Санкт-Петербургский Государственный Университет

Факультет математики и компьютерных наук

Многомерный анализ

Конспект основан на лекциях
Романа Викторовича Бессонова

17 сентября 2020 г.

Санкт-Петербургский
государственный
университет

Факультет
математики
и компьютерных
наук СПбГУ

Конспект основан на лекциях по многомерному анализу, прочитанных Романом Викторовичем Бессоновым студентам Факультета математики и компьютерных наук Санкт-Петербургского государственного университета в весеннем семестре 2019–2020 учебного года.

Материал конспекта соответствует части курса математического анализа, читаемой во 2-ом семестре бакалавриата.

Автор:

Михаил Опанасенко

Редактор:

Михаил Антоненко

Последняя версия и исходный код:

https://www.overleaf.com/read/dtbndhrbgmsd

Сайт СПБГУ: https://spbu.ru.
Сайт факультета МКН: https://math-cs.spbu.ru.

Оглавление

1 Введение . . . 1
2 Нормы . . . 3
2.1 Норма отображения . . . 5
2.2 Топология на пространстве линейных отображений . . . 6
3 Дифференцируемость . . . 8
3.1 Базовые свойства . . . 9
4 Дифференцируемые скалярные функции . . . 12
4.1 Производная по направлению . . . 12
4.2 Экстремумы и градиент . . . 16
4.3 Аналоги формулы Лагранжа . . . 17
4.4 Перестановочность частных производных . . . 20
4.5 Формула Тейлора . . . 21
4.6 Экстремумы и гессиан . . . 24
5 Гладко параметризованные многообразия . . . 27
5.1 Локальная билипшицевость . . . 27
5.2 Открытость отображения с невырожденным дифференциалом 29
5.3 Теорема об обратной функции . . . 30
5.4 Теорема о неявной функции . . . 31
5.5 Гладко параметризованные многообразия в $\mathbb{R}^n$ . . . 36
5.6 Касательное пространство . . . 41
5.7 ГПМ как множество уровня . . . 45
5.8 ГПМ и графики . . . 49
6 Множители Лагранжа . . . 49
6.1 Необходимое условие . . . 50
6.2 Достаточное условие . . . 52
7 Выпуклые функции . . . 55

Многомерный анализ

1 Введение

Определение. Множество X называется *линейным пространством* над полем $\mathbb{R}$, если заданы операции сложения $+ : X \times X \to X$ и умножения на скаляр $\cdot : \mathbb{R} \times X \to X$ такие, что выполняются условия:

(1) $(x+y)+z = x+(y+z)$;

(2) $x+y = y+x$;

(3) $\exists 0 \in X : x+0 = x$;

(4) $\forall x \in X,\ \exists! y = -x : x+y = 0$;

(5) $(\alpha+\beta)x = \alpha x + \beta x$;

(6) $(\alpha\beta)x = \alpha(\beta x)$;

(7) $\alpha(x+y) = \alpha x + \beta y$;

(8) $1 \cdot x = x$.

для всех $x, y, z \in X$ и $\alpha, \beta \in \mathbb{R}$.

Определение. *Базис* линейного пространства X — такой набор векторов

$$e_1, e_2, \ldots, e_n,$$

что любой вектор $x \in X$ представляется единственным образом в виде линейной комбинации

$$x = c_1 e_1 + \cdots + c_n e_n,$$

где $c_i \in \mathbb{R}$ для всех i.

Определение. Пространство X называется *конечномерным*, если у него есть базис из конечного числа элементов.

Определение. *Размерность* конечномерного линейного пространства X — это число элементов его базиса.

Из курса алгебры известно, что размерность линейного пространства не зависит от выбора базиса.

Евклидовы пространства

Большую часть времени мы будем изучать над n-кратное декартово произведение $\mathbb{R}$:

$$\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}.$$

В каком-то смысле $\mathbb{R}^n$ — это пара из вещественного векторного пространства и его базиса. Векторы здесь отождествлены со столбцами:

$$\mathbb{R}^n = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \middle| \; a_i \in \mathbb{R} \right\}.$$

Стандартный базис — это столбцы с единицей на одной позиции и нулями на остальных. Часто векторы стандартного базиса обозначаются через $e_1, \dots, e_n$. Операции в $\mathbb{R}^n$ задаются покоординатно:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}, \qquad \alpha \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \alpha a_1 \\ \alpha a_2 \\ \vdots \\ \alpha a_n \end{pmatrix}.$$

Линейные отображения

Определение. Отображение $T \colon \mathbb{R}^m \to \mathbb{R}^n$ называется *линейным*, если для всех $x, y \in X$ и $\alpha, \beta \in \mathbb{R}$ выполнено

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y).$$

Пространство линейных отображений из $\mathbb{R}^m$ в $\mathbb{R}^n$ будем обозначать через

$$\mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n).$$

Определение. Линейное отображение векторного пространства в себя называется *оператором*. Множество операторов на пространстве $\mathbb{R}^m$ будем обозначать через

$$\mathrm{End}(\mathbb{R}^m).$$

Операции векторного пространства на $\mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ задаются поточечно:

$$T_1 + T_2 := x \mapsto T_1 x + T_2 x, \qquad \alpha T := x \mapsto \alpha T x.$$

Поскольку мы уже выбрали базис, будем отождествлять линейные отображения из $\mathbb{R}^m$ в $\mathbb{R}^n$ с их матрицами размера $n \times m$.

2 Нормы

Определение. *Нормой* на векторном пространстве X над $\mathbb{R}$ или $\mathbb{C}$ называется отображение

$$\| \ \| \colon X \to \mathbb{R},$$

удовлетворяющее следующим свойствам для любых векторов x и y и скаляра α:

(1) $x \neq 0 \implies \|x\| > 0$;

(2) $\|\alpha x\| = |\alpha| \cdot \|x\|$;

(3) $\|x + y\| \leqslant \|x\| + \|y\|$.

Определение. Пусть X — векторное пространство над $\mathbb{R}$ или $\mathbb{C}$, а f — норма на X. Тогда пара (X, f) называется *нормированным* векторным пространством.

Как показывает следующее упражнение, нормированные векторные пространства являются топологическими пространствами.

Упражнение. Если f — норма на векторном пространстве X, то

$$d(x, y) := f(x - y)$$

— метрика на X.

В терминах упражнения выше, будем называть d метрикой, *индуцированной* нормой f.

Некоторые из вещей, которые мы докажем для евклидовых пространств $\mathbb{R}^n$, будут верными и для произвольных нормированных пространств. Например, почти без изменений останется определение дифференцируемости.

Для любого вектора $x \in \mathbb{R}^n$, определим

$$\|x\| := \sqrt{\sum_{k=1}^{n} x_k^2}. \tag{2.1}$$

Утверждение 2.1. Функция $\| \ \|$, определённая в (2.1) — это норма на $\mathbb{R}^n$.

Доказательство. Убедитесь, что первые два свойства действительно выполняются. Для третьего имеем неравенство

$$\sqrt{\sum_k \left(c_{1,k} + c_{2,k}\right)^2} \leqslant \sqrt{\sum_k c_{1,k}^2} + \sqrt{\sum_k c_{2,k}^2}$$

из предыдущего семестра. ∎

Определение. Будем называть норму $\| \ \|$ на $\mathbb{R}^n$ *стандартной*.

На самом деле, в некотором смысле достаточно ограничить себя рассмотрением этой нормы.

Определение. Две нормы f и g называются *эквивалентными*, если индуцированные ими метрики липшицево эквивалентны. Иначе говоря, существуют такие положительные константы c_1 и c_2, что для любого вектора x

$$c_1 f(x) \leqslant g(x) \leqslant c_2 f(x).$$

Упражнение. Докажите, что это отношение эквивалентности.

Лемма 2.2. Пусть f — норма на $\mathbb{R}^n$. Тогда f — липшицева функция.

Доказательство. Пусть $\{e_i\}$ — стандартный базис. Тогда для любого вектора $x \in \mathbb{R}^n$ верно

$$f(x) = \sum_{i=1}^{n} |x_i| \cdot f(e_i).$$

Это скалярное произведение двух векторов, которое можно оценить по неравенству КБШ:

$$f(x) \leqslant \|(|x_1|, \ldots, |x_n|)\| \cdot \|(f(e_1), \ldots, f(e_n))\|.$$

Первый из этих множителей — это стандартная норма $\|x\|$ вектора x; второй — какая-то константа c. Тогда для любых векторов $x, y \in \mathbb{R}^n$ выполняется требуемое неравенство:

$$|f(x) - f(y)| \leqslant |f(x - y)| \leqslant \|x - y\| \cdot c. \tag{2.2}$$

Здесь первое неравенство верно по неравенству треугольника для нормы f, второе — из оценки выше. ■

Следствие 2.3. Пусть f — норма на $\mathbb{R}^n$. Тогда f непрерывна.

Доказательство. Ясно, что если расстояние между векторами x и y стремится у нулю, то по (2.2) и число $|f(x) - f(y)|$ стремится к нулю. ■

Теорема 2.4. Пусть f и g — нормы на $\mathbb{R}^n$. Тогда f и g эквивалентны.

Доказательство. Достаточно доказать, что произвольная норма f на $\mathbb{R}^n$ эквивалентна стандартной. Заметим, что всякий вектор $x \in \mathbb{R}^n$ представляется в виде произведения некоторого положительного скаляра на вектор из S^{n-1}:

$$f(x) = f\left(\frac{x}{\|x\|} \cdot \|x\|\right) = \|x\| \cdot f\left(\frac{x}{\|x\|}\right). \tag{2.3}$$

Сфера S^{n-1} компактна, а норма f непрерывна по следствию 2.3. Поэтому по теореме Вейерштрасса f достигает на ней минимума m и максимума M. Таким образом, для любого вектора $x \in \mathbb{R}^n$ будет верно

$$m \leqslant f\left(\frac{x}{\|x\|}\right) \leqslant M.$$

Подставляя в это неравенство (2.3), получаем

$$\|x\| \cdot m \leqslant f(x) \leqslant \|x\| \cdot M,$$

что и требовалось. ■

2.1 Норма отображения

Отметим, что линейные отображения не могут растягивать вектора слишком сильно. Пусть T — линейное отображение на $\mathbb{R}^m$, $x \in \mathbb{R}^m$. Тогда верны следующие оценки:

$$\begin{aligned} \|Tx\| &= \|x_1 Te_1 + \dots + x_n Te_n\| \\ &\leqslant |x_1| \cdot \|Te_1\| + \dots + |x_n| \cdot \|Te_n\| \\ &\leqslant \|x\| \cdot \|(\|Te_1\|, \dots, \|Te_n\|)\|. \end{aligned}$$

Иначе говоря,

$$\|Tx\| \leqslant c\|x\|$$

для некоторой константы c, которая зависит только от отображения T. Посчитанное нами значение

$$c = \|(\|Te_1\|, \dots, \|Te_n\|)\|$$

на самом деле не оптимально. Но изучение нижней грани таких оценок приводит к полезным результатам.

Определение. Пусть T — линейное отображение. Его *норма* — это величина

$$\|T\| := \sup\left\{ \frac{\|Tx\|}{\|x\|} \,\middle|\, x \in \mathbb{R}^m \setminus 0 \right\}.$$

Как мы показали выше, этот супремум всегда конечен.

Заметим, что из-за линейности отображения T нам важно направление вектора x, но не его норма. То есть, супремум можно брать по точкам в проективизации $\mathbb{R}^m$ или, например, по векторам на сфере S^{m-1}:

$$\|T\| = \sup_{|x|=1} \|Tx\|.$$

Упражнение. Функция $T \mapsto \|T\|$ — норма на пространстве $\mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ линейных отображений из $\mathbb{R}^m$ в $\mathbb{R}^n$.

Утверждение 2.5. Линейные отображения липшицевы.

Доказательство. Действительно,

$$\begin{aligned} \|Tx - Ty\| &= \|T(x - y)\| \\ &\leqslant \|T\| \cdot \|x - y\| \end{aligned}$$

для любых $x, y \in \mathbb{R}^n$. ■

Утверждение 2.6. Норма композиции отображений A и B не превосходит произведения их норм:

$$\|AB\| \leqslant \|A\| \cdot \|B\|$$

Доказательство. Последовательно оценим норму образа через нормы отображений:

$$\begin{aligned}\|ABx\| &= \|A(Bx)\| \\ &\leqslant \|A\| \cdot \|Bx\| \\ &\leqslant \|A\| \cdot \|B\| \cdot \|x\|.\end{aligned}$$

Значит, и супремум не превосходит произведения норм. ■

2.2 Топология на пространстве линейных отображений

Определение. *Ядром* отображения A называется полный прообраз нуля:

$$\operatorname{Ker} A := \{x \mid Ax = 0\}.$$

Лемма 2.7. Пусть $A\colon U \to V$ — линейное отображение. Тогда следующие утверждения эквивалентны:

(1) $\dim U = \dim V$ и $\operatorname{Ker} A = \{0\}$;

(2) A биективно.

Доказательство. Пусть отображение A биективно. В частности, оно инъективно, и поэтому имеет нулевое ядро. По формуле для размерностей ядра и образа, имеем

$$\begin{aligned}\dim U &= \dim \operatorname{Im} A + \dim \operatorname{Ker} A \\ &= \dim \operatorname{Im} A.\end{aligned}$$

Поскольку A сюръективно, размерность его образа совпадает с размерностью V. Поэтому $\dim U = \dim V$, что и требовалось.

Пусть размерности U и V совпадают, и A имеет нулевое ядро. Сразу получаем, что A инъективно. Аналогично предыдущему абзацу,

$$\dim U = \dim \operatorname{Im} A.$$

По условию $\dim U = \dim V$, а потому

$$\dim \operatorname{Im} A = \dim V.$$

Любое подпространство V размерности $\dim V$ совпадает с самим V: иначе можно было бы дополнить до базиса и получить базис размером больше чем $\dim V$, что невозможно по теореме о равномощности базисов. Значит, A сюръективно. ■

Теорема 2.8 (близкий к обратимому оператор обратим). Пусть $A, B\colon \mathbb{R}^n \to \mathbb{R}^n$ — линейные операторы, причём

$$\|B\| < \frac{1}{\|A^{-1}\|}.$$

Тогда оператор $A - B$ обратим.

Доказательство. Вынесем за скобку обратимый A:

$$A - B = A(I - A^{-1}B).$$

Получается, достаточно доказать, что оператор

$$I - A^{-1}B$$

обратим. Обозначим

$$C := A^{-1}B$$

и посмотрим, может ли у оператора $I - C$ быть нетривиальное ядро. Вычислим его на некотором векторе x и оценим норму результата:

$$\begin{aligned} \|(I - C)x\| &\geqslant \|x\| - \|Cx\| \\ &\geqslant \|x\| - \|C\|\|x\| \\ &= (1 - \|C\|)\|x\|. \end{aligned}$$

При этом по условию

$$\begin{aligned} 1 - \|C\| &= 1 - \|A^{-1}B\| \\ &\geqslant 1 - \|A^{-1}\|\|B\| \\ &> 1 - 1 = 0. \end{aligned}$$

Таким образом, $\|(I - C)x\| > 0$ при $x \neq 0$, а это значит, что у оператора $I - C$ нулевое ядро, то есть он обратим. ■

Следствие 2.9. Множество обратимых операторов открыто в топологии, индуцированной метрикой

$$(A, B) \mapsto \|A - B\|.$$

Доказательство. В теореме 2.8 мы как раз нашли окрестность около обратимого оператора. ■

Утверждение 2.10 (билипшицевость обратимых операторов). Линейное отображение $A\colon \mathbb{R}^m \to \mathbb{R}^n$ обратимо тогда и только тогда, когда существует такая положительная константа c, что для всякого вектора $x \in \mathbb{R}^m$ верно

$$\|Ax\| \geqslant c\|x\|.$$

Доказательство. Пусть отображение A обратимо. Оценим через норму обратного отображения:

$$\begin{aligned} \|x\| &= \|A^{-1}Ax\| \\ &\leqslant \|A^{-1}\| \cdot \|Ax\|. \end{aligned}$$

Обратное утверждение следует из того, что у билипшицева отображения по определению не может быть нетривиального ядра. ■

Помимо этого, нам впоследствии пригодится следующее утверждение.

Утверждение 2.11 (непрерывность обращения оператора). Пусть последовательность $\{A_k\}$ обратимых операторов на $\mathbb{R}^m$ сходится по стандартной норме к оператору A. Тогда последовательность из их обратных $\{A_k^{-1}\}$ сходится к A^{-1}.

Доказательство. Заметим, что

$$A^{-1} - A_k^{-1} = A_k^{-1}(A_k - A)A^{-1}.$$

Поскольку A_k^{-1} удлиняет векторы максимум во столько раз, во сколько их укорачивает A,

$$\left\|A_k^{-1}\right\| \leqslant 1/c,$$

где c — константа, которая существует по билипшицевости обратимого A_k (утверждение 2.10). Нормы остальных множителей очевидно ограничены. ■

3 Дифференцируемость

Определение (дифференцируемость и дифференциал в точке). Пусть X — подмножество $\mathbb{R}^m$, $f\colon X \to \mathbb{R}^n$ — функция, $x \in \operatorname{Int} X$ — внутренняя точка области определения f. Будем говорить, что функция f *дифференцируема* в точке x, если существует линейное отображение $A \in \operatorname{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ такое, что

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - Ah\|}{\|h\|} = 0,$$

причём отображение A будем называть *дифференциалом* функции f *в точке* x и писать в этом случае

$$\mathrm{d}_x f = A.$$

Иначе говоря,

$$f(x+h) - f(x) = [\mathrm{d}_x f]h + r(h)$$

для некоторой векторнозначной функции $r\colon \mathbb{R}^m \to \mathbb{R}^n$ такой, что

$$\lim_{h\to 0} \frac{r(h)}{\|h\|} = 0.$$ [1]

Отметим, что функция r зависит от точки x. В следующем параграфе мы докажем, что дифференциал в точке единственен, в связи с чем корректно следующее определение.

Определение (дифференциал). Отображение

$$x \mapsto \mathrm{d}_x f,$$

действующее из $\mathbb{R}^m$ в $\operatorname{Hom}(\mathbb{R}^m, \mathbb{R}^n)$, называется *дифференциалом* функции f.

[1] То есть $r(h) = o(\|h\|)$.

Определение. Матрица линейного отображения $\mathrm{d}_x f$ в стандартном базисе называется *матрицей Якóби*.

Интуиция при работе с дифференцируемыми функциями заключается в том, что вблизи любой точки функция будет вести себя очень похоже на свой дифференциал. Например (и мы докажем это), обратимость матрицы Якоби влечёт локальную инъективность непрерывно-дифференцируемой функции.

Связь с одномерным случаем

Покажем, что в случае, когда f — вещественнозначная функция из подмножества $\mathbb{R}$, многомерное определение дифференцируемости соответствует одномерному. Запишем эти определения:

$$\begin{aligned}
\exists f'(x_0) &\iff \exists \lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} \\
&\iff \exists \lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h} \\
&\iff \frac{f(x_0+h)-f(x_0)}{h} = f'(x_0)+o(1),\ h\to 0, \\
&\iff f(x_0+h) = f(x_0)+f'(x_0)h+o(h),\ h\to 0.
\end{aligned}$$

Последнее условие равносильно многомерному определению, так как отображение

$$h \mapsto f'(x_0)h$$

задает линейный оператор на $\mathbb{R}$.

3.1 Базовые свойства

Единственность

Лемма 3.1. Если $A \in \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ и

$$\lim_{h\to 0} \frac{\|Ah\|}{\|h\|} = 0,$$

то $A = 0$.

Доказательство. Заметим, что значение под пределом не зависит от нормы вектора h. Зафиксируем направляющий вектор $h \neq 0$ и будем менять его норму. По условию

$$\frac{\|A(th)\|}{\|th\|} \to 0 \quad \text{при } t \to 0.$$

Но

$$\frac{\|A(th)\|}{\|th\|} = \frac{\|Ah\|}{\|h\|}$$

— константа. Значит,
$$\frac{\|Ah\|}{\|h\|} = 0,$$
откуда $Ah = 0$. Поскольку вектор h был произвольным, отсюда следует, что $A = 0$. ■

Утверждение 3.2 (единственность дифференциала). Пусть A_1 и A_2 — дифференциалы функции f в точке x. Тогда $A_1 = A_2$.

Доказательство. По определению,

$$f(x+h) = f(x) + A_1 h + r_1(h),$$
$$f(x+h) = f(x) + A_2 h + r_2(h).$$

Тогда, вычитая из первого уравнения второе, получаем

$$(A_1 - A_2)h = r_1(h) - r_2(h) = o(h) \quad \text{при } h \to 0,$$

по определению дифференциала. Значит, по лемме, $A_1 = A_2$. ■

Тривиально дифференцируемые отображения

Утверждение 3.3 (дифференцирование аффинных отображений). Пусть $A \in \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$, $b \in \mathbb{R}^n$,
$$f(x) = Ax + b.$$
Тогда функция f всюду дифференцируема с дифференциалом A:

$$\mathrm{d}_x f = A.$$

Доказательство. Проверим, что нужная разность вблизи x мала:

$$f(x_0 + h) - f(x_0) = b + A(x_0 + h) - b - Ax_0 = Ah.$$

Отображение A, таким образом, отвечает определению дифференциала в точке. По единственности дифференциала
$$A = \mathrm{d}_x f.$$
■

Утверждение 3.4 (линейность дифференциала). Пусть функции f, g дифференцируемы в x_0. Тогда для любых $\alpha, \beta \in \mathbb{R}$ имеет место равенство

$$\mathrm{d}_{x_0}(\alpha f + \beta g) = \alpha \mathrm{d}_{x_0} f + \beta \mathrm{d}_{x_0} g.$$

Доказательство. По определению дифференциала. ■

Дифференциал композиции

Теорема 3.5 (дифференцирование композиции). Пусть функция f дифференцируема в x_0, g дифференцируема в $f(x_0)$. Тогда функция $g \circ f$ дифференцируема в

x_0, причём
$$\mathrm{d}_{x_0}(g \circ f) = (\mathrm{d}_{f(x_0)} g) \cdot (\mathrm{d}_{x_0} f).$$

Доказательство. Напишем формулы для значения композиции $g \circ f$ в точке x_0 через известные дифференциалы:

$$\begin{aligned}
g\big(f(x_0+h)\big) &= g\big(f(x_0) + [\mathrm{d}_{x_0} f]h + r(h)\big) \\
&= \big[v := [\mathrm{d}_{x_0} f]h + r(h)\big] \\
&= g\big(f(x_0) + v\big) \\
&= g\big(f(x_0)\big) + \big[\mathrm{d}_{f(x_0)} g\big]\, v + \widetilde{r}(v) \\
&= g\big(f(x_0)\big) + \big[\mathrm{d}_{f(x_0)} g\big]\, [\mathrm{d}_{x_0} f]h + \big[\mathrm{d}_{f(x_0)} g\big]\, r(h) + \widetilde{r}(v) \\
&= \Big[R(h) := \big[\mathrm{d}_{f(x_0)} g\big]\, r(h) + \widetilde{r}(v)\Big] \\
&= g\big(f(x_0)\big) + \big[\mathrm{d}_{f(x_0)} g\big]\, [\mathrm{d}_{x_0} f]h + R(h).
\end{aligned}$$

Покажем, что
$$\|R(h)\| = o(\|h\|).$$

Для этого устремим h к нулю и покажем, что частное
$$\big[\mathrm{d}_{f(x_0)} g\big]\, \frac{r(h)}{\|h\|} + \frac{\widetilde{r}(v)}{\|h\|}$$
стремится к нулю. Во-первых,
$$\frac{\|r(h)\|}{\|h\|} \to 0$$
по определению r. Во-вторых,
$$\widetilde{r}(v) = \widetilde{p}(v) \cdot \|v\|$$
для некоторой стремящейся при $v \to 0$ к нулю функции $\widetilde{p}$. Тогда
$$\begin{aligned}
\frac{\widetilde{r}(h)}{\|h\|} &= \frac{\widetilde{p}(h) \cdot \|v\|}{\|h\|} \\
&\leqslant p(h) \cdot \left(\mathrm{d}_{x_0} f + \frac{r(h)}{\|h\|}\right).
\end{aligned}$$

Последнее выражение тоже стремится к нулю по определению функций p и r. ∎

Похожее доказательство того же факта можно получить, если использовать определение дифференцируемости через предел.

Дифференцирование координатных функций

Многие утверждения про векторнозначные функции можно сводить к скалярным.

Определение. Пусть $f\colon X \to \mathbb{R}^n$ — функция, $\{e_i\}$ — стандартный базис $\mathbb{R}^n$. Тогда

$$f = \sum_{i=1}^{n} e_i f_i$$

для некоторых функций

$$f_i\colon X \to \mathbb{R},$$

которые называются *координатными* функциями f.

Утверждение 3.6. Функция $f\colon X \to \mathbb{R}^n$ дифференцируема в точке x_0 только и только тогда, когда все координатные функции дифференцируемы в x_0.

Доказательство. Пусть координатные функции $\{f_k\}$ дифференцируемы, $\{e_i\}$ — стандартный базис $\mathbb{R}^n$. Тогда

$$f = \sum e_i f_i.$$

Векторы базиса можно отождествить с линейными функциями из $\mathbb{R}$ в $\mathbb{R}^n$. Значит, функция f — линейная комбинация композиций дифференцируемых функций.

Пусть теперь f дифференцируема в x_0. Координатная функция f_i записывается в виде композиции дифференцируемых функций:

$$f_i = e_i^{\mathrm{T}} f. \qquad \blacksquare$$

4 Дифференцируемые скалярные функции

4.1 Производная по направлению

Теперь мы можем сфокусироваться на изучении скалярных функций на векторных пространствах.

Определение. Пусть $f\colon X \to \mathbb{R}$ — вещественнозначная функция, $x_0 \in \operatorname{Int} X$ — внутренняя точка её области определения. *Производной f по направлению u* в точке x_0 называется предел

$$\partial_u f(x_0) = D_u f(x_0) = \lim_{\substack{t \to 0 \\ x_0 + tu \in X}} \frac{f(x_0 + tu) - f(x_0)}{t},$$

если он существует.

Утверждение 4.1. Пусть функция $f\colon X \to \mathbb{R}^n$ дифференцируема в x_0. Тогда по любому направлению $u \in \mathbb{R}^m$ существует производная, равная произведению дифференциала на это направление:

$$\partial_u f(x_0) = [\mathrm{d}_{x_0} f] u.$$

Доказательство. Посчитаем по определению. Для всех скаляров $t \in \mathbb{R}$ таких, что $x_0 + tu \in X$, имеем

$$\frac{f(x_0+tu)-f(x_0)}{t} = \frac{[\mathrm{d}_{x_0} f](tu)+r(tu)}{t} = [\mathrm{d}_{x_0} f]u + \frac{r(tu)}{t}.$$

Если t стремится к нулю, то по определению r

$$\frac{\|r(tu)\|}{t} = \frac{\|\alpha(tu)\| \cdot \|tu\|}{t} = \|\alpha(tu)\| \cdot \|u\|$$

где $\alpha\colon X \to \mathbb{R}^n$, $\|\alpha(tu)\| \to 0$ при $t \to 0$. ■

Частные производные

Часто достаточно рассматривать производные по направлениям базисных векторов.

Определение. Пусть $f\colon \Omega \to \mathbb{R}^n$ — функция, $e_1, \ldots, e_n$ — стандартный базис пространства $\mathbb{R}^n$. Тогда *частной производной* функции f в точке x по k-ой переменной называется производная по направлению e_k:

$$\partial_k f(x) := \partial_{e_k} f(x).$$

В случае, когда переменные нумеровать неудобно, будем также использовать обозначение

$$\frac{\partial f}{\partial x_k}(x) = \partial_k f(x).$$

Если расписать это определение более явно, получится

$$\partial_k f(x) = \lim_{\varepsilon \to 0} \frac{f(x_1, \ldots, x_{k-1}, x_k + \varepsilon, x_{k+1}, \ldots, x_m) - f(x_1, \ldots, x_m)}{\varepsilon}.$$

Следствие 4.2. Если f дифференцируема в точке x, в этой точке у f существуют все частные производные.

Доказательство. Если функция дифференцируема в точке, она дифференцируема в ней по любому направлению (утверждение 4.1). ■

Примеры.

1. Пусть $f = (t_1, t_2) \mapsto t_1^3 \cdot t_2^2 + 3$. Тогда у неё будут такие частные производные:

$$\partial_1 f = 3t_1^2 t_2^2, \quad \partial_2 f = 2t_1^3 t_2.$$

2. Пусть

$$F(x,y) = \begin{cases} 0, & (x-1)^2 + y^2 \leqslant 1, \\ 0, & x = 0, \\ 0, & (x+1)^2 + y^2 \leqslant 1, \\ 1, & \text{иначе}. \end{cases}$$

Можно понять, что в точке ноль F имеет производные по любому направлению, причём они равны нулю. Однако F не непрерывна в нуле, а потому не дифференцируема в нем.

Дифференцируемость и частные производные

В прошлом разделе мы поняли, что дифференцируемости функции достаточно для существования всех частных производных. Однако чтобы это работало в обратную сторону, на частные производные необходимо накладывать дополнительные условия.

Теорема 4.3. Пусть X — подмножество $\mathbb{R}^m$, $f\colon X \to \mathbb{R}^n$ — функция. Предположим, все частные производные f в некоторой окрестности внутренней точки $x \in \operatorname{Int} X$ существуют и непрерывны. Тогда f дифференцируема в точке x.

Доказательство. Можно считать, что f — скалярная функция, поскольку дифференцируемость в точке эквивалентна дифференцируемости всех координатных функций. Пусть l — строка из частных производных функции f в точке x. Мы хотим показать, что $l = \mathrm{d}_x f$, то есть найти такую функцию r, что

$$f(x+h) - f(x) = lh + r(h),$$

где

$$\frac{\|r(h)\|}{\|h\|} \to 0 \quad \text{при} \;\; h \to 0.$$

Пусть вектор

$$h_1 := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m$$

стремится к нулю. Определим последовательность векторов в пространстве области определения $\mathbb{R}^m$, каждый из которых несколько ближе к нулю, чем h_1:

$$h_2 := \begin{pmatrix} 0 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, \quad \ldots, \quad h_m := \begin{pmatrix} 0 \\ \vdots \\ 0 \\ y_m \end{pmatrix}.$$

Теперь

$$\begin{aligned} f(x+h_1) - f(x) = f(x+h_1) - f(x+h_2) + f(x+h_2) - f(x+h_3) + \cdots + \\ - f(x+h_m) + f(x+h_m) - f(x). \end{aligned}$$

Все значения функции f здесь определены для правильно подобранного h_1, поскольку x — внутренняя точка области определения f.

Параметризуем отрезок $[x + h_1, x + h_2]$, чтобы воспользоваться теоремой Лагранжа. Определим вещественнозначную функцию

$$F_1(t) := f((x + h_1)t + (x + h_2)(1 - t))$$

на отрезке $[0, 1]$. По определению x — внутренняя точка области определения f, поэтому можно считать, что F_1 действительно корректно определена для всех значений аргумента.

Итак, по теореме Лагранжа хочется найти значение $t = \xi$ такое, что

$$F_1'(\xi) = f(x + h_1) - f(x + h_2).$$

Можно видеть, что из-за устройства векторов $\{h_i\}$ здесь требуется только существование частной производной f по первой переменной. Она действительно существует по условию.

Вычислим производную F_1' другим способом. Пусть

$$v_1 := (x + h_1)\xi + (x + h_2)(1 - \xi).$$

Тогда

$$\begin{aligned} F_1'(\xi) &= (\mathrm{d}_{v_1} f) \cdot (x + h_1 - x - h_2) \\ &= (\mathrm{d}_{v_1} f) \cdot (h_1 - h_2) \\ &= (\mathrm{d}_{v_1} f) \cdot e_1 \cdot y_1 \\ &= (\partial_1 f)(v_1) \cdot y_1. \end{aligned}$$

Таким образом,

$$f(x + h_1) - f(x + h_2) = (\partial_1 f)(v_1) \cdot y_1.$$

Расписав аналогичным образом остальные разности, получаем

$$f(x + h_1) - f(x) = (\partial_1 f)(v_1) \cdot y_1 + \cdots + (\partial_m f)(v_m) \cdot y_m.$$

Справа стоит почти нужное выражение. Теперь мы можем использовать непрерывность частных производных в точке x и разложить каждую из них в сумму

$$(\partial_i f)(v_i) =: (\partial_i f)(x) + \alpha_i(v_i - x),$$

где α_i — некоторая функция, которая стремится к нулю при стремящемся к нулю аргументе. Таким образом,

$$f(x + h_1) - f(x) = lh_1 + \left[\alpha_1(v_1 - x)y_1 + \cdots + \alpha_m(v_m - x)y_m\right] = lh_1 + r(h_1).$$

Осталось показать, что

$$\alpha_1(v_1 - x)y_1 + \cdots + \alpha_m(v_m - x)y_m = o(\|h_1\|).$$

По КБШ имеем

$$|\alpha_1(v_1 - x)y_1 + \dots + \alpha_m(v_m - x)y_m| \leqslant \|h_1\| \cdot \|(\alpha_1(v_1 - x), \dots, \alpha_m(v_m - x))\|.$$

В то же время функций $\{\alpha_i\}$ конечное число, и они все стремятся к нулю. Значит, выражение справа стремится к нулю при $h_1 \to 0$. ∎

4.2 Экстремумы и градиент

Определение. Пусть вещественнозначная функция f дифференцируема в точке $x \in \Omega$. Тогда её *градиентом* в точке x называется вектор из её частных производных в этой точке:

$$\operatorname{grad}_x f = \sum_{i=1}^{m} \partial_i f(x) e_i = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_m f(x) \end{pmatrix}.$$

Иначе говоря, градиент — просто транспонированная матрица Якоби скалярной функции.

Упражнение. Пусть f дифференцируема в x, v — любой вектор. Тогда

$$\partial_v f(x) = (\operatorname{grad}_x f)^{\mathrm{T}} v.$$

Производная по направлению максимальна в направлении градиента и минимальна в противоположном направлении:

Утверждение 4.4. Пусть f — скалярная функция, дифференцируемая в некоторой точке $x \in \mathbb{R}^n$, причём её градиент там не равен нулю. Определим

$$v(x) := \frac{\operatorname{grad}_x f}{\|\operatorname{grad}_x f\|}.$$

Тогда для любого единичного вектора $u \in S^{m-1}$ верно

$$\partial_u f(x) \leqslant \partial_{v(x)} f(x).$$

Доказательство. Оценим левую часть по КБШ:

$$\begin{aligned} \partial_u f(x) &= (\operatorname{grad}_x f)^{\mathrm{T}} u \\ &\leqslant \|\operatorname{grad}_x f\| \|u\| \\ &= \|\operatorname{grad}_x f\|. \end{aligned}$$

Но это ровно правая часть:

$$(\operatorname{grad}_x f)^{\mathrm{T}} v(x) = \|\operatorname{grad}_x f\|.$$ ∎

Определение. Говорят, что функция f имеет в x *локальный минимум*, если существует такое $\delta > 0$, что из условия $\|\widetilde{x} - x\| < \delta$ следует $f(\widetilde{x}) \geqslant f(x)$ для всех $\widetilde{x} \in \mathbb{R}^n$.

Аналогично определяется локальный максимум.

Утверждение 4.5. Пусть скалярная функция f дифференцируема в x и имеет там локальный экстремум. Тогда её градиент в точке x равен нулю.

Доказательство. Пусть f имеет в x локальный максимум. Предположим, $\operatorname{grad}_x f \neq 0$. Рассмотрим производную по направлению градиента, которая, по предыдущей оценке, максимальна среди всех производных по направлению. Определим, как раньше,

$$v(x) := \frac{\operatorname{grad}_x f}{\|\operatorname{grad}_x f\|}.$$

Тогда

$$\begin{aligned}
\|\operatorname{grad}_x f\| > 0 &\implies \lim_{t\to 0} \frac{f(x+tv(x)) - f(x)}{t} > 0 \\
&\implies \exists t_0 > 0\ \forall t \in (0, t_0) : \frac{f(x+tv(x)) - f(x)}{t} > 0 \\
&\implies \exists t_0 > 0\ \forall t \in (0, t_0) : f(x+tv(x)) > f(x).
\end{aligned}$$

Значит $f(x)$ — не локальный максимум. Противоречие.

В доказательстве для случая минимума нужно рассматривать производную *против* направления градиента. В остальном оно аналогично. ■

4.3 Аналоги формулы Лагранжа

Пример 4.1. Формула Лагранжа

$$f(x) - f(y) = (\mathrm{d}_v f)(x-y), \quad v \in [x, y],$$

вообще говоря, неверна для функций $f\colon \mathbb{R}^m \to \mathbb{R}^n$, если $n \neq 1$. Возьмем, например, стандартную намотку прямой на окружность

$$f := t \mapsto \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}.$$

Разность:

$$f(\pi) - f(0) = \begin{pmatrix} -1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

Дифференциал:

$$\mathrm{d}_\theta f = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

В частности,

$$(\mathrm{d}_\theta f)(\pi - 0) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}\pi.$$

Если бы формула была верна, то получилось бы, что

$$\begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}\pi.$$

Считая квадраты норм, получаем
$$4 = \pi^2,$$
что неверно (например, потому что π — это длина верхней полуокружности, а 2 — это длина вписанной в нее ломаной из двух звеньев с концами 0, i, -1, отсюда легко видеть, что $\pi > 2$).

Тем не менее, ничего не мешает нам при желании параметризовать отрезок в пространстве области определения и применить одномерную теорему Лагранжа на нём. Мы уже делали так в доказательстве достаточного условия дифференцируемости функции в терминах её частных производных; повторим это рассуждение и здесь.

Формула Лагранжа на отрезках

Теорема 4.6 (обобщение теоремы Лагранжа). Пусть скалярная функция $f\colon X \to \mathbb{R}$, где $X \subset \mathbb{R}^m$, дифференцируема в каждой точке отрезка $[x, y]$. Тогда существует вектор $v \in [x, y]$ такой, что
$$f(x) - f(y) = [\mathrm{d}_v f](x - y).$$

Доказательство. Рассмотрим композицию f с линейной параметризацией отрезка $[x, y]$:
$$\varphi(t) = f(tx + (1 - t)y).$$
Функция φ корректно определена и дифференцируема как композиция дифференцируемых отображений. По теореме Лагранжа выберем такое значение параметра $t_0 \in [0, 1]$, что
$$\varphi(1) - \varphi(0) = \varphi'(t_0).$$
С другой стороны, если
$$v := t_0 x + (1 - t_0)y,$$
то, по правилу дифференцирования композиции,
$$\varphi'(t_0) = [\mathrm{d}_v f](x - y),$$
что и требовалось. ■

Неравенство Лагранжа

Но формула Лагранжа позволяет получать оценки и для *векторнозначных* функций.

Теорема 4.7 (неравенство Лагранжа). Пусть векторнозначная функция $f\colon X \to \mathbb{R}^n$, где $X \subset \mathbb{R}^m$, дифференцируема в каждой точке отрезка $[x, y]$. Тогда
$$\|f(x) - f(y)\| \leqslant \sup_{v \in [x,y]} \|\mathrm{d}_v f\| \cdot \|x - y\|$$

В случае со стандартной намоткой эта теорема говорит, что длина хорды не больше длины дуги.

Доказательство. Хотим применить обобщённую теорему Лагранжа к проекции функции f на некоторое одномерное подпространство. Зафиксируем вектор

$$k := f(x) - f(y).$$

Рассмотрим скалярный квадрат k:

$$\|k\|^2 = \|f(x) - f(y)\|^2 = \langle f(x), k\rangle - \langle f(y), k\rangle.$$

Продифференцируем возникшее скалярное отображение:

$$\mathrm{d}(k^{\mathrm{T}} f(t)) = (\mathrm{d}k^{\mathrm{T}}) \cdot (\mathrm{d}f(t)) = k^{\mathrm{T}} \mathrm{d}f(t).$$

Теперь можно сказать, что найдётся точка z на отрезке $[x, y]$ такая, что

$$\langle f(x), k\rangle - \langle f(y), k\rangle = \langle \mathrm{d}_z f, k\rangle(x - y).$$

Используя КБШ, получаем оценку на модуль этого выражения:

$$\|\langle \mathrm{d}_z f, k\rangle(x - y)\| \leqslant \|\mathrm{d}_z f\| \cdot \|k\| \cdot \|x - y\|.$$

В итоге,

$$\|k\|^2 \leqslant \|\mathrm{d}_z f\| \cdot \|k\| \cdot \|x - y\|.$$

При $k = 0$ утверждение верно. Иначе, сокращая на $\|k\|$, как раз получаем, что для найденной точки z выполняется требуемая оценка. ∎

Определение. Открытое связное подмножество любого топологического пространства называется *областью*.

Следствие 4.8 (функция с нулевым дифференциалом постоянна). Пусть дана функция $f\colon \Omega \to \mathbb{R}^n$, где $\Omega \subset \mathbb{R}^m$ — область. Предположим, дифференциал f везде существует и равен нулю. Тогда f постоянна на области определения.

Доказательство. Возьмём любую точку $x_0 \in \Omega$ и положим $c := f(x_0)$. Рассмотрим *множество уровня*

$$U := \{x \in \Omega \mid f(x) = c\}.$$

Отметим, что U непусто, поскольку $x_0 \in U$. Покажем, что, кроме того, множество U открыто и замкнуто; тогда равенство $U = \Omega$ будет следовать из связности области Ω.

Начнём с открытости. Пусть $x \in U$. Поскольку область Ω открыта, точка x лежит в некотором шаре N в Ω. Пусть $x' \in N$ — другая точка той же окрестности. По неравенству Лагранжа,

$$\|f(x') - f(x)\| \leqslant \sup_{x'' \in [x, x']} \|\mathrm{d}_{x''} f\| \cdot \|x' - x\| = 0,$$

так как $\|\mathrm{d}_{x''} f\| = 0$. Значит, $x' \in U$, что и требовалось.

С замкнутостью проще: поскольку f дифференцируема, она непрерывна, и в любой предельной точке U значение f тоже будет равно c. ∎

4.4 Перестановочность частных производных

Определение. Пусть $f\colon X \to \mathbb{R}^n$, где $X \subset \mathbb{R}^m$; $i_1, \dots i_k \in \{1, 2, \dots, m\}$. Определим частные производные *высших порядков* по индукции:

$$\partial_{i_1 i_2 \dots i_k} f = \partial_{i_1 i_2 \dots i_{k-1}} (\partial_{i_k} f).$$

Пример 4.2. Пусть

$$f(x, y) := \begin{cases} xy\frac{x^2-y^2}{x^2+y^2}, & (x, y) \neq 0 \\ 0, & (x, y) = 0 \end{cases}$$

тогда можно проверить, что $\partial_{12} f(0) \neq \partial_{21} f(0)$.

Найдем достаточные условия для того, чтобы можно было менять порядок дифференцирования:

Теорема 4.9 (изменение порядка дифференцирования). Пусть дана функция $f\colon \Omega \to \mathbb{R}^n$, где $\Omega \subset \mathbb{R}^m$ — область. Предположим, что f везде имеет непрерывные частные производные $\partial_{kl} f$ и $\partial_{lk} f$. Тогда для любого $x \in \Omega$ выполнено

$$\partial_{kl} f(x) = \partial_{lk} f(x).$$

Доказательство. Пусть $k = 1$ и $l = 2$. Из доказательства будет видно, что можно считать, что $m = 2$. Так как задача сводится к координатным функциям, можно также считать, что $n = 1$. Зафиксируем точку $(x_1, x_2) \in \Omega$ и при малых $h_1, h_2 \in \mathbb{R}$ рассмотрим выражение

$$F(h_1, h_2) := f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) - f(x_1, x_2 + h_2) + f(x_1, x_2).$$

Докажем, что оно будет стремиться одновременно к каждой из двух частных производных второго порядка. Для начала заметим, что

$$F(h_1, h_2) = \varphi(1) - \varphi(0) = \widetilde{\varphi}(1) - \widetilde{\varphi}(0),$$

где

$$\begin{aligned} \varphi(t) &:= f(x_1 + th_1, x_2 + h_2) - f(x_1 + th_1, x_2), \\ \widetilde{\varphi}(t) &:= f(x_1 + h_1, x_2 + th_2) - f(x_1, x_2 + th_2) \end{aligned}$$

— гладкие функции на отрезке $[0, 1]$. Применяя к каждой из них теорему Лагранжа, получаем для некоторых $t_0, \widetilde{t_0} \in [0, 1]$

$$\begin{aligned} F(h_1, h_2) &= \varphi'(t_0) = \partial_1 f(x_1 + t_0 h_1, x_2 + h_2) \cdot h_1 - \partial_1 f(x_1 + t_0 h_1, x_2) \cdot h_1, \\ F(h_1, h_2) &= \widetilde{\varphi}'(\widetilde{t_0}) = \partial_2 f(x_1 + h_1, x_2 + \widetilde{t_0} h_2) \cdot h_2 - \partial_2 f(x_1, x_2 + \widetilde{t_0} h_1) \cdot h_2. \end{aligned}$$

Применяя теорему Лагранжа еще раз к функциям

$$s \mapsto \partial_1 f(x_1 + t_0 h_1, x_2 + sh_2) \quad \text{и} \quad s \mapsto \partial_2 f(x_1 + sh_1, x_2 + \widetilde{t_0} h_1),$$

найдем точки $s_0, \widetilde{s}_0 \in [0,1]$ такие, что

$$F(h_1,h_2) = \partial_{21} f(x_1 + t_0 h_1, x_2 + s_0 h_2) h_1 h_2,$$
$$F(h_1,h_2) = \partial_{12} f(x_1 + \tilde{s}_0 h_1, x_2 + \tilde{t}_0 h_2) h_2 h_1.$$

Осталось устремить h_1, h_2 к нулю и воспользоваться непрерывностью частных производных. ■

Утверждение 4.10. Пусть функция $f\colon \Omega \to \mathbb{R}^n$, где $\Omega \subset \mathbb{R}^m$ — область, имеет непрерывные частные производные $\partial_{\sigma(i_1)\dots\sigma(i_k)} f$ для любой перестановки σ некоторого набора индексов $i_1, \dots, i_k$. Тогда все производные $\partial_{\sigma(i_1)\dots\sigma(i_k)} f$ совпадают.

Доказательство. По предыдущему утверждению любые два индекса при дифференцировании можно переставлять. Тогда утверждение следует из того, что группа перестановок порождается транспозициями. ■

4.5 Формула Тейлора

Определение. Будем говорить, что функция $f\colon X \subset \mathbb{R}^m \to \mathbb{R}^n$ принадлежит:

- *классу гладкости* $C^0(Y)$, если она непрерывна на множестве Y;
- *классу гладкости* $C^k(Y)$, где $k \geqslant 1$, если она дифференцируема в любой точке множества Y, а её дифференциал $x \mapsto \mathrm{d}_x f$ принадлежит классу гладкости $C^{k-1}(Y)$.[2]

Если $Y = X$, то $C^k(Y)$ обычно сокращается до C^k. Отдельно выделяется C^∞, который называется классом *бесконечно гладких* или просто *гладких* функций.

Утверждение 4.11. Пусть $f\colon X \subset \mathbb{R}^m \to \mathbb{R}^n$ — функция класса C^k. Тогда все частные производные f существуют и непрерывны.

Доказательство. Любая частная производная функции f — это координатная функция дифференциала $x \mapsto \mathrm{d}_x f$. Дифференциал непрерывен в топологии нормы линейного отображения, а поскольку любые две нормы на конечномерном векторном пространстве эквивалентны (теорема 2.4), он непрерывен в топологии метрики, которая сопоставляет матрице сумму абсолютных значений её коэффициентов. Можно видеть, что из этого следует непрерывность любой частной производной. ■

Обозначение. Пусть

$$\alpha := (\alpha_1, \dots, \alpha_m)$$

— набор неотрицательных целых чисел. Обозначим

$$|\alpha| := \alpha_1 + \dots + \alpha_m,$$

[2] При $k = 1$ имеется в виду непрерывность в стандартной топологии. Поскольку любые две нормы на конечномерном вещественном пространстве задают одинаковые топологии (страница 4), нам не важно, измеряется непрерывность дифференциала стандартной нормой линейного отображения или какой-то ещё.

и

$$\partial_\alpha := \partial_1^{\alpha_1} \dots \partial_n^{\alpha_m}.$$

В этом смысле множество α называется *мультииндексом*.

Определение. Пусть $x_1, x_2 \in \mathbb{R}^m$. *Отрезком* между x_1 и x_2 называется множество

$$[x_1, x_2] := \big\{(1-t)x_1 + tx_2 \mid t \in [0,1]\big\}.$$

Теорема 4.12 (формула Тейлора для функций нескольких переменных). Пусть скалярная функция $f\colon X \subset \mathbb{R}^m \to \mathbb{R}$ принадлежит классу гладкости $C^{k+1}[x_0, x_1]$ для некоторого отрезка $[x_0, x_1]$, где $x_0, x_1 \in \mathbb{R}^m$. Тогда существует функция $r\colon \mathbb{R}^m \to \mathbb{R}$ такая, что для любого $x \in [x_0, x_1]$

$$f(x) = \sum_{|\alpha| \leqslant k} \frac{h_1^{\alpha_1} \dots h_m^{\alpha_m}}{|\alpha|!} \partial_\alpha f(x_0) + r(x - x_0),$$

где

$$h = (h_1, \dots, h_m)^{\mathrm{T}} := x - x_0,$$

и

$$\lim_{\substack{x \to x_0 \\ x \in [x_0, x_1]}} \frac{r(x - x_0)}{\|x - x_0\|^k} = 0.$$

Из доказательства будет видно, что для функции r можно получить несколько формул.

Доказательство. Про точку x_1 можно забыть. Параметризуем аргумент функции f, определив

$$\varphi(t) := f(x_0 + (x - x_0)t).$$

Функция φ принадлежит классу $C^{k+1}[0,1]$. Разложим φ по одномерной формуле Тейлора в нуле: для некоторой остаточной функции $\widetilde{r}$ верно

$$\varphi(1) = \varphi(0) + \frac{1}{1!}\varphi^{(1)}(0) + \dots + \frac{1}{k!}\varphi^{(k)}(0) + \widetilde{r}(1).$$

Найдём для произвольного l производную $\varphi^{(l)}(0)$. Если $l = 1$, то

$$\varphi^{(1)}(t) = (\mathrm{d}_{x_0 + (x - x_0)t} f) \cdot (x - x_0).$$

Хочется теперь продифференцировать эту композицию ещё раз:

$$\varphi^{(2)}(t) = (\mathrm{d}_t(\mathrm{d}_{x_0 + (x - x_0)t} f)) \cdot (x - x_0).$$

Если рассуждать в этом направлении, для дифференциалов старших порядков придётся научиться умножать гиперкубы из производных на столбцы координат. Но мы сделаем эквивалентное по смыслу действие: распишем скалярное произведение в производной $\varphi^{(1)}$ как

$$\varphi^{(1)}(t) = \sum_{i_1=1}^{m} \partial_{i_1} f(x_0 + (x - x_0)t) \cdot h_{i_1}.$$

Теперь видно, что

$$\varphi^{(2)}(t) = \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \partial_{i_2}\partial_{i_1} f(x_0 + (x - x_0)t) \cdot h_{i_1} h_{i_2},$$

и так далее вплоть до $\varphi^{(k)}$. Пользуясь коммутативностью операции взятия частных производных и суммируя по мультииндексам, получаем такую сумму:

$$\varphi(1) = \sum_{|\alpha| \leqslant k} \frac{h_1^{\alpha_1} \ldots h_m^{\alpha_m}}{|\alpha|!} \partial_\alpha f(x_0) + \widetilde{r}(1).$$

Это почти равенство из условия, ведь $\varphi(1) = f(x)$. При фиксированных концах отрезка $[x_0, x_1]$ и остальных обозначениях, величина $\widetilde{r}(1)$ зависит только от x, поэтому можно считать, что

$$\widetilde{r}(1) = r(x - x_0)$$

для некоторой скалярной функции r. Таким образом,

$$f(x) = \sum_{|\alpha| \leqslant k} \frac{h_1^{\alpha_1} \ldots h_m^{\alpha_m}}{|\alpha|!} \partial_\alpha f(x_0) + r(x - x_0).$$

Докажем теперь асимптотику остаточной функции r. Выведенные нами формулы для $f(x)$ верны, в частности, для $f(x_1)$, поэтому положим теперь для удобства

$$\varphi(t) := f(x_0 + (x_1 - x_0)t)$$

и устремим x к x_0. По оценкам для одномерной формулы Лагранжа,

$$\lim_{t \to 0} \frac{\widetilde{r}(t)}{t^k} = 0.$$

Но, при сделанных нами переобозначениях,

$$\widetilde{r}(t) = r((x_1 - x_0)t)$$

и

$$t = \frac{x - x_0}{x_1 - x_0} = (x - x_0) \cdot \mathrm{const},$$

а потому

$$\lim_{t \to 0} \frac{r(x - x_0)}{\|x - x_0\|^k} = 0,$$

что завершает доказательство. ■

Итак, теперь видно, что остаток $\widetilde{r}$ можно записать, например, в интегральной форме:

$$\widetilde{r}(1) = \frac{1}{k!}\int\limits_0^1 \varphi^{(k+1)}(y)(1-y)^k\,\mathrm{d}y.$$

Раскладывая $\varphi^{(k+1)}$, как мы уже это делали, получаем одну из возможных формул для r:

$$r(x-x_0) = \frac{1}{k!}\int\limits_0^1 \sum_{|\alpha|=k+1} h_1^{\alpha_1}\dots h_m^{\alpha_m}\partial_\alpha f(x_0+(x-x_0)y)\cdot(1-y)^k\,\mathrm{d}y.$$

Остаток в форме Пеано

Неудобство выведенной нами формулы Тейлора в том, что функция r не стремится к нулю равномерно по направлению $x-x_0$. Но это можно исправить, если записать остаток в форме Лагранжа и оценить результат.

Утверждение 4.13 (остаточный член в форме Пеано). В условиях теоремы 4.12, функцию $r\colon \mathbb{R}^m \to \mathbb{R}$ можно взять такой, что

$$\lim_{v\to x_0}\frac{r(v-x_0)}{\|v-x_0\|^k} = 0.$$

Доказательство. Запишем остаток в форме Лагранжа: существует скаляр $\theta \in [0,1]$ такой, что

$$r(x-x_0) = \frac{1}{(k+1)!}\sum_{|\alpha|=k+1} h_1^{\alpha_1}\dots h_m^{\alpha_m}\partial_\alpha f(x_0+(x-x_0)\theta).$$

Пусть

$$h_* := \max_j |h_j|,$$

тогда модуль суммы $r(x-x_0)$ можно оценить сверху как

$$|r(x-x_0)| \leqslant \frac{h_*^{k+1}}{(k+1)!}\sum_{|\alpha|=k+1}\partial_\alpha f(x_0+(x-x_0)\theta).$$

По непрерывности каждой частной производной порядка $k+1$,

$$|r(x-x_0)| = o(h_*^{k+1}) = o(\|x-x_0\|^{k+1}),$$

что и требовалось доказать. ∎

4.6 Экстремумы и гессиан

Определение. Оператор $A \in \mathrm{Hom}(\mathbb{R}^n,\mathbb{R}^n)$ называется *симметричным*, если

$$\forall x,y\in\mathbb{R}^n : \langle Ax,y\rangle = \langle x,Ay\rangle.$$

Замечание. Симметричный оператор A однозначно восстанавливается по своей квадратичной форме, так как

$$\langle Ax, y\rangle = \frac{1}{2}\Big(\langle A(x+y), (x+y)\rangle - \langle Ax, x\rangle - \langle Ay, y\rangle\Big).$$

Из этого равенства следует, что квадратичная форма определяет элементы матрицы

$$a_{ij} = \langle Ae_j, e_i\rangle$$

оператора A, а значит, сам оператор.

Упражнение. Следующие условия равносильны:

(1) оператор A симметричен;

(2) $a_{ij} = a_{ji}$ для всех i, j;

(3) $A^{\mathrm{T}} = A$.

Определение. Будем писать:

- $A \geqslant 0$ (A *неотрицателен*), если A симметричен и $\langle Ax, x\rangle \geqslant 0$ для всех $x \in \mathbb{R}^n$.
- $A > 0$ (A *положителен*), если A симметричен и $\langle Ax, x\rangle > 0$ для всех $x \in \mathbb{R}^n \setminus \{0\}$.

Определение. Отображение

$$x \mapsto \langle Ax, x\rangle$$

из $\mathbb{R}^n$ в $\mathbb{R}$ называется *квадратичной формой* оператора A.

Нам несколько раз понадобится следующая лемма.

Лемма 4.14. Пусть $A \in \mathrm{End}(\mathbb{R}^m)$ — положительный оператор. Тогда

$$\exists \delta > 0\ \forall x \in \mathbb{R}^m : \langle Ax, x\rangle \geqslant \delta \|x\|^2.$$

Доказательство. Пусть $x \in \mathbb{R}^m$ — произвольный вектор. Спроецируем его на сферу S^{m-1} и воспользуемся однородностью квадратичной формы:

$$\langle Ax, x\rangle = \|x\|^2 \left\langle \frac{Ax}{\|x\|}, \frac{x}{\|x\|} \right\rangle.$$

Значит, достаточно доказать, что для некоторого $\delta > 0$ и любого $x \in S^{m-1}$ выполняется

$$\langle Ax, x\rangle \geqslant \delta.$$

И действительно: сфера компактна; значит, непрерывная функция

$$x \mapsto \langle Ax, x\rangle$$

достигает на ней минимума. Если бы этот минимум был равен нулю, оператор A не мог бы быть положительным. ∎

Определение. Пусть функция f имеет все производные второго порядка в точке x. Тогда матрица
$$(\partial_{ij} f(x))_{ij}$$
из частных производных второго порядка в точке x называется *гессианом* функции f в точке x.

То есть, гессиан — это матрица дифференциала градиента.

Теорема 4.15 (условия экстремума в терминах гессиана). Пусть
$$f \colon \Omega \subset \mathbb{R}^m \to \mathbb{R}$$
(где Ω — область) — скалярная функция класса C^2, H — гессиан f в точке a. Тогда:

(1) если a — локальный минимум, то $\operatorname{grad}_a f = 0$ и $H \geqslant 0$;

(2) если a — локальный максимум, то $\operatorname{grad}_a f = 0$ и $H \leqslant 0$.

В обратную сторону:

(3) если $\operatorname{grad}_a f = 0$ и $H > 0$, то a — строгий локальный минимум;

(4) если $\operatorname{grad}_a f = 0$ и $H < 0$, то a — строгий локальный максимум.

Доказательство. По формуле Тейлора для функций нескольких переменных,
$$f(a+h) = f(a) + \left\langle \operatorname{grad}_a f, h \right\rangle + \tfrac{1}{2}\langle Hh, h\rangle + r(h),$$
где
$$\lim_{h\to 0} \frac{r(h)}{\|h\|^2} = 0.$$
Предположим, что a — точка локального экстремума. По утверждению 4.5,
$$\operatorname{grad}_a f = 0.$$
Пусть a — это точка локального минимума. Предположим, от противного, что $H \not\geqslant 0$. Тогда найдётся вектор $v \in \mathbb{R}^m$ такой, что
$$\langle Hv, v\rangle < 0.$$
Введём скалярный параметр t и устремим его к нулю. Подставляя vt в формулу Тейлора, получаем, с учётом равенства градиента нулю,
$$f(a+vt) - f(a) = \frac{1}{2}\langle Hv, v\rangle t^2 + r(vt).$$
Поделив обе части на t^2, получаем
$$\frac{f(a+vt)-f(a)}{t^2} \xrightarrow[t\to 0]{} \frac{1}{2}\langle Hv, v\rangle < 0,$$

поскольку
$$\frac{r(vt)}{\|t\|^2} \xrightarrow[t\to 0]{} 0.$$
Это приводит к противоречию с тем, что a — точка локального минимума. Доказательство для локального максимума аналогично.

Пусть теперь
$$\operatorname{grad}_a f = 0 \quad \text{и} \quad H > 0.$$
Для положительной квадратичной формы существует такая константа $c > 0$, что для любого вектора $v \in \mathbb{R}^m$
$$\langle Hv, v\rangle \geqslant c\|v\|^2$$
(лемма 4.14). Пусть $v \in \mathbb{R}^m \setminus 0$. Подставим v в формулу Тейлора:
$$f(a+v) - f(a) = \frac{1}{2}\langle Hv, v\rangle + r(v).$$
Поделим обе части на квадрат нормы v и устремим v к нулю:
$$\frac{f(a+v) - f(a)}{\|v\|^2} \xrightarrow[v\to 0]{} \frac{1}{2}\cdot\frac{\langle Hv, v\rangle}{\|v\|^2} \geqslant \frac{c}{2} > 0.$$
Значит, найдётся окрестность точки a такая, что для любого вектора $v \neq 0$ из этой окрестности выполняется
$$f(a+v) > f(a),$$
что и требовалось. ■

5 Гладко параметризованные многообразия

График дифференцируемой скалярной функции — это гладкая линия. А что такое график дифференцируемой функции в $\mathbb{R}^n$?

В этом разделе мы сфокусируемся на изучении основных свойств дифференцируемых отображений между евклидовыми пространствами. Нашей конечной целью станет теорема о множителях Лагранжа, позволяющая оптимизировать хорошо ведущие себя функции при дополнительных условиях; но по пути мы докажем несколько других результатов, имеющих самостоятельное значение. В частности, мы до сих пор не выяснили, при каких условиях можно дифференцировать обратные отображения.

5.1 Локальная билипшицевость

Теорема 5.1 (о локальной билипшицевости отображения с невырожденным дифференциалом). Предположим, что у функции
$$f\colon \Omega \subset \mathbb{R}^m \to \mathbb{R}^m,$$

где Ω — область, существуют непрерывные производные по каждой переменной, причём её матрица Якоби в некоторой точке a невырождена. Тогда существуют положительные константы $c, \widetilde{c} > 0$ и окрестность U точки a такие, что для любых векторов $x_1, x_2 \in U$ выполняется

$$c\|x_1 - x_2\| \leqslant \|f(x_1) - f(x_2)\| \leqslant \widetilde{c}\|x_1 - x_2\|.$$

Доказательство. Разложим функцию f по формуле Тейлора в окрестности точки a:

$$f(x) = f(a) + (\mathrm{d}_a f)(x - a) + r(x).$$

Обозначим $A := \mathrm{d}_a f$ и запишем

$$f(x_1) - f(x_2) = A(x_1 - x_2) + r(x_1) - r(x_2).$$

Применим к обеим частям обратный оператор A^{-1}:

$$A^{-1}(f(x_1) - f(x_2)) = x_1 - x_2 + A^{-1}(r(x_1) - r(x_2)).$$

С одной стороны,

$$\|A^{-1}(f(x_1) - f(x_2))\| \leqslant \|A^{-1}\| \cdot \|(f(x_1) - f(x_2))\|.$$

С другой,

$$\begin{aligned}\|A^{-1}(f(x_1) - f(x_2))\| &= \|x_1 - x_2 + A^{-1}(r(x_1) - r(x_2))\| \\ &\geqslant \|x_1 - x_2\| - \|A^{-1}(r(x_1) - r(x_2))\| \\ &\geqslant \|x_1 - x_2\| - \|A^{-1}\| \cdot \|(r(x_1) - r(x_2))\|.\end{aligned}$$

По определению функции r, разность $r(x_1) - r(x_2)$ при стремящихся к точке a точках x_1 и x_2 стремится к нулю быстрее, чем $x_1 - x_2$. Поэтому для произвольного положительного ε найдётся окрестность точки a такая, что для любых x_1 и x_2 в ней будет выполнено

$$\|A^{-1}\| \cdot \|f(x_1) - f(x_2)\| \geqslant (1 - \|A^{-1}\|\varepsilon) \cdot \|x_1 - x_2\|.$$

Так мы получили нижнюю оценку. Почти аналогично получим теперь верхнюю:

$$\begin{aligned}\|f(x_1) - f(x_2)\| &= \|A(x_1 - x_2) + r(x_1) - r(x_2)\| \\ &\leqslant \|A(x_1 - x_2)\| + \|r(x_1) - r(x_2)\| \\ &\leqslant (\|A\| + \varepsilon) \cdot \|x_1 - x_2\|.\end{aligned}$$

∎

5.2 Открытость отображения с невырожденным дифференциалом

Теорема 5.2 (об открытости отображения с невырожденным дифференциалом). Пусть функция
$$f\colon \Omega \subset \mathbb{R}^m \to \mathbb{R}^m,$$
где Ω — область, принадлежит классу C^1, причём дифференциал f в любой точке обратим. Тогда f — открытое отображение.

Доказательство. Пусть $U \subset \Omega$ — шар с центром в точке $x_0 \in \Omega$ и радиусом r_U. Предположим, y — произвольная точка в шаре V радиуса r_V около точки $y_0 := f(x_0)$. Покажем, что радиус r_V можно взять достаточно маленьким, чтобы у любой такой точки y был прообраз внутри шара U.

Минимизируем функцию
$$\varphi(t) := \|f(t) - y\|^2.$$

По теореме Вейерштрасса функция φ достигает минимума на замыкании $\overline{U}$.

Покажем, что можно выбрать радиус r_V таким, что минимум достигается во внутренней точке. Пусть минимум достигается на границе шара $\overline{U}$ в некоторой точке x:
$$\|x - x_0\| = r_U.$$

По теореме о локальной билипшицевости функции с обратимым дифференциалом (теорема 5.1), f билипшицева во всём шаре $\overline{U}$; поэтому для некоторой положительной константы c верны следующие оценки:
$$\begin{aligned}
\sqrt{\varphi(x)} &= \|f(x) - y\| \\
&= \|f(x) - f(x_0) + y_0 - y\| \\
&\geqslant \|f(x) - f(x_0)\| - \|y - y_0\| \\
&\geqslant c\|x - x_0\| - \|y - y_0\| \\
&= cr_U - \|y - y_0\| \\
&\geqslant cr_U - r_V.
\end{aligned}$$

Отметим, что
$$\varphi(x_0) \leqslant r_V^2,$$
поэтому минимум заведомо не больше r_V^2. Значит, если выбрать r_V таким, что
$$cr_U - r_V > r_V,$$
минимум будет достигаться во внутренней точке: $x \in U$.

Покажем, что этот минимум будет равен нулю. Для этого продифференцируем функцию φ (проверьте следующее равенство самостоятельно):
$$\mathrm{d}_t\varphi = 2(f(t) - y)^{\mathrm{T}}(\mathrm{d}_t f).$$

По определению, в любой точке $t \in \Omega$ оператор дифференциала $d_t f$ обратим. При этом в точке минимума x имеем $d_x\varphi = 0$ (утверждение 4.5), что влечёт

$$f(x) = y.$$

Таким образом, мы нашли прообраз точки y внутри шара U. ∎

5.3 Теорема об обратной функции

Определение. Пусть M_1 и M_2 — многообразия. Дифференцируемая биекция

$$f\colon M_1 \to M_2$$

называется *диффеоморфизмом*, если её обратная тоже дифференцируема.

Иначе говоря, диффеоморфизм — это дифференцируемый гомеоморфизм с дифференцируемой обратной.

Теорема 5.3 (об обратной функции). Пусть функция

$$f\colon \Omega \subset \mathbb{R}^m \to \mathbb{R}^m$$

(где Ω — область) класса C^1 имеет обратимый оператор дифференциала в любой точке области определения.[3] Пусть $x_0 \in \Omega$. Тогда существует окрестность U точки x_0 такая, что сужение функции f на U — диффеоморфизм между U и её образом $f(U)$, причём дифференциал обратного отображения тоже имеет класс C^1.

То есть, нам нужно доказать, что в некоторой окрестности U у функции f существует дифференцируемое обратное отображение. Если это установлено, то, дифференцируя композицию функции f с её обратной, получаем необходимое условие на дифференциал обратной функции:

$$(d_{f(x)}f^{-1})(d_x f) = I,$$

что эквивалентно

$$d_{f(x)}f^{-1} = (d_x f)^{-1}.$$

Этим мы и будем пользоваться при доказательстве теоремы. Убедитесь, что эта формула соответствует одномерному случаю.

Доказательство. Инъективность дифференциала $d_{x_0}f$ влечёт билипшицевость функции f в некоторой окрестности U точки x_0 (теорема 5.1). Для любых точек $x_1, x_2 \in U$, равенство $f(x_1) = f(x_2)$ влечёт

$$\|x_1 - x_2\| \leqslant \mathrm{const} \cdot \|f(x_1) - f(x_2)\| = 0,$$

[3]Это не очень сильное требование из-за открытости множества обратимых операторов и непрерывности отображения $x \mapsto d_x f$.

поэтому сужение f на U инъективно. Это верно для любого x_0, поэтому функция f инъективна на области определения. Обозначим обратную функцию через

$$g := f^{-1}.$$

По теореме об открытости отображения с невырожденным дифференциалом (теорема 5.2), f — гомеоморфизм между U и $f(U)$.

Проверим, что функция g дифференцируема в точке $y_0 := f(x_0)$. Для этого покажем, что дифференциалом $\mathrm{d}_{y_0}g$ будет матрица $(\mathrm{d}_{x_0}f)^{-1}$.

Рассмотрим разность

$$g(y_0+k) - g(y_0),$$

где вектор $k \in \mathbb{R}^m$ таков, что $y_0 + k \in f(U)$. Здесь для некоторого вектора $h \in \mathbb{R}^m$ верно

$$g(y_0+k) - g(y_0) = x_0 + h - x_0 = h.$$

Кроме того, имеет место равенство

$$f(x_0+h) - f(x_0) = (\mathrm{d}_{x_0}f)h + r(h)$$

для некоторой функции $r(h) = o(\|h\|)$. Тогда

$$\begin{aligned} g(y_0+k) - g(y_0) &= (\mathrm{d}_{x_0}f)^{-1}(f(x_0+h) - f(x_0) - r(h)) \\ &= (\mathrm{d}_{x_0}f)^{-1}(k - r(h)). \end{aligned}$$

Покажем, что

$$r(h) = o(\|k\|).$$

Это так из-за липшицевости f:

$$\|h\| \geqslant \|k\| \cdot \mathrm{const}.$$

Докажем, что отображение

$$y \mapsto \mathrm{d}_y g$$

непрерывно. Оно, в сущности, равно

$$y \mapsto (\mathrm{d}_{f^{-1}(y)}f)^{-1}.$$

Это композиция трёх непрерывных отображений по определению функции f и поскольку обращение оператора непрерывно (страница 8). ■

5.4 Теорема о неявной функции

Пример 5.1. Рассмотрим уравнение окружности

$$x^2 + y^2 - 1 = 0.$$

Само по себе оно не задаёт функцию, но во всех точках решения уравнения можно локально, непрерывно-дифференцируемо параметризовать одной из переменных x или y:

$$y = \pm\sqrt{1 - x^2}.$$

Не получится выразить y через x ни в одной окрестности точек $\{(\pm 1, 0)\}$, но и там можно выразить x через y.

Пример 5.2 (лемниската Бернулли). Рассмотрим уравнение

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

Введём полярные координаты:

$$x = r\cos\varphi, \quad y = r\sin\varphi.$$

После преобразований получим уравнение

$$r^2 = 2\cos 2\varphi.$$

Можно видеть, что это восьмёрка на боку. Здесь никакой функцией не получится параметризовать решения в окрестности центра симметрии.

Теорема о неявной функции даёт достаточные условия для того, чтобы часть переменных можно было функционально (и непрерывно-дифференцируемо) выразить через другие в окрестности некоторой точки графика.

Обозначение. Для пары столбцов x и y будем обозначать через

$$\begin{pmatrix} x \\ y \end{pmatrix}$$

столбец, получающийся приписыванием y к x снизу.

Обозначение. Пусть $m \geqslant n$ и

$$A\colon \mathbb{R}^m \to \mathbb{R}^n$$

— линейное отображение. Обозначим через

$$A_n \quad \text{и} \quad A_{m-n}$$

отображения, образованные, соответственно, первыми n и последними $m - n$ столбцами матрицы A.

Важно, что в таких обозначениях для любых $x \in \mathbb{R}^n$ и $y \in \mathbb{R}^{m-n}$ выполняется

$$A\begin{pmatrix} x \\ y \end{pmatrix} = A_n x + A_{m-n} y. \tag{5.1}$$

Теорема 5.4 (о неявной функции). Пусть $m \geqslant n$ и функция $f\colon \Omega \subset \mathbb{R}^m \to \mathbb{R}^n$ (где Ω — область) имеет класс C^1; для некоторой точки $\binom{a}{b}$ выполнено $f\binom{a}{b} = 0$.

Обозначим $A := f'\binom{a}{b}$. Пусть, в наших обозначениях, оператор A_n инъективен.

Тогда существует окрестность $U = U_a \times U_b$ (где $U_a \subset \mathbb{R}^n$ и $U_b \subset \mathbb{R}^{m-n}$) точки $\binom{a}{b}$ и функция $g\colon U_b \to U_a$ такие, что

$$U \cap \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \;\middle|\; f\begin{pmatrix} x \\ y \end{pmatrix} = 0 \right\} = \left\{ \begin{pmatrix} g(y) \\ y \end{pmatrix} \;\middle|\; y \in U_b \right\}.$$

Кроме того, g принадлежит классу C^1 и её дифференциал в точке $\binom{a}{b}$ считается по формуле

$$g'(b) = -A_n^{-1} A_{m-n}.$$

Пример 5.3. В случае с окружностью

$$f\begin{pmatrix} x \\ y \end{pmatrix} := x^2 + y^2 - 1 = 0,$$

получаем

$$f'\begin{pmatrix} x \\ y \end{pmatrix} = (2x, 2y).$$

Можно видеть, что эта матрица на окружности всегда имеет ранг 1. Это соответствует нашим наблюдением о том, что график всегда получается параметризовать.

Пример 5.4. Если

$$f\begin{pmatrix} x \\ y \end{pmatrix} := \left(x^2 + y^2\right)^2 - 2\left(x^2 - y^2\right),$$

то

$$f'\begin{pmatrix} x \\ y \end{pmatrix} = \left(2\left(x^2 + y^2\right) \cdot 2x - 4x,\ 2\left(x^2 + y^2\right) \cdot 2y + 4y\right).$$

Поймём, когда левая половина этой матрицы равна нулю, и переменную x не удастся выразить через y. Если $x = 0$, то $y^3 + y = 0$, откуда $y = 0$. Интуитивно очевидно, что в нуле действительно не получится функционально параметризовать график. Если же $x \neq 0$, то имеем $x^2 + y^2 = 1$. Вкупе с уравнением собственно лемнискаты, получаем систему:

$$\begin{cases} x_0^2 + y_0^2 = 1 \\ \left(x_0^2 + y_0^2\right)^2 - 2\left(x_0^2 - y_0^2\right) = 0 \end{cases} \iff \begin{cases} x_0^2 + y_0^2 = 1 \\ x_0^2 - y_0^2 = \frac{1}{2} \end{cases} \iff \begin{cases} x_0 = \pm\sqrt{\frac{1+\frac{1}{2}}{2}} = \pm\frac{\sqrt{3}}{2} \\ y_0 = \pm 1. \end{cases}$$

Это те точки, в которых касательная графику параллельна оси абсцисс.

Пример 5.5 (поведение простых корней многочлена при малых возмущениях коэффициентов). Пусть дан многочлен $p(x) = a_0 + a_1 x + \cdots + a_n x^n$. Предположим, что λ — его вещественный корень кратности 1. Тогда существуют $\varepsilon, \delta > 0$ такие, что если

$$q := b_0 + b_1 x + \ldots + b_n x^n$$

и

$$\max_{0\leqslant k\leqslant n} |a_k - b_k| < \varepsilon,$$

то существует ровно один корень μ многочлена q такой, что

$$|\mu - \lambda| < \delta.$$

Для доказательства достаточно рассмотреть функцию $F(t, y_0, \ldots, y_m) : \mathbb{R} \times \mathbb{R}^{n+1} \to \mathbb{R}$, заданную соотношением

$$F(t, y_0, \ldots, y_n) := y_0 + y_1 t + \ldots + y_n t^n$$

и применить к ней теорему о неявном отображении. Тот факт, что матрица F_t' размера 1×1 будет невырожденна в некоторой окрестности точки $(\lambda, a_0, \ldots, a_n)$ следует из того, что $p'(\lambda) \neq 0$.

Доказательство теоремы о неявной функции. Сведем доказательство к теореме об обратной функции. Определим

$$\widetilde{f} := \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} f\binom{x}{y} \\ y \end{pmatrix}.$$

Покажем, что дифференциал $\widetilde{f}$ в точке $\binom{a}{b}$ обратим. Разложим f по формуле Тейлора в точке $\binom{a}{b}$: поскольку $f\binom{a}{b} = 0$,

$$f\begin{pmatrix} a + h_a \\ b + h_b \end{pmatrix} = A\begin{pmatrix} h_a \\ h_b \end{pmatrix} + r(h),$$

где

$$h = \begin{pmatrix} h_a \\ h_b \end{pmatrix}.$$

Посчитаем аналогичную разность для $\widetilde{f}$:

$$\begin{aligned} \widetilde{f}\begin{pmatrix} a + h_a \\ b + h_b \end{pmatrix} - \widetilde{f}\begin{pmatrix} a \\ b \end{pmatrix} &= \begin{pmatrix} f\binom{a+h_a}{b+h_b} \\ h \end{pmatrix} \\ &= \begin{pmatrix} A\binom{h_a}{h_b} \\ h \end{pmatrix} + \begin{pmatrix} r(h) \\ 0 \end{pmatrix}. \end{aligned}$$

Значит, линейное отображение

$$t \mapsto \begin{pmatrix} At \\ t \end{pmatrix}$$

соответствует определению дифференциала $\widetilde{f}'\binom{a}{b}$. Из-за единственности дифференциала им оно и является. Оно инъективно, поскольку

$$\begin{pmatrix} At_1 \\ t_1 \end{pmatrix} = \begin{pmatrix} At_2 \\ t_2 \end{pmatrix} \implies t_1 = t_2.$$

Теперь мы можем применить теорему об обратной функции к $\widetilde{f}$, чтобы получить окрестность $U = U_a \times U_b$ точки $\binom{a}{b}$ такую, что функция $\widetilde{f}$ в ней инъективна. То есть, для любых $\binom{x_1}{y_1}$ и $\binom{x_2}{y_2}$ из окрестности U из равенства

$$\begin{pmatrix} f\binom{x_1}{y_1} \\ y_1 \end{pmatrix} = \begin{pmatrix} f\binom{x_2}{y_2} \\ y_2 \end{pmatrix}$$

следует

$$x_1 = x_2.$$

Покажем, что для любого y существует единственный $x \in U_a$ такой, что

$$f\begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

Предположим, $y \in U_b$, и для $x_1, x_2 \in U_a$ выполняется

$$f\begin{pmatrix} x_1 \\ y \end{pmatrix} = f\begin{pmatrix} x_2 \\ y \end{pmatrix} = 0.$$

Тогда, в частности,

$$\begin{pmatrix} f\binom{x_1}{y} \\ y \end{pmatrix} = \begin{pmatrix} f\binom{x_2}{y} \\ y \end{pmatrix},$$

а из этого по определению следует, что

$$x_1 = x_2.$$

Поэтому определим искомое отображение $g\colon U_b \to U_a$ формулой

$$g := y \mapsto \text{единственный } x \in U_a \text{ такой, что } f\begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

Проверим, что $g \in C^1$. По теореме об обратной функции, обратная функция к сужению $\widetilde{f}$ на U принадлежит классу C^1. Функцию g можно дифференцируемо выразить через неё:

$$\begin{pmatrix} g(y) \\ y \end{pmatrix} = (\widetilde{f}|_U)^{-1}\begin{pmatrix} 0 \\ y \end{pmatrix}.$$

Поэтому $g \in C^1$.

Нам осталось явно найти её дифференциал в $\binom{a}{b}$. Определим

$$\widetilde{g} := y \mapsto \begin{pmatrix} g(y) \\ y \end{pmatrix}.$$

Тогда, как мы поняли это для функции $\widetilde{f}$, дифференциал $\widetilde{g}$ будет прямой суммой дифференциалов $y \mapsto g(y)$ и $y \mapsto y$:

$$\widetilde{g}'(y) = \begin{pmatrix} g'(y) \\ I \end{pmatrix}.$$

В то же время, для любого $y \in U_b$ по определению верно

$$f(\widetilde{g}(y)) = 0,$$

что, в частности, подразумевает

$$f'(\widetilde{g}(y))\widetilde{g}'(y) = 0.$$

В случае, когда $y = b$, первый из этих множителей по определению равен A:

$$A\widetilde{g}'(y) = 0.$$

В то же время A распадается в сумму A_n и A_{n-m} (по определению со страницы 32):

$$0 = A\widetilde{g}'(y) = A_n g'(y) + A_{n-m},$$

откуда, из обратимости A_n,

$$g'(y) = -A_n^{-1} A_{m-n},$$

как и хотели. ■

5.5 Гладко параметризованные многообразия в $\mathbb{R}^n$

Гладкие многообразия

Напомним,

Определение. *Многообразие размерности* $n \in \mathbb{N}$ — хаусдорфово топологическое пространство со счётной базой, в котором у каждой точки существует окрестность, гомеоморфная $\mathbb{R}^n$.

Пример 5.6. Сфера $S_2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$ — многообразие размерности 2.

Заведём на многообразиях аналитическую структуру.

Определение (карты и атласы). Пусть M — многообразие размерности n; $m \in M$ — его точка. У m есть окрестность U_m такая, что для некоторого открытого подмножества $U'_m \subset \mathbb{R}^n$ существует гомеоморфизм

$$\varphi \colon U'_m \to U_m.$$

U'_m называется *картой* окрестности U_m, а гомеоморфизм φ — *локальной параметризацией*, соответствующей карте U'_m. Набор карт A такой, что их образы при соответствующих параметризациях покрывают всё многообразие M, называется *атласом*.

Определение. Если в атласе всего одна карта, то многообразие называется *простым*, а параметризация, соответствующая этой карте, — *глобальной*.

Определение. n-мерное многообразие M с атласом $A = \{U_\alpha\}_{\alpha \in I}$ называется *гладким* класса C^k, если для любых двух карт U_α и U_β пересекающихся окрестностей $\varphi_\alpha(U_\alpha)$ и $\varphi_\beta(U_\beta)$ *отображение перехода*

$$\varphi_\beta^{-1} \circ \varphi_\alpha \colon \{x \in U_\alpha \mid \varphi_\alpha(x) \in \varphi_\beta(U_\beta)\} \to \mathbb{R}^n$$

лежит в классе C^k.

Примеры

Пример 5.7. Сфера $S_1 = \{(x, y) \mid x^2 + y^2 = 1\} \subset \mathbb{R}^2$ — гладкое многообразие. Рассмотрим атлас из двух карт

$$U_1 = (0, 2\pi) \quad \text{и} \quad U_2 = (-\pi, \pi),$$

которым соответствуют параметризации

$$\varphi_1 = t \to (\cos t, \sin t) \quad \text{и} \quad \varphi_2 = r \mapsto (\cos t, \sin t).$$

Отображения

$$\varphi_1^{-1} \circ \varphi_2 \quad \text{и} \quad \varphi_2^{-1} \circ \varphi_1$$

тождественны всюду, где определены, а потому окружность — C^∞-гладкое многообразие.

Пример 5.8. Рассмотрим множество

$$Q := \{(x, y) : |x| = 1 \text{ или } |y| = 1 \text{ и } |x| \leqslant 1, |y| \leqslant 1\}$$

(граница квадрата). Впишем окружность в квадрат и запараметризуем границу квадрата точками окружности:

$$\psi(\cos t, \sin t) := \frac{(\cos t, \sin t)}{\max(|\cos t|, |\sin t|)}.$$

Тогда отображение ψ — гомеоморфизм, поскольку оно непрерывно и у него есть непрерывное обратное отображение (из квадрата в окружность):

$$\psi^{-1} = (x, y) \mapsto \left(\frac{x}{x^2 + y^2}, \frac{y}{x^2 + y^2} \right).$$

Пусть U_1, U_2 — как в предыдущем примере, тогда для $i \in \{1, 2\}$ выполнено

$$\varphi_i = \psi(\varphi_i).$$

Как мы уже заметили, отображения перехода $\varphi_1^{-1} \circ \varphi_2$, $\varphi_2^{-1} \circ \varphi_1$ тождественны на области свего задания, а потому Q — топологически гладкое многообразие класса C^∞.

Пример 5.9. Лемниската Бернулли — не многообразие (неформально — рассмотрим малую окрестность нуля в $\mathbb{R}^2$. Её пересечение с лемнискатой должно быть простым размерности 1; в частности, должно параметризоваться с помощью гомеоморфизма, заданного на открытом интервале. Но при выкидывании центральной точки лемнискаты ее пересечение с малой окрестностью нуля распадается на четыре части, а интервал — только на две).

Собственно ГПМ

Образ гладкой функции $\mathbb{R}^m \to \mathbb{R}^n$, в сущности, задаёт некоторую *гладкую поверхность* в $\mathbb{R}^n$. Мы вводим два уточнения этого понятия и будем пользоваться ими вперемежку.

Определение (первое определение ГПМ). Множество $M \subset \mathbb{R}^n$ называется *гладко параметризованным многообразием* размерности k, если для всякой точки $p \in M$ существует непрерывно дифференцируемый гомеоморфизм

$$\varphi\colon U \to M \cap V \subset \mathbb{R}^n$$

(где $U \subset \mathbb{R}^n$; V — окрестность точки p) такой, что для любой точки $x \in U$

$$\operatorname{rank} \varphi'(x) = k.$$

Определение (второе определение ГПМ). Множество $M \subset \mathbb{R}^n$ называется *гладко параметризованным многообразием* размерности k, если для всякой точки $p \in M$ существует диффеоморфизм

$$\Phi\colon S \to W \subset \mathbb{R}^n$$

(где $S \subset \mathbb{R}^n$; W — окрестность точки p) такой, что, для

$$\widetilde{S} := \{x \in S \mid \text{нижние } n-k \text{ элементов } x \text{ — нулевые}\},$$

верно

$$\Phi(\widetilde{S}) = M \cap W.$$

В первом определении вы рисуете вокруг точки шар, пересекаете его с многообразием и говорите, что получился диск (с точностью до непрерывно дифференцируемого гомеоморфизма). Во втором вы рисуете вокруг точки шар и сначала говорите, что это шар (с точностью до диффеоморфизма), а затем отмечаете, что многообразие пересекает этот шар по диску.

Эквивалентность определений

Теорема 5.5. Определения 1 и 2 гладко параметризованного многообразия эквивалентны.

Нужно проверить, что C^1-гомеоморфизм из первого определения можно продолжить до диффеоморфизма из второго определения и, наоборот, что диффеоморфизм можно сузить до C^1-гомеоморфизма.

Доказательство. (1) $\implies$ (2). Пусть $p \in M$ — любая точка, $\varphi\colon U \to V \subset M$ — соответствующий C^1-гомеоморфизм, ранг дифференциала которого равен k на области определения. Определим отображение $\Phi\colon U \times \mathbb{R}^{n-k} \to \mathbb{R}^n$ формулой

$$\Phi\begin{pmatrix} x_1 \\ \vdots \\ x_k \\ x_{k+1} \\ \vdots \\ x_n \end{pmatrix} = \varphi\begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x_{k+1} \\ \vdots \\ x_n \end{pmatrix}.$$

Матрица Якоби Φ в некоторой точке a имеет следующий вид:

$$\Phi'(a) = [\varphi'(a) \mid 0] + \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}.$$

Здесь $\varphi'(a)$ — высокая матрица $n \times k$, а I — единичная матрица $(n-k)\times(n-k)$. Ранг матрицы $\varphi'(a)$ по определению равен k. Без ограничения общности можно считать, что k независимых строк идут подряд сверху. Тогда матрица $\Phi'(a)$ обратима.

Применяя теорему об обратном отображении к Φ-прообразу точки p, получаем область

$$S' \subset U \times \mathbb{R}^{n-k},$$

такую, что сужение $\Phi|_{S'}$ — диффеоморфизм из S' на свой образ

$$W' := \Phi(S').$$

По определению, эта W' — область в $\mathbb{R}^n$, содержащая точку p. Пусть

$$\widetilde{S'} := \{x \in S' \mid \text{последние } n-k \text{ координат } x \text{ равны нулю}\}.$$

Тогда, из-за определения Φ,

$$\Phi(\widetilde{S'}) = \varphi(U \cap \widetilde{S'}).$$

Это открытое множество, поскольку пересечение $U \cap \widetilde{S'}$ открытых множеств открыто, а φ — открытое отображение. При этом множество $\Phi(\widetilde{S'})$ полностью лежит в M, поскольку $U \cap \widetilde{S'} \subset U$. Обозначим его за

$$W := \Phi(\widetilde{S'}).$$

Пусть

$$S := \Phi^{-1}(W).$$

Отметим, что, поскольку $\widetilde{S'} \subset S'$ по определению,

$$\Phi(S) = W = \Phi(\widetilde{S'}) \subset \Phi(S').$$

Покажем, что

$$S \subset S'.$$

Если $s \in S \setminus S'$, то $\Phi(s) \in \Phi(S')$. Значит, у $\Phi(s)$ есть прообраз $s' \in S'$. Очевидно, $s \neq s'$, поэтому у $\Phi(s)$ два прообраза. Этого не может быть, поскольку $\Phi(S) = W$ — Φ-образ $\widetilde{S}'$, $\widetilde{S}' \subset S'$, а сужение $\Phi|_{S'}$ — биекция по определению.

Поэтому, если мы по аналогии определяем

$$\widetilde{S} := \{x \in S \mid \text{последние } n-k \text{ координат } x \text{ равны нулю}\},$$

то

$$\widetilde{S} \subset \widetilde{S}'.$$

Далее,

$$\begin{aligned} \Phi(\widetilde{S}') = W \cap M \subset W &\rightsquigarrow \widetilde{S}' \subset S \\ &\rightsquigarrow \widetilde{S}' \subset \widetilde{S} \\ &\rightsquigarrow \widetilde{S}' = \widetilde{S} \\ &\rightsquigarrow \Phi(\widetilde{S}') = \Phi(\widetilde{S}) \\ &\rightsquigarrow W \cap M = \Phi(\widetilde{S}). \end{aligned}$$

Что и требовалось доказать.

(2) $\implies$ (1). Пусть дан диффеоморфизм $\Phi\colon S \to W$ как во втором определении. Определим

$$\varphi\begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} := \Phi\begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Это гомеоморфизм как сужение гомеоморфизма. Положим, в первом определении ГПМ, $V := M \cap W$ и $U := S$. Тогда матрица $\varphi'(u)$ везде совпадает с подматрицей $n \times k$ матрицы дифференциала отображения Φ, а последняя имеет полный ранг. Значит, $\operatorname{rank} \varphi'(u) = k$. ■

Локальная билипшицевость карт

Следствие 5.6. Пусть M — гладко параметризованное многообразие размерности k, $p \in M$ — точка оттуда, гомеоморфизм $\varphi\colon U \to V$ как в определении. Тогда φ локально билипшицево: существуют окрестность U_* точки $\varphi^{-1}(p)$ и $\widetilde{c}, c \in \mathbb{R}$ такие, что для любых $u_1, u_2 \in U_* \subset U$ верно

$$\widetilde{c}\|u_1 - u_2\| \leqslant \|\varphi(u_1) - \varphi(u_2)\| \leqslant c\|u_1 - u_2\|.$$

Доказательство. Достаточно применить теорему о локальной билипшицевости к отображению Φ из доказательства переформулировки. Заметим, что выполнено

$$\varphi(u_1) - \varphi(u_2) = \Phi(u_1') - \Phi(u_2'),$$

где

$$u_1' = \begin{pmatrix} u_{1,1} \\ \vdots \\ u_{1,k} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad u_2' = \begin{pmatrix} u_{2,1} \\ \vdots \\ u_{2,k} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Кроме того, здесь

$$\left\|u_1' - u_2'\right\|_{\mathbb{R}^n} = \left\|u_1 - u_2\right\|_{\mathbb{R}^k}.$$

■

Гладкие многообразия и ГПМ

Следствие 5.7. Гладко параметризованное многообразие M размерности k — гладкое многообразие размерности k.

Доказательство. Зададим атлас многообразия M как набор множеств $U_p \subset \mathbb{R}^k$ таких, что существует $\varphi\colon U_p \to V$ — гомеоморфизм, являющийся сужением диффеоморфизма $\Phi_p : U_p \times \mathbb{R}^{n-k} \to W$ из второго определения г.п.м. Если $\varphi_{p_1}(U_{p_1}) \cap \varphi_{p_2}(U_{p_2}) \neq \varnothing$, то имеет место равенство $\varphi_{p_1}^{-1} \circ \varphi_{p_2} = \Phi_{p_1}^{-1} \circ \varphi_{p_2}$. По теореме об обратном отображении, $\Phi_{p_1}^{-1} \in C^1(W) \implies \varphi_{p_1}^{-1} \circ \varphi_{p_2} \in C^1$ там, где эта суперпозиция определена. ■

5.6 Касательное пространство

Определение. Пусть M — гладко параметризованное многообразие размерности k, $p \in M$. Вектор $\tau \in \mathbb{R}^n$ называется *касательным* к M в точке p, если существуют путь $\gamma\colon [a, b] \to M$ и точка $c \in [a, b]$ такие, что

$$\gamma(c) = p$$

и

$$\gamma'(c) = \tau.$$

Определение. *Касательное пространство* $T_p(M)$ — это множество всех касательных векторов к M в p.

Образ дифференциала карты

Теорема 5.8 (касательное пространство — образ дифференциала карты). Пусть M — ГПМ, p — любая его точка, $\varphi\colon U \subset \mathbb{R}^k \to V \ni p$ — соответствующая карта, a —

φ-прообраз точки p. Тогда
$$T_p(M) = \operatorname{Im}\varphi'(a).$$
В частности, размерности многообразия и касательного пространства совпадают.

Доказательство. Проверим, что $\operatorname{Im} \mathrm{d}_a\varphi \subset T_pM$. Пусть $e \in \mathbb{R}^k$, где k — размерность M. Рассмотрим путь
$$\gamma = t \mapsto \varphi(a + te),$$
где $t \in [-\varepsilon, \varepsilon]$ и число ε выбрано так, чтобы $a + te \in U$ $\forall t \in [-\varepsilon, \varepsilon]$.
$$\gamma'(0) = \lim_{t\to 0} \frac{\gamma(t) - \gamma(0)}{t} = \lim_{t\to 0} \frac{\varphi(a+te) - \varphi(a)}{t} = [\mathrm{d}_a\varphi](e),$$
причем $\gamma(0) = \varphi(a) = p$.

Наоборот, пусть $e = \gamma'(c)$, – касательный вектор в точке p, где $\gamma : [\widetilde{a}, \widetilde{b}] \to M$ — из определения, $\gamma(c) = p$. Рассмотрим $\varphi\colon U \to V \ni p$. Пусть $u_t = \varphi^{-1}(\gamma(t))$, где $t \in [\widetilde{a}, \widetilde{b}]$. Тогда $u_c = a$, где $a : \varphi(a) = p$.
$$\gamma'(c) = \lim_{\varepsilon\to 0} \frac{\gamma(c+\varepsilon) - \gamma(c)}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{\varphi(u_{c+\varepsilon}) - \varphi(a)}{\varepsilon} = [\mathrm{d}_a\varphi]h,$$
где $u_{c+\varepsilon} \to u_c = a$ при $\varepsilon \to 0$, h — любой частичный предел отображения $\dfrac{u_{c+\varepsilon} - a}{\varepsilon}$. Какой-то частичный предел есть, поскольку $\|u_{c+\varepsilon} - a\| \leqslant c \cdot \varepsilon$, так как отображение φ локально билипшицево, следовательно, множество
$$\left\{\frac{u_{c+1/n} - a}{1/n},\ n \in \mathbb{N}\right\}$$
ограничено в $\mathbb{R}^n$ и имеет предельную точку. На самом деле из дифференцируемости γ в точке c следует, что предельная точка h — единственна, но здесь этот факт не используется, так как мы уже показали, что $\gamma(c) \in \operatorname{Im} \mathrm{d}_a\varphi$, и, значит, $T_pM \subset \operatorname{Im} \mathrm{d}_a\varphi$. ■

Определение. *Аффинное* касательное пространство к гладко параметризуемому многообразию M в точке $p \in M$ — это аффинное подпространство $p + T_p(M)$.

Поведение многообразия вблизи точки касания

Теорема 5.9. Пусть M — гладко параметризованное многообразие размерности k в $\mathbb{R}^n$, $p \in M$. Тогда среди всех аффинных подпространств $\mathbb{R}^n$ вида $p + L$ аффинное касательное подпространство однозначно характеризуется следующими свойствами:

1. Если $x \in p + L$ и $x \to p$, то $\operatorname{dist}(x, M) = o(\|x - p\|)$.

2. $\dim L \geqslant k$.

Лемма 5.10. Пусть $A\colon n \times k$ — матрица, $\operatorname{rank} A = k$, $k \leqslant n$. Тогда существует константа $c > 0$ такая, что для любой точки $x \in \mathbb{R}^k$ выполняется неравенство $\|Ax\| \geqslant c\|x\|$.

Доказательство. $\operatorname{rank} A = k \iff \dim \operatorname{Im} A = k \implies \operatorname{Ker} A = \{0\}$ (если $\exists e_1 \in \mathbb{R}^k : Ae_1 = 0$, то $\exists e_2, \dots, e_k : e_1, \dots, e_k$ — базис $\mathbb{R}^k$ и

$$k = \dim \operatorname{Im} A = \dim(A \operatorname{span}\{e_1, \dots, e_k\}) = \dim(A \operatorname{span}\{e_2, \dots, e_k\}) \leqslant k - 1$$

— противоречие). Так как $\operatorname{Ker} A = \{0\}$, то отображение $p\colon x \mapsto \|Ax\|$ — норма на $\mathbb{R}^k$. ($p(x) = 0 \iff x = 0$). Значит, $\exists c : p(x) \geqslant c\|x\|$ (т.к. все нормы эквивалентны). ■

доказательство теоремы. Сначала покажем, что если $L = T_p(M)$, то условия 1 и 2 выполнены. Рассмотрим стремящийся к нулю вектор x из $p + T_p(M)$ и оценим его расстояние до M. По определению M, найдутся окрестность V точки p в M, окрестность $U \subset \mathbb{R}^k$ и C^1-гомеоморфизм $\varphi\colon U \to V$ с рангом дифференциала k на области определения. Пусть $a := \varphi^{-1}(p)$. Не умаляя общности, будем считать, что разность $x - p$ достаточно мала, так что точка $a + [\mathrm{d}_a\varphi]^{-1}(x - p)$ лежит в U и величина $m_x := \varphi(a + [\mathrm{d}_a\varphi]^{-1}(x - p))$ определена. Тогда, дифференцируя φ в точке a, получаем оценку

$$\begin{aligned} \operatorname{dist}(x, M) &\leqslant \|x - m_x\| \\ &= \left\|x - \left(\varphi(a) + [\mathrm{d}_a\varphi][\mathrm{d}_a\varphi]^{-1}(x - p) + r(a, [\mathrm{d}_a\varphi]^{-1}(x - p)\right)\right\|, \\ &= o(\|[\mathrm{d}_a\varphi]^{-1}(x - p)\|). \end{aligned}$$

По лемме,

$$\|x - p\| = \|[\mathrm{d}_a f][\mathrm{d}_a f]^{-1}(x - p)\| \geqslant c\|[\mathrm{d}_a f]^{-1}(x - p)\|,$$

значит,

$$\operatorname{dist}(x, M) = o\left(\frac{1}{c}\|x - p\|\right) = o(\|x - p\|),$$

Итак, мы проверили пункт 1; пункт 2 следует из того, что $\dim T_p(M) = k$ по теореме на странице 41.

Пусть теперь L — какое-то подпространство $\mathbb{R}^n$, $p + L$ удовлетворяет свойствам 1 и 2. Поскольку $\dim L \geqslant k$, достаточно показать, что $L \subseteq T_p(M)$ — тогда $\dim L = \dim T_p(M)$, и эти подпространства совпадают. Пусть $e \in L$; покажем, что $e \in T_p(M)$, или, что то же самое, существует вектор $v \in \mathbb{R}^k$ такой, что $e = [\mathrm{d}_a\varphi]v$.

Рассмотрим последовательность

$$x_t = p + te, \quad t \in [0, 1]$$

векторов из пространства L. По определению L, для каждого t найдётся вектор $m_t \in M$ такой, что

$$\|x_t - m_t\| = o(\|x_t - p\|).$$

По определению x_t, здесь $x_t - p = te$, так что

$$\|x_t - m_t\| = o(t).$$

Отсюда сразу имеем, что, так как $x_t \to p$, то $m_t \to p$. После регулировки вектора $e \in L$ можно считать, что для любого $t \in [0,1]$ вектор m_t лежит в V. Тогда существуют вектора $u_t \in \mathbb{R}^k$, параметризующие прообразы m_t:

$$\varphi(a + u_t) = m_t.$$

Ясно, что $u_t \to 0$. Продифференцируем гомеоморфизм φ по последовательности векторов u_t:

$$\varphi(a + u_t) = \varphi(a) + [\mathrm{d}_a\varphi](u_t) + o(\|u_t\|), \ t \to 0.$$

Оценим e, используя отрезки $[m_t, x_t]$ и $[m_t, p]$, для каждого из которых некоторые оценки у нас уже есть:

$$\begin{aligned} e &= \frac{x_t - p}{t} \\ &= \frac{x_t - m_t}{t} + \frac{m_t - p}{t} \\ &= o(1) + \frac{\varphi(a+u_t) - \varphi(a)}{t} \\ &= [\mathrm{d}_a\varphi]\left(\frac{u_t}{t}\right) + o\left(1 + \left\|\frac{u_t}{t}\right\|\right). \end{aligned}$$

Напомним, мы хотим найти такой вектор v, что $e = [\mathrm{d}_a\varphi](v)$. По равенству непосредственно выше, достаточно показать, что последовательность u_t/t имеет предельную точку $v \in \mathbb{R}^k$. Для этого можно доказать, что $\|u_t/t\| \leqslant c$ для любых t в некотором промежутке $[0, t_0]$; если это готово, то за v можно взять предел любой (существующей) сходящейся подпоследовательности.

Хотим оценить u_t сверху. Как можно вспомнить, это разность между точкой $a \in U$ и φ-прообразом вектора m_t. Ну а гомеоморфизм φ — билипшицев из U в V (по утверждению на странице 40): существует $\widetilde{c} > 0$ такая, что

$$\|\varphi(a + u_t) - \varphi(a)\| \geqslant \widetilde{c}\|a + u_t - a\|.$$

Здесь, по определению,

$$\varphi(a + u_t) = m_t \quad \text{и} \quad \varphi(a) = p.$$

Значит, при достаточно малых $t > 0$, если

$$\|m_t - x_t\| \leqslant \|x_t - p\|,$$

то имеем:

$$\begin{aligned} \widetilde{c}\|u_t\| &= \|m_t - p\| \\ &\leqslant \|m_t - x_t\| + \|x_t - p\| \\ &\leqslant 2\|x_t - p\| \\ &= 2t \cdot \|e\|. \end{aligned}$$

Отсюда следует, что

$$\left\|\frac{u_t}{t}\right\| \leqslant \frac{2\|e\|}{\widetilde{c}}.$$

■

5.7 ГПМ как множество уровня

Теорема 5.11 (любое ГПМ — множество уровня). Пусть $M \subset \mathbb{R}^n,\ k \in \mathbb{N}$. Следующие условия равносильны:

1. M — гладко параметризованное многообразие размерности k.

2. для любой точки $p \in M$ существуют окрестность W точки p и функция

$$F\colon \mathbb{R}^n \to \mathbb{R}^{n-k},$$

непрерывно дифференцируемая на W, так что

$$\operatorname{rank} F'(p) = n - k$$

и

$$M \cap W = \{x \in \mathbb{R}^n \mid F(x) = 0\}.$$

3. для любой точки $p \in M$ существуют окрестность W точки p и функции

$$F_1, \dots, F_{n-k}\colon \mathbb{R}^n \to \mathbb{R}^{n-k},$$

непрерывно дифференцируемые на W, так что вектора

$$\operatorname{grad}_p F_1, \dots, \operatorname{grad}_p F_{n-k};$$

линейно независимы в $\mathbb{R}^n$ и

$$M \cap W = \{x \in \mathbb{R}^n \mid F_1(x) = \dots = F_{n-k}(x) = 0\}.$$

Соглашение. Будем называть функции F и $\{F_i\}$ в контексте теоремы выше *функциями условия*.

Доказательство. (2) $\implies$ (3). Можно взять координатные функции $\{F_i\}$. Условие на ранг означает, что в матрице $F'(p)$ найдётся $n-k$ линейно независимых строк. Но там всего $n-k$ строк, и это ровно $F_1'(p), \dots, F_{n-k}'(p)$.

(3) $\implies$ (2). Аналогично, возьмем

$$F(x) := \begin{pmatrix} F_1(x) \\ \vdots \\ F_{n-k}(x) \end{pmatrix}.$$

(1) $\implies$ (2). Пусть M — гладко параметризованное многообразие размерности k, p — любая точка оттуда. Воспользуемся второй переформулировкой определения ГПМ: найдутся открытое множество $S \subset \mathbb{R}^n$ и окрестность W точки p такие, что между ними существует диффеоморфизм

$$\Phi\colon S \to W,$$

для которого, кроме того, выполняется условие

$$M \cap W = \Phi(\widetilde{S}),$$

где

$$\widetilde{S} = \{u \in S \mid u = (u_1, \dots, u_k, 0, \dots, 0)\}.$$

Пусть $P\colon \mathbb{R}^n \to \mathbb{R}^{n-k}$ — проекция, сопоставляющая столбцу из $\mathbb{R}^n$ его последние $n-k$ элементов. Рассмотрим отображение $F\colon \mathbb{R}^n \to \mathbb{R}^{n-k}$ заданное по правилу

$$F := x \mapsto P\Phi^{-1}(x).$$

По определению, $F \in C^1(W)$ и

$$\operatorname{rank} F(p) = \operatorname{rank} \mathrm{d}_p F = \operatorname{rank} P \mathrm{d}_p \Phi^{-1}.$$

Так как $\mathrm{d}_p\Phi^{-1} \in \operatorname{Hom}(\mathbb{R}^n, \mathbb{R}^n)$ — обратимый оператор, то

$$\operatorname{rank} P \mathrm{d}_p \Phi^{-1} = \operatorname{rank} P = n - k.$$

Кроме того, из-за биективности Φ,

$$\begin{aligned} x \in M \cap W &\iff \Phi^{-1}(x) \in \widetilde{S} \\ &\iff P\Phi^{-1}(x) = 0 \\ &\iff F(x) = 0. \end{aligned}$$

Так что F соответствует требованиям.

(2) $\implies$ (1). Воспользуемся теоремой о неявной функции. Перенумеруем координаты так, чтобы первые $n-k$ столбцов матрицы $f'(p)$ были линейно независимы. Тогда найдутся окрестность $U = U_a \times U_b$ (где $U_a \subset \mathbb{R}^{n-k}$ и $U_b \subset \mathbb{R}^k$) точки p и непрерывно-дифференцируемая функция $g\colon U_b \to U_a$ такие, что для $\binom{x}{y} \in U$ равенство $F\binom{x}{y} = 0$ выполняется тогда и только тогда, когда $x = g(y)$. Положим

$$\varphi := y \mapsto \begin{pmatrix} g(y) \\ y \end{pmatrix}.$$

Это карта пересечения $M \cap U$. Легко видеть, что ранг её дифференциала равен k. ∎

Примеры

Пример 5.10. Покажем, что

$$M := \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \; \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \right\}$$

— гладко параметризуемое многообразие размерности 2 в $\mathbb{R}^3$. Для этого рассмотрим соотношение

$$F(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1.$$

Ясно, что $F \in C^\infty(\mathbb{R}^3)$. Посмотрим на матрицу Якоби этого отображения:

$$F'(x, y, z) = \left(\frac{2x}{a^2}, \frac{2y}{b^2}, \frac{2z}{c^2} \right).$$

Если $(x, y, z) \in M$, то одно из чисел $\{x, y, z\}$ отлично от нуля, так что

$$\operatorname{rank} F'(x, y, z) = 1.$$

Значит, мы попадаем в условие теоремы непосредственно выше, когда

$$n - k = 1.$$

Отсюда имеем

$$k = 2.$$

Пример 5.11. Покажем, что

$$M := \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x^2 + y^2 + z^2 = 1, x + y + z = 0 \}.$$

— ГПМ, и найдём его размерность. Определим

$$F(x, y, z) := \begin{pmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{pmatrix}.$$

Тогда

$$F'(x, y, z) := \begin{pmatrix} 2x & 2y & 2z \\ 1 & 1 & 1 \end{pmatrix}.$$

Если строки этой матрицы линейно зависимы, то

$$x = y = z.$$

Подстановкой можно убедиться, что эти равенства не выполняются для точек из M.

Значит, M — гладко параметризованное многообразие размерности

$$3 - 2 = 1.$$

Касательное пространство и функция условия

Функция условия и карта многообразия — двойственные понятия (сравните с теоремой на странице 41):

Теорема 5.12 (касательное пространство — ядро дифференциала условия). Пусть функция F и ГПМ

$$M := \{x \in \mathbb{R}^n \mid F(x) = 0\}$$

— такие же, как в предыдущей теореме (страница 45). Тогда

$$T_p(M) = \operatorname{Ker} F'(p).$$

Доказательство. Покажем, что $T_p(M) \subset \operatorname{Ker} F'(p)$. Пусть $v \in T_p(M)$ и γ — соответсвующий путь по многообразию M, проходящий через p в некоторый момент c и имеющий в c дифференциал v. В частности, для любого t по определению функции F верно

$$F(\gamma(t)) = 0.$$

Дифференцируя это равенство, получаем

$$F'(\gamma(t))\gamma'(t) = 0.$$

Если $t = c$, то

$$F'(p)v = 0.$$

Значит,

$$v \in \operatorname{Ker} F'(p).$$

Покажем, что $\dim T_p(M) \geqslant \dim \operatorname{Ker} F'(p)$. Касательное пространство $T_p(M)$ имеет ту же размерность, что и само многообразие M (страница 41)

$$\dim T_p(M) = \dim M = k.$$

Размерность ядра можно посчитать из размерности образа, по определению функции F равной $n - k$:

$$\dim \operatorname{Ker} F'(p) = n - \dim \operatorname{Im} F'(p) = k.$$

■

Пример 5.12. Найдём общий вид аффинной касательной плоскости к двумерной сфере S^2. Пусть

$$p := \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \in S^2.$$

Касательное пространство к S_2 в точке p по теореме выше имеет вид

$$T_p(S_2) = \operatorname{Ker}\begin{pmatrix} 2x_0 & 2y_0 & 2z_0 \end{pmatrix}.$$

Сдвинув начало отсчёта в точку p, получаем определяющее соотношение для точки $(x, y, z)^{\mathrm{T}}$ в аффинном подпространстве $p + T_p(S^2)$:

$$(x - x_0)x_0 + (y - y_0)y_0 + (z - z_0)z_0 = 0.$$

5.8 ГПМ и графики

Теорема 5.13 (ГПМ локально является графиком отображения). Пусть $M \subseteq \mathbb{R}^n$ — ГПМ размерности k, p — точка оттуда. Тогда существует R — окрестность нуля в касательном пространстве $T_p(M)$, C^1-отображение $\xi\colon R \to \big(T_p(M)\big)^\perp$ и окрестность $W \subset \mathbb{R}^n$ точки p такая, что

$$M \cap W = p + \{\xi(w) + w \mid w \in R\}.$$

Эта полезное замечание говорит, что можно поставить репер в точке p таким образом, что часть многообразия будет в системе координат этого репера графиком функции на касательном пространстве.

Доказательство. После замены базиса и перенумеровки векторов можно считать, что касательное пространство $T_p(M)$ порождено векторами $e_1, \dots, e_k$ (их действительно k по теореме на странице 41) и, соответственно, его ортогонал $\big(T_p(M)\big)^\perp$ — векторами $e_{k+1}, \dots, e_n$. Вспомним, что всё многообразие M задаётся как множество уровня какой-то непрерывно дифференцируемой функции F с рангом дифференциала везде $n - k$ (страница 45). Применим теорему о неявном отображении (страница 32) и получим функцию ξ и все окрестности к ней. Условие, что R является окрестностью нуля, эквивалентно тому, что окрестность $M \cap W$ нашего многообразия содержит точку p. ■

6 Множители Лагранжа

Определение. Пусть $f\colon \mathbb{R}^m \to \mathbb{R}$ — скалярная функция, $F\colon \mathbb{R}^m \to \mathbb{R}^n$ — векторнозначная функция. Обозначим

$$K := \{x \in \mathbb{R}^m \mid F(x) = 0\}.$$

Будем говорить, что f имеет локальный минимум в точке x_0 *при условии* F, если $x_0 \in K$ и сужение $f|_K$ имеет локальный минимум в x_0. В таком случае x_0 называется точкой *условного экстремума*.

Аналогично с максимумом и строгими экстремумами.

В этом разделе мы научимся находить условные экстремумы в случае, когда функция F задаёт гладко параметризованное многообразие, а функция f непрерывно дифференцируема.

6.1 Необходимое условие

Теорема 6.1 (метод множителей Лагранжа, часть 1: необходимое условие). Пусть $n \leqslant m$. Пусть $F\colon \mathbb{R}^m \to \mathbb{R}^n$ — функция условия некоторого ГПМ M размерности n, а $p \in M$ — точка оттуда. Если скалярная функция $f\colon \mathbb{R}^m \to \mathbb{R}$ класса C^1 имеет локальный экстремум в точке p при условии F, то

1. $\operatorname{grad}_p f \perp T_p(M)$, где M — многообразие, заданное функцией F;

2. $\operatorname{grad}_p f \in \operatorname{span}\{\operatorname{grad}_p F_i \mid i\}$;

3. существует вектор $\lambda \in \mathbb{R}^n$ такой, что *функция Лагранжа* из $\mathbb{R}^{m+n}$ в $\mathbb{R}$, определяемая формулой
$$L(x,\lambda) := f(x) - \lambda^{\mathrm{T}} F(x)$$
имеет нулевой градиент в точке (p,λ).

Доказательство. Пункт 1: покажем, что любой касательный вектор $v \in T_p(M)$ ортогонален градиенту функции f в точке p. По определению, существует путь γ по многообразию M, проходящий через p в момент c и имеющий v своим дифференциалом в этой точке. Поскольку функция f имеет экстремум в точке p, функция

$$f \circ \gamma$$

имеет экстремум в точке c. Значит, её дифференциал там равен нулю:

$$0 = f'(\gamma(c))\gamma'(c) = f'(p)v.$$

Что и требовалось доказать.

Выведем утверждение пункта 2 из пункта 1 с учётом равенства

$$T_p(M) = \operatorname{Ker} F'(p)$$

(страница 48). Из пункта 1 имеем

$$\operatorname{grad}_p f \in \big(T_p(M)\big)^{\perp}.$$

В то же время, очевидно,

$$\operatorname{Ker} F'(p) = \Big(\operatorname{span}\{\operatorname{grad}_p F_i \mid i\}\Big)^{\perp}.$$

Значит,

$$\operatorname{grad}_p f \in \operatorname{span}\{\operatorname{grad}_p F_i \mid i\}.$$

Пункт 3: продифференцировав функцию Лагранжа и приравняв к нулю, понимаем, что нам нужен такой вектор $\lambda \in \mathbb{R}^n$, что

$$f'(p) = \lambda^{\mathrm{T}} F'(p).$$

Можно видеть, что его существование и утверждает пункт 2. ■

Пример 6.1. Попробуем максимизировать квадратичную форму

$$f(x) := x^{\mathrm{T}} A x,$$

(где A — симметричная матрица $n \times n$) на сфере S^{n-1}. Из-за компактности последней максимум действительно достигается. Заведём функцию Лагранжа

$$L(x, \lambda) = f(x) - \lambda\Big(\sum x_k^2 - 1\Big).$$

Посчитаем её частные производные по каждой переменной:

$$\begin{aligned}
\partial_{x_j} L(x, \lambda) &= \partial_{x_j} f(x) - 2\lambda x_j \\
&= \left(\lim_{t \to 0} \frac{\langle A(x + te_j), x + te_j\rangle - \langle Ax, x\rangle}{t}\right) - 2\lambda x_j \\
&= 2\big\langle Ax, e_j \big\rangle - 2\lambda x_j;
\end{aligned}$$

$$\partial_{\lambda_j} L(x, \lambda) = \sum x_k^2 - 1.$$

Приравнивая эти выражения к нулю, получаем

$$\operatorname{grad}_{x,\lambda} L = 0 \iff \begin{cases} \forall j : x^{\mathrm{T}} A e_j = x^{\mathrm{T}} \lambda e_j, \\ \sum x_k^2 = 1. \end{cases} \iff \begin{cases} Ax = \lambda x, \\ \|x\|^2 = 1. \end{cases}$$

Величина λ и будет равна нашему максимуму:

$$\begin{aligned}
f(x) &= \langle Ax, x\rangle \\
&= \langle \lambda x, x\rangle \\
&= \lambda \langle x, x\rangle \\
&= \lambda \|x\|^2 \\
&= \lambda.
\end{aligned}$$

Таким образом, мы доказали, что

$$\sup_{\|x\|=1} f(x)$$

равен максимальному собственному числу матрицы A (аналогично,

$$\inf_{\|x\|=1} f(x)$$

равен минимальному собственному числу).

6.2 Достаточное условие

Лемма 6.2. Пусть M — гладко параметризованное многообразие, p — точка оттуда. Пусть $x \in M$ и $x \to p$. Обозначим разность $x - p$ за h и проекцию вектора h на касательное пространство $T_p(M)$ за $\widehat{h}$. Тогда

1. $\|h - \widehat{h}\| = o(\|h\|)$,
2. $\|h\| \sim \|\widehat{h}\|$.

В этой лемме довольно много геометрической интуиции.

Доказательство. Пункт 1. Покажем, что

$$\operatorname{dist}\big(x, p + T_p(M)\big) = o(\|h\|).$$

Тогда результат будет следовать из того, что, по определению $\widehat{h}$,

$$\operatorname{dist}\big(x, p + T_p(M)\big) = \|h - \widehat{h}\|.$$

Воспользуемся тем, что касательное пространство — это образ дифференциала карты (страница 41). Пусть φ — карта многообразия в окрестности точки p; $a := \varphi^{-1}(p)$. Разложим карту по формуле Тейлора в окрестности точки a: если $t := \varphi^{-1}(x) - a$, то

$$\varphi(a + t) - \underbrace{\big(\varphi(a) + \varphi'(a)t\big)}_{\in\, p + T_p(M)} = o(\|t\|).$$

Получается, достаточно показать, что

$$o(\|t\|) = o(\|\varphi(a + t) - \varphi(a)\|).$$

Это так по билипшицевости карты φ в некоторой окрестности точки a (страница 40): для какого-то $c \in \mathbb{R}$,

$$\|t\| \leqslant c\|\varphi(a + t) - \varphi(a)\|.$$

Так мы доказали, что

$$\|h - \widehat{h}\| = o(\|h\|).$$

Пункт 2. Оценим с помощью пункта 1:

$$\begin{aligned} \|h\| &\leqslant \|\widehat{h}\| + \|h - \widehat{h}\| \\ &= \|\widehat{h}\| + o(\|h\|). \end{aligned}$$

Поделив обе части на $\|h\|$, получаем

$$1 \leqslant \frac{\|\widehat{h}\|}{\|h\|} + o(1).$$

Но, поскольку вектор $\widehat{h}$ — это ортогональная проекция h,

$$1 \geqslant \frac{\|\widehat{h}\|}{\|h\|}.$$

Так что, действительно,

$$\|h\| \sim \|\widehat{h}\|.$$

■

Теорема 6.3 (метод множителей Лагранжа, часть 2: достаточное условие). Пусть f, F, M, p, L — как в первой части (страница 50), но вдобавок функции f и F принадлежат и классу C^2. Пусть нашёлся такой вектор $\lambda \in \mathbb{R}^n$, что $L'(p, \lambda) = 0$. Обозначим $\widetilde{L} := x \mapsto L(x, \lambda)$. Тогда

1. Если гессиан функции $\widetilde{L}$ положителен на пространстве $T_p(M) \leqslant \mathbb{R}^m$, то функция f имеет строгий локальный минимум в p при условиях F.

2. Если гессиан не знакоопределён на $T_p(M)$, то f не имеет условного экстремума в точке p.

Доказательство. Обозначим гессиан функции $\widetilde{L}$ в точке p за H, а его квадратичную форму — за Q. По формуле Тейлора, при $h \to 0$,

$$\widetilde{L}(p + h) - \widetilde{L}(p) = \frac{1}{2} Q(h) + o\left(\|h\|^2\right).$$

Первый дифференциал в этой формуле равен нулю, так как по условию $L'(p, \lambda) = 0$.

Разумеется, мы хотим теперь воспользоваться положительностью формы Q на касательном пространстве. Обозначим для вектора h его проекцию на касательное пространство $T_p(M)$ за $\widehat{h}$ и переразложим квадратичную форму в правой части формулы Тейлора:

$$\begin{aligned} Q(h) &= Q\left(\widehat{h} + (h - \widehat{h})\right) \\ &= Q(\widehat{h}) + Q(h - \widehat{h}) + 2\left(h - \widehat{h}\right)^{\mathrm{T}} H \widehat{h}. \end{aligned}$$

Используя первую теорему Вейерштрасса и однородность, получаем такие оценки:

$$\begin{aligned} Q(h - \widehat{h}) &= O\left(\|h - \widehat{h}\|^2\right), \\ 2\left(h - \widehat{h}\right)^{\mathrm{T}} H \widehat{h} &= O\left(\|h - \widehat{h}\| \|\widehat{h}\|\right). \end{aligned}$$

По лемме на странице 52, каждое из этих слагаемых имеет асимптотику

$$o\left(\|\widehat{h}\|^2\right),$$

как и, в общем-то, остаточный член в изначальной формуле Тейлора. Перепишем её в новом виде:

$$\widetilde{L}(p+h)-\widetilde{L}(p)=\frac{1}{2}Q(\widehat{h})+o\left(\|\widehat{h}\|^2\right).$$

Теперь мы, наконец, можем использовать положительность формы Q на подпространстве $T_p(M)$: существует такая константа $\alpha>0$, что для любого вектора $x\in T_p(M)$ выполняется неравенство $Q(x)\geqslant\alpha\|x\|^2$ (страница 25). Тогда

$$\widetilde{L}(p+h)-\widetilde{L}(p)\geqslant\|\widehat{h}\|^2\left(\frac{\alpha}{2}+o(1)\right).$$

Так как $\|\widehat{h}\|\sim\|h\|$, множитель

$$\|\widehat{h}\|^2$$

в некоторой окрестности точки p положителен. Поэтому из неравенства выше следует, что найдётся окрестность точки p, в которой для любого вектора $p+h$, если $h\neq 0$, верно

$$\widetilde{L}(p+h)-\widetilde{L}(p)>0.$$

Если $p, p+h\in M$, эта разность переписывается так:

$$\begin{aligned}\widetilde{L}(p+h)-\widetilde{L}(p)&=f(p+h)-\lambda^{\mathrm{T}}F(p+h)-f(p)+\lambda^{\mathrm{T}}F(p)\\&=f(p+h)-f(p).\end{aligned}$$

Значит, у функции f действительно строгий локальный минимум при условии F в точке p.

Предположим теперь, что форма Q не является знакоопределённой: нашлись два таких вектора $\widehat{h}_+$ и $\widehat{h}_-$ из касательного пространства $T_p(M)$, что

$$Q(\widehat{h}_-)<0<Q(\widehat{h}_+).$$

Многообразие M локально является графиком некоторой функции на касательном пространстве (страница 49), так что можно предполагать существование векторов h_+ и h_- таких, что точки $p+h_+$ и $p+h_-$ принадлежат многообразию M, а вектора $\widehat{h}_+,\widehat{h}_-$ являются образами h_+,h_- при проекции на касательное пространство $T_p(M)$. Устремим некоторый скаляр t к нулю. Выполнены равенства

$$\widetilde{L}(p+th_+)-\widetilde{L}(p)=\frac{t^2}{2}Q(\widehat{h}_+)+o\left(t^2\right),$$
$$\widetilde{L}(p+th_-)-\widetilde{L}(p)=\frac{t^2}{2}Q(\widehat{h}_-)+o\left(t^2\right).$$

Теперь ясно, что найдётся окрестность точки p, в которой функция $\widetilde{L}$, а потому и f, не может иметь условный экстремум. ■

7 Выпуклые функции

Определение. Множество называется *выпуклым*, если вместе с любыми двумя точками x, y оно содержит отрезок $[x, y]$ между ними.

Определение. Пусть $f\colon X \to Y$, тогда *графиком* f называется множество

$$G(f) = \{(x, f(x)) \mid x \in X\} \subset X \times Y.$$

Если $Y = \mathbb{R}$, то *надграфиком* f называется множество

$$\mathrm{Epi}(f) = \{(x, t) \mid x \in X, t \in \mathbb{R}, t \geqslant f(x)\} \subset X \times \mathbb{R}.$$

Определение. Функция называется *выпуклой*, если её надграфик — выпуклое множество.

Упражнение. Функция $f\colon S \subset \mathbb{R}^m \to \mathbb{R}$ выпукла тогда и только тогда, когда S — выпуклое множество и

$$\forall x, y \in S\, \forall \lambda \in [0,1] : f\big(\lambda x + (1-\lambda)y\big) \leqslant \lambda f(x) + (1-\lambda) f(y).$$

Утверждение 7.1. Если f_1, f_2 — выпуклые функции на выпуклом множестве $S \subseteq \mathbb{R}^m$, то $\max(f_1, f_2)$ — выпуклая функция на S.

Доказательство. $\mathrm{Epi}(\max(f_1, f_2)) = \mathrm{Epi}(f_1) \cap \mathrm{Epi}(f_2)$ — выпукло как пересечение выпуклых множеств. ■

Упражнение. Если $a \geqslant 0, b \geqslant 0, f_1, f_2$ — выпуклы на S, φ — возрастающая выпуклая функция на $\mathbb{R}$, то $af_1 + bf_2$ и $\varphi \circ f_1$ выпуклы на S.

Упражнение (неравенство Йенсена). Пусть $x_1, \ldots, x_n \in S$ — точки выпуклого множества $S \subset \mathbb{R}^m$; $\lambda_1, \ldots, \lambda_n$ — положительные скаляры, которые суммируются к единице:

$$\lambda_1 + \cdots + \lambda_n = 1, \quad \{\lambda_i\} > 0;$$

f — вещественнозначная выпуклая функция на S. Тогда

$$f(\lambda_1 x_1 + \ldots \lambda_n x_n) \leqslant \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n).$$

Примеры 7.1.

1. Функция $x \mapsto \|x\|$ выпукла для любой нормы $\|\ \ \|$.

2. Если φ — линейное отображение $\mathbb{R}^m \to \mathbb{R}$ (*линейный функционал*), то φ выпукло.

3. Отображение
$$x \mapsto \max\left(\sum a_k e^{\langle A_k x, y_k\rangle}, \sum b_k e^{\langle B_k x, z_k\rangle}\right)$$
(где $\{a_k\}, \{b_k\} \geqslant 0$; $\{A_k\}, \{B_k\}$ — линейные отображения; x, $\{y_k\}$, $\{z_k\}$ — вектора) выпукло.

Теорема 7.2. Функция $f\colon \Omega \subset \mathbb{R}^m \to \mathbb{R}$ (где Ω — выпуклая область) класса C^2 выпукла тогда и только тогда, когда её гессиан в любой точке области определения неотрицателен.

Доказательство. Предположим, гессиан неотрицателен. Пусть $x, x+h \in \Omega$. Определим функцию $\varphi\colon \mathbb{R} \to \mathbb{R}$ как
$$\varphi := t \mapsto f(x+ht).$$
Достаточно проверить, что функция φ выпукла. Известно, что скалярная функция выпукла тогда и только тогда, когда её вторая производная неотрицательна. Поэтому продифференцируем φ:

$$\varphi'(t) = f'(x+ht)h,$$
$$\varphi''(t) = h^{\mathrm{T}} f''(x+ht)h.$$

Теперь можно видеть, что
$$\varphi''(t) \geqslant 0$$
из-за неотрицательности гессиана
$$f''(x+ht).$$

Предположим, что функция f выпукла. Достаточно проверить, что гессиан неотрицателен по каждому направлению вектора h. Поскольку область Ω открыта, можно считать, что найдутся точки x и h такие, что $x, x+h \in \Omega$. Рассмотрим ту же функцию φ. Из выпуклости функции f следует выпуклость φ, так что
$$\varphi''(t) \geqslant 0,$$
поэтому
$$h^{\mathrm{T}} f''(x+ht)h \geqslant 0.$$
■