C1 Пререквизиты из линейной алгебры

Линейная алгебра играет фундаментальную роль во многих областях математики. Её методы и концепции лежат в основе таких дисциплин, как численные методы, теория управления, анализ данных и, в том числе, теория оптимизации — предмет, которому посвящено пособие. В рамках этого параграфа обсуждаются пререквизиты из линейной алгебры, необходимые для успешного усвоения материала по оптимизации. В частности, обсуждаются векторные и матричные нормы, а также разложения матриц.

C1.1 Векторные нормы

Определение C1.1. Рассмотрим функцию $\|\cdot\| : \mathbb{R}^d \to \mathbb{R}$. Она называется *векторной нормой*, если удовлетворяет следующим условиям:

- $\|x\| \geq 0$, $\|x\| = 0 \iff x = 0$;
- $\|\lambda x\| = |\lambda|\|x\|, \;\; \lambda \in \mathbb{R}$;
- $\|x + y\| \leq \|x\| + \|y\|$.

Пример C1.1. Функция $\|\cdot\| : \mathbb{R}^d \to \mathbb{R}$, определённая как

$$\|x\|_p = \sqrt[p]{\sum_{i=1}^{d} |x_i|^p}, \; p \geq 1,$$

является векторной нормой.

Доказательство. Первые два условия из Определения C1.1 не нуждаются в доказательстве. Покажем, что для произвольных $x, y \in \mathbb{R}^d$ выполнено третье условие. Рассмотрим $p \neq 1$, поскольку случай $p = 1$ тривиален. Нетрудно видеть, что

$$|x_i + y_i|^p \leq |x_i||x_i + y_i|^{p-1} + |y_i||x_i + y_i|^{p-1}.$$

Обозначив $q = \frac{p}{p-1}$, применим неравенство Гельдера (0.2) к каждому из слагаемых. Получим

$$\begin{aligned}
\|x + y\|_p^p &= \sum_{i=1}^{d} |x_i + y_i|^p \leq \sum_{i=1}^{d} |x_i||x_i + y_i|^{p-1} + \sum_{i=1}^{d} |y_i||x_i + y_i|^{p-1} \\
&\leq \left(\sum_{i=1}^{d} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{d} |x_i + y_i|^p\right)^{\frac{1}{q}} + \left(\sum_{i=1}^{d} |y_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{d} |x_i + y_i|^p\right)^{\frac{1}{q}}.
\end{aligned}$$

Сократив обе части на $\left(\sum_{i=1}^{d} |x_i + y_i|^p\right)^{\frac{1}{q}}$, учитывая, что $1 - \frac{1}{q} = \frac{1}{p}$, имеем

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p.$$

■

Замечание C1.1. Нормы, введённые в Примере C1.1, будем называть *гёльдеровыми нормами* или просто *p-нормами*.

В анализе некоторых алгоритмов иногда возникает так называемая ∞-норма.

Определение C1.2. Функцию $\|\cdot\|_\infty : \mathbb{R}^d \to \mathbb{R}$, определённую как

$$\|x\|_\infty = \max_{i=\overline{1,d}} |x_i|,$$

будем называть ∞-*нормой*.

Пример C1.2. ∞-норма является предельным случаем гёльдеровых норм при $p \to \infty$.

Доказательство. Во-первых, известно, что $|x_j| \le \max_{i=\overline{1,d}} |x_i|$. Это означает, что

$$\|x\|_p \le \left(d \max_{i=\overline{1,d}} |x_i|^p\right)^{\frac{1}{p}} = d^{\frac{1}{p}} \max_{i=\overline{1,d}} |x_i| \underset{p\to\infty}{\longrightarrow} \|x\|_\infty.$$

С другой стороны, имеем

$$\|x\|_p \ge \left(\max_{i=\overline{1,d}} |x_i|^p\right)^{\frac{1}{p}} = \max_{i=\overline{1,d}} |x_i| = \|x\|_\infty.$$

По теореме о трех последовательностях (см. Теорему 2 Параграфа 5 в [27]), имеем

$$\lim_{p\to\infty} \|x\|_p = \|x\|_\infty.$$

■

Теперь, когда мы ввели в рассмотрение множество примеров векторных норм, закономерно возникает вопрос о том, как они соотносятся между собой. Действительно, показав сходимость численного метода в некоторой норме, мы хотели бы иметь уверенность, что он сходится и в других нормах тоже. Из курса математического анализа известно утверждение.

Утверждение C1.1 (Теорема 2 Части 14 в [9])**.** Рассмотрим $\|\cdot\|_A, \|\cdot\|_B : \mathbb{R}^d \to \mathbb{R}$. Существуют такие $c_1, c_2 > 0$, что для любого вектора x из $\mathbb{R}^d$ выполняется соотношение:

$$c_1\|x\|_A \le \|x\|_B \le c_2\|x\|_A.$$

Таким образом, в конечномерном вещественном пространстве все нормы эквивалентны. Доказательство утверждения не конструктивно, однако в ряде задач мы хотели бы конкретизировать значения c_1 и c_2.

Пример C1.3. Для $\|\cdot\|_1$ и $\|\cdot\|_2$ выполнено соотношение:

$$\frac{1}{\sqrt{d}}\|x\|_1 \le \|x\|_2 \le \|x\|_1.$$

Доказательство. Начнём с верхней оценки. Запишем

$$\begin{aligned}\|x\|_2^2 &= |x_1|^2 + \ldots + |x_d|^2 \le |x_1|^2 + \ldots + |x_d|^2 + 2|x_1||x_2| + \ldots + 2|x_{d-1}||x_d| \\ &= (|x_1| + \ldots + |x_d|)^2 = \|x_1\|_1^2.\end{aligned}$$

Таким образом,
$$\|x\|_2 \le \|x\|_1.$$
Заметим, что неравенство переходит в равенство на векторе $x = e_1$. Чтобы оценить $\|x\|_2$ снизу, воспользуемся неравенством Гельдера (0.2), положив один из векторов в скалярном произведении равным единице:

$$\|x\|_1^2 = \left(\sum_{i=1}^{d} 1|x_i|\right)^2 \le \left(\sum_{i=1}^{d} 1^2\right)\left(\sum_{i=1}^{d} |x_i|^2\right) = d\|x\|_2^2.$$

Заметим, что неравенство переходит в равенство на векторе $x = \mathbf{1}$, где $\mathbf{1}$ — вектор из всех единиц. Комбинируя результаты, получаем

$$\frac{1}{\sqrt{d}}\|x\|_1 \le \|x\|_2 \le \|x\|_1.$$

■

C1.2 Матричные нормы

В анализе численных методов помимо векторов также приходится работать с матрицами. Существует естественное обобщение определения нормы.

Определение C1.3. Рассмотрим функцию $\|\cdot\| : \mathbb{R}^{n\times d} \to \mathbb{R}$. Она называется *матричной нормой*, если удовлетворяет следующим условиям:

- $\|A\| \ge 0, \quad \|A\| = 0 \iff A = 0;$
- $\|\lambda A\| = |\lambda|\|A\|, \ \lambda \in \mathbb{R};$
- $\|A + B\| \le \|A\| + \|B\|.$

Замечание C1.2. Для матриц помимо сложения также определена операция умножения. Ключевым для анализа численных методов является свойство

$$\|AB\| \le \|A\|\|B\|.$$

Тем не менее, оно выполнено не для всех матричных норм. Рассмотрим норму:

$$\|A\| = \max_{\substack{i=\overline{1,n}\\ j=\overline{1,d}}} |a_{ij}|$$

и матрицы

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Получим, что $\|AB\| = 2$, $\|A\| = 1$, $\|B\| = 1$, то есть свойство не выполняется.

Замечание C1.3. Если для $\|\cdot\| : \mathbb{R}^{n\times d} \to \mathbb{R}$ выполнено

$$\|AB\| \le \|A\|\|B\|,$$

то её называют *субмультипликативной* матричной нормой.

Может показаться, что мы вынуждены вручную проверять субмультипликативность, сталкиваясь с новой матричной нормой. Эту проблему можно решить, рассматривая более узкий класс норм, для которых это свойство выполняется автоматически. В дальнейшем окажется, что наиболее часто используемые матричные нормы принадлежат этому классу.

Определение C1.4. Рассмотрим $A \in \mathbb{R}^{n \times d}$, норму $\|\cdot\|_\alpha : \mathbb{R}^d \to \mathbb{R}$ и норму $\|\cdot\|_\beta : \mathbb{R}^n \to \mathbb{R}$. Матричной нормой, *подчинённой* $\|\cdot\|_\alpha$ *и* $\|\cdot\|_\beta$, называется функция $\|\cdot\|_{\alpha,\beta} : \mathbb{R}^{n \times d} \to \mathbb{R}$, заданная как

$$\|A\|_{\alpha,\beta} = \sup_{\|x\|_\alpha = 1} \|Ax\|_\beta.$$

Замечание C1.4. Мы будем рассматривать только случай, когда $\alpha = \beta$, тогда определение можно переписать следующим образом:

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

Замечание C1.5. Определение C1.4 можно переформулировать, представив единичный x как некоторый вектор после нормировки на единичную сферу. Тогда имеем

$$\|A\| = \sup_{\|x\|=1} \|Ax\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}.$$

Замечание C1.6. Из Замечания C1.5 следует неравенство

$$\frac{\|Ax\|}{\|x\|} \leq \|A\| \implies \|Ax\| \leq \|A\|\|x\|.$$

Это очень полезное свойство, которое будет неоднократно использовано в дальнейшем изложении.

Пример C1.4. Подчинённая матричная норма всегда субмультипликативна.

Доказательство. Рассмотрим матрицы A и B, имеющие правильные размерности. Записав определение подчинённой нормы, получим

$$\|AB\| = \sup_{\|x\|=1} \|ABx\| \leq \sup_{\|x\|=1} \|A\|\|Bx\| \leq \sup_{\|x\|=1} \|A\|\|B\|\|x\| = \|A\|\|B\|.$$

■

Использование подчинённой нормы также позволяет оценивать спектральный радиус матрицы.

Утверждение C1.2. Рассмотрим $A \in \mathbb{S}^d$, где $\mathbb{S}^d$ — множество симметричных матриц. Обозначим $\rho(A) = \max_{i=\overline{1,d}} |\lambda_i(A)|$ — *спектральный радиус*, где $\lambda_i(A)$ — собственные числа матрицы A. Тогда

$$\rho(A) \leq \|A\|,$$

где $\|\cdot\|$ — подчинённая норма.

Доказательство. Запишем уравнение на собственные векторы матрицы A:

$$Ax = \lambda x \implies \|Ax\| = |\lambda|\|x\|.$$

В Замечании C1.6 было получено неравенство

$$\|Ax\| \leq \|A\|\|x\|.$$

Тогда можем утверждать, что

$$|\lambda| \leq \|A\|.$$

Отсюда очевидно следует, что

$$\rho(A) = \max_{i=\overline{1,d}} |\lambda_i(A)| \leq \|A\|.$$

Заметим, что на I_d достигается равенство. ■

Помимо собственных чисел часто оказывается удобным использовать сингулярные. Например, если работаем с прямоугольной матрицей.

Определение C1.5. *Сингулярным числом* матрицы $A \in \mathbb{R}^{n\times d}$ называется

$$\sigma_i(A) = \sqrt{\lambda_i\left(A^\top A\right)}.$$

На практике часто приходится работать с квадратными симметричными матрицами. В этом случае сингулярные числа имеют более удобный вид.

Утверждение C1.3. Пусть $A \in \mathbb{S}^d$. Тогда сингулярные числа имеют вид

$$\sigma_i(A) = |\lambda_i(A)|.$$

Доказательство. Собственные векторы вещественной симметричной матрицы A ортогональны (см. Теорему 1 Параграфа 3 Главы 6 в [26]). Рассмотрев их в качестве базиса, можно привести A к диагональному виду. Запишем её преобразование при переходе от стандартного базиса в базис собственных векторов:

$$A = V\Lambda V^\top,$$

где $V \in \mathbb{R}^{d\times d}$ — ортогональная матрица, составленная из собственных векторов A, $\Lambda = \mathrm{diag}(\lambda_1(A), \ldots, \lambda_d(A))$. Теперь запишем

$$A^\top A = A^2 = (V\Lambda V^\top)(V\Lambda V^\top) = V\Lambda^2 V^\top.$$

Таким образом, доказали, что $\lambda_i(A^\top A) = \lambda_i(A)^2$, откуда следует, что $\sigma_i(A) = |\lambda_i(A)|$. ■

Таким образом, подчинённая матричная норма определена достаточно удачно и имеет ряд хороших свойств. Более того, она даёт связь с векторными нормами. Оказывается, что Определение C1.4 позволяет давать аналитические выражения для подсчета матричных норм.

Пример C1.5. Рассмотрим $A \in \mathbb{R}^{n\times d}$. Для подчинённых норм $\|\cdot\|_\infty, \|\cdot\|_1, \|\cdot\|_2 : \mathbb{R}^{n\times d} \to \mathbb{R}$, имеем следующие выражения:

- $\|A\|_\infty = \max\limits_{i=\overline{1,n}} \sum\limits_{j=1}^{d} |a_{ij}|$ — максимум сумм по строкам,

- $\|A\|_1 = \max\limits_{j=\overline{1,d}} \sum\limits_{i=1}^{n} |a_{ij}|$ — максимум сумм по столбцам.

- $\|A\|_2 = \max\limits_{i=\overline{1,d}} \sqrt{\lambda_i\left(A^\top A\right)} = \max\limits_{i=\overline{1,d}} \sigma_i(A)$ — максимальное сингулярное число.

Доказательство. Будем доказывать в том же порядке, в котором утверждения приведены в формулировке примера.

- Распишем:

$$\|Ax\|_\infty = \max_{i=\overline{1,n}} \left| \sum_{j=1}^{d} a_{ij} x_j \right| \leq \max_{i=\overline{1,n}} \sum_{j=1}^{d} |a_{ij} x_j| \leq \max_{k=\overline{1,d}} |x_k| \max_{i=\overline{1,n}} \sum_{j=1}^{d} |a_{ij}|$$
$$= \|x\|_\infty \max_{i=\overline{1,n}} \sum_{j=1}^{d} |a_{ij}|.$$

  Пусть максимум правой части достигается при $i = i_0$. Рассмотрим вектор $x = \left(\frac{a_{i_0 1}}{|a_{i_0 1}|}, \ldots, \frac{a_{i_0 d}}{|a_{i_0 d}|} \right)^\top$. Тогда, подставляя его в выражение, выше получаем равенство. Таким образом, существует вектор, на котором верхняя грань достигается.

- Будем расписывать почти как в предыдущем примере:

$$\|Ax\|_1 = \sum_{i=1}^{n} \left| \sum_{j=1}^{d} a_{ij} x_j \right| \leq \sum_{j=1}^{d} |x_j| \sum_{i=1}^{n} |a_{ij}| \leq \|x\|_1 \max_{j=\overline{1,d}} \sum_{i=1}^{n} |a_{ij}|.$$

  Пусть максимум правой части достигается при $j = j_0$. Рассмотрим вектор $x = e_{j_0}$. Тогда, подставляя его в выражение, выше получаем равенство. Таким образом, существует вектор, на котором верхняя грань достигается.

- Раскроем определение:

$$\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sup_{x \neq 0} \sqrt{\frac{\langle A^\top Ax, x \rangle}{\langle x, x \rangle}}.$$

  Разложим x по ортонормированному базису $\nu_1, \ldots, \nu_d$ собственных векторов, отвечающих собственным числам $\lambda_1, \ldots, \lambda_d$ матрицы $A^\top A$. Напомним, что собственные числа симметричной положительно полуопределённой матрицы неотрицательны. Тогда:

$$\|A\|_2 = \sup_{\|x\|_2 = 1} \sqrt{\frac{\sum_{j=1}^{d} \lambda_j c_j^2}{\sum_{j=1}^{d} c_j^2}} \leq \max_{i=\overline{1,d}} \sqrt{\lambda_i} = \max_{i=\overline{1,d}} \sigma_i(A),$$

где c_i — коэффициенты разложения x по базису $\nu_1, \ldots, \nu_d$:

$$x = \sum_{i=1}^{d} c_i \nu_i.$$

Рассмотрим нормированный вектор, соответствующий максимальному собственному значению матрицы $A^\top A$. Тогда, подставляя его в выражение, выше получаем равенство. Таким образом, существует вектор, на котором верхняя грань достигается.

∎

Пример C1.6. Матричные нормы $\|\cdot\|_\infty$, $\|\cdot\|_1$ и $\|\cdot\|_2$ связаны соотношением:

$$\|A\|_2^2 \le \|A\|_1 \|A\|_\infty.$$

Доказательство. Поскольку при транспонировании строки и столбцы меняем местами, верно равенство $\|A\|_1 = \|A^\top\|_\infty$. Запишем определение второй нормы матрицы и для того, чтобы ограничить собственное значение матрицы воспользуемся Утверждением C1.2:

$$\|A\|_2^2 = \max_{i=\overline{1,d}} \lambda_i \left(A^\top A\right) \le \|A^\top A\|_1 \le \|A^\top\|_1 \|A\|_1 = \|A\|_1 \|A\|_\infty.$$

∎

Пример C1.7. Матричная норма $\|A\|_2$ может быть определена иначе:

$$\|A\|_2 = \sup_{\|x\|_2=1, \|y\|_2=1} \left|y^\top Ax\right|.$$

Доказательство. Ранее уже отмечалось, что для матричных норм верно неравенство $\|Ax\| \le \|A\|\|x\|$. Пользуясь этим свойством и неравенством Коши-Буняковского-Шварца (0.3) для нормы $p = 2$, запишем:

$$\left|y^\top Ax\right| = |\langle y, Ax\rangle| \le \|y\|_2 \|Ax\|_2 \le \|y\|_2 \|A\|_2 \|x\|_2 = \|A\|_2.$$

Выберем единичный вектор x_*, на котором достигается $\|A\|_2$ и определим $y_* = \frac{Ax_*}{\|Ax_*\|_2}$. Тогда:

$$\left|y_*^\top Ax_*\right| = \left|\frac{x_*^\top A^\top Ax_*}{\|Ax_*\|_2}\right| = \frac{\|Ax_*\|_2^2}{\|Ax_*\|_2} = \|Ax_*\|_2 = \|A\|_2.$$

∎

Тем не менее, не все нормы, используемые на практике являются подчинёнными. Чтобы перейти к их рассмотрению, требуется ввести понятие ранга матрицы.

Определение C1.6. *Столбцовым рангом* матрицы $A \in \mathbb{R}^{n \times d}$ называется число $\mathrm{rank}_c(A)$, такое, что в A существует линейно независимая система из $\mathrm{rank}_c(A)$ столбцов и нет линейно независимой системы из большего числа столбцов.

Замечание C1.7. Существует множество способов определения ранга матрицы. Мы считаем известным из курса линейной алгебры утверждение об их эквивалентности (см. Теорему 1 Параграфа 3 Главы 5 в [25]). В связи с этим, будем пользоваться обозначением $\operatorname{rank}(A) = \operatorname{rank}_c(A)$.

Пример C1.8. Пусть $A \in \mathbb{R}^{n \times d}$. Матричная норма

$$\|A\|_F = \sqrt{\sum_{i=1}^{\operatorname{rank}(A)} \sigma_i^2(A)}$$

не подчинена никакой векторной норме.

Доказательство. Если бы норма была подчинённой, то для единичной матрицы I_d выполнялось бы

$$\|I_d\|_F = \sup_{\|x\|=1} \|I_d x\| = \sup_{\|x\|=1} \|x\| = 1.$$

Однако

$$\|I_d\|_F = \sqrt{\sum_{i=1}^{d} 1} = \sqrt{d}.$$

Таким образом, предложенная норма не подчинена ни одной из векторных норм. ∎

Определение C1.7. Пусть $A \in \mathbb{R}^{n \times d}$. Функцию $\|\cdot\|_F : \mathbb{R}^{n \times d} \to \mathbb{R}$, определённую как

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

будем называть *фробениусовой нормой.*

Замечание C1.8. Обратим внимание, что пространство матриц является конечномерным вещественным. Это означает, что матричные нормы эквивалентны как и векторные.

Пример C1.9. Для $\|\cdot\|_2$ и $\|\cdot\|_F$ выполнено соотношение:

$$\|A\|_2 \le \|A\|_F \le \sqrt{d}\|A\|_2.$$

Доказательство. Без ограничения общности расположим сингулярные числа в порядке убывания и положим $\operatorname{rank}(A) = r$. Поскольку $\|A\|_2 = \sigma_1$, то по определению фробениусовой нормы имеем

$$\|A\|_2 \le \|A\|_F.$$

С другой стороны, $\|A\|_F \le \sqrt{r}\sigma_1 \le \sqrt{d}\sigma_1 = \sqrt{d}\|A\|_2$. Объединяя неравенства, запишем

$$\|A\|_2 \le \|A\|_F \le \sqrt{d}\|A\|_2.$$

∎

Фробениусова норма может показаться сложной для вычисления по сравнению с рассмотренными ранее. В дальнейшем выяснится, что можно дать простое эквивалентное определение. Для этого нам потребуется ряд вспомогательных утверждений.

Теорема C1.1. Матрица $A \in \mathbb{R}^{n \times d}$ представима в виде SVD-разложения:

$$A = U\Sigma V^\top,$$

где $U \in \mathbb{R}^{n\times n}, V \in \mathbb{R}^{d \times d}$ — ортогональные, $\Sigma \in \mathbb{R}^{n \times d}$ — диагональная матрица, составленная из сингулярных чисел A, расположенных в порядке убывания.

Доказательство. $A^\top A$ неотрицательно определена и симметрична, поэтому её собственные числа неотрицательны и соответствующие им собственные векторы ортогональны. Отсюда из Утверждения C1.3 следует существование ортогональной матрицы V, такой что

$$V^\top A^\top AV = \operatorname{diag}\left(\sigma_1^2, \dots, \sigma_d^2\right).$$

Без ограничения общности, будем считать $\sigma_1^2 \geq \ldots \geq \sigma_d^2 \geq 0$. Поскольку $\operatorname{rank}(A) = r$, имеем

$$\sigma_i = 0, \ \forall i > r.$$

Обозначим $V_r = (v_1, \dots, v_r)$, $\Sigma_r = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$. Тогда имеем

$$V_r^\top A^\top AV_r = \Sigma_r^2.$$

Умножив равенство слева и справа на Σ_r^{-1}, получим

$$\left(\Sigma_r^{-1} V_r^\top A^\top\right)\left(AV_r\Sigma_r^{-1}\right) = I.$$

Обозначим $U_r = AV_r\Sigma_r^{-1}$. Из написанного выше следует $U_r^\top U_r = I$, то есть U_r матрица с ортонормированными столбцами. Поскольку систему линейно независимых векторов можно дополнить до базиса, присоединим к U_r произвольные ортонормированные столбцы и получим новую матрицу U. Тогда

$$A = U\Sigma V^\top.$$

■

Теперь мы готовы дать эквивалентное определение фробениусовой нормы.

Утверждение C1.4. Для матрицы $A \in \mathbb{R}^{n\times d}$ фробениусова норма может быть эквивалентно определена как

$$\left\|A\right\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{d} |a_{ij}|^2.$$

Доказательство. Для простоты изложения введем обозначение:

$$\left\|A\right\|_S^2 = \sum_{i=1}^{n}\sum_{j=1}^{d} |a_{ij}|^2.$$

Будем работать с квадратными матрицами $A \in \mathbb{R}^{d\times d}$. Во-первых, заметим, что для ортогональной матрицы Q выполнено

$$\left\|QA\right\|_S = \left\|A\right\|_S.$$

Действительно, рассмотрим A как совокупность вектор-столбцов: $A = (a_1, \ldots, a_d)$. Тогда имеем

$$\|QA\|_S^2 = \|(Qa_1, \ldots, Qa_d)\|_S^2 = \sum_{i=1}^{d} \sum_{j=1}^{d} |(Qa_i)_j|^2 = \sum_{i=1}^{d} \|Qa_i\|_2^2 = \sum_{i=1}^{d} \|a_i\|_2^2 = \|A\|_S^2.$$

Аналогично проверяется инвариантность относительно умножения на Q справа. Воспользуемся SVD-разложением C1.1 и запишем $\|A\|_S$, применив эту идею. Получим

$$\|A\|_S^2 = \|U\Sigma V^\top\|_S^2 = \|\Sigma\|_S^2 = \sum_{i=1}^{r} \sigma_i^2 = \|A\|_F^2.$$

■

Теперь мы получили простой способ подсчёта фробениусовой нормы. Поскольку она не подчинена ни одной из векторных норм, выполнение субмультипликативного свойства надо проверять вручную.

Пример C1.10. Для матриц $A \in \mathbb{R}^{d \times n}$ и $B \in \mathbb{R}^{n \times k}$ фробениусова норма субмультипликативна:

$$\|AB\|_F \le \|A\|_F \|B\|_F.$$

Доказательство. Пользуясь новым определением и неравенством Коши-Буняковского-Шварца (0.3), запишем

$$\begin{aligned} \|AB\|_F^2 &= \sum_{i=1}^{d} \sum_{j=1}^{k} \left| \sum_{t=1}^{n} a_{it} b_{tj} \right|^2 \le \sum_{i=1}^{d} \sum_{j=1}^{k} \left(\sum_{t=1}^{n} |a_{it}|^2 \right) \left(\sum_{m=1}^{n} |b_{mj}|^2 \right) \\ &= \left(\sum_{i=1}^{d} \sum_{t=1}^{n} |a_{it}|^2 \right) \left(\sum_{m=1}^{n} \sum_{j=1}^{k} |b_{mj}|^2 \right) = \|A\|_F^2 \|B\|_F^2. \end{aligned}$$

■

Фробениусова норма широко используется в анализе методов наряду с подчиненными. Действительно, её вычисление значительно проще, чем, например, второй нормы, при этом она субмультипликативна.

Замечание C1.9. Используя эквивалентное определение, можно заметить, что для квадратных матриц выполнено

$$\|A\|_F^2 = \mathrm{Tr}\left(A^\top A\right),$$

где $\mathrm{Tr}(A) = \sum_{i=1}^{d} a_{ii}$ — *след матрицы*. Таким образом, мы говорим, что фробениусова норма порождена скалярным произведением матриц:

$$\langle A, B \rangle = \mathrm{Tr}\left(A^\top B\right).$$

Это замечание окажется особенно полезным для дифференцирования функций, принимающих на вход матрицу. Напомним без доказательства ряд важных свойств следа.

Утверждение C1.5. Для $\mathrm{Tr} : \mathbb{R}^{d \times d} \to \mathbb{R}$ выполнены следующие равенства:

- $\mathrm{Tr}(A^\top) = \mathrm{Tr}(A)$,

- $\mathrm{Tr}(A+B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$,
- $\mathrm{Tr}(cA) = c\,\mathrm{Tr}(A)$,
- $\mathrm{Tr}(A_1 \dots A_n) = \mathrm{Tr}(A_n A_1 \dots A_{n-1})$.

C1.3 Сопряжённые нормы

Если на исходном пространстве задана норма, то на сопряжённом пространстве задана сопряжённая норма.

Определение C1.8. Пусть в $\mathbb{R}^d$ задана норма $\|\cdot\|$. Тогда *сопряжённая норма* $\|\cdot\|_* : \mathbb{R}^d \to \mathbb{R}$ определяется как

$$\|y\|_* = \sup_{\|x\| \le 1} x^\top y.$$

Введённая функция еще не обязана удовлетворять свойствам нормы, докажем их напрямую.

Утверждение C1.6. $\|\cdot\|_*$ является нормой в $\mathbb{R}^d$.

Доказательство. Рассмотрим свойства нормы. Однородность очевидна, покажем положительную определенность: для произвольного $y \neq 0$ можно взять $x = \frac{y}{\|y\|}$, для которого выполняется

$$x^\top y = \frac{y^\top y}{\|y\|} > 0.$$

А для $y = 0$ имеем

$$\|0\|_* = \sup_{\|x\| \le 1} x^\top 0 = 0.$$

Остается показать неравенство треугольника: $\forall y_1, y_2 \in \mathbb{R}^d$:

$$\|y_1 + y_2\|_* = \sup_{\|x\| \le 1} x^\top (y_1 + y_2) \le \sup_{\|x\| \le 1} x^\top y_1 + \sup_{\|x\| \le 1} x^\top y_2 = \|y_1\|_* + \|y_2\|_*.$$

Следовательно, сопряжённая норма является нормой. ■

Пример C1.11. Пусть $\frac{1}{p} + \frac{1}{q} = 1$, тогда сопряжённая норма к $\|\cdot\|_p$ имеет вид $\|\cdot\|_* = \|\cdot\|_q$.

Доказательство. Для начала покажем, что $\forall y \in \mathbb{R}^d$ $\|y\|_* \le \|y\|_q$. Из неравенства Гёльдера (0.2) $\forall x, y \in \mathbb{R}^d : \|x\|_p \le 1$:

$$x^\top y \le \|x\|_p \|y\|_q \le \|y\|_q.$$

Покажем, что равенство достигается. Пусть $y \neq 0$ и $x \in \mathbb{R}^d$, тогда:

$$x_i = \frac{|y_i|^{q-1} \cdot \mathrm{sign}(y_i)}{\|y\|_q^{q-1}}$$

Нетрудно проверить, что $\|x\|_p = 1$, кроме того имеем

$$x^\top y = \frac{\sum\limits_{i=1}^{n} |y_i|^q}{\|y\|_q^{q-1}} = \|y\|_q.$$

Таким образом, $\forall y \in \mathbb{R}^d$ $\|y\|_* = \|y\|_q$. ∎

C1.4 Квадратичные формы

Часто специальные свойства матриц играют ключевую роль в эффективности методов оптимизации и гарантируют некоторые теоретические оценки. Разберёмся с ними подробнее.

Определение C1.9. Пусть на $\mathbb{R}^d$ задана симметричная билинейная (линейная по обоим аргументам) функция $\alpha : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, тогда соответствующая ей *квадратичная форма* $Q : \mathbb{R}^d \to \mathbb{R}$ определяется следующим образом $\forall x \in \mathbb{R}^d$:

$$Q(x) = \sum_{i=1}^{d} \sum_{j=1}^{d} \alpha_{ij} x_i x_j, \ \alpha_{ij} = \alpha_{ji}.$$

Замечание C1.10. Заметим, что любой квадратичной форме в соответствие можно поставить матрицу $A \in \mathbb{S}^d$. То есть $A = \{\alpha\}_{ij}$, и $Q(x) = x^\top A x$.

Далее мы будем отождествлять квадратичную форму и соответствующую ей симметричную матрицу.

Определение C1.10. Квадратичная форма $A \in \mathbb{S}^d$ называется *положительно определённой (полуоопределённой)*, если $\forall x \in \mathbb{R}^d \setminus \{0\}$:

$$x^\top A x \underset{(\geq)}{>} 0.$$

Аналогично определяются *отрицательная определённость (полуопределённость)*.

Замечание C1.11. Множество положительно полуопределённых матриц обозначается $\mathbb{S}_+^d$. Если $A \in \mathbb{S}_+^d$, то $A \succeq 0$. Множество положительно определённых матриц обозначается $\mathbb{S}_{++}^d$. Если $A \in \mathbb{S}_{++}^d$, то $A \succ 0$.

Полезными фактами о квадратичных формах, которые пригодятся нам в дальнейшем, являются критерии Сильвестра, которые позволяют проверять матрицы на положительную/отрицательную определённость и полуопределённость.

Теорема C1.2. (Критерий Сильвестра)

- Квадратичная форма положительно определена $\iff$ все угловые миноры соответствующей ей матрицы положительны.
- Квадратичная форма положительно полуопределена $\iff$ все главные миноры соответствующей ей матрицы неотрицательны. *Главным минором* называется определитель подматрицы, симметричной относительно главной диагонали.

Пример C1.12. Покажите, что матрица

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

положительно определённая.

Решение. Угловой минор порядка 1: $2 > 0$. Угловой минор порядка 2:

$$\det \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} = 5 > 0.$$

Угловой минор порядка 3:

$$\det \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} = 4 > 0.$$

По критерию Сильвестра, матрица A положительно определена. ■

Пример C1.13. Покажите, что матрица

$$A = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

положительно полуопределённая.

Решение. Заметим, что все диагональные элементы неотрицательны. Рассмотрим теперь миноры порядка 2:

$$\det \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = 3 \geq 0,$$

$$\det \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = 1 \geq 0,$$

$$\det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1 \geq 0.$$

Теперь рассмотрим единственный главный минор порядка 3:

$$\det \begin{pmatrix} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = 0.$$

По критерию Сильвестра матрица A положительно полуопределена. Заметим, что она не является положительно определённой. ■

Замечание C1.12. Отметим, что неотрицательности только лишь угловых миноров недостаточно для положительной полуопределённости. Действительно, рассмотрим матрицу

$$A = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}.$$

Нетрудно видеть, что все угловые миноры матрицы равны нулю. Тем не менее, она не является положительно полуопределённой. Действительно,

$$(0,1)A\,(0,1)^\top = (0,1)\,(0,-1)^\top = -1.$$