Singular Value Decomposition

Recall linear algebra: Let \( \rho(A) = \{ \lambda : \det(A - \lambda I) = 0 \} \) denote the collection of all eigenvalues of \( A \), called spectrum, and \( v \in \ker(A - \lambda I)(\text{eigenspace}) \) is called an eigenvector associated with \( \lambda \). Trivially eigenvalue is not defined for a non-square matrix. Know the following facts:

- If \( A \) is symmetric, then \( \rho(A) \subseteq [0, \infty) \);
- If \( \lambda, \mu \in \rho(A), \lambda \neq \mu \), then \( \ker(A - \lambda I) \perp \ker(A - \mu I) \).

Therefore we can define called singular values of \( A \): \( \sigma(A) = \{ \sqrt{\lambda} : \lambda \in \rho(A^T A) \} = \{ \sqrt{\rho(A^T A)} \} \). So \( A_{m \times n} \) has \( n \) singular values, some maybe zeros. Let \( r \) be the number of positive singular values(counting multiplicity), then we can find an orthogonal \( U_{n \times n} \) and an orthogonal \( V_{m \times m} \) such that

\[
V^T A = \begin{bmatrix}
\Sigma & 0 \\
0 & 0
\end{bmatrix}_{m \times n} = \text{diag}_{m \times n}(\sigma_1, \cdots, \sigma_r, \sigma_{r+1}, \cdots, \sigma_k),
\]

where \( k = \min(m, n) \), \( \sigma_1, \cdots, \sigma_n \in \sigma(A) = \sqrt{\rho(A^T A)} \), \( \sigma_1 \geq \cdots \geq \sigma_r > 0 \).

\[
A = V \begin{bmatrix}
\Sigma & 0 \\
0 & 0
\end{bmatrix} U^T
\]

is called a Singular Value Decomposition of \( A \). Notice that \( r = \) number of positive singular values = rank\( (A_{m \times n}) \) \leq \min(m, n) \). \( U \) and \( V \) are not unique, but \( \sigma_i \)'s are. SVD can be accomplished by QR algorithm — find an orthogonal \( V_{m \times m} \) to make \( A \) upper triangular: \( V^T A = R \), then find an orthogonal \( U_{n \times n} \) to make \( R \) diagonal: \( R U = V^T A U = D \), then \( D = \begin{bmatrix}
\Sigma & 0 \\
0 & 0
\end{bmatrix}_{m \times n} = \text{diag}_{m \times n}(\sigma_1, \cdots, \sigma_r, 0, \cdots, 0) \).

Application — Least Square Problem

Let \( x \in \mathbb{R}^n \), \( b \in \mathbb{R}^m \). If \( A_{m \times n} \) and \( m > n \), usually \( A x = b \) has no solution unless \( b \) lies in the column space of \( A \). But we can find \( \hat{x} \in \mathbb{R}^n \) which solves this minimizing problem: \( \min_{x \in \mathbb{R}^n} \| A x - b \| \), i.e. the distance between \( b \in \mathbb{R}^m \) and \( A \hat{x} \in \text{col}(A) := \text{span}(a_1, \cdots, a_n) \subseteq \mathbb{R}^m \) is the shortest. Acturally, \( A \hat{x} \) shall be the projection of \( b \) onto \( \text{col}(A) \).

\[
r = \text{rank}(A). \quad \text{From the SVD result: } A_{m \times n} = V_{m \times m} \begin{bmatrix}
\Sigma_{r \times r} & 0 \\
0 & 0
\end{bmatrix} U_{n \times n}^T, \quad \text{let } V^T b = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \quad c_1 \in \mathbb{R}^r, \quad \text{and for any } x \in \mathbb{R}^n, \quad \text{denote } U^T x = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \quad z_2 \in \mathbb{R}^r. \quad \text{We have shown in class that } \hat{x} \text{ minimizes } \| A x - b \|^2 \text{ if and only if } U^T \hat{x} = \begin{bmatrix} \Sigma_{r \times r}^{-1} c_1 \\ z_2 \end{bmatrix} \quad (z_2 \text{ arbitrary}). \quad \text{Let } \hat{x}_{LS} := U \begin{bmatrix} \Sigma_{r \times r}^{-1} & 0 \\
0 & 0
\end{bmatrix} V^T b, \quad \text{then it is a solution of the minimal norm. If } r = n, \text{ that means that columns of } A \text{ form a basis of } \text{col}(A) \text{ so the projection of } b \text{ onto } \text{col}(A) \text{ can be written as a unique linear combination of columns of } A, \text{ therefore } \hat{x}_{LS} \text{ is the unique LS solution.}
\]