Mathematics in DL: **Jan 23 2018**

Lecture 2: Applied Linear Algebra

Lecturer: Juhee Lee *Scribes: Juhee Lee*

2.1 Review vectors, Matrices, Operations

지난 시간에는 기본적인 용어와 연산에 관하여 살펴보았습니다. 오늘 및 앞으로의 강의에서 사용되는 예제들은 cs231n을 기반으로 구성하였습니다.

vectors

For a vector $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$,

- vector의 성분은 행렬, 함수 등 같은 형식의 성분이 들어오도록 합니다.
- vector는 기본적으로 크기와 방향을 갖고 있습니다.

- vector의 크기는 1강에서의 $\|\cdot\|$ (norm) 에 의해 측정가능 하고 방향은 화살표로 표시합니다.

  ex. $\mathbf{a} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, 크기와 방향이 같으면 동일 벡터 i.e., $\mathbf{a} = \mathbf{b}$

- 연산의 정의에 따라 여러가지가 가능합니다

  1. $\mathbf{a} + \mathbf{b} = (a_1 + b_1, \cdots, a_n + b_n)$
  2. "scalar multiplication"
     $c \in \mathbb{R}, c\mathbf{a} = c\,(a_1, \cdots, a_n) = (ca_1, \cdots, ca_n)$
  - "scalar multiplication"[1] 의 기하학(geometry)적인 의미 "scaling"
  3. "dot product (Euclidean inner product)" ($\cos\theta$를 이용한 $\angle$ 각도 도입 : 하나의 실수값)

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= (a_1 b_1 + \cdots + a_n b_n) = \sum_{i=1}^{n} a_i b_i \\ &= \|\mathbf{a}\| \|\mathbf{a}\| \cos\theta \ \in \mathbb{R} \end{aligned}$$

note. $\cos 0° = 1$, $\cos 90° = 0$ $\cos\theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{a}\|}$

- If $\mathbf{a} \cdot \mathbf{b} = 0$, $\mathbf{a}$ 와 $\mathbf{b}$는 서로 직교(**orthogonal**)한다.

[1] scalar c 는 벡터공간에 따라 실수 혹은 복소수를 나타낼 수 있습니다. 이는 벡터공간의 정의에 따라 달라집니다. 실수에 대하여 정의된 공간은 실 벡터 공간이라고 부르고 이중 Euclidean space를 다룹니다(가장 익숙하고 쉬운 공간이므로). 복소수를 scalar로 취하는 벡터공간은 Quantum mechanics (양자역학)등에서 자주 다루는 공간입니다.

- $\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$ is an **orthogonal set** if

$$\mathbf{a}_i \neq 0 \text{ for each } i \text{ and } \mathbf{a}_i \cdot \mathbf{a}_j = 0 \text{ for all } i, j$$

- $\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$ is an **orthonormal set** if it is an orthogonal set consisting of unit vectors[2]

Theorem 2.1 *If $\boldsymbol{a}$ is any vector in an inner produce space[3], then*

$$\left.\begin{array}{l} 1.\ \boldsymbol{a}\cdot\boldsymbol{a} \geq 0 \textit{ for all vectors } \boldsymbol{a} \\ 2.\ \boldsymbol{a}\cdot\boldsymbol{a} = 0 \textit{ iff } \boldsymbol{a} = \boldsymbol{0} \end{array}\right\} \textit{Positive-definiteness}$$

3. $\boldsymbol{0}\cdot\boldsymbol{a} = \boldsymbol{0} = \boldsymbol{0}\cdot\boldsymbol{a}$

Theorem 2.2 *Cauchy-Schwarz Inequality*

$$|\boldsymbol{a}\cdot\boldsymbol{b}| \leq \|\boldsymbol{a}\|\|\boldsymbol{b}\|$$

Theorem 2.3 *Triangle Inequality*

$$\|\boldsymbol{a}+\boldsymbol{b}\| \leq \|\boldsymbol{a}\| + \|\boldsymbol{b}\|$$

Definition 2.4 *For n vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ and n scalars $c_1, \ldots, c_n \in \mathbb{R}$,*

- $c_1\boldsymbol{a}_1 + \cdots + c_n\boldsymbol{a}_n$ 을 *vectors* $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ 의 일차결합 ***(linear combination)***이라 합니다.
- *For* $S = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$, ***span***$(S) = c_1\boldsymbol{a}_1 + \cdots + c_n\boldsymbol{a}_n \ |c_1, \ldots, c_n \in \mathbb{R}\}$
- $c_1\boldsymbol{a}_1 + \cdots + c_n\boldsymbol{a}_n = 0$을 만족하는 $c_1 = \cdots = c_n = 0$이 유일할 경우, $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$을 일차 독립 ***(linearly independent)***라고 합니다. 그이외의 경우를 일차 종속 ***(linearly dependent)*** 라고 합니다.

Quiz. 일차종속 집합의 예를 구하여라.

Note. The **rank** of a matrix $\mathbf{A}$ is the maximum number of **linearly independent** row vectors of $\mathbf{A}$.

- n개의 행벡터들로 구성된 matrix $\mathbf{A}$의 **rank**가 n이면 n개의 행벡터들은 일차 독립입니다. 그러나, matrix $\mathbf{A}$의 **rank**가 n 보다 작으면, n개의 행벡터들은 일차 종속입니다.

Matrices

Quiz. 두 행렬 $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ 와 $\mathbf{B} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$의 $\mathbf{AB}$와 $\mathbf{BA}$를 구하여라.
Solution. $\mathbf{AB} = \mathbf{O}$, $\mathbf{BA} = \mathbf{B}$

Note. $\mathbf{AB} \neq \mathbf{BA}$, $\quad \mathbf{AB} = \mathbf{O} \not\Rightarrow \mathbf{A} = \mathbf{O}$ *or* $\mathbf{B} = \mathbf{O}$

[2] a vector of norm 1, we can normalize $\mathbf{a}$ using $\frac{\mathbf{a}}{\|\mathbf{a}\|}$
[3] vector space defined by inner product

Quiz. 행렬 $\mathbf{A} = \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix}$ 와 $\mathbf{B} = \begin{pmatrix} 1 & 0 & 2 \\ 3 & -2 & 1 \end{pmatrix}$, $\mathbf{C} = \begin{pmatrix} -1 & 0 \\ 3 & 1 \\ 2 & 4 \end{pmatrix}$ 의 $\mathbf{ABC}$를 구하여라.

Solution. $\mathbf{ABC} = \begin{pmatrix} 17 & 4 \\ 13 & 14 \end{pmatrix}$

Quiz. 행렬 $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 0 & -2 \\ -2 & -1 \end{pmatrix}$ 에 대하여 $\mathbf{AA^T}$를 구하여라.

Solution. $\mathbf{AA^T} = \begin{pmatrix} 10 & -6 & -5 \\ -6 & 4 & 2 \\ -5 & 2 & 5 \end{pmatrix}$ symmetric matrix

Quiz. 행렬 $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 0 & -2 \\ -2 & -1 \end{pmatrix}$, 행렬 $\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ 에 대하여 $\mathbf{AX}$를 구하여라.

Answer. $\mathbf{AX} = \begin{pmatrix} x_1 + 3x_2 \\ -2x_2 \\ -2x_1 - x_2 \end{pmatrix}$ matrix×matrix / matrix × vector

2.2 Linear Transformation

Definition 2.5 두 벡터공간[4] $\mathbb{R}^n$, $\mathbb{R}^m$ 사이에 정의된 함수 $f : \mathbb{R}^n \to \mathbb{R}^m$ 가 다음 성질을 만족할때 선형변환 *(linear Transformation)* 이라고 합니다.

For all $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ *and any scalar* $c \in \mathbb{R}$, $f(\boldsymbol{a} + \boldsymbol{b}) = f(\boldsymbol{a}) + f(\boldsymbol{b})$, $f(c\boldsymbol{a}) = cf(\boldsymbol{a})$.

- 모든 선형변환은 다음을 만족합니다.
  1. $f(\mathbf{0}) = \mathbf{0}$
  2. $f(c_1\mathbf{a} + c_2\mathbf{b}) = c_1 f(\mathbf{a}) + c_2 f(\mathbf{b})$ for $c_1, c_2 \in \mathbb{R}$
- The linear transformation given by a matrix

  Let $\mathbf{A} = (a_{ij})_{m\times n}$ be a matrix. The function f defined by $f(\mathbf{x}) = \mathbf{Ax}$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$

- ex. Given $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 0 & -2 \\ -2 & -1 \end{pmatrix}$, we define a function $f : \mathbb{R}^2 \to \mathbb{R}^3$ by $f(\mathbf{x}) = \mathbf{Ax}$.

- $\mathbf{v} = (x, y) = (r\cos\alpha,\ r\sin\alpha)$,

  "polar coordinates in $\mathbb{R}^2$", $(r, \alpha) = (\pm\sqrt{x^2 + y^2},\ \tan^{-1}\frac{y}{x} + n\pi)$

- Examples
  1. "Rotation in $\mathbb{R}^2$"

     $f(x, y) = (x\cos\theta - y\sin\theta,\ x\sin\theta + y\cos\theta)$ for any θ

[4] vector space is a set that satisfies axioms ; addition (closure, commutative and associative property, identity and inverse), scalar multiplication (closure, distributive and associative property and scalar identity)
ex. $\mathbb{R}^n$=set of all n-tuples, $\mathbb{C}(-\infty, \infty)$, $M_{m,n}$=set of all $m \times n$ matrices

$$\begin{aligned} f(\mathbf{v}) &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} r\cos\alpha \\ r\sin\alpha \end{pmatrix} \\ &= \begin{pmatrix} r\cos\alpha\cos\theta - r\sin\alpha\sin\theta \\ r\cos\alpha\sin\theta + r\sin\alpha\cos\theta \end{pmatrix} \\ &= \begin{pmatrix} r\cos(\theta+\alpha) \\ r\sin(\theta+\alpha) \end{pmatrix} \end{aligned}$$

Quiz. "Reflection in $\mathbb{R}^2$" f를 구하여라.
Solution. $f(x,y) = (x,-y)$

Quiz. "Projection in $\mathbb{R}^2$" f를 구하여라.
Solution. $f(x,y) = (x,0)$ or $f(x,y) = (0,y)$

Quiz. "A linear transformation from $\mathbf{A}_{m\times n}$ to $\mathbf{A}_{n\times m}$" f를 구하여라.
Solution. $f(\mathbf{A}) = \mathbf{A}^T$

Quiz. "Expansions and Contractions in $\mathbb{R}^2$" f를 구하여라.
Solution. $f(x,y) = (kx,y)$ or $f(x,y) = (x,ky)$ Contraction$(0<k<1)$, Expansion$(k>1)$

Composition of linear transformation

"product" (matrix multiplication)

$\mathbf{AB} = (a_{ij})_{m\times p}(b_{ij})_{p\times n} = (c_{ij})_{m\times n}$ where $c_{ij} = \sum_{k=1}^{p} a_{ik}b_{kj} = a_{i1}b_{1j} + \cdots + a_{in}b_{nj} = \mathbf{a}_i^T \cdot \mathbf{b}_j$

For $f_{\mathbf{A}} : \mathbb{R}^p \to \mathbb{R}^m,\ f_{\mathbf{B}} : \mathbb{R}^n \to \mathbb{R}^p$,

$$f_{\mathbf{A}} \circ f_{\mathbf{B}} = f_{\mathbf{AB}} : \mathbb{R}^n \to \mathbb{R}^m,\ f_{\mathbf{AB}}(\mathbf{x}) = \mathbf{ABx}$$

- For $f_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^n,\ f_{\mathbf{B}} : \mathbb{R}^n \to \mathbb{R}^n$,

$$f_{\mathbf{A}} \circ f_{\mathbf{B}}(\mathbf{x}) = f_{\mathbf{A}}(f_{\mathbf{B}}(\mathbf{x})) = \mathbf{x} \ \text{ and } \ f_{\mathbf{B}} \circ f_{\mathbf{A}}(\mathbf{x}) = f_{\mathbf{B}}(f_{\mathbf{A}}(\mathbf{x})) = \mathbf{x}$$

Then $f_{\mathbf{B}}(f_{\mathbf{A}})$ is the inverse of $f_{\mathbf{A}}(f_{\mathbf{B}})$

2.3 Application of Identity matrix and Inverse matrix

Review of Identity matrix and Inverse matrix

$\mathbf{I}_n$ is the identity matrix of n

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

The matrix inverse of $\mathbf{A}$ is denoted as $\mathbf{A}^{-1}$ and is defined as the matrix such as

$$\mathbf{AB} = \mathbf{I}_n = \mathbf{BA}, \quad \mathbf{B} = \mathbf{A}^{-1}$$

*역행렬은 존재하면 유일!

Inverse matrix 활용

- For two linear transformations $f_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^n$, $f_{\mathbf{B}} : \mathbb{R}^n \to \mathbb{R}^n$,

  $$f_{\mathbf{A}} \circ f_{\mathbf{B}}(\mathbf{x}) = f_{\mathbf{A}}(f_{\mathbf{B}}(\mathbf{x})) = \mathbf{x} \text{ and } f_{\mathbf{B}} \circ f_{\mathbf{A}}(\mathbf{x}) = f_{\mathbf{B}}(f_{\mathbf{A}}(\mathbf{x})) = \mathbf{x}$$

  Then $f_{\mathbf{B}}(f_{\mathbf{A}})$ is the inverse of $f_{\mathbf{A}}(f_{\mathbf{B}}) \implies \mathbf{AB} = \mathbf{I}_n = \mathbf{BA}, \quad \mathbf{B} = \mathbf{A}^{-1}$

- Example. $f(x, y, z) = (2x + 3y + z,\ 3x + 3y + z,\ 2x + 4y + z)$

  The matrix $\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 \\ 3 & 3 & 1 \\ 2 & 4 & 1 \end{pmatrix}$ for $f \implies \exists \mathbf{A}^{-1} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 6 & -2 & -3 \end{pmatrix}$,

  f is invertible and $f^{-1}(x, y, z) = (-x + y,\ -x + z,\ 6x - 2y - 3z)$

- For $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{2n-1} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn-1} & a_{mn} \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$

  linear system $\mathbf{Ax} = \mathbf{b}$[5] 의 해[6]를 구하는 방법!

  If $m = n$ and $\mathbf{A}^{-1}$ 가 존재하면, 해는 $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

  Quiz.

  $$\begin{array}{rrcr} x - 2y & +3z & = & 9 \\ -1 + 3y & & = & -4 \\ 2x - 5y & +5z & = & 17 \end{array}$$

  Solution. $x = 1, y = -1, z = 2$

  If not, $\mathbf{A}^{-1}$ 를 구할 수 없는 rectangular matrix($m \neq n$) 에서, $\mathbf{Ax} = \mathbf{b}$ 의 least square problem (find a vector $\mathbf{x}$ that minimize $\|\mathbf{Ax} - \mathbf{b}\|$) 은 normal equation 을 이용해서 해결합니다.

  - "Linear regression: Finding the line that best fits a set of data points"
    Example. 세 점$(1, 2), (2, 1), (3, 3)$을 지나는 선형방정식 $(y = mx + b)$을 찾는 문제.(Curve fitting)

    $$m + b = 2,\ 2m + b = 1,\ 3m + b = 3 \text{ i.e.}^{7},\ \mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} m \\ b \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

---

[5] $f_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$, Domain of $f_{\mathbf{A}}$ is $\mathbb{R}^n$, Range $R(f_{\mathbf{A}}) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n\}$, kernel or null space $N(f_{\mathbf{A}}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\}$.

[6] solution set $\{\mathbf{x} \mid \mathbf{Ax} = \mathbf{b}\}$

[7] id est 라틴어 약자, 영어로는 that is, 즉

"least square problem : Find a vector $\mathbf{x}$ that minimize $\|\mathbf{Ax}-\mathbf{b}\|$ "

Note. A vector $\hat{\mathbf{x}}$ is a solution to the normal equation $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ if and only if it is a least square solution to $\mathbf{Ax} = \mathbf{b}$

Solution.

$$\begin{aligned}
\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} m \\ b \end{pmatrix} &= \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \\
\begin{pmatrix} 14 & 6 \\ 6 & 3 \end{pmatrix}\begin{pmatrix} m \\ b \end{pmatrix} &= \begin{pmatrix} 13 \\ 6 \end{pmatrix} \\
\begin{pmatrix} m \\ b \end{pmatrix} &= \frac{1}{6}\begin{pmatrix} 3 & -6 \\ -6 & 14 \end{pmatrix}\begin{pmatrix} 13 \\ 6 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 3 \\ 6 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}
\end{aligned}$$

- For $\mathbf{Ax} = \mathbf{b}$, $x_1\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2\begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n\begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$ 으로도 표현될 수 있다.

Review of determinant of a square($n \times n$) matrix $\mathbf{A} = (a_{ij})_{n\times n}$

is defined as

$$\det(\mathbf{A}) = |A| = \sum_{j=1}^{n} \mathbf{a}_{1j}\mathbf{C}_{1j} = \mathbf{a}_{11}\mathbf{C}_{11} + \mathbf{a}_{22}\mathbf{C}_{22} + \cdots + \mathbf{a}_{nn}\mathbf{C}_{nn}$$

where the minor $\mathbf{M}_{ij}$ of the element $\mathbf{a}_{ij}$ is the determinant of the matrix obtained by deleting the ith row and jth column of $\mathbf{A}$. The cofactor $\mathbf{C}_{ij}$ is given by $\mathbf{C}_{ij} = (-1)^{i+j}\mathbf{M}_{ij}$

2.4 Application of Norm

Review of Norm

$$f(x) = 0 \Rightarrow x = 0,\ f(x+y) \le f(x) + f(y),\ f(ax) = |a|x$$

- vector

  $\|\mathbf{a}\|_1 = \sum_{i=1}^{n} |a_i|$

  $\|\mathbf{a}\|_\infty = \max_{1\le i\le n} |a_i|$

  $\|\mathbf{a}\|_2 = \left(\sum_{i=1}^{n} |a_i|^2\right)^{1/2}$

- matrix $\mathbf{A}_{m\times n}$

  $\|\mathbf{A}\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|$ maximum column norm

  $\|\mathbf{A}\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|$ maximum row norm

$\|\mathbf{A}\|_2 = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} = (trace(A^*A))^{1/2}$
where A^* =conjugate transpose, **Frobenius norm** $\|\mathbf{A}\|_2 = \|\mathbf{A}\|_F$

$\|\mathbf{A}\|_{2^*} = \sqrt{\lambda_{\max}(A^TA)}$ **spectral norm** (where $\lambda_{\max}(A^TA)$ is the largest eigenvalue[8] of A^TA)

ex. $\mathbf{A} = \begin{pmatrix} 1 & 3 & -6 \\ -2 & 4 & 2 \\ 2 & 1 & -1 \end{pmatrix}$

$\|\mathbf{A}\|_1 = \max(5, 8, 9) = 9$

$\|\mathbf{A}\|_\infty = \max(10, 8, 4) = 10$

$\|\mathbf{A}\|_2 = \sqrt{76} \approx 8.718$

$\|\mathbf{A}\|_{2^*} = \sqrt{\max\{3.07, 23.86, 49.06\}} \approx 7.0045$

행렬의 norm은 행렬에서의 작은 변화에 대한 선형시스템의 민감도를 추정하는데 사용될수 있다.

Example. Sensitivity of linear equations to data error

- Error analysis: Original linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{b} \in \mathbb{R}^n$$

1. linear system with $\Delta\mathbf{A} \in \mathbb{R}^{n\times n}, \Delta\mathbf{b} \in \mathbb{R}^n$,

$$(\mathbf{A} + \Delta\mathbf{A})\tilde{\mathbf{x}} = \mathbf{b} + \Delta\mathbf{b}$$

then the error is $\Delta\mathbf{x} = \tilde{\mathbf{x}} - \mathbf{x}$ i.e., $\tilde{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$

2. linear system with $\Delta\mathbf{b} \in \mathbb{R}^n$,

$$\mathbf{A}(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b}$$

- subtracting we have $\mathbf{A}\Delta\mathbf{x} = \Delta\mathbf{b}$, or $\Delta\mathbf{x} = \mathbf{A}^{-1}\Delta\mathbf{b}$
- take norms: $\|\Delta\mathbf{x}\| = \|\mathbf{A}^{-1}\Delta\mathbf{b}\| \le \|\mathbf{A}^{-1}\|\|\Delta\mathbf{b}\|$
- to estimate relative error, $\|\mathbf{b}\| = \|\mathbf{A}\mathbf{x}\| \le \|\mathbf{A}\|\|\mathbf{x}\| \implies \frac{1}{\|\mathbf{x}\|} \le \|\mathbf{A}\|\frac{1}{\|\mathbf{b}\|}$
- we have $\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \|\mathbf{A}\|\|\mathbf{A}^{-1}\|\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} = \kappa(\mathbf{A})\ \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}$

- $\kappa(\mathbf{A})$ is the condition number of a matrix($\mathbf{A}$)

- If $\kappa(\mathbf{A})$ is small, the linear system is well conditioned. Otherwise, it is ill conditioned
- $\kappa_\infty(\mathbf{A}) = \|\mathbf{A}\|_\infty\|\mathbf{A}^{-1}\|_\infty = \left\|\begin{pmatrix} 1 & 3 & -6 \\ -2 & 4 & 2 \\ 2 & 1 & -1 \end{pmatrix}\right\|_\infty \left\|\begin{pmatrix} -1/10 & -1/20 & 1/2 \\ 1/30 & 11/60 & 1/6 \\ -1/6 & 1/12 & 1/6 \end{pmatrix}\right\|_\infty = 10 \cdot 39/60 \approx 6.5$

ex. $\mathbf{A} = \begin{pmatrix} 1.01 & 0.99 \\ 0.99 & 1.01 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$

[8] •

- $\kappa_\infty(\mathbf{A}) = \|\mathbf{A}\|_\infty \|\mathbf{A}^{-1}\|_\infty = \left\| \begin{pmatrix} 1.01 & 0.99 \\ 0.99 & 1.01 \end{pmatrix} \right\|_\infty \left\| \begin{pmatrix} 25.25 & -24.75 \\ -24.75 & 25.25 \end{pmatrix} \right\|_\infty = 2 \cdot 50 = 100$

- The matrix $\mathbf{A}$ has a large condition number

  In fact, the solution $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, we take $\hat{\mathbf{b}} = \mathbf{b} + \Delta\mathbf{b} = \begin{pmatrix} 2.02 \\ 1.98 \end{pmatrix}$ then the solution $\hat{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$. A small change in $\mathbf{b}$ has caused a large change in the solution vector.

2.5 Application of Matrix decomposition

Let $\mathbf{A}_{m\times n}$ be a matrix.

A linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$[9] has no solutions, one solution, or infinitely many solutions.

- If $\mathbf{A}$ is invertible, then the matrix equation has a unique solution given by $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$

- If $\mathbf{A}$ is invertible, then the only solution to the homogeneous equation $\mathbf{A}\mathbf{x} = \mathbf{0}$ is the trivial solution $\mathbf{A}\mathbf{x} = \mathbf{b}$

- If $\mathbf{u}, \mathbf{v}$ are solutions to $\mathbf{A}\mathbf{x} = \mathbf{0}$, then the $\mathbf{u} + c\mathbf{v}$ is another solution for every scalar c.

Theorem 2.6 *Elementary row operations*
1. Interchange two rows
2. Multiply a row by a nonzero constant
3. Add a multiple of a row to another row

LU Decomposition

- $\mathbf{A}_{n\times n} = \mathbf{L}\mathbf{U}$[10] $= \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}$

  The most efficient algorithm for solving linear systems $\mathbf{A}\mathbf{x} = \mathbf{b}$

  Example. $\mathbf{A} = \begin{pmatrix} 1 & 3 & -6 \\ -2 & 4 & 2 \\ 2 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & -6 \\ 0 & 10 & -10 \\ 0 & 0 & 6 \end{pmatrix}$

  step 1. $R_2 + 2R_1 \to R_2, \quad \mathbf{E_1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

  step 2. $R_3 - 2R_1 \to R_2, \quad \mathbf{E_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$

[9] A system of the form $\mathbf{A}\mathbf{x} = \mathbf{0}$
[10] $\mathbf{L}$=: lower triangular matrix, $\mathbf{U}$=: upper triangular matrix

step 3. $R_3 + 1/2R_2 \to R_2$, $\quad \mathbf{E_3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/2 & 1 \end{pmatrix}$

then

$$
\begin{aligned}
(E_3\mathbf{E_2}\mathbf{E_1})\mathbf{A} = \mathbf{U} \quad &\Leftrightarrow \quad \mathbf{A} = (\mathbf{E_3}\mathbf{E_2}\mathbf{E_1})^{-1}\mathbf{U} \\
&\Leftrightarrow \quad \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 1/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & -6 \\ 0 & 10 & -10 \\ 0 & 0 & 6 \end{pmatrix} \\
&\Leftrightarrow \quad \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & -6 \\ 0 & 10 & -10 \\ 0 & 0 & 6 \end{pmatrix} \\
&\Leftrightarrow \quad \mathbf{A} = \mathbf{LU}
\end{aligned}
$$

- i.e.,

$$
\begin{array}{rrrl} x_1 & +3x_2 & -6x_3 & = 5 \\ -2x_1 & +4x_2 & +2x_3 & = 2 \\ 2x_1 & +x_2 & -x_3 & = 10 \end{array} \quad \Leftrightarrow \quad \mathbf{Ax} = \mathbf{b} \quad \Leftrightarrow \quad \begin{pmatrix} 1 & 3 & -6 \\ -2 & 4 & 2 \\ 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 10 \end{pmatrix}
$$

$$
\Leftrightarrow \quad \mathbf{LUx} = \mathbf{b} \quad \Leftrightarrow \quad \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & -6 \\ 0 & 10 & -10 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 10 \end{pmatrix}
$$

- let $\mathbf{Ux} = \mathbf{y}$ and then solve $\mathbf{Ly} = \mathbf{b}$

$$
\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 10 \end{pmatrix} \quad \Leftrightarrow \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 12 \\ 6 \end{pmatrix}
$$

- solve $\mathbf{Ux} = \mathbf{y}$

$$
\begin{pmatrix} 1 & 3 & -6 \\ 0 & 10 & -10 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 12 \\ 6 \end{pmatrix} \quad \Leftrightarrow \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 22/5 \\ 11/5 \\ 1 \end{pmatrix}
$$

Quiz. $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} m \\ b \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$ **LU** 분해를 이용하여 $\mathbf{x}$의 추정치를 구하여라..

- Solution.

$$
\begin{aligned}
(E_3\mathbf{E_2}\mathbf{E_1})\mathbf{A} = \mathbf{U} \quad &\Leftrightarrow \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & -1 \\ 0 & 0 \end{pmatrix} \\
&\Leftrightarrow \quad \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \\ 0 & 0 \end{pmatrix} \\
&\Leftrightarrow \quad \mathbf{A} = \mathbf{LU}
\end{aligned}
$$

-

$$\mathbf{LU}\hat{\mathbf{x}} = \mathbf{b} \quad \Rightarrow \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

- let $\mathbf{Ux} = \mathbf{y}$ and then solve $\mathbf{Ly} = \mathbf{b}$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \quad \Leftrightarrow \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ -6 \end{pmatrix}$$

- solve $\mathbf{Ux} = \mathbf{y}$

$$\begin{pmatrix} 1 & 1 \\ 0 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ -6 \end{pmatrix} \quad \Leftrightarrow \quad \mathbf{x} = \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$$

QR Decomposition

- $\mathbf{A}_{n\times n} = \mathbf{QR}$[11]

  Example. $\mathbf{A} = \begin{pmatrix} 1 & 3 & -6 \\ -2 & 4 & 2 \\ 2 & 1 & -1 \end{pmatrix} \approx \begin{pmatrix} 0.33 & 0.67 & -0.67 \\ -0.67 & 0.67 & 0.33 \\ 0.67 & 0.33 & 0.67 \end{pmatrix} \begin{pmatrix} 3 & -1 & -4 \\ 0 & 5 & -3 \\ 0 & 0 & 4 \end{pmatrix}$

  Using Gram-Schmidt Orthogonalization

  step 1. $\mathbf{A} \overset{\text{GSO}}{\longrightarrow} \mathbf{U}$

  step 2. $\mathbf{U} \overset{\text{normalize}}{\longrightarrow} \mathbf{Q}$

  step 3. $\mathbf{Q}^T\mathbf{A} = \mathbf{Q}^T(\mathbf{QR}) = \mathbf{R}$

- $\mathbf{Ax} = \mathbf{b}$의 least square problem (find a vector $\mathbf{x}$ that minimize $\|\mathbf{Ax} - \mathbf{b}\|$)

- Using "QR decomposition"

- $\begin{pmatrix} 1 & 3 & -6 \\ -2 & 4 & 2 \\ 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 10 \end{pmatrix} \implies \hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \overset{QR}{\implies} \hat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q}^T\mathbf{b}$

-

$$\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^T\mathbf{b} \quad \Rightarrow \quad \begin{pmatrix} 3 & -1 & -4 \\ 0 & 5 & -3 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0.33 & -0.67 & 0.67 \\ 0.67 & 0.67 & 0.33 \\ -0.67 & 0.33 & 0.67 \end{pmatrix} \begin{pmatrix} 5 \\ 2 \\ 10 \end{pmatrix} = \begin{pmatrix} 7.0100 \\ 7.9900 \\ 4.0100 \end{pmatrix}$$

$$\Rightarrow \quad \begin{pmatrix} 3x_1 - x_2 - 4x_3 \\ 5x_2 - 3x_3 \\ 4x_3 \end{pmatrix} \approx \begin{pmatrix} 7 \\ 8 \\ 4 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \approx \begin{pmatrix} 22/5 \\ 11/5 \\ 1 \end{pmatrix}$$

Quiz. $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} m \\ b \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$ **QR**분해를 이용하여 $\mathbf{x}$의 추정치를 구하시오.

[11] $\mathbf{Q}$: orthogonal matrix($\mathbf{Q}^{-1} = \mathbf{Q}^T$), $\mathbf{R}$: upper triangular matrix

- Solution.

$$\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^T\mathbf{b} \quad \Rightarrow \quad \begin{pmatrix} -3.7417 & -1.6036 \\ 0 & 0.6547 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} -0.2673 & -0.5345 & -0.8018 \\ 0.8729 & 0.2182 & -0.4364 \\ 0.0482 & -0.8165 & 0.4082 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} -3.7417\ m - 1.6036\ b \\ 0.6547\ b \\ 0 \end{pmatrix} = \begin{pmatrix} -3.4745 \\ 0.6548 \\ 1.2245 \end{pmatrix} \Rightarrow \begin{pmatrix} m \\ b \end{pmatrix} \approx \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}$$

- 실제 QR기법에서는 GSO에 의한 분해 보다는 닮음변환 (Hessenberg행렬) 과 하우스홀더 법 (Householder 거울변환)을 이용하여 실행한다.
- eigen value 추정

Eigendecomposition (Spectral decomposition)

- $\mathbf{A}_{n\times n} = \mathbf{PDP}^{-1} = \begin{pmatrix} \mathbf{v}_1(\lambda_1) & \mathbf{v}_2(\lambda_2) & \mathbf{v}_3(\lambda_3) \end{pmatrix} \begin{pmatrix} {\sigma_1}^{12} & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1(\lambda_1) & \mathbf{v}_2(\lambda_2) & \mathbf{v}_3(\lambda_3) \end{pmatrix}^{-1}$

  λ_i=eigen value, $\sigma_i = \sqrt{\lambda_i}$

  Definition 2.7

  1. *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $n \times n$ *matrices, then* $\mathbf{A}$ *is* ***similar*** *to* $\mathbf{B}$ *if* $\exists$ *an invertible matrix* $\mathbf{P}$ *such that* $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}$
  2. *If* $\mathbf{B}$ *is a diagonal matrix, then the matrix* $\mathbf{A}$ *is* ***diagonalizable*** *if either* $\mathbf{D} = \mathbf{P}^{-1}\mathbf{AP}$ [13] *or* $\mathbf{A} = \mathbf{PDP}^{-1}$
  3. *A number* λ *is called an* ***eigenvalue*** *of* $\mathbf{A}$ *if* $\exists$ *a nonzero vector* $\boldsymbol{v}$ *in* $\mathbb{R}^n$ *such that* $\mathbf{A}\boldsymbol{v} = \lambda\boldsymbol{v}$
     The ***spectrum*** *of* $\mathbf{A}$ *is a set of eigenvalues of* $\mathbf{A}$.
  4. *Every nonzero vector satisfying above equation is called an* ***eigenvector*** *of* $\mathbf{A}$ *corresponding to the eigenvalue* λ

  Example. Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}$

$$\begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = -1\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\lambda_1 = 1 \to \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \lambda_2 = -1 \to \mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

  Theorem 2.8 *An* $n \times n$ *matrix* $\mathbf{A}$ *is diagonalizable iff* $\mathbf{A}$ *has* n *linearly independent eigenvectors. Moreover, if* $\mathbf{D} = \mathbf{P}^{-1}\mathbf{AP}$ *with* $\mathbf{D}$ *a diagonal matrix, then the diagonal entries of* $\mathbf{D}$ *are the eigenvalues of* $\mathbf{A}$ *and the column vectors of* $\mathbf{P}$ *are the corresponding eigenvectors.*

  Example. For $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}$,

$$\mathbf{A} = \mathbf{PDP}^{-1} \Leftrightarrow \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$$

  Note. If $\mathbf{A}$ is symmetric then $\mathbf{A} = \mathbf{Q}$[14]$\mathbf{DQ}^T$

[13] $\mathbf{P}$: matrix formed from eigenvector of A, $\mathbf{D}$: diagonal matrix with eigenvalue
[14] orthogonal marix

Singular Value Decomposition(SVD)

- $\mathbf{A}_{m\times n} = \mathbf{U}_{m\times m}\sum_{m\times n}\mathbf{V}^T_{n\times n}$
  - $\mathbf{A}^T_{m\times n}\mathbf{A}_{m\times n} = \mathbf{V}_{n\times n}\sum^T_{m\times n}\mathbf{U}^T_{m\times m}\mathbf{U}_{m\times m}\sum_{m\times n}\mathbf{V}^T_{n\times n} = \mathbf{V}_{n\times n}\mathbf{D}_1\mathbf{V}^T_{n\times n}$
  - The **singular value** σ_i of $\mathbf{A}$ for $1 \le i \le n$, are the positive square roots of the eigenvalues $\lambda_1, \dots, \lambda_n$ of $\mathbf{A}\mathbf{A}^T$
  - $\mathbf{A}_{m\times n}\mathbf{A}^T_{m\times n} = \mathbf{U}_{m\times m}\sum_{m\times n}\mathbf{V}^T_{n\times n}\mathbf{V}_{n\times n}\sum^T_{m\times n}\mathbf{U}^T_{m\times m} = \mathbf{U}_{m\times m}\mathbf{D}_2\mathbf{U}^T_{m\times m}$

ex. $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \approx \begin{pmatrix} \frac{-10+\sqrt{265}}{15} & 1 & -\frac{10+\sqrt{265}}{15} \\ \frac{5+\sqrt{265}}{30} & -2 & -\frac{-5+\sqrt{265}}{30} \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{-11-\sqrt{265}}{12} & 1 \\ \frac{11-\sqrt{265}}{12} & 1 \end{pmatrix}$

step 1. $\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 14 & 6 \\ 6 & 3 \end{pmatrix}$, $\sigma_1 = \sqrt{\lambda_1}$ $\sigma_2 = \sqrt{\lambda_2}$ where λ_1, λ_2 such that $\det(\mathbf{A}^T\mathbf{A} - \lambda I) = 0$

$\lambda_1 = \frac{17+\sqrt{265}}{2}$, $\lambda_2 = \frac{17-\sqrt{265}}{2}$ such that $\lambda^2 - 17\lambda + 6 = 0$, $\sigma_1 \ge \sigma_2$

step 2. Find $\mathbf{v}_1(\lambda_1), \mathbf{v}_2(\lambda_2)$ such that $(\mathbf{A}^T\mathbf{A} - \lambda I)(\mathbf{v}) = \mathbf{0}$

$$\mathbf{v}_1 = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad \text{s.t} \quad \left(\begin{pmatrix} 14 & 6 \\ 6 & 3 \end{pmatrix} - \frac{17+\sqrt{265}}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \frac{11-\sqrt{265}}{12} & 6 \\ 6 & \frac{-11-\sqrt{265}}{12} \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow v_1 = \frac{-11-\sqrt{265}}{12}v_2$$

$$\Rightarrow \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \frac{-11-\sqrt{265}}{12} \\ 1 \end{pmatrix} y, \text{ for any } y \ne 0$$

$$\mathbf{v}_2 = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad \text{s.t} \quad \left(\begin{pmatrix} 14 & 6 \\ 6 & 3 \end{pmatrix} - \frac{17-\sqrt{265}}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \frac{11+\sqrt{265}}{12} & 6 \\ 6 & \frac{-11+\sqrt{265}}{12} \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow v_1 = \frac{11-\sqrt{265}}{12}v_2$$

$$\Rightarrow \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \frac{11-\sqrt{265}}{12} \\ 1 \end{pmatrix} y, \text{ for any } y \ne 0$$

step 3. $\mathbf{A}\mathbf{A}^T = \begin{pmatrix} 2 & 3 & 4 \\ 3 & 5 & 7 \\ 4 & 7 & 10 \end{pmatrix}$, $\lambda_1, \lambda_2, \lambda_3$ such that $\det(\mathbf{A}\mathbf{A}^T - \lambda I) = 0$

step 4. Find $\mathbf{u}_1(\lambda_1), \mathbf{u}_2(\lambda_2), \mathbf{u}_3(\lambda_3)$ such that $(\mathbf{A}^T\mathbf{A} - \lambda I)(\mathbf{u}) = \mathbf{0}$

$\lambda_1 = \frac{17+\sqrt{265}}{2}$, $\lambda_2 = 0$, $\lambda_3 = \frac{17-\sqrt{265}}{2}$ such that $-\lambda^3 + 17\lambda^2 - 6\lambda = 0$

$$\mathbf{u}_1 = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} \frac{-10+\sqrt{265}}{15} \\ \frac{5+\sqrt{265}}{30} \\ 1 \end{pmatrix} y, \text{ for any } y \ne 0$$

$$\mathbf{u}_2 = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} y, \text{ for } \text{ any } y \neq 0$$

$$\mathbf{u}_3 = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} -\frac{10+\sqrt{265}}{15} \\ -\frac{-5+\sqrt{265}}{30} \\ 1 \end{pmatrix} y, \text{ for } \text{ any } y \neq 0$$

- "inverse" for $\mathbf{A}_{m\times n}$: **pseudo-inverse (Moore-Penrose inverse)**

$$\begin{aligned} \mathbf{A}_{m\times n} &= \mathbf{U}_{m\times m}\sum\nolimits_{m\times n}\mathbf{V}^T_{n\times n} = \mathbf{U}_{m\times m}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\mathbf{V}^T_{n\times n} \\ \mathbf{A}^{-1}_{m\times n} &= \mathbf{V}_{n\times n}\sum\nolimits^{-1}_{m\times n}\mathbf{U}^T_{m\times m} = \mathbf{V}_{n\times n}\begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix}^T\mathbf{U}^T_{m\times m} \end{aligned}$$

Quiz. $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} m \\ b \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$ **SVD**를 이용하여 $\mathbf{x}$의 추정치를 구하시오.

- solution. $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} \frac{-10+\sqrt{265}}{15} & 1 & -\frac{10+\sqrt{265}}{15} \\ \frac{5+\sqrt{265}}{30} & -2 & -\frac{-5+\sqrt{265}}{30} \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \frac{-11-\sqrt{265}}{12} & 1 \\ \frac{11-\sqrt{265}}{12} & 1 \end{pmatrix}$

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{A}^{\dagger}\mathbf{b}^{15} \\ \Rightarrow & \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} \frac{-11-\sqrt{265}}{12} & \frac{11-\sqrt{265}}{12} \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 \end{pmatrix}\begin{pmatrix} \frac{-10+\sqrt{265}}{15} & \frac{5+\sqrt{265}}{30} & 1 \\ 1 & -2 & 1 \\ -\frac{10+\sqrt{265}}{15} & -\frac{-5+\sqrt{265}}{30} & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \\ \Rightarrow & \begin{pmatrix} m \\ b \end{pmatrix} \approx \begin{pmatrix} 1/2 \\ 1 \end{pmatrix} \end{aligned}$$

Principal component analysis (PCA)

- feature extraction method, dimensionality reduction(cf. CNN pooling)
- 주성분 분석은 고차원 데이터를 저차원 데이터로로 변환해서 보는 기법이다.

  추출한 데이터에서 데이터의 구조(길이나 각도)는 유지한채 적은 수의 특징만으로 특정 현상을 설명하고자 하는 것이 목적이다

- 입력값보다 적은 출력값을 내는 함수의 정의 $f : \mathbb{R}^m \to \mathbb{R}^n$ $(n \leq m)$
- 분산이 큰 방향의 벡터(주성분)에 데이터 정사영한다.

  Example. 3-variate dataset with 10 observations.

step 1.

$$\mathbf{X} = \begin{pmatrix} 7 & 4 & 6 & 8 & 8 & 7 & 5 & 9 & 7 & 8 \\ 4 & 1 & 3 & 6 & 5 & 2 & 3 & 5 & 4 & 2 \\ 3 & 8 & 5 & 1 & 7 & 9 & 3 & 8 & 5 & 2 \end{pmatrix}$$

compute the correlation matrix $\sum = \begin{pmatrix} 1 & 0.67 & -0.10 \\ 0.67 & 1 & -0.29 \\ -0.10 & -0.29 & 1.00 \end{pmatrix}$

step 2. Matrix $\sum$ decomposition $\Rightarrow$ $\lambda_1 = 1.7969,\ \lambda_2 = 0.927, \lambda_3 = 0.304$

$$\begin{aligned}
\lambda_1 = 1.7969 \quad &\Rightarrow \quad \mathbf{v}_1 = \begin{pmatrix} 0.64 & 0.69 & -0.34 \end{pmatrix}^T \\
\lambda_2 = 0.927 \quad &\Rightarrow \quad \mathbf{v}_2 = \begin{pmatrix} 0.38 & 0.10 & 0.91 \end{pmatrix}^T \\
\lambda_3 = 0.304 \quad &\Rightarrow \quad \mathbf{v}_3 = \begin{pmatrix} -0.66 & 0.72 & 0.20 \end{pmatrix}^T
\end{aligned}$$

$$\begin{aligned}
\sum \mathbf{P} &= \begin{pmatrix} \mathbf{v}_1(\lambda_1) & \mathbf{v}_2(\lambda_2) & \mathbf{v}_3(\lambda_3) \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} = \mathbf{PD} \\
&= \begin{pmatrix} 0.64 & 0.38 & -0.66 \\ 0.69 & 0.10 & 0.72 \\ -0.34 & 0.91 & 0.20 \end{pmatrix} \begin{pmatrix} 1.33 & 0 & 0 \\ 0 & 0.96 & 0 \\ 0 & 0 & 0.55 \end{pmatrix} = \begin{pmatrix} 0.85 & 0.37 & -0.37 \\ 0.91 & 0.10 & 0.40 \\ -0.45 & 0.88 & 0.11 \end{pmatrix} \\
\text{Let } \mathbf{y}_1 &= \mathbf{v}_1^T \mathbf{X} \\
&= \begin{pmatrix} 0.64 & 0.69 & -0.34 \end{pmatrix} \begin{pmatrix} 7 & 4 & 6 & 8 & 8 & 7 & 5 & 9 & 7 & 8 \\ 4 & 1 & 3 & 6 & 5 & 2 & 3 & 5 & 4 & 2 \\ 3 & 8 & 5 & 1 & 7 & 9 & 3 & 8 & 5 & 2 \end{pmatrix} \\
&= 0.64 \begin{pmatrix} 7 & 4 & 6 & 8 & 8 & 7 & 5 & 9 & 7 & 8 \end{pmatrix} + 0.69 \begin{pmatrix} 4 & 1 & 3 & 6 & 5 & 2 & 3 & 5 & 4 & 2 \end{pmatrix} \\
&\quad -0.34 \begin{pmatrix} 3 & 8 & 5 & 1 & 7 & 9 & 3 & 8 & 5 & 2 \end{pmatrix} \\
&= \begin{pmatrix} 6.22 & 0.53 & 4.21 & 8.92 & 6.19 & 2.80 & 4.25 & 6.49 & 5.54 & 5.82 \end{pmatrix} \\
&\quad \mathbf{y}_2 = \mathbf{v}_2^T \mathbf{X}, \quad \mathbf{y}_3 = \mathbf{v}_3^T \mathbf{X} \\
\text{then, } \mathbf{Y} &= \mathbf{PX} \\
&= \begin{pmatrix} 0.64 & 0.69 & -0.34 \\ 0.38 & 0.10 & 0.91 \\ -0.66 & 0.72 & 0.20 \end{pmatrix} \begin{pmatrix} 7 & 4 & 6 & 8 & 8 & 7 & 5 & 9 & 7 & 8 \\ 4 & 1 & 3 & 6 & 5 & 2 & 3 & 5 & 4 & 2 \\ 3 & 8 & 5 & 1 & 7 & 9 & 3 & 8 & 5 & 2 \end{pmatrix} \\
&= \begin{pmatrix} 6.22 & 0.53 & 4.21 & 8.92 & 6.19 & 2.80 & 4.25 & 6.49 & 5.54 & 5.82 \\ 5.79 & 8.90 & 7.13 & 4.55 & 9.91 & 11.05 & 4.93 & 11.20 & 7.61 & 5.06 \\ -1.14 & -0.32 & -0.80 & -0.76 & -0.28 & -1.38 & -0.54 & -0.74 & -0.74 & -3.44 \end{pmatrix}
\end{aligned}$$

- Quiz. PCA 구하기 3 variable dataset

$$\mathbf{X} \quad = \quad \begin{pmatrix} 62 & 8 & 1 & 2 \\ 1 & 4 & 4 & 8 \\ 3 & 7 & 4 & 2 \end{pmatrix}$$

Solution. step1. $Cov(X)$ 행렬 구하기.

$Cov(X,Y) = (\sum(x-\mu)(y-\nu))/(n-1)$

$$\begin{aligned}\mathbf{Cov}(\mathbf{X}) &= \begin{pmatrix} Cov(x_1,x_1) & Cov(x_1,x_2) & Cov(x_1,x_3) \\ Cov(x_2,x_1) & Cov(x_2,x_2) & Cov(x_2,x_3) \\ Cov(x_3,x_1) & Cov(x_3,x_2) & Cov(x_3,x_3) \end{pmatrix} \\ &= \begin{pmatrix} 10.25 & -0.41666667 & 6 \\ -0.41666667 & 8.25 & -1.66666667 \\ 6 & -1.66666667 & 4.66666667 \end{pmatrix}\end{aligned}$$

step 2. eigenvectors 구하기.

$$\begin{pmatrix} \lambda_1 = 0.6541291 & \lambda_2 = 14.33608183 & \lambda_3 = 8.17645574 \end{pmatrix}$$
$$\Rightarrow \begin{pmatrix} 0.51879659 & 0.81567701 & 0.255970936 \\ -0.15597286 & -0.20408152 & 0.96644876 \\ -0.84054897 & 0.54131484 & -0.02134668 \end{pmatrix}$$

$\lambda_2 = 14.33608183 \geq \lambda_3 = 8.17645574 \geq \lambda_1 = 0.6541291$ 이고 $\frac{\lambda_2+\lambda_3}{\lambda_1+\lambda_2+\lambda_3} \approx 0.97$ 이므로 97% 변화량 반영하면서 3차원을 2차원으로 축소가능

$$\mathbf{X}' = \begin{pmatrix} 1 & 4 & 4 & 8 \\ 3 & 7 & 4 & 2 \end{pmatrix}$$

References

[1] http://cs229.stanford.edu/section/cs229-linalg.pdf

[2] Advanced Engineering Mathematics ch7-8, Erwin Kreyszig, Wiely

[3] Elementary Linear Algebra, 7th ed., Ron Larson

[4] Deep Learning ch2, Ian Goodfellow and Yoshua Bengio and Aaron Courville, The MIT Press

[5] https://en.wikipedia.org