

第四章 代数特征值问题

殷东生

dyin@math.tsinghua.edu.cn

清华大学数学科学系

2018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 $\mathbf{x} \in \mathbb{C}^n$ 和 $\lambda \in \mathbb{C}$, 满足

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

称 λ 为 特征值, \mathbf{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

美国数学家斯特兰 (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

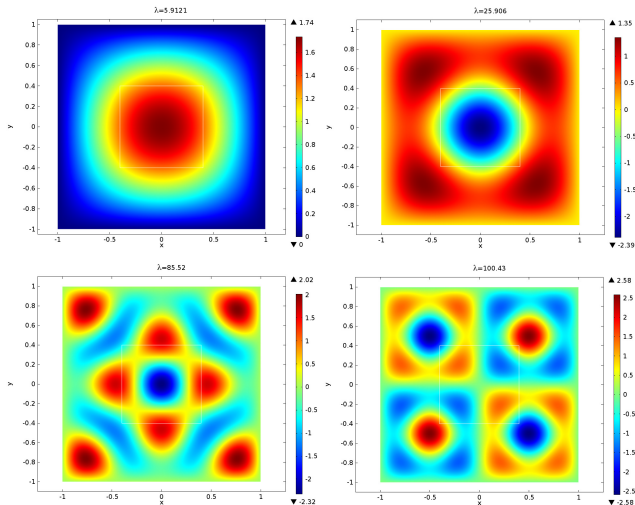
- 量子力学中的一个重要问题是求Schrödinger方程的特征值

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

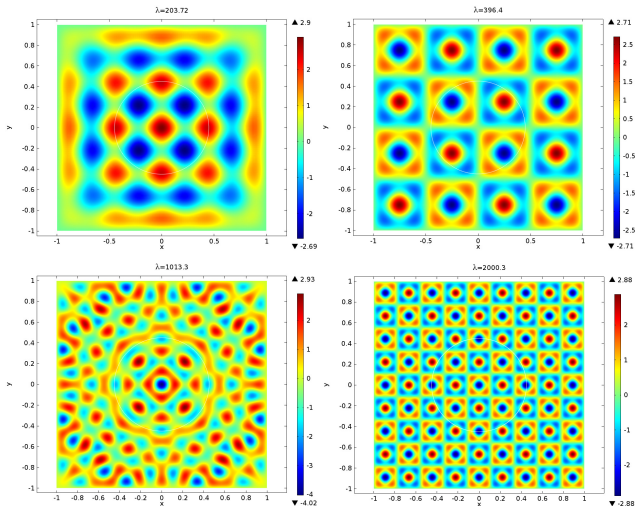
这里 H 是Hamilton量, E 是特征值, 代表能量, ψ_E 是对应的特征函数, 即波函数。

- 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。

Schrödinger方程的特征值和特征函数



Schrödinger方程的特征值和特征函数



主成分分析(PCA, Principal Component Analysis)

主成分分析经常用于减少数据集的维数，同时保持数据集中的对方差贡献最大的特征。其方法主要是通过对协方差矩阵进行特征分解，以得出数据的主成分（即特征向量）与它们的权值（即特征值或奇异值）。

如果想给出一条坐标轴可以尽量清晰的描述关心的数据，即更容易把它们分类。直观来看，肯定会选择与数据方差最大的那条直线，才能最大化数据的差异性。实际的做法就是将数据的高维坐标投影到这条直线，也就是向量上去，然后最大化投影后的方差。

原始数据的中心点

$$\tilde{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n$$

定义投影向量 \mathbf{u}_1 ，且 $\|\mathbf{u}_1\|_2 = 1$ 。

协方差矩阵(Covariance Matrix)

则每个 \mathbf{x}_n 的投影为 $\mathbf{u}_1^T \mathbf{x}_n$, 投影后的方差为

$$\frac{1}{N} \sum_{n=1}^N (\mathbf{u}_1^T \mathbf{x}_n - \mathbf{u}_1^T \tilde{\mathbf{x}})^2 = \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1$$

这里 \mathbf{S} 是数据的协方差矩阵。

协方差(Covariance)用于衡量两个变量的总体误差。设两个随机变量 X 和 Y 的期望值分别为 $E(X)$ 和 $E(Y)$, 则其协方差定义为:

$$\text{cov}(X, Y) = E((X - E(X))(Y - E(Y)))$$

方差是协方差的一种特殊情况, 即当两个变量是相同时:

$$\text{var}(X) = \text{cov}(X, X) = E((X - E(X))^2)$$

协方差从随机标量推广到随机向量, 则得到协方差矩阵定义:

$$\text{cov}(X) = E((X - E(X))(X - E(X))^T).$$

因此PCA是求解如下的约束优化问题

$$\begin{aligned} \max \quad & \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1 \\ \text{s.t.} \quad & \|\mathbf{u}_1\|_2 = 1. \end{aligned}$$

利用拉格朗日乘子法，引入标量 λ_1 变为如下无约束优化问题：

$$\max \{ \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1 + \lambda_1 (1 - \mathbf{u}_1^T \mathbf{u}_1) \}$$

其极值条件为梯度为零，即

$$\mathbf{S} \mathbf{u}_1 = \lambda_1 \mathbf{u}_1$$

这就转换成一个矩阵特征值分解问题。将特征值从大到小排列，保留前 M 个特征向量，即实现了从原来 N 维空间到 M 维新空间的降维。

注记

若矩阵 \mathbf{S} 不是方阵，则需要求奇异值。

- 在因素分析中，一个协变矩阵的特征向量对应于因素，而特征值是因素负载。因素分析是一种统计学技术，用于社会科学和市场营销、产品管理、运筹规划和其他处理大量数据的应用科学。其目标是用称为因素的少量的不可观测随机变量来解释在一些可观测随机变量中的变化。可观测随机变量用因素的线性组合来建模，再加上“残差项”。
- 在图像处理中，脸部图像的处理可以看作分量为每个像素的灰度的向量。该向量空间的维数是像素的个数。一个标准化面部图形的一个大型数据集合的协变矩阵的特征向量称为特征脸。它们对于将任何面部图像表达为它们的线性组合非常有用。特征脸提供了一种用于识别目的的数据压缩的方式。在这个应用中，一般只取最大那些特征值所对应的特征脸。

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 $r_i = \sum_{j \neq i}^n |a_{ij}|$ 为半径的圆盘, 即

$$D_i \stackrel{\text{def}}{=} \left\{ z \in \mathbb{C} \mid |z - a_{ii}| \leq r_i \right\}, \quad i = 1, \dots, n$$

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

$$B^{-1}AB = \begin{pmatrix} \frac{a_{ij}\alpha_j}{\alpha_i} \end{pmatrix}$$

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

$$D_1 = \{z \mid |z - 0.9| \leq 0.13\}, \quad D_2 = \{z \mid |z - 0.8| \leq 0.14\}$$

$$D_3 = \{z \mid |z - 0.4| \leq 0.03\}.$$

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 $B = \text{diag}(1, 1, 0.1)$, 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 $\bar{D}_1 = \{z \mid |z - 0.9| \leq 0.022\}, \bar{D}_2 = \{z \mid |z - 0.8| \leq 0.023\}, \bar{D}_3 = \{z \mid |z - 0.4| \leq 0.3\}$ 。

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 $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ 是 \mathbb{R}^n 的一组基。

- 对任意的初始向量

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

有

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

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

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

- 对任意 $\alpha \in \mathbb{R}$, $\mathbf{z} \in \mathbb{R}^n$, 有 $\max(\alpha \mathbf{z}) = \alpha \max(\mathbf{z})$ 。

幂迭代算法

幂迭代法的算法可写为

$$\left\{ \begin{array}{lcl} \boldsymbol{\nu}^{(0)} & = & \frac{\boldsymbol{\mu}}{\max \boldsymbol{\mu}} \\ \boldsymbol{z}^{(k)} & = & A\boldsymbol{\nu}^{(k-1)} \\ m_k & = & \max(\boldsymbol{z}^{(k)}) \\ \boldsymbol{\nu}^{(k)} & = & \frac{\boldsymbol{z}^{(k)}}{m_k}, \end{array} \right. \quad k = 1, 2, \dots$$

因为

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

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

幂迭代法的收敛性

而 $\max(\boldsymbol{\nu}^{(k)}) = 1$, 则

$$\begin{aligned}\boldsymbol{\nu}^{(k)} &= \frac{A^k \boldsymbol{\nu}^{(0)}}{\max(A^k \boldsymbol{\nu}^{(0)})} \\ \Rightarrow \max(A^k \boldsymbol{\nu}^{(0)}) &= \prod_{i=1}^k m_i\end{aligned}$$

$$\boldsymbol{\nu}^{(k)} \rightarrow \frac{\mathbf{x}_1}{\max \mathbf{x}_1},$$

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

$$m_k \rightarrow \max \left(A \frac{\mathbf{x}_1}{\max \mathbf{x}_1} \right) = \max \left(\frac{\lambda_1 \mathbf{x}_1}{\max \mathbf{x}_1} \right) = \lambda_1$$

收敛速度：线性

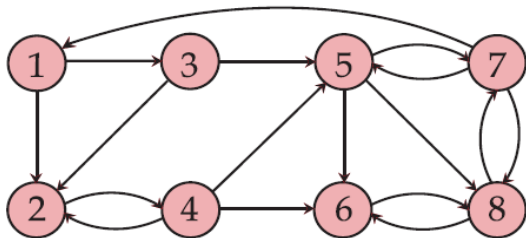
$$\begin{aligned} m_k &= \frac{\max(A^k \boldsymbol{\nu}^{(0)})}{\max(A^{k-1} \boldsymbol{\nu}^{(0)})} = \lambda_1 \frac{\max \left(\alpha_1 \mathbf{x}_1 + \sum_{i=2}^n \alpha_i \left(\frac{\lambda_i}{\lambda_1} \right)^k \mathbf{x}_i \right)}{\max \left(\alpha_1 \mathbf{x}_1 + \sum_{i=2}^n \alpha_i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

所以
$$\frac{|m_{k+1} - \lambda_1|}{|m_k - \lambda_1|} \approx \left| \frac{\lambda_2}{\lambda_1} \right|.$$

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 $\boldsymbol{\nu}^{(0)}$ 不含主特征值对应的特征向量 \mathbf{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。

PageRank的转移矩阵



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

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	...	I^{60}	I^{61}
1	0	0	0	0.0287	...	0.06	0.06
0	0.5	0.25	0.1667	0.0833	...	0.0675	0.0675
0	0.5	0	0	0	...	0.03	0.03
0	0	0.5	0.25	0.1667	...	0.0675	0.0675
0	0	0.25	0.1667	0.1111	...	0.0975	0.0975
0	0	0	0.25	0.1806	...	0.2025	0.2025
0	0	0	0.0833	0.0972	...	0.18	0.18
0	0	0	0.0833	0.3333	...	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

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

逆幂法的收敛性

对逆幂迭代法

$$\boldsymbol{\nu}^{(k)} \rightarrow \frac{\mathbf{x}_n}{\max \mathbf{x}_n},$$

$$m_k = \lambda_n^{-1} \left[1 + O \left(\left| \frac{\lambda_n}{\lambda_{n-1}} \right|^k \right) \right] \rightarrow \lambda_n^{-1}$$

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

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

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

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

则对 $A - qI$ 进行逆幂迭代

$$\left\{ \begin{array}{ll} \boldsymbol{\nu}^{(0)} &= \boldsymbol{\mu} / \max \boldsymbol{\mu} \\ (A - qI)\mathbf{z}^{(k)} &= \boldsymbol{\nu}^{(k-1)} \\ m_k &= \max(\mathbf{z}^{(k)}) \\ \boldsymbol{\nu}^{(k)} &= \frac{\mathbf{z}^{(k)}}{m_k}, \end{array} \right. \quad k = 1, 2, \dots$$

可得

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

幂法求特征值的思路：

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

定义 (Householder 变换)

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

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

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 $P^T P = I$, 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 $\|P\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ 。
- ⑤ 对任意 \mathbf{x} 和 \mathbf{v} , 满足 $\|\mathbf{x}\|_2 = \|\mathbf{v}\|_2$, 令

$$\mathbf{w} = \frac{\mathbf{x} - \mathbf{v}}{\|\mathbf{x} - \mathbf{v}\|_2} \implies P\mathbf{x} = \mathbf{v}.$$

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

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

可取

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

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

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

由于 $|k| = \|\mathbf{x}\|_2$. 可取

$$k = -\operatorname{sgn}(x_1)\|\mathbf{x}\|_2$$

则有

$$\mathbf{u} = \left(x_1 + \mathbf{sgn}(x_1)\|\mathbf{x}\|_2, x_2, \dots, x_n \right)$$

$$\|\mathbf{u}\|_2^2 = 2\|\mathbf{x}\|_2 \left(\|\mathbf{x}\|_2 + |x_1| \right),$$

$$\beta = \left[\|\mathbf{x}\|_2 (\|\mathbf{x}\|_2 + |x_1|) \right]^{-1}$$

$$\mathbf{P} = \mathbf{I} - \frac{2\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|_2^2} = \mathbf{I} - \beta\mathbf{u}\mathbf{u}^T$$

类似可将接连几个分量变为零, 设

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

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

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

Householder矩阵

其中

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

于是

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

由此确定的Householder矩阵为

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

并且

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

例

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

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

$$\mathbf{u} = \mathbf{x} - k\mathbf{e}_1 = (9, 5, 1, 1)^T,$$

$$\|\mathbf{u}\|_2^2 = 108,$$

$$\beta = [\|\mathbf{x}\|_2(\|\mathbf{x}\|_2 + |x_1|)]^{-1} = \frac{1}{54}$$

所以相应的Householder矩阵为

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

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 θ 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 $\text{sgn}(x_2)(x_1, x_2)$ 旋转 $\theta \in (0, \pi)$ 角, 变为平行于 x -轴的向量, 于是

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

因而若记

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

则

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

- 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

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

因而若记

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

则

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

现在将它推广到 n 维情况

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

称为**Givens矩阵**或**Givens变换**。显然其为正交矩阵，
设 $\mathbf{x} \in \mathbb{R}^n$ ， $\mathbf{y} = J(i, k, \theta)\mathbf{x}$ ，则 \mathbf{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 $x_k \neq 0, |x_k| \geq |x_i|$,

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

③ 若 $x_k \neq 0, |x_k| < |x_i|$,

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

笔记

$J(i, k, \theta)x$ 只改变 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 $\mathbf{x} = (1, 2, 3, 4)^T \in \mathbb{R}^4$ 有 $J(2, 4, \theta)\mathbf{x}$ 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

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

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

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

则令 $\bar{Q} = QD$ 为正交阵, $\bar{R} = D^{-1}R$ 为对角元是 $|r_{ii}|$ 的上三角阵。 $A = \bar{Q}\bar{R}$ 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 $R = P_2 P_1 A$ 是一个上三角阵, 且对角线元素为正, 则

$$Q = (P_2 P_1)^T = \begin{bmatrix} \frac{4}{5} & 0 & \frac{3}{5} \\ \frac{3}{5} & 0 & -\frac{4}{5} \\ 0 & 1 & 0 \end{bmatrix}$$

满足 $A = QR$.

也可用Householder变换, 对 A 的第1列 \mathbf{a}_1 , 有

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

则

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

则 $P_1 A$ 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 $\bar{P} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ 所以

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}, P_2 P_1 A = \begin{bmatrix} -5 & -5 & \frac{3}{5} \\ 0 & -1 & -1 \\ 0 & 0 & \frac{4}{5} \end{bmatrix}$$

$R = P_2 P_1 A$ 是上三角阵, 而

$$Q = (P_2 P_1)^T = \begin{bmatrix} -\frac{4}{5} & 0 & \frac{3}{5} \\ -\frac{3}{5} & 0 & -\frac{4}{5} \\ 0 & -1 & 0 \end{bmatrix}$$

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

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

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

$$X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$$

则

$$AX = XD \iff A\mathbf{x}_i = \lambda_i\mathbf{x}_i, \quad 1 \leq i \leq n.$$

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

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

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 :

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

若 X 非常病态, 即

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

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

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

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

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

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

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

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

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

使得

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

其中

$$\alpha = (a_{j+1,j}^2 + \dots + a_{nj}^2)^{1/2}$$

$$\mathbf{u}_j = (0, \dots, 0, a_{j+1,j} + \mathbf{sgn}(a_{j+1,j})\alpha, a_{j+2,j}, \dots, a_{nj})^T$$

因而如此逐次左乘 P_1, P_2, \dots , 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 $j = 1, 2, \dots, n-2$,

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

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

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

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

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

从上面的定理可以看出, 矩阵 A 的上Hessenberg形式不是唯一的, 但是可以证明下面的定理

定理

设 $A \in \mathbb{R}^{n \times n}$, $Q = [q_1, q_2, \dots, q_n]$ 及 $V = [v_1, v_2, \dots, v_n]$ 都是正交矩阵(q_i, v_i 分别是 Q 和 V 的列向量), 使 $Q^T A Q = H$ 与 $V^T A V = G$ 都是上Hessenberg矩阵, 其中 $H = [h_{ij}]$, $G = [g_{ij}]$ 。记 k 为使 $h_{k+1,k} = 0$ 的最小整数, 如果 $q_1 = v_1$, 则有

$$\begin{cases} q_i = \pm v_i, \\ |h_{i,i-1}| = |g_{i,i-1}|, \end{cases} \quad i = 2, 3, \dots, k$$

且若 $k < n$, 则 $g_{k+1,k} = 0$ 。

定理说明若 $Q^T A Q = H$ 与 $V^T A V = G$ 均为上Hessenberg矩阵, 且 Q 与 V 的第一列相同, 那么 G 和 H “本质上相同”记 $G = D^{-1} H D$, 其中 D 是对角元为 ± 1 的对角阵。

基本QR算法

注记

QR算法是计算中小型矩阵全部特征值的一种有效的方法。一般先把矩阵相似变换化为上**Hessenberg**形式（接近实**Schur**分解的形式），再用基于**QR**分解的**QR**迭代法计算。在一定的条件下迭代产生的矩阵序列趋于一种实**Schur**分解的形式。

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 **$B = RQ$** ，则 **$B = Q^T A Q$** ，**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

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

同步迭代法

幂法每次只能计算出一个特征值和特征向量，考虑一次能同时计算几个特征值和特征向量的方法：最简单的方式就是对几个不同的初值向量作幂法。设 $\mathbf{x}_1^{(0)}, \mathbf{x}_2^{(0)}, \dots, \mathbf{x}_p^{(0)}$ 是 p 个线性无关的初始向量，它们构成 $X_0 \in \mathbb{R}^{n \times p}$ 的列， $\text{rank}(X_0) = p$ 。对这些向量使用幂法称为同步迭代法。

① 给定 $X_0 \in \mathbb{R}^{n \times p}, \text{rank}(X_0) = p$,

② 对 $k = 1, 2, \dots$ 计算

$$X_k = AX_{k-1} \cdot \{\text{生成下一个矩阵}\}$$

设 $S_0 = \text{span}(X_0)$ ， S 是 A 的前 p 个按模最大的特征值 $\lambda_1, \lambda_2, \dots, \lambda_p$ 对应的特征向量 $\mathbf{v}_1, \dots, \mathbf{v}_p$ 张成的不变子空间。若 $\mathbf{x}_i, i = 1, \dots, p$ 不与 S 正交，则当 $|\lambda_p| > |\lambda_{p+1}|$ 时，对任意的 $k > 0, X_k = A^k X_0$ 构成了 $S_k = A^k S_0$ 的基底，子空间 S_k 将收敛到 S 。

正交迭代

同步迭代的 X_k 的各列在每次迭代时要重新调整防止上溢或者下溢，而且方法本质上是对 X_0 的每一列单独使用幂法，所以 X_k 的每一列收敛到 A 的主特征向量的倍数，则 X_k 的逐渐成为 S_k 的病态基底。为了克服这些困难，可在每一步迭代的时候对 X_k 作QR分解使得它的列正交化，这种方法称为**正交迭代法**。

- 1 给定 $X_0 \in \mathbb{R}^{n \times n}$, $\text{rank}(X_0) = p$,
- 2 对 $k = 1, 2, \dots$, 计算约化的QR分解, {**正交化**}。

$$Q_k R_k = X_{k-1},$$

- 3 $X_k = A Q_k$, {**生成下一个矩阵**}。

其中 $Q_k \in \mathbb{R}^{n \times p}$ 的列正交， $R_k \in \mathbb{R}^{p \times p}$ 是上三角阵。 $X_k \rightarrow X$ ， X 各列构成了 S 的基底。而 $Q_k \rightarrow Q$ ， Q 的各列构成了 S 的正交基。

原则上如果取 $p = n$, 则同步迭代法可以求出 $A \in \mathbb{R}^{n \times n}$ 的所有特征值和特征向量。如果对 \mathbb{R}^n 的正交基底同步迭代, 取基底为 I 的列, 即 $X_0 = I$, 定义

$$A_{k+1} = \bar{Q}_k^T A \bar{Q}_k,$$

希望 A_k 收敛到上三角矩阵, 然而在计算的时候最好给出一种计算 A_k 的递归的方法, 而不是每次迭代都用原始矩阵 A 形成 $\bar{Q}_k^T A \bar{Q}_k$ 。

如果取 $X_0 = I$, 则由QR分解

$$\bar{Q}_0 R_0 = X_0 \implies \bar{Q}_0 = R_0 = I$$

所以 $X_1 = A \bar{Q}_0 = A$ 。

为进行下一步迭代计算QR分解:

$$\bar{Q}_1 R_1 = X_1 = A$$

所以有

$$\begin{aligned} A_1 &= \bar{Q}_1^T A \bar{Q}_1 = \bar{Q}_1^T (\bar{Q}_1 R_1) \bar{Q}_1 \\ &= R_1 \bar{Q}_1. \end{aligned}$$

同步迭代的下一步需要计算

$$X_2 = A\bar{Q}_1 \text{ 及QR分解 } X_2 = \bar{Q}_2 R_2$$

但是可以通过计算 $Q_2 R_2 = A_1$ 达到同样的效果, 所以

$$\begin{aligned} X_2 &= A\bar{Q}_1 = \bar{Q}_1 \bar{Q}_1^T A\bar{Q}_1 \\ &= \bar{Q}_1 A_1 = \bar{Q}_1 (Q_2 R_2) \\ &= (\bar{Q}_1 Q_2) R_2 \\ &= \bar{Q}_2 R_2. \end{aligned}$$

这样不用形成和分解 X_2 , 只需要将 A_1 的QR分解反序就可直接计算出 $A_2 = R_2 Q_2$ 。显然, 用这种方式, 在不形成矩阵序列 $\{X_k\}$ 或不直接计算它的QR分解的情况下, 就可以将迭代进行下去。这种对同步迭代法所作的简单而精巧的改进所得的算法就是QR算法。

二者的等价性可由下面的定理给出

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 $\bar{R}_k = R_k \cdots R_2 R_1$ 。

注记

实际上 QR 迭代实际上是对 $A^k, k = 0, 1, \dots$ 进行 QR 分解, 因此 \bar{Q}_k 的列构成由 A^k 的列所张成的子空间的正交基, 而这些基是将单位矩阵的列作为初始向量对 A 进行同步迭代的结果。

注记

一般来说 QR 迭代可收敛到 A 的 $Schur$ 标准型, 如果 A 对称, 则 QR 迭代仍保持对称性, 此时 QR 迭代一定会收敛到对角阵。

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

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

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

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

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

基于此思想, 有如下的原点位移方法

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

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 。

❶ 第1种原点位移的QR迭代

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

收敛判别准则可取如下两者之一

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

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

❷ 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

$$A_2 = A - \mu_1 I = \begin{bmatrix} d_1 & b_2^{(1)} & 0 \\ b_2^{(1)} & d_2 & b_3^{(1)} \\ 0 & b_3^{(1)} & d_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

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

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

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

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

而矩阵A的特征值的准确值为

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

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 $O(n^2)$ ，而对对称矩阵为 $O(n)$ 。
- 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

对称阵 \longrightarrow 三对角阵 \longrightarrow 对角阵
一般矩阵 \longrightarrow Hessenberg矩阵 \longrightarrow 上三角阵

QR方法的计算量

- 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。