计算方法

刘景铖

计算机软件新技术国家重点实验室
南京大学

回顾与补充

- **Richardson iteration**
  - 两个视角：矩阵多项式、梯度最速下降
  - 思考：如果换作其它范数下面的最速下降？
  - 收敛性：当且仅当$\boldsymbol{\rho(I-\alpha A)} = \max\{|1-\alpha\lambda_1|, |1-\alpha\lambda_n|\} < \mathbf{1}$
  - 记$\boldsymbol{x_k = p_k(A)b \in span\{b, Ab, A^2b, \dots, A^{k-1}b\}}$；**Krylov**子空间
  - $\boldsymbol{x_* - x_k = (I - p_k(A)A)x_* = q_k(A)x_*}$
  - 收敛所需迭代次数正比于条件数$\boldsymbol{\frac{\lambda_n}{\lambda_1}}$
- **Conjugate gradients** （共轭梯度方法）
  - 记**Krylov**子空间$\boldsymbol{K_0 = \{0\}, K_i = span\{b, Ab, \dots A^{i-1}b\}}$
  - $\boldsymbol{x_i = arg\ min_{x \in K_i} \|x - x_*\|_A^2}$，其中$\boldsymbol{x_*}$满足$\boldsymbol{Ax_* = b}$
  - **CG**的收敛所需迭代次数正比于$\sqrt{\boldsymbol{\frac{\lambda_n}{\lambda_1}}}$，依然取决于条件数
  - 精确数值：最多迭代**n+1**次；有限精度：也需要选择合适的停止条件
- 对稀疏矩阵，或者有好的初始猜测时，迭代比高斯消元更加适合

预条件(Preconditioning)

要解$Ax = b$，选取矩阵$\boldsymbol{M}$，改为解$M^{-1}Ax = M^{-1}b$

$$cond(M^{-1}A) \neq cond(A)$$

注意：**Richardson iteration**和**CG**都需要正定矩阵
$M^{-1}A$可能不再是正定的，甚至可能不是对称的

假设$\boldsymbol{M}$是对称正定，则有**Cholesky**分解：$\boldsymbol{M} = \boldsymbol{E}\boldsymbol{E}^{\boldsymbol{T}}$，并且

$$cond(M^{-1}A) = cond(E^{-1}AE^{-T})$$

预条件(Preconditioning)

假设$\boldsymbol{M}$是对称正定，则有Cholesky分解： $\boldsymbol{M} = \boldsymbol{E}\boldsymbol{E}^T$，并且

$$cond(M^{-1}A) = cond(E^{-1}AE^{-T})$$

证明： $E^{-1}AE^{-T}$是对称正定的，有完整的特征基，因此只需要证明$M^{-1}A$和 $E^{-1}AE^{-T}$有着一样的特征值即可。任取$E^{-1}AE^{-T}$的特征向量和特征值

$$E^{-1}AE^{-T}v = \lambda v \Rightarrow E^{-T}E^{-1}AE^{-T}v = \lambda E^{-T}v \Rightarrow M^{-1}Ay = \lambda y$$

因为它们的奇异值一样，它们的条件数也是一样的

因此，可以先解$E^{-1}AE^{-T}y = E^{-1}b$，这是对称正定的系统

然后解得$x = E^{-T}y$

计算特征值与特征向量

最小化$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A}\, \boldsymbol{x}$, subject to $x^T x = 1$

$\det(\boldsymbol{xI} - \boldsymbol{A})$:关于$\boldsymbol{x}$的多项式

- n个根， $\boldsymbol{A}$的特征值
- 可能不稳定：多项式$(\boldsymbol{x}-\mathbf{1})(\boldsymbol{x}-\mathbf{2})..(\boldsymbol{x}-\mathbf{20})$
- 注：$\mathbf{det}(\boldsymbol{\lambda I} - \boldsymbol{A})$非常接近0并不意味着$\boldsymbol{\lambda}$是一个近似特征值
- 牛顿法求根：需要导数
- 截弦法求根：也需要计算$\det(\boldsymbol{xI} - \boldsymbol{A})$

迭代法？

- 解$Ax = b$：写出不动点方程，使用不动点迭代
- 解$Ax = \lambda x$：同时有两个变量： λ和x

对特征值的估计

Gershgorin circle theorem

- 令 $\boldsymbol{A}$为$\boldsymbol{n}\times\boldsymbol{n}$矩阵， $\boldsymbol{R_i} := \sum_{j\neq i}|\boldsymbol{a_{ij}}|$
- 则 $\boldsymbol{A}$的所有特征值都在复平面上的某个圆盘 $\boldsymbol{D(a_{ii}, R_i)}$ 之中

证明：任取A的一个特征值λ对应特征向量v，设i使得v_i为v 中绝对值最大的元素。

由$Av = \lambda v \Rightarrow \sum_j a_{ij}\, v_j = \lambda v_i \Rightarrow \sum_{j\neq i} a_{ij} v_j = (\lambda - a_{ii})v_i$

因此

$$|(\lambda - a_{ii})| = \left|\sum_{j\neq i} \frac{a_{ij}v_j}{v_i}\right| \leq \sum_{j\neq i} \left|\frac{a_{ij}v_j}{v_i}\right| \leq \sum_{j\neq i} |a_{ij}| = R_i$$

幂迭代 Power method

考虑可对角化的方阵$\boldsymbol{A}$. 假设$\boldsymbol{A}$有$\boldsymbol{n}$个特征值$\boldsymbol{\lambda_1, \lambda_2, \dots, \lambda_n}$满足$|\boldsymbol{\lambda_1}| > |\boldsymbol{\lambda_2}| \geq |\boldsymbol{\lambda_3}| \geq \cdots \geq |\boldsymbol{\lambda_n}|$。记它们对应的特征向量为$\boldsymbol{v_1, v_2, \dots, v_n}$。假设它们是线性无关的。

- $\boldsymbol{\lambda_1}$被称为占优特征值 (**dominating eigenvalue**)
- $\boldsymbol{v_1}$：占优特征向量

考虑任意向量$\boldsymbol{x} \in \boldsymbol{R^n}$，由于$\boldsymbol{v_1, v_2, \dots, v_n}$是线性无关的，所以可以展开

$$\boldsymbol{x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n}$$
$$\boldsymbol{Ax = \lambda_1 \alpha_1 v_1 + \lambda_2 \alpha_2 v_2 + \cdots + \lambda_n \alpha_n v_n}$$
$$\boldsymbol{A^2 x = \lambda_1^2 \alpha_1 v_1 + \lambda_2^2 \alpha_2 v_2 + \cdots + \lambda_n^2 \alpha_n v_n}$$
$$\boldsymbol{A^3 x = \lambda_1^3 \alpha_1 v_1 + \lambda_2^3 \alpha_2 v_2 + \cdots + \lambda_n^3 \alpha_n v_n}$$
$$\vdots$$
$$\boldsymbol{A^k x = \lambda_1^k \alpha_1 v_1 + \lambda_2^k \alpha_2 v_2 + \cdots + \lambda_n^k \alpha_n v_n}$$
$$\boldsymbol{= \lambda_1^k \left(\alpha_1 v_1 + \frac{\lambda_2^k}{\lambda_1^k} \alpha_2 v_2 + \cdots + \frac{\lambda_n^k}{\lambda_1^k} \alpha_n v_n \right)}$$

随着$\boldsymbol{k \to \infty, A^k x \to \lambda_1^k \alpha_1 v_1}$，如果$\boldsymbol{\alpha_1 \neq 0}$

幂迭代 Power method

$$A^k x = \lambda_1^k \alpha_1 v_1 + \lambda_2^k \alpha_2 v_2 + \cdots + \lambda_n^k \alpha_n v_n$$
$$= \lambda_1^k \left(\alpha_1 v_1 + \frac{\lambda_2^k}{\lambda_1^k} \alpha_2 v_2 + \cdots + \frac{\lambda_n^k}{\lambda_1^k} \alpha_n v_n \right)$$

随着$k \to \infty, A^k x \to \lambda_1^k \alpha_1 v_1$，如果$\alpha_1 \neq 0$

收敛速度：$|\lambda_2| / |\lambda_1|$

可能导致浮点数上溢或下溢：如果$|\lambda_1| > 1, A^k x \to \infty$; 如果$|\lambda_1| < 1$, $A^k x \to 0$

解决办法：每次迭代都进行归一化

$$w_{k+1} = A\, x_k\,,$$
$$x_{k+1} = \frac{w_{k+1}}{\|w_{k+1}\|}\,.$$

思考：归一化可以使用什么样的范数？为什么不除上λ_1^k进行归一化？

Power method

$$A^k x = \lambda_1^k \alpha_1 v_1 + \lambda_2^k \alpha_2 v_2 + \cdots + \lambda_n^k \alpha_n v_n$$
$$= \lambda_1^k \left(\alpha_1 v_1 + \frac{\lambda_2^k}{\lambda_1^k} \alpha_2 v_2 + \cdots + \frac{\lambda_n^k}{\lambda_1^k} \alpha_n v_n \right)$$

随着$k \to \infty, A^k x \to \lambda_1^k \alpha_1 v_1$

随着幂迭代的进行，得到了一个近似的特征向量方向之后，如何求解近似的特征值？

即，给定A和近似特征向量x，求λ使得$Ax \approx \lambda x$

最小二乘法：求λ以最小化$\|Ax - \lambda x\|_2^2$

$$\|Ax - \lambda x\|_2^2 = \|Ax\|_2^2 + \lambda^2 \|x\|_2^2 - 2\lambda x^T A x$$

法线方程$\lambda = x^T A x / x^T x$, **Rayleigh quotient!**

练习：使用扰动方法，从$\|Ax - (\lambda + \delta\lambda) x\|_2^2 \geq \|Ax - \lambda x\|_2^2, \forall \delta\lambda \in R$推导

Power method

最小二乘法：求λ以最小化$\|Ax-\lambda x\|_2^2$

$$\|Ax-\lambda x\|_2^2=\|Ax\|_2^2+\lambda^2\|x\|_2^2-2\lambda x^T Ax$$

法线方程$\lambda = x^T Ax/x^T x$, **Rayleigh quotient!**

定理：对于实数对称矩阵A，假设近似特征向量x满足$\|x\|_2=1$，并且实数λ满足$\|Ax-\lambda x\|_2 < \epsilon$，则有$min_{1\le j\le n}|\lambda_j-\lambda| < \epsilon$

证明：对于实数对称矩阵，可以选择特征向量$v_1, v_2, \dots, v_n$使得它们是正交且归一化的。展开$x=\alpha_1 v_1+\alpha_2 v_2+\cdots+\alpha_n v_n$

$$\begin{aligned}\|Ax-\lambda x\|_2^2 &= \left\|\sum_{j=1}^n \alpha_j(\lambda_j-\lambda)v_j\right\|_2^2\\ &=\sum_{j=1}^n |\alpha_j|^2|\lambda_j-\lambda|^2\|v_j\|_2^2\\ &=\sum_{j=1}^n |\alpha_j|^2|\lambda_j-\lambda|^2\\ &\ge min_{1\le j\le n}|\lambda_j-\lambda|^2\sum_{j=1}^n|\alpha_j|^2\end{aligned}$$

其中$\sum_{j=1}^n|\alpha_j|^2=\|x\|_2^2=1$，所以$\epsilon > \|Ax-\lambda x\|_2 \ge min_{1\le j\le n}|\lambda_j-\lambda|$

Inverse Power method

幂迭代告诉我们如何找到$\boldsymbol{\lambda_1}$与$\boldsymbol{v_1}$
其它特征值和特征向量呢？

给定方阵$\boldsymbol{A}$. 假设$\boldsymbol{A}$有$\boldsymbol{n}$个特征值$\boldsymbol{\lambda_1, \lambda_2, \ldots, \lambda_n}$和线性无关的特征向量为$\boldsymbol{v_1, v_2, \ldots, v_n}$

$\boldsymbol{A^{-1}}$**的特征值?**

$$\frac{\mathbf{1}}{\boldsymbol{\lambda_1}}, \frac{\mathbf{1}}{\boldsymbol{\lambda_2}}, \ldots, \frac{\mathbf{1}}{\boldsymbol{\lambda_n}}$$

特征向量？

$$\boldsymbol{v_1, v_2, \ldots, v_n}$$

迭代需要使用到$\boldsymbol{A^{-1}}$？只需要解线性方程：**LU**分解

Inverse Power method

幂迭代告诉我们如何找到$\boldsymbol{\lambda_1}$与$\boldsymbol{v_1}$
其它特征值和特征向量呢？

给定方阵$\boldsymbol{A}$. 假设$\boldsymbol{A}$有$\boldsymbol{n}$个特征值$\boldsymbol{\lambda_1, \lambda_2, \ldots, \lambda_n}$和线性无关的特征向量为$\boldsymbol{v_1, v_2, \ldots, v_n}$

$\boldsymbol{(A - qI)^{-1}}$的特征值?

$$\frac{\mathbf{1}}{\boldsymbol{\lambda_1 - q}}, \frac{\mathbf{1}}{\boldsymbol{\lambda_2 - q}}, \ldots, \frac{\mathbf{1}}{\boldsymbol{\lambda_n - q}}$$

特征向量？

$$\boldsymbol{v_1, v_2, \ldots, v_n}$$

Inverse Power method

给定方阵$\boldsymbol{A}$. 假设$\boldsymbol{A}$有$\boldsymbol{n}$个特征值$\boldsymbol{\lambda_1, \lambda_2, \dots, \lambda_n}$和线性无关的特征向量为$\boldsymbol{v_1, v_2, \dots, v_n}$

$(\boldsymbol{A-qI})^{-1}$的特征值?

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

特征向量?

$$v_1, v_2, \dots, v_n$$

如果对$(\boldsymbol{A-qI})^{-1}$使用幂迭代呢?
即$\boldsymbol{x_k} = (\boldsymbol{A-qI})^{-1}\boldsymbol{x_{k-1}}$
注意：可以通过高斯消元，先求出矩阵B = $(\boldsymbol{A-qI})^{-1}$的LU分解，再进行迭代；对于迭代次数远远比n小的稀疏矩阵，也可以尝试每一次迭代都解一次线性方程组

收敛到$\frac{1}{\lambda_k - q} = max_{1\le j\le n} \frac{1}{|\lambda_j - q|}$

如何选择q? 可以结合**Rayleigh quotient ⇒ Rayleigh quotient iteration**

Rayleigh Quotient Iteration

$$\boldsymbol{\sigma}_k = \frac{\boldsymbol{x}_{k-1}^T \boldsymbol{A} \boldsymbol{x}_{k-1}}{\boldsymbol{x}_{k-1}^T \boldsymbol{x}_{k-1}},$$

$$\boldsymbol{w}_k = (\boldsymbol{A} - \boldsymbol{\sigma}_k \boldsymbol{I})^{-1} \boldsymbol{x}_{k-1},$$

$$\boldsymbol{x}_k = \frac{\boldsymbol{w}_k}{\|\boldsymbol{w}_k\|}$$

注意$(\boldsymbol{A} - \boldsymbol{\sigma}_k \boldsymbol{I})$不再是固定的

迭代次数更少，但是每一次迭代的计算代价更多

Power method

幂迭代告诉我们如何找到$\boldsymbol{\lambda_1}$与$\boldsymbol{v_1}$

其它特征值和特征向量呢？

$$\boldsymbol{x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n}$$

如果$\boldsymbol{\alpha_1 \neq 0}$，随着$\boldsymbol{k \to \infty, A^k x \to \lambda_1^k \alpha_1 v_1}$

如果$\boldsymbol{\alpha_1 = 0}$呢？

- 如果$\boldsymbol{\alpha_2 \neq 0}$，随着$\boldsymbol{k \to \infty, A^k x \to \lambda_2^k \alpha_2 v_2}$
- 如果$\boldsymbol{\alpha_2 = 0, \alpha_3 \neq 0}$，随着$\boldsymbol{k \to \infty, A^k x \to \lambda_3^k \alpha_3 v_3}$

Power method

幂迭代告诉我们如何找到$\boldsymbol{\lambda_1}$与$\boldsymbol{v_1}$
其它特征值和特征向量呢？

如果$\boldsymbol{\alpha_1 \neq 0}$，随着$\boldsymbol{k \to \infty, A^k x \to \lambda_1^k \alpha_1 v_1}$
如果$\boldsymbol{\alpha_1 = 0}$呢？

每一次迭代，通过投影确保与已知的特征向量都是正交的

QR分解与同时迭代

如何同时找出所有的特征值与特征向量？

假设A是实数对称矩阵。同时进行幂迭代，并保持向量两两正交。
回忆：正交化，即QR分解

最开始

$$\left[A\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} \middle| A\begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix} \middle| \cdots \middle| A\begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix} \right] = \left[\overline{q}_1^1 | \cdots | \overline{q}_m^1\right] \begin{bmatrix} r_{11}^1 & r_{12}^1 & \cdots & r_{1m}^1 \\ & r_{22}^1 & & \vdots \\ & & \ddots & \vdots \\ & & & r_{mm}^1 \end{bmatrix}$$

$$A\overline{Q}_1 = \left[A\overline{q}_1^1 | A\overline{q}_2^1 | \cdots | A\overline{q}_m^1\right]$$

$$= \left[\overline{q}_1^2 | \overline{q}_2^2 | \cdots | \overline{q}_m^2\right] \begin{bmatrix} r_{11}^2 & r_{12}^2 & \cdots & r_{1m}^2 \\ & r_{22}^2 & & \vdots \\ & & \ddots & \vdots \\ & & & r_{mm}^2 \end{bmatrix}$$

$$= \overline{Q}_2 R_2.$$

QR分解与同时迭代

如何同时找出所有的特征值与特征向量？

假设A是实数对称矩阵。同时进行幂迭代，并保持向量两两正交。

$$
\begin{aligned}
A\overline{Q}_1 &= \left[A\overline{q}_1^1 | A\overline{q}_2^1 | \cdots | A\overline{q}_m^1\right] \\
&= \left[\overline{q}_1^2 | \overline{q}_2^2 | \cdots | \overline{q}_m^2\right]
\begin{bmatrix}
r_{11}^2 & r_{12}^2 & \cdots & r_{1m}^2 \\
 & r_{22}^2 & & \vdots \\
 & & \ddots & \vdots \\
 & & & r_{mm}^2
\end{bmatrix} \\
&= \overline{Q}_2 R_2.
\end{aligned}
$$

算法：

Set $\overline{Q}_0 = I$
for $j = 1, 2, 3, \ldots$
$\quad A\overline{Q}_j = \overline{Q}_{j+1} R_{j+1}$
end

QR分解与同时迭代

对于一般的可对角化矩阵，$\boldsymbol{A}$和$\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}$有相同的特征值

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

QR迭代的等价形式：

- $\boldsymbol{A}_1 = \boldsymbol{A}$
- 分解 $\boldsymbol{A}_k = \boldsymbol{Q}_k\boldsymbol{R}_k$
- $\boldsymbol{A}_{k+1} = \boldsymbol{R}_k\,\boldsymbol{Q}_k$

如果收敛，则 $\boldsymbol{A}_\infty = \boldsymbol{Q}_\infty\boldsymbol{R}_\infty = \boldsymbol{R}_\infty\,\boldsymbol{Q}_\infty$

奇异值 (singular values)

对于一般的$\boldsymbol{m}\times\boldsymbol{n}$实数非方阵$\boldsymbol{A}$，可以有奇异值分解

$$\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{S}\,\boldsymbol{V}^{\boldsymbol{T}}$$

$\boldsymbol{A}$ 的形状：$\boldsymbol{m}\times\boldsymbol{n}$
$\boldsymbol{U}$的形状：$\boldsymbol{m}\times\boldsymbol{m}$
$\boldsymbol{S}$的形状：$\boldsymbol{m}\times\boldsymbol{n}$
$\boldsymbol{V}$的形状：$\boldsymbol{n}\times\boldsymbol{n}$

$\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{U} = \boldsymbol{I}_{\boldsymbol{m}}$,　　$\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V} = \boldsymbol{I}_{\boldsymbol{n}}$

观察：

$$\boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}} = \boldsymbol{U}\boldsymbol{S}\;\;\boldsymbol{S}^{\boldsymbol{T}}\boldsymbol{U}^{\boldsymbol{T}}$$
$$\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}^{\boldsymbol{T}}\boldsymbol{S}\boldsymbol{V}^{\boldsymbol{T}}$$

$\boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}}$和$\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$　都是实数对称阵，可做特征值分解！

定理：$\boldsymbol{AA^T}$和$\boldsymbol{A^TA}$ 有着相同的非零特征值。

证明：考虑$\boldsymbol{AA^T}$的特征值$\boldsymbol{\lambda \neq 0}$和特征向量$\boldsymbol{v}$：$\boldsymbol{AA^Tv = \lambda v}$

$$\boldsymbol{A^TA(A^Tv) = A^T(AA^Tv) = A^T(\lambda v) = \lambda(A^Tv)}$$

可见，只要$\boldsymbol{A^Tv \neq 0}$，则$\boldsymbol{A^Tv}$为 $\boldsymbol{A^TA}$的一个特征向量，特征值也是$\boldsymbol{\lambda}$。不妨假设$\boldsymbol{A^Tv = 0}$，则$\boldsymbol{AA^Tv = A(A^Tv) = 0}$，与 $\boldsymbol{\lambda \neq 0}$矛盾。

类似地，考虑$\boldsymbol{A^TA}$的特征值$\boldsymbol{\lambda \neq 0}$和特征向量$\boldsymbol{v}$：$\boldsymbol{A^TAv = \lambda v}$，也有

$$\boldsymbol{AA^T(Av) = A(A^TAv) = A(\lambda v) = \lambda(Av)}.$$

同理有 $\boldsymbol{Av \neq 0}$，因此$\boldsymbol{Av}$为 $\boldsymbol{AA^T}$的一个特征向量，特征值也是$\boldsymbol{\lambda}$。

Operator norm and Condition number

2 to 2 operator norm of A: $\sigma_{max}(A)$

2-condition number of A: $\frac{\sigma_{max}(A)}{\sigma_{min}(A)}$

Corollary. If matrices A and B have the same non-zero singular values, they also have the same condition number