Mathematics for Machine Learning

— Linear Algebra: Eigenvalues, Eigenvectors, Eigenspaces, Cholesky Decomposition & Diagonalization

Joseph Chuang-Chieh Lin

Department of Computer Science & Information Engineering,
Tamkang University

Fall 2023

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

Outline

Characteristic Polynomial

Characteristic Polynomial

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

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

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

Note that

- $c_0 = \det(\boldsymbol{A})$.
- $c_{n-1} = (-1)^{n-1}\mathrm{tr}(\boldsymbol{A})$.

back

Example

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

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

Example

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

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

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

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

Eigenvalue Equation

Eigenvalues & Eigenvectors

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

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

if $\boldsymbol{A}\mathbf{x} = \lambda\mathbf{x}$.

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

Equivalent statements:

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

Remark

Eigenvectors are NOT unique.

Remark

Eigenvectors are NOT unique.

Suppose $\mathbf{x}$ is an eigenvector of $\boldsymbol{A}$ w.r.t. eigenvalue λ, then for any $c \in \mathbb{R} \setminus \mathbf{0}\}$

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

Theorems (or Definitions)

Theorem

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

Theorems (or Definitions)

Theorem

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

Algebraic Multiplicity

Let a square matrix $\boldsymbol{A}$ have an eigenvalue λ_i. The algebraic multiplicity of λ_i is the number of times the root appears in the characteristic polynomial.

Theorems (or Definitions)

Theorem

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

Algebraic Multiplicity

Let a square matrix $\boldsymbol{A}$ have an eigenvalue λ_i. The algebraic multiplicity of λ_i is the number of times the root appears in the characteristic polynomial.

Eigenspace

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

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

The Case of the Identity Matrix

The Case of the Identity Matrix

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

- what is $p_{\boldsymbol{I}}(\lambda)$?
- What are its eigenvalues and the associated eigenvectors?
- What are the eigenspaces?

Useful Properties (1/4)

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

Useful Properties (1/4)

- $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ possess the same eigenvaluesbut not necessarily the same eigenvectors.

Useful Properties (1/4)

- $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ possess the same eigenvaluesbut not necessarily the same eigenvectors.
- The eigenspace E_λ is null$(\boldsymbol{A} - \lambda \boldsymbol{I})$.

Useful Properties (1/4)

- $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ possess the same eigenvaluesbut not necessarily the same eigenvectors.
- The eigenspace E_λ is $\text{null}(\boldsymbol{A} - \lambda \boldsymbol{I})$.

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

Useful Properties (1/4)

- $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ possess the same eigenvaluesbut not necessarily the same eigenvectors.
- The eigenspace E_λ is null$(\boldsymbol{A} - \lambda \boldsymbol{I})$.

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

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

Useful Properties (2/4)

Theorem (4.13)

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

Theorem (4.14)

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

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

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

Useful Properties (3/4)

Theorem

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

Proof.

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

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

$\square$

Useful Properties (4/4)

Theorem (4.15; Spectral Theorem)

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

Useful Properties (4/4)

Theorem (4.15; Spectral Theorem)

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

Theorem (4.16)

For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, we have $\det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i$, where λ_i's are the eigenvalues of $\boldsymbol{A}$.

Useful Properties (4/4)

Theorem (4.15; Spectral Theorem)

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

Theorem (4.16)

For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, we have $\det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i$, where λ_i's are the eigenvalues of $\boldsymbol{A}$.

Theorem (4.17)

For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, we have $\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i$, where λ_i's are the eigenvalues of $\boldsymbol{A}$ recall?.

A Practical Example

- Google uses the eigenvector corresponding to the maximal eigenvalue of a matrix $\boldsymbol{A}$ 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

- Google uses the eigenvector corresponding to the maximal eigenvalue of a matrix $\boldsymbol{A}$ 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 $\boldsymbol{A}$ (prob.): modeling the navigation behavior of a user.
- **Goal:** $\mathbf{x}$, $\boldsymbol{A}\mathbf{x}$, $\boldsymbol{A}^2\mathbf{x}$, $\ldots$, $\mathbf{x}^*$

A Practical Example

- Google uses the eigenvector corresponding to the maximal eigenvalue of a matrix $\boldsymbol{A}$ 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 $\boldsymbol{A}$ (prob.): modeling the navigation behavior of a user.
- **Goal:** $\mathbf{x}, \boldsymbol{A}\mathbf{x}, \boldsymbol{A}^2\mathbf{x}, \ldots, \mathbf{x}^* \;\Rightarrow\; \boldsymbol{A}\mathbf{x}^* = \mathbf{x}^* \;\Rightarrow\;$ Turning to probabilities (normalization).

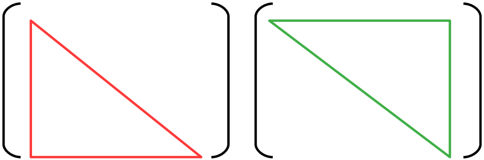
Outline

Cholesky Decomposition

Cholesky Decomposition

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

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



Example of Cholesky Factorization

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{L}\boldsymbol{L}^\top = \begin{bmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{bmatrix} \begin{bmatrix} \ell_{11} & \ell_{21} & \ell_{31} \\ 0 & \ell_{22} & \ell_{32} \\ 0 & 0 & \ell_{33} \end{bmatrix}.$$

We have

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

Finally, solve $\ell_{11}, \ldots, \ell_{33}$.

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.
  - $\det(\boldsymbol{A}) = \det(\boldsymbol{L})\det(\boldsymbol{L}^\top) = \det(\boldsymbol{L})^2$.
  - Note: $\det(\boldsymbol{L})$ can be computed efficiently ($\because$ triangular).

Outline

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

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

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

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

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

Similarity

Similarity

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

Similarity

Similarity

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

Diagonalizable

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

Similarity

Similarity

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

Diagonalizable

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

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

Eigenvectors & Diagonalization

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

We can show that

$$\boldsymbol{AP} = \boldsymbol{PD}.$$

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

Proof of the Claim

We can see that

$$\boldsymbol{AP} = \boldsymbol{A}[\mathbf{p}_1, \ldots, \mathbf{p}_n] = [\boldsymbol{A}\mathbf{p}_1, \ldots, \boldsymbol{A}\mathbf{p}_n],$$

and

Proof of the Claim

We can see that

$$\boldsymbol{AP} = \boldsymbol{A}[\mathbf{p}_1, \ldots, \mathbf{p}_n] = [\boldsymbol{A}\mathbf{p}_1, \ldots, \boldsymbol{A}\mathbf{p}_n],$$

and

$$\boldsymbol{PD} = [\mathbf{p}_1, \ldots, \mathbf{p}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Proof of the Claim

We can see that

$$\boldsymbol{AP} = \boldsymbol{A}[\mathbf{p}_1, \ldots, \mathbf{p}_n] = [\boldsymbol{A}\mathbf{p}_1, \ldots, \boldsymbol{A}\mathbf{p}_n],$$

and

$$\boldsymbol{PD} = [\mathbf{p}_1, \ldots, \mathbf{p}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = [\lambda_1 \mathbf{p}_1, \ldots, \lambda_n \mathbf{p}_n].$$

Proof of the Claim

We can see that

$$\boldsymbol{AP} = \boldsymbol{A}[\mathbf{p}_1, \ldots, \mathbf{p}_n] = [\boldsymbol{A}\mathbf{p}_1, \ldots, \boldsymbol{A}\mathbf{p}_n],$$

and

$$\boldsymbol{PD} = [\mathbf{p}_1, \ldots, \mathbf{p}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = [\lambda_1 \mathbf{p}_1, \ldots, \lambda_n \mathbf{p}_n].$$

Thus,

$$\begin{aligned} \boldsymbol{A}\mathbf{p}_1 &= \lambda_1 \mathbf{p}_1 \\ &\vdots \\ \boldsymbol{A}\mathbf{p}_n &= \lambda_n \mathbf{p}_n \end{aligned}$$

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

Eigendecomposition

Theorem [Eigendecomposition]

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

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

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

if and only if

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

Put it concisely

Theorem

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

- $\boldsymbol{A}$ is diagonalizable.
- $\boldsymbol{A}$ 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 of a symmetric matrix* $\boldsymbol{S} \in \mathbb{R}^{n \times n}$.

Remark

The spectral theorem tells us that:

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

Theorem

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

Example

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

1. Compute the eigenvalues and eigenvectors.

Example

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

1. Compute the eigenvalues and eigenvectors.

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

Example

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

1. Compute the eigenvalues and eigenvectors.

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

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

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

Example

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

1. Compute the eigenvalues and eigenvectors.

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

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

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

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

Example

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

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

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

Example

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

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

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

4. Construct $\boldsymbol{P}$:

Example

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

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

3. Check for independency of $\{\mathbf{p}_1, \mathbf{p}_2\}$. $\Longrightarrow \checkmark$
4. Construct $\boldsymbol{P}$: $\Longrightarrow \boldsymbol{P} = [\mathbf{p}_1, \ \mathbf{p}_2] = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$.

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

Example

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

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

3. Check for independency of $\{\mathbf{p}_1, \mathbf{p}_2\}$. $\Longrightarrow$ ✓
4. Construct $\boldsymbol{P}$: $\Longrightarrow \boldsymbol{P} = [\mathbf{p}_1,\ \mathbf{p}_2] = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$.

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

Example

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

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

3. Check for independency of $\{\mathbf{p}_1, \mathbf{p}_2\}$. $\Longrightarrow$ ✓
4. Construct $\boldsymbol{P}$: $\Longrightarrow \boldsymbol{P} = [\mathbf{p}_1,\ \mathbf{p}_2] = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$.

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

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

Remark On the Efficiency

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

Remark On the Efficiency

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

Remark On the Efficiency

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

Remark On the Efficiency

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

Remark On the Efficiency

- $\boldsymbol{A}^k = (\boldsymbol{PDP}^{-1})^k = (\boldsymbol{PDP}^{-1})(\boldsymbol{PDP}^{-1}) \cdots (\boldsymbol{PDP}^{-1}) = \boldsymbol{PD}^k\boldsymbol{P}^{-1}$.
- $\det(\boldsymbol{A}) = \det(\boldsymbol{PDP}^{-1}) = \det(\boldsymbol{P}) \det(\boldsymbol{D}) \det(\boldsymbol{P}^{-1})$

Remark On the Efficiency

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

Discussions