Technische Universität Braunschweig

Institute for Numerical Analysis

Ramp Up Mathematics — Numerical Analysis Ramp Up for Data Science

Matthias Bollhöfer, SS 2024

Contents I

Contents II

Definition and Properties

Definition 5.1

Given $A \in \mathbb{R}^{m,n}$, we call $A = U\Sigma V^T$ (the) singular value decomposition (SVD) of A, whenever $U \in \mathbb{R}^{m,m}$ and $V \in \mathbb{R}^{n,n}$ are orthogonal, $\Sigma \in \mathbb{R}^{m,n}$ is diagonal with nonnegative diagonal entries in decreasing order.

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix}, \sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$$

- $r = \operatorname{rank} A \Rightarrow \sigma_1 \geqslant \cdots \geqslant \sigma_r > 0 = \sigma_{r+1} = \cdots \sigma_{\min\{m,n\}}$
- columns of $U = [u_1, \ldots, u_m]$ left singular vectors,
  columns of $V = [v_1, \ldots, v_n]$ right singular vectors
- $Av_i = u_i\sigma_i$, $i = 1, \ldots, r$, particularly $Av_i = 0$, $i > r$.
  Likewise $A^T u_i = v_i\sigma_i$, $i = 1, \ldots, r$, $A^T u_i = 0$, $i > r$
- SVD and economy size SVD in MATLAB `svd(A)`, `svd(A,'econ')`

Basic Properties

Theorem 5.1 (SVD)

There exists an/the SVD of A

Further properties

$$A = [U_1\, U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} [V_1\, V_2]^T = U_1 \Sigma_1 V_1^T \tag{1}$$

$$A = U\Sigma V^T \Rightarrow A = [u_1, \dots, u_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} [v_1, \dots, v_r]^T = u_1 \sigma_1 v_1^T + \cdots + u_r \sigma_r v_r^T \tag{2}$$

- $A^T A = V\Sigma^T \Sigma V^T$, $AA^T = U\Sigma\Sigma^T U^T$
- both matrices $A^T A$ and AA^T are symmetric and positive semidefinite (nonnegative eigenvalues), σ_i^2 are the eigenvalues of $A^T A$, AA^T
- columns of U (V) are orthonormal eigenvectors of AA^T ($A^T A$)

Basic Properties

Example 5.1

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \Rightarrow A^T A = \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}$$

$$\det(A^T A - \lambda I) = 0 \Rightarrow \lambda_1 = \quad , \lambda_2 = \quad \Rightarrow \sigma_1 = \quad , \sigma_2 =$$

$$(A^T A - 45I)x = 0 \Rightarrow \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix} x = 0 \Rightarrow x = \alpha \begin{bmatrix} \quad \end{bmatrix}, v_1 := \quad \begin{bmatrix} \quad \end{bmatrix}$$

$$(A^T A - 5I)x = 0 \Rightarrow \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix} x = 0 \Rightarrow x = \alpha \begin{bmatrix} \quad \end{bmatrix}, v_2 := \quad \begin{bmatrix} \quad \end{bmatrix}$$

$$V = \quad \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}, \Sigma = \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}$$

In principle: $\lambda_{1,2}$ *eigenvls. of* AA^T*, compute* $u_{1,2}$ *from there. Better:* $u_i = Av_i/\sigma_i$

$$u_1 = Av_1/\sigma_1 = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \quad \begin{bmatrix} \quad \end{bmatrix} / \quad = \quad \begin{bmatrix} \quad \end{bmatrix}$$

$$u_2 = Av_2/\sigma_2 = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \quad \begin{bmatrix} \quad \end{bmatrix} / \quad = \quad \begin{bmatrix} \quad \end{bmatrix} \Rightarrow U = \quad \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}$$

The Key Property

Most important result:

Theorem 5.2 (Eckart-Young-Schmidt-Mirsky for 2-norm)

$A \in \mathbb{R}^{m,n}$, $A = U\Sigma V^T$ SVD, $A_k = [u_1, \ldots, u_k] \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_k \end{bmatrix} [v_1, \ldots, v_k]^T$. Then

$$A_k = \underset{\operatorname{rank} \hat{A}_k = k}{\operatorname{argmin}} \|A - \hat{A}_k\|_2 = \sigma_{k+1}.$$

Example 5.2

example1.m Display sing. vals. explain Thr 5.2.

Application Image Compression

Example 5.3

1. *Gray-scale picture* $X \sim 2000 \times 1500$, *matrix values in* $\{0, 1, 2, \ldots, 255\}$
   example1.m show picture, SVD rank 5,10, 1000, image compression
   use "best" rank-r approximation
   $X \approx U\Sigma V^T$, $U \in \mathbb{R}^{2000,r}$, $\Sigma \in \mathbb{R}^{r,r}$ *dgl. nonneg.*, $V \in \mathbb{R}^{1500,r}$, U, V *orth. cols.*
   Display pictures and diagonal of Σ
2. *imagesvd.m mandrill, reduce rank, colored picture*

The Key Property

$$A \in \mathbb{R}^{m,n},\ \|A\|_F := \sqrt{\sum_{i,j} |a_{ij}|^2} \tag{3}$$

Analogous result holds for F-norm.

Theorem 5.3 (Eckart-Young-Schmidt-Mirsky for F-norm)

$A \in \mathbb{R}^{m,n}$, $A = U\Sigma V^T$ SVD, $A_k = [u_1, \dots, u_k] \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_k \end{bmatrix} [v_1, \dots, v_k]^T$. Then

$$A_k = \operatorname*{argmin}_{\operatorname{rank} \hat{A}_k = k} \|A - \hat{A}_k\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_{\min\{m,n\}}^2}.$$

The Key Property

For 2-, F- norm we have

- $\|I\|_F = \sqrt{\min\{m,n\}}$, $\|A\|_2 \leqslant \|A\|_F$
- $\|AB\|_F \leqslant \left\{ \begin{smallmatrix} \|A\|_2\|B\|_F \\ \|A\|_F\|B\|_2 \end{smallmatrix} \right.$, $\|A\|_2 \leqslant \|A\|_F$
- P, Q orthogonal, $\|PAQ\|_F = \|A\|_F$
- $\|A\|_2 = \|\Sigma\|_2 = \sigma_1$
- $\|A\|_F = \|\Sigma\|_F = \sqrt{\sum_i \sigma_i^2}$
- F-norm advantageous, can be taken by columns or by rows and treat them separately

$$A = [a_1, \ldots, a_n] = \begin{bmatrix} \hat{a}_1^T \\ \vdots \\ \hat{a}_m^T \end{bmatrix}, \quad \|A\|_F^2 = \sum_j \|a_j\|_2^2 = \sum_i \|\hat{a}_i\|_2^2$$