Singular Value Decomposition

The singular value decomposition (SVD) is one of the most versatile decompositions in numerical linear algebra.

The SVD, among other things:

- characterizes the idea of matrix rank
- characterizes key norms
- is a valuable tool for perturbation analysis
- characterizes how near a matrix is to a rank deficient matrix
- provides a means of handling rank estimation rigorously
- solves linear least square problems independently of rank
- characterizes the key subspaces associated with arbitrary linear transformations, i.e., rectangular matrices with arbitrary rank
- can be used to relate and compare subspaces
- gives insight into when we can truncate “small” components of a matrix
**DEFINITION:** Given a matrix $A \in \mathbb{C}^{m \times n}$ with $m \geq n$ there exists unitary matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ such that

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^H$$

where $\Sigma \in \mathbb{R}^{n \times n}$ is a diagonal matrix with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_1 \geq 0$. 
The factorization is essentially unique.

- A pair of left and right singular vectors $Ue_i$ and $Ve_i$ respectively can be multiplied by an arbitrary multiple of modulus 1.

- If several singular values are equal then one can take the corresponding right singular vectors as any orthonormal basis for the associated singular space. The corresponding left singular vectors are specified by the identity $AV = U\Sigma$ for the nonzero singular values.
**LEMMA:** If $A \in \mathbb{C}^{m \times n}$ is has rank $k$ then $\sigma_{k+1} = \cdots = \sigma_n = 0$.

Therefore we have more structure:

$$A = \begin{pmatrix} U_k & U_{m-k} \end{pmatrix} \begin{pmatrix} \Sigma_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_k^H \\ V_{n-k}^H \end{pmatrix}$$

$$= U_k \Sigma_k V_k^H$$

The last form is called the **thin SVD** since it removes all directions from the factors except those corresponding to nonzero singular directions.

**LEMMA:** Given the thin SVD we have

$$\mathcal{R}(A) = \mathcal{R}(U_k)$$

So we have an orthonormal basis for $\mathcal{R}(A)$. 
SVD and Norms

For the 2-norm we have a simple relationship to the SVD

\[
\|A\|_2 = \max_{\|v\|_2 = 1} \|Av\|_2 = \sigma_1
\]

For the Frobenius norm we have

\[
\|A\|_F^2 = \sum_{i,j} |\alpha_{i,j}|^2
\]

\[
= \sum_{i=1}^{\text{rank}(A)} \sigma_i^2
\]

\[
= \sum_{i=1}^{\text{rank}(A)} \sigma_i^2
\]
Optimal Matrix Approximations

Suppose $A \in \mathbb{C}^{m \times n}$ has rank $k$ and let $A_i = U_i \Sigma_i V_i^H$ be defined by the first $i$ left and right singular vectors and the associated singular values.

$$
\|A - A_i\|_2 = \sigma_{i+1}
$$

$$
\|A - A_i\|_F^2 = \sum_{j=i+1}^{k} \sigma_j^2
$$

We also have optimality. If $Y_i$ is a matrix of rank $i$ then

$$
\|A - Y_i\|_2 \geq \sigma_{i+1}
$$

$$
\|A - Y_i\|_F \geq \sum_{j=i+1}^{k} \sigma_j^2
$$

Equality for each is when $Y_i = A_i$. 
So

- we have an optimal rank $i$ approximation to $A$

- if $A$ is uncertain then truncating the SVD to the $\sigma_i$ of the level of the uncertainty is acceptable.

- If $A \in \mathbb{C}^{m \times k}$ has rank $k$ then $\sigma_k$ is a measure of the distance to a matrix that does not have full column rank
The Pseudo-Inverse

Suppose $A$ is $m \times n$ and has rank $k$ such that

\[
A = U \left( \begin{array}{c} \Sigma \\ 0 \end{array} \right) V^H \\
= \left( \begin{array}{c} U_k \\ U_{m-k} \end{array} \right) \left( \begin{array}{cc} \Sigma_k & 0 \\ 0 & 0 \end{array} \right) \left( \begin{array}{c} V_k^H \\ V_{n-k}^H \end{array} \right) \\
= U_k \Sigma_k V_k^H
\]

**DEFINITION:** The pseudo-inverse $A^\dagger$

\[
A^\dagger = \left( \begin{array}{c} U_k \\ U_{m-k} \end{array} \right) \left( \begin{array}{cc} \Sigma_k^{-1} & 0 \\ 0 & 0 \end{array} \right) \left( \begin{array}{c} V_k^H \\ V_{n-k}^H \end{array} \right) \\
= V_k \Sigma_k^{-1} U_k^H
\]
Square nonsingular Case

We have $m = n = k$

\[
A = U\Sigma V^H \\
A^{-1} = V\Sigma^{-1}U^H \\
A^\dagger = V\Sigma^{-1}U^H
\]

So $Ax = b$ and $x = A^\dagger b$
Rectangular Full Column Rank

We have \( m > n = k \)

\[
\begin{align*}
A &= U_n \Sigma_n V_n^H \\
A^\dagger &= V_n \Sigma_n^{-1} U_n^H \\
(A^H A)^{-1} A^H &= (V_n \Sigma_n U_n^H U_n \Sigma_n V_n^H)^{-1} V_n \Sigma_n U_n^H \\
&= (V_n \Sigma_n^2 V_n^H)^{-1} V_n \Sigma_n U_n^H \\
&= V_n \Sigma_n^{-2} V_n^H V_n \Sigma_n U_n^H \\
&= V_n \Sigma_n^{-1} U_n^H \\
&= A^\dagger
\end{align*}
\]

So \( \|b - Ax\|_2 \) and \( x_{min} = A^\dagger b = (A^H A)^{-1} A^H b \)

where \( x_{min} \) is the unique minimizer.
Rectangular Column Rank Deficient

We have $m > n > k$

\[
A = U_k \Sigma_k V_k^H \\
A^\dagger = V_k \Sigma_k^{-1} U_k^H
\]

We have a projection onto $S = \mathcal{R}(A)$

\[
AA^\dagger = (U_k \Sigma_k V_k^H)(V_k \Sigma_k^{-1} U_k^H) \\
= U_k(\Sigma_k V_k^H V_k \Sigma_k^{-1}) U_k^H \\
= U_k U_k^H \\
= P_S
\]

So for the least squares problem $\|b - Ax\|_2$

\[
x_S = A^\dagger b \\
b_S = AA^\dagger b
\]
What does this have to do with the solution of the least squares problem? Solution not unique.

\[ \|b - Ax_S\|_2 = \|b - A(x_S + x_N(A))\|_2 \]

**THEOREM:** For the least squares problem \( \|b - Ax\|_2 \) where \( A \) is not full column rank the vector

\[ x_S = A^\dagger b \]

has the minimal residual and

\[ \|x_S\|_2 < \|x'\|_2 \]

where

\[ \|b - Ax_S\|_2 = \|b - Ax'\|_2 \]
• This is a minimal norm solution regularization

• Given that SVD is a robust numerical way of determining numerical rank, this is a very reliable method for solving near-column-rank deficient problems

We must consider the computational effort involved in determining the portion of the full SVD needed.

• not finite

• iterative approach that converges to the solution reliably and in an acceptable expected number of steps