THE SINGULAR VALUE DECOMPOSITION (Cont.)

- The Pseudo-inverse
- Use of SVD for least-squares problems
- Application to regularization
- Numerical rank
Pseudo-inverse of an arbitrary matrix

Let $A = U \Sigma V^T$ which we rewrite as

$$A = (U_1 \quad U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 \Sigma_1 V_1^T$$

Then the pseudo inverse of $A$ is

$$A^\dagger = V_1 \Sigma_1^{-1} U_1^T = \sum_{j=1}^{r} \frac{1}{\sigma_j} v_j u_j^T$$

The pseudo-inverse of $A$ is the mapping from a vector $b$ to the solution $\min_x \|Ax - b\|^2_2$ that has minimal norm (to be shown)

In the full-rank overdetermined case, the normal equations yield

$$x = (A^T A)^{-1} A^T b$$
Least-squares problem via the SVD

\( \text{Pb: } \min \| b - Ax \|_2 \) in general case. Consider SVD of \( A \):

\[
A = (U_1 \ U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = \sum_{i=1}^{r} \sigma_i v_i u_i^T
\]

Then left multiply by \( U^T \) to get

\[
\| Ax - b \|_2^2 = \left\| \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} U_1^T \\ U_2^T \end{pmatrix} b \right\|_2^2
\]

with \( \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} x \)

What are all least-squares solutions to the system? Among these which one has minimum norm?
**Answer:** From above, must have $y_1 = \Sigma_1^{-1}U_1^Tb$ and $y_2 =$ anything (free).

- Recall that $x = Vy$ and write

  \[
  x = [V_1, V_2] \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = V_1y_1 + V_2y_2 \\
  = V_1\Sigma_1^{-1}U_1^Tb + V_2y_2 \\
  = A^\dagger b + V_2y_2
  \]

- Note: $A^\dagger b \in \text{Ran}(A^T)$ and $V_2y_2 \in \text{Null}(A)$.

- Therefore: least-squares solutions are of the form $A^\dagger b + w$ where $w \in \text{Null}(A)$.

- Smallest norm when $y_2 = 0$. 

10-4 AB: 1.1. 2.2. 2.4; TB: 4-5; GvL 2.4, 5.4-5 – SVD1
Minimum norm solution to $\min_x \|Ax - b\|_2^2$ satisfies $\Sigma_1 y_1 = U_1^T b$, $y_2 = 0$. It is:

$$x_{LS} = V_1 \Sigma_1^{-1} U_1^T b = A^\dagger b$$

If $A \in \mathbb{R}^{m \times n}$ what are the dimensions of $A^\dagger$, $A^\dagger A$, $AA^\dagger$?

Show that $A^\dagger A$ is an orthogonal projector. What are its range and null-space?

Same questions for $AA^\dagger$. 
Moore-Penrose Inverse

The pseudo-inverse of $A$ is given by

$$A^\dagger = V \begin{pmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T = \sum_{i=1}^{r} \frac{v_i u_i^T}{\sigma_i}$$

Moore-Penrose conditions:

The pseudo inverse of a matrix is uniquely determined by these four conditions:

(1) $AXA = A$
(2) $XAX = X$
(3) $(AX)^H = AX$
(4) $(XA)^H =XA$

In the full-rank overdetermined case, $A^\dagger = (A^T A)^{-1} A^T$
SVD can give much information about solving overdetermined and underdetermined linear systems.

Let $A$ be an $m \times n$ matrix and $A = U\Sigma V^T$ its SVD with $r = \text{rank}(A)$, $V = [v_1, \ldots, v_n]$ $U = [u_1, \ldots, u_m]$. Then

$$x_{LS} = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i$$

minimizes $\|b - Ax\|_2$ and has the smallest 2-norm among all possible minimizers. In addition,

$$\rho_{LS} \equiv \|b - Ax_{LS}\|_2 = \|z\|_2$$ with $z = [u_{r+1}, \ldots, u_m]^T b$
A restatement of the first part of the previous result:

Consider the general linear least-squares problem

$$\min_{x \in S} \|x\|_2, \quad S = \{x \in \mathbb{R}^n | \|b - Ax\|_2 \text{ min}\}.$$ 

This problem always has a unique solution given by

$$x = A^+ b$$
Consider the matrix:

\[ A = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 0 & -2 & 1 \end{pmatrix} \]

- Compute the thin SVD of \( A \)
- Find the matrix \( B \) of rank 1 which is the closest to the above matrix in the 2-norm sense.
- What is the pseudo-inverse of \( A \)?
- What is the pseudo-inverse of \( B \)?
- Find the vector \( x \) of smallest norm which minimizes \( \|b - Ax\|_2 \) with \( b = (1, 1)^T \)
- Find the vector \( x \) of smallest norm which minimizes \( \|b - Bx\|_2 \) with \( b = (1, 1)^T \)
Ill-conditioned systems and the SVD

- Let $A$ be $m \times m$ and $A = U\Sigma V^T$ its SVD.

- Solution of $Ax = b$ is $x = A^{-1}b = \sum_{i=1}^{m} \frac{u_i^Tb}{\sigma_i} v_i$.

- When $A$ is very ill-conditioned, it has many small singular values. The division by these small $\sigma_i$'s will amplify any noise in the data. If $\tilde{b} = b + \epsilon$ then

$$A^{-1}\tilde{b} = \sum_{i=1}^{m} \frac{u_i^Tb}{\sigma_i} v_i + \sum_{i=1}^{m} \frac{u_i^T\epsilon}{\sigma_i} v_i$$

- Result: solution could be completely meaningless.
Remedy: SVD regularization

Truncate the SVD by only keeping the $\sigma_i'$s that are $\geq \tau$, where $\tau$ is a threshold

Gives the Truncated SVD solution (TSVD solution:)

$$x_{TSVD} = \sum_{\sigma_i \geq \tau} \frac{u_i^Tb}{\sigma_i} v_i$$

Many applications [e.g., Image and signal processing,..]
**Numerical rank and the SVD**

- Assuming the original matrix $A$ is exactly of rank $k$ the computed SVD of $A$ will be the SVD of a nearby matrix $A + E$ – Can show: $|\hat{\sigma}_i - \sigma_i| \leq \alpha \sigma_1 u$

- Result: zero singular values will yield small computed singular values and $r$ larger sing. values.

- Reverse problem: *numerical rank* – The $\epsilon$-rank of $A$:
  
  $$r_\epsilon = \min\{\text{rank}(B) : B \in \mathbb{R}^{m \times n}, \|A - B\|_2 \leq \epsilon\},$$

1. Show that $r_\epsilon$ equals the number sing. values that are $> \epsilon$
2. Show: $r_\epsilon$ equals the number of columns of $A$ that are linearly independent for any perturbation of $A$ with norm $\leq \epsilon$.

- Practical problem: How to set $\epsilon$?
Pseudo inverses of full-rank matrices

Case 1: \( m > n \) Then \( A^\dagger = (A^T A)^{-1} A^T \)

Thin SVD is \( A = U_1 \Sigma_1 V_1^T \) and \( V_1, \Sigma_1 \) are \( n \times n \). Then:

\[
(A^T A)^{-1} A^T = (V_1 \Sigma_1^2 V_1^T)^{-1} V_1 \Sigma_1 U_1^T \\
= V_1 \Sigma_1^{-2} V_1^T V_1 \Sigma_1 U_1^T \\
= V_1 \Sigma_1^{-1} U_1^T \\
= A^\dagger
\]

Example: Pseudo-inverse of

\[
\begin{pmatrix}
0 & 1 \\
1 & 2 \\
2 & -1 \\
0 & 1
\end{pmatrix}
\]
is?
Case 2: $m < n$

Then $A^\dagger = A^T(AA^T)^{-1}$

Thin SVD is $A = U_1 \Sigma_1 V_1^T$. Now $U_1$, $\Sigma_1$ are $m \times m$ and:

$$A^T(AA^T)^{-1} = V_1 \Sigma_1 U_1^T[U_1 \Sigma_1^2 U_1^T]^{-1}$$
$$= V_1 \Sigma_1 U_1^T U_1 \Sigma_1^{-2} U_1^T$$
$$= V_1 \Sigma_1 \Sigma_1^{-2} U_1^T$$
$$= V_1 \Sigma_1^{-1} U_1^T$$
$$= A^\dagger$$

Example: Pseudo-inverse of

$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 2 & -1 & 1 \end{pmatrix}$

is?

Mnemonic: The pseudo inverse of $A$ is $A^T$ completed by the inverse of the smallest of $(A^TA)^{-1}$ or $(AA^T)^{-1}$ where it fits (i.e., left or right)