1 Singular Value Decomposition

Definition 1. A matrix $M \in \mathbb{R}^{n \times n}$ is called orthogonal if all rows and columns of the matrix are mutually orthogonal. Moreover, if all rows and columns have a unit norm, the matrix is called orthonormal. So for an orthonormal matrix, we have $M^T M = M M^T = I$, where $I$ is an identity matrix of size $n \times n$.

Theorem 2. Any $m \times n$ matrix can be factored into $A = U \Sigma V^\top$, where $U$ is an $m \times m$ orthogonal matrix, $V$ is an $n \times n$ orthogonal matrix, and $\Sigma$ has the form

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix}$$

where $\text{rk} A = r$ and $\sigma_1, \ldots, \sigma_r$ are the singular values of $A$.

Consider. The SVD allows to describe the effect of a matrix on a vector (via the matrix-vector product), as a three-step process: a first rotation in the input space; a simple positive scaling that takes a vector in the input space to the output space; and another rotation in the output space. In particular, to get $Ax$, where $x \in \mathbb{R}^n$, we first form $\tilde{x} := V^T x \in \mathbb{R}^n$. Since $V$ is an orthogonal matrix, $V^T$ is also orthogonal, and $\tilde{x}$ is just a rotated version of $x$, which still lies in the input space. Then we act on the rotated vector $\tilde{x}$ by scaling its elements. Precisely, the first $r$ elements of $\tilde{x}$ are scaled by the singular values $\sigma_1, \ldots, \sigma_r$; the remaining $n - r$ elements are set to zero. This step results in a new vector $\tilde{y}$ which now belongs to the output space $\mathbb{R}^m$. The final step consists in rotating the vector $\tilde{y}$ by the orthogonal matrix $U$, which results in $y = U \tilde{y} = Ax$.

To do the proof/construction of the SVD, we need the following result:

Lemma 3. The columns of $U$ are orthonormal eigenvectors of $AA^\top$, the columns of $V$ are orthonormal eigenvectors of $A^\top A$ and $\sigma_i^2$s are the eigenvalues of $AA^\top$ (or $A^\top A$).
Problem 1. Find the SVD of

\[ A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \]

Proof.

\[ A A^\top = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix} \]

Now we need to find the eigenvectors of \( A A^\top \), i.e. solve \( A A^\top x = \lambda x \).

\[ \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \]

\[ \text{det} \left( \begin{bmatrix} 11 - \lambda & 1 \\ 1 & 11 - \lambda \end{bmatrix} \right) = 0 \]

solving gives \( \lambda_2 = 10, \lambda_1 = 12 \). Plugging these \( \lambda \)'s back in we get

\[ (11 - \lambda_2) x_1 + x_2 = 0 \quad \Rightarrow \quad x_1 = -x_2 \quad \text{(for } \lambda_2 = 10 \text{)} \quad \text{and} \quad x_1 = x_2 \quad \text{(for } \lambda_1 = 12 \text{)} \]

For \( \lambda_2 = 10 \) case, we can choose \( x_1 = \frac{1}{\sqrt{2}}, \ x_2 = -\frac{1}{\sqrt{2}} \). Similarly, for \( \lambda_1 = 12 \) case, we can choose \( x_1 = \frac{1}{\sqrt{2}}, \ x_2 = \frac{1}{\sqrt{2}} \). Thus we have the two orthonormal eigenvectors of \( A A^\top \). Therefore,

\[ U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \]

Similarly for \( V \) we can compute the orthonormal eigenvectors of \( A^\top A \).

\[ V = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{30}} \\ 0 & -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{30}} \end{bmatrix} \]

Then,

\[ \Sigma = \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \]

Hence,

\[ A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V^\top_{n \times n} \]

\[ \square \]
Problem 2. Matrix 2-norm. Prove that

\[ \|A\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2 = \sigma_1 \]

2 Lipschitz Continuity

Definition 4. \( f \) is globally Lipschitz continuous (LC) if there exists \( L \) such that

\[ \|f(x) - f(y)\| \leq L \|x - y\|, \]

for all \( x, y \in \mathbb{R}^n \).

\( f \) is locally Lipschitz continuous in \( U \subset \mathbb{R}^n \), if for every \( x, y \in U \), the Lipschitz property above is satisfied.

Consider. Which norm should we use to check the Lipschitz condition?

Problem 3. (Local or global Lipschitz condition.) Consider the following system of differential equations:

\[
\begin{align*}
\dot{x}_1 &= x_1^2 + x_2^2 \\
\dot{x}_2 &= x_1^2 - x_2^2
\end{align*}
\]

Prove that this system is locally Lipschitz, but not globally Lipschitz.
3 Fundamental Theorem

**Theorem 5** (Fundamental Theorem of Differential Equations). Consider the following ordinary differential equation (ODE):

\[
\dot{x} = f(x,t), \\
x(t_0) = x_0,
\]

with the vector field \( f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n \). If \( f \) is

- piecewise continuous in \( t \)
- Lipschitz continuous in \( x \),

then the ODE admits a unique solution, which is differentiable almost everywhere except at points where \( f \) is discontinuous with respect to \( t \).

**Problem 4.** (Linear systems) Consider the following linear system:

\[
\dot{x} = A(t)x(t) + B(t)u(t), \\
x(t_0) = x_0.
\]

Provide a sufficient condition for the linear system to have a unique solution.