第三章 Krylov 子空间方法

设有求解线性方程组

$$AX = b, \quad \text{其中 } A = (a_{ij}) \in \mathbb{R}^{n\times n} \text{ 可逆}, \tag{1}$$

的某迭代法，其初始迭代向量为 X_0，如记残差 $r_0 = b - AX_0$，则由向量 $r_0, Ar_0, \cdots, A^k r_0$ 所张成的线性空间

$$\mathcal{K}(A, r_0, k+1) = \mathrm{Span}\{r_0, Ar_0, \cdots, A^k r_0\}$$

称为**Krylov 子空间**。基于该空间，人们已经构造了多种求解线性方程组 (1)的迭代方法，其中有共轭梯度法、预优共轭梯度法、残量极小化方法、残量正交化方法等。这些方法被统称为**Krylov 子空间方法**，其对求解大型稀疏方程组尤为有效。本节将主要介绍最速下降法、共轭梯度法及预优共轭梯度法。

§3.1 最速下降法

为引入共轭梯度法，我们首先介绍最速下降法。考虑线性方程组 (1)，设其系数阵为对称正定阵。令引入二次泛函

$$\Phi(X) = X^T AX - 2b^T X, \quad X = (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n. \tag{2}$$

该二次泛函使得线性方程组的求解问题可转化为最优化问题。

定理 3.1 设 A 为对称正定阵，则 X^* 为方程组 (1) 的解当且仅当 X^* 为二次泛函 $\Phi(X)$ 的极小点。

证明 直接计算可得

$$\frac{\partial \Phi}{\partial x_i} = 2\Big(\sum_{j=1}^{n} a_{ij}x_j - b_i\Big), \quad i = 1, 2, \cdots, n.$$

记 $r = b - AX$，则 $\Phi(X)$ 的梯度

$$\nabla \Phi(X) = 2(AX - b) = -2r.$$

若 $\Phi(X)$ 在某点 X^* 达到极小，则必有 $\nabla \Phi(X^*) = 0$，即 $AX^* = b$，从而 X^* 是方程组 (1) 的解。

反之，若 X^* 是方程组 (1) 的解，则 $\forall Y \in \mathbb{R}^n$ 有

$$\Phi(X^* + Y) = \Phi(X^*) + Y^TAY \geq \Phi(X^*).$$

故 X^* 为二次泛函 $\Phi(X)$ 的极小点。 ■

定理 3.1 表明，线性方程组的求解问题可转化为求二次泛函 $\Phi(X)$ 的极小点问题。该极小点问题可采用**“盲人下山法”**解决，其具体做法如下：

Step 1. 任给初始点 X_0，确定一个下山方向 P_0，沿着过 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

$$\Phi(X_1) \leq \Phi(X_0 + \alpha P_0), \quad \forall \alpha \in \mathbb{R};$$

Step 2. 以 X_1 作为新的出发点，确定一个下山方向 P_1，沿着直线：$X = X_1 + \alpha P_1$ 找一个点 $X_2 = X_1 + \alpha_1 P_1$，使得

$$\Phi(X_2) \leq \Phi(X_1 + \alpha P_1), \quad \forall \alpha \in \mathbb{R};$$

Step 3. 重复上述步骤，则可获得步长序列 $\{\alpha_k\}$ 及搜索方向向量序列 $\{P_k\}$，其满足

$$\begin{cases} X_{k+1} = X_k + \alpha_k P_k, \quad X_0\text{给定}, \\ \Phi(X_{k+1}) \leq \Phi(X_k + \alpha P_k), \quad \forall \alpha \in \mathbb{R}. \end{cases} \tag{3}$$

上述算法中，步长序列 $\{\alpha_k\}$ 及搜索方向向量序列 $\{P_k\}$ 可确定如下：

为确定步长序列 $\{\alpha_k\}$，我们令 $f(\alpha) = \Phi(X_k + \alpha P_k)$，即

$$f(\alpha) = \alpha^2 P_k^T A P_k - 2\alpha r_k^T P_k + \Phi(X_k), \quad \text{其中 } r_k = b - AX_k.$$

由微分学理论可知

$$f'(\alpha) = 2\alpha P_k^T A P_k - 2r_k^T P_k = 0,$$

从而

$$\alpha_k = \frac{r_k^T P_k}{P_k^T A P_k}. \tag{4}$$

且当 $r_k^T P_k \neq 0$ 时，有

$$\Phi(X_{k+1}) - \Phi(X_k) = \alpha_k^2 P_k^T A P_k - 2\alpha_k r_k^T P_k = -\frac{(r_k^T P_k)^2}{P_k^T A P_k} < 0,$$

即 $\Phi(X_{k+1}) < \Phi(X_k)$。又由于一个函数的梯度方向是其函数值增长最快的方向，因此我们可取

$$P_k = -\frac{1}{2}\nabla\Phi(X_k) = r_k. \tag{5}$$

至此，我们获得了一个系数阵为对称正定阵的线性方程组的求解方法，称之为**最速下降法**。该算法的计算程序如下：

算法 3.1 最速下降法

```
function X=sdescent(A,b)
 n=length(b); X0=zeros(n,1); r0=b-A*X0; k=0;
 while norm(r0)>=10^(-12)
    k=k+1; alpha0=r0'*r0/(r0'*A*r0);
    X1=X0+alpha0*r0; r0=b-A*X1; X0=X1;
 end
X=X0
```

例 3.1 试应用最速下降法求解线性方程组

$$\begin{bmatrix} 4 & -2 & 4 & 2 \\ -2 & 10 & -2 & -7 \\ 4 & -2 & 8 & 4 \\ 2 & -7 & 4 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \\ 16 \\ 6 \end{bmatrix}.$$

解 应用算法 3.1，迭代 520 步可得其方程组的解

$$X = (x_1, x_2, x_3, x_4)^T = (1.0000,\ 2.0000,\ 1.0000,\ 2.0000)^T. \quad \blacksquare$$

为获得最速下降法的收敛性估计，我们首先引入如下引理。

引理 3.1 设正定阵 A 的特征值为 $0<\lambda_1\le\lambda_2\le\cdots\le\lambda_n$，$P(t)$是实系数多项式，则

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

其中 $\|X\|_A=\sqrt{X^TAX}$.

证明 设$\{y_i\}_{i=1}^n$是A 的一组线性无关的标准正交化特征向量，其中y_i从属于特征值λ_i。从而$\forall X\in\mathbb{R}^n$有$X=\sum\limits_{i=1}^n\beta_iy_i$，且

$$\begin{aligned}\|P(A)X\|_A^2&=[P(A)X]^TA[P(A)X]=[\sum_{i=1}^n\beta_iP(A)y_i]^TA[\sum_{j=1}^n\beta_jP(A)y_j]\\&=[\sum_{i=1}^n\beta_iP(\lambda_i)y_i]^TA[\sum_{j=1}^n\beta_jP(\lambda_j)y_j]=\sum_{i=1}^n\sum_{j=1}^n\lambda_j\beta_i\beta_jP(\lambda_i)P(\lambda_j)y_i^Ty_j\\&=\sum_{i=1}^n\lambda_i\beta_i^2P^2(\lambda_i)\le\max_{1\le i\le n}|P(\lambda_i)|^2\sum_{j=1}^n\lambda_j\beta_j^2\\&=\max_{1\le i\le n}|P(\lambda_i)|^2(X^TAX)=\max_{1\le i\le n}|P(\lambda_i)|^2\|X\|_A^2.\end{aligned}$$

故结论成立。 ■

定理 3.2 设方程组系数阵 A 的特征值为 $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$，则由最速下降法产生的迭代序列 $\{X_k\}$ 收敛于其方程组的精确解 X^*，且满足

$$\|X_k - X^*\|_A \le \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^k \|X_0 - X^*\|_A.$$

证明 由于$\Phi(X_k) \le \Phi(X_{k-1} + \alpha r_{k-1})\ (\forall \alpha \in \mathbb{R})$ 及

$$\Phi(X) + (X^*)^T A X^* = (X - X^*)^T A (X - X^*). \tag{6}$$

则

$$\begin{aligned} & (X_k - X^*)^T A (X_k - X^*) \\ \le\ & (X_{k-1} + \alpha r_{k-1} - X^*)^T A (X_{k-1} + \alpha r_{k-1} - X^*) \\ =\ & [(I - \alpha A)(X_{k-1} - X^*)]^T A [(I - \alpha A)(X_{k-1} - X^*)], \quad \forall \alpha \in \mathbb{R} \end{aligned}$$

记$P_\alpha(t) = 1 - \alpha t$，则由上式及引理 3.1得

$$\|X_k - X^*\|_A \le \|P_\alpha(A)(X_{k-1} - X^*)\|_A \le \max_{1 \le i \le n} |P_\alpha(\lambda_i)| \|X_{k-1} - X^*\|_A.$$

在上式中取$\alpha = 2/(\lambda_1 + \lambda_n)$，并由此递推即得结论。 ■

由此可见，最速下降法的收敛速度主要取决于比值 $(\lambda_n-\lambda_1)/(\lambda_n+\lambda_1)$ 及初始向量 X_0 的选取。当 $\lambda_n \gg \lambda_1$或初始向量 X_0 选取不恰当时，最速下降法的收敛速度将非常缓慢。因此，该方法不适宜应用于病态方程.

§3.2 基本共轭梯度法

为改进最速下降法，本节引入共轭梯度法。我们注意到最速下降法所取的下山方向 - 负梯度方向就局部而言是最佳的，但从全局而言则不一定是最佳的。下述计算过程给出全局最佳下山方向：

Step 1. 给定初始向量 X_0，第一步仍取负梯度方向为下山方向，即有

$$P_0=r_0,\quad \alpha_0=\frac{r_0^Tr_0}{P_0^TAP_0},\quad X_1=X_0+\alpha_0P_0,\quad r_1=b-AX_1;$$

Step 2. 第 $k+1$ 步（$k \geq 1$）不再取 $r_k := b - AX_k$ 作为下山方向，而是先作过点 X_k 且由向量 r_k, P_{k-1} 张成的超平面

$$\Pi: \quad X = X_k + \xi r_k + \eta P_{k-1}, \ \ \xi, \eta \in \mathbb{R},$$

在该平面上找出 $\Phi(X)$ 递减最快的方向 P_k。记 $\Psi(\xi,\eta) = \Phi(X_k + \xi r_k + \eta P_{k-1})$，并令

$$\begin{cases} \frac{\partial \Psi}{\partial \xi} = 2(\xi r_k^T A r_k + \eta r_k^T A P_{k-1} - r_k^T r_k) = 0, \\ \frac{\partial \Psi}{\partial \eta} = 2(\xi r_k^T A P_{k-1} + \eta P_{k-1}^T A P_{k-1}) = 0. \end{cases} \tag{7}$$

解该方程组可得 $\Phi(X)$ 在平面 Π 上的唯一极小点 $\hat{X} = X_k + \hat{\xi} r_k + \hat{\eta} P_{k-1}$，其中 $\hat{\xi}, \hat{\eta}$ 满足方程 (7)。由 $r_k \neq 0$ 可知 $\hat{\xi} \neq 0$，因此可取

$$P_k = \frac{1}{\hat{\xi}}(\hat{X} - X_k) = r_k + \frac{\hat{\eta}}{\hat{\xi}} P_{k-1}$$

作为新的下山方向。显然，这是平面 Π 上的最佳下山方向。

由 (7) 中第二式有

$$\beta_{k-1} := \frac{\hat{\eta}}{\hat{\xi}} = -\frac{r_k^T A P_{k-1}}{P_{k-1}^T A P_{k-1}}.$$

P_k 确定后，步长 α_k 仍可由 (4) 给定。至此，我们得到一个改进的最速下降法，称之为**共轭梯度法**，其计算公式为

$$\begin{cases} \alpha_k = \frac{r_k^T P_k}{P_k^T A P_k}, \quad X_{k+1} = X_k + \alpha_k P_k, \\ r_{k+1} = b - AX_{k+1}, \ \ \beta_k = -\frac{r_{k+1}^T A P_k}{P_k^T A P_k}, \ \ P_{k+1} = r_{k+1} + \beta_k P_k. \end{cases} \tag{8}$$

为节省计算量，上述公式还可进一步简化，其需引入下列结论。

定理 3.3 公式 (8) 中的向量组 $\{r_i\}, \{P_i\}$ 满足如下性质：

(1) $P_i^T r_j = 0, \quad 0 \le i < j \le k;$

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

(3) $P_i^T A P_j = 0, \quad i \neq j,\ 0 \le i, j \le k$ (称$P_0, P_1, \ldots,\ P_k$ **是相互共轭正交的**);

(4) $\mathcal{K}(A, r_0, k+1) = \mathrm{Span}\{r_0, r_1, \cdots, r_k\} = \mathrm{Span}\{P_0, P_1, \cdots, P_k\}.$

证明 (用数学归纳法) 当$k=1$时，由

$$P_0=r_0,\ r_1=r_0-\alpha_0 AP_0,\ P_1=r_1+\beta_0 P_0$$

有 $$P_0^T r_1=r_0^T r_1=r_0^T(r_0-\alpha_0 Ar_0)=r_0^T r_0-\alpha_0 r_0^T Ar_0=0,$$

$$P_0^T AP_1=P_1^T AP_0=(r_1+\beta_0 r_0)^T Ar_0=r_1^T Ar_0-\frac{r_1^T Ar_0}{r_0^T Ar_0}(r_0^T Ar_0)=0,$$

$$\mathrm{Span}\{r_0,Ar_0\}=\mathrm{Span}\{r_0,r_1\}=\mathrm{Span}\{P_0,P_1\}.$$

因此$k=1$时结论成立。

今设结论当$k\le m$时成立，下证其在$k=m+1$时也成立。
(1) 当$0\le i\le m-1$时，由归纳假设及等式

$$r_{m+1}=b-AX_{m+1}=b-A(X_m+\alpha_m P_m)=r_m-\alpha_m AP_m \qquad (9)$$

得 $$P_i^T r_{m+1}=P_i^T(r_m-\alpha_m AP_m)=P_i^T r_m-\alpha_m P_i^T AP_m=0.$$

且据(8)的第一式及(9)有

$$P_m^T r_{m+1}=P_m^T r_m-\alpha_m P_m^T AP_m=P_m^T r_m-\frac{r_m^T P_m}{P_m^T AP_m}(P_m^T AP_m)=0.$$

因此结论(1)在$k=m+1$时也成立。

(2) 由归纳假设

$$\text{Span}\{r_0, r_1, \ldots, r_m\} = \text{Span}\{P_0, P_1, \ldots, P_m\}.$$

而由(1)可知r_{m+1}与上述空间正交，因此结论(2)在$k = m+1$时也成立。

(3) 当$0 \le i \le m-1$时，由(8)的最后一式,(9),归纳假设及结论(2)有

$$P_i^T A P_{m+1} = P_i^T A(r_{m+1} + \beta_m P_m) = \frac{1}{\alpha_i} r_{m+1}^T (r_i - r_{i+1}) + \beta_m P_i^T A P_m = 0.$$

又由(8)的最后一式及其第4式有

$$P_{m+1}^T A P_m = (r_{m+1} + \beta_m P_m)^T A P_m = r_{m+1}^T A P_m - \frac{r_{m+1}^T A P_m}{P_m^T A P_m}(P_m^T A P_m) = 0.$$

因此结论(3)在$k = m+1$时也成立。

(4) 由归纳假设

$$r_m, P_m \in \mathcal{K}(A, r_0, m+1) = \text{Span}\{r_0, Ar_0, \cdots, A^m r_0\}.$$

因此

$$r_{m+1} = r_m - \alpha_m AP_m \in \mathrm{Span}\{r_0, Ar_0, \cdots, A^{m+1}r_0\},$$
$$P_{m+1} = r_{m+1} + \beta_m AP_m \in \mathrm{Span}\{r_0, Ar_0, \cdots, A^{m+1}r_0\}.$$

而由结论(2), (3)可知$r_0, r_1, \cdots, r_{m+1}$和$P_0, P_1, \cdots, P_{m+1}$均是线性无关的，即$\{r_i\}_{i=0}^{m+1}$和$\{P_i\}_{i=0}^{m+1}$ 同为空间Span$\{r_0, Ar_0, \cdots, A^{m+1}r_0\}$的一组基，故结论(4)在$k = m+1$时也成立。 ■

该定理表明 $\{r_i\}_{i=0}^{k}, \{P_i\}_{i=0}^{k}$ 分别是 Krylov 子空间 $\mathcal{K}(A, r_0, k+1)$ 的正交基和共轭正交基。据此，当应用共轭梯度法求解 n 维线性方程组时，理论上至多 n 步即可获得方程组的精确解。故共轭梯度法既是一种迭代法也是一种直接方法。此外，由(9)可推得等式 $AP_k = (r_k - r_{k+1})/\alpha_k$。据该等式及定理 3.3，我们进一步有

$$r_{k+1}^T AP_k = r_{k+1}^T (r_k - r_{k+1})/\alpha_k = -r_{k+1}^T r_{k+1}/\alpha_k,$$

$$P_k^T AP_k = P_k^T (r_k - r_{k+1})/\alpha_k = P_k^T r_k/\alpha_k = r_k^T (r_k + \beta_{k-1} P_{k-1})/\alpha_k = r_k^T r_k/\alpha_k.$$

因此，公式 (8) 可置换为

$$\begin{cases} \alpha_k = \frac{r_k^T r_k}{P_k^T AP_k}, \quad X_{k+1} = X_k + \alpha_k P_k, \\ r_{k+1} = r_k - \alpha_k AP_k, \ \ \beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}, \ \ P_{k+1} = r_{k+1} + \beta_k P_k. \end{cases} \tag{10}$$

该算法的程序代码如下：

算法 3.2　共轭梯度法

```
function X=congrad(A,b)
 n=length(b); X0=zeros(n,1); r0=b-A*X0; P0=r0;
while norm(r0)>=10^(-12)
   alpha=r0'*r0/(P0'*A*P0); X1=X0+alpha*P0;
   r1=r0-alpha*A*P0; beta=r1'*r1/(r0'*r0);
   P1=r1+beta*P0; r0=r1; X0=X1; P0=P1;
end
X0
```

例 3.2 试应用共轭梯度法求解例 3.1 中的线性方程组

解 应用算法 3.2于例 3.1 中的线性方程组，迭代 4 步即可获得其方程组的解

$$X=(x_1,x_2,x_3,x_4)^T=(1.0000,\ 2.0000,\ 1.0000,\ 2.0000)^T.\quad\blacksquare$$

将例 3.2与例 3.1相比较可知：共轭梯度法的收敛速度要大大地快于最速下降法。

定理 3.4 设X^*为对称正定方程组$AX = b$的解，且记

$$\Phi(X) = X^TAX - 2b^TX, \quad \|X\|_A = \sqrt{X^TAX}, \quad X \in \mathbb{R}^n,$$

则共轭梯度法的第k次迭代值X_k满足

$$\Phi(X_k) = \min\{\Phi(X): \ X \in X_0 + \mathcal{K}(A, r_0, k)\}, \qquad (11)$$

$$\|X_k - X^*\|_A = \min\{\|X - X^*\|_A: \ X \in X_0 + \mathcal{K}(A, r_0, k)\}. \qquad (12)$$

证明 由恒等式(6) 可知，式(11)与(12)是等价的。因此，我们只需证明式(12)。若共轭梯度法迭代到第l步出现$r_l = 0$ $(l \le n)$，则有

$$X^* = X_l = X_{l-1} + \alpha_{l-1}P_{l-1} = X_0 + \sum_{i=0}^{l-1} \alpha_i P_i,$$

$$X_k = X_0 + \sum_{i=0}^{k-1} \alpha_i P_i \in X_0 + \mathcal{K}(A, r_0, k), \quad \forall k < l.$$

又由定理3.3有

$$X = X_0 + \sum_{i=0}^{k-1} \gamma_i P_i, \quad \forall X \in X_0 + \mathcal{K}(A, r_0, k).$$

因此

$$X^* - X = \sum_{i=0}^{k-1} (\alpha_i - \gamma_i) P_i + \sum_{i=k}^{l-1} \alpha_i P_i, \quad X^* - X_k = \sum_{i=k}^{l-1} \alpha_i P_i.$$

从而，进一步应用定理3.3得

$$\begin{aligned} \|X^* - X\|_A^2 &= \Big\| \sum_{i=0}^{k-1} (\alpha_i - \gamma_i) P_i \Big\|_A^2 + \Big\| \sum_{i=k}^{l-1} \alpha_i P_i \Big\|_A^2 \\ &\geq \Big\| \sum_{i=k}^{l-1} \alpha_i P_i \Big\|_A^2 = \|X^* - X_k\|_A^2. \end{aligned} \tag{13}$$

故定理获证。 ■

由该定理可进一步获得共轭梯度法具有如下误差估计。

定理 3.5 设方程组系数阵 A 的特征值为 $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$，则由共轭梯度法产生的迭代序列 $\{X_k\}$ 收敛于其方程组的精确解 X^*，且满足

$$\|X_k - X^*\|_A \le 2\left(\frac{\sqrt{\lambda_n}-\sqrt{\lambda_1}}{\sqrt{\lambda_n}+\sqrt{\lambda_1}}\right)^k \|X_0 - X^*\|_A, \tag{14}$$

其中 $\|X\|_A = \sqrt{X^TAX}$.

证明 $\forall X \in X_0 + \mathcal{K}(A, r_0, k)$ 有

$$\begin{aligned} X^* - X &= X^* - X_0 + \sum_{j=1}^{k} d_{kj}A^{j-1}r_0 \\ &= A^{-1}b - A^{-1}(b - r_0) + \sum_{j=1}^{k} d_{kj}A^{j-1}r_0 \\ &= A^{-1}(r_0 + \sum_{j=1}^{k} d_{kj}A^{j}r_0) \\ &= A^{-1}P_k(A)r_0, \quad \text{其中 } P_k(\lambda) = \sum_{j=0}^{k} d_{kj}\lambda^j,\ P_k(0) = 1. \end{aligned}$$

令 $Q_k = \{P_k : \ P_k\text{为次数不超过}k\text{的实系数多项式, 且}P_k(0) = 1\}$，则由定理 3.4及引理 3.1得

$$
\begin{aligned}
\|X^* - X_k\|_A &= \min\{\|X - X^*\|_A :\ X \in X_0 + \mathcal{K}(A, r_0, k)\} \\
&= \min_{P_k \in Q_k} \|A^{-1} P_k(A) r_0\|_A = \min_{P_k \in Q_k} \|P_k(A) A^{-1} r_0\|_A \\
&\le \min_{P_k \in Q_k} \max_{1 \le i \le n} |P_k(\lambda_i)| \|A^{-1} r_0\|_A \\
&\le \min_{P_k \in Q_k} \max_{a \le \lambda \le b} |P_k(\lambda)| \|X^* - X_0\|_A, \quad (\text{记 } a = \lambda_1,\ b = \lambda_n).
\end{aligned}
$$

又由Chebychev多项式逼近定理有

$$
\min_{P_k \in Q_k} \max_{a \le \lambda \le b} |P_k(\lambda)| = \max_{a \le \lambda \le b} |\tilde{P}_k(\lambda)|,
$$

这里，

$$
\tilde{P}_k(\lambda) = \frac{T_k(\frac{b+a-2\lambda}{b-a})}{T_k(\frac{b+a}{b-a})}, \quad T_k(x) = \frac{(x + \sqrt{x^2 - 1})^k + (x - \sqrt{x^2 - 1})^k}{2},
$$

其中$T_k(x)$称为Chebychev多项式，其是首项系数为2^{k-1}的k次多项式，且有

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

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

故不等式(14)成立。 ■

该定理表明共轭梯度法的收敛速度主要取决于比值 $\frac{\sqrt{\lambda_n}-\sqrt{\lambda_1}}{\sqrt{\lambda_n}+\sqrt{\lambda_1}}$ 及初始向量 X_0 的选取。此外，根据下列不等式

$$\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1} = \left(\frac{\sqrt{\lambda_n}-\sqrt{\lambda_1}}{\sqrt{\lambda_n}+\sqrt{\lambda_1}}\right)\left(1+\frac{2\sqrt{\lambda_n\lambda_1}}{\lambda_n+\lambda_1}\right) > \frac{\sqrt{\lambda_n}-\sqrt{\lambda_1}}{\sqrt{\lambda_n}+\sqrt{\lambda_1}},$$

我们从理论上进一步获知：共轭梯度法的收敛速度要比最速下降法快。

此外，共轭梯度法也可用于求解一般线性方程组$AX=b$。事实上，当系数阵A可逆时，该方程组等价于正定方程组$A^TAX=A^Tb$，应用共轭梯度法求解后者，我们即可获得原方程组的解X，该方法称为**正则化方法**。

§3.3 预优共轭梯度法

对于系数阵 A 为正定阵的 n 维线性方程组，当其最大特征值和最小特征值分别为 λ_n, λ_1 时，其条件数 $\mathrm{cond}(A)_\infty = \lambda_n/\lambda_1$。由此可知，当 $\lambda_n \gg \lambda_1$ 时，该方程组为病态的，否则其是良态的。鉴此及定理 3.5 中的误差估计式 (14) 可知，当方程组为良态时，共轭梯度法的收敛速度较快；当方程组为病态时，共轭梯度法的收敛速度缓慢。为提高共轭梯度法解病态方程组的计算效率，本节引入**预优共轭梯度法**。

预优共轭梯度法的基本思想是：首先通过引入一个 n 级对称正定阵 E 设法将病态方程组 $AX = b$ 转换为良态方程组

$$\hat{A}\hat{X} = \hat{b}, \tag{15}$$

其中

$$\hat{A} = E^{-1}AE^{-1}, \quad \hat{X} = EX, \quad \hat{b} = E^{-1}b;$$

然后，应用共轭梯度法于良态方程组 (15)，即产生下列计算方案

$$\begin{cases} \alpha_k = \dfrac{\hat{r}_k^T\hat{r}_k}{\hat{P}_k^T\hat{A}\hat{P}_k}, \quad \hat{X}_{k+1} = \hat{X}_k + \alpha_k\hat{P}_k, \\ \hat{r}_{k+1} = \hat{r}_k - \alpha_k\hat{A}\hat{P}_k, \quad \beta_k = \dfrac{\hat{r}_{k+1}^T\hat{r}_{k+1}}{\hat{r}_k^T\hat{r}_k}, \quad \hat{P}_{k+1} = \hat{r}_{k+1} + \beta_k\hat{P}_k. \end{cases} \tag{16}$$

为直接得到原方程组 $AX=b$ 的逼近解 X_k，令

$$\hat{X}_k = EX_k \ \ \hat{r}_k = E^{-1}r_k, \ \ \hat{P}_k = EP_k, \ \ M = E^2,$$

$$\omega_k = AP_k, \ \ \rho_k = r_k^T\zeta_k, \ \ \zeta_k = M^{-1}r_k,$$

且取初始迭代向量 X_0 及

$$r_0 = b - AX_0, \ \ \zeta_0 = M^{-1}r_0, \ \ \rho_0 = r_0^T\zeta_0,$$

则计算方案可等价地写成

$$\begin{cases} \omega_k = AP_k, \ \ \alpha_k = \frac{\rho_k}{P_k^T\omega_k}, \ \ X_{k+1} = X_k + \alpha_k P_k, \\ r_{k+1} = r_k - \alpha_k\omega_k, \ \ \zeta_{k+1} = M^{-1}r_{k+1}, \ \ \rho_{k+1} = r_{k+1}^T\zeta_{k+1}, \\ \beta_k = \frac{\rho_{k+1}}{\rho_k}, \ \ P_{k+1} = \zeta_{k+1} + \beta_k P_k. \end{cases} \tag{17}$$

该算法的程序代码如下：

算法 3.3 预优共轭梯度法

```
function X=precongrad(A,b,M)
 n=length(b); X0=zeros(n,1); r0=b-A*X0;
 zeta0=M\r0; rho0=r0'*zeta0;  P0=zeta0;
while norm(r0)>=10^(-12)
 omega=A*P0; alpha=rho0/(P0'*omega); X1=X0+alpha*P0;
 r1=r0-alpha*omega; zeta1=M\r1; rho1=r1'*zeta1;
 beta=rho1/rho0; P1=zeta1+beta*P0;
 r0=r1; X0=X1; P0=P1; rho0=rho1;
end
X0
```

算法中的矩阵 M 称为**预优矩阵**，其选择可基于下列要求：

- M 为稀疏的对称正定阵；
- $M^{-1}A$ 的特征值量级相差不大；
- 方程组 $M\zeta = r$ 易解。

对于预优矩阵 M的选择，人们已导出多种方法，但其中有些方法异常复杂，有的是针对非常特殊的方程组给出的。这里，我们仅介绍二种简单易行的方法：

[1] 若原方程组的系数阵为分块阵

$$A=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pp} \end{bmatrix},$$

其中 A_{ii} 均是易于求逆的方阵，则可取

$$M=\mathrm{diag}(A_{11},A_{22},\cdots,A_{pp}).$$

特别，当原方程组系数阵的对角元素非零且有较大差异时，则可取

$$M=\mathrm{diag}(a_{11},a_{22},\cdots,a_{nn}).$$

[2] 若原方程组的系数阵可作如下不完全 Cholesky 分解

$$A=LL^T+R,$$

其中 L 为单位下三角阵，且 $LL^T\approx A$，则可取 $M=LL^T$。

§3.4 其它Krylov子空间方法

由定理3.4可知，共轭梯度法实质上就是求$X_k \in X_0 + \mathcal{K}(A, r_0, k)$使得

$$\Phi(X_k) = \min\{\Phi(X) : \ X \in X_0 + \mathcal{K}(A, r_0, k)\}.$$

针对求解上述最优化问题，人们又导出各种类型的Krylov子空间方法，如：SYMMLQ方法, Arnoldi方法等(详见文献[SIAM J. Numer. Anal. 1975, 12: 617-629; Math. Comp. 1981, 37: 105-126])。

对于求解一般线性方程组$AX = b$的另一种重要Krylov子空间方法是**残量极小化方法**，其基本思想是：求$X_k \in X_0 + \mathcal{K}(A, r_0, k)$使得

$$\|b - AX_k\|_2 = \min\{\|b - AX\|_2 : \ \ X \in X_0 + \mathcal{K}(A, r_0, k)\}$$

从而获得原线性方程组的逼近解X_k。围绕求解上述最优化问题至今也已衍生出许多类型的Krylov子空间方法，如：GMRES方法、MINRES方法等(详见文献[SIAM J. Numer. Anal. 1975, 12: 617-629; Math. Comp. 1981, 37: 105-126])。