A Corgi dog and a yellow rubber duck are positioned in front of a chalkboard. The chalkboard features a graph titled 'GRADIENT FUNCTION' with a colorful, multi-peaked surface. Various mathematical notations and equations are written on the board, including 'Nabla', 'f → min', and several arithmetic problems. Two clapperboards are visible on the left side of the frame.

Вспоминаем линейную алгебру.

Даниил Меркулов

Методы оптимизации. МФТИ

Вспоминаем линейную алгебру

Векторы и матрицы

Мы будем считать, что все векторы являются столбцами по умолчанию. Пространство векторов длины n обозначается \mathbb{R}^n , а пространство матриц размера $m \times n$ с вещественными элементами обозначается $\mathbb{R}^{m \times n}$. То есть ¹:

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

¹Подробный вводный курс по прикладной линейной алгебре можно найти в книге Introduction to Applied Linear Algebra – Vectors, Matrices, and Least Squares - книга от Stephen Boyd & Lieven Vandenberghe, которая указана в источнике. Также полезен материал по линейной алгебре приведенный в приложении А книги Numerical Optimization by Jorge Nocedal Stephen J. Wright.

Векторы и матрицы

Мы будем считать, что все векторы являются столбцами по умолчанию. Пространство векторов длины n обозначается \mathbb{R}^n , а пространство матриц размера $m \times n$ с вещественными элементами обозначается $\mathbb{R}^{m \times n}$. То есть ¹:

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

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

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

Мы будем писать $x \geq 0$ и $x \neq 0$ для обозначения покомпонентных неравенств

¹Подробный вводный курс по прикладной линейной алгебре можно найти в книге Introduction to Applied Linear Algebra – Vectors, Matrices, and Least Squares - книга от Stephen Boyd & Lieven Vandenberghe, которая указана в источнике. Также полезен материал по линейной алгебре приведенный в приложении А книги Numerical Optimization by Jorge Nocedal Stephen J. Wright.



Рис. 1: Эквивалентные представления вектора

Матрица A называется симметричной, если $A = A^T$. Обозначается как $A \in \mathbb{S}^n$ (множество квадратных симметричных матриц размерности n). Заметим, что только квадратная матрица может быть симметричной по определению.

Матрица A называется симметричной, если $A = A^T$. Обозначается как $A \in \mathbb{S}^n$ (множество квадратных симметричных матриц размерности n). Заметим, что только квадратная матрица может быть симметричной по определению.

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

Матрица A называется симметричной, если $A = A^T$. Обозначается как $A \in \mathbb{S}^n$ (множество квадратных симметричных матриц размерности n). Заметим, что только квадратная матрица может быть симметричной по определению.

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

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

Question

Верно ли, что положительно определенная матрица имеет все положительные элементы?

Матрица A называется симметричной, если $A = A^T$. Обозначается как $A \in \mathbb{S}^n$ (множество квадратных симметричных матриц размерности n). Заметим, что только квадратная матрица может быть симметричной по определению.

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

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

Question

Верно ли, что положительно определенная матрица имеет все положительные элементы?

Question

Верно ли, что если матрица симметрична, то она должна быть положительно определенной?

Матрица A называется симметричной, если $A = A^T$. Обозначается как $A \in \mathbb{S}^n$ (множество квадратных симметричных матриц размерности n). Заметим, что только квадратная матрица может быть симметричной по определению.

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

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

Question

Верно ли, что положительно определенная матрица имеет все положительные элементы?

Question

Верно ли, что если матрица симметрична, то она должна быть положительно определенной?

Question

Верно ли, что если матрица положительно определена, то она должна быть симметричной?

Матричное умножение (matmul)

Пусть A - матрица размера $m \times n$, а B - матрица размера $n \times p$, тогда их произведение AB равно:

$$C = AB$$

Тогда C - матрица размера $m \times p$, элемент (i, j) которой равен:

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

Эта операция в наивной форме требует $\mathcal{O}(n^3)$ арифметических операций, где n обычно считается наибольшей размерностью матриц.

Матричное умножение (matmul)

Пусть A - матрица размера $m \times n$, а B - матрица размера $n \times p$, тогда их произведение AB равно:

$$C = AB$$

Тогда C - матрица размера $m \times p$, элемент (i, j) которой равен:

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

Эта операция в наивной форме требует $\mathcal{O}(n^3)$ арифметических операций, где n обычно считается наибольшей размерностью матриц.

Question

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

Умножение матрицы на вектор (matvec)

Пусть A - матрица размера $m \times n$, а x - вектор длины n , тогда i -й элемент произведения Ax равен:

$$z = Ax$$

равен:

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

Эта операция в наивной форме требует $\mathcal{O}(n^2)$ арифметических операций, где n обычно считается наибольшей размерностью входов.

Отметим, что:

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

Умножение матрицы на вектор (matvec)

Пусть A - матрица размера $m \times n$, а x - вектор длины n , тогда i -й элемент произведения Ax равен:

$$z = Ax$$

равен:

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

Эта операция в наивной форме требует $\mathcal{O}(n^2)$ арифметических операций, где n обычно считается наибольшей размерностью входов.

Отметим, что:

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

Умножение матрицы на вектор (matvec)

Пусть A - матрица размера $m \times n$, а x - вектор длины n , тогда i -й элемент произведения Ax равен:

$$z = Ax$$

равен:

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

Эта операция в наивной форме требует $\mathcal{O}(n^2)$ арифметических операций, где n обычно считается наибольшей размерностью входов.

Отметим, 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 $m \times n$, а x - вектор длины n , тогда i -й элемент произведения Ax равен:

$$z = Ax$$

равен:

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

Эта операция в наивной форме требует $\mathcal{O}(n^2)$ арифметических операций, где n обычно считается наибольшей размерностью входов.

Отметим, 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 коммутируют, то есть $AB = BA$, то $e^{A+B} = e^A e^B$)

Умножение матрицы на вектор (matvec)

Пусть A - матрица размера $m \times n$, а x - вектор длины n , тогда i -й элемент произведения Ax равен:

$$z = Ax$$

равен:

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

Эта операция в наивной форме требует $\mathcal{O}(n^2)$ арифметических операций, где n обычно считается наибольшей размерностью входов.

Отметим, 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 коммутируют, то есть $AB = BA$, то $e^{A+B} = e^A e^B$)
- $\langle x, Ay \rangle = \langle A^T x, y \rangle$

Пример. Простая, но важная идея о матричных вычислениях.

Предположим, у вас есть следующее выражение

$$b = A_1 A_2 A_3 x,$$

где $A_1, A_2, A_3 \in \mathbb{R}^{3 \times 3}$ - случайные квадратные плотные матрицы, и $x \in \mathbb{R}^n$ - вектор. Вам нужно вычислить b .

Какой способ лучше всего использовать?

1. $A_1 A_2 A_3 x$ (слева направо)

Проверьте простой  код после вашего интуитивного ответа.

Пример. Простая, но важная идея о матричных вычислениях.

Предположим, у вас есть следующее выражение

$$b = A_1 A_2 A_3 x,$$

где $A_1, A_2, A_3 \in \mathbb{R}^{3 \times 3}$ - случайные квадратные плотные матрицы, и $x \in \mathbb{R}^n$ - вектор. Вам нужно вычислить b .

Какой способ лучше всего использовать?

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

Проверьте простой  код после вашего интуитивного ответа.

Пример. Простая, но важная идея о матричных вычислениях.

Предположим, у вас есть следующее выражение

$$b = A_1 A_2 A_3 x,$$

где $A_1, A_2, A_3 \in \mathbb{R}^{3 \times 3}$ - случайные квадратные плотные матрицы, и $x \in \mathbb{R}^n$ - вектор. Вам нужно вычислить b .

Какой способ лучше всего использовать?

1. $A_1 A_2 A_3 x$ (слева направо)
2. $(A_1 (A_2 (A_3 x)))$ (справа налево)
3. Не имеет значения

Проверьте простой  код после вашего интуитивного ответа.

Пример. Простая, но важная идея о матричных вычислениях.

Предположим, у вас есть следующее выражение

$$b = A_1 A_2 A_3 x,$$

где $A_1, A_2, A_3 \in \mathbb{R}^{3 \times 3}$ - случайные квадратные плотные матрицы, и $x \in \mathbb{R}^n$ - вектор. Вам нужно вычислить b .

Какой способ лучше всего использовать?

1. $A_1 A_2 A_3 x$ (слева направо)
2. $(A_1 (A_2 (A_3 x)))$ (справа налево)
3. Не имеет значения
4. Результаты первых двух вариантов не будут одинаковыми.

Проверьте простой  код после вашего интуитивного ответа.

Нормы

Норма - это **количественная мера малости вектора** и обычно обозначается как $\|x\|$.

Норма должна удовлетворять определенным свойствам:

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

Нормы

Норма - это **количественная мера малости вектора** и обычно обозначается как $\|x\|$.

Норма должна удовлетворять определенным свойствам:

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

Нормы

Норма - это **количественная мера малости вектора** и обычно обозначается как $\|x\|$.

Норма должна удовлетворять определенным 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 **количественная мера малости вектора** и обычно обозначается как $\|x\|$.

Норма должна удовлетворять определенным 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 **количественная мера малости вектора** и обычно обозначается как $\|x\|$.

Норма должна удовлетворять определенным свойствам:

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

Расстояние между двумя векторами определяется как

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

Наиболее широко используемой нормой является **Евклидова норма**:

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

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

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

p -норма вектора

Существуют два очень важных частных случая. Бесконечность-норма, или норма Чебышева, определяется как максимальное абсолютное значение элемента вектора:

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

p -норма вектора

Существуют два очень важных частных случая. Бесконечность-норма, или норма Чебышева, определяется как максимальное абсолютное значение элемента вектора:

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

l_1 норма (или **манхэттенское расстояние**) определяется как сумма модулей элементов вектора x :

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

p -норма вектора

Существуют два очень важных частных случая. Бесконечность-норма, или норма Чебышева, определяется как максимальное абсолютное значение элемента вектора:

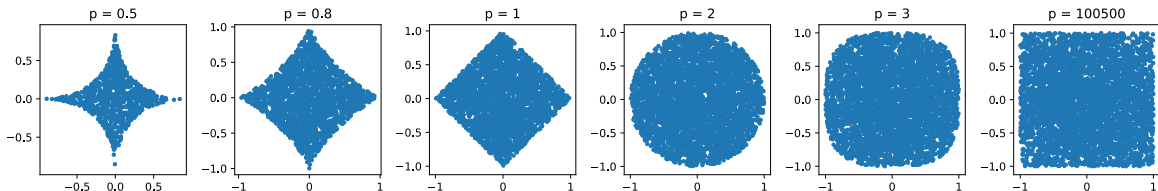
$$\|x\|_{\infty} = \max_i |x_i|$$

l_1 норма (или **манхэттенское расстояние**) определяется как сумма модулей элементов вектора x :

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

l_1 норма играет очень важную роль: она все связана с методами **compressed sensing**, которые появились в середине 00-х как одна из популярных тем исследований. Код для изображения ниже доступен [здесь](#):. Также посмотрите [это](#) видео.

Unit disk in the p -th norm



Матричные нормы

В некотором смысле между матрицами и векторами нет большой разницы (вы можете векторизовать матрицу), и здесь появляется самая простая матричная норма **Фробениуса**:

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

Матричные нормы

В некотором смысле между матрицами и векторами нет большой разницы (вы можете векторизовать матрицу), и здесь появляется самая простая матричная норма **Фробениуса**:

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

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

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

Она не может быть вычислена непосредственно из элементов с помощью простой формулы, как в случае нормы Фробениуса, однако, существуют эффективные алгоритмы для ее вычисления. Она напрямую связана с **сингулярным разложением** (SVD) матрицы. Для неё справедливо:

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

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

Скалярное произведение

Стандартное **скалярное произведение** между векторами x и y из \mathbb{R}^n равно:

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

Здесь x_i и y_i - i -ые компоненты соответствующих векторов.

Example

Докажите, что вы можете переставить матрицу внутри скалярного произведения с транспонированием:

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

Скалярное произведение матриц

Стандартное **скалярное произведение** между матрицами X и Y из $\mathbb{R}^{m \times n}$ равно:

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

Question

Существует ли связь между нормой Фробениуса $\|\cdot\|_F$ и скалярным произведением между матрицами $\langle \cdot, \cdot \rangle$?

Собственные вектора и собственные значения

Число λ является собственным значением квадратной 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 ни одно из её собственных значений не равно нулю. Собственные значения симметричных матриц являются вещественными числами, в то время как несимметричные матрицы могут иметь комплексные собственные значения. Если матрица положительно определена и симметрична, то все её собственные значения являются положительными вещественными числами.

Собственные вектора и собственные значения

i Theorem

$$A \succeq (\succ) 0 \Leftrightarrow \text{все собственные значения } A \geq (>) 0$$

Proof

1. \rightarrow Предположим, что некоторое собственное значение λ отрицательно, и пусть x обозначает соответствующий собственный вектор. Тогда

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

что противоречит условию $A \succeq 0$.

Собственные вектора и собственные значения

i Theorem

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

Proof

1. \rightarrow Предположим, что некоторое собственное значение λ отрицательно, и пусть x обозначает соответствующий 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 набор собственных векторов v_1, \dots, v_n , которые образуют ортонормированный базис в \mathbb{R}^n . Возьмем любой вектор $x \in \mathbb{R}^n$.

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

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

Спектральное разложение (eigendecomposition)

Пусть $A \in S_n$, т.е. A - вещественная симметричная матрица размера $n \times n$. Тогда A может быть разложена как

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

²Хорошая шпаргалка с разложением матриц доступна на сайте курса по линейной алгебре [website](#).

Спектральное разложение (eigendecomposition)

Пусть $A \in S_n$, т.е. 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, т.е. удовлетворяет $Q^T Q = I$, и $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. Вещественные числа λ_i являются собственными значениями A и являются корнями характеристического полинома $\det(A - \lambda I)$. Столбцы Q образуют ортонормированный набор собственных векторов A . Такое разложение называется спектральным.²

²Хорошая шпаргалка с разложением матриц доступна на сайте курса по линейной алгебре [website](#).

Спектральное разложение (eigendecomposition)

Пусть $A \in S_n$, т.е. 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, т.е. удовлетворяет $Q^T Q = I$, и $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. Вещественные числа λ_i являются собственными значениями A и являются корнями характеристического полинома $\det(A - \lambda I)$. Столбцы Q образуют ортонормированный набор собственных векторов A . Такое разложение называется спектральным.²

Мы обычно упорядочиваем вещественные собственные значения как $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Мы используем обозначение $\lambda_i(A)$ для обозначения i -го наибольшего собственного значения $A \in S$. Мы обычно пишем наибольшее или максимальное собственное значение как $\lambda_1(A) = \lambda_{\max}(A)$, и наименьшее или минимальное собственное значение как $\lambda_n(A) = \lambda_{\min}(A)$.

²Хорошая шпаргалка с разложением матриц доступна на сайте курса по линейной алгебре [website](#).

Собственные значения

Наибольшее и наименьшее вещественные собственные значения удовлетворяют

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

Собственные значения

Наибольшее и наименьшее вещественные собственные значения удовлетворяют

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

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

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

Собственные значения

Наибольшее и наименьшее вещественные собственные значения удовлетворяют

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

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

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

Число обусловленности невырожденной матрицы определяется как

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

Собственные значения

Наибольшее и наименьшее вещественные собственные значения удовлетворяют

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

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

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

Число обусловленности невырожденной матрицы определяется как

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

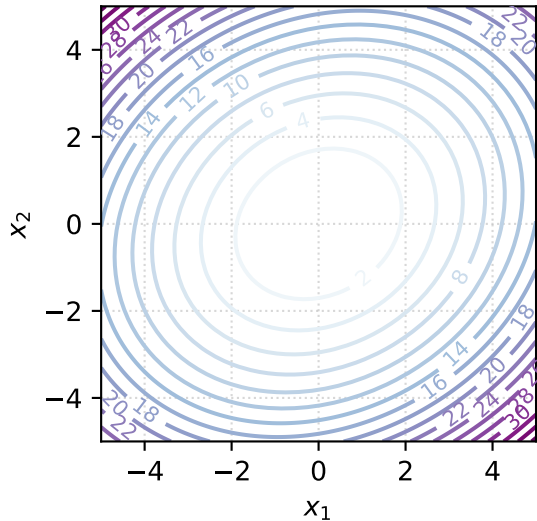
Если мы используем спектральную матричную норму, мы можем получить:

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

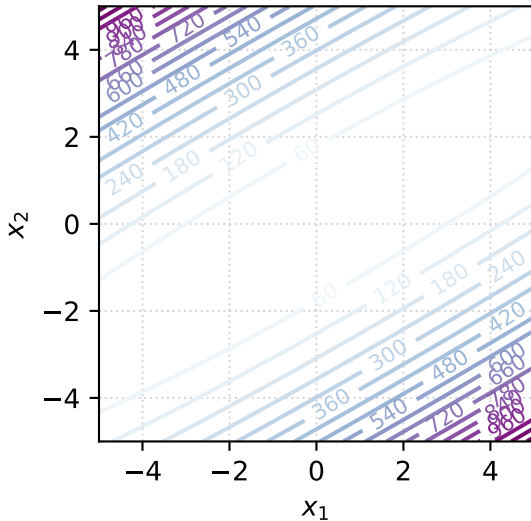
Если, кроме того, $A \in \mathbb{S}_{++}^n$: $\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$

Число обусловленности

$\kappa = 1.5$



$\kappa = 50$



Сингулярное разложение (SVD)

Пусть $A \in \mathbb{R}^{m \times n}$ с рангом $A = r$. Тогда A может быть разложена как

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

Сингулярное разложение (SVD)

Пусть $A \in \mathbb{R}^{m \times n}$ с рангом $A = r$. Тогда A может быть разложена 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 Σ является диагональной матрицей с $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$, такой что

Сингулярное разложение (SVD)

Пусть $A \in \mathbb{R}^{m \times n}$ с рангом $A = r$. Тогда A может быть разложена 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 Σ является диагональной матрицей с $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$, такой что

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

Сингулярное разложение (SVD)

Пусть $A \in \mathbb{R}^{m \times n}$ с рангом $A = r$. Тогда A может быть разложена 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 Σ является диагональной матрицей с $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$, такой что

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

Это разложение называется **сингулярным разложением (SVD)** матрицы A . Столбцы U называются левыми сингулярными векторами A , столбцы V называются правыми сингулярными векторами, и числа σ_i являются сингулярными значениями. Сингулярное разложение может быть записано как

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

где $u_i \in \mathbb{R}^m$ являются левыми сингулярными векторами, и $v_i \in \mathbb{R}^n$ являются правыми сингулярными векторами.

Сингулярное разложение

i Question

Пусть $A \in \mathbb{S}_{++}^n$. Что мы можем сказать о связи между его собственными значениями и сингулярными значениями?

Сингулярное разложение

i Question

Пусть $A \in \mathbb{S}_{++}^n$. Что мы можем сказать о связи между его собственными значениями и сингулярными значениями?

i Question

Как сингулярные значения матрицы связаны с её собственными значениями, особенно для симметричной матрицы?

Пример. Связь между Фробениусовой нормой и сингулярными значениями.

Пусть $A \in \mathbb{R}^{m \times n}$, и пусть $q := \min\{m, n\}$. Докажите, что

$$\|A\|_F^2 = \sum_{i=1}^q \sigma_i^2(A),$$

где $\sigma_1(A) \geq \dots \geq \sigma_q(A) \geq 0$ - сингулярные значения матрицы A . Подсказка: используйте связь между Фробениусовой нормой и скалярным произведением и SVD.

Ранговое разложение (Skeleton decomposition)

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

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

Ранговое разложение (Skeleton decomposition)

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

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

Последнее выражение относится к забавному факту: вы можете случайным образом выбрать r линейно независимых столбцов матрицы и любые r линейно независимых строк матрицы и хранить только их с возможностью точно (!) восстановить всю матрицу.

Ранговое разложение (Skeleton decomposition)

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

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

Последнее выражение относится к забавному факту: вы можете случайным образом выбрать r линейно независимых столбцов матрицы и любые r линейно независимых строк матрицы и хранить только их с возможностью точно (!) восстановить всю матрицу.

Применения для рангового разложения:

- Сжатие модели, сжатие данных и ускорение вычислений в численном анализе: для матрицы ранга r с $r \ll n, m$ необходимо хранить $\mathcal{O}((n+m)r) \ll nm$ элементов.

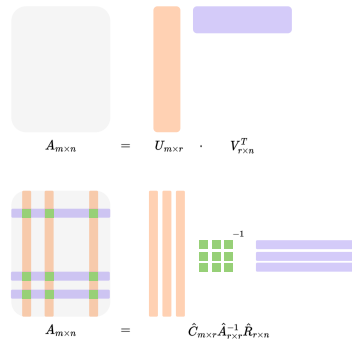


Рис. 3: Иллюстрация рангового разложения

Ранговое разложение (Skeleton decomposition)

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

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

Последнее выражение относится к забавному факту: вы можете случайным образом выбрать r линейно независимых столбцов матрицы и любые r линейно независимых строк матрицы и хранить только их с возможностью точно (!) восстановить всю матрицу.

Применения для рангового разложения:

- Сжатие модели, сжатие данных и ускорение вычислений в численном анализе: для матрицы ранга r с $r \ll n, m$ необходимо хранить $\mathcal{O}((n+m)r) \ll nm$ элементов.
- Извлечение признаков в машинном обучении

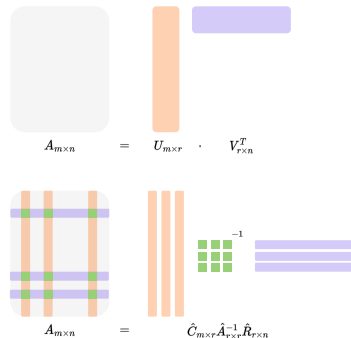


Рис. 3: Иллюстрация рангового разложения

Ранговое разложение (Skeleton decomposition)

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

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

Последнее выражение относится к забавному факту: вы можете случайным образом выбрать r линейно независимых столбцов матрицы и любые r линейно независимых строк матрицы и хранить только их с возможностью точно (!) восстановить всю матрицу.

Применения для рангового разложения:

- Сжатие модели, сжатие данных и ускорение вычислений в численном анализе: для матрицы ранга r с $r \ll n, m$ необходимо хранить $\mathcal{O}((n+m)r) \ll nm$ элементов.
- Извлечение признаков в машинном обучении
- Все приложения, где применяется SVD, так как ранговое разложение может быть преобразовано в форму усеченного SVD.

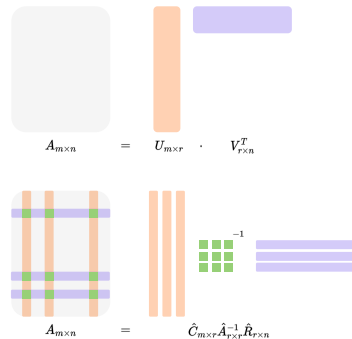


Рис. 3: Иллюстрация рангового разложения

Каноническое тензорное разложение

Можно рассмотреть обобщение рангового разложения на структуры данных более высокого порядка, такие как тензоры, что означает представление тензора в виде суммы r простых тензоров.

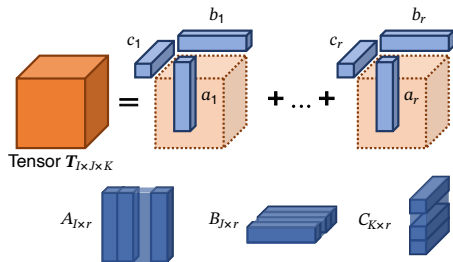


Рис. 4: Иллюстрация канонического тензорного разложения

Example

Заметьте, что существует множество тензорных разложений: каноническое, Таккера, тензорный поезд (ТТ), тензорное кольцо (ТК) и другие. В случае тензоров мы не имеем прямого определения *ранга* для всех типов разложений. Например, для разложения Тензорного поезда ранг является не скаляром, а вектором.

Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные значения

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

Определитель имеет несколько интересных свойств. Например,

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

Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные значения

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

Определитель имеет несколько интересных свойств. Например,

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

Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные значения

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

Определитель имеет несколько интересных свойств. Например,

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

Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные значения

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

Определитель имеет несколько интересных свойств. Например,

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

Определитель и след матрицы

Определитель и след матрицы могут быть выражены через собственные значения

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

Определитель имеет несколько интересных свойств. Например,

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

Не забывайте о циклическом свойстве следа для произвольных матриц A, B, C, D (предполагая, 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 значения

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

Определитель имеет несколько интересных свойств. Например,

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

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

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

Question

Как определитель матрицы связан с её обратимостью?

Задача. Найдите свое скалярное произведение.

Упростите следующее выражение:

$$\sum_{i=1}^n \langle S^{-1}a_i, a_i \rangle,$$

где $S = \sum_{i=1}^n a_i a_i^T, a_i \in \mathbb{R}^n, \det(S) \neq 0$

Пример. LoRA: Low-Rank Adaptation of Large Language Models (arXiv:2106.09685)

Поскольку современные LLM слишком большие, чтобы влезть в память среднего пользователя, мы используем некоторые трюки, чтобы сделать их потребление памяти меньше. Одним из наиболее популярных трюков является LoRA (Low-Rank Adaptation of Large Language Models).

Предположим, у нас есть матрица $W \in \mathbb{R}^{d \times k}$ и мы хотим выполнить следующее обновление:

$$W = W_0 + \Delta W.$$

Основная идея LoRA состоит в том, чтобы разложить обновление ΔW на две низкоранговые матрицы:

$$W = W_0 + \Delta W = W_0 + BA, \quad B \in \mathbb{R}^{d \times r}, A \in \mathbb{R}^{r \times k}, \\ \text{rank}(A) = \text{rank}(B) = r \ll \min\{d, k\}.$$

Проверьте 📁 ноутбук для примера реализации LoRA.

