Лекция 17. Производная

Прежде чем перейти к сути дела, дадим одно важное определение и докажем один простой технический результат.

Определение 17.1. Пусть V — векторное пространство над $\mathbb{R}$. *Нормой* на V называется функция из V в $\mathbb{R}$, обозначаемая $v \mapsto \|v\|$ и удовлетворяющая следующим аксиомам:

(1) $\|v\| \geqslant 0$ для всех $v \in V$, и $\|v\| = 0$ тогда и только тогда, когда $v = 0$.

(2) $\|\lambda v\| = |\lambda| \cdot \|v\|$ для любых $\lambda \in \mathbb{R}$ и $v \in V$.

(3) $\|v + w\| \leqslant \|v\| + \|w\|$ для любых $v, w \in V$.

Векторное пространство, на котором задана норма, называется *нормированным пространством.*

(Дословно то же определение можно дать с заменой $\mathbb{R}$ на $\mathbb{C}$.)

Примерами норм на $\mathbb{R}^n$ могут послужить L^p-метрики на $\mathbb{R}^n$ (норма — расстояние до начала координат). Другой пример — пространство ограниченных непрерывных функций с sup-нормой.

Нормированные пространства (в первую очередь бесконечномерные) играют важную роль в более продвинутых разделах анализа, но нам сейчас будет нужен только (очень простой) конечномерный случай.

Предложение 17.2. *Пусть V и W — коонечномерные пространтсва над $\mathbb{R}$; предположим, что на V выбрана норма $\|\cdot\|_1$, а на W — норма $\|\cdot\|_2$. Тогда для всякого линейного отображения $A\colon V \to W$ существует такое $C > 0$, что $\|Av\|_2 \leqslant C\|v\|_1$ для всех $v \in V$.*

(В бесконечномерном случае это утверждение совершенно неверно!)

Доказательство. Выберем раз и навсегда базис $(e_1, \ldots, e_m)$ в V и базис $(f_1, \ldots, f_n)$ в W; с помощью этих базисов отождествим V и W с $\mathbb{R}^m$ и $\mathbb{R}^n$ и введем на этих пространствах топологии, соответствующие при таких отождествлениях стандартным топологиям на $\mathbb{R}^m$ и $\mathbb{R}^n$. Если $v = x_1e_1 + \ldots + x_me_m \in V$, то

$$\|v\|_1 \leqslant \sum |x_i| \cdot \|e_i\| \leqslant C \max(|x_i|),$$

где $C = m \max(\|e_i\|_1)$; следовательно, функция $\|\cdot\|_1$ непрерывна на V. Множество

$$S = \{x_1e_1 + \ldots + x_me_m \mid \max(|x_i|) = 1\}$$

компактно и состоит из ненулевых векторов; поэтому функция $\|\cdot\|_1$ достигает на нем наименьшего значения $\lambda > 0$. Если теперь $v = x_1e_1 + \ldots + x_me_m \neq 0$, то $v/\max(|x_i|) \in S$, откуда

$$\|v\|_1 = \max(|x_i|)|v/\max(|x_i|)\|_1 \geqslant \lambda,$$

или

$$\max(|x_i|) \leqslant \lambda^{-1}\|v\|_1. \tag{17.1}$$

Если (a_{ij}) — матрица отображения A в указанных базисах и $v = x_1e_1 + \ldots + x_me_m$, то

$$\|Av\|_2 = \left\|\sum a_{ij}x_jf_i\right\|_2 \leqslant \sum |a_{ij}|\cdot|x_j|\cdot\|f_j\|_2 \leqslant$$
$$\leqslant mn\max(|a_{ij}|)\max(|x_j|)\max(\|f_i\|_2) \leqslant mn\max(|a_{ij}|)\max(\|f_i\|_2)\lambda^{-1}\|v\|_1,$$

где последнее неравенство вытекает из (17.1). Так как оэффициент при $\|v\|$ от v не зависит, все доказано. □

Следствие 17.3. *Пусть V — конечномерное векторное пространство над $\mathbb{R}$, и пусть $\|\cdot\|_1$ и $\|\cdot\|_2$ — две нормы на нем. Тогда существуют такие положительные числа m и M, что $m\|v\|_1 \leqslant \|v\|_2 \leqslant M\|v\|_1$ для всех $v \in V$.*

(В бесконечномерном случае это утверждение также неверно.)

Доказательство. В предложении 17.2 положите $V = W$, а в качестве A аозьмите тождественный оператор; затем поменяйте местами $\|\cdot\|_1$ и $\|\cdot\|_2$. □

Следствие 17.3 показывает, что, когда речь идет о функциях на подмножестве конечномерного пространства, запись типа $f(v) = o(\|v\|)$ имеет одинаковый смысл независимо от того, какая норма на этом пространстве выбрана.

Теперь мы можем дать одно из основных определений анализа.

Определение 17.4. Пусть E и F — конечномерные векторные пространства над $\mathbb{R}$ и $U \subset E$ — открытое подмножество. Тогда отображение $f\colon U \to F$ называется *дифференцируемым* в точке $p \in U$, если существует линейное отображение $A\colon E \to F$, для которого $\|f(p+h) - (f(p) + A(h))\| = o(\|h\|)$ при $h \to 0$ (через $\|\cdot\|$ обозначены нормы на E и F — все равно какие ввиду следствия 17.3).

При этом линейное отображение $A\colon E \to F$ называется *производной* отображения f в точке p (иногда говорят не «производная», а «дифференциал», но у нас этот термин будет иметь другой смысл).

Возможные обозначения для производной: $Df(p)$, $f'(p)$, T_pf, f_*; ни одно из них не является абсолютно общепринятым.

Если $E = F = \mathbb{R}$, то линейное отображение $A\colon \mathbb{R} \to \mathbb{R}$ — это просто умножение на некоторое число a, и ясно, что это число — не что иное, как известная нам производная $f'(a)$.

Поскольку величина, обозначенная через $o(\|h\|)$, стремится к нулю при $h \to 0$, и поскольку, ввиду непрерывности всякого линейного отображения конечномерных пространств, $\lim_{h\to 0}(f(p) + A(h)) = f(p)$, получаем, что $\lim_{h\to 0} f(p+h) = f(p)$, так что дифференцируемое в точке p отображение будет непрерывным в этой точке.

Перечислим простейшие свойства производной.

Предложение 17.5. *Если $A\colon E \to F$ — линейное отображение, то оно дифференцируемо в любой точке $p \in E$ и его производная равна A.*

Это очевидно.

Предложение 17.6. *Если, в обозначениях определения 17.4, отображение $f\colon U \to F$ дифференцируемо в точке $p \in U$, то его производная в точке p единственна.*

Доказательство. Если каждое из линейных отображений $A, B\colon E \to F$ является производной отображения f в точке p, то, вычитая друг из друга равенства $f(p + h) = f(p) + A(h) + \varphi(h)$ и $f(p+h) = f(p) + B(h) + \psi(h)$, получаем, что $A(h) - B(h) = \psi(h) - \varphi(h)$. Правая часть этого равенства есть $o(\|h\|)$ при $h \to 0$; если существует хотя бы одно v, для которого $A(v) - B(v) = w \neq 0$, то $\|A(tv) - B(tv)\| = t\|w\|$, и это не есть $o(\|tv\|)$. Значит, $A = B$. □

Следующее предложение — «правило дифференцирования сложной функции» в общем виде.

Предложение 17.7. *Пусть E, F и G — конечномерные векторные пространства, $U \subset E$ и $V \subset F$ — открытые подмножества, $f\colon U \to V$ и $g\colon V \to G$ — отображения. Если, для некоторой точки $p \in U$, отображение f дифференцируемо в точке p, а отображение g дифференцируемо в точке $f(p)$, то отображение $g \circ f\colon U \to G$ дифференцируемо в точке p, причем $D(g \circ f)(p) = Dg(f(p)) \circ Df(p)$.*

Доказательство. Зафиксируем какие-нибудь нормы на пространствах E, F и G; все эти нормы будем далее обозначать $\|\cdot\|$ — к путанице это не приведет.

Пусть $Df(p) = A$, $Dg(f(p)) = B$; положим

$$\tilde{f}(h) = f(p+h) - f(p) - Ah,$$
$$\tilde{g}(k) = g(f(p)+k) - g(f(p)) - Bk.$$

По определению производной, существует число $\varepsilon > 0$ и функции $\varphi, \psi\colon [0;\varepsilon] \to \mathbb{R}$, обладающие следующими свойствами:

$$\lim_{t\to 0} \varphi(t) = \lim_{t\to 0} \psi(t) = 0;\ \|\tilde{f}(h)\| \leqslant \varphi(\|h\|)\|h\|;\ \|\tilde{g}(k)\| \leqslant \psi(\|k\|)\|k\|$$

(неравенства выполняются при $\|h\|$ и $\|k\|$, не превосходящих ε). Кроме того, по предложению 17.2 существуют такие константы C_1 и C_2, что

$$\|Ah\| \leqslant C_1\|h\|, \quad \|Bk\| \leqslant C_2\|k\|$$

для всех $h \in E$, $k \in F$, Поскольку, наконец, из дифференцируемости следует непрерывность, для всех достаочно малых $\|h\|$ имеем $\|f(p+h) - f(p)\| \leqslant \varepsilon$. Для таких h имеем

$$g(f(p+h)) - g(f(p)) =$$
$$= g(f(p) + Ah + \tilde{f}(h)) - g(f(p)) = BAh + \underbrace{B\tilde{f}(h) + \tilde{g}(Ah + \tilde{f}(h))}_{R(h)}.$$

Оценим $R(h)$:

$$\|R(h)\| \leqslant \|B\tilde{f}(h)\| + \|\tilde{g}(Ah + \tilde{f}(h))\| \leqslant$$
$$\leqslant C_2\varphi(\|h\|)\|h\| + \psi(Ah + \tilde{f}(h))(\|Ah\| + \|\tilde{f}(h)\|) \leqslant$$
$$\leqslant C_2\varphi(\|h\|)\|h\| + C_1\psi(Ah + \tilde{f}(h))\|h\| + \psi(Ah + \tilde{f}(h))\varphi(\|h\|)\|h\|.$$

Поскольку все коэффициенты при $\|h\|$ стремятся к нулю при $\|h\| \to 0$, имеем $D(g \circ f)(p) = BA$, что и требовалось. □

Выясним, какой матрицей задается производная отображения.

Пусть $(u_1, \ldots, u_m)$ и $(v_1, \ldots, v_n)$ — стандартные базисы в $\mathbb{R}^m$ и $\mathbb{R}^n$: у вектора u_i (или v_i) на i-м мемте стоит единица, а на остальных — нули. В качестве норм в $\mathbb{R}^m$ и $\mathbb{R}^n$ выберем максимум модуля координат.

Пусть $U \subset \mathbb{R}^m$ — открытое множество, и пусть отображение $f\colon U \to \mathbb{R}^n$ действует по формуле

$$(x_1, \ldots, x_m) \mapsto (f_1(x_1, \ldots, x_m), \ldots, f_n(x_1, \ldots, x_m)).$$

Если f дифференцируемо в точке $p = (p_1, \ldots, p_m) \in U$ и $Df(p) = A\colon \mathbb{R}^m \to \mathbb{R}^n$, то $f(p+h) = f(p) + A(h) + o(\|h\|)$. Следовательно,

$$f(p+tu_j) = f(p) + A(tu_j) + o(\|tu_j\|) = f(p) + tA(u_j) + o(t),$$

откуда

$$A(u_j) = \lim_{t\to 0} t^{-1}(f(p+tu_j) - f(p)).$$

Если A записывается матрицей (a_{ij}), т. е. $A(u_j) = \sum a_{ij}v_i$, то, отделяя i-ую координату, получим, что

$$a_{ij} = \lim_{t\to 0} \frac{f_i(p_1, \ldots, p_{j-1}, p_j + t, p_{j+1}, \ldots, p_n) - f_i(p_1, \ldots, p_n)}{t}.$$

Иными словами, a_{ij} — производная от функции $y \mapsto f_i(p_1, \ldots, p_{j-1}, y, p_{j+1}, \ldots, p_n)$ в точке p_j. Для такого рода производных есть специальное название и обозначение:

Определение 17.8. Пусть $U \subset \mathbb{R}^m$ — открытое множество и $f\colon U \to \mathbb{R}$ — функция. Если $p = (p_1, \ldots, p_m) \in U$, то предел

$$\lim_{t\to 0} \frac{f(p_1, \ldots, p_{j-1}, p_j + t, p_{j+1}, \ldots, p_n) - f(p_1, \ldots, p_n)}{t}$$

(если он существует) называется *i-ой частной производной функции f в точке p* и обозначается $\frac{\partial f}{\partial x_i}(p)$ (иногда пишут $D_i f(p)$; применяются и другие обозначения).

Пользуясь этими обозначениями, можно подвести итоги нашего обсуждения так:

Предложение 17.9. *Пусть $U \subset \mathbb{R}^m$ — открытое множество, и пусть $f = (f_1, \ldots, f_n)$ — отображение из U в $\mathbb{R}^n$. Если f дифференцируемо в точке $p \in U$, то линейное отображение $Df(p)\colon \mathbb{R}^m \to \mathbb{R}^n$ в стандартных базисах записывается матрицей $(\partial f_i/\partial x_j)(p)$.*

Поскольку композиции линейных отображений соответствует произведение матриц, из предложения 17.7 вытекает такое

Следствие 17.10. *Пусть $U \subset \mathbb{R}^m$ и $V \subset \mathbb{R}^n$ — открытые множества, $f\colon U \to V$ — отображение, дифференцируемое всюду на U, и $g\colon V \to \mathbb{R}$ — отображение, дифференцируемое всюду на V. Если $f(x_1, \ldots, x_m) = (u_1(x_1, \ldots, x_m), \ldots, u_n(x_1, \ldots, x_m))$, то*

$$\frac{\partial g}{\partial x_i} = \sum_{j=1}^{n} \frac{\partial g}{\partial u_j}\frac{\partial u_j}{\partial x_i}.$$

Представляющийся на первый взгляд более общим случай, когда g — отображение не в $\mathbb{R}$, а в $\mathbb{R}^k$ для некоторого k, никаких новых формул для частных производных не дает.

Из теории функций одного переменного мы знаем, что полезно, наряду с дифференцируемыми функциями, рассматривать функции класса C^r. В нашем случае можно определить следующим образом.

Если E и F — векторные пространства, $U \subset E$ — открытое множество, и $f\colon U \to F$ — отображение, дифференцируемое в каждой точке U, то определено отображение $Df\colon U \to \operatorname{Hom}(E, F)$, действующее по формуле $p \mapsto Df(p)$ (напомним, что $\operatorname{Hom}(E, F)$ — множество линейных отображений из E в F, т.е. опять векторное пространство). Поэтому можно спросить, будет ли непрерывным или дифференцируемым отображение Df.

Определение 17.11. Пусть E и F — векторные пространства, $U \subset E$ — открытое множество, и $f\colon U \to F$ — отображение, дифференцируемое в каждой точке U. Говорят, что отображение f *принадлежит к классу* C^1, если отображение $Df\colon E \to \operatorname{Hom}(E, F)$ непрерывно; говорят, что f *принадлежит к классу* C^r, где $r \geqslant 2$ — целое число, если отображение $Df\colon E \to \operatorname{Hom}(E, F)$ принадлежит к классу C^{r-1}; говорят, наконец, что f *принадлежит к классу* C^∞, если f принадлежит к классу C^r для всякого $r \in \mathbb{N}$. Наконец, отображения класса C^r (для некоторого r) называют *гладкими*.

Предложение 17.12. *Пусть $U \subset \mathbb{R}^m$ — открытое множество, $f = (f_1, \ldots, f_n)$ — отображение из U в $\mathbb{R}^n$. Отображение f принадлежит классу C^1 тогда и только тогда, когда все частные производные $\partial f_i/\partial x_j$ существуют и непрерывны всюду на U.*

Из одного только существования частных производных дифференцируемость, вообще говоря, не следует.

Доказательство. «Только тогда»: при стандартном отождествлении $\operatorname{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ с пространством $m \times n$-матриц отображение $p \mapsto Df(p)$ переводит точку p в матрицу $((\partial f_i/\partial x_j)(p))$, и это отображение непрерывно тогда и только тогда, когда все $\partial f_i/\partial x_j$ непрерывны.

«Тогда»: пусть все $\partial f_i/\partial x_j$ непрерывны, и пусть $p \in U$. Достаточно установить дифференцируемость отображения f в точке p. Пусть $A\colon \mathbb{R}^m \to \mathbb{R}^n$ — линейное отображение с матрицей $a_{ij} = (\partial f_i/\partial x_j)(p)$. Согласно предложению 17.9, производная f в точке p, если она существует, обязана равняться A. Покажем, что так оно и есть.

В качестве нормы в обоих пространствах будем использовать максимум модуля координат. Тогда, расшифровывая определения получаем, что нам необходимо установить соотношения

$$|f_i(p+h) - f_i(p) - \sum_{j=1}^{m}(\partial f_i/\partial x_j)(p)h_j| = o(\|h\|)$$

для всякого i (в вышеприведенной формуле подразумевается, что $h = (h_1, \ldots, h_m)$). Для этого, в свою очередь, докажем такую лемму:

Лемма 17.13. *Пусть $U \subset \mathbb{R}^m$ — открытое множество, $f\colon U \to \mathbb{R}$ — функция, для которой в каждой точке множества U существуют все частные производные $\partial f/\partial x_j$. Если $\Delta \subset U$ — параллелепипед и $p \in \Delta$, то для всякого $h \in \mathbb{R}^m$ такого, что $p + h \in \Delta$, существуют такие точки $\pi_1, \ldots, \pi_m \in \Delta$, что*

$$f(p+h) - f(p) = \sum_{j=1}^{m}(\partial f/\partial x_j)(\pi_j)h_j.$$

Доказательство леммы. Положим $p = (p_1, \ldots, p_m)$, $h = (h_1, \ldots, h_m)$, и преобразуем разность $f(p+h) - f(p)$ так:

$$\begin{aligned}
&f(p+h) - f(p) = \\
&\qquad (f(p_1+h_1, p_2+h_2, \ldots, p_m+h_m) - f(p_1, p_2+h_2, \ldots, p_m+h_m))+ \\
&\qquad (f(p_1, p_2+h_2, \ldots, p_m+h_m) - f(p_1, p_2, p_3+h_3 \ldots, p_m+h_m))+ \\
&\qquad \cdots + \\
&\qquad (f(p_1, \ldots, p_{m-1}, p_m+h_m) - f(p_1, \ldots, p_{m-1}, p_m)).
\end{aligned}$$

Из теоремы Лагранжа вытекает, что j-ая скобка равна

$$(\partial f/\partial x_j)(p_1+h_1, \ldots, p_{j-1}+h_{j-1}, \xi_j, p_{j+1}, \ldots, p_n) \cdot h_j,$$

где число ξ_j лежит между p_j и $p_j + h_j$. Полагая

$$\pi_j = (p_1+h_1, \ldots, p_{j-1}+h_{j-1}, \xi_j, p_{j+1}, \ldots, p_n),$$

получаем утверждение леммы. $\square$

Применим доказанную лемму к функции f_i. Пусть $\Delta \subset U$ — параллелепипед, содержащий точку p. Если h таково, что $p + h \in \Delta$, то из леммы, примененной к функции f_i, вытекает, что

$$\left| f_i(p+h) - f_i(p) - \sum_{j=1}^{m} (\partial f_i/\partial x_j)(p) h_j \right| \leqslant \\ \left(\sum_{j=1}^{m} |(\partial f_i/\partial x_j)(\pi_j) - (\partial f_i/\partial x_j)(p)| \right) \cdot |h_j|. \quad (17.2)$$

Положим

$$M(\varepsilon) = \sup_{\substack{\|q-p\| \leqslant \varepsilon \\ 1 \leqslant j \leqslant m}} |(\partial f_i/\partial x_j)(q) - (\partial f_i/\partial x_j)(p)|.$$

Из неравенства (17.2) вытекает, что

$$|f_i(p+h) - f_i(p) - \sum_{j=1}^{m} (\partial f_i/\partial x_j)(p) h_j| \leqslant nM(\varepsilon) \cdot \|h\| \quad (17.3)$$

при $\|h\| \leqslant \varepsilon$; с другой стороны, из непрерывности всех частных производных в точке p следует, что $\lim_{\varepsilon \to 0} M(\varepsilon) = 0$. Значит, левая часть неравенства (17.3) есть $o(\|h\|)$, что и требовалось. $\square$