



香港中文大學(深圳)  
The Chinese University of Hong Kong, Shenzhen

# *Linear Algebra*

*MAT2040 Notebook*

*Prof. Tom Luo*

*The First Edition*

**A FIRST COURSE  
IN  
LINEAR ALGEBRA**

---

**A FIRST COURSE**  
**IN**  
**LINEAR ALGEBRA**  
**MAT2040 Notebook**

---

**Prof. Tom Luo**

*The Chinese University of Hong Kong, Shenzhen*

**Prof. Ruoyu Sun**

*University of Illinois Urbana-Champaign*



香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen





# Contents

<b>Preface</b>	<b>xi</b>
<b>Acknowledgments</b>	<b>xiii</b>
<b>Notations</b>	<b>xv</b>
<b>1 Week1</b>	<b>1</b>
<b>1.1 Tuesday</b>	<b>1</b>
1.1.1 Introduction	1
1.1.2 Gaussian Elimination	3
1.1.3 Complexity Analysis	11
1.1.4 Brief Summary	12
<b>1.2 Thursday</b>	<b>14</b>
1.2.1 Row-Echelon Form	14
1.2.2 Matrix Multiplication	16
1.2.3 Special Matrices	19
<b>1.3 Friday</b>	<b>21</b>
1.3.1 Matrix Multiplication	21
1.3.2 Elementary Matrix	22
1.3.3 Properties of Matrix	24
1.3.4 Permutation Matrix	26
1.3.5 LU decomposition	29
1.3.6 LDU decomposition	33
1.3.7 LU Decomposition with row exchanges	35
<b>1.4 Assignment One</b>	<b>36</b>

<b>2</b>	<b>Week2</b>	<b>39</b>
<b>2.1</b>	<b>Tuesday</b>	<b>39</b>
2.1.1	Review	39
2.1.2	Special matrix multiplication case	41
2.1.3	Inverse	44
<b>2.2</b>	<b>Wednesday</b>	<b>49</b>
2.2.1	Remarks on Gaussian Elimination	49
2.2.2	Properties of matrix	50
2.2.3	matrix transpose	53
<b>2.3</b>	<b>Assignment Two</b>	<b>56</b>
<b>2.4</b>	<b>Friday</b>	<b>57</b>
2.4.1	symmetric matrix	57
2.4.2	Interaction of inverse and transpose	58
2.4.3	Vector Space	59
<b>2.5</b>	<b>Assignment Three</b>	<b>68</b>
<b>3</b>	<b>Week3</b>	<b>71</b>
<b>3.1</b>	<b>Tuesday</b>	<b>71</b>
3.1.1	Introduction	71
3.1.2	Review of 2 weeks	72
3.1.3	Examples of solving equations	73
3.1.4	How to solve a general rectangular	78
<b>3.2</b>	<b>Thursday</b>	<b>83</b>
3.2.1	Review	83
3.2.2	Remarks on solving linear system equations	86
3.2.3	Linear dependence	88
3.2.4	Basis and dimension	90
<b>3.3</b>	<b>Friday</b>	<b>96</b>
3.3.1	Review	96

3.3.2	More on basis and dimension	97
3.3.3	What is rank?	99
<b>3.4</b>	<b>Assignment Four</b>	<b>107</b>
<b>4</b>	<b>Midterm</b>	<b>109</b>
<b>4.1</b>	<b>Sample Exam</b>	<b>109</b>
<b>4.2</b>	<b>Midterm Exam</b>	<b>116</b>
<b>5</b>	<b>Week4</b>	<b>123</b>
<b>5.1</b>	<b>Friday</b>	<b>123</b>
5.1.1	Linear Transformation	123
5.1.2	Example: differentiation	125
5.1.3	Basis Change	130
5.1.4	Determinant	132
<b>5.2</b>	<b>Assignment Five</b>	<b>138</b>
<b>6</b>	<b>Week5</b>	<b>141</b>
<b>6.1</b>	<b>Tuesday</b>	<b>141</b>
6.1.1	Formulas for Determinant	141
6.1.2	Determinant by Cofactors	146
6.1.3	Determinant Applications	147
6.1.4	Orthogonality	150
<b>6.2</b>	<b>Thursday</b>	<b>154</b>
6.2.1	Orthogonality	154
6.2.2	Least Squares Approximations	160
6.2.3	Projections	163
<b>6.3</b>	<b>Friday</b>	<b>166</b>
6.3.1	Orthonormal basis	166
6.3.2	Gram-Schmidt Process	170

6.3.3	The Factorization $A = QR$ .	175
6.3.4	Function Space	177
6.3.5	Fourier Series	179
<b>6.4</b>	<b>Assignment Six</b>	<b>180</b>
<b>7</b>	<b>Week6</b>	<b>181</b>
<b>7.1</b>	<b>Tuesday</b>	<b>181</b>
7.1.1	Summary of previous weeks	181
7.1.2	Eigenvalues and eigenvectors	185
7.1.3	Products and Sums of Eigenvalue	189
7.1.4	Application: Page Rank and Web Search	190
<b>7.2</b>	<b>Thursday</b>	<b>193</b>
7.2.1	Review	193
7.2.2	Similarity	193
7.2.3	Diagonalization	195
7.2.4	Powers of $A$	200
7.2.5	Nondiagonalizable Matrices	201
<b>7.3</b>	<b>Friday</b>	<b>205</b>
7.3.1	Review	205
7.3.2	Fibonacci Numbers	205
7.3.3	Imaginary Eigenvalues	207
7.3.4	Complex Numbers and vectors	208
7.3.5	Spectral Theorem	214
7.3.6	Hermitian matrix	216
<b>7.4</b>	<b>Assignment Seven</b>	<b>219</b>
<b>8</b>	<b>Week7</b>	<b>223</b>
<b>8.1</b>	<b>Tuesday</b>	<b>223</b>
8.1.1	Quadratic form	223
8.1.2	Convex Optimization Preliminaries	225



8.1.3	Positive Definite Matrices . . . . .	228
<b>8.2</b>	<b>Thursday</b>	<b>237</b>
8.2.1	SVD: Singular Value Decomposition . . . . .	237
8.2.2	Remark on SVD decomposition . . . . .	241
8.2.3	Best Low-Rank Approximation . . . . .	246
<b>8.3</b>	<b>Assignment Eight</b>	<b>250</b>
<b>9</b>	<b>Final Exam . . . . .</b>	<b>253</b>
<b>9.1</b>	<b>Sample Exam</b>	<b>253</b>
<b>9.2</b>	<b>Final Exam</b>	<b>260</b>
<b>10</b>	<b>Solution . . . . .</b>	<b>267</b>
<b>10.1</b>	<b>Assignment Solutions</b>	<b>267</b>
10.1.1	Solution to Assignment One . . . . .	267
10.1.2	Solution to Assignment Two . . . . .	273
10.1.3	Solution to Assignment Three . . . . .	277
10.1.4	Solution to Assignment Four . . . . .	283
10.1.5	Solution to Assignment Five . . . . .	294
10.1.6	Solution to Assignment Six . . . . .	300
10.1.7	Solution to Assignment Seven . . . . .	308
10.1.8	Solution to Assignment Eight . . . . .	318
<b>10.2</b>	<b>Midterm Exam Solutions</b>	<b>325</b>
10.2.1	Sample Exam Solution . . . . .	325
10.2.2	Midterm Exam Solution . . . . .	335
<b>10.3</b>	<b>Final Exam Solutions</b>	<b>343</b>
10.3.1	Sample Exam Solution . . . . .	343
10.3.2	Final Exam Solution . . . . .	354



# Preface

This book is intended for the foundation course MAT2040, which is the first course on the linear algebra. It aims to cover basic linear algebra knowledge and its simple applications. This book was first written in 2017, and it is reviewed and revised in 2018. We have corrected several mistakes shown in the previous book and modified some proofs a little bit to give readers better insights of linear algebra. During the modification, we also refer to many reading materials, which are also recommended for you:

- ENGG 5781 Course Notes by Prof. Wing-Kin (Ken) Ma, CUHK, Hongkong, China, <http://www.ee.cuhk.edu.hk/~wkma/engg5781>
- Roger A. Horn and Charles R. Johnson, *Matrix Analysis* (Second Edition), Cambridge University Press, 2012.
- S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra (Vectors, Matrices, and Least Squares)*, Cambridge University Press, 2018.

The whole book can cover a semester course in a 14-week, each section in which corresponds to a 2-hour lecture. If you read the whole book, and work some mini-exercises, you will learn a lot. We hope you will get the insights on linear algebra and apply them in your own subject.

CUHK(SZ)

August 29, 2018



# Acknowledgments

This book is from the MAT2040 in summer semester, 2017. It is revised in 2018 to correct some mistakes, and revise some proofs to give readers better insights on linear algebra.

CUHK(SZ)





# Notations and Conventions

$\mathbb{R}^n$	$n$ -dimensional real space
$\mathbb{C}^n$	$n$ -dimensional complex space
$\mathbb{R}^{m \times n}$	set of all $m \times n$ real-valued matrices
$\mathbb{C}^{m \times n}$	set of all $m \times n$ complex-valued matrices
$x_i$	$i$ th entry of column vector $\mathbf{x}$
$a_{ij}$	$(i, j)$ th entry of matrix $\mathbf{A}$
$\mathbf{a}_i$	$i$ th column of matrix $\mathbf{A}$
$\mathbf{a}_i^T$	$i$ th row of matrix $\mathbf{A}$
$\mathbb{S}^n$	set of all $n \times n$ real symmetric matrices, i.e., $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $a_{ij} = a_{ji}$ for all $i, j$
$\mathbb{H}^n$	set of all $n \times n$ complex Hermitian matrices, i.e., $\mathbf{A} \in \mathbb{C}^{n \times n}$ and $\bar{a}_{ij} = a_{ji}$ for all $i, j$
$\mathbf{A}^T$	transpose of $\mathbf{A}$ , i.e, $\mathbf{B} = \mathbf{A}^T$ means $b_{ji} = a_{ij}$ for all $i, j$
$\mathbf{A}^H$	Hermitian transpose of $\mathbf{A}$ , i.e, $\mathbf{B} = \mathbf{A}^H$ means $b_{ji} = \bar{a}_{ij}$ for all $i, j$
$\text{trace}(\mathbf{A})$	sum of diagonal entries of square matrix $\mathbf{A}$
$\mathbf{1}$	A vector with all 1 entries
$\mathbf{0}$	either a vector of all zeros, or a matrix of all zeros
$\mathbf{e}_i$	a unit vector with the nonzero element at the $i$ th entry
$\mathcal{C}(\mathbf{A})$	the column space of $\mathbf{A}$
$\mathcal{R}(\mathbf{A})$	the row space of $\mathbf{A}$
$\mathcal{N}(\mathbf{A})$	the null space of $\mathbf{A}$
$\text{Proj}_{\mathcal{M}}(\mathbf{A})$	the projection of $\mathbf{A}$ onto the set $\mathcal{M}$



## 8.2. Thursday

Three ways for matrix decomposition are significant in linear algebra:

$$\left\{ \begin{array}{l} \text{LU (from Gaussian elimination)} \\ \text{QR (from Orthogonalization)} \\ \text{SVD (from eigenvalues and eigenvectors)} \end{array} \right.$$

We have learnt the first two decomposition. And the third way is increasingly significant in the information age.

In the last lecture we learnt that any real symmetric matrix admits *diagonalization*, i.e., *eigendecomposition*. However, can we get some **universal** decomposition, i.e., Is there any decomposition that can be applied to all matrices?

The answer is yes. The key idea behind is to do *symmetrization*. We have to consider  $\mathbf{A}\mathbf{A}^T$  and  $\mathbf{A}^T\mathbf{A}$ .

### 8.2.1. SVD: Singular Value Decomposition

**Theorem 8.3 — SVD.** Given any matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , there exists a 3-tuple  $(\mathbf{U}, \Sigma, \mathbf{V}) \in \mathbb{R}^{m \times m} \times \mathbb{R}^{m \times n} \times \mathbb{R}^{n \times n}$  such that

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

where  $\mathbf{U}, \mathbf{V}$  are **orthogonal**, and  $\Sigma$  takes the form

$$\Sigma_{ij} = \begin{cases} \sigma_i, & i = j \\ 0, & i \neq j \end{cases},$$

with  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$  and with  $p = \min\{m, n\}$ .



- If  $\mathbf{V} = \mathbf{U}$ , this decomposition is exactly **eigen-decomposition**.

- Specifically speaking,
  - $\mathbf{U} \in \mathbb{R}^{m \times m}$  such that its columns are eigenvectors of  $\mathbf{A}\mathbf{A}^T$
  - $\mathbf{V} \in \mathbb{R}^{n \times n}$  such that its columns are eigenvectors of  $\mathbf{A}^T\mathbf{A}$
  - $\Sigma \in \mathbb{R}^{m \times n}$  looks like a diagonal matrix, i.e., it has the form

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_n & \\ 0 & \dots & 0 & \\ \vdots & \ddots & \vdots & \\ 0 & \dots & 0 & \end{pmatrix} \text{ if } m \geq n$$

$$\Sigma = \begin{pmatrix} \sigma_1 & & 0 & \dots & 0 \\ & \ddots & \vdots & \ddots & \vdots \\ & & \sigma_m & 0 & \dots & 0 \end{pmatrix} \text{ if } m < n.$$

with  $\sigma_i = \sqrt{\lambda_i}$  for  $i = 1, 2, \dots, \min\{m, n\}$ , where  $\lambda_i$ 's are eigenvalues of  $\mathbf{A}\mathbf{A}^T$  (if  $m < n$ ) or  $\mathbf{A}^T\mathbf{A}$ . (if  $m \geq n$ )

**Definition 8.7** [SVD] The above decomposition is called the **singular value decomposition (SVD)**

- $\sigma_i$  is called the  $i$ th **singular value**
- The columns of  $\mathbf{U}$  and  $\mathbf{V}$ ,  $\mathbf{u}_i$  and  $\mathbf{v}_i$  are called the  $i$ th **left and right singular vectors**, respectively.
- $(\sigma_i, \mathbf{u}_i)$  are the eigen-pairs of  $\mathbf{A}\mathbf{A}^T$ ;  $(\sigma_i, \mathbf{v}_i)$  are the eigen-pairs of  $\mathbf{A}^T\mathbf{A}$  for  $i = 1, 2, \dots, \min\{m, n\}$ .
- The following notations may be used to denote the singular values of  $\mathbf{A}$ :

$$\sigma_{\max}(\mathbf{A}) \geq \sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \dots \geq \sigma_p(\mathbf{A}) = \sigma_{\min}(\mathbf{A})$$

The proof for the SVD decomposition is constructive. To see the insights of the proof, let's study the case  $m = n$  first, then we extend the proof for general case:

**Proposition 8.2** SVD always exists for any **real square nonsingular** matrix.

*Proof.* For  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , you may verify that  $\mathbf{A}\mathbf{A}^T$  is PD, thus it admits the eigen-decomposition:

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T, \text{ with } \lambda_1 \geq \dots \geq \lambda_n > 0. \quad (8.8)$$

We define  $\mathbf{\Sigma} := \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$  and  $\mathbf{V} := \mathbf{A}^T \mathbf{U} \mathbf{\Sigma}^{-1}$ .

You may verify that  $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{A}$  and  $\mathbf{V}^T \mathbf{V} = \mathbf{I}$ , i.e.,  $\mathbf{V}$  is orthogonal. The proof is complete. ■

**Proposition 8.3** SVD always exists for any **real** matrix.

*Proof.* • Firstly, consider the matrix product  $\mathbf{A}\mathbf{A}^T$ . Since  $\mathbf{A}\mathbf{A}^T \in \mathbb{S}^m$  and  $\mathbf{A} \succeq 0$ , we can decompose  $\mathbf{A}\mathbf{A}^T$  as

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T & \mathbf{u}_2^T \end{bmatrix} = \mathbf{u}_1 \tilde{\mathbf{\Sigma}} \mathbf{u}_1^T \quad (8.9)$$

where:

- we assume that the eigenvalues are ordered, i.e.,

$$\lambda_1 \geq \dots \geq \lambda_r > 0, \text{ and } \lambda_{r+1} = \dots = \lambda_p = 0$$

with  $r$  being the number of nonzero eigenvalues

- $\mathbf{U} \in \mathbb{R}^{m \times m}$  denotes an orthogonal matrix, and its columns are the corresponding eigenvectors
- We partition  $\mathbf{U}$  as

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}, \quad \mathbf{u}_1 \in \mathbb{R}^{m \times r}, \mathbf{u}_2 \in \mathbb{R}^{m \times (m-r)},$$

and  $\tilde{\mathbf{\Sigma}} = \text{diag}(\lambda_1, \dots, \lambda_r)$ .

- Secondly, we show that

$$\mathbf{U}_2^T \mathbf{A} = \mathbf{0} \quad (8.10)$$

Since  $\mathbf{U}$  is orthogonal, we obtain:

$$\mathbf{U}^T \mathbf{U} = \begin{bmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{U}_1^T \mathbf{U}_1 & \mathbf{U}_1^T \mathbf{U}_2 \\ \mathbf{U}_2^T \mathbf{U}_1 & \mathbf{U}_2^T \mathbf{U}_2 \end{bmatrix} = \mathbf{I} \implies \mathbf{U}_2^T \mathbf{U}_1 = \mathbf{0}.$$

Substituting Eq.(8.9) into  $\mathbf{U}_2^T \mathbf{A} (\mathbf{U}_2^T \mathbf{A})^T$ , we obtain:

$$\mathbf{U}_2^T \mathbf{A} (\mathbf{U}_2^T \mathbf{A})^T = (\mathbf{U}_2^T \mathbf{U}_1) \tilde{\Sigma} \mathbf{U}_1^T \mathbf{U}_2 = \mathbf{0} \quad (8.11)$$

By Eq.(8.11) and the simple result that  $\mathbf{B} \mathbf{B}^T = \mathbf{0}$  implies  $\mathbf{B} = \mathbf{0}$  (write  $\mathbf{B}$  into column vectors form to verify it), we conclude that  $\mathbf{U}_2 \mathbf{A} = \mathbf{0}$

- Thirdly, we construct the following matrices:

$$\hat{\Sigma} = \tilde{\Sigma}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r}), \quad \mathbf{V}_1 = \mathbf{A}^T \mathbf{U}_1 \hat{\Sigma}^{-1} \in \mathbb{R}^{n \times r}.$$

Combining it with Eq.(8.9), we can verify that  $\mathbf{V}_1^T \mathbf{V}_1 = \mathbf{I}$ . Furthermore, there exists a matrix  $\mathbf{V}_2 \in \mathbb{R}^{n \times (n-r)}$  such that  $\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix}$  is orthogonal. Moreover, we can verify that

$$\mathbf{U}_1^T \mathbf{A} \mathbf{V}_1 = \hat{\Sigma}, \quad \mathbf{U}_1^T \mathbf{A} \mathbf{V}_2 = \mathbf{0} \quad (8.12)$$

- Fourthly, consider the matrix product  $\mathbf{U}^T \mathbf{A} \mathbf{V}$ . From Eq.(8.12) and Eq.(8.10), we have

$$\begin{aligned} \mathbf{U}^T \mathbf{A} \mathbf{V} &= \begin{bmatrix} \mathbf{U}_1^T \mathbf{A} \mathbf{V}_1 & \mathbf{U}_1^T \mathbf{A} \mathbf{V}_2 \\ \mathbf{U}_2^T \mathbf{A} \mathbf{V}_1 & \mathbf{U}_2^T \mathbf{A} \mathbf{V}_2 \end{bmatrix} \\ &= \begin{bmatrix} \hat{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} := \Sigma \end{aligned}$$

By multiplying the above equation on the left by  $\mathbf{U}$  and on the right by  $\mathbf{V}^T$ , we



obtain the desired result  $A = \mathbf{U}\Sigma\mathbf{V}^T$ . The proof is complete. ■

## 8.2.2. Remark on SVD decomposition

### 8.2.2.1. Remark 1: Different Ways of Writing out SVD

**Definition 8.8** [Partitioned form of SVD] let  $r$  be the number of nonzero singular values, and note that  $\sigma_1 \geq \dots \geq \sigma_r > 0, \sigma_{r+1} = \dots = \sigma_p = 0$ . Then from the standard form, we derive the **partitioned form of SVD**:

$$\mathbf{A} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \tilde{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} \quad (8.13)$$

where:

- $\tilde{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_r)$
- $\mathbf{U}_1 = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_r \end{bmatrix} \in \mathbb{R}^{m \times r}, \mathbf{U}_2 = \begin{bmatrix} \mathbf{u}_{r+1} & \dots & \mathbf{u}_m \end{bmatrix} \in \mathbb{R}^{m \times (m-r)}$
- $\mathbf{V}_1 = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_r \end{bmatrix} \in \mathbb{R}^{n \times r}, \mathbf{V}_2 = \begin{bmatrix} \mathbf{v}_{r+1} & \dots & \mathbf{v}_n \end{bmatrix} \in \mathbb{R}^{n \times (n-r)}$

Note that  $\mathbf{U}_1, \mathbf{U}_2, \mathbf{V}_1, \mathbf{V}_2$  are semi-orthogonal, i.e., they all have orthonormal columns. ■

**Definition 8.9** [Thin SVD] We can re-write Eq.(8.13) as the **thin form of SVD**:

$$\mathbf{A} = \mathbf{U}_1 \tilde{\Sigma} \mathbf{V}_1^T \quad (8.14)$$

**Definition 8.10** [Outer-product form] By expanding the Eq.(8.14), we derive the **outer-product form of SVD**:

$$\mathbf{A} = \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad (8.15)$$

### 8.2.2.2. Remark 2: SVD and Eigen-decomposition

The eigenvalues for  $\mathbf{A}^T \mathbf{A}$  and  $\mathbf{A} \mathbf{A}^T$  are the same for first  $p$  terms.

**Proposition 8.4** Suppose  $\mathbf{A}$  admits the SVD  $\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T$ , then we have:

$$\mathbf{A} \mathbf{A}^T = \mathbf{U} \mathbf{D}_1 \mathbf{U}^T, \quad \mathbf{D}_1 = \Sigma \Sigma^T = \text{diag}(\sigma_1^2, \dots, \sigma_p^2, \underbrace{0, \dots, 0}_{m-p \text{ zeros}}) \quad (8.16)$$

$$\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{D}_2 \mathbf{V}^T, \quad \mathbf{D}_2 = \Sigma^T \Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2, \underbrace{0, \dots, 0}_{n-p \text{ zeros}}) \quad (8.17)$$

*Proof.* Just apply the SVD form and the orthogonality of  $\mathbf{U}$  and  $\mathbf{V}$ . ■

### 8.2.2.3. Remark 3: SVD and Subspace

We are curious about how many singular values of  $\mathbf{A}$  are nonzero.

**Proposition 8.5** The following properties hold:

1.  $\mathcal{C}(\mathbf{A}) = \mathcal{C}(\mathbf{U}_1), \mathcal{C}(\mathbf{A})^\perp = \mathcal{C}(\mathbf{U}_2);$
2.  $\mathcal{C}(\mathbf{A}^T) = \mathcal{C}(\mathbf{V}_1), \mathcal{C}(\mathbf{A}^T)^\perp = \mathcal{N}(\mathbf{A}) = \mathcal{C}(\mathbf{V}_2);$
3.  $\text{rank}(\mathbf{A}) = r$ , i.e., the number of nonzero singular values.

*Proof.* The above properties are easily seen to be true using SVD. Also, you should apply the definition for column space and null space. You should verify these properties by yourself.

$$\mathcal{C}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = \mathbf{A} \mathbf{x}, \mathbf{x} \in \mathbb{R}^n\} \quad (8.18a)$$

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A} \mathbf{x} = \mathbf{0}\} \quad (8.18b)$$

For the third part of proposition(8.5), since  $\text{rank}(\mathbf{A}) = \dim(\mathcal{C}(\mathbf{A})) = \dim(\mathcal{C}(\mathbf{U}_1))$ , and  $\mathbf{U}_1$  has  $r$  orthonormal columns, we derive that  $\dim(\mathcal{C}(\mathbf{U}_1)) = r = \text{rank}(\mathbf{A})$ . ■

For the SVD decomposition

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

we can convert it into the following two forms:

$$AV = U\Sigma V^T V = U\Sigma$$

$$A = U\Sigma V^T \implies A^T = V\Sigma U^T \implies A^T U = V\Sigma U^T U = V\Sigma.$$

If we write it into vector forms, we obtain:

$$\begin{cases} Av_j = \sigma_j u_j \\ A^T u_j = \sigma_j v_j \end{cases}, \quad j = 1, 2, \dots, r. \quad (8.19)$$

The columns of  $U$  ( $u_j$ ) are called the **left singular vector** of  $A$ ; the columns of  $V$  ( $v_j$ ) are called the **right singular vector** of  $A$ ;  $\sigma_j$  is called the **singular value**.

We can easily understand the proposition(8.5) and Eq.(8.19) by the following graph:

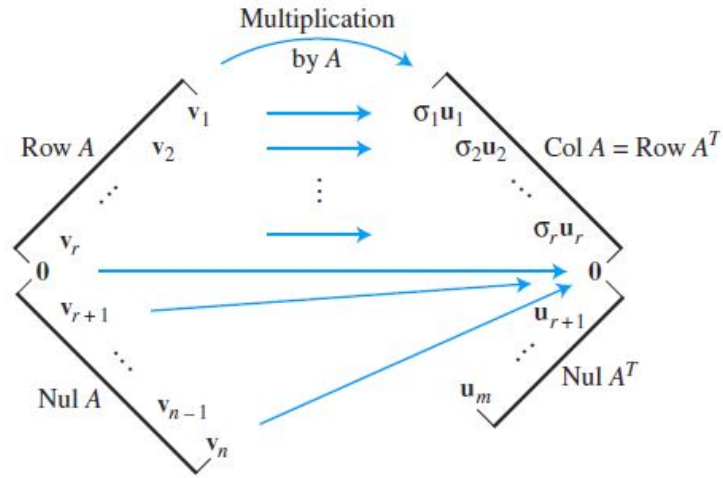


Figure 8.2: The fundamental spaces and the action of  $A$ .

#### Explanation:

- When  $\{v_1, \dots, v_r\}$  are multiplied by  $A$ , they are converted into  $\{\sigma_1 u_1, \dots, \sigma_r u_r\}$ ; when  $\{v_{r+1}, \dots, v_n\}$  are multiplied by  $A$ , they are converted into  $0$ .
- The first  $r$  columns of  $V$  forms the basis for the row space of  $A$ , i.e.,  $\mathcal{C}(V_1) = \mathcal{C}(A^T)$ .
- The last  $n - r$  columns of  $V$  forms the basis for the null space of  $A$ , i.e.,  $\mathcal{C}(V_2) =$

$\mathcal{N}(\mathbf{A})$ .

- The first  $r$  columns of  $\mathbf{U}$  forms the basis for the column space of  $\mathbf{A}$ , i.e.,  $\mathcal{C}(\mathbf{U}_1) = \mathcal{C}(\mathbf{A})$ .
- The last  $m - r$  columns of  $\mathbf{U}$  forms the basis for the null space of  $\mathbf{A}^\top$ , i.e.,  $\mathcal{C}(\mathbf{U}_2) = \mathcal{N}(\mathbf{A}^\top)$

Recall the outer-product form of SVD,

$$\mathbf{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^\top$$

where  $r = \text{rank}(\mathbf{A}) = \text{number of nonzero singular values}$ , which is the third meaning for the rank:

**R** Up till now,  $\text{rank}(\mathbf{A})$  has three meanings:

- $\text{rank}(\mathbf{A}) = \dim(\text{row}(\mathbf{A}))$
- $\text{rank}(\mathbf{A}) = \dim(\text{col}(\mathbf{A}))$
- $\text{rank}(\mathbf{A}) = \text{number of nonzero singular values of } \mathbf{A}$ .

**R** However,  $\text{rank}(\mathbf{A}) \neq \text{number of nonzero eigenvalues}$ . Let me raise a counterexample:

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

then eigenvalues are  $\lambda_1 = \lambda_2 = 0$ , and  $\text{rank}(\mathbf{A}) = 1$ .

**R** Also, note that many properties can be easily proved by **thin** or **outer-product** form of SVD. For example,  $\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ . If you have no ideas of a proof in exam, you may try SVD.

### 8.2.2.4. Compact SVD

Due to the outer-product form of SVD, i.e., any matrix with rank  $r$  can be factorized into

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\Sigma\mathbf{V}^T \\ &= \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_r \end{bmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix}, \end{aligned}$$

we obtain the following corollary:

**Corollary 8.2** Every rank  $r$  matrix can be written as the sum of  $r$  rank 1 matrices. Moreover, these matrices could be perpendicular!

What's the meaning of perpendicular?

**Definition 8.11** [perpendicular for matrix] For two real  $n \times n$  matrix  $\mathbf{A}$  and  $\mathbf{B}$ , they are said to be **perpendicular (orthogonal)** if the inner product between  $\mathbf{A}$  and  $\mathbf{B}$  is zero:

$$\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{B}^T \mathbf{A}) = \sum_{i,j=1}^n \mathbf{A}_{ij} \mathbf{B}_{ij} = 0.$$

Decompose  $\mathbf{A} := \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ . If we set  $\mathbf{A}_i = \mathbf{u}_i \mathbf{v}_i^T \sigma_i$ , let's show  $\mathbf{A}_i$ 's are perpendicular:

$$\begin{aligned} \langle \mathbf{A}_i, \mathbf{A}_j \rangle &= \text{trace}(\mathbf{A}_j^T \mathbf{A}_i) \\ &= \text{trace}(\sigma_i \sigma_j \mathbf{v}_j \mathbf{u}_j^T \mathbf{u}_i \mathbf{v}_i^T) = \sigma_i \sigma_j \text{trace}(\mathbf{v}_j \mathbf{u}_j^T \mathbf{u}_i \mathbf{v}_i^T) \\ &= \sigma_i \sigma_j \text{trace}(\mathbf{v}_j (\mathbf{u}_j^T \mathbf{u}_i) \mathbf{v}_i^T) = \sigma_i \sigma_j \text{trace}(\mathbf{v}_j \mathbf{0} \mathbf{v}_i^T) \\ &= 0. \end{aligned}$$

How many rank 1 matrices do we need to pick to construct matrix  $\mathbf{A}$ ? In fact, this

number has no upper bound. For example, if we obtain

$$\mathbf{A} = \mathbf{u}_1 \mathbf{v}_1^T + \mathbf{u}_2 \mathbf{v}_2^T$$

Then we can always decompose any rank 1 matrix into 2 rank 1 matrices:

$$\mathbf{A} = \mathbf{u}_1 \mathbf{v}_1^T + \frac{1}{2} \mathbf{u}_2 \mathbf{v}_2^T + \frac{1}{2} \mathbf{u}_2 \mathbf{v}_2^T.$$

But this number has a lower bound, that is rank. In other words,  $\text{rank}(\mathbf{A}) =$  smallest number of rank 1 matrices with sum  $\mathbf{A}$ .

### 8.2.3. Best Low-Rank Approximation

Given matrix  $\mathbf{A}$ . What is the *best rank  $k$  approximation*? In other words, given matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , what is the optimal solution for the optimization:

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{A} - \mathbf{Z}\|_F^2 \\ \text{s.t.} \quad & \text{rank}(\mathbf{Z}) = k \\ & \mathbf{Z} \in \mathbb{R}^{m \times n} \end{aligned}$$

Firstly let's introduce the definition for Frobenius norm:

**Definition 8.12** [Frobenius norm] The Frobenius norm for  $m \times n$  matrix  $\mathbf{A}$  is given by

$$\|\mathbf{A}\|_F = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle} = \sqrt{\text{trace}(\mathbf{A}^T \mathbf{A})}.$$

**Theorem 8.4** Suppose the SVD for  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is given by

$$\mathbf{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T.$$



with  $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$ .

Then the best rank  $k$  ( $k \leq r$ ) approximation of  $\mathbf{A}$  is

$$\mathbf{A}_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \dots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T.$$

For example,  $\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T$  is the best rank 1 approximation of  $\mathbf{A}$ .

### 8.2.3.1. Analogy with least square problem

For least squares problem, the key is to do approximation for  $\mathbf{b} \in \mathbb{R}^m$ . In other words, we just do a projection from  $\mathbf{b}$  to the plane  $\{\mathbf{Ax} | \mathbf{x} \in \mathbb{R}^n\}$ :

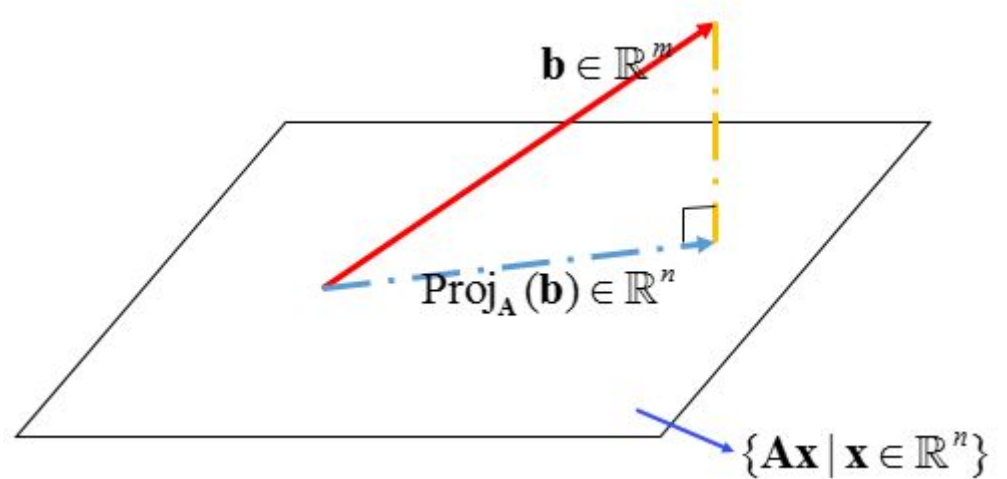


Figure 8.3: Least square problem: find  $\mathbf{x}$  such that  $\mathbf{Ax} = \text{Proj}_{\mathcal{C}(\mathbf{A})}(\mathbf{b})$ .

**R** For the least squares problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{Ax} - \mathbf{b}\|^2 \\ \text{s.t.} \quad & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

with  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{b} \in \mathbb{R}^m$ , the key is to do the projection of  $\mathbf{b}$  onto  $\mathcal{C}(\mathbf{A})$ , thus it suffices to solve the **equality**

$$\mathbf{Ax} = \text{Proj}_{\mathcal{C}(\mathbf{A})}(\mathbf{b}).$$

Similarly, the best rank  $k$  approximation could be viewed as a projection from  $\mathbf{A}$  with rank  $r$  to the “plane” that contains all rank  $k$  matrices:

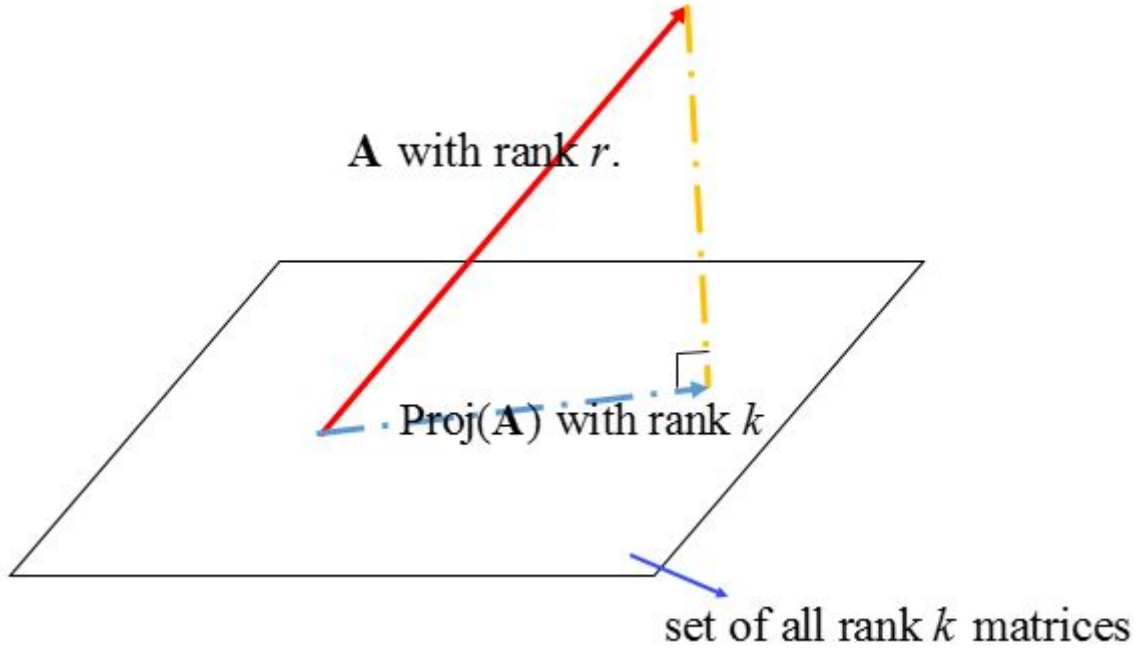


Figure 8.4: Best rank  $k$  approximation: find the projection from matrix  $\mathbf{A}$  with rank  $r$  onto the plane that contains all rank  $k$  matrices

Ⓡ Similarly, for the best rank  $k$  approximation problem

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{A} - \mathbf{Z}\|_F^2 \\ \text{s.t.} \quad & \text{rank}(\mathbf{Z}) = k \\ & \mathbf{Z} \in \mathbb{R}^{m \times n} \end{aligned}$$

with  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , the key is to do the projection of  $\mathbf{A}$  onto the set  $\mathcal{M} = \{\mathbf{M} \in \mathbb{R}^{m \times n} \mid \text{rank}(\mathbf{M}) = k\}$ , thus it suffices to solve the **equality**

$$\mathbf{Z} = \text{Proj}_{\mathcal{M}}(\mathbf{A}).$$

For some non-convex optimization problems, this idea is very useful. The

further reading is recommended:

Jain, Prateek, and P. Kar. "Non-convex Optimization for Machine Learning." *Foundations & Trends® in Machine Learning* 10.3-4(2017):142-336.

