

Лекция 1. Основы матричного анализа

М. В. Рахуба

20 января 2022 г.

1 Нормированные пространства

Определение 1. Пусть задано некоторое линейное пространство V над полем $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Функцию $\|\cdot\|: V \rightarrow \mathbb{R}$ будем называть нормой, если она удовлетворяет следующим четырем свойствам

1. $\|x\| \geq 0 \quad \forall x \in V$,
2. $\|x\| = 0 \Leftrightarrow x = 0$,
3. $\|\alpha x\| = |\alpha| \|x\| \quad \forall \alpha \in \mathbb{F}, \forall x \in V$,
4. $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in V$, (неравенство треугольника).

Пространство V , оснащенное нормой, называется нормированным пространством.

1.1 Векторные нормы

Пусть $V = \mathbb{F}^n$, где $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Элементы $x \in \mathbb{F}^n$ будем записывать в виде векторов-столбцов $x = [x_1, \dots, x_n]^T$. Введем p -норму:

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

для некоторого $p \geq 1$. Среди p -норм популярными являются случаи $p \in \{1, 2, \infty\}$:

- $\|x\|_1 = |x_1| + \dots + |x_n|$ (1-норма). В некоторых случаях является желаемым выбором, так как, например, помогает наложить условие разреженности¹ на векторы (подробнее дальше в курсе). Несмотря на это, она не является гладкой функцией, и построение эффективных алгоритмов может оказаться непростой задачей.
- $\|x\|_2 = (|x_1|^2 + \dots + |x_n|^2)^{1/2} = \sqrt{x^* x}$ (2- или евклидова норма). Часто является наиболее удобным выбором для анализа. Позволяет получать явные формулы и строить эффективные алгоритмы для решения некоторых задач оптимизации, например, для метода наименьших квадратов.

Запись $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ обозначает, что \mathbb{F} является либо полем действительных, либо комплексных чисел.

Обратите внимание, что $x = 0 \Rightarrow \|x\| = 0$ следует из 3-го свойства при подстановке $\alpha = 0$.

Первые 3 свойства из Определения 1 для $\|\cdot\|_p$ проверяются непосредственно. Неравенство треугольника верно при любом $p \geq 1$ и называется неравенством Минковского. Для случаев $p \in \{1, 2, \infty\}$ проверка неравенства треугольника является простым упражнением.

¹ Разреженными называются векторы, у которых большое количество компонент равно 0.

$A^* \equiv \bar{A}^T$ (транспонирование и поэлементное комплексное сопряжение) называется эрмитовым сопряжением A . Например,

$$\begin{bmatrix} 1 \\ i \end{bmatrix}^* = \begin{bmatrix} 1 & -i \end{bmatrix}.$$

- $\|x\|_\infty = \max_i |x_i|$ (бесконечная норма). В силу того, что $\|x\|_p \rightarrow \|x\|_\infty$ при $p \rightarrow \infty$, мы также относим $\|\cdot\|_\infty$ к p -нормам при $p = \infty$. Ошибки, измеренные в $\|x\|_\infty$, легко интерпретировать. Например, если известно, что \hat{x} приближает x с относительной точностью 10^{-d} , то есть,

$$\frac{\|\hat{x} - x\|_\infty}{\|x\|_\infty} \approx 10^{-d}, \quad d \in \mathbb{N},$$

то каждая компонента \hat{x} имеет как минимум d верных знаков после запятой.

На Рисунке 1 представлены единичные сферы $\|x\|_p = 1$ для различных p . Обратите внимание, что в случае $p = \frac{1}{2}$ функция $\|\cdot\|_p$ не является нормой (используя неравенство треугольника, несложно убедиться, что множество $\{x \in \mathbb{R}^2: \|x\|_p \leq 1\}$ должно быть выпуклым).

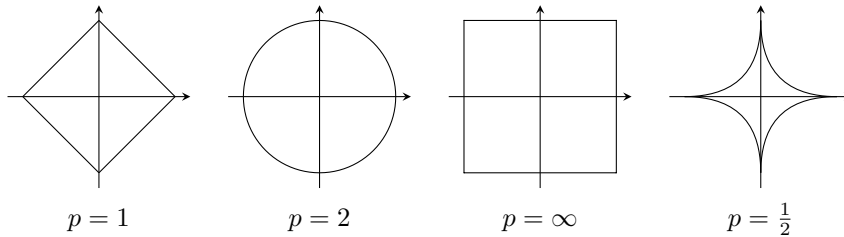


Рис. 1: Единичные сферы

$$\{x \in \mathbb{R}^2: \|x\|_p = 1\}$$

для различных значений p .

1.2 Эквивалентность норм и сходимости

Определение 2. Нормы $\|\cdot\|_a$ и $\|\cdot\|_b$ на нормированном пространстве V называются эквивалентными, если

$$C_1 \|x\|_a \leq \|x\|_b \leq C_2 \|x\|_a \quad \forall x \in V,$$

где $C_1, C_2 > 0$ – константы, не зависящие от x .

Известно, что все нормы на конечномерных пространствах являются эквивалентными. Например, для p -норм на \mathbb{F}^n справедливы следующие неравенства:

$$\begin{aligned} \|x\|_2 &\leq \|x\|_1 \leq \sqrt{n} \|x\|_2, \\ \|x\|_\infty &\leq \|x\|_2 \leq \sqrt{n} \|x\|_\infty, \\ \|x\|_\infty &\leq \|x\|_1 \leq n \|x\|_\infty. \end{aligned}$$

Будем говорить, что последовательность $\{x_k\}$ сходится к x , если $\|x_k - x\| \rightarrow 0$ при $k \rightarrow \infty$. Благодаря эквивалентности норм на конечномерных пространствах мы можем выбрать любую норму для проверки сходимости.

В случае с $\|\cdot\|_1$ или $\|\cdot\|_2$ это, вообще говоря, неверно. Пусть, например, $x \in \mathbb{R}^n$ – вектор из всех единиц, а \hat{x} отличается от x только в одной позиции: $\hat{x}_1 = 0$. Тогда $\|x - \hat{x}\|_1 / \|x\|_1 = 1/n$ мало при большом n , но $\|x - \hat{x}\|_\infty / \|x\|_\infty = 1$.

Константы C_1 и C_2 , вообще говоря, зависят от V . Например, для \mathbb{F}^n и p -норм мы получим зависимость C_1, C_2 от n .

2 Матричные нормы

Мы можем ввести нормы для матриц размера $m \times n$ как нормы из Определения 1 для векторного пространства $V = \mathbb{F}^{m \times n}$. Тогда по аналогии с p -нормами векторов для $p \in \{1, 2, \infty\}$ определим следующие нормы:

$$\|A\|_{\text{sum}} = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|,$$

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

$$\|A\|_C = \max_{i,j} |a_{ij}| \quad (\text{чебышёвская или max-норма}).$$

Полезными формулами также являются:

$$\|A\|_F^2 = \text{trace}(A^* A) = \text{trace}(A A^*).$$

Однако для того, чтобы называть норму *матричной*, мы потребуем еще одно условие — субмультипликативность.

Определение 3. Пусть $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Функция $\|\cdot\|$, которая каждой матрице A ставит в соответствие действительное число $\|A\|$, называется *матричной нормой*, если

1. Она является нормой на $V = \mathbb{F}^{m \times n}$ для любых $m, n \in \mathbb{N}$.
2. $\|AB\| \leq \|A\| \|B\|$ для всех $a \rightarrow b$, для которых определено AB (свойство субмультипликативности).

В литературе также встречается определение матричной нормы без требования субмультипликативности.

Можно проверить, что нормы $\|\cdot\|_{\text{sum}}$ и $\|\cdot\|_F$ удовлетворяют условию субмультипликативности. Для $\|\cdot\|_C$ это неверно:

$$A = B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \implies AB = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix},$$

то есть $\|AB\|_C = 2 > 1 = \|A\|_C \|B\|_C$.

2.1 Операторные нормы

Матричные нормы можно также ввести с помощью векторных норм.

Определение 4. Пусть задана матрица $A \in \mathbb{F}^{m \times n}$, и на \mathbb{F}^n задана норма $\|\cdot\|_a$, а на \mathbb{F}^m — норма $\|\cdot\|_b$. Операторной (или индуцированной) нормой A , подчиненной векторным нормам $\|\cdot\|_a$ и $\|\cdot\|_b$, называется следующее выражение:

$$\|A\|_{a \rightarrow b} = \sup_{x \neq 0} \frac{\|Ax\|_b}{\|x\|_a}.$$

По сути выражение $\|Ax\|_b / \|x\|_a$ является коэффициентом растяжения вектора x при умножении на матрицу A . Операторная норма является наибольшим коэффициентом растяжения по всем направлениям.

Замечание 1.

1. Для $\|A\|_{a \rightarrow b}$ можно также записать:

$$\|A\|_{a \rightarrow b} = \sup_{x \neq 0} \left\| A \left(\frac{x}{\|x\|_a} \right) \right\|_b = \sup_{\|x\|_a = 1} \|Ax\|_b.$$

2. В определении $\|A\|_{a \rightarrow b}$ супремум можно заменить на максимум благодаря теореме Вейерштрасса: конечномерная сфера является компактным множеством, а норма — непрерывной функцией.

3. Операторные нормы являются нормами, но не любых двух норм $\|\cdot\|_a$ и $\|\cdot\|_b$ являются матричными нормами.

Важным частным случаем операторных норм являются p -нормы матриц, подчиненные векторным p -нормам:

$$\|A\|_p \equiv \|A\|_{p \rightarrow p} = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}.$$

Несложно убедиться, что такие нормы являются матричными, так как

$$\begin{aligned} \|AB\|_p &= \sup_{x \neq 0} \frac{\|ABx\|_p}{\|x\|_p} = \sup_{\substack{x \neq 0 \\ Bx \neq 0}} \frac{\|ABx\|_p}{\|Bx\|_p} \frac{\|Bx\|_p}{\|x\|_p} \leq \\ &\leq \sup_{\substack{x \neq 0 \\ Bx \neq 0}} \frac{\|ABx\|_p}{\|Bx\|_p} \sup_{\substack{x \neq 0 \\ Bx \neq 0}} \frac{\|Bx\|_p}{\|x\|_p} \leq \|A\|_p \|B\|_p. \end{aligned}$$

В качестве упражнения покажите, что

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|, \quad \|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|,$$

то есть $\|A\|_1$ есть максимальная из сумм модулей элементов в столбцах, а $\|A\|_\infty$ — в строках. На следующей лекции мы покажем, что $\|A\|_2 = \sigma_1(A)$ — старшее сингулярное число матрицы A , которое также совпадает с $\sqrt{\lambda_1(A^*A)}$ и $\sqrt{\lambda_1(AA^*)}$ — максимальными собственными значениями матриц A^*A и AA^* соответственно. Поэтому 2-норма матриц иногда называется *спектральной*.

Фробениусова норма $\|A\|_F$, а также $\|A\|_p$ при $p \in \{1, 2, \infty\}$ будут регулярно встречаться в нашем курсе.

2.2 Эквивалентность матричных норм и сходимости

По аналогии с векторным случаем будем говорить, что последовательность матриц $\{A_k\}$ сходится к матрице A , если

$\|A_k - A\| \rightarrow 0$ при $k \rightarrow \infty$. Так как на конечномерных пространствах все нормы эквивалентны, то для проверки сходимости можно выбрать любую из норм (не обязательно матричную).

3 Унитарные матрицы

Особую роль для построения эффективных и устойчивых матричных алгоритмов играют унитарные матрицы.

Определение 5. Матрица $U \in \mathbb{C}^{n \times n}$, удовлетворяющая

$$U^{-1} = U^*,$$

называется унитарной.

Из определения следует, что для унитарных матриц выполняется

$$UU^* = U^*U = I,$$

где I обозначает единичную матрицу. То есть унитарность матрицы подразумевает ортонормированность ее строк и столбцов. Действительно, для столбцов u_1, \dots, u_n матрицы $U = [u_1, \dots, u_n]$ имеем:

$$U^*U = \begin{bmatrix} u_1^* \\ \vdots \\ u_n^* \end{bmatrix} \begin{bmatrix} u_1 & \dots & u_n \end{bmatrix} = \begin{bmatrix} u_1^*u_1 & \dots & u_1^*u_n \\ \vdots & \ddots & \vdots \\ u_n^*u_1 & \dots & u_n^*u_n \end{bmatrix} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix},$$

а значит $u_i^*u_j = \delta_{ij}$ — дельта Кронекера. Следовательно, $u_i \perp u_j$ при $i \neq j$ и $u_i^*u_i = \|u_i\|_2^2 = 1$.

Одним из важных свойств векторной 2-нормы является ее унитарная инвариантность, то есть неизменность значения нормы вектора при умножении на унитарную матрицу.

Утверждение 1. Пусть $U \in \mathbb{C}^{n \times n}$ — унитарная. Тогда

$$\|Ux\|_2 = \|x\|_2 \quad \forall x \in \mathbb{C}^n.$$

Доказательство.

$$\|Ux\|_2^2 = (Ux)^*Ux = x^*U^*Ux = x^*x = \|x\|_2^2.$$

□

Аналогичные утверждения можно сформулировать для фробениусовой и 2-нормы матриц:

В случае $U \in \mathbb{R}^{n \times n}$ матрицы называются ортогональными.

На матрицу $U = [u_1, \dots, u_n]$ можно смотреть как на блочную строку. При умножении блочных матриц мы можем работать с блоками как с числами и использовать правило “строка на столбец”.

Векторы $u, v \in \mathbb{C}^n$ будем называть ортогональными, если $u^*v = 0$.

В обратную сторону утверждение также остается верным: если $\|Ux\|_2 = \|x\|_2 \quad \forall x \in \mathbb{C}^n$, то U — унитарная. Доказательство оставляется в качестве упражнения.

Утверждение 2. Пусть $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$ — унитарные матрицы. Тогда

$$\|UAV\|_F = \|A\|_F,$$

$$\|UAV\|_2 = \|A\|_2,$$

для любой $A \in \mathbb{C}^{m \times n}$.

Доказательство. Доказательство для $\|\cdot\|_F$ следует из формулы $\|A\|_F^2 = \text{trace}(A^*A)$ и перестановочности матриц под следом.

Докажем для $\|\cdot\|_2$:

$$\begin{aligned} \|UAV\|_2 &= \sup_{x \neq 0} \frac{\|UAVx\|_2}{\|x\|_2} \stackrel{(\text{утв.1})}{=} \sup_{x \neq 0} \frac{\|AVx\|_2}{\|Vx\|_2} = \\ &\stackrel{y=Vx}{=} \sup_{y \neq 0} \frac{\|Ay\|_2}{\|y\|_2} = \|A\|_2. \end{aligned}$$

□

4 Разложение Шура

4.1 Собственное разложение и жорданова форма

Давайте вспомним некоторые разложения из классического курса линейной алгебры. Во-первых, запишем *собственное разложение* матрицы $A \in \mathbb{C}^{n \times n}$:

$$A = S\Lambda S^{-1},$$

где $S = [s_1, \dots, s_n] \in \mathbb{C}^{n \times n}$ — невырожденная матрица, столбцы которой являются собственными векторами A , а $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ — матрица, на диагонали которой стоят собственные значения A . Действительно:

$$A = S\Lambda S^{-1} \iff AS = S\Lambda \iff As_i = \lambda_i s_i.$$

Собственное разложение существует не для любой квадратной матрицы. Однако любую квадратную матрицу можно привести к *жордановой нормальной форме* (ЖНФ):

$$A = PJP^{-1},$$

где $P \in \mathbb{C}^{n \times n}$ — некоторая невырожденная матрица, а $J \in \mathbb{C}^{n \times n}$ — блочно-диагональная матрица вида

$$J = \begin{bmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_m}(\lambda_m) \end{bmatrix}, \quad J_{n_k}(\lambda_k) = \begin{bmatrix} \lambda_k & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix}_{n_k \times n_k},$$

где $n_1 + \dots + n_m = n$, n_k — алгебраическая кратность собственного значения λ_k . ЖНФ является полезным инструментом при теоретическом анализе во многих приложениях, например, при анализе систем дифференциальных уравнений, матричных функций, интерационных процессов и т.д. Несмотря на это, ее применение в вычислениях является ограниченным. Одним из неприятных свойств является то, что любую матрицу можно диагонализировать, возмущив ее элементы сколько угодно малым образом. Это означает, что сколь угодно малые изменения в матрице могут привести к большим изменениям в ЖНФ: J может стать диагональной при возмущении.

4.2 Разложение Шура

В вычислениях, например, при вычислении матричных функций используется другое разложение — *разложение Шура*. Оказывается, любая квадратная матрица подобна верхнетреугольной, и более того, это подобие унитарное. Справедлива следующая теорема:

Теорема 1 (Теорема Шура). Для любой матрицы $A \in \mathbb{C}^{n \times n}$ найдется унитарная $U \in \mathbb{C}^{n \times n}$ и верхнетреугольная матрица $T \in \mathbb{C}^{n \times n}$, такие что

$$A = UTU^* \quad (\text{разложение Шура}),$$

причем на диагонали T находятся собственные значения A .

Доказательство. Будем доказывать теорему по индукции. Для $n = 1$ доказательство очевидно. Пусть теорема доказана для матриц порядка $n - 1$. Так как мы рассматриваем поле комплексных чисел, то у любой матрицы найдется как минимум одна собственная пара:

$$Av_1 = \lambda_1 v_1, \quad \|v_1\|_2 = 1.$$

Рассмотрим матрицу $V = [v_1, v_2, \dots, v_n] \in \mathbb{C}^{n \times n}$, где v_2, \dots, v_n выбраны так, чтобы столбцы матрицы V были ортонормированы. Значит, V — унитарная (Секция 3). Получим:

$$V^*AV = \begin{bmatrix} v_1^* \\ \dots \\ v_n^* \end{bmatrix} \begin{bmatrix} Av_1 & \dots & Av_n \end{bmatrix} = \begin{bmatrix} v_1^*Av_1 & \dots & v_1^*Av_n \\ \vdots & \ddots & \vdots \\ v_n^*Av_1 & \dots & v_n^*Av_n \end{bmatrix}.$$

Так как $Av_1 = \lambda_1 v_1$ и $v_1 \perp v_i$ при $2 \leq i \leq n$, матрица V^*AV

Рассмотрим матрицу

$$A(\varepsilon) = \begin{bmatrix} 0 & 1 \\ 0 & \varepsilon \end{bmatrix},$$

у которой при $\varepsilon > 0$ два разных с.з.: 0 и ε , а значит матрица диагонализуема. Однако при $\varepsilon = 0$ матрица не является диагонализуемой.

Напомним, что квадратные матрицы $a \rightarrow b$ одного размера называются подобными, если существует S :

$$A = SBS^{-1}.$$

Обратите внимание, что $A = UTU^* = UTU^{-1}$ в силу унитарности U , а значит матрицы A и T подобны.

имеет следующий вид:

$$V^*AV = \begin{bmatrix} \lambda_1 & \star \\ 0 & A_1 \end{bmatrix}.$$

По предположению индукции $A_1 \in \mathbb{C}^{(n-1) \times (n-1)}$ имеет разложение Шура: $A_1 = U_1 T_1 U_1^*$. Следовательно,

$$V^*AV = \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \star \\ 0 & T_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix}^*$$

Вводя обозначение

$$U = V \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix}, \quad T = \begin{bmatrix} \lambda_1 & \star \\ 0 & T_1 \end{bmatrix},$$

получаем искомое разложение $A = UTU^*$. \square

Замечание 2. Разложение Шура можно записать и над \mathbb{R} , но с небольшой модификацией. Для любой $A \in \mathbb{R}^{n \times n}$ справедливо

$$A = UTU^T,$$

где $U \in \mathbb{R}^{n \times n}$ – ортогональная, а T – блочно-верхнетреугольная:

$$T = \begin{bmatrix} T_{11} & T_{12} & \dots & T_{1m} \\ 0 & T_{22} & \dots & T_{2m} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & T_{mm} \end{bmatrix} \in \mathbb{R}^{n \times n},$$

где T_{ii} , $i = 1, \dots, m$ являются блоками размера 1×1 или 2×2 .

Список литературы

- [1] Е. Е. Тьртышников. *Методы численного анализа*. Академия, 2007.
- [2] Е. Е. Тьртышников. *Матричный анализ и линейная алгебра*. Физматлит, 2007.
- [3] G.H. Golub and C.F. Van Loan. *Matrix Computations, 4th Edition*. The Johns Hopkins University Press, Baltimore, 2013.
- [4] <https://github.com/oseledets/nla2020>.

Символом “ \star ” мы будем обозначать не интересующие нас элементы в матрице. Будем также писать 0, подразумевая некоторую нулевую подматрицу подходящего размера. В данном случае 0 является вектором-столбцом длины $n - 1$.

Несложно убедиться, что произведение унитарных матриц P и Q также является унитарной:

$$(PQ)^*(PQ) = Q^*P^*PQ = Q^*Q = I.$$