Iraqi Journal of Statistical Sciences (25) 2013
The 6th Scientific Conference of the College of Computer Sciences & Mathematics
pp [80-88]

Modified PRP Method in Unconstrained Optimization Method

Aseel Muayad Qasim*

Abstract

we suppose in this paper a new scalar β_k^* which is a modification to PRP method derived from the quadratic function , and we compute the numerical value of the conjugacy factor tto achieve to a new parameter β_k^{**} with computed value t from the conjugacy condition using inexact line search and combine it with β_k^{**} in order to achieved the global convergence for this method.

الملخص

في هذا البحثتم اشتقاق معلمة جديدة β_k^*والتي هي تطوير لطريقة PRP و مشتقة من الدالة التربيعية ومن ثم وقمنا بحساب هذه القيمة العددية لعامل الترافق tمن اجل الحصول على β_k^{**}جديدة وبمعلمة t محسوبة من شرط الترافق وليست افتراضية باستخدام خط بحث غير مضبوط ومن ثم دمج هذه القيمة مع قيمة β_k^{**} المقترحة من اجل الوصول الى التقارب الشامل لهذه الطريقة.

1.Introduction

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

$$\min\{ f(x) : x \in R^n \} \qquad \text{.......................(1)}$$

Where $f : R^n \rightarrow R$ 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:

$$x_{k+1} = x_k + \alpha_k d_k \qquad \text{..............................(2)}$$

* Assist Lecturer/ Dept. Mathematics / Education College / Mosul University

Received Date 6/10/2013 —— Accept Date 21/11/2013

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: $d_{k+1} = -g_{k+1} + \beta_k d_k$ ……………………(3)

Where β_k is a scalar , $f(x_k)$ 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), Fletcher–Reeves (FR)(Fletcher , 1964) Polak–Ribiere(PR)(Polak and Ribiere,1969) methods which are given, respectively by:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \quad \text{……………………………….(4)}$$

$$\beta_k^{FR} = \frac{\|g_{k+1}^T\|^2}{\|g_k\|^2} \quad \text{……………………………………(5)}$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \quad \text{……………………………………(6)}$$

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 some line searchs , one of them is strong Wolfe conditions:

$$f(x_k) - f(x_k + \alpha_k d_k) \ge -\delta\alpha_k g_k^T d_k, \text{………………….(7)}$$

$$\left|g_{k+1}^T d_k\right| \le -\sigma g_k^T d_k \text{……………………………………………(8)}$$

Where $0 < \delta \le \sigma < 1$. 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) :

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

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

2. New nonlinear conjugate gradient methods:

The new nonlinear conjugate gradient methods is depend on the idea of using the conjugacy condition :

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

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 equation (11) , but if the direction is not exact then $y_k^T d_{k+1}=-t g_{k+1}^T s_k$; Assume that our new parameter which is denoted by β_k^* is a modification to the numerator of the PRP update parameter with the Conjugacy condition to obtain this new form :

$$\beta_k^*=\frac{g_{k+1}^T y_k + t\, g_{k+1}^T s_k}{\|g_k\|^2} = \beta_k^{PRP}+t\frac{g_{k+1}^T s_k}{\|g_k\|^2} \qquad (12)$$

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 β_k^* is reduced to PRP method. But if the line search is inexact then we can compute t by multiplying equation (3) with y_k and using (11), we obtain the following formula for computing t

$$y_k^T d_{k+1} = -y_k^T g_{k+1} + \beta_k y_k^T d_k \qquad (13)$$

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

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

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

$$-t g_{k+1}^T s_k \|g_k\|^2 - t\, g_{k+1}^T s_k y_k^T d_k = -y_k^T g_{k+1}\|g_k\|^2 + g_{k+1}^T y_k y_k^T d_k$$

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

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

now substitute the value of t in (14) in equation (12) we get :

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

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

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

$$\beta_k^{**}=\frac{2g_{k+1}^T y_k}{\|g_k\|^2+y_k^T d_k} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(15)$$

and we use the last β_k^{**} in equation (15) to prove the convergence analysis of our algorithms.

3. 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 $\xi=\{x \mid f(x)\le f(x_1)\}$ 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 (Gilbert J.C. and Nocedal J. ,1992)

Theorem (2)

Suppose that d_{k+1} is given by (3) and β_k^{**} 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 $g_k^T d_k < 0 \quad \forall \quad k \geq 2$.

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

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

It follows from strong wolfe condition (7) and (8) that:

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

$$g_{k+1}^T d_{k+1} + \| g_{k+1} \|^2 \leq \frac{2 y_k^T g_{k+1}}{y_k^T d_k + \|g_k\|^2} (-\sigma g_k^T d_k)$$ dividing both side by $\|g_{k+1}\|^2$ and invert the inequality:

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

Now also it follows from strong wolfe condition (7) and (8) that $g_k^T d_k \leq \frac{-y_k^T d_k}{(\sigma + 1)}$

$$\Rightarrow -g_k^T d_k \geq \frac{y_k^T d_k}{(\sigma + 1)}$$

$$\frac{\| g_{k+1} \|^2}{g_{k+1}^T d_{k+1} + \| g_{k+1} \|^2} \geq \frac{(y_k^T d_k + \|g_k\|^2)\| g_{k+1} \|^2 (\sigma + 1)}{2\|y_k\| \|g_{k+1}\| (\sigma y_k^T d_k)}$$

$$\frac{\| g_{k+1} \|^2}{g_{k+1}^T d_{k+1} + \| g_{k+1} \|^2} \geq \frac{(y_k^T d_k + \|g_k\|^2)\| g_{k+1} \|(\sigma + 1)}{2\|y_k\|(\sigma y_k^T d_k)}$$

and let $\frac{(y_k^T d_k + \|g_k\|^2)\| g_{k+1} \|(\sigma + 1)}{2\|y_k\|(\sigma y_k^T d_k)} = \delta > 0$

$$\frac{\| g_{k+1} \|^2}{g_{k+1}^T d_{k+1} + \| g_{k+1} \|^2} \geq \delta \Rightarrow \frac{g_{k+1}^T d_{k+1} + \| g_{k+1} \|^2}{\| g_{k+1} \|^2} \leq \frac{1}{\delta} \Rightarrow$$

$$g_{k+1}^{T} d_{k+1} \leq \tfrac{1}{\delta}\| g_{k+1} \|^{2} - \| g_{k+1} \|^{2} \Rightarrow g_{k+1}^{T} d_{k+1} \leq -(1-\tfrac{1}{\delta})\| g_{k+1} \|^{2}$$

and if we assume $1-\dfrac{1}{\delta}=c$ and $\dfrac{1}{\delta}\in(0,1)$ then we complete the proof

$$g_{k+1}^{T} d_{k+1} \leq -c\| g_{k+1} \|^{2}$$

4. Global Convergence Theorem:

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

Lemma (2) : 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 : $\sum_{k=1}^{\infty} \dfrac{1}{\|d_k\|^2} = \infty$

Theorem (3) : 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

$$\lim_{k\to\infty} \inf \|g_k\| = 0$$

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 $\|d_{k+1}\| \leq \|g_{k+1}\| + |\beta_k| \|d_k\|$ and since: $|\beta_k^{**}| = \left| \dfrac{2 g_{k+1}^T y_k}{\|g_k\|^2 + y_k^T d_k} \right|$

$$\Rightarrow |\beta_k^{**}| \leq \left| \frac{2 g_{k+1}^T y_k}{\|g_k\|^2 + (\sigma-1) g_k^T d_k} \right|$$

$$|\beta_k^{**}| \leq \left| \frac{2 g_{k+1}^T y_k}{\|g_k\|^2 - (\sigma-1)\|g_k\|^2} \right| \Rightarrow |\beta_k^{**}| \leq \left| \frac{2\|g_{k+1}\|\|y_k\|}{(2-\sigma)\|g_k\|^2} \right|$$

$$|\beta_k^{**}| \leq \left| \frac{2\gamma\mu}{(2-\sigma)\zeta^2} \right| = b$$

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

and with this contradiction we complete the prove that is

$$\sum_{i=1}^{k} \frac{1}{\|d_i\|^2} \geq \frac{1}{(\gamma+b\eta)^2} \sum_{k=1}^{\infty} 1 = \infty$$

5. Numerical experiments :

Now we present a numerical experiments whose objective function is compared with PRP 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 $\|g_k\|_\infty \leq 10^{-5}$.

Tables (1) ,(2) and (3) show the Comparison of algorithms w.r.s to NOI and NOF for n=10, n=100, n=500,n=1000,,n=10000 respectively.

Table(1)

Test problems	PRP N=10	New N=10	PRP N=100	New N=100
	NOF(NOI)	NOF(NOI)	NOF(NOI)	NOF(NOI)
Shallow	8(21)	8(21)	8(21)	8(21)
Wolfe	36(73)	36(73)	44(89)	44(89)
Strait	6(14)	6(14)	6(14)	6(14)
Edger	5(14)	5(14)	5(14)	5(14)
Nondiagonal	28(73)	27(71)	27(73)	27(71)
Cubic	14(40)	13(39)	15(44)	13(39)
Rosen	27(77)	30(81)	27(77)	30(81)
Beal	12(30)	11(28)	12(30)	12(30)
Powell	43(105)	31(79)	50(136)	34(95)
Fred	6(19)	6(20)	9(25)	6(20)
Sum	7(41)	7(41)	13(61)	13(61)
Recp	8(25)	8(25)	8(25)	8(25)
Total	200(532)	188(506)	224(609)	206(560)

Table(2)

Test problems	PRPN=500	NewN=500	PRPN=1000	NewN=1000
	NOF(NOI)	NOF(NOI)	NOF(NOI)	NOF(NOI)
Shallow	8(21)	8(21)	9(24)	9(24)

Wolfe	47(95)	55(111)	64(129)	62(125)
Strait	6(14)	6(14)	6(14)	6(14)
Edger	6(16)	6(16)	6(16)	6(16)
Nondiagonal	27(73)	26(67)	27(73)	27(71)
Cubic	15(44)	13(39)	15(44)	13(39)
Rosen	27(77)	30(81)	27(77)	30(81)
Beal	12(30)	12(30)	12(30)	12(30)
Powell	50(136)	39(110)	54(164)	39(110)
Fred	10(27)	6(20)	10(27)	6(20)
Sum	19(102)	18(85)	21(106)	22(112)
Recp	8(25)	8(25)	8(25)	8(25)
Total	235(660)	227(619)	259(729)	240(667)

Table(3)

Test problems	PRP N=10000	NewN=10000
	NOF(NOI)	NOF(NOI)
Shallow	9(24)	9(24)
Wolfe	271(551)	267(542)
Sum	35(127)	35(166)
Edger	6(16)	6(16)
Nondiagonal	27(73)	28(73)
Cubic	15(44)	13(39)
Beal	12(30)	12(30)
Wood	30(69)	30(69)
Powell	56(168)	39(110)
Fred	10(28)	9(26)
Recp	8(25)	8(25)
Osp	699(2683)	631(2485)
Total	1178(3838)	1087(3605)

6. Conclusion:

From tables (1),(2) and (3) which is denoted above we note clearly that the comparison result for the new β_k which is denoted by β_k^* with PRP method for n=10 ,100 , 500, 1000 and 10000 the result is more effective and effietient than the PRPmethod .

7.References:

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

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

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

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

[5] Glibert, J.C. and Nocedal,J. , 1992, Global Convergence Properties of Conjugate gradient methods for optimization , SIAM J. optimization, vol . 2 , no.1 ,pp. 21-42.

[6] Guoyin Li, Chunming Tang and ZengxinWei. , 2007. New conjugacy condition and related new conjugate gradientmethods for unconstrained optimization . Journal of Computational and Applied Mathematics , 202 : 523 – 539.

[7] Hestenes M. R. and E , Stiefel . , 1952 . Methods of conjugate gradients for solving linear systems. Journalof Research of the National Bureau of Standards, 49: 409–436.

[8] 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.

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

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