The Singular Value Decomposition and Least Squares Problems

Tom Lyche

University of Oslo
Norway
Applications of SVD

1. solving over-determined equations
2. statistics, principal component analysis
3. numerical determination of the rank of a matrix
4. search engines (Google,...)
5. theory of matrices
6. and lots of other applications...
Singular Value Decomposition

1. Works for any matrix $A \in \mathbb{C}^{m,n}$

2. $A = U \Sigma V^H$ with $U$, $V$ unitary and $\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{C}^{m,n}$

3. $\Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_r)$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$,

4. $r$ is the rank of $A$.

5. We define $\sigma_{r+1} = \cdots \sigma_n = 0$ if $r < n$ and call $\sigma_1, \ldots, \sigma_n$ the singular values of $A$.

6. The columns $u_1, \ldots, u_m$ of $U$ and $v_1, \ldots, v_n$ of $V$ are called left- and right singular vectors respectively.
Relation to eigenpairs for $A^T A$ and $AA^T$

1. $A^T A v_i = \sigma_i^2 v_i$ for $i = 1, \ldots, n$.

2. The columns of $V$ are orthonormal eigenvectors of $A^T A$.

3. The columns of $U$ are orthonormal eigenvectors of $AA^T$. 
Three forms of SVD

Suppose \( A \in \mathbb{R}^{m,n} \), \( A = U\Sigma V^T \) is the SVD of \( A \in \mathbb{R}^{m,n} \) and let \( r := \#\Sigma_1 \). We partition \( U \) and \( V \) as follows

- \( U = [U_1, U_2] \), \( U_1 \in \mathbb{R}^{m,r} \), \( U_2 \in \mathbb{R}^{m,m-r} \)
- \( V = [V_1, V_2] \), \( V_1 \in \mathbb{R}^{n,r} \), \( V_2 \in \mathbb{R}^{n,n-r} \).

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

The three forms

1. \( A = U\Sigma V^T \) \hspace{5cm} \text{full form}
2. \( A = U_1 \Sigma_1 V_1^T \) \hspace{5cm} \text{compact form}
3. \( A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^T \) \hspace{5cm} \text{outer product form}
Subspaces of $A$

column space and the null space of a matrix

$$\text{span}(A) := \{ y \in \mathbb{R}^m : y = Ax, \text{ for some } x \in \mathbb{R}^n \},$$

$$\ker(A) := \{ x \in \mathbb{R}^n : Ax = 0 \}.$$

- $\text{span}(A)$ is a subspace of $\mathbb{R}^m$.
- $\ker(A)$ is a subspace of $\mathbb{R}^n$.

We say that $A$ is a basis for a subspace $S$ of $\mathbb{R}^m$ if

1. $S = \text{span}(A)$,

2. $A$ has linearly independent columns, i.e., $\ker(A) = \{0\}$.

Recall the four fundamental subspaces $\text{span}(A)$, $\text{span}(A^T)$, $\ker(A)$, $\ker(A^T)$. 
The 4 fundamental Subspaces

Let \( A = U \Sigma V^T \) be the SVD of \( A \in \mathbb{R}^{m,n} \). Then \( A^T = V \Sigma^T U^T \) and \( AV = U \Sigma, A^T U = V \Sigma^T \) or

\[
A [v_1, v_2] = [u_1, u_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A^T [u_1, u_2] = [v_1, v_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}.
\]

\[
AV_1 = U_1 \Sigma_1, \quad U_1 \text{ is an orthonormal basis for } \text{span}(A)
\]
\[
A^T U_2 = 0, \quad U_2 \text{ is an orthonormal basis for } \ker(A^T)
\]
\[
A^T U_1 = V_1 \Sigma_1, \quad V_1 \text{ is an orthonormal basis for } \text{span}(A^T)
\]
\[
AV_2 = 0, \quad V_2 \text{ is an orthonormal basis for } \ker(A).
\]

We obtain the fundamental relations

1. \( \dim(\text{span}(A)) + \dim(\ker(A)) = \#A := \text{number of columns of } A, \)
2. \( \dim(\text{span}(A^T)) = \dim(\text{span}(A)) =: \text{rank}(A) = \#\Sigma_1. \)
Existence of SVD

Theorem 1. Every matrix has an SVD.
Uniqueness

- If the SVD of $A$ is $A = U \Sigma V^T$ then $A^T A = V \Sigma^T \Sigma V^T$.
- Thus $\sigma_1^2, \ldots, \sigma_n^2$ are uniquely given as the eigenvalues of $A^T A$ arranged in descending order.
- Taking the positive square root uniquely determines the singular values.
- From the proof of the existence theorem it follows that the orthogonal matrices $U$ and $V$ are in general not uniquely given.
Application I, rank

- Gauss-Jordan cannot be used to determine rank numerically
- Use singular value decomposition numerically will normally find $\sigma_n > 0$.
- Determine minimal $r$ so that $\sigma_{r+1}, \ldots, \sigma_n$ are "close" to round off unit.
Given $A^{m,n}$ and $b \in \mathbb{R}^m$.

The system $Ax = b$ is over-determined if $m > n$.

This system has a solution if $b \in \text{span}(A)$, the column space of $A$, but normally this is not the case and we can only find an approximate solution.

A general approach is to choose a vector norm $\| \cdot \|$ and find $x$ which minimizes $\|Ax - b\|$. 

We will only consider the Euclidian norm here.
The Least Squares Problem

Given $A^{m,n}$ and $b \in \mathbb{R}^m$ with $m \geq n \geq 1$. The problem to find $x \in \mathbb{R}^n$ that minimizes $\|Ax - b\|_2$ is called the least squares problem.

A minimizing vector $x$ is called a least squares solution of $Ax = b$.

Several ways to analyze:
- Quadratic minimization
- Orthogonal Projections
- SVD
Quadratic minimization

- Define function $E : \mathbb{R}^n \to \mathbb{R}$ by $E(x) = \|Ax - b\|_2^2$

- $E(x) = (Ax - b)^T(Ax - b) = x^T B x - 2 c^T x + \alpha$, where

- $B := A^T A$, $c := A^T b$ and $\alpha := b^T b$.

- $B$ is positive semidefinite and positive definite if $A$ has rank $n$.

- Since the Hessian $H E(x) := \left( \frac{\partial^2 E(x)}{\partial x_i \partial x_j} \right) = 2 B$ we can find minimum by setting partial derivatives equal zero.

- $\nabla E(x) := \left( \frac{\partial E(x)}{\partial x_i} \right) = 2 (B x - c) = 0$

- Normal equations $A^T Ax = A^T b$. 
A simple example

\[ x_1 = 1 \]
\[ x_1 = 1, \quad A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad x = [x_1], \quad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \]
\[ x_1 = 2 \]

Quadratic minimization problem:
\[ \|Ax - b\|_2^2 = (x_1 - 1)^2 + (x_1 - 1)^2 + (x_1 - 2)^2. \]

Setting the first derivative with respect to \( x_1 \) equal to zero we obtain
\[ 2(x_1 - 1) + 2(x_1 - 1) + 2(x_1 - 2) = 0 \text{ or } 6x_1 - 8 = 0 \text{ or } x_1 = 4/3 \]

The second derivative is positive (it is equal to 6) and \( x = 4/3 \) is a global minimum.
Suppose \( S \) and \( T \) are subspaces of a vector space \((V, F)\). We define

1. **Sum**: \( X := S + T := \{s + t : s \in S \text{ and } t \in T\} \);

2. **Direct Sum**: If \( S \cap T = \{0\} \), then \( S \oplus T := S + T \).

3. **Orthogonal Sum**: Suppose \((V, F, \langle \cdot, \cdot \rangle)\) is an inner product space. Then \( S \oplus T \) is an orthogonal sum if \( \langle s, t \rangle = 0 \) for all \( s \in S \) and all \( t \in T \).

4. **Orthogonal complement**:
   \[ T = S^\perp := \{x \in X : \langle s, x \rangle = 0 \text{ for all } s \in S\}. \]
Lemma 1. Suppose $S$ and $T$ are subspaces of a vector space $(V, \mathbb{F})$.

1. $S + T = T + S$ and $S + T$ is a subspace of $V$.

2. $\dim(S + T) = \dim S + \dim T - \dim(S \cap T)$

3. $\dim(S \oplus T) = \dim S + \dim T$. Every $v \in S \oplus T$ can be decomposed uniquely as $v = s + t$, where $s \in S$ and $t \in T$. $s$ is called the projection of $v$ into $S$.

4. Pythagoras: If $\langle s, t \rangle = 0$ then $\|s + t\|^2 = \|s\|^2 + \|t\|^2$.

5. Here $\|v\| := \sqrt{\langle v, v \rangle}$. 

---

The Singular Value Decomposition and Least Squares Problems – p. 16/27
Column space of $A$ and null space of $A$:

- $\mathbb{R}^m = \text{span}(A) \oplus \ker(A^T)$ and this is an orthogonal sum.
- Thus $\ker(A^T) = \text{span}(A)^\perp$ the orthogonal complement of $\text{span}(A)$.
- Example

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{span}(A) = \text{span}(e_1, e_2), \quad \ker(A^T) = e_3.$$
Proof that $\mathbb{R}^m = \text{span}(A) \oplus \ker(A^T)$ using SVD

- $s^T t = 0$ for all $s \in \text{span}(A)$ and $t \in \ker(A^T)$.
- For if $s \in \text{span}(A)$ and $t \in \ker(A^T)$ then $s = Ax$ for some $x \in \mathbb{R}^n$ and $A^T t = 0$.
- But then $\langle s, t \rangle = (Ax)^T t = x^T (A^T t) = 0$
- Suppose $A = U \Sigma V^T = U_1 \Sigma_1 V_1^T$ is the SVD of $A$.
- Then $I = UU^T = [u_1 \ u_2] [U_1^1] = U_1 U_1^T + U_2 U_2^T$.
- For any $b \in \mathbb{R}^m$ we have $b = (U_1 U_1^T + U_2 U_2^T) b = b_1 + b_2$, where $b_1 := U_1 U_1^T b = AA^\dagger b$ with $A^\dagger = V_1 \Sigma_1^{-1} U_1^T$,
- and $b_2 := U_2 U_2^T$ belongs to $\ker(A^T)$ since $A^T b_2 = (V_1 \Sigma_1 U_1^T) U_2 U_2^T b = V_1 \Sigma_1 (U_1^T U_2) U_2^T b = 0$. 

The Singular Value Decomposition and Least Squares Problems – p. 18/27
Projections and pseudoinverse

- \( b_1 := AA^\dagger b \) is the projection of \( b \) into \( \text{span}(A) \).

- The matrix \( A^\dagger := V_1 \Sigma_1^{-1} U_1^T = V \Sigma^\dagger U^T \in \mathbb{R}^{n,m} \) is called the pseudoinverse of \( A = U \Sigma V^T \in \mathbb{R}^{m,n} \).

- \( \Sigma^\dagger := \begin{bmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n,m} \) is the pseudoinverse of \( \Sigma \).

- \( b_2 := (I - AA^\dagger) b \) is the projection of \( b \) into \( \ker(A^T) \).

- Example

  \[
  A = \begin{bmatrix}
  1 & 0 \\
  0 & 1 \\
  0 & 0 \\
  \end{bmatrix} = U \Sigma V^T = I_3 A I_2, \quad A^\dagger = I_2 \begin{bmatrix}
  1 & 0 & 0 \\
  0 & 1 & 0 \\
  \end{bmatrix} I_3 = \begin{bmatrix}
  1 & 0 & 0 \\
  0 & 1 & 0 \\
  \end{bmatrix}.
  \]

- \( b = \begin{bmatrix} b_1 \\
  b_2 \\
  b_3 \end{bmatrix}, \quad b_1 = AA^\dagger b = \begin{bmatrix}
  1 & 0 & 0 \\
  0 & 1 & 0 \\
  0 & 0 & 0 \\
  \end{bmatrix} b = \begin{bmatrix} b_1 \\
  b_2 \\
  0 \end{bmatrix} \)

- \( b_2 = (I_3 - AA^\dagger) b = \begin{bmatrix}
  0 & 0 & 0 \\
  0 & 0 & 0 \\
  0 & 0 & 1 \\
  \end{bmatrix} b = \begin{bmatrix} 0 \\
  0 \\
  b_3 \end{bmatrix} \)
The least squares problem always has a solution. The solution is unique if and only if \( A \) has linearly independent columns.

Proof. Let \( b = b_1 + b_2 \), where \( b_1 \in \text{span}(A) \) is the (orthogonal) projection of \( b \) into \( \text{span}(A) \) and \( b_2 \in \ker(A^T) \).

Since \( b_1 \in \text{span}(A) \) there is an \( x \in \mathbb{R}^n \) such that \( Ax = b_1 \). Thus \( b_2 = b - Ax \).

By Pythagoras, for any \( s \in \text{span}(A) \) with \( s \neq b_1 \)

\[
\|b - s\|^2 = \|b_1 - s\|^2 + \|b_2\|^2 = \|b_1 - s\|^2 + \|b - Ax\|^2 > \|b - Ax\|^2.
\]

Since the projection \( b_1 \) is unique, the least squares solution \( x \) is unique if and only if \( A \) has linearly independent columns.
The Normal Equations

**Theorem 3.** Any solution $x$ of the least squares problem is a solution of the linear system

$$A^T A x = A^T b.$$  

The system is nonsingular if and only if $A$ has linearly independent columns.

**Proof.** Since $b - A x \in \ker(A^T)$, we have $A^T (b - A x) = 0$ or $A^T A x = A^T b$.  

$A^T A$ is nonsingular. Suppose $A^T A x = 0$ for some $x \in \mathbb{R}^n$. Then

$$0 = x^T A^T A x = (A x)^T A x = \|A x\|_2^2.$$  

Hence $A x = 0$ which implies that $x = 0$ if and only if $A$ has linearly independent columns.

The linear system $A^T A x = A^T b$ is called the normal equations.
Linear Regression

\[ A = \begin{bmatrix}
1 & t_1 \\
1 & t_2 \\
\vdots & \vdots \\
1 & t_m
\end{bmatrix}, \quad b = \begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_m
\end{bmatrix}, \quad \min \sum_{i=1}^{m} (x_1 + t_i x_2 - y_i)^2. \]

\[ A = \begin{bmatrix}
1 & 1 & 2 & 1 & 4 & 1 \\
1 & 2 & 3 & 1 & 4 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & t_1 & t_2 & \cdots & t_m
\end{bmatrix}, \quad b = \begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_m
\end{bmatrix} \]

\[ A^T A = \begin{bmatrix}
5 & 15 \\
15 & 55
\end{bmatrix}, \quad c = A^T b = \begin{bmatrix}
16.0620 \\
58.6367
\end{bmatrix}, \quad 5x_1 + 15x_2 = 16.0620, \quad 15x_1 + 55x_2 = 58.6367 \]

\[ x_1 = 0.0772, \quad x_2 = 1.0451, \quad x = A \backslash b \]
Analysis of LSQ using $A = U \Sigma V^T$

Define $y := V^T x = \begin{bmatrix} V_1^T x \\ V_2^T x \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Recall $\|Uv\|_2 = \|v\|_2$ for any $U \in \mathbb{R}^{n,n}$ with $U^T U = I$ and any $v \in \mathbb{R}^n$.

$$\|b - Ax\|_2^2 = \|UU^T b - U\Sigma y\|_2^2 = \|U^T b - \Sigma y\|_2^2 = \|\begin{bmatrix} U_1^T b \\ U_2^T b \end{bmatrix} - \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\|_2^2$$

$$= \|\begin{bmatrix} U_1^T b - \Sigma_1 y_1 \\ U_2^T b \end{bmatrix}\|_2^2 = \|U_1^T b - \Sigma_1 y_1\|_2^2 + \|U_2^T b\|_2^2.$$

We have $\|b - Ax\|_2 \geq \|U_2^T b\|_2$ for all $x \in \mathbb{R}^n$ with equality if and only if

$$x = Vy = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} \Sigma_1^{-1} U_1^T b \\ y_2 \end{bmatrix} = V_1 \Sigma_1^{-1} U_1^T b + V_2 y_2, \text{ for all } y_2 \in \mathbb{R}^{n-r}.$$ (1)
The general solution of \( \min \| Ax - b \|_2 \)

- The columns of \( V_2 \) is a basis for \( \ker(A) \) so that
  \[ \ker(A) = \{ z = V_2 y_2 : y_2 \in \mathbb{R}^{n-r} \}. \]

- Therefore the solution set is
  \[ \{ x \in \mathbb{R}^n : \| Ax - b \|_2 \text{ is minimized} \} = A^\dagger b + \ker(A). \]

- If \( r = n \) then \( A \) has linearly independent columns and \( A^T A \) is nonsingular.

- Since \( A^T Ax = A^T b \) we obtain \( A^\dagger = (A^T A)^{-1} A^T \) in this case.
The Minimal Norm Solution

- Suppose $A$ is rank deficient ($r < n$).
- Let $x = A^\dagger b + V_2y_2$ be a solution of $\min \|Ax - b\|_2$.
- $A^\dagger b$ and $V_2y_2$ are orthogonal
- By Pythagoras $\|x\|_2^2 = \|A^\dagger b\|_2^2 + \|V_2y_2\|_2^2 \geq \|A^\dagger b\|_2^2$.
- The solution $x^* = A^\dagger b$ is called the minimal norm solution to the LSQ problem.
- Orthogonal. Since $V_2^TA^\dagger = (V_2^TV_1)\Sigma_1^{-1}U_1^T = 0$ we have $(V_2y_2)^TA^\dagger b = 0$ for any $y_2$. 
More on the pseudoinverse

- If $A$ is square and nonsingular then $A^\dagger = A^{-1}$.
- $A^\dagger$ is always defined.
- Thus $A^\dagger$ is a generalization of usual inverse.
- If $B \in \mathbb{R}^{n,m}$ satisfies
  1. $ABA = A$
  2. $BAB = B$
  3. $(BA)^T = BA$
  4. $(AB)^T = AB$
  then $B = A^\dagger$.
- Thus $A^\dagger$ is uniquely defined by these axioms.
Example

Show that the pseudoinverse of $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$ is $B = \frac{1}{4} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

We have $BA = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $AB = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Thus

1. $ABA = A$
2. $BAB = B$
3. $(BA)^T = BA$
4. $(AB)^T = AB$

and hence $A^\dagger = B$. 

The Singular Value Decomposition and Least Squares Problems – p. 27/27