

Правила дифференцирования

В прошлый раз рассмотрели правила дифференцирования в общем случае.

Теорема 1. Дифференцирование сложной функции: Пусть X, Y, Z - нормированные пространства, функция $f: X \rightarrow Y$ - дифференцируема в точке $a \in X$, функция $g: Y \rightarrow Z$ - дифференцируема в точке $f(a)$. Тогда $g(f(x))$ - дифференцируема в точке a и верно следующее:

$$dg \circ f = dg \circ df$$

Сложность заключается в переходе на конкретные примеры.

Частные случаи дифференцируемости сложных функций

(I) Функции из \mathbb{R}^n в \mathbb{R}^m , из \mathbb{R}^m в \mathbb{R}^k

Пусть $f: \mathbb{R}_x^n \rightarrow \mathbb{R}_y^m$, $g: \mathbb{R}_y^m \rightarrow \mathbb{R}^k$, тогда:

$$df(h) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = J_f \cdot h, \quad dg(v) = \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial y_1} & \cdots & \frac{\partial g_k}{\partial y_m} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix} = J_g \cdot v$$

В этом случае, композиция запишется следующим образом:

$$\begin{aligned} f: x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &\rightarrow \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}, \quad g: y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \rightarrow \begin{pmatrix} g_1(y) \\ \vdots \\ g_k(y) \end{pmatrix} \Rightarrow \\ \Rightarrow g(f(x)) &= \begin{pmatrix} g_1(f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)) \\ \vdots \\ g_k(f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)) \end{pmatrix} \end{aligned}$$

Как выглядит дифференциал в таком случае? С одной стороны, он будет иметь следующий вид:

$$d(g \circ f)(h) = dg(f)(h) = \begin{pmatrix} \frac{\partial g_1(f(x))}{\partial x_1} & \cdots & \frac{\partial g_1(f(x))}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k(f(x))}{\partial x_1} & \cdots & \frac{\partial g_k(f(x))}{\partial x_n} \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = J_{g(f)} \cdot h$$

Получили матрицу Якоби функции $g(f)$. С другой стороны, это должно быть композицией двух дифференциалов $df(h)$ и $dg(v)$, распишем их по определению:

$$(dg \circ df)(h) = dg(df(h)) = dg(J_f \cdot h) = J_g \cdot (J_f \cdot h)$$

Распишем дифференцирование сложной функции подробнее:

$$\begin{aligned} d(g \circ f)(h) &= J_{g(f)} \cdot h = J_g \cdot (J_f \cdot h) = (dg \circ df)(h) \\ \begin{pmatrix} \frac{\partial g_1(f(x))}{\partial x_1} & \cdots & \frac{\partial g_1(f(x))}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k(f(x))}{\partial x_1} & \cdots & \frac{\partial g_k(f(x))}{\partial x_n} \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} &= \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial y_1} & \cdots & \frac{\partial g_k}{\partial y_m} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \end{aligned}$$

Как устроены элементы матрицы Якоби $J_{g(f)}$, рассмотрим подробнее:

$$\frac{\partial g_l}{\partial x_s} = \frac{\partial g_l}{\partial y_1} \cdot \frac{\partial f_1}{\partial x_s} + \cdots + \frac{\partial g_l}{\partial y_m} \cdot \frac{\partial f_m}{\partial x_s}, \quad l = \overline{1, k}, \quad s = \overline{1, n}$$

(II) Функции из \mathbb{R} в \mathbb{R}^2 , из \mathbb{R}^2 в \mathbb{R}

Пусть $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ - дифференцируема в каждой точке. Рассмотрим отображение $\mathbb{R} \rightarrow \mathbb{R}^2$, которое устроено так: $t \mapsto (x_1(t), x_2(t))$, где x_i - дифференцируемые функции. На основе этого отображения построим новую функцию:

$$t \mapsto (x_1(t), x_2(t)) \mapsto g(x_1(t), x_2(t))$$

Отображение рисует некоторую кривую $(x_1(t), x_2(t))$, а затем мы смотрим, как функция g меняется вдоль этой кривой. Таким образом, нам нужно посмотреть как устроена функция: $f(t) = g(x_1(t), x_2(t))$.

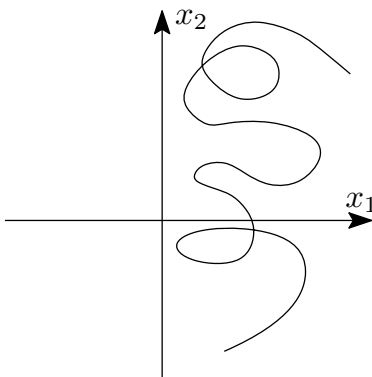


Рис. 1: Отображение $t \mapsto (x_1(t), x_2(t))$.

Первое, что обычно хотят понять, как устроена производная такой функции (заметим, функции одной переменной). Рассмотрим дифференциал f :

$$df(v) = dg \left(\left(\begin{pmatrix} \frac{\partial x_1}{\partial t} \\ \frac{\partial x_2}{\partial t} \end{pmatrix} v \right) \right) = dg \left(\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} v \right) = \left(\frac{\partial g}{\partial x_1} \quad \frac{\partial g}{\partial x_2} \right) \cdot \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} v$$

Перепишем это выражение в производных (опуская домножение на вектор v):

$$\frac{df}{dt} = \frac{\partial g}{\partial x_1} \cdot \dot{x}_1 + \frac{\partial g}{\partial x_2} \cdot \dot{x}_2$$

Докажем этот частный случай еще раз, для лучшего понимания.

□ Рассмотрим производную функции f по переменной t :

$$\begin{aligned} \left. \frac{df}{dt} \right|_{t=t_0} &= \left. \frac{d}{dt} g(x_1(t), x_2(t)) \right|_{t=t_0} = \lim_{t \rightarrow t_0} \frac{g(x_1(t), x_2(t)) - g(x_1(t_0), x_2(t_0))}{t - t_0} = \\ &= \lim_{t \rightarrow t_0} \left(\frac{\frac{\partial g}{\partial x_1}(x_1(t_0), x_2(t_0))(x_1(t) - x_1(t_0)) + \frac{\partial g}{\partial x_2}(x_1(t_0), x_2(t_0))(x_2(t) - x_2(t_0))}{t - t_0} + \right. \\ &\quad \left. + \frac{\alpha(x_1(t) - x_1(t_0), x_2(t) - x_2(t_0)) \cdot \sqrt{(x_1(t) - x_1(t_0))^2 + (x_2(t) - x_2(t_0))^2}}{t - t_0} \right) = \end{aligned}$$

Переопределим $\alpha(h)$ таким образом, что $\alpha(0) = 0$ и внесем $t - t_0$ под каждый из членов в отдельности:

$$\begin{aligned}
 &= \lim_{t \rightarrow t_0} \left(\frac{\partial g}{\partial x_1} \cdot \frac{x_1(t) - x_1(t_0)}{t - t_0} + \frac{\partial g}{\partial x_2} \cdot \frac{x_2(t) - x_2(t_0)}{t - t_0} + \right. \\
 &\quad \left. + \alpha(x_1(t) - x_1(t_0), x_2(t) - x_2(t_0)) \cdot \frac{1}{\operatorname{sgn}(t - t_0)} \cdot \sqrt{\left(\frac{x_1(t) - x_1(t_0)}{t - t_0} \right)^2 + \left(\frac{x_2(t) - x_2(t_0)}{t - t_0} \right)^2} \right) = \\
 &= \frac{\partial g}{\partial x_1} \cdot \dot{x}_1(t_0) + \frac{\partial g}{\partial x_2} \cdot \dot{x}_2(t_0) + 0 = \frac{\partial g}{\partial x_1} \cdot \dot{x}_1(t_0) + \frac{\partial g}{\partial x_2} \cdot \dot{x}_2(t_0)
 \end{aligned}$$

■

Частные случаи дифференцируемости обратной функции

Вспомним, о чем говорит теорема о дифференцируемости обратной функции. Если выполнены условия теоремы, то мы знаем, что дифференциал обратной функции df^{-1} есть обратное отображение к дифференциалу функции $(df)^{-1}$.

Теорема 2. О дифференцируемости обратной функции: Пусть X, Y - нормированные пространства, множества $\mathcal{U} \subset X, \mathcal{V} \subset Y$ - открытые. Если функция $f: \mathcal{U} \rightarrow \mathcal{V}$ - гомеоморфизм (биекция, отображение и обратное к нему - непрерывны), f - дифференцируема в точке $a \in \mathcal{U}$ и $df: X \rightarrow Y$ имеет обратный оператор $(df)^{-1}: Y \rightarrow X$ - непрерывный линейный оператор. Тогда функция $f^{-1}: \mathcal{V} \rightarrow \mathcal{U}$ является дифференцируемой в точке $f(a)$ и $df^{-1} = (df)^{-1}$.

Функции из \mathbb{R}^n в \mathbb{R}^n

Рассмотрим $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$. Пусть оно удовлетворяет условиям теоремы о дифференцируемости обратной функции (везде или в точке). Тогда, если $df(v) = J_f \cdot v$, то:

$$df^{-1}(v) = J_{f^{-1}} \cdot v = J_f^{-1} \cdot v = (df)^{-1}(v)$$

Как найти обратную к матрице Якоби? Надо найти матрицу Якоби обратного отображения:

$$J_{f^{-1}} = J_f^{-1}$$

По-другому, если отображение $f: y = y(x)$, то получить матрицу Якоби отображения $x = x(y)$ можно следующим образом:

$$J_f = \left(\frac{\partial y_i}{\partial x_j} \right)_{i,j} \Rightarrow J_{f^{-1}} = \left(\frac{\partial x_j}{\partial y_i} \right)_{i,j} = \left(\frac{\partial y_i}{\partial x_j} \right)_{i,j}^{-1} = (J_f)^{-1}$$

Типичная задача

Заданы уравнения $x = \varphi(u, v), y = \psi(u, v)$. То есть:

$$(x, y) = f(u, v) = (\varphi(u, v), \psi(u, v))$$

Предполагается, что это хорошая замена и можно выразить u, v через x, y и она удовлетворяет всем свойствам теоремы о дифференцируемости обратной функции. Нужно найти все частные производные $\frac{\partial u}{\partial x}, \frac{\partial v}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial v}{\partial y}$ этих уравнений. Как решать? Продифференцируем по x и получим:

$$1 = \frac{\partial \varphi}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial \varphi}{\partial v} \cdot \frac{\partial v}{\partial x}, \quad 0 = \frac{\partial \psi}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial \psi}{\partial v} \cdot \frac{\partial v}{\partial x}$$

Отсюда выражаются нужные частные производные и затем то же самое делается для y .

$$0 = \frac{\partial \varphi}{\partial u} \cdot \frac{\partial u}{\partial y} + \frac{\partial \varphi}{\partial v} \cdot \frac{\partial v}{\partial y}, \quad 1 = \frac{\partial \psi}{\partial u} \cdot \frac{\partial u}{\partial y} + \frac{\partial \psi}{\partial v} \cdot \frac{\partial v}{\partial y}$$

Но здесь на самом деле записано, что если взять матрицу Якоби отображения $f: (u, v) \rightarrow (x, y)$ и умножить на матрицу Якоби обратного отображения $f^{-1}: (x, y) \rightarrow (u, v)$, то получим единичную матрицу:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\partial \varphi}{\partial u} & \frac{\partial \varphi}{\partial v} \\ \frac{\partial \psi}{\partial u} & \frac{\partial \psi}{\partial v} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} = J_f \cdot J_{f^{-1}}$$

Это ровно то правило, которое утверждает, что матрица Якоби обратного отображения это обратная матрица Якоби исходного отображения.

Rm: 1. Дифференцирование обратной функции может быть полезно еще с одной стороны. Мы знаем, что если f дифференцируема, то при $x \approx a$ мы имеем локально аффинное отображение (т.е. сдвиг плюс линейное отображение):

$$f(x) \approx f(a) + J_f \cdot (x - a)$$

Теперь хотим найти обратное отображение к $y = f(x)$, то есть буквально выразить x через $y \Rightarrow$ как найти f^{-1} ? Нужно решить уравнение $y = f(x)$ и найти чему равен x , что в явном виде во многих случаях почти невозможно. Используя дифференцирование обратной функции можем использовать приближение выше, тогда при $y \approx f(a)$ получим:

$$f^{-1}(y) \approx a + (f'(a))^{-1} \cdot (y - f(a)) = a + J_f^{-1} \cdot (y - f(a))$$

И в этом случае система превращается в линейную, которая уже хорошо решается.

Rm: 2. При этом хочется отметить, что теорема о дифференцируемости обратной функции достаточно сложна в использовании из-за множества условий. Кроме того, она не утверждает, что обратная функция существует, это занесено в условие вместе с тем, что это непрерывная функция.

Теорема об обратной функции

Одномерный случай: \mathbb{R}

В первом семестре была теорема, в которой утверждалось существование обратной функции:

Теорема 3. (Об обратной функции в \mathbb{R}) Пусть $I \neq \emptyset$ промежуток и $f: I \rightarrow \mathbb{R}$ - непрерывная, строго монотонная функция. Тогда

- 1) $f(I) = J = \{y: \exists x \in I, f(x) = y\}$ - непустой промежуток;
- 2) $f: I \rightarrow J$ - биекция;
- 3) $f^{-1}: J \rightarrow I$ - строго монотонна и непрерывна;

Следствие 1. Пусть f - непрерывно дифференцируема в окрестности точки a и $f'(a) \neq 0$. Тогда существует интервалы $\mathcal{U}(a)$ и $\mathcal{V}(f(a))$ такие, что $f: \mathcal{U} \rightarrow \mathcal{V}$ - гомеоморфизм и f^{-1} непрерывно дифференцируема на \mathcal{V} .

Опр: 1. Если функция $f: \mathcal{U} \rightarrow \mathcal{V}$ - биекция, такая что f, f^{-1} - непрерывно дифференцируемы, то говорят, что задан диффеоморфизм.

Rm: 3. Таким образом, f локально это диффеоморфизм окрестностей точек a и $f(a)$, когда $f'(a) \neq 0$.

□ Докажем следствие, используя строгую монотонность.

Поскольку f - непрерывно дифференцируема в окрестности точки a (то есть $f'(x)$ - непрерывно в окрестности точки a) и $f'(a) \neq 0$, пусть $f'(a) > 0$, тогда $\exists (a - \delta, a + \delta) = I: f'(x) > 0$. Тогда на этом интервале f строго возрастает, то есть f строго монотонна на интервале $I \Rightarrow J = f(I)$ - интервал, $f: I \rightarrow J$ - биекция и обратная функция $f^{-1}: J \rightarrow I$ - непрерывна и строго монотонна.

Мы получили гомеоморфизм и $\forall x \in I, f'(x) > 0$, тогда обратное отображение в этой точке дифференцируемо и $(f^{-1})' = \frac{1}{f'}$ (по теореме о дифференцируемости обратной функции). Поскольку $\frac{1}{f'}$ - непрерывная функция, то мы получили не только дифференцируемость обратной функции, но еще и её непрерывную дифференцируемость. ■

В общем случае такого утверждения нет, поскольку нет простого свойства монотонности, нет отношения порядка на плоскости. Но хотелось бы какое-то простое условие, гарантирующее гомеоморфизм (хотя бы локально).

Теорема Банаха для общего случая: \mathbb{R}^n

Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ и мы хотим найти обратную функцию/понять, что она существует? У нас есть функция $y = f(x)$, а хотим найти $f^{-1}: y \rightarrow x$ или по-другому - решить уравнение $y = f(x)$, как уравнение на x относительно y . Таким образом, построить $f^{-1} \Leftrightarrow$ решить уравнение $y = f(x)$ относительно x .

$$\begin{cases} y_1 &= f_1(x_1, \dots, x_n) \\ &\vdots \\ y_n &= f_n(x_1, \dots, x_n) \end{cases}$$

Решать системы уравнений мы умеем, когда они линейные, но эта система - нелинейная и в лоб решить её не получится.

Идея: Свести решение уравнения к поиску неподвижной точки, то есть свести решение задачи к уравнению $x = F(x)$. Способ сведения будет следующим:

$$y = f(x) \Leftrightarrow x = x + Q(y - f(x))$$

где Q - невырожденная матрица. Почему это лучше? Потому что, как искать неподвижные точки мы знаем (чаще всего это задачи вида: $x_{n+1} = F(x_n)$).

Теорема 4. (Банаха) Пусть (X, ρ) - полное метрическое пространство и $F: X \rightarrow X$ - сжимающее отображение с коэффициентом сжатия $0 < q < 1$, то есть:

$$\rho(F(x_1), F(x_2)) \leq q \cdot \rho(x_1, x_2), \forall x_1, x_2$$

Тогда существует единственная точка $x_0 \in X: F(x_0) = x_0$, где x_0 называется неподвижной точкой F .

Rm: 4. Данное утверждение хорошо тем, что оно четко выделяет что нужно проверять, чтобы доказать, существование неподвижной точки. Более того, поскольку помимо решения уравнения нас интересует построение обратной функции (т.е. чтобы каждому y соответствовал единственный x), эта теорема дает не только существование, но и единственность такой точки.

□

Единственность: Пусть x_0 и y_0 - неподвижные точки, тогда:

$$\rho(x_0, y_0) = \rho(F(x_0), F(y_0)) \leq q \cdot \rho(x_0, y_0), \quad 0 < q < 1 \Rightarrow \underbrace{(1 - q)}_{>0} \cdot \rho(x_0, y_0) \leq 0 \Rightarrow \rho(x_0, y_0) = 0 \Rightarrow x_0 = y_0$$

Таким образом, такая точка только одна.

Существование: Возьмем последовательность $x_{n+1} = F(x_n)$, где x_1 - произвольная точка X , тогда получим последовательность вида:

$$x_1, F(x_1), F(F(x_1)), \dots$$

Докажем, что x_n сходится через критерий Коши, то есть покажем, что x_n - фундаментальна. Рассмотрим сначала “неправильную” оценку (поскольку для доказательства необходимо расстояние между любыми двумя членами):

$$\rho(x_{n+1}, x_n) = \rho(F(x_n), F(x_{n-1})) \leq q \cdot \rho(x_n, x_{n-1}) \leq q^2 \cdot \rho(x_{n-1}, x_{n-2}) \leq \dots \leq q^{n-1} \cdot \rho(x_2, x_1)$$

Оценим расстояние между любыми двумя членами последовательности $n > m$, используя неравенство треугольника:

$$\rho(x_n, x_m) \leq \rho(x_n, x_{n-1}) + \rho(x_{n-1}, x_{n-2}) + \dots + \rho(x_{m+1}, x_m)$$

Тогда, используя “неправильную” оценку мы получим следующее:

$$\begin{aligned} \rho(x_n, x_m) &\leq q^{n-2} \cdot \rho(x_2, x_1) + \dots + q^m \cdot \rho(x_2, x_1) + q^{m-1} \cdot \rho(x_2, x_1) \leq \\ &\leq q^{m-1} \cdot \rho(x_2, x_1) + q^m \cdot \rho(x_2, x_1) + \dots + q^{n-2} \cdot \rho(x_2, x_1) + q^{n-1} \cdot \rho(x_2, x_1) + \dots = \frac{\rho(x_2, x_1) \cdot q^{m-1}}{1 - q} \end{aligned}$$

Поскольку $0 < q < 1$, то для произвольного $\varepsilon > 0$ найдется такой достаточно большой N , что будет верно следующее неравенство:

$$\forall n, m > N, \quad \rho(x_n, x_m) \leq \frac{\rho(x_2, x_1) \cdot q^{N-1}}{1 - q} < \varepsilon$$

Следовательно x_n - фундаментальна и $\exists \lim_{n \rightarrow \infty} x_n = x_0$.

Rm: 5. На самом деле, мы доказали немного больше, чем существование предела, поскольку мы можем сказать, как быстро сходится к этому пределу последовательность. Если $n \rightarrow \infty$, то $\rho(x_0, x_m) \leq C \cdot q^m$.

Поскольку F - непрерывное отображение $\Rightarrow \lim_{n \rightarrow \infty} F(x_n) = F(x_0)$, тогда:

$$F(x_0) = \lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} x_{n+1} = x_0$$

и таким образом $F(x_0) = x_0$. ■

На самом деле, этого утверждения нам мало, поскольку мы хотим построить не просто обратную функцию, а обратную имеющие некие свойства, например, непрерывность. Нам необходимо понимать, как меняется x , когда меняется y . Правда ли, что если отображения отличаются мало, то и неподвижные точки отличаются мало?

Rm: 6. Пусть (X, ρ) - полное метрическое пространство, $F, G: X \rightarrow X$ - сжимающиеся отображения с коэффициентом сжатия q (если разные, то возьмем самый большой). Пусть x_f и x_g - это их неподвижные точки. Хотим оценить расстояние между этими двумя неподвижными точками. Применяя неравенство треугольника получим:

$$\rho(x_f, x_g) = \rho(F(x_f), G(x_g)) \leq \rho(F(x_f), F(x_g)) + \rho(F(x_g), G(x_g))$$

Поскольку F - сжимающее отображение, то получим следующее:

$$\rho(F(x_f), F(x_g)) + \rho(F(x_g), G(x_g)) \leq q \cdot \rho(x_f, x_g) + \rho(F(x_g), G(x_g))$$

Перенесем $\rho(x_f, x_g)$ в левую часть и разделим на коэффициент $(1 - q)$, получим оценку:

$$\rho(x_f, x_g) \leq \frac{1}{1 - q} \cdot \rho(F(x_g), G(x_g))$$

Таким образом, если отображения F и G мало отличаются во всех точках, то и точки x_f, x_g будут отличаться мало.