1. Singular Value Decomposition (SVD)

   Any m-by-n matrix A with $m \geq n$ can be written as

$$A = U\Sigma V^T,$$

   where U is m-by-n orthogonal matrix ($U^T U = I_n$) and V is n-by-n orthogonal matrix ($V^T V = I$), and $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

   Nonnegative scalar $\sigma_1, \sigma_2, \ldots, \sigma_n$ are called *singular values.* The columns $\{u_i\}$ of U are called *left singular vectors* of A. The columns $\{v_i\}$ of V are called *right singular vectors.*

   If $m < n$, the SVD can be defined by considering A^T.

2. Connection (difference) between eigenvalues and singular values.

   (a) Eigenvalues of $A^T A$ are σ_i^2 for $i = 1, 2, \ldots, n$. The corresponding eigenvectors are the right singular vectors v_i for $i = 1, 2, \ldots, n$.

   (b) Eigenvalues of AA^T are σ_i^2 for $i = 1, 2, \ldots, n$ and $m-n$ zeros. The left singular vectors u_i for $i = 1, 2, \ldots, n$ are corresponding eigenvectors for the eigenvalues σ_i^2. Any $m-n$ orthogonal vectors that are orthogonal to $u_1, u_2, \ldots, u_n$ as the eigenvectors for the zero eigenvalues.

3. Suppose that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0,$$

   Then

   (a) the rank of A is r,

   (b) the column space of A is spanned by $[u_1, u_2, \cdots, u_r]$.

   (c) the nullspace of A is spanned by $[v_{r+1}, v_{r+2}, \ldots, v_n]$.

4. The matrix norm $\|A\|_2$ induced the vector 2-norm

$$\|A\|_2 \equiv \max_{0 \neq x \in \mathbb{R}^n} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1 = \sqrt{\lambda_{\max}(A^T A)}.$$

5. Suppose that A has full column rank, then the pseudo-inverse can also be written as[1]

$$A^+ \equiv (A^T A)^{-1} A^T = V\Sigma^{-1} U^T.$$

6. The SVD of A can be rewritten as

$$A = E_1 + E_2 + \cdots + E_r$$

   where $r = \mathrm{rank}(A)$, E_k for $i = 1, 2, \ldots, r$ is a rank-one matrix of the form

$$E_k = \sigma_k u_k v_k^T,$$

[1] If $m < n$, then $A^+ = A^T (AA^T)^{-1}$.

and is referred to as the k-th *component* matrix. Component matrices are orthogonal to each other, i.e.,

$$E_j E_k^T = 0, \quad j \neq k.$$

Furthermore, since $\|E_k\|_2 = \sigma_k$, we know that

$$\|E_1\|_2 \geq \|E_2\|_2 \geq \cdots \geq \|E_r\|_2.$$

It means that the contribution each E_k makes to reproduce A is determined by the size of the singular value σ_k,

7. Optimal rank-k approximation (Eckart-Young Theorem):

$$\min_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank}(B) = k}} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1},$$

   where

$$A_k = E_1 + E_2 + \cdots + E_k = U_k \Sigma_k V_k^T,$$

   where $\Sigma_k = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k)$, U_k and V_k are the first k columns of U and V, respectively. A_k is called a *truncated SVD*.

8. The problem of applying the leading k components of A to analyze the data in the matrix A is called *Principal Component Analysis (PCA)*.

9. An application of PCA for lossy data compression.

   Note that A_k is represented by $mk + k + nk = (m + n + 1)k$ elements, in contrast, A is represented by mn elements. Therefore, we have

$$\text{compression ratio} = \frac{(m + n + 1)k}{mn}$$

   Matlab script: `svd4image.m`