ACTA MATHEMATICA VIETNAMICA 161
Volume 37, Number 2, 2012, pp. 161–169

AN IMPROVED DESCENT CONJUGATE GRADIENT METHOD AND ITS CONVERGENCE

MIN SUN AND JING LIU

Abstract. Recently, a new family of conjugate gradient methods with a Grippo-Lucidi type step length rule is proposed by Shi and Guo [A new family of conjugate gradient methods. Journal of Computational and Applied Mathematics, 2009, 224:444-457]. In this paper, we improve Shi and Guo's method by adopting an improved Grippo-Lucidi type step length rule, and the improvement is twofold: (1) We drop the local Lipschitz constant in the step length, which is beyond the problem data and has to be further estimated; (2) The search direction d_k only needs to satisfy the descent property instead of the sufficient descent property. The global convergence result of the modified method is established under some mild conditions. Preliminary numerical results are also reported to show the efficiency of the improved method.

1. Introduction

Consider the unconstrained nonlinear optimization problem

$$\min f(x),\ x \in \mathbb{R}^n, \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and its gradient $g(x)$ is available. Conjugate gradient method is very effective for solving large-scale unconstrained optimization problem (1.1) due to its low memory requirements, and its iterative formula is given by

$$x_{k+1} = x_k + \alpha_k d_k. \tag{1.2}$$

with

$$d_k = \begin{cases} -g_k,\ k = 1, \\ -g_k + \beta_k d_{k-1},\ k \geq 2, \end{cases} \tag{1.3}$$

where x_1 is a given initial point, α_k is a step-length along d_k which is computed by carrying out some line search, g_k denotes $g(x_k)$ and β_k is a suitable scalar

Received December 20, 2010; in revised form March 27, 2012.
2000 *Mathematics Subject Classification.* 90C30, 90C33.
Key words and phrases. Conjugate gradient method, descent direction, global convergence.
This work is supported by the Shandong Provincial Natural Science Foundation Under Grant ZR2010AL005, the Shandong Outstanding Young Scientists Award Project Under Grant BS2010SF031, the Humanity and Social Science Youth foundation of Ministry of Education of China Under Grant 12YJC630033.

given by different formulae which result in distinct conjugate gradient methods. β_k can be defined by

$$\beta_k^{\text{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2},\ \beta_k^{\text{PRP}} = \frac{g_k^\top(g_k - g_{k-1})}{\|g_{k-1}\|^2},\ \beta_k^{\text{DY}} = \frac{\|g_k\|^2}{d_{k-1}^\top(g_k - g_{k-1})},$$

$$\beta_k^{\text{CD}} = -\frac{\|g_k\|^2}{g_{k-1}^\top d_{k-1}},\ \beta_k^{\text{HS}} = \frac{g_k^\top(g_k - g_{k-1})}{d_{k-1}^\top(g_k - g_{k-1})},\ \beta_k^{\text{LS}} = \frac{g_k^\top(g_k - g_{k-1})}{-g_{k-1}^\top d_{k-1}},$$

where $\|\cdot\|$ stands for the Euclidean norm. The corresponding method is respectively called FR (Fletcher-Revees), PRP (Polyak-Ribiére-Polyak), DY (Dai-Yuan), CD (Conjugate Descent), HS (Hestenes-Stiefel), LS (Liu-Storey) conjugate gradient method.

The PRP method is globally convergent when the objective function is strictly convex and the line search are exact, but for a general objective function, the PRP method can cycle infinitely without approaching a solution point even when the line search is exact, and therefore many authors have contributed to this topic. Under several practical line searches, one can prove the global convergence of the PRP method[1, 2, 9].

Recently, Shi and Guo [1] proposed a new family of conjugate methods, and the parameter β_k is defined by

$$\beta_k^{\text{SG}} = \frac{g_k^\top(g_k - g_{k-1})}{(1-u)\|g_{k-1}\|^2 - u g_{k-1}^\top d_{k-1}},$$

where $u \in [0,1]$. It is obvious that if $u = 0$ then $\beta_k^{\text{SG}} = \beta_k^{\text{PRP}}$ and if $u = 1$ then $\beta_k^{\text{SG}} = \beta_k^{\text{LS}}$. Shi and Guo [1] proposed a new nonmonotone Grippo-Lucidi type line search, that is, for given constants $\rho \in (0,1)$, $c \in (0,1/2)$, $u \in [0,1]$, setting

$$s_k = \frac{1-c}{L_k} \frac{(1-u)\|g_k\|^2 - u g_k^\top d_k}{\|d_k\|^2},$$

then let

$$\alpha_k = \max\{\rho^j s_k; j = 0, 1, \ldots\},$$

such that α_k satisfies

$$f(x_k + \alpha_k d_k) \le \max_{0 \le j \le m(k)} f(x_{k-j}) + \mu \alpha_k [g_k^\top d_k + \frac{1}{2}\alpha_k L_k \|d_k\|^2],$$

and

$$g(x_k + \alpha_k d_k)^\top d(x_k + \alpha_k d_k) \le -c\|g(x_k + \alpha_k d_k)\|^2, \tag{1.4}$$

where

$$d(x_k + \alpha_k d_k) = -g(x_k + \alpha_k d_k) + \frac{g(x_k + \alpha_k d_k)^\top (g(x_k + \alpha_k d_k) - g_k)}{(1-u)\|g_k\|^2 - u g_k^\top d_k} d_k.$$

Shi and Guo [1] proved that the new method with the above nonmonotone line search is globally convergent for nonconvex minimization. In this paper, we improve Shi and Guo's method in twofold: (1) We drop the local Lipschitz constant in the step length α_k, which is beyond the problem data and has to be further estimated; (2) The direction d_k only needs to satisfy the descent property

$$g_k^\top d_k < 0 \tag{1.5}$$

instead of the sufficient descent property [4]. Obviously, the computation load of descent property (1.5) is less than the sufficient descent property (1.4).

From now on, we always suppose the following assumption holds.

Assumption 1.1. (H1): The level set $L(x_1) = \{x \in R^n | f(x) \leq f(x_1)\}$ is bounded, where x_1 is the initial point.

(H2): In some convex set B that contains the level set $L(x_1)$, the gradient g is Lipschitz continuous, i.e., there exists an $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|.$$

Under (H1), there exists a positive constant $\tau > 0$ such that

$$\|g_k\| \leq \tau. \tag{1.6}$$

The remainder of the paper is organized as follows. We describe the improved method and show it is well-defined in Section 2. In Section 3, we prove its global convergence under some mild conditions, and preliminary computational results are given in Section 4. Finally, Section 5 contains some conclusions.

2. The improved method

Motivated by Shi and Guo [1] and the convergence analysis for the PRP method by Yu *et al.* [2], we propose an improved conjugate gradient method, in which the scalar β_k is defined by

$$\beta_k^{\text{new}} = \frac{g_k^\top (g_k - g_{k-1})}{u_1 \|g_{k-1}\|^2 - u_2 g_{k-1}^\top d_{k-1}}, \tag{2.1}$$

where $u_1 \geq 0, u_2 \geq 0, u_1 + u_2 > 0$. Now we present the improved method as follows.

Algorithm 2.1. The improved method

Step 0: Given $\varepsilon > 0$. Choose $x_1 \in \mathbb{R}^n$, $c > 0$, $\rho \in (0,1)$, $\mu \in (0,1)$, $u_1 \geq 0, u_2 \geq 0, u_1 + u_2 > 0$.

Step 1: Compute g_1. If $\|g_1\| < \varepsilon$, then stop; else, set $d_1 = -g_1$, $k := 1$.

Step 2: Set

$$s_k = c\frac{u_1\|g_k\|^2 - u_2 g_k^\top d_k}{\|d_k\|^2},$$

and let

$$\alpha_k = \max\{\rho^j s_k ; j = 0, 1, \cdots\} \tag{2.2}$$

satisfies

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \mu\alpha_k (g_k^\top d_k - c\alpha_k \|d_k\|^2), \tag{2.3}$$

and

$$g(x_k + \alpha_k d_k)^\top d(x_k + \alpha_k d_k) < 0, \tag{2.4}$$

where

$$d(x_k + \alpha_k d_k) = -g(x_k + \alpha_k d_k) + \frac{g(x_k + \alpha_k d_k)^\top (g(x_k + \alpha_k d_k) - g_k)}{u_1 \|g_k\|^2 - u_2 g_k^\top d_k} d_k.$$

Set $x_{k+1} = x_k + \alpha_k d_k$, and compute g_{k+1}.

Step 3: If $\|g_{k+1}\| \leq \varepsilon$, then stop. Otherwise, set $k := k + 1$, and go to Step 2.

Remark 2.1. In fact, inequality (2.3) in the above line search is motivated by the line search in [2].

Lemma 2.1. *Suppose that $\|g_k\| \neq 0$ holds, then the line search (2.2)-(2.4) in Algorithm 2.1 is well defined.*

Proof. The proof easily follows from Lemma 2.1 in [2]. Thus it is omitted here. □

Remark 2.2. By induction, noting that as $g_1^\top d_1 = -\|g_1\|^2 < 0$ we will have $g_k^\top d_k < 0$ for all $k \geq 1$.

3. Global convergence

To establish the global convergence of Algorithm 2.1, we need the following result.

Lemma 3.1. *Assume that Assumption 1.1 holds, and Algorithm 2.1 generates an infinite sequence $\{x_k\}$, then*

$$\|d_k\| \leq \varrho \|g_k\| \ \forall \ k \geq 1, \tag{3.1}$$

where $\varrho > 0$ is a constant.

Proof. For $k = 1$, we have $\|d_k\| = \|g_k\|$. For $k > 1$, we have

$$\alpha_k \leq s_k \leq c \frac{u_1 \|g_k\|^2 - u_2 g_k^\top d_k}{\|d_k\|^2}.$$

By Cauchy-Schwarz inequality and the above inequality, we have

$$\begin{aligned} \|d_{k+1}\| &= \| - g_{k+1} + \beta_{k+1}^{\text{new}} d_k \| \\ &\leq \|g_{k+1}\| + \frac{|g_{k+1}^\top (g_{k+1} - g_k)|}{u_1 \|g_k\|^2 - u_2 g_k^\top d_k} \|d_k\| \\ &\leq \|g_{k+1}\| (1 + \alpha_k \frac{L \|d_k\|^2}{u_1 \|g_k\|^2 - u_2 g_k^\top d_k}) \\ &\leq (1 + Lc) \|g_{k+1}\|, \end{aligned}$$

then, taking $\varrho = 1 + Lc$ yields the desired result. □

Lemma 3.2. *Suppose that Assumption 1.1 holds and $\{x_k\}$ is generated by Algorithm 2.1. Then*

$$\lim_{k \to \infty} \alpha_k \|d_k\| = 0. \tag{3.2}$$

Proof. Since $\{f(x_k)\}$ is a decreasing sequence, it is clear that the sequence $\{x_k\}$ generated by Algorithm 2.1 is contained in the level set $L(x_1)$, and there exists a constant f^* such that

$$\lim_{k\to\infty} f(x_k) = f^*.$$

Hence, we have

$$\sum_{k=1}^{\infty}(f_k - f_{k+1}) = f_1 - f^* < +\infty. \tag{3.3}$$

By (2.3), we obtain

$$f_k - f_{k+1} \geq -\mu\alpha_k g_k^\top d_k + c\mu\alpha_k^2\|d_k\|^2 \geq c\mu(\alpha_k\|d_k\|)^2,$$

which, together with (3.3), implies that (3.2) holds. The proof is complete. □

The next result shows that Algorithm 2.1 is global convergent. If $u_1 = 0$, we assume $Lc < 1$.

Theorem 3.1. *In the setting of Lemma 3.1, we have*

$$\liminf_{k\to\infty} \|g_k\| = 0.$$

Proof. Suppose that the conclusion does not hold. Then there exists a constant $\varepsilon > 0$ such that for all k,

$$\|g_k\| > \varepsilon. \tag{3.4}$$

We will divide our proof into two cases: $\alpha_k = s_k$ and $\alpha_k < s_k$. In the first case, we have

$$\alpha_k \geq c\frac{u_1\|g_k\|^2 - u_2 g_k^\top d_k}{\|d_k\|^2}. \tag{3.5}$$

If $\alpha_k < s_k$, this implies that α_k/ρ violates one of the conditions (2.3) and (2.4). If α_k/ρ does not satisfy (2.3), we have

$$f(x_k + (\alpha_k/\rho)d_k) > f(x_k) + \mu(\alpha_k/\rho)[g_k^\top d_k - c(\alpha_k/\rho)\|d_k\|^2].$$

Using the mean value theorem in the above inequality, we obtain $\theta_k \in (0, 1)$, such that

$$g(x_k + \theta_k(\alpha_k/\rho)d_k)^\top d_k > \mu[g_k^\top d_k - c(\alpha_k/\rho)\|d_k\|^2].$$

Subtracting $g_k^\top d_k$ in both sides of the above inequality, we obtain

$$(g(x_k + \theta_k(\alpha_k/\rho)d_k) - g_k)^\top d_k > -(1-\mu)g_k^\top d_k - c(\alpha_k/\rho)\mu\|d_k\|^2,$$

which, together with (H1) shows that

$$L\theta_k(\alpha_k/\rho)\|d_k\|^2 > -(1-\mu)g_k^\top d_k - c(\alpha_k/\rho)\mu\|d_k\|^2.$$

Therefore

$$\alpha_k > \frac{(1-\mu)\rho}{L + c\mu}\frac{|g_k^\top d_k|}{\|d_k\|^2}. \tag{3.6}$$

If α_k/ρ does not satisfy (2.4), we have

$$\begin{aligned} 0 &\leq -\|g(x_k+(\alpha_k/\rho)d_k)\|^2 + \frac{g(x_k+(\alpha_k/\rho)d_k)^\top(g(x_k+(\alpha_k/\rho)d_k)-g_k)}{u_1\|g_k\|^2-u_2g_k^\top d_k} \\ &\quad \times g(x_k+(\alpha_k/\rho)d_k)^\top d_k \\ &\leq -\|g(x_k+(\alpha_k/\rho)d_k)\|^2 + \frac{\|g(x_k+(\alpha_k/\rho)d_k)\|^2}{u_1\|g_k\|^2-u_2g_k^\top d_k} L(\alpha_k/\rho)\|d_k\|^2. \end{aligned}$$

Dividing both sides of the above inequality by $\|g(x_k+(\alpha_k/\rho)d_k)\|^2$ yields

$$\alpha_k \geq \frac{\rho}{L}\frac{u_1\|g_k\|^2-u_2g_k^\top d_k}{\|d_k\|^2}. \tag{3.7}$$

Letting

$$M_1 = \frac{(1-\mu)\rho}{L+c\mu},\ M_2 = \min\{c, \frac{\rho}{L}\},$$

from (3.5), (3.6) and (3.7), we have

$$\alpha_k \geq M_1\frac{|g_k^\top d_k|}{\|d_k\|^2} \text{ or } \alpha_k \geq M_2\frac{u_1\|g_k\|^2-u_2g_k^\top d_k}{\|d_k\|^2},$$

and then we have two possible cases. The first case is the set $K_1 := \{k|\alpha_k \geq M_1|g_k^\top d_k|/\|d_k\|^2\}$ is infinite. From (1.6) and (3.1), there exists a constant $M > 0$ such that for all $k \in K_1$,

$$\|d_k\| \leq M. \tag{3.8}$$

From (3.2) and (3.8), we have

$$\lim_{k\in K_1, k\to\infty} \alpha_k\|d_k\|^2 = 0, \tag{3.9}$$

then by $\alpha_k \geq M_1|g_k^\top d_k|/\|d_k\|^2$ we have

$$\lim_{k\in K_1, k\to\infty} |g_k^\top d_k| = 0. \tag{3.10}$$

On the other hand, from (1.3), we have

$$g_k^\top d_k = -\|g_k\|^2 + \frac{g_k^\top(g_k-g_{k-1})}{u_1\|g_{k-1}\|^2-u_2g_{k-1}^\top d_{k-1}}g_k^\top d_{k-1},$$

i.e.,

$$\|g_k\|^2 = -g_k^\top d_k + \frac{g_k^\top(g_k-g_{k-1})}{u_1\|g_{k-1}\|^2-u_2g_{k-1}^\top d_{k-1}}g_k^\top d_{k-1}. \tag{3.11}$$

By (H1) we have

$$\|g_k\|^2 \leq |g_k^\top d_k| + \frac{\|g_k\|^2L\alpha_{k-1}\|d_{k-1}\|^2}{u_1\|g_{k-1}\|^2-u_2g_{k-1}^\top d_{k-1}}. \tag{3.12}$$

If $u_1 > 0$, from (1.6), (3.4), (3.12) and (H1), we have

$$\|g_k\|^2 \leq |g_k^\top d_k| + \frac{\|g_k\|^2L\alpha_{k-1}\|d_{k-1}\|^2}{u_1\|g_{k-1}\|^2} \leq |g_k^\top d_k| + \frac{\tau^2L\alpha_{k-1}\|d_{k-1}\|^2}{u_1\varepsilon^2}.$$

From (3.9), (3.10), and taking limits for the above inequality, we have

$$\lim_{k\in K_1,k\to\infty} \|g_k\| = 0,$$

which contradicts (3.4). If $u_1 = 0$, then $u_2 > 0$. From (3.4), (3.11), we have

$$\begin{aligned}
-g_k^\top d_k &= \|g_k\|^2 - \beta_k g_k^\top d_{k-1} \\
&\geq \|g_k\|^2 - \frac{\|g_k\|^2\|g_k - g_{k-1}\|}{-u_2 g_{k-1}^\top d_{k-1}}\|d_{k-1}\| \\
&\geq \|g_k\|^2 - \frac{L\alpha_{k-1}\|g_k\|^2}{-u_2 g_{k-1}^\top d_{k-1}}\|d_{k-1}\|^2 \\
&\geq \|g_k\|^2 - \frac{Ls_{k-1}\|g_k\|^2}{-u_2 g_{k-1}^\top d_{k-1}}\|d_{k-1}\|^2 \\
&= \|g_k\|^2 - \frac{L\|g_k\|^2}{-u_2 g_{k-1}^\top d_{k-1}}\frac{c(-u_2 g_{k-1}^\top d_{k-1})}{\|d_{k-1}\|^2}\|d_{k-1}\|^2 \\
&= \|g_k\|^2 - Lc\|g_k\|^2 \geq (1 - Lc)\varepsilon^2 := \varpi
\end{aligned}$$

Because we have assumed $Lc < 1$ for $u_1 = 0$, then $\varpi > 0$. The above inequality contradicts (3.10).

Consider the second case that $K_2 := \{k|\alpha_k \geq M_2(u_1\|g_k\|^2 - u_2 g_k^\top d_k)/\|d_k\|^2\}$ is an infinite set. If $u_1 = 0, u_2 > 0$, then $\alpha_k \geq M_2 u_2|g_k^\top d_k|/\|d_k\|^2$, and its proof is similar to the first case. If $u_1 > 0$, then $\alpha_k \geq M_2 u_1\|g_k\|^2/\|d_k\|^2$, i.e.,

$$\|g_k\|^2 \leq \frac{1}{M_2 u_1}\alpha_k\|d_k\|^2,$$

which combing with (3.9) yields a contradiction to (3.4). The proof is complete. □

4. Numerical results

In this section, we provide the implementation details of Algorithm 2.1 to verify its efficiency. The codes were written in Matlab 7.1 and run on a portable computer. For each problem, the limiting number of function evaluations is set to 10000. 'F' means the method failed.

Our numerical results are listed in the form NI/NF/NG, where the symbols NI, NF and NG mean the number of iterations, the number of function evaluations and the gradient evaluations, respectively. For PRP method, FR method, HS method and DY method, we use the same Armijo-line search as Algorithm 2.1. The parameters in the Armijo-line search were chosen to be $c = 1, u_1 = 0.3, u_2 = 0.7, \rho = 0.5$ and $\mu = 0.4$. For each test problem, the stopping criterion is

$$\|g_k\| \leq 10^{-5}.$$

Problem 4.1.

$$f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5(x_1 + x_2) - 21x_3 + 7x_4, x_1 = (1, 1, 1, 1)^\top.$$

Problem 4.2.

$$f(x) = (1 - x_1)^2 + (1 - x_{10})^2 + \sum_{i=1}^{9} (x_i^2 - x_{i+1})^2, x_1 = (-2, \cdots, -2)^\top.$$

Problem 4.3.

$$f(x) = e^{x_1} + x_1^2 + 2x_1x_2 + 4x_2^2, x_1 = (1, 1)^\top.$$

Problem 4.4.

$$f(x) = \sum_{i=1}^{n} (e^{x_i} - x_i), x_1 = (n/(n-1), \cdots, n/(n-1))^\top.$$

P	n	PRP	FR	HS	DY	Algorithm 2.1
P1	4	22/67/67	19/56/56	F	14/40/40	22/67/67
P2	10	353/702/702	F	F	F	308/530/530
P3	2	19/65/65	17/53/53	F	19/60/60	18/61/61
P4	10	16/35/35	23/47/47	47/2537/2537	23/47/47	15/34/34
	100	17/37/37	25/51/51	29/1422/1422	25/51/51	16/36/36
	500	18/39/39	26/53/53	30/1394/1394	27/55/54	17/39/38

Table 1. Numerical results of Problems 4.1-4.4

From Table 4 we can see that the average performances of the PRP method is better than FR and DY methods, and much better than the HS method, and the average performances of Algorithm 2.1 is a little better than the PRP method.

5. Conclusion

In this paper, we have proposed a new descent conjugate gradient method for solving unconstrained optimization problems. The new method is a modification of Shi and Guo's method by dropping the local Lipschitz constant. The global convergence result of the improved method is established under some mild conditions. Preliminary numerical results show that the performance of the improved method is a little more efficient than the PRP method for given test problems.

Acknowledgements

The authors are grateful to the referees for the insightful suggestions improving the presentation of this paper.

References

[1] Z. J. Shi and J. H. Guo, A new family of conjugate gradient methods, *J. Comput. Appl. Math.* **224** (2009), 444-457.
[2] G. H. Yu, L. T. Guan and Z. X. Wei, A globally convergent Polak-Ribiére-Polyak conjugate gradient method with Armijo-type line search, *Numer. Math. J. Chin. Univ. (Engl. Ser.)* **15**(4) (2006), 357-366.

[3] Z. J. Shi, A new memory gradient method under exact line search,*Asia-Pacific J. Oper. Res.* **20** (2003), 275–284.

[4] L. Zhang, W. J. Zhou, and D. H. Li, A descent modified Polak-Ribiére-Polyak conjugate gradient method and its global convergence, *IMA J. Numer. Anal.* **26** (2006), 629-640.

[5] W. Y. Cheng, A two-term PRP-based descent method, *Numer. Func. Anal. Optim.* **28** (2007), 1217-1230.

[6] Z. J. Shi and J. Shen, Convergence of Liu-Storey conjugate gradient method, *Euro. J. Oper. Res.* **182** (2007), 552-560.

[7] Z. J. Shi and J. H. Guo, Global convergence of conjugate gradient method, *Optimization* **58**(2) (2009), 163-179.

[8] Z. F. Li, J. Chen and N. Y. Deng, A new conjugate gradient method and its global convergence properties, *Systems Sci. Math. Sci.* **11**(1) (1998), 53-60.

[9] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.* **2** (1992), 21-42.

[10] L. Zhang, A new Liu-Storey type nonlinear conjugate gradient method for unconstrained optimization method. Journal of Computational and Applied Mathematics, 2009, 225: 146-157.

[11] Z. X. Wei, G. Li , and L. Qi, Global convergence of the Polak-Ribiére-Polyak conjugate gradient method with an Armijo-type inexact line search for nonconvex unconstrained optimization problems, *Math. Comp.* **77** (2008), 2173-2193.

[12] W. W. Hager and H. C. Zhang, A survey of nonlinear conjugate gradient methods, *Pac. J. Optim.* **2** (2006), 35-58.

School of Mathematics and Statistics,
Zaozhuang University, Shandong, 277160, China.
E-mail address: `sunmin_2008@yahoo.com.cn`

School of Mathematics and Statistics,
Zhejiang University of Finance and Economics,
Hangzhou, 310018, China.
E-mail address: `ljlj8899@163.com`