

# **Study Notes of Matrix and Tensor**

Pei Zhong

April 3, 2024

# Contents

<b>1</b>	<b>Matrix Algebra</b>	<b>3</b>
1.1	Notations and definitions . . . . .	3
1.2	Transposition and conjugate transposition . . . . .	4
1.3	Vector outer product and vectorization . . . . .	5
1.3.1	Vector outer product . . . . .	5
1.3.2	Vectorization . . . . .	5
1.4	Vector inner product and orthogonality . . . . .	5
1.4.1	Inner product . . . . .	5
1.4.2	Orthogonality . . . . .	6
1.5	Vector Norms . . . . .	6
1.5.1	Vector 2-norm . . . . .	6
1.5.2	Vector 1-norm . . . . .	7
1.5.3	Vector $\infty$ -norm . . . . .	7
1.5.4	Vector $p$ -norm . . . . .	7
1.6	Matrix Norms . . . . .	8
1.6.1	Frobenius norm . . . . .	8
1.6.2	Induced matrix norms . . . . .	8
1.7	Matrix multiplication . . . . .	9
1.8	Matrix trace, Matrix inner product . . . . .	10
1.8.1	Definition and properties of the trace . . . . .	10
1.9	Eigenvalues and eigenvectors . . . . .	10
1.10	Generalized inverse and pseudoinverse . . . . .	14
1.10.1	Generalized inverse . . . . .	15
1.10.2	property of generalized inverse and pseudoinverse . . . . .	15
1.10.3	Pseudoinverse . . . . .	15
1.11	Reference . . . . .	16
<b>2</b>	<b>Matrix Decompositions</b>	<b>17</b>
2.1	SVD . . . . .	17

2.1.1	Singular value decomposition (SVD)	17
2.1.2	Geometric interpretation of SVD	20
2.1.3	Applications of Singular-Value Decomposition	20
2.2	Reference	20
<b>3</b>	<b>Hadamard, Kronecker and Khatri-Rao Products</b>	<b>21</b>
3.1	Partitioned matrices	21
3.2	Notation	22
3.3	Hadamard product	22
3.3.1	Definition and identities	22
3.4	Kronecker product	23
3.4.1	Kronecker product of vectors	23
3.4.2	Kronecker product of matrices	23
3.5	Reference	23
<b>4</b>	<b>Tensor Operations</b>	<b>25</b>
4.1	Notation	25
4.2	Notion of slice	25
4.3	Mode combination	25
4.4	Matricization	25
4.5	Multiplication operations	25
<b>5</b>	<b>Tensor Decomposition</b>	<b>26</b>

# Preface

The notes mainly refer to:

- From Algebraic Structures to Tensors
- Matrix and Tensor Decompositions in Signal Processing

# Chapter 1

## Matrix Algebra

### 1.1 Notations and definitions

Scalars, column vectors, matrices, and hypermatrices/tensors of order higher than two will be denoted by lowercase letters ( $a, b, \dots$ ), bold lowercase letters ( $\mathbf{a}, \mathbf{b}, \dots$ ), bold uppercase letters ( $\mathbf{A}, \mathbf{B}, \dots$ ), and calligraphic letters ( $\mathcal{A}, \mathcal{B}, \dots$ ) respectively.

A matrix  $\mathbf{A}$  of dimensions  $I \times J$ , with  $I$  and  $J \in \mathbb{N}^*$ , denoted by  $\mathbf{A}(I, J)$ , is an array of  $IJ$  elements stored in  $I$  rows and  $J$  columns; the elements belong to a field  $\mathbb{K}$ . Its  $i$ th row and  $j$ th column, denoted by  $A_{i\cdot}$  and  $A_{\cdot j}$ , respectively, are called  $i$ th row vector and  $j$ th column vector. The element located at the intersection of  $A_{i\cdot}$  and  $A_{\cdot j}$  is designated by  $a_{ij}$ . We will use the notation  $\mathbf{A} = (a_{ij})$ , with  $a_{ij} \in \mathbb{K}$ ,  $i \in \langle I \rangle = \{1, 2, \dots, I\}$  and  $j \in \langle J \rangle = \{1, 2, \dots, J\}$ .

A matrix  $A \in \mathbb{K}^{I \times J}$  is written in the form:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1J} \\ a_{21} & a_{22} & \dots & a_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1} & a_{I2} & \dots & a_{IJ} \end{pmatrix}$$

The special cases  $I = 1$  and  $J = 1$  correspond respectively to row vectors of dimension  $J$  and to column vectors of dimension  $I$ :

$$\mathbf{v} = (v_1 \quad v_2 \quad \dots \quad v_J) \in \mathbb{K}^{1 \times J}, \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_I \end{pmatrix} \in \mathbb{K}^{I \times 1}.$$

In the following, for column vectors,  $\mathbb{K}^I$  will be used instead of  $\mathbb{K}^{I \times 1}$ .

Identity matrix (i.e., a matrix with all 1's on the diagonal and 0's everywhere else) is denoted by  $\mathbf{E}$ .  $\mathbf{e}_i^{(I)}$  is the column vector of dimension  $I$ , in which element is equal to 1 at position  $i$  and 0s elsewhere.  $\mathbf{E}_{ij}^{I \times J}$  is the matrix of dimension  $I \times J$  in which element is equal to 1 at position  $(i, j)$  and 0s elsewhere.

## 1.2 Transposition and conjugate transposition

Definition 1.1: transpose and the conjugate transpose of a column vector

The transpose and the conjugate transpose (also called transconjugate) of a column vector

$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_I \end{pmatrix} \in \mathbb{C}^I$ , denoted by  $\mathbf{u}^T$  and  $\mathbf{u}^H$ , respectively, are the row vectors defined as:

$$\mathbf{u}^T = \begin{pmatrix} u_1 & u_2 & \vdots & u_I \end{pmatrix} \text{ and } \mathbf{u}^H = (u_1^* \ u_2^* \ \dots \ u_I^*),$$

where  $u_i^*$  is the conjugate of  $u_i$  also denoted by  $\bar{u}_i$ .

Definition 1.2: transpose and the conjugate transpose of a matrix

The transpose of  $\mathbf{A} \in \mathbb{K}^{I \times J}$  is the matrix denoted by  $\mathbf{A}^T$ , of dimensions  $J \times I$ , such that  $\mathbf{A}^T = (a_{ji})$ , with  $i \in \langle I \rangle$  and  $j \in \langle J \rangle$ . In the case of a complex matrix, the conjugate transpose, also known as Hermitian transpose and denoted by  $\mathbf{A}^H$ , is defined as:  $\mathbf{A}^H = (\mathbf{A}^*)^T = (\mathbf{A}^T)^* = (a_{ji}^*)$ , where  $\mathbf{A}^* = (a_{ij}^*)$  is the conjugate of  $\mathbf{A}$ .

**Remark.** In mathematics, the complex conjugate of a complex number is the number with an equal real part and an imaginary part equal in magnitude but opposite in sign. That is, if  $a$  and  $b$  are real numbers then the complex conjugate of  $a + ib$  is  $a - ib$ . The complex conjugate of  $z$  is often denoted as  $\bar{z}$  or  $z^*$ .

Proposition 1.1

The operations of transposition and conjugate transposition satisfy:

$$\begin{aligned} (\mathbf{A}^T)^T &= \mathbf{A}, (\mathbf{A}^H)^H = \mathbf{A} \\ (\mathbf{A} + \mathbf{B})^T &= \mathbf{A}^T + \mathbf{B}^T, (\mathbf{A} + \mathbf{B})^H = \mathbf{A}^H + \mathbf{B}^H, \\ (\alpha \mathbf{A})^T &= \alpha \mathbf{A}^T, (\alpha \mathbf{A})^H = \alpha^* \mathbf{A}^H, \end{aligned}$$

for any matrix  $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{I \times J}$  and any scalar  $\alpha \in \mathbb{C}$ .

**Remark.** By decomposing  $\mathbf{A}$  using its real and imaginary parts, we have:

$$\mathbf{A} = \text{Re}(\mathbf{A}) + i\text{Im}(\mathbf{A}) \Rightarrow \begin{cases} \mathbf{A}^T = (\text{Re}(\mathbf{A}))^T + i(\text{Im}(\mathbf{A}))^T \\ \mathbf{A}^H = (\text{Re}(\mathbf{A}))^H - i(\text{Im}(\mathbf{A}))^H \end{cases}$$

## 1.3 Vector outer product and vectorization

### 1.3.1 Vector outer product

The outer product of two vectors  $\mathbf{u} \in \mathbb{K}^I$  and  $\mathbf{v} \in \mathbb{K}^J$ , denoted  $\mathbf{u} \circ \mathbf{v}$ , gives a matrix  $\mathbf{A} \in \mathbb{K}^{I \times J}$  such that  $a_{ij} = (\mathbf{u} \circ \mathbf{v})_{ij} = u_i v_j$ , and therefore,  $\mathbf{u} \circ \mathbf{v} = \mathbf{u} \mathbf{v}^T = (u_i v_j)$ , with  $i \in \langle I \rangle, j \in \langle J \rangle$ .

#### Example 1.1

For  $I = 2, J = 3$ , we have:

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \circ \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} = \begin{pmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \end{pmatrix}.$$

### 1.3.2 Vectorization

A very widely used operation in matrix computation is vectorization which consists of stacking the columns of a matrix  $\mathbf{A} \in \mathbb{K}^{I \times J}$  on top of each other to form a column vector of dimension  $J I$ :

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{\cdot 1} & \mathbf{A}_{\cdot 2} & \dots & \mathbf{A}_{\cdot J} \end{pmatrix} \in \mathbb{K}^{I \times J} \Rightarrow \text{vec}(\mathbf{A}) = \begin{pmatrix} \mathbf{A}_{\cdot 1} \\ \mathbf{A}_{\cdot 2} \\ \vdots \\ \mathbf{A}_{\cdot J} \end{pmatrix} \in \mathbb{K}^{J I}.$$

This operation defines an isomorphism between the space  $\mathbb{K}^{J I}$  of vectors of dimension  $J I$  and the space  $\mathbb{K}^{I \times J}$  of matrices  $I \times J$ . Indeed, the canonical basis of  $\mathbb{K}^{J I}$ , denoted by  $\{\mathbf{e}_{(j-1)I+i}^{(J I)}\}$ , allows us to write  $\text{vec}(\mathbf{A})$  as:

$$\mathbf{A} = \sum_{i=1}^I \sum_{j=1}^J a_{ij} \mathbf{e}_i^{(I)} \circ \mathbf{e}_j^{(J)} \Rightarrow \text{vec}(\mathbf{A}) = \sum_{i=1}^I \sum_{j=1}^J a_{ij} \mathbf{e}_{(j-1)I+i}^{(J I)},$$

with  $\mathbf{e}_{(j-1)I+i}^{(J I)} = \text{vec}(\mathbf{e}_i^{(I)} \circ \mathbf{e}_j^{(J)}) = \text{vec}(\mathbf{e}_i^{(I)} (\mathbf{e}_j^{(J)})^T)$ .

**Remark.** Since the operator  $\text{vec}$  satisfies  $\text{vec}(\alpha \mathbf{A} + \beta \mathbf{B}) = \alpha \text{vec}(\mathbf{A}) + \beta \text{vec}(\mathbf{B})$  for all  $\alpha, \beta \in \mathbb{K}$ , it is linear.

## 1.4 Vector inner product and orthogonality

### 1.4.1 Inner product

In this section, we recall the definition of the inner product(also called dot product) of two vectors  $\mathbf{a}, \mathbf{b} \in \mathbb{K}^I$ .

## Definition 1.3

If  $\mathbb{K} = \mathbb{R}$ , the vector inner product is defined as:

$$\langle \cdot, \cdot \rangle : \mathbb{R}^I \times \mathbb{R}^I \rightarrow \mathbb{R}$$

$$(\mathbf{a}, \mathbf{b}) \mapsto \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{b} = \sum_{i=1}^I a_i b_i.$$

In  $\mathbb{C}^I$ , the definition of the vector inner product is given by:

$$\langle \cdot, \cdot \rangle : \mathbb{C}^I \times \mathbb{C}^I \rightarrow \mathbb{C}$$

$$(\mathbf{a}, \mathbf{b}) \mapsto \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^H \mathbf{b} = \sum_{i=1}^I a_i^* b_i.$$

**Remark.** Whether  $\mathbb{K} = \mathbb{C}$  or  $\mathbb{R}$ ,  $\langle \mathbf{a}, \mathbf{a} \rangle \in \mathbb{R}$ .

### 1.4.2 Orthogonality

## Definition 1.4

Two vectors  $\mathbf{a}$  and  $\mathbf{b}$  of  $\mathbb{K}^I$  are said to be orthogonal if and only if  $\langle \mathbf{a}, \mathbf{b} \rangle = 0$ .

## 1.5 Vector Norms

## Definition 1.5

Let  $v : \mathbb{C}^n \rightarrow \mathbb{R}$ . Then  $v$  is a norm if for all  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$

- $\mathbf{x} \neq 0 \Rightarrow v(\mathbf{x}) > 0$ ,
- $v(\alpha \mathbf{x}) = |\alpha| v(\mathbf{x})$ , and
- $v(\mathbf{x} + \mathbf{y}) \leq v(\mathbf{x}) + v(\mathbf{y})$

**Remark.** often we will use  $\|\cdot\|$  to denote a vector norm.

### 1.5.1 Vector 2-norm

## Definition 1.6

The vector 2-norm  $\|\cdot\|_2 : \mathbb{C}^n \rightarrow \mathbb{R}$  is defined by

$$\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\mathbf{x}^H \mathbf{x}} = \sqrt{x_1^* x_1 + x_2^* x_2 + \dots + x_n^* x_n} = \sqrt{|x_1|^2 + \dots + |x_n|^2}.$$



## Theorem 1.1

Let  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ . Then  $|\langle \mathbf{x}, \mathbf{y} \rangle| = |\mathbf{x}^H \mathbf{y}| \leq \|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2$ .

## Proposition 1.2

The vector 2-norm is a norm.

**1.5.2 Vector 1-norm**

## Definition 1.7

The vector 1-norm  $\|\cdot\|_1 : \mathbb{C}^n \rightarrow \mathbb{R}$  is defined by

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \dots + |x_n|.$$

## Proposition 1.3

The vector 1-norm is a norm.

**1.5.3 Vector  $\infty$ -norm**

## Definition 1.8

The vector  $\infty$ -norm  $\|\cdot\|_\infty : \mathbb{C}^n \rightarrow \mathbb{R}$  is defined by  $\|\mathbf{x}\|_\infty = \max_i |x_i|$ .

## Proposition 1.4

The vector  $\infty$ -norm is a norm.

**1.5.4 Vector  $p$ -norm**

## Definition 1.9

The vector  $p$ -norm  $\|\cdot\|_p : \mathbb{C}^n \rightarrow \mathbb{R}$  is defined by

$$\|\mathbf{x}\|_p = \sqrt[p]{|x_1|^p + |x_2|^p + \dots + |x_n|^p}.$$

## Proposition 1.5

The vector  $p$ -norm is a norm.

## 1.6 Matrix Norms

It is not hard to see that vector norms are all measures of how "big" the vectors are. Similarly, we want to have measures for how "big" matrices are. We will start with one that are somewhat artificial and then move on to the important class of induced matrix norms.

### 1.6.1 Frobenius norm

#### Definition 1.10

The Frobenius norm  $\|\cdot\|_F : \mathbb{C}^{m \times n} \rightarrow \mathbb{R}$  is defined by

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2}.$$

**Remark.**  $\|\mathbf{A}\|_F = \|\text{vec}(\mathbf{A})\|_2$ .

#### Proposition 1.6

The Frobenius norm is a norm.

### 1.6.2 Induced matrix norms

#### Definition 1.11

Let  $\|\cdot\|_\mu : \mathbb{C}^m \rightarrow \mathbb{R}$  and  $\|\cdot\|_v : \mathbb{C}^n \rightarrow \mathbb{R}$  be vector norms. Define  $\|\cdot\|_{\mu,v} : \mathbb{C}^{m \times n} \rightarrow \mathbb{R}$  by

$$\|\mathbf{A}\|_{\mu,v} = \sup_{\mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_\mu}{\|\mathbf{x}\|_v}.$$

**Remark.** How "big"  $\mathbf{A}$  is, as measured by  $\|\mathbf{A}\|_{\mu,v}$ , is defined as the most that  $\mathbf{A}$  magnifies the length of nonzero vectors, where the length of the vector  $\mathbf{x}$  is measured with norm  $\|\cdot\|_v$  and the length of the transformed vector  $\mathbf{Ax}$  is measured with norm  $\|\cdot\|_\mu$ .

#### Proposition 1.7

Let  $\|\cdot\|_\mu : \mathbb{C}^m \rightarrow \mathbb{R}$  and  $\|\cdot\|_v : \mathbb{C}^n \rightarrow \mathbb{R}$  be vector norms.

$$\begin{aligned} \|\mathbf{A}\|_{\mu,v} &= \sup_{\mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_\mu}{\|\mathbf{x}\|_v} \\ &= \max_{\mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_\mu}{\|\mathbf{x}\|_v} \\ &= \max_{\|\mathbf{x}\|_v=1} \|\mathbf{Ax}\|_\mu. \end{aligned}$$

## Proposition 1.8

$\|\cdot\|_{\mu,v} : \mathbb{C}^{m \times n} \rightarrow \mathbb{R}$  is a norm.

## Definition 1.12

Define  $\|\cdot\|_p : \mathbb{C}^{m \times n} \rightarrow \mathbb{R}$  by

$$\begin{aligned}\|\mathbf{A}\|_p &= \max_{\mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_p}{\|\mathbf{x}\|_p} \\ &= \max_{\|\mathbf{x}\|_p=1} \|\mathbf{Ax}\|_p.\end{aligned}$$

## Proposition 1.9

For all  $\mathbf{A} \in \mathbb{C}^{m \times n}$ ,  $\mathbf{x} \in \mathbb{C}^n$ ,

$$\|\mathbf{Ax}\| \leq \|\mathbf{A}\|_p \cdot \|\mathbf{x}\|_p.$$

*Proof.* By the definition of  $\|\mathbf{A}\|_p$ ,

$$\frac{\|\mathbf{Ax}\|_p}{\|\mathbf{x}\|_p} \leq \|\mathbf{A}\|_p.$$

□

## Proposition 1.10

For any  $\mathbf{A} \in \mathbb{C}^{m \times k}$  and  $\mathbf{B} \in \mathbb{C}^{k \times n}$ ,

$$\begin{aligned}\|\mathbf{AB}\|_p &\leq \|\mathbf{A}\|_p \|\mathbf{B}\|_p \\ \|\mathbf{AB}\|_F &\leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F.\end{aligned}$$

## 1.7 Matrix multiplication

Suppose  $\mathbf{A} \in \mathbb{R}^{m \times r}$  and  $\mathbf{B} \in \mathbb{R}^{r \times n}$ , the matrix multiplication  $\mathbf{C} = \mathbf{AB}$  can be viewed from three different perspectives as follows:

Dot Product Matrix Multiply. Every element  $c_{ij}$  of  $\mathbf{C}$  is the dot product of row vector  $\mathbf{A}_i$  and

column vector  $\mathbf{B}_{\cdot j}$ .

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{1\cdot}^T \\ \vdots \\ \mathbf{A}_{m\cdot}^T \end{pmatrix}, \mathbf{A}_{k\cdot} \in \mathbb{R}^r$$

$$\mathbf{B} = (\mathbf{B}_{\cdot 1} \ \dots \ \mathbf{B}_{\cdot n}), \mathbf{B}_{\cdot k} \in \mathbb{R}^r$$

$$\mathbf{C} = (c_{ij}), c_{ij} = \mathbf{A}_{i\cdot}^T \mathbf{B}_{\cdot j} = \sum_{k=1}^r a_{ik} b_{kj}.$$

Column Combination Matrix Multiply. Every column  $\mathbf{C}_{\cdot j}$  of  $\mathbf{C}$  is a linear combination of column vector  $\mathbf{A}_{\cdot k}$  of  $\mathbf{A}$  with columns  $b_{kj}$  as the weight coefficients.

## 1.8 Matrix trace, Matrix inner product

### 1.8.1 Definition and properties of the trace

#### Definition 1.13

The trace of a square matrix  $\mathbf{A}$  of order  $I$  is defined as the sum of its diagonal elements:

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^I a_{ii}.$$

#### Proposition 1.11

The trace satisfies the following properties:

$$\begin{aligned} \text{tr}(\alpha \mathbf{A} + \beta \mathbf{B}) &= \alpha \text{tr}(\mathbf{A}) + \beta \text{tr}(\mathbf{B}), \\ \text{tr}(\mathbf{A}^T) &= \text{tr}(\mathbf{A}), \\ \text{tr}(\mathbf{A}^*) &= \text{tr}(\mathbf{A}^H) = (\text{tr}(\mathbf{A}))^*, \end{aligned}$$

## 1.9 Eigenvalues and eigenvectors

#### Definition 1.14

A real matrix  $\mathbf{A}$  is a symmetric matrix if it equals to its own transpose, that is  $\mathbf{A} = \mathbf{A}^T$ .

#### Definition 1.15

A complex matrix  $\mathbf{A}$  is a hermitian matrix if it equals to its own complex conjugate transpose, that is  $\mathbf{A} = \mathbf{A}^H$ .

## Definition 1.16

A real matrix  $\mathbf{Q}$  is an orthogonal matrix if the inverse of  $\mathbf{Q}$  equals to the transpose of  $\mathbf{Q}$ ,  $\mathbf{Q}^{-1} = \mathbf{Q}^T$ , that is  $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ .

## Definition 1.17

A complex matrix  $\mathbf{U}$  is a unitary matrix if the inverse of  $\mathbf{U}$  equals the complex conjugate transpose of  $\mathbf{U}$ ,  $\mathbf{U}^{-1} = \mathbf{U}^H$ , that is  $\mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}$ .

## Definition 1.18

A hermitian matrix  $\mathbf{Q}$  is positive semidefinite (abbreviated SPSD and denoted by  $\mathbf{Q} \succeq 0$ ) if

$$\mathbf{x}^H \mathbf{Q} \mathbf{x} \geq 0 \text{ for all } \mathbf{x} \in \mathbb{C}^n.$$

## Definition 1.19

A hermitian matrix  $\mathbf{Q}$  is positive semidefinite (abbreviated SPD and denoted by  $\mathbf{Q} \succ 0$ ) if

$$\mathbf{x}^H \mathbf{Q} \mathbf{x} > 0 \text{ for all } \mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq 0.$$

A number  $\lambda \in \mathbb{C}$  is an eigenvalue of  $\mathbf{M}$  if there exists a vector  $\bar{\mathbf{x}} \neq 0$  such that  $\mathbf{M}\bar{\mathbf{x}} = \lambda\bar{\mathbf{x}}$ .  $\bar{\mathbf{x}}$  is called an eigenvector of  $\mathbf{M}$  (and is called an eigenvector corresponding to  $\lambda$ ). Note that  $\lambda$  is an eigenvalue of  $\mathbf{M}$  if and only if there exists  $\bar{\mathbf{x}} \neq 0$  such that  $(\mathbf{M} - \lambda\mathbf{E})\bar{\mathbf{x}} = 0$  or, equivalently, if and only if  $\det(\mathbf{M} - \lambda\mathbf{E}) = 0$ .

Let  $g(\lambda) = \det(\mathbf{M} - \lambda\mathbf{E})$ . Then  $g(\lambda)$  is a polynomial of degree  $n$ , and so will have  $n$  roots that will solve the equation  $g(\lambda) = \det(\mathbf{M} - \lambda\mathbf{E}) = 0$ , including multiplicities. These roots are the eigenvalues of  $\mathbf{M}$ .

## Proposition 1.12

If  $\mathbf{Q}$  is a real symmetric matrix, all of its eigenvalues are real numbers.

## Proposition 1.13

If  $\mathbf{Q}$  is a complex hermitian matrix, all of its eigenvalues are real numbers.

## Proposition 1.14

If  $\mathbf{Q}$  is a hermitian matrix, its eigenvectors corresponding to different eigenvalues are orthogonal.

## Proposition 1.15

If  $\mathbf{Q}$  is SPD, the eigenvalues of  $\mathbf{Q}$  are nonnegative.

## Theorem 1.2

If  $\mathbf{A}$  is a real symmetric matrix, then  $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ , where  $\mathbf{Q}$  is an orthogonal matrix, the columns of  $\mathbf{Q}$  are an orthonormal basis of eigenvectors of  $\mathbf{A}$ , and  $\mathbf{\Lambda}$  is a diagonal matrix of the corresponding eigenvalues of  $\mathbf{A}$ .

## Theorem 1.3

If  $\mathbf{A}$  is a complex hermitian matrix, then  $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ , where  $\mathbf{U}$  is a unitary matrix, the columns of  $\mathbf{U}$  are an orthonormal basis of eigenvectors of  $\mathbf{A}$ , and  $\mathbf{\Lambda}$  is a diagonal matrix of the corresponding eigenvalues of  $\mathbf{A}$ .

## Definition 1.20

The square root of a matrix  $\mathbf{A}$  is a matrix  $\mathbf{A}^{1/2}$  such that

$$\mathbf{A} = \mathbf{A}^{1/2}\mathbf{A}^{1/2}.$$

## Proposition 1.16

If  $\mathbf{A}$  is SPSD, the  $\mathbf{A} = \mathbf{M}^H\mathbf{M}$  for some matrix  $\mathbf{M}$ .

*Proof.*  $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = (\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^H)(\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^H) = \mathbf{M}^H\mathbf{M}$ . □

## Definition 1.21

The Rayleigh quotient of the matrix  $\mathbf{A} \in \mathbb{C}^{n \times n}$  at the nonzero vector  $\mathbf{x} \in \mathbb{C}^n$  is the scalar

$$\frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}} \in \mathbb{C}.$$

**Remark.** If  $(\lambda, \mathbf{u})$  is an eigenpair for  $\mathbf{A}$ , then notice that

$$\frac{\mathbf{u}^H \mathbf{A} \mathbf{u}}{\mathbf{u}^H \mathbf{u}} = \frac{\mathbf{u}^H (\lambda \mathbf{u})}{\mathbf{u}^H \mathbf{u}} = \lambda,$$

so Rayleigh quotients generalize eigenvalues.

**Remark.** For Hermitian  $\mathbf{A} \in \mathbb{C}^{n \times n}$ , then there exists unitary matrix  $\mathbf{U}$  such that  $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ , for any  $\mathbf{x} \neq \mathbf{0}$ , it can be represented by  $\mathbf{x} = \mathbf{U}\mathbf{c} = \sum_{j=1}^n c_j \mathbf{u}_j$ , then

$$\frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}} = \frac{\mathbf{c}^H \mathbf{U}^H \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \mathbf{U} \mathbf{c}}{\mathbf{c}^H \mathbf{U}^H \mathbf{U} \mathbf{c}} = \frac{\mathbf{c}^H \mathbf{\Lambda} \mathbf{c}}{\mathbf{c}^H \mathbf{c}}$$

The diagonal structure of  $\mathbf{\Lambda}$  allows for an illuminating refinement,

$$\frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}} = \frac{\lambda_1 |c_1|^2 + \dots + \lambda_n |c_n|^2}{|c_1|^2 + \dots + |c_n|^2}.$$

As the numerator and denominator are both real, notice that the Rayleigh quotients for a Hermitian matrix is always real. We can say more: if the eigenvalues are ordered,  $\lambda_1 \leq \dots \leq \lambda_n$ ,

$$\frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}} = \frac{\lambda_1 |c_1|^2 + \dots + \lambda_n |c_n|^2}{|c_1|^2 + \dots + |c_n|^2} \geq \frac{\lambda_1 (|c_1|^2 + \dots + |c_n|^2)}{|c_1|^2 + \dots + |c_n|^2} = \lambda_1,$$

and similarly,

$$\frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}} = \frac{\lambda_1 |c_1|^2 + \dots + \lambda_n |c_n|^2}{|c_1|^2 + \dots + |c_n|^2} \leq \frac{\lambda_n (|c_1|^2 + \dots + |c_n|^2)}{|c_1|^2 + \dots + |c_n|^2} = \lambda_n.$$

#### Proposition 1.17

For a Hermitian matrix  $\mathbf{A} \in \mathbb{C}^{n \times n}$  with eigenvalues  $\lambda_1 \leq \dots \leq \lambda_n$ , the Rayleigh quotient for nonzero  $\mathbf{x} \in \mathbb{C}^{n \times n}$  satisfies

$$\frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}} \in [\lambda_1, \lambda_n].$$

#### Proposition 1.18

If  $\mathbf{A} \in \mathbb{C}^{m \times n}$  with  $m < n$  and  $\mathbf{A}$  has rank  $m$ , then

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^H \mathbf{A})} = \sqrt{\lambda_{\max}(\mathbf{A} \mathbf{A}^H)},$$

where  $\lambda_{\max}(\mathbf{M})$  denotes the largest eigenvalue of a matrix  $\mathbf{M}$ .

*Proof.*

$$\|\mathbf{A}\|_2^2 = \sup_{\mathbf{x} \in \mathbb{C}^n, \|\mathbf{x}\| \neq 0} \frac{\|\mathbf{A} \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} = \sup_{\mathbf{x} \in \mathbb{C}^n, \|\mathbf{x}\| \neq 0} \frac{\langle \mathbf{A} \mathbf{x}, \mathbf{A} \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \sup_{\mathbf{x} \in \mathbb{C}^n, \|\mathbf{x}\| \neq 0} \frac{\mathbf{x} \mathbf{A}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}} = \lambda_{\max}(\mathbf{A}^H \mathbf{A}).$$

□

#### Proposition 1.19

If  $\mathbf{A} \in \mathbb{C}^{n \times n}$  is a hermitian matrix, then

$$\|\mathbf{A}\|_2 = |\lambda_{\max}(\mathbf{A})|.$$

*Proof.* Since  $\mathbf{A}$  is a hermitian matrix, we have  $\mathbf{A} = \mathbf{U}^H \mathbf{\Lambda} \mathbf{U}$  where  $\mathbf{U}$  is an unitary matrix and  $\mathbf{\Lambda}$  is a diagonal matrix containing the eigenvalue of  $\mathbf{A}$ . Let  $\mathbf{y} = \mathbf{U} \mathbf{x}$ , then  $\|\mathbf{y}\|_2 = \|\mathbf{U} \mathbf{x}\|_2$ . Hence

$$\|\mathbf{A}\|_2^2 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A} \mathbf{x}\|_2^2 = \max_{\|\mathbf{x}\|_2=1} \langle \mathbf{A} \mathbf{x}, \mathbf{A} \mathbf{x} \rangle = \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^H \mathbf{A}^H \mathbf{A} \mathbf{x} = \max_{\|\mathbf{y}\|_2=1} \mathbf{y}^H \mathbf{\Lambda} \mathbf{y} = (\lambda_{\max}(\mathbf{A}))^2.$$

□

## Proposition 1.20

Suppose that  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is a SPSP. Then the following are equivalent:

- (a)  $h > 0$  satisfies  $\|\mathbf{A}^{-1}\|_2 \leq \frac{1}{h}$ .
- (b)  $h > 0$  satisfies  $\|\mathbf{A}\mathbf{x}\|_2 \geq h \cdot \|\mathbf{x}\|_2$  for any vector  $\mathbf{x}$
- (c)  $h > 0$  satisfies  $|\lambda_i(\mathbf{A})| \geq h$  for every eigenvalue  $\lambda_i(\mathbf{A})$  of  $\mathbf{A}$ ,  $i = 1, \dots, m$ .

*Proof.* By proposition 1.17, we have for all  $\mathbf{x} \neq 0$ ,

$$\frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \in [\lambda_{\min}(\mathbf{A}), \lambda_{\max}(\mathbf{A})].$$

Firstly, we claim that

$$\|\mathbf{A}^{-1}\|_2 = \frac{1}{\lambda_{\min}(\mathbf{A})}.$$

In fact,

$$\begin{aligned} \|\mathbf{A}^{-1}\|_2 &= \max_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}^{-1}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \max_{\mathbf{Ax} \neq 0} \frac{\|\mathbf{A}^{-1}\mathbf{Ax}\|_2}{\|\mathbf{Ax}\|_2} \\ &= \max_{\mathbf{Ax} \neq 0} \frac{\|\mathbf{x}\|_2}{\|\mathbf{Ax}\|_2} = \max_{\mathbf{x} \neq 0} \frac{\|\mathbf{x}\|_2}{\|\mathbf{Ax}\|_2} \\ &= \frac{1}{\min_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2}} = \frac{1}{\lambda_{\min}(\mathbf{A})}, \end{aligned}$$

the second and forth equality follows from the fact that  $\mathbf{x} \neq 0$  if and only if  $\mathbf{Ax} \neq 0$  since  $\mathbf{A}$  is nonsingular.

- (a)  $\Rightarrow$  (b)  $\frac{1}{\lambda_{\min}(\mathbf{A})} = \|\mathbf{A}^{-1}\|_2 \leq \frac{1}{h} \Rightarrow \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} \geq \lambda_{\min}(\mathbf{A}) \geq h, \forall \mathbf{x} \neq 0$ .
- (b)  $\Rightarrow$  (a)  $\|\mathbf{Ax}\|_2 \geq h \cdot \|\mathbf{x}\|_2, \forall \mathbf{x} \Rightarrow \lambda_{\min}(\mathbf{A}) \geq h \Rightarrow \frac{1}{\lambda_{\min}(\mathbf{A})} = \|\mathbf{A}^{-1}\|_2 \leq \frac{1}{h}$ .
- (b)  $\Leftrightarrow$  (c)  $\|\mathbf{Ax}\|_2 \geq h \cdot \|\mathbf{x}\|_2, \forall \mathbf{x} \Leftrightarrow \lambda_{\min}(\mathbf{A}) \geq h \Leftrightarrow \lambda_i(\mathbf{A}) \geq h, i = 1, \dots, m$ .

□

## 1.10 Generalized inverse and pseudoinverse

When  $\mathbf{A}$  is a square matrix of full rank, the system of linear equations  $\mathbf{Ax} = \mathbf{b}$  has a unique solution. When  $\mathbf{A}$  is a square of nonfull rank or rectangular matrix,  $\mathbf{A}^{-1}$  does not exist. Can we express the solution of the linear equation system  $\mathbf{Ax} = \mathbf{b}$  in a similar way as an inverse matrix?



### 1.10.1 Generalized inverse

#### Definition 1.22

Let  $\mathbf{A} \in \mathbb{C}^{m \times n}$  be any matrix. We call the matrix  $\mathbf{G} \in \mathbb{C}^{n \times m}$  a generalized inverse of  $\mathbf{A}$  if it satisfies

$$\mathbf{AGA} = \mathbf{A}$$

**Remark.** If  $\mathbf{A}$  is square and invertible, then it has one and only one generalized inverse which must coincide with the ordinary inverse  $\mathbf{A}^{-1}$ . To see this, first observe that  $\mathbf{A}^{-1}$  apparently satisfies the definition and thus is a generalized inverse. Conversely, if  $\mathbf{A}$  has a generalized inverse  $\mathbf{G}$ , then from the equation  $\mathbf{AGA} = \mathbf{A}$  we get

$$\mathbf{G} = \mathbf{A}^{-1}(\mathbf{AGA})\mathbf{A}^{-1} = \mathbf{A}^{-1}(\mathbf{A})\mathbf{A}^{-1} = \mathbf{A}^{-1}.$$

This thus justifies the term "generalized inverse".

**Remark.** For a general matrix  $\mathbf{A} \in \mathbb{C}^{m \times n}$ , its generalized inverse always exists but might not be unique.

### 1.10.2 property of generalized inverse and pseudoinverse

### 1.10.3 Pseudoinverse

Briefly speaking, the matrix pseudoinverse is a generalized inverse with more constraints.

#### Definition 1.23

Let  $\mathbf{A} \in \mathbb{C}^{m \times n}$ . We call the matrix  $\mathbf{B} \in \mathbb{C}^{n \times m}$  the pseudoinverse of  $\mathbf{A}$  if it satisfies all four conditions below:

- (a)  $\mathbf{ABA} = \mathbf{A}$  ( $\mathbf{B}$  is a generalized inverse of  $\mathbf{A}$ )
- (b)  $\mathbf{BAB} = \mathbf{B}$  ( $\mathbf{A}$  is a generalized inverse of  $\mathbf{B}$ )
- (c)  $(\mathbf{AB})^H = \mathbf{AB}$  ( $\mathbf{AB}$  is hermitian)
- (d)  $(\mathbf{BA})^H = \mathbf{BA}$  ( $\mathbf{BA}$  is hermitian)

#### Remark.

(1) If  $\mathbf{B}$  satisfies Condition (a), it is known as a generalized inverse of  $\mathbf{A}$ ; if  $\mathbf{B}$  satisfies Conditions (a) and (b), it is called a reflexive generalized inverse. Only when  $\mathbf{B}$  satisfies all 4 conditions, it is called the pseudoinverse of  $\mathbf{A}$ .

(2) It can be shown that for any matrix  $\mathbf{A} \in \mathbb{C}^{m \times n}$ , the pseudoinverse always exists and is unique. We denote the pseudoinverse of  $\mathbf{A}$  as  $\mathbf{A}^\dagger$ .

(3) A pseudoinverse is sometimes called the Moore-Penrose inverse, after the pioneering works by E. H. Moore and Roger Penrose.

(4) The hermitian form of the definition implies  $\mathbf{B} = \mathbf{A}^\dagger$  and  $\mathbf{A} = \mathbf{B}^\dagger$ , and thus,  $\mathbf{A} = (\mathbf{A}^\dagger)^\dagger$ .

## 1.11 Reference

- From Algebraic Structures to Tensors ch4 matrix algebra
- [Notes on Vector and Matrix Norms](#)
- [symmetric, positive definted, eigenvalues and eigenvectors](#)
- [Rayleigh quotient](#)
- [The Principal Axis Theorem](#)
- [linear algebra and geometry note](#)
- [eigenvalues application: ellipses](#)
- [Generalized inverses](#)
- [The Moore-Penrose Inverse and Least Squares](#)

# Chapter 2

## Matrix Decompositions

### 2.1 SVD

We have seen that hermitian matrices are always diagonalizable. What about general rectangular matrices? The following theorem shows that one can always find two matrix  $\mathbf{U}$  and  $\mathbf{V}$  so that the matrix  $\mathbf{U}^H \mathbf{A} \mathbf{V}$  will be diagonal. This can be written as  $\mathbf{U}^H \mathbf{A} \mathbf{V} = \mathbf{\Sigma}$  or equivalently,  $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$ .

$$\begin{array}{lcl}
 \text{SVD } (m > n) & \begin{array}{c} \boxed{\phantom{A}} \\ A \end{array} & = \begin{array}{c} \boxed{\phantom{U}} \\ U \end{array} \times \begin{array}{c} \boxed{\begin{array}{c} 0 \\ \diagdown \\ 0 \end{array}} \\ \Sigma \end{array} \times \begin{array}{c} \boxed{\phantom{V^H}} \\ V^H \end{array} \\
 \text{SVD } (m < n) & \begin{array}{c} \boxed{\phantom{A}} \\ A \end{array} & = \begin{array}{c} \boxed{\phantom{U}} \\ U \end{array} \times \begin{array}{c} \boxed{\begin{array}{c} \diagdown \\ 0 \end{array}} \\ \Sigma \end{array} \times \begin{array}{c} \boxed{\phantom{V^H}} \\ V^H \end{array}
 \end{array}$$

Figure 2.1

**Remark.**  $\mathbf{\Sigma} \in \mathbb{C}^{m \times n}$  is said to be diagonal if  $\Sigma_{ij} = 0$  for  $i \neq j$ . It has exactly  $p = \min\{m, n\}$  diagonal entries and can be denoted by  $\mathbf{\Sigma} = \text{diag}\{d_1, \dots, d_p\}$ .

#### 2.1.1 Singular value decomposition (SVD)

##### Proposition 2.1

For any matrix  $\mathbf{A} \in \mathbb{C}^{m \times n}$ ,  $\mathbf{A}^H \mathbf{A}$  and  $\mathbf{A} \mathbf{A}^H$  have non-negative eigenvalues.

*Proof.*  $\mathbf{A}^H \mathbf{A}$  and  $\mathbf{A} \mathbf{A}^H$  are hermitian and positive semi-definite. □

##### Proposition 2.2

For any matrix  $\mathbf{A} \in \mathbb{C}^{m \times n}$ ,  $r(\mathbf{A}) = r(\mathbf{A}^H \mathbf{A}) = r(\mathbf{A} \mathbf{A}^H)$ .

*Proof.* For  $\mathbf{x}$ ,  $\mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{A}^H \mathbf{A}\mathbf{x} = \mathbf{0}$ . For  $\mathbf{x}$ ,  $\mathbf{A}^H \mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x}^H \mathbf{A}^H \mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \|\mathbf{A}\mathbf{x}\|_2^2 = 0 \Rightarrow \mathbf{A}\mathbf{x} = \mathbf{0}$ . Hence,  $\dim(\ker \mathbf{A}) = \dim(\ker \mathbf{A}^H \mathbf{A})$  and so  $r(\mathbf{A}) = r(\mathbf{A}^H \mathbf{A})$ . The result follows since  $r(\mathbf{A}\mathbf{A}^H) = r(\mathbf{A}^H \mathbf{A})$ .  $\square$

### Theorem 2.1

Let  $\mathbf{A} \in \mathbb{C}^{m \times n}$  and denote  $p = \min\{m, n\}$ . Denote the rank of  $\mathbf{A}$  by  $r$  ( $0 \leq r \leq p$ ). Then there are unitary matrices  $\mathbf{U} \in \mathbb{C}^{m \times m}$  and  $\mathbf{V} \in \mathbb{C}^{n \times n}$  and a real diagonal matrix  $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_p\} \in \mathbb{R}^{m \times n}$  such that

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^H$$

and

$$\sigma_1 \geq \dots \geq \sigma_r \geq \sigma_{r+1} = \dots = \sigma_p = 0,$$

the matrix  $\Sigma$  is uniquely determined by  $\mathbf{A}$ .

**Proof Strategy:** Assume  $m > n$ , if  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^H$ , then

$$\begin{aligned} \mathbf{A}^H \mathbf{A} &= \mathbf{V}(\Sigma^H \Sigma)\mathbf{V}^H \\ &= \mathbf{V} \left( \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix} \right) \mathbf{V}^H & (\sigma_i^* = \sigma_i \text{ since } \sigma_i \in \mathbb{R}) \\ &= \mathbf{V} \begin{bmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_r^2 & \\ & & & 0 \end{bmatrix} \mathbf{V}^H \end{aligned}$$

Then we can let  $\mathbf{V}$  be matrix containing the eigenvectors of  $\mathbf{A}^H \mathbf{A}$  and  $\Sigma$  be matrix containing square roots of eigenvalues of  $\mathbf{A}^H \mathbf{A}$ . After we have found both  $\mathbf{V}$  and  $\Sigma$ , rewrite the matrix equation as

$$\mathbf{A}\mathbf{V} = \mathbf{U}\Sigma,$$

or in columns,

$$\mathbf{A} \begin{pmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_r & \mathbf{v}_{r+1} & \dots & \mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_r & \mathbf{u}_{r+1} & \dots & \mathbf{u}_n \end{pmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix}$$

By comparing columns, we obtain

$$\mathbf{X}\mathbf{v}_i = \begin{cases} \sigma_i \mathbf{u}_i, & 1 \leq i \leq r \text{ (nonzero singular values)} \\ \mathbf{0}, & r < i \leq d. \end{cases}$$

This tells us how to find the matrix  $\mathbf{U}$  :  $\mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{A}\mathbf{v}_i$  for  $1 \leq i \leq r$ .

*Proof.* Let  $\mathbf{C} = \mathbf{A}^H \mathbf{A} \in \mathbb{C}^{n \times n}$ . Then  $\mathbf{C}$  is square, hermitian, and positive semidefinite. Therefore,  $\mathbf{C} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^H$  for an unitary  $\mathbf{V} \in \mathbb{C}^{n \times n}$  and diagonal  $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_n\} \in \mathbb{R}^{n \times n}$  with  $\lambda_1 \geq \dots \geq \lambda_r > 0 = \lambda_{r+1} = \dots = \lambda_n$ . Let  $\sigma_i = \sqrt{\lambda_i}$  and correspondingly form the matrix  $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ :

$$\mathbf{\Sigma} = \begin{pmatrix} \text{diag}(\sigma_1, \dots, \sigma_r) & \mathbf{O}_{r \times (n-r)} \\ \mathbf{O}_{(m-r) \times r} & \mathbf{O}_{(m-r) \times (n-r)} \end{pmatrix}$$

Define also

$$\mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{A}\mathbf{v}_i \in \mathbb{C}^n, \text{ for each } 1 \leq i \leq r.$$

Then  $\mathbf{u}_1, \dots, \mathbf{u}_r$  are orthonormal vectors. To see this,

$$\begin{aligned} \mathbf{u}_i^H \mathbf{u}_j &= \left( \frac{1}{\sigma_i} \mathbf{A}\mathbf{v}_i \right)^H \left( \frac{1}{\sigma_j} \mathbf{A}\mathbf{v}_j \right) = \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^H \mathbf{A}^H \mathbf{A} \mathbf{v}_j = \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^H \mathbf{C} \mathbf{v}_j \\ &= \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^H (\lambda_j \mathbf{v}_j) & (\lambda_j = \sigma_j^2) &= \frac{\sigma_j}{\sigma_i} \mathbf{v}_i^H \mathbf{v}_j \\ &= \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases} \end{aligned}$$

Choose  $\mathbf{u}_{r+1}, \dots, \mathbf{u}_n \in \mathbb{C}^n$  (through basis completion) such that

$$\mathbf{U} = (\mathbf{u}_1 \quad \dots \quad \mathbf{u}_r \quad \mathbf{u}_{r+1} \quad \dots \quad \mathbf{u}_m) \in \mathbb{C}^{m \times m}$$

is an orthogonal matrix. It remains to verify that  $\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma}$ , i.e.,

$$\mathbf{A} \begin{pmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_r & \dots & \mathbf{v}_{r+1} & \dots & \mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_r & \dots & \mathbf{u}_{r+1} & \dots & \mathbf{u}_m \end{pmatrix} \begin{bmatrix} \sigma_1 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_r & & & & \\ & & & & & & \end{bmatrix}.$$

Consider two cases:

(a)  $1 \leq i \leq r$ :  $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$  by construction.

(b)  $i > r$ :  $\mathbf{A}\mathbf{v}_i = \mathbf{0}$ , which is due to  $\mathbf{A}^T \mathbf{A} \mathbf{v}_i = \mathbf{C} \mathbf{v}_i = 0 \mathbf{v}_i = \mathbf{0}$ .

Consequently, we have obtained that  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ .

Since  $\sigma_1, \dots, \sigma_p$  is uniquely determined by  $\mathbf{A}^T \mathbf{A}$ ,  $\mathbf{\Sigma}$  is uniquely determined by  $\mathbf{A}$ . □

**Remark.**  $\mathbf{U}$  and  $\mathbf{V}$  is not unique (this is mainly due to different choices of orthogonal basis for each eigenspace)

### 2.1.2 Geometric interpretation of SVD

#### Theorem 2.2: Real SVD

If  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , then it has a real SVD

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T,$$

where  $\mathbf{U} \in \mathbb{R}^{m \times m}$  and  $\mathbf{V} \in \mathbb{R}^{n \times n}$  are real orthogonal matrices.

*Proof.* Proof is just a repetition of the proof in theorem 2.1, except  $\mathbf{v}_1, \dots, \mathbf{v}_n$  and  $\mathbf{u}_1, \dots, \mathbf{u}_m$  must be real vectors.  $\square$

Given any matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , it defines a linear transformation:

$$f : \mathbb{R}^n \mapsto \mathbb{R}^m \text{ with } f(\mathbf{x}) = \mathbf{A}\mathbf{x}.$$

The SVD of  $\mathbf{A} \in \mathbb{R}^{m \times n}$  indicates that the linear transformation  $f$  can be decomposed into a sequence of three operations:

$$\underbrace{\mathbf{A}\mathbf{x}}_{\text{full transformation}} = \underbrace{\mathbf{U}}_{\text{rotation}} \cdot \underbrace{\mathbf{\Sigma}}_{\text{rescaling}} \cdot \underbrace{\mathbf{V}^H \mathbf{x}}_{\text{rotation}}.$$

### 2.1.3 Applications of Singular-Value Decomposition

Least square problem

## 2.2 Reference

- [Matrix Decomposition](#)
- [Mathematical Methods for Data Visualization](#)
- [SVD notes XMUT](#)
- [Singular Value Decomposition](#)
- [least squares svd](#)
- [svd application](#)
- [Low rank approximation l1](#)
- [Notes on Low-rank Matrix Factorization](#)

# Chapter 3

## Hadamard, Kronecker and Khatri-Rao Products

### 3.1 Partitioned matrices

Let  $\{\alpha_{m_1}, \dots, \alpha_{m_R}\}$  and  $\{\beta_{n_1}, \dots, \beta_{n_S}\}$  be partitions of the sets  $\{1, \dots, m\}$  and  $\{1, \dots, n\}$ , respectively, with  $m_r \in \langle m \rangle$  and  $n_s \in \langle n \rangle$ , such that  $\sum_{r=1}^R m_r = m$  and  $\sum_{s=1}^S n_s = n$ . It is said that matrices  $\mathbf{A}_{rs}$  of dimensions  $(m_r, n_s)$  form a partition of the matrix  $\mathbf{A} \in \mathbb{K}^{m \times n}$  into  $(R, S)$  blocks, or that  $\mathbf{A}$  is partitioned into  $(R, S)$  blocks, if  $\mathbf{A}$  can be written as:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{1S} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \dots & \mathbf{A}_{2S} \\ \dots & \dots & \dots & \dots \\ \mathbf{A}_{R1} & \mathbf{A}_{R2} & \dots & \mathbf{A}_{RS} \end{pmatrix} = (\mathbf{A}_{rs}), r \in \langle R \rangle, s \in \langle S \rangle.$$

All submatrices of the same row-block ( $r$ ) contain the same number ( $m_r$ ) of rows. Similarly, all submatrices of the same column-block ( $s$ ) contain the same number ( $n_s$ ) of columns, that is:

$$(\mathbf{A}_{r1} \ \mathbf{A}_{r2} \ \dots \ \mathbf{A}_{rS}) \in \mathbb{K}^{m_r \times n}, (\mathbf{A}_{1s} \ \mathbf{A}_{2s} \ \dots \ \mathbf{A}_{Rs}) \in \mathbb{K}^{m \times n_s}.$$

It is then said that the submatrices  $\mathbf{A}_{rs}$  are of compatible dimensions.

In the particular case where  $n = 1$ , the partitioned matrix becomes a block-column vector:

$$\mathbf{a} = (\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_R) \in \mathbb{K}^{m \times 1}, \mathbf{a}_r \in \mathbb{K}^{m_r \times 1}, r \in \langle R \rangle.$$

Similarly, when  $m = 1$ , the partitioned matrix becomes a block-row vector:

$$\mathbf{a}^T = (\mathbf{a}_1^T \ \mathbf{a}_2^T \ \dots \ \mathbf{a}_S^T) \in \mathbb{K}^{1 \times n}, \mathbf{a}_s \in \mathbb{K}^{n_s \times 1}, s \in \langle S \rangle.$$

## 3.2 Notation

We will write  $\mathbf{A}^*$ ,  $\mathbf{A}^T$ ,  $\mathbf{A}^H$ ,  $\mathbf{A}^\dagger$ ,  $\mathbf{A}_{i\cdot}$ ,  $\mathbf{A}_{\cdot j}$ ,  $r(\mathbf{A})$  and  $\det(\mathbf{A})$ , for the conjugate, the transpose, the transconjugate (also known as conjugate transpose or Hermitian transpose), the Moore-Penrose pseudo-inverse, the  $i$ th row, the  $j$ th column, the rank and the determinant of  $\mathbf{A} \in \mathbb{K}^{I \times J}$ , respectively. In the following literature, the Hadamard product is denoted by  $*$ . The symbols  $\otimes$  and  $\odot$  are used for the Kronecker and Khatri-Rao product, respectively.

The symbol  $\mathbf{1}_I$  denotes a column vector of size  $I$  whose elements are all equal to 1. The elements of the matrices  $\mathbf{0}_{I \times J}$  and  $\mathbf{1}_{I \times J}$  of size  $(I \times J)$  are all equal to 0 and 1, respectively. The symbols  $\mathbf{I}_N$  and  $\mathbf{e}_n^{(N)}$  denote the identity matrix of order  $N$  and the  $n$ th vector of the canonical basis of the vector space  $\mathbb{R}^N$ , respectively.

## 3.3 Hadamard product

### 3.3.1 Definition and identities

#### Definition 3.1

Let  $\mathbf{A}$  and  $\mathbf{B} \in \mathbb{K}^{I \times J}$  be two matrices of the same size. The Hadamard product of  $\mathbf{A}$  and  $\mathbf{B}$  is the matrix  $\mathbf{C} \in \mathbb{K}^{I \times J}$  defined as follows:

$$\mathbf{C} = \mathbf{A} * \mathbf{B} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} & \dots & a_{1J}b_{1J} \\ a_{21}b_{21} & a_{22}b_{22} & \dots & a_{2J}b_{2J} \\ \dots & \dots & \dots & \dots \\ a_{I1}b_{I1} & a_{I2}b_{I2} & \dots & a_{IJ}b_{IJ} \end{pmatrix},$$

i.e.  $c_{ij} = a_{ij}b_{ij}$ , and therefore  $\mathbf{C} = (a_{ij}b_{ij})$ , with  $i \in \langle I \rangle$ ,  $j \in \langle J \rangle$ .



## 3.4 Kronecker product

### 3.4.1 Kronecker product of vectors

#### Definition

##### Definition 3.2

(a) For  $\mathbf{u} \in \mathbb{K}^I$  and  $\mathbf{v} \in \mathbb{K}^J$ , we have:

$$\mathbf{x} = \mathbf{u} \otimes \mathbf{v} \in \mathbb{K}^{IJ} \Leftrightarrow x_{j+(i-1)J} = u_i v_j$$

or equivalently:

$$\mathbf{u} \otimes \mathbf{v} = (u_1 v_1 \quad \dots \quad u_1 v_J \quad u_2 v_1 \quad \dots \quad u_2 v_J \quad \dots \quad u_I v_1 \quad \dots \quad u_I v_J)^T = (u_1 \mathbf{v} \quad \dots \quad u_I \mathbf{v})^T.$$

(b) Similarly, for  $\mathbf{u} \in \mathbb{K}^I$ ,  $\mathbf{v} \in \mathbb{K}^J$ , and  $\mathbf{w} \in \mathbb{K}^K$ , we have:

$$\mathbf{x} = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} \in \mathbb{K}^{IJK} \Leftrightarrow x_{k+(j-1)K+(i-1)JK} = u_i v_j w_k.$$

**Remark.** By convention, the order of the dimensions in a product  $IJK$  follows the order of variation of the corresponding indices  $(i, j, k)$ . For example,  $\mathbb{K}^{IJK}$  means that the index  $i$  varies more slowly than  $j$ , which itself varies more slowly than  $k$ .

### 3.4.2 Kronecker product of matrices

#### Definitions and identities

##### Definition 3.3

Given two matrices  $\mathbf{A} \in \mathbb{K}^{I \times J}$  and  $\mathbf{B} \in \mathbb{K}^{M \times N}$  of arbitrary size, the right Kronecker product of  $\mathbf{A}$  by  $\mathbf{B}$  is the matrix  $\mathbf{C} \in \mathbb{K}^{IM \times JN}$  defined as follows:

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1J}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2J}\mathbf{B} \\ \dots & \dots & \dots & \dots \\ a_{I1}\mathbf{B} & a_{I2}\mathbf{B} & \dots & a_{IJ}\mathbf{B} \end{pmatrix} = (a_{ij}\mathbf{B}).$$

**Remark.** The Kronecker product is a matrix partitioned into  $(I, J)$  blocks, where the block  $(i, j)$  is given by the matrix  $a_{ij}\mathbf{B} \in \mathbb{K}^{M \times N}$ . The element  $a_{ij}b_{mn}$  is located at the position  $((i-1)M+m, (j-1)N+n)$  in  $\mathbf{A} \otimes \mathbf{B}$ .

## 3.5 Reference

- From Algebraic Structures to Tensors ch5
- Matrix and Tensor Decompositions in Signal Processing ch2

- [Kronecker Product and the vec Operator](#)
- [On Hadamard and Kronecker Products Over Matrix of Matrices](#)
- [HADAMARD, KHATRI-RAO, KRONECKER AND OTHER MATRIX PRODUCTS](#)

# Chapter 4

## Tensor Operations

### 4.1 Notation

Let  $\chi \in \mathbb{K}^{I_1}$

### 4.2 Notion of slice

### 4.3 Mode combination

### 4.4 Matricization

### 4.5 Multiplication operations

## Chapter 5

# Tensor Decomposition