数值代数课程总结

徐嘉声

Jason　Xu

北京师范大学　会同书院

2022 年 12 月 26 日

前言

这份笔记的制作初衷在于总结数值代数这门课程一个学期的所学与所感，并在数学知识的探索过程中留下自己的足迹，而非随着时间的推移让自己曾付出的努力全然销声匿迹。

首先，要感谢朱圣鑫老师的启发式教学以及对于这份课程作业的要求，让我进行了许多对于数值代数知识和算法优劣性的自主探索，并在期末之前总结好了课程知识点的笔记手写版。在整个学习周期中，我不仅理清了数值代数的基本框架，还与仝羽翔同学一起动手实践，对 SVD 算法进行一步步调优，实现了知识的深入了解，而非概念上的浅尝辄止。犹记得在选课结束前的周末，我曾一度因刚开始的 C 语言与 Matlab 的混合编译而打退堂鼓，害怕自己拿不到这个学分，但是在朱老师的鼓励下，我得到了一个学期满满的收获。

在课堂教学上，得益于小班制教学的缘故，我们每个人都拥有对数值代数重要知识点到黑板上进行推导的机会，但是由于我薄弱的高代基础，我往往没有办法把他们推导出来，但我总能深刻的记住最终的正确推导方法以及推导思路。在完成了数值代数基本知识的讲解后，最后几节课的代码分析讲解也令我收益匪浅，通过在课堂上一行行讲解、修改与优化同学们的 SVD 代码，我对这个算法的每一个细节都有了更深的理解，在追求代码效率的同时也掌握了不少提升算法精度的技巧。

在课堂内容上，课程的 PPT 以及参考文献大都是英文的，这也是我大学课程中第一次使用英文的资料，起初的不适是必然的，但是一个学期下来，这项阅读英文文献的能力已有了大大的提升，早已丢掉了课程伊始的那种恐惧感。在课后作业上，朱老师抛去了传统做题的方式，把重心放在了课堂中对知识点的理解上，并通过小组编写 SVD 算法代码的 Project，带着我们去深入算法的速度与精度，把后半学期所学的特征值知识点全部串了起来。

数值代数的课程学习脉络从整体上可以分为三个部分,分别是矩阵的存储与计算、求解线性方程组的直接方法以及矩阵特征值的计算。在第一个部分中主要探讨了大型稀疏矩阵的三种节省内存的存储方式，并通过 CSparse 的代码讲述了 Matlab 进行矩阵加减乘除运算的基本逻辑；在第二个部分中则是通过矩阵基础知识的复习引入高斯消元法、矩阵的 LU 三角分解法、对称矩阵的 Cholesky 和 LDLT 分解法以及矩阵的范数概念；在第三各部分则是研究矩阵的特征值问题，从特征值的估计算法和计算主特征值的幂法与反幂法，进入到了使用 Householder 变换和 Givens 变换的 QR 分解方法，并使用原点位移的 Shift 方式对 QR 方法进行进一步优化，最后通过 PCA 和 SVD 算法对整个特征值的计算进行了总结与深入。

以下正文部分，是本人结合自身的理解对于本学期所学核心重难点知识的整理。

徐嘉声

北京师范大学

2022 年 12 月 26 日

目录

第一章　稀疏存储

稀疏矩阵的存储主要有三个方式，分别是压缩行、压缩列以及三元组的方式，他们的核心思想都是把值为 0 的元素忽略，并通过不同的压缩方式来存储非零元的位置，下面用一个例子即可以帮助我们清晰的进行理解。

$$\text{e.g. } A = \begin{bmatrix} 1 & 0 & 5 & 0 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & 6 & 1 \\ 2 & 0 & 0 & 3 \end{bmatrix}$$

1. CSR (压缩稀疏行)

行指针	int p[ ]	0		2		4		6		8
列指标	int j[ ]	0	2	1	3	2	3	0	3	
非零元	int x[ ]	1	5	2	4	6	1	2	3	

2. CSC (压缩稀疏列)

列指针	int p[ ]	0		2	3		5			8
行指标	int i[ ]	0	3	1	0	2	1	2	3	
非零元	int x[ ]	1	2	2	5	6	4	1	3	

以上行指针代表元素个数，每当换行的时候进行存储，其中后面标红的**8**，代表着矩阵的结束，列指标代表该非零元所在的列，后面压缩稀疏列同理。

3. Zero Based Triplet（三元组存储法）

int x[ ]	1	5	2	4	6	1	2	3
int i[ ]	0	0	1	1	2	2	3	3
int j[ ]	0	2	1	3	2	3	0	3

第二章　线性方程组的解法

2.1　高斯消元

我们以如下的方程组为例：

$$\begin{cases} -3x_1 + 2x_2 + 6x_3 = 4 \\ 10x_1 - 7x_2 = 7 \\ 5x_1 - x_2 + 5x_3 = 6 \end{cases}$$

首先将他化为增广矩阵的形式，此时便可以进行高斯消元了

$$\Rightarrow \left[\begin{array}{ccc|c} -3 & 2 & 6 & 4 \\ 10 & -7 & 0 & 7 \\ 5 & -1 & 5 & 6 \end{array}\right] \xrightarrow{r_1 \leftrightarrow r_2} \left[\begin{array}{ccc|c} 10 & -7 & 0 & 7 \\ -3 & 2 & 6 & 4 \\ 5 & -1 & 5 & 6 \end{array}\right] \xrightarrow[r_2 + \frac{3}{10} r_1]{r_3 + \left(-\frac{5}{10}\right) r_1}$$

$$\begin{bmatrix} 10 & -7 & 0 & 7 \\ 0 & -\frac{1}{10} & 6 & \frac{61}{10} \\ 0 & \frac{5}{2} & 5 & \frac{5}{2} \end{bmatrix} \xrightarrow[r_3 + \frac{1}{25} r_2]{r_3 \leftrightarrow r_2} \begin{bmatrix} 10 & -7 & 0 & 7 \\ & \frac{5}{2} & 5 & \frac{5}{2} \\ & & \frac{31}{5} & \frac{31}{5} \end{bmatrix}$$

$$x_3 = \frac{\frac{31}{5}}{\frac{31}{5}} = 1, \qquad x_2 = \left(\frac{5}{2} - 5x_3\right) / \frac{5}{2} = -1, \qquad x_1 = 0$$

2.2　线性方程组解的存在条件

秩：增广矩阵化为行阶梯形时，系数矩阵的列数

$$\text{e.g. } (A,b) \sim \left[\begin{array}{ccc|c} 1 & 2 & 2 & -1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 \end{array}\right] \quad R(A) \neq R(A,b) \quad \text{无解，超定}$$

$$(A,b) \sim \left[\begin{array}{ccc|c} 1 & 2 & 2 & -1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad R(A) = R(A,b) < n \quad \text{无穷解，欠定}$$

$$(A,b) \sim \left[\begin{array}{ccc|c} 1 & 2 & 2 & -1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right] \quad R(A) = R(A,b) = n \quad \text{唯一解，正定}$$

2.3　LU 分解

此处不过多赘述 LU 分解中的定义和定理，只给出几个重要的证明。

1、LU 分解的唯一性证明

定理 2.3.1. 若 A 为 n 阶矩阵，那么 A 的顺序主子式 $D_i \neq 0, i = 1, 2, ..., n$，那么此时 A 可分为一个单位下三角 $L \times U$，且其是唯一的。

证：存在性在高斯消元过程中已经证明过了，下证唯一性：若 A 为非奇异矩阵，那么令 $A = LU = L'U'$，L, L' 为单位下三角矩阵，U, U' 为上三角矩阵，那我们便可以得到：

$$L^{-1}L' = UU'^{-1}$$

此时左式为下三角，右式为上三角，若他们相等，则均为单位矩阵，故有 $U = U', L = L'$ 成立，唯一性得证。

2、LU 中左乘矩阵的可逆性

定理 2.3.2. *若* $i \neq j, M(i,j,\lambda)^{-1} = M(i,j,-\lambda)$

证明：

注意到 $\mathbf{e}_j^T\mathbf{e}_i = \delta_{ij}$，故在 $i \neq j$ 时，$\mathbf{e}_j^T\mathbf{e}_i = 0$。我们首先对下三角矩阵的乘法进行拆解，消去 $\lambda\mathbf{e}_i\mathbf{e}_j^T$ 项，再根据本章节开始所提到的结合律，便能证明上述引理。

$$
\begin{aligned}
&M(i,j,\lambda)M(i,j,-\lambda) = \left(I + \lambda\mathbf{e}_j^T\right)\left(I - \lambda\mathbf{e}_i\mathbf{e}_j^T\right) \\
&= I + \lambda\mathbf{e}_i\mathbf{e}_j^T - \lambda\mathbf{e}_i\mathbf{e}_j^T + \lambda^2\mathbf{e}_i\mathbf{e}_j^T\mathbf{e}_i\mathbf{e}_j^T \\
&= I + \lambda^2\mathbf{e}_i\underbrace{\left(\mathbf{e}_j^T\mathbf{e}_i\right)}_{0}\mathbf{e}_j^T = I
\end{aligned}
$$

2.4　Cholesky 分解（对称正定矩阵的 LU 分解）

正定性的含义为：$\mathbf{x}^{\mathbf{T}}A\mathbf{x} > 0$ 对 $\forall x \neq 0$ 成立。

以下公式中，我们默认 $\mathbf{a_21}$ 为列向量。

基本: $\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} \sqrt{a} & 0 \\ u & v \end{bmatrix}\begin{bmatrix} \sqrt{a} & u \\ 0 & v \end{bmatrix} = \begin{bmatrix} a & u\sqrt{a} \\ u\sqrt{a} & u^2+v^2 \end{bmatrix}$

$\Rightarrow b = u\sqrt{a}, u = \dfrac{b}{\sqrt{a}}/c = u^2 + v^2 \rightarrow v = \sqrt{c - b^2/a}$

以下为计算中的一些特殊形式的 Cholesky 分解：

（1）**Right-Looking**（从左上向右下）

$$
\begin{bmatrix} l_{11} & \\ \overrightarrow{l_{21}} & L_{22} \end{bmatrix}\begin{bmatrix} l_{11} & \overrightarrow{l_{21}^{\top}} \\ & L_{22}^{\top} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11}\overrightarrow{l_{21}^{\top}} \\ l_{11}\overrightarrow{l_{21}} & L_{22}L_{22}^{\top} + l_{21}l_{21}^{\top} \end{bmatrix} = \begin{bmatrix} a_{11} & \overrightarrow{a_{21}} \\ \overrightarrow{a_{21}} & A_{22} \end{bmatrix}
$$

$$
\begin{cases} l_{11} = \sqrt{a_{11}} & (1) \\ l_{21} = \frac{a_{21}}{l_{11}} & (2) \quad \Rightarrow (n-1)\times(n-1) \\ L_{22}L_{22}^T = A_{22} - l_{21}l_{21}^{\top} & (3) \end{cases}
$$

(2) **Up-Loking** (从左下往右上)

$$\begin{bmatrix} R_{11}^\top & \\ \overrightarrow{r_{12}^\top} & r_{22} \end{bmatrix} \begin{bmatrix} R_{11} & \overrightarrow{r_{12}} \\ & r_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & \overrightarrow{a_{12}} \\ \overrightarrow{a_{12}^\top} & a_{22} \end{bmatrix}$$

$$\begin{cases} R_{11}^\top \cdot (R_{11} = A_{11} & \\ R_{11}^\top \vec{r}_{12} = \overrightarrow{a_{12}} & (1) \\ \vec{r}_{12}^\top \cdot \vec{r}_{12} + r_{22}^2 = a_{22} & (2) \end{cases}$$

$$\Rightarrow r_{22} = \sqrt{a_{22} - \overrightarrow{r_{12}^\top} \cdot \overrightarrow{r_{12}}}$$

(3) **Left-Looking**

$$\begin{bmatrix} L_{11} & & \\ \overrightarrow{l_{21}^\top} & l_{22} & \\ L_{31} & \overrightarrow{l_{32}} & L_{33} \end{bmatrix} \begin{bmatrix} L_1^\top & \overrightarrow{l_{12}} & L_{31}^\top \\ & l_{22} & \overrightarrow{l_{32}} \\ & & L_{33}^\top \end{bmatrix} = \begin{bmatrix} A_{11} & \overrightarrow{a_{12}} & A_{13}^\top \\ \overrightarrow{a_{12}}^\top & a_{22} & \overrightarrow{a_{32}}^\top \\ A_{31} & \overrightarrow{a_{32}} & A_{33} \end{bmatrix}$$

$$\vec{l}_{12}\vec{l}_{12} + l_{22}^2 = a_{22} \to l_{22} = \sqrt{a_{22} - \overrightarrow{l_{12}}\overrightarrow{l_{12}}}$$

$$L_{31}\vec{l}_{12} + l_{22} \cdot \overrightarrow{l_{22}}\vec{a}_{32} \to \overrightarrow{l_2} = \frac{(\overrightarrow{a_{32}} - l_{31}l_{21})}{l_{22}}$$

(4) **LDLT' 分解**

$$\begin{bmatrix} 1 & \\ \vec{l} & I \end{bmatrix} \begin{bmatrix} d_{11} & \\ & D \end{bmatrix} \begin{bmatrix} 1 & \overrightarrow{l^\top} \\ & I \end{bmatrix} = \begin{bmatrix} a_{11} & \overrightarrow{a_{12}^\top} \\ \overrightarrow{a_{21}} & A_{22} \end{bmatrix}$$

$d_{11} = a_{11}, d_{11} \cdot \vec{l} = \overrightarrow{a_{21}}, \to \vec{l} = \overrightarrow{a_{21}}/d_{11}$

$\vec{l}d_{11}\overrightarrow{l^\top} + D = A_{22} \to D = A_{22} - \vec{l}d_{11}\overrightarrow{l^T}$

对称正定的 A 可分解为 LDL^T 的形式，证明如下：

$\because$ A 可唯一分解为 LU，那么令 U 为一个对角矩阵和单位上三角矩阵的乘积，D 为对角阵，U_0 为单位上三角矩阵，那么 $A = LU = LDU_0$

因此 $A = A^\top = U_0^\top DL^\top \Rightarrow U_0^\top = L$

第三章　特征值的计算

3.1　特征值的估计：圆盘定理

对于 $A=(a_{ij})_{n\times n}$，Di 为 Gershgorin 圆盘

1) $r_i=\sum\limits_{j=1,j\neq i}^{n}|a_{ij}|, i=1,2,\cdots,n$

2) $D_i=\{z||z-a_{ii}|\leqslant r_i, z\in\mathbb{C}\}$

$\forall$ 特征值均在某一圆盘中，且 $|\lambda-a_{ii}|\leqslant r_i=\sum_{j=1 j\neq i}^{n}|a_{ij}|$

e.g. $A=\begin{pmatrix} 0 & 1 & 0 & i \\ 1 & 6 & 1 & i \\ \frac{i}{2} & i & 5i & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & -2 \end{pmatrix}$, 求 A 的特征值分布

$D_1=\{z:|z|\leq|a_{12}|+|a_{13}|+|a_{14}|=2\}$

$D_2=\{z:|z-6|\leq|a_{21}|+|a_{23}|+|a_{24}|=3\}$

$D_3=\left\{z:|z-5_i|\leq|a_{31}|+|a_{32}|+|a_{34}|=\frac{3}{2}\right\}$

$D_4=\{z:|z+2|\leq|a_{41}|+|a_{42}|+|a_{43}|=1\}$

3.2　幂法与反幂法 (power method)

作用是求解矩阵的最大特征值，核心是求解矩阵的特征向量，以下为幂法的一些基本定理：

定理 3.2.1. 若 $A\overrightarrow{x}=\lambda\overrightarrow{x}$，那么有

（*1*）$\forall a\neq 0$ 且 $a\in\mathbb{R}$

（*2*）$A-\mu I$ 的特征值为 $\lambda-\mu$

（*3*）A^{-1} 的特征值为 λ^{-1}

利用以上三条定理，我们便可以得到幂法、带位移的幂法以及反幂法

1、幂法

若 $A=(a_{ij})_{n\times n}$ 有完全特征向量组，且：$|\lambda_1|>|\lambda_2|\geqslant|\lambda_3|\geqslant\cdots\geqslant|\lambda_n|$

那么对任意的 $\vec{v}_0$ 有：$\begin{cases}\overrightarrow{v_1}=A\overrightarrow{v_0}\\ \overrightarrow{v_2}=A\overrightarrow{v_1}=A^2v_0\\ \vdots\\ \overrightarrow{v_k}=A\overrightarrow{v_{k-1}}=A^kV_0\end{cases}$

令 $\vec{v}_0=\alpha_1\vec{x}_1+\alpha_2\vec{x}_2+\cdots+\alpha_n\vec{x}_n\ (\ \alpha_1\neq 0)$

$\overrightarrow{v_k}=A\vec{v}_{k-1}=A^k\vec{v}_0=\alpha_1\lambda_1\vec{x}_1+\alpha_2\lambda_2^k\vec{x}_2+\cdots+\alpha_n{\lambda_n}^k\overrightarrow{x_n}$

$={\lambda_1}^k\left[\alpha_1\vec{x}_1+\sum_{i=2}^n\alpha_i\left(\lambda_i/\lambda_1\right)^k\vec{x}_i\right]={\lambda_1}^k\left(\alpha_1\vec{x}_1+\vec{\varepsilon}_k\right)$

其中 $\varepsilon_k=\sum_{i=2}^n\alpha_i\left(\lambda_i/\lambda_1\right)^k\overrightarrow{x_i}\ \ \because|\lambda_i/\lambda_1|<1$

故 $\lim_{k\to\infty}\varepsilon_k=0$ 即 $\lim_{k\to\infty}\frac{\nu_k}{\lambda_1^k}=\alpha_1\vec{x}_1$, 故 $\overrightarrow{v_k}\approx\alpha_1{\lambda_1}^k\overrightarrow{x_1}$

即迭代向量 $\overrightarrow{v_k}$ 为 λ_1 特征向量的近似值。

右式为计算 λ 的公式：$\dfrac{\left(\overrightarrow{V_{k+1}}\right)_i}{(\sqrt{k})_i}=\lambda_1\left\{\dfrac{\alpha_1\left(\vec{x}_1\right)_i+\left(\vec{\varepsilon}_{k+1}\right)_i\}}{\alpha_1\left(\vec{x}_1\right)_i+\left(\overrightarrow{\varepsilon_k}\right)\overline{i}}\right\}$

$\Rightarrow\lambda_1=\lim\limits_{k\to\infty}\dfrac{(\overrightarrow{\sqrt{k+1}})_i}{(\sqrt{k})_i}$

2、Shift

由于 A 的收敛速度是由 $r=\lambda_2/\lambda_1$ 所决定的，因此 $r\to 1$ 时收敛时间过程，因此我们令 $B=A-pI$，A 的特征值为 $\lambda_1,\lambda_2,...,\lambda_n$，B 为 $\lambda_1-p,...,\lambda_n-p$，A，B 的特征向量相同。

3、反幂法

反幂法的使用方法与幂法类似，他的作用在于求解矩阵的最小特征值问题。此时我们只需对 λ 取倒数便可以通过幂法的方式，求解出矩阵的最小特征值。

3.3　正交变换

3.3.1　Householder 变换 (一次消去一列)

对 $\overrightarrow{w} \in \mathbb{R}^n$，有 $\overrightarrow{w^T} \cdot \overrightarrow{w} = 1$，称 $H(w) = I - 2\overrightarrow{w}\overrightarrow{w^T}$ 为 Householder 变换，其中对任意的 $\overrightarrow{x}$，均有 $H\overrightarrow{x} = \overrightarrow{y}$，且 $||\overrightarrow{x}|| = ||\overrightarrow{y}||$ 的性质，如下为 Householder 变换的约化定理

$$\vec{x} = (x_1, x_2, x_3, \cdots x_n)^\top \neq 0, \exists H,\text{ 使} H\vec{x} = -\sigma\overrightarrow{e_1}$$

$$\begin{cases} H = I - \beta^{-1}\vec{u}\vec{u}^\top \\ \sigma = \operatorname{sgn}(x_1)\|\vec{x}\|_2 \\ \vec{u} = \vec{x} + \sigma\overrightarrow{e_1} \\ \beta = \dfrac{1}{2}\|\vec{u}\|_2^2 = \sigma(\sigma + x_1) \end{cases}$$

证明：令 $\vec{y} = -\sigma\vec{e}_1$, 令 $x \neq y, \sigma = \pm\|x\|_2$。故 $\|x\|_2 = \|y\|_2$

故存在 Householder 变换 $H = I - 2\vec{w} \cdot \vec{w}^\top$

其中 $W = \dfrac{\vec{x} + \sigma\vec{e}_1}{\|\vec{x} + \sigma\vec{e}_1\|_2} \cdot H\vec{x} = \vec{y} = -\sigma\vec{e}_1$ 取 $\vec{u} = \vec{x} + \sigma\overrightarrow{e_1}$,

故 $H = I - 2\dfrac{u \cdot u^\top}{\|u\|_2^2} = I - \beta^{-1}\vec{u} \cdot \vec{u}^\top$ $\vec{u} = (x_1 + \sigma \cdot x_2, \cdots x_n)^\top, \beta = \dfrac{1}{2}\|\vec{u}\|_2^2$

那么 $\beta = \dfrac{1}{2}\|\vec{u}\|_2^2 = \dfrac{1}{2}\left((x_1 + \sigma)^2 + x_2^2 + \cdots + x_n^2\right)$

$= \dfrac{1}{2}\left(x_1^2 + 2x_1\sigma + \sigma^2 + x_2^2 + \cdots + x_n^2\right)$

$= \dfrac{1}{2}(2\sigma^2 + 2x_1\sigma) = \sigma(\sigma + x_1)$

若 σ 和 x_1 异号，则 $(\sigma + x_1)$ 可能为 0，故取 $\sigma = \operatorname{sgn}(x_1)\|\vec{x}\|_2$

应用： $\vec{x}=(3,5,1,1)^\top, \|\vec{x}\|_2=\pm 6.\quad k_1=6, k_2=-6.$

$$\vec{u}_1=\vec{x}+6e_1=(9,5,1,1)^\top, \|\vec{u}\|_2^2=108, \beta=54$$
$$\vec{H}=I-\beta^{-1}uu^\top \quad ,Hx=(-6,0,0,0)$$

3.3.2 Givens 变换 (将某个位置的数字变为 0)

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \text{即}\vec{y}=P\vec{x}$$

令 $\vec{x}=(x_1,x_2,\cdots x_i,\cdots,x_j,\cdots,x_n)^\top$, 其中 x_i,x_j 不全为 0

那么有 $P(i,j,\theta)$，使 $.13P\vec{x}=(x_1,\cdots x_i',\cdots,0,\cdots x_n)^\top$, 其中

${x_i}'=\sqrt{x_i^2+x_j^2}$. $\theta=\arctan(x_j/x_i)$

proof: 取 $c=\cos\theta=x_i/x_i', s=\sin\theta=x_j/{x_i}'$

根据矩阵乘法，有

$$\begin{cases} x_i'=cx_i+sx_j \\ x_j'=-sx_i+cx_j \\ x_k'=x_k, k\neq i,j \end{cases} \Rightarrow \begin{cases} x_i'=\frac{x_i^2}{x_i'}+\frac{x_1^2}{x_i'} \\ x_j'=-\frac{x_ix_i}{x_i'}+\frac{x_ix_j}{x_i'}=0 \end{cases}$$

$$\Rightarrow x_i'=\sqrt{x_i^2+x_j^2}\cdot x_j'=0$$

3.4 QR 分解

设 A 为非奇异的，那么有正交的 P，使 PA=R，R 为上三角矩阵

$P_kA^{(k)}=P_kP_{k-1}\cdots P_1A=A^{(k+1)}\quad (P_k$ 是将 $(k,k+1:n)$ 变为 0)

$\Rightarrow P=P_{n-1}\cdots P_1\quad ,PA=R$。

由于 Q 为正交阵，R 为上三角阵，因此 $R\sim A$

那么 $D=\operatorname{diag}\left(\frac{r_n}{|r|r\ |},\cdots,\frac{r_{nn}}{|r_{mn}|}\right)$

$$\begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & * & * \\ 0 & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \\ 0 & 0 & * \end{pmatrix}$$

3.5 QR 方法

首先使用 QR 分解，将矩阵变为上 Hassenberg 矩阵或者三对角矩阵，后进行迭代。

普通 $QR:\begin{cases} A^{(k)}=Q_kR_k. \\ A^{(k+1)}=R_kQ_k. \end{cases},A^{(k+1)}=Q_k^{\top}AQ_k.$

下面进行 Shift QR 的证明

证：已知 Q 为正交矩阵，那么 $Q^T=Q^{-1}$

$$\begin{aligned} A^{(k+1)} &= R_kQ_k+\mu I=\left(Q_k^{\top}Q_kR_kQ_k+\mu I\right. \\ =&Q_k^{\top}\left(A^{(k)}-\mu I\right)Q_k+\mu I \\ =&Q_k^{\top}A^{(k)}Q_k-Q_k^{\top}\mu IQ_k+\mu I \\ =&Q_k^{\top}A^{(k)}Q_k \end{aligned}$$

3.6 奇异值分解 (SVD)

奇异值分解是整个特征值问题部分的集大成者, 他的目的是解决 $m\times n$ 矩阵正交分解问题的方法, 在算法上它利用了矩阵与其转置相乘的方法, 分别计算出了 U,Σ,V 的值，计算原理如下所示

$$\begin{aligned} M &= U\Sigma V^{\top} \\ MM^{\top} &= U\Sigma\Sigma^{\top}U^{\top} \\ M^{\top}M &= V\Sigma^{\top}\Sigma V^{\top} \end{aligned}$$

但是在计算机的计算过程之中，我们不能直接使用上述的算法，否则会降低整体的速度，优化后的算法如下所示：

1. 将整个矩阵进行 qr 分解得到上三角部分
2. 使用 householder 变换将其变为二对角矩阵
3. 矩阵相乘进行三对角化
4. 使用 Givens 变换对三对角化的矩阵进行 Shift-QR 迭代
5. 根据我们迭代过程中得到的值分别得出 U, Σ, V