



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

上节课:

- 离散傅里叶变换
- 三角级数近似
- 高斯消元



回顾高斯消元法

$$\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \end{array}$$

$$\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ 0 & a_{22} - \frac{a_{21}}{a_{11}}a_{12} & \dots & a_{2n} - \frac{a_{21}}{a_{11}}a_{1n} & b_2 - \frac{a_{21}}{a_{11}}b_1 \end{array}$$

可以一行一行地消，也可以一列一列地消
这里以行为例，主要涉及三种操作

- 交换行（交换两组方程）
- 给一行乘上一个数
- 在一行上减去另一行的一个倍数



回顾高斯消元法：交换行

令 $\sigma \in S_n$ 为 $[n]$ 上的一个置换

$$P_\sigma = \begin{bmatrix} \cdots & e_{\sigma(1)}^\top & \cdots \\ & \vdots & \\ \cdots & e_{\sigma(n)}^\top & \cdots \end{bmatrix}$$

其中向量 e_j 是只有在第 j 个分量上为 1 的单位向量



回顾高斯消元法：给一行乘上一个数

考虑对角矩阵

$$D = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & c_n \end{bmatrix}$$



回顾高斯消元法：在一行上减去另一行的一个倍数

在第 j 行减去第 i 行的 c 倍：

$$E = I - c e_j e_i^T$$

$e_j e_i^T$ 是一个矩阵，里面只有一个元素 (j, i) 是1

这是一个可逆的操作：

$$(I + c e_j e_i^T)(I - c e_j e_i^T) = I$$



回顾高斯消元法

$$Ax = b$$

$$E_1 Ax = E_1 b$$

$$E_2 E_1 Ax = E_2 E_1 b$$

...

$$E_k \cdots E_2 E_1 Ax = E_k \cdots E_2 E_1 b$$

如果一行一行地消，最后得到的将会是上三角阵
而且 $E_k \cdots E_2 E_1$ 是下三角阵

$$A = LU$$



回顾高斯消元法

- 期间可能需要交换行/列，以选择一个非零/更大的主元 (pivoting), 例子:

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

- 交换行: 左乘置换矩阵
- 交换列: 右乘置换矩阵
- 如果找不到非零主元, 则 A 一定是奇异的(singular, non-invertible)
- 每消去一个元素, 最多 $O(n)$ 次算术运算
- 最坏情况下总共需要 $O(n^3)$ 次算术运算

- BLAS (Basic Linear Algebra Subprograms)
- LAPACK



矩阵的范数

如何定义矩阵的范数？

- 方法一：把矩阵看成向量 $\|A\|_F := \sqrt{\sum_i \sum_j a_{ij}^2}$
 - Frobenius norm
 - 这样定义出来的范数用途有限，因为并没有利用上矩阵与向量在本质上的区别——线性映射
- 方法二：把矩阵看成线性映射
 - $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$
 - 给定 $x \in \mathbb{R}^n$, $Ax \in \mathbb{R}^m$
 - 向量范数导出的矩阵的算子范数：把 Ax 相对于 x 的大小作为 A 的大小
 - $\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$
 - 具体的应用：线性方程组的数值稳定性



线性方程组的数值稳定性

给定矩阵 $A \in \mathbb{R}^{m \times n}$ ，向量 $b \in \mathbb{R}^m$ ，求 $x \in \mathbb{R}^n$ 使得 $Ax = b$

- 假设向量 b 有误差，变成 $b + e$ ， x 会怎么变化？
- $x = A^{-1}b \rightarrow x = A^{-1}(b + e)$

- 该线性方程组的条件数 $\text{cond}(A) = \frac{x \text{ 的相对误差}}{b \text{ 的相对误差}}$

记 e 为 b 的误差，则 x 的误差为 $A^{-1}e$ ，任取一向量范数，

$$\begin{aligned} \text{cond}(A) &= \max_{e, b \neq 0} \frac{\|A^{-1}e\| / \|A^{-1}b\|}{\|e\| / \|b\|} = \max_{e, b \neq 0} \frac{\|A^{-1}e\|}{\|e\|} \frac{\|b\|}{\|A^{-1}b\|} \\ &= \max_{e \neq 0} \frac{\|A^{-1}e\|}{\|e\|} \max_{b \neq 0} \frac{\|b\|}{\|A^{-1}b\|} = \max_{e \neq 0} \frac{\|A^{-1}e\|}{\|e\|} \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} \end{aligned}$$

- 矩阵的算子范数 $\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$
- 因此条件数 $\text{cond}(A) = \|A\| \cdot \|A^{-1}\|$



矩阵算子范数 (operator norm)

$$\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\|$$

- 由定义可知: $\forall x \neq \mathbf{0}, \|A\| \geq \frac{\|Ax\|}{\|x\|}$
- 换言之 $\|Ax\| \leq \|A\| \|x\|$

由前面的推导可知 $\text{cond}(A) = \|A\| \cdot \|A^{-1}\|$

- 等号可以取到
- $\text{cond}(c \cdot A) = \text{cond}(A)$

- **注:** 这个条件数的定义适用于任意相容的矩阵范数 (特别地, 这包括所有由向量范数定义的)
- 相容性: $\|A\| \cdot \|A^{-1}\| \geq \|AA^{-1}\| = 1$



矩阵算子范数-例子

$$\|A\|_1 := \max_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \max_{\|x\|=1} \|Ax\| = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{i,j}|$$

“最大的列的和”

证明:

$$Ax = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ A_1 & A_2 & A_3 & \dots & A_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 A_1 + x_2 A_2 + \dots + x_n A_n$$

- 三角不等式: $\|Ax\|_1 \leq \|x_1 A_1\|_1 + \|x_2 A_2\|_1 + \dots + \|x_n A_n\|_1$
- 注意到 $\|x\| = 1$
- $\|Ax\|_1 \leq \max_j \|A_j\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{i,j}|$



矩阵算子范数-例子

$$\|A\|_1 := \max_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{i,j}|$$

$$\|A\|_\infty := \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{i,j}|$$

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

(Courant Fischer)

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

$$\|A\|_{p \rightarrow q} := \max_{x \neq 0} \frac{\|Ax\|_q}{\|x\|_p}$$



解线性方程的迭代方法

高斯消元法适用范围很广

但是计算量 $O(n^3)$ 一般来说太大了

实践中的问题可能有特殊的性质

- 矩阵是稀疏的: 只有 $O(n)$ 个非零元素(non-zeros)
 - 进行高斯消元可能会破坏掉稀疏性
- 对于解有一个比较好的估计范围: A 和 b 随时间动态变化
- 只需要一个近似的解

迭代方法



解线性方程的迭代方法-Jacobi

考虑方程组 $3u + v = 5, u + 2v = 5$

$$u = \frac{5 - v}{3}, \quad v = \frac{5 - u}{2}$$

$$\begin{bmatrix} u_0 \\ v_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} u_1 \\ v_1 \end{bmatrix} = \begin{bmatrix} \frac{5-v_0}{3} \\ \frac{5-u_0}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-0}{3} \\ \frac{5-0}{2} \end{bmatrix} = \begin{bmatrix} \frac{5}{3} \\ \frac{5}{2} \end{bmatrix}$$

$$\begin{bmatrix} u_2 \\ v_2 \end{bmatrix} = \begin{bmatrix} \frac{5-v_1}{3} \\ \frac{5-u_1}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-5/2}{3} \\ \frac{5-5/3}{2} \end{bmatrix} = \begin{bmatrix} \frac{5}{6} \\ \frac{5}{3} \end{bmatrix}$$

$$\begin{bmatrix} u_3 \\ v_3 \end{bmatrix} = \begin{bmatrix} \frac{5-5/3}{3} \\ \frac{5-5/6}{2} \end{bmatrix} = \begin{bmatrix} \frac{10}{9} \\ \frac{25}{12} \end{bmatrix}.$$

最后会收敛到(1,2)



解线性方程的迭代方法-Jacobi

考虑方程组 $u + 2v = 5$, $3u + v = 5$,

$$u = 5 - 2v, \quad v = 5 - 3u$$

$$\begin{bmatrix} u_0 \\ v_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} u_1 \\ v_1 \end{bmatrix} = \begin{bmatrix} 5 - 2v_0 \\ 5 - 3u_0 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} u_2 \\ v_2 \end{bmatrix} = \begin{bmatrix} 5 - 2v_1 \\ 5 - 3u_1 \end{bmatrix} = \begin{bmatrix} -5 \\ -10 \end{bmatrix}$$

$$\begin{bmatrix} u_3 \\ v_3 \end{bmatrix} = \begin{bmatrix} 5 - 2(-10) \\ 5 - 3(-5) \end{bmatrix} = \begin{bmatrix} 25 \\ 20 \end{bmatrix}.$$

这时候迭代是发散的



解线性方程的迭代方法-Jacobi

Jacobi迭代收敛的一个充分条件

我们称矩阵 A 为严格对角占优(strictly diagonally dominant), 如果它满足 $\forall 1 \leq i \leq n, |a_{i,i}| > \sum_{j \neq i} |a_{i,j}|$.

例子: $\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$

非例子: $\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}$

定理: 如果方阵 A 是严格对角占优的, 则 A 是可逆的, 而且对于所有向量 b 和初始猜测, 对线性方程组 $Ax=b$ 使用Jacobi迭代都会收敛到唯一解。



解线性方程的迭代方法-Jacobi

Jacobi迭代的不动点迭代形式

把矩阵 $A = L + D + U$ ，分解成下三角矩阵，对角阵，上三角矩阵

$Ax=b$ 可以重新写成 $(L + D + U)x = b$

$$Dx = b - (L + U)x$$

$$x = D^{-1}(b - (L + U)x)$$

不动点迭代:

$$x_0 = \text{initial vector}$$

$$x_{k+1} = D^{-1}(b - (L + U)x_k) \quad \text{for } k = 0, 1, 2, \dots$$



解线性方程的迭代方法-Gauss-Seidel

$$\begin{aligned} \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} &= \begin{bmatrix} \frac{5-v_0}{3} \\ \frac{5-u_1}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-0}{3} \\ \frac{5-5/3}{2} \end{bmatrix} = \begin{bmatrix} \frac{5}{3} \\ \frac{5}{3} \end{bmatrix} \\ \begin{bmatrix} u_2 \\ v_2 \end{bmatrix} &= \begin{bmatrix} \frac{5-v_1}{3} \\ \frac{5-u_2}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-5/3}{3} \\ \frac{5-10/9}{2} \end{bmatrix} = \begin{bmatrix} \frac{10}{9} \\ \frac{35}{18} \end{bmatrix} \\ \begin{bmatrix} u_3 \\ v_3 \end{bmatrix} &= \begin{bmatrix} \frac{5-v_2}{3} \\ \frac{5-u_3}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-35/18}{3} \\ \frac{5-55/54}{2} \end{bmatrix} = \begin{bmatrix} \frac{55}{54} \\ \frac{215}{108} \end{bmatrix}. \end{aligned}$$

注意：与Jacobi的区别在于，对 v_1 的计算中使用了 u_1 而不是 u_0

$$\begin{aligned} Ax=b &\text{可以重新写成 } (L + D + U)x = b \\ (L + D)x &= b - Ux \end{aligned}$$

x_0 = initial vector

$$x_{k+1} = D^{-1}(b - Ux_k - Lx_{k+1}) \quad \text{for } k = 0, 1, 2, \dots$$



解线性方程的迭代方法-Gauss-Seidel

注意区别在于，对 v_1 的计算中使用了 u_1 而不是 u_0

$$\begin{aligned}Ax=b & \text{可以重新写成 } (L + D + U)x = b \\ (L + D)x & = b - Ux\end{aligned}$$

$x_0 =$ initial vector

$$x_{k+1} = D^{-1}(b - Ux_k - Lx_{k+1}) \quad \text{for } k = 0, 1, 2, \dots$$

定理：如果方阵 A 是严格对角占优的，则 A 是可逆的，而且对于所有向量 b 和初始猜测，对线性方程组 $Ax=b$ 使用Gauss Seidel迭代都会收敛到唯一解。



线性迭代方法的收敛性

- 考虑迭代方程 $x_{k+1} = Ax_k + b$
- 令 x_* 为其中一个不动点
- $x_{k+1} - x_* = A(x_k - x_*)$
- $x_k - x_* = A^k(x_0 - x_*)$
- 什么时候收敛?
 - $A^k \rightarrow 0$ 当且仅当对任意初始值都收敛
 - x_0 选得好, 使得 $x_0 - x_*$ 落在一个好的线性子空间里面



线性迭代方法的收敛性

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 $A^k \rightarrow 0$ 的时候, 对任意初始值都收敛?

方法一: $\|x_k - x_*\| = \|A^k(x_0 - x_*)\| \leq \|A^k\| \|x_0 - x_*\|$

方法二: $x_k = Ax_{k-1} + b = A(Ax_{k-2} + b) + b = \dots$

- $x_k = A^k x_0 + (A^{k-1} + A^{k-2} + \dots + I)b$
 $= A^k x_0 + (I - A)^{-1}(I - A^k)b$
- $A^k \rightarrow 0$ 时, 对任意初始值都有 $x_k \rightarrow (I - A)^{-1}b$
- 存在唯一的不动点: $x_* = Ax_* + b \Rightarrow x_* = (I - A)^{-1}b$



线性迭代方法的收敛性

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 $A^k \rightarrow 0$?

- 对任意初始值都收敛，即 $\forall x_0, x_k - x_* = A^k(x_0 - x_*) \rightarrow 0$
- 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



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



线性迭代方法的收敛性-谱半径

如果 $\lambda \in \mathbb{C}$ 和向量 \boldsymbol{v} 满足 $A\boldsymbol{v} = \lambda\boldsymbol{v}$ ，则称 λ 为一个特征值，向量 \boldsymbol{v} 为对应的特征向量

注意 $A\boldsymbol{v} = \lambda\boldsymbol{v} \Leftrightarrow (A - \lambda I)\boldsymbol{v} = \mathbf{0}$ ，所以求特征值的问题可以转化为求多项式 $\det(A - \lambda I)$ 关于 $\lambda \in \mathbb{C}$ 的根

谱半径 $\rho(A) := \max\{|\lambda_1|, |\lambda_2|, \dots, |\lambda_n|\}$

定理：谱半径 $\rho(A) < 1$ 当且仅当 $\lim_{k \rightarrow \infty} A^k = 0$



线性迭代方法的收敛性-谱半径

谱半径 $\rho(A) := \max\{|\lambda_1|, |\lambda_2|, \dots, |\lambda_n|\}$

定理： 谱半径 $\rho(A) < 1$ 当且仅当 $\lim_{k \rightarrow \infty} A^k = 0$

证明： (\Leftarrow) 设 $\lambda \in \mathbf{C}$ 和向量 \boldsymbol{v} 满足 $A\boldsymbol{v} = \lambda\boldsymbol{v}$

$$\begin{aligned} 0 &= \left(\lim_{k \rightarrow \infty} A^k\right)\boldsymbol{v} = \lim_{k \rightarrow \infty} (A^k\boldsymbol{v}) = \lim_{k \rightarrow \infty} \lambda^k\boldsymbol{v} \\ &= \boldsymbol{v} \cdot \lim_{k \rightarrow \infty} \lambda^k \end{aligned}$$

即 $\lim_{k \rightarrow \infty} \lambda^k = 0$



线性迭代方法的收敛性-谱半径

谱半径 $\rho(A) := \max\{|\lambda_1|, |\lambda_2|, \dots, |\lambda_n|\}$

定理: 谱半径 $\rho(A) < 1$ 当且仅当 $\lim_{k \rightarrow \infty} A^k = 0$

证明(sketch): (\Rightarrow) 假设 A 的特征向量 v_1, v_2, \dots, v_n 可以张成 R^n , 则任意向量 u 可以写成 v_1, v_2, \dots, v_n 的线性组合

$$u = a_1 v_1 + a_2 v_2 + \dots + a_n v_n$$

进而 $A^k u = A^k (a_1 v_1 + a_2 v_2 + \dots + a_n v_n)$
 $= a_1 \lambda_1^k v_1 + \dots + a_n \lambda_n^k v_n \rightarrow 0$

注: 这里的证明仅讨论了 A 可以被对角化的情况。更一般的情况下, 可以使用 Jordan 标准形分解, 或者利用可对角化矩阵的稠密性质(dense);



线性迭代方法的收敛性

考虑迭代方程 $x_{k+1} = Ax_k + b$

令 x_* 为其中一个不动点

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

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

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



Jacobi迭代的谱半径

Jacobi迭代的谱半径

$x_0 =$ initial vector

$$x_{k+1} = D^{-1}(b - (L + U)x_k) \quad \text{for } k = 0, 1, 2, \dots$$

定理：对于对角占优矩阵， $\rho(D^{-1}(L + U)) < 1$

- 设 λ 为 $D^{-1}(L + U)$ 任意一个特征值，向量 v 为对应的特征向量，则 $D^{-1}(L + U)v = \lambda v$
- 或者说 $(L + U)v = \lambda Dv$
- 不失一般性地假设 $\|v\|_\infty = 1$
- 设 $v_m = 1, 1 \leq m \leq n$

记 $R = (L + U)$ ，考虑 $Rv = \lambda Dv$ 的第 m 个分量

- 右边 = $\lambda d_m v_m = \lambda d_m$
- 注意到左边矩阵 R 的对角线上的元素为0
- |左边| = $|r_{m1}v_1 + r_{m2}v_2 + \dots + r_{m,(m-1)}v_{m-1} + r_{m,(m+1)}v_{m+1} + \dots + r_{mn}v_n| \leq \sum_{j \neq m} |r_{mj}| < |d_m|$
- 左边=右边则意味着 $|\lambda| < 1$
- 因为 λ 是任取的一个特征值，所以谱半径 < 1



解线性方程的迭代方法-谱半径

Gauss-Seidel迭代的谱半径 $\rho((L + D)^{-1}U)$

$$(L + D)x = b - Ux$$

对严格对角占优的矩阵，也有

$$\rho\left((L + D)^{-1}U\right) < 1$$

证明思路类似



下节课

- 正定矩阵
- Courant-Fischer: 特征值的min-max刻画
- 矩阵的多项式