§4.4 矩阵的奇异值分解

(Singular Value Decomposition SVD)

奇异值分解在信号处理、统计学、最佳逼近问题、实验数据处理、数字图像存储中具有广泛的应用.

一、非零矩阵的奇异值及其性质

1. 矩阵 A^HA, AA^H 的性质

定理 令 $A \in C^{m\times n}$, 则 A^HA 和 AA^H 均为 Hermite 矩阵, 且

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

(2) A^HA 和 AA^H 的非零特征值相等;

(3) A^HA 和 AA^H 均半正定, 且当它们为满秩时正定.

Tips: (1) $AX = 0$ 和 $A^HAX = 0$ 同解;

(2) $A^HAX = \lambda X \neq 0 \Rightarrow AA^H(AX) = \lambda(AX) \neq 0$;

(3) $X^HA^HAX = (AX, AX)_{C^m} \geq 0$,

$$X^HA^HAX = 0 \Leftrightarrow AX = 0 \overset{r(A)=n}{\Leftrightarrow} X = 0.$$

注: 半正定(正定) Hermite 矩阵的特征值非负(大于0).

一、非零矩阵的奇异值及其性质

定义　对 $A \in C^{m\times n}$，若 $\text{rank}(A) = r > 0$ 且 $A^H A$ 的特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0,\ \lambda_{r+1} = \cdots = \lambda_n = 0$, 则正数 $\sigma_i = \sqrt{\lambda_i},\ 1 \leq i \leq r$，称为 A 的奇异值或奇值.

注: 由于 $\text{rank}(A^H A) = \text{rank}(A) = r$ 且 $A^H A = U\Lambda U^H$, $A^H A$ 必有 r 个正特征值, 即 A 必有 r 个奇异值.

二、非零矩阵的奇异值分解

定理　设 $A \in C^{m\times n}, r(A) = r, A$ 的奇异值 $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r$. 则存在酉矩阵 $U \in C^{m\times m}$ 和 $V \in C^{n\times n}$，使得

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

(酉等价到 Σ: $U^H AV = \Sigma$)

其中，$\Sigma = \begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}$，$\Delta = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$.

证明 由条件, $A^H A$ 是非负Hermite矩阵, 故可酉相似对角化

$$V^H A^H AV = \mathrm{diag}(\lambda_1, \cdots, \lambda_r; 0, \cdots, 0) = \begin{bmatrix} \Delta^2 & 0 \\ 0 & 0 \end{bmatrix}_{n\times n} = \Sigma^2,$$

其中 $\lambda_i = \sigma_i^2, i \le r$, 为$A^H A$ 的前 r 个特征值, 且 V 的列向量 $\{v_1, v_2, \cdots, v_n\}$为 $A^H A$ 对应于特征值的标准正交特征向量.

考察AV. 由 $\left(Av_i, Av_j\right) = v_j^H A^H Av_i = v_j^H \lambda_i v_i = 0,\ i \ne j \le r,$

且 $\| Av_i \|^2 = (Av_i, Av_i) = v_i^H \lambda_i v_i = \lambda_i = \sigma_i^2,\ i \le r,$

令 $u_i = \dfrac{Av_i}{\sigma_i}$，则 $\{u_1, u_2, \cdots, u_r\}$ 标准正交.

二、非零矩阵的奇异值分解

定理 设 $A \in C^{m\times n}, r(A)=r, A$ 的奇异值 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$.

则存在酉矩阵 $U \in C^{m\times m}$ 和 $V \in C^{n\times n}$，使得

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

(酉等价到 Σ: $U^H AV = \Sigma$)

其中，$\Sigma = \begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}$，$\Delta = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$.

证明 将其扩充为C^m 的标准正交基$\{u_1, u_2, \cdots, u_r; u_{r+1}, \cdots, u_m\}$.

则获得酉矩阵 $U = (u_1, u_2, \cdots, u_m)$, $u_i = \dfrac{Av_i}{\sigma_i}$, $i \leq r$.

由于 $Av_i = \sigma_i u_i$, $i \leq r$; $Av_i = 0$, $i > r$, 有

$$AV = (Av_1, \cdots, Av_r; 0) = (\sigma_1 u_1, \cdots, \sigma_r u_r; 0)$$

$$= (u_1, \cdots, u_r; u_{r+1}, \cdots, u_m)\begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix} = U\Sigma.$$

故 $A = U\Sigma V^H$. ■

二、非零矩阵的奇异值分解

定理　设 $A \in C^{m\times n}, r(A) = r$, A 的奇异值 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. 则存在酉矩阵 $U \in C^{m\times m}$ 和 $V \in C^{n\times n}$，使得

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

(酉等价到 Σ: $U^H AV = \Sigma$)

其中，$\Sigma = \begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}$, $\Delta = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$.

总结 (SVD的构造过程)

(1) 求 $A^H A$ 的特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = 0$,
即得 A 的奇异值 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ 和 Σ;

(2) $A^H A$ 的对应于特征值的线性无关特征向量标准正交化后排成 V;

(3) $U = (u_1, u_2, \cdots u_r; u_{r+1}, \cdots, u_m)$, $u_i = \frac{Av_i}{\sigma_i}$, $i \leq r$,
后面 $u_{r+1}, \cdots, u_m$ 为任意补齐的标准正交向量组.

则 $A = U\Sigma V^H$.

二、非零矩阵的奇异值分解 ★

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

解　(1) 求 A^HA 的特征值, 按大小排列 $\lambda_1=3,\lambda_2=1,\lambda_3=0$;

(2) 求一组对应于 λ_i 的特征向量(已正交), 标准化后得 V:

$$V=\begin{bmatrix}\frac{1}{\sqrt6}&\frac{1}{\sqrt2}&\frac{1}{\sqrt3}\\\frac{1}{\sqrt6}&\frac{-1}{\sqrt2}&\frac{1}{\sqrt3}\\\frac{2}{\sqrt6}&0&\frac{-1}{\sqrt3}\end{bmatrix}=(v_1,v_2,v_3);$$

(3) 构造 U: $\xrightarrow{u_i=\frac{Av_i}{\sigma_i}}$ $u_1=\frac{1}{\sqrt2}\begin{bmatrix}1\\1\\0\end{bmatrix}$, $u_2=\frac{1}{\sqrt2}\begin{bmatrix}1\\-1\\0\end{bmatrix}$, u_3 ?

取 $u_3=(0,0,1)$, 则 u_3 为单位向量且正交于 u_1,u_2.

令 $U=(u_1,u_2,u_3)$, 则有 $A=U\Sigma V^H$.

二、非零矩阵的奇异值分解 ★

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

解　(1) 求 B^HB 的特征值, 按大小排列 $\lambda_1=2,\lambda_2=1,\lambda_3=0$;

(2) 求一组对应于 λ_i 的特征向量(已正交), 标准化后得 V

$$v_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix},\quad v_2=\begin{bmatrix}0\\0\\1\end{bmatrix},\quad v_3=\frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\\0\end{bmatrix};$$

(3) 构造 U:

$$u_1=\frac{Bv_1}{\sigma_1}=\begin{bmatrix}1\\0\end{bmatrix},\quad u_2=\frac{Bv_2}{\sigma_2}=\begin{bmatrix}0\\1\end{bmatrix},$$

已得 U!

令 $U=(u_1,u_2)$, 则有 $B=U\Sigma V^H$. ■

求BB^H的特征值更方便，有相同非零特征值

二、非零矩阵的奇异值分解

奇异值展开式

对于矩阵 $A \in C^{m\times n}$，$r(A)=r$，设有奇异值分解

$$A = U\Sigma V^H = (u_1,\cdots,u_r;u_{r+1},\cdots,u_m)\begin{bmatrix}\Delta & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}v_1^H\\ v_2^H\\ \vdots\\ v_n^H\end{bmatrix}$$

$$= \sigma_1 u_1 v_1^H + \sigma_2 u_2 v_2^H + \cdots + \sigma_r u_r v_r^H.$$

定理 设 $A \in C^{m\times n}$, A 的奇异值 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, 则 A 有奇异值展开式

$$A = \sigma_1 u_1 v_1^H + \sigma_2 u_2 v_2^H + \cdots + \sigma_r u_r v_r^H.$$

r 个秩为 1 的矩阵之和

二、非零矩阵的奇异值分解

例 图像的数字化技术与矩阵的奇异值分解

1) 计算机处理图像技术的第一步是图像的数字化存储技术，即将图像转换成矩阵来存储.

转换的原理是将图形分解成像素（pixels）的一个矩形的数阵, 其中的信息就可以用一个矩阵 $A=\left(a_{ij}\right)_{m\times n}$ 来存储. 矩阵 A 的元素 a_{ij} 是一个正的数, 它相应于像素的灰度水平（gray level） 的度量值.

2) 压缩数字化图形存储量的方法主要是应用矩阵的奇异值分解和矩阵范数下的逼近.

若图像矩阵 A 的奇异值展开式为 $A=\sigma_1 u_1 v_1^H+\sigma_2 u_2 v_2^H+\cdots+\sigma_r u_r v_r^H$, 取

$$A_k=\sigma_1 u_1 v_1^H+\sigma_2 u_2 v_2^H+\cdots+\sigma_k u_k v_k^H, \qquad k\leq r,$$

则在所有秩不超过 k 的矩阵中, A_k 与 A 的 Frobenius 距离最小 .

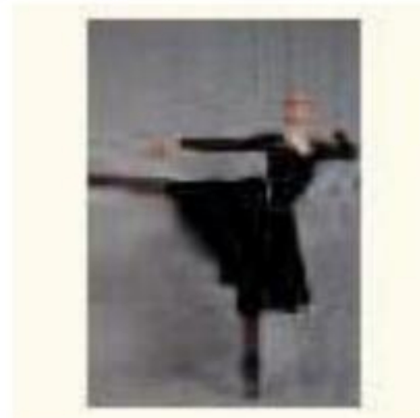
3) 压缩效果.

存储 A_k 只需要存储 k 个奇异值，k 个 m 维向量 u_i 和 k 个 n 维向量 v_i，共 $k+km+kn$ 个数值, 相比原来 mn 个数值要少得多. 例如 $m=n=1000$, $k=100$ 时，A_k 的数据存储量仅为 200100, 减少 80%.

Singular Value Decomposition in Image Processing

Original

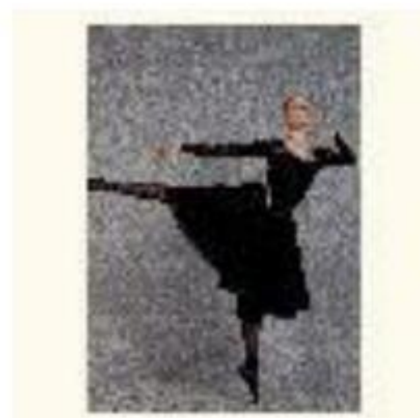
5 values



10 values

30 values

52 values



303 values

图像数据的奇异值分解压缩：秩从4到128

三、矩阵 A 的奇异值分解与线性变换 T_A

U, V 的空间性质

设 $r(A_{m\times n}) = r$，A 有SVD分解 $U^H AV = \Sigma.$ 由于 $U_{m\times m}$，$V_{n\times n}$ 为酉阵，其列向量分别构成 C^m，C^n 的标准正交基.

将 U,V 按前 r 列分块, 有

$$AV = U\Sigma \Rightarrow A(V_1, V_2) = (U_1, U_2)\begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} = (U_1\Delta, 0),$$

从而

$$AV_1 = U_1\Delta = (\sigma_1 u_1, \sigma_2 u_2, \cdots, \sigma_r u_r), \qquad AV_2 = 0.$$

由此,

(1) U_1 的列向量构成 $R(A)$ 的一组标准正交基;

(2) U_2 的列向量构成 $R^{\perp}(A)$ 的一组标准正交基;

(3) V_2 的列向量构成 $N(A)$ 的一组标准正交基;

(4) V_1 的列向量构成 $N^{\perp}(A)$ 的一组标准正交基.

三、矩阵 A 的奇异值分解与线性变换 T_A

对 $A \in C^{m\times n}$, 可定义线性变换 T_A: $C^n \to C^m$, $T_A(X) = AX$.

若有奇异值分解 $A = U\Sigma V^H$, 有

$$T_A(v_1, v_2, \cdots, v_n) = A(v_1, v_2, \cdots, v_n) = (u_1, u_2, \cdots, u_m)\Sigma.$$

将 V 和 U 的列分别视作 C^n, C^m 的基, 则 T_A 的变换矩阵为 Σ.

故若 $\alpha \in C^n$ 在基 $\{v_1, \cdots, v_n\}$ 下的坐标 $X = (x_1, \cdots, x_n)^\top$, 则

$T_A(\alpha)$在基$\{u_1, \cdots, u_m\}$下坐标 $Y = \Sigma X = (\sigma_1 x_1, \cdots, \sigma_r x_r; \mathbf{0})^\top$.

总之，$r(A_{m\times n}) = r \;\Rightarrow\; T_A$: $C^n \to C^m$ 满足

原像 α 在基 $\{v_1, \cdots, v_n\}$ 下的坐标　　　像 $T\alpha$ 在基 $\{u_1, \cdots, u_m\}$ 下的坐标

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{bmatrix} \xrightarrow{\quad T_A \quad} T_A(X) = \begin{bmatrix} \sigma_1 x_1 \\ \vdots \\ \sigma_r x_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

三、矩阵 A 的奇异值分解与线性变换 T_A

例如, 如果 α 来自 R^n 的单位球面: 在基 $\{v_1,\cdots,v_n\}$ 下的坐标

$$\sum_{1\le i\le n} x_i^2 = 1,$$

则像 $T\alpha$ 在 C^m 的基 $\{u_1,\cdots,u_m\}$ 下的坐标满足

$$\left(\frac{y_1}{\sigma_1}\right)^2+\left(\frac{y_2}{\sigma_2}\right)^2+\cdots+\left(\frac{y_r}{\sigma_r}\right)^2\le 1,\ \ r\le n,$$

且当 $r=n$ 时, 等号成立.

定理 设 $A\in R^{m\times n}$, $r(A)=r$, 则 R^n 中的单位球面在 T_A 下的像是 (1) R^m 中的 n 维椭球面, 如果 $r=n$;
(2) R^m 中的 r 维椭球体, 如果 $r<n$.

注：图形的形状不依赖于标准正交基的选取！

三、矩阵 A 的奇异值分解与线性变换 T_A

例2　设 $A=\begin{bmatrix}1&0&1\\0&1&1\\0&0&0\end{bmatrix}$, 求 R^3 的单位球面在 T_A 下的像的图形.

解　A 有奇异值 $\sigma_1=\sqrt{3}$, $\sigma_2=1$, 且有 SVD

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

故单位球面在 T_A 下的像在基 $\{u_1,u_2,u_3\}$ 下的坐标满足

$$\left(\frac{y_1}{\sqrt{3}}\right)^2+\left(\frac{y_2}{1}\right)^2\le 1,$$

即单位球面的像为 R^3 中 $L\{u_1,u_2\}$ 平面上的2维的实心椭圆. ■

四、方阵的极分解 (Polar Decomposition)

定理 设 $A \in C^{n\times n}$, $r(A) = r$. 则 A 有**极分解**

$$A = P_{n\times n} Q_{n\times n},$$

拉伸 ∘ 旋转

其中, P半正定Hermite阵且 $r = n$ 时正定; Q酉矩阵.

证明 设 A 的奇异值分解 $A = U\Sigma V^H = (U\Sigma U^H)(UV^H)$.

令

$$\boxed{P = U\Sigma U^H, \qquad Q = UV^H.}$$

则 P 是 Hermite 阵, 且相似于 Σ, 故以 A 的奇异值或 0 为特征值, 非负, 从而 P 半正定.

当 $r = n$ 时, A 有 n 个正的奇异值,

故 P 特征值全为正, 正定.

Q 为酉矩阵之积, 为酉矩阵. ■

四、方阵的极分解 (Polar Decomposition)

例 3　求 $A=\begin{bmatrix}\sqrt{3} & 2\\ 0 & \sqrt{3}\end{bmatrix}$ 的极分解, 并分析 T_A 的几何意义.

解　易求得 A 的奇异值为 $\sigma_1=3,\sigma_2=1$, 及奇异值分解

$$A=\underbrace{\begin{bmatrix}\frac{\sqrt{3}}{2} & -\frac{1}{2}\\ \frac{1}{2} & \frac{\sqrt{3}}{2}\end{bmatrix}}_{U}\begin{bmatrix}3 & \\ & 1\end{bmatrix}\underbrace{\begin{bmatrix}\frac{1}{2} & \frac{\sqrt{3}}{2}\\ -\frac{\sqrt{3}}{2} & \frac{1}{2}\end{bmatrix}}_{V^H}.$$

令

$$P=U\Sigma U^H=\begin{bmatrix}\frac{5}{2} & \frac{\sqrt{3}}{2}\\ \frac{\sqrt{3}}{2} & \frac{5}{2}\end{bmatrix},\quad Q=UV^H=\begin{bmatrix}\frac{\sqrt{3}}{2} & \frac{1}{2}\\ -\frac{1}{2} & \frac{\sqrt{3}}{2}\end{bmatrix}.$$

得极分解 $A=PQ$, 且

$$T_A(X)=AX=P(QX)=T_P\circ T_Q(X).$$

四、方阵的极分解 (Polar Decomposition)

例 3　求 $A=\begin{bmatrix}\sqrt{3} & 2\\ 0 & \sqrt{3}\end{bmatrix}$ 的极分解, 并分析 T_A 的几何意义.

解

$$P=U\Sigma U^H=\begin{bmatrix}\frac{5}{2} & \frac{\sqrt{3}}{2}\\ \frac{\sqrt{3}}{2} & \frac{5}{2}\end{bmatrix},\quad Q=UV^H=\begin{bmatrix}\frac{\sqrt{3}}{2} & \frac{1}{2}\\ -\frac{1}{2} & \frac{\sqrt{3}}{2}\end{bmatrix}.$$

考察 T_Q，由于

$$Q=\begin{bmatrix}\frac{\sqrt{3}}{2} & \frac{1}{2}\\ -\frac{1}{2} & \frac{\sqrt{3}}{2}\end{bmatrix}=\begin{bmatrix}\cos\frac{\pi}{6} & \sin\frac{\pi}{6}\\ -\sin\frac{\pi}{6} & \cos\frac{\pi}{6}\end{bmatrix},$$

故 T_Q 为绕原点顺时针旋转 $\frac{\pi}{6}$ 的变换;

四、方阵的极分解 (Polar Decomposition)

例 3　求 $A=\begin{bmatrix}\sqrt{3} & 2\\ 0 & \sqrt{3}\end{bmatrix}$ 的极分解, 并分析 T_A 的几何意义.

解

而 $P=U\Sigma U^H=\begin{bmatrix}\frac{5}{2} & \frac{\sqrt{3}}{2}\\ \frac{\sqrt{3}}{2} & \frac{5}{2}\end{bmatrix}$ 的特征值为 3 和 1,

其对应的标准正交的特征向量为

$$u_1=\left(\frac{\sqrt{3}}{2},\ \frac{1}{2}\right)^{\mathrm{T}},\quad u_2=\left(-\frac{1}{2},\ \frac{\sqrt{3}}{2}\right)^{\mathrm{T}}.$$

故对任意 $X\in R^2, X=k_1u_1+k_2u_2$, 有

$$\begin{aligned}T_P(X)=PX&=k_1Pu_1+k_2Pu_2\\&=3k_1u_1+k_2u_2,\end{aligned}$$

即 T_P 是一个拉伸变换: 将 X 沿 u_1 方向拉伸 3 倍, u_2 方向保持不变. ■