$Ax=b$

$\Rightarrow A_{1,:}x = b_1$

$A_{2,:}x = b_2$

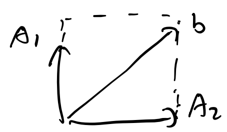
$\vdots$

$A_{m,:}x = b_m$

· 逆矩阵：A^{-1}　　def: $A^{-1}A = I$　　(对于方阵，右逆=左逆)

· 对于 $Ax=b$：

- A^{-1} 存在：对于每个b 恰好存在一解
- 可能无解/多解
- x, y 为 $Ax=b$ 解 $\Rightarrow \alpha x + (1-\alpha) y$ 也是 $Ax=b$ 的解.
- $Ax=b$ 可以看成将A中列向量组合如果达到b.（线性组合）



· linear combination: $\sum_i c_i v^{(i)}$

· span (生成子空间): set of every linear combination for some set of vectors.

· $Ax=b$ 有解 $\Leftrightarrow$ $\vec{b}$ 在A列向量的span中（即 $\vec{b} \in col(A)$ or A的值域）

- $Ax=b$ 对任意 b 有解：$col(A) = \mathbb{R}^m$

  必要条件：$n \geq m$. 充分条件：恰有m个 linearly independent 的向量 $\in col(A)$

- $Ax=b$ 唯一解：A可逆方阵.

· 范数：(norm)

- $\|x_p\| = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$, $p \in \mathbb{R}$, $p \geq 1$
- 作用：vector 映射到 scalar.
- 性质：① $f(x) = 0$

  ② $f(x+y) \leq f(x) + f(y)$

③ $\forall \alpha \in R$，$f(\alpha x)=|\alpha| f(x)$.

· p=2时：为Euclidean norm.

· $L^2 \stackrel{\text{def}}{=} \|x\|$

· p=1时：$L^1=\sum_i |x_i|$

当0与非0元素之间差异重要时使用.

· L^0范数：表示非零元素数量.

· L^∞范数：max norm（最大范数）

$$\|x\|_\infty = \max_i |x_i|$$

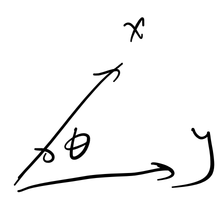
（表示绝对值最大的元素）

· Frobenius范数：$\|A\|_F=\sqrt{\sum_{i,j} A_{ij}^2}$

· 用于矩阵

· 点积可以用范数表示：

$$x\cdot y = \|x\|_2 \|y\|_2 \cos\theta$$



· 特殊矩阵，向量：

· 对角矩阵：

· $D_{ij} \stackrel{\text{def}}{=} \begin{cases} d_{ii}, & i=j \\ 0, & i\neq j \end{cases}$

· diag(v)：表示以向量v构造的对角矩阵.

- 好处：乘法计算高效

↳ 因为 $diag(v)x = v \odot x$，复杂度 $O(n)$

且 $diag(v)^{-1} = diag([1/v_1, \cdots, 1/v_n]^T)$

⇕

$$\begin{pmatrix} \frac{1}{v_1} & & & \\ & \frac{1}{v_2} & & \\ & & \cdots & \\ & & & \frac{1}{v_n} \end{pmatrix}$$

- 并非所有矩阵都是方阵。

ex：

$$\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & & & 0 & 0 \\ & 1 & & 0 & 0 \\ & & 1 & 0 & 0 \end{pmatrix}$$

也能高效计算

- 对称矩阵：
  - def：$A = A^T$
  - 用于不依赖顺序的双参数函数。
    如 A_{ij} 表示 i 到 j 的距离，$A_{ij} = A_{ji}$.
  - 一定能找到一组正交单位 eigenvector
  - 性质：① 特征值均为实数
    ② 特征向量正交.（同一特征值能选择正交）
  - 定理：主元有几个正数，特征值就有几个正数

- 单位向量 (unit vector)：
  - $\|x\|_2 = 1$

- 正交矩阵：

orthonormal：① 向量之间相互正交
② norm = 1

  - def：行向量、列向量分别标准正交的矩阵

- $A^TA = AA^T = I$
- $A^{-1} = A^T$
- 好处：求逆代价非常小

- 特征分解：(eigendecomposition)
  - def：将矩阵分解成一组特征向量和特征值
  - 目的：将矩阵分解为特征向量与特征值。

$$Av = \lambda v$$

（同样可以定义：$v^TA = v^T\lambda$）

- v可以缩放，所以我们只考虑单位特征向量。
- 假设 A 有 n个线性无关的特征向量 $\{v^{(1)}, \dots, v^{(n)}\}$

  对应特征值 $\{\lambda_1, \dots, \lambda_n\}$

令 $V = [v^{(1)} \dots v^{(n)}]$，$\lambda = [\lambda_1; \dots; \lambda_n]^T$

则 $AV = V\,\mathrm{diag}(\lambda) \Rightarrow A = V\,\mathrm{diag}(\lambda)V^{-1}$

- 每个实对称矩阵可以分解成实特征向量与实特征值：

$$A = Q\Lambda Q^T$$

对称矩阵　　正交矩阵（特征向量）　　对角矩阵（特征值）

☆ $A\,diag(\lambda)$：A列向量线性组合 —— $(\,[\,][\,][\,]\,)\begin{pmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_L\end{pmatrix}$

$diag(\lambda)\cdot A$ ：归一化处理

· 奇异值分解：(SVD: singular value decomposition)

· 将矩阵分解分解为奇异向量与奇异值 (singular vector、singular value)

· 每个实数矩阵都有一个奇异值分解。(但不一定有特征分解)

· $\underset{m\times n}{A} = \underset{m\times m}{U}\,\underset{m\times n}{D}\,\underset{n\times n}{V^T}$ (U、V为正交矩阵，D为对角矩阵)

U的列向量：左奇异向量，是 AA^T 的特征向量

V的列向量：右奇异向量。 A^TA 的特征向量

· A的奇异值 $\overset{def}{=}$ square roots of eigenvalues of A^TA

· SVD最有用的性质是拓展矩阵求逆到非方矩阵上.

· $A^TA = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T\Sigma V^T$

$AA^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T$

· 补充：

· orthogonally diagonalizable：exist othogonal matrix P ($P^{-1}=P^T$)、diagonal matrix D

such that $A = PDP^T = PDP^{-1}$

· Thm：Orthogonally diagnalizable $\Longleftrightarrow$ symmetric

· SVD 推导:

· 背景:

A : $m \times n$ matrix

A^TA is symmetric and othogononally diagonalized

$\{v_1, \cdots, v_n\}$: an othonormal basis for $\mathbb{R}^n$ consisisting of eigenvectors for A^TA

$\lambda_1 \cdots \lambda_n$: eigenvalues of $\mathbb{R}^n$

· $\|Av_i\|^2 = (Av_i)^T(Av_i) = v_i^TA^TAv_i = v_i^T\lambda_iv_i = \lambda_i$

$\therefore$ eigenvalues of A^TA are nonnegative.

let $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n \geq 0$

· singular value : $\sigma_i = \sqrt{\lambda_i}$, $\sigma_1 \cdots \sigma_n$

· singular value σ_i is the length of Av_i

· Thm : suppose A has r nonzero singular values.

Then $\{Av_1, \cdots Av_r\}$ is an othogonormal basis for Col A

, rank $= r$

· thm : $D = diag(\sigma)$, $\sigma := [\sigma_1, \cdots, \sigma_r]$, $\sigma_1 \geq \cdots \geq \sigma_r > 0$.

there exists an $m \times m$ othogonormal matrix U

$n \times n$ othogonormal matrix V

such that : $A = U\Sigma V^T$

· SVD 计算：

1、$A^TA = V\Sigma^T\Sigma V^T \Rightarrow V$

2、$AA^T = U\Sigma\Sigma^TU^T \Rightarrow U$.

· Moore-Penrose 伪逆：

· 问题：解 $Ax=y$，希望得到左逆 B，使得 $x=By$

· A的伪逆定义：

$$A^+ = \lim_{\alpha \to 0}(A^TA+\alpha I)^{-1}A^T$$

· 计算：

$A^+ = VD^+U^T$ （奇异值分解得到 V, D, U）

$\hookrightarrow D^+ = diag([1/\sigma_1, \cdots, 1/\sigma_n])^T$

· $m<n$ 时，$x=A^+y$ 为所有可行解中 $\|x\|_2$ 最小的

$m>n$ 时，可能无解，此时得到的解 x 使得 Ax 与 y 欧几里德距离 $\|Ax-y\|_2$ 最小

迹运算：

· def：$Tr(A) = \sum_i A_{i,i}$

· 迹运算描述矩阵 Frobenius 范数：

$$\|A\|_F = \sqrt{Tr(AA^T)}$$

· $Tr(A) = Tr(A^T)$

· $Tr(X_1X_2\cdots X_n) = Tr(X_nX_1\cdots X_{n-1})$

将最后一个矩阵移最前面，迹不变.

- 即使置换后矩阵形状变了，迹仍然不变

  $A \in R^{m\times n}$, $B \in R^{n\times m}$, $AB \in R^{m\times m}$, $BA \in R^{n\times n}$

  但 $Tr(AB) = Tr(BA)$

- a 为标量，则 $a = Tr(a)$

- 行列式：

  - $\det(A)$，是个将方阵A映射到实数的函数.

  - 行列式等于特征值乘积

  - 行列式的绝对值用来衡量矩阵参与矩阵乘法后空间扩大/缩小了多少

  - $\det(A)=0$，那么空间至少沿着某一维完全收缩，失去了所有体积

  - $\det(A)=1$，保持空间体积不变

※ 半正定矩阵 (positive semidefine matrix)

- def：设A为对称矩阵，若对任意向量 x ($x\neq0$)，都有 $x^TAx\geq0$，则称A为半正定矩阵。

- 性质：I、A为半正定 $\Rightarrow \det A\geq0$

  II、A、B为半正定 $\Rightarrow$ A+B为半正定

  III、$a\geq0$，A半正定 $\Rightarrow$ $a\cdot A$为半正定。

  scalar

- 判定：若A对称：

  A半正定 $\Leftrightarrow$ A的特征值非负 $\Leftrightarrow \exists\, C\in R^{n\times n},\ A=C^TC \Leftrightarrow \mathrm{Rank}(B)=r,\ B\in R^{r\times n},\ A=B^TB$.

※ 初等矩阵：(elementary matrix)

1. 行互换
2. 行倍数
3. j行倍数 + i行

I、行互换：i行、j行所有元素互换。

$$T_{ij}=\begin{bmatrix}1 & & & & \\ & \ddots & & & \\ & & 0 & 1 & \\ & & 1 & 0 & \\ & & & & \ddots \\ & & & & & 1\end{bmatrix}$$

$T_{ij}^{-1}=T_{ij}$

$\mathrm{Det}(T_{ij})=-1$

$\therefore |T_{ij}A|=|A|$

II、行倍数：$T_i(m)$ i行所有元素乘m.

$$T_i(m)=\begin{bmatrix}1 & & & & \\ & \ddots & & & \\ & & m & & \\ & & & 1 & \\ & & & & \ddots \\ & & & & & 1\end{bmatrix}$$

$T_i(m)^{-1}=T_i(\frac{1}{m})$

$\det(T_i(m))=m$

$\therefore |T_i(m)A|=m|A|$

III、i行 + m×(j行)

$$T_{ij}(m)=\begin{bmatrix}1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \ddots & \\ & & m & & 1 & \\ & & & & & \ddots \\ & & & & & & 1\end{bmatrix}$$

$T_{ij}(m)^{-1}=T_{ij}(-m)$

$\det(T_{ij}(m))=1$

$\therefore |T_{ij}(m)\cdot A|=|A|$