

A modified form of conjugate direction for general nonlinear function and its convergence

Sumit Saha and Biprajit Nath

*Department of Mathematics,
Assam University, Silchar,
India-788011*

Abstract

In this paper we have derived a new expression of the conjugate gradient parameter β for the nonlinear conjugate gradient method. The convergence of the CG method with the proposed β has been discussed. Also, the proposed β has been tested for a good number of nonlinear test functions.

AMS subject classification: 65, 49.

Keywords: Conjugate Gradient method, Line search, descent property, sufficient descent condition, global convergence.

1. Introduction

Conjugate Gradient (C.G) method has a very important role in the field of large-scale unconstrained optimization problems. In this paper we have obtained a new conjugate direction for general unconstrained nonlinear function with its convergence analysis. The result is then verified with different standard test functions and compared with other existing results.

$$\min \{f(x) : x \in \mathbb{R}^n\} \quad (1)$$

Where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function especially if the dimension n is large.

The conjugate gradient method to solve the general nonlinear problem defined by (1) is of the form

$$x_{k+1} = x_k + \alpha_k d_k \quad (2)$$

Where α_k is a step size obtained by a line search and d_k is the search direction defined by

$$d_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \beta_k d_{k-1}, & k \geq 1 \end{cases} \quad (3)$$

Where β_k is a parameter and g_k denotes $\nabla f(x_k)$ where the gradient of f at x_k is a row vector and g_k is a column vector.

Different C.G methods correspond to different choices for the scalar β_k are as follows

$$\beta_k^{HS} = \frac{g_{k+1}^T}{d_k^T y_k}$$

was proposed in 1952 in the original (linear) CG paper of Hestenes and Stiefel, in 1964 Fletcher and Reeves proposed the first nonlinear Conjugate Gradient method

$$\beta_k^{FR} = \frac{\|g_{k+1}^T\|^2}{\|g_k^T\|^2},$$

in 1967 Daniel proposed

$$\beta_k^D = \frac{g_{k+1}^T \nabla^2 f(x_k) d_k}{d_k^T \nabla^2 f(x_k) d_k},$$

which requires evaluation of the Hessian, $\nabla^2 f(x_k)$, Polak and Ribiere and by Polyak in 1969 proposed

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k^T\|^2},$$

in the year 1987

$$\beta_k^{CD} = \frac{\|g_{k+1}^T\|^2}{-d_k^T g_k},$$

was proposed by Fletcher, CD stands for ‘‘Conjugate Descent’’, a new beta

$$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k}$$

was proposed by Liu and Storeyin, 1991, Dai and Yuan proposed

$$\beta_k^{DY} = \frac{\|g_{k+1}^T\|^2}{d_k^T y_k}$$

in 1999,

$$\beta_k^N = \frac{1}{d_k^T y_k} \left(y_k - 2d_k \frac{\|y_k\|^2}{d_k^T y_k} \right)^T g_{k+1}$$

Conjugate direction for general nonlinear function and its convergence

was proposed by Hager and Zhang in 2005.

It is known from (2) and (3) that only the step size α_k and the parameter β_k remain to be determined in the definition of Conjugate Gradient method. In this case that if f is a convex quadratic, the choice of β_k should be such that the method (2)-(3) reduces to the linear Conjugate Gradient method if the line search is exact namely

$$\alpha_k = \operatorname{argmin} \{ f(x_k + \alpha d_k); \alpha > 0 \} \quad (4)$$

For non linear functions, different formulae for the parameter β_k result in different Conjugate Gradient methods and their properties can be significantly different. To differentiate the linear Conjugate Gradient method, sometimes we call the Conjugate Gradient method for unconstrained optimization by nonlinear Conjugate Gradient method. Meanwhile the parameter β_k is called Conjugate Gradient parameter. In general the nonlinear Conjugate Gradient method without restarts is only linearly convergent (Crowder and Wolfe [12]) while n -step quadratic convergence rate can be established if the method is restarted along the negative gradient every n -step. (See Cohen [9] and McCormick and Ritter [8]).

In 1964 the method has been extended to nonlinear problems by Fletcher and Reeves [14], which is usually considered as the first nonlinear Conjugate Gradient algorithm. Since then a large number of variations of Conjugate Gradient algorithms have been suggested. A survey on their definition including 40 nonlinear Conjugate Gradient algorithms for unconstrained optimization is given by Andrei [4]. Since the exact line search is usually expensive and impractical, the strong Wolfe line search is often considered the implementation of the nonlinear Conjugate Gradient methods. It aims to find a step size satisfying the strong Wolfe conditions.

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \rho \alpha_k g_k^T d_k \quad (5)$$

$$|g(x_k + \alpha_k d_k)^T| \leq -\sigma g_k^T d_k, \text{ where } 0 < \rho < \sigma < 1 \quad (6)$$

The strong Wolfe line search is often regarded as a suitable extension of the exact line search since it reduces to the latter. If σ is equal to zero, in practical computation a typical choice for σ that controls the inexactness of the line search is $\sigma = 0.1$. On the other hand general non linear function, one may be satisfied with a step size satisfying the standard Wolfe conditions, namely (5) and

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \text{ where } 0 < \sigma < 1 \quad (7)$$

As is well known the standard Wolfe line search is normally used in the implementation of Quasi-Newton methods, another important class of methods for unconstrained optimization. The work of Dai and Yuan indicates that the use of standard Wolfe line search is possible in the nonlinear Conjugate Gradient field. A requirement for an optimization method to use the above line searches is that, the search direction d_k must have descent property namely

$$g_k^T d_k < 0 \quad (8)$$

For Conjugate Gradient method, by multiplying (3) with g_k^T , we have

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

Thus if the line search is exact, we have $g_k^T d_k = -\|g_k\|^2$ since $g_k^T d_{k-1} = 0$. Consequently d_k is descent provided $g_k \neq 0$. In this paper we say that a Conjugate Gradient method is descent if (8) holds for all k and is sufficient descent if the sufficient descent condition

$$g_k^T d_k \leq -c \|g_k\|^2$$

holds for all k and some constant $c > 0$. However we have to point out that the borderlines between these Conjugate Gradient methods are not strict.

If $s_k = x_{k+1} - x_k$ and in the following $y_k = g_{k+1} - g_k$. Different Conjugate Gradient algorithms corresponds to different choices for the parameter β_k . Either of the following assumptions are often utilised in convergence analysis for C.G algorithms.

1.1 Lipschitz Assumption: In some neighbourhood N of the level set

$$\mathbb{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$$

The gradient $\nabla f(x)$ is Lipschitz continuous. That is there exist a constant $L < \infty$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \forall x, y \in N$$

1.2 Boundedness Assumption: The level set \mathbb{L} is bounded. That is there exist a constant $B < \infty$ such that

$$\|x\| \leq B \forall x \in \mathbb{L}$$

The conclusion of the following theorem often called the Zoutendijk condition, is used to prove the global convergence of the non linear CG methods; it was originally given by Zoutendijk [11] and Wolfe [10,13].

The objective of this paper is to investigate the conjugacy condition of the general nonlinear functions taking an arbitrary initial search direction although the convergence for general linear as well as nonlinear functions with an initial search direction already exist.

In this paper, we have obtain a new conjugacy condition of the general nonlinear functions taking an arbitrary initial search direction as discussed by Sumit Saha and Biprajit Nath [1]. Saha and Nath, have taken the initial direction as $d_0 = -g_0 + \gamma g_0$ instead of $d_0 = -g_0$ where $\gamma \in (0, 1)$. Then

$$d_k = \begin{cases} -g_k + \gamma g_k, & k = 0 \\ (-g_k + \gamma g_k) + \beta_k d_{k-1}, & k \geq 1 \end{cases}$$

Therefore,

$$d_k^T = \{(-g_k + \gamma g_k) + \beta_k d_{k-1}\}^T \quad (9)$$

$$\Rightarrow d_k^T g_{k-1} = -g_k^T g_{k-1} + \gamma g_k^T g_{k-1} + \beta_k d_{k-1}^T g_{k-1} \quad (10)$$

$$= (\gamma - 1) g_k^T g_{k-1} + \beta_k d_{k-1}^T g_{k-1} \quad (11)$$

Conjugate direction for general nonlinear function and its convergence

Again, from the findings of Dai, Liao [7]

$$d_k^T g_{k-1} = -t g_k^T s_{k-1} \quad (12)$$

$$\Rightarrow g_k^T g_{k-1} + \beta_k d_{k-1}^T g_{k-1} = -t g_k^T s_{k-1} \quad (13)$$

$$\Rightarrow d_{k-1}^T g_{k-1} = \frac{-t g_k^T s_{k-1} - (\gamma - 1) g_k^T g_{k-1}}{\beta_k} \quad (14)$$

$$\Rightarrow \beta_k = \frac{-t g_k^T s_{k-1} - (\gamma - 1) g_k^T g_{k-1}}{d_k^T g_{k-1}} \quad (15)$$

Where $t \geq 0$ is a scalar.

2. Convergence Analysis

From assumptions 1.1 and 1.2 we have $|g_k| \leq BL < \infty$. Also, from sufficient descent condition $g_k^T d_k \leq -c \|g_k\|^2$, holds for all k and some constant $c > 0$. From (15),

$$\beta_k d_k^T g_{k-1} + (\gamma - 1) g_k^T g_{k-1} = -t g_k^T s_{k-1} \quad (16)$$

$$\Rightarrow s_{k-1} = \left\{ \frac{\beta_k d_k^T g_{k-1} + (\gamma - 1) g_k^T g_{k-1}}{-t \|g_k\|^2} \right\} g_k \quad (17)$$

$$\Rightarrow |s_{k-1}| < \beta_k c BL + \frac{|1 - \gamma|}{t} \quad (18)$$

$$\Rightarrow |s_{k-1}| < \frac{|1 - \gamma| BL}{t} \quad (19)$$

Therefore, the CG method fails for $t = 0$ and converges for $t > |1 - \gamma|$.

Table 1: Experimental Results for various Test Functions

Problem	Dim	Proposed β (NI)	β^{FR} (NI)	β^{HS} (NI)
Rosenbrock	100/200/300/400/500	10/12/13/15/19	15/17/19/22/23	18/20/23/26/28
Sphere	100/200/300/400/500	11/12/14/18/20	13/16/17/18/19	14/16/18/19/23
Hyper Ellipsoid	100/200/300/400/500	19/22/24/28/31	20/25/29/33/39	20/24/27/20/35
Zakharov	100/200/300/400/500	20/23/27/29/34	21/26/28/32/37	19/25/29/36/41
Trigonometric	100/200/300/400/500	20/23/27/29/34	21/26/28/32/37	19/25/29/36/41
Extended Powell	100/200/300/400/500	54/63/71/88/97	61/77/86/95/108	69/82/98/104/112
Broyden tridiagonal	100/200/300/400/500	37/36/34/31/30	41/39/38/36/35	44/41/38/37/35
Broyden banded	100/200/300/400/500	21/22/26/27/28	/39/38/36/35/33	39/38/37/35/32

3. Result and Discussion

The experiment has been carried out in MATLAB 7 and the optimum value is correct upto three decimal places. NI indicates the number of iterations. In the experiments the value of γ is taken as in the interval (0,1) and close to 1. the value of t is taken from the interval [1,2].

From the above table it is clear that, with the new expression of β given in (15), the efficiency of the CG method improved in comparison with the other values of β .

References

- [1] Sumit Saha, Biprajit Nath, A new initial search direction for nonlinear conjugate gradient method, International Journal of Mathematics Research, ISSN 0976-5840 Volume 6, Number 2 (2014), pp. 183–190
- [2] Z. J. Shi and J. Guo, A new family of Conjugate Gradient methods, Journal of Computational and Applied Mathematics 224 (2009), pp. 444–457.
- [3] J. L. Nazaeth, Conjugate Gradient method, Wiley Interdisciplinary Reviews: Computational Statistics 1:3 (2009), pp. 348–353.
- [4] N. Andrei, 40 conjugate Gradient algorithms for unconstrained optimization. A survey on their definition. ICI Technical Report No.13/08, Mach 14, 2008.
- [5] N. Andei, A scaled BFGS preconditioned Conjugate Gradient algorithm for unconstrained optimization, Appl. Math. Lett. 20(2007), no. 6. 645–650.
- [6] W. W. Hager and H. Zhang, A new Conjugate Gradient method with guaranteed descent and an efficient line search, SIAM Journal on Optimization 16:1(2005), pp. 170–192.
- [7] Y.H. Dai and L.Z. Liao, New Conjugacy Conditions and Related Nonlinear Conjugate Gradient Methods, Appl Math Optim 43:87–101 (2001).
- [8] P. McCormick and K. Ritter, Alternative proofs of the convergence properties of the Conjugate Gradient method, JOTA 13:5(1975) PP. 497–518.
- [9] A. Cohen. Rate of convergence of several Conjugate Gradient method algorithms, SIAM Journal on Numerical Analysis 9(1972), pp. 248–259.
- [10] P. Wolfe, Convergence conditions for ascent methods, some corrections, SIAM Review, 13(1971), pp. 185–188.
- [11] G. Zoutendijk, Nonlinear Programming Computational methods, in integer and nonlinear programming, J. Abadie, ed., North-Holland, Amsterdam, 1970, pp. 37–86.
- [12] H.P. Crowder and P. Wolfe, Linear convergence of the Conjugate Gradient method, IBM J. Res. Dec. 16(1969), pp. 431–433.
- [13] P. Wolfe, Convergence conditions for ascent methods, SIAM Review, 11(1969), pp. 226–235.

Conjugate direction for general nonlinear function and its convergence

- [14] R. Fletcher and C.M. Reeves, Function minimization by Conjugate Gradients, *Comput. J.* 7s (1964), 149–154.
- [15] M.R. Hestenes and E. Stiefel, Methods of Conjugate Gradients for solving linear systems, *J. Research Nat. Bur. Standards* 49(1952), 409–436(1953).

