

# Numerical Linear Algebra

Chenghao Dong

November 15, 2023

# CONTENTS

<b>1</b>	<b>Basic LA Review</b>	<b>2</b>
1.1	Notations and Basic Concepts . . . . .	2
1.2	Matrices Calculations . . . . .	3
1.3	Matrix Norms . . . . .	5
1.4	Structured Matrices and Common Theories . . . . .	8
1.5	Subspaces and Dimensions . . . . .	10
<b>2</b>	<b>Singular Value Decomposition</b>	<b>12</b>
2.1	Singular Values . . . . .	12
2.2	Singular Value Decomposition . . . . .	13
2.3	Low-rank approximation via truncated SVD . . . . .	16
2.4	Courant-Fisher Minimax Theorem . . . . .	18
<b>3</b>	<b>Solving Linear Systems</b>	<b>21</b>
3.1	LU Factorization . . . . .	21
3.2	Algorithm for LU Factorizations . . . . .	24
3.3	Further Improvements for LU Factorizations . . . . .	26

# 1 Basic LA Review

## 1.1 Notations and Basic Concepts

在本笔记中, 规定如下记号 (其他常见记号比如  $A^{-1}$ ,  $A^T \dots$  在此处省略); 此外, 未经特殊说明, 所有的向量  $\mathbf{x} \in \mathbb{R}^n$  均指代列向量  $\mathbf{x} \in \mathbb{R}^{n \times 1}$

- 1)  $\lambda_i(A)$  表示按照自然顺序 (递减) 排列的第  $i$  个最大的特征值。  $\lambda(A)$  表示矩阵  $A$  的所有特征值;
- 2)  $\sigma_i(A)$  表示第  $i$  个 (最大) 的奇异值。  $\sigma(A)$  表示矩阵  $A$  的所有奇异值;
- 3)  $\text{diag}(A)$  表示矩阵  $A$  的对角元素;  $\text{span}(A)$  表示矩阵  $A$  的列空间;
- 4)  $\overline{A}$  表示矩阵  $A$  的共轭阵 (conjugate) s.t.  $(\overline{A})_{ij} = \overline{A_{ij}}$ ;
- 5)  $A^H$  表示矩阵  $A$  的共轭转置 (conjugate transpose) s.t.  $A^H = (\overline{A})^T = \overline{A^T}$ ;
- 6)  $\|\mathbf{x}\|_p$  表示向量  $\mathbf{x}$  的  $p$ -范数, 当  $p = 2$  时, 该范数为常见的欧氏范数 (Euclidean norm), 简单记作  $\|\mathbf{x}\| := \|\mathbf{x}\|_2 = \mathbf{x}^H \mathbf{x}$ ;
- 7)  $\|A\|_p$  表示矩阵  $A$  的诱导  $p$ -范数 (Induced  $p$ -norm), 当  $p = 2$  时, 该范数称为矩阵的谱范数 (spectral norm), 简单记作  $\|A\| := \|A\|_2$ ;
- 8)  $\|A\|_F$  表示矩阵  $A$  的诱导  $F$ -范数 (Frobenius norm);
- 9)  $A \succ 0$  表示矩阵  $A$  为正定矩阵 (positive definite);  $A \succ B$  表示  $A - B \succ 0$  (resp.  $\prec$  表示负定);
- 10)  $A \succeq 0$  表示矩阵  $A$  为半正定矩阵 (positive semi-definite);  $A \succeq B$  表示  $A - B \succeq 0$  (resp.  $\preceq$  表示半负定);
- 11)  $\Lambda/U/L$  未经特殊说明  $\Lambda$  均指代对角矩阵;  $U$  均指代上三角矩阵;  $L$  均指代下三角矩阵;
- 12)  $a_{i.}^T / a_{.j}$  未经特殊说明  $a_{i.}^T = e_i^T A$  均指代矩阵  $A$  的第  $i$  行;  $a_{.j} = A e_j$  均指代矩阵  $A$  的第  $j$  列,  $a_{i.}, a_{.j}$  均视为列向量; 所以一个  $m \times n$  矩阵  $A$  可以视为:

$$A = \begin{bmatrix} a_{1.}^T \\ \vdots \\ a_{m.}^T \end{bmatrix} \quad \text{or} \quad [a_{.1} \quad \cdots \quad a_{.n}]$$

- 13)  $\odot$  表示对应位置的 element-wise 点乘; 与一般的矩阵或向量乘区分;
- 14)  $|a|$  表示  $a \in \mathbb{C}$  的模 modulus, 当  $a$  为实数时等于绝对值。

## 1.2 Matrices Calculations

下面回顾一些常见的**矩阵运算的通式**，主要涉及四类常见的运算：

- 1)  $AB$  表示矩阵乘法，对于  $A \in \mathbb{F}^{m \times l}$ ,  $B \in \mathbb{F}^{l \times n}$  其通式

$$(AB)_{ij} = \sum_{k=1}^l a_{ik} b_{kj}$$

此外通过将矩阵进行行列拆分，矩阵乘法还可以化为

$$AB = A \begin{bmatrix} b_{\cdot 1} & \cdots & b_{\cdot n} \end{bmatrix} = \begin{bmatrix} Ab_{\cdot 1} & \cdots & Ab_{\cdot n} \end{bmatrix} \quad \text{s.t. } Ab_{\cdot j} \in \mathbb{F}^m$$

或者

$$AB = \begin{bmatrix} a_{1\cdot}^T \\ \vdots \\ a_{m\cdot}^T \end{bmatrix} B = \begin{bmatrix} a_{1\cdot}^T B \\ \vdots \\ a_{m\cdot}^T B \end{bmatrix} \quad \text{s.t. } a_{i\cdot}^T B \in \mathbb{F}^{1 \times n}$$

这分别说明了  $AB$  的列是  $A$  的列的线性组合；行是  $B$  的行的线性组合，于是显然有  $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$ ；如果  $A = \Lambda \in \mathbb{F}^{l \times l}$  为对角阵， $\Lambda B$  相当于给  $B$  的**行**乘上对角阵中对应位置的值

$$\Lambda B = \begin{bmatrix} \Lambda b_{\cdot 1} & \cdots & \Lambda b_{\cdot n} \end{bmatrix} = \begin{bmatrix} \lambda_1 b_{1\cdot}^T \\ \vdots \\ \lambda_l b_{l\cdot}^T \end{bmatrix} \quad \text{s.t. } \Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_l \end{bmatrix}$$

如果  $B = \Lambda \in \mathbb{F}^{l \times l}$  为对角阵， $A\Lambda$  相当于给  $A$  的**列**乘上对角阵中对应位置的值

$$A\Lambda = \begin{bmatrix} a_{1\cdot}^T \Lambda \\ \vdots \\ a_{m\cdot}^T \Lambda \end{bmatrix} = \begin{bmatrix} \lambda_1 a_{1\cdot} & \cdots & \lambda_l a_{1\cdot} \end{bmatrix} \quad \text{s.t. } \Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_l \end{bmatrix}$$

- 2)  $\mathbf{y}^T A \mathbf{x}$  双线性型 (bilinear form) 常出现于二次型运算，其结果为一常数，有通式

$$\mathbf{y}^T A \mathbf{x} = \sum_{ij} y_i a_{ij} x_j$$

- 3)  $X\Lambda Y$  在  $X \in \mathbb{F}^{m \times l}$ ,  $\Lambda \in \mathbb{F}^{l \times l}$ ,  $Y \in \mathbb{F}^{l \times n}$ ，其中  $\Lambda$  为对角矩阵时，结果是一个  $\mathbb{F}^{m \times n}$  中的矩阵，且可以分解为  $X$  的列  $\mathbf{x}_{\cdot k}$  与  $Y$  中对应行  $\mathbf{y}_{k\cdot}^T$  对于相应对角值  $\lambda_k$  的加权和

$$X\Lambda Y = \begin{bmatrix} \mathbf{x}_{\cdot 1} & \cdots & \mathbf{x}_{\cdot l} \end{bmatrix} \Lambda Y = \begin{bmatrix} \mathbf{x}_{\cdot 1} & \cdots & \mathbf{x}_{\cdot l} \end{bmatrix} \begin{bmatrix} \lambda_1 \mathbf{y}_{1\cdot}^T \\ \vdots \\ \lambda_l \mathbf{y}_{l\cdot}^T \end{bmatrix} = \sum_{k=1}^l \lambda_k \mathbf{x}_{\cdot k} \mathbf{y}_{k\cdot}^T$$

其中第  $(i, j)$  个元素等于  $(X\Lambda Y)_{ij} = \sum_{k=1}^l \lambda_k x_{ik} y_{kj} = \mathbf{x}_{i\cdot}^T \Lambda \mathbf{y}_{\cdot j}$

- 4)  $XAY$  在  $X \in \mathbb{F}^{m \times l}$ ,  $A \in \mathbb{F}^{l \times k}$ ,  $Y \in \mathbb{F}^{k \times n}$  时, 结果是一个  $\mathbb{F}^{m \times n}$  中的矩阵, 且第  $(i, j)$  个元素等于  $X$  的  $i$  行  $\mathbf{x}_{i.}^T$  与  $Y$  中  $j$  列  $\mathbf{y}_{.j}$  对于矩阵  $A$  的双线性型:

$$(XAY)_{ij} = \mathbf{x}_{i.}^T A \mathbf{y}_{.j}$$

一个简单的证明如下

$$XAY = \begin{bmatrix} \mathbf{x}_{1.}^T \\ \vdots \\ \mathbf{x}_{m.}^T \end{bmatrix} \begin{bmatrix} A\mathbf{y}_{.1} & \cdots & A\mathbf{y}_{.n} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{1.}^T A\mathbf{y}_{.1} & \cdots & \mathbf{x}_{1.}^T A\mathbf{y}_{.n} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{m.}^T A\mathbf{y}_{.1} & \cdots & \mathbf{x}_{m.}^T A\mathbf{y}_{.n} \end{bmatrix}$$

- 5)  $\text{trace}(AB)$  若  $AB$  为方阵, 则不难得出:  $\text{trace}(AB) = \text{trace}(BA)$
- 6) Neumann series 只要方阵  $\|X\| < 1$  对任何一项范数成立, 则在该范数下:

$$(I - X)^{-1} = 1 + X + X^2 + X^3 + \cdots$$

- 7) Matrix Exponentials 定义矩阵指数  $e^X$  s.t.  $X \in \mathbb{F}^{n \times n}$  为方阵, 则

$$e^X = 1 + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \cdots$$

这一般可以通过对角化简化运算。

- 8) 规范正交化 Othonormalization 如果一个矩阵满足  $A^H A = \Lambda$  ( $\Lambda$  为对角阵且对角线元素不为 0), 说明其各列相互正交。此时存在一个对角矩阵  $\Sigma$ , 与一个列规范正交矩阵  $\tilde{A}$  s.t.  $\tilde{A}^H \tilde{A} = I$ , 使得  $A = \tilde{A} \Sigma$ , 即

$$\begin{bmatrix} a_{.1} & \cdots & a_{.n} \end{bmatrix} = \begin{bmatrix} \frac{a_{.1}}{\|a_{.1}\|} & \cdots & \frac{a_{.n}}{\|a_{.n}\|} \end{bmatrix} \begin{bmatrix} \|a_{.1}\| & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|a_{.n}\| \end{bmatrix}$$

这对于行正交矩阵同理, 只不过变成左乘系数矩阵。

## 1.3 Matrix Norms

接着我们定义几个**常见的向量与矩阵范数**，从范数的基本定义开始。

**Definition 1.3.1. 范数 (norms)** 线性空间  $(V, \mathbb{F})$  里的一个范数  $\|\cdot\|$  应当满足如下几条性质

- 1)  $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$
- 2)  $\|\mathbf{0}\| \geq 0$  and  $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$
- 3)  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

**Definition 1.3.2. 向量的 p-范数 (p-norms)** 矢量  $\mathbf{x} \in \mathbb{C}^n$  的  $p$ -范数定义为，对  $1 \leq p \leq \infty$

$$\|\mathbf{x}\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad \text{or} \quad \|\mathbf{x}\|_\infty = \max_i |x_i|$$

对于复数来说， $|x_i|$  表示其模长 *modulus*。当  $p = 2$  时，该范数为常见的欧氏范数 (*Euclidean norm*)，省略下标简单记作  $\|\mathbf{x}\| := \|\mathbf{x}\|_2 = \mathbf{x}^T \mathbf{x}$ 。

**Definition 1.3.3. 矩阵的诱导 p-范数 (induced p-norms)** 矩阵  $A$  的诱导  $p$ -范数定义为，对  $1 \leq p \leq \infty$

$$\|A\|_p = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = \max_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p$$

当  $p = 2$  时，该范数称为矩阵的谱范数 (*spectral norm*)，省略下标简单记作  $\|A\| := \|A\|_2 = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$ 。诱导范数实际上衡量了矩阵在对应向量范数下能对其施加的最大影响。

**Definition 1.3.4. 矩阵的 F-范数 (F-norms)** 矩阵  $A$  的  $F$ -范数 (*Frobenius norm*) 定义为

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\text{trace}(A^H A)}$$

**Definition 1.3.5. 矩阵的迹范数/核范数 (trace/nuclear-norms)** 矩阵  $A$  的迹范数/核范数定义为

$$\|A\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(A) = \max_{Q \text{ orthonormal}} \{\text{trace}(Q^T A)\}$$

**Definition 1.3.6. 酉不变范数 (unitarily invariant)** 向量的一个范数  $\|\cdot\|$  是酉不变的当且仅当对于任意酉矩阵  $A^H A = A A^H = I$ , 其满足  $\forall \mathbf{x}: \|A\mathbf{x}\| = \|\mathbf{x}\|$ ; 矩阵的一个范数  $\|\cdot\|$  是酉不变的当且仅当对于任意酉矩阵  $U, V$  (注意这里的  $U$  不是上三角), 其满足  $\forall A: \|UAV\| = \|A\|$

### Theorem 1.3.1. 常见向量范数性质

- 1)  $\frac{1}{n^{1/p}} \|\mathbf{x}\|_p \leq \|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_p$  证明仅使用了简单的放缩, 略;
- 2)  $\frac{1}{\sqrt{n}} \|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$  前一个不等号使用 *Cauchy-Schwarz*, 后一个不等号直接平方加多项式定理展开。

### Theorem 1.3.2. 常见矩阵范数性质 对于矩阵范数与 $A \in \mathbb{C}^{m \times n}$ :

- 1)  $\|AB\mathbf{x}\|_p \leq \|A\|_p \|B\mathbf{x}\|_p$  (定义)
- 2)  $\|AB\|_p \leq \|A\|_p \|B\|_p$ ;  $\|AB\|_F \leq \|A\|_F \|B\|_F$
- 3)  $\|A\|_1 = \max_j \sum_i |A_{ij}|$  为最大的列元素模和;
- 4)  $\|A\|_\infty = \max_i \sum_j |A_{ij}| = \|A^T\|_1$  为最大的行元素模和;
- 5)  $\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_2 \leq \sqrt{m} \|A\|_\infty$  or  $\frac{1}{\sqrt{m}} \|A\|_1 \leq \|A\|_2 \leq \sqrt{n} \|A\|_1$
- 6)  $\|A\|_2 \leq \|A\|_F \leq \sqrt{\min(m, n)} \|A\|_2$
- 7)  $\|A\|_2 = \|A^T\|_2$ ;  $\|A\|_F = \|A^T\|_F$

### Theorem 1.3.3. 常见范数的酉不变性

向量的  $\|\mathbf{x}\|_2$ , 矩阵的  $\|A\|_2, \|A\|_F, \|A\|_*$  均为酉不变范数; 特别地,  $\|W\mathbf{x}\|_2 = \|\mathbf{x}\|_2$  对列规范正交矩阵  $W$  s.t.  $W^H W = I$  也同样成立。

我们简单证明几个结论:

#### Proof Theorem 1.2.2. (2) 对于第一个不等式有

$$\forall \|\mathbf{x}\|_p = 1 : \|AB\mathbf{x}\|_p \leq \|A\|_p \|B\mathbf{x}\|_p \leq \|A\|_p \|B\|_p$$

由最大值或上确界的定义:  $\|AB\|_p = \sup_{\|\mathbf{x}\|_p=1} \|AB\mathbf{x}\|_p \leq \|A\|_p \|B\|_p$

对于第二个不等式有

$$\begin{aligned} \|AB\|_F^2 &= \sum_{i,j} |a_{i.}^T b_{.j}|^2 \leq \sum_{i,j} \|a_{i.}\|^2 \cdot \|b_{.j}\|^2 \\ &= \sum_i \|a_{i.}\|^2 \cdot \sum_j \|b_{.j}\|^2 = \|A\|_F^2 \|B\|_F^2 \end{aligned}$$

(3) 记  $\gamma = \operatorname{argmax}_j \sum_i |A_{ij}|$ ,  $\Gamma = \max_j \sum_i |A_{ij}|$ , 我们先证:  $\forall \|\mathbf{x}\|_1 = 1 : \|A\mathbf{x}\|_1 \leq \Gamma$ , 即

$$\begin{aligned} \|A\mathbf{x}\|_1 &= \sum_i \left| \sum_j A_{ij} x_j \right| = \sum_i \left| \sum_j A_{ij} x_j \right| \\ &\leq \sum_i \sum_j |A_{ij}| |x_j| = \sum_j |x_j| \sum_i |A_{ij}| \\ &\leq \sum_j |x_j| \cdot \Gamma = \Gamma \sum_j |x_j| = \Gamma \|\mathbf{x}\|_1 = \Gamma \end{aligned}$$

于是  $\|A\|_1 = \max_{\|\mathbf{x}\|_1=1} \|A\mathbf{x}\|_1 \leq \Gamma$ , 而显然取  $\mathbf{x} = e_\gamma$  时等号  $\|A\mathbf{x}\|_1 = \Gamma$  成立, 故

$$\|A\|_1 = \Gamma = \max_j \sum_i |A_{ij}|$$

注意 (4) 的证明与之几乎完全类似, 故省略。

(5) 只证明前一个不等式。不妨记:  $\gamma = \operatorname{argmax}_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2$ , 由 thm 1.2.1. (1) 知道: 对于任意的  $\|\mathbf{x}\|_2 = 1$ ,  $A\mathbf{x} \in \mathbb{C}^m$  总满足

$$\|A\mathbf{x}\|_\infty \leq \|A\mathbf{x}\|_2 \leq \sqrt{m} \|A\mathbf{x}\|_\infty$$

于是

$$\|A\|_2 = \|A\gamma\|_2 \leq \sqrt{m} \|A\gamma\|_\infty \leq \sqrt{m} \|A\|_\infty$$

左侧不等号同理。注意第二个不等式的证明与之几乎完全类似, 故省略。

(6) 先证左侧不等号: 对任意  $\|\mathbf{x}\|_2 = \sqrt{\sum_j^n |x_j|^2} = 1$ :

$$\begin{aligned} \|A\|_F^2 &= \sum_i^m \sum_j^n |a_{ij}|^2 = \sum_i^m \left( \sum_j^n |a_{ij}|^2 \right) \cdot 1 = \sum_i^m \left( \sum_j^n |a_{ij}|^2 \right) \cdot \|\mathbf{x}\|_2^2 \\ &= \sum_i^m \left\{ \left( \sum_j^n |a_{ij}|^2 \right) \left( \sum_j^n |x_j|^2 \right) \right\} \geq \sum_i^m \left\{ \left| \sum_j^n a_{ij} x_j \right|^2 \right\} \quad \text{Cauchy - Schwartz} \\ &= \sum_i^m |(A\mathbf{x})_i|^2 = \|A\mathbf{x}\|_2^2 \implies \|A\|_F \geq \sup_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \|A\|_2 \end{aligned}$$

再证右侧不等号: 显然  $\|A\|_2^2 = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2^2 \geq \|Ae_j\|_2^2$  for  $j = 1 \cdots n$ , 于是

$$n \|A\|_2^2 \geq \sum_{j=1}^n \|Ae_j\|_2^2 = \|A\|_F^2$$

由 (7) 知道还有  $m \|A^T\|_2^2 = m \|A\|_2^2 \geq \|A\|_F^2$ , 合并得到  $\|A\|_F \leq \min(m, n) \|A\|_2$

(7) 对于 F-norm 显然因为其只不过是矩阵的所有元素一维化后使用向量的 2-norm; 对于  $\|A\|_2 = \|A^T\|_2$  的证明见奇异值分解 SVD 章节。

**Proof Theorem 1.2.3.** 先证  $\|\mathbf{x}\|_2$ , 余下的留待在 SVD 后证明。令  $U$  为 Unitary:

$$\|U\mathbf{x}\|_2 = (U\mathbf{x})^T (U\mathbf{x}) = \mathbf{x}^T (U^T U) \mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|_2$$



## 1.4 Structured Matrices and Common Theories

下面是几种常见的**方阵**  $A \in \mathbb{C}^{n \times n}$  结构和与它们有关的常见结论。

- 1) **正规矩阵 (normal)** 若方阵满足  $A^H A = A A^H$  则称其为正规矩阵, 一个方阵正规当且仅当其在复数域  $\mathbb{C}$  可以酉对角化, 即  $\exists U$  whose columns consist of eigenvectors in  $\mathbb{C}^n$  s.t.  $U^H U = U U^H = I$  unitary and  $A = U \Lambda U^H$ ; 或等价地, 存在由  $n$  个特征向量 (from  $\mathbb{C}^n$ ) 可以构成的规范正交基底使其张成为  $\mathbb{C}^n$ 。
- 2) **酉矩阵 (unitary)** 若方阵满足  $A^H A = A A^H = I$  则称其为酉矩阵; 若为实数阵, 则  $A^T A = A A^T = I$  称为规范正交矩阵 (orthogonal, 注意和**列正交** have orthogonal columns  $A^H A = \Lambda$  区别); 一个**正规矩阵**为酉矩阵当且仅当其所有特征值的模均为 1, 即  $|\lambda| \equiv 1$ 。注意正交矩阵的特征值也不一定是实特征值。
- 3) **自伴/厄米矩阵 (self-adjoint/Hermitian)** 若复数方阵满足  $A^H = A$  则称其为自伴/厄米矩阵; 若为实数阵, 则  $A^T = A$  称为实对称矩阵 (real symmetry); 一个**正规矩阵**为厄米矩阵当且仅当其所有特征值为实数, 即  $\lambda \in \mathbb{R}$ 。
- 4) **斜厄米矩阵 (skew-Hermitian)** 若复数方阵满足  $A^H = -A$  则称其为斜厄米矩阵; 若为实数阵, 则  $A^T = -A$  称为斜对称矩阵 (skew-symmetry)。

下面是两种常见的**非方阵**  $A \in \mathbb{C}^{m \times n}$  结构。

- 1) **Hessenberg** 若矩阵满足  $A_{ij} = 0$  if  $i > j + 1$  则称其为 Hessenberg 矩阵 (下辅对角线以下都为 0)。
- 2) **列规范正交矩阵 (orthonormal)** 若矩阵满足  $A^H A = I$  if  $m > n$  (tall) 则称其为列规范正交矩阵, 或称其为 matrix with orthonormal columns; 注意此时  $A^H A = I \not\Rightarrow A A^H = I$ 。
- 3) **稀疏矩阵 (sparse)** 若矩阵中的大量元素为 0, 则称其为稀疏矩阵, 否则为 dense。用  $\text{nnz}(A)$  表示矩阵中非 0 元素的个数。
- 4) **高/宽矩阵 (tall/fat)** 若矩阵满足行多于列  $m > n$ , 则称起为高矩阵, 反之 ( $m < n$ ) 称为宽矩阵。

由于**实对称**矩阵有很多很好的性质，我们取出来单独讨论。在此之前先做一些定义：

**Definition 1.4.1. 正/负定 (positive/negative definite)**  $n$  阶实方阵  $A \succ 0$  ( $A \prec 0$ ) 为正/负定的，是指其二次型  $\forall \mathbf{x} \in \mathbb{R}^n \setminus \{0\} : \mathbf{x}^T A \mathbf{x} > 0$  ( $< 0$ )；

**半正/负定 (positive/negative semi-definite)**  $n$  阶实方阵  $A \succeq 0$  ( $A \preceq 0$ ) 为半正/负定的，是指其二次型  $\forall \mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T A \mathbf{x} \geq 0$  ( $\leq 0$ )。

**Theorem 1.4.1. 实对称矩阵的性质** 若  $A^T = A \in \mathbb{R}^{n \times n}$  为实对称矩阵，则

- 1) **特征值分解** 所有实对称矩阵都能进行正交对角化： $A = V \Lambda V^T$ ，其中  $V V^T = V^T V = I$  为正交矩阵。
- 2) **特征值和特征向量** 所有实对称矩阵的**特征值均为实数**，**特征向量也均来自且能张成实数空间  $\mathbb{R}^n$** （由特征向量的求法自然可得）。
- 3) **正/负定性** 所有实对称矩阵，下列命题等价：

$$\begin{aligned}
 A \succ 0 &\iff \forall \lambda_i : \lambda_i > 0 \iff \forall d_i : d_i > 0 \\
 A \prec 0 &\iff \forall \lambda_i : \lambda_i < 0 \iff \forall d_i : (-1)^i d_i > 0 \\
 A \succeq 0 &\iff \forall \lambda_i : \lambda_i \geq 0 \iff \forall M_i : M_i \geq 0 \\
 A \preceq 0 &\iff \forall \lambda_i : \lambda_i \leq 0 \iff \forall M_i : (-1)^i M_i \geq 0
 \end{aligned}$$

其中  $d_i$  与  $M_i$  分别表示矩阵  $A$  的顺序主子式 *leading principle minor* 和主子式 *principle minor*。 $d_i$  为前  $i$  行  $i$  列分块子矩阵的行列式； $M_i$  为所有任取  $i$  行与相同下标的  $i$  列合并成的子矩阵的行列式。

- 4) **实对称矩阵的转化** 对任何  $B \in \mathbb{C}^{m \times n}$ ，都有  $B^H B$  与  $B B^H$  实对称。

## 1.5 Subspaces and Dimensions

### Definition 1.5.1. 向量张成与张成空间的维度

假设有一组列向量  $\{v_i\}_{i=1}^n \subseteq \mathbb{C}^m$ , 我们称其中极大线性无关组的大小  $d$  为这些向量张成空间  $\text{span}\{v_i\}_{i=1}^n$  的维数, 这些构成极大线性无关组的向量为张成空间的基底 *basis*。将这些向量依次排开成一个矩阵  $V = [v_{.1} \cdots v_{.n}]$ , 则可等价地记这些向量的张成为  $\text{span}(V)$ , 这些向量的任何线性组合都可以有等价的矩阵表示:

$$V\mathbf{c} = \sum_{i=1}^n c_i v_i$$

其中  $\mathbf{c} = [c_1 \cdots c_n]^T$  为系数向量。不难由定义得到维数  $d = \text{rank}(V)$ 。

下面是一些线性代数中关于维度的常见定理 (所在数域  $\mathbb{C}, \mathbb{R}$  不影响结论):

**Theorem 1.5.1. 向量张成与张成空间的维度** 假设有两组列向量 (构成的矩阵)  $V_1 \in \mathbb{C}^{n \times d_1}$  与  $V_2 \in \mathbb{C}^{n \times d_2}$  使得  $d_1 + d_2 > n$ , 则两个张成空间  $\text{span}(V_1)$  与  $\text{span}(V_2)$  必有非  $\mathbf{0}$  的重合元素:

$$(\text{span}(V_1) \cap \text{span}(V_2)) \setminus \{\mathbf{0}\} \neq \emptyset$$

证明见 *Lecture Notes p.13*。

**Theorem 1.5.2. 维度定理 (dimension theorem)** 假设有一个矩阵  $V \in \mathbb{C}^{m \times n} : \mathbb{C}^n \mapsto \mathbb{C}^m$ , 则:

$$\text{nullity}(V) + \text{rank}(V) = n$$

回顾一下:  $\text{nullity}(V) = \dim(N(V))$ , 其中  $N(V) = \{\mathbf{x} \in \mathbb{C}^n : V\mathbf{x} = \mathbf{0}\}$  称为矩阵的零空间 *null space*;  $\text{rank}(V) = \dim(R(V))$ , 其中  $R(V) = \text{span}(V) \subseteq \mathbb{C}^m$  称为矩阵的列空间 *column space*, 或者值域 *range*。证明见高等代数笔记。

关于特征多项式、特征值还有特征空间还有几个常见的结论如下:

### Theorem 1.5.3. 特征多项式/特征值/特征空间的性质

记  $A \in \mathbb{C}^{n \times n}$  为方阵, 其有  $k \leq n$  个相异的特征值, 特征多项式为

$$f(\lambda) = \det(\lambda I - A) = \prod_{i=1}^k (\lambda - \lambda_i)^{m_i}$$

我们称  $m_i$  为特征值对应的代数重数 *algebraic multiplicity*;  $\dim(E_{\lambda_i}) = \dim(A - \lambda_i I)$  为特征值对应的几何重数 *geometric multiplicity*, 我们有如下结论:

- 1) **行列式与迹的计算** 有行列式计算公式  $\det(A) = \prod_{i=1}^k \lambda_i^{m_i}$  与迹的计算公式  $\text{trace}(A) = \sum_{i=1}^k m_i \lambda_i$  (同时对  $f(\lambda)$  的两种形式展开即可)。
- 2) **特征值与可逆**  $A$  可逆  $\iff$  特征值均不为 0  $\iff$  奇异值均不为 0  $\iff A^H A \succ 0$ 。第一条: 存在 0 特征值当且仅当  $Ax = 0$  有非平凡解; 第二、三条详见后文 *thm 2.1.1* 与 *def 2.1.1*, 非 0 奇异值的个数等于矩阵的秩。
- 3)  **$A^H$  的特征值** 若  $\lambda_i$  为  $A$  的特征值, 则  $\bar{\lambda}_i$  必为  $A^H$  的特征值。一个简单的证明如下:  

$$f_{A^H}(\bar{\lambda}_i) = \det(\bar{\lambda}_i I - A^H) = \det(\bar{\lambda}_i I - \overline{A}) = \overline{\det(\lambda_i I - A)} = \overline{f_A(\lambda_i)} = 0$$
- 4) **重数与对角化** 几何重数不超过代数重数  $1 \leq \dim(E_{\lambda_i}) \leq m_i$ , 且当且仅当  $\forall \lambda_i : \dim(E_{\lambda_i}) = m_i$  时  $A$  才可以对角化。
- 5) **相似矩阵的特征值** 相似矩阵  $A, B$  s.t.  $A = PBP^{-1}$ , 有相同的特征多项式 (因此特征值、迹、行列式均相同, 秩也相同)。
- 6) **Cayley Hamilton Theorem** 特征多项式是矩阵的零化多项式  $f(A) = 0$ 。

**Lemma 1.5.1.** 设  $A \in \mathbb{C}^{m \times n}$ , 则

$$\begin{bmatrix} I_m & A_{mn} \\ O & I_n \end{bmatrix}^{-1} = \begin{bmatrix} I_m & -A_{mn} \\ O & I_n \end{bmatrix}$$

**Theorem 1.5.4.** 设  $A \in \mathbb{C}^{m \times n}$ ,  $B \in \mathbb{C}^{n \times m}$ , 则  $AB$  与  $BA$  有相同的特征值。特征多项式关系:  $f_{AB}(\lambda) = \lambda^{m-n} f_{BA}(\lambda)$ 。

**Proof Theorem 1.5.4.** 观察发现:

$$\begin{bmatrix} I_m & -A_{mn} \\ O & I_n \end{bmatrix} \begin{bmatrix} AB & O \\ B_{nm} & O_n \end{bmatrix} \begin{bmatrix} I_m & A_{mn} \\ O & I_n \end{bmatrix} = \begin{bmatrix} O_m & O \\ B & BA \end{bmatrix}$$

也即矩阵  $C_1 = \begin{bmatrix} AB & O \\ B_{nm} & O_n \end{bmatrix}$  与  $C_2 = \begin{bmatrix} O_m & O \\ B & BA \end{bmatrix}$  相似且为分块三角矩阵, 故有相同的特征多项式。由分块三角矩阵行列式公式, 他们的特征多项式可以化为:

$$\left. \begin{aligned} f_{C_1}(\lambda) &= |\lambda I - C_1| = |\lambda I - AB| \lambda^n = \lambda^n f_{AB}(\lambda) \\ f_{C_2}(\lambda) &= |\lambda I - C_2| = |\lambda I - BA| \lambda^m = \lambda^m f_{BA}(\lambda) \\ f_{C_1}(\lambda) &= f_{C_2}(\lambda) \end{aligned} \right\} \Rightarrow f_{AB}(\lambda) = \lambda^{m-n} f_{BA}(\lambda)$$

## 2 Singular Value Decomposition

### 2.1 Singular Values

在本章节我们讨论奇异值的定义合法性与一些前置的理论知识及证明。

**Theorem 2.1.1.** 设  $A \in \mathbb{C}^{m \times n}$ , 则  $A^H A$  与  $A A^H$  均为实对称且:

- 1)  $A^H A$  与  $A A^H$  均有实特征根
- 2)  $A^H A$  与  $A A^H$  的特征根均非负  $\lambda \geq 0$ , 即均为半正定矩阵
- 3)  $A$ ,  $A^H A$ ,  $A A^H$  的秩相同, 即  $\text{rank}(A) = \text{rank}(A^H A) = \text{rank}(A A^H)$
- 4)  $A^H A$ ,  $A A^H$  有相同的非 0 特征值

我们先证明结论 2/3, 结论 1 与由结论 4 分别由实对称的性质与 thm 1.5.4 显然可得。

**Proof Theorem 2.1.1. (2)** 假设  $A^H A$  存在特征根  $\lambda \neq 0$ , 与对应的特征向量  $\mathbf{x}_\lambda$ , 则考虑  $\|A\mathbf{x}_\lambda\|^2$ , 有

$$0 \leq \|A\mathbf{x}_\lambda\|^2 = (A\mathbf{x}_\lambda)^H A\mathbf{x}_\lambda = \mathbf{x}_\lambda^H (A^H A) \mathbf{x}_\lambda = \lambda \|\mathbf{x}_\lambda^H\|^2$$

故有:  $\lambda \|\mathbf{x}_\lambda^H\|^2 \geq \|A\mathbf{x}_\lambda\|^2 \Rightarrow \lambda \geq 0$

**(3)** 先证明前一个等号: 只要证明  $\text{nullity}(A) = \text{nullity}(A^H A)$  即可由 dimension theorem 自然得到等式结果, 实际上:  $N(A) = N(A^H A)$ 。显然  $N(A) \subseteq N(A^H A)$ , 我们的目标是证明:  $N(A^H A) \subseteq N(A)$ 。现任取:  $\mathbf{x} \in N(A^H A)$ ,

$$\|A\mathbf{x}\|^2 = \mathbf{x}^H (A^H A \mathbf{x}) = \mathbf{x}^H \mathbf{0} = 0 \Rightarrow \|A\mathbf{x}\| = 0 \iff A\mathbf{x} = \mathbf{0}$$

即  $\mathbf{x} \in N(A)$ , 所以  $\text{rank}(A) = \text{rank}(A^H A)$ , 同理

$$\text{rank}(A A^H) = \text{rank}(A^H) = \text{rank}(A)$$

**Definition 2.1.1. 矩阵的奇异值 (singular values)** 任给一个矩阵  $A \in \mathbb{C}_r^{m \times n}$  s.t.  $\text{rank}(A) = r$ , 记  $\lambda_i : i = 1 \cdots r$  为  $A^H A$  按照降序排列的特征值。定义  $A$  的奇异值为:

$$\sigma_i(A) = \sqrt{\lambda_i} \text{ s.t. } \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$$

上述的定理 2.1.1 中的 1 与 2 确保了开根号的合理性, 3 确保了非 0 奇异值的个数等于  $A^H A$  的秩 (相似对角化), 从而又等于原矩阵的秩。

## 2.2 Singular Value Decomposition

本节中我们将要讨论一种最常见的矩阵分解，它可以将任意形状的实、复数矩阵分解成一个对角矩阵与两个列正交或酉矩阵的乘积，称为奇异值分解 SVD。奇异值分解有很多不同的形式，一种形式如下：

### Theorem 2.2.1. 奇异值分解 SVD

任何  $A \in \mathbb{C}^{m \times n}$  s.t.  $m \geq n$  tall，都可以分解成如下的形式 (thin SVD)

$$A = U \Sigma V^H$$

其中  $U^H U = V^H V = I_n$  为列规范正交矩阵； $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$  s.t.  $\sigma_1 \geq \dots \geq \sigma_n \geq 0$ 。注意这几个矩阵的维数： $U \in \mathbb{R}^{m \times n}$ ， $\Sigma \in \mathbb{R}^{n \times n}$ ， $V \in \mathbb{R}^{n \times n}$  称  $U$ ， $V$  的列分别为左、右奇异向量 singular vectors。这种方法中  $U$  不一定是方阵；

或者可以分解成如下形式 (full SVD)，记任何  $A \in \mathbb{C}^{m \times n}$  s.t.  $\text{rank}(A) = r$ ，不一定为 tall matrix

$$A = U \Sigma V^H = U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix} V^H$$

其中  $U^H U = I_m$ ， $V^H V = I_n$  为酉方阵； $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$  s.t.  $\sigma_1 \geq \dots \geq \sigma_r > 0$ 。矩阵的维数： $U \in \mathbb{R}^{m \times m}$ ， $\Sigma \in \mathbb{R}^{m \times n}$ ， $V \in \mathbb{R}^{n \times n}$  称  $U$ ， $V$  的列分别为左、右奇异向量。这种方法中  $\Sigma$  不一定是方阵

对于 full SVD 更详细地，有：

$$A = U \Sigma V^H = [\tilde{U}_r \quad \tilde{U}_{m-r}^\perp] \begin{bmatrix} \Sigma_r & O_1 \\ O_2 & O_3 \end{bmatrix}_{m \times n} V^H$$

其中  $A^H A = V \Lambda V^H$  为  $A^H A$  的正交对角化； $\tilde{U}_r$  为  $AV$  前  $r$  列正规化后按序合并成的子矩阵（除以 norm, i.e., 对应的奇异值  $\sigma_i > 0$ ）得到；最后  $\tilde{U}_{m-r}^\perp$  为  $\text{span}(\tilde{U}_r)$  在  $\mathbb{C}^m$  中的规范正交补，不妨取  $\text{span}(\tilde{U}_{m-r}^\perp) = N(\tilde{U}_r^H)$ ：

$$AV = U \Sigma = [\tilde{U}_r \Sigma_r \quad O_{n-r}] = [\tilde{U}_r \quad \tilde{U}_{m-r}^\perp] \begin{bmatrix} \Sigma_r & O_1 \\ O_2 & O_3 \end{bmatrix}_{m \times n}$$

**Theorem 2.2.2. 列空间的正交补** 任何  $A \in \mathbb{C}^{m \times n}$  都有  $\{R(A)\}^\perp = N(A^H)$

**Proof Theorem 2.2.1.** 对任意的  $A \in \mathbb{C}^{m \times n}$  s.t.  $\text{rank}(A) = r$ , 都存在对实对称矩阵  $A^H A \in \mathbb{C}^{n \times n}$  的正交对角化 (特征值从大到小)  $A^H A = V \Lambda V^H$ , 不难证明  $AV$  为部分列正交矩阵 (可能含有  $\mathbf{0}$  向量):

$$(AV)^H(AV) = V^H(A^H A)V = V^H(V \Lambda V^H)V = \Lambda$$

于是, 我们可以对  $AV$  进行分解, 考虑由相似  $A \sim \Lambda$ , 可以有

$$\text{rank}(A) = \text{rank}(\Lambda) = \text{number of positive eigenvalues} = r$$

所以  $AV$  的前  $r$  个列向量  $u_i$  s.t.  $1 \leq i \leq r$  不为 0, 且其范数就是奇异值  $\|u_i\| = \sigma_i$ , 即有

$$\begin{aligned} AV &= [u_{.1} \ \cdots \ u_{.r} \ 0 \ \cdots \ 0] \\ &= [\tilde{U}_r \Sigma_r \ O_{m \times (n-r)}] \\ &= [\tilde{U}_r \ \tilde{U}_{m \times (m-r)}^\perp] \begin{bmatrix} \Sigma_r & O_{r \times (n-r)} \\ O_{(m-r) \times r} & O_{(m-r) \times (n-r)} \end{bmatrix} \quad (*) \\ &= U \Sigma \end{aligned}$$

注意由于  $\dim(\text{span}(\tilde{U}_r)) = r \leq m = \dim(\mathbb{C}^m)$ , 必然能找到  $\text{span}(\tilde{U}_r)$  的一个  $(m-r)$  维正交补, 其上任何一组大小为  $(m-r)$  的线性无关向量均可以化为一组规范正交基底, 作为列向量排好成  $\tilde{U}_{m \times (m-r)}^\perp$  即可。对于满秩矩阵  $\text{rank}(A) = r = \min\{m, n\}$ , 这一步甚至可以省略。

特别地, 我们总可以找  $N(\tilde{U}_r^H) = \{\text{span}(\tilde{U}_r)\}^\perp$ , 即列空间的正交补为共轭转置的零空间。最后两边同时乘上  $V^H$  即可得到分解结果  $A = U \Sigma V^H$

**Proof Theorem 2.2.2.**

证明一般情况  $\{R(A)\}^\perp = N(A^H)$ : 假设  $\mathbf{x} \in R(A)$ ,  $\mathbf{y} \in N(A^H)$ , 则

$$\langle \mathbf{x}, \mathbf{y} \rangle = (A\mathbf{c})^H \mathbf{y} = \mathbf{c}^H (A^H \mathbf{y}) = \mathbf{c}^H \cdot \mathbf{0} = 0$$

下面介绍 SVD 的具体操作流程。



### SVD 的具体操作流程

**STEP 1** 对  $A^H A$  进行正交对角化, 特征值从大到小排列  $A^H A = V \Lambda V^H$  这里回顾一下 Gram-Schmidt 正交化的流程, 假设要对向量组  $\{\mathbf{x}_i\}_{i=1}^k$  进行规范正交化  $\{\mathbf{x}_i\}_{i=1}^k \mapsto \{\mathbf{y}_i\}_{i=1}^k$ , 记  $\mathbf{x}^1 = \frac{\mathbf{x}}{\|\mathbf{x}\|}$  为一个向量的单位方向向量, 则按照如下两步迭代即可:

$$\mathbf{y}_1 = \mathbf{x}_1^1 \Rightarrow \mathbf{y}_{i+1} = \left( \mathbf{x}_{i+1} - \sum_{k=1}^i \langle \mathbf{x}_{i+1}, \mathbf{y}_k \rangle \cdot \mathbf{y}_k \right)^1$$

**STEP 2** 计算  $AV$  并将其中的非零列向量规范正交化, 系数作为奇异值矩阵提出  $\Sigma_r$ , 视情况而定是否需要添加规范正交补列向量 ( $AV$  无零列向量或  $A$  行满秩都不需要):

$$\begin{aligned} AV &= \begin{bmatrix} u_{.1} & \cdots & u_{.r} & 0 & \cdots & 0 \end{bmatrix} \\ &= \begin{bmatrix} \widetilde{U}_r \Sigma_r & O_{m \times (n-r)} \end{bmatrix} \xrightarrow{n=r} \widetilde{U}_r \Sigma_r \\ &= \begin{bmatrix} \widetilde{U}_r & \widetilde{U}_{m \times (m-r)}^\perp \end{bmatrix} \begin{bmatrix} \Sigma_r & O_{r \times (n-r)} \\ O_{(m-r) \times r} & O_{(m-r) \times (n-r)} \end{bmatrix} \xrightarrow{m=r} \widetilde{U}_r \begin{bmatrix} \Sigma_r & O_{r \times (n-r)} \end{bmatrix} \\ &= U \Sigma \end{aligned}$$

如果需要找  $\widetilde{U}_r$  的正交补  $\widetilde{U}_{m \times (m-r)}^\perp$ , 则解方程组:  $\widetilde{U}_r^H \mathbf{x} = 0$  并取解空间的规范正交基底即可。

**STEP 3** 回乘  $V^H$ :  $A = U \Sigma V^H$

### Theorem 2.2.3. SVD 的应用

设  $A \in \mathbb{C}^{m \times n}$ , 其秩  $\text{rank}(A) = r$  对其进行 full SVD 使得  $A = U \Sigma V^H$ , 则有下面的结论。证明比较简单, 省略:

- 1)  $A = U \Sigma V^H = \sum_{i=1}^r \sigma_i u_{.i} v_{.i}^H$
- 2)  $\text{span}(A) = \text{span}([u_{.1} \cdots u_{.r}]); \text{span}(A^H) = \text{span}([v_{.1} \cdots v_{.r}])$
- 3)  $N(A) = \text{span}([v_{.(r+1)} \cdots v_{.n}])$
- 4)  $Av_{.i} = \sigma_i u_{.i}$  且  $u_{.i}^H A = \sigma_i u_{.i} v_{.i}^H$ ,  $(\sigma_i, u_{.i}, v_{.i})$  称为  $i_{th}$  singular triplet
- 5)  $V$  是  $A^H A$  的特征向量矩阵;  $U$  的非零部分  $\widetilde{U}_r$  是  $AA^H$  的非零特征值的特征向量矩阵。



## 2.3 Low-rank approximation via truncated SVD

**Theorem 2.3.1. 矩阵范数与奇异值** 设  $A \in \mathbb{C}^{m \times n}$ , 其秩  $\text{rank}(A) = r$ , 则

- 1)  $\|A\| = \max_{\|x\|=1} \|Ax\| = \sigma_1(A)$ , 特别地, 若  $A^H A = A A^H$  normal,  $\sigma_1(A) = \max_i |\lambda_i| = \rho(A)$ ,  $|\cdot|$  表示 modulus,  $\rho(A)$  为谱半径。
- 2)  $\|A^{-1}\|^{-1} = \min_{\|x\|=1} \|Ax\| = \sigma_r(A)$
- 3)  $\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\sum_{i=1}^r \sigma_i^2}$
- 4)  $\|A\|_* = \max_{Q \text{ orthonormal}} \{\text{trace}(Q^T A)\} = \sum_{i=1}^r \sigma_i$
- 5) 条件数 condition number  $\kappa(A) := \|A\| \cdot \|A^{-1}\| = \sigma_1(A)/\sigma_r(A)$

现在, 对  $A \in \mathbb{C}^{m \times n}$ ,  $\text{rank}(A) = r$  我们考虑找一个秩为  $k \leq r$  的矩阵  $A_k$  使得其可以尽可能多地保留  $A$  的特征, 即使得:  $A_k = \arg\min_{\text{rank}(A') \leq k} \|A - A'\|$  我们有如下定理

**Theorem 2.3.2. Truncated SVD Approximation Theorem**

对  $A \in \mathbb{C}^{m \times n}$ , 其秩  $\text{rank}(A) = r$ , 考虑其 full SVD 展开,  $A = U \Sigma V^H$ , 记

$$A_k = U_k \Sigma_k V_k^H = \sum_{i=1}^k \sigma_i u_i v_i^H$$

其中  $U_k, V_k$  分别只包含了前  $k$  个对应的奇异向量, 中间的  $\Sigma_k$  为  $k \times k$  方阵, 则对  $\forall k \leq r$

$$\|A - A_k\| = \min_{\text{rank}(A') \leq k} \|A - A'\| = \sigma_{k+1}(A)$$

其中  $\sigma_{r+1}(A) = 0$ 。实际上, 这个估计  $A_k \sim A$  对上述关系在任何酉不变矩阵范数下均成立 (不做证明), 即我们也可以把上述的矩阵范数换成  $\|\cdot\|_F$  或者  $\|\cdot\|_*$ 。  $A$  与  $A_k$  的关系如下 (奇异值顺序递增) 显然每一项得到的矩阵的秩均为 1:

$$\begin{aligned} A &= \sum_{i=1}^k \sigma_i \begin{bmatrix} * \\ \vdots \\ * \end{bmatrix}_i \begin{bmatrix} * & \cdots & * \end{bmatrix}_i + \sum_{i=k+1}^r \sigma_i \begin{bmatrix} * \\ \vdots \\ * \end{bmatrix}_i \begin{bmatrix} * & \cdots & * \end{bmatrix}_i \\ &= A_k + \sum_{i=k+1}^r \sigma_i \begin{bmatrix} * \\ \vdots \\ * \end{bmatrix}_i \begin{bmatrix} * & \cdots & * \end{bmatrix}_i \end{aligned}$$

实际上主成分分析 PCA 也是利用了这个原理。比如我们将  $n$  个  $m$  维数据放在  $m \times n$  矩阵  $A$  中, 如果想要降至  $k$  维, 我们提取  $A_k$ , 将  $U_k$  的各列作为  $k$  个新的方向, 将  $\Sigma_k V_k^H$  的各列 (下面公式的红色部分) 作为新的数据点 (在  $k$  个新方向  $u_i$  上的坐标), 即可将维度从  $m$  降至  $k$

$$A_k = \begin{bmatrix} u_{\cdot 1} & \cdots & u_{\cdot k} \end{bmatrix} \begin{bmatrix} \sigma_1 v_{11} & \cdots & \sigma_1 v_{1n} \\ \vdots & \vdots & \vdots \\ \sigma_k v_{k1} & \cdots & \sigma_k v_{kn} \end{bmatrix}$$

**Proof Theorem 2.3.1.** (1) 对  $\forall : \|\mathbf{x}\| = 1$ , 考虑  $A$  的 full SVD s.t.  $A = U\Sigma V^T$ , 由向量 2-范数的酉不变性质, 有

$$\begin{aligned} \|A\mathbf{x}\| &= \|U\Sigma V^T \mathbf{x}\| = \|\Sigma(V^T \mathbf{x})\| = \|\Sigma \mathbf{y}\| \\ &= \sqrt{\sum_{i=1}^r \sigma_i^2 y_i^2} \leq \sqrt{\sum_{i=1}^r \sigma_1^2 y_i^2} \leq \sigma_1 \cdot \sqrt{\sum_{i=1}^n y_i^2} = \sigma_1 \|\mathbf{y}\| = \sigma_1 \end{aligned}$$

取  $\mathbf{x} = v_{\cdot 1}$  同理可证 (2), 此处略过。

(3) 考虑  $A$  的 full SVD s.t.  $A = U\Sigma V^T$

$$\begin{aligned} \|A\|_F^2 &= \text{trace}(A^H A) = \text{trace}(V \Sigma^T \Sigma V^T) \\ &= \text{trace}(\Sigma V^T V \Sigma^T) = \text{trace}(\Sigma \Sigma^T) = \sum_{i=1}^r \sigma_i^2 \end{aligned}$$

**Proof Theorem 2.3.2** 任取  $A' \in \mathbb{C}^{m \times n}$  s.t.  $\text{rank}(A') \leq k \leq r$ , 由 dimension theorem,  $\text{nullity}(A') = n - \text{rank}(A') \geq n - k$ , 即存在 ONB (规范正交基底)  $W \in \mathbb{C}^{n \times (n-k)}$  s.t.  $A'W = 0$  且  $W^H W = I$ 。注意到此时  $W$  和  $V_{k+1}$  的列恰好构成  $[W \ V_{k+1}] \in \mathbb{C}^{n \times (n+1)}$ , 所以这些列必然线性相关。即:  $\exists \mathbf{x}_1, \mathbf{x}_2 \neq \mathbf{0}$  s.t.  $W\mathbf{x}_1 - V_{k+1}\mathbf{x}_2 = 0$ , 我们可以对  $\mathbf{x}_1$  和  $\mathbf{x}_2$  的长度进行放缩, 使得有  $\|W\mathbf{x}_1\| = \|V_{k+1}\mathbf{x}_2\| = \|\mathbf{x}_1\| = \|\mathbf{x}_2\| = 1$  (矢量 2-范数的酉不变性)。取这么一个  $\mathbf{x}_1$ , 我们有:

$$\begin{aligned} \|A - A'\| &\geq \|(A - A')W\| = \|AW\| = \|U\Sigma(V^H W)\| \\ &\geq \|U\Sigma(V^H W)\mathbf{x}_1\| = \|\Sigma V^H V_{k+1}\mathbf{x}_2\| \\ &= \|\Sigma_{k+1}\mathbf{y}_2\| \geq \sigma_{k+1}(A) \end{aligned}$$

第一个不等号是因为, 由矢量乘法的酉不变性, 对于任意的列规范正交矩阵  $W$  和一个适配乘法的矩阵  $B$ ,  $\{\|BW\mathbf{x}\| : \|\mathbf{x}\| = 1\} \subseteq \{\|B\mathbf{x}\| : \|\mathbf{x}\| = 1\}$ 。因此集合上的最大值应当符合关系  $\|BW\| \leq \|B\|$ 。最后  $\|A - A'\| \geq \sigma_{k+1}(A)$  在  $A' = A_k$  时取等号, 证毕。

## 2.4 Courant-Fisher Minimax Theorem

### Theorem 2.4.1. Courant-Fisher Minimax Theorem (CF)

记矩阵  $A \in \mathbb{C}^{m \times n}$ , 其秩  $\text{rank}(A) = r$ ,  $S_n \subseteq \mathbb{C}^n$  为一子 (线性) 空间, 则

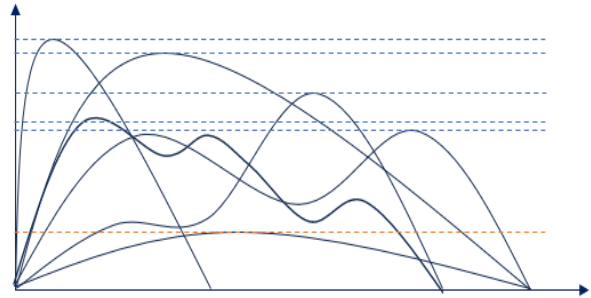
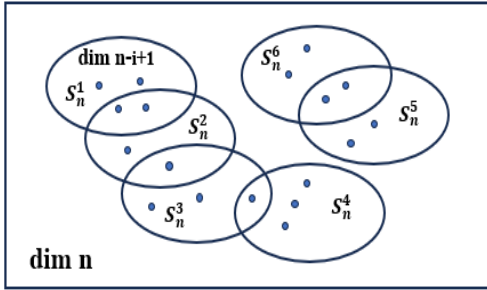
$$\begin{aligned}\sigma_i(A) &= \min_{\dim(S_n)=n+1-i} \max_{\mathbf{x} \in S_n, \|\mathbf{x}\|=1} \|A\mathbf{x}\| \\ &= \max_{\dim(S_n)=i} \min_{\mathbf{x} \in S_n, \|\mathbf{x}\|=1} \|A\mathbf{x}\|\end{aligned}$$

可以理解为所有  $n+1-i$  维子空间中, 矩阵能实现最大幅度变换的下限; 或者, 所有  $i$  维子空间中, 矩阵能实现最小幅度变换的上限。注意, 对于集合  $\{\|A\mathbf{x}\| : \mathbf{x} \in S_n, \|\mathbf{x}\|=1, \dim(S_n)=i\}$  (第二个最值的部分), 我们总可以有如下的替换

$$\begin{aligned}&\{\|A\mathbf{x}\| : \mathbf{x} \in S_n, \|\mathbf{x}\|=1, \dim(S_n)=i\} \\ &= \{\|AQ\mathbf{y}\| : \mathbf{y} \in \mathbb{C}^i, \|\mathbf{y}\|=1, Q \in \mathbb{C}^{n \times i}, Q^H Q = I_i\}\end{aligned}$$

类似地, 对于降序排列的对称矩阵  $A$  的特征值

$$\lambda_i(A) = \min_{\dim(S_n)=n+1-i} \max_{\mathbf{x} \in S_n} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \max_{\dim(S_n)=i} \min_{\mathbf{x} \in S_n} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$



**Proof Theorem 2.4.1** 我们先证明后一个等号。考虑  $A$  的 full SVD  $A = U\Sigma V^H$ , 令  $V_i^* = [v_i \cdots v_n]$  观察到  $\dim(S_n) + \dim(V_i^*) = n - i + 1 + i = n + 1 \geq n$ , 所以  $\exists \mathbf{w} \in S_n \cap \text{span}(V_i^*)$  s.t.  $\mathbf{w} = V_i^* \mathbf{y}$  且  $\|\mathbf{w}\| = \|\mathbf{y}\| = 1$ 。于是对任意的  $\dim(S_n) = i$ , 总存在  $\|\mathbf{w}\| = \|\mathbf{y}\| = 1$  使得

$$\|A\mathbf{w}\| = \|U\Sigma V^H \mathbf{w}\| = \|\Sigma V^H V_i^* \mathbf{y}\| = \|\text{diag}(\sigma_i \cdots \sigma_n) \mathbf{y}\| \leq \sigma_i$$

即, 对任意的  $\dim(S_n) = i$ ,  $\sigma_i \geq \|A\mathbf{w}\| \geq \min_{\mathbf{x} \in S_n, \|\mathbf{x}\|=1} \|A\mathbf{x}\|$ 。当取  $S_n = V_i = [v_1 \cdots v_i]$  时, 等号成立。前一个等号及对称矩阵特征值相关的证明与这个完全类似, 前者取  $V_i = [v_1 \cdots v_i]$ ; 后者用规范正交对角化的  $Q$  取代前面的  $V$  即可。

**Theorem 2.4.2. Weyl's Theorem**

- 1) 任何矩阵  $A$  即期奇异值:  $|\sigma_i(A+E) - \sigma_i(A)| \leq \|E\|_2 = \sigma_1(E)$
- 2) 当  $i = 1$ :  $\|A\|_2 - \|E\|_2 \leq \|A+E\|_2 \leq \|A\|_2 + \|E\|_2$
- 3) 实对称矩阵的特征值:  $|\lambda_i(A+E) - \lambda_i(A)| \leq \|E\|_2$

这基本说明奇异值和对称矩阵的特征值对于扰动不敏感 (*well-conditioned*)。

**Theorem 2.4.3. 奇异值的额外性质**

- 1)  $\sigma_i(A^H) = \sigma_i(A)$  if  $1 \leq i \leq \text{rank}(A) = \text{rank}(A^H)$
- 2)  $\sigma_{\min(m,n)}(A)\sigma_i(B) \leq \sigma_i(AB) \leq \sigma_1(A)\sigma_i(B)$  if  $A \in \mathbb{C}^{m \times n}$
- 3)  $\sigma_i(A)\sigma_{\min(n,k)}(B) \leq \sigma_i(AB) \leq \sigma_i(A)\sigma_1(B)$  if  $B \in \mathbb{C}^{n \times k}$
- 4)  $\sigma_i\left(\begin{bmatrix} A_1 \\ A_2 \end{bmatrix}\right) \geq \max(\sigma_i(A_1), \sigma_i(A_2))$
- 5)  $\sigma_i\left(\begin{bmatrix} A_1 & A_2 \end{bmatrix}\right) \geq \max(\sigma_i(A_1), \sigma_i(A_2))$

后两个性质证明见 Lecture Notes p.23

**Proof Theorem 2.4.2 (1)** 我们只证明  $\sigma_i(A+E) \leq \sigma_i(A) + \|E\|_2$ , 另一侧的证明 (包括特征值的情形) 均与之类似。记  $S_n^i \subseteq \mathbb{C}^n$ ,  $\dim(S_n^i) = i$

$$\begin{aligned}
\sigma_i(A+E) &= \max_{S_n^i} \left\{ \min_{\mathbf{x} \in S_n^i, \|\mathbf{x}\|=1} \|(A+E)\mathbf{x}\| \right\} \\
&\leq \max_{S_n^i} \left\{ \min_{\mathbf{x} \in S_n^i, \|\mathbf{x}\|=1} (\|A\mathbf{x}\| + \|E\mathbf{x}\|) \right\} \\
&\leq \max_{S_n^i} \left\{ \min_{\mathbf{x} \in S_n^i, \|\mathbf{x}\|=1} \|A\mathbf{x}\| + \max_{\mathbf{x} \in S_n^i, \|\mathbf{x}\|=1} \|E\mathbf{x}\| \right\} \\
&\leq \max_{S_n^i} \min_{\mathbf{x} \in S_n^i, \|\mathbf{x}\|=1} \|A\mathbf{x}\| + \max_{S_n^i} \max_{\mathbf{x} \in S_n^i, \|\mathbf{x}\|=1} \|E\mathbf{x}\| \\
&\leq \max_{S_n^i} \min_{\mathbf{x} \in S_n^i, \|\mathbf{x}\|=1} \|A\mathbf{x}\| + \max_{\|\mathbf{x}\|=1} \|E\mathbf{x}\| = \sigma_i(A) + \|E\|_2
\end{aligned}$$

**Proof Theorem 2.4.3 (2)** (1) 显然, 我们只证明 (2)/(3)。记  $A \in \mathbb{C}^{m \times n}$ ,  $B \in \mathbb{C}^{n \times k}$ ,  $S_i \subseteq \mathbb{C}^k$ ,  $\dim(S_i) = i$ , 我们有关系:

$$\min_{\mathbf{x} \in S_i} \frac{\|AB\mathbf{x}\|}{\|B\mathbf{x}\|} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} \geq \min_{\mathbf{x} \in S_i} \frac{\|AB\mathbf{x}\|}{\|B\mathbf{x}\|} \cdot \min_{\mathbf{x} \in S_i} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} \geq \sigma_{\min(m,n)}(A) \min_{\mathbf{x} \in S_i} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|}$$

$$\max_{\mathbf{x} \in S_{k+1-i}} \frac{\|AB\mathbf{x}\|}{\|B\mathbf{x}\|} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} \leq \max_{\mathbf{x} \in S_{k+1-i}} \frac{\|AB\mathbf{x}\|}{\|B\mathbf{x}\|} \cdot \max_{\mathbf{x} \in S_{k+1-i}} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} \leq \sigma_1(A) \max_{\mathbf{x} \in S_{k+1-i}} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|}$$

于是, 对  $AB \in \mathbb{C}^{m \times k}$ , 有

$$\begin{aligned} \sigma_i(AB) &= \max_{S_i} \min_{\mathbf{x} \in S_i} \frac{\|AB\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{S_i} \min_{\mathbf{x} \in S_i} \frac{\|AB\mathbf{x}\|}{\|B\mathbf{x}\|} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} \\ &\geq \max_{S_i} \left\{ \sigma_{\min(m,n)}(A) \min_{\mathbf{x} \in S_i} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} \right\} \\ &= \sigma_{\min(m,n)}(A) \max_{S_i} \min_{\mathbf{x} \in S_i} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} = \sigma_{\min(m,n)}(A) \sigma_i(B) \end{aligned}$$

$$\begin{aligned} \sigma_i(AB) &= \min_{S_{k+1-i}} \max_{\mathbf{x} \in S_{k+1-i}} \frac{\|AB\mathbf{x}\|}{\|\mathbf{x}\|} = \min_{S_{k+1-i}} \max_{\mathbf{x} \in S_{k+1-i}} \frac{\|AB\mathbf{x}\|}{\|B\mathbf{x}\|} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} \\ &\leq \min_{S_{k+1-i}} \left\{ \sigma_1(A) \max_{\mathbf{x} \in S_{k+1-i}} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} \right\} \\ &= \sigma_1(A) \min_{S_{k+1-i}} \max_{\mathbf{x} \in S_{k+1-i}} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} = \sigma_1(A) \sigma_i(B) \end{aligned}$$

故 (2) 证毕, 有

$$\sigma_{\min(m,n)}(A) \sigma_i(B) \leq \sigma_i(AB) \leq \sigma_1(A) \sigma_i(B)$$

要证 (3), 对  $AB$  取共轭转置即可

$$\sigma_{\min(n,k)}(B^H) \sigma_i(A^H) \leq \sigma_i(B^H A^H) \leq \sigma_1(B^H) \sigma_i(A^H)$$

由 (1)  $\sigma_i(B^H A^H) = \sigma_i(AB)$ ,  $\sigma_i(A^H) = \sigma_i(A)$  与  $\sigma_i(B^H) = \sigma_i(B)$  易得

$$\sigma_{\min(n,k)}(B) \sigma_i(A) \leq \sigma_i(AB) \leq \sigma_1(B) \sigma_i(A)$$

### 3 Solving Linear Systems

本章中我们将讨论如何利用算法求线性方程组  $A\mathbf{x} = \mathbf{b}$  的准确或近似解。一类常见的求解算法是对系数矩阵  $A$  进行分解  $A = BC$ ，其中  $B, C$  是两个结构相对简单的矩阵。接着通过分别求解  $B\mathbf{y} = \mathbf{b}$  与  $C\mathbf{x} = \mathbf{y}$  得到方程组的解  $\mathbf{x}$ 。在本章的后续内容中，我们将主要讨论如何实现这些分解。

#### 3.1 LU Factorization

在之前所有的解线性方程组  $A\mathbf{x} = \mathbf{b}$  的实践中，我们的通用步骤是使用行初等变换，i.e., 高斯消元法，从而将线性方程组  $A\mathbf{x} = \mathbf{b}$  转化为等价的  $U\mathbf{x} = \tilde{\mathbf{b}}$ ，使得新的系数矩阵  $U$  为一个上三角矩阵（不一定 square）。我们知道，所有的行初等变换实际上都可以通过左乘一个行初等变换矩阵实现：

**Definition 3.1.1. 行初等变换矩阵** 作用在矩阵  $A \in \mathbb{F}^{m \times n}$  上的所有的三类行初等变换，i.e., 倍加 *replacement*，互换 *interchange*，伸缩 *scaling* 都可以通过左乘一个行初等变换矩阵  $E \in \mathbb{F}^{m \times m}$  得到。定义如下的记号

变换	作用	记号	逆矩阵
倍加	将第 $i$ 行 $\times k$ 加到第 $j$ 行	$E_{ij}(k)$	$E_{ij}(-k)$
互换	将第 $i$ 行与第 $j$ 行互换位置	$E_{ij}$	$E_{ji} = E_{ij}$
伸缩	将第 $i$ 行 $\times k$	$E_i(k)$	$E_i(k^{-1})$

行初等变换矩阵可以通过对单位矩阵  $I_m$  执行其所对应的操作得到。更进一步，行初等矩阵的乘法结果  $E_n \cdots E_1$  可以通过对  $I_m$  依次执行  $E_1 \sim E_n$  的操作得到。用同样的方法可以得到列初等变换矩阵（列初等变换等价于右乘列初等矩阵）。

于是高斯消元法等价于同时在方程的两侧同时左乘若干（有限）个行初等变换矩阵  $E_1 \sim E_n$

$$E_n \cdots E_2 E_1 A \mathbf{x} = U \mathbf{x} = E_n \cdots E_2 E_1 \mathbf{b} \Rightarrow E_n \cdots E_2 E_1 A = U$$

因此我们有如下关于  $A$  的分解

$$A = E_n^{-1} \cdots E_2^{-1} E_1^{-1} U = LU \quad \text{s.t.} \quad L = E_n^{-1} \cdots E_2^{-1} E_1^{-1}$$

如果我们对高斯消元的流程作如下限制：不涉及行互换操作，且倍加  $E_{ij}(k)$  时总有  $i < j$ ，如果在此流程下仍能得到预期的上三角矩阵  $U$ （注意这不一定总是可行的），则我们有上述的  $E_1 \sim E_n$ （及其逆矩阵）为一系列下三角矩阵，从而  $L$  也是一个下三角矩阵。我们称  $A = LU$  为矩阵  $A$  的 LU 分解。

**Definition 3.1.2. LU Factorization**

对于矩阵  $A \in \mathbb{F}^{m \times n}$ , 若存在  $A = LU$  使得  $L \in \mathbb{F}^{m \times m}$  为下三角矩阵;  $U \in \mathbb{F}^{m \times n}$  为上三角矩阵, 则称该分解为  $A$  的 LU 分解。**不是所有的矩阵都能通过高斯消元法得到其 LU 分解。**

**Example 3.1.1. LU 分解不存在的情形** 尝试使用高斯消元法对下列矩阵进行 LU 分解

$$A = \begin{bmatrix} 1 & 2 & -3 & 1 \\ 2 & 4 & 0 & 7 \\ -1 & 3 & 2 & 0 \end{bmatrix}$$

首先进行行初等变换

$$\begin{aligned} A &= \begin{bmatrix} 1 & 2 & -3 & 1 \\ 2 & 4 & 0 & 7 \\ -1 & 3 & 2 & 0 \end{bmatrix} \xrightarrow{-2r_1 + r_2} \begin{bmatrix} 1 & 2 & -3 & 1 \\ 0 & 0 & 6 & 5 \\ -1 & 3 & 2 & 0 \end{bmatrix} \xrightarrow{r_1 + r_3} \\ &\begin{bmatrix} 1 & 2 & -3 & 1 \\ 0 & 0 & 6 & 5 \\ 0 & 5 & -1 & 1 \end{bmatrix} \xrightarrow{r_2 \leftrightarrow r_3} \begin{bmatrix} 1 & 2 & -3 & 1 \\ 0 & 5 & -1 & 1 \\ 0 & 0 & 6 & 5 \end{bmatrix} = U \end{aligned}$$

即我们有  $E_3 E_2 E_1 A = E_{23} \cdot E_{13}(1) \cdot E_{12}(-2) A = U$ , 即

$$\tilde{L} = (E_3 E_2 E_1)^{-1} = E_{12}(-2)^{-1} \cdot E_{13}(1)^{-1} \cdot E_{23}^{-1} = E_{12}(2) \cdot E_{13}(-1) \cdot E_{23}$$

将对应的行初等变换矩阵从右往左作用于单位矩阵  $I$  得到

$$\begin{aligned} I &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{r_2 \leftrightarrow r_3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \xrightarrow{-r_1 + r_3} \\ &\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} \xrightarrow{2r_1 + r_2} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} = \tilde{L} \end{aligned}$$

可以发现, 由于在高斯消元的过程中使用了行互换, 导致初等变换矩阵乘积的逆并不是一个下三角矩阵, 因此这个矩阵没有 LU 分解。但经过观察, 我们总是可以找到一个排列矩阵  $P$  使得  $P\tilde{L} = L$  为下三角矩阵, 从而得到  **$PA = P\tilde{L}U = LU$  为  $A$  的 PLU 分解。**

**Definition 3.1.3. 排列矩阵 (permutation matrix)** 对于一个给定的排列  $\tau(i) : \{1 \cdots n\} \mapsto \{1 \cdots n\}$ , 作用在矩阵  $A \in \mathbb{F}^{m \times n}$  上的排列矩阵为

$$P = \begin{bmatrix} e_{\tau(1)}^T \\ \vdots \\ e_{\tau(n)}^T \end{bmatrix}$$

其中  $e_i$  为第  $i$  个单位向量。左乘  $P$  相当于对行做相应的排列；右乘  $P^T$  相当于对列做相应的排列。由简单的抽象代数知识，所有 *permutation* 都可以分解成若干个二元 *transpose* 的复合，故而  $P$  实际上是一系列（行/列）互换初等矩阵  $E_{ij}$  的复合。显然，所有的排列矩阵  $P$  都是 *orthogonal matrix*  $P^T P = P P^T = I$ 。

### Definition 3.1.4. PLU Factorization

对于矩阵  $A \in \mathbb{F}^{m \times n}$ , 使用高斯消元法总可以将其转化为  $PA = LU$  的形式，使得  $P \in \mathbb{F}^{m \times m}$  *orthogonal* 为一个规范正交排列矩阵； $L \in \mathbb{F}^{m \times m}$  为下三角矩阵； $U \in \mathbb{F}^{m \times n}$  为上三角矩阵，称该分解为  $A$  的 PLU 分解。此时  $Ax = b$  等价于求  $LUx = Pb$ 。有时称这一分解为  $A$  的 *pivoted LU Factorization*。

下面给出利用高斯消元法手动求解矩阵  $A$  的 PLU 分解的具体流程

### PLU Factorization Using Gaussian Elimination

**STEP 1** 对  $A$  照常进行 Gaussian Elimination, 得到  $U$  并记录下每一步对应的行初等变换矩阵  $E_i$  s.t.

$$E_n \cdots E_1 A = U$$

**STEP 2** 对  $E_n \cdots E_1$  取逆，并将逆变换  $E_1^{-1} \cdots E_n^{-1}$  对应的行操作执行于单位矩阵  $I$  得到  $\tilde{L} = PL$  s.t.

$$E_1^{-1} \cdots E_n^{-1} I = \tilde{L}$$

**STEP 3** 将  $\tilde{L}$  整理成下三角得到  $L$ , 对比二者的行得到  $P$  对应的排列  $\tau$  使得  $\tau(\text{row index in } \tilde{L}) = \text{row index in } L$  即可，根据  $\tau$  得到对应的排列矩阵  $P_\tau$  使得  $P_\tau \tilde{L} = L$ , 最后有

$$P_\tau A = LU$$



### 3.2 Algorithm for LU Factorizations

本节介绍如何使用规范化的数值方法对矩阵  $A$  进行 (R)LU 分解。首先我们介绍一种最简单的 LU 分解作为引入，暂时先不妨假设矩阵  $A$  存在 LU 分解且为  $n$  阶方阵。为了统一流程，做如下的记号规定：

记  $R_1 := A \in \mathbb{F}^{n \times n}$ ，数域可为实数或复数域；对  $R_1$ ，即  $A$ ，以及随后递推产生的所有  $\{R_i\}_{i=1}^n$  ( $n$  等于矩阵  $A$  的阶)，用  $L_i$  和  $U_i^T$  分别表示他们的第  $i$  列/行；用  $r_i$  表示  $R_i$  的第  $(i, i)$  号元素（由推导也可以看出是  $R_i$  非零子阵的左上顶点元素）。我们用如下递推关系产生出余下的  $\{R_i\}_{i=1}^n$

$$i = 1 \cdots (n-1) \quad : \quad R_{i+1} = R_i - r_i^{-1} L_i U_i^T$$

**Lemma 3.2.1.** 在如上记号规定下，若  $A \in \mathbb{F}^{n \times n}$  存在 LU 分解，则

$$R_{n+1} := R_n - r_n^{-1} L_n U_n^T = 0$$

上述引理是一个显然的结果，实际上，由于  $r_1^{-1} L_1 U_1^T$  产生的矩阵的第一行一列与  $R_1$ ，也就是  $A$  完全相同。这导致  $R_2$ ，也即  $R_1$  与  $r_1^{-1} L_1 U_1^T$  的差矩阵的第一行一列为全 0 元素。实际上由归纳法不难看出，若用  $*$  表示矩阵位置上可能不为 0 的元素， $R_i$  与  $r_i^{-1} L_i U_i^T$  分别有如下结构

$$R_i = \begin{bmatrix} O & O & O & O \\ O & r_i & \cdots & r_{in} \\ O & \vdots & * & * \\ O & r_{ni} & * & * \end{bmatrix}$$

$$r_i^{-1} L_i U_i^T = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ r_i/r_i \\ \vdots \\ r_{ni}/r_i \end{bmatrix} \begin{bmatrix} 0 & \cdots & 0 & r_i & \cdots & r_{in} \end{bmatrix} = \begin{bmatrix} O & O & O & O \\ O & r_i & \cdots & r_{in} \\ O & \vdots & *' & *' \\ O & r_{ni} & *' & *' \end{bmatrix}$$

每一次迭代作差，都会使一行一列消失变为 0，直至到  $R_n$  时有  $R_n = r_n^{-1} L_n U_n^T$ 。之所以要求  $A$  存在 LU 分解，是因为需要确保总有  $r_i > 0$ 。上述引理也可以推广到非方阵的情形：

**Lemma 3.2.2.** 在如上记号规定下, 若  $A \in \mathbb{F}^{m \times n}$  存在 LU 分解, 则

$$d = \min(m, n) \quad \text{s.t.} \quad R_{d+1} := R_d - r_d^{-1} L_d U_d^T = 0$$

在上述引理的支持下, 有如下的 LU 分解定理及算法

**Theorem 3.2.1. Algorithm for LU Factorization**

在如上记号规定下, 若  $A \in \mathbb{F}^{m \times n}$  存在 LU 分解,  $d = \min(m, n)$ , 则有

$$A = LU = \begin{bmatrix} L_1/r_1 & \cdots & L_d/r_d \end{bmatrix} \begin{bmatrix} U_1^T \\ \vdots \\ U_d^T \end{bmatrix}$$

给出如下伪代码

---

**Algorithm 1:** LU Factorization

---

```

Initialize:  $d \leftarrow \min(m, n)$  ;
Initialize:  $R_1 \leftarrow A$  ;  $L \leftarrow []$  ;  $U \leftarrow []$  ;
1 for  $i = 1 : d$  do
2    $r_i, L_i, U_i \leftarrow R_i(i, i), R_i(:, i), R_i(i, :)$  ;
3   if  $r_i \neq 0$  then
4      $R_{i+1} \leftarrow R_i - r_i^{-1} L_i U_i^T$  ;
5   Append  $r_i^{-1} L_i$  to  $L$  as columns ; append  $U_i^T$  to  $U$  as rows ;

```

---

**Proof Theorem 3.2.1** 由递推关系, 我们有

$$\begin{aligned}
 R_2 &= R_1 - r_1^{-1} L_1 U_1^T \\
 R_3 &= R_2 - r_2^{-1} L_2 U_2^T \\
 &\vdots \\
 R_{d+1} &= R_d - r_d^{-1} L_d U_d^T
 \end{aligned}$$

两侧相加, 又有  $R_1 = A$ ,  $R_{d+1} = 0$ , 于是

$$A = r_1^{-1} L_1 U_1^T + \cdots + r_d^{-1} L_d U_d^T = \begin{bmatrix} L_1/r_1 & \cdots & L_d/r_d \end{bmatrix} \begin{bmatrix} U_1^T \\ \vdots \\ U_d^T \end{bmatrix} = LU$$

显然  $L, U$  分别为下三角和上三角矩阵。

### 3.3 Further Improvements for LU Factorizations

对于上一节的 Algorithm 1, 如果为  $n$  阶方阵, 第  $i$  个循环要实际执行  $n-i+1$  步除法,  $(n-i)^2$  步 (矩阵) 乘法与  $(n-i)^2$  步 (矩阵) 减法, 总计  $n$  步, 时间复杂度约为  $\frac{2}{3}n^3 \sim O(n^3)$ 。不过显然, 上述讨论的算法是有缺陷的, 比如并不是所有的矩阵都存在 LU 分解。我们可以使用一种名为 **Pivoting** 的技巧在上述算法的基础上稍加改进, 从而实现对任何可逆矩阵都能执行的 **PLU 分解**。

假设  $A \in \mathbb{F}^{n \times n}$  为可逆方阵, 仍然记  $L_i$ ,  $U_i^T$  和  $r_i$  分别为  $\{R_i\}_{i=1}^n$  的第  $i$  列/行以及第  $(i, i)$  号元素。不过这次对  $R_i$  的递推关系做如下改进:

$$i = 1 \cdots (n-1) \quad : \quad R_{i+1} = P_{i+1} (R_i - r_i^{-1} L_i U_i^T) \quad \text{where } R_1 = P_1 A$$

其中  $P_i$  为一个行互换初等矩阵  $P_i = E_{ij}$  for  $j \geq i$ , 它将原本余项  $R'_i = R_{i-1} - r_{i-1}^{-1} L_{i-1} U_{i-1}^T$  中的 leading pivot  $r'_i$  换为其所在列的 (modulus) 最大元素  $r'_j$  s.t.  $j \geq i$ 。这一步骤称为 **Pivoting**。在此基础上, 我们有

$$\begin{aligned} R_1 &= P_1 A \\ R_2 &= P_2 (R_1 - r_1^{-1} L_1 U_1^T) \\ R_3 &= P_3 (R_2 - r_2^{-1} L_2 U_2^T) \\ &\vdots \\ R_n &= P_n (R_{n-1} - r_{n-1}^{-1} L_{n-1} U_{n-1}^T) = r_n^{-1} L_n U_n^T \end{aligned}$$

于是, 从最后一个式子开始回代, 我们有

$$\begin{aligned} P_n \cdots P_2 P_1 A &= \sum_{i=1}^{n-1} P_n P_{n-1} \cdots P_{i+1} r_i^{-1} L_i U_i^T + r_n^{-1} L_n U_n^T \\ PA &= \begin{bmatrix} L'_1/r_1 & \cdots & L'_n/r_n \end{bmatrix} \begin{bmatrix} U_1^T \\ \vdots \\ U_n^T \end{bmatrix} = LU \end{aligned}$$

其中有  $P = P_n \cdots P_2 P_1$ , 以及  $L'_i = P_n P_{n-1} \cdots P_{i+1} L_i$ ,  $L'_n = L_n$ 。由于对列向量  $L_i$ ,  $P_k = E_{kj}$  for  $j \geq k$  只影响  $j, k$  两行, 而  $j \geq k \geq i+1 > i$ 。即  $P_n P_{n-1} \cdots P_{i+1}$  作用于  $L_i$  不会改变其形状 (从第  $i$  个元素开始不为 0)。因此  $L$  仍然是一个下三角矩阵。此外显然  $P = P_n \cdots P_2 P_1$  是一个排列矩阵。

**Theorem 3.3.1. Algorithm for Pivoted LU Factorization**

在如上记号规定下, 若  $A \in \mathbb{F}^{n \times n}$  为可逆方阵, 总存在 PLU 分解

$$PA = \begin{bmatrix} L'_1/r_1 & \cdots & L'_n/r_n \end{bmatrix} \begin{bmatrix} U_1^T \\ \vdots \\ U_d^T \end{bmatrix} = LU$$

其中有  $P = P_n \cdots P_2 P_1$ , 以及  $L'_i = P_n P_{n-1} \cdots P_{i+1} L_i$ ,  $L'_n = L_n$ 。  $P_i$  为一个行互换初等矩阵  $P_i = E_{ij}$  for  $j \geq i$ , 将原本余项  $R'_i = R_{i-1} - r_{i-1}^{-1} L_{i-1} U_{i-1}^T$  中的 *leading pivot*  $r'_i$  换为其所在列的 (*modulus*) 最大元素  $r'_j$  s.t.  $j \geq i$ 。算法伪代码如下

**Algorithm 2:** Pivoted LU Factorization (PLU)

---

```

Initialize:  $R \leftarrow P_1 A$  ;  $P \leftarrow I_n$ 
Initialize:  $L \leftarrow []$  ;  $U \leftarrow []$  ;

1 for  $i = 1 : n$  do
2      $r_i, L_i, U_i \leftarrow R(i, i), R(:, i), R(i, :)$  ;
3     Append  $r_i^{-1} L_i$  to  $L$  as columns ; append  $U_i^T$  to  $U$  as rows ;
4     if  $i < n$  then
5          $R \leftarrow R - r_i^{-1} L_i U_i^T$  ;
6          $R \leftarrow P_{i+1} R$  ;
7          $P \leftarrow P_{i+1} P$  ;
8         for  $col$  in  $L$  do
9              $col \leftarrow P_{i+1} col$  ;

10  $P \leftarrow P P_1$  ;
```

---

注意上述伪代码中,  $P_1$  必须保留至最后, 因为在第十行之前, 循环结束后伪代码给出的  $P = P_n \cdots P_2$ 。在第六行之前, 也最好检查一下目标列是不是全 0 列。该算法的复杂度 (不考虑遍历寻找  $P_i$ ) 仍然为  $\frac{2}{3}n^3 \sim O(n^3)$ 。使用 Pivoting 后算法稳定性有所提升, 但仍可能出现不稳定的情形。