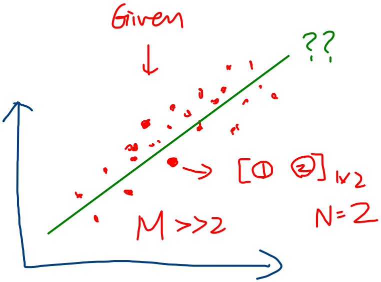
Outline

1. Motivations
2. Jordan Canonical Form
   - Definition and Examples
   - The Integer Power of a Matrix
3. Singular Value Decomposition (SVD)
   - Definition and Properties
   - Matrix Norms and SVD
4. Principal Component Analysis (PCA)



$y = ax + b$

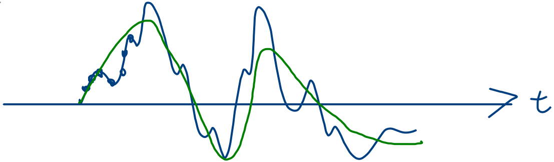
The Data Vectors

- Consider a set of data vectors (row vectors)

$$\mathbf{x}_m = \begin{bmatrix} x_{m,1} & x_{m,2} & x_{m,3} & \dots & x_{m,N} \end{bmatrix}_{1 \times N}, \tag{87}$$

for $m = 1, 2, \dots \boxed{M}$.

- The number of data vectors: M
- The length of a data vector: N
- Usually $M \gg N$.
- Applications
  - Audio signals
  - Images
  - Communication signals
  - Array signal processing (linear arrays or planar arrays)



Mean Subtraction

- The mean vector $\overline{\mathbf{x}}$ (as a row vector) is

$$\overline{\mathbf{x}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{x}_m. \tag{88}$$

$\underline{x}_1 =$

$\underline{x}_2 =$

$\vdots$

$\underline{x}_M =$

average

$\overline{\underline{x}} =$ $1 \times N$

- The new data vector $\mathbf{a}_m$ after subtracting the mean vector from $\mathbf{x}_m$

$$\mathbf{a}_m \triangleq \mathbf{x}_m - \overline{\mathbf{x}}. \tag{89}$$

$m = 1 \sim M$

↳ row vectors

$1 \times N$

The Data Matrix

$a_m = \boxed{}$ $1 \times N$

$m = 1 \sim M$

$A = \boxed{\; \cdot \quad \cdot \; \cdots \;}$

- The data matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$

(89)

$$\mathbf{A} \triangleq \begin{bmatrix} \boxed{\phantom{\mathbf{a}_1}} \\ \mathbf{a}_1 \\ \mathbf{a}_2 \boxed{} \\ \vdots \\ \mathbf{a}_M \boxed{} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 - \overline{\mathbf{x}} \\ \mathbf{x}_2 - \overline{\mathbf{x}} \\ \vdots \\ \mathbf{x}_M - \overline{\mathbf{x}} \end{bmatrix}. \tag{90}$$

$M \times N$

- The data vector $\mathbf{x}_m$ can be expressed as

to be compressed

$$\mathbf{x}_m = \mathbf{e}_m^{\mathsf{T}} \mathbf{A} + \overline{\mathbf{x}}, \tag{91}$$

where $\mathbf{e}_m \in \mathbb{C}^M$ satisfies

$$[\mathbf{e}_m]_i = \begin{cases} 1 & \text{if } i = m, \\ 0 & \text{if } i \neq m. \end{cases} \tag{92}$$

$$e_m = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow m^{th}$$

$$\underline{\underline{A}} = \begin{bmatrix} \underline{a}_1 \\ \vdots \\ \underline{a}_M \end{bmatrix}$$

$$(*) \quad \boxed{\underline{a}_1}_{\,1\times N} = \boxed{\begin{matrix} 1 & 0 & 0 & \cdots & 0 \end{matrix}}_{\,1\times M} \boxed{\underline{\underline{A}}}_{\,M\times N}$$

The m^{th}

$$\underbrace{\underline{x}_m - \underline{\bar{x}}} = \underline{a}_m = \boxed{\begin{matrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{matrix}} \underline{\underline{A}} = \underbrace{\underline{e}_m^T \underline{\underline{A}}}$$

SVD of $\mathbf{A}$

$A = \begin{bmatrix} - & - & - \\ - & \times & \zeta \\ & \times & \times \end{bmatrix}$

- According to Page 37, the SVD of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{H}} \quad (93)$$

components

$$= \sum_{i=1}^{N} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}} \quad (94)$$

"N terms"

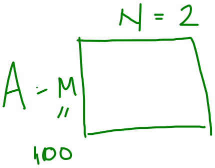
$$= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathsf{H}} + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^{\mathsf{H}} + \sigma_3 \mathbf{u}_3 \mathbf{v}_3^{\mathsf{H}} + \cdots + \sigma_N \mathbf{u}_N \mathbf{v}_N^{\mathsf{H}}. \quad (95)$$

- The singular values satisfy

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_N \geq 0. \quad (96)$$

- The ith component of $\mathbf{A}$ is $\sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}}$.

Dimensionality Reduction (1/2)



- We approximate the matrix $\mathbf{A}$ by L components:

$$\widehat{\mathbf{A}} \triangleq \sum_{i=1}^{L} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}} \tag{97}$$

$$= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathsf{H}} + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^{\mathsf{H}} + \sigma_3 \mathbf{u}_3 \mathbf{v}_3^{\mathsf{H}} + \cdots + \sigma_L \mathbf{u}_L \mathbf{v}_L^{\mathsf{H}}. \tag{98}$$

- Dimensional reduction: $L \leq N$.

$L = 1$: $\hat{A} = \sigma_1 \, u_1 \, \boxed{v_1^H}$

1x1 100x1 1x2

σ_1 : 1 number

u_1 : 100 numbers

v_1 : 2 numbers

Dimensionality Reduction (2/2)

- According to (91) and (97), we define the approximated data vectors

$$\widehat{\mathbf{x}}_m \triangleq \mathbf{e}_m^{\mathsf{T}}\widehat{\mathbf{A}} + \overline{\mathbf{x}} = \left(\sum_{i=1}^{L} \overset{\text{scalar}}{\boxed{\sigma_i\left(\mathbf{e}_m^{\mathsf{T}}\mathbf{u}_i\right)}}\overset{\text{row vector.}}{\boxed{\mathbf{v}_i^{\mathsf{H}}}}\right) + \overline{\mathbf{x}}. \quad (99)$$

M×1

  - $\boxed{\mathbf{e}_m^{\mathsf{T}}\mathbf{u}_i}$ is the mth entry of $\mathbf{u}_i$.
  - $\sigma_i\left(\mathbf{e}_m^{\mathsf{T}}\mathbf{u}_i\right)$ is the combination coefficient.
  - The set $\{\mathbf{v}_1^{\mathsf{H}}, \mathbf{v}_2^{\mathsf{H}}, \ldots, \mathbf{v}_L^{\mathsf{H}}\}$ contains the axes.
- A general form of the approximated data vectors is

$$\left(\sum_{i=1}^{L} c_i\mathbf{v}_i^{\mathsf{H}}\right) + \overline{\mathbf{x}}, \quad (100)$$

where $c_i \in \mathbb{C}$ for $i = 1, 2, \ldots L$.

An Example of the PCA (1/4)

Problem

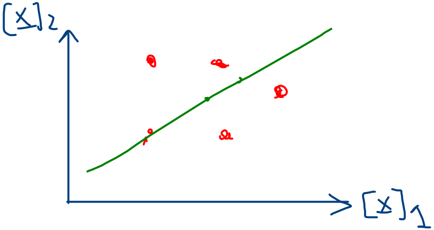
Use the PCA with $L = 1$ to find a regression line that approximates the points in $\mathbb{R}^2$

$$\mathbf{x}_1 = \begin{bmatrix} 7 & 8 \end{bmatrix}_{1\times 2}, \quad \mathbf{x}_2 = \begin{bmatrix} 9 & 8 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 10 & 10 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 11 & 12 \end{bmatrix}, \quad \mathbf{x}_5 = \begin{bmatrix} 13 & 12 \end{bmatrix}.$$

We assume that the combination coefficients are real numbers.

- (Solution) The number of data $M = 5$.
- The length of the data vector $N = 2$.
- The mean vector

$c = 0$
$c = 1$
$c =$



$$\overline{\mathbf{x}} = \begin{bmatrix} 10 & 10 \end{bmatrix}.$$

An Example of the PCA (2/4)

- The new data vectors

$$\underline{\mathbf{a}_1} = \begin{bmatrix} -3 & -2 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} -1 & -2 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 & 0 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 1 & 2 \end{bmatrix}, \quad \underline{\mathbf{a}_5} = \begin{bmatrix} 3 & 2 \end{bmatrix}.$$

- The data matrix $\mathbf{A}$ and its SVD

$$\mathbf{A} = \begin{array}{cl} \begin{bmatrix} -3 & -2 \\ -1 & -2 \\ 0 & 0 \\ 1 & 2 \\ 3 & 2 \end{bmatrix}. & \begin{array}{l} \underline{a_1} \\ \underline{a_2} \\ \\ \\ \underline{a_5} \end{array} \end{array} \tag{101}$$

5×2

An Example of the PCA (3/4)

- The SVD of $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{H}}$, where

$$
\mathbf{U} = \begin{array}{c} \begin{bmatrix} -0.6116 & 0.3549 & 0 & 0.0393 & 0.7060 \\ -0.3549 & -0.6116 & 0 & 0.7060 & -0.0393 \\ 0 & 0 & 1.0000 & 0 & 0 \\ 0.3549 & 0.6116 & 0 & 0.7060 & -0.0393 \\ 0.6116 & -0.3549 & 0 & 0.0393 & 0.7060 \end{bmatrix} \\ \begin{array}{ccccc} \underline{u}_1 & \quad\underline{u}_2 & \quad\underline{u}_3 & \quad\underline{u}_4 & \quad\underline{u}_5 \end{array} \end{array}, \quad \mathbf{\Sigma} = \begin{bmatrix} \boxed{5.8416} & 0 \\ 0 & \boxed{1.3695} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},
$$

$\sigma_1 = 5.8416$, $\sigma_2 = 1.3695$

$$
\mathbf{V} = \begin{array}{c} \begin{bmatrix} 0.7497 & -0.6618 \\ 0.6618 & 0.7497 \end{bmatrix} \\ \begin{array}{cc} \underline{v}_1 & \quad\underline{v}_2 \end{array} \end{array}.
$$

An Example of the PCA (4/4)

- For $L = 1$ in (97), we obtain

$$\widehat{\mathbf{A}} = \underbrace{(5.8416)}_{\sigma_1} \underbrace{\begin{bmatrix} -0.6116 \\ -0.3549 \\ 0 \\ 0.3549 \\ 0.6116 \end{bmatrix}}_{\mathbf{u}_1} \underbrace{\begin{bmatrix} 0.7497 & -0.6618 \end{bmatrix}}_{\mathbf{v}_1^{\mathsf{H}}}.$$

- According to (100) and page 61, an approximation of the data points is

$$\underbrace{\begin{bmatrix} 10 & 10 \end{bmatrix}}_{\bar{x}} + c \underbrace{\begin{bmatrix} 0.7497 & -0.6618 \end{bmatrix}}_{v_1^H}, \rightarrow$$

$[\underline{x}]_1 = 10 + 0.7497c$

$[\underline{x}]_2 = 10 - 0.6618c$

where $\boxed{c \in \mathbb{R}}$.

Selected Topics in Engineering Mathematics: Least Squares Problems

LS

Chun-Lin Liu (劉俊麟)

Department of Electrical Engineering
Graduate Institute of Communication Engineering
National Taiwan University

May 28, 2024

Reference

1. R. A. Horn and C. R. Johnson, Matrix Analysis, 2nd ed., New York: Cambridge University Press, 2013.
[HJ2013]
2. G. H. Golub and C. F. Van Loan, Matrix Computations, 4th ed., Baltimore: The Johns Hopkins University Press, 2013.
[GVL2013]
3. J.-J. Ding. (2023). Selected Topics in Engineering Mathematics [PowerPoint slides].

Outline

Motivation

$N \times 1$

- Find a vector $\mathbf{x} \in \mathbb{C}^N$ such that

$$\mathbf{A}\mathbf{x} = \mathbf{b}. \tag{1}$$

  - The data matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$ is given.
  - The observation vector $\mathbf{b} \in \mathbb{C}^{M}$ is given. $M \times 1$
- The number of equations is $\boxed{M}$.
- The number of unknowns is N.
- Underdetermined systems: $M < N$
- Overdetermined systems: $M > N$

$M<N$

$M=N$

$M>N$

Questions

- How many solutions to (1)?

Examples of (1)

rank($\underline{\underline{A}}$)

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} \overset{1}{x_1} \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Underdetermined Systems

$$\underbrace{\begin{bmatrix} 1 & 2 \end{bmatrix}}_{\mathbf{A}\ 1\times 2} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}\ 2\times 1} = \underbrace{\begin{bmatrix} 0 \end{bmatrix}}_{\mathbf{b}\ 1\times 1}. \tag{2}$$

The solutions to (2) are

$$\mathbf{x} = \begin{bmatrix} -2c \\ c \end{bmatrix},$$

→ infinitely many solution

where $c \in \mathbb{C}$.

Overdetermined Systems

$$\underbrace{\begin{bmatrix} 1 \\ 3 \end{bmatrix}}_{\substack{\mathbf{A} \\ 2\times 1}} \underbrace{\begin{bmatrix} x_1 \end{bmatrix}}_{\mathbf{x}\ 1\times 1} = \underbrace{\begin{bmatrix} 3 \\ 2 \end{bmatrix}}_{\mathbf{b}\ 2\times 1}. \tag{3}$$

↗ $\frac{1}{3}$

There are no solutions to (3).

Usually, an overdetermine system has no exact solution.

The Least Squares Problem (1/2)

$A\underline{x} = \underline{b}$

- We aim to find a solution such that

$x = ?$

$$\mathbf{A}\mathbf{x} \approx \mathbf{b}. \qquad (4)$$

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

- The vector p-norm measures the proximity of $\mathbf{A}\mathbf{x}$ to $\mathbf{b}$.

$\| \cdot \|_p \in \mathbb{R}$

$3 \geq 2$

$$\|\boxed{\mathbf{A}\mathbf{x} - \mathbf{b}}\|_p, \quad \in \mathbb{R} \qquad (5)$$

where $p \in [1, \infty)$.

$p = ?$ $\begin{cases} 1 \\ 2 \\ \infty \end{cases}$

The Least Squares Problem (2/2)

The Least Squares (LS) Problem $(p = 2)$

minimize

$$\min_{\mathbf{x}\in\mathbb{C}^N} \overbrace{\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2}^{\text{objective function}} \rightarrow \ell_2 \text{ norm} \qquad (6)$$

- The LS problem (6) is tractable for two reasons

convex function ($\|\cdot\|_p \quad p \in [1, \infty)$

  1. The solutions to (6) can be found readily.
     - Completion of squares
     - The (complex) derivatives of the objective function $\frac{\partial}{\partial x^*}$

SVD

$U \Sigma V^H$

  2. The ℓ_2 norm is invariant under unitary transformations. Namely,

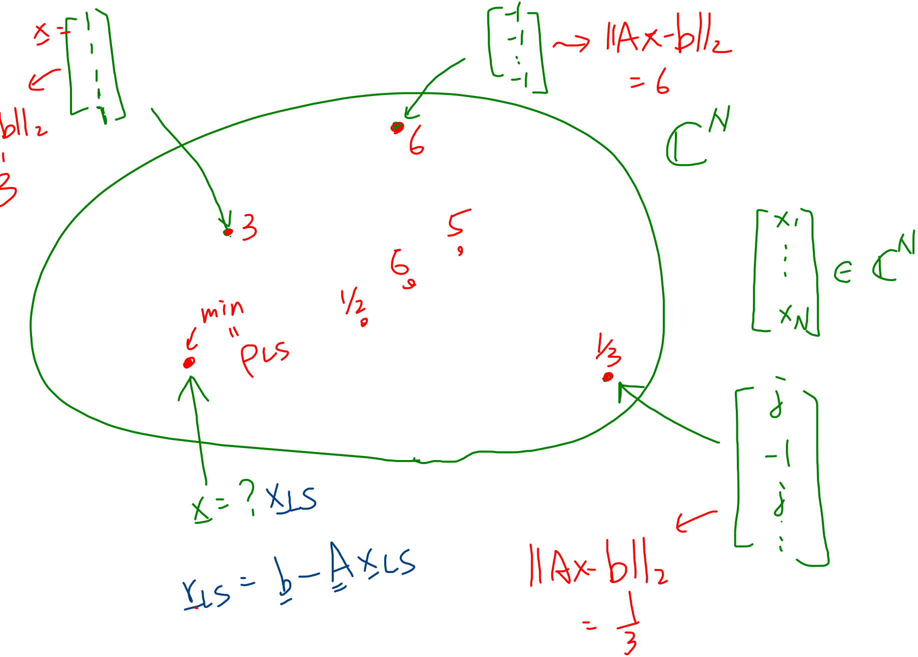
$$\underline{\|\mathbf{U}\mathbf{v}\|_2 = \|\mathbf{v}\|_2}\,, \qquad (7)$$

v $\quad$ Uv

for a unitary matrix $\mathbf{U}$.

LS:

$$\min_{\underline{x} \in \mathbb{C}^N} \|\underline{\underline{A}}\underline{x} - \underline{b}\|_2$$



$$\underline{r}_{LS} = \underline{b} - \underline{\underline{A}}\,\underline{x}_{LS}$$

Outline

$$\| \underline{\underline{A}} \underline{x} - \underline{b} \|_2$$

rank.

The LS Solution(s)

$A = \begin{pmatrix} \vdots \end{pmatrix}$

$$\min_{\mathbf{x}\in\mathbb{C}^N} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \tag{8}$$

- Let $\mathbf{x}_{\mathrm{LS}}$ be a solution to the LS problem (6).

Questions

- Does $\mathbf{x}_{\mathrm{LS}}$ exist?
- How do we find $\mathbf{x}_{\mathrm{LS}}$?
- Is the LS solution $\mathbf{x}_{\mathrm{LS}}$ unique?

The Normal Equation

x_{LS}

$A = M \times N$

$\text{rank}(A) = N$

Normal Equation

If **A has full column rank**, then there is a unique LS solution $\mathbf{x}_{\mathrm{LS}}$, and it satisfies

A

$$\boxed{\mathbf{A}^{\mathsf{H}}\mathbf{A}\mathbf{x}_{\mathrm{LS}}} = \boxed{\mathbf{A}^{\mathsf{H}}\mathbf{b}}. \tag{9}$$

- See Section 5.3.1 in [GVL2013] for the complete arguments
- The minimum residual $\mathbf{r}_{\mathrm{LS}}$

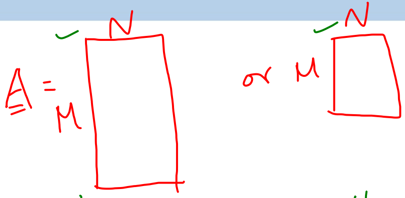
$$\underline{\underline{A}}\,\underline{x} \underset{=}{\approx} \underline{b}$$

$$\mathbf{r}_{\mathrm{LS}} \triangleq \mathbf{b} - \mathbf{A}\mathbf{x}_{\mathrm{LS}}. \tag{10}$$

- The size of $\mathbf{r}_{\mathrm{LS}}$

$$\rho_{\mathrm{LS}} \triangleq \left\|\mathbf{A}\mathbf{x}_{\mathrm{LS}} - \mathbf{b}\right\|_2 . \tag{11}$$

Remarks on the Normal Equation



- Assume that $\mathbf{A} \in \mathbb{C}^{M\times N}$ and $M \geq N$.
- If $\mathbf{A}$ has full column rank, then
  - $\operatorname{rank}(\mathbf{A}) = N$.
  - $\operatorname{rank}(\boxed{\mathbf{A}^{\mathsf{H}}\mathbf{A}}) = N$. (full rank)
  - $\mathbf{A}^{\mathsf{H}}\mathbf{A}$ is invertible.

- If $\mathbf{A}$ has full column rank, then the LS solution can be uniquely found by

$$\mathbf{x}_{\mathrm{LS}} \triangleq \boxed{\left(\mathbf{A}^{\mathsf{H}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathsf{H}}}\mathbf{b}. \tag{12}$$

A^+

- Interpretations of $\mathbf{x}_{\mathrm{LS}}$
  - Wiener-Hopf equation in Adaptive Signal Processing
  - Singular values and singular vectors of $\mathbf{A}$

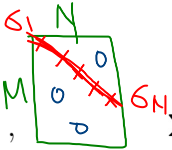
σ_i SVD u_i v_i

The LS Solution and the SVD (1/4)

- We assume that $\text{rank}(\mathbf{A}) = N$. (full column rank)
- The SVD of $\mathbf{A}$ is denoted by

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}} = \sum_{i=1}^{N} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}}. \tag{13}$$

- The matrix $\boldsymbol{\Sigma}$ is



$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_N \\ \mathbf{0}_{(M-N)\times N} \end{bmatrix}, \qquad \boldsymbol{\Sigma}_N = \text{diag}\left(\sigma_1, \sigma_2, \ldots, \sigma_N\right). \tag{14}$$

- The singular values satisfy $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N \boxed{> 0.}$
- The unitary matrices $\mathbf{U}$ and $\mathbf{V}$ comprise left and right singular vectors.

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_M \end{bmatrix}, \qquad \mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_N \end{bmatrix}. \tag{15}$$

The LS Solution and the SVD (2/4)

- The unitary matrices $\mathbf{U}$ and $\mathbf{V}$ satisfy

$$\mathbf{U}^{\mathsf{H}}\mathbf{U} = \mathbf{I}_M, \qquad \mathbf{V}^{\mathsf{H}}\mathbf{V} = \mathbf{I}_N. \tag{16}$$

- Substituting (13) into (12) leads to

$$\mathbf{x}_{\mathrm{LS}} = \left(\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}\right)^{\mathsf{H}} \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}\right)\right)^{-1} \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}\right)^{\mathsf{H}} \mathbf{b} \tag{17}$$

$$= \left(\mathbf{V}\boldsymbol{\Sigma}^{\mathsf{H}}\overbrace{\mathbf{U}^{\mathsf{H}}\mathbf{U}}^{I}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}\right)^{-1} \mathbf{V}\boldsymbol{\Sigma}^{\mathsf{H}}\mathbf{U}^{\mathsf{H}}\mathbf{b} \tag{18}$$

$$= \mathbf{V}\left(\boldsymbol{\Sigma}^{\mathsf{H}}\boldsymbol{\Sigma}\right)^{-1} \overbrace{\mathbf{V}^{\mathsf{H}}\mathbf{V}}^{I}\boldsymbol{\Sigma}^{\mathsf{H}}\mathbf{U}^{\mathsf{H}}\mathbf{b} \tag{19}$$

$$= \mathbf{V}\left(\boldsymbol{\Sigma}^{\mathsf{H}}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Sigma}^{\mathsf{H}}\boxed{\mathbf{U}^{\mathsf{H}}\mathbf{b}.} \tag{20}$$

The LS Solution and the SVD (3/4)

- From (14), the matrix associated with $\mathbf{\Sigma}$ can be expressed as

$$\left(\mathbf{\Sigma}^{\mathsf{H}}\mathbf{\Sigma}\right)^{-1}\mathbf{\Sigma}^{\mathsf{H}} = \left(\begin{bmatrix}\mathbf{\Sigma}_N \\ \mathbf{0}_{(M-N)\times N}\end{bmatrix}^{\mathsf{H}}\begin{bmatrix}\mathbf{\Sigma}_N \\ \mathbf{0}_{(M-N)\times N}\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{\Sigma}_N \\ \mathbf{0}_{(M-N)\times N}\end{bmatrix}^{\mathsf{H}} \tag{21}$$

$$= \left(\mathbf{\Sigma}_N^{\mathsf{H}}\mathbf{\Sigma}_N\right)^{-1}\begin{bmatrix}\mathbf{\Sigma}_N^{\mathsf{H}} & \mathbf{0}_{N\times(M-N)}\end{bmatrix} \tag{22}$$

$$= \begin{bmatrix}\mathbf{\Sigma}_N^{-1} & \mathbf{0}_{N\times(M-N)}\end{bmatrix} \tag{23}$$

$$= \begin{bmatrix} \sigma_1^{-1} & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^{-1} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N^{-1} & 0 & \dots & 0 \end{bmatrix}. \tag{24}$$

The LS Solution and the SVD (4/4)

Gram-Schmidt process

$\underline{w} = \alpha_1 \underline{v}_1 + \alpha_2 \underline{v}_2$

- Substituting (24) and (15) into (20) gives

$$\mathbf{x}_{\mathrm{LS}} = \sum_{i=1}^{N} \frac{\mathbf{u}_i^{\mathsf{H}}\mathbf{b}}{\sigma_i}\mathbf{v}_i, \tag{25}$$

singular values

- $\mathbf{x}_{\mathrm{LS}}$ is a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$.
- Two factors influence the combination coefficients
  1. The inner product $\langle \mathbf{b}, \mathbf{u}_i \rangle \triangleq \mathbf{u}_i^{\mathsf{H}}\mathbf{b}$
  2. The singular value σ_i

If $\underline{v}_1 \perp \underline{v}_2$

$\langle \underline{w}, \underline{v}_1 \rangle$

$\underline{w} \approx \frac{\langle \underline{w}, \underline{v}_1 \rangle}{\langle \underline{v}_1, \underline{v}_1 \rangle} \underline{v}_1 +$

$\alpha_1 = \frac{\langle \underline{w}, \underline{v}_1 \rangle}{\langle \underline{v}_1, \underline{v}_1 \rangle}$

$= \alpha_1 \langle v_1, v_1 \rangle + \alpha_2 \cancel{\langle v_2, v_1 \rangle} = 0$

The Size of the Minimum Residual

- (Exercise) It can be shown that the size of the minimum residual (denoted by ρ_{LS}) satisfies

$\rho_{LS} = \| \underline{\underline{A}} x_{LS} - b \|_2$

$$\rho_{\mathrm{LS}}^2 = \sum_{i=N+1}^{M} \left| \mathbf{u}_i^{\mathsf{H}} \mathbf{b} \right|^2. \tag{26}$$

error

Outline

A full column rank

Motivation

- (The normal equation of LS problems) If $\overset{p}{\boxed{\mathbf{A} \text{ has full column rank}}}$, then there is an unique LS solution $\mathbf{x}_{\mathrm{LS}}$ and

$$\mathbf{A}^{\mathsf{H}}\mathbf{A}\mathbf{x}_{\mathrm{LS}} = \mathbf{A}^{\mathsf{H}}\mathbf{b}. \tag{27}$$

- What if $\mathbf{A}$ is rank-deficient? Namely, $\mathbf{A} \in \mathbb{C}^{M\times N}$, and

$\sim p \to \ldots$

$$\mathrm{rank}(\mathbf{A}) = r < N. \tag{28}$$

$\min_{x \in \mathbb{C}^{N}} \|Ax - b\|_2$

- Logical reasoning:

$$p \to q \qquad \equiv \qquad \sim q \to \sim p \tag{29}$$

Example 1

- We consider the following equations

$$\underbrace{\begin{bmatrix} 1 & 2 \end{bmatrix}}_{\mathbf{A}_{1\times 2}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}_{2\times 1}} = \underbrace{\begin{bmatrix} 1 \end{bmatrix}}_{\mathbf{b}_{1\times 1}}. \tag{30}$$

- The associated LS problem is cast as

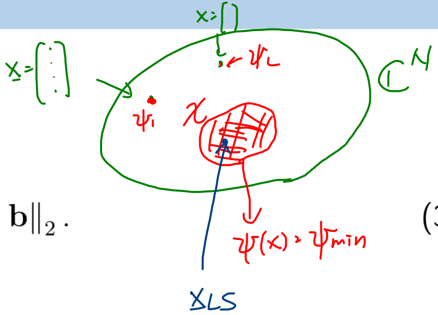
$$\min_{\mathbf{x}\in\mathbb{C}^N} \overbrace{\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2}^{=0} \tag{31}$$

$\psi_{min} = 0$

Observations

- There are infinitely many solutions to (30).
- If $\mathbf{x}^\star$ is a solution to (30), then $\|\mathbf{A}\mathbf{x}^\star - \mathbf{b}\|_2 = 0$.
- The LS problem (31) has an infinite number of solutions.

The Minimum 2-Norm Solution



- We define the objective function

$$\psi(\mathbf{x}) \triangleq \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 . \qquad (32)$$

- The minimum of $\psi(\mathbf{x})$ is denoted by $\psi_{\min}$.
- The set of all minimizers

$$\mathcal{X} \triangleq \{\mathbf{x} \in \mathbb{C}^N \mid \psi(\mathbf{x}) = \psi_{\min}\}. \qquad (33)$$

- The set $\mathcal{X}$ is convex [GVL2013, Section 5.5.1].
- Among the vectors in $\mathcal{X}$, we select the unique element with the minimum 2-norm:

$$\mathbf{x}_{\mathrm{LS}} \triangleq \arg\min_{\mathbf{x}\in\mathcal{X}} \|\mathbf{x}\|_2 \qquad (34)$$

The Rank-Deficient LS Solution with the Minimum 2-Norm

Theorem (Revised from Theorem 5.5.1 in [GVL2013])

Let the SVD of $\mathbf{A}$ *be* $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}} \in \mathbb{C}^{M\times N}$ *with* $\operatorname{rank}(\mathbf{A}) = r$. *The singular vectors satisfy*

$\sigma_1 \sim \sigma_r > 0$

$$\mathbf{U} \triangleq \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_M \end{bmatrix}, \qquad \mathbf{V} \triangleq \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_N \end{bmatrix}. \tag{35}$$

Assume that $\mathbf{b} \in \mathbb{C}^M$. *Then*

$$\mathbf{x}_{\mathrm{LS}} = \sum_{i=1}^{r} \frac{\mathbf{u}_i^{\mathsf{H}}\mathbf{b}}{\sigma_i}\mathbf{v}_i \tag{36}$$

$= A^+ b$

$= (V\Sigma^+ \boxed{U^H) b}$

(25)

minimizes $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$ *and has the smallest 2-norm of all minimizers.*

The LS Solution in Example 1

- We consider the matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 \end{bmatrix}$ in (30).
- The rank of $\mathbf{A}$ is 1.
- The SVD of $\mathbf{A}$

$$\mathbf{u}_1 = 1, \qquad \sigma_1 = \sqrt{5}, \qquad \mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}. \tag{37}$$

- The set of minimizers

real /imag

$Ax = b$

$$\mathcal{X} = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{C}^2 \;\middle|\; x_1 + 2x_2 = 1 \right\}. \tag{38}$$

$1 + j0$

The LS Solution in Example 1

$x_1, x_2 \in \mathbb{C}$

- The rank-deficient LS solution with the minimum 2-norm

$$\mathbf{x}_{\mathrm{LS}} \triangleq \arg\min_{\boxed{\mathbf{x}\in\mathcal{X}}} \|\mathbf{x}\|_2 = \arg\min_{\mathbf{x}\in\mathcal{X}} \sqrt{|x_1|^2 + |x_2|^2} \tag{39}$$

- We decompose the elements x_1 and x_2 into the real and imaginary parts:

$$x_1 = \mathrm{Re}\{x_1\} + \jmath \overset{0}{\mathrm{Im}\{x_1\}}, \tag{40}$$

$$x_2 = \mathrm{Re}\{x_2\} + \jmath \underset{0}{\mathrm{Im}\{x_2\}}. \tag{41}$$

- The LS solution $\mathbf{x}_{\mathrm{LS}}$

$$\mathbf{x}_{\mathrm{LS}} \triangleq \arg\min_{\mathbf{x}\in\mathcal{X}} \sqrt{(\mathrm{Re}\{x_1\})^2 + (\underset{0}{\cancel{\mathrm{Im}\{x_1\}}})^2 + (\mathrm{Re}\{x_2\})^2 + (\underset{0}{\cancel{\mathrm{Im}\{x_2\}}})^2} \tag{42a}$$

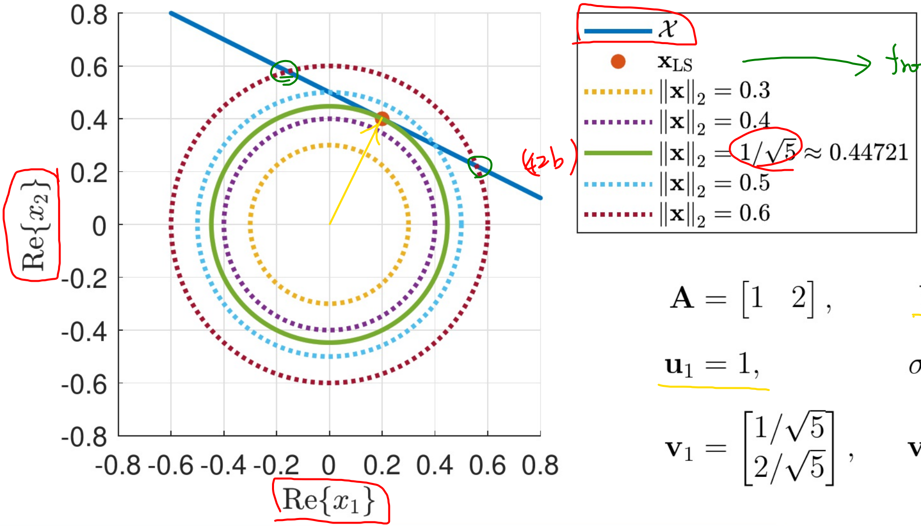
subject to

$$\mathrm{Re}\{x_1\} + 2\mathrm{Re}\{x_2\} = 1, \tag{42b}$$

$$\underset{0}{\cancel{\mathrm{Im}\{x_1\}}} + 2\underset{0}{\cancel{\mathrm{Im}\{x_2\}}} = 0. \tag{42c}$$

Illustration of the LS Solution

(42b) : $\mathrm{Re}\{x_1\} + 2\,\mathrm{Re}\{x_2\} = 1$



from (36)

$$\sum_{i=1}^{r} \frac{u_i^H b}{\sigma_i} v_i$$

$(r = 1)$

$$= \frac{u_1^H b}{\sigma_1} v_1$$

coef 1, $\sqrt{5}$ vec

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \end{bmatrix}, \qquad \mathbf{b} = 1,$$

$$\mathbf{u}_1 = 1, \qquad \sigma_1 = \sqrt{5},$$

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}.$$

Outline

Normal eq. $\rightarrow$ $x_{LS} = (A^H A)^{-1} A^H b$

$$A^H A x = A^H b$$

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

LS $\quad Ax \approx b$

$$x_{LS} = A^{\dagger} b$$

$Ax = b$

If A^{-1} exists, then $x = A^{-1} b$

Pseudo-inverse Using the SVD

$\sigma_1 \sim \sigma_r > 0$

- Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}} \in \mathbb{C}^{M\times N}$ where $\operatorname{rank}(\mathbf{A}) = r \leq \min\{M, N\}$ (c.f. page 21).
- We define a matrix $\boldsymbol{\Sigma}^{\dagger}$ (c.f. page 14)

$$\Sigma = \begin{array}{cc} & N \\ M & \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & \ddots \; 0 \end{bmatrix} \end{array}$$

$$\boldsymbol{\Sigma}^{\dagger} \triangleq \begin{array}{cc} & M \\ N & \begin{bmatrix} \sigma_1^{-1} & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^{-1} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r^{-1} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \end{array} \in \mathbb{C}^{N\times M}. \tag{43}$$

- The pseudo-inverse of $\mathbf{A}$ is defined as

$$\mathbf{A}^{\dagger} \triangleq \mathbf{V}\boldsymbol{\Sigma}^{\dagger}\mathbf{U}^{\mathsf{H}} \in \mathbb{C}^{N\times M}. \tag{44}$$

$$\underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^H = \square_{M \times N}$$

$$\underline{\underline{A}}^{\dagger} = \underline{\underline{V}}\,\underline{\underline{\Sigma}}^{\dagger}\,\underline{\underline{U}}^H = \square_{N \times M}$$

Example of the Pseudo-Inverse

$\underline{\underline{A}}\,\underline{\underline{A}}^{+} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1/5 \\ 2/5 \end{bmatrix}$

$= 1$

- We consider the matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 \end{bmatrix}$ in (30).
- The rank of $\mathbf{A}$ is 1.
- The SVD of $\mathbf{A}$

$$\mathbf{u}_1 = 1, \qquad \sigma_1 = \sqrt{5}, \qquad \mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}. \tag{45}$$

- The pseudo-inverse of $\mathbf{A}$

$\underline{\underline{V}} \quad \underline{\underline{\Sigma}}^{+} \quad \underline{\underline{U}}^{H}$

$$\mathbf{A}^\dagger = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} \sigma_1^{-1} \\ 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \end{bmatrix}^{\mathsf{H}} = \begin{bmatrix} 1/5 \\ 2/5 \end{bmatrix}. \tag{46}$$

$\underline{\underline{A}}^{+}\underline{\underline{A}} = \begin{bmatrix} 1/5 \\ 2/5 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} x & x \\ x & x \end{bmatrix}$

$x_{LS} \overset{?}{=} \underline{\underline{A}}^{+} b$

Properties of the Pseudo-Inverse (1/5)

$[\underline{A}\underline{x} \approx \underline{b}]$

LS: $\min \|\underline{A}\underline{x} - \underline{b}\|_2$

- Let $\mathbf{A} \in \mathbb{C}^{M \times N}$
- Let $\mathbf{A}^\dagger$ be the pseudo-inverse of $\mathbf{A}$
- Let $\mathbf{b} \in \mathbb{C}^{M}$.
- The LS solution $\mathbf{x}_{\mathrm{LS}}$ satisfies

$$\mathbf{x}_{\mathrm{LS}} = \mathbf{A}^\dagger \mathbf{b}. \tag{47}$$

- Remarks
  - Comparison: (25) and (36).
  - Initially, we aim to solve $\mathbf{Ax} = \mathbf{b}$. $\rightsquigarrow$ $x = \underline{A}^{-1} \underline{b}$

Properties of the Pseudo-Inverse (2/5)

- If $\mathrm{rank}(\mathbf{A}) = N$, then

full column rank

(12)

$$\mathbf{A}^{\dagger} = \left(\mathbf{A}^{\mathsf{H}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathsf{H}}. \tag{48}$$

- If $M = N = \mathrm{rank}(\mathbf{A})$, then

full rank

N

A = M

$$\begin{aligned}
\mathbf{A}^{\dagger} &= \left(\mathbf{A}^{\mathsf{H}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathsf{H}} && (49)\\
&= \mathbf{A}^{-1}\left(\cancel{\mathbf{A}^{\mathsf{H}}}\right)^{-1}\cancel{\mathbf{A}^{\mathsf{H}}} && (50)\\
&= \mathbf{A}^{-1}. && (51)
\end{aligned}$$

Properties of the Pseudo-Inverse (3/5)

$A^\dagger = V \Sigma^\dagger U^H$

- The pseudo-inverse $\mathbf{A}^\dagger$ satisfies the four Moore-Penrose conditions:

Inverse A^{-1} (exists)

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

$A^{-1} A = I$

$A A^\dagger \overset{?}{\neq} I$

$$\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}, \tag{52}$$

$$\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger, \tag{53}$$

$$\left(\mathbf{A}\mathbf{A}^\dagger\right)^{\mathsf{H}} = \mathbf{A}\mathbf{A}^\dagger, \tag{54}$$

$$\left(\mathbf{A}^\dagger\mathbf{A}\right)^{\mathsf{H}} = \mathbf{A}^\dagger\mathbf{A}. \tag{55}$$

- (Exercise) Prove the four Moore-Penrose conditions.

Properties of the Pseudo-Inverse (4/5)

- The matrix $\mathbf{A}\mathbf{A}^\dagger$ can be expressed as

$$\mathbf{A}\mathbf{A}^\dagger = \sum_{i=1}^{r} \mathbf{u}_i \mathbf{u}_i^{\mathsf{H}}, \tag{56}$$

  where $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ are the left singular vectors of $\mathbf{A}$.

- The matrix $\mathbf{A}^\dagger\mathbf{A}$ can be expressed as

$$\mathbf{A}^\dagger\mathbf{A} = \sum_{i=1}^{r} \mathbf{v}_i \mathbf{v}_i^{\mathsf{H}}, \tag{57}$$

  where $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are the right singular vectors of $\mathbf{A}$.

Properties of the Pseudo-Inverse (5/5)

- The size of the minimum residual satisfies

$$\rho_{\mathrm{LS}} = \left\| \left(\mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}\right) \mathbf{b} \right\|_2 . \tag{58}$$

Outline

Concluding Remarks

- The LS problem

$$\min_{\mathbf{x}\in\mathbb{C}^N} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2$$

  - Normal equations
  - Full-rank LS
  - Rank-deficient LS
  - Pseudo inverse A^+
- Extensions
  - Weighted least squares (WLS)
  - Total least squares (TLS)
  - Constrained least squares (CLS)
  - Recursive least squares (RLS)

$Ax \approx b$