# On New Conjugate Pair Method

**Abbas Y. Al-Bayati**       **Nidhal Al-Assady**       **Ban Ahmed Mitras**
profabbasalbayati@yahoo.com                    dr.banah.mitras@gmail.com
*College of Computer Sciences and Mathematics*
*University of Mosul, Iraq*

## ABSTRACT

This paper involves the combination between the conjugate pair and hybrid conjugate gradient methods. The new combined algorithm is based on exact line search and it is examined by using different nonlinear test functions in various dimensions. Experimental results indicate that the updated algorithm is more effective than of the two original algorithms.

**Keyword:** conjugate gradient method, Conjugate pair method.

حول طريقة التدرج المزدوجة الجديدة

**عباس يونس البياتي**       **نضال الأسدي**       **بان احمد متراس**
*كلية علوم الحاسبات والرياضيات، جامعة الموصل*

**تاريخ استلام البحث: 2002/08/05**       **تاريخ قبول البحث: 2003/03/15**

الملخص

تم في هذا البحث ربط خوارزميتين الأولى هي خوارزمية المتجهات المترافقة المزدوجة مع خوارزمية هجينة للمتجهات المترافقة. الخوارزمية الجديدة تستخدم خط البحث التام لحل دوال غير خطية وذات ابعاد مختلفة لفحص كفاءة الخوارزمية المقترحة. واثبتت النتائج العملية أن الخوارزمية المقترحة اكثر كفاءة من كلا الخوارزميتين الاصليتين.

**الكلمات المفتاحية:** خوارزمية التدرج المترافق، خوارزمية التدرج المزدوجة.

## 1. Introduction:

A very important concept in the theory of the unconstrained optimization of a function $f(x)$, $x \in R^n$ is that of conjugate directions.

Many algorithms for solving this problem consist basically of the following iteration. Starting with an initial approximation $x_0$ of the minimum point $x^*$ of $f(x)$,

$$x_{i+1} = x_i + \lambda_i d_i \qquad , i = 0, 1, .... \qquad ...(1)$$

the parameter $\lambda_i$ is chosen to minimize $f(x_i + \lambda_i d_i)$ as a function of $\lambda$, and the vector $d_i$ can be interpreted as a direction in which we move from $x_i$ to $x_{i+1}$, the optimal point in this direction. It follows immediately that the gradient vector at the point $x_{i+1}$ is orthogonal to $d_i$.

## Definition 1:

A set of $n$ nonzero vectors $u_0, u_1, ..., u_{n-1}$ are said to be conjugate to each other with respect to a given positive definite symmetric matrix $A$ if

$$u_i^T A u_j = 0 \quad \text{when} \quad i \neq j \qquad \ldots(2)$$

As an obvious consequence of the positive definiteness of *A*,

$$u_i^T A u_j > 0 \quad \text{when} \quad i = j \qquad \ldots(3)$$

There exists at least one set of *n* vectors satisfying this definition, via, the eigenvectors of *A*, then conjugate vectors are also linearly independent.

A general method to produce a set of *A*-conjugate vectors is the Gram-Shemidt orthogonalization procedure, which we will now describe briefly. Let ($v^T$) be a set of *n* linearly independent vectors. Let $u_1 = v_1$.
Choose $u_2$ as a linear combination of $v_2$ and $v_1$, that is $u_2 = v_2 + \lambda_{21} u_1$;
$u_2$ will be *A*-conjugate to $u_1$ if

$$0 = u_1^T A u_2 = u_1^T A v_2 + \lambda_{21} u_1^T A u_1,$$

or $\lambda_{21} = -(u_1^T A v^2 / u_1^T A u_1)$.

In general, having selected $u_1, u_2, \ldots, u_k$, we can choose $u_{k+1}$ as

$$u_{k+1} = v_{k+1} + \lambda_{k+1,1} u_1 + \lambda_{k+1,2} u_2 + \cdots + \lambda_{k+1,k} u_k$$

It will be conjugate to $u_r$, $r \leq k$, if

$$0 = u_r^T A u_{k+1} = u_r^T A v_{k+1} + \lambda_{k+1;r} u_r^T A u_r$$

or $\lambda_{k1,r} = -(u_r^T A v_{k+1} / u_r^T A u_r)$.

Thus, the conjugate set *{$u_i$}* can be developed from the set *{$v_i$}* by application of the recursion formula

$$u_{k+1} = v_{k+1} - \left\{ \left( u_1^T A v_{k+1} / u_1^T A u_1 \right) u_1 + \ldots + \left( u_k^T A v_{k+1} / u_k^T A u_k \right) u_k \right\} \qquad \ldots(4)$$

(VanWyk, 1977).

## 2. Conjugate Pair Method (CP):

Stewart (1977) introduced a generalization of the notation of conjugacy, leading to a variety of finitely terminating iterations for solving systems of linear equations. An adaptation of Stewart's ideas to minimization problems confirms not only the above-mentioned suspicion, but establishes a method with an even wider scope of generality.

We note that the definition of conjugancy can also be phrased as follows: If the vectors $u_0, u_1, \ldots, u_{n-1}$ are the columns of an *nxn* matrix *V*, then $u_0, u_1, \ldots, u_{n-1}$ are *A*-conjugate if $U^T A U$ is a diagonal (and of course nonsingular ). The generalization is achieved by introducing a second set of vectors $v_0, \ldots, v_{n-1}$.

## Definition 2:

let *A,U*, and *V* be non singular *nxn* matrices. Then *(U,V)* is an *A*-conjugate pair if $L = U^T A V$ is lower triangular.

The generalized algorithm for solving the equations

Gx+b=0

$g_{k-1}=Gx_{k-1}+b$

$$\mu_{k-1} = -v_{k-1}^T g_{k-1} / v_{k-1}^T G u_{k-1} \qquad \text{...(5)}$$

$x_k=x_{k-1}+\mu_{k-1}u_{k-1}$

Where $k=1,2,...,n$, and where

$U=[u_0,...,u_{n-1}]$ and $V=[v_0,...,v_n]$

## 3. Formulation of A Generalized-Conjugate Pair (GCP) Method:

Stewart (1977) developed an algorithm for constructing an *A*-conjugate pair *(U,V)* as follows. Given nonsingular matrices *V, A* and *P*, the vector $u_k$ is determined as a linear combination of $p_0,p_1,...,p_k$, $k=0,1,...,n-1$, such that *U* and *V* are *A*-conjugate. The resulting algorithm is as follows:

$u_0=d_0p_0,$

$u_1=d_1[\ p_1 - \left( v_0^T A p_1\ /\ v_0^T A u_0 \right)\ u_0\ ]\ ,$

$\vdots$

$\vdots$

$$u_k=d_k[\ p_k - \left( v_0^T A p_k\ /\ v_0^T A u_0 \right) u_0 - \left( v_1^T A p_k\ /\ v_1^T A u_1 \right) u_1 - ... - \left( v_{k-1}^T A p_k\ /\ v_{k-1}^T A u_{k-1} \right) u_{k-1}\ ] \quad \text{...(6)}$$

The constant $d_k$ are chosen to give $u_k$ some predetermined scaling.

We will now formulate the analogous generalized conjugate-direction method for the minimization of a function *f(x)*.

Suppose that *U* and *V* form a conjugate pair set.

$x_0$=arbitrary, $\qquad g_0=g(x_0)$

For $i=0,1,...,$compute

$$x_{i+1}=x_i+\lambda_i v_i\ , \qquad\qquad \text{...(7a)}$$

where $\lambda_i$ minimizes $f(x_i+\lambda_i v_i)$ as a function of $\lambda_i$ ,

$$g_i=g(x_i), \qquad g_{i+1}=g(x_{i+1}) \qquad\qquad \text{...(7b)}$$

$$\beta_i= -\lambda_i \left[ v_i^T g_i\ /\ u_i^T \left( g_{i+1} - g_i \right) \right] \qquad\qquad \text{...(7c)}$$

$$x_{i+1}= x_i+\beta_i u_i \qquad\qquad \text{...(7d)}$$

If *f* is quadratic, the $\beta_i$ in eq.(7) are the same as the $\mu_i$ in eq.(5). In fact if *f* takes the form:

$$\beta_i= -\lambda_i \left[ v_i^T g_i\ /\ u_i^T \left( g_{i+1} - g_i \right) \right] = \lambda_i \left[ v_i^T \left( g_{i+1} - g_i \right) / u_i^T \left( g_{i+1} - g_1 \right) \right]$$

$$= \lambda_i \left[ v_i^T \left( G x_{i+1} + b - G x_i - b \right) / u_i^T \left( G x_{i+1} + b - G x_i - b \right) \right]$$

$$= \lambda_i \left[ v_i^T G \left\{ (1/\lambda_i)(x_{i+1} - x_i) \right\} / u_i^T G \left\{ (1/\lambda_i)(x_{i+1} - x_i) \right\} \right]$$

$$= \lambda_i \left[ v_i^T G v_i\ /\ u_i^T G v_i \right] = -\left( v_i^T g_i\ /\ v_i^T G u_i \right) \left( v_i^T G v_i\ /\ u_i^T G v_i \right)$$

$$= \mu_i$$

**Theorem (1):**

If the iteration (7) is applied to the quadratic function where *(U,V)* form a generalized conjugate pair, the minimum is found in at most *n* iterations; moreover. $x_n$ lies in the subspase generated by $x_0$ and $v_0, v_1, ..., v_{n-1}$. For the proof see (VanWyk 1977)

## 4. Special case of CG-pair method:

By varying the choice of vectors $v_i$ and $p_i$ in the conjugation algorithm (6), one may obtain from (7) various finitely terminating iterations of minimization. When (6) is applied to a general function, the Hessian G can be eliminated from the conjugation algorithm by substituting $(1/\lambda_i)(g_{i+1}-g_i)^T$ for $v_1^T G$. Putting the scaling constants equal to 1, the choice *V=U* reduces (6) to the Gram Schmidt procedure and (7) becomes an ordinary conjugate direction algorithm. We will conclude by showing that, in this case, variation of the vectors $p_i$ leads to some well-known algorithms.

### 4.1 Fletcher-Reeves Algorithms:

We follow the way of induction to show that, if the columns of *P* are chosen successively,

$P=[-g_0, -g_1, ..., -g_{k-1}]$,

The directions in (6) reduce to

$u_k = -g_k + \left(g_k^T g_k / g_{k-1}^T g_{k-1}\right) u_{k-1}$.

We have

$u_0 = p_0 = -g_0$

$u_1 = -g_1 + \left(u_0^T G g_1 / u_0^T G u_0\right) u_0$

$\quad = -g_1 + [(1/\beta_0)(x_1-x_0)^T G g_1 / (1/\beta_0)(x_1-x_0)^T G u_0] u_0$

$\quad = -g_1 + [(g_1-g_0)^T g_1 / (g_1-g_0)^T u_0] u_0 = -g_1 + \left(g_1^T g_1 / g_0^T g_0\right) u_0$.

We note that

$g_2^T g_0 = 0 = g_2^T g_1$ .

Now, suppose that

$u_{k-1} = -g_{k-1} + \left(g_{k-1}^T g_{k-1} / g_{k-2}^T g_{k-2}\right) u_{k-2}$

$g_k^T g_1 = 0$,                j = 0,...,k - 1 .

Then

$u_k = -g_k + [(g_1-g_0)^T g_k / (g_1-g_0)^T u_0] u_0 + ... + [(g_k-g_{k-1})^T g_k / (g_k-g_{k-1})^T u_{k-1}] u_{k-1}$

$\quad = -g_k + \{ g_k^T g_k / (g_k - g_{k-1})^T \left[-g_{k-1} + \left(g_{k-1}^T g_{k-1} / g_{k-1}^T g_{k-2}\right) u_{k-2}\right] \} u_{k-1}$

$\quad = -g_k + \left[g_k^T g_k / g_{k-1}^T g_{k-1}\right] u_{k-1}$.

While, for *i=0,...,k,*

$\quad g_{k+1}^T g_i - g_{k+1}^T \left[-u_i + \left(g_i^T g_i / g_{i-1}^T g_{i-1}\right) u_{i-1}\right] = 0$ .         ...(8)

(Bazaraa, 1993)

**4.2 Fletcher-Powell Algorithm:**

Here the direction are defined as:

$$s_i = -H_i g_i,$$

Where $H_i$ is initially ($i=0$) any positive-definite, symmetric matrix; therefore

$$H_i = H_{i-1} + A_{i-1} + B_{i-1}$$

with

$$A_{i-1} = \beta_i s_{i-1} g_{i-1}^T \, / \, g_{i-1}^T (g_i - g_{i-1})$$

$$B_{i-1} = H_{i-1} (g_i - g_{i-1})(g_i - g_{i-1})^T \, /(g_i - g_{i-1})^T H_{i-1}(g_i - g_{i-1})$$

$B_i$ being the step length. By using the fact that the gradients $g_i$ are eigen vectors with unit eigen value of $H_i$ for $i < j < n$.

The directions in the Fletcher-Powell and Fletcher-Reeves algorithms are respectively scalar multiples of each other.

**Theorem (2):**

If the minimum of $f(x)$, that is

$f(x) = a + b^T x + \dfrac{1}{2} x^T G x$, in the direction $u$ from the point $x_0$ is at $x_i$, $i=0,1$, then

$x_1 - x_0$ is conjugate to $u$.

**Proof :**

for $i=0$ and $1$,

$(d/d\lambda) f(x_i + \lambda u) = 0$ at $\lambda = 0$

giving $u^T(b + Gx_i) = 0$

subtraction for $i=0$ and $1$ gives

$u^T G(x_1 - x_0) = 0$

(Fletcher, 1980)

**5. Hybrid Conjugate Gradient Algorithm:**

Despite the numerical superiority of Polak-Riebiere (PR) algorithm over Fletcher-Reeves (FR) algorithm, the later has better theoretical properties than the former. Under certain conditions FR-method can be shown to have global convergence with exact line Search (Powell, 1986) and also with inexact line search satisfying the strong Wolfe-Powell condition (see Al-Baali, 1985). Normally this leads to speculation on the best way to choose $\beta_k$.

Touati –Ahmed and Storey in 1995 proposed the following hybrid algorithm.

**Step1:** If $\lambda \|g_{i+1}\|^2 \le (2\mu)^{i+1}$ , with $\dfrac{1}{2} > \mu > \sigma$ and $\lambda > 0$, go to step 2

Otherwise set $\beta_i = 0$ .

**Step2:** If $\beta_i^{PR} < 0$ set $\beta_i = \beta_i^{FR}$

$\beta_i^{PR} \le \left(\dfrac{1}{2}\mu\right)\|g_{i+1}\|^2 / \|g_i\|^2$, otherwise, go to step 3.

**Step3: If** with   $\mu > \sigma$ ,set   $\beta_i = \beta_i^{PR}$

otherwise   set   $\beta_i = \beta_i^{FR}$

Here $\mu$, $\sigma$ and $\lambda$ are supplied parameters. This hybridization was shown to be globally convergent under both exact and inexact line searches and to be quite competitive with PR-algorithm and FR-algorithm. (Touati and Storey, 1995).

Lui and Storey investigated the rate of this proposed algorithm in 1991, and they proved the following result concerning global convergence of conjugate gradient algorithms. ( Lui and Storey, 1991).

## 6. Modification of the Pair Conjugate Algorithm:

Touati and Storey in 1995 suggested the following algorithm to compute the conjugancy coefficient $\beta_i$:

**Step 1**: If $\beta_i^{PR} < 0$, then $\beta_i = \beta_i^{FR}$; return to main program. Otherwise, go to step 2.

**Step 2:** If ( $0 \le g_{i+1}^T g_i \le \|g_{i+1}\|^2$ ), then $\beta_i = \beta_i^{PR}$; return to main program.
Otherwise go to step 3.

**Step 3:** If $(cos^2(\theta_i) \ge \gamma_i^2$;   where $\gamma_i^2 = \dfrac{\tau}{\|g_i\|^2 \sum\limits_{l=1}^{i}\|g_l\|^{-2}}$ ) holds, then $\beta_i = \beta_i^{PR}$;

return to main program. Otherwise set   $\beta_i = \beta_i^{FR}$; return to main program.

In order to improve the conjugate pair (CP) algorithm, we used the hybrid approach.

Now we make a combined algorithm between CP algorithm and hybrid CG algorithm as follows:

## 7. The Outlines of the New Algorithm:

**Step (1):** *i=1*.

**Step (2):** Compute $u_i = -g_i$, line search along $d_i$ to get $x_{k+1} = x_k + \beta_k u_k$.

**Step (3):** If at *x_{i+1}* the stopping criterion $\|g_{i+1}\| \le 1 \times 10^{-5}$ is satisfied, then terminate.

**Step (4):** Check for restarting criterion if *i=n* then go to step (1). Else go to step (5).

**Step (5):** If $\lambda\|g_{i+1}\|^2 \le (2\mu)^{i+1}$   , with   $\tfrac{1}{2} > \mu > \sigma$ and $\lambda>0$ continue.
Otherwise set $\beta_i=0$ .

If $\beta_i^{PR} < 0$ set $\beta_i = \beta_i^{FR}$ .Otherwise continue.

If $\quad \beta_i^{PR} \le \left(\frac{1}{2}\mu\right)\|g_{i+1}\|^2 / \|g_i\|^2$

with $\quad \mu > \sigma \quad$, set $\beta_i = \beta_i^{PR}$

Otherwise $\quad$ set $\quad \beta_i = \beta_i^{FR}$

and compute $u_{i+1} = -g_{i+1} + \beta_i u_i$ , where $\beta_i = -\lambda_i \left[ \dfrac{v_i^T g_i}{y_i^T ( g_{i+1} - g_i )} \right]$

**Step (6):** Set *i=i+1*.

**Step (7):** If *i > 1000*, stop. Else go to step 2.

## 8. Conclusion and Results:

Several standard test functions were minimized to compare the new algorithm with standard CP algorithm. The same line search was employed in each of algorithms, and the cubic interpolation is used. We tabulate for all the algorithms the number of functions evaluations (NOF), the number of iterations (NOI). Overall totals are also given for NOF and NOI with each algorithm.

Table (1) gives the comparison between the standard CP algorithm and the new algorithm, this table indicates that the new algorithm is better than the standard CP algorithm.

**Table (1)**

**Comparative Performance of the Two Algorithms for Group of Test Functions**

| Test function | N | CP Algorithm NOI (NOF) | New Algorithm NOI (NOF) |
|---|---|---|---|
| Powell | 4 | 100 (218) | 93 (201) |
|  | 100 | 120 (300) | 105 (220) |
|  | 1000 | 215 (452) | 110 (311) |
| Wood | 4 | 223 (340) | 217 (325) |
|  | 100 | 451 (502) | 399 (480) |
| Sum | 100 | 50 (104) | 24 (72) |
| Dixon | 100 | 73 (159) | 20 (42) |
| Rosen | 100 | 53 (120) | 33 (78) |
|  | 1000 | 60 (144) | 47 (120) |
| Cubic | 100 | 40 (105) | 17 (51) |
|  | 1000 | 58 (122) | 24 (95) |
| Tri | 100 | 70 (145) | 45 (88) |
| Total |  | 1513 (2711) | 1134 (2077) |

## 9. Appendix:

1- Generalized Powell Function:

$$f = \sum_{i=1}^{n/4} [(x_{4i-3} - 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4],$$

$x_0 = (3,-1,0,1;...)^T$.

2- Generalized Wood Function:

$$f = \sum_{i=1}^{n/4} 100[(x_{4i-2} - x_{4i-3}^2)^2] + (1 - x_{4i-3})^2 + 9(x_{4i} - x_{4i-1}^2)^2 + (1 - x_{4i-1}^2)^2$$

$$+ 10.1[(x_{4i-2} - 1)^2 + (x_{4i} - 1)^2] + 19.8(x_{4i-2} - 1)^2(x_{4i} - 1),$$

$x_0 = (-3,-1,-3,-1;...)^T$.

3- Generalized Sum of Quadratics Function:

$$f = \sum_{i=1}^{n} (x_i - i)^4,$$

$x_0 = (2;...)^T$.

4- Generalized Dixon Function:

$$f = \sum_{i=1}^{n} [(1 - x_i)^2 + (1 - x_n)^2 + \sum_{i=1}^{n-1} (x_i^2 - x_{i-1})^2],$$

$x_0 = (-1;...)^T$.

5- Generalized Rosenbrock Function:

$$f = \sum_{i=1}^{n/2} [100(x_{2i} - x_{2i-1}^2)^2 + (1 - x_{2i-1})^2],$$

$x_0 = (-1.2,1;...)^T$.

6- Generalized Cubic Function:

$$f = \sum_{i=1}^{n/2} [100(x_{2i} - x_{2i-1}^3)^2 + (1 - x_{2i-1})^2],$$

$x_0 = (-1.2,1;...)^T$

7- Generalized Tri Function:

$$f = \sum_{i=1}^{n} (ix_i^2)^2,$$

$x_0 = (-1;...)^T$.

## ***REFERENCES***

[1]     Al-Baali, M. (1985), "Descent Property and Global Convergence of the Fletcher-Reeves Method with Inexact Line Search" IMA Journal of Numerical Analysis, Vol. 5, pp. 121-124.

[2]     Bazaraa, M.S. (1993) " Nonlinear Programming", England, Universities Press, London.

[3]     Fletcher, R. (1980) "Practical Methods of Optimization", Unconstrained Optimization, John Wiley and Sons, Chichester. Newyork.

[4]     Liu, Y. and Storey, C. (1991), " Efficient Generalized Conjugate Gradient Algorithms" Part 1, Theory, Journal of Optimization, Theory and Applications, Vol. 69, pp. 129-137.

[5]     Touati-Ahmed, D. and Storey, C. (1995), " Efficient Hybrid Conjugate Gradient Techniques" Plenum Publishing Corporation, England.

[6]     VanWyk, D.J. (1977) " Generalization of Conjugate Direction in the Optimization of Function". JOTA, Vol. 21, No.4.