5. Дифференциальное исчисление

5.1. Нормированные и банаховы пространства

Определение 5.1. Пусть X — векторное пространство (над полем вещественных чисел). *Нормой* в X называется функция $\|\cdot\| : X \to \mathbb{R}_+$, обладающая следующими свойствами:

1) $\|x\| \geq 0 \, \forall x \in X$ и $\|x\| = 0 \Leftrightarrow x = 0$.

2) $\|\lambda x\| = |\lambda| \|x\| \, (\lambda \in \mathbb{R}, x \in X)$.

3) $\|x + y\| \leq \|x\| + \|y\|$.

Замечание. Иногда будут использоваться векторные пространства над полем комплексных чисел. Норма в таких пространствах обладает теми же свойствами, что и для вещественных, где $|\lambda|$ — это, разумеется, модуль комплексного числа λ.

Пространство X, оснащенное нормой $\|\cdot\|$ называется *нормированным*. Нормированное пространство $(X, \|\cdot\|)$ становится *метрическим*, если положить $d(x, x') = \|x - x'\|$; тем самым оно оснащается и топологией.

Упражнение 5.1. Докажите, что функция $d(\cdot, \cdot)$ удовлетворяет аксиомам метрики.

Пример 5.1. $X = \mathbb{R}$ или $X = \mathbb{C}$, $\|x\| = |x|$ (Это нормирование приводит к стандартной метрике на $\mathbb{R}$ или $\mathbb{C}$.)

Пример 5.2. $X = \mathbb{R}^n$, $\|x\|_p := (\sum_{k=1}^n |x_k|^p)^{1/p}$, $1 \leq p < \infty$, $\|x\|_\infty = \max_k |x_k|$. Важнейшие случаи: $p = 2$ — евклидова норма, $p = \infty$ — равномерная (или чебышевская) норма и $p = 1$ — l_1("эль один")-норма. Пространство $\mathbb{R}^n$ с нормой $\|\cdot\|_p$ будем обозначать l_p^n.

Пример 5.3. Пусть T —компакт. Пространство непрерывных функций на T с нормой $\|x(\cdot)\| := \max_{t \in T} |x(t)|$ обозначается $C(T)$.

Пример 5.4. Рассмотрим пространство бесконечных последовательностей $\{x = (x_1, x_2, \ldots) \mid \sum_{n \in \mathbb{N}} |x_n|^2 < \infty\}$ с евклидовой нормой $\|x\| := (\sum_{n \in \mathbb{N}} |x_n|^2)^{1/2}$.

Это пространство называется пространством l_2 (эль два). Оно является бесконечномерным обобщением евклидова пространства $\mathbf{E}^n$. Пространства $\mathbf{E}^n$ и l_2 во многом сходны, но есть важное отличие: *единичный шар пространства l_2, несмотря на то, что он ограничен и замкнут, компактным не является.*

Во всех приведённых примерах возможно рассматривать пространства над полем комплексных чисел.

Упражнение 5.2. Проверьте, что введенные в примерах 1 — 4 функции $\|\cdot\|$ удовлетворяют аксиомам нормы.

Определение 5.2. Нормированное пространство, полное относительно введенной метрики $d(x, x') = \|x - x'\|$ называется *банаховым* пространством.

Упражнение 5.3. Докажите, что пространства примеров 1 — 4 являются банаховыми.

Примером нормированного, но не банахова пространства является пространство $C_2([0,1])$ непрерывных функций с евклидовой нормой

$$\|x(\cdot)\| := (\int_0^1 x^2(t)dt)^{1/2}.$$

Упражнение 5.4. Докажите это утверждение.

Замечание. Определенные выше пространства $C([a,b])$ и l_2 играют в анализе исключительную роль: пространство $C([0,1])$ является *универсальным* сепарабельным (т. е. имеющим счётное всюду плотное множество) полным нормированным пространством (любое такое пространство изометрически вкладывается в $C([0,1])$ — теорема Банаха), а пространство l_2 является *уникальным* — любое бесконечномерное (замкнутое) подпространство l_2 изометрично l_2 и в любом бесконечномерном нормированном пространстве имеются подпространства "почти" изометричные l_2 (так называемая "теорема Дворецкого").

Определение 5.3. Пусть X — нормированное пространство. Множество

$$BX := \{\ x \in X \mid \|x\| \leq 1\}$$

называется *единичным шаром* в X (с центром в нуле).

Шар Bl_2^n — это обычный евклидов шар в $\mathbb{R}^n$; Bl_∞^n — это n-мерный куб с ребром 2; множество Bl_1^n называют n-мерным октаэдром.

Напомним, что множество A в векторном пространстве называется *выпуклым*, если $\forall x, x' \in A$, $\forall \lambda \in [0,1]$ $\lambda x + (1-\lambda)x' \in A$.

Из определений сразу вытекает, что единичный шар является выпуклым, ограниченным симметричным, поглощающим ($\Leftrightarrow\ \forall x \in X\ \exists \lambda > 0 : x \in \lambda BX$) множеством, не содержащим прямых. Верно и обратное утверждение: *всякое выпуклое, симметричное, поглощающее, ограниченное множество B в конечномерном векторном пространстве X может служить шаром некоторого нормированного пространства* (в бесконечномерном случае надо вместо ограниченности потребовать, чтобы пересечение B с любой прямой, проходящей через начало координат, было отрезком). Этот факт был обнаружен Г.Минковским. Норма в X, порожденная множеством B, вводится *с помощью функции Минковского* следующей формулой:

$$\|x\|_B := \inf\{t > 0 \mid x \in tB\}.$$

5.2. Непрерывность и линейность отображений в нормированных пространствах.

Сделаем ещё один шаг к разрешению вопроса: что же такое дифференциал отображения. Дифференциал — это *главная линейная часть* отображения. В первой части нашего курса мы уже сталкивались с понятием линейности.

Здесь нам придется вернуться к обсуждению этого понятия, ибо оно лежит в основе понятия дифференциала. Сначала — напоминание.

Пусть X и Y — векторные простарнства. Отображение $\Lambda : X \to Y$ называется *линейным* , если $\Lambda(\lambda x + \lambda' x') = \lambda\Lambda x + \lambda'\Lambda x'$ для любых $\lambda, \lambda' \in \mathbb{R}$ и $x, x' \in X$. Если $Y = \mathbb{R}$, то линейное отображение l из X и $\mathbb{R}$ называют *линейным функционалом.*

Пусть теперь X и Y — нормированные векторные пространства. Как было сказано, тогда, они являются метрическими пространствами, а значит определено понятие *непрерывности* отображения $F : X \to Y$. Линейное отображение $\Lambda : X \to Y$ непрерывно в нуле, если для любого $\varepsilon > 0$ найдётся число $\delta > 0$ такое, что если $\|x\|_X < \delta$, то $\|\Lambda x\|_Y < \varepsilon$. Это равносильно тому, что существует константа $C > 0$ такая, что

$$\|\Lambda x\|_Y \le C\|x\|_X, \quad x \in X. \tag{1}$$

Нетрудно понять, что линейный оператор, непрерывный в нуле непрерывен в любой точке $x \in X$, т. е. непрерывен в X (так что соотношение (1) есть определение непрерывности оператора Λ). Совокупность линейных непрерывных отображений из нормированного пространства X в нормированное пространство Y обозначают $\mathcal{L}(X, Y)$. Совокупность линейных непрерывных функционалов на X образует линейное пространство, называемое *сопряженным* пространством. Его обозначают X^* (а совокупность *всех* линейных функционалов на X будем обозначать X'). Действие линейного функционала $x^* \in X^*$ ($x' \in X'$) на элемент $x \in X$ обозначаем $\langle x^*, x\rangle$ ($\langle x', x\rangle$).

Величину

$$\|\Lambda x\|_{\mathcal{L}(X,Y)} = \sup_{x\neq 0} \frac{\|\Lambda x\|_Y}{\|x\|_X}$$

(конечную для любого непрерывного линейного отображения $\Lambda : X \to Y$) называют *нормой* оператора Λ. Будем обозначать эту величину $\|\Lambda\|_{\mathcal{L}(X,Y)}$. Из определения вытекает важное неравенство

$$\|\Lambda x\|_Y \le \|\Lambda\|_{\mathcal{L}(X,Y)}\|x\|_X. \tag{1'}$$

Пример 5.5. На $\mathbb{R}$ линейные функции имеют вид $\Lambda x = ax$, где a — некоторое вещественное число.

Пример 5.6. На $\mathbb{R}^n$ линейные функции имеют такой вид
$a \cdot x = \sum_{k=1}^n a_k x_k \quad (a \in (\mathbb{R}^n)')$, а линейные отображения A из $\mathbb{R}^n$ в $\mathbb{R}^m$ описываются *матрицами*:
$Ax = (y_1, \dots, y_m)^T, y_i = \sum_{j=1}^n a_{ij} x_j$.

В этих двух примерах любые линейные функции непрерывны. В бесконечномерных пространствах это не так, но мы будем иметь дело только с непрерывными функционалами.

Пример 5.7. Примером линейных непрерывных функционалов в пространстве $C([0,1])$ могут служить функционалы типа скалярного произведения:

$$l(x(\cdot)) = \int_0^1 a(t)x(t)\ dt,$$

где $a(\cdot)$ — непрерывная функция, а также выражения, которые в виде скалярного произведения не записываются, например, $l(x(\cdot)) = x(\tau)$, где τ фиксированное число из $[0,1]$.

Пример 5.8. В пространстве l_2 любая линейная непрерывная функция имеет вид скалярного произведения:

$$l(x(\cdot)) = \langle a, x\rangle = \sum_{n\in\mathbb{N}} a_k x_k, a = (a_1, \dots, a_n \dots) \in l_2.$$

Задача 5.1 (на доказательство). *Докажите это утверждение.*

Наконец, мы подошли к цели, и теперь можем дать определение производной и дифференциала.

5.3. Производная и дифференциал

Пусть X и Y — нормированные пространства, U — окрестность точки $x_0 \in X$, $F : U \to Y$.

Определение 5.4. Отображение F называется *дифференцируемым в точке x_0 (по Фреше)*, если существует линейный непрерывный оператор Λ ($\Lambda \in \mathcal{L}(X,Y)$), такой, что

$$F(x_0 + x) = F(x_0) + \Lambda x + r(x), \text{ где } r(x) = o(x) \Leftrightarrow \lim_{x\to 0} \frac{\|r(x)\|_Y}{\|x\|_X} = 0.$$

При этом мы будем писать $F \in \mathcal{D}^1(x_0, Y)$. Оператор Λ в определении 5.4 определяется однозначно. Его называют *производной Фреше* F в точке x_0 и обозначают $F'(x_0)$. Линейное отображение $x \to F'(x_0)x$ называют *дифференциалом* отображения F в точке x_0.

Пример 5.9. Пусть $X = Y = \mathbb{R}$. Тогда, как легко понять, наше определение равносильно классическому (восходящему к Ньютону и принадлежащему Коши) определению производной функции f одного переменного:

$$f'(x_0) = \lim_{x\to 0} \frac{f(x_0 + x) - f(x_0)}{x}.$$

Пример 5.10. Пусть f — функция n переменных, т. е. $X = \mathbb{R}^n$, $Y = \mathbb{R}$. Тогда (в соответствии с определением 5. 4) $f'(x_0)$ есть линейный функционал на $\mathbb{R}^n$, т. е. $f'(x_0)\cdot x = \sum_{k=1}^n a_k x_k$. Числа a_k определяются равенствами $a_k = \lim_{h\to 0} \frac{f(x_0+he_i)-f(x_0)}{h}$ ($e_i := (0, ..., 1, ...0)$, единица — на i-том месте), называются *частными производными* функции f в точке x_0 и обозначаются $\frac{\partial f(x_0)}{\partial x_i}$ (так что $f'(x_0) = (\frac{\partial f(x_0)}{\partial x_1}, \dots \frac{\partial f(x_0)}{\partial x_n})$.

Если $X = \mathbf{E}^n$ или l_2 (у них сопряжённые пространства отождествлятся с ними самими ($\Leftrightarrow$ изоморфно им)), производная $f'(x_0)$ (функционала $f : X \to \mathbb{R}$) это элемент самого пространства X. Он называется *градиентом* f в точке x_0 и обозначается $\mathrm{grad} f(x_0)$.

Пример 5.11. Пусть $X = \mathbb{R}^n, Y = \mathbb{R}^m$. Тогда $F'(x_0)$ есть линейный оператор, задаваемый матрицей $F'(x_0) = (\frac{\partial f_i(x_0)}{\partial x_j}),\ 1 \le i \le m, 1 \le j \le n$.

Исторический комментарий. Приведенное выше определение было дано французским математиком (учеником Адамара) Морисом Фреше в 1912 году. Судя по всему в традициях французской школы не было корректного определения производной функции многих переменных. Корректное определение дифференциала в многомерном случае дал (в восьмидесятых годах 19-го столетия) Вейерштрасс. Его труды были опубликованы лишь в тридцатых годы нашего столетия, но идеи его были распространены в Германии и Англии и содержались в некоторых немецких и английских учебниках (Штольца и Юнга), написанных на рубеже столетий.

5.4. Основные теоремы дифференциального исчисления.

Принято выделять три важнейших теоремы классического дифференциального исчисления: теорему о суперпозиции (восходящую к Лейбницу), теорему о среднем (обобщение теоремы Лагранжа) и теорему об обратной функции (первый результат подобного рода был получен, судя по всему, Дини). Теорема о суперпозиции одинаково доказывается в конечномерном и бесконечномерном случаях и мы её доказываем сразу в бесконечномерном случае. Две других теоремы доказываются в конечномерном случае.

Теорема 5.1 (о суперпозиции). *Пусть X, Y, Z — нормированные пространства, U — окрестность точки x_0 в X, $\varphi : U \to Y, \varphi(x_0) = y_0$, V — окрестность точки y_0 в Y, $\psi : V \to Z, f = \psi\circ\varphi : U \to Z, \varphi \in \mathcal{D}^1(x_0, Y)$, $\psi \in \mathcal{D}^1(y_0, Z)$. Тогда $f \in \mathcal{D}^1(x_0, Z)$ и*

$$f'(x_0) \ = \ \psi'(y_0) \circ \varphi'(x_0).$$

Доказательство.

$$\varphi(x_0+x) \overset{def}{=} \varphi(x_0) \ + \ \varphi'(x_0)x \ + \ r_1(x), \|x\|_X < \delta \Rightarrow \|r_1(x)\|_Y < \varepsilon_1\|x\|_X,\ (i)$$

$$\psi(\varphi(x_0) + \varphi'(x_0)x + r_1(x)) \overset{def,(i)}{=} \psi(\varphi(x_0)) + \psi'(\varphi(x_0))(\varphi'(x_0)x + r_1(x)) +$$
$$r_2(\varphi'(x_0)x + r_1(x)) \overset{Id}{=} \psi(\varphi(x_0)) + \psi'(y_0)\varphi'(x_0)x + r_3(x),$$

где

$$\|r_2(y)\|_Z < \varepsilon_2\|y\|_Y,\ \|r_3(x)\|_Z \overset{Id}{=} \|\psi'(y_0)r_1(x) \ + \ r_2(\varphi'(x_0)x \ + \ r_1(x))\|_Z \le$$

(в силу нер. тр-ка) $\|\psi'(y_0)\|\varepsilon_1\|x\|_X + \varepsilon_2\|\varphi'(x_0)\|\|x\|_X + \varepsilon_1\varepsilon_2\|x\|_X$, откуда из определения 5.4 сразу следует утверждение теоремы. $\square$

Теорема 5.2 (о среднем). *Пусть $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$, U — окрестность в X, содержащая отрезок $[a,b]$, $f \in \mathcal{D}^1(U,Y)$. Тогда выполнено неравенство:*

$$|f(b) - f(a)| \le \sup_{c\in[a,b]} \|f'(c)\|_{\mathcal{L}(X,Y)}|b-a|.$$

Доказательство. 0.1. Формула Лагранжа конечных приращений.

Обозначим $y := \frac{(f(b)-f(a))^T}{|f(b)-f(a)|}$, $\varphi(t) := y\cdot f(a+t(b-a))$. Тогда

$$y\cdot(f(b)-f(a)) \overset{Id}{=} |f(b)-f(a)|, \tag{i}$$

и в силу теоремы 5.1

$$\varphi'(t) = y\cdot f'(a+t(b-a))(b-a), \tag{ii}$$

а из 0.1 получаем

$$\varphi(1)-\varphi(0) = \varphi'(\theta),\ \theta\in(0,1). \tag{iii}$$

В итоге:

$$\begin{aligned}|f(b)-f(a)| &\overset{(i)}{=} y\cdot(f(b)-f(a)) \overset{def}{=} \varphi(1)-\varphi(0) \overset{(ii),(iii)}{=}\\ &y\cdot f'(a+\theta(b-a))(b-a) \overset{|y|=1}{\le} |f'(a+\theta(b-a))(b-a)| \overset{(1)}{\le}\\ &\|f'(a+\theta(b-a))\|_{\mathcal{L}(X,Y)}|b-a| \overset{def}{\le} \sup_{c\in[a,b]}\|f'(c)\|_{\mathcal{L}(X,Y)}|b-a|.\end{aligned}$$

□

Следствие. *Пусть выполнены условия теоремы о среднем и $\Lambda\in\mathcal{L}(X,Y)$. Тогда*

$$|f(b)-f(a)-\Lambda(b-a)| \le \sup_{c\in[a,b]}\|f'(c)-\Lambda\|\cdot|b-a|.$$

Доказательство. Надо применить теорему о среднем к отображению

$$g(x) := f(x) - \Lambda x.$$

□

Замечание. Прямое обобщение формулы Лагранжа в многомерном случае места не имеет. Например, отображение $f:[0,2\pi]\to\mathbb{R}^2$, $f(t) = (\cos t, \sin t)$ таково, что $0 = f(2\pi)-f(0)$. В то время, как $\|f'(t)\| = 1$, $\forall t\in[0,2\pi]$, т. е. $\sup_{c\in[0,2\pi]}\|f'(c)\|(2\pi-0) = 2\pi \ne 0$.

Теорема 5.3 (об обратном отображении). *Пусть $X, Y = \mathbb{R}^n$, U — окрестность начала координат в X, $F: U\to Y$, $F(0)=0$, причем F — непрерывно дифференцируема на U и, кроме того, $F'(0)$ — обратимая матрица (иначе говоря $\det(\frac{\partial f_i(0)}{\partial x_j})\ne 0$).*

Тогда найдутся $\delta>0$ и $K>0$ такие, что для любого $y: |y|<\delta$ существует единственный элемент $x = x(y)$ такой, что $F(x)=y$ и $|x|\le K|y|$.

Доказательство. Далее мы используем такой факт: *если* $A = (a_{ij})$ — *матрица размеров* $n \times n$ *и её норма определена равенством* $\|A\| = (\sum_{ij} a_{ij}^2)^{1/2}$, *то*

$$|Ay| \le \|A\||y|. \tag{0}$$

(этот факт следует из неравенства Коши – Буняковского).

Обозначим $C = \|F'(0)^{-1}\|$. Выберем $\varepsilon > 0$ так, чтобы (в соответствии со следствием из теоремы о среднем и в силу непрерывности F' в нуле) при $|x| < \varepsilon,\ |x'| < \varepsilon$ выполнялось соотношение

$$|F(x) - F(x') - F'(0)(x - x')| \le \frac{1}{2C}|x - x'|. \tag{i}$$

Выбрав $\delta > 0$ так, чтобы

$$|y| < \ \delta \ \Rightarrow |y| < \frac{\varepsilon}{8C} \tag{ii}$$

Зафиксируем $y : |y| < \delta$. Наша цель построить $x = x(y) : \ f(x) = y$. Определим следующую итеративную процедуру (называемую *модифицированным методом Ньютона*):

$$x_0 = 0,\ x_{k+1} \ = \ x_k + (F'(0))^{-1}(y - F(x_k)),\ k \in \mathbb{Z}_+. \tag{iii}$$

Докажем по индукции, что $|x_k| < \varepsilon$ для всех k. Имеем

$$|x_1| \overset{(iii)}{=} |(F'(0))^{-1}y| \le C|y| \overset{(ii)}{<} \varepsilon/2. \tag{iv}$$

Пусть $|x_i| < \varepsilon$ при $i = 1, ..., k$. Покажем, что $|x_{k+1}| < \varepsilon$. Имеем

$$-F'(0)(x_i - x_{i-1}) + y - F(x_{i-1}) \overset{(iii)}{=} 0,\ 1 \le i \le k. \tag{v}$$

Следовательно,

$$|x_{i+1} - x_i| \overset{(iii),(0)}{\le} C|y - F(x_i)| \overset{(v)}{=} C| - y + F(x_i) + y - F(x_{i-1}) -$$
$$F'(0)(x_i - x_{i-1})| \overset{(i)}{\le} \frac{1}{2}|x_i - x_{i-1}|,\ 1 \le i \le n,$$

Значит,

$$|x_{i+1} - x_i| \le 2^{-i}|x_1| \overset{(ii)}{<} 2^{-i-1}\varepsilon \tag{vi}$$

$$\Rightarrow |x_{k+1}| \overset{\mathrm{Id}}{=} |x_{k+1} - x_k + x_{k-1} - \ldots + x_1| \le |x_{k+1} - x_k| + \ldots + |x_1| \overset{(vi)}{<}$$
$$|x_1|(1/2 + 1/4 + \cdot + 1/2^k) + |x_1| \overset{(ii)}{<} \varepsilon. \quad (vii)$$

Индукция завершена. Аналогично $|x_{k+m} - x_n| \overset{(vi)}{\le} |x_{k+1} - x_k|(1 + 1/2 + \ldots 1/2^m) \overset{(vi)}{<} 2^{k-1}|x_1|$, т. е. $\{x_k\}$ — фундаментальная последовательность, сходящаяся в силу полноты X к некоторому $x(y)$. В силу (v) $y = F(x(y))$, а из-за (vii) $|x(y)| \le \lim_{k\to\infty} |x_k| \overset{(vii)}{\le} 2|x_1| \overset{(iv)}{\le} 2C|y|$. □

Замечание. Вместо требования гладкости $F \in C^1$ можно ограничиться меньшим (и к тому же "точечным") требованием.

Пусть X и Y — нормированные пространства, U — окрестность точки $x_0 \in X$, $F : U \to Y$. Отображение F называется *строго дифференцируемым в точке* x_0, если оно дифференцируемо по Фреше в этой точке и для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что если $\|x'\| < \delta, \|x\| < \delta$, то выполнено следующее неравенство:

$$\|F(x') - F(x) - F'(x_0)(x' - x)\|_Y \le \varepsilon \|x' - x\|_X$$

Если отображение F строго дифференцируемо в точке x_0, мы будем писать $F \in S\mathcal{D}^1(x_0), Y)$. Если функция строго дифференцируема в некоторой точке, то она (по определению) непрерывна (и даже липшицева) в окрестности этой точки. Отсюда следует, что функция $f(x) = x^2 \cdot D(x)$, где $D(x)$ — функция Дирихле, доставляет пример дифференцируемой по Фреше, но не строго дифференцируемой функции.

Скажем несколько слов о дифференцируемости комплексных функций. Пусть $f : \mathbb{C} \to \mathbb{C}$ —комплексная функция. Ее можно рассматривать, как отображение $z = x + iy \to f(z) = u(x, y) + iv(x, y)$. Скажем, что функция f дифференцируема в точке $z_0 = x_0 + iy_0$, если

$$f(z_0 + z) = f(z_0) + Az + o(|z|), \quad A = \alpha + i\beta \in \mathbf{C}.$$

Откуда, положив $z = t,\ t \in \mathbb{R}$ и перейдя к пределу, получаем: $\alpha = \frac{\partial u}{\partial x}$, $\beta = \frac{\partial v}{\partial x}$. Аналогично, положив $z = it,\ t \in \mathbb{R}$ и снова переходя к пределу, получаем: $\alpha = \frac{\partial v}{\partial y}$, $\beta = -\frac{\partial u}{\partial y}$. Мы доказали, что необходимым условием дифференцируемости комплексной функции $f = u + iv$ в точке $z_0 = u_0 + iv_0$ являются следующие соотношения:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \ \frac{\partial v}{\partial y} = -\frac{\partial u}{\partial x}$$

Эти соотношения называются *условиями Коши-Римана.* Нетрудно доказать, что они являются и достаточными условиями дифференцируемости функции f в точке z_0 (при условии, что функции u и v непрерывно дифференцируемы в окрестности точки z_0).

Исторический комментарий. Производную функции одного переменного определили Ньютон и Лейбниц. Первое определение дифференциала функции многих переменных дал (об этом было сказано) — Вейерштрасс. Нечто вроде производной в бесконечномерном случае определил впервые Лагранж. Он ввел понятие *вариации* функционала. Современную расшифровку того, что мог иметь ввиду Лагранж, дали ученики Адамара — Гато и П. Леви. Более сильную производную определил Фреше (определение в 5.4), но в сороковые годы выяснилось, что для многих естественных задач анализа существование производной Фреше лишь в точке является недостаточным. Определение строгой дифференцируемости по-видимому ввёл Бурбаки. Определение производной комплексной функции восходит к Коши.

5.5. Частные производные. Старшие производные.

Пусть X, Y, Z — нормированные пространства, U — окрестность точки (x_0, y_0) в $X \times Y$, $F : U \to Z$.

Если отображение $x \mapsto F(x, y_0)$ дифференцируемо в точке x_o по Фреше, то его производная называется *частной производной по* x отображения F в точке (x_0, y_0) и обозначается $F_x(x_0, y_0)$ или $\frac{\partial F(x_0, y_0)}{\partial x}$.

Аналогично определяется частная производная по y $F_y(x_0, y_0) = \frac{\partial F(x_0, y_0)}{\partial y}$.

Теорема 5.4 (о полном дифференциале). *Пусть X, Y и Z — нормированные пространства, U — окрестность точки $(\widehat{x}, \widehat{y})$ в $X \times Y$, $F : U \to Z$ — отображение, имеющее в каждой точке $(x, y) \in U$ частные производные $F_x(x, y)$ и $F_y(x, y)$.*

Если отображения $(x, y) \mapsto F_x(x, y)$ и $(x, y) \mapsto F_y(x, y)$ непрерывны в точке $(\hat{x}, \hat{y}) \in U$, то F строго дифференцируема в этой точке и при этом

$$F'(\hat{x}, \hat{y})[\xi, \eta] \;=\; F_x(\hat{x}, \hat{y})\xi \;+\; F_y(\hat{x}, \hat{y})\eta.$$

Доказательство этой теоремы носит чисто технический характер и принципиально не отличается от доказательства того же результата для вещественной функции двух переменных. Мы его не приводим.

Следствие. *Для того, чтобы $F \in C^1(U)$ необходимо и достаточно, чтобы в U частные производные F_x и F_y были непрерывны.*

Переходим к определению производных высших порядков. Пусть X и Y — нормированные пространства, U — открытое подмножество в X. Если отображение $f : U \to Y$ дифференцируемо в каждой точке $x \in U$, то определено отображение $f' : U \to \mathcal{L}(X, Y)$. Поскольку $\mathcal{L}(X, Y)$ также является нормированным пространством, можно ставить вопрос о существовании второй производной

$$f''(x) = (f')'(x) \in \mathcal{L}(X, \mathcal{L}(X, Y)).$$

По индукции определяются производные высших порядков.

При этом требуется некоторое уточнение. Отображение $\Pi(x_1, ..., x_n)$ декартова произведения $X^n = X \times \cdots X$ в Y называется *полилинейным*, если при каждом i отображение

$$x_i \mapsto \Pi(\hat{x}_1, ..., \hat{x}_{i-1}, x_i, \hat{x}_{i+1}, ..., \hat{x}_n)$$

линейно. Если X и Y нормированны, то линейное пространство непрерывных полилинейных отображений обозначим $\mathcal{L}^n(X, Y)$. Непрерывность Π равносильна его ограниченности, т.е. конечности числа

$$\|\Pi\| \;=\; \sup_{\|x_i\| \le 1, 1 \le i \le n} \|\Pi(x_1, ..., x_n)\|.$$

$\mathcal{L}^n(X, Y)$ становится нормированным пространством, если норму задать так, как это было определено выше. При этом выполняется неравенство

$$\|\Pi(x_1, ...x_n)\| \le \|\Pi\| \; \|x_1\| ... \|x_n\|.$$

Нетрудно установить, что пространство $\mathcal{L}(X, \mathcal{L}(X, Y))$ изометрично изоморфно $\mathcal{L}^2(X, Y), ..., \mathcal{L}(X, \mathcal{L}(X, , , , \mathcal{L}(X, Y)...))$ изоморфно $\mathcal{L}^n(X, Y)$. Таким образом,

$$f^{(2)} \in \mathcal{L}^2(X, Y), ..., f^{(n)} \in \mathcal{L}^n(X, Y).$$

Приведем без доказательства следующий результат: *если X и Y — нормированные пространства, U — окрестность точки $\hat{x}$ и для функции $f : U \to Y$ существует вторая производная $f''(\hat{x})$, то соответствующая ей билинейная форма симметрична, т.е. для всех $x_1, x_2 \in X$*

$$f''(\hat{x})[x_1, x_2] \;=\; f''(\hat{x})[x_2, x_1]$$

Если $\Pi(x_1, x_2)$ — симметричная билинейная форма, то функция $\Pi(x, x)$ называется соответствующей ей квадратичной формой.

Теорема (формула Тейлора). *Пусть X и Y — нормированные пространства, U — окрестность точки $\hat{x}$, $f : U \to Y$. Если $f''(\hat{x})$ существует, то*

$$f(\hat{x} + x) = f(\hat{x}) + f'(\hat{x})[x] + \frac{1}{2}f''(\hat{x})[x,x] + o(\|x\|_X^2).$$

Доказательство. Положим

$$g(x) = f(\hat{x} + x) - f(\hat{x}) - f'(\hat{x})[x] - \frac{1}{2}f''(\hat{x})[x,x] \qquad (i).$$

Тогда, как нетрудно понять, $g(0) = g'(0) = 0$ и, кроме того, квадратичная форма $g''(0)[x,x] \equiv 0$. Из тождества

$$g''(0)[t_1x_1 + t_2x_2, t_1x_1 + t_2x_2] = t_1^2 g''(0)[x_1,x_1] + 2g''(0)[x_1,x_2] + t_2^2 g''(0)[x_2,x_2]$$

заключаем, что и билинейная форма $g''(0)$ нулевая.

Докажем, что $\|g(x)\| \le \varepsilon\|x\|^2 (ii)$ при $\|x\| < \delta$. Положим $h(x) = g'(x)$. Тогда $h(0) = h'(0) = 0$, откуда (по определению) $\|h(x)\| \le \varepsilon\|x\|$ при $\|x\| < \delta$. Тогда по теореме о среднем

$$\|g(x)\| = \|g(x) - g(0) - g'(0)[x]\| \le \sup_{\xi\in[0,x]} \|g'(\xi)\| \, \|x\| = \sup_{\xi\in[0,x]} \|h(\xi)\| \, \|x\| \overset{(ii)}{<} \varepsilon\|x\|^2.$$

Заменяя $g(x)$ его выражением (i) получаем требуемое. $\square$

Читателю можно предложить сформулировать и доказать аналогичный результат для n-той производной. Нам он не понадобится.

5.6. Начала теории экстремума.

Обычно задачи на максимум и минимум изначально формализуются на языке той области математики, в которой они возникают. Например, в геометрии ставится задача: *среди всех треугольников заданного периметра найти треугольник наибольшей площади.* А в теории экстремума задачи решаются средствами математического анализа. Для того, чтобы это было возможно, нужно, чтобы они были записаны в аналитических терминах, иначе говоря – *формализованы.* В общем виде формализованная задача выглядит так: *найти экстремум (т.е. максимум или минимум) функции* $f : X \to \bar{\mathbb{R}}$ (определенной на некотором множестве X и принимающей значения в расширенной области вещественной прямой $\overline{\mathbb{R}} = \mathbb{R} \cup \pm\infty$) *при некотором ограничении* $x \in C (C \subset X)$. Кратко будем писать так

$$f(x) \to \min(\max), \ x \in C. \qquad (P)$$

Если надо найти и максимум и минимум, пишем $f(x) \to \text{extr}$. При формализации надо, вообще говоря, точно описать все три элемента: *функцию* f, *область определения* X и *ограничение* C.

Решением задачи (P) называется такая точка $\widehat{x} \in C$, что для всех $x \in C$ выполнено неравенство $f(x) \ge f(\widehat{x})$ $(f(x) \le f(\widehat{x}))$. Кроме глобальных минимумов и максимумов, будем искать *локальные экстремумы.* При этом X должно быть топологическим пространством. Точка $\hat{x}$ называется точкой *локального минимума (максимума)*, если существует окрестность U этой точки такая, что $f(x) \ge f(\widehat{x})$ $(f(x) \le f(\widehat{x})) \forall x \in U$.

Теорема 5.5 (Ферма). *Пусть X — нормированное пространство, U — окрестность точки $\widehat{x}$, $f : U \to \mathbb{R}, f \in \mathcal{D}^1(\widehat{x}), \widehat{x}$ — локальный экстремум (максимум или минимум) функции f. Тогда*

$$f'(\widehat{x}) \;=\; 0.$$

Доказательство. Пусть $x \in X, x \neq 0$. Положим $\varphi_x(t) \;:= f(\hat{x} + tx)$. Из определения производной следует тогда, что при любом x функция $\varphi_x(\cdot)$ дифференцируема и $\varphi'_x(0) = \langle f'(\widehat{x}), x\rangle$. Если допустить, что $\langle f'(\widehat{x}), \bar{x}\rangle \neq 0$ для некоторого $\bar{x}$, то по "одномерной" теореме Ферма, доказанной в первой части курса, $\widehat{x}$ не является локальным минимумом. □

Рассмотрим задачу *с ограничениями типа равенств*:

$$f_0(x) \to \mathrm{extr}, f_i(x) = 0,\ i = 1, \ldots, m,\ x = (x_1, \ldots, x_n) \qquad (P_1)$$

Функция $\mathcal{L}(x, \lambda) \;:=\; \sum_{i=0}^{m} \lambda_i f_i(x)$ называется *функцией Лагранжа* задачи (P_1). Вектор $\lambda = (\lambda_0, ..., \lambda_m)$ называется вектором *множителей Лагранжа*

Теорема (правило множителей Лагранжа). *Пусть все функции в задаче P_1 непрерывно дифференцируемы и $\hat{x}$ доставляет локальный экстремум в (P_1). Тогда существует ненулевой вектор λ множителей Лагранжа такой, что*

$$\frac{\partial\mathcal{L}(\hat{x}, \lambda)}{\partial x_i} = 0,\ 1 \le i \le n \Leftrightarrow \sum_{i=0}^{m} \lambda_i f'_i(\hat{x}) \;=\; 0.$$

Доказательство. 0.1. Теорема о ранге матрицы. 0.2. Теорема об обратном отображении.

1. Имеется альтернатива: (I) $\{f'_i(\hat{x})\}_{i=0}^m$ линейно зависимы и (II) они линейно независимы. В первом случае утверждение теоремы сразу следует из определения линейной зависимости. Покажем, что второго случая быть не может. Действительно, по определению, ранг матрицы равен наибольшему числу линейно независимых строк матрицы. Но у матрицы $A = \left(\frac{\partial f_i(\hat{x})}{\partial x_j}\right)_{0 \le i \le m, 1 \le j \le m+1}$ имеется всего $m+1$ строка и, согласно (II) все они линейно независимы. Значит, $rank A = m + 1$.

В силу 0.1 имеется минор матрицы A порядка $m+1$, отличный от нуля. Не ограничивая себя в общности, можно считать, что этот минор лежит на пересечении первых $m + 1$ столбцов и строк матрицы A. Положим

$$y = (y_1, ..., y_{m+1}), F(y) = (F_0(y), ..., F_m(y))$$

$$F_i(y) \;=\; f_i(\hat{x}_1 + y_1, ..., \hat{x}_{m+1} + y_{m+1}, \hat{x}_{m+2}, ..., \hat{x}_n)$$

Тогда, согласно нашему допущению о миноре отличном от нуля, отображение F удовлетворяет 0.2 и, значит, при малом ε разрешима система уравнений $F(y_\varepsilon) = (\varepsilon, 0, ..., 0) \Leftrightarrow f_0(\hat{x}) = \varepsilon, f_i(\hat{x}) = 0,\ 1 \le i \le m$. Но это противоречит тому, что $\hat{x}$ — локальный экстремум в задаче.□

5.7. Приложения

Приложения дифференциального исчисления огромны, и не представляется возможным дать о них достаточно полное представление в рамках этих лекций. Здесь мы расскажем о двух приложениях теории экстремума к анализу и алгебре: к доказательству классических неравенств и к приведению квадратичной формы к главным осям.

1. Неравенства для степенных и средних степенных. Для $x \in \mathbb{R}^n$ ввудём обозначения

$$\Sigma_p(x) := \left(\frac{\sum_{i=1}^n |x_i|^p}{n}\right)^{1/p}, p \neq 0,\ \Sigma_0(x) := \left(\Pi_{i=0}^n |x_i|\right)^{1/n},$$

$$\sigma_p(x) := \left(\sum_{i=1}^n |x_i|^p\right)^{1/p},\ p \neq 0.$$

Величины $\Sigma_p(x)$, $p \neq 0$ называются *средними степенными*, $\sigma_p(x)$, $p \neq 0$ —*степенными*, Σ_0 — *средним геометрическим*, Σ_1 —*средним арифметическим*, Σ_{-1} — *средним гармоническим.*

Имеет место следующее

Предложение 1. *Если $-\infty < p < q < \infty$, то $\Sigma_p(x) \leq \Sigma_q(x), \sigma_q(x) \leq \sigma_p(x)$ (для любого x из $\mathbb{R}^n$).*

Для доказательства надо рассмотреть задачу

$$\sum_{k=1}^n |x_k|^p \to \sup, \quad \sum_{k=1}^n |x_k|^q \leq 1$$

и применить правило множителей Лагранжа.

Продемонстрируем, как это делается, в простейшем случае:

$$x_1^2 + x_2^2 \to \text{extr}; \quad x_1^4 + x_2^4 = 1.$$

В силу компактности ограничения и из принципа Вейерштрасса решение существует, и значит, оно находится методом множителей Лагранжа.

Функция Лагранжа: $\mathcal{L} = \lambda_0(x_1^2 + x_2^2) + \lambda_1 x_1^4 + x_2^4 - 1)$.

Составляем уравнения

$$\mathcal{L}_x = 0 \Leftrightarrow 2\lambda_0 x_1 + 2\lambda_1 x_1^3 = 0, 2\lambda_0 x_2 + 2\lambda_1 x_2^3 = 0.$$

Если $\lambda_0 = 0$, то $\lambda_1 \neq 0$ (т.к. не все множители Лагранжа нули), и тогда $x_1 = x_2 = 0$, что не совместимо с ограничением. Значит, $\lambda_0 \neq 0$ и можно положить $\lambda_0 = 1$. Тогда либо $x_1 = 0, x_2 = \pm 1$, либо $x_1 = \pm 1, x_2 = 0$, либо $x_1^2 = x_2^2 = -1/(2\lambda_1)$,т. е. $|x_1| = |x_2| = 2^{-1/4}$.

В итоге первые две пары доставляют абсолютный минимум, равный единице, а четыре варианта третьей пары доставляют абсолютный максимум, равный $\sqrt{2}$. Из соображений однородности приходим к неравенствам

$$\left(\frac{x_1^2 + x_2^2}{2}\right)^{1/2} \leq \left(\frac{x_1^4 + x_2^4}{2}\right)^{1/4}, \quad (x_1^4 + x_2^4)^{1/4} \leq (x_1^2 + x_2^2)^{1/2}$$

Аналогично доказываются и другие неравенства о средних и средних степенных.

Упражнение 1. Доказать неравенство Коши–Буняковского

$$\left|\sum_{k=1}^{n} x_k y_k\right| \le \left(\sum_{k=1}^{n} x_k^2\right)^{1/2} \left(\sum_{k=1}^{n} y_k^2\right)^{1/2}.$$

Упражнение 2. Доказать неравенство Гельдера

$$\left|\sum_{k=1}^{n} x_k y_k\right| \le \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p} \left(\sum_{k=1}^{n} |x_k|^{p'}\right)^{1/p'}, \quad p \ge 1, 1/p + 1/p' = 1.$$

Упражнение 3. Доказать неравенство Минковского

$$\left(\sum_{k=1}^{n} |x_k + y_k|^p\right)^{1/p} \le \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p} + \left(\sum_{k=1}^{n} |y_k|^p\right)^{1/p} \quad p \ge 1.$$

А2. Экстремальные задачи геометрии.

1. Пусть $X = (x_i^j)_{i,j=1}^n$ произвольная квадратная матрица порядка n. Тогда имеет место *неравенство Адамара:*

$$(\det X)^2 \le \prod_{i=1}^{n} \quad (\sum_{j=1}^{n} (x_i^j)^2).$$

Здесь надо проименить правило множителей Лагранжа для задачи

$$\det(x_i^j)_{i,j=1}^n \to \sup, \quad |x_i|^2 = 1, \quad 1 \le i \le n.$$

2. Расстояние от точки до подпространства. Пусть $\{x^1, \ldots, x^k\}$ — линейно независимые элементы из $\mathbb{R}^n$. Тогда расстояние d от точки x^0 до подпространства, натянутого на $\{x^1, \ldots, x^k\}$ выражается формулой

$$d = d(x^0, \mathrm{span}\{x^1, ...x^k\}, R^n) = \left(\frac{(\det(\langle x^i, x^j\rangle)_{i,j=0}^k}{(\det(\langle x^i, x^j\rangle)_{i,j=1}^k}\right)^{1/2} \tag{1}$$

Здесь надо рассмотреть экстремальную задачу без ограничений

$$f(\xi) = |x^0 + \sum_{i=1}^{n} \xi_i x^i|^2 \to \inf$$

При этом: $f(\xi) = \langle A_k\xi, \xi\rangle + 2\langle a, \xi\rangle + b$, где $A_k = ((\langle x^i, x^j\rangle)_{i,j=1}^k$, $a = (\langle x^0, x^1\rangle, \ldots, \langle x^0, x^k\rangle), b = \langle 0x^0, x^0\rangle, \xi = (\xi_1, \ldots, \xi_k)$

Доказываем, используя компактность, существование решения $\widehat{\xi}$ и применяем теорему Ферма

$$f'(\widehat{\xi}) = 0$$

что равносильно (1).

Замечание. Определители $\det A_k$ и $\det A_{k-1}$ называются *определителями Грама.*

Упражнение 5.5. Вписать в единичный шар пространства $\mathbb{R}^n$ цилиндр наибольшего объема.

Упражнение 5.6. Вписать в единичный шар пространства $\mathbb{R}^n$ параллелипипед наибольшего объема.

Упражнение 5.7. Вписать в единичный шар пространства $\mathbb{R}^n$ конус наибольшего объема.