天津科技大学学报
Journal of Tianjin University of Science & Technology

DOI:10.13364/j.issn.1672-6510.20190283

求解非光滑问题的修正 HS 共轭梯度法

胡亚萍，王玉杰，刘丽英
（天津科技大学理学院，天津 300457）

摘　要：结合 Moreau-Yosida 正则化和非单调线搜索技术，提出一种求解非光滑问题的修正 HS 共轭梯度算法. 推导出搜索方向自动满足充分下降条件，证明该算法在适当条件下具有全局收敛性. 数值算例验证了该算法能够高效地处理非光滑极小化问题.

关键词：非光滑问题；共轭梯度算法；Moreau-Yosida 正则化；全局收敛性
中图分类号：O224　　**文献标志码**：A　　**文章编号**：1672-6510（0000）00-0000-00

A Modified HS Conjugate Gradient Algorithm for Nonsmooth Minimizations

HU Yaping，WANG Yujie，LIU Liying
（College of Science，Tianjin University of Science & Technology，Tianjin 300457，China）

Abstract：By using Moreau-Yosida regularization and nonmonotone line search，a new modified HS conjugate gradient algorithm is proposed for nonsmooth minimization. The search direction of the proposed algorithm has sufficient descending property and the global convergence property of the proposed algorithm is justified under suitable assumptions. Numerical results show that the new algorithm is effective for solving nonsmooth minimizations.

Key words：nonsmooth problem；conjugate gradient algorithm；Moreau-Yosida regularization；global convergence

考虑无约束优化问题 $\min\{f(x)\,|\,x\in\mathbb{R}^n\}$，其中 $f:\mathbb{R}^n\to\mathbb{R}$ 为非光滑凸函数. 非光滑问题中的目标函数是连续不可微函数，传统的优化算法不能直接用于求解该问题. 与非光滑凸优化问题紧密相关的是目标函数 Moreau-Yosida 正则化[1]，正则化函数 $F(x)$ 是定义在整个空间 $\mathbb{R}^n$ 上的可微的凸函数，并且与原非光滑优化问题拥有相同的解集合. 求解非光滑优化问题的常用算法有 Bundle 法和信赖域法[2-4]. 近年来，Yuan 等[5-6]和 Hu[7-8]提出的梯度类算法在求解非光滑问题时表现较好. 其中文献[6]基于 BFGS 修正技术提出的修正 PRP 共轭梯度法需要较大的存储空间和计算量，它每步迭代时的计算量和内存需求均大于共轭梯度类算法. 本文结合 Moreau-Yosida 正则化和非单调线搜索技术，提出了修正的 HS 共轭梯度算法求解非光滑优化问题. 新算法具有满足共轭性条件、自动具有充分下降性、给出近似参数选取方式、克服存储需求大与算法复杂等特点. 数值结果表明，与文献[6]的算法相比，新算法具有收敛速度快、精度高的优点.

1　算　法

记 $p(x)=\operatorname{argmin}\{\theta(z)\,|\,z\in\mathbb{R}^n\}$，且定义 $\theta(z)=f(z)+\|z-x\|^2/(2\lambda)$，由于 $\theta(z)$ 是一个强凸函数，极小值点 $p(x)$ 存在且唯一. 于是非光滑凸函数 $f(x)$ 的 Moreau-Yosida 正则函数 $F(x)$ 表示为

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

正则化函数 $F(x)$ 是连续可微的凸函数，但同时注意到 $F(x)$ 未必二次可微. $F(x)$ 在点 x 处的梯度为

收稿日期：2019–11–27；**修回日期**：2020–05–01
基金项目：天津市教委科研计划项目（2019KJ232）
作者简介：胡亚萍（1984—），女，河北邯郸人，讲师，huyaping@tust.edu.cn

$g(x)=\nabla F(x)=(x-p(x))/\lambda$. 然而 $\theta(z)$ 的极小值点 $p(x)$ 一般很难甚至不可能精确求解，这便不能直接利用 $p(x)$ 的精确值来确定函数值 $F(x)$ 和梯度值 $g(x)$. 但是对任意 $x\in\mathbb{R}^n$ 和任意的近似参数 $\varepsilon>0$，存在近似值 $p^{\alpha}(x,\varepsilon)\in\mathbb{R}^n$ 满足

$$f(p^{\alpha}(x,\varepsilon))+\frac{1}{2\lambda}\|p^{\alpha}(x,\varepsilon)-x\|^2\leqslant F(x)+\varepsilon \tag{2}$$

于是，可以利用 $p^{\alpha}(x,\varepsilon)$ 来确定 $F(x)$ 和 $g(x)$ 的近似值，即

$$F^{\alpha}(x,\varepsilon)=f(p^{\alpha}(x,\varepsilon))+\frac{1}{2\lambda}\|p^{\alpha}(x,\varepsilon)-x\|^2 \tag{3}$$

$$g^{\alpha}(x,\varepsilon)=\left[x-p^{\alpha}(x,\varepsilon)\right]/\lambda \tag{4}$$

一些用于求解近似极小值点 $p^{\alpha}(x,\varepsilon)$ 的算法见文献[9]，近似值 $F^{\alpha}(x,\varepsilon)$ 和 $g^{\alpha}(x,\varepsilon)$ 满足下面的性质[9]：

$$F(x)\leqslant F^{\alpha}(x,\varepsilon)\leqslant F(x)+\varepsilon \tag{5}$$

$$\|p^{\alpha}(x,\varepsilon)-p(x)\|\leqslant\sqrt{2\lambda\varepsilon} \tag{6}$$

$$\|g^{\alpha}(x,\varepsilon)-g(x)\|\leqslant\sqrt{2\varepsilon/\lambda} \tag{7}$$

本文提出修正 HS 共轭梯度算法，简记为 MHS 算法，令

$$d_{k+1}=\begin{cases}-g_{k+1}^{\alpha} & k=0\\ -g_{k+1}^{\alpha}+\dfrac{g_{k+1}^{\alpha}y_kd_k-d_k^{\mathrm T}g_{k+1}^{\alpha}y_k}{\max\{r\|d_k\|\|y_k\|,d_k^{\mathrm T}y_k,\|g_k^{\alpha}\|^2\}} & k\geqslant1\end{cases} \tag{8}$$

其中，g_{k+1}^{α} 是正则函数 $F(x)$ 在点 x_{k+1} 处的梯度近似值 $g^{\alpha}(x_{k+1},\varepsilon_{k+1})$，$y_k=g^{\alpha}(x_{k+1},\varepsilon_{k+1})-g^{\alpha}(x_k,\varepsilon_k)$.

算法 MHS 的步骤如下：

步骤 0：令 $k=0$，给定初始点 $x_0\in\mathbb{R}^n$，$s>0$，$\xi\in(0,1)$，$\sigma\in(0,1)$，$\lambda>0$，$\rho>0$，$E_0=1$，一个严格下降的正序列 $\{\tau_k\}$ 满足 $\tau_0\leqslant1$ 且 $\lim\limits_{k\to0}\tau_k=0$，$\varepsilon_0=\tau_0$，$J_0=F^{a}(x_0,\varepsilon_0)$，$d_0=-g^{\alpha}(x_0,\varepsilon_0)$.

步骤 1：若 $\|g^{\alpha}(x_k,\varepsilon_k)\|<\xi$，则停止.

步骤 2：选取 ε_{k+1} 满足

$$0<\varepsilon_{k+1}\leqslant\min\left\{\tau_k,\tau_k\|g^{\alpha}(x_k,\varepsilon_k)\|^2\right\}$$

由非单调 Armijo-型线搜索确定步长 α_k：

$$F^{\alpha}(x_k+\alpha_kd_k,\varepsilon_{k+1})-J_k\leqslant\sigma\alpha_kg^{\alpha}(x_k,\varepsilon_k)^{\mathrm T}d_k \tag{9}$$

其中，$\alpha_k=s2^{-i_k}$，$i_k\in\{1,2,\cdots\}$.

步骤 3：令 $x_{k+1}=x_k+\alpha_kd_k$. 若 $\|g^{\alpha}(x_{k+1},\varepsilon_{k+1})\|<\xi$，则算法停止.

步骤 4：由下面公式更新 J_{k+1}

$$E_{k+1}=\rho E_k+1,$$
$$J_{k+1}=\left[\rho E_kJ_k+F^{\alpha}(x_k+\alpha_kd_k,\varepsilon_{k+1})\right]/E_{k+1} \tag{10}$$

步骤 5：由式(8)计算搜索方向 d_{k+1}.

步骤 6：令 $k=k+1$，转步骤 1.

2　全局收敛性

本节讨论修正 HS 共轭梯度算法用于求解非光滑凸优化问题时的收敛性. 为此，需要文献[5–7]中的假设条件.

假设 A. 序列 $\{\boldsymbol V_k\}$ 有界，即存在常数 $M>0$ 使得

$$\|\boldsymbol V_k\|\leqslant M,\quad\forall k \tag{11}$$

其中矩阵 $\boldsymbol V_k\in\partial_Bg(x_k)$.

假设 B. 正则化函数 F 有下界.

引理 1 由式(8)的定义，搜索方向满足性质

$$g^{\alpha}(x_k,\varepsilon_k)^{\mathrm T}d_k=-\|g^{\alpha}(x_k,\varepsilon_k)\|^2 \tag{12}$$

$$\|d_k\|\leqslant\left(1+\frac{2}{\gamma}\right)\|g^{\alpha}(x_k,\varepsilon_k)\| \tag{13}$$

证明： 当 $k=0$ 时，$d_0=-g^{\alpha}(x_0,\varepsilon_0)$，式(11)、式(12)显然成立.

当 $k\geqslant1$ 时

$$d_{k+1}^{\mathrm T}g^{\alpha}(x_{k+1},\varepsilon_{k+1})=-\|g^{\alpha}(x_{k+1},\varepsilon_{k+1})\|^2+\left[\frac{g_{k+1}^{\alpha\ \mathrm T}y_kd_k-d_k^{\mathrm T}g_{k+1}^{\alpha}y_k}{\max\{\gamma\|d_k\|\|g_k^{\alpha}\|,-d_k^{\mathrm T}g_k^{\alpha},\|g_k^{\alpha}\|\}}\right]^{\mathrm T}g_{k+1}^{\alpha}=-\|g^{\alpha}(x_{k+1},\varepsilon_{k+1})\|^2$$

故(12)成立.

$$\|d_{k+1}\|=\left\|-g_{k+1}^{\alpha}+\frac{g_{k+1}^{\alpha\ \mathrm T}y_kd_k-d_k^{\mathrm T}g_{k+1}^{\alpha}y_k}{\max\{\gamma\|d_k\|\|g_k^{\alpha}\|,-d_k^{\mathrm T}g_k^{\alpha},\|g_k^{\alpha}\|\}}\right\|\leqslant\|g_{k+1}^{\alpha}\|+\frac{\|g_{k+1}^{\alpha}\|\|y_k\|\|d_k\|-\|d_k\|\|g_{k+1}^{\alpha}\|\|y_k\|}{\max\{\gamma\|d_k\|\|g_k^{\alpha}\|,-d_k^{\mathrm T}g_k^{\alpha},\|g_k^{\alpha}\|\}}\leqslant\left(1+\frac{2}{\gamma}\right)\|g^{\alpha}(x_{k+1},\varepsilon_{k+1})\|$$

故(13)成立. 证毕.

根据假设 B 和修正 HS 算法中的步骤 5，提出下面的引理. 引理表明该搜索是适定的，证明方法与文献[10]中的引理 1 类似，故省略.

引理 2 若假设 B 成立. 序列 $\{x_k\}$ 由算法 MHS 产生，则 $F^{a}(x_k,\varepsilon_k)\leqslant J_k\leqslant C_k$ 对每一个 k 成立，其中 $C_k=\dfrac{1}{k+1}\sum\limits_{i=0}^{k}F^{a}(x_i,\varepsilon_i)$. 另外，存在 α_k 满足线搜索中 Armijo 条件.

由假设 A，类似于文献[5]中的引理 4.2，可以得

到下面的引理.

引理 3 若假设 A 成立. 序列$\{(x_k,\varepsilon_k)\}$由算法 MHS 产生. 假设$\varepsilon_k=o(\alpha_k^2\|d_k\|^2)$成立. 则存在常数$m_0>0$，满足$\alpha_k\geqslant m_0$.

定理 1 若假设 A，假设 B 和引理 3 的条件成立，序列$\{x_k\}$由算法 MHS 产生，则有$\lim\limits_{k\to\infty}\|g(x_k)\|=0$，且序列$\{x_k\}$的每一个聚点都是非光滑凸优化问题(1)的最优解.

证明：先用反证法证明$\lim\limits_{k\to\infty}\|g^\alpha(x_k,\varepsilon_k)\|=0$. 假设存在常数$\epsilon_0>0$和$k_0>0$使得$\|g^\alpha(x_k,\varepsilon_k)\|\geqslant\epsilon_0$对所有的$k>k_0$成立. 由式(5)和假设 B，知$F^a(x_k,\varepsilon_k)$有下界. 结合引理 2，得到$J_k$有下界，且

$$\sum_{k>k_0}(J_k-J_{k+1})<\infty \tag{14}$$

另一方面，由式(9)和引理 3，有

$$J_k-F^a(x_{k+1},\varepsilon_{k+1})\geqslant-\sigma\alpha_k g^a(x_k,\varepsilon_k)^{\mathrm{T}}d_k=$$
$$\sigma\alpha_k\|g^a(x_k,\varepsilon_k)\|^2\geqslant\sigma m_0\epsilon_0,\ \forall k>k_0$$

因此，上式结合式(10)可推出

$$\sum_{k>k_0}(J_k-J_{k+1})=$$
$$\sum_{k>k_0}\frac{E_{k+1}J_k-\rho E_kJ_k-F^a(x_{k+1},\varepsilon_{k+1})}{E_{k+1}}=$$
$$\sum_{k>k_0}\frac{J_k-F^a(x_{k+1},\varepsilon_{k+1})}{1+\rho+\cdots+\rho^{k+1}}\geqslant\sum_{k>k_0}\frac{\sigma m_0\epsilon_0}{k+2}=+\infty$$

这与式(14)矛盾. 所以$\lim\limits_{k\to\infty}\|g^\alpha(x_k,\varepsilon_k)\|=0$.

由$\{\varepsilon_k\}$的定义和式(7)，有$\lim\limits_{k\to\infty}\|g(x_k)\|=0$. 令$x^*$是序列$\{x_k\}$的一个聚点，不妨设存在一个子列$\{x_k\}_K$，使得$\lim\limits_{k\to\infty,k\in K}x_k=x^*$.

由正则化函数$F(x)$的定义，有

$$g(x_k)=\frac{(x_k-p(x_k))}{\lambda}$$

式中令$k\to\infty$，有$x^*=p(x^*)$成立. 因此x^*是非光滑优化问题(1)的最优解. 证毕.

3 数值实验

算法 MHS、MPRP[6]和 BT[11]的数值结果见表 1. 非光滑测试函数信息可参考文献[5]的表 1. 在实验中，取参数$s=\lambda=1$，$\rho=0.75$，$\sigma=0.9$，$\varepsilon_k=1/(k+1)^2$，终止准则为$\|g^a(x,\varepsilon)\|\leqslant10^{-5}$. 表 1 中$f(x)$表示算法终止时的函数值；$f_{\mathrm{ops}}(x)$表示目标函数的最优值.

从表 1 中迭代次数、函数值计算次数和算法终止时的函数值三方面综合来看，修正 HS 共轭梯度算法对求解非光滑问题是有效的.

表 1 不同算法的数值结果

Tab. 1 Numerical results of different algorithms

测试函数的编号	算法总迭代次数			函数值的计算次数			$f(x)$			$f_{\mathrm{ops}}(x)$
	MHS	MPRP	BT	MHS	MPRP	BT	MHS	MPRP	BT	
1	31	46	79	34	48	88	4.17×10^{-7}	7.09×10^{-7}	0.13×10^{-11}	0
2	9	11	24	11	13	27	2.75×10^{-5}	6.74×10^{-5}	0.94×10^{-6}	0
3	9	12	13	10	14	16	1.952 225	1.952 225	1.952 225	1.952 224 5
4	3	2	13	7	6	21	2.000 047	2.000 098	2.0	2.0
5	5	4	9	7	6	13	–2.999 991	–2.999 866	–3.0	–3
6	10	10	12	12	12	17	7.200 000	7.200 011	7.200 009	7.20
7	3	2	10	4	3	11	–1.414 213 53	–1.414 214	–1.414 214	–1.414 213 6
8	3	4	49	6	6	74	–0.998 438 2	–0.991 981 5	–1.0	–1.0
9	10	20	6	11	23	13	–0.999 973 5	–0.999 992 5	–1.0	–1.0
10	7	—	—	10	—	—	–7.999 998	—	—	–8
11	11	28	22	14	58	32	–43.999 90	–43.999 86	–43.999 98	–44
12	21	33	29	25	91	30	22.600 167	22.600 23	–22.600 16	22.600 162

4 结 语

非光滑优化问题是最优化理论与方法的重要分支，其求解也是优化领域的难题之一. 本文结合 Moreau-Yosida 正则化和非单调线搜索技术提出了非线性修正 HS 共轭梯度算法用于求解非光滑优化问题. 在适当条件下，证明了该算法具有全局收敛性. 数值结果表明新算法在求解非光滑优化问题方面是有效的.

参考文献：

［1］ Hiriart-Urruty J B，Lemmaréchal C. Convex Analysis and Minimization Algorithms［M］. Berlin：Springer Press，1993.

［2］ Monjezi N H，Nobakhtian S. A new infeasible proximal bundle algorithm for nonsmooth nonconvex constrained optimization[J]. Computational Optimization and Applications，2019，74：443–480.

[3] Sheng Z，Yuan G. An effective adaptive trust region algorithm for nonsmooth minimization[J]. Computational Optimization and Applications，2018，71：251–271.

[4] 沈洁，胡盼，李函阳，等. 一种求解非光滑无约束凸规划的混合束方法[J]. 吉林师范大学学报：自然科学版，2019，40(2)：58–62.

[5] Yuan G，Wei Z. The Barzilai and Borwein gradient method with nonmonotone line search for nonsmooth convex optimization problems[J]. Mathematical Modelling & Analysis，2012，17(2)：203–216.

[6] Yuan G，Wei Z，Li G. A modified Polak-Ribière-Polyak conjugate gradient algorithm for nonsmooth convex programs[J]. Journal of Computational and Applied Mathematics，2014(255)：86–96.

[7] Hu Y P. Multivariate spectral gradient algorithm for nonsmooth convex optimization problems[J]. Mathematical Problems in Engineering，2015，145323：1–7.

[8] 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，2015.

[9] Fukushima M，Qi L. A globally and superlinearly convergent algorithm for nonsmooth convex minimization [J]. Siam Journal on Optimization，1996，6(4)：1106–1120.

[10] Zhang H，Hager W. A nonmonotone line search technique and its application to unconstrained optimization [J]. SIAM Journal on Optimization，2004，14(4)：1043–1056.

[11] Schramm H，Zowe J. A version of the bundle idea for minimizing a nonsmooth function：Conceptual idea，convergence analysis，numerical results[J]. SIAM Journal on Optimization，1992，2(1)：121–152.

责任编辑：周建军