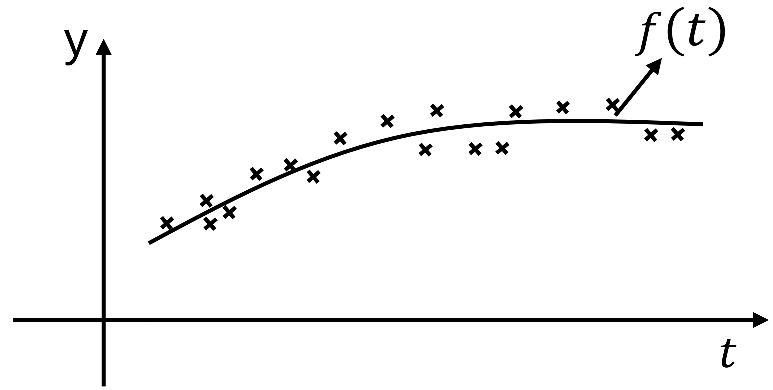
线性最小二乘

- 内容
  - 线性最小二乘法问题
  - 解的存在性和惟一性
  - 问题的敏感性和病态性
  - 求解方法
    - 正规方程组法
    - 增广方程组法
    - 正交变换法
    - 奇异值分解法

线性最小二乘问题

曲线拟合问题

- **Motivation**
  - 发现数据的规律, “回归分析”
  - 由于数据可能有误差, 逼近曲线不必通过所有点



- **问题描述**
  - 数据点(t_i, y_i), $(i=1,\cdots,m)$, 用函数$f(t)$来拟合
  - 拟合要求: $\sum_{i=1}^{m}[f(t_i)-y_i]^2$最小, “最小二乘”
  - 若$f(t)=\sum_{j=1}^{n} x_j \varphi_j(t)$, 求系数$x_j$, <u>线性</u>最小二乘

    通常$m > n$
  - 多项式拟合 $f(t,\boldsymbol{x}) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$

<u>线性:</u> 函数 f 是参数向量 $\boldsymbol{x}$ 的分量的线性函数

线性最小二乘 (Linear Least Squares)

- 线性最小二乘问题的矩阵表述

设$\boldsymbol{b} = [y_1, \cdots, y_m]^T$ ，　　$\boldsymbol{x} = [x_1, \cdots, x_m]^T$

$$\boldsymbol{A} = \begin{bmatrix} \varphi_1(t_1) & \varphi_2(t_1) & \cdots & \varphi_n(t_1) \\ \varphi_1(t_2) & \varphi_2(t_2) & \cdots & \varphi_n(t_2) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_1(t_m) & \varphi_2(t_m) & \cdots & \varphi_n(t_m) \end{bmatrix}$$

$$f(t) \approx \sum_{j=1}^{n} x_j \varphi_j(t)$$

$\sum_{i=1}^{m}[f(t_i) - y_i]^2$最小

$\boldsymbol{A}\boldsymbol{x} \cong \boldsymbol{b}$　　求$\boldsymbol{x}$, 使$\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$最小

解$\boldsymbol{x}$称为该问题的最小二乘解, $f(t) = \sum_{j=1}^{n} x_j \varphi_j(t)$

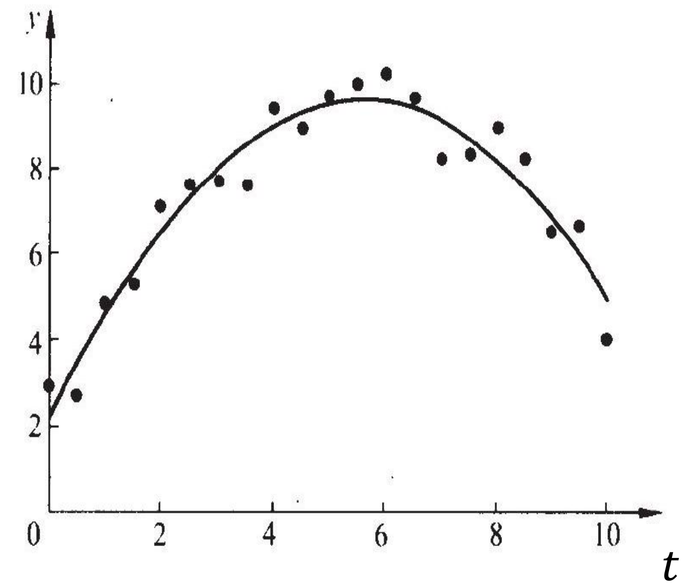
A是$m \times n$矩阵，且$m>n$, 超定方程组, 即方程的个数多于未知数的个数

例　对 21个数据点的二次多项式拟合

二次多项式$f(t,\boldsymbol{x}) = x_1 + x_2 t + x_3 t^2$

矩阵 $\boldsymbol{A}$ 是 21×3

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{21} & t_{21}^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cong \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{21} \end{bmatrix} = \boldsymbol{b}$$

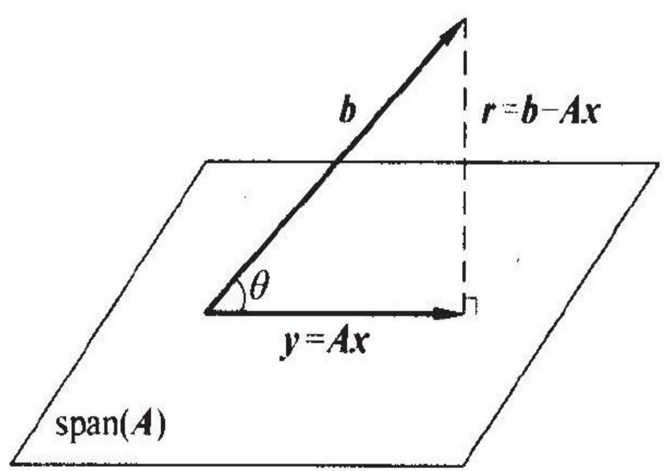


这种矩阵称为范德蒙 (Vandermonde) 矩阵, 它的列 (或行) 是一些独立变量的顺序幂次.

求得的多项式是在所有二次多项式中对给定数据的**最佳拟合**, 即在所有的二次多项式中, 这个多项式与数据点的竖直距离的平方和最小.

线性最小二乘解的存在性和惟一性

- 线性最小二乘问题 $\boldsymbol{Ax} \cong \boldsymbol{b}$ 的解的存在性是有保证的
- 线性最小二乘问题 $\boldsymbol{Ax} \cong \boldsymbol{b}$ 的解惟一的充要条件是 $\boldsymbol{A}$ 列满秩, 即 $\mathrm{rank}(\boldsymbol{A}) = n$.
- 如果 $\mathrm{rank}(\boldsymbol{A}) < n$, 称 $\boldsymbol{A}$ 为**秩亏损**的. 相应的最小二乘问题的解存在,但不惟一



线性最小二乘问题的敏感性和病态性

- $\boldsymbol{A}$ 是列满秩，$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 非奇异，$\boldsymbol{A}$ 的广义逆 $\boldsymbol{A}^{+}=\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}$
- $m\times n$ 阶列满秩矩阵的条件数为 $\mathrm{cond}(\boldsymbol{A})=\|\boldsymbol{A}\|_2\cdot\|\boldsymbol{A}^{+}\|_2$
  - 非方阵的条件数度量了与秩亏损阵的接近程度
- 最小二乘解 $\boldsymbol{x}$ 关于 $\boldsymbol{b}$ 的扰动$\Delta\boldsymbol{b}$的条件数不仅与 $\mathrm{cond}(\boldsymbol{A})$ 有关, 还与 $\boldsymbol{b}$ 和 $\boldsymbol{Ax}$ 的夹角 θ 有关

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}\leqslant\mathrm{cond}(\boldsymbol{A})\cdot\frac{1}{\cos\theta}\cdot\frac{\|\Delta\boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2}$$

- 最小二乘解 $\boldsymbol{x}$ 关于矩阵 $\boldsymbol{A}$ 的扰动$\boldsymbol{E}$的条件数不仅与 $\mathrm{cond}(\boldsymbol{A})$ 有关, 还与 $\boldsymbol{b}$ 和 $\boldsymbol{Ax}$ 的夹角 θ 有关

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}\lesssim\left([\mathrm{cond}(\boldsymbol{A})]^2\tan\theta+\mathrm{cond}(\boldsymbol{A})\right)\frac{\|\boldsymbol{E}\|_2}{\|\boldsymbol{A}\|_2}$$

例 测量员要测量在某个基准点上 3 个山头的高度. 首先从基准点观测, 测得高度分别为 $x_1 = 1237\text{ft}, x_2 = 1914\text{ft}, x_3 = 2417\text{ft}$. (ft :英尺)。为进一步确认初始的测量数据, 测量员爬上第一座小山, 测得第二座小山相对于第一座小山的高度为 $x_2 - x_1 = 711\text{ft}$, 第三座相对于第一座的高度为 $x_3 - x_1 = 1177\text{ft}$. 最后, 测量员爬上第二座小山, 测得第三座小山相对于第二座小山的高度是 $x_3 - x_2 = 475\text{ft}$. 这些测量值对应一个超定方程组

$$\boldsymbol{Ax} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cong \begin{bmatrix} 1237 \\ 1941 \\ 2417 \\ 711 \\ 1177 \\ 475 \end{bmatrix} = \boldsymbol{b},$$

$$\boldsymbol{A}^{+} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}} = \frac{1}{4}\begin{bmatrix} 2 & 1 & 1 & -1 & -1 & 0 \\ 1 & 2 & 1 & 1 & 0 & -1 \\ 1 & 1 & 2 & 0 & 1 & 1 \end{bmatrix}$$

$$\| \boldsymbol{A} \|_2 = 2, \ \|\boldsymbol{A}^{+}\|_2 = 1, \ \ \mathrm{cond}(\boldsymbol{A}) = \| \boldsymbol{A} \|_2 \cdot \|\boldsymbol{A}^{+}\|_2 = 2.$$

$$\cos\theta = \frac{\| \boldsymbol{Ax} \|_2}{\| \boldsymbol{b} \|_2} = \frac{\| \boldsymbol{y} \|_2}{\| \boldsymbol{b} \|_2} \approx \frac{3640.8761}{3640.8809} \approx 0.99999868,$$

夹角大约为 0.001625, 由于条件数和角度 θ 都非常小,所以这个最小二乘问题是良态的.

线性最小二乘的求解

正规方程组法

目标是残差向量 $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$ 的欧几里得范数的平方取得最小值

目标函数是 $\phi: \mathbb{R}^n \to \mathbb{R}$,

$$\phi(\boldsymbol{x}) = \| \boldsymbol{r} \|_2^2 = \boldsymbol{r}^{\mathrm{T}}\boldsymbol{r} = (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})$$
$$= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} - 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$$

取最小值的必要条件是 $\boldsymbol{x}$ 点的梯度向量 $\nabla\phi(\boldsymbol{x})$ 为零, $\nabla\phi(\boldsymbol{x})$ 的第 i 个分量为 $\partial\phi(\boldsymbol{x})/\partial x_i$,

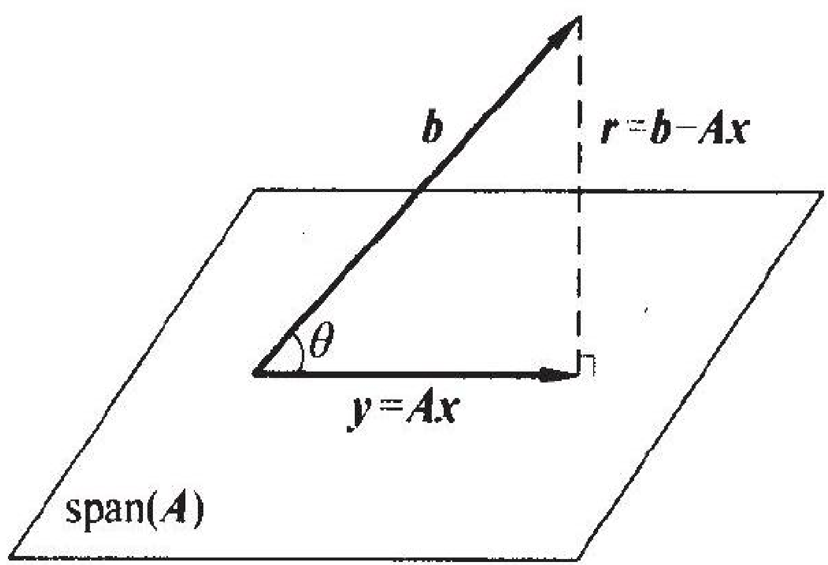
$$\boldsymbol{0} = \nabla\phi(\boldsymbol{x}) = 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

因此 ϕ 的最小值点 $\boldsymbol{x}$ 一定满足 $n \times n$ 的对称线性方程组(**正规方程组**)

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

对于 $m > n$ 的最小二乘问题, m 维向量 $\boldsymbol{b}$ 一般不属于 span$(\boldsymbol{A})$, 子空间 span$(\boldsymbol{A})$ 的维数最大为 n. 当残差向量 $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$ 垂直于 span$(\boldsymbol{A})$ 时, 向量 $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in$ span$(\boldsymbol{A})$ 在欧几里得范数意义下是最接近 $\boldsymbol{b}$ 的. 这样, 对于最小二乘解 $\boldsymbol{x}$, 残差向量 $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$一 定垂直于 $\boldsymbol{A}$ 的各列, 因此

$$\boldsymbol{0} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{r} = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}), \quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$



例

$$\boldsymbol{Ax}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\-1&1&0\\-1&0&1\\0&-1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\cong\begin{bmatrix}1237\\1941\\2417\\711\\1177\\475\end{bmatrix}=\boldsymbol{b}$$

正规方程组为

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\begin{bmatrix}3&-1&-1\\-1&3&-1\\-1&-1&3\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}-651\\2177\\4069\end{bmatrix}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

利用楚列斯基分解得到

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}},$$

其中 $\boldsymbol{L}$ 是下三角阵，求解 $\boldsymbol{Ly}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 和 $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{y}$ 得到

$\boldsymbol{x}^{\mathrm{T}}=[1236,1943,2416]^{\mathrm{T}}$，对应的残差平方和的最小值 $\|\boldsymbol{r}\|_2^2=35.$

问题转换的过程: 长方阵 → 方阵 → 三角阵.

理论上合理的转换方法数值上不一定可靠. 理论上, 由正规方程组可以得到线性最小二乘问题的精确解, 但实际上这种方法得到的解往往令人失望.

- 形成矩阵$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$和右端向量$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$时会产生信息丢失. 例如,

$$\boldsymbol{A}=\begin{bmatrix}1 & 1\\ \varepsilon & 0\\ 0 & \varepsilon\end{bmatrix}, \text{其中 } 0<\varepsilon<\sqrt{\varepsilon_{\text{mach}}}\,.$$

利用浮点计算, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\begin{bmatrix}1+\boldsymbol{\varepsilon}^2 & 1\\ 1 & 1+\varepsilon^2\end{bmatrix}=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$, 这是一个奇异矩阵.

- 矩阵$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$的条件数是原矩阵$\boldsymbol{A}$条件数的平方,

$$\operatorname{cond}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)=[\operatorname{cond}(\boldsymbol{A})]^2,$$

这个条件数决定了正规方程组解的敏感性. 即使在拟合良好, 并且残差也较小的情况下, 正规方程组也会出现条件数平方效应, 这就使得计算解的敏感性比原来的最小二乘问题更强. 在这个意义上**正规方程组解法是不稳定的**.

增广方程组法

令残差向量 $\boldsymbol{r}$ 与 $\boldsymbol{A}$ 的各列正交, 得到方程组

$$\boldsymbol{r} + \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$
$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{r} = \boldsymbol{0},$$

写成矩阵形式, 称为 $(m+n)\times(m+n)$ **增广方程组**

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{A} \\ \boldsymbol{A}^{\mathrm{r}} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{bmatrix}.$$

由这个方程既可得到所需的解 $\boldsymbol{x}$, 又可得到这个解的残差向量 $\boldsymbol{r}$.

- 增广方程组对称但不正定, 而且方程组变大，尽管可以利用增广矩阵的特殊结构(稀疏性)
- 直接选主元会再次产生正规方程组, 数值稳定性很差. 可以采用其他的选主元策略, 以保证数值稳定.

上三角最小二乘问题的求解

$$\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{x} \cong \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix}, \quad \boldsymbol{R} \text{ 是 } n \times n \text{ 上三角阵}$$

$$\| \boldsymbol{r} \|_2^2 = \|\boldsymbol{c}_1 - \boldsymbol{R}\boldsymbol{x}\|_2^2 + \|\boldsymbol{c}_2\|_2^2.$$

式中第二项与 $\boldsymbol{x}$ 无关, 可以选择使第一项为零的 $\boldsymbol{x}$,使 $\| \boldsymbol{r} \|_2^2$ 取极小值. 此时, $\boldsymbol{x}$ 应满足上三角方程组

$$\boldsymbol{R}\boldsymbol{x} = \boldsymbol{c}_1,$$

可以用回代法求解出 $\boldsymbol{x}$. 这样就得到了最小二乘解 $\boldsymbol{x}$ 及 $\| \boldsymbol{r} \|_2^2 = \|\boldsymbol{c}_2\|_2^2$.

正交变换法

用高斯消去法将矩阵约化成三角阵的形式并不合适, 因为这种变换并不能保证欧几里得范数不变, 不能保持最小二乘解不变.

需要一种能保持欧几里得范数不变的线性变换.

- 如果一个实方阵 $\boldsymbol{Q}$ 的各列正交, 即 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{I}$, 则称 $\boldsymbol{Q}$ 为正交阵.
- **正交变换不改变向量的欧几里得范数**:

$$\|\boldsymbol{Qv}\|_2^2=(\boldsymbol{Qv})^{\mathrm{T}}\boldsymbol{Qv}=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Qv}=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}=\|\boldsymbol{v}\|_2^2$$

正交阵可以对向量作各种方式的变换, 像旋转、反射, 但它并不改变向量的欧几里得长度.

- 若将线性最小二乘方程组两端同时乘上正交阵,最小二乘解并不改变.

QR 分解

- $m \times n$, 且 $m > n$ 的矩阵 $\boldsymbol{A}$ 的QR 分解:

$$\boldsymbol{A} = \boldsymbol{Q}\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix}$$

其中 $\boldsymbol{Q}$ 是 $m \times m$ 正交矩阵, $\boldsymbol{R}$ 是 $n \times n$ 上三角阵.

- QR分解将线性最小二乘问题 $\boldsymbol{Ax} \cong \boldsymbol{b}$ 转化成**上三角最小二乘问题**,但解不变。

设变换后的右端向量为 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix}$,

$$\| \boldsymbol{r} \|_2^2 = \| \boldsymbol{b} - \boldsymbol{Ax} \|_2^2 = \left\| \boldsymbol{b} - \boldsymbol{Q}\begin{bmatrix} \boldsymbol{R} \\ 0 \end{bmatrix}\boldsymbol{x} \right\|_2^2 = \left\| \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} - \begin{bmatrix} \boldsymbol{R} \\ 0 \end{bmatrix}\boldsymbol{x} \right\|_2^2 = \| \boldsymbol{c}_1 - \boldsymbol{Rx} \|_2^2 + \| \boldsymbol{c}_2 \|_2^2$$

其中 $\boldsymbol{c}_1$ 是 n 维向量. 解 $\boldsymbol{x}$ 满足 $n \times n$ 三角线性方程组 $\boldsymbol{Rx} = \boldsymbol{c}_1$

最小残差的范数 $\| \boldsymbol{r} \|_2 = \|\boldsymbol{c}_2\|_2$.

矩阵奇异值/奇异向量

- m×n实矩阵$\boldsymbol{A}$的奇异值 (可看成非方阵的“特征值”)

$$\begin{cases} \boldsymbol{A}\boldsymbol{v} = \sigma\boldsymbol{u} \\ \boldsymbol{A}^T\boldsymbol{u} = \sigma\boldsymbol{v} \end{cases}$$

非负实数σ为奇异值(singular value)

非零向量u, v为左/右奇异向量(singular vector)

$$\Updownarrow$$

$$\boldsymbol{u}^T\boldsymbol{A} = \sigma\boldsymbol{v}^T$$

奇异向量对有无穷多个. 一般取单位化的向量: $\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1$

- 设m≥n, 矩阵$\boldsymbol{A}$可以分解为$\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^{\mathrm{T}}$, $\boldsymbol{U}, \boldsymbol{V}$为正交阵, $\boldsymbol{\Sigma}$为对角阵, 其对角元$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & \end{bmatrix}$$

$\Rightarrow$ $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\Sigma$, $\boldsymbol{A}\boldsymbol{v}_k = \sigma_k\boldsymbol{u}_k$, $k = 1, \cdots, n$

同时 $\boldsymbol{A}^T = \boldsymbol{V}\Sigma^T\boldsymbol{U}^T$, $\boldsymbol{A}^T\boldsymbol{U} = \boldsymbol{V}\Sigma^T$, $\boldsymbol{A}^T\boldsymbol{u}_k = \sigma_k\boldsymbol{v}_k$

u_k/v_k为左/右奇异向量(能找到n对),且$\{u_k\}$和$\{v_k\}$各自两两正交

- 当$\boldsymbol{A}$为实对称半正定阵时, 奇异值分解 = 特征值分解

例　奇异值分解

$\boldsymbol{A}=\begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\\ 10 & 11 & 12\end{bmatrix}$的奇异值分解为

$$\boldsymbol{U\Sigma V}^{\mathrm{T}}=\begin{bmatrix}0.141 & 0.825 & -0.420 & -0.351\\ 0.344 & 0.426 & 0.298 & 0.782\\ 0.547 & 0.028 & 0.664 & -0.509\\ 0.750 & -0.371 & -0.542 & 0.079\end{bmatrix}\begin{bmatrix}25.5 & 0 & 0\\ 0 & 1.29 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}0.504 & 0.574 & 0.644\\ -0.761 & -0.057 & 0.646\\ 0.408 & -0.816 & 0.408\end{bmatrix}$$

$\sigma_1=25.5, \sigma_2=1.29, \sigma_3=0.$

有个奇异值为零, 说明矩阵是秩亏损的.

矩阵的秩等于其非零奇异值的个数, 矩阵A的秩为 2 .

奇异值分解(SVD)法

设 $\boldsymbol{A}$ 是 $m\times n$满秩矩阵, 且 $\operatorname{rank}(\boldsymbol{A})=n$, $\boldsymbol{A}$ 的 SVD 为

$$\boldsymbol{A}=\boldsymbol{U\Sigma V}^{\mathrm{T}}=[\boldsymbol{U}_1\quad \boldsymbol{U}_2]\begin{bmatrix}\boldsymbol{\Sigma}\\ \boldsymbol{0}\end{bmatrix}\boldsymbol{V}^{\mathrm{T}}=\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}^{\mathrm{T}},$$

其中 $\boldsymbol{U}_1$ 是 $m\times n$ 矩阵, $\boldsymbol{\Sigma}_1$ 是 $n\times n$ 非奇异矩阵.

最小二乘问题 $\boldsymbol{Ax}\cong\boldsymbol{b}$ 的解满足: $\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\cong\boldsymbol{b}$

$$\boldsymbol{x}=\boldsymbol{V}\boldsymbol{\Sigma}_1{}^{-1}\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{b}=\sum_i\frac{\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{b}}{\sigma_i}\boldsymbol{v}_i$$

其证明只需将 $\boldsymbol{A}$ 的 SVD 代入正规方程.

方法间的比较

- **正规方程组法**易于实现, 只需要矩阵乘法和楚列斯基分解，产生的解的相对误差与 $[\mathrm{cond}(\boldsymbol{A})^2]$ 成比例. 当问题充分良态时,正规方程组方法可提供充分精确的解。
- 对于接近方阵即 $m \approx n$ 的问题, 正规方程组方法和正交变换法所需的工作量大约相同. 但对于高度超定即 $m \gg n$ 的问题,**正交变换法**的工作量大约为正规方程组方法的两倍，但精度更高，应用更广泛。
- **SVD**计算量大，具有极强稳健性和可靠性，在问题特别敏感的情形才使用．

矩阵的QR分解

- 豪斯霍尔德(Householder)变换（初等反射）
- 吉文斯(Givens)变换（平面旋转）
- 格拉姆－施密特(Gram-Schmidt)正交化

格拉姆－施密特(Gram-Schmidt)正交化

$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$线性无关

$$\boldsymbol{\beta}_1 = \boldsymbol{\alpha}_1,\ \boldsymbol{\varepsilon}_1 = \frac{\boldsymbol{\beta}_1}{|\boldsymbol{\beta}_1|},$$

$$\boldsymbol{\beta}_2 = \boldsymbol{\alpha}_2 - \langle \boldsymbol{\alpha}_2, \boldsymbol{\varepsilon}_1 \rangle \boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_2 = \frac{\boldsymbol{\beta}_2}{|\boldsymbol{\beta}_2|},$$

$\{\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n\}$ 两两正交，

……

且是单位向量。

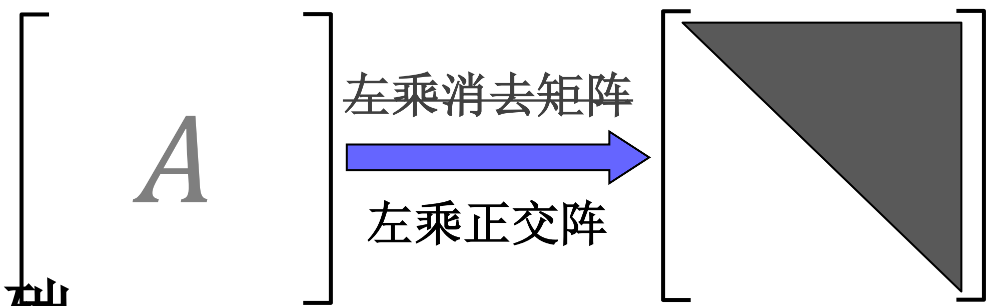
$$\boldsymbol{\beta}_n = \boldsymbol{\alpha}_n - \langle \boldsymbol{\alpha}_n, \boldsymbol{\varepsilon}_1 \rangle \boldsymbol{\varepsilon}_1 - \cdots - \langle \boldsymbol{\alpha}_n, \boldsymbol{\varepsilon}_{n-1} \rangle \boldsymbol{\varepsilon}_{n-1}, \boldsymbol{\varepsilon}_n = \frac{\boldsymbol{\beta}_n}{|\boldsymbol{\beta}_n|}$$

$$\begin{cases} \boldsymbol{\alpha}_1 = \boldsymbol{\beta}_1 = |\boldsymbol{\beta}_1| \boldsymbol{\varepsilon}_1, \\ \boldsymbol{\alpha}_2 = \langle \boldsymbol{\alpha}_2, \boldsymbol{\varepsilon}_1 \rangle \boldsymbol{\varepsilon}_1 + \boldsymbol{\beta}_2 = \langle \boldsymbol{\alpha}_2, \boldsymbol{\varepsilon}_1 \rangle \boldsymbol{\varepsilon}_1 + |\boldsymbol{\beta}_2| \boldsymbol{\varepsilon}_2, \\ \quad \vdots \\ \boldsymbol{\alpha}_n = \langle \boldsymbol{\alpha}_n, \boldsymbol{\varepsilon}_1 \rangle \boldsymbol{\varepsilon}_1 + \cdots + \langle \boldsymbol{\alpha}_n, \boldsymbol{\varepsilon}_{n-1} \rangle \boldsymbol{\varepsilon}_{n-1} + \boldsymbol{\beta}_n \\ \quad = \langle \boldsymbol{\alpha}_n, \boldsymbol{\varepsilon}_1 \rangle \boldsymbol{\varepsilon}_1 + \cdots + \langle \boldsymbol{\alpha}_n, \boldsymbol{\varepsilon}_{n-1} \rangle \boldsymbol{\varepsilon}_{n-1} + |\boldsymbol{\beta}_n| \boldsymbol{\varepsilon}_n. \end{cases}$$

$$[\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \quad \cdots \quad \boldsymbol{\alpha}_n] = [\boldsymbol{\varepsilon}_1 \quad \boldsymbol{\varepsilon}_2 \quad \cdots \quad \boldsymbol{\varepsilon}_n] \boldsymbol{R} \qquad \boldsymbol{R} = \begin{bmatrix} |\boldsymbol{\beta}_1| & \langle \boldsymbol{\alpha}_2, \boldsymbol{\varepsilon}_1 \rangle & \cdots & \langle \boldsymbol{\alpha}_n, \boldsymbol{\varepsilon}_1 \rangle \\ 0 & |\boldsymbol{\beta}_2| & \cdots & \langle \boldsymbol{\alpha}_n, \boldsymbol{\varepsilon}_2 \rangle \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \langle \boldsymbol{\alpha}_n, \boldsymbol{\varepsilon}_{n-1} \rangle \\ 0 & 0 & \cdots & |\boldsymbol{\beta}_n| \end{bmatrix}$$

Householder变换

- 矩阵的正交三角化
  - 高斯消去过程?
  - 可用正交阵来乘吗?
  - 是矩阵特征值计算、曲线拟合等解法的基础



- Householder矩阵

$\|\boldsymbol{w}\|_2 = 1$

  - 定义 $\boldsymbol{w} \in \mathbb{R}^n$且$\boldsymbol{w}^T\boldsymbol{w} = 1$, 称$\boldsymbol{H}(\boldsymbol{w}) = \boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T$为Householder矩阵 (初等反射阵)
  - $\boldsymbol{H}(\boldsymbol{w}) = \boldsymbol{H}(-\boldsymbol{w})$
  - $\boldsymbol{H}$为对称阵、正交阵
  - $\boldsymbol{H}\boldsymbol{x}$实现Householder变换

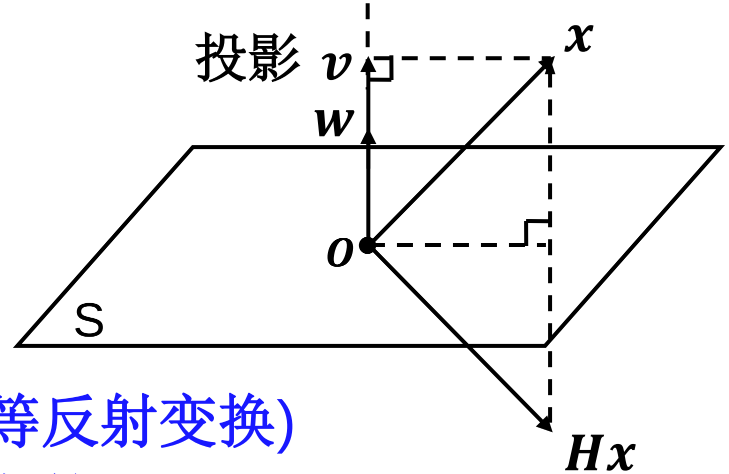
$$\boldsymbol{H} = \begin{bmatrix} 1-2w_1^2 & -2w_1w_2 & \cdots & -2w_1w_n \\ -2w_2w_1 & 1-2w_2^2 & \cdots & -2w_2w_n \\ \vdots & \vdots & \cdots & \vdots \\ -2w_nw_1 & -2w_nw_2 & \cdots & 1-2w_n^2 \end{bmatrix}$$

Householder变换

- Householder变换的几何意义
  - $\boldsymbol{Hx}$: 以$\boldsymbol{w}$为法向画出超平面S

$$\boldsymbol{Hx} = \left(\boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T\right)\boldsymbol{x} = \boldsymbol{x} - 2\boldsymbol{w}\boldsymbol{w}^T\boldsymbol{x}$$

$$\boldsymbol{w}\boldsymbol{w}^T\boldsymbol{x} = \left(\boldsymbol{w}^T\boldsymbol{x}\right)\boldsymbol{w} = \boldsymbol{v} \Longrightarrow \boldsymbol{Hx} = \boldsymbol{x} - 2\boldsymbol{v}$$



  - $\boldsymbol{Hx}$为$\boldsymbol{x}$关于平面S的镜像　(初等反射变换)
  - $\boldsymbol{Hx}$与$\boldsymbol{x}$的2-范数相等, 属于正交变换
- **Th5.18** 设$\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n, \boldsymbol{x} \neq \boldsymbol{y}, \|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2$, 则存在Householder矩阵$\boldsymbol{H}$, 使$\boldsymbol{Hx} = \boldsymbol{y}$

  几何的启示: $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{y}$, $\boldsymbol{w} = \boldsymbol{v}/\|\boldsymbol{v}\|_2$，构造矩阵$\boldsymbol{H}$
- **Th5.19** 可将定理**5.18**中的$\boldsymbol{y}$设为 $\begin{bmatrix} \pm\|\boldsymbol{x}\|_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ 这样就用正交变换实现“消元”！

  取$\boldsymbol{y} = -\sigma\boldsymbol{e}_1$, $\sigma = \mathrm{sign}(x_1)\|\boldsymbol{x}\|_2$较好

  构造$\boldsymbol{H}$时, $\boldsymbol{v} = \boldsymbol{x} + \sigma\boldsymbol{e}_1$

Householder变换

$$x \xrightarrow{H} -\sigma e_1 \qquad \begin{bmatrix} * \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\sigma = \mathrm{sign}(x_1)\|x\|_2$$

- 正交变换实现消元
  - 对向量x做Householder变换, 结果$-\sigma e_1$中的负号是为了数值稳定

例: 确定一个Householder变换, 对向量实现消元操作

$$a = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \Longrightarrow \begin{bmatrix} * \\ 0 \\ 0 \end{bmatrix}$$

解: $\sigma = \mathrm{sign}(a_1)\|a\|_2 = 3$， 构造$v = a + \sigma e_1 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}$

取$w = v/\|v\|_2$，则实现变换的矩阵为$H = I - 2ww^T$

验证: $Ha = a - 2(w^T a)w = a - 2\dfrac{v^T a}{v^T v}v = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} - 2 \times \dfrac{15}{30} \times \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 0 \end{bmatrix}$

用v或w计算Hx时只算向量内积

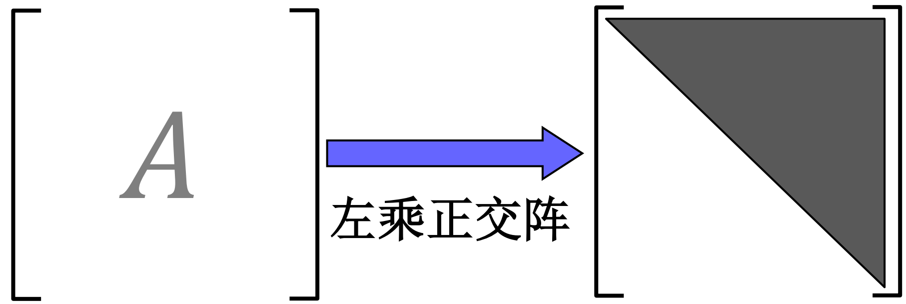
矩阵的QR分解

- 用Householder变换实现正交三角化

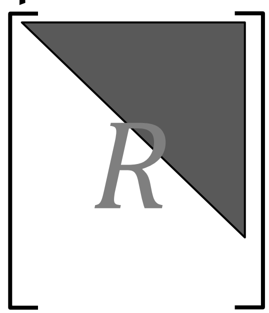
  - □ $\boldsymbol{H}_k \cdots \boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A} = \boldsymbol{R}$

    $\boldsymbol{A} \in \mathbb{R}^{m\times n}$不一定是方阵,
    上三角阵$\boldsymbol{R}$也同样

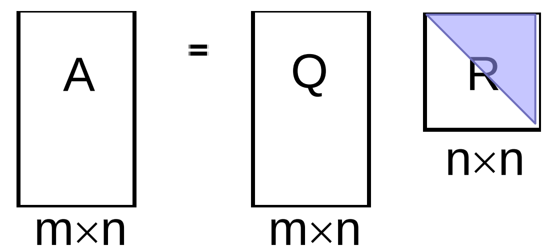
$m = n$的情况:



$m > n$的情况:



经济的QR分解



  - □ $\boldsymbol{A} = \boldsymbol{H}_1\boldsymbol{H}_2 \cdots \boldsymbol{H}_k\boldsymbol{R} = \boldsymbol{QR}$, 其中$\boldsymbol{Q}$为正交阵, 称为QR分解

Th5.20 对任意实矩阵$\boldsymbol{A}$, 一定存在QR分解;

若$\boldsymbol{A}$为非奇异方阵, 且要求$\boldsymbol{R}$的对角元都>0, 此分解唯一.

矩阵的QR分解

- 怎么实现: $\boldsymbol{H}_k \cdots \boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A} = \boldsymbol{R}$
  - 设$\boldsymbol{A} \in \mathbb{R}^{m\times n}, m \geq n$, $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n]$, 其中$\boldsymbol{a}_j$为$m$维向量

构造m阶反射阵$\boldsymbol{H}_1$

消$\boldsymbol{A}$的第1列:

$$\boldsymbol{A}^{(2)} = \boldsymbol{H}_1\boldsymbol{A} = \begin{bmatrix} -\sigma_1 & a_{12}^{(2)} & \cdots & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{m2}^{(2)} & \cdots & a_{mn}^{(2)} \end{bmatrix} = \begin{bmatrix} -\sigma_1 & | & & | \\ 0 & \boldsymbol{H}_1\boldsymbol{a}_2 & \cdots & \boldsymbol{H}_1\boldsymbol{a}_n \\ \vdots & | & & | \\ 0 & | & & | \end{bmatrix}$$

用$\boldsymbol{H}_2$对$\boldsymbol{A}^{(2)}$第2列消元, 同时不影响第1列

令$\boldsymbol{H}_2 = \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{H}_2' \end{bmatrix}$ $\quad \boldsymbol{H}_2\boldsymbol{A}^{(2)} = \begin{bmatrix} -\sigma_1 & \boldsymbol{r}_1^T \\ \boldsymbol{0} & \boldsymbol{H}_2'\boldsymbol{A}'^{(2)} \end{bmatrix} = \left[\begin{array}{c:cccc} -\sigma_1 & & \boldsymbol{r}_1^T & & \\ \hdashline 0 & -\sigma_2 & | & & | \\ 0 & 0 & \boldsymbol{H}_2'\boldsymbol{a}_2'^{(2)} & \cdots & \boldsymbol{H}_2'\boldsymbol{a}_{n-1}'^{(2)} \\ \vdots & \vdots & | & & | \\ 0 & 0 & | & & | \end{array}\right]$

记$\boldsymbol{A}^{(2)} = \begin{bmatrix} -\sigma_1 & \boldsymbol{r}_1^T \\ \boldsymbol{0} & \boldsymbol{A}'^{(2)} \end{bmatrix}$ $\qquad$ 构造反射阵$\boldsymbol{H}_2'$

$\boldsymbol{H}_2$也是Householder阵(其向量$\boldsymbol{v}$的第一个分量为0).后续$\boldsymbol{H}_j$可类似构造

算法5.3：基于Householder变换的矩阵正交三角化 $(\text{设} m \geq n)$

输入: $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n]$；输出: $\boldsymbol{A}; \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$.

For k=1, 2, …, *n*

  $\sigma_k := \mathrm{sign}(a_{kk}) \sqrt{\sum_{j=k}^{m} a_{jk}^2}$; {下三角部分第k列的2-范数 }

  If $\sigma_k = a_{kk}$ **then** {第k列对角线下方已经全为0 }

    Continue with next k ;

  End

  $\boldsymbol{v}_k := [0, \cdots, 0, a_{kk}, \cdots, a_{mk}]^T + \sigma_k \boldsymbol{e}_k$; {构造$\boldsymbol{H}_k$的向量}

  $\beta_k := \boldsymbol{v}_k^T \boldsymbol{v}_k$;

  For j=k, k+1, …, n {对剩余各列作Householder变换 }

    $\gamma_j := \boldsymbol{v}_k^T \boldsymbol{a}_j$;

    $\boldsymbol{a}_j := \boldsymbol{a}_j - (2\gamma_j / \beta_k) \boldsymbol{v}_k$; $\{ \boldsymbol{H}_k \boldsymbol{a}_j^{(k)} = \boldsymbol{a}_j^{(k)} - 2\left(\boldsymbol{v}_k^T \boldsymbol{a}_j^{(k)} / \boldsymbol{v}_k^T \boldsymbol{v}_k\right) \boldsymbol{v}_k \}$

  End

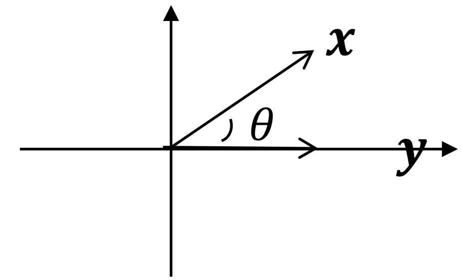
End

- 算法执行完, $\boldsymbol{A}$变成$\boldsymbol{R}$, 得到构造$\boldsymbol{H}_1, \boldsymbol{H}_2, \cdots, \boldsymbol{H}_k$所需的$\boldsymbol{v}$向量
- 总乘法次数: $(2n+1) \cdot m + (2n-1) \cdot m + \cdots + 3m \approx mn^2$
- 没有生成$\boldsymbol{Q}$矩阵. 如果生成, 则计算量为$O(mn^2)$

Matlab: [Q,R] = qr(A); R = qr(A)

Givens旋转变换

- 二维平面旋转变换
  - □ 将向量$\boldsymbol{x}$顺时针旋转θ角度后得到$\boldsymbol{y}$



$$\boldsymbol{y}=\begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}\boldsymbol{x}$$

  - □ 二阶Givens矩阵为: $\boldsymbol{G}=\begin{bmatrix}c & s\\ -s & c\end{bmatrix}$, 其中$c=\cos\theta, s=\sin\theta$
  - □ $\boldsymbol{G}$是2x2的正交阵, 实施平面旋转(称为Givens旋转)
  - □ 实现消元: 选c和s值可使$\boldsymbol{G}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}\alpha\\ 0\end{bmatrix}$, $\boxed{c=\frac{x_1}{\alpha}, s=\frac{x_2}{\alpha}}$
- n阶Givens旋转阵

将2阶Givens阵
嵌入n阶单位阵:
$\boldsymbol{G}$仍是正交阵

$$\underbrace{\begin{bmatrix}1 & 0 & 0 & 0 & 0\\ 0 & \boldsymbol{c} & 0 & \boldsymbol{s} & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & -\boldsymbol{s} & 0 & \boldsymbol{c} & 0\\ 0 & 0 & 0 & 0 & 1\end{bmatrix}}_{G}\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\\ x_5\end{bmatrix}=\begin{bmatrix}x_1\\ \boldsymbol{\alpha}\\ x_3\\ \boldsymbol{0}\\ x_5\end{bmatrix}$$

$$c=\frac{x_2}{\alpha}$$

$$s=\frac{x_4}{\alpha}$$

Givens旋转变换

例: 用Givens旋转变换来消元 $\boldsymbol{a}=\begin{bmatrix}2\\0\\1\\2\end{bmatrix} \Longrightarrow \begin{bmatrix}*\\0\\0\\0\end{bmatrix}$

解: 先针对第1, 3分量构造二阶旋转矩阵,

$$\boldsymbol{G}_1'=\begin{bmatrix}c_1 & s_1\\-s_1 & c_1\end{bmatrix} \qquad c_1=\frac{2}{\sqrt{2^2+1}}=2/\sqrt{5}, \qquad s_1=1/\sqrt{5}$$

则: $$\boldsymbol{G}_1\boldsymbol{a}=\begin{bmatrix}c_1 & 0 & s_1 & 0\\0 & 1 & 0 & 0\\-s_1 & 0 & c_1 & 0\\0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}2\\0\\1\\2\end{bmatrix}=\begin{bmatrix}\sqrt{5}\\0\\0\\2\end{bmatrix}$$

接着对第1, 4分量旋转

求出 $c_2=\sqrt{5}/3,\ s_2=2/3$，$$\boldsymbol{G}_2\boldsymbol{G}_1\boldsymbol{a}=\begin{bmatrix}c_2 & 0 & 0 & s_2\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\\-s_2 & 0 & 0 & c_2\end{bmatrix}\begin{bmatrix}\sqrt{5}\\0\\0\\2\end{bmatrix}=\begin{bmatrix}3\\0\\0\\0\end{bmatrix}$$

- 每个Givens旋转用参数c, s刻画
- 每次旋转仅影响向量两个元素, 仅影响矩阵的两行
- 可实现矩阵的**QR**分解, 适合于稀疏矩阵

线性最小二乘

例: 一组数据如下表, 用适当的函数对它们进行拟合

t_i	1.00	1.25	1.50	1.75	2.00
y_i	5.10	5.79	6.53	7.45	8.46
$\tilde{y}_i$	1.6292	1.7561	1.8764	2.0082	2.1353

解: 在直角坐标系里绘出这些数据点, 根据其分布趋势, 采用指数函数来描述: $y \approx x_1 e^{x_2 t}$

不能直接用线性最小二乘, 需做变换 $\ln y \approx \ln x_1 + x_2 t$

$$\tilde{y} = \tilde{x}_1 + x_2 t$$

$$A = \begin{bmatrix} 1 & 1.00 \\ 1 & 1.25 \\ 1 & 1.5 \\ 1 & 1.75 \\ 1 & 2.00 \end{bmatrix}, \quad b = \begin{bmatrix} 1.6292 \\ 1.7561 \\ 1.8764 \\ 2.0082 \\ 2.1353 \end{bmatrix}$$

要解 $\boldsymbol{A}^T \boldsymbol{A} \boldsymbol{x} = \boldsymbol{A}^T \boldsymbol{b}$, 即

$$\begin{bmatrix} 5 & 7.5 \\ 7.5 & 11.875 \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 9.4052 \\ 14.4239 \end{bmatrix}$$

解得:

$$\begin{bmatrix} \tilde{x}_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.1225 \\ 0.5057 \end{bmatrix}$$

$$x_1 = e^{\tilde{x}_1} = 3.0725$$

最后的解为 $f(t) = 3.0725 e^{0.5057t}$

线性最小二乘

解$\boldsymbol{Ax} \cong \boldsymbol{b}$

$$\boldsymbol{A}=\begin{bmatrix}1 & 1.00\\ 1 & 1.25\\ 1 & 1.5\\ 1 & 1.75\\ 1 & 2.00\end{bmatrix},\quad \boldsymbol{b}=\begin{bmatrix}1.6292\\ 1.7561\\ 1.8764\\ 2.0082\\ 2.1353\end{bmatrix}\qquad \tilde{y}=\tilde{x}_1+x_2 t$$

用Householder变换对$\boldsymbol{A}$作正交三角化

$\sigma_1=\sqrt{5}$ 第一个变换用的 $\boldsymbol{v}_1=\begin{bmatrix}1+\sqrt{5}\\ 1\\ 1\\ 1\\ 1\end{bmatrix}$, 则 $\begin{bmatrix}-2.236 & -3.354\\ 0 & -0.095\\ 0 & 0.155\\ 0 & 0.405\\ 0 & 0.655\end{bmatrix}\boldsymbol{x}\cong\begin{bmatrix}-4.206\\ -0.047\\ 0.073\\ 0.205\\ 0.332\end{bmatrix}$

再做第二个变换得 $\begin{bmatrix}-2.236 & -3.354\\ 0 & 0.791\\ 0 & 0\\ 0 & 0\\ 0 & 0\end{bmatrix}\boldsymbol{x}\cong\begin{bmatrix}-4.206\\ 0.400\\ -0.005\\ 0.001\\ 0.002\end{bmatrix}$

解得:

$$\begin{bmatrix}\tilde{x}_1\\ x_2\end{bmatrix}=\begin{bmatrix}1.1225\\ 0.5057\end{bmatrix}$$

线性最小二乘

- **Matlab命令**
  - 线性最小二乘问题$\boldsymbol{Ax} \cong \boldsymbol{b}$,　$\boldsymbol{A}$是$m \times n$矩阵, $m > n$

```
>> x=A \ b
```

  - 用不超过n次的多项式拟合离散点(x_i, y_i)

```
>> p = polyfit (x, y, n)
```

拟合多项式的系数存于向量p, 相关命令还有polyval

矩阵奇异值分解

矩阵奇异值/奇异向量

- m×n实矩阵$\boldsymbol{A}$的奇异值　　(可看成非方阵的"特征值")

$$\begin{cases} \boldsymbol{A}\boldsymbol{v} = \sigma\boldsymbol{u} \\ \boldsymbol{A}^T\boldsymbol{u} = \sigma\boldsymbol{v} \end{cases}$$

非负实数σ为奇异值(singular value)

非零向量u, v为左/右奇异向量(singular vector)

$$\boldsymbol{u}^T\boldsymbol{A} = \sigma\boldsymbol{v}^T$$

奇异向量对有无穷多个. 一般考虑单位化的:　$\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1$

- □ **Th5.23**设m≥n, 矩阵$\boldsymbol{A}$可以分解为 $\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^T$, $\boldsymbol{U}, \boldsymbol{V}$为正交阵, $\boldsymbol{\Sigma}$为对角阵, 其对角元$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\Sigma, \ \boldsymbol{A}\boldsymbol{v}_k = \sigma_k\boldsymbol{u}_k, \ k = 1, \cdots, n \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & \end{bmatrix}$$

同时 $\boldsymbol{A}^T = \boldsymbol{V}\Sigma^T\boldsymbol{U}^T$, $\boldsymbol{A}^T\boldsymbol{U} = \boldsymbol{V}\Sigma^T$, $\boldsymbol{A}^T\boldsymbol{u}_k = \sigma_k\boldsymbol{v}_k$

u_k/v_k为左/右奇异向量(能找到n对),且$\{u_k\}$和$\{v_k\}$各自两两正交

- □ 当$\boldsymbol{A}$为实对称半正定阵时, 奇异值分解 = 特征值分解

奇异值分解(SVD)

"线性代数领域的瑞士军刀"

- **Th5.23** (奇异值分解定理)

  矩阵A可以分解为$A = U\Sigma V^T$, U, V为正交阵, Σ为对角阵, 其对角元$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

  - 只需证明m≥n的情况

思路: $\underset{n\times n}{A^T A} = V\Sigma^T\Sigma V^T = V\begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix}V^T$

$A^T A$对称半正定, 设非零特征值为 $\sigma_1^2, \sigma_2^2, \cdots, \sigma_r^2$形成对角阵 Σ_r^2

(不妨设r<n)

$$A^T A = \underset{\text{(正交阵)}}{[V_1 \quad V_2]}\begin{bmatrix} \Sigma_r^2 & O \\ O & O \end{bmatrix}\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \Rightarrow \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} A^T A [V_1 \quad V_2] = \begin{bmatrix} \Sigma_r^2 & O \\ O & O \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} V_1^T A^T A V_1 & \cdots \\ \cdots & V_2^T A^T A V_2 \end{bmatrix}$$

$$\Rightarrow \Sigma_r^{-1} V_1^T A^T \boxed{A V_1 \Sigma_r^{-1}} = I_r \Rightarrow U_1^T U_1 = I_r$$

$$\Rightarrow A V_2 = O$$

U_1为m×r列正交阵

$$\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} A^T A [V_1 \quad V_2] = \begin{bmatrix} \Sigma_r^2 & O \\ O & O \end{bmatrix} \Rightarrow AV_2 = O$$

设 $U_1 = AV_1\Sigma_r^{-1}$，$U_1^T U_1 = I_r$

根据 U_1 补齐正交单位向量得到正交阵 $U = [U_1 \quad U_2]$

$$U^T AV = \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} A[V_1 \quad V_2] = \begin{bmatrix} U_1^T AV_1 & U_1^T AV_2 \\ U_2^T AV_1 & U_2^T AV_2 \end{bmatrix} = \begin{bmatrix} U_1^T AV_1 & O \\ U_2^T AV_1 & O \end{bmatrix}$$

$$U_1 = AV_1\Sigma_r^{-1} \Rightarrow AV_1 = U_1\Sigma_r \Rightarrow U_1^T AV_1 = \Sigma_r$$

$$U_2^T AV_1 = U_2^T U_1 \Sigma_r = O$$

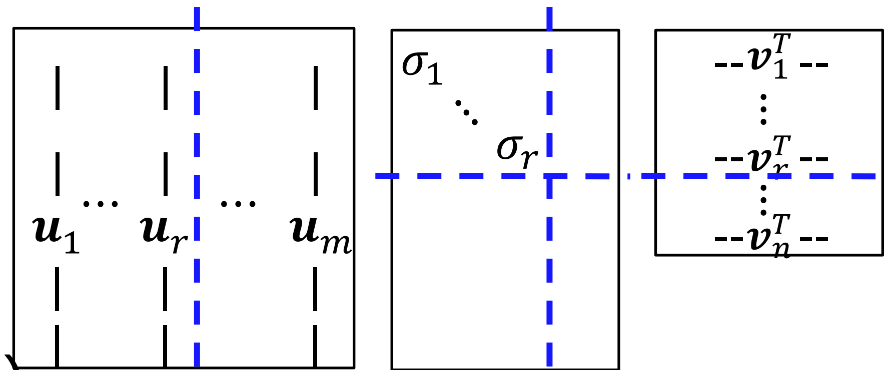
所以，$U^T AV = \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix} = \Sigma$　　矩阵Σ是唯一确定的

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

奇异值分解(SVD)的简化形式

Matlab: [U, S, V] = svd(A);
S= svd(A);

$$A = U_{m\times m}\Sigma_{m\times n}V^T_{n\times n} =$$

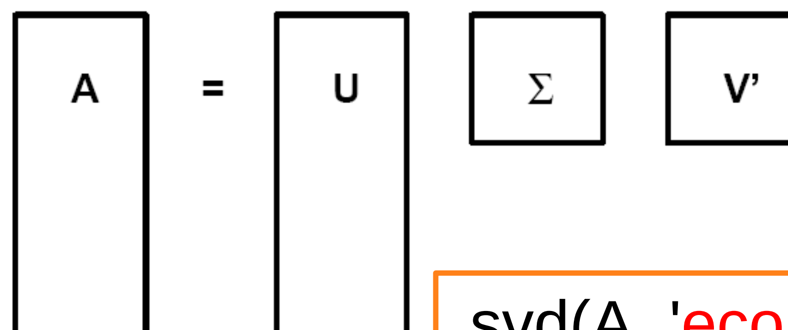


- 简化SVD分解

$$A = U_{r \atop m\times r}\Sigma_{r \atop r\times r}V^T_{r \atop r\times n}$$

- 一般秩r不确定, ≤min(m,n)
  简化SVD为 (以m>n为例)

$$A = U_n\Sigma_n V^T$$



$$A^{\mathrm{T}}A = V\Sigma_n^{\mathrm{T}}\Sigma_n V^{\mathrm{T}} = V\Sigma_n^2 V^{\mathrm{T}}$$

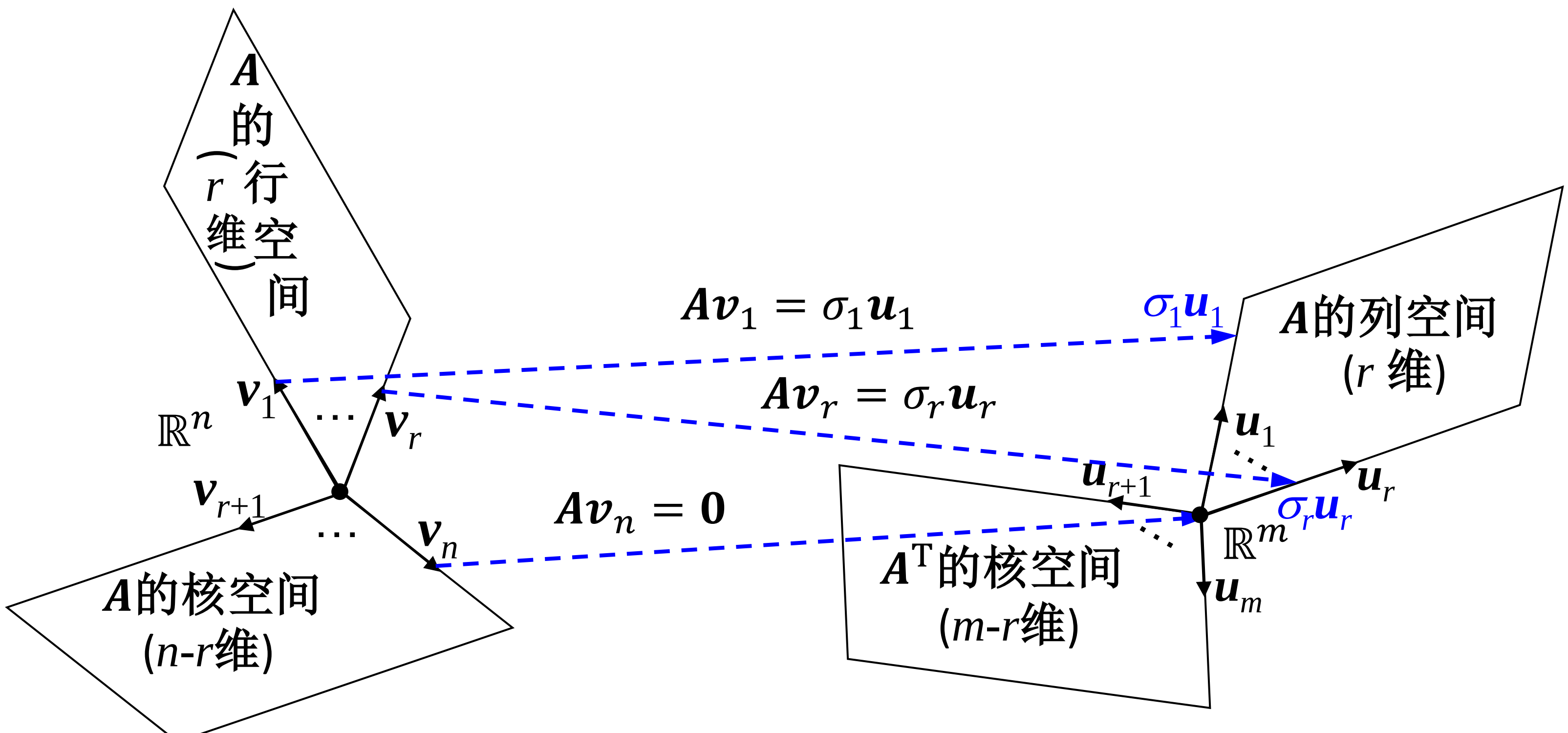
svd(A, 'econ')

- 奇异值σ_k是AA^{T}或$A^{\mathrm{T}}A$的特征值的算术平方根

奇异值分解(SVD)

$$\underset{m\times m}{\boldsymbol{A} = \boldsymbol{U}}\underset{m\times n}{\Sigma}\underset{n\times n}{\boldsymbol{V}^T} \quad \begin{cases} \boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\Sigma \\ \boldsymbol{A}^T\boldsymbol{U} = \boldsymbol{V}\Sigma^T \end{cases}$$

- 奇异向量的意义
  - $\boldsymbol{A}$的秩为r，$\boldsymbol{A}\boldsymbol{v}_1=\sigma_1\boldsymbol{u}_1, \cdots, \boldsymbol{A}\boldsymbol{v}_r=\sigma_r\boldsymbol{u}_r, \boldsymbol{A}\boldsymbol{v}_{r+1}=\boldsymbol{0}, \cdots, \boldsymbol{A}\boldsymbol{v}_n=\boldsymbol{0}$
  - $\boldsymbol{A}^\mathrm{T}=\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^\mathrm{T}$, $\boldsymbol{A}^\mathrm{T}\boldsymbol{u}_1=\sigma_1\boldsymbol{v}_1,\cdots, \boldsymbol{A}^\mathrm{T}\boldsymbol{u}_r=\sigma_r\boldsymbol{v}_r, \boldsymbol{A}^\mathrm{T}\boldsymbol{u}_{r+1}=\boldsymbol{0},\cdots, \boldsymbol{A}^\mathrm{T}\boldsymbol{u}_m=\boldsymbol{0}$
  - 四个重要子空间的单位正交基



奇异值分解(SVD)

- 奇异值与2-范数/F-范数

$\boldsymbol{y} = [y_1 \quad 0 \quad \cdots \quad 0]^{\mathrm{T}}$

最大
奇异值

$$\|A\|_2 = \max_{x\neq 0}\frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \max_{x\neq 0}\frac{\|U\Sigma \boldsymbol{V}^T\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \max_{x\neq 0}\frac{\|\Sigma \boldsymbol{V}^T\boldsymbol{x}\|_2}{\|\boldsymbol{V}^T\boldsymbol{x}\|_2} = \max_{y\neq 0}\frac{\|\Sigma \boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2} = \boldsymbol{\sigma}_1$$

若$\boldsymbol{A}$非奇异方阵，
其奇异值均>0

$$\|\boldsymbol{A}^{-1}\|_2 = 1/\boldsymbol{\sigma}_n \Rightarrow \mathrm{cond}(A)_2 = \frac{\sigma_1}{\sigma_n}$$

奇异值范围反映矩阵接近奇异的程度

$$\|\boldsymbol{A}\|_{\mathrm{F}} = \left(\sum_k \sum_j a_{kj}^2\right)^{1/2}$$

$$\|\boldsymbol{A}\|_{\mathrm{F}} = \sqrt{\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})} = \sqrt{\sum_{k=1}^{n}\lambda_k(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})} = \sqrt{\sum_{k=1}^{min(m,n)}\sigma_k^2}$$

$$\|\boldsymbol{A}\|_{\mathrm{F}} \geqslant \|\boldsymbol{A}\|_2$$

奇异值分解(SVD)

- SVD的向量外积和形式

$$\boldsymbol{A} = [\boldsymbol{u}_1 \cdots \boldsymbol{u}_m]\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & \end{bmatrix}\begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_n^T \end{bmatrix} = \begin{bmatrix} \sigma_1 \boldsymbol{u}_1 \cdots \sigma_n \boldsymbol{u}_n \end{bmatrix}\begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_n^T \end{bmatrix} = \sum_{k=1}^{n} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T$$

$+ \cdots$

$$\boldsymbol{A}_r = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T$$

Eckart-Young定理: 所有秩为r的矩阵中$\boldsymbol{A}_r$最接近$\boldsymbol{A}$

$$\boldsymbol{A}_r \equiv \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^{\mathrm{T}} = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}}$$

$$\|\boldsymbol{A} - \boldsymbol{A}_r\|_2 = \sigma_{r+1} \qquad \|\boldsymbol{A} - \boldsymbol{A}_r\|_{\mathrm{F}} = \sqrt{\sum_{k=r+1}^{\min(m,n)} \sigma_k^2}$$