# LEAST SQUARE METHODS

# The Singular Value Decomposition ( SVD ) Minimum Norm Solution

Dr. Yogananda Isukapalli

- Review of some results from Linear Algebra

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

$$\underset{\sim}{A}\underline{x}\approx\underline{b}$$

then define a residual vector

$$\underline{e}=\underline{b}-\underset{\sim}{A}\underline{x}$$

also define the norm of $\underline{e}$ by $\|\underline{e}\|$ that satisfies： $\|\underline{e}\|>0$

for $\underline{e}\neq 0$ and $\|0\|=0$

$$\|c\underline{e}\|=|c|\cdot\|\underline{e}\|$$

the norm definition must also satisfy the triangle inequality

$$\|\underline{e}+\underline{s}\| \leq \|\underline{e}\| + \|\underline{s}\|$$

the norm is generally computed

$$\|\underline{e}\|_E = (\underline{e}^T \underline{e})^{1/2} \quad \text{Euclidean Norm}$$

Follows :

$$\underline{e}^T \underline{e} = \sum_{i=1}^{m} e_i^2 \qquad \text{scalar}$$

known as the sum of the error squares.

The least-squares solution $\underline{x}$ of $\underset{\sim}{A}\underline{x} \approx \underline{b}$ is that set of parameters which minimizes this sum of squares where rank ( $\underset{\sim}{A}$ ) $< n$ ($n$ equations in $m$ unknowns ). The solution is not unique.

- The Linear Least-Squares Problem

Start from $\underline{e} = \underline{b} - \underset{\sim}{A}\underline{x}$

the minimization of $\underline{e}^T \underline{e}$ with respect to $\underline{x}$ will yield

$$\boxed{\underset{\sim}{A}^T \underset{\sim}{A}\underline{x} = \underset{\sim}{A}^T \underline{b}}$$ Normal Equations

which $\underline{x}$ must satisfy :

- Inverse of a Matrix

For square matrices, $\underset{\sim}{A}^{-1}$ is defined :

$$\underset{\sim}{A}^{-1}\underset{\sim}{A} = \underset{\sim}{A}\underset{\sim}{A}^{-1} = \underset{\sim}{I}_n$$

$A^{-1}$ exists only if $\underset{\sim}{A}$ has full rank. Then

$$\underline{x} = \underset{\sim}{A}^{-1}\underline{b}$$

when $\underset{\sim}{A}$ is rectangular,

$$\underline{x} = \underset{\sim}{A}^{+} \underline{b}$$

that is $\underset{\sim}{A}^{+} \underset{\sim}{A} = I_n$

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

$$\underset{\sim}{A}^{+} \underset{\sim}{A} = \begin{bmatrix} \underset{\sim}{I}_K & & & \\ & 0 & & \\ & & \cdot & \\ & & & 0 \end{bmatrix}$$

with the last $n-k$ diagonal entries marked by a brace.

$\underline{x}$ in this case is not unique. In which case

$$\underline{x} = \underset{\sim}{A}^{+} \underline{b} + (\underset{\sim}{I}_n - \underset{\sim}{A}^{+} \underset{\sim}{A}) \underline{g} \qquad \text{(A)}$$

where $\underline{g}$ is any vector of order $n$.

The normal equations must still be satisfied.
For full rank case

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

In the rank - deficient case : Use (A)

$$\underset{\sim}{A}^{T}\underset{\sim}{A}\underline{x} = \underset{\sim}{A}^{T}\underset{\sim}{A}\underset{\sim}{A}^{+}\underline{b} + (\underset{\sim}{A}^{T}\underset{\sim}{A} - \underset{\sim}{A}^{T}\underset{\sim}{A}\underset{\sim}{A}^{+}\underset{\sim}{A})\underline{g}$$
$$= \underset{\sim}{A}^{T}\underline{b}$$

This equality is true if

$$\underset{\sim}{A}^{T}\underset{\sim}{A}\underset{\sim}{A}^{+} = \underset{\sim}{A}^{T} \qquad \text{(B)}$$

By requiring $\underset{\sim}{A}^{+}$ to satisfy :

$$\underset{\sim}{A}\underset{\sim}{A}^{+}\underset{\sim}{A} = \underset{\sim}{A} \qquad \text{(C)}$$
$$(\underset{\sim}{A}\underset{\sim}{A}^{+})^{T} = \underset{\sim}{A}\underset{\sim}{A}^{+} \qquad \text{(D)}$$

(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 $\underline{x}^T \underline{x}$ be minimum .

From (A)

$$\underline{x}^T \underline{x} = \underline{b}^T (\underset{\sim}{A}^+)^T \underset{\sim}{A}^+ \underline{b} + \underline{g}^T (\underset{\sim}{I} - \underset{\sim}{A}^+ \underset{\sim}{A})^T (\underset{\sim}{I} - \underset{\sim}{A}^+ \underset{\sim}{A}) \underline{g} + 2\underline{g}^T (\underset{\sim}{I} - \underset{\sim}{A}^+ \underset{\sim}{A})^T \underset{\sim}{A}^+ \underline{b}$$

It attains minimum at $\underline{g} = 0$

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

This will imply two more conditions :

$$\underset{\sim}{A}^+ \underset{\sim}{A} \underset{\sim}{A}^+ = \underset{\sim}{A}^+ \qquad \text{(E)}$$

$$(\underset{\sim}{A}^+ \underset{\sim}{A})^T = \underset{\sim}{A}^+ \underset{\sim}{A} \qquad \text{(F)}$$

$A^+$ is the generalised inverse proposed by Moore-Penrose and must satisfy (C), (D) ,(E) and (F).

- The Singular Value Decomposition

Consider transforming a m x n matrix $\underset{\sim}{A}$ into another real m x n matrix $\underset{\sim}{B}$ whose columns are orthogonal.

Find $\underset{\sim}{V}$ such that

$$\underset{\sim}{B} = \underset{\sim}{A}\underset{\sim}{V} = (\underline{b}_1, \underline{b}_2, \ldots \ldots \underline{b}_n)$$

where $\quad \underline{b}_i^T \underline{b}_i = \sigma_i^2 \delta_{ij}$

and $\quad \underset{\sim}{V}\underset{\sim}{V}^T = \underset{\sim}{V}^T \underset{\sim}{V} = \underset{\sim}{I}_n$

$$\delta_{ij} = \begin{vmatrix} 0 & i \neq j \\ 1 & i = j \end{vmatrix}$$

$\sigma_i$ may be either positive or negative since $\sigma_1^2$ is only defined by $\underline{b}_i^T \underline{b}_i$. If $\sigma_i$ are taken positive, then they are called singular values of the matrix $\underset{\sim}{A}$.

the vectors $\underline{u}_j = \underline{b}_j / \sigma_j$

when $\sigma_j$ are not zero, are unit orthogonal vectors.

Define now :

$$\underset{\sim}{B} = \underset{\sim}{U}\,\underset{\sim}{\Sigma}$$

where $\underset{\sim}{U}^T \underset{\sim}{U} = I_n$

If we choose the first $k$ of the singular values to be the non-zero ones, then

$$\underset{\sim}{U}^T \underset{\sim}{U} = \begin{pmatrix} \underset{\sim}{I}_k & \\ & \underset{\sim}{0}_{n-k} \end{pmatrix}$$

If we sort $\sigma_i$'s such that

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq \sigma_k \geq \ldots \geq \sigma_n$$

then

$$\underset{\sim}{A} = \sum_{j=1}^{n} u_j \sigma_j u_j^T$$

Partial sums of this series give a sequence of approximations

$$\underset{\sim}{\widetilde{A}}_1, \underset{\sim}{\widetilde{A}}_2, \ldots\ldots, \underset{\sim}{\widetilde{A}}_n$$

where $\underset{\sim}{\widetilde{A}}_n = \underset{\sim}{A}$

Finally

$$\underset{\sim}{A}\underset{\sim}{V} = \underset{\sim}{U}\underset{\sim}{\Sigma}$$

then the orthogonality of $\underset{\sim}{V}$ implies :

$$\underset{\sim}{A} = \underset{\sim}{U}\underset{\sim}{\Sigma}\underset{\sim}{V}^T$$

which is the SVD of $\underset{\sim}{A}$

---

- The SVD and the Least-Squares Filter

Starting from Normal Equations

$$\underset{\sim}{A}^H \underset{\sim}{A}\underline{w} = \underset{\sim}{A}^H \underline{d}$$

we have solved :

$$\underline{\hat{w}} = (\underset{\sim}{A}^H \underset{\sim}{A})^{-1} \underset{\sim}{A}^H \underline{d}$$

where $\underline{\hat{w}}$ is the least-square estimate of the tap-weight vector of a transversal filter model. $\underset{\sim}{A}$ is the data matrix and $\underline{d}$ is the desired data vector.

- The SVD

Start from : $\underset{\sim}{A}\underline{\hat{w}}=\underline{d}$

$\underset{\sim}{A}$ is a K-by-M matrix

$\underline{d}$ is a K-by-1 vector

Given a K x M rectangular data matrix $\underset{\sim}{A}$, the SVD says that there exists two Unitary matrices $\underset{\sim}{V}$ and $\underset{\sim}{U}$, such that we may write

$$\underset{\sim}{U}\,\underset{\sim}{A}\,\underset{\sim}{V}=\begin{bmatrix}\underset{\sim}{\Sigma} & \underline{0}\\ \underline{0} & \underline{0}\end{bmatrix}$$

where $\underset{\sim}{\Sigma}$ is a diagonal matrix :

$$\underset{\sim}{\Sigma}=\mathrm{diag}(\sigma_1,\sigma_2,\ldots\ldots,\sigma_W)$$

where $\sigma_1 \ge \sigma_2 \ge \ldots\ldots \ge \sigma_W > 0$

$\mathrm{W}=\mathrm{rank}\,(\underset{\sim}{A})$ the rank $\mathrm{W} \le \min(\mathrm{K},\mathrm{M})$

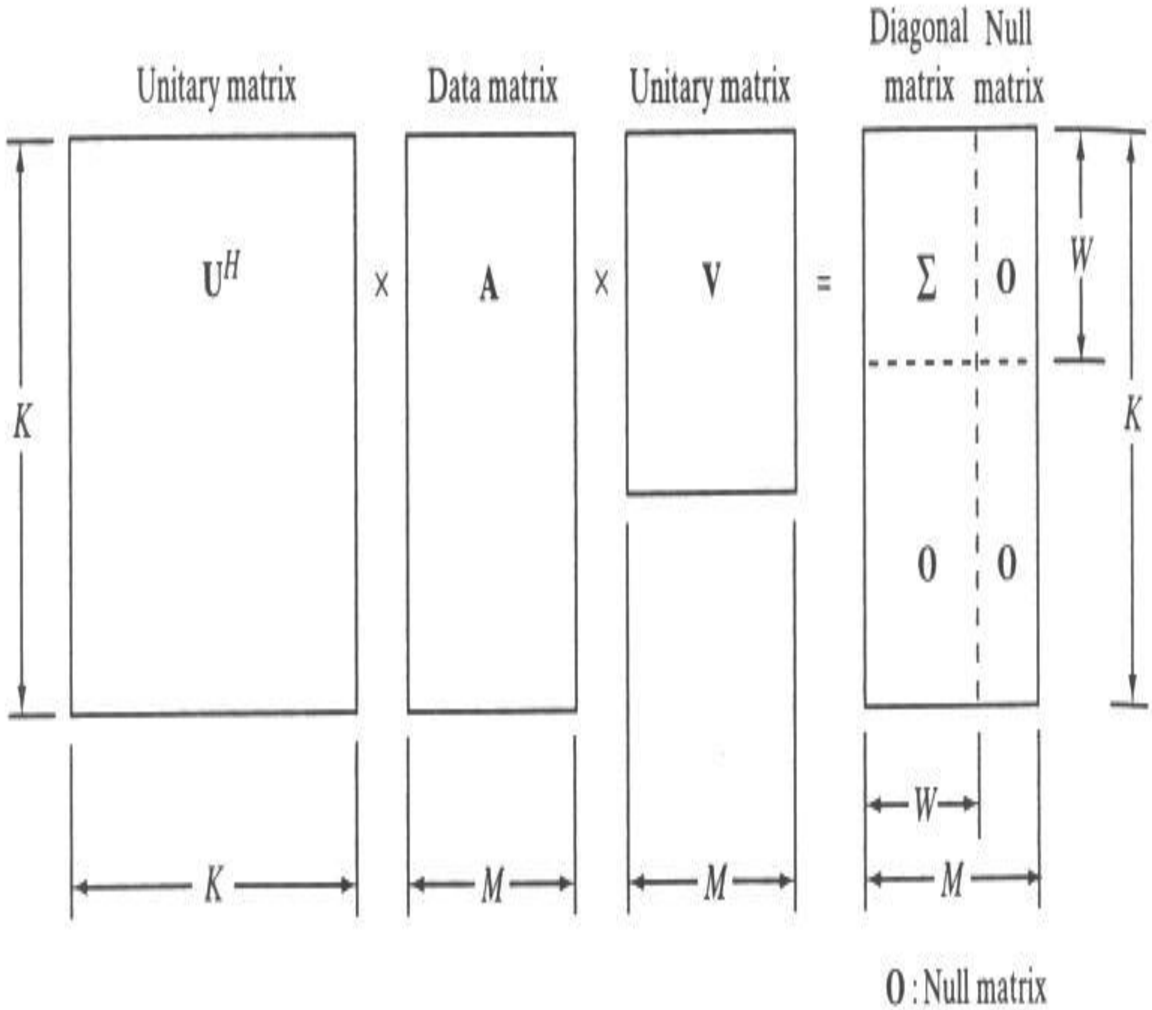


FIGURE 8.8 Diagrammatic interpretation of the singular-value decomposition theorem.

- The pseudoinverse of the matrix $\underset{\sim}{A}$

Given $\underset{\sim}{A}$ is K by M

then :

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

where :

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

$\underline{a}$. Over determined System

$\underset{\sim}{A}$ has a full rank

K > M and the rank W = M

Then $\underset{\sim}{A}^{+} = (\underset{\sim}{A}^{H} \underset{\sim}{A})^{-1} \underset{\sim}{A}^{H}$

$\underline{b}$. Underdetermined system

M > K and the rank W = K

then $\underset{\sim}{A}^{+} = \underset{\sim}{A}^{H}(\underset{\sim}{A}\underset{\sim}{A}^{H})^{-1}$

- The Least-Squares problem : Minimum Norm Solution

$$\hat{\underline{w}} = \underset{\sim}{A}^{+}\underline{d}$$

where $$\underset{\sim}{A}^{+} = \underset{\sim}{V}\begin{bmatrix} \underset{\sim}{\Sigma}^{-1} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix} U^{H}$$

the solution is unique in that the shortest length possible in the Euclidean sense even when null ( A ) ≠ 0 ( or rank-deficient case ).

Review :

$$\underline{\hat{w}} = (\underset{\sim}{A}^H \underset{\sim}{A})^{-1} \underset{\sim}{A}^H \underline{d}$$

$$\varepsilon_{\min} = \underline{d}^H \underline{d} - \underline{d}^H \underset{\sim}{A} (\underset{\sim}{A}^H \underset{\sim}{A})^{-1} \underset{\sim}{A}^H \underline{d}$$

Start From :

$$\varepsilon_{\min} = \underline{d}^H \underline{d} - \underline{d}^H \underset{\sim}{A} \underline{\hat{w}}$$

$$= \underline{d}^H (\underline{d} - \underset{\sim}{A} \underline{\hat{w}})$$

$$= \underline{d}^H \underset{\sim}{U} \underset{\sim}{U}^H (\underline{d} - \underset{\sim}{A} \underset{\sim}{V} \underset{\sim}{V}^H \underline{\hat{w}})$$

Now Define :
$$\underset{\sim}{V}^H \underline{\hat{w}} = \underline{z} = \begin{bmatrix} \underline{z}_1 \\ \underline{z}_2 \end{bmatrix}$$

$$\underset{\sim}{U}^H \underline{d} = \underline{c} = \begin{bmatrix} \underline{c}_1 \\ \underline{c}_2 \end{bmatrix}$$

$\underline{z}_1$ and $\underline{c}_1$ are W by 1 vectors.

Now :

$$\varepsilon_{\min} = \underline{d}^H \underset{\sim}{U}(\underset{\sim}{U}^H \underline{d} - \underset{\sim}{U}^H \underset{\sim}{A}\underset{\sim}{V}\underset{\sim}{V}^H \hat{\underline{w}})$$

$$= \underline{d}^H \underset{\sim}{U}\left(\begin{bmatrix} \underline{c}_1 \\ \underline{c}_2 \end{bmatrix} - \begin{bmatrix} \underset{\sim}{\Sigma} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}\begin{bmatrix} \underline{z}_1 \\ \underline{z}_2 \end{bmatrix}\right)$$

$$= \underline{d}^H \underset{\sim}{U}\begin{pmatrix} \underline{c}_1 - \underset{\sim}{\Sigma}\underline{z}_1 \\ \underline{c}_2 \end{pmatrix}$$

$\varepsilon_{\min}$ is independent of $\underline{z}_2$ and can be arbitrary.

For $\varepsilon_{\min}$ to be minimum, set

$$\underline{c}_1 = \underset{\sim}{\Sigma}\underline{z}_1$$

*or*

$$\underline{z}_1 = \underset{\sim}{\Sigma}^{-1}\underline{c}_1$$

If we set $\underline{z}_2 = 0$,

$$\hat{\underline{w}} = \underset{\sim}{V}\,\underline{z} = \underset{\sim}{V}\begin{bmatrix} \underset{\sim}{\Sigma}^{-1}\underline{c}_1 \\ \underline{0} \end{bmatrix}$$

*or*

$$\hat{\underline{w}} = \underset{\sim}{V}\begin{bmatrix} \underset{\sim}{\Sigma}^{-1} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}\begin{bmatrix} \underline{c}_1 \\ \underline{c}_2 \end{bmatrix}$$

$$= \underset{\sim}{V}\begin{bmatrix} \underset{\sim}{\Sigma}^{-1} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}\underset{\sim}{U}^H \underline{d}$$

$$= \underset{\sim}{A}^+ \underline{d}$$

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 $\underset{\sim}{V}\,\underset{\sim}{V}^H = \underset{\sim}{I}$

$$\left\|\hat{\underline{w}}\right\|^2 = \left\|\underset{\sim}{\Sigma}^{-1}\underline{c}_1\right\|^2$$

Consider a second solution :

$$\underline{w}' = \underset{\sim}{V}\begin{bmatrix} \underset{\sim}{\Sigma}^{-1}\underline{c}_1 \\ \underline{z}_2 \end{bmatrix} \qquad \underline{z}_2 \neq 0$$

then $\|\underline{w}'\|^2 = \|\underset{\sim}{\Sigma}^{-1}\underline{c}_1\|^2 + \|\underline{z}_2\|^2$

For any $\underline{z}_2 \neq 0$, Follows :

$$\|\underline{\hat{w}}\| < \|\underline{w}'\|$$

$\hat{w}$ 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

$$A^{+} = \underset{\sim}{V} \begin{bmatrix} \underset{\sim}{\Sigma}^{-1} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix} \underset{\sim}{U}^{H} \qquad (a)$$

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

$\text{W} = \text{rank of } \underset{\sim}{A}$

Over determined case : K > M

Form $\underbrace{\underset{\sim}{A}^{H} \underset{\sim}{A}}$ M x M Matrix

↓

Hermitian and non-negative definite
eigen values are real and non negative numbers

Denote eigenvalues of $\underset{\sim}{A}^H \underset{\sim}{A}$ :

$$\sigma_1^2, \sigma_2^2, \ldots \ldots \sigma_M^2 \text{ where}$$

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

and $$\sigma_{W+1} = \sigma_{W+2} = \ldots \ldots = \sigma_M = 0$$

$\underset{\sim}{A}^H \underset{\sim}{A}$ has the same rank as $\underset{\sim}{A}$

The eigen value -eigen vector decomposition of the matrix $\underset{\sim}{A}^H \underset{\sim}{A}$:

$$\underset{\sim}{V}^H \underset{\sim}{A}^H \underset{\sim}{A} \underset{\sim}{V} = \begin{bmatrix} \underset{\sim}{\Sigma}^2 & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}$$

Now partition the Unitary matrix $\underset{\sim}{V}$ :

$$\underset{\sim}{V} = [\underset{\sim}{V}_1, \underset{\sim}{V}_2]$$

$$\underset{\sim}{V}_1 = [v_1, v_2, \ldots \ldots v_W] \quad \text{M by W matrix}$$

$$\underset{\sim}{V}_2 = [v_{W+1}, v_{W+2}, \ldots \ldots v_M] \text{ M by (M} - \text{W) matrix}$$

Note there are W non zero eigen values :

$$\underset{\sim}{V}_1^H \underset{\sim}{V}_2 = 0$$

Follows that :

$$\underset{\sim}{V}_1^H \underset{\sim}{A}^H \underset{\sim}{A}\underset{\sim}{V}_1 = \underset{\sim}{\Sigma}^2 \qquad \text{( b )}$$

or $$\underset{\sim}{\Sigma}^{-1} \underset{\sim}{V}_1^H \underset{\sim}{A}^H \underset{\sim}{A}\underset{\sim}{V}_1 \underset{\sim}{\Sigma}^{-1} = I \qquad (c)$$

$$\underset{\sim}{V}_2^H \underset{\sim}{A}^H \underset{\sim}{A}\underset{\sim}{V}_2 = \underline{0} \qquad (d)$$

$$\underset{\sim}{A}\underset{\sim}{V}_2 = \underline{0}$$

Define K by W matrix

$$\underset{\sim}{U}_1 = \underset{\sim}{A}\underset{\sim}{V}_1 \underset{\sim}{\Sigma}^{-1} \qquad (e)$$

From ( c )

$$\underset{\sim}{U}_1^H \underset{\sim}{U}_1 = \underset{\sim}{I} \qquad (f)$$

Define $\underset{\sim}{U}_2$ K by ( K - W ) matrix

$$\underset{\sim}{U} = \left[\underset{\sim}{U}_1 \quad \underset{\sim}{U}_2\right] \text{ is a unitary matrix}$$

$$\underset{\sim}{U}_1^H \underset{\sim}{U}_2 = \underline{0}$$

Now : $\underline{\hat{w}} = \underset{\sim}{A}^+ \underline{d}$

Use ( a) $\underline{\hat{w}} = \underset{\sim}{V} \begin{bmatrix} \underset{\approx}{\Sigma}^{-1} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix} \underset{\sim}{U}^H \underline{d}$

Using $\underset{\sim}{V} = \left[\underset{\sim}{V}_1 \quad \underset{\sim}{V}_2\right]$ and (e)

$$\underline{\hat{w}} = (\underset{\sim}{V}_1 \underset{\approx}{\Sigma}^{-1})(\underset{\sim}{A} \underset{\sim}{V}_1 \underset{\approx}{\Sigma}^{-1})^H \underline{d}$$

$$= \underset{\sim}{V}_1 \underset{\approx}{\Sigma}^{-1} \underset{\approx}{\Sigma}^{-1} \underset{\sim}{V}_1^H \underset{\sim}{A}^H \underline{d}$$

$$= \underset{\sim}{V}_1 \underset{\approx}{\Sigma}^{-2} \underset{\sim}{V}_1 \underset{\sim}{A}^H \underline{d}$$

$$\boxed{\underline{\hat{w}} = \sum_{i=1}^{W} \frac{\underline{v}_i}{\sigma_i^2} \underline{v}_i^H \underset{\sim}{A}^H \underline{d}} \qquad \text{( I )}$$

Underdetermined case :

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

we can get similar equation :

$$\underline{\hat{w}} = \sum_{i=1}^{W} \frac{\underline{u}_i^H \underline{d}}{\sigma_i^2} \underset{\sim}{A}^H u_i \qquad \text{(II)}$$

where $\underset{\sim}{U}_1 = [\underline{u}_1, \underline{u}_2, \ldots\ldots \underline{u}_W]$

$\underset{\sim}{U}_2 = [\underline{u}_{W+1}, \underline{u}_{W+2}, \ldots\ldots \underline{u}_K]$

$\underset{\sim}{U} = [\underset{\sim}{U}_1, \underset{\sim}{U}_2]$ Unitary Matrix

$$\underset{\sim}{U}_1^H \underset{\sim}{U}_2 = \underline{0}$$

$$\underset{\sim}{U}^H \underset{\sim}{A} \underset{\sim}{A}^H \underset{\sim}{U} = \begin{bmatrix} \underset{\sim}{\Sigma}^2 & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}$$

$\underset{\sim}{A} \underset{\sim}{A}^H$ is now K by K matrix

Computation of ( I ) and ( II )

Step 1. Compute the SVD of the data matrix $\underset{\sim}{A}$, that is find the singular values $\sigma_1, \sigma_2, \ldots \sigma_W$ and associated right-singular vectors $\underline{v}_1, \underline{v}_2, \ldots \underline{v}_W$ and the left-singular vectors $\underline{u}_1, u_2, \ldots \underline{u}_W$.

Step 2. Compute $\hat{\underline{w}}$ 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 $\hat{\underline{w}}$.