

第三讲 线性最小二乘问题

奇异值分解 (SVD)



目录

3.1 问题介绍

3.2 Householder 变换和 Givens 变换

3.3 QR 分解

3.4 奇异值分解

3.5 线性最小二乘问题的求解方法

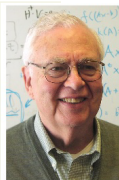
Professor SVD

BY CLEVE MOLER

Stanford computer science professor Gene Golub has done more than anyone to make the singular value decomposition one of the most powerful and widely used tools in modern matrix computation.



Gene Golub's license plate, photographed by Professor P. M. Kroonenberg of Leiden University.



Gene Golub

Popularized numerical computing with matrices via the informal "Golub thesis"

"anything worth computing can be stated as a matrix problem"

**THE \$25,000,000,000* EIGENVECTOR
THE LINEAR ALGEBRA BEHIND GOOGLE**

KURT BRYAN[†] AND TANYA LEISE[‡]

William Kahan

Formalized IEEE-754 floating point arithmetic.

Make it possible to compute with probabilities as "real numbers" instead of discrete counts.



3-4 | 奇异值分解

3.4 奇异值分解

3.4.1 奇异值与奇异值分解

3.4.2 奇异值的性质

潘建瑜 @MATH.ECNU

<https://math.ecnu.edu.cn/~jypan/Teaching/NA>

SVD 不仅是矩阵计算中的重要工具之一, 也是图像处理、压缩感知、机器学习、数据科学等领域的重要技术.

3-4-1 | 奇异值与奇异值分解

定理 设 $A \in \mathbb{C}^{m \times n}$ ($m \geq n$), 则存在酉矩阵 $U \in \mathbb{C}^{m \times m}$ 和 $V \in \mathbb{C}^{n \times n}$ 使得

$$U^*AV = \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} \quad \text{或} \quad A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*, \quad (3.1)$$

其中 $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) \in \mathbb{R}^{n \times n}$, 且

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0.$$

分解 (3.1) 称为 A 的**奇异值分解 (SVD)**, 而 $\sigma_1, \sigma_2, \dots, \sigma_n$ 则称为 A 的**奇异值**.

(板书)

3-4-1 | 奇异值与奇异值分解

定理 设 $A \in \mathbb{C}^{m \times n}$ ($m \geq n$), 则存在酉矩阵 $U \in \mathbb{C}^{m \times m}$ 和 $V \in \mathbb{C}^{n \times n}$ 使得

$$U^*AV = \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} \quad \text{或} \quad A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*, \quad (3.1)$$

其中 $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) \in \mathbb{R}^{n \times n}$, 且

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0.$$

分解 (3.1) 称为 A 的**奇异值分解 (SVD)**, 而 $\sigma_1, \sigma_2, \dots, \sigma_n$ 则称为 A 的**奇异值**.

(板书)



如果 $A \in \mathbb{R}^{m \times n}$ 是实矩阵, 则 U, V 也都可以是实矩阵.

奇异值与特征值

$$A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*$$

 $\Rightarrow$

$$A^*A = V\Sigma^*\Sigma V^*, \quad AA^* = U \begin{bmatrix} \Sigma\Sigma^* & 0 \\ 0 & 0 \end{bmatrix} U^*$$

奇异值与特征值

$$A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*$$

$\Rightarrow$

$$A^*A = V\Sigma^*\Sigma V^*, \quad AA^* = U \begin{bmatrix} \Sigma\Sigma^* & 0 \\ 0 & 0 \end{bmatrix} U^*$$

A 的奇异值就是 A^*A 的特征值的平方根.

奇异值与特征值

$$A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*$$

$\Rightarrow$

$$A^*A = V\Sigma^*\Sigma V^*, \quad AA^* = U \begin{bmatrix} \Sigma\Sigma^* & 0 \\ 0 & 0 \end{bmatrix} U^*$$

A 的奇异值就是 A^*A 的特征值的平方根.



若 $\text{rank}(A) = r \leq n$, 则有

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0.$$



特别地, 如果 $\text{rank}(A) = n$, 则奇异值都是正的, 此时对角矩阵 Σ 非奇异.

奇异向量

$$A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*$$

 $\Rightarrow$

$$AV = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix}, \quad A^*U = [\Sigma, 0]V$$

$$Av_i = \sigma_i u_i, \quad i = 1, 2, \dots, n,$$

$$A^*u_i = \sigma_i v_i, \quad i = 1, 2, \dots, n,$$

$$A^*u_i = 0, \quad i = n+1, n+2, \dots, m.$$

 矩阵 $U = [u_1, u_2, \dots, u_m]$ 的列向量称为 A 的**左奇异向量**

 矩阵 $V = [v_1, v_2, \dots, v_n]$ 的列向量称为 A 的**右奇异向量**

降阶 SVD

$$A = [u_1, \dots, u_m] \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} [v_1, \dots, v_n]^* = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \dots + \sigma_n u_n v_n^* = \sum_{i=1}^n \sigma_i u_i v_i^*$$

降阶 SVD

$$A = [u_1, \dots, u_m] \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} [v_1, \dots, v_n]^* = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \dots + \sigma_n u_n v_n^* = \sum_{i=1}^n \sigma_i u_i v_i^*$$

✿ 记 $U_n = [u_1, u_2, \dots, u_n] \in \mathbb{C}^{m \times n}$, 则 U_n 是单位列正交矩阵 (即 $U_n^* U_n = I_{n \times n}$), 且

$$A = U_n \Sigma V^* \quad (3.2)$$

这就是 **细 SVD (thin SVD)** 或 **降阶 SVD (reduced SVD)**.

📖 有的文献将 (3.2) 称为奇异值分解.

截断 SVD

 设 $k < n$, 我们称

$$A_k = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots + \sigma_k u_k v_k^* = \sum_{i=1}^k \sigma_i u_i v_i^*$$

为 A 的 **截断 SVD** (**truncated SVD**)

截断 SVD

 设 $k < n$, 我们称

$$A_k = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots + \sigma_k u_k v_k^* = \sum_{i=1}^k \sigma_i u_i v_i^*$$

为 A 的 **截断 SVD (truncated SVD)**

如果 $\text{rank}(A) = r < n$, 则 $\sigma_i = 0, i = r + 1, \dots, n$, 此时

$$A = A_r = \sum_{i=1}^r \sigma_i u_i v_i^*.$$

奇异值的应用

- 计算矩阵范数和条件数
- 计算矩阵的数值秩 (numerical rank)
- 求解最小二乘问题
- 矩阵和张量的低秩分解
- 图像处理, 压缩感知, 主成分分析, 数据分析, ...

3-4-2 | 奇异值的性质

定理 设 $A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*$ 是 $A \in \mathbb{C}^{m \times n}$ ($m \geq n$) 的奇异值分解, 则下面结论成立:

- (1) A^*A 的特征值是 σ_i^2 , 对应的特征向量是 $v_i, i = 1, 2, \dots, n$;
- (2) AA^* 的特征值是 σ_i^2 和 $m - n$ 个零, 对应的特征向量是 $u_i, i = 1, 2, \dots, m$;
- (3) $\|A\|_2 = \sigma_1, \|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}$;
- (4) 若 $\text{rank}(A) = r \leq n$, 则

$$\text{Ran}(A) = \text{span}\{u_1, u_2, \dots, u_r\}, \quad \text{Ker}(A) = \text{span}\{v_{r+1}, v_{r+2}, \dots, v_n\};$$

- (5) (酉不变性) 设 $X \in \mathbb{C}^{m \times m}$ 和 $Y \in \mathbb{C}^{n \times n}$ 是酉矩阵, 则 $\sigma_i(X^*AY) = \sigma_i(A)$.

(留作练习)

奇异值基本性质 (续)

定理 设 $A = U\Sigma V^*$ 是 $A \in \mathbb{C}^{n \times n}$ 的奇异值分解, 则下面结论成立:

(1) $|\det(A)| = \sigma_1 \sigma_2 \cdots \sigma_n$;

(2) 若 A 非奇异, 则 $\|A^{-1}\|_2 = \sigma_n^{-1}$, $\kappa_2(A) = \sigma_1/\sigma_n$;

(留作练习)

SVD 的重要应用: 低秩逼近

定理 设 $A = U_n \Sigma V^*$ 是 $A \in \mathbb{C}^{m \times n}$ 的降阶 SVD. 令 $A_k = \sum_{i=1}^k \sigma_i u_i v_i^*$, 则 A_k 是

$$\min_{B \in \mathbb{C}^{m \times n}, \text{rank}(B)=k} \|A - B\|_2 \quad (3.3)$$

的一个解, 且

$$\|A - A_k\|_2 = \sigma_{k+1}.$$

此时, 我们称 A_k 是 A 的一个秩- k 逼近.

(板书)

SVD 的重要应用: 低秩逼近

定理 设 $A = U_n \Sigma V^*$ 是 $A \in \mathbb{C}^{m \times n}$ 的降阶 SVD. 令 $A_k = \sum_{i=1}^k \sigma_i u_i v_i^*$, 则 A_k 是

$$\min_{B \in \mathbb{C}^{m \times n}, \text{rank}(B)=k} \|A - B\|_2 \quad (3.3)$$

的一个解, 且

$$\|A - A_k\|_2 = \sigma_{k+1}.$$

此时, 我们称 A_k 是 A 的一个秩- k 逼近.

(板书)

 定理中的 (3.3) 式也可以改写为

$$\min_{B \in \mathbb{C}^{m \times n}, \text{rank}(B) \leq k} \|A - B\|_2 \quad (3.4)$$

谢谢
THANK YOU

