最速下降法和共轭梯度法

教师: 胡俊

北京大学数学科学学院

October 24, 2019

目录

目录

最速下降法

考虑线性方程组

$$Ax = b \tag{1.1}$$

的求解问题，其中矩阵A是对称正定的. 定义二次泛函

$$\varphi(x) = x^T Ax - 2b^T x. \tag{1.2}$$

定理： 设A是对称正定的，求方程组$Ax = b$的解等价于求解二次泛函$\varphi(x)$的极小值点.

证明： 直接计算可得

$$\frac{\partial \varphi}{\partial x_i} = 2(a_{i1}x_1 + \cdots + a_{in}x_n) - 2b_i, i = 1, \cdots, n.$$

令$r = b - Ax$，则有

$$\mathrm{grad}\varphi(x) = 2(Ax - b) = -2r.$$

若$\varphi(x)$在x_*达到极小，则必有$\mathrm{grad}\varphi(x_*) = 0$，从而有$Ax_* = b$.

反之，若x_*是方程的解，则对任意向量y，有

$$\begin{aligned}\varphi(x_*+y) &= (x_*+y)^T A(x_*+y) - 2b^T(x_*+y) \\ &= x_*^T A x_* - 2b^T x_* + y^T A y = \varphi(x_*) + y^T A y.\end{aligned}$$

当A正定时，有$y^T Ay > 0$，因此

$$\varphi(x_*+y) \geq \varphi(x_*)$$

即x_*使得$\varphi(x)$到极小. 证毕.

类盲人下山法

先任意给定一个初始向量x_0，确定一个下山的方向p_0，沿着经过点x_0而方向为p_0 的直线$x = x_0 + \alpha p_0$找一个点

$$x_1 = x_0 + \alpha_0 p_0,$$

使得对所有实数α, 有

$$\varphi(x_0 + \alpha_0 p_0) \leq \varphi(x_0 + \alpha p_0).$$

也就是说，在这条直线上，x_1使$\varphi(x)$到达极小. 然后从x_1出发，再确定一个下山的方向p_1, 沿着直线

$$x = x_1 + \alpha p_1$$

再跨一步，即找到α_1使得$\varphi(x)$在$x_2 = x_1 + \alpha_1 p_1$达到极小：

$$\varphi(x_1 + \alpha_1 p_1) \leq \varphi(x_1 + \alpha p_1).$$

如此下去，得到一个序列：

$$(p_0, \alpha_0), (p_1, \alpha_1), (p_2, \alpha_2), \cdots, .$$

称p_k为搜索方向，α_k为步长. 一般地，先在x_k点招一个下山方向p_k, 再在直线$x = x_k + \alpha p_k$上确定步长α_k使得

$$\varphi(x_k + \alpha_k p_k) \leq \varphi(x_k + \alpha p_k).$$

最后求出$x_{k+1} = x_k + \alpha_k p_k$.

步长的确定

令

$$
\begin{aligned}
f(\alpha) &= \varphi(x_k + \alpha p_k) \\
&= (x_k + \alpha p_k)^T A(x_k + \alpha p_k) - 2b^T(x_k + \alpha p_k) \\
&= \alpha^2 p_k^T A p_k - 2\alpha r_k^T p_k + \varphi(x_k),
\end{aligned}
$$

其中$r_k = b - Ax_k$.

$$f'(\alpha) = 2\alpha p_k^T A p_k - 2 r_k^T p_k = 0.$$

因此

$$\alpha_k = \frac{r_k^T p_k}{p_k^T A p_k}, x_{k+1} = x_k + \alpha_k p_k.$$

注：由$\alpha_k p_k^T A p_k - r_k^T p_k = 0$,
和$x_{k+1} = x_k + \alpha_k p_k$即$r_{k+1} = r_k - \alpha_k A p_k$，可得$p_k^T r_{k+1} = 0$.

下山方向的确定

由于

$$\begin{aligned}\varphi(x_{k+1}) - \varphi(x_k) &= \alpha_k^2 p_k^T A p_k - 2\alpha_k r_k^T p_k \\ &= -\frac{(r_k^T p_k)^2}{p_k^T A p_k}.\end{aligned}$$

因此，只要$r_k^T p_k \neq 0$, 就有$\varphi(x_{k+1}) < \varphi(x_k)$.
因为梯度方向是增加最快的方向，因此负梯度方向是下降最快的方向. 因而，取$p_k = r_k$.

算法及收敛性

算法(解对称正定方程组: **最速下降法**)

x_0=初值

$r_0 = b - Ax_0$; k=0

While $r_k \neq 0$

$k = k + 1$

$\alpha_{k-1} = r_{k-1}^T r_{k-1} / r_{k-1}^T A r_{k-1}$

$x_k = x_{k-1} + \alpha_{k-1} r_{k-1}$

$r_k = b - Ax_k$

End

最速下降方法的收敛性

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

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

证明： 设$y_1, y_2, \cdots, y_n$是A的对应于$\lambda_1, \lambda_2, \cdots, \lambda_n$的特征向量所构成的$\mathbb{R}^n$ 一组标准正交基, 则对任意$x \in \mathbb{R}^n$, 有$x = \sum\limits_{i=1}^{n} \beta_i y_i$, 从而有

$$x^T P(A) A P(A) x = \left(\sum_{i=1}^{n} \beta_i P(\lambda_i) y_i \right)^T A \left(\sum_{i=1}^{n} \beta_i P(\lambda_i) y_i \right)$$

$$= \sum_{i=1}^{n} \lambda_i \beta_i^2 P^2(\lambda_i) \le \max_{1\le i\le n} P^2(\lambda_i) \sum_{i=1}^{n} \lambda_i \beta_i^2 = \max_{1\le i\le n} P^2(\lambda_i) x^T A x.$$

定理： 设A的特征值为$0 < \lambda_1 \leq \cdots \leq \lambda_n$, 则上述算法产生的序列$\{x_k\}$满足

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

证明：

$$\varphi(x_k) \leq \varphi(x_{k-1} + \alpha p_{k-1}), \forall \alpha \in \mathbb{R}.$$

并注意到

$$\varphi(x) + x_*^T A x_* = (x - x_*)^T A (x - x_*),$$

有

$$\begin{aligned}
&(x_k - x_*)^T A (x_k - x_*) \\
&\leq (x_{k-1} + \alpha r_{k-1} - x_*)^T A (x_{k-1} + \alpha r_{k-1} - x_*) \\
&= [(I - \alpha A)(x_{k-1} - x_*)]^T A [(I - \alpha A)(x_{k-1} - x_*)]
\end{aligned}$$

对任意$\alpha \in \mathbb{R}$成立.

记$P_\alpha(t) = 1 - \alpha t$, 由上述引理，有

$$\|x_k - x_*\|_A \leq \min_\alpha \max_{1\leq i\leq n} |P_\alpha(\lambda_i)| \|x_{k-1} - x_*\|_A.$$

因为

$$\max_{1\leq i\leq n} |P_\alpha(\lambda_i)| = \max(|1-\alpha\lambda_1|, |1-\alpha\lambda_n|)$$

显然，当

$$1 - \alpha\lambda_1 = \alpha\lambda_n - 1$$

时，上述最大值取得最小，此时

$$\alpha = \frac{2}{\lambda_1 + \lambda_n}.$$

这样

$$\|x_k - x_*\|_A \leq \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \|x_{k-1} - x_*\|_A.$$

证毕.

目录

共轭梯度法

对最速下降法做一简单的分析就会发现，负梯度方向虽从局部来看是最佳的下山方向，但从整体来看并非最佳. 这就促使人们去寻求更好的下山方向. 当然，我们自然希望每步确定新的下山方向所付出的代价不要太大. **共轭梯度法**就是根据这一思想设计的，其具体计算过程如下：

给定初始向量x_0，第一步仍选负梯度方向为下山方向，即$p_0 = r_0$，于是有

$$\alpha_0 = \frac{r_0^T r_0}{p_0^T A p_0}, \quad x_1 = x_0 + \alpha_0 p_0, \quad r_1 = b - A x_1.$$

对以后各步，例如，第$k+1 (k \geq 1)$步，下山方向就不再取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

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

内找出使函数φ下降最快的方向作为新的下山方向p_k.

考虑φ在π_2上的限制：

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

直接计算可得

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

其中最后一式用到了$r_k^T p_{k-1}=0$, 这可由r_k的定义直接验证. 令

$$
\frac{\partial\psi}{\partial\xi}=\frac{\partial\psi}{\partial\eta}=0,
$$

即知φ在π_2内有唯一的极小值点

$$
\tilde{x}=x_k+\xi_0 r_k+\eta_0 p_{k-1},
$$

其中ξ_0和η_0满足

$$\begin{cases} \xi_0 r_k^T A r_k + \eta_0 r_k^T A p_{k-1} = r_k^T r_k, \\ \xi_0 r_k^T A p_{k-1} + \eta_0 p_{k-1}^T A p_{k-1} = 0. \end{cases} \tag{5.2.1}$$

注意，上式蕴含着$r_k \neq 0$必有$\xi_0 \neq 0$，因此我们可取

$$p_k = \frac{1}{\xi_0}(\tilde{x} - x_k) = r_k + \frac{\eta_0}{\xi_0} p_{k-1}$$

作为新的下山方向. 显然这是在平面π_2内可得到的最佳下山方向. 令$\beta_{k-1} = \frac{\eta_0}{\xi_0}$，则由(5.2.1)式的第二个方程得

$$\beta_{k-1} = -\frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}.$$

注意这样确定的p_k满足$p_k^T A p_{k-1} = 0$，即所谓的p_k 与p_{k-1}是相互共轭的.

p_k确定以后, α_k的确定仍用前面的公式, 然后计算$x_{k+1}=x_k+\alpha_k p_k$. 总结上面的讨论, 可得如下的计算公式:

$$\begin{aligned} \alpha_k &= \frac{r_k^T p_k}{p_k^T A p_k}, & x_{k+1} &= x_k+\alpha_k p_k \\ & & r_{k+1} &= b - A x_{k+1}, \\ \beta_k &= -\frac{r_{k+1}^T A p_k}{p_k^T A p_k}, & p_{k+1} &= r_{k+1}+\beta_k p_k. \end{aligned} \tag{5.2.2}$$

在实际计算中, 常将上述公式进一步简化, 从而得到一个形式上更为简单而且对称的计算公式. 首先来简化r_{k+1}的计算公式:

$$\begin{aligned} r_{k+1} &= b - A x_{k+1} = b - A(x_k+\alpha_k p_k) \\ &= r_k - \alpha_k A p_k. \end{aligned} \tag{5.2.3}$$

因为Ap_k在计算α_k时已经求出, 所以计算r_{k+1}时可以不必将x_{k+1}代入方程去计算, 而是从递推关系(5.2.3)得到.

再来简化α_k和β_k的计算公式. 我们需要用到下面的关系式:

$$r_k^T r_{k+1} = r_k^T p_{k-1} = r_{k+1}^T p_k = 0, \ \ k = 1, 2, \cdots . \tag{5.2.4}$$

这些关系式的证明包含在定理5.2.1的证明中. 从(5.2.4)式和(5.2.3)式可导出

$$r_{k+1}^T A p_k = \frac{1}{\alpha_k} r_{k+1}^T (r_k - r_{k+1}) = -\frac{1}{\alpha_k} r_{k+1}^T r_{k+1},$$

$$\begin{aligned} p_k^T A p_k &= \frac{1}{\alpha_k} p_k^T (r_k - r_{k+1}) = \frac{1}{\alpha_k} p_k^T r_k \\ &= \frac{1}{\alpha_k} r_k^T (r_k + \beta_{k-1} p_{k-1}) = \frac{1}{\alpha_k} r_k^T r_k. \end{aligned} \tag{5.2.5}$$

由此可得

$$\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}, \ \ \beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}. \tag{5.2.6}$$

综合上面的讨论, 可得下面的算法:

算法5.2.1(解对称正定方程组: 共轭梯度法)

$x_0 =$ 初值

$r_0 = b - Ax_0;\ k = 0$

while $r_k \neq 0$

  $k = k + 1$

  if $k = 1$

    $p_0 = r_0$

  else

    $\beta_{k-2} = r_{k-1}^T r_{k-1} / r_{k-2}^T r_{k-2}$

    $p_{k-1} = r_{k-1} + \beta_{k-2} p_{k-2}$

  end

  $\alpha_{k-1} = r_{k-1}^T r_{k-1} / p_{k-1}^T A p_{k-1}$

  $x_k = x_{k-1} + \alpha_{k-1} p_{k-1}$

  $r_k = r_{k-1} - \alpha_{k-1} A p_{k-1}$

end

$x = x_k$

注意, 该算法每迭代一次仅需使用系数矩阵A做一次矩阵–向量运算.

共轭梯度法的基本性质

定理 (5.2.1)

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) $p_i^T r_j = 0,\ 0 \le i < j \le k;$

(2) $r_i^T r_j = 0,\ i \ne j,\ 0 \le i, j \le k;$

(3) $p_i^T A p_j = 0,\ i \ne j,\ 0 \le i, j \le k;$

(4) $\mathrm{span}\{r_0, \cdots, r_k\} = \mathrm{span}\{p_0, \cdots, p_k\} = \mathcal{K}(A, r_0, k+1)$, 其中

$$\mathcal{K}(A, r_0, k+1) = \mathrm{span}\{r_0, Ar_0, \cdots, A^k r_0\}, \qquad (5.2.6b)$$

通常称之为**Krylov**子空间.

证明 用数学归纳法. 当$k=1$时, 因为

$$p_0 = r_0,\ r_1 = r_0 - \alpha_0 A p_0,\ p_1 = r_1 + \beta_0 p_0,$$
$$r_1^T r_0 = r_0^T (r_0 - \alpha_0 A r_0) = r_0^T r_0 - \alpha_0 r_0^T A r_0 = 0,$$
$$p_1^T A p_0 = (r_1 + \beta_0 r_0)^T A r_0 = r_1^T A r_0 - \frac{r_1^T A r_0}{r_0^T A r_0} r_0^T A r_0 = 0,$$

所以定理的结论成立. 现在假设定理得结论对k成立, 我们来证明其对$k+1$也成立.

(1) 利用等式$r_{k+1} = r_k - \alpha_k A p_k$及归纳法假设, 有

$$p_i^T r_{k+1} = p_i^T r_k - \alpha_k p_i^T A p_k = 0,\ 0 \le i \le k-1.$$

又由于

$$p_k^T r_{k+1} = p_k^T r_k - \frac{p_k^T r_k}{p_k^T A p_k} p_k^T A p_k = 0,$$

故定理的结论(1)对$k+1$亦成立.

(2)利用归纳法假设有

$$\mathrm{span}\{r_0, \cdots, r_k\} = \mathrm{span}\{p_0, \cdots, p_k\}$$

而由(1)所证知, r_{k+1}与上述子空间正交, 从而定理的结论(2)对$k+1$也成立.

(3)利用等式

$$p_{k+1} = r_{k+1} + \beta_k p_k \text{ 和} r_{i+1} = r_i - \alpha_i A p_i,$$

并利用归纳法假设和(2)所证的结论, 就有

$$p_i^T A p_{k+1} = \frac{1}{\alpha_i} r_{k+1}^T (r_i - r_{i+1}) + \beta_k p_i^T A p_k = 0$$

对$i = 0, 1, \cdots, k-1$成立, 而由β_k的定义得

$$\begin{aligned} p_{k+1}^T A p_k &= (r_{k+1} + \beta_k p_k)^T A p_k \\ &= r_{k+1}^T A p_k - \frac{r_{k+1}^T A p_k}{p_k^T A p_k} p_k^T A p_k = 0. \end{aligned}$$

这样, 定理的结论(3)对$k+1$也成立.

(4)由归纳法假设知

$$r_k, p_k \in \mathcal{K}(A, r_0, k+1) = \mathrm{span}\{r_0, Ar_0, \cdots, A^k r_0\},$$

于是

$$\begin{aligned} r_{k+1} &= r_k - \alpha_k A p_k \in \mathcal{K}(A, r_0, k+2) = \mathrm{span}\{r_0, Ar_0, \cdots, A^{k+1} r_0\}, \\ p_{k+1} &= r_{k+1} + \beta_k p_k \in \mathcal{K}(A, r_0, k+2) = \mathrm{span}\{r_0, Ar_0, \cdots, A^{k+1} r_0\}. \end{aligned}$$

再注意到(2)和(3)所证的结论表明，向量组$r_0, \cdots, r_{k+1}$和$p_0, \cdots, p_{k+1}$都是线性无关的，因此定理的结论(4)对$k+1$同样成立.

综上所述，由归纳法原理知定理得证.

定理5.2.1表明，向量组$r_0, \cdots, r_k$和$p_0, \cdots, p_k$分别是Krylov子空间$\mathcal{K}(A, r_0, k+1)$的正交基和共轭正交基. 由此可知，利用共轭梯度法最多n步便可得到方程组的解x_*. 因此，理论上来讲，共轭梯度法是直接法.

定理 (5.2.2)

用共轭梯度法计算得到的近似解x_k满足

$$\varphi(x_k) = \min\{\varphi(x) : \ x \in x_0 + \mathcal{K}(A, r_0, k)\} \tag{5.2.7}$$

或

$$\|x_k - x_*\|_A = \min\{\|x - x_*\|_A : \ x \in x_0 + \mathcal{K}(A, r_0, k)\}, \tag{5.2.8}$$

其中$\|x\|_A = \sqrt{x^T A x}$, x_*是方程组$Ax = b$的解, $\mathcal{K}(A, r_0, k)$是(5.2.6b)式所定义的*Krylov*子空间.

证明 利用(5.1.4)式立即知道(5.2.7)式和(5.2.8)式是等价的, 因此我们下面只证明(5.2.8)式成立.

假设共轭梯度法计算到ℓ步出现$r_\ell = 0$, 那么有

$$
\begin{aligned}
x_* = x_\ell &= x_{\ell-1} + \alpha_{\ell-1} p_{\ell-1} \\
&= x_{\ell-2} + \alpha_{\ell-2} p_{\ell-2} + \alpha_{\ell-1} p_{\ell-1} \\
&= \cdots \\
&= x_0 + \alpha_0 p_0 + \alpha_1 p_1 + \cdots, \alpha_{\ell-1} p_{\ell-1}.
\end{aligned}
$$

此外, 对计算过程中的任一步$k < \ell$, 有

$$
x_k = x_0 + \alpha_0 p_0 + \alpha_1 p_1 + \cdots + \alpha_{k-1} p_{k-1} \in x_0 + \mathcal{K}(A, r_0, k).
$$

设x是属于$x_0 + \mathcal{K}(A, r_0, k)$的任一向量, 则由定理5.2.1 的(4)知, x可以表示为

$$
x = x_0 + \gamma_0 p_0 + \gamma_1 p_1 + \cdots + \gamma_{k-1} p_{k-1}.
$$

于是

$$
\begin{aligned}
x_* - x =& (\alpha_0 - \gamma_0) p_0 + \cdots + (\alpha_{k-1} - \gamma_{k-1}) p_{k-1} \\
& + \alpha_k p_k + \cdots + \alpha_{\ell-1} p_{\ell-1},
\end{aligned}
$$

而$x_* - x_k = \alpha_k p_k + \cdots + \alpha_{\ell-1} p_{\ell-1}$, 再利用定理5.2.1的(3)就可以推出

$$
\begin{aligned}
\|x_* - x\|_A^2 &= \|(\alpha_0 - \gamma_0) p_0 + \cdots + (\alpha_{k-1} - \gamma_{k-1}) p_{k-1}\|_A^2 \\
&\quad + \|\alpha_k p_k + \cdots + \alpha_{\ell-1} p_{\ell-1}\|_A^2 \\
&\geq \|\alpha_k p_k + \cdots + \alpha_{\ell-1} p_{\ell-1}\|_A^2 = \|x_* - x_k\|_A^2,
\end{aligned}
$$

从而定理得证.

目录

实用共轭梯度法

算法(解对称正定方程组: **实用共轭梯度法**)

x=初值

$r = b - Ax$; $\rho = r^T r$; k=0

While $(\sqrt{\rho} > \epsilon \|b\|_2)$ **and** $(k < k_{max})$

  $k = k + 1$

  if $k = 1$

    $p = r$

  else

    $\beta = \rho / \tilde{\rho}$; $p = r + \beta p$

  end

  $w = Ap$; $\alpha = \rho / p^T w$; $x = x + \alpha p$

  $r = r - \alpha w$; $\tilde{\rho} = \rho$; $\rho = r^T r$

end

实用共轭梯度法的收敛性分析

定理 (5.3.1)

如果$A = I + B$，而且$rank(B) = k$，则共轭梯度法至多迭代$k+1$步即得到方程的精确解.

证明： 注意到$\mathrm{rank}(B) = k$蕴涵着子空间

$$\mathrm{span}\{r_0, Ar_0, \cdots, A^k r_0\} = \mathrm{span}\{r_0, Br_0, \cdots, B^k r_0\}$$

的维数不会超过$k+1$. 证毕

定理 (5.3.2)

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

证明： 对任意$x \in x_0 + \mathcal{K}(A, r_0, k)$, 有

$$\begin{aligned} x_* - x &= x_* - x_0 + a_{k1}r_0 + a_{k2}Ar_0 + \cdots + a_{kk}A^{k-1}r_0 \\ &= A^{-1}(r_0 + a_{k1}Ar_0 + a_{k2}A^2r_0 + \cdots + a_{kk}A^k r_0) \\ &= A^{-1}P_k(A)r_0, \end{aligned}$$

其中$P_k(\lambda) = 1 + \sum\limits_{j=1}^{k} a_{kj}\lambda^j$.

令$\mathcal{P}_k$为所有满足$P_k(0)=1$ 且次数不超过k的实系数多项式的全体. 这样

$$\begin{aligned}\|x_*-x_k\|_A &= \min\{\|x-x_*\|_A : x\in x_0+\mathcal{K}(A,r_0,k)\}\\ &= \min_{P_k\in\mathcal{P}_k}\|A^{-1}P_k(A)r_0\|_A = \min_{P_k\in\mathcal{P}_k}\|P_k(A)A^{-1}r_0\|_A\\ &\le \min_{P_k\in\mathcal{P}_k}\max_{1\le i\le n}|P_k(\lambda_i)|\|A^{-1}r_0\|_A\\ &\le \min_{P_k\in\mathcal{P}_k}\max_{a\le\lambda\le b}|P_k(\lambda)|\|x_0-x_*\|_A,\end{aligned}$$

其中$0<a=\lambda_1\le\cdots\le\lambda_n=b$是$A$的特征值. 由著名的Chebyshev多项式逼近定理知，最优化问题

$$\min_{P_k\in\mathcal{P}_k}\max_{a\le\lambda\le b}|P_k(\lambda)|$$

有唯一的解

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

其中$T_k(z)$是k次Chebyshev多项式. 由Chebyshev多项式的性质知

$$\max_{a\le\lambda\le b}|\tilde{P}_k(\lambda)| = \frac{1}{T_k(\frac{b+a}{b-a})} \le 2\left(\frac{\sqrt{\kappa_2}-1}{\sqrt{\kappa_2}+1}\right)^k.$$

于是，我们有

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

因此，定理得证.

$$T_k\left(\frac{b+a}{b-a}\right) = 2^{k-1}\frac{1}{2^k}\left\{(x+\sqrt{x^2-1})^k + (x-\sqrt{x^2-1})^k\right\}$$

$$\text{（其中}x = \frac{b+a}{b-a}\text{）}$$

$$= \frac{1}{2}\left\{\left(\frac{(b+a)+\sqrt{(b+a)^2-(b-a)^2}}{b-a}\right)^k + \left(\frac{(b+a)-\sqrt{(b+a)^2-(b-a)^2}}{b-a}\right)^k\right\}$$

$$= \frac{1}{2}\left\{\left(\frac{(b+a)+2\sqrt{ab}}{b-a}\right)^k + \left(\frac{(b+a)-2\sqrt{ab}}{b-a}\right)^k\right\}$$

因为，

$$b+a+2\sqrt{ab}=(\sqrt{b}+\sqrt{a})^2, b+a-2\sqrt{ab}=(\sqrt{b}-\sqrt{a})^2,$$

所以

$$T_k(\frac{b+a}{b-a})=\frac{1+\sigma^{2k}}{2\sigma^k}$$

其中$\sigma=(\sqrt{b}-\sqrt{a})/(\sqrt{b}+\sqrt{a})$. 注意到

$$\frac{\sigma^k}{1+\sigma^{2k}}\leq\left(\frac{\sqrt{b}-\sqrt{a}}{\sqrt{b}+\sqrt{a}}\right)^k$$

目录

先将方程组转化为

$$\tilde{A}\tilde{x} = \tilde{b},$$

其中$\tilde{A} = C^{-1}AC^{-1}$, $\tilde{x} = Cx$, $\tilde{b} = C^{-1}b$, 这里要求C是对称正定的， 目的是通过C的选取, 使$\tilde{A}$具有良好性质.

$$\begin{aligned} \alpha_k &= \frac{\tilde{r}_k^T\tilde{r}_k}{\tilde{p}_k^T\tilde{A}\tilde{p}_k}, & \tilde{x}_{k+1} &= \tilde{x}_k + \alpha_k\tilde{p}_k \\ & & \tilde{r}_{k+1} &= \tilde{r}_k - \alpha_k\tilde{A}\tilde{p}_k, \\ \beta_k &= \frac{\tilde{r}_{k+1}^T\tilde{r}_{k+1}}{\tilde{r}_k^T\tilde{r}_k}, & \tilde{p}_{k+1} &= \tilde{r}_{k+1} + \beta_k\tilde{p}_k. \end{aligned} \tag{5.4.2}$$

$$\tilde{x}_k = Cx_k, \tilde{r}_k = C^{-1}r_k, \tilde{p}_k = Cp_k.$$

记$M = C^2$

$$
\begin{aligned}
w_k &= Ap_k & \alpha_k &= \rho_k/(p_k^T w_k) \\
x_{k+1} &= x_k + \alpha_k p_k & r_{k+1} &= r_k - \alpha_k w_k \\
z_{k+1} &= M^{-1} r_{k+1} & \rho_{k+1} &= r_{k+1}^T z_{k+1} \\
\beta_k &= \rho_{k+1}/\rho_k & p_{k+1} &= z_{k+1} + \beta_k p_k.
\end{aligned}
$$

其中x_0是任意给定的初始向量， $r_0 = b - Ax_0$，$z_0 = M^{-1} r_0$，$\rho_0 = r_0^T z_0$，$p_0 = z_0$.

算法

(解对称正定方程组: **预优共轭梯度法**)

x=初值

$r = b - Ax$; k=0

While $(\sqrt{r^T r} > \epsilon \|b\|_2)$ and $(k < k_{max})$

  求解$Mz = r$得z

  $k = k + 1$

  if $k = 1$

    $p = z$; $\rho = r^T z$

  else

    $\tilde{\rho} = \rho$; $\rho = r^T z$

    $\beta = \rho/\tilde{\rho}$; $p = z + \beta p$

  end

  $w = Ap$; $\alpha = \rho/p^T w$;

  $x = x + \alpha p$; $r = r - \alpha w$

end

这一算法也称为预条件共轭梯度法，简称PCG, 其中的矩阵M称做预优矩阵. 利用共轭梯度法的性质容易导出预优共轭梯度法如下性质：

1. 残量r_k是相互M^{-1}正交的，即$r_i^T M^{-1} r_j = 0$, $i \neq j$;
2. 方向向量p_k是相互A正交的，即$p_i^T A p_j = 0$, $i \neq j$;
3. 近似解向量x_k满足

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

其中$\kappa = \lambda_n/\lambda_1$, λ_n和λ_1是$M^{-1}A$的最大和最小特征值.

目录

广义极小剩余法

给定初始值x_0, 残量为$r_0 = b - Ax_0$. 定义Krylov子空间

$$\kappa(A, r_0, k) = \mathsf{span}\{r_0, Ar_0, \cdots, A^{k-1}r_0\}.$$

在它上求解一个向量$\tilde{x}_k$使得剩余向量$r_k = b - A(\tilde{x}_k + x_0)$ 满足

$$\|r_k\|_2 = \|b - A(\tilde{x}_k + x_0)\|_2 = \min = \|r_0 - A\tilde{x}_k\|_2$$

设

$$\kappa(A, r_0, m) = \mathsf{span}\{r_0, Ar_0, \cdots, A^{m-1}r_0\} = \mathsf{span}\{v_1, v_2, \cdots, v_m\}, \quad m = 1, 2, \cdots, k+1,$$

则有

$$\begin{aligned} Av_m \in \kappa(A, r_0, m+1) &= \mathsf{span}\{r_0, Ar_0, \cdots, A^{m-1}r_0, A^m r_0\} \\ &= \mathsf{span}\{v_1, v_2, \cdots, v_m, v_{m+1}\}. \end{aligned}$$

$$Av_m = h_{1m}v_1 + h_{2m}v_2 + \cdots + h_{mm}v_m + h_{m+1,m}v_{m+1}.$$

令

$$V_m = [v_1, v_2, \cdots, v_m]$$

$$H_m = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1m} \\ h_{21} & h_{22} & \cdots & h_{2m} \\ & h_{32} & \ddots & \vdots \\ & & \ddots & h_{mm} \\ & & & h_{m+1,m} \end{bmatrix}$$

则有

$$AV_k = V_{k+1}H_k.$$

注意到

$$r_0 = \alpha v_1.$$

若记$d_k = \alpha e_1$, 其中e_1是k阶单位矩阵的第一列，则上述最小值问题等价于求k 维向量y_k使得

$$\|V_{k+1}(d_k - H_k y_k)\|_2 = \min$$

若v_1, v_2, $\cdots$, v_{k+1}是空间$\kappa(A, r_0, k+1)$的标准正交基，则有$V_{k+1}^T V_{k+1} = I$，于是上式即为求一个k维向量y_k满足

$$\|d_k - H_k y_k\|_2 = \min .$$

这个系数矩阵为上Hessenberg矩阵，其相应的最小二乘问题很容易求得.

正交基v_1, v_2, $\cdots$, v_{k+1}和系数阵H_k的元素h_{ij}可以用Gram–Schmidt正交化逐步求得：

$$v_1 = \frac{r_0}{\|r_0\|_2}, h_{ij} = v_i^T A v_j, i = 1, 2, \cdots, j,$$

$$\tilde{v}_{j+1} = A v_j - \sum_{i=1}^{j} h_{ij} v_i,$$

$$h_{j+1,j} = \|\tilde{v}_{j+1}\|, v_{j+1} = \frac{\tilde{v}_{j+1}}{h_{j+1,j}}.$$

这个迭代称作Arnoldi迭代.

将

$$x_k = x_0 + V_k y_k$$

作为线性方程组的近似解. 这一方法称为广义极小剩余发(GMRES).

当k很大时，由于整个计算过程需要保存所有的向量$v_1, v_2, \cdots, v_{k+1}$, 因此需要占用大量的存储单元，以至于使得所处理问题的规模受到很大的限制. 为了克服(GMRES)的这个缺点，通常使用时，先选择一个适当的正整数m, 计算到$k = m$时，就去求x_m; 然后在以x_m为初始向量重新开始；这样循环直到求出满足精度的近似解. 这种循环的GMRES，习惯上称为GMRES(m).

GMRES(m)算法

1. 输入A, b, x_0, m, 以及精度要求ϵ;
2. $r_0 = b - Ax_0$, $v_1 = \frac{r_0}{\|r_0\|_2}$, $k := 1$;
3. $h_{ik} = v_i^T A v_k$, $i = 1, 2, \cdots, k$,
   $\tilde{v}_{k+1} = Av_k - (h_{1k}v_1 + \cdots + h_{kk}v_k)$
   $h_{k+1,k} = \|\tilde{v}_{k+1}\|_2$;
4. 若$h_{k+1,k} < \epsilon$ 或者$k = m$, 则转步(5);否则$v_{k+1} = \frac{\tilde{v}_{k+1}}{h_{k+1,k}}$,
   $k := k + 1$,转步(3);
5. 求解最小二乘问题
   $$\|H_k y_k - d_k\|_2 = \min$$
   并计算$x_k = x_0 + V_k y_k$.
6. 若$\|Ax_k - b\| < \epsilon$, 则输出x_k, 结束;否则$x_0 := x_m$, 转步(2).

目前， 这一方法是求解大型稀疏非对称线性方程组最有效的方法之一. 若$\frac{1}{2}(A+A^T)$是正定的，对任意m, GMRES(m)产生的向量序列是收敛的.