



第四讲 矩阵的奇异值分解

李宇峰

liyf@nju.edu.cn

人工智能学院



矩阵的奇异值分解

- 矩阵计算和数值分析算法是稳定的吗？
- 长方形矩阵的特征如何分析？
- 矩阵如何逼近？小矩阵逼近大矩阵？



矩阵的奇异值分解

- 数值稳定性与条件数
- 奇异值分解
- 矩阵逼近

数值稳定性和条件数

- 数据处理和数值分析的算法是稳定的么？

为方便计，先假定只存在向量 \mathbf{b} 的扰动 $\delta \mathbf{b}$ ，而矩阵 \mathbf{A} 是稳定不变的。此时，精确的解向量 \mathbf{x} 会扰动为 $\mathbf{x} + \delta \mathbf{x}$ ，即有

$$\mathbf{A}(\mathbf{x} + \delta \mathbf{x}) = \mathbf{b} + \delta \mathbf{b} \quad (4.1.1)$$

这意味着

$$\delta \mathbf{x} = \mathbf{A}^{-1} \delta \mathbf{b} \quad (4.1.2)$$

因为 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 。对式 (4.1.2) 应用矩阵范数的性质，得

$$\|\delta \mathbf{x}\|_2 \leq \|\mathbf{A}^{-1}\|_2 \|\delta \mathbf{b}\|_2 \quad (4.1.3)$$

对线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 也使用矩阵范数的相同性质，又有

$$\|\mathbf{b}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{x}\|_2 \quad (4.1.4)$$

由式 (4.1.3) 和式 (4.1.4)，立即得到

$$\frac{\|\delta \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq (\|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2) \frac{\|\delta \mathbf{b}\|_2}{\|\mathbf{b}\|_2} \quad (4.1.5)$$

数值稳定性和条件数

- 数据处理和数值分析的算法是稳定的么？

然后，考虑扰动 $\delta\mathbf{A}$ 的影响。此时，线性方程组变为

$$(\mathbf{A} + \delta\mathbf{A})(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b}$$

由上式可推导出

$$\begin{aligned}\delta\mathbf{x} &= [(\mathbf{A} + \delta\mathbf{A})^{-1} - \mathbf{A}^{-1}]\mathbf{b} \\ &= \{\mathbf{A}^{-1}[\mathbf{A} - (\mathbf{A} + \delta\mathbf{A})](\mathbf{A} + \delta\mathbf{A})^{-1}\}\mathbf{b} \\ &= -\mathbf{A}^{-1}\delta\mathbf{A}(\mathbf{A} + \delta\mathbf{A})^{-1}\mathbf{b} \\ &= -\mathbf{A}^{-1}\delta\mathbf{A}(\mathbf{x} + \delta\mathbf{x})\end{aligned}\tag{4.1.6}$$

由此得

$$\|\delta\mathbf{x}\|_2 \leq \|\mathbf{A}^{-1}\|_2 \|\delta\mathbf{A}\|_2 \|\mathbf{x} + \delta\mathbf{x}\|_2$$

即有

$$\frac{\|\delta\mathbf{x}\|_2}{\|\mathbf{x} + \delta\mathbf{x}\|_2} \leq (\|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2) \frac{\|\delta\mathbf{A}\|_2}{\|\mathbf{A}\|_2}\tag{4.1.7}$$

数值稳定性和条件数

- 数据处理和数值分析的算法是稳定的么？

式 (4.1.5) 和式 (4.1.7) 表明，解向量 \mathbf{x} 的相对误差与数值

$$\text{cond}(\mathbf{A}) = \|\mathbf{A}\|_2 \cdot \|\mathbf{A}^{-1}\|_2 \quad (4.1.8)$$

成正比。式中， $\text{cond}(\mathbf{A})$ 称为矩阵 \mathbf{A} 的条件数，有时也用符号 $\kappa(\mathbf{A})$ 表示。

当系数矩阵 \mathbf{A} 一个很小的扰动只引起解向量 \mathbf{x} 很小的扰动时，就称矩阵 \mathbf{A} 是“良态”矩阵 (well-conditioned matrix)。若系数矩阵 \mathbf{A} 一个很小的扰动会引起解向量 \mathbf{x} 很大的扰动，则称矩阵 \mathbf{A} 是“病态”矩阵 (ill-conditioned matrix)。条件数刻画了求解线性方程组时，误差经过矩阵 \mathbf{A} 的传播扩大为解向量的误差的程度，因此是衡量线性方程组数值稳定性的一个重要指标。

矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$ 的条件数是该矩阵的最大奇异值 $\sigma_{\max}(\mathbf{A})$ 与最小奇异值 $\sigma_{\min}(\mathbf{A})$ 的比值，即

$$\text{cond}(\mathbf{A}) = \frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})} \quad (4.1.9)$$

数值稳定性和条件数

- 数据处理和数值分析的算法是稳定的么？

由此易知条件数具有以下性质：

(1) 条件数 $\text{cond}(\mathbf{A}) \geq 1$ 。

(2) 正交矩阵或者酉矩阵的条件数等于 1，因为两种矩阵的所有奇异值都等于 1。

(3) 奇异矩阵的条件数为无穷大，因为其最小奇异值等于零。

(4) 若 $\mathbf{A} \in \mathbb{C}^{m \times n}$ 是满列秩矩阵，则 $\text{cond}(\mathbf{A}^H \mathbf{A}) = [\text{cond}(\mathbf{A})]^2$ 。类似地，对于满行秩矩阵 \mathbf{A} ，则 $\text{cond}(\mathbf{A} \mathbf{A}^H) = [\text{cond}(\mathbf{A})]^2$ 。

性质 (2) 表明，用酉矩阵（或者正交矩阵）左乘或者右乘任何一个矩阵，都不会改变该矩阵的条件数。

利用条件数的上述性质，考虑超定方程组 $\mathbf{A} \mathbf{x} = \mathbf{b}$ 的以下两种解：

(1) 最小二乘解 $\mathbf{x}_{\text{LS}} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}$ 与矩阵 \mathbf{A} 的条件数的平方成正比。

(2) 若利用 \mathbf{A} 的 QR 分解 $\mathbf{A} = \mathbf{Q} \mathbf{R}$ 求解超定方程组 $\mathbf{A} \mathbf{x} = \mathbf{b}$ ，其中 \mathbf{Q} 为正交矩阵， \mathbf{R} 为上三角矩阵，则方程的解由 $\mathbf{R} \mathbf{x} = \mathbf{Q}^H \mathbf{b}$ 给出。因此， \mathbf{x} 只受 \mathbf{b} 和 \mathbf{R} 的扰动的影响。由于

$$\text{cond}(\mathbf{Q}) = 1, \quad \text{cond}(\mathbf{A}) = \text{cond}(\mathbf{Q}^H \mathbf{A}) = \text{cond}(\mathbf{R}) \quad (4.1.10)$$

即 \mathbf{R} 与 \mathbf{A} 具有相同的条件数，故 \mathbf{b} 的扰动和 \mathbf{R} 的扰动分别与 \mathbf{A} 的条件数成正比。

矩阵的奇异值分解

奇异值分解 (singular Value decomposition, 简称SVD) 是一种矩阵分解方法。

奇异值分解的概念在理论讨论中同样十分重要：在很多数值代数问题中，我们首先会思考：如果把奇异值分解用进去，会得到什么结果？这往往会使我们解决问题的思路得到开拓。

4.1 奇异值分解

历史： 奇异值分解已经有一百多年的历史。1873年Beltrami从双线性函数 $f(x, y) = x^T A y, A \in R^{n \times n}$ 出发，引入线性变换 $x = U\xi, y = V\eta$ ，这样双线性函数变为 $f(x, y) = \xi^T S \eta$ ，其中

$$S = U^T A V.$$

Beltrami观测到，如果约束 U 和 V 为正交矩阵，则它们的选择各存在 $n^2 - n$ 个自由度。他提出利用这些自由度使矩阵 S 的对角线以外的元素全部为零，即矩阵

$$S = \Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$$

为对角矩阵。于是用 U 和 V^T 分别左乘和右乘式 $S = U^T A V$

4.1 奇异值分解

立即得到

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

这就是Beltrami在1873年得到的实正方矩阵的奇异值分解。1874年Jordan也独立地推导出了实正方矩阵的奇异值分解。

后来，Autonne于1902年把奇异值分解推广到复正方矩阵；**Eckart与Young于1936年又进一步把奇异值分解推广到一般的复长方形矩阵。**

4.1 奇异值分解

定理4.1 (实矩阵的奇异值分解) 令 $A \in R^{m \times n}$, 则存在正交矩阵 $U \in R^{m \times m}$ 和 $V \in R^{n \times n}$ 使得

$$A = U\Sigma V^T \quad (1)$$

其中 $\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}$, 且 $\Sigma_1 = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$, 对角线元素按照从大到小的次序排列, 即

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0, \quad r = \text{rank}(A)$$

数值 $\sigma_1, \sigma_2, \dots, \sigma_r$ 连同 $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_{\min(m,n)} = 0$ 称为 A 的奇异值。

4.1 奇异值分解

定理4.2 (复矩阵的奇异值分解) 令 $A \in \mathbb{C}^{m \times n}$, 则存在酉矩阵

$U \in \mathbb{C}^{m \times m}$ 和 $V \in \mathbb{C}^{n \times n}$ 使得

$$A = U \Sigma V^H \quad (2)$$

其中 $\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}$, 且 $\Sigma_1 = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$, 对角线元素

按照从大到小的次序排列, 即

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0, \quad r = \text{rank}(A)$$

数值 $\sigma_1, \sigma_2, \dots, \sigma_r$ 连同 $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_{\min(m,n)} = 0$ 称为 A 的奇异值。

4.1 奇异值分解

下面是关于奇异值和奇异值分解的几点解释和标记。

(1) $n \times n$ 矩阵 V 为酉矩阵, 用 V 右乘式 (4.2.3), 得 $AV = U\Sigma$, 其列向量形式为

$$Av_i = \begin{cases} \sigma_i u_i, & i = 1, 2, \dots, r \\ \mathbf{0}, & i = r + 1, r + 2, \dots, \min\{m, n\} \end{cases} \quad (4.2.5)$$

因此, V 的列向量 v_i 称为矩阵 A 的右奇异向量 (right singular vector), V 称为 A 的右奇异向量矩阵 (right singular vector matrix)。

(2) $m \times m$ 矩阵 U 是酉矩阵, 用 U^H 左乘式 (4.2.3), 得到 $U^H A = \Sigma V$, 其列向量形式

$$u_i^H A = \begin{cases} \sigma_i v_i^T, & i = 1, 2, \dots, r \\ \mathbf{0}, & i = r + 1, r + 2, \dots, \min\{m, n\} \end{cases} \quad (4.2.6)$$

因此, U 的列向量 u_i 称为矩阵 A 的左奇异向量 (left singular vector), 并称 U 为 A 的左奇异向量矩阵 (left singular vector matrix)。

4.1 奇异值分解

(3) 矩阵 \mathbf{A} 的奇异值分解式 (4.2.3) 可以改写成向量表达形式

$$\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^H \quad (4.2.7)$$

这种表达有时称为 \mathbf{A} 的并向量 (奇异值) 分解 (dyadic decomposition) [46]。

(4) 当矩阵 \mathbf{A} 的秩 $r = \text{rank}(\mathbf{A}) < \min\{m, n\}$ 时, 由于奇异值 $\sigma_{r+1} = \cdots = \sigma_h = 0$, $h = \min\{m, n\}$, 故奇异值分解式 (4.2.3) 可以简化为

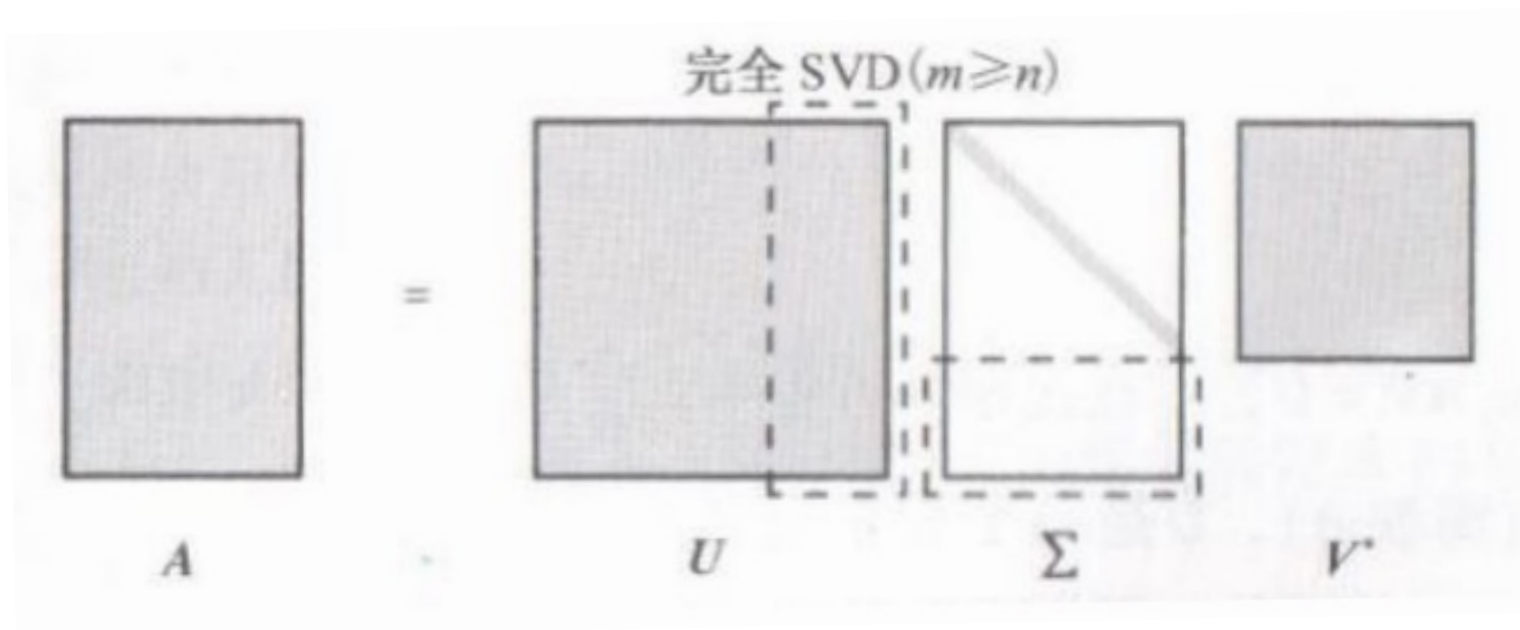
$$\mathbf{A} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^H \quad (4.2.8)$$

式中

$$\mathbf{U}_r = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r], \quad \mathbf{V}_r = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r], \quad \mathbf{\Sigma}_r = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$$

式 (4.2.8) 称为矩阵 \mathbf{A} 的截尾奇异值分解 (truncated SVD) 或薄奇异值分解 (thin SVD)。与之形成对照, 式 (4.2.3) 则称为全奇异值分解 (full SVD)。

4.1 奇异值分解



图中的虚线指出了 U 中“不起作用”的列和 Σ 中“不起作用”部分的行。

4.1 奇异值分解

(5) 用 \mathbf{u}_i^H 左乘式 (4.2.5), 并注意到 $\mathbf{u}_i^H \mathbf{u}_i = 1$, 易得

$$\mathbf{u}_i^H \mathbf{A} \mathbf{v}_i = \sigma_i, \quad i = 1, 2, \dots, \min\{m, n\} \quad (4.2.9)$$

或用矩阵形式写成

$$\mathbf{U}^H \mathbf{A} \mathbf{V} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}, \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix} \quad (4.2.10)$$

(6) 由式 (4.2.3) 易得

$$\mathbf{A} \mathbf{A}^H = \mathbf{U} \boldsymbol{\Sigma}^2 \mathbf{U}^H \quad (4.2.11)$$

这表明, $m \times n$ 矩阵 \mathbf{A} 的奇异值 σ_i 是矩阵乘积 $\mathbf{A} \mathbf{A}^H$ 的特征值 (这些特征值是非负的) 的正平方根。

4.1 奇异值分解

(7) 如果矩阵 $\mathbf{A}_{m \times n}$ 具有秩 r , 则有

- ① $m \times m$ 酉矩阵 \mathbf{U} 的前 r 列组成矩阵 \mathbf{A} 的列空间的标准正交基。
- ② $n \times n$ 酉矩阵 \mathbf{V} 的前 r 列组成矩阵 \mathbf{A} 的行空间或 \mathbf{A}^H 的列空间的标准正交基。
- ③ \mathbf{V} 的后 $n - r$ 列组成矩阵 \mathbf{A} 的零空间的标准正交基。
- ④ \mathbf{U} 的后 $m - r$ 列组成矩阵 \mathbf{A}^H 的零空间的标准正交基。

前两者对应了矩阵的像空间，后两者对应了矩阵的零空间

推论 2.4.6 假设 \mathbf{A} 有 r 个正奇异值, 则 $\text{rank}(\mathbf{A}) = r$ 且

$$\text{null}(\mathbf{A}) = \text{span}\{v_{r+1}, \dots, v_n\},$$

$$\text{ran}(\mathbf{A}) = \text{span}\{u_1, \dots, u_r\}.$$

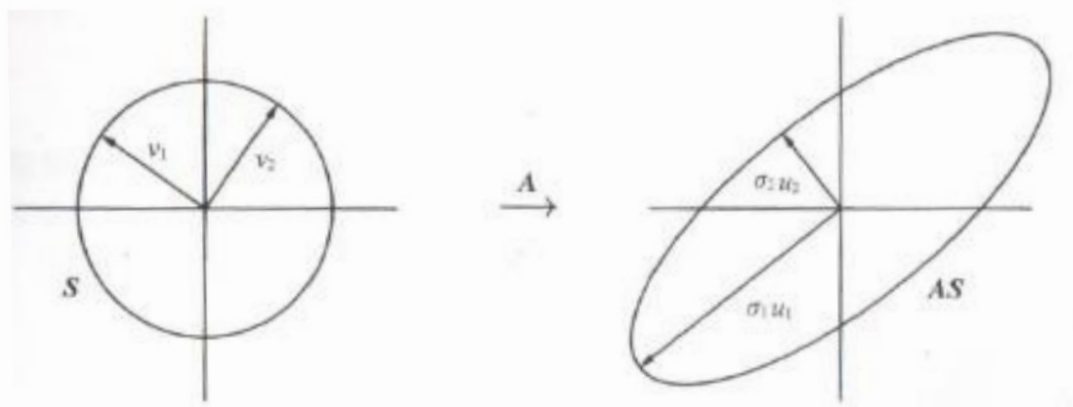
证明 对角矩阵的秩等于主对角线上非零元素的个数, 因此我们有 $\text{rank}(\mathbf{A}) = \text{rank}(\Sigma) = r$. 利用推论 2.4.2 即可得出有关零空间及其值域的断言. \square

奇异值分解的几何解释

我们也可以从几何的观点来看 SVD ，这里其实用到一个基本事实，就是任意一个 $m \times n$ 矩阵将 n 维空间的单位球面映射为 m 维空间一个超球面。

令 S 为 R^n 中的单位球面，取任意的 $A \in R^{m \times n}$ ，其中 $m \geq n$ ，为方便起见，现设 $rank(A) = n$ ，这时 A 的像 AS 是 R^m 中的一个超椭圆，这样可以根据 AS 的形状确定 A 的性质。

我们以下面的示意图说明有关概念。



奇异值分解的几何解释

这是 2×2 矩阵的SVD，对于2维空间的单位球面 S ，给定 2×2 矩阵后， S 经 A 的映射后 AS 就是一个2维椭球，这样长轴 $\sigma_1 u_1$ 和短轴 $\sigma_2 u_2$ 就确定了，而 $\|u_1\|_2 = \|u_2\|_2 = 1$ ，而 $\sigma_1 u_1$ 对应于 S 上的原像为 v_1 ， $\sigma_2 u_2$ 对应于 S 上的原像为 v_2 ，也就是

$$Av_1 = \sigma_1 u_1, Av_2 = \sigma_2 u_2.$$

推广到高维空间，设 $A \in R^{n \times n}$ ， $\text{rank}(A) = n$ ， S 是 R^n 中的单位球面， S 经 A 的映射后，其像 AS 在 R^n 中为一椭球面，它的 n 个主半轴的长度记成 $\sigma_1, \sigma_2, \dots, \sigma_n$ ，习惯上假设奇异值以降序编号， $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$ 。

其次我们定义 AS 的 n 个主半轴方向的单位向量 u_1, u_2, \dots, u_n ，为左奇异向量，其编号对应于奇异值。向量 $\sigma_i u_i$ 就是 AS 的第 i 位大的主半轴。

奇异值的性质

矩阵的各种变形与奇异值的变化有以下关系：

(1) $m \times n$ 矩阵 \mathbf{A} 的共轭转置 \mathbf{A}^H 的奇异值分解为

$$\mathbf{A}^H = \mathbf{V} \mathbf{\Sigma}^T \mathbf{U}^H \quad (4.2.12)$$

即矩阵 \mathbf{A} 和 \mathbf{A}^H 具有完全相同的奇异值。

(2) $\mathbf{A}^H \mathbf{A}$, $\mathbf{A} \mathbf{A}^H$ 的奇异值分解分别为

$$\mathbf{A}^H \mathbf{A} = \mathbf{V} \mathbf{\Sigma}^T \mathbf{\Sigma} \mathbf{V}^H, \quad \mathbf{A} \mathbf{A}^H = \mathbf{U} \mathbf{\Sigma} \mathbf{\Sigma}^T \mathbf{U}^H \quad (4.2.13)$$

其中

$$\mathbf{\Sigma}^T \mathbf{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2, \overbrace{0, \dots, 0}^{n-r}) \quad (4.2.14)$$

$$\mathbf{\Sigma} \mathbf{\Sigma}^T = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2, \overbrace{0, \dots, 0}^{m-r}) \quad (4.2.15)$$

即是说, $\mathbf{A}^H \mathbf{A}$ 和 $\mathbf{A} \mathbf{A}^H$ 的非零奇异值的个数都等于 r (矩阵 \mathbf{A} 的秩), 且数值分别相同, 只是零奇异值的个数不同而已。

奇异值的性质

(3) P 和 Q 分别为 $m \times m$ 和 $n \times n$ 酉矩阵时, PAQ^H 的奇异值分解由

$$PAQ^H = \tilde{U}\Sigma\tilde{V}^H \quad (4.2.16)$$

给出, 其中, $\tilde{U} = PU$, $\tilde{V} = QV$ 。也就是说, 矩阵 PAQ^H 与 A 具有相同的奇异值, 即奇异值具有酉不变性, 但奇异向量不同。

(4) 若 $m \times n$ 矩阵 A 的奇异值分解为 $A = U\Sigma V^H$, 则 $n \times m$ 维 Moore-Penrose 广义逆矩阵 A^\dagger 的奇异值分解为

$$A^\dagger = V\Sigma^\dagger U^H \quad (4.2.17)$$

虽然 U 和 V 相对于 A 不是唯一确定的, 但广义逆矩阵 A^\dagger 是唯一确定的。特别地, 若 A 是一个正方的非奇异矩阵, 则 $A^\dagger = A^{-1}$ 。因此, 在这一情况下, 如果 A 的奇异值是 $\sigma_1, \sigma_2, \dots, \sigma_n$, 那么 A^{-1} 的奇异值就是 $1/\sigma_1, 1/\sigma_2, \dots, 1/\sigma_n$ 。

奇异值的性质

矩阵的奇异值与矩阵范数、行列式、特征值有密切关系：

1. 奇异值与范数的关系

(1) 矩阵 A 的2-范数(谱范数)等于 A 的最大奇异值，即

$$\|A\|_2 = \sigma_1 = \sigma_{max}$$

这是因为 A 的2-范数 $\|A\|_2 = \sqrt{\lambda_{max}(A^H A)} = \sigma_1 = \sigma_{max}$. 如果 A 是实对称矩阵，则 $A^H A = A^2$ 所以

$$\|A\|_2 = \sqrt{\lambda_{max}(A^2)} = |\lambda_{max}(A)|$$

奇异值的性质

(2) 由于 A 的 F -范数 $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$ 是酉不变范数, 即 $\|U^H A V\|_F = \|A\|_F$, 因此就有

$$\begin{aligned}\|A\|_F &= \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \|U^H A V\|_F = \|\Sigma\|_F \\ &= \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}\end{aligned}$$

也就是说任何一个矩阵的 F -范数等于该矩阵所有奇异值的平方和的正平方根。从这里也可以看出 $\|A\|_2 \leq \|A\|_F$.

奇异值的性质

2. 奇异值与行列式的关系

如 A 是 $n \times n$ 的正方矩阵，由于酉矩阵的行列式等于1，所以

$$\begin{aligned} |\det(A)| &= |\det(U\Sigma V^H)| = |\det(U)| \times |\det(V^H)| \times |\det(\Sigma)| \\ &= |\det(\Sigma)| = \sigma_1 \sigma_2 \cdots \sigma_n \end{aligned}$$

这就是说，在 A 的 n 个奇异值中只要有一个奇异值为零，则 A 就是奇异矩阵，即 $|\det(A)| = 0$ ，若奇异值都不等于零，则 A 为非奇异矩阵。

奇异值的性质

3. 奇异值与条件数的关系

对于一个 $m \times n$ 矩阵 \mathbf{A} , 其条件数也可以利用奇异值定义为

$$\text{cond}(\mathbf{A}) = \sigma_1 / \sigma_p, \quad p = \min\{m, n\} \quad (4.2.21)$$

由定义式 (4.2.21) 可以看出, 条件数是一个大于或等于 1 的正数, 因为 $\sigma_1 \geq \sigma_p$ 。显然, 由于奇异矩阵至少有一个奇异值 $\sigma_p = 0$, 故其条件数为无穷大。而条件数虽然不是无穷大, 但很大时, 就称 \mathbf{A} 是接近奇异的。这意味着, 当条件数很大时, \mathbf{A} 的行向量或列向量的线性相关性很强。另由定义式 (4.1.8) 易知, 正交或酉矩阵 \mathbf{V} 的条件数等于 1。从这个意义上讲, 正交或酉矩阵是“理想条件”的。式 (4.2.21) 也可用作条件数 $\text{cond}(\mathbf{A})$ 的评价。

奇异值的性质

3. 奇异值与特征值的关系

讲特征值只能是正方形矩阵

(1) 设 A 是 $n \times n$ 的实对称矩阵, 且其特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$ ($|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$), 奇异值为 $\sigma_1, \sigma_2, \dots, \sigma_n$ ($|\sigma_1| \geq |\sigma_2| \geq \dots \geq |\sigma_n|$), 则 $\sigma_i = |\lambda_i|, i = 1, 2, \dots, n$.

(2) 如果 A 为 $m \times n$ 矩阵, $p = \min\{m, n\}$, 且 $\sigma_1, \sigma_2, \dots, \sigma_p$ 是 A 的奇异值, 则

$$\text{tr}(A^H A) = \sum_i^p \sigma_i^2$$

(3) A 的谱范数 $\|A\|_2 = \sigma_{max}$, 矩阵 A 的谱条件数

$$\text{cond}_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_{max}}{\sigma_{min}}$$

奇异值的性质

奇异值被矩阵E扰动的分析

我们将在 8.6 节证明，如果矩阵 A 被矩阵 E 扰动，那么奇异值的变化不会超过 $\|E\|_2$ 。以下推论确定了此结果的两个有用实例。

推论 2.4.4 假设 $A \in \mathbb{R}^{m \times n}$ 且 $E \in \mathbb{R}^{m \times n}$ ，则

$$\sigma_{\max}(A + E) \leq \sigma_{\max}(A) + \|E\|_2,$$

$$\sigma_{\min}(A + E) \geq \sigma_{\min}(A) - \|E\|_2.$$

证明 利用推论 2.4.2 易证

$$\sigma_{\min}(A) \cdot \|x\|_2 \leq \|Ax\|_2 \leq \sigma_{\max}(A) \cdot \|x\|_2.$$

由此得到所需的不等式。 □

推论 2.4.2 假设 $U^T A V = \Sigma$ 是 $A \in \mathbb{R}^{m \times n}$ 的 SVD，且 $m \geq n$ ，则对于 $i = 1:n$ 有 $A v_i = \sigma_i u_i$ 和 $A^T u_i = \sigma_i v_i$ 。

奇异值的性质

奇异值被矩阵E扰动的分析

推论 2.4.5 假设 $A \in \mathbb{R}^{m \times n}$, $m > n$, $z \in \mathbb{R}^m$, 则

$$\sigma_{\max}([A \mid z]) \geq \sigma_{\max}(A),$$

$$\sigma_{\min}([A \mid z]) \leq \sigma_{\min}(A).$$

证明 设 $A = U\Sigma V^T$ 是 A 的 SVD, 令 $x = V(:, 1)$, $\tilde{A} = [A \mid z]$. 利用推论 2.4.4, 我们有

$$\sigma_{\max}(A) = \|Ax\|_2 = \left\| \tilde{A} \begin{bmatrix} x \\ 0 \end{bmatrix} \right\|_2 \leq \sigma_{\max}(\tilde{A}).$$

同理可证 $\sigma_{\min}(A) \geq \sigma_{\min}(\tilde{A})$.

4.2 矩阵的低秩逼近(low rank approximation)

理论研究价值

- **深化矩阵理论理解**：矩阵逼近是矩阵理论的重要组成部分，学习它有助于深入理解矩阵的性质和结构。通过研究如何用一个“较好”的矩阵去逼近另一个矩阵，可以更清晰地认识矩阵之间的关系以及矩阵空间的特性，如通过奇异值分解来进行矩阵逼近，能让我们从奇异值和奇异向量的角度理解矩阵的内在结构。
- **为其他数学理论提供工具**：矩阵逼近在泛函分析、数值分析等数学领域有重要地位。在泛函分析中，矩阵逼近可用于研究算子的性质和空间的结构；在数值分析中，是许多数值算法的理论基础，如在求解线性方程组的迭代法中，矩阵逼近理论能帮助分析迭代矩阵的性质，进而判断算法的收敛性和稳定性。

实际应用价值

- **数据处理与分析**
 - **数据降维**：在大数据分析中，数据常以高维矩阵形式存在，维度高会导致计算复杂、存储困难及“维数灾难”等问题。矩阵逼近可通过低秩逼近等方法，用低维矩阵近似高维矩阵，在保留关键信息的同时降低数据维度，如主成分分析（PCA）利用矩阵逼近实现数据降维，广泛应用于图像识别、生物信息学等领域。
 - **噪声处理**：实际数据常含噪声，影响后续分析和处理。矩阵逼近可用于去除噪声，如通过对含噪数据矩阵进行奇异值阈值处理等逼近方法，将小奇异值对应的成分视为噪声并去除，从而提高数据质量，在信号处理、医学影像等领域有重要应用。

4.2 矩阵的低秩逼近(low rank approximation)

- 优化问题求解

- **约束优化**: 在许多工程和科学问题中, 常遇到约束条件下的优化问题, 目标函数和约束条件可能涉及矩阵。矩阵逼近可将原问题转化为更易处理的逼近问题, 如在二次规划问题中, 若目标函数的系数矩阵不满足某些条件, 可通过矩阵逼近构造一个满足条件的近似矩阵, 进而求解优化问题。
- **非凸优化**: 对于非凸的矩阵优化问题, 直接求解困难, 矩阵逼近可提供一种有效的求解策略, 如通过交替方向乘子法 (ADMM) 等方法, 利用矩阵逼近将非凸问题分解为多个较易处理的子问题, 逐步逼近最优解, 在机器学习的模型训练等方面有广泛应用。

- 模型简化与近似

- **系统建模**: 在控制理论和系统工程中, 建立系统的数学模型时, 可能得到复杂的高阶矩阵模型, 不利于分析和控制。矩阵逼近可用于将高阶模型简化为低阶近似模型, 在保持系统主要动态特性的前提下, 降低模型复杂度, 便于控制器设计和系统分析。
- **数值模拟**: 在计算流体力学、有限元分析等数值模拟领域, 求解大规模矩阵方程组计算成本高。矩阵逼近可构造近似矩阵, 在保证计算精度的同时减少计算量, 提高计算效率, 如采用不完全 LU 分解等矩阵逼近方法加速迭代求解过程。

4.2 矩阵的低秩逼近(low rank approximation)

矩阵的低秩逼近:

假如 A 是一个 $m \times n$ 矩阵, 设 $\text{rank}(A) = p$, 低秩逼近问题就是要在所有的 $m \times n$ 矩阵类中, 确定一个 rank 为 $k < p$ 的 $m \times n$ 矩阵 A_k , 使其是在一定的度量意义下与 A 最接近。

$$\begin{aligned} & \min_{B \in R^{m \times n}} \|A - B\|_2 \\ & s. t. \text{rank}(B) = k < p \end{aligned}$$

4.2 矩阵的低秩逼近(low rank approximation)

定理 (The Eckart-Young Theorem): 设 $A \in R^{m \times n}$, $rank(A) = p$,
 A 的奇异值分解为

$$A = \sum_{i=1}^p \sigma_i u_i v_i^T$$

令 $k < p$, 则其低秩逼近矩阵为

$$A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$$

逼近质量为

$$\min_{B \in R^{m \times n}, rank(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$$

4.2 矩阵的低秩逼近(low rank approximation)

证明 由于 $U^T A_k V = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k, 0, \dots, 0)$, 故推得

$\text{rank}(A_k) = k$, 而 $U^T (A - A_k) V = \text{diag}(0, \dots, 0, \sigma_{k+1}, \dots, \sigma_p, 0, \dots, 0)$. 因此 $\|A - A_k\|_2 = \sigma_{k+1}$.

下面证明对任何 $B \in R^{m \times n}$, $\text{rank}(B) = k$, 都有 $\|A - B\|_2 \geq \sigma_{k+1}$.

假设 $B \in R^{m \times n}$, $\text{rank}(B) = k$. 这就推得 B 的零空间 $N(B)$ 的维数为 $n-k$, 设 x_1, x_2, \dots, x_{n-k} 是 $N(B)$ 的一组标准正交基, 即

$$N(B) = \text{span}\{x_1, x_2, \dots, x_{n-k}\}$$

4.2 矩阵的低秩逼近(low rank approximation)

由于

$$\dim(\text{span}\{x_1, x_2, \dots, x_{n-k}\}) + \dim(\text{span}\{v_1, v_2, \dots, v_{k+1}\}) > n,$$

所以 $\text{span}\{x_1, x_2, \dots, x_{n-k}\} \cap \text{span}\{v_1, v_2, \dots, v_{k+1}\} \neq \{0\}$.

设 $z \in \text{span}\{x_1, x_2, \dots, x_{n-k}\} \cap \text{span}\{v_1, v_2, \dots, v_{k+1}\}$, 且 $\|z\|_2 = 1$.

由于 $Bz = 0$, $(A - B)z = Az$, 又由于 $z \in \text{span}\{v_1, v_2, \dots, v_{k+1}\}$

故 z 可以表示成

$$\begin{aligned} z &= \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_{k+1} v_{k+1} \\ &= (v_1^T z) v_1 + (v_2^T z) v_2 + \dots + (v_{k+1}^T z) v_{k+1} \end{aligned}$$

4.2 矩阵的低秩逼近(low rank approximation)

因此

$$\begin{aligned}Az &= A[(v_1^T z)v_1 + (v_2^T z)v_2 + \cdots + (v_{k+1}^T z)v_{k+1}] \\ &= \sum_{i=1}^{k+1} \sigma_i (v_i^T z) u_i\end{aligned}$$

$$\|A - B\|_2^2 \geq \|(A - B)z\|_2^2 = \|Az\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 (v_i^T z)^2 \geq \sigma_{k+1}^2$$

这就证明了 A_k 是使得满足定理条件的低秩逼近矩阵。

4.2 矩阵的低秩逼近(low rank approximation)

前面讨论的是2-范数，对于F-范数，我们有

定理 对任意满足 $0 \leq k \leq p = \text{rank}(A)$, 上述给出的 A_k 也满足

$$\|A - A_k\|_F = \min_{B \in \mathbb{R}^{m \times n}, \text{rank}(B)=k} \|A - B\|_F = \sqrt{\sigma_{k+1}^2 + \dots + \sigma_p^2}$$

这些用低秩的矩阵 A_k 去逼近一个有噪声或者扰动的矩阵A时，其逼近的误差由那些被舍弃的小的奇异值的平方和的平方根来确定。

4.2 矩阵的低秩逼近(low rank approximation)

那么，如何确定哪些奇异值是小到可以舍弃的呢？换言之，使用多大的低秩逼近一个矩阵才最合适呢？这涉及一个矩阵的“有效秩”的确定。

有效秩确定有以下两种常用方法。

1. 归一化奇异值方法

计算归一化奇异值

$$\bar{\sigma}_i = \frac{\hat{\sigma}_i}{\hat{\sigma}_1} \quad (4.2.24)$$

选择满足准则

$$\bar{\sigma}_i \geq \epsilon \quad (4.2.25)$$

的最大整数作为有效秩的估计值 \hat{r} 。显然，这一准则等价于选择满足

$$\hat{\sigma}_i \geq \epsilon \hat{\sigma}_1 \quad (4.2.26)$$

的最大整数 \hat{r} 。式中， ϵ 是某个很小的正数，它根据计算机精度与 (或) 数据精度选取。例如，选取 $\epsilon = 0.01$ 或者 $\epsilon = 0.05$ 等。这相当于将那些小于或者等于最大奇异值的 1% 或者 5% 的奇异值舍弃。

4.2 矩阵的低秩逼近(low rank approximation)

2. 范数比方法

令 $m \times n$ 矩阵 \mathbf{A}_k 是原 $m \times n$ 矩阵 \mathbf{A} 的秩 k 近似, 定义该近似矩阵与原矩阵的 Frobenius 范数比为

$$\nu(k) = \frac{\|\mathbf{A}_k\|_F}{\|\mathbf{A}\|_F} = \frac{\sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_k^2}}{\sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_q^2}}, \quad q = \min\{m, n\} \quad (4.2.27)$$

并选择满足

$$\nu(k) \geq \alpha \quad (4.2.28)$$

的最大整数作为有效秩估计 \hat{r} , 其中, α 是接近于 1 的阈值, 例如 $\alpha = 0.997$ 等。

采用以上两种准则确定出有效秩 \hat{r} 后, 可将

$$\hat{\mathbf{x}}_{\text{LS}} = \sum_{i=1}^{\hat{r}} (\hat{\mathbf{u}}_i^H \mathbf{b} / \hat{\sigma}_i) \hat{\mathbf{v}}_i \quad (4.2.29)$$

看作是真实最小二乘解 \mathbf{x}_{LS} 的一个合理近似。显而易见, 这种解就是线性方程组 $\mathbf{A}_{\hat{r}} \mathbf{x} = \mathbf{b}$ 的最小二乘解, 其中

$$\mathbf{A}_{\hat{r}} = \sum_{i=1}^{\hat{r}} \sigma_i \mathbf{u}_i \mathbf{v}_i^H \quad (4.2.30)$$

4.2 矩阵的低秩逼近(low rank approximation)

例 计算以下 10×8 矩阵 A 的SVD, $rank(A) = 8$.

A =

4.9304	-0.1289	-0.1778	-0.2162	-0.2443	-0.2620	-0.3057	-0.3494
-0.1197	3.7814	-0.2967	-0.3541	-0.3906	-0.4064	-0.4742	-0.5419
-0.1501	-0.2691	2.6432	-0.4134	-0.4389	-0.4332	-0.5053	-0.5775
-0.1609	-0.2803	-0.3581	1.6056	-0.3891	-0.3422	-0.3993	-0.4563
-0.1522	-0.2524	-0.3006	-0.2969	0.7587	-0.1337	-0.1560	-0.1782
-0.1238	-0.1852	-0.1843	-0.1210	0.0046	0.1926	0.2246	0.2567
-0.1444	-0.2160	-0.2150	-0.1412	0.0053	0.2246	0.2621	0.2995
-0.1650	-0.2469	-0.2457	-0.1613	0.0061	0.2567	0.2995	0.3423
-0.1856	-0.2778	-0.2764	-0.1815	0.0069	0.2888	0.3369	0.3850
-0.2063	-0.3086	-0.3071	-0.2017	0.0077	0.3209	0.3743	0.4278

4.2 矩阵的低秩逼近(low rank approximation)

由SVD分解得到, $A = USV^T$:

U =

0.9948	-0.0104	-0.0156	-0.0208	-0.0260	-0.0312	-0.0364	-0.0416	-0.0468	-0.0519
-0.0104	0.9792	-0.0312	-0.0416	-0.0519	-0.0623	-0.0727	-0.0831	-0.0935	-0.1039
-0.0156	-0.0312	0.9532	-0.0623	-0.0779	-0.0935	-0.1091	-0.1247	-0.1403	-0.1558
-0.0208	-0.0416	-0.0623	0.9169	-0.1039	-0.1247	-0.1455	-0.1662	-0.1870	-0.2078
-0.0260	-0.0519	-0.0779	-0.1039	0.8701	-0.1558	-0.1818	-0.2078	-0.2338	-0.2597
-0.0312	-0.0623	-0.0935	-0.1247	-0.1558	0.8130	-0.2182	-0.2494	-0.2805	-0.3117
-0.0364	-0.0727	-0.1091	-0.1455	-0.1818	-0.2182	0.7455	-0.2909	-0.3273	-0.3636
-0.0416	-0.0831	-0.1247	-0.1662	-0.2078	-0.2494	-0.2909	0.6675	-0.3740	-0.4156
-0.0468	-0.0935	-0.1403	-0.1870	-0.2338	-0.2805	-0.3273	-0.3740	0.5792	-0.4675
-0.0519	-0.1039	-0.1558	-0.2078	-0.2597	-0.3117	-0.3636	-0.4156	-0.4675	0.4805

V =

0.9902	-0.0196	-0.0294	-0.0392	-0.0490	-0.0588	-0.0686	-0.0784
-0.0196	0.9608	-0.0588	-0.0784	-0.0980	-0.1176	-0.1373	-0.1569
-0.0294	-0.0588	0.9118	-0.1176	-0.1471	-0.1765	-0.2059	-0.2353
-0.0392	-0.0784	-0.1176	0.8431	-0.1961	-0.2353	-0.2745	-0.3137
-0.0490	-0.0980	-0.1471	-0.1961	0.7549	-0.2941	-0.3431	-0.3922
-0.0588	-0.1176	-0.1765	-0.2353	-0.2941	0.6471	-0.4118	-0.4706
-0.0686	-0.1373	-0.2059	-0.2745	-0.3431	-0.4118	0.5196	-0.5490
-0.0784	-0.1569	-0.2353	-0.3137	-0.3922	-0.4706	-0.5490	0.3725

4.2 矩阵的低秩逼近(low rank approximation)

$$S = \text{diag}(5, 4, 3, 2, 1, 0.000067, 0.000045, 0.000021)$$

由于奇异值对角矩阵 S 从第6个奇异值开始突减，故可以认为是噪音引起，可以用低秩逼近，取 $k = 5$ ，取矩阵 U 的前5列为 $U_{11} \in R^{10 \times 5}$ 的列正交矩阵、矩阵 V 的前5列为 $V_{11} \in R^{8 \times 5}$ 的列正交矩阵。

U11 =

-0.9948	0.0104	-0.0156	-0.0208	0.0260
0.0104	-0.9792	-0.0312	-0.0416	0.0519
0.0156	0.0312	0.9532	-0.0623	0.0779
0.0208	0.0416	-0.0623	0.9169	0.1039
0.0260	0.0519	-0.0779	-0.1039	-0.8701
0.0312	0.0623	-0.0935	-0.1247	0.1558
0.0364	0.0727	-0.1091	-0.1455	0.1818
0.0416	0.0831	-0.1247	-0.1662	0.2078
0.0468	0.0935	-0.1403	-0.1870	0.2338
0.0519	0.1039	-0.1558	-0.2078	0.2597

S11 =

5.0000	0	0	0	0
0	4.0000	0	0	0
0	0	3.0000	0	0
0	0	0	2.0000	0
0	0	0	0	1.0000

4.2 矩阵的低秩逼近(low rank approximation)

$A_5 = AA = U_{11} \times S_{11} \times V_{11}^T$, 这是所有秩为5的10*8的矩阵中欧氏长度意义下A的最佳逼近矩阵。

AA =

4.9304	-0.1289	-0.1778	-0.2162	-0.2443	-0.2620	-0.3057	-0.3494
-0.1197	3.7814	-0.2967	-0.3541	-0.3906	-0.4064	-0.4742	-0.5419
-0.1501	-0.2691	2.6432	-0.4134	-0.4389	-0.4332	-0.5053	-0.5775
-0.1609	-0.2803	-0.3581	1.6056	-0.3891	-0.3422	-0.3993	-0.4563
-0.1522	-0.2524	-0.3006	-0.2969	0.7587	-0.1337	-0.1560	-0.1783
-0.1238	-0.1852	-0.1843	-0.1210	0.0046	0.1925	0.2246	0.2567
-0.1444	-0.2160	-0.2150	-0.1412	0.0053	0.2246	0.2620	0.2995
-0.1650	-0.2469	-0.2457	-0.1613	0.0061	0.2567	0.2995	0.3422
-0.1856	-0.2778	-0.2764	-0.1815	0.0069	0.2888	0.3369	0.3850
-0.2063	-0.3086	-0.3071	-0.2017	0.0076	0.3209	0.3743	0.4278

A =

4.9304	-0.1289	-0.1778	-0.2162	-0.2443	-0.2620	-0.3057	-0.3494
-0.1197	3.7814	-0.2967	-0.3541	-0.3906	-0.4064	-0.4742	-0.5419
-0.1501	-0.2691	2.6432	-0.4134	-0.4389	-0.4332	-0.5053	-0.5775
-0.1609	-0.2803	-0.3581	1.6056	-0.3891	-0.3422	-0.3993	-0.4563
-0.1522	-0.2524	-0.3006	-0.2969	0.7587	-0.1337	-0.1560	-0.1783
-0.1238	-0.1852	-0.1843	-0.1210	0.0046	0.1926	0.2246	0.2567
-0.1444	-0.2160	-0.2150	-0.1412	0.0053	0.2246	0.2621	0.2995
-0.1650	-0.2469	-0.2457	-0.1613	0.0061	0.2567	0.2995	0.3423
-0.1856	-0.2778	-0.2764	-0.1815	0.0069	0.2888	0.3369	0.3850
-0.2063	-0.3086	-0.3071	-0.2017	0.0077	0.3209	0.3743	0.4278

4.2 矩阵的低秩逼近(low rank approximation)

3. 数值计算误差

由于奇异值分解采取了数值计算的方法, 所以一定存在计算值与真实值之间的误差。在数值算法中, 迭代或者是计算停止的时刻都是通过设定一个误差门限来得到的: 如果在计算过程中的某一步, 实际值与预期值的差的绝对值小于一个特定的门限, 则认为已经得到了最终结果。SVD 数值计算算法的误差主要来自两个方面: 停止计算的误差门限产生的截断误差以及计算机的存储误差。下面仅对 Linpack 中的 SVD 算法进行简单的分析:

(1) 在整个算法过程中, 若某个次对角线元素 e_j 满足 $|e_j| \leq \epsilon(|s_{j+1}| + |s_j|)$, 则认为 e_j 为 0。若对角线元素 s_j 满足 $|s_j| \leq \epsilon(|e_{j-1}| + |e_j|)$, 则认为 s_j 为 0 (即为零奇异值)。其中, ϵ 为给定的精度要求。根据 Householder 和 Givens 变换可以大致估算这种奇异值分解算法存在的误差为 $O(\epsilon^2)$ 量级。

(2) 假定计算过程中基本运算带来的舍入误差为零, 从算法可以看出, 矩阵元素在存储过程中产生的舍入误差在迭代运算过程中并不会被放大。因此, 在对结果精度要求远低于计算机存储精度时, 舍入误差的影响可以忽略不计。

综上所述, 可以看出 SVD 算法在计算矩阵奇异值分解的过程中产生的误差主要来自误差限的舍入误差, 误差大致为 $O(\epsilon^2)$ 量级。

4.2 矩阵的低秩逼近(low rank approximation)

第 10 章 大型稀疏特征值问题	503
10.1 对称 Lanczos 方法	504
10.2 Lanczos 方法、求积和近似	513
10.3 实用 Lanczos 方法	519
10.4 大型稀疏 SVD 方法	527
10.5 非对称问题的 Krylov 方法	534
10.6 Jacobi-Davidson 方法及相 关方法	543

展开阅读、思考分析、实践~

矩阵的奇异值分解

- **数据压缩**

- 在图像、视频等数据处理中，奇异值分解可以用于数据压缩。图像或视频数据通常可以表示为矩阵形式，通过对矩阵进行奇异值分解，可以将矩阵分解为三个矩阵的乘积，其中奇异值代表了矩阵的重要特征信息。只保留较大的奇异值及其对应的奇异向量，而忽略较小的奇异值，就可以在损失较少信息的情况下对数据进行压缩。
- 例如，在图像压缩中，将图像的像素矩阵进行奇异值分解后，根据设定的压缩比，保留一定数量的较大奇异值，然后利用这些保留的信息来重构图像。这样可以大大减少数据存储量和传输带宽，同时仍然能够保持图像的主要视觉特征，使得在有限的存储空间和传输速度下能够高效地处理图像数据。

- **矩阵近似**

- 对于一个给定的矩阵，奇异值分解提供了一种最佳的低秩近似方法。在很多实际问题中，数据矩阵往往是高维且复杂的，但可能存在一种低秩结构，即矩阵的大部分信息可以由少数几个主要成分来表示。
- 例如，在推荐系统中，用户对商品的评分矩阵通常是稀疏且高维的。通过奇异值分解，可以将这个评分矩阵近似为一个低秩矩阵，其中低秩矩阵的行和列分别对应于用户和商品的潜在特征。这种低秩近似能够捕捉到用户和商品之间的主要关系，去除噪声和冗余信息，从而实现更准确的推荐。

矩阵的奇异值分解

- **特征提取与降维**

- 在机器学习和数据分析中，奇异值分解常用于特征提取和降维。对于高维数据矩阵，奇异值分解可以将其分解为奇异值和奇异向量的形式，奇异值的大小反映了数据在相应奇异向量方向上的方差大小，而方差较大的方向通常包含了数据的主要特征。
- 例如，在主成分分析（PCA）中，通过对数据矩阵进行奇异值分解，可以找到数据的主成分，即对应于较大奇异值的奇异向量。这些主成分构成了一个新的特征空间，能够在保留数据主要信息的同时降低数据的维度，从而减少计算量、提高模型训练速度，并有助于发现数据中的潜在结构和规律。

- **求解线性方程组**

- 在数值计算中，奇异值分解为求解线性方程组提供了一种稳定且有效的方法。对于线性方程组，其中是系数矩阵，是未知向量，是常数向量。当是病态矩阵（即矩阵的条件数很大，对输入数据的微小变化非常敏感）时，直接求解可能会导致数值不稳定和误差较大。
- 利用奇异值分解，可以更稳定地求解，特别是对于存在噪声或数据不准确的情况，奇异值分解能够有效地抑制噪声的影响，得到更可靠的解。

矩阵的奇异值分解

- 数值稳定性与条件数
- 奇异值分解
- 矩阵逼近