Содержание

1 Векторы. Линейная зависимость системы векторов. Базис линейного пространства. Скалярное произведение векторов.

Вектор — это направленный отрезок, который характеризует величину и направление. Геометрически его можно представить как стрелку от начала координат до некоторой точки.

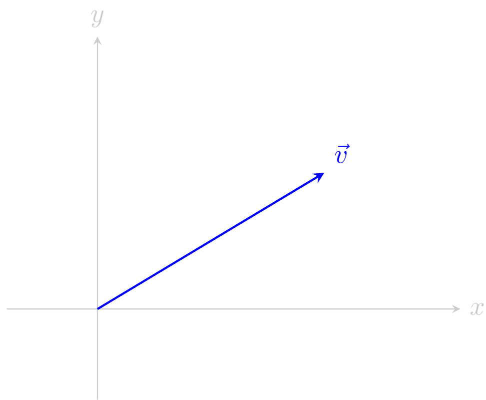
Алгебраически вектор в n-мерном пространстве $\mathbb{R}^n$ — это упорядоченный набор чисел:

$$\vec{v} = (v_1, v_2, \ldots, v_n)$$

Примеры:

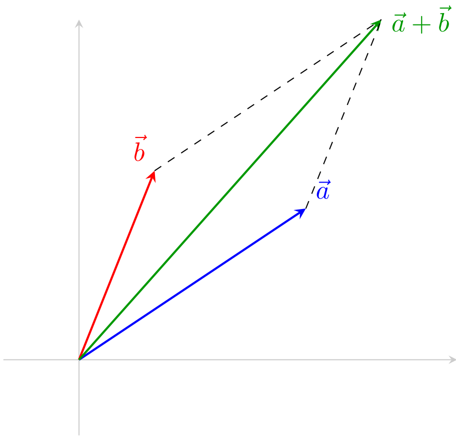
$$\vec{a} = (3, -1), \quad \vec{b} = (1, 2, 4)$$

Геометрическая интерпретация:



Основные операции:

- **Сложение:** $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$
- **Умножение на скаляр:** $\lambda \cdot (x, y) = (\lambda x, \lambda y)$
- **Нулевой вектор:** $\vec{0} = (0, 0, \ldots, 0)$



Линейная зависимость и независимость векторов:

Рассмотрим векторы $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$. Если существует набор коэффициентов $\lambda_1, \dots, \lambda_k$ (не все нули), такой что:

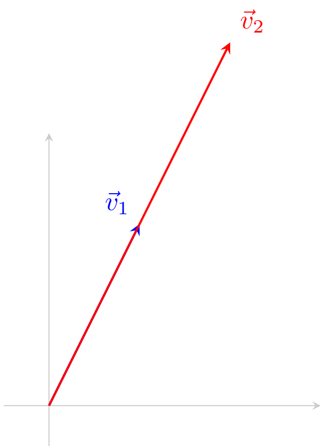
$$\lambda_1 \vec{v}_1 + \cdots + \lambda_k \vec{v}_k = \vec{0}$$

то векторы — **линейно зависимы**.

Если единственное решение — тривиальное ($\lambda_1 = \cdots = \lambda_k = 0$), то векторы — **линейно независимы**.

Пример:

$$\vec{v}_1 = (1, 2), \quad \vec{v}_2 = (2, 4) \Rightarrow \vec{v}_2 = 2\vec{v}_1 \Rightarrow \text{зависимы}$$

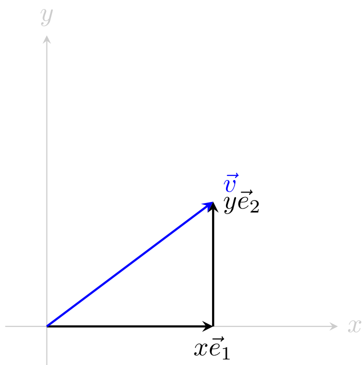


Базис линейного пространства:

Базис — это система линейно независимых векторов, которая порождает всё пространство. Любой вектор пространства выражается через базис как линейная комбинация.

В $\mathbb{R}^2$ стандартный базис:

$$\vec{e}_1 = (1, 0), \quad \vec{e}_2 = (0, 1) \Rightarrow \vec{v} = x\vec{e}_1 + y\vec{e}_2$$



Размерность пространства — это количество векторов в базисе. Например, в $\mathbb{R}^3$ базис содержит 3 вектора.

Скалярное произведение векторов:

Скалярное произведение двух векторов $\vec{a} = (a_1, \dots, a_n)$ и $\vec{b} = (b_1, \dots, b_n)$:

$$\langle \vec{a}, \vec{b} \rangle = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

Пример:

$$\vec{a} = (1, 2), \quad \vec{b} = (3, 4) \Rightarrow \langle \vec{a}, \vec{b} \rangle = 1 \cdot 3 + 2 \cdot 4 = 11$$

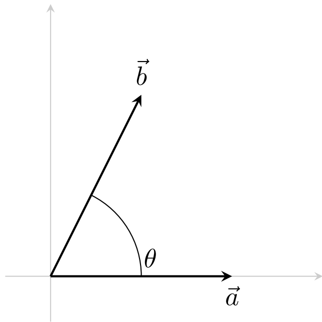
Свойства:

- Симметрия: $\langle \vec{a}, \vec{b} \rangle = \langle \vec{b}, \vec{a} \rangle$
- Линейность: $\langle \alpha\vec{a} + \beta\vec{c}, \vec{b} \rangle = \alpha\langle \vec{a}, \vec{b} \rangle + \beta\langle \vec{c}, \vec{b} \rangle$
- $\langle \vec{a}, \vec{a} \rangle = \|\vec{a}\|^2$

Геометрически:

$$\langle \vec{a}, \vec{b} \rangle = \|\vec{a}\| \|\vec{b}\| \cos\theta$$

где θ — угол между векторами. Если $\langle \vec{a}, \vec{b} \rangle = 0$ — векторы ортогональны (перпендикулярны).



Длина вектора (норма):

$$\|\vec{a}\| = \sqrt{\langle \vec{a}, \vec{a} \rangle}$$

Угол между векторами:

$$\cos\theta = \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\|\|\vec{b}\|}$$

—

Итоги:

- Векторы — фундаментальные объекты в линейной алгебре.
- Линейная зависимость помогает понимать структуру пространства.
- Базис — минимальный набор независимых векторов, порождающих всё пространство.
- Скалярное произведение связывает векторы с геометрией — углами и длинами.

2 Матрицы. Их свойства. Транспонированная матрица. Ранг матрицы

Матрица — это прямоугольная таблица чисел, организованная в строки и столбцы. Она записывается в виде:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

где a_{ij} — элемент матрицы на i-й строке и j-м столбце.

Обозначения и размерность

Матрицу обозначают заглавной латинской буквой (A, B, C и т.д.). Размерность матрицы — это количество строк и столбцов. Если в матрице m строк и n столбцов, её размер обозначают как $m \times n$.

Примеры:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 3 & 5 \end{pmatrix}$$

Здесь A — квадратная матрица 2×2, B — прямоугольная матрица 2×3.

—

Основные типы матриц

- **Нулевая матрица:** все элементы равны нулю.
- **Единичная матрица I_n:** квадратная матрица с единицами на главной диагонали и нулями вне её.

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- **Диагональная матрица:** все элементы вне главной диагонали равны нулю.
- **Квадратная матрица:** одинаковое число строк и столбцов.
- **Столбец (вектор-столбец):** матрица размером $m \times 1$.
- **Строка (вектор-строка):** матрица размером $1 \times n$.

—

Операции с матрицами

1. **Сложение:** складываются поэлементно. Возможно только для матриц одинакового размера.
$$A + B = (a_{ij} + b_{ij})$$
2. **Умножение на скаляр:**
$$\lambda A = (\lambda \cdot a_{ij})$$
3. **Умножение матриц:** если A — матрица размера $m \times n$, а B — $n \times k$, то их произведение $C = AB$ будет размером $m \times k$:
$$c_{ij} = \sum_{r=1}^{n} a_{ir} \cdot b_{rj}$$

4. **Транспонирование (см. ниже)** — замена строк и столбцов.

—

Свойства операций

- Коммутативность сложения: $A + B = B + A$
- Ассоциативность: $(A + B) + C = A + (B + C)$
- Дистрибутивность: $\lambda(A + B) = \lambda A + \lambda B$
- $(AB)^T = B^T A^T$ — важное свойство транспонирования

—

Транспонированная матрица

Матрица A^T (читается: «A транспонированная») получается из A заменой строк на столбцы. Формально:

$$(A^T)_{ij} = A_{ji}$$

Пример:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \quad \Rightarrow \quad A^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$$

Свойства транспонирования:

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(\lambda A)^T = \lambda A^T$
- $(AB)^T = B^T A^T$

—

Ранг матрицы

Ранг матрицы — это максимальное число линейно независимых строк (или столбцов) в матрице.

Обозначается: $\mathrm{rank}(A)$.

Интуитивно: ранг показывает, сколько ”уникальной”информации содержится в строках или столбцах.

Пример:

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \Rightarrow \text{строки линейно зависимы} \Rightarrow \mathrm{rank}(A) = 1$$

Другой пример:

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow \text{rank}(B) = 3$$

Как находить ранг?

Обычно с помощью преобразования матрицы к **ступенчатому виду** методом Гаусса. Количество ненулевых строк после преобразования и будет рангом.

Пример пошагово:

Дана матрица:

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 3 & 6 & 3 \end{pmatrix}$$

Видим: вторая и третья строки — кратные первой. После приведения:

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \text{rank}(A) = 1$$

—

Геометрическая интерпретация ранга

Векторы-строки (или столбцы) матрицы можно представить как векторы в пространстве. Ранг говорит о том, какое пространство они натягивают:

- Ранг 1: все лежат на одной прямой
- Ранг 2: в одной плоскости
- Ранг 3: в трёхмерном пространстве и т.д.

—

Важность ранга

Ранг используется в:

- Исследовании решений линейных систем: число решений зависит от ранга матрицы коэффициентов.
- Анализе линейной зависимости строк/столбцов.
- Проверке обратимости матрицы: квадратная матрица обратима $\iff$ её ранг равен размерности.

—

Выводы по теме

- Матрицы — основа линейной алгебры. Они обобщают векторы, храня данные и операции.
- Транспонирование меняет строки и столбцы местами.
- Ранг показывает, сколько независимых строк/столбцов содержит матрица.
- Если ранг меньше полной размерности — значит, матрица "выражает"только подпространство.

3 Сложение, умножение матрицы на число, умножение матриц, транспонирование матриц. Обратная матрица

Сложение матриц

Две матрицы A и B одинакового размера ($m \times n$) можно сложить, если у них совпадают размеры. Сложение происходит поэлементно:

$$(A + B)_{ij} = A_{ij} + B_{ij}$$

Пример:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \Rightarrow A + B = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix}$$

Свойства сложения:

- Коммутативность: $A + B = B + A$
- Ассоциативность: $(A + B) + C = A + (B + C)$
- Существование нулевой матрицы O (нулевая поэлементно): $A + O = A$

—

Умножение матрицы на число

Если $\lambda \in \mathbb{R}$ — число (скаляр), то умножение $\lambda \cdot A$ означает умножение каждого элемента матрицы на это число:

$$(\lambda A)_{ij} = \lambda \cdot A_{ij}$$

Пример:

$$A = \begin{pmatrix} 2 & -1 \\ 0 & 3 \end{pmatrix}, \quad \lambda = 4 \Rightarrow 4A = \begin{pmatrix} 8 & -4 \\ 0 & 12 \end{pmatrix}$$

Свойства:

- $\lambda(\mu A) = (\lambda\mu)A$
- $(\lambda + \mu)A = \lambda A + \mu A$
- $\lambda(A + B) = \lambda A + \lambda B$

—

Умножение матриц

Матрицы A и B можно перемножить, если **число столбцов в** A равно **числу строк в** B.

Если A — размера $m \times n$, а B — $n \times k$, то произведение $C = AB$ — это матрица $m \times k$, где:

$$C_{ij} = \sum_{r=1}^{n} A_{ir} \cdot B_{rj}$$

То есть: элемент C_{ij} получается как скалярное произведение i-й строки A и j-го столбца B.

Пример:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}$$

$$AB = \begin{pmatrix} 1 \cdot 0 + 2 \cdot 2 & 1 \cdot 1 + 2 \cdot 3 \\ 3 \cdot 0 + 4 \cdot 2 & 3 \cdot 1 + 4 \cdot 3 \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ 8 & 15 \end{pmatrix}$$

Важно: $AB \neq BA$ в общем случае! Умножение матриц **не коммутативно**.

Свойства:

- Ассоциативность: $A(BC) = (AB)C$
- Дистрибутивность: $A(B + C) = AB + AC$
- $(AB)^T = B^T A^T$ — транспонирование произведения

—

Транспонирование матрицы

Транспонирование — это операция, при которой строки становятся столбцами, а столбцы — строками.

Для любой матрицы A, её транспонированная матрица A^T определяется как:

$$(A^T)_{ij} = A_{ji}$$

Пример:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \Rightarrow A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

Свойства транспонирования:

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(\lambda A)^T = \lambda A^T$
- $(AB)^T = B^T A^T$

Эта операция часто используется при симметризации, а также в определениях симметрических и ортогональных матриц.

—

Обратная матрица

Обратная матрица A^{-1} к квадратной матрице A определяется как:

$$A \cdot A^{-1} = A^{-1} \cdot A = I$$

где I — единичная матрица той же размерности.

Условия существования:

- Матрица должна быть **квадратной**.
- Её **определитель не должен равняться нулю** ($\det A \neq 0$).
- Ранг A должен равняться её размерности: $\text{rank}(A) = n$

Пример:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{\det A} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \frac{1}{(1 \cdot 4 - 2 \cdot 3)} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 1.5 & -0.5 \end{pmatrix}$$

Способы нахождения обратной матрицы

1. Для 2×2-матриц:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

2. Для больших матриц:

    - Через присоединённую матрицу (алгебраические дополнения + транспонирование + деление на определитель)
    - Метод Гаусса: расширение A до $[A|I]$ и приведение к $[I|A^{-1}]$

—

Свойства обратной матрицы

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^T)^{-1} = (A^{-1})^T$

Важно: не все матрицы имеют обратную. Такие матрицы называются **вырожденными**.

—

Применения обратной матрицы

- Решение систем уравнений: $A\vec{x} = \vec{b} \Rightarrow \vec{x} = A^{-1}\vec{b}$
- Вывод формул в статистике и машинном обучении
- Нормализация линейных преобразований
- Преобразование координат

—

Выводы

- Операции над матрицами (сложение, умножение, транспонирование) формируют алгебраическую структуру.
- Умножение матриц — основа линейных отображений и систем уравнений.
- Транспонирование — полезная симметризующая операция.
- Обратная матрица существует только у невырожденных квадратных матриц и даёт способ обращения линейных операторов.

4 Аппроксимация и интерполяция функций

Аппроксимация и **интерполяция** — это два метода приближённого описания функций, основанные на наборе дискретных точек.

Интерполяция — это построение функции, которая точно проходит через заданные точки. **Аппроксимация** — это построение функции, которая приближённо описывает данные, но может не проходить через все точки.

—

Постановка задачи

Пусть дана таблица значений:

$$(x_0, y_0),\ (x_1, y_1),\ \ldots,\ (x_n, y_n)$$

Наша цель — построить функцию $f(x)$ такую, что:

- **Для интерполяции:** $f(x_i) = y_i$ для всех i
- **Для аппроксимации:** $f(x_i) \approx y_i$

—

Интерполяция: идея и цель

Интерполяция позволяет восстанавливать значение функции в промежуточных точках, не выходя за пределы интервала $[x_0, x_n]$.

Пример: если известно, что

$$f(1) = 2, \quad f(2) = 4, \quad f(3) = 6$$

можем интерполировать $f(x)$, скажем, через многочлен второй степени и вычислить $f(1.5)$.

—

Линейная интерполяция

Между двумя точками (x_0, y_0) и (x_1, y_1) интерполяционная функция задаётся по формуле:

$$f(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0)$$

Это уравнение прямой, проходящей через две точки. Очень просто, но недостаточно точно для сложных функций.

—

Полиномиальная интерполяция

Если заданы $n + 1$ точек, можно построить единственный многочлен степени не выше n, который проходит через все точки.

Формула Лагранжа:

$$P_n(x) = \sum_{i=0}^{n} y_i \cdot L_i(x), \quad \text{где } L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j}$$

Каждое $L_i(x)$ — базисный многочлен Лагранжа, равный 1 в точке x_i и 0 в остальных x_j.

Проблема: при увеличении числа узлов интерполяция может сильно колебаться (эффект Рунге), особенно на концах интервала.

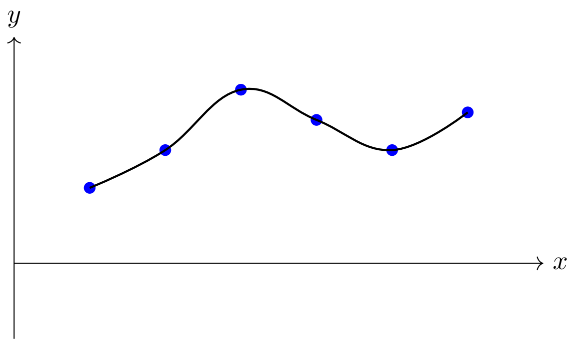
—

Сплайны (кубическая интерполяция)

Сплайн-интерполяция делит интервал на участки и на каждом строит многочлен степени 3 (кубический сплайн), с условием сглаженности в стыках.

- Гладкость первого и второго порядка: C^2-непрерывность
- Сплайны хорошо подходят для графиков, траекторий и данных с шумом

Визуально:



—

Аппроксимация: общая идея

Аппроксимация применяется, когда функция неизвестна, но имеются измеренные значения с шумом. Здесь уже не требуется точное прохождение через точки.

Идея: найти «наилучшую» функцию $f(x)$, которая *приблизительно* соответствует данным.

Часто ищут функцию в виде:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

Аппроксимация методом наименьших квадратов (МНК)

Пусть есть точки (x_i, y_i), и нужно найти параметры $a_0, a_1, \ldots, a_n$, минимизирующие отклонение:

$$S(a_0, a_1, \ldots, a_n) = \sum_{i=0}^{n} \left(f(x_i) - y_i\right)^2$$

Минимум достигается при решении системы нормальных уравнений, которая получается из частных производных S по параметрам a_k.

Частный случай — линейная аппроксимация:

$$f(x) = a_0 + a_1 x$$

Тогда минимизируется:

$$S(a_0, a_1) = \sum_{i=1}^{n} \left(a_0 + a_1 x_i - y_i\right)^2$$

Решение:

$$\begin{cases} n a_0 + a_1 \sum x_i = \sum y_i \\ a_0 \sum x_i + a_1 \sum x_i^2 = \sum x_i y_i \end{cases}$$

—

Сравнение: интерполяция vs аппроксимация

Критерий	Интерполяция	Аппроксимация
Проходит через точки	Да	Не обязательно
Чувствительность к шуму	Высокая	Устойчивая
Сложность вычислений	Средняя–высокая	Низкая–средняя
Гладкость	Может не быть	Обычно есть

—

Практические применения

- **Интерполяция**:
  - Таблицы и справочники
  - Заполнение пропущенных значений
  - Графическая визуализация

- **Аппроксимация**:
  - Обработка измерений с шумом
  - Моделирование реальных процессов
  - Предсказания, тренды

—

Выводы по теме

- Интерполяция позволяет точно восстановить функцию внутри диапазона, но может колебаться.
- Аппроксимация — более устойчивая техника, особенно с шумными или неточными данными.
- Полиномы Лагранжа и кубические сплайны — мощные методы интерполяции.
- Метод наименьших квадратов — классический способ аппроксимации, широко используемый в статистике и машинном обучении.

5 Производные. Необходимое и достаточное условия дифференцируемости функции. Частные и полные производные

Понятие производной функции одной переменной

Пусть $f(x)$ определена в окрестности точки x_0. Производная функции f в точке x_0 — это предел отношения приращения функции к приращению аргумента:

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

Если предел существует, то говорят, что функция **дифференцируема** в точке x_0.

Геометрический смысл: производная — это угловой коэффициент касательной к графику функции в данной точке.

Дифференцируемость и непрерывность

- Если функция дифференцируема в точке x_0, то она непрерывна в этой точке.

- Обратное неверно: непрерывность не гарантирует существование производной.

Пример (не дифференцируема):

$$f(x) = |x| \Rightarrow f'(0) \text{ не существует, хотя } f \text{ непрерывна в } 0$$

—

Производная по направлению и частные производные

Пусть $f(x, y)$ — функция двух переменных.

Производную по направлению можно определить через вектор направления $\vec{l} = (l_1, l_2)$:

$$D_{\vec{l}} f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + hl_1, y_0 + hl_2) - f(x_0, y_0)}{h}$$

Частные производные — это производные по отдельным переменным:

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

$$\frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}$$

Обозначения:

$$f_x(x, y), \quad f_y(x, y), \quad \text{или } \partial_x f, \ \partial_y f$$

Пример: пусть $f(x, y) = x^2 y + \sin(xy)$

$$\frac{\partial f}{\partial x} = 2xy + y\cos(xy), \quad \frac{\partial f}{\partial y} = x^2 + x\cos(xy)$$

—

Необходимое и достаточное условия дифференцируемости функции многих переменных

Функция $f(x, y)$ называется **дифференцируемой в точке** (x_0, y_0), если приращение можно представить в виде:

$$\Delta f = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) = A\Delta x + B\Delta y + o(\sqrt{\Delta x^2 + \Delta y^2})$$

где A и B — постоянные (зависят от точки), а $o(\rho)$ — бесконечно малая по сравнению с ρ.

Формально: f дифференцируема в (x_0, y_0), если существует линейное отображение L, приближающее Δf.

Необходимое условие дифференцируемости

Если f дифференцируема в (x_0, y_0), то:

- Частные производные $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ существуют
- $\Delta f = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + o(\sqrt{\Delta x^2 + \Delta y^2})$

Достаточное условие дифференцируемости

Если:

- Частные производные существуют в окрестности точки
- И они непрерывны в точке (x_0, y_0)

то f дифференцируема в этой точке.

—

Градиент и направление наибольшего роста

Градиент — это вектор, составленный из всех частных производных:

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

Производная функции по направлению вектора $\vec{l}$ выражается как скалярное произведение:

$$D_{\vec{l}} f = \nabla f \cdot \vec{l}$$

Свойства:

- Направление градиента — направление наибольшего возрастания функции.
- Если $\nabla f = \vec{0}$, то это критическая точка (возможно максимум, минимум или седло).

—

Полный дифференциал

Если функция $f(x, y)$ дифференцируема, то её полное приращение можно выразить через полный дифференциал:

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy$$

Пример: $f(x, y) = x^2 y + \sin(xy)$

$$df = (2xy + y\cos(xy))dx + (x^2 + x\cos(xy))dy$$

Полный дифференциал используется:

- В оценке приращений функции
- В дифференциальной геометрии и анализе ошибок
- При переходе к новым координатам

—

Итоги

- Производная — это мера изменения функции.
- Частные производные — изменения по осям координат.
- Дифференцируемость функции двух переменных требует не только существования производных, но и их «согласованного» поведения.
- Градиент показывает направление роста функции.
- Полный дифференциал — обобщение производной на многомерные функции.

6 Частные производные. Градиент функции. Производная по направлению

Частные производные функции двух переменных

Пусть $f(x,y)$ — функция двух переменных, определённая в окрестности точки (x_0, y_0). Тогда:

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0+h, y_0) - f(x_0, y_0)}{h}$$

$$\frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0, y_0+h) - f(x_0, y_0)}{h}$$

Частные производные — это производные функции по одной переменной при фиксированной другой.

Геометрический смысл: производная по x — это скорость изменения функции вдоль оси x, при фиксированном y.

Обозначения:

$$f_x = \frac{\partial f}{\partial x}, \quad f_y = \frac{\partial f}{\partial y}$$

Пример: Пусть $f(x, y) = x^2 y + \sin(xy)$. Тогда:

$$\frac{\partial f}{\partial x} = 2xy + y\cos(xy)$$

$$\frac{\partial f}{\partial y} = x^2 + x\cos(xy)$$

—

Частные производные высших порядков

Можно вычислять производные второго порядка и выше:

$$\frac{\partial^2 f}{\partial x^2}, \quad \frac{\partial^2 f}{\partial y^2}, \quad \frac{\partial^2 f}{\partial x \partial y}, \quad \frac{\partial^2 f}{\partial y \partial x}$$

Если f дважды непрерывно дифференцируема, то:

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

(теорема Шварца о равенстве смешанных производных)

—

Градиент функции

Пусть $f(x, y)$ — дифференцируемая функция. Тогда **градиент** функции f в точке (x_0, y_0) — это вектор:

$$\nabla f(x_0, y_0) = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0) \right)$$

Если f зависит от n переменных, то градиент — вектор из n компонент:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

Геометрический смысл:

- Направление градиента — это направление *наибольшего роста функции*.
- Его длина — скорость наибольшего изменения.
- Если $\nabla f = \vec{0}$, то точка является критической (возможный экстремум).

Пример: Пусть $f(x, y) = x^2 + y^2 \Rightarrow \nabla f = (2x, 2y)$ — вектор, указывающий от начала координат.

—

Производная функции по направлению

Пусть функция $f(x, y)$ определена в точке (x_0, y_0), и задан единичный вектор направления:

$$\vec{l} = (\cos\alpha, \cos\beta), \quad \|\vec{l}\| = 1$$

Производная функции f по направлению вектора $\vec{l}$ в точке (x_0, y_0):

$$D_{\vec{l}} f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h\cos\alpha, y_0 + h\cos\beta) - f(x_0, y_0)}{h}$$

Свойство: Если функция f дифференцируема, то производная по направлению вычисляется через градиент:

$$D_{\vec{l}} f = \nabla f \cdot \vec{l} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \cdot (\cos\alpha, \cos\beta)$$

Это **скалярное произведение** векторов: градиента и направления.
Следствия:

- Наибольшая производная по направлению достигается в направлении градиента.
- Если $\vec{l} \perp \nabla f$, то производная по направлению равна нулю (функция не меняется вдоль этого направления).

Пример:

Пусть $f(x, y) = x^2 y + y$, точка $M = (1, 2)$, направление $\vec{l} = \frac{1}{\sqrt{2}}(1, 1)$.

$$\nabla f = \big(2xy, x^2 + 1\big) \Rightarrow \nabla f(1, 2) = (4, 2)$$

$$D_{\vec{l}} f = \nabla f \cdot \vec{l} = \frac{1}{\sqrt{2}}(4 + 2) = \frac{6}{\sqrt{2}} = 3\sqrt{2}$$

—

Итоги

- Частные производные показывают, как функция меняется по каждой координате.
- Градиент — главный вектор изменения, указывает направление наибольшего роста.
- Производная по направлению обобщает понятие производной на произвольное направление.
- Всё вместе используется в оптимизации, градиентном спуске, анализе поверхности.

7 Численные методы решения задач оптимизации. Метод Ньютона и секущей. Методы покоординатного и градиентного спуска

Оптимизация — это процесс нахождения минимума или максимума функции при заданных ограничениях (или без них). В задачах прикладной математики часто встречаются функции, которые слишком сложны для аналитического нахождения экстремума, поэтому используют численные методы.

В данной секции мы рассмотрим:

- Метод Ньютона;
- Метод секущей;
- Метод покоординатного спуска;
- Метод градиентного спуска.

7.1 Постановка задачи оптимизации

Пусть дана функция:

$$f : \mathbb{R}^n \to \mathbb{R}, \quad f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$$

Необходимо найти точку $\mathbf{x}^* \in \mathbb{R}^n$, такую что:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n$$

(для задачи минимизации).

Для поиска минимума часто используют производные:

- **Необходимое условие экстремума:** $\nabla f(\mathbf{x}^*) = 0$
- **Достаточное условие минимума:** матрица Гессе $H_f(\mathbf{x}^*)$ положительно определена.

7.2 Метод Ньютона

Метод Ньютона — это численный метод, который использует разложение функции в ряд Тейлора второго порядка для приближения к экстремуму.

7.2.1 Алгоритм в одномерном случае

Для уравнения $f'(x) = 0$:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

Здесь:

- $f'(x_k)$ — первая производная функции в точке x_k;
- $f''(x_k)$ — вторая производная.

7.2.2 Многомерный случай

Для векторной функции:

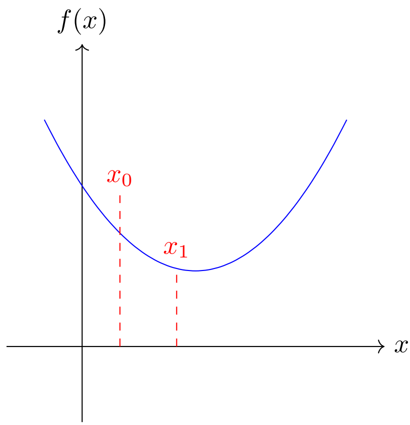
$$\mathbf{x}_{k+1} = \mathbf{x}_k - H_f^{-1}(\mathbf{x}_k)\nabla f(\mathbf{x}_k)$$

где $H_f(\mathbf{x}_k)$ — матрица Гессе.

7.2.3 Плюсы и минусы

- **Плюсы:** Быстрая сходимость (обычно квадратичная).
- **Минусы:** Нужно вычислять и инвертировать матрицу Гессе, что дорого для больших n.

7.2.4 Графическая иллюстрация



7.3 Метод секущей

Метод секущей — это упрощённая версия метода Ньютона, в которой вторую производную заменяют приближением по разностям.

$$x_{k+1} = x_k - f'(x_k) \cdot \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})}$$

- **Плюс:** Не нужно вычислять вторую производную.
- **Минус:** Скорость сходимости ниже, чем у метода Ньютона.

7.4 Метод покоординатного спуска

Метод покоординатного спуска оптимизирует функцию, изменяя одну координату за раз, оставляя остальные фиксированными.

7.4.1 Алгоритм

1. Выбираем начальную точку $\mathbf{x}_0$.
2. Для каждой координаты i ищем минимум функции по x_i при фиксированных остальных координатах.
3. Повторяем процесс до сходимости.

7.4.2 Особенности

- Подходит для задач, где минимизация по одной переменной проста.
- Может сходиться медленно, если переменные сильно связаны.

7.5 Метод градиентного спуска

Метод градиентного спуска — один из самых популярных методов оптимизации.

7.5.1 Основная идея

Движемся из текущей точки в направлении, противоположном градиенту функции (т.к. градиент указывает направление наибольшего роста).

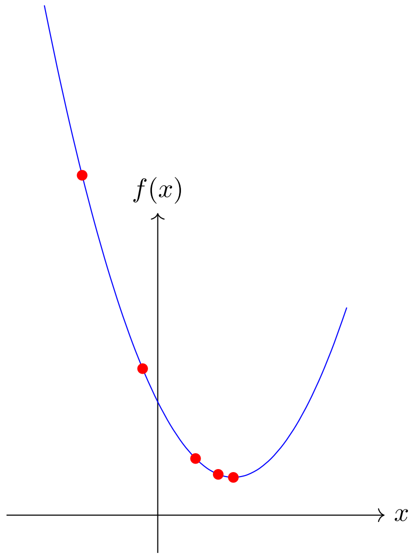
$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k)$$

где $\alpha > 0$ — шаг обучения.

7.5.2 Выбор шага

- Слишком большой α — можем «перепрыгнуть» минимум.
- Слишком маленький α — сходимость медленная.

7.5.3 Графическая иллюстрация



7.5.4 Варианты

- **С постоянным шагом**
- **С адаптивным шагом** (например, Adam, RMSprop)

7.6 Заключение

Выбор метода оптимизации зависит от свойств задачи:

- Метод Ньютона — быстрый, но требует вычислений второй производной.
- Метод секущей — компромисс, подходит для случаев, когда вторая производная недоступна.
- Покоординатный спуск — полезен при раздельной оптимизации переменных.
- Градиентный спуск — универсальный, особенно в задачах машинного обучения.