

# Mathematics for Machine Learning

## — Linear Algebra

Eigenvalues, Eigenvectors, Eigenspaces, Cholesky Decomposition & Diagonalization

Joseph Chuang-Chieh Lin

Department of Computer Science & Engineering,  
National Taiwan Ocean University

Fall 2025

## Credits for the resource

- The slides are based on the textbooks:
  - *Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong: Mathematics for Machine Learning. Cambridge University Press. 2020.*
  - *Howard Anton, Chris Rorres, Anton Kaul: Elementary Linear Algebra. Wiley. 2019.*
- We could partially refer to the monograph:  
*Francesco Orabona: A Modern Introduction to Online Learning.*  
<https://arxiv.org/abs/1912.13213>

- Matrix decomposition or matrix factorization.

- Matrix decomposition or matrix factorization.
- Three matrix decompositions will be introduced.

# Outline

- 1 Eigenvalues & Eigenvectors
- 2 Cholesky Decomposition
- 3 Eigendecomposition & Diagonalization

# Outline

- 1 Eigenvalues & Eigenvectors
- 2 Cholesky Decomposition
- 3 Eigendecomposition & Diagonalization

# Characteristic Polynomial

## Characteristic Polynomial

For  $\lambda \in \mathbb{R}$  and a square matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,

$$\begin{aligned} p_{\mathbf{A}}(\lambda) &:= \det(\mathbf{A} - \lambda \mathbf{I}) \\ &= (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n) \\ &= c_0 + c_1 \lambda + \cdots + c_{n-1} \lambda^{n-1} + (-1)^n \lambda^n, \end{aligned}$$

for  $c_0, \dots, c_{n-1} \in \mathbb{R}$ , is called the **characteristic polynomial** of  $\mathbf{A}$ .

Note that

- $c_0 = \det(\mathbf{A})$

# Characteristic Polynomial

## Characteristic Polynomial

For  $\lambda \in \mathbb{R}$  and a square matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,

$$\begin{aligned} p_{\mathbf{A}}(\lambda) &:= \det(\mathbf{A} - \lambda \mathbf{I}) \\ &= (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n) \\ &= c_0 + c_1 \lambda + \cdots + c_{n-1} \lambda^{n-1} + (-1)^n \lambda^n, \end{aligned}$$

for  $c_0, \dots, c_{n-1} \in \mathbb{R}$ , is called the **characteristic polynomial** of  $\mathbf{A}$ .

Note that

- $c_0 = \det(\mathbf{A}) = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$ .
- $c_{n-1} = (-1)^{n-1} \text{tr}(\mathbf{A})$



# Characteristic Polynomial

## Characteristic Polynomial

For  $\lambda \in \mathbb{R}$  and a square matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,

$$\begin{aligned} p_{\mathbf{A}}(\lambda) &:= \det(\mathbf{A} - \lambda \mathbf{I}) \\ &= (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n) \\ &= c_0 + c_1 \lambda + \cdots + c_{n-1} \lambda^{n-1} + (-1)^n \lambda^n, \end{aligned}$$

for  $c_0, \dots, c_{n-1} \in \mathbb{R}$ , is called the **characteristic polynomial** of  $\mathbf{A}$ .

Note that

- $c_0 = \det(\mathbf{A}) = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$ .
- $c_{n-1} = (-1)^{n-1} \operatorname{tr}(\mathbf{A}) = (-1)^{n-1} (\lambda_1 + \lambda_2 + \cdots + \lambda_n)$ .

back

## Example

Given  $\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$ ,

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \begin{vmatrix} 3 - \lambda & 0 \\ 8 & -1 - \lambda \end{vmatrix} = (3 - \lambda)(-1 - \lambda) = \lambda^2 - 2\lambda - 3.$$

## Example

Given  $\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$ ,

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \begin{vmatrix} 3 - \lambda & 0 \\ 8 & -1 - \lambda \end{vmatrix} = (3 - \lambda)(-1 - \lambda) = \lambda^2 - 2\lambda - 3.$$

Given  $\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$ ,

$$\det(\mathbf{B} - \lambda \mathbf{I}) = \begin{vmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ 4 & -17 & 8 - \lambda \end{vmatrix} = -\lambda^3 + 8\lambda^2 - 17\lambda + 4.$$

# Eigenvalue Equation

## Eigenvalues & Eigenvectors

Let  $\mathbf{A} \in \mathbb{R}^{n \times n}$  be a square matrix. Then

- $\lambda \in \mathbb{R}$  is an **eigenvalue** of  $\mathbf{A}$  and
- $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$  is the corresponding **eigenvector** of  $\mathbf{A}$

if  $\mathbf{Ax} = \lambda\mathbf{x}$ .

$$\mathbf{Ax} = \lambda\mathbf{x} \iff (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$$

Equivalent statements:

- $\lambda$  is an eigenvalue of  $\mathbf{A} \in \mathbb{R}^{n \times n}$ .
- There exists an  $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$  with  $\mathbf{Ax} = \lambda\mathbf{x}$  (i.e.,  $(\mathbf{A} - \lambda\mathbf{I}_n)\mathbf{x} = \mathbf{0}$ ) that can be solved non-trivially (i.e.,  $\mathbf{x} \neq \mathbf{0}$ ).
- $\text{rank}(\mathbf{A} - \lambda\mathbf{I}_n) < n$ .
- $\det(\mathbf{A} - \lambda\mathbf{I}_n) = 0$ .

## Remark

- Eigenvectors are NOT unique.

## Remark

- Eigenvectors are NOT unique.
- Suppose  $\mathbf{x}$  is an eigenvector of  $\mathbf{A}$  w.r.t. eigenvalue  $\lambda$ , then for any  $c \in \mathbb{R} \setminus \{0\}$

$$\mathbf{A}(c\mathbf{x}) = c\mathbf{A}\mathbf{x} = c\lambda\mathbf{x} = \lambda(c\mathbf{x}).$$

# Theorems (or Definitions)

## Theorem

$\lambda \in \mathbb{R}$  is an eigenvalue of  $\mathbf{A} \in \mathbb{R}^{n \times n}$  if and only if  $\lambda$  is a root of the characteristic polynomial  $p_{\mathbf{A}}(\lambda)$  of  $\mathbf{A}$ .



# Theorems (or Definitions)

## Theorem

$\lambda \in \mathbb{R}$  is an eigenvalue of  $\mathbf{A} \in \mathbb{R}^{n \times n}$  if and only if  $\lambda$  is a root of the characteristic polynomial  $p_{\mathbf{A}}(\lambda)$  of  $\mathbf{A}$ .

## Algebraic Multiplicity

Suppose that matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  has an eigenvalue  $\lambda_i$ . The **algebraic multiplicity** of  $\lambda_i$  is the number of times the root appears in the characteristic polynomial.

- Denoted by  $\text{am}(\lambda_i)$

# Theorems (or Definitions)

## Theorem

$\lambda \in \mathbb{R}$  is an eigenvalue of  $\mathbf{A} \in \mathbb{R}^{n \times n}$  if and only if  $\lambda$  is a root of the characteristic polynomial  $p_{\mathbf{A}}(\lambda)$  of  $\mathbf{A}$ .

## Algebraic Multiplicity

Suppose that matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  has an eigenvalue  $\lambda_i$ . The **algebraic multiplicity** of  $\lambda_i$  is the number of times the root appears in the characteristic polynomial.

- Denoted by  $\text{am}(\lambda_i)$

# Theorems (or Definitions)

## Eigenspace

For  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , the set of all eigenvectors of  $\mathbf{A}$  associated with the eigenvalue  $\lambda$  spans the **eigenspace** of  $\mathbf{A}$  (denoted by  $E_\lambda$ ).

## Geometric Multiplicity

$\dim(E_\lambda)$  is called the **geometric multiplicity** of  $\lambda$ .

- Denoted by  $\text{gm}(\lambda)$ .

## Eigenspectrum (Spectrum)

The set of all eigenvalues of  $\mathbf{A}$  is called the **eigenspectrum** (or spectrum) of  $\mathbf{A}$ .

## Relation b/w $\text{am}(\lambda)$ & $\text{gm}(\lambda)$

Geometric multiplicity  $\leq$  Algebraic multiplicity

Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$  and assume that  $\lambda$  is an eigenvalue of  $\mathbf{A}$ , then

$$\text{gm}(\lambda) \leq \text{am}(\lambda).$$

- Assume that  $\dim(E_\lambda) = k \leq n$  and let  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  is a basis of  $E_\lambda$  such that  $\mathbf{v}_1, \dots, \mathbf{v}_k$  are eigenvectors w.r.t.  $\lambda$ .

## Relation b/w $\text{am}(\lambda)$ & $\text{gm}(\lambda)$

Geometric multiplicity  $\leq$  Algebraic multiplicity

Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$  and assume that  $\lambda$  is an eigenvalue of  $\mathbf{A}$ , then

$$\text{gm}(\lambda) \leq \text{am}(\lambda).$$

- Assume that  $\dim(E_\lambda) = k \leq n$  and let  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  is a basis of  $E_\lambda$  such that  $\mathbf{v}_1, \dots, \mathbf{v}_k$  are eigenvectors w.r.t.  $\lambda$ .
- $\exists \mathbf{v}_{k+1}, \dots, \mathbf{v}_n$  such that  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  is a basis of  $\mathbb{R}^n$ .

## Relation b/w $\text{am}(\lambda)$ & $\text{gm}(\lambda)$

Geometric multiplicity  $\leq$  Algebraic multiplicity

Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$  and assume that  $\lambda$  is an eigenvalue of  $\mathbf{A}$ , then

$$\text{gm}(\lambda) \leq \text{am}(\lambda).$$

- Assume that  $\dim(E_\lambda) = k \leq n$  and let  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  is a basis of  $E_\lambda$  such that  $\mathbf{v}_1, \dots, \mathbf{v}_k$  are eigenvectors w.r.t.  $\lambda$ .
- $\exists \mathbf{v}_{k+1}, \dots, \mathbf{v}_n$  such that  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  is a basis of  $\mathbb{R}^n$ .
- Denote by  $\mathbf{P} = [\mathbf{U} \ \mathbf{V}]$  for  $\mathbf{U} = [\mathbf{v}_1 \cdots \mathbf{v}_k]$  and  $\mathbf{V} = [\mathbf{v}_{k+1} \cdots \mathbf{v}_n]$  (note:  $\mathbf{P}$  is invertible).

## Relation b/w $\text{am}(\lambda)$ & $\text{gm}(\lambda)$ (contd.)

- $\because \mathbf{P}$  is invertible  $\Rightarrow$  Let  $\mathbf{P}^{-1} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$ , where  $\mathbf{X} \in \mathbb{R}^{k \times n}$  and  $\mathbf{Y} \in \mathbb{R}^{(n-k) \times n}$ .

## Relation b/w $\text{am}(\lambda)$ & $\text{gm}(\lambda)$ (contd.)

- $\because P$  is invertible  $\Rightarrow$  Let  $P^{-1} = \begin{bmatrix} X \\ Y \end{bmatrix}$ , where  $X \in \mathbb{R}^{k \times n}$  and  $Y \in \mathbb{R}^{(n-k) \times n}$ .
- Then,

$$\begin{bmatrix} I_k & O \\ O & I_{n-k} \end{bmatrix} = P^{-1}P = \begin{bmatrix} X \\ Y \end{bmatrix} [U \ V] = \begin{bmatrix} XU & XV \\ YU & YV \end{bmatrix}$$



## Relation b/w $\text{am}(\lambda)$ & $\text{gm}(\lambda)$ (contd.)

- $\because \mathbf{P}$  is invertible  $\Rightarrow$  Let  $\mathbf{P}^{-1} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$ , where  $\mathbf{X} \in \mathbb{R}^{k \times n}$  and  $\mathbf{Y} \in \mathbb{R}^{(n-k) \times n}$ .

- Then,

$$\begin{bmatrix} \mathbf{I}_k & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_{n-k} \end{bmatrix} = \mathbf{P}^{-1}\mathbf{P} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} [\mathbf{U} \ \mathbf{V}] = \begin{bmatrix} \mathbf{XU} & \mathbf{XV} \\ \mathbf{YU} & \mathbf{YV} \end{bmatrix}$$

- Note that  $\mathbf{AU} = \mathbf{A}[\mathbf{v}_1 \cdots \mathbf{v}_k] = [\mathbf{Av}_1 \cdots \mathbf{Av}_k] = [\lambda \mathbf{v}_1 \cdots \lambda \mathbf{v}_k] = \lambda \mathbf{U}$ .

## Relation b/w $\text{am}(\lambda)$ & $\text{gm}(\lambda)$ (contd.)

$$P^{-1}AP = \begin{bmatrix} X \\ Y \end{bmatrix} A \begin{bmatrix} U & V \end{bmatrix} = \begin{bmatrix} XAU & XAV \\ YAU & YAV \end{bmatrix}$$

## Relation b/w $\text{am}(\lambda)$ & $\text{gm}(\lambda)$ (contd.)

$$\begin{aligned} P^{-1}AP &= \begin{bmatrix} X \\ Y \end{bmatrix} A[U \ V] = \begin{bmatrix} XAU & XAV \\ YAU & YAV \end{bmatrix} \\ &= \begin{bmatrix} \lambda XU & XAV \\ \lambda YU & YAV \end{bmatrix} = \begin{bmatrix} \lambda I_k & XAV \\ \mathbf{0} & YAV \end{bmatrix} \end{aligned}$$

## Relation b/w $\text{am}(\lambda)$ & $\text{gm}(\lambda)$ (contd.)

$$\begin{aligned} P^{-1}AP &= \begin{bmatrix} X \\ Y \end{bmatrix} A \begin{bmatrix} U & V \end{bmatrix} = \begin{bmatrix} XAU & XAV \\ YAU & YAV \end{bmatrix} \\ &= \begin{bmatrix} \lambda XU & XAV \\ \lambda YU & YAV \end{bmatrix} = \begin{bmatrix} \lambda I_k & XAV \\ \mathbf{0} & YAV \end{bmatrix} \end{aligned}$$

$$\det(P^{-1}AP - zI) = \det \begin{bmatrix} \lambda I_k - zI_k & XAV \\ \mathbf{0} & YAV - zI_{n-k} \end{bmatrix}$$

## Relation b/w $\text{am}(\lambda)$ & $\text{gm}(\lambda)$ (contd.)

$$\begin{aligned} P^{-1}AP &= \begin{bmatrix} X \\ Y \end{bmatrix} A[U \ V] = \begin{bmatrix} XAU & XAV \\ YAU & YAV \end{bmatrix} \\ &= \begin{bmatrix} \lambda XU & XAV \\ \lambda YU & YAV \end{bmatrix} = \begin{bmatrix} \lambda I_k & XAV \\ \mathbf{0} & YAV \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \det(P^{-1}AP - zI) &= \det \begin{bmatrix} \lambda I_k - zI_k & XAV \\ \mathbf{0} & YAV - zI_{n-k} \end{bmatrix} \\ &= (\lambda - z)^k \det(YAV - zI_{n-k}). \end{aligned}$$

• **Note:**  $\det(A - zI) = \det(P^{-1}AP - zI)$ .

## Remark

$$\begin{aligned}P^{-1}AP - zI &= P^{-1}AP - zP^{-1}P \\&= P^{-1}AP - P^{-1}(zI)P \\&= P^{-1}(A - zI)P.\end{aligned}$$

## Remark

$$\begin{aligned} P^{-1}AP - zI &= P^{-1}AP - zP^{-1}P \\ &= P^{-1}AP - P^{-1}(zI)P \\ &= P^{-1}(A - zI)P. \end{aligned}$$

Therefore,

$$\det(\mathbf{A} - z\mathbf{I}) = \det(\mathbf{P}^{-1}\mathbf{A}\mathbf{P} - z\mathbf{I}).$$

# The Case of the Identity Matrix

## Exercise: The Case of the Identity Matrix

For  $I_n \in \mathbb{R}^{n \times n}$ ,

- what is  $p_I(\lambda)$ ?
- What are its eigenvalues and the associated eigenvectors?
- What are the eigenspaces?



## Useful Properties (1/4)

- $\mathbf{A}$  and  $\mathbf{A}^\top$  possess the same eigenvalues

## Useful Properties (1/4)

- $\mathbf{A}$  and  $\mathbf{A}^\top$  possess the same eigenvalues but not necessarily the same eigenvectors.

## Useful Properties (1/4)

- $\mathbf{A}$  and  $\mathbf{A}^\top$  possess the same eigenvalues but not necessarily the same eigenvectors.
- The eigenspace  $E_\lambda$  is  $\ker(\mathbf{A} - \lambda \mathbf{I})$ .

## Useful Properties (1/4)

- $\mathbf{A}$  and  $\mathbf{A}^\top$  possess the same eigenvalues but not necessarily the same eigenvectors.
- The eigenspace  $E_\lambda$  is  $\ker(\mathbf{A} - \lambda \mathbf{I})$ .

$$\begin{aligned}\mathbf{Ax} = \lambda \mathbf{x} &\Leftrightarrow \mathbf{Ax} - \lambda \mathbf{x} = \mathbf{0} \\ &\Leftrightarrow (\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \mathbf{x} \in \ker(\mathbf{A} - \lambda \mathbf{I}).\end{aligned}$$

## Useful Properties (1/4)

- $\mathbf{A}$  and  $\mathbf{A}^\top$  possess the same eigenvalues but not necessarily the same eigenvectors.
- The eigenspace  $E_\lambda$  is  $\ker(\mathbf{A} - \lambda \mathbf{I})$ .

$$\begin{aligned}\mathbf{Ax} = \lambda \mathbf{x} &\Leftrightarrow \mathbf{Ax} - \lambda \mathbf{x} = \mathbf{0} \\ &\Leftrightarrow (\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \mathbf{x} \in \ker(\mathbf{A} - \lambda \mathbf{I}).\end{aligned}$$

- Symmetric, positive definite matrices always have positive, real eigenvalues.

## Useful Properties (1/4)

- $\mathbf{A}$  and  $\mathbf{A}^\top$  possess the same eigenvalues but not necessarily the same eigenvectors.
- The eigenspace  $E_\lambda$  is  $\ker(\mathbf{A} - \lambda \mathbf{I})$ .

$$\begin{aligned}\mathbf{Ax} = \lambda \mathbf{x} &\Leftrightarrow \mathbf{Ax} - \lambda \mathbf{x} = \mathbf{0} \\ &\Leftrightarrow (\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \mathbf{x} \in \ker(\mathbf{A} - \lambda \mathbf{I}).\end{aligned}$$

- Symmetric, positive definite matrices always have positive, real eigenvalues.
  - $\mathbf{x}^\top \mathbf{Ax} = \mathbf{x}^\top \lambda \mathbf{x}$

## Useful Properties (1/4)

- $\mathbf{A}$  and  $\mathbf{A}^\top$  possess the same eigenvalues but not necessarily the same eigenvectors.
- The eigenspace  $E_\lambda$  is  $\ker(\mathbf{A} - \lambda \mathbf{I})$ .

$$\begin{aligned}\mathbf{Ax} = \lambda \mathbf{x} &\Leftrightarrow \mathbf{Ax} - \lambda \mathbf{x} = \mathbf{0} \\ &\Leftrightarrow (\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \mathbf{x} \in \ker(\mathbf{A} - \lambda \mathbf{I}).\end{aligned}$$

- Symmetric, positive definite matrices always have positive, real eigenvalues.
  - $\mathbf{x}^\top \mathbf{Ax} = \mathbf{x}^\top \lambda \mathbf{x} = \lambda \|\mathbf{x}\|^2 > 0$ .

## Useful Properties (2/4)

### Theorem (4.13)

The eigenvectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$  of a matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  with  $n$  distinct eigenvalues  $\lambda_1, \dots, \lambda_n$  are linearly independent.

### Theorem (4.14)

Given a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , we can always obtain a symmetric, positive **semidefinite** matrix  $\mathbf{S} \in \mathbb{R}^{n \times n}$  by defining

$$\mathbf{S} := \mathbf{A}^\top \mathbf{A}.$$

If  $\text{rank}(\mathbf{A}) = n$ , then  $\mathbf{S} := \mathbf{A}^\top \mathbf{A}$  is symmetric, positive definite.



## Useful Properties (3/4)

### Theorem

If  $\mathbf{A}$  is symmetric, then eigenvectors to different eigenvalues are orthogonal.

### Proof.

- Assume that  $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$  and  $\mathbf{A}\mathbf{w} = \mu\mathbf{w}$  for two eigenvectors  $\mathbf{v}, \mathbf{w} \in V$  corresponding to eigenvalues  $\lambda$  and  $\mu$  such that  $\lambda \neq \mu$ .
- $$\begin{aligned}\lambda\langle\mathbf{v}, \mathbf{w}\rangle &= \langle\lambda\mathbf{v}, \mathbf{w}\rangle = \langle\mathbf{A}\mathbf{v}, \mathbf{w}\rangle = (\mathbf{A}\mathbf{v})^\top \mathbf{w} = \mathbf{v}^\top \mathbf{A}^\top \mathbf{w} = \langle\mathbf{v}, \mathbf{A}^\top \mathbf{w}\rangle \\ &= \langle\mathbf{v}, \mathbf{A}\mathbf{w}\rangle = \langle\mathbf{v}, \mu\mathbf{w}\rangle = \mu\langle\mathbf{v}, \mathbf{w}\rangle.\end{aligned}$$

The equalities hold only if  $\langle\mathbf{v}, \mathbf{w}\rangle = 0$ .



## Useful Properties (4/4)

### Theorem (4.15; Spectral Theorem)

If  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is symmetric, then there exists an orthonormal basis, consisting of eigenvectors of  $\mathbf{A}$ , of the corresponding vector space  $V$ , and each eigenvalue is real.

## Useful Properties (4/4)

### Theorem (4.15; Spectral Theorem)

If  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is symmetric, then there exists an orthonormal basis, consisting of eigenvectors of  $\mathbf{A}$ , of the corresponding vector space  $V$ , and each eigenvalue is real.

### Example

Consider

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$$

## Useful Properties (4/4)

### Theorem (4.15; Spectral Theorem)

If  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is symmetric, then there exists an orthonormal basis, consisting of eigenvectors of  $\mathbf{A}$ , of the corresponding vector space  $V$ , and each eigenvalue is real.

### Example

Consider

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$$

Compute  $p_{\mathbf{A}}(\lambda) = -(\lambda - 1)^2(\lambda - 7)$

## Useful Properties (4/4)

### Theorem (4.15; Spectral Theorem)

If  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is symmetric, then there exists an orthonormal basis, consisting of eigenvectors of  $\mathbf{A}$ , of the corresponding vector space  $V$ , and each eigenvalue is real.

### Example

Consider

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$$

Compute  $p_{\mathbf{A}}(\lambda) = -(\lambda - 1)^2(\lambda - 7) \Rightarrow \lambda_1 = 1$  (repeated),  $\lambda_2 = 7$ .

- $E_1 = \text{span}(\mathbf{x}_1, \mathbf{x}_2)$ , where  $\mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ ,  $\mathbf{x}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ ,  
 $E_7 = \text{span}(\mathbf{x}_3)$ , where  $\mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ .

- $E_1 = \text{span}(\mathbf{x}_1, \mathbf{x}_2)$ , where  $\mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ ,  $\mathbf{x}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ ,

$$E_7 = \text{span}(\mathbf{x}_3), \text{ where } \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

- $\mathbf{x}_3 \perp \mathbf{x}_1, \mathbf{x}_2$  but  $\mathbf{x}_1 \not\perp \mathbf{x}_2$ .

- $E_1 = \text{span}(\mathbf{x}_1, \mathbf{x}_2)$ , where  $\mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ ,  $\mathbf{x}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ ,

$$E_7 = \text{span}(\mathbf{x}_3), \text{ where } \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

- $\mathbf{x}_3 \perp \mathbf{x}_1, \mathbf{x}_2$  but  $\mathbf{x}_1 \not\perp \mathbf{x}_2$ .
- **Note:** Any linear combination of  $\mathbf{x}_1$  and  $\mathbf{x}_2$  is also an eigenvector of  $\mathbf{A}$  w.r.t.  $\lambda_1$ .



- $E_1 = \text{span}(\mathbf{x}_1, \mathbf{x}_2)$ , where  $\mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ ,  $\mathbf{x}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ ,

$$E_7 = \text{span}(\mathbf{x}_3), \text{ where } \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

- $\mathbf{x}_3 \perp \mathbf{x}_1, \mathbf{x}_2$  but  $\mathbf{x}_1 \not\perp \mathbf{x}_2$ .
- **Note:** Any linear combination of  $\mathbf{x}_1$  and  $\mathbf{x}_2$  is also an eigenvector of  $\mathbf{A}$  w.r.t.  $\lambda_1$ .

$$\mathbf{A}(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) = \mathbf{A}\mathbf{x}_1\alpha + \mathbf{A}\mathbf{x}_2\beta = \lambda(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2).$$

- Use Gram-Schmidt algorithm to construct an orthogonal basis for  $\text{span}(\mathbf{x}_1, \mathbf{x}_2)$ !

Take  $\mathbf{u}_1 = \mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ ,

$$\text{Take } \mathbf{u}_1 = \mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},$$

$$\mathbf{u}_2 = \mathbf{x}_2 - \frac{\mathbf{u}_1 \mathbf{u}_1^\top}{\|\mathbf{u}_1\|^2} \mathbf{x}_2 =$$

$$\text{Take } \mathbf{u}_1 = \mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},$$

$$\begin{aligned} \mathbf{u}_2 = \mathbf{x}_2 - \frac{\mathbf{u}_1 \mathbf{u}_1^\top}{\|\mathbf{u}_1\|^2} \mathbf{x}_2 &= \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \\ &= -\frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}. \end{aligned}$$

## Why is each eigenvalue real?

- **Note:** Define  $\mathbf{x}^* = [a_1 - b_1i, a_2 - b_2i]^\top$  for  $\mathbf{x} = [a_1 + b_1i, a_2 + b_2i]$ . And, define  $\mathbf{A}^* = \{\overline{\mathbf{A}}_{ji}\}_{(i,j) \in N \times M}$  for  $\mathbf{A} \in \mathbb{C}^{m \times n}$ .
- **Note:**  $\mathbf{x}^* \mathbf{x} \geq 0$ .

Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$  and  $\mathbf{A}$  is symmetric:

- Assume that  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$  with  $\mathbf{x} \neq \mathbf{0}$ .

## Why is each eigenvalue real?

- **Note:** Define  $\mathbf{x}^* = [a_1 - b_1i, a_2 - b_2i]^\top$  for  $\mathbf{x} = [a_1 + b_1i, a_2 + b_2i]$ . And, define  $\mathbf{A}^* = \{\overline{\mathbf{A}}_{ji}\}_{(i,j) \in N \times M}$  for  $\mathbf{A} \in \mathbb{C}^{m \times n}$ .
- **Note:**  $\mathbf{x}^* \mathbf{x} \geq 0$ .

Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$  and  $\mathbf{A}$  is symmetric:

- Assume that  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$  with  $\mathbf{x} \neq \mathbf{0}$ .

Note that  $\lambda$  could be a complex number and  $\mathbf{x}$  is possibly complex.

## Why is each eigenvalue real?

- **Note:** Define  $\mathbf{x}^* = [a_1 - b_1i, a_2 - b_2i]^\top$  for  $\mathbf{x} = [a_1 + b_1i, a_2 + b_2i]$ . And, define  $\mathbf{A}^* = \{\overline{\mathbf{A}}_{ji}\}_{(i,j) \in N \times M}$  for  $\mathbf{A} \in \mathbb{C}^{m \times n}$ .
- **Note:**  $\mathbf{x}^* \mathbf{x} \geq 0$ .

Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$  and  $\mathbf{A}$  is symmetric:

- Assume that  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$  with  $\mathbf{x} \neq \mathbf{0}$ .

Note that  $\lambda$  could be a complex number and  $\mathbf{x}$  is possibly complex.

- $\mathbf{x}^* \mathbf{A} \mathbf{x} = \mathbf{x}^* \lambda \mathbf{x} = \lambda \mathbf{x}^* \mathbf{x}$ .

## Why is each eigenvalue real?

- **Note:** Define  $\mathbf{x}^* = [a_1 - b_1i, a_2 - b_2i]^\top$  for  $\mathbf{x} = [a_1 + b_1i, a_2 + b_2i]$ . And, define  $\mathbf{A}^* = \{\overline{\mathbf{A}}_{ji}\}_{(i,j) \in N \times M}$  for  $\mathbf{A} \in \mathbb{C}^{m \times n}$ .
- **Note:**  $\mathbf{x}^* \mathbf{x} \geq 0$ .

Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$  and  $\mathbf{A}$  is symmetric:

- Assume that  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$  with  $\mathbf{x} \neq \mathbf{0}$ .

Note that  $\lambda$  could be a complex number and  $\mathbf{x}$  is possibly complex.

- $\mathbf{x}^* \mathbf{A} \mathbf{x} = \mathbf{x}^* \lambda \mathbf{x} = \lambda \mathbf{x}^* \mathbf{x}$ .
- Also,  $(\mathbf{A}\mathbf{x})^* \mathbf{x} = \mathbf{x}^* \mathbf{A}^* \mathbf{x} = \mathbf{x}^* \mathbf{A}^\top \mathbf{x} = \mathbf{x}^* \mathbf{A} \mathbf{x}$  ( $\because \mathbf{A}$  is a real & symmetric).



## Why is each eigenvalue real?

- **Note:** Define  $\mathbf{x}^* = [a_1 - b_1i, a_2 - b_2i]^\top$  for  $\mathbf{x} = [a_1 + b_1i, a_2 + b_2i]$ . And, define  $\mathbf{A}^* = \{\bar{\mathbf{A}}_{ji}\}_{(i,j) \in N \times M}$  for  $\mathbf{A} \in \mathbb{C}^{m \times n}$ .
- **Note:**  $\mathbf{x}^* \mathbf{x} \geq 0$ .

Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$  and  $\mathbf{A}$  is symmetric:

- Assume that  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$  with  $\mathbf{x} \neq \mathbf{0}$ .

Note that  $\lambda$  could be a complex number and  $\mathbf{x}$  is possibly complex.

- $\mathbf{x}^* \mathbf{A} \mathbf{x} = \mathbf{x}^* \lambda \mathbf{x} = \lambda \mathbf{x}^* \mathbf{x}$ .
- Also,  $(\mathbf{A}\mathbf{x})^* \mathbf{x} = \mathbf{x}^* \mathbf{A}^* \mathbf{x} = \mathbf{x}^* \mathbf{A}^\top \mathbf{x} = \mathbf{x}^* \mathbf{A} \mathbf{x}$  ( $\because \mathbf{A}$  is a real & symmetric).
  - Note:  $(\mathbf{A}\mathbf{x})^* = (\lambda\mathbf{x})^* = \bar{\lambda}\mathbf{x}^*$ , so  $(\mathbf{A}\mathbf{x})^* \mathbf{x} = \bar{\lambda} \mathbf{x}^* \mathbf{x}$ .

## Why is each eigenvalue real?

- **Note:** Define  $\mathbf{x}^* = [a_1 - b_1i, a_2 - b_2i]^\top$  for  $\mathbf{x} = [a_1 + b_1i, a_2 + b_2i]$ . And, define  $\mathbf{A}^* = \{\bar{\mathbf{A}}_{ji}\}_{(i,j) \in N \times M}$  for  $\mathbf{A} \in \mathbb{C}^{m \times n}$ .
- **Note:**  $\mathbf{x}^* \mathbf{x} \geq 0$ .

Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$  and  $\mathbf{A}$  is symmetric:

- Assume that  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$  with  $\mathbf{x} \neq \mathbf{0}$ .

Note that  $\lambda$  could be a complex number and  $\mathbf{x}$  is possibly complex.

- $\mathbf{x}^* \mathbf{A} \mathbf{x} = \mathbf{x}^* \lambda \mathbf{x} = \lambda \mathbf{x}^* \mathbf{x}$ .
- Also,  $(\mathbf{A}\mathbf{x})^* \mathbf{x} = \mathbf{x}^* \mathbf{A}^* \mathbf{x} = \mathbf{x}^* \mathbf{A}^\top \mathbf{x} = \mathbf{x}^* \mathbf{A} \mathbf{x}$  ( $\because \mathbf{A}$  is a real & symmetric).
  - Note:  $(\mathbf{A}\mathbf{x})^* = (\lambda\mathbf{x})^* = \bar{\lambda}\mathbf{x}^*$ , so  $(\mathbf{A}\mathbf{x})^* \mathbf{x} = \bar{\lambda} \mathbf{x}^* \mathbf{x}$ .
- Thus,  $(\lambda - \bar{\lambda})\mathbf{x}^* \mathbf{x} = 0$ . We have  $\lambda = \bar{\lambda}$  since  $\mathbf{x}^* \mathbf{x} > 0$ .  $\Rightarrow \lambda \in \mathbb{R}$ .

## A Practical Example [Page et al. 1999]

- Google uses the eigenvector corresponding to the maximal eigenvalue of a matrix  $\mathbf{A}$  to determine the rank of a page for search.
  - The **PageRank** algorithm was developed at Stanford University by Larry Page and Sergey Brin in 1996.

## A Practical Example [Page et al. 1999]

- Google uses the eigenvector corresponding to the maximal eigenvalue of a matrix  $\mathbf{A}$  to determine the rank of a page for search.
  - The **PageRank** algorithm was developed at Stanford University by Larry Page and Sergey Brin in 1996.
- Websites are represented as a huge directed graph (pages: vertices; links: edges).
- Compute the weight (importance)  $x_i \geq 0$  for a website  $a_i$  and get  $\mathbf{x}$ .
  - The number of pages pointing to  $a_i$ .
- A transition matrix  $\mathbf{A}$  (prob.): modeling the navigation behavior of a user.
- **Goal:**  $\mathbf{x}, \mathbf{Ax}, \mathbf{A}^2\mathbf{x}, \dots, \mathbf{x}^*$

## A Practical Example [Page et al. 1999]

- Google uses the eigenvector corresponding to the maximal eigenvalue of a matrix  $\mathbf{A}$  to determine the rank of a page for search.
  - The **PageRank** algorithm was developed at Stanford University by Larry Page and Sergey Brin in 1996.
- Websites are represented as a huge directed graph (pages: vertices; links: edges).
- Compute the weight (importance)  $x_i \geq 0$  for a website  $a_i$  and get  $\mathbf{x}$ .
  - The number of pages pointing to  $a_i$ .
- A transition matrix  $\mathbf{A}$  (prob.): modeling the navigation behavior of a user.
- **Goal:**  $\mathbf{x}, \mathbf{Ax}, \mathbf{A}^2\mathbf{x}, \dots, \mathbf{x}^* \Rightarrow \mathbf{Ax}^* = \mathbf{x}^* \Rightarrow$  Turning to probabilities (normalization).

# Outline

- 1 Eigenvalues & Eigenvectors
- 2 Cholesky Decomposition
- 3 Eigendecomposition & Diagonalization

# Cholesky Decomposition

## Cholesky Decomposition

A symmetric, positive definite matrix  $\mathbf{A}$  can be factorized into a product  $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$ , where  $\mathbf{L}$  is a lower-triangular matrix with positive diagonal elements.

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} \text{red triangle} \end{bmatrix} \begin{bmatrix} \text{green triangle} \end{bmatrix}$$

## Example of Cholesky Factorization

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \mathbf{L}\mathbf{L}^T = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}.$$

We have

$$\mathbf{A} = \begin{bmatrix} l_{11}^2 & l_{21}l_{11} & l_{31}l_{11} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{31}l_{21} + l_{32}l_{22} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Finally, solve  $l_{11}, \dots, l_{33}$ .



## Example Steps for Cholesky Factorization

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \ell_{11}^2 & \ell_{21}\ell_{11} & \ell_{31}\ell_{11} \\ \ell_{21}\ell_{11} & \ell_{21}^2 + \ell_{22}^2 & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} \\ \ell_{31}\ell_{11} & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} & \ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2 \end{bmatrix}$$

- $\ell_{11} = \sqrt{a_{11}},$

## Example Steps for Cholesky Factorization

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \ell_{11}^2 & \ell_{21}\ell_{11} & \ell_{31}\ell_{11} \\ \ell_{21}\ell_{11} & \ell_{21}^2 + \ell_{22}^2 & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} \\ \ell_{31}\ell_{11} & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} & \ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2 \end{bmatrix}$$

- $\ell_{11} = \sqrt{a_{11}}, \quad \ell_{21} = \frac{a_{21}}{\ell_{11}},$

## Example Steps for Cholesky Factorization

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \ell_{11}^2 & \ell_{21}\ell_{11} & \ell_{31}\ell_{11} \\ \ell_{21}\ell_{11} & \ell_{21}^2 + \ell_{22}^2 & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} \\ \ell_{31}\ell_{11} & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} & \ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2 \end{bmatrix}$$

- $\ell_{11} = \sqrt{a_{11}}, \quad \ell_{21} = \frac{a_{21}}{\ell_{11}}, \quad \ell_{22} = \sqrt{a_{22} - \ell_{21}^2},$

## Example Steps for Cholesky Factorization

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \ell_{11}^2 & \ell_{21}\ell_{11} & \ell_{31}\ell_{11} \\ \ell_{21}\ell_{11} & \ell_{21}^2 + \ell_{22}^2 & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} \\ \ell_{31}\ell_{11} & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} & \ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2 \end{bmatrix}$$

- $\ell_{11} = \sqrt{a_{11}}, \quad \ell_{21} = \frac{a_{21}}{\ell_{11}}, \quad \ell_{22} = \sqrt{a_{22} - \ell_{21}^2}, \quad \ell_{31} = \frac{a_{31}}{\ell_{11}},$

$$\ell_{32} = \frac{a_{32} - \ell_{31}\ell_{21}}{\ell_{22}},$$

## Example Steps for Cholesky Factorization

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \ell_{11}^2 & \ell_{21}\ell_{11} & \ell_{31}\ell_{11} \\ \ell_{21}\ell_{11} & \ell_{21}^2 + \ell_{22}^2 & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} \\ \ell_{31}\ell_{11} & \ell_{31}\ell_{21} + \ell_{32}\ell_{22} & \ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2 \end{bmatrix}$$

- $\ell_{11} = \sqrt{a_{11}}, \quad \ell_{21} = \frac{a_{21}}{\ell_{11}}, \quad \ell_{22} = \sqrt{a_{22} - \ell_{21}^2}, \quad \ell_{31} = \frac{a_{31}}{\ell_{11}},$

$$\ell_{32} = \frac{a_{32} - \ell_{31}\ell_{21}}{\ell_{22}}, \quad \ell_{33} = \sqrt{a_{33} - \ell_{31}^2 - \ell_{32}^2}.$$

# Motivations of Using Cholesky Decomposition

- Symmetric positive definite matrices require frequent manipulation.
  - E.g., Covariance matrix of a multivariate Gaussian variable.
  - The Cholesky factorization of the covariance matrix allows us to **generate samples from a Gaussian distribution**.
- Computing gradients in deep stochastic models such as variational auto-encoder (VAE).

# Motivations of Using Cholesky Decomposition

- Symmetric positive definite matrices require frequent manipulation.
  - E.g., Covariance matrix of a multivariate Gaussian variable.
  - The Cholesky factorization of the covariance matrix allows us to **generate samples from a Gaussian distribution**.
- Computing gradients in deep stochastic models such as variational auto-encoder (VAE).
- Compute determinants efficiently.
  - $\det(\mathbf{A}) = \det(\mathbf{L}) \det(\mathbf{L}^\top) = \det(\mathbf{L})^2$ .
  - Note:  $\det(\mathbf{L})$  can be computed efficiently ( $\because$  triangular).

# Why Cholesky enables Gaussian sampling

**Goal:** Generate  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,  $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$  is symmetric positive definite (SPD).

❶ **Start with i.i.d. standard normals.**

Draw  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$  (components independent  $N(0, 1)$ ).

❷ **Target distribution.**

We want  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  with given mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ .

❸ **Cholesky factorization of  $\boldsymbol{\Sigma}$ .**

Since  $\boldsymbol{\Sigma}$  is SPD, there exists a lower-triangular  $\mathbf{L}$  such that  $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^\top$ .

❹ **Build the desired correlations.**

Set  $\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\mathbf{z}$ . Then  $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu} + \mathbf{L}\mathbb{E}[\mathbf{z}] = \boldsymbol{\mu}$ ,

$$\text{Cov}(\mathbf{x}) = \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top] = \mathbb{E}[\mathbf{L}\mathbf{z}\mathbf{z}^\top \mathbf{L}^\top] = \mathbf{L}\mathbf{I}_n\mathbf{L}^\top = \mathbf{L}\mathbf{L}^\top = \boldsymbol{\Sigma}.$$



# Outline

- 1 Eigenvalues & Eigenvectors
- 2 Cholesky Decomposition
- 3 Eigendecomposition & Diagonalization**

# Motivation of Diagonalization

- Diagonalization is an important application of basis change and eigenvalues.

# Motivation of Diagonalization

- Diagonalization is an important application of basis change and eigenvalues.
- Diagonalization allow fast computation of determinants, powers and inverses of matrices.

# Motivation of Diagonalization

- Diagonalization is an important application of basis change and eigenvalues.
- Diagonalization allow fast computation of determinants, powers and inverses of matrices. A **diagonal matrix** is like

$$\mathbf{D} = \begin{bmatrix} c_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_n \end{bmatrix}.$$

# Motivation of Diagonalization

- Diagonalization is an important application of basis change and eigenvalues.
- Diagonalization allow fast computation of determinants, powers and inverses of matrices. A **diagonal matrix** is like

$$\mathbf{D} = \begin{bmatrix} c_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_n \end{bmatrix}.$$

- **Question:** What are the determinant, cubic, and inverse of  $\mathbf{D}$ ?

# Similarity

## Similarity

Two matrices  $\mathbf{A}$  and  $\mathbf{B} \in \mathbb{R}^{n \times n}$  are **similar** if there exists an invertible matrix  $\mathbf{S} \in \mathbb{R}^{n \times n}$  such that  $\mathbf{A} = \mathbf{S}^{-1}\mathbf{B}\mathbf{S}$ .

# Similarity

## Similarity

Two matrices  $\mathbf{A}$  and  $\mathbf{B} \in \mathbb{R}^{n \times n}$  are **similar** if there exists an invertible matrix  $\mathbf{S} \in \mathbb{R}^{n \times n}$  such that  $\mathbf{A} = \mathbf{S}^{-1}\mathbf{B}\mathbf{S}$ .

## Diagonalizable

A matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is **diagonalizable** if it is *similar* to a *diagonal* matrix..

# Similarity

## Similarity

Two matrices  $\mathbf{A}$  and  $\mathbf{B} \in \mathbb{R}^{n \times n}$  are **similar** if there exists an invertible matrix  $\mathbf{S} \in \mathbb{R}^{n \times n}$  such that  $\mathbf{A} = \mathbf{S}^{-1}\mathbf{B}\mathbf{S}$ .

## Diagonalizable

A matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is **diagonalizable** if it is *similar* to a *diagonal* matrix..

- $\exists \mathbf{D} \in \mathbb{R}^{n \times n}$ , such that  $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ .



# Eigenvectors & Diagonalization

- Let  $\mathbf{A} \in \mathbb{R}^{n \times n}$  and  $\lambda_1, \dots, \lambda_n$  be a set of scalars.
- Let  $\mathbf{p}_1, \dots, \mathbf{p}_n$  be a set of vectors in  $\mathbb{R}^n$ .
- Let  $\mathbf{D} \in \mathbb{R}^{n \times n}$  be a diagonal matrix with diagonal entries  $\lambda_1, \dots, \lambda_n$ .

We can show that

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}.$$

if and only if  $\lambda_1, \dots, \lambda_n$  are the eigenvalues of  $\mathbf{A}$  and  $\mathbf{p}_1, \dots, \mathbf{p}_n$  are the corresponding eigenvectors of  $\mathbf{A}$ .

# Proof Sketch of the Claim

We can see that

$$\mathbf{A}\mathbf{P} = \mathbf{A}[\mathbf{p}_1, \dots, \mathbf{p}_n] = [\mathbf{A}\mathbf{p}_1, \dots, \mathbf{A}\mathbf{p}_n],$$

and

# Proof Sketch of the Claim

We can see that

$$\mathbf{A}\mathbf{P} = \mathbf{A}[\mathbf{p}_1, \dots, \mathbf{p}_n] = [\mathbf{A}\mathbf{p}_1, \dots, \mathbf{A}\mathbf{p}_n],$$

and

$$\mathbf{P}\mathbf{D} = [\mathbf{p}_1, \dots, \mathbf{p}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

# Proof Sketch of the Claim

We can see that

$$\mathbf{A}\mathbf{P} = \mathbf{A}[\mathbf{p}_1, \dots, \mathbf{p}_n] = [\mathbf{A}\mathbf{p}_1, \dots, \mathbf{A}\mathbf{p}_n],$$

and

$$\mathbf{P}\mathbf{D} = [\mathbf{p}_1, \dots, \mathbf{p}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = [\lambda_1 \mathbf{p}_1, \dots, \lambda_n \mathbf{p}_n].$$

# Proof Sketch of the Claim

We can see that

$$\mathbf{A}\mathbf{P} = \mathbf{A}[\mathbf{p}_1, \dots, \mathbf{p}_n] = [\mathbf{A}\mathbf{p}_1, \dots, \mathbf{A}\mathbf{p}_n],$$

and

$$\mathbf{P}\mathbf{D} = [\mathbf{p}_1, \dots, \mathbf{p}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = [\lambda_1 \mathbf{p}_1, \dots, \lambda_n \mathbf{p}_n].$$

Thus,

$$\mathbf{A}\mathbf{p}_1 = \lambda_1 \mathbf{p}_1$$

$$\vdots$$

$$\mathbf{A}\mathbf{p}_n = \lambda_n \mathbf{p}_n$$

Therefore, the columns of  $\mathbf{P}$  are eigenvectors of  $\mathbf{A}$ .

# Are $\mathbf{p}_1, \dots, \mathbf{p}_n$ linearly independent?

- If  $n = 1$ , we only get  $\mathbf{p}_1 (\neq \mathbf{0}) \Rightarrow$  linearly independent.

# Are $\mathbf{p}_1, \dots, \mathbf{p}_n$ linearly independent?

- If  $n = 1$ , we only get  $\mathbf{p}_1 (\neq \mathbf{0}) \Rightarrow$  linearly independent.
- Assume that  $\mathbf{p}_1, \dots, \mathbf{p}_{n-1}$  are linearly independent.

# Are $\mathbf{p}_1, \dots, \mathbf{p}_n$ linearly independent?

- If  $n = 1$ , we only get  $\mathbf{p}_1 (\neq \mathbf{0}) \Rightarrow$  linearly independent.
- Assume that  $\mathbf{p}_1, \dots, \mathbf{p}_{n-1}$  are linearly independent.
- Suppose that  $\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n = \mathbf{0}$ . (\*)



# Are $\mathbf{p}_1, \dots, \mathbf{p}_n$ linearly independent?

- If  $n = 1$ , we only get  $\mathbf{p}_1 (\neq \mathbf{0}) \Rightarrow$  linearly independent.
- Assume that  $\mathbf{p}_1, \dots, \mathbf{p}_{n-1}$  are linearly independent.
- Suppose that  $\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n = \mathbf{0}$ . (\*)

$$\Rightarrow \mathbf{A}(\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n) = \mathbf{A}\mathbf{0} = \mathbf{0}.$$

# Are $\mathbf{p}_1, \dots, \mathbf{p}_n$ linearly independent?

- If  $n = 1$ , we only get  $\mathbf{p}_1 (\neq \mathbf{0}) \Rightarrow$  linearly independent.

- Assume that  $\mathbf{p}_1, \dots, \mathbf{p}_{n-1}$  are linearly independent.

- Suppose that  $\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n = \mathbf{0}$ . (\*)

$$\Rightarrow \mathbf{A}(\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n) = \mathbf{A}\mathbf{0} = \mathbf{0}.$$

$$\Rightarrow \alpha \lambda_1 \mathbf{p}_1 + \alpha_2 \lambda_2 \mathbf{p}_2 + \dots + \alpha_n \lambda_n \mathbf{p}_n = \mathbf{0}. (**)$$

$$\Rightarrow (**)-\lambda_n \cdot (*):$$

$$\alpha_1(\lambda_1 - \lambda_n) \mathbf{p}_1 + \alpha_2(\lambda_2 - \lambda_n) \mathbf{p}_2 + \dots + \alpha_{n-1}(\lambda_{n-1} - \lambda_n) \mathbf{p}_{n-1} = \mathbf{0}.$$

# Are $\mathbf{p}_1, \dots, \mathbf{p}_n$ linearly independent?

- If  $n = 1$ , we only get  $\mathbf{p}_1 (\neq \mathbf{0}) \Rightarrow$  linearly independent.

- Assume that  $\mathbf{p}_1, \dots, \mathbf{p}_{n-1}$  are linearly independent.

- Suppose that  $\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n = \mathbf{0}$ . (\*)

$$\Rightarrow \mathbf{A}(\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n) = \mathbf{A}\mathbf{0} = \mathbf{0}.$$

$$\Rightarrow \alpha \lambda_1 \mathbf{p}_1 + \alpha_2 \lambda_2 \mathbf{p}_2 + \dots + \alpha_n \lambda_n \mathbf{p}_n = \mathbf{0}. (**)$$

$$\Rightarrow (**)-\lambda_n \cdot (*):$$

$$\alpha_1(\lambda_1 - \lambda_n) \mathbf{p}_1 + \alpha_2(\lambda_2 - \lambda_n) \mathbf{p}_2 + \dots + \alpha_{n-1}(\lambda_{n-1} - \lambda_n) \mathbf{p}_{n-1} = \mathbf{0}.$$

$$\Rightarrow \alpha_i(\lambda_i - \lambda_n) = 0 \text{ for each } i = 1, 2, \dots, n-1.$$

$$\Rightarrow \alpha_i = 0 \text{ for each } i = 1, 2, \dots, n-1.$$

# Are $\mathbf{p}_1, \dots, \mathbf{p}_n$ linearly independent?

- If  $n = 1$ , we only get  $\mathbf{p}_1 (\neq \mathbf{0}) \Rightarrow$  linearly independent.

- Assume that  $\mathbf{p}_1, \dots, \mathbf{p}_{n-1}$  are linearly independent.

- Suppose that  $\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n = \mathbf{0}$ . (\*)

$$\Rightarrow \mathbf{A}(\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n) = \mathbf{A}\mathbf{0} = \mathbf{0}.$$

$$\Rightarrow \alpha \lambda_1 \mathbf{p}_1 + \alpha_2 \lambda_2 \mathbf{p}_2 + \dots + \alpha_n \lambda_n \mathbf{p}_n = \mathbf{0}. (**)$$

$$\Rightarrow (**) - \lambda_n \cdot (*):$$

$$\alpha_1(\lambda_1 - \lambda_n) \mathbf{p}_1 + \alpha_2(\lambda_2 - \lambda_n) \mathbf{p}_2 + \dots + \alpha_{n-1}(\lambda_{n-1} - \lambda_n) \mathbf{p}_{n-1} = \mathbf{0}.$$

$$\Rightarrow \alpha_i(\lambda_i - \lambda_n) = 0 \text{ for each } i = 1, 2, \dots, n-1.$$

$$\Rightarrow \alpha_i = 0 \text{ for each } i = 1, 2, \dots, n-1. \Rightarrow \alpha_n = 0.$$

# Are $\mathbf{p}_1, \dots, \mathbf{p}_n$ linearly independent?

- If  $n = 1$ , we only get  $\mathbf{p}_1 (\neq \mathbf{0}) \Rightarrow$  linearly independent.

- Assume that  $\mathbf{p}_1, \dots, \mathbf{p}_{n-1}$  are linearly independent.

- Suppose that  $\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n = \mathbf{0}$ . (\*)

$$\Rightarrow \mathbf{A}(\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \dots + \alpha_n \mathbf{p}_n) = \mathbf{A}\mathbf{0} = \mathbf{0}.$$

$$\Rightarrow \alpha \lambda_1 \mathbf{p}_1 + \alpha_2 \lambda_2 \mathbf{p}_2 + \dots + \alpha_n \lambda_n \mathbf{p}_n = \mathbf{0}. (**)$$

$$\Rightarrow (**)-\lambda_n \cdot (*):$$

$$\alpha_1(\lambda_1 - \lambda_n) \mathbf{p}_1 + \alpha_2(\lambda_2 - \lambda_n) \mathbf{p}_2 + \dots + \alpha_{n-1}(\lambda_{n-1} - \lambda_n) \mathbf{p}_{n-1} = \mathbf{0}.$$

$$\Rightarrow \alpha_i(\lambda_i - \lambda_n) = 0 \text{ for each } i = 1, 2, \dots, n-1.$$

$$\Rightarrow \alpha_i = 0 \text{ for each } i = 1, 2, \dots, n-1. \Rightarrow \alpha_n = 0.$$

- $\therefore \alpha_1 = \alpha_2 = \dots = \alpha_n = 0$ .

# Eigendecomposition (Diagonalization)

## Theorem [Eigendecomposition (Diagonalization)]

A square matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  can be factored into

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1},$$

where  $\mathbf{P} \in \mathbb{R}^{n \times n}$  and  $\mathbf{D}$  is a diagonal matrix whose diagonal entries are the eigenvalues of  $\mathbf{A}$

if and only if

the eigenvectors of  $\mathbf{A}$  form a basis of  $\mathbb{R}^n$ .

# Put it concisely

## Theorem

For a square matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , the following statements are equivalent:

- $\mathbf{A}$  is diagonalizable.
- $\mathbf{A}$  has  $n$  linearly independent eigenvectors.

## Remark

The spectral theorem tells us that:

*We can find an orthonormal basis of the corresponding vector space consisting of eigenvectors of a symmetric matrix  $\mathbf{S} \in \mathbb{R}^{n \times n}$ .*



## Remark

The spectral theorem tells us that:

*We can find an orthonormal basis of the corresponding vector space consisting of eigenvectors of a symmetric matrix  $\mathbf{S} \in \mathbb{R}^{n \times n}$ .*

## Theorem

A symmetric matrix  $\mathbf{S} \in \mathbb{R}^{n \times n}$  can be always diagonalized.

## Example

Compute the eigendecomposition of  $\mathbf{A} = \frac{1}{2} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ .

- 1 Compute the eigenvalues and eigenvectors.

## Example

Compute the eigendecomposition of  $\mathbf{A} = \frac{1}{2} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ .

- 1 Compute the eigenvalues and eigenvectors.

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \det \left( \begin{bmatrix} \frac{5}{2} - \lambda & -1 \\ -1 & \frac{5}{2} - \lambda \end{bmatrix} \right) =$$

## Example

Compute the eigendecomposition of  $\mathbf{A} = \frac{1}{2} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ .

- 1 Compute the eigenvalues and eigenvectors.

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \det \left( \begin{bmatrix} \frac{5}{2} - \lambda & -1 \\ -1 & \frac{5}{2} - \lambda \end{bmatrix} \right) = \left( \lambda - \frac{7}{2} \right) \left( \lambda - \frac{3}{2} \right).$$

Set  $\lambda_1 = \frac{7}{2}, \lambda_2 = \frac{3}{2}$ .

- 2 Solving  $\mathbf{A}\mathbf{p}_1 = \frac{7}{2}\mathbf{p}_1$  and  $\mathbf{A}\mathbf{p}_2 = \frac{3}{2}\mathbf{p}_2$ .

## Example

Compute the eigendecomposition of  $\mathbf{A} = \frac{1}{2} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ .

- 1 Compute the eigenvalues and eigenvectors.

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \det \left( \begin{bmatrix} \frac{5}{2} - \lambda & -1 \\ -1 & \frac{5}{2} - \lambda \end{bmatrix} \right) = \left( \lambda - \frac{7}{2} \right) \left( \lambda - \frac{3}{2} \right).$$

$$\text{Set } \lambda_1 = \frac{7}{2}, \lambda_2 = \frac{3}{2}.$$

- 2 Solving  $\mathbf{A}\mathbf{p}_1 = \frac{7}{2}\mathbf{p}_1$  and  $\mathbf{A}\mathbf{p}_2 = \frac{3}{2}\mathbf{p}_2$ .

$$\mathbf{p}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{p}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

## Example

Compute the eigendecomposition of  $\mathbf{A} = \frac{1}{2} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ .

$$\mathbf{p}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{p}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

③ Check for independency of  $\{\mathbf{p}_1, \mathbf{p}_2\}$ .  $\Rightarrow$

## Example

Compute the eigendecomposition of  $\mathbf{A} = \frac{1}{2} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ .

$$\mathbf{p}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{p}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

- ③ Check for independency of  $\{\mathbf{p}_1, \mathbf{p}_2\}$ .  $\implies \checkmark$
- ④ Construct  $\mathbf{P}$ :

## Example

Compute the eigendecomposition of  $\mathbf{A} = \frac{1}{2} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ .

$$\mathbf{p}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{p}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

③ Check for independency of  $\{\mathbf{p}_1, \mathbf{p}_2\}$ .  $\implies \checkmark$

④ Construct  $\mathbf{P}$ :  $\implies \mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2] = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ .

★ Note that  $\{\mathbf{p}_1, \mathbf{p}_2\}$  forms an **orthonormal basis**



## Example

Compute the eigendecomposition of  $\mathbf{A} = \frac{1}{2} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ .

$$\mathbf{p}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{p}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

③ Check for independency of  $\{\mathbf{p}_1, \mathbf{p}_2\}$ .  $\implies \checkmark$

④ Construct  $\mathbf{P}$ :  $\implies \mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2] = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ .

★ Note that  $\{\mathbf{p}_1, \mathbf{p}_2\}$  forms an **orthonormal basis**  $\mathbf{P}^{-1} = \mathbf{P}^\top$ .  
(Exercise)

## Example

Compute the eigendecomposition of  $\mathbf{A} = \frac{1}{2} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ .

$$\mathbf{p}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{p}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

③ Check for independency of  $\{\mathbf{p}_1, \mathbf{p}_2\}$ .  $\implies \checkmark$

④ Construct  $\mathbf{P}$ :  $\implies \mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2] = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ .

★ Note that  $\{\mathbf{p}_1, \mathbf{p}_2\}$  forms an **orthonormal basis**  $\mathbf{P}^{-1} = \mathbf{P}^\top$ .  
(Exercise)

Finally we obtain  $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ .

# Remark On the Efficiency

- $A^k = (PDP^{-1})^k$

## Remark On the Efficiency

- $A^k = (PDP^{-1})^k = (PDP^{-1})(PDP^{-1})\dots(PDP^{-1})$

## Remark On the Efficiency

- $\mathbf{A}^k = (\mathbf{PDP}^{-1})^k = (\mathbf{PDP}^{-1})(\mathbf{PDP}^{-1})\dots(\mathbf{PDP}^{-1}) = \mathbf{PD}^k\mathbf{P}^{-1}.$

## Remark On the Efficiency

- $\mathbf{A}^k = (\mathbf{PDP}^{-1})^k = (\mathbf{PDP}^{-1})(\mathbf{PDP}^{-1}) \dots (\mathbf{PDP}^{-1}) = \mathbf{PD}^k \mathbf{P}^{-1}$ .
- $\det(\mathbf{A}) = \det(\mathbf{PDP}^{-1})$

## Remark On the Efficiency

- $\mathbf{A}^k = (\mathbf{P}\mathbf{D}\mathbf{P}^{-1})^k = (\mathbf{P}\mathbf{D}\mathbf{P}^{-1})(\mathbf{P}\mathbf{D}\mathbf{P}^{-1})\dots(\mathbf{P}\mathbf{D}\mathbf{P}^{-1}) = \mathbf{P}\mathbf{D}^k\mathbf{P}^{-1}.$
- $\det(\mathbf{A}) = \det(\mathbf{P}\mathbf{D}\mathbf{P}^{-1}) = \det(\mathbf{P})\det(\mathbf{D})\det(\mathbf{P}^{-1})$

## Remark On the Efficiency

- $\mathbf{A}^k = (\mathbf{P}\mathbf{D}\mathbf{P}^{-1})^k = (\mathbf{P}\mathbf{D}\mathbf{P}^{-1})(\mathbf{P}\mathbf{D}\mathbf{P}^{-1})\dots(\mathbf{P}\mathbf{D}\mathbf{P}^{-1}) = \mathbf{P}\mathbf{D}^k\mathbf{P}^{-1}.$
- $\det(\mathbf{A}) = \det(\mathbf{P}\mathbf{D}\mathbf{P}^{-1}) = \det(\mathbf{P})\det(\mathbf{D})\det(\mathbf{P}^{-1}) = \det(\mathbf{D}) = \prod_i d_{ii}.$



# Another Application: Using Maclaurin Series

## Maclaurin Series

A Maclaurin Series expansion of  $f(x)$  is

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \cdots$$

# Another Application: Using Maclaurin Series

## Maclaurin Series

A Maclaurin Series expansion of  $f(x)$  is

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \cdots$$

For example, suppose  $f(x) = e^x$ , then

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

## Another Application: Using Maclaurin Series

### Maclaurin Series

A Maclaurin Series expansion of  $f(x)$  is

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \cdots$$

For example, suppose  $f(x) = e^x$ , then

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

Thus, for  $\mathbf{A} \in \mathbb{R}^{n \times n}$  such that  $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ , we may define

$$f(\mathbf{A}) = \mathbf{P}f(\mathbf{D})\mathbf{P}^{-1}.$$

## Another Application: Using Maclaurin Series

### Maclaurin Series

A Maclaurin Series expansion of  $f(x)$  is

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \cdots$$

For example, suppose  $f(\mathbf{A}) = e^{\mathbf{A}}$ , then

$$f(\mathbf{A}) = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \cdots$$

## Another Application: Using Maclaurin Series

### Maclaurin Series

A Maclaurin Series expansion of  $f(x)$  is

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \cdots$$

For example, suppose  $f(\mathbf{A}) = e^{\mathbf{A}}$ , then

$$f(\mathbf{A}) = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \cdots$$

Thus, for  $\mathbf{A} \in \mathbb{R}^{n \times n}$  such that  $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ , we may define

$$e^{\mathbf{A}} = \mathbf{P}e^{\mathbf{D}}\mathbf{P}^{-1}.$$

Indeed,

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \cdots$$

Indeed,

$$\begin{aligned}e^{\mathbf{A}} &= \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \cdots \\&= \mathbf{P}\mathbf{I}\mathbf{P}^{-1} + \mathbf{P}\mathbf{D}\mathbf{P}^{-1} + \frac{\mathbf{P}\mathbf{D}^2\mathbf{P}^{-1}}{2!} + \frac{\mathbf{P}\mathbf{D}^3\mathbf{P}^{-1}}{3!} + \cdots\end{aligned}$$

Indeed,

$$\begin{aligned}e^{\mathbf{A}} &= \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \cdots \\&= \mathbf{P}\mathbf{I}\mathbf{P}^{-1} + \mathbf{P}\mathbf{D}\mathbf{P}^{-1} + \frac{\mathbf{P}\mathbf{D}^2\mathbf{P}^{-1}}{2!} + \frac{\mathbf{P}\mathbf{D}^3\mathbf{P}^{-1}}{3!} + \cdots \\&= \mathbf{P}\left(\mathbf{I} + \mathbf{D} + \frac{\mathbf{D}^2}{2!} + \frac{\mathbf{D}^3}{3!} + \cdots\right)\mathbf{P}^{-1}\end{aligned}$$



Indeed,

$$\begin{aligned}e^{\mathbf{A}} &= \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \cdots \\&= \mathbf{P}\mathbf{I}\mathbf{P}^{-1} + \mathbf{P}\mathbf{D}\mathbf{P}^{-1} + \frac{\mathbf{P}\mathbf{D}^2\mathbf{P}^{-1}}{2!} + \frac{\mathbf{P}\mathbf{D}^3\mathbf{P}^{-1}}{3!} + \cdots \\&= \mathbf{P}\left(\mathbf{I} + \mathbf{D} + \frac{\mathbf{D}^2}{2!} + \frac{\mathbf{D}^3}{3!} + \cdots\right)\mathbf{P}^{-1} \\&= \mathbf{P}e^{\mathbf{D}}\mathbf{P}^{-1}\end{aligned}$$

Indeed,

$$\begin{aligned}
 e^{\mathbf{A}} &= \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \cdots \\
 &= \mathbf{P}\mathbf{I}\mathbf{P}^{-1} + \mathbf{P}\mathbf{D}\mathbf{P}^{-1} + \frac{\mathbf{P}\mathbf{D}^2\mathbf{P}^{-1}}{2!} + \frac{\mathbf{P}\mathbf{D}^3\mathbf{P}^{-1}}{3!} + \cdots \\
 &= \mathbf{P} \left( \mathbf{I} + \mathbf{D} + \frac{\mathbf{D}^2}{2!} + \frac{\mathbf{D}^3}{3!} + \cdots \right) \mathbf{P}^{-1} \\
 &= \mathbf{P}e^{\mathbf{D}}\mathbf{P}^{-1}
 \end{aligned}$$

Note that if  $\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ , we have  $\mathbf{I} + \mathbf{D} + \frac{\mathbf{D}^2}{2!} + \frac{\mathbf{D}^3}{3!} + \cdots$

$$\begin{aligned}
 &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} + \frac{1}{2!} \begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix} + \frac{1}{3!} \begin{bmatrix} \lambda_1^3 & 0 \\ 0 & \lambda_2^3 \end{bmatrix} + \cdots \\
 &= \begin{bmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{bmatrix}
 \end{aligned}$$

## Example

Compute  $e^{\mathbf{A}}$  for  $\mathbf{A} = \begin{bmatrix} -2 & -6 \\ 1 & 3 \end{bmatrix}$ .

- Eigenvalues: 0, 1.

## Example

Compute  $e^{\mathbf{A}}$  for  $\mathbf{A} = \begin{bmatrix} -2 & -6 \\ 1 & 3 \end{bmatrix}$ .

- Eigenvalues: 0, 1.
- Corresponding eigenvectors:  $[-3 \ 1]^T$ ,  $[-2 \ 1]^T$ .

## Example

Compute  $e^{\mathbf{A}}$  for  $\mathbf{A} = \begin{bmatrix} -2 & -6 \\ 1 & 3 \end{bmatrix}$ .

- Eigenvalues: 0, 1.
- Corresponding eigenvectors:  $[-3 \ 1]^T$ ,  $[-2 \ 1]^T$ .
- Take  $\mathbf{P} = \begin{bmatrix} -3 & -2 \\ 1 & 1 \end{bmatrix}$ , and we have  $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ .

## Example

Compute  $e^{\mathbf{A}}$  for  $\mathbf{A} = \begin{bmatrix} -2 & -6 \\ 1 & 3 \end{bmatrix}$ .

- Eigenvalues: 0, 1.
- Corresponding eigenvectors:  $[-3 \ 1]^T$ ,  $[-2 \ 1]^T$ .
- Take  $\mathbf{P} = \begin{bmatrix} -3 & -2 \\ 1 & 1 \end{bmatrix}$ , and we have  $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ .
- Thus,  $e^{\mathbf{A}} = \mathbf{P}f(\mathbf{D})\mathbf{P}^{-1}$ , and note that

$$f(\mathbf{D}) =$$

## Example

Compute  $e^{\mathbf{A}}$  for  $\mathbf{A} = \begin{bmatrix} -2 & -6 \\ 1 & 3 \end{bmatrix}$ .

- Eigenvalues: 0, 1.
- Corresponding eigenvectors:  $[-3 \ 1]^T$ ,  $[-2 \ 1]^T$ .
- Take  $\mathbf{P} = \begin{bmatrix} -3 & -2 \\ 1 & 1 \end{bmatrix}$ , and we have  $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ .
- Thus,  $e^{\mathbf{A}} = \mathbf{P}f(\mathbf{D})\mathbf{P}^{-1}$ , and note that

$$f(\mathbf{D}) = \begin{bmatrix} f(0) & 0 \\ 0 & f(1) \end{bmatrix}$$

## Example

Compute  $e^{\mathbf{A}}$  for  $\mathbf{A} = \begin{bmatrix} -2 & -6 \\ 1 & 3 \end{bmatrix}$ .

- Eigenvalues: 0, 1.
- Corresponding eigenvectors:  $[-3 \ 1]^T$ ,  $[-2 \ 1]^T$ .
- Take  $\mathbf{P} = \begin{bmatrix} -3 & -2 \\ 1 & 1 \end{bmatrix}$ , and we have  $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ .
- Thus,  $e^{\mathbf{A}} = \mathbf{P}f(\mathbf{D})\mathbf{P}^{-1}$ , and note that

$$f(\mathbf{D}) = \begin{bmatrix} f(0) & 0 \\ 0 & f(1) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & e \end{bmatrix}.$$



# Remarks

- A square matrix can have a zero eigenvalue, but never has zero eigenvectors.
- A zero matrix  $\mathbf{O}$  is diagonalizable, too.
  - All nonzero vectors are eigenvectors since all vectors  $\mathbf{v}$  satisfy  $\mathbf{O}\mathbf{v} = \mathbf{0}\mathbf{v}$ .

# Discussions