第 39 卷第 2 期　　河 北 科 技 大 学 学 报　　Vol.39,No.2
2018 年 4 月　　Journal of Hebei University of Science and Technology　　Apr. 2018

文章编号:1008-1542(2018)02-0142-07　　doi:10.7535/hbkd.2018yx02007

求解非光滑优化问题的修正 HS 三项共轭梯度法

黎　勇,王松华

(百色学院数学与统计学院,广西百色　533000)

摘　要: 为了提高大规模非光滑优化问题的求解效率,克服其他方法存储需求大、算法复杂等缺点,提出求解非光滑优化问题的一种修正 HS 共轭梯度算法。在经典 HS 三项共轭梯度法的基础上提出一种新的搜索方向,并利用 Moreau-Yosida 正则化技术和 Armijo-type 线搜索技术进行设计。新算法满足充分下降条件,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。初步的数值实验表明新算法在求解非光滑无约束优化问题方面比 LMBM 方法更有效。新算法不仅具有较好的收敛性质,而且数值表现良好,为更加高效地求解非光滑优化问题提供了新的方法。

关键词: 最优化;非光滑优化;共轭梯度法;充分下降条件;信赖域;全局收敛性

中图分类号: O224　　**MSC(2010)主题分类:** 49J25　　**文献标志码:** A

A modified three-term HS conjugate gradient method for solving nonsmooth minimizations

LI Yong, WANG Songhua

(School of Mathematics and Statistics, Baise University, Baise, Guangxi 533000, China)

Abstract: To improve the efficiency for large-scale nonsmooth optimization problems and overcome the large storage requirements and complex computation of other algorithms, a modified HS conjugate gradient algorithm for nonsmooth optimization problems is proposed. A new search direction based on the classical HS conjugate gradient method is given, then the Moreau-Yosida regularization technique and the Armijo-type line search technique are used to design the algorithm. The sufficient descent condition and the trust region are satisfied for this algorithm. Under suitable conditions, the global convergence of the new algorithm is proved. The preliminary numerical experiments show that the new algorithm is more efficient than the LMBM method for nonsmooth unconstrained optimization problems. The presented algorithm is efficiently for solving nonsmooth optimization problems since it has good convergence property and good numerical performance.

Keywords: optimization; nonsmooth optimization; conjugate gradient method; sufficient descent; trust region; global convergence

收稿日期:2017-08-21;修回日期:2017-12-10;责任编辑:张　军

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

第一作者简介:黎　勇(1973—),男,广西靖西人,副教授,硕士,主要从事最优化理论与方法方面的研究。

E-mail:liyong3922@163.com

黎勇,王松华.求解非光滑优化问题的修正 HS 三项共轭梯度法[J].河北科技大学学报,2018,39(2):142-148.

LI Yong, WANG Songhua. A modified three-term HS conjugate gradient method for solving nonsmooth minimizations[J]. Journal of Hebei University of Science and Technology, 2018, 39(2): 142-148.

非线性共轭梯度法被广泛应用于最优化问题的求解方面，该算法简单、存储需求小。在长期的研究过程中，学者们设计出了 PRP[1-2]，HS[3]，LS[4]，FR[5] 等一批经典的共轭梯度算法。研究结果表明，充分下降条件是共轭梯度法的一个重要性质，对算法的全局收敛等性质的影响非常关键[6-10]。近年来共轭梯度法的研究新成果不断出现，如 WEI 等[11] 提出的 WYL 共轭梯度法，该算法不仅收敛性质较好，而且数值方面的表现也比较出色。YAO 等[12] 对 WYL 算法做了进一步推广，提出了一种修正的 HS 共轭梯度法，其公式定义如下：

$$\boldsymbol{\beta}_k^{MHS}=\frac{\boldsymbol{g}_k^{\mathrm{T}}\left(\boldsymbol{g}_k-\dfrac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{g}_{k-1}\|}\boldsymbol{g}_{k-1}\right)}{(\boldsymbol{g}_k-\boldsymbol{g}_{k-1})^{\mathrm{T}}\boldsymbol{d}_{k-1}}。$$

ZHANG 等[13] 为了保证搜索方向自动具有充分下降性质，提出了一种三项 PRP 共轭梯度法，该算法的搜索方向定义如下：

$$\boldsymbol{d}_k=\begin{cases}-\boldsymbol{g}_k, & \text{if } k=0,\\ -\boldsymbol{g}_k+\boldsymbol{\beta}_k^{PRP}\boldsymbol{d}_{k-1}-\boldsymbol{\theta}_k\boldsymbol{y}_{k-1}, & \text{if } k\geqslant 1,\end{cases}$$

式中 $\boldsymbol{\theta}_k=\dfrac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\|\boldsymbol{g}_{k-1}\|^2}$，$\boldsymbol{y}_{k-1}=\boldsymbol{g}_k-\boldsymbol{g}_{k-1}$，迭代公式为 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+t_k\boldsymbol{d}_k$，$t_k$ 是步长。这样建立的算法能自动满足充分下降条件，因而具有良好的收敛性质。

共轭梯度法目前主要被应用于光滑优化问题的求解，特别是对大规模优化问题，其效果相对于牛顿法、拟牛顿法、信赖域法等优化方法更为明显。针对大规模非光滑凸优化问题，如金融、图形图像、生物工程、最优控制、信息技术等方面研究，能否充分利用共轭梯度法的特点来有效解决非光滑优化问题，研究成果还比较少。文献[14]和文献[15]分别提出了新的修正 Polak-Ribière-Polyak 和修正 Hestenes-Stiefel 共轭梯度法，这些算法通过利用 Moreau-Yosida 正则化技术，并结合 Armijo-type 线搜索，能够有效解决非光滑凸优化问题。胡亚萍[16] 结合 Moreau-Yosida 正则化技术、邻近点算法提出求解无约束非光滑凸优化问题的修正 LS 算法和修正 WYL 算法。本研究将在文献[12]和文献[13]基础上，利用文献[14－16]的思想设计一种新的修正 Hestenes-Stiefel 共轭梯度算法，讨论其在求解大规模非光滑凸优化问题中的收敛性质和数值表现。

1　预备知识

为方便后面的讨论，给出非光滑凸优化问题的相关理论知识。

针对非光滑无约束优化问题：

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

其中目标函数 $f:\mathbf{R}^n\to\mathbf{R}$ 是非光滑凸函数。

对目标函数 f 施行 Moreau-Yosida 正则化后得到新函数 $F:\mathbf{R}^n\to\mathbf{R}$，

$$F(x)=\min_{z\in\mathbf{R}^n}\{f(z)+\frac{1}{2\mu}\|z-x\|^2\},\tag{2}$$

这里 $\|\cdot\|$ 指欧几里得范数，参数 $\mu>0$。令 $Q(z,x)=f(z)+\dfrac{1}{2\mu}\|z-x\|^2$，且令 $h(x)=\arg\min_z Q(z,x)$，由于 $Q(z,x)$ 对每一个固定的 x 强凸，故 $h(x)$ 唯一。因此 $F(x)=f(h(x))+\dfrac{1}{2\mu}\|h(x)-x\|^2$。

由文献[17—19]可知 $F(x)$ 具有以下性质：

i) $F(x)$ 是有界凸函数，且处处可微。令：

$$g(x)=\nabla F(x)=\frac{x-h(x)}{\mu},\tag{3}$$

则 $g(x)$ Lipschitz 连续，即

$$\|g(x)-g(y)\|\leqslant\frac{1}{\mu}\|x-y\|,\quad \forall x,y\in\mathbf{R}^n。\tag{4}$$

ii) x 是问题(1)的最优解，当且仅当 $\nabla F(x)=0$ 时，有 $h(x)=x$。

从上述讨论容易知道，只要求出 $\arg\min_z Q(z,x)$ 的最优解，$F(x)$ 和 $g(x)$ 就可以确定下来。然而求出 $\arg\min_z Q(z,x)=h(x)$ 的精确解是非常困难的，甚至是不可能完成的任务。因此在实际计算中不可能用精

确的 $h(x)$ 去定义 $F(x)$ 和 $g(x)$。文献[20]给出了如下近似替换的思路与方法。

对 $\forall x\in\mathbf{R}^n,\varepsilon>0$,存在向量 $\boldsymbol{h}^\alpha(x,\varepsilon)\in\mathbf{R}^n$,使得:

$$f(\boldsymbol{h}^\alpha(x,\varepsilon))+\frac{1}{2\mu}\|\boldsymbol{h}^\alpha(x,\varepsilon)-x\|^2\leqslant F(x)+\varepsilon,\tag{5}$$

当 ε 充分小时可用 $\boldsymbol{h}^\alpha(x,\varepsilon)$ 近似定义 $F(x)$ 和 $g(x)$ 表示如下:

$$\boldsymbol{F}^\alpha(x,\varepsilon)=f(\boldsymbol{h}^\alpha(x,\varepsilon))+\frac{1}{2\mu}\|\boldsymbol{h}^\alpha(x,\varepsilon)-x\|^2,\tag{6}$$

$$\boldsymbol{g}^\alpha(x,\varepsilon)=\frac{x-\boldsymbol{h}^\alpha(x,\varepsilon)}{\mu}。\tag{7}$$

从文献[20]可知,$\boldsymbol{F}^\alpha(x,\varepsilon)$ 和 $\boldsymbol{g}^\alpha(x,\varepsilon)$ 具有以下性质。

命题1 假设式(5)—式(7)成立,则有:

$$F(x)\leqslant\boldsymbol{F}^\alpha(x,\varepsilon)\leqslant F(x)+\varepsilon,\tag{8}$$

$$\|\boldsymbol{h}^\alpha(x,\varepsilon)-h(x)\|\leqslant\sqrt{2\mu\varepsilon},\tag{9}$$

$$\|\boldsymbol{g}^\alpha(x,\varepsilon)-g(x)\|\leqslant\sqrt{2\varepsilon/\mu}。\tag{10}$$

命题说明,当 ε 充分小,$\boldsymbol{F}^\alpha(x,\varepsilon)$ 和 $\boldsymbol{g}^\alpha(x,\varepsilon)$ 可以无限接近 $F(x)$ 和 $g(x)$。其证明详见文献[20]。

2 修正HS三项共轭梯度法

在文献[12-15]基础上,笔者提出一种求解非光滑无约束凸优化问题(1)的修正HS三项共轭梯度法,其搜索方向定义如下:

$$\boldsymbol{d}_{k+1}=\begin{cases}-\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})+\dfrac{\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})^{\mathrm{T}}\boldsymbol{y}_k^*\boldsymbol{d}_k-\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\boldsymbol{y}_k^*}{\max\{2c\|\boldsymbol{d}_k\|\|\boldsymbol{y}_k^*\|,|\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{y}_k|\}}, & \text{if } k\geqslant1,\\ -\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1}), & \text{if } k=0,\end{cases}\tag{11}$$

式中:$\boldsymbol{d}_k$ 是搜索方向,且 $\boldsymbol{y}_k^*=\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})-\dfrac{\|\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\|}{\|\boldsymbol{g}^\alpha(\boldsymbol{x}_k,\varepsilon_k)\|}\boldsymbol{g}^\alpha(\boldsymbol{x}_k,\varepsilon_k)$,$\boldsymbol{y}_k=\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})-\boldsymbol{g}^\alpha(\boldsymbol{x}_k,\varepsilon_k)$,常数 $c>0$。在此基础上提出新算法。

算法1

Step1 选择初始值 $x_0\in\mathbf{R}^n$,取 $\sigma\in(0,1),c>0,s>0,\mu>0,\boldsymbol{d}_0=-\boldsymbol{g}^\alpha(x_0,\varepsilon_0),\gamma\in[0,1]$,令 $k=0$。

Step2 如果 $\|\boldsymbol{g}^\alpha(\boldsymbol{x}_k,\varepsilon_k)\|<\gamma$,则停止运算,否则进行下一步。

Step3 选择一个 ε_{k+1},满足 $0<\varepsilon_{k+1}<\varepsilon_k$,通过非单调Armijo-type线搜索[21]确定步长 t_k:

$$\boldsymbol{F}^\alpha(\boldsymbol{x}_k+t_k\boldsymbol{d}_k,\varepsilon_{k+1})-\boldsymbol{F}^\alpha(\boldsymbol{x}_k,\varepsilon_k)\leqslant\sigma t_k\boldsymbol{g}^\alpha(\boldsymbol{x}_k,\varepsilon_k)^{\mathrm{T}}\boldsymbol{d}_k,\tag{12}$$

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

Step4 定义 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+t_k\boldsymbol{d}_k$,如果 $\|\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\|<\gamma$,则停止运算;否则,进行下一步。

Step5 利用式(11)计算 $\boldsymbol{d}_{k+1}$。

Step6 令 $k:=k+1$,返回Step2。

引理1 对所有 $k\in\mathbf{N}\cup\{0\}$,有:

$$\boldsymbol{g}^\alpha(\boldsymbol{x}_k,\varepsilon_k)^{\mathrm{T}}\boldsymbol{d}_k=-\|\boldsymbol{g}^\alpha(\boldsymbol{x}_k,\varepsilon_k)\|^2。\tag{13}$$

证明

当 $k=0$ 时,$\boldsymbol{d}_0=-\boldsymbol{g}^\alpha(x_0,\varepsilon_0)$,命题成立。

当 $k\geqslant1$ 时,式(11)两边同时乘以 $\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})$ 得:

$$\begin{aligned}\boldsymbol{d}_{k+1}^{\mathrm{T}}\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})=&-\|\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\|^2+\\&\left[\frac{\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})^{\mathrm{T}}\boldsymbol{y}_k^*\boldsymbol{d}_k-\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\boldsymbol{y}_k^*}{\max\{2c\|\boldsymbol{d}_k\|\|\boldsymbol{y}_k^*\|,|\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{y}_k|\}}\right]^{\mathrm{T}}\boldsymbol{g}^\alpha(\boldsymbol{x}_{k+1},\varepsilon_{k+1})=\\&-\|\boldsymbol{g}^\alpha(x_{k+1},\varepsilon_{k+1})\|^2。\end{aligned}$$

命题成立。

综上所述,命题对所有 $k\in\mathbf{N}\cup\{0\}$ 都成立。

引理 2　对 $\forall k \in \mathbf{N} \cup \{0\}$,有:

$$\|\boldsymbol{d}_k\| \leqslant (1+\frac{1}{c})\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)\|。\tag{14}$$

证明　由于 $\max\{2c\|\boldsymbol{d}_k\|\|\boldsymbol{y}_k^*\|,|\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{y}_k|\} \geqslant 2c\|\boldsymbol{d}_k\|\|\boldsymbol{y}_k^*\|$,所以式(11)两边同取范数得:

$$\begin{aligned}\|\boldsymbol{d}_{k+1}\| &= \left\| -\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1}) + \frac{\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})^{\mathrm{T}}\boldsymbol{y}_k^*\boldsymbol{d}_k - \boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\boldsymbol{y}_k^*}{\max\{2c\|\boldsymbol{d}_k\|\|\boldsymbol{y}_k^*\|,|\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{y}_k|\}} \right\| \leqslant \\
&\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\| + \left\| \frac{\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})^{\mathrm{T}}\boldsymbol{y}_k^*\boldsymbol{d}_k - \boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\boldsymbol{y}_k^*}{\max\{2c\|\boldsymbol{d}_k\|\|\boldsymbol{y}_k^*\|,|\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{y}_k|\}} \right\| \leqslant \\
&\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\| + \frac{\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\|\|\boldsymbol{y}_k^*\|\|\boldsymbol{d}_k\| + \|\boldsymbol{d}_k\|\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\|\|\boldsymbol{y}_k^*\|}{\max\{2c\|\boldsymbol{d}_k\|\|\boldsymbol{y}_k^*\|,|\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{y}_k|\}} \leqslant \\
&\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\| + \frac{\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\|\|\boldsymbol{y}_k^*\|\|\boldsymbol{d}_k\| + \|\boldsymbol{d}_k\|\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\|\|\boldsymbol{y}_k^*\|}{2c\|\boldsymbol{d}_k\|\|\boldsymbol{y}_k^*\|} \leqslant \\
&(1+\frac{1}{c})\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})\|。\end{aligned}$$

引理 1 说明本研究所提出的算法不需要线搜索即可满足充分下降条件。引理 2 说明搜索方向具有有界性,即搜索方向具有信赖域性质。

3　全局收敛性

以下是算法 1 的全局收敛性讨论中需要用到的假设。

假设 A　i)对 $\forall \xi_k \in [\boldsymbol{x}_k,\boldsymbol{x}_{k+1}]$和 $\forall k \in \mathbf{N}$,存在一个正的常数 ρ,使得:

$$\|\nabla^2 F(\xi_k)\| \leqslant \rho,\tag{15}$$

其中 F 是目标函数 f 经 Moreau-Yosida 正则化后所得到的函数。

ii)函数 F 有下界。

iii)序列$\{\varepsilon_k\}$收敛于 0。

引理 3　设序列$\{\boldsymbol{x}_k\}$由算法 1 产生,假设 A 成立,如果 $\varepsilon_k = o(t_k^2\|\boldsymbol{d}_k\|^2)$,则对充分大的 k,存在一个正的常数 l,使

$$t_k \geqslant l。\tag{16}$$

证明　反证法。

假设结论不成立,则存在 $t'_k = \frac{t_k}{2}$,使式(12)不成立,即:

$$\boldsymbol{F}^{\alpha}(\boldsymbol{x}_k + t'_k\boldsymbol{d}_k,\varepsilon_{k+1}) - \boldsymbol{F}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k) > \sigma t'_k\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)^{\mathrm{T}}\boldsymbol{d}_k。$$

根据式(8)和假设 A,并利用泰勒展开式,有:

$$\begin{aligned}\sigma t'_k\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)^{\mathrm{T}}\boldsymbol{d}_k &< \boldsymbol{F}^{\alpha}(\boldsymbol{x}_k + t'_k\boldsymbol{d}_k,\varepsilon_{k+1}) - \boldsymbol{F}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k) \leqslant \\
&F(\boldsymbol{x}_k + t'_k\boldsymbol{d}_k) + \varepsilon_{k+1} - \boldsymbol{F}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k) \leqslant \\
&F(\boldsymbol{x}_k + t'_k\boldsymbol{d}_k) - F(\boldsymbol{x}_k) + \varepsilon_{k+1} = \\
&F(\boldsymbol{x}_k) + t'_k\boldsymbol{d}_k^{\mathrm{T}}g(\boldsymbol{x}_k) + \frac{\nabla F^2(\xi_k)(t'_k\boldsymbol{d}_k)^2}{2} - F(\boldsymbol{x}_k) + \varepsilon_{k+1} = \\
&t'_k\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}_k) + \frac{(t'_k)^2\boldsymbol{d}_k^{\mathrm{T}}\ \nabla F^2(\xi_k)\boldsymbol{d}_k}{2} + \varepsilon_{k+1} \leqslant \\
&t'_k\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}_k) + \frac{\rho}{2}(t'_k)^2\|\boldsymbol{d}_k\|^2 + \varepsilon_{k+1},\end{aligned}\tag{17}$$

式中 $\xi_k = \boldsymbol{x}_k + at'_k\boldsymbol{d}_k$, $a \in (0,1)$。从式(17)出发,根据式(10)、式(13)和式(14)及 $\varepsilon_{k+1} \leqslant \varepsilon_k$,结合$\varepsilon_k = o(t_k^2\|\boldsymbol{d}_k\|^2)$,有:

$$\frac{t_k}{2}=t'_k \geqslant \frac{\sigma \boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)^{\mathrm{T}}\boldsymbol{d}_k-\boldsymbol{d}_k^{\mathrm{T}}g(\boldsymbol{x}_k)-\frac{\varepsilon_{k+1}}{t'_k}}{\frac{\rho}{2}\|\boldsymbol{d}_k\|^2}>$$

$$\left[\frac{(\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)-g(\boldsymbol{x}_k))^{\mathrm{T}}\boldsymbol{d}_k-(1-\sigma)\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)^{\mathrm{T}}\boldsymbol{d}_k-\frac{\varepsilon_{k+1}}{t'_k}}{\|\boldsymbol{d}_k\|^2}\right]\frac{2}{\rho}\geqslant$$

$$\left[\frac{(1-\sigma)\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)\|^2-\sqrt{\frac{2\varepsilon_k}{\mu}}\|\boldsymbol{d}_k\|-\frac{\varepsilon_k}{t'_k}}{\|\boldsymbol{d}_k\|^2}\right]\frac{2}{\rho}=$$

$$\left[\frac{(1-\sigma)\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)\|^2}{\|\boldsymbol{d}_k\|^2}-\frac{o(t_k)}{\sqrt{\mu}}-o(t_k)\right]\frac{2}{\rho},$$

两边同时除以 t_k，并取极限得：

$$\frac{1}{2}\geqslant\lim_{k\to\infty}\left(\frac{2(1-\sigma)}{\rho(1+\frac{1}{c})^2}-o(t_k)\right)\frac{1}{t_k}\geqslant+\infty,$$

矛盾，故结论成立。

定理 1　设序列$\{x_k\}$，$\{t_k\}$由算法 1 产生，假设 A 成立，$\varepsilon_k=o(t_k^2\|\boldsymbol{d}_k\|^2)$，则$\lim\limits_{k\to\infty}\|\boldsymbol{g}(\boldsymbol{x}_k)\|=0$，且序列$\{\boldsymbol{x}_k\}$的聚点是问题(1)的最优解。

证明　要证明$\lim\limits_{k\to\infty}\|g(\boldsymbol{x}_k)\|=0$，只需证明

$$\lim_{k\to\infty}\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)\|=0 \tag{18}$$

即可，采用反证法。假设式(18)不成立，则存在正的常数 η_0 和 k_0，使得对 $\forall k>k_0$，有：

$$\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)\|\geqslant\eta_0。 \tag{19}$$

从式(12)出发，利用式(13)、式(16)、式(19)，得：

$$\boldsymbol{F}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1})-\boldsymbol{F}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)\leqslant\sigma t_k\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)^{\mathrm{T}}\boldsymbol{d}_k=-\sigma t_k\|\boldsymbol{g}^{\alpha}(x_k,\varepsilon_k)\|^2\leqslant-\sigma l\eta_0,\forall k>k_0,$$

所以有：$\sum\limits_{k>k_0}(\boldsymbol{F}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)-\boldsymbol{F}^{\alpha}(\boldsymbol{x}_{k+1},\varepsilon_{k+1}))\geqslant\sum\limits_{k>k_0}\sigma l\eta_0$，由此易知，当$k\to\infty$时，$\boldsymbol{F}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)\to\infty$，与假设 A 中条件 ii) 矛盾。因此式(18) 成立。

下面证明第 2 个结论。

由式(10) 有：

$$\|\boldsymbol{g}^{\alpha}(\boldsymbol{x}_k,\varepsilon_k)-g(\boldsymbol{x}_k)\|\leqslant\sqrt{\frac{2\varepsilon_k}{\mu}},$$

根据假设 A 中条件 iii)，可得：

$$\lim_{k\to\infty}\|g(\boldsymbol{x}_k)\|=0。 \tag{20}$$

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

$$\lim_{k\in K,k\to\infty}\boldsymbol{x}_k=\boldsymbol{x}^*。 \tag{21}$$

根据式(3)以及 $F(x)$的性质易知 $g(\boldsymbol{x}_k)=\dfrac{\boldsymbol{x}_k-h(x_k)}{\mu}$，因此由式(20)和式(21)可知 $\boldsymbol{x}^*=h(\boldsymbol{x}^*)$成立，故 $\boldsymbol{x}^*$ 是问题(1)的最优解。

4　数值结果

笔者通过数值试验来考查本研究所提出的算法解决非光滑优化问题的有效性。试验的计算机环境为 Windows7+Fortran90，内存 2.0 GB，各参数选取为 $s=\mu=1$，$\sigma=0.8$，$\varepsilon_k=1/(k+2)^2$；终止条件为 $\|\boldsymbol{g}^{\alpha}(\boldsymbol{x},\varepsilon)\|\leqslant10^{-15}$或者迭代次数 $N_i>10^4$。测试问题选自文献[12]，测试程序是利用 Fortran 语言在文献[12]所提供程序基础上修改而得，新方法的实验结果将与文献[22]所提出的 LMBM(limited memory bundle method)方法从 N_i，N_f 和 $f(x)$等几方面进行比较。其中 N_i 为迭代次数，N_f 表示函数值的计算次

数，$f(x)$ 为近似最优点的函数值。

测试问题名称及初始点在表 1 中列出，2 种方法数值试验的结果在表 2 中列出，其中 Problem 表示测试问题的名称，x_0 表示初始点，Dim 表示维数。

表 1　测试问题

Tab.1　Test problems

No.	Problem	x_0
1	Generalization of MAXQ	$(1,2,\cdots,n/2,\cdots-(n/2+1),\cdots,-n)$
2	Generalization of MXHILB	$(1,1,\cdots)$
3	Chained LQ	$(-0.5,-0.5,\cdots)$
4	Number of active faces	$(1,1,\cdots)$
5	Nonsmooth generalization of Brown function 2	$(1,0,\cdots)$
6	Chained Mifflin2	$(-1,-1,\cdots)$

表 2　数值结果

Tab.2　Numerical results

No.	Dim	MHS $N_i/N_f/f(x)$	LMBM $N_i/N_f/f(x)$
1	1 000	133/2 735/2.760×10^{-7}	21 492/22 259/6.710×10^{-6}
	3 000	149/2 990/1.485×10^{-6}	94 144/96 680/1.134×10^{-5}
	5 000	167/3 381/3.107×10^{-7}	191 470/196 034/3.450×10^{-5}
	10 000	152/3 060/ 2.021×10^{-5}	512 415/523 351/5.835×10^{-5}
2	1 000	67/1 086/4.280×10^{-2}	441/861/6.166×10^{-3}
	3 000	88/1 392/$3.013\ 5\times10^{-7}$	209/579/5.872×10^{-2}
	5 000	104/1 564/1.238×10^{-11}	1 258/2 487/3.521×10^{-2}
	10 000	107/1 681/1.797×10^{-12}	7 027/7 810/5.116×10^{-2}
3	1 000	4/30/1.129×10^{-8}	300/1 824/-1.413×10^{5}
	3 000	7/51/-1.096×10^{-2}	275/1 373/-4.241×10^{5}
	5 000	2/11/1.201×10^{-4}	365/2 198/-7.070×10^{5}
	10 000	3/152/ -9.655×10^{-3}	376/2 281/-1.414×10^{6}
4	1 000	35/583/2.458×10^{-4}	523/569/1.377×10^{-14}
	3 000	51/778/4.076×10^{-8}	1 576/1 577/1.555×10^{-10}
	5 000	50/782/1.599×10^{-3}	2 585/2 586/1.213×10^{-10}
	10 000	59/945/9.789×10^{-9}	5 069/5 073/5.384×10^{-10}
5	1 000	4/30/2.258×10^{-8}	467/3 873/4.058×10^{-9}
	3 000	7/40/2.118×10^{-7}	542/5 245/6.227×10^{-8}
	5 000	2/12/5.273×10^{-9}	453/4 073/1.080×10^{-8}
	10 000	3/20/1.083×10^{-10}	736/7 453/2.522×10^{-8}
6	1 000	4/30/-2.498×10^{4}	1 254/7 355/-7.065×10^{4}
	3 000	7/51/-7.498×10^{4}	411/2 353/-2.121×10^{5}
	5 000	2/11/-1.250×10^{5}	219 /7 382/-3.535×10^{5}
	10 000	3/125/ -2.500×10^{5}	267/743/ -7.070×10^{5}

分别对 6 个测试函数的 4 种不同维数进行测试比较，从表 2 可以看出，本研究所提出的新算法与求解非光滑问题的传统 LMBM 算法相比较优势明显。综上所述，新算法不仅具有较好的收敛性质，而且数值表现良好，因此可以认为它能够有效求解非光滑优化问题。

5　结　论

非线性共轭梯度法算法简单，存储需求小，是一种重要的优化方法，HS 方法是其中被广泛讨论和运用的一种方法。本文针对非光滑优化问题的求解，在经典 HS 共轭梯度法的基础上，通过利用 Moreau-Yosida 正则化技术和文献[21]所提出的 Armijo-type 线搜索技术，设计了一种修正 HS 共轭梯度算法。笔者所设计的算法不仅可以自动具有充分下降性，而且相应的搜索方向属于信赖域。在适当条件下，证明了新算法是全局收敛的。初步的数值试验表明新算法在求解非光滑无约束优化问题方面是有效的。

在接下来的工作中，还有很多相关问题值得做进一步的思考和讨论，如新算法的收敛速度，各种参数的不同选择对算法效率的影响等，以及新算法在解决金融、图形图像、生物工程、最优控制、信息技术等实际问题所蕴含的非光滑问题时的效果，这些都有待在以后工作中继续检验和测试。

参考文献/References：

[1] POLAK E, RIBIERE G. Note surla convergence de method de directions conjugees[J]. Rev Francaise informat Recherche Opèrationelle, 3e Annèe, 1969, 16(16): 35-43.

[2] POLYAK B T. The conjugate gradient method in extremal problems[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 systems[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 Applications, 1991, 69(1): 129-137.

[5] FLETCHER R, REEVES C. Function minimization by conjugate gradients[J]. Computer Journal, 1964, 7(2): 149-154.

[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] HU Y F, STOREY C. Global convergence result for conjugate gradient methods[J]. Journal of Optimization Theory & Applications, 1991, 71(2): 399-405.

[8] ZENG X W, GUO Y L, LI Q Q. Global convergence of the Polak-Ribière-Polyak conjugate gradient methods with inexact line search for nonconvex unconstrained optimization problems[J]. Mathematics of Computation, 2008, 77(264): 2173-2193.

[9] TOUATI-AHMED D, STOREY C. Efficient hybrid conjugate gradient techniques[J]. Journal of Optimization Theory & Applications, 1990, 64 (2): 379-397.

[10] ALBAALI M. Descent property and global convergence of the Flecher-Reves method with inexact line search[J]. Ima Journal of Numerical Analysis, 1985, 5(1): 121-124.

[11] 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.

[12] YAO S W, 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.

[13] ZHANG L, ZHOU W J, LI D H. A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence[J]. Ima Journal of Numerical Analysis, 2006, 26(4): 629-640.

[14] YUAN G L, WEI Z X, LI G Y. A modified Polak-Ribière-Polyak conjugate gradient algorithm for nonsmooth covex programs[J]. Journal of Computational and Applied Mathematics, 2014, 255(1): 86-96.

[15] YUAN G L, MENG Z H, LI Y. A modified Hestenes and Stiefel conjugate gradient algorithm for large-scale nonsmooth minimizations and nonlinear equations[J]. Journal of Optimization Theory and Applications, 2016, 168(1): 129-152.

[16] 胡亚萍.非线性单调方程组和非光滑优化问题的算法研究[D].上海：华东理工大学，2014.
HU Yaping. Numerical Methods for Nonlinear Monotone Equations and Nonsmooth Optimization Problems[D]. Shanghai: East China University of Science and Technology, 2014.

[17] BONNAN J F, GILBERT J C, LEMARECHAL C, et al. A family of variable metric proximal methods[J]. Mathematical Programming, 1995, 68(1): 15-47.

[18] CORREA R, LEMARÉCHAL C. Convergence of some algorithms for convex minimization[J]. Mathematical Programming, 1993, 62 (1/2/3): 261-273.

[19] HIRIART-URRUTY J B, LEMARECHAL C. Convex Analysis and Minimization Algorithms Ⅱ[M]. Berlin: Springer, 1993.

[20] FUKUSHIMA M, QI L. A globally and superlinearly convergent algorithm for nonsmooth convex minimization[J]. Siam Journal on Optimization, 1996, 6(4): 1106-1120.

[21] ZHANG H C, HARGER W W. A nonmonotone line search technique and its application to unconstrained optimization[J]. Siam Journal on Optimization, 2006, 14(4): 1043-1056.

[22] HAARALA M, MIETTINEN K, MÄKELÄ M M. New limited memory bundle method for large-scale nonsmooth optimization[J]. Optimization Methods & Software, 2004, 19(6): 673-692.