第五章 共轭梯度法

杜磊
dulei@dlut.edu.cn

大连理工大学 数学科学学院
创新园大厦 B1207

2019 年 10 月 22 日

内容提要

1. 最速下降法
2. 共轭梯度法及其基本性质
3. 实用共轭梯度法及其收敛性
4. 预优共轭梯度法
5. Krylov 子空间法

最速下降法

定理

设 A 对称正定, 求方程组 $Ax = b$ 的解等价于求二次泛函 $\phi(x) = x^{\mathrm{T}}Ax - 2b^{\mathrm{T}}x$ 的极小值点.

最速下降法

定理

设 A 对称正定，求方程组 $Ax=b$ 的解等价于求二次泛函 $\phi(x)=x^{\mathrm{T}}Ax-2b^{\mathrm{T}}x$ 的极小值点.

Algorithm 2 最速下降法

设定初始向量 x_0;
$r_0=b-Ax_0$, $k=0$;
while $r_k\neq 0$ **do**
  $k=k+1$;
  $\alpha_{k-1}=\dfrac{r_{k-1}^{\mathrm{T}}r_{k-1}}{r_{k-1}^{\mathrm{T}}Ar_{k-1}}$;
  $x_k=x_{k-1}+\alpha_{k-1}r_{k-1}$;
  $r_k=b-Ax_k$; $(r_k=r_{k-1}-\alpha_{k-1}Ar_{k-1})$
end while

最速下降法

定理

设 A 的特征值为 $0 < \lambda_1 \leq \cdots \leq \lambda_n$, 则由最速下降法产生的序列 $\{x_k\}$ 满足

$$\|x_k - x_*\|_A \leq (\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1})^k \|x_0 - x_*\|_A ,$$

其中 $x_* = A^{-1}b, \|x\|_A = \sqrt{x^{\mathrm{T}} A x}.$

最速下降法

定理

设 A 的特征值为 $0 < \lambda_1 \le \cdots \le \lambda_n$，则由最速下降法产生的序列 $\{x_k\}$ 满足

$$\|x_k - x_*\|_A \le \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^k \|x_0 - x_*\|_A\,,$$

其中 $x_* = A^{-1}b$, $\|x\|_A = \sqrt{x^{\mathrm{T}}Ax}$.

引理

设 A 的特征值为 $0 < \lambda_1 \le \cdots \le \lambda_n$，$P(t)$ 是一个 t 的多项式，则

$$\|P(A)x\|_A \le \max_{1\le i\le n} |P(\lambda_i)|\,\|x\|_A\,,\ x \in \mathbb{R}^n.$$

Chebyshev 多项式及其性质

当权函数 $\rho(x)=\frac{1}{\sqrt{1-x^2}}$, 区间为 $[-1,1]$ 时, 由序列 $\{1,x,\cdots,x^n,\cdots\}$ 正交化得到的正交多项式为 Chebyshev 多项式, 表示为

$$T_n(x)=\cos(n\arccos x),\ \ |x|\le 1.$$

Chebyshev 多项式及其性质

当权函数 $\rho(x)=\frac{1}{\sqrt{1-x^2}}$, 区间为 $[-1,1]$ 时, 由序列 $\{1,x,\cdots,x^n,\cdots\}$ 正交化得到的正交多项式为 Chebyshev 多项式, 表示为

$$T_n(x)=\cos(n\arccos x),\ \ |x|\leq 1.$$

Chebyshev 多项式性质:

- $T_{n+1}(x)=2xT_n(x)-T_{n-1}(x),\ n=1,2,\cdots,\ T_0(x)=1,\ T_1(x)=x.$
- $\int_{-1}^{1}\frac{T_n(x)T_m(x)}{\sqrt{1-x^2}}\mathrm{d}x=\begin{cases}0, n\neq m;\\ \dfrac{\pi}{2}, n=m\neq 0;\\ \pi, n=m=0.\end{cases}$
- $T_{2k}(x)$ 只含 x 的偶次幂; $T_{2k}(x)$ 只含 x 的偶次幂; $T_{2k+1}(x)$ 只含 x 的偶次幂;
- $T_n(x)$ 在区间 $[-1,1]$ 上有 n 个零点 $x_k=\cos\frac{2k-1}{2n}\pi$;

Chebyshev 多项式及其性质

若考虑 $|x| \geq 1$ 情况, Chebyshev 多项式可表示为

$$T_n(x) = \frac{(x+\sqrt{x^2-1})^n + (x-\sqrt{x^2-1})^k}{2}.$$

若考虑 $|x| \geq 1$ 情况，Chebyshev 多项式可表示为

$$T_n(x) = \frac{(x+\sqrt{x^2-1})^n + (x-\sqrt{x^2-1})^k}{2}.$$

定理

设 $a, b, c \in \mathbb{R}$ 满足 $a < b$ 且 $|c| \notin [a, b]$，则下面的最小最大问题

$$\min_{p(\lambda)\in\mathbb{P}_k, p(c)=1} \max_{a\leq\lambda\leq b} |p(\lambda)|$$

的唯一解为

$$\tilde{P}_k(\lambda) = \frac{T_k(\frac{b+a-2\lambda}{b-a})}{T_k(\frac{b+a-2c}{b-a})}.$$

内容提要

1 最速下降法

2 共轭梯度法及其基本性质

3 实用共轭梯度法及其收敛性

4 预优共轭梯度法

5 Krylov 子空间法

共轭梯度法

给定初始向量 x_0, 第一步仍选负梯度方向 $p_0 = r_0$. 以后各步不在选取负梯度方向 r_k, 而是在过点 x_k 由向量 r_k 和 p_{k-1} 所张成的二维平面

$$\pi_2 = \{x = x_k + \xi r_k + \eta p_{k-1} : \xi, \eta \in \mathbb{R}\}$$

内找出使函数 ϕ 下降最快的方向作为新的下山方向 p_k, 即

$$\begin{aligned}\psi(\xi,\eta) &= \phi(x_k + \xi r_k + \eta p_{k-1}) \\ &= (x_k + \xi r_k + \eta p_{k-1})^{\mathrm{T}} A (x_k + \xi r_k + \eta p_{k-1}) \\ &- 2b^{\mathrm{T}}(x_k + \xi r_k + \eta p_{k-1}).\end{aligned}$$

分别对 ξ, η 求偏导可得

$$\begin{aligned}\frac{\partial\psi}{\partial\xi} &= 2(\xi r_k^{\mathrm{T}} A r_k + \eta r_k^{\mathrm{T}} A p_{k-1} - r_k^{\mathrm{T}} r_k) \\ \frac{\partial\psi}{\partial\eta} &= 2(\xi r_k^{\mathrm{T}} A p_{k-1} + \eta p_{k-1}^{\mathrm{T}} A p_{k-1} - \color{red}{r_k^{\mathrm{T}} p_{k-1}})\end{aligned}$$

共轭梯度法

Algorithm 3 共轭梯度法

设定初始向量 x_0;
$r_0 = b - Ax_0,\ k = 0$;
while $r_k \neq 0$ **do**
    $\alpha_k = \frac{r_k^{\mathrm{T}} p_k}{p_k^{\mathrm{T}} A p_k}$;
    $x_{k+1} = x_k + \alpha_k p_k$;
    $r_{k+1} = b - Ax_{k+1}$;
    $\beta_k = -\frac{r_{k+1}^{\mathrm{T}} A p_k}{p_k^{\mathrm{T}} A p_k}$;
    $p_{k+1} = r_{k+1} + \beta_k p_k$;
    $k = k + 1$;
end while

共轭梯度法

Algorithm 4 共轭梯度法

```
设定初始向量 x_0;
r_0 = b − A x_0, k = 0; p_0 = r_0;
while r_k ≠ 0 do
    α_k = (r_k^T r_k) / (p_k^T A p_k);
    x_{k+1} = x_k + α_k p_k;
    r_{k+1} = r_k − α_k A p_k
    β_k = (r_{k+1}^T r_{k+1}) / (r_k^T r_k);
    p_{k+1} = r_{k+1} + β_k p_k;
    k = k + 1;
end while
```

基本性质

定理

由共轭梯度法得到的向量组 $\{r_i\}$ 和 $\{p_i\}$ 具有下面的性质:

1. $p_i^{\mathrm{T}} r_j = 0,\ 0 \le i < j \le k;$
2. $r_i^{\mathrm{T}} r_j = 0,\ i \neq j, 0 \le i, j \le k;$
3. $p_i^{\mathrm{T}} A p_j = 0,\ i \neq j, 0 \le i, j \le k;$
4. $\mathrm{span}\{r_0, \cdots, r_k\} = \mathrm{span}\{p_0, \cdots, p_k\} = \mathcal{K}(A, r_0, k+1)$, 其中$\mathcal{K}(A, r_0, k+1) = \mathrm{span}\{r_0, Ar_0, \cdots, A^k r_0\}$, 通常称之为Krylov子空间.

基本性质

定理

由共轭梯度法得到的向量组 $\{r_i\}$ 和 $\{p_i\}$ 具有下面的性质:

1. $p_i^{\mathrm{T}} r_j = 0,\ 0 \le i < j \le k;$
2. $r_i^{\mathrm{T}} r_j = 0,\ i \ne j, 0 \le i, j \le k;$
3. $p_i^{\mathrm{T}} A p_j = 0,\ i \ne j, 0 \le i, j \le k;$
4. $\mathrm{span}\{r_0, \cdots, r_k\} = \mathrm{span}\{p_0, \cdots, p_k\} = \mathcal{K}(A, r_0, k+1)$, 其中$\mathcal{K}(A, r_0, k+1) = \mathrm{span}\{r_0, Ar_0, \cdots, A^k r_0\}$, 通常称之为Krylov子空间.

定理

用共轭梯度法计算得到的近似解 x_k 满足

$$\phi(x_k) = \min\{\phi(x) : x \in x_0 + \mathcal{K}(A, r_0, k)\}$$

或

$$\|x_k - x_*\|_A = \min\{\|x - x_*\|_A : x \in x_0 + \mathcal{K}(A, r_0, k)\},$$

其中 $\|x\|_A = \sqrt{x^{\mathrm{T}} A x}$, $Ax_* = b$, $\mathcal{K}(A, r_0, k) = \mathrm{span}\{r_0, Ar_0, \cdots, A^k r_0\}$.

内容提要

1 最速下降法

2 共轭梯度法及其基本性质

3 实用共轭梯度法及其收敛性

4 预优共轭梯度法

5 Krylov 子空间法

实用共轭梯度法

Algorithm 5 实用共轭梯度法

```
设定初始向量 x;
r = b − Ax, k = 0; ρ = rᵀr;
while √ρ > ϵ‖b‖₂ and k < k_max do
    k = k + 1;
    if k=1 then
        p = r;
    else
        β = ρ/ρ̃; p = r + βp;
    end if
    w = Ap; α = ρ/pᵀw; x = x + αp;
    r = r − αw; ρ̃ = ρ; ρ = rᵀr;
end while
```

收敛性分析

定理

如果 $A = I + B$, 而且 $\mathrm{rank}(B) = r$, 则共轭梯度法至多迭代 $r+1$ 步即可得到方程组 $Ax = b$ 的精确解.

收敛性分析

定理

如果 $A = I + B$, 而且 $\text{rank}(B) = r$, 则共轭梯度法至多迭代 $r + 1$ 步即可得到方程组 $Ax = b$ 的精确解.

定理

用共轭梯度法求得的 x_k 有如下的误差估计:

$$\|x_k - x_*\|_A \le 2\left(\frac{\sqrt{\kappa_2} - 1}{\sqrt{\kappa_2} + 1}\right)^k \|x_0 - x_*\|_A ,$$

其中 $\kappa_2 = \kappa_2(A) = \|A\|_2 \left\|A^{-1}\right\|_2$.

内容提要

1 最速下降法

2 共轭梯度法及其基本性质

3 实用共轭梯度法及其收敛性

4 预优共轭梯度法

5 Krylov 子空间法

预优共轭梯度法 (PCG)

设 L 非奇异，$L^{-1}AL^{-\mathrm{T}}L^{\mathrm{T}}x = L^{-1}b$，记 $M = LL^{\mathrm{T}}$.

Algorithm 6 共轭梯度法

设定初始向量 x_0;
$r_0 = b - Ax_0$, $k = 0$; $p_0 = r_0$;
while $r_k \neq 0$ **do**
  $\alpha_k = \dfrac{r_k^{\mathrm{T}} r_k}{p_k^{\mathrm{T}} A p_k}$;
  $x_{k+1} = x_k + \alpha_k p_k$;
  $r_{k+1} = r_k - \alpha_k A p_k$
  $\beta_k = \dfrac{r_{k+1}^{\mathrm{T}} r_{k+1}}{r_k^{\mathrm{T}} r_k}$;
  $p_{k+1} = r_{k+1} + \beta_k p_k$;
  $k = k + 1$;
end while

Algorithm 7 预优共轭梯度法

设定初始向量 x_0;
$r_0 = b - Ax_0$, $k = 0$; $p_0 = r_0$;
while $r_k \neq 0$ **do**
  $\tilde{\alpha}_k = \dfrac{r_k^{\mathrm{T}} M^{-1} r_k}{p_k^{\mathrm{T}} A p_k}$;
  $x_{k+1} = x_k + \tilde{\alpha}_k p_k$;
  $r_{k+1} = r_k - \tilde{\alpha}_k A p_k$
  $\tilde{\beta}_k = \dfrac{r_{k+1}^{\mathrm{T}} M^{-1} r_{k+1}}{r_k^{\mathrm{T}} M^{-1} r_k}$;
  $p_{k+1} = M^{-1} r_{k+1} + \tilde{\beta}_k p_k$;
  $k = k + 1$;
end while

预优矩阵 M 选取

- 基本准则:
  - M 是对称正定的;
  - M 是稀疏的;
  - $\lambda(M^{-1}A)$ 分布集中;
  - 方程组 $Mz = r$ 易于求解.
- 常用预优矩阵:
  - 对角矩阵;
  - 不完全 Cholesky 因子分解;
  - 多项式预优矩阵;
  - $\cdots\cdots$

内容提要

1 最速下降法

2 共轭梯度法及其基本性质

3 实用共轭梯度法及其收敛性

4 预优共轭梯度法

5 Krylov 子空间法

Krylov 子空间

定义

设 $A \in \mathbb{R}^{n\times n}, r \in \mathbb{R}^n$, 则由 A 和 r 生成的 m 阶 Krylov 子空间为

$$\mathcal{K}_m(A, r) = \mathrm{span}\{r, Ar, A^2r, \cdots, A^{m-1}r\}.$$

构造 Krylov 子空间法的两个关键问题:

- 如何基底基底?
- 如何构造近似解?

Arnoldi 过程

Algorithm 8 Arnoldi 过程 (Gram-Schmidt)

设定 $v_1 = r/\|r\|_2$;
for $j = 1, 2, \ldots, m$ **do**
  计算 $h_{ij} = (Av_j, v_i), i = 1, 2, \ldots, j$;
  计算 $w_j = Av_j - \sum_{i=1}^{j} h_{ij} v_i$;
  $h_{j+1,j} = \|w_j\|_2$;
  如果 $h_{j+1,j} = 0$, 循环停止.
  $v_{j+1} = w_j / h_{j+1,j}$;
end for

引理

如果程序 *(8)* 前 m 步没有中断, 则 $v_1, v_2, \ldots, v_m$ 为 Krylov 子空间 $\mathcal{K}_m(A, r) = \mathrm{span}\{r, Ar, A^2r, \cdots, A^{m-1}r\}$ 的一组标准正交基底.

引理

记 $V_m = [v_1, v_2, \ldots, v_m]$,

$$H_m = \begin{bmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1m} \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2m} \\ & h_{32} & h_{33} & \cdots & h_{3m} \\ & & \ddots & \ddots & \vdots \\ & & & h_{m-1,m} & h_{mm} \end{bmatrix}, \tilde{H}_m = \begin{bmatrix} H_m \\ h_{m+1,m} \end{bmatrix},$$

则有下列关系成立:

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^{\mathrm{T}} = V_{m+1} \tilde{H}_m,$$

$$V_m^{\mathrm{T}} A V_m = H_m.$$

Arnoldi 过程

Algorithm 9 Arnoldi 过程 (Modified Gram-Schmidt)

设定 $v_1 = r/\|r\|_2$;
for $j = 1, 2, \ldots, m$ **do**
  计算 $w_j = Av_j$;
  for $i = 1, 2, \ldots, j$ **do**
    $h_{ij} = (w_j, v_i)$;
    $w_j = w_j - h_{ij}v_i$;
  end for
  $h_{j+1,j} = \|w_j\|_2$;
  如果 $h_{j+1,j} = 0$, 循环停止.
  $v_{j+1} = w_j/h_{j+1,j}$;
end for

Lanczos 过程

如果 A 为对称矩阵，则 H_m 为对称三对角矩阵

引理

记 $V_m = [v_1, v_2, \ldots, v_m]$,

$$H_m = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \ddots & \ddots & \\ & \ddots & \ddots & \beta_{m-1} \\ & & \beta_{m-1} & \alpha_m \end{bmatrix}, \tilde{H}_m = \begin{bmatrix} H_m \\ \beta_m \end{bmatrix},$$

则有下列关系成立:

$$\begin{aligned} A V_m &= V_m H_m + \beta_m v_{m+1} e_m^{\mathrm{T}} = V_{m+1} \tilde{H}_m, \\ V_m^{\mathrm{T}} A V_m &= H_m. \end{aligned}$$

Lanczos 过程

Algorithm 10 Lanczos 过程

```
设定 v_0 = 0, β_0 = 0, v_1 = r/‖r‖_2;
for j = 1, 2, ..., m − 1 do
    计算 z = Av_j;
    α_j = v_j^T z;
    z = z − α_j v_j − β_{j−1} v_{j−1};
    β_j = ‖z‖_2;
    如果 β_j = 0, 循环停止.
    v_{j+1} = z/β_j;
end for
```

相关文献

D.P. O'Leary
Some History of Conjugate Gradients and Other Krylov Subspace Methods
https://archive.siam.org/meetings/la09/talks/oleary.pdf.

G.H. Golub & D.P. O'Leary
Some History of the Conjugate Gradient and Lanczos Algorithms: 1948–1976
SIAM Review, 31:1(1989), 50-102.