



计算方法

刘景铨

计算机软件新技术国家重点实验室
南京大学



回顾：线性迭代方法的收敛性

通知：本次作业为take-home midterm

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

- $\Leftrightarrow A^k \rightarrow \mathbf{0}$
- $\Leftrightarrow \rho(A) < 1$

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

作为应用，我们看到了迭代法解线性方程组的例子

解 $Ax = b$ 时：Jacobi和Gauss-Seidel迭代都通过写出不动点方程，使用不动点迭代

这节课：

- 理论上，能通过某种“线性组合”表示出来的问题都适合使用线性迭代方程
- 解 $Ax = \lambda x$ ：同时有两个变量： λ 和 x 。严格来说，它们的乘积并不是线性关系，但（线性）迭代法依然适用



(对称) 正定矩阵

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

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

要解 $Ax = b$, 理论上可以转化为解 $A^T A x = A^T b$

- $A^T A \geq 0$
- $A^T A$ 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



回顾：特征值特征向量

$$Av = \lambda v$$

- v 为 A 的一个特征向量 (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 \mathbf{R}^n , 且所有特征值均为实数



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：

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

$$v^T Av = \lambda v^T v > 0$$

又因为 $v^T v > 0$ ，因此 $\lambda > 0$ 。

(\Leftarrow) 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, \dots, v_n ，则任意向量 x 可以表达成 v_1, v_2, \dots, v_n 的线性组合：

$$\begin{aligned} x^T Ax &= (a_1 v_1 + \dots + a_n v_n)^T A (a_1 v_1 + \dots + a_n v_n) \\ &= \lambda_1 a_1^2 + \dots + \lambda_n a_n^2 > 0 \end{aligned}$$

因此 A 是正定的。



Loewner order

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

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）



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

$$I = \sum_i v_i v_i^\top$$

$$A v_i = \lambda_i v_i \Rightarrow A = \sum_i \lambda_i v_i v_i^\top$$

$$A^{-1} = \sum_i \lambda_i^{-1} v_i v_i^\top, \quad \text{if } A^{-1} \text{ exists.}$$

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



特征值常用的等式

- $\text{tr}(A) = \sum_{i=1}^n A_{i,i} = \sum_{i=1}^n \lambda_i$
- $\det(A) = \prod_{i=1}^n \lambda_i$
- $\|A\|_F^2 = \text{tr}(A^\top A)$



对特征值的估计

Gershgorin circle theorem

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

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



特征值的min-max刻画 (Courant-Fischer)

- 对实数对称矩阵 A ，最大的特征值

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

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

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, \dots, v_n

$$\begin{aligned} x^\top A x &= (a_1 v_1 + \dots + a_n v_n)^\top A (a_1 v_1 + \dots + a_n v_n) \\ &= \lambda_1 a_1^2 + \dots + \lambda_n a_n^2 \leq \lambda_n (a_1^2 + \dots + a_n^2) \end{aligned}$$

$$x^\top x = (a_1 v_1 + \dots + a_n v_n)^\top (a_1 v_1 + \dots + a_n v_n) = a_1^2 + \dots + a_n^2$$

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



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：

$$\begin{aligned} \lambda_k(A) &= \min_{x \neq 0, x^T v_i = 0, \forall i \in \{1, \dots, k-1\}} \frac{x^T A x}{x^T x} \\ &= \min_{U: \dim(U)=k} \max_{x \in U, x \neq 0} \frac{x^T A x}{x^T x} \end{aligned}$$

$$\begin{aligned} \lambda_k(A) &= \max_{x \neq 0, x^T v_i = 0, \forall i \in \{k+1, \dots, n\}} \frac{x^T A x}{x^T x} \\ &= \max_{U: \dim(U)=n-k+1} \min_{x \in U, x \neq 0} \frac{x^T A x}{x^T x} \end{aligned}$$



正定矩阵

如果 A 是正定阵,

$$\begin{aligned}\|A\|_2 &:= \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2 \leq 1} \sqrt{x^\top A^\top Ax} \\ &= \sqrt{\lambda_{\max}(A^\top A)} = \lambda_{\max}(A)\end{aligned}$$

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



矩阵与多项式

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

给定多项式 $p(x) = c_0 + c_1x + \cdots + c_nx^n$, 考虑矩阵

$$p(A) = c_0I + c_1A + \cdots + c_nA^n$$

则 $p(A)v = p(\lambda)v$

特征空间如何变化?

– $A \rightarrow p(A)$

– 特征空间 $\{\lambda_1, \lambda_2, \dots, \lambda_n\} \rightarrow \{p(\lambda_1), p(\lambda_2), \dots, p(\lambda_n)\}$



矩阵与多项式

考虑任意向量 $x \in R^n$ ，假设 v_1, v_2, \dots, v_n 能张成 R^n ，所以可以展开

$$x = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$$

$$Ax = \lambda_1 \alpha_1 v_1 + \lambda_2 \alpha_2 v_2 + \dots + \lambda_n \alpha_n v_n$$

$$A^2 x = \lambda_1^2 \alpha_1 v_1 + \lambda_2^2 \alpha_2 v_2 + \dots + \lambda_n^2 \alpha_n v_n$$

$$A^3 x = \lambda_1^3 \alpha_1 v_1 + \lambda_2^3 \alpha_2 v_2 + \dots + \lambda_n^3 \alpha_n v_n$$

⋮

$$A^k x = \lambda_1^k \alpha_1 v_1 + \lambda_2^k \alpha_2 v_2 + \dots + \lambda_n^k \alpha_n v_n$$

$$p(A)x = p(\lambda_1)\alpha_1 v_1 + p(\lambda_2)\alpha_2 v_2 + \dots + p(\lambda_n)\alpha_n v_n$$



一个神奇的多项式

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

记 $q(x) = 1 - xp(x)$, 则 $q(A)$ 为全零矩阵。

这样的多项式 $q(x)$ 一定存在!

考虑由 $\{I, A, A^2, A^3, \dots, A^{n^2}\}$ 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 n 次多项式即可



Cayley-Hamilton

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

假设 $q_A(x) = c_0 + c_1x + \cdots + c_nx^n$

类似地定义 $q_A(B) = c_0I + c_1B + \cdots + c_nB^n$

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

非平凡的: 注意 $\det(A - xI)$ 仅仅为一个数



Cayley-Hamilton (选讲)

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

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

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

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

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

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



Cayley-Hamilton (选讲)

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

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

证明(cont'd, sketch): 对于 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 x 的 n 次多项式 $q_A(x) = \det(A - xI)$

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

注意 $q_A(0) = \det(A) \neq \mathbf{0}$

$$A^{-1} = \frac{(-1)^{n-1}}{\det(A)} (A^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1I)$$

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



计算特征值与特征向量

最小化 $f(x) = x^T A x$, subject to $x^T x = 1$

$\det(xI - A)$: 关于 x 的多项式

- n 个根, A 的特征值
- 可能不稳定: 多项式 $(x - 1)(x - 2) \dots (x - 20)$
- 注: $\det(\lambda I - A)$ 非常接近 0 并不意味着 λ 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 $\det(xI - A)$

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



幂迭代 Power method

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

- λ_1 被称为占优特征值 (dominating eigenvalue)
- v_1 : 占优特征向量

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

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

如果 $\alpha_1 \neq 0$ 且 $\lambda_1 = 1$, 随着 $k \rightarrow \infty, A^k x \rightarrow \alpha_1 v_1$



幂迭代 Power method

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

随着 $k \rightarrow \infty$, $A^k x \rightarrow \lambda_1^k \alpha_1 v_1$, 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 $|\lambda_1| > 1$, $A^k x \rightarrow \infty$; 如果 $|\lambda_1| < 1$, $A^k x \rightarrow 0$

解决办法: 每次迭代都进行归一化

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

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?



从特征值到特征向量

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

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

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 $\|Ax - \lambda x\|_2^2$

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

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

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



从特征值到特征向量

最小二乘法：求 λ 以最小化 $\|Ax - \lambda x\|_2^2$

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

法线方程 $\lambda = x^\top Ax / x^\top x$, 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 $\min_{1 \leq j \leq n} |\lambda_j - \lambda| < \epsilon$

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 $x = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$

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

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