Jacobi 方法

张晓平

2018 年 10 月 28 日

1 Schur 分解

首先我们来介绍 Schur 分解，它是 QR 方法的理论基础，而 QR 方法是求解一般方阵（可以是实方阵、复方阵、对称阵或非对称阵）特征值问题的一种迭代方法。

Schur 分解有两个版本，分别针对实方阵和复方阵。

定理 (复 Schur 分解)**.** *设 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$，则 $\boldsymbol{A}$ 可分解成*

$$\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{R}\boldsymbol{Q}^{-1}=\boldsymbol{Q}\boldsymbol{R}\boldsymbol{Q}^{H}, \tag{1}$$

其中 $\boldsymbol{Q}\in\mathbb{C}^{n\times n}$ 为酉矩阵，即 $\boldsymbol{Q}^{-1}=\boldsymbol{Q}^{H}$，$\boldsymbol{Q}^{H}$ 表示 $\boldsymbol{Q}$ 的共轭转置，$\boldsymbol{R}\in\mathbb{C}^{n\times n}$ 为上三角阵，其对角元为 $\boldsymbol{A}$ 的各个特征值。

证明. 对 n 用数学归纳法。

1. 当 $n=1$ 时，结论显然成立。
2. 假设结论对 $n-1$ 成立，以下证明结论对 n 也成立。令 $\boldsymbol{v}$ 为 $\boldsymbol{A}$ 对应于特征值 λ 的特征向量，且 $\|\boldsymbol{v}\|^2=\boldsymbol{v}^H\boldsymbol{v}=1$。现构造一个酉矩阵

   $$\boldsymbol{Q}_0=[\boldsymbol{v},\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{n-1}]=[\boldsymbol{v},\boldsymbol{W}]$$

   使得

   $$\boldsymbol{Q}_0\boldsymbol{Q}_0^H=\boldsymbol{I},$$

   其中 $\boldsymbol{W}=[\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{n-1}]\in\mathbb{C}^{n\times(n-1)}$ 是一个向量组，它们两两正交并且都与 $\boldsymbol{v}$ 正交（可通过 Gram-Schmidt 正交化过程得到）。于是

   $$\boldsymbol{Q}_0^H\boldsymbol{A}\boldsymbol{Q}_0=\begin{bmatrix}\boldsymbol{v}^H\\ \boldsymbol{W}^H\end{bmatrix}\boldsymbol{A}\begin{bmatrix}\boldsymbol{v} & \boldsymbol{W}\end{bmatrix}=\begin{bmatrix}\boldsymbol{v}^H\boldsymbol{A}\boldsymbol{v} & \boldsymbol{v}^H\boldsymbol{A}\boldsymbol{W}\\ \boldsymbol{W}^H\boldsymbol{A}\boldsymbol{v} & \boldsymbol{W}^H\boldsymbol{A}\boldsymbol{W}\end{bmatrix}:=\begin{bmatrix}\lambda & *\\ 0 & \boldsymbol{A}_1\end{bmatrix}$$

   这里用到了 $\boldsymbol{W}^H\boldsymbol{A}\boldsymbol{v}=\lambda\boldsymbol{W}^H\boldsymbol{v}=0$，并注意 $\boldsymbol{A}_1=\boldsymbol{W}^H\boldsymbol{A}\boldsymbol{W}\in\mathbb{C}^{(n-1)\times(n-1)}$。由归纳假设，$A_1$ 可分解为

   $$\boldsymbol{A}_1=\boldsymbol{Q}_1\boldsymbol{R}_1\boldsymbol{Q}_1^H,\ \text{i.e.,}\ \boldsymbol{R}_1=\boldsymbol{Q}_1^H\boldsymbol{A}_1\boldsymbol{Q}_1.$$

   令

   $$\widetilde{\boldsymbol{Q}}_1=\begin{bmatrix}1 & 0\\ 0 & \boldsymbol{Q}_1\end{bmatrix}$$

   并定义 $\boldsymbol{Q}=\boldsymbol{Q}_0\widetilde{\boldsymbol{Q}}_1$，可得

   $$\boldsymbol{Q}^H\boldsymbol{A}\boldsymbol{Q}=\widetilde{\boldsymbol{Q}}_1^H\boldsymbol{Q}_0^H\boldsymbol{A}\boldsymbol{Q}_0\widetilde{\boldsymbol{Q}}_1=\begin{bmatrix}1 & 0\\ 0 & \boldsymbol{Q}_1^H\end{bmatrix}\begin{bmatrix}\lambda & *\\ 0 & \boldsymbol{A}_1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & \boldsymbol{Q}_1\end{bmatrix}=\begin{bmatrix}\lambda & *\\ 0 & \boldsymbol{Q}_1^H\boldsymbol{A}_1\boldsymbol{Q}_1\end{bmatrix}=\begin{bmatrix}\lambda & *\\ 0 & \boldsymbol{R}_1\end{bmatrix}:=\boldsymbol{R}.$$

$\square$

定理. 实矩阵的所有复特征值构成共轭对。

证明. 设 $\alpha+i\beta$ 是 $\boldsymbol{A}$ 的特征值，$\boldsymbol{u}+i\boldsymbol{v}$ 为对应的特征向量，则

$$\boldsymbol{A}(\boldsymbol{u}+i\boldsymbol{v})=(\alpha+i\beta)(\boldsymbol{u}+i\boldsymbol{v}),$$

即

$$\boldsymbol{A}\boldsymbol{u}+i\boldsymbol{A}\boldsymbol{v}=(\alpha\boldsymbol{u}-\beta\boldsymbol{v})+i(\beta\boldsymbol{u}+\alpha\boldsymbol{v}),$$

于是有

$$\boldsymbol{A}\boldsymbol{u}-i\boldsymbol{A}\boldsymbol{v}=(\alpha\boldsymbol{u}-\beta\boldsymbol{v})-i(\beta\boldsymbol{u}+\alpha\boldsymbol{v}),$$

亦即

$$\boldsymbol{A}(\boldsymbol{u}-i\boldsymbol{v})=(\alpha-i\beta)(\boldsymbol{u}-i\boldsymbol{v}).$$

这说明 $\alpha-i\beta$ 也是 A 的特征值，与其对应的特征向量为 $\boldsymbol{u}-i\boldsymbol{v}$。 $\square$

定理 (实 Schur 分解)**.** 设 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$，则 $\boldsymbol{A}$ 可分解成

$$\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{R}\boldsymbol{Q}^{-1}=\boldsymbol{Q}\boldsymbol{R}\boldsymbol{Q}^{T}, \tag{2}$$

其中 $\boldsymbol{Q}\in\mathbb{R}^{n\times n}$ 为正交矩阵，$\boldsymbol{R}\in\mathbb{R}^{n\times n}$ 为拟上三角阵，形如

$$\boldsymbol{R}=\begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} & \cdots & \boldsymbol{R}_{1m} \\ 0 & \boldsymbol{R}_{22} & \cdots & \boldsymbol{R}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{R}_{mm} \end{bmatrix}$$

其中 $\boldsymbol{R}_{ii}$ 要么为实数（$\boldsymbol{A}$ 的实特征值），要么为 2×2 的矩阵（其特征值为共轭对，也是 $\boldsymbol{A}$ 的特征值）。

证明. 对 n 用数学归纳法。

1. 当 $n=1$ 时，结论显然成立。

2. 假设结论对 $k\le n$ 成立，以下证明结论对 n 也成立。设 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$，若 A 有一个实特征值 λ，则类似于复 Schur 分解，可以将 A 做相似变换约化成

$$\boldsymbol{A}\sim\begin{bmatrix} \lambda & * \\ 0 & \boldsymbol{A}_1 \end{bmatrix},\quad \boldsymbol{A}_1\in\mathbb{R}^{(n-1)\times(n-1)}$$

利用归纳假设即可完成证明；若 $\boldsymbol{A}$ 有一个互为共轭的特征值对 $\alpha\pm i\beta(\beta\neq0)$，其对应的特征向量为 $\boldsymbol{u}\pm i\boldsymbol{v}$，则

$$\boldsymbol{A}(\boldsymbol{u}+i\boldsymbol{v})=(\alpha+i\beta)(\boldsymbol{u}+i\boldsymbol{v})\quad\Rightarrow\quad \boldsymbol{A}(\boldsymbol{u}\ \boldsymbol{v})=(\boldsymbol{u}\ \boldsymbol{v})\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}.$$

$\beta\neq0$ 意味着 $\boldsymbol{u},\boldsymbol{v}$ 线性无关，因此可以利用 Gram-Schimt 正交化过程构造一个正交向量组

$$\boldsymbol{Q}=\left(\boldsymbol{u},\ \boldsymbol{w},\ \boldsymbol{q}_1,\ \cdots,\ q_{n-2}\right):=\left(\boldsymbol{u},\ \boldsymbol{w},\ \widetilde{\boldsymbol{Q}}\right),$$

其中

$$\boldsymbol{w}=\frac{a\boldsymbol{u}+\boldsymbol{v}}{b},\quad a=-\frac{(x,y)}{(x,x)},\quad b=\|a\boldsymbol{u}+\boldsymbol{v}\|_2.$$

于是

$$\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q}=\begin{pmatrix}\boldsymbol{u}^T\\ \boldsymbol{w}^T\\ \widetilde{\boldsymbol{Q}}^T\end{pmatrix}\boldsymbol{A}\left(\boldsymbol{u},\ \boldsymbol{w},\ \widetilde{\boldsymbol{Q}}\right)=\begin{pmatrix}\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u} & \boldsymbol{u}^T\boldsymbol{A}\boldsymbol{w} & \boldsymbol{u}^T\boldsymbol{A}\widetilde{\boldsymbol{Q}}\\ \boldsymbol{w}^T\boldsymbol{A}\boldsymbol{u} & \boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w} & \boldsymbol{w}^T\boldsymbol{A}\widetilde{\boldsymbol{Q}}\\ \widetilde{\boldsymbol{Q}}^T\boldsymbol{A}\boldsymbol{u} & \widetilde{\boldsymbol{Q}}^T\boldsymbol{A}\boldsymbol{w} & \widetilde{\boldsymbol{Q}}^T\boldsymbol{A}\widetilde{\boldsymbol{Q}}\end{pmatrix}:=\begin{pmatrix}\boldsymbol{R}_{11} & \widetilde{\boldsymbol{R}}_{12}\\ 0 & \widetilde{\boldsymbol{R}}_{22}\end{pmatrix},$$

其中 $\boldsymbol{R}_{11}\in\mathbb{R}^{2\times 2}$，特征值为 $\alpha\pm i\beta$，这里用到了

$$\widetilde{\boldsymbol{Q}}^T\boldsymbol{A}\boldsymbol{u}=\widetilde{\boldsymbol{Q}}^T(\alpha\boldsymbol{u}-\beta\boldsymbol{v})=\widetilde{\boldsymbol{Q}}^T(\alpha\boldsymbol{u}-\beta(a\boldsymbol{u}-b\boldsymbol{w}))=(\alpha-a\beta)\widetilde{\boldsymbol{Q}}^T\boldsymbol{u}+b\beta\widetilde{\boldsymbol{Q}}^T\boldsymbol{v}=0.$$

利用归纳假设即可完成证明。

□

注 1. *Schur 分解定理仅说明了 $\boldsymbol{Q}$ 和 U 的存在性，但并未给出方法去求它们。*

2 基本 QR 方法

对给定的 $\boldsymbol{A}_0=\boldsymbol{A}\in\mathbb{C}^{n\times n}$，QR 算法的基本迭代格式为：

> 对 $k=1,2,\cdots$，
>
> - 对 $\boldsymbol{A}_{k-1}$ 进行 QR 分解，即 $\boldsymbol{A}_{k-1}=\boldsymbol{Q}_k\boldsymbol{R}_k$；
> - 计算 $\boldsymbol{A}_k=\boldsymbol{R}_k\boldsymbol{Q}_k$
>
> 其中 $\boldsymbol{Q}_k\in\mathbb{C}^{n\times n}$ 为酉矩阵，$\boldsymbol{R}_k$ 为上三角阵。

经过 k 次迭代后，可得

$$\begin{aligned}\boldsymbol{A}_k=\boldsymbol{R}_k\boldsymbol{Q}_k&=\boldsymbol{Q}_k^H\boldsymbol{A}_{k-1}\boldsymbol{Q}_k=\cdots\\ &=\boldsymbol{Q}_k^H\cdots\boldsymbol{Q}_1^H\boldsymbol{A}\boldsymbol{Q}_1\cdots\boldsymbol{Q}_k\\ &:=(\boldsymbol{Q}^{(k)})^H\boldsymbol{A}\boldsymbol{Q}^{(k)},\end{aligned}$$

其中

$$\boldsymbol{Q}^{(k)}=\boldsymbol{Q}_1\cdots\boldsymbol{Q}_k$$

为正交阵，这说明矩阵序列 $\{\boldsymbol{A}_k\}$ 中的每一个矩阵都相似于 $\boldsymbol{A}$。

将 $\boldsymbol{A}_k=\boldsymbol{Q}_{k+1}\boldsymbol{R}_{k+1}$ 代入上式可得

$$\boldsymbol{Q}_{k+1}\boldsymbol{R}_{k+1}=(\boldsymbol{Q}^{(k)})^H\boldsymbol{A}\boldsymbol{Q}^{(k)}$$

即

$$\boldsymbol{Q}^{(k)}\boldsymbol{Q}_{k+1}\boldsymbol{R}_{k+1}=\boldsymbol{A}\boldsymbol{Q}^{(k)}$$

从而有

$$\boldsymbol{Q}^{(k)}\boldsymbol{Q}_{k+1}\boldsymbol{R}_{k+1}\boldsymbol{R}_k\cdots\boldsymbol{R}_1=\boldsymbol{A}\boldsymbol{Q}^{(k)}\boldsymbol{R}_k\cdots\boldsymbol{R}_1$$

即

$$\boldsymbol{Q}^{(k+1)}\boldsymbol{R}^{(k+1)}=\boldsymbol{A}\boldsymbol{Q}^{(k)}\boldsymbol{R}^{(k)}$$

其中 $\boldsymbol{R}^{(k)}=\boldsymbol{R}_k\cdots\boldsymbol{R}_1$。

在适当条件下，$\boldsymbol{A}_k$ 的所有的或大部分的对角线以下的元素都将趋于零。

定理. *设 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$ 的 n 个特征值满足 $|\lambda_1|>|\lambda_2|>\cdots>|\lambda_n|$。令*

$$\boldsymbol{A}=\boldsymbol{Y}^{-1}\boldsymbol{D}\boldsymbol{Y},\quad \boldsymbol{D}:=\mathrm{diag}(\lambda_1,\cdots,\lambda_n).$$

如果 $\boldsymbol{Y}$ 有 LU 分解，则有

1. *$\boldsymbol{A}_k$ 主对角元以下的元素以 $O(t^k)$ 的速度趋于零，其中*

$$t:=\max\left\{\left|\frac{\lambda_2}{\lambda_1}\right|,\cdots,\left|\frac{\lambda_n}{\lambda_{n-1}}\right|\right\}<1.$$

2. *$\boldsymbol{A}_k$ 的主对角线收敛到 $\boldsymbol{D}$。*

2.1 特征向量的求解

因 $\boldsymbol{Q}^H\boldsymbol{A}\boldsymbol{Q}=\boldsymbol{R}$，故 $\boldsymbol{A}$ 与 $\boldsymbol{R}$ 有相同的特征值，即为 $\boldsymbol{R}$ 的主对角元素。那如何计算 A 的特征向量呢?

- 如果 $\boldsymbol{A}$ 对称，则上三角阵 $\boldsymbol{R}$ 将退化成对角阵，且各对角元为 $\boldsymbol{A}$ 的各个特征值。此时正交阵 $\boldsymbol{Q}$ 的各列即为 $\boldsymbol{A}$ 的特征向量。

- 如果 $\boldsymbol{A}$ 不对称，则 $\boldsymbol{R}$ 仅为上三角阵。设 λ 为 $\boldsymbol{A}$ 的任意特征值，$\boldsymbol{v}$ 为其对应的特征向量，我们可以通过以下两种方式来计算 $\boldsymbol{v}$:

  1. 求解

  $$(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{v}_{k+1}=\boldsymbol{v}_k$$

  注意到若 λ 为 $\boldsymbol{A}$ 的某个特征值，则 $\boldsymbol{A}-\lambda\boldsymbol{I}$ 不可逆。我们可以引入一个扰动 δ，一方面我们希望 δ 足够大以使得 $\boldsymbol{A}-(\lambda+\delta)\boldsymbol{I}$ 变得可逆，另一方面我们又希望 δ 足够小以使得 $\lambda+\delta$ 能充分靠近 λ。

  2. 令 $\boldsymbol{\Lambda}$ 和 $\widetilde{\boldsymbol{V}}=[\widetilde{\boldsymbol{v}}_1,\cdots,\widetilde{\boldsymbol{v}}_n]$ 为 $\boldsymbol{R}$ 的特征值和特征向量，即

  $$\boldsymbol{R}=\widetilde{\boldsymbol{V}}\boldsymbol{\Lambda}\widetilde{\boldsymbol{V}}^{-1}$$

  由 Schur 分解 $\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{R}\boldsymbol{Q}^{-1}$ 可知

  $$\boldsymbol{A}=\boldsymbol{Q}\widetilde{\boldsymbol{V}}\boldsymbol{\Lambda}\widetilde{\boldsymbol{V}}^{-1}\boldsymbol{Q}^{-1}=(\boldsymbol{Q}\widetilde{\boldsymbol{V}})\boldsymbol{\Lambda}(\boldsymbol{Q}\widetilde{\boldsymbol{V}})^{-1}$$

  故 $\boldsymbol{A}$ 的特征向量矩阵为

  $$[\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n]=\boldsymbol{V}=\boldsymbol{Q}\widetilde{\boldsymbol{V}}=\boldsymbol{Q}[\widetilde{\boldsymbol{v}}_1,\cdots,\widetilde{\boldsymbol{v}}_n]=[Q\widetilde{\boldsymbol{v}}_1,\cdots,Q\widetilde{\boldsymbol{v}}_n]$$

  我们可以通过求解如下齐次方程组

  $$(\boldsymbol{R}-\lambda_i\boldsymbol{I})\widetilde{\boldsymbol{v}}_i=0$$

  来计算 $\boldsymbol{R}$ 对应于 λ_i 的特征向量 $\widetilde{\boldsymbol{v}}_i$，然后计算

  $$\boldsymbol{v}_i=\boldsymbol{Q}\widetilde{\boldsymbol{v}}_i$$

  来计算 $\boldsymbol{A}$ 对应于 λ_i 的特征向量 $\boldsymbol{v}_i$.

3 Hessenberg 矩阵

定义. 设 $\boldsymbol{H}=(h_{ij})\in\mathbb{R}^{n\times n}$，若 $h_{ij}=0,\ j<i-1$，则称 H 为上 *Hessenberg* 矩阵；若 $h_{ij}=0,\ j>i+1$，则称 $\boldsymbol{H}$ 为下 *Hessenberg* 矩阵。上 *Hessenberg* 矩阵形如

$$\begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ & \ddots & \ddots & \vdots \\ & & h_{n,n-1} & h_{nn} \end{pmatrix},$$

下 *Hessenberg* 矩阵形如

$$\begin{pmatrix} h_{11} & h_{12} & & \\ h_{21} & h_{22} & \ddots & \\ \vdots & \ddots & \ddots & h_{n-1,n} \\ h_{n,1} & \cdots & h_{n,n-1} & h_{nn} \end{pmatrix}$$

3.1 上 Hessenberg 矩阵的 QR 迭代

为什么要用到上 Hessenberg 矩阵?

- 可证明，若 $\boldsymbol{H}_{m-1}$ 为上 Hessenberg 矩阵，则经过一次 QR 迭代，即

$$\boldsymbol{H}_{m-1}=\boldsymbol{Q}_m\boldsymbol{R}_m,\quad \boldsymbol{H}_m=\boldsymbol{R}_m\boldsymbol{Q}_m$$

  后，$\boldsymbol{H}_m$ 仍为上 Hessenberg 矩阵。

- 对上 Hessenberg 矩阵做 QR 分解的工作量大大减小。

下面举例说明：考虑 5 阶上 Hessenberg 矩阵

$$\boldsymbol{H}=\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & * & * \end{pmatrix}$$

可依次确定平面旋转变换 $\boldsymbol{P}_{12},\boldsymbol{P}_{23},\boldsymbol{P}_{34},\boldsymbol{P}_{45}$ 使得

$$\boldsymbol{P}_{45}\boldsymbol{P}_{34}\boldsymbol{P}_{23}\boldsymbol{P}_{12}\boldsymbol{H}=\begin{pmatrix} * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & * & * \\ & & & & * \end{pmatrix}$$

要完成一次 QR 迭代，还需计算

$$\widetilde{\boldsymbol{H}}=\boldsymbol{R}\boldsymbol{Q}=\boldsymbol{R}\boldsymbol{P}_{12}^T\boldsymbol{P}_{23}^T\boldsymbol{P}_{34}^T\boldsymbol{P}_{45}^T$$

而

$$
\boldsymbol{R}\boldsymbol{P}_{12}^T = \begin{pmatrix} * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & * & * \\ & & & & * \end{pmatrix} \begin{pmatrix} * & * & & & \\ * & * & & & \\ & & & & \\ & & & & \\ & & & & \end{pmatrix} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & & * & * & * \\ & & & * & * \\ & & & & * \end{pmatrix}
$$

$$
\boldsymbol{R}\boldsymbol{P}_{12}^T\boldsymbol{P}_{23}^T = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & & * & * & * \\ & & & * & * \\ & & & & * \end{pmatrix} \begin{pmatrix} 0 & & & & \\ & * & * & & \\ & * & * & & \\ & & & & \\ & & & & \end{pmatrix} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & & * & * \\ & & & & * \end{pmatrix}
$$

$$
\boldsymbol{R}\boldsymbol{P}_{12}^T\boldsymbol{P}_{23}^T\boldsymbol{P}_{34}^T = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & & * & * \\ & & & & * \end{pmatrix} \begin{pmatrix} 0 & & & & \\ & 0 & & & \\ & & * & * & \\ & & * & * & \\ & & & & \end{pmatrix} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & & * \end{pmatrix}
$$

$$
\widetilde{\boldsymbol{H}} = \boldsymbol{R}\boldsymbol{P}_{12}^T\boldsymbol{P}_{23}^T\boldsymbol{P}_{34}^T\boldsymbol{P}_{45}^T = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & & * & * \\ & & & & * \end{pmatrix} \begin{pmatrix} 0 & & & & \\ & 0 & & & \\ & & 0 & & \\ & & & * & * \\ & & & * & * \end{pmatrix} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & * & * \end{pmatrix}
$$

由此可看出，经过一次 QR 迭代，所得到的 $\widetilde{\boldsymbol{H}}$ 仍是一个 Hessenberg 矩阵。

注 2. *对 Hessenberg 矩阵进行一次 QR 迭代的运算量为* $O(n^2)$*，而对一般方阵进行一次 QR 迭代的运算量为* $O(n^3)$*。*

利用 Givens 变换实现 Hessenberg 矩阵的一次 QR 迭代

```
function [H, Q] = qr_iter(H)
n = size(H, 1);
c = zeros(n-1, 1); s = c;
for k = 1:n-1
    [c(k), s(k)] = givens_rotation(H(k, k), H(k+1, k));
    G = [c(k) -s(k); s(k) c(k)];
    H(k:k+1, k:n) = G' * H(k:k+1, k:n);
end
Q = eye(n);
for k = 1:n-1
    G = [c(k) -s(k); s(k) c(k)];
    H(1:k+1, k:k+1) = H(1:k+1, k:k+1) * G;
    Q(1:k+1, k:k+1) = Q(1:k+1, k:k+1) * G;
end
end
```

3.2 一般方阵的上 Hessenberg 化

对于一般方阵 $\boldsymbol{A}$，我们可以通过相似变换将其约化为上 Hessenberg 矩阵，即构造一个非奇异矩阵 $\boldsymbol{Q}$ 使得

$$\widetilde{A} = \boldsymbol{QAQ}^{-1}$$

为上 Hessenberg 矩阵。该过程可通过 Householder 变换来实现。

以下仍举例来说明整个过程。设 $\boldsymbol{A} \in \mathbb{R}^{5\times 5}$，形如

$$\boldsymbol{A} = \begin{pmatrix} * & * & * & * & * \\ \Delta & * & * & * & * \\ \Delta & * & * & * & * \\ \Delta & * & * & * & * \\ \Delta & * & * & * & * \end{pmatrix}.$$

1. 构造 Householder 变换 $\boldsymbol{H}_1$，使得

$$\boldsymbol{H}_1\boldsymbol{A} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & * & * & * & * \\ & * & * & * & * \end{pmatrix},$$

此时 $\boldsymbol{H}_1$ 应该形如

$$\boldsymbol{H}_1 = \begin{pmatrix} 1 & 0 \\ 0 & \widetilde{\boldsymbol{H}}_1 \end{pmatrix},$$

其中 $\widetilde{\boldsymbol{H}}_1 \in \mathbb{R}^{4\times 4}$ 使得

$$\widetilde{\boldsymbol{H}}_1 \begin{pmatrix} \Delta \\ \Delta \\ \Delta \\ \Delta \end{pmatrix} = \begin{pmatrix} * \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

为完成相似变换，还需计算 $\boldsymbol{H}_1\boldsymbol{AH}_1$，即

$$\boldsymbol{H}_1\boldsymbol{AH}_1 = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & * & * & * & * \\ & * & * & * & * \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \widetilde{\boldsymbol{H}}_1 \end{pmatrix} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & * & * & * & * \\ & * & * & * & * \end{pmatrix},$$

即通过第一次相似变换，$\boldsymbol{H}_1\boldsymbol{AH}_1$ 第一列中的后 3 个元素置为 0.

2. 构造 Householder 变换 $\boldsymbol{H}_2$，使得

$$\boldsymbol{H}_2(\boldsymbol{H}_1\boldsymbol{AH}_1) = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & * & * & * \end{pmatrix},$$

此时 $\boldsymbol{H}_2$ 应该形如

$$\boldsymbol{H}_2 = \begin{pmatrix} \boldsymbol{I}_2 & 0 \\ 0 & \widetilde{\boldsymbol{H}}_2 \end{pmatrix}, \quad \widetilde{\boldsymbol{H}}_2 \in \mathbb{R}^{3\times 3}.$$

为完成相似变换，还需计算 $\boldsymbol{H}_2(\boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1)\boldsymbol{H}_2$，即

$$\boldsymbol{H}_2(\boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1)\boldsymbol{H}_2=\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & * & * & * \end{pmatrix}\begin{pmatrix} \boldsymbol{I}_2 & 0 \\ 0 & \widetilde{\boldsymbol{H}}_2 \end{pmatrix}=\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & * & * & * \end{pmatrix},$$

即通过第二次相似变换，$\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1\boldsymbol{H}_2$ 第二列中的后 2 个元素置为 0.

第三步，构造 Householder 变换 $\boldsymbol{H}_3$，使得

$$\boldsymbol{H}_3(\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1\boldsymbol{H}_2)=\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & * & * \end{pmatrix},$$

此时 $\boldsymbol{H}_3$ 应该形如

$$\boldsymbol{H}_3=\begin{pmatrix} \boldsymbol{I}_3 & 0 \\ 0 & \widetilde{\boldsymbol{H}}_3 \end{pmatrix},\quad \widetilde{\boldsymbol{H}}_3\in\mathbb{R}^{2\times 2}.$$

为完成相似变换，还需计算 $\boldsymbol{H}_3(\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1\boldsymbol{H}_2)\boldsymbol{H}_3$，即

$$\boldsymbol{H}_3(\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1\boldsymbol{H}_2)\boldsymbol{H}_3=\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & * & * \end{pmatrix}\begin{pmatrix} \boldsymbol{I}_3 & 0 \\ 0 & \widetilde{\boldsymbol{H}}_3 \end{pmatrix}=\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & * & * \end{pmatrix},$$

即通过第三次相似变换，$\boldsymbol{H}_3\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1\boldsymbol{H}_2\boldsymbol{H}_3$ 第三列中的最后一个元素置为 0。至此，经过三次相似变换，可构造出一个正交矩阵

$$\boldsymbol{Q}_0=\boldsymbol{H}_1\boldsymbol{H}_2\boldsymbol{H}_3,$$

使得 $\boldsymbol{Q}_0^T\boldsymbol{A}\boldsymbol{Q}_0=\boldsymbol{H}$ 为上 Hessenberg 矩阵。

一般地，对于方阵 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$，可找到 $n-2$ 个 Householder 变换 $\boldsymbol{H}_1,\cdots,\boldsymbol{H}_{n-2}$ 使得

$$\boldsymbol{H}_{n-2}\cdots\boldsymbol{H}_1\boldsymbol{A}\boldsymbol{H}_1\cdots\boldsymbol{H}_{n-2}$$

为上 Hessenberg 矩阵。记 $\boldsymbol{Q}_0=\boldsymbol{H}_1\cdots\boldsymbol{H}_{n-2}$，则 $\boldsymbol{Q}_0^T\boldsymbol{A}\boldsymbol{Q}_0$ 为上 Hessenberg 矩阵。该过程称为 A 的上 Hessenberg 化。

矩阵 $\boldsymbol{A}$ 的上 Hessenberg 化

```
function [A, Q] = hessenberg(A)
n = size(A, 1); Q = eye(n);
d = zeros(n-2, 1);
for k = 1:n-2
    [v, beta] = householder(A(k+1:end, k));
    H = blkdiag(eye(k), eye(n-k)-beta*v*v');
    A(k+1:end, k:end) = A(k+1:end, k:end) - beta * v * (v' * A(k+1:end, k:
    end));
```

```
    A(:, k+1:end) = A(:, k+1:end) - beta * (A(:, k+1:end) * v) * v';
    Q = H * Q;
end
end
```

4 实用 QR 方法

实际计算中，我们通过以下方式来实现 QR 方法：设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$，

1. 将 $\boldsymbol{A}$ 上 Hessenberg 化，即构造正交矩阵 $\boldsymbol{Q}_0$ 使得

$$\boldsymbol{Q}_0^T \boldsymbol{A} \boldsymbol{Q}_0 = \boldsymbol{H},$$

   需要的工作量为 $O(n^3)$；

2. 对上 Hessenberg 矩阵 $\boldsymbol{H}$ 进行 QR 迭代，即重复进行以下步骤

$$\boldsymbol{H} = \boldsymbol{Q}\boldsymbol{R}, \quad \widetilde{\boldsymbol{H}} = \boldsymbol{R}\boldsymbol{Q}, \quad \boldsymbol{H} = \widetilde{\boldsymbol{H}},$$

   直至收敛。每次迭代所需的工作量为 $O(n^2)$。

5 代码

```
%% find eigenvalue of quasi-upper-triangular matrix
function lambda = eigvalue(H)
n = size(H, 1); lambda = zeros(n, 1);
k = 1;
while 1
    if k > n
        break;
    end
    if k == n
        lambda(k) = H(k,k);
        break;
    end
    if abs(H(k+1, k)) < 1e-10
        lambda(k) = H(k,k);
        k = k+1;
    else
        a = H(k, k); b = H(k, k+1); c = H(k+1, k); d = H(k+1, k+1);
        lambda(k) = (a+d+sqrt((a-d)^2+4*b*c))/2;
        lambda(k+1) = (a+d-sqrt((a-d)^2+4*b*c))/2;
        k = k+2;
    end
end
end
```

```
%% find eigenvector of quasi-upper-triangular matrix
function v = eigvector(H, lambda)
n = size(H, 1);
I = eye(n); v = zeros(n);
for k = 1:n
    H1 = H - lambda(k)*I;
    H2 = H1([1:k-1,k+1:end],[1:k-1,k+1:end]);
    v(k, k) = 1;
    v([1:k-1,k+1:end], k) = -H2\H1([1:k-1,k+1:end], k);
    v(:, k) = v(:, k)/norm(v(:, k));
end
end
```

```
n = 5; tol = 1e-10;
A = rand(n); A1 = A;
lambda1 = eig(A);

%% Upper Hessenberg
[A, Q0] = hessenberg(A);

%% QR iterations
H = triu(A,-1);
[H, Q1] = qr_iter(H);
d1 = diag(H);
for k = 1:1000
    [H, Q] = qr_iter(H);
    d2 = diag(H);
    if norm(d1-d2)/norm(d2) < tol
       break;
    end
    d1 = d2;
    Q1 = Q1 * Q;
end
fprintf('After %d iteration, QR iteration converges!\n\n', k);

%% find the eigenvalues
lambda = eigvalue(H);
fprintf('Eigen Values:\n');
lambda


%% find the eigenvectors
%  i.e., solve the homegenous equation (A - lambda I)x = 0
v = eigvector(H, lambda);
```

```
Q = Q1' * Q0;
v = Q' * v;
fprintf('Eigen Vectors:\n');
v

%A1*v - v*diag(lambda)
```