Applied Mathematics E-Notes, 11(2011), 139-147 © ISSN 1607-2510
Available free at mirror sites of http://www.math.nthu.edu.tw/∼amen/

A Conjugate Gradient Method With Sufficient Descent And Global Convergence For Unconstrained Nonlinear Optimization*

Hailin Liu†, Sui Sun Cheng‡, Xiaoyong Li§

Received 9 March 2011

Abstract

In this paper a new conjugate gradient method for unconstrained optimization is introudced, which is sufficient descent and globally convergent and which can also be used with the Dai-Yuan method to form a hybrid algorithm. Our methods do not require the strong convexity condition on the objective function. Numerical evidence shows that this new conjugate gradient algorithm may be considered as one of the competitive conjugate gradient methods.

1 Introduction

There are now many conjugate gradient schemes for solving unconstrained optimization problems of the form

$$\min \{f(x) : x \in R^n\}$$

where f is a continuously differentiable function of n real variables with gradient $g = \nabla f$. An essential feature of these schemes is to arrive at the desired extreme points through the following nonlinear conjugate gradient algorithm

$$x^{(k+1)} = x^{(k)} + \alpha_k d_k \tag{1}$$

where α_k is the stepsize, and d_k is the conjugate search direction defined by

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

where $g_k = \nabla f\left(x^{(k)}\right)$ and β_k is an update parameter. In a recent survey paper by Hager and Zhang [5], a number of choices of the parameter β_k are given in chronological order. Two well known choices are recalled here for later use:

*Mathematics Subject Classifications: 15A99, 39A10.

†School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, Guangdong 510665, P. R. China

‡Department of Mathematics, Tsing Hua University, Taiwan 30043, R. O. China

§School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, Guangdong 510665, P. R. China

$$\text{Dai-Yuan: } \beta_k^{DY} = \frac{||g_k||^2}{d_{k-1}^T y_k} \tag{3}$$

$$\text{Hager-Zhang: } \beta_k^{HZ} = \frac{1}{d_{k-1}^T y_k} \left(y_k - 2d_{k-1} \frac{||y_k||^2}{d_{k-1}^T y_k} \right)^T g_k, \tag{4}$$

where $||\cdot||$ is the Euclidean norm and $y_k = g_k - g_{k-1}$. In (1) the stepsize α_k is obtained through the exact linear search (i.e., $g\left(x^{(k)} + \alpha_k d_k\right)^T d_k = 0$) or inexact linear search with Wolfe's criterion defined by

$$f(x^{(k)} + \alpha_k d_k) \le f(x^{(k)}) + \rho \alpha_k g_k^T d_k, \tag{5}$$

and

$$g(x^{(k)} + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k \tag{6}$$

where $0 < \rho < \sigma < 1$.

In 1999, Dai and Yuan [2] proposed the DY conjugate gradient method using β_k defined by (3). In 2001 [1], they introduced an updated formula of β_k with three parameters, which may be regarded as a convex combination of several earlier choices of β_k listed in [5]; but the three parameters are restricted in small intervals. Based on the ideas of Dai-Yuan, Andrei in [3] presents yet another sufficient descent and global convergence algorithm that avoided the strongly convex condition on the objective function $f(x)$ assumed by Hager and Zhang [4] incoporating β_k^{HZ} in (4) (to be named the HZ method in the sequel).

However, the method by Andrei requires some additional conditions (see the statement following the proof of Theorem 1, and also the additional conditions such as $g_{k+1}^T(g_{k+1} - g_k) > 0$ and $0 < \omega \le \theta_k \le \Omega$ in Theorem 2 of [3]). Therefore it is of interest to find further alternate methods that are as competitive, yet neither the strong convexity of the objective function nor the above mentioned conditions are required.

In this note, we introduce a new formulation of the update parameter β_k defined by

$$\beta_k^{NEW} = \frac{||g_k||^2}{\mu |d_{k-1}^T g_k| + d_{k-1}^T y_k}. \tag{7}$$

Note that if we use the exact line search, our new algorithm reduces to the algorithm of Dai and Yuan. In this paper, however, we consider general nonlinear functions and an inexact line search. By means of our β_k^{NEW} and the β_k^{DY} in (3), we may then introduce a hybrid algorithm for finding the extreme values of f.

Global convergence of our methods will be established and numerical evidence will be listed to support our findings.

2 New Algorithm and Convergence

As in [2], we assume that the continuously differentiable function f is bounded in the level set $L_1 = \{x | f(x) \le f(x^{(1)})\}$, where $x^{(1)}$ is the starting point; and that $g(x)$ is Lipschitz continuous in L_1, i.e., there exists a constant $L > 0$ such that $||g(x) - g(y)|| \le$

$L||x-y||$ for all $x, y \in L_1$. We remark that in Andrei [3], it is required that the level set L_1 be bounded instead of the slightly weaker condition of Dai-Yuan.

Also, we use the same algorithm in [2] which is restated here for the sake of convenience:

Step 1. Initialize starting point $x^{(1)}$, and $\mu > 1$, a very small positive $\varepsilon > 0$. Compute $d_1 = -g_1$. Set $k = 1$.

Step 2. If $||g_k|| < \varepsilon$, then stop and output $x^{(k)}$, else go to step 3.

Step 3. Compute $x^{(k+1)} = x^{(k)} + \alpha_k d_k$ through inexact linear search by (5) and (6).

Step 4. Compute d_{k+1} by (2) and (7). Compute g_{k+1}. Set $k = k+1$ and go to step 2.

In order to consider convergence, we first notice, by (6), that

$$d_{k-1}^T(g_k - g_{k-1}) \geq \sigma d_{k-1}^T g_{k-1} - d_{k-1}^T g_{k-1} = (\sigma - 1) d_{k-1}^T g_{k-1}. \tag{8}$$

LEMMA 1. If $\mu > 1$, then $g_k^T d_k < -(1 - \frac{1}{\mu})||g_k||^2 < 0$ for $k = 1, 2, \ldots$.

PROOF. If $k = 1$ then $d_1 = -g_1$ and $g_1^T d_1 = -||g_1||^2 < -(1 - \frac{1}{\mu})||g_1||^2 < 0$ since $\mu > 1$. Assume by induction that $g_{k-1}^T d_{k-1} < -(1 - \frac{1}{\mu})||g_{k-1}||^2 < 0$. By (2), (6), (7) and (8), we have

$$\begin{aligned}
g_k^T d_k &= -||g_k||^2 + \beta_k^{NEW} g_k^T d_{k-1} = -||g_k||^2 + \frac{||g_k||^2}{\mu|d_{k-1}^T g_k| + d_{k-1}^T(g_k - g_{k-1})} g_k^T d_{k-1} \\
&\leq -||g_k||^2 + \frac{||g_k||^2}{|\mu|d_{k-1}^T g_k| + d_{k-1}^T(g_k - g_{k-1})|} |g_k^T d_{k-1}| \\
&\leq -||g_k||^2 + \frac{||g_k||^2}{|\mu|d_{k-1}^T g_k| + (\sigma - 1) d_{k-1}^T g_{k-1}|} |g_k^T d_{k-1}| \\
&\leq -||g_k||^2 + \frac{||g_k||^2}{\mu|d_{k-1}^T g_k|} |g_k^T d_{k-1}| \\
&= -\left(1 - \frac{1}{\mu}\right) ||g_k||^2.
\end{aligned}$$

The proof is complete.

We remark that our Lemma 1 implies that d_k is a sufficient descent direction.

LEMMA 2 (see [2]). If the sequence $\{x^{(k)}\}$ is generated by (1) and (2), the stepsize α_k satisfies (5) and (6), and d_k is a descent direction, f is bounded and $g(x)$ is Lipschitz in the level set, then

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{||d_k||^2} < \infty. \tag{9}$$

THEOREM 1 (Global convergence). If $\mu > 1$ in (7), f is bounded and $g(x)$ is Lipschitz in the level set, then our algorithm either terminates at a stationary point or $\liminf ||g_k|| = 0$.

Proof. If our conclusion does not hold, then there exists a real number $\varepsilon > 0$ such that $||g_k|| > \varepsilon$, for all $k = 1, 2, \dots$. Since $d_k + g_k = \beta_k d_{k-1}$, we have

$$||d_k||^2 = \beta_k^2 ||d_{k-1}||^2 - ||g_k||^2 - 2g_k^T d_k. \tag{10}$$

By (8) and Lemma 1, we have

$$\begin{aligned} g_k^T d_k &= -||g_k||^2 + \beta_k^{NEW} g_k^T d_{k-1} \\ &= -||g_k||^2 + \frac{||g_k||^2}{\mu |d_{k-1}^T g_k| + d_{k-1}^T (g_k - g_{k-1})} g_k^T d_{k-1} \\ &= \frac{-\mu |d_{k-1}^T g_k| + d_{k-1}^T \ g_{k-1}}{\mu |d_{k-1}^T g_k| + d_{k-1}^T (g_k - g_{k-1})} ||g_k||^2 \\ &\leq \frac{d_{k-1}^T g_{k-1}}{\mu |d_{k-1}^T g_k|} ||g_k||^2. \end{aligned}$$

Since $d_{k-1}^T g_{k-1} < 0$ and $d_k^T g_k < 0$, we see that

$$||g_k||^2 \leq \frac{\mu |d_{k-1}^T g_k| |g_k^T d_k|}{|d_{k-1}^T g_{k-1}|},$$

that is,

$$\beta_k^{NEW} = \frac{||g_k||^2}{\mu |d_{k-1}^T g_k| + d_{k-1}^T (g_k - g_{k-1})} \leq \frac{||g_k||^2}{\mu |d_{k-1}^T g_k|} \leq \frac{|d_k^T \ g_k|}{|d_{k-1}^T \ g_{k-1}|}.$$

Replace β_k in (10) with β_k^{NEW}, we get

$$\begin{aligned} \frac{||d_k||^2}{(g_k^T d_k)^2} &\leq \frac{||d_{k-1}||^2}{(g_{k-1}^T d_{k-1})^2} - \frac{||g_k||^2}{(g_k^T d_k)^2} - 2\frac{1}{g_k^T d_k} \\ &= \frac{||d_{k-1}||^2}{(g_{k-1}^T d_{k-1})^2} - (\frac{||g_k||}{g_k^T d_k} + \frac{1}{||g_k||})^2 + \frac{1}{||g_k||^2} \\ &\leq \frac{||d_{k-1}||^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{||g_k||^2} \leq \frac{||d_{k-1}||^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\varepsilon^2} \end{aligned}$$

since $d_1 = -g_1$, so that

$$\frac{||d_k||^2}{(g_k^T d_k)^2} < \frac{||d_1||^2}{(g_1^T d_1)^2} + \frac{k-1}{\varepsilon^2} = \frac{1}{||g_1||^2} + \frac{k-1}{\varepsilon^2} < \frac{1}{\varepsilon^2} + \frac{k-1}{\varepsilon^2} = \frac{k}{\varepsilon^2}.$$

Thus

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{||d_k||^2} > \sum_{k=1}^{\infty} \frac{\varepsilon^2}{k} = +\infty$$

which is contrary to Lemma 2. The proof is complete.

3 Hybrid Algorithm

We may build a hybrid algorithm (see discussions on hybrid algorithms in [5] for background information) based on β_k^{DY} and our β_k^{NEW} as follows: First we let

$$\beta_k^{mix} = \begin{cases} \beta_k^{DY} & \text{if } |\beta_k^{DY}| \leq \beta_k^{NEW} \text{ and } g_k^T d_{k-1} < 0 \\ \beta_k^{NEW} & \text{otherwise} \end{cases}, \tag{11}$$

and then we replace β_k^{NEW} with β_k^{mix} at step 4 of the algorithm in the last section.

THEOREM 2. For $k = 1, 2, ...$, we have $g_k^T d_k < -(1 - \frac{1}{\mu})||g_k||^2$ (so that our new method is also sufficient descent).

PROOF. When $n = 1$, since $\mu > 1$ and $d_1 = -g_1$, we have $g_1^T d_1 = -||g_1||^2 < -(1 - \frac{1}{\mu})||g_1||^2 < 0$. Assume by induction that $g_{k-1}^T d_{k-1} < -(1 - \frac{1}{\mu})||g_{k-1}||^2 < 0$. If $\beta_k^{mix} = \beta_k^{DY}$, then $g_k^T d_{k-1} \geq 0$. Therefore, in case where $\beta_k^{mix} = \beta_k^{DY}$ or $\beta_k^{mix} = \beta_k^{NEW}$, we have

$$g_k^T d_k = -||g_k||^2 + \beta_k^{mix} g_k^T d_{k-1} \leq -||g_k||^2 + \beta_k^{NEW} g_k^T d_{k-1}.$$

From the proof of Lemma 1, we can then get $g_k^T d_k < -(1 - \frac{1}{\mu})||g_k||^2$.

THEOREM 3. (Global convergence). If $\mu > 1$, f is bounded and $g(x)$ is Lipschitz in the level set, then our algorithm either terminates at a stationary point or $\liminf ||g_k|| = 0$.

Proof. As in the proof of Theorem 1, if our conclusion does not hold, then we have

$$\begin{aligned} g_k^T d_k &= -||g_k||^2 + \beta_k^{mix} g_k^T d_{k-1} \\ &\leq -||g_k||^2 + \beta_k^{NEW} g_k^T d_{k-1} \\ &= \frac{-\mu |d_{k-1}^T g_k| + d_{k-1}^T \ g_{k-1}}{\mu |d_{k-1}^T g_k| + d_{k-1}^T (g_k - g_{k-1})} ||g_k||^2 \\ &\leq \frac{||g_k||^2}{\mu |d_{k-1}^T g_k| + d_{k-1}^T (g_k - g_{k-1})} d_{k-1}^T \ g_{k-1} \\ &= \beta_k^{NEW} d_{k-1}^T g_{k-1}. \end{aligned}$$

Since $g_k^T d_k < 0$ for all $k \geq 1$, therefore,

$$\beta_k^{NEW} \leq \frac{g_k^T d_k}{d_{k-1}^T g_{k-1}} = \frac{|g_k^T d_k|}{|d_{k-1}^T g_{k-1}|}. \tag{12}$$

On the other hand, by (12) and (10), we have

$$\begin{aligned} ||d_k||^2 &= (\beta_k^{mix})^2 ||d_{k-1}||^2 - ||g_k||^2 - 2 g_k^T d_k \\ &\leq (\beta_k^{NEW})^2 ||d_{k-1}||^2 - ||g_k||^2 - 2 g_k^T d_k \\ &\leq \frac{(g_k^T d_k)^2 ||d_{k-1}||^2}{(g_{k-1}^T d_{k-1})^2} - ||g_k||^2 - 2 g_k^T d_k \end{aligned}$$

The remaining proof is the same as the proof of Theorem 1.

4 Numerical Evidences

In this section, we will test the DY, HZ, the ANDREI (see [3]) and our NEW as well as HYBRID conjugate methods with weak Wofle line search. For each method, we take $\rho = 0.2, \sigma = 0.3, \mu = 1.1$ in (5)-(6), and the termination condition is $\|g_k\| \leq \varepsilon = 10^{-6}$. The test problems are extracted from [6]. Since the computational procedures are similar to those in [6] and in [7], we will not bother with the detailed descriptions of the numerical data. Instead, we prepare a Table which provides conclusions of our numerical comparisons. More specifically, in this table, the terms Problem, Dim, NI, NF, NG, -, * have the following meaning:

Problem: the name of the test problem;
Dim: the dimension of the problem;
NI: the total number of iterations;
NF: the number of the function evaluations;
NG: the number of the gradient evaluations;
-: method not applicable;
*: the best method.

Our Table (see the last two pages) shows that our new methods in some test problems outperform the Dai-Yuan method. Although the HZ method is also performing well, but this method requires that the objective function is strongly convex and the level set is bounded, so HZ method may not be applicable (such as the Gulf research problem). In conclusion, our new methods are competitive among the well known conjugate gradient methods for unconstrained optimization.

Acknowledgment. Research supported by the Natural Science Foundation of Guangdong Province of P. R. China under grant number 9151008002000012.

References

[1] Y. H. Dai and Y. Yuan, A Three-parameter family of nonlinear conjugate gradient methods, Math. Comp., 70(2001), 1155-1167.

[2] Y. H. Dai and Y. Yuan, A nonlinear conjugate gradient with a strong global convergence property, SIAM Journal of Optimization, 10(1999), 177–182.

[3] N. Andrei, A Dai-Yuan conjugate gradient algorithm with sufficient descent and conjugacy conditions for unconstrained optimization, Appl. Math. Lett., 21 (2008) 165–171

[4] W. W. Hager and H. C. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, Siam J. Optim., 16(1)(2005), 170–192.

[5] W. W. Hager and H. C. Zhang, A survey of nonlinear conjugate gradient methods. Pac. J. Optim., 2(1)(2006), 35–58.

		DY	NEW	HYBRID	HZ	ANDREI
Problem	Dim	NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG
Gaussian	3	4/7/5	4/7/5	4/7/5*	5/7/5	4/9/5
Linear-full rank	10	3/27/3	3/27/3	3/27/3*	3/27/3	4/26/12
	10000	3/27/8	3/27/8	3/27/8*	3/27/8	4/26/15
Linear-rank 1	10	4/33/14	4/33/12	4/33/12*	5/38/12	6/42/19
	10000	4/38/3	4/38/3	4/38/3*	4/38/3	5/51/7
Broden tridiagonal	10	31/94/31	26/79/26	26/79/26*	27/81/27	47/146/50
	10000	91/275/94	65/195/67	64/194/67	72/219/76	54/171/63*
Broden banded	10	16/51/18	18/58/22	14/45/16*	20/66/23	37/121/43
	10000	22/66/23	18/55/79*	22/66/23	21/67/23	37/112/43
Dis. boundary value	1000	2/22/1	2/22/1	2/22/1*	2/22/1	2534/10003/2602
	10000	2/22/1	2/22/1	2/22/1*	2/22/1	–
Dis. integral equation	10	4/29/19	5/27/4	5/27/4	4/26/3*	7/40/22
	10000	5/31/7	5/27/4	5/27/4	4/26/3*	6/32/18
Box 3-dimensional	3	2364/7121/2386	136/257/165	58/123/73	39/96/56*	–
Rosenbrock	2	53/180/61	96/325/123	41/153/69	36/127/51*	102/482/115
Variably dimensioned	200	8/42/11	8/42/11	8/42/11*	8/46/12	12/86/24
	10000	15/143/28*	16/151/31	17/160/33	18/173/37	20/198/44
Watson	31	2507/10002/2508	2504/10003/2503	2507/10001/2506	2592/10002/2594	294/1868/311*
Penalty I	500	37/85/56	38/77/63	32/73/51	29/64/46	20/93/36*
	10000	314/1211/505	67/147/108	53/144/108	26/68/42*	692/3405/737

		DY	NEW	HYBRID	HZ	ANDREI
Problem	Dim	NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG
Penalty II	100	102/324/117	76/214/87*	78/229/92	89/263/103	–
	10000	–	–	–	–	–
Brown badly scaled	2	1667/10003/1676	1668/10005/1674	1666/10001/1677	15/92/21	4/16/4*
Brown and Dennis	4	63/287/81	37/181/57*	57/270/83	41/212/72	50/288/80
Gulf research	3	316/970/327	944/2836/946	764/2302/764	–	84/543/99*
Trigoonometric	10	78/151/80	41/58/54	41/58/50*	49/65/60	394/1839/411
	10000	225/473/248	70/88/85	65/67/67*	79/119/104	53/156/76
Ext. Rosenbrock	500	57/191/65	120/397/147	44/161/72	27/130/52*	1611/10000/1613
	10000	57/191/65	141/460/168	48/171/76	39/127/53*	74/319/72
Ext. Powell singular	20	3225/9678/3229	3183/8548/3187	3334/10002/3337	104/290/115*	1456/10002/1462
	10000	3334/10002/3337	3336/10002/3339	3334/10002/3337	152/414/164	56/231/72*
Beale	2	17/42/20	15/38/20*	27/69/29	16/42/21	50/151/65
Wood	4	2327/9399/2373	113/374/123	80/252/85*	127/394/133	711/4587/733
Chebyquad	20	2483/10002/2526	166/522/172	153/461/155*	155/473/162	484/1939/489
	10000	2/22/1	2/22/1	2/22/1*	2/22/1	2/22/1
Brown almost-linear	20	6/34/19	14/57/27	6/34/19*	6/36/19	7/44/26
	10000	6/67/28	6/87/30	6/67/28	6/64/19*	7/61/29

[6] N. Andrei, Test functions for unconstrained optimization, http://www.ici.ro/neculai/SCALCG/evalfg.for.

[7] W. W. Hager and H. Zhang, Algorithm 851: CG DESCENT, a conjugate gradient method with guaranteed descent, ACM Trans. Math. Software, 32(2006), 113–137.