文章编号:1000-5471(2012)04-0053-06

充分下降的共轭梯度法簇及其全局收敛性①

黄 海, 林穗华, 潘义前

广西民族师范学院 数学与计算机科学系,广西 崇左 532200

摘要:在双参数共轭梯度法的基础上,给出一类具有充分下降性的共轭梯度法簇,证明了相应的方法在非单调线搜索及弱 Wolfe 线搜索下对非凸目标函数全局收敛,并用数值实验表明该方法具有良好的数值结果.

关 键 词:共轭梯度法;充分下降;全局收敛

中图分类号:O224 **文献标志码**:A

共轭梯度法是求解无约束优化问题 $\min f(\boldsymbol{x})$ 的一类常用方法,尤其对于维数很大的目标函数 f(f: $\mathbf{R}^n \longrightarrow \mathbf{R}$ 为连续可微的非线性函数)的问题求解更具有优势. 共轭梯度法迭代形式为:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t_k \boldsymbol{d}_k \tag{1}$$

其中 $\boldsymbol{x}_k$ 为当前迭代点,$t_k > 0$ 为步长因子,$\boldsymbol{d}_k$ 为按形式

$$\boldsymbol{d}_k = \begin{cases} -\boldsymbol{g}_k & k = 1 \\ -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1} & k \geqslant 2 \end{cases} \tag{2}$$

产生的搜索方向,其中 $\boldsymbol{g}_k$ 表示 $f(\boldsymbol{x})$ 在点 $\boldsymbol{x}_k$ 的梯度 $\nabla f(\boldsymbol{x}_k)$,不同的标量参数 $\beta_k \in \mathbf{R}$ 决定不同的共轭梯度法.

著名的共轭梯度法有 PRP,HS,FR,DY,CD,LS 等方法[1],文献[2-3]在这些方法的基础上提出了新的 β_k 公式:

$$\beta_k^{\mathrm{WYL}} = \frac{\boldsymbol{g}_k^{\mathrm{T}} \bar{\bar{\boldsymbol{y}}}_{k-1}}{\boldsymbol{g}_{k-1}^{\mathrm{T}} \boldsymbol{g}_{k-1}} \qquad \beta_k^{\mathrm{MHS}} = \frac{\boldsymbol{g}_k^{\mathrm{T}} \bar{\bar{\boldsymbol{y}}}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}} \qquad \beta_k^{\mathrm{MLS}} = \frac{\boldsymbol{g}_k^{\mathrm{T}} \bar{\bar{\boldsymbol{y}}}_{k-1}}{-\boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{g}_{k-1}} \tag{3}$$

其中 $\boldsymbol{y}_{k-1} = \boldsymbol{g}_k - \boldsymbol{g}_{k-1}$,$\bar{\bar{\boldsymbol{y}}}_{k-1} = \boldsymbol{g}_k - \dfrac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}\boldsymbol{g}_{k-1}$. 文献[4]在公式(3)的基础上讨论了带两个参数的共轭梯度法簇

$$\beta_k(\mu_1, \mu_2) = \frac{\boldsymbol{g}_k^{\mathrm{T}} \bar{\bar{\boldsymbol{y}}}_{k-1}}{-\mu_1 \boldsymbol{g}_{k-1}^{\mathrm{T}} \boldsymbol{d}_{k-1} + \mu_2 \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1} + (1-\mu_1-\mu_2)\|\boldsymbol{g}_{k-1}\|^2} \tag{4}$$

方法(4)及其扩展为 DL 型的共轭梯度法[5],在强 Wolfe 线搜索及适当的条件下,对一般的目标函数具有全局收敛性. 文献[6-7]提出了新的修正 PRP 公式

$$\beta_k^{\mathrm{DPRP}} = \beta_k^{\mathrm{PRP}} - \frac{c\|\boldsymbol{y}_{k-1}\|^2}{\|\boldsymbol{g}_{k-1}\|^4}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1} \qquad \beta_k^{\mathrm{MPRP}} = \beta_k^{\mathrm{PRP}} - \min\left\{\beta_k^{\mathrm{PRP}}, \frac{c\|\boldsymbol{y}_{k-1}\|^2}{\|\boldsymbol{g}_{k-1}\|^4}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}\right\} \tag{5}$$

其中 $c > 4^{-1}$ 为常数. 文献[7]还将这一修正思想应用于对 FR,CD,LS,DY,HS 公式的修正,所得的方法均具有充分下降性,在步长 $t_k \geqslant t^* > 0$ 的假设下得到了修正方法的全局收敛性.

① 收稿日期:2011-03-02
基金项目:广西壮族自治区教育厅科研项目(201012MS215);广西民族师范学院科研项目(200909).
作者简介:黄 海(1969-),男,广西武鸣人,副教授,主要从事最优化理论与方法的研究.

共轭梯度法第 k 次迭代中的步长因子 t_k 通常由线搜索产生，非单调线搜索不要求每次迭代目标函数都单调下降，放宽了对线搜索的要求，因此非单调线搜索可能比单调线搜索有效. 文献[8]提出了新型的非单调线搜索：

$$\begin{cases} f(\boldsymbol{x}_k + t_k\boldsymbol{d}_k) \leqslant C_k + \delta t_k \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k \\ g(\boldsymbol{x}_k + t_k\boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k \geqslant \sigma \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k \end{cases} \tag{6}$$

其中 $\delta \in (0, 2^{-1})$，$\sigma \in (\delta, 1)$，且

$$C_k = \frac{\eta_{k-1}Q_{k-1}C_{k-1} + f_k}{Q_k} \qquad Q_k = \eta_{k-1}Q_{k-1} + 1 \tag{7}$$

$$\eta_k \in [\eta_{\min}, \eta_{\max}] \qquad 0 \leqslant \eta_{\min} \leqslant \eta_{\max} \leqslant 1 \qquad C_1 = f_1 \qquad Q_1 = 1$$

显然 C_k 是 C_{k-1} 与 f_k 的凸组合，从而 C_k 是目标函数值 $f_1, f_2, \cdots, f_k$ 的凸组合. 若取 $\eta_k \equiv 0$，则 $C_k \equiv f_k$，非单调线搜索(6)退化为单调的弱 Wolfe 线搜索：

$$\begin{cases} f(\boldsymbol{x}_k + t_k\boldsymbol{d}_k) \leqslant f_k + \delta t_k \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k \\ g(\boldsymbol{x}_k + t_k\boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k \geqslant \sigma \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k \end{cases} \tag{8}$$

充分下降性条件为：存在 $c \in (0, 1)$，使对任意 $k \geqslant 1$，有

$$\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k \leqslant -c \| \boldsymbol{g}_k \|^2 \tag{9}$$

1 算法及其充分下降性

受文献[6-7,9-10]的启发，我们结合方法(4)和(5)提出新的共轭梯度法簇：

$$\beta_k(\mu_1, \mu_2) = \frac{\boldsymbol{g}_k^{\mathrm{T}} \bar{\bar{\boldsymbol{y}}}_{k-1}}{R_{k-1}} - \min\left\{\frac{\boldsymbol{g}_k^{\mathrm{T}} \bar{\bar{\boldsymbol{y}}}_{k-1}}{R_{k-1}}, \frac{\mu_2 \| \bar{\boldsymbol{y}}_{k-1} \|^2}{R_{k-1}^2} \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}\right\} \tag{10}$$

其中 $\mu_1 \in [0, 1]$，$\mu_2 > 4^{-1}$，且

$$R_{k-1} = -\mu_1 \boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1} + (1-\mu_1)\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1} \tag{11}$$

显然

$$\beta_k(0, \mu_2) = \beta_k^{\mathrm{MHS}} - \min\left\{\beta_k^{\mathrm{MHS}}, \frac{\mu_2 \| \bar{\boldsymbol{y}}_{k-1} \|^2}{(\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1})^2} \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}\right\}$$

$$\beta_k(1, \mu_2) = \beta_k^{\mathrm{MLS}} - \min\left\{\beta_k^{\mathrm{MLS}}, \frac{\mu_2 \| \bar{\boldsymbol{y}}_{k-1} \|^2}{(\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{g}_{k-1})^2} \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}\right\}$$

下面给出方法(10)的采用非单调线搜索技术解无约束优化问题的新算法(DNCG 算法)：

Step 0 给定初值 $\boldsymbol{x}_1 \in \mathbf{R}^n$，$\mu_1 \in [0, 1]$，$\mu_2 > 4^{-1}$，$1 > \varepsilon \geqslant 0$，$\delta \in (0, 2^{-1})$，$\sigma \in (\delta, 1)$，$0 \leqslant \eta_{\min} \leqslant \eta_{\max} < 1$，$\boldsymbol{d}_1 = -\boldsymbol{g}_1 = -\nabla f(\boldsymbol{x}_1)$，$C_1 = f(\boldsymbol{x}_1)$，$Q_1 = 1$，$k = 1$，若 $\| \boldsymbol{g}_1 \| \leqslant \varepsilon$，停止；

Step 1 计算满足非单调线搜索(6)的步长因子 t_k；

Step 2 由迭代(1)计算下一迭代点 $\boldsymbol{x}_{k+1}$，$\boldsymbol{g}_{k+1} = g(\boldsymbol{x}_{k+1})$，若 $\| \boldsymbol{g}_{k+1} \| \leqslant \varepsilon$，停止；

Step 3 由方法(10)计算参数 β_k，由(2)式计算下一搜索方向 $\boldsymbol{d}_{k+1}$；

Step 4 $k := k+1$，转 Step 1.

定理 1 设 $\{\boldsymbol{g}_k, \boldsymbol{d}_k\}$ 为由 DNCG 算法生成的无穷序列，则 $\{\boldsymbol{g}_k, \boldsymbol{d}_k\}$ 满足充分下降性条件(9)，且对任意 $k > 1$，有 $R_{k-1} > 0$.

证 令 $c = 1 - \dfrac{1}{4\mu_2}$，因 $\mu_2 > \dfrac{1}{4}$，则 $c \in (0, 1)$. 由 $\boldsymbol{d}_1 = -\boldsymbol{g}_1$，知 $\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{d}_1 = -\| \boldsymbol{g}_1 \|^2 \leqslant -c \| \boldsymbol{g}_1 \|^2 < 0$. 假设 $\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1} \leqslant -c \| \boldsymbol{g}_{k-1} \|^2 < 0$，则由公式(6),(11)，可得

$$R_{k-1} = -\mu_1 \boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1} + (1-\mu_1)\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1} \geqslant -[1+\sigma(1-\mu_1)]\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1} \geqslant \tag{12}$$

$$c[1+\sigma(1-\mu_1)] \| \boldsymbol{g}_{k-1} \|^2 > 0 \tag{13}$$

设 $\boldsymbol{u} = \dfrac{\sqrt{R_{k-1}}}{\sqrt{2\mu_2}}\boldsymbol{g}_k$，$\boldsymbol{v} = \dfrac{\sqrt{2\mu_2}\,\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\sqrt{R_{k-1}}}\bar{\boldsymbol{y}}_{k-1}$，分两种情况讨论：

若 $\dfrac{\boldsymbol{g}_k^{\mathrm{T}}\bar{\boldsymbol{y}}_{k-1}}{R_{k-1}} < \dfrac{\mu_2\|\bar{\boldsymbol{y}}_{k-1}\|^2}{R_{k-1}^2}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}$，由公式(2)和(10)，可得 $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k = -\|\boldsymbol{g}_k\|^2 \leqslant -c\|\boldsymbol{g}_k\|^2$；

若 $\dfrac{\boldsymbol{g}_k^{\mathrm{T}}\bar{\boldsymbol{y}}_{k-1}}{R_{k-1}} \geqslant \dfrac{\mu_2\|\bar{\boldsymbol{y}}_{k-1}\|^2}{R_{k-1}^2}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}$，由公式(2)和(10)及不等式 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} \leqslant \dfrac{1}{2}(\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2)$，可得

$$\begin{aligned}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k &= -\|\boldsymbol{g}_k\|^2 + \left(\frac{\boldsymbol{g}_k^{\mathrm{T}}\bar{\boldsymbol{y}}_{k-1}}{R_{k-1}} - \frac{\mu_2\|\bar{\boldsymbol{y}}_{k-1}\|^2}{R_{k-1}^2}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}\right)\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1} = \\ &\frac{1}{R_{k-1}}\left[(\boldsymbol{g}_k^{\mathrm{T}}\bar{\boldsymbol{y}}_{k-1})(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}) - \|\boldsymbol{g}_k\|^2R_{k-1} - \frac{\mu_2\|\bar{\boldsymbol{y}}_{k-1}\|^2}{R_{k-1}}(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1})^2\right] = \\ &\frac{1}{R_{k-1}}\left[\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} - \frac{1}{2}(\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2) - \left(1-\frac{1}{4\mu_2}\right)\|\boldsymbol{g}_k\|^2R_{k-1}\right] \leqslant \\ &-\left(1-\frac{1}{4\mu_2}\right)\|\boldsymbol{g}_k\|^2 = -c\|\boldsymbol{g}_k\|^2\end{aligned}$$

综上所述，由数学归纳法，定理 1 得证.

2　算法的收敛性

假设 A (i)　设水平集 $\Omega = \{\boldsymbol{x} \in \mathbf{R}^n \mid f(\boldsymbol{x}) \leqslant f(\boldsymbol{x}_1)\}$ 有界，且 $f(\boldsymbol{x})$ 在 Ω 中有下界；

(ii)　设 $f(\boldsymbol{x})$ 的梯度 $g(\boldsymbol{x})$ 在 Ω 上 Lipschitz 连续，即存在常数 $L>0$，使对任意 $\boldsymbol{x},\boldsymbol{y} \in \Omega$，有

$$\|g(\boldsymbol{y}) - g(\boldsymbol{x})\| \leqslant L\|\boldsymbol{y}-\boldsymbol{x}\|$$

由假设 A 知，存在 $B>0$，$\bar{\gamma}>0$，使对任意 $\boldsymbol{x} \in \Omega$，有 $\|\boldsymbol{x}\| \leqslant B$，$\|g(\boldsymbol{x})\| \leqslant \bar{\gamma}$.

引理 1　若假设 A 成立，$\{\boldsymbol{g}_k, \boldsymbol{d}_k\}$ 为由 DNCG 算法生成的序列，则

$$\sum_{k=1}^{\infty}\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2} < +\infty \tag{14}$$

证　由定理 1 知，对任意 $k \geqslant 1$，有 $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k < 0$. 由公式(6),(13)，可得

$$-(1-\sigma)\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k \leqslant (\boldsymbol{g}_{k+1}-\boldsymbol{g}_k)^{\mathrm{T}}\boldsymbol{d}_k \leqslant Lt_k\|\boldsymbol{d}_k\|^2$$

从而可得

$$t_k \geqslant \frac{1-\sigma}{L}\frac{(-\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)}{\|\boldsymbol{d}_k\|^2} \tag{15}$$

由非单调线搜索(6)中 Q_k 的定义(7)，再利用 $Q_1=1$ 及 $\eta_{\max}<1$，可得

$$Q_k = \sum_{i=1}^{k-1}\left(\prod_{j=1}^{i}\eta_{k-i}\right)+1 \leqslant \sum_{i=1}^{k-1}(\eta_{\max})^i+1 = \sum_{i=0}^{k-1}(\eta_{\max})^i < \sum_{i=0}^{\infty}(\eta_{\max})^i = \frac{1}{1-\eta_{\max}}$$

再由公式(6),(7),(15)可得

$$\begin{aligned}C_k &= \frac{\eta_{k-1}Q_{k-1}C_{k-1}+f(\boldsymbol{x}_{k-1}+t_{k-1}\boldsymbol{d}_{k-1})}{Q_k} \leqslant \\ &\frac{\eta_{k-1}Q_{k-1}C_{k-1}+C_{k-1}-\dfrac{\delta(1-\sigma)}{L}\dfrac{(\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1})^2}{\|\boldsymbol{d}_{k-1}\|^2}}{Q_k} \leqslant \\ &C_{k-1} - \frac{\delta(1-\sigma)(1-\eta_{\max})}{L}\frac{(\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1})^2}{\|\boldsymbol{d}_{k-1}\|^2}\end{aligned} \tag{16}$$

由公式(16)知 $C_k \leqslant C_{k-1}$，再由公式(7)得到

$$C_k = \frac{\eta_{k-1}Q_{k-1}C_{k-1}+f_k}{Q_k} \geqslant \frac{\eta_{k-1}Q_{k-1}C_k+f_k}{Q_k}$$

从而 $C_k(Q_k-\eta_{k-1}Q_{k-1}) \geqslant f_k$，即 $C_k \geqslant f_k$. 再依假设 A(i)，f_k 有下界，可知 $\{C_k\}$ 单调递减且有下界，即存在 $\lim\limits_{k\to\infty} C_k$. 对不等式(16)两边分别求和得到

$$\frac{\delta(1-\sigma)(1-\eta_{max})}{L}\sum_{k=1}^{\infty}\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2} \leqslant C_1 - \lim_{k\to\infty}C_k < +\infty$$

即(14)式成立,引理 1 得证.

引理 2 若假设 A 及

$$0 < \gamma \leqslant \|\boldsymbol{g}_k\| \leqslant \bar{\gamma} \tag{17}$$

成立,$\{\boldsymbol{g}_k, \boldsymbol{d}_k\}$ 为由 DNCG 算法生成的序列,则 β_k 满足文献[1]的性质(*).

证 易计算得 $\|\bar{\boldsymbol{y}}_{k-1}\|^2 = 2\boldsymbol{g}_k^{\mathrm{T}}\bar{\boldsymbol{y}}_{k-1}$ 且

$$0 \leqslant \boldsymbol{g}_k^{\mathrm{T}}\bar{\boldsymbol{y}}_{k-1} = \|\boldsymbol{g}_k\|^2\left(1 - \frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_{k-1}}{\|\boldsymbol{g}_k\|\,\|\boldsymbol{g}_{k-1}\|}\right) \leqslant 2\|\boldsymbol{g}_k\|^2$$

若 $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1} \geqslant 0$,则由方法(10)知 $0 \leqslant \beta_k \leqslant \dfrac{\boldsymbol{g}_k^{\mathrm{T}}\bar{\bar{\boldsymbol{y}}}_{k-1}}{R_{k-1}}$;若 $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1} < 0$,由方法(10)、公式(6)和(12),可得

$$0 \leqslant \beta_k = \frac{\boldsymbol{g}_k^{\mathrm{T}}\bar{\bar{\boldsymbol{y}}}_{k-1}}{R_{k-1}} - \frac{\mu_2\|\bar{\bar{\boldsymbol{y}}}_{k-1}\|^2}{R_{k-1}^2}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1} \leqslant$$

$$\left(1 - \frac{2\sigma\mu_2}{R_{k-1}}\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}\right)\frac{\boldsymbol{g}_k^{\mathrm{T}}\bar{\bar{\boldsymbol{y}}}_{k-1}}{R_{k-1}} \leqslant \left[1 + \frac{2\sigma\mu_2}{1+\sigma(1-\mu_1)}\right]\frac{\boldsymbol{g}_k^{\mathrm{T}}\bar{\bar{\boldsymbol{y}}}_{k-1}}{R_{k-1}}$$

类似文献[4]的引理 3,易证引理 2 成立.

引理 3 设假设 A 成立,$\{\boldsymbol{g}_k, \boldsymbol{d}_k\}$ 为由 DNCG 算法生成的序列. 如果存在常数 $\gamma > 0$,使对任意 $k \geqslant 1$,有 $\|\boldsymbol{g}_k\| \geqslant \gamma$,则 $\sum\limits_{k\geqslant 2}\|\boldsymbol{u}_k - \boldsymbol{u}_{k-1}\|^2 < \infty$,其中 $\boldsymbol{u}_k = \dfrac{\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|}$.

证 由定理 1 及 $\|\boldsymbol{g}_k\| \geqslant \gamma > 0$ 知,对任意 $k \geqslant 1$,有 $\boldsymbol{d}_k \neq \boldsymbol{0}$. 记

$$\delta_k = \beta_k \frac{\|\boldsymbol{d}_{k-1}\|}{\boldsymbol{d}_k} \qquad \boldsymbol{r}_k = -\frac{\boldsymbol{g}_k}{\boldsymbol{d}_k}$$

则由方法(2)和(10)可得 $\boldsymbol{u}_k = \boldsymbol{r}_k + \delta_k\boldsymbol{u}_{k-1}$. 利用 $\|\boldsymbol{u}_k\| = \|\boldsymbol{u}_{k-1}\| = 1$,可得

$$\|\boldsymbol{r}_k\| = \|\boldsymbol{u}_k - \delta_k\boldsymbol{u}_{k-1}\| = \|\delta_k\boldsymbol{u}_k - \boldsymbol{u}_{k-1}\|$$

因 $\beta_k \geqslant 0$,故 $\delta_k \geqslant 0$,从而可得

$$\|\boldsymbol{u}_k - \boldsymbol{u}_{k-1}\| \leqslant \|(1+\delta_k)(\boldsymbol{u}_k - \boldsymbol{u}_{k-1})\| \leqslant \|\boldsymbol{u}_k - \delta_k\boldsymbol{u}_{k-1}\| + \|\delta_k\boldsymbol{u}_k - \boldsymbol{u}_{k-1}\| = 2\|\boldsymbol{r}_k\| \tag{18}$$

由定理 1、引理 1 及 $\|\boldsymbol{g}_k\| \geqslant \gamma > 0$,可得

$$\sum_{k=1}^{\infty}\|\boldsymbol{r}_k\|^2 \leqslant \gamma^{-2}\sum_{k=1}^{\infty}\|\boldsymbol{r}_k\|^2\|\boldsymbol{g}_k\|^2 = \gamma^{-2}\sum_{k=1}^{\infty}\frac{\|\boldsymbol{g}_k\|^4}{\|\boldsymbol{d}_k\|^2} \leqslant \gamma^{-2}c^{-1}\sum_{k=1}^{\infty}\frac{(\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2} < +\infty \tag{19}$$

由(18),(19)式,可得 $\sum\limits_{k\geqslant 2}\|\boldsymbol{u}_k - \boldsymbol{u}_{k-1}\|^2 < \infty$. 引理 3 得证.

设

$$\kappa_{k,\Delta}^{\lambda} = \{i \in \mathbf{Z}_+ \mid k \leqslant i \leqslant k + \Delta - 1,\ \|\boldsymbol{s}_{i-1}\| > \lambda\}$$

$|\kappa_{k,\Delta}^{\lambda}|$ 表示 $\kappa_{k,\Delta}^{\lambda}$ 的元素个数. 类似文献[1]的引理 3.2.2,利用引理 2 可得到:

引理 4 设假设 A 成立,$\{\boldsymbol{g}_k, \boldsymbol{s}_k\}$ 为由 DNCG 算法生成的序列. 若(17)式成立,则存在 $\lambda > 0$,使得对任意 $\Delta \in \mathbf{Z}_+$ 和指标 k_0,均存在指标 $k \geqslant k_0$,满足 $|\kappa_{k,\Delta}^{\lambda}| > \dfrac{\Delta}{2}$.

根据假设 A、定理 1、引理 1-4,类似文献[1]的定理 3.3.3,可得到如下结论:

定理 2 若假设 A 成立,$\{g_k\}$ 为由 DNCG 算法生成的序列,则 $\lim\limits_{k\to\infty}\inf\|\boldsymbol{g}_k\| = 0$.

注意到,若 DNCG 算法中取参数 $\eta_k \equiv 0$,则非单调线搜索(6)为弱 Wolfe 线搜索(7),由定理 2 知共轭梯度法簇(10)采用弱 Wolfe 线搜索时也具有全局收敛性.

3 数值实验

我们对 DNCG 算法进行数值试验. 实验平台为 Windows XP + Matlab 7.3,算法程序参数为 $\delta = 0.01$,$\sigma = 0.1$,$\varepsilon = 10^{-5}$,终止条件为 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$ 或 $\|\boldsymbol{g}_k\| \leqslant \varepsilon(1 + |f_k|)$. 测试的算法表示如下:

PRP+:采用弱 Wolfe 线搜索的 PRP+方法(即 DNCG 算法中 $\beta_k = \max\{\beta_k^{\mathrm{PRP}}, 0\}$,$\eta_k \equiv 0$);

DNCG1：DNCG 算法，参数 $\mu_1=1$，$\mu_2=2$，$\eta_k\equiv 0$；

DNCG2：DNCG 算法，参数 $\mu_1=0.5$，$\mu_2=0.5$，$\eta_k\equiv 0.1$.

记 Problem 为测试问题的名称，Dim 为对应的维数，测试目标函数集[11] 见表 1.

表 1　测试目标函数集

(Proplem：Dim)
(Rose：2)；(Froth：2)；(Badscp：2)；(Badscb：2)；(Beale：2)；(Jensam：2)；(Helix：3)；(Bard：3)；
(Gauss：3)；(Meyer：3)；(Gulf：3)；(Box：3)；(Sing：4)；(Wood：4)；(Kowosb：4)；(Bd：4)；
(Osb1：5)；(Biggs：6)；(Osb2：11)；(Watson：20)；(Rosex：8，50，100)；(Singx：8)；(Pen1：2)；
(Pen2：4，50)；(Vardim：2，50)；(Trig：3，50，100)；(Bv：3，10)；(Ie：3，50，100，200，500)；
(Trid：3，50，100，200)；(Band：3，50，100，200)；(Lin：2，50，500，1000)；(Lin1：2，10)

采用文献[12]的方法，我们对算法的数值结果分别比较目标函数及其梯度的计算次数. 相对于这些指标的算法性能关系见图 1 和图 2：

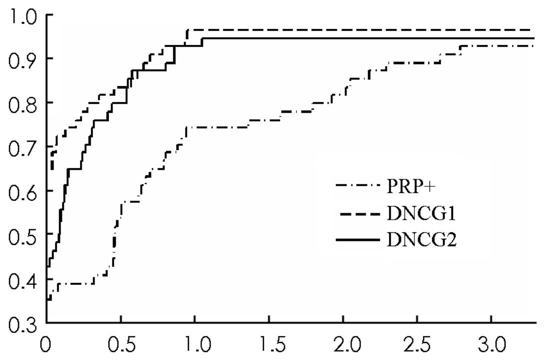

图 1　基于目标函数计算次数的性能比较

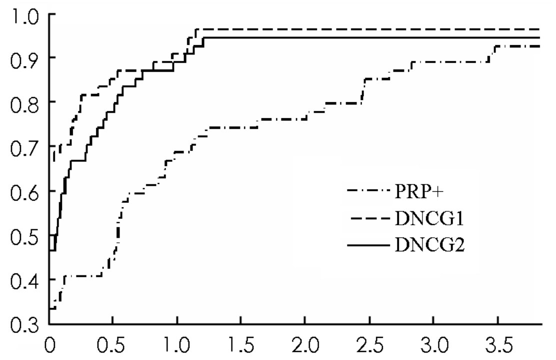

图 2　基于目标梯度计算次数的性能比较

从图 1、图 2 可看出，DNCD1，DNCD2，PRP＋算法对 54 个目标函数测试集的成功求解率分别约为 97%，94%，92%；目标函数及其梯度计算次数最少的是 DNCD1 算法，其次是 DNCD2 算法，计算量最大的是 PRP＋算法.

4　结束语

本文基于文献[5]的双参数共轭梯度法和文献[7]的算法下降性的修正思想，提出了具充分下降性和在非单调线搜索及弱 Wolfe 线搜索下全局收敛的共轭梯度法簇，数值实验表明该算法的计算性能优于 PRP＋方法.

参考文献：

[1]　戴彧虹，袁亚湘. 非线性共轭梯度法 [M]. 上海：上海科学技术出版社，2001：14－77.

[2]　WEI Zeng-xin，YAO Sheng-wei，LIU Li-ying. The Convergence Properties of Some New Conjugate Gradient Methods [J]. Applied Mathematics and Computation，2006，183(2)：1341－1350.

[3]　YAO Sheng-wei，WEI Zeng-xin，HUANG Hai. A Notes about WYL's Conjugate Gradient Method and Its Applications [J]. Applied Mathematics and Computation，2007，191(2)：381－388.

[4]　林穗华，黄　海. 一个双参数的共轭梯度法簇 [J]. 西南师范大学学报：自然科学版，2007，32(6)：43－47.

[5]　黄　海，林穗华. 改进的多参数非线性共轭梯度法的全局收敛性 [J]. 西南师范大学学报：自然科学版，2010，35(2)：76－80.

[6]　喻高航，关履泰. 具有充分下降性的修正 PRP 算法及其收敛性 [J]. 中山大学学报：自然科学版，2006，45(4)：11－14.

[7]　YUAN Gong-lin. Modified Nonlinear Conjugate Gradient Methods with Sufficient Descent Property for Large-scale Opti-

mination Problems [J]. Optimization Letters, 2009(3): 11-21.
[8] ZHANG Hong-chao, HAGER W W. A Nonmonotone Line Search Technique and Its Application to Unconstrained Optimization [J]. SIAM Journal on Optimization, 2004, 14: 1043-1056.
[9] 莫利柳. 一类新型的杂交共轭梯度法 [J]. 西南大学学报: 自然科学版, 2010, 32(9): 1-5.
[10] 黎 勇. 一类修正PRP共轭梯度法的全局收敛性及其数值试验结果 [J]. 西南大学学报: 自然科学版, 2011, 33(11): 23-28.
[11] MORE J J, GARBOW B S, HILLSTROME K E. Testing Unconstrained Optimization Software [J]. ACM Trans Math Software, 1981(7): 17-41.
[12] DOLAN E D, MORE J J. Benchmarking Optimization Software with Performance Profiles [J]. Math Program, 2002, 91: 201-213.

A Conjugate Gradient Method Family with Sufficient Descent Condition and Its Global Convergence

HUANG Hai, LIN Sui-hua, PAN Yi-qian

Department of Mathematics and Computer Science, Guangxi Normal University for Nationalities, Chongzuo Guangxi 532200, China

Abstract: Based on the updated formula with two paramerters, a new class of conjugate gradient methods which possess the sufficient descent property were proposed in this paper. The authors prove that the algorithm with nonmonotone line search and weak Wolfe line search for nonconvex function possess global convergence property. Preliminary numerical results show that those methods are effective.

Key words: conjugate gradient method; sufficient descent; global convergence

责任编辑 廖 坤