0)$, удовлетворяющую условию $y_0=\varphi(x_0)$, имеющую непрерывную производную в каждой точке $x\in u(x_0)$, причем

$$\varphi'(x)=-\frac{f_x'(x,y)}{f_y'(x,y)}. \tag{3}$$

(Для вычисления производной функции $\varphi(x)$ по формуле (3) в правую ее часть подставляется значение $\varphi(x)$)

▶ По условию, $f_y'(x_0,y_0)\neq 0$. Будем считать для определенности, что $f_y'(x_0,y_0)>0$. Поскольку частная производная $f_y'(x,y)$ непрерывна в окрестности $u(x_0,y_0)$, по теореме о сохранении знака (теорема 5 п.3° §2 гл.1) существует такая окрестность точки (x_0,y_0), для всех точек которой также выполняется неравенство $f_y'(x,y)>0$.

Выберем и зафиксируем замкнутый прямоугольник с центром в точке (x_0,y_0) (рис.2.7):

$$\overline{D}=\left\{(x,y)\;\middle|\;|x-x_0|\le\xi,\ |y-y_0|\le\eta\right\}$$

так, чтобы он целиком содержался в этой окрестности. Функция $f_y'(x,y)$, таким образом, непрерывна и принимает только положительные значения на компакте $\overline{D}$. В таком случае по теореме Вейерштрасса (теорема 6 п.3° §2 гл.1) эта функция достигает на $\overline{D}$ своего наименьшего значения *m*, которое также положительно:

$$m=\inf_{(x,y)\in\overline{D}}f_y'(x,y)=\min_{(x,y)\in\overline{D}}f_y'(x,y)>0. \tag{4}$$

Этот факт будет использован позже.

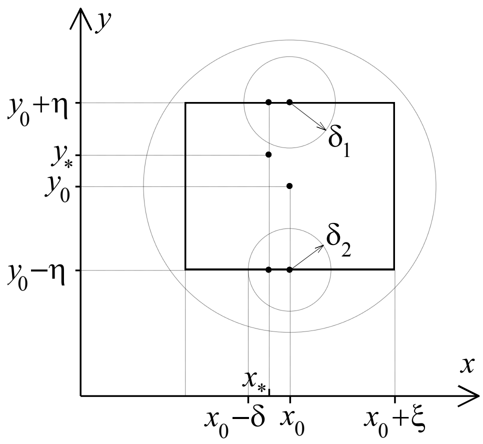

Рис. 2.8. К доказательству теоремы 1.

Из условий, наложенных на функцию $f(x,y)$, следует, в частности, ее непрерывность на прямоугольнике $\overline{D}$. Функция $f(x_0,y)$, рассматриваемая как функция одной переменной *y*, имеет положительную производную и, следовательно, непрерывна и строго возрастает на отрезке $[y_0-\eta,\ y_0+\eta]$. Кроме того, эта функция обращается в нуль в точке y_0 (по условию, $f(x_0,y_0)=0$). В таком случае на концах отрезка она принимает значения противоположных знаков, а именно:

$$f(x_0,y_0+\eta)>0,\quad f(x_0,y_0-\eta)<0.$$

Функция $f(x,y_0+\eta)$, рассматриваемая как функция одной переменной *x*, непрерывна в точке x_0 и принимает положительное значение в этой точке. По теореме о сохранении знака для функций одной переменной в некоторой одномерной окрестности $u_{\delta_1}(x_0) = \left\{ x \,\middle|\, |x - x_0| < \delta_1 \right\}$ выполнено неравенство $f(x, y_0 + \eta) > 0$. Аналогично, найдется такая окрестность $u_{\delta_2}(x_0)$, что $f(x, y_0 - \eta) < 0$ для всех $x \in u_{\delta_2}(x_0) = \left\{ x \,\middle|\, |x - x_0| < \delta_2 \right\}$. Обозначим

$$u(x_0) = u_\delta(x_0) = u_{\delta_1}(x_0) \bigcap u_{\delta_2}(x_0),$$

где $\delta = \min(\delta_1, \delta_2)$. Тогда для любого фиксированного $x_* \in u(x_0)$ функция $f(x_*, y)$ на отрезке $[y_0 - \eta, y_0 + \eta]$ непрерывна, строго возрастает и принимает на концах значения разных знаков:

$$f(x_*, y_0 - \eta) < 0, \quad f(x_*, y_0 + \eta) > 0.$$

По теореме о промежуточном значении, существует и притом единственное число $y_* \in u_\eta(y_0) = u(y_0)$, такое, что

$$f(x_*, y_*) = 0. \tag{5}$$

Таким образом, однозначно определена следующая функция: каждому значению $x \in u(x_0)$ поставлено в соответствие единственное число $y \in u(y_0)$. Обозначим эту функцию $y = \varphi(x)$. Согласно ее определению,

$$f\big(x, \varphi(x)\big) = 0 \tag{6}$$

при любом $x \in u(x_0)$. Так как для каждого $x_* \in u(x_0)$ значение y_*, обладающее свойством (5), единственно, то существует лишь одна такая функция, определенная на $u(x_0)$ со значениями в $u(y_0)$ и удовлетворяющая условию (6). В частности, поскольку по условию $f(x_0, y_0) = 0$, значение функции $\varphi(x)$ в точке x_0 совпадает с y_0, т.е. $y_0 = \varphi(x_0)$ (см. рис.2.9).

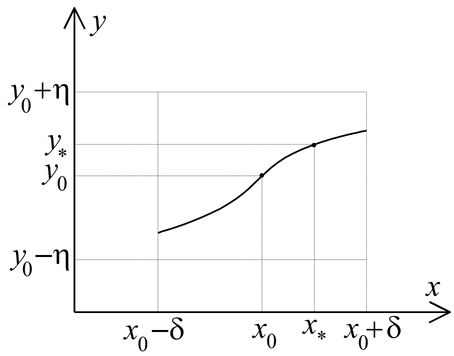

Рис. 2.9. Однозначная разрешимость уравнения (1) в окрестности точки (x_0, y_0).

Перейдем к установлению непрерывности функции $y = \varphi(x)$ на множестве $u(x_0)$. Пусть x – произвольное число из $u(x_0)$, и $\Delta x \neq 0$ – такое приращение, что $x + \Delta x \in u(x_0)$. Пусть, далее, $\Delta y = \varphi(x + \Delta x) - \varphi(x)$ и, следовательно, $y + \Delta y = \varphi(x + \Delta x)$. Пользуясь формулой Тейлора (при $m = 0$) для функции $f(x, y)$ (формула (18) п.3° §2) и учитывая равенства

$$f(x, y) = f\big(x, \varphi(x)\big) = 0 \quad \text{и} \quad f(x + \Delta x, y + \Delta y) = f\big(x + \Delta x, \varphi(x + \Delta x)\big) = 0,$$

имеем

$$0 = f(x+\Delta x,\ y+\Delta y) - f(x,y) = f'_x(x+\theta\Delta x,\ y+\theta\Delta y)\,\Delta x + f'_y(x+\theta\Delta x,\ y+\theta\Delta y)\,\Delta y, \qquad (7)$$

где θ – некоторое число из промежутка $(0,1)$. Поскольку $x+\Delta x \in u(x_0)$ и $y+\Delta y \in u(y_0)$, точка $(x+\theta\Delta x,\ y+\theta\Delta y) \in \overline{D}$ и, значит, $f'_y(x+\theta\Delta x,\ y+\theta\Delta y) > 0$. Из равенства (7) получаем

$$\frac{\Delta y}{\Delta x} = -\frac{f'_x(x+\theta\Delta x,\ y+\theta\Delta y)}{f'_y(x+\theta\Delta x,\ y+\theta\Delta y)}. \qquad (8)$$

Функция $f'_x(x,y)$ непрерывна на $\overline{D}$ по условию, что по теореме Вейерштрасса (см. теорему 6 п.3° §2 гл.1) влечет за собой ее ограниченность на данном множестве, т.е. существует такое число $M>0$, что $\left|f'_x(x,y)\right| \le M$ для всех точек $(x,y)\in\overline{D}$ (в частности, для всех $x\in u(x_0)$, $y\in u(y_0)$). Учитывая это обстоятельство, а также то, что $f'_y(x,y) \ge m > 0$ для всех $(x,y)\in\overline{D}$ (формула (4)), из равенства (8) получаем

$$\left|\frac{\Delta y}{\Delta x}\right| = \frac{\left|f'_x(x+\theta\Delta x,\ y+\theta\Delta y)\right|}{f'_y(x+\theta\Delta x,\ y+\theta\Delta y)} \le \frac{M}{m}.$$

Переписывая это неравенство в виде

$$0 \le |\Delta y| \le \frac{M}{m}|\Delta x|$$

и применяя теорему о сжатой переменной при $\Delta x \to 0$, приходим к равенству $\lim\limits_{\Delta x\to 0}|\Delta y| = 0$, которое равносильно равенству $\lim\limits_{\Delta x\to 0}\Delta y = 0$. Полученное означает непрерывность функции $y=\varphi(x)$ в произвольной точке $x\in u(x_0)$.

Снова обратимся к равенству (8) и перейдем в нем к пределу при $\Delta x\to 0$ (по доказанному, при этом и $\Delta y \to 0$):

$$\varphi'(x) = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} = -\frac{f'_x(x,y)}{f'_y(x,y)} = -\frac{f'_x\big(x,\varphi(x)\big)}{f'_y\big(x,\varphi(x)\big)}. \qquad (9)$$

Непрерывность функции $\varphi'(x)$ на $u(x_0)$ вытекает из равенства (9) и из теоремы о непрерывности суперпозиции (теорема 4 п.3° §2 гл.1): функция $y=\varphi(x)$ непрерывна на $u(x_0)$, имеет значения в $u(y_0)$, а $f'_x(x,y)$ и $f'_y(x,y)$ непрерывны на прямоугольнике $\overline{D} \supset \left\{(x,y)\ \middle|\ x\in u(x_0);\ y\in u(y_0)\right\}$. Таким образом, неявная функция $y=\varphi(x)$ однозначно определенная в окрестности $u(x_0)$ и удовлетворяющая условию $f\big(x,\varphi(x)\big)=0$ при любом $x\in u(x_0)$, имеет непрерывную производную в каждой точке этой окрестности. ◄

Вернемся к примеру, рассмотренному в начале параграфа. Функция $f(x,y) = x^2+y^2-1$

имеет непрерывные частные производные $f'_x(x,y)=2x$, $f'_y(x,y)=2y$. Следовательно, для любой точки (x_0, y_0), удовлетворяющей условию $x_0^2+y_0^2=1$ и такой, что $y_0 \neq 0$, частная производная функции $f(x,y)$ по y отлична от нуля: $f'_y(x_0, y_0) \neq 0$. Из теоремы 1 вытекает, что существует такая окрестность точки x_0, в которой однозначно определена неявная функция $y=\varphi(x)$, имеющая непрерывную производную

$$\varphi'(x) = -\frac{f'_x(x,y)}{f'_y(x,y)} = -\frac{2x}{2y} = -\frac{x}{y}.$$

Этой функцией является $\varphi_1(x)=\sqrt{1-x^2}$ (если $y_0>0$) или $\varphi_2(x)=-\sqrt{1-x^2}$ (если $y_0<0$).

Формула (3) дает явное выражение для производной неявной функции $y=\varphi(x)$. Это позволяет, в частности, получить уравнение касательной к кривой, заданной уравнением (1). Если $f(x_0, y_0)=0$ и в точке (x_0, y_0) выполняются условия теоремы 1, то для неявной функции, однозначно определенной уравнением (1), имеем:

$$y_0=\varphi(x_0); \quad \varphi'(x_0) = -\frac{f'_x(x_0,y_0)}{f'_y(x_0,y_0)}.$$

Подставляя эти выражения в уравнение касательной $y=\varphi'(x_0)(x-x_0)+\varphi(x_0)$, получим уравнение касательной к кривой, заданной уравнением (1):

$$f'_x(x_0, y_0)(x-x_0)+f'_y(x_0, y_0)(y-y_0)=0.$$

Обратимся теперь к более общему случаю, который рассматривается аналогично. Пусть функция $f(\boldsymbol{x},y)=f(x_1,\dots,x_n,y)$ определена на некотором множестве $D\subset \mathbf{R}^{n+1}$. Функция n переменных $\varphi(\boldsymbol{x})=\varphi(x_1,\dots,x_n)$, определенная на некотором множестве $X\subset\mathbf{R}^n$, такая, что для всякого $\boldsymbol{x}\in X$ выполняется

1) $(\boldsymbol{x},\varphi(\boldsymbol{x}))\in D$,

2) $f(\boldsymbol{x},\varphi(\boldsymbol{x}))=0$,

называется *неявной функцией, определенной уравнением* $f(\boldsymbol{x},y)=0$. При этом функцию $\varphi(\boldsymbol{x})$ называют *решением этого уравнения*.

Теорема 2. Пусть функция $f(\boldsymbol{x},y)$ в окрестности точки (x_0, y_0) такой, что $f(\boldsymbol{x}_0, y_0)=0$, имеет непрерывные частные производные по каждой из переменных x_1, …, x_n, y, и пусть $f'_y(\boldsymbol{x}_0, y_0)\neq 0$. Тогда найдутся такие окрестности $u(\boldsymbol{x}_0)$ и $u(y_0)$ точек $\boldsymbol{x}_0$ и y_0 соответственно, что для любой точки $\boldsymbol{x}\in u(\boldsymbol{x}_0)$ существует единственное решение $y\in u(y_0)$ уравнения

$$f(\boldsymbol{x},y)=0.$$

Это решение, обозначаемое $y = \varphi(\boldsymbol{x})$, представляет собой функцию n переменных, удовлетворяющую условию $\varphi(\boldsymbol{x}_0) = y_0$ и имеющую непрерывные частные производные в окрестности $u(\boldsymbol{x}_0)$, которые могут быть найдены по формуле

$$\varphi'_{x_i}(\boldsymbol{x}) = -\frac{f'_{x_i}(\boldsymbol{x}, y)}{f'_y(\boldsymbol{x}, y)} = -\frac{f'_{x_i}\big(\boldsymbol{x}, \varphi(\boldsymbol{x})\big)}{f'_y\big(\boldsymbol{x}, \varphi(\boldsymbol{x})\big)}.$$

Резюме к главе 2

В этой главе понятие дифференцируемой функции переносится на случай функции нескольких переменных.

В §1 вводится и анализируется понятие частной производной, а также его обобщение – производная по направлению. Далее определяются функции, дифференцируемые в точке, и устанавливается связь таких функций с существованием у них частных производных, а также выводятся
правила вычисления частных производных сложной функции. Вводится понятие дифференциала, определяются касательная плоскость и градиент и выясняется его связь с производной по направлению.

Второй параграф посвящен частным производным и дифференциалам высших порядков. Доказывается теорема о достаточных условиях равенства смешанных производных. Выводится формула Тейлора и рассматриваются различные формы ее записи.

В третьем параграфе исследуются функции, заданные неявно.

Задание на самостоятельную работу

1. Покажите, что производная $\dfrac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{y}}$ функции $f(\boldsymbol{x})$ в точке $\boldsymbol{x}$ по направлению $\boldsymbol{y}$ связана с производной $\dfrac{\partial f(\boldsymbol{x})}{\partial (\lambda \boldsymbol{y})}$ по направлению $\lambda \boldsymbol{y}$ (λ – не равное нулю вещественное число) соотношением $\dfrac{\partial f(\boldsymbol{x})}{\partial (\lambda \boldsymbol{y})} = \lambda \dfrac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{y}}$ (использовать связь между производной по направлению и градиентом (теорема 5 п.6° §1)).

2. Покажите, что функция $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ непрерывна в точке $\mathbf{0}$, имеет производные по любому направлению в этой точке, но не является при этом дифференцируемой.

3. Сколько частных производных порядка k имеет функция n переменных

$f(\boldsymbol{x}) = f(x_1, \dots, x_n)$?

4. Как выглядит матрица Гессе функции трех переменных $f(x,y,z) = \dfrac{xy^2}{z}$ в точке $(2, 1, -1)$?

5. Чему равен дифференциал $df(\boldsymbol{x}_0)$ функции n переменных $f(\boldsymbol{x})$ в точке $\boldsymbol{x}_0$, если ее градиент в этой точке равен нулю?

6. Выпишите формулу Тейлора с остаточным членом в форме Лагранжа для функции двух переменных $f(x,y)$ в точке (x_0, y_0) при $m = 0$ (частный случай формулы (18)) и при $m = 1$ (частный случай формулы (19)).

7. Выпишите формулу Тейлора с остаточным членом в форме Пеано для функции двух переменных $f(x,y)$ в точке (x_0, y_0) при $m = 1$ (частный случай формулы (20)) и при $m = 2$ (частный случай формулы (22)).

8. Решите задачи из сборника задач [3]: 7/55 – 7.73, 7.79 – 7.156.

Глава 3. Экстремумы

Глава посвящена задачам отыскания максимальных и/или минимальных значений функций нескольких переменных. Такого рода задачи часто встречаются в практике. Важнейшим инструментом изучения и решения подобных задач являются необходимые и/или достаточные условия экстремума, которые рассматриваются ниже. Изложение предваряет параграф, посвященный выпуклым множествам и вогнутым (выпуклым) функциям нескольких переменных.

§ 1. Выпуклые множества. Вогнутые и выпуклые функции

Здесь вводится важный с точки зрения теории экстремальных задач класс выпуклых множеств в пространстве $\mathbf{R}^n$. Приводится ряд примеров выпуклых множеств. На выпуклом множестве определяются вогнутые и выпуклые функции, которые также играют важную роль в теории экстремальных задач. Изучаются некоторые свойства этих функций.

1°. Выпуклые множества. Пусть имеются две различные точки $\boldsymbol{x}'$ и $\boldsymbol{x}''$ пространства $\mathbf{R}^n$. Напомним, что *отрезок, соединяющий* $\boldsymbol{x}'$ *и* $\boldsymbol{x}''$, состоит из всех точек вида

$$\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}'' \tag{1.1}$$

при $0 \le \lambda \le 1$.

Определение 1.1.Пусть $X \subset \mathbf{R}^n$. Множество X называют *выпуклым*, если оно вместе с каждой парой своих точек содержит и весь отрезок, их соединяющий, т.е. если для каждой пары точек $\boldsymbol{x}', \boldsymbol{x}'' \in X$ включение $\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}'' \in X$ выполнено при всех $\lambda \in [0,1]$.

Пустое множество и одноэлементное множество не содержат ни одной пары точек; тем не менее, оба этих множества удобно причислять к классу выпуклых множеств.

Анализ определения 1.1 показывает, что в случае $n = 1$ выпуклым множеством является промежуток (и только он!), т.е. отрезок, интервал или полуинтервал, причем последние два множества могут быть и неограниченными.

Рис.1.1 иллюстрирует определение 1.1 в случае $n = 2$.

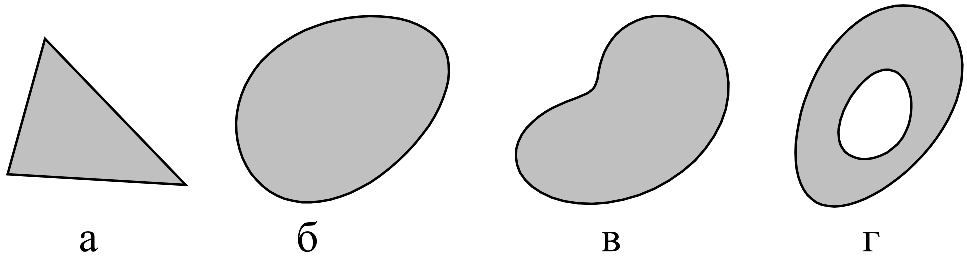

Рис. 1.1. Примеры плоских выпуклых (а, б) и невыпуклых (в, г) множеств.

Перейдем к рассмотрению примеров выпуклых множеств в пространстве $\mathbf{R}^n$.

Пример 1.1. Пространство $\mathbf{R}^n$ является выпуклым множеством.

▶ Действительно, для любых векторов $\boldsymbol{x}', \boldsymbol{x}'' \in \mathbf{R}^n$ и для любого $\lambda \in [0, 1]$ выражение (1.1) представляет собой некоторый *n*-мерный вектор, который, разумеется, принадлежит пространству $\mathbf{R}^n$. ◀

Пример 1.2. Отрезок, соединяющий две произвольные точки пространства $\mathbf{R}^n$, является выпуклым множеством.

▶ Пусть имеются две точки $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{R}^n$, $\boldsymbol{x} \neq \boldsymbol{y}$. Выберем на отрезке, соединяющем $\boldsymbol{x}$ и $\boldsymbol{y}$, две произвольные точки $\boldsymbol{x}' = \lambda'\boldsymbol{x} + (1-\lambda')\boldsymbol{y}$, $\boldsymbol{x}'' = \lambda''\boldsymbol{x} + (1-\lambda'')\boldsymbol{y}$ при некоторых $\lambda', \lambda'' \in [0, 1]$. Для любого $\lambda \in [0, 1]$ имеем

$$= \left[\lambda\lambda' + (1-\lambda)\lambda''\right]\boldsymbol{x} + \left[1 - \lambda\lambda' - (1-\lambda)\lambda''\right]\boldsymbol{y} = \overline{\lambda}\boldsymbol{x} + (1-\overline{\lambda})\boldsymbol{y},$$

где $0 \le \overline{\lambda} = \lambda\lambda' + (1-\lambda)\lambda'' \le 1$. Полученное означает, что отрезок, соединяющий точки $\boldsymbol{x}'$ и $\boldsymbol{x}''$, целиком содержится в отрезке, соединяющем точки $\boldsymbol{x}$ и $\boldsymbol{y}$. ◀

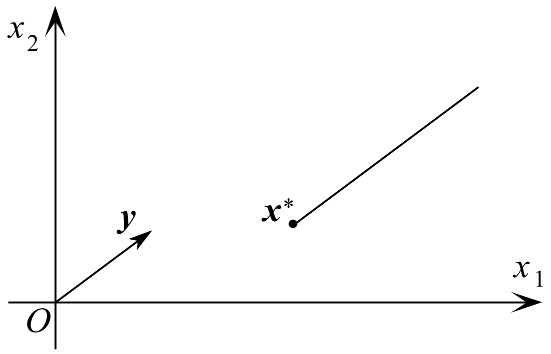

Рис. 1.2. Луч на плоскости

Пример 1.3. Выпуклым является множество

$$\left\{\boldsymbol{x} \in \mathbf{R}^n \,\middle|\, \boldsymbol{x} = \boldsymbol{x}^* + \lambda\boldsymbol{y} \text{ при некотором } \lambda \ge 0\right\},$$

называемое *лучом, исходящим из точки* $\boldsymbol{x}^* \in \mathbf{R}^n$ *в направлении ненулевого вектора* $\boldsymbol{y} \in \mathbf{R}^n$ (рис.1.2).

▶ Рассмотрим две произвольные точки луча: $\boldsymbol{x}' = \boldsymbol{x}^* + \lambda'\boldsymbol{y}$, $\boldsymbol{x}'' = \boldsymbol{x}^* + \lambda''\boldsymbol{y}$ при некоторых $\lambda', \lambda'' \ge 0$. Для каждого $\lambda \in [0, 1]$ можно записать:

$$\lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{x}'' = \lambda(\boldsymbol{x}^* + \lambda'\boldsymbol{y}) + (1-\lambda)(\boldsymbol{x}^* + \lambda''\boldsymbol{y}) = \boldsymbol{x}^* + \left(\lambda\lambda' + (1-\lambda)\lambda''\right)\boldsymbol{y},$$

где $\lambda\lambda' + (1-\lambda)\lambda'' \ge 0$. Следовательно, отрезок, соединяющий точки $\boldsymbol{x}'$ и $\boldsymbol{x}''$, целиком лежит на луче, исходящем из точки $\boldsymbol{x}^*$ в направлении вектора $\boldsymbol{y}$. ◀

Пример 1.4. Пусть $\boldsymbol{x}^* \in \mathbf{R}^n$, $\varepsilon > 0$. ε-окрестность точки $\boldsymbol{x}^*$, т. е. множество

$$u_\varepsilon(\boldsymbol{x}^*) = \left\{\boldsymbol{x} \in \mathbf{R}^n \,\middle|\, \left\|\boldsymbol{x} - \boldsymbol{x}^*\right\| < \varepsilon\right\},$$

является выпуклым.

▶ В самом деле, выберем две произвольные точки $\boldsymbol{x}', \boldsymbol{x}'' \in u_\varepsilon(\boldsymbol{x}^*)$ так, что $\left\|\boldsymbol{x}' - \boldsymbol{x}^*\right\| < \varepsilon$, $\left\|\boldsymbol{x}'' - \boldsymbol{x}^*\right\| < \varepsilon$. Рассмотрим точку $\lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''$ при каждом $\lambda \in [0, 1]$. Имеем

$$\left\|\lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{x}'' - \boldsymbol{x}^*\right\| = \left\|\lambda(\boldsymbol{x}' - \boldsymbol{x}^*) + (1-\lambda)(\boldsymbol{x}'' - \boldsymbol{x}^*)\right\| \le \left\|\lambda(\boldsymbol{x}' - \boldsymbol{x}^*)\right\| + \left\|(1-\lambda)(\boldsymbol{x}'' - \boldsymbol{x}^*)\right\| =$$

$$= \lambda \left\| (\boldsymbol{x}' - \boldsymbol{x}^*) \right\| + (1-\lambda) \left\| (\boldsymbol{x}'' - \boldsymbol{x}^*) \right\| < \lambda\varepsilon + (1-\lambda)\,\varepsilon = \varepsilon .$$

Здесь было использовано неравенство треугольника и свойство нормы вектора (см. п.1° §1 гл.1). Полученное в итоге неравенство $\left\| \lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{x}'' - \boldsymbol{x}^* \right\| < \varepsilon$ означает выполнение включения $\lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{x}'' \in u_\varepsilon(\boldsymbol{x}^*)$. ◀

2°. Вогнутые и выпуклые функции.

Определение 1.2.Функция *f*, заданная на выпуклом множестве X, $X \subset \mathbf{R}^n$, называется *вогнутой* на этом множестве, если неравенство

$$f\big(\lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''\big) \ge \lambda f(\boldsymbol{x}') + (1-\lambda) f(\boldsymbol{x}'') \tag{1.2}$$

выполняется для всех пар точек $\boldsymbol{x}', \boldsymbol{x}'' \in X$ и всех чисел $\lambda \in [0,1]$.

В левой части неравенства (1.2) записано значение функции, вычисленное в точке $\lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''$ для всех $\boldsymbol{x}', \boldsymbol{x}'' \in X$ и всех $\lambda \in [0,1]$. Таким образом, функция f должна быть определена в каждой точке отрезка, соединяющего произвольную пару точек множества *X*. Отсюда следует, что область определения вогнутой функции обязательно должна быть выпуклым множеством, иначе левая часть неравенства (1.2) при некоторых $\boldsymbol{x}', \boldsymbol{x}'' \in X$ и некотором $\lambda \in [0,1]$ может не иметь смысла.

Следует отметить, что при $\lambda = 0$, а также при $\lambda = 1$ неравенство (1.2) выполняется как равенство при всех $\boldsymbol{x}', \boldsymbol{x}'' \in X$. Поэтому указанные значения λ можно было бы из определения 2 исключить и потребовать выполнения неравенства (1.2) лишь для всех $\lambda \in (0,1)$. Модифицированное таким образом определение вогнутой функции будет эквивалентно исходному определению 1.2.

Очевидно, если функция f вогнута на некотором множестве *X*, то она является вогнутой и на любом выпуклом подмножестве множества *X*.

Определение 1.3.Функцию *f*, заданную на выпуклом множестве X, $X \subset \mathbf{R}^n$, называют *выпуклой* на этом множестве, если на нем функция $-f$ является вогнутой. Иначе говоря, функция f *выпукла* на *X*, если неравенство

$$f\big(\lambda\boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''\big) \le \lambda f(\boldsymbol{x}') + (1-\lambda) f(\boldsymbol{x}'') \tag{1.3}$$

имеет место для всех $\boldsymbol{x}', \boldsymbol{x}'' \in X$ и всех $\lambda \in [0,1]$.

Неравенство (1.3), записанное для функции $-f$, равносильно неравенству (2) для *f*, поэтому f вогнута, если функция $-f$ выпукла. Таким образом, умножая вогнутую (выпуклую) функцию на -1, придем к выпуклой (соответственно вогнутой) функции. Учитывая эту простую связь между вогнутыми и выпуклыми функциями, любой результат, полученный для одного класса функций, можно легко переформулировать для другого класса.

Пример 1.5. Аффинная функция $f(\boldsymbol{x}) = \langle \boldsymbol{c}, \boldsymbol{x} \rangle + \alpha$, где $\boldsymbol{c} \in \mathbf{R}^n$, $\boldsymbol{c} \neq \mathbf{0}$, $\alpha \in \mathbf{R}$, является одновременно вогнутой и выпуклой на $\mathbf{R}^n$.

▶ Действительно, для любых $\boldsymbol{x}', \boldsymbol{x}'' \in \mathbf{R}^n$ и всех λ имеем

$$f\big(\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''\big) = \langle \boldsymbol{c}, \lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}'' \rangle + \alpha = \langle \boldsymbol{c}, \lambda \boldsymbol{x}' \rangle + \langle \boldsymbol{c}, (1-\lambda)\boldsymbol{x}'' \rangle + \alpha =$$

$$= \lambda \langle \boldsymbol{c}, \boldsymbol{x}' \rangle + \lambda\alpha + (1-\lambda)\langle \boldsymbol{c}, \boldsymbol{x}'' \rangle + (1-\lambda)\alpha = \lambda f(\boldsymbol{x}') + (1-\lambda) f(\boldsymbol{x}'').$$ ◀

Пример 1.6. Евклидова норма $f(\boldsymbol{x}) = \|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$ является выпуклой функцией на $\mathbf{R}^n$.

▶ Проверка выпуклости данной функции опирается на неравенство треугольника и свойство нормы вектора (см. п.1° §1 гл.1):

$$f\big(\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''\big) = \|\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''\| \le \|\lambda \boldsymbol{x}'\| + \|(1-\lambda)\boldsymbol{x}''\| =$$

$$= \lambda\|\boldsymbol{x}'\| + (1-\lambda)\|\boldsymbol{x}''\| = \lambda f(\boldsymbol{x}') + (1-\lambda) f(\boldsymbol{x}'')$$

для всех $\boldsymbol{x}', \boldsymbol{x}'' \in \mathbf{R}^n$ и каждого $\lambda \in [0,1]$. ◀

При $n = 1$ функция $y = \|\boldsymbol{x}\|$ превращается в модуль: $y = |x_1|$. В случае $n = 2$ графиком функции $y = \|\boldsymbol{x}\|$ является поверхность кругового конуса (рис.1.3), а ее линии уровня (см. п.1° §2 гл.1) представляют собой семейство окружностей в плоскости x_1Ox_2 с центром в начале координат.

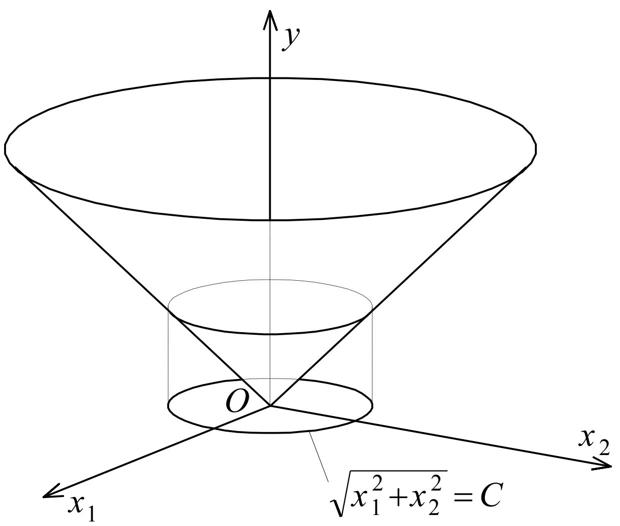

Рис. 1.3. График функции $y = \sqrt{x_1^2 + x_2^2}$ и ее линия уровня.

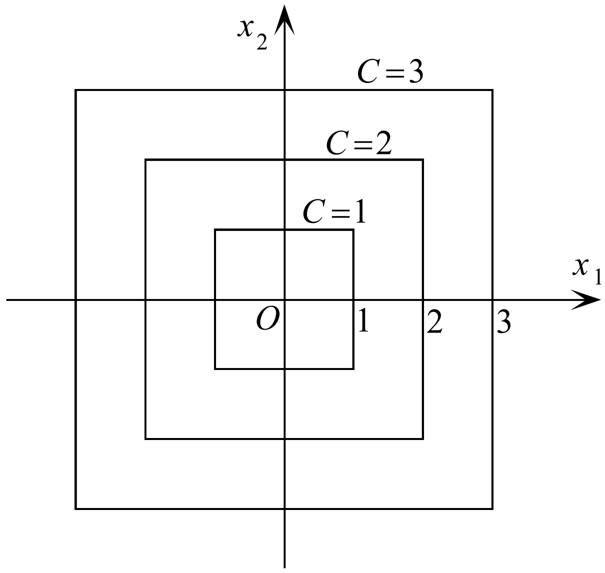

Рис. 1.4. Линии уровня функции $y = \max\{|x_1|;\ |x_2|\}$.

Пример 1.7. Функция $f(\boldsymbol{x}) = \max\{|x_1|; \ldots; |x_n|\}$ является выпуклой на пространстве $\mathbf{R}^n$.

▶ Выберем произвольно две точки $\boldsymbol{x}', \boldsymbol{x}'' \in \mathbf{R}^n$ и число $\lambda \in [0,1]$. Имеем:

$$f\left(\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''\right) = \max\left\{ \left| \lambda x_1' + (1-\lambda)x_1'' \right|; \ldots ; \left| \lambda x_n' + (1-\lambda)x_n'' \right| \right\} = \left| \lambda x_i' + (1-\lambda)x_i'' \right|$$

при некотором $i \in \{1, 2, \ldots, n\}$. Отсюда с использованием неравенства $|a+b| \le |a| + |b|$ получаем:

$$f\left(\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''\right) \le \lambda |x_i'| + (1-\lambda)|x_i''| \le$$

$$\le \lambda \max\left\{ |x_1'|; \ldots, |x_n'| \right\} + (1-\lambda)\max\left\{ |x_1''|; \ldots, |x_n''| \right\} = \lambda f(\boldsymbol{x}') + (1-\lambda) f(\boldsymbol{x}'') . \quad ◀$$

Функция $y = \max\left\{ |x_1|, \ldots, |x_n| \right\}$ при $n = 1$ также превращается в модуль: $y = |x_1|$. Таким образом, при $n = 1$ обе функции из двух последних примеров совпадают, чего нельзя сказать о случае $n > 1$, так как линии уровня второй функции при $n = 2$ задаются уравнением $\max\left\{ |x_1|; |x_2| \right\} = C$, где C – const и имеют вид квадратов (рис.1.4) со сторонами, параллельными координатным осям и с центром в начале координат.

3°. Свойства вогнутых (выпуклых) функций. Как показывает нижеследующая теорема, линейная комбинация вогнутых (выпуклых) функций с положительными коэффициентами представляет собой вогнутую (соответственно, выпуклую) функцию.

Теорема 1.1.Если f_1, f_2, ... , f_m – набор вогнутых (выпуклых) функций, заданных на выпуклом множестве X , $X \subset \mathbf{R}^n$, и λ_1, λ_2, ... , λ_m – произвольные положительные числа, то функция $f(\boldsymbol{x}) = \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x})$ является вогнутой (выпуклой) на множестве X.

▶ Выберем произвольно две точки $\boldsymbol{x}'$, $\boldsymbol{x}'' \in X$ и число $\lambda \in [0, 1]$. Используя вогнутость функций f_1, f_2, ... , f_m, имеем:

$$f\left(\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''\right) = \sum_{i=1}^{m} \lambda_i f_i\left(\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''\right) \ge \sum_{i=1}^{m} \lambda_i \left(\lambda f_i(\boldsymbol{x}') + (1-\lambda) f_i(\boldsymbol{x}'')\right) =$$

$$= \lambda \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}') + (1-\lambda) \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}'') = \lambda f(\boldsymbol{x}') + (1-\lambda) f(\boldsymbol{x}'') .$$

При доказательстве теоремы 1.1 в случае выпуклых функций неравенство “≥” в приведенной выше цепочке соотношений следует заменить неравенством “≤”. ◀

Теорема 1.1 указывает простой путь конструирования новых вогнутых и выпуклых функций на базе уже известных функций.

Пример 1.9. Пусть λ_1, λ_2, ... , λ_m – положительные числа. Функция одной переменной $y = x_i^2$ выпукла на всей вещественной оси, $i = 1, 2, \ldots, n$. Эту же функцию можно рассматривать как функцию *n* переменных, заданную на всем пространстве $\mathbf{R}^n$ (при каждом $i = 1, 2, \ldots, n$), причем она также будет выпуклой. Поэтому в силу теоремы 1.1 новая функция

n переменных (она является квадратичной) $y = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \ldots + \lambda_n x_n^2$ будет выпуклой на всем пространстве $\mathbf{R}^n$. Если все числа λ_1, λ_2, ... , λ_n отрицательны, то указанная функция вогнута.

4°. Критерии вогнутости (выпуклости) функций. Первый критерий формулируется для дифференцируемых функций.

Теорема 1.2. Предположим, что функция f дифференцируема в каждой точке открытого выпуклого множества X, $X \subset \mathbf{R}^n$. Для того, чтобы f была вогнутой (выпуклой) на множестве X, необходимо и достаточно, чтобы неравенство

$$\langle \nabla f(\boldsymbol{x}''), \boldsymbol{x}' - \boldsymbol{x}'' \rangle \geq (\leq) f(\boldsymbol{x}') - f(\boldsymbol{x}'') \tag{1.4}$$

выполнялось для всех $\boldsymbol{x}', \boldsymbol{x}'' \in X$.

▶ Доказательство теоремы достаточно провести лишь для вогнутой функции, поскольку умножение неравенства (1.4) (со знаком "$\geq$") на -1 приведет к неравенству (1.4) (со знаком "$\leq$") для функции $-f$, которая выпукла.

Необходимость. Так как по условию функция f вогнута на множестве X, то для любых $\boldsymbol{x}', \boldsymbol{x}'' \in X$ и всех $\lambda \in [0, 1]$ имеет место неравенство

$$f\big(\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}''\big) \geq \lambda f(\boldsymbol{x}') + (1-\lambda) f(\boldsymbol{x}''). \tag{1.5}$$

Отсюда с использованием равенства $\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}'' = \boldsymbol{x}'' + \lambda(\boldsymbol{x}' - \boldsymbol{x}'')$ получаем

$$\frac{f\big(\boldsymbol{x}'' + \lambda(\boldsymbol{x}' - \boldsymbol{x}'')\big) - f(\boldsymbol{x}'')}{\lambda} \geq f(\boldsymbol{x}') - f(\boldsymbol{x}'') \tag{1.6}$$

для всех $\lambda \in (0, 1]$. Благодаря дифференцируемости функции f в точке $\boldsymbol{x}''$ левая часть неравенства (1.6) имеет вид

$$\langle \nabla f(\boldsymbol{x}''), \boldsymbol{x}' - \boldsymbol{x}'' \rangle + \frac{o(\lambda)}{\lambda}.$$

Поэтому, переходя к пределу при $\lambda \to 0+$ в неравенстве (11), с учетом сказанного получим неравенство (1.4) (со знаком "$\geq$").

Достаточность. Выберем произвольно точки $\boldsymbol{x}', \boldsymbol{x}'' \in X$ и число $\lambda \in [0, 1]$. Введем точку $\boldsymbol{z} = \lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}'' \in X$. По условию неравенство (1.4) (со знаком "$\geq$") выполняется для любых точек $\boldsymbol{x}'$ и $\boldsymbol{x}''$ множества X, поэтому можно записать следующие два неравенства

$$\langle \nabla f(\boldsymbol{z}), \boldsymbol{x}' - \boldsymbol{z} \rangle \geq f(\boldsymbol{x}') - f(\boldsymbol{z}),$$

$$\langle \nabla f(\boldsymbol{z}), \boldsymbol{x}'' - \boldsymbol{z} \rangle \geq f(\boldsymbol{x}'') - f(\boldsymbol{z}).$$

Умножая первое неравенство на λ, а второе – на $(1-\lambda)$ и складывая их почленно, получим в левой части нуль, так как

$$\lambda\langle\nabla f(z), \boldsymbol{x}'-z\rangle+(1-\lambda)\langle\nabla f(z), \boldsymbol{x}''-z\rangle=\langle\nabla f(z), \lambda\boldsymbol{x}'+\boldsymbol{x}''-\lambda\boldsymbol{x}''-z\rangle=\langle\nabla f(z), \mathbf{0}\rangle=0 .$$

Следовательно, в результате сложения придем к неравенству

$$0\geq\lambda f(\boldsymbol{x}')+(1-\lambda)f(\boldsymbol{x}'')-f(z),$$

которое совпадает с неравенством (1.5) из определения вогнутой функции. ◀

Следующий критерий вогнутости (выпуклости) относится к дважды дифференцируемым функциям.

Теорема 1.3. Будем считать, что функция f определена на открытом выпуклом множестве X, $X\subset\mathbf{R}^n$, и в каждой точке $\boldsymbol{x}$ этого множества имеет непрерывные частные производные второго порядка, образующие матрицу Гессе (см. п.2° §2 гл.2)

$$H(\boldsymbol{x})=\left(\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i\partial x_j}\right)_{n\times n}.$$

Для того, чтобы функция f была вогнутой (выпуклой) на множестве X, необходимо и достаточно, чтобы для каждого $\boldsymbol{x}\in X$ соответствующая этой функции матрица Гессе $H(\boldsymbol{x})$ была неположительно (неотрицательно) определена, т.е. имело место неравенство

$$\boldsymbol{y}\cdot H(\boldsymbol{x})\cdot\boldsymbol{y}^T\leq(\geq)\,0 \quad \text{при всех} \quad \boldsymbol{y}\in\mathbf{R}^n .$$

Заметим, что в левой части этого неравенства $\boldsymbol{y}$ – вектор-строка, $\boldsymbol{y}^T$ – вектор-столбец, а умножение производится по правилу перемножения матриц, принятому в линейной алгебре.

▶ Очевидно, матрицы Гессе функций f и $-f$ отличаются лишь знаком, поэтому теорему 7 достаточно доказать, например, лишь для вогнутых функций.

Достаточность. Выберем произвольно две точки $\boldsymbol{x}',\boldsymbol{x}''\in X$ и запишем формулу Тейлора для функции f с остаточным членом в форме Лагранжа до членов второго порядка:

$$f(\boldsymbol{x}')=f(\boldsymbol{x}'')+\langle\nabla f(\boldsymbol{x}''),\boldsymbol{x}'-\boldsymbol{x}''\rangle+\frac{1}{2}(\boldsymbol{x}'-\boldsymbol{x}'')\cdot H\big(\boldsymbol{x}''+\theta(\boldsymbol{x}'-\boldsymbol{x}'')\big)\cdot(\boldsymbol{x}'-\boldsymbol{x}'')^T$$

при некотором $\theta\in(0,1)$. По условию матрица Гессе H неположительно определена, а значит, последнее слагаемое в написанной выше формуле неположительно. Поэтому

$$f(\boldsymbol{x}')\leq f(\boldsymbol{x}'')+\langle\nabla f(\boldsymbol{x}''),\boldsymbol{x}'-\boldsymbol{x}''\rangle.$$

Нетрудно видеть, что полученное неравенство равносильно неравенству (1.4) (со знаком “≥”). Поэтому в соответствии с теоремой 1.2 функция f вогнута на множестве X.

Необходимость. Пусть функция f вогнута на множестве X. Для доказательства неположительной определенности матрицы Гессе выберем произвольно $\boldsymbol{x}\in X$ и $\boldsymbol{y}\in\mathbf{R}^n$. По условию, множество X – открытое; значит, каждая его точка является внутренней. В частности, такой будет выбранная точка $\boldsymbol{x}$. Это означает, что она входит в множество X вместе с некоторой своей окрестностью, которую обозначим через $u(\boldsymbol{x})$: $u(\boldsymbol{x})\subset X$.

Рассмотрим точку $\boldsymbol{x}+t\boldsymbol{y}$ при $t\in\mathbf{R}$. Так как

$$\|\boldsymbol{x}-(\boldsymbol{x}+t\boldsymbol{y})\|=\|t\boldsymbol{y}\|=|t|\cdot\|\boldsymbol{y}\|\xrightarrow[t\to 0]{} 0,$$

то рассматриваемая точка $\boldsymbol{x}+t\boldsymbol{y}$ при всех достаточно малых (по абсолютной величине) значениях t будет принадлежать окрестности $u(\boldsymbol{x})$. Иными словами, существует положительное число ξ, такое, что $\boldsymbol{x}+t\boldsymbol{y}\in X$ для всех $t\in(-\xi,\xi)$.

На интервале $(-\xi,\xi)$ введем функцию Φ одной переменной t по формуле $\Phi(t)=f(\boldsymbol{x}+t\boldsymbol{y})$. Благодаря существованию всех частных производных второго порядка функции *f*, введенная функция Φ является дважды дифференцируемой, причем, по правилу дифференцирования сложной функции (см. п.4° §1 гл.2) имеем

$$\Phi'(t)=\sum_{i=1}^{n}\frac{\partial f(\boldsymbol{x}+t\boldsymbol{y})}{\partial x_i}y_i, \qquad -\xi<t<\xi,$$

$$\Phi''(t)=\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{\partial^2 f(\boldsymbol{x}+t\boldsymbol{y})}{\partial x_i\,\partial y_j}y_i y_j, \qquad -\xi<t<\xi,$$

$$\Phi''(0)=\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i\,\partial y_j}y_i y_j=\boldsymbol{y}\cdot H(\boldsymbol{x})\cdot\boldsymbol{y}^T. \tag{1.7}$$

Кроме того, вогнутость функции f влечет вогнутость функции Φ на $(-\xi,\xi)$:

$$\Phi\big(\lambda t'+(1-\lambda)t''\big)=f\Big(\boldsymbol{x}+\big(\lambda t'+(1-\lambda)t''\big)\cdot\boldsymbol{y}\Big)=f\big(\lambda(\boldsymbol{x}+t'\boldsymbol{y})+(1-\lambda)(\boldsymbol{x}+t''\boldsymbol{y})\big)\geq$$
$$\geq\lambda f(\boldsymbol{x}+t'\boldsymbol{y})+(1-\lambda)f(\boldsymbol{x}+t''\boldsymbol{y})=\lambda\Phi(t')+(1-\lambda)\Phi(t'')$$

для всех $t',t''\in(-\xi,\xi)$ и любого $\lambda\in[0,1]$.

Запишем формулу Маклорена для функции Φ с остаточным членом в форме Пеано:

$$\Phi(t)=\Phi(0)+\Phi'(0)\cdot t+\frac{1}{2}\Phi''(0)\cdot t^2+o(t^2)$$

при $t\in(-\xi,\xi)$. Отсюда находим:

$$\frac{1}{2}\Phi''(0)\cdot t^2+o(t^2)=\Phi(t)-\Phi(0)-\Phi'(0)\cdot t.$$

Так как функция Φ вогнута на $(-\xi,\xi)$, то в силу теоремы 1.2 (при $n=1$) правая часть последнего равенства не положительна. Такой же будет и левая часть:

$$\frac{1}{2}\Phi''(0)\cdot t^2+o(t^2)\leq 0, \quad \text{или} \quad \Phi''(0)+\frac{2\cdot o(t^2)}{t^2}\leq 0$$

для всех $t\in(-\xi,\xi)$, $t\neq 0$. Переходя в последнем неравенстве к пределу при $t\to 0$, получим $\Phi''(0)\leq 0$, что с учетом (1.7) означает, неположительную определенность матрицы Гессе $H(\boldsymbol{x})$.◀

Пример 1.10. Рассмотрим квадратичную функцию $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x} \cdot Q \cdot \boldsymbol{x}^T + \boldsymbol{c} \cdot \boldsymbol{x}^T + \alpha$, где Q – симметричная матрица размера $n \times n$, $\boldsymbol{x}$ и $\boldsymbol{c}$ – n-мерные векторы, записанные в виде строки, и

α – некоторое число. Матрица Гессе квадратичной функции совпадает с Q: $H(\boldsymbol{x}) = Q$. Поэтому на основании теоремы 1.3 получаем, что квадратичная функция вогнута (выпукла) на всем пространстве $\mathbf{R}^n$ в том и только в том случае, когда ее матрица Q неположительно (неотрицательно) определена, т.е. неравенство $\boldsymbol{y} \cdot Q \cdot \boldsymbol{y}^T \leq (\geq)\, 0$ выполняется для всех $\boldsymbol{y} \in \mathbf{R}^n$.

§ 2. Основные понятия, связанные с экстремумами

В этом параграфе после рассмотрения прикладной задачи о выборе параметров функции дается постановка экстремальной задачи и обсуждаются различные типы экстремумов. Изучаются случаи, когда экстремальная задача разрешима, приводится классификация экстремальных задач.

1°. Задача выбора параметров функции. Предположим, что имеются две, например, физические скалярные величины x и y, зависимость между которыми выражается равенством $y = y(x)$. Для простоты будем считать, что зависимость эта – линейная, т.е.

$$y = ax + b \tag{1.1}$$

при некоторых числовых коэффициентах a и b, которые заранее не известны. Таким образом, хотя характер зависимости между x и y известен, конкретный ее вид не задан.

Предположим, что имеется возможность для частного значения x_i аргумента x измерить (вычислить) соответствующее значение y_i, $i = 1, 2, \ldots, m$ неизвестной функции (1.1). Пусть результаты подобных измерений занесены в таблицу 1.

*Таблица 1.1.*Таблица значений функции.

x	x_1	x_2	. . .	x_m
y	y_1	y_2	. . .	y_m

Естественно считать, что измерения были произведены с некоторой ошибкой, и поэтому если точки (x_i, y_i), $i = 1, 2, \ldots, m$, изобразить на декартовой плоскости, то они, вообще говоря, не будут лежать на одной прямой (см. рис.2.1.6).

Введем вектор ошибок измерений $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_m)$, где $\delta_i = y_i - ax_i - b$, $i = 1, 2, \ldots, m$. В идеальном случае, когда измерения выполнены без ошибок, должны найтись такие (“истинные”) значения параметров a и b, при которых все компоненты вектора ошибок δ равны нулю. При этих значениях параметров табл.1.1 в точности соответствует функции $y = ax + b$, и точки (x_i, y_i), $i = 1, 2, \ldots, m$, декартовой плоскости оказываются расположенными на одной прямой. На практике идеальный случай обычно не реализуется, так как измерения содержат ошибки. Поэтому, как правило, не существует таких значений a и b, при которых все компоненты вектора ошибок обращаются в нуль. В этом случае табл.1.1 не соответствует ни одной функции вида (1.1) и, тем самым, имеется “рассогласование” между функциональной зависимостью (1.1) и данными из табл.1.1.

Задача выбора параметров функции (1.1) заключается в отыскании таких значений коэффициентов a и b, при которых эмпирические данные табл.1.1 оказываются наилучшим образом согласованными с функциональной зависимостью (1.1). Геометрически наилучшая согласованность означает, что точки (x_i, y_i), $i = 1, 2, \ldots, m$, “наименее уклоняются” от прямой, являющейся графиком функции $y = ax + b$ (рис.2.1), поэтому задача выбора параметров функции превращается в задачу проведения соответствующей прямой на плоскости. При этом ясно, что решение этой задачи в сильной степени зависит от выбора меры “отклонения” точек от прямой.

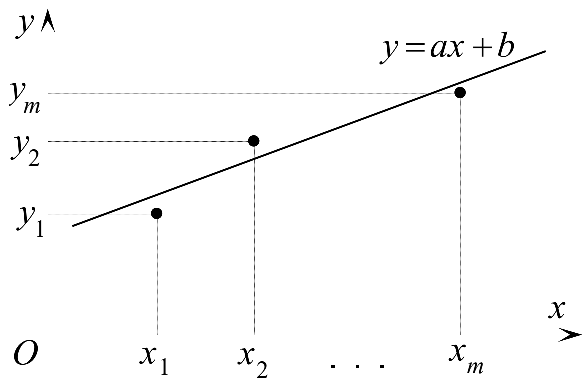

Рис. 2.1. Геометрия задачи о выборе параметров функции.

Рассмотрим два возможных подхода к решению поставленной задачи. Первый подход носит название *метода наименьших квадратов* и состоит в выборе параметров a и b таким образом, чтобы квадрат нормы вектора ошибок δ при этих значениях a и b был бы как можно меньше (как можно ближе к нулю). Если ввести функцию $F(a,b) = \|\delta\|^2$ двух переменных a и b, то соответствующая задача принимает вид

$$F(a,b) = \|\delta\|^2 = \sum_{i=1}^{m} (y_i - ax_i - b)^2 \to \min_{a,b}. \tag{2.2}$$

Второй подход связан с реализацией идеи *равномерного* (*чебышевского*) *приближения*. Здесь параметры a и b стремятся подобрать так, чтобы абсолютная величина самой большой ошибки оказалась как можно меньше. В этом случае добиваются “равномерного” приближения абсолютных величин ошибок к нулю. Соответствующая этому подходу математическая задача имеет вид

$$G(a,b) = \max\left\{|\delta_1|; \ldots; |\delta_m|\right\} = \max\left\{|y_1 - ax_1 - b|; \ldots; |y_m - ax_m - b|\right\} \to \min_{a,b}. \tag{2.3}$$

Итак, рассмотрены два возможных подхода к решению поставленной выше задачи выбора параметров функции, один из которых приводит к задаче (2.2), а второй – к задаче (2.3). В обеих задачах требуется минимизировать некоторую функцию двух переменных. Задачи подобного типа называются *экстремальными*. Изучению и решению таких задач посвящено последующее рассмотрение.

2°. Постановка экстремальной задачи. Различные типы экстремумов. Всякая *экстремальная задача* содержит два объекта: множество X, относительно которого будем считать, что оно является подмножеством пространства $\mathbf{R}^n$ (т.е. $X \subset \mathbf{R}^n$), и функцию f, заданную на этом множестве X. Экстремальную задачу нередко называют также *задачей оптимизации*, когда хотят подчеркнуть ее определенную прикладную направленность.

Элементы множества X, как и раньше, будем называть *точками* или *векторами*; кроме того, для этой цели будем использовать термин *решение*. Само множество X именуют *множеством допустимых* (или *возможных*) *решений* или, короче, – *допустимым множеством*. Что касается функции f, заданной на множестве X, то ее нередко называют *целевой функцией* или *критерием оптимальности*. Так, например, в рассмотренных выше экстремальных задачах (2.2) и (2.3) целевыми функциями являются $F(a,b)$ и $G(a,b)$, а допустимым множеством в обеих задачах служит все пространство $\mathbf{R}^2$.

Различают два вида экстремальных задач – *задачу максимизации* и *задачу минимизации*.

Решить *задачу максимизации* функции f на множестве X означает найти такой элемент $\boldsymbol{x}^* \in X$ (и соответствующее значение $f(\boldsymbol{x}^*)$), для которого неравенство

$$f(\boldsymbol{x}^*) \geq f(\boldsymbol{x}) \tag{2.4}$$

выполняется для всех $\boldsymbol{x} \in X$; при этом пишут

$$f(\boldsymbol{x}^*) = \max_{\boldsymbol{x} \in X} f(\boldsymbol{x}),$$

а для постановки задачи максимизации используют запись

$$f(\boldsymbol{x}) \to \max_{\boldsymbol{x} \in X}.$$

Здесь элемент $\boldsymbol{x}^*$ называют *точкой максимума*, а значение $f(\boldsymbol{x}^*)$ – *максимумом* или *наибольшим значением* функции f на множестве X.

Постановка *задачи минимизации* имеет вид

$$f(\boldsymbol{x}) \to \min_{\boldsymbol{x} \in X}.$$

Эта задача заключается в нахождении элемента $\boldsymbol{x}^* \in X$ (и соответствующего значения $f(\boldsymbol{x}^*)$), при котором неравенство

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) \tag{2.5}$$

имеет место для всех $\boldsymbol{x} \in X$. В таком случае пишут

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x} \in X} f(\boldsymbol{x}),$$

элемент $\boldsymbol{x}^*$ называют *точкой минимума*, а значение $f(\boldsymbol{x}^*)$ – *минимумом* или *наименьшим значением* функции f на множестве X.

Еще одно название для задачи максимизации и задачи минимизации – *задача оптимизации*. При этом точку максимума, а также точку минимума именуют *оптимальной точкой*, *точкой экстремума* или *оптимальным решением* данной оптимизационной задачи. Значение функции f, вычисленное в точке экстремума, называют *экстремумом, оптимальным значением* или *оптимумом*.

Как указано выше, решить задачу оптимизации означает найти оптимальную точку $\boldsymbol{x}^*$ и оптимальное значение $f(\boldsymbol{x}^*)$. Если найдена оптимальная точка, то определение оптимального значения обычно не составляет труда. Если же найдено лишь оптимальное значение $f(\boldsymbol{x}^*)$, то для отыскания оптимальной точки следует решить уравнение $f(\boldsymbol{x}) = f(\boldsymbol{x}^*)$, что нередко составляет сложную вычислительную задачу.

Использование введенных понятий нередко сопровождают дополнительным прилагательным “глобальный”, поскольку все эти понятия имеют свой “локальный” аналог.

Элемент $\boldsymbol{x}^* \in X$ называют *точкой локального минимума* (*локального максимума*), если найдется такая окрестность $u(\boldsymbol{x}^*)$ точки $\boldsymbol{x}^*$, для которой неравенство (2.4) (соответственно (2.5)) выполняется при всех $\boldsymbol{x} \in X \bigcap u(\boldsymbol{x}^*)$. При этом значение $f(\boldsymbol{x}^*)$, вычисленное в точке $\boldsymbol{x}^*$ локального максимума (локального минимума) называют *локальным максимумом* (*локальным минимумом*). Для введенных точек используют также термины – *точка локального экстремума* (*оптимума*), а для значений – *локальный экстремум* (*оптимум*).

Всякая точка глобального экстремума является точкой локального экстремума, так как из выполнения неравенства (2.4) или (2.5) для всех элементов $\boldsymbol{x}$ множества X вытекает выполнение этих неравенств для всех элементов $\boldsymbol{x}$ разве что более узкого множества $X \bigcap u(\boldsymbol{x}^*)$, где $u(\boldsymbol{x}^*)$ – окрестность точки $\boldsymbol{x}^*$, например, единичного радиуса.

Что касается обратного утверждения, то оно имеет место не всегда. В следующей теореме указан класс задач оптимизации, в котором локальные и глобальные экстремумы совпадают.

Теорема 2.1.Пусть непустое множество X, $X \subset \mathbf{R}^n$, выпукло. Если функция f, заданная на множестве X, является вогнутой (выпуклой), то всякая ее точка локального максимума (локального минимума) будет точкой глобального максимума (соответственно – точкой глобального
минимума.

▶ Приведем доказательство для случая вогнутой целевой функции (случай выпуклой функции рассматривается аналогично).

Пусть $\boldsymbol{x}^* \in X$ – точка локального максимума вогнутой функции f на выпуклом множестве X. Это означает, что найдется такая окрестность $u(\boldsymbol{x}^*)$, что

$$f(\boldsymbol{x}^*) \geq f(\boldsymbol{x}) \quad \text{для всех} \quad \boldsymbol{x} \in X \bigcap u(\boldsymbol{x}^*) . \tag{2.6}$$

Выберем произвольную точку $\bar{\boldsymbol{x}} \in X$. В силу выпуклости множества X имеем $\boldsymbol{x} = \lambda \bar{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^* \in X$ для всех $\lambda \in (0,1)$. Более того, так как $\left\| \boldsymbol{x} - \boldsymbol{x}^* \right\| = \lambda \left\| \bar{\boldsymbol{x}} - \boldsymbol{x}^* \right\| \xrightarrow[\lambda \to 0+]{} 0$, то $\boldsymbol{x} = \lambda \bar{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^* \in X \bigcap u(\boldsymbol{x}^*)$ при достаточно малом $\lambda \in (0,1)$. Выписывая для такой точки $\boldsymbol{x}$ неравенство (2.6), с использованием вогнутости f получим:

$$f(\boldsymbol{x}^*) \geq f\big(\lambda \bar{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^*\big) \geq \lambda f(\bar{\boldsymbol{x}}) + (1-\lambda) f(\boldsymbol{x}^*) .$$

Отсюда после несложных преобразований следует неравенство $f(\boldsymbol{x}^*) \geq f(\bar{\boldsymbol{x}})$, где $\bar{\boldsymbol{x}}$ – произвольная точка множества X. Таким образом, $\boldsymbol{x}^*$ – точка глобального максимума функции f на множестве X. ◀

3°. Условия разрешимости экстремальных задач. Сформулированная в предыдущем пункте экстремальная задача имеет оптимальное решение не при любых целевых функциях и допустимых множествах. Это относится как к локальному, так и к глобальному вариантам постановки задачи оптимизации. Рассмотрим условия, при которых экстремальная задача имеет оптимальное решение в глобальном смысле.

Если допустимое множество данной оптимальной задачи не содержит ни одного элемента, то, очевидно, оптимальных точек на существует. Поэтому при рассмотрении экстремальных задач обычно считают, что допустимое множество не пусто.

В случае конечного допустимого множества имеется конечный набор значений целевой функции, вычисленных в допустимых точках. В этом конечном наборе всегда имеется максимальное, а также минимальное число. Первое из них является глобальным максимумом, второе – глобальным минимумом. Таким образом, *если допустимое множество не пусто и конечно, то любая экстремальная задача* (*с любой целевой функцией*) имеет *оптимальное решение на этом множестве*.

Для бесконечного (возможно, неограниченного) допустимого множества справедлив следующий результат, который представляет собой упоминавшуюся ранее теорему Вейерштрасса (см. §2 гл.1).

Теорема 2.2. Предположим, что функция f определена на непустом компактном множестве X, $X \subset \mathbf{R}^n$, и непрерывна на нем. Тогда функция f достигает на множестве X своего глобального максимума (соответственно – глобального минимума).

4°. Классы экстремальных задач. Из выполнения неравенства $f(\boldsymbol{x}^*) \le f(\boldsymbol{x})$ для всех $\boldsymbol{x} \in X$ следует выполнение неравенства $h(\boldsymbol{x}^*) \ge h(\boldsymbol{x})$ для всех $\boldsymbol{x} \in X$, где $h(\boldsymbol{x}) = -f(\boldsymbol{x})$. Это означает, что всякая оптимальная точка задачи минимизации $f(\boldsymbol{x}) \to \min\limits_{\boldsymbol{x}\in X}$ является оптимальной точкой в задаче максимизации $-f(\boldsymbol{x}) \to \max\limits_{\boldsymbol{x}\in X}$, а оптимальные значения обеих задач связаны равенством

$$\min_{\boldsymbol{x}\in X} f(\boldsymbol{x}) = -\max_{\boldsymbol{x}\in X}\left(-f(\boldsymbol{x})\right).$$

Благодаря указанному обстоятельству любая задача минимизации может быть сведена к эквивалентной ей задаче максимизации на том же самом допустимом множестве, если целевую функцию умножить на -1. Более того, нетрудно понять, что отмеченная эквивалентность сохраняет свою силу и в тех случаях, когда речь идет о локальном оптимальном решении.

На основании сказанного последующее рассмотрение будет, в основном, ограничено кругом задач максимизации. Все полученные для этих задач результаты могут быть очевидным образом переформулированы применительно к задачам минимизации.

Итак, рассмотрим задачу максимизации $f(\boldsymbol{x}) \to \max\limits_{\boldsymbol{x}\in X}$. Если в этой задаче $X = \mathbf{R}^n$, то говорят о *задаче безусловной максимизации*, или о *задаче максимизации без ограничений*. Если же допустимое множество не совпадает со всем пространством, т.е. $X \ne \mathbf{R}^n$, то используют термин *задача условной максимизации,* или *задача с ограничениями*. При этом ко всем введенным ранее понятиям экстремальных точек и экстремумов добавляют прилагательное “безусловный”, или “условный”. Например, тот факт, что $\boldsymbol{x}^* \in X$ является точкой локального условного максимума функции f на множестве *X*, означает существование такой окрестности $u(\boldsymbol{x}^*)$ точки $\boldsymbol{x}^*$, что выполнено неравенство $f(\boldsymbol{x}^*) \ge f(\boldsymbol{x})$ для всех $\boldsymbol{x} \in X \bigcap u(\boldsymbol{x}^*)$, причем $X \ne \mathbf{R}^n$.

Нередко в задаче условной оптимизации допустимое множество представляет собой множество решений некоторой системы уравнений

$$X = \left\{\boldsymbol{x} \in \mathbf{R}^n \,\middle|\, g_j(\boldsymbol{x}) = 0,\ j = 1, 2, \ldots, k \right\},$$

системы неравенств

$$X = \left\{\boldsymbol{x} \in \mathbf{R}^n \,\middle|\, g_j(\boldsymbol{x}) \ge 0,\ j = 1, 2, \ldots, k \right\},$$

или – в общем случае – системы, содержащей как равенства, так и неравенства. Здесь $g_j :\ \mathbf{R}^n \to \mathbf{R}$, $j = 1, 2, \ldots, k$, – некоторые числовые функции. В этом случае говорят о *задаче математического программирования*.

Среди задач математического программирования наиболее проста и хорошо изучена *задача линейного программирования*, в которой все участвующие в ее постановке функции являются линейными (точнее говоря, аффинными):

$$f(\boldsymbol{x}) = \langle \boldsymbol{c}, \boldsymbol{x} \rangle, \quad g_j(\boldsymbol{x}) = \langle \boldsymbol{a}^j, \boldsymbol{x} \rangle - b_j, \quad j = 1, 2, \dots, k,$$

где $\boldsymbol{c}, \boldsymbol{a}^j \in \mathbf{R}^n$, $b_j \in \mathbf{R}$, $j = 1, 2, \dots, k$.

§ 3. Безусловный экстремум

Здесь формулируются утверждения, с помощью которых появляется возможность решать задачи на безусловный экстремум.

1°. Необходимое условие первого порядка. Следующая теорема представляет собой распространение на случай функции нескольких переменных известной теоремы Ферма о том, что в точке локального экстремума производная целевой функции равна нулю.

Теорема 1. Пусть $\boldsymbol{x}^*$ – точка локального безусловного экстремума функции f и в этой точке существуют все частные производные первого порядка указанной функции. Тогда они обращаются в нуль, т.е.

$$\frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} = 0, \quad i = 1, 2, \dots, n.$$

▶ По условию найдется окрестность $u(\boldsymbol{x}^*)$, для которой выполнено неравенство

$$f(\boldsymbol{x}^*) \geq (\leq) f(\boldsymbol{x}) \quad \text{при всех} \quad \boldsymbol{x} \in u(\boldsymbol{x}^*). \tag{1}$$

Выберем $i = 1$ и зафиксируем значения всех переменных функции $f(x_1, x_2, \dots, x_n)$, кроме первой, следующим образом: $f(x_1, x_2^*, \dots, x_n^*)$, где x_2^*, … , x_n^* – соответствующие компоненты точки $\boldsymbol{x}^*$ локального безусловного экстремума. В итоге будет получена некоторая функция $f(x_1, x_2^*, \dots, x_n^*)$ одной переменной x_1.

Рассмотрим окрестность $u(x_1^*)$ точки x_1^* на числовой оси, имеющую такой же радиус, что и упомянутая выше окрестность $u(\boldsymbol{x}^*)$. Нетрудно понять, что выполняется включение $(x_1, x_2^*, \dots, x_n^*) \in u(\boldsymbol{x}^*)$ при всех $x_1 \in u(x_1^*)$. Благодаря этому включению из (1) следует неравенство $f(x_1^*, x_2^*, \dots, x_n^*) \geq (\leq) f(x_1, x_2^*, \dots, x_n^*)$ при всех $x_1 \in u(x_1^*)$, которое означает, что x_1^* – точка локального максимума (локального минимума) введенной функции одной переменной. Согласно теореме Ферма производная этой функции в точке x_1^* должна обращаться в нуль:

$$\left. \frac{df(x_1, x_2^*, \ldots, x_n^*)}{dx_1} \right|_{x_1 = x_1^*} = 0 .$$

Но записанная в левой части равенства производная функции одной переменной совпадает с частной производной функции n переменных f, вычисленной в точке $\boldsymbol{x}^*$, поэтому

$$\frac{\partial f(\boldsymbol{x}^*)}{\partial x_1} = 0 .$$

Аналогично устанавливается равенство нулю всех остальных частных производных функции f. ◄

Градиент $\nabla f(\boldsymbol{x}^*)$ функции f в точке $\boldsymbol{x}^*$ – это n-мерный вектор, составленный из частных производных функции f, вычисленных в точке $\boldsymbol{x}^*$. Поэтому необходимое условие экстремума теоремы 1 в векторной форме принимает вид:

$$\nabla f(\boldsymbol{x}^*) = \mathbf{0} . \tag{2}$$

Это условие называют условием *первого порядка* из-за того, что в его формулировке участвуют частные производные первого порядка.

Определение. Всякое решение системы уравнений

$$\begin{cases} \dfrac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_1} = 0, \\ \dfrac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_2} = 0, \\ . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad ., \\ \dfrac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_n} = 0 \end{cases} \tag{3}$$

называют *стационарной точкой* функции f.

С использованием введенного понятия в предположении существования всех частных производных функции f в $\boldsymbol{x}^*$ утверждение теоремы 1 можно сформулировать следующим образом: *для того, чтобы $\boldsymbol{x}^*$ была точкой локального безусловного экстремума функции f, необходимо, чтобы она являлась стационарной точкой этой функции*.

На основе последнего утверждения можно предложить следующий путь решения экстремальной задачи на безусловный экстремум. В предположении, что точки экстремума существуют, следует найти все стационарные точки целевой функции. Для этого необходимо решить систему уравнений (3). Найденные точки будут “подозрительными” на оптимальные. В случае, если полученное множество точек окажется “обозримым”, сравнив значения целевой функции в найденных точках, выделить среди них те, в которых целевая функция принимает наибольшее (наименьшее) значение.

Здесь, однако, необходимо сделать следующие замечания. Во-первых, для реализации описанного способа решения экстремальной задачи условие существования экстремальных точек является существенным (в этом убеждает простой пример функции одной переменной $y = x^3$). Во-вторых, может статься, что процесс решения системы уравнений (3) окажется слишком сложным с вычислительной точки зрения. Наконец, множество стационарных точек не всегда является “обозримым”; в таких случаях процесс выделения оптимальных точек среди стационарных точек может сопровождаться значительными вычислительными трудностями.

2°. Достаточное условие первого порядка. Существует класс функций, для которых полученное в предыдущем пункте необходимое условие экстремума является и достаточным. Этот класс функций указан в следующей теореме.

Теорема 2. Предположим, что функция f дифференцируема и вогнута (выпукла) на всем пространстве $\mathbf{R}^n$. Если в точке $\boldsymbol{x}^* \in \mathbf{R}^n$ выполнено необходимое условие экстремума (2), т.е. если $\boldsymbol{x}^*$ – стационарная точка функции f, то эта точка является точкой безусловного глобального максимума (соответственно минимума) функции f.

▶ Доказательство немедленно следует из теоремы 6 §1 при $X = \mathbf{R}^n$, $\boldsymbol{x}'' = \boldsymbol{x}^*$ и произвольном $\boldsymbol{x}' = \boldsymbol{x} \in \mathbf{R}^n$. ◀

В предыдущем параграфе была сформулирована задача выбора параметров функции. Подход к решению этой задачи, носящий название *метода наименьших квадратов*, сводится к минимизации целевой функции

$$F(a,b) = \sum_{i=1}^{m} (y_i - ax_i - b)^2$$

двух переменных a и b на всем пространстве $\mathbf{R}^2$.

Рассмотрим следующую функцию двух переменных

$$f_i(a,b) = (y_i - ax_i - b)^2 .$$

Найдем ее частные производные второго порядка:

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

Матрица Гессе для рассматриваемой функции имеет вид

$$H = \begin{pmatrix} 2x_i^2 & 2x_i \\ 2x_i & 2 \end{pmatrix}.$$

Поэтому

$$\boldsymbol{y}^T \cdot H \cdot \boldsymbol{y} = 2x_i^2 y_1^2 + 2y_2^2 + 4x_i y_1 y_2 = 2(x_i y_1 + y_2)^2 \ge 0$$

для всех $\boldsymbol{y} = (y_1, y_2) \in \mathbf{R}^2$. Отсюда на основании теоремы 7 §1 следует выпуклость функции

$f_i(a,b)$ на всем пространстве $\mathbf{R}^2$. Но сумма выпуклых функций – выпуклая функция (см. теорему 2 §1). Следовательно, функция $F(a,b)=\sum_{i=1}^{m} f_i(a,b)$ выпукла на $\mathbf{R}^2$, а значит, каждая ее стационарная точка является точкой безусловного глобального минимума.

Таким образом, для того, чтобы решить задачу выбора параметров функции в соответствии с методом наименьших квадратов следует приравнять обе частные производные функции $F(a,b)$ нулю. Решая полученную систему двух линейных уравнений относительно a и b, будет найдено оптимальное решение задачи выбора параметров функции.

Пример. Предположим, что имеется следующая таблица вычисленных значений неизвестной функции $y = ax + b$.

Таблица 2. Таблица значений функции.

x	0	1	2	3
y	3.0	2.5	1.5	0.3

Составляем целевую функцию

$$F(a,b) = (3-b)^2 + (2.5-a-b)^2 + (1.5-2a-b)^2 + (0.3-3a-b)^2 .$$

Находим ее частные производные:

$$\frac{\partial F}{\partial a} = 28a + 12b - 12.8,$$
$$\frac{\partial F}{\partial b} = 12a + 8b - 14.6.$$

Приравнивая найденные частные производные нулю, получаем систему линейных уравнений

$$\begin{cases} 7a + 3b = 3.2, \\ 3a + 2b = 3.65, \end{cases}$$

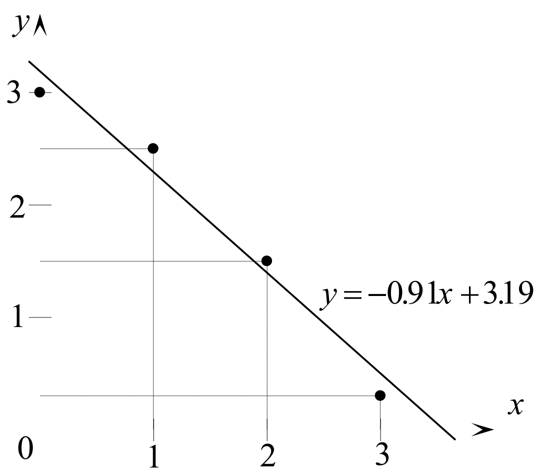

Рис. 3.7. Геометрия задачи выбора параметров функции.

имеющую единственное решение $a^* = -0.91$, $b^* = 3.19$. В силу теоремы 2 найденная стационарная точка (a^*, b^*) является точкой глобального безусловного минимума функции $F(a,b)$. Следовательно, решением данной задачи выбора параметров является функция $y = -0.91x + 3.19$ (рис.3.7).

Необходимо отметить, что разобранный способ решения задачи минимизации функции $F(a,b)$ с использованием теорем 1 и 2 не приемлем при реализации второго подхода к решению задачи выбора параметров функции, так как соответствующая этому подходу целевая функция

$$G(a,b) = \max\left\{\left| y_1 - ax_1 - b\right| ; \ldots ; \left| y_m - ax_m - b\right|\right\}$$

в общем случае не является дифференцируемой в каждой точке пространства $\mathbf{R}^2$.

3°. Оптимальные точки квадратичной функции. Рассмотрим квадратичную функцию

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}\cdot Q\cdot\boldsymbol{x}^T + \boldsymbol{c}\cdot\boldsymbol{x}^T + \alpha\,, \tag{4}$$

где Q – симметричная числовая матрица размера $n\times n$, $\boldsymbol{x}$ и $\boldsymbol{c}$ – n-мерные векторы, записанные в виде строк, и α – некоторое число.

Ранее было найдено (п.6° §1 гл.2), что вектор-строка частных производных первого порядка для этой квадратичной функции имеет вид:

$$\nabla f(\boldsymbol{x}) = \boldsymbol{x}\cdot Q + \boldsymbol{c}\,.$$

Таким образом, в данном случае необходимое условие экстремума первого порядка выглядит следующим образом:

$$\boldsymbol{x}\cdot Q + \boldsymbol{c} = \mathbf{0}\,,$$

или после транспонирования с учетом симметричности матрицы Q:

$$Q\cdot\boldsymbol{x}^T = -\boldsymbol{c}^T\,. \tag{5}$$

Равенство (5) имеет вид системы линейных уравнений относительно вектора неизвестных $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.

В §1 было установлено, что квадратичная функция (4) вогнута (выпукла) на всем пространстве $\mathbf{R}^n$ тогда и только тогда, когда матрица Q неположительно (неотрицательно) определена, т.е. когда выполнено неравенство $\boldsymbol{y}\cdot Q\cdot\boldsymbol{y}^T \le (\ge)\, 0$ для всех $\boldsymbol{y}\in\mathbf{R}^n$. В соответствии с этим на основании теоремы 2 приходим к следующему выводу.

Пусть матрица Q *квадратичной функции* (4) *неположительно* (*неотрицательно*) *определена. Точка* $\boldsymbol{x}\in\mathbf{R}^n$ *является точкой безусловного глобального максимума* (*минимума*) *квадратичной функции* (4) *тогда и только тогда, когда она удовлетворяет системе линейных уравнений* (5). В частности, если система линейных уравнений (5) несовместна, то оптимальных точек у данной квадратичной функции не существует.

4°. Необходимое условие второго порядка. В этом условии участвуют элементы матрицы Гессе целевой функции, т.е. частные производные второго порядка. Потому его и называют условием второго порядка.

Теорема 3. Предположим, что функция f имеет непрерывные частные производные второго порядка в каждой точке некоторой окрестности точки $\boldsymbol{x}^*\in\mathbf{R}^n$. Если $\boldsymbol{x}^*$ – точка локального безусловного максимума (минимума) функции f, то имеет место равенство (2) и, кроме того, матрица Гессе $H(\boldsymbol{x}^*)$, вычисленная в точке $\boldsymbol{x}^*$, неположительно (неотрицательно) определена, т.е.

$$\boldsymbol{y}\cdot H(\boldsymbol{x}^*)\cdot\boldsymbol{y}^T \le (\ge)\, 0 \quad \text{для всех} \quad \boldsymbol{y}\in\mathbf{R}^n .$$

▶ Тот факт, что в точке $\boldsymbol{x}^*$ локального безусловного экстремума градиент обращается в нулевой вектор, следует прямо из теоремы 1.

Докажем неположительную определенность матрицы Гессе в точке $\boldsymbol{x}^*$ локального безусловного максимума функции *f*. Для доказательства предположим противное: существует вектор $\boldsymbol{y}\in\mathbf{R}^n$, для которого выполняется противоположное неравенство $\boldsymbol{y}\cdot H(\boldsymbol{x}^*)\cdot\boldsymbol{y}^T > 0$. С учетом равенства (2) запишем для функции f в окрестности точки $\boldsymbol{x}^*$ формулу Тейлора с остаточным членом в форме Пеано:

$$f(\boldsymbol{x}^*+t\boldsymbol{y}) = f(\boldsymbol{x}^*)+\frac{1}{2}(t\boldsymbol{y})\cdot H(\boldsymbol{x}^*)\cdot(t\boldsymbol{y})^T+\|\, t\boldsymbol{y}\|^2\cdot\varepsilon(t\boldsymbol{y}),$$

где $\varepsilon(t\boldsymbol{y})\to 0$ при $t\to 0+$ и положительное число t выбирается настолько малым, чтобы точка $\boldsymbol{x}^*+t\boldsymbol{y}$ принадлежала окрестности точки $\boldsymbol{x}^*$, о которой идет речь в формулировке теоремы и в которой существуют частные производные второго порядка функции *f*. Последнее равенство можно переписать так:

$$f(\boldsymbol{x}^*+t\boldsymbol{y})-f(\boldsymbol{x}^*)=\frac{t^2}{2}\Big[\,\boldsymbol{y}\cdot H(\boldsymbol{x}^*)\cdot\boldsymbol{y}^T+2\|\,\boldsymbol{y}\|^2\cdot\varepsilon(t\boldsymbol{y})\,\Big].$$

Так как первое слагаемое, записанное в квадратных скобках, положительно и $\lim\limits_{t\to 0+}\varepsilon(t\boldsymbol{y})=0$, то для всех достаточно малых положительных t правая часть рассматриваемого равенства будет положительна. Положительность левой части того же равенства противоречит тому, что $\boldsymbol{x}^*$ – точка локального максимума. Для задачи максимизации теорема доказана.

Утверждение теоремы 3 для задачи минимизации следует из доказанного применительно к функции $-f$. ◀

Если функция f имеет лишь одну переменную, т.е. $n=1$, то $H(x_1^*)=f''(x_1^*)$, и условие неположительной определенности матрицы Гессе записывается следующим образом:

$$f''(x_1^*)\cdot y_1^2\le 0 \quad \text{для всех} \quad y_1\in\mathbf{R}.$$

Нетрудно понять, что это условие равносильно неположительности производной второго порядка $f''(x_1^*)$ в точке максимума x_1^*. Таким образом, в частном случае, когда функция f имеет одну переменную, теорема 3 совпадает с полученным ранее необходимым условием экстремума второго порядка.

5°. Достаточные условия второго порядка. Выделим еще один тип экстремальных точек.

Определение. Пусть функция f определена на множестве $X \subset \mathbf{R}^n$. Элемент $\boldsymbol{x}^* \in X$ называют *точкой строгого максимума* (*строгого локального максимума*) функции f, если неравенство $f(\boldsymbol{x}^*) > f(\boldsymbol{x})$ выполняется для всех $\boldsymbol{x} \in X$, $\boldsymbol{x} \neq \boldsymbol{x}^*$ (соответственно для всех $\boldsymbol{x} \in X \bigcap u(\boldsymbol{x}^*)$, $\boldsymbol{x} \neq \boldsymbol{x}^*$, где $u(\boldsymbol{x}^*)$ – некоторая окрестность точки $\boldsymbol{x}^*$).

Аналогичным образом определяются точки строгого минимума (строгого максимума) и строгого локального минимума (строгого локального максимума). Все эти точки называют *точками строгого* (локального) *экстремума*.

Нетрудно понять, что точка строгого максимума (и минимума) единственна.

Точка обычного экстремума не всегда бывает четко выраженной в том смысле, что иногда неясно, является ли она точкой максимума или точкой минимума. Например, каждая точка постоянной функции $f(\boldsymbol{x}) = \mathrm{const}$ одновременно будет как точкой максимума, так и точкой минимума этой функции. Что касается точек строгого экстремума, то они лишены указанного недостатка. Точка строгого (локального) максимума не может быть одновременно и точкой строгого (локального) минимума; это следует прямо из определений этих точек.

Теорема 1. Предположим, что функция f имеет непрерывные частные производные второго порядка в каждой точке некоторой окрестности точки $\boldsymbol{x}^* \in \mathbf{R}^n$. Кроме того, предположим, что точка $\boldsymbol{x}^*$ является стационарной точкой функции f. Тогда

1) если матрица Гессе

$$H(\boldsymbol{x}^*) = \begin{pmatrix} \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1^2} & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2 \partial x_n} \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot\ \cdot \\ \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_n^2} \end{pmatrix}$$

целевой функции f, вычисленная в точке $\boldsymbol{x}^*$, отрицательно определена, т.е. выполнено $\boldsymbol{y} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{y}^T < 0$ для всех $\boldsymbol{y} \neq \boldsymbol{0}$, то $\boldsymbol{x}^*$ – точка строгого локального максимума функции f;

2) если матрица Гессе $H(\boldsymbol{x}^*)$ положительно определена, т.е. выполнено $\boldsymbol{y} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{y}^T > 0$ для всех $\boldsymbol{y} \neq \boldsymbol{0}$, то $\boldsymbol{x}^*$ – точка строгого локального минимума функции f;

3) если матрица Гессе $H(\boldsymbol{x}^*)$ такова, что при некоторых векторах $\boldsymbol{y}, \boldsymbol{z} \in \mathbf{R}^n$ имеют место одновременно оба неравенства $\boldsymbol{y} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{y}^T < 0$, $\boldsymbol{z} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{z}^T > 0$, то $\boldsymbol{x}^*$ не является ни точкой локального максимума, ни точкой локального минимума функции f.

* ▶ Докажем первую часть теоремы. По условию матрица Гессе $H(\boldsymbol{x}^*)$ отрицательно определена. Покажем, что в этом случае функция f имеет отрицательно определенную матрицу Гессе $H(\boldsymbol{x})$ не только в точке $\boldsymbol{x}^*$, но и в каждой точке $\boldsymbol{x}$ некоторой окрестности точки $\boldsymbol{x}^*$.

Допустим противное: в любой проколотой окрестности точки $\boldsymbol{x}^*$, в которой f имеет частные производные второго порядка, матрица Гессе $H(\boldsymbol{x})$ не является отрицательно определенной.

Рассмотрим произвольную последовательность точек $\{\boldsymbol{x}^{(k)}\}$, сходящуюся к точке $\boldsymbol{x}^*$, так что $\lim\limits_{k\to\infty}\boldsymbol{x}^{(k)}=\boldsymbol{x}^*$. Благодаря сделанному допущению, для каждого члена $\boldsymbol{x}^{(k)}$ указанной последовательности существует такой ненулевой вектор $\boldsymbol{y}^{(k)}$, что выполняется неравенство $\boldsymbol{y}^{(k)}\cdot H(\boldsymbol{x}^{(k)})\cdot(\boldsymbol{y}^{(k)})^T\geq 0$, $k=1,2,\ldots$. Разделив обе части этого неравенства на $\left\|\boldsymbol{y}^{(k)}\right\|^2$, получим

$$\boldsymbol{u}^{(k)}\cdot H(\boldsymbol{x}^{(k)})\cdot(\boldsymbol{u}^{(k)})^T\geq 0,\quad k=1,2,\ldots\ , \tag{6}$$

где $\boldsymbol{u}^{(k)}=\dfrac{\boldsymbol{y}^{(k)}}{\left\|\boldsymbol{y}^{(k)}\right\|}$. Для векторов $\boldsymbol{u}^{(k)}$ выполнено равенство

$$\left\|\boldsymbol{u}^{(k)}\right\|=\left\|\frac{\boldsymbol{y}^{(k)}}{\left\|\boldsymbol{y}^{(k)}\right\|}\right\|=\frac{1}{\left\|\boldsymbol{y}^{(k)}\right\|}\cdot\left\|\boldsymbol{y}^{(k)}\right\|=1,\quad k=1,2,\ldots\ ,$$

из которого следует ограниченность последовательности $\{\boldsymbol{u}^{(k)}\}$. По теореме Больцано–Вейерштрасса (п.3 §1 гл.1) из ограниченной последовательности $\{\boldsymbol{u}^{(k)}\}$ можно выделить сходящуюся подпоследовательность, которую обозначим через $\{\boldsymbol{u}^{(k_l)}\}$. Пусть $\lim\limits_{l\to\infty}\boldsymbol{u}^{(k_l)}=\boldsymbol{u}^*$, причем, как нетрудно понять, $\left\|\boldsymbol{u}^*\right\|=1$. Для членов указанной подпоследовательности имеет место неравенство (6), т.е. выполнено $\boldsymbol{u}^{(k_l)}\cdot H\left(\boldsymbol{x}^{(k_l)}\right)\cdot\left(\boldsymbol{u}^{(k_l)}\right)^T\geq 0$, $l=1,2,\ldots$. Переходя здесь к пределу при $l\to\infty$ с учетом непрерывности элементов матрицы Гессе $H(\boldsymbol{x})$, получим неравенство

$$\boldsymbol{u}^*\cdot H(\boldsymbol{x}^*)\cdot(\boldsymbol{u}^*)^T\geq 0,\quad \boldsymbol{u}^*\neq\boldsymbol{0}.$$

Оно не совместимо с условием доказываемой теоремы о том, что матрица $H(\boldsymbol{x}^*)$ в точке $\boldsymbol{x}^*$ отрицательно определена. Следовательно, сделанное в начале доказательства допущение не справедливо и, значит, матрица Гессе $H(\boldsymbol{x})$ является отрицательно определенной в каждой точке некоторой окрестности точки $\boldsymbol{x}^*$, которую обозначим через $u(\boldsymbol{x}^*)$.

Теперь выпишем формулу Тейлора для функции f в окрестности стационарной точки $\boldsymbol{x}^*$ для произвольного $\boldsymbol{x} \in u(\boldsymbol{x}^*)$ $(\boldsymbol{x} \neq \boldsymbol{x}^*)$:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*) \cdot H\big(\boldsymbol{x}^* + \theta(\boldsymbol{x}-\boldsymbol{x}^*)\big) \cdot (\boldsymbol{x}-\boldsymbol{x}^*)^T,$$

где $0 < \theta < 1$. Так как $\boldsymbol{x}^* + \theta(\boldsymbol{x}-\boldsymbol{x}^*) = \theta\boldsymbol{x} + (1-\theta)\boldsymbol{x}^* \in u(\boldsymbol{x}^*)$ при всех $\theta \in (0,1)$, то в силу доказанного выше матрица Гессе $H(\boldsymbol{x})$ в каждой точке $\boldsymbol{x}^* + \theta(\boldsymbol{x}-\boldsymbol{x}^*)$, $0 < \theta < 1$, отрицательно определена. Поэтому

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*) \cdot H\big(\boldsymbol{x}^* + \theta(\boldsymbol{x}-\boldsymbol{x}^*)\big) \cdot (\boldsymbol{x}-\boldsymbol{x}^*)^T < 0$$

для всех $\boldsymbol{x} \in u(\boldsymbol{x}^*)$, $\boldsymbol{x} \neq \boldsymbol{x}^*$. Полученное означает, что $\boldsymbol{x}^*$ – точка строгого локального максимума функции f. Первая часть теоремы доказана.

Вторая часть теоремы вытекает из первой, если вместо задачи максимизации функции f рассмотреть ей эквивалентную задачу максимизации функции $-f$.

Перейдем к доказательству третьей части теоремы. Пусть вектор $\boldsymbol{y} \in \mathbf{R}^n$ таков, что для него по условию выполняется неравенство $\boldsymbol{y} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{y}^T < 0$. С учетом стационарности точки $\boldsymbol{x}^*$ запишем формулу Тейлора в окрестности точки $\boldsymbol{x}^*$:

$$f(\boldsymbol{x}^* + t\boldsymbol{y}) = f(\boldsymbol{x}^*) + \frac{1}{2}(t\boldsymbol{y}) \cdot H(\boldsymbol{x}^*) \cdot (t\boldsymbol{y})^T + \| t\boldsymbol{y} \|^2 \cdot \varepsilon(t\boldsymbol{y}),$$

$$f(\boldsymbol{x}^* + t\boldsymbol{y}) - f(\boldsymbol{x}^*) = \frac{t^2}{2}\Big[\, \boldsymbol{y} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{y}^T + 2\| \boldsymbol{y} \|^2 \cdot \varepsilon(t\boldsymbol{y}) \Big],$$

где положительные числа t таковы, что точка $\boldsymbol{x}^* + t\boldsymbol{y}$ принадлежит той окрестности точки $\boldsymbol{x}^*$, в которой существуют частные производные второго порядка функции f. По условию первое слагаемое в квадратных скобках последнего равенства отрицательно и $\lim\limits_{t \to 0+} \varepsilon(t\boldsymbol{y}) = 0$. Поэтому для всех достаточно малых положительных t правая часть указанного равенства отрицательна. Отрицательность левой части того же равенства означает, что точка $\boldsymbol{x}^*$ не может быть точкой локального максимума функции f.

Аналогично с использованием неравенства $\boldsymbol{z} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{z}^T > 0$ можно установить, что $\boldsymbol{x}^*$ не может быть также и точкой локального минимума функции f. ◀

На практике проверку условия положительной или отрицательной определенности матрицы Гессе осуществляют с помощью *критерия Сильвестра*, доказанного в курсе линейной алгебры. Он представляет собой необходимое и достаточное условие положительной (отрицательной) определенности квадратичной формы $\varphi = \boldsymbol{x} \cdot Q \cdot \boldsymbol{x}^T$, где Q –

симметричная квадратная матрица размера $n \times n$, а $\boldsymbol{x}$ – n-мерный вектор, записанный в виде строки.

Напомним, что указанная квадратичная форма называется *положительно* (*отрицательно*) *определенной*, если $\boldsymbol{x} \cdot Q \cdot \boldsymbol{x}^T > (<) 0$ для всех $\boldsymbol{x} \neq \mathbf{0}$.

Нетрудно видеть, что квадратичная форма $\boldsymbol{x} \cdot Q \cdot \boldsymbol{x}^T$ положительно (отрицательно) определена в том и только в том случае, когда ее матрица Q является положительно (отрицательно) определенной. Поэтому критерий Сильвестра применительно к матрице Гессе $H(\boldsymbol{x}^*)$ функции f может быть сформулирован следующим образом:

Для того, чтобы матрица $H(\boldsymbol{x}^*)$ *была положительно определенной, необходимо и достаточно, чтобы все угловые миноры этой матрицы были положительны,* т.е.

$$\Delta_k = \begin{vmatrix} \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1^2} & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1 \partial x_2} & \ldots & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1 \partial x_k} \\ \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2^2} & \ldots & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2 \partial x_k} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_k \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_k \partial x_2} & \ldots & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_k^2} \end{vmatrix} > 0, \quad k = 1, 2, \ldots, n.$$

Для того, чтобы матрица $H(\boldsymbol{x}^*)$ *была отрицательно определенной, необходимо и достаточно, чтобы знаки угловых миноров этой матрицы чередовались следующим образом*: $\Delta_1 < 0$, $\Delta_2 > 0$, $\Delta_3 < 0$, … , $(-1)^n \cdot \Delta_n > 0$.

Если f – функция двух переменных x_1 и x_2, то

$$H(\boldsymbol{x}^*) = \begin{pmatrix} \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1^2} & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2^2} \end{pmatrix},$$

причем $\dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1 \partial x_2} = \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2 \partial x_1}$. В этом случае

$$\Delta_1 = \frac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1^2}, \quad \Delta_2 = \begin{vmatrix} \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1^2} & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2^2} \end{vmatrix}.$$

Обозначим $A = \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1^2}$, $B = \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_1 \partial x_2} = \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2 \partial x_1}$, $C = \dfrac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_2^2}$. Тогда

$$\Delta_1 = A\,, \quad \Delta_2 = \begin{vmatrix} A & B \\ B & C \end{vmatrix},$$

и из теоремы 4 получаем следующий результат.

Следствие. Пусть функция f двух переменных имеет непрерывные частные производные второго порядка в каждой точке некоторой окрестности стационарной точки $\boldsymbol{x}^* \in \mathbf{R}^2$. Тогда

1) если $A < 0$, $AC - B^2 > 0$, то $\boldsymbol{x}^*$ – точка строгого локального максимума функции *f*;

2) если $A > 0$, $AC - B^2 > 0$, то $\boldsymbol{x}^*$ – точка строгого локального минимума функции *f*;

3) если $AC - B^2 < 0$, то $\boldsymbol{x}^*$ не является ни точкой локального максимума, ни точкой локального минимума функции *f*.

* ▶ Первые две части следствия вытекают прямо из соответствующих частей теоремы 4. Убедимся в справедливости третьей части следствия.

Имеем

$$\boldsymbol{y} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{y}^T = \begin{pmatrix} y_1 & y_2 \end{pmatrix} \cdot \begin{pmatrix} A & B \\ B & C \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = Ay_1^2 + 2By_1y_2 + Cy_2^2 .$$

Если $A = C = 0$, то выражение $\boldsymbol{y} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{y}^T = 2By_1y_2$ может принимать как положительные, так и отрицательные значения в зависимости от y_1 и y_2. В соответствии с третьей частью теоремы 4 из этого следует, что $\boldsymbol{x}^*$ не является ни точкой локального максимума, ни точкой локального минимума.

Если $A \neq 0$, то, как легко проверить, имеет место представление

$$\boldsymbol{y} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{y}^T = \frac{1}{A}\left[(Ay_1 + By_2)^2 + (AC - B^2)y_2^2\right].$$

Из этого представления видно, что при $y_1 \neq 0$, $y_2 = 0$ знак выражения $\boldsymbol{y} \cdot H(\boldsymbol{x}^*) \cdot \boldsymbol{y}^T$ совпадает со знаком числа *A*. Если же положить $y_1 = B$, $y_2 = -A$, то знак указанного выражения будет противоположен знаку числа A из-за того, что по условию выполнено $AC - B^2 < 0$. Таким образом, на основании третьей части теоремы 4 снова делаем вывод о том, что $\boldsymbol{x}^*$ не является ни точкой локального максимума, ни точкой локального минимума функции *f*.

Оставшийся случай $A = 0$, $C \neq 0$ рассматривается аналогично случаю $A \neq 0$. ◀

Пример. Рассмотрим целевую функцию двух переменных $f(x_1, x_2) = x_1^3 + x_2^3 - 3x_1x_2$. Вычисляем частные производные этой функции и приравниваем их нулю:

$$\begin{cases} \dfrac{\partial f}{\partial x_1} = 3x_1^2 - 3x_2 = 0, \\ \dfrac{\partial f}{\partial x_2} = 3x_2^2 - 3x_1 = 0. \end{cases}$$

Полученная система из двух уравнений имеет два решения – $(1,1)$ и $(0,0)$. Находим

$$\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} = 6x_1, \quad \frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2} = -3, \quad \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} = 6x_2.$$

Для стационарной точки $(1,1)$ имеем $A = 6$, $B = -3$, $C = 6$, так что $A > 0$, $AC - B^2 = 27 > 0$. Поэтому $(1,1)$ – точка строгого локального минимума.

Для второй стационарной точки $(0,0)$ получаем $A = C = 0$, $B = -3$. И так как $AC - B^2 = -9 < 0$, то эта точка не является ни точкой локального максимума, ни точкой локального минимума.

Замечание. В доказанном следствии разобраны все возможности как при $\Delta_2 = AC - B^2 < 0$, так и при $\Delta_2 > 0$. Однако в этом следствии ничего не говорится о том случае, когда $\Delta_2 = 0$. Оказывается (простые примеры убеждают в этом), при $\Delta_2 = 0$ ничего определенного о точке $\boldsymbol{x}^*$ сказать нельзя, так как для одних целевых функций она в данном случае может оказаться точкой экстремума, тогда как для других – она таковой не будет.

§ 4. Условный экстремум

В этом параграфе рассматривается задача на экстремум с ограничениями в форме равенств, и для нее формулируются соответствующие условия экстремума в терминах так называемой функции Лагранжа.

1°. Функция Лагранжа. Необходимое условие экстремума. Предметом изучения данного параграфа является задача на экстремум функции f при ограничениях $g_j(\boldsymbol{x}) = 0$, $j = 1, 2, \dots, m$, где $g_j : \mathbf{R}^n \to \mathbf{R}$, $j = 1, 2, \dots, m$. Иначе говоря, рассматриваемая задача заключается в максимизации (минимизации) целевой функции f на множестве $X \subset \mathbf{R}^n$, которое задается равенством

$$X = \left\{ \boldsymbol{x} \in \mathbf{R}^n \;\middle|\; g_j(\boldsymbol{x}) = 0,\ j = 1, 2, \dots, m \right\}. \tag{1}$$

Тот факт, что экстремум ищется не на всем пространстве $\mathbf{R}^n$, а лишь на некотором его подмножестве X, не позволяет прямо использовать необходимые и достаточные условия экстремума, полученные в предыдущем параграфе. Тем не менее, для сформулированной задачи на условный экстремум удается получить условия экстремума, сходные с условиями безусловного экстремума для некоторой специальной функции, в которой участвует как целевая функция f, так и функции g_1, g_2, … , g_m, задающие множество X. Эта функция носит название *функции Лагранжа* и определяется равенством

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

где числа λ_1, λ_2, … , λ_m образуют вектор $\boldsymbol{\lambda}$ и называются *множителями Лагранжа*. Нетрудно видеть, что функция Лагранжа зависит от $n+m$ переменных x_1, x_2, … , x_n, λ_1, λ_2, … , λ_m.

Теорема 1. Допустим, что функции f, g_1, g_2, … , g_m, непрерывны вместе со своими частными производными первого порядка в каждой точке некоторой окрестности точки $\boldsymbol{x}^* \in \mathbf{R}^n$, которая является точкой локального условного экстремума функции f на множестве X вида (1). Тогда существует такой набор множителей Лагранжа λ_0^*, λ_1^*, λ_2^*, … , λ_m^*, что

$$\lambda_0^* \cdot \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j^* \cdot \nabla g_j(\boldsymbol{x}^*) = \mathbf{0}, \qquad (2)$$

причем $\left(\lambda_0^*, \lambda_1^*, \lambda_2^*, \dots, \lambda_m^*\right) \neq \mathbf{0}$.

Если дополнительно предположить выполнение *условия регулярности*, состоящее в требовании линейной независимости векторов $\nabla g_1(\boldsymbol{x}^*)$, $\nabla g_2(\boldsymbol{x}^*)$, ... , $\nabla g_m(\boldsymbol{x}^*)$, то в равенстве (2) можно положить $\lambda_0^* = 1$.

***** ▶ Для определенности будем считать, что $\boldsymbol{x}^*$ – точка локального условного максимума.

Выберем положительное число ξ настолько малым, чтобы замкнутый шар $\overline{U} = \left\{ \boldsymbol{x} \in \mathbf{R}^n \,\middle|\, \left\| \boldsymbol{x} - \boldsymbol{x}^* \right\| \le \xi \right\}$ целиком содержался в той окрестности точки $\boldsymbol{x}^*$, о которой идет речь в формулировке теоремы. Ясно, что $\boldsymbol{x}^*$ является точкой глобального условного максимума функции f на множестве $X \cap \overline{U}$.

Введем функции

$$f_k(\boldsymbol{x}) = f(\boldsymbol{x}) - k \cdot \sum_{j=1}^{m} g_j^2(\boldsymbol{x}) - \left\| \boldsymbol{x} - \boldsymbol{x}^* \right\|^2, \quad k = 1, 2, \dots .$$

Каждая из этих функций непрерывна на непустом замкнутом множестве $\overline{U}$ и, значит, по теореме Вейерштрасса достигает на нем своего наибольшего значения. Обозначим через $\boldsymbol{x}^{(k)}$

точку максимума функции f_k на множестве $\overline{U}$:

$$f_k(\boldsymbol{x}^{(k)}) = \max_{x \in \overline{U}} f_k(\boldsymbol{x}), \quad k = 1, 2, \ldots \,. \tag{3}$$

Тем самым получена последовательность точек $\{\boldsymbol{x}^{(k)}\}$ в $\mathbf{R}^n$. Докажем сходимость этой последовательности к точке $\boldsymbol{x}^*$. Благодаря (3) имеем $f_k(\boldsymbol{x}^{(k)}) \ge f_k(\boldsymbol{x}^*)$, $k = 1, 2, \ldots$. Отсюда с учетом равенств $g_j(\boldsymbol{x}^*) = 0$, $j = 1, 2, \ldots, m$, следует неравенство

$$f(\boldsymbol{x}^{(k)}) - k\sum_{j=1}^{m} g_j^2(\boldsymbol{x}^{(k)}) - \left\| \boldsymbol{x}^{(k)} - \boldsymbol{x}^* \right\|^2 \ge f(\boldsymbol{x}^*)\,. \tag{4}$$

Поэтому

$$\sum_{j=1}^{m} g_j^2(\boldsymbol{x}^{(k)}) \le \frac{1}{k} \cdot \left[f(\boldsymbol{x}^{(k)}) - f(\boldsymbol{x}^*) - \left\| \boldsymbol{x}^{(k)} - \boldsymbol{x}^* \right\|^2 \right],$$

а значит,

$$0 \le \left| \sum_{j=1}^{m} g_j^2(\boldsymbol{x}^{(k)}) \right| \le \frac{1}{k} \cdot \left[\left| f(\boldsymbol{x}^{(k)}) \right| + \left| f(\boldsymbol{x}^*) \right| + \left\| \boldsymbol{x}^{(k)} - \boldsymbol{x}^* \right\|^2 \right] \tag{5}$$

для всех $k = 1, 2, \ldots$. Из непрерывности функции f на множестве $\overline{U}$ по теореме Вейерштрасса следует ограниченность функции f на $\overline{U}$. Отсюда вытекает ограниченность последовательности точек $\left\{ \left| f(\boldsymbol{x}^{(k)}) \right| \right\}$. Кроме того, согласно включению $\boldsymbol{x}^{(k)} \in \overline{U}$ выполнено $\left\| \boldsymbol{x}^{(k)} - \boldsymbol{x}^* \right\| \le \xi$, $k = 1, 2, \ldots$. На этом основании ограниченной является и последовательность точек, k-й член которой записан в квадратных скобках правой части неравенства (5). Поэтому, переходя в (5) к пределу при $k \to \infty$, с использованием теоремы о сжатой переменной получим равенство

$$\lim_{k \to \infty} \left| \sum_{j=1}^{m} g_j^2(\boldsymbol{x}^{(k)}) \right| = 0,$$

откуда следует

$$\lim_{k \to \infty} \sum_{j=1}^{m} g_j^2(\boldsymbol{x}^{(k)}) = 0\,. \tag{6}$$

По теореме Больцано – Вейерштрасса из ограниченной последовательности $\{\boldsymbol{x}^{(k)}\}$ можно выделить сходящуюся подпоследовательность. Обозначим произвольную такую подпоследовательность через $\{\boldsymbol{x}^{(k_l)}\}$, а предел этой подпоследовательности через $\hat{\boldsymbol{x}}$: $\lim\limits_{l \to \infty} \boldsymbol{x}^{(k_l)} = \hat{\boldsymbol{x}} \in \overline{U}$. Переходя в равенстве (6) при $k = k_l$ от предела суммы к сумме пределов, получим, что сумма неотрицательных слагаемых равна нулю. Отсюда следует, что каждое слагаемое этой суммы равно нулю, так что $\lim\limits_{l\to\infty} g_j\left(\boldsymbol{x}^{(k_l)}\right) = g_j(\hat{\boldsymbol{x}}) = 0$, $j = 1, 2, \dots, m$.

Из неравенства (4) следует

$$f\left(\boldsymbol{x}^{(k)}\right) - \left\| \boldsymbol{x}^{(k)} - \boldsymbol{x}^* \right\|^2 \ge f(\boldsymbol{x}^*), \quad k = 1, 2, \dots .$$

Отсюда при $k = k_l$ и $l \to \infty$ получаем

$$f(\hat{\boldsymbol{x}}) - \left\| \hat{\boldsymbol{x}} - \boldsymbol{x}^* \right\|^2 \ge f(\boldsymbol{x}^*) .$$

С другой стороны, так как $\boldsymbol{x}^*$ – точка максимума функции f на множестве $X \bigcap \overline{U}$ и $\hat{\boldsymbol{x}} \in X \bigcap \overline{U}$, то

$$f(\boldsymbol{x}^*) \ge f(\hat{\boldsymbol{x}}) .$$

Складывая почленно последние два неравенства, получим, что $\left\| \hat{\boldsymbol{x}} - \boldsymbol{x}^* \right\|^2 \le 0$. Отсюда вытекает равенство $\hat{\boldsymbol{x}} = \boldsymbol{x}^*$. Тем самым доказано, что любая сходящаяся подпоследовательность ограниченной последовательности $\{\boldsymbol{x}^{(k)}\}$ сходится к точке $\boldsymbol{x}^*$. В таком случае сама последовательность $\{\boldsymbol{x}^{(k)}\}$ сходится к точке $\boldsymbol{x}^*$ (в этом можно убедиться, рассуждая “от противного”)*) .

Обозначим через K номер, начиная с которого точки последовательности $\{\boldsymbol{x}^{(k)}\}$ “попадают” в окрестность $u_\varepsilon(\boldsymbol{x}^*)$: $\boldsymbol{x}^{(k)} \in u_\varepsilon(\boldsymbol{x}^*)$, $k = K, K+1, \dots$. Множество $u_\varepsilon(\boldsymbol{x}^*)$ является открытым, значит, каждая точка $\boldsymbol{x}^{(k)}$, $k = K, K+1, \dots$, входит в него с которой своей окрестностью. Являясь точкой максимума функции f_k на $\overline{U}$, $\boldsymbol{x}^{(k)}$ будет также точкой максимума на указанной своей окрестности. Следовательно, каждая из точек $\boldsymbol{x}^{(k)}$, $k = K, K+1, \dots$, представляет собой точку локального безусловного максимума функции f_k, и для них в соответствии с теоремой 1 предыдущего параграфа выполняется равенство $\nabla f_k\left(\boldsymbol{x}^{(k)}\right) = \boldsymbol{0}$, или, в развернутой форме,

---

*) Наметим ход этих рассуждений. Исходная последовательность $\{\boldsymbol{x}^{(k)}\}$ ограничена, поэтому из нее можно выделить сходящуюся подпоследовательность, которую обозначим через $\{\boldsymbol{x}^{(l)}\}$: $\lim\limits_{l\to\infty} \boldsymbol{x}^{(l)} = \boldsymbol{x}^*$. Убедимся в том, что последовательность $\{\boldsymbol{x}^{(k)}\}$ сходится к той же самой точке $\boldsymbol{x}^*$. Для этого предположим противное: $\boldsymbol{x}^*$ не является пределом последовательности $\{\boldsymbol{x}^{(k)}\}$. Отсюда следует, что найдется некоторая окрестность точки $\boldsymbol{x}^*$, за пределами которой окажется *бесконечное* число членов последовательности $\{\boldsymbol{x}^{(k)}\}$, образующих некоторую подпоследовательность исходной последовательности $\{\boldsymbol{x}^{(k)}\}$. Из полученной подпоследовательности снова можно выделить сходящуюся подпоследовательность. Обозначим ее через $\{\boldsymbol{x}^{(m)}\}$: $\lim\limits_{m\to\infty} \boldsymbol{x}^{(m)} = \hat{\boldsymbol{x}}$. Так как все члены подпоследовательности $\{\boldsymbol{x}^{(m)}\}$ лежат вне некоторой указанной выше окрестности точки $\boldsymbol{x}^*$, то $\hat{\boldsymbol{x}} \ne \boldsymbol{x}^*$. Это противоречит тому, что всякая сходящаяся подпоследовательность последовательности $\boldsymbol{x}^*$ имеет один и тот же предел.

$$\nabla f\left(\boldsymbol{x}^{(k)}\right)-2k\cdot\sum_{j=1}^{m}g_j\left(\boldsymbol{x}^{(k)}\right)\cdot\nabla g_j\left(\boldsymbol{x}^{(k)}\right)-2\left(\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\right)=\mathbf{0}, \qquad (7)$$

так как

$$\frac{\partial g_j^2\left(\boldsymbol{x}^{(k)}\right)}{\partial x_i}=2g_j\left(\boldsymbol{x}^{(k)}\right)\cdot\frac{\partial g_j\left(\boldsymbol{x}^{(k)}\right)}{\partial x_i},$$

$$\frac{\partial}{\partial x_i}\sum_{i=1}^{n}(x_i^{(k)}-x_i^*)^2=2\,(x_i^{(k)}-x_i^*),\quad i=1,2,\ldots,n.$$

Разделив обе части равенства (7) на положительное число $M_k=\sqrt{1+4k^2\sum_{j=1}^{m}g_j^2\left(\boldsymbol{x}^{(k)}\right)}$, получим

$$\mu_k\nabla f\left(\boldsymbol{x}^{(k)}\right)+\sum_{j=1}^{m}\mu_{kj}\nabla g_j\left(\boldsymbol{x}^{(k)}\right)=2\cdot\mu_k\left(\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\right), \qquad (8)$$

где $\mu_k=\dfrac{1}{M_k}$, $\mu_{kj}=-\dfrac{2kg_j\left(\boldsymbol{x}^{(k)}\right)}{M_k}$, $k=1,2,\ldots$. Непосредственной проверкой устанавливается равенство $\mu_k^2+\sum_{j=1}^{m}\mu_{kj}^2=1$, которое свидетельствует об ограниченности последовательностей $\{\mu_k\}$ и $\{\mu_{kj}\}$, $j=1,2,\ldots,m$.

Выделим из этих последовательностей сходящиеся подпоследовательности:

$$\lim_{l\to\infty}\mu_{k_l}=\lambda_0^*,\quad \lim_{l\to\infty}\mu_{k_l j}=\lambda_j^*,\quad j=1,2,\ldots,m,$$

причем $(\lambda_0^*)^2+\sum_{j=1}^{m}(\lambda_j^*)^2=1$, т.е. $\left(\lambda_0^*,\lambda_1^*,\ldots,\lambda_m^*\right)\neq\mathbf{0}$. Полагая в равенстве (8) $k=k_l$ и переходя в нем к пределу при $l\to\infty$, получим требуемое равенство (2).

Остается проверить последнюю часть теоремы 1. В случае выполнения условия регулярности верно $\lambda_0^*\neq 0$. Действительно, если бы выполнялось равенство $\lambda_0^*=0$, то (2) превратилось бы в равенство $\sum_{j=1}^{m}\lambda_j^*\nabla g_j(\boldsymbol{x}^*)=\mathbf{0}$, которое в силу $\left(\lambda_1^*,\lambda_2^*,\ldots,\lambda_m^*\right)\neq\mathbf{0}$ приводило к линейной зависимости векторов $\nabla g_1(\boldsymbol{x}^*)$, $\nabla g_2(\boldsymbol{x}^*)$, … , $\nabla g_m(\boldsymbol{x}^*)$, противоречащей условию регулярности. Поэтому $\lambda_0^*\neq 0$ и, разделив на λ_0^* обе части равенства (2), придем к равенству, в котором коэффициент при $\nabla f(\boldsymbol{x}^*)$ равен 1. ◀

Замечание. При доказательстве теоремы 1 были использованы вспомогательные функции f_k. Их называют *штрафными функциями*. Как оказалось, точки локального безусловного максимума этих функций образуют последовательность, сходящуюся к точке

$\boldsymbol{x}^*$ условного максимума функции f. Таким образом, исходной задаче условной оптимизации можно поставить в соответствие последовательность вспомогательных задач безусловной оптимизации. Решая эти вспомогательные задачи, получим последовательность точек, сходящихся к решению исходной задачи на условный экстремум. Описанная идея находит свое применение при построении численных методов условной оптимизации, называемых *методами штрафных функций*.

В предположении выполнения условия регулярности выпишем необходимое условие экстремума теоремы 1

$$\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j^* \cdot \nabla g_j(\boldsymbol{x}^*) = \boldsymbol{0} \tag{9}$$

и условие принадлежности точки $\boldsymbol{x}^*$ множеству X:

$$g_j(\boldsymbol{x}^*) = 0\,, \quad j = 1, 2, \dots, m\,. \tag{10}$$

Оба этих условия можно объединить в формуле

$$\nabla L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{0}\,, \tag{11}$$

где градиент функции Лагранжа состоит из частных производных по всем ее $m+n$ переменным и вычисляется следующим образом:

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_j}\,, \quad i = 1, 2, \dots, n\,,$$

$$\frac{\partial L}{\partial \lambda_j} = g_j\,, \quad j = 1, 2, \dots, m\,.$$

В соответствии с этим теорему 1 можно сформулировать так: *если* $\boldsymbol{x}^*$ *– точка локального условного экстремума функции* f *на множестве* X (1), *то при определенных предположениях найдутся такие множители Лагранжа* $\lambda_1^*,\ \lambda_2^*,\ \dots,\ \lambda_m^*$, *образующие вектор* $\boldsymbol{\lambda}^*$, *что* $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ *является стационарной точкой функции Лагранжа* $L(\boldsymbol{x}, \boldsymbol{\lambda})$, *т.е. выполнено равенство* (11).

Пример. Рассмотрим задачу оптимизации функции трех переменных

$$f(x_1, x_2, x_3) = x_1 x_2 + x_2 x_3$$

при ограничениях

$$-2x_1 + x_2^2 = 1,$$

$$x_2 + x_3 = 1.$$

Составляем функцию Лагранжа

$$L(x_1, x_2, x_3, \lambda_1, \lambda_2) = x_1 x_2 + x_2 x_3 + \lambda_1(-2x_1 + x_2^2 - 1) + \lambda_2(x_2 + x_3 - 1)$$

и приравниваем ее частные производные нулю:

$$
\begin{cases}
\dfrac{\partial L}{\partial x_1} = x_2 - 2\lambda_1 = 0, \\
\dfrac{\partial L}{\partial x_2} = x_1 + x_3 + 2\lambda_1 x_2 + \lambda_2 = 0, \\
\dfrac{\partial L}{\partial x_3} = x_2 + \lambda_2 = 0, \\
\dfrac{\partial L}{\partial \lambda_1} = -2x_1 + x_2^2 - 1 = 0, \\
\dfrac{\partial L}{\partial \lambda_2} = x_2 + x_3 - 1 = 0.
\end{cases}
$$

Полученная система уравнений имеет два решения

$$
\boldsymbol{x}^{(1)} = (0, 1, 0), \quad \boldsymbol{\lambda}^{(1)} = \left(\frac{1}{2}, -1\right), \quad \boldsymbol{x}^{(2)} = \left(-\frac{4}{9}, \frac{1}{3}, \frac{2}{3}\right), \quad \boldsymbol{\lambda}^{(2)} = \left(\frac{1}{6}, -\frac{1}{3}\right).
$$

Таким образом, найденные две точки $\boldsymbol{x}^{(1)}$ и $\boldsymbol{x}^{(2)}$ являются “подозрительными” на экстремальные. Для решения вопроса о том, будут ли эти точки действительно экстремальными, следует использовать достаточные условия условного экстремума, излагаемые в п.3°.

2°. Геометрическая интерпретация необходимого условия. Рассмотрим самый простой случай, когда целевая функция f и единственная функция g, входящая в ограничение, имеют лишь две переменные. В этом случае равенство (9) принимает вид

$$
\nabla f(\boldsymbol{x}^*) + \lambda \cdot \nabla g(\boldsymbol{x}^*) = \boldsymbol{0},
$$

или

$$
\nabla f(\boldsymbol{x}^*) = -\lambda \cdot \nabla g(\boldsymbol{x}^*). \tag{12}
$$

Изобразим на плоскости (см. рис.3.8) точку экстремума $\boldsymbol{x}^*$, две линии уровня функций f и g, проходящие через эту точку, и векторы-градиенты $\nabla f(\boldsymbol{x}^*)$ и $\nabla g(\boldsymbol{x}^*)$, приложенные к точке $\boldsymbol{x}^*$.

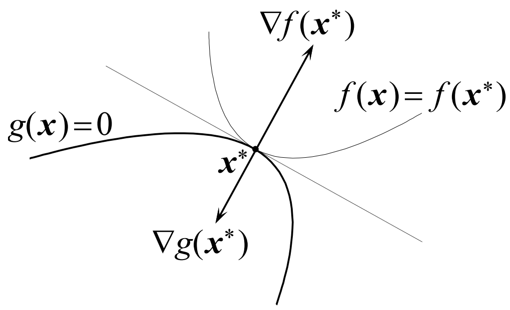

Рис. 3.8. Геометрическая интерпретация необходимого условия экстремума

Линия уровня функции g, проходящая через точку $\boldsymbol{x}^*$, задается уравнением $g(\boldsymbol{x}) = 0$, так как точка $\boldsymbol{x}^*$ удовлетворяет равенству $g(\boldsymbol{x}^*) = 0$. Таким образом, задача на экстремум функции f при ограничении $g(\boldsymbol{x}) = 0$ геометрически состоит в отыскании среди всех точек кривой, задаваемой уравнением $g(\boldsymbol{x}) = 0$, таких точек, в которых целевая функция принимает максимальное или минимальное значение.

Необходимое условие (12) геометрически означает коллинеарность векторов $\nabla f(\boldsymbol{x}^*)$ и $\nabla g(\boldsymbol{x}^*)$. Таким образом, *если точка* $\boldsymbol{x}^*$ *является точкой (локального) условного экстремума, то градиенты целевой функции* $\nabla f(\boldsymbol{x}^*)$ *и функции-ограничения* $\nabla g(\boldsymbol{x}^*)$, *вычисленные в этой точке экстремума, должны быть коллинеарными*. На рис.3.8 указанные два вектора имеют противоположное направление, хотя коллинеарность допускает и их сонаправленность.

Как известно (см. п.6 §1 гл.2), градиент функции перпендикулярен линии уровня данной функции, проходящей через точку, в которой вычисляется этот градиент. Точнее говоря, он перпендикулярен касательной к линии уровня, проходящей через указанную точку. С учетом этого можно дать еще одну геометрическую интерпретацию необходимого условия (12): *линии уровня целевой функции и функции, задающей ограничение в форме равенства, которые проходят через точку (локального) экстремума, имеют одну и ту же касательную в этой точке экстремума*.

*** 3°. Условия второго порядка.** Для функции Лагранжа

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

матрицу Гессе (по переменным $x_1,\ x_2,\ \dots,\ x_n$) будем обозначать через $H_L(\boldsymbol{x},\boldsymbol{\lambda})$, т.е.

$$H_L(\boldsymbol{x},\boldsymbol{\lambda}) = \begin{pmatrix} \dfrac{\partial^2 L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_1^2} & \dfrac{\partial^2 L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_1 \partial x_2} & \dots & \dfrac{\partial^2 L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_2 \partial x_1} & \dfrac{\partial^2 L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_2^2} & \dots & \dfrac{\partial^2 L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_2 \partial x_n} \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot \\ \dfrac{\partial^2 L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_n \partial x_1} & \dfrac{\partial^2 L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_n \partial x_2} & \dots & \dfrac{\partial^2 L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_n^2} \end{pmatrix}.$$

В терминах этой матрицы в следующей теореме, доказательство которой можно найти в [ ], формулируются необходимое, а также достаточное условия экстремума для задачи оптимизации функции f на множестве X вида (1).

Теорема 2. Предположим, что все функции $f,\ g_1,\ g_2,\ \dots,\ g_m$ имеют непрерывные частные производные второго порядка в каждой точке некоторой окрестности точки $\boldsymbol{x}^* \in X$, причем векторы $\nabla g_j(\boldsymbol{x}^*)$, $j = 1, 2, \dots, m$, образуют линейно независимую систему.

Необходимое условие экстремума. Если $\boldsymbol{x}^*$ является точкой локального условного максимума (минимума) функции f на множестве X (1), то существуют множители Лагранжа $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*)$ такие, что имеет место равенство (9) и, кроме того, неравенство

$$\boldsymbol{y}\cdot H_L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\cdot\boldsymbol{y}^T\le(\ge)\,0$$

выполняется для всех векторов-строк $\boldsymbol{y}\in Y$, где $Y=\left\{\boldsymbol{y}\in\mathbf{R}^n\,\middle|\,\langle\nabla g_j(\boldsymbol{x}^*),\,\boldsymbol{y}\rangle=0,\ \ j=1,2,\dots,m\right\}$.

Достаточное условие экстремума. Если для точки $\boldsymbol{x}^*\in X$ и некоторого вектора $\boldsymbol{\lambda}^*=(\lambda_1^*,\lambda_2^*,\dots,\lambda_m^*)$ справедливо равенство (9) и, кроме того, неравенство

$$\boldsymbol{y}\cdot H_L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\cdot\boldsymbol{y}^T<(>)\,0$$

выполняется для всех $\boldsymbol{y}\in Y$, $\boldsymbol{y}\ne\boldsymbol{0}$, то $\boldsymbol{x}^*$ – точка локального условного максимума (миниму-
ма) функции f на множестве X.

Пример. Продолжим начатое в п.1° рассмотрение задачи оптимизации функции

$$f(x_1,x_2,x_3)=x_1x_2+x_2x_3$$

при ограничениях

$$-2x_1+x_2^2=1,$$
$$x_2+x_3=1.$$

Вычислим матрицу Гессе H_L для функции Лагранжа:

$$H_L(x_1,x_2,x_3,\lambda_1,\lambda_2)=\begin{pmatrix}0&1&0\\1&2\lambda_1&1\\0&1&0\end{pmatrix}.$$

Напомним, что раньше были найдены две “подозрительные” на экстремальные точки

$$\boldsymbol{x}^{(1)}=(0,1,0)\quad\text{и}\quad\boldsymbol{x}^{(2)}=\left(-\frac{4}{9},\frac{1}{3},\frac{2}{3}\right)$$

вместе с соответствующими векторами

$$\boldsymbol{\lambda}^{(1)}=\left(\frac{1}{2},-1\right)\quad\text{и}\quad\boldsymbol{\lambda}^{(2)}=\left(\frac{1}{6},-\frac{1}{3}\right).$$

Для первой точки имеем
$$H_L\left(\boldsymbol{x}^{(1)},\boldsymbol{\lambda}^{(1)}\right)=\begin{pmatrix}0&1&0\\1&1&1\\0&1&0\end{pmatrix},$$

$$\nabla g_1\left(\boldsymbol{x}^{(1)}\right)=\left(-2,2,0\right),$$

$$\nabla g_2\left(\boldsymbol{x}^{(1)}\right)=\left(0,1,1\right).$$

Множество Y, которое участвует в формулировке теоремы 2, представляет собой множество решений системы из двух линейных уравнений

$$\begin{cases} -2y_1 + 2y_2 = 0, \\ \quad y_2 + y_3 = 0. \end{cases}$$

С использованием этих равенств можно записать

$$\boldsymbol{y} \cdot H_L\left(\boldsymbol{x}^{(1)}, \boldsymbol{\lambda}^{(1)}\right) \cdot \boldsymbol{y}^T = 2y_1 y_2 + y_2^2 + 2y_2 y_3 = y_2^2 > 0$$

для всех $y_2 \neq 0$. Следовательно, точка $\boldsymbol{x}^{(1)}$ является точкой локального условного минимума.

Аналогично для второй точки $\boldsymbol{x}^{(2)}$ получаем

$$H_L\left(\boldsymbol{x}^{(2)}, \boldsymbol{\lambda}^{(2)}\right) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1/3 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\nabla g_1\left(\boldsymbol{x}^{(2)}\right) = \left(-2, 2/3, 0\right),$$

$$\nabla g_2\left(\boldsymbol{x}^{(2)}\right) = \left(0, 1, 1\right).$$

В этом случае множество Y задается следующей системой линейных уравнений

$$\begin{cases} -2y_1 + \dfrac{2}{3} y_2 = 0, \\ \quad y_2 + y_3 = 0. \end{cases}$$

Поэтому

$$\boldsymbol{y} \cdot H_L\left(\boldsymbol{x}^{(2)}, \boldsymbol{\lambda}^{(2)}\right) \cdot \boldsymbol{y}^T = 2y_1 y_2 + \frac{1}{3} y_2^2 + 2y_2 y_3 = -y_2^2 < 0$$

для всех $y_2 \neq 0$. Согласно достаточному условию теоремы 2 это означает, что $\boldsymbol{x}^{(2)}$ – точка локального условного максимума.

Замечание. Разобранный пример задачи на условный экстремум иллюстрирует применение результатов теорем 1 и 2. Следует отметить, что изложенный способ решения данной задачи не является рациональным. Значительно более простое решение состоит в следующем. Используя ограничения-равенства, переменные x_1 и x_3 можно выразить через x_2:

$$x_1 = \frac{x_2^2 - 1}{2}, \quad x_3 = 1 - x_2$$

Подставив найденные выражения в формулу, задающую целевую функцию f, получим

$$f = x_1 x_2 + x_2 x_3 = \frac{x_2^3}{2} - x_2^2 + \frac{x_2}{2}$$

в виде функции одной переменной x_2. Таким образом, исходная задача на условный экстремум сведена к задаче на безусловный экстремум функции одной переменной.

Приравнивая производную этой функции нулю, получим квадратное уравнение

$$f' = \frac{3}{2}x_2^2 - 2x_2 + \frac{1}{2} = 0,$$

имеющее два корня $x_2' = 1$, $x_2'' = \frac{1}{3}$. Поскольку $f''(1) = 1 > 0$, $f''(1/3) = -1 < 0$, то в первой точке достигается локальный минимум, а во второй – локальный максимум.

Резюме к главе 3

Прежде всего были введены понятия выпуклого множества в $\mathbf{R}^n$, а также вогнутой и выпуклой функций нескольких переменных, которые играют важную роль в теории экстремальных задач. Введенные понятия проиллюстрированы соответствующими примерами.

После постановки прикладной задачи выбора параметров функции сформулирована в общем виде экстремальная задача и перечислены возможные типы ее решений. Разобраны различные классы экстремальных задач.

Далее для задач без ограничений, а также для задач с ограничениями в форме равенств были получены различного рода необходимые и/или достаточные условия экстремума.

Задание на самостоятельную работу

1. Сформулируйте определение выпуклого множества в $\mathbf{R}^n$.
2. Сформулируйте определение вогнутой и выпуклой функций.
3. Приведите примеры вогнутых и выпуклых функций n переменных.
4. Установите выпуклость функции

$$f(x_1, x_2, x_3) = 2x_1^2 + 3x_2^2 + 4x_3^2 + 2x_1x_2 + x_1x_3 + x_2x_3 - 2x_1 - 3x_2 - x_3.$$

5. Приведите постановку задачи безусловной (условной) максимизации (минимизации) целевой функции f.
6. Дайте определение точек глобального (локального) безусловного (условного) максимума (минимума).
7. Сформулируйте задачу линейного программирования.
8. Напишите необходимые (достаточные) условия безусловного экстремума первого и второго порядков.
9. Сформулируйте необходимое условие условного экстремума.

Решите задачи:

16.101 – 16.106 [ ]

7.187 – 7.198, 7.201 – 7.209, 7.211 – 7.227 [ ]

Литература

1. Ильин В.А., Позняк В.Г. Основы математического анализа, ч.1. – М.: Наука, 1971.
2. Бугров Я.С., Никольский С.М. Дифференциальное и интегральное исчисление. – М.: Наука, 1988.
3. Сборник задач по математике. Т.1. Линейная алгебра и основы математического анализа. Под ред. Ефимова А.В., Демидовича Б.П. – 2-е изд. – М.: Наука, 1986.