CS 229, Fall 2018
Problem Set #0: Linear Algebra and Multivariable Calculus

Notes: (1) These questions require thought, but do not require long answers. Please be as concise as possible. (2) If you have a question about this homework, we encourage you to post your question on our Piazza forum, at `https://piazza.com/stanford/fall2018/cs229` (3) If you missed the first lecture or are unfamiliar with the collaboration or honor code policy, please read the policy on Handout #1 (available from the course website) before starting work. (4) This specific homework is *not graded*, but we encourage you to solve each of the problems to brush up on your linear algebra. Some of them may even be useful for subsequent problem sets. It also serves as your introduction to using Gradescope for submissions.

1. **[0 points] Gradients and Hessians**

   Recall that a matrix $A \in \mathbb{R}^{n\times n}$ is *symmetric* if $A^T = A$, that is, $A_{ij} = A_{ji}$ for all i, j. Also recall the gradient $\nabla f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$, which is the n-vector of partial derivatives

f: [ ]

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \quad \text{where} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

   The hessian $\nabla^2 f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the $n \times n$ symmetric matrix of twice partial derivatives,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n^2} f(x) \end{bmatrix}.$$

   (a) Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. What is $\nabla f(x)$?

   (b) Let $f(x) = g(h(x))$, where $g : \mathbb{R} \to \mathbb{R}$ is differentiable and $h : \mathbb{R}^n \to \mathbb{R}$ is differentiable. What is $\nabla f(x)$?

   (c) Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is symmetric and $b \in \mathbb{R}^n$ is a vector. What is $\nabla^2 f(x)$?

   (d) Let $f(x) = g(a^T x)$, where $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and $a \in \mathbb{R}^n$ is a vector. What are $\nabla f(x)$ and $\nabla^2 f(x)$? (*Hint:* your expression for $\nabla^2 f(x)$ may have as few as 11 symbols, including $'$ and parentheses.)

2. **[0 points] Positive definite matrices**

   A matrix $A \in \mathbb{R}^{n\times n}$ is *positive semi-definite* (PSD), denoted $A \succeq 0$, if $A = A^T$ and $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$. A matrix A is *positive definite*, denoted $A \succ 0$, if $A = A^T$ and $x^T Ax > 0$ for all $x \neq 0$, that is, all non-zero vectors x. The simplest example of a positive definite matrix is the identity I (the diagonal matrix with 1s on the diagonal and 0s elsewhere), which satisfies $x^T Ix = \|x\|_2^2 = \sum_{i=1}^n x_i^2$.

(a) Let $z \in \mathbb{R}^n$ be an n-vector. Show that $A = zz^T$ is positive semidefinite.

(b) Let $z \in \mathbb{R}^n$ be a *non-zero* n-vector. Let $A = zz^T$. What is the null-space of A? What is the rank of A?

(c) Let $A \in \mathbb{R}^{n \times n}$ be positive semidefinite and $B \in \mathbb{R}^{m \times n}$ be arbitrary, where $m, n \in \mathbb{N}$. Is BAB^T PSD? If so, prove it. If not, give a counterexample with explicit A, B.

3. **[0 points] Eigenvectors, eigenvalues, and the spectral theorem**

The eigenvalues of an $n \times n$ matrix $A \in \mathbb{R}^{n \times n}$ are the roots of the characteristic polynomial $p_A(\lambda) = \det(\lambda I - A)$, which may (in general) be complex. They are also defined as the the values $\lambda \in \mathbb{C}$ for which there exists a vector $x \in \mathbb{C}^n$ such that $Ax = \lambda x$. We call such a pair (x, λ) an *eigenvector, eigenvalue* pair. In this question, we use the notation $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ to denote the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$, that is,

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

(a) Suppose that the matrix $A \in \mathbb{R}^{n \times n}$ is diagonalizable, that is, $A = T\Lambda T^{-1}$ for an invertible matrix $T \in \mathbb{R}^{n \times n}$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. Use the notation $t^{(i)}$ for the columns of T, so that $T = [t^{(1)} \ \cdots \ t^{(n)}]$, where $t^{(i)} \in \mathbb{R}^n$. Show that $At^{(i)} = \lambda_i t^{(i)}$, so that the eigenvalues/eigenvector pairs of A are $(t^{(i)}, \lambda_i)$.

A matrix $U \in \mathbb{R}^{n \times n}$ is orthogonal if $U^T U = I$. The spectral theorem, perhaps one of the most important theorems in linear algebra, states that if $A \in \mathbb{R}^{n \times n}$ is symetric, that is, $A = A^T$, then A is *diagonalizable by a real orthogonal matrix.* That is, there are a diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ and orthogonal matrix $U \in \mathbb{R}^{n \times n}$ such that $U^T A U = \Lambda$, or, equivalently,

$$A = U\Lambda U^T.$$

Let $\lambda_i = \lambda_i(A)$ denote the ith eigenvalue of A.

(b) Let A be symmetric. Show that if $U = [u^{(1)} \ \cdots \ u^{(n)}]$ is orthogonal, where $u^{(i)} \in \mathbb{R}^n$ and $A = U\Lambda U^T$, then $u^{(i)}$ is an eigenvector of A and $Au^{(i)} = \lambda_i u^{(i)}$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

(c) Show that if A is PSD, then $\lambda_i(A) \geq 0$ for each i.

$f(x) = \frac{1}{2} \underbrace{x^T \cdot A \cdot x}_{\text{Scalar}} + \underbrace{b^T x}_{\text{Scalar}}$ Where $A^T = A$ $b \in \mathbb{R}^n$, $x \in \mathbb{R}^n$ $\nabla f(x)$?

$b^T x = b_1 x_1 + \cdots + b_n x_n$

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\nabla f(x) = Ax + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\begin{bmatrix} a_{11}x_1 + a_{21}x_2, & a_{12}x_1 + a_{22}x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$a_{11}x_1^2 + a_{21}x_1x_2 + a_{12}x_1x_2 + a_{22}x_2^2$$

$$\boxed{\nabla f(x) = Ax + b}$$

$$\begin{bmatrix} a_{11}x_1 + a_{21}x_2 + a_{31}x_3 + \cdots + a_{n1}x_n, & a_{12}x_1 + a_{22}x_2 + a_{32}x_3 + \cdots + a_{n2}x_n, & \cdots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$a_{11}x_1^2 + a_{21}x_1x_2 + a_{31}x_1x_3 + \cdots + a_{n1}x_1x_n$

$+ \quad a_{12}x_1x_2 + a_{22}x_2^2 +$

$a_{13}x_3x_1$

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots \\ \end{bmatrix}$$

How? 어케 했음?

$$\frac{1}{2} x^T \cdot A \cdot x = \frac{1}{2} \sum_i \sum_j A_{ij} x_i x_j$$

$$\frac{\partial}{\partial x_k} \frac{1}{2} x^T \cdot A \cdot x = \frac{\partial}{\partial x_k} \frac{1}{2} \sum_{i=1, i \neq k}^{n} A_{ik} \cdot x_i x_k + \frac{\partial}{\partial x_k} \frac{1}{2} \sum_{j=1, j \neq k}^{n} A_{kj} x_k x_j$$

※ 1-(1) 해설.

$$\frac{1}{2} x^T \cdot A \cdot x = \frac{1}{2} \sum_{i}^{n} \sum_{j}^{n} A_{ij} x_i x_j$$

$$\frac{\partial}{\partial x_k} \frac{1}{2} x^T A x = \frac{\partial}{\partial x_k} \frac{1}{2} \cdot \sum_{i=1, i \neq k}^{n} A_{ik} x_i x_k + \frac{\partial}{\partial x_k} \frac{1}{2} \sum_{j=1, j \neq k}^{n} A_{kj} x_k x_j$$

$$+ \frac{\partial}{\partial x_k} \frac{1}{2} A_{kk} x_k x_k$$

$$= \frac{1}{2} \sum_{i=1, i \neq k}^{n} A_{ik} \cdot x_i + \frac{1}{2} \sum_{j=1, j \neq k}^{n} A_{kj} x_j + A_{kk} x_k.$$

symmetric

$$= \sum_{i=1}^{n} A_{ki} x_i$$

1×1

(b) Let $f(x) = g(h(x))$, where $g : \mathbb{R} \to \mathbb{R}$ is differentiable and $h : \mathbb{R}^n \to \mathbb{R}$ is differentiable. What is $\nabla f(x)$?

$$h : \overbrace{\begin{bmatrix} h_1 & \cdots & h_n \end{bmatrix}}^{n} \quad \nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\frac{\partial}{\partial x_i} f(x) = \frac{\partial}{\partial x_i} g(h(x))$$

$$= \frac{\partial}{\partial x_i} g(h_1 x_1 + h_2 x_2 + \cdots + h_n x_n)$$

$$= g'(h_1 x_1 + \cdots + h_n x_n) \cdot h_i$$

$$= g'(h(x)) \cdot h_i$$

$$\nabla f(x) = \begin{array}{c} \\ n \end{array} \overbrace{\begin{bmatrix} g'(h(x)) \cdot h_1 \\ g'(h(x)) \cdot h_2 \\ \vdots \\ g'(h(x)) \cdot h_n \end{bmatrix}}^{1} = n \overbrace{\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}}^{1} \times \overbrace{\begin{bmatrix} g'(h(x)) \end{bmatrix}}^{1}$$

$=$

1-(b) Sol.

$$\nabla f(x) = g'(h(x)) \cdot \nabla h(x).$$

$$\frac{\partial}{\partial x_i} f(x) = \frac{\partial}{\partial x_i} g(h(x)) = g'(h(x)) \cdot \frac{\partial}{\partial x_i} h(x)$$

$$\nabla f(x) = \begin{bmatrix} g'(h(x)) \cdot \frac{\partial}{\partial x_1} h(x) \\ g'(h(x)) \cdot \frac{\partial}{\partial x_2} h(x) \\ \vdots \\ g'(h(x)) \cdot \frac{\partial}{\partial x_n} h(x) \end{bmatrix} = g'(h(x)) \cdot \nabla h(x)$$

(c) Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is symmetric and $b \in \mathbb{R}^n$ is a vector. What is $\nabla^2 f(x)$?

$A: \mathbb{R}^{n \times n}$.

by (a). $\nabla f(x) = Ax + b$.

$b: n \times 1$

$x: n \times 1$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} = \begin{bmatrix} A^{(1)T} \cdot x + b^{(1)} \\ A^{(2)T} x + b^{(2)} \\ \vdots \\ A^{(n)T} \cdot x + b^{(n)} \end{bmatrix} = \begin{bmatrix} \Sigma A_{1j} x_j + b^{(1)} \\ \Sigma A_{2j} x_j + b^{(2)} \\ \vdots \\ \Sigma A_{nj} x_j + b^{(n)} \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial^2}{\partial x_1 x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 x_n} f(x) \\ \frac{\partial^2}{\partial x_2 x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2}{\partial x_n x_1} f(x) & \cdots & & \frac{\partial^2}{\partial x_n^2} f(x) \end{bmatrix}$$

$$= \begin{bmatrix} A_{11} & A_{12} & & \\ A_{21} & \ddots & & \\ & & \ddots & \\ & & & A_{nn} \end{bmatrix} = A.$$

(d) Let $f(x) = g(a^T x)$, where $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and $a \in \mathbb{R}^n$ is a vector. What are $\nabla f(x)$ and $\nabla^2 f(x)$? (*Hint:* your expression for $\nabla^2 f(x)$ may have as few as 11 symbols, including $'$ and parentheses.)

g'' 존재.

$a : \begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad x : \begin{bmatrix} \\ \\ \\ \end{bmatrix}$

$$\frac{\partial}{\partial x_i} f(x) = g'(a^T x) \cdot \frac{\partial}{\partial x_i}(a^T x) = g'(a^T \cdot x) \cdot a_i$$

$$\therefore \nabla f(x) = \begin{bmatrix} g'(a^T x) \cdot a_1 \\ g'(a^T \cdot x) \cdot a_2 \\ \vdots \\ g'(a^T x) \cdot a_n \end{bmatrix} = g'(a^T x) \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = g'(a^T \cdot x) \cdot \vec{a}.$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial^2}{\partial x_1 x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 x_n} f(x) \\ \frac{\partial^2}{\partial x_2 x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n x_1} f(x) & \cdots & \cdots & \frac{\partial^2}{\partial x_n x_n} f(x) \end{bmatrix} = \begin{bmatrix} a_1^2 \cdot g'(a^T x) & a_{12} g'(a^T x) \\ a_{21} g'(a^T x) & a_2^2 g'(a^T x) \\ & \end{bmatrix}$$

$$\boxed{\vec{a}\,\vec{a}^T \cdot g''(a^T x)}$$

$n \times 1 \quad 1 \times n$

2. **[0 points] Positive definite matrices** $\sum_i \sum_j A_{ij} x_i x_j$

A matrix $A \in \mathbb{R}^{n\times n}$ is *positive semi-definite* (PSD), denoted $A \succeq 0$, if $A = A^T$ and $\underbrace{x^TAx \geq 0}$ for all $x \in \mathbb{R}^n$. A matrix A is *positive definite*, denoted $A \succ 0$, if $A = A^T$ and $x^TAx > 0$ for all $x \neq 0$, that is, all non-zero vectors x. The simplest example of a positive definite matrix is the identity I (the diagonal matrix with 1s on the diagonal and 0s elsewhere), which satisfies $x^TIx = \|x\|_2^2 = \sum_{i=1}^n x_i^2$.

(a) Let $z \in \mathbb{R}^n$ be an n-vector. Show that $A = zz^T$ is positive semidefinite.

(b) Let $z \in \mathbb{R}^n$ be a *non-zero* n-vector. Let $A = zz^T$. What is the null-space of A? What is the rank of A?

(c) Let $A \in \mathbb{R}^{n\times n}$ be positive semidefinite and $B \in \mathbb{R}^{m\times n}$ be arbitrary, where $m, n \in \mathbb{N}$. Is BAB^T PSD? If so, prove it. If not, give a counterexample with explicit A, B.

$$z = \begin{bmatrix} \\ \\ \\ \end{bmatrix} \qquad \underbrace{zz^T = \begin{bmatrix} z_i^2 & z_1 z_2 & & z_1 z_n \\ z_2 z_1 & z_2 z_2 & & \\ \vdots & & \ddots & \\ z_n z_1 & & & z_n z_n \end{bmatrix}}_{\text{symmetric}}$$

for all $x \in \mathbb{R}^n$: $\quad x^TAx = \sum_i \sum_j A_{ij} x_i x_j \underset{\sim}{\geq} 0$?

$$= x^T z z^T x$$

$$\sum\sum z_i z_j x_i x_j$$

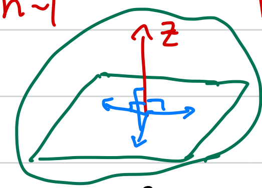
$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \cancel{\emptyset} \quad (x^Tz)^2 \geq 0$$

$$\underbrace{(x^T \cdot z)^2 \geq 0}_{\cancel{\emptyset}}$$

(2).

$z \in \mathbb{R}^n$ $\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \neq 0$. $A = zz^T$.

$\underbrace{\dim(\text{null } A)}_{n-1} + \underbrace{\dim(\text{Im } A)}_{1} = n$

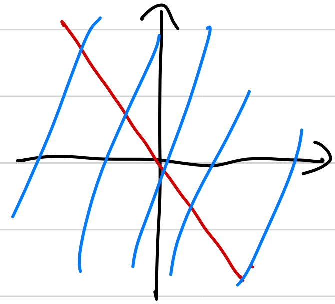


⇒ z와 이 평면만 있으면 $\mathbb{R}^n$ 구현 가능 ⇓ $(n-1)$

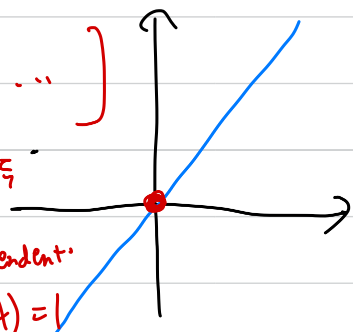
null space of A? Rank of A?

$\underline{Ax = 0}$ $\quad z_1 \cdot z_1 \cdot x_1 + z_1 z_2 \cdot x_2$.

$$zz^Tx = \begin{bmatrix} z_1^2 & z_1 z_2 & z_1 z_n \\ z_2 z_1 & z_2 z_2 & \\ \vdots & & \ddots \\ z_n z_1 & & z_n z_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} | & | & | & \\ z_1 z & z_2 z & z_3 z & \cdots \\ | & | & | & \end{bmatrix}$$



⇓ squish



$$= \begin{bmatrix} z_1 \cdot \sum_i^n z_i x_i \\ z_2 \cdot \sum_i^n z_i x_i \\ \vdots \\ z_n \cdot \sum_i^n z_i x_i \end{bmatrix} = 0.$$

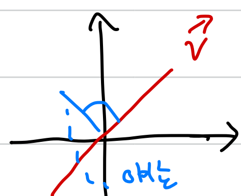
즉, z에 모두 상수배!!

⇒ Linearly dependent.

&& $\underline{\dim(\text{Im } A) = 1}$

0 되려면 $\underline{\sum_i^n z_i x_i = 0}$.

$z^T \cdot x = 0$ ⇐



애는 2차원 i, j로 표현되지만 이 직선만 보면 1차원.

∴ z에 수직인 애들이 nullspace!!

(3) $A \in \mathbb{R}^{n\times n}$ PSD. $B \in \mathbb{R}^{m\times n}$. $x \in \mathbb{R}^{m}$

BAB^T 只要: $\underbrace{x^T \cdot B}A \cdot \underbrace{B^T x} \geq 0$

Let $\underset{n\times m \;\; m\times 1}{v = B^T x} \in \mathbb{R}^{n}$

$$\boxed{v^T \cdot A \cdot v \geq 0}$$

$$\mathrm{diag}(\lambda_1, \ldots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

(a) Suppose that the matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable, that is, $A = T\Lambda T^{-1}$ for an invertible matrix $T \in \mathbb{R}^{n\times n}$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. Use the notation $t^{(i)}$ for the columns of T, so that $T = [t^{(1)} \; \cdots \; t^{(n)}]$, where $t^{(i)} \in \mathbb{R}^n$. Show that $At^{(i)} = \lambda_i t^{(i)}$, so that the eigenvalues/eigenvector pairs of A are $(t^{(i)}, \lambda_i)$.

$n \times n$

$A = \underset{\mathbb{R}^{n\times n}}{T} \underset{\mathrm{diag}(\lambda_1, \ldots, \lambda_n)}{\Lambda} T^{-1} \Rightarrow At^{(i)} = \lambda_i t^{(i)}$. 즉. eigen value.

$A\vec{v} = \lambda\vec{v}$

$(A - \lambda I)\cdot\vec{v} = 0$

$\det(A - \lambda I) = 0$

T is invertible.

$I_{n\times n} = T^{-1} \times T = T^{-1}[t^{(1)} \; t^{(2)} \; \cdots \; t^{(n)}]$

$= [T^{-1}t^{(1)} \; T^{-1}t^{(2)} \; \cdots \; T^{-1}t^{(n)}]$

$\Rightarrow T^{-1}t^{(1)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad T^{-1}t^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix} \quad \cdots \quad T^{-1}t^{(i)} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \} \; i-1 \\ \leftarrow i \text{ 번째} \\ \} \; n-i \end{array}$

$T^{-1}t^{(i)} = [0 \cdots 0 \; 1 \; 0 \cdots 0]^T$

standard basis vector

$\therefore \Lambda T^{-1}t^{(i)} = \Lambda \cdot \underbrace{e^{(i)}}_{\text{standard basis vector}} = \lambda_i \cdot e^{(i)}$

$T\Lambda T^{-1}t^{(i)} = T\cdot\lambda_i\cdot e^{(i)} = \lambda_i \cdot t^{(i)}$

(a) Suppose that the matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable, that is, $A = T\Lambda T^{-1}$ for an invertible matrix $T \in \mathbb{R}^{n\times n}$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. Use the notation $t^{(i)}$ for the columns of T, so that $T = [t^{(1)} \ \cdots \ t^{(n)}]$, where $t^{(i)} \in \mathbb{R}^n$. Show that $At^{(i)} = \lambda_i t^{(i)}$, so that the eigenvalues/eigenvector pairs of A are $(t^{(i)}, \lambda_i)$.

A matrix $U \in \mathbb{R}^{n\times n}$ is orthogonal if $U^T U = I$. The spectral theorem, perhaps one of the most important theorems in linear algebra, states that if $A \in \mathbb{R}^{n\times n}$ is symetric, that is, $A = A^T$, then A is *diagonalizable by a real orthogonal matrix*. That is, there are a diagonal matrix $\Lambda \in \mathbb{R}^{n\times n}$ and orthogonal matrix $U \in \mathbb{R}^{n\times n}$ such that $U^T AU = \Lambda$, or, equivalently,

$$A = U\Lambda U^T.$$

Let $\lambda_i = \lambda_i(A)$ denote the ith eigenvalue of A.

(b) Let A be symmetric. Show that if $U = [u^{(1)} \ \cdots \ u^{(n)}]$ is orthogonal, where $u^{(i)} \in \mathbb{R}^n$ and $A = U\Lambda U^T$, then $u^{(i)}$ is an eigenvector of A and $Au^{(i)} = \lambda_i u^{(i)}$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

(c) Show that if A is PSD, then $\lambda_i(A) \geq 0$ for each i.

① $A = A^T$. $A \in \mathbb{R}^{n\times n}$.

② $U^T A U = \Lambda$ $\qquad U^T \in \mathbb{R}^{n\times n}$ s.t. $U^T U = I$.

$U A U^T = \Lambda$ $\qquad U^T = U^{-1}$.

$U = [u^{(1)}.\ u^{(2)} \ \cdots \ u^{(n)}]$

(b). U is orthogonal $\Rightarrow$ $U^T = U^{-1}$. 즉, U는 invertible.

$A = U\Lambda U^T = U\Lambda U^{-1}$ $\quad$ U는 invertible, Λ는 $\text{diag}(\lambda_1, \ldots, \lambda_n)$ by (a). $\Rightarrow$ $u^{(i)}$ eigen vector, λ_i eigen value.

$I = U^{-1} \cdot U = U^{-1} \cdot [u^{(1)} \ u^{(2)} \cdots u^{(i)} \cdots u^{(n)}] = [U^{-1} u^{(1)} \cdots U^{-1} u^{(i)} \cdot U^{-1} u^{(n)}]$

$\therefore U^{-1} \cdot u^{(i)} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$ $\quad U\Lambda U^{-1} \cdot u^{(i)} = U\Lambda \cdot e^{(i)} = U\lambda_i \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \lambda_i \cdot u^{(i)}$

$$\underbrace{A = A^T \quad x^T A x \geq 0 \text{ for all } x \in \mathbb{R}^n}$$

$$A \Rightarrow \lambda_i(A) \overset{= \lambda_i}{\geq} 0 \text{ for each } i$$

$$\lambda_i(\underbrace{U \Lambda U^{-1}}_{A}) \geq 0$$

$$U^T U = I$$

$$U^{-1} U = I$$

$$A = A^T \quad U \Lambda U^T = U^T \Lambda^T U = U^T \Lambda U$$

$$U \Lambda U^{-1} = U^{-1} \Lambda U$$

$$\underbrace{x^T U^T}_{B} \Lambda \underbrace{U x}_{B}$$

$$B^T \Lambda B \geq 0$$

$$U^{-1} \underbrace{u^{(i)}}_{\text{eigen vector}} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

$$U^T u^{(i)} = e^{(i)} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

$$0 \leq u^{(i)T} \cdot A \cdot u^{(i)} = u^{(i)T} \cdot U \cdot \Lambda U^T \cdot u^{(i)}$$

$$= (U^T \cdot u^{(i)})^T \Lambda (U^T \cdot u^{(i)}) = e^{(i)T} \cdot \Lambda e^{(i)}$$

$$= e^{(i)T} \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \lambda_i \\ \vdots \\ 0 \end{bmatrix} = \lambda_i$$

$$\therefore \boxed{\lambda_i \geq 0}$$