

A Variant Hybrid Conjugate Gradient Algorithm for Large-Scale Unconstrained Optimization

Zhijun Luo^{1,a}, Lirong Wang^{2,b} and Guohua Chen^{3,c}

¹School of Mathematics and Finance, Hunan University of Humanities, Science and Technology, Loudi, 417000, P.R.China

²School of Information science, Hunan University of Humanities, Science and Technology, Loudi, 417000, P.R.China

^aldljz123@163.com, ^bldwlr1234@163.com (Corresponding author), ^chnlcdgh@163.com

Keywords: Conjugate gradient; Unconstrained optimization; Global convergence

Abstract. In this paper, we have presented a new hybrid conjugate gradient algorithm for solving unconstrained optimization problems. The parameter b is a convex combination of the PRP and FR conjugate gradient methods. Under general wolfe line search conditions, we proved the global convergence of the algorithm. The numerical results show that the proposed methods are effective.

Introduction

Many practical problems in engineering can be translated into the following unconstrained optimization problem

$$\min f(x), x \in R^n \quad (1)$$

where R^n denotes an n -dimensional Euclidean space and $f(x)$ is continuously differentiable function. To solve the problem (1), starting from an initial point $x_0 \in R^n$, the conjugate gradient method generates a sequence $\{x_k\} \subset R^n$ such that $x_{k+1} = x_k + a_k d_k$, $k = 0, 1, 2, \dots$, where $a_k > 0$ is a step size obtained by line search, and the directions d_k are generated as

$$d_k = \begin{cases} -g_k, & \text{if } k = 0 \\ -g_k + b_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \quad (2)$$

where $g_k = \nabla f(x_k)$, and b_k is known as the conjugate gradient parameter. For general nonlinear functions, different choices of b_k lead to different conjugate gradient methods. There are some well-known formulas for b_k which are given below (see [1-6]):

$$b_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \quad b_k^{DY} = \frac{g_k^T g_k}{d_{k-1}^T y_{k-1}}, \quad b_k^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}, \quad b_k^{PRP} = \frac{g_k^T y_{k-1}}{g_{k-1}^T g_{k-1}}, \quad b_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \quad b_k^{LS} = -\frac{g_k^T y_{k-1}}{d_{k-1}^T g_{k-1}},$$

or by other formulae (where $y_{k-1} = g_k - g_{k-1}$, the symbol $\|\cdot\|$ stands for the Euclidean norm). In general, the FR, DY and CD methods have strong convergence properties, but they may have modest practical performance due to jamming for general objective function. On the other hand, the methods of PRP, LS and HS in general may not be convergent, but they often have better computational performances. In recent years, many researchers devoted to the hybrid or mixed conjugate gradient methods which have better computational performances and strong convergence properties. Such as, combining between PRP and DY conjugate gradient methods, N. Andrei [7] proposed the following hybrid method: $\beta = (1 - \theta)\beta^{PRP} + \theta\beta^{DY}$, where the parameter in the convex combination is computed in such a way that the conjugacy condition is satisfied, independently of the line search. In order to take advantage of the attractive features of the HS and DY conjugate gradient methods, [8] proposed two hybridizations methods based on Andrei's approach of hybridizing the CG parameters convexly and Powell's approach of nonnegative restriction of the CG parameters.

In this paper, we focus on hybrid conjugate gradient methods as a convex combination of the PRP and FR conjugate gradient methods. We selected these two methods to combine in a hybrid conjugate gradient algorithm because FR has strong convergence properties, on one side, and PRP has good computational properties, on the other side. In general PRP method performs better in practice than FR, and we consider this in order to have a good practical conjugate algorithm. Further, if the line

search fulfils the general wolfe conditions, we establish the global convergence of our method. Numerical experiments show that our proposed method is preferable and in general superior to the classical conjugate gradient methods in terms of efficiency.

Description of algorithm

Due to the parameter b_k is very important to analyze the global convergence of the conjugate gradient methods. In order to avoid the drawbacks of FR and PRP methods, in our algorithm, let $\beta = u\beta^{FR} + (1-u)\beta^{PRP}$, $0 \leq u \leq 1$, where u is a scalar parameter. Obviously, if $u = 1$, then $b = b^{PRP}$, and if $u = 0$, then $\beta = \beta^{FR}$. On the other hand, if $0 < u < 1$, then, β is a convex combination of β^{PRP} and β^{FR} . Hence, from $y_{k-1}^T d_k = 0$ (the conjugate condition), we get

$u_k = \frac{(y_{k-1}^T g_k)(\|g_{k-1}\|^2 - y_{k-1}^T d_{k-1})}{(g_k^T g_{k-1})(y_{k-1}^T d_{k-1})}$. Then, we constructed the following parameters

$$b_k = \begin{cases} u_k \frac{\|g_k\|^2}{\|g_{k-1}\|^2} + (1-u_k) \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} = u_k b^{FR} + (1-u_k) b^{PRP}, & \text{if } 0 \leq u < 1, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Now we state our algorithm as follows.

Algorithm A:

Step 0 Initialization: Given a starting point $x_0 \in R^n$, choose parameters $0 < e < 1$, $0 < d < \frac{1}{2}$,

$d < s_1 \leq s_2 < 1$. Set $k := 0$;

Step 1 If $\|g_k\| < e$, STOP, else go to Step 2;

Step 2 Compute the search direction d_k by (2) and generate b_k by (3)

Step 3 Compute step size a_k such that,

$$f(x_k + a_k d_k) \leq f(x_k) + d a_k g_k^T d_k, \quad (4)$$

$$s_1 g_k^T d_k \leq g_k^T (x_k + a_k d_k) \leq -s_2 g_k^T d_k. \quad (5)$$

Step 4 Let $x_{k+1} = x_k + a_k d_k$, $k := k + 1$, and go to Step 2

Remark:

1) If $u_k < 0$ or $u_k \geq 1$, then $b_k = 0$, the algorithm performs the restart strategy. i.e. let x_k be the initial point, select the steepest descent direction as the current search direction and reuse (3) to start the iteration.

2) If $u_k \in [0, 1)$, we can show that one of the following two groups of inequality holds, $g_k^T g_{k-1} \geq 0, y_{k-1}^T g_k \geq 0$, or $g_k^T g_{k-1} < 0, g_k^T d_{k-1} \leq 0$.

Global convergence of algorithm

The following basic assumptions on the objective function are assumed, which have been widely used in the literature to analyze the global convergence of the conjugate gradient methods. Throughout this paper, the symbol $\| \cdot \|$ denotes the Euclidean norm.

H3.1 i) The objective function $f(x)$ is continuously differentiable and has a lower bound on the level set $L_0 = \{x \in R^n \mid f(x) \leq f(x_0)\}$, where x_0 is the starting point.

ii) The gradient $g(x)$ of $f(x)$ is Lipschitz continuous in some neighborhood U of L_0 , namely, there exists a constant $L > 0$ such that $\|g(x) - g(y)\| \leq L \|x - y\|, \forall x, y \in U$.

Lemma 3.1[9] Suppose that Assumption H3.1 holds. Consider any method in the form (2), where a_k satisfies the generalized Wolfe conditions (5) and (4), and d_k is a descent direction (i.e. $g_k^T d_k < 0$).

Then we have that $\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty$.

Theorem 3.1 Suppose that Assumption H3.1 holds and the sequence $\{x_k\}$ is generated by Algorithm A, then $g_k^T d_k < 0$.

Theorem 3.2 Suppose that Assumption H3.1 holds and the sequence $\{x_k\}$ is generated by Algorithm A. If there exists a constant $M > 0$ such that $\|g_k\| \leq M$ for all $k \geq 1$, then $\liminf_{k \rightarrow \infty} \|g_k\| = 0$.

Numerical experiments

In this section, we give the numerical results of Algorithm A to show that the method is efficient for unconstrained optimization problems. The problems that we tested are from [10] and [11]. We stop the iteration if the inequality $\|g_k\| \leq 10^{-6}$ is satisfied. **Table 1** show the computation results, where the columns have the following meanings:

Dim—the dimension of the problems; NI—the number of iterations;

NF—the number of function evaluations; NG—the dimension of gradient evaluations;

NaN—means the number of iterations more than 5000 or the method fails.

Table 1 Comparative numerical results of Algorithm A, PRP and FR

Problem	Dim	Algorithm A	PRP	FR
		NI/NF/NG	NI/NF/NG	NI/NF/NG
Rosen	2	35/1068/90	758/28440/1408	NaN
Penalty1	4	4/52/9	5/54/11	7/172/34
	10	5/90/10	7/90/14	9/250/34
Variably	3	5/120/10	5/116/10	9/280/34
Trigonometric	2	14/210/27	24/344/49	107/3632/128
	100	23/180/84	107/1492/215	127/4262/180
	500	10/32/21	50/610/101	73/2402/89
Brown	2	14/210/27	24/344/49	269/8304/366
	25	6/240/7	6/240/9	11/440/12
Axis hyper	2	10/80/21	9/58/19	19/410/31
	500	162/5564/325	299/1686/479	23903/887840/45033
DeJong	2	8/34/17	2/12/5	13/270/41
	500	5/26/11	3/18/7	19/402/50
	5000	7/34/15	9/32/19	29/620/45
	30000	9/40/19	3/18/7	23/490/56
Rastrigin	2	10/336/24	19/690/33	13/446/28
	500	16/534/33	30/1024/59	222/8858/233
	5000	19/632/39	29/1042/49	93/3698/106
	30000	20/698/92	32/1096/63	109/4328/124

From Table 2, under the same computing environment, we see that the performance of the Algorithm A is better than the PRP, and the FR methods for some problems. Especially for solving

problems DeJong, Axis hyper and Rastrigin. Therefore our numerical experiments show that the algorithm is efficient.

Conclusion

In this paper, a new hybrid conjugate gradient algorithm is proposed in order to avoid the drawbacks of the PRP and FR conjugate gradient methods. The parameter b is a combination of the ideas of PRP and FR. The numerical results for some classical unconstrained optimization problems show that the proposed methods are effective.

Acknowledgements

This work was supported in part by Scientific Research Fund of Hunan Provincial Education Department (NO. 14C0609 and 16C0829).

References

- [1] R. Fletcher, C. Reeves. Function minimization by conjugate gradients, The computer journal. vol. 7, pp. 149-154, 1964.
- [2] E. Polak and G. Ribière, Note sur la convergence de méthodes de directions conjuguées, Revue Française de Recherche Opérationnelle, no. 16, pp. 35–43, 1969.
- [3] Y. H. Dai and Y. Yuan. A nonlinear conjugate gradient method with a strong global convergence property, SIAM Journal on Optimization, vol.10(1), pp. 177–182, 1999.
- [4] R. Fletcher. Unconstrained Optimization: Practical Methods of Optimization, vol. 1, John Wiley & Sons, New York, NY, USA, 1987.
- [5] Y. Liu and C. Storey. Efficient generalized conjugate gradient algorithms, Part 1, Journal of Optimization Theory and Applications, vol. 69, no. 1, pp. 129–137, 1991.
- [6] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems, Journal of Research of the National Bureau of Standards, vol. 49, pp. 409–436, 1952.
- [7] N. Andrei. A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan. Studies in Informatics and Control, vol.17, no. 4, pp. 55–70, 2008.
- [8] S. Babaie-Kafaki, R. Ghanbari. Two hybrid nonlinear conjugate gradient methods based on a modified secant equation, Optimization, vol. 63, no.7, pp.1027-1042, 2014.
- [9] X. W. Du, L. Q. Ye. “Global convergence of a class of unconstrained optimal methods included conjugate descent method,” Chinese Journal of Engineering Mathematics, Vol.18, no.2, pp. 119-121, 2001.
- [10] M. Molga, C. Smutnicki. Test functions for optimization needs. <http://www.zsd.ict.pwr.wroc.pl/files/docs/functions.pdf>, 2005
- [11] W. Hock and K. Schittkowski. Test examples for nonlinear programming codes. Journal of Optimization Theory and Applications, vol. 30, no. 1, pp. 127–129, 1981.