Лекция 14. Производная

Сейчас мы дадим одно из основных определений анализа.

Определение 14.1. Пусть E и F — нормированные векторные пространства над $\mathbb{R}$, $U \subset E$ — открытое подмножество, $f\colon U \to F$ — отображение. Говорят, что отображение f является *дифференцируемым* в точке $p \in U$, если существует непрерывное линейное отображение $A\colon E \to F$, обладающее тем свойством, что $\|f(p+h) - (f(p) + A(h))\| = o(\|h\|)$ при $h \to 0$. 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; ни одно из них не является абсолютно общепринятым.

Комментарии. 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 $f'(a)$ (см. предложение 7.2).

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

Если пространства E и F конечномерны, то непрерывность линейного отображения A, требуемая согласно определению производной, будет выполнена автоматически (предложение 13.9), плюс у нас нет надобности уточнять, какие именно нормы выбраны в пространствах E и F (замечание 13.8). В нашем курсе мы сосредоточим внимание на конечномерном случае.

Наконец, отметим, что дословно такое же определение производной можно дать для случая, когда E и F — векторные пространства над $\mathbb{C}$, а не над $\mathbb{R}$. Однако аналогия между вещественным и комплексным случаем заходит не слишком далеко, и изучать (очень важный) комплексный случай лучше не параллельно с вещественным, а последовательно.

Перечислим простейшие свойства производной.

Предложение 14.2. *Если $A\colon E \to F$ — непрерывное линейное отображение, то оно дифференцируемо в любой точке $p \in E$, и его производная равна A.*

Это очевидно.

Предложение 14.3. *Если, в обозначениях определения 14.1, 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 $h \to 0$; из предложения 13.11 теперь следует, что линейное отображение $h \mapsto A(h) - B(h)$ есть тождественный нуль, что и требовалось. $\square$

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

Предложение 14.4. *Пусть E, F и G — нормированные 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. Обозначив $Df(p) = A$, $Dg(f(p)) = B$, запишем равенства $f(p + h) = f(p) + A(h) + \varphi(h)$ и $g(f(p) + k) = g(f(p)) + B(k) + \psi(k)$, где $\|vf(h)\| = o(\|h\|)$ и $\|\psi(k)\| = o(\|k\|)$. Из этих равенств имеем

$$\begin{aligned} g(f(x+h)) &= g(f(p)) + B(A(h) + \varphi(h)) + \psi(A(h) + \varphi(h)) = \\ &= g(f(p)) + B(A(h)) + B(\varphi(h)) + \psi(A(h) + \varphi(h)). \end{aligned}$$

Поскольку линейное отображение $h \mapsto B(A(h))$ непрерывно как композиция двух непрерывных (в конечномерном случае эта оговорка, ввиду предложения 13.9, вообще не нужна), для доказательства предложения нам надо лишь убедиться, что $\|B(\varphi(h)) + \psi(A(h) + \varphi(h))\| = o(\|h\|)$. Для этого, в свою очередь, достаточно установить два соотношения: $\|B(\varphi(h)\| = o(\|h\|)$ и $\|\psi(A(h) + \varphi(h))\| = o(\|h\|)$.

Чтобы установить первое соотношение, заметим, что, ввиду непрерывности линейного отображения B, существует такое число $Q > 0$, что $\|B(y)\| \leqslant Q\|y\|$ для любого $y \in F$. Стало быть, $\|B(\varphi(h))\| \leqslant Q\|\varphi(h)\|$. Поскольку $\|\varphi(h)\| = o(\|h\|)$, первое из наших соотношений доказано. Чтобы установить второе соотношение, заметим, что, ввиду непрерывности линейного отображения A, существует такое число $P > 0$, $\|A(h)\| \leqslant P\|h\|$ для любого $h \in E$. Кроме того, поскольку $\|\varphi(h)\| = o(\|h\|)$, существует такое $\delta > 0$, что $\|\varphi(h)\| \leqslant \|h\|$ при всех $\|h\| < \delta$ (для этой грубой оценки достаточно было бы и более слабого условия $\|\varphi(h)\| = O(\|h\|)$, но нам этого хватит; «на всю катушку» условием $\|\varphi(h)\| = o(\|h\|)$ мы воспользовались на предыдущем шаге). Стало быть, $\|A(h) + \varphi(h)\| \leqslant (P+1)\|h\|$, как только $\|h\| < \delta$. Так как $\|\psi(k)\| = o(\|k\|)$, для любого $\varepsilon > 0$ существует такое $\delta_1 > 0$, что $\|\psi(k)\| < \varepsilon\|k\|$ при $\|k\| < \delta_1$. Стало быть, если $\|h\|$ меньше как δ, так и $\delta_1/(P+1)$, то

$$\|\psi(A(h) + \varphi(h))\| < \varepsilon\|A(h) + \varphi(h)\| \leqslant (Q+1)\varepsilon\|h\|.$$

Поскольку Q постоянно, а число $\varepsilon > 0$ произвольно, это означает, что $\|\psi(A(h) + \varphi(h))\| = o(\|h\|)$, что и завершает доказательство. $\square$

В конечномерном случае линейные отображения удобно задавать матрицами, так что грех не поинтересоваться, какой матрицей задается производная отображения.

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

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

Пусть f дифференцируемо в точке $p = (p_1, \dots, p_m) \in U$, и пусть $Df(p) = A\colon \mathbb{R}^m \to \mathbb{R}^n$. Тогда, по определению, $f(p+h) = f(p) + A(h) + \varphi(h)$, где $\|\varphi(h)\| = o(\|h\|)$. Пусть $(u_1, \dots, u_m)$ и $(v_1, \dots, v_n)$ — стандартные базисы в $\mathbb{R}^m$ и $\mathbb{R}^n$; в качестве нормы в $\mathbb{R}^m$ и $\mathbb{R}^n$ выберем максимум модуля координат в этих базисах. Если отображение A записывается матрицей (a_{ij}), так что $A(u_j) = a_{1j}v_1 + \dots + a_{nj}v_n$, то для малых по модулю $t \in \mathbb{R}$ имеем

$$f(p+tu_j) = f(p) + tA(u_j) + \varphi(tu_j) = f(p) + t(a_{1j}v_1 + \dots + a_{nj}v_n) + \varphi(tu_j);$$

отделяя i-ую координату, получим

$$f_i(p_1, \dots, p_{j-1}, p_j + t, p_{j+1}, \dots, p_n) = f_i(p_1, \dots, p_n) + ta_{ij} + \varphi_i(tu_j). \tag{14.1}$$

Поскольку $|\varphi_i(tu_j)| \leqslant \|\varphi(tu_j)\|$ и $\|tu_j\| = |t|$, получаем, что $|\varphi_i(tu_j)| = o(|t|)$. Стало быть, равенство (14.1) означает, что a_{ij} — производная от функции $y \mapsto f_i(p_1, \dots, p_{j-1}, y, p_{j+1}, \dots, p_n)$ в точке p_j. Для такого рода производных есть специальное название и обозначение:

Определение 14.5. Пусть $U \subset \mathbb{R}^m$ — открытое множество и $f\colon U \to \mathbb{R}$ — функция. Если $p = (p_1, \dots, p_m) \in U$, то предел

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

называется (если он существует) *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 14.6. *Пусть $U \subset \mathbb{R}^m$ — открытое множество, и пусть $f = (f_1, \dots, f_n)$ — 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, предложение 14.4 дает в конечномерном случае такое

Следствие 14.7. *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* $f(x_1, \dots, x_m) = (u_1(x_1, \dots, x_m), \dots, u_n(x_1, \dots, x_m))$, *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, не дает никаких новых формул для частных производных.

Из теории функций одного переменного мы знаем, что полезно, наряду с дифференцируемыми функциями, рассматривать функции класса C^r. В нашем случае эти понятия также очень полезны, но соответствующие определения дать чуть сложнее. Вот как это делается.

Если E и F — нормированные векторные пространства, $U \subset E$ — открытое множество, и $f\colon U \to F$ — отображение, дифференцируемое в каждой точке U, то определено отображение $Df\colon U \to \mathrm{Hom}(E, F)$, действующее по формуле $p \mapsto Df(p)$ (напомним, что $\mathrm{Hom}(E, F)$ — множество линейных отображений из E в F; в бесконечномерном случае здесь и далее вместо $\mathrm{Hom}(E, F)$ надо рассматривать пространство $\mathcal{L}(E, F)$, состоящее из непрерывных линейных отображений). Однако $\mathrm{Hom}(E, F)$ — также векторное пространство (в бесконечномерном случае $\mathcal{L}(E, F)$ — *нормированное* векторное пространство, так как его можно снабдить определенной в предыдущей лекции операторной нормой). Поэтому можно спросить, будет ли непрерывным или дифференцируемым отображение Df. Этот ход мысли приводит к такому определению, которое мы дадим сразу для конечномерного случая (в бесконечномерном случае надо только заменить Hom на $\mathcal{L}$):

Определение 14.8. Пусть E и F — конечномерные 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 $Df\colon E \to \mathrm{Hom}(E, F)$ непрерывно; говорят, что f *принадлежит к классу* C^r, где $r \geqslant 2$ — целое число, если отображение $Df\colon E \to \mathrm{Hom}(E, F)$ 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.*

До сих пор у нас не было ни одного нетривиального примера дифференцируемого отображения. Следующее предложение снабжает нас большим запасом дифференцируемых отображений (и отчасти объясняет, почему отображения классов C^r важнее, чем дифференцируемые).

Предложение 14.9. *Пусть* $U \subset \mathbb{R}^m$ *— открытое множество,* $f = (f_1, \dots, f_n)$ *— отображение из* U *в* $\mathbb{R}^n$. *Отображение* f *принадле-*

жит 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. Мы не будем приводить приличествующие случаю контрпримеры.

Доказательство. Часть «только тогда» мгновенно следует из предыдущих обсуждений. В самом деле, отождествим стандартным образом $\mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ с пространством $m \times n$-матриц. Если f дифференцируемо всюду на U, то отображение $p \mapsto Df(p)$ переводит точку p в матрицу $((\partial f_i/\partial x_j)(p))$, и ясно, что это отображение непрерывно тогда и только тогда, когда все $\partial f_i/\partial x_j$ непрерывны.

Доказательство в противоположную сторону менее тривиально. Именно, пусть все $\partial f_i/\partial x_j$ непрерывны, и пусть $p \in U$. Ввиду уже доказанного достаточно установить дифференцируемость функции f в точке p. Положим $a_{ij} = (\partial f_i/\partial x_j)(p)$, и пусть $A\colon \mathbb{R}^m \to \mathbb{R}^n$ — линейное отображение с матрицей (a_{ij}). Согласно предложению 14.6, производная f в точке p, если она вообще 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 14.10. *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:

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

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. □

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

$$\Bigl|f_i(p+h)-f_i(p)-\sum_{j=1}^{m}(\partial f_i/\partial x_j)(p)h_j\Bigr|\leqslant \Bigl(\sum_{j=1}^{m}|(\partial f_i/\partial x_j)(\pi_j)-(\partial f_i/\partial x_j)(p)|\Bigr)\cdot|h_j|. \quad (14.2)$$

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 (14.2) вытекает, что

$$|f_i(p+h)-f_i(p)-\sum_{j=1}^{m}(\partial f_i/\partial x_j)(p)h_j|\leqslant nM(\varepsilon)\cdot\|h\| \quad (14.3)$$

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 (14.3) есть $o(\|h\|)$, что и требовалось. □

В заключение этой лекции введем три важных термина. Ниже словосочетание «гладкое отображение» будет означать «отображение класса C^r»; в качестве r может выступать любое положительное целое число или ∞.

Во-первых, пусть $U,V\subset\mathbb{R}^n$ — открытые множества. Отображение $f\colon U\to V$ называется *диффеоморфизмом,* если оно гладко, взаимно однозначно, и обратное к нему отображение также гладко (того же класса гладкости). Если $f\colon U\to V$ — диффеоморфизм, то из предложения 14.4 следует, что $Df(p)$ — изоморфизм для всякой $p\in U$, так что диффеоморфизм между открытыми подмножествами в пространствах разных размерностей невозможен.

Пусть теперь $U\subset\mathbb{R}^m$ — открытое подмножество. Гладкое отображение $f\colon U\to\mathbb{R}^n$ называется *иммерсией* (синоним: *погружение*), если линейное отображение $Df(p)$ инъективно для всякого $p\in U$. (Из определения ясно, что при этом $m\leqslant n$).

Если же, в тех же условиях, $Df(p)$ сюръективно для всех $p\in U$, то отображение f называется *субмерсией.* (При этом, как явствует из определения, обязательно $m\geqslant n$.)