奇异值分解

颜文斌

清华大学

内容提要

- **奇异值分解**
- 主成分分析
- 奇异值分解和几何

矩阵对角化

- 矩阵对角化有很多应用：简化计算、解方程……
- 不是所有矩阵都可以对角化的
  - 可对角化矩阵例：对称矩阵
- 特征值和特征向量只使用于方阵
- 对于一般的mxn矩阵A，有没有类似的操作？
  - 回忆方程$Ax = b$不一定有解，但是$A^T Ax = A^T b$有解
  - 考虑方阵$A^T A$和AA^T，他们都是**半正定**矩阵，所以可以对角化而且特征值大于等于0

对角化A^TA和AA^T

- mxn矩阵A， A^TA和AA^T都是**半正定**矩阵，
  - 证明：$\forall \boldsymbol{x}, \boldsymbol{x}^T A^T A \boldsymbol{x} = \|A\boldsymbol{x}\|^2 \geq 0$，所以$A^TA$是半正定的。$AA^T$同理
- A^TA和AA^T都可以对角化
  - $A^TA = V\Lambda_1 V^T$， $AA^T = U\Lambda_2 U^T$
  - $V^TA^TAV = (AV)^T A(V) = \Lambda_1$
  - $U^TAA^TU = (U^T)A(A^TU) = \Lambda_2$
- 猜测：找到正交矩阵U和V使得mxn矩阵U^TAV可以写成Σ？其中Σ是某种意义上的“对角”矩阵

奇异值（singular value）

- mxn的实矩阵A， $A^T A$是一个nxn的对称矩阵，$\{\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n\}$是由$A^T A$的特征向量构成的$\mathbb{R}^n$中的正交归一基，对应的实特征值为$\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$，假设$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$
- $A^T A$的所有特征值都非负：$\|A\boldsymbol{q}_i\|^2 = \boldsymbol{q}_i^T A^T A\boldsymbol{q}_i = \lambda_i \geq 0$
- 矩阵A的奇异值定义为$A^T A$的特征值的平方根：$\sigma_i = \sqrt{\lambda_i}$
  - $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$
  - $\sigma_i = \|A\boldsymbol{q}_i\|$

例

- 求以下矩阵的奇异值

$$A=\begin{bmatrix}4 & 11 & 14\\ 8 & 7 & -2\end{bmatrix} \qquad A^TA=\begin{bmatrix}4 & 8\\ 11 & 7\\ 14 & -2\end{bmatrix}\begin{bmatrix}4 & 11 & 14\\ 8 & 7 & -2\end{bmatrix}=\begin{bmatrix}80 & 100 & 40\\ 100 & 170 & 140\\ 40 & 140 & 200\end{bmatrix}$$

- 特征方程：$\det\begin{bmatrix}80-\lambda & 100 & 40\\ 100 & 170-\lambda & 140\\ 40 & 140 & 200-\lambda\end{bmatrix}$

$$=-\lambda(\lambda-90)(\lambda-360)$$

- 特征值：360，90，0。奇异值：$6\sqrt{10}$，$3\sqrt{10}$，0

A的秩和 A^TA的秩

- **定理**： mxn矩阵A，$\operatorname{rank}(A^TA) = \operatorname{rank}(A)$
- 证明：
  - $\operatorname{rank}(A^TA) = n - \dim N(A^TA)$，$\operatorname{rank}(A) = n - \dim N(A)$，所以我们只需证明$\dim N(A^TA) = \dim N(A)$
  - 如果$\boldsymbol{x} \in N(A)$，则$A\boldsymbol{x} = 0$，等号两边同时左乘A^T得到$A^TA\boldsymbol{x} = 0$，所以$\boldsymbol{x} \in N(A^TA)$
  - 如果$\boldsymbol{x} \in N(A^TA)$，则$A^TA\boldsymbol{x} = 0$，等号两边同时左乘$\boldsymbol{x}^T$得到$\boldsymbol{x}^TA^TA\boldsymbol{x} = 0$，所以$\|A\boldsymbol{x}\| = 0$，所以$A\boldsymbol{x} = 0$，所以$\boldsymbol{x} \in N(A)$
  - $N(A)$和$N(A^TA)$存在一一映射，所以$\dim N(A^TA) = \dim N(A)$，
- 推论：$\operatorname{rank}(A) = \operatorname{rank}\left(A^{\mathrm{T}}\right) = \operatorname{rank}(AA^T) = \operatorname{rank}(A^TA)$

非零奇异值的数量

- **定理**：mxn的实矩阵A的非零奇异值$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ 的个数r等于A的秩rank(A)
- 证明：
  - 设$\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$为$\mathbb{R}^n$中可以把$A^TA$对角化的正交归一基，$\{\lambda_1, \cdots, \lambda_n\}$为对应的特征值，则$\{A\boldsymbol{v}_1, \cdots, A\boldsymbol{v}_n\}$是一个**正交向量集合**，也就是说$\forall\ i \neq j$，$(A\boldsymbol{v}_i)^T(A\boldsymbol{v}_j) = \boldsymbol{v}_i^T A^T A\boldsymbol{v}_j = \boldsymbol{v}_i^T(\lambda_j \boldsymbol{v}_j) = 0$
  - 假设$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$是所有正特征值，$\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$是所有0特征值，则$\{A\boldsymbol{v}_{r+1}, \cdots, A\boldsymbol{v}_n\}$都是零向量
  - 因为$\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$为$\mathbb{R}^n$中可以把$A^TA$对角化的正交归一基，所以对于任意的$\boldsymbol{x} \in \mathbb{R}^n$，$\boldsymbol{x}$可以写成$\boldsymbol{x} = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n$。

非零奇异值的数量（续）

- **定理**：mxn的实矩阵A的非零奇异值$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ 的个数r等于A的秩rank(A)
- 证明（续）：
  - 因为$\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$为$\mathbb{R}^n$中可以把$A^TA$对角化的正交归一基，所以对于任意的$\boldsymbol{x} \in \mathbb{R}^n$， $\boldsymbol{x}$可以写成$\boldsymbol{x} = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n$。
  - 我们需要证明A的秩为r，只需要证明列空间$C(A)$的维度为r，需要找到列空间的一组基
  - 因为$C(A)$是由A所有列的线性组合得到的，所以$\forall\ \mathbf{y} \in C(A)$， $\boldsymbol{y} = A\boldsymbol{x} = c_1A\boldsymbol{v}_1 + \cdots + c_nA\boldsymbol{v}_n = c_1A\boldsymbol{v}_1 + \cdots + c_rA\boldsymbol{v}_r$，也就是说$C(A)$中的任何向量都可以写成$\{A\boldsymbol{v}_1, \cdots, A\boldsymbol{v}_r\}$的线性组合
  - 所以$\{A\boldsymbol{v}_1, \cdots, A\boldsymbol{v}_r\}$是$C(A)$的一组正交基，所以$r = \mathrm{rank}(A)$

奇异值分解：陈述

- 广义“对角”矩阵：mxn矩阵Σ

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

  - D是一个rxr的对角矩阵，Σ所有大于r的行和列都是0

- **定理（奇异值分解）**：mxn矩阵A的秩为r。则存在一个形状如上的mxn矩阵Σ且D的对角元是A的前r个（非零）的奇异值，mxm的正交矩阵U和nxn的正交矩阵V，而且以上矩阵满足关系

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

- 可以用两个正交矩阵把任意矩阵A变成简单形式
- 考虑rank(A)=1，矩阵可以写成两个向量的乘积（压缩）

奇异值分解：证明

- 证明：直接构造相应的矩阵
  - A^TA是对称矩阵，由谱定理可知存在一组$\mathbb{R}^n$中的正交归一基，将A^TA对角化
  - 设$\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n\}$为$\mathbb{R}^n$中可以把$A^TA$对角化的正交归一基，$\{\lambda_1,\cdots,\lambda_n\}$为对应的特征值。由之前定理$\{\lambda_1,\cdots,\lambda_r\}$是正特征值，$\{\lambda_{r+1},\cdots,\lambda_n\}$是零特征值
  - 因为$\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n\}$之间是正交的，则$\forall\, i\neq j$，$(A\boldsymbol{v}_i)^T\left(A\boldsymbol{v}_j\right)=\boldsymbol{v}_i^TA^TA\boldsymbol{v}_j=\boldsymbol{v}_i^T\left(\lambda_j\boldsymbol{v}_j\right)=0$，所以$\{A\boldsymbol{v}_1,\cdots,A\boldsymbol{v}_r\}$之间也是正交的，$\{A\boldsymbol{v}_{r+1},\cdots,A\boldsymbol{v}_n\}$是零向量
  - 令$u_i=\frac{1}{\|Av_i\|}Av_i=\frac{1}{\sigma_i}Av_i$，$i=1,\cdots,r$，则$\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r\}$是$C(A)$的一组正交归一基。$\sigma_i=\sqrt{\lambda_i}$

奇异值分解：证明（续）

- 证明：直接构造相应的矩阵
  - 令$u_i = \frac{1}{\|Av_i\|}Av_i = \frac{1}{\sigma_i}Av_i$，$i = 1, \cdots, r$，则$\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$是$C(A)$的一组正交归一基。$\sigma_i = \sqrt{\lambda_i}$
  - 再设$\{\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m\}$是$N(A^T)$ 中的一组正交归一基。因为$N(A^T)$是$C(A)$的正交补，则$\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m\}$是$\mathbb{R}^m$的一组正交归一基
  - 设矩阵$U = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m)$和$V = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$，$U$和$V$都是正交的
  - $AV = (A\boldsymbol{v}_1, \cdots, A\boldsymbol{v}_n) = (\sigma_1\boldsymbol{u}_1, \cdots, \sigma_r\boldsymbol{u}_r, 0, \ldots, 0) = U\Sigma$
  - 所以$A = U\Sigma V^T$，我们得到了A的奇异值分解$A = U\Sigma V^T = \sum_{i=1}^r \sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^T$

奇异值分解：例

- 矩阵$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$的奇异值分解
- $A^T A = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$
- $A^T A$的特征值：360，90，0。A 的奇异值：$6\sqrt{10}$，$3\sqrt{10}$，0
- $V = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}$

奇异值分解——例

- 矩阵$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$的奇异值分解
- A 的奇异值：$6\sqrt{10}$，$3\sqrt{10}$，0
- $V = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}$，$AV = \begin{bmatrix} 4 & 11 \\ 8 & 7 \end{bmatrix}$
- $\boldsymbol{u}_1 = \frac{1}{\sigma_1} A\boldsymbol{v}_1 = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$，$\boldsymbol{u}_2 = \frac{1}{\sigma_2} A\boldsymbol{v}_2 = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$
- $U = (\boldsymbol{u}_1 \quad \boldsymbol{u}_2)$

奇异值分解——例

- 矩阵$A=\begin{bmatrix}4 & 11 & 14\\ 8 & 7 & -2\end{bmatrix}$的奇异值分解 $A=U\Sigma V^T$
- A 的奇异值：$6\sqrt{10}$，$3\sqrt{10}$，0，$\Sigma=\begin{bmatrix}6\sqrt{10} & 0 & 0\\ 0 & 3\sqrt{10} & 0\end{bmatrix}$
- $V=\begin{bmatrix}1/3 & -2/3 & 2/3\\ 2/3 & -1/3 & -2/3\\ 2/3 & 2/3 & 1/3\end{bmatrix}$，$U=\begin{bmatrix}3/\sqrt{10} & 1/\sqrt{10}\\ 1/\sqrt{10} & -3/\sqrt{10}\end{bmatrix}$
- $A=\begin{bmatrix}3/\sqrt{10} & 1/\sqrt{10}\\ 1/\sqrt{10} & -3/\sqrt{10}\end{bmatrix}\begin{bmatrix}6\sqrt{10} & 0 & 0\\ 0 & 3\sqrt{10} & 0\end{bmatrix}\begin{bmatrix}1/3 & 2/3 & 2/3\\ -2/3 & -1/3 & 2/3\\ 2/3 & -2/3 & 1/3\end{bmatrix}$

A和A^T的奇异值分解

- 练习：矩阵$A^T = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix}$的奇异值分解
- 思考：已知A的奇异值分解，求矩阵A^T的奇异值分解？
- 结论：若A的奇异值分解为$U\Sigma V^T$， A^T的奇异值分解为$A^T = V\Sigma^T U^T$
- **推论**：$\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m\}$是矩阵$AA^T$的特征向量
- 证明：
  - $AA^T = U\Sigma V^T V\Sigma^{\mathrm{T}} U^T = U\Sigma\Sigma^T U^T$，所以$AA^T U = U\Sigma\Sigma^T$
- V和U分别是将$A^T A$和AA^T对角化的正交矩阵

A和A^T的奇异值分解

- **定理**：A^TA和AA^T的非零特征值都相同
- 证明：可以直接用A的奇异值分解证明，我们这里直接证明
  - 假设x_i是A^TA的特征值为$\lambda_i \neq 0$的特征向量，$(\lambda_i I - A^TA)x_i = 0$
  - 等号两端同时左乘：$0 = A(\lambda_i I - A^TA)x_i = (\lambda_i A - AA^TA)x_i = (\lambda_i I - AA^T)(Ax_i)$。又因为$x^TA^TAx = \lambda_i x^Tx > 0$，所以$Ax_i \neq 0$。所以$Ax_i$ 是 AA^T 的特征值为 λ_i 的特征向量
  - 同理，如果x_i是AA^T的特征值为$\lambda_i \neq 0$的特征向量， A^Tx_i是A^TA的特征值为λ_i的特征向量
  - 这样，我们就有A^TA和AA^T非零特征值和非零特征向量的一一对应

应用：四个子空间的正交归一基

- $A = U\Sigma V^T$
  - $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r\}$是$C(A^T)$的正交归一基，$V_r = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r)$
  - $\{\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_n\}$是$N(A)$的正交归一基， $V_{n-r} = (\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_n)$
  - $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$是$C(A)$的正交归一基，$U_r = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r)$
  - $\{\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m\}$是$N(A^T)$ 的正交归一基， $U_{n-r} = (\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m)$
- 证明？

应用：数据压缩

- 假设rank(A)<min(m,n)，则

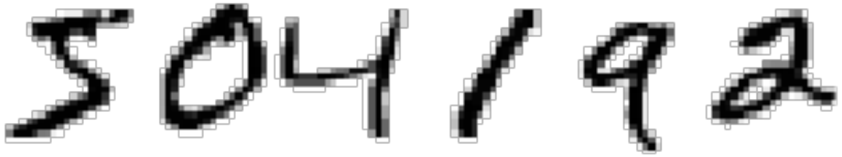
$$A=(U_r,U_{m-r})\begin{bmatrix}D & 0\\0 & 0\end{bmatrix}\begin{bmatrix}V_r^T\\V_{n-r}^T\end{bmatrix}=U_rDV_r^T=\sum_{i=1}^{r}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^T$$

- 可以用U_r,D,V_r这三个矩阵的r(m+1+n)个分量完全决定A（mn）
- **图像压缩**
    - 先考虑黑白图片，可以用一个mxn的矩阵描述，每个元素是该像素的灰度（0-255之间的整数，0是黑，255是白）
    - 如果r(m+1+n)<mn，我们可以只储存或者传输U_r,D,V_r（无损）。例如矩阵秩为1的时候我们只需要储存一个行向量和一个列向量
    - 甚至可以把很小的奇异值当成零忽略，进一步压缩图片（有损）

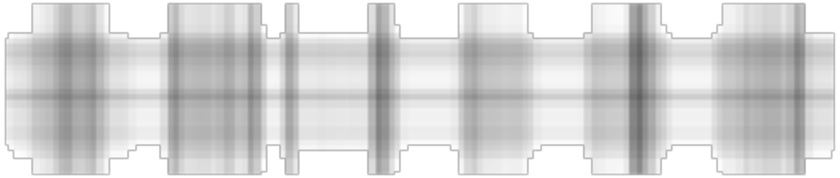
数据压缩的误差估计

- $A=\sum_{i=1}^{r}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^T$，我们需要估算取前k个奇异值得到的矩阵和真实矩阵之间的误差
  - 设$A(k)=\sum_{i=1}^{k}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^T$，则$\delta A=A-A(k)=\sum_{l=k+1}^{r}\sigma_l\boldsymbol{u}_l\boldsymbol{v}_l^T$
  - δA的ij分量的绝对值：
$$\left|\delta A_{ij}\right|=\left|\sum_{l=k+1}^{r}\sigma_l(\boldsymbol{u}_l)_i\left(\boldsymbol{v}_l^T\right)_j\right|\le\sum_{l=k+1}^{r}\sigma_l|(\boldsymbol{u}_l)_i|\left|\left(\boldsymbol{v}_l^T\right)_j\right|$$
  - $\boldsymbol{u}_l$，$\boldsymbol{v}_l$的长度都是1，所以$|(\boldsymbol{u}_l)_i|\le 1$，$\left|\left(\boldsymbol{v}_l^T\right)_j\right|\le 1$
  - 所以$\left|\delta A_{ij}\right|\le\sum_{l=k+1}^{r}\sigma_l$
- 结论：误差由忽略的奇异值控制，取得奇异值越多误差越小

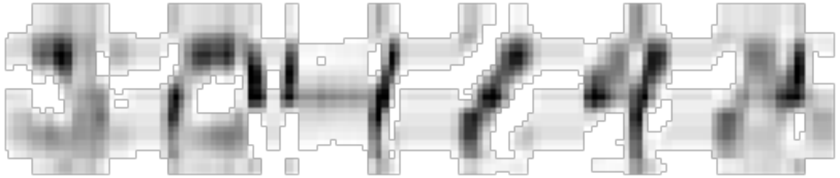
图像压缩例



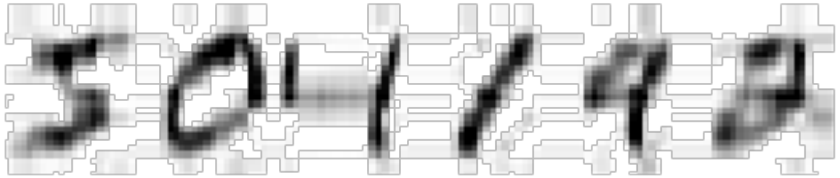
- 一个128x623=79744像素的黑白图片
  - 秩为24。奇异值分解后有效信息24*(128+623)=18024
  - 取前k个奇异值的近似



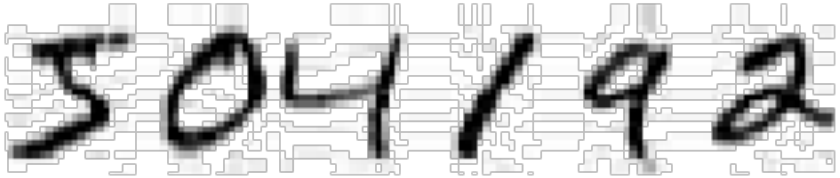
k=1



k=3



k=5

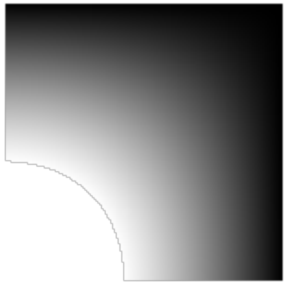


k=10

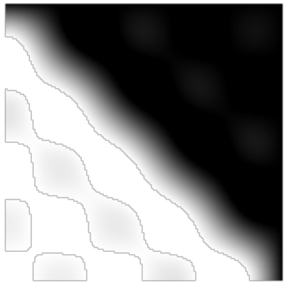
图像压缩例

- 考虑一个300x300的下三角矩阵，对角线和下面的元素全是255

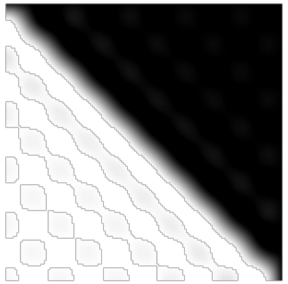
原图：300x300



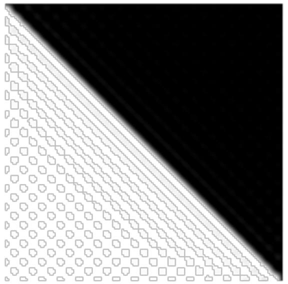
前1个奇异值 ~ 600



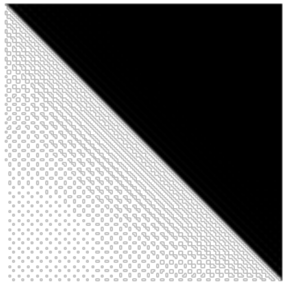
前5个 ~ 3000



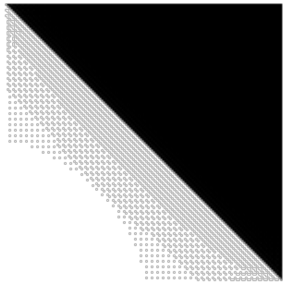
前10个 ~ 6000



前30个 ~ 18000



前60个 ~ 36000



前100个 ~ 60000

小结

1. The SVD factors A into $U\Sigma V^{\mathrm{T}}$, with r singular values $\sigma_1 \geq \ldots \geq \sigma_r > 0$.
2. The numbers $\sigma_1^2, \ldots, \sigma_r^2$ are the nonzero eigenvalues of AA^{T} and $A^{\mathrm{T}}A$.
3. The orthonormal columns of U and V are eigenvectors of AA^{T} and $A^{\mathrm{T}}A$.
4. Those columns hold orthonormal bases for the four fundamental subspaces of A.
5. Those bases diagonalize the matrix: $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$ for $i \leq r$. This is $\boldsymbol{AV} = \boldsymbol{U\Sigma}$.
6. $A = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}$ and σ_1 is the maximum of the ratio $||A\boldsymbol{x}|| \,/\, ||\boldsymbol{x}||$.

内容提要

- 奇异值分解
- **主成分分析**
- 奇异值分解和几何

统计知识

- 假设一组数据来源于n个样本$\{\mu_1, \mu_2, \cdots, \mu_n\}$
- 例：所有同学的期中考试成绩
- **平均值（mean）**：$\bar{\mu} = \frac{\sum_i \mu_i}{n}$
- **标准差（standard deviation）**：$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(\mu_i - \bar{\mu})^2}{n-1}}$
  - n-1个自由度，因为平均值也是一个自由度
  - 数据的分散程度，标准差越大，数据越分散：

统计知识

- 假设n个样本，每个样本i我们得到两个数据μ_i和ρ_i
- 例：所有同学的期中考试成绩和平时作业成绩
- **协方差**：$\mathrm{cov}(\mu,\rho)=\frac{\sum_{i=1}^{n}(\mu_i-\overline{\mu})(\rho_i-\overline{\rho})}{\mathrm{n}-1}$
  - 描述了μ和ρ之间的相关性
  - $\mathrm{cov}(\mu,\rho)>0$正相关，$\mathrm{cov}(\mu,\rho)<0$负相关

协方差矩阵

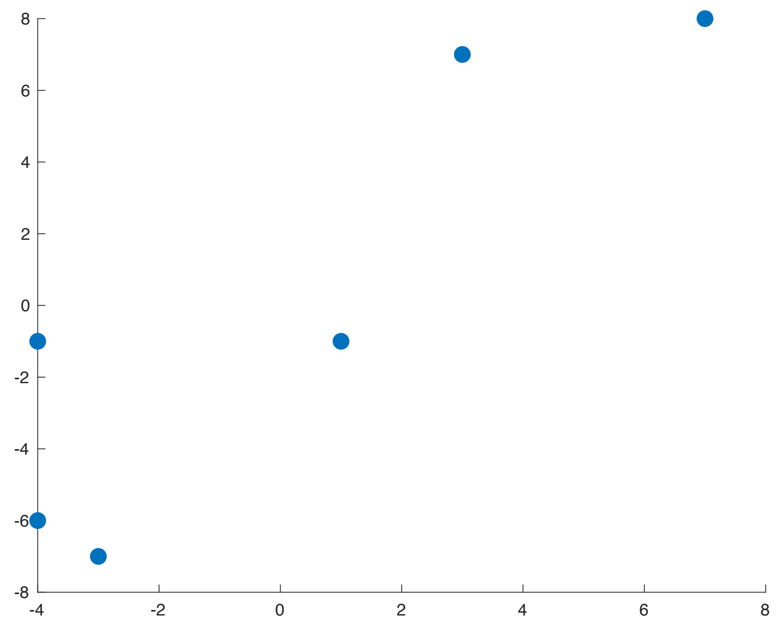
- 将数据存在一个mxn的矩阵A_0中，每一行对应一种数据，每一列代表一个样本
- 矩阵A:
  - 由A_0的每一个元素减去它所在行的平均值得到$A_{ij} = (A_0)_{ij} - \frac{\sum_{k=1}^{n}(A_0)_{ik}}{n}$
  - A中每一行的数据都是以0为中心分布
- **协方差矩阵（covariance matrix）**：$S = \frac{AA^T}{n-1}$
  - 样本**方差**：$S_{ii} = \sigma_i^2$，第i种数据的**标准差**平方。S_{ij}：第i种和第j种数据的**协方差**
  - **总方差（total variance）**：$\mathrm{tr}S = \sum_i S_{ii} = \sum_i \sigma_i^2$

例：

- 6个同学的数学和历史成绩（已经减去平均值）

$$\boldsymbol{A} = \begin{bmatrix} 3 & -4 & 7 & 1 & -4 & -3 \\ 7 & -6 & 8 & -1 & -1 & -7 \end{bmatrix}$$

$$\boldsymbol{S} = \frac{AA^{\mathrm{T}}}{5} = \begin{bmatrix} 20 & 25 \\ 25 & 40 \end{bmatrix}$$



主成分分析（PCA）

- 一般来说数据i和数据j可能会有相关，也就是说它们之间的协方差S_{ij}不等于0
- 主成分分析：找到原有数据的一系列线性组合作为新的数据，新数据之间的协方差为0
  - 奇异值分解：$A = U\Sigma V^T$
- 定义新的数据矩阵：$B = U^T A = \Sigma V^T$， B的协方差矩阵为$\frac{\Sigma^T V^T V\Sigma}{n-1} = \frac{\Sigma^T\Sigma}{n-1}$， 因为$\Sigma^T\Sigma$是对角矩阵，$B$的数据之间协方差为0
  - 总方差不变：$\mathrm{tr}\frac{BB^T}{n-1} = \mathrm{tr}\frac{U^T AA^T U}{n-1} = \frac{\mathrm{tr}U^T AA^T U}{n-1} = \frac{\mathrm{tr}A^T UU^T A}{n-1} = \frac{\mathrm{tr}AA^T}{n-1}$

主成分分析

- 原数据矩阵：$A = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$，第i列向量$\boldsymbol{a}_i$对应样本i的数据
- 新数据矩阵：$B = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n) = U^T A = (U^T \boldsymbol{a}_1, \cdots, U^T \boldsymbol{a}_n)$
  - B的第i列向量$\boldsymbol{b}_i$对应样本i的数据，这些数据由$\boldsymbol{a}_i$的分量决定：$\boldsymbol{b}_i = U^T \boldsymbol{a}_i$
  - 因为U是正交矩阵$UU^T = U^T U = I$，$\boldsymbol{a}_i = U\boldsymbol{b}_i$。
- B的协方差矩阵$\frac{BB^T}{n-1} = \frac{\Sigma^T V^T V \Sigma}{n-1} = \frac{\Sigma^T \Sigma}{n-1}$是对角矩阵
  - 非对角元为0，新数据之间互不相关
  - 新数据的方差=A的奇异值平方/(n-1)。
  - 原方差$\frac{(AA^T)_{ii}}{n-1} = \frac{(UBB^T U^T)_{ii}}{n-1} = \frac{\sum_{k=1}^{r} \sigma_k^2 U_{ik}^2}{n-1}$

主成分分析

- 原数据矩阵： $A=(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n)$，第i列向量$\boldsymbol{a}_i$对应样本i的数据
- 新数据矩阵： $B=(\boldsymbol{b}_1,\cdots,\boldsymbol{b}_n)=U^TA=(U^T\boldsymbol{a}_1,\cdots,U^T\boldsymbol{a}_n)$
- B的协方差矩阵$\frac{BB^T}{n-1}=\frac{\Sigma^TV^TV\Sigma}{n-1}=\frac{\Sigma^T\Sigma}{n-1}$是对角矩阵
  - 新数据的方差=A的奇异值平方/(n-1)。
  - A的非零奇异值的数量是A的秩r，r+1到m的新数据的方差是0
  - 所有的数据都在$\mathbb{R}^m$的m-r个平面$\sum_{j=1}^m U_{ji}x_j=0, i=r+1,\cdots,m$的交集上
  - 所有数据点分布在一个r维的空间中，这个空间由$\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r\}$张成（$C(A)$的正交归一基）
  - 如果第i个奇异值很接近0，说明数据很靠近平面$\sum_{j=1}^m U_{ji}x_j=0$。

主成分：$\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$

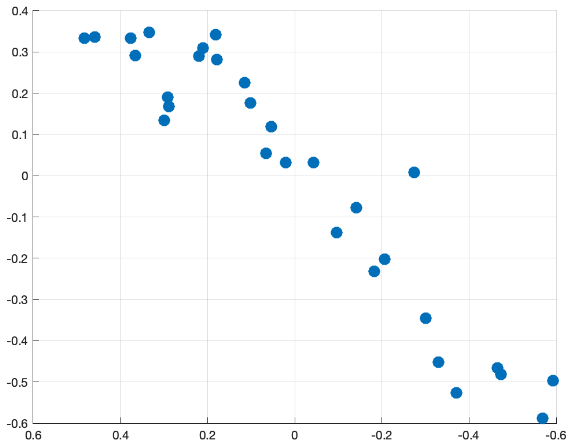
- 原数据矩阵： $A = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$，第i列向量$\boldsymbol{a}_i$对应样本i的数据
- A的奇异值分解： $A = U\Sigma V^T$
- 新数据矩阵： $B = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n) = U^T A = (U^T\boldsymbol{a}_1, \cdots, U^T\boldsymbol{a}_n)$
- 所有数据点分布在一个r维的空间中，这个空间由$\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$张成（$C(A)$的正交归一基）
  - $\boldsymbol{u}_1$是所有数据变化最大的方向（对应的方差最大）， $\boldsymbol{u}_2$ 次之。。。
  - $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$被称作主成分（principle component）
  - 主成分是描述整组数据最重要的线性组合，而且互相独立
  - 由于r小于等于m，所以虽然每个样本测了m个数据，里面只有r个是独立的

$\{\sigma_1 \boldsymbol{v}_1, \cdots, \sigma_r \boldsymbol{v}_r\}$的意义

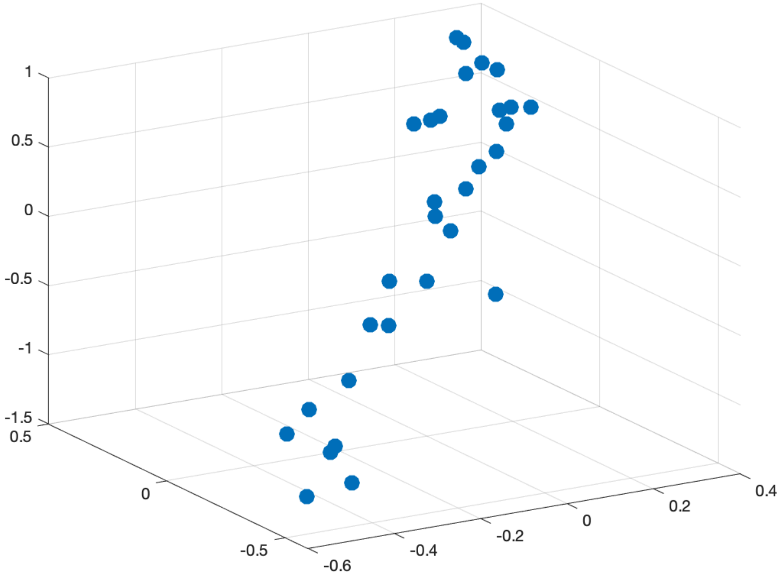
- 原数据矩阵： $A = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$，第i列向量$\boldsymbol{a}_i$对应样本i的数据
- A的奇异值分解： $A = U\Sigma V^T = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_i \boldsymbol{v}_i^T$
  - 主成分$\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r\}$是描述整组数据最重要的线性组合，而且互相独立
- $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r\}$是什么？
  - $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r\}$都是n维向量，每个分量对应一个样本
  - 第一主成分的数值：$\boldsymbol{u}_1^T A = \boldsymbol{u}_1^T(\sum_{k=1}^{r} \sigma_k \boldsymbol{u}_i \boldsymbol{v}_i^T) = \sigma_1 \boldsymbol{v}_1^T$
  - $\sigma_1 \boldsymbol{v}_1$的第i个分量是第i个样本的第一主成分的值，同理$\sigma_j \boldsymbol{v}_j$的第i个分量是第i个样本的第j个主成分的值
  - $\boldsymbol{v}_j$是单位向量，所以每个分量的绝对值小于等于1，数据的分散程度取决于σ_j

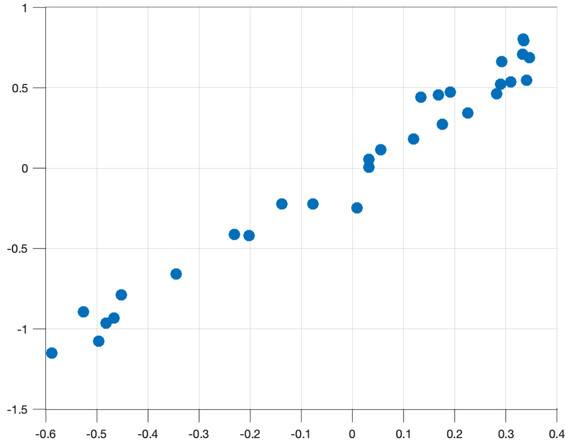
例

- 30个样本，每个样本测三个数据x_1, x_2, x_3

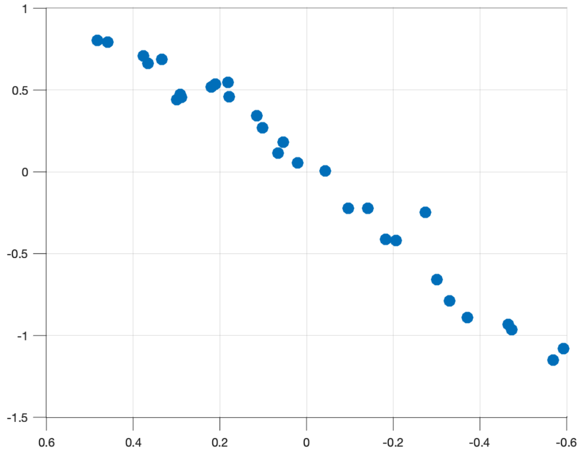


$x_1 - x_2$





$\boldsymbol{x_1 - x_3}$



$x_2 - x_3$

例（续）

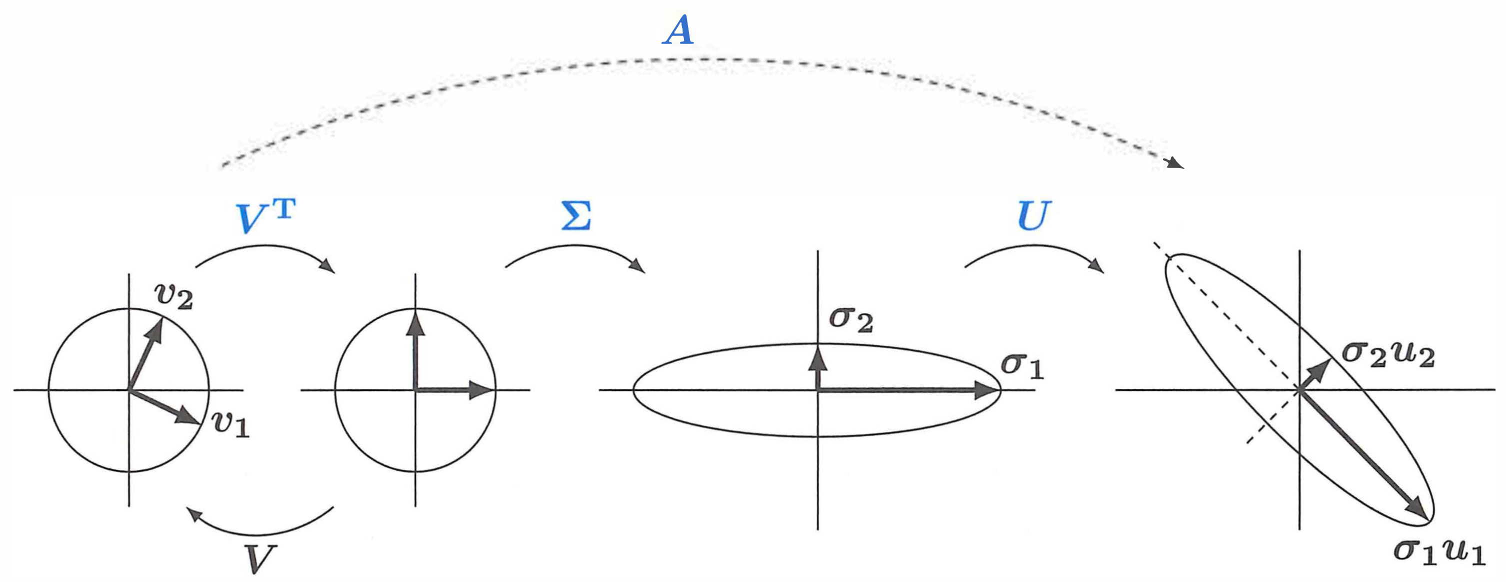
- 协方差矩阵：$\frac{AA^T}{29} = \begin{bmatrix} 0.0987 & 0.0941 & 0.1940 \\ 0.0941 & 0.1004 & 0.1951 \\ 0.1940 & 0.1951 & 0.3908 \end{bmatrix}$
- A的奇异值分解： $A = U\Sigma V^T$
  - $U = \begin{bmatrix} -0.4057 & 0.6838 & 0.6065 \\ -0.4084 & -0.7292 & 0.5492 \\ -0.8177 & 0.0249 & -0.5751 \end{bmatrix}$
  - $D = \begin{bmatrix} 4.1171 & 0 & 0 \\ 0 & 0.3967 & 0 \\ 0 & 0 & 0.0252 \end{bmatrix}$

内容提要

- 奇异值分解
- 主成分分析
- **奇异值分解和几何**

奇异值分解的几何图像

- 考虑2x2的情况 $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} = U\Sigma V^{\mathrm{T}}.$



- 一般情况：（转动+反射）x（拉伸）x（转动+反射）

矩阵的模

- 我们用内积定义了向量的模（长度）：$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}$
- 矩阵的模：$\|A\| = \max\limits_{x \neq 0} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \sigma_1$
- 证明：
  - $\|A\boldsymbol{x}\|^2 = \boldsymbol{x}^T A^T A\boldsymbol{x} = \boldsymbol{x}^T V\Sigma^T\Sigma V^T\boldsymbol{x} = \sum_{k=1}^{r} \boldsymbol{x}^T\boldsymbol{v}_k\sigma_k^2\boldsymbol{v}_k^T\boldsymbol{x}$
  - $\sum_{k=1}^{r} \boldsymbol{x}^T\boldsymbol{v}_k\sigma_k^2\boldsymbol{v}_k^T\boldsymbol{x} = \sigma_1^2\sum_{k=1}^{n} \boldsymbol{x}^T\boldsymbol{v}_k\boldsymbol{v}_k^T\boldsymbol{x} - \sum_{k=2}^{r}(\sigma_1^2 - \sigma_k^2)\boldsymbol{x}^T\boldsymbol{v}_k\boldsymbol{v}_k^T\boldsymbol{x}$
  - $\sigma_1^2\sum_{k=1}^{n} \boldsymbol{x}^T\boldsymbol{v}_k\boldsymbol{v}_k^T\boldsymbol{x} = \sigma_1^2\boldsymbol{x}^T VV^T\boldsymbol{x} = \sigma_1^2\boldsymbol{x}^T\boldsymbol{x} = \sigma_1^2\|\boldsymbol{x}\|^2$
  - $\|A\boldsymbol{x}\|^2 = \sigma_1^2\sum_{k=1}^{n} \boldsymbol{x}^T v_k v_k^T\boldsymbol{x} \leq \sigma_1^2\|\boldsymbol{x}\|^2$
  - 所以$\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \frac{\sigma_1\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \sigma_1$

矩阵的模的性质

- 由矩阵模的定义可知：$\forall\, \boldsymbol{x} \neq 0, \|A\boldsymbol{x}\| \leq \|A\|\|\boldsymbol{x}\|$
- **三角不等式**：$\|A+B\| \leq \|A\| + \|B\|$
- 证明：利用向量模的三角不等式
  - $\|(A+B)\boldsymbol{x}\| = \|A\boldsymbol{x} + B\boldsymbol{x}\| \leq \|A\boldsymbol{x}\| + \|B\boldsymbol{x}\| \leq \|A\|\|\boldsymbol{x}\| + \|B\|\|\boldsymbol{x}\|$
  - $\frac{\|(A+B)\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \|A\| + \|B\|$
  - 所以 $\max\limits_{\boldsymbol{x}\neq 0} \frac{\|(A+B)\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \|A\| + \|B\|$
  - 也就是说：$\|A+B\| \leq \|A\| + \|B\|$
- **乘积不等式**：$\|AB\| \leq \|A\|\|B\|$

矩阵模的应用

- 两个向量差的模代表了两个向量端点的距离
- 类似：两个矩阵差的模量度了两个矩阵之间的“差距”
- **Eckart-Young-Mirsky定理**：同矩阵A最接近的秩为k的矩阵是$A_k = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T$
- 证明：
  - 我们需要证明对于任意秩为k的矩阵B，都有$\|A - B\| \geq \|A - A_k\| = \sigma_{k+1}$

矩阵模的应用

- **Eckart-Young-Mirsky定理**：同矩阵A最接近的秩为k的矩阵是$A_k = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T$
- 证明：
  - 我们需要证明对于任意秩为k的矩阵B，都有$\|A-B\| \geq \|A-A_k\| = \sigma_{k+1}$
  - $w = c_1 \boldsymbol{v}_1 + \cdots + c_{k+1} \boldsymbol{v}_{k+1}$，$Bw = c_1 B \boldsymbol{v}_1 + \cdots + c_{k+1} B \boldsymbol{v}_{k+1}$
  - 因为的秩为k，所以k+1个向量$B\boldsymbol{v}_1, \cdots, B\boldsymbol{v}_{k+1}$线性相关，所以必然存在非零的$\{c_i\}$使的$Bw = 0$。我们可以再假设$\|w\| = 1$
  - $\|A-B\|^2 \geq \|(A-B)w\|^2 = \|Aw\|^2 = \sigma_1^2 c_1^2 + \cdots + \sigma_{k+1}^2 c_{k+1}^2 \geq \sigma_{k+1}^2 \left(c_1^2 + \cdots + c_{k+1}^2\right) = \sigma_{k+1}^2$

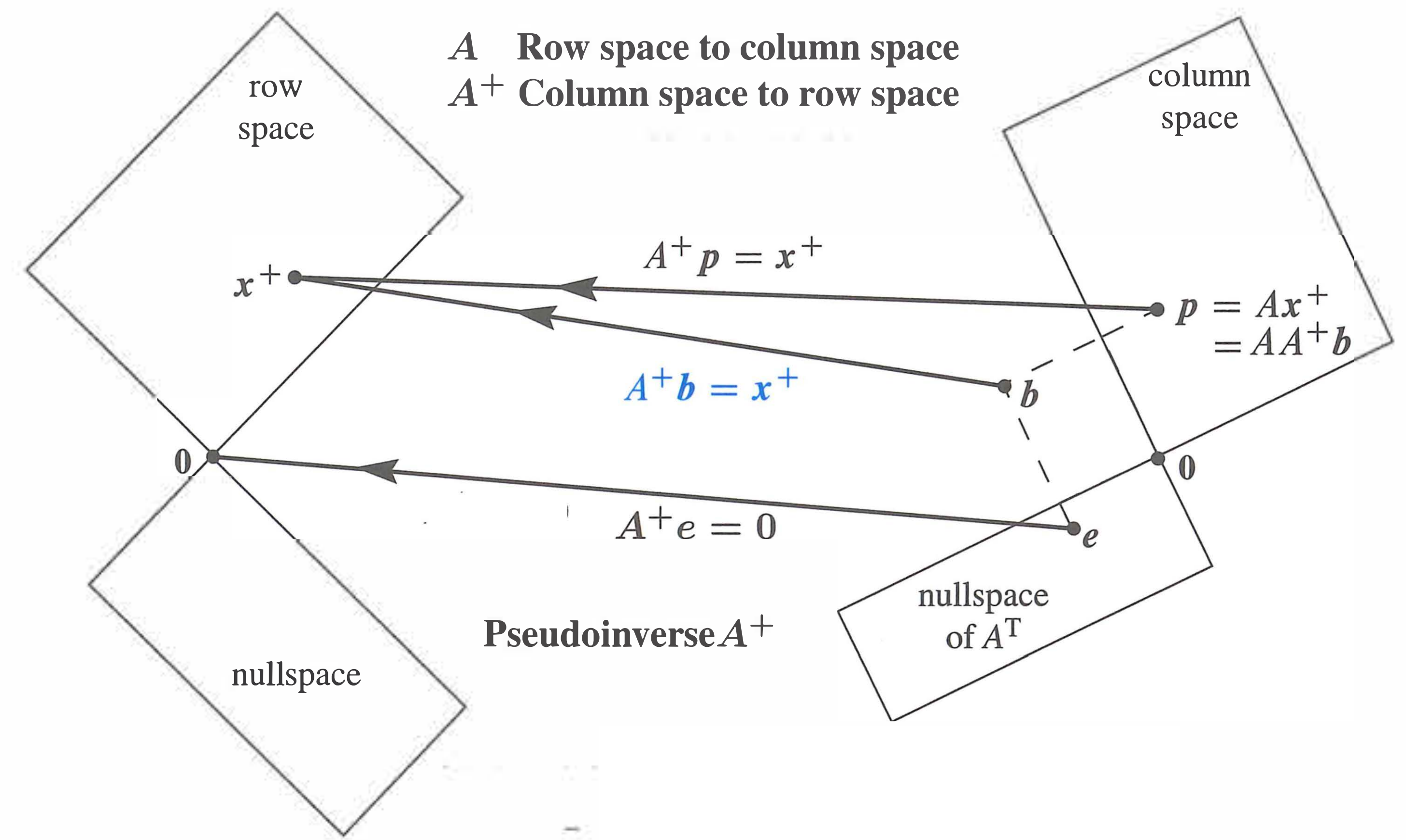
极分解$A = QS$

- 非零复数：$x + iy = re^{i\theta}$，r>0是一个1x1的正定矩阵， $e^{i\theta}$是1x1的幺正矩阵（正交矩阵在复数域上的推广）
- 实方阵的极分解：$A = U\Sigma V^T = (UV^T)(V\Sigma V^T) = QS$
  - $Q = UV^T$是一个正交矩阵
  - S是一个半正定矩阵（因为Σ可能有0特征值）
- 如果A可逆，则Q和S都可逆，这时候S是一个正定矩阵

伪逆（pseudoinverse）

- mxn矩阵$A = U\Sigma V^T$
- 伪逆：$A^+ = V\Sigma^+ U^T = \sum_{k=1}^{r} \boldsymbol{v}_k \sigma_k^{-1} \boldsymbol{u}_k^T$
- 其中$\Sigma^+ = \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}$是一个nxm的矩阵
- $A^+A = V\Sigma^+U^TU\Sigma V^T = V\begin{bmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}V^T$是投影到$C(A^T)$的矩阵
- $AA^+ = U\Sigma V^TV\Sigma^+U^T = U\begin{bmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}U^T$ 是投影到$C(A)$的矩阵
- $AA^+A = A,\quad A^+AA^+ = A^+$

伪逆



A Row space to column space
A^+ Column space to row space
row space
column space
$A^+p = x^+$
x^+
$p = Ax^+$
$= AA^+b$
$A^+b = x^+$
b
0
0
$A^+e = 0$
e
nullspace of A^T
nullspace
Pseudoinverse A^+

伪逆的应用

- 最小二乘法：$A^T A\boldsymbol{x} = A^T\boldsymbol{b}$
- 解：$\boldsymbol{x}^+ = A^+\boldsymbol{b} = V\Sigma^+U^T\boldsymbol{b}$
- 证明：
  - $A^T A\boldsymbol{x}^+ = V\Sigma^T U^T U\Sigma V^T V\Sigma^+ U^T\boldsymbol{b} = V\Sigma^T U^T\boldsymbol{b} = A^T\boldsymbol{b}$