LINEAR ALGEBRA I

JIN YOUNG CHOI

ECE, SEOUL NATIONAL UNIVERSITY

Outline

basics
- matrix and vector operation
- Gaussian elimination: inverse of matrix
- linear equations

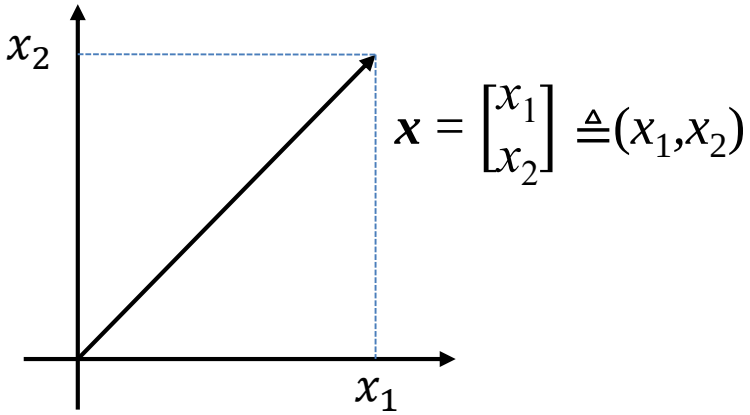
vector space
- def., dimension, basis, subspace
- linear combination, linearly independent
- vector representation

linear transformation
- matrix representation of LT
- null space and range space
- solution set of linear equation
- neural transformation

useful tricks
- vector norm, matrix norm
- eigenvalue and eigenvector
- singular value decomposition
- pseudo inverse
- derivative of vector function

n-tuple Vector

- n-Tuple Vector



- Inner Product, Dot Product

for real field

$$< x, y >= x \cdot y = x^T y = [x_1 \ x_2] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = x_1 y_1 + x_2 y_2$$

for complex field

$$< x, y >= x \cdot y = y^* x = [\bar{y}_1 \ \bar{y}_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \bar{y}_1 + x_2 \bar{y}_2$$

Matrix

- Matrix Notation

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

→ A is an $m \times n$ matrix (m by n matrix)

m: number of rows; n: number of columns

- Matrix Addition, Scalar Multiplication

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 2 & 3 \\ 4 & 5 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} 3 & 5 \\ 7 & 9 \\ 11 & 13 \end{bmatrix},$$

$$5 \begin{bmatrix} 0 & 1 & 0 \\ 2 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 5 & 0 \\ 10 & 15 & 10 \end{bmatrix}$$

$3 \times 2 \ + \ 3 \times 2 \ = 3 \times 2$

Matrix

- Matrix Multiplication
  - — Matrix & Vector Multiplication

$$Ax = \begin{bmatrix} 1 & 1 & 6 \\ 3 & 0 & 3 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix} = \begin{bmatrix} r_1^T \\ r_2^T \\ r_3^T \end{bmatrix} x = \begin{bmatrix} r_1^T x \\ r_2^T x \\ r_3^T x \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 1 \cdot 5 + 6 \cdot 0 \\ 3 \cdot 2 + 0 \cdot 5 + 3 \cdot 0 \\ 1 \cdot 2 + 1 \cdot 5 + 4 \cdot 0 \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \\ 7 \end{bmatrix}$$

  - — Matrix & Matrix Multiplication

$$AB = \begin{bmatrix} 1 & 1 & 6 \\ 3 & 0 & 3 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 5 & 4 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} r_1^T \\ r_2^T \\ r_3^T \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} r_1^T x_1 & r_1^T x_2 \\ r_2^T x_1 & r_2^T x_2 \\ r_3^T x_1 & r_3^T x_2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \cdot 2 + 1 \cdot 5 + 6 \cdot 0 & 1 \cdot 3 + 1 \cdot 4 + 6 \cdot 1 \\ 3 \cdot 2 + 0 \cdot 5 + 3 \cdot 0 & 3 \cdot 3 + 0 \cdot 4 + 3 \cdot 1 \\ 1 \cdot 2 + 1 \cdot 5 + 4 \cdot 0 & 1 \cdot 3 + 1 \cdot 4 + 4 \cdot 1 \end{bmatrix} = \begin{bmatrix} 7 & 13 \\ 6 & 12 \\ 7 & 11 \end{bmatrix}$$

Matrix

- Properties of matrix multiplication

  (1) Matrix multiplication is associative: $(AB)C = A(BC) \leftarrow n \times m \cdot m \times p \cdot p \times q = n \times q$

  (2) Matrix operations are distributive:

  $$A(B + C) = AB + AC \; and \; (B + C)D = BD + CD$$

  (3) Matrix multiplication is not commutative, i.e., $AB \neq BA$ in general

- $Def.$ The identity matrix $\mathbf{I}$ is the matrix that leaves every matrix unchanged when multiplied to that matrix.

  $\rightarrow \mathbf{I}$ is an $n \; x \; n$ matrix with 1's on the diagonal and 0's everywhere else

  $\rightarrow \mathbf{I}$ is commutative under matrix multiplication: $\mathbf{I}A = A\mathbf{I}$

$$\mathbf{I} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

Matrix

- The transpose of a matrix A is denoted by A^T
  - $A^T_{ij} = A_{ji}$
  - The i-th row of A^T = the row vector from the i-th column of A
  - A: $m \times n$, then A^T: $n \times m$
- Some important results
  - $(AB)^T = (B)^T(A)^T$
  - $(A^{-1})^T = (A^T)^{-1}$
- *Def*: A is called a symmetric matrix if $A^T = A$

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & 3 & 4 \end{bmatrix} \quad A^T = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 1 & 1 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 3 & 2 \\ 4 & 5 & 7 \end{bmatrix} \quad B^T = \begin{bmatrix} 1 & 4 \\ 3 & 5 \\ 2 & 7 \end{bmatrix}$$

Exercise

- For $X^T \coloneqq [x_1 \quad x_2]$, where $x_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, x_2 = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$, Calculate $x_1^T X^T$ and $X x_1$.

- Sol.

Exercise

- For $X^T := [x_1 \quad x_2]$, where $x_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, x_2 = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$, Calculate $x_1^T X^T$ and Xx_1.

- Sol.

$$x_1^T X^T = x_1^T [x_1 \quad x_2] = [x_1^T x_1 \quad x_1^T x_2] = [5 \quad -10]$$

$$Xx_1 = \begin{bmatrix} x_1^T \\ x_2^T \end{bmatrix} x_1 = \begin{bmatrix} x_1^T x_1 \\ x_2^T x_1 \end{bmatrix} = \begin{bmatrix} 5 \\ -10 \end{bmatrix}$$

Exercise

- For $X^T := [x_1 \quad x_2]$, where $x_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, x_2 = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$, Calculate $X^T X$ and XX^T.

- Sol.

Exercise

- For $X^T := [x_1 \quad x_2]$, where $x_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, x_2 = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$, Calculate $X^T X$ and XX^T.

- Sol.

$$X^T X = [x_1 \quad x_2] \begin{bmatrix} x_1^T \\ x_2^T \end{bmatrix} = x_1 x_1^T + x_2 x_2^T = \begin{bmatrix} 1 \\ -2 \end{bmatrix} [1 \quad -2] + \begin{bmatrix} -4 \\ 3 \end{bmatrix} [-4 \quad 3]$$

$$= \begin{bmatrix} 1 \times 1 & 1 \times (-2) \\ (-2) \times 1 & (-2) \times (-2) \end{bmatrix} + \begin{bmatrix} 16 & -12 \\ -12 & 9 \end{bmatrix} = \begin{bmatrix} 17 & -14 \\ -14 & 13 \end{bmatrix}$$

$$XX^T = \begin{bmatrix} x_1^T \\ x_2^T \end{bmatrix} [x_1 \quad x_2] = \begin{bmatrix} x_1^T x_1 & x_1^T x_2 \\ x_2^T x_1 & x_2^T x_2 \end{bmatrix} = \begin{bmatrix} 5 & -10 \\ -10 & 25 \end{bmatrix}$$

Gaussian Elimination

- Linear equation

$$y + z = 2$$
$$2x + 2y + 5z = 9$$
$$2x + 3y + 4z = 9$$

- Sol:

$$\left[\begin{array}{ccc|c} 0 & 1 & 1 & 2 \\ 2 & 2 & 5 & 9 \\ 2 & 3 & 4 & 9 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 2 & 2 & 5 & 9 \\ 0 & 1 & 1 & 2 \\ 2 & 3 & 4 & 9 \end{array}\right]$$

Gaussian Elimination

- Linear equation

$$y + z = 2$$
$$1x + 2y + 5z = 9$$
$$2x + 3y + 4z = 9$$

- Sol:

$$\left[\begin{array}{ccc|c} 0 & 1 & 1 & 2 \\ 2 & 2 & 5 & 9 \\ 2 & 3 & 4 & 9 \end{array}\right] \xrightarrow{permu.} \left[\begin{array}{ccc|c} 1 & 2 & 5 & 9 \\ 0 & 1 & 1 & 2 \\ 2 & 3 & 4 & 9 \end{array}\right] \xrightarrow{elo1} \left[\begin{array}{ccc|c} 2 & 4 & 10 & 18 \\ 0 & 1 & 1 & 2 \\ 2 & 3 & 4 & 9 \end{array}\right]$$

$$\xrightarrow{elo2} \left[\begin{array}{ccc|c} 2 & 4 & 10 & 18 \\ 0 & 1 & 1 & 2 \\ 0 & -1 & -6 & -9 \end{array}\right] \xrightarrow{elo2} \left[\begin{array}{ccc|c} 2 & 4 & 10 & 18 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -5 & -7 \end{array}\right]$$

$\therefore z = 7/5, y = 3/5, x = 4/5$

→ This is the usual variable elimination

Gaussian Elimination

- Linear equation $\begin{bmatrix}0 & 1 & 1\\ 1 & 2 & 5\\ 2 & 3 & 4\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\begin{bmatrix}2\\ 9\\ 9\end{bmatrix}$

- Sol:

$$E_1:\begin{bmatrix}2 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\left[\begin{array}{ccc|c}1 & 2 & 5 & 9\\ 0 & 1 & 1 & 2\\ 2 & 3 & 4 & 9\end{array}\right]=\left[\begin{array}{ccc|c}2 & 4 & 10 & 18\\ 0 & 1 & 1 & 2\\ 2 & 3 & 4 & 9\end{array}\right]$$

$$E_2:\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -1 & 0 & 1\end{bmatrix}\left[\begin{array}{ccc|c}2 & 4 & 10 & 18\\ 0 & 1 & 1 & 2\\ 2 & 3 & 4 & 9\end{array}\right]=\left[\begin{array}{ccc|c}2 & 4 & 10 & 18\\ 0 & 1 & 1 & 2\\ 0 & -1 & -6 & -9\end{array}\right]$$

$$E_3:\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 1\end{bmatrix}\left[\begin{array}{ccc|c}2 & 4 & 10 & 18\\ 0 & 1 & 1 & 2\\ 0 & -1 & -6 & -9\end{array}\right]=\left[\begin{array}{ccc|c}2 & 4 & 10 & 18\\ 0 & 1 & 1 & 2\\ 0 & 0 & -5 & -7\end{array}\right]$$

$$E_3E_2E_1A=U\rightarrow A=LU,\qquad L=(E_3E_2E_1)^{-1}$$
$$Ax=y\rightarrow x=A^{-1}y,\ A^{-1}=U^{-1}L^{-1}$$

$$E_i=\begin{bmatrix}1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & -l & 1 & \\ & & & & 1\end{bmatrix}$$

$$\Rightarrow {E_i}^{-1}=\begin{bmatrix}1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & l & 1 & \\ & & & & 1\end{bmatrix}$$

Gaussian Elimination

- Factorization $PA = LDU \rightarrow A = LDU$

  For $A = LU'$ where U' is invertible,

$$U' = \begin{bmatrix} d_1 & u_{12} & \cdots & u_{1n} \\ 0 & d_2 & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} 1 & u_{12}/d_1 & \cdots & u_{1n}/d_1 \\ 0 & 1 & \cdots & u_{2n}/d_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = DU$$

  $\therefore$ We can obtain $A = LDU$, where L and U have 1's on the diagonal and D is the diagonal matrix

$$A = LU' = LDU,\ L = (E_3E_2E_1)^{-1}, U = (E_6E_5E_4)^{-1}$$
$$Ax = y \rightarrow x = A^{-1}y, A^{-1} = U^{-1}D^{-1}L^{-1}$$

$$E_i = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & -l & 1 & \\ & & & & 1 \end{bmatrix}$$

$$\Rightarrow E_i^{-1} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & l & 1 & \\ & & & & 1 \end{bmatrix}$$

Gaussian Elimination

- Gauss-Jordan method to find A^{-1}

  $AA^{-1} = \mathrm{I}$

  We can perform elementary operations to A to get $E_n \cdots E_2 E_1 A = \mathrm{I}$

  By applying A^{-1} to both side, $E_n \cdots E_2 E_1 \mathrm{I} = A^{-1}$

- One simple way to find A^{-1}

  By applying elementary operations to $[A \mid \mathrm{I}] \rightarrow [\mathrm{I} \mid A^{-1}]$

Gaussian Elimination

- Example

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}. \quad \text{Find } A^{-1}.$$

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 4 & -6 & 0 & 0 & 1 & 0 \\ -2 & 7 & 2 & 0 & 0 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 8 & 3 & 1 & 0 & 1 \end{array}\right]$$

Gaussian Elimination

- Example

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}. \quad \text{Find } A^{-1}.$$

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 4 & -6 & 0 & 0 & 1 & 0 \\ -2 & 7 & 2 & 0 & 0 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 8 & 3 & 1 & 0 & 1 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 2 & -1 & -1 \\ 0 & -8 & 0 & -4 & 3 & 2 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right]$$

Gaussian Elimination

- Example

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}. \quad \text{Find } A^{-1}.$$

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 4 & -6 & 0 & 0 & 1 & 0 \\ -2 & 7 & 2 & 0 & 0 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 8 & 3 & 1 & 0 & 1 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 2 & -1 & -1 \\ 0 & -8 & 0 & -4 & 3 & 2 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 2 & -1 & -1 \\ 0 & 1 & 0 & 1/2 & -3/8 & -1/4 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 3/2 & -5/8 & -3/4 \\ 0 & 1 & 0 & 1/2 & -3/8 & -1/4 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right]$$

Gaussian Elimination

- Example

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}. \quad \text{Find } A^{-1}.$$

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 4 & -6 & 0 & 0 & 1 & 0 \\ -2 & 7 & 2 & 0 & 0 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 8 & 3 & 1 & 0 & 1 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 2 & -1 & -1 \\ 0 & -8 & 0 & -4 & 3 & 2 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 2 & -1 & -1 \\ 0 & 1 & 0 & 1/2 & -3/8 & -1/4 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 3/2 & -5/8 & -3/4 \\ 0 & 1 & 0 & 1/2 & -3/8 & -1/4 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 3/4 & -5/16 & -3/8 \\ 0 & 1 & 0 & 1/2 & -3/8 & -1/4 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right] \quad \therefore A^{-1} = \begin{bmatrix} 3/4 & -5/16 & -3/8 \\ 1/2 & -3/8 & -1/4 \\ -1 & 1 & 1 \end{bmatrix}$$

Gaussian Elimination

- Computational Complexity of A^{-1}

$$A=\begin{bmatrix}2 & 1 & 1\\ 4 & -6 & 0\\ -2 & 7 & 2\end{bmatrix}.\ \text{Find } A^{-1}.$$

$$\left[\begin{array}{ccc|ccc}2 & 1 & 1 & 1 & 0 & 0\\ 4 & -6 & 0 & 0 & 1 & 0\\ -2 & 7 & 2 & 0 & 0 & 1\end{array}\right]\Rightarrow\left[\begin{array}{ccc|ccc}2 & 1 & 1 & 1 & 0 & 0\\ 0 & -8 & -2 & -2 & 1 & 0\\ 0 & 8 & 3 & 1 & 0 & 1\end{array}\right]$$

$$\Rightarrow\left[\begin{array}{ccc|ccc}2 & 1 & 1 & 1 & 0 & 0\\ 0 & -8 & -2 & -2 & 1 & 0\\ 0 & 0 & 1 & -1 & 1 & 1\end{array}\right]\Rightarrow\left[\begin{array}{ccc|ccc}2 & 1 & 0 & 2 & -1 & -1\\ 0 & -8 & 0 & -4 & 3 & 2\\ 0 & 0 & 1 & -1 & 1 & 1\end{array}\right]$$

$$\Rightarrow\left[\begin{array}{ccc|ccc}2 & 1 & 0 & 2 & -1 & -1\\ 0 & 1 & 0 & 1/2 & -3/8 & -1/4\\ 0 & 0 & 1 & -1 & 1 & 1\end{array}\right]\Rightarrow\left[\begin{array}{ccc|ccc}2 & 0 & 0 & 3/2 & -5/8 & -3/4\\ 0 & 1 & 0 & 1/2 & -3/8 & -1/4\\ 0 & 0 & 1 & -1 & 1 & 1\end{array}\right]$$

$$\Rightarrow\left[\begin{array}{ccc|ccc}1 & 0 & 0 & 3/4 & -5/16 & -3/8\\ 0 & 1 & 0 & 1/2 & -3/8 & -1/4\\ 0 & 0 & 1 & -1 & 1 & 1\end{array}\right]\quad\therefore A^{-1}=\begin{bmatrix}3/4 & -5/16 & -3/8\\ 1/2 & -3/8 & -1/4\\ -1 & 1 & 1\end{bmatrix}$$

$2n$ multiplications
$2n$ additions
$\rightarrow 4n$ flops or 2n flops

n times operations for U
n times operations for L
$\rightarrow 4n(n)(n)$ flops
$\approx 4n^3 \rightarrow O(n^3)$

Vector space

- Vector space V is a set of vectors such that

A. $\forall x, y \in V,\ x + y$ and $a \cdot x = ax$ are unique in V.

B. $\forall x, y, z \in V$, the following hold:

✓ Commutativity: $x + y = y + x$

✓ Associativity: $(x + y) + z = x + (y + z)$

✓ Identity: $\exists 0 \in V$ such that $x + 0 = x$

✓ Inverse: $\exists u \in V$ such that $x + u = 0$

✓ Identity: $1x = x$

✓ Associativity: $(ab)x = a(bx)$

✓ Distributivity: $a(x + y) = ax + ay$

✓ Distributivity: $(a + b)x = ax + bx$.

Vector Space

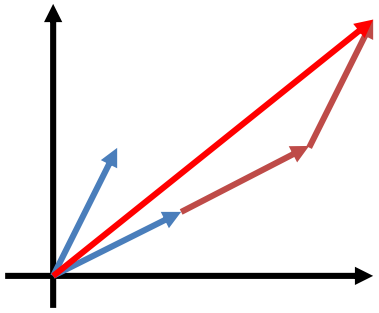
- Linear Combination

  For *vectors* $\boldsymbol{v}, \boldsymbol{w} \in X$ and *scalars* $\alpha, \beta \in F$ $(R \text{ or } C)$,

  $\alpha \boldsymbol{v} + \beta \boldsymbol{w}$ is a linear combination of $\boldsymbol{v}$ and $\boldsymbol{w}$.

Example :

$$2\begin{bmatrix}2\\1\end{bmatrix} + 1\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}5\\4\end{bmatrix} \in X$$



Vector Space

- Linear Combination

  For *vectors* $\boldsymbol{v}, \boldsymbol{w} \in X$ and *scalars* $\alpha, \beta \in F\ (R\ or\ C)$,

  $\alpha\boldsymbol{v} + \beta\ \boldsymbol{w}$ is a linear combination of $\boldsymbol{v}$ and $\boldsymbol{w}$.

- Vector Space (simple check)

  For any *vectors* $\boldsymbol{v}, \boldsymbol{w} \in X$ and *any scalars* $\alpha, \beta \in F\ (R\ or\ C)$,

  if $\alpha\boldsymbol{v} + \beta\ \boldsymbol{w} \in X$, becomes a vector space under the normal operation assumption

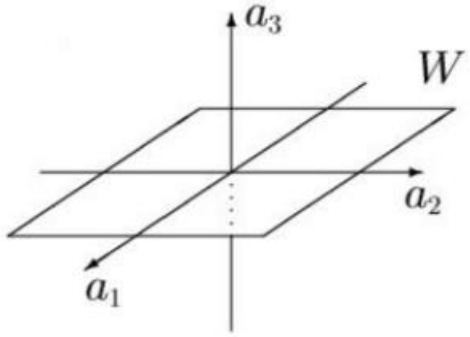
- Subspace : V is a vector space and $W \subseteq V$. Then W is a subspace of V if and only if

  $x, y \in W \Rightarrow ax + by \in W$,

  where $a, b \in F$ [closed under $linear\ combination$].

Vector Space

- Example of Vector Space
  - $V = F^n = \{(a_1, \cdots, a_n) : a_i \in F\}$,
    $W = \{(a_1, a_2, 0, \cdots, 0) : a_i \in F\}$
  - $V = M_{n\times n}(F)$,
    $W = \{symmetric\ matrices\}$
  - $V = R^{\infty}$
    $W = \{convergent\ sequences\}$



  - A set of n-order polynomials $V = \{p_n(x) | p_n(x) = \sum_{k=0}^{n} a_k x^k, a_k \in F\}$
    $W = \{p_m(x) | p_m(x) = \sum_{k=0}^{m} a_k x^k, a_k \in F,\ m < n\}$
  - A set of analytic functions $V = \{f(x) | f(x) = \sum_{k=0}^{\infty} a_k x^k, a_k \in F\}$
    $W = \{f(x) | f(x): \text{continuous function}\}$
  - A set of periodic functions $V = \{s(t) | s(t) = \sum_{k=-\infty}^{\infty} a_k e^{i2\pi k f_0 t}, a_k \in F\}$
    $W = \{s(t) | s(t) = \sum_{k=-K}^{K} a_k e^{i2\pi k f_0 t}, a_k \in F\}$

Vector Space

- Linearly independent (LI): For *vectors* $\boldsymbol{v_i} \in X$, $i = 1, \dots, n$ and *scalars* $\alpha_i \in F$, $\{\boldsymbol{v_i}\}$ are LI when a linear combination $\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v_2} + \cdots + \alpha_n \boldsymbol{v_n} = \mathbf{0}$ if only if $\alpha_1 = \alpha_2 = \cdots = \alpha_n = \mathbf{0}$. Otherwise, $\{\boldsymbol{v_i}\}$ are linearly dependent.

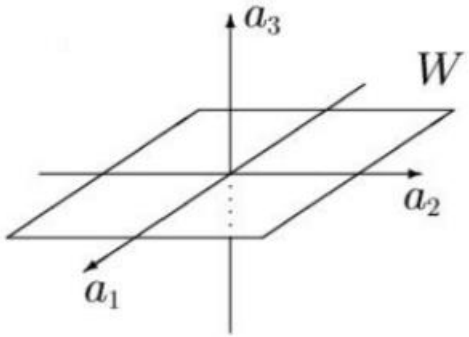
- Example
  - A set of n-order polynomials $V = \{p_n(x) | p_n(x) = \sum_{k=0}^{n} a_k x^k, a_k \in F\}$. For $1, x, x^2, x^3, \cdots, x^n \in V, \sum_{k=0}^{n} a_k x^k = 0, a_k = 0, k = 0, \dots, n$. Thus $\{x^i, i = 0, \cdots, n\}$ are LI.
  - $V = M_{2\times 2}(F)$, for $\left\{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right\} \in M_{2\times 2}(F)$,

    $a_1 \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + a_2 \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + a_3 \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + a_4 \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix} = \begin{bmatrix}a_1 & a_2\\a_3 & a_3\end{bmatrix} = \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$ implies

    $a_i = 0, i = 1, 2, 3, 4$. Thus $\left\{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right\}$ are LI.

Vector Space

- Dimension of a vector space X : maximum number of linearly independent vectors in a vector space X.



- Example
  - 3-dim real space: R^3

    $(a, b, c) = a(1, 0, 0) + b(0, 1, 0) + c(0, 0, 1)$
  - $V = M_{2\times 2}(F)$

    $$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_3 \end{bmatrix} = a_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + a_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$
  - $V = \{p_2(x) | p_2(x) = \sum_{k=0}^{2} a_k x^k = a_0 + a_1 x + a_2 x^2$

Vector Space

- Basis is defined by a set of maximum number of linearly independent vectors in a vector space V.

- Example
  - 3-dim real space: R^3

    $(a, b, c) = a(1, 0, 0) + b(0, 1, 0) + c(0, 0, 1)$

    Basis= $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ or $\{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$
  - $V = M_{2\times 2}(F)$

    $$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_3 \end{bmatrix} = a_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + a_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

    Basis= $\left\{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right\}$
  - $V = \{p_2(x) | p_2(x) = \sum_{k=0}^{2} a_k x^k = a_0 + a_1 x + a_2 x^2$

    Basis= $\{1, x, x^2\}$ or $\{x + 1, 2x + 1, x^2 + 2\}$

Vector Space

- Generating set of V: $\{Basis \cup any\ subset\ of\ V\}$, which can span V by linear combinations.

- Vector representations of a vector x

$$x = \begin{bmatrix}5\\4\end{bmatrix} = 5\begin{bmatrix}1\\0\end{bmatrix} + 4\begin{bmatrix}0\\1\end{bmatrix} = 5e_1 + 4e_2 = [e_1\ \ e_2]\begin{bmatrix}5\\4\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}5\\4\end{bmatrix}$$

$$x = \begin{bmatrix}5\\4\end{bmatrix} = 2\begin{bmatrix}2\\1\end{bmatrix} + 1\begin{bmatrix}1\\2\end{bmatrix} = 2b_1 + 1b_2 = [b_1\ \ b_2]\begin{bmatrix}2\\1\end{bmatrix} = \begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}\begin{bmatrix}2\\1\end{bmatrix}$$

$$[x]_E = \begin{bmatrix}5\\4\end{bmatrix}, [x]_B = \begin{bmatrix}2\\1\end{bmatrix} \rightarrow [x]_B = B^{-1}E\ [x]_E = T[x]_E,\ T\text{: Isomorphism}$$

Vector Space

- Exercise:

  Are (3, 0, 0), (4, 1, 0), (2, 5, 2) L.I. ?

$$\alpha_1 \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix} = 0$$

Vector Space

- Linearly independent (LI): For *vectors* $\boldsymbol{v_i} \in X$, $i = 1, \dots, n$ and *scalars* $\alpha_i \in F$, $\{\boldsymbol{v_i}\}$ are LI when a linear combination $\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v_2} + \cdots + \alpha_n \boldsymbol{v_n} = \mathbf{0}$ if only if $\alpha_1 = \alpha_2 = \cdots = \alpha_n = \mathbf{0}$.
  Otherwise, $\{\boldsymbol{v_i}\}$ are linearly dependent.

- Example:

  Are (3, 0, 0), (4, 1, 0), (2, 5, 2) L.I. ?

$$\alpha_1 \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix} = 0$$

$$\begin{bmatrix} 3 & 4 & 2 \\ 0 & 1 & 5 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \mathbf{0} \rightarrow \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \mathbf{0}$$

  → (3, 0, 0), (4, 1, 0), (2, 5, 2) are LI.

Vector Space

- Example:

  Are (2 1 0), (4 2 0), (2 5 2) L.I. ?

$$\begin{bmatrix}2 & 4 & 2\\ 1 & 2 & 5\\ 0 & 0 & 2\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{bmatrix} = \mathbf{0} \underset{ero}{\longrightarrow} \begin{bmatrix}2 & 4 & 2\\ 0 & 0 & 8\\ 0 & 0 & 2\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{bmatrix} = \mathbf{0},$$

$$\rightarrow \alpha_3 = 0,\ \alpha_1 + 2\alpha_2 = 0,$$

$$\rightarrow \begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{bmatrix} = \begin{bmatrix}2\\ -1\\ 0\end{bmatrix}, \begin{bmatrix}3\\ -1.5\\ 0\end{bmatrix}, \begin{bmatrix}-4\\ 2\\ 0\end{bmatrix}, \ \dots\dots$$

$$\rightarrow Solution\ set = \left\{ a\begin{bmatrix}2\\ -1\\ 0\end{bmatrix} \mid\ a \in R \right\}$$

$$= Null\ space\ of\ A = \{x | Ax = 0\} \rightarrow \text{1-dim vector space}$$

→ (2 1 0), (4 2 0), (2 5 2) are not LI, i.e., LD.

Vector Space

- What is the rank of A, i.e., the number of LI columns?

$$A = \begin{bmatrix} 3 & 2 & 3 & -2 & 1 \\ 1 & 1 & 1 & 0 & 3 \\ 1 & 2 & 1 & -1 & 2 \\ 5 & 5 & 5 & -3 & 6 \end{bmatrix}$$

Vector Space

- Reduced row echelon form:

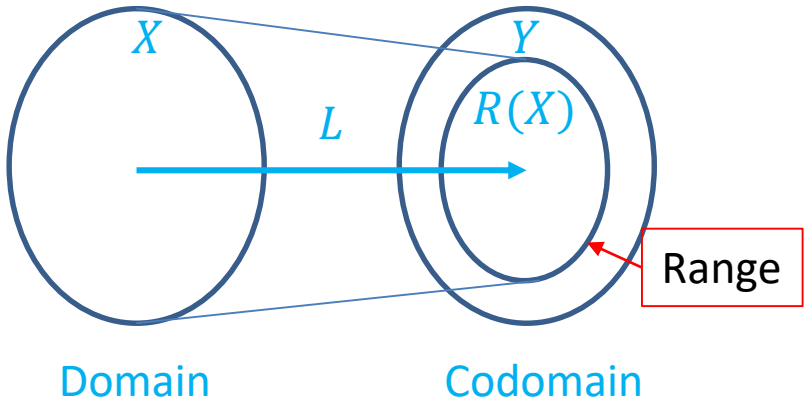
$$\begin{bmatrix} 3 & 2 & 3 & -2 & 1 \\ 1 & 1 & 1 & 0 & 3 \\ 1 & 2 & 1 & -1 & 2 \\ 5 & 5 & 5 & -3 & 6 \end{bmatrix} \xrightarrow[ero]{} \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

key columns key columns

$$\begin{bmatrix} 3 \\ 1 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ -1 \\ -3 \end{bmatrix}$$ are linear independent

Linear Transformation

- Transformation $L: X \rightarrow Y$



- *Def*: Let L be a transformation from v.s. X to Y. Then L is called a *linear transformation* if $L(cx + dy) = cL(x) + dL(y)$ for all $x, y \in X$ and all scalars c and d.
- If $\dim(X) = \dim(Y)$, L is called *linear operator*.
- Ex) convolution: $L(h, \alpha_1 u_1 + \alpha_2 u_2) = h * (\alpha_1 u_1 + \alpha_2 u_2)$

$$= \int h(t-\tau)\,(\alpha_1 u_1(\tau) + \alpha_2 u_2(\tau))d\tau$$

$$= \alpha_1 \int h(t-\tau)u_1(\tau)\,d\tau + \alpha_1 \int h(t-\tau)u_1(\tau)\,d\tau = \alpha_1 L(h, u_1) + \alpha_2 L(h, u_2)$$

Linear Transformation

- Matrix representation of linear transformation

  Q: What is the matrix representation of $L = \frac{d}{dt}$ for P_3.

Linear Transformation

Q: What is the matrix representation of $L = \frac{d}{dt}$ for P_3.

A:

Basis β for P_3 : $p_1 = 1, \quad p_2 = t, \quad p_3 = t^2, \quad p_4 = t^3$

Action of $\frac{d}{dt}$: $Ap_1 = 0, \quad Ap_2 = 1, \quad Ap_3 = 2t, \quad Ap_4 = 3t^2$

Matrix representation of L: $[L]_\beta = A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

For $p(t) = 2 + t - t^2 - t^3 \rightarrow p = \begin{bmatrix} 2 \\ 1 \\ -1 \\ -1 \end{bmatrix}$; $\quad Lp(t) = Ap = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ -3 \\ -0 \end{bmatrix}$

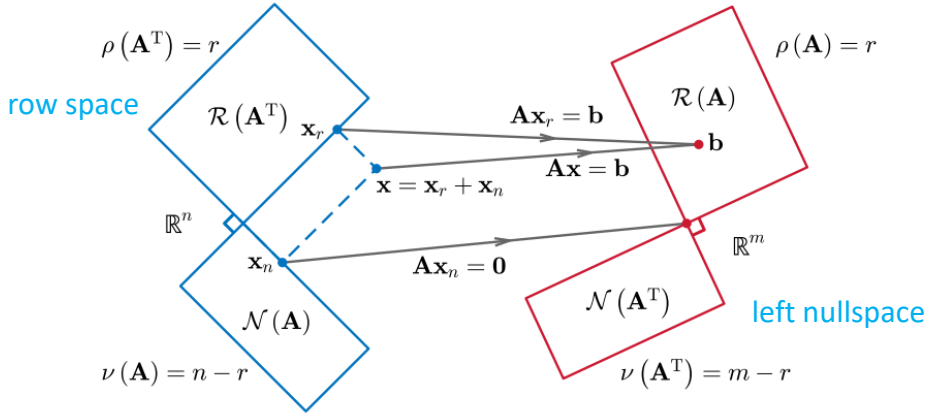
$\frac{d}{dt}(2 + t - t^2 - t^3) = 1 - 2t - 3t^2$

Linear Transformation

- Null Space : $\{x|Ax = 0\}$

$$\begin{bmatrix}2 & 4 & 2\\ 1 & 2 & 5\\ 0 & 0 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \mathbf{0} \rightarrow \begin{bmatrix}2 & 4 & 2\\ 0 & 0 & 8\\ 0 & 0 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \mathbf{0} \rightarrow \begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} \neq \mathbf{0}, \begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \begin{bmatrix}2\\ -1\\ 0\end{bmatrix}, \begin{bmatrix}3\\ -1.5\\ 0\end{bmatrix}, \begin{bmatrix}-4\\ 2\\ 0\end{bmatrix}, \dots\dots$$

$$\rightarrow solution\ set = \left\{\beta\begin{bmatrix}2\\ -1\\ 0\end{bmatrix} \mid \beta \in R\right\} = \{x|Ax = 0\}$$



Linear Transformation

- Null Space : $\{x|Ax = 0\}$

$$\begin{bmatrix}2 & 4 & 2\\ 1 & 2 & 5\\ 0 & 0 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \mathbf{0} \rightarrow \begin{bmatrix}2 & 4 & 2\\ 0 & 0 & 8\\ 0 & 0 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \mathbf{0} \rightarrow \begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} \neq \mathbf{0}, \begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \begin{bmatrix}2\\ -1\\ 0\end{bmatrix}, \begin{bmatrix}3\\ -1.5\\ 0\end{bmatrix}, \begin{bmatrix}-4\\ 2\\ 0\end{bmatrix}, \ \dots\dots$$

$$\rightarrow solution\ set = \left\{\beta\begin{bmatrix}2\\ -1\\ 0\end{bmatrix} \mid \beta \in R\right\} = \{x|Ax = 0\}$$

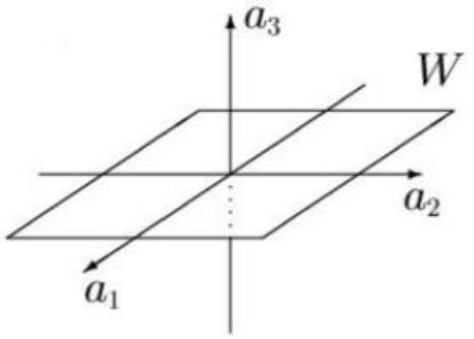
$\rightarrow Null\ space\ of\ A = N(A)$: 1-dim vector space

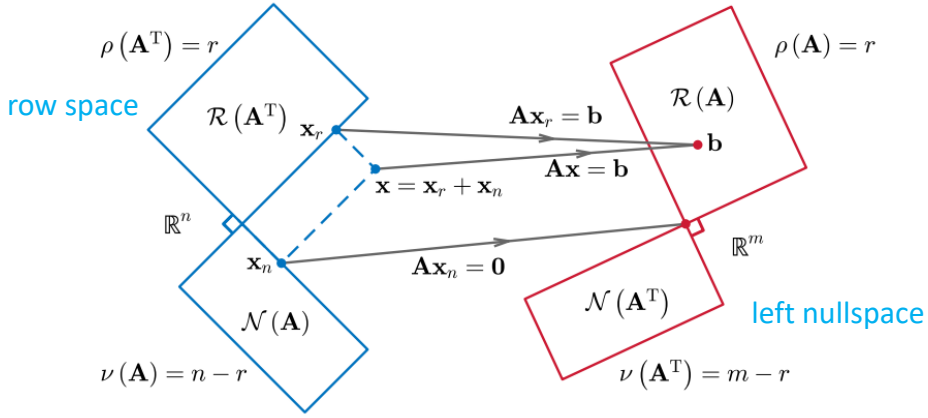
Linear Transformation

- Range(Column) Space: $C(A) = R(A) = \{b|\ Ax = b, \forall x \in V\}$

$$\begin{bmatrix} 2 & 4 & 2 \\ 1 & 2 & 5 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = (x_1 + 2x_2) \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix}$$

$$\rightarrow R(A) = \{b|\ Ax = b, \forall x \in V\} = \left\{ \beta_1 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \beta_2 \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix} | \beta_1, \beta_2 \in R \right\}$$





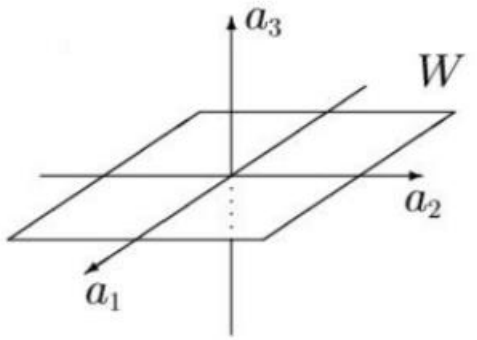
Linear Transformation

- Range(Column) Space: $C(A) = R(A) = \{b|\ Ax = b, \forall x \in V\}$

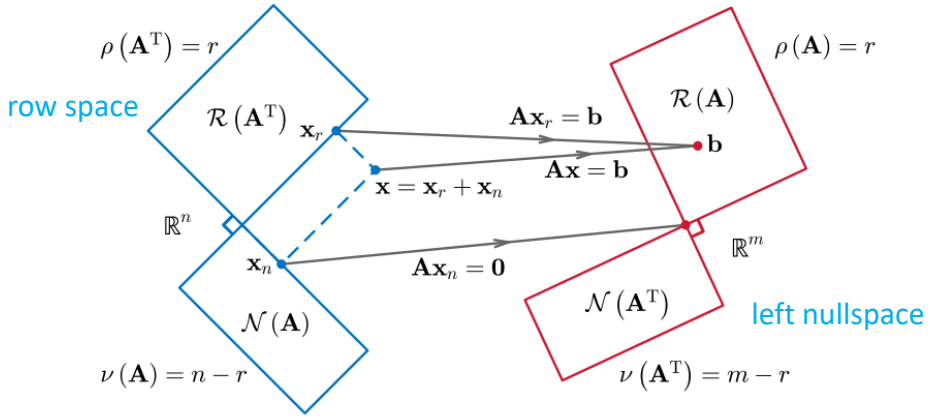
$$\begin{bmatrix} 2 & 4 & 2 \\ 1 & 2 & 5 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = (x_1 + 2x_2) \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix}$$

$$\rightarrow R(A) = \{b|\ Ax = b, \forall x \in V\} = \left\{ \beta_1 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \beta_2 \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix} | \beta_1, \beta_2 \in R \right\}$$

$\rightarrow Range\ space\ of\ A = R(A)$: 2-dim vector space



- For $x \in V$, $\rho(V) = \rho(N(A) + \rho(R(A))$



Linear Transformation

- General Solution of Linear Equation: $Ax = b, \rho(N(A)) \neq 0$

$$\begin{bmatrix} 2 & 4 & 2 \\ 1 & 2 & 5 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 2 \end{bmatrix} \underset{ero}{\longrightarrow} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 0 & 8 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 2 \end{bmatrix}$$

$$\rightarrow x_3 = 1, x_1 + 2x_2 = 3 \rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 - 2t \\ t \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, t \in R$$

Linear Transformation

- General Solution of Linear Equation: $Ax = b, \rho(N(A)) \neq 0$

$$\begin{bmatrix} 2 & 4 & 2 \\ 1 & 2 & 5 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 2 \end{bmatrix} \underset{reo}{\longrightarrow} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 0 & 8 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 2 \end{bmatrix}$$

$$\rightarrow x_3 = 1, x_1 + 2x_2 = 3 \rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 - 2t \\ t \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, t \in R$$

$$\begin{bmatrix} 2 & 4 & 2 \\ 1 & 2 & 5 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 2 \end{bmatrix} \rightarrow \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$$ is particular solution.

$$\begin{bmatrix} 2 & 4 & 2 \\ 1 & 2 & 5 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} -2t \\ t \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow N(A) = \left\{ t \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \mid t \in R \right\}$$

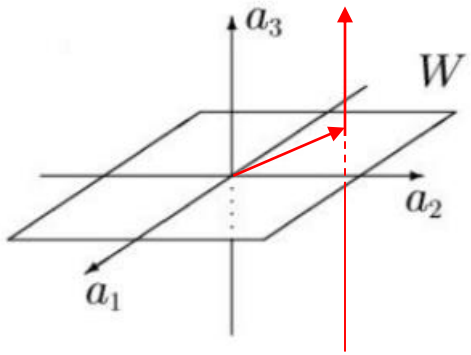
$$\rightarrow \text{solution set} = \left\{ \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \right\} + \left\{ t \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \mid t \in R \right\} = \{x_p\} + N(A)$$ (is it v. space?)

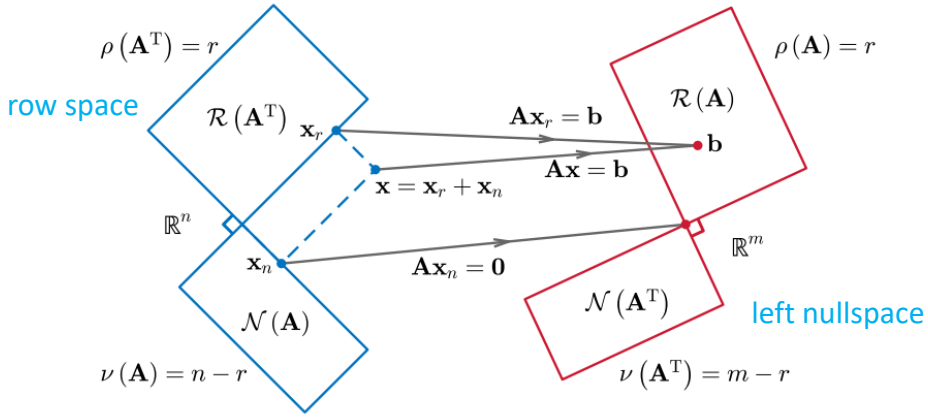
Linear Transformation

- General Solution of Linear Equation:

$$\text{Solution set} = \left\{ \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \right\} + \left\{ t \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \mid t \in R \right\} = \{x_p\} + N(A)$$

Is it v. space? No.





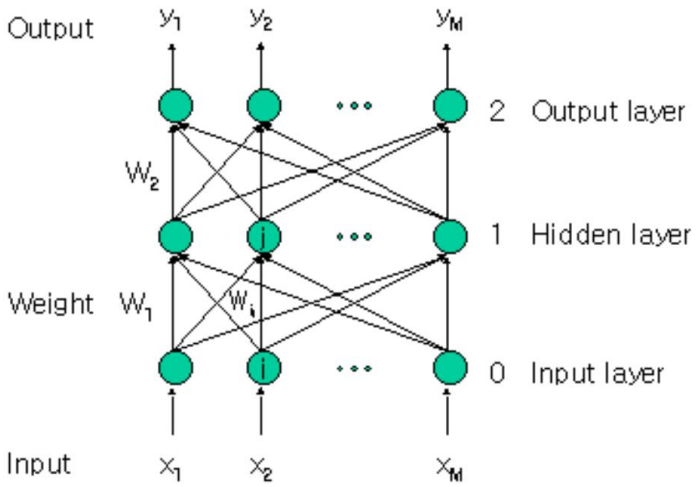
Linear Transformation

- $n \times m$ matrix $W \in R^{n \times m}$ is a matrix representation for a linear transformation from R^m to R^n, i.e., $W: R^m \rightarrow R^n$.
- For $\boldsymbol{y} \in R^n, \boldsymbol{x} \in R^m$, $\boldsymbol{y} = W\boldsymbol{x}$.

  $n \times 1 = \quad n \times m \quad \cdot \quad m \times 1$

- $$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

  $$= x_1 \begin{bmatrix} w_{11} \\ w_{21} \\ \vdots \\ w_{n1} \end{bmatrix} + x_2 \begin{bmatrix} w_{12} \\ w_{22} \\ \vdots \\ w_{n2} \end{bmatrix} + \cdots + x_m \begin{bmatrix} w_{1m} \\ w_{2m} \\ \vdots \\ w_{nm} \end{bmatrix}$$



- Neural network: $y = \sigma(Wx) \leftarrow non\text{-}linear\ transformation$

Linear Transformation

- $n \times m$ matrix $W \in R^{n \times m}$ is a matrix representation for a linear transformation from R^m to R^n, i.e., $W: R^m \rightarrow R^n$.
- For $\boldsymbol{y} \in R^n, \boldsymbol{x} \in R^m$, $\boldsymbol{y} = W\boldsymbol{x}$.

  $(n \times 1 = \quad n \times m \quad \cdot \quad m \times 1)$
- Associate Memory

  for training samples $\{(\boldsymbol{x}_i, \boldsymbol{y}_i) | i = 1, \dots, p\}$

  $\boldsymbol{y}_i = W\boldsymbol{x}_i$, $i = 1, \dots, p$

  $[\boldsymbol{y}_1 \quad \cdots \quad \boldsymbol{y}_p] = W[\boldsymbol{x}_1 \quad \cdots \quad \boldsymbol{x}_p]$

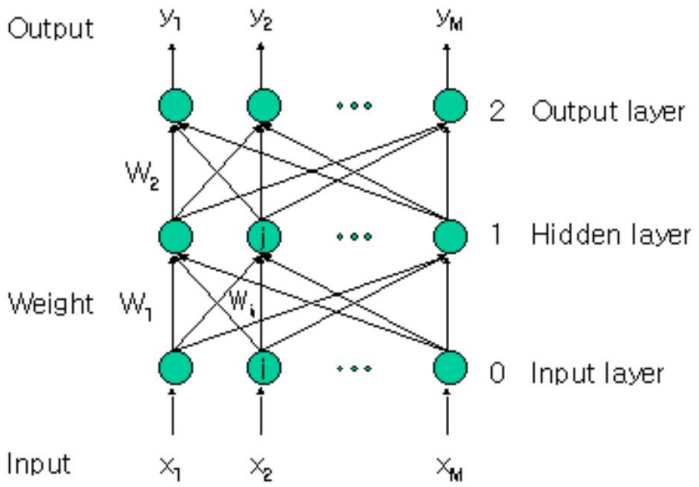
  $(n \times p = \quad n \times m \quad \cdot \quad m \times p)$

  $Y = WX$

  $YX^T = WXX^T$

  $(n \times p \cdot p \times m = \quad n \times m \cdot \ m \times p \cdot p \times m)$

  $W = YX^T(XX^T)^{-1}$



Summary

basics
- matrix and vector operation
- Gaussian elimination: inverse of matrix
- linear equations

vector space
- def., dimension, basis, subspace
- linear combination, linearly independent
- vector representation

linear transformation
- matrix representation of LT
- null space and range space
- solution set of linear equation
- neural transformation

useful tricks
- vector norm, matrix norm
- eigenvalue and eigenvector
- singular value decomposition
- pseudo inverse
- derivative of vector function

LINEAR ALGEBRA II

JIN YOUNG CHOI

ECE, SEOUL NATIONAL UNIVERSITY

Outline

basics
- matrix and vector operation
- Gaussian elimination: inverse of matrix
- linear equations

vector space
- def., dimension, basis, subspace
- linear combination, linearly independent
- vector representation

linear transformation
- matrix representation of LT
- null space and range space
- solution set of linear equation
- neural transformation

useful tricks
- vector norm, matrix norm
- eigenvalue and eigenvector
- singular value decomposition
- pseudo inverse
- derivative of vector function

Determinant

- Main uses of determinants
  1. Test of invertibility
     $\det(A) \neq 0 \quad \Leftrightarrow \quad A$ is invertible
  2. $\det(A)$ equals the volume of a box in n-dimensional space.

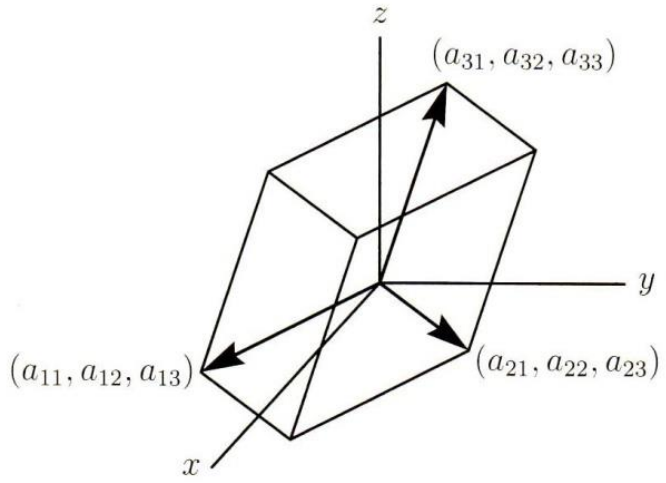


Figure 4.1 The box formed from the rows of A: volume = |determinant|.

Determinant

- Calculation rule

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$$

$$= a_{11}(-1)^{1+1}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}(-1)^{1+2}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}(-1)^{1+3}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$= a_{11}\,(a_{22}a_{33} - a_{23}a_{32}) - a_{12}\,(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

$$= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

Determinant

- Key property of Determinant
  - $\det(AB) = \det(A)\det(B)$
  - $A \xrightarrow[ero1]{} B \Rightarrow \det(B) = -\det(A)$
  - $A \xrightarrow[ero2]{} B \Rightarrow \det(B) = k\det(A)$
  - $A \xrightarrow[ero3]{} B \Rightarrow \det(B) = \det(A)$
  - $\det(A^T) = \det(A)$
  - $Ax = b \rightarrow x_k = \frac{1}{\det(A)}\det(a_1, a_2, \cdots, b, \cdots, a_n)$,

    where $a_i \coloneqq i-th\ column\ of\ A$

Vector Operations

- Norm: Length(magnitude) of a vector $x = [x_1 \quad \cdots \quad x_n]^T$

  $\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}$

  $\|x\|_\infty = \max_i |x_i|$

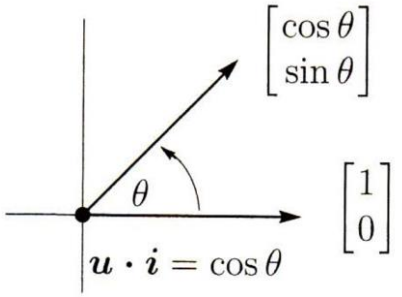
- Inner product (scalar product or dot product) of x and y :

$$x^T y = [x_1 \quad \cdots \quad x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

- x and y are orthogonal if and only if

  $x^T y = 0.$

  $x^T y = \|x\|\|y\| cos\theta \rightarrow x^T x = \|x\|_2^2$



Matrix Norm

- **Induced Norm**

$$\|A\|_p = \sup_{\|x\|_p=1} \|Ax\|_p \qquad \|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}| \qquad \|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|$$

- **Schatten norm**

$$\|A\|_p = \left(\sum_{i=1}^{\min\{m,\,n\}} \sigma_i^p(A) \right)^{1/p}$$

- **Nuclear norm**

$$\|A\|_* = \operatorname{trace}\left(\sqrt{A^* A}\right) = \sum_{i=1}^{\min\{m,\,n\}} \sigma_i(A)$$

- **Frobenius Norm**

$$\|A\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\operatorname{trace}\left(A^{\mathsf{T}} A\right)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(A)}$$

- **Spectral (maximum singular value) norm**

$$f(A) = \|A\|_2 = \sigma_{\max}\ (A) = (\lambda_{\max}\ (A^T A))^{1/2}$$

$$\begin{aligned}
\|A\|_2 &= \sup_{\|x\|_2=1} \|Ax\|_2 \\
&= \sup_{\|x\|_2=1} (x^T A^T A x)^{1/2} \\
&= \sup_{\|x\|_2=1} (x^T U^T \Lambda U x)^{1/2} \\
&= \sup_{\|y\|_2=1} (y^T \Lambda y)^{1/2} \Leftarrow y^T y = x^T U^T U x \\
&= \sup_{\|y\|_2=1} (\textstyle\sum {y_i}^2 \lambda_i)^{1/2} \\
&= (\lambda_{\max}\ (A^T A))^{1/2}
\end{aligned}$$

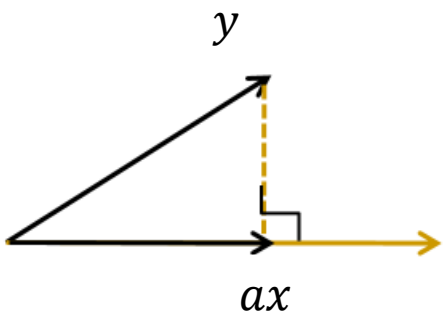
Vector Operations

- Projection of y onto x:

$$x \perp (y - ax) \leftrightarrow x^T(y - ax) = 0 \leftrightarrow x^T y = ax^T x$$

$$\therefore a = \frac{x^T y}{x^T x} = \frac{\|x\|\|y\|cos\theta}{\|x\|^2}$$

$$\therefore ax = \frac{x^T y}{x^T x} x = \frac{\|y\|cos\theta}{\|x\|} x$$



- Transpose of inner products:

$$(Ax)^T y = x^T A^T y = x^T (A^T y)$$
$$(ABx)^T y = (Bx)^T A^T y = x^T (B^T A^T) y$$

Orthogonalization

- Orthogonalization:

$$\boldsymbol{v_i^T v_j} = \begin{cases} a > 0 & if\ i = j \\ 0 & otherwise \end{cases}$$

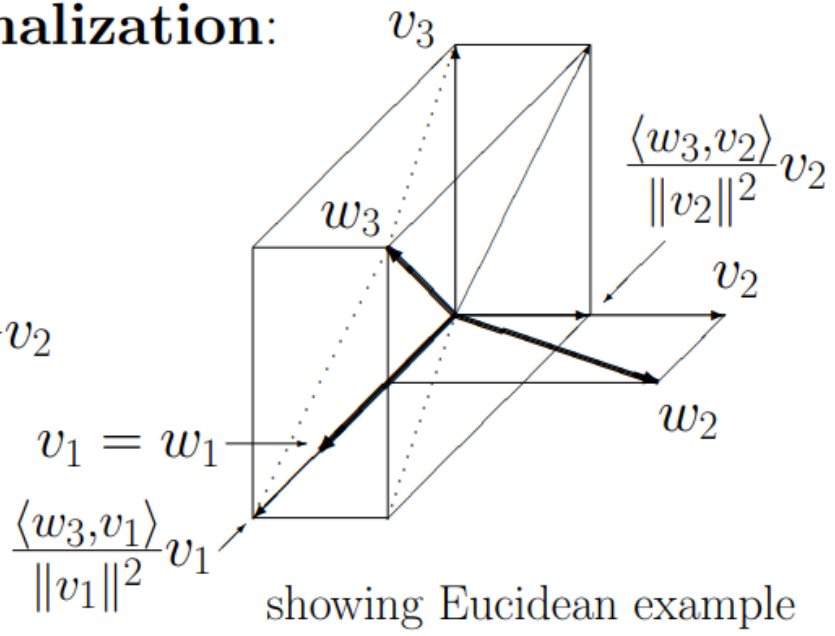
Gram-Schmidt orthogonalization:

$$v_1 = w_1$$

$$v_2 = w_2 - \frac{\langle w_2, v_1 \rangle}{\|v_1\|^2} v_1$$

$$v_3 = w_3 - \frac{\langle w_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle w_3, v_2 \rangle}{\|v_2\|^2} v_2$$

$$\cdots$$



showing Eucidean example

Eigenvalues and Eigenvectors

- Eigenvectors and eigenvalues of a matrix:

$$\boldsymbol{A v_i} = \lambda_i \boldsymbol{v_i}, \qquad \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{N}$$

$[\boldsymbol{A} - \lambda_i \mathbf{I}]\boldsymbol{v_i} = 0 \implies |\boldsymbol{A} - \lambda_i \mathbf{I}| = 0$ has N λ_i (can be duplicated)
$\implies \exists\ \boldsymbol{v_i} \neq \mathbf{0}$ since $\rho[\boldsymbol{A} - \lambda_i \mathbf{I}] \leq N - 1 \ \ \forall \lambda_i$
$\implies \exists\ N$ linearly independent eigenvectors $\{\boldsymbol{v_i}\}$
$\implies \{\boldsymbol{v_i}\}$ becomes a basis of the eigenspace E of $\boldsymbol{A}$
$\implies \{\boldsymbol{v_i}\}$ span the eigenspace of $\boldsymbol{A}$
$\implies \boldsymbol{A}$ is diagonalizable if λ_i are distinctive.

Note: E_i is the eigenspace for λ_i, spanned by $\boldsymbol{v_i}$.
Note: If λ_i is not distinct, $\boldsymbol{A}$ may not be diagonalizable

Diagonalization

- Eigenvalue and Eigenvector

Define $\boldsymbol{v_i^T} = [v_{i1} \ \cdots \ v_{ii} \ \cdots v_{iN}]$,

$\boldsymbol{Av_i} = \lambda_i \boldsymbol{v_i}$;

$$\Rightarrow \ \boldsymbol{A}[\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \cdots \quad \boldsymbol{v}_N] = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \cdots \quad \boldsymbol{v}_N] \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & \lambda_i & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & \lambda_N \end{bmatrix}$$

$$\Rightarrow \ \boldsymbol{AV} = \boldsymbol{V\Lambda} \ \Rightarrow \ \boldsymbol{A} = \boldsymbol{V\Lambda V^{-1}} \ \Rightarrow \ \boldsymbol{\Lambda} = \boldsymbol{V^{-1}AV}$$

Diagonalization

- **Self-adjoint (Hermitian)** $\boldsymbol{A}$: $\boldsymbol{A^T = A}$ or $\boldsymbol{A^* = A}$
  - Every eigenvalues are real
  - $\boldsymbol{A}$ has orthonormal eigenvectors

By Schur's Theorem, ∃ an orthonormal basis β $s.t.$
$A = [T]_\beta$ is upper triangular, and
$A^* = [T^*]_\beta = [T]_\beta = A$,
Thus A is a diagonal matrix and so β consists of eigenvectors

$$\boldsymbol{Av_i = \lambda_i v_i;}\ \ \boldsymbol{v_i^T v_j} = \begin{cases} 1 & if\ i = j \\ 0 & otherwise \end{cases}$$

$\boldsymbol{V}$: Orthonomal Matrix for real matrix $\boldsymbol{A}$

$$\boldsymbol{V^T V = I} \Rightarrow \boldsymbol{V^T = V^{-1}} \Rightarrow \boldsymbol{A = V\Lambda V^{-1} = V\Lambda V^T}$$

$\boldsymbol{V}$: Unitary Matrix for complex matrix $\boldsymbol{A}$

$$\boldsymbol{V^* V = I} \Rightarrow \boldsymbol{V^* = V^{-1}} \Rightarrow \boldsymbol{A = V\Lambda V^{-1} = V\Lambda V^*}$$

Eigenvalues and Eigenvectors

- Example

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \rightarrow |A - \lambda I| = (2-\lambda)(4-\lambda) - 3$$
$$\rightarrow \lambda^2 - 6\lambda + 5 = 0 \;\rightarrow\; \lambda = 1, 5$$

$$\lambda = 1: (A - \lambda I)x = 0 \leftrightarrow \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = 0$$

$\therefore$ eigenvector v_1 associated with eigenvalue 1 is $c \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$$\lambda = 5: (A - \lambda I)x = 0 \leftrightarrow \begin{bmatrix} -3 & 1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = 0$$

$\therefore$ eigenvector v_2 associated with eigenvalue 5 is $c \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

Eigenvalues and Eigenvectors

- Example

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}, \lambda = 1, 5$$

$$v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ for } \lambda_1, v_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \text{ for } \lambda_2$$

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = 5 \cdot \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

$$AV = V\Lambda \leftrightarrow A = V\Lambda V^{-1} \leftrightarrow \Lambda = V^{-1}AV$$

Exercise

Find the null space of $\begin{bmatrix} 1 & 0 \\ 5 & 4 \\ 2 & 4 \end{bmatrix}$.

Exercise

Find the null space of $\begin{bmatrix} 1 & 0 \\ 5 & 4 \\ 2 & 4 \end{bmatrix}$.

Sol:

$$\begin{bmatrix} 1 & 0 \\ 5 & 4 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$Solution: \quad u = 0, \quad v = 0$$

$$\text{For } B = \begin{bmatrix} A & 1 \\ & 9 \\ & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 5 & 4 & 9 \\ 2 & 4 & 6 \end{bmatrix},$$

its null space contains the vector (1, 1, -1).

$$N(B) = \{(c,c,-c)\}$$

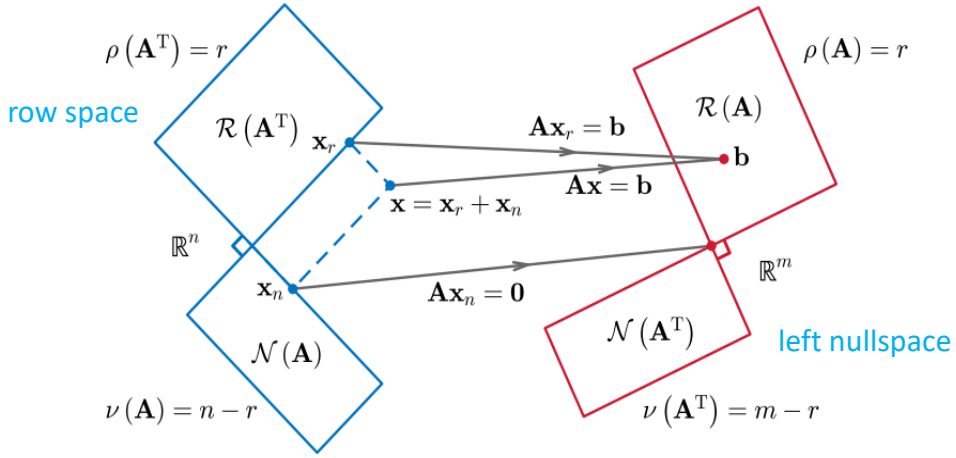
Exercise

The column(range) space of A: $C(A)$ = a subspace spanned by the columns of A

The row space of A: $C(A^T)$ = a subspace spanned by the columns of A^T

→ a subspace spanned by the rows of A

Problem: Find $C(A), C(A^T), N(A), N(A^T)$ of $A = \begin{bmatrix} 1 & -1 & 3 \\ 2 & -2 & 6 \end{bmatrix}$.



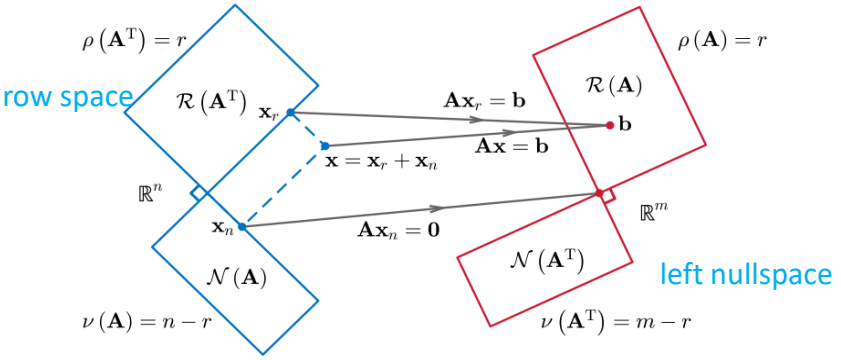
Exercise

Pro.: Find $C(A), C(A^T), N(A), N(A^T)$ of $A = \begin{bmatrix} 1 & -1 & 3 \\ 2 & -2 & 6 \end{bmatrix}$.

Sol.:

$$C(A) = \left\{ c \begin{bmatrix} 1 \\ 2 \end{bmatrix} | c \in R \right\}, \quad \text{dimension} = 1, \text{basis} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

$$C(A^T) = \left\{ c \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} | c \in R \right\}, \quad \text{dimension} = 1, \text{basis} = \left\{ \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} \right\}$$



$$N(A) = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} | x - y + 3z = 0 \right\} = \left\{ s \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} | s, t \in R \right\}, \quad \text{dim.} = 2, \text{basis} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

$$N(A^T) = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} | x + 2y = 0 \right\} = \left\{ s \begin{bmatrix} -2 \\ 1 \end{bmatrix} | s \in R \right\}, \quad \text{dimension} = 1, \text{basis} = \left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}$$

Exercise

Pro.: Find $N(L), C(L)$ of $L = \frac{d}{dt}$.

Sol.:

Exercise

Pro.: Find $N(L), C(L)$ of $L = \frac{d}{dt}$.

Sol.:

$$Lp(t) = \frac{d}{dt}(a_0 + a_1 t + \ldots + a_n t^n)$$
$$= a_1 + \ldots\ldots + n a_n t^{n-1}$$
$$N(L) = \{\text{constants}\}$$
$$C(L) = P_{n-1}$$

Exercise

Pro.: Find $N(L), C(L)$ of $L = \int_0^t p(t)dt$.

Sol.:

$$Lp(t) = \int_0^t (a_0 + a_1 t + \ldots + a_n t^n)dt$$
$$= a_0 t + \ldots\ldots + \frac{a_n}{n+1} t^{n+1}$$
$$N(L) = \{0\}$$
$$C(L) = P_{n+1}$$

Exercise

Pro.: Is it linear transformation to multiply a fixed polynormal to a vector in P_n ?

Sol.:

Exercise

Pro.: Is it linear transformation to multiply a fixed polynormal to a vector in P_n ?

Sol.:

$$L\big(\alpha p_1(t) + \beta p_2(t)\big) = \alpha(2+3t)(a_0 + \ldots + a_n t^n) + \beta(2+3t)(a_0 + \ldots + a_n t^n)$$
$$= (2+3t)[\alpha(a_0 + \ldots + a_n t^n) + \beta(a_0 + \ldots + a_n t^n)] \in C(L) = P_{n+1}$$

Jordan Canonical Form

- Generalized Eigenvector v_{ip} for λ_i: $(A-\lambda_i \mathbf{I})^p v_{ip} = 0$

$$A = \begin{bmatrix} 3 & 1 & -2 \\ -1 & 0 & 5 \\ -1 & -1 & 4 \end{bmatrix} \Rightarrow f(t) = \det(A - t\mathbf{I}) = -(t-3)(t-2)^2$$

$$\lambda_1 = 3 : (A - 3\mathbf{I})v_1 = 0 \Rightarrow v_1 = (-1, 2, 1)^t$$

$$\lambda_2 = 2 : (A - 2\mathbf{I})v_2 = 0 \Rightarrow v_2 = (1, -3, -1)^t \; ; \; only\ one$$

$$(A - 2\mathbf{I})^2 v_3 = 0 \Rightarrow v_3 = (-1, 2, 0)^t \; ; \; generalized\ eigenvector$$

$$V = [v_1, v_2, v_3] = \begin{bmatrix} -1 & 1 & -1 \\ 2 & -3 & 2 \\ 1 & -1 & 0 \end{bmatrix}$$

$$\Rightarrow A = VJV^{-1}, J = V^{-1}AV = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

Singular Value Decomposition(SVD)

- Diagonalization & SVD

Linear Operator	Linear Transformation
$T : X \to X$	$T : X \to Y$
$[T]_\beta \in M_{n\times n}$	$[T]_\beta \in M_{m\times n}$
$eigenvalue$	$singular\ value$
$[T]_\beta = A = V\Lambda V^*$	$[T]_\beta = A = U\Sigma V^*$
$\Lambda = V^* AV$	$\Sigma = U^* AV$

- Assume $\rho(A) = r, then\ \sigma_1 \geq \cdots \geq \sigma_r > 0$

$$\Lambda_{n\times n} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & 0 \\ 0 & \cdots & \lambda_r & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix} \quad | \quad \Sigma_{m\times n} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \sigma_r & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Singular Value Decomposition(SVD)

- Diagonalization & SVD

$$[T]_\beta = A = V\Lambda V^* \quad | \quad [T]_\beta = A = U\Sigma V^*$$

- V & U

$V = [v_1 \quad v_2 \quad \cdots \quad v_n]$ with orthonormal $\{v_i\}$ such that

$$A^*Av_i = \begin{cases} \sigma_i^2 v_i, & 1 \le i \le r \\ 0 & i > r \end{cases}.$$

$U = [u_1 \quad u_2 \quad \cdots \quad u_m]$ with orthonormal $\{u_i\}$ such that

$$Av_i = \begin{cases} \sigma_i u_i, & 1 \le i \le r \\ 0 & i > r \end{cases} \leftarrow u_i = \frac{1}{\sigma_i} Av_i.$$

- Proof of $A = UDV^*$

$$AV = [Av_1 \quad Av_2 \quad \cdots \quad Av_n] = [\sigma_1 u_1 \quad \cdots \quad \sigma_r u_r \quad 0u_{r+1} \quad \cdots \quad 0u_n]$$
$$= U\Sigma$$
$$\Rightarrow A = UDV^*$$

- Diagonalization & SVD

$$\Lambda = V^*AV \quad | \quad \Sigma = U^*AV$$

$$\Sigma_{m\times n} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \sigma_r & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Singular Value Decomposition(SVD)

- Diagonalization & SVD

$$[T]_\beta = A = V\Lambda V^* \quad | \quad [T]_\beta = A = U\Sigma V^*$$

- V & U

$V = [v_1 \quad v_2 \quad \cdots \quad v_n]$ with orthonormal $\{v_i\}$ such that

$$A^*Av_i = \begin{cases} \sigma_i^2 v_i, & 1 \le i \le r \\ 0 & i > r \end{cases}.$$

$U = [u_1 \quad u_2 \quad \cdots \quad u_m]$ with orthonormal $\{u_i\}$ such that

$$Av_i = \begin{cases} \sigma_i u_i, & 1 \le i \le r \\ 0 & i > r \end{cases} \leftarrow u_i = \frac{1}{\sigma_i} Av_i.$$

- Proof of $A = UDV^*$

$$AV = [Av_1 \quad Av_2 \quad \cdots \quad Av_n] = [\sigma_1 u_1 \quad \cdots \quad \sigma_r u_r \quad 0u_{r+1} \quad \cdots \quad 0u_n]$$
$$= U\Sigma$$
$$\Rightarrow A = UDV^*$$

- Diagonalization & SVD

$$\Lambda = V^*AV \quad | \quad \Sigma = U^*AV$$

For $AA^*u_i = \sigma_i^2 u_i$,

Let $v_i = \frac{1}{\sigma_i} A^* u_i$

$$\rightarrow A^* u_i = \sigma_i v_i$$

$$AV = [Av_1 \quad Av_2 \quad \cdots \quad Av_n]$$
$$= \left[\cdots \quad \frac{1}{\sigma_i} AA^*u_i \quad \cdots\right]$$
$$= [\cdots \quad \sigma_i u_i \quad \cdots] = U\Sigma$$

$$\Sigma_{m\times n} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \sigma_r & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Singular Value Decomposition(SVD)

- Example: $A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \rightarrow A^*A = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}, AA^* = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$
- Eigenvalues of AA^*: $\lambda_1 = 12, \lambda_2 = 10 \rightarrow \sigma_1 = \sqrt{12}, \sigma_2 = \sqrt{10}$
- Eigenvectors of AA^*, normalized: $u_1 = \frac{1}{\sqrt{2}}(1,1)^T, u_2 = \frac{1}{\sqrt{2}}(1,-1)^T$
- Eigenvalues ofA^*A: $\lambda_1 = 12, \lambda_2 = 10,\ \lambda_3 = 0 \rightarrow \sigma_1 = \sqrt{12}, \sigma_2 = \sqrt{10}, \sigma_3 = 0$
- Eigenvectors of A^*A, normalized:

$$v_1 = \frac{1}{\sigma_1}A^*u_1 = \frac{1}{\sqrt{12}}\begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}}(1,2,1)^T,$$

$$v_2 = \frac{1}{\sigma_2}A^*u_2 = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{5}}(2,-1,0)^T,$$

$v_3 = \frac{1}{\sqrt{30}}(1,2,-5)^T$ (should calculate v_3 directly) $\rightarrow A = U\Sigma V^*$

Singular Value Decomposition(SVD)

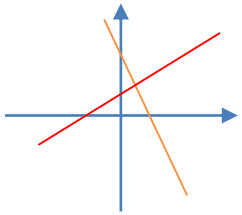
- $A = U\Sigma V^*$

$$A = U\Sigma V^* = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & -\frac{5}{\sqrt{30}} \end{bmatrix}$$

Linear Equation

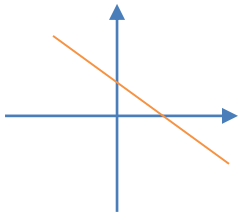
- Linear equation: $Ax = b$
- Invertible linear equation:

$$\begin{bmatrix} 4 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$ →unique solution



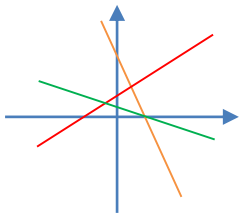
- Non-invertible linear equation:

$$\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \rightarrow 2x_1 + x_2 = 1$$ →many solutions (부정)



→ pseudo inverse (minimum norm)

$$\begin{bmatrix} 4 & 2 \\ 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$ → no solutions (불능)



→ projection (approximation)
→ least squares (minimum error)

Pseudo Inverse

- Non-invertible linear equation (many solution):

$$Ax = b, \qquad \rho(A) < n, \qquad x \in R^n$$

- Solution set:

$$\{x_r + x_n \mid x_r : a\ particular\ solution, x_n \in N(A)\}$$

- Minimal solution (unique) with minimum norm $\|x\|$:

$$x = A^{\dagger}b, \quad A^{\dagger}: \text{pseudo inverse}$$

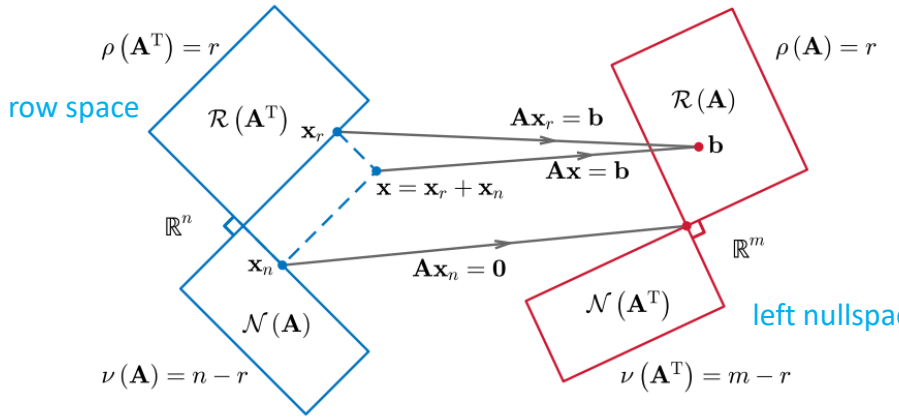
- Pseudo inverse $A^{\dagger}$ of A

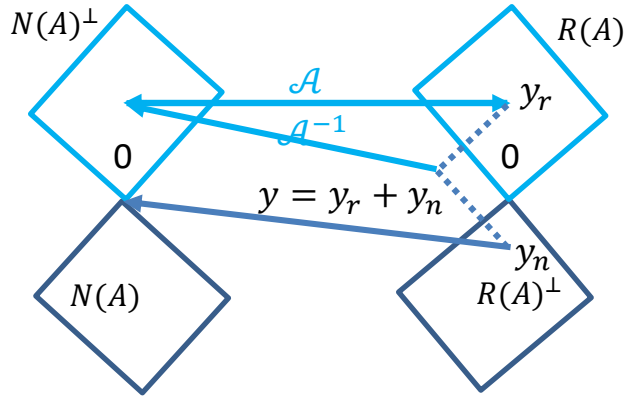
$$A: X \to Y,\ A^{\dagger}: Y \to X,\ \mathcal{A}: N(A)^{\perp} \to R(A),\ \mathcal{A}^{-1}: R(A) \to N(A)^{\perp}$$

$$A^{\dagger}(y) = \begin{cases} \mathcal{A}^{-1}(y) & y \in R(A) \\ 0 & y \in R(A)^{\perp} \end{cases}$$

$$A^{\dagger}(y_r + y_n) = \mathcal{A}^{-1}(y_r)$$

$$\rightarrow A^{\dagger} = V\Sigma^{\dagger}U^{*}, \quad A = U\Sigma V^{*}$$





Pseudo Inverse

- Example: Find x for $Ax = b$, where $A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
- Sol.

$$A = U\Sigma V^* = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & -\frac{5}{\sqrt{30}} \end{bmatrix}$$

$$A^\dagger = V\Sigma^\dagger U^* = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{6} & 0 & -\frac{5}{\sqrt{30}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{12}} & 0 \\ 0 & \frac{1}{\sqrt{10}} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$x = A^\dagger b$$

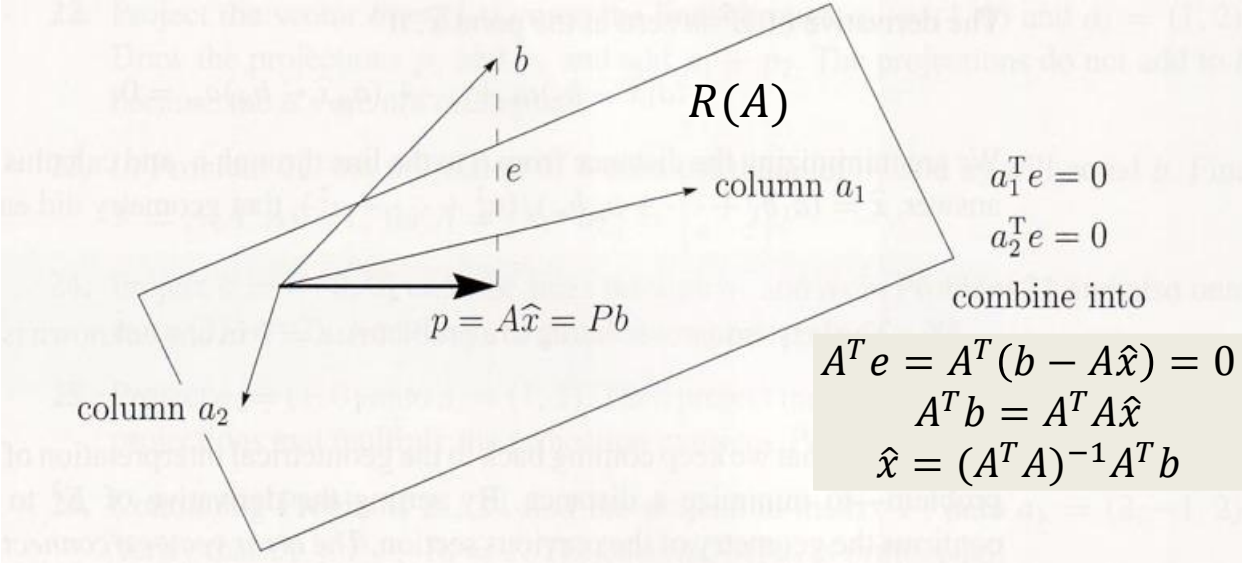
Projection

- Non-invertible linear equation (no solution):

$$Ax = b, \qquad b \notin R(A)$$

$\rightarrow Ax - b \neq 0$

$\rightarrow$ What is the best solution minimizing errors between Ax and b?



Least Squares

- Non-invertible linear equation (no solution):

$$Ax = b, \qquad b \notin R(A)$$

$\rightarrow Ax - b \neq 0$

$\rightarrow$ What is the best solution minimizing errors between Ax and b?

- Optimization and Gradient

$$\hat{x} = arg\min_x \|Ax - b\|_2^2 \triangleq E(x)$$

$$\nabla_x E(x) = \frac{dE(x)}{dx} = \frac{d}{dx}(Ax - b)^T(Ax - b) = 2A^T(Ax - b) = 0$$

(??)

$$\hat{x} = (A^T A)^{-1} A^T b$$

Vector Derivative

Identities: vector-by-vector $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

Condition	Expression	Numerator layout, i.e. by y and $\mathbf{x}^\mathsf{T}$	Denominator layout, i.e. by $\mathbf{y}^\mathsf{T}$ and x
a is not a function of **x**	$\frac{\partial \mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{0}$	
	$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{I}$	
A is not a function of **x**	$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{A}$	$\mathbf{A}^\top$
A is not a function of **x**	$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} =$	$\mathbf{A}^\top$	$\mathbf{A}$
a is not a function of **x**, **u** = **u**(**x**)	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	
v = *v*(**x**), **a** is not a function of **x**	$\frac{\partial v\mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{a}\frac{\partial v}{\partial \mathbf{x}}$	$\frac{\partial v}{\partial \mathbf{x}}\mathbf{a}^\top$
v = *v*(**x**), **u** = **u**(**x**)	$\frac{\partial v\mathbf{u}}{\partial \mathbf{x}} =$	$v\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}\frac{\partial v}{\partial \mathbf{x}}$	$v\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}\mathbf{u}^\top$
A is not a function of **x**, **u** = **u**(**x**)	$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{A}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}^\top$
u = **u**(**x**), **v** = **v**(**x**)	$\frac{\partial (\mathbf{u}+\mathbf{v})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	
u = **u**(**x**)	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
u = **u**(**x**)	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$

Derivative of Matrix-Vector Function

- Gradient of $f(x)$: $\nabla_x f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$
- Hessian of $f(x)$: $H(f)$

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\, \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\, \partial x_n} \\ \frac{\partial^2 f}{\partial x_2\, \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\, \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\, \partial x_1} & \frac{\partial^2 f}{\partial x_n\, \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix},$$

Derivative of Matrix-Vector Function

- Jacobian of $\mathbf{f}(x)$: First derivative of a vector function $\mathbf{f}(x) = [f_1(x), \dots, f_m(x)]^T$

$$\mathbf{f}(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \rightarrow \quad \mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

- Example

$$\mathbf{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix} = \begin{bmatrix} x^2 y \\ 5x + \sin y \end{bmatrix}$$

$$\mathbf{J}_{\mathbf{f}}(x,y) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2xy & x^2 \\ 5 & \cos y \end{bmatrix}$$

Derivative of Matrix-Vector Function

- Gradient of $f(x)$: $\nabla_x f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$

- Hessian of $f(x)$: $H(f)$

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix},$$

- Jacobian of $\mathbf{f}(x)$: $\mathbf{f}(x) = [f_1(x), \ldots, f_m(x)]^T$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

$$H(f) = J \cdot \nabla_x f(x)$$

Exercise

- For $f(x) = 21x_1^2 + 14x_1x_2 + 6x_2^2 + 2x_1 + x_2$
  find Gradient $\nabla_x f$, Hessian $H(f)$.

Exercise

- For $f(x) = 21x_1^2 + 14x_1x_2 + 6x_2^2 + 2x_1 + x_2$
  find Gradient $\nabla_x f$, Hessian $H(f)$.

- Gradient
  $$\nabla_x f(\mathrm{x}) = \begin{bmatrix} 42x_1 + 14x_2 + 2 \\ 14x_1 + 12x_2 + 1 \end{bmatrix}$$

- Hessian
  $$H = \begin{bmatrix} 42 & 14 \\ 14 & 12 \end{bmatrix}$$

Derivative of Matrix-Vector Function

- Jacobian of $\mathbf{f}(x)$: First derivative of a vector function $\mathbf{f}(x) = [f_1(x), \dots, f_m(x)]^T$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

- Example: $\mathbf{f}(x) = Ax, A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \rightarrow \frac{d}{dx} Ax = ?$

Derivative of Matrix-Vector Function

- Jacobian of $\mathbf{f}(x)$: First derivative of a vector function $\mathbf{f}(x) = [f_1(x), \dots, f_m(x)]^T$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

- Example: $\mathbf{f}(x) = Ax, A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \rightarrow \frac{d}{dx} Ax = ?$

$$\mathbf{f}(x) = Ax = \begin{bmatrix} x_1 + 2x_2 \\ 4x_1 + 3x_2 \end{bmatrix}$$

$$\mathbf{J}[\mathbf{f}(x)] = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \Rightarrow \boxed{\frac{d}{dx} Ax = A}$$

Derivative of Matrix-Vector Function

- Example: $\mathbf{f}(x) = x^T A^T, A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \rightarrow \frac{d}{dx} x^T A^T = ?$

Derivative of Matrix-Vector Function

- Example: $\mathbf{f}(x) = x^T A^T, A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \rightarrow \frac{d}{dx} x^T A^T = ?$

$$\mathbf{f}(x) = [x_1 \quad x_2] \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} = [x_1 + 2x_2 \quad 4x_1 + 3x_2]$$

$$\mathbf{J}[\mathbf{f}(x)] = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} \Rightarrow \boxed{\frac{d}{dx} x^T A^T = (\frac{d}{dx} Ax)^T = A^T}$$

- Jacobian of $\mathbf{f}(x)$: First derivative of a vector function $\mathbf{f}(x) = [f_1(x), \dots, f_m(x)]^T$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \Rightarrow \boxed{\mathbf{J}(\mathbf{f}^T) = [\mathbf{J}(\mathbf{f})]^T}$$

Derivative of Matrix-Vector Function

- How to find derivative of matrix-vector form or $x = (x_1, x_2)$

  $f(x) = \|Ax\|_2^2 + b^T x$, where $A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

- Gradients

  $\nabla_x f(x) = \frac{d}{dx}[(Ax)^T(Ax) + b^T x] = A^T Ax + [(Ax)^T A]^T + b = 2A^T Ax + b$

Derivative of Matrix-Vector Function

- How to find derivative of matrix-vector form or $x = (x_1, x_2)$

  $f(x) = \|Ax\|_2^2 + b^T x$, where $A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

- Gradients

  $\nabla_x f(x) = \frac{d}{dx}[(Ax)^T(Ax) + b^T x] = A^T Ax + [(Ax)^T A]^T + b = 2A^T Ax + b$

- Check!

$$f(x) = (Ax)^T(Ax) + b^T x = \begin{bmatrix} x_1 + 2x_2 \\ x_1 + x_2 \end{bmatrix}^T \begin{bmatrix} x_1 + 2x_2 \\ x_1 + x_2 \end{bmatrix} + x_1 + 2x_2$$

$$f(x) = 2x_1^2 + 6x_1x_2 + 5x_2^2 + x_1 + 2x_2$$

$$\nabla_x f(x) = \begin{bmatrix} \partial f(x)/\partial x_1 \\ \partial f(x)/\partial x_2 \end{bmatrix} = \begin{bmatrix} 4x_1 + 6x_2 + 1 \\ 6x_1 + 10x_2 + 2 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 6 \\ 6 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 2A^T Ax + b$$

Hessian of $f(x)$

$$H = J[\nabla_x f(x)] = \begin{bmatrix} 4 & 6 \\ 6 & 10 \end{bmatrix} = 2A^T A$$

Derivative of Matrix-Vector Function

- More examples

  - $f(x) = x^T Ax$

    $\nabla_x f(x) = Ax + A^T x$

    $= 2Ax$ if A is symmetric.

  - $f(x) = \|Ax - b\|_2^2 + c^T x,\ f(x) = (Ax - b)^T (Ax - b) + c^T x$

    $\nabla_x f(x) = \nabla_x (x^T A^T - b^T)(Ax - b) + c$

    $= A^T (Ax - b) + [(x^T A^T - b^T)A]^T + c$

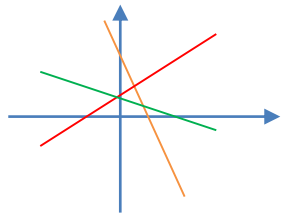
    $= A^T (Ax - b) + A^T (Ax - b) + c$

    $= 2A^T (Ax - b) + c$

Least Squares

- Non-invertible linear equation (no solution):

$$Ax = b, \qquad b \notin R(A)$$



$$\begin{bmatrix} 4 & 2 \\ 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \rightarrow$$ no solutions (불능) (sol by projection)

$\rightarrow Ax - b \neq 0$

$\rightarrow$ What is the best solution minimizing errors between Ax and b?

- Optimization and Gradient

$$\hat{x} = arg\min_x \|Ax - b\|_2^2 \triangleq E(x)$$

$$\nabla_x E(x) = \frac{dE(x)}{dx} = \frac{d}{dx}(Ax - b)^T(Ax - b) = 2A^T(Ax - b) = 0$$

$$\hat{x} = (A^T A)^{-1} A^T b$$

$$A^* b - A^* A\hat{x} = 0$$
$$A^* A\hat{x} = A^* b$$
$$\hat{x} = (A^* A)^{-1} A^* b = A^\dagger b$$

$$A^T b - A^T A\hat{x} = 0$$
$$A^T A\hat{x} = A^T b$$
$$\hat{x} = (A^T A)^{-1} A^T b = A^\dagger b$$

Positive Definite Matrices

- Definitness of Matrix $f(x) = x^T Hx + b^T x$
  - If $H > 0$ is positive definite matrix, all $\lambda_i(H) > 0$
  - If $H \geq 0$ is positive semi-definite matrix, all $\lambda_i(H) \geq 0$
  - If $H < 0$ is negative definite matrix, all $\lambda_i(H) < 0$
  - If $H \leq$ is negative semi-definite matrix, all $\lambda_i(H) \leq 0$
  - If H is indefinite matrix, some $\lambda_i(H) > 0$, some $\lambda_i(H) < 0$

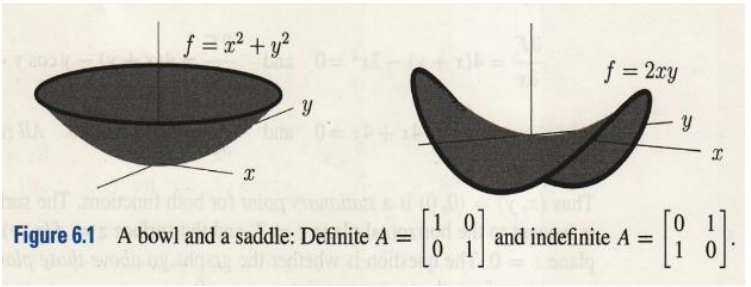


Figure 6.1 A bowl and a saddle: Definite $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and indefinite $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

- Example

$$f = \frac{1}{2}(x^2 + y^2), \quad H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; v^T Hv = v_1^2 + v_2^2 > 0 \rightarrow H > 0$$

$$f = \frac{1}{2}(x^2 + 2xy + y^2), \; H = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \; v^T Hv = (v_1 + v_2)^2 \geq 0$$
$$\rightarrow H \geq 0 \rightarrow f = (x + y)^2 \text{ has minimum at } x = -y$$

$$f = xy, \qquad H = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \; v^T Hv = 2v_1 \, v_2 \; \not\leq \not\geq 0 \rightarrow H \not\leq \not\geq 0$$

Positive Definite Matrices

- Optimization problem

$$\min_{(x,y)} f(x,y), \quad \text{where } f(x,y) = [x\ y]\begin{bmatrix}2 & 2\\ 2 & 3\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} + \begin{bmatrix}-2\\ 0\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}$$

- Necessary condition

$$\nabla_{(x,y)} f(x,y) = \begin{bmatrix}4 & 4\\ 4 & 6\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} - \begin{bmatrix}2\\ 0\end{bmatrix} = 0$$

$$\begin{bmatrix}4 & 4\\ 4 & 6\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}2\\ 0\end{bmatrix} \rightarrow \begin{bmatrix}4 & 4\\ 0 & -2\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}2\\ 2\end{bmatrix} \rightarrow y = -1, x = \frac{3}{2}$$

- Necessary and Sufficient condition (Convexity)

$$H(f) = \begin{bmatrix}4 & 4\\ 4 & 6\end{bmatrix} > 0 \ \leftarrow \ [v_1 \quad v_2]\begin{bmatrix}4 & 4\\ 4 & 6\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix} = 4v_1^2 + 8v_1v_2 + 6v_2^2$$

$$\nabla_{(x,y)} f(x,y) = 0, \qquad = 4(v_1 + v_2)^2 + 2v_2^2 > 0$$

- A is positive definite matrix, i.e., $A > 0$ iff $v^T A v > 0$ for $\forall v \neq 0$.
- A is positive semi-definite matrix, i.e., $A \geq 0$ iff $v^T A v \geq 0$ for $\forall v \neq 0$.

Trace

- $tr(A) = \sum_{i=1}^{n} A_{ii}$.

- $tr(A + B) = tr(A) + tr(B)$
  $tr(A) = tr(A^T)$
  $tr(A^T B) = tr(AB^T) = tr(B^T A) = tr(BA^T)$

Derivative of Trace

$$\frac{d}{dX} tr(AXB) = ?$$

Derivative of Trace

$$\frac{d}{dX}tr(AXB)=\frac{d}{dX}tr(BAX) = (BA)^T = A^T B^T \ (m \times p)$$

$$n \times m \cdot m \times p \cdot p \times n \rightarrow p \times n \cdot n \times m \cdot m \times p$$

Summary

basics
- matrix and vector operation
- Gaussian elimination: inverse of matrix
- linear equations

vector space
- def., dimension, basis, subspace
- linear combination, linearly independent
- vector representation

linear transformation
- matrix representation of LT
- null space and range space
- solution set of linear equation
- neural transformation

useful tricks
- vector norm, matrix norm
- eigenvalue and eigenvector
- singular value decomposition
- pseudo inverse
- derivative of vector function