7. Discrete Vector Set Approximation

Section 7.1 Discrete Orthogonal Vector Set Expansion

Section 7.2 Non-Orthogonal Discrete Vector Set Expansion

Section 7.3 Generalized Inverse

$M \times N$

M: number of equations

Section 7.4 Discrete Orthogonal Polynomials (只教不考) N= number of unknowns

$$\mathbf{A}\mathbf{x} \cong \mathbf{y}$$

if A is a square matrix ($N \times N$)

and A^{-1} exists ($\det(A) \neq 0$)

$x = A^{-1}y$

A and **y** are known.

Problem: How do we find **x** such that

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}\| \qquad (L_2 \text{ norm of } \mathbf{y}-\mathbf{A}\mathbf{x})$$

is minimized?

Q: How to solve x when $M \neq N$ or $\det(A) = 0$?

$A_{M \times N}$

- Discrete Vector Set Expansion
  - Orthogonal
    - $M=N$, $\det(A) \neq 0$
      Complete (Case 1) (Sec. 7-1-2)
    - $M<N$
      Incomplete (Case 2) (Sec. 7-1-3)
    - Discrete Orthogonal Transforms (Secs. 6-7, 7.1.1)
    - Discrete Orthogonal Polynomials (Sec. 7-4)
  - Non-Orthogonal
    - Independent Set (Sec. 7-2)
      - $M=N$, $\det(A) \neq 0$
        Complete (Case 3) (Secs. 7.2.1, 7.2.2)
      - $M<N$ or $\det(A)=0$
        Incomplete (Case 4) (Secs. 7.2.2, 7.2.3)
    - Dependent Set (Case 5) (Sec. 7-3)

7.1 Discrete Orthogonal Vector Set Expansion

7.1.1 Discrete Orthogonal Matrix

[Orthogonal (Column Form)]

$$\mathbf{A} = \begin{bmatrix} \phi_1[1] & \phi_2[1] & \phi_3[1] & \cdots & \phi_N[1] \\ \phi_1[2] & \phi_2[2] & \phi_3[2] & \cdots & \phi_N[2] \\ \phi_1[3] & \phi_2[3] & \phi_3[3] & \cdots & \phi_N[3] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1[M] & \phi_2[M] & \phi_3[M] & \cdots & \phi_N[M] \end{bmatrix}$$

If

Note!

$$\sum_{m=1}^{M} \phi_n[m]\phi_k^*[m] = \begin{cases} 0 & for\ n \neq k \\ d_n & for\ n = k \end{cases}$$

then

$$\mathbf{A}^{\mathbf{H}}\mathbf{A} = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_N \end{bmatrix}$$

[Orthogonal (Column Form)]

Suppose that **A** is an *M*x*N* matrix. If all the columns of **A** are orthogonal, then

$$\mathbf{A}^{\mathbf{H}}\mathbf{A} = \mathbf{D}$$

where **D** is an *N*x*N* orthogonal matrix. Moreover, if all the columns of **A** are orthonormal, then

$$(d_n = 1 \quad \textit{for all } n) \qquad \mathbf{A}^{\mathbf{H}}\mathbf{A} = \mathbf{I}$$

where **I** is an *N*x*N* identity matrix.

(Note: An orthonormal matrix is also called a unitary matrix.)

[Orthogonal (Row Form)]

$$\mathbf{A}=\begin{bmatrix}\phi_1[1] & \phi_1[2] & \phi_1[3] & \cdots & \phi_1[N]\\ \hdashline \phi_2[1] & \phi_2[2] & \phi_2[3] & \cdots & \phi_2[N]\\ \hdashline \phi_3[1] & \phi_3[2] & \phi_3[3] & \cdots & \phi_3[N]\\ \hdashline \vdots & \vdots & \vdots & \ddots & \vdots\\ \hdashline \phi_M[1] & \phi_M[2] & \phi_M[3] & \cdots & \phi_M[N]\end{bmatrix}$$

If $$\sum_{n=1}^{N}\phi_m[n]\phi_k^*[n]=\begin{cases}0 & for\ m\neq k\\ d_m & for\ m=k\end{cases}$$

then $$\mathbf{AA}^{\mathbf{H}}=\begin{bmatrix}d_1 & 0 & 0 & \cdots & 0\\ 0 & d_2 & 0 & \cdots & 0\\ 0 & 0 & d_3 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & d_M\end{bmatrix}$$

ex: $\begin{bmatrix}1 & 1 & 1\\ 1 & -2 & 1\\ 1 & 0 & -1\end{bmatrix}$

is row-form orthogonal but not column form orthogonal.

normalize: $\begin{bmatrix}1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3}\\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6}\\ 1/\sqrt{2} & 0 & -1/\sqrt{2}\end{bmatrix}$

is both row-form orthonormal and column form orthonormal.

[Orthogonal (Row Form)]

Suppose that **A** is an MxN matrix. If all the rows of **A** are orthogonal, then

$$\mathbf{A}\mathbf{A}^{\mathbf{H}} = \mathbf{D}$$

where **D** is an MxM orthogonal matrix. Moreover, if all the rows of **A** are orthonormal, then

$$\mathbf{A}\mathbf{A}^{\mathbf{H}} = \mathbf{I} \Rightarrow A^H = A^{-1} \Rightarrow \begin{matrix} A^{-1}A = I \\ A^H A = I \end{matrix}$$

where **I** is an MxM identity matrix.

(Note: If a set of vectors is orthogonal, then these vectors should be linearly independent. Therefore, if the rows of A are orthogonal, then $M \leq N$ should be satisfied.)

★ orthogonal (row form) ≠ orthogonal (column form)
orthonormal (row form) = orthonormal (column form)

[Inverse of an Orthogonal Matrix]

If **A** is a square matrix (i.e., $M = N$)

(1) If all the columns of **A** are orthogonal, $\mathbf{A}^{\mathbf{H}}\mathbf{A} = \mathbf{D}$, then

$D^{-1}A^{H}A = I$

$$\mathbf{A}^{-1} = \mathbf{D}^{-1}\mathbf{A}^{\mathbf{H}}$$

(2) If all the columns of **A** are orthonormal, $\mathbf{A}^{\mathbf{H}}\mathbf{A} = \mathbf{I}$, then

$$\mathbf{A}^{-1} = \mathbf{A}^{\mathbf{H}}$$

(3) If all the rows of **A** are orthogonal, $\mathbf{A}\mathbf{A}^{\mathbf{H}} = \mathbf{D}$, then

$$\mathbf{A}^{-1} = \mathbf{A}^{\mathbf{H}}\mathbf{D}^{-1}$$

(4) If all the rows of **A** are orthonormal, $\mathbf{A}\mathbf{A}^{\mathbf{H}} = \mathbf{I}$, then

$$\mathbf{A}^{-1} = \mathbf{A}^{\mathbf{H}}$$

[Example of Orthogonal Matrix]

- DFT
- Discrete Cosine Transform
- Walsh (Hadamard Transform)

both row-form and column-form orthogonal

- Haar Transform (row-form orthogonal)
- Discrete Orthogonal Polynomial Matrices (row-form orthogonal)

[**Example 1**]

Walsh

$$\mathbf{W}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \qquad \mathbf{W}_4^{-1} = \frac{1}{4}\mathbf{I}\mathbf{W}_4^{H} = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & -1/4 & -1/4 \\ 1/4 & -1/4 & -1/4 & 1/4 \\ 1/4 & -1/4 & 1/4 & -1/4 \end{bmatrix}$$

$$\mathbf{W}_4^{\mathbf{H}}\mathbf{W}_4 = 4\mathbf{I}$$

[Duality Property of Orthogonal Matrices]

If all the columns of a square matrix **A** are orthonormal, then all the rows of **A** are orthonormal, too.

(Proof): If

$$\mathbf{A}^{\mathbf{H}}\mathbf{A} = \mathbf{I}$$

then since $\mathbf{A}^{\mathbf{H}} = \mathbf{A}^{-1}$, we have

$$\mathbf{A}\mathbf{A}^{\mathbf{H}} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

Therefore, all the rows of **A** are orthonormal, too.

[**Example 2**] Note that, if

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & 1 & -1 \end{bmatrix}$$

then the columns of **A** are orthogonal. However, the rows of **A** are not orthogonal.

If we perform normalization for the columns **A** and obtain **B**:

$$\mathbf{B} = \left[\begin{array}{c:c:c} 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \end{array}\right]$$

then both the columns and the rows of **B** are orthonormal:

$$\mathbf{B}^H\mathbf{B} = \mathbf{I}, \qquad \mathbf{B}\mathbf{B}^H = \mathbf{I}$$

7.1.2 Discrete Orthogonal Vector Set Expansion of the Complete Case (Case 1)

Suppose that $b_1[n]$, $b_2[n]$, …. $b_N[n]$ forms a complete and orthogonal set in C^N:

$$\sum_{n=1}^{N} b_m[n] b_k^*[n] = \begin{cases} 0 & for\ m \neq k \\ d_m & for\ m = k \end{cases}$$

If we want to expand $y[n]$ by a linear combination of $b_m[n]$ (m = 1, 2, …, N):

$$y[n] = \sum_{m=1}^{N} x_m b_m[n]$$

then, analogous to page 277,

$$x_m = \frac{\sum_{n=1}^{N} y[n] b_m^*[n]}{\sum_{n=1}^{N} b_m[n] b_m^*[n]}$$

From the view point of the matrix

If

$$\mathbf{A} = \left[\begin{array}{c:c:c:c:c} b_1[1] & b_2[1] & b_3[1] & \cdots & b_N[1] \\ b_1[2] & b_2[2] & b_3[2] & \cdots & b_N[2] \\ b_1[3] & b_2[3] & b_3[3] & \cdots & b_N[3] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_1[N] & b_2[N] & b_3[N] & \cdots & b_N[N] \end{array}\right]$$

$$\mathbf{x} = [x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_N]^T \qquad \mathbf{y} = [y[1] \quad y[2] \quad y[3] \quad \cdots \quad y[N]]^T$$

then the problem can be re-expressed as

$$\mathbf{Ax} = \mathbf{y}$$

Since

$$\mathbf{A}^H\mathbf{A} = \mathbf{D} \qquad \text{where} \quad D[m,n] = \begin{cases} 0 & \text{if } m \neq n \\ \underbrace{\sum_{k=1}^{N} b_m[k] b_m^*[k]}_{:\, d_m} & \text{if } m = n \end{cases}$$

we have

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} = \mathbf{D}^{-1}\mathbf{A}^H\mathbf{y}, \qquad x_m = \frac{\sum_{n=1}^{N} y[n] b_m^*[n]}{\sum_{n=1}^{N} b_m[n] b_m^*[n]}$$

[Parseval's Theorem for Discrete Orthogonal Matrix]

analogous to page 301

If

$$\mathbf{A}\mathbf{x} = \mathbf{y}$$

square orthogonal matrix

and the columns of **A** are orthogonal, then

$$\sum_{n=1}^{N} |y[n]|^2 = \sum_{n=1}^{N} d_n |x[n]|^2 \quad \text{where} \quad d_n = \sum_{k=1}^{N} |A[k,n]|^2$$

nth column

(Proof):

$$\mathbf{y}^H \mathbf{y} = \mathbf{x}^H \mathbf{A}^H \mathbf{A}\mathbf{x} = \mathbf{x}^H \mathbf{D}\mathbf{x}$$

[Example 3]

Parseval's theorem for the DFT and the Walsh transform:

$$\sum_{n=1}^{N} |y[n]|^2 = N \sum_{n=1}^{N} |x[n]|^2$$

Parseval's theorem for the DCT

$$\sum_{n=1}^{N} |y[n]|^2 = \sum_{n=1}^{N} |x[n]|^2$$

7.1.3 Discrete Orthogonal Basis Expansion of the Incomplete Case (Case 2)

$n = 1 \sim N, \quad m = 1 \sim M$

Suppose that $b_1[n]$, $b_2[n]$, …. $b_M[n]$ forms an incomplete and orthogonal set in C^N but $M < N$:

$$\sum_{n=1}^{N} b_m[n] b_k^*[n] = \begin{cases} 0 & for\ m \neq k \\ d_m & for\ m = k \end{cases}$$

If we want to expand $y[n]$ by a linear combination of $b_m[n]$ (m = 1, 2, …, M):

$$y[n] \cong \sum_{m=1}^{M} x_m b_m[n]$$

then

$$x_m = \frac{\sum_{n=1}^{N} y[n] b_m^*[n]}{\sum_{n=1}^{N} b_m[n] b_m^*[n]}$$

Same as Case 1

The formulas are similar to those of Case 1, except for that $y[n] =$ is replaced by $y[n] \cong$

Note:

(1) Since $b_1[n]$, $b_2[n]$, …. $b_M[n]$ can be viewed as a subset of a complete and orthogonal set $\{b_1[n], b_2[n], …., b_M[n], b_{M+1}[n], …, b_N[n]\}$, the method to determine the linear combination coefficients x_m is all the same as that of the complete case.

Note:

(2) Determine x_m by $x_m = \sum_{n=1}^{N} y[n] b_m^*[n] / \sum_{n=1}^{N} b_m[n] b_m^*[n]$ can minimize

$$\left\| y[n] - \sum_{m=1}^{M} x_m b_m[n] \right\| = \sqrt{\sum_{n=1}^{N} \left(y[n] - \sum_{m=1}^{M} x_m b_m[n] \right)^2}$$

$$b_m = \begin{bmatrix} b_m[1] \\ b_m[2] \\ \vdots \\ b_m[N] \end{bmatrix}$$

$$\left\| y[n] - \sum_{m=1}^{M} x_m b_m[n] \right\|^2 = \sum_{n=1}^{N} \left| \sum_{m=M+1}^{N} x_m b_m[n] \right|^2 = \sum_{m=M+1}^{N} d_m |x_m|^2$$

$$(x_{M+1} b_{M+1} + x_{M+2} b_{M+2} + \cdots x_N b_N)^H \times (x_{M+1} b_{M+1} + x_{M+2} b_{M+2} + \cdots + x_N b_N)$$

$$= \sum_{m=M+1}^{N} |x_m|^2 d_m$$

$$= \sum_{m=1}^{N} d_m |x_m|^2 - \sum_{m=1}^{M} d_m |x_m|^2$$

(from Parseval's theorem on page 618) where $d_m = \sum_{n=1}^{N} |b_m[n]|^2$

$$\left\| y[n] - \sum_{m=1}^{M} x_m b_m[n] \right\|^2 = \sum_{n=1}^{N} |y[n]|^2 - \sum_{m=1}^{M} d_m |x_m|^2 = \|y[n]\|_2^2 - \sum_{m=1}^{M} |x_m|^2 \|b_m[n]\|_2^2$$

[**Example 4**] Suppose that

$$\mathbf{y} = [1 \quad 1 \quad 5 \quad 5 \quad 6 \quad 6 \quad 5 \quad 4 \quad 4 \quad 3 \quad 3]^T$$

→ N=11

Try to expand **y** as a linear combination of

$$\mathbf{b_1} = [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]^T$$

1

and $\mathbf{b_2} = [-5 \quad -4 \quad -3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5]^T$

(n−6)

M=2

$x_1 \cdot 1 + x_2(n-6)$

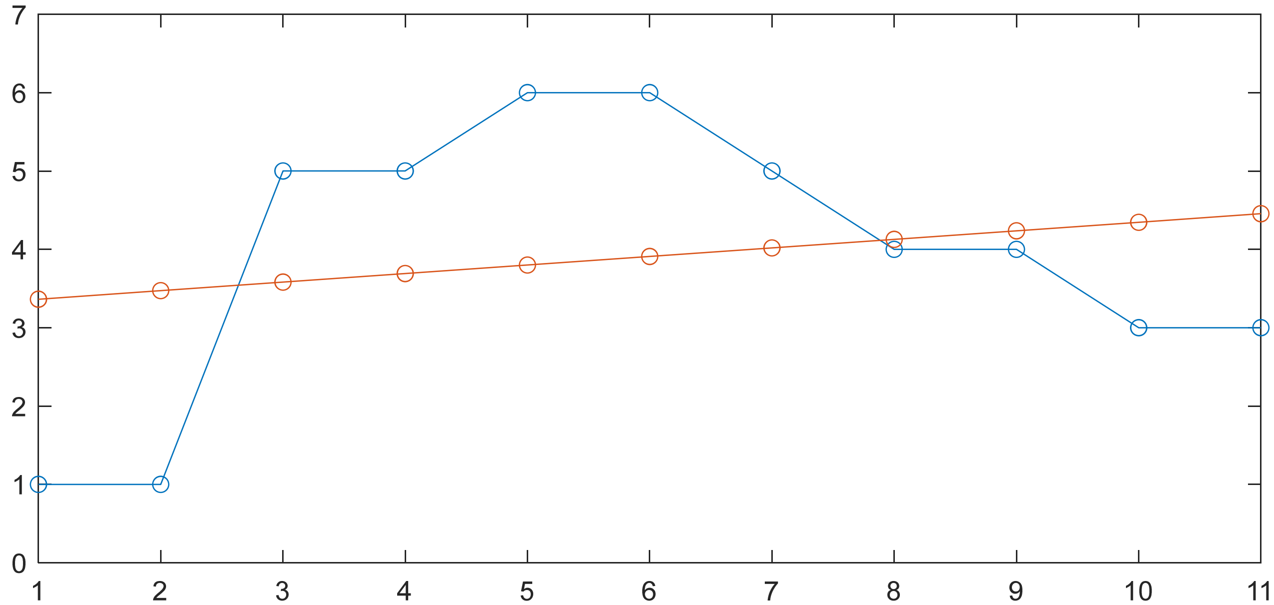
such that $\|\mathbf{y} - x_1\mathbf{b_1} - x_2\mathbf{b_2}\|$ is minimized.

incomplete

(Solution): It is obvious that $\mathbf{b_1}$ and $\mathbf{b_2}$ are orthogonal. Therefore,

$$x_1 = \frac{\sum_{n=1}^{11} y[n]b_1^*[n]}{\sum_{n=1}^{11} b_1[n]b_1^*[n]} = \frac{43}{11} \qquad x_2 = \frac{\sum_{n=1}^{11} y[n]b_2^*[n]}{\sum_{n=1}^{11} b_2[n]b_2^*[n]} = \frac{12}{110}$$

$$\mathbf{y} \cong \frac{43}{11}\mathbf{b_1} + \frac{6}{55}\mathbf{b_2}$$



Blue: $\mathbf{y}$ Red: $\frac{43}{11}\mathbf{b_1} + \frac{6}{55}\mathbf{b_2}$

$$\left\|\mathbf{y} - x_1\mathbf{b_1} - x_2\mathbf{b_2}\right\|^2 = \left\|\mathbf{y}\right\|^2 - \left|x_1\right|^2\left\|\mathbf{b_1}\right\|^2 - \left|x_2\right|^2\left\|\mathbf{b_2}\right\|^2 = 29.6$$

$$\left\|\mathbf{y} - x_1\mathbf{b_1} - x_2\mathbf{b_2}\right\| = 5.4406$$

7.2 Non-Orthogonal Discrete Basis Expansion

7.2.1 Method 1: Matrix Inverse Case 3

Suppose that $\{b_1[n], b_2[n], b_3[n], \ldots\ldots\ldots\ldots, b_N[n]\}$ are linearly independent and complete vector set in C^N but are not orthogonal. (Case 3)

To express $y[n] \in C^N$ by a linear combination of $b_1[n], b_2[n], b_3[n], \ldots\ldots\ldots\ldots, b_N[n]\}$

$$y[n] = \sum_{m=1}^{N} x_m b_m[n]$$

we first construct a matrix **A**:

$$\mathbf{A} = \begin{bmatrix} b_1[1] & b_2[1] & b_3[1] & \cdots & b_N[1] \\ b_1[2] & b_2[2] & b_3[2] & \cdots & b_N[2] \\ b_1[3] & b_2[3] & b_3[3] & \cdots & b_N[3] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_1[N] & b_2[N] & b_3[N] & \cdots & b_N[N] \end{bmatrix}$$

$\det(A) \neq 0$

A^{-1} exists

Then,

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$$

$x_m = \sum_{n=1}^{N} \phi_m^*[n]\, y[n]$

from page 626

where

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_N \end{bmatrix}^T \qquad \mathbf{y} = \begin{bmatrix} y[1] & y[2] & y[3] & \cdots & y[N] \end{bmatrix}^T$$

[Dual Orthogonal]

$\{b_1[n], b_2[n], b_3[n], \ldots\ldots\ldots., b_N[n]\}$ are dual orthogonal to $\{\phi_1[n], \phi_2[n], \phi_3[n], \ldots\ldots\ldots., \phi_N[n]\}$ if:

$$\sum_{m=1}^{N} b_m[n]\phi_k^*[n] = \begin{cases} 0 & if\ m \neq k \\ u_m & if\ m = k \end{cases}$$

In fact, they are also dual orthonormal if $u_m = 1$.

$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$

If

$$\mathbf{A} = \begin{bmatrix} b_1[1] & b_2[1] & b_3[1] & \cdots & b_N[1] \\ b_1[2] & b_2[2] & b_3[2] & \cdots & b_N[2] \\ b_1[3] & b_2[3] & b_3[3] & \cdots & b_N[3] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_1[N] & b_2[N] & b_3[N] & \cdots & b_N[N] \end{bmatrix}$$

conjugation

$$\overline{\mathbf{A}}^{-1} = \begin{bmatrix} \phi_1[1] & \phi_1[2] & \phi_1[3] & \cdots & \phi_1[N] \\ \phi_2[1] & \phi_2[2] & \phi_2[3] & \cdots & \phi_2[N] \\ \phi_3[1] & \phi_3[2] & \phi_3[3] & \cdots & \phi_3[N] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_N[1] & \phi_N[2] & \phi_N[3] & \cdots & \phi_N[N] \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} \phi_1^* \\ \phi_2^* \\ \vdots \\ \phi_N^* \end{bmatrix}$$

then $\{b_1[n], b_2[n], b_3[n], \ldots\ldots\ldots, b_N[n]\}$ are dual orthonormal to $\{\phi_1[n], \phi_2[n], \phi_3[n], \ldots\ldots\ldots, \phi_N[n]\}$:

$$\sum_{m=1}^{N} b_m[n]\phi_k^*[n] = \begin{cases} 0 & if\ m \neq k \\ 1 & if\ m = k \end{cases}$$

7.2.2 Method 2: Gram-Schmidt (Cases 3, 4)

Suppose that $\{b_1[n], b_2[n], \ldots\ldots, b_M[n]\}$ are linearly independent but not orthogonal. Then we can follow the Gram-Schmidt process to convert it into an orthogonal set $\{a_1[n], a_2[n], \ldots\ldots, a_M[n]\}$ and perform expansion.

(applicable for both complete and incomplete case)

$$a_1[n] = \frac{b_1[n]}{\|b_1[n]\|} \qquad m = 1$$

$$\Downarrow$$

$$g_m[n] = b_m[n] - \sum_{k=1}^{m-1}\left(\sum_{n=1}^{N} b_m[n]\,a_k^*[n]\right)a_k[n]$$

$$\Downarrow$$

$$a_m[n] = \frac{g_m[n]}{\|g_m[n]\|}$$

$$\Downarrow$$

$$m = m+1$$

Find $x_1, x_2, \ldots, x_M$ to minimize $\left\| \mathbf{y} - x_1\mathbf{b_1} - x_2\mathbf{b_2} - \cdots - x_M\mathbf{b_M} \right\|$

by the Gram-Schmidt method.

Step 1: Convert $\{b_1[n], b_2[n], \ldots\ldots, b_M[n]\}$ into an orthogonal set $\{a_1[n], a_2[n], \ldots\ldots, a_M[n]\}$ by the Gram-Schmidt method.

Step 2: Expand $y[n]$ by $\{a_1[n], a_2[n], \ldots\ldots, a_M[n]\}$

$$y[n] \cong \sum_{m=1}^{M} z_m a_m[n] \qquad z_m = \sum_{n=1}^{N} y[n] b_m^*[n] \qquad \text{(from page 619)}$$

Step 3: If

$$a_k[n] \cong \sum_{m=1}^{k} c_{k,m} b_m[n]$$

then

$$y[n] \cong \sum_{k=1}^{M} z_k \sum_{m=1}^{k} c_{k,m} b_m[n] = \sum_{m=1}^{M} \sum_{k=m}^{M} z_k c_{k,m} b_m[n] = \sum_{m=1}^{M} x_m b_m[n]$$

$$x_m = \sum_{k=m}^{M} z_k c_{k,m}$$

[Example 1] Suppose that

$$\mathbf{y} = \begin{bmatrix} 2 & 3 & 3 & 4 & 5 & 4 & 5 \end{bmatrix}^T$$

Try to express $\mathbf{y}$ as $x_1\mathbf{b_1} + x_2\mathbf{b_2} + x_3\mathbf{b_3}$ where

Case 4
non-orthogonal
incomplete

$$\mathbf{b_1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T$$

$$\mathbf{b_2} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{bmatrix}^T$$

$$\mathbf{b_3} = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{bmatrix}^T$$

such that

$$\left\| \mathbf{y} - x_1\mathbf{b_1} - x_2\mathbf{b_2} - x_3\mathbf{b_3} \right\| \text{ is minimized}$$

using the Gram-Schmidt method.

(Solution):

$$\mathbf{a_1} = \frac{\mathbf{b_1}}{\|\mathbf{b_1}\|} = \frac{1}{\sqrt{7}}\mathbf{b_1} = \frac{1}{\sqrt{7}}[1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$$

$$\mathbf{g_2} = \mathbf{b_2} - \mathbf{a_1}\sum_{n=1}^{7} b_2[n]a_1[n] = \mathbf{b_2} - 4\sqrt{7}\mathbf{a_1} = \mathbf{b_2} - 4\mathbf{b_1}$$

$$= [-3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3]$$

$$\mathbf{a_2} = \frac{\mathbf{g_2}}{\|\mathbf{g_2}\|} = \frac{\mathbf{g_2}}{2\sqrt{7}} = -\frac{2}{\sqrt{7}}\mathbf{b_1} + \frac{1}{2\sqrt{7}}\mathbf{b_2} = \frac{1}{2\sqrt{7}}[-3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3]$$

$$\mathbf{g_3} = \mathbf{b_3} - \mathbf{a_1}\sum_{n=1}^{7} b_3[n]a_1[n] - \mathbf{a_2}\sum_{n=1}^{7} b_3[n]a_2[n] = \mathbf{b_3} - \frac{1}{\sqrt{7}}\mathbf{a_1} - 0\mathbf{a_2} = \mathbf{b_3} - \frac{1}{7}\mathbf{b_1}$$

$$= \frac{2}{7}[3 \quad -4 \quad 3 \quad -4 \quad 3 \quad -4 \quad 3]$$

$$\mathbf{a_3} = \frac{\mathbf{g_3}}{\|\mathbf{g_3}\|} = \frac{7\mathbf{g_3}}{4\sqrt{21}} = \frac{-1}{4\sqrt{21}}\mathbf{b_1} + \frac{7}{4\sqrt{21}}\mathbf{b_3} = \frac{1}{2\sqrt{21}}[3 \quad -4 \quad 3 \quad -4 \quad 3 \quad -4 \quad 3]$$

Since

$$\sum_{n=1}^{7} y[n]a_1[n] = \frac{26}{\sqrt{7}} \qquad \sum_{n=1}^{7} y[n]a_2[n] = \frac{13}{2\sqrt{7}}$$

from page 619

$$\sum_{n=1}^{7} y[n]a_3[n] = \frac{1}{2\sqrt{21}}$$

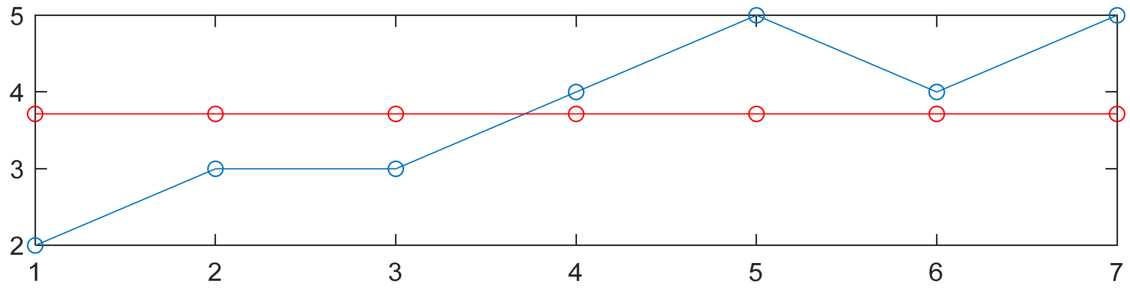
Therefore

$$y[n] \cong \frac{26}{\sqrt{7}}a_1[n] + \frac{13}{2\sqrt{7}}a_2[n] + \frac{1}{2\sqrt{21}}a_3[n]$$

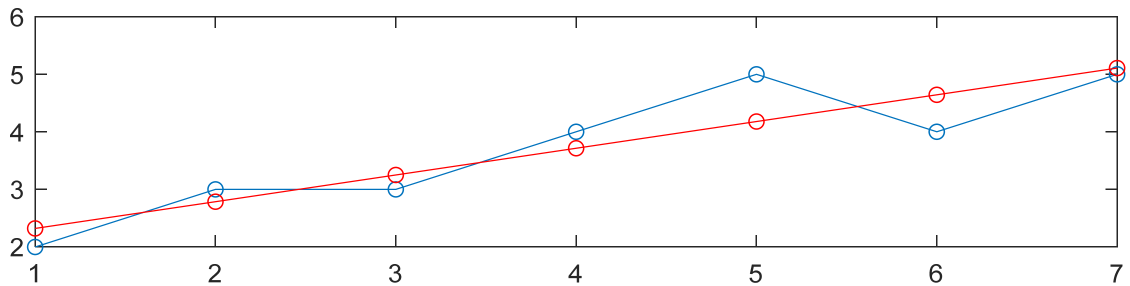
$$y[n] \cong \frac{311}{168}b_1[n] + \frac{13}{28}b_2[n] + \frac{1}{24}b_3[n]$$

$$= \begin{bmatrix} \frac{99}{42} & \frac{115}{42} & \frac{138}{42} & \frac{154}{42} & \frac{177}{42} & \frac{193}{42} & \frac{216}{42} \end{bmatrix}$$

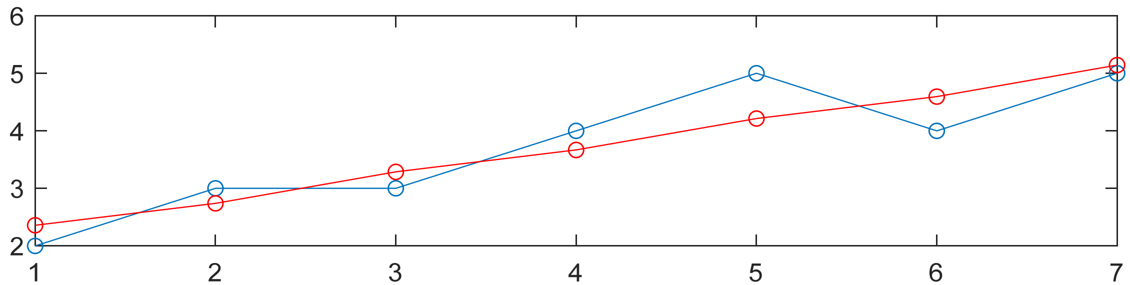
$$y[n] \cong \frac{26}{\sqrt{7}} a_1[n]$$



$$y[n] \cong \frac{26}{\sqrt{7}} a_1[n] + \frac{13}{2\sqrt{7}} a_2[n]$$



$$y[n] \cong \frac{26}{\sqrt{7}} a_1[n] + \frac{13}{2\sqrt{7}} a_2[n] + \frac{1}{2\sqrt{21}} a_3[n]$$



Least mean square error approximation
(LMSE approximation)

7.2.3 Method 3: Least Square Approximation

Suppose that $\{b_1[n], b_2[n], \ldots\ldots, b_M[n]\}$ are real and linearly independent but not orthogonal and incomplete. If we want to find x_m such that

$$E = \left\| \mathbf{y} - x_1\mathbf{b_1} - x_2\mathbf{b_2} - \cdots - x_M\mathbf{b}_M \right\|$$

is minimized, we can also apply the least square approximation method.

$$E^2 = \sum_{n=1}^{N}\left(y[n] - \sum_{k=1}^{M} x_k b_k[n] \right)^2$$

$$\frac{\partial}{\partial x_m}E^2 = \sum_{n=1}^{N}\left[\frac{\partial}{\partial x_m}\left(y[n] - \sum_{k=1}^{M} x_k b_k[n] \right)\right]2\left(y[n] - \sum_{k=1}^{M} x_k b_k[n] \right)$$

k=m

$$= \sum_{n=1}^{N} -2b_m[n]\left(y[n] - \sum_{k=1}^{M} x_k b_k[n] \right)$$

$$= -2\sum_{n=1}^{N} b_m[n]y[n] + 2\sum_{k=1}^{M} x_k \sum_{n=1}^{N} b_m[n]b_k[n] = 0$$

Therefore, if we want

$$\frac{\partial}{\partial x_m}E^2 = 0 \qquad \text{for } m = 1, 2, \ldots, M$$

then

$$\boxed{\sum_{k=1}^{M} x_k \sum_{n=1}^{N} b_m[n] b_k[n] = \sum_{n=1}^{N} b_m[n] y[n]} \qquad \text{for } m = 1, 2, \ldots, M$$

Therefore,

$$\mathbf{C}\mathbf{x} = \mathbf{z} \qquad\qquad \mathbf{x} = \mathbf{C}^{-1}\mathbf{z}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_M \end{bmatrix}^T \qquad \mathbf{z} = \begin{bmatrix} \sum_{n=1}^{N} b_1[n]y[n] & \sum_{n=1}^{N} b_2[n]y[n] & \cdots & \sum_{n=1}^{N} b_M[n]y[n] \end{bmatrix}^T$$

$$\mathbf{C} = \begin{array}{cc}
 & \begin{array}{cccc} \quad k=1 \quad & \quad k=2 \quad & & \quad k=M \quad \end{array} \\
\begin{array}{c} m=1 \\ m=2 \\ \\ m=N \end{array} &
\begin{bmatrix}
\sum_{n=1}^{N} b_1[n]b_1[n] & \sum_{n=1}^{N} b_1[n]b_2[n] & \cdots & \sum_{n=1}^{N} b_1[n]b_M[n] \\
\sum_{n=1}^{N} b_2[n]b_1[n] & \sum_{n=1}^{N} b_2[n]b_2[n] & \cdots & \sum_{n=1}^{N} b_2[n]b_M[n] \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{n=1}^{N} b_M[n]b_1[n] & \sum_{n=1}^{N} b_M[n]b_2[n] & \cdots & \sum_{n=1}^{N} b_M[n]b_M[n]
\end{bmatrix}
\end{array}$$

Also note that, if

$$\mathbf{A} = \begin{bmatrix} b_1[1] & b_2[1] & b_3[1] & \cdots & b_M[1] \\ b_1[2] & b_2[2] & b_3[2] & \cdots & b_M[2] \\ b_1[3] & b_2[3] & b_3[3] & \cdots & b_M[3] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_1[N] & b_2[N] & b_3[N] & \cdots & b_M[N] \end{bmatrix}$$

then

$$\mathbf{C} = \mathbf{A}^{\mathrm{T}}\mathbf{A}$$

$$\mathbf{z} = \mathbf{A}^{T}\mathbf{y} \quad \text{where} \quad \mathbf{y} = \begin{bmatrix} y[1] & y[2] & \cdots & y[M] \end{bmatrix}^{T}$$

Therefore, from $\mathbf{x} = \mathbf{C}^{-1}\mathbf{z}$, we have

☆ $$\boxed{\mathbf{x} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y}}$$ LMSE solution

[Example 2] Suppose that

$$\mathbf{y} = \begin{bmatrix} 2 & 3 & 3 & 4 & 5 & 4 & 5 \end{bmatrix}^T$$

Try to express $\mathbf{y}$ as $x_1\mathbf{b_1} + x_2\mathbf{b_2} + x_3\mathbf{b_3}$ where

$$\mathbf{b_1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T$$

$$\mathbf{b_2} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{bmatrix}^T$$

$$\mathbf{b_3} = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{bmatrix}^T$$

such that

$$\left\| \mathbf{y} - x_1\mathbf{b_1} - x_2\mathbf{b_2} - x_3\mathbf{b_3} \right\| \quad \text{is minimized}$$

using the least square approximation method.

First, we construct the matrix

$$\mathbf{A} = \begin{array}{c} \begin{array}{ccc} \mathbf{b_1} & \mathbf{b_2} & \mathbf{b_3} \end{array} \\ \left[\begin{array}{c:c:c} 1 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & 3 & 1 \\ 1 & 4 & -1 \\ 1 & 5 & 1 \\ 1 & 6 & -1 \\ 1 & 7 & 1 \end{array}\right] \end{array}$$

7x3 matrix

Since

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 7 & 28 & 1 \\ 28 & 140 & 4 \\ 1 & 4 & 7 \end{bmatrix} \qquad (\mathbf{A}^T\mathbf{A})^{-1} = \frac{1}{336}\begin{bmatrix} 241 & -48 & -7 \\ -48 & 12 & 0 \\ -7 & 0 & 49 \end{bmatrix}$$

$$(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \frac{1}{168}\begin{bmatrix} 93 & 76 & 45 & 28 & -3 & -20 & -51 \\ -18 & -12 & -6 & 0 & 6 & 12 & 18 \\ 21 & -28 & 21 & -28 & 21 & -28 & 21 \end{bmatrix}$$

therefore, from $\mathbf{x} = (\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}\mathbf{y}$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{168}\begin{bmatrix} 93 & 76 & 45 & 28 & -3 & -20 & -51 \\ -18 & -12 & -6 & 0 & 6 & 12 & 18 \\ 21 & -28 & 21 & -28 & 21 & -28 & 21 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ 3 \\ 4 \\ 5 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 311/168 \\ 13/28 \\ 1/24 \end{bmatrix}$$

$$y[n] \cong \frac{311}{168} b_1[n] + \frac{13}{28} b_2[n] + \frac{1}{24} b_3[n]$$

$$= \begin{bmatrix} \frac{99}{42} & \frac{115}{42} & \frac{138}{42} & \frac{154}{42} & \frac{177}{42} & \frac{193}{42} & \frac{216}{42} \end{bmatrix}$$

(the same as Example 1)

7.3 Generalized Inverse

Remember that, for the case where the vector sets are linearly independent and complete, one can use the matrix inverse method (pages 624, 625) to determine the linear combination coefficients:

If $\mathbf{y} = \mathbf{A}\mathbf{x}$

then $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ $x \cong A^{+}y$

However, when

(1) The vector sets are not linearly independent (i.e., det($\mathbf{A}$) = 0)

(2) The number of vector sets is smaller than the vector length

(i.e., $\mathbf{A}$ is not a square matrix)

$\mathbf{A}^{-1}$ is hard to be determined.

[Definition] Generalized Inverse

For an matrix $\mathbf{A}$, if there is a matrix $\mathbf{A}^+$ such that

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$$

then $\mathbf{A}^+$ is called the generalized inverse of $\mathbf{A}$.

We always use $\mathbf{A}^+$ to denote the generalized inverse of $\mathbf{A}$.

If A^{-1} exists, $A(A^{-1}A) = AI = A$

$A^+ = A^{-1}$

[Additional Definitions for Generalized Inverse]

(1) $\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}$

(2) $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+}$

(3) $\left(\mathbf{A}\mathbf{A}^{+}\right)^{H} = \mathbf{A}\mathbf{A}^{+}$

(4) $\left(\mathbf{A}^{+}\mathbf{A}\right)^{H} = \mathbf{A}^{+}\mathbf{A}$

If (1) is satisfied, then $\mathbf{A}^{+}$ is called the generalized inverse of $\mathbf{A}$.

If (1) and (2) are satisfied, then $\mathbf{A}^{+}$ is called the reflexive generalized inverse of $\mathbf{A}$.

If (1), (2), (3), and (4) are all satisfied, then $\mathbf{A}^{+}$ is called the pseudo inverse of $\mathbf{A}$.

pseudo inverse $\subset$ reflexive generalized inverse $\subset$ generalized inverse

$size(\mathbf{A}) = M \times N$

- $M = N$
  - $\det(\mathbf{A}) \neq 0$ (Case 1)
    (columns are independent)
  - $\det(\mathbf{A}) = 0$ (Case 3)
    (columns are dependent)
- $M > N$
  - columns are independent (Case 2)
  - columns are dependent (Case 4) SVD
- $M < N$
  - columns must be dependent (Case 4) SVD

[Case 1] If **A** is a square matrix and all the columns of **A** are linearly independent, then

$$\mathbf{A}^{+} = \mathbf{A}^{-1}$$

Note that, in this case,

$$\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}\mathbf{I} = \mathbf{A}$$

[Case 2] If **A** is an MxN matrix, $N < M$, and all the columns of **A** are linearly independent, then

$$\mathbf{A}^{+} = \underline{(\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}}$$ LMSE solution

Note that, in this case,

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

Also note that it is the same as the least square approximation method introduced in subsection 7-2-3

[Case 3] Suppose that **A** is a square matrix and some columns of **A** are dependent. Then, in this case

$$\det(\mathbf{A}) = 0$$

and some of the eigenvalues of **A** are equal to zero.

[Case 3-1] Suppose that the eigenvector-eigenvalue decomposition of **A** exists

$$\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^{-1}$$

(i.e., eigenvectors form a complete set)

where **D** is a diagonal matrix where the diagonal entries are the eigenvalues of **A**.

$$D[m,n] = \begin{cases} \lambda_n & if\ m = n \\ 0 & if\ m \neq n \end{cases}$$

If $\det(A) = 0$ some eigenvalue λ_n is zero

Then, the generalized inverse of **A** is

If $\det(A) \neq 0$, $A = ED^{-1}E^{-1}$

$$\mathbf{A}^{+} = \mathbf{E}\mathbf{D}^{+}\mathbf{E}^{-1} \quad \text{where} \quad D^{+}[m,n] = \begin{cases} 1/\lambda_n & if\ m = n\ and\ \lambda_n \neq 0 \\ 0 & if\ m = n\ and\ \lambda_n = 0 \\ 0 & if\ m \neq n \end{cases}$$

Note that, in this case,

$$\mathbf{AA^{+}A} = \mathbf{EDE^{-1}ED^{+}E^{-1}EDE^{-1}} = \mathbf{EDD^{+}DE^{-1}}$$

If

$$\mathbf{S} = \mathbf{DD^{+}D}$$

then

$$S[n,n] = \lambda_n \lambda_n^{-1} \lambda_n = \lambda_n \qquad \text{if } \lambda_n \neq 0$$

$$S[n,n] = \lambda_n 0 \lambda_n = 0 = \lambda_n \qquad \text{if } \lambda_n = 0$$

$$S[m,n] = 0 \qquad \text{if } m \neq n$$

Therefore,

$$\mathbf{S} = \mathbf{DD^{+}D} = \mathbf{D}$$

$$\mathbf{AA^{+}A} = \mathbf{EDE^{-1}} = \mathbf{A}$$

[Example 1] Suppose that

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

det(A)=0

Case 3

Determine the generalized inverse of **A**.

(Solution): The eigenvalues of **A** is $\lambda = 0, 1, 3$

The eigenvectors are

$$[1 \quad -1 \quad 1]^T \qquad \text{corresponding to } \lambda = 0$$

$$[1 \quad 0 \quad -1]^T \qquad \text{corresponding to } \lambda = 1$$

$$[1 \quad 2 \quad 1]^T \qquad \text{corresponding to } \lambda = 3$$

Therefore, the eigenvector-eigenvalue decomposition of **A** is

$$\mathbf{A} = \begin{array}{c} \begin{array}{ccc} \lambda=0 & \lambda=1 & \lambda=3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 2 \\ 1 & -1 & 1 \end{bmatrix} \end{array} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 2 \\ 1 & -1 & 1 \end{bmatrix}^{-1}$$

Since

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 2 \\ 1 & -1 & 1 \end{bmatrix} \overset{D}{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}} \begin{bmatrix} 1/3 & -1/3 & 1/3 \\ 1/2 & 0 & -1/2 \\ 1/6 & 2/6 & 1/6 \end{bmatrix}$$

we have

$$\mathbf{A}^{+} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 2 \\ 1 & -1 & 1 \end{bmatrix} \overset{D^{+}}{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/1 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}} \begin{bmatrix} 1/3 & -1/3 & 1/3 \\ 1/2 & 0 & -1/2 \\ 1/6 & 2/6 & 1/6 \end{bmatrix}$$

$$\mathbf{A}^{+} = \begin{bmatrix} 5/9 & 1/9 & -4/9 \\ 1/9 & 2/9 & 1/9 \\ -4/9 & 1/9 & 5/9 \end{bmatrix}$$

One can show that

$$\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \mathbf{A}$$

[Case 3-2]

[Generalized Inverse when the Eigenvectors are not Complete]

If $\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^{-1}$ where $\mathbf{D} = \begin{bmatrix} \mathbf{D_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D_K} \end{bmatrix}$

$$\mathbf{D_k} = \lambda_k, \quad \begin{bmatrix} \lambda_k & 0 & \cdots & 0 \\ 0 & \lambda_k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} \lambda_k & 1 & 0 & \cdots & 0 \\ 0 & \lambda_k & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k & 1 \\ 0 & 0 & \cdots & 0 & \lambda_k \end{bmatrix}$$

then

$\mathbf{A}^{+} = \mathbf{E}\mathbf{D}^{+}\mathbf{E}^{-1}$ where $\mathbf{D}^{+} = \begin{bmatrix} \mathbf{D_1^{+}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D_2^{+}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D_K^{+}} \end{bmatrix}$

When $\lambda_k \neq 0$

if $\mathbf{D_k} = \lambda_k$, then $\mathbf{D_k^+} = 1/\lambda_k$,

(1) If $\mathbf{D_k} = \begin{bmatrix} \lambda_k & 0 & \cdots & 0 \\ 0 & \lambda_k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$, then $\mathbf{D_k^+} = \begin{bmatrix} 1/\lambda_k & 0 & \cdots & 0 \\ 0 & 1/\lambda_k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\lambda_k \end{bmatrix}$,

$\lambda_k \neq 0$

(2) If $\mathbf{D_k} = \begin{bmatrix} \lambda_k & 1 & 0 & \cdots & 0 \\ 0 & \lambda_k & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k & 1 \\ 0 & 0 & \cdots & 0 & \lambda_k \end{bmatrix}$, $\mathbf{D_k^+} = \begin{bmatrix} \lambda_k^{-1} & -\lambda_k^{-2} & \lambda_k^{-3} & \cdots & (-1)^M \lambda_k^{-M} \\ 0 & \lambda_k^{-1} & -\lambda_k^{-2} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \lambda_k^{-3} \\ 0 & 0 & \cdots & \lambda_k^{-1} & -\lambda_k^{-2} \\ 0 & 0 & \cdots & 0 & \lambda_k^{-1} \end{bmatrix}$

(suppose that the size of $\mathbf{D_k}$ is MxM)

One can show that $\mathbf{D_k D_k^+} = \mathbf{I}$

$D_k^+ = D_k^{-1}$

When $\lambda_k = 0$

if $\quad \mathbf{D_k} = \lambda_k, \quad$ then $\quad \mathbf{D_k^+} = 0,$

(3) If $\quad \mathbf{D_k} = \begin{bmatrix} \lambda_k & 0 & \cdots & 0 \\ 0 & \lambda_k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}, \quad$ then $\quad \mathbf{D_k^+} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$

$\lambda_k = 0$

(4) If $\quad \mathbf{D_k} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad$ then $\quad \mathbf{D_k^+} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$

Note that if

$$\mathbf{D_k} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \qquad \mathbf{D_k^+} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

then $$\mathbf{D_k^+ D_k} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$\mathbf{D_k D_k^+ D_k} = \mathbf{D_k}$$

[Case 4] Suppose that **A** is an MxN matrix, when
(i) $M < N$ or
(ii) $N < M$ but some column vectors are not linearly independent,
the methods introduced in this chapter cannot be applied.

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

$\det(A^T A)$ may be zero

We can use the singular value decomposition (SVD) method introduced in Section 8.1 to solve the generalized inverse problem in Cases 1, 2, 3, and 4.

7.4 Discrete Orthogonal Polynomials

(只教不考)

[Definition of Discrete Orthogonal Polynomials]

Suppose that there is a set of discrete functions as follows

$$P_m[n] = \sum_{k=0}^{m} c_{m,k}(n)_k \qquad m = 0, 1, 2, \ldots\ldots$$

where $(n)_k$ is called the falling factorial function:

$$(n)_0 = 1, \qquad (n)_1 = n, \qquad (n)_2 = n(n-1),$$

$$(n)_k = n(n-1)(n-2)\cdots\cdots(n-k+1)$$

n^k 變形

If

$$\boxed{\sum_{n=n_0}^{n_1} w[n]P_m[n]P_s[n] = 0 \quad when\ m \neq s}$$

then we call $\{P_0[n], P_1[n], P_2[n], \ldots\ldots\}$ a discrete orthogonal polynomial set within $n \in [n_0, n_1]$ with the weight $w[n]$

Note that since

$$span\{(n)_0,(n)_1,(n)_2,\cdots,(n)_m\} = span\left\{1,n,n^2,\cdots,n^m\right\}$$

therefore, $P_m[n]$ can also be expressed as a linear combination of 1, n, n^2, …, n^m.

[Discrete Legendre Polynomials]

$$w[n]=1 \qquad n\in[0,N]$$

The Discrete Legendre Polynomial of Order m

$$\boxed{P_m[n]=\sum_{k=0}^{m}(-1)^k\binom{m}{k}\binom{m+k}{k}\frac{(n)_k}{(N)_k}}$$

$$\sum_{n=0}^{N}P_m[n]P_s[n]=\begin{cases}\dfrac{(N+m+1)!(N-m)!}{(2m+1)(N!)^2} & if\ m=s\\ 0 & if\ m\neq s\end{cases}$$

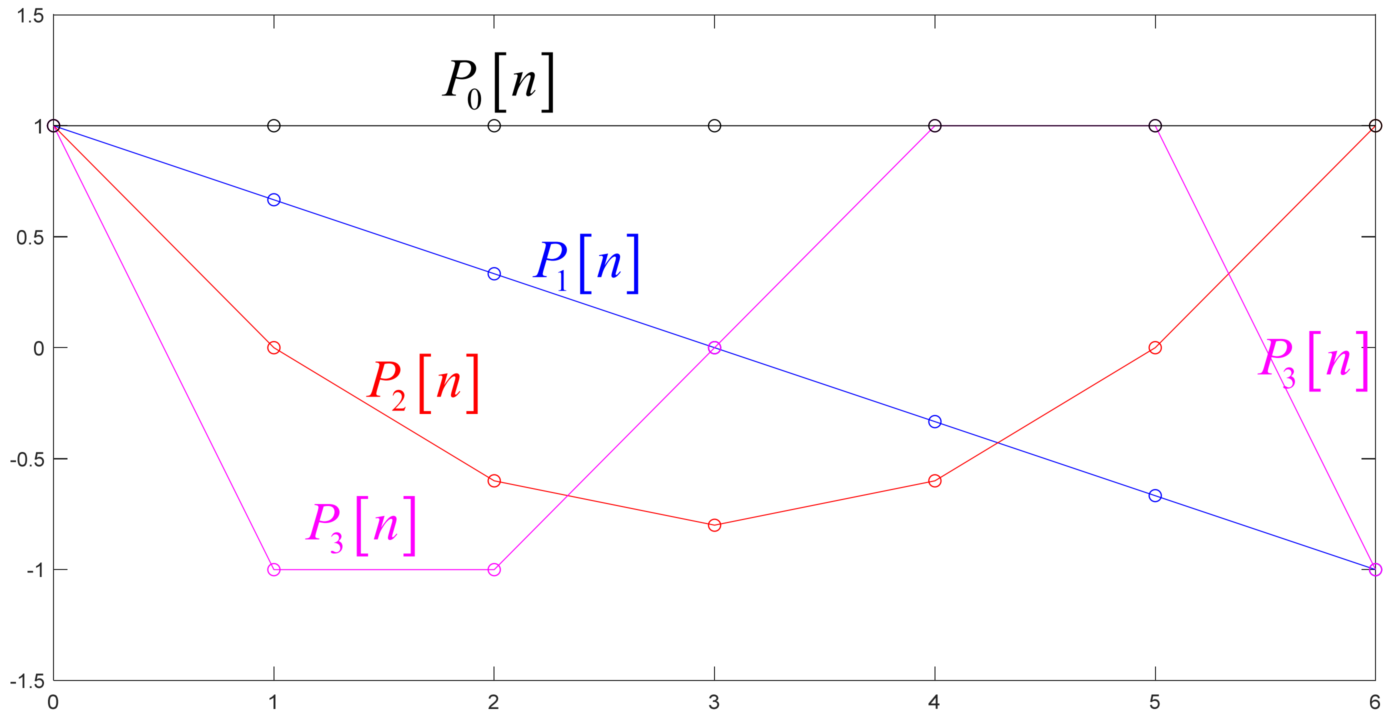
$$P_0[n]=1 \qquad P_1[n]=1-2\frac{n}{N}$$

$$P_2[n]=1-6\frac{n}{N}+6\frac{(n)_2}{(N)_2} \qquad P_3[n]=1-12\frac{n}{N}+30\frac{(n)_2}{(N)_2}-20\frac{(n)_3}{(N)_3}$$

$$P_4[n]=1-20\frac{n}{N}+90\frac{(n)_2}{(N)_2}-140\frac{(n)_3}{(N)_3}+70\frac{(n)_4}{(N)_4}$$

$(n)_4 = n(n-1)(n-2)(n-3)$

$N = 6$



discrete Jacobi

[Hahn Polynomials]

Two extra parameters: α, β

$$w[n] = \binom{n+\alpha}{n}\binom{N-n+\beta}{N-n} \qquad n \in [0, N]$$

When $\alpha = \beta = -1/2$, it is analogous to the continuous Chebyshev polynomial on page 319.

If α or β is not an integer, it can still be defined:

$$w[n] = \frac{\Gamma(n+\alpha+1)}{\Gamma(n+1)\Gamma(\alpha+1)}\frac{\Gamma(N-n+\beta+1)}{\Gamma(N-n+1)\Gamma(\beta+1)}$$

The Hahn Polynomial of Order m

$$\boxed{P_m[n] = {}_3F_2\left(\begin{matrix}-m, -n, m+\alpha+\beta+1; \\ \alpha+1, -N; 1\end{matrix}\right)}$$

$${}_pF_q\left(\begin{matrix}a_1,a_2,\cdots,a_p;\\ b_1,b_2,\cdots,b_q;z\end{matrix}\right):$$ hypergeometric function

$${}_pF_q\left(\begin{matrix}a_1,a_2,\cdots,a_p;\\ b_1,b_2,\cdots,b_q;z\end{matrix}\right)=\sum_{k=0}^{\infty}\frac{a_1^{(k)}a_2^{(k)}\cdots a_p^{(k)}}{b_1^{(k)}b_2^{(k)}\cdots b_q^{(k)}}\frac{z^k}{k!}$$

where $a^{(k)}$ is called the rising factorial function:

$$a^{(0)}=1$$
$$a^{(k)}=a(a+1)(a+2)\cdots\cdots(a+k-1)$$

discrete		continuous
Hahn polynomials	$\xrightarrow{\text{analogous}}$	Jacobi polynomials
Meixner polynomials	$\xrightarrow{\text{analogous}}$	Laguerre polynomials
Krawtchouk polynomials	$\xrightarrow{\text{analogous}}$	Hermite polynomials (refer to page 322)

Hahn polynomials α, β (discrete Jacobi polynomials) —— $\alpha = \beta$ ——> discrete ultraspherical polynomials

discrete ultraspherical polynomials:

- $\alpha = 0$: discrete Legendre polynomials
- $\alpha = -1/2$: discrete Chebyshev polynomials (I)
- $\alpha = 1/2$: discrete Chebyshev polynomials (II)

[Meixner Polynomials]

Two extra parameters: A, b

$$w[n] = A^n \frac{b^{(n)}}{n!} \qquad n \in [0, \infty)$$

The Meixner Polynomial of Order m

$$P_m[n] = {}_2F_1\begin{pmatrix} -m, -n; \\ b; 1 - \frac{1}{A} \end{pmatrix}$$

Note: When

$$A = e^{-\lambda}, \quad b = 1$$

then

$$w[n] = e^{-\lambda n}$$

(the same weight function as the continuous Laguarre polynomial)

When $A = e^{-\lambda}$, $b = 1$, it is analogous to the continuous Laguerre polynomial on page 320.

$b = 1$ $b^{(n)} = 1 \cdot 2 \cdot \cdots (n-1)$
$= (n-1)!$

If $A = e^{-\lambda}$, $b = 1$

$w[n] = \frac{e^{-\lambda n}}{n}$ $n\,w[n] = e^{-\lambda n}$

$$e^{-x} = \sum_{n=0}^{\infty} \frac{(-x)^n}{n!}$$

As shown on the next page, when p = 1/2, it is analogous to the continuous Hermite polynomial on page 322.

[Krawtchouk Polynomials]

One extra parameter: p

$$w[n] = p^n (1-p)^{N-n} \binom{N}{n} \qquad\qquad n \in [0, N]$$

(Similar to the Binomial distribution)

The Krawtchouk Polynomial of Order m

$$P_m[n] = {}_2F_1\left(\begin{matrix} -m, -n; \\ -N; \frac{1}{p} \end{matrix} \right)$$

$$w[n] = p^n (1-p)^{N-n} \binom{N}{n}$$

Note: When

$$p = 1/2$$

then

$$w[n] = \binom{N}{n}$$

Moreover, when $N \to \infty$

$$\lim_{N\to\infty} \binom{N}{n} \cong \frac{2^N}{\sqrt{N\pi/2}} \exp\left(-\frac{(n-N/2)^2}{N/2}\right)$$

which is near to the weight function of the continuous Hermite polynomial. Therefore, the Krawtchouk polynomial is also called the discrete Hermite polynomial.

附錄十 Approximation Using Other Norms

Until now, we discuss the approximation problem based on the L_2 norm, that is, to find $\mathbf{x}$ that can minimize

$$\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|$$

$$\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\| = \sqrt{\sum_{n=1}^{N}\left(y[n]-\sum_{m=1}^{M}A[n,m]x_m\right)^2}$$

However, how do we minimize the approximation problem based on the L_α norm, that is, to find $\mathbf{x}$ that can minimize

$$\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|_\alpha$$

$$\left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\| = \sqrt[\alpha]{\sum_{n=1}^{N}\left|y[n]-\sum_{m=1}^{M}A[n,m]x_m\right|^\alpha}$$

The problem of minimizing

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_\alpha$$

is always hard to solve if $\alpha \neq 2$.

However, when $\alpha \geq 1$, $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_\alpha$ is convex, which means that $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_\alpha$ has only one local minimum (i.e., local minimum = global minimum). Therefore, many numerical methods (the simplex algorithm, Golden search, gradient descent, Newton's method, …..) can be applied to minimize $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_\alpha$. We describe the general method to minimize $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_\alpha$ when $\alpha \geq 1$ as follows.

It is even harder to minimize $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_\alpha$ when $\alpha < 1$.

(Problem): Determine

$$\mathbf{x} = \arg\min_{\mathbf{x}} \left\| \mathbf{y} - \mathbf{A}\mathbf{x} \right\|_{\alpha}$$

It means that to find $\mathbf{x}$ that can minimize $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\alpha}$

Suppose that

size($\mathbf{A}$) = $N \times M$, length($\mathbf{y}$) = N, length($\mathbf{x}$) = M

$M < N$

(Step 1): Initial: $\mathbf{x} = \mathbf{0}$, $E_0 = \|\mathbf{y}\|_{\alpha}$, $c = 1$, $try = 0$

Set Δ (the threshold for error convergence)

Set T (the upper bound of times for no error reduction)

(Step 2): Choose the feasible direction as follows.

(Method 1): Assign the feasible direction **b** as the projection of $\mathbf{y} - \mathbf{A}\mathbf{x}$ on

$$span\left(\mathbf{A_1}, \mathbf{A_2}, \cdots\cdots, \mathbf{A_M}\right)$$

where $\mathbf{A_1}$, $\mathbf{A_2}$, …, $\mathbf{A_M}$ are columns of **A**.

(Method 2): If the projection is 0 or **c** = 0 (i.e., the adjusting step in the previous iteration is zero)
Generate d_m randomly.
Then, set the feasible direction **b** as

$$\mathbf{b} = \sum_{m=1}^{M} d_m \mathbf{A_m} / \left\| \mathbf{A_m} \right\|$$

(Step 3): Find c to minimize $\|\mathbf{y} - \mathbf{A}(\mathbf{x} + c\mathbf{b})\|_\alpha$

$$c = \arg\min_c \left\| \mathbf{y} - \mathbf{A}(\mathbf{x} + c\mathbf{b}) \right\|_\alpha$$

Then, update **x** as

$$\mathbf{x} \Leftarrow \mathbf{x} + c\left[e_1, e_2, \cdots, e_M\right] \qquad \text{if} \quad \mathbf{b} = e_1\mathbf{A_1} + e_2\mathbf{A_2} + \cdots e_M\mathbf{A_M}$$

(Step 4): Determine $E_1 = \|\mathbf{y} - \mathbf{Ax}\|_\alpha$. If

$$E_0 - E_1 < \Delta$$

then set

$$try \Leftarrow try + 1$$

Otherwise, set $try = 0$.

(Step 5): If $try \le T$:

Set $E_0 = E_1$ and return to (Step 2)

If $try > T$:

The process is terminated and the solution is obtained.

Flowchart

Initial: $\mathbf{x} = \mathbf{0}$

↓

Assign the feasible direction **b** as the projection of $\mathbf{y} - \mathbf{A}\mathbf{x}$ on *span*(Columns of **A**)
(If the projection is 0, assign **b** randomly)

↓

Find c iteratively to minimize $\|\mathbf{y} - \mathbf{A}(\mathbf{x} + c\mathbf{b})\|_\alpha$

↓

Update **x** by $\mathbf{x} \Leftarrow \mathbf{x} + c\left[e_1, e_2, \cdots, e_M\right]$

if $\mathbf{b} = e_1\mathbf{A_1} + e_2\mathbf{A_2} + \cdots e_M\mathbf{A_M}$

$A = [A_1 \vdots A_2 \vdots \cdots \vdots A_M]$

↓

Whether the error does not decrease for T times

No → back to "Assign the feasible direction **b**"

Yes ↓

Obtain the solution of **x**

[Example 1] Suppose that

$$\mathbf{y} = [2 \quad 3 \quad 3 \quad 4 \quad 5 \quad 4 \quad 5]$$

Try to express $\mathbf{y}$ as $x_1\mathbf{b_1} + x_2\mathbf{b_2} + x_3\mathbf{b_3}$ where

$$\mathbf{b_1} = [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$$

$$\mathbf{b_2} = [1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7]$$

$$\mathbf{b_3} = [1 \quad -1 \quad 1 \quad -1 \quad 1 \quad -1 \quad 1]$$

such that

L1 norm

$$\left\| \mathbf{y} - x_1\mathbf{b_1} - x_2\mathbf{b_2} - x_3\mathbf{b_3} \right\|_1 \quad \text{is minimized}$$

minimize $\sum_n \left| y[n] - x_1 b_1[n] - x_2 b_2[n] - x_3 b_3[n] \right|$

(Solution): (Step 1): Initially, set

$$[x_1, x_2, x_3] = [0, 0, 0]$$

$$E_0 = \left\| \mathbf{y} - x_1\mathbf{b_1} - x_2\mathbf{b_2} - x_3\mathbf{b_3} \right\|_1 = 26$$

(Step 2):

Then, we find the projection of $\mathbf{y} - 0\mathbf{b}_1 - 0\mathbf{b}_2 - 0\mathbf{b}_3 = \mathbf{y}$ on Span($\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$):

$$\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 \xrightarrow[\text{Gram-Schmidt}]{} \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$$

$$\mathbf{a}_1 = \frac{1}{\sqrt{7}}\begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1\end{bmatrix} \qquad \mathbf{a}_2 = \frac{1}{2\sqrt{7}}\begin{bmatrix}-3 & -2 & -1 & 0 & 1 & 2 & 3\end{bmatrix}$$

$$\mathbf{a}_3 = \frac{1}{2\sqrt{21}}\begin{bmatrix}3 & -4 & 3 & -4 & 3 & -4 & 3\end{bmatrix}$$

Since

$$\sum_n \mathbf{y}[n]\mathbf{a}_1[n] = 9.2871 \qquad \sum_n \mathbf{y}[n]\mathbf{a}_2[n] = 2.4568$$

$$\sum_n \mathbf{y}[n]\mathbf{a}_3[n] = 0.1091$$

the projection of $\mathbf{y}$ on Span($\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$) is

$$9.2871\mathbf{a}_1 + 2.4568\mathbf{a}_2 + 0.1091\mathbf{a}_3 = 1.8512\mathbf{b}_1 + 0.4643\mathbf{b}_2 + 0.0417\mathbf{b}_3$$

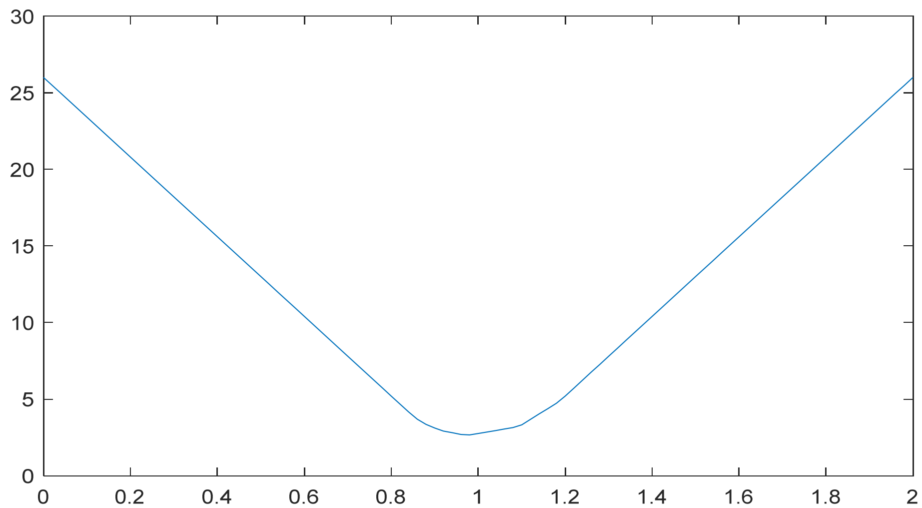
Therefore, we choose the feasible direction **b** as

$$\mathbf{b} = 1.8512\,\mathbf{b}_1 + 0.4643\,\mathbf{b}_2 + 0.0417\,\mathbf{b}_3$$
$$= [2.3571,\ 2.7381,\ 3.2857,\ 3.6667,\ 4.2143,\ 4.5952,\ 5.1429]$$

(Step 3): Find c to minimize $\|\mathbf{y} - \mathbf{A}(\mathbf{x} + c\mathbf{b})\|_1$

Newton's Method
Golden search
gradient descend

$\|\mathbf{y} - \mathbf{A}(\mathbf{x} + c\mathbf{b})\|_1$



c

The solution is $c = 0.9722$. Then, update **x** as

$$\mathbf{x} \Leftarrow \mathbf{x} + 0.9722\mathbf{b} = [1.7998,\ 0.4514,\ 0.0405]$$

(Step 4): Determine the residue

$$\mathbf{y} - \mathbf{A}\mathbf{x} = [-0.2917,\ 0.338,\ -0.1944,\ 0.4352,\ 0.9028,\ -0.4676,\ 0]$$

and calculate the error

$$E_1 = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_1 = 2.6296$$

(Step 5): Return to (Step 2)

$$\vdots$$

$$\vdots$$

$$\vdots$$

After 60-110 times of iterations, we obtain

$$\mathbf{x} = [1.75,\ 0.5,\ -0.25]$$

$$\mathbf{y} - \mathbf{A}\mathbf{x} = [0, 0, 0, 0, 1, -1, 0]$$

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_1 = 2$$

8. Component Analysis

8.1 Singular Value Decomposition

If **A** is a square matrix, then we can perform eigenvector-eigenvalue decomposition for **A**:

$$\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^{-1}$$

$$\mathbf{A} = \lambda_1 \mathbf{e_1}\mathbf{f_1^H} + \lambda_2 \mathbf{e_2}\mathbf{f_2^H} + \cdots + \lambda_{N-1}\mathbf{e_{N-1}}\mathbf{f_{N-1}^H} + \lambda_N \mathbf{e_N}\mathbf{f_N^H}$$

where E, E^{-1} have been normalized

$$\mathbf{E} = \begin{bmatrix}\mathbf{e_1} & \mathbf{e_2} & \cdots & \mathbf{e_N}\end{bmatrix}, \quad \mathbf{E}^{-1} = \begin{bmatrix}\mathbf{f_1} & \mathbf{f_2} & \cdots & \mathbf{f_N}\end{bmatrix}^{\mathbf{H}} = \begin{bmatrix} f_1^H \\ f_2^H \\ \vdots \\ f_N^H \end{bmatrix}$$

$$\mathbf{A}\mathbf{e_m} = \lambda_m \mathbf{e_m}$$

If $|\lambda_m|$ is the largest, then

$$\lambda_m \mathbf{e_m}\mathbf{f_m^H}$$

is the most important component of **A**.

8.1.1 Singular Value Decomposition Process

Q: How do we perform eigenvector-eigenvalue decomposition for **A** if **A** is not a square matrix?

$$size(\mathbf{A}) = \underset{5\times 2}{M \times N}, \quad M \neq N$$

We can apply the singular value decomposition (SVD) process as follows.

奇值

SVD result
≠ eigenvector-eigenvalue
decomposition result

(1) Generate **B** and **C**

$$\underset{2\times 2}{\mathbf{B} = \mathbf{A}^H\mathbf{A}} \qquad \underset{5\times 5}{\mathbf{C} = \mathbf{A}\mathbf{A}^H}$$

(Note): Since **B** is an *N*x*N* square matrix,
C is an *M*x*M* square matrix,
therefore, it is possible to derive the eigenvector sets for **B** and **C**.

$$\mathbf{B} = \mathbf{A}^H\mathbf{A} \qquad\qquad \mathbf{C} = \mathbf{A}\mathbf{A}^H$$

(2) Perform Eigenvector-Eigenvalue Decomposition for **B** and **C**

$$\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1} \qquad\qquad \mathbf{C} = \tilde{\mathbf{U}}\boldsymbol{\Omega}\tilde{\mathbf{U}}^{-1}$$

(Note): Since $\mathbf{B}^H = \mathbf{B}$, $\mathbf{C}^H = \mathbf{C}$, **B** and **C** have orthogonal eigenvector sets and $\tilde{\mathbf{U}}$ and **V** are orthogonal matrices.

(i) It is proper to normalize $\tilde{\mathbf{U}}$ and **V** properly such that

$$\mathbf{V}^H\mathbf{V} = \mathbf{I} \qquad\qquad \tilde{\mathbf{U}}^H\tilde{\mathbf{U}} = \mathbf{I}$$

then

$$V^{-1} = V^H \qquad\qquad \tilde{U}^{-1} = \tilde{U}^H$$

$$\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^H \qquad\qquad \mathbf{C} = \tilde{\mathbf{U}}\boldsymbol{\Omega}\tilde{\mathbf{U}}^H$$

(ii) It is proper to sort the eigenvalues of **B** and **C** from large to small.

The eigenvectors are also sorted according to eigenvalues.

(3) Then, we calculate

$$\underbrace{\mathbf{S_1}}_{5\times 2} = \underbrace{\tilde{\mathbf{U}}^{\mathbf{H}}}_{\substack{5\times5\\M\times M}}\underbrace{\mathbf{A}}_{\substack{5\times2\\M\times N}}\underbrace{\mathbf{V}}_{\substack{2\times2\\N\times N}}$$

$\mathbf{S_1}$ will be an *M*x*N* diagonal matrix

$$S_1[m,n] = 0 \quad if\ m \neq n$$

(4) Varying the sign of $\mathbf{S_1}$ and $\tilde{\mathbf{U}}$

$$S[m,n] = |S_1[m,n]|$$

$$U[m,n] = \tilde{U}[m,n] \quad \text{if } S_1[n, n] \geq 0,$$

$$U[m,n] = -\tilde{U}[m,n] \quad \text{if } S_1[n, n] < 0,$$

change the sign of nth column

(Note): With sign change,

$\mathbf{S} = \mathbf{U^H A V}$ and $\mathbf{C} = \mathbf{U \Omega U^{-1}}$

are still satisfied.

$S = U^H A V$

$U S V^H = U U^H A V V^H = I A I = A$

(5) Then, the SVD of **A** is

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

eigenvector matrix of $\mathbf{A}\mathbf{A}^{\mathbf{H}}$, size: MxM

diagonal matrix, size: MxN

eigenvector matrix of $\mathbf{A}^{\mathbf{H}}\mathbf{A}$, size: NxN

If

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \cdots & \mathbf{u_M} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} & \cdots & \mathbf{v_N} \end{bmatrix}$$

then

☆ Note: u_m, v_n should be normalized

$$\mathbf{A} = \underbrace{s_1\mathbf{u_1}\mathbf{v_1^H}}_{\text{most significant}} + \underbrace{s_2\mathbf{u_2}\mathbf{v_2^H}}_{\text{second significant}} + \cdots + s_{K-1}\mathbf{u_{K-1}}\mathbf{v_{K-1}^H} + s_K\mathbf{u_K}\mathbf{v_K^H}$$

$$s_1 \geq s_2 \geq \cdots \geq s_{K-1} \geq s_K$$

where $K = \min(M, N)$

$$s_n = S[n, n]$$

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_N \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

if $M > N$

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_M & 0 & \cdots & 0 \end{bmatrix}$$

if $M < N$

s_k is call the singular value

[Example 1] Perform the SVD for the following matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \\ -1 & 1 \end{bmatrix}$$

(Solution): First, we determine

$B = B^H$ $\quad B^H = (A^H A)^H = A^H (A^H)^H = A^H A = B$ $\quad C = C^H$

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} = \begin{bmatrix} 6 & 2 \\ 2 & 6 \end{bmatrix} \qquad \mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{H}} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix}$$

Then, we perform eigenvector-eigenvalue decomposition for **B** and **C**:

$$\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathbf{H}} \quad \text{where} \quad \mathbf{V} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} 8 & 0 \\ 0 & 4 \end{bmatrix}$$

$$\mathbf{C} = \tilde{\mathbf{U}}\boldsymbol{\Omega}\tilde{\mathbf{U}}^{\mathbf{H}}$$

where $\tilde{\mathbf{U}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$ $\quad \boldsymbol{\Omega} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Note: The eigenvectors should be (i) normalized and (ii) sorted according to the magnitudes of the eigenvalues.

Then,

$$\mathbf{S_1} = \tilde{\mathbf{U}}^{\mathbf{H}}\mathbf{AV} = \begin{bmatrix} \sqrt{8} & 0 \\ 0 & -2 \\ 0 & 0 \end{bmatrix}$$

Then,

$$\mathbf{S} = \begin{bmatrix} |\sqrt{8}| & 0 \\ 0 & |-2| \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{8} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}$$

Since $S_1[2, 2] < 0$, we change the sign of the 2[nd] column of $\tilde{\mathbf{U}}$ and obtain

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Note!

Therefore,

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$ where

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} \sqrt{8} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \quad \mathbf{V} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Note that

$$\mathbf{A} = s_1 \mathbf{u_1} \mathbf{v_1^H} + s_2 \mathbf{u_2} \mathbf{v_2^H}$$

$$\mathbf{A} = \sqrt{8}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} + 2\begin{bmatrix} 0 \\ -1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$$

principal component

minor component

$$\begin{bmatrix} 2 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 \\ -1 & 1 \\ -1 & 1 \end{bmatrix}$$

(Note):

(1) In fact, the eigenvalues of **B** and **C** has a close relation to the $\boxed{\text{singular values}}$ of **A**.

S對角線上值

$$\mathbf{S}^{\mathbf{H}}\mathbf{S} = \mathbf{D} \qquad \mathbf{S}\mathbf{S}^{\mathbf{H}} = \boldsymbol{\Omega}$$

$$S^2[n,n] = D[n,n] = \Omega[n,n]$$

Since

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} = \mathbf{V}\mathbf{S}^{\mathbf{H}}\cancel{\mathbf{U}}^{\mathbf{H}}\cancel{\mathbf{U}}\mathbf{S}\mathbf{V}^{\mathbf{H}} = \mathbf{V}\mathbf{S}^{\mathbf{H}}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

eigenvalue matrix of B

(Note):

(2) Even when $M = N$ (i.e., **A** is a square matrix), the SVD may not be the same as the eigenvector-eigenvalue decomposition.

For the SVD, **U** and **V** are both orthonormal matrices and the singular values are non-negative.

However, for a square matrix, the eigenvectors may not be orthogonal and the eigenvalues can be negative (even complex).

(3) Moreover, since **U** and **V** are usually different and $\mathbf{V}^{\mathbf{H}} \neq \mathbf{U}^{-1}$, one cannot use the SVD to compute the power of a matrix.

[Example 2] Determine the eigenvector-eigenvalue decomposition and the SVD of **A**.

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix}$$

(Solution): The eigenvalues of **A** are 2 and -1.

The eigenvectors corresponding to 2 is $[1\ 0]^T$
The eigenvectors corresponding to -1 is $[1, 3]^T$

Therefore,

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1/3 \\ 0 & 1/3 \end{bmatrix}$$

To perform SVD for **A**,

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} \qquad \mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{H}} = \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{B} = \underset{\mathbf{V}}{\begin{bmatrix} 0.8507 & 0.5257 \\ -0.5257 & 0.8507 \end{bmatrix}} \underset{\mathbf{D}}{\begin{bmatrix} 5.2361 & 0 \\ 0 & 0.7639 \end{bmatrix}} \underset{\mathbf{V}^{\mathbf{H}}}{\begin{bmatrix} 0.8507 & -0.5257 \\ 0.5257 & 0.8507 \end{bmatrix}}$$

$$\mathbf{C} = \underset{\tilde{\mathbf{U}}}{\begin{bmatrix} 0.9732 & -0.2298 \\ 0.2298 & 0.9732 \end{bmatrix}} \underset{\mathbf{\Omega}}{\begin{bmatrix} 5.2361 & 0 \\ 0 & 0.7639 \end{bmatrix}} \underset{\tilde{\mathbf{U}}^{\mathbf{H}}}{\begin{bmatrix} 0.9732 & 0.2298 \\ -0.2298 & 0.9732 \end{bmatrix}}$$

change sign since −0.874 < 0

$$\begin{bmatrix} 0.9732 & -0.2298 \\ 0.2298 & 0.9732 \end{bmatrix}^{\mathbf{H}} \mathbf{A} \begin{bmatrix} 0.8507 & 0.5257 \\ -0.5257 & 0.8507 \end{bmatrix} = \begin{bmatrix} 2.2882 & 0 \\ 0 & -0.8740 \end{bmatrix}$$

change sign

$$\mathbf{A} = \underset{U}{\begin{bmatrix} 0.9732 & 0.2298 \\ 0.2298 & -0.9732 \end{bmatrix}} \underset{S}{\begin{bmatrix} 2.2882 & 0 \\ 0 & 0.8740 \end{bmatrix}} \underset{V^H}{\begin{bmatrix} 0.8507 & -0.5257 \\ 0.5257 & 0.8507 \end{bmatrix}}$$

8.1.2 Generalized Inverse Using the SVD

Suppose that the SVD of **A** is

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

Then the generalized inverse of **A** is

$$\mathbf{A}^{+} = \mathbf{V}\mathbf{S}^{+}\mathbf{U}^{\mathbf{H}}$$

where

$$S^{+}[n,n] = 1/S[n,n] \quad if\ S[n,n] \neq 0$$

$$S^{+}[n,n] = 0 \quad if\ S[n,n] = 0$$

$$size\left(\mathbf{S}^{+}\right) = N \times M \quad if\ size\left(\mathbf{S}\right) = M \times N$$

(Proof):

$$\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \overbrace{\mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}}^{A}\overbrace{\mathbf{V}\mathbf{S}^{+}\mathbf{U}^{\mathbf{H}}}^{A^{+}}\overbrace{\mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}}^{A} = \mathbf{U}\mathbf{S}\mathbf{S}^{+}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

If

$$\mathbf{S}_{\mathbf{2}} = \mathbf{S}^{+}\mathbf{S}$$

then

$$S_2[n,n] = 1 \quad if\ S[n,n] \neq 0 \qquad\qquad S_2[n,n] = 0 \quad if\ S[n,n] = 0$$

Therefore,

$$\mathbf{S} = \mathbf{S}\mathbf{S}^{+}\mathbf{S}$$

$$\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}} = \mathbf{A}$$

(1) $\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}$ is satisfied.

Note: The generalized inverse derived from the SVD is in fact the pseudo inverse since

(2) $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+}$

(3) $\left(\mathbf{A}\mathbf{A}^{+}\right)^{H} = \mathbf{A}\mathbf{A}^{+}$

(4) $\left(\mathbf{A}^{+}\mathbf{A}\right)^{H} = \mathbf{A}^{+}\mathbf{A}$

are all satisfied.

(Try to prove them)

[Example 3] Determine the generalized inverse of the following matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 1 & -2 & 1 \\ -1 & 2 & -1 \end{bmatrix}$$

4×3 matrix
but columns are not
independent

Note: Since the 1st and the 3rd columns are dependent, we cannot use the method of

$$(\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}$$

to determine the generalized inverse. Instead, we should apply the SVD method.

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}} \qquad \mathbf{A}^{+} = \mathbf{V}\mathbf{S}^{+}\mathbf{U}^{\mathbf{H}}$$

(Solution): Since

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 1 & -2 & 1 \\ -1 & 2 & -1 \end{bmatrix}$$

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} = \begin{bmatrix} 10 & 4 & 10 \\ 4 & 16 & 4 \\ 10 & 4 & 10 \end{bmatrix} \qquad \mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{H}} = \begin{bmatrix} 12 & 12 & 0 & 0 \\ 12 & 12 & 0 & 0 \\ 0 & 0 & 6 & -6 \\ 0 & 0 & -6 & 6 \end{bmatrix}$$

$$\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathbf{H}}$$

where $\mathbf{V} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \end{bmatrix}$ $\quad \mathbf{D} = \begin{bmatrix} 24 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

normalized

$$\mathbf{C} = \tilde{\mathbf{U}}\boldsymbol{\Lambda}\tilde{\mathbf{U}}^{\mathbf{H}} \quad \text{where}$$

$$\tilde{\mathbf{U}} = \begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 0 & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix} \qquad \boldsymbol{\Lambda} = \begin{bmatrix} 24 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Then

$$\mathbf{S_1} = \tilde{\mathbf{U}}^{\mathbf{H}}\mathbf{AV} = \begin{bmatrix} \sqrt{24} & 0 & 0 \\ 0 & \sqrt{12} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Since all entries of $\mathbf{S_1}$ are non-negative,

$$\mathbf{S} = \mathbf{S_1} \qquad \mathbf{U} = \tilde{\mathbf{U}}$$

$$\mathbf{A} = \mathbf{USV}^{\mathbf{H}} \qquad \mathbf{U} = \begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 0 & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$$

$$\underset{4\times 3}{\mathbf{S}} = \begin{bmatrix} \sqrt{24} & 0 & 0 \\ 0 & \sqrt{12} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{V} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \end{bmatrix}$$

$$\underline{\mathbf{A}^{+}} = \underline{\mathbf{VS}^{+}\mathbf{U}^{\mathbf{H}}} \qquad \underset{3\times 4}{\mathbf{S}^{+}} = \begin{bmatrix} 1/\sqrt{24} & 0 & 0 & 0 \\ 0 & 1/\sqrt{12} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{A}^{+} = \begin{bmatrix} 1/12 & 1/12 & 1/12 & -1/12 \\ 1/12 & 1/12 & -1/6 & 1/6 \\ 1/12 & 1/12 & 1/12 & -1/12 \end{bmatrix}$$

8.2 Principal Component Analysis

Principal component analysis (PCA) is to find the principal component of a set of data.

Principal components: Corresponding to larger singular values for SVD

[Process of PCA]

Suppose that there is a set of data. The number of data is M and each data has the length of N.

$$\mathbf{x_m} = \begin{bmatrix} x_{m,1} & x_{m,2} & x_{m,3} & \cdots & x_{m,N} \end{bmatrix}$$

$m = 1, 2, \ldots, M$

(In usual, $M >> N$)

(1) First, we subtract each entry by $\bar{x}_n = \dfrac{1}{M}\sum_{m=1}^{M} x_{m,n}$

$$\mathbf{a_m} = \begin{bmatrix} a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,N} \end{bmatrix}$$

where $a_{m,n} = x_{m,n} - \bar{x}_n, \qquad \bar{x}_n = \dfrac{1}{M}\sum_{m=1}^{M} x_{m,n}$

(2) Then, construct an MxN matrix **A**:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \vdots \\ \mathbf{a_M} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

(3) Then, perform SVD for **A**

$$\mathbf{A} = \mathbf{USV^H}$$

second important

(N-1)th important

Nth important

(4) Then

$$\mathbf{A} = s_1\mathbf{u_1}\mathbf{v_1^H} + s_2\mathbf{u_2}\mathbf{v_2^H} + \cdots + s_{N-1}\mathbf{u_{N-1}}\mathbf{v_{N-1}^H} + s_N\mathbf{u_N}\mathbf{v_N^H}$$

most important

where $s_n = S[n,n]$

$$\mathbf{U} = [\mathbf{u_1} \quad \mathbf{u_2} \quad \cdots \quad \mathbf{u_M}], \quad \mathbf{V} = [\mathbf{v_1} \quad \mathbf{v_2} \quad \cdots \quad \mathbf{v_N}]$$

If we want to reduce the component from N to L due to the consideration of compression or feature selection, then

$$\mathbf{A} \cong \mathbf{A_1} = s_1\mathbf{u_1}\mathbf{v_1^H} + s_2\mathbf{u_2}\mathbf{v_2^H} + \cdots + \mathbf{s_L}\mathbf{u_L}\mathbf{v_L^H}$$

Note:

$$\mathbf{x_m} \cong c_{m,1}\mathbf{v_1^H} + c_{m,2}\mathbf{v_2^H} + \cdots + c_{m,L}\mathbf{v_L^H} + \begin{bmatrix}\bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_L\end{bmatrix}$$

where $c_{m,n} = s_n u_n[m]$ m^{th} entry of $\mathbf{u_n}$

$\mathbf{v_1^H}, \mathbf{v_2^H}, \cdots, \mathbf{v_L^H}$ can be viewed as the most important L axes

In general,

$$\mathbf{x} \cong c_1\mathbf{v_1^H} + c_2\mathbf{v_2^H} + \cdots + c_L\mathbf{v_L^H} + \begin{bmatrix}\bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_L\end{bmatrix}$$

$$c_n \in (-\infty, \infty)$$

Main Applications of the PCA

(1) Dimensionality reduction (i.e., feature selection) for pattern recognition and machine learning

(2) Data compression

(3) Data mining

(4) Identifying the principal axis of an object in an image

(5) Line approximation

Example of PCA

3. 在處理二維數據時，有種方法是將數據垂直投影到某一直線，並以該直線為數線，進而了解投影點所成一維數據的變異。下圖的一組二維數據，試問投影到哪一選項的直線，所得之一維投影數據的變異數會是最小？

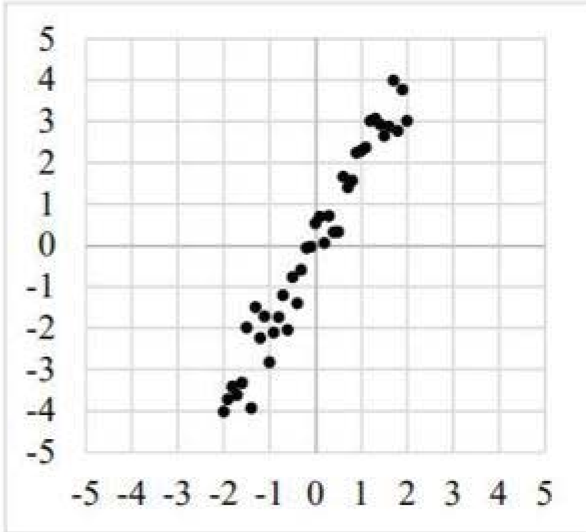
(1) $y = 2x$

(2) $y = -2x$

(3) $y = -x$

(4) $y = \frac{x}{2}$

(5) $y = -\frac{x}{2}$

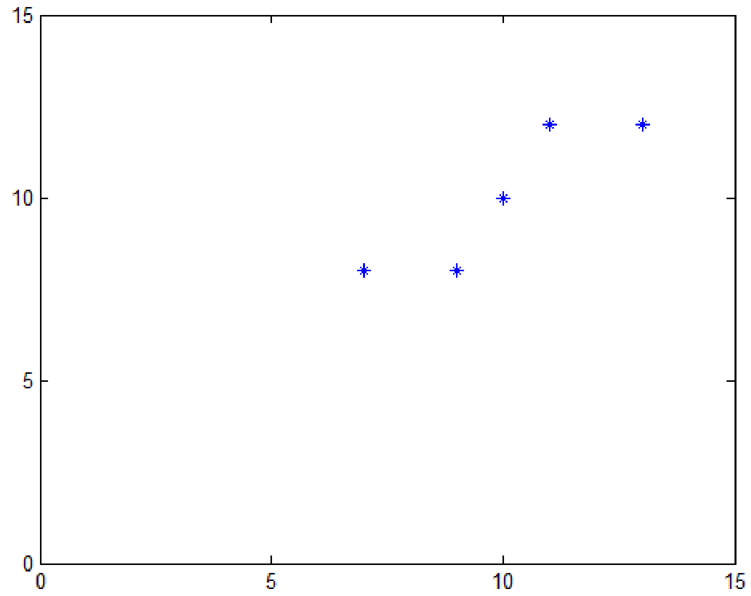


From 2022 大考中心官網

[Example 1] Suppose that there are 5 points in a 2-D space and their coordinates are

$$(7,8),\ (9,8),\ (10, 10),\ (11,12),\ (13,12)$$

Try to find a line that can approximate these points.



(Note): $M = 5$, $N = 2$

(Solution):

First, since the mean of these 5 points is

$$(10, 10),$$

we subtract these points by (10, 10) and obtain

$$(-3, -2),\ (-1\ -2),\ (0, 0),\ (1, 2),\ (3, 2)$$

(-3, -2), (-1 -2), (0, 0), (1, 2), (3, 2)

Then, we construct a 5x2 matrix **A**:

$$\mathbf{A} = \begin{bmatrix} -3 & -2 \\ -1 & -2 \\ 0 & 0 \\ 1 & 2 \\ 3 & 2 \end{bmatrix}$$

$US = [\, s_1 u_1 \;\vdots\; s_2 u_2 \,]$

$A = s_1 u_1 v_1^H + s_2 u_2 v_2^H$

Then, we perform SVD for **A**:

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

$$\mathbf{U} = \begin{array}{c} \begin{array}{ccccc} u_1 & u_2 & u_3 & u_4 & u_5 \end{array} \\ \begin{bmatrix} -0.6116 & 0.3549 & 0 & 0.0393 & 0.7060 \\ -0.3549 & -0.6116 & 0 & 0.7060 & -0.0393 \\ 0 & 0 & 1 & 0 & 0 \\ 0.3549 & 0.6116 & 0 & 0.7060 & -0.0393 \\ 0.6116 & -0.3549 & 0 & 0.0393 & 0.7060 \end{bmatrix} \end{array}$$

$$\mathbf{S} = \begin{bmatrix} \overset{s_1}{5.8416} & 0 \\ 0 & 1.3695^{\,s_2} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{V} = \begin{array}{c} \begin{array}{cc} v_1 & v_2 \end{array} \\ \begin{bmatrix} 0.7497 & -0.6618 \\ 0.6618 & 0.7497 \end{bmatrix} \end{array}$$

Then, **A** can be expanded by

$$\mathbf{A} = 5.8416\begin{bmatrix} -0.6116 \\ -0.3549 \\ 0 \\ 0.3549 \\ 0.6116 \end{bmatrix}\begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix} + 1.3695\begin{bmatrix} 0.3549 \\ -0.6116 \\ 0 \\ 0.6116 \\ -0.3549 \end{bmatrix}\begin{bmatrix} -0.6618 & 0.7497 \end{bmatrix}$$

principal component

secondary component

Therefore,

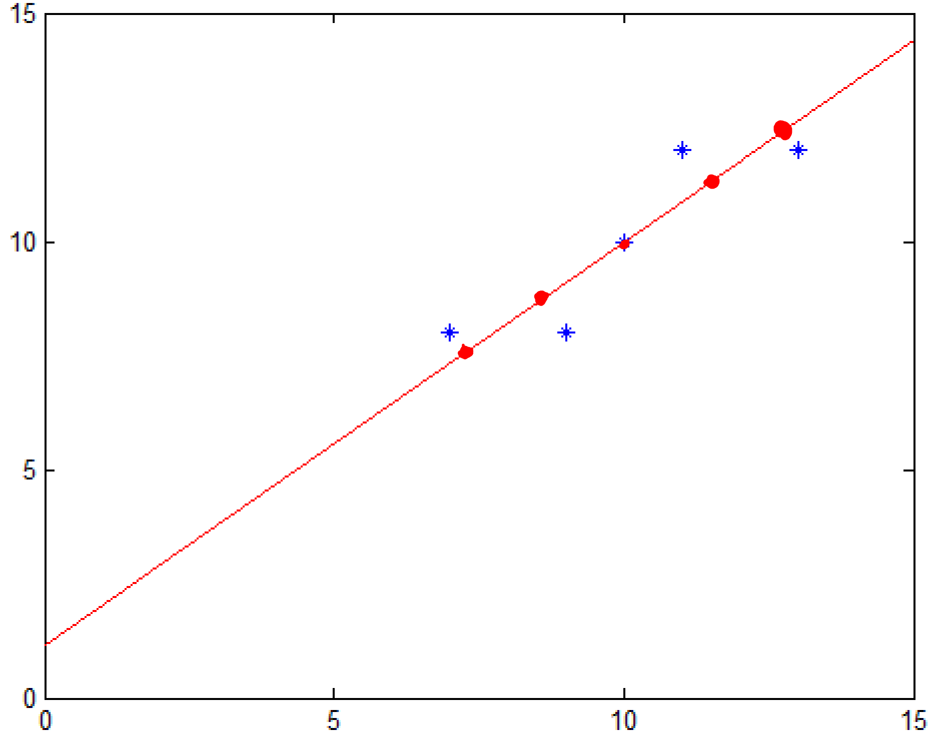
$$\mathbf{A} \cong 5.8416\begin{bmatrix} -0.6116 \\ -0.3549 \\ 0 \\ 0.3549 \\ 0.6116 \end{bmatrix}\begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix} = \begin{bmatrix} -3.5726 \\ -2.0733 \\ 0 \\ 2.0733 \\ 3.5726 \end{bmatrix}\begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix}$$

$$\begin{bmatrix} 7 & 8 \\ 9 & 8 \\ 0 & 0 \\ 11 & 12 \\ 13 & 12 \end{bmatrix} \cong \begin{bmatrix} -3.5726 \\ -2.0733 \\ 0 \\ 2.0733 \\ 3.5726 \end{bmatrix} \begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix} + \begin{bmatrix} 10 & 10 \\ 10 & 10 \\ 10 & 10 \\ 10 & 10 \\ 10 & 10 \end{bmatrix}$$

Approximation line:

$$\begin{bmatrix} 10 & 10 \end{bmatrix} + c \begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix}$$

$c \in (-\infty, \infty)$



[Simplification for Computation]

Suppose that we only want to find the most important L axes of the data. (It is usually the case for practical applications).

If M is very large, then the MxM matrix **U is unnecessary to be computed**. One only has to perform eigenvector-eigenvalue decomposition for **B** and obtain the NxN matrix **V**:

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} \qquad\qquad \mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$$

If $D[n, n]$ is larger than other diagonal entries of **D**, then the nth column of **V** is the principal axis.

附錄十一 Some Common Mathematical Notations

(1) Commutator

$$[\mathbf{A},\mathbf{B}] = \mathbf{AB} - \mathbf{BA}$$

(2) Trace

$$tr(\mathbf{A}) = \sum_{n=1}^{N} A(n,n)$$

(3) Bras and Kets Notations

$$\langle \mathbf{A} \mid \mathbf{B} \rangle = \underbrace{\begin{bmatrix} \mathbf{a}_1^* & \mathbf{a}_2^* & \cdots & \mathbf{a}_N^* \end{bmatrix}}_{\mathbf{A}^H} \underbrace{\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_N \end{bmatrix}}_{\mathbf{B}}$$

A and **B** are column vectors.

$$\langle \mathbf{A} | = \begin{bmatrix} \mathbf{a}_1^* & \mathbf{a}_2^* & \cdots & \mathbf{a}_N^* \end{bmatrix} \qquad | \mathbf{B} \rangle = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_N \end{bmatrix}$$

(4) sup: supremum (the least upper bound，上確界)

$$\sup\{x \mid 1 < x < 2\} = 2$$

$$\sup\left\{(-1)^n - 1/n \mid n \in N\right\} = 1$$

(5) inf: infimum (the greatest lower bound，下確界)

$$\inf\{x \mid 1 < x < 2\} = 1$$

$$\inf\left\{e^{-x} \mid x \in R\right\} = 0$$

(6) card: the number of elements in a set

$$card(\{x, y\}) = 2$$

$$card\left(\left\{x^2, y^2, xy, x, y, 1\right\}\right) = 6$$