# Positive Definite Matrix

Chia-Ping Chen

Professor
Department of Computer Science and Engineering
National Sun Yat-sen University

Linear Algebra

# Outline

- Quadratic form
- Function approximation
- Singular value decomposition
- Minimum principles

## Notation

- $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}$: quadratic form
- $f(\boldsymbol{x})$: multi-variate function
- $\boldsymbol{\nabla} f(\boldsymbol{x})$: gradient vector of $f(\boldsymbol{x})$
- $\boldsymbol{H}(\boldsymbol{x})$: Hessian matrix
- $\sigma$: singular value
- $\boldsymbol{\Sigma}$: singular value matrix
- $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^T$: singular value decomposition of $\boldsymbol{A}$
- $\boldsymbol{A}^+$: pseudo-inverse of $\boldsymbol{A}$

# Quadratic Function and Quadratic Form

## Definition (quadratic function)

A quadratic function of $n$ variables is a sum of second-order terms.

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_i x_j$$

## Definition (quadratic form)

Let $\boldsymbol{A}$ be a matrix of order $n \times n$.

- The quadratic form of $\boldsymbol{A}$ is $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}$
- $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}$ is a quadratic function of $n$ variables

## SYMMETRIC MATRICES SUFFICE

Consider

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_i x_j$$

Define matrix $\boldsymbol{A}$

$$a_{ij} = \frac{1}{2}(c_{ij} + c_{ji})$$

Then

$$f(x_1, \ldots, x_n) = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}$$

- $\boldsymbol{A}$ is symmetric
- Eigenvalues of $\boldsymbol{A}$ are real

## Example (Quadratic form)

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

- Quadratic form of $\boldsymbol{A}$

$$\begin{aligned} \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= a x_1^2 + 2b x_1 x_2 + c x_2^2 \end{aligned}$$

- Quadratic function of $(x_1, x_2)$

## Definition (positive definite matrix)

Let $\boldsymbol{A}$ be a real symmetric matrix.

- $\boldsymbol{A}$ is **positive definit**e if the quadratic form of $\boldsymbol{A}$ is positive
- Specifically

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$$

for any $\boldsymbol{x} \neq \mathbf{0}$

### Lemma (positive definite ⇒ positive eigenvalues)

*Let $\boldsymbol{A}$ be positive definite. The eigenvalues of $\boldsymbol{A}$ are positive.*

### Proof.

Let $\lambda$ be an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{s}$ be a corresponding eigenvector. Then

$$\boldsymbol{A}\boldsymbol{s} = \lambda \boldsymbol{s}$$

It follows that

$$\boldsymbol{s}^T \boldsymbol{A}\boldsymbol{s} = \lambda(\boldsymbol{s}^T \boldsymbol{s})$$

Hence

$$\lambda = \frac{\boldsymbol{s}^T \boldsymbol{A}\boldsymbol{s}}{\boldsymbol{s}^T \boldsymbol{s}} > 0$$

□

## Lemma (positive eigenvalues ⇒ positive definite)

*If all eigenvalues of $\boldsymbol{A}$ are positive, $\boldsymbol{A}$ is positive definite.*

## Proof.

Let $\boldsymbol{A}$ have spectral decomposition $\boldsymbol{A} = \boldsymbol{Q\Lambda Q}^T$ where $\boldsymbol{Q}$ is orthogonal. Consider the quadratic form of $\boldsymbol{A}$.

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \overbrace{\boldsymbol{x}^T\boldsymbol{Q}}^{\boldsymbol{y}^T}\boldsymbol{\Lambda}\overbrace{\boldsymbol{Q}^T\boldsymbol{x}}^{\boldsymbol{y}} = \boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{y} = \sum_i \lambda_i y_i^2$$

For $\boldsymbol{x} \neq \boldsymbol{0}$, we have $\boldsymbol{y} \neq \boldsymbol{0}$ and thus $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \sum_i \lambda_i y_i^2 > 0$. Hence $\boldsymbol{A}$ is positive definite. □

## Lemma (positive definite ⇒ positive determinant)

*Let $\boldsymbol{A}$ be positive definite. Every leading principal sub-matrix of $\boldsymbol{A}$ has a positive determinant.*

## Proof.

Consider $\boldsymbol{x}^T = \begin{bmatrix} \boldsymbol{x}_k^T & \boldsymbol{0}^T \end{bmatrix}$ with $\boldsymbol{x}_k \in \mathbb{R}^k$. For $\boldsymbol{x}_k \neq \boldsymbol{0}$

$$\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_k^T & \boldsymbol{0}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_k & \boldsymbol{B} \\ \boldsymbol{B}^T & \boldsymbol{C} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{x}_k^T \boldsymbol{A}_k \boldsymbol{x}_k > 0$$

So $\boldsymbol{A}_k$ is positive definite, the eigenvalues of $\boldsymbol{A}_k$ are positive, and

$$|\boldsymbol{A}_k| = \prod_{i=1}^{k} \lambda_{k,i} > 0$$

where $\lambda_{k,i}$ is an eigenvalue of $\boldsymbol{A}_k$. □

## Lemma (positive determinants ⇒ positive pivots)

*If every leading principal sub-matrix of $\boldsymbol{A}$ has positive determinant, the pivots of $\boldsymbol{A}$ are positive.*

## Proof.

Let $\boldsymbol{A}$ have LDU decomposition $\boldsymbol{A} = \boldsymbol{LDU}$. Then

$$\begin{bmatrix} \boldsymbol{A}_k & \boldsymbol{B} \\ \boldsymbol{B}^T & \boldsymbol{C} \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_k & \boldsymbol{0} \\ * & * \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_k & \boldsymbol{0} \\ \boldsymbol{0} & * \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_k & * \\ \boldsymbol{0} & * \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_k \boldsymbol{D}_k \boldsymbol{U}_k & * \\ * & * \end{bmatrix}$$

So $\boldsymbol{A}_k = \boldsymbol{L}_k \boldsymbol{D}_k \boldsymbol{U}_k$. Let $d_1, \ldots, d_n$ be the pivots of $\boldsymbol{A}$. Then

$$d_1 = a_{11} > 0, \ d_k = \frac{|\boldsymbol{D}_k|}{|\boldsymbol{D}_{k-1}|} = \frac{|\boldsymbol{A}_k|}{|\boldsymbol{A}_{k-1}|} > 0, \quad k = 2, \ldots, n$$

□

**Lemma (positive pivots ⇒ positive definite)**

*If the pivots of $\boldsymbol{A}$ are positive, $\boldsymbol{A}$ is positive definite.*

**Proof.**

Let $\boldsymbol{A}$ have LDU decomposition $\boldsymbol{A} = \boldsymbol{LDU}$.

$$\boldsymbol{A} = \boldsymbol{A}^T \Rightarrow \boldsymbol{LDU} = \boldsymbol{U}^T\boldsymbol{D}\boldsymbol{L}^T \Rightarrow \boldsymbol{U} = \boldsymbol{L}^T$$

Thus

$$\boldsymbol{A} = \boldsymbol{LDL}^T = \boldsymbol{L}\boldsymbol{D}^{1/2}\boldsymbol{D}^{1/2}\boldsymbol{L}^T = \boldsymbol{R}^T\boldsymbol{R}$$

where $\boldsymbol{R} = \boldsymbol{D}^{1/2}\boldsymbol{L}^T$ is non-singular. For $\boldsymbol{x} \neq \boldsymbol{0}$

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{x} = (\boldsymbol{R}\boldsymbol{x})^T(\boldsymbol{R}\boldsymbol{x}) = \|\boldsymbol{R}\boldsymbol{x}\|^2 > 0$$

Hence $\boldsymbol{A}$ is positive definite. □

## Theorem (conditions for positive definite)

*The following conditions are equivalent.*

1. $A$ *is positive definite*
2. *The eigenvalues of* $A$ *are positive*
3. *The determinants of the leading principal sub-matrices of* $A$ *are positive*
4. *The pivots of* $A$ *are positive*

The previous slides show

$$① \Leftrightarrow ②$$

and

$$① \Rightarrow ③ \Rightarrow ④ \Rightarrow ①$$

## Example (positive definite matrix)

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Quadratic form

$$\begin{aligned} \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} &= 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3 \\ &= 2\left(x_1 - \frac{1}{2}x_2\right)^2 + \frac{3}{2}\left(x_2 - \frac{2}{3}x_3\right)^2 + \frac{4}{3}x_3^2 \end{aligned}$$

Eigenvalues, determinants, pivots

$$\text{spectrum}(\boldsymbol{A}) = \{2, 2 \pm \sqrt{2}\}, \ |\boldsymbol{A}_1| = 2, \ |\boldsymbol{A}_2| = 3, \ |\boldsymbol{A}_3| = 4$$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & & \\ & \frac{3}{2} & \\ & & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix}$$

## ELLIPSOID

Let $\boldsymbol{A}$ be positive definite. $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = 1$ defines an ellipsoid.

- With a spectral decomposition $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{y} = \sum_i \lambda_i y_i^2$$

  where $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q}_1 & \dots & \boldsymbol{q}_n \end{bmatrix}$ and $\boldsymbol{y} = \boldsymbol{Q}^T\boldsymbol{x}$
- $\{\boldsymbol{q}_1, \dots \boldsymbol{q}_n\}$ is an orthonormal eigenbasis
- $y_i\boldsymbol{q}_i$ is the projection of $\boldsymbol{x}$ on $\boldsymbol{q}_i$
- With axes $\boldsymbol{q}_1, \dots \boldsymbol{q}_n$, the coordinates are $y_1, \dots, y_n$
- $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = 1 \Rightarrow \sum_i \lambda_i y_i^2 = 1$ is an ellipsoid
- The intercepts are

$$l_i = \pm\left(\sqrt{\lambda_i}\right)^{-1}$$

# Example (quadratic form and ellipse)

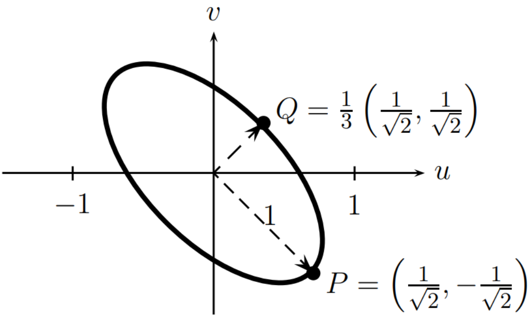


**Figure 6.2:** The ellipse $x^{\mathrm{T}}Ax = 5u^2 + 8uv + 5v^2 = 1$ and its principal axes.

**Definition (negative definite and semidefinite)**

Let $\boldsymbol{A}$ be a real symmetric matrix.

- $\boldsymbol{A}$ is negative definite if $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} < 0$ for $\boldsymbol{x} \neq \boldsymbol{0}$
- $\boldsymbol{A}$ is positive semidefinite if $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq 0$ for any $\boldsymbol{x}$
- $\boldsymbol{A}$ is negative semidefinite if $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \leq 0$ for any $\boldsymbol{x}$

# Approximation

## Definition (partial derivatives)

Let $f(x_1, \ldots, x_n)$ be a multi-variate function.

- First-order partial derivatives

$$f_{x_i} = \frac{\partial f}{\partial x_i},\ i = 1, \ldots, n$$

- Second-order partial derivatives

$$f_{x_i x_j} = \frac{\partial f_{x_i}}{\partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i},\ i, j = 1, \ldots, n$$

Note that

$$f_{x_i x_j} = f_{x_j x_i}$$

## Definition (gradient and Hessian)

Let $f(x_1, \ldots, x_n)$ be a function.

- Gradient vector

$$\boldsymbol{\nabla} f = \begin{bmatrix} f_{x_1} \\ \vdots \\ f_{x_n} \end{bmatrix}$$

- Hessian matrix

$$\boldsymbol{H} = \begin{bmatrix} f_{x_1 x_1} & \cdots & f_{x_1 x_n} \\ \vdots & \ddots & \vdots \\ f_{x_n x_1} & \cdots & f_{x_n x_n} \end{bmatrix}$$

## Function approximation

Let $f(\boldsymbol{x})$ be a function.

- First-order approximation near $\boldsymbol{x}_0$

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_0) + \boldsymbol{\nabla} f(\boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0)$$

- Second-order approximation near $\boldsymbol{x}_0$

$$\begin{aligned} f(\boldsymbol{x}) \approx f(\boldsymbol{x}_0) &+ \boldsymbol{\nabla} f(\boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0) \\ &+ \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T \boldsymbol{H}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \end{aligned}$$

## Definition (stationary points)

Let $f(\boldsymbol{x})$ be a function. $\boldsymbol{x}_0$ is a stationary point of $f(\boldsymbol{x})$ if

$$\boldsymbol{\nabla} f(\boldsymbol{x}_0) = \boldsymbol{0}$$

Let $\boldsymbol{x}_0$ be a stationary point of $f(\boldsymbol{x})$. Near $\boldsymbol{x}_0$, we have

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T \boldsymbol{H}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)$$

### Example (function approximation)

Approximate

$$f(x, y) = 2x^2 + 4xy + y^2$$

near $(0, 0)$.

$$\nabla f = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} 4x + 4y \\ 4x + 2y \end{bmatrix}, \quad \boldsymbol{H} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 2 \end{bmatrix}$$

$$f(\mathbf{0}) = 0, \quad \nabla f(\mathbf{0}) = \mathbf{0}, \quad \boldsymbol{H}(\mathbf{0}) = \begin{bmatrix} 4 & 4 \\ 4 & 2 \end{bmatrix}$$

$$f(\boldsymbol{x}) \approx f(\mathbf{0}) + \frac{1}{2}(\boldsymbol{x} - \mathbf{0})^T \boldsymbol{H}(\mathbf{0})(\boldsymbol{x} - \mathbf{0}) = 2x^2 + 4xy + y^2$$

## Example (function approximation)

Approximate

$$F(x,y) = 7 + 2(x+y)^2 - y\sin y - x^3$$

near $(0,0)$.

$$\nabla F = \begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} 4(x+y) - 3x^2 \\ 4(x+y) - \sin y - y\cos y \end{bmatrix}$$

$$\boldsymbol{H} = \begin{bmatrix} F_{xx} & F_{xy} \\ F_{yx} & F_{yy} \end{bmatrix} = \begin{bmatrix} 4 - 6x & 4 \\ 4 & 4 - 2\cos y + y\sin y \end{bmatrix}$$

$$F(\mathbf{0}) = 7, \quad \nabla F(\mathbf{0}) = \mathbf{0}, \quad \boldsymbol{H}(\mathbf{0}) = \begin{bmatrix} 4 & 4 \\ 4 & 2 \end{bmatrix}$$

$$F(\boldsymbol{x}) \approx F(\mathbf{0}) + \frac{1}{2}(\boldsymbol{x} - \mathbf{0})^T \boldsymbol{H}(\mathbf{0})(\boldsymbol{x} - \mathbf{0}) = 7 + 2x^2 + 4xy + y^2$$

### Definition (local minimum and local maximum)

Let $f(\boldsymbol{x})$ be a function.

- $\boldsymbol{x}_0$ is a **local minimum** if in a neighborhood of $\boldsymbol{x}_0$

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0)$$

- $\boldsymbol{x}_0$ is a **local maximum** if in a neighborhood of $\boldsymbol{x}_0$

$$f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)$$

### Optimality of a Stationary Point

Let $f(\boldsymbol{x})$ be a function. Let $\boldsymbol{x}_0$ be a stationary point and $\boldsymbol{H}$ be the Hessian matrix at $\boldsymbol{x}_0$.

- $\boldsymbol{x}_0$ is a local minimum if $\boldsymbol{H}$ is positive semidefinite
- $\boldsymbol{x}_0$ is a local maximum if $\boldsymbol{H}$ is negative semidefinite
- $\boldsymbol{x}_0$ is a saddle point if it is neither a local maximum nor a local minimum

For example, $(0, 0)$ is a saddle point of $F(x, y)$.

## bowl or saddle

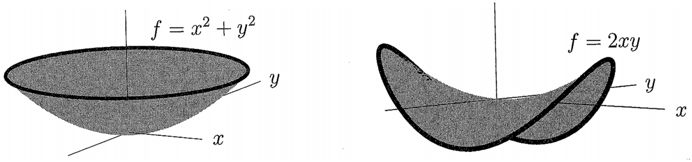


**Figure 6.1:** A bowl and a saddle: Definite $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and indefinite $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

## Singular Value Decomposition

**BASIC IDEA**

Let $\boldsymbol{A}$ be a real matrix.

- Real symmetric $\left(\boldsymbol{A}^T\boldsymbol{A}\right)$ and $\left(\boldsymbol{A}\boldsymbol{A}^T\right)$
- Decompose $\boldsymbol{A}$ with the eigenvalues and eigenvectors of $\left(\boldsymbol{A}^T\boldsymbol{A}\right)$ and $\left(\boldsymbol{A}\boldsymbol{A}^T\right)$
- An extension of eigen-decomposition

$$\left(\boldsymbol{A}^T\boldsymbol{A}\right)^T = \boldsymbol{A}^T\left(\boldsymbol{A}^T\right)^T = \left(\boldsymbol{A}^T\boldsymbol{A}\right)$$

$$\left(\boldsymbol{A}\boldsymbol{A}^T\right)^T = \left(\boldsymbol{A}^T\right)^T\boldsymbol{A}^T = \left(\boldsymbol{A}\boldsymbol{A}^T\right)$$

## $\left(A^T A\right)$ AND $\left(AA^T\right)$

Let $A$ be a real matrix.

- Positive semi-definite $\left(A^T A\right)$ and $\left(AA^T\right)$
- Non-negative eigenvalues
- Real and orthonormal eigenvectors

$$x^T \left(A^T A\right) x = (Ax)^T (Ax) = \|Ax\|^2 \geq 0$$

$$y^T \left(AA^T\right) y = \left(A^T y\right)^T \left(A^T y\right) = \|A^T y\|^2 \geq 0$$

## Definition (singular value and singular vector)

Let $\boldsymbol{A}$ be a real matrix. Square roots of the positive eigenvalues of $\left(\boldsymbol{A}^T\boldsymbol{A}\right)$ are the **singular values** of $\boldsymbol{A}$.

Let $\sigma$ be a singular value of $\boldsymbol{A}$.

- $\sigma^2$ is an eigenvalue of $\left(\boldsymbol{A}^T\boldsymbol{A}\right)$
- $\exists\, \boldsymbol{v} \neq \boldsymbol{0}$
$$\left(\boldsymbol{A}^T\boldsymbol{A}\right)\boldsymbol{v} = \sigma^2\boldsymbol{v}$$
- $\sigma^2$ is also an eigenvalue of $\left(\boldsymbol{A}\boldsymbol{A}^T\right)$
- $\exists\, \boldsymbol{u} \neq \boldsymbol{0}$
$$\left(\boldsymbol{A}\boldsymbol{A}^T\right)\boldsymbol{u} = \sigma^2\boldsymbol{u}$$
- $\boldsymbol{u}$ is **left singular vector** and $\boldsymbol{v}$ is **right singular vector**

## Definition (singular space)

Let $\boldsymbol{A}$ be a real matrix and $\sigma$ be a singular value of $\boldsymbol{A}$.

- **Right singular space**

$$\mathbb{R}_\sigma(\boldsymbol{A}) = \left\{ \boldsymbol{v} \,\middle|\, \left( \boldsymbol{A}^T \boldsymbol{A} \right) \boldsymbol{v} = \sigma^2 \boldsymbol{v} \right\}$$

- **Left singular space**

$$\mathbb{L}_\sigma(\boldsymbol{A}) = \left\{ \boldsymbol{u} \,\middle|\, \left( \boldsymbol{A} \boldsymbol{A}^T \right) \boldsymbol{u} = \sigma^2 \boldsymbol{u} \right\}$$

## Lemma (linearly independent singular vectors*)

*Let $\boldsymbol{A}$ be a real matrix of* **rank** $r$. *There exists a linearly independent set containing $r$ right singular vectors of $\boldsymbol{A}$.*

## Proof.

Suppose $\mathbb{N}(\boldsymbol{A})$ is of dimension $n-r$. $\mathbb{N}\left(\boldsymbol{A}^T\boldsymbol{A}\right) = \mathbb{N}(\boldsymbol{A})$ since

$$\left(\boldsymbol{A}^T\boldsymbol{A}\right)\boldsymbol{x} = \boldsymbol{0} \Rightarrow \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = 0 \Rightarrow \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \Rightarrow \left(\boldsymbol{A}^T\boldsymbol{A}\right)\boldsymbol{x} = \boldsymbol{0}$$

So eigenvalue 0 of $\left(\boldsymbol{A}^T\boldsymbol{A}\right)$ has multiplicity $(n-r)$, and the non-zero eigenvalues of $\left(\boldsymbol{A}^T\boldsymbol{A}\right)$ have total multiplicity

$$n-(n-r) = r$$

Thus, there are $r$ linearly independent right singular vectors. □

## Theorem (singular value decomposition)

*Let $\boldsymbol{A}$ be a real matrix of order $m \times n$.*

$$\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^T$$

- $\boldsymbol{\Sigma}$ *is an $m \times n$ "diagonal" matrix with the singular values of $\boldsymbol{A}$ as the leading diagonal elements*
- $\boldsymbol{U}$ *is an $m \times m$ orthogonal matrix with the eigenvectors of $\left(\boldsymbol{A}\boldsymbol{A}^T\right)$ as columns*
- $\boldsymbol{V}$ *is an $n \times n$ orthogonal matrix with the eigenvectors of $\left(\boldsymbol{A}^T\boldsymbol{A}\right)$ as columns*

## Matrix Construction

- Find singular values $\sigma_1 \ \dots \ \sigma_r$
- Find orthonormal right singular vectors $\boldsymbol{v}_1 \ \dots \ \boldsymbol{v}_r$
- Find orthonormal left singular vectors $\boldsymbol{u}_i = \frac{\boldsymbol{A}\boldsymbol{v}_i}{\sigma_i}$
- Find orthonormal $\boldsymbol{v}_{r+1} \ \dots \ \boldsymbol{v}_n$ in $\mathbb{E}_0\left(\boldsymbol{A}^T\boldsymbol{A}\right)$
- Find orthonormal $\boldsymbol{u}_{r+1} \ \dots \ \boldsymbol{u}_m$ in $\mathbb{E}_0\left(\boldsymbol{A}\boldsymbol{A}^T\right)$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & \dots & 0 & \\ \vdots & \ddots & \vdots & \mathbf{0} \\ 0 & \dots & \sigma_r & \\ & \mathbf{0} & & \mathbf{0} \end{bmatrix}$$

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_m \end{bmatrix}, \ \boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_n \end{bmatrix}$$

## PROOF OF SINGULAR VALUE DECOMPOSITION.

For $j = 1, \ldots, r$ and $i = 1, \ldots, m$

$$\left(\boldsymbol{A}\boldsymbol{A}^T\right)\boldsymbol{u}_j = \frac{\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_j}{\sigma_j} = \frac{\boldsymbol{A}\sigma_j^2\boldsymbol{v}_j}{\sigma_j} = \sigma_j^2\boldsymbol{u}_j$$

$$\boldsymbol{u}_i^T\boldsymbol{u}_j = \left(\frac{\boldsymbol{A}\boldsymbol{v}_i}{\sigma_i}\right)^T\left(\frac{\boldsymbol{A}\boldsymbol{v}_j}{\sigma_j}\right) = \frac{\boldsymbol{v}_i^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_j}{\sigma_i\sigma_j} = \delta_{ij}$$

$$(\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V})_{ij} = \boldsymbol{u}_i^T\boldsymbol{A}\boldsymbol{v}_j = \boldsymbol{u}_i^T(\sigma_j\boldsymbol{u}_j) = \sigma_j\delta_{ij}$$

For $j = r+1, \ldots, n$ and $i = 1, \ldots, m$

$$(\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V})_{ij} = \boldsymbol{u}_i^T\boldsymbol{A}\boldsymbol{v}_j = \boldsymbol{u}_i^T\boldsymbol{0} = 0$$

Thus

$$\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V} = \boldsymbol{\Sigma} \;\Rightarrow\; \boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

□

## Matrices in Singular Value Decomposition

Suppose $\boldsymbol{A}$ has SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$.

- $\boldsymbol{U}$ is eigenvector matrix of $\left(\boldsymbol{A}\boldsymbol{A}^T\right)$

$$\boldsymbol{A}\boldsymbol{A}^T = \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\right)\left(\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\right) = \boldsymbol{U}\overbrace{\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\right)}^{\text{diagonal}}\boldsymbol{U}^T$$

- $\boldsymbol{V}$ is eigenvector matrix of $\left(\boldsymbol{A}^T\boldsymbol{A}\right)$

$$\boldsymbol{A}^T\boldsymbol{A} = \left(\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\right)\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\right) = \boldsymbol{V}\overbrace{\left(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\right)}^{\text{diagonal}}\boldsymbol{V}^T$$

### SINGULAR VECTORS AND FUNDAMENTAL SUBSPACES*

Suppose $\boldsymbol{A}$ has SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$.

- The right singular vectors of $\boldsymbol{A}$ in $\boldsymbol{V}$ form an orthonormal basis of $\mathbb{C}\left(\boldsymbol{A}^T\right)$
- The left singular vectors of $\boldsymbol{A}$ in $\boldsymbol{U}$ form an orthonormal basis of $\mathbb{C}\left(\boldsymbol{A}\right)$

Suppose $\boldsymbol{A}$ has order $m \times n$ and rank $r$.

- $\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n$ are eigenvectors of $\left(\boldsymbol{A}^T\boldsymbol{A}\right)$ with eigenvalue 0
$$\left(\boldsymbol{A}^T\boldsymbol{A}\right)\boldsymbol{v}_j = 0\,\boldsymbol{v}_j = \boldsymbol{0}$$
so they form a basis of $\mathbb{N}(\boldsymbol{A}^T\boldsymbol{A}) = \mathbb{N}(\boldsymbol{A})$. It follows that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ form a basis of the orthogonal complement of $\mathbb{N}(\boldsymbol{A})$, i.e. $\mathbb{C}\left(\boldsymbol{A}^T\right)$.
- The first $r$ columns of $\boldsymbol{AV} = \boldsymbol{U\Sigma}$ means
$$\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$$
Thus $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ are vectors in $\mathbb{C}\left(\boldsymbol{A}\right)$. Since they are linearly independent, they form a basis of $\mathbb{C}(\boldsymbol{A})$.

## Example (singular value decomposition)

$$\boldsymbol{A} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

- Singular values

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

The eigenvalues are $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$. The singular values are

$$\sigma_1 = \sqrt{3}, \ \sigma_2 = 1$$

- Right singular vectors (and orthonormal eigenvectors)

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- Left singular vectors (and orthonormal eigenvectors)

$$\boldsymbol{u}_1 = \frac{\boldsymbol{A}\boldsymbol{v}_1}{\sigma_1} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \boldsymbol{u}_2 = \frac{\boldsymbol{A}\boldsymbol{v}_2}{\sigma_2} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- SVD

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

## SVD for approximation

Suppose $\boldsymbol{A}$ has rank $r$ and SVD $\boldsymbol{A} = \boldsymbol{U\Sigma V}^T$.

$$\begin{aligned}\boldsymbol{A} &= \boldsymbol{U\Sigma V}^T \\ &= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^T \\ &= \boldsymbol{A}_1 + \cdots + \boldsymbol{A}_r\end{aligned}$$

- $\boldsymbol{A}$ is the sum of $r$ matrices of rank 1
- An image of size $1000 \times 1000$ can be compressed with a rate of 90% when 50 terms are used

*Data Compression with SVD*

### Theorem (pseudo-inverse* and SVD)

*Suppose* $\boldsymbol{A}$ *has SVD* $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$.

- *The pseudo inverse of* $\boldsymbol{A}$ *is*

$$\boldsymbol{A}^+ = \boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^T$$

- *For any* $\boldsymbol{b}$, *the minimum-length least-squares solution to* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *is*

$$\boldsymbol{x}^+ = \boldsymbol{A}^+\boldsymbol{b}$$

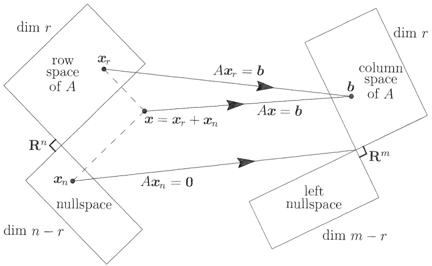


**Figure 3.4:** The true action $Ax = A(x_{\text{row}} + x_{\text{null}})$ of any $m$ by $n$ matrix.

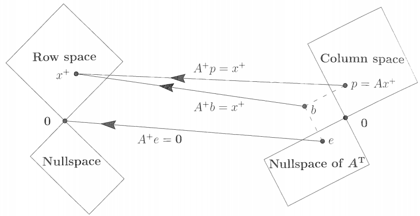


**Figure 6.3:** The pseudoinverse $A^+$ inverts $A$ where it can on the column space.

## Minimum Principles*

## MINIMUM PRINCIPLE FOR SQUARE SYSTEM

Let $\boldsymbol{A}$ be positive definite. Consider $\boldsymbol{Ax} = \boldsymbol{b}$.

- The system is non-singular
- It can be solved by minimum principle: $\boldsymbol{x}_0$ is a solution of $\boldsymbol{Ax} = \boldsymbol{b}$ if and only if it minimizes

$$P(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Ax} - \boldsymbol{x}^T\boldsymbol{b}$$

Note

$$\boldsymbol{\nabla}\left(\boldsymbol{x}^T\boldsymbol{Ax}\right) = 2\boldsymbol{Ax},\ \boldsymbol{\nabla}\left(\boldsymbol{x}^T\boldsymbol{b}\right) = \boldsymbol{b}$$

so

$$\boldsymbol{\nabla}P(\boldsymbol{x}) = \boldsymbol{Ax} - \boldsymbol{b}$$

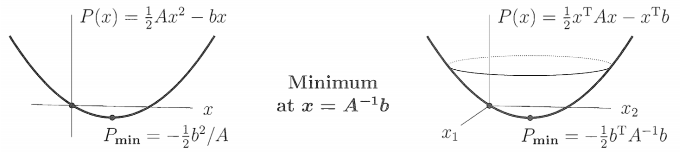


**Figure 6.4:** The graph of a positive quadratic $P(x)$ is a parabolic bowl.

## Minimum Principle for Over-determined System

Let $\boldsymbol{Ax} = \boldsymbol{b}$ be an over-determined system of linear equations. Such a system can be solved by minimum principle.

Specifically, the sum of squared errors as a function of $\boldsymbol{x}$ is

$$\begin{aligned} E(\boldsymbol{x}) &= \|\boldsymbol{Ax} - \boldsymbol{b}\|^2 \\ &= (\boldsymbol{Ax} - \boldsymbol{b})^T(\boldsymbol{Ax} - \boldsymbol{b}) \\ &= \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{b} \end{aligned}$$

An $\boldsymbol{x}_0$ that achieves the minimum of $E(\boldsymbol{x})$ satisfies

$$\boldsymbol{\nabla} E(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{x}_0} = \boldsymbol{0}$$

That is

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{A}^T\boldsymbol{b}$$

## Definition (Rayleigh quotient)

Let $\boldsymbol{A}$ be a symmetric matrix. The Rayleigh quotient of $\boldsymbol{A}$ is

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}$$

- Let $\mathcal{Q} = \boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ be an orthonormal eigenbasis of $\boldsymbol{A}$, corresponding to eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$
- For any $\boldsymbol{x} = \sum_i x_i \boldsymbol{q}_i$, we have

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} = \frac{\left(\sum_i x_i\boldsymbol{q}_i\right)^T \boldsymbol{A} \left(\sum_i x_i\boldsymbol{q}_i\right)}{\left(\sum_i x_i\boldsymbol{q}_i\right)^T \left(\sum_i x_i\boldsymbol{q}_i\right)}$$

$$= \frac{\sum_i \lambda_i x_i^2}{\sum_i x_i^2}$$

## Theorem (extremum of Rayleigh quotient)

*Let* $\boldsymbol{A}$ *be a symmetric matrix and* $\mathcal{Q} = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$ *be orthonormal eigenbasis of* $\boldsymbol{A}$, *corresponding to eigenvalues*

$$\lambda_1 \leq \cdots \leq \lambda_n$$

- *The global minimum of the Rayleigh quotient of* $\boldsymbol{A}$ *is* $\lambda_1$
- *The global maximum of the Rayleigh quotient of* $\boldsymbol{A}$ *is* $\lambda_n$
- *The minimum* $\lambda_1$ *is attained by* $\boldsymbol{q}_1$ *(i.e.* $[\boldsymbol{x}_{\mathcal{Q}}] = \{\delta_{i,1}\}$*)*

$$\lambda_1 = \min_{\boldsymbol{x}} R(\boldsymbol{x})$$

- *The maximum* $\lambda_n$ *is attained by* $\boldsymbol{q}_n$ *(i.e.* $[\boldsymbol{x}_{\mathcal{Q}}] = \{\delta_{i,n}\}$*)*

$$\lambda_n = \max_{\boldsymbol{x}} R(\boldsymbol{x})$$

## DIAGONAL ELEMENTS AND EIGENVALUES

Let $\boldsymbol{A}$ be a symmetric matrix.

- For a unit vector along coordinate axis, $R(\boldsymbol{e}_i) = a_{ii}$
- Thus $a_{ii}$ is bounded by eigenvalues

$$\lambda_1 \leq a_{ii} \leq \lambda_n$$

- In the cases of all positive eigenvalues for $\boldsymbol{A}$, we have

$$\frac{1}{\sqrt{\lambda_n}} \leq \frac{1}{\sqrt{a_{ii}}} \leq \frac{1}{\sqrt{\lambda_1}}$$

- The intercept of ellipsoid $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = 1$ along a coordinate axis is bounded

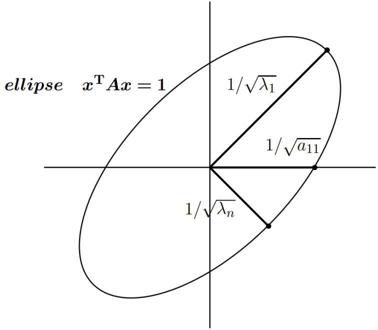


**Figure 6.6:** The farthest $x = x_1/\sqrt{\lambda_1}$ and the closet $x = x_n/\sqrt{\lambda_n}$ both give $x^T A x = x^T \lambda x = 1$. These are the major axes of the ellipse.

### Theorem (saddle points of Rayleigh quotient)

*Let $\boldsymbol{A}$ be a symmetric matrix and $\mathcal{Q} = \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_n\}$ be orthonormal eigenbasis of $\boldsymbol{A}$, corresponding to eigenvalues*

$$\lambda_1 \leq \cdots \leq \lambda_n$$

*The eigenvectors $\boldsymbol{q}_2, \dots, \boldsymbol{q}_{n-1}$ are saddle points of $R(\boldsymbol{x})$.*

Consider $\boldsymbol{q}_2$ for example.

- If we move from $\boldsymbol{q}_2$ along $\boldsymbol{q}_1$, $R(\boldsymbol{x})$ decreases
- If we move from $\boldsymbol{q}_2$ along $\boldsymbol{q}_2$, $R(\boldsymbol{x})$ does not change
- If we move from $\boldsymbol{q}_2$ along $\boldsymbol{q}_3$, $R(\boldsymbol{x})$ increases

## Theorem (Rayleigh quotient in a hyperplane)

*Let $\boldsymbol{A}$ be a symmetric matrix with orthonormal eigenvectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ and eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$.*

$$\lambda_2 = \max_{\boldsymbol{v}} \left[ \min_{\boldsymbol{x} \in \boldsymbol{v}^{\perp}} R(\boldsymbol{x}) \right]$$

Let $\boldsymbol{v}$ be a vector and consider $R(\boldsymbol{x})$ in the subspace $\boldsymbol{v}^{\perp}$.

- For $\boldsymbol{v} = \boldsymbol{q}_1$
$$\lambda_2 = \min_{\boldsymbol{x} \in \boldsymbol{q}_1^{\perp}} R(\boldsymbol{x})$$
- Given $\boldsymbol{v}$, $R(\boldsymbol{x})$ can be smaller as $\boldsymbol{x}$ can have component along $\boldsymbol{q}_1$
$$\lambda_2 \geq \min_{\boldsymbol{x} \in \boldsymbol{v}^{\perp}} R(\boldsymbol{x})$$

Thus

$$\lambda_2 = \max_{\boldsymbol{v}} \left[ \min_{\boldsymbol{x} \in \boldsymbol{v}^{\perp}} R(\boldsymbol{x}) \right]$$

## Corollary (Rayleigh quotient in a subspace)

- *For the maximum in a hyperplane, we have*

$$\lambda_{n-1} \le \max_{\boldsymbol{x} \in \boldsymbol{v}^{\perp}} R(\boldsymbol{x})$$

$$\lambda_{n-1} = \min_{\boldsymbol{v}} \left[ \max_{\boldsymbol{x} \in \boldsymbol{v}^{\perp}} R(\boldsymbol{x}) \right]$$

- *Let $\mathcal{V}$ be a subspace of dimension $j$. We have*

$$\lambda_{j+1} = \max_{\mathcal{V}} \left[ \min_{\boldsymbol{x} \in \mathcal{V}^{\perp}} R(\boldsymbol{x}) \right]$$

$$\lambda_{n-j} = \min_{\mathcal{V}} \left[ \max_{\boldsymbol{x} \in \mathcal{V}^{\perp}} R(\boldsymbol{x}) \right]$$

## Theorem (intertwining of eigenvalues)

*Let* $\boldsymbol{A}$ *be a real symmetric matrix and* $\boldsymbol{B}$ *be* $(n-1)\times(n-1)$ *matrix formed by stripping the last row and column of* $\boldsymbol{A}$.

$$\lambda_1(\boldsymbol{A}) \le \lambda_1(\boldsymbol{B}) \le \lambda_2(\boldsymbol{A}) \le \lambda_2(\boldsymbol{B}) \le \cdots \le \lambda_{n-1}(\boldsymbol{B}) \le \lambda_n(\boldsymbol{A})$$

## Example (intertwining of eigenvalues)

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \ \boldsymbol{B} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\lambda_1(\boldsymbol{A}) = 2-\sqrt{2}, \ \lambda_2(\boldsymbol{A}) = 2, \ \lambda_3(\boldsymbol{A}) = 2+\sqrt{2}$$

$$\lambda_1(\boldsymbol{B}) = 1, \ \lambda_2(\boldsymbol{B}) = 3$$