

Математический анализ

Дима Трушин

Семинар 1

Пределы

По видимому одна из самых важных тем в математическом анализе – это пределы. Она добротнo обросла непролазным формализмом, которым мучают студентов всех технических и математических вузов. И по этому поводу я хочу сказать несколько общих слов. Изначально пределы возникли как необходимый аппарат для определения производных, которые в свою очередь возникли в работах двух математиков Лейбница и Ньютона. Ньютон пытался решать физические задачи и производные у него служили для определения скоростей и ускорений. Лейбниц же больше тяготел к построению удобного мат аппарата, не вдаваясь в физический и сельскохозяйственный смысл построенных вещей.

Сегодня мы привыкли к страшным определениям пределов через ε и δ , написанным на языке машинного кода понятном любой символьной среде вычислений. Однако, если посмотреть на рассуждения Лейбница, то он рассуждал в терминах бесконечно больших и бесконечно малых величин. Позже Вейерштрасс, будучи недовольным такой нестрогостью, и придумал известный нам язык ε и δ для определения пределов строго. Он, можно сказать, поставил математический анализ на строгие рельсы, решив проблемы не строгости!?. Все бы хорошо, да только в 40-ые годы прошлого века Абрахам Робинсон продемонстрировал нестандартную модель вещественных чисел, в которой есть бесконечно малые и бесконечно большие величины. Кроме того, любая теорема, которая доказана в нестандартной модели, верна и в обычных вещественных числах и наоборот. Тем самым, Робинсон показал, что все рассуждения Лейбница были строгими, только он рассуждал в нестандартной модели. В 80-е годы в штатах даже проводили эксперимент и в высших учебных заведениях занятия по математическому анализу преподавали через бесконечно малые и бесконечно большие. Эксперимент был призван успешным и студенты осваивали анализ методами Лейбница ничуть не хуже, чем классическими методами Вейерштрасса.

Кроме того, чуть позже, когда появился функциональный анализ, люди поняли, что можно брать пределы даже от несходящихся, но ограниченных последовательностей. Правда, таких пределов бывает много разных. Методами теории моделей и алгебры можно описать вообще все подобные операции взятия пределов от любых последовательностей, правда с некоторыми неудобными сайд-эффектами. И вообще, математики постарались с придумыванием сложных и непонятных вещей за последние пару сотен лет.

В сухом остатке. Все изложенное выше говорит нам вот о чем. Есть такая процедура как предел. Она важна и нужна. Однако, эту процедуру можно определять очень по-разному так, что одно определение хуже другого. Я затрону привычный подход к пределам кратко, но минимизирую весь ненужный формализм. Кроме этого я постараюсь дать общий сеттинг для работы с пределами и производными в случае векторных и матричных величин.

Виды пределов

Какие бывают сеттинги для пределов? Классические, разбираемые в курсе математического анализа, – это пределы последовательностей и пределы функций. Давайте посмотрим на них. Что такое последовательность? Это счетный набор чисел $a_n \in \mathbb{R}$ при $n \in \mathbb{N}$. А что такое функция? Это отображение $f: U \rightarrow \mathbb{R}$, где $U \subseteq \mathbb{R}$. На самом деле, последовательность – это тоже отображение только вида $a: \mathbb{N} \rightarrow \mathbb{R}$. Потому изучать их можно с одинаковой точки зрения, просто у последовательности индекс n дискретный, а у функции индекс x непрерывный.

Для последовательностей определяется процедура $\lim_{n \rightarrow \infty}$, которая перерабатывает последовательность в число, называемое пределом. Для функций надо дополнительно фиксировать точку $x_0 \in U$ и определяется процедура $\lim_{x \rightarrow x_0}$, которая перерабатывает функцию в число, называемое пределом функции в точке x_0 . Такая разница связана с тем, что в множестве \mathbb{N} нет точек накопления кроме бесконечности, а в области

$U \subseteq \mathbb{R}$ вообще говоря может быть много точек накопления (возьмите интервал, вы можете накапливаться к любой его точке).

Геометрический смысл предела В начале я хочу помахать руками и до строгого определения дать интуицию происходящего. Пусть у нас есть некоторая последовательность $a_n \in \mathbb{R}$, где $n \in \mathbb{N}$, и пусть $a \in \mathbb{R}$ – некоторое число. Я хочу объяснить, что значит, что число a является пределом последовательности a_n . Идейно это значит, что с увеличением номера n число a_n становится все ближе и ближе к числу a . Например, если взять $a_n = \frac{1}{n}$, то с увеличением номера n число $\frac{1}{n}$ все ближе и ближе к 0. Или еще более дурацкий пример, если $a_n = 0$ для всех n , то число a_n все время совпадает с нулем (большей близости между $a = 0$ и a_n добиться сложно).

Теперь посмотрим на функции. Пусть у нас $f: U \rightarrow \mathbb{R}$ – некоторая функция в области $U \subseteq \mathbb{R}$ на прямой и $x_0 \in U$ – произвольная точка. Тогда можно определить понятие предела функции в точке $\lim_{x \rightarrow x_0} f(x)$. Неформально, число a будет пределом $f(x)$ в точке x_0 , если для всех точек x близких (но не равных) x_0 значения функции $f(x)$ будут близки к a .

Формальные определения Пусть $a_n \in \mathbb{R}$, где $n \in \mathbb{N}$, – некоторая последовательность. Тогда число $a \in \mathbb{R}$ называется ее пределом, если для любой окрестности $(a - \varepsilon, a + \varepsilon)$, существует номер $n_0 \in \mathbb{N}$, что все числа a_n при $n \geq n_0$ попадают в интервал $(a - \varepsilon, a + \varepsilon)$.

Пусть у нас $f: U \rightarrow \mathbb{R}$ – некоторая функция в области $U \subseteq \mathbb{R}$ на прямой и $x_0 \in U$ – произвольная точка. Число a называется пределом f в точке x_0 , если для любой окрестности $(a - \varepsilon, a + \varepsilon)$ точки a найдется окрестность $(x_0 - \delta, x_0 + \delta)$ точки x_0 такая, что проколота окрестность $(x_0 - \delta, x_0 + \delta) \setminus \{x_0\}$ отображается с помощью функции f внутрь окрестности $(a - \varepsilon, a + \varepsilon)$.¹

Общий сеттинг На пределы последовательностей и функций можно посмотреть с общей точки зрения такой, что это определение обобщается на случай произвольных отображений $f: V \rightarrow U$, где V и U – векторные пространства. Такой подход поможет нам понять как правильно работать с пределами, когда аргумент, например, матрица или вектор, или значение функции является таковым. Более того, можно обобщить понятие предела на случай $f: X \rightarrow Y$, где X и Y – произвольные множества. В этом случае надо зафиксировать некоторое семейство «окрестностей» в X и некоторое семейство «окрестностей» точки $y \in Y$. Тогда мы будем говорить, что y – это предел для f , если для любой «окрестности» U точки y , найдется «окрестность» V в X такая, что $f(V) \subseteq U$. При этом предел зависит от наборов «окрестностей» в X , выбора точки $y \in Y$ и ее набора «окрестностей». Меняя окрестности и точку, мы можем менять понятие предела. Если мы хотим определить «предел в точке», то обычно на X фиксируют точку $x \in X$ и семейство ее «проколотых окрестностей», а именно мы должны рассматривать окрестности вида $V \setminus \{x\}$. Это связано с тем, что мы не хотим учитывать поведение f в самой точке x , нас интересует только поведение в «близких» к x точках.

Теперь давайте разберемся с более конкретными примерами. Добавим формально новую точку к натуральным числам $\bar{\mathbb{N}} = \mathbb{N} \sqcup \{\infty\}$. И положим окрестностями бесконечности ∞ подмножества вида $\{n \in \mathbb{N} \mid n \geq M\} \sqcup \{\infty\}$ для некоторого $M \in \mathbb{N}$. Тогда проколотой окрестностью будет множество $\{n \in \mathbb{N} \mid n \geq M\}$. В этом случае определения предела даются в общей манере. Пусть X – это либо $\bar{\mathbb{N}}$ либо область $U \subseteq \mathbb{R}$ и $f: X \rightarrow \mathbb{R}$ – некоторое отображение. Пусть $x_0 \in X$ – точка для которой определены окрестности. Тогда число $a \in \mathbb{R}$ называется пределом f в точке x_0 , если для любой окрестности $(a - \varepsilon, a + \varepsilon)$ существует окрестность $E \subseteq X$ содержащая x_0 такая, что $f(E \setminus \{x_0\}) \subseteq (a - \varepsilon, a + \varepsilon)$.

Давайте подытожим сказанное выше. Мы видим, что можно изучать пределы для ситуаций, когда задано отображение $f: X \rightarrow Y$ и на множествах X и Y для всех точек заданы понятия окрестностей.² Существует общий подход в этом случае и имя ему – Топология. Однако я ограничусь простыми случаями: либо классические случаи для матанализа, либо X и Y являются векторными пространствами. В последнем случае можно справиться с помощью метрических пространств, а точнее их частного случая нормированных пространств. Но обо всем по порядку.

Пределы последовательностей

Выше я распинаялся про общие случаи и высокие материи. Давайте я теперь явно проговорю все детали для случая последовательностей. Пусть $a_n \in \mathbb{R}$, где $n \in \mathbb{N}$, – некоторая последовательность чисел.

¹Хочу отметить, что это определение обобщается на случай, когда точка $x_0 = \pm\infty$.

²На самом деле такой подход работает, когда вы определяете интегралы или любые другие конструкции в анализе требующие предельный переход.

Формальное определение Кроме определения предела в качестве числа, можно еще определить предел равным бесконечности (положительной, отрицательной или беззнаковой). Я повторю определение для числа $a \in \mathbb{R}$, чтобы все они были вместе.

- Пусть $a_n \in \mathbb{R}$, где $n \in \mathbb{N}$, – некоторая последовательность. Тогда число $a \in \mathbb{R}$ называется ее пределом, если для любого $\varepsilon > 0$, существует номер $n_0 \in \mathbb{N}$, что для всех $n \geq n_0$ выполнено $|a_n - a| < \varepsilon$.
Расшифровка определения следующая. Для любого интервала $(a - \varepsilon, a + \varepsilon)$ для числа a все члены последовательности a_n попадают в этот интервал начиная с некоторого номера n_0 .
- Это определение можно расширить на случай $a = +\infty$ или $a = -\infty$. А именно, мы будем говорить, что $+\infty$ является пределом a_n , если для любого $A \in \mathbb{R}$, существует $n \in \mathbb{N}$, что для всех $n \geq n_0$ выполнено $a_n > A$.
Расшифровка определения следующая. Для любого интервала $(A, +\infty)$ для $+\infty$ все члены последовательности a_n попадают в этот интервал начиная с некоторого номера n_0 .
- Аналогично дается определение для $a = -\infty$. А именно, мы будем говорить, что $-\infty$ является пределом a_n , если для любого $A \in \mathbb{R}$, существует $n \in \mathbb{N}$, что для всех $n \geq n_0$ выполнено $a_n < A$.
Расшифровка определения следующая. Для любого интервала $(-\infty, A)$ для $-\infty$ все члены последовательности a_n попадают в этот интервал начиная с некоторого номера n_0 .
- Мы будем говорить, что ∞ является пределом a_n , если для любого $A \in \mathbb{R}$, существует $n \in \mathbb{N}$, что для всех $n \geq n_0$ выполнено $|a_n| > A$.

Примеры поведения последовательностей Хочу обратить внимание, что не всякая последовательность сходится и потому не всякая имеет предел (даже $+\infty$, $-\infty$ или ∞). Я хочу продемонстрировать, какие ситуации бывают.

1. Последовательность a_n сходится.³ Например $a_n = 1 + \frac{1}{n}$ сходится к 1.
2. Последовательность a_n сходится в расширенном смысле.⁴ Например, $a_n = n$ сходится к ∞ , а $a_n = -n^2$ сходится к $-\infty$. Обратите внимание, что в литературе сходимости к ∞ или $-\infty$ часто тоже считается расходимостью, но специального вида.
3. Последовательность не сходится. Например $a_n = (-1)^n$. То есть по четным номерам a_n будет 1, а по нечетным номерам a_n будет -1 . В итоге нет одного числа, к которому накапливаются члены нашей последовательности. Но есть 2 числа вокруг которых вертятся члены нашей последовательности.
4. Последовательность не сходится, но при этом «крутится вокруг большого числа чисел». Например $a_n = \sin n$. Можно показать, что для любого числа $a \in [0, 1]$, можно найти подпоследовательность n_k такую, что $\sin n_k$ сходится к a . То есть эта последовательность все время находится близко ко всем точкам отрезка $[0, 1]$, постоянно возвращается к каждой точке, подходя к ней сколь угодно близко при любых больших номерах.⁵

Вычисления пределов

Самое главное для нас не определение предела, а вычисление пределов. Вычисления пределов устроены по следующей схеме. Для начала по определению люди доказывают руками, что какие-то последовательности сходятся к каким-то пределам. А после того, как у нас есть некоторый запас последовательностей, для которых мы знаем пределы, мы сводим вычисления любых других пределов к известным последовательностям с помощью свойств пределов.

³В классическом смысле сходимости имеется в виду к конечному пределу.

⁴Разрешены еще и бесконечные пределы.

⁵Это совершенно не очевидно и требует аккуратных доказательств.

Зверинец примеров Начнем с запаса примеров.

1. Есть куча примеров убегающих в бесконечность

(a) $p \in \mathbb{R}[x]$ – ненулевой многочлен. Тогда $\lim_{n \rightarrow \infty} p(n)$ будет бесконечностью, а знак совпадает со знаком старшего коэффициента.

(b) $n!$ убегает в $+\infty$.

(c) a^n при $|a| > 1$ убегает в бесконечность.

2. Если поделить на выражение бегущее в бесконечность (любую), то получим выражение идущее к нулю.

3. Есть еще популярный предел, по определению

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$$

Чуть более общо

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

4. Сравнение скоростей экспоненты и полинома

$$\lim_{n \rightarrow \infty} \frac{a_d n^d + \dots + a_1 n + a_0}{a^n} = 0$$

для любого $a > 1$ и любого многочлена.

5. Сравнение скорости логарифма и линейной

$$\lim_{n \rightarrow \infty} \frac{\ln n}{n} = 0$$

Свойства пределов

1. $\lim_{n \rightarrow \infty} a = a$.

2. $\lim_{n \rightarrow \infty} (a_n + b_n) = \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n$.

3. $\lim_{n \rightarrow \infty} (a_n b_n) = (\lim_{n \rightarrow \infty} a_n)(\lim_{n \rightarrow \infty} b_n)$.

4. $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{\lim_{n \rightarrow \infty} a_n}{\lim_{n \rightarrow \infty} b_n}$

5. Если $\lim_{n \rightarrow \infty} a_n = \infty$ (или $-\infty$), то $\lim_{n \rightarrow \infty} \frac{1}{a_n} = 0$.

6. Если $a_n \leq b_n \leq c_n$ и $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = b$, то $\lim_{n \rightarrow \infty} b_n = b$.

В свойствах (1)–(4) надо понимать, что если правая часть равенства существует, то и левая часть равенства существует и выполнено указанное равенство.

Обычно беда возникает, когда считается предел вида $\lim_{n \rightarrow \infty} \frac{a_n}{b_n}$ и при этом либо $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = 0$ либо $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = \pm\infty$. Вот тупой пример

$$\lim_{n \rightarrow \infty} \frac{n}{n^2} = \frac{\lim_{n \rightarrow \infty} n}{\lim_{n \rightarrow \infty} n^2} = \frac{\infty}{\infty} \quad \text{Беда!}$$

Тут надо пойти на ухищрения. Конкретно в этом случае мы видим, что не до сократили

$$\lim_{n \rightarrow \infty} \frac{n}{n^2} = \lim_{n \rightarrow \infty} \frac{1}{n} = 0$$

Но бывают ситуации и позапутаннее.

Примеры вычислений

1. Покажем, что $\lim_{n \rightarrow \infty} \frac{(-1)^n}{n} = 0$. Заметим, что

$$\frac{-1}{n} \leq \frac{(-1)^n}{n} \leq \frac{1}{n}$$

Но левая и правая части неравенства стремятся к нулю, значит и центральная часть идет к нулю.

2. Покажем, что $\lim_{n \rightarrow \infty} \frac{2n}{n^3+1} = 0$. Действительно,

$$\lim_{n \rightarrow \infty} \frac{2n}{n^3+1} = \lim_{n \rightarrow \infty} \frac{2}{n^2 + \frac{1}{n}} = \frac{2}{\lim_{n \rightarrow \infty} (n^2 + \frac{1}{n})} = \frac{2}{\infty} = 0$$

3. Покажем, что $\lim_{n \rightarrow \infty} \frac{\sqrt[3]{n^2} \sin(n!)}{n+1} = 0$. В начале заметим, что $-1 \leq \sin(n!) \leq 1$. Значит

$$-\frac{\sqrt[3]{n^2}}{n+1} \leq \frac{\sqrt[3]{n^2} \sin(n!)}{n+1} \leq \frac{\sqrt[3]{n^2}}{n+1}$$

Потому достаточно показать, что $\frac{\sqrt[3]{n^2}}{n+1}$ стремится к нулю. Теперь

$$\frac{\sqrt[3]{n^2}}{n+1} = \frac{n^{\frac{2}{3}}}{n} \cdot \frac{n}{n+1} = \frac{1}{n^{\frac{1}{3}}} \cdot \frac{1}{1 + \frac{1}{n}}$$

Теперь мы видим, что $\frac{1}{1 + \frac{1}{n}}$ стремится к 1, а $\frac{1}{n^{\frac{1}{3}}}$ стремится к 0.

4. Покажем, что $\lim_{n \rightarrow \infty} \frac{2^n}{n!} = 0$. Действительно,

$$0 \leq \frac{2^n}{n!} = \frac{2 \cdot 2 \dots 2}{1 \cdot 2 \dots n} = \frac{2}{1} \cdot \frac{2}{2} \cdot \frac{2}{3} \dots \frac{2}{n} \leq \frac{2}{1} \cdot \frac{2}{2} \cdot \frac{2}{3} \dots \frac{2}{3} = 2 \left(\frac{2}{3}\right)^{n-2}$$

Так как правая часть стремится к нулю, то и искомый предел тоже равен нулю.

Пределы функций

Пусть у нас $f: U \rightarrow \mathbb{R}$ – некоторая функция на открытой области $U \subseteq \mathbb{R}$ и $x_0 \in U$ – произвольная точка. Я уже определял понятие предела функции в точке. Про него можно еще думать так, когда вы подходите по любой последовательности к x_0 (которая при этом не попадает в x_0), то функция по этой последовательности подходит к числу a , являющемуся пределом.

Формальное определение Давайте я все же проговорю тут формальное определение. Пусть $U \subseteq \mathbb{R}$ – некоторая область в \mathbb{R} и $f: U \rightarrow \mathbb{R}$ – некоторая функция и фиксирована точка $x_0 \in U$. Тогда число $a \in \mathbb{R}$ называется пределом f в точке x_0 , если для любой окрестности $(a - \varepsilon, a + \varepsilon)$ существует окрестность $(x_0 - \delta, x_0 + \delta)$ содержащая x_0 такая, что $f((x_0 - \delta, x_0 + \delta) \setminus \{x_0\}) \subseteq (a - \varepsilon, a + \varepsilon)$. Или на языке ε - δ это звучит так. Для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что для всех x таких что $0 < |x - x_0| < \delta$ выполнено $|f(x) - f(x_0)| < \varepsilon$.

Приемы вычислений Как и в случае последовательностей для начала нам нужно запастись зверинцем примеров. В случае функций нам поможет понятие непрерывной функции. Функция $f: U \rightarrow \mathbb{R}$ называется непрерывной в точке $x_0 \in U$, если $\lim_{x \rightarrow x_0} f(x) = f(x_0)$. Непрерывные функции хороши тем, что их пределы легко считать, а именно, надо лишь подставить значение точки в функцию и все. Неформально их график изображается непрерывной линией. Вот список непрерывных функций:

- Любой многочлен $a_d x^d + \dots + a_1 x + a_0$ является непрерывным в любой точке $x_0 \in \mathbb{R}$.
- Рациональная функция вида $\frac{a_d x^d + \dots + a_1 x + a_0}{b_q x^q + \dots + b_1 x + b_0}$ является непрерывной в любой точке, кроме нулей знаменателя.
- Функции e^x , $\sin x$, $\cos x$ непрерывны на всей прямой.

- $\ln x$ непрерывен для всех положительных x .
- $\operatorname{tg} x$, $\operatorname{ctg} x$ непрерывны во всех точках, где они определены.
- Функции x^α непрерывны во всех точках, где они определены. При $\alpha \geq 0$ непрерывны при $x \geq 0$, а при $\alpha < 0$ непрерывны при $x > 0$.

Свойства у пределов функций такие же, как и у пределов последовательностей. Пределы можно брать у сумм, произведений, отношений и так далее. Все эти правила работают тогда, когда определены все элементы применяемых формул.

Пример Пределы легко считать через непрерывные функции.

$$\lim_{x \rightarrow \pi} \sqrt{\frac{e^{\sin(x)}}{1 + \cos(\ln x)}} = \sqrt{\lim_{x \rightarrow \pi} \frac{e^{\sin(x)}}{1 + \cos(\ln x)}} = \sqrt{\frac{\lim_{x \rightarrow \pi} e^{\sin(x)}}{\lim_{x \rightarrow \pi} (1 + \cos(\ln x))}} = \sqrt{\frac{e^{\lim_{x \rightarrow \pi} \sin(x)}}{1 + \cos(\lim_{x \rightarrow \pi} \ln x)}} = \frac{1}{\sqrt{1 + \cos(\ln(\pi))}}$$

Однако бывают и проблемы вида $\frac{0}{0}$ или $\frac{\infty}{\infty}$. Вот известный случай

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = \frac{\lim_{x \rightarrow 0} \sin x}{\lim_{x \rightarrow 0} x} = \frac{0}{0} \quad \text{Беда!}$$

Чтобы справиться с подобным случаем надо либо страдать, либо знать чуть более прогрессивные приемы, например, правило Лопиталья.

Правило Лопиталья Для этого правила надо знать понятие производной. Я считаю, что с этим то мы знакомы, потому нарушу логический формальный порядок изложения материала и объясню само правило до появления производных.

Правило Лопиталья можно применять для функций или для последовательностей, когда $a_n = f(n)$.⁶ Например, очень часто так бывает, что $a_n = f(n)$ для некоторой дифференцируемой функции и $b_n = g(n)$, где g — дифференцируемая функция и вы хотите посчитать предел для $\frac{a_n}{b_n}$. То вам достаточно посчитать $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)}$.

В общем случае, если функции f и g произвольные, возможны следующие варианты:

	$\lim f(x) = 0$	$\lim f(x) = a$	$\lim f(x) = \infty$
$\lim g(x) = 0$?	∞	∞
$\lim g(x) = b$	0	$\frac{a}{b}$	∞
$\lim g(x) = \infty$	0	0	?

Для разрешения указанных неопределенностей используется правило Лопиталья, а именно: пусть $f, g: \mathbb{R} \rightarrow \mathbb{R}$ — дифференцируемые функции такие, что либо $\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = 0$ или $\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = \infty$, тогда

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)} \quad \text{при условии } g'(x) \neq 0 \text{ в окрестности } x_0^7$$

В этом случае точка x_0 может быть равной ∞ или $-\infty$.

Правило Лопиталья объясняет почему $\frac{\sin x}{x}$ принимает значение 1 при $x = 0$. Действительно,

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \frac{\cos x}{1} = 1$$

Производные

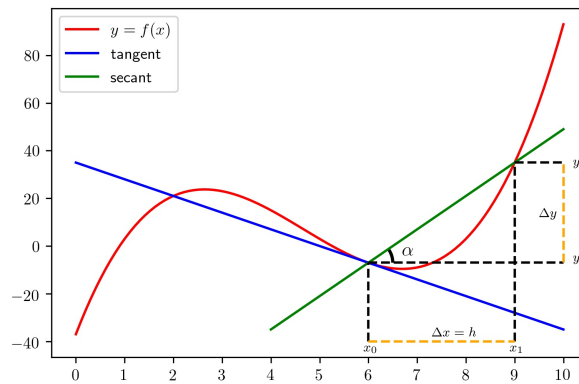
Понятие производной

Есть много всяких разных способов определять производную. Я начну с геометрического определения, которое объяснит зачем нужны производные. После уже обсудим как их вычислять и далее правило Лопиталья и приложение производных.

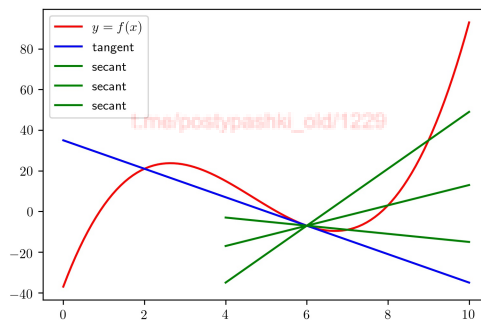
⁶Потому что в этом случае можно проверить предел для функций по всем числам, а значит будет верно и для последовательностей только по натуральным.

⁷Это условие никогда не является проблемой, ибо оно говорит, что функция $g(x)$ была константой, а в этом случае не было смысла применять правило Лопиталья.

Геометрический подход Давайте начнем с изучения картинки.



Пусть у нас есть некоторая функция $f: \mathbb{R} \rightarrow \mathbb{R}$ и некоторая точка $x_0 \in \mathbb{R}$. Обычно задача изучения производной начинается с построения касательной в точке x_0 . Как это делается. Надо отойти в сторону от точки x_0 , скажем в точку $x_0 + h$, где $h \in \mathbb{R}$ – вообще говоря любое число. В новой точке у нас будет значение $f(x_0 + h)$. Теперь у нас на плоскости есть две точки: $(x_0, f(x_0))$ и $(x_0 + h, f(x_0 + h))$ и мы можем провести через них прямую. Если мы будем выбирать число h все меньше и меньше, то прямая будет все ближе и ближе подходить к касательной.



Чтобы формально завершить данную процедуру нужно ввести числовую характеристику прямой, за которой мы будем следить и выполнить переход к пределу. Обычно смотрят на тангенс угла наклона прямой, а именно, давайте посмотрим на величину

$$\frac{\Delta y}{\Delta x} = \frac{f(x_0 + h) - f(x_0)}{h}$$

Это отношение длины противолежащего катета угла α с картинки к прилежащему катету, то есть – тангенс этого угла (по определению). Если мы устремим h к нулю, то наша секущая прямая превращается в касательную прямую и числовая характеристика $\frac{\Delta y}{\Delta x}$ стремится к числу, которое и называется производной f в точке x_0 и обозначается $f'(x_0)$. То есть формально

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}, \quad \text{если предел существует}$$

Из предыдущего описания должно быть понятно, что производная $f'(x_0)$ – это тангенс угла наклона касательной к графику $y = f(x)$ в точке x_0 . А как будет выглядеть уравнение касательной? Касательная – прямая, а потому ее уравнение будет иметь вид $y = k(x - x_0) + y_0$, где k – тангенс угла наклона, а (x_0, y_0) – какая-то одна точка на прямой. Глядя на это равенство, мы понимаем, что касательная к $y = f(x)$ в точке x_0 должна задаваться уравнением

$$y = f'(x_0)(x - x_0) + f(x_0)$$

Еще одно полезное замечание, напрямую вытекающее из геометрического подхода к производной – это описание поведения функции в зависимости от знака производной. Если $f'(x_0) > 0$, то функция точно растёт,

а если $f'(x_0) < 0$ то функция обязательно убывает в этой точке. То есть в этих точках понятно, что происходит. Потому интерес представляют обычно точки, где непонятно, что происходит, это точки x_0 такие, что $f'(x_0) = 0$. Такие точки называются критическими точками. Чуть ниже мы обсудим, что происходит в них.

Аналитический подход Давайте теперь геометрическую идею облачим в некоторые формулы. Пусть как и выше $f: \mathbb{R} \rightarrow \mathbb{R}$ – некоторая функция и x_0 – заданная точка. Давайте сравним уравнение касательной и уравнение функции, а именно, рассмотрим разность $\rho(x - x_0) = f(x) - f(x_0) - f'(x_0)(x - x_0)$. По определению мы видим, что $\rho(x - x_0) = o(x - x_0)$.⁸ Давайте перепишем это равенство по-другому и проанализируем полученное. Положим $h = x - x_0$ – расстояние, на которое мы отошли от точки x_0 , тогда

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \rho(h)$$

То есть если мы отойдем в сторону от точки x_0 на расстояние h (оно положительно если отходим вправо и отрицательно, если влево), то расстояние между графиками функции $f(x)$ и ее касательной в точке x_0 будет равно $\rho(h)$. Условие $\rho(h) = o(h)$ означает, что если отступ растет линейно, то расстояние будет медленнее, чем линейно. Это замечание очень полезно при малых h , а именно, когда $|h| < 1$. Давайте для простоты предположим, что $|\rho(h)| = |h|^2$. Тогда если мы отойдем вправо на расстояние $1/2$ от точки x_0 , то разница между графиком функции и касательной будет не больше $1/4$. Если мы отойдем на расстояние $1/10$, то разница между графиками будет уже не больше $1/100$ и т.д. То есть возле точки касания, график функции меняется медленнее, чем мы удаляемся от точки. В этом смысле касательная прямая является самой близкой прямой к данной функции в данной точке, то есть это лучшая замена функции на прямую возле данной точки.

Приближение квадратным многочленом Написанное выше выражение

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \rho(h)$$

можно понимать так: мы попытались заменить нашу функцию возле точки x_0 на линейную функцию самым лучшим образом (лучшесть здесь измеряется в терминах расстояния ρ , на сколько быстро оно падает). А что, если мы захотим заменить нашу функцию возле точки x_0 не на линейную функцию, а на многочлен степени 2 и тоже самым лучшим способом? Так тоже можно сделать. Оказывается, что полученное разложение будет иметь вид

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \rho(h), \text{ где } \rho(h) = o(h^2).$$

Давайте разберемся со всеми ингредиентами в формуле. Здесь $f: \mathbb{R} \rightarrow \mathbb{R}$ – некоторая функция. Теперь для каждого x_0 мы можем посчитать $f'(x_0)$, то есть мы можем определить еще одну функцию $f': \mathbb{R} \rightarrow \mathbb{R}$ по правилу $x \mapsto f'(x)$. А значит мы можем взять у нее производную в точке x_0 , а именно $f''(x_0)$ – это производная f' в точке x_0 . Чтобы получить еще лучшую интуицию, давайте думать про $f(x_0 + h)$ как про закон движения точки на прямой, где h – это время. Тогда $f(x_0)$ – это стартовая позиция, $f'(x_0)$ – это скорость движения, а $f''(x_0)$ – это ускорение.

Поведение в критических точках Предположим, мы хотим исследовать функцию $f: \mathbb{R} \rightarrow \mathbb{R}$ на наличие минимумов и максимумов. У минимумов и максимумов есть локальный и глобальный вариант. Точка $x_m \in \mathbb{R}$ называется глобальным максимумом, если значение в ней $f(x_m)$ больше, чем любое другое значение функции $f(x)$, то есть $f(x_m) \geq f(x)$ для любого $x \in \mathbb{R}$. Точка x_0 будет локальным максимумом, если значение в ней $f(x_0)$ будет самым большим не на всей прямой, а только возле этой точки. Формально, что найдется интервал $(x_0 - h, x_0 + h)$ вокруг точки x_0 , что $f(x_0) \geq f(x)$ для любой $x \in (x_0 - h, x_0 + h)$. Аналогично определяются точки глобального и локального минимума.

Глобальные минимумы и максимумы хороши тем, что мы бы их хотели найти, решая оптимизационные задачи. Однако, они очень плохо ищутся. Локальные минимумы и максимумы плохи тем, что они не оптимальны, но зато в отличие от глобальных их очень легко искать. А именно, если функция f в точке x_0 принимает минимум, то в $f'(x_0)$ обязана быть нулем, так как если она принимает положительное значение, то функция должна убывать слева от точки, а если $f'(x_0)$ отрицательная, то убывать справа от точки, что противоречило бы локальному минимуму. Оформим это в виде замечания.

⁸Здесь $o(x)$ обладает условием $o(x)/x \rightarrow 0$ при условии $x \rightarrow 0$.

Утверждение. Пусть $f: \mathbb{R} \rightarrow \mathbb{R}$ – некоторая функция и $x_0 \in \mathbb{R}$ – точка на прямой. Тогда если точка x_0 является точкой локального минимума или максимума, то $f'(x_0) = 0$, то есть x_0 является критической точкой для f .

Еще полезно думать про это условие следующим образом. Пусть x – это время, а $f(x)$ – это позиция точки на прямой. Тогда $f'(x)$ – это скорость в момент времени x . Если мы достигли локального максимума, то в перед ним мы двигались с положительной скоростью, а после него с отрицательной. Но в самом максимуме мы должны были остановиться, чтобы сменить знак у скорости. А условие остановки – это и есть условие равенства нулю производной.

Как вычислять производные Теперь я хочу потратить некоторое время на то, чтобы понять, как вычисляются производные. Первый вопрос: а как мы задаем функции. Вообще говоря, теоретически есть куча разных способов, но реально на практике мы используем формульные выражения для задания функций, например, $\sin x + e^{\lg x}$ или что-то в этом духе. Такие формулы обычно состоят из чисел, операций и имен функций. Потому, чтобы уметь дифференцировать подобные функции, нам надо знать как дифференцировать числа, функции, которые присутствуют в выражении и как дифференцировать результаты операций.

Для дифференцирования операций применяются следующие правила. Пусть $f, g: U \rightarrow \mathbb{R}$ – некоторые функции в области $U \subseteq \mathbb{R}$, тогда

1. $(f \pm g)' = f' \pm g'$
2. $(fg)' = f'g + fg'$
3. $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$
4. $f(g(x))' = f'(g(x))g'(x)$

А теперь несколько правил для дифференцирования знакомых нам функций:

1. $(\lambda)' = 0$ для любого $\lambda \in \mathbb{R}$
2. $x' = 1$
3. $(\ln x)' = \frac{1}{x}$
4. $(\sin x)' = \cos x$ и $(\cos x)' = -\sin x$
5. $(e^x)' = e^x$

Это базовый список функций, которых достаточно, чтобы продифференцировать любую функцию с учетом правил для дифференцирования операций. Действительно,

1. $(x^n)' = nx^{n-1}$, где $n \in \mathbb{Z}$, или более обще $(x^\alpha)' = \alpha x^{\alpha-1}$, где $\alpha \in \mathbb{R}$
2. $p(x) = a_0 + a_1x + \dots + a_nx^n$, тогда $p'(x) = a_1 + 2a_2x + \dots + na_nx^{n-1}$.
3. $(\operatorname{tg} x)' = \left(\frac{\sin x}{\cos x}\right)' = \frac{1}{\cos^2 x}$
4. $(a^x)' = (e^{\ln ax})' = \ln aa^x$

Все эти формулы и желание их комбинировать в минуту душевной невзгоды дают вам возможность посчитать производную от любой функции, то есть найти каким выражением будет задаваться производная, если вы знаете выражение для исходной функции. А геометрический смысл, мы уже обсудили, потому вы должны быть полностью вооружены и готовы к бою с функциями нескольких переменных.

Функции нескольких переменных

Частные производные Пусть теперь у нас есть функция от нескольких переменных $f: \mathbb{R}^n \rightarrow \mathbb{R}$, то есть правило по вектору $x = (x_1, \dots, x_n)$ ставящее число $f(x) = f(x_1, \dots, x_n)$. Пусть теперь у нас задана точка $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ и мы хотим исследовать ее в окрестности этой точки. Первое, что мы можем сделать, давайте зафиксируем все координаты равными координатам точки a , а одну координату сделаем переменной. Например

$$\varphi_i(t) = f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_n)$$

Тогда мы получим обычную функцию от переменной t . Давайте продифференцируем ее по t в точке a_i , результат называется частной производной f по переменной x_i в точке a , и обозначается все это вот так

$$\frac{\partial f}{\partial x_i}(a) = \varphi'_i(a_i)$$

В терминах предела определение пишется так:

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \rightarrow 0} \frac{f(a_1, \dots, a_{i-1}, a_i + h, a_{i+1}, \dots, a_n) - f(a_1, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_n)}{h}$$

Пусть теперь у нас есть вектор $e_i = (0, \dots, 0, 1, 0, \dots, 0)$, где единица стоит на i -ом месте. Тогда определение выше можно переписать так

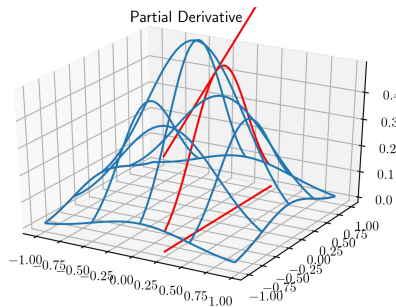
$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \rightarrow 0} \frac{f(a + he_i) - f(a)}{h}$$

Давайте поймем геометрический смысл проделанной работы на примере функции двух переменных $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ по правилу $(x, y) \mapsto f(x, y)$. Пусть $a = (a_1, a_2)$ – некоторая точка, мы зафиксируем a_2 , и рассмотрим $\varphi_1(t) = f(t, a_2)$. Заметим, что точки (t, a_2) лежат на прямой параллельной оси ОХ и пересекающей ось ОУ в точке a_2 . Тогда $\varphi_1(t)$ – это ограничение функции $f(x, y)$ на эту прямую. То есть мы хотим узнать, а как себя ведет данная функция вдоль прямой параллельной ОХ. Таким образом $\partial f / \partial x(a)$ будет показывать наклон касательной в точке a вдоль прямой параллельной ОХ и проходящей через a .

t.me/postypashki_old/1229

t.me/postypashki_old/1229

t.me/postypashki_old/1229



Правила вычисления частных производных Если вам надо посчитать частную производную $\frac{\partial f}{\partial x_i}(x)$ от некоторой функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$, то это делается ровно по тем же самым правилам, что и подсчет обычной производной. Тут надо думать так: вы смотрите на функцию $f(x_1, \dots, x_i, \dots, x_n)$ и считаете в ней все переменные кроме x_i какими-то числами и берете обычную производную по x_i как если бы вы дифференцировали по одной переменной.

Единственное полезное правило, которое стоит упомянуть в случае нескольких переменных – дифференцирование композиции. Предположим у нас есть несколько функций $f: \mathbb{R}^n \rightarrow \mathbb{R}$ по правилу $(y_1, \dots, y_n) \mapsto f(y_1, \dots, y_n)$ и $g_1, \dots, g_n: \mathbb{R}^m \rightarrow \mathbb{R}$ все по правилам $(x_1, \dots, x_m) \mapsto g_i(x_1, \dots, x_m)$. Я специально выбрал разные имена переменным для удобства. Теперь рассмотрим функцию

$$\varphi(x_1, \dots, x_m) = f(g_1(x_1, \dots, x_m), \dots, g_n(x_1, \dots, x_m))$$

Как надо считать ее частную производную по x_i ? Оказывается правило такое:

$$\frac{\partial \varphi}{\partial x_i}(x_1, \dots, x_m) = \sum_{k=1}^n \frac{\partial f}{\partial y_k}(g_1(x_1, \dots, x_m), \dots, g_n(x_1, \dots, x_m)) \frac{\partial g_k}{\partial x_i}(x_1, \dots, x_m)$$

Или кратко, это правило записывать так

$$\frac{\partial f(g_1, \dots, g_n)}{\partial x_i} = \sum_{k=1}^n \frac{\partial f}{\partial y_k} \frac{\partial g_k}{\partial x_i}$$

Вот пример, пусть у нас есть функция $f(y_1, y_2) = y_1^2 + \sin y_2$ и пусть $g_1(x) = \cos x$ и $g_2(x) = \ln x$. Тогда $\varphi(x) = f(g_1(x), g_2(x)) = (\cos x)^2 + \sin(\ln x)$. Мы хотим найти производную φ по x . Нам никто не мешает это сделать в лоб и получить

$$\varphi(x)' = -2 \cos x \sin x + \cos(\ln x) \frac{1}{x}$$

Однако, это можно сделать и по правилу дифференцирования композиции. Прежде чем считать, обратим внимание, так как функции g_1 и g_2 зависят только от одной переменной, то нет разницы между частной производной и обычной производной. А теперь посчитаем

$$\frac{\partial f}{\partial y_1} = 2y_1, \quad \frac{\partial f}{\partial y_2} = \cos y_2, \quad \frac{dg_1}{dx} = -\sin x, \quad \frac{dg_2}{dx} = \frac{1}{x}$$

Если подставить эти выражения в формулу для дифференцирования сложной функции, то получится

$$\varphi(x)' = \frac{\partial f}{\partial y_1}(g_1, g_2) \frac{dg_1}{dx} + \frac{\partial f}{\partial y_2}(g_1, g_2) \frac{dg_2}{dx} = (2 \cos x)(-\sin x) + (\cos \ln x) \left(\frac{1}{x} \right)$$

Обратите внимание, когда мы считаем частную производную $\frac{\partial f}{\partial y_1}$, то мы получаем выражение от y_1 и y_2 , после этого мы туда вместо y_1 подставляем $g_1(x)$, а вместо y_2 подставляем $g_2(x)$. Аналогично и для другой частной производной.

Производная по направлению Если как и выше мы возьмем функцию $f: \mathbb{R}^n \rightarrow \mathbb{R}$, некоторую точку $a \in \mathbb{R}^n$ и некоторый вектор $v \in \mathbb{R}^n$, то можно определить производную по направлению следующим образом

$$\nabla_v f(a) = \lim_{h \rightarrow 0} \frac{f(a + hv) - f(a)}{h}$$

Таким образом, частная производная – это лишь производная в направлении одного из векторов e_i . Смысл производной по направлению следующий. Если мы проведем через точку a прямую в направлении вектора v и ограничим на эту прямую функцию f , то получим функцию одной переменной $\varphi(t) = f(a + tv)$. Вот ее производная в нуле и будет производной по направлению. Это разумно, что мы хотим уметь брать производные в любом направлении, а не только вдоль координатных осей. Мало ли, что нам может пригодиться.

Градиент Как и выше у нас $f: \mathbb{R}^n \rightarrow \mathbb{R}$ – некоторая функция и $a \in \mathbb{R}^n$ – произвольная точка. Определим градиент функции f в точке a , как вектор, составленный из частных производных в этой точке:⁹

$$\text{grad}_a f = \left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right)$$

Если я теперь снабжу \mathbb{R}^n стандартным скалярным произведением $(v, u) = v^t u$, то вот есть какая полезная формула для вычисления производной по направлению. Пусть $v \in \mathbb{R}^n$ – произвольный вектор, тогда

$$\nabla_v f(a) = (\text{grad}_a f, v) = \frac{\partial f}{\partial x_1}(a) v_1 + \dots + \frac{\partial f}{\partial x_n}(a) v_n = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) v_i$$

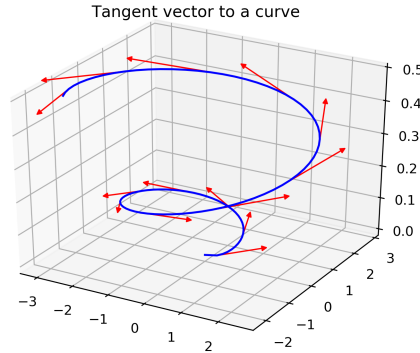
Вспомним, что геометрический смысл $\nabla_v f(a)$ – это скорость роста функции в направлении вектора v в точке a . С другой стороны

$$\nabla_v f(a) = (\text{grad}_a f, v) = |\text{grad}_a f| |v| \cos \angle(\text{grad}_a f, v)$$

Если я зафиксирую точку a и буду варьировать вектор v , то есть направление, в котором мы смотрим из точки a , то максимальное значение у выражения $\nabla_v f(a)$ будет достигнуто, когда вектор v сонаправлен градиенту. То есть градиент указывает в сторону наибольшего роста функции. Это самое важное свойство, которым обладает градиент. Он всегда говорит в какую сторону надо пойти, чтобы функция выросла. Однако, важно понимать, что градиент никогда не признается на сколько далеко надо пойти в указанном направлении. Он предпочитает оставить налет загадочности по этому вопросу.

⁹Для тех кому интересно, градиент надо считать вектор-строкой. Почему, ну потому что по-хорошему он является не вектором, а линейным отображением $\text{grad} f: \mathbb{R}^n \rightarrow \mathbb{R}$, но это уже отдельная песня. И мы на самом деле будем его трактовать то строкой, то столбцом в зависимости от удобства.

Касательный вектор Пусть у нас наоборот задана функция $g: \mathbb{R} \rightarrow \mathbb{R}^n$ по правилу $t \mapsto x = g(t)$. Про нее можно думать, как про траекторию движения точки в пространстве \mathbb{R}^n . Моя задача определить вектор скорости для этой кривой, он же будет касательный вектор, который будет смотреть вдоль кривой, а его длина будет говорить на сколько мы быстро движемся в этот момент.



Идея такая же, как и с касательным вектором. Пусть у нас есть момент времени t_0 , ему соответствует точка $x_0 = g(t_0)$. Возьмем следующий момент времени t_1 , в этот момент мы будем находиться в точке $x_1 = g(t_1)$. Тогда за время $t_1 - t_0$ мы переместились из x_0 в x_1 , а значит сдвинулись на вектор $x_1 - x_0$. Значит за этот момент времени наша средняя скорость равна отношению

$$v_{\text{средняя}} = \frac{x_1 - x_0}{t_1 - t_0} \in \mathbb{R}^n$$

Теперь надо время t_1 выбирать все ближе к времени t_0 и посмотреть предел средней скорости. Обратите внимание, что средняя скорость является вектором, потому, чтобы определить предел

$$v(t_0) = \lim_{t_1 \rightarrow t_0} \frac{x_1 - x_0}{t_1 - t_0}$$

нам надо понимать, что такое предел векторной последовательности. Тут можно поступить так: можно ввести норму на векторах и относительно нее определить предел, так как это делалось на прошлом занятии. Но я так же отмечал, что нормы можно выбирать разные, а понятие предела не изменится. Оказывается, можно выбрать норму так (например $\|x\|_\infty$), что сходимость будет просто покоординатной сходимостью. Последний предел так же обозначается производной по t , то есть скорость в момент времени t обозначается так $v(t) = g'(t)$.

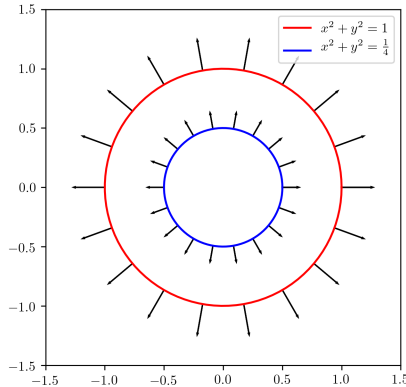
Давайте разберемся с примером. Пусть у нас есть кривая на плоскости $x(t) = (\cos t, \sin t)$. Тогда касательный вектор будет считаться, как производная по времени t . И оказывается эти операции можно сделать покоординатно, то есть $v(t) = ((\cos t)', (\sin t)') = (-\sin t, \cos t)$. То есть в точке $(\cos t, \sin t)$ касательный вектор будет $(-\sin t, \cos t)$. Аналогично все считается и в общем случае. Надо просто продифференцировать покоординатно вектор, задающий траекторию, и мы получим значение касательного вектора в данной точке. Его направление показывает куда в данный момент времени движется точка, а его длина показывает на сколько быстро мы движемся.

Поверхности и линии уровня Для функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и произвольной константы $c \in \mathbb{R}$ можно нарисовать в пространстве \mathbb{R}^n множество точек удовлетворяющих условию $f(x) = c$. Такую поверхность называют поверхностью уровня или изоповерхностью. На этой поверхности наша функция постоянна, а значит двигаясь по этой поверхности мы не меняем значение функции. В силу этого градиент оказывается всегда ортогонален поверхности уровня. Это можно понять так: можно запустить по изоповерхности произвольную кривую $x(t)$, тогда $f(x(t)) = c$ для любого момента времени t . Значит производная от этого выражения будет 0, так как константа, с другой стороны по правилам вычисления

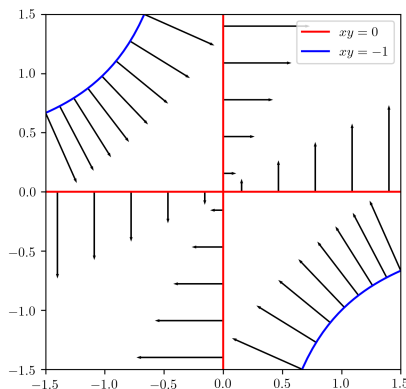
$$0 = f(x(t))' = \sum_{i=1}^n \frac{\partial f}{\partial x_i} x_i'(t) = (\text{grad}_{x(t)} f, x'(t))$$

Здесь $x'(t)$ – это касательный вектор к $x(t)$ в момент времени t , то есть в точке $x(t)$ и градиент берется в этой самой точке $x(t)$. То есть мы видим, что любой касательный вектор ортогонален градиенту для любого пути по поверхности уровня.

- Пусть нам задана функция $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ по правилу $(x, y) \mapsto x^2 + y^2$. Тогда ее линии уровня – это концентрические окружности с центром в начале координат. Ниже изображены две линии уровня и векторы градиента. Прошу обратить внимание, что для удобства векторы градиента нарочно укорочены в 4 раза.



- Пусть нам задана функция $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ по правилу $(x, y) \mapsto xy$. Тогда ее линии уровня – это либо крест из осей координат, либо гиперболы. Ниже изображены две линии уровня и векторы градиента. Прошу обратить внимание, что векторы градиента нарочно укорочены, причем с разными коэффициентами на разных линиях уровня.



Касательная гиперповерхность Как и в случае функции одного переменного, мы можем провести касательную гиперповерхность к графику функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ в произвольной точке $a \in \mathbb{R}^n$. Мы будем считать, что у нас на \mathbb{R}^n задано стандартное скалярное произведение. Теперь, если долго и пристально смотреть на случай функции одной переменной и определение градиента, то можно увидеть, что касательная плоскость будет задаваться в виде

$$y = (\text{grad}_a f, x - a) + f(a) = f(a) + \frac{\partial f}{\partial x_1}(a)(x_1 - a_1) + \dots + \frac{\partial f}{\partial x_n}(a)(x_n - a_n)$$

То есть если встать в точку $a \in \mathbb{R}^n$ и отойти на вектор $h \in \mathbb{R}^n$ в сторону от нее, то

$$f(a + h) = f(a) + (\text{grad}_a f, h) + \rho(h), \quad \text{где } |\rho(h)| \leq C|h|^2. \quad {}^{10}, {}^{11}$$

¹⁰Здесь я так же немного обманываю с оценкой справа для упрощения жизни.

¹¹Здесь $|h|$ означает длину вектора h относительно стандартного скалярного произведения.

Таким образом, градиент в точке a – это еще и вектор, который задает угловые коэффициенты касательной гиперповерхности в точке a для данной функции.

Исследование на минимумы и максимумы Как и в случае функций одной переменной мы можем сказать, что такое минимум или максимум для функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$, а так же, что такое локальный минимум и максимум. Если минимум и максимум определяются дословно, как точки, в которых значение функции самое маленькое (самое большое, соответственно), то для локальных минимумов и максимумов я должен уточнить в каком смысле понимается фраза «рядом с точкой».¹² Формально, точка $a \in \mathbb{R}^n$ называется локальным минимумом, если найдется шар радиуса r с центром в точке a , что для любой точки x из этого шара (то есть $|x - a| < r$) выполнено неравенство $f(a) \leq f(x)$. Аналогично дается определение локального максимума.

Скажем, что точка $a \in \mathbb{R}^n$ является критической для f , если $\text{grad}_a f = 0$, то есть все частные производные f в точке a равны нулю. Как и в случае одной переменной верно следующее полезное утверждение, помогающее найти точки локального минимума и максимума.

Утверждение. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}$ – некоторая функция, $a \in \mathbb{R}^n$ – точка. Тогда, если a является точкой локального минимума или максимума, то она является критической точкой для f , то есть $\text{grad}_a f = 0$.

Гессиан Как и в случае функций одного переменного это условие является необходимым, но не достаточным. Какие-то из критических точек могут оказаться не точками минимума или максимума. Чтобы лучше понять, как ведет себя график функции в окрестности точки, используют трюк аналогичный случаю одной переменной – смотрят на приближение второго порядка.

Пусть как и раньше $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Тогда для каждой точки $a \in \mathbb{R}^n$, мы можем определить $\partial f / \partial x_i(a)$. Таким образом получается новая функция $\partial f / \partial x_i: \mathbb{R}^n \rightarrow \mathbb{R}$ по правилу $x \mapsto \partial f / \partial x_i(x)$. А значит для этой новой функции тоже определены частные производные. В частности можно рассматривать частную производную по x_j от частной производной по x_i :

$$\frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i}(a)$$

Как вы видите получается безумно отвратительное выражение. Чтобы глаза не кровоточили лишний раз, подобное выражение сокращают так:

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(a) = \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i}(a)$$

Еще полезно понимать, что для хороших функций, а на практике это означает для любых, не важно в каком порядке брать частные производные – результат будет один и тот же, то есть

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a)$$

Теперь определим матрицу Гессе в точке $a \in \mathbb{R}^n$ для функции f , как матрицу составленную из частных производных:

$$H_a(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(a) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(a) \end{pmatrix}$$

Тогда Гессианом называется следующая квадратичная форма

$$Q_{a,f}(x) = x^t H_a(f) x$$

Оказывается, что хорошие послушные функции раскладываются следующим образом

$$f(a + h) = f(a) + (\text{grad}_a f, h) + \frac{1}{2} Q_{a,f}(h) + \rho(h), \text{ где } \rho(h) = o(|h|^2) \quad ^{13}$$

¹² Тут опять, можно понимать понятие рядом в смысле какой-нибудь нормы. Самая популярная норма – евклидова длина, мы и будем ее использовать. Но вообще говоря, можно взять любую норму, определение от этого не изменится.

¹³ Здесь $|h|$ – длина относительно стандартного скалярного произведения.

Обычно это равенство пишут в координатах в двух видах: в матричной форме и в координатной. Я напишу оба:

$$f(a+h) = f(a) + (\text{grad}_a f)^t h + \frac{1}{2} h^t H_a(f) h + \rho(h)$$

$$f(a+h) = f(a) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) h_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(a) h_i h_j + \rho(h)$$

Таким образом, если мы хотим проверить функцию на локальный минимум или максимум, то вначале надо найти точки $a \in \mathbb{R}^n$ обнуляющие градиент, то есть с условием $\text{grad}_a f = 0$. После этого, для каждой такой точки найти матрицу Гессе $H_a(f)$ и найти ее сигнатуру.¹⁴ Если она положительно определена, то точка a будет точкой строгого минимума, а если отрицательно определена, то точкой строгого максимума. Строгость в данном случае означает, что в случае минимума, она будет строго меньше всех точек рядом, а в случае максимума, она будет строго больше всех точек рядом. Если оказалось, что матрица Гессе имеет и отрицательный и положительный индекс инерции, то a точно не точка минимума или максимума. А вот если индексы инерции неположительные или неотрицательные, то тут могут быть разные чудеса. Но как мы помним, в практике чудес не бывает.

Градиентный спуск Давайте отметим популярное приложение градиента для решения оптимизационных задач. Пусть у нас дана функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и мы хотим найти ее точку локального минимума. Давайте в начале решим задачу чуть проще. Давайте найдем какую-нибудь критическую точку этой функции. Идея такая, возьмем любую случайную точку $a_1 \in \mathbb{R}^n$ и посчитаем градиент $\text{grad}_{a_1} f$, если он ноль, то мы уже нашли критическую точку, если нет, то мы знаем, что если мы чуть-чуть отступим в сторону $-\text{grad}_{a_1} f$, то значение функции должно уменьшиться. Выберем некоторое значение шага h_1 и шагнем в точку $a_2 = a_1 - h_1 \text{grad}_{a_1} f$. Если мы были аккуратны и выбрали h_1 достаточно малым, то получим $f(a_2) < f(a_1)$. Таким образом мы можем двигаться против градиента подбирая скорость и надеяться, что в конце концов придем к некоторой критической точке.

Теперь в этом идейном алгоритме есть несколько непонятных вещей: придем ли мы когда-нибудь в критическую точку? окажется ли она точкой локального минимума? как выбирать длины шагов h_i в этом алгоритме. Ситуация с этими вопросами следующая. Если у функции есть точка локального минимума и мы попали в хорошую стартовую точку, то мы обязательно придем к точке локального минимума. Если кроме локального минимума присутствуют локальные максимумы и седловые точки, то в численных вычислениях они не создают проблем. От максимума мы и так будем уходить, двигаясь против градиента, а седло является картинкой неустойчивой, то есть если чуть-чуть сбиться с траектории мы с большой вероятностью попадем в его убывающую зону и уйдем от седловой точки. Есть одна беда, мы можем о-о-очень долго топтаться рядом с седлом. Куда хуже точки с неотрицательной сигнатурой, там могут происходить всякие неприятные эффекты. Но с практической точки зрения надо понимать, что этот метод вас всегда приведет к точке локального минимума, если шаги подобраны так, что $f(a_n) < f(a_{n-1})$.

По поводу подбора длины шага. Тут у меня для вас плохие новости. Не существует какого-то единого супер способа подобрать длину шага. Потому существует куча разных методов градиентного спуска, все они отличаются именно способом того, как подбирается длина шага на каждом этапе. Самый простой способ подбора длины шага – метод наискорейшего градиентного спуска. Когда мы выбираем шаг h_i так, чтобы выражение $f(a_i - h_i \text{grad}_{a_i} f)$ было минимальным. Правда, бывает, что не всегда возможно решить явно эту задачу для данной функции f . Но и даже если мы ее решим и на каждом шаге будем находить оптимальное в этом смысле h_i , дела со скоростью стремления к минимуму у нас будут обстоять хреново. Со времен царя Гороха люди накопили огромное количество всяких дурацких примеров, когда наискорейший градиентный спуск ведет себя плохо. Важный вывод, который надо сделать – градиентный спуск это просто идейно и сложно технически, когда речь заходит о подборе длины шага. Данным вопросом занимается целая область математики.

Нормы в векторных пространствах

Пусть V – векторное пространство, тогда функция $|\cdot|: V \rightarrow \mathbb{R}$ называется нормой, если выполнены следующие свойства

¹⁴Сигнатуру можно проверять методом Якоби, симметричным Гауссом, а можно через поиск собственных значений матрицы Гессе. Данная задача как раз является той самой причиной, по которой нам важен метод определения сигнатуры.

1. $|v| \geq 0$ для любого $v \in V$, причем равенство достигается только на нулевом векторе.
2. $|\lambda v| = |\lambda| \cdot |v|$ для $v \in V$ и $\lambda \in \mathbb{R}$.
3. $|v + u| \leq |v| + |u|$.

Если в пространстве V задана норма, то можно определить расстояние между векторами по формуле $\phi(v, u) = |v - u|$.

Примеры норм

1. $x \in \mathbb{R}^n$, тогда

$$|x|_\infty = \max_i |x_i|$$

2. $x \in \mathbb{R}^n$, тогда

$$|x|_1 = \sum_i |x_i|$$

3. $x \in \mathbb{R}^n$, тогда

$$|x|_2 = \sqrt{\sum_i x_i^2}$$

Есть еще куча других норм, все они дают разные расстояния. Каждое из этих расстояний дает сходимость. А именно, последовательность $x_n \in \mathbb{R}^m$ сходится к $x \in \mathbb{R}^m$, если $\rho(x_n, x)$ стремится к нулю. И тут происходит очень любопытное чудо, оказывается, что все нормы в конечномерном пространстве эквивалентны. Это значит, что если $|\cdot|$ и $|\cdot|'$ – две нормы на \mathbb{R}^m , то найдутся такие положительные константы $a, b \in \mathbb{R}$, что

$$a|x| \leq |x|' \leq b|x|, \quad \text{для любого } x \in \mathbb{R}^m$$

А как несложно видеть из этого следует, что все нормы дают одну и ту же сходимость в том смысле, что если последовательность x_n сходится к x по расстоянию для одной нормы, то она сходится к x по расстоянию для любой другой нормы.

Производные в векторном пространстве

Пусть V и U – два нормированных векторных пространства, то есть на V и U заданы некоторые нормы и пусть $f: V \rightarrow U$ – некоторая функция. Мы хотим понять, что такое производная f в какой-то точке $v \in V$. Обычное определение для прямой не работает

$$\lim_{v \rightarrow v_0} \frac{f(v) - f(v_0)}{v - v_0} = f'(v_0)$$

потому что нельзя делить на векторы. Но вместо этого переделаем это определение в форму, в которой оно будет иметь смысла для случая векторного пространства. Тогда несуществующая функция ϕ :

$$\phi(v) = \frac{f(v) - f(v_0)}{v - v_0} - f'(v_0)$$

должна идти к нулю. Но теперь, формально умножив на $v - v_0$, можно переписать это так

$$f(v) = f(v_0) + f'(v_0)(v - v_0) + \phi(v)(v - v_0)$$

при этом $\phi(v) \rightarrow 0$ если $v \rightarrow v_0$.

Определение Теперь можно выдать из себя определение. Пусть $f: V \rightarrow U$ – функция между нормированными пространствами, $v_0 \in V$ – некоторый вектор и мы смогли записать

$$f(v) = f(v_0) + \varphi(v - v_0) + \alpha(v)$$

где $\varphi: V \rightarrow U$ – линейное отображение и $\alpha: V \rightarrow U$ – отображение такое, что $|\alpha(v)|/|v - v_0| \rightarrow 0$ при условии $v \rightarrow v_0$. Тогда отображение φ называется производной f в точке v_0 и обозначается $f'(v_0)$. В математике принят более пристойный термин – дифференциал f в точке v_0 , который обозначается через $d_{v_0}f$. Таким образом, мы можем переписать равенство выше в двух формах

$$f(v) = f(v_0) + f'(v_0)(v - v_0) + \alpha(v) = f(v_0) + d_{v_0}f(v - v_0) + \alpha(v)$$

Примеры

1. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и $\text{grad}_v f$ – градиент в точке v . Тогда дифференциал (или производная) в точке v_0 – это скалярное произведение с градиентом (относительно стандартного скалярного произведения). То есть

$$d_{v_0} f = f'(v_0): \mathbb{R}^n \rightarrow \mathbb{R} \\ x \rightarrow (\text{grad}_{v_0} f, x)$$

2. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $v_0 \in \mathbb{R}^n$ и $J = (\partial f / \partial x(v_0))$ – матрица якоби для f в точке v_0 . Тогда дифференциал в точке v_0 совпадает с линейным отображением заданным якобианом, то есть

$$d_{v_0} f = f'(v_0): \mathbb{R}^n \rightarrow \mathbb{R}^m \\ x \rightarrow \partial f / \partial x(v_0) x$$

Свойства дифференциала Для вычисления дифференциала используются правила аналогичные правилам для вычисления производных в одномерном случае:

1. Если $f, g: V \rightarrow U$ – две функции, $v \in V$, тогда $d_v(f + g) = d_v f + d_v g$.
2. Если $f: V \rightarrow U$ – две функции, $v \in V$, тогда для любого числа $\lambda \in \mathbb{R}$ имеем $d_v(\lambda f) = \lambda d_v f$.
3. Если $g: V \rightarrow U$ и $f: U \rightarrow W$ – две функции, $v \in V$, тогда $d_v(f \circ g) = d_{g(v)} f \circ d_v g$.

В случае функций со значениями в \mathbb{R} мы еще умеем их перемножать (еще и делить). В этом случае есть правило Лейбница (или его обобщение на дроби). В случае произвольного векторного пространства U операции умножения нет. Однако, в случае когда операцию умножения придумать можно, правило Лейбница тоже работает. Например в случае матричнозначных функций. Об этом поговорим дальше.

Матрицы от параметра Давайте рассмотрим матрицу A размера m на n такую, что все ее элементы – это функции от переменной t , то есть $A = (a_{ij}(t))$. Пусть производная функции a_{ij} по t обозначается как \dot{a}_{ij} . Тогда положим $\dot{A} = (\dot{a}_{ij})$ – матрица из всех производных по t . Такая операция удовлетворяет всем ожидаемым

t.me/postypashki_old/1229

t.me/postypashki_old/1229

t.me/postypashki_old/1229

1. $(A + B)^\cdot = \dot{A} + \dot{B}$ для $A, B \in M_{m \times n}(\mathbb{R})$.
2. $(\lambda A)^\cdot = \lambda \dot{A}$ для $A \in M_{m \times n}(\mathbb{R})$.
3. $(AB)^\cdot = \dot{A}B + A\dot{B}$ для $A \in M_{m \times k}(\mathbb{R})$ и $B \in M_{k \times n}(\mathbb{R})$.
4. $(A^n)^\cdot = \sum_{k=0}^{n-1} A^k \dot{A} A^{n-1-k}$ для $A \in M_m(\mathbb{R})$.
5. $(A^{-1})^\cdot = -A^{-1} \dot{A} A^{-1}$ для $A \in M_m(\mathbb{R})$.¹⁵
6. Пусть $A \in M_n(\mathbb{R})$ и пусть $A = (A_1 | \dots | A_n)$ – столбцы матрицы A . Тогда

$$\det(A)^\cdot = \sum_{k=1}^n \det(A_1 | \dots | \dot{A}_k | \dots | A_n)$$

Если разложить k -ое слагаемое по k -ому столбцу, то получим

$$\det(A_1 | \dots | \dot{A}_k | \dots | A_n) = \sum_{i=1}^n \dot{a}_{ik} A_{ik}$$

где A_{ik} – алгебраическое дополнение a_{ik} в матрице A . Тогда

$$\det(A)^\cdot = \sum_{k=1}^n \sum_{i=1}^n \dot{a}_{ik} A_{ik} = \sum_{k=1}^n \sum_{i=1}^n \dot{a}_{ik} (\hat{A})_{ki} = \sum_{i=1}^n (\hat{A} \dot{A})_{ii} = \text{tr}(\hat{A} \dot{A}) = \text{tr}(\dot{A} \hat{A})$$

Таким образом

$$\det(A)^\cdot = \text{tr}(\hat{A} \dot{A})$$

7. Для любой обратимой матрицы $A \in M_n(\mathbb{R})$ верно

$$(\ln \det A)^\cdot = \text{tr}(A^{-1} \dot{A})$$

¹⁵Здесь надо продифференцировать равенство $AA^{-1} = E$.

Вычисление дифференциалов от матричных аргументов Пусть $f: V \rightarrow U$ – функция между нормированными пространствами, $v_0 \in V$ – некоторый вектор и мы хотим найти $d_{v_0}f: V \rightarrow U$. В этом случае давайте посмотрим на выражение $f(v)$ и будем считать, что v – это вектор зависящий от параметра t . Тогда можно посчитать производную по t как было рассмотрено в предыдущем разделе и получим, что $(f(v))' = d_v f(\dot{v})$. Это дает практический способ вычисления производных в случае векторных и матричных аргументов. Давайте разберем серию примеров.

1. Пусть $f: M_n(\mathbb{R}) \rightarrow \mathbb{R}$ задана по правилу $X \mapsto \text{tr}(X^n)$ и мы хотим посчитать $d_A f: M_n(\mathbb{R}) \rightarrow \mathbb{R}$ в некоторой матрице A . Тогда рассмотрим

$$(\text{tr}(A^n))' = \text{tr}((A^n)') = \text{tr} \left(\sum_{k=0}^{n-1} A^k \dot{A} A^{n-1-k} \right) = \sum_{k=0}^{n-1} \text{tr} (A^k \dot{A} A^{n-1-k}) = \sum_{k=0}^{n-1} \text{tr} (A^{n-1} \dot{A}) = \text{tr}(n A^{n-1} \dot{A})$$

Значит $d_A f(X) = \text{tr}(n A^{n-1} X)$.

2. Если $f: M_n(\mathbb{R}) \rightarrow M_n(\mathbb{R})$ – это отображение по правилу $X \mapsto X^{-1}$, то $d_A f(X) = -A^{-1} X A^{-1}$.
3. Если $f: M_n(\mathbb{R}) \rightarrow \mathbb{R}$ – отображение заданное по правилу $X \mapsto \det X$, то $d_A f(X) = \text{tr}(\hat{A} X)$.

Вычисление градиента функции от матрицы Пусть теперь у нас есть функция $f: M_{m,n}(\mathbb{R}) \rightarrow \mathbb{R}$. Мы можем ее рассматривать как функцию $f(\dots, x_{ij}, \dots)$. Тогда для любой матрицы $A \in M_{m,n}(\mathbb{R})$ мы можем посчитать градиент $\text{grad}_A f = (\dots, \frac{\partial f}{\partial x_{ij}}(A), \dots)$. Часто вычислять эти частные производные в лоб неудобно. Удобнее воспользоваться матричными дифференцированиями. Для этого надо знать, как дифференциал связан с градиентом. Если мы для нашей функции $f(X)$ посчитали дифференциал в виде $d_A f: M_{m,n}(\mathbb{R}) \rightarrow \mathbb{R}$, то он обязательно представляется в виде $d_A f(X) = (S, X) = \text{tr}(S^t X)$,¹⁶ где $S \in M_{m,n}(\mathbb{R})$, тогда S – это и есть градиент! То есть $s_{ij} = \frac{\partial f}{\partial x_{ij}}(A)$.

Примеры Рассмотрим из предыдущих примеров случай функции.

1. Пусть $f: M_n(\mathbb{R}) \rightarrow \mathbb{R}$ задана по правилу $X \mapsto \text{tr}(X^n)$ и мы знаем, что отображение $d_A f: M_n(\mathbb{R}) \rightarrow \mathbb{R}$ в некоторой матрице A , задается по правилу $d_A f(X) = \text{tr}(n A^{n-1} X)$. Тогда мы видим, что $\text{grad}_A f = n(A^{n-1})^t$, то есть $\frac{\partial f}{\partial x_{ij}}(A) = n(A^{n-1})_{ji}$.
2. Если $f: M_n(\mathbb{R}) \rightarrow \mathbb{R}$ – отображение заданное по правилу $X \mapsto \det X$, то $d_A f(X) = \text{tr}(\hat{A} X)$. Тогда мы видим, что $\text{grad}_A f = (\hat{A})^t = A^+$, то есть $\frac{\partial f}{\partial x_{ij}}(A) = (A^+)_{ij} = A_{ij}$ – алгебраические дополнения в матрице A к ij -ому элементу.

Оптимизационная задача

Давайте рассмотрим следующую задачу: нам даны функции $f, g_1, \dots, g_k: \mathbb{R}^n \rightarrow \mathbb{R}$ и мы хотим найти минимум или максимум функции f при условии, что все остальные функции g_i равны нулю.¹⁷ То есть мы хотим решить следующую задачу

$$\begin{cases} f(x_1, \dots, x_n) \rightarrow \min \\ g_1(x_1, \dots, x_n) = 0 \\ \dots \\ g_k(x_1, \dots, x_n) = 0 \end{cases} \quad \text{или} \quad \begin{cases} f(x_1, \dots, x_n) \rightarrow \max \\ g_1(x_1, \dots, x_n) = 0 \\ \dots \\ g_k(x_1, \dots, x_n) = 0 \end{cases}$$

В качестве решения мы хотим предъявить точку $x \in \mathbb{R}^n$ такую, что $f(x) \leq f(y)$ (для второй задачи $f(x) \geq f(y)$) для любой $y \in \mathbb{R}^n$ такой, что $g_1(y) = \dots = g_k(y) = 0$. Стандартный рецепт решения такой задачи следующий:

1. Надо составить так называемый Лагранжиан

$$\mathcal{L} = \alpha f(x) - \lambda_1 g_1(x) - \dots - \lambda_k g_k(x)$$

где $\alpha, \lambda_1, \dots, \lambda_k \in \mathbb{R}$ – какие-то неизвестные константы.

¹⁶ Тут скобки означают стандартное скалярное произведение в матрицах, то есть $(A, B) = \text{tr}(A^t B)$.

¹⁷ Я для простоты предположил, что все функции определены в \mathbb{R}^n , но вообще говоря достаточно какой-то области $U \subseteq \mathbb{R}^n$. Технически задачи ничем не отличаются, однако, изложение становится проще и яснее.

- Далее говорится, что если точка $x \in \mathbb{R}^n$ является точкой минимума или максимума f при условии, что все g_i равны нулю, то найдутся такие константы $\alpha, \lambda_1, \dots, \lambda_k \in \mathbb{R}$, что точка x будет критической для \mathcal{L} .
- Это значит, что мы перебираем все возможные значения констант $\alpha, \lambda_1, \dots, \lambda_k$ и для них решаем уравнения:

$$\frac{\partial \mathcal{L}}{\partial x_1} = \dots = \frac{\partial \mathcal{L}}{\partial x_n} = 0$$

Самое удивительное во всем этом – оно работает! Но весь вопрос в том, а почему оно работает? Давайте сделаем пару замечаний, от которых у вас должны закрасться обоснованные сомнения. Например, что если мы рассмотрим случай $\alpha = 0$? Тогда Лагранжиан вообще не зависит от функции f и мы получаем какие-то странные условия на функции g_i . Особенно это звучит странно, если среди g_i есть повторяющиеся функции, мы получим очень странное условие, что все $x \in \mathbb{R}^n$ нам подойдут в качестве подозрительных точек. Давайте обсудим, что тут происходит, какой у всего этого геометрический смысл и как не запутаться в куче странных условий. Надеюсь, что после геометрического описания, станет понятно, почему эта штука вообще работает.

Наглядный частный случай

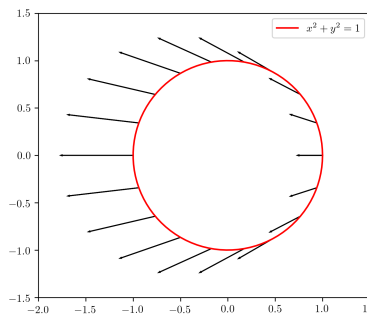
Движение по касательному градиенту Давайте для простоты объяснения рассмотрим случай всего двух переменных и одной вспомогательной функции, а именно

$$\begin{cases} f(x, y) \rightarrow \min \\ g(x, y) = 0 \end{cases} \quad \text{или} \quad \begin{cases} f(x, y) \rightarrow \max \\ g(x, y) = 0 \end{cases}$$

Более того, давайте рассмотрим конкретный пример

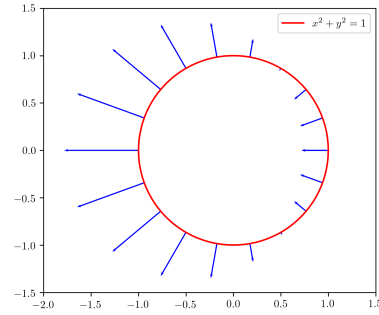
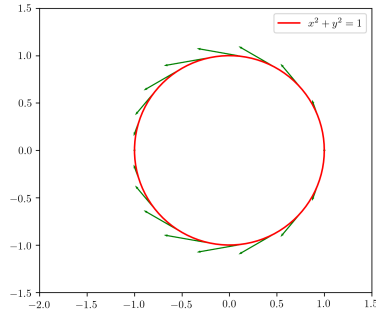
$$f(x, y) = (x - 2)^2 + y^2 \quad \text{и} \quad g(x, y) = x^2 + y^2 - 1$$

Тогда условие $g(x, y) = 0$ задает нам окружность радиуса 1 с центром в начале координат, а функция $f(x, y)$ имеет глобальный минимум в точке $(2, 0)$, которая не лежит на этой окружности. Давайте думать про нашу функцию $f(x, y)$ так: ее минимум – это источник жидкости, а жидкость из него вытекает и бежит вдоль градиента в сторону максимума (в реальной жизни жидкость бежит вниз к земле, а у нас она будет тянуться вверх). Давайте нарисуем поле градиента для функции $f(x, y)$, но не на всей плоскости а только на окружности $g(x, y) = 0$.¹⁸



Теперь думать надо так. Наша невесомая жидкость бежит из истока по градиентному полю на плоскости. А теперь мы поставим границы для нашей жидкости в виде равенства $g(x, y) = 0$ и разрешим бежать только по ней. В этом случае градиент $\text{grad } f$ будет иметь две составляющие: вдоль окружности (касательная компонента отмечена зеленым цветом) и перпендикулярно окружности (ортогональная компонента отмечена синим цветом). Вторая часть будет гаситься тем, что мы запрещаем жидкости покидать окружность, значит наша жидкость будет плыть только за счет касательной составляющей.

¹⁸Для читаемости на картинке изображено поле градиента для функции $\frac{1}{8}f(x, y)$, иначе векторы слишком длиннющие получаются.

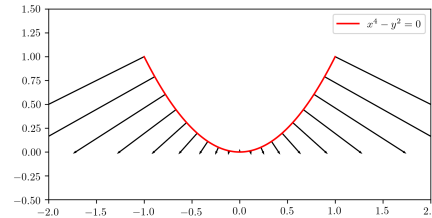
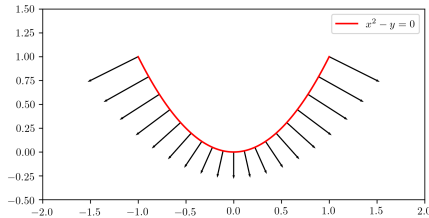


Как видно из рисунка справа, минимум и максимум достигаются там, где градиент функции f ортогонален окружности. Но с другой стороны, $\text{grad } g$ – это вектор, который тоже ортогонален окружности в каждой точке по его геометрическому смыслу (окружность – это линия уровня для g). А значит $f = \lambda \text{grad } g$ для некоторого числа λ .

Плохая параметризация Однако, может случиться неприятность. Что если мы плохо параметризовали нашу кривую и градиент g не всегда является ненулевым вектором? Давайте попробуем минимизировать функцию $f(x, y) = x^2 + (y + 1)^2$ на параболе заданной в первом случае с помощью $g_1(x, y) = x^2 - y$, а во втором случае с помощью $g_2(x, y) = x^4 - y^2$. То есть получаем две оптимизационные задачи

$$\begin{cases} f(x, y) \rightarrow \min \\ g_1(x, y) = 0 \end{cases} \quad \text{и} \quad \begin{cases} f(x, y) \rightarrow \min \\ g_2(x, y) = 0 \end{cases}$$

Заметим, что функция f , при ограничении на параболу, принимает минимум в начале координат. Но при этом $\text{grad } f$ не равен нулю в начале координат и смотрит вниз. Теперь давайте посмотрим, а что же происходит с градиентами $\text{grad } g_1$ и $\text{grad } g_2$. Они изображены на рисунках ниже:



В частности мы видим, что для g_1 градиент в точке $(0, 0)$ не равен нулю и мы можем любой вектор ортогональный параболу в этой точке представить в виде $\lambda \text{grad } g_1$. С другой стороны, если поглядеть на правый рисунок, то видно, что $\text{grad } g_2$ в точке $(0, 0)$ равен нулю. Потому мы никак не сможем представить $\text{grad } f$ в виде $\lambda \text{grad } g_2$ в начале координат.

Теперь давайте составим лагранжианы этих задач в наивном виде $\mathcal{L}_i = f - \lambda g_i$. Получим

$$\begin{cases} \mathcal{L}_1(x, y) = x^2 + (y + 1)^2 - \lambda(x^2 - y) \\ \frac{\partial \mathcal{L}}{\partial x}(x, y) = 2x - 2\lambda x = 0 \\ \frac{\partial \mathcal{L}}{\partial y}(x, y) = 2(y + 1) + \lambda = 0 \end{cases} \quad \text{и} \quad \begin{cases} \mathcal{L}_2(x, y) = x^2 + (y + 1)^2 - \lambda(x^4 - y^2) \\ \frac{\partial \mathcal{L}}{\partial x}(x, y) = 2x - 4\lambda x^3 = 0 \\ \frac{\partial \mathcal{L}}{\partial y}(x, y) = 2(y + 1) + 2\lambda y = 0 \end{cases}$$

Тогда мы видим, что для точки $(0, 0)$ можно подобрать $\lambda = -2$, чтобы она стала решением первой системы. С другой стороны, для второй системы при любом параметре λ точка $(0, 0)$ не является решением. Таким образом, когда Лагранжиан записывают в виде $\mathcal{L} = \alpha f - \lambda g$, это делается для того, чтобы исключить плохие

параметризации. Если $\alpha = 0$, $\text{grad } \mathcal{L} = 0$ превращается в условие $\text{grad } g = 0$ и это означает, что надо проверить точки, которые плохо запараметризованы с помощью g . А если $\alpha \neq 0$, то мы можем заменить его на 1 и рассматривать обычную функцию Лагранжа вида $\mathcal{L} = f - \lambda g$.

Интуиция для многомерного случая В общем виде, когда нам задана задача

$$\begin{cases} f(x) \rightarrow \min \\ g_1(x) = 0 \\ \dots \\ g_k(x) = 0 \end{cases} \quad \text{где } x \in \mathbb{R}^n$$

Условия $g_1(x) = \dots = g_k(x) = 0$ задают некоторую «поверхность». Если в точке x , удовлетворяющей этой системе, градиенты $\text{grad } g_1, \dots, \text{grad } g_k$ линейно независимы, то оказывается, в окрестности точки x эти уравнения задают поверхность размерности $n-k$ (гладко изогнутое подпространство размерности $n-k$). Последнее утверждение – это одна из форм теоремы о неявной функции. Потому, когда мы задаем Лагранжиан в виде $\mathcal{L} = \alpha f - \sum_{i=1}^k \lambda_i g_i$, мы подразумеваем, что у нас есть два случая:

- $\alpha = 0$. Тогда условие для критической точки функции Лагранжа $\text{grad } \mathcal{L} = 0$ означает, что для некоторых λ_i имеет место равенство $\lambda_1 \text{grad } g_1 + \dots + \lambda_k \text{grad } g_k = 0$. То есть это случай линейно зависимых градиентов, случай плохой параметризации.
- $\alpha \neq 0$, а значит можно считать $\alpha = 1$. В этом случае можно рассматривать задачу с функцией $\mathcal{L} = f - \lambda_1 g_1 - \dots - \lambda_k g_k$. Тогда условие для критической точки функции Лагранжа $\text{grad } \mathcal{L} = 0$ означает, что для некоторых λ_i имеет место равенство $\text{grad } f = \lambda_1 \text{grad } g_1 + \dots + \lambda_k \text{grad } g_k$. То есть это случай, когда $\text{grad } f$ ортогонален нашей поверхности. Если думать в терминах потока, то в этой точке поток не движется вдоль поверхности.

Полный набор уравнений для наглядности Пусть мы решаем задачу

$$\begin{cases} f(x) \rightarrow \min \\ g_1(x) = 0 \\ \dots \\ g_k(x) = 0 \end{cases} \quad \text{где } x \in \mathbb{R}^n$$

Тогда составим функцию $\mathcal{L} = \alpha f - \lambda_1 g_1 - \dots - \lambda_k g_k$. В этом случае точку минимума надо искать среди решений следующей системы

$$\begin{cases} \text{grad } \mathcal{L}(x) = 0 \\ g_1(x) = 0 \\ \dots \\ g_k(x) = 0 \end{cases} \quad \text{где } x \in \mathbb{R}^n$$

При этом мы перебираем все возможные значения параметров $\alpha, \lambda_1, \dots, \lambda_k$. Как говорилось выше, для параметра α можно считать, что его значение либо 0 либо 1.

Оптимизационная задача с неравенствами

Давайте рассмотрим следующую задачу: нам даны функции $f, g_1, \dots, g_k, h_1, \dots, h_s: \mathbb{R}^n \rightarrow \mathbb{R}$ и мы хотим найти минимум или максимум функции f при условии, что функции g_i равны нулю и выполнены неравенства $h_i \leq 0$. То есть мы хотим решить следующую задачу

$$\begin{cases} f(x_1, \dots, x_n) \rightarrow \min \\ g_1(x_1, \dots, x_n) = 0 \\ \dots \\ g_k(x_1, \dots, x_n) = 0 \\ h_1(x_1, \dots, x_n) \leq 0 \\ \dots \\ h_s(x_1, \dots, x_n) \leq 0 \end{cases} \quad \text{или} \quad \begin{cases} f(x_1, \dots, x_n) \rightarrow \max \\ g_1(x_1, \dots, x_n) = 0 \\ \dots \\ g_k(x_1, \dots, x_n) = 0 \\ h_1(x_1, \dots, x_n) \leq 0 \\ \dots \\ h_s(x_1, \dots, x_n) \leq 0 \end{cases}$$

В качестве решения мы хотим предъявить точку $x \in \mathbb{R}^n$ такую, что $f(x) \leq f(y)$ (для второй задачи $f(x) \geq f(y)$) для любой $y \in \mathbb{R}^n$ такой, что $g_1(y) = \dots = g_k(y) = 0$ и $h_1(y) = \dots = h_k(y) \leq 0$. Стандартный рецепт решения такой задачи следующий:

1. Надо составить так называемый Лагранжиан

$$\mathcal{L} = \alpha f(x) - \lambda_1 g_1(x) - \dots - \lambda_k g_k(x) - \mu_1 h_1(x) - \dots - \mu_s h_s(x)$$

где $\alpha, \lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_s \in \mathbb{R}$ – какие-то неизвестные константы. При этом для поиска максимума мы считаем, что $\alpha \geq 0$ и $\mu_i \geq 0$, а для поиска минимума мы считаем, что $\alpha \leq 0$ и $\mu_i \leq 0$.

2. Далее говорится, что если точка $x \in \mathbb{R}^n$ является точкой минимума или максимума f при условии, что все g_i равны нулю и все $h_i \leq 0$, то найдутся такие константы $\alpha, \lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_s \in \mathbb{R}$, что точка x будет удовлетворять системе (слева для минимума, а справа для максимума)

$$\left\{ \begin{array}{l} \frac{\partial \mathcal{L}}{\partial x}(x) = 0 \\ g_i(x) = 0 \\ \alpha \geq 0 \\ \mu_i h_i(x) = 0 \\ h_i(x) \leq 0 \\ \mu_i \leq 0 \end{array} \right. \quad \text{или} \quad \left\{ \begin{array}{l} \frac{\partial \mathcal{L}}{\partial x}(x) = 0 \\ g_i(x) = 0 \\ \alpha \leq 0 \\ \mu_i h_i(x) = 0 \\ h_i(x) \geq 0 \\ \mu_i \geq 0 \end{array} \right.$$

В этой задаче еще менее очевидно, почему подобное барахло хоть как-то работает. Особенно получилась куча каких-то странных неравенств на константы, которых не было в предыдущем случае, когда были только уравнения и не было неравенств в системе. Как и раньше, я предлагаю разобрать механизм работы этого метода на конкретном примере малой размерности.

Наглядный пример с неравенствами

Давайте для простоты объяснения рассмотрим случай всего двух переменных и одной вспомогательной функции с неравенством, а именно

$$\left\{ \begin{array}{l} f(x, y) \rightarrow \min \\ h(x, y) \leq 0 \end{array} \right. \quad \text{или} \quad \left\{ \begin{array}{l} f(x, y) \rightarrow \max \\ h(x, y) \leq 0 \end{array} \right.$$

Более того, давайте рассмотрим конкретный пример

$$f(x, y) = (x - 2)^2 + y^2 \quad \text{и} \quad h(x, y) = x^2 + y^2 - 1$$

В этом случае $h(x, y) \leq 0$ задает круг радиуса 1 с центром в нуле и теперь минимум и максимум функции мы ищем не только на окружности, но и внутри круга. По хорошему, это означает, что нам надо решить две разные задачи: 1) найти минимум внутри открытого круга, 2) найти минимум на окружности. После чего надо сравнить эти два минимума и выбрать меньший. Аналогично надо поступить в случае максимума.

Давайте начнем со случая поиска минимума (или максимума) внутри круга. Тогда мы решаем задачу

$$\left\{ \begin{array}{l} f(x, y) \rightarrow \min \\ h(x, y) < 0 \end{array} \right. \quad \text{или} \quad \left\{ \begin{array}{l} f(x, y) \rightarrow \max \\ h(x, y) < 0 \end{array} \right.$$

Так как условие $h(x, y) < 0$ задает открытое подмножество, то необходимое условие для минимума или максимума – градиент f равен нулю, то есть

$$\left\{ \begin{array}{l} \frac{\partial f}{\partial x}(x, y) = 0 \\ \frac{\partial f}{\partial y}(x, y) = 0 \\ h(x, y) < 0 \end{array} \right.$$

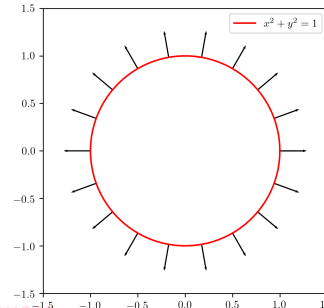
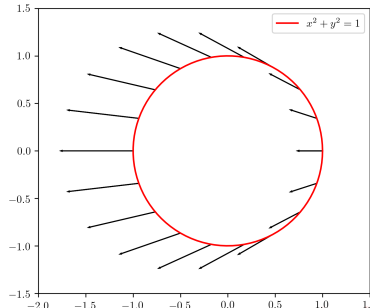
Теперь, если мы ищем минимум или максимум на границе, то задача превращается в предыдущую и мы решаем

$$\begin{cases} \frac{\partial f}{\partial x}(x, y) = \mu \frac{\partial h}{\partial x}(x, y) \\ \frac{\partial f}{\partial y}(x, y) = \mu \frac{\partial h}{\partial y}(x, y) \\ h(x, y) = 0 \end{cases}$$

Я здесь не пишу α , потому что мы уже знаем, что градиент на окружности не зануляется.

Теперь давайте обсудим странные неравенства и равенства, которые возникли в общем случае, на этом примере. Первое неравенство было $\alpha \geq 0$. Это условие нужно, чтобы критические точки функций f и αf не сменили своего характера. Точнее, при $\alpha > 0$ точка является минимум для одной из них тогда и только тогда, когда она минимум для другой. А при $\alpha = 0$ у нас от функции в лагранжиане ничего не зависит и этот случай нужен лишь для случая плохой параметризации.

Причина, по которой мы хотим иметь одинаковый характер поведения в критических точках для f и αf очень простой – мы можем распознать минимумы и максимумы в этом случае. Давайте посмотрим на градиенты f и h



Где будет максимум у функции f на окружности? Там где жидкость пытается утечь наружу. А минимум? Там где жидкость пытается затечь внутрь окружности. То есть в точке максимума на границе $\text{grad } f$ и $\text{grad } h$ смотрят в одну сторону, а в точке минимума $\text{grad } f$ и $\text{grad } h$ смотрят в разные стороны. Потому для минимума мы требуем $\mu \leq 0$, а для максимума $\mu \geq 0$. Случай равенства нулю нужен если эта точка оказалась стационарной для f (например точка была точкой максимума или минимума без ограничения $h = 0$).

Осталось понять зачем нужно условие $\mu h(x, y) = 0$. Но в этом случае все просто. В общем случае, когда у нас много неравенств, нам надо разобрать для каждого уравнения два случая $h_i < 0$ или $h_i = 0$. А как записать эти два условия одновременно?¹⁹ Условие $\mu h(x, y) = 0$ гарантирует, что либо $\mu = 0$ и значит условие минимизации превращается в $\text{grad } f = 0$ (то есть мы ищем точку минимума без ограничений). А если $\mu \neq 0$, то условие $\mu h(x, y) = 0$ означает, $h(x, y) = 0$ и значит мы вышли на границу и условие минимизации означает $\text{grad } f = \mu \text{grad } h$. И для того, чтобы в первом случае $\mu = 0$ мы учли неравенство $h(x, y) < 0$ в систему все равно добавляем уравнение $h(x, y) \leq 0$. Лишнее равенство нулю тут ничему не мешает формально, мы лишь учтем минимум на границе дважды, но ничего не потеряем.

¹⁹ Не то чтобы это поможет при решении. Так как решая составленное уравнение, вы все равно будете перебирать случаи. Это вопрос лишь краткой записи и так понятных условий.

Семинар 1

Задачи:

1. Найти предел

(a) $\lim_{x \rightarrow a} \left(\frac{\sin x}{\sin a} \right)^{\frac{1}{x-a}}$

(b) $\lim_{x \rightarrow 0} \frac{\ln(1+x^2)}{\sin(\cos x - 1)}$

2. Даны два семейства кривых на плоскости

$$y^2 = 4a(a - x) \quad (a > 0) \quad \text{и} \quad y^2 = 4b(b + x) \quad (b > 0)$$

Покажите, что в любой точке пересечения любой кривой из первого семейства с любой кривой из второго семейства касательные этих кривых перпендикулярны.

3. Исследовать на экстремумы (то есть на локальные минимумы и максимумы)

$$z = xy \ln(x^2 + y^2) \quad (x, y) \neq 0$$

4. Пусть $x_1, \dots, x_d \in \mathbb{R}^n$ – набор векторов и $U \subseteq M_n(\mathbb{R})$ – множество матриц с положительным определителем. Рассмотрим функцию

$$\varphi: U \rightarrow \mathbb{R} \quad \text{по правилу} \quad A \mapsto \frac{1}{d} \sum_{i=1}^d (Ax_i, x_i) - \ln \det A$$

где $(x, y) = x^t y$ – стандартное скалярное произведение. Найдите градиент φ в точке A .

5. Найти методом функции Лагранжа минимум и максимум функции $z = x^2 + y^2 - 12x + 16y$ при условии $x^2 + y^2 \leq 25$.

Математический анализ

Дима Трушин

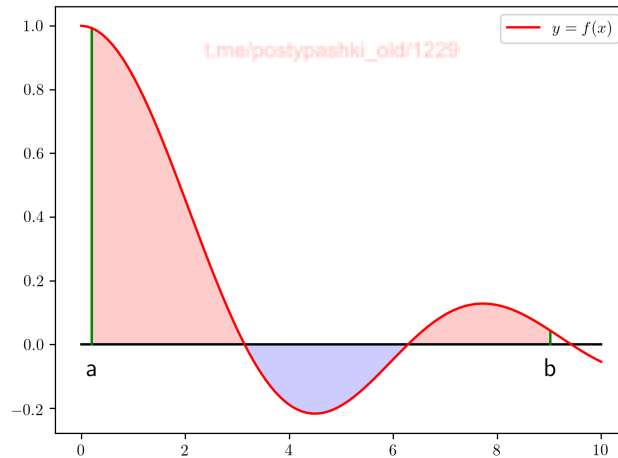
Семинар 2

Интеграл

Теперь давайте разберемся, что такое интеграл. Как обычно мы преследуем три цели:

- Понять сельскохозяйственный смысл интеграла.
- Освоить техническую составляющую о том, как вычислять интеграл.
- Посмотреть какие-то применения интегрирования.

Что такое интеграл Пусть у нас задана функция $f: \mathbb{R} \rightarrow \mathbb{R}$ и отрезок $[a, b] \subseteq \mathbb{R}$. И мы хотим найти площадь под графиком функции на отрезке $[a, b]$, причем если график лежит выше оси OX , то площадь будем считать положительной, а если ниже, то отрицательной. Например, ниже изображен график некоторой функции f . Площади, которые мы считаем положительными изображены красным цветом, а отрицательные синим.



Интеграл функции f на отрезке $[a, b]$ обозначается следующим образом

$$\int_a^b f(x) dx$$

Я выше взял функцию, которая определена на всей прямой, но это на самом деле не обязательно. Функция может быть определена в меньшей части, как, например, логарифм. Главное, чтобы отрезок $[a, b]$ попал в область определения функции. Еще можно рассматривать случаи $a = -\infty$ или $b = +\infty$. То есть можно считать интегралы вида

$$\int_{-\infty}^b f(x) dx \quad \int_a^{+\infty} f(x) dx \quad \int_{-\infty}^{+\infty} f(x) dx$$

Как бы вы ни определяли интеграл, вы всегда хотите выполнение следующих свойств

$$\int_{-\infty}^b f(x) dx = \lim_{a \rightarrow -\infty} \int_a^b f(x) dx \quad \int_a^{+\infty} f(x) dx = \lim_{b \rightarrow +\infty} \int_a^b f(x) dx \quad \int_{-\infty}^{+\infty} f(x) dx = \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_a^b f(x) dx$$

Кроме того, про интеграл можно думать вот как: если мы идем по отрезку $[a, b]$ слева на право (то есть $a < b$) то мы с плюсом берем площади выше оси OX , а с минусом ниже. А если мы идем по отрезку $[b, a]$ (опять же $a < b$), то мы наоборот, будем считать площади выше оси с минусом, а ниже с плюсом. Таким образом мы можем учитывать направления интегрирования, что дает формулу

$$\int_a^b f(x) dx = - \int_b^a f(x) dx$$

Таким образом интеграл учитывает не только знак функции, но и направления движения по прямой. Вычисления интеграла – это вычисление ориентированной площади.

Еще очень полезно понимать, что интеграл не всегда существует. Если функция $f(x)$ очень поганая, то вообще говоря, для нее может быть не определено понятие $\int_a^b f(x) dx$. Но это зависит от вида интеграла, который вы используете. Существует много разных определений интеграла и где не срабатывает одно, помогает другое. А есть ситуации, когда интеграл не существует по объективным причинам, а не из-за дефекта конструкции. Например, $\int_0^1 1/x dx = +\infty$ независимо от конструкции. То есть площадь под графиком этой функции бесконечна на отрезке $[0, 1]$.

Формальная сторона На самом деле, читая всякие умные книжки по разным видам анализов, вы ничего нового про интеграл не узнаете. Это все та же площадь под графиком функции. Почему же тогда существует так много разных видов интеграла? Вас наверняка уже пугали: интеграл Римана, интеграл Лебега, интеграл Римана-Стилтьеса, интеграл Макшейна, интеграл Курцвейла-Хинстока и т.д. Все это связано с тем, что математиков не устраивает слово «площадь» в формулировке выше. Надо строго сказать как ее считать. И вот для строгого определения площади есть вагон и маленькая тележка различных конструкций, которые на разные лады считают одну и ту же площадь. Обычно разница лишь в запасе функций, которые поддаются той или иной конструкции.

Сложно выделить самую лучшую конструкцию. Обычно одна хороша в одних обстоятельствах другая в других. А в случае приложений, так нам вообще плевать, какая из этих конструкций используется. Мы просто считаем площади по формулам и точно знаем, что полученные числа будут площадями. Но можно выделить две самые популярные: интеграл Римана и интеграл Лебега. Интеграл Римана пытается приблизить нашу функцию ступенчатой ломаной, считает площадь под ломаной, а потом переходит в некотором смысле к пределу, чтобы получить площадь под кривой. Думаю, что больше вам знать по этому вопросу не положено. А даже если и положено, то пользы оно вам не принесет. Прочитав 100500 страниц текста, вы лишь убедитесь, что знаете как считать площадь со знаком и все.

У моего решения не выбирать конкретный интеграл есть свои последствия. Я нигде не говорю точные условия существования интегралов ибо они все зависят от конкретного вида интеграла. На это можно смотреть как на полуобман. Ибо все равно никаких плохих функций в вашей жизни не встретится, а так хоть интегрироваться научитесь. В любом случае, если требовать, чтобы все функции встречаемые под знаком интеграла были непрерывны, то этого хватит всегда и для любых теорем, но оговаривать это я не буду ни в каком виде.

Вычисление интегралов

Тонкости вычисления интегралов можно разбить на несколько частей:

- Как считать интегралы у конкретных функций.
- Какие свойства есть у интегралов.
- Какие трюки в работе с интегралами можно использовать: различные оценки, связь с производной и т.д.

По поводу интегрирования и дифференцирования есть одна мудрость: дифференцировать можно научить и кролика, если за каждую правильно взятую производную ему давать морковку, а интегрировать – это уже искусство. Процесс дифференцирования полностью описывался алгоритмами к действию. Что бы вам ни скормили, все можно продифференцировать по правилам. А вот с интегрированием дела обстоят хуже. Вообще говоря существуют так называемые неберущиеся интегралы, то есть такие, для которых ответ нельзя записать формулой.¹ В связи с таким делом, в теории интегрирования есть лишь рецепты о том, как можно пытаться найти интеграл.

Удобный формализм Интеграл состоит из двух картинок самого значка интеграла \int_a^b и подынтегрального выражения $f(x) dx$. Можно придать смысл этим отдельным выражениям, но тогда придется объяснять всякую абстрактную лабуду, потому проще относиться к ним как к картинкам. И как полагается, с картинками можно удобно поиграться. Например, можно определить дифференциал функции следующим образом $df = f'(x) dx$. При этом будут выполняться следующие свойства:

1. $d(f + g) = df + dg$.
2. $d(\lambda f) = \lambda df$.
3. $d(fg) = f dg + g df$.

Свойства интегралов

1. Линейность по функции.

$$\int_a^b (f(x) + g(x)) dx = \int_a^b f(x) dx + \int_a^b g(x) dx, \quad \int_a^b \lambda f(x) dx = \lambda \int_a^b f(x) dx$$

2. Линейность по отрезку интегрирования.

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx, \quad \int_a^b f(x) dx = - \int_b^a f(x) dx$$

3. Формула Ньютона-Лейбница (она же одномерная формула Стокса).

$$\int_a^b f'(x) dx = f(x)|_a^b = f(b) - f(a)$$

Подчеркнем, что $f(x)|_a^b = f(b) - f(a)$ – это удобное сокращение и ничего более. В терминах дифференциалов то же самое правило превращается в

$$\int_a^b df = f(x)|_a^b = f(b) - f(a)$$

На него можно смотреть так: знак интегрирования убивает знак дифференциала и меняется на знак $|_a^b$.

4. Интегрирование по частям. По правилу Лейбница имеем $(fg)' = f'g + fg'$. Если проинтегрировать это равенство, получим

$$\int_a^b (fg)' dx = \int_a^b (f'g + fg') dx = \int_a^b f'g dx + \int_a^b fg' dx$$

¹У кого случился флеш-бэк с корнями многочленов пятой степени и выше, это не случайно. Именем Галуа потоптались и по теории дифференциально-интегральных уравнений и доказали кучу разных результатов о неинтегрируемости различных функций. Вообще, если вы возьмете случайно написанную функцию, то она наверняка не интегрируется в том смысле, что для интеграла не существует формулы.

Применив правило Ньютона-Лейбница из предыдущего пункта к левой части и перекинув слагаемые, это правило обычно пишут так

$$\int_a^b f g' dx = f(x)g(x)|_a^b - \int_a^b f' g dx$$

Или в терминах дифференциалов

$$\int_a^b f dg = f(x)g(x)|_a^b - \int_a^b g df$$

Запас живых интегралов Табличные интегралы берутся из известных нам производных. Вот примеры того, что стоит знать

$$1. \int_a^b x^n dx = \frac{1}{n+1} x^{n+1} \Big|_a^b$$

$$2. \int_a^b \frac{1}{x} dx = \ln x \Big|_a^b$$

$$3. \int_a^b \cos x dx = \sin x \Big|_a^b \quad \text{и} \quad \int_a^b \sin x dx = -\cos x \Big|_a^b$$

$$4. \int_a^b e^x dx = e^x \Big|_a^b$$

$$5. \int_a^b \frac{1}{\cos^2 x} dx = \operatorname{tg} x \Big|_a^b \quad \text{и} \quad \int_a^b \frac{1}{\sin^2 x} dx = -\operatorname{ctg} x \Big|_a^b$$

$$6. \int_a^b \frac{1}{\sqrt{1-x^2}} dx = \arcsin x \Big|_a^b \quad \text{и} \quad \int_a^b \frac{1}{1+x^2} dx = \operatorname{arctg} x \Big|_a^b$$

Примеры Как и с производными, мы обычно сводим неизвестные нам интегралы к известным с помощью свойств. Только в случае с интегралом правила менее гибкие и не всегда работают. Вот пара примеров.

- Интегрирование рациональной дроби.

$$\int_a^b \frac{p(x)}{q(x)} dx \quad \text{где } p \text{ и } q - \text{многочлены}$$

Вообще говоря все такие интегралы считаются и делается это с помощью разложения в простейшие дроби вида $\frac{a}{(x-b)^d}$ или $\frac{ax+b}{(x^2+cx+d)^f}$, где во второй дроби находится многочлен без вещественных корней. Оказывается, что любую рациональную функцию можно разложить в сумму многочлена и дробей указанного вида. А для них интегралы берутся с помощью табличных. Я не буду разбирать общий случай а приведу пару примеров. Пусть $[a, b] \subseteq (-1, 1)$, тогда

$$\begin{aligned} \int_a^b \frac{1}{1-x^2} dx &= \frac{1}{2} \int_a^b \left(\frac{1}{1+x} + \frac{1}{1-x} \right) dx = \frac{1}{2} \int_a^b \frac{1}{1+x} dx + \frac{1}{2} \int_a^b \frac{1}{1-x} dx = \\ &= \frac{1}{2} \int_a^b \frac{1}{1+x} d(x+1) - \frac{1}{2} \int_a^b \frac{1}{1-x} d(1-x) = \frac{1}{2} \int_a^b d \ln(x+1) - \frac{1}{2} \int_a^b d \ln(1-x) = \frac{1}{2} \ln(x+1) \Big|_a^b - \frac{1}{2} \ln(1-x) \Big|_a^b \end{aligned}$$

Другой пример интеграла

$$\int_a^b \frac{x}{1+x^2} dx = \frac{1}{2} \int_a^b \frac{1}{1+x^2} dx^2 = \frac{1}{2} \int_a^b \frac{1}{1+x^2} d(1+x^2) = \frac{1}{2} \int_a^b d \ln(1+x^2) = \frac{1}{2} \ln(1+x^2) \Big|_a^b$$

- Давайте найдем интеграл от логарифма, воспользовавшись интегрированием по частям

$$\int_1^y \ln x \, dx = x \ln x \Big|_1^y - \int_1^y x \, d \ln x = y \ln y - 1 \ln 1 - \int_1^y dx = y \ln y - x \Big|_1^y = y \ln y - y + 1$$

Оценки интегралов Из-за того, что интегралы вообще говоря не считаются, а если и считаются, то получившееся выражение не всегда помогает понять, а что же мы получили, бывает требуется оценить значения интеграла сверху или снизу. В этих случаях полезно знать, какие бывают оценки на интегралы. Вот некоторые из них

1. Если функция f ограничена на отрезке $[a, b]$ следующим образом $m \leq f(x) \leq M$ для $x \in [a, b]$, то верно

$$m(b-a) \leq \int_a^b f(x) \, dx \leq M(b-a)$$

2. В более общем виде, если для двух функций f и g выполнено неравенство $f(x) \leq g(x)$ для всех $x \in [a, b]$, то выполнено

$$\int_a^b f(x) \, dx \leq \int_a^b g(x) \, dx$$

3. Внесение модуля под интеграл

$$\left| \int_a^b f(x) \, dx \right| \leq \int_a^b |f(x)| \, dx$$

Это, наверное, самое важное неравенство на свете. На нем строится весь функциональный анализ и смежные дисциплины.

4. Неравенство Коши-Шварца²

$$\left(\int_a^b f(x)g(x) \, dx \right)^2 \leq \left(\int_a^b f(x)^2 \, dx \right) \left(\int_a^b g(x)^2 \, dx \right)$$

На это утверждение можно смотреть так. На интеграл от произведения можно смотреть как на скалярное произведение

$$(f, g) = \int_a^b f(x)g(x) \, dx$$

Тогда «длина» функции относительно этого скалярного произведения

$$|f| = \sqrt{(f, f)} = \sqrt{\int_a^b f(x)^2 \, dx}$$

Но мы знаем, что каким бы ни было пространство со скалярным произведением, для него выполнено $(f, g) = |f||g| \cos \alpha$, где α – угол между функциями f и g . Потому неравенство Коши-Шварца говорит, что косинус угла по модулю не превосходит единицы, что понятно.

5. Неравенство Гельдера. Пусть $p, q \geq 1$ такие, что $\frac{1}{p} + \frac{1}{q} = 1$.

$$\left| \int_a^b f(x)g(x) \, dx \right| \leq \left(\int_a^b |f(x)|^p \, dx \right)^{\frac{1}{p}} \left(\int_a^b |g(x)|^q \, dx \right)^{\frac{1}{q}}$$

Это как бы несимметричная версия неравенства Коши.

²В России очень популярно название неравенство Коши-Буняковского. Ну нам же нужны русские фамилии в названии математических утверждений.

6. Неравенство Минковского. Пусть $p \geq 1$, тогда

$$\left(\int_a^b |f(x) + g(x)|^p dx \right)^{\frac{1}{p}} \leq \left(\int_a^b |f(x)|^p dx \right)^{\frac{1}{p}} + \left(\int_a^b |g(x)|^p dx \right)^{\frac{1}{p}}$$

На это неравенство можно смотреть так. Мы можем ввести «длину функции», которая не происходит из скалярного произведения следующим образом

$$\|f\|_p = \left(\int_a^b |f(x)|^p dx \right)^{\frac{1}{p}}$$

Тогда для этой длины выполняется неравенство треугольника $\|f + g\|_p \leq \|f\|_p + \|g\|_p$. По нормальному, то есть по математическому, у этой длины есть более приличное название, подобные длины называются нормами. О них я поговорю чуть позже. Поверьте пригодится.

Замена переменной Пусть у нас есть два отрезка $[a, b]$ и $[c, d]$, координату на $[a, b]$ я буду обозначать за t , а координату на $[c, d]$ за x . Пусть у нас есть отображение $g: [a, b] \rightarrow [c, d]$ по правилу $x = g(t)$ причем $g(a) = c$ и $g(b) = d$. То есть мы переводим один отрезок в другой и при этом начало переходит в начало и конец в конец. На это можно смотреть так: мы задали параметризацию отрезка $[c, d]$ с помощью функции g определенной на отрезке $[a, b]$.

Пусть теперь у нас есть функция f на отрезке $[c, d]$. В этом случае интеграл f на отрезке $[c, d]$ можно посчитать с помощью параметризации следующим образом

$$\int_c^d f(x) dx = \boxed{\int_{g(a)}^{g(b)} f(x) dx = \int_a^b f(g(t)) d(g(t))} = \int_a^b f(g(t)) g'(t) dt$$

Важная часть формулы обведена в рамку. Левая и правая части от рамки – это всего лишь переобозначения. Думать про эту формулу можно так: если мы интегрируем по отрезку $[g(a), g(b)]$, то можно делать замену переменной, то есть надо вместо x подставить $g(t)$.

Дифференцирование интеграла Давайте рассмотрим интеграл с переменным верхним пределом, а именно, для произвольной функции f рассмотрим функцию

$$F(x) = \int_a^x f(t) dt$$

В этом случае оказывается, что производная F это f , то есть

$$F'(x) = \left(\int_a^x f(t) dt \right)' = f(x)$$

То есть если вам надо продифференцировать интеграл по верхнему пределу, то вы должны стереть значок интеграла и подставить значение верхнего предела в подынтегральную функцию. Если верхний предел сложно зависит от параметра, то дифференцирование идет по правилу дифференцирования сложной функции

$$\left(\int_a^{g(x)} f(t) dt \right)' = F(g(x))' = F'(g(x)) g'(x) = f(g(x)) g'(x)$$

Дифференцирование по нижнему пределу происходит аналогично и следует из того, что мы можем сменить нижний и верхний предел, поменяв знак интеграла

$$\left(\int_x^b f(x) dx \right)' = - \left(\int_b^x f(x) dx \right)' = -f(x)$$

Есть еще одна ситуация, которую стоит иметь в виду, когда функция f зависит от двух параметров x и t , в этом случае получается правило

$$\left(\int_a^b f(x, t) dt \right)' = \int_a^b \frac{\partial}{\partial x} f(x, t) dt$$

здесь значок $\frac{\partial}{\partial x}$ означает частную производную по x , то есть вы смотрите на функцию $f(x, t)$ как на функцию от x , считая t постоянной величиной и дифференцируете в обычном смысле по x . Результат будет функцией, которая зависит и от x и от t , вот ее и называют $\frac{\partial}{\partial x} f(x, t)$. Такую формулу гарантированно можно применить, когда $\frac{\partial}{\partial x} f(x, t)$ непрерывна как функция двух переменных. В зависимости от выбранного интеграла этот результат можно усиливать. Но в целом можно особенно не задумываться и пользоваться этим правилом.

Если же надо продифференцировать интеграл сразу по всему одновременно, то получится следующее

$$\left(\int_{h(x)}^{g(x)} f(x, t) dt \right)' = f(x, g(x))g'(x) - f(x, h(x))h'(x) + \int_{h(x)}^{g(x)} \frac{\partial}{\partial x} f(x, t) dt$$

Почему это именно так, я расскажу чуть позже, когда будут функции от многих переменных, а сейчас лишь скажу, что надо на это смотреть так. Мы вводим функцию трех переменных

$$F(u, v, w) = \int_v^u f(w, t) dt$$

А нам надо посчитать производную функции

$$\phi(x) = F(g(x), h(x), x)$$

Это делается по правилу

$$\phi(x)' = \frac{\partial}{\partial u} F(g(x), h(x), x)g'(x) + \frac{\partial}{\partial v} F(g(x), h(x), x)h'(x) + \frac{\partial}{\partial w} F(g(x), h(x), x)$$

Больше я пока не хочу говорить про частные производные, потому пока надо запомнить правило, а понять его лучше получится на следующем занятии.

Неопределенный интеграл Неопределенным интегралом функции f называется функция F такая, что $F' = f$. Таких функций много, но все они отличаются на константу, то есть если $F_1' = f$ и $F_2' = f$, то $F_2 = F_1 + c$, где $c \in \mathbb{R}$. Неопределенный интеграл обозначается так

$$\int f(x) dx = F(x) + c$$

Понятно, что какую-то функцию F можно посчитать по правилу

$$F(x) = \int_a^x f(t) dt$$

Идея неопределенного интеграла в том, чтобы явно не считать константу c . Например, посчитаем интеграл $\int_1^x \ln t dt$. Давайте посмотрим на соответствующий неопределенный интеграл

$$\int \ln x dx = x \ln x - \int x d \ln x = x \ln x - \int x \frac{1}{x} dx = x \ln x - \int dx = x \ln x - x + c$$

Так как неопределенный интеграл – это определенный интеграл с переменным верхним пределом плюс любая константа, то для его вычисления можно использовать любые правила вычисления определенных интегралов и еще в качестве бонуса можно откидывать все константы вне интегралов.

Многомерные интегралы

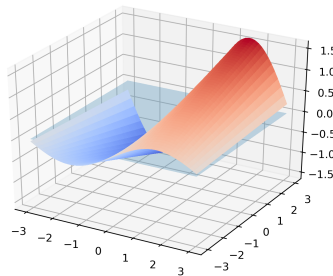
Мы с вами уже изучили аналог производных в случае функции нескольких переменных. Теперь давайте поймем, что происходит с интегралами. Пусть у нас есть функция нескольких переменных $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и пусть в \mathbb{R}^n у нас задана какая-нибудь фигура $A \subseteq \mathbb{R}^n$. Самый простой вид фигуры – параллелепипед

$$\Pi = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid a_1 \leq x_1 \leq b_1, \dots, a_n \leq x_n \leq b_n\} = [a_1, b_1] \times \dots \times [a_n, b_n]$$

Или может быть шар

$$B_r(a) = \{x \in \mathbb{R}^n \mid |x - a| \leq r\}$$

Тогда мы хотим посчитать ориентированный $n + 1$ -мерный объем под графиком функции $y = f(x_1, \dots, x_n)$ над фигурой A . Причем объемы над плоскостью $y = 0$ мы будем считать с плюсом, а под ней с минусом. На картинке ниже я изобразил положительную часть в красном цвете, а отрицательную часть в синем. В качестве фигуры выступает квадрат.



Объем под графиком функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ над фигурой $A \subseteq \mathbb{R}^n$ называется интегралом (кратным интегралом, многомерным интегралом и т.д.) и обозначается следующим образом

$$\int_A f(x_1, \dots, x_n) dx_1 \dots dx_n$$

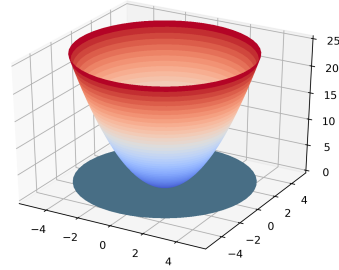
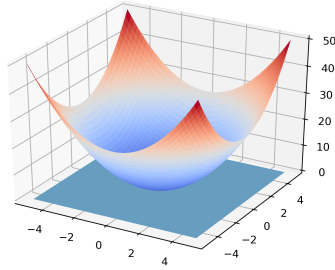
Еще используется обозначение

$$\int \dots \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n$$

Фигура $A \subseteq \mathbb{R}^n$ называется областью интегрирования.

Замечание про область интегрирования Я тут говорю слова про «фигура» и не уточняю, что это значит. Связано это с тем, что класс доступных множеств из \mathbb{R}^n по которым можно интегрировать функции зависит от вида используемого интеграла. Но мы не будем заморачиваться с конкретной конструкцией. А в практике, все что встречается будет вполне конкретной фигурой: прямоугольники, треугольники, многоугольники, диски, параллелепипеды, шары, объединения этих фигур, области задаваемые какими-то функциями и т.д. Обычно на практике всегда понятно, что за область нам дана и как с ней работать. Потому нет особого смысла разводить большую теорию на этот счет.

Примеры Для большей наглядности ниже изображена функция двух переменных $z = x^2 + y^2$. В первом случае в качестве фигуры выбран квадрат $A_1 = [-5, 5] \times [-5, 5]$, а во втором случае диск радиуса 5, то есть $A_2 = \{(x, y) \mid x^2 + y^2 \leq 5^2\}$. В данном случае мы хотим посчитать объем (3-мерный) под графиком поверхности $z = x^2 + y^2$ в зоне ограниченной фигурой A_1 в первом случае и фигурой A_2 во втором.



Как считать интегралы

Как обычно надо знать пару рецептов. Основная идея в том, что кратные интегралы можно сводить к обычным интегралам от одной переменной, а еще можно делать хитрые замены переменных. Но обо всем по порядку, начнем с простых свойств.

Свойства

1. Для любой функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$, любого числа $\lambda \in \mathbb{R}$ и любой фигуры $A \subseteq \mathbb{R}^n$ выполнено

$$\int_A \lambda f(x_1, \dots, x_n) dx_1 \dots dx_n = \lambda \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n$$

2. Для любых функций $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$ и любой фигуры $A \subseteq \mathbb{R}^n$ выполнено

$$\int_A (f(x_1, \dots, x_n) + g(x_1, \dots, x_n)) dx_1 \dots dx_n = \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n + \int_A g(x_1, \dots, x_n) dx_1 \dots dx_n$$

3. Для любой функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и любых двух не пересекающихся фигур $A, B \subseteq \mathbb{R}^n$ выполнено

$$\int_{A \cup B} f(x_1, \dots, x_n) dx_1 \dots dx_n = \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n + \int_B f(x_1, \dots, x_n) dx_1 \dots dx_n$$

Сведение к одномерным интегралам

Теперь я хочу объяснить как устроено сведение интеграла к последовательному вычислению одномерных интегралов. В начале я сформулирую общее правило только для прямоугольных областей, а потом в случае двух переменных объясню как поступать с произвольными областями.

Общий случай Пусть нам дана функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ и прямоугольный параллелепипед $A = [a_1, b_1] \times \dots \times [a_n, b_n]$. Тогда

$$\int_A f(x_1, \dots, x_n) dx_1 \dots dx_n = \int_{a_n}^{b_n} \left(\dots \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1, \dots, x_n) dx_1 \right) dx_2 \dots \right) dx_n$$

Если вы задумываетесь когда такое работает, то правило следующее, если многомерный интеграл по A существует, то существуют и одномерные интегралы справа и формула работает. Но как показать, что многомерный интеграл существует? Оказывается, если взять модуль функции f , то из существования одномерных интегралов справа следует существование интеграла слева. Но опять же, я бы не парился и просто применял все эти правила. Очень часто просто существует интеграл, который работает в вашем конкретном случае.

Давайте забудем этот кошмар и я продемонстрирую ситуацию на серии более или менее конкретных примеров. Пусть, например, у нас задана функция двух переменных $f(x, y)$ и мы хотим проинтегрировать ее по прямоугольнику $A = [a, b] \times [c, d]$, то есть ищем

$$\int_A f(x, y) dx dy$$

Тогда

$$\int_A f(x, y) dx dy = \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x, y) dx \right) dy$$

Кроме того, верно и в другом порядке

$$= \int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x, y) dy \right) dx$$

Я на всякий случай распифую первое равенство. Мы сначала смотрим на функцию $f(x, y)$ как на функцию от x , а y считаем фиксированным параметром. И интегрируем по отрезку $[a_1, b_1]$ эту функцию, получаем

$$\varphi(y) = \int_{a_1}^{b_1} f(x, y) dx$$

Далее мы полученную функцию от y интегрируем по $[a_2, b_2]$, получаем

$$\int_{a_2}^{b_2} \varphi(y) dy = \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x, y) dx \right) dy$$

Кстати, интеграл по такому прямоугольнику часто обозначают так

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x, y) dx dy$$

Причем \int и dy сгруппированы как скобки, внешний интеграл соответствует внешнему дифференциалу dy .

Функция двух переменных на квадрате Вот конкретный пример. Пусть у нас $A = [0, 1] \times [0, 1]$, а $f(x, y) = x^2 + y^2$. Тогда

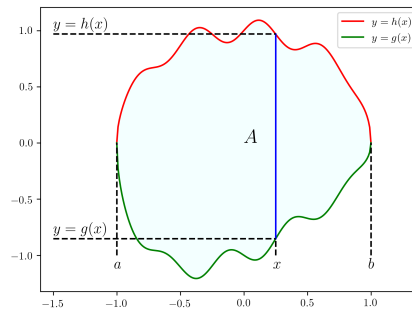
$$\begin{aligned} \int_A (x^2 + y^2) dx dy &= \int_0^1 \left(\int_0^1 (x^2 + y^2) dx \right) dy = \int_0^1 \left(\left(\frac{x^3}{3} + y^2 x \right) \Big|_0^1 \right) dy = \\ &= \int_0^1 \left(\frac{1}{3} + y^2 \right) dy = \left(\frac{1}{3} y + \frac{y^3}{3} \right) \Big|_0^1 = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \end{aligned}$$

Функция двух переменных на диске Пусть теперь в качестве фигуры выступает диск радиуса 1, то есть $A = \{(x, y) \mid x^2 + y^2 \leq 1\}$, а функция та же самая $f(x, y) = x^2 + y^2$. Тогда надо посмотреть какое наибольшее и наименьшее значение может принимать первая переменная x . Мы видим, что он может меняться от -1 до 1 . Но тогда для каждого фиксированного $x \in [-1, 1]$, число y может меняться от $-\sqrt{1-x^2}$ до $\sqrt{1-x^2}$. В этом случае мы можем посчитать наш интеграл следующим образом

$$\int_A (x^2 + y^2) dx dy = \int_{-1}^1 \left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} (x^2 + y^2) dy \right) dx$$

Обратите внимание, что переменная с которой мы начинали – x , будет самой внешней переменной интегрирования, далее идет y . Теперь лишь остается техническая задача посчитать данные интегралы. Я не буду никого мучить вычислениями. Тем более, что тут можно проделать вычисление сильно проще.

Функция двух переменных на произвольном множестве В общем случае если ваша фигура задана как-то более сложно, применяется точно такой же подход как и в предыдущем примере. Пусть у нас опять функция двух переменных $f(x, y)$ и фигура $A \subseteq \mathbb{R}^2$ как на картинке ниже.



Верхняя граница фигуры задана уравнением $y = h(x)$, а нижняя условием $y = g(x)$. При этом x меняется в диапазоне от a до b . При фиксированном x , переменная y бежит по синему вертикальному отрезку от $g(x)$ до $h(x)$. Потому интеграл f по A можно посчитать так

$$\int_A f(x, y) dx dy = \int_a^b \left(\int_{g(x)}^{h(x)} f(x, y) dy \right) dx$$

Если при этом $f(x, y) = 1$, то мы посчитаем площадь нашей сложной фигуры A .

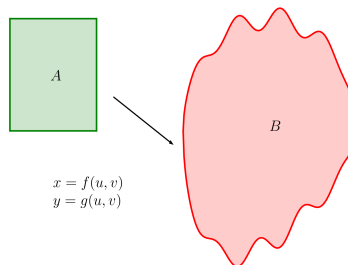
Замена переменных

Очень полезный прием вычисления интегралов от нескольких переменных – сделать замену области интегрирования с помощью замены переменных.

Предположим у нас есть фигура $A \subseteq \mathbb{R}^2$ и на этой плоскости координаты u и v . И еще есть фигура $B \subseteq \mathbb{R}^2$ и на этой плоскости координаты x и y . И предположим у нас есть отображение $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ заданное по правилу

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f(u, v) \\ g(u, v) \end{pmatrix}$$

Которое почти взаимно однозначно переводит фигуру A в B . Под почти взаимно однозначным отображением я имею в виду отображение, которое однозначно на внутренней части фигуры и лишь на границе может быть не однозначным. Изобразим это на рисунке ниже.



Предположим, что у нас есть функция $h(x, y)$ и мы хотим посчитать интеграл

$$\int_B h(x, y) dx dy$$

Однако фигура B очень сложная и по ней считать интеграл не удобно. Оказывается можно сделать замену переменных с помощью отображения φ и сменить фигуру B на A . Для этого надо в интеграле вместо x и y подставить $f(u, v)$ и $g(u, v)$. Тогда фигура B заменится на фигуру A , функция $h(x, y)$ заменится на $h(f(u, v), g(u, v))$ и осталось понять, на что заменится $dxdy$. Для начала составим следующую матрицу, называемую якобианом φ

$$J(\varphi) = \begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{pmatrix}$$

Оказывается, что $dxdy = |\det(J(\varphi))| dudv$.³ Таким образом интеграл считается по правилу

$$\int_B h(x, y) dxdy = \int_A h(f(u, v), g(u, v)) |\det(J(\varphi))| dudv$$

Применение замены переменных Давайте посчитаем интеграл

$$\int_{x^2+y^2 \leq a^2} (x^2 + y^2) dxdy$$

с помощью замены переменных. Этот интеграл ведется по кругу радиуса a . Давайте сделаем следующую замену переменных

$$\begin{cases} x = r \cos t \\ y = r \sin t \end{cases}$$

Тогда если t пробегает от 0 до 2π , а r пробегает от 0 до a , то (x, y) пробегает весь диск. Это отображение не однозначно лишь на границе квадрата. Построим якобиан для нашей замены

$$J = \begin{pmatrix} \cos t & -r \sin t \\ \sin t & r \cos t \end{pmatrix}$$

Определитель якобиана $\det J = r$. Значит

$$\int_{x^2+y^2 \leq a^2} (x^2 + y^2) dxdy = \int_{[0, a] \times [0, 2\pi]} r^2 |r| dr dt$$

А теперь посчитаем с помощью одномерных интегралов

$$\int_0^{2\pi} \left(\int_0^a r^3 dr \right) dt = \int_0^{2\pi} \frac{r^4}{4} \Big|_0^a dt = \int_0^{2\pi} \frac{a^4}{4} dt = 2\pi \frac{a^4}{4} = \frac{\pi a^4}{2}$$

Равномерная сходимость и интегралы

Пусть $f_n(x)$ – последовательность непрерывных функций на отрезке $[0, 1]$.⁴ И пусть она поточечно сходится к некоторой непрерывной функции $f(x)$ на этом отрезке. Тогда оказывается, что f_n сходится равномерно к f . Это очень полезное утверждение. Главный бонус от него заключается вот в чем. Если на отрезке $[0, 1]$ последовательность непрерывных функций f_n равномерно сходится к функции f , то функция f тоже непрерывна и верно равенство

$$\lim_{n \rightarrow \infty} \int_0^1 f_n(x) dx = \int_0^1 \lim_{n \rightarrow \infty} f_n(x) dx = \int_0^1 f(x) dx$$

³На самом деле у многомерного интеграла есть ориентированная версия, в которой не нужен модуль определителя, не нужна однозначность параметризации, но он будет менять знак от перестановки dx_i . Если вы встретите такие слова как интегрирование дифференциальных форм, теорема Стокса и т.д., то знайте, это про многомерный ориентированный интеграл.

⁴Можно взять любой отрезок, $[0, 1]$ выбран исключительно из эстетических соображений.

Преобразование Фурье

Есть несколько разных контекстов, в которых возникают преобразования Фурье. К классическим относятся преобразования Фурье на окружности (ну или на отрезке) и на прямой. Классические формулы на прямой работают только для очень небольшого класса функций, что приводит к необходимости рассмотреть неклассические преобразования Фурье, однако это требует так называемых «обобщенных функций».⁵ Кроме того, если мы хотим что-то посчитать, то нужно уметь непрерывную задачу привести к дискретному случаю, так возникает дискретная версия преобразования Фурье, для которой есть эффективный алгоритм вычисления под названием «быстрое преобразование Фурье». Я постараюсь рассказать про идею преобразования Фурье, после чего опишу в общих чертах классические случаи и дискретный случай на окружности.

Идея преобразования Фурье

Внимание, сейчас будет неприятно. **ОСТОРОЖНО: Физика.** Давайте представим себе задачу колебания чего-либо на прямой. Представьте, что вы закрепили шнур конечной длины с одной стороны, а с другой руками дергаете его и смотрите на то, какую форму он принимает и при этом тоже не двигаетесь с места. Если остановить время, то в каждый момент времени шнур будет давать вам график какой-то функции на прямой. Чем дальше точка на шнуре от состояния покоя, тем сильнее она отклонена, чем ближе, тем менее она отклонена от покоя. Таким образом на такой график можно смотреть как на сигнал в пространстве, где по оси x идет координата точек шнура, а по оси y идет степень отклонения точки от положения покоя (то есть «сила» колебания). Аналогично можно бросить камень в бассейн с водой и смотреть на форму поверхности воды сбоку. Тогда это будет уже не колебание шнура, а колебание воды. Аналогично можно думать и про распространение звука на прямой – это колебание молекул воздуха. Точно так же распространение света – это колебание электромагнитного поля. Такая картинка со шнуром или водой очень помогает думать про абстрактные колебания.

В нашей ситуации выше мы брали шнур конечной длины, это соответствует колебанию на отрезке. Про такие колебания еще полезно думать, как колебания струны музыкального инструмента. Однако, полезно рассматривать еще колебания «бесконечного» шнура в обе стороны. Эти две задачи являются предметом изучения классического случая для преобразования Фурье. Давайте рассмотрим их отдельно.

Отрезок (или окружность) В случае отрезка $[0, 2\pi]$ можно рассмотреть «элементарные колебания» вида $\sin(kx)$ и $\cos(mx)$ для $k, m \in \mathbb{Z}$, $k > 0$ и $m \geq 0$. Эти функции отвечают колебаниям с дискретными частотами k и m . Если думать про колебание струны гитары, то можно думать, что такое элементарное колебание дает звук определенной частоты. То есть в случае конечной струны у нас есть лишь дискретный набор частот, которые могут возникать. Теперь интересная задача заключается в том, чтобы определить, а какие частоты входят в сигнал и с какой силой они входят. По сути это задача разложения функции на отрезке в ряд из синусов и косинусов с коэффициентами.

Прямая В случае прямой к «элементарным колебаниям» уже относятся колебания с произвольными частотами вроде $\sin(\lambda x)$ и $\cos(\mu x)$, где $\lambda, \mu \in \mathbb{R}$ и $\lambda > 0$ и $\mu \geq 0$. Тут хорошо думать про колебание как про световой сигнал в пространстве на прямой. Мы знаем его мощность в каждой точке пространства. А элементарные частоты отвечают цвету (на самом деле тут надо использовать комплексные функции, чтобы правильно говорить про свет, так как у него есть поляризация, но я сейчас рассказываю на пальцах). Частоты в этом случае отвечают различным цветам, которые мы можем видеть или же инфракрасному излучению (то есть тепловому излучению), или ультрафиолетовому излучению, или радио сигналу и т.д. Правильно рассматривать колебания электромагнитного поля, загадочной субстанции, которая все объясняет. Потому, если мы хотим понять какое излучение присутствует в нашем сигнале, нам надо уметь вычислять вклад частот в наш сигнал.⁶ Так как на прямой частот континуум, то в виде суммы таких элементарных сигналов функцию уже не представишь, потому разложение в ряд заменяется на интеграл.

⁵Обобщенные функции – это феномен русского образования. Во всем мире их называют просто распределениями.

⁶Например, мы знаем, что разные атомы разных веществ излучают сигналы на разных частотах. Потому если мы хотим узнать состав далекой звезды, то нам бы хотелось принять свет от нее и разложить его в набор таких частот, что дает нам знание о том, что же там за элементы. Кроме того, на этом подходе основаны методы определения состава сплавов и проверки качества материалов.

Преобразование Фурье на отрезке (окружности)

Формально преобразование Фурье на отрезке начинается так. Давайте рассмотрим функции

$$\{f: [0, 2\pi] \rightarrow \mathbb{C}\}$$

Это то же самое, что рассмотреть 2π периодические функции на прямой или что то же самое, рассмотреть функции на окружности, где $\varphi \in [0, 2\pi]$ – угол параметризующий точку на окружности. При этом мы не предполагаем, что функции непрерывны или обладают какими-то еще дополнительными свойствами. Про них мы поговорим позже. Так же я рассматриваю комплекснозначные функции, потому что это делает теорию технически более простой и очень хорошо подходит для случая электромагнитных колебаний.

В этом случае в качестве «элементарных колебаний» мы рассматриваем функции $\varphi_k(x) = e^{ikx} = \cos(kx) + i \sin(kx)$, где $k \in \mathbb{Z}$. Это функции имеющие период $2\pi/|k|$. То есть все элементарные колебания параметризованы целыми числами. Преобразованием Фурье от функции f называется набор коэффициентов $a_k \in \mathbb{C}$ таких, что

$$a_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx$$

Чтобы такой интеграл имел смысл нам достаточно предположить, что существует интеграл $\|f\|_1 = \int_0^{2\pi} |f(x)| dx$. Если мы думаем про $f(x)$ как про отклонение колеблющегося материала от состояния покоя (то есть нуля), то чем больше $|f(x)|$ тем, мощнее это отклонение и тем больше энергии содержится в нашем сигнале в точке x . Таким образом на интеграл $\|f\|_1$ можно смотреть как на среднюю энергию в нашем сигнале.⁷ Таким образом мы построили отображение

$$F: L_1[0, 2\pi] \rightarrow C_0(\mathbb{Z}), \quad f \mapsto \{a_k\}_{k \in \mathbb{Z}}$$

где $L_1[0, 2\pi]$ – функции на отрезке $[0, 2\pi]$ с интегрируемым модулем,⁸ $C_0(\mathbb{Z})$ – последовательности $a_k \in \mathbb{C}$ при $k \in \mathbb{Z}$, такие, что $a_k \rightarrow 0$ при $k \rightarrow \pm\infty$. Обратим внимание, что при этом отображении e^{imx} идет в последовательность $\{a_k = \delta_k^m\}$, то есть единичка в m -ом месте и нули в остальных местах. Про коэффициент a_k таким образом надо думать как про вклад элементарного колебания e^{ikx} в сигнал f . Условие того, что a_k стремятся к нулю на бесконечности означает, что колебания с высокими частотами вносят все меньше и меньше вклад в сигнал.

Свойства преобразования F Отображение F называется прямым преобразованием Фурье. Оно инъективно в следующем смысле, если две функции дают один и тот же набор коэффициентов, то они совпадают «почти всюду» на прямой (то есть множество точек $\{x \mid f(x) \neq g(x)\}$ имеет «длину» ноль).⁹ Давайте приведу главный пример. Если вы измените функцию f в одной точке, то интегралы от нее не изменятся. Аналогично, вы можете поменять f в счетном числе точек и это никак не отразится на коэффициентах a_k . Но если функции f и g непрерывны и у них совпадают преобразования Фурье, то они совпадают всюду на отрезке. Это отображение не является сюръективным, но в некотором смысле «почти сюръективно». Если ввести $\|a_k\|_\infty = \max_k |a_k|$, то мы получим понятие расстояния между двумя последовательностями $\rho(a_k, b_k) = \|a_k - b_k\|_\infty$. Тогда образ F плотен в $C_0(\mathbb{Z})$, то есть если даже какая-то последовательность не лежит в образе рядом с ней в равномерной метрике сколь угодно близко можно найти последовательность из образа. По видимому лучше описать образ тут нельзя.

Введенную выше норму $\|a_k\|_\infty = \max_k |a_k|$, можно интерпретировать так: $\|a_k\|_\infty$ – самый большой вклад элементарного колебания по мощности. По определению можно показать, что $\|F(f)\|_\infty \leq \|f\|_1$. То есть максимальная энергия сигнала элементарного колебания не превосходит средней энергии всего сигнала. Что не удивительно, однако мы бы хотели какой-то закон сохранения энергии и возможность восстановить функцию f по набору коэффициентов a_k . Обратным преобразованием Фурье называется формула

$$\{a_k\} \mapsto \sum_k a_k e^{ikx}$$

То есть мы складываем все элементарные колебания с соответствующими мощностями. Есть одна большая проблема, как мы знаем сходимость a_k к нулю является необходимым, но не достаточным условием для сходимости ряда. Потому такая формула вообще говоря дает расходящиеся ряды. И возникает вопрос, а в каком

⁷Внимание, тут я пользуюсь словами «мощность», «энергия» и т.д. неформально. Физический смысл всему этому можно аккуратно придать лишь когда обсуждается конкретная физическая задача, а не общий математический формализм. Тем не менее, эти слова полезны для формирования какой-то интуиции, что же мы тут считаем.

⁸Интеграл Лебега в этом плане очень удобен, потому что функция интегрируема тогда и только тогда, когда интегрируем модуль функции.

⁹Более подробно про размеры множеств я поговорю на теории вероятностей.

смысле вообще такие ряды сходятся и можно ли найти такой класс функций, которые будут восстанавливаться по своим коэффициентам.

Преобразование в L_2 Давайте рассмотрим следующие множества

$$L_2[0, 2\pi] = \left\{ f: [0, 2\pi] \rightarrow \mathbb{C} \mid \exists \int_0^{2\pi} |f(x)|^2 dx \right\}$$

$$l_2(\mathbb{Z}) = \left\{ a_k \mid k \in \mathbb{Z}, \exists \sum_k |a_k|^2 \right\}$$

Первое определяет запас функций – такие функции, что их модуль в квадрате интегрируем, а второе – запас последовательностей. Тогда можно показать, что прямое и обратное отображения Фурье являются взаимно обратными отображениями:

$$F: L_2[0, 2\pi] \rightarrow l_2(\mathbb{Z}) \quad F^{-1}: l_2(\mathbb{Z}) \rightarrow L_2[0, 2\pi]$$

$$f \mapsto \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) e^{-ikx} dx \quad a_k \mapsto \frac{1}{\sqrt{2\pi}} \sum_k a_k e^{ikx}$$

Обратите внимание на перенормировку в преобразовании Фурье. Она сделана вот по какой причине. Дело в том, что оба пространства содержат естественное скалярное произведение, а преобразования Фурье с такой перенормировкой его сохраняют. Точнее, на пространстве функций мы определим следующее скалярное произведение

$$(f, g) = \int_0^{2\pi} f(x) \overline{g(x)} dx$$

А на пространстве последовательностей

$$(a_k, b_k) = \sum_k a_k \overline{b_k}$$

Обратите внимание на комплексное сопряжение над вторым аргументом. Я не обсуждал формально скалярные произведения в комплексном случае. Оказывается, что вот так как написано выше делать правильно, но я не хочу останавливаться на этом подробно. С такими скалярными произведениями верно равенство $(F(f), F(g)) = (f, g)$. Более того, на величины

$$\|f\|_2 = \sqrt{\int_0^{2\pi} |f(x)|^2 dx} \quad \text{и} \quad \|a_k\|_2 = \sqrt{\sum_k |a_k|^2}$$

можно смотреть как на «среднюю энергию» в f и a_k соответственно. Так как это просто длины векторов в терминах скалярных произведений и F сохраняет скалярное произведение, то длины тоже сохраняются. А значит $\|f\|_2 = \|F(f)\|_2$, то есть мы получили желаемый закон «сохранения энергии». Оказывается для сохранения энергии надо считать не площадь под графиком $|f|$, а площадь под графиком $|f|^2$. И вот в таком виде энергия сохраняется.

Тот факт, что прямое и обратное преобразование Фурье являются взаимно обратными, означает, что любую функцию можно разложить в ряд Фурье:

$$f(x) = \frac{1}{2\pi} \sum_k \left(\int_0^{2\pi} f(x) e^{-ikx} dx \right) e^{ikx}$$

Однако, надо уточнить очень важную вещь – в каком смысле сходится ряд в правой части. Вообще говоря поточечной сходимости может не быть у ряда в правой части. Однако, сходимость будет в смысле расстояния в $L_2[0, 2\pi]$. А именно, если обозначить

$$f_N(x) = \frac{1}{2\pi} \sum_{|k| \leq N} \left(\int_0^{2\pi} f(x) e^{-ikx} dx \right) e^{ikx}$$

то $\|f - f_N\|_2 \rightarrow 0$ при $N \rightarrow \infty$.

Еще один способ смотреть на композицию прямого и обратного преобразования Фурье следующий. Функции e^{ikx} при $k \in \mathbb{Z}$ образуют ортонормированный «базис» пространства $L_2[0, 2\pi]$ в следующем смысле. Они образуют ортогональную систему, то есть $(e^{ikx}, e^{imx}) = 2\pi\delta^{km}$. И при этом любая функция однозначно представляется в виде ряда $\sum_k a_k e^{ikx}$ (где сходимость ряда имеется в виду в смысле расстояния и равенство функций имеется в виду почти всюду). Тогда формула разложения функции в ряд Фурье можно трактовать как формулу вычисления коэффициентов при ортонормированном базисе, то есть

$$f = \sum_k \frac{(f, e^{ikx})}{(e^{ikx}, e^{ikx})} e^{ikx} = \frac{1}{2\pi} \sum_k (f, e^{ikx}) e^{ikx} = \frac{1}{2\pi} \sum_k \left(\int_0^{2\pi} f(x) e^{-ikx} dx \right) e^{ikx}$$

Тогда равенство $\|f\|_2 = \|F(f)\|_2$ – это просто теорема Пифагора в ортонормированном базисе $\frac{1}{\sqrt{2\pi}} e^{ikx}$.

Случай \mathbb{R} vs \mathbb{C} Если мы представили функцию f в виде ряда Фурье, то есть

$$f(x) = \frac{1}{2\pi} \sum_k a_k e^{ikx}$$

То коэффициенты $a_k \in \mathbb{C}$ отвечают за «амплитуду» и «поляризацию» для элементарного колебания e^{ikx} . Под амплитудой я понимаю $|a_k|$ – по сути она показывает на сколько мы сильно отклонились от нуля по расстоянию. А за поляризацию отвечает аргумент комплексного числа a_k . Поляризация просто так не видна глазу, но она дает вклад в эффект интерференции, когда при наложении волны могут друг друга усиливать, если у них одинаковая поляризация или наоборот могут гасить друг друга, если у них противоположная поляризация.

Не у всех колебаний есть поляризация, потому тогда естественно рассматривать не комплексный, а вещественный случай. В этом случае элементарными колебаниями являются $\sin(kx)$ при $k > 0$ и $\cos(kx)$ при $k \geq 0$ и при этом k целое. Тогда функция f на отрезке $[0, 2\pi]$ дает две серии коэффициентов

$$b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) dx \quad \text{и} \quad a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx \quad \text{и} \quad a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx$$

Тогда ряд Фурье будет иметь вид

$$a_0 + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx))$$

В случае квадратично интегрируемых функций f , то есть таких, что существует $\int_0^{2\pi} |f(x)|^2 dx$ ряд Фурье сходится к f в смысле метрики $\|\cdot\|_2$. Опять же, набор функций $1, \sin(kx), \cos(kx)$ при $k > 0$ является ортогональным базисом в $L_2[0, 2\pi]$ и потому можно применять обычные формулы в евклидовом пространстве для вычисления коэффициентов при базисных векторах

$$a_0 = \frac{(f, 1)}{(1, 1)}, \quad a_k = \frac{(f, \cos(kx))}{(\cos(kx), \cos(kx))}, \quad b_k = \frac{(f, \sin(kx))}{(\sin(kx), \sin(kx))}$$

где скалярное произведение задано $(f, g) = \int_0^{2\pi} f(x)g(x) dx$.

Преобразование Фурье на прямой

Для преобразования Фурье на прямой надо рассматривать функции

$$\{f: \mathbb{R} \rightarrow \mathbb{C}\}$$

В этом случае элементарными колебаниями являются $e^{i\lambda x}$, при $\lambda \in \mathbb{R}$. Преобразование Фурье тогда задано формулой¹⁰

$$f(x) \mapsto \hat{f}(\lambda) = \int_{-\infty}^{\infty} f(x) e^{-i\lambda x} dx$$

¹⁰Стоит сказать, что существуют разные нормировки этой формулы с помощью констант, например с помощью $\frac{1}{2\pi}$.

Чтобы эта формула имела смысла достаточно потребовать существования $\int_{\mathbb{R}} |f(x)| dx$. Функция $\hat{f}(\lambda)$ будет обязательно непрерывна и стремится к нулю при $\lambda \rightarrow \pm\infty$. То есть как и в случае отрезка мы получаем отображение

$$F: L_1(\mathbb{R}) \rightarrow C_0(\mathbb{R}), \quad f \mapsto \hat{f}$$

где левое пространство – это функции для которых модуль интегрируем по всей прямой, а правое пространство – непрерывные функции стремящиеся к нулю на обеих бесконечностях. Это отображение так же является инъективным в существенном смысле, что если две функции дают одно и то же преобразование Фурье, то они равны почти всюду, в частности на непрерывных функциях отображение инъективно. Так же это отображение не сюръективно и имеет плотный образ в равномерной норме.¹¹

Преобразование в L_2 Если так же сменить пространство на $L_2(\mathbb{R})$, то получим так называемую теорему «Планшереля». Пусть

$$L_2(\mathbb{R}) = \left\{ f: \mathbb{R} \rightarrow \mathbb{C} \mid \exists \int_{\mathbb{R}} |f(x)|^2 dx \right\}$$

Тогда прямое и обратное преобразование Фурье заданные формулами ниже будут взаимно обратными отображениями¹²

$$F: L_2(\mathbb{R}) \rightarrow L_2(\mathbb{R})$$

$$f(x) \mapsto \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\lambda x} dx$$

$$F^{-1}: L_2(\mathbb{R}) \rightarrow L_2(\mathbb{R})$$

$$\hat{f}(\lambda) \mapsto \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(\lambda) e^{i\lambda x} d\lambda$$

Единственное надо сказать небольшое замечание о том, как считать интегралы выше, потому что они вообще говоря могут не посчитаться. Надо взять

$$\psi_N(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-N}^N f(x) e^{-i\lambda x} dx$$

Тогда последовательность $\phi_N(\lambda)$ сходится к $\hat{f}(\lambda)$ в смысле метрики $\|-\|_2$ заданной скалярным произведением. То есть вообще говоря поточечной сходимости для каждого λ может просто не быть. Для хороших функций формулы конечно будут работать, но вот иногда при плохой скорости сходимости интеграла $\int_{\mathbb{R}} |f(x)| dx$ приходится считать в том смысле в котором указано выше. Тем не менее, для удобства (или наоборот для путаницы) люди все это обозначают в виде интегралов.

Как и в случае отрезка F сохраняет скалярное произведение $(f, g) = \int_{\mathbb{R}} f(x) \overline{g(x)} dx$, то есть $(F(f), F(g)) = (f, g)$. А значит и сохраняет длины векторов. То есть $\|f\|_2 = \|\hat{f}\|_2$. А это означает как раз «закон сохранения энергии». Существенное отличие тут от случая отрезка в том, что мы уже не можем просто так трактовать $e^{i\lambda x}$ как ортогональную систему. Хотя бы потому что эти функции не лежат в пространстве $L_2(\mathbb{R})$ и при подстановке их в интегралы мы получаем расходящийся результат.

Дискретное преобразование Фурье

К дискретному преобразованию Фурье можно прийти из разных соображений. Одно из них – мы хотим посчитать преобразование Фурье на отрезке на компьютере. Для этого надо на отрезке $[0, 2\pi]$ задать сетку с некоторым фиксированным шагом и после этого ограничить функции на этот отрезок. То есть если я хочу разбить отрезок на n частей, то я беру точки вида

$$\frac{2\pi \cdot 0}{n}, \frac{2\pi \cdot 1}{n}, \dots, \frac{2\pi \cdot k}{n}, \dots, \frac{2\pi \cdot (n-1)}{n}$$

точку 2π мы пропускаем, потому что мы хотим как бы склеить отрезок в окружность, где $0 = 2\pi$. Потому значение в 2π должно совпасть со значением в 0 .

¹¹Я позволю себе не расшифровывать эти слова и понадеюсь на вас, что вы продлите по аналогии случай отрезка.

¹²Удивительно, что в случае прямой тоже понадобился множитель $\frac{1}{\sqrt{2\pi}}$, хотя тут нет никакого ясного геометрического смысла

вроде длины окружности (или отрезка). Однако при такой нормировке функция $e^{-\frac{x^2}{2}}$ будет неподвижным вектором, то есть собственным вектором с собственным значением 1. Этот факт сразу должен вас насторожить, если вы знаете про нормальное распределение. Это в частности означает, что нормальное распределение должно играть какую-то особенную роль. Так и будет, мы об этом узнаем на теории вероятностей.

Таким образом если была функция $f: [0, 2\pi] \rightarrow \mathbb{C}$, то по ней можно построить вектор значений на сетке

$$f_{\bullet} = \left(f(0), f\left(\frac{2\pi}{n}\right), \dots, f\left(\frac{2\pi k}{n}\right), \dots, f\left(\frac{2\pi(n-1)}{n}\right) \right)$$

Как будут выглядеть элементарные колебания в этом случае? Давайте ограничим функцию e^{imx} на сетку. Получим последовательность

$$e^{im0}, e^{im\frac{2\pi}{n}}, \dots, e^{im\frac{2\pi k}{n}}, \dots, e^{im\frac{2\pi(n-1)}{n}}$$

Если ввести обозначения $\xi_m = e^{i\frac{2\pi m}{n}}$, то ξ_m будет корнем n степени из 1, то есть $\xi_m^n = 1$. Более того, ограничение функции e^{imx} превращается в последовательность

$$\psi_m = (1, \xi_m, \xi_m^2, \dots, \xi_m^k, \dots, \xi_m^{n-1})$$

В такой задаче мы хотим разложить произвольный вектор f_{\bullet} по векторам ψ_1, \dots, ψ_n . Преобразование Фурье будет вычислять коэффициенты разложения, а обратное преобразование по коэффициентам строить исходный вектор. Преобразование Фурье тогда задается такими формулами

$$a_k = \frac{1}{n} \sum_{r=0}^{n-1} f_r \xi_k^{-r}$$

Тогда мы имеем отображение

$$F: \mathbb{C}^n \rightarrow \mathbb{C}^n, \quad f_{\bullet} \mapsto \{a_k\}$$

Если ввести обозначения $\|f_{\bullet}\|_1 = \sum_{k=0}^{n-1} |f_k|$ и $\|a_k\|_{\infty} = \max_k |a_k|$, то верно утверждение $\|F(f_{\bullet})\|_{\infty} \leq \|f_{\bullet}\|_1$. То есть выполняется то же самое неравенство, что и в предыдущих случаях. Это неравенство лишь говорит, что в таком виде «энергия не сохраняется».

Случай L_2 Если рассмотреть аналог случая L_2 , то мы получим два отображения:

$$F: \mathbb{C}^n \rightarrow \mathbb{C}^n$$

$$F^{-1}: \mathbb{C}^n \rightarrow \mathbb{C}^n$$

$$f_{\bullet} \mapsto a_k = \frac{1}{\sqrt{n}} \sum_{r=0}^{n-1} f_r \xi_k^{-r}$$

$$a_k \mapsto \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} a_k \psi_k$$

Если ввести стандартное скалярное произведение в \mathbb{C}^n по правилу $(x, y) = \sum_{k=0}^{n-1} x_k \overline{y_k}$, то отображение F будет сохранять это скалярное произведение $(F(f_{\bullet}), F(g_{\bullet})) = (f_{\bullet}, g_{\bullet})$. А значит и будет сохранять длины векторов. То есть $\|F(f_{\bullet})\|_2 = \|f_{\bullet}\|_2$. В этом смысле исходный вектор и его преобразование Фурье несут «одинаковое количество энергии».

Как и в классических случаях, на эту задачу можно смотреть так. Векторы ψ_k при $0 \leq k < n$ являются ортогональным базисом, потому композиция преобразования Фурье и с обратным – это всего лишь разложение вектора f_{\bullet} по ортогональному базису, для которого годятся формулы

$$f_{\bullet} = \sum_{k=0}^{n-1} \frac{(f_{\bullet}, \psi_k)}{(\psi_k, \psi_k)} \psi_k = \frac{1}{n} \sum_{k=0}^{n-1} (f_{\bullet}, \psi_k) \psi_k = \frac{1}{n} \sum_{k=0}^{n-1} \left(\sum_{m=0}^{n-1} f_m \xi_k^{-m} \right) \psi_k$$

При этом мы видим, что $\|\psi_k\|_2 = \sqrt{n}$. Для подсчета каждого коэффициента a_k требуется $O(n)$ операций. Всего у нас n коэффициентов. Потому чтобы посчитать такое разложение нужно выполнить $O(n^2)$ операций. Оказывается, что вычисления можно выполнить за $O(n \log n)$ операций. Такой алгоритм называется Быстрое преобразование Фурье.

Семинар 2

Задачи:

1. Посчитайте неопределенный интеграл $\int \left(1 - \frac{1}{x^2}\right) \sqrt{x\sqrt{x}} dx$

2. Посчитайте определенный интеграл $\int_1^2 x \ln x dx$

3. Найдите предел

$$\lim_{x \rightarrow 0} x \int_x^1 \frac{\cos t}{t^2} dt$$

4. Найдите интеграл

$$\int_{x^2+y^2 \leq a^2} |xy| dx dy$$

5. Пусть $f: [-\pi, \pi] \rightarrow \mathbb{R}$ задана по правилу $f(x) = x$. Будем рассматривать f как элемент $L_2[-\pi, \pi]$, со скалярным произведением $(g, h) = \int_{-\pi}^{\pi} g(x)h(x) dx$.

(а) Покажите, что f ортогонально функциям $\cos(kx)$ при $k \in \mathbb{Z}$ и $k \geq 0$.

(б) Найдите разложение $f(x) = \sum_{k=1}^{\infty} a_k \sin(kx)$ посчитав коэффициенты Фурье a_k .

(с) Найдите длины векторов $\|f\|_2$ и $\|a_k\|_2$. Зная эти длины, найдите сумму ряда $\sum_{k=1}^{\infty} \frac{1}{k^2}$.