

强 Wolfe 线搜索下一个新的优化算法

朱铁锋

(内蒙古财经大学统计与数学学院, 内蒙古 呼和浩特 010070)

摘要: 本文主要研究了一个新的优化算法. 首先, 利用给出的新的公式和强 Wolfe 线搜索, 证明了该算法在不要求搜索方向满足共轭性条件下具有充分下降性和全局收敛性; 其次, 利用目标函数为一致凸函数的假设, 证明了该算法具有线性收敛速率; 最后, 利用数值试验, 验证了新算法是有效的、可行的.

关键词: 无约束优化; 共轭梯度法; 强 Wolfe 线搜索; 全局收敛性

MR(2010) 主题分类号: 65K10; 90C30 中图分类号: O224

文献标识码: A 文章编号: 0255-7797(2023)03-0267-10

1 引言

优化技术在众多领域有着广泛的应用, 受到高度重视^[1-3]. 无约束最优化问题的一般形式为

$$\min f(x), x \in R^n, \quad (1.1)$$

$f: R^n \rightarrow R$ 是连续可微函数. 共轭梯度法因其存储量小, 仅需一阶导数信息, 不需大规模矩阵求逆运算等优点, 被认为是求解 (1.1) 的一类非常有效的方法^[4-6]. 共轭梯度法求解 (1.1) 的一般迭代公式为:

$$x_{k+1} = x_k + \alpha_k d_k, \quad (1.2)$$

其中 α_k 为步长因子, 可通过精确线搜索或非精确线搜索求得. d_k 为 $f(x)$ 在 x_k 点的下降方向, 通常定义为

$$d_k = \begin{cases} -g_1 & k = 1 \\ -g_k + \beta_k d_{k-1} & k \geq 2, \end{cases} \quad (1.3)$$

其中 $g_k = \nabla f(x_k)$, 一般情况下要求 $g_k^T d_k < 0$, 更进一步要求 d_k 满足充分下降性^[7], 即满足

$$g_k^T d_k \leq -c \|g_k\|^2. \quad (1.4)$$

β_k 是一标量, β_k 的不同选取, 对应不同的共轭梯度法. 较著名的有

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \text{(Fletcher-Reeves[8], 1964),}$$

*收稿日期: 2022-09-25 接收日期: 2023-02-14

基金项目: 内蒙古自治区自然科学基金资助 (2020MS01001; 2022MS07006); 国家自然科学基金资助 (72163022).

作者简介: 朱铁锋 (1979-), 男, 副教授, 主要研究方向: 最优化理论与方法, 数理统计. E-mail: tfzhu2016@163.com.

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \text{ (Polak-Ribière-Polyak[9], 1969),}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \text{ (Dai-Yuan[10], 1999),}$$

等, $y_{k-1} = g_k - g_{k-1}$. 为得到上述方法的全局收敛性, 一般需步长 α_k 满足强 Wolfe 条件, 即满足

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k g_k^T d_k, \quad (1.5)$$

$$\sigma g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq -\sigma g_k^T d_k, \quad (1.6)$$

其中 $0 < \rho < \sigma < 1$.

如果优化函数不是二次函数时, 这些算法的计算效率有较大差异. 一般地, 基于非精确线搜索的 DY 和 FR 方法具有良好的收敛性能, 但计算效率不如 PRP 方法. 然而对于一般目标函数, 基于非精确线搜索的 PRP 方法无法保证产生的搜索方向为下降方向, 无法保证算法的全局收敛性. 为了建立兼具良好数值表现和收敛性质的方法, 近年来, 许多学者提出了改进的共轭梯度法, 这些算法在满足强 Wolfe 线搜索准则条件下具有充分下降性和全局收敛性. 例如, 文献 [11] 通过分段函数, 提出一种混合的 PRP-WYL 共轭梯度法. 文献 [12] 提出了一种带两个参数的三项共轭梯度法. 文献 [13] 在 PRP 公式中引入调节因子, 并据此提出了一个自调节 PRP 共轭梯度法. Yousif^[14] 提出一个改进的 RMIL 共轭梯度法. Sellami 和 Sellami^[15] 提出一个改进的 FR 共轭梯度法. 文献 [16] 提出一个三参数族共轭梯度法. Dong^[17] 提出一种改进的 PRP 共轭梯度法. 其他的改进共轭梯度法见文献 [18–20]. 这些共轭梯度法一般要求线搜索方向近似满足 Dai-Liao 共轭条件^[21], 即 $d_k^T y_{k-1} = -\xi g_k^T s_{k-1}$ ($\xi \geq 0$), 其中 $s_{k-1} = x_k - x_{k-1}$. 当搜索方向不能近似满足共轭条件时, 算法的充分下降和收敛性质将不再成立.

受这些文献的启发, 本文给出一个新的 d_k 和 β_k 的计算公式,

$$d_k = \begin{cases} -g_1 & k = 1 \\ (\beta_k - 1)g_k - \beta_k d_{k-1} & k \geq 2, \end{cases} \quad (1.7)$$

$$\beta_k = \frac{\|g_k\|^2}{\|g_k\|^2 - (g_k + g_{k-1})^T d_{k-1}}, \quad (1.8)$$

并证明了这类方法在不要求搜索方向为共轭方向, 不要求满足 Dai-Liao 共轭条件的情况下, 使用强 Wolfe 线搜索时仍具有充分下降和全局收敛性质, 且当目标函数为一致凸函数时, 新算法具有线性收敛速率. 最后, 我们通过数值试验来验证新算法的优越性.

2 算法及其性质

为证明本文所提出方法的全局收敛性, 需要如下两个常规假设.

(H1) $f(x)$ 在水平集 $\Omega := \{x \in R^n | f(x) \leq f(x_1)\}$ 上有下界, 其中 x_1 为算法初始点.

(H2) $f(x)$ 在水平集 Ω 的一个邻域 N 内连续可微, 且其梯度函数 g 满足 Lipschitz 条件, 即存在常数 $L > 0$, 满足 $\|g(x) - g(y)\| \leq L\|x - y\|$, $\forall x, y \in N$.

基于公式 (1.7) 和 (1.8), 建立如下算法 (简称算法 A):

Step 1: 给定初值 $x_1 \in R^n$, $\varepsilon > 0$, $\rho \in (0, 1)$, $\sigma \in (\rho, 1)$, 令 $d_1 = g_1$, $k := 1$.

Step 2: 如果 $\|g_k\| < \varepsilon$, 则停止, 否则转 Step 3.

Step 3: 由强 Wolfe 准则计算 α_k .

Step 4: 由 (1.7) 和 (1.8) 计算 d_k , β_k , $x_{k+1} = x_k + \alpha_k d_k$.

Step 5: 令 $k := k + 1$, 转 Step 2.

下面的引理 2.1 表明由算法 A 产生的搜索方向满足充分下降性质 (1.4).

引理 2.1 设目标函数 $f(x)$ 连续可微, 则算法 A 产生的搜索方向 d_k 满足 $g_k^T d_k \leq -c\|g_k\|^2$.

证 当 $k = 1$ 时, $g_1^T d_1 = -\|g_1\|^2 < -\frac{1}{\sigma+1}\|g_1\|^2 = -c\|g_1\|^2$, 这里 $c = \frac{1}{\sigma+1} > 0$. 假设对 $k-1$ 的情形 $g_{k-1}^T d_{k-1} \leq -c\|g_{k-1}\|^2$ 成立. 下面证明对 k 的情形也成立.

$$\begin{aligned}
 g_k^T d_k &= g_k^T [(\beta_k - 1)g_k - \beta_k d_{k-1}] \\
 &= (\beta_k - 1)\|g_k\|^2 - \beta_k g_k^T d_{k-1} \\
 &= \left(\frac{\|g_k\|^2}{\|g_k\|^2 - (g_k + g_{k-1})^T d_{k-1}} - 1 \right) \|g_k\|^2 - \frac{\|g_k\|^2}{\|g_k\|^2 - (g_k + g_{k-1})^T d_{k-1}} g_k^T d_{k-1} \\
 &= \|g_k\|^2 \left(\frac{\|g_k\|^2 - [\|g_k\|^2 - (g_k + g_{k-1})^T d_{k-1}] - g_k^T d_{k-1}}{\|g_k\|^2 - (g_k + g_{k-1})^T d_{k-1}} \right) \\
 &= \|g_k\|^2 \left(\frac{g_{k-1}^T d_{k-1}}{\|g_k\|^2 - (g_k + g_{k-1})^T d_{k-1}} \right) \\
 &= \|g_k\|^2 \left(\frac{g_{k-1}^T d_{k-1}}{\|g_k\|^2 - g_k^T d_{k-1} - g_{k-1}^T d_{k-1}} \right) \\
 &\leq \|g_k\|^2 \left(\frac{g_{k-1}^T d_{k-1}}{-g_k^T d_{k-1} - g_{k-1}^T d_{k-1}} \right) \\
 &= -\|g_k\|^2 \left(\frac{g_{k-1}^T d_{k-1}}{g_k^T d_{k-1} + g_{k-1}^T d_{k-1}} \right) \\
 &\leq -\|g_k\|^2 \left(\frac{g_{k-1}^T d_{k-1}}{\sigma g_{k-1}^T d_{k-1} + g_{k-1}^T d_{k-1}} \right) \\
 &= -\frac{1}{\sigma+1} \|g_k\|^2 \\
 &= -c\|g_k\|^2.
 \end{aligned}$$

因此引理得证.

引理 2.2 设目标函数 $f(x)$ 连续可微, 则算法 A 产生的标量 β_k 满足 $0 < \beta_k < 1$.

证 在强 Wolfe 准则下, 根据引理 2.1, β_k 的分母

$$\begin{aligned}
 \|g_k\|^2 - (g_k + g_{k-1})^T d_{k-1} &\geq \|g_k\|^2 + \sigma g_{k-1}^T d_{k-1} - g_{k-1}^T d_{k-1} \\
 &= \|g_k\|^2 + (\sigma - 1)g_{k-1}^T d_{k-1} \\
 &\geq \|g_k\|^2 + c(1 - \sigma)\|g_{k-1}\|^2 \\
 &\geq \|g_k\|^2.
 \end{aligned}$$

由于 β_k 的分子分母皆大于零, 且分母大于分子, 因此引理得证.

引理 2.3 设目标函数 $f(x)$ 连续可微, 则算法 A 产生的搜索方向 d_k 满足 $\|d_k\|^2 \leq \|g_k\|^2 + \|d_{k-1}\|^2$.

证 由 (1.7) 可知, 当 $k = 1$ 时, 引理显然成立. 当 $k \geq 2$ 时, $d_k - (\beta_k - 1)g_k = -\beta_k d_{k-1}$, 上式两端取模的平方, 并移项得到

$$\|d_k\|^2 = 2(\beta_k - 1)g_k^T d_k - (\beta_k - 1)^2 \|g_k\|^2 + (\beta_k)^2 \|d_{k-1}\|^2.$$

由引理 2.1, $g_k^T d_k < 0$, 引理 2.2, $0 < \beta_k < 1$, 所以上式第一项大于零, 因此上式 $\leq -(\beta_k - 1)^2 \|g_k\|^2 + (\beta_k)^2 \|d_{k-1}\|^2$, 由三角不等式, 上式 $\leq (\beta_k - 1)^2 \|g_k\|^2 + (\beta_k)^2 \|d_{k-1}\|^2 \leq \|g_k\|^2 + \|d_{k-1}\|^2$. 因此引理得证.

引理 2.4 设目标函数 $f(x)$ 连续可微, 则算法 A 产生的步长 α_k 满足 $\alpha_k \geq \frac{(\sigma-1)g_k^T d_k}{L\|d_k\|^2}$.

证 一方面, 根据假设 H2 可知,

$$\begin{aligned} (g_{k+1} - g_k)^T d_k &= (g(x_k + \alpha_k d_k) - g(x_k))^T d_k \\ &\leq \|g(x_k + \alpha_k d_k) - g(x_k)\| \|d_k\| \\ &\leq \|L(x_k + \alpha_k d_k - x_k)\| \|d_k\| \\ &= \alpha_k L \|d_k\|^2. \end{aligned}$$

另一方面, 根据强 Wolfe 准则可知, $(g_{k+1} - g_k)^T d_k = g_{k+1}^T d_k - g_k^T d_k \geq (\sigma - 1)g_k^T d_k$. 结合上述两方面, 可得

$$\alpha_k L \|d_k\|^2 \geq (\sigma - 1)g_k^T d_k \alpha_k \geq \frac{(\sigma - 1)g_k^T d_k}{L\|d_k\|^2}.$$

因此引理得证.

引理 2.5 若 H1 和 H2 成立, $\{x_k\}$ 是有算法 A 产生迭代点列, 则 $\sum_{k=2}^{\infty} \frac{\|g_k\|^4}{\|g_k\|^2 + \|d_{k-1}\|^2} < +\infty$

证 由 (1.5) 及引理 2.1, 2.3, 2.4 有

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\geq -\rho \alpha_k g_k^T d_k \\ &\geq -\rho \frac{(\sigma - 1)(g_k^T d_k)^2}{L\|d_k\|^2} \\ &\geq \rho \frac{(1 - \sigma)c^2 \|g_k\|^4}{L(\|g_k\|^2 + \|d_{k-1}\|^2)} \\ &= \theta \frac{\|g_k\|^4}{(\|g_k\|^2 + \|d_{k-1}\|^2)}, \end{aligned}$$

这里 $\theta = \frac{\rho(1-\sigma)c^2}{L}$. 因为 $\{f(x_k)\}$ 单调不增且有下界, 故知 $\sum_{k=2}^{\infty} (f(x_k) - f(x_{k+1})) < +\infty$, 从而

知 $\sum_{k=2}^{\infty} \frac{\|g_k\|^4}{\|g_k\|^2 + \|d_{k-1}\|^2} < +\infty$, 故引理得证.

3 算法的全局收敛性

定理 3.1 若**H1**和**H2**成立, $\{x_k\}$ 是由算法 A 产生迭代点列, 则或者 $\liminf_{k \rightarrow \infty} \|g_k\| = 0$, 或者 x_k 无界.

证 若 $\liminf_{k \rightarrow \infty} \|g_k\| = 0$ 不成立, 则存在常数 $\varepsilon > 0$, 对任意 $k \geq 1$, 有 $\|g_k\| \geq \varepsilon$. 由 $\Phi(\alpha) = \frac{\alpha^2}{\alpha + \|d_{k-1}\|^2}$ 为单调增函数知

$$\frac{\varepsilon^4}{\varepsilon^2 + \|d_{k-1}\|^2} \leq \frac{\|g_k\|^4}{\|g_k\|^2 + \|d_{k-1}\|^2}. \quad (3.1)$$

若 $\{x_k\}$ 有界, 则由**H2**知 $\{\|g_k\|^2\}$ 有界. 即存在 $M > 0$, 对任意 $k \geq 2$ 有 $\|g_k\|^2 \leq M$. 由引理 2.3 知 $\|d_k\|^2 \leq \|g_k\|^2 + \|d_{k-1}\|^2 \leq M + \|d_{k-1}\|^2 \leq \dots \leq kM$. 由上式 (3.1) 及引理 2.5 知

$$\sum_{k=2}^{\infty} \frac{\varepsilon^4}{\varepsilon^2 + kM} \leq \sum_{k=2}^{\infty} \frac{\varepsilon^4}{\varepsilon^2 + \|d_{k-1}\|^2} \leq \sum_{k=2}^{\infty} \frac{\|g_k\|^4}{\|g_k\|^2 + \|d_{k-1}\|^2} \leq +\infty.$$

而 $\sum_{k=2}^{\infty} \frac{\varepsilon^4}{\varepsilon^2 + kM} = +\infty$, 故矛盾, 从而知 $\{x_k\}$ 无界.

定理 3.2 若**H1**和**H2**成立, 且水平集 Ω 有界, $\{x_k\}$ 是由算法 A 产生迭代点列, 则 $\liminf_{k \rightarrow \infty} \|g_k\| = 0$.

证 由水平集 Ω 有界知 $\{x_k\}$ 有界, 由定理 3.1 证明可知 $\liminf_{k \rightarrow \infty} \|g_k\| = 0$.

4 线性收敛速率

为证明本文所提出方法的收敛速率, 需要如下假设.

(H3) $f(x)$ 是二次连续可微的一致凸函数. 即存在 $m > 0$, $M > 0$, 满足 $m\|y\|^2 \leq y^T \nabla^2 f(x)y \leq M\|y\|^2$, $\forall x, y \in R^n$. 若 **(H3)** 成立, 根据文献 [22] 可得到下面两个引理.

引理 4.1 若 **(H3)** 成立, 则 $f(x)$ 具有下列性质:

- (1) $f(x)$ 在 R^n 上有唯一的极小点 x^* .
- (2) 水平集 Ω 有界.
- (3) $\frac{m}{2}\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{M}{2}\|x - x^*\|^2$, $m\|x - x^*\| \leq \|g(x)\| \leq M\|x - x^*\|$.
- (4) 假设 **(H1)**, **(H2)** 成立.

引理 4.2 若**H3**成立, $\{x_k\}$ 是由算法 A 产生迭代点列, 且满足 $\liminf_{k \rightarrow \infty} \|g_k\| = 0$, 则 $x_k \rightarrow x^*$ ($k \rightarrow \infty$). 这里 x^* 为 $f(x)$ 的唯一极小点.

定理 4.1 若**H3**成立, $\{x_k\}$ 是由算法 A 产生迭代点列, 则 $x_k \rightarrow x^*$ ($k \rightarrow \infty$). 进一步, 则或者存在一个无穷子列 $\{x_k\}_{k \in K \subset \{1, 2, \dots\}}$, 使 $\lim_{k \in K} \frac{\|g_k\|}{\|d_k\|} = 0$; 或者 $\{x_k\}$ 线性收敛于 x^* .

证 由定理 3.2 知 $\liminf_{k \rightarrow \infty} \|g_k\| = 0$, 从而由引理 4.2 知 $x_k \rightarrow x^*$ ($k \rightarrow \infty$). 若 $\{\frac{\|d_k\|}{\|g_k\|}\}$ 无界, 则必存在无穷点列 $\{x_k\}_{k \in K \subset \{1, 2, \dots\}}$, 使 $\lim_{k \in K} \frac{\|d_k\|}{\|g_k\|} = +\infty$, 从而知 $\lim_{k \in K} \frac{\|g_k\|}{\|d_k\|} = 0$. 若 $\{\frac{\|d_k\|}{\|g_k\|}\}$ 有界, 即存在 $\frac{1}{\mu} > 0$, 对任意 $k \geq 1$, 有 $\frac{\|d_k\|}{\|g_k\|} \leq \frac{1}{\mu}$, 从而有

$$\frac{\|g_k\|}{\|d_k\|} \geq \mu. \quad (4.1)$$

由 $g_{k+1} - g_k = \alpha_k \int_0^1 \nabla^2 f(x_k + \tau \alpha_k d_k) d_k d\tau$, 和假设 **H3** 知

$$(g_{k+1} - g_k)^T d_k = \alpha_k \int_0^1 d_k^T \nabla^2 f(x_k + \tau \alpha_k d_k) d_k d\tau \leq \alpha_k M \|d_k\|^2. \quad (4.2)$$

又根据 Wolfe 条件得

$$(g_{k+1} - g_k)^T d_k \geq (\sigma - 1) g_k^T d_k, \quad (4.3)$$

故由 (4.2) 和 (4.3) 知

$$\alpha_k \geq \frac{(\sigma - 1) g_k^T d_k}{M \|d_k\|^2} \geq \frac{(\sigma - 1) g_k^T d_k}{2M \|d_k\|^2}. \quad (4.4)$$

由引理 2.1 和 (4.1) 知

$$\frac{-g_k^T d_k}{\|g_k\| \|d_k\|} \geq \frac{c \|g_k\|^2}{\|g_k\| \|d_k\|} = \frac{c \|g_k\|}{\|d_k\|} \geq c\mu. \quad (4.5)$$

由 (1.5), (4.4) 和 (4.5) 知

$$f(x_k) - f(x_{k+1}) \geq -\rho \alpha_k g_k^T d_k \geq \frac{\rho(1-\sigma)}{2M} \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 = \frac{\rho(1-\sigma)}{2M} \left(\frac{g_k^T d_k}{\|d_k\| \|g_k\|} \right)^2 \|g_k\|^2 \geq \frac{\rho(1-\sigma)c^2\mu^2}{2M} \|g_k\|^2.$$

令 $\eta = \frac{\rho(1-\sigma)c^2\mu^2}{2M}$, 由引理 4.1(3) 知

$$f(x_k) - f(x_{k+1}) \geq \eta \|g_k\|^2 \geq \eta m^2 \|x_k - x^*\|^2 \geq \frac{\eta m^2}{M} (f(x_k) - f(x^*)).$$

令 $\omega = \frac{\eta m^2}{M} = \frac{m^2 \rho(1-\sigma)c^2\mu^2}{M^2}$, 则

$$f(x_k) - f(x_{k+1}) \geq \omega (f(x_k) - f(x^*)),$$

即

$$f(x_{k+1}) - f(x_k) \leq -\omega (f(x_k) - f(x^*)). \quad (4.6)$$

下面证 $0 < \omega < 1$. 由 Cauchy-Schwartz 不等式及引理 2.1 有 $\|g_k\| \|d_k\| \geq -g_k^T d_k \geq c \|g_k\|^2$, 从而由 (4.5) 知 $c\mu \leq 1$, 故

$$\omega = \frac{m^2 \rho(1-\sigma)c^2\mu^2}{M^2} \leq \frac{m^2 \rho(1-\sigma)}{M^2} < 1.$$

又由 (4.6) 知

$$\begin{aligned} f(x_k) - f(x^*) &= f(x_k) - f(x_{k-1}) + f(x_{k-1}) - f(x^*) \\ &\leq -\omega (f(x_{k-1}) - f(x^*)) + (f(x_{k-1}) - f(x^*)) \\ &= (1 - \omega) (f(x_{k-1}) - f(x^*)) \\ &\leq \vdots \\ &\leq (1 - \omega)^{k-1} (f(x_1) - f(x^*)), \end{aligned}$$

从而由引理 4.1(3) 和上式知 $\|x_k - x^*\|^2 \leq \frac{2}{m} (f(x_k) - f(x^*)) \leq (1 - \omega)^{k-1} \frac{2(f(x_1) - f(x^*))}{m}$. 因为 x^* 为 $f(x)$ 的唯一极小点, 故 $f(x_1) - f(x^*) > 0$. 令 $p = \sqrt{\frac{2(f(x_1) - f(x^*))}{m}}$, $q = \sqrt{1 - \omega}$ ($0 < q < 1$), 则有 $\|x_k - x^*\| \leq pq^{k-1}$, 从而知 $\{x_k\}$ 线性收敛于 x^* . 定理得证.

5 数值试验

为了检验本文所提出算法的数值效果, 我们从无约束优化问题测试集中选取了部分算例进行测试 (具体算例见文献 [4]), 并将数值结果与 FR, PRP 和 DY 经典算法进行比较. 测试环境为 Windows 10 操作系统, CPU1.6GHZ 4GB RAM, 所有程序由 Matlab R2016a 软件实现. 参数设置为 $\rho = 0.04$, $\sigma = 0.6$, 算法的终止条件为 $\|g_k\| \leq 10^{-4}$ 或迭代次数超过 1000. 测试函数和数值结果见表 1 和表 2, 其中 Functions 表示测试函数的名称, NaN/NI/ $x_0/\bar{x}/f(\bar{x})/x^*/f(x^*)$ 分别代表程序运行无结果, 迭代次数, 算法初值, 算法终止时的迭代值, 算法终止时的函数值, 函数最优解, 函数最优值.

从表 2 结果可以看出, FR, PRP 和 DY 算法在数值试验中的表现并不理想, 较多函数寻优失效. 本文提出的新算法具有良好的数值表现, 总体上优于比较的算法, 是可行的, 有效的.

表 1 测试函数和 $x_0/x^*/f(x^*)$

Functions	$x_0/x^*/f(x^*)$
Ewh	(2, 2, 2, ..., 2, 2)/(1, 1, ..., 1, 1)/0
Raydan1	(10, 10)/(0, 0)/0.3
Wood	(-3, 10, -1, 3)/(1, 1, 1, 1)/0
Ewood	(1.2, 1, 1.2, 1)/(1, 1, 1, 1)/0
Emic	(2, 1, ..., 2, 1, 2, 2)/(0, 1, 1, 1, 0, 1, 1, 1)/0
Erosen	(-1.2, 1, 1.2, 1, ..., 1.2, 1)/(1, 1, 1, 1, ..., 1, 1)/0
LIARWHD	(4.1, 4.1, ..., 4.1)/(1, 1, 1, ..., 1, 1)/0
POWER	(1, 2, 3, 4, 5, 1, 2, 3, 4, 5, ..., 1, 2, 3, 4, 5)/(0, 0, ..., 0, 0)/0
pqd	(-0.5, 0.5, -0.6, 0.8, -0.9)/(0, 0, ..., 0, 0)/0
Etri	(0.1, 0.1, ..., 0.1)/(0, 0, ..., 0, 0)/0
Epow	(3, -1, 0, 1)/(0, 0, ..., 0, 0)/0
Perq	(1, 2, 3, 4, 5)/(0, 0, 0, 0, 0)/0
Etri1	(1, 2, ..., 10, 1, 2, ..., 10, ..., 1, 2, ..., 10)/(1, 2, ..., 1, 2)/0
QUARTC	(3, 3, ..., 3)/(1, 1, ..., 1)/0
Staircase1	(3, 4, 5, 6)/(0, 0, ..., 0)/0
Diagonal4	(-5, 5, 5, 5)/(0, 0, ..., 0)/0

表 2: $\bar{x}/f(\bar{x})/NI$ 的数值结果

Functions	Algorithms	$\bar{x}/f(\bar{x})/NI$
Ewh	FR	NaN
	PRP	NaN
	DY	NaN
	A	(0.9934, 0.9804, ..., 0.9934, 0.9804)/1.3019 * 10 ⁻⁴ /1000

Raydan1	FR	NaN
	PRP	NaN
	DY	NaN
	A	$10^{-3} * (-0.1377, -0.4749)/0.3/226$
Wood	FR	$(1, 1, 1, 1)/4.2158 * 10^{-10}/1000$
	PRP	NaN
	DY	NaN
	A	$(1, 1.0001, 1, 0.9999)/5.6609 * 10^{-9}/872$
Ewood	FR	NaN
	PRP	NaN
	DY	NaN
	A	$(1.1910, 1.4182, 0.7646, 0.5836)/0.0920/1000$
Emic	FR	NaN
	PRP	NaN
	DY	NaN
	A	$(0.2775, 1.3201, 1.3692, 1.3468, -0.0965, 0.6564, 0.6453, 0.5912)/0.0634/1000$
Erosen	FR	NaN
	PRP	NaN
	DY	NaN
	A	$(1, 1, 1, 1, \dots, 1, 1)/2.1835 * 10^{-5}/1000$
LIARWHD	FR	NaN
	PRP	NaN
	DY	NaN
	A	$(1, 1, \dots, 1)/2.4854 * 10^{-6}/1000$
POWER	FR	NaN
	PRP	NaN
	DY	NaN
	A	$10^{-4} * (-0.1703, -0.0785, 0.0013, -0.0005, 0.0010, 0.0009, 0.0029, 0.0035, -0.0002, 0.0080, -0.0022, -0.0061, 0.0013, 0.0000, 0.0059, -0.0033, 0.0033, -0.0009, -0.0012, 0.0043)/5.4962 * 10^{-10}/182$
pqd	FR	$10^{-3} * (-0.1580, 0.0813, 0.0295, 0.0170, 0.0113)/4.2593 * 10^{-8}/748$
	PRP	$10^{-3} * (-0.1001, 0.0482, 0.0162, 0.0084, 0.0049)/1.5856 * 10^{-8}/644$
	DY	$10^{-3} * (-0.9056, 0.4197, 0.0437, 0.2990, 0.1416)/1.6360 * 10^{-6}/50$
	A	$(-0.3362, 0.5199, -0.3280, 0.5519, -0.3655)/2.8628/1000$
Etri	FR	NaN
	PRP	$(0, 0, 0, 0, 0, 0, 0, 0, 0.1993, 0.1813)/7.8768 * 10^{-10}/21$
	DY	NaN
	A	$(0, 0, 0, 0, 0, 0, 0, 0.1993, 0.1813)/1.0050 * 10^{-10}/699$
Epow	FR	$(1, 0, 0, 1)/3.1294 * 10^{-11}/26$

	PRP	$(1, 0, 0, 1)/4.0827 * 10^{-11}/274$
	DY	NaN
	A	$(1, 0, 0, 1)/1.6725 * 10^{-11}/144$
Perq	FR	$10^{-5} * (0.0019, 0.0094, 0.7387, 0.5957, -0.0757)/3.1012 * 10^{-10}/44$
	PRP	$10^{-4} * (0.2696, -0.0030, -0.0014, -0.0001, 0.0748)/1.0186 * 10^{-9}/38$
	DY	$10^{-4} * (-0.0054, -0.1048, 0.1159, 0.0043, 0.0121)/6.3116 * 10^{-10}/27$
	A	$10^{-4} * (0.2321, -0.0017, -0.0005, -0.0004, -0.0047)/5.4478 * 10^{-10}/28$
Etri1	FR	NaN
	PRP	NaN
	DY	NaN
	A	$(1, 2, \dots, 1, 2)/0/115$
QUARTC	FR	$(0.4817, 0.4817, \dots, 0.4817, 0.4817)/0.3609/1000$
	PRP	$(1.8340, 1.8340, \dots, 1.8340, 1.8340)/2.4189/1000$
	DY	$(1, 1, \dots, 1, 1)/1.0187 * 10^{-23}/42$
	A	$(1, 1, \dots, 1, 1)/1.0851 * 10^{-26}/37$
Staircase1	FR	$10^{-4} * (0.1905, -0.0045, -0.2293, -0.0840)/8.8970 * 10^{-10}/40$
	PRP	$10^{-4} * (0.0656, -0.2782, 0.2894, -0.2157)/7.4715 * 10^{-10}/45$
	DY	$10^{-4} * (0.2651, -0.0922, -0.1942, -0.0118)/1.0172 * 10^{-9}/41$
	A	$10^{-4} * (-0.0364, 0.0920, -0.1080, 0.0716)/7.5355 * 10^{-10}/40$
Diagonal4	FR	$10^{-4} * (-0.6431, -0.0020, 0.6431, -0.0020)/4.1398 * 10^{-9}/54$
	PRP	$10^{-4} * (0.1127, -0.0060, -0.1127, -0.0060)/1.6268 * 10^{-10}/102$
	DY	$10^{-4} * (-0.4919, -0.0587, 0.4919, -0.0587)/5.8623 * 10^{-11}/75$
	A	$10^{-4} * (-0.6875, 0.0007, 0.6875, 0.0007)/4.7266 * 10^{-10}/49$

参 考 文 献

- [1] Gu R, Yuan Y X. A partial first-order affine-scaling method[J]. Acta Mathematica Sinica, English Series, 2019, 35: 1–16.
- [2] Gu C, Zhu D. Convergence of a three-dimensional dwindling filter algorithm without feasibility restoration phase[J]. Numerical Functional Analysis and Optimization, 2016, 37: 324–341.
- [3] 王珏钰, 顾超. 一种非单调滤子信赖域算法解线性不等式约束优化 [J]. 数学学报, 2020, 63(6): 601–620.
- [4] Zhu T F, Yan Z Z, Peng X Y. A modified nonlinear conjugate gradient method for engineering computation[J]. Mathematical Problems in Engineering, 2017, 2017: 1–11.
- [5] Goncalves M L N, Prudente L F. On the extension of the hager-zhang conjugate gradient method for vector optimization[J]. Computational Optimization and Applications, 2020, 76(11): 889–916.
- [6] 张迎春, 吕全义, 肖曼玉. 复线性方程组的预处理 mcg 算法 [J]. 工程数学学报, 2018, 35(3): 308–318.

- [7] 简金宝, 尹江华, 江羨珍. 一个充分下降的有效共轭梯度法 [J]. 计算数学, 2015, 37(4): 415–424.
- [8] Fletcher R, Reeves C. Function minimization by conjugate gradients[J]. Computer Journal, 1964, 7(2): 149–154.
- [9] Polyak B T. The conjugate gradient method in extreme problems[J]. Ussr. Comput. Math. Phys., 1969, 9: 94–112.
- [10] Dai Y H, Yuan Y X. A nonlinear conjugate gradient with a strong global convergence property[J]. SIAM journal of optimization, 1999, 10: 177–182.
- [11] 张鹏, 杜学武. 强 Wolfe 线搜索下一种混合的 PRP-WYL 共轭梯度法 [J]. 重庆师范大学学报 (自然科学版), 2020, 171(1): 46–56.
- [12] 夏师, 袁功林, 王博朋, 王晓亮. 一种具有充分下降性的三项共轭梯度法 [J]. 数学的实践与认识, 2018, 48(23): 96–102.
- [13] 江羨珍. 一个自调节 Polak-Ribiere-Polyak 型共轭梯度法 [J]. 应用数学位, 2017, (40)3: 449–459.
- [14] Yousif O O O. The convergence properties of rml+ conjugate gradient method under the strong wolfe line search[J]. Applied Mathematics and Computation, 2020, 367: 1–8.
- [15] Sellami B, Sellami M C E. Global convergence of a modified fletcher-reeves conjugate gradient method with wolfe line search[J]. Asian-European Journal of Mathematics, 2019, 13(4): 1–10.
- [16] 景书杰, 赵海燕. 带参数共轭梯度法簇的全局收敛性 [J]. 应用数学与计算数学学报, 2014, 28(3): 281–290.
- [17] Dong X. A modified nonlinear polak-ribière-polyak conjugate gradient method with sufficient descent property[J]. Calcolo, 2020, 57(3): 1–14.
- [18] 夏丽娜, 朱志斌. Wolfe 线搜索下的两类修正 FR 谱共轭梯度法 [J]. 应用数学, 2021, 34(3): 647–656.
- [19] 段复建, 杨冲, 李向利. 一个自适应 Dai-Liao 共轭梯度法 [J]. 应用数学, 2022, 35(2): 336–342.
- [20] 祝子长, 刘丽平. 基于 BFGS 修正的 HSDY 混合共轭梯度 [J]. 数学杂志, 2022, 42(3): 246–258.
- [21] Dai Y H, Liao L Z. New conjugacy conditions and related nonlinear conjugate gradient methods[J]. Applied Mathematics and Optimization, 2001, 43(1): 87–101.
- [22] 袁亚湘, 孙文瑜. 最优化理论与方法 [M]. 北京: 科学出版社, 1997.

A NEW OPTIMIZATION ALGORITHM UNDER STRONG WOLFE LINE SEARCH

ZHU Tie-feng

(School of Statistics and Mathematics, Inner Mongolia University of Finance and Economics,
Hohhot 010070, China)

Abstract: In this paper, a new optimization algorithm is studied. Firstly, by using the new formula and strong Wolfe line search, it is proved that the algorithm has sufficient descent and global convergence without requiring the search direction to satisfy the conjugation. Secondly, using the assumption that the objective function is uniformly convex, it is proved that the algorithm has linear convergence rate. Finally, numerical experiments show that the new algorithm is effective and feasible.

Keywords: unconstrained optimization; CG method; strong Wolfe line search; global convergence

2010 MR Subject Classification: 65K10; 90C30