A Modified Three-Term Conjugate Gradient Algorithm for Large-Scale Nonsmooth Convex Optimization

Wujie Hu¹, Gonglin Yuan^{1,*} and Hongtruong Pham²

Abstract: It is well known that Newton and quasi-Newton algorithms are effective to small and medium scale smooth problems because they take full use of corresponding gradient function's information but fail to solve nonsmooth problems. The perfect algorithm stems from concept of 'bundle' successfully addresses both smooth and nonsmooth complex problems, but it is regrettable that it is merely effective to small and medium optimization models since it needs to store and update relevant information of parameter's bundle. The conjugate gradient algorithm is effective both large-scale smooth and nonsmooth optimization model since its simplicity that utilizes objective function's information and the technique of Moreau-Yosida regularization. Thus, a modified three-term conjugate gradient algorithm was proposed, and it has a sufficiently descent property and a trust region character. At the same time, it possesses the global convergence under mild assumptions and numerical test proves it is efficient than similar optimization algorithms.

Keywords: Conjugate gradient, large-scale, trust region, global convergence.

1 Introduction

Optimization model is one of the most important problems since it was widely used in all aspects [Liu, Yong, Gao et al. (2018); Liu, Chen, Ji et al. (2017); Melro and Jensen (2017)], where the nonsmooth mathematical model is more and more famous since it influences various fields and is a urgent as well as complex problem [Bonettini, Loris and Porta (2015); Li, Li and Yang et al. (2017); Shi, Tuan, Su et al. (2017); Zhang, Wu and Nguyen (2014)]. Writes and scholars paid much attention to this optimization program and obtained fruitful achievements. In Birge et al. [Birge, Qi and Wei (1998)], an algorithm on the basis of a classical proximal point was proposed and numeral test proves it is an effective algorithm because proximal point is helpful to simplify the process of calculating objective function's value. Sagara et al. [Sagara and Fukushima (2017)] and Yuan et al. [Yuan, Wei and Wang (2013)] proposed a trust region algorithm that mainly consider the search region and radius by relevant parameters. The concept of 'bundle' was proposed, and scholars introduced bundle algorithm into solving optimization model. The bundle algorithm sufficiently takes full use the bundle's information of the limiting variable conditions and has wide applications [Albaali, Spedicato and Maggioni (2014); Hall, Zimbro, Maduro et al. (2017); Luo, Tang and Zhou (2008); Kim, Um, Suh et al.

¹ College of Mathematics and Information Science, Guangxi University, Nanning, China.

² Thai Nguyen University of Economics and Business Administration, Thai Nguyen, Vietnam.

^{*} Corresponding Author: Gonglin Yuan. Email: yuangl0417@126.com.

(2016); Xu, Zhang, Du et al. (2016)]. To small and medium scale smooth and nonsmooth problems, bundle algorithm is effective but fails to solve large-scale problems. Conjugate gradient algorithm is more and more popular because of its simplicity and high efficiency [Dai, Liu, Zhang et al. (2017); Fasi, Langou and Robert (2016); Li, Zhang and Dong (2016); Wang and Zhu (2016); Yuan and Sheng (2017); Yuan, Meng and Li (2016); Yuan and Zhang (2015); Yuan, Wei and Li (2014); Yuan (2009); Yuan and Lu (2009)] and wide applications (see Baggio et al. [Baggio, Franceschini, Spiezia et al. (2017); Bernaschi, Bisson, Fantozzi et al. (2016); Janna (2016); Liu and Weng (2016); Sarkar (2016); Tarzanagh, Nazari and Peyghami (2016); Zhang and Liu (2016)]). For general conjugate gradient algorithm, it divides into two parts: one is inexactly or exactly line technique, the other is search direction, thus conjugate gradient algorithm is various since different formulas of search technique and search direction. Conjugate gradient algorithm addresses not only small and medium scale smooth and nonsmooth problems, but also solves large scale optimization model, which avoids calculating and store complex matrices and information of limiting variable boundary.

On the basis of above discussion, this paper proposes a new three-term conjugate gradient algorithm to solve large-scale nonsmooth optimization models. Objective algorithm has following perfect properties:

- 1. Proposed algorithm combines steepest descent method with conjugate gradient algorithm by introducing a parameter into search direction.
- 2. The search direction not only possesses a descent character but also has a trust region trait.
- 3. The objective algorithm has the ability of obtaining the global convergence.
- 4. The numeral test expresses the objective algorithm is superior than relevant optimization algorithms.

In this paper, next section expresses the model of nonsmooth convex optimization. The objective algorithm was introduced in Section 3 and its outstanding characters were expressed in Section 4. Numeral results were introduced in Section 5 and relevant references were listed in Section 6.

2 The model of nonsmooth convex functions

Consider general nonsmooth optimization model

$$\min_{x\in R^n} v(x) \tag{1}$$

where v(x) sometimes does not satisfy the property of smooth. It means that Newton

and quasi-Newton algorithms fail to solve it because of the nonsmooth property. Thus, a perfect technique was named 'Moreau-Yosida' regularization introduce this paper into solving the projective model, where the main idea of 'Moreau-Yosida' regularization is objective model exchanges a new constant with the variable, so objective model is translated into a smooth mathematical convex model. In general, the model's formula of 'Moreau-Yosida' regularization is

$$v^{M}(x) = \min_{t \in \mathbb{R}^{n}} \left(v(t) + \frac{\|t - x\|^{2}}{2\varphi} \right)$$
(2)

where φ is a positive constant, ||.|| presents the Euclidean norm. This paper denotes $v^{M}(x)$ as the 'Moreau-Yosida' regularization of objective model and similar papers proved it is outstanding [Wang and Zhu (2016)].

This paper assumes objective functions of (1) are convex and the objective model's regularization has solution of $\omega(x) = \arg \min v^{M}(t)$. On the basis of the mathematic knowledge, then obtain

$$\nabla v^{M}(x) = \frac{\omega(x) - x}{\varphi}$$
(3)

where $\nabla v^{M}(x)$ is the objective function's gradient function. In virtue of relevant papers, the $\nabla v^{M}(x)$ not only satisfies the smooth property, but also its gradient function is Lipschitz continuous. It is easy to find that (1) and (2) are equivalent because they have identical solution, thus this paper mainly considers the model of (2) since it is an optimization model that has perfect properties.

3 Three-term conjugate gradient algorithm

It is well known that a wonderful optimization algorithm not only has a perfect search direction but also outstanding exactly or inexactly line search technique. The three-term conjugate gradient algorithm has seen extensive study and obtained extremely good theoretical results. In view of Albaali et al. [Albaali (1985); Gilbert and Nocedal (1990); Touati-Ahmed and Storey (1990)] on conjugate gradient methods, the sufficient descent condition and trust region character are crucial to the global convergence. From this, a famous formula for the search direction d_{k+1} is expressed as follows: Zhang et al. [Zhang, Zhou and Li (2006)] proposed the following formula:

$$d_{k+1} = \begin{cases} -g_{k+1} + \frac{g_{k+1}^T y_k d_k - d_k^T g_{k+1} y_k}{g_k g_k^T} & k \ge 1 \\ -g_{k+1} & k = 0 \end{cases}$$
(4)

where $y_k = g_{k+1} - g_k$. In Nazareth [Nazareth (1977)], Nazareth proposed another search formula that is computed by

$$d_{k+1} = -y_k + \frac{y_k^T y_k}{y_k^T d_k} d_k + \frac{y_{k-1}^T y_k}{y_{k-1}^T d_{k-1}} d_{k-1},$$
(5)

where $y_k = g_{k+1} - g_k$, g_k is the gradient function value at the point x_k and $d_0 = d_{-1} = 0$.

3.2 Objective algorithm

In virtue of above discussion, this paper proposes a new three-term conjugate gradient algorithm and the formula is

$$d_{k+1} = \begin{cases} -\beta_{1} \nabla v^{M} (x_{k+1}) + \frac{\left(\nabla v^{M} (x_{k+1}) \right)^{T} y_{k} d_{k} - d_{k}^{T} \left(\nabla v^{M} (x_{k+1}) \right) y_{k}}{\max \left\{ \beta_{2} \| d_{k} \| \| y_{k} \|, \beta_{3} \| y_{k} \|^{2} \right\} + \nabla v^{M} (x_{k+1})^{2}} \\ -\nabla v^{M} (x_{k+1}) \end{cases}$$
(6)

where $\beta_i > 0$, i = 1, 2, 3 are positive constant. At the same time, we propose an inexactly line search technique of

$$v^{M}(x_{k+1}) \leq v^{M}(x_{k}) + \lambda \alpha_{k} \nabla v^{M}(x_{k})^{T} d_{k}$$

$$+ \alpha_{k} \min \left[-\lambda_{1} \nabla v^{M}(x_{k})^{T} d_{k}, \lambda \alpha_{k} \| d_{k} \|^{2} / 2 \right]$$

$$\text{where } \lambda, \gamma \in (0,1), \lambda_{1} \in (0,\lambda), \ \alpha_{k} \text{ is the largest number of } \{ \gamma^{k} \mid k = 0,1,2,\cdots \}.$$

$$(7)$$

Then we list the specific algorithm (denoted as Algorithm 2.1)

Step 1: (Initiation) Choose an initial point

$$x_{0}, \gamma, \lambda \in (0,1), \lambda_{1} \in (0,\lambda), \beta_{i} > 0, (i = 1,2,3\cdots),$$

positive constants $\varepsilon \in (0,1)$. Let $k = 0, d_0 = -\nabla v^M(x_0)$.

Step 2: If $\left\| \nabla v^{\scriptscriptstyle M}(x_{\scriptscriptstyle k}) \right\| \leq \varepsilon$, then stop.

Step 3: Calculation α_k in virtue of (7).

Step 4: Set new iteration point of $x_{k+1} = x_k + \alpha_k d_k$.

Step 5: Update the search direction by (6).

Step 6: If $\|\nabla v^{M}(x_{k+1})\| \leq \varepsilon$ holds, the algorithm stops, otherwise go to next step.

Step 7: Let k = k + 1 and go to Step 2.

4 Important characteristics and global convergence

Now we list the search direction has a trust region trait and a sufficient descent character without any conditions and this section lists them and expresses relevant proof.

Lemma 2.1: If the search direction d_k is generated by (6), then have

$$\left(\nabla v^{M}(\boldsymbol{x}_{k})\right)^{T}\boldsymbol{d}_{k} = -\beta_{1}\left\|\nabla^{M}(\boldsymbol{x}_{k})\right\|^{2}$$

$$\tag{8}$$

and

A Modified Three-Term Conjugate Gradient Algorithm

$$\left\|\boldsymbol{d}_{k}\right\| \leq \left(\boldsymbol{\beta}_{1} + \frac{2}{\boldsymbol{\beta}_{2}}\right) \left\|\nabla \boldsymbol{v}^{\boldsymbol{M}}\left(\boldsymbol{x}_{k}\right)\right\|$$

$$\tag{9}$$

holds.

To save space, we merely list but omit its proof. The formulas of (8) and (9) express the algorithm has a sufficient descent character and a trust region trait.

In the rest of this section, we express the specific process of proposed algorithm (algorithm 2.1) and prove the existence and necessity of step length α_k in inexactly line search.

Theorem 4.1. If Assumption (i-ii) are true, then has a constant α_k that satisfy the formula of (7).

proof: To prove relevant conclusion, we introduce following function $v(\alpha) = v^{M}(x_{k} + \alpha d_{k}) - v^{M}(x_{k}) - \lambda \alpha \nabla v^{M}(x_{k})^{T} d_{k}$ $-\alpha \min\left[-\lambda_{1} \nabla v^{M}(x_{k})^{T} d_{k}, \lambda \alpha \|d_{k}\|^{2} / 2\right]$

Considering sufficiently small parameter lpha , then has

$$v(\alpha) = v^{M}(x_{k} + \alpha d_{k}) - v^{M}(x_{k}) - \lambda \alpha \nabla v^{M}(x_{k})^{T} d_{k}$$
$$- \alpha \min\left[-\lambda_{1} \nabla v^{M}(x_{k})^{T} d_{k}, \lambda \alpha \|d_{k}\|^{2} / 2\right]$$
$$= \alpha (1 + \lambda_{1} - \lambda) \nabla v^{M}(x_{k})^{T} d_{k} + o(\alpha) < 0$$

It means that exists a constant $0 < \alpha_0 < 1$ satisfies the formula of $v(\alpha_0) < 0$. At the same time, the formula of v(0) = 0 holds. On the basis of the property of objective functions, we safely arrive a conclusion that exist a constant α_1 satisfies following formula:

$$v(\alpha_{1})=v^{M}(x_{k}+\alpha_{1}d_{k})-v^{M}(x_{k})-\lambda\alpha_{1}\nabla v^{M}(x_{k})^{T}d_{k}\\-\alpha_{1}\min\left[-\lambda_{1}\nabla v^{M}(x_{k})d_{k},\lambda\alpha_{1}d_{k}^{T}d_{k}/2\right]<0$$

thus the formula of

$$v^{M}(x_{k} + \alpha_{1}d_{k}) < v^{M}(x_{k}) + \lambda\alpha_{1}\nabla v^{M}(x_{k})^{T}d_{k} + \alpha_{1}\min[-\lambda_{1}\nabla v^{M}(x_{k})d_{k}, \lambda\alpha_{1}d_{k}^{T}d_{k}/2]$$

holds, it means that modified Armijo line search technique is well defined.

From above discussion, the Algorithm 2.1 has a sufficient descent trait and a trust region character and the line search technique is reasonable and necessary, thus we express the theorem of global convergence.

Assumption:

i) the level set $\pi = \{x \mid v^{M}(x) \leq v^{M}(x_{0})\}$ is bounded.

ii) function $v^{M}(x) \in C^{2}$, is bounded from below.

In virtue of the technique of 'Moreau-Yosida' regularization, we arrive a conclusion that function $\nabla v^{M}(x)$ is Lipschitz continuous. It means that exists a positive constant κ satisfy the formula of

$$\left\|\nabla v^{\scriptscriptstyle M}(x) - \nabla v^{\scriptscriptstyle M}(y)\right\| \le \kappa \|x - y\| \tag{10}$$

Theorem 4.2. If relevant assumptions hold, the relative sequences $\{x_k\}, \{\alpha_k\}, \{d_k\}, \{v^{M}(x_k)\}$ are generated in virtue of the Algorithm 2.1, then we obtain that

$$\lim_{k \to \infty} \left\| \nabla v^{M} \left(x_{k} \right) \right\| = 0 \tag{11}$$

proof: We will prove it by contradiction. Suppose above conclusion does not hold, it means that exist a constant \mathcal{E}_* and a corresponding index k_* , which satisfies

$$\left\|\nabla v^{M}(x_{k})\right\| \geq \varepsilon_{*}, \forall k > k_{*}.$$
(12)

On the basis of line search technique (7), we have

$$v^{M}(x_{k} + \alpha_{k}d_{k}) \leq v^{M}(x_{k}) + \lambda \alpha_{k} \nabla v^{M}(x_{k})^{T}d_{k}$$
$$+ \alpha_{k} \min\left[-\lambda_{1} \nabla v^{M}(x_{k})^{T}d_{k}, \lambda \alpha_{k} \|d_{k}\|^{2}/2\right]$$
$$\leq v^{M}(x_{k}) + (\lambda - \lambda_{1})\alpha_{k} \nabla v^{M}(x_{k})^{T}d_{k}$$

Then we sum above inequalities from k = 0 to $k = \infty$ and obtain

$$\sum_{k=0}^{k=\infty} (\lambda_1 - \lambda) \alpha_k \nabla v^M (x_k)^T d_k \le v^M (x_0) - v^M (x_\infty) < \infty$$
(13)

It is notable that the formula of (13) express

$$\lim_{k\to\infty}\alpha_k\nabla v^M(x_k)^Td_k=0$$

In virtue of (8), we obtain

$$\lim_{k \to \infty} -\beta_{I} \alpha_{k} \left\| \nabla v^{M} (x_{k}) \right\|^{2} = 0$$
⁽¹⁴⁾

Then we divide above conclusion into two parts:

If $\lim_{k \to \infty} \alpha_{_k} = 0\,$, it means that exist an infinity index set M , such that

$$\lim_{k \in M, \ k \to \infty} \alpha_k = 0 \tag{15}$$

In virtue of the algorithm's process, the constant $\frac{\alpha_k}{\gamma}$ does not satisfies (7), it means that

792

A Modified Three-Term Conjugate Gradient Algorithm

$$v^{M}(x_{k} + \alpha_{k} / \gamma d_{k}) - v^{M}(x_{k}) > -\lambda \alpha_{k} \beta_{1} / \gamma \|\nabla v^{M}(x_{k})\|^{2} + \alpha_{k} / \gamma \min\left[-\lambda_{1} \beta_{1} \|\nabla v^{M}(x_{k})\|^{2}, \lambda \alpha_{k} \|d_{k}\|^{2} / (2*\gamma)\right]$$

Considering the formulas of (8) and (9), we obtain

$$\beta_{1} \left\| \nabla v^{M} \left(x_{k} \right) \right\| \leq \left\| d_{k} \right\| \leq \left(\beta_{1} + \frac{2}{\beta_{2}} \right) \left\| \nabla v^{M} \left(x_{k} \right) \right\|$$

$$(16)$$

Thus has a positive constant $\lambda^* < \lambda_{\perp}$ such that

$$v^{M}(x_{k} + \alpha_{k} / \gamma d_{k}) - v^{M}(x_{k}) > -(\lambda - \lambda^{*})\alpha_{k} / \gamma \|\nabla v^{M}(x_{k})\|^{2}$$

On the basis of the mean-value theorem, the property of descent and objective function's continuity, there has a positive $\tau_k \in (0,1)$ meets the demand of

$$v^{M}(x_{k} + \alpha_{k} / \gamma d_{k}) - v^{M}(x_{k}) = \alpha_{k} / \gamma \nabla v^{M}(x_{k} + \tau_{k}\alpha_{k} / \gamma d_{k})^{T} d_{k}$$

$$= \alpha_{k} / \gamma [\nabla v^{M}(x_{k}) + \nabla v^{M}(x_{k} + \tau_{k}\alpha_{k} / \gamma d_{k}) - \nabla v^{M}(x_{k})]^{T} d_{k}$$

$$\leq -\beta_{1}\alpha_{k} / \gamma \|\nabla v^{M}(x_{k})\|^{2} + \tau_{k}\kappa(\alpha_{k} / \gamma)^{2} \|d_{k}\|^{2}$$

Compare with above formulas, we have

$$\alpha_{k} \geq \frac{\gamma(\beta_{1} - \lambda - \lambda^{*})\beta_{2}^{2}}{\tau_{k}\kappa(2 + \beta_{1}\beta_{2})}$$

It is obvious that above conclusion contradicts with $\lim_{k\to\infty} \alpha_k = 0$, thus end up with first part's proof.

If $\lim_{k \to \infty} \alpha_k > 0$, then have $\lim_{k \to \infty} \|\nabla v^M(x_k)\| = 0$, which contradicts with (12). Thus original conclusion holds, we complete the proof.

5 Numerical results of nonsmooth functions

Related content is presented in this section and consists of two parts: test problems and corresponding numerical results. To measure the algorithm's efficiency, we compare Algorithm 2.1 with algorithm 1 in terms of NI, NF, and CPU on the test problems listed in Tab. 1, where NI, NF, and CPU express the number of iterations, the calculation's frequency of the objective function, and the calculation time needed to solve various test problems (in seconds), respectively. Algorithm 1 is different from the objective algorithm in the formula of d_{k+1} that was determined by (4), and the remainder of algorithm 1 is identical to Algorithm 2.1.

Stopping rule: If the formula of $\|\nabla v(x_k)\| \le \varepsilon$ holds, the whole iteration number is greater than 10000 and the interior iteration frequency of step length $\lambda = 1000$ stops.

Initiation: $\varepsilon = 1e - 7$, $\gamma = 0.9$, $\lambda = 0.95$, $\lambda_1 = 0.1$, $\beta_1 = 1.6$, $\beta_2 = 0.01$, $\beta_3 = 0.001$.

Dimension: 30000, 60000, 90000, 120000, 150000, 180000, 210000.

Calculation environment: The calculation environment is a computer with 2 GB of memory, a Pentium(R) Dual-Core CPU E5800@3.20 GHz, the 64-bit Windows 7 operation system and the software of FORTRAN 90.

Other case: The uppercase letter of 'F' means the relative algorithm fails to solve corresponding problem.

The numerical results with the corresponding problem's index are listed in Tab. 2. Then, on the basis of the technique in Dolan [Dolan (2001)], plots of the corresponding figures are presented for two discussed algorithms.

No.	Problem	No	Problem
1	Chained LQ	5	Chained Mifflin 2
2	Chained CB3	6	Chained crescent 2
3	Chained CB3 2	7	Generalization of MAXQ
4	Chained crescent	8	Number of active faces

Table 1: The test problems

Table 2: Numeral results									
No	Dim	Algorithm 2.1			Algorithm 1				
INO.	Dim	NI	NF	CPU	NI	NF	CPU		
1	30000	32	81	0.184750	36	108	0.227750		
2	30000	4	43	0.091375	F	F	F		
3	30000	18	68	0.137625	F	F	F		
4	30000	31	81	0.171813	39	171	0.271875		
5	30000	32	81	0.203563	36	108	0.221438		
6	30000	3	9	0.015313	4	12	0.018438		
7	30000	23	429	0.454750	243	5067	5.206563		
8	30000	32	176	0.574000	101	1581	4.106250		
9	30000	33	84	1.953688	37	111	2.396875		
1	60000	47	112	0.527188	51	139	0.641000		
2	60000	4	43	0.209938	F	F	F		
3	60000	18	68	0.321000	F	F	F		
4	60000	46	112	0.584125	54	202	0.906750		
5	60000	47	112	0.598688	51	139	0.709688		

A Modified Three-Term Conjugate Gradient Algorithm

6	60000	4	11	0.051188	4	12	0.059500	-
7	60000	24	450	0.939125	254	5298	12.881630	
8	60000	33	179	1.301687	106	1686	10.241310	
9	60000	47	112	5.320000	66	170	7.524125	
1	90000	47	112	0.702000	66	170	0.966875	-
2	90000	4	43	0.308875	F	F	F	
3	90000	18	68	0.507375	F	F	F	
4	90000	46	112	0.761750	54	202	1.173063	
5	90000	47	112	0.868938	52	142	0.988250	
6	90000	5	14	0.086125	5	15	0.100188	
7	90000	24	450	1.541687	260	5424	17.023250	
8	90000	33	183	1.866562	110	1752	15.112440	
9	90000	48	114	8.018375	81	201	13.118440	
1	120000	47	112	0.920563	64	166	1.231375	-
2	120000	4	43	0.389688	F	F	F	
3	120000	18	68	0.605063	F	F	F	
4	120000	47	114	1.017250	55	205	1.589687	
5	120000	48	114	1.130062	66	170	1.547562	
6	120000	3	9	0.069625	4	12	0.094938	
7	120000	24	450	1.934188	265	5529	23.471190	
8	120000	34	188	2.491688	113	1815	20.853120	
9	120000	63	145	13.644500	81	201	17.501560	
1	150000	48	114	1.196125	67	173	1.673750	-
2	150000	4	43	0.512813	F	F	F	
3	150000	18	68	0.813188	F	F	F	
4	150000	47	114	1.263875	55	205	2.048750	
5	150000	48	114	1.407938	67	173	2.013375	
6	150000	3	9	0.086250	4	12	0.126875	
7	150000	25	471	2.608250	268	5592	30.071310	
8	150000	34	188	3.216687	114	1836	26.802130	
9	150000	63	145	17.122940	82	204	22.271810	
1	180000	48	114	1.407938	81	201	2.144063	-
2	180000	4	43	0.585688	F	F	F	
								-

3	180000	18	68	1.010750	F	F	F	
4	180000	47	114	1.515875	69	233	2.718875	
5	180000	48	114	1.676563	81	201	2.642313	
6	180000	3	9	0.112250	5	15	0.188125	
7	180000	25	471	3.063563	271	5655	36.615120	
8	180000	34	200	3.907125	116	1878	32.505370	
9	180000	63	145	20.321870	93	225	29.400560	
1	210000	63	145	1.942188	81	201	2.502875	
2	210000	4	43	0.730875	F	F	F	
3	210000	18	68	1.123500	F	F	F	
4	210000	48	117	1.776500	69	233	3.152000	
5	210000	49	117	1.904750	81	201	3.137625	
6	210000	3	9	0.112063	5	15	0.217750	
7	210000	25	471	3.309000	273	5697	42.916940	
8	210000	34	200	4.745500	117	1899	38.607690	
9	210000	63	145	23.970500	96	232	35.303310	



Figure 1: The iteration number of two algorithms



Figure 2: The calculation's frequency of objective function



Figure 3: The calculation's time of two algorithms

On the basis of numeral tests, it is obvious that discussed algorithms successfully address large scale nonsmooth problems and are valuable in addressing complex problems. In Figs. 1- 3, the red curve of proposed algorithm (Algorithm 2.1) is above the others since the iteration number (NI), the calculation frequency (NF), and calculation time (CPU) in solving problems are outstanding than the others. The initial points in three Figures are close to 1.0, which denotes the proposed algorithm is effective in solving challenging problems. It is noted that the red curve increases rapidly and is very smooth, which is important for a good optimization algorithm. Proposed algorithm avoids the denominator of the formula for search direction is 0, thus proposed algorithm enables efficiently to solve more complex problems.

Acknowledgement: The authors wish to express their appreciation to the reviewers for their helpful suggestions which greatly improved the presentation of this paper. This work is supported by the National Natural Science Foundation of China (Grant No. 11661009), the

Guangxi Science Fund for Distinguished Young Scholars (No. 2015GXNSFGA139001), the Guangxi Natural Science Key Fund (No. 2017GXNSFDA198046), and Innovation Project of Guangxi Graduate Education (No. YCSW2018046).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Albaali, M. (1985): Descent property and global convergence of the fletcher-reeves method with inexact line search. *IMA Journal of Numerical Analysis*, vol. 5, no. 1, pp. 121-124.

Albaali, M.; Spedicato, E.; Maggioni, F. (2014): Broyden's quasi-newton methods for a nonlinear system of equations and unconstrained optimization: a review and open problems. *Optimization Methods and Software*, vol. 29, no. 5, pp. 937-954.

Baggio, R.; Franceschini, A.; Spiezia, N.; Janna, C. (2017): Rigid body modes deflation of the preconditioned conjugate gradient in the solution of discretized structural problems. *Computers and Structures*, vol. 185, pp. 15-26.

Bernaschi, M.; Bisson, M.; Fantozzi, C.; Janna, C. (2016): A factored sparse approximate inverse preconditioned conjugate gradient solver on graphics processing units. *Siam Journal on Scientific Computing*, vol. 38, no. 1, pp. 53-72.

Birge, J.; Qi, L.; Wei, Z. (1998): Convergence analysis of some methods for minimizing a nonsmooth convex function. *Journal Optimization Theory Application*, vol. 97, no. 2, pp. 357-383.

Bonettini, S.; Loris, I.; Porta, F. (2015): Variable metric inexact line-search based methods for nonsmooth optimization. *Siam Journal on Optimization*, vol. 26, no. 2, pp. 891-921.

Dai, X.; Liu, Z.; Zhang, L; Zhou, A. (2017): A conjugate gradient method for electronic structure calculations. *Siam Journal on Scientific Computing*, vol. 39, no. 6.

Dolan, D.; Moré, J. (2001): Benchmarking optimization software with performance profiles. *Mathematical Programming*, vol. 91, no. 2, pp. 201-213.

Fasi, M.; Langou, J.; Robert, Y. (2016): A backward/forward recovery approach for the preconditioned conjugate gradient method. *Journal of Computational Science*, vol. 17, pp. 522-534.

Gilbert, J.; Nocedal, J. (1990): Global convergence properties of conjugate gradient methods for optimization. *Siam Journal on Optimization*, vol. 2, no. 1, pp. 21-42.

Hall, D.; Zimbro, H.; Maduro, R.; Morgan, M. (2017): Impact of a restraint management bundle on restraint use in an intensive care unit. *Journal of Nursing Care Quality*, vol. 33, no. 2, pp. 143-148.

Kim, B.; Um, T.; Suh, C.; Park, F. (2016): Tangent bundle RRT: a randomized algorithm for constrained motion planning. *Robotica*, vol. 34, no. 1, pp. 202-225.

Li, J.; Li, X.; Yang, B.; Sun, X. (2017): Segmentation-based image copy-move forgery detection scheme. *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 507-518.

798

Liu, Q.; Weng, C. (2016): Applications of the conjugate gradient fast fourier hankel transfer method with an improved fast hankel transform algorithm. *Radio Science*, vol. 29, no. 4, pp. 1009-1022.

Liu, W.; Chen, Z.; Ji, S.; Wang, H.; Zhang, J. (2017): Multi-party semi-quantum key agreement with delegating quantum computation. *International Journal of Theoretical Physics*, vol. 56, no. 10, pp. 3164-3174.

Liu, W.; Yong, X.; Yang, C.; Gao, P.; Yu, W. (2018): An efficient and secure arbitrary N-Party quantum key agreement protocol using bell states. *International Journal of Theoretical Physics*, vol. 57, pp. 195-207.

Li, X.; Zhang, W.; Dong, X. (2016): A class of modified FR conjugate gradient method and applications to non-negative matrix factorization. *Computers and Mathematics with Applications*, vol. 73, no. 2, pp. 270-276.

Luo, Y.; Tang, G.; Zhou, L. (2008): Hybrid approach for solving systems of nonlinear equations using chaos optimization and quasi-newton method. *Applied Soft Computing*, vol. 8, no. 2, pp. 1068-1073.

Melro, L.; Jensen, L. (2004): On second-order properties of the moreau-yosida regularization for constrained nonsmooth convex programs. *Numerical Functional Analysis and Optimization*, vol. 25, no. 5-6, pp. 515-529.

Nazareth, L. (1977): A conjugate direction algorithm without line searches. *Journal of Optimization Theory and Applications*, vol. 23, no. 3, pp. 373-387.

Sagara, N.; Fukushima, M. (2017): Trust region method for nonsmooth convex optimization. *Journal of Industrial and Management Optimization*, vol. 1, no. 2, pp. 171-180.

Sarkar, T. (2016): The application of the conjugate gradient method for the solution of operator equations arising in electromagnetic scattering from wire antennas. *Radio Science*, vol. 19, no. 5, pp. 1156-1172.

Shi, Y.; Tuan, H.; Su, S.; Tam, H. (2017): Nonsmooth optimization for optimal power flow over transmission networks. *Journal of Global Optimization*, vol. 69, pp. 1-16.

Tarzanagh, D.; Nazari, T.; Peyghami, P. (2016): A nonmonotone PRP conjugate gradient method for solving square and under-determined systems of equations. *Computers and Mathematics with Applications*, vol. 73, no. 2, pp. 339-354.

Touati-Ahmed, D.; Storey, C. (1990): Efficient hybrid conjugate gradient techniques. *Journal of Optimization Theory and Applications*, vol. 64, no. 2, pp. 379-397.

Wang, J.; Zhu, D. (2016): Derivative-free restrictively preconditioned conjugate gradient path method without line search technique for solving linear equality constrained optimization. *Computers and Mathematics with Applications*, vol. 73, pp. 277-293.

Xu, M.; Zhang, L.; Du, B.; Zhang, L. (2016): An image-based endmember bundle extraction algorithm using reconstruction error for hyperspectral imagery. *Neurocomputing*, vol. 173, pp. 397-405.

Yuan, G. (2009): A conjugate gradient method for unconstrained optimization problems. *International Journal of Mathematics and Mathematical Sciences*, vol. 2009, no. 16, pp. 1-14.

Yuan, G. (2009): Modified nonlinear conjugate gradient methods with sufficient descent

property for large-scale optimization problems. *Optimization Letters*, vol. 3, no. 1, pp. 11-21. **Yuan, G.; Lu, Y.** (2009): A modified PRP conjugate gradient method. *Annals of Operations Research*, vol. 166, no. 1, pp. 73-90.

Yuan, G.; Meng, Z.; Li, Y. (2016): A modified hestenes and stiefel conjugate gradient algorithm for large-scale nonsmooth minimizations and nonlinear equations. *Journal of Optimization Theory and Applications*, vol. 168, no. 1, pp. 1-24.

Yuan, G.; Sheng, Z. (2017): Optimization Algorithm of Nonsmooth. Science Press, China.

Yuan, G.; Wei, Z.; Li, G. (2014): A modified polak-ribière-polyak conjugate gradient algorithm for nonsmooth convex programs. *Journal of Computational and Applied Mathematics*, vol. 255, pp. 86-96.

Yuan, G.; Wei, Z.; Wang, Z. (2013): Gradient trust region algorithm with limited memory BFGS update for nonsmooth convex minimization. *Computational Optimization and Applications*, vol. 54, no. 1, pp. 15-64.

Yuan, G.; Zhang, M. (2015): A three-terms polak-ribière-polyak conjugate gradient algorithm for large-scale nonlinear equations. *Journal of Computational and Applied Mathematics*, vol. 286, pp. 186-195.

Zhang, H.; Wu, Q.; Nguyen, T. (2014): Synthetic aperture radar image segmentation by modified student's t-mixture model. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 7, pp. 4391-4403.

Zhang, L.; Zhou, W.; Li, D. (2006): A descent modified polak-ribière-polyak conjugate gradient method and its global convergence. *Ima Journal of Numerical Analysis*, vol. 26, no. 4, pp. 629-640.

Zhang, Z.; Liu, Q. (2016): Simulation of induction-logging response using conjugate gradient method with nonuniform fast fourier and fast hankel transforms. *Radio Science*, vol. 36, no. 4, pp. 599-608.

800