第五章　矩阵特征值问题的数值方法

数值分析

南京理工大学数统学院

5.1 特征值的估计与扰动

5.1.1 特征值的估计

1. **定理5.1:**（盖尔圆定理）设 $\boldsymbol{A} = (a_{ij})_{n\times n} \in \mathbb{C}^{n\times n}$，则 $\boldsymbol{A}$ 的特征值有

$$\lambda \in \bigcup_{i=1}^{n} D_i$$

其中

$$D_i = \{z : |z - a_{ii}| \leq \sum_{j=1, j\neq i}^{n} |a_{ij}|\} \quad (i = 1, 2, \cdots n)$$

为复平面上的圆盘，D_i 也称为矩阵 $\boldsymbol{A}$ 的第 i 个盖尔圆，$R_i = \sum_{j=1, j\neq i}^{n} |a_{ij}|$ 为盖尔圆 D_i 的半径。

证明： 设 λ 是 $\boldsymbol{A}$ 的任意特征值，$\boldsymbol{x}$ 为 $\boldsymbol{A}$ 对应于 λ 的特征向量，设 x_i 是特征向量 $\boldsymbol{x}$ 中按模最大的分量，由于 $(\boldsymbol{Ax})_i = \lambda x_i$，

即 $\lambda x_i = \sum\limits_{j=1}^{n} a_{ij}x_j$，则有 $(\lambda - a_{ii})x_i = \sum\limits_{j=1,j\neq i}^{n} a_{ij}x_j$。

注意到 $|x_j| \leq |x_i| (j \neq i)$，则由上式得

$$|\lambda - a_{ii}| \leq \sum_{j=1,j\neq i}^{n} |a_{ij}| \frac{|x_j|}{|x_i|} \leq \sum_{j=1,j\neq i}^{n} |a_{ij}|.$$

于是 $\lambda \in D_i$。证毕■

注： $\boldsymbol{A}$ 的全体特征值也都在 $\boldsymbol{A}^T$ 的 n 个盖尔圆构成的并集中，称 $\boldsymbol{A}^T$ 的盖尔圆为 $\boldsymbol{A}$ 的列盖尔圆。

注： 定理5.1说明$\boldsymbol{A}$的n个特征值均落在n 个圆盘的并集中, 但未说明特征值的具体分布情况, 并不是指每个圆盘内都有特征值。

注： $\boldsymbol{A}$ 的全体特征值也都在 $\boldsymbol{A}^T$ 的 n 个盖尔圆构成的并集中，称 $\boldsymbol{A}^T$ 的盖尔圆为 $\boldsymbol{A}$ 的列盖尔圆。

注： 定理5.1说明$\boldsymbol{A}$的n个特征值均落在n 个圆盘的并集中，但未说明特征值的具体分布情况，并不是指每个圆盘内都有特征值。

定理5.2: 设矩阵的 n 个盖尔圆圆盘中，有 m 个圆盘相交构成一个连通域 S，且 S 与其余 $n-m$ 个圆盘严格分离，则S 中恰有 $\boldsymbol{A}$ 的 m 个特征值，其中重根特征值的个数按其重数计算。

例： 估计下面矩阵的特征值范围

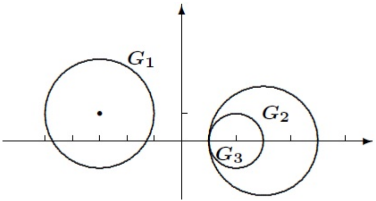
$$\boldsymbol{A}=\begin{pmatrix} -3+i & 0 & 2 \\ -1 & 3 & i \\ 0 & -1 & 2 \end{pmatrix}$$

解： 矩阵 $\boldsymbol{A}$ 的 3 个盖尔圆为

$G_1 : |z+3-i| \le 2$

$G_2 : |z-3| \le 2$

$G_3 : |z-2| \le 1,$



矩阵 $\boldsymbol{A}$ 的特征值在 G_1 中有一个，在 $G_2 \cup G_3$ 中有两个；

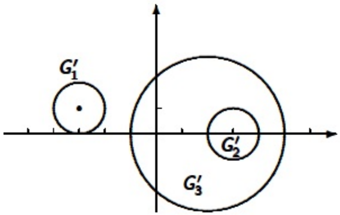
矩阵 A 的 3 个列盖尔圆为

$G_1': |z+3-i| \le 1,$

$G_2': |z-3| \le 1,$

$G_3': |z-2| \le 3,$

故 A 的 3 个特征值有一个在 G_1' 中，两个在 $G_2' \cup G_3'$ 中。



在这里，矩阵 A 的其中一个特征值独占一个盖尔圆，是可以分出来的，而另外两个特征值只能判断是在 $G_2 \cup G_3$ 或$G_2' \cup G_3'$ 中，无法判断其位置（有可能都在小圆外面，而小圆内一个都没有）。

盖尔圆隔离的基本思想: 如果每个盖尔圆都是独立的，则每个盖尔圆内有一个特征值。

1. 结合列盖尔圆实现特征值隔离.

例: 应用盖尔圆定理隔离

$$\boldsymbol{A} = \begin{pmatrix} 20 & 3 & 2 \\ 2 & 10 & 4 \\ 4 & 0.5 & 0 \end{pmatrix}$$

的特征值。

解： $\boldsymbol{A}$ 的三个盖尔圆为

$$G_1:|z-20|\le 5,\ G_2:|z-10|\le 6,\ G_3:|z|\le 4.5$$

显然，G_1, G_2, G_3 相交。

而 $\boldsymbol{A}$ 的三个列盖尔圆为

$$\overline{G}_1:|z-20|\le 6,\ \overline{G}_2:\ |z-10|\le 3.5,\ \overline{G}_3:|z|\le 6,$$

它们相互分离。

故在 $\overline{G}_1$，$\overline{G}_2$，$\overline{G}_3$ 中各有 $\boldsymbol{A}$ 的一个特征值。

2. 利用相似变换调节盖尔圆半径，实现特征值隔离.

对于 $\boldsymbol{A}=(a_{ij})_{n\times n}$，设 $\boldsymbol{D}=\mathrm{diag}(d_1,d_2,\cdots,d_n)$，$(d_i>0,\ i=1,2,\cdots,n)$，则

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

此时 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 有相同特征值。

注意到：$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 具有相同的对角线元素（即盖尔圆中心不变），但 $\boldsymbol{B}$ 的盖尔圆盘半径为

$$R_i=\sum_{j=1,j\neq i}^{n}\left|\frac{a_{ij}d_j}{d_i}\right|,$$

发生了改变，其具体大小可通过 $\boldsymbol{D}$ 的对角元素取值来调节。

通过适当选取 D 的元素值，可以使某个圆盘半径相对减少或增加。

一般地，要使 A 的第 i 个盖尔圆半径放大，其余盖尔圆适量缩小（相对于 A 的同序号盖尔圆），我们可取 $d_i < 1$，其它元素值取 1;

相反，若使 A 的第 i 个盖尔圆半径缩小，其余盖尔圆适量放大（相对于 A 的同序号盖尔圆），我们可取$d_i > 1$，其它元素值取 1.

例： 应用盖尔圆定理隔离证明

$$\boldsymbol{A}=\begin{pmatrix}2 & 2 & -1\\ 1 & 10 & -1\\ 8 & 2 & 20\end{pmatrix}$$

有三个互异的实数特征值。

例： 应用盖尔圆定理隔离证明

$$A = \begin{pmatrix} 2 & 2 & -1 \\ 1 & 10 & -1 \\ 8 & 2 & 20 \end{pmatrix}$$

有三个互异的实数特征值。

解： A 的三个盖尔圆为

$$G_1 : |z-2| \le 3,\ G_2 : |z-10| \le 2,\ G_3 : |z-20| \le 10.$$

G_2 与 G_3 相交。

取 $\boldsymbol{D}=\mathrm{diag}(\frac{1}{2},1,1)$，则

$$\boldsymbol{B}=\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}=\begin{pmatrix} 2 & 4 & -2 \\ 0.5 & 10 & -1 \\ 4 & 2 & 20 \end{pmatrix}.$$

则 $\boldsymbol{B}$ 的三个盖尔圆为

$$\overline{G}_1:|z-2|\leq 6,\ \overline{G}_2:\ |z-10|\leq 1.5,\ \overline{G}_3:|z-20|\leq 6,$$

它们相互分离。

若实矩阵 $\boldsymbol{A}$ 的特征值 λ 为复数，则 $\overline{\lambda}$ 也为特征值，但 $\overline{G}_1,\overline{G}_2,\overline{G}_3$ 关于实轴对称，则 $\boldsymbol{A}$ 特征值必皆为实数，故在区间 $[-4,8],[8.5,11.5],[14,26]$ 中各有 $\boldsymbol{A}$ 的一个实特征值。

5.1.2 特征值的扰动

定理5.3: 设$\boldsymbol{E}$为矩阵$\boldsymbol{A}$的扰动矩阵, μ是矩阵$\boldsymbol{A}+\boldsymbol{E}\in\mathbb{C}^{n\times n}$的一个特征值, 且存在$\boldsymbol{X}\in\mathbb{C}^{n\times n}$, $\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}=\boldsymbol{D}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$, 则

$$\min_{\lambda\in\sigma(\boldsymbol{A})}|\lambda-\mu|\le ||\boldsymbol{X}^{-1}||_p||\boldsymbol{X}||_p||\boldsymbol{E}||_p,$$

其中$||\cdot||_p$为矩阵的$p-$范数, $p=1,2,\infty$.

证明: 若$\mu\in\sigma(\boldsymbol{A})$, 结论显然成立. 今设$\mu\notin\sigma(\boldsymbol{A})$, 由$\boldsymbol{A}$ 可对角化

$$\boldsymbol{A}+\boldsymbol{E}-\mu\boldsymbol{I}=\boldsymbol{X}[\boldsymbol{D}-\mu\boldsymbol{I}+\boldsymbol{X}^{-1}\boldsymbol{E}\boldsymbol{X}]\boldsymbol{X}^{-1},$$

μ 为$\boldsymbol{A}+\boldsymbol{E}$的特征值, 所以$\boldsymbol{D}-\mu\boldsymbol{I}+\boldsymbol{X}^{-1}\boldsymbol{E}\boldsymbol{X}$ 奇异, 存在非零向量$\boldsymbol{y}\in\mathbb{C}^n$, 使

$$(\boldsymbol{D}-\mu\boldsymbol{I})\boldsymbol{y}=-(\boldsymbol{X}^{-1}\boldsymbol{E}\boldsymbol{X})\boldsymbol{y},$$

由此得

$$\boldsymbol{y}=-(\boldsymbol{D}-\mu\boldsymbol{I})^{-1}(\boldsymbol{X}^{-1}\boldsymbol{E}\boldsymbol{X})\boldsymbol{y},$$

两边取$p-$范数有

$$||(\boldsymbol{D}-\mu\boldsymbol{I})^{-1}||_p||\boldsymbol{X}^{-1}||_p||\boldsymbol{E}||_p||\boldsymbol{X}||_p\geq 1,$$

由$||(\boldsymbol{D}-\mu\boldsymbol{I})^{-1}||_p=\dfrac{1}{\min\limits_{\lambda\in\sigma(\boldsymbol{A})}|\lambda-\mu|}$, 定理得证.

$||\boldsymbol{X}^{-1}||_p||\boldsymbol{X}||_p$称为矩阵$\boldsymbol{A}$关于特征值问题的条件数.

5.2 幂迭代法与逆幂迭代法

5.2.1 幂迭代法

假设矩阵 A 可对角化，其特征值满足

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$$

且 $x_1, x_2, \cdots, x_n$ 为相应线性无关的特征向量，其中 A 按模最大的特征值 λ_1 称为主特征值，对应 x_1 称为主特征向量。

幂迭代法目的：求出主特征值及对应的主特征向量。

算法思想： 由于特征向量 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 构成 $\mathbb{R}^n$ 的一组基，因此对于 $\mathbb{R}^n$ 中的任意元素 $\boldsymbol{v}^{(0)}$，有

$$\boldsymbol{v}^{(0)} = \alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \cdots + \alpha_n \boldsymbol{x}_n.$$

从而得到

$$\boldsymbol{A}^k \boldsymbol{v}^{(0)} = \lambda_1^k \left[\alpha_1 \boldsymbol{x}_1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1} \right)^k \boldsymbol{x}_i \right].$$

若记 $\boldsymbol{v}^{(k)} = \boldsymbol{A}^k \boldsymbol{v}^{(0)}$，则有 $\boldsymbol{v}^{(k+1)} = \boldsymbol{A} \boldsymbol{v}^{(k)}$。

当 $k \to \infty$，$\left|\frac{\lambda_i}{\lambda_1}\right|^k \to 0$。且当 $\alpha_1 \neq 0$ 时，$\boldsymbol{v}^{(k)} \approx \lambda_1^k \alpha_1 \boldsymbol{x}_1$。

但注意到，

当 $|\lambda_1| > 1$ 时，$\boldsymbol{v}^{(k)} \approx \lambda_1^k \alpha_1 \boldsymbol{x}_1$ 的分量都趋向于无穷；

当 $|\lambda_1| < 1$ 时，$\boldsymbol{v}^{(k)} \approx \lambda_1^k \alpha_1 \boldsymbol{x}_1$ 的分量都趋向于 0。

为避免这种情况，通常在迭代过程中加上规范化步骤，即在每一次计算中都要求 $\boldsymbol{v}^{(k)}$ 的按模最大分量为 1，这样 $||\boldsymbol{v}^{(k)}||_\infty = 1$。

幂迭代格式： 选取初始向量 $\boldsymbol{v}^{(0)} \neq 0$，对 $k = 0, 1, 2, \cdots$，按照如下迭代格式计算

$$\begin{cases} \boldsymbol{u}^{(k+1)} = \boldsymbol{A}\boldsymbol{v}^{(k)}, \\ m_{k+1} = \max(\boldsymbol{u}^{(k+1)}), \\ \boldsymbol{v}^{(k+1)} = \dfrac{\boldsymbol{u}^{(k+1)}}{m_{k+1}}. \end{cases}$$

定理5.4： 设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$，$\boldsymbol{A}$ 存在 n 个线性无关的特征向量 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，对应的特征值满足

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|,$$

且初始向量 $\boldsymbol{v}^{(0)}$ 在 $\boldsymbol{x}_1$ 方向上的投影非零，则有

$$\lim_{k\to\infty} \boldsymbol{v}^{(k)} = \frac{\boldsymbol{x}_1}{\max\{\boldsymbol{x}_1\}}, \quad \lim_{k\to\infty} m_k = \lambda_1.$$

证明: 特征向量 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 构成 $\mathbb{R}^n$ 的一组基，因此对于 $\mathbb{R}^n$ 中的任意元素 $\boldsymbol{v}^{(0)}$，有 $\boldsymbol{v}^{(0)} = \alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \cdots + \alpha_n\boldsymbol{x}_n$，从而得到

$$\boldsymbol{A}^k\boldsymbol{v}^{(0)} = \lambda_1^k\left[\alpha_1\boldsymbol{x}_1 + \sum_{i=2}^{n}\alpha_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{x}_i\right]. \ (\lambda_1\text{为按模最大特征值})$$

当 $\boldsymbol{v}^{(0)}$ 在 $\boldsymbol{x}_1$ 方向上的投影非零，即 $\alpha_1 \neq 0$，则有

$$\boldsymbol{v}^{(k)} = \frac{\boldsymbol{A}\boldsymbol{v}^{(k-1)}}{m_k} = \frac{\boldsymbol{A}}{m_k}\frac{\boldsymbol{A}\boldsymbol{v}^{(k-2)}}{m_{k-1}} = \cdots = \frac{\boldsymbol{A}^k\boldsymbol{v}^{(0)}}{m_k\cdots m_1} = \frac{\boldsymbol{A}^k\boldsymbol{v}^{(0)}}{\max(\boldsymbol{A}^k\boldsymbol{v}^{(0)})} \to \frac{\boldsymbol{x}_1}{\max(\boldsymbol{x}_1)}$$

$$m_k = \max(\boldsymbol{A}\boldsymbol{v}^{(k-1)}) = \frac{\max(\boldsymbol{A}^k\boldsymbol{v}^{(0)})}{\max(\boldsymbol{A}^{k-1}\boldsymbol{v}^{(0)})}$$

$$= \lambda_1\left[\max\left(\alpha_1\boldsymbol{x}_1 + \sum_{i=2}^{n}\alpha_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{x}_i\right)\Big/\max\left(\alpha_1\boldsymbol{x}_1 + \sum_{i=2}^{n}\alpha_i\left(\frac{\lambda_i}{\lambda_1}\right)^{k-1}\boldsymbol{x}_i\right)\right] \to \lambda_1.$$

注1: 幂法收敛速度取决于$\left|\frac{\lambda_2}{\lambda_1}\right|$, 当$|\lambda_2| \ll |\lambda_1|$ 时, 收敛较快, 而当$|\lambda_2| \approx |\lambda_1|$时, 收敛速度很慢.

注2: 实际计算中, 可设置$||\boldsymbol{v}^{(k)} - \boldsymbol{v}^{(k+1)}|| \leq \epsilon$作为迭代终止准则.

注3: 对于不满足$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$的情形, 如按模最大特征值为$r$重根的情形($r \geq 2$), 但$\boldsymbol{A}$仍满足可对角化条件, 那么$m_k = \max(\boldsymbol{u}_k) \to \lambda_1$仍成立.

注4: 若选取的$\boldsymbol{v}^{(0)}$满足$\alpha_1 = 0$, 则在中间计算过程无任何误差的情况下, 幂法计算得到的是次模最大的特征值; 若中间过程有舍入误差, 则很可能某次迭代开始成立$\alpha_1 \neq 0$, 但α_1 仍很小, 有可能造成迭代很多次后收敛到$\boldsymbol{x}_1$, λ_1. 通常取$\boldsymbol{v}^{(0)} = (1, 1, \cdots, 1)^T$, 以尽可能使$\alpha_1 \neq 0$的可能性大些.

例: 计算$A=\begin{pmatrix} 2 & 4 & 6 \\ 3 & 9 & 15 \\ 4 & 16 & 36 \end{pmatrix}$的按模最大特征值及对应特征向量。

解: 取 $\boldsymbol{v}^{(0)}=(1,1,1)^T$，按规范化幂法计算结果见表

k	$\boldsymbol{u}^{(k)T}$			$\boldsymbol{v}^{(k)T}$			m_k
0	–	–	–	1	1	1	
1	12	27	56	0.2143	0.4820	1	56
2	8.357	19.98	55.57	0.1857	0.4483	1	55.57
3	8.168	19.60	43.92	0.1860	0.4463	1	43.92
4	8.157	19.57	43.88	0.1859	0.4460	1	43.88
5	8.156	19.57	43.88	0.1859	0.4460	1	43.88

所以按模最大特征值为 $\lambda_1=43.88$，其对应的特征向量为

$\boldsymbol{x}_1=(0.1859,0.4460,1.0000)^T$。

5.2.2 幂法的加速

原点平移法

当$\left|\frac{\lambda_2}{\lambda_1}\right|$接近1时，幂法收敛可能很慢，可采用加速收敛的技巧.

引进矩阵$\boldsymbol{B} = \boldsymbol{A} - p\boldsymbol{I}$，其中$p$为待定参数，设$\boldsymbol{A}$特征值为$\lambda_1, \lambda_2, \cdots, \lambda_n$，则$B$特征值为$\lambda_1 - p, \lambda_2 - p, \cdots, \lambda_n - p$，且$\boldsymbol{A}, \boldsymbol{B}$特征向量相同.

如果计算$\boldsymbol{A}$的主特征值，需要适当选择p，使得

$$|\lambda_1 - p| > |\lambda_2 - p| \geq \cdots \geq |\lambda_n - p|$$
$$\left|\frac{\lambda_2 - p}{\lambda_1 - p}\right| < \left|\frac{\lambda_2}{\lambda_1}\right|.$$

问题： 若已知$\boldsymbol{A}$的特征值为实数，如何选取p使幂法得到加速?

注： 设计一个自动选择适当参数p的算法是困难的.

瑞利商加速

定义： 设$\boldsymbol{A}$为n阶实对称矩阵，对于任一非零向量$\boldsymbol{x}$，称

$$R(x) = \frac{(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}$$

为对应于向量$\boldsymbol{x}$的瑞利(Rayleigh)商.

定理： 设$\boldsymbol{A} \in \boldsymbol{R}^{n\times n}$为对称矩阵，特征值为$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$，则

(1) $\lambda_n \leq \dfrac{(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})} \leq \lambda_1$,

(2) $\lambda_1 = \max\limits_{\boldsymbol{x}\in\boldsymbol{R}^n, \boldsymbol{x}\neq\boldsymbol{0}} \dfrac{(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}$,

(3) $\lambda_n = \min\limits_{\boldsymbol{x}\in\boldsymbol{R}^n, \boldsymbol{x}\neq\boldsymbol{0}} \dfrac{(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}$.

定理： 设$\boldsymbol{A} \in \boldsymbol{R}^{n\times n}$为对称矩阵，特征值满足$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$，对应的特征向量满足$(\boldsymbol{x}_i, \boldsymbol{x}_j) = \delta_{ij}$，则应用幂法计算$\boldsymbol{A}$的主特征值$\lambda_1$，规范化向量$\boldsymbol{v}^{(k)}$的瑞利商满足

$$\frac{(\boldsymbol{A}\boldsymbol{v}^{(k)}, \boldsymbol{v}^{(k)})}{(\boldsymbol{v}^{(k)}, \boldsymbol{v}^{(k)})} = \lambda_1 + O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\right).$$

5.2.3 逆幂迭代法

应用 1：计算非奇异矩阵 $\boldsymbol{A}$ 按模最小特征值及相应特征向量。

设 $\boldsymbol{A}$ 非奇异，且 $0 < |\lambda_n| < |\lambda_{n-1}| \leq \cdots |\lambda_2| \leq |\lambda_1|$，则 $\boldsymbol{A}$ 的按模最小特征值 λ_n 的倒数恰好为 $\boldsymbol{A}^{-1}$ 的按模最大特征值。对 $\boldsymbol{A}^{-1}$ 应用幂法即可求得最小特征值及其对应的特征向量。

算法：选取 $\boldsymbol{v}^{(0)} \neq 0$，对 $k=1,2,\cdots$，按照如下迭代格式计算

$$\begin{cases} \boldsymbol{u}^{(k+1)} = \boldsymbol{A}^{-1}\boldsymbol{v}^{(k)}, \\ m_{k+1} = \max(\boldsymbol{u}^{(k+1)}), \quad k=1,2,\cdots \\ \boldsymbol{v}^{(k+1)} = \dfrac{\boldsymbol{u}^{(k+1)}}{m_{k+1}}. \end{cases}$$

则有 $m_k \to \dfrac{1}{\lambda_n}$，$\boldsymbol{v}^{(k)} \to \dfrac{\boldsymbol{x}_n}{\max(\boldsymbol{x}_n)}$。

为了避免计算 $\boldsymbol{A}^{-1}$，我们往往将求 $\boldsymbol{u}^{(k+1)} = \boldsymbol{A}^{-1}\boldsymbol{v}^{(k)}$ 变为求解线性方程组 $\boldsymbol{A}\boldsymbol{u}^{(k+1)} = \boldsymbol{v}^{(k)}$($LU$分解法)。

应用 2： 给定一个常数 p，求$\boldsymbol{A}$的最接近于 p 的特征值及其特征向量。

若矩阵 $(\boldsymbol{A}-p\boldsymbol{I})^{-1}$ 存在，则其特征值为

$$\frac{1}{\lambda_1 - p}, \frac{1}{\lambda_2 - p}, \cdots, \frac{1}{\lambda_n - p}$$

对应的特征向量仍为 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$。

设 λ_i 为最接近 p 的特征值，则有

$$|\lambda_i - p| < |\lambda_j - p|, \qquad j \neq i$$

即 $(\lambda_i - p)^{-1}$ 为 $(\boldsymbol{A}-p\boldsymbol{I})^{-1}$ 的按模最大特征值。

⋆ 对$\boldsymbol{A}-p\boldsymbol{I}$ 实行逆幂迭代法计算λ_i:

$$\begin{cases} (\boldsymbol{A}-p\boldsymbol{I})\boldsymbol{u}^{(k+1)} = \boldsymbol{v}^{(k)}, \\ m_{k+1} = \max(\boldsymbol{u}^{(k+1)}), \\ \boldsymbol{v}^{(k+1)} = \dfrac{\boldsymbol{u}^{(k+1)}}{m_{k+1}}. \end{cases}$$

- 收敛性:

$$\boldsymbol{v}^{(k)} \to \frac{\boldsymbol{x}_i}{\max(\boldsymbol{x}_i)}, \quad m_k \to \frac{1}{\lambda_i - p}.$$

得$k \to \infty$ 时, $\dfrac{1}{m_k} + p \to \lambda_i$ 。

例： 用逆幂法求矩阵 $\boldsymbol{A}=\begin{pmatrix}2&1&0\\1&3&1\\0&1&4\end{pmatrix}$ 在 1 附近的特征值及对应特征向量（准确值 $\lambda=3-\sqrt{3}$）。

解： 计算可得

$$(\boldsymbol{A}-\boldsymbol{I})^{-1}=\frac{1}{2}\begin{pmatrix}5&-3&1\\-3&3&-1\\1&-1&1\end{pmatrix}.$$

取 $\boldsymbol{v}^{(0)} = (1,1,1)^T$，计算结果见表

k	$\boldsymbol{u}^{(k)T}$			$\boldsymbol{v}^{(k)T}$			$p + \frac{1}{\max(\boldsymbol{u}^{(k)})}$
0				1	1	1	
1	1.5	-0.5	0.5	1	-0.3333	0.3333	1.667
2	3.1666	-2.1666	1.6666	1	-0.6842	0.4545	1.2727
3	3.7536	-2.7536	1.0694	1	-0.7336	0.2849	1.2664
4	3.7429	-2.7429	1.0093	1	-.07328	0.2697	1.2671
5	3.7341	-2.7341	1.0013	1	-0.7322	0.2681	1.2678
6	3.7324	-2.7324	1.0001	1	-0.7321	0.2680	1.2679
7	3.7322	-2.7322	1.0001	1	-0.7321	0.2680	1.2679

从而 $\lambda_j \approx 1.2679$，特征向量 $(1, -0.7321, 0.2680)^T$。

5.3 QR 算法

结合对矩阵特征值的估计，理论上可以利用逆幂迭代法计算所有的特征值，但这种方法计算量很大。能否用方便的方法求出所有特征值？

问题1：什么样的矩阵的特征值最好求？

回答1： 对角矩阵，上（下）三角矩阵，特征值就是主对角线元素。

问题2： 如何将一个矩阵化为对角阵或上（下）三角矩阵，而且还保持特征值不变？

回答2： 利用正交相似变换，即使对于矩阵 A，找一个正交矩阵 Q，使得 $QAQ^T = B$，其中 B 是一个对角阵，或者上(下)三角矩阵，而 B 的对角线元素就是特征值。

如何构造？ **QR算法**（二十世纪十大算法占2席）

5.3.1 矩阵的 QR 分解

1. 定义(QR分解)： 设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$，如果存在 n 阶正交矩阵 $\boldsymbol{Q}$ 和 n 阶上三角矩阵 $\boldsymbol{R}$，使得

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$$

则称该分解为矩阵 $\boldsymbol{A}$ 的 $\boldsymbol{QR}$ 分解。

5.3.1 矩阵的 QR 分解

1. 定义(QR分解)： 设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$，如果存在 n 阶正交矩阵 $\boldsymbol{Q}$ 和 n 阶上三角矩阵 $\boldsymbol{R}$，使得

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$$

则称该分解为矩阵 $\boldsymbol{A}$ 的 $\boldsymbol{Q}\boldsymbol{R}$ 分解。

写成分量形式，即

$$(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) = (\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots \boldsymbol{q}_n)\begin{pmatrix} r_{11} & r_{1,2} & \cdots & r_{1,n} \\ & r_{22} & & r_{2,n} \\ & & \ddots & \vdots \\ & & & r_{n,n} \end{pmatrix},$$

由此可得，

$$
\begin{cases}
\boldsymbol{a}_1 = r_{1,1} \cdot \boldsymbol{q}_1, \\
\vdots \\
\boldsymbol{a}_k = r_{1,k} \cdot \boldsymbol{q}_1 + r_{2,k} \cdot \boldsymbol{q}_2 + \cdots + r_{k,k} \cdot \boldsymbol{q}_k, \\
\vdots \\
\boldsymbol{a}_n = r_{1,n} \cdot \boldsymbol{q}_1 + r_{2,n} \cdot \boldsymbol{q}_2 + \cdots + r_{n,n} \cdot \boldsymbol{q}_n.
\end{cases}
$$

由此可见，矩阵 $\boldsymbol{A}$ 的 $\boldsymbol{QR}$ 分解，实际上是要找到一组正交基 $\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n$ 来对 $\boldsymbol{A}$ 的列向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ 进行线性表示，其中 $\boldsymbol{R}$ 的列向量就是对应的表示系数。

2. Schmidt 正交化与 QR 分解

定理5.5： 设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 为非奇异矩阵，存在 n 阶正交矩阵 $\boldsymbol{Q}$ 和 n 阶上三角矩阵 $\boldsymbol{R}$，使得

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R},$$

且当 $\boldsymbol{R}$ 的对角元素为正时，该分解是唯一的。

证明： 由于矩阵 $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n)$ 为非奇异矩阵，则 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ 是线性无关的。利用 Schmidt 正交化和向量标准化，可得到标准正交向量组 $\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n$。

首先，根据 Gram-Schmidt 正交化过程：

$$\begin{cases} \boldsymbol{b}_1 = \boldsymbol{a}_1, \quad \boldsymbol{q}_1 = \dfrac{\boldsymbol{b}_1}{||\boldsymbol{b}_1||_2}, \\ \boldsymbol{b}_k = \boldsymbol{a}_k - \displaystyle\sum_{j=1}^{k-1} l_{j,k}\boldsymbol{q}_j, \quad l_{j,k} = (\boldsymbol{a}_k, \boldsymbol{q}_j), \\ \boldsymbol{q}_k = \dfrac{\boldsymbol{b}_k}{||\boldsymbol{b}_k||_2}, \qquad k = 2, 3, \cdots, n. \end{cases}$$

于是，有

$$\begin{cases} \boldsymbol{a}_1 = ||\boldsymbol{b}_1||_2\boldsymbol{q}_1, \quad ||\boldsymbol{b}_1||_2 = ||\boldsymbol{a}_1||_2, \\ \boldsymbol{a}_k = \displaystyle\sum_{j=1}^{k-1} l_{j,k}\boldsymbol{q}_j + ||\boldsymbol{b}_k||_2\boldsymbol{q}_k, \\ ||\boldsymbol{b}_k||_2 = \sqrt{||\boldsymbol{a}_k||_2^2 - \displaystyle\sum_{j=1}^{k-1} l_{j,k}^2}, \quad k = 2, 3, \cdots, n. \end{cases}$$

上式可用矩阵形式表示为

$$\boldsymbol{A}=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)=\underbrace{(\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{q}_n)}_{Q}\underbrace{\begin{pmatrix}||\boldsymbol{b}_1||_2 & l_{1,2} & \cdots & l_{1,n}\\ 0 & ||\boldsymbol{b}_2||_2 & \cdots & l_{2,n}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & ||\boldsymbol{b}_n||_2\end{pmatrix}}_{\mathbf{R}}=\boldsymbol{Q}\boldsymbol{R},$$

这里 $\boldsymbol{Q}$ 满足 $\boldsymbol{Q}\boldsymbol{Q}^T=\boldsymbol{I}$ 为正交矩阵，$\boldsymbol{R}$ 为对角线元素为正的上三角矩阵。由此，$\boldsymbol{Q}\boldsymbol{R}$ 分解的存在性得以证明。

下面证明 QR 分解的唯一性。

假设存在两个 QR 分解：

$$\boldsymbol{A}=\boldsymbol{Q}_1\boldsymbol{R}_1=\boldsymbol{Q}_2\boldsymbol{R}_2,$$

由于 $\boldsymbol{R}_1$ 和 $\boldsymbol{R}_2$ 的主对角线元素为正，因此 $\boldsymbol{R}_1$ 和 $\boldsymbol{R}_2$ 可逆，由此可得

$$\boldsymbol{Q}_1\boldsymbol{R}_1=\boldsymbol{Q}_2\boldsymbol{R}_2 \quad\Rightarrow\quad \boldsymbol{Q}_1=\boldsymbol{Q}_2(\boldsymbol{R}_2\boldsymbol{R}_1^{-1})=\boldsymbol{Q}_2\boldsymbol{D},$$

其中 $\boldsymbol{D}=\boldsymbol{R}_2\boldsymbol{R}_1^{-1}$ 也是一个上三角矩阵。

由于

$$\boldsymbol{D}^T\boldsymbol{D}=\boldsymbol{D}^T\boldsymbol{Q}_2^T\boldsymbol{Q}_2\boldsymbol{D}=(\boldsymbol{Q}_2\boldsymbol{D})^T(\boldsymbol{Q}_2\boldsymbol{D})=\boldsymbol{Q}_1{}^T\boldsymbol{Q}_1=\boldsymbol{I}.$$

因此，D 是一个正交矩阵，又由于 D 是一个上三角矩阵，因此 D 是单位矩阵，从而得到

$$\boldsymbol{Q}_1 = \boldsymbol{Q}_2\boldsymbol{D} = \boldsymbol{Q}_2, \boldsymbol{R}_2 = \boldsymbol{D}\boldsymbol{R}_1 = \boldsymbol{R}_1,$$

因此，矩阵 $\boldsymbol{A}$ 的 $\boldsymbol{QR}$ 分解是唯一的。 证毕■

定理的证明不仅证明了 $\boldsymbol{QR}$ 分解的存在性和唯一性，同时也给出了基于 Schmidt 正交化的 $\boldsymbol{QR}$ 分解实现的方法。

注：由于舍入误差的影响，Gram-Schmidt正交化方法数值不稳定.

改进: 修正的Gram-Schmidt 正交化方法（比GS 有改善, 但效果也不太好）.

修正的 Gram-Schmidt 正交化过程：

$$
\begin{cases}
\boldsymbol{b}_1 = \boldsymbol{a}_1, \ \cdots, \ \boldsymbol{b}_n = \boldsymbol{a}_n, \\
\boldsymbol{q}_1 = \dfrac{\boldsymbol{b}_1}{||\boldsymbol{b}_1||_2}, \ \boldsymbol{b}_k = \boldsymbol{b}_k - l_{1,k}\boldsymbol{q}_1, \quad l_{1,k} = (\boldsymbol{b}_k, \boldsymbol{q}_1), \ k = 2, \cdots, n, \\
\boldsymbol{q}_2 = \dfrac{\boldsymbol{b}_2}{||\boldsymbol{b}_2||_2}, \ \boldsymbol{b}_k = \boldsymbol{b}_k - l_{2,k}\boldsymbol{q}_2, \quad l_{2,k} = (\boldsymbol{b}_k, \boldsymbol{q}_2), \ k = 3, \cdots, n, \\
\boldsymbol{q}_3 = \dfrac{\boldsymbol{b}_3}{||\boldsymbol{b}_2||_2}, \ \cdots\cdots \\
\cdots\cdots \\
\boldsymbol{q}_n = \dfrac{\boldsymbol{b}_n}{||\boldsymbol{b}_n||_2}.
\end{cases}
$$

例： 利用 修正的Gram-Schmidt 正交化求矩阵 $\boldsymbol{A}$ 的 $\boldsymbol{QR}$ 分解，其中

$$\boldsymbol{A}=\begin{pmatrix} 0 & 3 & 1 \\ 0 & 4 & -2 \\ 2 & 1 & 1 \end{pmatrix}$$

解： 矩阵 $\boldsymbol{A}$ 的列向量

$$\boldsymbol{a}_1=(0,0,2)^T,\quad \boldsymbol{a}_2=(3,4,1)^T,\quad \boldsymbol{a}_3=(1,-2,1)^T,$$

修正的Gram-Schmidt 正交化, 首先令

$$\boldsymbol{b}_1=\boldsymbol{a}_1,\quad \boldsymbol{b}_2=\boldsymbol{a}_2,,\quad \boldsymbol{b}_3=\boldsymbol{a}_3,$$

第一步，取向量$\boldsymbol{q}_1 = \frac{\boldsymbol{b}_1}{||\boldsymbol{b}_1||_2} = (0, 0, 1)^T$,

$$\boldsymbol{b}_2 = \boldsymbol{b}_2 - (\boldsymbol{b}_2, \boldsymbol{q}_1)\boldsymbol{q}_1 = (3, 4, 0)^T,$$

$$\boldsymbol{b}_3 = \boldsymbol{b}_3 - (\boldsymbol{b}_3, \boldsymbol{q}_1)\boldsymbol{q}_1 = (1, -2, 0)^T,$$

第二步，取向量$\boldsymbol{q}_2 = \frac{\boldsymbol{b}_2}{||\boldsymbol{b}_2||_2} = (3/5, 4/5, 0)^T$,

$$\boldsymbol{b}_3 = \boldsymbol{b}_3 - (\boldsymbol{b}_3, \boldsymbol{q}_2)\boldsymbol{q}_2 = (8/5, -6/5, 0)^T, \ \boldsymbol{q}_3 = \frac{\boldsymbol{b}_3}{||\boldsymbol{b}_3||_2} = (4/5, -3/5, 0)^T,$$

由此得到 $\boldsymbol{A}$ 的 $\boldsymbol{QR}$ 分解为：

$$\boldsymbol{Q} = \begin{pmatrix} 0 & 3/5 & 4/5 \\ 0 & 4/5 & -3/5 \\ 1 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{R} = \boldsymbol{Q}^T\boldsymbol{A} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 5 & -1 \\ 0 & 0 & 2 \end{pmatrix}.$$

3. Householder 变换与 QR 分解

定义： 设向量 $\boldsymbol{w} \in \mathbb{R}^n$，且$||\boldsymbol{w}||_2 = 1$，称矩阵

$$\boldsymbol{H}(\boldsymbol{w}) = \boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T$$

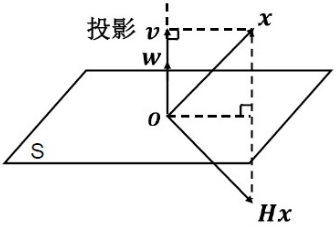
为 Householder 矩阵，或者 Householder 变换。Householder 变换也称为镜像变换。对任意向量 $\boldsymbol{x}$，$\boldsymbol{Hx}$ 是 $\boldsymbol{x}$ 在以 $\boldsymbol{w}$ 为法线"镜面"S中的像。

记$S = \{\boldsymbol{u} | \boldsymbol{\omega}^T \boldsymbol{u} = 0,\ \forall \boldsymbol{u} \in R^n\}$,

则$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2,\ \boldsymbol{x}_1 \in S,\ \boldsymbol{x}_2 = \alpha\boldsymbol{\omega} \in \text{span}\{\boldsymbol{w}\}$.

$\boldsymbol{Hx} = (\boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T)\boldsymbol{x} = \boldsymbol{x}_1 - \boldsymbol{x}_2$,

其中 $\alpha = (\boldsymbol{x}^T\boldsymbol{w})$。



定理5.6： Householder 矩阵 $\boldsymbol{H}$ 满足下面的性质：

(1) $\boldsymbol{H}$ 为对称正交矩阵，即 $\boldsymbol{H}^T=\boldsymbol{H}$，$\boldsymbol{H}^{-1}=\boldsymbol{H}$。

(2) $\begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{H} \end{pmatrix}$ 也是 Householder 矩阵。

(3) 对任意两个长度相等的非零向量（即 $||\boldsymbol{x}||_2=||\boldsymbol{y}||_2$），取 $\boldsymbol{w}=\frac{\boldsymbol{x}-\boldsymbol{y}}{||\boldsymbol{x}-\boldsymbol{y}||_2}$，则 $\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{w}\boldsymbol{w}^T$ 为一个 Householder 变换，使得 $\boldsymbol{H}\boldsymbol{x}=\boldsymbol{y}$。

(4) 设 $\boldsymbol{u}$ 是一个单位向量，则对于任意 $\boldsymbol{x}\in\mathbb{C}^n$，存在 Householder 矩阵 $\boldsymbol{H}$，使得 $\boldsymbol{H}\boldsymbol{x}=a\boldsymbol{u}$，其中 $|a|=||\boldsymbol{x}||_2$ 且 $a\boldsymbol{x}^H\boldsymbol{u}$ 为实数。

（取 $\boldsymbol{w}=\frac{\boldsymbol{x}-a\boldsymbol{u}}{||\boldsymbol{x}-a\boldsymbol{u}||_2}$ 即可）

对于任意非零向量 $\boldsymbol{x} \in \mathbb{R}^n$，求 Householder 变换 $\boldsymbol{H}$ ，使得

$$\boldsymbol{H}\boldsymbol{x} = k\boldsymbol{e}_1, \ \boldsymbol{e}_1 = (1, 0, 0, \cdots, 0)^T.$$

根据性质(4)，构造$\boldsymbol{w} = \dfrac{\boldsymbol{u}}{||\boldsymbol{u}||_2}, \boldsymbol{u} = \boldsymbol{x} - k\boldsymbol{e}_1$, 其中$k = -\mathrm{sgn}(x_1)||\boldsymbol{x}||_2$,

$$\mathrm{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ -1 & x < 0 \end{cases}.$$

$$\boldsymbol{u} = \boldsymbol{x} - (-\mathrm{sgn}(x_1)||\boldsymbol{x}||_2)\boldsymbol{e}_1 = (x_1 + \mathrm{sgn}(x_1)||\boldsymbol{x}||_2, x_2, \cdots, x_n)^T,$$

则有$||\boldsymbol{u}||_2^2 = 2||\boldsymbol{x}||_2(||\boldsymbol{x}||_2 + |x_1|)$ 。

从而，得到 Householder 变换为

$$\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T = \boldsymbol{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^T}{||\boldsymbol{u}||_2^2} = \boldsymbol{I} - \beta\boldsymbol{u}\boldsymbol{u}^T, \quad \beta = [||\boldsymbol{x}||_2(||\boldsymbol{x}||_2 + |x_1|)]^{-1}.$$

基于 Householder 变换的 QR 分解

基本思想：利用 Householder 变换，对矩阵 $\boldsymbol{A}$ 的每一列进行逐列消元，变为一个上三角矩阵。

对 $\boldsymbol{A}=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)$，记 $\boldsymbol{A}^{(1)}=\boldsymbol{A}$,

(1) 对于 $\boldsymbol{a}_1^{(1)}=(a_{11}^{(1)},a_{21}^{(1)},\cdots,a_{n1}^{(1)})^T$,

构造

$$\begin{cases} r_{11}=-\mathrm{sgn}(a_{11}^{(1)})||\boldsymbol{a}_1^{(1)}||_2, \\ \boldsymbol{u}_1=\boldsymbol{a}_1^{(1)}-r_{11}\boldsymbol{e}_1=(a_{11}^{(1)}+\mathrm{sgn}(a_{11}^{(1)})||\boldsymbol{a}_1^{(1)}||_2,a_{21}^{(1)},\cdots,a_{n1}^{(1)})^T, \\ \beta_1=\left[||\boldsymbol{a}_1^{(1)}||_2(||\boldsymbol{a}_1^{(1)}||_2+|a_{11}^{(1)}|)\right]^{-1}, \\ \boldsymbol{H}_1=\boldsymbol{I}-\beta_1\boldsymbol{u}_1\boldsymbol{u}_1^T. \end{cases}$$

则

$$\boldsymbol{H}_1\boldsymbol{A}^{(1)} = \begin{pmatrix} r_{11} & * & \cdots & * \\ 0 & a_{22}^{(2)} & & a_{2n}^{(2)} \\ 0 & & & \\ 0 & a_{n2}^{(2)} & & a_{nn}^{(2)} \end{pmatrix} = \begin{pmatrix} r_{11} & * \\ \boldsymbol{0} & \boldsymbol{A}^{(2)} \end{pmatrix},$$

其中 $\boldsymbol{A}^{(2)} = (\boldsymbol{a}_2^{(2)}, \boldsymbol{a}_3^{(2)}, \cdots, \boldsymbol{a}_n^{(2)})$，$\boldsymbol{a}_k^{(2)} = (a_{2k}^{(2)}, a_{3k}^{(2)}, \cdots, a_{nk}^{(2)})^T$，$k = 2, 3, \cdots, n$。

然后，下面再考虑对 $\boldsymbol{A}^{(2)}$ 的第一列向量进行消元。

假设经过 $k-1$ 次消元后得到：

$$\boldsymbol{H}_{k-1}\cdots\boldsymbol{H}_1\boldsymbol{A}=\begin{pmatrix} r_{1,1} & * & \cdots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & r_{k-1,k-1} & * \\ 0 & \cdots & 0 & \boldsymbol{A}^{(k)} \end{pmatrix},$$

$$\boldsymbol{A}^{(k)}=\left(\boldsymbol{a}_k^{(k)},\boldsymbol{a}_{k+1}^{(k)},\cdots,\boldsymbol{a}_n^{(k)}\right),$$

$$\boldsymbol{a}_j^{(k)}=\left(a_{k,j}^{(k)},a_{k+1,j}^{(k)},\cdots,a_{n,j}^{(k)}\right)^T.$$

(2) 对于 $\boldsymbol{a}_k^{(k)} = (a_{k,k}^{(k)}, a_{k+1,k}^{(k)}, \cdots, a_{n,k}^{(k)})^T$，构造

$$\begin{cases} r_{k,k} = -\operatorname{sgn}(a_{k,k}^{(k)})\|\boldsymbol{a}_k^{(k)}\|_2, \\ \boldsymbol{u}_k = \boldsymbol{a}_k^{(k)} - r_{k,k}\boldsymbol{e}_1^{(k)} = (a_{k,k}^{(k)} + \operatorname{sgn}(a_{k,k}^{(k)})\|\boldsymbol{a}_k^{(k)}\|_2, a_{k+1,k}^{(1)}, \cdots, a_{n,k}^{(1)})^T, \\ \beta_k = \left[\|\boldsymbol{a}_k^{(k)}\|_2(\|\boldsymbol{a}_k^{(k)}\|_2 + |a_{k,k}^{(k)}|)\right]^{-1}, \\ \boldsymbol{H}_k = \begin{pmatrix} \boldsymbol{I}_{(k-1)\times(k-1)} & 0 \\ 0 & \boldsymbol{I}_{(n-k+1)\times(n-k+1)} - \beta_k \boldsymbol{u}_k \boldsymbol{u}_k^T \end{pmatrix}. \end{cases}$$

其中 $k = 2, 3, \cdots, n$.

根据上述方法，可以构造出 $n-1$ 个 Householder 矩阵 $\boldsymbol{H}_1, \boldsymbol{H}_2, \cdots, \boldsymbol{H}_{n-1}$，使得

$$\boldsymbol{H}_{n-1}\cdots\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A} = \boldsymbol{R} = \begin{pmatrix} r_{1,1} & * & \cdots & * \\ & r_{2,2} & \ddots & \vdots \\ & & \ddots & * \\ & & & r_{n,n} \end{pmatrix}.$$

由于 $\boldsymbol{H}_1, \boldsymbol{H}_2, \cdots, \boldsymbol{H}_{n-1}$ 都是正交矩阵，因此 $\boldsymbol{Q} = (\boldsymbol{H}_{n-1}\cdots\boldsymbol{H}_1)^T$ 也是正交矩阵，从而得到矩阵 $\boldsymbol{A}$ 的 $\boldsymbol{QR}$ 分解为

$$\boldsymbol{A} = \boldsymbol{QR}.$$

例： 利用 Householder 变换求矩阵

$$\boldsymbol{A}=\begin{pmatrix}0 & 3 & 1\\ 0 & 4 & -2\\ 2 & 1 & 1\end{pmatrix}$$

的 $\boldsymbol{QR}$ 分解。

解： (1) 对 $\boldsymbol{a}_1^{(1)}=(0,0,2)^T$，$||\boldsymbol{a}_1^{(1)}||_2=2$，则

$$r_{11}=-2,\quad \boldsymbol{u}_1=(2,0,2)^T,\quad \beta_1=\frac{1}{4}$$

$$\boldsymbol{H}_1=\boldsymbol{I}-\frac{1}{4}\begin{pmatrix}2\\ 0\\ 2\end{pmatrix}(2,0,2)=\begin{pmatrix}0 & 0 & -1\\ 0 & 1 & 0\\ -1 & 0 & 0\end{pmatrix},$$

从而得到：

$$
\boldsymbol{H}_1\boldsymbol{A}=\begin{pmatrix}0&0&-1\\0&1&0\\-1&0&0\end{pmatrix}\begin{pmatrix}0&3&1\\0&4&-2\\2&1&1\end{pmatrix}=\begin{pmatrix}-2&-1&-1\\0&4&-2\\0&-3&-1\end{pmatrix}.
$$

(2) 对 $\boldsymbol{a}_2^{(2)}=(4,-3)^T$，$||\boldsymbol{a}_2^{(2)}||_2=5$，则 $r_{22}=-5$, $\boldsymbol{u}_2=(9,-3)^T$, $\beta_2=\frac{1}{45}$

$$
\tilde{\boldsymbol{H}}_2=\boldsymbol{I}-\frac{1}{45}\begin{pmatrix}9\\-3\end{pmatrix}\begin{pmatrix}9&-3\end{pmatrix}=\frac{1}{5}\begin{pmatrix}-4&3\\3&4\end{pmatrix},
$$

$$
\boldsymbol{H}_2=\begin{pmatrix}\boldsymbol{I}&\\&\tilde{\boldsymbol{H}}_2\end{pmatrix}=\frac{1}{5}\begin{pmatrix}5&0&0\\0&-4&3\\0&3&4\end{pmatrix},
$$

从而得到

$$\tilde{\boldsymbol{H}}_2\boldsymbol{A}^{(2)} = \frac{1}{5}\begin{pmatrix} -4 & 3 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 4 & -2 \\ -3 & -1 \end{pmatrix} = \begin{pmatrix} -5 & 1 \\ 0 & -2 \end{pmatrix},$$

$$\boldsymbol{H}_2(\boldsymbol{H}_1\boldsymbol{A}) = \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \tilde{\boldsymbol{H}}_2 \end{pmatrix}\left(\begin{array}{c|cc} -2 & -1 & -1 \\ \hline 0 & & \\ 0 & & \boldsymbol{A}^{(2)} \end{array}\right) = \left(\begin{array}{c|cc} -2 & -1 & -1 \\ \hline 0 & & \\ 0 & & \tilde{\boldsymbol{H}}_2\boldsymbol{A}^{(2)} \end{array}\right)$$

$$= \begin{pmatrix} -2 & -1 & -1 \\ & -5 & 1 \\ & & {\color{red}-2} \end{pmatrix}.$$

$\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A}$已经是一个上三角阵，若要得到对角元全为正的上三角阵，令$\boldsymbol{D} = \mathrm{diag}(-1,-1,-1)$, 则

$$\boldsymbol{R} = \boldsymbol{D}\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A} = \begin{pmatrix} 2 & -1 & -1 \\ & 5 & -1 \\ & & 2 \end{pmatrix},$$

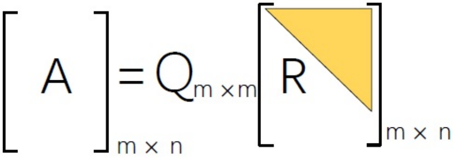
于是

$$\boldsymbol{Q} = \boldsymbol{H}_1\boldsymbol{H}_2\boldsymbol{D}^{-1} = \begin{pmatrix} 0 & 0.6 & 0.8 \\ 0 & 0.8 & -0.6 \\ 1 & 0 & 0 \end{pmatrix}.$$

关于 Householder 变换几点说明:

注1 对矩阵进行$\boldsymbol{QR}$分解时， Householder 变换与 Schmidt 正交化方法计算量相当$O(n^3)$, 但Householder 变换数值稳定性良好;

注2 由于 Householder 变换是按列逐列操作的，因此实际上可以处理不是方阵的情况：如对于 $\boldsymbol{A} \in \mathbb{R}^{m\times n}, m > n$，也存在正交矩阵$\boldsymbol{Q}$，使得



从基的表示角度看，$\boldsymbol{A} = \boldsymbol{QR}$ 实际上是找到一组正交基（即 $\boldsymbol{Q}$ 的列向量）来对 $\boldsymbol{A}$ 中的列向量进行表示，表示系数则是对应的 $\boldsymbol{R}$ 中的列。

4. Givens 变换与 QR 分解

我们知道，在平面 $\mathbb{R}^2$ 上，将一个向量 $\boldsymbol{x}$ 绕原点顺时针旋转 θ 角度得到向量 $\boldsymbol{y}$，该旋转变换可表示为

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \boldsymbol{T}\boldsymbol{x} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

这里，$\boldsymbol{T} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ 被称为旋转变换。

将二维空间中的旋转变换推广到 n 维空间，就得到 Givens 变换。

定义(Givens变换)： 称下面的 n 阶正交矩阵为 Givens 矩阵（变换）

$$\boldsymbol{J}(i,k,\theta)=\begin{pmatrix}
1 &&&&&&&&&& \\
& \ddots &&&&&&&&& \\
&& 1 &&&&&&&& \\
&&& \cos\theta &&&& \sin\theta &&& \\
&&&& 1 &&&&&& \\
&&&&& \ddots &&&&& \\
&&&&&& 1 &&&& \\
&&& -\sin\theta &&&& \cos\theta &&& \\
&&&&&&&& 1 && \\
&&&&&&&&& \ddots & \\
&&&&&&&&&& 1
\end{pmatrix}
\begin{matrix} \\ \\ \\ \text{第i 行} \\ \\ \\ \\ \text{第k 行} \\ \\ \\ \\ \end{matrix}$$

对于任意向量 $\boldsymbol{x} \in \mathbb{R}^n$，对其进行 Givens 变换 $\boldsymbol{y} = \boldsymbol{J}(i,k,\theta)\boldsymbol{x}$ ，可得

$$\begin{cases} y_i = x_i \cos\theta + x_k \sin\theta, \\ y_k = x_k \cos\theta - x_i \sin\theta, \\ y_j = x_j, & j \neq i, j \neq k \end{cases}$$

若要使得 $y_k = 0$，只需要取 $\theta = \theta_{i,k}$，其中

$$\cos\theta_{i,k} = \frac{x_i}{\sqrt{x_i^2 + x_k^2}}, \quad \sin\theta_{i,k} = \frac{x_k}{\sqrt{x_i^2 + x_k^2}},$$

此时 $y_k = 0, y_i = \sqrt{x_i^2 + x_k^2}$，而其他的分量都不变。

说明： Givens 变换可以利用向量中的某一个分量，把另外一个分量变为0，而保持其他分量都不变，这意味着可以用正交变换逐分量消元！

利用 Givens 变换进行 QR 分解

类似于高斯消去法，在第 k 步，利用 $\boldsymbol{A}^{(k)}$ 的当前第 k 列的对角线元素 $a_{k,k}^{(k)}$，利用相应的 Givens 变换 $\boldsymbol{J}^{(k)}(k,j,\theta_{k,j})$ 将当前的 $a_{j,k}^{(k)}, j>k$ 依次变为零。需要注意的是，$\boldsymbol{J}^{(k)}(k,j,\theta_{k,j})\boldsymbol{A}^{(k)}$ 在把 $a_{j,k}^{(k)}$ 变为 0 的同时，还会改变第 j 行（$j>k$）的其他元素。

第1列消元：

$$\underbrace{\boldsymbol{J}^{(1)}(1,n,\theta_{1,n})\cdots\boldsymbol{J}^{(1)}(1,3,\theta_{1,3})\boldsymbol{J}^{(1)}(1,2,\theta_{1,2})}_{\boldsymbol{P}_1}\boldsymbol{A}^{(1)}=\boldsymbol{P}_1\boldsymbol{A},$$

第2列消元：

$$\underbrace{\boldsymbol{J}^{(2)}(2,n,\theta_{2,n})\cdots\boldsymbol{J}^{(2)}(2,4,\theta_{2,4})\boldsymbol{J}^{(2)}(2,3,\theta_{2,3})}_{\boldsymbol{P}_2}(\boldsymbol{P}_1\boldsymbol{A})=\boldsymbol{P}_2\boldsymbol{P}_1\boldsymbol{A},$$

第k列消元：

$$\underbrace{\boldsymbol{J}^{(k)}(k,n,\theta_{k,n})\cdots\boldsymbol{J}^{(k)}(k,k+2,\theta_{k,k+2})\boldsymbol{J}^{(k)}(k,k+1,\theta_{k,k+1})}_{\boldsymbol{P}_k}(\boldsymbol{P}_{k-1}\cdots\boldsymbol{P}_1\boldsymbol{A})$$

$$=\boldsymbol{P}_k\cdots\boldsymbol{P}_1\boldsymbol{A},$$

由此得到

$$(\boldsymbol{P}_{n-1}\cdots\boldsymbol{P}_1)\boldsymbol{A}=\boldsymbol{R},$$

其中 $\boldsymbol{R}$ 是一个上三角矩阵。

由于所有的 Givens 变换都是正交的，因此所有的 $\boldsymbol{P}_k$ 也都是正交矩阵，记 $\boldsymbol{Q}=\boldsymbol{P}_1^T\boldsymbol{P}_2^T\cdots\boldsymbol{P}_{n-1}^T$，则 $\boldsymbol{Q}$ 是一个正交矩阵，使得

$$\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{R}.$$

例： 利用 Givens 变换求矩阵

$$\boldsymbol{A}=\begin{pmatrix}4 & 4 & 0\\ 3 & 3 & -1\\ 0 & 1 & 1\end{pmatrix}$$

的 $\boldsymbol{QR}$ 分解, 使$\boldsymbol{R}$的对角元素为正数。

Givens 变换比较适合于稀疏矩阵的分解，而且有利于并行计算，实际计算中最常采用的还是 Householder 变换，Householder 形式化的矩阵分解方法也被认为是二十世纪十大算法之一。

对于线性方程组 $\boldsymbol{Ax} = \boldsymbol{b}$，如果得到系数矩阵 $\boldsymbol{A}$ 的 $\boldsymbol{QR}$ 分解，则可以将方程组求解转化为

$$\boldsymbol{Rx} = \boldsymbol{Q}^T\boldsymbol{b},$$

然后采用“回代”过程就可以得到方程组的解。

另外，$\boldsymbol{QR}$ 分解在计算特征值的问题中扮演着重要角色，能稳定计算矩阵特征值的 $\boldsymbol{QR}$ 算法也是二十世纪十大算法之一。

5.3.2 特征值求解的 QR 算法

1. 基本QR算法

假设矩阵 $\boldsymbol{A}$ 有 $\boldsymbol{QR}$ 分解

$$\boldsymbol{A} = \boldsymbol{QR},$$

其中 $\boldsymbol{Q}$ 为正交矩阵，$\boldsymbol{R}$ 为上三角矩阵。若记 $\boldsymbol{B} = \boldsymbol{RQ}$，则可得到

$$\boldsymbol{B} = \boldsymbol{Q}^T\boldsymbol{AQ},$$

即 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 是相似矩阵，其特征值相同。继续对 $\boldsymbol{B}$ 进行 $\boldsymbol{QR}$ 分解，重复以上步骤，则可得到**基本 QR 算法：**

给定初值 $\boldsymbol{A}_1 = \boldsymbol{A}$，对 $k = 1, 2, \cdots$

$$\begin{cases} \boldsymbol{A}_k = \boldsymbol{Q}_k\boldsymbol{R}_k, \\ \boldsymbol{A}_{k+1} = \boldsymbol{R}_k\boldsymbol{Q}_k. \end{cases}$$

定理5.7： 由基本 QR 算法产生的迭代序列 $\{A_k\}$满足：

(1) $A_{k+1} = Q_k^T A_k Q_k$ ；

(2) $A_{k+1} = \tilde{Q}_k^T A \tilde{Q}_k$；

(3) $A^k = \tilde{Q}_k \tilde{R}_k$;

其中 $\tilde{Q}_k = Q_1 Q_2 \cdots Q_k, \tilde{R}_k = R_k R_{k-1} \cdots R_1$。

定义： 矩阵序列 $\{A_k\}$，当 $k \to \infty$ 时，对角线元素均收敛，严格下三角部分收敛到零，则称 $\{A_k\}$ 基本收敛到上三角矩阵。（严格上三角部分不一定收敛）

定理5.8(QR方法的收敛性)： 已知 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$，设 $\boldsymbol{A}$ 的 n 个特征值满足

$$|\lambda_1|>|\lambda_2|>\cdots>|\lambda_n|>0,$$

方阵 $\boldsymbol{X}=[x_1,x_2,\cdots,x_n]$ 由 λ_i 对应的特征向量 x_i 构成（$i=1,2,\cdots,n$）。设 $\boldsymbol{X}^{-1}$ 可分解为 $\boldsymbol{X}^{-1}=\boldsymbol{LU}$，其中 $\boldsymbol{L}$ 为单位下三角矩阵，$\boldsymbol{U}$ 为上三角矩阵，则基本 $\boldsymbol{QR}$ 算法产生的序列 $\{\boldsymbol{A}_k\}$ 基本收敛到上三角矩阵，其对角线元素极限为

$$\lim_{k\to\infty}a_{i,i}^{(k)}=\lambda_i,\quad i=1,2,\cdots,n$$

定理5.9： 若 $\boldsymbol{A}$ 为对称矩阵且满足上述定理条件，则基本 $\boldsymbol{QR}$ 算法产生的迭代序列 $\{\boldsymbol{A}_k\}$ 收敛于 $\boldsymbol{D}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$.

如果矩阵 $\boldsymbol{A}$ 具有重实根或者多重复特征值, 则由 $\boldsymbol{QR}$ 算法产生的迭代序列 $\{\boldsymbol{A}_k\}$ 基本收敛于分块上三角矩阵(对角块为一阶和二阶子块), 且对角块中的每个二阶子块给出 $\boldsymbol{A}$ 的一对共轭复特征值, 每一个一阶子块给出 $\boldsymbol{A}$ 的实特征值:

$$\boldsymbol{A}_k \to \begin{pmatrix} \lambda_1 & \cdots & * & * & \cdots & * \\ & \ddots & \vdots & \vdots & & \vdots \\ & & \lambda_m & * & & * \\ & & & \boldsymbol{B}_1 & \cdots & * \\ & & & & \ddots & \vdots \\ & & & & & \boldsymbol{B}_l \end{pmatrix}, \qquad m+2l=n,$$

其中 $\boldsymbol{B}_i = \begin{pmatrix} b_{11}^{(i)} & b_{12}^{(i)} \\ b_{21}^{(i)} & b_{22}^{(i)} \end{pmatrix}$ 的特征值为 $\boldsymbol{A}$ 的一对共轭复特征值。

2. Hessenberg矩阵的QR算法

定义: 若$\boldsymbol{B} = [b_{ij}] \in \mathbb{R}^{n\times n}$, 当$i > j+1$时有$b_{ij} = 0$, 称$\boldsymbol{B}$为上Hessenberg矩阵, 如果还有$b_{i,i-1} \neq 0$, $(i = 2, \cdots, n)$, 则称$\boldsymbol{B}$为不可约上Hessenberg阵.

上Hessenberg阵的QR算法:

对上Hessenberg矩阵 $\boldsymbol{H}_1 = \boldsymbol{H}$，对 $k = 1, 2, \cdots$

$$\begin{cases} \boldsymbol{H}_k = \boldsymbol{Q}_k \boldsymbol{R}_k, \\ \boldsymbol{H}_{k+1} = \boldsymbol{R}_k \boldsymbol{Q}_k. \end{cases}$$

注1: $\{\boldsymbol{H}_k\}$均为上Hessenberg阵;

注2： 每步计算量可降低为$O(n^2)$。

$$\begin{pmatrix} * & * & \cdots & * & * \\ * & * & \cdots & * & * \\ & * & \cdots & * & * \\ & & \ddots & \vdots & \vdots \\ & & & * & * \end{pmatrix} \to \begin{pmatrix} \times & * & \cdots & * & * \\ 0 & * & \cdots & * & * \\ & * & \cdots & * & * \\ & & \ddots & \vdots & \vdots \\ & & & * & * \end{pmatrix} \to \begin{pmatrix} \times & * & \cdots & * & * \\ 0 & \times & \cdots & * & * \\ & 0 & \cdots & * & * \\ & & \ddots & \vdots & \vdots \\ & & & * & * \end{pmatrix}$$

定理5.12: 对任意矩阵$A \in \mathbb{R}^{n\times n}$, 存在正交矩阵$Q$, 使

$$\boldsymbol{B} = \boldsymbol{Q}^T \boldsymbol{A}\boldsymbol{Q}$$

为一个上Hessenberg阵.(可通过Householder相似变换实现.)

$$(\boldsymbol{H}_1 \cdots \boldsymbol{H}_{n-3}\boldsymbol{H}_{n-2})^T \boldsymbol{A}(\boldsymbol{H}_1\boldsymbol{H}_2 \cdots \boldsymbol{H}_{n-2}) \to \boldsymbol{B} = \begin{pmatrix} {\color{red}*} & * & \cdots & * & * \\ {\color{blue}*} & {\color{red}*} & \cdots & * & * \\ & {\color{blue}*} & \cdots & * & * \\ & & \ddots & \vdots & \vdots \\ & & & {\color{blue}*} & {\color{red}*} \end{pmatrix}.$$

一般矩阵的 QR 算法:

(1) 利用 Houeseholder 变换，将 A 作相似转化为上 Hessenberg 矩阵 B:

(2) 对上 Hessenberg 矩阵 B 执行 QR 算法，其中 QR 分解可用 Givens 变换或者 Householder 变换实现（此时，每一列只需要消除一个元素即可）。

一般矩阵的 QR 算法：

(1) 利用 Houeseholder 变换，将 A 作相似转化为上 Hessenberg 矩阵 B:

(2) 对上 Hessenberg 矩阵 B 执行 QR 算法，其中 QR 分解可用 Givens 变换或者 Householder 变换实现（此时，每一列只需要消除一个元素即可）。

QR 算法的计算量? 如何提高收敛速度?

设 $\boldsymbol{A}$ 的 n 个特征值满足

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0,$$

且 $\boldsymbol{H}$ 为 $\boldsymbol{A}$ 经正交相似变换转化成的上Hessenberg阵，对 $\boldsymbol{H}$ 的 $\boldsymbol{QR}$ 算法，收敛快慢取决于$\max\left|\frac{\lambda_{j+1}}{\lambda_j}\right|$, 若能找到$\mu$, 使得$\left|\frac{\lambda_{j+1}-\mu}{\lambda_j-\mu}\right| < \left|\frac{\lambda_{j+1}}{\lambda_j}\right|$, 则可以提高收敛速度.

设 $\boldsymbol{A}$ 的 n 个特征值满足

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0,$$

且 $\boldsymbol{H}$ 为 $\boldsymbol{A}$ 经正交相似变换转化成的上Hessenberg阵，对 $\boldsymbol{H}$ 的 $\boldsymbol{QR}$ 算法，收敛快慢取决于$\max\left|\frac{\lambda_{j+1}}{\lambda_j}\right|$，若能找到$\mu$，使得$\left|\frac{\lambda_{j+1}-\mu}{\lambda_j-\mu}\right| < \left|\frac{\lambda_{j+1}}{\lambda_j}\right|$，则可以提高收敛速度.

带原点位移的$\boldsymbol{QR}$ 算法：

设上Hessenberg矩阵 $\boldsymbol{H}_1 = \boldsymbol{H}$，对 $k = 1, 2, \cdots$

$$\begin{cases} \boldsymbol{H}_k - \mu \boldsymbol{I} = \boldsymbol{Q}_k \boldsymbol{R}_k, \\ \boldsymbol{H}_{k+1} = \boldsymbol{R}_k \boldsymbol{Q}_k + \mu \boldsymbol{I}. \end{cases}$$

如果$\boldsymbol{A}$的特征值排序为：$|\lambda_1 - \mu| \geq |\lambda_2 - \mu| \geq \cdots \geq |\lambda_n - \mu|$，

则$\boldsymbol{H}_k$的第j个次对角元素的收敛快慢取决于$\left|\frac{\lambda_{j+1}-\mu}{\lambda_j-\mu}\right|^k$.

定理: 设$\boldsymbol{H} \in R^{n\times n}$为不可约上Hessenberg阵, μ为$\boldsymbol{H} = \boldsymbol{H}_1$ 的一个特征值, 则对如下方法得到的$\boldsymbol{H}_2$满足: $h_{nn}^{(2)} = \mu,\ h_{n,n-1}^{(2)} = 0$.

$$\begin{cases} \boldsymbol{H}_1 - \mu \boldsymbol{I} = \boldsymbol{Q}\boldsymbol{R}, \\ \boldsymbol{H}_2 = \boldsymbol{R}\boldsymbol{Q} + \mu \boldsymbol{I}. \end{cases}$$

定理: 设$\boldsymbol{H} \in R^{n\times n}$为不可约上Hessenberg阵, μ为$\boldsymbol{H} = \boldsymbol{H}_1$ 的一个特征值, 则对如下方法得到的$\boldsymbol{H}_2$满足: $h_{nn}^{(2)} = \mu,\ h_{n,n-1}^{(2)} = 0.$

$$\begin{cases} \boldsymbol{H}_1 - \mu\boldsymbol{I} = \boldsymbol{Q}\boldsymbol{R}, \\ \boldsymbol{H}_2 = \boldsymbol{R}\boldsymbol{Q} + \mu\boldsymbol{I}. \end{cases}$$

带原点位移的QR 算法:

设上Hessenberg矩阵 $\boldsymbol{H}_1 = \boldsymbol{H}$，对 $k = 1, 2, \cdots$, 取$\mu_k = h_{nn}^{(k)}$,

$$\begin{cases} \boldsymbol{H}_k - \mu_k\boldsymbol{I} = \boldsymbol{Q}_k\boldsymbol{R}_k, \\ \boldsymbol{H}_{k+1} = \boldsymbol{R}_k\boldsymbol{Q}_k + \mu_k\boldsymbol{I}. \end{cases}$$

5.4 矩阵的奇异值分解

定义： 设$\boldsymbol{A} \in \mathbb{C}_r^{m\times n}\ (r>0)$, $\boldsymbol{A}^H\boldsymbol{A}$ 的特征值为

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$$

则称$\sigma_i = \sqrt{\lambda_i}\ (i=1,2,\cdots,n)$ 为$\boldsymbol{A}$ 的奇异值.

定义： 设$\boldsymbol{A}, \boldsymbol{B} \in C^{m\times n}$，若存在$m$ 阶酉矩阵$\boldsymbol{U}$ 和n 阶酉矩阵$\boldsymbol{V}$，使得

$$\boldsymbol{U}^H\boldsymbol{A}\boldsymbol{V} = \boldsymbol{B}$$

则称$\boldsymbol{A}$ 与$\boldsymbol{B}$ 酉等价。

定理5.13: 酉等价矩阵有相同的奇异值。

定理5.14： 设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}\ (r>0)$，则存在 m 阶酉矩阵 $\boldsymbol{U}$ 和 n 阶酉矩阵 $\boldsymbol{V}$，使得

$$\boldsymbol{U}^H\boldsymbol{A}\boldsymbol{V} = \begin{pmatrix} \boldsymbol{\Sigma} & 0 \\ 0 & 0 \end{pmatrix},$$

其中 $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_{\mathrm{r}})$，这里 σ_i 为 $\boldsymbol{A}$ 的非零奇异值。而

$$\boldsymbol{A} = \boldsymbol{U}\begin{pmatrix} \boldsymbol{\Sigma} & 0 \\ 0 & 0 \end{pmatrix}\boldsymbol{V}^H,$$

称之为 $\boldsymbol{A}$ 的奇异值分解。

证明： 存在 n 阶酉矩阵 $\boldsymbol{V}$ 使得

$$\boldsymbol{V}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{V} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \begin{pmatrix} \boldsymbol{\Sigma}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

将 $\boldsymbol{V}$ 分块为

$$\boldsymbol{V} = (\boldsymbol{V}_1, \boldsymbol{V}_2), \quad (\boldsymbol{V}_1 \in \mathbb{C}^{n\times r},\ \boldsymbol{V}_2 \in \mathbb{C}^{n\times(n-r)}),$$

得$\boldsymbol{V}_1^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{V}_1 = \boldsymbol{\Sigma}^2$, $\boldsymbol{V}_2^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{V}_2 = \mathbf{0}$, 于是

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_1^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1} = \boldsymbol{I}_r, \ \ (\boldsymbol{A}\boldsymbol{V}_2)^H(\boldsymbol{A}\boldsymbol{V}_2) = \mathbf{0},$$

从而 $\boldsymbol{A}\boldsymbol{V}_2 = \mathbf{0}$.

又记 $\boldsymbol{U}_1=\boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1}$，则 $\boldsymbol{U}_1^H\boldsymbol{U}_1=\boldsymbol{I}_r$，即 $\boldsymbol{U}_1$ 的 r 个列是两两正交的单位向量，取 $\boldsymbol{U}_2\in\mathbb{C}^{m\times(m-r)}$ 使 $\boldsymbol{U}=(\boldsymbol{U}_1,\boldsymbol{U}_2)$ 为 m 阶酉矩阵，即

$$\boldsymbol{U}_2^H\boldsymbol{U}_1=\mathbf{0},\ \boldsymbol{U}_2^H\boldsymbol{U}_2=\boldsymbol{I}_{m-r},$$

则有

$$\boldsymbol{U}^H\boldsymbol{A}\boldsymbol{V}=\begin{pmatrix}\boldsymbol{U}_1^H\\ \boldsymbol{U}_2^H\end{pmatrix}\boldsymbol{A}(\boldsymbol{V}_1,\boldsymbol{V}_2)=\begin{pmatrix}\boldsymbol{U}_1^H\boldsymbol{A}\boldsymbol{V}_1 & \boldsymbol{U}_1^H\boldsymbol{A}\boldsymbol{V}_2\\ \boldsymbol{U}_2^H\boldsymbol{A}\boldsymbol{V}_1 & \boldsymbol{U}_2^H\boldsymbol{A}\boldsymbol{V}_2\end{pmatrix},$$

$$=\begin{pmatrix}\boldsymbol{U}_1^H(\boldsymbol{U}_1\boldsymbol{\Sigma}) & \mathbf{0}\\ \boldsymbol{U}_2^H(\boldsymbol{U}_1\boldsymbol{\Sigma}) & \mathbf{0}\end{pmatrix}=\begin{pmatrix}\boldsymbol{\Sigma} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}.$$

求 $\boldsymbol{A}$ 的奇异值分解步骤：

(1) 求酉矩阵 $\boldsymbol{V}$ 使 $\boldsymbol{V}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{V} = \begin{pmatrix} \boldsymbol{\Sigma}^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$;

(2) 将 $\boldsymbol{V}$ 分块为

$$\boldsymbol{V} = (\boldsymbol{V}_1, \boldsymbol{V}_2), \quad (\boldsymbol{V}_1 \in \mathbb{C}^{n\times r}, \ \boldsymbol{V}_2 \in \mathbb{C}^{n\times(n-r)}),$$

计算 $\boldsymbol{U}_1 = \boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1}$;

(3) 取 $\boldsymbol{U}_2$ 使 $\boldsymbol{U} = (\boldsymbol{U}_1, \boldsymbol{U}_2)$ 为 m 阶酉矩阵，则

$$\boldsymbol{A} = \boldsymbol{U}\begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{V}^H$$

为 $\boldsymbol{A}$ 的奇异值分解.

例： 求$\boldsymbol{A}=\begin{pmatrix}1&0&1\\1&1&0\end{pmatrix}$的奇异值分解.

解： $\boldsymbol{A}^T\boldsymbol{A}=\begin{pmatrix}2&1&1\\1&1&0\\1&0&1\end{pmatrix}$，所以$\boldsymbol{A}^T\boldsymbol{A}$的特征值为 $\lambda_1=3$，$\lambda_2=1$，$\lambda_3=0$，对应的特征向量为 $\boldsymbol{p}_1=(2,1,1)^T$, $\boldsymbol{p}_2=(0,-1,1)^T$, $\boldsymbol{p}_3=(-1,1,1)^T$，经过标准化后，可得

$$\boldsymbol{V}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3)=\begin{pmatrix}\frac{2}{\sqrt{6}}&0&-\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{6}}&-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{6}}&\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{3}}\end{pmatrix},$$

使得 $\boldsymbol{V}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{V}=\mathrm{diag}(3,1,0)$。

计算

$$\boldsymbol{U}_1 = \boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix},$$

取 $\boldsymbol{U} = \boldsymbol{U}_1$ 是酉矩阵。

故 $\boldsymbol{A}$ 的奇异值分解为

$$\boldsymbol{A} = \boldsymbol{U}\begin{pmatrix} \boldsymbol{\Sigma} & 0 \end{pmatrix}\boldsymbol{V}^H = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$