Fundamental content after midterm- A brief review

MAT2040 Linear Algebra

Brief Review- Linear transformation

1. Definition for linear transformation: Let V, W be two vector spaces, and the mapping L from V to W is said to be a linear transformation if the following condition is satisfied:

$$L(\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2) = \alpha_1 L(\mathbf{v}_1) + \alpha_2 L(\mathbf{v}_2), \quad \forall \alpha_1,\ \alpha_2 \in \mathbb{R}, \quad \forall \mathbf{v}_1,\ \mathbf{v}_2 \in V. (*)$$

V is called the **domain** of the linear transformation, and W is called the **codomain** of the linear transformation.

2. Definition for Kernel of L, denoted by $\ker(L)$ is defined as

$$\ker(L) = \{\mathbf{v} \in V | L(\mathbf{v}) = \mathbf{0}_W\}.$$

3. Definition of range. Let S be a subspace of V, the **image** of S, denoted by $L(S)$, is defined by

$$L(S) = \{\mathbf{w} \in W | \exists\ \mathbf{v} \in S,\ s.t.\ L(\mathbf{v}) = \mathbf{w}\}$$

The image of the entire vector space V, i.e., $L(V)$ is called the **range** of L.

Brief Review- Linear transformation

(**Matrix Representation for linear transformation between Eulerian vector spaces w.r.t. standard bases**) If L is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, there is a $m \times n$ matrix A such that

$$L(\mathbf{x}) = A\mathbf{x}$$

for each $\mathbf{x} \in \mathbb{R}^n$. In fact, the jth column vector of $A = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$ is given by

$$\mathbf{a}_j = L(\mathbf{e}_j), \quad j = 1, 2, \cdots, n$$

where $\mathcal{E}_n = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is the standard basis of $\mathbb{R}^n$.

Brief Review- Linear transformation

(**Matrix Representation for General Vector Spaces**) If $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a basis for vector space V and $\mathcal{W} = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$ is a basis for vector space W, and L is a linear transformation mapping from vector space V to vector space W, then there is a $m \times n$ matrix A such that

$$[L(\mathbf{u})]_{\mathcal{W}} = A[\mathbf{u}]_{\mathcal{V}}, \quad \forall \mathbf{u} \in V$$

And in fact, the jth column of A is given by

$$\mathbf{a}_j = [L(\mathbf{v}_j)]_{\mathcal{W}}$$

and $A = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$.

Brief Review-Orthogonality

Let $\mathbf{x} = [x_1, \cdots, x_n]^T, \ \mathbf{y} = [y_1, \cdots, y_n]^T \in \mathbb{R}^n$, then

$$< \mathbf{x}, \mathbf{y} >= \mathbf{x}^{\mathbf{T}}\mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

$$\| \mathbf{x} \| = \sqrt{\mathbf{x}^T \mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

$$\mathbf{x}^T \mathbf{y} = \| \mathbf{x} \| \| \mathbf{y} \| \cos\theta, \quad 0 \leq \theta \leq \pi.$$

Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are said to be **orthogonal** if $\mathbf{x}^T \mathbf{y} = 0$. Denote $\mathbf{x} \perp \mathbf{y}$.

Brief Review-Orthogonality

(**Orthogonal Subspaces in** $\mathbb{R}^n$) Two subspaces X and Y of $\mathbb{R}^n$ are said to be orthogonal if

$$\mathbf{x}^T\mathbf{y} = 0,\ \forall\ \mathbf{x} \in X,\ \ \mathbf{y} \in Y.$$

Denoted by $X \perp Y$.

$$Y^{\perp} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^T\mathbf{y} = 0, \forall\ \mathbf{y} \in Y\}$$

Brief Review-Orthogonality

$A \in \mathbb{R}^{m \times n}$, (1)Null(A)$=Col(A^T)^{\perp}=Row(A)^{\perp}$
(2)Null(A^T)$=Col(A)^{\perp}=Row(A^T)^{\perp}$

Brief Review-Orthogonality

Theorem If S is a subspace of $\mathbb{R}^n$, then

$$\dim S + \dim S^{\perp} = n.$$

Furthermore, if $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ is a basis for S and $\{\mathbf{u}_{r+1}, \cdots, \mathbf{u}_n\}$ is a basis for $S^{\perp}$, then $\{\mathbf{u}_1, \cdots, \mathbf{u}_r, \mathbf{u}_{r+1}, \cdots, \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$.

$$\mathbb{R}^n = S \oplus S^{\perp}$$

(**Definition for Least square solution**) Given linear system $A\mathbf{x} = \mathbf{b}(A \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m)$, a vector $\hat{\mathbf{x}}(\mathbf{x} \in \mathbb{R}^n)$ that satisfies the minimum residual condition

$$\| r(\hat{\mathbf{x}}) \| = \min_{\mathbf{x}} \| r(\mathbf{x}) \|$$

is called the least square solution for $A\mathbf{x} = \mathbf{b}$, where $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$.

Brief Review-Orthogonality

(**Normal equations for the linear system**) Given the linear system $A\mathbf{x} = \mathbf{b}$ ($A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$), let the projection of $\mathbf{b}$ onto the subspace $Col(A)$ is $\mathbf{p}$, then there exists a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$, s.t.
$\mathbf{p} = A\hat{\mathbf{x}} \in Col(A), \mathbf{b} - A\hat{\mathbf{x}} \in Col(A)^\perp = Null(A^T)$ and
$\| \mathbf{b} - A\mathbf{x} \| \geq \| \mathbf{b} - A\hat{\mathbf{x}} \|$ for any $\mathbf{x} \in \mathbb{R}^n$.
$\mathbf{b} - A\hat{\mathbf{x}} \in Col(A)^\perp = Null(A^T)$ gives the condition

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$$

which is called the **normal equation**, and it is a $n \times n$ linear system.

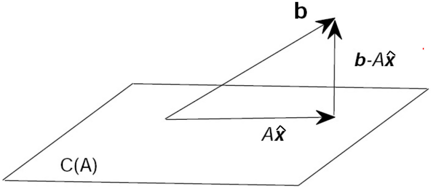


Figure: Projection of $\mathbf{b} \in V$ onto column space $Col(A)$.

Brief Review-Orthogonality

(**Unique Solution Condition for the Normal Equations**) If A is a $m \times n$ matrix of rank n, the normal equations

$$A^T A\hat{\mathbf{x}} = A^T \mathbf{b}$$

have a unique solution

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

and $\hat{\mathbf{x}}$ is the unique least square solution for the linear system $A\hat{\mathbf{x}} = \mathbf{b}$. The projection vector is given by $\mathbf{p} = A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T \mathbf{b}$ where $P = A(A^T A)^{-1} A^T$ is called the **projection matrix**

Brief Review-Orthogonality

1. (**Orthogonal Set in $\mathbb{R}^n$**) Let $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ be nonzero vectors from $\mathbb{R}^n$. If $\mathbf{u}_i^T \mathbf{u}_j = \langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ when $i \neq j$, then $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ is said to be an **orthogonal set**.
2. **Theorem** (**Orthogonal vectors are linearly independent**)
3. (**Orthonormal Set in $\mathbb{R}^n$**) Let $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ be nonzero vectors from $\mathbb{R}^n$. $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ is said to be the **orthonormal set** if
$$< \mathbf{u}_i, \mathbf{u}_j >= \mathbf{u}_i^T \mathbf{u}_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$
4. (**Orthonormal basis for $\mathbb{R}^n$**) $B = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ if $B = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is an orthonormal set in $\mathbb{R}^n$.

Brief Review-Orthogonality

1. (**Orthogonal Matrix**)
Let $Q \in \mathbb{R}^{n \times n}$, Q is said to be the orthogonal matrix if the column vectors of Q form an orthonormal set in $\mathbb{R}^n$ (also form an orthonormal basis for $\mathbb{R}^n$).

2. (**Equivalent Condition for Orthogonal Matrix**) Let $Q \in \mathbb{R}^{n \times n}$, Q is an orthogonal matrix if and only if $Q^T Q = I_n$. Q is an orthogonal matrix if and only if $Q^{-1} = Q^T$.

3. **Property for Orthogonal Matrix**)
Let $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, then

(a) $\| Q\mathbf{x} \| = \| \mathbf{x} \|$, $\forall \ \mathbf{x} \in \mathbb{R}^n$

(b) $\langle Q\mathbf{x}, Q\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$, $\forall \ \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

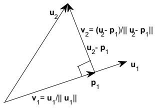
Brief Review-Gram-Schmidt process in $\mathbb{R}^n$

Question: Given an linearly independent set $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ in $\mathbb{R}^n$, how can we find an orthonormal set $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ such that $\mathrm{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_m) = \mathrm{Span}(\mathbf{v}_1, \cdots, \mathbf{v}_m)$?

(**Gram-Schmidt Process for m=3**)

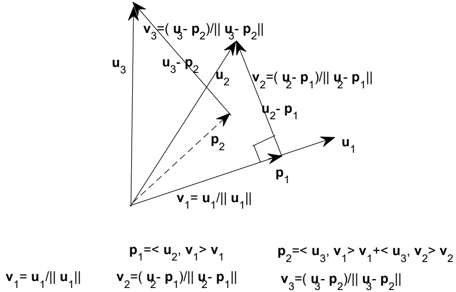
Step 1: normalize $\mathbf{u}_1$ to get $\mathbf{v}_1$, i.e., $\mathbf{v}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$

Step 2: project $\mathbf{u}_2$ onto $\mathrm{Span}(\mathbf{v}_1)$ to get $\mathbf{p}_1 = \langle \mathbf{u}_2, \mathbf{v}_1 \rangle \mathbf{v}_1$, then $\mathbf{r}_1 = \mathbf{u}_2 - \mathbf{p}_1 \perp \mathrm{Span}(\mathbf{u}_1)$. Set $\mathbf{v}_2 = \frac{\mathbf{r}_1}{\|\mathbf{r}_1\|} = \frac{\mathbf{u}_2 - \mathbf{p}_1}{\|\mathbf{u}_2 - \mathbf{p}_1\|}$, then $\{\mathbf{v}_1, \mathbf{v}_2\}$ are orthonormal set and $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$=$\mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2)$.



$v_1 = u_1/\|u_1\|$ $v_2 = (u_2 - p_1)/\|u_2 - p_1\|$

Brief Review-Gram-Schmidt process in $\mathbb{R}^n$



Step 3: project $\mathbf{u}_3$ onto $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$ to get $\mathbf{p}_2 = \langle \mathbf{u}_3, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}_3, \mathbf{v}_2 \rangle \mathbf{v}_2$, then $\mathbf{r}_2 = \mathbf{u}_3 - \mathbf{p}_2 \perp \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$, set $\mathbf{v}_3 = \frac{\mathbf{r}_2}{\|\mathbf{r}_2\|} = \frac{\mathbf{u}_3 - \mathbf{p}_2}{\|\mathbf{u}_3 - \mathbf{p}_2\|}$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ are orthonormal set and $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$.

QR decomposition

Let $A = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$ is a real matrix, whose the column vector set $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ is linearly independent. Gram-Schmidt process gives following orthonormal set

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{\| \mathbf{a}_1 \|}, \quad \mathbf{r}_1 = \mathbf{a}_2 - \mathbf{p}_1 \quad (\mathbf{p}_1 =< \mathbf{a}_2, \mathbf{q}_1 > \mathbf{q}_1), \quad \mathbf{q}_2 = \frac{\mathbf{r}_1}{\| \mathbf{r}_1 \|}$$

$$\mathbf{r}_2 = \mathbf{a}_3 - \mathbf{p}_2 \quad (\mathbf{p}_2 =< \mathbf{a}_3, \mathbf{q}_1 > \mathbf{q}_1 + < \mathbf{a}_3, \mathbf{q}_2 > \mathbf{q}_2), \quad \mathbf{q}_3 = \frac{\mathbf{r}_2}{\| \mathbf{r}_2 \|}$$

$(\mathbf{q}_2 \perp \mathrm{Span}(\mathbf{q}_1) = \mathrm{Span}(\mathbf{a}_1), \quad \mathbf{q}_3 \perp \mathrm{Span}(\mathbf{q}_1, \mathbf{q}_2) = \mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2))$
The above relations can be rewritten as

$$\begin{aligned}
\mathbf{a}_1 &= \|\mathbf{a}_1\|\mathbf{q}_1, \\
\mathbf{a}_2 &=< \mathbf{a}_2, \mathbf{q}_1 > \mathbf{q}_1 + \| \mathbf{r}_1 \| \mathbf{q}_2, \\
\mathbf{a}_3 &=< \mathbf{a}_3, \mathbf{q}_1 > \mathbf{q}_1 + < \mathbf{a}_3, \mathbf{q}_2 > \mathbf{q}_2 + \| \mathbf{r}_2 \| \mathbf{q}_3.
\end{aligned}$$

QR decomposition

This gives

$$A = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3] = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] \begin{bmatrix} \|\mathbf{a}_1\| & < \mathbf{a}_2, \mathbf{q}_1 > & < \mathbf{a}_3, \mathbf{q}_1 > \\ 0 & \|\mathbf{r}_1\| & < \mathbf{a}_3, \mathbf{q}_2 > \\ 0 & 0 & \|\mathbf{r}_2\| \end{bmatrix}$$
$$\triangleq QR.$$

This is called the QR factorization. Here $Q = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3]$, and

$$R = \begin{bmatrix} \|\mathbf{a}_1\| & < \mathbf{a}_2, \mathbf{q}_1 > & < \mathbf{a}_3, \mathbf{q}_1 > \\ 0 & \|\mathbf{r}_1\| & < \mathbf{a}_3, \mathbf{q}_2 > \\ 0 & 0 & \|\mathbf{r}_2\| \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1^T \mathbf{a}_1 & \mathbf{q}_1^T \mathbf{a}_2 & \mathbf{q}_1^T \mathbf{a}_3 \\ 0 & \mathbf{q}_2^T \mathbf{a}_2 & \mathbf{q}_2^T \mathbf{a}_3 \\ 0 & 0 & \mathbf{q}_3^T \mathbf{a}_3 \end{bmatrix}$$

since $< \mathbf{q}_1, \mathbf{a}_1 >= \mathbf{q}_1^T \mathbf{a}_1 = \|\mathbf{a}_1\|$, $< \mathbf{q}_2, \mathbf{a}_2 >= \mathbf{q}_2^T \mathbf{a}_2 = \|\mathbf{r}_1\|$, $< \mathbf{q}_3, \mathbf{a}_3 >= \mathbf{q}_3^T \mathbf{a}_3 = \|\mathbf{r}_2\|$.

Brief Review-Eigenvalue and Eigenvectors

1. Let A be a square matrix with size $n \times n$ ($A \in \mathbb{R}^{n\times n}$ or $A \in \mathbb{C}^{n\times n}$), if there exists a scalar λ ($\lambda \in \mathbb{R}$ or $\lambda \in \mathbb{C}$) and nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$, then λ is called the **eigenvalue** (or **characteristic value**) and $\mathbf{x}$ is called the **eigenvector** (or **characteristic vector**) w.r.t λ.
2. (**Characteristic Polynomial**) Let A is a $n \times n$ matrix ($A \in \mathbb{R}^{n\times n}$ or $A \in \mathbb{C}^{n\times n}$) and λ is a variable, then $p_A(\lambda) = \det(A - \lambda I)$ is the characteristic polynomial with degree n. The roots of the characteristic polynomial are the eigenvalues of A, the number of eigenvalues (counting with multiplicity) are n.
3. (**Product and Sum of Eigenvalues**) Let $A = (a_{ij})_{n\times n}$ be a square matrix ($A \in \mathbb{R}^{n\times n}$ or $A \in \mathbb{C}^{n\times n}$), $\lambda_i (i = 1, 2, \cdots, n)$ are the eigenvalues, then (1) $\det(A) = \prod_{i=1}^{n} \lambda_i$, (2) $\sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i$, where $\sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn} = Trace(A)$ is called the trace of A.

Brief Review-Eigenvalue and Eigenvectors

1. (**Diagonalizable**) A $n \times n$ matrix A is said to be diagonalizable if there exists a nonsingular matrix X and a diagonal matrix D such that $X^{-1}AX = D$.

Brief Review-Eigenvalue and Eigenvectors

2. **Theorem (Sufficient and Necessary Condition for Diagonalization**) A $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

3. **Corollary** Let A be a matrix with size $n \times n$, if A has n distinct eigenvalues, then A is diagonalizable.
However, if A has eigenvalues with multiplicity ≥ 2, then A may or may not be diagonalizable

Brief Review-Diagonalization for real symmetric matrix

1. **Theorem** Let $A \in \mathbb{R}^{n \times n}$ be the real symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$, then
(1) $\lambda_i \in \mathbb{R}, \ \forall\ i = 1, \cdots, n$. (**The eigenvalues of real symmetric matrices are real numbers.**)
(2) If $\lambda_i \neq \lambda_j$, $\mathbf{x}_i$ is the eigenvectors w.r.t λ_i, $\mathbf{x}_j$ is the is the eigenvectors w.r.t λ_j, then $\mathbf{x}_i$, $\mathbf{x}_j$ are orthogonal.(**For real symmetric matrices, the eigenvectors belonging to different eigenvalues are orthogonal.**)

2. **Theorem** (**Spectral Theorem (eigen decomposition theorem) for Real Symmetric Matrix**) If A is a real symmetric matrix, then there exists an orthogonal matrix Q that diagonalizes A, i.e.,
$Q^T A Q = Q^{-1} A Q = \Lambda$ (Λ is a diagonal matrix)

Brief Review-Quadratic form

Let $\mathbf{x} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$ be symmetric, then

(1) The quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is called **positive definite** if $f(\mathbf{x}) > 0$ for any $\mathbf{x} \neq \mathbf{0}$. And correspondingly, A is called **positive definite matrix**.

(2) The quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is called **positive semidefinite** if $f(\mathbf{x}) \geq 0$ for any $\mathbf{x} \neq \mathbf{0}$. And correspondingly, A is called **positive semidefinite matrix**.

(3) The quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is called **indefinite** if $f(\mathbf{x})$ takes different signs.

Brief Review-Quadratic form

The negative definite and negative semidefinite can defined as follows:

(4) The quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is called **negative definite** if $f(\mathbf{x}) < 0$ for any $\mathbf{x} \neq \mathbf{0}$. And correspondingly, A is called **negative definite matrix**.

(5) The quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is called **negative semidefinite** if $f(\mathbf{x}) \geq 0$ for any $\mathbf{x} \neq \mathbf{0}$. And correspondingly, A is called **negative semidefinite matrix**.

Brief Review-Quadratic form

Theorem:
Let $A \in \mathbb{R}^{n\times n}$ be symmetric, then A is positive definite if only if all eigenvalues are positive.

Proof. Since A is symmetric, by spectral theorem for real symmetric matrix, there exists an orthogonal matrix Q such that $Q^{-1}AQ = Q^TAQ = D$, where D is the diagonal matrix. Let $\hat{\mathbf{x}} = Q^T\mathbf{x}$ then $\mathbf{x} = Q\hat{\mathbf{x}}$ and $\mathbf{x}^TA\mathbf{x} = (Q\hat{\mathbf{x}})^TAQ\hat{\mathbf{x}} = \hat{\mathbf{x}}^TQ^TAQ\hat{\mathbf{x}} = \hat{\mathbf{x}}^TD\hat{\mathbf{x}}$. Since Q is invertible and $\hat{\mathbf{x}} = Q^T\mathbf{x}$, thus

$$\mathbf{x}^TA\mathbf{x} > 0,\ \forall\mathbf{x} \neq \mathbf{0} \Leftrightarrow \hat{\mathbf{x}}^TD\hat{\mathbf{x}} > 0,\ \forall\hat{\mathbf{x}} \neq \mathbf{0}$$

Thus, A is positive definite $\Leftrightarrow$ the entries in diagonal elements of D are all positive$\Leftrightarrow$ all eigenvalues of A are positive.

Brief Review-Quadratic form

Remark.

1. Let $A \in \mathbb{R}^{n \times n}$ be symmetric, then A is negative definite if only if all eigenvalues are negative.
2. Let $A \in \mathbb{R}^{n \times n}$ be symmetric, then A is indefinite if only if eigenvalues have different signs.

Theorem:

A ($A \in \mathbb{R}^{n \times n}$) is a symmetric positive definite matrix if and only if all leading principal submatrices have positive determinants.