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，$\vec{a}$、$\boldsymbol{a}$都表示列向量。

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

pdf版参见：https://github.com/zyj-nova/LinearAlgebraNotes

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：

$E_1E_2 \cdots E_nA$；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：$AE_1E_2 \cdots E_n$。

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$\begin{pmatrix}I&F\end{pmatrix}\begin{pmatrix}x_{pivot}\\x_{free}\end{pmatrix}=0\Rightarrow x_{pivot}=-Fx_{free}$

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。

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

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：$P(AB) = I \Rightarrow PABB^{-1} = IB^{-1} \Rightarrow PA = B^{-1} \Rightarrow P = B^{-1} A^{-1}$
- 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 $x_1 c_1 + x_2 c_2 + ...$ 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 $x_1 r_1 + x_2 r_2 + ...$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$A: \boldsymbol{a_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_m}$，如果存在**不全为0**的数$k_1, k_2, \cdots, k_m$，使

$$k_1\boldsymbol{a_1} + k_2\boldsymbol{a_2} + \cdots + k_m\boldsymbol{a_m} = \boldsymbol{0}$$

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，$R(A)=2=$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：

$$\begin{aligned}a_{11}x_1+a_{12}x_2+a_{13}x3=0\\a_{21}x_1+a_{22}x_2+a_{23}x3=0\\a_{31}x_1+a_{32}x_2+a_{33}x3=0\end{aligned}$$

可以写成矩阵形式：

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

$$Ax=0$$

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$Ax=0$.

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

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

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。

$N(A) = c\boldsymbol{x}, c \in R$。**注意**：当求零空间的基向量时才是$\boldsymbol{x}$。任意倍的$\vec{\boldsymbol{x}}$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$Ax = 0$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

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

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

求解步骤：

- 将所有自由变量设为0，带入$Ax = b$，求解主元的值，得到一个特解x_p；$(Ax_p = b)$
- 求解$Ax = 0$，即求A的零空间x_n（$Ax_n = 0$）
- 得到通解$x_n + x_p$（$Ax_n + Ax_p = b$）这就是非其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$R(A) = R(A, b)$）：

可知0 $<= R(A) <= \min\{m, n\}$

- 若$R(A) = R(A, b) = n$，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 $N(A) = \boldsymbol{0}$，$Ax = b$仅有一解 x_p
- 若$R(A) = R(A, b) < n$，此时存在自由变量，因此有无限解，$x_p + x_n$，x_n是一系列向量的线性组合。
- 若$R(A) = R(A, b) = m < n$，即**行满秩**，合第二种情况相同，无限解

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

- 例子

给定

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

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

给定

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

求解$A\boldsymbol{x}=\boldsymbol{b}$,

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$\boldsymbol{x}_p=(-1,2,0)^T$

当$x_3=1$，得$(-2,1,1)^T$.

因此

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

5. 四个基本子空间

给定矩阵$A_{m\times n}$

列空间$C(A)$

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

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)$N(A^T)$：$A^T\boldsymbol{x}=\vec{0},\boldsymbol{x}\in\mathbb{R}^m$或者写成$\boldsymbol{x}^TA=0$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$\vec{v} \in C(A)$，$\boldsymbol{v}=A\boldsymbol{x}$，$\boldsymbol{v} \in \mathbb{R}^m$，即列空间的线性组合，$\vec{u} \in N(A^T), A^T u=\vec{0}$，$\vec{u} \in \mathbb{R}^m v^T u=(Ax)^T u=x^T A^T u=x^T(A^T u)=0$.所以 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 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$N(A)=\{x|A\boldsymbol{x}=\boldsymbol{0}\}$，即 $\boldsymbol{x}$ 与 A 中每个行向量正交$\boldsymbol{r_i}\boldsymbol{x}=0$，$\boldsymbol{r_i}$是行向量。且行空间由行向量的线性组合得到，因此 $\boldsymbol{x}$ 与行空间任意向量正交。即 $\boldsymbol{a}=c_1\boldsymbol{r_1}+c_2\boldsymbol{r_2} \in C(A^T), \boldsymbol{a}^T\boldsymbol{x}=(c_1\boldsymbol{r_1}+c_2\boldsymbol{r_2})\boldsymbol{x}=0$.

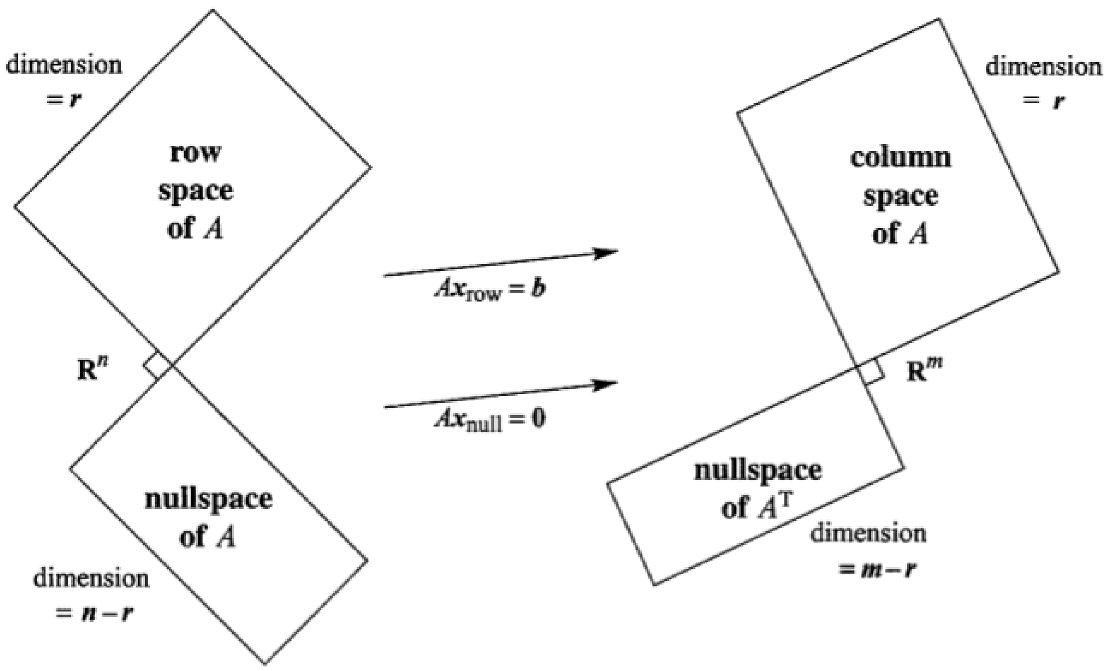


Figure 4.2: Two pairs of orthogonal subspaces. The dimensions add to n and add to m. **This is an important picture**—one pair of subspaces is in $\mathbf{R}^n$ and one pair is in $\mathbf{R}^m$.

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**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：

$$\boldsymbol{p}=x\boldsymbol{a},x\in R$$
$$\boldsymbol{a}\perp\boldsymbol{e}\Rightarrow\boldsymbol{a}^T(\boldsymbol{b}-x\boldsymbol{a})=0\Rightarrow x\boldsymbol{a}^T\boldsymbol{a}=\boldsymbol{a}^T\boldsymbol{b}$$

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

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

我们称

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

为由 $\boldsymbol{b}$ 到 $\boldsymbol{a}$ 的投影变换矩阵，即 $\boldsymbol{p}=P\boldsymbol{b}$。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，$P(P\boldsymbol{b})=P\boldsymbol{b}$
- $P^T=P$，对称

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

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

那么问题随之而来，如何得到 $\hat{\boldsymbol{b}}$？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$\boldsymbol{e}=\boldsymbol{b}-A\hat{\boldsymbol{x}}$，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，$\boldsymbol{e}\in N(A^T)$，有：

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

$$A^T\boldsymbol{b}=A^TA\hat{\boldsymbol{x}}$$

我们希望这个方程有解，也就是求出$\hat{\boldsymbol{x}}=(A^TA)^{-1}A^T\boldsymbol{b}$。问题是$A^TA$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^TA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^TA可逆**

（A为$m\times n$的矩阵，A^TA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$A^TA\boldsymbol{x}=\boldsymbol{0}$，即有：$\boldsymbol{x}^TA^TA\boldsymbol{x}=\boldsymbol{0}\Rightarrow(A\boldsymbol{x})^TA\boldsymbol{x}=\boldsymbol{0}$，即 $||A\boldsymbol{x}||=0\Rightarrow A\boldsymbol{x}=\boldsymbol{0}$

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

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

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

这里的投影矩阵

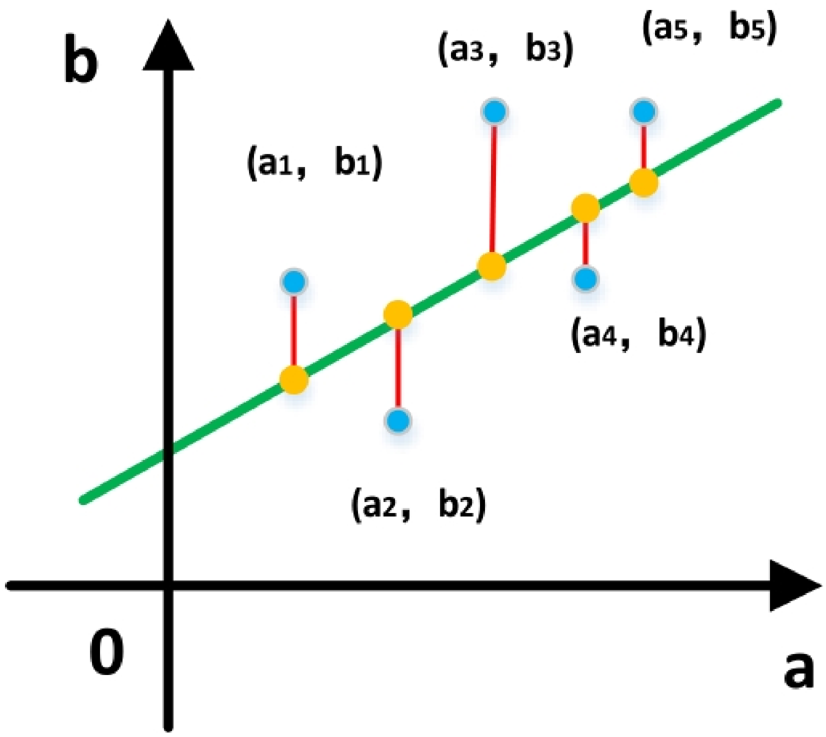
$$P=A(A^TA)^{-1}A^T$$

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

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**。



$$\begin{gathered}kx_1+b=y_1\\kx_2+b=y_2\\\cdots\\kx_m+b=y_m\end{gathered}$$

写成矩阵形式就是

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

求解方程组$A\boldsymbol{x}=\boldsymbol{b}$。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$\boldsymbol{x}$，让该等式"尽量成立"，引入残差平方和函数H：

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

（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 $\boldsymbol{x}$ 取最优值的时候，$A\boldsymbol{x}$恰好对应图中线上橙色的点，而 $\boldsymbol{b}$ 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 $\boldsymbol{x}$？我们可以求关于$\boldsymbol{x}$的导数，并令其结果等于0，即：

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

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

即最优$\boldsymbol{x}$满足$A^TA\boldsymbol{x}=A^T\boldsymbol{b}$，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 x 使得 H 最小。而从向量投影的角度来看这个问题，$H=||\boldsymbol{b}-A\boldsymbol{x}||^2$ 就是向量 $\boldsymbol{e}$ 长度的平方(参考6.3的图)，如何才能使 $\boldsymbol{e}$ 的长度最小呢？b 和x_1，x_2都是固定的，当然是e 垂直x_1，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$Ax=b$没有解，我们就把 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$A^TA=I$，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$A^T=A^{-1}$。

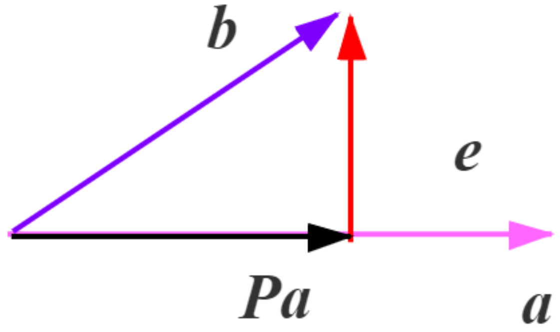
举个正交矩阵的例子：$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

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

考虑$A^TA\hat{\boldsymbol{x}}=A^T\boldsymbol{b}$当$A$为正交矩阵时，$\hat{\boldsymbol{x}}=A^T\boldsymbol{b}$，$\hat{\boldsymbol{x}}_i=\boldsymbol{a}_i\boldsymbol{b}$，$\hat{\boldsymbol{x}}$的第$i$个分量等于$A$中第$i$列的列向量与$\boldsymbol{b}$的内积。可以看到当$A$为正交矩阵时，求解$A\boldsymbol{x}=\boldsymbol{b}$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$\boldsymbol{a},\boldsymbol{b}$，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$\boldsymbol{p},\boldsymbol{q}$使得$\boldsymbol{p}\perp\boldsymbol{q}$?

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

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

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

$$\boldsymbol{b}-\frac{<\boldsymbol{a},\boldsymbol{b}>}{<\boldsymbol{a},\boldsymbol{a}>}\boldsymbol{a}$$

如果三个向量呢，第三个向量减去$\boldsymbol{a}$、$\boldsymbol{b}$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A列空间$C(A)$的一组标准正交基。

7.3 QR分解

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

$$R=\begin{pmatrix}\boldsymbol{q}_1^T\boldsymbol{x}_1 & \boldsymbol{q}_1^T\boldsymbol{x}_2 & .. & \boldsymbol{q}_1^T\boldsymbol{x}_n\\ 0 & \boldsymbol{q}_2^T\boldsymbol{x}_2 & .. & \boldsymbol{q}_2^T\boldsymbol{x}_n\\ .. & .. & .. & ..\\ 0 & 0 & .. & \boldsymbol{q}_n^T\boldsymbol{x}_n\end{pmatrix}$$
$$Q=\begin{pmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & .. & \boldsymbol{q}_n\end{pmatrix}$$

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 $\boldsymbol{x} \in \mathbb{R}^n$，以及实数 $\lambda \in \mathbb{R}$，有:

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

则称λ为A的特征值（eigenvalue），非零向量$\boldsymbol{x}$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$P = A(A^TA)^{-1}A^T$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的，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。

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

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

很重要的两个定理:

- n阶方阵A的特征值之和等于对角线元素之和（迹trace）。
- $\lambda_1 + \lambda_2 + \ldots \lambda_n = |A|$ 特征值之和等于$|A|$

如何求解方阵的特征值和特征向量

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

即对于上述方程称之为**特征方程**，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$\boldsymbol{x}_1 = (1,1)^T$;

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

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

A与A^T的特征值相同：$|A - \lambda I| = |(A - \lambda I)^T| = |A^T - \lambda I| = 0$

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

当 A 为一个上三角矩阵时，根据行列式性质 $|A-\lambda I|$ 的值恰好为 $(a_{11}-\lambda_1)(a_{22}-\lambda_2)\ldots(a_{nn}-\lambda_n)=0$，因此上/下三角矩阵的特征值即为对角线元素。

9.2 相似矩阵

在线性空间中，相似矩阵就是同一个矩阵的不同基下的表示。

设A,B是n阶矩阵，若存在可逆矩阵P，使得$P^{-1}AP=B$成立，称A与B相似。

$$A\boldsymbol{x}=\lambda\boldsymbol{x}, B=P^{-1}AP$$

$$P^{-1}APP^{-1}\boldsymbol{x}=\lambda P^{-1}\boldsymbol{x}\Rightarrow B(P^{-1}\boldsymbol{x})=\lambda(P^{-1}\boldsymbol{x})$$

由此可以看出，相似矩阵的特征值相同，$\boldsymbol{x}$为A的特征向量，那么B的特征向量等于$M^{-1}\boldsymbol{x}$。

9.3 矩阵对角化和矩阵幂乘

讨论下面内容的**大前提**：n阶方阵A存在n个线性无关的特征向量$\Leftrightarrow A$有n个互不相等的特征值（没有重复的特征值）。如果存在重复的特征值，**可能存在也可能不存在n个线性无关的**特征向量。（存在相同的k个特征值，求出来的特征向量可能是同一个，或者个数小于k）

$S^{-1}AS=\Lambda$（A,Λ 相似），S是 A 对应的特征向量按列组成的矩阵，Λ是由特征值组成的对角阵。这一过程被称为矩阵的对角化。需要注意：**矩阵A可以被对角化的充要条件是A存在n个线性无关的特征向量（S可逆）**。谨记!

$$AS=A\begin{pmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & .. & \boldsymbol{x}_n\end{pmatrix}=\begin{pmatrix}A\boldsymbol{x}_1 & A\boldsymbol{x}_2 & .. & A\boldsymbol{x}_n\end{pmatrix}=\begin{pmatrix}\lambda_1\boldsymbol{x}_1 & \lambda_2\boldsymbol{x}_2 & .. & \lambda_n\boldsymbol{x}_n\end{pmatrix}=S\Lambda$$

$$\Lambda=\begin{pmatrix}\lambda_1 & 0 & 0 & .. & 0\\ 0 & \lambda_2 & 0 & .. & 0\\ .. & .. & .. & .. & ..\\ 0 & 0 & 0 & .. & \lambda_n\end{pmatrix}$$

$$AS=S\Lambda\Rightarrow S^{-1}AS=\Lambda\Leftrightarrow A=S\Lambda S^{-1}$$

矩阵幂乘

$$A^2=AA=S\Lambda S^{-1}S\Lambda S^{-1}=S\Lambda^2S^{-1}$$

$$A^n=S\Lambda^nS^{-1}$$

若$A\boldsymbol{x}=\lambda\boldsymbol{x}$，那么$A^2\boldsymbol{x}=\lambda A\boldsymbol{x}=\lambda^2\boldsymbol{x}$。即$A$的$n$次幂对应的特征向量没有发生改变，特征值为$\Lambda$的$n$次幂。

10. 对称矩阵与正定矩阵

10.1 对称矩阵的特征值与特征向量

$A=A^T$，则称n阶方阵为对称矩阵。

- 对称矩阵的特征值都是实数

- 若$\lambda_1 \neq \lambda_2$，$\boldsymbol{x}_1, \boldsymbol{x}_2$为其对应特征向量，则 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 正交，也就是说特征值不同的特征向量两两正交，当特征值重复的时候，求出的特征向量未必正交（但是与其他不同特征值的特征向量仍然正交），但可以通过正交化，化为正交向量。也就是说，特征矩阵总存在一个由特征向量组成的正交矩阵。

$$
\begin{gathered}
A\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1, A\boldsymbol{x}_2 = \lambda_2\boldsymbol{x}_2 \\
(A\boldsymbol{x}_1)^T = \boldsymbol{x}_1^T A^T = \lambda_1\boldsymbol{x}_1^T \\
\lambda_1\boldsymbol{x}_1^T\boldsymbol{x}_2 = \boldsymbol{x}_1^T A\boldsymbol{x}_2 = \lambda_2\boldsymbol{x}_1^T\boldsymbol{x}_2 \\
(\lambda_1 - \lambda_2)x_1^T\boldsymbol{x}_2 = 0
\end{gathered}
$$

若$\lambda_1 \neq \lambda_2$,$x_1^T\boldsymbol{x}_2 = 0$，即$\boldsymbol{x}_1, \boldsymbol{x}_2$正交。

在将矩阵对角化时：$S^{-1}AS = \Lambda \Rightarrow A = S\Lambda S^{-1}$，若$A$为对称矩阵，那么$S$中列向量两两正交，1可知$S^TS = I$，进而$S^T = S^{-1}$。因此，$A = S\Lambda S^{-1} = S\Lambda S^T$。这就是**谱分解**。

例子：图的拉普拉斯矩阵L就是一个对称矩阵，$L = U\Lambda U^T$，利用U可以对节点进行图上傅里叶变换和反傅里叶变换。

10.2 正定矩阵（positive definite matrix）

讨论完对称矩阵的特征值，我们知道一定是实数，我们另一个比较关心的是特征值的正负问题，这就讲到了正定矩阵。下面给出正定矩阵的定义。

给定对称矩阵A，非零向量$\boldsymbol{x}$，若满足：

$$
\boldsymbol{x}^T A\boldsymbol{x} > 0, \boldsymbol{x} \neq \mathbf{0}
$$

则称A为正定矩阵。当$\boldsymbol{x}^T A\boldsymbol{x} \geq 0$，我们称$A$为半正定矩阵(positive semi-definite matrix)。$\boldsymbol{x}^T A\boldsymbol{x}$为二次型。

正定矩阵的性质：

- 特征值均为正数，也就意味着行列式值为正数
- A, B为正定矩阵，$A + B$也为正定矩阵

若$A_{m\times n}$，考虑A^TA是n阶方阵，是对称矩阵，且至少是半正定矩阵（矩阵元素全为0的时候）。

$$
\boldsymbol{x}^T A^T A\boldsymbol{x} = (A\boldsymbol{x})^T(A\boldsymbol{x}) = ||A\boldsymbol{x}||^2 \geq 0
$$

11. 奇异值分解

12 线性变换

设T是一个从V_n到U_m的映射，若T满足：

- $T(\lambda v) = \lambda T(v)$
- $T(u + v) = T(u) + T(v)$

就称从V_n到V_m的线性映射或线性变换。

坐标（coordinate）：$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$是线性空间$V_n$的一组基，对于$\boldsymbol{u} \in V_n$，$\boldsymbol{u} = \alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \ldots + \alpha_n\boldsymbol{x}_n$（基向量的线性组合），$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T$称为这个基下的坐标。

例子：对于$T(\boldsymbol{x}) = A\boldsymbol{x}, \boldsymbol{x} \in \mathbb{R}^n$，$A_{m\times n}$，就是一个线性变换，考虑$A$是如何构成的?

对于$\boldsymbol{x} \in \mathbb{R}^n$，有一组基$\boldsymbol{v}_i$描述，坐标为$\boldsymbol{\alpha}$，对于$T(\boldsymbol{x}) \in \mathbb{R}^m$，有一组基$\boldsymbol{w}_i$描述，坐标为$\boldsymbol{\beta}$。

可知$\boldsymbol{x} = \boldsymbol{\alpha}^T \boldsymbol{v}, T(\boldsymbol{v}_1) = \beta_{11}\boldsymbol{w}_1 + \beta_{21}\boldsymbol{w}_2 + \ldots$，即坐标构成了第一列，同理第二列、第三列....这是线性变换的矩阵形式。

注意：求导也是一种线性变换。

13. 矩阵与向量微分

根据函数值类型的不同可以分为：标量函数、向量函数、矩阵函数；根据输入类型的不同：标量、向量、矩阵，两两组合可以分为：标量函数对标量的导数、标量对向量的偏导（比如梯度）、标量对矩阵的偏导；向量对标量的偏导、向量对向量的偏导、向量对矩阵的偏导等等。

关于向量微分，转置与否对结果有很大影响，本章讨论时，统一以输入的形状为准，即当输入是列向量，偏导结果仍为列向量；输入是行向量，偏导结果仍为行向量。

13.1 标量对向量的偏导

$f(\boldsymbol{x}) \in \mathbb{R}, \boldsymbol{x} \in \mathbb{R}^n$，$\frac{\partial f}{\partial \boldsymbol{x}} = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots \frac{\partial f}{\partial x_n})^T$称为梯度向量。

黑塞矩阵 **Hessian Matrix**

$\frac{\partial f}{\partial \boldsymbol{x}^2}$称为黑塞矩阵，即$f$对多元变量的二阶偏导所构成的矩阵。黑塞矩阵与多元函数的极值判定密切相关。

$$H = \begin{pmatrix} \frac{\partial f^2}{\partial x_1^2} & \frac{\partial f^2}{\partial x_1 \partial x_2} & \cdots & \frac{\partial f^2}{\partial x_1 \partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f^2}{\partial x_n \partial x_1} & \frac{\partial f^2}{\partial x_n \partial x_2} & \cdots & \frac{\partial f^2}{\partial x_n^2} \end{pmatrix}$$

即：$H_{ij} = \frac{\partial f^2}{\partial x_i \partial x_j}$.

13.2 向量对向量的偏导

注意向量对向量的偏导结果是一个矩阵。

$f(\boldsymbol{x}) \in \mathbb{R}^m, \boldsymbol{x} \in \mathbb{R}^n$，$\frac{\partial f}{\partial \boldsymbol{x}} = \frac{\partial(y_1, y_2, \ldots, y_m)}{\partial(x_1, x_2, \ldots, x_n)}$称为雅可比矩阵

雅可比矩阵 **Jacobi Matrix**

$$\begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

可知雅可比矩阵的形状：$m \times n$。

二次型的偏导

$f(\boldsymbol{x}) = \boldsymbol{x}^T P \boldsymbol{x} + q^T \boldsymbol{x} + r, r \in \mathbb{R}$