

A Corgi dog is sitting on the right side of the frame, looking towards the left. In the foreground, a yellow rubber duck is positioned on the left. Behind them is a chalkboard filled with mathematical content. At the top, a clapperboard graphic reads 'GRADIENT FUNCTION'. The board contains various mathematical notations including  $\nabla$ ,  $\frac{\partial}{\partial x}$ , and  $\frac{\partial}{\partial y}$ . There are also some handwritten numbers and symbols. A semi-transparent white box is overlaid on the center of the image, containing the main title and author information.

Основы линейной алгебры. SVD, Skeleton.  
Градиент. Гессиан. Матричное  
дифференцирование.

Даниил Меркулов

Методы оптимизации. МФТИ

# Основы линейной алгебры

## Векторы и матрицы

По умолчанию мы будем рассматривать все векторы как векторы-столбцы. Линейное пространство вещественных векторов длины  $n$  обозначается  $\mathbb{R}^n$ , а пространство вещественных матриц размера  $m \times n$  обозначается  $\mathbb{R}^{m \times n}$ . То есть: <sup>1</sup>

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad x^T = [x_1 \quad x_2 \quad \dots \quad x_n] \quad x \in \mathbb{R}^n, x_i \in \mathbb{R} \quad (1)$$

---

<sup>1</sup>A full introduction to applied linear algebra can be found in Introduction to Applied Linear Algebra -- Vectors, Matrices, and Least Squares - book by Stephen Boyd & Lieven Vandenberghe, which is indicated in the source. Also, a useful refresher for linear algebra is in Appendix A of the book Numerical Optimization by Jorge Nocedal Stephen J. Wright.

## Векторы и матрицы

По умолчанию мы будем рассматривать все векторы как векторы-столбцы. Линейное пространство вещественных векторов длины  $n$  обозначается  $\mathbb{R}^n$ , а пространство вещественных матриц размера  $m \times n$  обозначается  $\mathbb{R}^{m \times n}$ . То есть: <sup>1</sup>

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad x^T = [x_1 \quad x_2 \quad \dots \quad x_n] \quad x \in \mathbb{R}^n, x_i \in \mathbb{R} \quad (1)$$

Аналогично, если  $A \in \mathbb{R}^{m \times n}$  мы обозначаем транспонированную матрицу как  $A^T \in \mathbb{R}^{n \times m}$ :

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix} \quad A \in \mathbb{R}^{m \times n}, a_{ij} \in \mathbb{R}$$

Будем писать  $x \geq 0$  и  $x \neq 0$ , подразумевая покомпонентные соотношения.

---

<sup>1</sup>A full introduction to applied linear algebra can be found in Introduction to Applied Linear Algebra -- Vectors, Matrices, and Least Squares - book by Stephen Boyd & Lieven Vandenbergh, which is indicated in the source. Also, a useful refresher for linear algebra is in Appendix A of the book Numerical Optimization by Jorge Nocedal Stephen J. Wright.



Рис. 1: Equivalent representations of a vector

Матрица называется симметричной(симметрической), если  $A = A^T$ . Это обозначается как  $A \in \mathbb{S}^n$  (множество симметричных матриц размера  $n \times n$ ). Из определения следует, что только квадратные матрицы могут быть симметричными.

Матрица называется симметричной(симметрической), если  $A = A^T$ . Это обозначается как  $A \in \mathbb{S}^n$  (множество симметричных матриц размера  $n \times n$ ). Из определения следует, что только квадратные матрицы могут быть симметричными.

Матрица  $A \in \mathbb{S}^n$  называется **положительно (отрицательно) определенной**, если для всех  $x \neq 0 : x^T A x > (<) 0$ . Мы обозначаем это как  $A \succ (<) 0$ . Множество таких матриц обозначается  $\mathbb{S}_{++}^n (\mathbb{S}_{--}^n)$ .

Матрица называется симметричной(симметрической), если  $A = A^T$ . Это обозначается как  $A \in \mathbb{S}^n$  (множество симметричных матриц размера  $n \times n$ ). Из определения следует, что только квадратные матрицы могут быть симметричными.

Матрица  $A \in \mathbb{S}^n$  называется **положительно (отрицательно) определенной**, если для всех  $x \neq 0 : x^T A x > (<)0$ . Мы обозначаем это как  $A \succ (<)0$ . Множество таких матриц обозначается  $\mathbb{S}_{++}^n (\mathbb{S}_{--}^n)$ .

Матрица  $A \in \mathbb{S}^n$  называется **положительно (отрицательно) полуопределенной**, если для всех  $x : x^T A x \geq (\leq)0$ . Мы обозначаем это как  $A \succeq (\preceq)0$ . Множество таких матриц обозначается  $\mathbb{S}_+^n (\mathbb{S}_-^n)$

### Question

Верно ли, что у положительно определенной матрицы все элементы положительны?



Матрица называется симметричной(симметрической), если  $A = A^T$ . Это обозначается как  $A \in \mathbb{S}^n$  (множество симметричных матриц размера  $n \times n$ ). Из определения следует, что только квадратные матрицы могут быть симметричными.

Матрица  $A \in \mathbb{S}^n$  называется **положительно (отрицательно) определенной**, если для всех  $x \neq 0 : x^T A x > (<)0$ . Мы обозначаем это как  $A \succ (<)0$ . Множество таких матриц обозначается  $\mathbb{S}_{++}^n (\mathbb{S}_{--}^n)$ .

Матрица  $A \in \mathbb{S}^n$  называется **положительно (отрицательно) полуопределенной**, если для всех  $x : x^T A x \geq (\leq)0$ . Мы обозначаем это как  $A \succeq (\preceq)0$ . Множество таких матриц обозначается  $\mathbb{S}_+^n (\mathbb{S}_-^n)$

### Question

Верно ли, что у положительно определенной матрицы все элементы положительны?

### Question

Верно ли, что симметричная матрица должна быть положительно определенной?

Матрица называется симметричной(симметрической), если  $A = A^T$ . Это обозначается как  $A \in \mathbb{S}^n$  (множество симметричных матриц размера  $n \times n$ ). Из определения следует, что только квадратные матрицы могут быть симметричными.

Матрица  $A \in \mathbb{S}^n$  называется **положительно (отрицательно) определенной**, если для всех  $x \neq 0 : x^T A x > (<)0$ . Мы обозначаем это как  $A \succ (<)0$ . Множество таких матриц обозначается  $\mathbb{S}_{++}^n (\mathbb{S}_{--}^n)$ .

Матрица  $A \in \mathbb{S}^n$  называется **положительно (отрицательно) полуопределенной**, если для всех  $x : x^T A x \geq (\leq)0$ . Мы обозначаем это как  $A \succeq (\preceq)0$ . Множество таких матриц обозначается  $\mathbb{S}_+^n (\mathbb{S}_-^n)$

### Question

Верно ли, что у положительно определенной матрицы все элементы положительны?

### Question

Верно ли, что симметричная матрица должна быть положительно определенной?

### Question

Верно ли, что положительно определенная матрица должна быть симметричной?

## Матричное умножение (matmul)

Пусть  $A$  - матрица размера  $m \times n$ ,  $B$  - матрица размера  $n \times p$ . Рассмотрим их произведение  $AB$ :

$$C = AB,$$

где  $C$  - матрица размера  $m \times p$  с элементами  $(i, j)$  заданными следующим образом:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}.$$

Наивная реализация матричного умножения требует  $\mathcal{O}(n^3)$  арифметических операций, где в качестве  $n$  берётся наибольший из размеров матриц.

### Question

Возможно ли перемножить две матрицы быстрее, чем за  $\mathcal{O}(n^3)$  операций? Что насчет  $\mathcal{O}(n^2)$ ,  $\mathcal{O}(n)$ ?

## Умножение матрицы на вектор (matvec)

Пусть  $A$  - матрица размера  $m \times n$ ,  $x$  - вектор размера  $n$  (или, что то же самое,  $x$  - матрица размера  $n \times 1$ ). Тогда  $i$ -я компонента произведения:

$$z = Ax$$

задается выражением:

$$z_i = \sum_{k=1}^n a_{ik} x_k$$

Наивная реализация матричного умножения требует  $\mathcal{O}(n^2)$  арифметических операций, где в качестве  $n$  берётся наибольший из размеров матриц.

Следует помнить, что:

- $C = AB \quad C^T = B^T A^T$

## Умножение матрицы на вектор (matvec)

Пусть  $A$  - матрица размера  $m \times n$ ,  $x$  - вектор размера  $n$  (или, что то же самое,  $x$  - матрица размера  $n \times 1$ ). Тогда  $i$ -я компонента произведения:

$$z = Ax$$

задается выражением:

$$z_i = \sum_{k=1}^n a_{ik} x_k$$

Наивная реализация матричного умножения требует  $\mathcal{O}(n^2)$  арифметических операций, где в качестве  $n$  берётся наибольший из размеров матриц.

Следует помнить, что:

- $C = AB \quad C^T = B^T A^T$
- $AB \neq BA$

## Умножение матрицы на вектор (matvec)

Пусть  $A$  - матрица размера  $m \times n$ ,  $x$  - вектор размера  $n$  (или, что то же самое,  $x$  - матрица размера  $n \times 1$ ). Тогда  $i$ -я компонента произведения:

$$z = Ax$$

задается выражением:

$$z_i = \sum_{k=1}^n a_{ik} x_k$$

Наивная реализация матричного умножения требует  $\mathcal{O}(n^2)$  арифметических операций, где в качестве  $n$  берётся наибольший из размеров матриц.

Следует помнить, что:

- $C = AB \quad C^T = B^T A^T$
- $AB \neq BA$
- $e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$

## Умножение матрицы на вектор (matvec)

Пусть  $A$  - матрица размера  $m \times n$ ,  $x$  - вектор размера  $n$  (или, что то же самое,  $x$  - матрица размера  $n \times 1$ ). Тогда  $i$ -я компонента произведения:

$$z = Ax$$

задается выражением:

$$z_i = \sum_{k=1}^n a_{ik} x_k$$

Наивная реализация матричного умножения требует  $\mathcal{O}(n^2)$  арифметических операций, где в качестве  $n$  берётся наибольший из размеров матриц.

Следует помнить, что:

- $C = AB \quad C^T = B^T A^T$
- $AB \neq BA$
- $e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$
- $e^{A+B} \neq e^A e^B$  (но если  $A$  и  $B$  - коммутирующие матрицы, то есть  $AB = BA$ , тогда  $e^{A+B} = e^A e^B$ )

## Умножение матрицы на вектор (matvec)

Пусть  $A$  - матрица размера  $m \times n$ ,  $x$  - вектор размера  $n$  (или, что то же самое,  $x$  - матрица размера  $n \times 1$ ). Тогда  $i$ -я компонента произведения:

$$z = Ax$$

задается выражением:

$$z_i = \sum_{k=1}^n a_{ik} x_k$$

Наивная реализация матричного умножения требует  $\mathcal{O}(n^2)$  арифметических операций, где в качестве  $n$  берётся наибольший из размеров матриц.

Следует помнить, что:

- $C = AB \quad C^T = B^T A^T$
- $AB \neq BA$
- $e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$
- $e^{A+B} \neq e^A e^B$  (но если  $A$  и  $B$  - коммутирующие матрицы, то есть  $AB = BA$ , тогда  $e^{A+B} = e^A e^B$ )
- $\langle x, Ay \rangle = \langle A^T x, y \rangle$



## Пример. Простой, но важный сюжет про матричное умножение

Предположим, имеется следующее выражение

$$b = A_1 A_2 A_3 x,$$

где  $A_1, A_2, A_3 \in \mathbb{R}^{3 \times 3}$  - какие-либо квадратные плотные (почти не содержащие нулевых элементов) матрицы и  $x \in \mathbb{R}^n$  - какой-то вектор. Необходимо вычислить  $b$ .

Каким образом лучше это сделать?

1.  $A_1 A_2 A_3 x$  (слева направо)

Проверьте ответ на  примере кода.

## Пример. Простой, но важный сюжет про матричное умножение


Предположим, имеется следующее выражение

$$b = A_1 A_2 A_3 x,$$

где  $A_1, A_2, A_3 \in \mathbb{R}^{3 \times 3}$  - какие-либо квадратные плотные (почти не содержащие нулевых элементов) матрицы и  $x \in \mathbb{R}^n$  - какой-то вектор. Необходимо вычислить  $b$ .

Каким образом лучше это сделать?

1.  $A_1 A_2 A_3 x$  (слева направо)
2.  $(A_1 (A_2 (A_3 x)))$  (справа налево)

Проверьте ответ на  примере кода.

## Пример. Простой, но важный сюжет про матричное умножение

Предположим, имеется следующее выражение

$$b = A_1 A_2 A_3 x,$$

где  $A_1, A_2, A_3 \in \mathbb{R}^{3 \times 3}$  - какие-либо квадратные плотные (почти не содержащие нулевых элементов) матрицы и  $x \in \mathbb{R}^n$  - какой-то вектор. Необходимо вычислить  $b$ .

Каким образом лучше это сделать?

1.  $A_1 A_2 A_3 x$  (слева направо)
2.  $(A_1 (A_2 (A_3 x)))$  (справа налево)
3. Без разницы

Проверьте ответ на  примере кода.

## Пример. Простой, но важный сюжет про матричное умножение

Предположим, имеется следующее выражение

$$b = A_1 A_2 A_3 x,$$

где  $A_1, A_2, A_3 \in \mathbb{R}^{3 \times 3}$  - какие-либо квадратные плотные (почти не содержащие нулевых элементов) матрицы и  $x \in \mathbb{R}^n$  - какой-то вектор. Необходимо вычислить  $b$ .

Каким образом лучше это сделать?

1.  $A_1 A_2 A_3 x$  (слева направо)
2.  $(A_1 (A_2 (A_3 x)))$  (справа налево)
3. Без разницы
4. Результаты, полученные первыми двумя предложенными способами, будут различаться.

Проверьте ответ на  примере кода.

# Нормы

Норма - это **качественная мера малости вектора**, обычно обозначаемая  $\|x\|$ .

Норма должна удовлетворять следующим свойствам:

1.  $\|\alpha x\| = |\alpha| \|x\|, \alpha \in \mathbb{R}$

# Нормы

Норма - это **качественная мера малости вектора**, обычно обозначаемая  $\|x\|$ .

Норма должна удовлетворять следующим свойствам:

1.  $\|\alpha x\| = |\alpha| \|x\|$ ,  $\alpha \in \mathbb{R}$
2.  $\|x + y\| \leq \|x\| + \|y\|$  (неравенство треугольника)

# Нормы

Норма - это **качественная мера малости вектора**, обычно обозначаемая  $\|x\|$ .

Норма должна удовлетворять следующим свойствам:

1.  $\|\alpha x\| = |\alpha| \|x\|$ ,  $\alpha \in \mathbb{R}$
2.  $\|x + y\| \leq \|x\| + \|y\|$  (неравенство треугольника)
3. Если  $\|x\| = 0$ , то  $x = 0$

# Нормы

Норма - это **качественная мера малости вектора**, обычно обозначаемая  $\|x\|$ .

Норма должна удовлетворять следующим свойствам:

1.  $\|\alpha x\| = |\alpha| \|x\|$ ,  $\alpha \in \mathbb{R}$
2.  $\|x + y\| \leq \|x\| + \|y\|$  (неравенство треугольника)
3. Если  $\|x\| = 0$ , то  $x = 0$



# Нормы

Норма - это **качественная мера малости вектора**, обычно обозначаемая  $\|x\|$ .

Норма должна удовлетворять следующим свойствам:

1.  $\|\alpha x\| = |\alpha| \|x\|$ ,  $\alpha \in \mathbb{R}$
2.  $\|x + y\| \leq \|x\| + \|y\|$  (неравенство треугольника)
3. Если  $\|x\| = 0$ , то  $x = 0$

Тогда расстояние между двумя векторами определяется как:

$$d(x, y) = \|x - y\|.$$

Наиболее известная и широко используемая норма - это **Евклидова норма**:

$$\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2},$$

которая соответствует расстоянию в нашей реальной жизни. Если векторы имеют комплексные компоненты, мы используем их модуль. Евклидова норма, или 2-норма, - подкласс важного класса  $p$ -норм:

$$\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}.$$

## $p$ -норма вектора

Имеется два очень важных особых случая. Бесконечная норма, или норма Чебышева, определяется как максимальный модуль компоненты вектора  $x$ :

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

## $p$ -норма вектора

Имеется два очень важных особых случая. Бесконечная норма, или норма Чебышева, определяется как максимальный модуль компоненты вектора  $x$ :

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

$L_1$  норма (или **Манхэттенское расстояние, расстояние городских кварталов**), которая определяется как сумма модулей элементов  $x$ :

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

## $p$ -норма вектора

Имеется два очень важных особых случая. Бесконечная норма, или норма Чебышева, определяется как максимальный модуль компоненты вектора  $x$ :

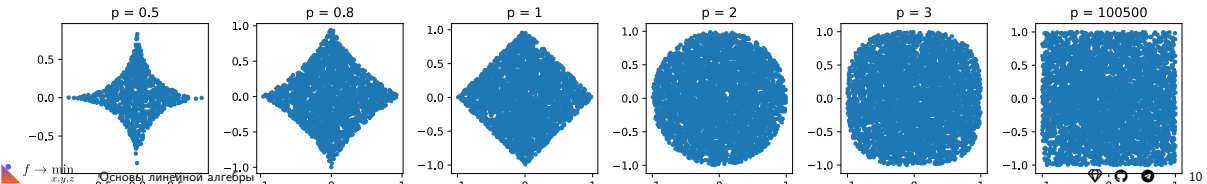
$$\|x\|_{\infty} = \max_i |x_i|$$

$L_1$  норма (или **Манхэттенское расстояние, расстояние городских кварталов**), которая определяется как сумма модулей элементов  $x$ :

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

$L_1$  норма играет очень важную роль: она относится к методам **compressed sensing**, которые появились в середине 00-х годов как одна из самых популярных исследовательских тем. Код для картинок снизу доступен [здесь](#):. Также посмотрите [это](#) видео.

Unit disk in the  $p$ -th norm



## Матричные нормы

В каком-то смысле нет сильных отличий между матрицами и векторами (вы можете векторизовать матрицу). Отсюда и получается простейшая матричная норма - **Фробениусова норма**:

$$\|A\|_F = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}$$

## Матричные нормы

В каком-то смысле нет сильных отличий между матрицами и векторами (вы можете векторизовать матрицу). Отсюда и получается простейшая матричная норма - **Фробениусова норма**:

$$\|A\|_F = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}$$

Спектральная норма,  $\|A\|_2$  является одной из наиболее используемых матричных норм (наряду с Фробениусовой нормой).

$$\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2},$$

Она не может быть посчитана напрямую через элементы с использованием какой-либо простой формулы, как, например, Фробениусова норма, однако, существуют эффективные алгоритмы для ее вычисления. Это напрямую связано с **сингулярным разложением** (SVD) матрицы. Из него следует

$$\|A\|_2 = \sigma_1(A) = \sqrt{\lambda_{\max}(A^T A)}$$

где  $\sigma_1(A)$  - наибольшее сингулярное число матрицы  $A$ .

# Скалярное произведение

Стандартное **скалярное произведение (inner product)** векторов  $x$  и  $y$  из  $\mathbb{R}^n$  задается как:

$$\langle x, y \rangle = x^T y = \sum_{i=1}^n x_i y_i = y^T x = \langle y, x \rangle$$

Здесь  $x_i$  и  $y_i$  - значения  $i$ -й компоненты соответствующего вектора.

## Example

Докажите, что можно переносить матрицу с транспонированием внутри скалярного произведения :

$$\langle x, Ay \rangle = \langle A^T x, y \rangle \text{ и } \langle x, yB \rangle = \langle xB^T, y \rangle$$

## Скалярное произведение матриц

Стандартное **скалярное произведение (inner product)** матриц  $X$  и  $Y$  из  $\mathbb{R}^{m \times n}$  задается как:

$$\langle X, Y \rangle = \text{tr}(X^T Y) = \sum_{i=1}^m \sum_{j=1}^n X_{ij} Y_{ij} = \text{tr}(Y^T X) = \langle Y, X \rangle$$

### Question

Существует ли какая-то связь между Фробениусовой нормой  $\|\cdot\|_F$  и скалярным произведением матриц  $\langle \cdot, \cdot \rangle$ ?



# Собственные числа и собственные вектора

Скаляр  $\lambda$  называется собственным числом матрицы  $A$  размера  $n \times n$ , если существует ненулевой вектор  $q$  такой, что

$$Aq = \lambda q.$$

вектор  $q$  называется собственным вектором матрицы  $A$ . Матрица  $A$  невырожденная, если все ее собственные числа отличны от нуля. Все собственные числа симметричной матрицы являются вещественными, в то время как у несимметричной матрицы могут быть комплексные собственные числа. Если матрица положительно определена, а значит, и симметрична, ее собственные числа - вещественные.

# Собственные числа и собственные вектора

## i Theorem

$$A \succeq (>)0 \Leftrightarrow \text{все собственные числа матрицы } A \geq (>)0$$

## i Proof

1.  $\rightarrow$  Предположим, что собственное число  $\lambda$  отрицательно и пусть  $x$  - соответствующий ему собственный вектор. Тогда

$$Ax = \lambda x \rightarrow x^T Ax = \lambda x^T x < 0$$

что противоречит условию  $A \succeq 0$ .

# Собственные числа и собственные вектора

## i Theorem

$A \succeq (>)0 \Leftrightarrow$  все собственные числа матрицы  $A \geq (>)0$

## i Proof

1.  $\rightarrow$  Предположим, что собственное число  $\lambda$  отрицательно и пусть  $x$  - соответствующий ему собственный вектор. Тогда

$$Ax = \lambda x \rightarrow x^T Ax = \lambda x^T x < 0$$

что противоречит условию  $A \succeq 0$ .

2.  $\leftarrow$  Для любой симметричной матрицы мы можем выбрать множество собственных векторов  $v_1, \dots, v_n$ , составляющих ортогональный базис в  $\mathbb{R}^n$ . Рассмотрим любой  $x \in \mathbb{R}^n$ .

$$\begin{aligned} x^T Ax &= (\alpha_1 v_1 + \dots + \alpha_n v_n)^T A(\alpha_1 v_1 + \dots + \alpha_n v_n) \\ &= \sum \alpha_i^2 v_i^T A v_i = \sum \alpha_i^2 \lambda_i v_i^T v_i \geq 0 \end{aligned}$$

здесь мы использовали тот факт, что  $v_i^T v_j = 0$  для  $i \neq j$ .

# Спектральное разложение матрицы

Предположим, что  $A \in S_n$ , то есть  $A$  вещественная симметричная матрица размера  $n \times n$ . Тогда  $A$  может быть разложена как

$$A = Q\Lambda Q^T,$$

---

<sup>2</sup>A good cheat sheet with matrix decomposition is available at the NLA course website.

# Спектральное разложение матрицы

Предположим, что  $A \in S_n$ , то есть  $A$  вещественная симметричная матрица размера  $n \times n$ . Тогда  $A$  может быть разложена как

$$A = Q\Lambda Q^T,$$

где  $Q \in \mathbb{R}^{n \times n}$  ортогональная, то есть удовлетворяет соотношению  $Q^T Q = I$ , и матрица  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ . (Вещественные) числа  $\lambda_i$  - собственные числа матрицы  $A$  и корни характеристического многочлена  $\det(A - \lambda I)$ . Столбцы  $Q$  образуют ортонормированный базис из собственных векторов  $A$ . Такая факторизация матрицы  $A$  называется спектральным разложением, или разложением матрицы на основе собственных векторов.<sup>2</sup>

---

<sup>2</sup>A good cheat sheet with matrix decomposition is available at the NLA course website.

# Спектральное разложение матрицы

Предположим, что  $A \in S_n$ , то есть  $A$  вещественная симметричная матрица размера  $n \times n$ . Тогда  $A$  может быть разложена как

$$A = Q\Lambda Q^T,$$

где  $Q \in \mathbb{R}^{n \times n}$  ортогональная, то есть удовлетворяет соотношению  $Q^T Q = I$ , и матрица  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ . (Вещественные) числа  $\lambda_i$  - собственные числа матрицы  $A$  и корни характеристического многочлена  $\det(A - \lambda I)$ . Столбцы  $Q$  образуют ортонормированный базис из собственных векторов  $A$ . Такая факторизация матрицы  $A$  называется спектральным разложением, или разложением матрицы на основе собственных векторов.<sup>2</sup>

Обычно собственные числа упорядочивают следующим образом:  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ . Будем использовать нотацию  $\lambda_i(A)$  для обозначения  $i$ -го по значению собственного числа матрицы  $A \in S$ . Обычно будем обозначать наибольшее собственное число как  $\lambda_1(A) = \lambda_{\max}(A)$ , а наименьшее как  $\lambda_n(A) = \lambda_{\min}(A)$ .

---

<sup>2</sup>A good cheat sheet with matrix decomposition is available at the NLA course website.

## Ещё о собственных значениях

Максимальное и минимальное собственное значение удовлетворяют соотношениям

$$\lambda_{\min}(A) = \inf_{x \neq 0} \frac{x^T A x}{x^T x}, \quad \lambda_{\max}(A) = \sup_{x \neq 0} \frac{x^T A x}{x^T x}$$

## Ещё о собственных значениях

Максимальное и минимальное собственное значение удовлетворяют соотношениям

$$\lambda_{\min}(A) = \inf_{x \neq 0} \frac{x^T A x}{x^T x}, \quad \lambda_{\max}(A) = \sup_{x \neq 0} \frac{x^T A x}{x^T x}$$

и, следовательно,  $\forall x \in \mathbb{R}^n$  (отношение Рэлея):

$$\lambda_{\min}(A)x^T x \leq x^T A x \leq \lambda_{\max}(A)x^T x$$



## Ещё о собственных значениях

Максимальное и минимальное собственное значение удовлетворяют соотношениям

$$\lambda_{\min}(A) = \inf_{x \neq 0} \frac{x^T A x}{x^T x}, \quad \lambda_{\max}(A) = \sup_{x \neq 0} \frac{x^T A x}{x^T x}$$

и, следовательно,  $\forall x \in \mathbb{R}^n$  (отношение Рэлея):

$$\lambda_{\min}(A)x^T x \leq x^T A x \leq \lambda_{\max}(A)x^T x$$

**Число обусловленности** невырожденной матрицы вводится следующим образом

$$\kappa(A) = \|A\| \|A^{-1}\|$$

## Ещё о собственных значениях

Максимальное и минимальное собственное значение удовлетворяют соотношениям

$$\lambda_{\min}(A) = \inf_{x \neq 0} \frac{x^T A x}{x^T x}, \quad \lambda_{\max}(A) = \sup_{x \neq 0} \frac{x^T A x}{x^T x}$$

и, следовательно,  $\forall x \in \mathbb{R}^n$  (отношение Рэля):

$$\lambda_{\min}(A) x^T x \leq x^T A x \leq \lambda_{\max}(A) x^T x$$

**Число обусловленности** невырожденной матрицы вводится следующим образом

$$\kappa(A) = \|A\| \|A^{-1}\|$$

Если мы используем спектральную норму, то можно получить:

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

Если, более того,  $A \in \mathbb{S}_{++}^n$ :  $\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$

## Сингулярное разложение

Предположим,  $A \in \mathbb{R}^{m \times n}$  и  $\text{rank } A = r$ . Тогда  $A$  может быть представлена как

$$A = U \Sigma V^T$$

## Сингулярное разложение

Предположим,  $A \in \mathbb{R}^{m \times n}$  и  $\text{rank } A = r$ . Тогда  $A$  может быть представлена как

$$A = U \Sigma V^T$$

где  $U \in \mathbb{R}^{m \times r}$  удовлетворяет  $U^T U = I$ ,  $V \in \mathbb{R}^{n \times r}$  удовлетворяет  $V^T V = I$  и  $\Sigma$  - диагональная матрица  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$  такая, что

## Сингулярное разложение

Предположим,  $A \in \mathbb{R}^{m \times n}$  и  $\text{rank } A = r$ . Тогда  $A$  может быть представлена как

$$A = U \Sigma V^T$$

где  $U \in \mathbb{R}^{m \times r}$  удовлетворяет  $U^T U = I$ ,  $V \in \mathbb{R}^{n \times r}$  удовлетворяет  $V^T V = I$  и  $\Sigma$  - диагональная матрица  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$  такая, что

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0.$$

## Сингулярное разложение

Предположим,  $A \in \mathbb{R}^{m \times n}$  и  $\text{rank } A = r$ . Тогда  $A$  может быть представлена как

$$A = U \Sigma V^T$$

где  $U \in \mathbb{R}^{m \times r}$  удовлетворяет  $U^T U = I$ ,  $V \in \mathbb{R}^{n \times r}$  удовлетворяет  $V^T V = I$  и  $\Sigma$  - диагональная матрица  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$  такая, что

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0.$$

Эта факторизация называется **сингулярным разложением (SVD)** матрицы  $A$ . Столбцы  $U$  называются левыми сингулярными векторами  $A$ , столбцы  $V$  - правыми сингулярными векторами, и  $\sigma_i$  являются сингулярными числами. Сингулярное разложение может быть записано как

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T,$$

где  $u_i \in \mathbb{R}^m$  - левые сингулярные векторы, а  $v_i \in \mathbb{R}^n$  - правые сингулярные векторы.

# Сингулярное разложение

## Question

Предположим, матрица  $A \in \mathbb{S}_{++}^n$ . Что можно сказать о связи собственных чисел с сингулярными числами?

# Сингулярное разложение

## i Question

Предположим, матрица  $A \in \mathbb{S}_{++}^n$ . Что можно сказать о связи собственных чисел с сингулярными числами?

## i Question

Как сингулярные числа матрицы связаны с собственными числами, главным образом, для симметричных матриц?



## Ранговое разложение (Skeleton)

Простое, но очень интересное разложение матрицы - ранговое разложение, которое может быть записано в двух формах:

$$A = UV^T \quad A = \hat{C}\hat{A}^{-1}\hat{R}$$

## Ранговое разложение (Skeleton)

Простое, но очень интересное разложение матрицы - ранговое разложение, которое может быть записано в двух формах:

$$A = UV^T \quad A = \hat{C}\hat{A}^{-1}\hat{R}$$

Последнее выражение основано на забавном факте: можно случайно выбрать  $r$  линейно независимых столбцов матрицы и любые  $r$  линейно независимых строк матрицы и только по ним восстановить исходную матрицу.

## Ранговое разложение (Skeleton)

Простое, но очень интересное разложение матрицы - ранговое разложение, которое может быть записано в двух формах:

$$A = UV^T \quad A = \hat{C}\hat{A}^{-1}\hat{R}$$

Последнее выражение основано на забавном факте: можно случайно выбрать  $r$  линейно независимых столбцов матрицы и любые  $r$  линейно независимых строк матрицы и только по ним восстановить исходную матрицу.

Случаи применения рангового разложения:

- Сокращение моделей, сжатие данных и ускорение вычислений в численных методах: заданная матрица с  $\text{rank } r$ , где  $r \ll n, m$  требует для хранения  $\mathcal{O}((n+m)r) \ll nm$  элементов.



Рис. 3: Illustration of Skeleton decomposition

## Ранговое разложение (Skeleton)

Простое, но очень интересное разложение матрицы - ранговое разложение, которое может быть записано в двух формах:

$$A = UV^T \quad A = \hat{C}\hat{A}^{-1}\hat{R}$$

Последнее выражение основано на забавном факте: можно случайно выбрать  $r$  линейно независимых столбцов матрицы и любые  $r$  линейно независимых строк матрицы и только по ним восстановить исходную матрицу.

Случаи применения рангового разложения:

- Сокращение моделей, сжатие данных и ускорение вычислений в численных методах: заданная матрица с  $\text{rank } r$ , где  $r \ll n, m$  требует для хранения  $\mathcal{O}((n+m)r) \ll nm$  элементов.
- Выделение признаков в машинном обучении, где это также известно как матричная факторизация



Рис. 3: Illustration of Skeleton decomposition

## Ранговое разложение (Skeleton)

Простое, но очень интересное разложение матрицы - ранговое разложение, которое может быть записано в двух формах:

$$A = UV^T \quad A = \hat{C}\hat{A}^{-1}\hat{R}$$

Последнее выражение основано на забавном факте: можно случайно выбрать  $r$  линейно независимых столбцов матрицы и любые  $r$  линейно независимых строк матрицы и только по ним восстановить исходную матрицу.

Случаи применения рангового разложения:

- Сокращение моделей, сжатие данных и ускорение вычислений в численных методах: заданная матрица с  $\text{rank } r$ , где  $r \ll n, m$  требует для хранения  $\mathcal{O}((n+m)r) \ll nm$  элементов.
- Выделение признаков в машинном обучении, где это также известно как матричная факторизация
- Все приложения, где применяется SVD, поскольку ранговое разложение может быть преобразовано в усеченную форму SVD.



Рис. 3: Illustration of Skeleton decomposition

## Каноническое тензорное разложение

Можно рассмотреть обобщение ранговой декомпозиции на структуры данных более высокого порядка, такие как тензоры, что подразумевает представление тензора в виде суммы  $r$  примитивных тензоров.

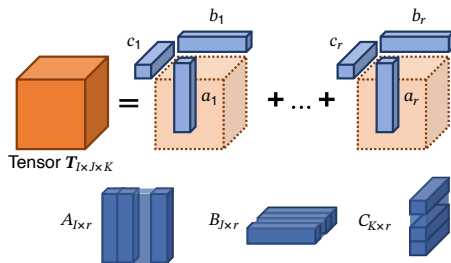


Рис. 4: Illustration of Canonical Polyadic decomposition

### **i** Example

Обратите внимание, что существует множество тензорных разложений: Canonical, Tucker, Tensor Train (TT), Tensor Ring (TR) и другие. В тензорном случае у нас нет прямого определения *ранга* для всех типов разложений. Например, для разложения TT ранг - это не скаляр, а вектор.

## Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные числа матрицы

$$\det A = \prod_{i=1}^n \lambda_i, \quad \operatorname{tr} A = \sum_{i=1}^n \lambda_i$$

Определитель обладает несколькими привлекательными (и показательными) свойствами. Например,

- $\det A = 0$  тогда и только тогда, когда  $A$  вырожденная;

## Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные числа матрицы

$$\det A = \prod_{i=1}^n \lambda_i, \quad \operatorname{tr} A = \sum_{i=1}^n \lambda_i$$

Определитель обладает несколькими привлекательными (и показательными) свойствами. Например,

- $\det A = 0$  тогда и только тогда, когда  $A$  вырожденная;
- $\det AB = (\det A)(\det B)$ ;



## Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные числа матрицы

$$\det A = \prod_{i=1}^n \lambda_i, \quad \operatorname{tr} A = \sum_{i=1}^n \lambda_i$$

Определитель обладает несколькими привлекательными (и показательными) свойствами. Например,

- $\det A = 0$  тогда и только тогда, когда  $A$  вырожденная;
- $\det AB = (\det A)(\det B)$ ;
- $\det A^{-1} = \frac{1}{\det A}$ .

## Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные числа матрицы

$$\det A = \prod_{i=1}^n \lambda_i, \quad \operatorname{tr} A = \sum_{i=1}^n \lambda_i$$

Определитель обладает несколькими привлекательными (и показательными) свойствами. Например,

- $\det A = 0$  тогда и только тогда, когда  $A$  вырожденная;
- $\det AB = (\det A)(\det B)$ ;
- $\det A^{-1} = \frac{1}{\det A}$ .

## Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные числа матрицы

$$\det A = \prod_{i=1}^n \lambda_i, \quad \operatorname{tr} A = \sum_{i=1}^n \lambda_i$$

Определитель обладает несколькими привлекательными (и показательными) свойствами. Например,

- $\det A = 0$  тогда и только тогда, когда  $A$  вырожденная;
- $\det AB = (\det A)(\det B)$ ;
- $\det A^{-1} = \frac{1}{\det A}$ .

Не забывайте о циклическом свойстве следа для произвольных матриц  $A, B, C, D$  (при условии, что все размерности согласованы):

$$\operatorname{tr}(ABCD) = \operatorname{tr}(DABC) = \operatorname{tr}(CDAB) = \operatorname{tr}(BCDA)$$

## Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные числа матрицы

$$\det A = \prod_{i=1}^n \lambda_i, \quad \operatorname{tr} A = \sum_{i=1}^n \lambda_i$$

Определитель обладает несколькими привлекательными (и показательными) свойствами. Например,

- $\det A = 0$  тогда и только тогда, когда  $A$  вырожденная;
- $\det AB = (\det A)(\det B)$ ;
- $\det A^{-1} = \frac{1}{\det A}$ .

Не забывайте о циклическом свойстве следа для произвольных матриц  $A, B, C, D$  (при условии, что все размерности согласованы):

$$\operatorname{tr}(ABCD) = \operatorname{tr}(DABC) = \operatorname{tr}(CDAB) = \operatorname{tr}(BCDA)$$

### Question

Как определитель матрицы связан с ее обратимостью?

# Аппроксимация Тейлора первого порядка

Аппроксимация Тейлора первого порядка, также известная как линейная аппроксимация, рассматривается вокруг некоторой точки  $x_0$ . Если  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  - дифференцируемая функция, то ее аппроксимация Тейлора первого порядка задается:

$$f_{x_0}^I(x) = f(x_0) + \nabla f(x_0)^T(x - x_0)$$

Где:

- $f(x_0)$  - значение функции в точке  $x_0$ .

# Аппроксимация Тейлора первого порядка

Аппроксимация Тейлора первого порядка, также известная как линейная аппроксимация, рассматривается вокруг некоторой точки  $x_0$ . Если  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  - дифференцируемая функция, то ее аппроксимация Тейлора первого порядка задается:

$$f_{x_0}^I(x) = f(x_0) + \nabla f(x_0)^T(x - x_0)$$

Где:

- $f(x_0)$  - значение функции в точке  $x_0$ .
- $\nabla f(x_0)$  - градиент функции в точке  $x_0$ .

# Аппроксимация Тейлора первого порядка

Аппроксимация Тейлора первого порядка, также известная как линейная аппроксимация, рассматривается вокруг некоторой точки  $x_0$ . Если  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  - дифференцируемая функция, то ее аппроксимация Тейлора первого порядка задается:

$$f_{x_0}^I(x) = f(x_0) + \nabla f(x_0)^T(x - x_0)$$

Где:

- $f(x_0)$  - значение функции в точке  $x_0$ .
- $\nabla f(x_0)$  - градиент функции в точке  $x_0$ .

# Аппроксимация Тейлора первого порядка

Аппроксимация Тейлора первого порядка, также известная как линейная аппроксимация, рассматривается вокруг некоторой точки  $x_0$ . Если  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  - дифференцируемая функция, то ее аппроксимация Тейлора первого порядка задается:

$$f_{x_0}^I(x) = f(x_0) + \nabla f(x_0)^T(x - x_0)$$

Где:

- $f(x_0)$  - значение функции в точке  $x_0$ .
- $\nabla f(x_0)$  - градиент функции в точке  $x_0$ .

Для простоты анализа некоторых подходов часто удобно заменить  $f(x)$  на  $f_{x_0}^I(x)$  вблизи точки  $x_0$ .



Рис. 5: First order Taylor approximation near the point  $x_0$



## Аппроксимация Тейлора второго порядка

Аппроксимация Тейлора второго порядка, также известная как квадратичная аппроксимация, учитывает кривизну функции. Для дважды дифференцируемой функции  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , ее аппроксимация Тейлора второго порядка в окрестности некоторой точки  $x_0$  имеет вид:

$$f_{x_0}^{II}(x) = f(x_0) + \nabla f(x_0)^T(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

Где  $\nabla^2 f(x_0)$  - Гессиан функции  $f$  в точке  $x_0$ .

## Аппроксимация Тейлора второго порядка

Аппроксимация Тейлора второго порядка, также известная как квадратичная аппроксимация, учитывает кривизну функции. Для дважды дифференцируемой функции  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , ее аппроксимация Тейлора второго порядка в окрестности некоторой точки  $x_0$  имеет вид:

$$f_{x_0}^{II}(x) = f(x_0) + \nabla f(x_0)^T(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

Где  $\nabla^2 f(x_0)$  - Гессиан функции  $f$  в точке  $x_0$ .

Когда использование линейного приближения функции недостаточно, можно рассмотреть замену  $f(x)$  на  $f_{x_0}^{II}(x)$  вблизи точки  $x_0$ . В общем случае аппроксимации Тейлора дают нам возможность локально аппроксимировать функции. Аппроксимация первого порядка представляет собой плоскость, касательную к функции в точке  $x_0$ , а аппроксимация второго порядка включает кривизну и представлена параболой. Эти аппроксимации особенно полезны в оптимизации и численных методах, так как обеспечивают удобный способ работы со сложными функциями.



Рис. 6: Second order Taylor approximation near the point  $x_0$

# Матричное дифференцирование

# Градиент

Пусть  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ . Тогда вектор, содержащий все частные производные первого порядка:

$$\nabla f(x) = \frac{df}{dx} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

# Градиент

Пусть  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ . Тогда вектор, содержащий все частные производные первого порядка:

$$\nabla f(x) = \frac{df}{dx} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

называется градиентом функции  $f(x)$ . Этот вектор указывает направление самого крутого подъема. Таким образом, вектор  $-\nabla f(x)$  соответствует направлению наикрутейшего спуска функции в точке. Более того, вектор градиента всегда ортогонален линии уровня (изолинии) в точке.

## i Example

Для функции  $f(x, y) = x^2 + y^2$  градиент равен:

$$\nabla f(x, y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

Этот вектор указывает направление наибольшего роста функции в точке.

## i Question

Как величина градиента связана с крутизной функции?

## Гессиан

Пусть  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ . Тогда матрица, содержащая все частные производные второго порядка:

$$f''(x) = \nabla^2 f(x) = \frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

## Гессиан

Пусть  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ . Тогда матрица, содержащая все частные производные второго порядка:

$$f''(x) = \nabla^2 f(x) = \frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

На самом деле, Гессиан может быть тензором в таком случае:  
( $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ) - это просто 3d-тензор, где каждый срез - это Гессиан соответствующей скалярной функции  
( $\nabla^2 f_1(x), \dots, \nabla^2 f_m(x)$ ).

### i Example

Для функции  $f(x, y) = x^2 + y^2$   
Гессиан:

$$H_f(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Эта матрица содержит информацию о кривизне функции по разным направлениям.

### i Question

Как Гессиан может быть использован для определения вогнутости или выпуклости функции?

## Теорема Шварца

Пусть функция  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Если смешанные частные производные  $\frac{\partial^2 f}{\partial x_i \partial x_j}$  и  $\frac{\partial^2 f}{\partial x_j \partial x_i}$  непрерывны на открытом множестве, содержащем точку  $a$ , то они равны в точке  $a$ . Таким образом,

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



## Теорема Шварца

Пусть функция  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Если смешанные частные производные  $\frac{\partial^2 f}{\partial x_i \partial x_j}$  и  $\frac{\partial^2 f}{\partial x_j \partial x_i}$  непрерывны на открытом множестве, содержащем точку  $a$ , то они равны в точке  $a$ . Таким образом,

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

По теореме Шварца, если смешанные частицы непрерывны на открытом множестве, то Гессиан симметричен. Это означает, что элементы над главной диагональю равны элементам, которые зеркально симметричны относительно главной диагонали:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \quad \nabla^2 f(x) = (\nabla^2 f(x))^T$$

Эта симметрия упрощает вычисления и анализ с использованием Гессиана в различных приложениях, особенно в оптимизации.

### i Контрпример Шварца

$$f(x, y) = \begin{cases} \frac{xy(x^2 - y^2)}{x^2 + y^2} & \text{для } (x, y) \neq (0, 0), \\ 0 & \text{для } (x, y) = (0, 0). \end{cases}$$

### Counterexample ♣



Можно убедиться, что  $\frac{\partial^2 f}{\partial x \partial y}(0, 0) \neq \frac{\partial^2 f}{\partial y \partial x}(0, 0)$ , хотя смешанные частные производные существуют, и в любой другой точке симметрия сохраняется.

# Якобиан

Расширение понятия градиента многомерной функции  
 $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  - это следующая матрица:

$$J_f = f'(x) = \frac{df}{dx^T} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Эта матрица предоставляет информацию о скорости изменения функции по отношению к ее входам.

## i Question

Можем ли мы как-то связать эти три определения (градиент, Якобиан и Гессиан) одним корректным утверждением?

## i Example

Для функции

$$f(x, y) = \begin{bmatrix} x + y \\ x - y \end{bmatrix},$$

Якобиан равен:

$$J_f(x, y) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

## i Question

Как матрица Якоби связана с градиентом для скалярно-значных функций?

$$f(x) : X \rightarrow Y; \quad \frac{\partial f(x)}{\partial x} \in G$$

| X                         | Y              | G                         | Name  |
|---------------------------|----------------|---------------------------|---|
| $\mathbb{R}$              | $\mathbb{R}$   | $\mathbb{R}$              | $f'(x)$ (производная)                         |
| $\mathbb{R}^n$            | $\mathbb{R}$   | $\mathbb{R}^n$            | $\frac{\partial f}{\partial x_i}$ (градиент)  |
| $\mathbb{R}^n$            | $\mathbb{R}^m$ | $\mathbb{R}^{n \times m}$ | $\frac{\partial f_i}{\partial x_j}$ (Якобиан) |
| $\mathbb{R}^{m \times n}$ | $\mathbb{R}$   | $\mathbb{R}^{m \times n}$ | $\frac{\partial f}{\partial x_{ij}}$          |

## i Theorem

Пусть  $x \in S$  - внутренняя точка множества  $S$ , и пусть  $D : U \rightarrow V$  - линейный оператор. Мы говорим, что функция  $f$  дифференцируема в точке  $x$  с производной  $D$ , если для всех достаточно малых  $h \in U$  имеет место следующее разложение:

$$f(x + h) = f(x) + D[h] + o(\|h\|)$$

Если для любого линейного оператора  $D : U \rightarrow V$  функция  $f$  не дифференцируема в точке  $x$  с производной  $D$ , то мы говорим, что  $f$  не дифференцируема в точке  $x$ .

Введя дифференциальное обозначение  $df$ , мы можем получить градиент по следующей формуле:

$$df(x) = \langle \nabla f(x), dx \rangle$$

Введя дифференциальное обозначение  $df$ , мы можем получить градиент по следующей формуле:

$$df(x) = \langle \nabla f(x), dx \rangle$$

Тогда, если у нас есть дифференциал указанной выше формы и нам нужно вычислить вторую производную матрицы/векторной функции, мы рассматриваем «старый»  $dx$  как константу  $dx_1$ , а затем вычисляем  $d(df) = d^2 f(x)$ .

$$d^2 f(x) = \langle \nabla^2 f(x) dx_1, dx \rangle = \langle H_f(x) dx_1, dx \rangle$$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$



# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$
- $d(X^T) = (dX)^T$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$
- $d(X^T) = (dX)^T$
- $d(XY) = (dX)Y + X(dY)$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$
- $d(X^T) = (dX)^T$
- $d(XY) = (dX)Y + X(dY)$
- $d\langle X, Y \rangle = \langle dX, Y \rangle + \langle X, dY \rangle$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$
- $d(X^T) = (dX)^T$
- $d(XY) = (dX)Y + X(dY)$
- $d\langle X, Y \rangle = \langle dX, Y \rangle + \langle X, dY \rangle$
- $d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$
- $d(X^T) = (dX)^T$
- $d(XY) = (dX)Y + X(dY)$
- $d\langle X, Y \rangle = \langle dX, Y \rangle + \langle X, dY \rangle$
- $d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$
- $d(\det X) = \det X \langle X^{-T}, dX \rangle$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$
- $d(X^T) = (dX)^T$
- $d(XY) = (dX)Y + X(dY)$
- $d\langle X, Y \rangle = \langle dX, Y \rangle + \langle X, dY \rangle$
- $d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$
- $d(\det X) = \det X \langle X^{-T}, dX \rangle$
- $d(\operatorname{tr} X) = \langle I, dX \rangle$



# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$
- $d(X^T) = (dX)^T$
- $d(XY) = (dX)Y + X(dY)$
- $d\langle X, Y \rangle = \langle dX, Y \rangle + \langle X, dY \rangle$
- $d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$
- $d(\det X) = \det X \langle X^{-T}, dX \rangle$
- $d(\operatorname{tr} X) = \langle I, dX \rangle$
- $df(g(x)) = \frac{df}{dg} \cdot dg(x)$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$
- $d(X^T) = (dX)^T$
- $d(XY) = (dX)Y + X(dY)$
- $d\langle X, Y \rangle = \langle dX, Y \rangle + \langle X, dY \rangle$
- $d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$
- $d(\det X) = \det X \langle X^{-T}, dX \rangle$
- $d(\operatorname{tr} X) = \langle I, dX \rangle$
- $df(g(x)) = \frac{df}{dg} \cdot dg(x)$
- $H = (J(\nabla f))^T$

# Свойства дифференциалов

Пусть  $A$  и  $B$  - постоянные матрицы, в то время как  $X$  и  $Y$  - переменные (или матричные функции).

- $dA = 0$
- $d(\alpha X) = \alpha(dX)$
- $d(AXB) = A(dX)B$
- $d(X + Y) = dX + dY$
- $d(X^T) = (dX)^T$
- $d(XY) = (dX)Y + X(dY)$
- $d\langle X, Y \rangle = \langle dX, Y \rangle + \langle X, dY \rangle$
- $d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$
- $d(\det X) = \det X \langle X^{-T}, dX \rangle$
- $d(\operatorname{tr} X) = \langle I, dX \rangle$
- $df(g(x)) = \frac{df}{dg} \cdot dg(x)$
- $H = (J(\nabla f))^T$
- $d(X^{-1}) = -X^{-1}(dX)X^{-1}$

# Матричное дифференцирование. Пример 1

## Example

Найдите  $df, \nabla f(x)$ , if  $f(x) = \langle x, Ax \rangle - b^T x + c$ .

## Матричное дифференцирование. Пример 2

### Example

Найдите  $df, \nabla f(x)$ , if  $f(x) = \ln \langle x, Ax \rangle$ .

1. Необходимо, чтобы  $A$  была положительно определенной, потому что она стоит в показателе логарифма. Значит,  $A \in \mathbb{S}_{++}^n$ . Найдем дифференциал:

$$\begin{aligned} df &= d(\ln \langle x, Ax \rangle) = \frac{d(\langle x, Ax \rangle)}{\langle x, Ax \rangle} = \frac{\langle dx, Ax \rangle + \langle x, d(Ax) \rangle}{\langle x, Ax \rangle} = \\ &= \frac{\langle Ax, dx \rangle + \langle x, Adx \rangle}{\langle x, Ax \rangle} = \frac{\langle Ax, dx \rangle + \langle A^T x, dx \rangle}{\langle x, Ax \rangle} = \frac{\langle (A + A^T)x, dx \rangle}{\langle x, Ax \rangle} \end{aligned}$$

## Матричное дифференцирование. Пример 2

### i Example

Найдите  $df, \nabla f(x)$ , if  $f(x) = \ln \langle x, Ax \rangle$ .

1. Необходимо, чтобы  $A$  была положительно определенной, потому что она стоит в показателе логарифма. Значит,  $A \in \mathbb{S}_{++}^n$ . Найдём дифференциал:

$$\begin{aligned} df &= d(\ln \langle x, Ax \rangle) = \frac{d(\langle x, Ax \rangle)}{\langle x, Ax \rangle} = \frac{\langle dx, Ax \rangle + \langle x, d(Ax) \rangle}{\langle x, Ax \rangle} = \\ &= \frac{\langle Ax, dx \rangle + \langle x, Adx \rangle}{\langle x, Ax \rangle} = \frac{\langle Ax, dx \rangle + \langle A^T x, dx \rangle}{\langle x, Ax \rangle} = \frac{\langle (A + A^T)x, dx \rangle}{\langle x, Ax \rangle} \end{aligned}$$

2. Заметим, что наша главная цель - получить формулу вида  $df = \langle \cdot, dx \rangle$

$$df = \left\langle \frac{2Ax}{\langle x, Ax \rangle}, dx \right\rangle$$

Следовательно, градиент равен:  $\nabla f(x) = \frac{2Ax}{\langle x, Ax \rangle}$

## Матричное дифференцирование. Пример 3

### Example

Найдите  $df, \nabla f(X)$ , если  $f(X) = \langle S, X \rangle - \log \det X$ .