DOI: 10.13656/j.cnki.gxkx.2009.02.010
广西科学 Guangxi Sciences 2009, 16(2): 109~112

一个新的解非线性对称方程组的非单调共轭梯度方法*

A New Nonmonotone Conjugate Gradient Method for Symmetric Nonlinear Equations

袁功林,李向荣
YUAN Gong-lin, LI Xiang-rong

(广西大学数学与信息科学学院,广西南宁 530004)
(College of Mathmatics and Information Sciences, Guangxi University, Nanning, Guangxi, 530004, China)

摘要: 给出一个新的解非线性对称方程组: $g(x)=0(x\in R^n, g: R^n\rightarrow R^n$ 连续可微,并且其雅克比矩阵 $\nabla g(x)$ 在 $x\in R^n$ 上对称)的非单调共轭梯度方法,分析新方法的全局收敛性,并用数值实验来检验其有效性.新方法全局收敛,在不执行任意线搜索的条件下能够确保搜索方向的下降性,而且初始点的选择与维数的增加并不明显影响检验结果.

关键词: 共轭梯度方法 非单调 对称方程组
中图法分类号: O221 文献标识码: A 文章编号: 1005-9164(2009)02-0109-04

Abstract A new nonmonotone conjugate gradient method is presented for solving symmetric nonlinear equations $g(x)=0(x\in R^n, g: R^n\rightarrow R^n$ is continuously differentiable and its Jacobian $\nabla g(x)$ is symmetric for all $x\in R^n$). The global convergence of the method is established under suitable conditions. Numerical results show that this method is effective. The search direction is descent without any line search technique. Moreover, the initial points and the increase of dimension don't influence the performance of the presented method.

Key words conjugate gradient method, nonmonotone, symmetric equations

有许多方法用来求解无约束优化问题[1~5]:

$$\min_{x\in R^n} f(x). \qquad (0.1)$$

其中共轭梯度方法是非常重要的一类,特别是求解大规模优化问题.该方法是用下列迭代公式来求解问题(0.1).

$$x_{k+1}=x_k+T_k d_k, \qquad (0.2)$$

$$d_k=-g_k+U_k d_{k-1}, \qquad (0.3)$$

其中 $T_k>0$ 是步长,d_k 是搜索方向,并且 $d_0=-\nabla f(x_0)=-g_0$,$U_k\in R$.许多著名的 U_k 计算公式可以参见文献[6~11].

近年来,非单调技术已被广泛应用于求解无约束优化问题(0.1)[12~20].文献[13]给出一个非单调共轭梯度类型方法来求解问题(0.1),并且在适当的假设条件下建立了全局收敛性结果.受文献[13]的启发,本文给出一个新的求解下述非线性对称方程组的非单调共轭梯度方法.

$$g(x)=0, x\in R^n, \qquad (0.4)$$

其中 $g: R^n\rightarrow R^n$ 连续可微,并且其雅克比矩阵 $\nabla g(x)$ 在 $x\in R^n$ 上是对称的.所给方法在不执行任意线搜索的条件下能确保搜索方向的下降性,并具有全局收敛性.问题(0.4)可来源于无约束优化问题[21].易知,非线性方程组问题(0.4)等价于全局优化问题:

$$\min\theta(x), x\in R^n, \qquad (0.5)$$

其中 $\theta(x)=\frac{1}{2}\|g(x)\|^2$.文中 $\|\cdot\|$ 指欧氏向量或矩阵范数,$g_k=g(x_k)$,$\theta_k=\theta(x_k)$.

1 算法

算法 1

步骤 0 给定初始点 $x_0\in R^n$,常数 $r\in(0,1)$,$0<W_1<W_2<1$.置 $d_0=-\nabla\theta(x_0)$,且 $k=0$;

步骤 1 如果 $\nabla\theta_0=0$ 停止;否则,转步骤 2;

收稿日期: 2008-09-15
作者简介: 袁功林(1976-),男,副教授,主要从事优化理论与方法的研究.
* Supported by China NSF Grands 10761001 and the Scientific Research Foundation of Guangxi University (Grant No. X081082).

步骤 2 令 i_k 是满足下式最小的非负整数 i,

$$\theta(x_k+\lambda d_k)\leqslant \theta(x_{l(k)})+W_1\lambda\nabla\theta(x_k)^T d_k, \qquad (1.1)$$

$$\nabla\theta(x_k+\lambda d_k)^T d_k\geqslant W_2\nabla\theta(x_k)^T d_k, \qquad (1.2)$$

其中 $\lambda = r^i$, $\theta_{l(k)} = \max_{0\leqslant j\leqslant M(k)}\{\theta_{k-j}\}$, $M(k) = \min\{M,k\}$, $k = 0,1,2,\cdots$, $M\geqslant 0$ 是一个给定的整常数, $\nabla\theta(x) = \nabla g(x)^T g(x)$. 令 $\lambda_k = r^{i_k}$, $x_{k+1} = x_k+\lambda_k d_k$. 转步骤 3;

步骤 3 计算

$$d_{k+1} = -\nabla\theta_{k+1}+U_{k+1}d_k, \qquad (1.3)$$

其中

$$U_{k+1} = \frac{\|\nabla\theta_{k+1}\|^2}{\max\{-\nabla\theta_k^T d_k, d_k^T y_k\}}; \qquad (1.4)$$

$y_k = \nabla\theta_{k+1}-\nabla\theta_k$.

步骤 4 Let $k := k+1$, 转步骤 1.

2 全局收敛性分析

设 K 是如下定义的水平集.

$$K = \{x|\theta(x)\leqslant\theta(x_0)\}. \qquad (2.1)$$

给出假设条件 A: (i) g 在包含于 K 的开凸集 K_1 上是连续可微的,并且 $\|g(x)\|$ 在 K 是下有界的. (ii) g 的雅克比矩阵是对称的且存在常数 $M>0$ 满足

$$\|\theta(x)-\theta(y)\|\leqslant M\|x-y\|, \forall x,y\in K_1. \qquad (2.2)$$

引理 2.1 若假设条件 A 满足并且 $\{\lambda_k, x_{k+1}, g(x_{k+1}), d_{k+1}\}$ 由算法 1 产生. 则算法 1 在有限步内将产生下一迭代点 $x_{k+1} = x_k+\lambda_k d_k$.

证明 利用 Wolfe 准则和文献 [22] 中的 Lemma3.8, 在有限步内, λ_k 满足 $\theta(x_k+\lambda d_k)\leqslant\theta(x_k)+W_1\lambda\nabla\theta(x_k)^T d_k$ 和 $\nabla\theta(x_k+\lambda d_k)^T d_k\geqslant W_2\nabla\theta(x_k)^T d_k$. 又由于 $\theta(x_k)\leqslant\theta(x_{l(k)})$, 则结论成立.

注 利用 (1.2) 式和 (2.2) 式, 容易获得

$$\lambda_k\geqslant c\frac{-\nabla\theta_k^T d_k}{\|d_k\|^2}, \qquad (2.3)$$

其中 $c = \frac{1-W_2}{M}$.

引理 2.2 设 $\{\lambda_k, x_{k+1}, g(x_{k+1}), d_{k+1}\}$ 由算法 1 产生. 则对所有 k,

$$\nabla\theta_k^T d_k < 0. \qquad (2.4)$$

证明 当 $k = 0$ 时, 利用 $d_0 = -\nabla\theta_0$, 容易获得 (2.4) 式. 假设 (2.4) 式对 $k\geqslant 1$ 成立, (1.3) 式子的两边同时乘上 $\nabla\theta_{k+1}^T$ 并利用 (1.4) 式, 可得

$$\nabla\theta_{k+1}^T d_{k+1} = \frac{\nabla\theta_{k+1}^T d_k-\max\{-d_k^T\nabla\theta_k, d_k^T y_k\}}{\max\{-d_k^T\nabla\theta_k, d_k^T y_k\}}\|\nabla\theta_{k+1}\|^2. \qquad (2.5)$$

情形 1 如果 $\nabla\theta_{k+1}^T d_k\geqslant 0$, 容易获得 $d_k^T y_k>0$ 并且 (2.5) 重写为

$$\nabla\theta_{k+1}^T d_{k+1} = \frac{\nabla\theta_k^T d_k}{d_k^T y_k}\|\nabla\theta_{k+1}\|^2. \qquad (2.6)$$

这表明 $\nabla\theta_{k+1}^T d_{k+1}<0$ 成立.

情形 2 如果 $\nabla\theta_{k+1}^T d_k<0$, (2.5) 式重写为

$$\nabla\theta_{k+1}^T d_{k+1} = \frac{-\nabla\theta_k^T d_{k-1}-\nabla\theta_k^T d_k}{d_k^T\nabla\theta_k}\|\nabla\theta_{k+1}\|^2. \qquad (2.7)$$

这意味着 $\nabla\theta_{k+1}^T d_{k+1}<0$ 成立.

因此两种情形下 (2.4) 式对 $k+1$ 都满足. 所以由归纳法可知 (2.4) 式对所有 k 都满足.

引理 2.3 设 $\{\lambda_k, x_{k+1}, g(x_{k+1}), d_{k+1}\}$ 由算法 1 产生且假设条件 A 满足. 则

$$\sum_{\ell\geqslant 0}\min_{i=1,2,\cdots,M}\left\{\frac{(\nabla\theta_{M\ell+i}^T d_{M\ell+i})^2}{\|d_{M\ell+i}\|^2}\right\}<+\infty. \qquad (2.8)$$

证明 对 $\ell\geqslant 0$, 定义

$$T_\ell = \max\{\|g_{M\ell+1}\|^2, \|g_{M\ell+2}\|^2, \cdots, \|g_{M(\ell+1)}\|^2\}. \qquad (2.9)$$

利用归纳法证明

$$\|g_{M\ell+i}\|^2\leqslant T_{\ell-1}+W_1\lambda_{M\ell+i}\nabla\theta_{M\ell+i}^T d_{M\ell+i} \qquad (2.10)$$

对所有 $i = 1,2,\cdots,M$ 都成立. 由 (1.1) 式, 当 $i = 1$ 时, (2.10) 式显然成立. 假设 (2.10) 式对所有 $1\leqslant i\leqslant M-1$ 都成立. 利用引理 2.2, 可得

$$\|g_{M\ell+i}\|^2\leqslant T_{\ell-1}. \qquad (2.11)$$

利用 (1.1) 式和 (2.11) 式, 有

$$\|g_{M\ell+i+1}\|^2\leqslant\max\{\|g_{M\ell+i}\|^2, T_{\ell-1}\}+W_1\lambda_{M\ell+i+1}\nabla\theta_{M\ell+i+1}^T d_{M\ell+i+1}\leqslant T_{\ell-1}+W_1\lambda_{M\ell+i+1}\nabla\theta_{M\ell+i+1}^T d_{M\ell+i+1}. \qquad (2.12)$$

因此, (2.10) 式对 $i+1$ 也成立. 所以 (2.10) 式对 $i = 1,2,\cdots,M$ 都满足.

利用 (2.9) 式, (2.10) 式和 (2.3) 式, 可获得

$$T_\ell\leqslant T_{\ell-1}-W\min_{i=1,2,\cdots,M}\left\{\frac{(\nabla\theta_{M\ell+i}^T d_{M\ell+i})^2}{\|d_{M\ell+i}\|^2}\right\}. \qquad (2.13)$$

利用假设条件 A(i), 知道 $\|g(x)\|$ 是下有界的, 这意味着 $\{T_\ell\}$ 也是有界的. 以 ℓ 为下标将 (2.13) 式相加, 就可以得到 (2.8) 式.

定理 2.1 设 $\{\lambda_k, x_{k+1}, g(x_{k+1}), d_{k+1}\}$ 由算法 1 产生, 假设 A 满足. 则

$$\liminf_{k\to\infty}\|\nabla\theta(x_k)\| = 0. \qquad (2.14)$$

证明 假设 $\liminf_{k\to\infty}\|\nabla\theta(x_k)\|\neq 0$, 则存在一个正常数 X_0 满足

$$\|\nabla\theta(x_k)\|>X_0, \forall k\geqslant 0. \qquad (2.15)$$

利用 (2.7) 式, (1.4) 式重写为

$$U_{k+1} = \frac{\nabla\theta_{k+1}^T d_{k+1}}{(\max\{\nabla\theta_{k+1}^T d_k, 0\}-\nabla\theta_{k+1}^T d_k)-\nabla\theta_k^T d_k}. \qquad (2.16)$$

联立(2.4)式和(2.16)式,可得

$$\frac{\nabla\theta_{k+1}^T d_{k+1}}{d_k^T\nabla\theta_k} \geqslant \beta_{k+1} \geqslant 0. \qquad (2.17)$$

又由(1.3)式,得到 $d_k + \nabla\theta_k = \beta_k d_{k-1}$,两边同时平方并利用(2.17)式,得到

$$\|d_k\|^2 = -2\nabla\theta_k^T d_k - \|\nabla\theta_k\|^2 + \beta_k^2\|d_{k-1}\|^2 \leqslant$$
$$-2\nabla\theta_k^T d_k - \|\nabla\theta_k\|^2 + \left[\frac{\nabla\theta_k^T d_k}{d_{k-1}^T\nabla\theta_{k-1}}\right]^2\|d_{k-1}\|^2. \qquad (2.18)$$

将(2.18)式除以 $(\nabla\theta_k^T d_k)^2$,得

$$\frac{\|d_k\|^2}{(\nabla\theta_k^T d_k)^2} \leqslant -\left(\frac{\|\nabla\theta_k\|^2}{(\nabla\theta_k^T d_k)^2} + \frac{2}{\nabla\theta_k^T d_k} + \frac{1}{\|\nabla\theta_k\|^2}\right) + \frac{1}{\|\nabla\theta_k\|^2} + \frac{\|d_{k-1}\|^2}{(\nabla\theta_{k-1}^T d_{k-1})^2} = -\left[\frac{\|\nabla\theta_k\|}{\nabla\theta_k^T d_k} + \frac{1}{\|\nabla\theta_k\|}\right]^2 + \frac{1}{\|\nabla\theta_k\|^2} + \frac{\|d_{k-1}\|^2}{(\nabla\theta_{k-1}^T d_{k-1})^2} \leqslant \frac{1}{\|\nabla\theta_k\|^2} + \frac{\|d_{k-1}\|^2}{(\nabla\theta_{k-1}^T d_{k-1})^2}. \qquad (2.19)$$

利用(2.15)式和(2.19)式,可得到

$$\frac{\|d_k\|^2}{(\nabla\theta_k^T d_k)^2} \leqslant \sum_{j=0}^{k}\frac{1}{\|\nabla\theta_j\|^2} \leqslant \frac{k+1}{\gamma_0^2}. \qquad (2.20)$$

因此

$$\frac{(\nabla\theta_k^T d_k)^2}{\|d_k\|^2} \geqslant \frac{\gamma_0^2}{k+1}. \qquad (2.21)$$

利用(2.21)式,容易推出

$$\sum_{k\geqslant 0}\min_{i=1,2,\cdots,M}\left\{\frac{(\nabla\theta_{Mk+i}^T d_{Mk+i})^2}{\|d_{Mk+i}\|^2}\right\} = +\infty.$$

这和(2.8)式产生矛盾.因此(2.14)式成立.

3 数值检验

问题 选取与文献[23]类似的离散两点边界值问题:

$$g(x) = Ax + \frac{1}{(n+1)^2}F(x) = 0,$$

其中 A 是 $n\times n$ 三对角矩阵,定义为

$$A = \begin{bmatrix} 4 & -1 & & & & \\ -1 & 4 & -1 & & & \\ & -1 & 4 & -1 & & \\ & & \cdots & \cdots & \cdots & \\ & & & \cdots & \cdots & -1 \\ & & & & -1 & 4 \end{bmatrix},$$

$F(x) = (F_1(x), F_2(x), \cdots, F_n(x))^{\mathrm{T}}$, $F_i(x) = \sin x_i - 1$, $i = 1,2,\cdots,n$ 进行数值检验.检验中,算法1中的参量选为 $r = 0.05$, $\sigma_1 = 0.1$, $\sigma_2 = 0.9$. 利用 MATLAB6.5.1运行程序.当条件 $\|\nabla\theta(x)\| \leqslant 10^{-6}$ 时程序停止.表1表2中各栏意义为: Dim是问题的维数, NI是总的迭代次数, NG是函数值的迭代次数, GG是结束时函数值.

从表1和表2的结果可以看出,利用算法1对所给问题的检验是有效的,并且初始点的选择与维数的增加并不明显影响检验结果.

表**1** 小规模检验

Table 1 Small-scale tests

Dim (n)	NI/NG/GG					
	(1,⋯,1)	(50,⋯,50)	(100,⋯,100)	(−1,⋯,−1)	(−50,⋯,−50)	(−100,⋯,−100)
10	38/77/7.663105e−07	47/95/7.324909e−07	48/97/9.438427e−07	38/77/7.721902e−07	47/95/7.325242e−07	48/97/9.438180e−07
50	43/87/7.882965e−07	53/107/6.717607e−07	54/109/8.945024e−07	43/87/7.885965e−07	53/107/6.717657e−07	54/109/8.945056e−07
100	44/89/7.897887e−07	54/109/6.801967e−07	55/111/9.064276e−07	44/89/7.898657e−07	54/109/6.801980e−07	55/111/9.064285e−07
Dim (n)	NI/NG/GG					
	(1,0,1,0,⋯)	(50,0,50,0,⋯)	(100,0,100,0,⋯)	(−1,0,−1,0,⋯)	(−50,0,−50,0,⋯)	(−100,0,−100,0,⋯)
10	37/75/6.989783e−07	46/93/6.674146e−07	47/95/8.629571e−07	37/75/7.088681e−07	46/93/6.675344e−07	47/95/8.629878e−07
50	42/85/6.939959e−07	51/103/8.885944e−07	53/107/7.872924e−07	42/85/6.945328e−07	51/103/8.886079e−07	53/107/7.872983e−07
100	43/87/7.119706e−07	52/105/9.203034e−07	54/109/8.169964e−07	43/87/7.121099e−07	52/105/9.203070e−07	54/109/8.169979e−07

表**2** 大规模检验

Table 2 Large-scale tests

Dim (n)	NI/NG/GG					
	(1,⋯,1)	(50,⋯,50)	(100,⋯,100)	(−1,⋯,−1)	(−50,⋯,−50)	(−100,⋯,−100)
400	46/93/7.283977e−07	55/111/9.461489e−07	57/115/8.407951e−07	46/93/7.284022e−07	55/111/9.461490e−07	57/115/8.407951e−07
700	46/93/9.682958e−07	56/113/8.389176e−07	58/117/7.455928e−07	46/93/9.682977e−07	56/113/8.389176e−07	58/117/7.455928e−07
1000	47/95/7.729904e−07	57/115/6.698745e−07	58/117/8.931203e−07	47/95/7.729911e−07	57/115/6.698745e−07	58/117/8.931203e−07
Dim (n)	NI/NG/GG					
	(1,0,1,0,⋯)	(50,0,50,0,⋯)	(100,0,100,0,⋯)	(−1,0,−1,0,⋯)	(−50,0,−50,0,⋯)	(−100,0,−100,0,⋯)
400	45/91/6.749606e−07	54/109/8.767142e−07	56/113/7.790873e−07	45/91/6.749688e−07	54/109/8.767144e−07	56/113/7.790874e−07
700	54/114/7.898739e−07	64/134/6.839158e−07	65/136/9.117921e−07	54/114/7.898818e−07	64/134/6.839160e−07	65/136/9.117922e−07
1000	50/106/6.962799e−07	59/124/8.854997e−07	61/128/7.845766e−07	50/106/6.962862e−07	59/124/8.854998e−07	61/128/7.845767e−07

参考文献:

[1] Yuan G L, Lu X W. A new line search method with trust region for unconstrained optimization[J]. Communications on Applied Nonlinear Analysis, 2008, 15(1): 35-49.

[2] Yuan G L, Lu X W. A modified PRP conjugate gradient method [J]. Annals of Operations Research, 2009, 166 73-90.

[3] Yuan G L, Wei Z X. New line search methods for unconstrained optimization[J]. Journal of the Korean Statistical Society, 2009, 38 29-39.

[4] Yuan G, Lu X, Wei Z X. New two-point stepsize gradient methods for solving unconstrained optimization problems [J]. Natural Science Journal of Xiangtan University, 2007, 29(1): 13-15.

[5] 袁功林,韦增欣.一个新的 BFGS信赖域方法 [J].广西科学, 2004, 11: 195-196.

[6] Fletcher R, Reeves C. Function minimization buconjugate gradients[J]. Compute J, 1964, 7 149-154.

[7] Dai Y, Yuan Y. A nonlinear conjugate gradient with a strong global convergence properties[J]. SIAM Journal of Optimization, 2000, 10 177-182.

[8] Fletcher R. Practical method of optimization, VolⅠ: unconstrained optimization[M]. 2nd Edition. New York Wiley, 1997.

[9] Hestenes M R, Stiefel E. Method of conjugate gradient for solving linear equations [J]. J Res Nat Bur Stand, 1952, 49 409-436.

[10] Polak E, Ribiere G. Note sur la xonvergence dedirections conjugees [J]. Rev Francaise Informat Recherche Operatinelle, 3e Annee, 1969, 16 35-43.

[11] Liu Y, Storey C. Effcient generalized conjugate gradient algorithms, part 1: theory [J]. Journal of Optimization Theory and Application, 1992, 69 17-41.

[12] Dai Y. On the nonmonotone line search[J]. J Optim Theory Appl, 2002, 112 315-330.

[13] Dai Y. A nonmonotone conjugate gradient algorithm for unconstrained optimization [J]. J Syst Sci Complex, 2002, 15 139-145.

[14] Ranier E R, Tits A L. Avoiding the maratos effect by means of a nonmonotone line search[J]. SIAM J Numer Anal, 1991, 28 1183-1195.

[15] Raydan M. The Barzilai and Borwein gradient methodfor the large scale unconstrained minimization problem [J]. SIAM J Optim, 1997, 7: 26-33.

[16] Toint P L. An assessment of non-monotone line search techniques for unconstrained optimization [J]. SIAM J Sci Comput, 1996, 17 725-739.

[17] Grippo L, Lampariello F, Lucidi S. A nonmonotone line search technique for Newton's method [J]. SIAM J Numer Anal, 1986, 23 707-716.

[18] Grippo L, Lampariello F, Lucidi S. A truncated Newton method with nonmonotone line search for unconstrained optimization[J]. J Optim Theory Appl, 1989, 60 401-419.

[19] Yuan G L, Wei Z X. The superlinear convergence analysis of a nonmonotone BFGS algorithm on convex objective functions [J]. Acta Mathematica Sinica, English Series, 2008, 24(1): 35-42.

[20] Zhang H C, Hager W W. A nonmonotone line search technique and its application to unconstrained optimization [J]. SIAM J Optim, 2004, 4(4): 1043-1056.

[21] Li D, Fukushima M. A global and superlinear convergent Gauss-Newton-based BFGS method for symmetric nonlinear equations [J]. SIAM Journal on Numerical Analysis, 1999, 37: 152-172.

[22] Brown P N, Saad Y. Convergence theory of nonlinear Newton Kryloy algorithm[J]. SIAM J Optim, 1994(4): 297-330.

[23] Ortega J M, Rheinboldt W C. Iterative solution of nonlinear equations in several variables[M]. New York: Academic Press, 1970.

(责任编辑: 尹 闯)