Лекция 32

Запись дифференциала с помощью частных производных

Рассмотрим функцию $f : \mathbb{R}^d \to \mathbb{R}$, дифференцируемую в точке $\boldsymbol{x}$. Пусть в пространстве $\mathbb{R}^d$ задан базис $\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ ...,\ \boldsymbol{e}_d$. Записывая определение дифференцируемости через координаты $\boldsymbol{x}$ и $\boldsymbol{h}$, а также учитывая, что матрица дифференциала $df(\boldsymbol{x})$ представляют собой строку чисел $(A_1(\boldsymbol{x}), ..., A_k(\boldsymbol{x}))$ (где $df(\boldsymbol{x})\boldsymbol{e}_i = A_i(\boldsymbol{x})$ $(i = 1,\ ...,\ d)$, то есть эти числа являются образами базисных векторов $\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ ...,\ \boldsymbol{e}_d$ при отображении $df(\boldsymbol{x})$), имеем равенство

$$f(x_1 + h_1, ..., x_k + h_k) - f(x_1, ..., x_k) = A_1(\boldsymbol{x})h_1 + ... + A_k(\boldsymbol{x})h_k + o(\|h\|), \|h\| \to 0.$$

Найдём числа $A_j(\boldsymbol{x})$ $(j = 1, ..., k)$. Для этого рассмотрим смещение на вектора

$$\boldsymbol{h} = (0, ..., 0, h_j, 0, ..., 0).$$

Тогда

$$f(x_1, ..., x_{j-1}, x_j + h_j, x_{j+1}, ..., x_k) - f(x_1, ..., x_k) = A_j(\boldsymbol{x})h_j + o(|h_j|), |h_j| \to 0.$$

Разделив обе части на h_j и перейдя к пределу при $h_j \to 0$, видим, что $A_j(\boldsymbol{x}) = \frac{\partial f}{\partial x_j}(\boldsymbol{x})$, то есть *числа $A_j(\boldsymbol{x})$ являются частными производными функции f в точке $\boldsymbol{x}$.*

Рассмотрим функцию $\pi_j : \mathbb{R}^k \to \mathbb{R}$, сопоставляющую точке $\boldsymbol{x} \in \mathbb{R}^k$ её j-ю координату, то есть $\pi_j(\boldsymbol{x}) = x_j$. Тогда $\pi_j(\boldsymbol{x} + \boldsymbol{h}) - \pi_j(\boldsymbol{x}) = x_j + h_j - x_j = h_j$, то есть приращение функции π_j является линейной функцией, поэтому $d\pi_j(\boldsymbol{x})\boldsymbol{h} = h_j$ в любой точке $\boldsymbol{x} \in \mathbb{R}^k$. Дифференциал $d\pi_j(\boldsymbol{x})$ обозначается dx_j, поэтому выражение для дифференциала $df(\boldsymbol{x})$ записывается в виде

$$df(\boldsymbol{x}) = \sum_{j=1}^{k} \frac{\partial f}{\partial x_j}(\boldsymbol{x})dx_j. \tag{1}$$

Можно сказать и по-другому. Из линейной алгебры известно, что все линейные функции на векторном пространстве V над полем $\mathbb{R}$ также образуют линейное пространство над $\mathbb{R}$, которое называется *сопряженным пространством* и обозначается через V^*. Линейные функции π_j $(j = 1,\ ...,\ d)$, заданные на векторном пространстве $\mathbb{R}^d$ образуют базис в соответствующем сопряженном пространстве, который называется *дуальным или двойственным* к базису $\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ ...,\ \boldsymbol{e}_d$. Действие дифференциала на вектор мы можем записать так:

$$df(\boldsymbol{x})\boldsymbol{h} = \sum_{j=1}^{d} \frac{\partial f}{\partial x_j}(\boldsymbol{x})h_j = \sum_{j=1}^{d} \frac{\partial f}{\partial x_j}(\boldsymbol{x})\pi_j(\boldsymbol{h}),$$

то есть *сам дифференциал $df(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}$ можно представить в виде линейной комбинации линейных функций из дуального базиса, в которой коэффициенты являются частными производными в точке $\boldsymbol{x}$.* Однако $\pi_j(\boldsymbol{h}) = h_j = d\pi_j(\boldsymbol{x})\boldsymbol{h}$, поэтому, принимая обозначения $\pi_j = dx_j$, снова приходим к формуле 1.

Пример 1. *Отметим, что* ***наличие частных производных функции в точке ещё не означает, что функция в этой точке дифференцируема****. Рассмотрим, например, функцию $f(x, y) = \sqrt{|xy|}$. Если $x^2 + y^2 \neq 0$, то обе частные производные существуют. В нуле есть частные производные f и по x, и по y, но f не дифференцируема в этой точке. Действительно,*

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{h_1 \to 0} \frac{f(h_1, 0) - f(0, 0)}{h_1} = 0$$

и, аналогично, $\frac{\partial f}{\partial y}(0,0)=0$. *Если же вектор смещения* $\boldsymbol{h}=\begin{pmatrix} h \\ h \end{pmatrix}$, *то есть имеет равные координаты, и* $h>0$, *то* $f(h,h)-f(0,0)=h$. *Однако если бы функция* f *была дифференцируема в начале координат, то её приращение по определению дифференциала было бы равно* $o(h)$ *при* $h\to 0$.

Можно привести и более простой пример: пусть функция $f:\mathbb{R}^2\to\mathbb{R}$ *принимает значение* 1 *во всех точках осей* Ox *и* Oy, *а в остальных точках плоскости* Oxy *она равна* 0. *Докажите, что у* f *есть обе частные производные в точке* $O(0,0)$, *но при этом она не дифференцируема в этой точке.*

Градиент функции и матрица Якоби отображения

Пусть в пространствах, между которыми действуют отображения и функции, заданы стандартные базисы. Ниже мы определим градиент функции. Хотя он будет определён с использованием координат, но после будет показано, что этот вектор существует независимо от того, какой базис выбран в пространстве.

Определение 1. *Вектор* $\nabla f(\boldsymbol{x})=(\frac{\partial f}{\partial x_1}(\boldsymbol{x}),...,\frac{\partial f}{\partial x_d}(\boldsymbol{x}))$ *называется градиентом функции* f *в точке* $\boldsymbol{x}$.

Отметим, что выше мы увидели, что компоненты вектора градиента составляют матрицу линейного отображения $df(\boldsymbol{x})$, то есть дифференциала, а в определении выше мы рассматриваем тот же набор частных производных как набор координат вектора, то есть один и тот же набор чисел мы рассматриваем в двух разных смыслах. Роль этого набора частных производных как матрицы дифференциала мы прояснили в прошлом пункте, а свойства того же набора как вектора будет ясна из предложения ниже. При этом важно ещё раз повторить предложение, сформулированное и доказанное на прошлой лекции перед определением производной по направлению.

Предложение 1. *Пусть функция* $f:\mathbb{R}^d\mapsto\mathbb{R}$ *дифференцируема в точке* $\boldsymbol{x}$ *и* $df(\boldsymbol{x})$ *не является тождественно нулевым отображением (то есть* $\nabla f(\boldsymbol{x})\neq\boldsymbol{0}$.*) Пусть* $\boldsymbol{l}$ *– вектор единичной длины. Тогда* $\frac{\partial f}{\partial \boldsymbol{l}}(\boldsymbol{x})$ *принимает наибольшее значение, если* $\boldsymbol{l}=\frac{\nabla f(\boldsymbol{x})}{\|\nabla f(\boldsymbol{x})\|}$.

Доказательство. Из предложения о производной по направлению дифференцируемой функции следует, что $\frac{\partial f}{\partial \boldsymbol{l}}(\boldsymbol{x})=df(\boldsymbol{x})\boldsymbol{l}$, откуда, используя неравенство Коши – Буняковского и равенство (1), получим:

$$df(\boldsymbol{x})\boldsymbol{l}=(\nabla f(\boldsymbol{x}),\boldsymbol{l})\leq\|\nabla f(\boldsymbol{x})\|\cdot\|\boldsymbol{l}\|=\|\nabla f(\boldsymbol{x})\|.$$

Равенство достигается, когда вектора пропорциональны, поэтому наибольшее значение производная по направлению $\boldsymbol{l}$ принимает, когда $\boldsymbol{l}=\frac{\nabla f(\boldsymbol{x})}{\|\nabla f(\boldsymbol{x})\|}$. □

Таким образом, градиент – это вектор, в направлении которого производная функции в точке принимает наибольшее значение (и оно равно длине этого вектора). Именно это свойство градиента и можно взять в качестве его определения. В случае функции двух переменных можно вернуться к иллюстрации из конспекта прошлой лекции: градиент будет указывать то направление, в котором поверхность в точке $\boldsymbol{x}\in\mathbb{R}^2$ растёт быстрее всего, то есть $\operatorname{tg}\varphi$ максимален.

Если функция дифференцируема, то в направлении, противоположном градиенту, функция убывает с наибольшей скоростью, а в направлении, перпендикулярном направлению градиента, функция не возрастает и не убывает.

Рассмотрим снова для большей наглядности функцию $f : \mathbb{R}^2 \to \mathbb{R}$ двух переменных. *Множество точек плоскости* Oxy, *в которых функция* f *принимает постоянное значение* C, *называется линией* C*-уровня функции* f (ниже см. иллюстрацию).

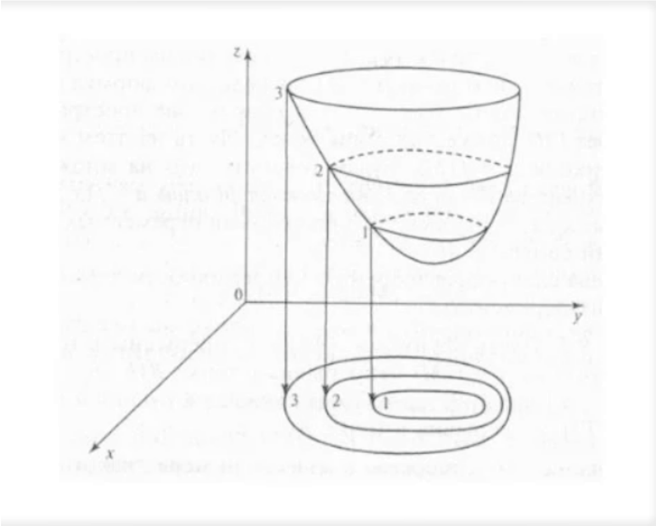

Рис. 1: Линии уровня функции

В этом случае *вектор градиента в любой точке перпендикулярен вектору, касательному к линии уровня в этой точке.* Соответственно, в противоположном градиенту направлении при выборе подходящего шага мы можем быстрее всего достичь минимального значения функции (см. рисунок 3).

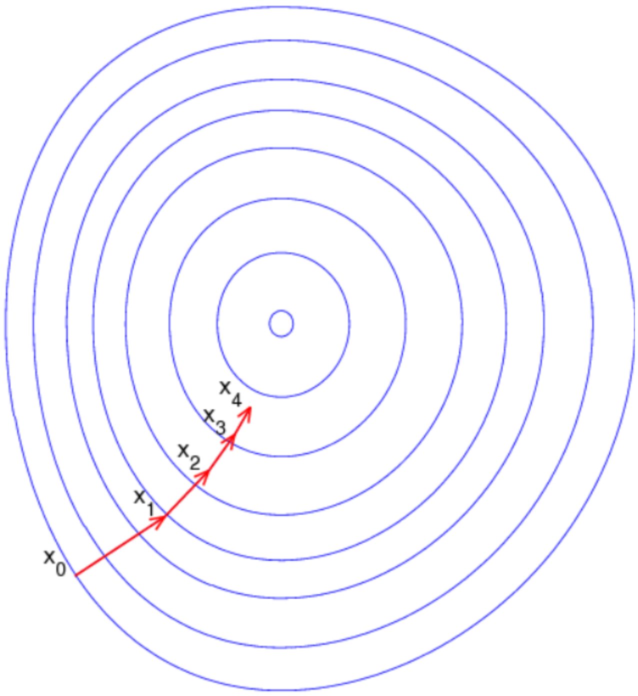

Рис. 2: Линии уровня функции синие, а красные вектора "указывают"на минимум.

На этой идее основан **метод градиентного спуска.** Суть его в том, что если известно, что дифференцируемая функция имеет один минимум, то, мы считаем градиент этой функции в какой-то точке, откладываем найденный вектор, противоположный градиенту, от этой точки, в новой точке снова считаем градиент и так далее. Получается итерационный процесс, который сходится к точке минимума функции. В точке минимума градиент будет равен нулю. Существуют различные модификации этого метода. Так как многие прикладные задачи сводятся к нахождению экстремальных значений, то метод градиентного спуска широко применяется при решении задач, например, в машинном обучении (см. риманов градиентный спуск). Отметим, однако, что если у функции несколько точек, в которых достигаются минимумы, то такой метод может и не привести к верному ответу, поэтому существуют различные его модификации.

Теперь рассмотрим дифференцируемое в точке $\boldsymbol{x} \in \mathbb{R}^d$ отображение f из $\mathbb{R}^d$ в $\mathbb{R}^m$. В этом случае дифференциал, как и любое линейное отображение из $\mathbb{R}^d$ в $\mathbb{R}^m$, задаётся матрицей, в которой m строк и d столбцов. При этом, как мы помним, необходимо и достаточно, чтобы были дифференцируемы все координатные функции

$$f_j : \mathbb{R}^d \to \mathbb{R} \ (j = 1,\ 2,\ ...,\ m)$$

отображения f, а для координатных функций, так как это функции из $\mathbb{R}^d$ в $\mathbb{R}$, мы уже знаем, как устроены дифференциалы. Зная матрицу дифференциала (или, что то же самое, вектор градиента) функции f_j, мы знаем $j-$ые координаты образов базисных векторов пространства $\mathbb{R}^d$ в пространстве $\mathbb{R}^m$. Получаем, что в $j-$ой строке матрицы отображения $df(\boldsymbol{x})$ стоит градиент координатной функции f_j. Таким образом, приходим к следующему определению.

Определение 2. *Матрицей Якоби отображения $f : \mathbb{R}^d \mapsto \mathbb{R}^m$ в точке $\boldsymbol{x}$ называется матрица, по строкам которой стоят градиенты координатных функций отображения f, то есть*

$$J_f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_1}{\partial x_2}(\boldsymbol{x}) & ... & \frac{\partial f_1}{\partial x_d}(\boldsymbol{x}) \\ \frac{\partial f_2}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_2}{\partial x_2}(\boldsymbol{x}) & ... & \frac{\partial f_2}{\partial x_d}(\boldsymbol{x}) \\ ... & ... & ... & ... \\ \frac{\partial f_m}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_m}{\partial x_2}(\boldsymbol{x}) & ... & \frac{\partial f_m}{\partial x_d}(\boldsymbol{x}) \end{pmatrix},$$

где $J_f(\boldsymbol{x})$ обозначает матрицу Якоби.

Ещё раз отметим, что из равенства (1) и определения дифференциала следует, что для дифференцируемого в точке $\boldsymbol{x}$ отображения f матрица Якоби является матрицей $df(\boldsymbol{x})$ в заданных базисах пространств $\mathbb{R}^d$ и $\mathbb{R}^m$. Полученная в примере, приведённом после определения дифференцируемости (см. прошлые лекции), матрица как раз и является матрицей Якоби линейного отображения из этого примера (проверьте это!).

Достаточные условия дифференцируемости

Дадим более общее определение окрестности точки, чем в предыдущих лекциях.

Определение 3. *Окрестностью точки $\boldsymbol{x} \in \mathbb{R}^d$ называется такое множество $U \subset \mathbb{R}^d$, что вместе с $\boldsymbol{x}$ в этом множестве содержится и некоторый шар $B_r(\boldsymbol{x})$.*

Для функций одной переменной, как мы знаем, существование производной функции в точке равносильно дифференцируемости функции в этой точке. В случае функции многих переменных, как мы видели в примере выше, это не так. Сейчас, однако, мы сформулируем *достаточные условия дифференцируемости* функции многих переменных.

Предложение 2. *Пусть у функции $f : \mathbb{R}^d \mapsto \mathbb{R}$ существуют все частные производные в некоторой окрестности U точки $\boldsymbol{x}$, а в самой точке $\boldsymbol{x}$ они непрерывны. Тогда функция f дифференцируема в точке $\boldsymbol{x}$.*

Доказательство. Мы рассмотрим случай $f : \mathbb{R}^2 \mapsto \mathbb{R}$, так как рассуждения для общего случая такие же. Пусть $\boldsymbol{h} = (h_1, h_2)$ – вектор смещения, причём $\boldsymbol{x} + \boldsymbol{h} \in U$ тогда в силу существования в U частных производных и теоремы Лагранжа имеем:

$$\begin{aligned} f(x_1+h_1, x_2+h_2) - f(x_1, x_2) &= f(x_1+h_1, x_2+h_2) - f(x_1, x_2+h_2) + f(x_1, x_2+h_2) - f(x_1, x_2) = \\ &= f'_{x_1}(\xi_1, x_2+h_2)h_1 + f'_{x_2}(x_1, \xi_2)h_2 = f'_{x_1}(x_1, x_2)h_1 + f'_{x_2}(x_1, x_2)h_2 + \alpha(\boldsymbol{h})\|\boldsymbol{h}\|, \end{aligned}$$

где $\xi_1 \in (x_1, x_1+h_1),\ \xi_2 \in (x_2, x_2+h_2)$ и

$$\alpha(\boldsymbol{h}) = (f'_{x_1}(\xi_1, x_2+h_2) - f'_{x_1}(x_1, x_2))\frac{h_1}{\|\boldsymbol{h}\|} + (f'_{x_2}(x_1, \xi_2) - f'_{x_2}(x_1, x_2))\frac{h_2}{\|\boldsymbol{h}\|}.$$

В силу непрерывности частных производных в точке $\boldsymbol{x}$ $\alpha(\boldsymbol{h}) \to 0$ при $\|\boldsymbol{h}\| \to 0$, что и даёт дифференцируемость функции f в точке $\boldsymbol{x}$. □

Упражнение 1. Запишите доказательства для случая, когда переменных d. Для этого нужно понять, как в приращении прибавлять и вычитать функции так, чтобы возникали приращения только по одному аргументу.

Следующее упражнение показывает, что существуют дифференцируемые в точке функции, частные производные которых разрывны в этой точке.

Упражнение 2. Показать, что функция

$$f(x, y) = \begin{cases} (x^2+y^2)\cos\frac{\pi}{x^2+y^2},\ x^2+y^2 \neq 0, \\ 0,\ x^2+y^2 = 0 \end{cases}$$

дифференцируема в точке $(0, 0)$, но её частные производные разрывны в этой точке.