

求解大规模非光滑优化问题的一种修正 Hestenes-Stiefel 共轭梯度算法^①

黎 勇¹, 袁功林²

1. 百色学院 数学与统计学院, 广西 百色 533000; 2. 广西大学 数学与信息科学学院, 南宁 530004

摘要: 利用 Moreau-Yosida 正则化技术和非单调线搜索技术, 设计了一种针对大规模非光滑优化问题的修正 Hestenes-Stiefel 共轭梯度算法. 该算法的搜索方向不仅自动满足充分下降条件, 而且属于信赖域. 在适当条件下, 新算法全局收敛. 初步的数值实验也表明新算法对于求解大规模非光滑无约束凸优化问题是有效的.

关键词: 非光滑; 大规模; 共轭梯度法; 充分下降条件; 全局收敛性

中图分类号: O224

文献标志码: A

文章编号: 1673-9868(2018)05-0081-08

非线性共轭梯度法由于具有算法简单、所需的存储需求小等特点, 被广泛应用于大规模光滑无约束优化问题的求解, 出现了如 PRP^[1-2], HS^[3], LS^[4], FR^[5] 等一批经典的共轭梯度算法. 然而在实际生活中, 我们经常会碰到大规模的非光滑优化问题, 如图形图像处理、生物工程等等. 充分利用共轭梯度法的特点来求解非光滑凸优化问题是一个有意义的问题, 目前这方面的研究成果还比较少. 文献[6]提出了一种新的修正 Polak-Ribière-Polyak 共轭梯度法, 通过利用 Moreau-Yosida 正则化技术, 并结合非单调线搜索, 新算法能够有效解决非光滑凸优化问题. 同时, 文献[6]也提出了一个修正 Hestenes-Stiefel(HS)参数公式, 但是并未进一步设计相应的共轭梯度法, 也未对方法的收敛性质和数值效果做讨论. 本文将以修正 HS 参数公式为基础建立相应的共轭梯度算法, 并讨论其相关性质.

1 非光滑凸问题及其相关理论

考虑如下非光滑无约束优化问题:

$$\min\{f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\} \quad (1)$$

其中 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是凸函数(可能非光滑).

目标函数 f 经过 Moreau-Yosida 正则化后记作 $F: \mathbb{R}^n \rightarrow \mathbb{R}$:

$$F(\mathbf{x}) = \min_{\mathbf{z} \in \mathbb{R}^n} \{f(\mathbf{z}) + \frac{1}{2\lambda} \|\mathbf{z} - \mathbf{x}\|^2\} \quad (2)$$

这里参数 $\lambda > 0$, $\|\cdot\|$ 指欧氏范数. 令

$$\phi(\mathbf{z}, \mathbf{x}) = f(\mathbf{z}) + \frac{1}{2\lambda} \|\mathbf{z} - \mathbf{x}\|^2$$

① 收稿日期: 2016-12-20

基金项目: 国家自然科学基金项目(11661001, 11661009); 广西自然科学基金项目(2014GXNSFAA118030); 广西教育厅科研项目(YB2014389); 广西中青年教师能力提升项目(KY2016YB417).

作者简介: 黎 勇(1973-), 男, 壮族, 副教授, 主要从事最优化理论与方法的研究.

并定义

$$\mathbf{h}(\mathbf{x}) = \operatorname{argmin}_{\mathbf{z}} \psi(\mathbf{z}, \mathbf{x})$$

因为 $\psi(\mathbf{z}, \mathbf{x})$ 对每一个固定的 $\mathbf{x}$ 强凸, 所以 $\mathbf{h}(\mathbf{x})$ 唯一. 因此函数 $F(\mathbf{x})$ 可以表示为

$$F(\mathbf{x}) = f(\mathbf{h}(\mathbf{x})) + \frac{1}{2\lambda} \|\mathbf{h}(\mathbf{x}) - \mathbf{x}\|^2$$

根据文献[7-9] 容易知道 F 具有以下重要性质:

(i) F 是有限凸函数, 且处处可微:

$$\mathbf{g}(\mathbf{x}) = \nabla F(\mathbf{x}) = \frac{\mathbf{x} - \mathbf{h}(\mathbf{x})}{\lambda} \quad (3)$$

进一步地, $\mathbf{g}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ Lipschitz 连续, 即

$$\|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y})\| \leq \frac{1}{\lambda} \|\mathbf{x} - \mathbf{y}\| \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad (4)$$

(ii) $\mathbf{x}$ 是(1) 式的最优解当且仅当 $\nabla F(\mathbf{x}) = 0$, 即 $\mathbf{h}(\mathbf{x}) = \mathbf{x}$.

从以上定义容易看出, $F(\mathbf{x})$ 和 $\mathbf{g}(\mathbf{x})$ 可以通过 $\operatorname{argmin}_{\mathbf{z} \in \mathbb{R}^n} \psi(\mathbf{z}, \mathbf{x})$ 的最优解来获得. 然而 $\operatorname{argmin}_{\mathbf{z} \in \mathbb{R}^n} \psi(\mathbf{z}, \mathbf{x}) = \mathbf{h}(\mathbf{x})$ 很难精确求出, 甚至不可能. 因此在实际计算中, 不能用精确的 $\mathbf{h}(\mathbf{x})$ 去定义 $F(\mathbf{x})$ 和 $\mathbf{g}(\mathbf{x})$. 由于对 $\forall \mathbf{x} \in \mathbb{R}^n, \epsilon > 0$, 存在向量 $\mathbf{h}^a(\mathbf{x}, \epsilon) \in \mathbb{R}^n$, 使得

$$f(\mathbf{h}^a(\mathbf{x}, \epsilon)) + \frac{1}{2\lambda} \|\mathbf{h}^a(\mathbf{x}, \epsilon) - \mathbf{x}\|^2 \leq F(\mathbf{x}) + \epsilon \quad (5)$$

因此, 当 ϵ 充分小的时候, 可用 $\mathbf{h}^a(\mathbf{x}, \epsilon)$ 近似定义 $F(\mathbf{x})$ 和 $\mathbf{g}(\mathbf{x})$:

$$F^a(\mathbf{x}, \epsilon) = f(\mathbf{h}^a(\mathbf{x}, \epsilon)) + \frac{1}{2\lambda} \|\mathbf{h}^a(\mathbf{x}, \epsilon) - \mathbf{x}\|^2 \quad (6)$$

$$\mathbf{g}^a(\mathbf{x}, \epsilon) = \frac{\mathbf{x} - \mathbf{h}^a(\mathbf{x}, \epsilon)}{\lambda} \quad (7)$$

这样定义的 $F^a(\mathbf{x}, \epsilon)$ 和 $\mathbf{g}^a(\mathbf{x}, \epsilon)$ 有以下重要性质:

命题 1 设 $\mathbf{h}^a(\mathbf{x}, \epsilon)$ 满足(5) 式, $F^a(\mathbf{x}, \epsilon)$ 和 $\mathbf{g}^a(\mathbf{x}, \epsilon)$ 分别由(6), (7) 式定义, 则有

$$F(\mathbf{x}) \leq F^a(\mathbf{x}, \epsilon) \leq F(\mathbf{x}) + \epsilon \quad (8)$$

$$\|\mathbf{h}^a(\mathbf{x}, \epsilon) - \mathbf{h}(\mathbf{x})\| \leq \sqrt{2\lambda\epsilon} \quad (9)$$

$$\|\mathbf{g}^a(\mathbf{x}, \epsilon) - \mathbf{g}(\mathbf{x})\| \leq \sqrt{\frac{2\epsilon}{\lambda}} \quad (10)$$

以上命题说明, 可以通过选取充分小的参数 ϵ , 使得 $F^a(\mathbf{x}, \epsilon)$ 和 $\mathbf{g}^a(\mathbf{x}, \epsilon)$ 无限接近 $F(\mathbf{x})$ 和 $\mathbf{g}(\mathbf{x})$. 命题 1 及其证明详见文献[10].

2 求解非光滑问题的修正 HS 共轭梯度法

针对非光滑无约束凸优化问题(1), Yuan 等在文献[6] 提出一个修正 HS 参数公式:

$$\mathbf{d}_{k+1} = \begin{cases} -\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1}) + \frac{\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1})^T \mathbf{y}_k^* \mathbf{d}_k - \mathbf{d}_k^T \mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1}) \mathbf{y}_k^*}{\max\{2c \|\mathbf{d}_k\| \|\mathbf{y}_k^*\|, |\mathbf{d}_k^T \mathbf{y}_k^*|\}} & \text{if } k \geq 1 \\ -\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1}) & \text{if } k = 0 \end{cases} \quad (11)$$

这里 $\mathbf{y}_k^* = \mathbf{y}_k + \gamma_k^* \mathbf{s}_k$, $\mathbf{y}_k = \mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1}) - \mathbf{g}^a(\mathbf{x}_k, \epsilon_k)$, $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, $\mathbf{d}_k$ 是搜索方向, 常数 $c > 0$, 并且

$$\gamma_k^* = \frac{(\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1}) + \mathbf{g}^a(\mathbf{x}_k, \epsilon_k))^T \mathbf{s}_k + 2(F^a(\mathbf{x}_k, \epsilon_k) - F^a(\mathbf{x}_{k+1}, \epsilon_{k+1}))}{\|\mathbf{s}_k\|^2}$$

但相应算法并未建立. 本文将在参数公式(11) 基础上, 结合非单调 Armijo-type 线搜索, 提出以下包含函数

值信息的求解非光滑问题的修正 HS 共轭梯度法. 通过研究, 该算法不仅自动满足充分下降条件, 而且搜索方向属于信赖域. 在适当条件下, 证明了新算法全局收敛. 初步的数值实验也表明新算法在求解非光滑无约束凸优化问题方面是有效的.

算法 1

步骤 1 取 $\mathbf{x}_0 \in \mathbb{R}^n$, $\gamma_0 \in (0, 1)$, $\sigma \in (0, 1)$, $c > 0$, $s > 0$, $\lambda > 0$, $\rho \in [0, 1]$, $Q_0 = 1$, $R_0 = F^a(\mathbf{x}_0, \epsilon_0)$, $\mathbf{d}_0 = -\mathbf{g}^a(\mathbf{x}_0, \epsilon_0)$, $\gamma \in (0, 1)$, 令 $k = 0$.

步骤 2 若 $\|\mathbf{g}^a(\mathbf{x}_k, \epsilon_k)\| < \gamma$, 则停止; 否则进行步骤 3.

步骤 3 选取一个标量 ϵ_{k+1} 满足 $0 < \epsilon_{k+1} < \epsilon_k$, 利用以下非单调 Armijo-type 线搜索计算步长 t_k :

$$F^a(\mathbf{x}_k + t_k \mathbf{d}_k, \epsilon_{k+1}) - R_k \leq \sigma t_k \mathbf{g}^a(\mathbf{x}_k, \epsilon_k)^\top \mathbf{d}_k \quad (12)$$

其中 $t_k = s2^{-i_k}$, $i_k \in \{0, 1, 2, \dots\}$.

步骤 4 令 $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$, 若 $\|\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1})\| < \gamma$, 则停止; 否则进行步骤 5.

步骤 5 利用以下公式校正 R_k :

$$Q_{k+1} = \rho Q_k + 1 \quad R_{k+1} = \frac{\rho Q_k R_k + F^a(\mathbf{x}_k + t_k \mathbf{d}_k, \epsilon_{k+1})}{Q_{k+1}} \quad (13)$$

步骤 6 利用(12) 式计算搜索方向 $\mathbf{d}_{k+1}$.

步骤 7 令 $k := k + 1$, 转步骤 2.

算法中的非单调 Armijo-type 线搜索(12) 选自文献[11], ρ 是控制非单调程度的参数. 当 $\rho = 0$ 时, 线搜索(12) 就是普通的单调 Armijo-type 线搜索.

引理 1 对所有 $k \in \mathbb{N} \cup \{0\}$, 以下充分下降条件成立:

$$\mathbf{g}^a(\mathbf{x}_k, \epsilon_k)^\top \mathbf{d}_k = -\|\mathbf{g}^a(\mathbf{x}_k, \epsilon_k)\|^2 \quad (14)$$

证 当 $k = 0$ 时,

$$\mathbf{d}_0 = -\mathbf{g}^a(\mathbf{x}_0, \epsilon_0)$$

(14) 式显然成立.

当 $k \geq 1$ 时, 由(11) 式, 有

$$\begin{aligned} \mathbf{d}_{k+1}^\top \mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1}) &= -\|\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1})\|^2 + \\ &\quad \left[\frac{\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1})^\top \mathbf{y}_k^* \mathbf{d}_k - \mathbf{d}_k^\top \mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1}) \mathbf{y}_k^*}{\max\{2c \|\mathbf{d}_k\| \|\mathbf{y}_k^*\|, |\mathbf{d}_k^\top \mathbf{y}_k^*|\}} \right]^\top \mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1}) = \\ &\quad -\|\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1})\|^2 \end{aligned}$$

(14) 式成立. 证毕.

引理 2 对所有 $k \in \mathbb{N} \cup \{0\}$, 以下信赖域性质成立:

$$\|\mathbf{d}_k\| \leq \left(1 + \frac{1}{c}\right) \|\mathbf{g}^a(\mathbf{x}_k, \epsilon_k)\| \quad (15)$$

证 由(11) 式得

$$\begin{aligned} \|\mathbf{d}_{k+1}\| &= \left\| -\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1}) + \frac{\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1})^\top \mathbf{y}_k^* \mathbf{d}_k - \mathbf{d}_k^\top \mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1}) \mathbf{y}_k^*}{\max\{2c \|\mathbf{d}_k\| \|\mathbf{y}_k^*\|, |\mathbf{d}_k^\top \mathbf{y}_k^*|\}} \right\| \leq \\ &\|\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1})\| + \frac{\|\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1})\| \|\mathbf{y}_k^*\| \|\mathbf{d}_k\| + \|\mathbf{d}_k\| \|\mathbf{g}^a(\mathbf{x}_{k+1}, \epsilon_{k+1})\| \|\mathbf{y}_k^*\|}{\max\{2c \|\mathbf{d}_k\| \|\mathbf{y}_k^*\|, |\mathbf{d}_k^\top \mathbf{y}_k^*|\}} \end{aligned}$$

因为

$$\max\{2c \|\mathbf{d}_k\| \|\mathbf{y}_k^*\|, |\mathbf{d}_k^\top \mathbf{y}_k^*|\} \geq 2c \|\mathbf{d}_k\| \|\mathbf{y}_k^*\|$$

所以

$$\begin{aligned} \|d_{k+1}\| &\leq \|g^a(x_{k+1}, \epsilon_{k+1})\| + \\ &\frac{\|g^a(x_{k+1}, \epsilon_{k+1})\| \|y_k^*\| \|d_k\| + \|d_k\| \|g^a(x_{k+1}, \epsilon_{k+1})\| \|y_k^*\|}{2c \|d_k\| \|y_k^*\|} \leq \\ &\left(1 + \frac{1}{c}\right) \|g^a(x_{k+1}, \epsilon_{k+1})\| \end{aligned}$$

证毕.

3 全局收敛性

算法 1 的全局收敛性的证明需要用到以下假设:

假设 A (i) 设目标函数 f 经 Moreau-Yosida 正则化后得到 F , 则存在常数 $M > 0$, 使得

$$\|\nabla^2 F(u_k)\| \leq M \quad \forall u_k \in [x_k, x_{k+1}], \forall k \in \mathbb{N} \quad (16)$$

(ii) F 有下界.

(iii) 序列 $\{\epsilon_k\}$ 收敛于 0.

根据文献 [11], 我们得到以下引理:

引理 3 若假设 A 成立, 针对算法 1 中的每一个 k , 有

$$F^a(x_k, \epsilon_k) \leq R_k \leq C_k$$

其中

$$C_k = \frac{1}{k+1} \sum_{i=0}^k F^a(x_i, \epsilon_i)$$

而且满足 Armijo-type 线搜索(12) 的步长 t_k 存在.

引理 4 若假设 A 成立, 序列 $\{(x_k, t_k)\}$ 由算法 1 产生, 假设

$$\epsilon_k = o(t_k^2 \|d_k\|^2)$$

则存在常数 $m > 0$, 使得步长 t_k 满足:

$$t_k \geq m \quad (17)$$

证 设步长 t_k 满足线搜索(12) 式. 下面用反证法证明.

假设 $t_k \rightarrow 0$, 则存在

$$t'_k = \frac{t_k}{2}$$

满足

$$F^a(x_k + t'_k d_k, \epsilon_{k+1}) - R_k > \sigma t'_k g^a(x_k, \epsilon_k)^T d_k$$

把引理 3 中的 $F^a(x_k, \epsilon_k) \leq R_k \leq C_k$ 代入上式, 可得

$$F^a(x_k + t'_k d_k) - F^a(x_k, \epsilon_k) > F^a(x_k + t'_k d_k, \epsilon_{k+1}) - R_k > \sigma t'_k g^a(x_k, \epsilon_k)^T d_k \quad (18)$$

由(8), (18) 式及假设 A, 并利用泰勒展开式, 有

$$\begin{aligned} \sigma t'_k g^a(x_k, \epsilon_k)^T d_k &< F^a(x_k + t'_k d_k, \epsilon_{k+1}) - F^a(x_k, \epsilon_k) \leq \\ &F(x_k + t'_k d_k) + \epsilon_{k+1} - F^a(x_k, \epsilon_k) \leq \\ &F(x_k + t'_k d_k) - F(x_k) + \epsilon_{k+1} = \\ &F(x_k) + t'_k d_k^T g(x_k) + \frac{\nabla F^2(\xi_k)(t'_k d_k)^2}{2} - F(x_k) + \epsilon_{k+1} = \\ &t'_k d_k^T g(x_k) + \frac{(t'_k)^2 d_k^T \nabla F^2(\xi_k) d_k}{2} + \epsilon_{k+1} \leq \\ &t'_k d_k^T g(x_k) + \frac{M}{2} (t'_k)^2 \|d_k\|^2 + \epsilon_{k+1} \end{aligned} \quad (19)$$

因为

$$\xi_k = x_k + \iota t'_k d_k \quad \iota \in (0, 1)$$

根据(10),(14) 式及 $\epsilon_{k+1} \leq \epsilon_k$, 由(19) 式可得

$$\begin{aligned} \frac{t_k}{2} = t'_k &\geq \frac{\sigma g^a(x_k, \epsilon_k)^T d_k - d_k^T g(x_k) - \frac{\epsilon_{k+1}}{t'_k}}{\frac{M}{2} \|d_k\|^2} > \\ &\left[\frac{(g^a(x_k, \epsilon_k) - g(x_k))^T d_k - (1 - \sigma) g^a(x_k, \epsilon_k)^T d_k - \frac{\epsilon_{k+1}}{t'_k}}{\|d_k\|^2} \right] \frac{2}{M} \geq \\ &\left[\frac{(1 - \sigma) \|g^a(x_k, \epsilon_k)\|^2 - \sqrt{\frac{2\epsilon_k}{\lambda}} \|d_k\| - \frac{\epsilon_k}{t'_k}}{\|d_k\|^2} \right] \frac{2}{M} \end{aligned}$$

再由 $\epsilon_k = o(t_k^2 \|d_k\|^2)$ 和(15) 式, 上式变为

$$\begin{aligned} t'_k &> \left[\frac{(1 - \sigma) \|g^a(x_k, \epsilon_k)\|^2}{\|d_k\|^2} - \frac{o(t_k)}{\sqrt{\lambda}} - o(t_k) \right] \frac{2}{M} \geq \\ &\left[\frac{(1 - \sigma)}{(1 + \frac{1}{c})^2} - \frac{o(t_k)}{\sqrt{\lambda}} - o(t_k) \right] \frac{2}{M} \end{aligned} \quad (20)$$

不等式两边除以 t_k 并同取极限, 得

$$\frac{1}{2} \geq \lim_{k \rightarrow \infty} \left(\frac{2(1 - \sigma)}{\left(1 + \frac{1}{c}\right)^2 M} - o(t_k) \right) \frac{1}{t_k} = +\infty$$

矛盾. 故假设不成立. 命题得证.

根据引理 1-4 证明算法 1 的全局收敛性:

定理 1 设引理 4 的条件都成立, 则

$$\lim_{k \rightarrow \infty} \|g(x_k)\| = 0$$

且序列 $\{x_k\}$ 的聚点即为问题(1) 的最优解.

证 先证明

$$\lim_{k \rightarrow \infty} \|g^a(x_k, \epsilon_k)\| = 0 \quad (21)$$

用反证法. 假设(21) 式不成立, 则存在正数 $\gamma_0 > 0$ 和 $k_0 > 0$, 使得

$$\|g^a(x_k, \epsilon_k)\| \geq \gamma_0 \quad \forall k > k_0 \quad (22)$$

利用(12),(14),(17),(22) 式, 得

$$F^a(x_{k+1}, \epsilon_{k+1}) - R_k \leq \sigma t_k g^a(x_k, \epsilon_k)^T d_k = -\sigma t_k \|g^a(x_k, \epsilon_k)\|^2 \leq -\sigma m \gamma_0 \quad \forall k > k_0$$

由(13) 式可得:

$$\begin{aligned} R_{k+1} &= \frac{\rho Q_k R_k + F^a(x_k + t_k d_k, \epsilon_{k+1})}{Q_{k+1}} \leq \\ &\frac{\rho Q_k R_k + R_k - \sigma m \gamma_0}{Q_{k+1}} = \\ &\frac{R_k (\rho Q_k + 1) - \sigma m \gamma_0}{Q_{k+1}} = \\ &\frac{R_k Q_{k+1} - \sigma m \gamma_0}{Q_{k+1}} = \end{aligned}$$

$$R_k - \frac{\sigma m \gamma_0}{Q_{k+1}}$$

(23)

由假设 A 知 F 有下界, 因而 $F^a(\mathbf{x}, \epsilon)$ 有下界. 又由引理 3 知对所有的 k , $F^a(\mathbf{x}_k, \epsilon_k) \leq R_k$, 因此不难知道 R_k 有下界. 因而根据(23) 式, 有

$$\sum_{k=k_0}^{\infty} \frac{\sigma m \gamma_0}{Q_{k+1}} < \infty$$

(24)

但由于根据 Q_{k+1} 的定义(13) 式容易得到 $Q_{k+1} \leq k + 2$, 因此

$$\sum_{k=k_0}^{\infty} \frac{\sigma m \gamma_0}{Q_{k+1}} \geq \sum_{k=k_0}^{\infty} \frac{\sigma m \gamma_0}{k + 2} = +\infty$$

这与(24) 式矛盾. 因此(21) 式成立.

另一方面, 由(10) 式有

$$\| \mathbf{g}^a(\mathbf{x}_k, \epsilon_k) - \mathbf{g}(\mathbf{x}_k) \| \leq \sqrt{\frac{2\epsilon_k}{\lambda}}$$

因而结合假设 A(III), 即可得到

$$\lim_{k \rightarrow \infty} \| \mathbf{g}(\mathbf{x}_k) \| = 0$$

(25)

令 $\mathbf{x}^*$ 是序列 $\{\mathbf{x}_k\}$ 的聚点, 则存在子序列 $\{\mathbf{x}_k\}_K$ 满足

$$\lim_{k \in K, k \rightarrow \infty} \mathbf{x}_k = \mathbf{x}^*$$

(26)

根据(3) 式以及 $F(\mathbf{x})$ 的性质易知

$$\mathbf{g}(\mathbf{x}_k) = \frac{\mathbf{x}_k - \mathbf{h}(\mathbf{x}_k)}{\lambda}$$

因此由(25) 和(26) 式可知

$$\mathbf{x}^* = \mathbf{h}(\mathbf{x}^*)$$

成立, 故 $\mathbf{x}^*$ 是问题(1) 的最优解. 证毕.

4 数值结果

为了考查本文所提算法对大规模非光滑优化问题的数值表现, 本文对该方法进行数值实验, 并将数值结果与 LMBM(limited memory bundle method)方法^[12] 结果进行比较. 测试问题选自文献[12], 所采用的程序是在文献[12] 的程序基础上利用 Fortran 语言修改而得, 操作系统为 Windows 7, 软件平台为 Fortran90, 处理器为 Pentium Dual E5800 3.20 GHz, 内存为 2.0 GB, 参数选取如下: $s=\lambda=1$, $\rho=0.5$, $\sigma=0.8$, $\epsilon_k=\frac{1}{(k+2)^2}$; 终止条件为 $\| \mathbf{g}^a(\mathbf{x}, \epsilon) \| \leq 10^{-15}$, 迭代次数 $N_i > 10^4$ 时算法失效.

表 1 列出了测试问题名称及初始点, 表 2 则列出两种方法数值实验的结果, 其中 $\mathbf{x}_0$ 表示初始点, N_i 为迭代次数, N_f 表示函数值的计算次数, $f(\mathbf{x})$ 为近似最优点的函数值.

表 1 测试问题

序号	测试问题	x_0
1	Generalization of MAXQ	$\left(1, 2, \dots, \frac{n}{2}, -\frac{n}{2+1}, \dots, -n\right)$
2	Chained LQ	$(-0.5, -0.5, \dots)$
3	Number of active faces	$(1, 1, \dots)$
4	Nonsmoothgeneralization of Brown function 2	$(1, 0, \dots)$
5	Chained crescent	$(-1.5, 2, \dots)$
6	Chained crescent 2	$(1, 0, \dots)$

表 2 数值结果

序号	维数	MHS	LMBM
		$N_i/N_f/f(x)$	$N_i/N_f/f(x)$
1	3 000	239/4971/6.955 650 353 842 63 $\times 10^{-8}$	94 144/96 680/1.134 215 823 070 18 $\times 10^{-5}$
	6 000	253/5272/6.726 138 641 769 14 $\times 10^{-8}$	245 066/250 174/565628885146208 $\times 10^{-5}$
	20 000	272/5697/6.433 270 442 007 30 $\times 10^{-8}$	1 419 971/1 450 334/1.563 108 595 262 89 $\times 10^{-4}$
	50 000	286/5991/6.599 447 300 897 05 $\times 10^{-8}$	4 999 996/5 000 000/5.776 229 182 256 93 $\times 10^2$
2	3 000	11/55/2.793 039 194 309 49 $\times 10^{-6}$	275/1373/−4.241 179 314 955 21 $\times 10^5$
	6 000	11/56/2.793 504 856 062 3 $\times 10^{-6}$	314/1847/−8.483 777 056 063 14 $\times 10^5$
	20 000	12/59/4.656 385 164 455 58 $\times 10^{-6}$	572/3 665/−2.828 279 647 277 23 $\times 10^6$
	50 000	22/79/1.164 132 379 879 49 $\times 10^{-5}$	582/2 998/−7.070 920 052 968 46 $\times 10^6$
3	3 000	60/1156/1.461 567 777 860 02 $\times 10^{-6}$	1576/1 577/1.554 743 000 887 91 $\times 10^{-10}$
	6 000	65/1261/2.116 923 651 896 82 $\times 10^{-6}$	3085/3 086/1.548 308 148 238 18 $\times 10^{-10}$
	20 000	75/1471/3.700 806 349 456 93 $\times 10^{-6}$	10 090/100 91/4.714 519 884 484 46 $\times 10^{-10}$
	50 000	82/1618/5.888 908 804 180 25 $\times 10^{-6}$	25 070/25 101/1.463 351 438 952 38 $\times 10^{-9}$
4	3 000	12/58/2.793 039 473 613 58 $\times 10^{-6}$	542/5145/6.227 226 749 432 59 $\times 10^{-8}$
	6 000	12/59/2.793 505 135 412 75 $\times 10^{-6}$	951/8278/8.841 189 430 325 71 $\times 10^{-4}$
	20 000	13/62/4.656 385 630 092 92 $\times 10^{-6}$	746/7446/2.711 495 095 809 97 $\times 10^{-8}$
	50 000	23/82/1.164 132 496 293 23 $\times 10^{-5}$	1293/10995/7.265 354 767 529 77 $\times 10^{-7}$
5	3 000	15/122/4.325 183 552 422 73 $\times 10^{-7}$	199/591/3.844 034 424 105 30 $\times 10^{-9}$
	6 000	16/124/8.641 743 994 841 63 $\times 10^{-7}$	135/177/5.965 544 059 455 29 $\times 10^{-2}$
	20 000	17/127/1.439 280 695 292 31 $\times 10^{-6}$	198/279/2.417 616 116 197 71 $\times 10^2$
	50 000	26/146/1.798 779 903 361 41 $\times 10^{-6}$	391/725/4.756 805 394 023 90 $\times 10^{-9}$
6	3 000	13/61/2.618 998 658 832 08 $\times 10^{-6}$	757/6162/2.669 729 354 299 88 $\times 10^{-3}$
	6 000	13/62/2.619 173 781 637 23 $\times 10^{-6}$	1163/11352/1.068 332 415 193 75 $\times 10^{-3}$
	20 000	23/82/8.730 995 607 830 75 $\times 10^{-6}$	1353/12333/3.886 503 606 216 34 $\times 10^{-4}$
	50 000	42/121/1.091 389 351 826 07 $\times 10^{-5}$	4657/61720/8.010 423 366 732 19 $\times 10^{-4}$

从表 2 容易看出, 新算法不仅能够有效求解 50 000 维的非光滑问题, 而且相比求解非光滑问题的传统 LMBM 算法优势明显. 由此可以认为, 新算法对大规模非光滑无约束凸优化问题的求解是有效的.

参考文献:

[1] POLAK E, RIBIERE G. Note Sur La Convergence de Directions Conjugées [J]. Revue Francaise Informatique Recherche Operationnelle, 3e Année, 1969, 16(3): 35—43.

[2] POLYAK B T. The Conjugate Gradient Method in Extreme Problem [J]. USSR Computational Mathematics and Mathematical Physics, 1969, 9(4): 94—112.

[3] HESTENES M R, STIEFEL E. Method of Conjugate Gradient for Solving Linear Equation [J]. Journal of Research of the National Bureau of Standards, 1952, 49(6): 409—436.

[4] LIU Y, STOREY C. Efficient Generalized Conjugate Gradient Algorithms, Part 1: Theory [J]. Journal of Optimization Theory and Application, 1991, 69(1): 129—137.

[5] FLETCHER R, REEVES C. Function Minimization by Conjugate Gradients [J]. The Computer Journal, 1964, 7(12): 149—154.

[6] 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(1): 86—96.

[7] BONNAN J F, GILBERT J C, LEMARECHAL C, et al. A Family of Variable Metric Proximal Methods [J]. Mathe-

- mathematical Programming, 1995, 68(1): 15—47.
- [8] CORREA R, LEMARECHAL C. Convergence of some Algorithms for Convex Minimization [J]. Mathematical Programming, 1993, 62(1): 261—273.
- [9] HIRIART-URRUTY J B, LEMARECHAL C. Convex Analysis and Minimization Algorithms II [M]. Berlin: Springer, 1993.
- [10] FUKUSHIMA M, QI L. A Global and Superlinearly Convergent Algorithm for Nonsmooth Convex Minimization [J]. SIAM Journal on Optimization, 1996, 6(4): 1106—1120.
- [11] ZHANG H, HARGER W W. A Nonmonotone Line Search Technique and Its Application to Unconstrained Optimization [J]. SIAM Journal on Optimization, 2006, 14(4): 1043—1056.
- [12] HAARALA M, MIETTINEN K, MAKELA M M. New Limited Memory Bundle Method for Large-Scale Nonsmooth Optimization [J]. Optimization Methods and Software, 2004, 19(6): 673—692.

A Modified Hestenes-Stiefel Conjugate Gradient Algorithm for Large-Scale Nonsmooth Optimization Problems

LI Yong¹, YUAN Gong-lin²

1. School of Mathematics and Statistics, Baise University, Baise Guangxi 533000, China;

2. College of Mathematics and Information Science, Guangxi University, Nanning 530004, China

Abstract: This paper gives a modified Hestenes-Stiefel conjugate gradient algorithm for large-scale nonsmooth optimization problems, using the Moreau-Yosida regulation approach in combination with the nonmonotone line search technique. The search direction of the algorithm not only possesses the sufficient descent property but also belongs to a trust region. The new algorithm has the global convergence under proper conditions. A preliminary numerical experiment shows that the algorithm proposed herein is more effective than the normal method for large-scale non-smooth unconstrained convex optimization problems.

Key words: nonsmooth; large-scale; conjugate gradient method; sufficient descent condition; global convergence

责任编辑 张 枸

