SVD 定义：设 A 是 m*n 矩阵，A^HA 的特征值为 $\lambda_1 \ge \lambda_2 \ge ..... \ge \lambda_r \ge \lambda_{r+1} = ... = \lambda_n = 0$，则称 $\sigma_i = \sqrt{\lambda_i} ......(i = 1,2,3....r)$ 为矩阵 A 的奇异值，r 为 A 的秩。存在 m 阶酉矩阵 U 和 N 阶酉矩阵 V，使得 $A = U\begin{bmatrix} \sum & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times(N-r)} \end{bmatrix}V$，其中 $\sum = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & .... & \\ & & & \sigma_r \end{bmatrix}$。

应用：

很多情况下，线性方程组 Ax=b 没有解，因此我们计算其最小二乘解，即使得$||Ax-b||^2$ 最小的 x，设 A 的 SVD 分解为 $A = U\begin{bmatrix} \sum & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times(N-r)} \end{bmatrix}V$，由于 2-范数具有酉不便性，因此$||Ax-b||_2$=$\left\| U\begin{bmatrix} \sum & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times(N-r)} \end{bmatrix}Vx - b \right\| = \left\| \begin{bmatrix} \sum & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times(N-r)} \end{bmatrix}Vx - U^Hb \right\|$，由此 Ax=b 的最小二乘解即是 $\begin{bmatrix} \sum & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times(N-r)} \end{bmatrix}Vx = U^Hb$ 的最小二乘解。

令 $y = Vx$，$c = U^Hb$，$\begin{bmatrix} \sum & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times(N-r)} \end{bmatrix}y = c$ 的最小二乘解为 $y = [\frac{c_1}{\sigma_1}, \frac{c_2}{\sigma_2}, ... \frac{c_r}{\sigma_r}, 0,0....,0]$，所以原方程组的最小二乘解为：$x = V^H y$。

其他相关知识：

最小二乘：

- 最小二乘估计起源于 1795 年，当时高斯运用这种估计方法研究行星运动。
- 最小二乘估计不需要任何先验知识，只需估计量的观测信号模型。

设被估计矢量θ，x 是观测值矩阵，n 为观测噪声矢量矩阵，H 是观测矩阵

$$\boldsymbol{x} = \boldsymbol{H\theta} + \boldsymbol{n}$$
$$J(\hat{\boldsymbol{\theta}}) = (\boldsymbol{x} - \boldsymbol{H}\hat{\boldsymbol{\theta}})^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{H}\hat{\boldsymbol{\theta}})$$

J（$\hat{\theta}$ 达到最小，这就是线性最小二乘估计量的构造规则。

$$\left.\frac{\partial J(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\mathrm{ls}}} = \boldsymbol{0}$$
$$\hat{\boldsymbol{\theta}}_{\mathrm{ls}} = (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{x}$$
$$J_{\min}(\hat{\boldsymbol{\theta}}_{\mathrm{ls}}) = \boldsymbol{x}^{\mathrm{T}}[\boldsymbol{I} - \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}]\boldsymbol{x}$$

所求的的最小二乘估计误差为$J_{\min}(\hat{\theta}_{ls}$