

计算方法

Computational Methods

Lecture 8: 奇异值分解与低秩逼近

Singular Value Decomposition and Low-Rank Approximation

刘景铎

南京大学 · 计算机学院

课程信息

计算方法 Fall 2026

授课教师 刘景铎 liu@nju.edu.cn

上课 周四 5-6 节 · 1-18 周 · 仙 I-319

答疑 周四下午 4 点

助教 周宇恒 · 侯哲

作业 nm_nju_2026@163.com

课程主页 TCS Wiki · Fall 2026

QQ 群 1054896504

入群备注: 姓名、专业、学号

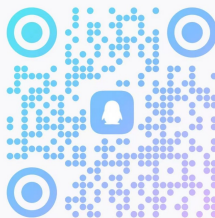
教材: Timothy Sauer, 《数值分析》(原书第 2 版)。

最新通知以课程主页为准。



计算方法 Fall 2026

群号: 1054896504



扫一扫二维码, 加入群聊



回顾：第七讲的谱定理

对实对称矩阵 $B \in \mathbb{R}^{n \times n}$ ，存在正交矩阵 V 和实对角矩阵 Λ ，使得

$$B = V\Lambda V^\top, \quad Bv_i = \lambda_i v_i.$$

如果 B 还满足 $B \succeq 0$ ，则所有 $\lambda_i \geq 0$ 。

上节课的结论：对称矩阵有正交的特征向量基，且 Rayleigh 商能够刻画其最大、最小特征值。

问题：一般的矩阵 $A \in \mathbb{R}^{m \times n}$ 可能不是方阵，也不对称；还能找到这样的几何分解吗？

为什么研究一般矩阵的分解？

在数据分析与数值计算中，矩阵往往**不是方阵**：

$$A \in \mathbb{R}^{m \times n}, \quad x \in \mathbb{R}^n, \quad Ax \in \mathbb{R}^m.$$

几何问题

单位球面经 $x \mapsto Ax$ 会变成什么形状？
哪些输入方向被放大得最多？

近似问题

能否用低秩矩阵 B 近似 A ？
如何量化舍弃的信息？

共同的答案：奇异值分解（Singular Value Decomposition, SVD）。

从第六讲的条件数说起

回顾：对任意 $A \in \mathbb{R}^{m \times n}$ ，诱导二范数定义为

$$\|A\|_2 := \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

等价地，寻找单位向量 x 使 $\|Ax\|_2$ 最大：

$$\|Ax\|_2^2 = x^\top A^\top A x, \quad \|x\|_2 = 1.$$

关键观察： $A^\top A$ 是**实对称且半正定**的矩阵，可直接使用第七讲的谱定理。

今天从这里出发，得到 A 在所有方向上的伸缩规律。

定理：奇异值分解

SVD 定理：任意实矩阵 $A \in \mathbb{R}^{m \times n}$ ，都可写为

$$A = U \Sigma V^{\top},$$

其中 $U \in \mathbb{R}^{m \times m}$ 、 $V \in \mathbb{R}^{n \times n}$ 为正交矩阵； $\Sigma \in \mathbb{R}^{m \times n}$ 仅在主对角线上有非负元素。

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0, \quad p = \min(m, n).$$

这些 σ_i 称为 A 的**奇异值**； U 、 V 的列称为左、右奇异向量。**接下来证明：**为何一定存在这样的分解？

关键矩阵: $A^\top A$ 与 $AA^\top$

$$(A^\top A)^\top = A^\top A, \quad x^\top A^\top A x = \|Ax\|_2^2 \geq 0.$$

因此存在正交矩阵 V , 使

$$A^\top A = V \operatorname{diag}(\lambda_1, \dots, \lambda_n) V^\top, \quad \lambda_i \geq 0.$$

同样, $AA^\top$ 是 $m \times m$ 的实对称半正定矩阵。

定义: 非零特征值 λ_i 的平方根 $\sigma_i = \sqrt{\lambda_i}$, 将成为 A 的正奇异值。

但为何 $A^\top A$ 与 $AA^\top$ 具有相同的正特征值?

共同的非零特征值：一个方向

设 $AA^\top u = \lambda u$, $\lambda > 0$, 且 $u \neq 0$ 。则

$$A^\top A(A^\top u) = A^\top (AA^\top u) = \lambda A^\top u.$$

并且 $A^\top u \neq 0$ ：否则

$$0 = A(A^\top u) = AA^\top u = \lambda u,$$

与 $\lambda > 0$ 矛盾。

因此 λ 也是 $A^\top A$ 的特征值，其特征向量可取 $A^\top u$ 。

想一想：反过来从 $A^\top A$ 的正特征值出发呢？

共同的非零特征值：反方向

设 $A^\top Av = \lambda v$, $\lambda > 0$, $v \neq 0$ 。于是

$$AA^\top(Av) = A(A^\top Av) = \lambda Av.$$

同理 $Av \neq 0$, 因为否则 $A^\top Av = 0$, 与 $\lambda v \neq 0$ 矛盾。

结论： $A^\top A$ 与 $AA^\top$ 的**正特征值**相同，且具有相同的重数。

为什么重数也相同？ 映射

$$v \mapsto Av, \quad u \mapsto A^\top u$$

分别将两侧的 λ -特征空间映到另一侧；两次映射复合都等于 λI ，因此它们是同构，维数（重数）相同。

由右奇异向量构造左奇异向量

令 $r = \text{rank}(A)$ 。对 $A^\top A$ 的正特征值选择标准正交特征向量 $v_1, \dots, v_r$ ：

$$A^\top A v_i = \sigma_i^2 v_i, \quad \sigma_i > 0.$$

定义

$$u_i := \frac{A v_i}{\sigma_i} \quad (1 \leq i \leq r).$$

由此立即得到

$$A v_i = \sigma_i u_i, \quad A^\top u_i = \sigma_i v_i.$$

剩下要证： u_i 为什么两两正交、且都为单位向量？

构造出的向量为何标准正交？

用 $A^\top Av_j = \sigma_j^2 v_j$ 和 $v_i^\top v_j = \delta_{ij}$ ：

$$u_i^\top u_j = \frac{v_i^\top A^\top Av_j}{\sigma_i \sigma_j} = \frac{\sigma_j^2}{\sigma_i \sigma_j} \delta_{ij} = \delta_{ij}.$$

于是 $u_1, \dots, u_r$ 是 $\mathbb{R}^m$ 中的标准正交组。

奇异向量的成对关系：

$$Av_i = \sigma_i u_i, \quad A^\top u_i = \sigma_i v_i.$$

这两个关系比直接记忆 $U\Sigma V^\top$ 更能解释分解的含义。

零奇异值：哪些方向被压成零？

若 $A^\top Av = 0$ ，则

$$0 = v^\top A^\top Av = \|Av\|_2^2 \implies Av = 0.$$

故

$$\ker(A^\top A) = \ker(A), \quad \dim \ker(A) = n - r.$$

取 $v_{r+1}, \dots, v_n$ 为 $\ker(A)$ 的标准正交基；取 $u_{r+1}, \dots, u_m$ ，将 $u_1, \dots, u_r$ 补成 $\mathbb{R}^m$ 的标准正交基。

零奇异值需要区别对待：不能在 $\sigma_i = 0$ 时使用 $u_i = Av_i/\sigma_i$ 。

特别注意宽矩阵：例如 $A = \begin{pmatrix} 2 & 0 \end{pmatrix} \in \mathbb{R}^{1 \times 2}$ ，通常只列出一个奇异值 $\sigma_1 = 2$ ，但 $\ker(A) = \text{span}\{(0, 1)^\top\} \neq \{0\}$ 。

SVD 的完整证明

由上一页标准正交基构成正交矩阵

$$U = [u_1 \ \cdots \ u_m], \quad V = [v_1 \ \cdots \ v_n].$$

因 $Av_i = \sigma_i u_i$ ($i \leq r$), 且 $Av_i = 0$ ($i > r$), 有

$$AV = U\Sigma, \quad \Sigma = \begin{pmatrix} \text{diag}(\sigma_1, \dots, \sigma_r) & 0 \\ 0 & 0 \end{pmatrix}_{m \times n}.$$

因此

$$\boxed{A = U\Sigma V^\top.}$$

证明只用了：实对称谱定理、正交基扩充、以及 $A^\top A$ 的结构。

紧致形式与秩一展开

当 $r = \text{rank}(A)$ 时，只保留非零奇异值和对应向量：

$$A = U_r \Sigma_r V_r^\top, \quad U_r \in \mathbb{R}^{m \times r}, \quad \Sigma_r \in \mathbb{R}^{r \times r}, \quad V_r \in \mathbb{R}^{n \times r}.$$

也可写为

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

每个 $u_i v_i^\top$ 都是秩一矩阵；大矩阵可以拆成一系列有序的**秩一成分**。

这既是对 A 的描述，也提供了之后的低秩逼近方法。

例子：一个非方阵的 SVD

考虑 $A \in \mathbb{R}^{3 \times 2}$:

$$A = \begin{pmatrix} 0 & 2 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A^\top A = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}.$$

取 $v_1 = e_2$, $v_2 = e_1$, 则 $\sigma_1 = 2$, $\sigma_2 = 1$;

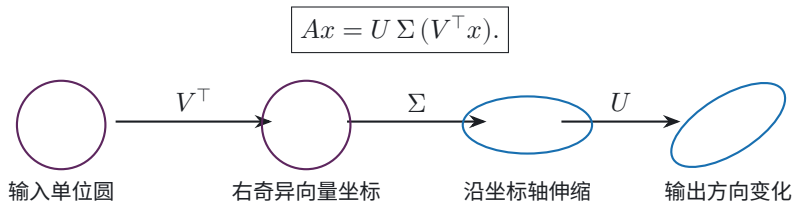
$$u_1 = \frac{Ae_2}{2} = e_1, \quad u_2 = \frac{Ae_1}{1} = e_2.$$

因此

$$U = I_3, \quad \Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad A = U\Sigma V^\top.$$

问题：矩阵既非方阵也不对称，但伸缩方向仍被准确找出。

几何解释：正交变换 + 伸缩 + 正交变换



正交矩阵保持长度与角度，且**可能包含反射**；只有 Σ 改变长度（并可能降维）。

图示为二维满秩情形。对一般 $m \times n$ 矩阵，输入、输出空间甚至可能维数不同。

奇异向量就是主伸缩方向

对 $1 \leq i \leq r = \text{rank}(A)$ ，沿右奇异向量 v_i 有

$$Av_i = \sigma_i u_i, \quad \|v_i\|_2 = \|u_i\|_2 = 1.$$

而 $i > r$ 的右奇异方向属于 $\ker(A)$ ，满足 $Av_i = 0$ 。因此 σ_i 就是该单位方向的伸缩倍数。对单位向量 $x = \sum_{i=1}^n \alpha_i v_i$ ，

$$\|Ax\|_2^2 = \sum_{i=1}^r \sigma_i^2 \alpha_i^2, \quad \sum_{i=1}^n \alpha_i^2 = 1.$$

最大的伸缩倍数是 σ_1 ；若 $n > r$ ，则存在非零输入方向被 A 映成零（即 $\ker(A) \neq \{0\}$ ）。

这也解释了为什么 SVD 与条件数有直接联系。

SVD 给出算子范数

令 $x = Vy$ ，由 U 、 V 的正交性，

$$\frac{\|Ax\|_2^2}{\|x\|_2^2} = \frac{\|\Sigma y\|_2^2}{\|y\|_2^2} \leq \sigma_1^2.$$

取 $x = v_1$ 时等号成立。因此

$$\|A\|_2 = \sigma_1 = \sqrt{\lambda_{\max}(A^\top A)}.$$

回到第六讲：算子 2-范数为何等于 $\sqrt{\lambda_{\max}(A^\top A)}$ ，现在终于有了完整证明。

奇异值、条件数与病态问题

对可逆方阵 A ,

$$\|A\|_2 = \sigma_1, \quad \|A^{-1}\|_2 = \frac{1}{\sigma_n}, \quad \boxed{\kappa_2(A) = \frac{\sigma_1}{\sigma_n}}.$$

特别地, 当 A 实对称正定时, $\sigma_i = \lambda_i(A)$, 故 $\kappa_2(A) = \lambda_{\max}(A)/\lambda_{\min}(A)$ 。
对满列秩的长方矩阵, 常用相应的二范数条件数 σ_1/σ_n (此时 $m \geq n$)。

小奇异值

存在方向被压得很短; 逆向恢复会放大误差。

非平凡零空间

当 $r < n$, 存在非零 x 使 $Ax = 0$; 即使宽矩阵列出的奇异值全为正, 也可能如此。

为什么不直接解正规方程？

第四讲：最小二乘常归结为正规方程

$$A^{\top}Ax = A^{\top}b.$$

若 A 满列秩，其奇异值为 $\sigma_1 \geq \dots \geq \sigma_n > 0$ ，则 $A^{\top}A$ 的特征值是 $\sigma_1^2, \dots, \sigma_n^2$ 。

$$\kappa_2(A^{\top}A) = \kappa_2(A)^2.$$

数值后果：形成正规方程可能使条件数平方化；第四讲引入 QR 来求解最小二乘并非只是换一种写法。

下一问：既然 SVD 揭示了全部伸缩方向，能否只保留最重要的几个？

奇异值的极值刻画

回顾第七讲 Courant–Fischer 定理，应用于 $B = A^\top A$ ：

$$\sigma_1 = \max_{\|x\|_2=1} \|Ax\|_2.$$

更一般地，对 $1 \leq k \leq p = \min(m, n)$,

$$\sigma_k = \max_{\substack{S \subseteq \mathbb{R}^n \\ \dim S = k}} \min_{\substack{x \in S \\ \|x\|_2=1}} \|Ax\|_2.$$

检查下标： $k = 1$ 时最大化一个方向上的伸缩；宽矩阵的 $A^\top A$ 还有额外的 $n - p$ 个零特征值，不属于上面列出的 p 个奇异值。

低秩逼近：保留前 k 个方向

SVD 的秩一展开为

$$A = \sum_{i=1}^r \sigma_i u_i v_i^\top, \quad \sigma_1 \geq \dots \geq \sigma_r > 0.$$

只保留前 $k < r$ 项：

$$A_k := \sum_{i=1}^k \sigma_i u_i v_i^\top, \quad \text{rank}(A_k) = k.$$

问题：在所有秩不超过 k 的矩阵中， A_k 是否最接近 A ？

这里的“最接近”要指定范数： $\|\cdot\|_2$ 还是 $\|\cdot\|_F$ ？

截断 SVD 的误差究竟有多大？

$A - A_k = \sum_{i=k+1}^r \sigma_i u_i v_i^\top$ 。由于奇异向量标准正交，

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

$$\|A - A_k\|_F^2 = \sum_{i=k+1}^r \sigma_i^2.$$

解释：二范数关心**最坏输入方向**上的最大误差；Frobenius 范数累计**所有方向**上的平方误差。

接下来证明没有其它秩不超过 k 的矩阵能做得更好。

Eckart-Young-Mirsky: 二范数下的最优性

任取 $B \in \mathbb{R}^{m \times n}$, 满足 $\text{rank}(B) \leq k$ 。在空间 $S = \text{span}\{v_1, \dots, v_{k+1}\}$ 上, 线性映射 $B|_S$ 必有非零核。因此可取 $x \in S$, $\|x\|_2 = 1$, $Bx = 0$ 。

$$\|A - B\|_2 \geq \|(A - B)x\|_2 = \|Ax\|_2.$$

而在 S 中, $x = \sum_{i=1}^{k+1} \alpha_i v_i$, 所以

$$\|Ax\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 \alpha_i^2 \geq \sigma_{k+1}^2.$$

Eckart-Young-Mirsky: 截断 SVD 在二范数和 Frobenius 范数下均最优。以上证明二范数情形; Frobenius 的证明列为课堂机动。

令 P 为到 $\text{col}(B)$ 的正交投影, $\text{rank}(P) \leq k$ 。由第四讲的正交分解,

$$\|A - B\|_F^2 = \|(I - P)A\|_F^2 + \|PA - B\|_F^2 \geq \|(I - P)A\|_F^2.$$

另一方面, 由 $AA^\top = \sum_i \sigma_i^2 u_i u_i^\top$,

$$\|(I - P)A\|_F^2 = \|A\|_F^2 - \text{tr}(PAA^\top) = \sum_{i=1}^r \sigma_i^2 (1 - \|Pu_i\|_2^2).$$

因此只需证明: 一个秩不超过 k 的投影最多能保留前 k 个奇异值的平方和。

记 $p_i = \|Pu_i\|_2^2$ 。因为 P 是正交投影，

$$0 \leq p_i \leq 1, \quad \sum_{i=1}^r p_i \leq \text{tr}(P) \leq k.$$

又因为 $\sigma_1^2 \geq \dots \geq \sigma_r^2$,

$$\sum_{i=1}^r \sigma_i^2 p_i \leq \sum_{i=1}^k \sigma_i^2.$$

代回上一页的不等式：

$$\|A - B\|_F^2 \geq \sum_{i=k+1}^r \sigma_i^2 = \|A - A_k\|_F^2.$$

Eckart-Young-Mirsky: A_k 同时是二范数与 Frobenius 范数下的**最优秩 k 逼近**。

一个数字例子：舍弃多少信息？

假设矩阵的非零奇异值为 $\sigma_1 = 5, \sigma_2 = 2, \sigma_3 = 1$ 。那么 A_1 是最好的秩一逼近， A_2 是最好的秩二逼近：

	$\ A - A_k\ _2$	$\ A - A_k\ _F$
$k = 1$	2	$\sqrt{2^2 + 1^2} = \sqrt{5}$
$k = 2$	1	1

压缩思路：只存储 U_k 、 Σ_k 、 V_k ，大约需要 $k(m + n + 1)$ 个数，而不是 mn 个。

什么时候值得压缩？ 当 $k \ll \min(m, n)$ 且奇异值下降得足够快。

与 PCA 的联系：若视为数据矩阵，前两项保留的平方能量占比为

$$\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2} = \frac{29}{30} \approx 96.7\%.$$

从最佳低秩逼近走向 PCA

设 $X \in \mathbb{R}^{N \times d}$ 是**已中心化**的数据矩阵 ($N \geq 2$): 每行是一个样本, 列均值为 0。写成

$$X = U\Sigma V^\top.$$

则样本**协方差矩阵**为

$$C := \frac{1}{N-1} X^\top X = V \operatorname{diag}(c_1, \dots, c_d) V^\top,$$
$$c_i = \frac{\sigma_i^2}{N-1} \quad (1 \leq i \leq r), \quad c_i = 0 \quad (r < i \leq d).$$

主成分分析 (PCA): 右奇异向量 $v_1, v_2, \dots$ 给出数据方差从大到小的正交方向; 保留前 k 个方向等价于最优的秩 k 数据近似。

降维后的坐标: $T_k = XV_k \in \mathbb{R}^{N \times k}$; **投影重构:** $X_k = T_k V_k^\top = XV_k V_k^\top$ 。

回扣第四讲: $V_k V_k^\top$ 是正交投影; X_k 正是 Eckart-Young 的最佳秩 k 逼近。

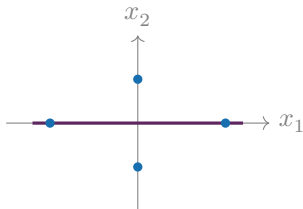
注意: 未经中心化时, 最大奇异方向可能主要反映均值, 而非波动。

PCA 例子：选择投影方向

考虑二维数据点 $\{(2, 0), (-2, 0), (0, 1), (0, -1)\}$ ，均值为 0。

$$X^T X = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}.$$

因此第一主成分是水平轴 $v_1 = e_1$ 。



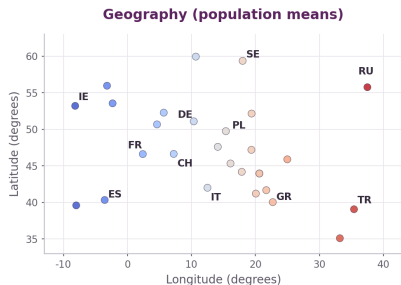
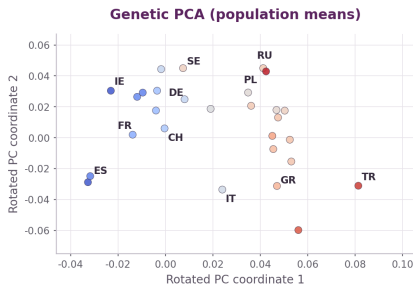
保留一个维度时：

- ▶ 沿 e_1 投影，丢失平方误差 2；
- ▶ 沿 e_2 投影，丢失平方误差 8。

最佳的一维表示： 保留水平方向，重构误差为 2；保留的平方能量占比为 $8/(8+2) = 80\%$ 。

PCA 的真实应用：基因数据与地理结构

真实数据中的现象：欧洲人群的遗传 PCA 坐标与地理位置呈现明显的对应关系。



如何读图？ 根据原作者公开的数据，选取 28 个人群组（每组至少 3 人，共 1,376 人），比较各组的**地理位置均值**与**旋转后的主成分坐标均值**；颜色表示经度。

基于 Novembre et al., *Nature* (2008) 的[公开数据](#)重新绘图（人群均值，非原文图 1）；[研究原文](#)。PCA 不保证恢复地理位置。

SVD 的应用：哪些结论能直接使用？

完整矩阵：最佳低秩逼近

PCA / Eigenfaces: 将中心化的数据投影到前 k 个右奇异向量所张成的子空间。

潜在语义分析 (LSA): 对词语-文档矩阵做截断 SVD，提取主要潜在方向。

部分观测：低秩模型

推荐系统 / 矩阵补全: 仅知道部分评分或条目；可假设未知的完整矩阵具有低秩结构。

重要区别: 缺失值不能直接当零后套用 Eckart-Young 定理。

部分观测时的一个典型目标（背景）：

$$\min_{\text{rank}(B) \leq k} \sum_{(i,j) \in \Omega} (B_{ij} - A_{ij})^2, \quad \Omega = \text{已观察条目的位置集合}.$$

其它联系（背景）: Max-Cut 的近似算法也用到谱松弛，但**并非**直接对邻接矩阵应用截断 SVD。

证明与应用背后的两条主线

分解与几何

$A^T A$ 的特征向量

⇓

奇异向量与伸缩方向

⇓

算子范数、条件数

近似与压缩

秩一分解

⇓

截断奇异值

⇓

Eckart-Young 与 PCA

共同主题：正交坐标揭示矩阵的主伸缩方向；**当奇异值衰减较快时**，少数方向便能很好地近似整个矩阵。

下一讲：分解存在，怎样真正计算？

第七讲给出对称矩阵的**谱定理**；本讲由此得到**奇异值分解**。但完整计算 SVD 需要求解特征方向，至少概念上需要知道：

$$A^{\top} A v_i = \sigma_i^2 v_i.$$

注意数值稳定性：上述等式是证明 SVD 存在性的工具，不表示实际计算 SVD 时应总是先显式形成 $A^{\top} A$ ；这样做可能使条件数平方化。

Lecture 9：回到第七讲的幂迭代，研究逆迭代、位移、同时迭代和 QR 特征值算法。**更远的联系：**高效计算特征信息之后，第十、十一讲将转入线性方程组的迭代求解与 Krylov 方法。

补充 用 SVD 构造伪逆 (pseudoinverse)

已知 $A = U\Sigma V^\top$ ，且 $r = \text{rank}(A)$ 。定义

$$A^+ = V\Sigma^+U^\top, \quad \Sigma_{ii}^+ = \begin{cases} 1/\sigma_i, & i \leq r, \\ 0, & i > r. \end{cases}$$

等价地

$$A^+ = \sum_{i=1}^r \frac{1}{\sigma_i} v_i u_i^\top.$$

- ▶ 当 A 是可逆方阵， $A^+ = A^{-1}$ ；
- ▶ 一般情况下，即使 A 不可逆或不是方阵， A^+ 仍有定义。

与第四讲的联系：投影与最小二乘问题可以用这一统一公式描述。

设 $A = U\Sigma V^\top$, 考虑 $\min_x \|Ax - b\|_2$, 令 $x = Vy$ 、 $c = U^\top b$ 。则问题变成

$$\min_y \|\Sigma y - c\|_2^2 = \min_y \left(\sum_{i=1}^r (\sigma_i y_i - c_i)^2 + \sum_{i=r+1}^m c_i^2 \right).$$

对 $1 \leq i \leq r$ 取 $y_i = c_i/\sigma_i$ 即可最小化残差; 对剩余的自由坐标 $r < i \leq n$ 取 $y_i = 0$, 又使 $\|x\|_2$ 最小。

$$x_* = A^+ b.$$

对于超定、欠定、秩亏问题, $A^+ b$ 都是**二范数最小的最小二乘解**。

另一个推导 $\|A\|_2$ 的方法，是研究约束最优化问题

$$\max_{x \in \mathbb{R}^n} x^\top A^\top A x \quad \text{s.t. } x^\top x = 1.$$

其拉格朗日函数为

$$\mathcal{L}(x, \lambda) = x^\top A^\top A x - \lambda(x^\top x - 1).$$

一阶必要条件：

$$\nabla_x \mathcal{L} = 2A^\top A x - 2\lambda x = 0 \quad \implies \quad A^\top A x = \lambda x.$$

说明：这是谱定理和 Rayleigh 商的另一种解释；后续优化课程会系统讲解拉格朗日乘数法。

SVD 寻找**依赖数据**的低维结构；还有一种思路：使用**随机线性映射**近似保留距离。

Johnson-Lindenstrauss 引理 (仅陈述): 给定 $0 < \varepsilon < 1$ 和有限点集 $X \subset \mathbb{R}^d$, $|X| = N$, 取 $k = O(\varepsilon^{-2} \log N)$, 存在映射 $f: \mathbb{R}^d \rightarrow \mathbb{R}^k$, 使任意 $x, y \in X$ 有

$$(1 - \varepsilon) \|x - y\|_2^2 \leq \|f(x) - f(y)\|_2^2 \leq (1 + \varepsilon) \|x - y\|_2^2.$$

这也是压缩的数学机制，但**不依赖**先计算 SVD。概率构造与证明留作课外延伸。