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 $B_{k}^{NB}$ 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 $B_{k}^{NB}$ 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:

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

Where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 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:

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

Where $\alpha_k$ 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:

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

Where $\beta_k$ is a scalar, $f(x)$ is a strictly convex quadratic function, if $\alpha_k$ 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 $\beta_k$ 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:

$$\beta_k^{HS} = \frac{\gamma_{k+1} T y_k}{d_{k}^{T} y_k}$$

$$\beta_k^{FR} = \frac{\gamma_{k+1} T y_k}{\|x_{k+1}\|}$$

$$\beta_k^{PR} = \frac{\gamma_{k+1} T y_k}{\|x_{k}\|}$$

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 $\alpha_k$ should satisfy the strong Wolfe conditions:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\sigma_k g_k^T d_k,$$

$$\|g_k + \beta_k g_{k-1}\| \leq -\sigma_k g_k^T d_k$$

Where $0 < \sigma_0 \leq 1$. Some convergence analysis even require that the $\alpha_k$ be computed by the exact line search, that is $\alpha_k = \arg \min_{\alpha > 0} f(x_k + \alpha d_k)$ . 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):
New Nonlinear Conjugate Gradient Methods:

We know if any algorithm uses ELS then \( \gamma_k T d_{k+1} = 0 \) and this satisfies when we put \( t=0 \) in the Conjugacy condition

\[
d^T_{k+1} y_k = -s^T_{k+1} y_k
\]

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

\[
\rho_{N_{k+1}} = \frac{2(f_k - f_{k+1}) + g^T_k s_k + g^T_k y_k + g^T_{k+1} y_k}{\| s_k \|^2} - \frac{2(f_k - f_{k+1}) + g^T_k s_k + g^T_k y_k + g^T_{k+1} y_k}{\| s_{k+1} \|^2} y^T_{k+1} d_k
\]

Assume that our new parameter which is denoted by \( \rho_{k+1}^{NB} \) is a modification to the numerator of the Wu and Chen update parameter to obtain:

\[
\rho_{k+1}^{NB} = \frac{2(f_k - f_{k+1}) + g^T_k s_k + g^T_k y_k + g^T_{k+1} y_k}{\| s_k \|^2} - \frac{2(f_k - f_{k+1}) + g^T_k s_k + g^T_k y_k + g^T_{k+1} y_k}{\| s_{k+1} \|^2} y^T_{k+1} d_k
\]

where \( s_k = \alpha_k d_k \) and \( \tau > 0 \) is a constant, for an exact line search \( s_{k+1} \) is orthogonal to \( s_k \) hence, the \( \rho_{k+1}^{NB} \) 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 \( t \):

\[
y^T_{k+1} d_k = -y^T_k s_{k+1} + \rho_k^{NB} y^T_k d_k
\]

Now if the direction is in exact (ILS) then \( y^T_{k+1} d_{k+1} = -s^T_{k+1} y_k \) and so we have

\[
-t = -\| s_k \|^2 g^T_k s_{k+1} + 2(f_k - f_{k+1}) + g^T_k s_k + g^T_k y_k + g^T_{k+1} y_k
\]

\[
t = \frac{\| s_k \|^2 g^T_k s_{k+1} + 2(f_k - f_{k+1}) + g^T_k s_k + g^T_k y_k + g^T_{k+1} y_k}{\| s_k \|^2 - \| s_{k+1} \|^2} y^T_{k+1} d_k
\]

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

\[
\rho_{k+1}^{NI} = \frac{2(f_k - f_{k+1}) + g^T_k s_k + g^T_k y_k + g^T_{k+1} y_k}{\| s_{k+1} \|^2(t^T_{k+1} d_k + \| s_k \|^2)} - \frac{2(f_k - f_{k+1}) + g^T_k s_k + g^T_k y_k + g^T_{k+1} y_k}{\| s_{k+1} \|^2(t^T_{k+1} d_k + \| s_k \|^2)} y^T_{k+1} d_k
\]
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 $\{x : f(x) \leq f(x_0)\}$ is bounded, namely, there exists a constant $B > 0$ such that $\|x\| \leq B$ for all $x \in \xi$.

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

**Theorem (2):**

Suppose that $d_{k+1}$ is given by (3) and $\beta_{k+1}^N$, which is defined in (15), then, the following result is satisfies:

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

**Proof:**

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

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

It follows from (8) and (10) that $d_k^T g_k = d_k^T (g_k - g_k) \leq 2 \|g_k\|\|g_k\|$ (Dai and Yuan, 1999) and from $d_k^T g_k = -\|g_k\|^2$ (Anwa, Zhibin, Hao and Qian, 2011) that:

$$g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 + 2(k_{k+1} - k_k) g_{k+1}^T d_k$$

And since $\sigma_2 > 0$ is small and negative then we assume that $\sigma_2 = -\kappa$ and noting from (7) that:

$$g_k^T s_k = \sigma_k d_k$$

and then:

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

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

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

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

**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 \( \alpha_k \) 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 \( \{x_k, k = 1, 2, \ldots\} \) 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 \( \varepsilon \) such that
\[
\|g_k\| \geq \varepsilon
\]
for all \( k \). Since \( d_{k+1} = -\beta_k d_k \) which is can be written as \( d_{k+1} = -\beta_k d_k \) and since:
\[
\beta_k = \frac{-2f(x_{k+1}) - \gamma_k x_{k+1} + \gamma_k x_k + 2\gamma_k s_{k+1}}{(\gamma_k s_k + \gamma_k s_{k+1})}
\]
From (7) we have:
\[
f(x_k + \alpha_k d_k) - f(x_k) \leq \gamma_k s_k d_k \quad \Rightarrow \quad \beta_k = \frac{-2\gamma_k d_k + \gamma_k d_k}{(\gamma_k + \gamma_k d_k)}
\]
\[
\beta_k \leq \frac{(2\gamma_k - \varepsilon)^2}{(\gamma_k + \varepsilon)^2} \leq \frac{\gamma_k}{\gamma_k + \varepsilon} \Rightarrow \|g_k\| < \varepsilon
\]
and with this contradiction complete the prove that is
\[
\sum_{k=1}^{\infty} \frac{1}{\|g_k\|^2} = \frac{1}{(\gamma + \varepsilon)^2} \sum_{i=1}^{\infty} i = \infty
\]

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
\[
\|g_k\| \leq 10^{-5}
\]

<table>
<thead>
<tr>
<th>Test problems</th>
<th>New (N=100)</th>
<th>Wu &amp; Chen</th>
<th>New (N=1000)</th>
<th>Wu &amp; Chen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miele</td>
<td>116(386)</td>
<td>133(404)</td>
<td>140(485)</td>
<td>158(494)</td>
</tr>
<tr>
<td>Wolfe</td>
<td>44(99)</td>
<td>44(99)</td>
<td>66(133)</td>
<td>64(129)</td>
</tr>
<tr>
<td>Strait</td>
<td>7(16)</td>
<td>7(16)</td>
<td>7(16)</td>
<td>7(16)</td>
</tr>
<tr>
<td>Rosen</td>
<td>26(74)</td>
<td>31(86)</td>
<td>27(76)</td>
<td>31(86)</td>
</tr>
<tr>
<td>Nondiagonal</td>
<td>26(104)</td>
<td>27(73)</td>
<td>24(67)</td>
<td>27(73)</td>
</tr>
<tr>
<td>Cubic</td>
<td>16(46)</td>
<td>15(44)</td>
<td>10(46)</td>
<td>15(44)</td>
</tr>
<tr>
<td>Beal</td>
<td>12(30)</td>
<td>14(34)</td>
<td>14(34)</td>
<td>14(34)</td>
</tr>
<tr>
<td>Wood</td>
<td>29(67)</td>
<td>30(68)</td>
<td>29(67)</td>
<td>30(68)</td>
</tr>
<tr>
<td>Total</td>
<td>276(812)</td>
<td>301(814)</td>
<td>317(924)</td>
<td>346(944)</td>
</tr>
</tbody>
</table>

Table 2: Comparison of algorithms w.r.t NOI and NOF for n=5000, n=10000
<table>
<thead>
<tr>
<th>Test problems</th>
<th>New(N=5000)</th>
<th>Wu &amp; Chen</th>
<th>New(N=10000)</th>
<th>Wu &amp; Chen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miele</td>
<td>173(608)</td>
<td>183(601)</td>
<td>168(607)</td>
<td>275(879)</td>
</tr>
<tr>
<td>Wolfe</td>
<td>191(392)</td>
<td>197(401)</td>
<td>277(362)</td>
<td>305(618)</td>
</tr>
<tr>
<td>Strait</td>
<td>7(16)</td>
<td>7(16)</td>
<td>7(16)</td>
<td>7(16)</td>
</tr>
<tr>
<td>Rosen</td>
<td>27(76)</td>
<td>31(86)</td>
<td>27(76)</td>
<td>31(86)</td>
</tr>
<tr>
<td>Shallo</td>
<td>10(26)</td>
<td>10(26)</td>
<td>10(26)</td>
<td>9(24)</td>
</tr>
</tbody>
</table>
Conclusion:

From tables (1) and (2) which is denoted above we note clearly that the comparison result for the new $\beta_k$ which is denoted by $\beta_N^k$ with Wu and Chen method for n=100, 1000, 5000 and 10000 is more effective and efficient than the Wu and Chen method as we shown.

REFERENCES


