LEAST SQUARE METHODS

The Singular Value Decomposition (SVD)
Minimum Norm Solution

Dr. Yogananda Isukapalli
• Review of some results from Linear Algebra

Suppose the vector $\hat{x}$ does not satisfy exactly the algebraic equation $A\hat{x}=b$, then define an approximate problem:

$$A\hat{x} \approx b$$

then define a residual vector

$$e = b - A\hat{x}$$

also define the norm of $e$ by $|e|$ that satisfies: $|e| > 0$

for $e \neq 0$ and $|0| = 0$

$$|ce| = |c| \cdot |e|$$
the norm definition must also satisfy the triangle inequality
\[ \|e+s\| \leq \|e\| + \|s\| \]

the norm is generally computed
\[ \|e\|_E = (e^T e)^{1/2} \]  Euclidean Norm

Follows:
\[ e^T e = \sum_{i=1}^{m} e_i^2 \]  scalar

known as the sum of the error squares.

The least-squares solution \( x \) of \( Ax \approx b \) is that set of parameters which minimizes this sum of squares where \( \text{rank} (A) < n \) (\( n \) equations in \( m \) unknowns).

The solution is not unique.
• The Linear Least-Squares Problem

Start from \( e = b - Ax \)

the minimization of \( e^T e \) with respect to \( x \) will yield

\[
\mathbb{A}^T \mathbb{A}x = \mathbb{A}^T b
\]

Normal Equations

which \( x \) must satisfy:

• Inverse of a Matrix

For square matrices, \( \mathbb{A}^{-1} \) is defined:

\[
\mathbb{A}^{-1} \mathbb{A} = \mathbb{A} \mathbb{A}^{-1} = \mathbb{I}_n
\]

\( \mathbb{A}^{-1} \) exists only if \( \mathbb{A} \) has full rank. Then

\[
x = \mathbb{A}^{-1} b
\]
when $\mathcal{A}$ is rectangular,

$$x = \mathcal{A}^+ b$$

that is

$$\mathcal{A}^+ \mathcal{A} = I_n$$

But when $\mathcal{A}$ has only $k$ linearly independent columns, then

$$\mathcal{A}^+ \mathcal{A} = \begin{bmatrix} \mathcal{L}K & 0 \\ 0 \\ \overbrace{0}^{n-k} \end{bmatrix}$$

$x$ in this case is not unique. In which case

$$x = \mathcal{A}^+ b + (I_n - \mathcal{A}^+ \mathcal{A})g \quad (A)$$

where $g$ is any vector of order $n$. 
The normal equations must still be satisfied. For full rank case

$$A^+ = (A^T A)^{-1} A^T$$

In the rank - deficient case : Use (A)

$$A^T A x = A^T A A^+ b + (A^T A - A^T A A^+ A) g = A^T b$$

This equality is true if

$$A^T A A^+ = A^T$$  \hspace{1cm} (B)

By requiring $A^+$ to satisfy :

$$A A^+ A = A$$  \hspace{1cm} (C)

$$(A A^+)^T = A A^+$$  \hspace{1cm} (D)

(B) is satisfied if (C) and (D) are true.
In addition to (C) and (D), For (A) to be minimum length least-squares solution, It is necessary also that $x^T x$ be minimum.

From (A)

$$x^T x = b^T (A^+)^T A^+ b + g^T (I - A^+ A) T (I - A^+ A) g$$

$$+ 2g^T (I - A^+ A) T A^+ b$$

It attains minimum at $g = 0$

if $(I - A^+ A)^T$ is the annihilator of $A^+ b$ which ensures that two contributions from $b$ and $g$ to $x^T x$ are orthogonal.

This will imply two more conditions:

$$A^+ A A^+ = A^+ \quad \text{(E)}$$

$$(A^+ A)^T = A^+ A \quad \text{(F)}$$
$A^+$ is the generalised inverse proposed by Moore-Penrose and must satisfy (C), (D), (E) and (F).

- The Singular Value Decomposition

Consider transforming a $m \times n$ matrix $A$ into another real $m \times n$ matrix $B$ whose columns are orthogonal.

Find $\mathcal{V}$ such that

$$B = A\mathcal{V} = (b_1, b_2, \ldots, b_n)$$

where

$$b_i^T b_i = \sigma_i^2 \delta_{ij}$$

and

$$\mathcal{V} \mathcal{V}^T = \mathcal{V}^T \mathcal{V} = I_n$$

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$
$\sigma_i$ may be either positive or negative since $\sigma_i^2$ is only defined by $b_i^T b_i$. If $\sigma_i$ are taken positive, then they are called singular values of the matrix $A$.

The vectors $u_j = \frac{b_j}{\sigma_j}$ when $\sigma_j$ are not zero, are unit orthogonal vectors.

Define now:

$$B = U\Sigma$$

where $$U^T U = I_n$$

If we choose the first $k$ of the singular values to be the non-zero ones, then
\( \mathcal{U}^T \mathcal{U} = \begin{bmatrix} I_k \\ 0_{n-k} \end{bmatrix} \)

If we sort \( \sigma \)'s such that
\[
\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq \sigma_k \geq \ldots \geq \sigma_n
\]
then
\[
\mathcal{A} = \sum_{j=1}^{n} u_j \sigma_j u_j^T
\]

Partial sums of this series give a sequence of approximations
\[
\tilde{\mathcal{A}}_1, \tilde{\mathcal{A}}_2, \ldots , \tilde{\mathcal{A}}_n
\]
where \( \tilde{\mathcal{A}}_n = \mathcal{A} \)
Finally

\[ AV = U \Sigma \]

then the orthogonality of \( V \) implies:

\[ A = U \Sigma V^T \]

which is the SVD of \( A \)
• The SVD and the Least-Squares Filter

Starting from Normal Equations

\[ \mathcal{A}^H \mathcal{A} \hat{\mathcal{w}} = \mathcal{A}^H \mathcal{d} \]

we have solved:

\[ \hat{\mathcal{w}} = (\mathcal{A}^H \mathcal{A})^{-1} \mathcal{A}^H \mathcal{d} \]

where \( \hat{\mathcal{w}} \) is the least-square estimate of the tap-weight vector of a transversal filter model. \( \mathcal{A} \) is the data matrix and \( \mathcal{d} \) is the desired data vector.
• The SVD

Start from:  \( A\hat{w} = d \)

- \( A \) is a \( K \)-by-\( M \) matrix
- \( d \) is a \( K \)-by-1 vector

Given a \( K \times M \) rectangular data matrix \( A \), the SVD says that there exists two Unitary matrices \( V \) and \( U \), such that we may write

\[
UAV = \begin{bmatrix}
\Sigma & 0 \\
0 & 0
\end{bmatrix}
\]

where \( \Sigma \) is a diagonal matrix:

\[
\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_W)
\]

where \( \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_W > 0 \)

\( W = \text{rank}(A) \) the rank \( W \leq \min(K, M) \)
FIGURE 8.8 Diagrammatic interpretation of the singular-value decomposition theorem.

$$U^H \times A \times V = \Sigma \begin{array}{c} \mathbf{O} \\ \mathbf{O} \end{array}$$

$O$ : Null matrix
The pseudoinverse of the matrix $\mathcal{A}$

Given $\mathcal{A}$ is $K$ by $M$ then:

$$A^+ = \mathcal{V} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathcal{U}^H$$

where:

$$\Sigma^{-1} = \text{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_W^{-1})$$

\(a\). Over determined System

$\mathcal{A}$ has a full rank

$K > M$ and the rank $W = M$

Then

$$A^+ = (\mathcal{A}^H \mathcal{A})^{-1} \mathcal{A}^H$$
b. Underdetermined system
\[ M > K \quad \text{and the rank } W = K \]
then \[ A^+ = A^H (A^A)^{-1} \]

• The Least-Squares problem: Minimum Norm Solution
\[ \hat{w} = A^+ d \]

where \[ A^+ = V \left[ \begin{array}{cc} \Sigma^{-1} & 0 \\ 0 & 0 \end{array} \right] U^H \]

the solution is unique in that the shortest length possible in the Euclidean sense even when null (A) ≠ 0 (or rank-deficient case).
Review:
\[
\hat{w} = (A^H A)^{-1} A^H d \\
\varepsilon_{\min} = d^H d - d^H A (A^H A)^{-1} A^H d
\]

Start From:
\[
\varepsilon_{\min} = d^H d - d^H A \hat{w} \\
= d^H (d - A \hat{w}) \\
= d^H U U^H (d - A \hat{V} \hat{V}^H \hat{w})
\]

Now Define:
\[
\hat{V}^H \hat{w} = z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}
\]
\[
U^H d = c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}
\]

\(z_1\) and \(c_1\) are \(W\) by 1 vectors.
Now:

\[ \varepsilon_{\min} = d^H \mathcal{U} (\mathcal{U}^H d - \mathcal{U}^H A \mathcal{V} \mathcal{V}^H \hat{w}) \]

\[ = d^H \mathcal{U} \left( \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - \begin{bmatrix} \Sigma \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right) \]

\[ = d^H \mathcal{U} \begin{bmatrix} c_1 - \Sigma z_1 \\ c_2 \end{bmatrix} \]

\[ \varepsilon_{\min} \text{ is independent of } z_2 \text{ and can be arbitrary.} \]

For \( \varepsilon_{\min} \) to be minimum, set

\[ c_1 = \Sigma z_1 \]

or

\[ z_1 = \Sigma^{-1} c_1 \]
If we set $z_2 = 0$,

\[ \hat{w} = \Sigma^{-1} \begin{bmatrix} c_1 \\ 0 \end{bmatrix} \]

or

\[ \hat{w} = \Sigma^{-1} \begin{bmatrix} 0 \\ c_1 \\ c_2 \end{bmatrix} \]

\[ = \Sigma^{-1} \begin{bmatrix} 0 \\ 0 \\ c_2 \end{bmatrix} U^H d \]

\[ = A^+ d \]

which is the desired result, the pseudoinverse solution to the least-squares problem when data matrices are rectangular matrices.

Unique and has a minimum norm:

Since \( \Sigma \Sigma^H = I \)

\[ \| \hat{w} \|^2 = \| \Sigma^{-1} c_1 \|^2 \]
Consider a second solution:

\[
\begin{bmatrix}
\bar{w}' \\
\Sigma^{-1} c_1
\end{bmatrix}
\bar{z}_2 
eq 0
\]

Then

\[
|\bar{w}'|^2 = |\Sigma^{-1} c_1|^2 + |\bar{z}_2|^2
\]

For any \( \bar{z}_2 \neq 0 \), follows:

\[
|\hat{w}| < |\bar{w}'|
\]

\( \hat{w} \) is the minimum norm-solution to a linear transversal filter problem even when null( A ) \( \neq 0 \).
Other Representations of the Minimum-Norm Solution

Start From

\[
A^+ = U \left[ \begin{array}{ccc} \Sigma^{-1} & 0 \\ 0 & 0 \end{array} \right] U^H \quad (a)
\]

\[\Sigma^{-1} = \text{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_W^{-1})\]

\[W = \text{rank of } A\]

Over determined case: \( K > M \)

Form \( A^H A \) \( M \times M \) Matrix

\(\begin{array}{c}
A^H \\
A
\end{array}\)

Hermitian and non-negative definite eigen values are real and non negative numbers
Denote eigenvalues of $A^H A$:

$$\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2$$

where

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_W > 0$$

and

$$\sigma_W = \sigma_W = \ldots = \sigma_M = 0$$

$A^H A$ has the same rank as $A$

The eigenvalue-eigen vector decomposition of the matrix $A^H A$:

$$\Lambda^H A^H A \Lambda = \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

Now partition the Unitary matrix $\Lambda$:

$$\Lambda = [\Lambda_1, \Lambda_2]$$

$$\Lambda_1 = [\nu_1, \nu_2, \ldots, \nu_W] \text{ M by W matrix}$$

$$\Lambda_2 = [\nu_{W+1}, \nu_{W+2}, \ldots, \nu_M] \text{ M by (M-W) matrix}$$
Note there are W non zero eigen values:
\[ \mathcal{V}_1^H \mathcal{V}_2 = 0 \]

Follows that:
\[ \mathcal{V}_1^H \mathcal{A}^H \mathcal{A} \mathcal{V}_1 = \Sigma^2 \quad (b) \]

or
\[ \Sigma^{-1} \mathcal{V}_1^H \mathcal{A}^H \mathcal{A} \mathcal{V}_1 \Sigma^{-1} = I \quad (c) \]

\[ \mathcal{V}_2^H \mathcal{A}^H \mathcal{A} \mathcal{V}_2 = 0 \quad (d) \]
\[ \mathcal{A} \mathcal{V}_2 = 0 \]

Define K by W matrix
\[ \mathcal{U}_1 = \mathcal{A} \mathcal{V}_1 \Sigma^{-1} \quad (e) \]

From (c)
\[ \mathcal{U}_1^H \mathcal{U}_1 = \mathcal{L} \quad (f) \]
Define $\mathcal{U}_2$ by $(\mathbf{K} - \mathbf{W})$ matrix
\[
\mathcal{U} = \begin{bmatrix} \mathcal{U}_1 & \mathcal{U}_2 \end{bmatrix}
\]
is a unitary matrix.

\[
\mathcal{U}_1^H \mathcal{U}_2 = 0
\]

Now:
\[
\hat{\mathbf{w}} = \mathbf{A}^+ \mathbf{d}
\]

Use (a)
\[
\hat{\mathbf{w}} = \mathcal{V} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathcal{U}^H \mathbf{d}
\]

Using
\[
\mathcal{V} = \begin{bmatrix} \mathcal{V}_1 & \mathcal{V}_2 \end{bmatrix}
\]
and (e)

\[
\hat{\mathbf{w}} = (\mathcal{V}_1 \Sigma^{-1})(\mathbf{A} \mathcal{V}_1 \Sigma^{-1})^H \mathbf{d}
\]

\[
= \mathcal{V}_1 \Sigma^{-1} \Sigma^{-1} \mathcal{V}_1^H \mathbf{A}^H \mathbf{d}
\]

\[
= \mathcal{V}_1 \Sigma^{-2} \mathcal{V}_1 \mathbf{A}^H \mathbf{d}
\]

\[
\hat{\mathbf{w}} = \sum_{i=1}^W \frac{\mathbf{v}_i}{\sigma_i^2} \mathbf{v}_i^H \mathbf{A}^H \mathbf{d}
\]  

(1)
Underdetermined case:

K (no. of equations) < M (the no. of unknowns)

we can get similar equation:

\[
\hat{w} = \sum_{i=1}^{W} \frac{u_i^H d}{\sigma_i^2} A^H u_i
\]

where

\[ \mathcal{U}_1 = [u_{-1}^i, u_2^i, \ldots, u_{W}^i] \]
\[ \mathcal{U}_2 = [u_{W+1}^i, u_{W+2}^i, \ldots, u_{K}^i] \]
\[ \mathcal{U} = [\mathcal{U}_1, \mathcal{U}_2] \quad \text{Unitary Matrix} \]
\[ \mathcal{U}_1^H \mathcal{U}_2 = 0 \]
\[ \mathcal{U}^H A A^H \mathcal{U} = \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \]

\( A A^H \) is now K by K matrix
Computation of (I) and (II)

Step 1. Compute the SVD of the data matrix $\mathcal{A}$, that is find the singular values $\sigma_1, \sigma_2, \ldots, \sigma_W$ and associated right-singular vectors $v_1, v_2, \ldots, v_W$ and the left-singular vectors $u_1, u_2, \ldots, u_W$.

Step 2. Compute $\hat{w}$ by (I) for ($K > M$) over determined case and by (II) for the underdetermined case ($K < M$).

the SVD provides a numerically stable solution for $\hat{w}$. 