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：解线性方程的迭代方法

<u>**通知：** 下周四的作业将作为**take-home midterm**</u>

回顾迭代方程$x_{k+1} = Ax_k + b$
令x_***为其中一个不动点**

- $x_{k+1} - x_* = A(x_k - x_*)$
- $x_k - x_* = A^k(x_0 - x_*)$

什么时候收敛？
对任意初始值x_0，**迭代方程**$x_{k+1} = Ax_k + b$ **都收敛到唯一的不动点**

- $\Leftrightarrow A^k \to 0$
- $\Leftrightarrow \rho(A) < 1$

也可能是因为x_0选得好，使得$x_0 - x_*$落在一个好的线性子空间里面

回顾与补充：解线性方程的迭代方法

考虑迭代方程$x_{k+1} = Ax_k + b$
令x_*为其中一个不动点

- $x_{k+1} - x_* = A(x_k - x_*)$
- $x_k - x_* = A^k(x_0 - x_*)$

为什么$A^k \to 0$（即$\rho(A) < 1$）的时候对任意初始值都收敛？

- $\|x_k - x_*\| = \|A^k(x_0 - x_*)\| \le \|A^k\| \|x_0 - x_*\|$
- $x_k = Ax_{k-1} + b = A(Ax_{k-2} + b) + b = \cdots$
- $x_k = A^k x_0 + (A^{k-1} + A^{k-2} + \cdots + I)b$
  $= A^k x_0 + (I - A)^{-1}(I - A^k)b$
- 存在唯一的不动点：$x_* = Ax_* + b \Rightarrow x_* = (I - A)^{-1}b$

回顾与补充：解线性方程的迭代方法

考虑迭代方程$x_{k+1} = Ax_k + b$
令x_*为其中一个不动点

- $x_{k+1} - x_* = A(x_k - x_*)$
- $x_k - x_* = A^k(x_0 - x_*)$

反过来，如果对任意初始值都收敛，为什么$A^k \to 0$（即$\rho(A) < 1$）？

- **对任意初始值都收敛，即$\forall x_0$，$x_k - x_* = A^k(x_0 - x_*) \to 0$**
- **由于 x_0是任意的，$x_0 - x_*$可以取到任意的向量**
- **如果一个矩阵乘上任意一个向量，都是可以任意接近0的，**则矩阵必须是零矩阵

（对称）正定矩阵

如果对所有向量$x \neq 0$，$x^T Ax > 0$，则称A为正定矩阵(positive definite matrix)。亦记为$A \succ 0$

如果对所有向量$x \neq 0$，$x^T Ax \geq 0$，则称A为半正定矩阵(positive semidefinite matrix， PSD matrix) 。亦记为$A \succcurlyeq 0$

要解$Ax = b$，理论上可以转化为解$A^T Ax = A^T b$

- $A^T A \succcurlyeq 0$
- $A^T A$是对称的
- 如果A可逆，则$A^T A$也可逆
- 但是条件数可能会变坏：2-条件数会变为原来的平方

定理：实数对称矩阵A是正定的，当且仅当其所有特征值为正
定理：实数对称矩阵A是半正定的，当且仅当其所有特征值为非负

正定矩阵与高斯消元

回顾：如果对矩阵一行一行地进行高斯消元，则存在**是下三角阵**L，上三角阵U**使得**

$$A = LU$$

$$L^{-1}A = U$$

如果A**是对称的话：**对矩阵的行进行消元的每一步操作，对矩阵的列也同样可以消元！

$$L^{-1}A(L^{-1})^T = ?$$

若进一步假设A**是正定的，**则存在Cholesky分解：

$$A = C^T C$$

其中 C**为上三角矩阵**（对角线元素非零）

正定矩阵与特征值

定理：实数对称矩阵A是正定的，当且仅当其所有特征值为正。

证明：

(=>) 设 $\boldsymbol{\lambda} \in \boldsymbol{R}$和非零向量$\boldsymbol{v}$满足$\boldsymbol{Av} = \boldsymbol{\lambda v}$，则由正定性，

$$v^T \boldsymbol{Av} = \boldsymbol{\lambda v^T v} > \mathbf{0}$$

又因为$\boldsymbol{v^T v} > \mathbf{0}$，因此 $\boldsymbol{\lambda} > \mathbf{0}$。

(<=) 因为 $\boldsymbol{A}$ 是实数对称矩阵，可以设 $\boldsymbol{A}$的两两正交的特征向量为$\boldsymbol{v_1, v_2, .., v_n}$，则任意向量$x$可以表达成$\boldsymbol{v_1, v_2, .., v_n}$ 的线性组合:

$$x^T Ax = (\boldsymbol{a_1 v_1} + \cdots + \boldsymbol{a_n v_n})^T \boldsymbol{A}(\boldsymbol{a_1 v_1} + \cdots + \boldsymbol{a_n v_n})$$

$$= \boldsymbol{\lambda_1 a_1^2} + \cdots + \boldsymbol{\lambda_n a_n^2} > \mathbf{0}$$

因此 $\boldsymbol{A}$是正定的。

Loewner order

对于同样大小的实数对称矩阵A, B，如果$A - B \succcurlyeq 0$，也记为$A \succcurlyeq B$

一个充分条件： A的最小特征值也比 B的最大特征值要大（或者一样大）

如果它们拥有同样的特征向量，则每个对应的特征值都是大于等于的关系

注意：这仅仅是一个偏序（partial order）

特征值特征向量

$$Av = \lambda v$$

- v为A的特征向量 (eigenvector)
- λ为对应于特征向量v的特征值(eigenvalue)
- A的特征多项式（characteristic polynomial ）为$\det(A - xI)$
- $\det(A - xI) = 0$ 的零点是A的所有特征值

特征值λ的代数重数， algebraic multiplicity : λ作为重根出现的次数
特征值λ的几何重数， geometric multiplicity : λ对应的特征空间的维度
对于可对角化的矩阵， 代数重数=几何重数

定理. 令 A 为$n\times n$实数对称矩阵. 则存在正规正交（orthonormal）特征向量，它们能张成$\boldsymbol{R}^n$，且所有特征值均为实数

实数对称矩阵的谱分解

令 A 为$n \times n$实数对称矩阵， V为正规正交（orthonormal）的特征向量组成的矩阵

注意到$V^T V = I$

$$Av_i = \lambda_i v_i \Rightarrow AV = VD \Rightarrow A = VDV^T$$

实数对称矩阵的谱分解

任意一个向量都可以表示成一组正规正交基 $\{v_i\}$ 的线性组合

$$I = \sum_i v_i v_i^T$$

$$Av_i = \lambda_i v_i \Rightarrow A = \sum_i \lambda_i v_i v_i^T$$

$$A^{-1} = \sum_i \lambda_i^{-1} v_i v_i^T, \qquad if\ A^{-1}\ exists.$$

$$A^{+} = \sum_{i:\lambda_i \neq 0} \lambda_i^{-1} v_i v_i^T$$

特征值常用的等式

- $\mathrm{tr}(A) = \sum_{i=1}^{n} A_{i,i} = \sum_{i=1}^{n} \lambda_i$
- $\det(A) = \prod_{i=1}^{n} \lambda_i$

特征值的min-max刻画 (Courant-Fischer)

- **对实数对称矩阵A，** 最大的特征值

$$\lambda_n(A) = max_{x\neq 0}\ \frac{x^T A\ x}{x^T x}$$

证明：等号可以取到。只需要证明$\frac{x^T A\ x}{x^T x} \leq \lambda_n(A)$。

因为 A 是实数对称矩阵，可以设 A的两两正交的特征向量为 $v_1, v_2, .., v_n$

$$x^T A x = (a_1 v_1 + \cdots + a_n v_n)^T A (a_1 v_1 + \cdots + a_n v_n)$$
$$= \lambda_1 a_1^2 + \cdots + \lambda_n a_n^2 \leq \lambda_n (a_1^2 + \cdots + a_n^2)$$

$$x^T x = (a_1 v_1 + \cdots + a_n v_n)^T (a_1 v_1 + \cdots + a_n v_n) = a_1^2 + \cdots + a_n^2$$

因此$\frac{x^T A\ x}{x^T x} \leq \lambda_n$

特征值的min-max刻画 (Courant-Fischer)

对实数对称矩阵A，最大的特征值

$$\lambda_n(A) = max_{x\neq 0}\ \frac{x^T A\, x}{x^T x}$$

最小特征值

$$\lambda_1(A) = min_{x\neq 0}\ \frac{x^T A\, x}{x^T x}$$

更一般地：

$$\lambda_k(A) = min_{x\neq 0, x^T v_i = 0, \forall i\in\{1,\dots,k-1\}}\ \frac{x^T A\, x}{x^T x}$$

$$= min_{U: dim(U)=k}\, max_{x\in U, x\neq 0}\ \frac{x^T A\, x}{x^T x}$$

$$\lambda_k(A) = max_{x\neq 0, x^T v_i = 0, \forall i\in\{k+1,\dots,n\}}\ \frac{x^T A\, x}{x^T x}$$

$$= max_{U: dim(U)=n-k+1}\, min_{x\in U, x\neq 0}\ \frac{x^T A\, x}{x^T x}$$

正定矩阵

如果 $\boldsymbol{A}$是正定阵，

$$\|A\|_2 := \max_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2\leq 1}\sqrt{x^T A^T A x}$$

$$= \sqrt{\lambda_{\max}(A^T A)} = \lambda_{\max}(A)$$

则$cond_2(A) = \|A\|_2 \cdot \|A^{-1}\|_2 = \lambda_{\max}(A)/\lambda_{\min}(A)$

矩阵与多项式

- 设$\boldsymbol{\lambda} \in \boldsymbol{R}$和非零向量$\boldsymbol{v}$满足$\boldsymbol{A}\boldsymbol{v} = \boldsymbol{\lambda}\boldsymbol{v}$
- $\boldsymbol{A}^2\boldsymbol{v} = \boldsymbol{\lambda}^2\boldsymbol{v}$
- $\boldsymbol{A}^3\boldsymbol{v} = \boldsymbol{\lambda}^3\boldsymbol{v}$
- …

给定多项式$\boldsymbol{p}(\boldsymbol{x}) = \boldsymbol{c}_0 + \boldsymbol{c}_1\boldsymbol{x} + \cdots + \boldsymbol{c}_n\boldsymbol{x}^n$，考虑矩阵
$\boldsymbol{p}(\boldsymbol{A}) = \boldsymbol{c}_0\boldsymbol{I} + \boldsymbol{c}_1\boldsymbol{A} + \cdots + \boldsymbol{c}_n\boldsymbol{A}^n$
则$\boldsymbol{p}(\boldsymbol{A})\,\boldsymbol{v} = \boldsymbol{p}(\boldsymbol{\lambda})\boldsymbol{v}$

特征空间如何变化？

– $A \to p(A)$
–特征空间$\{\lambda_1, \lambda_2, \cdots, \lambda_n\} \to \{p(\lambda_1), p(\lambda_2), \cdots, p(\lambda_n)\}$

矩阵与多项式

考虑任意向量$\boldsymbol{x} \in \boldsymbol{R}^n$，假设$\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$能张成$\boldsymbol{R}^n$，所以可以展开

$$\boldsymbol{x} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_n \boldsymbol{v}_n$$

$$\boldsymbol{A}\boldsymbol{x} = \lambda_1 \alpha_1 \boldsymbol{v}_1 + \lambda_2 \alpha_2 \boldsymbol{v}_2 + \cdots + \lambda_n \alpha_n \boldsymbol{v}_n$$

$$\boldsymbol{A}^2\boldsymbol{x} = \lambda_1^2 \alpha_1 \boldsymbol{v}_1 + \lambda_2^2 \alpha_2 \boldsymbol{v}_2 + \cdots + \lambda_n^2 \alpha_n \boldsymbol{v}_n$$

$$\boldsymbol{A}^3\boldsymbol{x} = \lambda_1^3 \alpha_1 \boldsymbol{v}_1 + \lambda_2^3 \alpha_2 \boldsymbol{v}_2 + \cdots + \lambda_n^3 \alpha_n \boldsymbol{v}_n$$

$$\vdots$$

$$\boldsymbol{A}^k\boldsymbol{x} = \lambda_1^k \alpha_1 \boldsymbol{v}_1 + \lambda_2^k \alpha_2 \boldsymbol{v}_2 + \cdots + \lambda_n^k \alpha_n \boldsymbol{v}_n$$

$$p(\boldsymbol{A})\boldsymbol{x} = p(\lambda_1^k) \alpha_1 \boldsymbol{v}_1 + p(\lambda_2^k) \alpha_2 \boldsymbol{v}_2 + \cdots + p(\lambda_n^k) \alpha_n \boldsymbol{v}_n$$

一个神奇的多项式

要解$Ax=b$，可否求出多项式$\boldsymbol{p}(\boldsymbol{A})$ 使得 $\boldsymbol{p}(\boldsymbol{A})\boldsymbol{A}=\boldsymbol{I}$?

记$\boldsymbol{q}(\boldsymbol{x})=\mathbf{1}-\boldsymbol{x}\boldsymbol{p}(\boldsymbol{x})$，则$\boldsymbol{q}(\boldsymbol{A})$ 为全零矩阵。

这样的多项式$\boldsymbol{q}(\boldsymbol{x})$一定存在！

考虑由$\left\{\boldsymbol{I},\boldsymbol{A},\boldsymbol{A}^2,\boldsymbol{A}^3,\ldots,\boldsymbol{A}^{n^2}\right\}$张成的向量空间，维度是多少？练习：证明它们是线性相关的。

Cayley-Hamilton: 只需要$\boldsymbol{n}$次多项式即可

Cayley-Hamilton

考虑关于$\boldsymbol{x}$的$\boldsymbol{n}$次多项式$q_A(x) = \det(\boldsymbol{A} - \boldsymbol{xI})$

假设$\boldsymbol{q_A(x) = c_0 + c_1 x + \cdots + c_n x^n}$

类似地定义$\boldsymbol{q_A(B) = c_0 I + c_1 B + \cdots + c_n B^n}$

定理**(Cayley-Hamilton):** $\boldsymbol{q_A(A) = 0}$为全零矩阵

非平凡的：注意$\det(\boldsymbol{A} - \boldsymbol{xI})$仅仅为一个数

Cayley-Hamilton

考虑关于$\boldsymbol{x}$的$\boldsymbol{n}$次多项式$q_A(x) = \det(\boldsymbol{A} - \boldsymbol{xI})$

定理(Cayley-Hamilton): $\boldsymbol{q_A(A) = 0}$为全零矩阵

证明(sketch)：先考虑$\boldsymbol{A}$是可以对角化的。注意$\boldsymbol{q_A(A)}$也一定是可以对角化的。

取$\boldsymbol{A}$的任一特征值$\boldsymbol{\lambda} \in \boldsymbol{R}$和特征向量$\boldsymbol{v}$满足$\boldsymbol{Av = \lambda v}$。由前面的观察可知，$\boldsymbol{v}$也是$\boldsymbol{q_A(A)}$的特征向量，而且特征值满足$\boldsymbol{q_A(A)v = q_A(\lambda)v = 0}$

可见$\boldsymbol{q_A(A)}$的每个特征向量对应的特征值均为0

又因为$\boldsymbol{q_A(A)}$可以对角化，此时$\boldsymbol{q_A(A)}$只能是全零矩阵

Cayley-Hamilton

考虑关于$\boldsymbol{x}$的$\boldsymbol{n}$次多项式$q_A(x) = \det(\boldsymbol{A} - \boldsymbol{xI})$

定理(Cayley-Hamilton): $\boldsymbol{q_A(A) = 0}$**为全零矩阵**

证明(cont'd, sketch)： 对于$\boldsymbol{A}$不可以对角化的情况，可以使用Jordan标准形分解；或者利用可对角化矩阵在复数域上的稠密性(任意一个矩阵，均可表示为可对角化矩阵的极限）

Cayley-Hamilton

考虑关于$\boldsymbol{x}$的$\boldsymbol{n}$次多项式$q_A(x) = \det(\boldsymbol{A} - \boldsymbol{xI})$

定理(Cayley-Hamilton): $\boldsymbol{q_A}(\boldsymbol{A}) = \mathbf{0}$为全零矩阵

注意$q_A(0) = \det(\boldsymbol{A}) \neq \mathbf{0}$

$$\boldsymbol{A}^{-1} = \frac{(-\mathbf{1})^{\boldsymbol{n}-\mathbf{1}}}{\det(\boldsymbol{A})}(\boldsymbol{A}^{\boldsymbol{n}-\mathbf{1}} + \boldsymbol{c}_{\boldsymbol{n}-\mathbf{1}}\boldsymbol{A}^{\boldsymbol{n}-\mathbf{2}} + \cdots + \boldsymbol{c}_{\mathbf{1}}\boldsymbol{I})$$

$$\boldsymbol{A}^{-1} = \boldsymbol{p}(\boldsymbol{A})$$