

DOI: 10.13718/j.cnki.xdzk.2021.07.011

改进共轭梯度法的收敛性

林穗华

广西民族师范学院 教育科学学院, 广西 崇左 532200

摘要: 提出一类 PRP, HS, LS 共轭梯度法的修正参数公式, 改进方法的搜索方向自动充分下降. 在标准 WWP 线搜索和新型 MWWP 线搜索下, 证明了算法的全局收敛性. 数值实验表明算法结果是有效的.

关键词: 共轭梯度法; 充分下降性; WWP 线搜索; 全局收敛性

中图分类号: O221.2 **文献标志码:** A **文章编号:** 1673-9868(2021)07-0081-08

解极小化问题 $\min\{f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\}$ 的迭代法有多种, 其中共轭梯度法由于无须用到 $\nabla^2 f(\mathbf{x})$ 等 n 阶矩阵数据, 存储需求少而特别适合大规模优化问题^[1-18]. 传统共轭梯度法的迭代格式为:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k \mathbf{d}_k & (1) \\ \mathbf{d}_k &= \begin{cases} -\mathbf{g}_k & k=1 \\ -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1} & k \geq 2 \end{cases} & (2) \end{aligned}$$

其中: $\mathbf{d}_k$ 为搜索方向, α_k 为步长因子, β_k 为参数, $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$, $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$. 不同的 β_k 参数公式对应不同的共轭梯度法. 著名的 PRP, HS, LS, FR, DY, CD 方法的参数公式如下:

$$\begin{aligned} \beta_k^{\text{PRP}} &= \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\|\mathbf{g}_{k-1}\|^2} & \beta_k^{\text{HS}} &= \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} & \beta_k^{\text{LS}} &= \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}}{-\mathbf{g}_{k-1}^T \mathbf{d}_{k-1}} \\ \beta_k^{\text{FR}} &= \frac{\|\mathbf{g}_k\|^2}{\|\mathbf{g}_{k-1}\|^2} & \beta_k^{\text{DY}} &= \frac{\|\mathbf{g}_k\|^2}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} & \beta_k^{\text{CD}} &= \frac{\|\mathbf{g}_k\|^2}{-\mathbf{g}_{k-1}^T \mathbf{d}_{k-1}} \end{aligned}$$

其中: $\mathbf{y}_{k-1} = \mathbf{g}_k - \mathbf{g}_{k-1}$, $\|\cdot\|$ 为欧氏范数. 基于 $\beta_k \geq 0$ 在某些类型共轭梯度法收敛性分析中的重要性, 文献[5-14] 均对 β_k 采取相应的非负修正策略, 如文献[5] 对 β_k^{PRP} 截值得到 $\beta_k^{\text{PRP}+} = \max\{\beta_k^{\text{PRP}}, 0\} \geq 0$; 文献[7-8] 则从 β_k^{PRP} 与 β_k^{FR} , β_k^{HS} 与 β_k^{DY} , β_k^{LS} 与 β_k^{CD} 凸组合的方式, 得到 β_k^{PRP} , β_k^{HS} , β_k^{LS} 的非负修正公式如下:

$$\beta_k^{\text{WYL}} = \frac{\mathbf{g}_k^T \bar{\mathbf{y}}_{k-1}}{\|\mathbf{g}_{k-1}\|^2} \quad \beta_k^{\text{MHS}} = \frac{\mathbf{g}_k^T \bar{\mathbf{y}}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} \quad \beta_k^{\text{MLS}} = \frac{\mathbf{g}_k^T \bar{\mathbf{y}}_{k-1}}{-\mathbf{g}_{k-1}^T \mathbf{d}_{k-1}}$$

其中

$$\bar{\mathbf{y}}_{k-1} = \mathbf{g}_k - \frac{\|\mathbf{g}_k\|}{\|\mathbf{g}_{k-1}\|} \mathbf{g}_{k-1}$$

相应的 WYL, MHS, MLS 方法都具有较好的收敛性.

类似地, β_k^{PRP} , β_k^{HS} , β_k^{LS} 也可修正为:

$$\beta_k^{\text{VPRP}} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}^*}{\|\mathbf{g}_{k-1}\|^2} \quad \beta_k^{\text{VHS}} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}^*}{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}} \quad \beta_k^{\text{VLS}} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}^*}{-\mathbf{g}_{k-1}^T \mathbf{d}_{k-1}} \quad (3)$$

其中:

$$\mathbf{y}_{k-1}^* = \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \mathbf{g}_k - \mathbf{g}_{k-1}$$

这一修改方式能确保 $\beta_k \geq 0$, 但却无法满足充分下降条件: $\exists c \in (0, 1)$, 使

$$\mathbf{g}_k^T \mathbf{d}_k \leq -c \|\mathbf{g}_k\|^2 \quad \forall k \geq 1 \quad (4)$$

然而某些类型共轭梯度法的收敛性依赖于充分下降条件, 因此文献[11] 定理 1 只得将“满足充分下降条件”作为预设前提. 事实上 VPRP, VHS, VLS 方法不满足充分下降性, 其全局收敛性仍无法保证.

受文献[9-10, 16-17] 充分下降性策略的启发, 我们修改(3) 式的分母, 得到

$$\beta_k^{V_1} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}^*}{\max\{\|\mathbf{g}_{k-1}\|^2, \mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}| \}} \quad (5)$$

$$\beta_k^{V_2} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}^*}{\max\{\mathbf{d}_{k-1}^T \mathbf{y}_{k-1}, \mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}| \}} \quad (6)$$

$$\beta_k^{V_3} = \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}^*}{\max\{-\mathbf{g}_{k-1}^T \mathbf{d}_{k-1}, \mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}| \}} \quad (7)$$

其中常数 $\mu > 2$. 我们将讨论(1), (2) 式迭代法的收敛性, 其中参数 β_k 取自(5), (6) 或(7) 式, 步长因子 α_k 考虑采用经典的 weak Wolfe-Powell(WWP) 线搜索^[2] 及文献[18] 针对 BFGS 和 PRP 方法设计的 modified weak Wolfe-Powell(MWWP) 线搜索.

记参数为 δ, σ 的 WWP 线搜索条件为 WWP(δ, σ) 条件:

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) - f_k \leq \delta \alpha_k \mathbf{g}_k^T \mathbf{d}_k \quad (8)$$

$$\mathbf{g}(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k \geq \sigma \mathbf{g}_k^T \mathbf{d}_k \quad (9)$$

其中: $\delta \in (0, \frac{1}{2})$, $\sigma \in (\delta, 1)$.

记参数为 δ_1, δ, σ 的 MWWP 线搜索条件为 MWWP(δ_1, δ, σ) 条件:

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) - f_k \leq \delta \alpha_k \mathbf{g}_k^T \mathbf{d}_k + \alpha_k \min\{-\delta_1 \mathbf{g}_k^T \mathbf{d}_k, \delta \frac{\alpha_k}{2} \|\mathbf{d}_k\|^2\} \quad (10)$$

$$\mathbf{g}(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k \geq \sigma \mathbf{g}_k^T \mathbf{d}_k + \min\{-\delta_1 \mathbf{g}_k^T \mathbf{d}_k, \delta \alpha_k \|\mathbf{d}_k\|^2\} \quad (11)$$

其中: $\delta \in (0, \frac{1}{2})$, $\delta_1 \in (0, \delta)$, $\sigma \in (\delta, 1)$.

将 $\beta_k^{V_1}, \beta_k^{V_2}, \beta_k^{V_3}$ 统称为 β_k^V , 对应的算法 V_1, V_2 和 V_3 也统称为算法 V.

1 算法及性质

算法 V-WWP

步骤 1 设定初值 $\mathbf{x}_1 \in \mathbb{R}^n$, $\mu > 2$, $0 < \delta_1 < \delta < \sigma < 1$, $\epsilon > 0$, $\mathbf{d}_1 = -\mathbf{g}_1$, $k=1$. 若 $\|\mathbf{g}_k\| \leq \epsilon$, 则停止.

步骤 2 计算 α_k 满足 WWP(δ, σ) 条件(8) 和(9) 式.

步骤 3 由(1) 式计算 $\mathbf{x}_{k+1}$. 若 $\|\mathbf{g}_{k+1}\| \leq \epsilon$, 则停止.

步骤 4 由(5), (6) 或(7) 式计算 β_{k+1} , 由(2) 式计算 $\mathbf{d}_{k+1}$.

步骤 5 置 $k = k + 1$, 转步骤 2.

在算法 V-WWP 框架中, 将步骤 2 改为计算 α_k 满足 MWWP(δ_1, δ, σ) 条件(10) 和(11) 式, 则得到 V-MWWP 算法.

定理 1 算法 V-WWP 生成的序列 $\beta_k, \mathbf{g}_k, \mathbf{d}_k$ 满足 $0 \leq \beta_k$ 和充分下降条件(4).

证 $\forall k \geq 2$, 由 $-\|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\| \leq \mathbf{g}_k^T \mathbf{g}_{k-1} \leq \|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|$, 可得

$$0 \leq \mathbf{g}_k^T \mathbf{y}_{k-1}^* = \mathbf{g}_k^T \left(\frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \mathbf{g}_k - \mathbf{g}_{k-1} \right) = \|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\| - \mathbf{g}_k^T \mathbf{g}_{k-1} \leq 2 \|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\| \quad (12)$$

又有

$$\max \left\{ \|\mathbf{g}_{k-1}\|^2, \mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{g}_{k-1}| \right\} \geq \mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}| \geq 0 \quad (13)$$

$$\max \left\{ \mathbf{d}_{k-1}^T \mathbf{y}_{k-1}, \mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}| \right\} \geq \mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}| \geq 0 \quad (14)$$

$$\max \left\{ -\mathbf{g}_{k-1}^T \mathbf{d}_{k-1}, \mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}| \right\} \geq \mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}| \geq 0 \quad (15)$$

从而, 可得 $0 \leq \beta_k^{V_1}$, $0 \leq \beta_k^{V_2}$, $0 \leq \beta_k^{V_3}$. 根据算法 V-WWP 步骤(4) β_k 的取法, 可得 $0 \leq \beta_k$.

取 $c = 1 - \frac{2}{\mu}$, 则 $c \in (0, 1)$. 当 $k=1$ 时, 显然

$$\mathbf{g}_1^T \mathbf{d}_1 = -\|\mathbf{g}_1\|^2 \leq -c \|\mathbf{g}_1\|^2$$

对 $k \geq 2$ 情形, 若 $\mathbf{g}_k^T \mathbf{d}_{k-1} = 0$, 将 $\mathbf{g}_k^T$ 与(2)式两端作内积, 可得

$$\mathbf{g}_k^T \mathbf{d}_k = -\|\mathbf{g}_k\|^2 + \beta_k \mathbf{g}_k^T \mathbf{d}_{k-1} = -\|\mathbf{g}_k\|^2 \leq -c \|\mathbf{g}_k\|^2$$

若 $\mathbf{g}_k^T \mathbf{d}_{k-1} \neq 0$, 则由(12)–(15)式, 可得

$$\begin{aligned} \beta_k^{V_1} &\leq \frac{2 \|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|}{\mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}|} = \frac{2 \|\mathbf{g}_k\|^2}{\mu |\mathbf{g}_k^T \mathbf{d}_{k-1}|} \\ \beta_k^{V_2} &\leq \frac{2 \|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|}{\mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}|} = \frac{2 \|\mathbf{g}_k\|^2}{\mu |\mathbf{g}_k^T \mathbf{d}_{k-1}|} \\ \beta_k^{V_3} &\leq \frac{2 \|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|}{\mu \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} |\mathbf{g}_k^T \mathbf{d}_{k-1}|} = \frac{2 \|\mathbf{g}_k\|^2}{\mu |\mathbf{g}_k^T \mathbf{d}_{k-1}|} \end{aligned}$$

从而可得

$$0 \leq \beta_k \leq \frac{2 \|\mathbf{g}_k\|^2}{\mu |\mathbf{g}_k^T \mathbf{d}_{k-1}|}$$

进一步可得

$$\mathbf{g}_k^T \mathbf{d}_k = -\|\mathbf{g}_k\|^2 + \beta_k \mathbf{g}_k^T \mathbf{d}_{k-1} \leq -\|\mathbf{g}_k\|^2 + \beta_k |\mathbf{g}_k^T \mathbf{d}_{k-1}| \leq -\left(1 - \frac{2}{\mu}\right) \|\mathbf{g}_k\|^2 = -c \|\mathbf{g}_k\|^2$$

定理 1 得证.

2 原理与假设

证明算法的收敛性需要用到以下假设和引理.

假设:

- (i) $f(\mathbf{x})$ 在水平集 $\Omega = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \leq f(\mathbf{x}_1)\}$ 下方有界, Ω 有界.
- (ii) $\nabla f(\mathbf{x})$ 在 Ω 的某邻域 N 上 Lipschitz 连续, 即存在常数 $L > 0$, 使得

$$\|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\| \quad \forall \mathbf{x}, \mathbf{y} \in N \quad (16)$$

引理 1^[3,6] 考虑满足如下条件的任一共轭梯度法(1)–(2):

- (a) $\beta_k \geq 0$.
- (b) 搜索方向满足充分下降条件(4).
- (c) 以下 Zoutendijk 条件成立:

$$\sum_{k \geq 1} \frac{(\mathbf{g}_k^T \mathbf{d}_k)^2}{\|\mathbf{d}_k\|^2} < +\infty \quad (17)$$

(d) β_k 满足性质(*): 设 $0 < r \leq \|g_k\| \leq \bar{r}$, 存在常数 $b > 1$ 和 $\lambda > 0$, 对 $\forall k \geq 2$ 有

$$|\beta_k| < b \text{ 和 } \|s_{k-1}\| \leq \lambda \Rightarrow |\beta_k| \leq \frac{1}{2b}$$

若假设 (i) 和 (ii) 成立, 则该迭代全局收敛.

以下算法收敛性分析中均假设 $\|g_k\| \neq 0$, 否则算法已得到 $f(x)$ 的稳定点而终止.

3 全局收敛性

定理 2 若假设 (i) 和 (ii) 成立, 则算法 V-WWP 生成的序列 g_k, d_k 满足 Zoutendijk 条件(17).

证 由(8)式和定理 1, 知 $f_{k+1} \leq f_k + \delta \alpha_k g_k^T d_k < f_k$, 递推可得 $f_{k+1} < f_k < \dots < f_1$. 再由假设 (i) 可知, 序列 $\{f_k\}$ 单调有界从而收敛, 即 $\lim_{k \rightarrow \infty} f_{k+1}$ 为常数.

由(9)式和假设 (ii), 可得

$$L \alpha_k \|d_k\|^2 \geq d_k^T (g_{k+1} - g_k) \geq (\sigma - 1) g_k^T d_k$$

从而可得

$$\alpha_k \geq \frac{(\sigma - 1)}{L} \frac{g_k^T d_k}{\|d_k\|^2} \quad (18)$$

由(8), (18) 式可得

$$f_k - f_{k+1} \geq -\delta \alpha_k g_k^T d_k \geq \frac{\delta(1 - \sigma)}{L} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \quad (19)$$

(19) 式两端对 $k = 1, 2, \dots$ 求和, 可得

$$f_1 - \lim_{k \rightarrow \infty} f_{k+1} \geq \frac{\delta(1 - \sigma)}{L} \sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2}$$

从而可知(17)式成立, 定理 2 得证.

定理 3 若假设 (i) 和 (ii) 成立, β_k, g_k, d_k 为算法 V-WWP 生成的序列, 则 β_k 满足性质(*).

证 由(9)式和定理 1, 可得

$$\begin{aligned} & \max\{d_{k-1}^T y_{k-1}, \mu \frac{\|g_{k-1}\|}{\|g_k\|} |g_k^T d_{k-1}|\} \geq \\ & d_{k-1}^T y_{k-1} \geq (1 - \sigma) g_{k-1}^T d_{k-1} \geq c(1 - \sigma) \|g_{k-1}\|^2 \\ & \max\{-g_{k-1}^T d_{k-1}, \mu \frac{\|g_{k-1}\|}{\|g_k\|} |g_k^T d_{k-1}|\} \geq \\ & -g_{k-1}^T d_{k-1} \geq c \|g_{k-1}\|^2 > c(1 - \sigma) \|g_{k-1}\|^2 \\ & \max\{\|g_{k-1}\|^2, \mu \frac{\|g_{k-1}\|}{\|g_k\|} |g_k^T d_{k-1}|\} \geq \\ & \|g_{k-1}\|^2 \geq c(1 - \sigma) \|g_{k-1}\|^2 \end{aligned}$$

从而可得

$$\begin{aligned} 0 & \leq \beta_k^{V_1} \leq \frac{g_k^T y_{k-1}^*}{c(1 - \sigma) \|g_{k-1}\|^2} \\ 0 & \leq \beta_k^{V_2} \leq \frac{g_k^T y_{k-1}^*}{c(1 - \sigma) \|g_{k-1}\|^2} \\ 0 & \leq \beta_k^{V_3} \leq \frac{g_k^T y_{k-1}^*}{c(1 - \sigma) \|g_{k-1}\|^2} \end{aligned}$$

根据算法 V-WWP 步骤(4) β_k 的取法, 可知

$$0 \leq \beta_k \leq \frac{g_k^T y_{k-1}^*}{c(1 - \sigma) \|g_{k-1}\|^2} \quad (20)$$

假设 $\forall k \geq 1, 0 < r \leq \|g_k\| \leq \bar{r}$, 其中 r 和 $\bar{r}$ 为常数. 取

$$b = \frac{2\bar{r}}{c(1-\sigma)r} \quad \lambda = \frac{c^2(1-\sigma)^2 r^2}{8L\bar{r}}$$

则由 $\sigma \in (0, 1)$, $c = 1 - \frac{2}{\mu} \in (0, 1)$, 知 $b > 1$, $\lambda > 0$.

由(12) 和(20) 式, 可得

$$|\beta_k| \leq \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}^*}{c(1-\sigma) \|\mathbf{g}_{k-1}\|^2} \leq \frac{2 \|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|}{c(1-\sigma) \|\mathbf{g}_{k-1}\|^2} = \frac{2 \|\mathbf{g}_k\|}{c(1-\sigma) \|\mathbf{g}_{k-1}\|} \leq \frac{2\bar{r}}{c(1-\sigma)r} = b$$

设 $\|\mathbf{s}_{k-1}\| \leq \lambda$, 则由(16), (20) 式及

$$\begin{aligned} \|\mathbf{y}_{k-1}^*\| &= \left\| \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \mathbf{g}_k - \mathbf{g}_{k-1} \right\| = \\ &= \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \left\| (\mathbf{g}_k - \mathbf{g}_{k-1}) + \mathbf{g}_{k-1} \left(1 - \frac{\|\mathbf{g}_k\|}{\|\mathbf{g}_{k-1}\|} \right) \right\| \leq \\ &= \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \left(\|\mathbf{g}_k - \mathbf{g}_{k-1}\| + \|\mathbf{g}_{k-1}\| \left| 1 - \frac{\|\mathbf{g}_k\|}{\|\mathbf{g}_{k-1}\|} \right| \right) = \\ &= \frac{\|\mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} (\|\mathbf{g}_k - \mathbf{g}_{k-1}\| + |\|\mathbf{g}_{k-1}\| - \|\mathbf{g}_k\||) \leq \\ &= \frac{2 \|\mathbf{g}_{k-1}\| \|\mathbf{g}_k - \mathbf{g}_{k-1}\|}{\|\mathbf{g}_k\|} \end{aligned}$$

可得

$$\begin{aligned} |\beta_k| &\leq \frac{\mathbf{g}_k^T \mathbf{y}_{k-1}^*}{c(1-\sigma) \|\mathbf{g}_{k-1}\|^2} \leq \frac{\|\mathbf{g}_k\| \|\mathbf{y}_{k-1}^*\|}{c(1-\sigma) \|\mathbf{g}_{k-1}\|^2} \leq \frac{2 \|\mathbf{g}_k - \mathbf{g}_{k-1}\|}{c(1-\sigma) \|\mathbf{g}_{k-1}\|} \leq \\ &= \frac{2L \|\mathbf{s}_{k-1}\|}{c(1-\sigma)r} \leq \frac{2L\lambda}{c(1-\sigma)r} = \frac{1}{2b} \end{aligned}$$

定理 3 得证.

由定理 1—3 以及引理 1, 可得如下定理 4.

定理 4 若假设 (i) 和 (ii) 成立, $\mathbf{g}_k$ 为算法 V-WWP 生成的序列, 则 $\liminf_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0$.

定理 5 若 α_k 满足 MWWP(δ_1, δ, σ) 条件, 则 α_k 也满足 WWP($\delta - \delta_1, \sigma$) 条件.

证 设 α_k 由 MWWP(δ_1, δ, σ) 线搜索产生, 其中 $\delta \in (0, \frac{1}{2})$, $\delta_1 \in (0, \delta)$, $\sigma \in (\delta, 1)$, 则由(10)

和(11) 式, 可得

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) - f_k \leq \delta \alpha_k \mathbf{g}_k^T \mathbf{d}_k + \alpha_k \min\{-\delta_1 \mathbf{g}_k^T \mathbf{d}_k, \delta \frac{\alpha_k}{2} \|\mathbf{d}_k\|^2\} \leq$$

$$\delta \alpha_k \mathbf{g}_k^T \mathbf{d}_k - \delta_1 \alpha_k \mathbf{g}_k^T \mathbf{d}_k = (\delta - \delta_1) \alpha_k \mathbf{g}_k^T \mathbf{d}_k \quad (21)$$

$$\mathbf{g}(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k \geq \sigma \mathbf{g}_k^T \mathbf{d}_k + \min\{-\delta_1 \mathbf{g}_k^T \mathbf{d}_k, \delta \alpha_k \|\mathbf{d}_k\|^2\} \geq \sigma \mathbf{g}_k^T \mathbf{d}_k \quad (22)$$

显然 $\delta - \delta_1 \in (0, \frac{1}{2})$, 由(21) 和(22) 式, 可知 α_k 也满足参数为 $\delta - \delta_1, \sigma$ 的 WWP 线搜索条件. 定

理 5 得证.

由定理 5, 类似定理 4 的证明过程, 可得算法 V-MWWP 全局收敛, 即如下定理 6 成立.

定理 6 若假设 (i) 和 (ii) 成立, $\mathbf{g}_k$ 为算法 V-MWWP 生成的序列, 则 $\liminf_{k \rightarrow \infty} \|\mathbf{g}_k\| = 0$.

4 数值实验

利用文献[19]的部分测试问题, 维数分别取 1 000, 5 000, 10 000, 对算法 PRP(采用 MWWP 搜索) 和算法 V-MWWP 进行数值试验. 试验环境为: ThinkPad E580, Intel(R) Core(TM) i5-8250U CPU@ 1.60GHz, RAM 8.00GB, Windows 10, Matlab R2016a. 算法参数 $\delta_1 = 0.05$, $\delta = 0.1$, $\sigma = 0.9$, $\mu = 5.1$,

$\epsilon=10^{-6}$. 终止条件为 $\| \mathbf{g}_k \| \leqslant 10^{-6}$ 或迭代次数 $N \geqslant 1\,000$. 计算结果见表 1, 其中 t 表示算法所耗 CPU 时间, NaN 表示算法终止时未满足 $\| \mathbf{g}_k \| \leqslant \epsilon$.

表 1 数值结果

问 题	维数	PRP		V1		V2		V3	
		N	t/s	N	t/s	N	t/s	N	t/s
Extended Trigonometric Function	1 000	147	0.094	138	0.078	138	0.109	128	0.078
	5 000	178	2.438	187	2.438	185	2.531	189	2.406
	10 000	178	4.922	154	4.578	168	4.188	156	4.234
Raydan 2 Function	1 000	NaN	NaN	37	0	37	0.016	37	0.016
	5000	NaN	NaN	33	0.125	33	0.125	33	0.172
	10 000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Diagonal 5 Function (MatrixRom)	1 000	5	0	5	0	5	0	5	0
	5 000	5	0.031	8	0.094	8	0.141	8	0.031
	10 000	4	0.078	9	0.172	9	0.172	9	0.188
Extended Himmelblau Function	1 000	61	0	23	0.016	23	0.016	23	0
	5 000	65	0.172	38	0.109	38	0.109	38	0.109
	10 000	66	0.344	25	0.188	24	0.125	25	0.063
Extended Block Diagonal BD1 Function	1 000	119	0.031	76	0.016	74	0.016	76	0.016
	5 000	148	0.828	100	0.516	99	0.609	100	0.641
	10 000	159	1.891	65	0.75	60	0.594	60	0.859
Extended Quadratic Penalty QP1 Function	1 000	77	0.016	93	0.031	73	0.016	73	0.078
	5 000	90	0.406	78	0.516	78	0.344	78	0.375
	10 000	105	0.875	100	1.547	93	1.172	112	1.969
A Quadratic Function QF2	1 000	35	0.016	25	0.016	25	0.016	25	0
	5 000	35	0.109	26	0.109	26	0.063	26	0.109
	10 000	35	0.172	25	0.109	25	0.109	25	0.109
DQDRTIC (CUTE)	1 000	NaN	NaN	59	0.016	NaN	NaN	143	0.016
	5 000	NaN	NaN	99	0.281	NaN	NaN	91	0.281
	10 000	NaN	NaN	542	3.266	141	0.984	87	0.609
DIXMAANA (CUTE)	1 000	51	0.078	38	0.063	38	0.063	38	0.063
	5 000	54	1.641	35	1.031	35	1.172	35	1.313
	10 000	55	2.859	36	1.094	36	1.141	36	2.031
DIXMAANC (CUTE)	1 000	137	0.219	122	0.188	122	0.188	122	0.172
	5 000	141	4.469	125	3.781	125	4.078	125	4.031
	10 000	141	7.547	127	5.641	127	5.375	127	6.578
Broyden Tridiagonal	1 000	129	0.094	77	0.047	45	0.047	58	0.031
	5 000	144	1.938	108	1.469	99	1.313	122	1.75
	10 000	191	5.141	128	3.766	130	3.875	129	3.875
DIXMAAND (CUTE)	1 000	43	0.078	33	0.047	33	0.063	33	0.047
	5 000	44	1.391	33	0.953	33	1.063	33	1.125
	10 000	45	2.5	34	1.313	34	1.625	34	1.5

续表 1

问 题	维数	PRP		V1		V2		V3	
		N	t/s	N	t/s	N	t/s	N	t/s
COSINE (CUTE)	1 000	NaN	NaN	35	0.016	35	0.031	35	0.016
	5 000	NaN	NaN	39	0.516	36	0.375	36	0.484
	10 000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Extended DENSCHNB (CUTE)	1 000	75	0.016	60	0.078	60	0	60	0
	5 000	80	0.172	64	0.219	64	0.172	64	0.219
	10 000	81	0.547	65	0.438	65	0.438	65	0.297
Extended DENSCHNF (CUTE)	1 000	135	0.016	34	0	34	0	34	0.016
	5 000	76	0.5	45	0.234	58	0.359	33	0.125
	10 000	80	0.844	43	0.344	38	0.391	52	0.563

应用文献[20]的技术得到算法 PRP 与算法 V 关于迭代次数和 CPU 时间比较的效能图(图 1 和图 2)。可见算法 V 对这些测试问题的数值表现比算法 PRP 更好些。

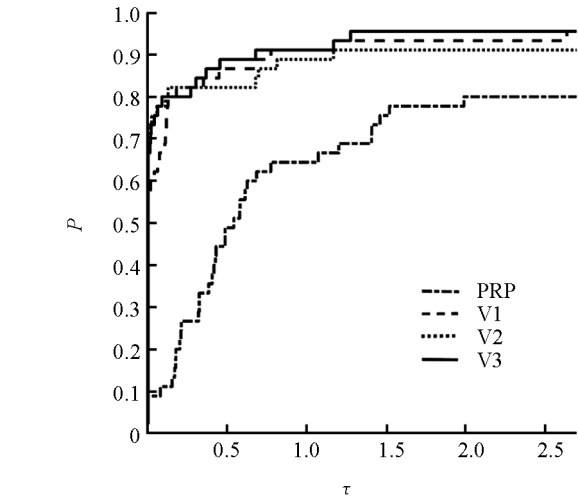


图 1 算法迭代次数效能图

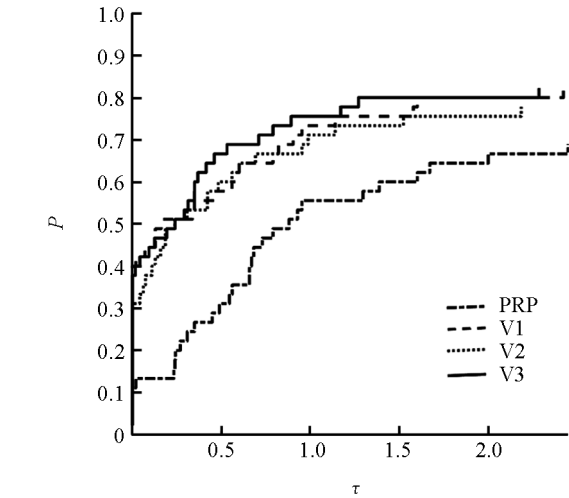


图 2 算法 CPU 时间效能图

参考文献：

[1] HESTENES M R, STIEFEL E. Methods of Conjugate Gradients for Solving Linear Systems [J]. Journal of Research of the National Bureau of Standards, 1952, 49(6): 409.

[2] 戴或虹. 非线性共轭梯度法 [M]. 上海: 上海科学技术出版社, 2000.

[3] HAGER W W W, ZHANG H C. A Survey of Nonlinear Conjugate Gradient Methods [J]. Pacific Journal of Optimization, 2006, 2(1): 35-58.

[4] HAGER W W, ZHANG H C. A New Conjugate Gradient Method with Guaranteed Descent and an Efficient Line Search [J]. SIAM Journal on Optimization, 2005, 16(1): 170-192.

[5] POWELL M J D. Convergence Properties of Algorithms for Nonlinear Optimization [J]. SIAM Review, 1986, 28(4): 487-500.

[6] GILBERT J C, NOCEDAL J. Global Convergence Properties of Conjugate Gradient Methods for Optimization [J]. SIAM Journal on Optimization, 1992, 2(1): 21-42.

[7] WEI Z X, YAO S W, LIU L Y. The Convergence Properties of Some New Conjugate Gradient Methods [J]. Applied Mathematics and Computation, 2006, 183(2): 1341-1350.

[8] SHENGWEI Y, WEI Z X, HUANG H. A Note about WYL's Conjugate Gradient Method and Its Applications [J]. Applied Mathematics and Computation, 2007, 191(2): 381-388.

- [9] HUANG H, LIN S H. A Modified Wei-Yao-Liu Conjugate Gradient Method for Unconstrained Optimization [J]. Applied Mathematics and Computation, 2014, 231: 179-186.
- [10] 黎 勇, 韦增欣. 一种自动充分下降的共轭梯度法 [J]. 西南师范大学学报(自然科学版), 2016, 41(5): 36-40.
- [11] 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.
- [12] 林穗华. 一类充分下降共轭梯度法的全局收敛性 [J]. 吉林大学学报(理学版), 2017, 55(4): 874-880.
- [13] 林穗华. 基于 Wolfe 线搜索的修正共轭梯度算法 [J]. 安徽大学学报(自然科学版), 2018, 42(2): 47-53.
- [14] 关 哲, 于宪伟. 标准 Wolfe 线搜索下修正的 DY 共轭梯度法 [J]. 西南师范大学学报(自然科学版), 2016, 41(1): 31-34.
- [15] 赛·闹尔再, 张慧玲. 修正 LS 共轭梯度方法及其收敛性 [J]. 西南师范大学学报(自然科学版), 2016, 41(7): 20-26.
- [16] 李春念, 袁功林. 求解无约束问题的修正 PRP 共轭梯度算法 [J]. 西南大学学报(自然科学版), 2018, 40(9): 67-75.
- [17] YUAN G L, WEI Z X, LI G Y. A Modified Polak-Ribière-Polyak Conjugate Gradient Algorithm for Nonsmooth Convex Programs [J]. Journal of Computational and Applied Mathematics, 2014, 255: 86-96.
- [18] YUAN G L, WEI Z X, LU X W. Global Convergence of BFGS and PRP Methods under a Modified Weak Wolfe-Powell Line Search [J]. Applied Mathematical Modelling, 2017, 47: 811-825.
- [19] ANDREI N. An Unconstrained Optimization Test Functions Collection [J]. Advanced Modeling and Optimization, 2008, 10(1): 147-161.
- [20] DOLAN E D, MORE J J. Benchmarking Optimization Software with Performance Profiles [J]. Mathematical Programming, 2002, 91(2): 201-213.

Global Convergence of Some Improved Conjugate Gradient Methods

LIN Sui-hua

School of Education Science, Guangxi Normal University for Nationalities, Chongzuo Guangxi 532200, China

Abstract: A class of modified parameter formulas of PRP, HS and LS conjugated gradient methods is proposed, and their search direction automatically possesses the sufficient descent property. The global convergence of the algorithms is proved under the standard WWP line search and the new modified WWP line search. Preliminary numerical experiments show that these algorithms are effective.

Key words: conjugate gradient method; sufficient descent property; WWP line search; global convergence

责任编辑 张 枸