A NEW CONJUGATE GRADIENT METHOD WITH STRONGLY GLOBAL CONVERGENCE AND SUFFICIENT DESCENT CONDITION

DONG Xiao-liang¹, HE Yu-bo², KONG Xiang-yu¹, LI Wei-jun³

(1. School of Mathematics and Information, Beifang University of Nationalities, Yinchuan, 710021, China)

 (2.Department of Mathematics and Application Mathematics, Huaihua University, Huaihua, 418008, China)
 (3.Network Information Technology Center, Beifang University of Nationalities, Yinchuan, 710021, China)

Abstract: In this paper, we study the WYL conjugate gradient method for unconstrained optimization problems. By making use of the modified iterative scheme, the sufficient descent conditions are satisfied at each iteration independent of the line search used. Also, by removing the original restriction on the parameter of the Wolfe conditions, we establish the strongly global convergence property for the general function. Numerical results illustrate that our method is efficient for the test problems.

Keywords: conjugate gradient method; sufficient descent condition; strongly global convergence; Wolfe line search

 2010 MR Subject Classification:
 90C30; 65K10

 Document code:
 A
 Article ID:
 0255-7797(2017)02-0231-08

1 Introduction

Consider the following unconstrained optimization problem

$$\min f(x),\tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth and nonlinear function, and its gradient $g_k = \nabla f(x_k)$ is available. The conjugate gradient (CG) methods represent an important class of unconstrained optimization algorithms with strong local and global convergence properties and modest memory requirements.

^{*} Received date: 2014-05-02 Accepted date: 2015-03-16

Foundation item: Supported by National Natural Science Foundation of China (11601012; 11661002); Ningxia Natural Science Foundation (NZ13095; NZ16093); Scientific Research Foundation of the Higher Education Institutions of Ningxia (NGY2016134); Beifang University of Nationalities Foundation (2016SXKY06; 2014XBZ09; 2014XBZ01; 2013XYZ028).

Biography: Dong Xiaoliang (1981–), male, born at Lingwu, Ningxia, lecturer, major in nonlinear optimization.

The iterative formula of the CG method is given by

$$x_{k+1} = x_k + s_k, s_k = \alpha_k d_k.$$
(1.2)

In (1.2), d_k is a search direction updated by

$$d_k = \begin{cases} -g_k, & k = 1, \\ -g_k + \beta_k d_{k-1}, & k \ge 2, \end{cases}$$
(1.3)

and the steplength $\alpha_k > 0$ is commonly chosen to satisfy certain line search conditions. Among them, the so-called Wolfe conditions have attracted special attention in the convergence analyses and the implementations of CG methods, requiring that

$$f(x_k + \alpha_k d_k) - f(x_k) \le \rho \alpha_k g_k^T d_k, \tag{1.4}$$

$$g\left(x_k + \alpha_k d_k\right)^T d_k \ge \sigma g_k^T d_k,\tag{1.5}$$

where $0 < \rho < \sigma < 1$ are often imposed on the line search.

The well-known formulas β_k are Fletcher-Reeves, Hestenes–Stiefel, Polak-Ribière-Polyak and Dai-Yuan formulas, which are specified by

$$\beta_{k}^{FR} = \frac{\|g_{k}\|^{2}}{\|g_{k-1}\|^{2}}, \beta_{k}^{HS} = \frac{g_{k}^{T}(g_{k} - g_{k-1})}{d_{k-1}^{T}y_{k-1}},$$

$$\beta_{k}^{PRP} = \frac{g_{k}^{T}(g_{k} - g_{k-1})}{\|g_{k-1}\|^{2}}, \beta_{k}^{DY} = \frac{\|g_{k}\|^{2}}{d_{k-1}^{T}y_{k-1}},$$
(1.6)

where $y_{k-1} = g_k - g_{k-1}$. Please refer to [1] for more details of their convergence properties.

Little is known concerning global convergence of the HS method for nonconvex minimization problems. However, the HS method, which resembles the PRP method in practical computation and is often recommended due to its superior numerical properties.

Recently, various modifications of the HS method received growing interests, in which sufficient descent condition is important in the convergence analysis of CG method.

Hager and Zhang [2] proposed the CG_DESCENT method and proved its convergence with the Wolfe search. Based on the MBFGS method [3], Zhang [4] and Dai [5] introduced modified HZ methods with the Armijo search for nonconvex objective functions. Zhang, Zhou and Li [6] proposed a three-term modified PRP method, in which the property $-d_k^T g_k = ||g_k||^2$ always holds without any line searches. To seek the CG direction that is closest to the direction of the scaled memoryless BFGS method, Dai and Kou [7] proposed a family of CGOPT methods. Yu et al. [8, 9] proposed several modified spectral CG methods. For further study on the some recent advances, we may refer to [10–15].

In this paper, we will continue to study the HS-type method. One feature of our method is that our method guarantee sufficient descent condition and strongly global convergence.

The rest of this paper is organized as follows. In the next section, the new method is presented formally. Meanwhile, we prove the proposed method possesses sufficient descent and strongly global convergence properties with the Wolfe line search. In Section 4, we report numerical comparisons with existing methods by using some test problems.

2 A New CG Method and Its Properties

Recently, Wei, Yao and Huang [16] proposed the MHS, WYL and MLS methods, in which the parameters β_k are given by

$$\beta_k^{MHS} = \frac{g_k^T \overline{y_{k-1}}}{d_{k-1}^T y_{k-1}}, \beta_k^{WYL} = \frac{g_k^T \overline{y_{k-1}}}{||g_{k-1}||^2}, \beta_k^{MLS} = -\frac{g_k^T \overline{y_{k-1}}}{d_{k-1}^T g_{k-1}},$$
(2.1)

where $\overline{y_{k-1}} = g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}$. The global convergence of these methods with the strong Wolfe line search was proved for cases that the parameter σ in (1.5) is constrained to $\sigma \in \left[0, \frac{1}{3}\right], \sigma \in \left[0, \frac{1}{4}\right]$ and $\sigma \in \left[0, \frac{1}{4}\right]$, respectively. Furthermore, with the same restriction on the parameter σ , these methods above possessed the sufficient descent condition.

The above discussion prompts naturally us to raise the following question:

Is it possible to modify the direction of the MHS method in a suitable way, thereby enlarging its sufficient descent properties and ensuring the global convergence for the general functions?

By taking the theoretical advantage of the MHS method into consideration, we give another method to guarantee sufficient descent condition, in which strongly global convergence is satisfied. With our choices, the additional restriction on the parameter σ in (1.5) can be removed. The direction d_k in our method is given by

$$d_k^{DHS} = \begin{cases} -g_k, & d_{k-1}^T y_{k-1} \le \varepsilon_1 \|y_{k-1}\| \|d_{k-1}\|, \\ -g_k + \beta_k^{DHS} d_{k-1} + \varphi_k g_{k-1}, & d_{k-1}^T y_{k-1} > \varepsilon_1 \|y_{k-1}\| \|d_{k-1}\|, \end{cases}$$
(2.2)

where $\varepsilon_1 > 0$ is a constant.

In (2.2), the parameters β_k^{DHS} and φ_k are chosen as

$$\beta_k^{DHS} = \frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right)}{\max\{d_{k-1}^T y_{k-1}, \lambda | d_{k-1}^T g_k |\}},$$
(2.3)

$$\varphi_k = \frac{g_k^T d_{k-1}}{\max\{d_{k-1}^T y_{k-1}, \lambda | d_{k-1}^T g_k |\}} \cdot \frac{\|g_k\|}{\|g_{k-1}\|},$$
(2.4)

where $\lambda > 1$ and $\varepsilon_1 > 0$ are two given constants.

For convenience, we call our method (2.2) as DHS method in later part of this paper and formally state the steps of this method as follows.

Algorithm 2.1 (DHS method)

Step 1 Choosing constants $\lambda > 1$, $\varepsilon_1 > 0$ and $\varepsilon > 0$. Select an initial point $x_1 \in \mathbb{R}^n$, set k = 1.

Step 2 Test a criterion for stopping the iterations. If $||g_k|| < \varepsilon$, then stop, otherwise calculate d_k by (2.2).

Step 3 Determine the steplength by the Wolfe conditions.

Step 4 Calculate the new iterate by $x_{k+1} = x_k + \alpha_k d_k$, set k = k+1 and goto Step 2. The descent property is an indispensable factor in the convergence analysis of CG method. More exactly, if there exists a constant $c_1 > 0$ such that

$$d_k^T g_k \le -c_1 \|g_k\|^2, \forall k \in N$$
(2.5)

holds for all $k \in N$, then the so-called sufficient descent condition holds.

Property (2.5) is guaranteed in our method, as proved in the following lemma.

Lemma 2.1 Let $\{x_k\}$ and $\{d_k\}$ be generated by the DHS method. Then the direction d_k satisfies the sufficient descent condition (2.5), which is independent of any line search.

Proof For k = 1, it follows that $d_1^T g_1 = -||g_1||^2$. Now we mainly consider the case where

$$d_{k-1}^T y_{k-1} > \varepsilon_1 ||y_{k-1}|| ||d_{k-1}||$$

is satisfied. It follows from (2.2) that

$$g_{k}^{T}d_{k} = -||g_{k}||^{2} + \frac{g_{k}^{T}d_{k-1}}{\max\{d_{k-1}^{T}y_{k-1},\lambda|d_{k-1}^{T}g_{k}|\}}||g_{k}||^{2}$$

$$\leq -||g_{k}||^{2} + \frac{|g_{k}^{T}d_{k-1}|}{\max\{d_{k-1}^{T}y_{k-1},\lambda|d_{k-1}^{T}g_{k}|\}}||g_{k}||^{2}$$

$$\leq -\left(1 - \frac{1}{\lambda}\right)||g_{k}||^{2}.$$
(2.6)

Then the conclusion holds by letting $c_1 = 1 - \lambda^{-1}$.

3 Convergence Analysis

In this section, we prove the global convergence of the proposed CG method. We need the following assumptions, which are generally assumed in the literature.

Assumption 3.1

Boundedness Assumption: the level set defined by $\Omega = \{x \in \mathbb{R}^n | f(x) \le f(x_1)\}$ with x_1 to be the initial point, is bounded;

Lipschitz Assumption: in some neighborhood Ω_0 of Ω , the objective function f is continuously differentiable, and its gradient g is Lipschitz continuous, namely, there exist a constant L > 0 such that

$$||g(x) - g(y)|| \le L||x - y||, \forall x, y \in \Omega_0.$$
(3.1)

Theorem 3.1 Suppose that Assumption 3.1 holds. Let $\{x_k\}$ and $\{d_k\}$ be generated by Algorithm 2.1. Then there exists a constant \overline{c} such that

$$||d_k|| \le \overline{c}||g_k||, \ \forall k \in N.$$

$$(3.2)$$

Proof Since $d_1 = -g_1$, the result obviously holds for k = 1.

If k > 1, it suffice to consider the case where $d_{k-1}^T y_{k-1} > \varepsilon_1 ||y_{k-1}|| ||d_{k-1}||$ holds. It follows form the definition β_k^{DHS} in (2.3) and the Cauchy inequality that

$$\begin{aligned} |\beta_k^{DHS}| &\leq ||g_k|| \frac{||g_k - g_{k-1}|| + \left\| g_{k-1} - \frac{||g_k||}{||g_{k-1}||} g_{k-1} \right\|}{d_{k-1}^T (g_k - g_{k-1})} \\ &\leq 2 \frac{||g_k|| ||y_{k-1}||}{|d_{k-1}^T y_{k-1}|} \leq \frac{2}{\varepsilon_1} \frac{||g_k||}{||d_{k-1}||}. \end{aligned}$$

Also, we can estimate the upper bound for $|\varphi_k|$, presented by

$$|\varphi_k| = \left| \frac{g_k^T d_{k-1}}{\max\{d_{k-1}^T y_{k-1}, \lambda | d_{k-1}^T g_k|\}} \right| \cdot \frac{\|g_k\|}{\|g_{k-1}\|} \le \frac{1}{\lambda} \frac{\|g_k\|}{\|g_{k-1}\|}.$$
(3.3)

Combining this with (3.3) yields

$$||d_{k}|| \leq ||g_{k}|| + |\beta_{k}^{DHS}|||d_{k-1}|| + |\varphi_{k}|||g_{k-1}|| \\ \leq \left(1 + \frac{2}{\varepsilon_{1}} + \frac{1}{\lambda}\right) ||g_{k}||.$$
(3.4)

Let $\overline{c} = 1 + \frac{2}{\varepsilon_1} + \frac{1}{\lambda}$, we complete the proof.

The conclusion of the following theorem, called the Zoutendijk condition, is often used to prove global convergence of nonlinear CG method.

Theorem 3.2 (see [17]) Suppose that Assumption 3.1 holds. Consider the general CG method, where d_k is a descent direction and α_k satisfies the Wolfe conditions. Then we have

$$\sum_{k\ge 1} \frac{(g_k^T d_k)^2}{||d_k||^2} < \infty.$$
(3.5)

For the general function, we can develop a strongly global convergence result as follows. **Theorem 3.3** Suppose that Assumption 3.1 holds. Let $\{x_k\}$ be generated by Algorithm 2.1. Then

$$\lim_{k \to \infty} ||g_k|| = 0.$$
(3.6)

Proof The bound for $||d_k||$ in (3.2) coupled with (2.5) indicates that

$$\sum_{k\geq 1} (\bar{c})^{-2} ||g_k||^2 \le \sum_{k\geq 1} \frac{||g_k||^4}{||d_k||^2} < \left(1 - \frac{1}{\lambda}\right)^{-2} \sum_{k\geq 1} \frac{(g_k^T d_k)^2}{||d_k||^2} < +\infty.$$
(3.7)

Equation (3.7) leads to (3.6).

4 Numerical Experiments

In this section, we provide the implementation detail of the new methods to verify the numerical performance. Our tests problems come form Moré [18].

We stop the iteration if the inequality $||g_k|| \leq 10^{-6}$ is satisfied. We use "F" to denote the numbers of iteration exceed 2000. All algorithms use exactly the same implementation of the strong Wolfe conditions with $\rho = 10^{-3}, \sigma = 0.5$. The parameters in (2.2) are specified by $\varepsilon_1 = 10^{-12}, \mu = 10$.

In order to rank the CG methods, we subsequently use the method of Dai and Ni [19] to compare the performance of our methods with that of the other methods in [16].

First, we compute the total number of the function and its gradient evaluations by the formula $N_{\text{total}} = NF + m * NG$, where m=5.

Second, for all our two methods, we evaluate their efficiency with respect to the WYL method in the following way. For each problem i, compute the ratio

$$r_i = \frac{N_{\text{total},i}(\text{a method in [16]})}{N_{\text{total},i}(\text{DHS})}$$

and the geometric mean of these ratios over all the test problems $r_{\text{total}} = (\prod_{i \in S} r_i)^{\frac{1}{|S|}}$, where S denotes the set of the problems and |S| the number of elements in S.

Problem	Ν	DHS	MHS	WYL	MLS
		NI/NF/NG/TIME	NI/NF/NG/TIME	NI/NF/NG/TIME	NI/NF/NG/TIME
ROSE	2	36/120/102/0.094	93/254/218 /0.03	105/282/248/0.034	61/176/151 /0.021
FROTH	2	19/76/64/0.0086	24/95/79/0.01	39/147/125/0.016	25/102/85/0.01
BADSCP	2	48/245/219/0.025	\mathbf{F}	\mathbf{F}	902/2548/2345/0.34
BADSCB	2	13/56/45/0.0061	\mathbf{F}	\mathbf{F}	253/1165/836/0.12
BEALE	2	14/47/37/0.0069	16/57/43/0.0073	17/61/47/0.0073	26/86/69/0.01
JENSAM	2	13/43/31/0.0061	15/51/39/0.0064	19/63/50/0.0087	18/61/47/0.0085
HELIX	3	55/149/121/0.026	100/262/221/0.037	134/364/312/0.05	43/121/95/0.016
BARD	3	28/88/77/0.018	22/69/58/0.014	25/81/68/0.016	37/106/85/0.02
GAUSS	3	4/9/5/0.0023	4/9/5/0.0018	4/9/5/0.0026	4/9/5/0.0018
MEYER	3	\mathbf{F}	\mathbf{F}	\mathbf{F}	\mathbf{F}
GULF	3	70/340/301/0.059	29/131/107/0.023	31/140/119/0.025	30/122/99/0.022
BOX	3	31/125/99/0.029	19/67/54/0.011	174/525/456/0.085	99/303/259/0.049
SING	4	188/588/514/0.07	324/1041/923/0.12	425/1354/1201/0.16	525/1625/1433/0.22
WOOD	4	79/288/246/0.034	407/1093/989/0.14	377/1003/923/0.13	343/893/830/0.12
KOWOSB	4	211/650/580/0.1	51/165/146/0.027	57/173/149/0.028	63/231/201/0.034
BD	4	64/866/372/0.27	\mathbf{F}	57/491/256/0.14	53/770/355/0.2
OSB1	5	794/2574/2290/0.67	\mathbf{F}	\mathbf{F}	336/1172/1027/0.29
BIGGS	6	75/264/232/0.046	142/486/429/0.087	113/385/335/0.068	184/542/487/0.11
OSB2	11	215/559/509/0.31	183/481/426/0.27	217/589/527/0.33	222/596/534/0.34
WATSON	20	1039/3521/3102/4.6	\mathbf{F}	\mathbf{F}	\mathbf{F}
ROSEX	1000	39/132/112/2	119/361/320/5.7	64/206/185/3.3	84/264/228/4.1
SINGX	1000	361/1175/1029/19	1509/5135/4486/78	265/832/733/13	1832/5899/5237/90
PEN1	20	141/625/524/0.21	100/402/348/0.14	100/404/349/0.14	120/461/404/0.16
PEN2	20	144/658/580/0.13	150/617/537/0.13	107/443/376/0.087	347/1191/1036/0.25
VARDIM	100	10/90/59/0.028	7/92/61/0.027	9/84/55/0.032	9/84/55/0.026
TRIG	100	49/123/109/0.3	50/128/113/0.31	50/128/116/0.31	49/125/108/0.3
BV	1000	0/1/1/0.022	0/1/1/0.021	0/1/1/0.02	0/1/1/0.02
IE	500	6/14/9/1.7	6/14/9/1.7	6/14/9/1.7	6/14/9/1.7
BAND	100	18/54/33/0.038	18/57/35/0.043	19/59/36/0.041	19/59/36/0.041
LIN	100	1/8/8/0.015	1/8/8/0.014	1/8/8/0.014	1/8/8/0.014
LIN	500	1/9/9/0.28	1/9/9/0.29	1/9/9/0.28	1/9/9/0.29

Table 4.1: Numerical Results

Finally, we present the r_{total} in the Table 4.2:

Table 4.2: Comparison of efficiency with other algorithm

Alg	orithm	DHS	MHS	WYL	MLS
1	total	1	1.1216	1.1642	1.3124

References

- Dai Y H, Yuan Y X. Nonlinear conjugate gradient methods[M]. Shanghai: Shanghai Science and Technology Press, 2000.
- [2] Hager W W, Zhang H C. A new conjugate gradient method with guaranteed descent and an efficient line search[J]. SIAM J. Optim., 2005, 16(1): 170–192.
- [3] Li D H, Fukushima M. A modified BFGS method and its global convergence in nonconvex minimizationp[J]. J. Comput. Appl. Math., 2001, 129(1): 15–35.
- [4] Zhang L, Zhou W J. On the global convergence of Hager–Zhang conjugate gradient with Armijo line search[J]. Math. Numer. Sin., 2008, 28(5): 840–845.
- [5] Dai Z F, Wen F H. Global convergence of a modified Hestenes-Stiefel nonlinear conjugate gradient method with Armijo line search[J]. Numer. Algor., 2012, 59(1): 79–93.
- [6] Zhang L, Zhou W J, Li D H. A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence[J]. IMA J. Numer. Anal., 2006, 26(4): 629–640.
- [7] Dai Y H, Kou C X. A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search[J]. SIAM J. Optim., 2013, 23(1): 296–320.
- [8] Yu G H, Guan L T, Chen W F. Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization[J]. Optim. Meth. Softw., 2008, 23(2): 275–293.
- Yu G H. New descent nonlinear conjugate gradient methods for large-scale unconstrained optimization, Tech. Rep. Department of scientific computation and computer applications[D]. Guangzhou: Sun Yat-Sen University, 2005.
- [10] Dong X, Liu H, He Y, Yang X. A modified Hestenes–Stiefel conjugate gradient method with sufficient descent condition and conjugac condition[J]. J. Comp. Appl. Math., 2015, 281: 239–249.
- [11] Dong X, Liu H, Xu Y, Yang X. Some nonlinear conjugate gradient methods with sufficient descent condition and global convergence[J]. Optim. Lett., 2015, 9(7): 1421–1432.
- [12] Dong X. Comment on "A new three-term conjugate gradient method for unconstrained problem" [J]. Numer. Algorithms, 2016, 72(1): 173–179.
- [13] Dong X, Liu H, He Y. A self-adjusting conjugate gradient method with sufficient descent condition and conjugacy condition[J]. J. Optim. Theory Appl., 2015, 165(1): 225–241.
- [14] Dong X, Li C, He Y, Yu G. A relaxed two-stage multi-spllting algorithm for bi-obstacle problems[J].
 J. Math., 2011, 31(2): 323–330.
- [15] He Y, Dong X. On the convergence of Levenberg–Marquardt method for nonlinear inequalities[J]. J. Math., 2012, 32(1): 25–34.
- [16] Yao S W, Wei Z X, Huang H. A note about WYL's conjugate gradient method and its applications[J]. Appl. Math. Comput., 2007, 191(2): 381–388.
- [17] Wolfe, P. Convergence conditions for ascent methods[J]. SIAM Rev., 1969, 11(2): 226–235
- [18] Moré J J, Garbow B S, Hillstrom K E. Testing unconstrained optimization software[J]. ACM Trans. Math. Softw., 1981, 7(1): 17–41.

[19] Dai Y H, Ni Q. Testing different conjugate gradient methods for large-scale unconstrained optimization[J]. J. Comput. Math., 2003, 21(3): 311–320.

一类新的具有充分下降条件和强收敛性的共轭梯度法

董晓亮^{1,2},何郁波³,孔翔宇¹,李卫军¹ (1.北方民族大学数学与信息学院,宁夏 银川 750021) (2.怀化学院数学与应用数学系,湖南 怀化 418008)

(3.北方民族大学网络信息技术中心, 宁夏 银川 750021)

摘要: 本文研究了求解无约束优化问题的WYL共轭梯度法.利用修正迭代格式,得到了算法在每步迭 代能产生不依赖于搜索条件的充分下降方向.同时,在原算法中关于Wolfe条件中参数去掉的情况下,获得了 本文算法是强收敛的.数值实验说明本文算法可以有效求解测试问题.

关键词: 共轭梯度法;充分下降条件;强收敛性;Wolfe搜索 MR(2010)主题分类号: 90C30;65K10 中图分类号: O224