# 15 奇异值分解

从矩阵的几何作用到低秩近似

# 学习目标

- 解释三个因子的维度、正交性和作用次序，区分输入方向与输出方向，以及完整、薄和截断 SVD
- 说明怎样从单位输入方向的像长度出发，依次得到输入方向、伸缩强度和输出方向，并识别被压到零的方向
- 解释怎样把矩阵拆成逐方向的秩一分量，怎样截断这些分量，并在指定误差度量下理解“最佳近似”
- 说明压缩、降噪、PCA 和伪逆分别处理什么对象、对方向做什么操作，以及这些操作在什么条件下成立
- 面对给定矩阵与任务，依据对象含义和任务目标选择保留、舍弃、投影或反演方向，并检查误差定义、中心化要求和敏感性

# 1 奇异值分解的几何意义与分解形式

按输入方向、伸缩强度与输出方向理解矩阵作用

# 1.1 奇异值分解的几何解释

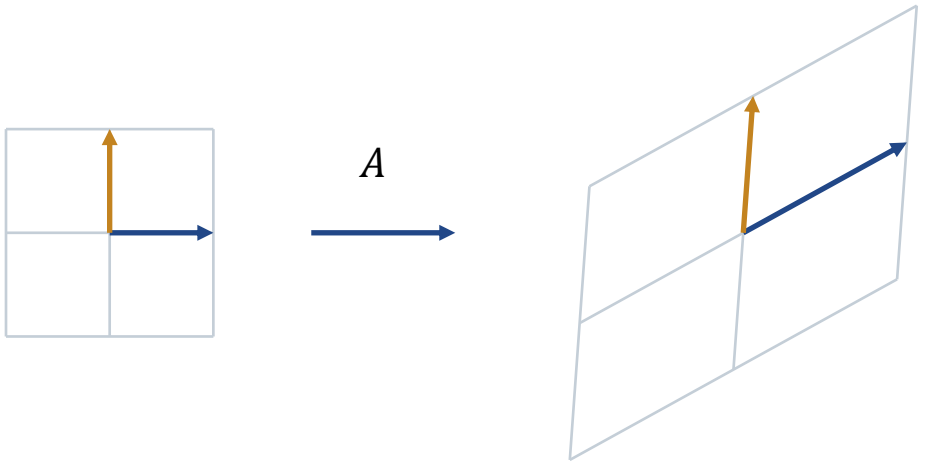
以二维矩阵为例，奇异值分解用正交输入方向、对应的伸缩系数和输出方向描述矩阵变换。任意输入都可以分解为这些输入方向上的分量；矩阵分别变换各分量，再将结果相加，就得到输出。由此可以写出矩阵的三因子分解。

# 为什么需要 SVD

- $A \in \mathbb{R}^{m\times n}$ 将 $x \in \mathbb{R}^n$ 映为 $Ax \in \mathbb{R}^m$

矩阵元素给出输入坐标到输出坐标的组合规则。我们还希望知道：各个输入方向经过矩阵后，长度怎样变化、输出指向哪里？当 $m \neq n$ 时，输入与输出属于不同维度的空间。

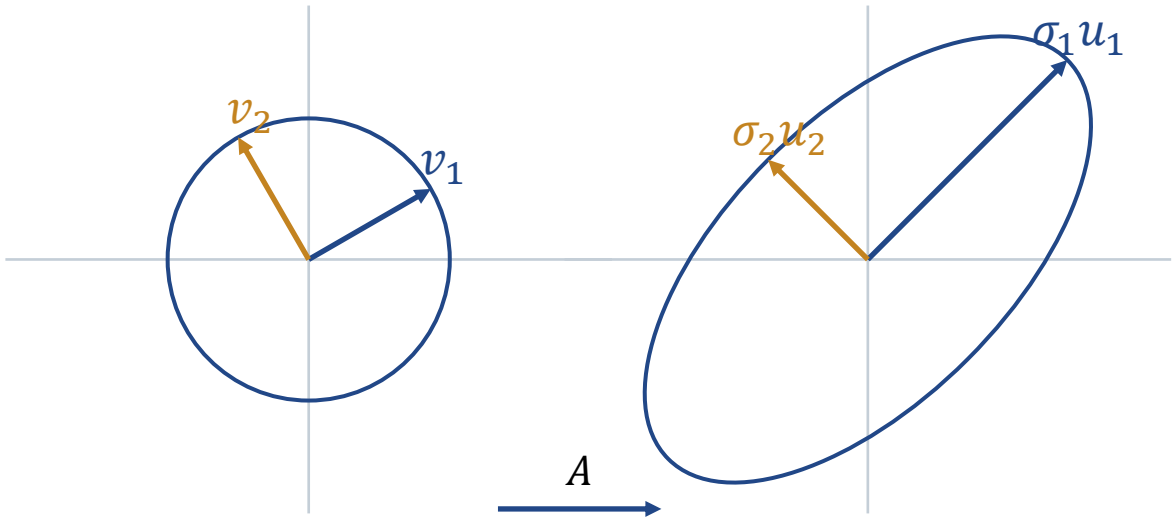
- 能否选择**正交归一的输入方向**，使非零像也沿正交输出方向？
  - 正交归一指长度为 1、内积为零
  - 任意正交输入基的像未必正交
- 任意实矩阵均有此配对，零像表示方向被压到零
  - 后续解释这些方向怎样决定任意输入的作用




原坐标网格　　像的方向发生混合

# 单位圆的像与奇异向量

- 二维单位圆包含所有单位输入；像为椭圆，也可退化为线段或点
- SVD 选取两侧正交归一基，使 $Av_i = \sigma_i u_i$，$\sigma_1 \geq \sigma_2 \geq 0$
- **奇异向量与奇异值**
  - $v_i$：**右奇异向量**，输入方向
  - $u_i$：**左奇异向量**，输出侧配对方向
  - $\sigma_i$：**奇异值**，沿 $v_i$ 的长度缩放系数；也等于单位输入 $v_i$ 的像长度 $\|Av_i\|$，$\|\ \cdot\ \|$ 为长度
    - $\sigma_i > 0$ 时像沿 $u_i$；$\sigma_i = 0$ 时像为零，无实际指向
- 半轴长度对应最大、最小像长度；输入方向须由 $A$ 的配对关系确定

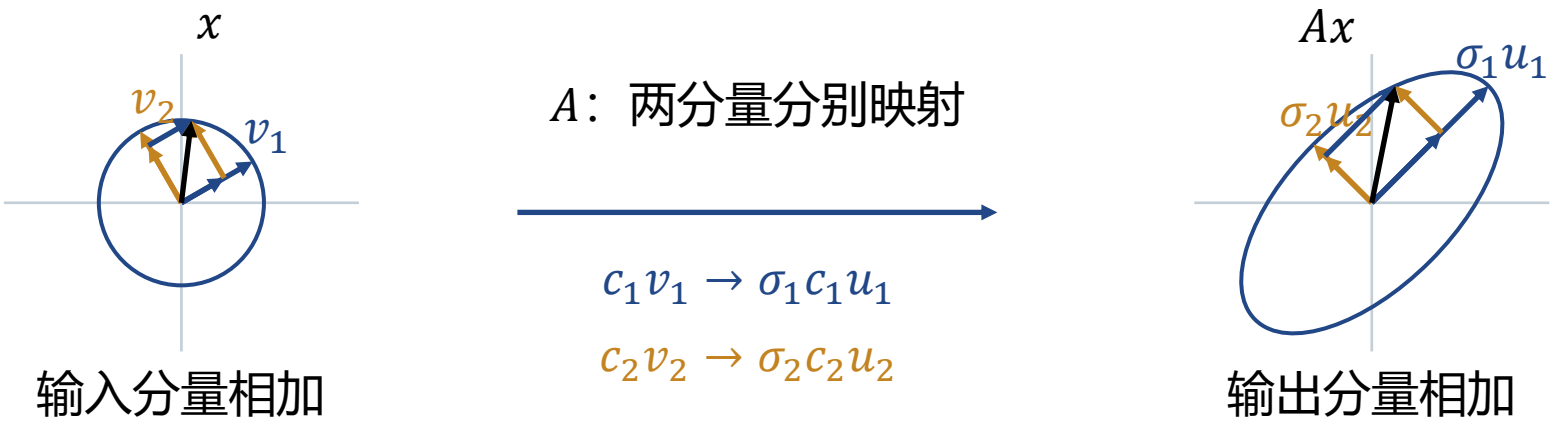



输入：单位圆　　输出：椭圆

$Av_i = \sigma_i u_i$；$\sigma_1 = 2, \sigma_2 = 1$（示意）

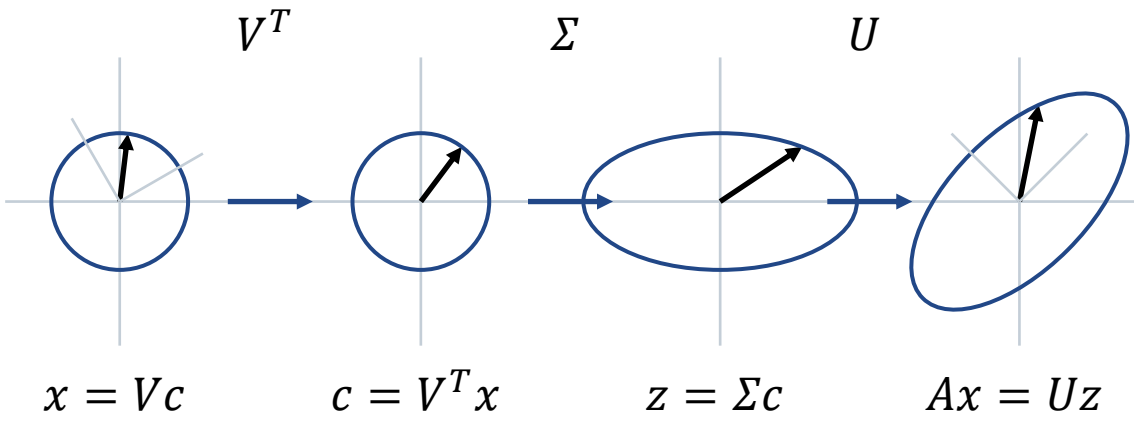
# 任意输入向量的分解与变换

- 任意输入在正交基下都有唯一坐标
  - $x = c_1 v_1 + c_2 v_2$，$c_i = v_i^T x$ 是有符号坐标；取内积时自身项保留，另一项因正交为零
- 由配对关系与线性性，**各输入分量分别映射**
  - $A(c_i v_i) = \sigma_i c_i u_i$
- 由线性性，**向量和的像等于各分量的像之和**
  - $Ax = \sigma_1 c_1 u_1 + \sigma_2 c_2 u_2$，故基方向的作用确定任意输入
- 对单位输入，$c_1^2 + c_2^2 = 1$，伸缩后坐标 $(\sigma_1 c_1, \sigma_2 c_2)$ 描出椭圆



# 由向量变换得到三因子分解

- 将方向按列组成矩阵
  - $V = [v_1\ v_2]$，$U = [u_1\ u_2]$，$\Sigma = diag(\sigma_1, \sigma_2)$；diag 表示对角矩阵
  - $x = Vc$，$c = (c_1, c_2)^T$ 是输入坐标
- **三个因子的作用**
  - $c = V^T x$：读取坐标
  - $z = \Sigma c$：各坐标乘对应 $\sigma_i$
  - $Uz = \sum_i \sigma_i\, c_i u_i$：组合输出
- 对所有 $x$，$Ax = U\Sigma V^T x$，故 $A = U\Sigma V^T$
- $U, V$ 保持长度与夹角（旋转或反射）
- $\Sigma$ 改变长度



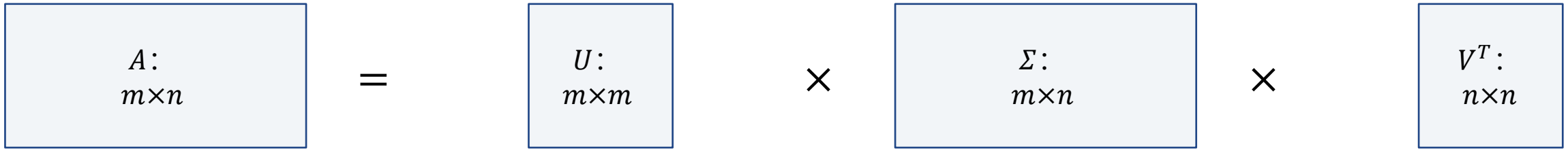

同一输入：$c = (3/5, 4/5)^T$

伸缩后：$z = (6/5, 4/5)^T$

## 1.2 分解形式与奇异向量关系

本节将二维分解推广到一般尺寸的矩阵，先核对各因子的空间与尺寸，再比较完整、薄与截断分解保留的内容。随后利用奇异向量的配对关系，说明非零奇异值下左右奇异向量如何相互求得，以及零奇异值时的限制。

# 一般矩阵的 SVD 与因子尺寸




- 对任意 $A \in \mathbb{R}^{m\times n}$，均有上图的完整 SVD；令 $p = min(m,n)$
- **输入与输出空间的正交基**
  - $V = [v_1 \cdots v_n] \in \mathbb{R}^{n\times n}$，$V^TV = I_n$：输入空间的 $n$ 个方向
  - $U = [u_1 \cdots u_m] \in \mathbb{R}^{m\times m}$，$U^TU = I_m$：输出空间的 $m$ 个方向；$I$ 表示单位矩阵
- $\Sigma \in \mathbb{R}^{m\times n}$ 前 $p$ 个对角元素为 $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$，其他元素为零
- **矩形 $\Sigma$ 的坐标变换**
  - $c = V^Tx \in \mathbb{R}^n$，$z = \Sigma c \in \mathbb{R}^m$，$Ax = Uz$；前 $p$ 项满足 $z_i = \sigma_i c_i$
  - $n > p$ 时额外输入方向映为零；$m > p$ 时额外输出坐标恒为零
  - 例：$m = 2, n = 3$，$(c_1, c_2, c_3)^T$ 变为 $(\sigma_1 c_1, \sigma_2 c_2)^T$，$v_3$ 不贡献输出
- 一般矩阵按配对方向组合输出：$Ax = \sum_{i=1}^{p} \sigma_i \, (v_i^T x) u_i$

# 完整、薄与截断 SVD

令 $p = min(m,n)$、$r = rank(A)$。正交因子不改变秩，故 $r$ 等于非零奇异值个数

| 形式 | 保留对象 | $U,\Sigma,V$ 尺寸 | 表示 |
|---|---|---|---|
| 完整 | 两侧完整正交基 | $m\times m, m\times n, n\times n$ | $A = U\Sigma V^T$，精确 |
| 薄 | 前 $p$ 个配对方向，含零值 | $m\times p, p\times p, n\times p$ | $A = U_p\Sigma_p V_p^T$，精确 |
| 截断 | 前 $k$ 个非零配对，$0 \le k \le r$ | $m\times k, k\times k, n\times k$ | $A_k = U_k\Sigma_k V_k^T$ |

- 下标表示取前若干列和相应对角元素；完整到薄只删除额外零贡献，$2\times3$ 例中删去 $v_3$，仍保留 $v_2$，即使 $\sigma_2 = 0$
- 薄到截断：$k < r$ 舍弃非零作用，得到近似；$k = r$ 精确重构；$k = 0$ 约定 $A_0 = 0$

# 奇异三元组与转置矩阵

- **奇异三元组**：将第 $i$ 组配对记为 $(u_i, \sigma_i, v_i)$，$i = 1, \ldots, p$，其正向关系为
  - $Av_i = \sigma_i u_i$
- 由 $A^T = V\Sigma^T U^T$，正交归一性使 $U^T u_i$ 的第 $i$ 项为 1、其他项为零；$\Sigma^T$ 将其第 $i$ 项变为 $\sigma_i$，所以
  - $A^T u_i = \sigma_i v_i$：转置矩阵沿同一组方向反向配对；这不表示 $A^T$ 是 $A$ 的逆
- $\sigma_i = 0$ 时，$Av_i = 0$、$A^T u_i = 0$；零空间指被相应矩阵映为零的向量集合，零像不能通过除以 $\sigma_i$ 确定方向
- **奇异向量的非唯一性**
  - 奇异值序列由 $A$ 确定；非零配对向量可同时变号
  - 对重复的正奇异值，两侧对应的基必须按相同的正交变换更换，才能保持配对
  - 两侧零空间的基可分别选择并保持正交归一

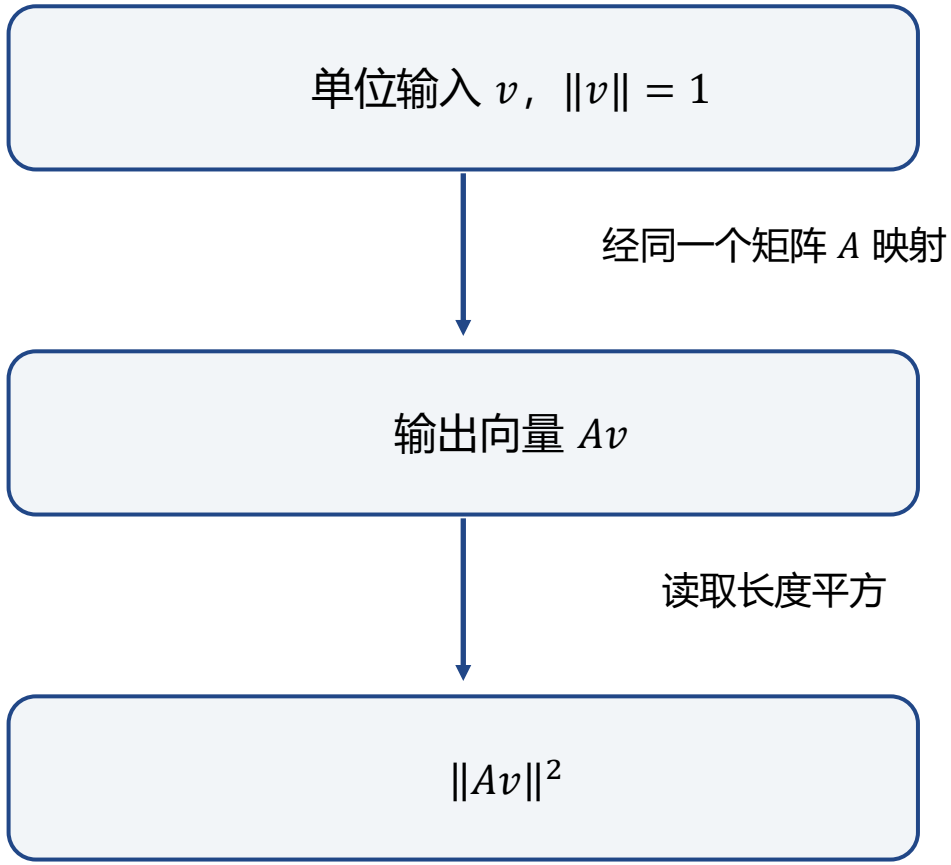
# 2 奇异值与奇异向量的构造

从像长度寻找方向，再由非零像确定输出配对

# 由 $A^TA$ 确定右奇异向量

前面已理解输入方向、伸缩和输出方向；现在用 $A^TA$ 找到这些几何对象，再由非零像配对输出。

- **比较单位输入**
  - $A(tv) = tAv$，像长随输入长度变化；固定 $\|v\| = 1$，才能比较方向的伸缩差异
- **像长度与输入二次型**
  - $\|Av\|^2 = (Av)^T(Av) = v^T(A^TA)v$
  - 二次型将输入和矩阵组合为标量；$A^TA$ 记录同一长度平方，原映射仍是 $A$
- **输入侧的正交特征基**
  - $(A^TA)^T = A^TA$，且 $z^TA^TAz = \|Az\|^2 \geq 0$，故对称半正定
  - 谱定理保证可选正交归一特征基：$A^TAv_i = \lambda_i v_i, \lambda_i \geq 0$
  - 下一页检查这些方向怎样分离长度贡献



# 奇异值与单位输入的像长度

用二次型比较不同单位输入的像长度平方，沿用伸缩系数为 2、1 的几何示意。

- 将单位输入沿正交特征方向展开为 $v = \sum_i c_i v_i$，$c_i = v_i^T v$ 是输入坐标，$\sum_i c_i^2 = 1$
  - $\|Av\|^2 = \sum_i \lambda_i c_i^2$
  - 特征关系与正交性使交叉项为零；$\lambda_i c_i^2$ 是各分量的长度平方贡献

| 单位输入 | 像长度平方 |
| --- | --- |
| 沿 $v_1$ 或 $v_2$ | 分别为 4 或 1 |
| $v = c_1 v_1 + c_2 v_2$，$c_1^2 + c_2^2 = 1$ | $4c_1^2 + c_2^2$，介于 1 与 4 |

- 像长平方为特征值的加权平均，**最大值由最强纯方向取得**
- 奇异值是单位输入的实际像长度，即 $\sigma_i = \sqrt{\lambda_i} = \|Av_i\|$，按 $\lambda_i$ 降序也就是按 $\sigma_i$ 降序
  - 前 $p = min(m, n)$ 个方向构成配对范围；额外输入方向若存在，其特征值为零，映为零

输入正交不自动保证像正交，下一步核对非零像。

# 左奇异向量的构造与零奇异值

- $\sigma_i > 0$ 时，$Av_i$ 给出输出方向与长度；通过**归一化**去掉长度，得到单位方向
  - $u_i = Av_i/\sigma_i, Av_i = \sigma_i u_i$
  - $Av_i$ 沿 $u_i$，不要求沿输入方向 $v_i$
  - $\sigma_i = 0$ 时 $Av_i = 0$，$v_i$ 属于零空间；零像无可归一化方向，不能除以 $\sigma_i$ 构造 $u_i$
- 对 $i \neq j$，沿用输入侧特征关系
  - $\sigma_i \sigma_j u_i^T u_j = v_i^T A^T A v_j = \sigma_j^2 v_i^T v_j = 0$
  - 因 $\sigma_i, \sigma_j > 0$，有 $u_i^T u_j = 0$；输入特征基既分离长度贡献，也使非零像正交，完成方向配对
- 删去薄 SVD 的零项，非零方向组成 $U_r, V_r$，强度组成 $\Sigma_r$
  - $A = U_r \Sigma_r V_r^T$，$r = rank(A)$ 是非零奇异值数
  - 因子尺寸：$U_r: m \times r$，$\Sigma_r: r \times r$，$V_r: n \times r$
  - 共同非零谱：由 $A^T A v_i = \sigma_i^2 v_i$ 得 $AA^T u_i = \sigma_i^2 u_i$，两侧共用非零特征值，无须另求输出方向

下一单元将此乘积展开成逐方向的秩一分量。

# 二维 SVD 算例

用下面的二维矩阵核对奇异值与奇异向量的构造；本例与前面的几何示意使用不同矩阵。

| 构造步骤 | 本例的结果 | 对应的几何角色 |
| --- | --- | --- |
| 输入侧谱 | $A=\begin{bmatrix}0 & 3\\ 2 & 0\end{bmatrix}$，$A^TA=diag(4,9)$ | 在输入空间读取像长度平方 |
| 按降序排列方向与强度 | $v_1=(0,1)^T,\sigma_1=3;v_2=(1,0)^T,\sigma_2=2$ | 两个单位输入分别伸成长度 3、2 |
| 非零像归一化 | $u_1=Av_1/3=(1,0)^T;u_2=Av_2/2=(0,1)^T$ | 输出单位方向也正交，与输入方向配对 |
| 组成因子 | $U=[u_1\ u_2]=I$，$\Sigma=diag(3,2)$，$V=[v_1\ v_2]=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$ | 将方向、强度和配对写成矩阵 |

$$U\Sigma V^T=\begin{bmatrix}0 & 3\\ 2 & 0\end{bmatrix}=A$$

输入侧谱确定方向与强度，非零像确定输出配对，因子乘积重构 $A$；下一单元展开逐方向贡献。

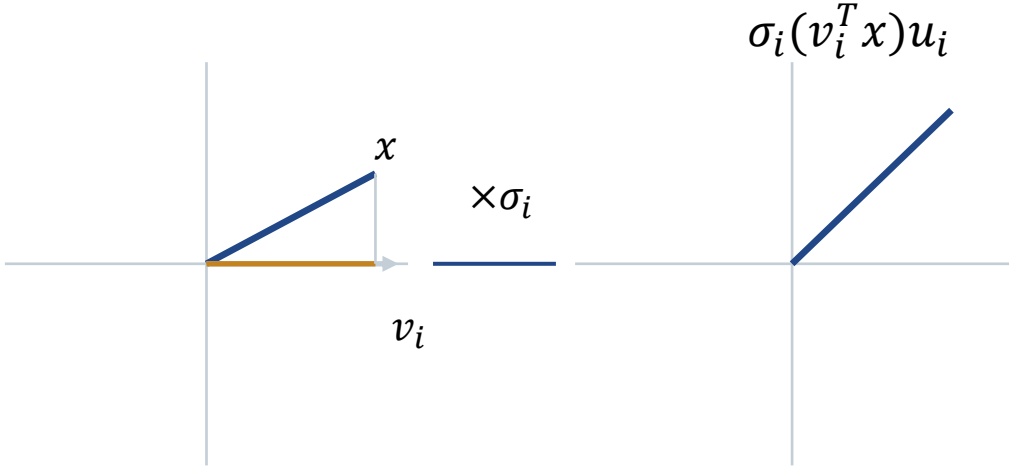
# 3 秩一分量与低秩近似

将单方向作用写成秩一分量，并按误差口径截断

# SVD 的秩一展开

前面将任意输入分解后逐方向变换；把每条方向的作用写成矩阵，就得到秩一分量。

- **一个方向对应一个秩一矩阵**
  - 把已知的“读取一个坐标、按强度伸缩、组合到输出方向”写成
    - $(\sigma_i u_i v_i^T)x = \sigma_i (v_i^T x) u_i$
  - $v_i^T x$ 是标量坐标，$u_i v_i^T$ 是列向量与行向量相乘得到的外积矩阵；它对任何输入的输出都在 $u_i$ 张成的一维空间中
  - 由于 $i \le r$ 时 $\sigma_i > 0$，每个 $\sigma_i u_i v_i^T$ 都是秩一矩阵
- **所有方向相加重构矩阵**
  - $A = \sum_{i=1}^{r} \sigma_i \, u_i v_i^T$
  - 对任意输入，所有单方向输出相加得到 $Ax$；相应的秩一矩阵相加得到 $A$




投影坐标：$v_i^T x$　　输出的一项

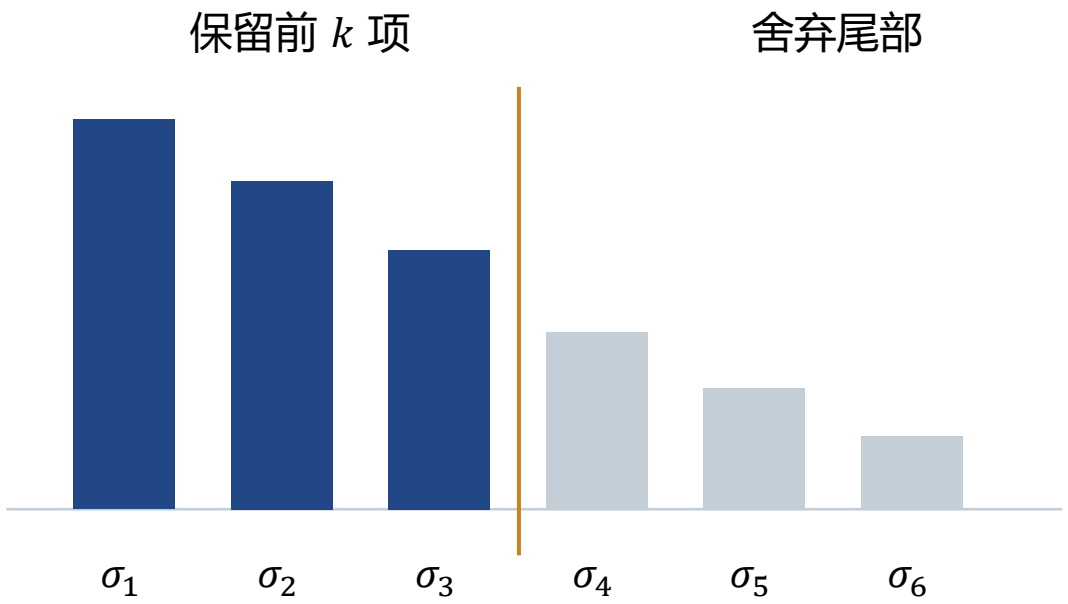
只沿一个输出方向产生贡献

# 截断 SVD 与残差

取 $0 \le k \le r = rank(A)$，截断只省略尾部方向的作用。

- **保留项形成截断矩阵**
  - $A_k = \sum_{i \le k} \sigma_i u_i v_i^T$：保留最大的 $k$ 项
  - 秩至多为 $k$；$k = 0$ 时空和为零
- **舍弃项形成残差**
  - $E_k = A - A_k = \sum_{i > k} \sigma_i u_i v_i^T$
  - $k < r$ 残差非零，$k = r$ 精确重构
- **同一残差的两种误差度量**
  - 谱范数 $\|E_k\|_2 = max_{\|x\|_2 = 1} \|E_k x\|_2$：最坏单位输入误差
  - Frobenius 范数 $\|E_k\|_F = (\sum_{a,b} (E_k)_{ab}^2)^{1/2}$：元素误差的合成大小

先算当前误差，再按同一口径比较候选。




截断位置 $k$（图示取 3）

示意谱，非具体矩阵数据

# 截断 SVD 的误差计算

- 上一页已选定 $A_k$，$E_k = \sum_{i>k} \sigma_i u_i v_i^T$
  - $k = r$ 时残差为零；超出配对范围的奇异值按零处理
- 计算**整体元素误差**时，Frobenius 内积是对应元素乘积之和
  - 单位外积满足 $\langle u_i v_i^T, u_j v_j^T \rangle_F = (u_i^T u_j)(v_i^T v_j)$：同项为 1、异项为零
  - 因此交叉项为零，每个舍弃项贡献 $\sigma_i^2$，得到 $\|E_k\|_F^2 = \sum_{i>k} \sigma_i^2$
- 计算**最坏方向误差**时，对单位输入 $x$
  - $\|E_k x\|_2^2 = \sum_{i>k} \sigma_i^2 (v_i^T x)^2$
  - $k < r$ 时至多为 $\sigma_{k+1}^2$，在 $x = v_{k+1}$ 时取得，因此 $\|E_k\|_2 = \sigma_{k+1}$

已算出当前截断损失；是否有其他秩不超过 $k$ 的矩阵做得更好，还需比较全部候选。

# 最佳低秩近似

- 保留上一页的 $A_k$ 和误差度量，将它与全部满足 $rank(B) \le k$ 的候选矩阵 $B$ 比较

| 误差度量 | 当前截断 $A_k$ 的误差 | 全部候选中的最小误差 |
| --- | --- | --- |
| 谱范数 | $\|A - A_k\|_2 = \sigma_{k+1}$ | $min_{rank(B)\le k}\|A - B\|_2 = \sigma_{k+1}$ |
| Frobenius 范数平方 | $\|A - A_k\|_F^2 = \sum_{i>k} \sigma_i^2$ | $min_{rank(B)\le k}\|A - B\|_F^2 = \sum_{i>k} \sigma_i^2$ |

- **Eckart–Young–Mirsky 定理**：$A_k$ 分别达到谱范数和 Frobenius 范数下的最小误差
- 两列误差相同才说明当前截断达到全局最优；只算出当前残差不能代替定理结论
- 这里的“最佳”受秩约束和上述误差度量限定，不保证下游预测效果最佳
- 当 $k = r$ 时残差为零；超出配对范围的奇异值按零处理

# 累计能量与截断秩的选择

- **保留的 Frobenius 能量比例**：这里“能量”指奇异值平方和；当 $\sum_i \sigma_i^2 > 0$ 时
  - $\rho_k = \dfrac{\sum_{i \le k} \sigma_i^2}{\sum_i \sigma_i^2}$
  - 由上一页的舍弃平方和，$1 - \rho_k = \|E_k\|_F^2 / \|A\|_F^2$：保留比例越高，相对残差能量越小
  - 零矩阵可用秩零精确表示，不使用能量比例
- **阈值选取规则**：给定 $0 < \eta \le 1$，取累计能量首次达到阈值的最小 $k$
  - $k = min\{j : \rho_j \ge \eta\}$
- 能量比例只反映**矩阵近似误差**。最终的 $k$ 还需结合任务约束；存在下游任务时，还应检查验证集性能

# 4 奇异值分解的应用与使用条件

用任务和结构假设决定保留、舍弃、投影或反演

# 图像的低秩近似

- **灰度图像的存储代价**
  - 灰度图 $A \in \mathbb{R}^{m\times n}$ 含 $mn$ 个像素值
  - $U_k, V_k$ 及奇异值共 $mk + nk + k = k(m + n + 1)$ 个标量
  - 参数优势：$k(m + n + 1) < mn$；文件大小另受量化、编码影响
- 比较同一张 160×240 灰度图的原图及 $k = 5,30$ 截断结果
  - 参数量：标量数量
  - 能量比例：保留能量
  - 相对误差：$\|A - A_k\|_F / \|A\|_F$
- **矩阵误差与视觉感知质量**
  - 谱衰减越快，小 $k$ 越可能保留高能量
  - 矩阵误差与感知质量分别评价



**原图**

160×240
38,400 像素



$k = 5$

参数 2,005
能量 91.9%
相对误差 28.4%



$k = 30$

参数 12,030
能量 97.5%
相对误差 15.7%

同一张图、同一次实际 SVD 计算

# 低秩降噪的条件

- 观测矩阵 $Y = S + E$，其中 $S$ 是待恢复的有用信号，$E$ 是扰动；希望从 $Y$ 估计 $S$

| 比较维度 | 低秩压缩 | 低秩降噪 |
| --- | --- | --- |
| 目标对象 | 原始矩阵 | 观测中的有用信号 $S$ |
| 选择依据 | 权衡表示代价与矩阵重构误差 | 有用结构集中在少数强方向，部分扰动分散在弱方向 |
| 判断边界 | 参数量与文件大小、矩阵误差与感知质量分别评价 | 强弱只按奇异值大小区分；小奇异值不必然是噪声，有用信号也可能位于弱方向，阈值需任务验证 |

- 保留强方向、舍弃候选弱方向，重构较低秩的 $Y_k$；变化的是要逼近哪个对象，以及凭什么决定舍弃
- $Y_k$ 对观测矩阵 $Y$ 的近似最优，不保证它对真实信号 $S$ 的恢复误差也最小
- 本单元把矩阵直接视为待近似对象；下一单元改把矩阵的行、列解释为样本与特征，方向也随之获得数据含义

# 数据矩阵的奇异向量与样本得分

约定 $X \in \mathbb{R}^{N\times d}$：每行对应一个样本，每列对应一个特征。

- **从一个样本得分到全部样本得分**
  - $v_i \in \mathbb{R}^d$ 是特征空间中的方向
  - 数据表第 $j$ 行与 $v_i$ 的内积给出该样本在此方向上的得分
  - 全部样本的得分组成 $Xv_i = \sigma_i u_i \in \mathbb{R}^N$，向量的第 $j$ 个元素就是第 $j$ 个样本的得分
- **得分与方向组成秩一数据分量**
  - $\sigma_i u_i v_i^T$ 是“样本得分向量 × 特征方向”形成的秩一分量
  - 同一个秩一项既描述线性变换中的单方向作用也描述数据矩阵中“样本得分沿某一特征方向展开”的单个分量

**共同秩一项：** $\sigma_i u_i v_i^T$

变换：$x \to v_i^T x \to \sigma_i (v_i^T x) u_i$

数据表第 $j$ 行 $x_j^T$ ⟶ $x_j^T v_i = \sigma_i u_{ji}$

全部样本得分 $\sigma_i u_i$

× 特征方向 $v_i^T$

= 秩一数据分量

# 数据中心化与 PCA 主方向

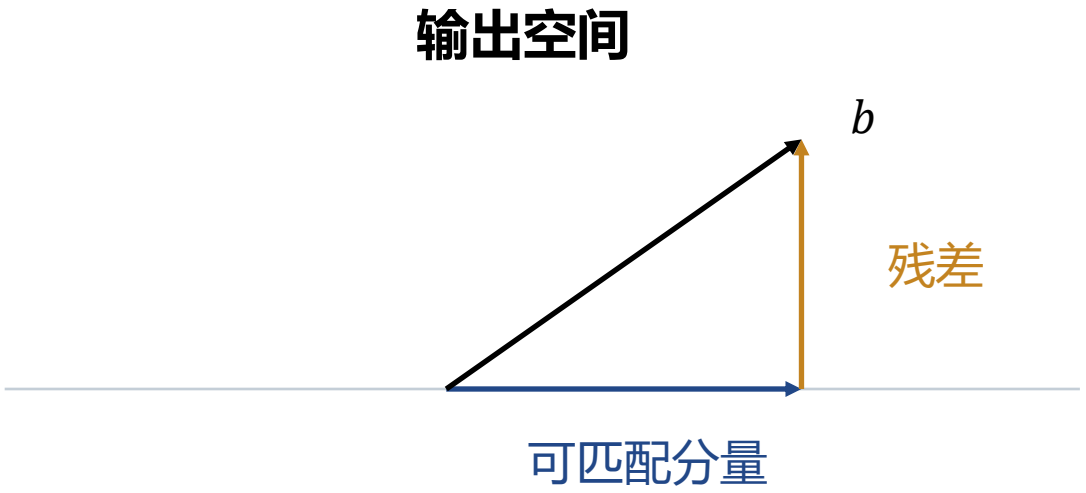
- 约定 $X \in \mathbb{R}^{N\times d}$，样本为行、特征为列，且 $N > 1$
  - $\mu = \frac{1}{N} X^T \mathbf{1} \in \mathbb{R}^d$ 为各特征的样本均值，$\mathbf{1} \in \mathbb{R}^N$ 为全 1 列向量
  - $X_c = X - \mathbf{1}\mu^T$：每个样本减去同一个特征均值向量，使每列的样本均值为零
- **SVD 与协方差特征方向**
  - $X_c = U\Sigma V^T$
  - $\frac{X_c^T X_c}{N-1} = V \frac{\Sigma^T \Sigma}{N-1} V^T$
  - 因而配对范围内的 $v_i$ 是特征协方差矩阵的特征方向，对应特征值 $\sigma_i^2/(N-1)$；额外特征方向若存在则对应零特征值
- 按奇异值降序取得特征主方向，也就是按协方差特征值降序排列的方向
  - 使用这一标准 PCA 接口前，应先按特征中心化数据

# 主方向投影与降维坐标

- 沿用中心化数据 $X_c \in \mathbb{R}^{N\times d}$ 及 $X_c = U\Sigma V^T$
  - 取 $1 \le k \le min(N,d)$，按降序选前 $k$ 个右奇异方向，组成 $V_k \in \mathbb{R}^{d\times k}$
- 取得**降维坐标**时，复用单方向 $Xv_i$ 的得分读法，对中心化数据同时取 $k$ 个方向
  - $Z = X_c V_k = U_k \Sigma_k \in \mathbb{R}^{N\times k}$
  - 每行是一个样本的低维坐标，每列是一个主方向上的全部样本得分
- **坐标列的方差**
  - 中心化使第 $i$ 列坐标均值为零，该列长度平方为 $\sigma_i^2$，所以 $Var(Z_i) = \frac{\sigma_i^2}{N-1}$
- 中心化数据，用 SVD 求特征方向，再投影得到低维坐标
  - PCA 的选维方法留到下一讲；这里处理的是保留或投影方向，下一单元转向可恢复方向的反演

# 伪逆与最小二乘

- 给定 $b \in \mathbb{R}^m$，求 $x \in \mathbb{R}^n$
  - 先最小化 $\|Ax - b\|_2$，再在这些解中取最小范数
- **非零方向反演，组成伪逆**
  - 读取 $u_i^T b$，除以非零 $\sigma_i$，沿 $v_i$ 组合；零值不能取倒数
  - $A^+ = V\Sigma^+ U^T$；$A^+, \Sigma^+$ 均为 $n \times m$，$\Sigma^+$ 非零值取倒数，其余置零
- **最小残差与最小范数的选择**
  - 列空间含全部可达 $Ax$；匹配可达分量，其余为最小残差
  - 输入零空间分量正交于恢复分量，只增范数、不改输出；取零使范数最小
- 对任意 $b$，最小范数最小二乘解为
  - $x = A^+ b$
  - 有精确解时，残差为零且取最小范数解




非零方向：读取 → 除以 $\sigma_i$

↓

**输入空间：** $x = A^+ b$

输入零空间的自由分量取零

# 小奇异值与扰动放大

- 固定 $A$，只把目标沿 $u_i$ 改变 $\Delta b = \varepsilon u_i$
  - 当 $\sigma_i > 0$，$\Delta x = (\varepsilon/\sigma_i)v_i$：$1/\sigma_i$ 为放大倍数，小奇异值方向更敏感

| 考察范围 | 条件数或伸缩比 | 条件与含义 |
| --- | --- | --- |
| 全部满秩配对方向 | $\kappa_2(A) = \sigma_1/\sigma_p$ | $r = p = min(m,n)$，$\sigma_p > 0$ |
| 秩亏的全部配对方向 | 完整求解范围记为无穷大 | $r < p$，含无法反演的零方向 |
| 可恢复的非零方向 | $\sigma_1/\sigma_r$ | $r > 0$，只比较非零方向，不表示存在双侧逆 |

- 零奇异值方向无法反演；SVD 揭示敏感性，本身不消除不稳定性
- 下一单元将四类用法放在同一选择框架中检查

# 匹配应用任务与方向操作

- 依据各情境的任务重点，与“保留、舍弃、投影、反演”匹配
  1. 用较少参数近似灰度图像矩阵，并控制矩阵重构误差
  2. 在有用结构集中于少数强方向的假设下减弱候选扰动
  3. 把中心化的样本—特征矩阵变成 $k$ 维坐标
  4. 对 $Ax \approx b$ 求最小二乘解中的最小范数解

- **操作选项**
  - 对非零奇异值取倒数
  - 投影到前 $k$ 个右奇异方向
  - 保留前 $k$ 个秩一分量
  - 在结构假设支持时舍弃候选弱方向
- 每个选项只描述方向操作。匹配时还需结合输入对象与期望输出，奇异值大小本身不足以确定答案
- **说明理由**
  - 任选一种情境，按“输入对象—方向操作—输出”解释匹配理由，并说明一项成立条件或潜在风险

# 应用任务的操作匹配与使用条件

| 任务 | 方向操作 | 输出 | 必要边界 |
| --- | --- | --- | --- |
| 图像压缩 | 保留前 $k$ 项 | 近似图像 $A_k$ | 误差与感知质量、参数量与文件大小分别评价 |
| 数据降噪 | 有假设支持时舍弃弱方向 | 低秩重构 | 小值不自动表示噪声，阈值需验证 |
| PCA | 中心化后投影到前 $k$ 个右方向 | $Z = X_c V_k$ | 先按特征中心化 |
| 最小二乘 | 对非零奇异值取倒数 | $x = A^{+}b$ | 小值倒数放大对应方向的扰动 |

- 压缩与降噪可用同一截断：“保留”强调表示与误差，“舍弃”强调在结构假设下减弱扰动；操作名称不能保证任务目标
- 先分离方向并排序，再由任务选择操作
- 核对操作产生的输出是否符合任务、条件是否支持操作，据此修正自己的匹配与理由

# 本讲总结

- SVD 用输入方向、伸缩强度和输出方向描述实矩阵；**完整 SVD 和薄 SVD** 给出精确表示，截断 SVD 给出低秩近似
- $A^TA$ 由输入方向的像长度确定右奇异方向和奇异值；非零方向经 $A$ 映射并归一化后得到左奇异方向，零奇异值对应被压到零的输入方向
- 矩阵可写成按奇异值排序的**秩一分量之和**；在秩不超过 $k$ 的约束下，保留前 $k$ 项在谱范数和 Frobenius 范数下达到最小矩阵误差
- 压缩与降噪可采用同一截断运算，**任务目标与结构假设**决定怎样评价结果；PCA 投影到特征主方向，最小二乘反演非零奇异方向，分别得到低维坐标与最小范数解
- 先识别矩阵的**对象含义与任务输出**，再选择方向操作，最后检查相应的误差度量、中心化要求、低秩结构假设和小奇异值敏感性；矩阵近似最优不等同于任意下游任务最优