Singular Value Decomposition
(matrix factorization)
Singular Value Decomposition

The SVD is a factorization of a $m \times n$ matrix into

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

where $U$ is a $m \times m$ orthogonal matrix, $V^T$ is a $n \times n$ orthogonal matrix and $\Sigma$ is a $m \times n$ diagonal matrix.

For a square matrix ($m = n$):

$$A = \begin{pmatrix} u_1 & \ldots & u_n \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} v_1 & \ldots & v_n \end{pmatrix}^T$$

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots$$
Reduced SVD

What happens when $A$ is not a square matrix?

1) $m > n$

$$A = U \Sigma V^T = \begin{pmatrix} \vdots & \cdots & \vdots & \cdots & \vdots \\ u_1 & \cdots & u_n & \cdots & u_m \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_n & & \\ & & & 0 & \vdots \\ & & & \vdots & 0 \end{pmatrix} \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$m \times m$  $m \times n$  $n \times n$
Reduced SVD

2) \( n > m \)

\[
A = U \Sigma V^T = \begin{pmatrix}
\vdots & \cdots & \vdots \\
u_1 & \cdots & u_m \\
\vdots & \cdots & \vdots \\
\end{pmatrix}
\begin{pmatrix}
\sigma_1 & \cdots & 0 \\
\vdots & \sigma_m & \vdots \\
0 & \cdots & 0 \\
\end{pmatrix}
\begin{pmatrix}
\vdots & v_1 \\
\vdots & \vdots \\
\vdots & v_m \\
\vdots & \vdots \\
\vdots & \vdots \\
\vdots & \vdots \\
\vdots & v_n \\
\end{pmatrix}
\]

\( n \times m \) \( m \times n \) \( n \times n \)
Let’s take a look at the product $\Sigma^T \Sigma$, where $\Sigma$ has the singular values of a $A$, a $m \times n$ matrix.

$m > n$

$$\Sigma^T \Sigma = \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_n \end{pmatrix} = \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_n^2 \end{pmatrix}$$

$n > m$

$$\Sigma^T \Sigma = \begin{pmatrix} \sigma_1 & \cdots & \sigma_m \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_m \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots \\ 0 & \ddots & \vdots \\ \vdots & \cdots & 0 \end{pmatrix}$$
Assume \( A \) with the singular value decomposition \( A = U \Sigma V^T \). Let’s take a look at the eigenpairs corresponding to \( A^T A \):
In a similar way,

$$AA^T = (U \Sigma V^T) (U \Sigma V^T)^T$$

$$= (U \Sigma V^T)(V^T)^T (\Sigma )^T U^T$$

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

$$= U \Sigma \Sigma^T U^T$$

Hence $$AA^T = U \Sigma^2 U^T$$

Recall that columns of $$U$$ are all linear independent (orthogonal matrices), then from diagonalization ($$B = XDX^{-1}$$), we get:

- The columns of $$U$$ are the eigenvectors of the matrix $$AA^T$$
How can we compute an SVD of a matrix $A$?

1. Evaluate the $n$ eigenvectors $v_i$ and eigenvalues $\lambda_i$ of $A^T A$

2. Make a matrix $V$ from the normalized vectors $v_i$. The columns are called “right singular vectors”.

$$V = \begin{pmatrix} \vdots & \ldots & \vdots \\ v_1 & \ldots & v_n \\ \vdots & \ldots & \vdots \end{pmatrix}$$

3. Make a diagonal matrix from the square roots of the eigenvalues.

$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \quad \sigma_i = \sqrt{\lambda_i} \quad \text{and} \quad \sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots$$

4. Find $U$: $A = U \Sigma V^T \Rightarrow U \Sigma = A V$. The columns are called the “left singular vectors”.
True or False?

\( A \) has the singular value decomposition \( A = U \Sigma V^T \).

- The matrices \( U \) and \( V \) are not singular
- The matrix \( \Sigma \) can have zero diagonal entries
- \( \| U \|_2 = 1 \)
- The SVD exists when the matrix \( A \) is singular
- The algorithm to evaluate SVD will fail when taking the square root of a negative eigenvalue
Singular values are always non-negative

- A matrix is positive definite if $x^T B x > 0$ for $\forall x \neq 0$
- A matrix is positive semi-definite if $x^T B x \geq 0$ for $\forall x \neq 0$
Cost of SVD

The cost of an SVD is proportional to $mn^2 + n^3$ where the constant of proportionality constant ranging from 4 to 10 (or more) depending on the algorithm.

![Graph showing the cost of SVD, matmat, and LU algorithms as a function of matrix size.]

\[
C_{\text{SVD}} = \alpha (mn^2 + n^3) = O(n^3) \\
C_{\text{matmat}} = n^3 = O(n^3) \\
C_{\text{LU}} = 2n^3 / 3 = O(n^3)
\]
SVD summary:

- The SVD is a factorization of a $m \times n$ matrix into $A = U \Sigma V^T$ where $U$ is a $m \times m$ orthogonal matrix, $V^T$ is a $n \times n$ orthogonal matrix and $\Sigma$ is a $m \times n$ diagonal matrix.

- In reduced form: $A = U_R \Sigma_R V_R^T$, where $U_R$ is a $m \times k$ matrix, $\Sigma_R$ is a $k \times k$ matrix, and $V_R$ is a $n \times k$ matrix, and $k = \min (m, n)$.

- The columns of $V$ are the eigenvectors of the matrix $A^T A$, denoted the right singular vectors.

- The columns of $U$ are the eigenvectors of the matrix $AA^T$, denoted the left singular vectors.

- The diagonal entries of $\Sigma^2$ are the eigenvalues of $A^T A$. $\sigma_i = \sqrt{\lambda_i}$ are called the singular values.

- The singular values are always non-negative (since $A^T A$ is a positive semi-definite matrix, the eigenvalues are always $\lambda \geq 0$).
Singular Value Decomposition (applications)
1) Determining the rank of a matrix

Suppose $A$ is a $m \times n$ rectangular matrix where $m > n$:

$$A = \begin{pmatrix} u_1 & \ldots & u_n & \ldots & u_m \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_n & & \\ & & & 0 & \\ & & & & 0 \end{pmatrix} \begin{pmatrix} v_1^T & \ldots & \\ \vdots & \ddots & \\ \vdots & \ddots & v_n^T \end{pmatrix}$$

$$A = \begin{pmatrix} u_1 & \ldots & u_n \\ \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \sigma_1 v_1^T & \ldots & \\ \vdots & \ddots & \\ \vdots & \ddots & \sigma_n v_n^T \end{pmatrix}$$
Rank of a matrix

For general rectangular matrix $A$ with dimensions $m \times n$, the reduced SVD is:

$$A = U_R \Sigma_R V^T$$
Rank of a matrix

- The rank of \( \mathbf{A} \) equals the number of non-zero singular values which is the same as the number of non-zero diagonal elements in \( \mathbf{\Sigma} \).
- Rounding errors may lead to small but non-zero singular values in a rank deficient matrix, hence the rank of a matrix determined by the number of non-zero singular values is sometimes called “effective rank”.
- The right-singular vectors (columns of \( \mathbf{V} \)) corresponding to vanishing singular values span the null space of \( \mathbf{A} \).
- The left-singular vectors (columns of \( \mathbf{U} \)) corresponding to the non-zero singular values of \( \mathbf{A} \) span the range of \( \mathbf{A} \).
2) Pseudo-inverse

- **Problem:** if $A$ is rank-deficient, $\Sigma$ is not be invertible

- **How to fix it:** Define the Pseudo Inverse

- **Pseudo-Inverse of a diagonal matrix:**

  $$(\Sigma^+)_i = \begin{cases} \frac{1}{\sigma_i}, & \text{if } \sigma_i \neq 0 \\ 0, & \text{if } \sigma_i = 0 \end{cases}$$

- **Pseudo-Inverse of a matrix $A$:**

  $$A^+ = V\Sigma^+ U^T$$
3) Matrix norms

The Euclidean norm of an orthogonal matrix is equal to 1

\[ \|U\|_2 = \max_{\|x\|_2=1} \|Ux\|_2 = \max_{\|x\|_2=1} \sqrt{(Ux)^T(Ux)} = \max_{\|x\|_2=1} \sqrt{x^T x} = \max_{\|x\|_2=1} \|x\|_2 = 1 \]

The Euclidean norm of a matrix is given by the largest singular value

\[ \|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \max_{\|x\|_2=1} \|U \Sigma V^T x\|_2 = \max_{\|x\|_2=1} \|\Sigma V^T x\|_2 \]

\[ = \max_{\|v^T x\|_2=1} \|\Sigma V^T x\|_2 = \max_{\|y\|_2=1} \|\Sigma y\|_2 \]

Where we used the fact that \( \|U\|_2 = 1, \|V\|_2 = 1 \). Since \( \Sigma \) is diagonal we get:

\[ \|A\|_2 = \max(\sigma_i) = \sigma_{\text{max}} \]

\( \sigma_{\text{max}} \) is the largest singular value.
4) Norm for the inverse of a matrix

The Euclidean norm of the inverse of a square-matrix is given by:

Assume here $A$ is full rank, so that $A^{-1}$ exists

$$||A^{-1}||_2 = \max_{||x||_2=1} ||(U \Sigma V^T)^{-1}x||_2$$

$$||A^{-1}||_2 = \max_{||x||_2=1} ||V \Sigma^{-1}U^T x||_2$$

Since $||U||_2 = 1$, $||V||_2 = 1$ and $\Sigma$ is diagonal then

$$||A^{-1}||_2 = \frac{1}{\sigma_{min}}$$

$\sigma_{min}$ is the smallest singular value
5) Norm of the pseudo-inverse matrix

The norm of the pseudo-inverse of a $m \times n$ matrix is:

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

$$\|A^+\|_2 = \frac{1}{\sigma_r}$$

where $\sigma_r$ is the smallest non-zero singular value. This is valid for any matrix, regardless of the shape or rank.

Note that for a full rank square matrix, $\|A^+\|_2$ is the same as $\|A^{-1}\|_2$.

Zero matrix: If $A$ is a zero matrix, then $A^+$ is also the zero matrix, and $\|A^+\|_2 = 0$. 
6) Condition number of a matrix

The condition number of a matrix is given by

\[ cond_2(A) = \|A\|_2 \|A^+\|_2 \]

If the matrix is full rank: \( rank(A) = \min(m, n) \)

\[ cond_2(A) = \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}} \]

where \( \sigma_{\text{max}} \) is the largest singular value and \( \sigma_{\text{min}} \) is the smallest singular value

If the matrix is rank deficient: \( rank(A) < \min(m, n) \)

\[ cond_2(A) = \infty \]
7) Low-Rank Approximation

We will again use the SVD to write the matrix $A$ as a sum of outer products (of left and right singular vectors) – here for $m > n$ without loss of generality:

$$A = \begin{pmatrix} u_1 & \cdots & u_m \end{pmatrix} \begin{pmatrix} \sigma_1 & \vdots & \cdots & \sigma_n \end{pmatrix} \begin{pmatrix} v_1^T & \cdots & v_n^T \end{pmatrix}$$

$$= \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} \sigma_1 v_1^T & \cdots \end{pmatrix}$$

$$= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_n u_n v_n^T$$
7) Low-Rank Approximation (cont.)

\[ A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_n u_n v_n^T \quad \sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots \geq 0 \]

What is the rank of \( \sigma_1 u_1 v_1^T \)?

What is the rank of \( \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T \)?
7) Low-Rank Approximation (cont.)

The best rank-\(k\) approximation for a \(m \times n\) matrix \(A\), (where \(k \leq \min(m, n)\)) is the one that minimizes the following problem:

\[
\min_{A_k} \|A - A_k\|
\]

such that \(\text{rank}(A_k) \leq k\).

When using the induced 2-norm, the best rank-k approximation is given by:

\[
A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T
\]

\[
\sigma_1 \geq \sigma_2 \geq \sigma_3 \cdots \geq 0
\]

Note that \(\text{rank}(A) = n\) and \(\text{rank}(A_k) = k\) and the norm of the difference between the matrix and its approximation is
Example: Image compression

1417

500

Image using rank-50 approximation
8) Using SVD to solve square system of linear equations

If $A$ is a $n \times n$ square matrix and we want to solve $Ax = b$, we can use the SVD for $A$ such that