斜投影方法收敛速度的估计

颜世建

(南京师范大学数学与计算机科学学院,南京,210097)

[摘要] 对斜投影法的收敛速度给出了上下界.

[关键词] 线性方程组,最小二乘解,斜投影法

[中图分类号]O24; [文献标识码]A; [文章编号]1001-4616(2001)01-0011-05

0 引言

设 $A\in R_r^{m\times n}$,$b\in R^m$,解线性方程组 $Ax=b$ 的斜投影法为:

(1)初始点 x_1,取 $W_1=(w_1)\in R_1^{m\times 1}$,$V_1=(v_1)\in R_1^{n\times 1}$ 满足 $\det(W_1^TAV_1)\neq 0$,置 $k=1$.

(2)求 $x_{k+1}=x_1+V_ky_k$,满足 $W_k^T(Ax_{k+1}-b)=0$,即计算

$x_{k+1}=x_1+V_k(W_k^TAV_k)^{-1}W_k^T(b-Ax_1)$.

(3)置 $W_{k+1}=(W_k,w_{k+1})\in R_{k+1}^{m\times(k+1)}$,$V_{k+1}=(V_k,v_k)\in R_{k+1}^{n\times(k+1)}$,要求它们满足 $\det(W_{k+1}^TAV_{k+1})\neq 0$,置 $k=k+1$,转(2).

如何选择 W_k 与 V_k 使 $\det(W_k^TAV_k)\neq 0$ 且 $(W_k^TAV_k)^{-1}$易于计算?这个问题在文献[2]中我们已作了详细讨论.对斜投影法来讲,另一个重要问题是如何选择 W_k 与 V_k,使 $x_{k+1}=x_1+V_k(W_k^TAV_k)^{-1}W_k^T(b-Ax_1)$能快速收敛到解 x_e.显然,这个问题依赖于对 $\|x_{k+1}-x_e\|$ 的估计.本文的目标就是对 $\|x_{x+1}-x_e\|_2$ 给出一些估计.

1 $\|x_{k+1}-x_e\|_2$ 的估计

定理 1 设 $A\in R_r^{m\times n}$,$Ax=b$ 相容,则斜投影法至多 r 步必可求得 $Ax=b$ 的一个解 x_e,记 $P_k=P_{R(V_k),N(V_k^T)}=V_k(V_k^TV_k)^{-1}V_k^T$,$\varepsilon_k=\|(I-P_k)(x_e-x_1)\|_2$,则有如下估计

$$\varepsilon_k\leqslant\|x_{k+1}-x_e\|_2\leqslant\left[1+\|P_{R(V_k),N(W_k^TA)}-P_k\|_2^2\right]^{\frac{1}{2}}\varepsilon_k.$$

证明 如果 $x_1,x_2,\cdots,x_r$ 均不为 $Ax=b$ 的解,令 $G=V_r(W_r^TAV_r)^{-1}W_r^T$,由 $\mathrm{rank}(W_r^TAV_r)=\mathrm{rank}(A)=\mathrm{rank}(W_r)=\mathrm{rank}(V_r)=r$,知 $G\in A\{1,2\}$,故 $x_{r+1}=x_1+G(b-Ax_1)=Gb+(I-GA)x_1$ 必为 $Ax=b$ 的解,这个解记为 x_e.

$$W_k^TA(x_{k+1}-x_e)=W_k^T(Ax_{k+1}-Ax_e)=W_k^T(Ax_{k+1}-b)=0,$$

由 $x_{k+1}-x_1\in R(V_k)$知$(I-P_{R(V_k),N(W_k^TA)})(x_{k+1}-x_1)=0$,于是

收稿日期:2000-03-28

作者简介:颜世建,1946—,南京师范大学数学与计算机科学学院副教授,主要从事计算数学的教学与研究.

$$x_{k+1}-x_e=(I-P_{R(V_k),N(W_k^TA)})(x_{k+1}-x_1+x_1-x_e)=-(I-P_{R(V_k),N(W_k^TA)})(x_e-x_1),$$

又由 $I-P_{R(V_k),N(W_k^TA)}=I-P_k-P_{R(V_k),N(W_k^TA)}+P_k$,以及 $P_k=P_{R(V_k),N(W_k^TA)}\cdot P_k$ 得

$$I-P_{R(V_k),N(W_k^TA)}=(I-P_{R(V_k),N(W_k^TA)})(I-P_k).$$

注意到 $(I-P_k)(x_e-x_1)$ 与 $P_{R(V_k),N(W_k^TA)}(I-P_k)(x_e-x_1)$ 正交,得

$$\|x_{k+1}-x_e\|_2^2=\|-(I-P_k)(x_e-x_1)+P_{R(V_k),N(W_k^TA)}\cdot(I-P_k)(x_e-x_1)\|_2^2$$

$$=\|(I-P_k)(x_e-x_1)\|_2^2+\|P_{R(V_k),N(W_k^TA)}(I-P_k)(x_e-x_1)\|_2^2$$

$$=\|(I-P_k)(x_e-x_1)\|_2^2+\|(P_{R(V_k),N(W_k^TA)}-P_k)(I-P_k)(x_e-x_1)\|_2^2$$

$$\leqslant\varepsilon_k^2+\|P_{R(V_k),N(W_k^TA)}-P_k\|_2^2\varepsilon_k^2.$$

由 $\|x_{k+1}-x_e\|_2^2=\|(I-P_k)(x_e-x_1)\|_2^2+\|P_{R(V_k),N(W_k^TA)}(I-P_k)(x_e-x_1)\|_2^2$ 知

$$\|x_{k+1}-x_e\|_2^2\geqslant\|(I-P_k)(x_e-x_1)\|_2^2\geqslant\varepsilon_k^2.$$

因而

$$\varepsilon_k\leqslant\|x_{k+1}-x_e\|_2\leqslant[1+\|P_{R(V_k),N(W_k^TA)}-P_k\|_2^2]^{\frac{1}{2}}\varepsilon_k.$$

注 1 定理 1 中的下界 ε_k 是可以达到的,只需取 $v_1=A^Tw_1,v_2=A^Tw_2,\cdots,v_k=A^Tw_k$,就有 $P_{R(V_k),N(W_k^TA)}=P_{R(V_k),N(V_k^T)}=P_k$,这样就成立 $\|x_{k+1}-x_e\|_2=\varepsilon_k$.

注 2 显然 $\varepsilon_k=\|(I-P_k)(x_e-x_1)\|_2$ 随 k 增大而减少.虽然不会对 $Ax=b$ 的每一个解 x_e 有 $\varepsilon_k\to0$,但必存在一个解 x_e 及 $k\leqslant r$ 使 $\varepsilon_k=0$.

定理 2 设 $A\in R_r^{m\times n}$,$Ax=b$ 相容,x_e 为 $Ax=b$ 的极小范数解,若 $x_1\in R(A^T)$,$R(V_k)\subseteq R(A^T)$,$k\leqslant r$,则斜投影法产生的 $\{x_k\}$ 至多 r 步收敛于 x_e,且有估计

$$\varepsilon_k\leqslant\|x_{k+1}-x_e\|_2\leqslant[1+\|P_{R(V_k),N(W_k^TA)}-P_k\|_2^2]^{\frac{1}{2}}\varepsilon_k,$$

特别,若取 $V_k=A^TW_k$,$k\leqslant r$,则有 $\|x_{k+1}-x_e\|_2=\varepsilon_k$.

证明 若 $x_1,x_2,\cdots,x_r$ 均不等于 x_e,由定理 1 知 x_{r+1} 必使 $Ax_{r+1}=b$.由 $x_1\in R(A^T)$,$R(V_r)\subseteq R(A^T)$,知 $x_{r+1}=A^Ty$,由 $AA^Ty=b$ 知 $y=(AA^T)^-b$,$x_{r+1}=A^T(AA^T)^-b=x_e$,其余结论由定理 1 与注 1 立得.

定理 3 设 $A\in R_r^{m\times n}$,$Ax=b$ 不相容,x_e 为 $Ax=b$ 的极小范数最小二乘解.如果 $x_1\in R(A^T)$,$R(V_k)\subseteq R(A^T)$,$R(W_k)\subseteq R(A)$,$k\leqslant r$,则由斜投影法产生的 $\{x_k\}$ 至多 r 步必收敛到 x_e.若记 $S_k=P_{R(V_k),N(W_k^T)}$,$M_k=S_k^TS_k+(I-S_k)^T(I-S_k)$,用 $\lambda_{\max}^{(k)},\lambda_{\min}^{(k)}$ 分别表示 M_k 的最大与最小特征值,P_k 以及 ε_k 与定理 1 所给相同,则有以下估计:

$$\min_{x\in x_1+R(V_k)}\|Ax-b\|_2\leqslant\|Ax_{k+1}-b\|_2\leqslant\left[\frac{\lambda_{\max}^{(k)}}{\lambda_{\min}^{(k)}}\right]^{\frac{1}{2}}\min_{x\in x_1+R(V_k)}\|Ax-b\|_2,$$

$$\varepsilon_k\leqslant\|x_{k+1}-x_e\|_2\leqslant[1+\|P_{R(V_k),N(W_k^TA)}-P_k\|_2^2]^{\frac{1}{2}}\varepsilon_k.$$

证明 如果 $x_1,x_2,\cdots,x_r$ 均不等于 x_e,令 $G=V_r(W_r^TAV_r)^{-1}W_r^T$,由 $\mathrm{rank}(W_r^TAV_r)=\mathrm{rank}(A)=\mathrm{rank}(W_r)=\mathrm{rank}(V_r)=r$ 知 $G=A_{R(V_r),N(W_r^T)}^{(1,2)}$,由 $R(V_r)=R(A^T)$,$N(W_r^T)=N(A^T)$,故 $G=A_{R(A^T),N(A^T)}^{(1,2)}=A^+$,而 $x_1\in R(A^T)$,因而 $(I-GA)x_1=0$,这样 $x_{r+1}=x_1+G(b-Ax_1)=Gb+(I-GA)x_1=Gb=A^+b=x_e$.

其次,对 $\forall\, y\in R^m$ 且 $y\neq 0$,$y=S_ky+(I-S_k)y$ 且 S_ky 与 $(I-S_k)y$ 不同时为 0,故 $y^TM_ky>0$,这说明 M_k 对称正定,$\lambda_{\min}^{(k)}>0$.

令 $G_k=(W_k^TAV_k)^{-1}W_k^T$. 由 $\mathrm{rank}(W_k^TAV_k)=\mathrm{rank}(AV_k)=\mathrm{rank}(W_k)=K$ 知 $G_k\in AV_k\{1\}$,显然 $AV_kG_k=AV_k(W_k^TAV_k)^{-1}W_k=S_k$,因而 $(M_kAV_kG_k)^T=S_k^TS_k=M_kAV_kG_k$,这说明 $G_k=(AV_k)^-_{l(M_k)}$. 因而 $y_k=G_kr_1$ 为 $\min\limits_{y\in R^k}\|AV_ky-r_1\|_{M_k}^2$ 的解,从

$$\lambda_{\min}^{(k)}\|AV_ky_k-r_1\|_2^2\leqslant\|AV_ky_k-r_1\|_{M_k}^2\leqslant\lambda_{\max}^{(k)}\|AV_ky_k-r_1\|_2^2$$

可知,对 $\forall\, y\in R^k$ 有

$$\|Ax_{k+1}-b\|_2^2=\|AV_ky_k-r_1\|_2^2\leqslant(\lambda_{\min}^{(k)})^{-1}\|AV_ky_k-r_1\|_{M_k}^2\leqslant(\lambda_{\min}^{(k)})^{-1}\|AV_ky-r_1\|_{M_k}^2$$

$$\leqslant(\lambda_{\min}^{(k)})^{-1}\lambda_{\max}^{(k)}\|AV_ky-r_1\|_2^2=\left[\frac{\lambda_{\max}^{(k)}}{\lambda_{\min}^{(k)}}\right]\|A(x_1+V_ky)-b\|_2^2$$

故有,$\|Ak_{k+1}-b\|_2^2\leqslant\left[\dfrac{\lambda_{\max}^{(k)}}{\lambda_{\min}^{(k)}}\right]\min\limits_{x\in x_1+R(V_k)}\|Ax-b\|_2^2$.

由 $x_{k+1}\in x_1+R(V_k)$ 知 $\min\limits_{x\in x_1+R(V_k)}\|Ak-b\|_2\leqslant\|Ax_{k+1}-b\|_2^2$. 因而

$$\min_{x\in x_1+R(V_k)}\|Ax-b\|_2\leqslant\|Ax_{x+1-b}\|_2\leqslant\left[\frac{\lambda_{\max}^{(k)}}{\lambda_{\min}^{(k)}}\right]^{\frac{1}{2}}\min_{x\in x_1+R(V_k)}\|Ax-b\|_2.$$

最后,由 $A^T(Ax_e-b)=0$,$R(W_k)\subseteq R(A)$,$k\leqslant r$ 知 $W_k^T(Ax_e-b)=0$,注意到 $W_k^T(Ax_{k+1}-b)=0$ 知 $W_k^TA(x_{k+1}-x_e)=0$,以下与定理 1 的证明完全一样,可得

$$\varepsilon_k\leqslant\|x_{k+1}-x_e\|_2\leqslant[1+\|P_{R(V_k),N(W_k^TA)}-P_k\|_2^2]^{\frac{1}{2}}\varepsilon_k.$$

2 $A\in R_n^{n\times n}$ 且 $W_k=V_k$ 时的估计

$A\in R_n^{n\times n}$ 且 $W_k=V_k$ 时,斜投影法就称为投影法,对某些特殊的 A,投影法收敛速度的估计可以进一步深入讨论.

定理 4 设 $A\in R_n^{n\times n}$,A 有奇异值分解 $A=U\mathrm{diag}(\sigma_1,\cdots,\sigma_n)V^T$,记 $\bar{A}=V\mathrm{diag}(\sigma_1,\cdots,\sigma_n)V^T$,若 $\|A-\bar{A}\|_2<\sigma_n$,则投影法有如下估计

$$\varepsilon_k\leqslant\|x_{k+1}-x_e\|_2\leqslant[1+(\sigma_n-\|A-\bar{A}\|_2)^{-2}\|A-I\|_2^2]^{\frac{1}{2}}\varepsilon_k.$$

证明 设 $V_k=\tilde{V}_kR_k$. 其中 $\tilde{V}_k$ 的列标准正交,$R_k\in R_k^{k\times k}$ 为上三角阵,则 $P_{R(V_k),N(V_k^TA)}=\tilde{V}_k(\tilde{V}_k^TA\tilde{V}_k)^{-1}\tilde{V}_k^T$,$P_k=\tilde{V}_k\tilde{V}_k^T$,于是

$$P_{R(V_k),N(V_k^TA)}-P_k=\tilde{V}_k(\tilde{V}_k^TA\tilde{V}_k)^{-1}\tilde{V}_k^T(A-I)(I-\tilde{V}_k\tilde{V}_k^T),$$

$$\|P_{R(V_k),N(V_k^TA)}-P_k\|_2\leqslant\|(\tilde{V}_k^TA\tilde{V}_k)^{-1}\|_2\|A-I\|_2.$$

用 α_i,$\bar{\alpha}_i$,$i=1,\cdots,k$,分别表示 $\tilde{V}_k^TA\tilde{V}_k$ 与 $\tilde{V}_k^T\bar{A}\tilde{V}_k$ 的奇异值,并按递减顺序排列,由

$$\|\alpha_k-\bar{\alpha}_k\|\leqslant\|\tilde{V}_k^TA\tilde{V}_k-\tilde{V}_k^T\bar{A}\tilde{V}_k\|_2\leqslant\|A-\bar{A}\|_2<\sigma_n.$$

可知 $\alpha_k\geqslant\bar{\alpha}_k-\|A-\bar{A}\|_2$,再由 $\tilde{V}_k$ 的列标准正交知 $\bar{\alpha}_k\geqslant\sigma_n$,故 $\alpha_k\geqslant\sigma_n-\|A-\bar{A}\|_2>0$. 因而有

$$\varepsilon_k\leqslant\|x_{k+1}-x_e\|_2\leqslant[1+\|(\tilde{V}_k^TA\tilde{V}_k)^{-1}\|_2^2\|A-I\|_2^2]^{\frac{1}{2}}\varepsilon_k$$

$$=[1+\alpha_k^{-2}\|A-I\|_2^2]^{\frac{1}{2}}\varepsilon_k\leqslant[1+(\sigma_n-\|A-\bar{A}\|_2)^{-2}\|A-I\|_2^2]^{\frac{1}{2}}\varepsilon_k.$$

推论 设 A 为对称正定矩阵,则投影法有

$$\varepsilon_k \leqslant \| x_{k+1} - x_e \|_2 \leqslant [1 + \| A^{-1} \|_2^2 \| A - I \|_2^2]^{\frac{1}{2}} \varepsilon_k .$$

证明　这时 $A = \bar{A}$，$\sigma_n^{-1} = \| A^{-1} \|_2$，由定理 4 立得.

定理 5　设 $A \in R_n^{n \times n}$ 为正实阵，则投影法有

$$\varepsilon_k \leqslant \| x_{k+1} - x_e \|_2 \leqslant \left[1 + \| A - I \|_2^2 \left(\frac{A + A^T}{2} \right)^{-1} \|_2^2 [1 + \left(1 - \| A \|_2^{-2} \| \left(\frac{A + A^T}{2} \right)^{-1} \|_2^{-2} \right)^{\frac{1}{2}}]^2 \right]^{\frac{1}{2}} \varepsilon_k .$$

证明　记 $\lambda_1 > 0$ 为 $A^T A$ 最大的特征值，$\tilde{\lambda}_n > 0$ 为 $\frac{A + A^T}{2}$ 最小的特征值，$\bar{\alpha} = \tilde{\lambda}_n / \lambda_1$，则

$$\| \bar{\alpha} A - I \|_2^2 = \max_{\| y \|_2 = 1} y^T \left[\bar{\alpha}^2 A^T A + I - 2\bar{\alpha} \left(\frac{A + A^T}{2} \right) \right] y$$
$$\leqslant \bar{\alpha}^2 \lambda_1 + 1 - 2\bar{\alpha} \tilde{\alpha}_n = 1 - \tilde{\lambda}_n^2 / \lambda_1 .$$

如果 $\tilde{\lambda}_n^2 / \lambda_1 = 1$ 则 $A = \frac{1}{2} I$，由前述推论可知本定理结论成立，以下设 $\tilde{\lambda}_n^2 / \lambda_1 \neq 1$.

由于 $\tilde{\lambda}_i \left(\frac{A + A^T}{2} \right) \leqslant [\lambda_i (A^T A)]^{\frac{1}{2}}$，$i = 1, \cdots, n$，故 $1 - \frac{\tilde{\lambda}_n^2}{\lambda_1} \in (0, 1)$. 由定理 4 的证明知

$$\| P_{R(V_k), N(V_k^T A)} - P_k \|_2 \leqslant \| (\tilde{V}_k^T A \tilde{V}_k)^{-1} \|_2 \| A - I \|_2 .$$

设 σ_k 为 $\tilde{V}_k^T (\bar{\alpha} A) \tilde{V}_k$ 的最小奇异值，则

$$\| \sigma_k - 1 \| \leqslant \| \tilde{V}_k^T (\bar{\alpha} A - I) \tilde{V}_k \|_2 \leqslant \| \bar{\alpha} A - I \|_2 \leqslant \left(1 - \frac{\tilde{\lambda}_n^2}{\lambda_1} \right)^{\frac{1}{2}} < 1,$$

因而 $\sigma_k^{-1} \leqslant \left[1 - \left(1 - \frac{\tilde{\lambda}_n^2}{\lambda_1} \right)^{\frac{1}{2}} \right]^{-1}$，于是

$$\| (\tilde{V}_k^T A \tilde{V}_k)^{-1} \|_2 = \bar{\alpha} \| (\tilde{V}_k^T (\bar{\alpha} A) \tilde{V}_k)^{-1} \|_2 \leqslant \bar{\alpha} \sigma_k^{-1} \leqslant \bar{\alpha} \left[1 - \left(1 - \frac{\tilde{\lambda}_n^2}{\lambda_1} \right)^{\frac{1}{2}} \right]^{-1} .$$

这样

$$\| P_{R(V_k), N(V_k^T A)} - P_k \|_2 \leqslant \| A - I \|_2 \cdot \bar{\alpha} \left[1 + \left(1 - \frac{\tilde{\lambda}_n^2}{\lambda_1} \right)^{\frac{1}{2}} \right] / (\tilde{\lambda}_n^2 / \lambda_1)$$
$$= \| A - I \|_2 \left[1 + \left(1 - \frac{\tilde{\lambda}_n^2}{\lambda_1} \right)^{\frac{1}{2}} \right] \tilde{\lambda}_n^{-1} .$$

注意到 $\tilde{\lambda}_n^{-1} = \| \left(\frac{A + A^T}{2} \right)^{-1} \|_2$，$\lambda_1 = \| A \|_2^2$，则有

$$\varepsilon_k \leqslant \| x_{k+1} - x_e \|_2 \leqslant \left[1 + \| A - I \|_2^2 \| \left(\frac{A + A^T}{2} \right)^{-1} \|_2^2 [1 + \left(1 - \| A \|_2^{-2} \| \left(\frac{A + A^T}{2} \right)^{-1} \|_2^{-2} \right)^{\frac{1}{2}}]^2 \right]^{\frac{1}{2}} \varepsilon_k .$$

[参考文献]

[1] Saad Y. The Lanczos Biorthogonalization Algorithm and Other Oblique Projection Methods for Solving Large Unsymmetric Systems[J]. SIAMJ. Numer. Anal.,1982,19(3):458—505.

[2] Yan Shijian,Huang Kaibin. Constructions of Bases for Extended Oblique Projection Method[J]. Numer. Math. J. Chin. Univ. English Series,1995,4(2):176—182.

[3] Huang Kaibin,Yan Shijian. On uniqueness condition of mid-solution for oblique projection methods[J]. Bulletin of Science,P. R. China,1992,37(20):1919—1920.

The Estimate of Convergence Rate for the Oblique Projection Method

Yan Shijian

(College of Mathematics and Computer Sciences,Nanjing Normal University,Nanjing,210097,PRC)

Abstract:In this paper,we present the error estimate of the approximate solution obtained by the oblique projection method for solving linear equations $Ax = b$.

Key words:linear equations;least squares solution;oblique projection method

[责任编辑:陆炳新]