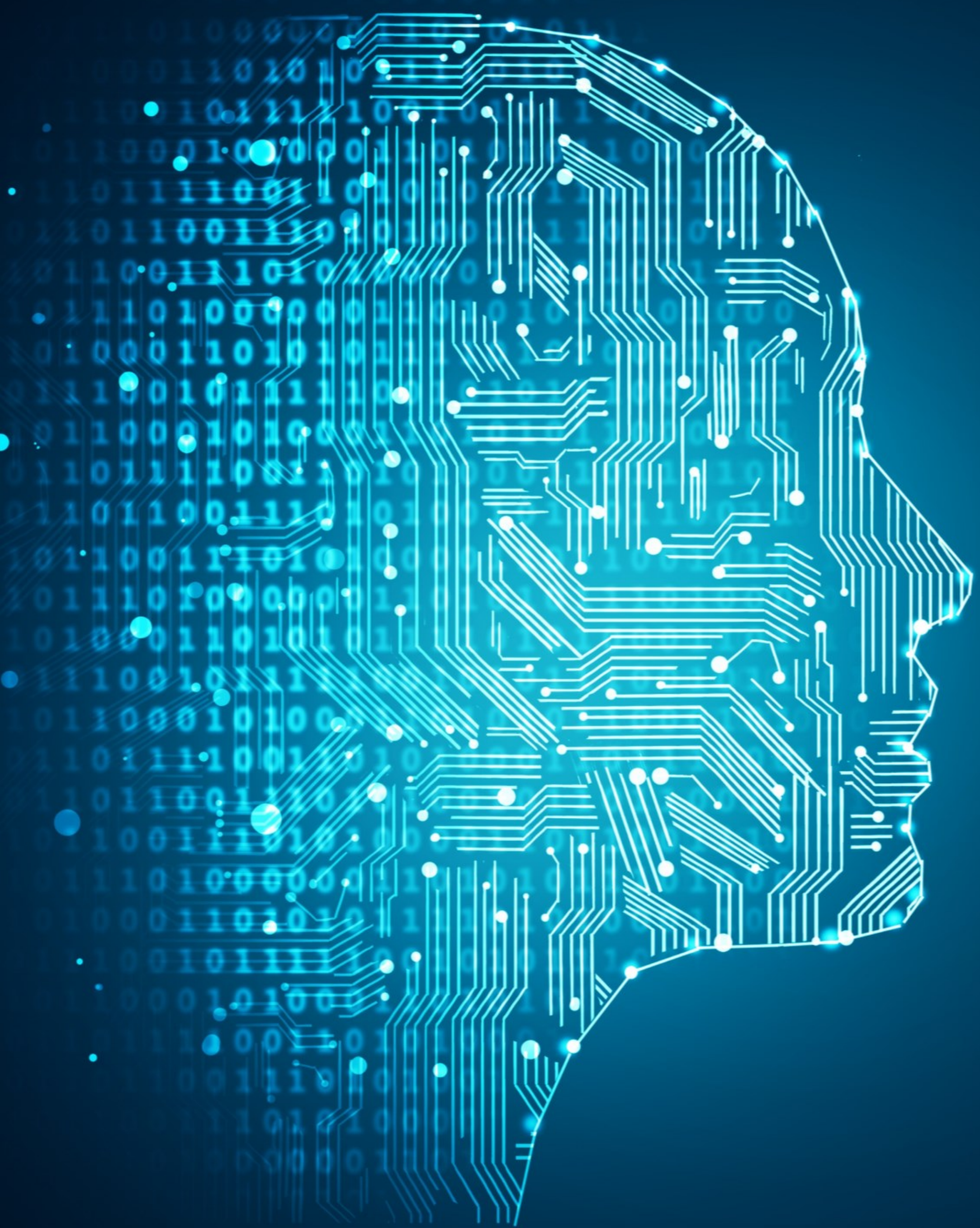


AI6103 Linear Algebra

Boyang Li, Albert

School of Computer Science and Engineering

Nanyang Technological University

Orthonormal Matrix

- A matrix A is orthonormal if its rows and columns have unit length and are orthogonal to each other
- Take the rows for example. If $A = \begin{bmatrix} a_1^\top \\ a_2^\top \\ a_3^\top \end{bmatrix}$, that means $a_1^\top a_2 = 0, a_2^\top a_3 = 0, a_1^\top a_3 = 0.$
- And $a_1^\top a_1 = 1, a_2^\top a_2 = 1, a_3^\top a_3 = 1$
- Thus, $AA^\top = I$
- Since A has an inverse, $A^\top = A^{-1}$
- Thus, $AA^\top = A^\top A = I$

Below are a few examples of small orthogonal matrices and possible interpretations.

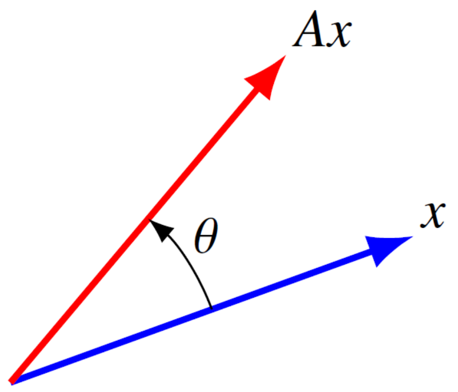
- $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (identity transformation)
- $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 0.96 & -0.28 \\ 0.28 & 0.96 \end{bmatrix}$ (rotation by 16.26°)
- $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ (reflection across *x*-axis)
- $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ (permutation of coordinate axes)

Geometric Interpretation

- A matrix represents a transformation of a vector
- For example, $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ performs a stretch operation for the first dimension and does nothing to the second dimension

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x \\ y \end{bmatrix}$$

- $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ performs a rotation operation
- $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ flips the vector x (i.e. 180-degree rotation)



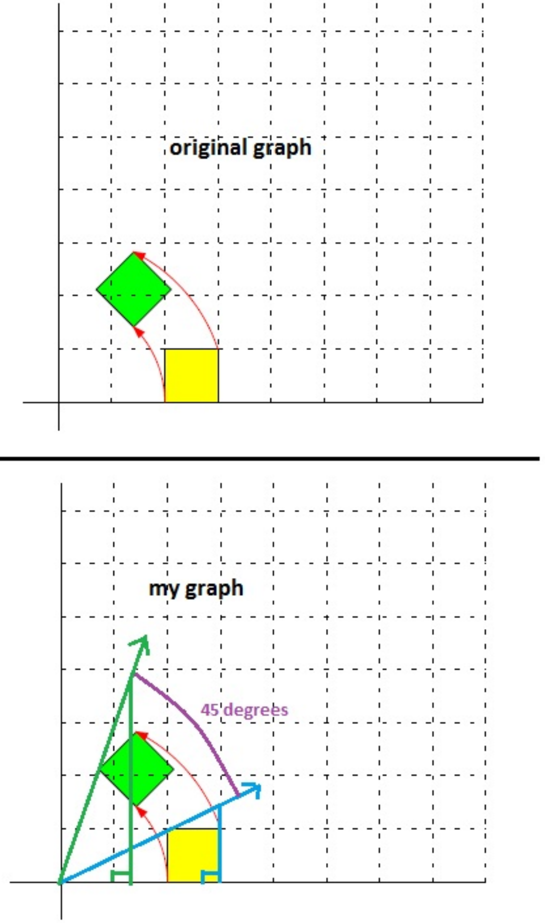
Example: Rotation Matrix

To rotate 45° about the origin, we apply the matrix

$$\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

Note: $\frac{\sqrt{2}}{2} = \cos 45^\circ = \sin 45^\circ$, so this is the same as

$$\begin{pmatrix} \cos 45^\circ & -\sin 45^\circ \\ \sin 45^\circ & \cos 45^\circ \end{pmatrix}$$



Counter Clockwise $\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$

Clockwise $\begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}$

Matrix As Transformation

- A matrix can be a combination of multiple transformations.
- $\begin{bmatrix} \cos\frac{\pi}{2} & -\sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ rotates a vector by 90 degrees counterclockwise.
- Therefore, $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}$ performs rotation first and the stretching afterwards.

Eigenvalues and Eigenvectors

- A non-zero vector v is an eigenvector of square matrix A if
$$Av = \lambda v$$
- λ is called the eigenvalue associated with v.
- Generically speaking, Av does not change the direction of v (except when $\lambda = 0$).
- This is very special. Most vectors change direction after multiplication with a matrix A.
- Thus, v is an important characteristic of the transformation represented by A.

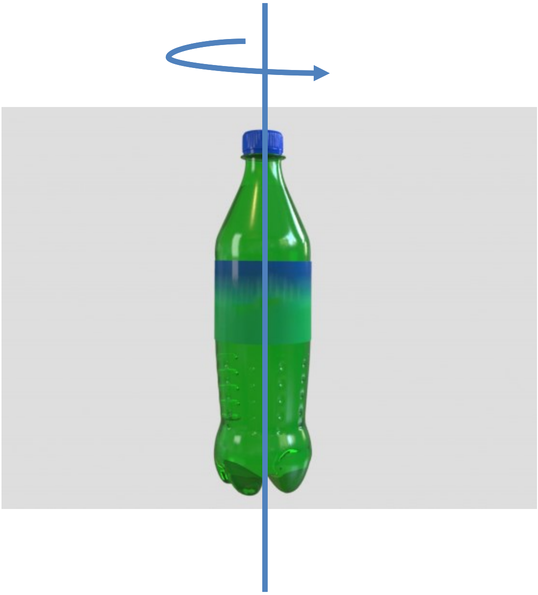
Examples

- The identity matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ does not change anything, so its eigenvectors can be any non-zero vector and the eigenvalue is always 1.

- The matrix $\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$ stretches the first dimension and compresses the second dimension, so that any vector have non-zero values in both dimension will change direction.

- Thus, the eigenvectors are $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and their multiples. The eigenvalues are 2 and 0.5 respectively.

Example



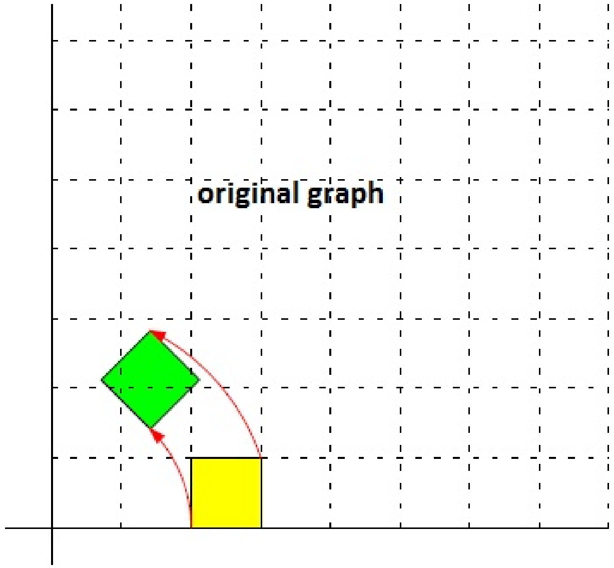
A matrix can represent rotation around an axis.
The only vector that does not change direction is the direction of the axis.

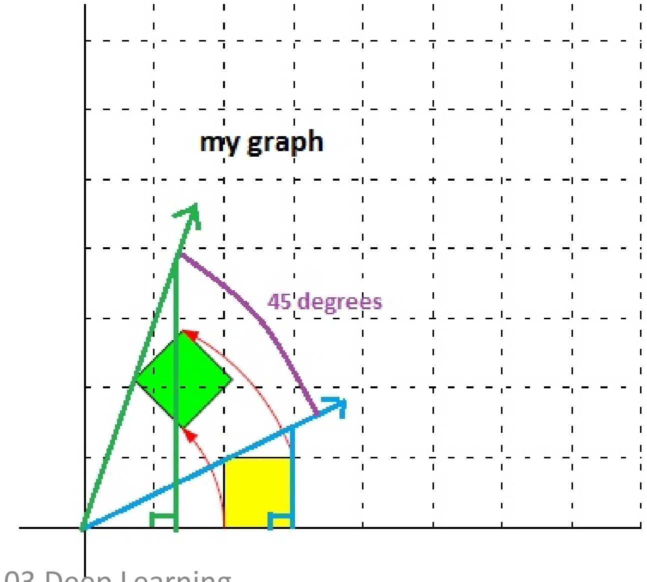
Consider matrix $\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$, which has the eigenvector $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and the eigenvalue 1. The rotation happens around the z-axis.

Other cases may not be so obvious. Rotating 30^o around the z-axis and 45^o around the y-axis yields the matrix

$\begin{bmatrix} 0.6123 & -0.356 & 0.707 \\ 0.5 & 0.866 & 0 \\ -0.6123 & 0.356 & 0.707 \end{bmatrix}$, which has the rotational axis $\begin{bmatrix} 0.22 \\ 0.82 \\ 0.53 \end{bmatrix}$

Example





- If we perform rotation along all N axes in N-dimensional space, the only direction that does not change is the 0 direction.
- However, an eigenvector must not be zero.
- In general, rotational matrices have complex eigenvectors and eigenvalues.

Eigendecomposition

- A square $n \times n$ matrix M can be written as
$$M = Q\Lambda Q^{-1}$$
- Q is a $n \times n$ matrix whose i-th column is the eigenvector q_i. These eigenvectors are usually normalized to length 1. (but not always so, as we can always have cancellation from Q^{-1})
- Λ is a diagonal matrix with the eigenvalues on the diagonal.
- It requires M to be diagonalizable (in real or complex domains).

Eigendecomposition

Eigendecomposition reveals inherent characteristics of the matrix. We will see its use in the optimization of neural networks.

- A square $n \times n$ matrix M can be written as

$$M = Q \Lambda Q^{-1}$$

- Derivation
- For eigenvector q, we have $Mq = \lambda q$
- For multiple eigenvectors $q_1, \ldots, q_n$, we can write

$$M[q_1, \ldots, q_n] = [q_1, \ldots, q_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \text{ or } MQ = Q\Lambda$$

- For a diagonalizable matrix, Q is full rank. Thus, we can multiple its inverse on both side, arriving at the equation.

The Order of Eigenvalues

- An $n \times n$ real matrix M can be written as

$$M = Q\Lambda Q^{-1} = [q_1\ q_2\ q_3\ \cdots q_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} v_1^\top \\ v_2^\top \\ v_3^\top \\ \vdots \\ v_n^\top \end{bmatrix}$$

$$= \sum_{i=1}^{n} \lambda_i q_i v_i^\top$$

This sum does not care about the ordering of λ

- where q_i are the columns of Q and v_i are the columns of V.
- We can freely switch the positions of eigenvalues, as long as we keep (λ_i, q_i, v_i) together
- Many math texts assume the eigenvalues are sorted in either descending or ascending order

Symmetric Matrix

- A square matrix A is symmetric iff $A_{ij} = A_{ji}, \forall i, j$

$$\begin{bmatrix} 1 & 3 & 7 \\ 3 & 2 & 6 \\ 7 & 6 & 8 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & -4 & 38 \\ 2 & -1 & 71 & 2 \\ -4 & 71 & 9 & 56 \\ 38 & 2 & 56 & 30 \end{bmatrix}$$

- A square matrix is anti-symmetric (or skew-symmetric) iff $A_{ij} = -A_{ji}, \forall i \neq j$

$$\begin{bmatrix} 1 & -3 & -7 \\ 3 & 2 & -6 \\ 7 & 6 & 8 \end{bmatrix}$$

- If a square matrix A is symmetric, the following statements are true
  - $A = A^\top$
  - A^k is symmetric for integer k
  - A^{-1} is symmetric, if it exists
- Any matrix X can be written as the sum of a symmetric and an anti-symmetric matrix.
  - Symmetric part: $\frac{1}{2}(X + X^\top)$
  - Anti-symmetric part: $\frac{1}{2}(X - X^\top)$

Hessian Matrix

- Consider the function $f(x)$ that takes a scalar x, its derivative $\frac{\partial f(x)}{\partial x}$ is a scalar.
- Consider the function $g(\boldsymbol{x})$ that takes a p-dimensional vector $\boldsymbol{x}$, its derivative $\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}$ (a.k.a, the Jacobian) is a vector.

$$\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left[\frac{\partial g(\boldsymbol{x})}{\partial x_1}, \frac{\partial g(\boldsymbol{x})}{\partial x_2}, \dots, \frac{\partial g(\boldsymbol{x})}{\partial x_p}\right]^\top$$

- The second derivative is a matrix, known as the Hessian matrix.
- The Hessian is a symmetric matrix.

$$H = \begin{bmatrix} \frac{\partial^2 g}{\partial x_1^2} & \frac{\partial^2 g}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 g}{\partial x_1 \partial x_p} \\ \frac{\partial^2 g}{\partial x_2 \partial x_1} & \frac{\partial^2 g}{\partial x_2^2} & \cdots & \frac{\partial^2 g}{\partial x_2 \partial x_p} \\ \vdots & \vdots & \cdots & \cdots \\ \frac{\partial^2 g}{\partial x_p \partial x_1} & \frac{\partial^2 g}{\partial x_p \partial x_2} & \cdots & \frac{\partial^2 g}{\partial x_p^2} \end{bmatrix}$$

Eigendecomposition

- A square $n \times n$ matrix M can be written as
$$M = Q\Lambda Q^{-1}$$
- For real symmetric matrices,
$$M = Q\Lambda Q^{\top}$$
  – All eigenvalues are real. The eigenvectors are real and usually chosen to be orthonormal. That is, $Q^{-1} = Q^{\top}$

Positive Definiteness

- A square matrix A is symmetric iff $A_{ij} = A_{ji}, \forall i, j$

$$\begin{bmatrix} 1 & 3 & 7 \\ 3 & 2 & 6 \\ 7 & 6 & 8 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & -4 & 38 \\ 2 & -1 & 71 & 2 \\ -4 & 71 & 9 & 56 \\ 38 & 2 & 56 & 30 \end{bmatrix}$$

- A square matrix is anti-symmetric (or skew-symmetric) iff $A_{ij} = -A_{ji}, \forall i \neq j$

$$\begin{bmatrix} 1 & -3 & -7 \\ 3 & 2 & -6 \\ 7 & 6 & 8 \end{bmatrix}$$

- A matrix A is positive definite (PD) iff for all vector $x \neq 0$, $x^\top A x > 0$
- Recall $x^\top A x = \sum_i \sum_j A_{ij} x_i x_j$

$$x^\top \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix} x = x_1^2 + x_2^2 + x_3^2 + (x_1 - x_2)^2 + (x_3 - x_2)^2 > 0, \forall x \neq 0$$

- Similarly,

$$x^\top \begin{bmatrix} 2 & -2 & 0 \\ 0 & 3 & 0 \\ 0 & -2 & 2 \end{bmatrix} x = x_1^2 + x_2^2 + x_3^2 + (x_1 - x_2)^2 + (x_3 - x_2)^2 > 0, \forall x \neq 0$$

Definite Matrix

- A square matrix A is symmetric iff $A_{ij} = A_{ji}, \forall i, j$
- A square matrix is anti-symmetric iff $A_{ij} = -A_{ji}, \forall i \neq j$

- A matrix A is positive definite (PD) iff for all vector $x \neq 0$, $x^\top A x > 0$

Some textbooks require all PD matrices to be symmetric.

Consider a matrix C that is PD but not symmetric. It can be made symmetric by shifting values between off-diagonal terms.

$$x^\top \begin{bmatrix} 2 & \boxed{-2} & 0 \\ \boxed{0} & 3 & 0 \\ 0 & -2 & 1 \end{bmatrix} x = x^\top \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 1 \end{bmatrix} x$$

That is, we can write $C = A + B$ where A is a symmetric PD matrix and B is an anti-symmetric matrix with zero diagonal.

Definite Matrix

- A square matrix A is symmetric iff $A_{ij} = A_{ji}, \forall i, j$
- A square matrix is anti-symmetric iff $A_{ij} = -A_{ji}, \forall i \neq j$
- A matrix A is positive definite (PD) iff for all vector $x \neq 0$, $x^\top A x > 0$
- Non-symmetric PD matrices can be made symmetric by adding an anti-symmetric zero-diagonal matrix to it.
- If A is PD and B is PD, $A + B$ is PD
$$x^\top (A + B)x = x^\top A x + x^\top B x > 0$$
- If A is PD and scalar α is positive, αA is PD
$$x^\top (\alpha A)x = \alpha x^\top A x > 0$$
- If A is symmetric and PD, all eigenvalues of A are positive.

Definite Matrix

- **Positive definite**: for all vector $x \neq 0$, $x^\top Ax > 0$
- **Positive semi-definite**: $x^\top Ax \geq 0$
- **Negative definite**: for all vector $x \neq 0$, $x^\top Ax < 0$
- **Negative semi-definite**: $x^\top Ax \leq 0$

- **Positive definite**: All eigenvalues > 0
- **Positive semi-definite**: All eigenvalues ≥ 0
- **Negative definite**: All eigenvalues < 0
- **Negative semi-definite**: All eigenvalues ≤ 0

Some matrices do not belong to any of these categories. For example, matrices with both positive and negative eigenvalues.

Definite Matrix

- **Positive definite**: for all vector $x \neq 0$, $x^\top A x > 0$
- **Positive semi-definite**: $x^\top A x \geq 0$
- **Negative definite**: for all vector $x \neq 0$, $x^\top A x < 0$
- **Negative semi-definite**: $x^\top A x \leq 0$

- **Positive definite**: All eigenvalues > 0

Proof:

Let $x \neq 0$ be an eigenvector of A, $x^\top A x = \lambda x^\top x$

Since A is PD, $\lambda x^\top x > 0$.

Note $x^\top x > 0$, so the eigenvalue λ must be greater than 0.

Another example

$$A = \begin{bmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{bmatrix}$$

is positive semi-definite.

Proof:

$$A = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix}$$

$$x^\top A x = \left(x^\top \begin{bmatrix} a \\ b \\ c \end{bmatrix} \right) ([a\ b\ c]x) = z^\top z \geq 0$$

Any matrix in the form of $M^\top M$ is PSD.