第六章 代数特征值问题

般东生
yindongsheng@tsinghua.edu.cn

清华大学数学科学系

2023年秋季学期

代数特征值问题

定义

设$A \in \mathbb{C}^{n\times n}$，若存在非零向量$\boldsymbol{x} \in \mathbb{C}^n$和$\lambda \in \mathbb{C}$，满足

$$A\boldsymbol{x} = \lambda \boldsymbol{x},$$

称λ为特征值，$\boldsymbol{x}$为相应的特征向量。

由于

$$f(\lambda) = \det(\lambda I - A) = \lambda^n + c_1\lambda^{n-1} + \cdots + c_{n-1}\lambda + c_n$$

若λ是A的特征值的充分必要条件是λ是$f(\lambda) = 0$的根。

定理

n阶复系数多项式方程在复数域上有n个根。

一般不能用直接法求解多项式方程的根，因此特征值问题的数值方法只能是迭代法。反过来，为了求多项式

$$q(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

的根，可以把$q(\lambda)$看成它的友矩阵

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ & 1 & \cdots & 0 & -a_2 \\ & & \ddots & \vdots & \vdots \\ & & & 1 & -a_{n-1} \end{bmatrix}$$

的特征多项式。它是一个上Hessenberg矩阵，可以用QR迭代求解。
特征值在结构分析、电子结构的计算等领域有广泛的应用。
Google的PageRank归结为求解其转移矩阵的最大特征值对应的特征向量。

力学中的特征值问题

美国数学家斯特让（G..Strang）在其经典教材《线性代数及其应用》中这样介绍了特征值作为频率的物理意义，他说：

大概最简单的例子（我从不相信其真实性，虽然据说1831年有一桥梁毁于此因）是一队士兵通过桥梁的例子。传统上，他们要停止齐步前进而要散步通过。这个理由是因为他们可能以等于桥的特征值之一的频率齐步行进，从而将发生共振。就像孩子的秋千那样，你一旦注意到一个秋千的频率，和此频率相配，你就使频率荡得更高。一个工程师总是试图使他的桥梁或他的火箭的自然频率远离风的频率或液体燃料的频率；而在另一种极端情况，一个证券经纪人则尽毕生精力于努力到达市场的自然频率线。特征值是几乎任何一个动力系统的最重要的特征。

- 应力分析：主应力是应力张量矩阵的特征值，主方向是特征向量的方向。
- 对多自由度机械结构作振动分析时，其特征值给出振动的自然频率，特征向量给出振动行为。

塔科马海峡吊桥（Tacoma Narrows Bridge)

塔科马海峡吊桥是位于美国华盛顿州塔科马的悬索桥。第一条桥于1938年开始建造，为了减低造价，把桥面设计的厚度从25呎减至8呎（2.4米），该桥于1940年7月1日通车，但在启用后数个星期，桥面便开始出现上下摆动。有鉴及此，有关人员在支柱上安装摄录机，以便观测摆动。同时也吸引了不少驾车人士慕名而来，感受其振荡威力的刺激，一些大风的日子，其桥面摆动幅度甚至可达五英尺之多。其后桥面的波动幅度不断增加，工程人员尝试加建缆索及液压缓冲装置去试图减低波动，但不成功。

在持续数个月的摆动之下，桥梁最终于同年11月7日倒塌，其过程获得全程记录。当天早上，桥面的上下摆动突然停止，取而代之的是出现左右的扭力摆动，当时有两人被困在桥上，后来也成功逃离现场，然后桥面在数分钟内陆续崩塌。

量子力学中的特征值问题

- 量子力学中的一个重要问题是求Schrödinger方程的特征值

$$H\psi_E = E\psi_E,$$

这里H是Hamilton量，E是特征值，代表能量，ψ_E是对应的特征函数，即波函数。

- 原子物理和分子物理中，在Hartree-Fock理论下，原子轨道和分子轨道可以定义为Fock算子的特征向量。相应的特征值通过Koopmans定理可以解释为电离势能。方程通常采用迭代程序求解，称为自洽场方法。在量子化学中，经常会把Hartree-Fock方程通过非正交基集合来表达。这个特定地表达是一个广义特征值问题称为Roothaan方程。

定理 (Gershgorin圆盘定理)

设 $A=(a_{ij})\in\mathbb{C}^{n\times n}$，则特征值

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

其中D_i为复平面上以a_{ii}为中心$r_i=\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|$ 为半径的圆盘，即

$$D_i\overset{def}{=}\left\{z\in\mathbb{C}\,\middle|\,|z-a_{ii}|\leq r_i\right\},\ i=1,\cdots,n$$

注记

圆盘定理可以从矩阵的元素估计特征值所在的范围。矩阵的n个特征值均落在n个圆盘上，但是不一定每个圆盘都有一个特征值。

定理 (第2圆盘定理)

设上述定理的n个圆盘中有m个圆盘构成一个连通域S，且S与其余$n-m$个圆盘严格分离，则S中恰有m个特征值，其中重特征值按其重数重复计算。

注记

设$A=(a_{ij})$，$B=\mathrm{diag}(\alpha_1,\cdots,\alpha_n)$，其中$\alpha_i\neq 0$，则$A$与

$$B^{-1}AB=\left(\frac{a_{ij}\alpha_j}{\alpha_i}\right)$$

有相同的特征值。适当的选取B，可使某个圆盘半径相对地减小，更易估计。当然也可能使其他的某些圆盘半径变大。

例 (圆盘估计的变化)

$$A=\begin{bmatrix}0.9 & 0.01 & 0.12\\ 0.01 & 0.8 & 0.13\\ 0.01 & 0.02 & 0.4\end{bmatrix}$$

解: A的三个圆盘为

$$\begin{aligned} D_1 &= \{z \big| |z-0.9| \le 0.13\},\ D_2=\{z\big|z-0.8|\le 0.14|\} \\ D_3 &= \{z\big||z-0.4|\le 0.03\}. \end{aligned}$$

D_1, D_2和 D_3严格分离，则D_3必有一实特征值。为估计另外两个特征值，缩小D_1, D_2的半径。取$B=\mathrm{diag}(1,1,0.1)$，则

$$B^{-1}AB=\begin{bmatrix}0.9 & 0.01 & 0.012\\ 0.01 & 0.8 & 0.013\\ 0.1 & 0.2 & 0.4\end{bmatrix}$$

则新的圆盘$\overline{D}_1=\left\{z\Big||z-0.9|\le 0.022\right\}, \overline{D}_2=\left\{z\Big||z-0.8|\le 0.023\right\}, \overline{D}_3=\left\{z\Big||z-0.4|\le 0.3\right\}$。

幂法

- 设A是可对角化的，其特征值满足

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

对应的特征向量 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 是$\mathbb{R}^n$的一组基。

- 对任意的初始向量

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

有

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

- 对任意向量$\boldsymbol{z}=(z_1,\cdots,z_n)^T$，定义

$$\max(\boldsymbol{z})=z_i,\ |z_i|=\|\boldsymbol{z}\|_\infty,\ |z_j|<\|\boldsymbol{z}\|_\infty,\ \forall j<i.$$

- 对任意$\alpha\in\mathbb{R}$，$\boldsymbol{z}\in\mathbb{R}^n$，有

$$\max(\alpha\boldsymbol{z})=\alpha\max(\boldsymbol{z}).$$

幂迭代法的算法可写为

$$\begin{cases} \boldsymbol{v}^{(0)} & = & \dfrac{\boldsymbol{\mu}}{\max \boldsymbol{\mu}} \\ \boldsymbol{z}^{(k)} & = & A\boldsymbol{v}^{(k-1)} \\ m_k & = & \max(\boldsymbol{z}^{(k)}) \\ \boldsymbol{v}^{(k)} & = & \dfrac{\boldsymbol{z}^{(k)}}{m_k}, \end{cases} \qquad k = 1, 2, \cdots$$

因为

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

所以 $\boldsymbol{v}^{(k)}$和$A^k\boldsymbol{v}^{(0)}$平行。

幂迭代法的收敛性

而$\max(\boldsymbol{\nu}^{(k)}) = 1$，则

$$\boldsymbol{\nu}^{(k)} = \frac{A^k\boldsymbol{\nu}^{(0)}}{\max(A^k\boldsymbol{\nu}^{(0)})}$$

$$\Rightarrow \max(A^k\boldsymbol{\nu}^{(0)}) = \prod_{i=1}^{k} m_i$$

$$\boldsymbol{\nu}^{(k)} \to \frac{\boldsymbol{x}_1}{\max \boldsymbol{x}_1}$$

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

$$m_k \to \max\left(A\frac{\boldsymbol{x}_1}{\max \boldsymbol{x}_1}\right) = \max\left(\frac{\lambda_1\boldsymbol{x}_1}{\max \boldsymbol{x}_1}\right) = \lambda_1.$$

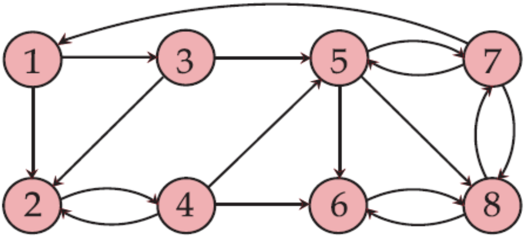
收敛速度

$$
\begin{aligned}
m_k &= \frac{\max(A^k \boldsymbol{\nu}^{(0)})}{\max(A^{k-1}\boldsymbol{\nu}^{(0)})} \\
&= \lambda_1 \frac{\max\left(\alpha_1 \boldsymbol{x}_1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \boldsymbol{x}_i\right)}{\max\left(\alpha_1 \boldsymbol{x}_1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k-1} \boldsymbol{x}_i\right)} \\
&= \lambda_1 \left[1 + O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)\right] \to \lambda_1
\end{aligned}
$$

$$
\Longrightarrow \quad \frac{|m_{k+1} - \lambda_1|}{|m_k - \lambda_1|} \approx \left|\frac{\lambda_2}{\lambda_1}\right|, \quad \text{线性收敛！}
$$

注记

幂迭代法一般用来求解主特征值。如果矩阵没有主特征值，则方法无法使用。如果 迭代初值$\boldsymbol{v}^{(0)}$不含主特征值对应的特征向量$\boldsymbol{x}_1$，即$\alpha_1 = 0$，则算法亦不适用。 更多的时候，幂迭代法是在已知主特征值的情况下用来求解主特征向量。



$$H=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix},$$

PangRank的计算

主特征值1对应的主特征向量经过迭代可得

I^0	I^1	I^2	I^3	I^4	$\cdots$	I^{60}	I^{61}
1	0	0	0	0.0287	$\cdots$	0.06	0.06
0	0.5	0.25	0.1667	0.0833	$\cdots$	0.0675	0.0675
0	0.5	0	0	0	$\cdots$	0.03	0.03
0	0	0.5	0.25	0.1667	$\cdots$	0.0675	0.0675
0	0	0.25	0.1667	0.1111	$\cdots$	0.0975	0.0975
0	0	0	0.25	0.1806	$\cdots$	0.2025	0.2025
0	0	0	0.0833	0.0972	$\cdots$	0.18	0.18
0	0	0	0.0833	0.3333	$\cdots$	0.295	0.295

这就是对应网页的PageRank值。

逆幂迭代法

如果矩阵$A \in \mathbb{R}^{n\times n}$特征值满足

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{n-1}| > |\lambda_n| > 0$$

则A^{-1}的特征值满足

$$|\lambda_n^{-1}| > |\lambda_{n-1}^{-1}| \geq \cdots \geq |\lambda_1^{-1}|$$

则有逆幂迭代法

$$\left\{\begin{array}{rcl} \boldsymbol{\nu}^{(0)} & = & \boldsymbol{\mu}/\max \boldsymbol{\mu} \\ A\boldsymbol{z}^{(k)} & = & \boldsymbol{\nu}^{(k-1)} \\ m_k & = & \max(\boldsymbol{z}^{(k)}) \\ \boldsymbol{\nu}^{(k)} & = & \boldsymbol{z}^{(k)}/m_k, \end{array}\right. \qquad k = 1, 2, \cdots .$$

逆幂法的收敛性

对逆幂迭代法

$$
\begin{aligned}
\boldsymbol{v}^{(k)} &\to \frac{\boldsymbol{x}_n}{\max \boldsymbol{x}_n}, \\
m_k &= \lambda_n^{-1}\left[1 + O\left(\left|\frac{\lambda_n}{\lambda_{n-1}}\right|^k\right)\right] \to \lambda_n^{-1}
\end{aligned}
$$

Q: 逆幂迭代法可以求最小的特征值，如何求中间的特征值？

A: 若中间的特征值与其他的特征值分离，则可通过原点位移技巧来求！

原点位移技巧

设通过分析知A有一单特征值λ_i与一个确定的数q最接近，即

$$0 < |\lambda_i - q| \ll |\lambda_j - q|, \; j \neq i.$$

而$(A - qI)^{-1}$的特征值为

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

此时$1/(\lambda_i - q)$是$(A - qI)^{-1}$的主特征值。则对$\boldsymbol{\nu}^{(0)} = \sum_{j=1}^{n} \alpha_j \boldsymbol{x}_j$进行逆幂迭代可得

$$(A - qI)^{-k} \boldsymbol{\nu}^{(0)} = \sum_{j=1}^{n} \frac{\alpha_j}{(\lambda_j - q)^k} \boldsymbol{x}_j$$

则对$A-qI$进行逆幂迭代

$$\begin{cases} \boldsymbol{\nu}^{(0)} &= \boldsymbol{\mu}/\max\boldsymbol{\mu} \\ (A-qI)\boldsymbol{z}^{(k)} &= \boldsymbol{\nu}^{(k-1)} \\ m_k &= \max(\boldsymbol{z}^{(k)}) \\ \boldsymbol{\nu}^{(k)} &= \dfrac{\boldsymbol{z}^{(k)}}{m_k}, \end{cases} \quad k=1,2,\cdots$$

可得

$$\begin{aligned} m_k &\to (\lambda_i-q)^{-1}, \\ \boldsymbol{\nu}^{(k)} &\to \frac{\boldsymbol{x}_i}{\max(\boldsymbol{x}_i)}. \end{aligned}$$

幂法求特征值的思路：

- 确定q：圆盘定理，或若$\boldsymbol{x}$为近似特征向量，取$q=\dfrac{(A\boldsymbol{x},\boldsymbol{x})}{(\boldsymbol{x},\boldsymbol{x})}$。
- 原点位移的逆幂迭代法。

定义 (Householder 变换)

设$\boldsymbol{w} \in \mathbb{R}^n$，且$\|\boldsymbol{w}\|_2 = 1$，则称

$$P = I - 2\boldsymbol{w}\boldsymbol{w}^T$$

为Householder矩阵，其表示变换称为 Householder变换。

Householder矩阵有如下性质

1. $P^T = P$，Householder矩阵为对称阵；
2. $P^TP = I$，Householder矩阵为正交阵；
3. 几何解释：镜面反射矩阵，初等反射矩阵；
4. $\|P\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$。
5. 对任意$\boldsymbol{x}$和$\boldsymbol{v}$，满足$\|\boldsymbol{x}\|_2 = \|\boldsymbol{v}\|_2$，令

$$\boldsymbol{w} = \frac{\boldsymbol{x} - \boldsymbol{v}}{\|\boldsymbol{x} - \boldsymbol{v}\|_2} \Longrightarrow P\boldsymbol{x} = \boldsymbol{v}.$$

对$\boldsymbol{x} \neq 0$，可找到Householder变换，使得

$$P\boldsymbol{x} = k\boldsymbol{e}_1$$

可取

$$\begin{aligned} \boldsymbol{u} &= \boldsymbol{x} - k\boldsymbol{e}_1, \\ \boldsymbol{w} &= \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|_2}. \end{aligned}$$

设$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T$，则

$$\boldsymbol{u} = (x_1 - k, x_2, \cdots, x_n)^T$$

由于$|k| = \|\boldsymbol{x}\|_2$。可取

$$k = -\mathrm{sgn}(x_1)\|\boldsymbol{x}\|_2$$

则有

$$\begin{aligned}
\boldsymbol{u} &= \left(x_1 + \mathbf{sgn}(x_1)\|\boldsymbol{x}\|_2, x_2, \cdots, x_n\right) \\
\|\boldsymbol{u}\|_2^2 &= 2\|\boldsymbol{x}\|_2\left(\|\boldsymbol{x}\|_2 + |x_1|\right), \\
\beta &= \left[\|\boldsymbol{x}\|_2(\|\boldsymbol{x}\|_2 + |x_1|)\right]^{-1} \\
P &= I - \frac{2\boldsymbol{u}\boldsymbol{u}^T}{\|\boldsymbol{u}\|_2^2} = I - \beta\boldsymbol{u}\boldsymbol{u}^T
\end{aligned}$$

类似可将接连几个分量变为零，设

$$\boldsymbol{x} = (x_1, \cdots, x_{j-1}, x_j, \cdots, x_k, x_{k+1}, \cdots, x_n)^T$$

欲将其变成一个从$j+1$指标到k指标都为零的向量，可以取

$$\boldsymbol{u} = (0, \cdots, 0, x_j + \mathbf{sgn}(x_j)\alpha, x_{j+1}, \cdots, x_k, 0, \cdots, 0)^T$$

Householder矩阵

其中

$$\alpha = (x_j^2 + \cdots + x_k^2)^{1/2}$$

于是

$$\|\boldsymbol{u}\|_2^2 = 2\alpha(\alpha + |x_j|)$$

由此确定的Householder矩阵为

$$P = I - \frac{2\boldsymbol{u}\boldsymbol{u}^T}{\|\boldsymbol{u}\|_2^2} = \begin{bmatrix} I_{j-1} & 0 & 0 \\ 0 & \bar{P} & 0 \\ 0 & 0 & I_{n-k} \end{bmatrix}$$

并且

$$P\boldsymbol{x} = (x_1, \cdots, x_{j-1}, -\mathbf{sgn}(x_j)\alpha, 0, \cdots, 0, x_{k+1}, \cdots, x_n)^T$$

例

求*Householder*矩阵P使得$P\boldsymbol{x} = k\boldsymbol{e}_1$，其中$\boldsymbol{x} = (3,5,1,1)^T$。

解:计算$\|\boldsymbol{x}\|_2 = 6$，则取$k = -6$，而

$$
\begin{aligned}
\boldsymbol{u} &= \boldsymbol{x} - k\boldsymbol{e}_1 = (9,5,1,1)^T, \\
\|\boldsymbol{u}\|_2^2 &= 108, \\
\beta &= [\|\boldsymbol{x}\|_2(\|\boldsymbol{x}\|_2 + |x_1|)]^{-1} = \frac{1}{54}
\end{aligned}
$$

所以相应的Householder矩阵为

$$
P = I - \beta\boldsymbol{u}\boldsymbol{u}^T = \frac{1}{54}\begin{pmatrix} -27 & -45 & -9 & -9 \\ -45 & 29 & -5 & -5 \\ -9 & -5 & 53 & -1 \\ 9 & -5 & -1 & 53 \end{pmatrix}
$$

对某个θ，记$s=\sin\theta$，$c=\cos\theta$，

$$J(\theta)=\begin{pmatrix} c & s \\ -s & c \end{pmatrix}=\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

为一正交阵。若$\mathbf{x}\in\mathbb{R}^2$,则$J\mathbf{x}$表示将向量$\mathbf{x}$顺时针旋转$\theta$角所得到的向量。

如何将一个向量(x_1,x_2)变成另一个向量(y_1,y_2)，使得$y_2=0$。

- 若$x_2=0$，则$c=1$，$s=0$;
- 若$x_2\neq 0$，$|x_2|\geq|x_1|$，考虑将向量$\mathbf{sgn}(x_2)(x_1,x_2)$旋转 $\theta\in(0,\pi)$角，变为平行于x-轴的向量，于是

$$\begin{aligned} c &= \cos\theta=\frac{\mathbf{sgn}(x_2)x_1}{(x_1^2+x_2^2)^{1/2}}, \\ s &= \sin\theta=\frac{\mathbf{sgn}(x_2)x_2}{(x_1^2+x_2^2)^{1/2}}>0, \end{aligned}$$

因而若记

$$t = c/s = \frac{x_1}{x_2}$$

则

$$s = (1+t^2)^{-\frac{1}{2}},\ c = st$$

- 若$x_2 \neq 0$，$|x_2| < |x_1|$，考虑将向量$\mathbf{sgn}(x_1)(x_1, x_2)$旋转$\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$角，变为平行于$x$-轴的向量，于是

$$c = \cos\theta = \frac{\mathbf{sgn}(x_1)x_1}{(x_1^2 + x_2^2)^{1/2}} > 0,\ s = \sin\theta = \frac{\mathbf{sgn}(x_1)x_2}{(x_1^2 + x_2^2)^{1/2}},$$

因而若记

$$t = s/c = \frac{x_2}{x_1}$$

则

$$c = (1+t^2)^{-\frac{1}{2}},\ s = ct$$

现在将它推广到n维情况

$$J(i,k,\theta)=\begin{array}{cc}\begin{bmatrix} 1 &&&&&&&&&& \\ & \ddots &&&&&&&&& \\ && 1 &&&&&&&& \\ &&& c && \cdots && s &&& \\ &&&& 1 &&&&&& \\ &&& \vdots && \ddots && \vdots &&& \\ &&&&&& 1 &&&& \\ &&& -s && \cdots && c &&& \\ &&&&&&&& 1 && \\ &&&&&&&&& \ddots & \\ &&&&&&&&&& 1 \end{bmatrix} & \begin{array}{c} \\ \\ \\ i \\ \\ \\ \\ k \\ \\ \\ \\ \end{array}\end{array}$$

称为Givens矩阵或Givens变换。显然其为正交矩阵，
设$\boldsymbol{x}\in\mathbb{R}^n$，$\boldsymbol{y}=J(i,k,\theta)\boldsymbol{x}$，则$\boldsymbol{y}$的分量为

$$(y_i,y_k)^T=J(x_i,x_k)^T,\quad y_j=x_j(j\neq i,k)$$

于是当$x_i^2+x_k^2\neq 0$时，要将第k个分量变为零，可以左乘$J(i,k,\theta)$，其中c和s确定如下

1. 若$x_k=0$，则$c=1, s=0$;
2. 若$x_k\neq 0$，$|x_k|\geq|x_i|$，

$$t=c/s=\frac{x_i}{x_k},\ s=(1+t^2)^{-\frac{1}{2}},\ c=st$$

3. 若$x_k\neq 0, |x_k|<|x_i|$，

$$t=s/c=\frac{x_k}{x_i},\ c=(1+t^2)^{-\frac{1}{2}},\ s=ct$$

注记

$J(i,k,\theta)\boldsymbol{x}$只改变$\boldsymbol{x}$的第$i,k$分量， $J(i,k,\theta)A$只改变A 的第i,k行。同理$AJ(i,k,\theta)^T$只改变A的第i,k列。$J(i,k,\theta)AJ(i,k,\theta)^T$只改变$A$的第 i,k行和第i,k列。

例

求Givens矩阵$J(2,4,\theta)$使得对$\boldsymbol{x}=(1,2,3,4)^T\in\mathbb{R}^4$ 有$J(2,4,\theta)\boldsymbol{x}$的第4个分量为零。

解：取$i=2,k=4$，则

$$t=\frac{1}{2}, s=\sin\theta=\frac{2}{\sqrt{5}}, c=\cos\theta=\frac{1}{\sqrt{5}},$$

所以

$$J(2,4,\theta)=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & \frac{1}{\sqrt{5}} & 0 & \frac{2}{\sqrt{5}}\\ 0 & 0 & 1 & 0\\ 0 & -\frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}}\end{bmatrix}, J(2,4,\theta)\boldsymbol{x}=\begin{bmatrix}1\\ 2\sqrt{5}\\ 3\\ 0\end{bmatrix}$$

定理 (QR分解定理)

设$A \in \mathbb{R}^{n\times n}$，则存在正交阵$Q$，使得$A = QR$，且当$R$的对角元均为正时，分解唯一。

注记

实际上按照*Givens*变换或者*Householder*变换方法作出的分解$A = QR$中R的元素不一定是正的。设 $R = [r_{ij}]$，只要令

$$D = \mathrm{diag}\left(\frac{r_{11}}{|r_{11}|}, \cdots, \frac{r_{nn}}{|r_{nn}|}\right),$$

则令$\overline{Q} = QD$为正交阵，$\overline{R} = D^{-1}R$为对角元是$|r_{ii}|$的上三角阵。$A = \overline{Q}\overline{R}$便是唯一的*QR*分解。

例

求矩阵$A=\begin{bmatrix}4&4&0\\3&3&-1\\0&1&1\end{bmatrix}$的$QR$分解，使 R的对角线元素为正。

解:先用Givens变换，A的第1列为$(4,3,0)^T$，所以

$$J(1,2,\theta_{12})=\begin{bmatrix}\frac{4}{5}&\frac{3}{5}&0\\-\frac{3}{5}&\frac{4}{5}&0\\0&0&1\end{bmatrix},J(1,2,\theta_{12})A=\begin{bmatrix}5&5&-\frac{3}{5}\\0&0&-\frac{4}{5}\\0&1&1\end{bmatrix}$$

则$P_1=J(1,2,\theta)$，P_1A的第2列后两个元素为$(0,1)$，因此

$$P_2=J(2,3,\theta_{23})=\begin{bmatrix}1&0&0\\0&0&1\\0&-1&0\end{bmatrix},P_2P_1A=\begin{bmatrix}5&5&-\frac{3}{5}\\0&1&1\\0&0&\frac{4}{5}\end{bmatrix}$$

此时$R=P_2P_1A$是一个上三角阵，且对角线元素为正，则

$$Q=(P_2P_1)^T=\begin{bmatrix}\frac{4}{5} & 0 & \frac{3}{5}\\ \frac{3}{5} & 0 & -\frac{4}{5}\\ 0 & 1 & 0\end{bmatrix}$$

满足$A=QR$.
也可用Householder变换，对A的第1列$\boldsymbol{a}_1$，有

$$k=-5,\boldsymbol{u}=(4+5,3,0)^T=(9,3,0)^T,\|\boldsymbol{u}\|_2^2=90,\beta=\frac{1}{45}$$

则

$$P_1=I-\beta\boldsymbol{u}\boldsymbol{u}^T=\begin{bmatrix}-\frac{4}{5} & -\frac{3}{5} & 0\\ -\frac{3}{5} & \frac{4}{5} & 0\\ 0 & 0 & 1\end{bmatrix},P_1A=\begin{bmatrix}-5 & -5 & \frac{3}{5}\\ 0 & 0 & -\frac{4}{5}\\ 0 & 1 & 1\end{bmatrix},$$

则P_1A的第1列对角线以下元素为0。

对$\begin{bmatrix}0 & -\frac{4}{5}\\ 1 & 1\end{bmatrix}$的第1列做Householder变换有 $\overline{P}=\begin{bmatrix}0 & -1\\ -1 & 0\end{bmatrix}$ 所以

$$P_2=\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & -1\\ 0 & -1 & 0\end{bmatrix},P_2P_1A=\begin{bmatrix}-5 & -5 & \frac{3}{5}\\ 0 & -1 & -1\\ 0 & 0 & \frac{4}{5}\end{bmatrix}$$

$R=P_2P_1A$是上三角阵，而

$$Q=(P_2P_1)^T=\begin{bmatrix}-\frac{4}{5} & 0 & \frac{3}{5}\\ -\frac{3}{5} & 0 & -\frac{4}{5}\\ 0 & -1 & 0\end{bmatrix}$$

满足$A=QR$，但是R的对角线元素并非都是正的，
令$\overline{D}=diag(-1,-1,1)$，则

$$\overline{Q}=\begin{bmatrix}\frac{4}{5} & 0 & \frac{3}{5}\\ \frac{3}{5} & 0 & -\frac{4}{5}\\ 0 & 1 & 0\end{bmatrix},\overline{R}=\overline{D}^{-1}R=\begin{bmatrix}5 & 5 & -\frac{3}{5}\\ 0 & 1 & 1\\ 0 & 0 & \frac{4}{5}\end{bmatrix}$$

定理 (谱分解)

设$A \in \mathbb{R}^{n\times n}$可对角化，即存在非奇异的$X$使得

$$X^{-1}AX = D, \quad D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

当且仅当A有n个线性无关的特征向量，令

$$X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]$$

则

$$AX = XD \Longleftrightarrow A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i, \quad 1 \le i \le n.$$

定理 (Schur 分解)

对任意的矩阵$A \in \mathbb{R}^{n\times n}$，存在一个酉矩阵

$$U \in \mathbb{C}^{n\times n}: \quad U^H U = UU^H = I$$

使得

$$U^H A U = \begin{pmatrix} * & * & * \\ & * & * \\ & & * \end{pmatrix} = R \in \mathbb{C}^{n\times n}$$

酉相似变换不畏惧舍入误差，不放大A中已有误差：

$$U^H(A+E)U = U^H A U + U^H E U$$

而

$$\|U^H E U\|_2 = \|E\|_2$$

而对一般的相似变换矩阵X：

$$\begin{aligned} X^{-1}(A+E)X &= X^{-1}AX + X^{-1}EX \\ \|X^{-1}EX\| &\le \|X^{-1}\|\|E\|\|X\|. \end{aligned}$$

若X非常病态，即

$$\|X\|\|X^{-1}\| \gg 1 \Longrightarrow \|X^{-1}EX\| \gg \|E\|$$

一般的相似变换具有潜在的不稳定性，应该尽量避免。

但是复矩阵U有不少缺点：

- 计算量和存储量显著增加，
- 会破坏问题本身的性质和结构。

定理 (实Schur分解定理)

设$A \in \mathbb{R}^{n\times n}$，则存在正交阵 $Q \in \mathbb{R}^{n\times n}$，使得

$$Q^TAQ = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ & R_{22} & \cdots & R_{2m} \\ & & \ddots & \vdots \\ & & & R_{mm} \end{pmatrix}$$

其中每个R_{ii}是1×1或2×2的。若是1×1的，其元素就是A的特征值。若是2×2的，R_{ii}的特征值是A的一对共轭的复特征值。

通过正交相似变换可以把矩阵A化为一种特殊的形式，它更接近实Schur分解的形式，对这种 特殊形式的矩阵用数值方法求特征值会减少计算量。

定义 (Hessenberg矩阵)

若$B = (b_{ij}) \in \mathbb{R}^{n\times n}$，当$i > j + 1$时有

$$b_{ij} = 0$$

则 称B为上Hessenberg矩阵。而且若B的 次对角元素

$$b_{i,i-1} \neq 0, i = 2, \cdots, n,$$

则称其为 不可约的上Hessenberg矩阵。

若$A \in \mathbb{R}^{n\times n}$，它的第$j$列元素为

$$\boldsymbol{a}^{(j)} = (a_{1j}, \cdots, a_{jj}, a_{j+1,j}, \cdots, a_{nj})^T$$

可以选择$n-j$阶的Householder阵$\bar{P}_j$，使得

$$\bar{P}_j \begin{pmatrix} a_{j+1,j} \\ a_{j+2,j} \\ \vdots \\ a_{nj} \end{pmatrix} = \begin{pmatrix} * \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

于是有n阶Householder阵P_j

$$P_j = \mathbf{diag}(I_j, \bar{P}_j) = I - \frac{2\boldsymbol{u}_j\boldsymbol{u}_j^T}{\|\boldsymbol{u}_j\|_2^2}$$

使得

$$P_j\boldsymbol{a}^{(j)} = (a_{1j}, \cdots, a_{jj}, -\mathbf{sgn}(a_{j+1,j})\alpha, 0, \cdots, 0)^T$$

其中

$$\begin{aligned} \alpha &= (a_{j+1,j}^2 + \cdots + a_{nj}^2)^{1/2} \\ \boldsymbol{u}_j &= (0, \cdots, 0, a_{j+1,j} + \mathbf{sgn}(a_{j+1,j})\alpha, a_{j+2,j}, \cdots, a_{nj})^T \end{aligned}$$

因而如此逐次左乘$P_1, P_2, \cdots$，可将A次对角线以下的元素都化为零。为求A的特征值，应对A作相似变换。第一次作$P_1AP_1^{-1} = P_1AP_1$，其形式如

$$
\begin{aligned}
(P_1A)P_1 &= \begin{pmatrix} * & * & * & \cdots & * & * \\ * & * & * & \cdots & * & * \\ 0 & * & * & \cdots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & * & * & \cdots & * & * \\ 0 & * & * & \cdots & * & * \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & * & * & \cdots & * & * \\ 0 & * & * & \cdots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & * & * & \cdots & * & * \\ 0 & * & * & \cdots & * & * \end{pmatrix} \\
&= \begin{pmatrix} * & * & * & \cdots & * & * \\ * & * & * & \cdots & * & * \\ 0 & * & * & \cdots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & * & * & \cdots & * & * \\ 0 & * & * & \cdots & * & * \end{pmatrix}
\end{aligned}
$$

设$k-1$步Householder正交相似变换后

$$
\begin{aligned}
A^{(k)} &= (P_1 \cdots P_{k-1})^T A (P_1 \cdots P_{k-1}) \\
&= \begin{pmatrix} B_{11} & B_{12} & c_1 & c_2 \\ B_{21} & B_{22} & c_3 & c_4 \\ 0 & b_1 & d_1 & d_2 \\ 0 & b_2 & d_3 & d_4 \end{pmatrix}
\end{aligned}
$$

其中

$$
\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}
$$

为上Hessenberg矩阵。
设$\bar{P}_k$为$n-k$阶Householder阵，使得

$$
\bar{P}_k \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \hat{b}_1 \\ 0 \end{pmatrix}
$$

则n阶Householder矩阵$P_k = \mathbf{diag}(I, \bar{P}_k)$使

$$A^{(k+1)} = P_k^T A^{(k)} P_k$$

形式为

$$\begin{aligned}
A^{(k+1)} &= (P_1 \cdots P_k)^T A (P_1 \cdots P_k) \\
&= \begin{pmatrix} B_{11} & B_{12} & \hat{c}_1 & \hat{c}_2 \\ B_{21} & B_{22} & \hat{c}_3 & \hat{c}_4 \\ 0 & \hat{b}_1 & \hat{d}_1 & \hat{d}_2 \\ 0 & 0 & \hat{d}_3 & \hat{d}_4 \end{pmatrix} \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix} \\
&= \begin{pmatrix} B_{11} & B_{12} & \tilde{c}_1 & \tilde{c}_2 \\ B_{21} & B_{22} & \tilde{c}_3 & \tilde{c}_4 \\ 0 & \hat{b}_1 & \tilde{d}_1 & \tilde{d}_2 \\ 0 & 0 & \tilde{d}_3 & \tilde{d}_4 \end{pmatrix}
\end{aligned}$$

此过程可以一直进行至$n-2$步即将A通过相似变换化为上Hessenberg矩阵

$$A^{(n-1)} = \begin{pmatrix} a_{11}^{n-1} & a_{12}^{n-1} & \cdots & a_{1,n-1}^{n-1} & a_{1n}^{n-1} \\ a_{21}^{n-1} & a_{22}^{n-1} & \cdots & a_{2,n-1}^{n-1} & a_{2n}^{n-1} \\ & a_{32}^{n-1} & \cdots & a_{3,n-1}^{n-1} & a_{3n}^{n-1} \\ & & \ddots & \vdots & \vdots \\ & & & a_{n,n-1}^{n-1} & a_{nn}^{n-1} \end{pmatrix}$$

如果A是对称阵，则$A^{(n-1)}$是一个对称的三对角阵。

定理

对任意的矩阵$A \in \mathbb{R}^{n\times n}$，存在正交矩阵$Q$使

$$B = Q^T A Q$$

为一个上*Hessenberg*矩阵。

用Householder矩阵约化矩阵$A \in \mathbb{R}^{n\times n}$为上Hessenberg形式的算法可以写成：

对$j = 1, 2, \cdots, n-2$,

确定$n-j$阶Householder阵$\bar{P}_j$，使得

$$\bar{P}_j \begin{pmatrix} a_{j+1,j} \\ a_{j+2,j} \\ \vdots \\ a_{nj} \end{pmatrix} = \begin{pmatrix} * \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$P_j = \mathbf{diag}(I_j, \bar{P}_j)$$

$$P_j^T A P_j \Rightarrow A$$

此算法计算量约为$\frac{5n^3}{3}$。

例

把矩阵

$$A = \begin{pmatrix} 1 & 5 & 7 \\ 3 & 0 & 6 \\ 4 & 3 & 1 \end{pmatrix}$$

变为上*Hessenberg*形式。

解：选Householder阵$P \in R^{3\times 3}$使得

$$P(1,3,4)^T = (1,*,0)^T.$$

则取$\alpha = (3^2+4^2)^{\frac{1}{2}} = 5$，$\boldsymbol{u} = (0,8,4)^T$，$\|\boldsymbol{u}\|_2^2 = 80$,

$$P = I - \frac{\boldsymbol{u}\boldsymbol{u}^T}{40} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -0.6 & -0.8 \\ 0 & -0.8 & 0.6 \end{pmatrix}$$

$$PAP = \begin{pmatrix} 1.00 & -8.60 & 0.20 \\ -5.00 & 4.96 & -0.72 \\ 0 & 2.28 & -3.96 \end{pmatrix}$$

也可对矩阵做Givens变换，对A的第一列构造Givens矩阵$J(2,3,\theta)$使得$J(2,3,\theta)(1,3,4)^T=(1,*,0)^T$，则

$$J(2,3,\theta)=\begin{pmatrix}1&0&0\\0&0.6&0.8\\0&-0.8&0.6\end{pmatrix},J(2,3,\theta)\begin{pmatrix}1\\3\\4\end{pmatrix}=\begin{pmatrix}1\\5\\0\end{pmatrix}$$

令$Q=[J(2,3,\theta)]^T$，则

$$Q=\begin{pmatrix}1&0&0\\0&0.6&-0.8\\0&0.8&0.6\end{pmatrix}$$

有

$$Q^TAQ=\begin{pmatrix}1&8.6&0.2\\5&4.96&0.72\\0&-2.28&-3.96\end{pmatrix}$$

实矩阵A可分解为

$$A = QR$$

其中Q为正交阵，R为上三角阵。若规定R 的对角元为正，则QR 分解是唯一的。
令

$$B = RQ \Rightarrow B = Q^T AQ = Q^{-1}AQ,$$

B和A相似，有相同的特征值。如果对B继续做如上运算，则有
基本QR算法

$$\begin{cases} A_1 &= A, \\ A_k &= Q_k R_k, \quad (A_k\text{的QR分解}) \\ A_{k+1} &= R_k Q_k, \quad \text{逆序相乘} \end{cases}$$

定理

QR算法产生的序列$\{A_k\}$满足

1. $A_{k+1} = Q_k^T A_k Q_k$；
2. $A_{k+1} = \bar{Q}_k^T A \bar{Q}_k$,其中$\bar{Q}_k = Q_1 Q_2 \cdots Q_k$；
3. $A^k = \bar{Q}_k \bar{R}_k$，其中$\bar{R}_k = R_k \cdots R_2 R_1$。

对上Hessenberg 阵H，选择θ_i，用Givens矩阵$J(i,i+1,\theta_i)$左乘H，将第$i+1$行第i列的元素变为0。

$$H=\begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1i} & \cdots & h_{1,n-1} & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2i} & \cdots & h_{2,n-1} & h_{2n} \\ & \ddots & \ddots & \vdots & & \vdots & \vdots \\ & & \ddots & h_{ii} & \cdots & h_{i,n-1} & h_{in} \\ & & & h_{i+1,i} & \cdots & h_{i+1,n-1} & h_{i+1,n} \\ & & & & \ddots & \vdots & \vdots \\ & & & & & h_{n,n-1} & h_{nn} \end{pmatrix}$$

即

$$J(\theta_i)\begin{pmatrix} h_{ii} \\ h_{i+1,i} \end{pmatrix}=\begin{pmatrix} * \\ 0 \end{pmatrix}$$

如此$n-1$次操作之后得到上三角阵R

$$U^TH=R,\ U^T=J(n-1,n,\theta_{n-1})\cdots J(1,2,\theta_1)$$

注记

容易验证H为一上*Hessenberg*矩阵，由于$H = UR$，作*QR*迭代时，下一步计算 $H_2 = RU$，容易验证RU是一个上*Hessenberg*矩阵。因而*QR*算法保持*H*的上*Hessenberg*结构。

同样考虑上Hessenberg矩阵，假设每次QR迭代中产生的H_k都是不可约的，不然

$$H_k = \begin{pmatrix} H_{11} & H_{12} \\ 0 & H_{22} \end{pmatrix}$$

可分为两个小规模矩阵的特征值问题。
假设H的特征值排列成

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

则H_k的第j个次对角元收敛速度决定于$\left|\frac{\lambda_{j+1}}{\lambda_j}\right|$，其越小，收敛越快。

因而可以借鉴幂法的想法，考虑$H-\mu I$的特征值，使得将其特征值重新排列后

$$|\lambda_1-\mu| \geq |\lambda_2-\mu| \geq \cdots \geq |\lambda_n-\mu|$$

$\left|\dfrac{\lambda_{j+1}-\mu}{\lambda_j-\mu}\right|$尽可能小。

基于此思想，有如下的原点位移方法

$$\begin{cases} H_k-\mu I=U_kR_k, (H_k-\mu I\text{的QR分解}) \\ \\ H_{k+1}=R_kU_k+\mu I,\ k=1,2,\cdots \end{cases}$$

若μ为H的一个特征值，则

$$R_k=U_k^T(H_k-\mu I)$$

为奇异阵。若H_k不可约，则$r_{nn}^{(k)}=0$，从而H_{k+1}的最后一行为

$$(0,\cdots,0,\mu)$$

于是求得其一个特征值μ。

1. 第1种原点位移的QR迭代

$$\begin{cases} \mu_k = h_{nn}^{(k)}, \\ H_k - \mu_k I = U_k R_k, (H_k - \mu I\text{的QR分解}) \\ H_{k+1} = R_k U_k + \mu_k I,\ k = 1, 2, \cdots \end{cases}$$

收敛判别准则可取如下两者之一

$$|h_{n,n-1}^{(k)}| \leq \epsilon ||H_1||_\infty$$

$$|h_{n,n-1}^{(k)}| \leq \epsilon(|h_{nn}^{(k)}| + |h_{n-1,n-1}^{(k)}|)$$

2. 第2种原点位移的QR迭代选μ_k为

$$\begin{pmatrix} h_{n-1,n-1}^{(k)} & h_{n-1,n}^{(k)} \\ h_{n,n-1}^{(k)} & h_{n,n}^{(k)} \end{pmatrix}$$

的特征值中接近$h_{nn}^{(k)}$中的一个。

用第2种原点位移的QR迭代求解矩阵

$$A=\begin{bmatrix}3&1&0\\1&3&1\\0&1&3\end{bmatrix}=\begin{bmatrix}a_1^{(1)}&b_2^{(1)}&0\\b_2^{(1)}&a_2^{(1)}&b_3^{(1)}\\0&b_3^{(1)}&a_3^{(1)}\end{bmatrix}$$

的特征值。

解：A已经是一个上Hessenberg矩阵的形式，首先求出

$$\begin{bmatrix}a_2^{(1)}&b_3^{(1)}\\b_3^{(1)}&a_3^{(1)}\end{bmatrix}=\begin{bmatrix}3&1\\1&3\end{bmatrix}$$

的特征值$\lambda_1=4,\lambda_2=2$。则两个特征值和$a_3^{(1)}$的距离一样，选取$\mu_1=2$来做位移，则

$$A_2=A-\mu_1I=\begin{bmatrix}d_1&b_2^{(1)}&0\\b_2^{(1)}&d_2&b_3^{(1)}\\0&b_3^{(1)}&d_3\end{bmatrix}=\begin{bmatrix}1&1&0\\1&1&1\\0&1&1\end{bmatrix}$$

用Givens变换作QR分解，首先把$b_2^{(1)}$变为零，则Givens矩阵

$$J(1,2,\theta)A_2=\begin{bmatrix}\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0\\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 1 & 0\\ 1 & 1 & 1\\ 0 & 1 & 1\end{bmatrix}=\begin{bmatrix}\sqrt{2} & \sqrt{2} & \frac{\sqrt{2}}{2}\\ 0 & 0 & \frac{\sqrt{2}}{2}\\ 0 & 1 & 1\end{bmatrix}=A_2^{(1)}$$

下一步的Givens变换

$$J(2,3,\theta)A_2^{(1)}=\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & -1 & 0\end{bmatrix}\begin{bmatrix}\sqrt{2} & \sqrt{2} & \frac{\sqrt{2}}{2}\\ 0 & 0 & \frac{\sqrt{2}}{2}\\ 0 & 1 & 1\end{bmatrix}=\begin{bmatrix}\sqrt{2} & \sqrt{2} & \frac{\sqrt{2}}{2}\\ 0 & 1 & 1\\ 0 & 0 & -\frac{\sqrt{2}}{2}\end{bmatrix}$$

则 $R^{(1)}=J(2,3,\theta)A_2^{(1)}$，而

$$Q^{(1)}=J(1,2,\theta)^TJ(2,3,\theta)^T=\begin{bmatrix}\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2}\\ 0 & 1 & 0\end{bmatrix}$$

所以

$$A^{(2)} = R^{(1)}Q^{(1)} = \begin{bmatrix} 2 & \frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & 1 & -\frac{\sqrt{2}}{2} \\ 0 & -\frac{\sqrt{2}}{2} & 0 \end{bmatrix}$$

上述过程完成一步QR迭代。下一步迭代先确定位移因子为

$$\begin{bmatrix} a_2^{(2)} & b_3^{(2)} \\ b_3^{(2)} & a_3^{(2)} \end{bmatrix} = \begin{bmatrix} 1 & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & 0 \end{bmatrix}$$

的特征值$\lambda_2 = \frac{1}{2} \pm \frac{\sqrt{3}}{2}$，所以取$\mu_2 = \frac{1}{2} - \frac{\sqrt{3}}{2}$为最靠近$a_3^{(2)} = 0$的特征值，再做一次QR迭代有

$$A^{(3)} = \begin{bmatrix} 2.6720277 & 0.37597448 & 0 \\ 0.37597448 & 1.4736080 & 0.030396964 \\ 0 & 0.030396964 & -0.047559530 \end{bmatrix}$$

此时$b_3^{(3)}=0.030396964$可以认为很小，接近于零。
则$\lambda_3\approx 1.5864151=\mu_1+\lambda_2=2+\frac{1-\sqrt{3}}{2}$。删除第3行和第三列有

$$B^{(3)}=\begin{bmatrix}2.6720277 & 0.37597448\\ 0.37597448 & 1.4736080\end{bmatrix}$$

它的特征值为$\mu_1=2.7802140,\mu_2=1.3654218$。把位移加上可得

$$\lambda_1\approx 4.4141886,\ \ \lambda_2\approx 2.9993964$$

而矩阵A的特征值的准确值为

$$\lambda_1=4.41420,\ \ \lambda_2=3.00000,\ \ \lambda_3=1.58579$$

仅仅两次迭代就有4位有效数字的精度。

把矩阵化为上Hessenberg矩阵的形式计算量为$O(n^3)$，而接下来的用Givens 变换对矩阵进行 QR分解的计算量为$O(n^2)$。对称阵可以化成三对角阵。所以预先化成上Hessenberg矩阵的优点：

1. 对一般的矩阵，每步QR迭代的工作量为$O(n^2)$，而实对称矩阵为$O(n)$。
2. 矩阵接近于上三角阵，只需要几步QR迭代。
3. 若第一次对角线上有零元素，则Hessenberg矩阵为分块的，可以将此问题分解为规模更小的问题。

由此可见QR迭代可分为两种情况

$$
\begin{array}{rcl}
\text{对称阵} & \longrightarrow & \text{三对角阵} \longrightarrow \text{对角阵} \\
\text{一般矩阵} & \longrightarrow & \text{Hessenberg矩阵} \longrightarrow \text{上三角阵}
\end{array}
$$

- 对实对称矩阵若只计算特征值工作量为$4n^3/3$，如果还要计算特征向量则还需$9n^3$的计算量。
- 对一般的矩阵，只计算特征值的计算量为$10n^3$，如果还需计算特征向量，则工作量大约为$25n^3$。

QR迭代是计算矩阵全部特征值及特征向量的有效方法，但是它有一定的缺点，即在约化和后续迭代的过程中必须对整个矩阵进行操作。

1. 当n很大时，计算量太大。
2. 当n很大时，如果只求几个特征值和特征向量，方法的优势不明显。
3. 矩阵为大型稀疏矩阵时，由于相似变换产生了许多要存储的非零元素，方法的存储量太大。

对称矩阵特征值问题

设$A \in \mathbb{R}^{n\times n}$，$A = A^T$为对称阵，则$A$个特征值都是实数，可排列为

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$$

并可找到对应的单位特征向量构成正交向量组

$$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$$

若记

$$\begin{aligned} \Lambda &= \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n), \\ U &= (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) \end{aligned}$$

则

$$AU = U\Lambda, \qquad \Lambda = U^T A U$$

并有
$$\lambda_i = \frac{(A\boldsymbol{x}_i, \boldsymbol{x}_i)}{(\boldsymbol{x}_i, \boldsymbol{x}_i)} = \min_{\boldsymbol{x}\in \mathrm{span}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_i)} \frac{(A\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}$$
$$= \max_{\dim W = i} \min_{\boldsymbol{x}\in W, \boldsymbol{x}\neq \boldsymbol{0}} \frac{(A\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})} = \max_{\boldsymbol{x}\in \mathrm{span}(\boldsymbol{x}_i,\cdots,\boldsymbol{x}_n)} \frac{(A\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}$$
$$= \min_{\dim W = n-i+1} \max_{\substack{x\in W \\ \boldsymbol{x}\neq \boldsymbol{0}}} \frac{(A\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}$$

$$\Longrightarrow \lambda_n \leq \frac{(A\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})} \leq \lambda_1, \ \forall \boldsymbol{x} \in \mathbb{R}^n, \quad \boldsymbol{x} \neq \boldsymbol{0}.$$

定义 (Rayleigh商)

$(A\boldsymbol{x}, \boldsymbol{x})/(\boldsymbol{x}, \boldsymbol{x})$称为*Rayleigh商*,

$$\lambda_1 = \max_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{(A\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}, \quad \lambda_n = \min_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{(A\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}.$$

设对称矩阵A的特征值满足$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$，则可以用幂迭代法求$\lambda_1$，但是

$$m_k = \lambda_1[1 + O(\left|\frac{\lambda_2}{\lambda_1}\right|^k)]$$

如果幂迭代法的初始向量为

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

其中$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$是规范正交的特征向量，$\boldsymbol{v}^{(k)}$的Rayleigh商为

$$\begin{aligned} R(\boldsymbol{v}^{(k)}) &= \frac{(A\boldsymbol{v}^{(k)}, \boldsymbol{v}^{(k)})}{(\boldsymbol{v}^{(k)}, \boldsymbol{v}^{(k)})} = \frac{(A^{k+1}\boldsymbol{v}^{(0)}, A^k\boldsymbol{v}^{(0)})}{(A^k\boldsymbol{v}^{(0)}, A^k\boldsymbol{v}^{(0)})} \\ &= \frac{\sum_{j=1}^n \alpha_j^2 \lambda_j^{2k+1}}{\sum_{j=1}^n \alpha_j^2 \lambda_j^{2k}} = \lambda_1[1 + O(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k})]. \end{aligned}$$

从而$R(\boldsymbol{v}^{(k)})$的收敛速度要比m_k快，利用$R(\boldsymbol{v}^{(k)})$的方法称为Rayleigh商加速迭代方法。

注记

*Rayleigh*商$R(\boldsymbol{x})$使得$\|(A-\lambda I)\boldsymbol{x}\|_2$达到最小。所以如果知道$\boldsymbol{v}$为近似特征向量，则可以用$R(\boldsymbol{v})$作为特征值的一个较好的近似；反之若$\mu$为近似特征值，则可以用反幂迭代解$(A-\mu I)\boldsymbol{y}=\boldsymbol{v}$得到特征向量的近似。

Rayleigh商迭代算法

1. 给定$\boldsymbol{v}^{(0)}$满足于$\|\boldsymbol{v}^{(0)}\|=1$,
2. 对$k=0,1,\cdots$,

$$\mu_k=R(\boldsymbol{v}^{(k)}),$$

   解$(A-\mu I)\boldsymbol{y}^{(k+1)}=\boldsymbol{v}^{(k)}$得到$\boldsymbol{y}^{(k+1)}$。

3.

$$\boldsymbol{v}^{(k+1)}=\frac{\boldsymbol{y}^{(k+1)}}{\|\boldsymbol{y}^{(k+1)}\|_2}.$$

Jacobi 方法

Jacobi 方法利用实对称矩阵可通过正交相似变换约化为对角阵，通过一系列的平面旋转变换（Givens变换）逐步约化实对称矩阵，使其非对角元收敛到零。和对称矩阵的 QR 方法不同，Jacobi 方法的收敛速度很慢，但是它编程简单，容易实现并行，仍受到重视。

设 $c=\cos\theta, s=\sin\theta$，把 Givens 矩阵称为 Jacobi 旋转矩阵，即：

$$J(p,q,\theta)=\begin{array}{cc}\begin{bmatrix}1 &&&&&&&& \\ &\ddots&&&&&&& \\ &&1&&&&&& \\ &&&c&\cdots&s&&& \\ &&&\vdots&\ddots&\vdots&&& \\ &&&-s&\cdots&c&&& \\ &&&&&&1&& \\ &&&&&&&\ddots& \\ &&&&&&&&1\end{bmatrix} & \begin{array}{c} \\ \\ \\ p \\ \\ q \\ \\ \\ \\ \end{array}\end{array}$$

对称矩阵 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$，其非对角元的平方和记为

$$\mathrm{off}(A)=\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}a_{ij}^2=\|A\|_F^2-\sum_{i=1}^{n}a_{ii}^2.$$

$$令\ \ B=J(p,q,\theta)^TAJ(p,q,\theta).$$

F-范数在正交变换下保持不变，而 B 只有 p,q 行和列发生变化，则

$$\begin{aligned}
\mathrm{off}(B) &= \|B\|_F^2-\sum_{i=1}^{n}b_{ii}^2\\
&= \|A\|_F^2-\sum_{i\neq p,q}b_{ii}^2-(b_{pp}^2+b_{qq}^2)\\
&= \|A\|_F^2-\sum_{i\neq p,q}a_{ii}^2-(a_{pp}^2+2a_{pq}^2+a_{qq}^2)\\
&= \|A\|_F^2-\sum_{i=1}^{n}a_{ii}^2-2a_{pq}^2\\
&= \mathrm{off}(A)-2a_{pq}^2<\mathrm{off}(A).
\end{aligned}$$

因此，实对称矩阵通过Jacobi旋转矩阵做相似变换后，其非对角线部分会“衰减”。
由 Jacobi旋转矩阵的性质，确定 c, s 使得

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix}^T \begin{bmatrix} a_{pp} & a_{pq} \\ a_{pq} & a_{qq} \end{bmatrix} \begin{bmatrix} c & s \\ -s & c \end{bmatrix} = \begin{bmatrix} b_{pp} & 0 \\ 0 & b_{qq} \end{bmatrix}$$

由上面的矩阵关系式可得

$$0 = b_{pq} = a_{pq}(c^2 - s^2) + (a_{pp} - a_{qq})cs.$$

记

$$\tau = \frac{a_{pp} - aqq}{2a_{pq}}, \quad t = \frac{s}{c},$$

则 t 满足二次方程

$$t^2 + 2\tau t - 1 = 0.$$

取绝对值较小的根

$$t = \frac{\operatorname{sgn}(\tau)}{|\tau| + \sqrt{1 + \tau^2}}$$

利用关系式

$$\frac{s}{c} = t, \quad s^2 + c^2 = 1,$$

可得

$$c = \frac{1}{\sqrt{1+t^2}}, \quad s = ct.$$

选取较小的根是为了保证 $|\theta| \le \pi/4$，因为

$$\|B - A\|_F^2 = 4(1-c) \sum_{i \neq p,q} (a_{ip}^2 + a_{iq}^2) + 2a_{pq}^2/c^2$$

希望 c 尽可能大，亦即 t 尽可能小。

算法

经典 *Jacobi* 方法：找到 使得 $|a_{pq}|, p \neq q$ 最大的 p, q，利用单个 *Jacobi* 旋转使得 $a_{pq} = 0$，重复这个过程直到 $\mathrm{off}(A)$ 充分小。

实对称矩阵 $A \in \mathbb{R}^{n\times n}$ 的非对角元元素个数的一半为 $N = \frac{n(n-1)}{2}$， N 次 Jacobi 旋转称为一个 sweep. 因为在每次迭代中， a_{pq} 是绝对值最大的非对角元，因此

$$\mathrm{off}(A) \le 2Na_{pq}^2.$$

而

$$\mathrm{off}(B) = \mathrm{off}(A) - 2a_{pq}^2$$

我们有

$$\mathrm{off}(B) \le \left(1 - \frac{1}{N}\right) \mathrm{off}(A).$$

方法是线性收敛的。但是可以证明对充分大的 k，存在常数 C 使得

$$\mathrm{off}(A^{(k+N)}) \le C * \mathrm{off}(A^{(k)})^2,$$

此处 $A^{(k)}$ 是 k 次 Jacobi 迭代后的矩阵，这说明经典的 Jacobi 算法二阶收敛。且在实际计算中，只需 $\log n$ 次 sweep。