

Strang 教授线性代数笔记

说明：单位矩阵： I ；声明的向量都是列向量， \vec{a} 、 \mathbf{a} 都表示列向量。

课程视频：<https://www.youtube.com/playlist?list=PL6839449936471E0C>

1. 矩阵消元、矩阵的秩与转置（基础）

1.1 矩阵的初等变换

矩阵初等行变换：

- 对换两行
- 以数 $k \neq 0$ 乘某一行中所有元（数乘）
- 把某一行所有元的 k 倍加到另一行对应的元上去

把“行”换成“列”，即为矩阵的初等列变换。矩阵的初等行变换和初等列变换统称为初等变换。

初等矩阵：由单位矩阵 I 经过一次初等变换得到的矩阵成为初等矩阵。

性质：设矩阵 A 是一个 $m \times n$ 矩阵，则矩阵 A 的初等行变换可以表示为左乘多个 m 阶初等矩阵 E ，即：

$E_1 E_2 \cdots E_n A$ ；矩阵 A 的初等列变换可以表示为右乘多个 n 阶初等矩阵 E ，即：
 $A E_1 E_2 \cdots E_n$ 。

注意：这里的左右指矩阵 A 的左右。

例子：

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ ，交换单位阵的两行得到初等矩阵 $E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ，则 $EA = \begin{bmatrix} c & d \\ a & b \end{bmatrix}$ ，即交换了矩阵 A 的两行。

E 也可以视为交换单位阵的两列得到，那么 $AE = \begin{bmatrix} b & a \\ d & c \end{bmatrix}$ ，即交换 A 的两列。

行阶梯形矩阵

非零矩阵满足：

1. 非零行在零行（一行全为0）的上面
2. 非零行的第一个非零元素所在列在上一行（如果存在）的第一个非零元素所在列的右面。

例如： $B = \begin{pmatrix} 1 & 1 & -2 & 1 & 4 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

行最简形矩阵

在阶梯形矩阵中，若非零行的第一个非零元素全是**1**，且非零行的第一个元素**1**所在列的其余元素全为零，就称该矩阵为行最简形矩阵。

利用初等行变换，把一个矩阵化为行阶梯形矩阵和行最简形矩阵，是一种很重要的运算。

例如对上述行阶梯矩阵 B 进行初等行变换得到：

$$\begin{pmatrix} 1 & 0 & -1 & 0 & 4 \\ 0 & 1 & -1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

行最简形矩阵(或者行阶梯形矩阵)的非零行个数即为矩阵的秩 (**rank**)。

若通过初等行变换可以得到形如 $R = \begin{pmatrix} I_{r \times r} & F_{r \times (n-r)} \\ O & O \end{pmatrix}$ 的简化矩阵 (上述 B 就不符合以下情况)，那么就有如下结论：

很重要!!!!: $RN = 0 \Rightarrow N = \begin{pmatrix} -F \\ I \end{pmatrix}$, I 的阶数等于自由变量的个数。可以看出对系数矩阵初等变换为形如 R ，解向量就可以写成 N 的形式。

接着考虑 $(I \quad F) \begin{pmatrix} x_{pivot} \\ x_{free} \end{pmatrix} = 0 \Rightarrow x_{pivot} = -Fx_{free}$

利用初等行变换求方阵的逆矩阵 (高斯-约旦消元法) (Lecture 10的内容)

A 为可逆 n 阶方阵，将矩阵 (A, I) 经过初等行变换可以得到 (I, E) ， I 为单位矩阵，则 E 为矩阵的逆矩阵。实际上，矩阵 E 是由单位阵变换而来的，在对 (A, I) 进行行变换的过程中， E 记录了整个变换过程。

$$\begin{aligned} R(A, I) &= (I, E) \\ RA = I, RI = E &\Rightarrow R = E = A^{-1} \end{aligned}$$

R 表示初等行变换矩阵。

从初等变换角度理解矩阵乘法: $A_{m \times n} B_{n \times p} = C_{m \times p}$

从 A 的角度，相当于 A 右乘矩阵 B 中的每个列向量， C 中每列都是 A 中列向量的线性组合。

从 B 的角度，相当于 B 左乘 A 中每个行向量， C 中每行都是 B 中行向量的线性组合。

注意：矩阵的初等行变化不会改变线性方程组的解，初等列变换会改变解（交换两列相当于把两个未知数的系数交换了。方程组也就变了）。

1.2 矩阵的转置

A 为 $m \times n$ 的矩阵， A^T 称为 A 的转置矩阵，且 $(A)_{ij} = (A^T)_{ji}$ 。（行变列，列变行）。

性质：

- $(A + B)^T = A^T + B^T$
- $(\lambda A)^T = \lambda A^T$
- $(AB)^T = B^T A^T$
- 若 A 为方阵，且 $A = A^T$ ，则 A 为对称矩阵
- $A^T A$ 是一个 n 阶对称阵， $(A^T A)^T = A^T (A^T)^T = A^T A$

1.3 矩阵的逆

定义： $AB = BA = I$ ，则 A 、 B 互为逆矩阵。

可逆矩阵又称为非奇异矩阵，不可逆矩阵被称为奇异矩阵。

方阵 A 可逆的充要条件是 $|A| \neq 0$ 。

方阵 A 可逆的充要条件是 A 可以通过初等行变换得到单位阵。

- 证明： A 可逆 \Leftrightarrow 存在可逆矩阵 P ， $PA = I \Leftrightarrow A$ 行变换得到 I

逆矩阵的性质

若 n 阶矩阵 A 可逆，有

- $(A^{-1})^{-1} = A$
- 数 $\lambda \neq 0$ ， $(\lambda A)^{-1} = \frac{A}{\lambda}$
- B 与 A 同阶数，且可逆， $(AB)^{-1} = B^{-1}A^{-1}$
证明： $P(AB) = I \Rightarrow PAB B^{-1} = I B^{-1} \Rightarrow PA = B^{-1} \Rightarrow P = B^{-1}A^{-1}$

- A^T 亦可逆，且 $(A^T)^{-1} = (A^{-1})^T$
- $|A| \neq 0$
- $|A^{-1}| = \frac{1}{|A|}$

证明： A 可逆 $\Rightarrow AA^{-1} = E \Rightarrow |A| * |A^{-1}| = |E| = 1 \Rightarrow |A^{-1}| = \frac{1}{|A|}$

- A 满秩矩阵。由于 $|A| \neq 0$ ，则 A 的最高阶子式为自身且不为 0，因此 $R(A) = n$ ，即满秩。

1.4 矩阵与向量乘法

- Ax 相当于 A 的列向量的线性组合，结果是列向量 $x_1c_1 + x_2c_2 + \dots$ 对应列乘以 \vec{x} 的分量然后所有列求和
- xA 相当于 A 的行向量的线性组合，结果是行向量 $x_1r_1 + x_2r_2 + \dots$ 对应行乘以 \vec{x} 的分量然后所有行求和

1.5 矩阵的秩

2. 向量空间、列空间、零空间（进阶）

给定 $m \times n$ 的矩阵 A

2.1 向量空间的定义（只讨论实数）

定义：以向量为元素的集合 V 称为向量空间（vector space）。

向量空间必须满足：

- 对加法运算封闭，任意两个向量相加得到的向量仍在空间内
- 对数乘运算封闭，乘以任意实数仍在空间内
- 包含一个零向量（当非零向量数乘 0 得到零向量，因此必须包含）

\mathbb{R}^n 表示含有 n 个元素的向量且基向量个数为 n （维度为 n ）组成的向量空间。

子空间 (**subspace**) 满足以上三条性质却又不包含所有向量的空间。(向量空间内的向量空间)

例子: \mathbb{R}^2 的子空间: 整个二维平面; 二维平面内, 过原点的任意直线; 零向量 $\vec{0} = (0, 0)$ 。

向量空间 P , 两个 S 中子空间 S, I 。两个子空间的并并不是一个子空间, 两个子空间的交是一个子空间。

即: $S \cap I \in P, S \cup I \notin P$ 。

2.2 列空间

定义: 矩阵 A 的列向量的线性组合 (数乘和加法运算) 构成了矩阵 A 的列空间, 记为 $C(A)$ 。

同理, 定义矩阵 A 的行向量的线性组合, 构成了 A 的行空间, 记为 $R(A)$ 。

考虑 $A = \begin{pmatrix} 1 & 3 \\ 2 & 3 \\ 4 & 1 \end{pmatrix}$, $C(A)$ 构成了三维空间中的一个过原点的平面。

考虑 $A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix}$, $C(A)$ 是 \mathbb{R}^4 的一个子空间 (每一列含有 4 个元素)。 $C(A)$ 的维数

为 2, 是四维空间的一个平面 (第三列是前两列的线性组合, 因此前两列作为列空间的基)。

考虑线性方程组 $Ax = b$, 是否始终有解? 答案是当 \vec{b} 不在 $C(A)$ 内时, 方程无解。因为 Ax 相当于矩阵列向量的线性组合, 若 \vec{b} 不在列空间, 必然无解。只有 b 是 A 中各列的线性组合时, 即 $\vec{b} \in C(A)$, 方程才有解。

2.3 零空间 (核) null space

定义: $Null(A) = \{x | Ax = 0, x \in \mathbb{R}^n\}$, 齐次线性方程组的解构成了 A 的零空间。

定理: 零空间的维数 $\dim N(A) = n - R(A) = n - \dim C(A)$, n 为列数; 也就是自由变量的个数, 主元的个数就是矩阵的秩。

3. 向量组的线性相关性、基、维数

3.1 向量组的线性相关性 (independence)

定义: 给定向量组 $A: a_1, a_2, \dots, a_m$, 如果存在不全为 0 的数 k_1, k_2, \dots, k_m , 使

$$k_1 a_1 + k_2 a_2 + \dots + k_m a_m = 0$$

则称向量组 A 是线性相关的 (linearly dependent), 否则称它线性无关 (linearly independent)。

若向量组线性无关, 则向量组向量的线性组合始终不为 $\vec{0}$ (不包括系数全为 0 的情况)

向量组能线性组合得到零向量: 相关; 线性组合得不到零向量: 无关。

随之可以与线性方程组的解联系起来。设向量组构成的矩阵 $A_{n \times m}$ ，线性方程组 $Ax = \vec{0}$ ， $x \in \mathbb{R}^m$

- 线性相关： $N(A)$ 有非零解。 $R(A) < m$ ，含有自由变量。
- 线性无关： $N(A)$ 只包含一个零向量。 $R(A) = m$ ，所有分量都是主元，没有自由变量。

3.2 基、维数

定理：对于给定的向量空间，基向量的个数都是一样的。

向量空间 V 中一组线性无关的向量组称为 V 的一个基，基向量的个数称为向量空间的维数。

因此我们说 $A = \begin{pmatrix} 1 & 3 \\ 2 & 3 \\ 4 & 1 \end{pmatrix}$ 的列空间 $C(A)$ 是三维空间的一个平面，因为 $\dim C(A) = 2$ 。

“基向量的个数不多也不少。”

基向量的性质：

- 线性无关
- 线性组合生成(spanning)整个空间

若向量组构成的矩阵为方阵，若向量组线性无关，则方阵可逆（满秩矩阵 $|A| \neq 0$ ）。

例子：求矩阵 $A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & -2 & -1 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \end{pmatrix}$ 的零空间的一组（个）基（a basis）。

就是求 $Ax = 0$ 的一个解。

对其进行行变换最终得到： $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ ，可知向量 $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$ 可以作为 $N(A)$ 的一个基向量

（不需要乘常数）

基向量个数为1，因此 $\dim N(A) = 1, R(A) = 2$ 。

注意！！！！教授称之为**Great theorem**的：

矩阵的秩 = 主元个数 = 矩阵列空间的维数，因为主列可以构成一组基，所以主元个数就是列空间维数。

例子： $A = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{pmatrix}$ ，显然第一列与第二列线性无关可以作为列空间的一组基。

经过行变换可以得到

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

可以看出第一列与第二列是主列， $R(A) = 2 = \text{主元个数}$ 。

4. 线性方程组的解

4.1 齐次线性方程组

考虑下列方程组：

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= 0 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= 0 \end{aligned}$$

可以写成矩阵形式：

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} + x_3 \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix} = \vec{0}$$

$$Ax = \vec{0}$$

其中 $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$

对 A 进行初等变换为行阶梯形矩阵，那么非零行第一个非零元素对应的列对应的未知变量我们称为主元 (pivot)，其余列称为自由变量。

例子：若方程系数矩阵 $A = \begin{pmatrix} 1 & 1 & -2 & 1 & 4 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ ，那么主元就是 x_1, x_2, x_4 ，自由变

量就是 x_3, x_5 。

- 例子：给定 A ，求 $Ax = \vec{0}$ 。

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 2 & 6 & 8 \\ 2 & 8 & 10 \end{pmatrix}$$

$$A \sim \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 4 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

可知 x_1, x_2 为主元， x_3 为自由变量，令 $x_3 = 1$ （为 0 时解为 0 向量无意义），得到 $x = (-1, -1, 1)^T$ 。

记得初等变换讲过的解的形式： $N = \begin{pmatrix} -F \\ I \end{pmatrix}$ ，这儿的 I 为1阶，即只有元素1。 I 的阶数等于自由变量的个数。

$N(A) = c\mathbf{x}, c \in R$ 。注意：当求零空间的基向量时才是 \mathbf{x} 。任意倍的 \vec{x} 生成了零空间。

求解步骤：

- 将自由变量依次设置为1，其余自由变量为0；带入 $A\mathbf{x} = 0$ 求出主元的值
- 有几个自由变量，零空间就有几个向量线性组合生成
- 自由变量个数 $n - R(A)$

4.2 非齐次线性方程组

$$A\mathbf{x} = \mathbf{b}$$

增广矩阵 (A, \mathbf{b}) ，若 $R(A) = R(A, \mathbf{b}) \Leftrightarrow$ 线性方程组有解。

求解步骤：

- 将所有自由变量设为0，带入 $A\mathbf{x} = \mathbf{b}$ ，求解主元的值，得到一个特解 \mathbf{x}_p ；($A\mathbf{x}_p = \mathbf{b}$)
- 求解 $A\mathbf{x} = 0$ ，即求 A 的零空间 \mathbf{x}_n ($A\mathbf{x}_n = 0$)
- 得到通解 $\mathbf{x}_n + \mathbf{x}_p$ ($A\mathbf{x}_n + A\mathbf{x}_p = \mathbf{b}$) 这就是非齐次线性方程组的解结构

注意：特解不需要乘常数 c

矩阵的秩与解的个数（仅讨论有解的情况即 $R(A) = R(A, \mathbf{b})$ ）：

可知 $0 \leq R(A) \leq \min\{m, n\}$

- 若 $R(A) = R(A, \mathbf{b}) = n$ ，即列满秩，此时自由变量个数为0。非零行个数为 n ，相当于 n 个方程， n 个未知变量，且 $N(A) = \mathbf{0}$ ， $A\mathbf{x} = \mathbf{b}$ 仅有一解 \mathbf{x}_p
- 若 $R(A) = R(A, \mathbf{b}) < n$ ，此时存在自由变量，因此有无限解， $\mathbf{x}_p + \mathbf{x}_n$ ， \mathbf{x}_n 是一系列向量的线性组合。
- 若 $R(A) = R(A, \mathbf{b}) = m < n$ ，即行满秩，合第二种情况相同，无限解

考虑当 $R(A) \neq R(A, \mathbf{b})$ 的情形，此时 A 中非零行个数不等于 \mathbf{b} 的非零行个数，方程无解。

- 例子

给定

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & -1 \\ 1 & 3 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ -2 \\ 5 \end{pmatrix}$$

求 $A\mathbf{x} = \mathbf{b}$ 的解。

$$\begin{pmatrix} 1 & 1 & 2 & 1 \\ 1 & 2 & -1 & -2 \\ 1 & 3 & 1 & 5 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & -3 & -3 \\ 0 & 0 & 1 & 2 \end{pmatrix} (r_3 - r_2, r_2 - r_1, r_3 - r_2, r_3/2) \sim \begin{pmatrix} 1 & 0 & 0 & -6 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

得 $\mathbf{x} = (-6, 3, 2)^T$ 。列满秩矩阵只有一个特解。

给定

$$A = \begin{pmatrix} 2 & 3 & 1 \\ 1 & -2 & 4 \\ 3 & 8 & -2 \\ 4 & -1 & 9 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 4 \\ -5 \\ 13 \\ -6 \end{pmatrix}$$

求解 $A\mathbf{x} = \mathbf{b}$,

经过行变换可得

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

主元为 x_1, x_2 , 自由变量为 x_3 , 令 $x_3 = 0$, 得 $x_1 = -1, x_2 = 2$, 特解 $\mathbf{x}_p = (-1, 2, 0)^T$

当 $x_3 = 1$, 得 $(-2, 1, 1)^T$.

因此

$$\mathbf{x} = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} + c \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}, c \in \mathbb{R}$$

5. 四个基本子空间

给定矩阵 $A_{m \times n}$

列空间 $C(A)$

零空间 $N(A)$: $A\mathbf{x} = \vec{0}, \mathbf{x} \in \mathbb{R}^n$

行空间 $C(A^T)$: 矩阵 A 经过初等变换后, 非零行即可作为行空间的一组基。

A^T 的零空间(左零空间) $N(A^T)$: $A^T\mathbf{x} = \vec{0}, \mathbf{x} \in \mathbb{R}^m$ 或者写成 $\mathbf{x}^T A = \mathbf{0}$ 这也是为什么叫左零空间的原因。

考虑 A 到其行最简形矩阵 R 的变换过程, 现在我们想知道 A 乘以一个什么矩阵可以变到 R , 即 $EA = R$, 求 E ?

注意 A 是任意矩阵 (方阵或者长方形矩阵) !

考虑求逆矩阵时在方阵后加上一个同阶方阵然后行变换的过程, 同样在 A 后加一个 m 阶单位阵, 有

$$(A_{m \times n}, I_{m \times m}) \rightarrow (R_{m \times n}, E_{m \times m})$$

R 为矩阵 A 的行阶梯矩阵, 此时 E 就是我们要求的那个变换矩阵, 即 $EA = R$, 此时 R 中的零行对应矩阵 E 的行就是 $N(A^T)$ 的一个基向量。

四个子空间的正交性(orthogonality)

- $C(A) \perp N(A^T)$

证明：设 $\vec{v} \in C(A)$, $\mathbf{v} = A\mathbf{x}$, $\mathbf{v} \in \mathbb{R}^m$, 即列空间的线性组合,
 $\vec{u} \in N(A^T)$, $A^T u = \vec{0}$, $\vec{u} \in \mathbb{R}^m$ $\mathbf{v}^T u = (A\mathbf{x})^T u = \mathbf{x}^T A^T u = \mathbf{x}^T (\vec{0}) = 0$. 所以 u 和 v 正交, 由于两个空间的任意向量都相互正交, 因此这两个空间正交。

• $C(A^T) \perp N(A)$

证明：显然 $N(A) = \{\mathbf{x} | A\mathbf{x} = \mathbf{0}\}$, 即 \mathbf{x} 与 A 中每个行向量正交 $\mathbf{r}_i \mathbf{x} = 0$, \mathbf{r}_i 是行向量。且行空间由行向量的线性组合得到, 因此 \mathbf{x} 与行空间任意向量正交。即 $\mathbf{a} = c_1 \mathbf{r}_1 + c_2 \mathbf{r}_2 \in C(A^T)$, $\mathbf{a}^T \mathbf{x} = (c_1 \mathbf{r}_1 + c_2 \mathbf{r}_2) \mathbf{x} = 0$.

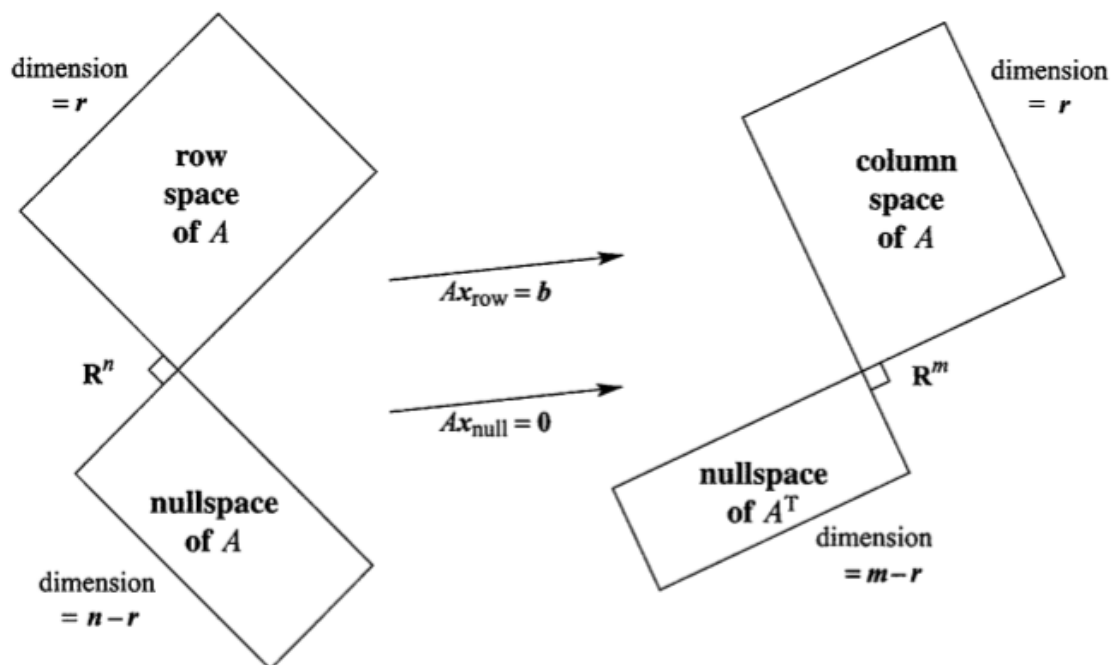


Figure 4.2: Two pairs of orthogonal subspaces. The dimensions add to n and add to m . This is an important picture—one pair of subspaces is in \mathbb{R}^n and one pair is in \mathbb{R}^m .

6. 正交向量与投影

6.1 正交向量

给定列向量 $\vec{u}, \vec{v} \in \mathbb{R}^n$, 若 $\vec{u}^T \vec{v} = \vec{v}^T \vec{u} = 0$, 则 \vec{u}, \vec{v} 正交。零向量 $\mathbf{0}$ 与任意向量正交 (orthogonal)。

$\|\vec{u}\|$ 表示向量 \vec{u} 的模长, 即 $\|\vec{u}\| = \sqrt{\vec{u}^T \vec{u}}$ 。注意: $\vec{u} \vec{u}^T$ 结果是一个 $n \times n$ 的矩阵!

6.2 正交子空间与正交基

子空间 S, T , 若 S 与 T 正交, 则 S 中任意向量与 T 的任意向量正交。四个基本子空间的正交性上一章已经讨论过。

若空间的一组基向量两两正交, 就是一组正交基, 在此条件上, 若每个向量的模长为1, 则称这组基向量为标准正交基。

例子: \mathbb{R}^3 的一组标准正交基: $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

6.3 投影与投影矩阵



根据上图可得：

$$\begin{aligned} \mathbf{p} &= x\mathbf{a}, x \in R \\ \mathbf{a} \perp \mathbf{e} &\Rightarrow \mathbf{a}^T(\mathbf{b} - x\mathbf{a}) = 0 \Rightarrow x\mathbf{a}^T\mathbf{a} = \mathbf{a}^T\mathbf{b} \end{aligned}$$

$$x = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$$

$$\mathbf{p} = \mathbf{a} \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$$

我们称

$$P = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}$$

为由 \mathbf{b} 到 \mathbf{a} 的投影变换矩阵，即 $\mathbf{p} = P\mathbf{b}$ 。直观上理解，任意向量与 P 相乘，结果都会在 P 的列空间内。

投影矩阵 P 的性质：

- $P^2 = P$ ，幂等， $P(P\mathbf{b}) = P\mathbf{b}$
- $P^T = P$ ，对称

Q: 现在考虑这个问题：线性方程组 $A\mathbf{x} = \mathbf{b}$ 无解的时候，即 $\mathbf{b} \notin C(A)$ ，如何求一个近似（最优解）？

A: 有了投影我们就可以把 \mathbf{b} 投影到 $C(A)$ 中得到 $\hat{\mathbf{b}}$ ，从而 $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ 有解， $\hat{\mathbf{x}}$ 即为一个近似解。

那么问题随之而来，如何得到 $\hat{\mathbf{b}}$ ？也就是问这里的投影矩阵 P 如何求得？

观察上图，我们发现有个误差向量 $\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$ ，在本例中，它是垂直于 $C(A)$ 的，在讲四个子空间时我们说过 $C(A) \perp N(A^T)$ ，因此， $\mathbf{e} \in N(A^T)$ ，有：

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$$

$$A^T\mathbf{b} = A^T A\hat{\mathbf{x}}$$

我们希望这个方程有解，也就是求出 $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$ 。问题是 $A^T A$ 的可逆性需要进一步讨论，即可能是可逆的也可能是不可逆的。那么 $A^T A$ 这个方阵可逆的条件是什么？答案是当且仅当 A 各列线性无关时， $A^T A$ 可逆

（ A 为 $m \times n$ 的矩阵， $A^T A$ 就是一个 $n \times n$ 的对称方阵）。下面给出证明。

证明：

不妨先假设 $A^T A\mathbf{x} = \mathbf{0}$ ，即有： $\mathbf{x}^T A^T A\mathbf{x} = 0 \Rightarrow (A\mathbf{x})^T A\mathbf{x} = 0$ ，即 $\|A\mathbf{x}\| = 0 \Rightarrow A\mathbf{x} = \mathbf{0}$

又 A 各列线性无关，所以 $A\mathbf{x} = \mathbf{0}$ 只有零解（线性无关的定义），所以 $\mathbf{x} = \mathbf{0}$ 。即 $A^T A\mathbf{x} = \mathbf{0}$ 只有零解， $A^T A$ 中各列线性无关， $R(A^T A) = n$ ，满秩矩阵一定可逆（秩的定义， $|A^T A| \neq 0$ ），得证。

好了，我们已经得到了 $\hat{\mathbf{x}}$ 的表达式，现在就可以求出投影矩阵了。

$$\hat{\mathbf{b}} = A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T \mathbf{b}$$

这里的投影矩阵

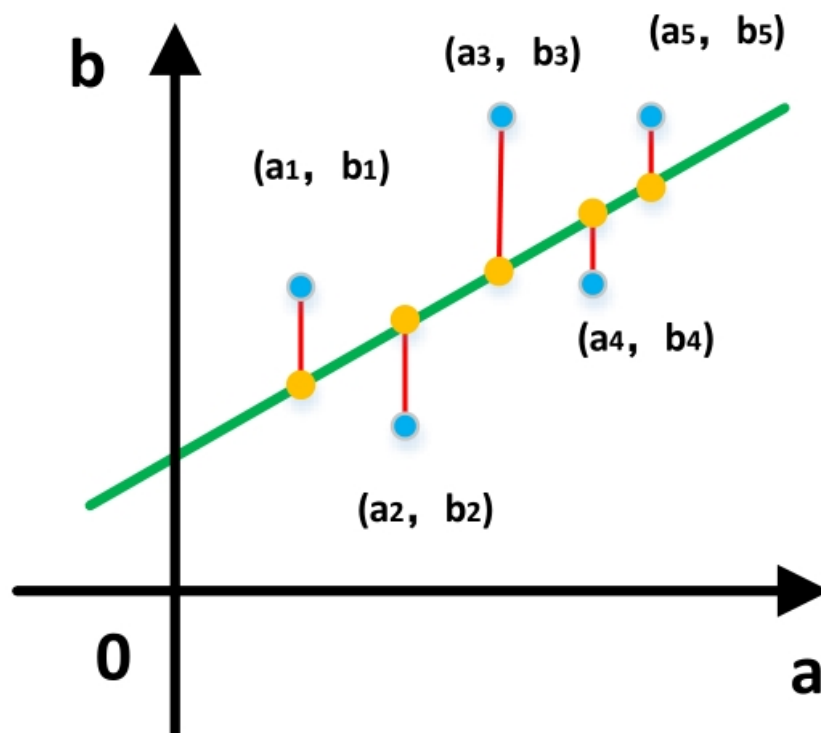
$$P = A(A^T A)^{-1} A^T$$

P 依旧满足投影矩阵的两个性质。特别的，当 $\mathbf{b} \perp C(A)$ ， $P\mathbf{b} = \mathbf{0}$ 。

6.4 最小二乘(Least Squares)

最小二乘的思想最小化理论测与观测值的平方和： $\min \sum_0^n y - y_i$

给定一系列二维平面上的点，要拟合一条直线 $y = kx + b$ ， k, b 未知；就用到了最小二乘。下面先讨论使用最小二乘拟合直线，再进一步探讨最小二乘与投影的奇妙联系。



$$\begin{aligned} kx_1 + b &= y_1 \\ kx_2 + b &= y_2 \\ &\dots \\ kx_m + b &= y_m \end{aligned}$$

写成矩阵形式就是

$$A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_m & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{pmatrix}, \mathbf{x} = \begin{pmatrix} k \\ b \end{pmatrix}$$

求解方程组 $A\mathbf{x} = \mathbf{b}$ 。接下来我们讨论这个方程无解的时候该怎么办？即根本没有一条直线同时经过上图中所有蓝色的点！为了选取最合适的 \mathbf{x} ，让该等式“尽量成立”，引入残差平方和函数 H ：

$$\min(H) = \arg \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2^2$$

（下标表示2范式，上标表示平方）。这也就是最小二乘法的思想。当 \mathbf{x} 取最优值的时候， $A\mathbf{x}$ 恰好对应图中线上橙色的点，而 \mathbf{b} 则对应图中蓝色的点， e 的值则应红色的线长。

现在的问题是如何求最优的 \mathbf{x} ？我们可以求关于 \mathbf{x} 的导数，并令其结果等于0，即：

$$y = \|A\mathbf{x} - \mathbf{b}\|_2^2 = (A\mathbf{x} - \mathbf{b})^T (A\mathbf{x} - \mathbf{b}) = \mathbf{x}^T A^T A \mathbf{x} - \mathbf{x}^T A^T \mathbf{b} - \mathbf{b}^T A \mathbf{x} + \mathbf{b}^T \mathbf{b}$$

$$\frac{\partial y}{\partial \mathbf{x}} = 2A^T A \mathbf{x} - 2A^T \mathbf{b} = 0$$

即最优 \mathbf{x} 满足 $A^T A \mathbf{x} = A^T \mathbf{b}$ ，是不是有点眼熟？

投影与最小二乘的联系

最小二乘的思想是想如何选取参数 \mathbf{x} 使得 H 最小。而从向量投影的角度来看这个问题， $H = \|\mathbf{b} - A\mathbf{x}\|^2$ 就是向量 \mathbf{e} 长度的平方(参考6.3的图)，如何才能使 \mathbf{e} 的长度最小呢？ \mathbf{b} 和 $\mathbf{x}_1, \mathbf{x}_2$ 都是固定的，当然是 \mathbf{e} 垂直 $\mathbf{x}_1, \mathbf{x}_2$ 平面的时候长度最小！换句话说：最小二乘法的解与矩阵投影时对变量求解的目标是一致的！最小二乘的求解实质上就是 $A\mathbf{x} = \mathbf{b}$ 没有解，我们就把 \mathbf{b} 投影到 $C(A)$ 内，此时误差向量或者说 H 最小（垂直于 $C(A)$ ）。

7. 正交矩阵与正交化法

7.1 正交矩阵（只讨论实矩阵）

注意：Strang教授在视频中讲正交矩阵时，谈到了非方阵的正交矩阵，而张贤达的《矩阵分析与应用》一书中对正交矩阵的定义仅限于方阵。下面以教授视频教授的为主。

定义： A 为 $m \times n$ ，若 $A^T A = I$ ，则 A 为正交矩阵，即 A 中的列向量两两正交且模长为1！（列向量是标准正交向量组）当 A 为 n 阶方阵，可得 $A^T = A^{-1}$ 。

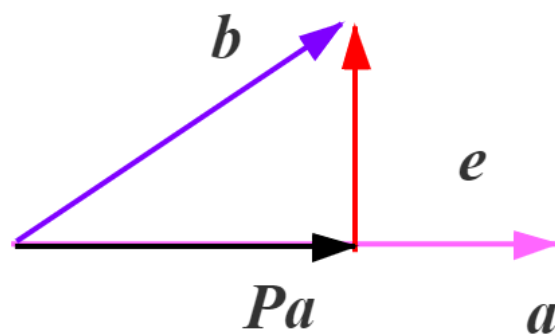
举个正交矩阵的例子：
$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

当某个向量投影到 $C(A)$ 中时，投影矩阵 $P = A(A^T A)^{-1} A^T = A A^T$ ，当 A 为方阵， $A^T A = A A^T = P = I$ ，投影矩阵是单位阵。

考虑 $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$ 当 A 为正交矩阵时， $\hat{\mathbf{x}} = A^T \mathbf{b}$ ， $\hat{\mathbf{x}}_i = \mathbf{a}_i^T \mathbf{b}$ ， $\hat{\mathbf{x}}$ 的第 i 个分量等于 A 中第 i 列的列向量与 \mathbf{b} 的内积。可以看到当 A 为正交矩阵时，求解 $A\mathbf{x} = \mathbf{b}$ 形式非常简单，那给定一个非正交矩阵 A ，如何把它化为正交矩阵呢？这就是下一节讨论的正交化法。

7.2 Gram-Schmidt 正交化法

首先考虑两个不垂直的向量 \mathbf{a}, \mathbf{b} ，如图所示：



现在的问题是，在这两个向量构成的空间中，求出一组标准正交基，即求 \mathbf{p}, \mathbf{q} 使得 $\mathbf{p} \perp \mathbf{q}$?

不妨令 $\mathbf{p} = \mathbf{a}$ ，那么与 \mathbf{a} 垂直的向量该如何求？之前投影讲到， \mathbf{b} 投影到 \mathbf{a} 上时，误差向量 \mathbf{e} 是垂直于 \mathbf{a} 的。

$$\mathbf{q} = \mathbf{e} = \mathbf{b} - \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}\mathbf{b} = \mathbf{b} - \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}\mathbf{a}$$

$\mathbf{q} \perp \mathbf{a}$

$\mathbf{a}^T\mathbf{b}$ 是一个常数，那么 \mathbf{a} 放在左右也就无所谓了，写成内积形式即：

$$\mathbf{b} - \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\langle \mathbf{a}, \mathbf{a} \rangle} \mathbf{a}$$

如果三个向量呢，第三个向量减去 \mathbf{a} 、 \mathbf{b} 上的分量就可以了。之后对所得向量除以模长即可得到标准正交向量组。这就是 Gram-Schmidt 正交化。

我们可以看到，本质上正交化就是求给定矩阵 \mathbf{A} 列空间 $C(\mathbf{A})$ 的一组标准正交基。

7.3 QR分解

给定向量组 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ ，向量组构成的矩阵 \mathbf{A} ，对向量组正交化后得到 $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ ，正交向量组组成矩阵 \mathbf{Q} ，那 \mathbf{A} 与 \mathbf{Q} 有何关联？事实上 $\mathbf{A} = \mathbf{Q}\mathbf{R}$ ，矩阵 \mathbf{R} 是一个上三角矩阵，由于 \mathbf{q}_i 是 $C(\mathbf{A})$ 的一组标准正交基，因此 \mathbf{A} 中的每个列向量都是由 \mathbf{q}_i 的线性组合构成， \mathbf{R} 恰好是 \mathbf{q}_i 线性组合的系数。

$$\mathbf{R} = \begin{pmatrix} \mathbf{q}_1^T \mathbf{x}_1 & \mathbf{q}_1^T \mathbf{x}_2 & \dots & \mathbf{q}_1^T \mathbf{x}_n \\ 0 & \mathbf{q}_2^T \mathbf{x}_2 & \dots & \mathbf{q}_2^T \mathbf{x}_n \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mathbf{q}_n^T \mathbf{x}_n \end{pmatrix}$$

$$\mathbf{Q} = (\mathbf{q}_1 \quad \mathbf{q}_2 \quad \dots \quad \mathbf{q}_n)$$

8. 行列式及其性质

9. 方阵的特征值与特征向量

注意：特征值与特征向量都是针对方阵而言！

9.1 方阵的特征值

给定 n 阶方阵 A ，若存在非零向量 $\mathbf{x} \in \mathbb{R}^n$ ，以及实数 $\lambda \in \mathbb{R}$ ，有：

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

则称 λ 为 A 的特征值 (eigenvalue)，非零向量 \mathbf{x} 为 A 对应于 λ 的特征向量 (eigenvector)。特征向量经过矩阵 A 的变换后，方向没有发生改变，只是模长发生了改变，这就是特征向量的神奇之处。

给定 n 阶方阵 A ，现在考虑任意向量到 $C(A)$ 的投影矩阵 $P = A(A^T A)^{-1} A^T$ 的特征向量与特征值？上面讲过特征向量是经过矩阵变化后方向不发生改变，那什么向量经过 P 的变换方向不变？答案就是任意位于 $C(A)$ 内的向量，很容易理解在 $C(A)$ 内的向量到 $C(A)$ 的投影是向量本身，而且特征值为 1。

$$P\mathbf{x} = \mathbf{x}, \mathbf{x} \in C(A)$$

考虑置换矩阵 $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ 的特征值和对应的特征向量。

显然 $\mathbf{x} = (1, 1)^T$ 是，交换行后方向没发生变化， $\lambda = 1$ ； $\mathbf{x} = (1, -1)^T$ ， $\lambda = -1$ 。

很重要的一条定理：

- n 阶方阵 A 的特征值之和等于对角线元素之和（迹 trace）。

如何求解方阵的特征值和特征向量

$$(A - \lambda I)\mathbf{x} = \mathbf{0}, \mathbf{x} \neq \mathbf{0}$$

即对于上述方程，若存在非零解，矩阵 $A - \lambda I$ 务必不是满秩矩阵（必须存在自由变量），降秩矩阵也就意味着是奇异矩阵（不可逆矩阵），即 $|A - \lambda I| = 0$ ，现在只需计算行列式的值就可求出 λ ，之后再带回原方程，得到对应的特征向量。

- 例子：给定 $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$ ，求对应的特征值与特征向量。

$$|A - \lambda I| = \begin{vmatrix} 3 - \lambda & 1 \\ 1 & 3 - \lambda \end{vmatrix} = 0, \lambda_1 = 4, \lambda_2 = 2$$

当 $\lambda_1 = 4$ ，对应的方程：

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

解得 $\mathbf{x}_1 = (1, 1)^T$ ；

当 $\lambda_2 = 2$ ，解得 $\mathbf{x}_2 = (1, -1)^T$ 。

显然，若 \mathbf{x}_i 是矩阵 A 对应于 λ_i 的特征向量，则 $k\mathbf{x}_i (k \neq 0)$ 也是对应于 λ_i 的特征向量。

旋转矩阵的特征值与特征向量

上三角矩阵和下三角矩阵的特征值

9.2 矩阵对角化和矩阵幂乘

10. 对称矩阵与正定矩阵

11. 奇异值分解

12. 矩阵与向量微分

阅读中发现错误请联系: nefuzyj@163.com, 不胜感激!