求解对称非线性方程组的最速下降 –DY 混合共轭梯度法

The Steepest Descent-DY Hybrid Conjugate Gradient Method for Solving Symmetric Nonlinear Equations

沈冬梅

Dongmei Shen

南昌工学院 教育学院　中国·江西 南昌 330108

Nanchang Institute of Science and Technology, Nanchang, Jiangxi, 330108, China

摘　要： 针对对称非线性方程组的求解问题，论文提出了一种具有下降性质的最速下降 -DY 混合无导数型共轭梯度法，并在较弱的条件下，证明了算法的全局收敛性。最后，通过数值实验证明了该算法的有效性。

Abstract: This paper proposes a hybrid derivative-free conjugate gradient method that is a combination of the steepest method and DY method for solving the problem of symmetric nonlinear equations. Under weaker conditions, we prove the global convergence of the algorithm. Finally, the numerical experiments prove the effectiveness of the algorithm.

关键词： 对称非线性方程组；无导数算法；全局收敛性；混合共轭梯度法

Keywords: symmetric nonlinear system of equations; derivative-free algorithm; global convergence; hybrid conjugate gradient method

基金项目： 江西省教育厅科学技术研究项目：求解对称非线性方程组的几类修正共轭梯度法；项目编号：GJJ191110。

DOI: 10.12346/sde.v4i3.6060

1 引言

对称非线性方程组的求解已吸引了很多研究人员的兴趣，并得到广泛研究。文献 [1-2] 分别利用各类修正的共轭梯度法求解对称非线性方程组，并证明了算法的全局收敛性。Zhou 和 Shen[3] 提出了两类无修正的 PRP 型无导数共轭梯度法用于求解对称非线性方程组，并分别得到了算法的全局收敛性。Sabi’u 等 [4] 提出利用修正的 PRP 共轭梯度法来求解大规模的非线性对称方程组，并证明其全局收敛性。

论文主要研究如下对称非线性方程组的求解问题：

$$F(x)=0 \tag{1}$$

其中， $F: R^n \to R^n$ 是连续可微函数，它的 Jacobian 矩阵 $J(x)$ 是对称的，即 $J(x)=J(x)^T$ 。

张丽等 [5] 提出了一种最速下降 -DY 混合共轭梯度法，并用于求解无约束极小化问题。最速下降 -DY 混合共轭梯度法显著特点是在 Armijo 线性搜索下产生下降方向，即 $\nabla f(x_k)^T d_k<0$ ，其迭代方式如下：

$$x_{k+1}=x_k+\lambda_k d_k$$

$$d_k=\begin{cases}-\nabla f(x_k)+\beta_k^{DY} d_{k-1}, & \text{如果} d_{k-1}^T y_{k-1} \ge \varepsilon_1 \|d_{k-1}\|\|\nabla f(x_{k-1})\| \\ -\nabla f(x_k), & \text{否则}\end{cases} \tag{2}$$

其中，ε_1 是一个很小的正常数，$\nabla f(x_k)$ 为 $f(x)$ 在 x_k 处的梯度， $y_{k-1}=\nabla f(x_k)-\nabla f(x_{k-1})$ ， $\hat{a}_k^{DY}=\frac{\|\nabla f(x_k)\|^2}{d_{k-1}^T y_{k-1}}$ 。

受文献 [5] 启发，论文提出求解对称非线性方程组的最速下降 -DY 混合无导数型共轭梯度法。论文首先构造求解对称非线性方程组的最速下降 -DY 混合无导数型共轭梯度法，然后在较弱的条件下证明算法的全局收敛性，最后通过数值实验对所提算法进行测试。

2 算法

通常我们将式（1）的求解转化为如下无约束极小化问题：

$$min \quad f(x)=\frac{1}{2}\|F(x)\|^2, \quad x \in R^n \tag{3}$$

【作者简介】沈冬梅（1988-），女，中国湖北广水人，硕士，讲师，从事最优化理论与控制研究。

然而，由式（3）定义的函数 $f(x)$ 的梯度$\nabla f(x)=J(x)^T F(x)=J(x)F(x)$计算较复杂，不适合用于求解大规模问题，Li 和 Fukushima[6] 通过利用式（4）：

$$g_k=\frac{F(x_k+\lambda_{k-1}F_k)-F(x_k)}{\lambda_{k-1}} \tag{4}$$

替代精确梯度 $\nabla f(x_k)$，成功避免了复杂梯度的计算。显然，当 λ 充分小时：

$$g_k\approx\nabla f(x_k)=F'(x_k)F(x_k)=\lim_{\lambda\to 0}\frac{F(x+\lambda F(x))-F(x)}{\lambda} \tag{5}$$

论文将利用近似梯度函数 gk 的计算方式建立求解对称非线性方程组的最速下降 -DY 混合无导数型共轭梯度法。我们将通过以下方式得到最速下降 -DY 混合无导数型共轭梯度法的搜索方向和步长。

第一，最速下降 -DY 混合无导数型共轭梯度法搜索方向：

将式（4）运用到最速下降 -DY 混合无导数型共轭梯度法中，用近似梯度 $-gk_{-1}$、$-gk$ 代替精确梯度 $-\nabla f(x_{k-1})$、$-\nabla f(x_k)$，得到如下搜索方向：

$$d_k=\begin{cases}-g_k+\beta_k^{DY}d_{k-1}, & \text{如果} d_{k-1}^T y_{k-1}\ge\varepsilon_1\|d_{k-1}\|\|g_{k-1}\| \\ -g_k, & \text{否则}\end{cases}$$

其中，$\hat{a}_k^{DY}=\dfrac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}$，$y_{k-1}=g_k-g_{k-1}$。

由式（5）知，当 $\lambda>0$ 且充分小时，$g_k\approx\nabla f(x_k),\quad k\ge 1$，即 $\lim\limits_{\lambda\to 0}g_k(\lambda)=\nabla f(x_k)$，

$$\text{从而有}\lim_{\lambda\to 0}g_k^T(\lambda)d_k=\nabla f(x_k)^T d_k<0 \tag{6}$$

第二，最速下降 -DY 混合无导数型共轭梯度法的步长因子 [3]：

对于搜索方向$d_k(\lambda)$中的步长 λ，令 $\lambda=max\{\rho^j, j=0,1,2\cdots\}$ 满足下列不等式：

$$f(x_k+\lambda d_k)\le f(x_k)-\sigma_1\|\lambda F(x_k)\|^2-\sigma_2\|\lambda d_k(\lambda)\|^2+\eta_k f(x_k) \tag{7}$$

其中，$0<p<1$，$\sigma_1>0$ 和 $\sigma_2>0$，正数序列 $\{\eta_k\}$满足 $\sum\limits_{k=0}^{\infty}\eta_k\le\eta<\infty$。

下面建立求解对称非线性方程组的最速下降 -DY 混合无导数型共轭梯度算法。

算法 1：（最速下降 -DY 混合无导数型共轭梯度法）

步骤 1：选取初始点$x_0\in R^n$，参数 $0<p<1$，$\sigma_1>0$ 和 $\sigma_2>0$，$\varepsilon_1>0$，令 $k:=0$；

步骤 2：按照下式计算搜索方向 d_k：

$$d_k=\begin{cases}-g_k+\beta_k^{DY}d_{k-1}, & \text{如果} k\in K \\ -g_k, & \text{否则}\end{cases} \tag{8}$$

其中，$K=\{k\,|\,d_{k-1}^T y_{k-1}\ge\varepsilon_1\|d_{k-1}\|\|g_{k-1}\|\}$，$\beta_k^{DY}=\dfrac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}$，$y_{k-1}=g_k-g_{k-1}$；

步骤 3：由式（7）确定搜索步长 λ_k；

步骤 4：令 $x_{k+1}=x_k+\lambda_k d_k$；

步骤 5：令 $k:=k+1$，转步骤 2。

3 算法的全局收敛性

为证明算法的全局收敛性，我们给出如下假设条件。

假设 A：

①水平集 $\Omega=\left\{x\in R^n\,\middle|\,f(x)\le e^{\frac{\eta}{2}}f(x_0)\right\}$有界；

② $F(x)=0$ 为对称非线性方程组；

③在 Ω 的某个邻域 N 内，是连续可微的且其梯度是 Lipschitz 连续的，即存在常数 $L>0$ 使得，$\|F'(x)-F'(y)\|\le L\|x-y\|,\quad\forall x,y\in N$。

由假设 A 知，对 $\forall x\in N$ 有，存在 γ_1，$\gamma>0$，使得 $\|F(x)\|\ge\gamma_1$，$\|\nabla f(x)\|\le\gamma$。

由假设 A 知，对 $\forall x,y\in N$ 有，存在 L_1，$L_2>0$，使得 $\|\nabla f(x)-\nabla f(y)\|\le L_1\|x-y\|$，$\|F(x)-F(y)\|\le L_2\|x-y\|$。

引理 1[7]：设序列 $\{x_k\}$ 由最速下降 -DY 混合无导数型共轭梯度法产生，则 $\forall k\ge 0$，函数值序列 $\{\|F_k\|\}$ 收敛且序列 $\{x_k\}\subset\Omega$。

引理 2[7]：设序列 $\{x_k\}$ 由最速下降 -DY 混合无导数型共轭梯度法产生，则

$$\sum_{k=0}\|\lambda_k F(x_k)\|^2<\infty,\quad\sum_{k=0}\|\lambda_k d_k\|^2<\infty$$

引理 2 表明：$\|\lambda_k d_k\|$ 和$\|\lambda_k F(x_k)\|$ 有界。下面的引理表明搜索方向 d_k 有界。

引理 3：设 $\{x_k\}$ 由最速下降 -DY 混合无导数型共轭梯度法产生，若存在常量 $\varepsilon>0$，使得对 $\forall x\in\Omega$，有 $\|\nabla f(x)\|\ge\varepsilon$，则存在一个常数 $M>0$ 使得 $\|d_k\|\le M$，$\forall k\ge 0$。

证明：由 g_k 的定义易有 $\|g_k\|=\left\|\dfrac{F(x_k+\lambda_{k-1}F(x_k))-F(x_k)}{\lambda_{k-1}}\right\|\le L_2 F(x_k)\le\gamma_1 L_2=c$，

由 d_k 的定义易有，当 $k\in K, k\to\infty$ 时，

$$\|d_k\|\le\|g_k\|+|\beta_k^{DY}|\|d_{k-1}\|$$

$$\le\|g_k\|+\frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}\|d_{k-1}\|\le\|g_k\|+\frac{\|g_k\|^2}{\varepsilon_1\|d_{k-1}^T\|\|g_{k-1}\|}\|d_{k-1}\|$$

由引理的假设条件知，对 $j>0$，有$\|g_{j-1}\|\ge\varepsilon$，因而易得：

$$\|d_k\|\le c+\frac{c^2}{\gamma\varepsilon}=M$$

下面的定理表明最速下降 -DY 混合无导数型共轭梯度法具有全局收敛性。

定理 4：若假设 A 成立，由最速下降 -DY 混合无导数型共轭梯度法产生，则有：

$$\liminf_{k\to\infty}\|\nabla f(x_k)\|=0 \tag{9}$$

证明：如果 K 是有限集合，则由最速下降法的收敛性质，

易得结论成立。现设 K 是无限集合，运用反证法证明原结论成立。假设（9）式不成立，则存在一个正常数 τ 使得：

$$\|\nabla f(x_k)\| \geq \tau,\quad \forall k \geq 0 \qquad (10)$$

由 $\nabla f(x_k) = F'(x_k)^T F(x_k)$ 和式（10）知，存在常数 $\tau_1 > 0$，使得 $\|F(x_k)\| \geq \tau_1$。

第一，若 $\lim_{k\in K, k\to\infty} \sup \ddot{e}_k > 0$，由引理 2，我们易得：

$$\lim_{k\in K, k\to\infty} \inf \|F(x_k)\| = 0 \qquad (11)$$

由（11）式和引理 1 得 $\lim_{k\in K, k\to\infty} \|F(x_k)\| = 0$，这与 $\|F(x_k)\| \geq \tau_1$ 矛盾。

第二，若 $\lim_{k\in K, k\to\infty} \sup \lambda_k = 0$，由 $\lambda_k \geq 0$ 知，则存在无限集合 $\bar{K} \subset K$，使得：

$$\lim_{k\in \bar{K}, k\to\infty} \lambda_k = 0 \qquad (12)$$

由（12）式和算法 1 知 $\lambda'_k = \dfrac{\lambda_k}{\rho}$ 不满足（7）式，即：

$$\frac{f(x_k+\rho^{-1}\lambda_k d_k) - f(x_k)}{\lambda'_k} = \nabla f(x_k + t_k \lambda'_k d_k)^T d_k$$

$$> -\sigma_1 \|\lambda'_k F(x_k)\|^2 - \sigma_2 \|\lambda'_k d_k\|^2$$

这意味着：

$$\frac{f(x_k+\rho^{-1}\lambda_k d_k) - f(x_k)}{\rho^{-1}\lambda_k} \geq -\sigma_1\rho^{-1}\lambda_k f(x_k) - \sigma_2\rho^{-1}\lambda_k \|d_k\|^2 \qquad (13)$$

另一方面，由中值定理，存在一个常数 $t_k \in (0,1)$，使得：

$$f(x_k+\rho^{-1}\lambda_k d_k) - f(x_k) = \rho^{-1}\lambda_k \nabla f(x_k + t_k\rho^{-1}\lambda_k d_k)^T d_k$$

$$= \rho^{-1}\lambda_k \nabla f(x_k)^T d_k + \rho^{-1}\lambda_k (\nabla f(x_k + t_k\rho^{-1}\lambda_k d_k) - \nabla f(x_k))^T d_k$$

$$\leq \rho^{-1}\lambda_k \nabla f(x_k)^T d_k + L_1\rho^{-2}\lambda_k^2 \|d_k\|^2 \qquad (14)$$

由式（13）和式（14）有：

$$\nabla f(x_k)^T d_k + L_1\rho^{-2}\lambda_k^2 \|d_k\|^2 \geq -\sigma_1\rho^{-1}\lambda_k f(x_k) - \sigma_2\rho^{-1}\lambda_k \|d_k\|^2$$

由 $\|d_k\|$、$\|F_k\|$ 有界及式（6）知，对充分大的 $k \in \bar{K}$，有：

$$\lambda_k > \frac{-\nabla f(x_k)^T d_k}{(L_1+\sigma_2)\rho^{-1}\|d_k\|^2 + \sigma_2\rho^{-1}\|F_k\|^2} > 0$$

这与 $\lim_{k\in \bar{K}, k\to\infty} \ddot{e}_k = 0$ 矛盾，故假设不成立，即证。

论文用 Matlab 实现最速下降 -DY 混合无导数型共轭梯度法的编程。数值实验表明，该算法具有占用内存少、消耗时间短的优势，同时也反映出最速下降 -DY 混合无导数型共轭梯度法数值表现与近似 PRP 算法 1[7] 很类似，是求解大规模对称非线性方程组的有效算法。

算例 1：$F(x) = Ax + f(x)$，其中 $f(x) = (e^{x_1}-1, \cdots, e^{x_n}-1)^T$，矩阵 A 由文献 [7] 给出。

算例 2：$F_1(x) = x_1(x_1^2 + x_2^2) - 1$，$F_i(x) = x_i(x_{i-1}^2 + 2x_i^2 + x_{i+1}^2) - 1, i = 2,3,\cdots,n-1$，$F_n(x) = x_n(x_{n-1}^2 + x_n^2)$。

有关参数选取及初始值如下：$\sigma_1 = \sigma_2 = 10^{-4}$，$\lambda_0 = 0.01$，$\rho = 0.2$，$\varepsilon_1 = 0.01$；$\eta_k = \dfrac{1}{(k+1)^2}$，初始值 $x_0 = (0.1, 0.1, \cdots 0.1)$，算法终止条件为 $\|F(x_k)\| \leq 10^{-3}$。Dim：测试问题的维数，Iter：算法迭代次数，Time：迭代所需时间，$\|F_k\|$：$\|F(x_k)\|$ 的最终值，算法的数值结果见表 1。

表 1 算法的数值结果

		最速下降 -DY 混合无导数型共轭梯度法			近似 PRP 算法 1		
算例	Dim	Iter	Time	$\|F_k\|$	Iter	Time	$\|F_k\|$
算例 1	10	37	0.0350	9.0636e-04	27	0.0379	9.9566e-04
	50	47	0.0708	9.7129e-04	26	0.0404	8.5321e-04
	100	50	0.0944	9.7589e-04	28	0.0622	5.2109e-04
	500	48	0.7836	9.8409e-04	36	0.8107	8.9261e-04
	1000	51	3.2648	8.8978e-04	31	1.8118	7.9903e-04
	2000	53	9.7286	9.3731e-04	33	6.4984	7.9339e-04
算例 2	50	114	0.0164	9.9710e-04	45	0.0147	9.5281e-04
	100	117	0.0168	9.9692e-04	43	0.0141	8.9002e-04
	500	118	0.0266	9.8169e-04	43	0.0155	8.5609e-04
	1000	118	0.0294	9.8173e-04	46	0.0199	8.8368e-04
	2000	118	0.0393	9.8187e-04	52	0.0255	9.1798e-04
	5000	119	0.0972	9.3286e-04	54	0.0845	8.9205e-04

4 结语

论文通过利用近似梯度代替精确梯度，构造了求解对称非线性方程组的最速下降 -DY 混合无导数型共轭梯度法，并在较弱的条件下证明了算法的全局收敛性。理论分析和数值试验均表明，最速下降 -DY 混合无导数型共轭梯度法对求解对称非线性方程组是有效的。

参考文献

[1] Li D H, Wang X L. A modefied Fletcher-Reeves-type derivative-free method for symmetric nonlinear equations[J]. Numer. Alge. Cont. Optim,2011(1):71-82.

[2] 李灿.求解对称非线性方程组的MPRP型Derivative-free算法[J].西南大学学报:自然科学版,2014,36(1):67-71.

[3] Zhou W J, Shen D M. An inexact PRP conjugate gradient method for symmetric nonlinear equations[J]. Numerical Functional Analysis and Optimization,2014(35):370-388.

[4] SABI'U J, MUANGCHOO K, SHAH A, et al. A Modified PRP-CG Type Derivative-Free Algorithm with Optimal Choices for Solving Large-Scale Nonlinear Symmetric Equations[J]. Symmetry,2021(13):234.

[5] Zhang L, Zhou W J, Li D H. Global convergence of the DY conjugate gradient method with Armijo line search for unconstrained optimization problems[J]. Optimization Methods and Software,2007(22):511-517+3.

[6] Li D H, Fukushima M. A globally and superlinearly convergent Gauss-Newton-based BFGS method for symmetric nonlinear equations[J]. SIAM Journal on Numerical Analysis,1999,37(1):152-172.

[7] 沈冬梅.求解对称非线性方程组PRP型算法研究[D].长沙:长沙理工大学,2014.