Учебник по Эконометрике

Лекция 2: Линейная Алгебра и Матрицы

Джон Стачурски

Лекции: Акшай Шенкер

Перевел: Алексей Кедо

27 сентября 2020 г.

Матрицы

Обычная $N \times K$ **матрица**:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix}$$

Символ a_{nk} означает элемент, стоящий в n-ой строке k-ого столбца

$N \times K$ матрица также называется

- **вектором строки**, если $N = 1$
- **вектором столбца**, если $K = 1$

Если $N = K$, то матрицу $\mathbf{A}$ называют **квадратной**

Если $\mathbf{A}$ квадратная и $a_{nk} = a_{kn}$ для любых k и n, то $\mathbf{A}$ называют **симметричной**

Часто элементы матрицы $\mathbf{A}$ представляют собой коэффициенты в системе линейных уравнений

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1K}x_K = b_1 \\ \vdots \\ a_{N1}x_1 + a_{N2}x_2 + \cdots + a_{NK}x_K = b_N \end{gathered}$$

Для матрицы $\mathbf{A}$ применяются обозначения:

- $\mathrm{row}_n\,\mathbf{A}$ означает n-ую строку $\mathbf{A}$
- $\mathrm{col}_k\,\mathbf{A}$ означает k-ый столбец $\mathbf{A}$

Символы $\mathbf{0}$ и $\mathbf{1}$ представляют собой матрицы, все элементы которых равны нулю и единице соответственно

Для квадратной матрицы $\mathbf{A}$, элементы a_{nn} называют **главной диагональю**:

$$\begin{pmatrix} \color{red}{a_{11}} & a_{12} & \cdots & a_{1N} \\ a_{21} & \color{red}{a_{22}} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & \color{red}{a_{NN}} \end{pmatrix}$$

Единичная матрица:

$$\mathbf{I} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Заметим, что $\mathrm{col}_n \, \mathbf{I} = \mathbf{e}_n$ — n-ый канонический базис $\mathbb{R}^N$

Алгебраические операции над матрицами

Операции над матрицами:

- Умножение на скаляр
- Сложение
- Умножение матриц

Умножение на скаляр выполняется поэлементно, как с векторами:

$$\gamma \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix} := \begin{pmatrix} \gamma a_{11} & \gamma a_{12} & \cdots & \gamma a_{1K} \\ \gamma a_{21} & \gamma a_{22} & \cdots & \gamma a_{2K} \\ \vdots & \vdots & & \vdots \\ \gamma a_{N1} & \gamma a_{N2} & \cdots & \gamma a_{NK} \end{pmatrix}$$

Сложение тоже выполняется поэлементно:

$$\begin{pmatrix} a_{11} & \cdots & a_{1K} \\ a_{21} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1K} \\ b_{21} & \cdots & b_{2K} \\ \vdots & \vdots & \vdots \\ b_{N1} & \cdots & b_{NK} \end{pmatrix}$$

$$:= \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1K}+b_{1K} \\ a_{21}+b_{21} & \cdots & a_{2K}+b_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1}+b_{N1} & \cdots & a_{NK}+b_{NK} \end{pmatrix}$$

Заметим, что матрицы должны быть одинакового размера

Умножение матриц:

Произведение $\mathbf{AB}$: i, j-ый элемент — скалярное умножение i-ой строки $\mathbf{A}$ и j-ого столбца $\mathbf{B}$

$$c_{ij} = \langle \mathrm{row}_i\,\mathbf{A}, \mathrm{col}_j\,\mathbf{B} \rangle = \sum_{k=1}^{K} a_{ik} b_{kj}$$

Для $i = j = 1$:

$$\begin{pmatrix} a_{11} & \cdots & a_{1K} \\ a_{21} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1J} \\ b_{21} & \cdots & b_{2J} \\ \vdots & \vdots & \vdots \\ b_{K1} & \cdots & b_{KJ} \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1J} \\ c_{21} & \cdots & c_{2J} \\ \vdots & \vdots & \vdots \\ c_{N1} & \cdots & c_{NJ} \end{pmatrix}$$

Здесь

$$c_{11} = \langle \mathrm{row}_1(\mathbf{A}), \mathrm{col}_1(\mathbf{B}) \rangle = \sum_{k=1}^{K} a_{1k} b_{k1}$$

Предположим, что размер $\mathbf{A}$ — $N \times K$, $\mathbf{B}$ — $J \times M$

- $\mathbf{AB}$ определена, только если $K = J$
- Размер $\mathbf{AB}$ — $N \times M$

Запомните правило:

Произведение $N \times K$ и $K \times M$ равно $N \times M$

Умножение не коммутативно: $\mathbf{AB} \neq \mathbf{BA}$

Заметим, что произведение $\mathbf{BA}$ определено, только если $N = M$ соблюдается

Факт. (??) Для согласованных матриц $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ и скаляра α, верно

1. $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$,
2. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$,
3. $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$,
4. $\mathbf{A}\alpha\mathbf{B} = \alpha\mathbf{AB}$, и
5. $\mathbf{AI} = \mathbf{A}$ и $\mathbf{IA} = \mathbf{A}$, где $\mathbf{I}$ — единичная матрица.

Здесь “согласованость” значит, что операция определена при заданных размерностях матриц

k-**ая мощность** квадратной матрицы $\mathbf{A}$ определяется как

$$\mathbf{A}^k := \underbrace{\mathbf{A} \cdots \mathbf{A}}_{k \text{ terms}}$$

Если матрица $\mathbf{B}$ такая, что $\mathbf{B}^2 = \mathbf{A}$, то $\mathbf{B}$ называется **квадратным корнем** матрицы $\mathbf{A}$ и записывается как $\sqrt{\mathbf{A}}$

Возьмем матрицу $\mathbf{A}$ размерности $N \times K$ и $K \times 1$ вектор столбца $\mathbf{x}$, перемножение $\mathbf{Ax}$:

$$\mathbf{Ax} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix}$$

$$= x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{N1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{N2} \end{pmatrix} + \cdots + x_K \begin{pmatrix} a_{1K} \\ a_{2K} \\ \vdots \\ a_{NK} \end{pmatrix}$$

$$= \sum_{k=1}^{K} x_k \operatorname{col}_k \mathbf{A}$$

Матрицы как отображения

Можно размышлять о матрице $\mathbf{A}$ размерна $N \times K$ как об отображении из $\mathbb{R}^K$ в $\mathbb{R}^N$:

$$\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$$

Такое отображение линейно

Как насчет примеров линейных функций, не использующих матрицы?

...на самом деле, таких не существует!

Множеству линейных функций из $\mathbb{R}^K$ в $\mathbb{R}^N$ инъективно соответствует множество матриц $N \times K$:

Теорема. (??) Пусть T является функцией из $\mathbb{R}^K$ в $\mathbb{R}^N$. Следующее эквивалентно:

1. T линейна.
2. Существует матрица $\mathbf{A}$ размера $N \times K$, такая что $T\mathbf{x} = \mathbf{A}\mathbf{x}$ для всех $\mathbf{x} \in \mathbb{R}^K$.

Доказательство. (1. $\implies$ 2.)

Пусть $T\colon \mathbb{R}^K \to \mathbb{R}^N$ линейна

Мы собираемся построить матрицу $\mathbf{A}$ размерности $N \times K$, такую что

$$T\mathbf{x} = \mathbf{A}\mathbf{x}, \qquad \forall\, \mathbf{x} \in \mathbb{R}^K$$

Как обычно, пусть $\mathbf{e}_k$ — k-ый канонический базисный вектор в $\mathbb{R}^K$

Определим матрицу $\mathbf{A}$ как $\operatorname{col}_k(\mathbf{A}) = T\mathbf{e}_k$. Возьмем любой $\mathbf{x} = (x_1, \dots, x_K) \in \mathbb{R}^K$. По линейности, получается

$$T\mathbf{x} = T\left[\sum_{k=1}^{K} x_k \mathbf{e}_k\right] = \sum_{k=1}^{K} x_k T\mathbf{e}_k = \sum_{k=1}^{K} x_k \operatorname{col}_k(\mathbf{A}) = \mathbf{A}\mathbf{x}$$

Доказательство. (2 $\implies$ 1) Возьмем матрицу $\mathbf{A}$ размера $N \times K$ и пусть T определяется

$$T \colon \mathbb{R}^K \to \mathbb{R}^N, \qquad T\mathbf{x} = \mathbf{A}\mathbf{x}$$

Возьмем любые $\mathbf{x}$, $\mathbf{y}$ в $\mathbb{R}^K$, и любые скаляры α и β

Правила матричной арифметики говорят нам, что

$$T(\alpha\mathbf{x} + \beta\mathbf{y}) := \mathbf{A}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{A}\mathbf{x} + \beta\mathbf{A}\mathbf{y} =: \alpha T\mathbf{x} + \beta T\mathbf{y}$$

Рассмотрим возможность решения системы линейных уравнений, например, $\mathbf{Ax} = \mathbf{b}$

Существование: можем ли мы найти $\mathbf{x}$, удовлетворяющий уравнению при любых заданных $\mathbf{b}$?

- является ли линейное отображение $T\mathbf{x} = \mathbf{Ax}$ сюръекцией?
- эквивалентно, равен ли $\operatorname{rng} T$ всему $\mathbb{R}^N$?

Пространство столбцов

Диапазон T — все вектора вида $T\mathbf{x} = \mathbf{A}\mathbf{x}$, где $\mathbf{x}$ варьируется в $\mathbb{R}^K$

Для $\mathbf{x} \in \mathbb{R}^K$ получается

$$\mathbf{A}\mathbf{x} = \sum_{k=1}^{K} x_k \operatorname{col}_k \mathbf{A}$$

Таким образом, $\operatorname{rng} T$ равен **пространству столбцов** $\mathbf{A}$ — линейная оболочка столбцов $\mathbf{A}$

$$\operatorname{colspace} \mathbf{A} := \operatorname{span}\{\operatorname{col}_1 \mathbf{A}, \ldots, \operatorname{col}_K \mathbf{A}\}$$

В итоге,

$$\operatorname{colspace} \mathbf{A} = \operatorname{rng} T = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^K\}$$

Ранг

Эквивалентные вопросы

- Насколько велик диапазон линейного отображения $T\mathbf{x} = \mathbf{A}\mathbf{x}$?
- Насколько велико пространство столбцов $\mathbf{A}$?

Очевидной мерой размера линейного подпространства является его размерность

Размерность colspace $\mathbf{A}$ известна как **ранг** $\mathbf{A}$

$$\operatorname{rank} \mathbf{A} := \dim \operatorname{colspace} \mathbf{A}$$

Так как colspace $\mathbf{A}$ — линейная оболочка K векторов, получается

$$\operatorname{rank} \mathbf{A} = \dim \operatorname{colspace} \mathbf{A} \leq K$$

$\mathbf{A}$ имеет **полный ранг системы столбцов**, если

$$\operatorname{rank}\mathbf{A} = \text{ количество столбцов } \mathbf{A}$$

Факт. (??) Для любой матрицы $\mathbf{A}$, следующие утверждения эквивалентны:

1. $\mathbf{A}$ является полным рангом системы столбцов
2. Столбцы $\mathbf{A}$ линейно независимы
3. Если $\mathbf{Ax} = \mathbf{0}$, то $\mathbf{x} = \mathbf{0}$

Квадратные матрицы и обратимость

Рассмотрим случай с матрицей $\mathbf{A}$ размера $N \times N$

Мы ищем условия, при которых для каждого $\mathbf{b} \in \mathbb{R}^N$, существует только один $\mathbf{x} \in \mathbb{R}^N$, такой что $\mathbf{A}\mathbf{x} = \mathbf{b}$

Пусть T является линейным отображением $T\mathbf{x} = \mathbf{A}\mathbf{x}$

- Когда каждая точка $\mathbf{b} \in \mathbb{R}^N$ имеет только один прообраз в T?
- Эквивалентно, когда T — биекция?

Напомним, линейные биекции называются несингулярными функциями.

Факт. (??) Для матрицы $\mathbf{A}$ размера $N \times N$ следующее эквивалентно:

1. Столбцы $\mathbf{A}$ линейно независимы.
2. Столбцы $\mathbf{A}$ формируют базис $\mathbb{R}^N$.
3. $\operatorname{rank} \mathbf{A} = N$.
4. $\operatorname{colspace} \mathbf{A} = \mathbb{R}^N$.
5. $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{y} \implies \mathbf{x} = \mathbf{y}$.
6. $\mathbf{A}\mathbf{x} = \mathbf{0} \implies \mathbf{x} = \mathbf{0}$.
7. Для каждого $\mathbf{b} \in \mathbb{R}^N$, уравнение $\mathbf{A}\mathbf{x} = \mathbf{b}$ имеется решение.
8. Для каждого $\mathbf{b} \in \mathbb{R}^N$, уравнение $\mathbf{A}\mathbf{x} = \mathbf{b}$ имеет единственное решение.

Если любое из эквивалентных условий факта **??** верное, мы назовем **несингулярной** не только отображение T, но и основную матрицу $\mathbf{A}$

Если хоть одно — и, следовательно, все — из этих условий не выполняются, то $\mathbf{A}$ называется **сингулярной**

Любая биекция имеет обратную функцию (смотрите §?? в ET)

Любое несингулярное отображение T имеет несингулярное обратное T^{-1} (факт **??** на странице **??**)

- если T создано матрицей $\mathbf{A}$, обратное T^{-1} также связано с матрицей, называемой обратной $\mathbf{A}$

Теорема. (??) Для несингулярной $\mathbf{A}$ следующие утверждения верны:

1. Существует квадратная матрица $\mathbf{B}$, такая что $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, где $\mathbf{I}$ — единичная матрица. Матрица $\mathbf{B}$ называется **обратной** $\mathbf{A}$, и записывается как $\mathbf{A}^{-1}$.
2. Для каждого $\mathbf{b} \in \mathbb{R}^N$, единственное решение задачи $\mathbf{Ax} = \mathbf{b}$ записывается как

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Пример. Расстмотрим линейный спрос для N товаров

$$q_n = \sum_{k=1}^{N} a_{nk} p_k + b_n, \quad n = 1, \dots, N$$

где q_n и p_n — количество и цена n-ого товара

Мы хотим вычислить обратную функцию спроса, которая показывает зависимость цены от количества

Записшем систему в матричной форме: $\mathbf{q} = \mathbf{A}\mathbf{p} + \mathbf{b}$

Если столбцы $\mathbf{A}$ линейно независимы, то система обратима — единственное решение существует для любых фиксированных $\mathbf{q}$ и $\mathbf{b}$:

$$\mathbf{p} = \mathbf{A}^{-1}(\mathbf{q} - \mathbf{b})$$

Для матриц $\mathbf{A}$ и $\mathbf{B}$ размера $N \times N$, если

- $\mathbf{B}$ является **левой обратной**, в том смысле, что $\mathbf{BA} = \mathbf{I}$
- или $\mathbf{B}$ является **правой обратной**, в том смысле, что $\mathbf{AB} = \mathbf{I}$

Тогда $\mathbf{A}$ обратима и $\mathbf{B}$ является обратной матрицей $\mathbf{A}$

Факт. (??) Пусть $\mathbf{A}$ и $\mathbf{B}$ — квадратные матрицы размера $N \times N$. Если $\mathbf{B}$ является левой или правой обратной матрицы $\mathbf{A}$, то $\mathbf{A}$ несингулярна и $\mathbf{B}$ — ее обратная матрица.

Факт. (??) Если $\mathbf{A}$ и $\mathbf{B}$ несингулярны и $\alpha \neq 0$, то

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$,
2. $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$, и
3. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Определитель

Определитель — уникальное число для любой квадратной матрицы $\mathbf{A}$

Пусть $S(N)$ является множеством всех биекций из $\{1, \ldots, N\}$ в саму себя

Для $\pi \in S(N)$ определим **функцию знака** π

$$\operatorname{sgn}(\pi) := \prod_{m<n} \frac{\pi(m) - \pi(n)}{m - n}$$

Определитель матрицы $\mathbf{A}$ размера $N \times N$ — это

$$\det \mathbf{A} := \sum_{\pi \in S(N)} \operatorname{sgn}(\pi) \prod_{n=1}^{N} a_{\pi(n)n}$$

Факт. (??) Если $\mathbf{I}$ — единичная матрица размера $N \times N$, $\mathbf{A}$ и $\mathbf{B}$ являются марицами размера $N \times N$ и $\alpha \in \mathbb{R}$, то

1. $\det \mathbf{I} = 1$,
2. $\mathbf{A}$ несингулярна тогда и только тогда, когда $\det \mathbf{A} \neq 0$,
3. $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$,
4. $\det(\alpha \mathbf{A}) = \alpha^N \det(\mathbf{A})$, и
5. $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$.

В случае матрицы размера 2×2 можно показать, что

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

Доказательство. (Факт ??)

Возьмем кважратную матрицу $\mathbf{A}$ и предположим, что правая обратная матрица $\mathbf{B}$ существует:

$$\mathbf{AB} = \mathbf{I}$$

Тогда обе матрицы $\mathbf{A}$ и $\mathbf{B}$ несингулярны, так как по правилам факта ??:

$$\det(\mathbf{A})\det(\mathbf{B}) = 1$$

Оба определителя $\det \mathbf{A}$ и $\det \mathbf{B}$ ненулевые и, следовательно, обе матрицы несингулярные. Далее, $\mathbf{AB} = \mathbf{I}$, значит

$$\mathbf{BAB} = \mathbf{B}$$

Умножим на $\mathbf{B}^{-1}$, получается $\mathbf{BA} = \mathbf{I}$

Диагональные и треугольные матрицы

Квадратная матрица назвается **нижней треугольной**, если каждый элемент строго выше главной диагонали равен нулю

Пример.

$$\mathbf{L} := \begin{pmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 1 \end{pmatrix}$$

Квадратная матрица назвается **верхней треугольной**, если каждый элемент строго ниже главной диагонали равен нулю

Пример.

$$\mathbf{U} := \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 1 \end{pmatrix}$$

Квадратная матрица назвается **треугольной**, если она верхняя или нижняя треугольная

Факт. (??) Если $\mathbf{A} = (a_{mn})$ треугольная, то $\det \mathbf{A} = \prod_{n=1}^{N} a_{nn}$.

Связанные линейные уравнения просто решить

Пример.
$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

становится

$$\begin{gathered} x_1 = b_1 \\ 2x_1 + 5x_2 = b_2 \\ 3x_1 + 6x_2 + x_3 = b_3 \end{gathered}$$

Верхнее уравнение включает только x_1, так что его можно решить напрямую

Подставьте это значение во второе уравнение и решите для x_2 и т.д.

Рассмотрим квадратную матрицу $\mathbf{A}$ размера $N \times N$

N элементов a_{nn} называются **главной диагональю**

$$\begin{pmatrix} \color{red}{a_{11}} & a_{12} & \cdots & a_{1N} \\ a_{21} & \color{red}{a_{22}} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & \color{red}{a_{NN}} \end{pmatrix}$$

Квадратная матрица $\mathbf{D}$ называется **диагональной**, если все значения вне главной диагонали равны нулю

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_N \end{pmatrix}$$

Часто записывается как

$$\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_N)$$

Диагональные системы очень легко решить

Пример.

$$\begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

эквивалентно

$$\begin{aligned} d_1 x_1 &= b_1 \\ d_2 x_2 &= b_2 \\ d_3 x_3 &= b_3 \end{aligned}$$

Факт. (??) Если $\mathbf{C} = \mathrm{diag}(c_1, \ldots, c_N)$ и $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_N)$, то

1. $\mathbf{C} + \mathbf{D} = \mathrm{diag}(c_1 + d_1, \ldots, c_N + d_N)$
2. $\mathbf{CD} = \mathrm{diag}(c_1 d_1, \ldots, c_N d_N)$
3. $\mathbf{D}^k = \mathrm{diag}(d_1^k, \ldots, d_N^k)$ для любых $k \in \mathbb{N}$
4. $d_n \geq 0$ для всех $n \implies \mathbf{D}^{1/2}$ существует и равняется
$$\mathrm{diag}(\sqrt{d_1}, \ldots, \sqrt{d_N})$$
5. $d_n \neq 0$ для всех $n \implies \mathbf{D}$ несингулярна и
$$\mathbf{D}^{-1} = \mathrm{diag}(d_1^{-1}, \ldots, d_N^{-1})$$

Доказательства: Проверьте 1 и 2 напрямую. Части 3-5 следуют из 1 и 2.

След, транспонирование, симметрия

След матрицы определяется как

$$\operatorname{trace}\begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix} = \sum_{n=1}^{N} a_{nn}$$

Факт. (??) Если $\mathbf{A}$ и $\mathbf{B}$ являются квадратными матрицами и $\alpha, \beta \in \mathbb{R}$, то

$$\operatorname{trace}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha \operatorname{trace}(\mathbf{A}) + \beta \operatorname{trace}(\mathbf{B})$$

Если $\mathbf{A}$ размера $N \times M$ и $\mathbf{B}$ размера $M \times N$, то $\operatorname{trace}(\mathbf{AB}) = \operatorname{trace}(\mathbf{BA})$

Транспонирование превращает матрицу $\mathbf{A}$ размера $N \times K$ в матрицу $\mathbf{A}^\mathsf{T}$ размера $K \times N$, определяемую как

$$\mathrm{col}_n(\mathbf{A}') = \mathrm{row}_n(\mathbf{A})$$

Пример. Если

$$\mathbf{A} := \begin{pmatrix} 10 & 40 \\ 20 & 50 \\ 30 & 60 \end{pmatrix} \quad \text{, то} \quad \mathbf{A}' = \begin{pmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \end{pmatrix}$$

Если

$$\mathbf{B} := \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \quad \text{, то} \quad \mathbf{B}' := \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

Квадратная матрица $\mathbf{A}$ называется **симметричной**, если $\mathbf{A}^\intercal = \mathbf{A}$

- эквивалентно, $a_{nk} = a_{kn}$ для всех k и n

Матрицы $\mathbf{A}^\intercal\mathbf{A}$ и $\mathbf{A}\mathbf{A}^\intercal$ всегда корректно определены и симметричны

Факт. (??) Для согласующихся матриц $\mathbf{A}$ и $\mathbf{B}$ транспонирование удовлетворяет

1. $(\mathbf{A}^\intercal)^\intercal = \mathbf{A}$,
2. $(\mathbf{AB})^\intercal = \mathbf{B}^\intercal\mathbf{A}^\intercal$,
3. $(\mathbf{A}+\mathbf{B})^\intercal = \mathbf{A}^\intercal + \mathbf{B}^\intercal$, и
4. $(c\mathbf{A})^\intercal = c\mathbf{A}^\intercal$ для любых постоянных c.

Факт. (??) Для любой квадратной матрицы $\mathbf{A}$, верно

1. $\operatorname{trace}(\mathbf{A}) = \operatorname{trace}(\mathbf{A}^\intercal)$ и
2. $\det(\mathbf{A}^\intercal) = \det(\mathbf{A})$.
3. Если $\mathbf{A}$ несингулярна, то $\mathbf{A}^\intercal$ тоже несингулярна, и $(\mathbf{A}^\intercal)^{-1} = (\mathbf{A}^{-1})^\intercal$.

Если $\mathbf{a}$ и $\mathbf{b}$ — векторы размера $N \times 1$, умножение матриц $\mathbf{a}^\intercal \mathbf{b} = \mathbf{b}^\intercal \mathbf{a}$ равняется $\sum_{n=1}^{N} a_n b_n$,

- так же, как скалярное произведение $\langle \mathbf{a}, \mathbf{b} \rangle$

Собственные значения и собственные векторы

Возьмем матрицу $\mathbf{A}$ размера $N \times N$

Обычно $\mathbf{A}$ отображает $\mathbf{x}$ в какое-то произвольное новое место $\mathbf{Ax}$

Но иногда $\mathbf{x}$ будет только масштабироваться:

$$\mathbf{Ax} = \lambda\mathbf{x} \quad \text{для некоторого скаляра } \lambda \tag{1}$$

Если (1) выполняется и $\mathbf{x}$ is ненулевой, то

1. $\mathbf{x}$ называется **собственным вектором** $\mathbf{A}$ и λ называется **собственным значением**
2. $(\mathbf{x}, \lambda)$ называется **собственной парой**

Ясно, что $(\mathbf{x}, \lambda)$ — собственная пара $\mathbf{A} \implies (\alpha\mathbf{x}, \lambda)$ — собственная пара $\mathbf{A}$ для любых ненулевых α

Пример. Пусть

$$\mathbf{A} := \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix}$$

Тогда

$$\lambda = 2 \quad \text{и} \quad \mathbf{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

формируют собственную пару, потому что $\mathbf{x} \neq \mathbf{0}$ и

$$\mathbf{Ax} = \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \lambda \mathbf{x}$$

Пример. Рассмотрим матрицу

$$\mathbf{R} := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Вызывает вращение против часовой стрелки в любой точке на 90°

Значит ни одна точка $\mathbf{x}$ не масштабируется

Значит не существует пары $\lambda \in \mathbb{R}$ и $\mathbf{x} \neq \mathbf{0}$, такой что

$$\mathbf{R}\mathbf{x} = \lambda \mathbf{x}$$

- Другими словами, не существует вещественных собственных пар

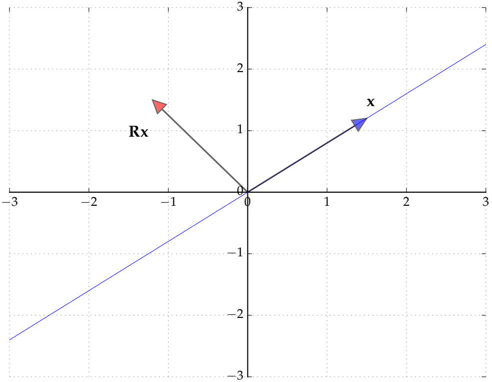

Рис.: Матрица $\mathbf{R}$ поворачивает точки на 90°

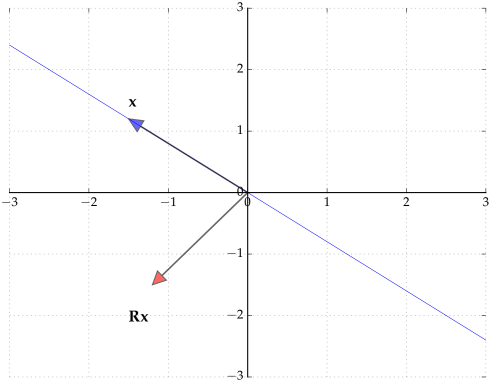

Рис.: Матрица $\mathbf{R}$ поворачивает точки на 90°

Но $\mathbf{R}\mathbf{x} = \lambda\mathbf{x}$ может выполняться, если мы допускаем комплексные числа

Пример.

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

То есть

$$\mathbf{R}\mathbf{x} = \lambda\mathbf{x} \quad \text{для} \quad \lambda := i \quad \text{и} \quad \mathbf{x} := \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

Тогда $(\mathbf{x}, \lambda)$ является собственной парой при условии, что мы допускаем комплексные числа

Факт. (??) для любой квадратной матрицы $\mathbf{A}$

$$\lambda \text{ является собственным значением } \mathbf{A} \iff \det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

Доказательство. Пусть $\mathbf{A}$ — матрица размера $N \times N$ и $\mathbf{I}$ — единичная матрица размера $N \times N$

Получается

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{I}) = 0 &\iff \mathbf{A} - \lambda\mathbf{I} \text{ сингулярно} \\ &\iff \exists \mathbf{x} \neq \mathbf{0} \text{ s.t. } (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \\ &\iff \exists \mathbf{x} \neq \mathbf{0} \text{ s.t. } \mathbf{A}\mathbf{x} = \lambda\mathbf{x} \\ &\iff \lambda \text{ является собственным значением } \mathbf{A} \end{aligned}$$

Пример. В случае матрицы 2×2,

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Longrightarrow \quad \mathbf{A} - \lambda \mathbf{I} = \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix}$$

$$\begin{aligned} \therefore \quad \det(\mathbf{A} - \lambda \mathbf{I}) &= (a-\lambda)(d-\lambda) - bc \\ &= \lambda^2 - (a+d)\lambda + (ad - bc) \end{aligned}$$

Значит собственные значения $\mathbf{A}$ являются двумя корнями уравнения

$$\lambda^2 - (a+d)\lambda + (ad - bc) = 0$$

Эквивалентно,

$$\lambda^2 - \mathrm{trace}(\mathbf{A})\lambda + \det(\mathbf{A}) = 0$$

Existence of Eigenvalues

Fix $N \times N$ matrix $\mathbf{A}$

Факт. There exist complex numbers $\lambda_1, \ldots, \lambda_N$ such that

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \prod_{n=1}^{N} (\lambda_n - \lambda)$$

Each such λ_i is an eigenvalue of $\mathbf{A}$ because

$$\det(\mathbf{A} - \lambda_i \mathbf{I}) = \prod_{n=1}^{N} (\lambda_n - \lambda_i) = 0$$

Important: not all eigenvalues are necessarily distinct — there can be repeats

Факт. (??) Given $N \times N$ matrix $\mathbf{A}$ with eigenvalues $\lambda_1, \ldots, \lambda_N$ we have

1. $\det(\mathbf{A}) = \prod_{n=1}^{N} \lambda_n$
2. $\text{trace}(\mathbf{A}) = \sum_{n=1}^{N} \lambda_n$
3. если $\mathbf{A}$ is symmetric, then $\lambda_n \in \mathbb{R}$ для всех n
4. если $\mathbf{A}$ is nonsingular, then eigenvalues of $\mathbf{A}^{-1} = 1/\lambda_1, \ldots, 1/\lambda_N$
5. если $\mathbf{A} = \text{diag}(d_1, \ldots, d_N)$, then $\lambda_n = d_n$ для всех n

Hence $\mathbf{A}$ is nonsingular $\iff$ all eigenvalues are nonzero

Quadratic Forms

Fix $N \times N$ matrix $\mathbf{A}$

The **quadratic function** or **quadratic form** on $\mathbb{R}^N$ associated with $\mathbf{A}$ is the map Q defined by

$$Q(\mathbf{x}) := \mathbf{x}^\top \mathbf{A} \mathbf{x} = \sum_{j=1}^{N} \sum_{i=1}^{N} a_{ij} x_i x_j$$

Пример. Let $N = 2$ и let $\mathbf{A}$ be the identity matrix $\mathbf{I}$. In this case,

$$Q(\mathbf{x}) = \|\mathbf{x}\|^2 = x_1^2 + x_2^2$$

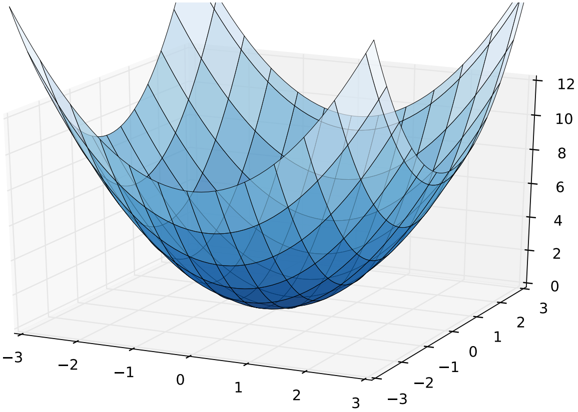

Рис.: Quadratic function $Q(\mathbf{x}) = x_1^2 + x_2^2$

Notice:

- The graph for $Q(\mathbf{x}) = \|\mathbf{x}\|^2 = x_1^2 + x_2^2$ lies everywhere above zero

Matrix $\mathbf{A}$ with Quadratic form with the above property $Q(\mathbf{x}) \geq 0$ называется *positive definite*

More generally, an $N \times N$ symmetric matrix $\mathbf{A}$ называется

- **nonnegative definite** если $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \geq 0$ для всех $\mathbf{x} \in \mathbb{R}^N$,
- **positive definite** если $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} > 0$ для всех $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$,
- **nonpositive definite** если $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \leq 0$ для всех $\mathbf{x} \in \mathbb{R}^N$, и
- **negative definite** если $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} < 0$ для всех $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$.

если $\mathbf{A}$ fits none of these categories, then $\mathbf{A}$ называется **indefinite**

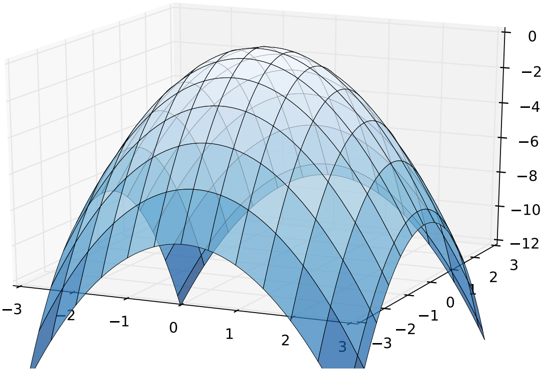

Рис.: Quadratic function $Q(\mathbf{x}) = -x_1^2 - x_2^2$

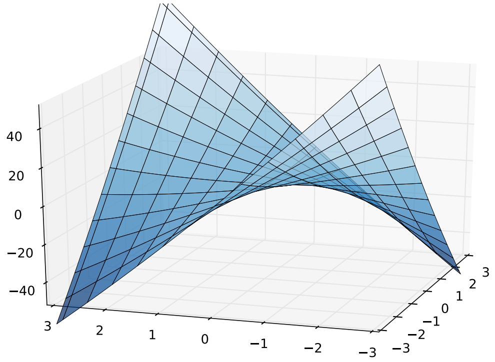

Рис.: Quadratic function $Q(\mathbf{x}) = x_1^2/2 + 8x_1x_2 + x_2^2/2$

When the matrix $\mathbf{A}$ is diagonal:

$$\mathbf{A} = \operatorname{diag}(d_1, \ldots, d_N) \quad \text{implies} \quad Q(\mathbf{x}) = d_1 x_1^2 + \cdots + d_N x_N^2$$

A diagonal matrix is positive definite если и only если all diagonal elements are positive

Факт. (??) Let $\mathbf{A}$ be any symmetric matrix. $\mathbf{A}$ is

1. positive definite если и only если its eigenvalues are all positive,
2. negative definite если и only если its eigenvalues are all negative,

...и similarly for nonpositive и nonnegative definite

Факт. (??) если $\mathbf{A}$ is positive definite, then $\mathbf{A}$ is nonsingular, with $\det \mathbf{A} > 0$

A necessary (but not sufficient) condition for each kind of definiteness:

Факт. (??) если $\mathbf{A}$ is positive definite, then each element a_{nn} on the principal diagonal is positive, и the same for nonnegative, nonpositive и negative.

Projection Matrices

Recall given any subspace of $\mathbb{R}^N$, S, the corresponding projection $\mathbf{P} = \operatorname{proj} S$ is a linear map from $\mathbb{R}^N$ to $\mathbb{R}^N$

Recall Theorem **??**: there exists an $N \times N$ matrix $\hat{\mathbf{P}}$ such that $\mathbf{Px} = \hat{\mathbf{P}}\mathbf{x}$ для всех $\mathbf{x} \in \mathbb{R}^N$

- from now on $\mathbf{P}$ will also represent the corresponding matrix

What does this matrix look like?

Теорема. (??) Let S be a subspace of $\mathbb{R}^N$. если $\mathbf{P} = \operatorname{proj} S$, then

$$\mathbf{P} = \mathbf{B}(\mathbf{B}^\intercal \mathbf{B})^{-1}\mathbf{B}^\intercal \tag{2}$$

for every matrix $\mathbf{B}$ such that the columns of $\mathbf{B}$ form a basis of S

See exercise ?? for proof

- The matrix $\mathbf{M} = \mathbf{I} - \mathbf{P}$ denotes the residual projection (see page ??)

Пример. Recall example **??** on page **??**

We found the projection of $\mathbf{y} \in \mathbb{R}^N$ onto $\text{span}\{\mathbf{1}\}$ is $\bar{y}\mathbf{1}$

Same result using Theorem (**??**):

- Since $\mathbf{1}$ is a basis for $\text{span}\{\mathbf{1}\}$:

$$\mathbf{P} = \text{proj span}\{\mathbf{1}\} \implies \mathbf{P} = \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top = \frac{1}{N}\mathbf{1}\mathbf{1}^\top$$

- Thus, $\mathbf{P}\mathbf{y} = \bar{y}\mathbf{1}$, as expected
- Corresponding residual projection is

$$\mathbf{M}_c = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$$

Факт. (??) In the setting of theorem **??**, we have

1. $\mathbf{MB} = \mathbf{0}$
2. $\mathbf{PB} = \mathbf{B}$

Proof is an exercise (ex. **??** in ET)

Easy to see $\mathbf{M}_c$ in the previous example maps $\mathbf{1}$ to $\mathbf{0}$

A square matrix $\mathbf{A}$ is **idempotent** если $\mathbf{AA} = \mathbf{A}$

Факт. (??) Both $\mathbf{P}$ и $\mathbf{M}$ are symmetric и idempotent

(Exercise: check by direct calculation)

Intuition: projecting onto a subspace twice is the same as projecting once – recall fact **??** on page **??**

Факт. (??) Let S be a linear subspace of $\mathbb{R}^N$. если $\mathbf{P} = \operatorname{proj} S$ и $\mathbf{M}$ is the residual projection, then

1. $\operatorname{rank} \mathbf{P} = \operatorname{trace} \mathbf{P} = \dim S$ и
2. $\operatorname{rank} \mathbf{M} = \operatorname{trace} \mathbf{M} = N - \dim S$

Доказательство.

- The rank of a linear map is the dimension of its range. When $\mathbf{P} = \operatorname{proj} S$, the range of $\mathbf{P}$ is exactly S
- To show that $\operatorname{trace} \mathbf{P} = \dim S$ also holds, use fact **??**– $\operatorname{trace} \mathbf{P} = \dim S$,
- It follows that $\operatorname{trace} \mathbf{M} = N - \dim S$, because

$$\operatorname{trace} \mathbf{M} = \operatorname{trace}(\mathbf{I} - \mathbf{P}) = \operatorname{trace} \mathbf{I} - \operatorname{trace} \mathbf{P} = N - \dim S$$

Overdetermined Systems of Equations

Рассмотрим systems of equations of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$ when:

- The matrix $\mathbf{A}$ is $N \times K$ и has full column rank
- The vector $\mathbf{x}$ is $K \times 1$
- The vector $\mathbf{b}$ is $N \times 1$
- $K \leq N$

Taking $\mathbf{A}$ и $\mathbf{b}$ as given, we seek $\mathbf{x} \in \mathbb{R}^K$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$

если $K = N$, then system has precisely one solution

When $N > K$, the system of equations said to be **overdetermined**:

- number of equations $>$ number of unknowns
- number of constraints $>$ degrees of freedom

May not be able find a $\mathbf{b}$ that satisfies all N equations

Recall the linear map $T\colon \mathbb{R}^K \to \mathbb{R}^N$ corresponding to $\mathbf{A}$ is $T\mathbf{x} = \mathbf{A}\mathbf{x}$

The following statements are equivalent:

1. there exists an $\mathbf{x} \in \mathbb{R}^K$ with $\mathbf{A}\mathbf{x} = \mathbf{b}$
2. the vector $\mathbf{b} \in \operatorname{colspace} \mathbf{A}$
3. the vector $\mathbf{b} \in \operatorname{rng} T$

Theorem **??** on page **??**: when $K < N$, the function T cannot be onto – possible $\mathbf{b}$ lies outside the range of T

When $K < N$, the scenario $\mathbf{b} \in \operatorname{colspace} \mathbf{A}$ is “very rare"because:

- the point $\mathbf{b}$ is an arbitrary point in $\mathbb{R}^N$
- the space $\operatorname{colspace} \mathbf{A}$ has dimension K
- K-dimensional subspaces of $\mathbb{R}^N$ have “Lebesgue measure zero” – the “chance” of $\mathbf{b}$ happening to lie in this subspace is tiny

Stиard approach: admit an exact solution may not exist

Focus on finding $\mathbf{x} \in \mathbb{R}^K$ to make $\mathbf{A}\mathbf{x}$ as close to $\mathbf{b}$ as possible

- close in terms of ordinary Euclidean norm

The minimization problem, called the **least squares problem**

$$\hat{\mathbf{x}} := \operatorname*{argmin}_{\mathbf{x} \in \mathbb{R}^K} \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \tag{3}$$

Assuming $\mathbf{A}$ is $N \times K$ with $K \leq N$ и $\mathbf{b}$ is $N \times 1$, we can use the orthogonal projection theorem to solve (3)

Теорема. (??) если $\mathbf{A}$ has full column rank, then (3) has the unique solution

$$\hat{\mathbf{x}} := (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \tag{4}$$

Доказательство.

Let:

- $\mathbf{A}$ и $\mathbf{b}$ be as in the statement of the theorem
- $\hat{\mathbf{x}}$ be as in (4) и
- $S := \operatorname{colspace} \mathbf{A}$

By full column rank assumption, the columns of $\mathbf{A}$ form a basis for S. Applying theorem **??**, orthogonal projection of $\mathbf{b}$ onto S is

$$\mathbf{P}\mathbf{b} := \mathbf{A}(\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}\mathbf{b} = \mathbf{A}\hat{\mathbf{x}} \tag{5}$$

Since the orthogonal projection theorem gives a unique minimizer in terms of the closest point in S to $\mathbf{b}$,

$$\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\| < \|\mathbf{b} - \mathbf{y}\| \quad \text{для всех } \mathbf{y} \in S,\ \mathbf{y} \neq \mathbf{A}\hat{\mathbf{x}} \tag{6}$$

Доказательство. (cont.) Pick any $\mathbf{x} \in \mathbb{R}^K$ such that $\mathbf{x} \neq \hat{\mathbf{x}}$

We have $\mathbf{A}\mathbf{x} \in S$

In addition, since $\mathbf{x} \neq \hat{\mathbf{x}}$, и since $\mathbf{A}$ has full column rank, it must be that $\mathbf{A}\mathbf{x} \neq \mathbf{A}\hat{\mathbf{x}}$ (ex. **??**)

Hence

$$\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\| < \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \quad \text{для всех } \mathbf{x} \in \mathbb{R}^K,\ \mathbf{x} \neq \hat{\mathbf{x}}$$

In other words, $\hat{\mathbf{x}}$ is the unique solution to (3)

In (4), the matrix $(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}$ называется the pseudoinverse of $\mathbf{A}$

если $K = N$, then the least squares solution $\hat{\mathbf{x}}$ in (4) reduces to:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

What happens если the columns of $\mathbf{A}$ are not linearly independent?

- the set $\operatorname{colspace} \mathbf{A}$ is still a linear subspace и the orthogonal projection theorem still gives us a closest point $\mathbf{Pb}$ to $\mathbf{b}$ in $\operatorname{colspace} \mathbf{A}$
- since $\mathbf{Pb} \in \operatorname{colspace} \mathbf{A}$, there still exists a vector $\hat{\mathbf{x}}$ such that $\mathbf{Pb} = \mathbf{A}\hat{\mathbf{x}}$
- but there exists an infinity of such vectors

See Exercise ??

QR Decomposition

The **QR decomposition** of a given matrix $\mathbf{A}$ is a product of the form $\mathbf{QR}$

- first matrix has orthonormal columns и
- the second is upper triangular

Applications include least squares problems и the computation of eigenvalues

Теорема. (**??**) если $\mathbf{A}$ is an $N \times K$ matrix with full column rank, then there exists a factorization $\mathbf{A} = \mathbf{QR}$ where

1. $\mathbf{R}$ is $K \times K$, upper triangular и nonsingular, и
2. $\mathbf{Q}$ is $N \times K$, with orthonormal columns

See page **??** in ET for a proof

Given the decomposition $\mathbf{A} = \mathbf{QR}$, the least squares solution $\hat{\mathbf{x}}$ defined in (4) can also be written as:

$$\hat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q}^{\intercal}\mathbf{b}$$

See Ex. **??**

Premultiplying by $\mathbf{R}$:

$$\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^{\intercal}\mathbf{b}$$

Diagonalisation и Spectral Theory

если $f\colon A \to A$ и $g\colon B \to B$, then g is said to be **topologically conjugate** to f whenever there exists a continuous bijection $\tau\colon B \to A$ such that

$$f = \tau \circ g \circ \tau^{-1}$$

Can be beneficial если g is somehow simpler than f

A square matrix $\mathbf{A}$ is said to be **similar** to another matrix $\mathbf{B}$ если there exists an invertible matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$

$$
\begin{tikzcd}
\bullet \arrow[r, "\mathbf{B}"] & \bullet \arrow[d, "\mathbf{P}"] \\
\mathbf{x} \arrow[u, "\mathbf{P}^{-1}"] \arrow[r, "\mathbf{A}"] & \mathbf{Ax}
\end{tikzcd}
$$

Рис.: $\mathbf{A}$ is similar to $\mathbf{B}$

если $\mathbf{A}$ is similar to a diagonal matrix, then $\mathbf{A}$ is called **diagonalizable**

We are interested in similarity to simple matrices, и diagonal matrices are the simplest kind

Факт. (??) если $\mathbf{A}$ is similar to $\mathbf{B}$, then $\mathbf{A}^t$ is similar to $\mathbf{B}^t$ для всех $t \in \mathbb{N}$

Пример. We want to calculate $\mathbf{A}^t$ for some given $t \in \mathbb{N}$

если $\mathbf{A} = \mathbf{P\Lambda P}^{-1}$ for some $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, then by fact ?? и fact ??, we have

$$\mathbf{A}^t = \mathbf{P}\,\mathrm{diag}(\lambda_1^t, \ldots, \lambda_N^t)\mathbf{P}^{-1}$$

Diagonalization и Eigenvalues

Факт. (??) если $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$ for some $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, then $(\mathrm{col}_n \mathbf{P}, \lambda_n)$ is an eigenpair of $\mathbf{A}$ for each n

Доказательство. Observe $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$ implies $\mathbf{A}\mathbf{P} = \mathbf{P}\boldsymbol{\Lambda}$

Equating the nth column on each side gives

$$\mathbf{A}\mathbf{p}_n = \lambda_n \mathbf{p}_n$$

Where $\mathbf{p}_n := \mathrm{col}_n \mathbf{P}$

Note $\mathbf{p}_n$ is not the zero vector because $\mathbf{P}$ is invertible

But when is $\mathbf{A}$ diagonalizable?

Факт. (3.3.7) An $N \times N$ matrix $\mathbf{A}$ is diagonalizable если и only если it has N linearly independent eigenvectors

In some cases, we can get an even simpler matrix decomposition если the matrix $\mathbf{P}$ has orthogonal columns

These kinds of matrices are called **orthogonal matrices**

Факт. (??) если $\mathbf{Q}$ и $\mathbf{P}$ are $N \times N$ orthogonal matrices, then

1. $\mathbf{Q}^\mathsf{T}$ is orthogonal и $\mathbf{Q}^{-1} = \mathbf{Q}^\mathsf{T}$,
2. $\mathbf{QP}$ is orthogonal, и
3. $\det \mathbf{Q} \in \{-1, 1\}$.

если $\mathbf{A} = \mathbf{Q\Lambda Q}^{-1}$ и $\mathbf{Q}$ has orthonormal columns, then

$$\mathbf{A} = \mathbf{Q\Lambda Q}^{\mathsf{T}}$$

Clearly, $\mathbf{A}$ must be symmetric. Next theorem shows symmetry of $\mathbf{A}$ is also sufficient

Теорема. (??) если $\mathbf{A}$ is symmetric, then $\mathbf{A}$ can be diagonalized as $\mathbf{A} = \mathbf{Q\Lambda Q}^{\mathsf{T}}$, where $\mathbf{Q}$ is an orthogonal matrix и $\mathbf{\Lambda}$ is the diagonal matrix formed from the eigenvalues of $\mathbf{A}$

Above theorem was a version of the spectral decomposition theorem

$\mathbf{Q\Lambda Q}^\mathsf{T}$ называется the symmetric eigenvalue decomposition of $\mathbf{A}$ – action of $\mathbf{A}$ on an $N \times 1$ vector $\mathbf{x}$:

$$\mathbf{Ax} = \sum_{n=1}^{N} \lambda_n (\mathbf{u}_n^\mathsf{T}\mathbf{x})\mathbf{u}_n$$

where λ_n is the nth eigenvalue of $\mathbf{A}$ и $\mathbf{u}_n = \mathrm{col}_n\, \mathbf{Q}$

Compare with $\mathbf{x} = \sum_{n=1}^{N} (\mathbf{u}_n^\mathsf{T}\mathbf{x})\mathbf{u}_n$

Факт. (??) если $\mathbf{A}$ is nonnegative definite, then $\sqrt{\mathbf{A}}$ exists и equals $\mathbf{Q}\sqrt{\mathbf{\Lambda}}\mathbf{Q}^{\intercal}$. The matrix $\sqrt{\mathbf{\Lambda}}$ is given by $\mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_N})$

Факт. (??) если $\mathbf{A}$ is positive definite, then there exists a nonsingular, upper triangular matrix $\mathbf{R}$ such that $\mathbf{A} = \mathbf{R}^{\intercal}\mathbf{R}$

This decomposition называется the **Cholesky decomposition**

Доказательство. (Cholesky decomposition) We can write:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\mathsf{T} = \mathbf{Q}\sqrt{\mathbf{\Lambda}}\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\mathsf{T} = (\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\mathsf{T})^\mathsf{T}\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\mathsf{T}$$

Then apply the QR decomposition to $\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\mathsf{T}$:

$$\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\mathsf{T} = \tilde{\mathbf{Q}}\mathbf{R}$$

where $\mathbf{R}$ is nonsingular и upper triangular, и $\tilde{\mathbf{Q}}$ has orthonormal columns

Because the columns of $\tilde{\mathbf{Q}}$ are orthonormal,

$$\mathbf{A} = (\tilde{\mathbf{Q}}\mathbf{R})^\mathsf{T}\tilde{\mathbf{Q}}\mathbf{R} = \mathbf{R}^\mathsf{T}\tilde{\mathbf{Q}}^\mathsf{T}\tilde{\mathbf{Q}}\mathbf{R} = \mathbf{R}^\mathsf{T}\mathbf{R}$$

Norms и Continuity

Given vector sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^K$ и any point $\mathbf{x} \in \mathbb{R}^K$, we say that $\{\mathbf{x}_n\}$ **converges** to $\mathbf{x}$ и write $\mathbf{x}_n \to \mathbf{x}$ если, для любых $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that $\|\mathbf{x}_n - \mathbf{x}\| < \epsilon$ whenever $n \geq N$

Equivalently, the real-valued sequence $z_n := \|\mathbf{x}_n - \mathbf{x}\|$ converges to zero in $\mathbb{R}$ as $n \to \infty$

Факт. (3.3.11) The following results hold:

1. если $\mathbf{x}_n \to \mathbf{x}$ и $\mathbf{y}_n \to \mathbf{y}$, then $\mathbf{x}_n + \mathbf{y}_n \to \mathbf{x} + \mathbf{y}$.
2. если $\mathbf{x}_n \to \mathbf{x}$ и $\alpha \in \mathbb{R}$, then $\alpha \mathbf{x}_n \to \alpha \mathbf{x}$.
3. $\mathbf{x}_n \to \mathbf{x}$ если и only если $\mathbf{a}^\top \mathbf{x}_n \to \mathbf{a}^\top \mathbf{x}$ для всех $\mathbf{a} \in \mathbb{R}^K$.

We want to extend notion of convergence to matrices

The **matrix norm** of $N \times K$ matrix $\mathbf{A}$:

$$\|\mathbf{A}\| := \max \left\{ \|\mathbf{A}\mathbf{x}\| \ : \ \mathbf{x} \in \mathbb{R}^K, \ \|\mathbf{x}\| = 1 \right\} \tag{7}$$

The value of the matrix norm is not easy to solve for in general

However, the matrix norm behaves like the vector norm

Факт. (3.3.12) для любых conformable matrices $\mathbf{A}$ и $\mathbf{B}$, the matrix norm satisfies

1. $\|\mathbf{A}\| \geq 0$ и $\|\mathbf{A}\| = 0$ если и only если all entries of $\mathbf{A}$ are zero,
2. $\|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\|$ для любых scalar α,
3. $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$, и
4. $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$.

Факт. (??) для любых $J \times K$ matrix $\mathbf{A}$ with elements a_{jk}, we have

$$\|\mathbf{A}\| \leq \sqrt{JK} \max_{jk} |a_{jk}|$$

если every element of $\mathbf{A}$ is close to zero then $\|\mathbf{A}\|$ is also close to zero

Neumann Series

Later on, we study dynamic systems of the form

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{b}$$

Does there exist a “stationary"vector $\mathbf{x} \in \mathbb{R}^N$, in the sense that $\mathbf{x}_t = \mathbf{x}$ implies $\mathbf{x}_{t+1} = \mathbf{x}$?

We seek an $\mathbf{x} \in \mathbb{R}^N$ that solves the system of equations

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b} \qquad (\mathbf{A} \text{ is } N \times N \text{ и } \mathbf{b} \text{ is } N \times 1) \tag{8}$$

Рассмотрим the scalar case $x = ax + b$

если $|a| < 1$, then there is a unique solution

$$\bar{x} = \frac{b}{1-a} = b \sum_{k=0}^{\infty} a^k$$

The Neumann series Lemma helps generalise to $\mathbb{R}^N$

Теорема. (??) если $\mathbf{A}$ is square и $\|\mathbf{A}^j\| < 1$ for some $j \in \mathbb{N}$, then $\mathbf{I} - \mathbf{A}$ is invertible, и

$$(\mathbf{I} - \mathbf{A})^{-1} = \sum_{i=0}^{\infty} \mathbf{A}^i$$

When the condition of the Neumann series lemma holds, (8) has the unique solution

$$\bar{\mathbf{x}} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \sum_{i=0}^{\infty} \mathbf{A}^i \mathbf{b}$$

To test the condition, we use the **spectral radius** of $\mathbf{A}$:

$$\varrho(\mathbf{A}) := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } \mathbf{A}\}$$

$|\lambda|$ is the **modulus** of the possibly complex number λ

Факт. если $\varrho(\mathbf{A}) < 1$, then $\|\mathbf{A}^j\| < 1$ for some $j \in \mathbb{N}$

Why is $\varrho(\mathbf{A}) < 1$ is sufficient?

We need $\sum_{i=0}^{t} \mathbf{A}^i(\mathbf{I} - \mathbf{A})$ to close be $\mathbf{I}$ for large t

We have:

$$\sum_{i=0}^{t} \mathbf{A}^i(\mathbf{I} - \mathbf{A}) = \sum_{i=0}^{t} \mathbf{A}^i - \sum_{i=0}^{t} \mathbf{A}^{i+1} = \mathbf{I} - \mathbf{A}^{t+1}$$

Simpleслиу to the case where $\mathbf{A}$ is diagonalizable:

$$\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues $\lambda_1, \ldots, \lambda_N$ of $\mathbf{A}$ on its principal diagonal

Now, use fact **??**,

$$\mathbf{A}^t = \mathbf{P} \begin{pmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_N^t \end{pmatrix} \mathbf{P}^{-1}$$

если $\varrho(\mathbf{A}) < 1$, then $|\lambda_n| < 1$ для всех n, и hence $\lambda_n^t \to 0$ as $t \to \infty$. It follows that $\mathbf{A}^t \to \mathbf{0}$