回顾上节课内容:

1. 奇异值分解定理(SVD):

给定$A \in M_{m\times n}(\mathbb{R})$秩为$r$，存在$m$阶正交矩阵$U$和$n$阶正交矩阵$V$使得 $A = U\Sigma V^T$,

其中，$\Sigma = \begin{bmatrix} \Sigma_r & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix}$, $\Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \in M_r(\mathbb{R})$

且$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$ ⬅ 按重要性顺序对奇异值排序

$\sigma_1, \dots, \sigma_r$称为$A$的奇异值

步骤：1）对半正定对称矩阵$A^T A$做谱分解, 特征值λ, 特征向量$\boldsymbol{v}$

2）令$\sigma = \sqrt{\lambda}$, 得到奇异值

3）令$\boldsymbol{u} = \frac{A\boldsymbol{v}}{\sigma}$

4）将$\boldsymbol{u}$扩充为$\mathbb{R}^m$的一组标准正交基

2. 简化奇异值分解:

$A \in M_{m\times n}(\mathbb{R})$秩为$r$. 由奇异值分解

$$A = U\Sigma V^T = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_m]\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}\begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_n^T \end{bmatrix}$$

$$= \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^T + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^T + \cdots + \sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^T.$$

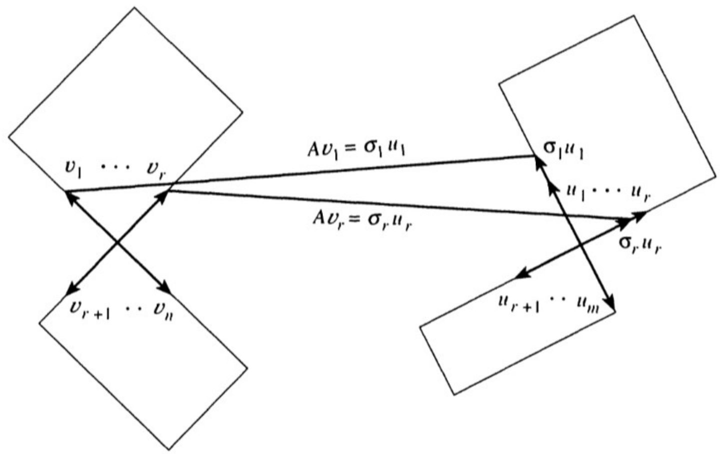
令$U_r = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_r], \Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}, V_r = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_r]$

$A = U_r\Sigma_r V_r^T$ 称为A的**简化奇异值分解**

3. 奇异值分解的几何的解释

1）映射复合的角度

2）空间分解的角度



4. 广义逆与最小二乘法

广义逆的定义：$A \in M_{m\times n}(\mathbb{R}) \rightsquigarrow A^{+} \in M_{n\times m}(\mathbb{R})$

线性映射角度 $A\colon \mathbb{R}^n \to \mathbb{R}^m \rightsquigarrow A^{+}\colon \mathbb{R}^m \to \mathbb{R}^n$

设$\Sigma = \begin{bmatrix} \Sigma_r & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix} \in M_{m\times n}(\mathbb{R})$, 其中$\Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \in M_r(\mathbb{R})$

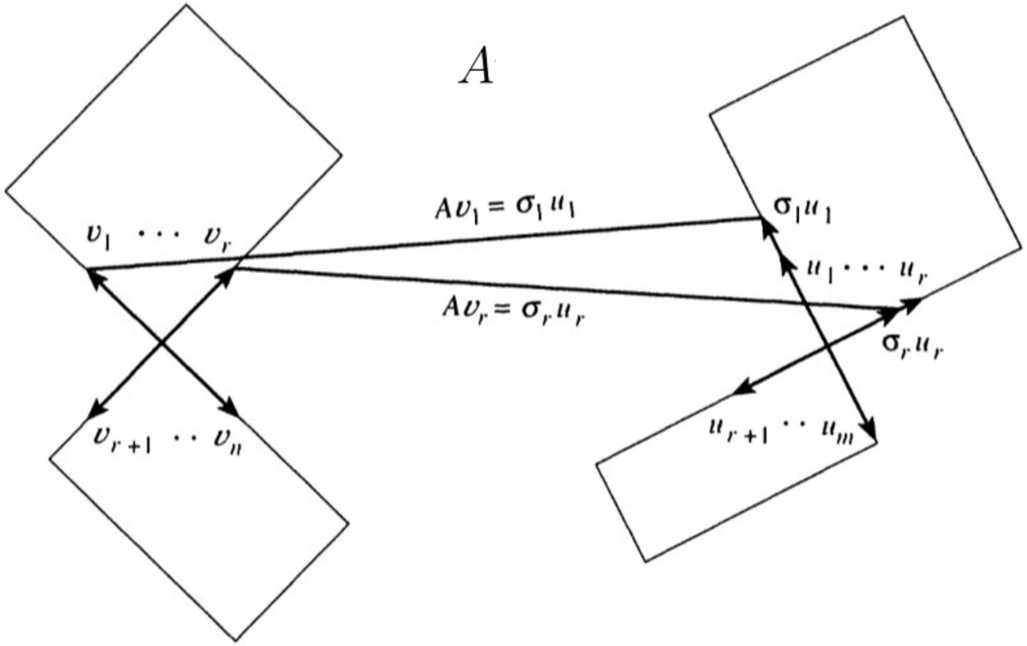
令$\Sigma^{+} := \begin{bmatrix} \Sigma_r^{+} & O_{r\times(m-r)} \\ O_{(n-r)\times r} & O_{(n-r)\times(m-r)} \end{bmatrix} \in M_{n\times m}(\mathbb{R})$, 其中$\Sigma_r^{+} = \begin{bmatrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_r^{-1} \end{bmatrix} \in M_r(\mathbb{R})$

$A \in M_{m\times n}(\mathbb{R})$，有奇异值分解$A = U\Sigma V^T$。定义$A$的广义逆：

$A^{+} = V\Sigma^{+}U^T \in M_{n\times m}(\mathbb{R})$.

于是对于可逆方阵，广义逆就是矩阵的逆。

广义逆与正交投影:



$$A^+(\sigma_1 \boldsymbol{u}_1) = \boldsymbol{v}_1, \dots, A^+(\sigma_r \boldsymbol{u}_r) = \boldsymbol{v}_r; A^+(\boldsymbol{u}_{r+1}) = \cdots = A^+(\boldsymbol{u}_m) = \boldsymbol{0}.$$

问题：A^+A在$\mathcal{R}(A^T)$和$\mathcal{N}(A)$上的作用是什么?

AA^+在$\mathcal{R}(A)$和$\mathcal{N}(A^T)$上的作用是什么?

命题:

(1) A^+A为在A的行空间上的正交投影矩阵。

(2) $I_n - A^+A$为在A的零空间上的正交投影矩阵。

(3) AA^+为在A的列空间上的正交投影矩阵。

(4) $I_m - AA^+$为在A的左零空间上的正交投影矩阵。

小结:

计算在A的列空间上的正交投影矩阵

(1) 列满秩矩阵: $A\left(A^T A\right)^{-1} A^T$

(2) 一般矩阵:

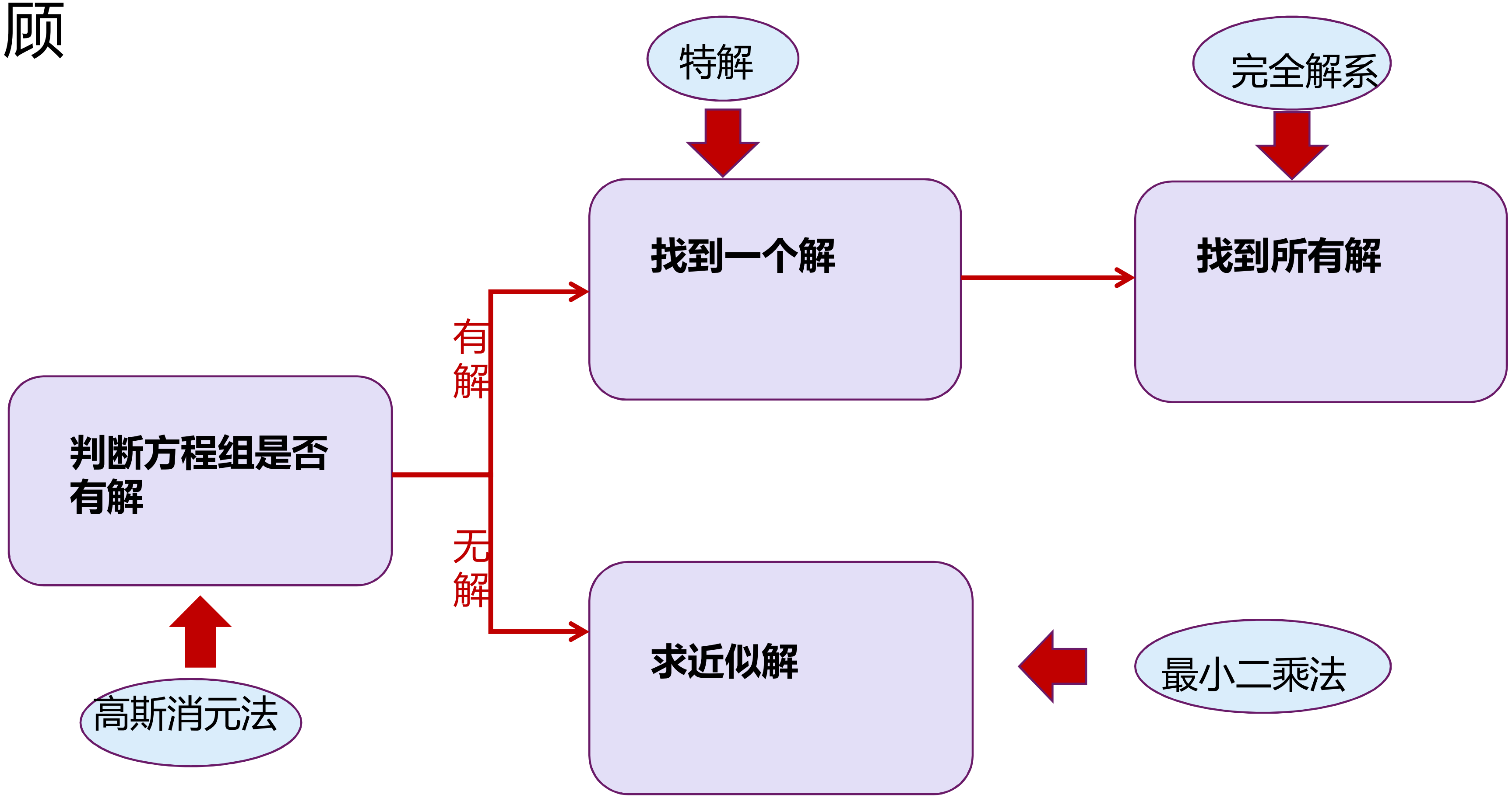
a. 高斯消元找主列, 利用主列构成的列满秩矩阵计算

b. AA^+, 需要计算奇异值分解

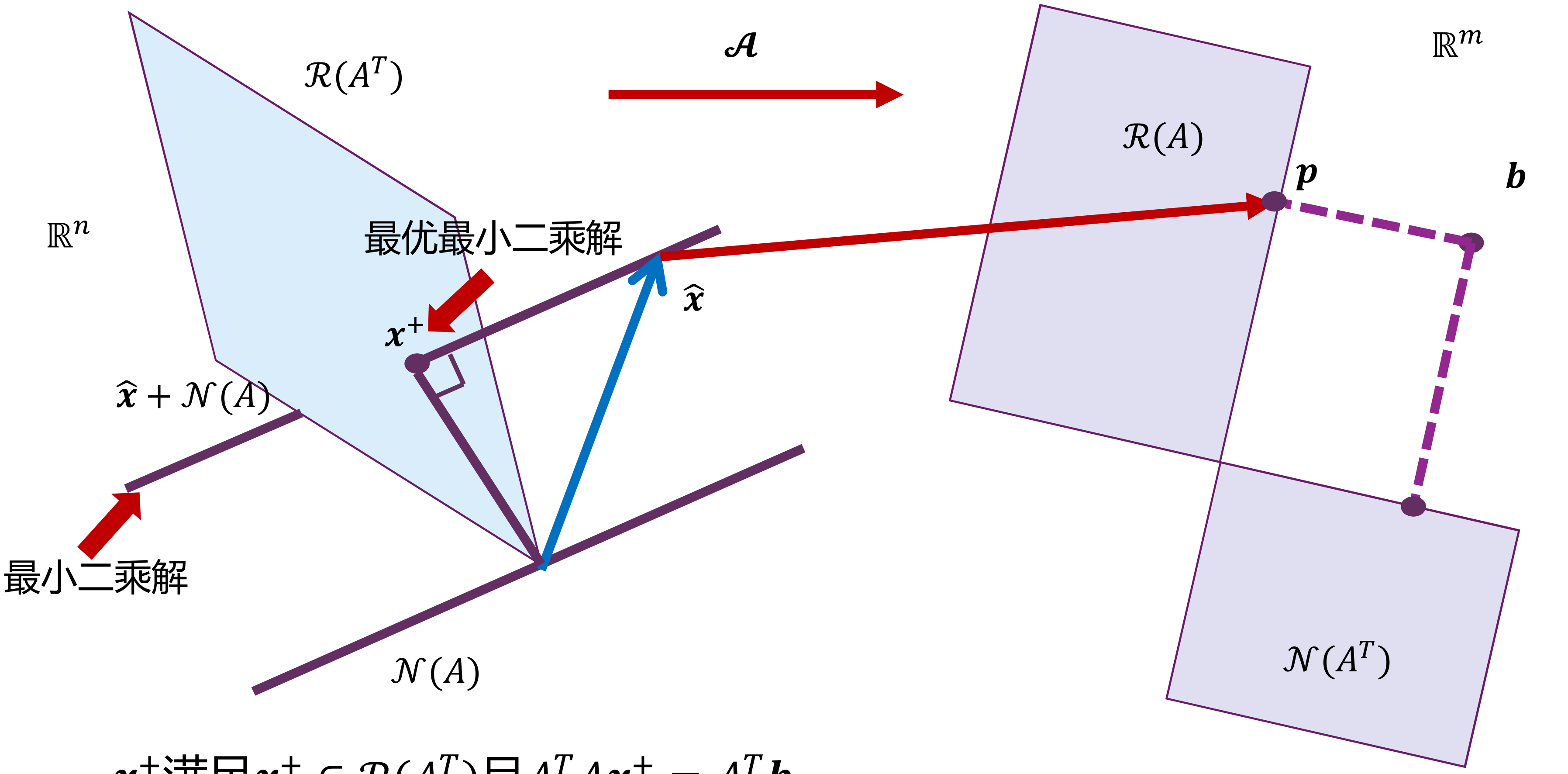
广义逆与最优最小二乘解:

回顾

解$A\boldsymbol{x} = \boldsymbol{b}$



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



$\boldsymbol{x}^+$满足$\boldsymbol{x}^+ \in \mathcal{R}(A^T)$且$A^T A \boldsymbol{x}^+ = A^T \boldsymbol{b}$

广义逆与最优最小二乘解:

$\boldsymbol{x}^+ = A^+\boldsymbol{b}$为方程$A\boldsymbol{x} = \boldsymbol{b}$的最优最小二乘解。

证明:

只需说明(1) $\boldsymbol{x}^+ = A^+\boldsymbol{b}$满足$A^T A\boldsymbol{x}^+ = A^T\boldsymbol{b}$ (2)$\boldsymbol{x}^+ \in \mathcal{R}(A^T)$

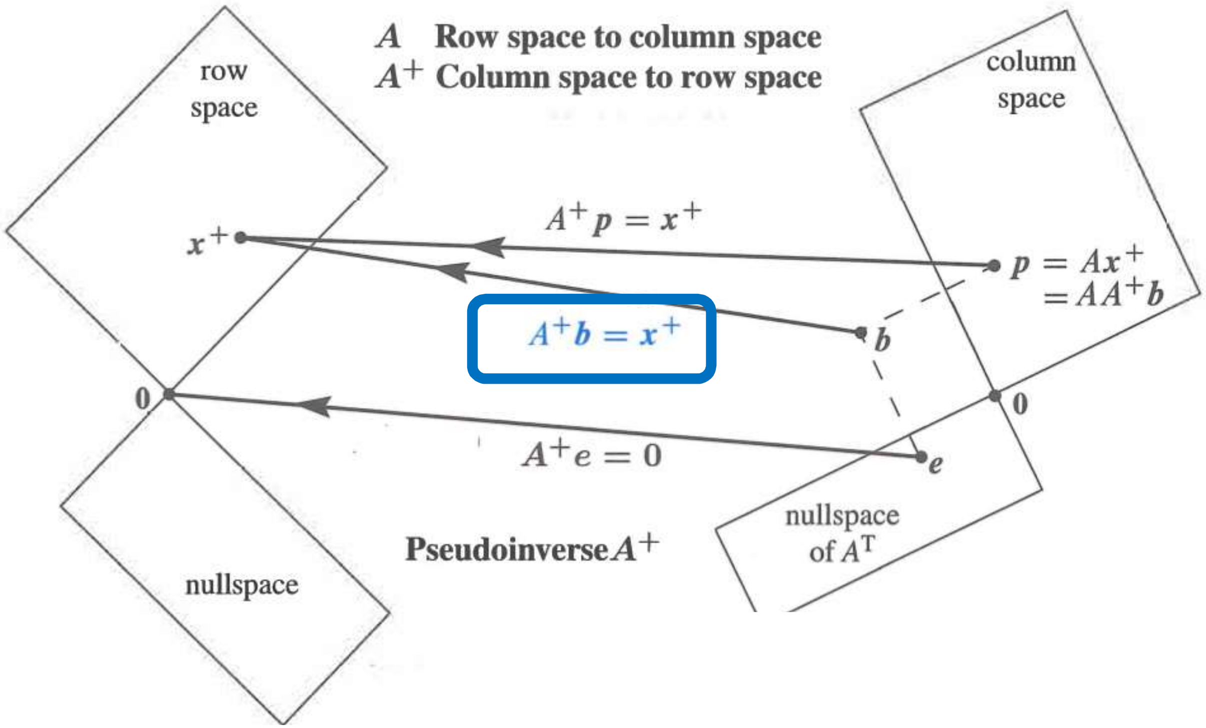
(1) 首先说明$A^+\boldsymbol{b}$为最小二乘解，即$A^+\boldsymbol{b}$满足$A(A^+\boldsymbol{b}) = \boldsymbol{p}$

$A(A^+\boldsymbol{b}) = (AA^+)\boldsymbol{b}$

AA^+为在$\mathcal{R}(A)$上的正交投影矩阵，因此，$(AA^+)\boldsymbol{b} = \boldsymbol{p}$

π

(2) $\boldsymbol{x}^+ = A^+\boldsymbol{b} = A^+(\boldsymbol{p}+\boldsymbol{e}) = A^+\boldsymbol{p} + A^+\boldsymbol{e} = A^+\boldsymbol{p} \in \mathcal{R}(A^T)$.



图片来源: Gilbert Strang, Introduction to linear algebra, 5th edition

(d) 矩阵的谱范数（考试内容）与低秩逼近（非考试内容）

矩阵的低秩逼近在实际中具有广泛应用

回顾在奇异值分解中，$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$

$$A = U\Sigma V^T = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_m] \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_n^T \end{bmatrix}$$

$$= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^T.$$

对$k \leq r$，定义$A_k = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T$

令$U_k = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_k], \Sigma_k = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix}, V_k = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_k].$

$$A_k = U_k \Sigma_k V_k^T$$

容易看出： $A_k = U_k\Sigma_k V_k^T = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_m]\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & \\ & & & \\ & & & \end{bmatrix}\begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_n^T \end{bmatrix}$

由于U, V可逆，$rank(A_k) = k.$

A_k称为矩阵A的秩为k的逼近。

Eckart-Young-Mirsky定理：A_k是所有秩小于等于k的$m \times n$阶矩阵中“距离”A最近的矩阵。

下面我们给出距离的定义

回顾：实数(或复数)有了绝对值(或模)，我们定义两个数a, b之间的距离为$|a - b|$;

矩阵范数与矩阵距离:

一个函数$\|-\|: M_{m\times n}(\mathbb{R}) \to \mathbb{R}_{\geq 0}$称为是一个范数(norm)，如果$\|-\|$满足

(1) $\|A+B\| \leq \|A\| + \|B\|$

(2) $\|cA\| = |c|\|A\|, c \in \mathbb{R}$,

(3) $\|A\| = 0$当且仅当$A = O_{m\times n}$.

固定了一个矩阵范数，可以定义两个矩阵的距离：

对$A, B \in M_{m\times n}(\mathbb{R})$，定义$d(A,B) = \|A-B\|$.

根据范数的性质我们容易得到：

(1) $d(A,B) = 0$当且仅当$A = B$;

(2) $d(A,B) = d(B,A)$;

(3) $d(A,C) \leq d(A,B) + d(B,C)$（三角不等式）

谱范数的定义:

定义谱范数$\|\cdot\|: M_{m\times n}(\mathbb{R}) \to \mathbb{R}_{\geq 0}$,

$$\|A\| = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \max_{\boldsymbol{x},\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|,$$

这里$\|A\boldsymbol{x}\|, \|\boldsymbol{x}\|$分别为向量$A\boldsymbol{x}, \boldsymbol{x}$的长度。

当$n = 1$时，A是列向量，$\|A\|$就是向量的长度。

命题：谱范数是个范数，即

(1) $\|A+B\| \leq \|A\|+\|B\|$.

(2) $\|cA\| = |c|\|A\|$,

(3) $\|A\| = 0$当且仅当$A = O_{m\times n}$,

(4) $\|AB\| \leq \|A\|\|B\|$. 矩阵乘法在谱范数下是连续的

证明:

(1) $\frac{\|(A+B)\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \frac{\|A\boldsymbol{x}+B\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \frac{\|A\boldsymbol{x}\|+\|B\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} + \frac{\|B\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$,

$$\max\frac{\|(A+B)\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \max\left(\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}+\frac{\|B\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\right) \leq \max\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}+\max\frac{\|B\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$$

因此，$\|A+B\| \leq \|A\|+\|B\|$

(2) $\|cA\| = |c|\|A\|$

证明：$\frac{\|(cA)\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \frac{\|cA\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = |c|\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$,

因此，$\|cA\| = |c|\|A\|$

(3) $\|A\| = 0$当且仅当$A = O_{m\times n}$,

证明：如果$A = O_{mn}$，则$\|A\| = 0$。

反之，如果$\|A\| = 0$，那么$\|A\boldsymbol{x}\| = 0, \forall \boldsymbol{x} \neq \mathbf{0}$.

因此，$A\boldsymbol{x} = 0, \forall \boldsymbol{x} \neq \mathbf{0}$, 于是$A = O_{mn}$.

(4) $\|AB\| \leq \|A\|\|B\|$

证明：

$\|AB\| = \max\frac{\|(AB)\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \max\left(\frac{\|AB\boldsymbol{x}\|}{\|B\boldsymbol{x}\|}\frac{\|B\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\right) \leq \max\left(\frac{\|AB\boldsymbol{x}\|}{\|B\boldsymbol{x}\|}\right)\max\left(\frac{\|B\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\right) \leq$
$\|A\|\|B\|$.

Eckart-Young-Mirsky定理:

令A_k为A的秩k 逼近，则对任意秩$\leq k$ 的矩阵$B \in M_{m\times n}(\mathbb{R})$

$d(A, A_k) = \|A - A_k\| \leq \|A - B\| = d(A, B)$.

为了证明这个定理, 我们建立矩阵谱范数与奇异值的关系

工具：Rayleigh商

谱范数与奇异值:

命题：$\|A\| = \sigma_1$是A的最大奇异值。

回顾Rayleigh商：给定实对称矩阵S，特征值为$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$,

相应的特征向量为$\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$，即$S = Q\Lambda Q^T$，则$\lambda_1 = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}^T S\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}$

命题的证明:

令$S = A^T A = V\Lambda V^T, \Lambda = Diag\left(\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0\right), V = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_n]$.

由Rayleigh商，$\sigma_1^2 = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}^T A^T A\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|A\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2}$.

于是$\sigma_1 = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \|A\|$.

推论:

$$\|A-A_k\|=\sigma_{k+1}.$$

证明:

$$A-A_k=\sigma_{k+1}\boldsymbol{u}_{k+1}\boldsymbol{v}_{k+1}^T+\cdots+\sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^T$$

是$A-A_k$的简化奇异值分解,

其最大奇异值为σ_{k+1}.

Eckart-Young-Mirsky定理:

对任意秩$\leq k$的$m \times n$阶矩阵B，$\|A - A_k\| \leq \|A - B\|$

证明:

设$B \in M_{m\times n}(\mathbb{R})$，$rank(B) \leq k$.

首先证明存在非零向量$\boldsymbol{x}_0 \in \mathcal{N}(B) \cap \mathcal{R}\left(A_{k+1}^T\right)$.

由于$rank(B) \leq k, dim\mathcal{N}(B) \geq n - k$. 因此,
$dim\mathcal{N}(B) + dim\mathcal{R}\left(A_{k+1}^T\right) \geq n - k + k + 1 = n + 1$.

由维数公式

$dim\mathcal{N}(B) \cap \mathcal{R}\left(A_{k+1}^T\right) = dim\mathcal{N}(B) + dim\mathcal{R}\left(A_{k+1}^T\right) - \dim(\mathcal{N}(B) + \mathcal{R}\left(A_{k+1}^T\right))$

$\geq n + 1 - n = 1$.

因此，存在非零向量$\boldsymbol{x}_0 \in \mathcal{N}(B) \cap \mathcal{R}\left(A_{k+1}^T\right)$.

由于$\boldsymbol{v}_1, \dots, \boldsymbol{v}_{k+1}$构成$\mathcal{R}\left(A_{k+1}^T\right)$的一组标准正交基,

存在$c_1, \dots, c_{k+1} \in \mathbb{R}$使得$\boldsymbol{x}_0 = c_1\boldsymbol{v}_1 + \cdots + c_{k+1}\boldsymbol{v}_{k+1}$.

$$\boldsymbol{x}_0 \in \mathcal{N}(B)$$

$$\|A - B\| \geq \frac{\|(A - B)\boldsymbol{x}_0\|}{\|\boldsymbol{x}_0\|} = \frac{\|A\boldsymbol{x}_0\|}{\|\boldsymbol{x}_0\|}$$

$$\boxed{\begin{aligned} A &= U\Sigma V^T \\ AV &= U\Sigma \end{aligned}}$$

$$= \frac{\|A(c_1\boldsymbol{v}_1 + \cdots + c_{k+1}\boldsymbol{v}_{k+1})\|}{\|c_1\boldsymbol{v}_1 + \cdots + c_{k+1}\boldsymbol{v}_{k+1}\|}$$

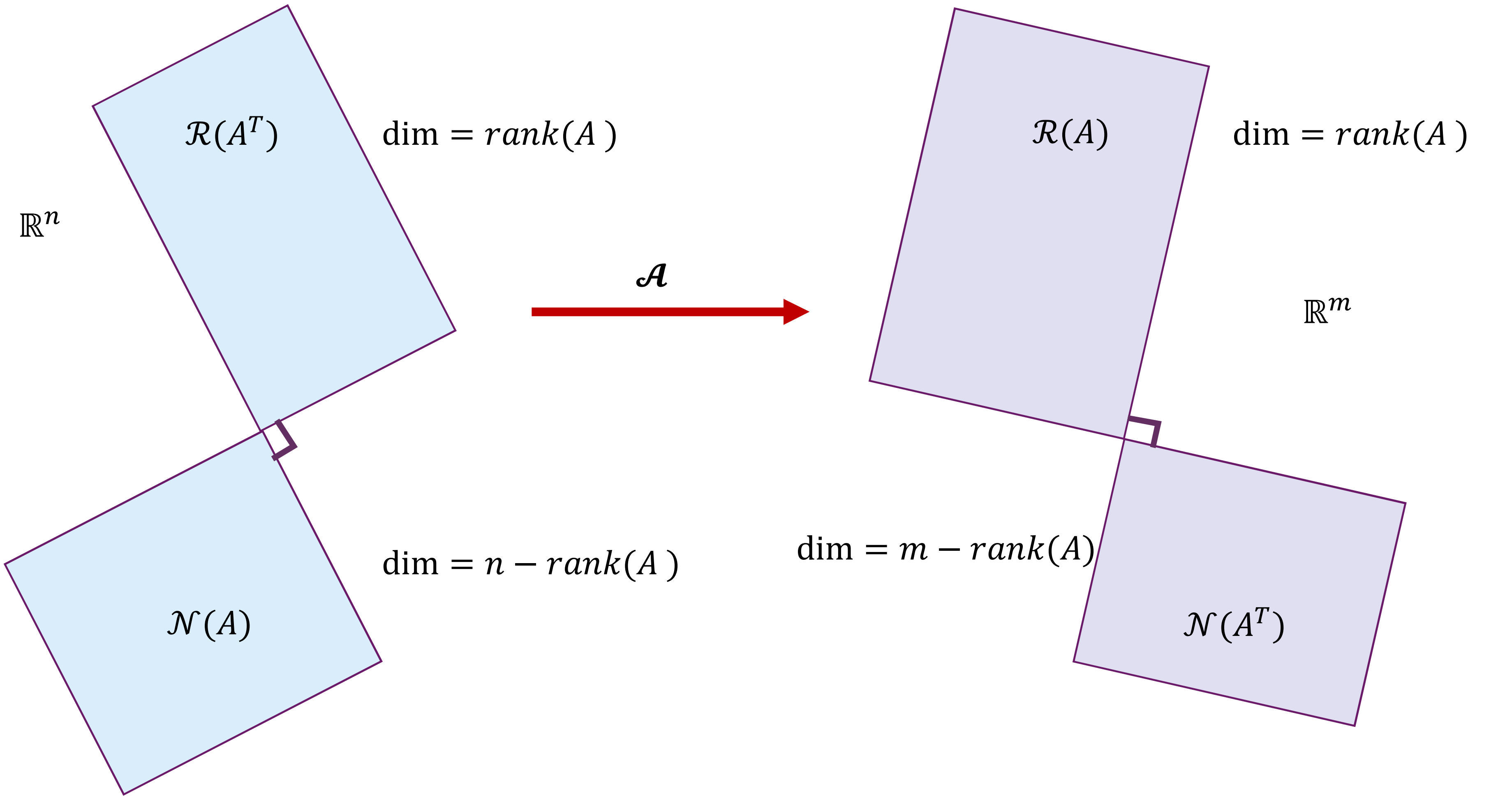
$$= \frac{\|\sigma_1 c_1\boldsymbol{u}_1 + \cdots + \sigma_{k+1}c_{k+1}\boldsymbol{u}_{k+1}\|}{\|c_1\boldsymbol{v}_1 + \cdots + c_{k+1}\boldsymbol{v}_{k+1}\|}$$

$$= \frac{\sqrt{|\sigma_1 c_1|^2 + \cdots + |\sigma_{k+1}c_{k+1}|^2}}{\sqrt{c_1^2 + \cdots + c_{k+1}^2}}$$

$$\geq \frac{\sqrt{|\sigma_{k+1} c_1|^2 + \cdots + |\sigma_{k+1}c_{k+1}|^2}}{\sqrt{c_1^2 + \cdots + c_{k+1}^2}}$$

$$= \sigma_{k+1} = \|A - A_k\|.$$

总复习I： 线性代数基本定理:



线性代数基本定理(v.2.0):

$A \in M_{m\times n}(\mathbb{R})$

(1) $rank(A) \le \min(m,n)$.

(2) 四个基本子空间的维数

行空间$\mathcal{R}(A^T)$: $dim\mathcal{R}(A^T) = rank(A)$

零空间$\mathcal{N}(A)$: $dim\mathcal{N}(A) = n - rank(A)$

列空间$\mathcal{R}(A)$: $dim\mathcal{R}(A) = rank(A)$

左零空间$\mathcal{N}(A^T)$: $dim\mathcal{N}(A^T) = m - rank(A)$

(3) 四个基本子空间的一组基

由$rref(A)$的非零行给出

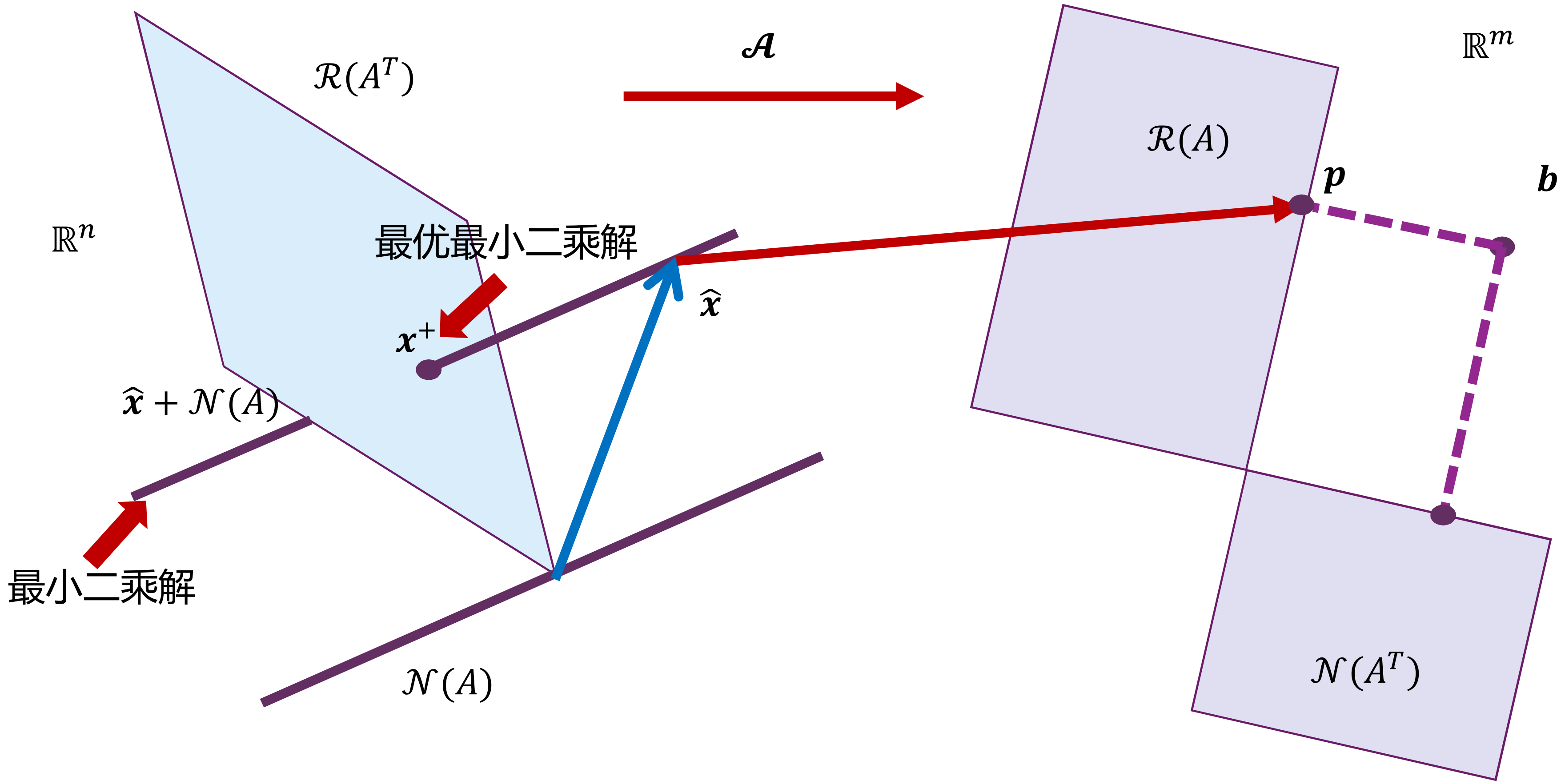
由规范基础解系给出

由A的主元列给出

由$EA = rref(A)$中E的后$m - rank(A)$行给出

(4) 空间分解： $\mathbb{R}^n = \mathcal{N}(A) + \mathcal{R}(A^T), \mathbb{R}^m = \mathcal{N}(A^T) + \mathcal{R}(A)$

(5) $\mathcal{N}(A) = \mathcal{R}(A^T)^{\perp}, \mathcal{N}(A^T) = \mathcal{R}(A)^{\perp}$



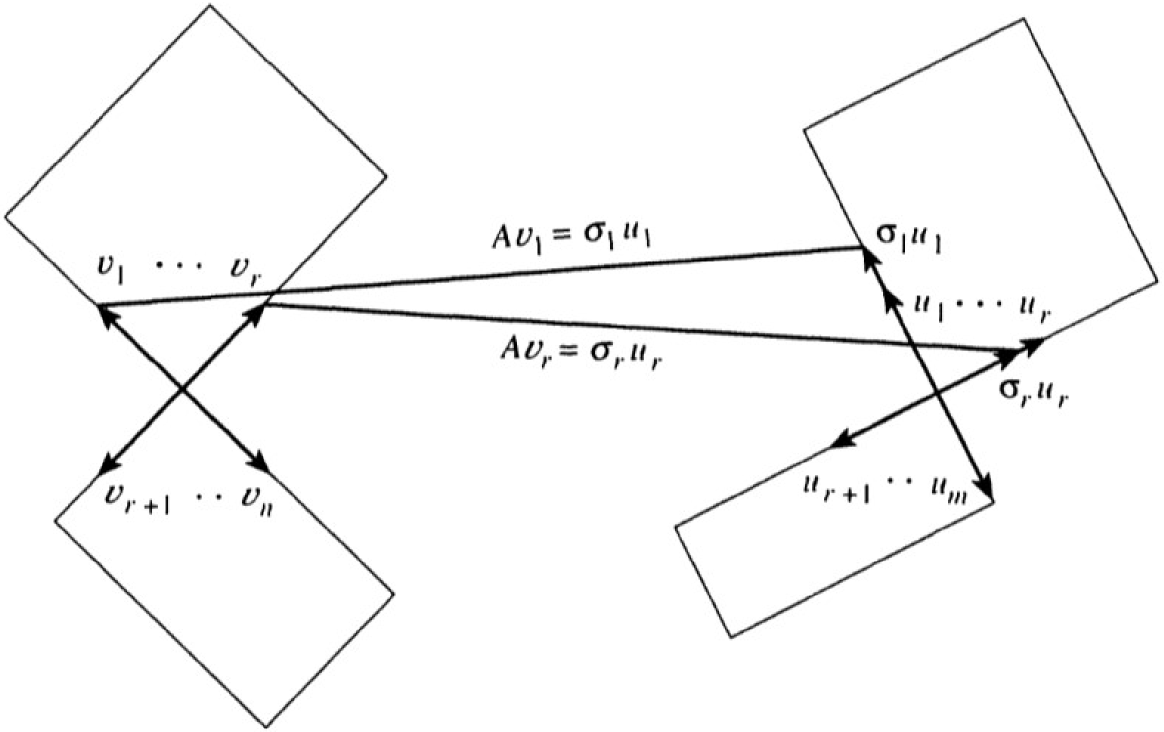
$$A^T A\boldsymbol{x} = A^T\boldsymbol{b} \Leftrightarrow A\boldsymbol{x} = \boldsymbol{p}$$

(6) 奇异值分解：$\mathcal{R}\left(A^T\right), \mathcal{N}(A), \mathcal{R}(A), \mathcal{N}\left(A^T\right)$存在标准正交基

$\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_r\}, \{\boldsymbol{v}_{r+1}, \dots, \boldsymbol{v}_n\}, \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_r\}, \{\boldsymbol{u}_{r+1}, \dots, \boldsymbol{u}_m\}$使得有$\sigma_i > 0, 1 \le i \le r$满足

$A\boldsymbol{v}_1 = \sigma_1 \boldsymbol{u}_1, A\boldsymbol{v}_2 = \sigma_2 \boldsymbol{u}_2, \dots, A\boldsymbol{v}_r = \sigma_r \boldsymbol{u}_r, A\boldsymbol{v}_{r+1} = \boldsymbol{0}, \dots, A\boldsymbol{v}_n = \boldsymbol{0}$

$$\boxed{A = U\Sigma V^T}$$



(7) 令$U=[\boldsymbol{u}_1,\dots,\boldsymbol{u}_m]$, $V=[\boldsymbol{v}_1,\dots,\boldsymbol{v}_n]$.

$A^+=V\Sigma^+U^T\in M_{n\times m}(\mathbb{R})$为$A$的广义逆。

(8) $\boldsymbol{b}\in\mathbb{R}^m$在$\mathcal{R}(A)$上的正交投影：

如果A为列满秩矩阵，正交投影$\boldsymbol{p}=A\left(A^TA\right)^{-1}A^T\boldsymbol{b}$.

对于一般矩阵A，$\boldsymbol{p}=AA^+\boldsymbol{b}$.

(9) $A\boldsymbol{x}=\boldsymbol{b}$的最小二乘解与最优最小二乘解：

当A为列满秩矩阵时，$\left(A^TA\right)^{-1}A^T\boldsymbol{b}$是最小二乘解也是最优最小二乘解。

对于一般矩阵A，最优最小二乘解$\boldsymbol{x}^+=A^+\boldsymbol{b}$.

集合$\boldsymbol{x}^++\mathcal{N}(A)$是所有最小二乘解构成的集合。