Course Objectives & Structure

The course provides an introduction to the mathematics of data analysis and a detailed overview of statistical models for inference and prediction.

Course Structure:
- Weekly lectures [notes: www.ece.udel.edu/~arce/Courses]
- Homework & computer assignments [30%]
- Midterm & Final examinations [70%]

Textbooks:
- Papoulis and Pillai, Probability, random variables, and stochastic processes.
- Hastie, Tibshirani and Friedman, The elements of statistical learning.
- Haykin, Adaptive Filter Theory.
The singular-value decomposition (SVD) of a matrix is one of the most elegant algorithms in numerical algebra for providing quantitative information about the structure of a system of linear equations. Given a data matrix $A \in \mathbb{R}^{K \times M}$, there are two unitary matrices $U$ and $V$, such that

$$U^H AV = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \quad (1)$$

where

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_W) \quad (2)$$

is a diagonal matrix, where $W$ is the rank of $A$. The $\sigma$s are ordered as $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_W \geq 0$. 

Gonzalo R. Arce (ECE, Univ. of Delaware)
SVD Theorem

Singular-Value Decomposition Theorem

The equation

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

is a mathematical statement of the *singular-value decomposition theorem*.
Consider the system of linear equations described by

$$A \hat{w} = d$$  \hspace{1cm} (3)

where, \(d \in \mathbb{R}^{K \times 1}\).
\(\hat{w} \in \mathbb{R}^{M \times 1}\) is a vector representing an estimate of the unknown parameter vector.
If \(K > M\), we have a *overdetermined system* (more equations than unknowns).
If \(K < M\), we have a *underdetermined system*.
We want to prove the SVD in both cases.
Form a Hermitian and nonnegative definite matrix \( \mathbf{A}^H \mathbf{A} \in \mathbb{R}^{M \times M} \), with eigenvalues \( \sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2 \), where \( \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_W \geq 0 \) and \( \sigma_{W+1} = \sigma_{W+2} = \ldots = \sigma_M = 0 \).

Let \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M \) denote a set of orthonormal eigenvectors of \( \mathbf{A}^H \mathbf{A} \) associated with the eigenvalues \( \sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2 \). Let \( \mathbf{V} \in \mathbb{R}^{M \times M} \) denote the unitary matrix whose columns are these eigenvectors.

The eigendecomposition of the matrix \( \mathbf{A}^H \mathbf{A} \):

\[
\mathbf{V}^H \mathbf{A}^H \mathbf{A} \mathbf{V} = \begin{bmatrix}
\Sigma^2 & 0 \\
0 & 0
\end{bmatrix}
\]  

(4)
Let the unitary matrix \( V \) be partitioned as

\[
V = [V_1, V_2],
\]

(5)

where

\[
V_1 = [v_1, v_2, \ldots, v_W]
\]

(6)
is an \( M \) by \( W \) matrix and

\[
V_2 = [v_{W+1}, v_{W+2}, \ldots, v_M]
\]

(7)
is an \( M \) by \( M - W \) matrix with

\[
V_1^H V_2 = 0
\]

(8)
We now have

\[ V^H A^H A V = \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad V = [V_1, V_2] \]

We may take the following two deductions:

- For matrix \( V_1 \), we have

  \[ V_1^H A^H A V_1 = \Sigma^2 \]

  Consequently,

  \[ \Sigma^{-1} V_1^H A^H A V_1 \Sigma^{-1} = I. \]  \( (9) \)

- For matrix \( V_2 \), we have

  \[ V_2^H A^H A V_2 = 0 \]

  Consequently,

  \[ AV_2 = 0. \]  \( (10) \)
Overdetermined System

Notice: \( \mathbf{A} \in \mathbb{R}^{K \times M}; \mathbf{V}_1 \in \mathbb{R}^{M \times W}; \mathbf{\Sigma}^{-1} \in \mathbb{R}^{W \times W}. \)

\[ \mathbf{\Sigma}^{-1} \mathbf{V}_1^H \mathbf{A}^H \mathbf{A} \mathbf{V}_1 \mathbf{\Sigma}^{-1} = \mathbf{I}. \]

We now define a new \( K \) by \( W \) matrix

\[ \mathbf{U}_1 = \mathbf{A} \mathbf{V}_1 \mathbf{\Sigma}^{-1}. \] (13)

Then, it follows that

\[ \mathbf{U}_1^H \mathbf{U}_1 = \mathbf{I} \] (14)

which means that the columns of the matrix \( \mathbf{U}_1 \) are orthonormal.

We choose another \( \mathbf{U}_2 \in \mathbb{R}^{K \times (K-W)} \) such that

\[ \mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2], \] (15)

is a unitary matrix, which means

\[ \mathbf{U}_1^H \mathbf{U}_2 = 0 \] (16)
Overdetermined System

Remember we already have:

\[ AV_2 = 0, \]
\[ V_1^H A^H AV_1 = \Sigma^2. \]

We can write:

\[
U^H AV = \begin{bmatrix}
U_1^H \\
U_2^H
\end{bmatrix} A \begin{bmatrix}
V_1 & V_2
\end{bmatrix}
\]
\[
= \begin{bmatrix}
U_1^H AV_1 & U_1^H AV_2 \\
U_2^H AV_1 & U_2^H AV_2
\end{bmatrix}
\]
\[
= \begin{bmatrix}
(\Sigma^{-1} V_1^H A^H ) AV_1 & U_1^H 0 \\
U_2^H (U_1 \Sigma) & U_2^H 0
\end{bmatrix}
\]
\[
= \begin{bmatrix}
\Sigma & 0 \\
0 & 0
\end{bmatrix}
\]

which proves Eq.(1) for the overdetermined case.
Underdetermined System

Form a Hermitian and nonnegative definite matrix $\mathbf{A} \mathbf{A}^H \in \mathbb{R}^{K \times K}$, with eigenvalues $\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2$, where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_W \geq 0$ and $\sigma_{W+1} = \sigma_{W+2} = \ldots = \sigma_K = 0$.

Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K$ denote a set of orthonormal eigenvectors of $\mathbf{A} \mathbf{A}^H$ associated with the eigenvalues $\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2$. Let $\mathbf{U} \in \mathbb{R}^{K \times K}$ denote the unitary matrix whose columns are these eigenvectors.

The eigendecomposition of the matrix $\mathbf{A} \mathbf{A}^H$:

$$\mathbf{U}^H \mathbf{A} \mathbf{A}^H \mathbf{U} = \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$ (17)
Let the unitary matrix $\mathbf{U}$ be partitioned as

$$
\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2], \quad (18)
$$

where

$$
\mathbf{U}_1 = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_W], \quad (19)
$$

$$
\mathbf{U}_2 = [\mathbf{u}_{W+1}, \mathbf{u}_{W+2}, \ldots, \mathbf{u}_K] \quad (20)
$$

and

$$
\mathbf{U}_1^H \mathbf{U}_2 = \mathbf{0} \quad (21)
$$
We now have

\[ U^H A A^H U = \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad U_1^H U_2 = 0 \]

We may take the following two deductions:

- For matrix \( U_1 \), we have
  \[ U_1^H A A^H U_1 = \Sigma^2 \]  
  Consequently,
  \[ \Sigma^{-1} U_1^H A A^H U_1 \Sigma^{-1} = I. \]  

- For matrix \( U_2 \), we have
  \[ U_2^H A A^H U_2 = 0 \]  
  Consequently,
  \[ A^H U_2 = 0. \]
**SVD Theorem**

**Underdetermined System**

Notice: \( \mathbf{A}^H \in \mathbb{R}^{M \times K}; \mathbf{U}_1 \in \mathbb{R}^{K \times W}; \Sigma^{-1} \in \mathbb{R}^{W \times W}. \)

\[ \Sigma^{-1} \mathbf{U}_1^H \mathbf{A} \mathbf{A}^H \mathbf{U}_1 \Sigma^{-1} = \mathbf{I}. \]

We now define a \( M \) by \( W \) matrix

\[ \mathbf{V}_1 = \mathbf{A}^H \mathbf{U}_1 \Sigma^{-1}. \tag{26} \]

Then, it follows that

\[ \mathbf{V}_1^H \mathbf{V}_1 = \mathbf{I} \tag{27} \]

which means that the columns of the matrix \( \mathbf{V}_1 \) are orthonormal.

We choose another matrix \( \mathbf{V}_2 \in \mathbb{R}^{M \times (M-W)} \) such that

\[ \mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2], \tag{28} \]

is a unitary matrix, which means

\[ \mathbf{V}_2^H \mathbf{V}_1 = \mathbf{0} \tag{29} \]
Underdetermined System

Remember we already have

\[ \mathbf{A}^H \mathbf{U}_2 = \mathbf{0}, \]
\[ \mathbf{V}_1 = \mathbf{A}^H \mathbf{U}_1 \Sigma^{-1}. \]

We can write

\[
\mathbf{U}^H \mathbf{A} \mathbf{V} = \begin{bmatrix}
\mathbf{U}^H_1 \\
\mathbf{U}^H_2
\end{bmatrix} \mathbf{A} \begin{bmatrix}
\mathbf{V}_1 & \mathbf{V}_2
\end{bmatrix}
\]
\[
= \begin{bmatrix}
\mathbf{U}^H_1 \mathbf{A} \mathbf{V}_1 & \mathbf{U}^H_1 \mathbf{A} \mathbf{V}_2 \\
\mathbf{U}^H_2 \mathbf{A} \mathbf{V}_1 & \mathbf{U}^H_2 \mathbf{A} \mathbf{V}_2
\end{bmatrix}
\]
\[
= \begin{bmatrix}
\mathbf{U}^H_1 \mathbf{A} (\mathbf{A}^H \mathbf{U}_1 \Sigma^{-1}) & (\Sigma \mathbf{V}^H_1) \mathbf{V}_2 \\
0 \mathbf{V}_1 & 0 \mathbf{V}_2
\end{bmatrix}
\]
\[
= \begin{bmatrix}
\Sigma & 0 \\
0 & 0
\end{bmatrix}
\]

which proves Eq.(1) for the underdetermined case.

And the proof of the SVD theorem is completed.
Terminology and Relation to Eigenanalysis

- **Singular values** of $\mathbf{A}$: $\sigma_1, \sigma_2, \ldots, \sigma_W$, which are equal to the square root of the eigenvalues of $\mathbf{A}^H \mathbf{A}$ or $\mathbf{A} \mathbf{A}^H$.

- **Right singular vector** of $\mathbf{A}$: the columns of $\mathbf{V}$, (i.e., $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M$), which are the eigenvectors of $\mathbf{A}^H \mathbf{A}$.

- **Left singular vector** of $\mathbf{A}$: the columns of $\mathbf{U}$, (i.e., $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K$), which are the eigenvectors of $\mathbf{A} \mathbf{A}^H$. 
Since $\mathbf{U}\mathbf{U}^H = \mathbf{I} \in \mathbb{R}^{K \times K}$, and $\mathbf{V}\mathbf{V}^H = \mathbf{I} \in \mathbb{R}^{M \times M}$,

$$A = \mathbf{U}\mathbf{U}^H \mathbf{A} \mathbf{V}\mathbf{V}^H = \mathbf{U} \begin{bmatrix} \sum & 0 \\ 0 & 0 \end{bmatrix} \mathbf{V}^H.$$  \hspace{1cm} (30)

Therefore, we can write $\mathbf{A}$ in the expanded form

$$A = \sum_{i=1}^{W} \sigma_i \mathbf{u}_i \mathbf{v}_i^H.$$  \hspace{1cm} (31)
Define the pseudoinverse of the data matrix $A$ (Stewart, 1973; Golub and Van Loan, 1996) as

$$A^+ = V \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^H$$  \hspace{1cm} (32)

where

$$\Sigma^{-1} = \text{diag}(\sigma_1^{-1}, \sigma_2^{-1}, ..., \sigma_W^{-1})$$  \hspace{1cm} (33)

and $W$ is the rank of $A$.

The pseudoinverse $A^+$ may be expressed in the expanded form

$$A^+ = \sum_{i=1}^{W} \frac{1}{\sigma_i} v_i u_i^H.$$  \hspace{1cm} (34)
Pseudoinverse in Overdetermined System

In this case, $K > M$, $(A^H A)^{-1}$ exists. The pseudoinverse of $A$ is defined by

$$A^+ = (A^H A)^{-1} A^H.$$  \hspace{1cm} (35)

Remember we have

$$V_1^H A^H A V_1 = \Sigma^2$$

thus,

$$(A^H A)^{-1} = V_1 \Sigma^{-2} V_1^H$$ \hspace{1cm} (36)

Remember we have

$$U_1 = A V_1 \Sigma^{-1}$$

thus

$$A^H = V_1 \Sigma U_1^H$$ \hspace{1cm} (37)
We now have:

\[(A^H A)^{-1} = V_1 \Sigma^{-2} V_1^H\]

\[A^H = V_1 \Sigma U_1^H\]

Therefore, we may express the right hand side of Eq.(35)

\[(A^H A)^{-1} A^H = (V_1 \Sigma^{-2} V_1^H)(V_1 \Sigma U_1^H)\]

\[= V_1 \Sigma^{-1} U_1^H\]

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

\[= A^+.

Notice that in this situation, \(A^+ A = I\).
Pseudoinverse in Underdetermined System

In this case, $K < M$, $(\mathbf{AA}^H)^{-1}$ exists. The pseudoinverse of $\mathbf{A}$ is defined by

$$\mathbf{A}^+ = \mathbf{A}^H (\mathbf{AA}^H)^{-1}. \quad (38)$$

Remember we have

$$\mathbf{U}_1^H \mathbf{A} \mathbf{A}^H \mathbf{U}_1 = \Sigma^2$$

thus

$$(\mathbf{A} \mathbf{A}^H)^{-1} = \mathbf{U}_1 \Sigma^{-2} \mathbf{U}_1^H \quad (39)$$

Remember we have

$$\mathbf{V}_1 = \mathbf{A}^H \mathbf{U}_1 \Sigma^{-1}.$$ 

thus

$$\mathbf{A}^H = \mathbf{V}_1 \Sigma \mathbf{U}_1^H \quad (40)$$
Pseudoinverse in Underdetermined System

We now have

\[(\mathbf{AA}^H)^{-1} = \mathbf{U}_1 \mathbf{\Sigma}^{-2} \mathbf{U}^H_1\]

\[\mathbf{A}^H = \mathbf{V}_1 \mathbf{\Sigma} \mathbf{U}^H_1\]

Therefore, we may express the right hand side of Eq.(38)

\[\mathbf{A}^H (\mathbf{AA}^H)^{-1} = (\mathbf{V}_1 \mathbf{\Sigma} \mathbf{U}^H_1)(\mathbf{U}_1 \mathbf{\Sigma}^{-2} \mathbf{U}^H_1)\]

\[= \mathbf{V}_1 \mathbf{\Sigma}^{-1} \mathbf{U}^H_1\]

\[= \mathbf{V} \begin{bmatrix} \mathbf{\Sigma}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{U}^H\]

\[= \mathbf{A}^+.\]

Notice that in this situation, \(\mathbf{A}\mathbf{A}^+ = \mathbf{I}\).
Pseudoinverse

We have $A \in \mathbb{R}^{K \times M}$, $A^+ \in \mathbb{R}^{M \times K}$, $V = [V_1, V_2]$, $V_1 \in \mathbb{R}^{M \times W}$, $V_2 \in \mathbb{R}^{M \times (M-W)}$, $W$ is the rank of $A$.

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

So we have

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

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

$$= V \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} V^H$$

$$= \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^H \\ V_2^H \end{bmatrix}$$

$$= V_1 V_1^H = V_1 V_1^H$$
Pseudoinverse

We have $A \in \mathbb{R}^{K \times M}$, $A^+ \in \mathbb{R}^{M \times K}$, $U = [U_1, U_2]$, $U_1 \in \mathbb{R}^{K \times W}$, $U_2 \in \mathbb{R}^{K \times (K-W)}$, $W$ is the rank of $A$

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

So we have

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

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

$$= U \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} U^H$$

$$= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^H \\ U_2^H \end{bmatrix}$$

$$= U_1 U_1^H = U_1 U_1^H$$

Gonzalo R. Arce (ECE, Univ. of Delaware)
Consider $\mathbf{A} \in \mathbb{R}^{K \times M}$ is overdetermined. We define $\mathbf{y} \in \mathbb{R}^{K \times 1}$ and $\mathbf{x} \in \mathbb{R}^{M \times 1}$:

$$\mathbf{y} = \mathbf{A}\mathbf{x}.$$  
(41)

$\mathbf{x}$ is constrained to have a Euclidean norm of unity

$$||\mathbf{x}|| = 1.$$  
(42)

Solving Eq.(41) for $\mathbf{x}$
Interpretation of Singular Values and Singular Vectors

\[ x = A^+ y \]
\[ = \sum_{i=1}^{W} \frac{1}{\sigma_i} v_i u_i^H y \]
\[ = \sum_{k=1}^{W} \frac{u_i^H y}{\sigma_i} v_i \]

Imposing the constraint of \( \|x\| = 1 \), remember \( v_1, v_2, \ldots, v_W \) are orthonormal, so we have

\[ \sum_{i=1}^{W} \frac{|y^H u_i|^2}{\sigma_i^2} = 1 \]

(43)

Indeed, this is the equation of a hyperellipsoid.
Interpretation of Singular Values and Singular Vectors

\[
\sum_{i=1}^{W} \frac{|y^H u_i|^2}{\sigma_i^2} = 1
\]

To see this in a better way, we define the complex scalar

\[
\zeta_i = y^H u_i = \sum_{k=1}^{K} y_k^* u_{ik}
\] (44)

\(\zeta_i\) is referred to as the span of \(u_i\), and

\[
\sum_{i=1}^{W} \frac{|\zeta_i|^2}{\sigma_i^2} = 1
\] (45)

This is the equation of hyperellipsoid with coordinates \(|\zeta_1|, \ldots, |\zeta_W|\) and with semiaxes whose lengths are singular values \(\sigma_1, \ldots, \sigma_W\), respectively.
When $W = 2$ and $\sigma_1 > \sigma_2$, assuming the data matrix $A$ is real, we can draw:
Solution to Least Square Problem

Remember, the solution to the least square problem:

\[
\hat{w} = (A^H A)^{-1} A^H d
\]  \hspace{1cm} (46)

where, \( A \) is the data matrix representing the time evolution of the tap input vectors.
\( d \) is desired data vector representing the time evolution of the desired response.
\( \hat{w} \) is the least square estimate of the unknown parameter vector of a multiple regression model.

Using the SVD, the result can be written as

\[
\hat{w} = A^+ d
\]  \hspace{1cm} (47)
Solution to Least Square Problem

Remember the minimum sum of error squares can be written as

\[ E = \mathbf{d}^H \mathbf{d} - \mathbf{d}^H \mathbf{A} \hat{\mathbf{w}} \]  \hspace{1cm} (48)

And we know \( \mathbf{V} \mathbf{V}^H = \mathbf{I} \) and \( \mathbf{U} \mathbf{U}^H = \mathbf{I} \)

\[ E = \mathbf{d}^H \mathbf{d} - \mathbf{d}^H \mathbf{A} \hat{\mathbf{w}} \\
= \mathbf{d}^H (\mathbf{d} - \mathbf{A} \hat{\mathbf{w}}) \\
= \mathbf{d}^H \mathbf{U} \mathbf{U}^H (\mathbf{d} - \mathbf{A} \mathbf{V} \mathbf{V}^H \hat{\mathbf{w}}) \\
= \mathbf{d}^H \mathbf{U} (\mathbf{U}^H \mathbf{d} - \mathbf{U}^H \mathbf{A} \mathbf{V} \mathbf{V}^H \hat{\mathbf{w}}).

Remember that

\[ \mathbf{U}^H \mathbf{A} \mathbf{V} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \]
Solution to Least Square Problem

Let

$$V^H \hat{w} = b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$  \hfill (49)$$

and

$$U^H d = c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$  \hfill (50)$$

where $c_1, b_1 \in \mathbb{R}^{W \times 1}$

$$
\mathcal{E} = d^H U (U^H d - U^H AVV^H \hat{w}) \\
= d^H U \left( \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right) \\
= d^H U \begin{bmatrix} c_1 - \Sigma b_1 \\ c_2 \end{bmatrix}
$$
\[ E = d^H U \left[ c_1 - \Sigma b_1 \right] \]

For \( E \) to be minimum, we require

\[ c_1 = \Sigma b_1 \]

or equivalently,

\[ b_1 = \Sigma^{-1} c_1 \] (51)

Remember

\[ V^H \hat{w} = b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \]

Notice, \( E \) is independent of \( b_2 \). We can let \( b_2 = 0 \) and we get

\[ \hat{w} = Vb = V \begin{bmatrix} \Sigma^{-1} c_1 \\ 0 \end{bmatrix} \] (52)
Solution to Least Square Problem

We may also express \( \hat{w} \) as

\[
\hat{w} = Vb = V \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ 0 \end{bmatrix} = V \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^H d = A^+ d
\]

\( A^+ \) does indeed solve the linear least square problem.
Moreover, $\hat{w}$ is unique and has the minimum Euclidean norm.

$$\hat{w} = Vb = V\begin{bmatrix} \Sigma^{-1}c_1 \\ 0 \end{bmatrix}$$

Since $VV^H = I$, we have

$$||\hat{w}||^2 = ||\Sigma^{-1}c_1||^2.$$  \hfill (53)

Consider another solution

$$w' = V\begin{bmatrix} \Sigma^{-1}c_1 \\ b_2 \end{bmatrix}, b_2 \neq 0.$$ \hfill (54)

The squared Euclidean norm

$$||w'||^2 = ||\Sigma^{-1}c_1||^2 + ||b_2||^2.$$ \hfill (55)

Therefore,

$$||\hat{w}|| < ||w'||$$ \hfill (56)
Another Formulation of the minimum-Norm Solution

Case 1: Overdetermined

Remember that
\[ U_1 = AV_1 \Sigma^{-1} \text{ and } \hat{w} = V_1 \Sigma^{-1} c_1 \]

We may write
\[ \hat{w} = V_1 \Sigma^{-1} U_1^H d \]
\[ = (V_1 \Sigma^{-1})(AV_1 \Sigma^{-1})^H d \]
\[ = V_1 \Sigma^{-1} \Sigma^{-1} V_1^H A^H d \]
\[ = V_1 \Sigma^{-2} V_1^H A^H d \]

Remember,
\[ V_1 = [v_1, v_2, ..., v_W] \]
\[ \hat{w} = \sum_{i=1}^{W} v_i \sigma_i^{-2} v_i^H A^H d = \sum_{i=1}^{W} \frac{(v_i^H A^H d)}{\sigma_i^2} v_i \] (57)
Case 2: Underdetermined

\[ V_1 = A^H U_1 \Sigma^{-1} \text{ and } \hat{w} = V_1 \Sigma^{-1} c_1 \]

We may write

\[
\hat{w} = V_1 \Sigma^{-1} U_1^H d \\
= (A^H U_1 \Sigma^{-1})(\Sigma^{-1} U_1^H d) \\
= A^H U_1 \Sigma^{-2} U_1^H d
\]

Remember,

\[ U_1 = [u_1, u_2, \ldots, u_W] \]

\[
\hat{w} = \sum_{i=1}^{W} A^H u_i \sigma^{-2} u_i^H d = \sum_{i=1}^{W} \frac{(u_i^H d)}{\sigma_i^2} A^H u_i
\]  \hspace{1cm} (58)

Notice that

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

contains both cases, and is more preferred for computing the least-square estimate \( \hat{w} \).
Application in Normalized LMS Solution

Remember we have the error equation

\[ \varepsilon(n) = d(n) - \hat{w}^H(n+1)u(n), \]  

(59)

where \(d(n)\) is a desired response
\(u(n)\) is a tap-input vector
\(\hat{w}(n+1)\) is the tap weight vector measured at time \(n+1\) We want to find \(\hat{w}(n+1)\) such that

\[ \delta\hat{w}(n+1) = \hat{w}(n+1) - \hat{w}(n) \]  

(60)

is minimized, subject to the constraint

\[ \varepsilon(n) = 0 \]  

(61)
We may reformulate the error as

\[ \varepsilon(n) = d(n) - \hat{w}^H(n+1)u(n) \]

\[ = d(n) - \hat{w}^H(n)u(n) - \delta \hat{w}^H(n+1)u(n). \]

Recognize the definition of the estimation error

\[ e(n) = d(n) - \hat{w}^H(n)u(n) \quad (62) \]

Hence, we have

\[ \varepsilon(n) = e(n) - \delta \hat{w}^H(n+1)u(n). \quad (63) \]

The constraint can be written as

\[ u^H(n) \delta \hat{w}(n+1) = e^*(n) \quad (64) \]
We may restate the constrained minimization problem as follows:
Find the minimum-norm solution $\delta \hat{w}(n+1)$ that satisfies the constraint

$$u^H(n)\delta \hat{w}(n+1) = e^*(n)$$

Notice: This is an underdetermined linear least-square estimation problem.

<table>
<thead>
<tr>
<th></th>
<th>$LS$ estimation</th>
<th>Normalized LMS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data matrix</strong></td>
<td>$A$</td>
<td>$u^H(n)$</td>
</tr>
<tr>
<td><strong>Desired data vector</strong></td>
<td>$d$</td>
<td>$e^*(n)$</td>
</tr>
<tr>
<td><strong>Parameter vector</strong></td>
<td>$\hat{w}$</td>
<td>$\delta \hat{w}(n+1)$</td>
</tr>
<tr>
<td><strong>Rank</strong></td>
<td>$W$</td>
<td>$1$</td>
</tr>
<tr>
<td><strong>Singular value</strong></td>
<td>$\sigma_i, i = 1, \ldots, W$</td>
<td>$</td>
</tr>
<tr>
<td><strong>Left singular vector</strong></td>
<td>$u, i = 1, \ldots, W,$</td>
<td>$1$</td>
</tr>
</tbody>
</table>
So we apply SVD,

\[ \delta \hat{w}(n+1) = \frac{1}{||u(n)||^2} u(n)e^*(n). \]  \hspace{1cm} (65)

As before, we introduce a scaling factor \( \tilde{\mu} \)

\[ \delta \hat{w}(n+1) = \frac{\tilde{\mu}}{||u(n)||^2} u(n)e^*(n). \]  \hspace{1cm} (66)

Equivalently, we may write

\[ \hat{w}(n+1) = \hat{w}(n) + \frac{\tilde{\mu}}{||u(n)||^2} u(n)e^*(n). \]  \hspace{1cm} (67)

Thus, we have the tap-weight vector update for the normalized LMS algorithm.
Example 1

Given the complex matrix

\[
A = \begin{bmatrix}
1 + j & 1 + 0.5j \\
0.5 - j & 1 - j
\end{bmatrix}
\]

We can construct:

\[
A^H A = \begin{bmatrix}
3.25 & 3 \\
3 & 3.25
\end{bmatrix}
\]

and

\[
AA^H = \begin{bmatrix}
3.25 & 3j \\
-3j & 3.25
\end{bmatrix}
\]

Use

\[
|D - \lambda I| = 0
\]

to calculate the eigenvalues of \(A^H A\) and \(AA^H\)
We get
\[ \lambda_1 = 6.25, \lambda_2 = 0.25; \]

And the singular values of \( A \) are
\[ \sigma_1 = 2.5, \sigma_2 = 0.5 \]

Observe the form of \( A^H A \) and \( AA^H \), notice we have
\[ Dq_i = \lambda_i q_i \]

Thus we can simply write the form of \( V \) and \( U \):
\[
V = \begin{bmatrix} x & y \\ x & -y \end{bmatrix}, \\
U = \begin{bmatrix} jx & y \\ x & -jy \end{bmatrix},
\]

Solve for \( V \): \( x = y \) and \( x^2 + y^2 = 1 \)
Solve for \( U \): \( x^2 = y^2 \) and \( x^2 + y^2 = 1 \)
Then, take arbitrary values of $x$ and $y$ when solving $U$

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, U = \frac{1}{\sqrt{2}} \begin{bmatrix} j & 1 \\ 1 & j \end{bmatrix},$$

The SVD of $A$ is given by

$$B = U \Sigma V^H = \begin{bmatrix} 1.25 + j0.25 & -1.25 + j0.25 \\ 0.25 + j1.25 & 0.25 - j1.25 \end{bmatrix}$$

Notice that $B \neq A$, but $B^H B = A^H A$ and $BB^H = AA^H$.

If we want $B = A$, the arbitrariness of choosing eigenvectors of $A^H A$ and $AA^H$ should be removed.
We have
\[ V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \]

Then define \( U \) as
\[ U = AV\Sigma^{-1} = \begin{bmatrix} 0.5657 + j0.4243 & j0.7071 \\ 0.4243 - j0.5657 & -0.7071 \end{bmatrix} \]

We can easily verify that the new \( U \) is indeed composed of eigenvectors of \( AA^H \).

Clearly, the new \( U \) and the old \( U \) are similar:
\[ U_{\text{new}} = W^H U_{\text{old}} W \quad (68) \]

where \( W \) is a unitary matrix.
Example 2

A two-tap LS filter, given data matrix

\[ A = \begin{bmatrix} u_2 & u_1 \\ u_3 & u_2 \\ u_4 & u_3 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 2 \\ -1 & 1 \end{bmatrix} \]

and desired data vector

\[ d = \begin{bmatrix} d_2 \\ d_3 \\ d_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1/34 \end{bmatrix} \]

Compute the two tap weights using the pseudoinverse \( A^+ \).
First, let’s compute

$$A^H A = \begin{bmatrix} 6 & 7 \\ 7 & 14 \end{bmatrix}$$

The eigenvalues of $A^H A$:

$$\lambda_1 = 1.94, \lambda_2 = 18.06.$$  

Take square root:

$$\sigma_1 = 1.393, \sigma_2 = 4.25.$$  

Calculate the eigenvectors using $D q_i = \lambda_i q_i$ and form the orthonormal matrix:

$$V = \begin{bmatrix} 0.87 & -0.505 \\ -0.505 & -0.87 \end{bmatrix}$$
Then, let’s compute

\[
AA^H = \begin{bmatrix}
13 & 8 & 1 \\
8 & 5 & 1 \\
1 & 1 & 2
\end{bmatrix}
\]

The eigenvalues of \( AA^H \):

\[
\lambda_1 = 1.94, \lambda_2 = 18.06, \lambda_3 = 0
\]

Calculate the eigenvectors and form:

\[
U = \begin{bmatrix}
0.167 & -0.845 & 0.5071 \\
0.198 & -0.524 & -0.8425 \\
-0.98 & -0.085 & 0.169
\end{bmatrix}
\]
Now we can write the SVD of \( \mathbf{A} \)

\[
\mathbf{A} = \begin{bmatrix}
0.167 & -0.845 & 0.5071 \\
0.198 & -0.524 & -0.8425 \\
-0.98 & -0.085 & 0.169
\end{bmatrix}
\begin{bmatrix}
0.1393 & 0 \\
0 & 4.25 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
0.87 & -0.505 \\
-0.505 & -0.87
\end{bmatrix}
\]

As a check we can calculate

\[
\mathbf{A} = \begin{bmatrix}
2.016 & 3.0069 \\
0.9925 & 2.0142 \\
-1.0065 & 1.0044
\end{bmatrix}
\]

which is closed to the original data matrix.
The pseudoinverse $\mathbf{A}^+$ is

$$\mathbf{A}^+ = \sum_{i=1}^{2} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^T$$

The LS weights vector is

$$\hat{\mathbf{w}} = \mathbf{A}^+ \mathbf{d} = \sum_{i=1}^{2} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^T \mathbf{d} = \sum_{i=1}^{2} \frac{1}{\sigma_i} \mathbf{v}_i (\mathbf{u}_i^T \mathbf{d})$$

$$= \sum_{i=1}^{2} \mathbf{v}_i \frac{\mathbf{u}_i^T \mathbf{d}}{\sigma_i} = \sum_{i=1}^{2} \frac{\mathbf{u}_i^T \mathbf{d}}{\sigma_i} \mathbf{v}_i$$

$$= \begin{bmatrix} 0.167 & 0.109 & -0.981 \\ 1.393 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1/34 \end{bmatrix} + \begin{bmatrix} 0.845 & 0.524 & -0.085 \\ 4.250 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1/34 \end{bmatrix}$$

$$= \begin{bmatrix} 0.3859 \\ 0.3859 \end{bmatrix}$$
Notice that in Example 2, we have

\[
A = \begin{bmatrix}
  u_2 & u_1 \\
  u_3 & u_2 \\
  u_4 & u_3
\end{bmatrix} = \begin{bmatrix}
  2 & 3 \\
  1 & 2 \\
 -1 & 1
\end{bmatrix}
\]

which is an overdetermined case.

Even if we have an underdetermined case, \( \hat{w} = A^+d \) will give us a unique solution to the LS problem that

- It produces the minimal sum of error squares.
- It has the smallest Euclidean norm.

This special solution is so called the \textit{minimal norm solution}.
An $M$ by $N$ image $\mathbf{A}$ with rank $W$. Its SVD is:

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^H = \sum_{i=1}^{W} \sigma_i \mathbf{u}_i \mathbf{v}_i^H$$

where, $\sigma_i \mathbf{u}_i \mathbf{v}_i^H$ are called eigenimages

The Lenna image and its $\mathbf{U}$, $\mathbf{V}$, $\Sigma$:
Image SVD Transform

\[ A = U \Sigma V^H = \sum_{i=1}^{W} \sigma_i u_i v_i^H \]

The first 10 eigenimages of Lenna image:
Image SVD Transform

\[ A = U \Sigma V^H = \sum_{i=1}^{W} \sigma_i u_i v_i^H \]

The first 10 partial sum images:
Image SVD Transform

\[ A = U \Sigma V^H = \sum_{i=1}^{W} \sigma_i u_i v_i^H \]

Eigenimages from 10 to 100 with increment of 10:
Image SVD Transform

\[ A = U\Sigma V^H = \sum_{i=1}^{W} \sigma_i u_i v_i^H \]

The partial sum images (from 10 to 100 eigen-images with increment 10):
**Application in Image Compression**

We have matrix

\[ A = [a_{i,j}]_{N \times N} = [a_1, \ldots, a_N] \]

where \( a_i \) is the \( i \)th column vector of \( A \).

The total amount of energy contained in \( A \):

\[
\mathcal{E} = \| A \| = tr[A^T A] = \sum_{i=1}^{N} a_i^T a_i = \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}
\]

Image compression using only the first \( M \) eigenimages:

\[ A_M = \sum_{i=1}^{M} \sigma_i u_i v_i^H \]

with error

\[ E_M = A - A_M = \sum_{i=M+1}^{N} \sigma_i u_i v_i^T \]
After compression, the energy (information) contained in $A$ is:

$$
||A_M|| = tr[A_M^T A_M] = tr[\sum_{i=1}^{M} \sigma_i v_i u_i^T][\sum_{j=1}^{M} \sigma_j u_j v_j^T]
$$

$$
= tr[\sum_{i=1}^{M} \sum_{j=1}^{M} \sigma_i \sigma_j v_i u_i^T u_j v_j^T]
$$

$$
= tr[\sum_{i=1}^{M} \sigma_i^2 v_i v_i^T] = \sum_{i=1}^{M} \sigma_i^2 tr[v_i v_i^T]
$$

$$
= \sum_{i=1}^{M} \sigma_i^2 tr[v_i^T v_i] = \sum_{i=1}^{M} \sigma_i^2
$$
Application in Image Compression

The total amount of energy (information) contained in the original image $A$ is

$$\text{energy in } A = \sum_{i=1}^{N} \sigma_i^2$$  \hspace{1cm} (69)

and the energy (information) lost (contained in $E_M$) is

$$\text{energy in } E_M = \sum_{i=M+1}^{N} \sigma_i^2$$  \hspace{1cm} (70)

Minimum energy is lost if we range $\sigma_i$’s as

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N.$$
Image SVD Transform

Reconstruction with $M = 512, 256, 128$ (first row), $64, 32, 16$ (second row):
Image SVD Transform

Difference between original and reconstruction image with $M = 256, 128, 64$ (first row), $32, 16$ (second row):
Conservation of Degrees of Freedom

The degrees of freedom (the number of independent variables in the signal) are conserved in the SVD transform.

- In spatial domain the d.o.f. of the image matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ (assuming $M = N$) is $N^2$.

- In the transform domain, both $\mathbf{U}$ and $\mathbf{V}$ have the same d.o.f:
  - The first column vector has $N$ elements subject to normalization, i.e., $N – 1$ d.o.f.;
  - The second vector is the same except it also has to be orthogonal to the first one, and therefore has $N – 2$ d.o.f.;
  - The third vector has to be orthogonal to the first two vectors and therefore has $N – 3$ d.o.f.; etc.
Now the total d.o.f. of all $N$ vectors are:

$$(N - 1) + (N - 2) + \ldots + 1 = (N - 1)N/2 = (N^2 - N)/2$$

Together with the $N$ d.o.f. in $\Sigma$, the overall d.o.f. is

$$2(N - 1)N/2 + N = N^2.$$ 

This indicates that the signal, in either the original spatial domain ($A$) or the transform domain ($U, V$ and $\Sigma$), always has the same degrees of freedom.