Методы оптимизации в машинном обучении, ВМК+Физтех, осень 2017

Семинар 1: Скорости сходимости. Матрично-векторные скалярные произведения и нормы

1 Скорости сходимости

Ключевой характеристикой сходящейся последовательности является ее предел. При этом, как известно, разные последовательности могут иметь один и тот же предел. Возникает естественный вопрос: как понять, *насколько быстро* та или иная последовательность сходится к своему пределу? Общепринятой здесь является классификация последовательностей по *скорости сходимости*.

Традиционно скорости сходимости вводятся только для последовательностей неотрицательных чисел, сходящихся к нулю. Чтобы распространить эти понятия на произвольную последовательность объектов $(a_k)_{k=m}^{\infty}$ (например, чисел, многомерных векторов, матриц и т. д.), сходящуюся к пределу a, поступают следующим образом. По $(a_k)_{k=m}^{\infty}$ определяется новая последовательность невязок $(r_k)_{k=m}^{\infty}$, которая состоит из неотрицательных чисел и сходится к нулю. Например, если $(a_k)_{k=m}^{\infty}$ — это последовательность чисел, то в качестве $(r_k)_{k=m}^{\infty}$ обычно берут $r_k := |a_k - a|$ или $r_k := |a_k - a|^2$; аналогичным образом, заменяя модуль на некоторую норму, можно ввести $(r_k)_{k=m}^{\infty}$ для последовательности многомерных векторов или матриц. Говоря о скорости сходимости $(a_k)_{k=m}^{\infty}$, в итоге подразумевают скорость сходимости $(r_k)_{k=m}^{\infty}$.

Базовым понятием является линейная сходимость, которая основана на сравнении последовательности с геометрической прогрессией:

Определение 1.1 (Линейная сходимость)**.** Пусть $(r_k)_{k=m}^{\infty}$ — последовательность неотрицательных чисел. Говорят, что $(r_k)_{k=m}^{\infty}$ *линейно сходится с параметром* $0 < q < 1$, если найдется $C > 0$, такое, что

$$r_k \leq Cq^k$$

для всех $k \geq m$. Если существует хотя бы одно $0 < q < 1$, такое, что $(r_k)_{k=m}^{\infty}$ линейно сходится с параметром q, то говорят, что $(r_k)_{k=m}^{\infty}$ имеет *линейную сходимость*. При этом точная нижняя грань множества всех q, для которых $(r_k)_{k=m}^{\infty}$ линейно сходится с параметром q, называется *константной линейной сходимости* последовательности $(r_k)_{k=m}^{\infty}$.

Замечание 1.1. Поскольку $Cq^k \to 0$ при $k \to \infty$, то из этого определения автоматически следует, что последовательность $(r_k)_{k=m}^{\infty}$ обязана сходиться к нулю. Таким образом, слово «сходимость» в названии означает сходимость последовательности к нулю. Другими словами, если последовательность $(r_k)_{k=m}^{\infty}$ не сходится к нулю, то она не может иметь линейную сходимость.

Пример 1.1. Пусть $(r_k)_{k=1}^{\infty}$ — последовательность $r_k := 1/3^k$, и пусть $(z_k)_{k=1}^{\infty}$ — последовательность $z_k := 4/3^k$. Обе последовательности $(r_k)_{k=1}^{\infty}$ и $(z_k)_{k=1}^{\infty}$ линейно сходятся с любым параметром $1/3 \leq q < 1$. При этом константа линейной сходимости обеих последовательностей равна 1/3. Покажем, это, например, для последовательности $(r_k)_{k=1}^{\infty}$; для второй последовательности это делается аналогично. Предположим противное, т. е. что константа линейной сходимости $(r_k)_{k=1}^{\infty}$ строго меньше 1/3. Тогда, согласно определению, для некоторого $0 < q < 1/3$ найдется $C > 0$, такое, что $1/3^k \leq Cq^k$ для всех $k \geq 1$. Но отсюда следует $C^{-1} \leq (3q)^k$ для всех $k \geq 1$. Переходя к пределу при $k \to \infty$ получаем, что $C^{-1} \leq 0$ (поскольку $0 \leq 3q < 1$), что невозможно.

Упражнение 1.1. Покажите, используя определение, что последовательность $(1/k)_{k=1}^{\infty}$, хотя и сходится к нулю, линейной сходимостью не обладает.

Следующее утверждение показывает, что линейная сходимость не зависит от конечного числа начальных элементов последовательности:

Утверждение 1.1. *Пусть $(r_k)_{k=m}^{\infty}$ — последовательность неотрицательных чисел, и пусть $s \geq 0$ — целое число. Тогда последовательность $(r_k)_{k=m+s}^{\infty}$ сходится линейно с параметром $0 < q < 1$, если и только если последовательность $(r_k)_{k=m}^{\infty}$ также сходится линейно с параметром q. (В частности, константы линейной сходимости последовательностей $(r_k)_{k=m}^{\infty}$ и $(r_k)_{k=m+s}^{\infty}$ совпадают.)*

Доказательство. В обратную сторону утверждение является очевидным. Поэтому проведем доказательство лишь в прямую сторону. Поскольку при $s = 0$ утверждение является бессодержательным, будем считать, что $s \geq 1$.

Пусть $(r_k)_{k=m+s}^{\infty}$ сходится линейно с параметром q. Тогда, согласно определению, найдется $C > 0$, такое, что $r_k \leq Cq^k$ для всех $k \geq m + s$. Положим

$$C' := \max\left\{C, \frac{r_m}{q^m}, \dots, \frac{r_{m+s-1}}{q^{m+s-1}}\right\}.$$

Тогда нетрудно видеть, что $r_k \leq C'q^k$ для всех $k \geq m$. Таким образом, $(r_k)_{k=m}^{\infty}$ сходится линейно с параметром q. □

Теперь перейдем к рассмотрению двух других типов сходимости.

Определение 1.2 (Сублинейная и сверхлинейная сходимости)**.** Пусть $(r_k)_{k=m}^{\infty}$ — последовательность неотрицательных чисел. Если $(r_k)_{k=m}^{\infty}$ сходится к нулю, но при этом не обладает линейной сходимостью, то говорят, что $(r_k)_{k=m}^{\infty}$ имеет *сублинейную сходимость*. Если же, наоборот, $(r_k)_{k=m}^{\infty}$ обладает линейной сходимостью, и при этом константа линейной сходимости равна 0, то говорят, что имеет место *сверхлинейная сходимость*.

Говоря неформально, сублинейная сходимость означает, что последовательность сходится медленнее любой (даже «самой медленной») геометрической прогрессии; сверхлинейная сходимость, наоборот, означает, что последовательность сходится быстрее любой (даже «самой быстрой») геометрической прогрессии.

Пример 1.2. Последовательность $(1/k)_{k=1}^{\infty}$ имеет сублинейную сходимость (согласно упражнению 1.1). Последовательность $(1/3^{k^2})_{k=1}^{\infty}$ имеет сверхлинейную сходимость. Действительно, пусть $0 < q < 1$ — произвольное число. Поскольку $1/3^k \to 0$ при $k \to \infty$, то найдется $N \geq 1$, такое, что $1/3^k \leq q$ для всех $k \geq N$. Отсюда $1/3^{k^2} = (1/3^k)^k \leq q^k$ для всех $k \geq N$. Но это означает, что $(1/3^{k^2})_{k=1}^{\infty}$ сходится линейно с параметром q (согласно утверждению 1.1, не важно, что неравенство выполнено только с номера N). В силу произвольности q, это означает, что константа линейной сходимости $(1/3^{k^2})_{k=1}^{\infty}$ равна 0.

Замечание 1.2. Согласно введенным определениям, если последовательность имеет сверхлинейную сходимость, то она также имеет и линейную сходимость. Таким образом, класс сверхлинейно сходящихся последовательностей является подклассом линейно сходящихся последовательностей. При желании эти классы можно было бы отделить друг от друга, введя понятие *истинно линей-*

ной сходимости как линейной сходимости с константой, отличной от нуля. Однако мы не будем этого делать.

Из утверждения 1.1 сразу же вытекает аналогичное утверждение относительно сублинейной и сверхлинейной сходимостей:

Следствие 1.1. *Пусть $(r_k)_{k=m}^{\infty}$ — последовательность неотрицательных чисел, и пусть $s \geq 0$ — целое число. Тогда последовательность $(r_k)_{k=m+s}^{\infty}$ имеет сублинейную (соответственно сверхлинейную) сходимость, если и только если последовательность $(r_k)_{k=m}^{\infty}$ также имеет сублинейную (соответственно сверхлинейную) сходимость.*

Таким образом, тип сходимости не меняется при удалении из последовательности конечного числа начальных элементов.

Скорость сходимости удобно определять с помощью следующего теста, аналогичного признаку Коши сходимости числовых рядов:

Теорема 1.1 (Тест корней)**.** *Пусть $(r_k)_{k=m}^{\infty}$ — последовательность неотрицательных чисел, сходящаяся к нулю, и пусть $\alpha := \limsup_{k\to\infty} r_k^{1/k}$. (Заметим, что $\alpha \geq 0$.)*

(a) Если $0 \leq \alpha < 1$, то $(r_k)_{k=m}^{\infty}$ имеет линейную сходимость с константной α.

(b) В частности, если $\alpha = 0$, то $(r_k)_{k=m}^{\infty}$ имеет сверхлинейную сходимость.

(c) Если $\alpha = 1$, то $(r_k)_{k=m}^{\infty}$ имеет сублинейную сходимость.

(d) Случай $\alpha > 1$ невозможен.

Доказательство. Во-первых, покажем, что, если $(r_k)_{k=m}^{\infty}$ имеет линейную сходимость с константой $0 \leq \beta < 1$, то непременно $\alpha \leq \beta$. Действительно, по определению константы линейной сходимости, для любого $\varepsilon > 0$, удовлетворяющего $\beta + \varepsilon < 1$, найдется $C > 0$, такое, что $r_k \leq C(\beta + \varepsilon)^k$ для всех $k \geq m$. Отсюда $r_k^{1/k} \leq C^{1/k}(\beta + \epsilon)$ для всех $k \geq m$. Переходя к верхнему пределу и используя $C^{1/k} \to 1$ при $k \to \infty$, получаем $\alpha \leq \beta + \varepsilon$. В силу произвольности ε, отсюда следует, что $\alpha \leq \beta$.

Таким образом, в случае $\alpha = 1$ последовательность $(r_k)_{k=m}^{\infty}$ не может иметь линейную сходимость согласно установленному выше результату (доказывается от противного). Поскольку, тем не менее, $(r_k)_{k=m}^{\infty}$ сходится к нулю, то $(r_k)_{k=m}^{\infty}$ имеет сублинейную сходимость.

Рассмотрим теперь случай $0 \leq \alpha < 1$. Пусть $\varepsilon > 0$ — произвольное число, такое, что $\alpha + \varepsilon < 1$. Согласно свойствам верхнего предела, найдется $N \geq m$, такое, что $r_k^{1/k} \leq \alpha + \varepsilon$ для всех $k \geq N$. Отсюда $r_k \leq (\alpha + \varepsilon)^k$ для всех $k \geq N$. Таким образом, $(r_k)_{k=m}^{\infty}$ сходится линейно с параметром $\alpha + \varepsilon$ (согласно утверждению 1.1, не важно, что неравенство выполнено только с номера N). В силу произвольности ε, это означает, что константа линейной сходимости $(r_k)_{k=m}^{\infty}$ не превосходит α. Поскольку, как было показано выше, константа линейной сходимости не может быть меньше α, это означает, что константа линейной сходимости $(r_k)_{k=m}^{\infty}$ в точности равна α.

Наконец, покажем, что случай $\alpha > 1$ невозможен. Действительно, пусть $\alpha > 1$. Тогда из определения верхнего предела следует, что для любого $N \geq m$ найдется $k \geq N$, такое, что $r_k^{1/k} \geq 1$, и, в частности, $r_k \geq 1$. Но это означает, что r_k имеет подпоследовательность, которая отделена от нуля. Значит, $(r_k)_{k=m}^{\infty}$ не может сходиться к нулю, что противоречит условию. □

Замечание 1.3. В отличие от признака Коши для сходимости числовых рядов, в данном случае тест корня не имеет ситуаций, в которых ответ может быть неоднозначным.

Иногда бывает проще вычислить предел отношений, чем предел корней. Следующая лемма показывает, как связаны эти два предела:

Лемма 1.1. *Пусть* $(r_k)_{k=m}^{\infty}$ *— последовательность строго положительных чисел. (Строгая положительность необходима для того, чтобы отношения* r_{k+1}/r_k*, появляющиеся ниже, были корректно определены.) Тогда*

$$\liminf_{k\to\infty}\frac{r_{k+1}}{r_k}\le\liminf_{k\to\infty}r_k^{1/k}\le\limsup_{k\to\infty}r_k^{1/k}\le\limsup_{k\to\infty}\frac{r_{k+1}}{r_k}.$$

Доказательство. Среднее неравенство следует из того, что нижний предел любой последовательности всегда не превосходит ее верхнего предела. Докажем последнее неравенство; первое доказывается аналогично.

Обозначим $L:=\limsup_{k\to\infty}\frac{r_{k+1}}{r_k}$. Если $L=+\infty$, то неравенство, очевидно, верное, поэтому далее будем считать, что L конечное. Заметим, что $L\ge 0$, поскольку отношение $\frac{r_{k+1}}{r_k}$ положительное для всех $k\ge m$. Пусть $\varepsilon>0$ — произвольное число. Согласно свойствам верхнего предела, найдется $N\ge m$, такое, что $\frac{r_{k+1}}{r_k}\le L+\varepsilon$ для всех $k\ge N$. Отсюда $r_{k+1}\le(L+\varepsilon)r_k$ для всех $k\ge N$. Применяя индукцию, получаем, что $r_k\le(L+\varepsilon)^{k-N}r_N$ для всех $k\ge N$. Обозначим $C:=(L+\varepsilon)^{-N}r_N$. Тогда $r_k\le C(L+\varepsilon)^k$ для всех $k\ge N$, откуда $r_k^{1/k}\le C^{1/k}(L+\epsilon)$. Переходя к верхнему пределу и используя $C^{1/k}\to 1$ при $k\to\infty$, получаем $\limsup_{k\to\infty}r_k^{1/k}\le L+\varepsilon$. В силу произвольности ε, отсюда следует $\limsup_{k\to\infty}r_k^{1/k}\le L$. $\square$

Следствием теоремы 1.1 и леммы 1.1 является тест отношений, который можно рассматривать как аналог признака Даламбера сходимости числовых рядов.

Следствие 1.2 (Тест отношений)**.** *Пусть* $(r_k)_{k=m}^{\infty}$ *— последовательность строго положительных чисел, сходящаяся к нулю.*

(a) Если существует $\alpha:=\lim_{k\to\infty}\frac{r_{k+1}}{r_k}$ *и при этом* $0\le\alpha<1$*, то* $(r_k)_{k=m}^{\infty}$ *имеет линейную сходимость с константой* α*.*

(b) В частности, если $\lim_{k\to\infty}\frac{r_{k+1}}{r_k}=0$*, то* $(r_k)_{k=m}^{\infty}$ *имеет сверхлинейную сходимость.*

(c) Если $\lim_{k\to\infty}\frac{r_{k+1}}{r_k}$ *не существует, но при этом* $q:=\limsup_{k\to\infty}\frac{r_{k+1}}{r_k}<1$*, то* $(r_k)_{k=m}^{\infty}$ *имеет линейную сходимость с константой, не превосходящей* q*.*

(d) Если $\liminf_{k\to\infty}\frac{r_{k+1}}{r_k}=1$*, то* $(r_k)_{k=m}^{\infty}$ *имеет сублинейную сходимость.*

(e) Ситуация $\liminf_{k\to\infty}\frac{r_{k+1}}{r_k}>1$ *невозможна.*

(f) Во всех остальных случаях (т. е. когда $\liminf_{k\to\infty}\frac{r_{k+1}}{r_k}<1\le\limsup_{k\to\infty}\frac{r_{k+1}}{r_k}$*) нельзя утверждать что-либо конкретное о скорости сходимости* $(r_k)_{k=m}^{\infty}$*.*

Упражнение 1.2. Определите скорость сходимости последовательности $(1/k!)_{k=1}^{\infty}$.

В отличие от теста корней, в тесте отношений возможны случаи, когда ответ может быть неоднозначным:

Упражнение 1.3. Пусть $(r_k)_{k=1}^{\infty}$, $(z_k)_{k=1}^{\infty}$, $(w_k)_{k=1}^{\infty}$ — последовательности

$$r_k:=\begin{cases}\frac{1}{3^k}, & k \text{ нечетное},\\ \frac{1}{3^{2k}}, & \text{иначе},\end{cases}\qquad z_k:=\begin{cases}\frac{1}{k}, & k \text{ нечетное},\\ \frac{1}{k^2}, & \text{иначе},\end{cases}\qquad w_k:=\begin{cases}\frac{1}{k^k}, & k \text{ нечетное},\\ \frac{1}{k^{2k}}, & \text{иначе}.\end{cases}$$

Покажите, что

$$\liminf_{k\to\infty} \frac{r_{k+1}}{r_k} = \liminf_{k\to\infty} \frac{z_{k+1}}{z_k} = \liminf_{k\to\infty} \frac{w_{k+1}}{w_k} = 0$$

и

$$\limsup_{k\to\infty} \frac{r_{k+1}}{r_k} = \limsup_{k\to\infty} \frac{z_{k+1}}{z_k} = \limsup_{k\to\infty} \frac{w_{k+1}}{w_k} = +\infty,$$

в то время как $(r_k)_{k=1}^{\infty}$ имеет линейную сходимость, $(z_k)_{k=1}^{\infty}$ имеет сублинейную сходимость, $(w_k)_{k=1}^{\infty}$ имеет сверхлинейную сходимость. (*Подсказка*: используйте тест корней.)

Помимо рассмотренных линейной, сублинейной и сверхлинейной сходимостей традиционно также выделяют еще один тип сходимости:

Определение 1.3 (Сходимость порядка p)**.** Пусть $(r_k)_{k=m}^{\infty}$ — последовательность неотрицательных чисел, сходящаяся к нулю, и пусть $p > 1$. Говорят, что $(r_k)_{k=m}^{\infty}$ имеет *сходимость порядка* p, если найдется $M > 0$, такое, что

$$r_{k+1} \leq M r_k^p$$

для всех $k \geq m$. Сходимость порядка 2 называется *квадратичной сходимостью.*

Пример 1.3. Пусть $p > 1$, и пусть $(r_k)_{k=1}^{\infty}$ — последовательность $r_k := p^{p^k}$. Поскольку $r_{k+1} = r_k^p$ для всех $k \geq 1$, то, по определению, $(r_k)_{k=1}^{\infty}$ имеет сходимость порядка p. В частности, последовательность $(2^{2^k})_{k=1}^{\infty}$ имеет квадратичную сходимость.

Класс последовательностей, имеющих сходимость порядка p, является подклассом сверхлинейно сходящихся последовательностей. Другими словами, сходимость порядка p — это частный случай сверхлинейной сходимости.

Упражнение 1.4. Покажите, что, если последовательность имеет сходимость порядка p, то она сходится сверхлинейно. (*Подсказка*: используйте тест отношений.)

Упражнение 1.5. Пусть $(r_k)_{k=m}^{\infty}$ — последовательность неотрицательных чисел, сходящаяся к нулю, и пусть $s \geq 0$ — целое число. Покажите, что последовательность $(r_k)_{k=m+s}^{\infty}$ имеет сходимость порядка $p > 1$, если и только если последовательность $(r_k)_{k=m}^{\infty}$ также имеет сходимость порядка p.

Упражнение 1.6 (Тест отношений для сходимости порядка p)**.** Пусть $(r_k)_{k=m}^{\infty}$ — последовательность строго положительных чисел, и пусть $p > 1$. Покажите, что $(r_k)_{k=m}^{\infty}$ имеет сходимость порядка p, если и только если $\limsup_{k\to\infty} \frac{r_{k+1}}{r_k^p} < +\infty$.

Упражнение 1.7. Пусть $M > 0$, $r_0 \geq 0$, и пусть $(r_k)_{k=0}^{\infty}$ — последовательность, определенная рекуррентно по правилу $r_{k+1} := M r_k^2$ для $k \geq 0$. Установите необходимое и достаточное условие для M и r_0, при котором последовательность $(r_k)_{k=0}^{\infty}$ будет сходиться к нулю.

2 Матрично-векторные скалярные произведения

В дальнейшем мы регулярно будем пользоваться различными матрично-векторными скалярными произведениями и нормами. В связи с этим, кратко напомним основные понятия и факты из этой области.

Всюду в дальнейшем будут использоваться следующие стандартные обозначения:

(a) $\mathbb{R}$ обозначает множество вещественных чисел;

(b) $\mathbb{R}^n$ обозначает множество всех n-мерных вещественных вектор-столбцов;

(c) $\mathbb{R}^{m\times n}$ обозначает множество всех вещественных матриц с m строками и n столбцами;

(d) $\mathbb{S}^n$ обозначает множество всех $n \times n$ вещественных симметричных матриц.

(e) $\mathbb{S}^n_+$ и $\mathbb{S}^n_{++}$ обозначают множество всех $n \times n$ вещественных симметричных положительно полуопределенных и положительно определенных матриц соответственно;

(f) I_n обозначает единичную матрицу размера n.

Заметим, что под векторами из $\mathbb{R}^n$ всюду будут подразумеваться именно вектор-столбцы (а не, например, вектор-строки); таким образом, $\mathbb{R}^n = \mathbb{R}^{n\times 1}$, но $\mathbb{R}^n \neq \mathbb{R}^{1\times n}$. Напомним, что $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{m\times n}$ и $\mathbb{S}^n$ являются вещественными векторными пространствами (со стандартными операциями сложения и умножения на число).

Для матрицы $A \in \mathbb{R}^{n\times n}$ символ $\mathrm{Tr}(A) := \sum_{i=1}^n a_{ii}$ обозначает ее след.

Упражнение 2.1 (Циклическое свойство следа)**.** Покажите, что для любых матриц $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times m}$ выполнено

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA).$$

Прежде, чем переходить к конкретным примерам, напомним общее определение скалярного произведения.

Определение 2.1 (Скалярное произведение)**.** Пусть V — вещественное векторное пространство. Функция $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$, которая каждой паре x, y векторов в V ставит в соответствие вещественное число $\langle x, y\rangle$, называется вещественным *скалярным произведением*, если она удовлетворяет следующим аксиомам:

(a) (Положительность) Для любого $x \in V$ выполнено $\langle x, x\rangle \geq 0$. Более того, $\langle x, x\rangle = 0$ тогда и только тогда, когда $x = 0$.

(b) (Симметричность) Для любых $x, y \in V$ выполнено $\langle x, y\rangle = \langle y, x\rangle$.

(c) (Линейность) Для любых $x, y, z \in V$ выполнено $\langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$. Для любых $x, y \in V$ и любого $\alpha \in \mathbb{R}$ выполнено $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$.

Векторное пространство с заданным на нем скалярным произведением называется *пространством со скалярным произведением* или *предгильбертовым пространством*. Конечномерное вещественное пространство со скалярным произведением также называют *евклидовым пространством*.

Пример 2.1 (Стандартное скалярное произведение в $\mathbb{R}$)**.** В пространстве $\mathbb{R}$ скалярное произведение можно ввести как обычное произведение чисел: $\langle x, y\rangle := xy$. Это скалярное произведение называется *стандартным скалярным произведением в* $\mathbb{R}$.

Пример 2.2. Стандартное скалярное произведение в $\mathbb{R}$ не является единственно возможным. Например, в $\mathbb{R}$ также можно ввести нестандартное скалярное произведение $\langle x, y\rangle' := 5xy$, которое отличается от стандартного скалярного произведения лишь постоянным множителем. Оказывается, что таким образом устроено любое скалярное произведение в $\mathbb{R}$ (см. упражнение 2.2).

Упражнение 2.2. Пусть $\langle \cdot, \cdot\rangle'$ — произвольное скалярное произведение в $\mathbb{R}$. Покажите, что для некоторого $a > 0$ выполнено $\langle x, y\rangle' = axy$ для всех $x, y \in \mathbb{R}$. Таким образом, с точностью до постоянного множителя стандартное скалярное произведение является единственно возможным в $\mathbb{R}$. (*Подсказка*: положите $a := \langle 1, 1\rangle'$.)

Пример 2.3 (Стандартное скалярное произведение в $\mathbb{R}^n$)**.** В пространстве $\mathbb{R}^n$ вещественных n-мерных вектор-столбцов *стандартное скалярное произведение* задается формулой

$$\langle x, y\rangle := \mathrm{Tr}(x^T x) = \sum_{i=1}^{n} x_i y_i.$$

Упражнение 2.3 (Общий вид скалярного произведения в $\mathbb{R}^n$)**.** Снова рассмотрим пространство $\mathbb{R}^n$ со стандартным скалярным произведением $\langle \cdot, \cdot\rangle$.

(a) Пусть $A \in \mathbb{S}^n_{++}$ — симметричная положительно определенная матрица. Покажите, что функция $\langle \cdot, \cdot\rangle_A : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, определенная по формуле

$$\langle x, y\rangle_A := \langle Ax, y\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i y_j,$$

задает скалярное произведение в $\mathbb{R}^n$.

(b) Покажите, что любое скалярное произведение в $\mathbb{R}^n$ обязательно имеет указанный выше вид для некоторой матрицы $A \in \mathbb{S}^n_{++}$. (*Подсказка*: рассмотрите произвольный базис $e_1, \dots, e_n$ в $\mathbb{R}^n$ и разложите векторы x, y по этому базису.)

Пример 2.4 (Стандартное скалярное произведение в $\mathbb{R}^{m\times n}$)**.** В пространстве $\mathbb{R}^{m\times n}$ матриц можно ввести *фробениусово скалярное произведение*

$$\langle A, B\rangle := \mathrm{Tr}(A^T B) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}.$$

Это скалярное произведение называется *стандартным скалярным произведением в* $\mathbb{R}^{m\times n}$.

Замечание 2.1. Напомним, что, согласно договоренности, сделанной в самом начале, $\mathbb{R}^{n\times 1} = \mathbb{R}^n$. Таким образом, в примере 2.4 было переопределено стандартное скалярное произведение в $\mathbb{R}^n$, введенное в примере 2.3. Однако нетрудно видеть, что никакой проблемы в этом нет, поскольку оба определения дают одинаковый результат.

Пример 2.5. Пусть V — вещественное векторное пространство со скалярным произведением $\langle \cdot, \cdot \rangle : V \times V$, и пусть U — подпространство V. Тогда сужение $\langle \cdot, \cdot \rangle|_{U\times U}$ скалярного произведения $\langle \cdot, \cdot \rangle$ задает скалярное произведение в U. Таким образом, скалярное произведение можно наследовать на подпространство.

Пример 2.6 (Стандартное скалярное произведение в $\mathbb{S}^n$)**.** Наследуя фробениусово скалярное произведение из пространства $\mathbb{R}^{n\times n}$ на подпространство симметричных матриц $\mathbb{S}^n$, получаем *фробениусово скалярное произведение в* $\mathbb{S}^n$:

$$\langle A, B \rangle := \operatorname{Tr}(AB).$$

(В отличие от примера 2.4, знак транспонирования можно опустить в силу симметричности.) Это скалярное произведение называется *стандартным скалярным произведением в* $\mathbb{S}^n$.

В дальнейшем, используя обозначение $\langle x, y \rangle$, где x, y являются объектами одного из пространств $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{m\times n}$, $\mathbb{S}^n$, если не оговорено иное, будем иметь в виду именно соответствующее стандартное скалярное произведение из примеров 2.1, 2.3, 2.4, 2.6.

При работе со стандартными матрично-векторными скалярными произведениями часто оказываются полезными следующие свойства сопряженности, которые связывают между собой скалярные произведения в различных пространствах:

Утверждение 2.1 (Тождества сопряженности)**.** *Для любых матриц* $A \in \mathbb{R}^{m\times k}$, $B \in \mathbb{R}^{k\times n}$, $C \in \mathbb{R}^{m\times n}$ *имеет место*

$$\langle AB, C \rangle = \langle B, A^T C \rangle = \langle A, CB^T \rangle.$$

Доказательство. Согласно определению, $\langle AB, C \rangle = \operatorname{Tr}((AB)^T C)$. Поскольку $(AB)^T = B^T A^T$, то $\operatorname{Tr}((AB)^T C) = \operatorname{Tr}(B^T A^T C)$. Но, опять же, по определению, $\operatorname{Tr}(B^T A^T C) = \langle B, A^T C \rangle$, что доказывает первое равенство. Для доказательства второго равенства остается воспользоваться циклическим свойством следа (см. упражнение 2.1), чтобы получить $\operatorname{Tr}(B^T A^T C) = \operatorname{Tr}(A^T C B^T)$. □

Упражнение 2.4. Пусть $x, y \in \mathbb{R}^n$. Покажите, что $\langle xx^T, yy^T \rangle = \langle x, y \rangle^2$.

В заключение отметим крайне важное неравенство Коши–Буняковского, которое в общем случае справедливо для произвольного скалярного произведения в произвольном векторном пространстве.

Утверждение 2.2 (Неравенство Коши–Буняковского)**.** *Пусть V — вещественное векторное пространство со скалярным произведением $\langle \cdot, \cdot \rangle$. Тогда для любых $x, y \in V$ справедливо*

$$|\langle x, y \rangle| \le \langle x, x \rangle^{1/2} \langle y, y \rangle^{1/2}, \tag{2.1}$$

причем неравенство переходит в равенство тогда и только тогда, когда либо $y = 0$, либо $x = \alpha y$ для некоторого $\alpha \in \mathbb{R}$.

Доказательство. Для $y = 0$ утверждение, очевидно, верное. Поэтому далее будем считать, что $y \neq 0$.

Пусть $\alpha \in \mathbb{R}$ — произвольное число (которое будет выбрано позже). Согласно аксиомам скалярного произведения, имеем

$$0 \leq \langle x - \alpha y, x - \alpha y\rangle = \langle x, x\rangle - 2\alpha\langle x, y\rangle + \alpha^2\langle y, y\rangle. \tag{2.2}$$

Отсюда

$$2\alpha\langle x, y\rangle \leq \langle x, x\rangle + \alpha^2\langle y, y\rangle.$$

Поскольку $y \neq 0$, то $\langle y, y\rangle \neq 0$. Полагая $\alpha := \langle x, y\rangle/\langle y, y\rangle$, получаем

$$\langle x, y\rangle^2 \leq \langle x, x\rangle\langle y, y\rangle,$$

что дает (2.1) после извлечения корня.

Нетрудно видеть, что неравенство (2.1) переходит в равенство тогда и только тогда, когда неравенство (2.2) переходит в равенство. Согласно аксиомам скалярного произведения, последнее возможно, если и только если $x = \alpha y$ (где $\alpha = \langle x, y\rangle/\langle y, y\rangle$). □

В частности, рассматривая пространство $\mathbb{R}^n$ со стандартным скалярным произведением, получаем классическое неравенство Коши–Буняковского для конечных сумм:

Следствие 2.1 (Неравенство Коши–Буняковского для конечных сумм)**.** *Пусть $x_1, \ldots, x_n \in \mathbb{R}$ и $y_1, \ldots, y_n \in \mathbb{R}$. Тогда*

$$\left|\sum_{i=1}^n x_i y_i\right| \leq \left(\sum_{i=1}^n x_i^2\right)^{1/2} \left(\sum_{i=1}^n y_i^2\right)^{1/2},$$

причем неравенство переходит в равенство, если и только если $y_1 = \cdots = y_n = 0$ или $x_1 = \alpha y_1, \ldots, x_n = \alpha y_n$ для некоторого $\alpha \in \mathbb{R}$.

3 Матрично-векторные нормы

Теперь перейдем к рассмотрению матрично-векторных норм. Как и раньше, начнем с общего определения.

Определение 3.1 (Норма)**.** Пусть V — вещественное векторное пространство. Функция $\|\cdot\| : V \to [0, +\infty)$, которая каждому вектору $x \in V$ ставит в соответствие неотрицательное вещественное число $\|x\|$, называется *нормой*, если она удовлетворяет следующим аксиомам:

(a) (Положительность) Для любого $x \in V$ выполнено $\|x\| \geq 0$. Более того, $\|x\| = 0$ тогда и только тогда, когда $x = 0$.

(b) (Абсолютная однородность) Для любого $x \in V$ и любого $\alpha \in \mathbb{R}$ выполнено $\|\alpha x\| = |\alpha|\|x\|$.

(c) (Неравенство треугольника) Для любых $x, y \in V$ выполнено $\|x + y\| \leq \|x\| + \|y\|$.

Векторное пространство с заданной на нем нормой называется *нормированным пространством.*

Пример 3.1 (Стандартная норма в $\mathbb{R}$)**.** В пространстве $\mathbb{R}$ норму можно ввести как модуль числа: $\|x\| := |x|$. Эта норма называется *стандартной нормой в* $\mathbb{R}$. Как показывает упражнение 3.1, с точностью до постоянного множителя стандартная норма является единственно возможной нормой в пространстве $\mathbb{R}$.

Упражнение 3.1. Покажите, что любая норма $\|\cdot\|'$ в пространстве $\mathbb{R}$ обязательно имеет вид $\|x\|' = a|x|$ для некоторого $a > 0$. (*Подсказка*: положите $a := |1|$.)

Пример 3.2 (Евклидова норма)**.** Если V — вещественное векторное пространство со скалярным произведением $\langle\cdot,\cdot\rangle$, то в V можно ввести *евклидову норму* $\|x\| := \langle x, x\rangle^{1/2}$. Таким образом, любое пространство со скалярным произведением может быть превращено в нормированное пространство с помощью введения евклидовой нормы.

Пример 3.3 (Стандартная евклидова норма в $\mathbb{R}^n$)**.** Рассматривая в примере 3.2 в качестве V пространство $\mathbb{R}^n$ со стандартным скалярным произведением, получаем *стандартную евклидову норму*

$$\|x\|_2 := \langle x, x\rangle^{1/2} = \left(\sum_{i=1}^n x_i^2\right)^{1/2}.$$

Эта норма также известна как *l^2-норма*. В дальнейшем для краткости стандартную евклидову норму на $\mathbb{R}^n$ будем обозначать символом $|\cdot|$ (таким образом, $|x| := \|x\|_2$ для $x \in \mathbb{R}^n$).

Евклидова норма в $\mathbb{R}^n$ соответствует геометрическому представлению о *длине* вектора. В частности, при повороте (ортогональном преобразовании) вектора его евклидова норма не изменяется:

Упражнение 3.2. Пусть $x \in \mathbb{R}^n$, и пусть $Q \in \mathbb{R}^{n\times n}$ — ортогональная матрица (т. е. $Q^TQ = QQ^T = I_n$). Покажите, что $|Qx| = |x|$.

Пример 3.4 (Общий вид евклидовой нормы в $\mathbb{R}^n$)**.** Снова вернемся к примеру 3.2 и рассмотрим в качестве V пространство $\mathbb{R}^n$, но на этот раз рассмотрим нестандартное скалярное произведение $\langle x, y\rangle_A := \langle Ax, y\rangle$, где $A \in \mathbb{S}^n_{++}$ — симметричная положительно определенная матрица (см. упражнение 2.3). Тогда функция $\|\cdot\|_A : \mathbb{R}^n \to \mathbb{R}$, определенная по формуле

$$\|x\|_A := \langle Ax, x\rangle^{1/2} = \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j\right)^{1/2},$$

задает нестандартную евклидову норму в пространстве $\mathbb{R}^n$ (которая переходит в стандартную евклидову норму, если $A = I_n$). Согласно упражнению 2.3, любая евклидова норма в пространстве $\mathbb{R}^n$ обязательно имеет указанный вид.

Пример 3.5 (l^1-норма)**.** Помимо евклидовой нормы в пространстве $\mathbb{R}^n$ также можно задать и неевклидову норму. Например, функция $\|\cdot\|_1 : \mathbb{R}^n \to \mathbb{R}$, определенная по формуле

$$\|x\|_1 := \sum_{i=1}^n |x_i|,$$

задает норму, которая называется l^1-*нормой*.

Пример 3.6 (l^∞-норма)**.** Еще одной популярной нормой в пространстве $\mathbb{R}^n$ является l^∞-*норма*

$$\|x\|_\infty := \max_{1\le i\le n} |x_i|.$$

Эта норма также известна как *равномерная норма* или *норма Чебышева*.

Следующее упражнение показывает, как связаны между собой нормы l^2, l^1 и l^∞:

Упражнение 3.3. Пусть $x \in \mathbb{R}^n$. Докажите следующие неравенства:

$$\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\|x\|_\infty \qquad \text{и} \qquad \frac{1}{\sqrt{n}}\|x\|_1 \le \|x\|_2 \le \|x\|_1.$$

(*Подсказка*: для первой части второго неравенства используйте неравенство Коши–Буняковского.)

Замечание 3.1. Нормы l^2, l^1 и l^∞, рассмотренные в примерах 3.3, 3.5 и 3.6, являются частными случаями более общего семейства l^p-*норм*

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

где $p \in [1, +\infty]$ (при $p = +\infty$ правая часть полагается равной соответствующему пределу при $p \to +\infty$).

Пример 3.7 (Фробениусова норма в $\mathbb{R}^{m\times n}$)**.** Снова вернемся к примеру 3.2 и рассмотрим теперь в качестве V пространство матриц $\mathbb{R}^{m\times n}$ со стандартным скалярным произведением. Соответствующая евклидова норма в данном случае называется *фробениусовой нормой* и задается формулой

$$\|A\|_F := \langle A, A\rangle^{1/2} = [\mathrm{Tr}(A^T A)]^{1/2} = \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2\right)^{1/2}.$$

Эта норма также известна как *норма Гильберта–Шмидта*.

Упражнение 3.4. Пусть $A \in \mathbb{R}^{m\times n}$, и пусть $Q_1 \in \mathbb{R}^{m\times m}$, $Q_2 \in \mathbb{R}^{n\times n}$ — ортогональные матрицы (т. е. $Q_1^T Q_1 = Q_1 Q_1^T = I_m$ и $Q_2^T Q_2 = Q_2 Q_2^T = I_n$). Покажите, что

$$\|Q_1 A\|_F = \|A Q_2\|_F = \|A\|_F.$$

Таким образом, фробениусова норма инвариантна к ортогональным преобразованиям.

Упражнение 3.5. Пусть $A \in \mathbb{R}^{m\times n}$, и пусть $q := \min\{m, n\}$. Покажите, что

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

где $\sigma_1(A) \geq \cdots \geq \sigma_q(A) \geq 0$ — сингулярные числа матрицы A. (*Подсказка*: воспользуйтесь сингулярным разложением и упражнением 3.4.)

Пример 3.8. Пусть V — вещественное векторное пространство с определенной на нем нормой $\|\cdot\| : V \to \mathbb{R}$, и пусть U — подпространство V. Тогда сужение $\|\cdot\||_U$ нормы $\|\cdot\|$ на подпространство U задает норму в этом подпространстве. Таким образом, норму можно наследовать на подпространство.

Пример 3.9 (Фробениусова норма в $\mathbb{S}^n$)**.** Наследуя фробениусову норму из пространства $\mathbb{R}^{n\times n}$ на подпространство симметричных матриц $\mathbb{S}^n$, получаем *фробениусову норму в пространстве* $\mathbb{S}^n$:

$$\|A\|_F := \langle A, A\rangle^{1/2} = [\mathrm{Tr}(A^2)]^{1/2}.$$

Упражнение 3.6. Пусть $A \in \mathbb{S}^n$. Покажите, что

$$\|A\|_F = \left(\sum_{i=1}^{n} \lambda_i^2(A)\right)^{1/2},$$

где $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ — собственные значения матрицы A. (См. также упражнение 3.5.)

Пример 3.10 (Операторная норма в $\mathbb{R}^{m\times n}$)**.** Помимо фробениусовой нормы важным примером матричной нормы в пространстве $\mathbb{R}^{m\times n}$ является *операторная норма*:

$$\|A\|_{\mathrm{op}} := \max_{x\in\mathbb{R}^n : |x|=1} |Ax|.$$

Эта норма также известна как *спектральная норма*.

Упражнение 3.7. Пусть $v \in \mathbb{R}^n = \mathbb{R}^{n\times 1}$. Покажите, что $\|v\|_{\mathrm{op}} = |v|$.

Упражнение 3.8. Пусть D — диагональная матрица с элементами $d_1, \ldots, d_n \in \mathbb{R}$ на диагонали. Покажите, что $\|D\|_{\mathrm{op}} = \max_{1\leq i\leq n} |d_i|$.

Упражнение 3.9. Пусть $A \in \mathbb{R}^{m\times n}$, и пусть $Q_1 \in \mathbb{R}^{m\times m}$, $Q_2 \in \mathbb{R}^{n\times n}$ — ортогональные матрицы. Покажите, что

$$\|Q_1 A\|_{\text{op}} = \|AQ_2\|_{\text{op}} = \|A\|_{\text{op}}.$$

Таким образом, операторная норма инвариантна к ортогональным преобразованиям. (*Подсказка*: используйте результат упражнения 3.2.)

Упражнение 3.10. Пусть $A \in \mathbb{R}^{m\times n}$. Покажите, что $\|A\|_{\text{op}} = \sigma_{\max}(A)$, где $\sigma_{\max}(A)$ — максимальное сингулярное число матрицы A. (*Подсказка*: используйте сингулярное разложение и упражнения 3.9 и 3.8.)

Следующее упражнение показывает, как связаны операторная норма и фробениусова норма, рассмотренная ранее:

Упражнение 3.11. Пусть $A \in \mathbb{R}^{m\times n}$. Покажите, что

$$\|A\|_{\text{op}} \leq \|A\|_F \leq \sqrt{\min\{m,n\}}\|A\|_{\text{op}}. \tag{3.1}$$

(*Подсказка*: воспользуйтесь результатами упражнений 3.5 и 3.10.)

Пример 3.11 (Операторная норма в $\mathbb{S}^n$)**.** Наследуя операторную норму из пространства $\mathbb{R}^{n\times n}$ на подпространство $\mathbb{S}^n$, получаем *операторную норму в пространстве* $\mathbb{S}^n$:

$$\|A\|_{\text{op}} := \max_{x\in\mathbb{R}^n:|x|=1} |Ax|.$$

Упражнение 3.12. Пусть $A \in \mathbb{S}^n$. Покажите, что

$$\|A\|_{\text{op}} = \max\{\lambda_{\max}(A), -\lambda_{\min}(A)\},$$

где $\lambda_{\min}(A)$ и $\lambda_{\max}(A)$ — минимальное и максимальное собственное значения матрицы A. (См. также упражнение 3.10.)

Как показывает упражнение 3.3, каждая из норм l^2, l^1 и l^∞ в пространстве $\mathbb{R}^n$ может быть ограничена снизу и сверху любой другой с точностью до постоянного множителя; упражнение 3.11 показывает, что аналогичная связь существует также между операторной и фробениусовой нормой в пространстве матриц $\mathbb{R}^{m\times n}$. Оказывается, это не случайно, и подобное утверждение справедливо не только для рассмотренных выше норм, но и вообще для любых двух норм в *конечномерном* пространстве.

Утверждение 3.1 (Эквивалентность норм в конечномерном пространстве)**.** *Пусть V — конечномерное вещественное векторное пространство, и пусть $\|\cdot\|_{(1)}$ и $\|\cdot\|_{(2)}$ — нормы в пространстве V. Тогда найдутся $c_1, c_2 > 0$, такие, что*

$$c_1\|x\|_{(2)} \leq \|x\|_{(1)} \leq c_2\|x\|_{(2)}.$$

Доказательство. Нам особо не понадобится это утверждение, поэтому примем его без доказательства. Ограничимся лишь упоминанием, что доказательство опирается на (a) теорему Вейерштрасса о достижении непрерывной функции, заданной на компакте, своих точных нижней и верхней грани; (b) непрерывность нормы; (c) компактность единичной сферы в конечномерном пространстве (в силу ограниченности и замкнутости). □

Замечание 3.2. Как показывают неравенства 3.3 и 3.1, константы c_1 и c_2, вообще говоря, зависят от размерности пространства V. Неформально говоря, именно по этой причине утверждение 3.1 перестает быть верным в случае бесконечномерного пространства V.

При взаимодействии матрицы $A \in \mathbb{R}^{m\times n}$ и вектора $x \in \mathbb{R}^n$ получается новый вектор $Ax \in \mathbb{R}^m$. Если $A = 0$ или $x = 0$, то $Ax = 0$. Но что если матрица A не в точности равна нулю, но при этом близка к нулю (т. е. $\|A\|$ близка к нулю) можно ли утверждать, что норма $\|Ax\|$ также будет близка к нулю? Аналогично, если норма $\|x\|$ близка к нулю, можно ли утверждать, что норма $\|Ax\|$ также будет близка к нулю? Формализуя это рассуждение, хотелось бы иметь неравенство $\|Ax\| \le \|A\|\|x\|$. Оказывается, что подобное неравенство выполнено не для любых норм, что мотивирует следующее определение:

Определение 3.2 (Согласованность норм)**.** Пусть $\|\cdot\|_{(m\times n)}$, $\|\cdot\|_{(m)}$, $\|\cdot\|_{(n)}$ — нормы в пространствах $\mathbb{R}^{m\times n}$, $\mathbb{R}^m$ и $\mathbb{R}^n$ соответственно. Говорят, что матричная норма $\|\cdot\|_{(m\times n)}$ *согласована с векторными нормами* $\|\cdot\|_{(m)}$ и $\|\cdot\|_{(n)}$, если для всех $x \in \mathbb{R}^n$ выполнено

$$\|Ax\|_{(m)} \le \|A\|_{(m\times n)}\|x\|_{(n)}.$$

Упражнение 3.13. Покажите, что операторная и фробениусова нормы в $\mathbb{R}^{n\times n}$ согласованы с векторной евклидовой нормой в $\mathbb{R}^n$. (*Подсказка*: сперва покажите это для операторной нормы, а затем воспользуйтесь (3.1).) Согласованы ли эти нормы, например, с l^1 нормой в $\mathbb{R}^n$?

Аналогичным образом, для двух матриц A, B справедливость неравенства $\|AB\| \le \|A\|\|B\|$ зависит от используемой нормы:

Упражнение 3.14. (a) Покажите, что операторная и фробениусова матричные нормы *субмультипликативны*, т. е. $\|AB\| \le \|A\|\|B\|$ для всех $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times k}$. (*Подсказка*: для фробениусовой нормы примените неравенство Коши–Буняковского.)

(b) Покажите, что $\|A\|_{\max} := \max_{1\le i\le m; 1\le j\le n} |a_{ij}|$ задает норму в пространстве $\mathbb{R}^{m\times n}$, но эта норма не является субмультипликативной.