С. Р. Насыров

ДИФФЕРЕНЦИРОВАНИЕ ФУНКЦИЙ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

Казань – 2018

УДК 515.1

Рекомендовано к опубликованию и размещению
на сайте Казанского (Приволжского) федерального университета
Учебно-методической комиссии Института математики и механики
им. Н. И. Лобачевского, Протокол № 4 от 15 февраля 2018 г.

Рецензент
кандидат физ.-мат. наук, доцент Р. Н. Гумеров

Насыров С. Р. Дифференцирование функций нескольких переменных: Учебное пособие / С. Р. Насыров. – Казань: Казанский (Приволжский) федеральный университет, 2018. – 80 с.

Данное учебно-методическое пособие предназначено для студентов, обучающихся в бакалавриате по направлениям «Математика», «Математика и компьютерные науки». В нем излагаются основы дифференциального исчисления функций нескольких вещественных переменных. Материал соответствует программе курса «Математический анализ» для студентов-математиков, действующей в Казанском (Приволжском) федеральном университете.

1 Пределы и непрерывность отображений в пространстве $\mathbb{R}^n$ (сводка результатов)

1.1 Топология и метрика в $\mathbb{R}^n$

Обозначим через $\mathbb{R}^n$ вещественное евклидово n-мерное пространство, т. е. линейное векторное пространство над полем $\mathbb{R}$, элементами которого являются упорядоченные наборы x из n вещественных чисел:

$$x = (x^1, x^2, \ldots, x^n).$$

(Здесь верхний индекс означает не показатель степени, а номер соответствующей координаты. Мы будем пользоваться такими обозначениями в случаях, когда есть возможность путаницы с последовательностями (x_n), в других случаях индексы будем писать внизу.) Элементы x также называются *точками* или *векторами* пространства $\mathbb{R}^n$. Напомним, как вводятся операции на множестве $\mathbb{R}^n$, превращающие его в линейное векторное пространство.

1) *Сложение векторов.* Если

$$x = (x^1, x^2, \ldots, x^n) \quad \text{и} \quad y = (y^1, y^2, \ldots, y^n)$$

два произвольных вектора из $\mathbb{R}^n$, то их суммой называется вектор

$$x + y := (x^1 + y^1, x^2 + y^2, \ldots, x^n + y^n).$$

Относительно сложения $\mathbb{R}^n$ является абелевой группой. Единицей в этой группе является нулевой вектор $(0, 0, \dots, 0)$, который в отличие от скалярного нуля будем обозначать θ.

2) *Умножение на скаляры.* Если

$$x = (x^1, x^2, \dots, x^n), \quad \alpha \in \mathbb{R},$$

то их *произведением* называется вектор

$$\alpha x := (\alpha x^1, \alpha x^2, \dots, \alpha x^n).$$

Умножение вектора x на скаляр α удовлетворяет естественным условиям ассоциативности и дистрибутивности.

В $\mathbb{R}^n$ можно ввести *скалярное произведение векторов* по правилу

$$\langle x, y \rangle := \sum_{k=1}^{n} x^k y^k,$$

и *норму*

$$\|x\| := \sqrt{\sum\nolimits_{k=1}^{n} (x^k)^2}.$$

Очевидно, что $\|x\| = \sqrt{\langle x, x \rangle}$.

Для любых векторов x, $y \in \mathbb{R}^n$ имеет место *неравенство Коши-Буняковского*

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|, \quad x, y \in \mathbb{R}^n,$$

причем равенство достигается тогда и только тогда, когда вектора x и y линейно зависимы.

Расстояние между точками в $\mathbb{R}^n$ (метрика) вводится по формуле

$$d(x, y) := \|x - y\| = \sqrt{\sum\nolimits_{k=1}^{n} (x^k - y^k)^2}.$$

Среди свойств метрики отметим особо неравенство треугольника

$$\|x - z\| \leq \|x - y\| + \|y - z\|.$$

Окрестностью точки $x \in \mathbb{R}^n$ (или r-окрестностью x) называется множество

$$B_r(x) := \{y \in \mathbb{R}^n \mid \|y - x\| < r\} \quad (r > 0).$$

Ясно, что $B_r(x)$ — это (открытый) шар радиуса r с центром в точке x.

Проколотая окрестность (или r-окрестностью) точки x – это множество $\overset{\vee}{B}_r(x) := B_r(x) \setminus \{x\}$.

Любые две различные точки в $\mathbb{R}^n$ обладают непересекающимися окрестностями.

В $\mathbb{R}^n$ можно ввести и другую норму

$$\|x\|_1 = \max_{1 \le k \le n} |x^k|$$

и расстояние

$$d_1(x, y) := \|x - y\|_1 = \max_{1 \le k \le n} |x^k - y^k|.$$

Справедлива оценка:

$$\|x\|_1 \le \|x\| \le \sqrt{n}\, \|x\|_1, \quad x \in \mathbb{R}^n.$$

1.2 Открытые и замкнутые множества в $\mathbb{R}^n$

Точка $a \in \mathbb{R}^n$ называется *граничной точкой* множества $A \subset \mathbb{R}^n$, если для любого $\varepsilon > 0$ выполняются условия: $B_\varepsilon(a) \cap A \neq \emptyset$ и $B_\varepsilon(a) \cap A^c \neq \emptyset$. Точка $a \in \mathbb{R}^n$ называется внутренней точкой множества A, если $\exists\, \varepsilon > 0$: $B_\varepsilon(a) \subset A$. Точка $a \in \mathbb{R}^n$ называется *внешней точкой* множества A, если $\exists\, \varepsilon > 0$: $B_\varepsilon(a) \subset A^c$.

Множество граничных точек множества A или его *граница* обозначается A^Γ или ∂A, множество внутренних точек или *внутренность* — A°. Из определений следует, что $\mathbb{R}^n = A^\circ \cup A^\Gamma \cup (A^c)^\circ$.

Точка $a \in \mathbb{R}^n$ называется *точкой прикосновения* множества A, если $\forall \varepsilon > 0$ $(B_\varepsilon(a) \cap A) \neq \emptyset$. Множество точек прикосновения множества A называется *замыканием* A и обозначается $\overline{A}$ или A^-. Ясно, что справедливо равенство $\overline{A} = A \cup A^\Gamma$.

Множество A называется *открытым*, если любая точка A является внутренней точкой этого множества.

Множество A открыто $\iff$ $A = A^\circ$ $\iff$ $A \cap A^\Gamma = \emptyset$.

Множество A называется *замкнутым*, если оно содержит все свои точки прикосновения. Множество A замкнуто $\iff$ $A = \overline{A}$ $\iff$ $A^\Gamma \subset A$.

Множество A замкнуто тогда и только тогда, когда его дополнение A^c открыто.

Примеры. 1) Имеем $(\mathbb{R}^n)^\Gamma = \emptyset$, поэтому $(\mathbb{R}^n)^\circ = (\mathbb{R}^n)^- = \mathbb{R}^n$. Множество $\mathbb{R}^n$ открыто и замкнуто одновременно.

2) Имеем $(\emptyset)^\Gamma = \emptyset$, поэтому $(\emptyset)^\Gamma = (\emptyset)^- = \emptyset$. Множество $\emptyset$ открыто и замкнуто одновременно.

Упражнение. Докажите, что других множеств в $\mathbb{R}^n$, кроме пустого множества и самого $\mathbb{R}^n$, которые были бы открытыми и замкнутыми одновременно, нет.

1.3 Свойства открытых и замкнутых множеств в $\mathbb{R}^n$

Сначала отметим свойства открытых множеств в $\mathbb{R}^n$.

1) Множества $\mathbb{R}^n$ и $\emptyset$ открыты.

2) Объединение любого семейства открытых множеств открыто.

3) Пересечение конечного числа открытых множеств в открыто.

Аналогичными свойствами обладают замкнутые множества.

1) Множества $\mathbb{R}^n$ и $\emptyset$ замкнуты.

2) Пересечение любого семейства замкнутых множеств замкнуто.

3) Объединение конечного числа замкнутых множеств замкнуто.

Примеры открытых и замкнутых множеств.

1) Шар $B_r(a) = \{x \in \mathbb{R}^n \mid \|x - a\| < r\}$ является открытым множеством. Его граница $(B_r(a))^\Gamma = \{x \in \mathbb{R}^n \mid \|x - a\| = r\}$ является замкнутым множеством. Замыкание $(B_r(a))^- = \{x \in \mathbb{R}^n \mid \|x - a\| \leq r\}$ является замкнутым множеством и называется замкнутым шаром радиуса r с центром в точке a.

2) Рассмотрим множество $A = B_r(a) \cup C$, где

$$B_r(a) = \{(x^1, x^2) \in \mathbb{R}^2 \mid (x^1 - a^1)^2 + (x^2 - a_2)^2 < r^2\},$$

$$C = \{(x^1, x^2) \in \mathbb{R}^2 \mid (x^1 - a^1)^2 + (x^2 - a_2)^2 = r^2,\ x^1 \geq a^1\}.$$

Множество A не является ни открытым, ни замкнутым в $\mathbb{R}^2$.

1.4 Предельные и изолированные точки

Точка a называется *предельной точкой* множества $A \subset \mathbb{R}^n$, если для любого $\varepsilon > 0$ имеем $\overset{\vee}{B}_\varepsilon(a) \cap A \neq \emptyset$. Точка a называется *изолированной точкой* множества A, если существует $\varepsilon > 0$ такое, что $B_\varepsilon(a) \cap A = \{a\}$. Ясно что замыкание $\overline{A}$ множества A делится на два непересекающихся множества: множество предельных и множество изолированных точек A.

Пример. Рассмотрим на плоскости множество

$$A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\} \cup \left\{ \left(1 + \frac{1}{n}, 0\right) \mid n \in \mathbb{N} \right\}.$$

Любая точка вида $\left(1 + \frac{1}{n}, 0\right)$ является изолированной точкой A, множество предельных точек A состоит из замкнутого круга $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$.

1.5 Последовательности в $\mathbb{R}^n$ и их пределы

Последовательностью (x_m) в пространстве $\mathbb{R}^n$ называется отображение $x_{(\,\cdot\,)} : \mathbb{N} \to \mathbb{R}^n$, которое каждому натуральному числу m сопоставляет вектор $x_m \in \mathbb{R}^n$.

Точка a называется пределом последовательности (x_m), если

$$\forall \varepsilon > 0 \quad \exists N \in \mathbb{N} : \forall m \geq N \quad (\|x_m - a\| < \varepsilon).$$

Отметим свойства пределов последовательностей.

1) Единственность предела. Если $(x_m) \to a$ и $(x_m) \to b$, то $a = b$.

2) Если $(x_m) \to a$, то и любая подпоследовательность $(x_{m_k}) \to a$. Наоборот, если из любой подпоследовательности (x_{m_k}) можно выделить подпоследовательность $(x_{m_{k_j}})$, сходящуюся к a, то и вся последовательность $(x_m) \to a$.

3) Любая сходящаяся последовательность ограничена, т. е. если $(x_m) \to a$, то существует константа $M > 0$ такая, что $\|x_m\| \leq M$, $m \geq 1$.

Теорема 1 (характеризация точек прикосновения и предельных точек через последовательности). 1) *Точка a является точкой прикосновения множества $A \subset \mathbb{R}^n$ тогда и только тогда, когда существует последовательность (x_m) точек множества A, сходящаяся к a.*

2) *Точка a является предельной точкой множества $A \subset \mathbb{R}^n$ тогда и только тогда, когда существует последовательность (x_m) точек множества $A \setminus \{a\}$, сходящаяся к a.*

Теорема 2. *Пусть $x_m = (x_m^1, x_m^2, \dots, x_m^n)$ — последовательность в $\mathbb{R}^n$. Последовательность точек x_m сходится к $a = (a^1, a^2, \dots, a^n)$ тогда и только тогда, когда для любого k, $1 \le k \le n$, имеет место сходимость $x_m^k \to a^k$.*

Последовательность x_m называется фундаментальной, если

$$\forall \varepsilon > 0 \quad \exists N \in \mathbb{N} : \forall\, l, m \ge N \quad (\|x_l - x_m\| < \varepsilon).$$

Теорема 3 (критерий Коши). *Последовательность x_m сходится тогда и только тогда, когда она фундаментальна.*

Теорема Больцано-Вейерштрасса для $\mathbb{R}^n$. *Из любой ограниченной последовательности в $\mathbb{R}^n$ можно выделить сходящуюся подпоследовательность.*

Доказательство. Пусть $x_m = (x_m^1, x_m^2, \dots, x_m^n)$ ограничена. Тогда существует константа $M > 0$ такая, что $\|x_m\| \le M$, $m \ge 1$. В силу свойств норм справедливо неравенство $\|x_m\|_1 \le \|x_m\|$, откуда получаем $\|x_m\|_1 \le M$, т. е. $\max |x_m^k| \le M$, $m \ge 1$. Значит, для любого k числовая последовательность x_m^k ограничена. Так как последовательность x_m^1 ограничена, то по теореме Больцано-Вейерштрасса для $\mathbb{R}$ существует ее подпоследовательность $x_{m_1}^1$, сходящаяся к некоторому числу a^1. Последовательность $x_{m_1}^2$ ограничена, поэтому по теореме Больцано-Вейерштрасса для $\mathbb{R}$ существует ее подпоследовательность $x_{m_2}^2$, сходящаяся к некоторому числу a^2. Продолжая этот процесс, в результате построим такую подпоследовательность x_{m_n}, для которой $x_{m_n}^k$ сходится к некоторому числу a^k для всех k, $1 \le k \le n$. По теореме 2 $x_{m_n} \to a = (a^1, a^2, \dots, a^n)$.

1.6 Компактные множества в $\mathbb{R}^n$

Семейство $(U_i)_{i\in I}$ подмножеств $\mathbb{R}^n$ называется *покрытием* множества $A\subset\mathbb{R}^n$, если $A\subset\cup_{i\in I}U_i$. Если к тому же все U_i — открытые множества, то *покрытие* $(U_i)_{i\in I}$ называется *открытым*. Если I — конечное множество, то *покрытие* $(U_i)_{i\in I}$ называется *конечным*.

Множество A называется *компактным*, если из любого его открытого покрытия можно выделить конечное подпокрытие.

Множество A называется *секвенциально компактным*, если из любой последовательности x_m элементов множества A можно выделить подпоследовательность x_{m_k}, сходящуюся к некоторому элементу множества A.

Множество называется замкнутым, если оно содержится в некотором шаре. В силу неравенства треугольника можем считать, что центр шара располагается в точке θ.

Теорема. *Пусть $A\subset\mathbb{R}^n$. Следующие условия эквивалентны:*
(1) *множество A компактно;*
(2) *множество A секвенциально компактно;*
(3) *множество A ограничено и замкнуто.*

Диаметром множества A называется число

$$d(A):=\sup_{x,y\in A}\|x-y\|.$$

Нетрудно видеть, что множество A ограничено тогда и только тогда, когда $d(A)<+\infty$.

Примеры. 1) Диаметр шара $B_r(a)$ равен $d(B_r(a))=2r$.

2) Пусть $Q=[0,1]^n$ — единичный куб в пространстве $\mathbb{R}^n$. Его диаметр $d(Q)=\sqrt{n}$.

Теорема. *Пусть $A_1\supset A_2\supset A_3\supset\ldots$ — убывающая последовательность непустых компактных множеств в $\mathbb{R}^n$, причем $d(A_k)\to 0$, $k\to\infty$. Тогда существует одна и только одна точка x_0, принадлежащая всем A_k, т. е. $\cap_{k=1}^{\infty}A_k=\{x_0\}$.*

Доказательство. Для любого k фиксируем точку $x_k \in A_k$. Если $l > k$, то $x_l \in A_l \subset A_k$. Значит, точки x_l, $x_k \in A_k$. Тогда $\|x_k - x_l\| \le d(A_k) \to 0$, $k \to \infty$. Это означает, что последовательность x_k фундаментальна в $\mathbb{R}^n$, следовательно по критерию Коши сходится к некоторому x_0. Так как $x_l \in A_k$, $l \ge k$, то $x_0 = \lim_{l\to\infty} x_l \in A_k^-$. Множество A_k компактно, поэтому замкнуто, т. е. $A_k^- = A_k$. Итак, $x_0 \in A_k$, $k \ge 1$.

Осталось показать, что такая точка единственна. Предположим противное. Если две различные точки x_0, x_0' принадлежат всем A_k, то $d(A_k) \ge \|x_0 - x_0'\| > 0$, $k \ge 1$. Но тогда $d(A_k) \not\to 0$, $k \to \infty$. Полученное противоречие полностью доказывает теорему.

1.7 Отображения многомерных пространств

Пусть n, $m \in \mathbb{N}$, множество $A \subset \mathbb{R}^n$. Рассмотрим некоторую функцию $f : A \to \mathbb{R}^m$. Запишем f покомпонентно:

$$y = f(x) = (f^1(x), f^2(x), \ldots, f^m(x)) =$$

$$= (f^1(x_1, x_2, \ldots, x_n), f^2(x_1, x_2, \ldots, x_n), \ldots, f^m(x_1, x_2, \ldots, x_n)).$$

Функция f представляет собой *векторную функцию векторного переменного* или, по-другому, векторную функцию n переменных x_1, $x_2, \ldots, x_n$. Если $m = 1$, то f называют *скалярной функцией векторного переменного* $x = (x_1, x_2, \ldots, x_n)$ или (скалярной) функцией n переменных x_1, $x_2, \ldots, x_n$. Если $n = 1$, то f называют *векторной функцией скалярного переменного* x или одного вещественного переменного. Часто вместо термина «векторная функция» будем употреблять термин «отображение».

Примеры. 1) Функция, действующая по правилу

$$f : (x, y) \mapsto (xy, \sin xe^y, x^2 + y^2),$$

отображает $\mathbb{R}^2$ в $\mathbb{R}^3$.

2) Функция $f(x, y) = \ln(x^2 + y^2)$ — вещественная функция двух переменных x и y.

3) Функция $x \mapsto (\arcsin x, \arccos x, x^2)$ является векторной функцией одного вещественного переменного x.

1.8 Предел функции в точке

Пусть $A \subset \mathbb{R}^n$, $f : A \to \mathbb{R}^m$. Пусть точка x_0 является предельной точкой множества A. Пусть $y \in \mathbb{R}^m$. Следующие условия эквивалентны:

1) $\forall \varepsilon > 0 \ \exists \delta > 0 : \ \forall x \in A \ \ (0 < \|x - x_0\| < \delta \Rightarrow \|f(x) - y_0\| < \varepsilon)$.

2) Для любой последовательности точек (x_k) из множества $A \setminus \{x_0\}$ имеет место импликация $x_k \to x_0 \Rightarrow f(x_k) \to y_0$.

3) $\forall \varepsilon > 0 \ \exists \delta > 0 : \ f(A \cap \overset{\vee}{B}_\delta(x_0)) \subset B_\varepsilon(y_0)$.

Любое из этих трех условий можно принять за *определение предела функции в точке*: $\lim_{x \to x_0} f(x) = y_0$.

Для пределов векторных функций векторного аргумента справедливы свойства, аналогичные свойствам пределов вещественных функций вещественного аргумента: теоремы о пределе линейной комбинации, произведения и частного (при $m = 1$), сложной функции. Имеет место

Теорема (критерий Коши). *Пусть $A \subset \mathbb{R}^n$, $f : A \to \mathbb{R}^m$ и точка x_0 является предельной точкой множества A. Для того чтобы существовал предел $\lim_{x \to x_0} f(x)$, необходимо и достаточно, чтобы*

$$\forall \varepsilon > 0 \ \exists \delta > 0 : \ \forall x, x' \in A \setminus \{x_0\} \ \ (\|x' - x''\| < \delta \Rightarrow \|f(x') - f(x'')\| < \varepsilon).$$

Если $y = f(x) = (f^1(x), f^2(x), \ldots, f^m(x))$, то $\lim_{x \to x_0} f(x)$ существует тогда и только тогда, когда для любого l, $1 \leq l \leq m$, существует $\lim_{x \to x_0} f^l(x)$. При этом

$$\lim_{x \to x_0} f(x) = (\lim_{x \to x_0} f^1(x), \lim_{x \to x_0} f^2(x), \ldots, \lim_{x \to x_0} f^m(x)).$$

1.9 Непрерывность отображений в точке

Отображение $f : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$, называется *непрерывным в точке* $x_0 \in A$, если выполняется одно из двух эквивалентных условий:

1) $\forall \varepsilon > 0\ \exists \delta > 0 :\ \forall x \in A\ \ (\|x - x_0\| < \delta \Rightarrow \|f(x) - f(x_0)\| < \varepsilon)$.

2) Для любой последовательности точек (x_k) из множества A имеет место импликация $x_k \to x_0 \Rightarrow f(x_k) \to f(x_0)$.

Пусть $y = f(x) = (f^1(x), f^2(x), \dots, f^m(x))$. Отображение f непрерывно в точке x_0 тогда и только тогда, когда для любого l, $1 \le l \le m$, скалярные функции f^l непрерывны в точке x_0.

Пусть $x_0 = (x_0^1, x_0^2, \dots, x_0^n) \in A$. Для любого p, $1 \le p \le n$, рассмотрим векторную функцию скалярного аргумента t:

$$g_p(t) = f(x_0^1, x_0^2, \dots, x_0^{p-1}, t, x_0^{p-1}, \dots, x_0^n).$$

Если для любого p, $1 \le p \le n$, функция g_p непрерывна в точке x_0^p, то говорят, что f *непрерывна в точке* x_0 *по каждой переменной.* В отличие от непрерывности по каждой переменной, обычную непрерывность называют *непрерывностью по совокупности переменных.*

Ясно, что из непрерывности отображения по совокупности переменных следует непрерывность по каждой переменной. Обратное неверно.

Пример. Рассмотрим функцию

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

Очевидно, что функция f непрерывна на множестве $\mathbb{R}^2 \setminus \{(0,0)\}$. Так как $f(x,0) \equiv 0$, $f(0,y) \equiv 0$, то функция f непрерывна в точке $(0,0)$ по каждой переменной в отдельности.

Покажем, что функция f не является, тем не менее, непрерывной по совокупности переменных в точке $(0,0)$. Для этого рассмотрим последовательность точек $(\frac{1}{k}, \frac{1}{k})$, стремящуюся к $(0,0)$. Имеем

$$f\left(\frac{1}{k}, \frac{1}{k}\right) = \frac{1}{2} \not\to f(0,0) = 0.$$

Таким образом, f не непрерывна в точке $(0,0)$.

Как и в случае вещественных функций вещественного переменного вводится понятие непрерывности отображения на множестве.

Свойства отображений, непрерывных в точке (на множестве).

1) Линейная комбинация непрерывных отображений непрерывна.

2) Произведение и частное (если знаменатель не обращается в нуль) непрерывных отображений непрерывно.

3) Суперпозиция непрерывных отображений непрерывна.

1.10 Свойства непрерывных функций на компактном множестве

Теорема 1. *Если $f : A \to \mathbb{R}$ непрерывная функция, множество $A \subset \mathbb{R}^n$ компактно, то $f(A)$ — компактно.*

Как следствие получаем следующую теорему.

Теорема Вейерштрасса. *Непрерывная функция $f : A \to \mathbb{R}$, определенная на компактном множестве $A \subset \mathbb{R}^n$, является ограниченной и принимает на A свои наибольшее и наименьшее значения.*

Теорема Кантора. *Непрерывная функция $f : A \to \mathbb{R}$, определенная на компактном множестве $A \subset \mathbb{R}^n$, является равномерно непрерывной, т. е.*

$$\forall \varepsilon > 0 \; \exists \delta > 0 : \; \forall x', x'' \in A \quad (\|x' - x''\| < \delta \Rightarrow \|f(x') - f(x'')\| < \varepsilon).$$

Множество $A \subset \mathbb{R}^n$ называется *линейно связным*, если для любых точек a, $b \in A$ существует кривая в A, соединяющая точки a и b.

Теорема о промежуточном значении. *Пусть непрерывная функция $f : A \to \mathbb{R}$ определена на линейно связном множестве $A \subset \mathbb{R}^n$.*

Тогда для любых a, $b \in A$ *функция* f *принимает любое значение, лежащее между* $f(a)$ *и* $f(b)$.

Доказательство. Предположим, для определенности, что $f(a) \leq f(b)$. Если $f(a) = f(b)$, то доказательство тривиально. Если $f(a) < f(b)$, то пусть $\gamma \in [f(a), f(b)]$. Докажем, что существует точка c из множества A такая, что $f(c) = \gamma$. Так как A линейно связно, то существует кривая с представлением $x : [0,1] \to A$, соединяющая a и b, т. е. такая, что $x(0) = a$, $x(1) = b$. Рассмотрим сложную функцию $y = f(x(t))$, $0 \leq t \leq 1$. Эта функция является непрерывной на отрезке $[0,1]$, при этом $f(x(0)) = f(a)$, $f(x(1)) = f(b)$. По теореме Больцано-Коши о промежуточном значении существует точка $t_0 \in [0,1]$ такая, что $f(x(t_0)) = \gamma$. Следовательно, можно положить $c = x(t_0)$. Теорема доказана.

2 Дифференцирование скалярных функций нескольких переменных

2.1 Частные производные

Пусть функция $f : U \to R$, где U — открытое множество в $\mathbb{R}^n$. Рассмотрим стандартный базис из единичных векторов в $\mathbb{R}^n$:

$$\begin{array}{rcl} e_1 & = & (1, 0, 0, \ldots, 0), \\ e_2 & = & (0, 1, 0, \ldots, 0), \\ \ldots & \ldots & \ldots\ldots\ldots\ldots \\ e_n & = & (0, 0, 0, \ldots, 1). \end{array}$$

Для любого k, $1 \leq k \leq n$, определим *частичное приращение* функции f вдоль k-й координаты в точке $x = (x_1, x_2, \ldots, x_n)$:

$$\Delta_k f(t) := f(x + te_k) - f(x) =$$

$$= f(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, \ldots, x_n) - f(x_1, x_2, \ldots, x_n).$$

Отметим, что если t достаточно малое по модулю вещественное число, то точка $x + te_k$ лежит в U и выражение $\Delta_k f(t)$ имеет смысл.

Если существует предел $\lim_{t\to 0} \frac{\Delta_k f(t)}{t}$, то этот предел называется частной производной функции f в точке x по переменной x_k и обозначается $\frac{\partial f(x)}{\partial x_k}$. Таким образом,

$$\frac{\partial f(x)}{\partial x_k} = \lim_{t\to 0} \frac{\Delta_k f(t)}{t} = \lim_{t\to 0} \frac{f(x + te_k) - f(x)}{t}$$

или в более развернутом виде

$$\frac{\partial f(x)}{\partial x_k} = \lim_{t\to 0} \frac{f(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{t}.$$

Замечание 1. Отметим, что частичное приращение функции f вдоль k-й координаты и, следовательно, частная производная $\frac{\partial f(x)}{\partial x_k}$ зависят только от значений функции в точках вида $x + te_k$, попадающих в U, т. е. от значений функции в точках пересечения прямой

$$\mathcal{L}_x^k := \{x + te_k,\ t \in \mathbb{R}\},$$

проходящей через точку x и параллельной k-й координатной оси, и множества U.

Замечание 2. Если ввести функции

$$g_k(t) := f(x_1, x_2, \ldots, x_{k-1}, x_k + t, x_{k+1}, \ldots, x_n),$$

$$h_k(\tau) := f(x_1, x_2, \ldots, x_{k-1}, \tau, x_{k+1}, \ldots, x_n),$$

то

$$\frac{\partial f(x)}{\partial x_k} = g_k'(0) = h_k'(x_k).$$

Таким образом, можно вычислять частные производные функции как производные функции одной переменной, если считать другие переменные константами!

Примеры. 1) Пусть $f(x, y) = xy \sin x$. Тогда, считая y константой и дифференцируя по x, получаем

$$\frac{\partial f(x, y)}{\partial x} = y(\sin x + x \cos x).$$

Аналогично,

$$\frac{\partial f(x, y)}{\partial y} = x \sin x.$$

2) Пусть $f(x) = \|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$, $x \in \mathbb{R}^n$. Тогда

$$\frac{\partial f(x)}{\partial x_k} = \frac{1}{2} \cdot \frac{1}{\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}} \cdot 2x_k = \frac{x_k}{\|x\|}, \quad \|x\| \neq 0.$$

Пусть существуют все частные производные функции f в точке x, и эти производные являются конечными числами. В силу замечания 2) из этого вытекает дифференцируемость функций h_k в точках x_k. Отсюда следует, что функции h_k непрерывны в точках x_k, а это влечет непрерывность функции f по каждой переменной x_k в точке x. Однако, в силу замечания 1) из существования всех частных производных функции f в точке x не следует непрерывность функции f в точке x по совокупности переменных. Вообще ничего нельзя сказать о поведении функции в окрестности точки x вне $\cup_{k=1}^{n}\mathcal{L}_x^k$.

2.2 Дифференцируемость функций нескольких переменных

Напомним, что функция $f : D \to \mathbb{R}$, определенная на открытом множестве $D \subset \mathbb{R}$ *дифференцируема в точке* $x_0 \in D$, если

$$\Delta f(x_0) := f(x_0 + h) - f(x_0) = a \cdot h + o(h), \quad h \to 0.$$

Если f дифференцируема в точке x_0, то $a = f'(x_0)$. При этом функция $y = a(x-x_0)+y_0$, где $y_0 = f(x_0)$, является уравнением касательной к графику функции f в точке x_0. Соответствие $h \mapsto a \cdot h$ является линейным отображением, тесно связанным с уравнением касательной, которое перепишем в несколько измененном виде $y - y_0 = a(x - x_0)$ или $\Delta y = a \cdot \Delta x$ ($\Delta x = h$).

Приведенное выше определение допускает естественное обобщение на случай функций нескольких переменных. Как мы отмечали выше, из существования конечных частных производных в точке не следует даже непрерывности функции в этой точке. Поэтому при введении понятия дифференцируемости нужно рассматривать всевозможные малые сдвиги h точки x_0, а не только вдоль координатных осей.

Итак, пусть $f : U \to \mathbb{R}$, U — открытое множество в пространстве $\mathbb{R}^n$ и точка $x_0 = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}) \in U$. Функция f называется *дифференцируемой в точке* x_0, если существуют константы a_1, $a_2, \dots, a_n$ такие, что для любых $h = (h_1, h_2, \dots, h_n)$ таких, что $x_0 \in U$, имеет место равенство

$$\Delta f(x_0) := f(x_0+h) - f(x_0) = \sum_{k=1}^{n} a_k h_k + o(\|h\|), \quad \|h\| \to 0. \tag{2.1}$$

Таким образом, если функция дифференцируема в точке x_0, то ее приращение является суммой выражения $\sum_{k=1}^{n} a_k h_k$, линейно зависящего от приращения аргумента h, и величины порядка малости больше, чем $\|h\|$.

Выражение $\sum_{k=1}^{n} a_k h_k$ называется *дифференциалом функции* f в точке x_0 и обозначается $df(x_0)$. Компоненты h_k вектора h называются *дифференциалами независимых переменных* x_1, $x_2, \ldots, x_n$ и обозначаются dx_k. Таким образом,

$$df(x_0) = \sum_{k=1}^{n} a_k dx_k.$$

Дифференциал представляет из себя выражение, линейно зависящее от $dx_k = h_k$. Выясним смысл констант a_k, входящих в выражение для дифференциала. Покажем, что они совпадают с частными производными функции f в точке x_0 по переменным x_k, $1 \le k \le n$.

Теорема (необходимое условие дифференцируемости). *Если функция f дифференцируема в точке $x_0 \in U$, то f непрерывна в точке x_0 и в этой точке x_0 существуют все частные производные $\frac{\partial f(x_0)}{\partial x_k}$, причем $\frac{\partial f(x_0)}{\partial x_k} = a_k$, $1 \le k \le n$, т. е.*

$$df(x_0) = \sum_{k=1}^{n} \frac{\partial f(x_0)}{\partial x_k}\, dx_k.$$

Доказательство. Если функция f дифференцируема в точке x_0, то имеет место (2.1), откуда

$$\Delta f(x_0) = f(x_0+h) - f(x_0) = \sum_{k=1}^{n} a_k h_k + o(\|h\|), \quad \|h\| \to 0. \tag{2.2}$$

Если $\|h\| \to 0$, то все $h_k \to 0$ и $\sum_{k=1}^{n} a_k h_k \to 0$. Кроме того, по определению $o(\|h\|) \to 0$, $\|h\| \to 0$. В силу (2.2) $\Delta f(x_0) \to 0$, $\|h\| \to 0$, а это означает непрерывность функции f в точке x_0.

Положим теперь $h = te_k$, $t \in \mathbb{R}$. Тогда $\|h\| = |t|$, $h_j = 0$. $j \neq k$, $h_k = t$ и в силу (2.2)

$$f(x_0 + te_k) - f(x_0) = a_k t + o(|t|), \quad t \to 0.$$

Следовательно, функция

$$g_k(t) := f(x_0 + te_k) = f(x_1^{(0)}, x_2^{(0)}, \ldots, x_{k-1}^{(0)}, x_k^{(0)} + t, x_{k+1}^{(0)}, \ldots, x_n^{(0)})$$

дифференцируема в точке $t = 0$, ее производная равна a_k, т. е. существует частная производная $\frac{\partial f(x_0)}{\partial x_k} = a_k$. Теорема доказана.

2.3 Геометрический смысл дифференцируемости функции в точке

Дифференцируемость функции f в точке x_0 означает, что приращение функции f в этой точке ведет себя почти (т. е. с точностью до величин, бесконечно малых по сравнению с приращением) как линейная функция от приращений координат. Следовательно, график функции ведет себя почти как график функции

$$y = f(x_0) + \sum_{k=1}^{n} a_k \left(x_k - x_k^{(0)} \right), \quad \text{где} \quad \frac{\partial f(x_0)}{\partial x_k} = a_k.$$

График функции, задаваемой последним уравнением, представляет собой гиперплоскость в $\mathbb{R}^n$. Эта гиперплоскость называется *плоскостью, касательной к графику функции f в точке x_0*.

Пример. Рассмотрим функцию $f(x, y) = \sqrt{|xy|}$, $(x, y) \in \mathbb{R}^2$. У этой функции существуют нулевые частные производные в точке $(0, 0)$, так как

$$f(x, 0) \equiv 0, \quad f(0, y) \equiv 0.$$

Докажем, что f не дифференцируема в точке $(0, 0)$. В противном случае мы бы имели $df(0, 0) = 0 \cdot dx + 0 \cdot dy \equiv 0$ и тогда для приращения

$h = (h^1, h^2)$ выполнялось бы равенство

$$f(h) = f(h) - f(0) = df(0,0) + o(\|h\|) = o(\|h\|), \quad \|h\| \to 0.$$

Докажем, что

$$f(h) \neq o(\|h\|), \quad \|h\| \to 0. \tag{2.3}$$

Для этого рассмотрим последовательность $h_m = \left(\frac{1}{m}, \frac{1}{m}\right)$. При $m \to \infty$ имеем $h_m \to (0,0)$, но

$$\frac{f(h_m)}{\|h_m\|} = \frac{1}{\sqrt{2}} \not\to 0, \quad m \to \infty.$$

Это доказывает (2.3).

Теорема (достаточное условие дифференцируемости функции). *Пусть функция $f : U \to \mathbb{R}$, U — открытое множество в $\mathbb{R}^n$, и в некоторой δ-окрестности точки $x \in U$ существуют все частные производные функции f. Если эти производные являются непрерывными функциями в точке x, то тогда функция f дифференцируема в точке x.*

Доказательство. Так как множество U открыто и $x \in U$, существует $\delta > 0$ такое, что $B_\delta(x) \subset U$. Рассмотрим приращение функции f точке x, соответствующее вектору $h = \sum_{k=1}^n h_k e_k$. Будем считать, что $\|h\| < \delta$. Введем вектора

$$h^{(j)} = \sum_{k=1}^{j} h_k e_k = (h_1, h_2, \ldots, h_j, 0, 0, \ldots, 0), \quad 1 \leq j \leq n.$$

Ясно, что $h^{(n)} = h$. Положим $h^{(0)} = \theta$. Так как $\|h^{(j)}\| \leq \|h\| < \delta$, то для любого $1 \leq j \leq n$ имеем $x + h^{(j)} \in B_\delta(x) \subset U$. Тогда

$$\Delta f(x) = f(x+h) - f(x) = f(x + h^{(n)}) - f(x + h^{(0)}) =$$

$$= \sum_{j=1}^{n} \left(f(x + h^{(j)}) - f(x + h^{(j-1)}) \right). \tag{2.4}$$

Очевидно, что $h^{(j)} = h^{(j-1)} + h_j e_j$, $1 \leq j \leq n$. Точки $x + h^{(j-1)} + t e_j$ при $t \in [0; h_j]$ лежат на отрезке с концами $h^{(j-1)}$ и $h^{(j-1)}$. Поскольку концы отрезка лежат в $B_\delta(x)$, то и весь отрезок лежит в $B_\delta(x)$. Рассмотрим функцию $g_j(t) := f(x+h^{(j-1)}+te_j)$, $t \in [0; h_j]$. Так как функция f имеет частную производную $\partial f/\partial x_j$ во всех точках из $B_\delta(x)$, то функция g_j дифференцируема на отрезке $[0; h_j]$ и

$$g_j'(t) = \frac{\partial f(x + h^{(j-1)} + te_j)}{\partial x_j}, \quad t \in [0; h_j].$$

С использованием формулы конечных приращений Лагранжа получаем

$$f(x + h^{(j)}) - f(x + h^{(j-1)}) = g_j(h_j) - g_j(0) = g_j'(\xi_j)h_j =$$

$$= \frac{\partial f(x + h^{(j-1)} + \xi_j e_j)}{\partial x_j} h_j \tag{2.5}$$

для некоторого $\xi_j \in (0; h_j)$.

Если $\|h\| \to 0$, то $\xi_j \to 0$, $x + h^{(j-1)} + \xi_j e_j \to x$ и в силу непрерывности частных производных функции f в точке x

$$\frac{\partial f(x + h^{(j-1)} + \xi_j h_j e_j)}{\partial x_j} \to \frac{\partial f(x)}{\partial x_j},$$

поэтому

$$\frac{\partial f(x + h^{(j-1)} + \xi_j h_j e_j)}{\partial x_j} = \frac{\partial f(x)}{\partial x_j} + o(1). \tag{2.6}$$

Из (2.4), (2.5) и (2.6) получаем

$$\Delta f(x) = \sum_{j=1}^{n} \left(\frac{\partial f(x)}{\partial x_j} + o(1) \right) h_j =$$

$$= \sum_{j=1}^{n} \frac{\partial f(x)}{\partial x_j} h_j + \sum_{j=1}^{n} o(h_j) = \sum_{j=1}^{n} \frac{\partial f(x)}{\partial x_j} h_j + o(\|h\|).$$

Это доказывает дифференцируемость функции f в точке x. Теорема доказана.

Замечание. Непрерывность частных производных не является необходимым условием дифференцируемости функции f в точке x, даже если они существуют в некоторой окрестности точки x.

Пример. Рассмотрим функцию

$$f(x) = \begin{cases} \|x\|^2 \sin \dfrac{1}{\|x\|}, & x \neq \theta, \\ 0, & x = \theta. \end{cases}$$

Функция f дифференцируема в точке 0, так как

$$f(x) - f(\theta) = f(x) = \|x\|^2 \sin \frac{1}{\|x\|} = \|x\|^2 O(1) = o(\|x\|), \quad \|x\| \to 0.$$

Из последнего равенства следует, что дифференциал в точке $x = \theta$ равен нулю, поэтому все частные производные функции f в этой точке равны нулю. Частные производные существуют и во всех других точках но они не являются непрерывными в точке $x = \theta$. Действительно, если $x \neq \theta$, то

$$\frac{\partial f(x)}{\partial x_k} = 2x_k \sin \frac{1}{\|x\|} + \|x\|^2 \cos \frac{1}{\|x\|} \left(-\frac{1}{\|x\|^2}\right) \frac{\partial \|x\|}{\partial x_k} =$$

$$= 2x_k \sin \frac{1}{\|x\|} - \frac{x_k}{\|x\|} \cos \frac{1}{\|x\|}.$$

Однако не существует пределов частных производных при $x \to \theta$, так как, во-первых,

$$\lim_{x \to \theta} x_k \sin \frac{1}{\|x\|} = 0$$

в силу того, что $x_k \to 0$ и $\sin \frac{1}{\|x\|} = O(1)$, $x \to \theta$, а во-вторых, но не существует

$$\lim_{x \to \theta} \frac{x_k}{\|x\|} \cos \frac{1}{\|x\|}$$

(докажите это самостоятельно!).

Функция $f : U \to \mathbb{R}$ называется *дифференцируемой на множестве* U, если f дифференцируема в любой точке множества U. Функция f называется *непрерывно дифференцируемой* на U, если ее частные производные существуют и являются непрерывными функциями на U. Будем говорить, что f *непрерывно дифференцируема в точке* x, если ее частные производные существуют и являются непрерывными в некоторой окрестности точки x.

2.4 Дифференцирование сложной функции

Пусть функция $f : U \to \mathbb{R}$, U — открытое множество в $\mathbb{R}^n$, вектор-функция $x : I \to U$, где I — некоторый числовой промежуток в $\mathbb{R}$. Будем называть функцию $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))$ *дифференцируемой (непрерывно дифференцируемой) в точке* t, если все компоненты $x_k(t)$ — вещественные функции вещественного переменного — дифференцируемы (непрерывно дифференцируемы) в точке t. Обозначим $\widehat{f}(t) = f(x(t))$, $t \in I$.

Теорема 1. *Пусть функция $x : I \to U$ дифференцируема в точке t, функция $f : U \to \mathbb{R}$ дифференцируема в точке $x(t)$, тогда функция $\widehat{f}$ дифференцируема в точке t и*

$$\frac{d\widehat{f}(t)}{dt} = \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} \frac{dx_k(t)}{dt}. \tag{2.7}$$

Доказательство. Так как функция f дифференцируема в точке $x = x(t)$, то

$$\Delta f(x) = f(\overline{x}) - f(x) = \sum_{k=1}^{n} \frac{\partial f(x)}{\partial x_k}(\overline{x}_k - x_k) + o(\|\overline{x} - x\|), \quad \overline{x} \to x.$$

Пусть $\overline{x} = x(t + \Delta t)$, тогда в силу дифференцируемости функций $x_k(t)$ в точке t имеем

$$\overline{x}_k - x_k = x_k(t + \Delta t) - x(t) = \frac{dx_k(t)}{dt} \Delta t + o(\Delta t), \quad \Delta t \to 0.$$

Пусть $T := \{\Delta t \in \mathbb{R} \mid x(t + \Delta t) \neq x(t)\}$.

1) Если $\Delta t \in T$, то

$$\widehat{f}(t + \Delta t) - \widehat{f}(t) = f(x(t + \Delta t)) - f(x(t)) =$$

$$= \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} (x_k(t + \Delta t) - x_k(t)) + o(\|\overline{x} - x\|) =$$

$$= \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} \frac{dx_k(t)}{dt} \Delta t + \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} o(\Delta t) + o(\|\overline{x} - x\|), \tag{2.8}$$

$\Delta t \to 0$. Ясно, что

$$\sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k}\, o(\Delta t) = o(\Delta t), \quad \Delta t \to 0.$$

Покажем, что

$$o(\|\overline{x} - x\|) = o(\Delta t), \quad \Delta t \in T,\ \Delta t \to 0. \tag{2.9}$$

Имеем

$$\lim_{\Delta t \to 0, \Delta t \in T} \frac{o(\|\overline{x} - x\|}{|\Delta t|} = \lim_{\overline{x} \to x} \frac{o(\|\overline{x} - x\|}{\|\overline{x} - x\|} \lim_{\Delta t \to 0} \frac{\|\overline{x} - x\|}{|\Delta t|}.$$

Так как

$$\lim_{\overline{x} \to x} \frac{o(\|\overline{x} - x\|}{\|\overline{x} - x\|} = 0,$$

$$\lim_{\Delta t \to 0} \frac{\|\overline{x} - x\|}{|\Delta t|} = \lim_{\Delta t \to 0} \sqrt{\sum_{k=1}^{n} \left(\frac{\overline{x}_k - x_k}{\Delta t} \right)^2} = \sqrt{\sum_{k=1}^{n} (x_k'(t))^2},$$

то имеет место (2.9).

Из (2.8), (2.9) следует

$$\widehat{f}(t + \Delta t) - \widehat{f}(t) = \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} \frac{dx_k(t)}{dt} \Delta t + o(\Delta t), \quad \Delta t \in T,\ \Delta t \to 0.$$

2) Если $\Delta t = 0$ — изолированная точка множества T, то доказательство закончено. В противном случае имеем

$$\begin{aligned} 0 = \widehat{f}(t + \Delta t) - \widehat{f}(t) = f(x(t + \Delta t)) - f(x(t)) = \\ = \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} (x_k(t + \Delta t) - x_k(t)) = \\ = \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} \frac{dx_k(t)}{dt} \Delta t + \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} o(\Delta t) = \\ = \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} \frac{dx_k(t)}{dt} \Delta t + o(\Delta t), \quad \Delta t \notin T,\ \Delta t \to 0. \end{aligned} \tag{2.10}$$

Из (2.9) и (2.10) следует дифференцируемость функции $\widehat{f}$ и равенство (2.7). Теорема доказана.

Пусть $x : G \to U$, где $G \subset \mathbb{R}^m$, $U \subset \mathbb{R}^n$ — открытые множества. Пусть $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))$. Будем говорить, что функция $x = x(t)$ дифференцируема в точке t (на множестве G), если вещественнозначные функции $x_l(t)$, $1 \le l \le n$ дифференцируемы в точке t (на множестве G). аналогично вводится понятие непрерывной дифференцируемости.

Теорема 2. *Пусть $x : G \to U$, $f : U \to \mathbb{R}$. Если вектор-функция $x = x(t)$ дифференцируема (непрерывно дифференцируема) в точке $t \in G$ и функция f дифференцируема (непрерывно дифференцируема) в точке $x(t)$, то сложная функция $\widehat{f}(t) = f(x(t))$ дифференцируема (непрерывно дифференцируема) в точке t и*

$$\frac{\partial \widehat{f}(t)}{\partial t_l} = \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} \frac{\partial x_k(t)}{\partial t_l} \,. \tag{2.11}$$

Доказательство. Дифференцируемость функции $\widehat{f}$ доказывается аналогично тому, как это было сделано в предыдущей теореме. Точно так же как в теореме 1 доказывается (2.11); по-существу, (2.11) следует сразу из (2.7).

Если функции f и x непрерывно дифференцируемы в соответствующих точках, то частные производные $\frac{\partial f}{\partial x_k}$ непрерывны в точке $x(t)$, а частные производные $\frac{\partial x_k}{\partial t_l}$ — в точке t. Следовательно, правая часть равенства (2.11) непрерывна в точке t, так как получается из непрерывных функций с помощью арифметических операций и операции суперпозиции. Таким образом, левая часть $\frac{\partial \widehat{f}(t)}{\partial t_l}$ непрерывна в точке t. Теорема доказана.

2.5 Инвариантность формы первого дифференциала

Пусть выполняются условия предыдущей теоремы. Тогда для функции $\widehat{f}(t) = f(x(t))$ имеем

$$d\widehat{f}(t) = df(x(t)) = \sum_{l=1}^{m} \sum_{k=1}^{n} \frac{\partial \widehat{f}(t)}{\partial t_l}\, dt_l = \sum_{l=1}^{m} \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} \frac{\partial x_k(t)}{\partial t_l}\, dt_l =$$

$$= \sum_{k=1}^{n} \sum_{l=1}^{m} \frac{\partial f(x(t))}{\partial x_k} \frac{\partial x_k(t)}{\partial t_l}\, dt_l = \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k} \sum_{l=1}^{m} \frac{\partial x_k(t)}{\partial t_l}\, dt_l =$$

$$= \sum_{k=1}^{n} \frac{\partial f(x(t))}{\partial x_k}\, dx_k(t). \tag{2.12}$$

Если бы x была бы независимой переменной, то тогда бы

$$df(x) = \sum_{k=1}^{n} \frac{\partial f(x)}{\partial x_k}\, dx_k. \tag{2.13}$$

Сравнивая (2.12) и (2.13), делаем вывод, что форма первого дифференциала не зависит от того, является ли x независимой переменной или функцией от другой переменной t. Это свойство называется инвариантностью формы первого дифференциала.

Отметим некоторые свойства дифференциалов.

Теорема. *Пусть функции f, $g : U \to \mathbb{R}$ (U — открытое множество в $\mathbb{R}^n$) дифференцируемы в точке x (на множестве U). Тогда в точке x (на множестве U) дифференцируемы функции $\alpha f + \beta g$, где α, $\beta \in \mathbb{R}$; произведение fg; частное f/g, если функция g не обращается в нуль на U. При этом*

1) $d(\alpha f + \beta g) = \alpha df + \beta dg$,

2) $d(fg) = f\, dg + g\, df$,

3) $d\left(\dfrac{f}{g}\right) = \dfrac{g\, df - f\, dg}{g^2}$.

Докажем, для примера, 2). Рассмотрим функцию $h(u,v)=u\cdot v$. Имеем

$$\frac{\partial h}{\partial u}=v,\quad \frac{\partial h}{\partial v}=u.$$

Поэтому $dh(u,v)=v\,du+u\,dv$. Пользуясь инвариантностью формы первого дифференциала, получаем $d(fg)=dh(f,g)=f\,dg+g\,df$.

Примеры дифференцирования сложных функций.

1) Рассмотрим функцию $\psi(x)=\sin x^{\cos x}$. Рассмотрим вспомогательную функцию $h(u,v)=u^v$. Тогда $dh(u,v)=vu^{v-1}du+u^v\ln v dv$. Пользуясь инвариантностью формы первого дифференциала, получаем

$$d\psi(x)=\cos x\sin x^{\cos x-1}d\sin x+\sin x^{\cos x}\ln\sin x d\cos x=$$

$$=(\cos^2 x\sin x^{\cos x-1}-\sin x^{\cos x+1}\ln\sin x)dx.$$

2) Пусть $f(u,v,w)$ — дифференцируемая функция трех переменных. Рассмотрим сложную функцию $\omega(x,y):=f(x,y,g(x,y))$, где $g(x,y))$ — дифференцируемая функция двух переменных. Тогда

$$\frac{\partial\omega}{\partial x}=\frac{\partial f}{\partial u}\cdot 1+\frac{\partial f}{\partial v}\cdot 0+\frac{\partial f}{\partial w}\frac{\partial g}{\partial x}=\frac{\partial f}{\partial u}+\frac{\partial f}{\partial w}\frac{\partial g}{\partial x}.$$

Аналогично

$$\frac{\partial\omega}{\partial y}=\frac{\partial f}{\partial v}+\frac{\partial f}{\partial w}\frac{\partial g}{\partial y}.$$

2.6 Формула конечных приращений

Теорема. *Пусть функция $f:U\to\mathbb{R}$ (U открыто в R^n), отрезок с концами в точках x и $x+h$ лежит в U и f дифференцируема на этом отрезке. Тогда существует $\xi\in(0;1)$ такое, что*

$$f(x+h)-f(x)=\sum_{k=1}^{n}\frac{\partial f(x+\xi h)}{\partial x_k}h_k,$$

где $h=(h_1,h_2,\ldots,h_n)$.

Доказательство. Рассмотрим функцию одного вещественного переменного $\varphi(t):=f(x+th)$. Она определена и дифференцируема на отрезке $[0;1]$ как суперпозиция дифференцируемых функций. По формуле конечных приращений для функций одной переменной получаем, что существует $\xi \in (0;1)$ такое, что

$$f(x+h) - f(x) = \varphi(1) - \varphi(0) = \varphi'(\xi) =$$

$$= \sum_{k=1}^{n} \frac{\partial f(x+th)}{\partial x_k}\bigg|_{t=\xi} \frac{d(x_k + th_k)}{dt} = \sum_{k=1}^{n} \frac{\partial f(x+\xi h)}{\partial x_k} h_k.$$

Теорема доказана.

Замечание. Формулу конечных приращений можно записать в виде

$$\Delta f(x,h) = df(x+\xi h, h)$$

где $\Delta f(x,h)$ и $df(x+\xi h, h)$ — приращение и дифференциал функции, соответствующие приращению аргумента h и записанные в точках x и $x+\xi h$ соответственно.

Пусть $U \subset R^n$ — область, т. е. открытое линейно связное множество. Можно доказать, что любые две точки x, $y \in U$ можно соединить ломаной, целиком лежащей с области U.

Теорема (критерий постоянства функции). *Пусть функция $f : U \to \mathbb{R}^n$ дифференцируема в области U. Для того чтобы функция была постоянной в U, необходимо и достаточно, чтобы выполнялись равенства*

$$\frac{\partial f(x)}{\partial x_k} = 0, \quad x \in U,\ 1 \le k \le n. \tag{2.14}$$

Доказательство. Необходимость условий (2.14) очевидна. Докажем достаточность. Пусть имеет место (2.14). Рассмотрим любые точки x, $y \in U$. Пусть L — ломаная, соединяющая точки x и y в U с последовательными вершинами в точках $x^{(0)} = x$, $x^{(1)}$, $x^{(2)}, \ldots, x^{(N)} = y$. В силу формулы конечных приращений для любого j, $1 \le j \le N$, существует точка $a^{(j)}$, лежащая на отрезке с концами $x^{(j)}$ и $x^{(j-1)}$ такая, что

$$f(x^{(j)}) - f(x^{(j-1)}) = \sum_{k=1}^{n} \frac{\partial f(a^{(j)})}{\partial x_k} \left(x_k^{(j)} - x_k^{(j-1)}\right) = 0,$$

так как все частные производные функции f в U равны нулю в силу (2.14). Поэтому

$$f(x) = f(x^{(0)}) = f(x^{(1)}) = f(x^{(2)}) = \cdots = f(x^{(N)}) = f(y).$$

Итак, в любых двух точках из U значения функции f совпадают. Значит, $f \equiv \text{const}$ в U. Теорема доказана.

2.7 Частные производные и дифференциалы высших порядков

Пусть функция f имеет производную $\frac{\partial f}{\partial x_k}$ в окрестности точки x_0. Если существует частная производная $\frac{\partial}{\partial x_l}\left(\frac{\partial f}{\partial x_k}\right)$ в точке x_0, то она называется *частной производной второго порядка* функции f *по переменным* x_k и x_l в точке x_0 и обозначается $\frac{\partial^2 f(x_0)}{\partial x_k \partial x_l}$.

Аналогично по индукции определяются частные производные более высоких порядков. Если в окрестности точки x_0 существует частная производная $(k-1)$-го порядка $\frac{\partial^{k-1} f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_{k-1}}}$, где $1 \le i_j \le n$, $1 \le j \le k-1$, $k \ge 2$, и у нее существует частная производная в точке x_0 по переменной x_{i_k}, $1 \le i_k \le n$, то эта производная называется *k-й частной производной функции f в точке x_0 по переменным* x_{i_1}, $x_{i_3}, \ldots, x_{i_k}$ и обозначается $\frac{\partial^k f(x_0)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}}$.

Для сокращения записи вместо $\underbrace{\partial x_i \partial x_i \cdots \partial x_i}_{l \text{ раз}}$ пишут ∂x_i^l.

Пример. Пусть функция $z = f(x, y)$ является функцией двух вещественных переменных. У этой функции может существовать 4 вида различных частных производных 2-го порядка:

$$\frac{\partial^2 f}{\partial x^2} := \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \quad \frac{\partial^2 f}{\partial x \partial y} := \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right),$$

$$\frac{\partial^2 f}{\partial y \partial x} := \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right), \quad \frac{\partial^2 f}{\partial y^2} := \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right).$$

Функция f называется *дважды дифференцируемой* в точке $x_0 \in U$, если в некоторой окрестности точка x существует дифференциал

$df(x)$ и этот дифференциал является дифференцируемой в точке x_0 по переменной x функцией (при условии, что дифференциалы dx_1, $dx_2, \ldots, dx_n$ — константы. По определению, *второй дифференциал функции* f в точке x_0 или дифференциал второго порядка — это дифференциал от первого дифференциала в точке x_0, он обозначается $d^2f(x_0)$. Итак,

$$d^2f(x_0) = d(df)(x_0) = d\left(\sum_{k=1}^{n} \frac{\partial f(x)}{\partial x_k} dx_k\right)\Bigg|_{x=x_0} = \sum_{k=1}^{n} d\left(\frac{\partial f(x)}{\partial x_k}\right)\Bigg|_{x=x_0} dx_k =$$

$$= \sum_{k=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f(x_0)}{\partial x_k \partial x_j} dx_k dx_j = \sum_{k,j=1}^{n} \frac{\partial^2 f(x_0)}{\partial x_k \partial x_j} dx_k dx_j.$$

Пример. Для функции двух переменных $f(x,y)$ имеем

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

Аналогично по индукции определяются дифференцируемость и дифференциалы более высоких порядков.

Так, функция f называется *трижды дифференцируемой*, если дифференцируем второй дифференциал и *третий дифференциал* или дифференциал третьего порядка в точке x есть

$$d^3 f(x) := d(d^2 f(x)) = \sum_{k,j,l=1}^{n} \frac{\partial^3 f(x)}{\partial x_k \partial x_j \partial x_l}\, dx_k dx_j dx_l.$$

Пусть теперь $m \in \mathbb{N}$, $m \geq 2$, и определено понятие дифференцируемости и дифференциала $(m-1)$-го порядка.

Функция f называется *m раз дифференцируемой* в точке $x \in U$, если в некоторой окрестности точка x существует дифференциал $d^{m-1}f$ порядка $(m-1)$ и этот дифференциал является дифференцируемой в точке x функцией. По определению, *дифференциал m-го порядка* функции f в точке x или m-й дифференциал — это дифференциал от дифференциала $(m-1)$-го порядка, он обозначается $d^m f(x)$:

$$d^m f(x) := d(d^{m-1} f(x)) = \sum_{k_1,k_2,\ldots,k_m=1}^{n} \frac{\partial^m f(x)}{\partial x_{k_1} \partial x_{k_2} \ldots \partial x_{k_m}}\, dx_{k_1} dx_{k_2} \ldots dx_{k_m}.$$

Очень часто частные производные не зависят от порядка дифференцирования, т. е. для любого набора индексов $(j_1, j_2, \ldots, j_k)$, получающегося перестановкой из набора $(i_1, i_2, \ldots, i_k)$, имеем

$$\frac{\partial^k f(x_0)}{\partial x_{j_1} \partial x_{j_2} \ldots \partial x_{j_k}} = \frac{\partial^k f(x_0)}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_k}}.$$

Это равенство будет выполнено, в частности, если частные производные являются непрерывными функциями. В этом случае можно объединять слагаемые в выражении для дифференциала, которые соответствуют дифференцированию по одним и тем же переменным. Например, если для функции двух переменных $f(x, y)$ выполняется равенство

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

то в выражении для второго дифференциала можно объединить два слагаемых и получить

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

В общем случае, если есть частная производная

$$\frac{\partial^m f}{\partial x_1^{l_1} \partial x_2^{l_2} \cdots \partial x_n^{l_n}},$$

где $l_i \geq 0$, $l_1 + l_2 + \cdots + l_n = m$, то можно показать с использованием комбинаторики, что число различных частных производных, которые только порядком дифференцирования могут отличаются от данной, равно

$$C_m^{l_1, l_2, \ldots, l_n} := \frac{m!}{l_1!\, l_2! \, \cdots \, l_n!}.$$

Поэтому в случае, когда частные производные m-го порядка не зависят от порядка дифференцирования, выражение для m-го дифференциала принимает вид

$$d^m f(x) = \sum_{l_i \geq 0, l_1 + l_2 + \cdots + l_n = m} C_m^{l_1, l_2, \ldots, l_n} \frac{\partial^m f(x)}{\partial x_1^{l_1} \partial x_2^{l_2} \cdots \partial x_n^{l_n}}\, dx_1^{l_1} dx_2^{l_2} \cdots dx_n^{l_n}. \quad (2.15)$$

Отметим также, что в силу формулы

$$(x_1 + x_2 + \cdots + x_n)^m = \sum_{l_i \geq 0, l_1 + l_2 + \cdots + l_n = m} C_m^{l_1, l_2, \ldots, l_n} x_1^{l_1} x_2^{l_2} \cdots x_n^{l_n},$$

обобщающей бином Ньютона на случай произвольного числа слагаемых, можно записать формулу (2.15) в виде

$$d^m f = \left(\frac{\partial}{\partial x_1} dx_1 + \frac{\partial}{\partial x_2} dx_2 + \cdots + \frac{\partial}{\partial x_n} dx_n \right)^m f.$$

Примеры. 1) При $m = 3$ для функции $f(x, y)$ получаем

$$d^3 f = \left(\frac{\partial}{\partial x} dx + \frac{\partial}{\partial y} dy \right)^3 f = \left(\frac{\partial^3}{\partial x^3} dx^3 + 3 \frac{\partial^3}{\partial x^2 \partial y} + 3 \frac{\partial^3}{\partial x \partial y^2} + \frac{\partial^3}{\partial y^3} \right) f =$$

$$= \frac{\partial^3 f}{\partial x^3} dx^3 + 3 \frac{\partial^3 f}{\partial x^2 \partial y} + 3 \frac{\partial^3 f}{\partial x \partial y^2} + \frac{\partial^3 f}{\partial y^3}.$$

2) При $m = 4$ для функции $f(x, y)$ получаем

$$d^4 f = \left(\frac{\partial}{\partial x} dx + \frac{\partial}{\partial y} dy \right)^4 f =$$

$$= \left(\frac{\partial^4}{\partial x^4} dx^4 + 4 \frac{\partial^4}{\partial x^3 \partial y} + 6 \frac{\partial^4}{\partial x^2 \partial y^2} + 4 \frac{\partial^4}{\partial x \partial y^3} + \frac{\partial^4}{\partial y^4} \right) f =$$

$$= \frac{\partial^4 f}{\partial x^4} dx^4 + 4 \frac{\partial^4 f}{\partial x^3 \partial y} + 6 \frac{\partial^4 f}{\partial x^2 \partial y^2} + 4 \frac{\partial^4 f}{\partial x \partial y^3} + \frac{\partial^4 f}{\partial y^4}.$$

3) Пусть $m = 3$ и функция $f(x, y, z)$ зависит от трех переменных. Имеем

$$C_3^{3,0,0} = 1, \quad C_3^{2,1,0} = 3, \quad C_3^{1,1,1} = 6,$$

поэтому

$$(x+y+z)^3 = x^3 + y^3 + z^3 + 3x^2 y + 3x^2 z + 3xy^2 + 3y^2 z + 3xz^2 + 3xy^2 + 6xyz$$

и

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

$$+3\frac{\partial^3 f}{\partial x\partial y^2}\,dxdy^2+3\frac{\partial^3 f}{\partial z\partial y^2}\,dy^2dz+3\frac{\partial^3 f}{\partial x\partial z^2}\,dxdz^2+$$

$$+3\frac{\partial^3 f}{\partial y\partial z^2}\,dydz^2+6\frac{\partial^3 f}{\partial x\partial y\partial z}\,dxdydz.$$

Функция f называется *m раз непрерывно дифференцируемой* в точке $x\in U$, если в некоторой окрестности этой точки существуют все частные производные m-го порядка функции f и эти частные производные являются непрерывными в точке x.

Функция f называется *m раз дифференцируемой (непрерывно дифференцируемой) в* U, если она дифференцируема (непрерывно дифференцируема) в любой точке $x\in U$.

2.8 Независимость частных производных от порядка дифференцирования

Для изучения вопроса о независимости частных производных от порядка дифференцирования достаточно рассмотреть случай функции двух переменных $z=f(x,y)$. Найдем довольно общие достаточные условия, при которых выполняется равенство

$$\frac{\partial^2 f(a,b)}{\partial x\partial y}=\frac{\partial^2 f(a,b)}{\partial y\partial x} \tag{2.16}$$

в фиксированной точке (a,b).

Теорема 1. *Пусть функция f дважды дифференцируема в точке (a,b). Тогда справедливо равенство* (2.16).

Доказательство. Введем функции

$$g(t):=f(a+t,b+t)-f(a+t,b)-f(a,b+t)+f(a,b),$$

$$\varphi(x)=f(x,b+t)-f(x,b). \tag{2.17}$$

Функция g определена и дифференцируема в окрестности точки $t = 0$, а функция φ — в окрестности точки $x = a$. При достаточно малых $|t|$ имеем по формуле конечных приращений

$$g(t) = \varphi(a + t) - \varphi(a) = \varphi'(a + \xi t)t, \quad \xi \in (0; 1).$$

Следовательно,

$$g(t) = \left(\frac{\partial f(x, b + t)}{\partial x} - \frac{\partial f(x, b)}{\partial x}\right)\bigg|_{x=a+\xi t} \cdot t =$$

$$= \left[\frac{\partial f(a + \xi t, b + t)}{\partial x} - \frac{\partial f(a + \xi t, b)}{\partial x}\right] t = \left\{\left[\frac{\partial f(a + \xi t, b + t)}{\partial x} - \frac{\partial f(a, b)}{\partial x}\right] - \right.$$

$$\left. - \left[\frac{\partial f(a + \xi t, b)}{\partial x} - \frac{\partial f(a, b)}{\partial x}\right]\right\} \cdot t = (A - B) \cdot t,$$

где

$$A = \frac{\partial f(a + \xi t, b + t)}{\partial x} - \frac{\partial f(a, b)}{\partial x}, \quad B = \frac{\partial f(a + \xi t, b)}{\partial x} - \frac{\partial f(a, b)}{\partial x}.$$

Так как функция f дважды дифференцируема в точке (a, b), то ее частные производные первого порядка дифференцируемы в точке (a, b), в частности,

$$\frac{\partial f(a + h, b + k)}{\partial x} - \frac{\partial f(a, b)}{\partial x} =$$

$$= \frac{\partial^2 f(a, b)}{\partial x^2} h + \frac{\partial^2 f(a, b)}{\partial x \partial y} k + o(\sqrt{h^2 + k^2}), \quad (2.18)$$

$\sqrt{h^2 + k^2} \to 0$. Тогда с учетом (2.18) получаем

$$A = \frac{\partial^2 f(a, b)}{\partial x^2} \xi t + \frac{\partial^2 f(a, b)}{\partial x \partial y} t + o(t\sqrt{1 + \xi^2}), \quad t \to 0.$$

Аналогично

$$B = \frac{\partial^2 f(a, b)}{\partial x^2} \xi t + \frac{\partial^2 f(a, b)}{\partial x \partial y} \cdot 0 + o(\xi t), \quad t \to 0.$$

Вычитая последние два равенства и используя очевидные равенства $o(t\sqrt{1+\xi^2}) = o(t)$, $o(\xi t) = o(t)$, $t \to 0$, получаем

$$g(t) = (A-B)t = \left(\frac{\partial^2 f(a,b)}{\partial x \partial y} t + o(t)\right) t = \\ = \frac{\partial^2 f(a,b)}{\partial x \partial y} t^2 + o(t^2), \quad t \to 0. \quad (2.19)$$

Совершенно аналогично можно показать, что

$$g(t) = \frac{\partial^2 f(a,b)}{\partial y \partial x} t^2 + o(t^2), \quad t \to 0.$$

Для этого следует представить $g(t)$ в виде

$$g(t) = \psi(b+t) - \psi(b), \quad \text{где} \quad \psi(y) = f(a+t, y) - f(a, y). \quad (2.20)$$

Сравнивая (2.19) и (2.20), получаем

$$\frac{\partial^2 f(a,b)}{\partial x \partial y} + \frac{o(t^2)}{t^2} = \frac{\partial^2 f(a,b)}{\partial y \partial x} + \frac{o(t^2)}{t^2}, \quad t \to 0.$$

После перехода к пределу при $t \to 0$ получаем (2.16). Теорема 1 доказана.

Теорема 2. *Пусть смешанные частные производные* $\frac{\partial^2 f}{\partial x \partial y}$ *и* $\frac{\partial^2 f}{\partial y \partial x}$ *существуют в некоторой окрестности точки* (a,b) *и являются непрерывными в точке* (a,b). *Тогда справедливо* (2.16).

Доказательство. Как и при доказательстве предыдущей теоремы введем функции g и φ по формулам (2.17). В силу формулы конечных приращений имеем

$$g(t) = \varphi(a+t) - \varphi(a) = \varphi'(a+\xi_1 t)t, \quad \xi_1 \in (0;1).$$

Таким образом,

$$g(t) = \left[\frac{\partial f(a+\xi_1 t, b+t)}{\partial x} - \frac{\partial f(a+\xi_1 t, b)}{\partial x}\right] \cdot t.$$

применяя еще раз формулу конечных приращений, на этот раз к функции $\omega(\tau) = \frac{\partial f(a+\xi_1 t, b+\tau)}{\partial x}$, получаем

$$g(t) = \frac{\partial^2 f(a+\xi_1 t, b+\eta_1 t)}{\partial x \partial y} t^2, \quad \xi_1,\, \eta_1 \in (0;1). \quad (2.21)$$

Аналогично показывается, что существуют ξ_2, $\eta_2 \in (0;1)$ такие, что

$$g(t) = \frac{\partial^2 f(a+\xi_2 t, b+\eta_2 t)}{\partial y \partial x}\, t^2, \quad \xi_2,\, \eta_2 \in (0;1). \tag{2.22}$$

Для этого нужно использовать вместо функции φ функция ψ, введенную при доказательстве теоремы 1.

Сравнивая (2.21) и (2.22), получаем

$$\frac{\partial^2 f(a+\xi_1 t, b+\eta_1 t)}{\partial x \partial y} = \frac{\partial^2 f(a+\xi_2 t, b+\eta_2 t)}{\partial y \partial x}. \tag{2.23}$$

Устремим t к нулю. Учитывая ограниченность величин ξ_1, η_1, ξ_2, η_2, получаем, что $(a+\xi_j t, b+\eta_j t) \to (a,b)$ в $\mathbb{R}^2$ при $t \to 0$, $j = 1,\ 2$. В силу непрерывности смешанных частных производных из (2.23) получаем (2.16). Теорема 1 доказана.

Следствие 1. *Если функция m раз дифференцируема в точке x (на открытом множестве U), то смешанные частные производные m-го порядка в точке x (на U) не зависят от порядка дифференцирования.*

Следствие 2. *Если все частные производные m-го порядка непрерывны в точке x (на открытом множестве U), то в этой точке x (на U) они не зависят от порядка дифференцирования.*

2.9 Формула Тейлора

Теорема (Формула Тейлора с остаточным членом в форме Лагранжа). *Пусть функция $f : U \to \mathbb{R}$, где U — открытое множество в $\mathbb{R}^n$. Пусть отрезок с концами x, $x+dx$ лежит в U и f является $(k+1)$ раз дифференцируемой на этом отрезке. Тогда существует число $\xi \in (0;1)$ такое, что справедливо равенство*

$$f(x+dx) = f(x) + df(x) + \frac{1}{2!}\, d^2 f(x) + \cdots$$

$$\cdots + \frac{1}{k!}\, d^k f(x) + \frac{1}{(k+1)!}\, d^{k+1} f(x+\xi dx). \tag{2.24}$$

Доказательство. Рассмотрим функцию

$$g(t) = f(x + t\,dx), \quad 0 \le t \le 1.$$

Тогда функция g дифференцируема $(k+1)$ раз на $[0;1]$ и к ней применима формула Тейлора для функций одной переменной.

Имеем $g(0) = f(x)$,

$$g'(t) = \sum_{j=1}^{n} \frac{\partial f(x+tdx)}{\partial x_j} \frac{d(x_j + tdx_j)}{dt} = \sum_{j=1}^{n} \frac{\partial f(x+tdx)}{\partial x_j}\,dx_j = df(x+tdx).$$

Аналогично

$$g''(t) = \sum_{l=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f(x+tdx)}{\partial x_j \partial x_l}\,dx_j\,dx_l = d^2 f(x+tdx).$$

По индукции доказываем, что для любого m, $1 \le m \le k+1$, справедливо равенство

$$g^{(m)}(t) = d^m f(x+tdx).$$

Теперь запишем формулу Тейлора для функции g: для любого $t \in [0;1]$ существует $\xi = \xi_t \in (0;t)$ такое, что

$$g(t) = g(0) + g'(0)t + \frac{g''(0)}{2!}\,t^2 + \cdots + \frac{g^{(k)}(0)}{k!}\,t^k + \frac{g^{(k+1)}(\xi)}{(k+1)!}\,t^{k+1}.$$

Подставляя в последнее равенство $t = 1$, получаем (2.24). Теорема доказана.

Теорема (формула Тейлора с остаточным членом в форме Пеано). *Пусть функция $f : U \to \mathbb{R}$ является k раз непрерывно дифференцируемой в точке x. Тогда*

$$f(x+dx) = \sum_{j=0}^{k} \frac{1}{j!}\,d^j f(x) + o(\|dx\|^k), \quad \|dx\| \to 0.$$

Доказательство. Так как функция k раз непрерывно дифференцируемой в точке x, то можно применить предыдущую теорему. Пусть $r > 0$ таково, что $B_r(x) \subset U$. При $\|dx\| < r$ получаем

$$f(x+dx) = f(x) + df(x) + \frac{1}{2!}\,d^2 f(x) + \ldots + \frac{1}{(k-1)!}\,d^{k-1} f(x) + \frac{1}{k!}\,d^k f(x+\xi dx).$$

Осталось установить, что

$$d^k f(x+\xi dx) = d^k f(x) + o(\|dx\|^k), \quad \|dx\| \to 0.$$

Имеем

$$\frac{|d^k f(x+\xi dx) - d^k f(x)|}{\|dx\|^k} =$$

$$= \frac{1}{\|dx\|^k} \left| \sum_{i_1,i_2,\ldots,i_k=1}^{n} \left(\frac{\partial^k f(x+\xi dx)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} - \frac{\partial^k f(x)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} \right) dx_{i_1} dx_{i_2} \cdots dx_{i_k} \right| \leq$$

$$\leq \sum_{i_1,i_2,\ldots,i_k=1}^{n} \left| \frac{\partial^k f(x+\xi dx)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} - \frac{\partial^k f(x)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} \right| \frac{|dx_{i_1} dx_{i_2} \cdots dx_{i_k}|}{\|dx\|^k} \leq$$

$$\leq \sum_{i_1,i_2,\ldots,i_k=1}^{n} \left| \frac{\partial^k f(x+\xi dx)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} - \frac{\partial^k f(x)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} \right|.$$

Если $dx \to 0$, то в силу ограниченности ξ, $x+\xi dx \to x$. Так как частные производные k-го порядка непрерывны в точке x, то

$$\frac{\partial^k f(x+\xi dx)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} \to \frac{\partial^k f(x)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}}, \quad dx \to 0,$$

и из последнего неравенства следует, что

$$\lim_{dx \to 0} \frac{d^k f(x+\xi dx) - d^k f(x)}{\|dx\|^k} = 0.$$

Это завершает доказательство теоремы.

2.10 Производная по направлению. Градиент функции.

Пусть функция $f : U \to \mathbb{R}$, где U — открытое множество в $\mathbb{R}^n$. Рассмотрим $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ — некоторый вектор единичной длины в $\mathbb{R}^n$. Этот вектор задает некоторое направление в $\mathbb{R}^n$. Если существует предел

$$\lim_{t \to 0+} \frac{f(x+t\omega) - f(x)}{t},$$

то говорят, что существует производная $\frac{\partial f(x)}{\partial \omega}$ функции f по направлению ω в точке x и значение это производной есть

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

Если ω совпадает с координатным вектором e_k, то

$$\frac{\partial f(x)}{\partial \omega} = \frac{\partial f(x)}{\partial x_k}, \quad 1 \le k \le n.$$

Теорема. *Если функция f дифференцируема в точке x, то для любого единичного вектора $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ существует производная по направлению ω и*

$$\frac{\partial f(x)}{\partial \omega} = \sum_{j=1}^{n} \frac{\partial f(x)}{\partial x_j}\, \omega_j. \tag{2.25}$$

Доказательство. Так как функция f дифференцируема в точке x, то

$$f(x+h) - f(x) = \sum_{j=1}^{n} \frac{\partial f(x)}{\partial x_j}\, h_j + o(\|h\|), \quad \|h\| \to 0.$$

Следовательно,

$$f(x+t\omega) - f(x) = \sum_{j=1}^{n} \frac{\partial f(x)}{\partial x_j}\, t\omega_j + o(t), \quad t \to 0,$$

так как $\|t\omega\| = |t|$. Деля обе части последнего равенства на t и устремляя t к нулю, получаем утверждение теоремы.

Пусть в точке x существуют все частные производные первого порядка функции f. Градиентом функции f в точке x называется вектор

$$\nabla f(x) := \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

Наряду с обозначением $\nabla f(x)$ широко используется другое обозначение градиента — $\operatorname{grad} f(x)$.

Соотношение (2.25) можно записать с помощью операции скалярного произведения в виде

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

Напомним классическое неравенство Коши-Буняковского в $\mathbb{R}^n$:

$$|\langle a, b \rangle| \leq \|a\| \cdot \|b\|, \quad a, b \in \mathbb{R}^n,$$

причем равенство достигается тогда и только тогда, когда вектора a и b линейно зависимы. С использованием этого неравенства получаем следующий результат.

Теорема. *Если функция f дифференцируема в точке x, то для любого единичного вектора $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ производная по направлению ω удовлетворяет неравенству*

$$\left| \frac{\partial f(x)}{\partial \omega} \right| \leq \|\nabla f(x)\|.$$

Если $\nabla f(x) \neq \theta$, то равенство

$$\frac{\partial f(x)}{\partial \omega} = \|\nabla f(x)\|$$

имеет место тогда и только тогда, когда

$$\omega = \frac{\nabla f(x)}{\|\nabla f(x)\|}.$$

Пример. Пусть $z = z(x, y) = \sqrt{1 - x^2 - y^2}$, $x^2 + y^2 < 1$. В любой точке (x, y) из области определения

$$\nabla z(x, y) = -\left(\frac{x}{\sqrt{1 - x^2 - y^2}}, \frac{y}{\sqrt{1 - x^2 - y^2}} \right).$$

Если $\omega = (-\sqrt{2}/2, \sqrt{2}/2)$, то

$$\frac{\partial z(x, y)}{\partial \omega} = \frac{\sqrt{2}\,(x - y)}{2\sqrt{1 - x^2 - y^2}}$$

2.11 Экстремум функций нескольких переменных

Пусть функция f определена на множестве $G \subset \mathbb{R}^n$. Точка x_0 называется *точкой локального максимума (минимума)* функции f, если существует $\varepsilon > 0$ такое, что для любого $x \in B_\varepsilon(x_0) \cap G$ выполняется неравенство $f(x) \le f(x_0)$ $(f(x) \ge f(x_0))$. Если при этом при некотором $\varepsilon > 0$ в $x \in B_\varepsilon(x_0) \cap G \setminus \{x_0\}$ выполняется строгое неравенство $f(x) < f(x_0)$ $(f(x) > f(x_0))$, то точка x_0 называется *точкой строгого локального максимума (минимума)* функции f. Точки локального максимума (минимума) функции f называются *точками локального экстремума* f, точки строгого локального максимума (минимума) — *точками строгого локального экстремума* f.

Теорема (необходимое условие локального экстремума). *Пусть $f : U \to \mathbb{R}$, множество U открыто в $\mathbb{R}^n$ и x_0 — точка локального экстремума f. Если в точке x_0 существуют все частные производные первого порядка функции f, то*

$$\frac{\partial f(x_0)}{\partial x_1} = \frac{\partial f(x_0)}{\partial x_2} = \cdots = \frac{\partial f(x_0)}{\partial x_n} = 0.$$

Доказательство. Пусть $x_0 = (x_1^0, x_2^0, \ldots, x_n^0)$. Рассмотрим для любого j, $1 \le j \le n$, функцию $g(t) := f(x_1^0, x_2^0, \ldots, x_{j-1}^0, t, x_{j-1}^0, \ldots, x_n^0)$. Если x_0 — точка локального экстремума f, то функция g имеет локальный экстремум в точке $t = x_j^0$. В силу теоремы о необходимом условии экстремума для функций одной переменной имеем $g'(x_j^0) = 0$. Но $g'(x_j^0) = \frac{\partial f(x_0)}{\partial x_j}$. Теорема доказана.

Следствие. *Пусть f дифференцируема в точке $x_0 \in U$. Если x_0 — точка локального экстремума f, то $\nabla f(x_0) = \theta$ и $df(x_0) \equiv 0$.*

Квадратичной формой от переменных x_1, $x_2, \ldots, x_n$ называется выражение $q(x) = q(x_1, x_2, \ldots, x_n) = \sum_{i,j=1}^n a_{ij} x_i x_j$, где a_{ij} — некоторые константы. Можно всегда считать, что квадратичная форма симметрична, т. е. $a_{ij} = a_{ji}$ для всех i и j.

Например, квадратичную форму от двух переменных x и y можно записать в виде $Ax^2 + 2Bxy + Cy^2$, где A, B и C — константы.

Если функция дважды дифференцируема в точке локального экстремума x_0, то в силу следствия, приведенного выше, $df(x_0) \equiv 0$. Применяя формулу Тейлора, получаем

$$f(x) - f(x_0) = \frac{1}{2}\, d^2 f(x_0) + o(\|x - x_0\|^2), \quad x \to x_0. \tag{2.26}$$

Таким образом, если в окрестности точки x_0 правая часть (2.26) является неотрицательной или неположительной, то точка x_0 является точкой локального минимума (максимума). Эта правая часть состоит из двух слагаемых, одно из которых с точностью до множителя $\frac{1}{2}$ совпадает со вторым дифференциалом $d^2 f(x_0)$. Во многих случаях второе слагаемое $o(\|x - x_0\|^2)$ не оказывает влияния на знак суммы, поэтому важной задачей является определение знака $d^2 f(x_0)$. Поскольку второй дифференциал $d^2 f(x_0)$ является квадратичной формой относительно переменных $dx_k = \Delta x_k$, то следует привести известные из алгебры факты о знакопостоянных квадратичных формах.

Квадратичная форма $q(x) = q(x_1, x_2, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j$ называется *положительно определенной* (пишут $q \geq 0$), если для любого $x \in \mathbb{R}^n$ имеет место неравенство $q(x) \geq 0$.

Пример. Квадратичная форма

$$q(x_1, x_2) = x_1^2 + 2x_1 x_2 + x_2^2 = (x_1 + x_2)^2 \geq 0$$

является положительно определенной.

Квадратичная форма $q(x) = q(x_1, x_2, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j$ называется *строго положительно определенной* (пишут $q > 0$), если для любого $x \in \mathbb{R}^n$, $x \neq \theta$, имеет место неравенство $q(x) > 0$.

Квадратичная форма $q(x)$ называется *(строго) отрицательно определенной*, если форма $(-q(x))$ является (строго) положительно определенной. Отрицательную определенность квадратичной формы кратко записывают в виде $q \leq 0$, строгую отрицательную определенность — в виде $q < 0$.

Пример. Квадратичная форма $q(x) = (x_1+x_2+\dots+x_n)^2$ не является строго положительно определенной при $n \geq 2$, так как обращается в нуль во всех точках, где $\sum_{k=1}^n x_k = 0$.

В курсе алгебры доказывается

Теорема (критерий Сильвестра). *Квадратичная форма*

$$q(x) = \sum_{i,j=1}^n a_{ij} x_i x_j$$

является строго положительно определенной тогда и только тогда, когда выполняются неравенства

$$\Delta_1 := a_{11} > 0, \quad \Delta_2 := \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \Delta_3 := \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0, \dots$$

$$\dots, \quad \Delta_n := \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} > 0.$$

Определители Δ_j, $1 \leq j \leq n$, называются *главными минорами* матрицы

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

составленной из коэффициентов квадратичной формы.

Следствие. *Квадратичная форма*

$$q(x) = \sum_{i,j=1}^n a_{ij} x_i x_j$$

является строго отрицательно определенной тогда и только тогда, когда выполняются неравенства

$$\Delta_1 < 0, \quad \Delta_2 > 0, \quad \Delta_3 < 0, \quad \dots , \quad (-1)^n \Delta_n > 0,$$

где Δ_n — главные миноры матрицы квадратичной формы q.

Отметим также, что выполнение нестрогих неравенств

$$\Delta_j \geq 0, 1 \leq j \leq n,$$

являются необходимым и достаточным условием положительной определенности q, а нестрогих неравенств

$$(-1)^j \Delta_j \geq 0, 1 \leq j \leq n,$$

для отрицательной определенности.

Пример. Пусть $q(x,y) = Ax^2 + 2Bxy + Cy^2$ — квадратичная форма от двух переменных. Ее матрица есть

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix}.$$

Главные миноры

$$\Delta_1 = A, \quad \Delta_2 = \begin{vmatrix} A & B \\ B & C \end{vmatrix} = AC - B^2.$$

Таким образом, q строго положительно определена тогда и только тогда, когда $A > 0$ и $AC - B^2 > 0$; q строго отрицательно определена тогда и только тогда, когда $A < 0$ и $AC - B^2 > 0$.

Теорема (необходимое условие экстремума в терминах второго дифференциала). *Пусть функция f дважды непрерывно дифференцируема в точке x_0. Если x_0 — точка локального минимума f, то $df(x_0) = 0$ и $d^2 f(x_0) \geq 0$, если локального максимума, то $df(x_0) = 0$ и $d^2 f(x_0) \leq 0$.*

Доказательство. Рассмотрим для определенности случай, когда x_0 — точка локального минимума f. Для точек локального максимума достаточно рассмотреть вместо f функцию $(-f)$. Из условий теоремы следует, что для f справедлива формула Тейлора

$$f(x) = f(x_0) + df(x_0) + \frac{1}{2}\, d^2 f(x_0) + o(\|dx\|^2), \quad x \to x_0,$$

где $dx = x - x_0$. Поскольку в точке выполняется необходимое условие экстремума $df(x_0) = 0$, то в некоторой окрестности точки x_0

$$\frac{1}{2}\,d^2f(x_0) + o(\|dx\|^2) = f(x) - f(x_0) \geq 0.$$

Перепишем последнее неравенство в виде

$$\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}\,dx_i dx_j + o\left(\sum_{i=1}^{n} dx_i^2\right) \geq 0.$$

Фиксируем некоторый ненулевой вектор $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ и положим $dx = t\omega$, где t — достаточно малое число. Тогда

$$dx_i = t\omega_i, \quad \sum_{i=1}^{n} dx_i^2 = t^2$$

и

$$\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}\,t\omega_i t\omega_j + o(t^2) \geq 0,$$

откуда

$$\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}\,\omega_i \omega_j + \frac{o(t^2)}{t^2} \geq 0.$$

Устремляя t к нулю, получаем

$$\sum_{i,j=1}^{n}\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}\,\omega_i \omega_j \geq 0.$$

Это верно для любого набора $(\omega_1, \omega_2, \ldots, \omega_n) \neq \theta$. Заметим, что последнее неравенство верно также когда все $\omega_j = 0$, т. е. для любых действительных чисел ω_1, $\omega_2, \ldots, \omega_n$. Беря в качестве ω_j произвольные приращения переменных $\Delta x_j = dx_j$, получаем

$$d^2f(x_0) = \sum_{i,j=1}^{n}\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}\,dx_i dx_j \geq 0$$

для любых dx_j. Таким образом, $d^2f(x_0) \geq 0$ и теорема доказана.

Прежде, чем устанавливать достаточное условие локального экстремума в терминах второго дифференциала, установим лемму.

Лемма. *Пусть квадратичная форма* $q(x) = \sum_{i,j=1}^n a_{ij}x_ix_j$ *строго положительно определена. Тогда существует* $m > 0$ *такое, что для любого* $x \in \mathbb{R}^n$ *имеет место неравенство* $q(x) \geq m\|x\|^2$.

Доказательство. Неравенство $q(x) \geq m\|x\|^2$ при $x = \theta$ очевидно справедливо для любого m, а при $x \neq \theta$ эквивалентно неравенству

$$\sum_{i,j=1}^n a_{ij}\frac{x_i}{\|x\|}\frac{x_j}{\|x\|} \geq m$$

или

$$q(y) = \sum_{i,j=1}^n a_{ij}y_iy_j \geq m, \tag{2.27}$$

где $y = (y_1, y_2, \ldots, y_n)$, $y_i = \frac{x_i}{\|x\|}$. Нетрудно видеть, что $y = \frac{x}{\|x\|}$ лежит на единичной сфере $S = \{z \in \mathbb{R}^n \mid \|z\| = 1\}$. Так как S ограничено и замкнуто в $\mathbb{R}^n$, то S компактно. Функция q непрерывна на S. По теореме Вейерштрасса она принимает минимальное значение на S некоторой точке z_0. Пусть $q(z_0) = m$. Так как $z_0 \in S$, то $z_0 \neq \theta$, поэтому в силу строгой положительной определенности квадратичной формы q имеем $m = q(z_0) > 0$. Наконец, для любого $y \in S$ имеем $q(y) \geq q(z_0) = m$ и (2.27) доказано, поэтому доказана и лемма.

Теорема (необходимое условие экстремума в терминах второго дифференциала). *Пусть функция* f *дважды непрерывно дифференцируема в точке* x_0 *и* $df(x_0) = 0$. *Если* $d^2f(x_0) > 0$, *то* x_0 *— точка строгого локального минимума функции* f, *а если* $d^2f(x_0) < 0$, *то строгого локального максимума.*

Доказательство. Рассмотрим для определенности случай, когда второй дифференциал $d^2f(x_0) > 0$. Как и при доказательстве предыдущей теоремы запишем с учетом того, что $df(x_0) = 0$, формулу Тейлора

$$f(x) = f(x_0) + \frac{1}{2}d^2f(x_0) + r(x), \quad x \to 0,$$

где $r(x) = o(\|dx\|^2)$, $dx = x - x_0 \to 0$. В силу леммы существует константа $m > 0$ такая, что для любого $x \in \mathbb{R}^n$

$$d^2f(x_0) \geq m\|dx\|^2.$$

Так как
$$\frac{r(x)}{\|dx\|^2} \to 0, \quad \|dx\| \to 0,$$
то существует достаточно малая окрестность $B_r(x_0)$, в которой выполняется неравенство
$$\frac{r(x)}{\|dx\|^2} < \frac{m}{4}, \quad \text{т. е.} \quad |r(x)| < \frac{m}{4}\|dx\|^2.$$
Если $x \in B_r(x_0)$, $x \neq x_0$, то
$$f(x) - f(x_0) = \frac{1}{2}\,d^2 f(x_0) + r(x) \geq \frac{m}{2}\|dx\|^2 - |r(x)| \geq$$
$$\geq \frac{m}{2}\|dx\|^2 - \frac{m}{4}\|dx\|^2 = \frac{m}{4}\|dx\|^2 = \frac{m}{4}\|x - x_0\|^2 > 0.$$
Это доказывает, что x_0 — точка строгого локального минимума. Теорема доказана.

3 Дифференцирование векторнозначных функций

3.1 Линейные отображения конечномерных пространств и их нормы

Отображение $A : \mathbb{R}^n \to \mathbb{R}^m$ называется *линейным*, если для любых точек x, $y \in \mathbb{R}^n$ и любых скаляров λ, $\mu \in \mathbb{R}$ имеет место равенство

$$A(\lambda x + \mu y) = \lambda Ax + \mu Ay.$$

Как принято в линейной алгебре, мы пишем Ax вместо $A(x)$, опуская скобки. Как правило, это не вызывает недоразумений.

Каждое линейное отображение $A : \mathbb{R}^n \to \mathbb{R}^m$ можно представить в виде

$$Ax = (A_1 x, A_2 x, \ldots, A_m x), \tag{3.1}$$

где $A : \mathbb{R}^n \to \mathbb{R}$ — линейные функции.

Выберем стандартный базис $\{e_j, 1 \leq j \leq n\}$ в $\mathbb{R}^n$. Любой элемент x однозначно представим в виде $x = \sum_{k=1}^n x_k e_k$, где $x_k \in \mathbb{R}$, $1 \leq k \leq n$. Тогда $A_i x = \sum_{k=1}^n x_k A_i e_k$ или

$$A_i x = \sum_{k=1}^{n} a_{ik} x_k, \tag{3.2}$$

где $a_{ik} = A_i e_k$. Отображению A можно сопоставить матрицу

$$[A] := \|a_{ik}\|_{1 \leq i \leq m, 1 \leq k \leq n}.$$

Обратно, любой матрице $\|a_{ik}\|$ размера $(m \times n)$ можно сопоставить линейное отображение, определяемой формулами (3.1) и (3.2).

Для любого линейного отображения $A : \mathbb{R}^n \to \mathbb{R}^m$ обозначим через $\|A\|$ величину

$$\|A\| := \sup_{\|x\|=1} \|Ax\|. \tag{3.3}$$

Заметим, что норма $\|Ax\|$ берется в пространстве $\mathbb{R}^m$, а норма $\|x\|$ — в пространстве $\mathbb{R}^n$, то есть при $m \neq n$ эти нормы различны, но мы для простоты обозначений не будем указывать дополнительно, в каком пространстве берется та или иная норма, это ясно из контекста.

Так как любое линейное отображение непрерывно и операция взятия нормы $y \mapsto \|y\|$ также непрерывна, то отображение $x \mapsto \|Ax\|$ непрерывно как суперпозиция непрерывных отображений. Кроме того, единичная сфера $S = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ является компактным множеством в $\mathbb{R}^n$. По теореме Вейерштрасса величина, определяемая в (3.3), является конечной и знак sup можно заменить на max.

Утверждение. Имеет место равенство

$$\|A\| = \sup_{x \neq \theta} \frac{\|Ax\|}{\|x\|},$$

т. е. норма $\|A\|$ — это наименьшая из констант α, для которых неравенство $\|Ax\| \leq \alpha\|x\|$ имеет место для любого $x \in \mathbb{R}^n$.

Доказательство. Если $x \in \mathbb{R}^n$, $x \neq \theta$, то $y = \frac{1}{\|x\|}x \in S$, поэтому имеем $\|Ay\| \leq \|A\|$ в силу определения $\|A\|$. Тогда

$$\|Ay\| = \left\|A\left(\frac{1}{\|x\|}\,x\right)\right\| = \left\|\frac{1}{\|x\|}Ax\right\| = \frac{1}{\|x\|}\|Ax\| \leq \|A\|,$$

откуда

$$\frac{\|Ax\|}{\|x\|} \leq \|A\|. \tag{3.4}$$

Пусть

$$\alpha_0 = \sup_{x \neq \theta} \frac{\|Ax\|}{\|x\|}.$$

Из (3.4) получаем $\alpha_0 \leq \|A\|$.

Докажем обратное неравенство. Если $x \in S$, то

$$\|Ax\| = \frac{\|Ax\|}{\|x\|} \le \alpha_0,$$

поэтому $\|A\| = \sup_{x\in S} \|Ax\| \le \alpha_0$. Итак, $\|A\| = \alpha_0$ и утверждение доказано.

Следствие. *Для любого* $x \in \mathbb{R}^n$ *имеем* $\|Ax\| \le \|A\| \, \|x\|$.

Обозначим через $L(\mathbb{R}^n, \mathbb{R}^m)$ *пространство всех линейных отображений* $A : \mathbb{R}^n \to \mathbb{R}^m$. Это пространство является линейном векторным пространством над $\mathbb{R}$ с операциями сложения и умножения на скаляр, определяемыми следующим образом:

1) *Суммой отображений* A и B называется отображение $A+B$ такое, что $(A+B)x := Ax + Bx$.

2) *Произведением отображения* A *на скаляр* λ называется отображение λA такое, что $(\lambda A)x := \lambda(Ax)$.

Упражнение. Докажите, что операция $A \mapsto \|A\|$ задает норму на пространстве $L(\mathbb{R}^n, \mathbb{R}^m)$, т. е.

1) $\|A\| \ge 0$, $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, причем $\|A\| = 0 \iff A = \Theta$ — нулевой оператор.

2) $\|\lambda A\| = |\lambda| \, \|A\|$, $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, $\lambda \in \mathbb{R}$.

3) $\|A + B\| \le \|A\| + \|B\|$, A, $B \in L(\mathbb{R}^n, \mathbb{R}^m)$.

Лемма 1. *Пусть* $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, $B \in L(\mathbb{R}^m, \mathbb{R}^k)$, *тогда* $B \circ A \in L(\mathbb{R}^n, \mathbb{R}^k)$ *и* $\|B \circ A\| \le \|B\| \, \|A\|$.

Доказательство. Для любого $x \in \mathbb{R}^n$ имеем

$$\|(B \circ A)x\| = \|B(Ax))\| \le \|B\| \, \|Ax\| \le \|B\| \, \|A\| \, \|x\|.$$

Если $x \ne \theta$, то

$$\frac{\|(B \circ A)x\|}{\|x\|} \le \|B\| \, \|A\|,$$

откуда

$$\|B \circ A\| = \sup_{x \ne \theta} \frac{\|(B \circ A)x\|}{\|x\|} \le \|B\| \, \|A\|.$$

Лемма 1 доказана.

Если $m = n$, то вместо $L(\mathbb{R}^n, \mathbb{R}^n)$ будем писать $L(\mathbb{R}^n)$. В курсе линейной алгебры доказывается, что линейное отображение $A \in L(\mathbb{R}^n)$ является биекцией тогда и только тогда, когда A инъективно. Последнее эквивалентно условию $\det[A] \neq 0$, где $[A]$ — матрица линейного отображения A, $\det[A]$ — ее определитель. Если $\det[A] \neq 0$, то существует обратное отображение $A^{-1} \in L(\mathbb{R}^n)$. Отображение A в этом случае называется обратимым.

Лемма 2. *Пусть $A \in L(\mathbb{R}^n)$ – обратимое линейное отображение. Тогда существует $m > 0$ такое, что для любого $x \in \mathbb{R}^n$ выполняется неравенство*

$$\|Ax\| \geq m\|x\|.$$

Доказательство. Пусть $x \in \mathbb{R}^n$, $y = Ax$. Тогда $x = A^{-1}y$ и

$$\|x\| = \|A^{-1}y\| \leq \|A^{-1}\| \, \|y\| = \|A^{-1}\| \, \|Ax\|,$$

откуда $\|Ax\| \geq m\|x\|$, где $m = 1/\|A\|$. Лемма 2 доказана.

3.2 Дифференцируемость вектор-функций

Пусть $f : U \to \mathbb{R}^m$, где U — открытое множество в $\mathbb{R}^n$. Пусть $a \in U$. Так как U открыто, то существует окрестность $B_\varepsilon(a) \subset U$. Следовательно, если $\|h\| < \varepsilon$, то $a + h \in U$.

Отображение f называется *дифференцируемым в точке* $a \in U$, если существует линейное отображение $A \in L(\mathbb{R}^n, \mathbb{R}^n)$ такое, что

$$f(a+h) - f(a) = Ah + r(h), \quad \text{где} \quad r(h) = o(\|h\|), \ h \to \theta, \tag{3.5}$$

т. е.

$$\lim_{h \to \theta} \frac{\|r(h)\|}{\|h\|} = 0.$$

Если имеет место (3.5), то отображение A называется *производным отображением* в точке a для отображения f и обозначается $f'(a)$.

Условие (3.5) можно записать в виде

$$f(a+h)-f(a)=f'(a)h+o(h), \quad h\to\theta,$$

или

$$f(a+dx)-f(a)=f'(a)dx+o(dx), \quad dx\to\theta.$$

Справедливо

Утверждение. *Любое линейное отображение* $A\in L(\mathbb{R}^n,\mathbb{R}^n)$ *дифференцируемо в любой точке* $a\in\mathbb{R}^n$ *и* $A'(a)=A$.

Доказательство. Имеем $A(a+h)-Aa=Ah=Ah+\theta=Ah+o(h)$, $\|h\|\to 0$, так как $\theta=o(h)$, $\|h\|\to 0$.

3.3 Необходимое и достаточное условия дифференцируемости вектор-функции. Матрица Якоби.

Пусть $f:U\to\mathbb{R}^m$, где U — открытое множество в $\mathbb{R}^n$. Тогда запишем $f=(f_1,f_2,\ldots,f_m)$, где $f_i:U\to\mathbb{R}$, $1\le i\le n$. Пусть $A\in L(\mathbb{R}^n,\mathbb{R}^m)$, тогда $A=(A_1,A_2,\ldots,A_m)$, где $A_i\in L(\mathbb{R}^n,\mathbb{R})$, $1\le i\le n$.

Теорема (критерий дифференцируемости). *Отображение* f *дифференцируемо в точке* $a\in U$ *тогда и только тогда, когда для любого* i, $1\le i\le n$, *функции* f_i *дифференцируемы в точке* a. *При этом,*

$$f'(a)=(f_1'(a),f_2'(a),\ldots,f_m'(a)).$$

Доказательство. Пусть f дифференцируемо в точке a. Тогда

$$f(a+h)-f(a)=Ah+r(h), \quad \text{где} \quad r(h)=o(h),\ h\to\theta. \tag{3.6}$$

При этом $A=f'(a)\in L(\mathbb{R}^n,\mathbb{R}^m)$. Равенство (3.6) — это равенство векторов в $\mathbb{R}^m$. Запишем его для i-й компоненты:

$$f_i(a+h)-f_i(a)=A_ih+r_i(h), \tag{3.7}$$

где $(A_1, A_2, \ldots, A_m) = A$, $(r_1(h), r_2(h), \ldots, r_m(h)) = r(h)$. Ясно, что линейные отображения $A_i \in L(\mathbb{R}^n, \mathbb{R})$. Кроме того, $|r_i(h)| \leq \|r(h)\|$, откуда $r_i(h) = o(\|h\|)$, $h \to \theta$. Из (3.7) тогда следует, что каждое f_i дифференцируемо в точке a и $f_i'(a) = A_i$.

Обратно, пусть имеют место условия (3.7), где $A_i \in L(\mathbb{R}^n, \mathbb{R})$, $r_i(h) = o(\|h\|)$, $h \to \theta$. Тогда имеет место (3.6), где

$$A = (A_1, A_2, \ldots, A_m), \quad r(h) = (r_1(h), r_2(h), \ldots, r_m(h)).$$

Ясно, что $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, $r(h) = o(h)$, $h \to \theta$. Следовательно, f дифференцируема в точке a и $f'(a) = A$. Теорема доказана.

Следствие. *Если отображение f дифференцируемо в точке a, то оно непрерывно в точке a.*

Пусть $f(x) = (f_1(x), f_2(x), \ldots, f_m(x))$, $x = (x_1, x_2, \ldots, x_n) \in U$. Пусть в точке a существуют все частные производные $\frac{\partial f_i}{\partial x_j}$. *Матрицей Якоби* $J_f(a)$ отображения f в точке a называется матрица

$$J_f(a) := \begin{pmatrix} \dfrac{\partial f_1(a)}{\partial x_1} & \dfrac{\partial f_1(a)}{\partial x_2} & \ldots & \dfrac{\partial f_1(a)}{\partial x_n} \\ \dfrac{\partial f_2(a)}{\partial x_1} & \dfrac{\partial f_2(a)}{\partial x_2} & \ldots & \dfrac{\partial f_2(a)}{\partial x_n} \\ \ldots & \ldots & \ldots & \ldots \\ \dfrac{\partial f_m(a)}{\partial x_1} & \dfrac{\partial f_m(a)}{\partial x_2} & \ldots & \dfrac{\partial f_m(a)}{\partial x_n} \end{pmatrix} = \left(\frac{\partial f_i(a)}{\partial x_j}\right)_{1 \leq i \leq m, 1 \leq j \leq n},$$

составленная из всех частных производных первого порядка всех компонент отображения f.

Теорема 2. *Пусть отображение*

$$f(x) = (f_1(x), f_2(x), \ldots, f_m(x)), \quad x = (x_1, x_2, \ldots, x_n) \in U,$$

дифференцируемо в точке $a \in U$. Тогда в точке a существуют все частные производные $\frac{\partial f_i}{\partial x_j}$, $1 \leq i \leq m$, $1 \leq j \leq n$, и матрица $[f'(a)]$ производного отображения $f'(a)$ совпадает с матрицей Якоби отображения f в точке a, т. е. $[f'(a)] = J_f(a)$.

Доказательство. Пусть отображение f дифференцируемо в точке a. Тогда по теореме 1 все компоненты f_i дифференцируемы в точке a и

$$f_i(a+h) - f_i(a) = A_i h + r_i(h),$$

где $r_i(h) = o(\|h\|)$, $\|h\| \to 0$, а $A_i \in L(\mathbb{R}^n, \mathbb{R})$. В силу линейности A_i существуют константы a_{i1}, $a_{i2}, \ldots, a_{in}$ такие, что $A_i h = \sum_{j=1}^{n} a_{ij} h_j$. Таким образом,

$$f_i(a+h) - f_i(a) = \sum_{j=1}^{n} a_{ij} h_j + r_i(h), \tag{3.8}$$

поэтому функция f_i дифференцируема как скалярная функция нескольких переменных. По теореме о необходимом условии дифференцируемости таких функций

$$a_{ij} = \frac{\partial f_i(a)}{\partial x_j}, \quad 1 \le j \le n.$$

Наконец, из скалярных равенств (3.8) следует векторное равенство (3.6), где линейное отображение $A = f'(a)$ имеет матрицу, составленную из элементов a_{ij}, т. е. $\frac{\partial f_i(a)}{\partial x_j}$. Таким образом, $[f'(a)] = J_f(a)$ и теорема доказана.

Теорема 3. *Если в некоторой окрестности точка a существуют все частные производные $\frac{\partial f_i(x)}{\partial x_j}$ и эти производные непрерывны в точке a, то отображение f дифференцируемо в точке a.*

Доказательство. Из условий теоремы и достаточного условия дифференцируемости скалярных функций следует дифференцируемость компонент f_i, $1 \le i \le m$. Остается применить теорему 1.

Отображение $f = (f_1, f_2, \ldots, f_m)$ называется *непрерывно дифференцируемым в точке* a, если все компоненты f_i, $1 \le i \le m$, непрерывно дифференцируемы в точке a.

Отображение f называется *дифференцируемым (непрерывно дифференцируемым) на множестве* U, если f дифференцируемо (непрерывно дифференцируемо) в любой точке $a \in U$.

3.4 Дифференциал вектор-функции

Если отображение f дифференцируемо в точке a, то

$$\Delta f(a) := f(a+h) - f(a) = f'(a)h + o(\|h\|), \quad \|h\| \to 0.$$

где $h = \Delta x := dx$ — дифференциал независимой переменной. Выражение $f'(a)h = f'(a)dx$ называется *дифференциалом отображения* f в точке a и обозначается $df(a)$.

Отметим некоторые частные случаи.

1) $m = n = 1$. В этом случае $df(a) = f'(a)dx$, где запись $f'(a)dx$ можно понимать двояко: как произведение двух вещественных чисел $f'(a)$ и dx или как результат применения линейного отображения $f'(a) \in L(\mathbb{R})$ к элементу $dx \in \mathbb{R}$. В первом случае $f'(a)$ понимается как число, во втором — как линейное отображение (умножение на это число).

2) $m = 1$, $n \geq 1$. Тогда $df(a) = \sum_{j=1}^{n} \frac{\partial f(a)}{\partial x_j} dx_j$.

3) $n = 1$, $m \geq 1$. Тогда $df(a) = (f_1'(a), f_2'(a), \ldots, f_m'(a))$.

4) Произвольный случай: $n \geq 1$, $m \geq 1$. Тогда

$$df(a) = \left(\sum_{j=1}^{n} \frac{\partial f_1(a)}{\partial x_j} dx_j, \sum_{j=1}^{n} \frac{\partial f_2(a)}{\partial x_j} dx_j, \ldots, \sum_{j=1}^{n} \frac{\partial f_m(a)}{\partial x_j} dx_j \right).$$

3.5 Производная суперпозиции отображений

Теорема 1. *Пусть $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ — открытые множества, $f : U \to V$, $g : V \to \mathbb{R}^k$. Если отображение f дифференцируемо в точке $a \in U$, а отображение g дифференцируемо в точке $b = f(a) \in V$, то $h = g \circ f$ дифференцируемо в точке a и $h'(a) = g'(b) \circ f'(a)$.*

Доказательство. Из дифференцируемости отображений f и g в точках a и b соответственно следует, что в достаточно малых окрестностях $B_r(a)$ и $B_s(b)$ этих точек

$$f(x) - f(a) = f'(a)(x-a) + \|x-a\|\alpha(x), \quad \text{где} \quad \alpha(x) \to \theta,\ x \to a, \tag{3.9}$$

$$g(y) - g(b) = g'(b)(y-b) + \|y-b\|\beta(y), \quad \text{где} \quad \beta(y) \to \theta,\ y \to b. \quad (3.10)$$

Уменьшая в случае необходимости r. считаем, что $f(B_r(a)) \subset B_s(b)$. Пусть $x \in B_r(a)$, тогда $y = f(x) \in B_s(b)$. Имеем в силу (3.9) и (3.10)

$$h(x)-h(a) = g(f(x))-g(f(a)) = g'(b)(f(x)-f(a))+\|f(x)-f(a)\|\beta(f(x))=$$

$$= g'(b)\big(f'(a)(x-a) + \|f(x)-f(a)\|\alpha(x)\big) + \|f(x)-f(a)\|\beta(f(x)) =$$

$$= \big(g'(b)\circ f'(a)\big)(x-a) + g'(b)\big(\|x-a\|\alpha(x)\big) + \|f(x)-f(a)\|\beta(f(x)). \quad (3.11)$$

Обозначим

$$r_1(x) = g'(b)\big(\|x-a\|\alpha(x)\big), \quad r_2(x) = \|f(x)-f(a)\|\beta(f(x)). \qquad (3.12)$$

Докажем, что

$$r_1(x) = o(\|x-a\|), \quad r_2(x) = o(\|x-a\|), \quad x \to a. \qquad (3.13)$$

Имеем $\|r_1(x)\| \le \|g'(b)\|\,\|x-a\|\,\|\alpha(x)\|$, поэтому

$$0 \le \frac{\|r_1(x)\|}{\|x-a\|} \le \|g'(b)\|\,\|\alpha(x)\|.$$

Так как $\alpha(x) \to \theta,\ x \to a$, то справедливо первое из соотношений (3.12). Докажем второе. Так как $\alpha(x) \to \theta,\ x \to a$, то $\alpha(x)$ ограничено в некоторой окрестности $B_r(a)$ точки a и без ограничения общности можно считать, что для некоторого $C > 0$ в $B_r(a)$ выполняется неравенство $\|\alpha(x)\| \le C$. Тогда $\forall x \in B_r(a)$

$$\|f(x)-f(a)\| = \|f'(a)(x-a) + \|x-a\|\alpha(x)\| \le$$

$$\le \|f'(a)(x-a)\| + \|x-a\|\,\|\alpha(x)\| \le (\|f'(a)\| + C)\,\|x-a\|.$$

Из непрерывности f в точке a следует, что

$$\lim_{x\to a}\beta(f(x)) = \lim_{y\to b}\beta(y) = \theta.$$

С учетом последнего первенства и определения $r_2(x)$ получаем, что справедливо и второе соотношение в (3.13).

Теперь из (3.11), (3.12) и (3.13) следует, что

$$h(x) - h(a) = \big(g'(b) \circ f'(a)\big)(x - a) + o(\|x - a\|), \quad x \to a.$$

Это, по определению, означает, что h дифференцируема в точке a и ее производная $h'(a)$ равна $g'(b) \circ f'(a)$. Теорема доказана.

Теорема 2. *В предположениях теоремы 1 матрица Якоби суперпозиции $g \circ f$ равна произведению матриц Якоби отображения f и отображения g:*

$$J_{g\circ f}(a) = J_g(b) J_f(a).$$

Доказательство. Обозначая квадратными скобками матрицы соответствующих линейных отображений, получаем

$$J_{g\circ f}(a) = J_h(a) = [h'(a)] = [g'(b) \circ f'(a)] = [g'(b)]\,[f'(a)] = J_g(b) J_f(a).$$

При этом мы воспользовались тем фактом, что матрица произведения (суперпозиции) линейных отображений равна произведению матриц этих отображений.

Теорема 3. *В предположениях теоремы 1, если отображение f непрерывно дифференцируемо в точке a, а отображение g непрерывно дифференцируемо в точке b, то $h = g \circ f$ непрерывно дифференцируемо в точке a.*

Доказательство. В силу теоремы 2 имеем в некоторой окрестности точки a $J_h(x) = J_g(y) J_f(x)$, $y = f(x)$, т. е.

$$\left(\frac{\partial h_i(x)}{\partial x_j}\right)_{\substack{1\le i\le k\\ 1\le j\le n}} = \left(\frac{\partial g_i(y)}{\partial y_l}\right)_{\substack{1\le i\le k\\ 1\le l\le n}} \left(\frac{\partial f_l(x)}{\partial x_j}\right)_{\substack{1\le l\le m\\ 1\le j\le n}}.$$

С учетом правила перемножения матриц получаем

$$\frac{\partial h_i(x)}{\partial x_j} = \sum_{l=1}^{m} \frac{\partial g_i(f(x))}{\partial y_l}\,\frac{\partial f_l(x)}{\partial x_j}. \tag{3.14}$$

Отображение f дифференцируемо в точке a, следовательно, непрерывно в точке a. Кроме того, по условию теоремы частные производные $\frac{\partial f_l(x)}{\partial x_j}$

непрерывны в точке a, а $\frac{\partial g_i(f(x))}{\partial y_l}$ — в точке b. Таким образом, частные производные $\frac{\partial h_i(x)}{\partial x_j}$ непрерывны в точке a. Теорема доказана.

Теорема 4. *Пусть $k \in \mathbb{N}$. В предположениях теоремы 1, если отображение f является k раз непрерывно дифференцируемым в точке a, а отображение g — k раз непрерывно дифференцируемым в точке b, то $h = g \circ f$ является k раз непрерывно дифференцируемым в точке a.*

Доказательство. В силу теоремы 3 имеем (3.14). Дифференцируя это соотношение $(k-1)$ убеждаемся по индукции, что частные производные k-го порядка компонент h_i в точке x являются многочленами от частных производных функций g_i в точке $f(x)$ и f_l в точке x порядка не выше k (докажите это строго!). Поскольку все такие производные являются непрерывными функциями и f непрерывно, то и частные производные k-го порядка компонент h_i в точке x являются непрерывными функциями.

3.6 Якобиан отображения

Пусть $f : U \to \mathbb{R}^n$, где U — открытое множество в $\mathbb{R}^n$. Если существует матрица Якоби отображения f в точке a, то она является квадратной матрицей. Ее определитель называется *якобианом* отображения f в точке a и обозначается $I_f(a)$. Итак,

$$I_f(a) := \det J_f(a) = \det[f'(a)].$$

Для якобиана применяется также обозначение

$$\frac{\partial(f_1, f_2, \ldots, f_n)}{\partial(x_1, x_2, \ldots, x_n)}.$$

Теорема 1. *Пусть U, $V \subset \mathbb{R}^n$ — открытые множества, $f : U \to V$, $g : V \to \mathbb{R}^n$. Если отображение f дифференцируемо в точке $a \in U$, а отображение g дифференцируемо в точке $b = f(a) \in V$, то*

$$I_{g \circ f}(a) = I_g(b) I_f(a).$$

Доказательство. Имеем $I_{g\circ f}(a) = \det I_{g\circ f}(a) = \det(J_g(b)J_f(a)) = = \det(J_g(b))\det(J_f(a)) = I_g(b)I_f(a)$.

Теорема 2. *Пусть U, $V \subset \mathbb{R}^n$ — открытые множества, $f : U \to V$ и существует обратное отображение $f^{-1} : V \to \mathbb{R}^n$. Если отображение f дифференцируемо в точке $a \in U$, а f^{-1} дифференцируемо в точке $b = f(a)$, то $f'(a)$ является обратимым линейным отображением в $\mathbb{R}^n$ и*

$$(f'(a))^{-1} = (f^{-1})'(b), \quad (J_f(a))^{-1} = J_{f^{-1}}(b), \quad (I_f(a))^{-1} = I_{f^{-1}}(b).$$

Доказательство. Имеем

$$f^{-1}\circ f(x) = x, \quad x \in U, \quad f\circ f^{-1}(y) = y, \quad y \in V.$$

Применяя теорему о дифференцировании суперпозиции, получаем

$$(f^{-1})'(b)\circ f'(a) = \mathrm{id}_{\mathbb{R}^n}, \quad f'(a)\circ (f^{-1})'(b) = \mathrm{id}_{\mathbb{R}^n}.$$

Из этих соотношений получаем, что линейные отображения $f'(a)$ и $(f^{-1})'(b)$ обратимы и являются взаимно-обратными. Таким образом, $(f'(a))^{-1} = (f^{-1})'(b)$. Переходя к их матрицам и определителям матриц, получаем утверждение теоремы.

3.7 Регулярные отображения

Пусть $U \subset \mathbb{R}^n$ — открытое множество, $f : U \to \mathbb{R}^n$. Отображение f называется *регулярным*, если оно непрерывно дифференцируемо в U и в любой точке x из U якобиан $I_f(x) \neq 0$.

Отметим некоторые свойства регулярных отображений.

1) Любое регулярное отображение f непрерывно. Это следствие дифференцируемости f.

2) Якобиан регулярного отображения f — непрерывная функция. Действительно, якобиан — это определитель матрицы с непрерывными элементами $\frac{\partial f_i(x)}{\partial x_j}$.

3) Если U линейно связно, то либо $I_f > 0$ в U, либо $I_f < 0$ в U. Действительно, если бы это было не так, то существовали бы две точки в U такие, что $I_f(a) > 0$, $I_f(b) < 0$. Тогда по теореме о промежуточном значении существовала бы точка $x \in U$, в которой $I_f(c) = 0$, что противоречит регулярности f.

Докажем локальную инъективность регулярных отображений.

Теорема 1. *Если $f : U \to \mathbb{R}^n$ — регулярное отображение, то для любого $x \in U$ существует окрестность $B_\delta(x) \subset U$, в которой f является инъективным.*

Доказательство. Пусть $x \in U$. Так как отображение f регулярно, то $I_f(x) \neq 0$. В силу непрерывности якобиана существует окрестность $B_\varepsilon(x)$ этой точки, в которой $I_f \neq 0$. Докажем, что при достаточно малых $\delta > 0$ отображение f инъективно на $B_\delta(x)$.

Предположим противное. Тогда для любого $\delta > 0$ существуют две точки a, $b \in B_\delta(x)$ такие, что $f(a) = f(b)$, т. е. $f_i(a) = f_i(b)$, $1 \leq i \leq n$. По формуле конечных приращений, примененной к f_i, получаем

$$0 = f_i(a) - f_i(b) = \sum_{j=1}^{n} \frac{\partial f_i(c_i)}{\partial x_j} (b_j - a_j)$$

для некоторой точки c_i, лежащей на отрезке Δ_{ab} с концами a и b. Здесь a_j, b_j — координаты векторов a и b. Отметим, что поскольку a, $b \in B_\delta(x)$ и шар $B_\delta(x)$ — выпуклое множество, то отрезок $\Delta_{ab} \subset B_\delta(x)$, следовательно, точки $c_i \in B_\delta(x)$.

Теперь положим $\delta = \frac{1}{n}$, $n \in \mathbb{N}$. В силу доказанного в шаре $B_{1/n}(x)$ существуют точки a, b, а также $c_i = c_i^{(n)}$, $1 \leq i \leq n$, такие, что

$$\sum_{j=1}^{n} \frac{\partial f_i(c_i^{(n)})}{\partial x_j} (b_j - a_j) = 0, \quad 1 \leq i \leq n. \tag{3.15}$$

(точки a, b тоже зависят от n, но мы не будем это явно указывать). Ясно, что $c_i^{(n)} \to x$, $n \to \infty$, для любого i.

Посмотрим на (3.15) как на систему n линейных уравнений относительно n неизвестных $b_j - a_j$, $1 \leq j \leq n$. Так как частные производные

$\frac{\partial f_i}{\partial x_j}$ являются непрерывными функциями в U, то при $n \to \infty$

$$\frac{\partial f_i(c_i^{(n)})}{\partial x_j} \to \frac{\partial f_i(x)}{\partial x_j},$$

следовательно, определитель $D^{(n)}$ системы (3.15) стремится к якобиану $I_f(x) \neq 0$. Поэтому при достаточно больших n определитель $D^{(n)} \neq 0$, следовательно, однородная система (3.15) имеет только нулевой решение: $b_j - a_j = 0,\ 1 \leq i \leq n$. Но тогда $a = b$ — противоречие, доказывающее теорему.

Отметим, что регулярные отображения не обязаны быть инъективными.

Пример. Отображение $f : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}^2$, заданное формулой

$$f(x, y) = (x^2 - y^2, 2xy)$$

является регулярным, так как оно непрерывно дифференцируемо и якобиан

$$I_f(x, y) = \begin{vmatrix} 2x & -2y \\ 2y & 2x \end{vmatrix} = 4(x^2 + 4y^2) \neq 0,$$

$(x, y) \in \mathbb{R}^2 \setminus \{0\}$. Однако f не инъективно, так как $f(-x, -y) = f(x, y)$ для любого $(x, y) \in \mathbb{R}^2 \setminus \{0\}$.

Лемма. *Если отображение $f : U \to \mathbb{R}^n$ (U открыто в $\mathbb{R}^n$) дифференцируемо в точке $x \in U$ и $f'(x)$ обратимо, то существуют $\varepsilon > 0$ и $m > 0$ такие, что для любого $x^* \in B_\varepsilon(x)$ выполняется неравенство*

$$\|f(x^*) - f(x)\| \geq m\|x^* - x\|.$$

Доказательство. В силу дифференцируемости отображения f в точке x имеем

$$f(x^*) - f(x) = f'(x)(x^* - x) + r(x^*), \quad r(x^*) = o(\|x^* - x\|), \quad x^* \to x. \quad (3.16)$$

Линейное отображение $f'(x)$ является обратимым, поэтому, как было показано ранее, существует константа $m > 0$ такая, что для любого x^*

$$\|f'(x)(x^* - x)\| \geq 2m\|x^* - x\|. \quad (3.17)$$

Так как
$$\frac{\|r(x^*\|}{\|x^*-x\|}\to 0,\quad x^*\to x,$$
то существует $\varepsilon>0$ такое, что для любого $x^*\in B_\varepsilon(x)$, $x^*\neq x$, выполняется неравенство
$$\frac{\|r(x^*\|}{\|x^*-x\|}<m,$$
откуда выводим, что
$$\|r(x^*\|<m\|x^*-x\|,\quad x^*\in B_\varepsilon(x). \tag{3.18}$$
Из (3.16), (3.17) и (3.18) с использованием неравенства треугольника получаем
$$\|f(x^*)-f(x)\|\geq\|f'(x)(x^*-x)\|-\|r(x^*\|>$$
$$>2m\|x^*-x\|-m\|x^*-x\|=m\|x^*-x\|.$$
Лемма доказана.

Теорема 2. *Регулярное отображение переводит открытые множества в открытые.*

Доказательство. Пусть $f:U\to\mathbb{R}^n$ регулярно и $G\subset U$ — открытое подмножество. Докажем, что $f(G)$ открыто.

Пусть $b\in f(G)$. Покажем, что существует $r>0$ такое, что $B_r(b)\subset G$. Так как $b\in f(G)$, то существует $a\in G$ такое, что $f(a)=b$. Отображение f является регулярным, следовательно, f дифференцируемо в точке a и $I_f(a)=\det[f'(a)]\neq 0$. Таким образом, $f'(a)$ — обратимо. В силу доказанной леммы существует $\delta>0$ такое, что $B_\delta(a)\subset G$ и для любого $x^*\in B_\delta(a)$ выполняется неравенство
$$\|f(x^*)-f(a)\|\geq m\|x^*-a\|.$$
Пусть $\|x-a\|=\delta/2$. Тогда $\|f(x)-f(a)\|\geq m\delta/2=2r$.

Таким образом, на границе шара $B_{\delta/2}(a)$ имеем $\|f(x)-f(a)\|\geq 2r$. Докажем, что справедливо вложение $B_r(b)\subset f(G)$. Для этого рассмотрим функцию $g(x):=\|f(x)-\widetilde{y}\|^2$, где $\widetilde{y}$ — произвольная фиксированная точка из $B_r(b)$. На границе $B_{\delta/2}(a)$ имеем

$$\|f(x)-\widetilde{y}\| \geq \|f(x)-f(a)\|-\|f(a)-\widetilde{y}\| = \|f(x)-f(a)\|-\|b-\widetilde{y}\| \geq 2r-r = r,$$

откуда $g(x) \geq r^2$. При $x = a$ имеем $g(a) = \|f(a) - \widetilde{y}\|^2 = \|b - \widetilde{y}\|^2 < r^2$, так как $\widetilde{y} \in B_r(b)$. Следовательно, минимальное значение функции g в $\overline{B_{\delta/2}(a)}$, которое существует по теореме Вейерштрасса, не может достигаться на границе $\partial B_{\delta/2}(a)$, т. е. существует точка минимума $\widetilde{x} \in B_{\delta/2}(a)$. В силу необходимого условия экстремума имеем

$$\frac{\partial g(\widetilde{x})}{\partial x_j} = 0, \quad 1 \leq j \leq n.$$

Так как $g(x) = \sum_{i=1}^n (f_i(x) - \widetilde{y}_i)^2$, то

$$\frac{\partial g(\widetilde{x})}{\partial x_j} = 2\sum_{i=1}^n \frac{\partial f_i(\widetilde{x})}{\partial x_j}(f_i(x) - \widetilde{y}_i) = 0, \quad 1 \leq i \leq n. \tag{3.19}$$

Якобиан

$$I_f(\widetilde{x}) = \det\left(\frac{\partial f_i(\widetilde{x})}{\partial x_j}\right)_{1 \leq i,j \leq n} \neq 0,$$

поэтому, рассматривая (3.19) как систему линейных уравнений относительно неизвестных $f_i(x) - \widetilde{y}_i$, получаем, что эта система имеет только тривиальное решение, т. е. $f_i(\widetilde{x}) - \widetilde{y}_i = 0$, $1 \leq i \leq n$, Отсюда следует, что $f(\widetilde{x}) = \widetilde{y}$. Итак, мы показали, что любая точка $\widetilde{y}$ из $B_r(b)$ лежит в $f(G)$. Значит любая точка $b \in f(G)$ является внутренней точкой множества $f(G)$, и $f(G)$ открыто. Теорема доказана.

В заключение напомним один критерий непрерывности отображений, заданных на открытом множестве.

Теорема 3. *Пусть $f : U \to \mathbb{R}^m$, где U — открытое множество в $\mathbb{R}^n$. Отображение f непрерывно на U тогда и только тогда, когда для любого открытого множества $V \subset \mathbb{R}^m$ его прообраз $f^{-1}(V)$ открыт в $\mathbb{R}^n$.*

3.8 Диффеоморфизмы

Пусть $f : U \to \mathbb{R}^n$, где U — открытое множество в $\mathbb{R}^n$. Отображение f называется *диффеоморфизмом,* если f регулярно и инъективно на U.

Отметим следующий легко проверяемый факт.

Лемма. *Если $f : U \to \mathbb{R}^n$, где U — открытое множество в $\mathbb{R}^n$, является диффеоморфизмом, то $f^{-1} : V \to U$ — непрерывное отображение.*

Доказательство сразу следует из теорем 2 и 3 предыдущего пункта.

Справедлива

Теорема о диффеоморфизме. *Пусть $f : U \to \mathbb{R}^n$, где U — открытое множество в $\mathbb{R}^n$, является диффеоморфизмом и $V = f(U)$. Тогда обратное отображение $f^{-1} : V \to U$ также является диффеоморфизмом и для любого $x \in U$*

$$(f^{-1})'(y) = (f'(x))^{-1}, \quad \text{где} \quad y = f(x).$$

Более того, если отображение f является k раз непрерывно дифференцируемым, то и f^{-1} является k раз непрерывно дифференцируемым.

Доказательство. Фиксируем точку $x \in U$. В силу леммы из предыдущего пункта существуют константы $\varepsilon > 0$ и $m > 0$ такие, что для любого $\widetilde{x} \in B_\varepsilon(x)$ выполняется неравенство

$$\|f(\widetilde{x}) - f(x)\| \leq m\|\widetilde{x} - x\|. \tag{3.20}$$

Так как f регулярно, то $f(B_\varepsilon(x))$ открыто, поэтому существует шар $B_r(y) \subset f(B_\varepsilon(x))$. Отображение f дифференцируемо в точке x, следовательно,

$$f(\widetilde{x}) - f(x) = A(\widetilde{x} - x) + r(\widetilde{x}), \tag{3.21}$$

где

$$r(\widetilde{x}) = o(\|\widetilde{x} - x\|), \quad \widetilde{x} \to x. \tag{3.22}$$

Здесь $A = f'(x)$ — линейное отображение. Так как отображение f регулярно, $\det[A] = I_f(x) \neq 0$, поэтому A обратимо. Перепишем (3.21) в виде

$$\widetilde{y} - y = A(\widetilde{x} - x) + r(\widetilde{x}), \quad \text{где} \quad \widetilde{y} = f(\widetilde{x}),\ y = f(x).$$

Применим к обеим частям последнего равенства линейный оператор A^{-1}. Получаем

$$A^{-1}(\widetilde{y} - y) = \widetilde{x} - x + A^{-1}r(\widetilde{x}),$$

следовательно, для любого $\widetilde{y} \in B_r(y)$

$$f^{-1}(\widetilde{y}) - f^{-1}(y) = A^{-1}(\widetilde{y} - y) - A^{-1}r(\widetilde{x}). \tag{3.23}$$

Докажем, что

$$A^{-1}r(\widetilde{x}) = o\left(\|\widetilde{y} - y\|\right), \quad \widetilde{y} \to y. \tag{3.24}$$

Так как $\|A^{-1}r(\widetilde{x})\| \leq \|A^{-1}\|\,\|r(\widetilde{x})\|$, то с учетом (3.20) и (3.22) получаем

$$\frac{\|A^{-1}r(\widetilde{x})\|}{\|\widetilde{y} - y\|} \leq \frac{\|A^{-1}r(\widetilde{x})\|}{m\,\|\widetilde{x} - x\|} \leq \frac{\|A^{-1}\|}{m}\,\frac{r(\widetilde{x})}{\|\widetilde{x} - x\|} = \frac{\|A^{-1}\|}{m}\,o(1) = o(1), \tag{3.25}$$

$\widetilde{x} \to x$. Так как при $\widetilde{y} \to y$

$$\widetilde{x} = f^{-1}(\widetilde{y}) \to f^{-1}(y) = x,$$

то из (3.25) следует (3.24).

Теперь из (3.23) и (3.24) получаем

$$f^{-1}(\widetilde{y}) - f^{-1}(y) = A^{-1}(\widetilde{y} - y) + o(\|\widetilde{y} - y\|), \quad \widetilde{y} \to y.$$

Следовательно, обратное отображение $f^{-1} : V \to U$ дифференцируемо в точке y и $(f^{-1})'(y) = A^{-1} = (f'(x))^{-1}$.

Обозначим $g := f^{-1}$. Докажем, что g является непрерывно дифференцируемым, т. е. что частные производные $\frac{\partial g_i}{\partial y_j}$ непрерывны на V. В силу свойств обратной матрицы имеем

$$J_g(y) = J_{f^{-1}}(y) = (J_f(x))^{-1} = \left(\left(\frac{\partial f_i(x)}{\partial x_j}\right)_{1 \leq i,j \leq n}\right)^{-1} = \frac{1}{I_f(x)}\,(A_{ij}(x))_{1 \leq i,j \leq n},$$

где $A_{ij}(x)$ — алгебраические дополнения элемента $\frac{\partial f_i(x)}{\partial x_j}$ в матрице Якоби. Следовательно,

$$\frac{\partial g_i(y)}{\partial y_j} = \varphi_{ij}(x) := \frac{1}{I_f(x)} A_{ij}(x), \quad \text{где} \quad x = g(y), \quad 1 \le i, j \le n. \tag{3.26}$$

Якобиан $I_f(x)$ — непрерывная функция в U, не обращающаяся в нуль. Поскольку $A_{ij}(x)$ являются полиномами от частных производных $\frac{\partial f_k(x)}{\partial x_l}$ (элементов матрицы Якоби), которые непрерывны, сами $A_{ij}(x)$ являются непрерывными функциями в U. Отсюда вытекает непрерывность в U правых частей $\varphi_{ij}(x)$ равенств (3.26). С учетом непрерывности обратного отображения $g = f^{-1}$ из (3.26) получаем, что левые части непрерывны в V как суперпозиции непрерывных функций, так как

$$\frac{\partial g_i(y)}{\partial y_j} = \varphi_{ij} \circ g(y), \quad 1 \le i, j \le n. \tag{3.27}$$

Якобиан

$$I_g(y) = \frac{1}{I_f(x)},$$

откуда следует, что $I_g(y) \neq 0$ в V. Следовательно, g — регулярное отображение. Оно инъективно, так как является обратным к f. Таким образом, f^{-1} — диффеоморфизм.

Осталось показать, что если f является k раз непрерывно дифференцируемым, то и g является k раз непрерывно дифференцируемым. Доказательство проведем методом математической индукции. При $k = 1$ это уже доказано. Предположим, что утверждение доказано при $k = l$, где $l \in N$. Докажем его при $k = l + 1$.

Пусть f непрерывно дифференцируемо $(l + 1)$ раз, тогда частные производные $\frac{\partial f_i(x)}{\partial x_j}$ являются l раз непрерывно дифференцируемыми в U. Это означает, что $I_f(x)$ и все $A_{ij}(x)$ как полиномы от этих функций также l раз непрерывно дифференцируемы в U, поэтому функции φ_{ij} являются l раз непрерывно дифференцируемыми. По предположению индукции обратное отображение $g = f^{-1}$ непрерывно дифференцируемо l раз. Из (3.27) вытекает, что функции $\frac{\partial g_i(y)}{\partial y_j}$ являются l раз непрерывно дифференцируемыми. Это означает, что g непрерывно дифференцируемо $(l + 1)$ раз.

3.9 Теоремы об обратной и о неявной функции

Теорема об обратной функции. *Пусть отображение $f : U \to \mathbb{R}^n$ непрерывно дифференцируема на U и $I_f(a) \neq 0$ в некоторой точке $a \in U$. Тогда существует окрестность V точки a и окрестность $W = f(V)$ точки $b := f(a)$ такие, что $f|_V : V \to W$ является диффеоморфизмом, т. е. существует $(f|_V)^{-1} : W \to V$. Более того, если отображение f является k раз непрерывно дифференцируемой, то и $(f|_V)^{-1}$ является k раз непрерывно дифференцируемой.*

Доказательство сразу следует из теоремы о диффеоморфизме и свойств регулярных отображений. Действительно, пусть, для определенности, $I_f(a) > 0$, тогда в силу непрерывности якобиана существует окрестность V_1 точки a, в которой выполняется неравенство $I_f(x) > 0$. Следовательно, отображение f является регулярным на V_1. По теореме о локальной инъективности регулярных отображений существует окрестность $V \subset V_1$ точки a, в которой f является инъективным. При этом $W := f(V)$ открыто. В итоге получаем, что $f|_V$ диффеоморфизм. Далее остается применить теорему о диффеоморфизме.

Пусть отображение $f : U \to \mathbb{R}^m$, где U — открытое множество в R^{n+m}. Обозначим элементы $\mathbb{R}^{n+m}$ через (x, y), где $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_m)$. Важной задачей является нахождение решения векторного уравнения

$$f(x, y) = \theta \tag{3.28}$$

относительно векторной неизвестной y.

Векторное уравнение (3.28) равносильно системе m скалярных уравнений

$$\begin{cases} f_1(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m) & = & 0, \\ f_2(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m) & = & 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \ldots & \ldots \\ f_m(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m) & = & 0. \end{cases}$$

относительно m вещественных неизвестных y_1, $y_2, \ldots, y_m$.

Пусть $(a,b)\in U$, $f(a,b)=\theta$. Если для любого x в малой окрестности V точки a существует единственное $y=g(x)$ из некоторой окрестности W точки b такое, что $f(x,g(x))=\theta$, то говорят, что $y=g(x)$ — это *неявная функция,* заданная в окрестности точки a соотношением $f(x,y)=\theta$.

Пример. Пусть $m=n=1$, $f(x,y)=x^2+y^2-1$. Если $a=0$, $b=1$, то уравнение $x^2+y^2-1=0$ задает неявную функцию $y=\sqrt{1-x^2}$ в окрестности точки $x=0$. Если $a=0$, $b=-1$, то неявная функция в окрестности нуля имеет вид $y=-\sqrt{1-x^2}$. Если $a=\pm1$, $b=0$, то неявной функции не существует ни в какой окрестности точки a.

Возникает естественный вопрос: когда существует неявная функция? В случае дифференцируемости f ответ дает теорема о неявной функции приводимая ниже.

Введем обозначение $f^x(y)=f(x,y)$, для фиксированного x эта функция определена на пересечении U с плоскостью $\{x=\mathrm{const}\}$. Если f дифференцируемо в точке (x,y), то существует якобиан

$$I_{f^x}(y)=\begin{vmatrix} \frac{\partial f_1(x,y)}{\partial y_1} & \frac{\partial f_1(x,y)}{\partial y_2} & \dots & \frac{\partial f_1(x,y)}{\partial y_m} \\ \frac{\partial f_2(x,y)}{\partial y_1} & \frac{\partial f_2(x,y)}{\partial y_2} & \dots & \frac{\partial f_2(x,y)}{\partial y_m} \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ \frac{\partial f_m(x,y)}{\partial y_1} & \frac{\partial f_m(x,y)}{\partial y_2} & \dots & \frac{\partial f_m(x,y)}{\partial y_m} \end{vmatrix}.$$

Теорема о неявной функции. *Пусть $f:U\to\mathbb{R}^m$, где U — открытое множество в $\mathbb{R}^{n+m}$, является непрерывно дифференцируемым в U. Пусть $(a,b)\in U$, $f(a,b)=\theta$ и $I_{f^a}(b)\neq 0$. Тогда существует окрестность V точки a в $\mathbb{R}^n$ и окрестность W точки b в $\mathbb{R}^m$ такие, что для любого $x\in V$ существует единственное $y=:g(x)\in W$ такие, что $f(x,y)=\theta$. При этом g непрерывно дифференцируемо в V. Если же f является k раз непрерывно дифференцируемым, то и g непрерывно дифференцируемо k раз.*

Доказательство. Рассмотрим отображение

$$F(x,y):=(x,f(x,y)),\quad F:U\to\mathbb{R}^{n+m}.$$

Имеем $F(a,b)=(a,\theta)$. Запишем компоненты F:

$$\begin{array}{lcl}
F_1\ \ (x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m) & = & x_1, \\
F_2\ \ (x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m) & = & x_2, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \ldots & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
F_n\ \ (x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m) & = & x_n, \\
F_{n+1}(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m) & = & f_1(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m), \\
F_{n+2}(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m) & = & f_2(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m), \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \ldots & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
F_{n+m}(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m) & = & f_m(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m).
\end{array}$$

Подсчитаем якобиан F:

$$I_F = \begin{vmatrix}
1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \\
\dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \ldots & \dfrac{\partial f_1}{\partial x_n} & \dfrac{\partial f_1}{\partial y_1} & \dfrac{\partial f_1}{\partial y_2} & \ldots & \dfrac{\partial f_1}{\partial y_m} \\
\dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \ldots & \dfrac{\partial f_2}{\partial x_n} & \dfrac{\partial f_2}{\partial y_1} & \dfrac{\partial f_2}{\partial y_2} & \ldots & \dfrac{\partial f_2}{\partial y_m} \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \ldots & \dfrac{\partial f_m}{\partial x_n} & \dfrac{\partial f_m}{\partial y_1} & \dfrac{\partial f_m}{\partial y_2} & \ldots & \dfrac{\partial f_m}{\partial y_m}
\end{vmatrix}.$$

Это — определитель матрицы, состоящей из четырех прямоугольных блоков. Блок в левом верхнем углу представляет собой единичную матрицу размера $(n\times n)$, в правом верхнем — нулевую матрицу размера $(n\times m)$. Следовательно, определитель I_F равен определителю блока размера $(m \times m)$, стоящего в нижнем правом углу:

$$I_F = \begin{vmatrix} \dfrac{\partial f_1}{\partial y_1} & \dfrac{\partial f_1}{\partial y_2} & \dots & \dfrac{\partial f_1}{\partial y_m} \\ \dfrac{\partial f_2}{\partial y_1} & \dfrac{\partial f_2}{\partial y_2} & \dots & \dfrac{\partial f_2}{\partial y_m} \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ \dfrac{\partial f_m}{\partial y_1} & \dfrac{\partial f_m}{\partial y_2} & \dots & \dfrac{\partial f_m}{\partial y_m} \end{vmatrix},$$

поэтому $I_F(a,b) = I_{f^a}(b) \neq 0$. В силу этого к отображению F можно применить теорему об обратной функции. По этой теореме существует окрестность $B_r(c)$ точки $c = (a,b)$ в $\mathbb{R}^{n+m}$, в которой F является диффеоморфизмом.

Пусть $\delta = \frac{r}{\sqrt{2}}$, $V_1 = B_\delta(a)$, $W = B_\delta(b)$. Тогда $V_1 \times W \subset B_r(c)$. Действительно, пусть $(x,y) \in V_1 \times W$, тогда $x \in V$, $y \in W$ и

$$\|(x,y) - (a,b)\|^2_{\mathbb{R}_{n+m}} = \|x-a\|^2_{\mathbb{R}_n} + \|y-b\|^2_{\mathbb{R}_m} < 2\delta^2 = r^2,$$

т. е. $(x,y) \in B_r(c)$.

Отображение F является диффеоморфизмом на $B_r(c)$, следовательно, на его подмножестве $V_1 \times W$ оно биективно, т. е. существует диффеоморфизм

$$F^{-1} : F(V_1 \times W) =: Z \to V_1 \times W.$$

(Здесь для простоты обозначений мы пишем F^{-1}, хотя на самом деле мы имеем в виду диффеоморфизм, обратный к сужению $F|_{V_1 \times W}$ отображения F на множество $V_1 \times W$.) Так как F сохраняет неизменными первые n координат, то и F^{-1} обладает этим свойством, т. е. можно написать $F^{-1}(x,y) = (x, h(x,y))$, где функция $h(x,y)$ непрерывно дифференцируема, так как F^{-1} — диффеоморфизм.

Так как $f(a,b) = \theta$, $F(a,b) = \widetilde{a} := (a,\theta)$, то точка $\widetilde{a} \in Z$. Множество Z является открытым, поэтому существует окрестность $B_\varepsilon(\widetilde{a}) \subset Z$. Пусть $V = B_\varepsilon(a)$. Покажем, что окрестности V и W являются искомыми.

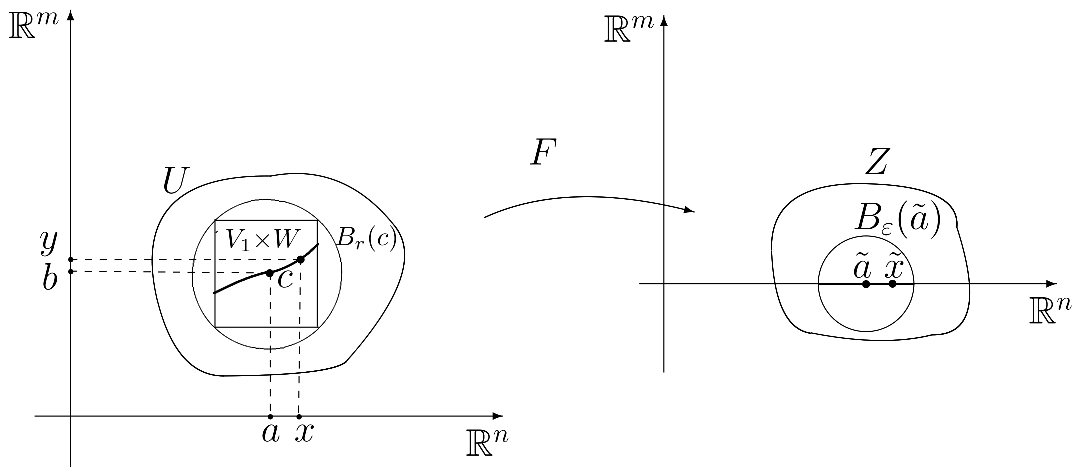

Рис. 1

Действительно, если $x \in V$, то точка $\widetilde{x} := (x, \theta) \in B_\varepsilon(\widetilde{a})$, так как

$$\|\widetilde{x} - \widetilde{a}\|_{\mathbb{R}^{n+m}} = \|x - a\|_{\mathbb{R}^n} < \varepsilon.$$

Отсюда следует, что $\widetilde{x} \in Z$. Положим

$$y = g(x) := h(x, \theta).$$

Тогда $(x, y) = F^{-1}(\widetilde{x}) \in V_1 \times W$, откуда $y \in W$. Поскольку

$$F^{-1}(\widetilde{x}) = F^{-1}(x, \theta) = (x, h(x, \theta)) = (x, y),$$

получаем $(x, f(x, y)) = F(x, y) = (x, \theta)$, следовательно, $f(x, y) = \theta$.

На рис. 1 график отображения $y = g(x)$, $x \in V$, изображен слева жирной линией; справа жирной линией изображено множество $V \times \{\theta\}$, соответствующее этому графику при отображении F.

Докажем единственность такого y. Если существует другое $y' \in W$ и $(x, y') \in V \times W$, $f(x, y') = \theta$, то $F(x, y') = (x, \theta) \in B_\varepsilon(\widetilde{a}) \subset F(V_1 \times W)$. Поскольку F биективно на $V \times W$, то

$$(x, y') = F^{-1}((x, \theta)) = (x, h(x, \theta)) = (x, y).$$

Итак, $y' = y$.

Таким образом, $g(x) = h(x, \theta)$. Это отображение непрерывно дифференцируемо, так как F^{-1} непрерывно дифференцируемо. Если же f является k раз непрерывно дифференцируемым, то F, F^{-1}, а, следовательно, h и g непрерывно дифференцируемы k раз. Теорема доказана.

3.10 Условный экстремум функций нескольких переменных

Начнем рассмотрение с примера задачи на условный экстремум.

Пример. Найти минимум функции $f(x,y) = x^2 + y^2$. Нетрудно видеть, что точка $(0,0)$ является точкой минимума функции f. Теперь найдем минимум функции среди всех точек (x,y), удовлетворяющих дополнительному условию $x+y=1$ (такой минимум называется условным). Имеем $y = 1-x$, откуда при таких (x,y) функция

$$f(x,y) = x^2 + (1-x)^2 = 2x^2 - 2x + 1.$$

Исследуя функцию $g(x) := 2x^2 - 2x + 1$, $x \in \mathbb{R}$, на экстремум, находим, что точка $x = 1/2$ является точкой минимума функции g. Таким образом, среди всех точек, удовлетворяющих условию $x + y = 1$, наименьшее значение функция f достигает в точке $(1/2, 1/2)$.

Теперь дадим общую постановку задачи на условный экстремум.

Пусть в области $U \subset \mathbb{R}^n$ заданы функции f, $g_1, \ldots, g_m : U \to \mathbb{R}$. Обозначим

$$\Gamma := \{x \in U \mid g_1(x) = g_2(x) = \ldots g_m(x) = 0\}.$$

Будем предполагать, что $\Gamma \neq \emptyset$.

Точка $x_0 \in \Gamma$ называется *точкой условного локального минимума* (*максимума*) *функции* f *при ограничениях* $g_j(x) = 0$, $1 \leq j \leq m$, если существует окрестность V точки x_0 такая, что для любого $x \in \Gamma \cap V$ имеет место неравенство $f(x) \geq f(x_0)$ $(f(x) \leq f(x_0))$.

В дальнейшем будем предполагать, что все эти функции f и g_j, $1 \leq j \leq m$, непрерывно дифференцируемы на U, $m < n$ и в любой точке

$x \in U$ матрица

$$\begin{pmatrix} \dfrac{\partial g_1(x)}{\partial x_1} & \dfrac{\partial g_1(x)}{\partial x_2} & \dots & \dfrac{\partial g_1(x)}{\partial x_n} \\ \dfrac{\partial g_2(x)}{\partial x_1} & \dfrac{\partial g_2(x)}{\partial x_2} & \dots & \dfrac{\partial g_2(x)}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \dfrac{\partial g_m(x)}{\partial x_1} & \dfrac{\partial g_m(x)}{\partial x_2} & \dots & \dfrac{\partial g_m(x)}{\partial x_n} \end{pmatrix} = \left(\frac{\partial g_j(x)}{\partial x_k} \right)_{1 \le j \le m, 1 \le k \le n}, \tag{3.29}$$

имеет ранг m.

Теорема Лагранжа. *Пусть точка x_0 является точкой локального условного экстремума функции f при ограничениях $g_j(x) = 0$, $1 \le j \le m$. Тогда существуют единственные константы λ_1, $\lambda_2, \dots, \lambda_m$ такие, что имеют место равенства*

$$\frac{\partial f(x_0)}{\partial x_k} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(x_0)}{\partial x_k} = 0, \quad 1 \le k \le n. \tag{3.30}$$

Доказательство. Пусть x_0 является точкой локального условного экстремума функции f при ограничениях $g_j(x) = 0$, $1 \le j \le m$. По предположению, сформулированному выше, ранг матрицы (3.29) равен m. Без ограничения общности можно считать, что определитель

$$\begin{vmatrix} \dfrac{\partial g_1(x)}{\partial x_{n-m+1}} & \dfrac{\partial g_1(x)}{\partial x_{n-m+2}} & \dots & \dfrac{\partial g_1(x)}{\partial x_n} \\ \dfrac{\partial g_2(x)}{\partial x_{n-m+1}} & \dfrac{\partial g_2(x)}{\partial x_{n-m+2}} & \dots & \dfrac{\partial g_2(x)}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \dfrac{\partial g_m(x)}{\partial x_{n-m+1}} & \dfrac{\partial g_m(x)}{\partial x_{n-m+2}} & \dots & \dfrac{\partial g_m(x)}{\partial x_n} \end{vmatrix} \neq 0. \tag{3.31}$$

По теореме о неявной функции в некоторой окрестности точки x_0 множество Γ можно задать в виде системы уравнений

$$\begin{array}{lcl} x_{n-m+1} & = & \varphi_1(x_1, x_2, \dots, x_{n-m}), \\ x_{n-m+2} & = & \varphi_2(x_1, x_2, \dots, x_{n-m}), \\ \dots\dots & \dots & \dots\dots\dots\dots\dots\dots \\ x_n & = & \varphi_m(x_1, x_2, \dots, x_{n-m}). \end{array} \tag{3.32}$$

Обозначим $s = (x_1, x_2, \ldots, x_{n-m})$, $t = (x_{n-m+1}, x_{n-m+2}, \ldots, x_n$. Тогда пару (s, t) можно отождествить с точкой $x = (x_1, x_2, \ldots, x_n)$. Пусть $x_0 = (s_0, t_0)$

Перепишем (3.32) в векторном виде как $t = \varphi(s)$, где s принадлежит некоторой окрестности точки s_0, $\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_m)$. Если s принадлежит малой окрестности точки s_0, то точка $x = (s, \varphi(s))$ лежит на Γ в малой окрестности точки x_0. Отсюда следует, что точка s_0 является точкой (безусловного) экстремума функции $F(s) := f(s, \varphi(s))$. Так как по условию функции f и g_j, $1 \le j \le m$, непрерывно дифференцируемы на U, то по теореме о неявной функции φ, а, следовательно, и F непрерывно дифференцируема в малой окрестности точки s_0. Запишем необходимое условие экстремума функции F. Имеем $\nabla F(s_0) = 0$ или, для любого $1 \le k \le n - m$

$$\frac{\partial F(s_0)}{\partial s_k} = \frac{\partial f(x_0)}{\partial x_k} + \sum_{j=n-m+1}^{n} \frac{\partial f(x_0)}{\partial x_j} \frac{\partial \varphi_j(s_0)}{\partial s_k} = 0. \tag{3.33}$$

Запишем (3.33) короче

$$\frac{\partial f(x_0)}{\partial s_k} + \left\langle \nabla_t f(x_0), \frac{\partial \varphi(s_0)}{\partial s_k} \right\rangle = 0, \tag{3.34}$$

где

$$\nabla_t f(x_0) = \left(\frac{\partial f(x_0)}{\partial x_{n-m+1}}, \frac{\partial f(x_0)}{\partial x_{n-m+2}}, \ldots, \frac{\partial f(x_0)}{\partial x_n} \right).$$

Так как $g_j(x) = 0$, $x \in \Gamma$, для любого j , то

$$g_j(s, \varphi(s)) = 0 \tag{3.35}$$

в окрестности точки s_0. Дифференцируя соотношения (3.35) по s_k, получаем

$$\frac{\partial g_j(x_0)}{\partial s_k} + \left\langle \nabla_t g_j(x_0), \frac{\partial \varphi(s_0)}{\partial s_k} \right\rangle = 0, \quad 1 \le j \le m. \tag{3.36}$$

Умножая (3.36) на (пока произвольные!) константы λ_j и складывая с (3.34), получаем

$$\frac{\partial f(x_0)}{\partial s_k} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(x_0)}{\partial s_k} + \left\langle \nabla_t f(x_0) + \sum_{j=1}^{m} \lambda_j \nabla_t g_j(x_0), \frac{\partial \varphi(s_0)}{\partial s_k} \right\rangle = 0. \tag{3.37}$$

Подберем константы λ_j таким образом, чтобы

$$\nabla_t f(x_0) + \sum_{j=1}^{m} \lambda_j \nabla_t g_j(x_0) = 0$$

или покоординатно

$$\frac{\partial f(x_0)}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(x_0)}{\partial x_i} = 0, \quad n - m + 1 \le i \le n. \tag{3.38}$$

Отметим, что равенства (3.38) являются частью равенств (3.30). Система (3.38) является системой из m линейных уравнений относительно неизвестных λ_j. Определитель системы в силу (3.31) не равен нулю. Следовательно существуют единственные λ_j, удовлетворяющие (3.38). Для таких λ_j в силу (3.37) имеют равенства (3.30). Теорема доказана.

Замечание 1. Соотношения (3.30) можно записать в виде

$$\nabla \widehat{f}(x_0) = 0,$$

где $\widehat{f} = f + \sum_{j=1}^{m} \lambda_j g_j$. Константы λ_1, $\lambda_2, \ldots, \lambda_m$ называются *множителями Лагранжа,* а функция $\widehat{f}$ — *функцией Лагранжа.*

Замечание 2. Соотношения (3.30) означают, что градиенты функций f и g_j, $1 \le j \le m$, в точке локального условного экстремума являются линейно зависимыми.

Замечание 3. Метод Лагранжа нахождения точек локального условного экстремума состоит в следующем.

1) Составляется функция Лагранжа $\widehat{f} = f + \sum_{j=1}^{m} \lambda_j g_j$.

2) Записывается система $\nabla \widehat{f}(x) = 0$, состоящая из n уравнений.

3) К этим уравнениям присоединяются m уравнений $g_j(x) = 0$, $1 \le j \le m$.

В результате получается система из $(n+m)$ уравнений с $(n+m)$ неизвестными x_1, $x_2, \ldots, x_n$, λ_1, $\lambda_2, \ldots, \lambda_n$. Решая ее, находим точки, подозрительные на экстремум.

Замечание 4. Теорема Лагранжа дает лишь необходимое условие условного экстремума. Чтобы получить достаточное условие, нужно исследовать на строгую положительность второй дифференциал функции f (или $\widehat{f}$, которая на Γ совпадает с f), но не для произвольных дифференциалов dx, а для тех, которые удовлетворяют равенствам

$$dg_j(x_0) = \sum_{k=1}^{n} \frac{\partial g_j(x_0)}{\partial x_k}\, dx_k = 0.$$

Пример. Найти экстремум функции

$$f(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \ldots + x_n$$

при ограничении $x_1^2 + x_2^2 + \ldots + x_n^2 = 1$. По теореме Вейерштрасса непрерывная функция f на единичной сфере (компакте) достигает своего минимального и максимального значений. Составим функцию Лагранжа ($m = 1$, вместо λ_1 пишем просто λ):

$$\widehat{f}(x) := x_1 + x_2 + \ldots + x_n + \lambda(x_1^2 + x_2^2 + \ldots + x_n^2 - 1).$$

В точке локального экстремума имеем $\frac{\partial \widehat{f}}{\partial x_j} = 1 + 2\lambda x_j = 0$, откуда находим $x_j = -1/(2\lambda)$. Кроме того, $x_1^2 + x_2^2 + \ldots + x_n^2 = 1$, поэтому $\lambda^2 = n/4$. Имеем два решения: либо все $x_j = 1/\sqrt{n}$ или все $x_j = -1/\sqrt{n}$. Нетрудно убедиться, что, в силу сделанного в начале замечания, в первом случае имеем точку максимума, а во втором — минимума.

Продемонстрируем, как исследовать, какая из точек будет точкой условного локального минимума, а какая — условного локального максимума, без использования теоремы Вейерштрасса. Поскольку $d^2 f \equiv 0$, рассмотрим второй дифференциал функции Лагранжа $d^2\widehat{f}(x) = 2\lambda \sum_{j=1}^{n} dx_j^2$ при условии, что

$$d(x_1^2 + x_2^2 + \ldots + x_n^2 - 1) = 2\sum_{j=1}^{n} x_j dx_j = 0.$$

Таким образом, с учетом того, что все x_k равны между собой в точках, подозрительных на экстремум, получаем $\sum_{j=1}^{n} dx_j = 0$. Выразим dx_n через остальные дифференциалы: $dx_n = -\sum_{j=1}^{n-1} dx_j$, откуда

$$d^2\widehat{f}(x) = 2\lambda \left[\sum_{j=1}^{n-1} dx_j^2 + \left(\sum_{j=1}^{n-1} dx_j \right)^2 \right].$$

Нетрудно видеть, что второй дифференциал функции $\widehat{f}$ на Γ является квадратичной формой от dx_1, $dx_2, \ldots, dx_{n-1}$. Эта форма строго положительно определена, если $\lambda > 0$, в этом случае точка с координатами $x_j = -1/\sqrt{n}$ является точкой условного минимума. Если $\lambda < 0$, то соответствующее решение ($x_j = 1/\sqrt{n}$) является точкой условного максимума, так как форма строго отрицательно определена.

Литература

[1] Никольский С.М. *Курс математического анализа, т. 1.* – М.: Наука, 1973. – 432 с.

[2] Зорич В.А. *Математический анализ, ч. I.* – М.: Наука, 1981.– 243 с.

[3] Шерстнев А.Н. *Конспект лекций по математическому анализу.* – Казань: КГУ, 2009. – 374 с.

[4] Кудрявцев Л.Д. *Математический анализ, т.1.* – М.: Высшая школа, 1973. – 614 с.

[5] Фихтенгольц Г.М. *Курс дифференциального и интегрального исчисления, т. 1.* – М.: Физматлит, 2001. – 616 с.

[6] Демидович Б.П. *Сборник задач по математическому анализу.* – М.: МГУ, ЧеРо, 1997. – 624 с.

[7] Гелбаум Б., Олмстед Дж. *Контрпримеры в анализе.* – М.: Мир, 1967. – 251 с.

Оглавление