

# 从两个视角来看奇异值分解

范潇 2254298

2024 年 6 月 16 日

## 1 引言

奇异值分解是一种功能强大，且适用面广的矩阵分解方法，它能够将任意矩阵  $\mathbf{A}$  分解为  $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ ，其中  $\mathbf{U}, \mathbf{V}$  为正交矩阵， $\mathbf{\Sigma}$  为  $r$  阶对角矩阵 ( $r = \text{Rank}(\mathbf{A})$ )。

本文将分别从代数视角和几何视角来理解奇异值分解。通过代数视角，我们能够理解奇异值分解过程的背后动机。通过几何视角，我们能够从奇异值分解的本质延申开来，得到矩阵的模、矩阵的广义逆、矩阵的极分解这三个概念。

## 2 代数视角

任给一个矩阵  $\mathbf{A}$ ，我们可以立即得到 4 个子空间。

定义. 设  $\mathbf{A}_{m \times n}$  的行向量为  $r_1, \dots, r_m$ ，列向量为  $c_1, \dots, c_n$ 。则

1.  $\mathbf{A}$  的行空间为  $\text{span}\{r_1, \dots, r_m\}$ ;
2.  $\mathbf{A}$  的零空间为  $\mathbf{A}$  的行空间的正交补;
3.  $\mathbf{A}$  的列空间为  $\text{span}\{c_1, \dots, c_n\}$ ;

4.  $\mathbf{A}$  的左零空间为  $\mathbf{A}$  的列空间的正交补。

显然,  $\mathbf{A}$  的行空间与列空间的维数均为  $r = \text{Rank}(\mathbf{A})$ 。奇异值分解便试图分别从  $\mathbf{A}$  的行空间和列空间中取出两组标准正交基  $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ ,  $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ , 使得行空间的基经过  $\mathbf{A}$  的转换后, 能够变为列空间的基。即

$$\mathbf{A}_{m \times n} \mathbf{V}_{n \times r} = \mathbf{U}_{m \times r}$$

当然, 想要求出这样的分解式要求过于苛刻, 我们可以将其修改为

$$\mathbf{A}_{m \times n} \mathbf{V}_{n \times r} = \mathbf{U}_{m \times r} \mathbf{\Sigma}_{r \times r}$$

其中  $\mathbf{\Sigma}$  是一个对角矩阵, 添加了这一矩阵后我们允许两组基的变换过程中额外经过一次伸缩变换。

更进一步, 我们想要让  $\mathbf{V}, \mathbf{U}$  成为两个正交矩阵。此时, 我们便需要利用  $\mathbf{A}$  的零空间和左零空间。从零空间取一组基  $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ , 并将其附加在  $\mathbf{V}$  的右侧; 从左零空间取一组基  $\mathbf{u}_{r+1}, \dots, \mathbf{u}_m$ , 并将其附加在  $\mathbf{U}$  的右侧。由于  $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$  经  $\mathbf{A}$  转换后得到的是零向量, 所以用于伸缩变换的矩阵  $\mathbf{\Sigma}$  中对应的对角元素  $\sigma_{r+1}, \sigma_{r+2}, \dots, \sigma_{\min\{m, n\}}$  为 0。最终我们有分解式

$$\mathbf{A}_{m \times n} \mathbf{V}_{n \times n} = \mathbf{U}_{m \times m} \mathbf{\Sigma}_{m \times n}$$

在确定了分解式的组成后, 便可以利用其组成的特性来反推分解的过程。对于奇异值分解而言, 我们需要利用  $\mathbf{V}, \mathbf{U}$  均为正交矩阵这一特点。有

$$\mathbf{A}^T \mathbf{A} = (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T)^T (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T) = \mathbf{V} \mathbf{\Sigma}^T \mathbf{\Sigma} \mathbf{V}^T,$$

其中  $\mathbf{\Sigma}^T \mathbf{\Sigma}$  是  $r$  阶对角矩阵  $\text{diag}\{\sigma_1^2, \dots, \sigma_r^2\}$ , 实际上  $\sigma_1^2, \dots, \sigma_r^2$  便是  $\mathbf{A}^T \mathbf{A}$  的特征值, 而  $\mathbf{V}$  便是由对应的特征向量组成的矩阵。这样, 我们便确定下来  $\mathbf{V}, \mathbf{\Sigma}$  的获取方式, 剩余的  $\mathbf{U}$  的获取方式便水到渠成。由矩阵乘法可知,

$$u_i = \mathbf{A} \mathbf{v}_i / \sigma_i (i = 1, \dots, r),$$

对于  $u_i (i = r + 1, \dots, m)$ , 则取  $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\}^\perp$  的基。

综上, 通过代数视角, 我们自然而然地得到了奇异值分解的过程:

1. 求  $\mathbf{A}^T \mathbf{A}$  的特征值  $\lambda_1, \dots, \lambda_n$ , 以及对应的特征向量  $\mathbf{v}_1, \dots, \mathbf{v}_n$ ;
2. 令  $\Sigma = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ , 其中  $\mathbf{D} = \text{diag}\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r}\}$ ;
3. 记  $\sigma_1 = \sqrt{\lambda_1}, \dots, \sigma_r = \sqrt{\lambda_r}$ , 令  $u_i = \mathbf{A}v_i/\sigma_i, (i = 1, \dots, r)$ ;
4. 将  $u_i = \mathbf{A}v_i/\sigma_i, (i = 1, \dots, r)$  扩充为标准正交基, 得到  $u_{r+1}, \dots, u_m$ 。
5. 令  $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n), \mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m)$ , 有分解式  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ 。

事实上, 由上述的讨论中可以得出  $\lambda_i = 0, i \geq r + 1$ 。

### 3 几何视角

有两类正交矩阵, 分别对应的是旋转变换和镜像变换, 它们的区别在于: 前者的行列式值为 1, 后者为-1。

对于奇异值分解

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$$

我们并不能够保证  $\mathbf{U}, \mathbf{V}$  是哪一类正交矩阵。例如分解式

$$\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix} \begin{bmatrix} 5 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix},$$

其中,  $\mathbf{U}, \mathbf{V}$  分别为第二类正交矩阵和第一类正交矩阵。又例如分解式

$$\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 5 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix},$$

其中,  $\mathbf{U}, \mathbf{V}$  均为第一类正交矩阵。在本节的讨论重点关注两个正交矩阵都是第一类正交矩阵的情况。

此时,  $\mathbf{A}$  的变换效果可以视为  $\mathbf{V}^T, \Sigma, \mathbf{U}$  三个矩阵的变换效果的依次叠加。 $\mathbf{V}^T = \mathbf{V}^{-1}$  便是  $\mathbf{V}$  的逆变换, 是一个旋转变换;  $\Sigma$  是一个伸缩变换;  $\mathbf{U}$  是一个旋转变换。因此,  $\mathbf{A}$  的变换效果便是“先旋转, 后伸缩, 再旋转”, 如图 1所示。

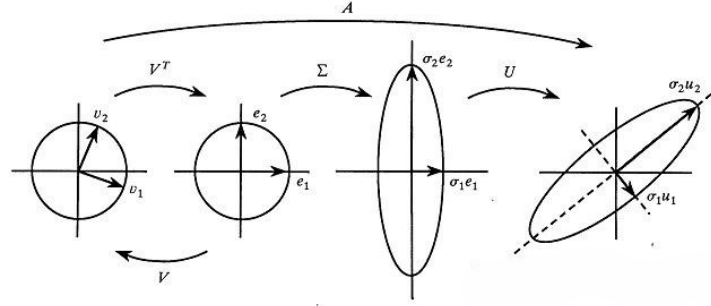


图 1: 几何视角

### 3.1 矩阵的模

任何矩阵  $\mathbf{A}$  都能进行奇异值分解, 无论得到的两个正交矩阵是否都是第一类正交矩阵,  $\mathbf{A}$  对于向量的长度的改变始终是由  $\Sigma$  决定的, 而变化的比例则是由  $\Sigma$  对角线上的元素值决定的。特别地, 变化的最大比例即  $\Sigma$  对角线上的最大元素值。矩阵  $\mathbf{A}$  的模便是从这一角度定义的。

**定义.** 对于任意矩阵  $\mathbf{A}$ , 定义它的模长为

$$\|\mathbf{A}\| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}.$$

**定理 3.1.** 矩阵的模满足正定性, 即

$$\|\mathbf{A}\| \geq 0, \|\mathbf{A} = \mathbf{0}\| \text{ 当且仅当 } \mathbf{A} = \mathbf{0}$$

**证明.** 当  $\mathbf{A} \neq \mathbf{0}$  时,  $\text{Rank}(\mathbf{A}) \neq 0$ , 从而  $\text{Rank}(\mathbf{A}^T \mathbf{A}) \neq 0$ , 因此  $\mathbf{A}^T \mathbf{A}$  至少有一个正特征值, 从而  $\|\mathbf{A}\| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} > 0$ 。

当  $\mathbf{A} = \mathbf{0}$  时, 显然  $\|\mathbf{A}\| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = 0$ 。 □

**定理 3.2.** 矩阵的模满足三角不等式

$$\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|.$$

证明.

$$\begin{aligned}
 \|A + B\| &= \max_{x \neq 0} \frac{\|Ax + Bx\|}{\|x\|} \leq \max_{x \neq 0} \frac{\|Ax\| + \|Bx\|}{\|x\|} \\
 &\leq \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} + \max_{x \neq 0} \frac{\|Bx\|}{\|x\|} \\
 &= \|A\| + \|B\|
 \end{aligned}$$

□

**定理 3.3.** 矩阵的模满足乘积不等式

$$\|AB\| \leq \|A\|\|B\|.$$

证明.

$$\|AB\| = \max_{x \neq 0} \frac{\|ABx\|}{\|x\|} \leq \max_{x \neq 0} \frac{\|A\|\|Bx\|}{\|x\|} = \|A\|\|B\|.$$

□

### 3.2 矩阵的广义逆

由于任何矩阵都可以进行奇异值分解, 该矩阵的变换效果可以视为三个矩阵  $V^T, \Sigma, U$  的变换效果的依次叠加。自然地, 我们可以把逆变换对应的矩阵定义为  $A$  的逆矩阵。因为  $V^T, U$  均可逆, 所以如此定义的广义逆便是

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

其中

$$\Sigma^+ = \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}, D = \text{diag}\{\sigma_1, \dots, \sigma_r\}.$$

事实上, 这样得到的广义逆与 Moore-Penrose 广义逆是一致的。但是通过借助几何视角下的奇异值分解, 这样的广义逆的定义变得更加自然。

### 3.3 矩阵的极分解

奇异值分解将矩阵  $\mathbf{A}$  分解为三个矩阵的乘积，同时也将其变换效果分解为三个矩阵的变换效果的叠加。但理论上，这三次变换的叠加等价于一次旋转变换和一次伸缩变化的叠加。事实上，这就像任一复数  $x + iy$ ，都可以写为极坐标的形式  $re^{i\theta}$ ，可以视为  $(1, 0)$  先经旋转变换旋转了  $\theta$ ，然后再以比例  $r$  进行伸缩。对于矩阵，我们有如下定理：

**定理 3.4.** 对于任意实方阵  $\mathbf{A}$ ，可以将其分解为  $\mathbf{A} = \mathbf{Q}\mathbf{S}$ ，其中  $\mathbf{Q}$  是正交矩阵， $\mathbf{S}$  是对称半正定矩阵。当  $\mathbf{A}$  可逆时， $\mathbf{S}$  是正定矩阵。

证明. 先对  $\mathbf{A}$  进行奇异值分解得到

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

利用等式  $\mathbf{V}^T\mathbf{V} = \mathbf{E}$ ，得到

$$\mathbf{A} = \mathbf{U}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}\mathbf{V}^T = (\mathbf{U}\mathbf{V}^T)(\mathbf{V}\mathbf{\Sigma}\mathbf{V}^T) = \mathbf{Q}\mathbf{S}.$$

因为正交矩阵的乘积仍是正交矩阵，所以  $\mathbf{Q}$  是正交矩阵。当  $\mathbf{A}$  不可逆时，由于  $\text{Rank}(\mathbf{A}^T\mathbf{A}) = \text{Rank}(\mathbf{A})$ ，所以  $\mathbf{A}^T\mathbf{A}$  也不可逆。由于  $|\mathbf{A}^T\mathbf{A}| = \prod_i \sigma_i^2 = 0$ ，所以  $\mathbf{\Sigma}$  对角线上有零值，从而  $\mathbf{S}$  为半正定矩阵；当  $\mathbf{A}$  可逆时，由于  $\text{Rank}(\mathbf{A}^T\mathbf{A}) = \text{Rank}(\mathbf{A})$ ，所以  $\mathbf{A}^T\mathbf{A}$  也可逆。由于  $|\mathbf{A}^T\mathbf{A}| = \prod_i \sigma_i^2 \neq 0$ ，所以  $\mathbf{\Sigma}$  对角线上均为正值，从而  $\mathbf{S}$  为正定矩阵； $\square$

从证明的过程可以看出，极分解中的  $\mathbf{Q}$  实际上便是奇异值分解中  $\mathbf{U}\mathbf{V}^T$  两次旋转变换的叠加。