Australian Journal of Basic and Applied Sciences, 7(7): 947-951, 2013
ISSN 1991-8178

A globally Convergent Conjugate Gradient Algorithms For Spectral Method

Aseel Muayad Qasim

Department of Mathematics, Numerical optimization Mosul University, Iraq.

Abstract: we suppose in this paper a new scalar β_k^{NB} from the quadratic function which is depend on the Conjugacy condition , then we compute the numerical value of the factor t from the Conjugacy condition using inexact line search and combine it with β_k^{NB} in order to achieved the global convergence for this method.

Key words: Unconstrained optimization, Conjugacy condition, Conjugate gradient method, Global convergence.

INTRODUCTION

The conjugate gradient method is designed to solve the following unconstrained optimization problem:

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

Where $f: R^n \to R$ is a smooth, nonlinear function whose gradient will be denoted by $g_k = \nabla f(x_k)$. More explicitly, It is well known that the linear conjugate gradient methods generate a sequence of search directions d_k such that the following condition holds:

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

Where α_k is a step length which is computed by carrying out a line search, and the search direction at the first iteration is the steepest descent direction i.e $d_0 = -g_0$.The consequent search direction can be defined by:

$$d_{k+1} = -g_{k+1} + \beta_k d_k \tag{3}$$

Where β_k is a scalar , $f(x)$ is a strictly convex quadratic function, if α_k is the exact one-dimensional minimize along the direction d_k, $\alpha_k = \arg\min_{\alpha>0}\{f(x_k + \alpha d_k)\}$ then (2),(3) are called the linear conjugate gradient method. Otherwise, (2), (3) are called the nonlinear conjugate gradient method (Guoyin Li, Chunming Tang and ZengxinWei, 2007). Some well-known formulas for β_k are the Hestense–Stiefel(HS)(Hestense and Stiefel , 1952), Polak–Ribiere(PR)(Polak and Ribiere, 1969) and Fletcher–Reeves (FR)(Fletcher, 1964) methods which are given, respectively, by:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \tag{4}$$

$$\beta_k^{FR} = \frac{\|g_{k+1}^T\|^2}{\|g_k\|^2} \tag{5}$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \tag{6}$$

The global convergence properties of the FR, PR and HS methods have been studied by many researches, including (Zoutendijk, 1970). To establish the convergence results of these methods, it is usually required that the step length α_k should satisfy the strong Wolfe conditions:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k, \tag{7}$$

$$\left|g_{k+1}^T d_k\right| \leq -\sigma g_k^T d_k \tag{8}$$

Where $0 < \delta \leq \sigma < 1$ .some convergence analysis even require that the α_k be computed by the exact line search, that is : $f(x_k + \alpha_k d_k) = \min_{\alpha>0} f(x_k + \alpha_k d_k)$. On the other hand, many other numerical methods for unconstrained optimization are proved to be convergent under the Wolfe conditions (Guoyin, Chunming Tang and ZengxinWei, 2007) :

Corresponding Author: Aseel Muayad Qasim, Department of Mathematics, Numerical optimization Mosul University, Iraq.
E-mail: qasim.aseel@yahoo.com

$$f(x_k)-f(x_k+\alpha_k d_k)\geq -\delta\alpha_k g_k^T d_k, \quad (9)$$

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k \quad (10)$$

These line search strategies require the descent condition $g_k^T d_k<0$ for all k.

New Nonlinear Conjugate Gradient Methods:

We know if any algorithm use ELS then $y_k^T d_{k+1}=0$ and this is satisfies when we put t=0 in the Conjugacy condition

$$d_{k+1}^T y_k = -t g_{k+1}^T s_k \quad (11)$$

but if the direction is not exact then $y_k^T d_{k+1}=-t g_{k+1}^T s_k$, (Wu and Chen, 2010) formula used to find the minimum value for the quadratic convex function which is denoted by

$$\beta_k^{wch}=\frac{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k}{\|g_k\|^2} \quad (12)$$

Assume that our new parameter which is denoted by β_k^{NB} is a modification to the numerator of the Wu and Chen update parameter to obtain :

$$\beta_k^{NB}=\frac{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k+t\; g_{k+1}^T s_k}{\|g_k\|^2} = \beta_k^{wch}+t\frac{g_{k+1}^T s_k}{\|g_k\|^2} \quad (13)$$

where $s_k=\alpha_k d_k$ and $t\geq 0$ is a constant , for an exact line search g_{k+1} is orthogonal to s_k hence , the β_k^{NB} is reduced to Wu and Chen method .And further more we can compute t by multiplying (3) with y_k and using (11) , we obtain the following formula for computing t : $y_k^T d_{k+1} = -y_k^T g_{k+1} + \beta_k y_k^T d_k$

Now if the direction is in exact) ILS) then $y_k^T d_{k+1}=-t g_{k+1}^T s_k$ and so we have

$$-t g_{k+1}^T s_k = -y_k^T g_{k+1}+\frac{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k+t g_{k+1}^T s_k}{\|g_k\|^2} y_k^T d_k$$

$$-t g_{k+1}^T s_k - \frac{t g_{k+1}^T s_k}{\|g_k\|^2} y_k^T d_k = -y_k^T g_{k+1} + \frac{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k}{\|g_k\|^2} y_k^T d_k$$

$$-t(g_{k+1}^T s_k y_k^T d_k + \|g_k\|^2 g_{k+1}^T s_k) = -\|g_k\|^2 y_k^T g_{k+1} + \{2(f_k - f_{k+1}) + g_k^T s_k + g_{k+1}^T y_k\} y_k^T d_k$$

$$t = \frac{-\|g_k\|^2 y_k^T g_{k+1}+\{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k\} y_k^T d_k}{-(g_{k+1}^T s_k y_k^T d_k+\|g_k\|^2 g_{k+1}^T s_k)}$$

$$t = \frac{-\|g_k\|^2 y_k^T g_{k+1}+\{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k\} y_k^T d_k}{-g_{k+1}^T s_k (y_k^T d_k+\|g_k\|^2)} \quad (14)$$

now substitute the value of t in (14) in equation (13) we get:

$$\beta_k^{N1} = \frac{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k+g_{k+1}^T s_k+\frac{-\|g_k\|^2 y_k^T g_{k+1}+\{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k\} y_k^T d_k}{-g_{k+1}^T s_k (y_k^T d_k+\|g_k\|^2)}}{\|g_k\|^2}$$

$$\beta_k^{N1}=\frac{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k-\frac{-\|g_k\|^2 y_k^T g_{k+1}+\{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k\} y_k^T d_k}{(y_k^T d_k+\|g_k\|^2)}}{\|g_k\|^2}$$

$$\beta_k^{N1}=\frac{(2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k)(y_k^T d_k+\|g_k\|^2)+\|g_k\|^2 y_k^T g_{k+1}-\{2(f_k-f_{k+1})+g_k^T s_k+g_{k+1}^T y_k\} y_k^T d_k}{\|g_k\|^2 (y_k^T d_k+\|g_k\|^2)}$$

$$\beta_k^{N1} = \frac{(2(f_k - f_{k+1}) + g_k^T s_k + g_{k+1}^T y_k)\|g_k\|^2 + \|g_k\|^2 y_k^T g_{k+1}}{\|g_k\|^2 (y_k^T d_k + \|g_k\|^2)}$$

$$\beta_k^{N1} = \frac{2(f_k - f_{k+1}) + g_k^T s_k + 2g_{k+1}^T y_k}{(y_k^T d_k + \|g_k\|^2)} \quad (15)$$

Convergence Analysis:

In order to establish the global convergence analysis, we make the following assumptions for the objective function *f*.

Assumption (1):

i. The level set $\xi = \{x | f(x) \le f(x_1)\}$ is bounded, namely, there exists a constant *B* >0 such that $\|x\| \le B$ for all $x \in \xi$

ii. In some neighborhood N *of* ξ , *f* is continuously differentiable, and its gradient is globally Lipschitz continuous, namely, there exists a constant *L*>0 such that $\|g(x) - g(y)\| \le L\|x - y\|$ for all x ,y$\in$N (Eom and Joan , 2011)

Theorem (2):

Suppose that d_{k+1} is given by (3) and β_k^{N1} which is defined in (15) then , the following result is satisfies : $g_{k+1}^T d_{k+1} < -c\|g_{k+1}\|^2$

Proof:

By induction for k=1 we have $d_1 = -g_1$ then $d_1^T g_1 < 0$, then we assume that $g_k^T d_k < 0 \quad \forall \quad k \ge 2$.

$$d_{k+1} = -g_{k+1} + \beta_k d_k$$

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}^T\|^2 + \frac{2(f_k - f_{k+1}) + g_k^T s_k + 2y_k^T g_{k+1}}{(y_k^T d_k + \|g_k\|^2)} g_{k+1}^T d_k$$

It follows from (8) and (10) that $d_k^T y_k = d_k^T (g_{k+1} - g_k) \ge (\sigma - 1) g_k^T d_k$ (Dai and Yuan, 1999) and from $d_k^T g_k = -\|g_k\|^2$ (Anwa, Zhibin, Hao and Qian , 2011) that:

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}^T\|^2 + \frac{-2(f_{k+1} - f_k) + g_k^T s_k + 2y_k^T g_{k+1}}{(\sigma - 1) g_k^T d_k - g_k^T d_k} (-\sigma g_k^T d_k)$$

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}^T\|^2 + \frac{2(f_{k+1} - f_k) - g_k^T s_k - 2y_k^T g_{k+1}}{(\sigma - 2)} (\sigma)$$

Now from Powell restart (Powell ,1977) and let $\gamma \in (0,1)$ we have:

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}^T\|^2 + \{2(f_{k+1} - f_k) - g_k^T s_k - 2\gamma \|g_{k+1}^T\|^2\} \frac{\sigma}{\sigma - 2}$$

And since $\frac{\sigma}{\sigma - 2}$ is small and negative then we assume that $\frac{\sigma}{\sigma - 2} = -\kappa$ and noting from (7) and $g_k^T s_k = \alpha g_k^T d_k$ that:

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}^T\|^2 - \alpha\kappa(2\delta - 1) g_k^T d_k + 2\gamma\kappa \|g_{k+1}^T\|^2$$

Because of $\alpha\kappa(2\delta - 1) g_k^T d_k > 0$ and $2\gamma\kappa \|g_{k+1}^T\|^2 > 0$ then this inequality is true:

$g_{k+1}^T d_{k+1} \le -\|g_{k+1}^T\|^2$ and this mean that $g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2$

Global Convergence Theorem:

Under Assumption ii, we give a useful lemma, which was essentially proved by (Zoutendijk, 1970):

Lemma (1):

Suppose that x_1 is a starting point for which Assumption (1) is satisfied. Consider any method of the form (2), where d_k is a descent direction and α_k satisfies Wolfe conditions (7) and (8) then we have: $\sum_{k=1}^{\infty} \frac{1}{\|d_k\|^2} = \infty$

Theorem (2):

Suppose that x_1 is a starting point for which Assumption (1) holds. Let $\{x_k$, k = 1,2,....$\}$ be generated by our method. Then the algorithm either terminates at a stationary point or converges in the sense that $\liminf_{k\to\infty} \|g_k\| = 0$

Proof:

Suppose that the conclusion does not hold, that is to say their exist appositive constant ε such that $\|g_k\| \geq \varepsilon$ for all k. Since $d_{k+1} = -g_{k+1} + \beta_k d_k$ which is can be written as $d_{k+1} + g_{k+1} = \beta_k d_k$ and since:

$$\beta_k^{N1} = \frac{2(f_k - f_{k+1}) + g_k^T s_k + 2y_k^T g_{k+1}}{(y_k^T d_k + \|g_k\|^2)} \qquad \Rightarrow \beta_k^{N1} = \frac{-2(f_{k+1} - f_k) + g_k^T s_k + 2y_k^T g_{k+1}}{(y_k^T d_k + \|g_k\|^2)}$$

From (7) we have $f(x_k + \alpha_k d_k) - f(x_k) \leq \delta\alpha_k g_k^T d_k$, $\Rightarrow \beta_k \leq \frac{-2\delta\alpha g_k^T d_k + \alpha g_k^T d_k + 2g_{k+1}^T y_k}{(g_{k+1}^T - g^T{}_k) d_k + \|g_k\|^2)}$

$$\beta_k \leq \frac{-2\delta\alpha g_k^T d_k + \alpha g_k^T d_k + 2\|g_{k+1}\|^2 - 2g_{k+1}^T g_k}{(g_{k+1}^T d_k - g_k^T d_k - g_k^T d_k)} \Rightarrow \beta_k \leq \frac{-2\delta\alpha g_k^T d_k + \alpha g_k^T d_k + 2\|g_{k+1}\|^2 + 2g_{k+1}^T d_k}{(\sigma g_k^T d_k - 2g_k^T d_k)}$$

$$\beta_k \leq \frac{(-2\delta\alpha + \alpha) g_k^T d_k - 2\sigma g_k^T d_k}{(\sigma - 2) g_k^T d_k} \qquad \Rightarrow \beta_k \leq \frac{(-2\delta\alpha + \alpha - 2\sigma) g_k^T d_k}{(\sigma - 2) g_k^T d_k} = \frac{\gamma g_k^T d_k}{(\sigma - 2) g_k^T d_k} = b \quad \Rightarrow |\beta_k| < |b|$$

such that b is a constant $\|d_{k+1}\| = \|-g_{k+1} + \beta_k d_k\| \leq \|g_{k+1}\| + |b|\|d_k\|$

$\leq \gamma + |b|\eta$ and with this contradiction complete the prove that is $\sum_{i=1}^{k} \frac{1}{\|d_i\|^2} \geq \frac{1}{(\gamma + b\eta)^2} \sum 1 = \infty$

Numerical Experiments:

Now we present a numerical experiment whose objective function is compared with Wu and Chen algorithms on the same set of unconstrained optimization test problem. For each test function (Andre, 2008). All algorithms implemented with the same line search and with the same parameters. The comparison is based on number of iteration (NOI), and number of function evaluation (NOF). Our algorithms has converged as soon as $\|g_k\|_\infty \leq 10^{-5}$.

Table 1: Comparison of algorithms w.r.s to NOI and NOF for n=100,n=1000

Test problems	New(N=100)	Wu&chen	New (N=1000)	Wu&chen
	NOF(NOI)	NOF(NOI)	NOF(NOI)	NOF(NOI)
Miele	116(386)	133(404)	140(485)	158(494)
Wolfe	44(89)	44(89)	66(133)	64(129)
Strait	7(16)	7(16)	7(16)	7(16)
Rosen	26(74)	31(86)	27(76)	31(86)
Nondiagonal	26(104)	27(73)	24(67)	27(73)
Cubic	16(46)	15(44)	10(46)	15(44)
Beal	12(30)	14(34)	14(34)	14(34)
Wood	29(67)	30(68)	29(67)	30(68)
Total	276(812)	301(814)	317(924)	346(944)

Table 2: Comparison of algorithms w.r.s to NOI and NOF for n=5000, n=10000

Test problems	New(N=5000)	Wu&chen	New(N=10000)	Wu&chen
	NOF(NOI)	NOF(NOI)	NOF(NOI)	NOF(NOI)
Miele	173(608)	183(601)	168(607)	275(879)
Wolfe	191(392)	197(401)	277(562)	305(618)
Strait	7(16)	7(16)	7(16)	7(16)
Rosen	27(76)	31(86)	27(76)	31(86)
Shallo	10(26)	10(26)	10(26)	9(24)

Nondiagonal	24(67)	27(73)	24(67)	27(73)
Cubic	16(4)	15(44)	16(46)	15(44)
Beal	12(30)	14(34)	12(30)	14(34)
Wood	29(67)	30(68)	29(67)	30(68)
Total	489(1277)	514(1349)	570(1497)	713(1842)

Conclusion:

From tables (1) and (2) which is denoted above we note clearly that the comparison result for the new β_k which is denoted by β_k^{N1} with Wu and Chen method for n=100, 1000, 5000 and 10000 is more effective and effietient than the Wu and Chen method as we shown .

REFERENCES

Andrei, N., 2008. An unconstrained optimization test Function Collection . Advanced Modeling and Optimization . Romania, 10(1): 147-161.

Anwa Zhou, Zhibin Zhu, Hao Fan and QianQing, 2011. Three New Hybrid Conjugate Methods for optimization . Applied Mathematics. china, 2: 303-308.

Dai, Y and Y. Yuan, 1999. A Nonlinear conjugate gradient method with a strong global convergence property . SIAM J. Optim, 10(1): 177-182.

Eom, K and R. Joan-Arinyo, 2011. A Descent Dai-Liao Conjugate GradientMethod Based on a Modified Secant Equation and Its Global Convergence . Ioannis E. Livieris and Panagiotis Pintelas . Patras Greece.

Fletcher, R. and C.M. Reeves, 1964. Function minimization by conjugate gradients. The Computer Journal, 7: 149-154.

Guoyin Li, Chunming Tang and ZengxinWei., 2007. New conjugacy condition and related new conjugate gradient methods for unconstrained optimization. Journal of Computational and Applied Mathematics, 202: 523 -539.

Hestenes, M.R. and E. Stiefel, 1952. Methods of conjugate gradients for solving linear systems. Journal of Research of the National Bureau of Standards, 49: 409-436.

powell, M.J.D., 1977. Restart procedure for the conjugate gradient method. mathematical programming, 2: 241-254.

Polak, E. and G., Ribiere, 1969. Note sur la convergence dem´ethodes de directions conjugu´ees. Revue Francais d'Informatique et de Recherche Operationnelle, 3(16): 35-43.

Wu, C.Y. and G.Q., Chen, 2010. New type of conjugate gradient algorithms for unconstrained optimization problems . journal of system engineering and electronics, 21(6): 1000-1007.

Zoutendijk, G., 1970. Nonlinear programming computational methods, in: Jabadie (Ed.). Integer and Nonlinear Programming. North-Holland. Amsterdam, pp: 37-86.