第三章 矩阵分解

3.9 奇 异 值 分 解

若$A \in \mathbb{C}^{n\times n}$有$n$个线性无关的特征向量，则$A$是单纯矩阵. 用于高次幂计算、线性微分方程求解等.

（1）$A \in \mathbb{C}^{m\times n}$?

（2）$A \in \mathbb{C}^{n\times n}$，线性无关的特征向量小于$n$个?

（3）A可相似对角化，$P^{-1}AP = \Lambda$.

$P = [\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n]$的列向量是$A$的特征向量，构成空间$\mathbb{C}^n$的一组基，但不是正交基.

$$P^{-1} = [\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_n]^T$$

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

$$\begin{aligned} A &= P\Lambda P^{-1} \\ &= [\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n]\Lambda[\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_n]^T \\ &= \lambda_1\boldsymbol{\varepsilon}_1{\boldsymbol{\eta}_1}^T + \cdots + \lambda_n\boldsymbol{\varepsilon}_n{\boldsymbol{\eta}_n}^T \end{aligned}$$

$$\begin{aligned} A &= \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -1 \end{bmatrix}\begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \end{bmatrix} + 4\begin{bmatrix} 1 \\ 2 \end{bmatrix}\begin{bmatrix} \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{aligned}$$

$$P^{-1}AP=\Lambda,\quad AP=P\Lambda$$

线性变换在**同一组基**下的表示矩阵！

两组(正交)基,则有**奇异值分解:**

对任意的$A\in\mathbb{C}^{m\times n}$,总存在酉矩阵$U\in\mathbb{C}^{m\times m}$与酉矩阵$V\in\mathbb{C}^{n\times n}$使得

$$A=U\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}V^H$$

其中 $\Sigma_r=\mathrm{diag}(\sigma_1,\cdots,\sigma_r)$.

酉矩阵U与V的列向量分别是AA^H与A^HA的特征向量，$\sigma_1,\cdots,\sigma_r$是AA^H和A^HA的正特征值的平方根.

设$A\in\mathbb{C}^{m\times n}$，则

（1）$\operatorname{rank}(A^HA)=\operatorname{rank}(AA^H)=\operatorname{rank}(A)$.

（2）A^HA与AA^H均是半正定Hermite矩阵.

注1: 设$A \in \mathbb{C}^{n\times n}$是半正定Hermite矩阵

（1）半正定Hermite矩阵的特征值均为非负实数；

（2）存在可逆矩阵P使得$P^H AP = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, 其中

$r = \mathrm{rank}(A)$.

半正定矩阵主子式>=0

余子式为去掉某行某列的行列式大小

正定矩阵顺序主子式大于0

代数余子式是余子式之前*(-1)^(i+j)次幂

引理3.9.1 设$A \in \mathbb{C}_r^{m\times n}$, 则$A^H A$与$AA^H$的所有非零特征值完全相同且非零特征值的个数均为$r$.

例3.9.1 设$A=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}$, 计算$A^HA$和$AA^H$的特征值.

解: $A^HA=\mathrm{diag}(1,1,0)$, $AA^H=I_2$.

A^HA的特征值为$\lambda_1=1,\lambda_2=1,\lambda_3=0$.

AA^H的特征值$\lambda_1=1,\lambda_2=1$.

注2: 在计算A^HA和AA^H的特征值时, 可优先考虑计算阶数较小的矩阵.

注3: 设$A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{n\times m}$, 则$AB$和$BA$具有相同的非零特征值.

法1：利用特征值定义

法2：利用相抵分解

$$A = P^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}Q，\ Q^{-1}BP = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

$$Q^{-1}BAQ = \begin{bmatrix} B_{11} & B_{12} \\ 0 & 0 \end{bmatrix},\ P^{-1}ABP = \begin{bmatrix} B_{11} & 0 \\ B_{21} & 0 \end{bmatrix}$$

注4: 设$A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{n\times m}$, 则$AB$和$BA$具有相同的非零特征值.

法3：利用

$$\begin{bmatrix} I_m & -A \\ 0 & I_n \end{bmatrix}\begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix}\begin{bmatrix} I_m & A \\ 0 & I_n \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix}$$

$$\lambda^n|\lambda I - AB| = \lambda^m|\lambda I - BA|$$

?

思考： AB和BA具有相同的特征值吗？

定义3.9.1(奇异值) 设$A \in \mathbb{C}_r^{m\times n}$, $A^H A$的特征值满足

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0, \lambda_{r+1} = \cdots = \lambda_n = 0$$

称$\sigma_i = \sqrt{\lambda_i}$为$A$的**奇异值**.

特别地, 称 σ_i, $i = 1, \cdots, r$, 为A的**正奇异值**.

若A为方阵, 存在$\sigma_i = 0$, 则A不可逆, 为奇异矩阵.

例3.9.2 计算$A=\begin{bmatrix}1 & 2\\ 0 & 1\\ \mathrm{i} & 0\end{bmatrix}$的正奇异值.

解: $A^HA=\begin{bmatrix}1 & 0 & -\mathrm{i}\\ 2 & 1 & 0\end{bmatrix}\begin{bmatrix}1 & 2\\ 0 & 1\\ i & 0\end{bmatrix}=\begin{bmatrix}2 & 2\\ 2 & 5\end{bmatrix}$.

$$|\lambda I-A^HA|=\begin{vmatrix}\lambda-2 & -2\\ -2 & \lambda-5\end{vmatrix}=\lambda^2-7\lambda+6.$$

求得A^HA的特征值: $\lambda_1=6, \lambda_2=1$.

由此, A的奇异值是$\sigma_1=\sqrt{\lambda_1}=\sqrt{6}, \sigma_2=\sqrt{\lambda_2}=1$.

思考: 复方阵的特征值与奇异值有何异同?

$$A=\begin{bmatrix}0 & 0\\0 & 2\end{bmatrix}, B=\begin{bmatrix}0 & k\\0 & 1\end{bmatrix}, C=\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}.$$

A的奇异值为0, 2; 完全相同

B的奇异值为$\sigma_1=\sqrt{k^2+1}, \sigma_2=0$; 部分相同

C的奇异值为 完全不同

$$\sigma_1=\sqrt{\frac{1}{2}\left(k^2+2+k\sqrt{k^2+4}\right)}, \sigma_2=\frac{1}{\sigma_1}$$

命 题 3.9.1 设 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ 和 $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$分别是$n$阶正规矩阵$A$的奇异值和特征值，则$\sigma_i = |\lambda_i|, i = 1, \cdots, n.$

命题3.9.2 设σ_i和$\lambda_i(i = 1, \cdots, n)$分别是$n$阶复方阵$A$的奇异值和特征值，则

$$\sum_{i=1}^{n} |\lambda_i|^2 \leq \sum_{i=1}^{n} {\sigma_i}^2$$

定理3.9.1(奇异值分解) 设$A \in \mathbb{C}_r^{m\times n}$, 则存在酉矩阵$U \in \mathbb{C}^{m\times m}$与酉矩阵$V \in \mathbb{C}^{n\times n}$使得

$$A = U\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^H$$

式中, $\Sigma_r = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$, $\sigma_i\ (i = 1, \cdots, r)$是矩阵$A$的正奇异值.

注5: 奇异值分解中, U的列向量为AA^H的特征向量, V的列向量为A^HA的特征向量.

$$V_1 = A^H U_1 \Sigma_r^{-1}, \quad U_1 = A V_1 \Sigma_r^{-1}.$$

小结

- $A^H A$与AA^H的性质
- 矩阵奇异值的定义
- 矩阵的奇异值分解

第三章 矩阵分解

3.9 奇 异 值 分 解

奇异值分解步骤:

（1）计算AA^H的m个特征值，确定对角矩阵Σ_r；

（2）计算AA^H的m个单位特征向量$\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_m$，以其为列构成酉矩阵$U$；

其中$\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$分别是属于$\sigma_1^2,\cdots,\sigma_r^2$的特征向量.

（3）计算$A^H\boldsymbol{\alpha}_1,\cdots,A^H\boldsymbol{\alpha}_r$，单位化，并将其扩为$\mathbb{C}^n$的一组标准正交基，构成酉矩阵$V$.

注5: 也可先构造V，然后再依据V构造U.

例3.9.3 求矩阵$A=\begin{bmatrix}2&1\\0&2\\1&0\end{bmatrix}$的奇异值分解.

解: 首先有

$$A^HA=\begin{bmatrix}2&0&1\\1&2&0\end{bmatrix}\begin{bmatrix}2&1\\0&2\\1&0\end{bmatrix}=\begin{bmatrix}5&2\\2&5\end{bmatrix}.$$

$$|\lambda I-A^HA|=\begin{vmatrix}\lambda-5&-2\\-2&\lambda-5\end{vmatrix}=(\lambda-7)(\lambda-3).$$

由此得到A^HA的特征值: $\lambda_1=7,\lambda_2=3.$

求得$A^H A$关于特征值$\lambda_1 = 7$的单位特征向量：

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}.$$

求得$A^H A$关于特征值$\lambda_2 = 3$的单位特征向量：

$$\boldsymbol{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}.$$

令$V = V_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}, \Sigma = \begin{bmatrix}\sqrt{7} & 0\\0 & \sqrt{3}\end{bmatrix},$

令$U_1 = AV_1\Sigma_r^{-1} = \begin{bmatrix} \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{14}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{6}} \end{bmatrix} = [\boldsymbol{u}_1 \quad \boldsymbol{u}_2].$

将$\boldsymbol{u}_1, \boldsymbol{u}_2$扩充为$\mathbb{C}^3$中的标基,

设$\boldsymbol{u}_3 = [x_1, x_2, x_3]^T \in \mathbb{C}^3$. 由$\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$相互正交, 且

$\boldsymbol{u}_3$是单位向量, 得到$\boldsymbol{u}_3 = \frac{1}{\sqrt{21}}\begin{bmatrix} 2 \\ -1 \\ -4 \end{bmatrix}.$

由此得到酉矩阵U:

$$U=\begin{bmatrix}\frac{3}{\sqrt{14}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{21}}\\ \frac{2}{\sqrt{14}} & -\frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{21}}\\ \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{6}} & -\frac{4}{\sqrt{21}}\end{bmatrix}$$

$$A=\begin{bmatrix}\frac{3}{\sqrt{14}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{21}}\\ \frac{2}{\sqrt{14}} & -\frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{21}}\\ \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{6}} & -\frac{4}{\sqrt{21}}\end{bmatrix}\begin{bmatrix}\sqrt{7} & 0\\ 0 & \sqrt{3}\\ 0 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}$$

例3.9.4 求$A=\begin{bmatrix}1 & 0 & 0 & -1\\ 0 & 1 & 0 & 1\end{bmatrix}$的奇异值分解.

解: $AA^H=\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}$,

特征值: $\lambda_1=3, \lambda_2=1.$

$\boldsymbol{u}_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\ -1\end{bmatrix}$, $\boldsymbol{u}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\ 1\end{bmatrix}$.

$U=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}, \Sigma_r=\mathrm{diag}(\sqrt{3},1)$

$A^H A$的特征值$\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = \lambda_4 = 0$. 求得单位正交的特征向量$\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$. 得到

$$V = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \boldsymbol{v}_3 \quad \boldsymbol{v}_4] = \begin{bmatrix} \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{3}} \\ 0 & 0 & 1 & 0 \\ \frac{2}{\sqrt{6}} & 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$$

$$U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix} V^H = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & -1 \end{bmatrix} \neq A$$

定理3.9.2(极分解) 任意复方阵A必有如下分解：

$$A = GW$$

式中, G为半正定Hermite矩阵, W为酉矩阵.

当A可逆时, G是正定Hermite矩阵, 此时该极分解式唯一.

$$A = U\Sigma V^H = U\Sigma U^H UV^H$$

注7: 定理3.9.2中分解式称为**左极分解**.

若$A = VG$, 其中V是酉矩阵, G是半正定Hermite矩阵, 则称分解式$A = VG$为**右极分解**.

应用: 奇异值分解在矩阵特征值、广义逆矩阵等矩阵分析和计算方面有着重要应用,而且在图像处理、机器学习等领域有着广泛应用.

（1）最小二乘问题

设$A \in \mathbb{C}^{m\times n}$, $\boldsymbol{b} \in \mathbb{C}^m$, 求$\boldsymbol{x}^* \in \mathbb{C}^n$使得

$$\|A\boldsymbol{x}^* - \boldsymbol{b}\| = \min_{\boldsymbol{x}\in\mathbb{C}^n}\|A\boldsymbol{x} - \boldsymbol{b}\|$$

一般而言$\boldsymbol{b} \notin R(A)$, 否则必存在$\boldsymbol{x}^* \in \mathbb{C}^n$使得$A\boldsymbol{x}^* = \boldsymbol{b}$.

设A的奇异值分解为:

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

于是

$$A\boldsymbol{x} - \boldsymbol{b} = U\Sigma V^H\boldsymbol{x} - \boldsymbol{b} = U(\Sigma V^H\boldsymbol{x} - U^H\boldsymbol{b}).$$

令$\boldsymbol{y} = V^H\boldsymbol{x}, \boldsymbol{c} = U^H\boldsymbol{b}$,则

$$A\boldsymbol{x} - \boldsymbol{b} = U(\Sigma\boldsymbol{y} - \boldsymbol{c}).$$

因为U是酉矩阵,不改变向量的长度,所以

$$\|A\boldsymbol{x} - \boldsymbol{b}\| = \|\Sigma\boldsymbol{y} - \boldsymbol{c}\|.$$

设$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $\boldsymbol{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}$

知

$$\Sigma\boldsymbol{y} = \begin{bmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

所以 $\Sigma \boldsymbol{y}-\boldsymbol{c}=$

$[\sigma_1 y_1-c_1, \quad \cdots, \quad \sigma_r y_r-c_r, \quad -c_{r+1}, \quad \cdots, \quad -c_m]^T$

当$\sigma_1 y_1-c_1=\cdots=\sigma_r y_r-c_r=0$时,

$\|\Sigma \boldsymbol{y}-\boldsymbol{c}\|$达到最小. 由此解出$y_1, y_2, \cdots, y_r$.

再由$\boldsymbol{y}=V^H \boldsymbol{x}$, 得 $\boldsymbol{x}=V \boldsymbol{y}$.

（2）图像压缩

问题: 存储矩阵$A \in \mathbb{C}^{m\times n}$, 须存储$mn$个数据,

希望用尽可能少的数据来逼近A

考察秩1的矩阵如何存储？

例如, 若$\text{rank}(A) = 1$, 则有

$$A = \boldsymbol{u}\boldsymbol{v}^T, \boldsymbol{u} \in \mathbb{C}^m, \boldsymbol{v} \in \mathbb{C}^n$$

即只要用$m + n$个数据就能表示A.

一幅图像对应一个矩阵：

$$A=\begin{bmatrix} f(1,1) & f(1,2) & \cdots & f(1,n) \\ f(2,1) & f(2,2) & \cdots & f(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ f(m,1) & f(m,2) & \cdots & f(m,n) \end{bmatrix}$$

其中$f(x,y)$表示点(x,y)处图像的灰度或强度.

设A的奇异值分解为

$$A = U\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} V^H$$

$$= \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^H \\ \vdots \\ \boldsymbol{v}_n^H \end{bmatrix}$$

$$= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^H + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^H$$

$$mn \longrightarrow r(m+n)$$

若从中选择$k(k < r)$个大奇异值以及这些奇异值对应的秩1矩阵逼近原图像，即

$$\tilde{A} = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^H$$

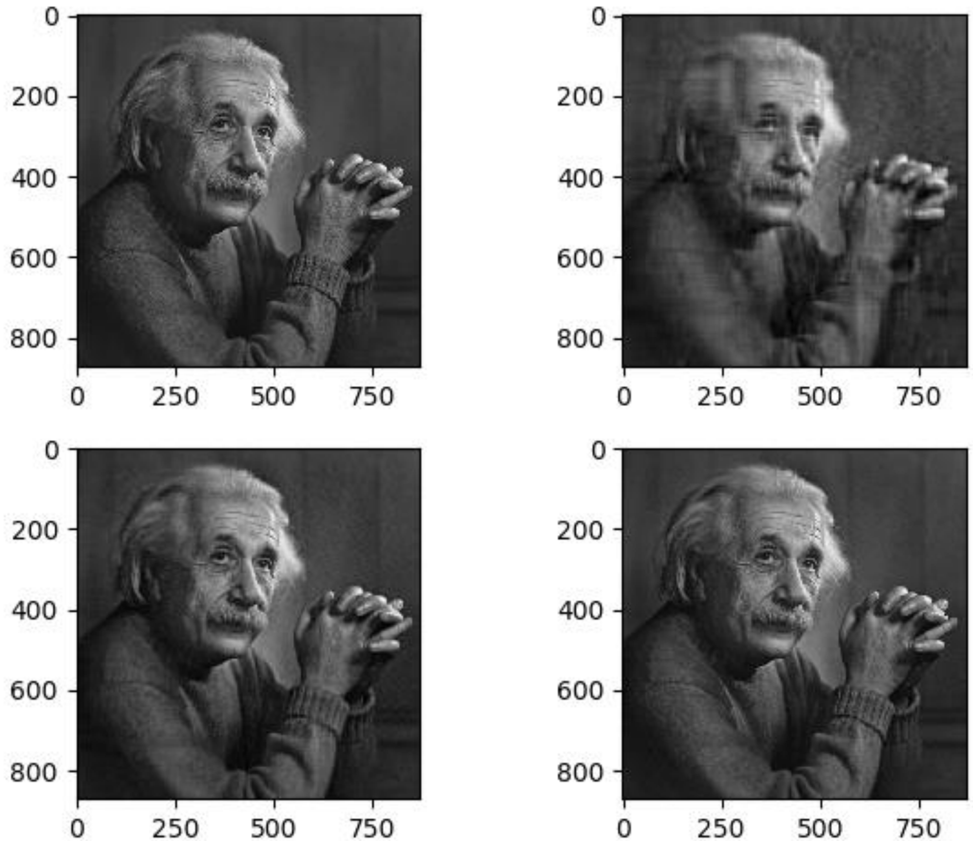
则可利用$\tilde{A}$"近似"A以实现图像的压缩.

k偏小，则重构的图像质量可能不满意；

k过大，降低图像压缩和传送的效率.

需要选择合适的k值以兼顾效率和重构质量.

第三章 矩阵分解——奇异值分解的例子及应用



$k = 20$

$k = 50$

$k = 100$

小结

- 矩阵奇异值分解的步骤与注意事项
- 方阵的极分解
- 矩阵奇异值分解的应用实例。

  最小二乘法,图像压缩

  主成分分析,信息检索