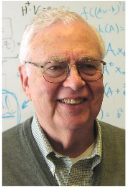
第三讲　线性最小二乘问题

— 奇异值分解 (SVD)

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

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.

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

什么是 SVD

SVD

奇异值分解 (SVD) 不仅仅是矩阵计算中非常有用的工具之一，也是图像处理、数据分析、压缩感知等领域的重要技术.

奇异值与奇异值分解

定理 (SVD) 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

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

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

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

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

(板书)

奇异值与奇异值分解

定理 (SVD) 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

$$U^* A V = \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} \quad 或 \quad A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*, \tag{3.1}$$

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

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

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

由 (3.1) 可知,

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

所以 $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ 是 A^*A 和 AA^* 的特征值.

由 (3.1) 可知,

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

所以 $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ 是 A^*A 和 AA^* 的特征值.

A 的奇异值就是 A^*A 的特征值的平方根.

由 (3.1) 可知,

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

所以 $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ 是 A^*A 和 AA^* 的特征值.

A 的奇异值就是 A^*A 的特征值的平方根.

若 $\mathrm{rank}(A) = r \le n$, 则有

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

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

奇异向量

矩阵 $U = [u_1, u_2, \ldots, u_m]$ 和 $V = [v_1, v_2, \ldots, v_n]$ 的列向量分别称为 A 的**左奇异向量**和**右奇异向量**, 即存在关系式

$$
\begin{aligned}
&Av_i = \sigma_i u_i, \quad i = 1, 2, \ldots, n, \\
&A^* u_i = \sigma_i v_i, \quad i = 1, 2, \ldots, n, \\
&A^* u_i = 0, \quad i = n+1, n+2, \ldots, m.
\end{aligned}
$$

降阶 SVD

$$A = U_{m\times m}\begin{bmatrix}\Sigma\\ 0\end{bmatrix}V^*_{n\times n} = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots + \sigma_n u_n v_n^* = \sum_{i=1}^{n} \sigma_i u_i v_i^*.$$

降阶 SVD

$$A = U_{m\times m}\begin{bmatrix}\Sigma \\ 0\end{bmatrix} V_{n\times n}^* = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots + \sigma_n u_n v_n^* = \sum_{i=1}^{n} \sigma_i u_i v_i^*.$$

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

$$A = U_n \Sigma V^*. \tag{3.2}$$

这就是所谓的**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).

奇异值的应用

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

奇异值基本性质

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, $i = 1, 2, \ldots, n$;

(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, $i = 1, 2, \ldots, m$;

(3) $\|A\|_2 = \sigma_1$, $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}$;

(4) 若 $\mathrm{rank}(A) = r \leq n$, 则

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

(5) 设 $x \in \mathbb{C}^n$ 且 $\|x\|_2 = 1$, 则 $\sigma_n \leq \|Ax\|_2 \leq \sigma_1$;

(6) (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 $\sigma_i(X^*AY) = \sigma_i(A)$.

(证明留作练习)

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 $\|A^{-1}\|_2 = \sigma_n^{-1}$, $\kappa_2(A) = \sigma_1/\sigma_n$;

(3) 若 A 是 Hermite 的, 且 $A = U\Lambda U^*$ 是 A 的酉特征值分解, 即 $U^*U = I$, $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. 设 $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$, 则 $A = U\Sigma V$ 是 A 的奇异值分解, 其中 $\sigma_i = |\lambda_i|$, $v_i = \mathrm{sign}(\lambda_i)u_i$, 若 $\lambda_i = 0$, 则取 $v_i = u_i$;

(4) 矩阵 $H = \begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}$ 的特征值是 $\pm\sigma_i$, 对应的单位特征向量为 $\dfrac{1}{\sqrt{2}}\begin{bmatrix} v_i \\ \pm u_i \end{bmatrix}$.

($A \in \mathbb{C}^{m\times n}$ 时也有类似结论)

(证明留作练习)

例 由奇异值的性质可知, 矩阵的谱条件数取决于最大奇异值和最小奇异值的比值.

那么, 谱条件数与特征值有没有直接关系?

如果矩阵 A 对称正定, 则其谱条件数就是最大特征值与最小特征值的比值.

但是, 对于一般的矩阵, 则没有这个性质. 如下面的矩阵:

$$A = \begin{bmatrix} 1 & -\frac{1}{2} & \cdots & -\frac{1}{2} \\ & 1 & \ddots & \vdots \\ & & \ddots & -\frac{1}{2} \\ & & & 1 \end{bmatrix} \in \mathbb{R}^{n\times n},$$

即对角线全是 1, 严格上三角部分全是 $-\frac{1}{2}$, 其它为 0.

易知 A 的所有特征值都是 1, 但其谱条件数却随着矩阵规模的增长而快速变大!

n	10	20	30	40	50	60	70	80
$\kappa_2(A)$	6.3e+01	7.6e+03	6.8e+05	5.3e+07	3.9e+09	2.7e+11	1.8e+13	1.2e+15

另外, 通过直接计算可知

$$A^{-1} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1.5}{2} & \frac{1.5^2}{2} & \cdots & \frac{1.5^{n-2}}{2} \\ & 1 & \frac{1}{2} & \frac{1.5}{2} & \ddots & \vdots \\ & & 1 & \ddots & \ddots & \frac{1.5^2}{2} \\ & & & \ddots & \frac{1}{2} & \frac{1.5}{2} \\ & & & & 1 & \frac{1}{2} \\ & & & & & 1 \end{bmatrix}.$$

因此, **$\kappa_1(A)$ 和 $\kappa_\infty(A)$ 都是矩阵维数 n 的幂函数**

SVD 的重要应用: 低秩逼近

定理 设 $A = U_n \Sigma V^*$ 是 $A \in \mathbb{C}^{m\times n}$ 的细奇异值分解. 令 $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^*$, 则 A_k 是

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

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

$$\min_{B\in\mathbb{C}^{m\times n},\ \mathrm{rank}(B)\le k} \|A-B\|_2 \tag{3.4}$$

✍ 对于 Frobenius 范数, 我们有类似的结论.

✍ 如果 $\mathrm{rank}(A)=l\le n$, 则 $\sigma_i=0,\ i=l+1,\dots,n$. 此时

$$A=\sum_{i=1}^{l}\sigma_i u_i v_i^* = A_l. \tag{3.5}$$