Matrix Analysis and Applications

Chapter 2: Eigenvalues and Eigenvectors

Instructor: Kai Lu

(http://seit.sysu.edu.cn/teacher/1801)

School of Electronics and Information Technology
Sun Yat-sen University

November 22, 2020

Table of Contents

Table of Contents

Eigenvalue problem: Given $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, find an n-dimensional vector $\boldsymbol{v}$, $\boldsymbol{v} \neq \mathbf{0}$, such that

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}, \tag{1}$$

for some scalar λ, where

- $\boldsymbol{v}$ is called an **eigenvector** of $\boldsymbol{A}$ (direction);
- λ is called an **eigenvalue** of $\boldsymbol{A}$ (scaling factor).

Eigenvalue problem: Given $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, find an n-dimensional vector $\boldsymbol{v}$, $\boldsymbol{v} \neq \boldsymbol{0}$, such that

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}, \tag{1}$$

for some scalar λ, where

- $\boldsymbol{v}$ is called an **eigenvector** of $\boldsymbol{A}$ (direction);
- λ is called an **eigenvalue** of $\boldsymbol{A}$ (scaling factor).

Matrix-vector multiplication for various vectors, where $\boldsymbol{A} = [4, 1; 1, 4]$.

Eigenvalue problem: Given $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, find an n-dimensional vector $\boldsymbol{v}$, $\boldsymbol{v} \neq \boldsymbol{0}$, such that

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}, \tag{1}$$

for some scalar λ, where

- $\boldsymbol{v}$ is called an **eigenvector** of $\boldsymbol{A}$ (direction);
- λ is called an **eigenvalue** of $\boldsymbol{A}$ (scaling factor).

Matrix-vector multiplication for various vectors, where $\boldsymbol{A} = [4, 1; 1, 4]$.

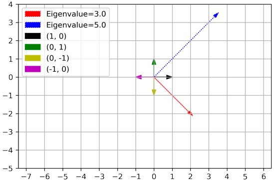


Figure 1: Original Vectors

Eigenvalue problem: Given $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, find an n-dimensional vector $\boldsymbol{v}$, $\boldsymbol{v} \neq \boldsymbol{0}$, such that

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}, \tag{1}$$

for some scalar λ, where

- $\boldsymbol{v}$ is called an **eigenvector** of $\boldsymbol{A}$ (direction);
- λ is called an **eigenvalue** of $\boldsymbol{A}$ (scaling factor).

Matrix-vector multiplication for various vectors, where $\boldsymbol{A} = [4, 1; 1, 4]$.

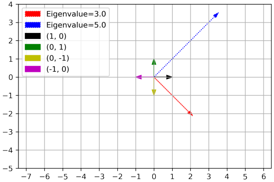


Figure 1: Original Vectors

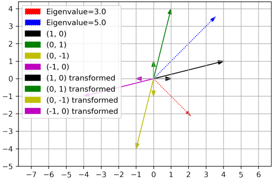


Figure 2: Transformed Vectors

Eigenvalue problem: Given $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, find an n-dimensional vector $\boldsymbol{v}$, $\boldsymbol{v} \neq \boldsymbol{0}$, such that

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}, \tag{1}$$

for some scalar λ, where

- $\boldsymbol{v}$ is called an **eigenvector** of $\boldsymbol{A}$ (direction);
- λ is called an **eigenvalue** of $\boldsymbol{A}$ (scaling factor).

Matrix-vector multiplication for various vectors, where $\boldsymbol{A} = [4, 1; 1, 4]$.

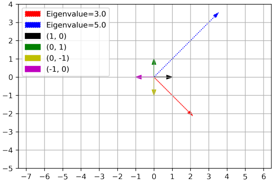


Figure 1: Original Vectors

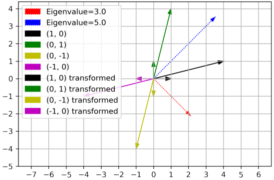


Figure 2: Transformed Vectors

Remark 1

The word eigenvalue is half-German, half-English. The German adjective **eigen** *means 'own' in the sense of characterizing an intrinsic property. Some mathematicians use the term* **characteristic value/vector** *instead of eigenvalue/eigenvector.*

Characteristic Polynomial

The eigenvalue problem can be rewritten as

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}. \tag{2}$$

The equation above holds if and only if $\boldsymbol{A} - \lambda \boldsymbol{I}$ is singular, thus

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0. \tag{3}$$

Characteristic Polynomial

The eigenvalue problem can be rewritten as

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = \mathbf{0}. \tag{2}$$

The equation above holds if and only if $\boldsymbol{A} - \lambda \boldsymbol{I}$ is singular, thus

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0. \tag{3}$$

- The equation above is called the **characteristic equation** of $\boldsymbol{A}$.
- Let $p(\lambda) \triangleq \det(\boldsymbol{A} - \lambda \boldsymbol{I})$ called the **characteristic polynomial** of $\boldsymbol{A}$.
- It is shown that $p(\lambda)$ is a polynomial of degree n. As a degree-n polynomial has exactly n roots, we can write

$$p(\lambda) = \prod_{i=1}^{n}(\lambda_i - \lambda), \tag{4}$$

where $\lambda_1, \cdots, \lambda_n$ are the roots.

Characteristic Polynomial (cont'd)

Remark 2 (Abel's impossibility theorem)

As we have seen, computing eigenvalues boils down to solving a polynomial equation. But determining solutions to polynomial equations can be a formidable task. It was proven in the nineteenth century that it's impossible to express the roots of a general polynomial of degree five or higher using radicals of the coefficients (i.e, Abel–Ruffini theorem, a.k.a. Abel's impossibility theorem[1]*). This means that there does not exist a generalized version of the quadratic formula for polynomials of degree greater than four, and general polynomial equations cannot be solved by a finite number of arithmetic operations involving* $+$, $-$, $\times$, $\div$, $\sqrt[n]{}$. *Unlike solving* $\boldsymbol{Ax}=\boldsymbol{b}$*, the eigenvalue problem generally requires an infinite algorithm, so all practical eigenvalue computations are accomplished by iterative methods. For more information, please refer to*[2].

[1] `https://en.wikipedia.org/wiki/Abel-Ruffini_theorem`

[2] V. B. Alekseev, *Abel's Theorem in Problems and Solutions*, Kluwer Academic Publishers, 2004.

Eigenvalues and Eigenvectors

Since the characteristic polynomial has n roots, the eigenvalue problem has n solutions

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, \quad i = 1, \cdots, n \tag{5}$$

where λ_i is the i^{th} eigenvalue, each being a root of $p(\lambda)$; $\boldsymbol{v}_i$ is the corresponding eigenvector of λ_i.

Eigenvalues and Eigenvectors

Since the characteristic polynomial has n roots, the eigenvalue problem has n solutions

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, \quad i = 1, \cdots, n \tag{5}$$

where λ_i is the i^{th} eigenvalue, each being a root of $p(\lambda)$; $\boldsymbol{v}_i$ is the corresponding eigenvector of λ_i.

Remark 3

1) *If $\boldsymbol{v}_i$ is an eigenvector, then, for any $\alpha \in \mathbb{R}$, $\alpha \boldsymbol{v}_i$ is also an eigenvector.*
   - *In the sequel, an eigenvector is assumed to be normalized such that $\|\boldsymbol{v}_i\|_2 = 1$.*
2) *λ_i may be complex-valued even though $\boldsymbol{A}$ is real (a polynomial $p(\lambda)$ with real coefficients can still have complex roots). As a result, $\boldsymbol{v}_i$ could be complex.*

Similarity

We consider complex-valued square matrices in the sequel.

Given $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{S} \in \mathbb{C}^{n\times n}$, $\boldsymbol{B}$ is said to be similar to $\boldsymbol{A}$ if there exists a nonsingular matrix $\boldsymbol{S}$ such that

$$\boldsymbol{B} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}. \tag{6}$$

Similarity

We consider complex-valued square matrices in the sequel.

Given $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{S} \in \mathbb{C}^{n \times n}$, $\boldsymbol{B}$ is said to be similar to $\boldsymbol{A}$ if there exists a nonsingular matrix $\boldsymbol{S}$ such that

$$\boldsymbol{B} = \boldsymbol{S}^{-1} \boldsymbol{A} \boldsymbol{S}. \tag{6}$$

Property 1

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are similar, then the characteristic polynomial of $\boldsymbol{A}$ is same as that of $\boldsymbol{B}$.

Property 2

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are similar, then they have the same eigenvalues, and the same determinant.

Diagonalizability

A matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ is said to be **diagonalizable** if it is similar to a diagonal matrix, i.e.,

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

where $\boldsymbol{D}$ is a diagonal matrix, and $\boldsymbol{S}$ is nonsingular.

Theorem 1

A matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ is diagonalizable if and only if there is a set of n linearly independent vectors, each of which is an eigenvector of $\boldsymbol{A}$.

Diagonalizability and Eigenvalues

Distinct eigenvalues:

Property 3

Suppose that $\{\lambda_1, \cdots, \lambda_k\}$, $k \leq n$, *is a set of distinct eigenvalues, i.e.,* $\lambda_i \neq \lambda_j$ *for* $i \neq j$, $\forall i, j \in \{1, 2, \cdots, k\}$. *Then, the corresponding set of eigenvectors* $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ *is linearly independent.*

- If all the eigenvalues of $\boldsymbol{A}$ are distinct, then, by Theorem 1, $\boldsymbol{A}$ is diagonalizable.

Repeated eigenvalues: In the case where there are, say, r repeated eigenvalues, then a linearly independent set of r eigenvectors for those eigenvalues exists, provided that

$$\text{rank}(\boldsymbol{A} - \lambda \boldsymbol{I}) = n - r. \tag{7}$$

Example 2

Show that $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ satisfies Eq. (7).

Diagonalizability and Eigenvalues

Distinct eigenvalues:

Property 3

Suppose that $\{\lambda_1, \cdots, \lambda_k\}$, $k \leq n$, *is a set of distinct eigenvalues, i.e.,* $\lambda_i \neq \lambda_j$ *for* $i \neq j$, $\forall i, j \in \{1, 2, \cdots, k\}$. *Then, the corresponding set of eigenvectors* $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ *is linearly independent.*

- If all the eigenvalues of $\boldsymbol{A}$ are distinct, then, by Theorem 1, $\boldsymbol{A}$ is diagonalizable.

Repeated eigenvalues: In the case where there are, say, r repeated eigenvalues, then a linearly independent set of r eigenvectors for those eigenvalues exists, provided that

$$\text{rank}(\boldsymbol{A} - \lambda \boldsymbol{I}) = n - r. \tag{7}$$

Example 2

Show that $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ satisfies Eq. (7).

Solution: $|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0 \Longrightarrow \lambda_1 = 1$ $(c_1 \boldsymbol{e}_1)$ and $\lambda_2 = \lambda_3 = 0$ $(c_2 \boldsymbol{e}_2 + c_3 \boldsymbol{e}_3)$.

Table of Contents

Multiplicities

For $\lambda \in \sigma(\boldsymbol{A}) = \{\lambda_1, \lambda_2, \cdots, \lambda_s\}$, we define

- The **algebraic multiplicity** of λ is the number of times it is repeated as a root of the characteristic polynomial. In other words, $\mathrm{alg\ mult}_{\boldsymbol{A}}(\lambda_i) = a_i$ if and only if $(x-\lambda_1)^{a_1}\cdots(x-\lambda_s)^{a_s} = 0$ is the characteristic equation for $\boldsymbol{A}$.

Multiplicities

For $\lambda \in \sigma(\boldsymbol{A}) = \{\lambda_1, \lambda_2, \cdots, \lambda_s\}$, we define

- The **algebraic multiplicity** of λ is the number of times it is repeated as a root of the characteristic polynomial. In other words, $\mathrm{alg\ mult}_{\boldsymbol{A}}(\lambda_i) = a_i$ if and only if $(x-\lambda_1)^{a_1}\cdots(x-\lambda_s)^{a_s} = 0$ is the characteristic equation for $\boldsymbol{A}$.
- When $\mathrm{alg\ mult}_{\boldsymbol{A}}(\lambda) = 1$, λ is called a **simple eigenvalue**.

Multiplicities

For $\lambda \in \sigma(\boldsymbol{A}) = \{\lambda_1, \lambda_2, \cdots, \lambda_s\}$, we define

- The **algebraic multiplicity** of λ is the number of times it is repeated as a root of the characteristic polynomial. In other words, $\text{alg mult}_{\boldsymbol{A}}(\lambda_i) = a_i$ if and only if $(x-\lambda_1)^{a_1}\cdots(x-\lambda_s)^{a_s} = 0$ is the characteristic equation for $\boldsymbol{A}$.
- When $\text{alg mult}_{\boldsymbol{A}}(\lambda) = 1$, λ is called a **simple eigenvalue**.
- The **geometric multiplicity** of λ is $\dim \mathcal{N}(\boldsymbol{A} - \lambda\boldsymbol{I})$. In other words, $\text{geo mult}_{\boldsymbol{A}}(\lambda)$ is the maximal number of linearly independent eigenvectors associated with λ.

Multiplicities

For $\lambda \in \sigma(\boldsymbol{A}) = \{\lambda_1, \lambda_2, \cdots, \lambda_s\}$, we define

- The **algebraic multiplicity** of λ is the number of times it is repeated as a root of the characteristic polynomial. In other words, $\text{alg mult}_{\boldsymbol{A}}(\lambda_i) = a_i$ if and only if $(x-\lambda_1)^{a_1} \cdots (x-\lambda_s)^{a_s} = 0$ is the characteristic equation for $\boldsymbol{A}$.
- When $\text{alg mult}_{\boldsymbol{A}}(\lambda) = 1$, λ is called a **simple eigenvalue**.
- The **geometric multiplicity** of λ is $\dim \mathcal{N}(\boldsymbol{A} - \lambda \boldsymbol{I})$. In other words, $\text{geo mult}_{\boldsymbol{A}}(\lambda)$ is the maximal number of linearly independent eigenvectors associated with λ.
- Eigenvalues such that $\text{alg mult}_{\boldsymbol{A}}(\lambda) = \text{geo mult}_{\boldsymbol{A}}(\lambda)$ are called **semisimple eigenvalues** of $\boldsymbol{A}$.

Diagonalizability and Multiplicities

A matrix $\boldsymbol{A}$ is **diagonalizable** if and only if

$$\text{alg mult}_{\boldsymbol{A}}(\lambda) = \text{geo mult}_{\boldsymbol{A}}(\lambda), \tag{8}$$

for each $\lambda \in \sigma(\boldsymbol{A})$, – i.e., if and only if every eigenvalue is semisimple.

Diagonalizability and Multiplicities

A matrix $\boldsymbol{A}$ is **diagonalizable** if and only if

$$\text{alg mult}_{\boldsymbol{A}}(\lambda) = \text{geo mult}_{\boldsymbol{A}}(\lambda), \tag{8}$$

for each $\lambda \in \sigma(\boldsymbol{A})$, – i.e., if and only if every eigenvalue is semisimple.

Special case: If no eigenvalue of $\boldsymbol{A}$ is repeated, then $\boldsymbol{A}$ is diagonalizable. However, the converse is not true.

Diagonalizability and Multiplicities

A matrix $\boldsymbol{A}$ is **diagonalizable** if and only if

$$\text{alg mult}_{\boldsymbol{A}}(\lambda) = \text{geo mult}_{\boldsymbol{A}}(\lambda), \tag{8}$$

for each $\lambda \in \sigma(\boldsymbol{A})$, – i.e., if and only if every eigenvalue is semisimple.

Special case: If no eigenvalue of $\boldsymbol{A}$ is repeated, then $\boldsymbol{A}$ is diagonalizable. However, the converse is not true.

Remark 4

The criteria described above is compatible with Theorem 1.

Theorem 1: A matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ is diagonalizable if and only if there is a set of n linearly independent vectors, each of which is an eigenvector of $\boldsymbol{A}$.

Eigendecomposition

Eigendecomposition: Given $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ being symmetric, and assuming that all eigenvectors of $\boldsymbol{A}$ are linearly independent, $\boldsymbol{A}$ can be decomposed as

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H, \tag{9}$$

where $\boldsymbol{V} = \left[\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\right] \in \mathbb{C}^{n\times n}$ (being an orthonormal matrix) and $\boldsymbol{\Lambda} = \mathrm{Diag}(\lambda_1, \cdots, \lambda_n) \in \mathbb{C}^{n\times n}$ (being a diagonal matrix) are called the **eigenvector matrix** and **eigenvalue matrix** of $\boldsymbol{A}$, respectively.

Eigendecomposition

Eigendecomposition: Given $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ being symmetric, and assuming that all eigenvectors of $\boldsymbol{A}$ are linearly independent, $\boldsymbol{A}$ can be decomposed as

$$\boldsymbol{A} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^H, \tag{9}$$

where $\boldsymbol{V} = \left[\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\right] \in \mathbb{C}^{n \times n}$ (being an orthonormal matrix) and $\boldsymbol{\Lambda} = \mathrm{Diag}(\lambda_1, \cdots, \lambda_n) \in \mathbb{C}^{n \times n}$ (being a diagonal matrix) are called the **eigenvector matrix** and **eigenvalue matrix** of $\boldsymbol{A}$, respectively.

- The above eigendecomposition is a direct consequence of Theorem 1.
- Eigendecomposition is a.k.a. **spectral decomposition**
  - Spectrum of $\boldsymbol{A} \in \mathbb{C}^{n \times n}$: $\sigma(\boldsymbol{A}) \triangleq \{\lambda_1, \lambda_2, \cdots, \lambda_s\}$, $s \leq n$.
  - Spectral radius $\triangleq \max\{|\lambda_1|, \cdots, |\lambda_n|\}$.
- Not every $\boldsymbol{A}$ can have eigendecomposition, e.g., the defective matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ cannot be diagonalized (algebraic multiplicity vs. geometric multiplicity).
- In some cases, the decomposition exists but involves complex rather than real numbers.

Gershgorin Circles Theorem for Estimating Eigenvalues

- The eigenvalues of $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ are contained the union $\mathcal{G}_r$ of the n **Gershgorin circles** defined by

$$|\lambda - a_{ii}| \leq \sum_{j\neq i}^{n} |a_{ij}|, \text{ for } i = 1, 2, \cdots, n \tag{10}$$

- Furthermore, if a union $\mathcal{U}$ of k Gershgorin circles does not touch any of the other $n-k$ circles, then there are exactly k eigenvalues (counting multiplicities) in the circles in $\mathcal{U}$.
- Since $\sigma(\boldsymbol{A}^T) = \sigma(\boldsymbol{A})$, the deleted absolute row sums in (10) can be replaced by deleted absolute column sums, so the eigenvalues of $\boldsymbol{A}$ are also contained in the union $\mathcal{G}_c$ of the circles defined by

$$|\lambda - a_{jj}| \leq \sum_{i\neq j}^{n} |a_{ij}| \text{ for } j = 1, 2, \cdots, n \tag{11}$$

- Combining (10) and (11) means that the eigenvalues of $\boldsymbol{A}$ are contained in the intersection $\mathcal{G}_r \cap \mathcal{G}_c$.

Gershgorin Circles Theorem (cont'd)

Example 3

$$\boldsymbol{A} = \begin{bmatrix} 10 & -1 & 0 & 1 \\ 0.2 & 8 & 0.2 & 0.2 \\ 1 & 1 & 2 & 1 \\ -1 & -1 & -1 & -11 \end{bmatrix}$$

Solution:

- Four circles: $C(10, 2)$, $C(8, 0.6)$, $C(2, 3)$, and $C(-11, 3)$.
- To improve the accuracy, we can use the last two circles pertaining to the columns of $\boldsymbol{A}$: $C(2, 1.2)$ and $C(-11, 2.2)$.
- The eigenvalues are 9.8218, 8.1478, 1.8995, and -10.86.

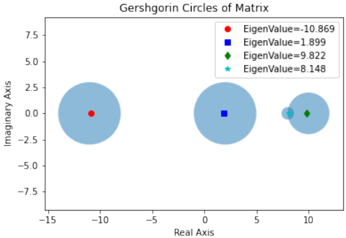
Gershgorin Circles Theorem (cont'd)

Example 3

$$\boldsymbol{A} = \begin{bmatrix} 10 & -1 & 0 & 1 \\ 0.2 & 8 & 0.2 & 0.2 \\ 1 & 1 & 2 & 1 \\ -1 & -1 & -1 & -11 \end{bmatrix}$$

Solution:

- Four circles: $C(10,2)$, $C(8,0.6)$, $C(2,3)$, and $C(-11,3)$.
- To improve the accuracy, we can use the last two circles pertaining to the columns of $\boldsymbol{A}$: $C(2,1.2)$ and $C(-11,2.2)$.
- The eigenvalues are 9.8218, 8.1478, 1.8995, and -10.86.



Orthogonality

Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ (either real or complex) are said to be **orthogonal** if

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0. \tag{12}$$

- A set of vectors $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ is called **orthogonal** if $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = 0$ for all $i \neq j$.
- A set of vectors $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ is called **orthonormal** if $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = 0$ for all $i \neq j$ and $\|\boldsymbol{x}_i\|_2^2 = 1$ for all i.
- Property: an orthogonal set of vectors is linearly independent.

Unitary and Orthogonal Matrices

1) A matrix $\boldsymbol{U} \in \mathbb{C}^{n \times n}$ is called **unitary** if

$$\boldsymbol{U}^H \boldsymbol{U} = \boldsymbol{I},$$

and a matrix $\boldsymbol{U} \in \mathbb{R}^{n \times n}$ is called **orthogonal** if $\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{I}$.

2) A unitary (orthogonal) matrix is a matrix where its columns form an orthonormal set of vectors.

3) Some properties:
    - $\boldsymbol{U}^{-1} = \boldsymbol{U}^H$.
    - $\boldsymbol{U}\boldsymbol{U}^H = \boldsymbol{I}$.
    - The rows of $\boldsymbol{U}$ form an orthonormal set of vectors.

Table of Contents

Symmetric and Hermitian Matrices

1) A square matrix $\boldsymbol{A}$ is called **symmetric** if

$$\boldsymbol{A} = \boldsymbol{A}^T, \tag{13}$$

or equivalently, $a_{ij} = a_{ji}$ for any i, j.

2) A square matrix $\boldsymbol{A}$ is called **Hermitian** if

$$\boldsymbol{A} = \boldsymbol{A}^H, \tag{14}$$

or equivalently, $a_{ij} = a_{ji}^*$ for any i, j.

Symmetric and Hermitian Matrices (cont'd)

Remark 5

1) *Usually, real symmetric matrices and complex Hermitian matrices are considered.*
2) *Complex Hermitian matrices subsume real symmetric matrices (a real symmetric matrix lies in the set of all complex Hermitian matrices, but the converse is not true).*
3) *The eigendecomposition of real symmetric and complex Hermitian matrices deserve particular attention, and have relevance to many engineering applications.*

Eigendcomposition for Hermitian Matrices - Simple Version

Property 4

The eigenvalues of a Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ are real.

Property 5

Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ be a Hermitian matrix, and suppose that all its eigenvalues are distinct. Then, the eigenvectors of $\boldsymbol{A}$ are mutually orthogonal.

Eigendcomposition for Hermitian Matrices - Simple Version

Property 4

The eigenvalues of a Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *are real.*

Property 5

Let $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *be a Hermitian matrix, and suppose that all its eigenvalues are distinct. Then, the eigenvectors of* $\boldsymbol{A}$ *are mutually orthogonal.*

By Properties 4-5, any Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ with distinct eigenvalues can be eigen-decomposed as

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H, \quad (\text{or } \boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T \text{ for real symmetric } \boldsymbol{A}) \tag{15}$$

where $\boldsymbol{V} \in \mathbb{C}^{n\times n}$ is the eigenvector matrix and is unitary; $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ is the eigenvalue matrix.

Eigendcomposition for Hermitian Matrices - Simple Version

Property 4

The eigenvalues of a Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ are real.

Property 5

Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ be a Hermitian matrix, and suppose that all its eigenvalues are distinct. Then, the eigenvectors of $\boldsymbol{A}$ are mutually orthogonal.

By Properties 4-5, any Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ with distinct eigenvalues can be eigen-decomposed as

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H, \quad (\text{or } \boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T \text{ for real symmetric } \boldsymbol{A}) \tag{15}$$

where $\boldsymbol{V} \in \mathbb{C}^{n \times n}$ is the eigenvector matrix and is unitary; $\boldsymbol{\Lambda} \in \mathbb{R}^{n \times n}$ is the eigenvalue matrix.

Question: Does any Hermitian matrix admit the eigendecomposition above?

Eigendcomposition for Hermitian Matrices - Simple Version

Property 4

The eigenvalues of a Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ are real.

Property 5

Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ be a Hermitian matrix, and suppose that all its eigenvalues are distinct. Then, the eigenvectors of $\boldsymbol{A}$ are mutually orthogonal.

By Properties 4-5, any Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ with distinct eigenvalues can be eigen-decomposed as

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H, \quad (\text{or } \boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T \text{ for real symmetric } \boldsymbol{A}) \tag{15}$$

where $\boldsymbol{V} \in \mathbb{C}^{n \times n}$ is the eigenvector matrix and is unitary; $\boldsymbol{\Lambda} \in \mathbb{R}^{n \times n}$ is the eigenvalue matrix.

Question: Does any Hermitian matrix admit the eigendecomposition above?

Answer: Yes, by the Schur triangularization to be shown below.

Eigendcomposition for Hermitian Matrices - General Version

Theorem 4 (Schur triangularization)

Given $\boldsymbol{A} \in \mathbb{C}^{n \times n}$, *there exists a unitary matrix* $\boldsymbol{U} \in \mathbb{C}^{n \times n}$ *such that*

$$\boldsymbol{U}^H \boldsymbol{A} \boldsymbol{U} = \boldsymbol{T}, \tag{16}$$

where $\boldsymbol{T}$ *is* **an upper triangular matrix** *with* $\mathrm{diag}(\boldsymbol{T}) = \left[\lambda_1, \cdots, \lambda_n\right]^T$. *If* $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ *is Hermitian, it is evident that*

$$\boldsymbol{T}^H = (\boldsymbol{U}^H \boldsymbol{A} \boldsymbol{U})^H = \boldsymbol{U}^H \boldsymbol{A}^H \boldsymbol{U} = \boldsymbol{U}^H \boldsymbol{A} \boldsymbol{U} = \boldsymbol{T}, \tag{17}$$

that is, $\boldsymbol{T}$ *is diagonal and* $\boldsymbol{T} = \boldsymbol{\Lambda}$.

Eigendcomposition for Hermitian Matrices - General Version

Theorem 4 (Schur triangularization)

Given $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, *there exists a unitary matrix* $\boldsymbol{U} \in \mathbb{C}^{n\times n}$ *such that*

$$\boldsymbol{U}^H \boldsymbol{A}\boldsymbol{U} = \boldsymbol{T}, \tag{16}$$

where $\boldsymbol{T}$ *is* **an upper triangular matrix** *with* $\operatorname{diag}(\boldsymbol{T}) = \left[\lambda_1, \cdots, \lambda_n\right]^T$. *If* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *is Hermitian, it is evident that*

$$\boldsymbol{T}^H = (\boldsymbol{U}^H \boldsymbol{A}\boldsymbol{U})^H = \boldsymbol{U}^H \boldsymbol{A}^H \boldsymbol{U} = \boldsymbol{U}^H \boldsymbol{A}\boldsymbol{U} = \boldsymbol{T}, \tag{17}$$

that is, $\boldsymbol{T}$ *is diagonal and* $\boldsymbol{T} = \boldsymbol{\Lambda}$.

Theorem 5

A Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *can always be diagonalized as*

$$\boldsymbol{V}^H \boldsymbol{A}\boldsymbol{V} = \boldsymbol{\Lambda} \ \ (\textit{or } \boldsymbol{V}^T \boldsymbol{A}\boldsymbol{V} = \boldsymbol{\Lambda} \textit{ for real symmetric } \boldsymbol{A}), \tag{18}$$

where $\boldsymbol{V} \in \mathbb{C}^{n\times n}$ *and* $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ *are the eigenvector and eigenvalue matrices of* $\boldsymbol{A}$, *resp.*

(Nice) Results from Hermitian Eigendecomposition

Let $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ be a Hermitian matrix.

- $\boldsymbol{A}^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^H$, assuming a nonsingular $\boldsymbol{A}$.
  Note: $\boldsymbol{\Lambda}^{-1} = \mathrm{Diag}\left(\frac{1}{\lambda_1}, \cdots, \frac{1}{\lambda_n}\right)$.
- rank($\boldsymbol{A}$) is the number of nonzero eigenvalues.
- det($\boldsymbol{A}$) $= \prod_{i=1}^{n} \lambda_i$.
- tr($\boldsymbol{A}$) $= \sum_{i=1}^{n} \lambda_i$.

Note: The last two equalites also hold for general $\boldsymbol{A}$, although they can be easily proven in the Hermitian case via eigendecomposition.

Table of Contents

Matrix Norms

The definition of a matrix norm is the same as that of a vector norm. A **matrix norm** is a function $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ that satisfies

1) $\|\boldsymbol{X}\| \geq 0$ for all $\boldsymbol{X} \in \mathbb{C}^{m\times n}$; (non-negetivity)
2) $\|\boldsymbol{X}\| = 0$ if and only if $\boldsymbol{X} = \mathbf{0}$; (separate points)
3) $\|\boldsymbol{X} + \boldsymbol{Y}\| \leq \|\boldsymbol{X}\| + \|\boldsymbol{Y}\|$ for any $\boldsymbol{X}$, $\boldsymbol{Y} \in \mathbb{C}^{m\times n}$; (triangle inequality)
4) $\|\alpha\boldsymbol{X}\| = |\alpha|\,\|\boldsymbol{X}\|$ for any $\alpha \in \mathbb{C}, \boldsymbol{X} \in \mathbb{C}^{m\times n}$; (absolute homogeneity)

Matrix Norms

The definition of a matrix norm is the same as that of a vector norm. A **matrix norm** is a function $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ that satisfies

1) $\|\boldsymbol{X}\| \geq 0$ for all $\boldsymbol{X} \in \mathbb{C}^{m\times n}$; (non-negetivity)
2) $\|\boldsymbol{X}\| = 0$ if and only if $\boldsymbol{X} = \mathbf{0}$; (separate points)
3) $\|\boldsymbol{X}+\boldsymbol{Y}\| \leq \|\boldsymbol{X}\| + \|\boldsymbol{Y}\|$ for any $\boldsymbol{X}$, $\boldsymbol{Y} \in \mathbb{C}^{m\times n}$; (triangle inequality)
4) $\|\alpha\boldsymbol{X}\| = |\alpha|\,\|\boldsymbol{X}\|$ for any $\alpha \in \mathbb{C}, \boldsymbol{X} \in \mathbb{C}^{m\times n}$; (absolute homogeneity)

Example 6 (Frobenius norm)

$$\|\boldsymbol{A}\|_F \triangleq \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}. \tag{19}$$

- A (straight) matrix version of the vector ℓ_2-norm.
- $\|\boldsymbol{A}\|_F = \left[\mathrm{tr}\left(\boldsymbol{A}^H\boldsymbol{A}\right)\right]^{1/2}$.

Case I: Matrix Norms Induced by Vector Norms

A vector norm that is defined on $\mathbb{C}^l$ for $l = m, n$ induces a matrix norm on $\mathbb{C}^{m\times n}$ by setting

$$\|\boldsymbol{A}\| \triangleq \sup_{\boldsymbol{x}\neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|, \ \forall \boldsymbol{A} \in \mathbb{C}^{m\times n}, \boldsymbol{x} \in \mathbb{C}^{n\times 1}. \tag{20}$$

- It's apparent that an induced matrix norm is compatible with its underlying vector norm in the sense that

$$\|\boldsymbol{A}\boldsymbol{x}\| \leq \|\boldsymbol{A}\| \ \|\boldsymbol{x}\|. \tag{21}$$

- When $\boldsymbol{A}$ is nonsingular,

$$\min_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\| = \frac{1}{\|\boldsymbol{A}^{-1}\|}. \tag{22}$$

Case I: Matrix Norms Induced by Vector Norms

A vector norm that is defined on $\mathbb{C}^l$ for $l = m, n$ induces a matrix norm on $\mathbb{C}^{m\times n}$ by setting

$$\|\boldsymbol{A}\| \triangleq \sup_{\boldsymbol{x}\neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|,\ \forall \boldsymbol{A} \in \mathbb{C}^{m\times n}, \boldsymbol{x} \in \mathbb{C}^{n\times 1}. \tag{20}$$

- It's apparent that an induced matrix norm is compatible with its underlying vector norm in the sense that

$$\|\boldsymbol{A}\boldsymbol{x}\| \leq \|\boldsymbol{A}\|\ \|\boldsymbol{x}\|. \tag{21}$$

- When $\boldsymbol{A}$ is nonsingular,

$$\min_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\| = \frac{1}{\|\boldsymbol{A}^{-1}\|}. \tag{22}$$

Note: An induced norm of $\boldsymbol{A}$ represents the maximum extent to which a vector on the unit sphere can be **stretched** by $\boldsymbol{A}$, and $\frac{1}{\|\boldsymbol{A}^{-1}\|}$ measures the extent to which a nonsingular matrix $\boldsymbol{A}$ can **shrink** vectors on the unit sphere.

Case I: Matrix Norms Induced by Vector Norms (cont'd)

In summary, it is not hard to prove that

$$\frac{\|\boldsymbol{x}\|}{\|\boldsymbol{A}^{-1}\|} \leq \|\boldsymbol{A}\boldsymbol{x}\| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{x}\|. \tag{23}$$

Case I: Matrix Norms Induced by Vector Norms (cont'd)

In summary, it is not hard to prove that

$$\frac{\|\boldsymbol{x}\|}{\|\boldsymbol{A}^{-1}\|} \leq \|\boldsymbol{A}\boldsymbol{x}\| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{x}\|. \tag{23}$$

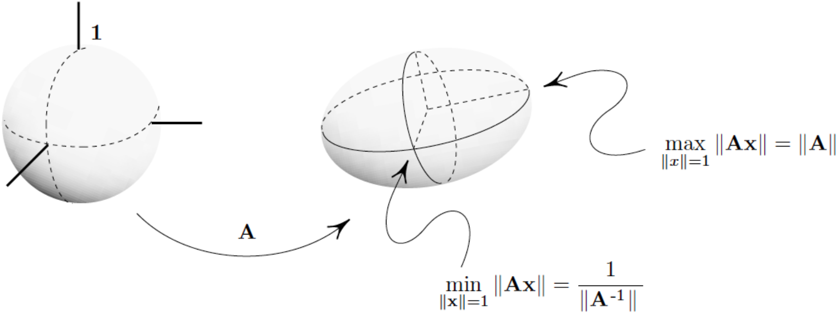


Figure 3: The induced matrix 2-norm in $\mathbb{R}^3$.

Special Case: Matrix p-Norms

In particular, if the p-norm for vectors ($p \geq 1$) is used for both spaces $\mathbb{C}^l$, $l = m, n$, then the corresponding **Matrix p-norms** is defined as

$$\|\boldsymbol{A}\|_p \triangleq \sup_{\boldsymbol{x} \neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p} = \max_{\|\boldsymbol{x}\|_p = 1} \|\boldsymbol{A}\boldsymbol{x}\|_p. \tag{24}$$

1) 1-norm

$$\|\boldsymbol{A}\|_1 \triangleq \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}| \quad \text{(maximum column sum)} \tag{25}$$

2) 2-norm (a.k.a. spectral norm)

$$\|\boldsymbol{A}\|_2 = \sqrt{\lambda_{\max}(\boldsymbol{A}^H \boldsymbol{A})} = \sigma_1, \tag{26}$$

where $\lambda_{\max}(\boldsymbol{X})$ denotes the largest eigenvalue of $\boldsymbol{X}$, and σ_1 is the maximum singular value of $\boldsymbol{A}$ (Theorem 2.3.1, p.73, *Matrix Computations*).

3) ∞-norm

$$\|\boldsymbol{A}\|_\infty \triangleq \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}| \quad \text{(maximum row sum)} \tag{27}$$

Properties of Matrix p-Norms

Remark 6 (Some properties of matrix norms)

- $\frac{1}{\sqrt{m}}\|\boldsymbol{A}\|_1 \leq \|\boldsymbol{A}\|_2 \leq \sqrt{n}\ \|\boldsymbol{A}\|_1$
- $\|\boldsymbol{A}\|_2 \leq \|\boldsymbol{A}\|_F \leq \sqrt{\min\{m,n\}}\ \|\boldsymbol{A}\|_2$
- $\frac{1}{\sqrt{n}}\|\boldsymbol{A}\|_\infty \leq \|\boldsymbol{A}\|_2 \leq \sqrt{m}\ \|\boldsymbol{A}\|_\infty$
- $\|\boldsymbol{A}\|_2 \leq \sqrt{\|\boldsymbol{A}\|_1\ \|\boldsymbol{A}\|_\infty}$
- $\|\boldsymbol{A}\boldsymbol{x}\|_p \leq \|\boldsymbol{A}\|_p\,\|\boldsymbol{x}\|_p$
- $\|\boldsymbol{A}\boldsymbol{B}\|_p \leq \|\boldsymbol{A}\|_p\ \|\boldsymbol{B}\|_p$ *(Hölder inequality)*
- *Let $\boldsymbol{Q}$ and $\boldsymbol{Z}$ be unitary matrices of appropriate sizes, we have*

$$\|\boldsymbol{Q}\boldsymbol{A}\boldsymbol{Z}\|_2 = \|\boldsymbol{A}\|_2, \tag{28}$$

$$\|\boldsymbol{Q}\boldsymbol{A}\boldsymbol{Z}\|_F = \|\boldsymbol{A}\|_F. \tag{29}$$

Case II: "Entrywise" Matrix Norms

These norms treat an $m \times n$ matrix as a vector of size $m \cdot n$ and use one of the familiar vector norms. For example, using the p-norm for vectors, $p \geq 1$, we get:

$$\|\boldsymbol{A}\|_p = \|\text{vec}(\boldsymbol{A})\|_p = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^p \right)^{1/p}. \tag{30}$$

The special case $p = 2$ is the Frobenius norm, and $p = \infty$ is the maximum norm.

- $L_{2,1}$ Norm
  Let $(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ be the columns of matrix $\boldsymbol{A}$. The $L_{2,1}$ norm is the sum of the Euclidean norms of the columns of the matrix:

$$\|\boldsymbol{A}\|_{2,1} = \sum_{j=1}^{n} \|\boldsymbol{a}_j\|_2 = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} |a_{ij}|^2 \right)^{1/2}. \tag{31}$$

[3] [R3] Z. Gong, P. Zhong, and W. Hu "Diversity in Machine Learning," *IEEE Access*, vol. 7, pp. 64323–64350, May 2019.

Case II: "Entrywise" Matrix Norms

These norms treat an $m \times n$ matrix as a vector of size $m \cdot n$ and use one of the familiar vector norms. For example, using the p-norm for vectors, $p \geq 1$, we get:

$$\|\boldsymbol{A}\|_p = \|\text{vec}(\boldsymbol{A})\|_p = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^p \right)^{1/p}. \tag{30}$$

The special case $p = 2$ is the Frobenius norm, and $p = \infty$ is the maximum norm.

- $L_{2,1}$ Norm

  Let $(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ be the columns of matrix $\boldsymbol{A}$. The $L_{2,1}$ norm is the sum of the Euclidean norms of the columns of the matrix:

$$\|\boldsymbol{A}\|_{2,1} = \sum_{j=1}^{n} \|\boldsymbol{a}_j\|_2 = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} |a_{ij}|^2 \right)^{1/2}. \tag{31}$$

  Application: The $L_{2,1}$ norm as an error function is more robust since the error for each data point (a column) is not squared. It is widely used in robust data analysis and sparse coding since it can take advantage of the group-wise correlation and obtain a group-wise sparse representation of $\boldsymbol{a}_j$, $j = 1, 2, \cdots, n$, see e.g., Section IV-A-3-e of [R3].[3]

[3] [R3] Z. Gong, P. Zhong, and W. Hu "Diversity in Machine Learning," *IEEE Access*, vol. 7, pp. 64323–64350, May 2019.

Case II: "Entrywise" Matrix Norms (cont'd)

- $L_{p,q}$ Norm

  The $L_{2,1}$ norm defined in (31) can be generalized to the $L_{p,q}$ norm, $p, q \geq 1$, defined by

$$\|\boldsymbol{A}\|_{p,q} = \left(\sum_{j=1}^{n} \left(\sum_{i=1}^{m} |a_{ij}|^p \right)^{q/p} \right)^{1/q}. \tag{32}$$

- Frobenius Norm (i.e., $p = q = 2$ in (32))

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\operatorname{tr}\left(\boldsymbol{A}^T \boldsymbol{A}\right)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(\boldsymbol{A})}. \tag{33}$$

- Max Norm (i.e., $p = q = \infty$ in (32))

$$\|\boldsymbol{A}\|_{\max} = \max_{i,j} |a_{ij}|. \tag{34}$$

Case III: Schatten Norms

The Schatten p-norms arise when applying the p-norm to the vector of singular values of a matrix:

$$\|\boldsymbol{A}\|_p = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i^p(\boldsymbol{A}) \right)^{1/p}. \tag{35}$$

- Nuclear Norm (a.k.a. Schatten 1-norm or the trace norm)

$$\|\boldsymbol{A}\|_* = \operatorname{tr}\left(\sqrt{\boldsymbol{A}^H \boldsymbol{A}}\right) = \sum_{i=1}^{\min\{m,n\}} \sigma_i(\boldsymbol{A}). \tag{36}$$

Note: Here $\sqrt{\boldsymbol{A}^H \boldsymbol{A}}$ denotes a positive semidefinite matrix $\boldsymbol{B}$ such that $\boldsymbol{B}\boldsymbol{B} = \boldsymbol{A}^H \boldsymbol{A}$. More precisely, since $\boldsymbol{A}^H \boldsymbol{A}$ is a positive semidefinite matrix, its square root is well-defined.

[4] [R4] Y. Hu, D. Zhang, H. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35. no. 9, pp. 2117–2130, 2013.

Case III: Schatten Norms

The Schatten p-norms arise when applying the p-norm to the vector of singular values of a matrix:

$$\|\boldsymbol{A}\|_p = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i^p(\boldsymbol{A}) \right)^{1/p}. \tag{35}$$

- Nuclear Norm (a.k.a. Schatten 1-norm or the trace norm)

$$\|\boldsymbol{A}\|_* = \mathrm{tr}\left(\sqrt{\boldsymbol{A}^H\boldsymbol{A}}\right) = \sum_{i=1}^{\min\{m,n\}} \sigma_i(\boldsymbol{A}). \tag{36}$$

Note: Here $\sqrt{\boldsymbol{A}^H\boldsymbol{A}}$ denotes a positive semidefinite matrix $\boldsymbol{B}$ such that $\boldsymbol{B}\boldsymbol{B} = \boldsymbol{A}^H\boldsymbol{A}$. More precisely, since $\boldsymbol{A}^H\boldsymbol{A}$ is a positive semidefinite matrix, its square root is well-defined.

Application: In general, $\|\boldsymbol{A}\|_*$ is a convex approximation to the rank of $\boldsymbol{A}$, like ℓ_1-norm is a convex approximation to ℓ_0-norm of a vector. For example, an application of nuclear norm refers to [R4].[4]

[4] [R4] Y. Hu, D. Zhang, H. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35. no. 9, pp. 2117–2130, 2013.

Thank you for your attention!

Kai Lu

E-mail: lukai86@mail.sysu.edu.cn

Web: http://seit.sysu.edu.cn/teacher/1801