Journal of Iraqi Al-Khwarizmi (JIKh) Volume:6 Issue:2 Year: 2022 pages: 14-26

Combined Conjugate Gradient and Quasi-Newton Methods for Unconstrained Optimization

Ivan S. Latif[1] and Mehiddin Al-Baali[2]

[1]Department of Mathematics, College of Education, Salahaddin University-Erbil, Iraq
[2]Department of Mathematics and Statistics, College of Science, Sultan Qaboos University, Oman
[1]*ivan.latif@su.edu.krd*
[2]*albaali@squ.edu.om*

Abstract

In this paper, we propose an algorithm for solving nonlinear unconstrained optimization problems by combining an extended conjugate gradient method and the damped-technique of Al Baali-Powell for the BFGS method in a sense to be defined. The combination will be considered not only at the current iteration, but also at a number of pervious iterations, some convergence properties for the proposed algorithm will be investigated and some numerical result will be reported.

Key words: Unconstrained optimization, conjugate gradient methods, quasi-Newton methods, damped and line search techniques.

Published online 5-12-2022

Introduction

Consider solving the unconstrained optimization problem

$$\min f(x), \tag{1}$$

where $f:\Re^n \to \Re$ is a continuously differentiable function, by iterative methods of the form

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

where α_k is a positive steplength and d_k is a search direction. Denoting the function value $f(x_k)$, computed at the current point x_k, by f_k and its gradient $\nabla f(x_k)$ by g_k, we consider methods which define the search direction such the descent property $d_k^T g_k < 0$ hold which insures that there exists $\alpha_k > 0$ such that f_{k+1} is sufficiently smaller than f_k.

If we take $d_k = -g_k$, we obtain the gradient method which has wide applications to large- scale optimization (see for example Nocedal and Wright, 1999). In general, the class of conjugate gradient (CG) methods is useful for solving large-scale optimization problems, because it requires storage of a few vectors. This class defines the search direction by

$$d_k = \begin{cases} -g_k & \text{if} \quad \mathrm{k} = 1 \\ -g_k + \beta_k d_{k-1} & \text{if} \quad \mathrm{k} \geq 2, \end{cases} \tag{3}$$

where β_k is a parameter that determines the different CG methods (see for example Fletcher, 1987). Well known choices of β_k are

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \beta_k^{\mathrm{PR}} = \frac{g_k^{\mathrm{T}}(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \qquad \beta_k^{\mathrm{HS}} = \frac{g_k^{\mathrm{T}}(g_k - g_{k-1})}{d_{k-1}^{\mathrm{T}} g_{k-1}} \tag{4}$$

which correspond to the FR (Fletcher-Reeves, 1964), PR (Polak-Ribière, 1969) and HS (Hestenes Stiefel, 1952), respectively. The CG methods with exact line search have a finite termination if the objective function is quadratic and strictly convex. However, in general, inexact line search technique is used in practice which yields that a method may not converge to a solution of (1).

To improve the performance of a CG method, other methods have been combined with it. In particular, Shi-Shen (2004) has combined gradient methods and Al-Bayati and Latif (2011) have combined quasi-Newton methods, which we define below, and reported encouraging numerical results. A new combined method between CG and BFGS was hybrid BFGS, that presented by Ibrahim, et al. (2014), and denoted by HBFGS. Moreover, Ibrahim, Mamet and Leong (2014) hybridized BFGS with CG, referred by BFGS-CG, which improves the performance for methods in solving large scaled unconstrained optimization problems. Forsgren and Odland (2015) studied the conditions on the basic matrix of quasi-Newton and its update correction, that belong to Broyden family with one parameter to make the method parallel in direction to CG on quadratic programming. Li, M. (2018) modified a nonlinear CG method that closed to momeryless BFGS methods in terms of search direction, which author's modification reduces to Hestenes Stiefel (1952) with exact line search. A new hybridization proposed by Aini, et al. (2019), that was between quasi-Newton methods with a modified CG named by Aini-Rivaie-Mustafa (ARM) method investigated in 2018 under exact line search. Finally, Li, X. et al. (2019) gave a new spectral CG method based on quasi-Newton direction and equation for unconstrained optimization problems.

The theoretical and practical merits of the quasi Newton methods for unconstrained optimization have been systematically explored. On each iteration of these methods, an estimate x_k of a solution to problem (1) and a positive definite Hessian approximation B_k are used to compute the search direction $d_k = -B_k^{-1} g_k$. Hence, a new estimate x_{k+1} is computed by (2). For the next iteration, B_k is updated to a new matrix in terms of the differences

$$s_k = x_{k+1} - x_k \tag{5}$$

and

$$y_k = g_{k+1} - g_k \,. \tag{6}$$

Several updating formulae have been proposed with some advantages and disadvantages (see for example Fletcher, 1987, Dennis and Schnabel, 1996, and Nocedal and Wright, 1999). We consider the most attractive update is that of BFGS, given by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^{\mathrm{T}} B_k}{s_k^{\mathrm{T}} B_k s_k} + \frac{y_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}. \tag{7}$$

This update is maintained positive definite if the curvature condition $s_k^{\mathrm{T}} y_k > 0$ holds. Whether this condition holds or not, Al-Baali and Grandinetti (2009) have shown that the performance of the BFGS method can be improved if y_k is modified before updating to the damped vector

$$\hat{y}_k = \varphi_k y_k + (1-\varphi_k) B_k s_k \,, \tag{8}$$

where $0 < \varphi_k \le 1$ is a parameter which is chosen such that $s_k^{\mathrm{T}} \hat{y}_k$ is sufficiently positive. Thus this technique maintains the damped BFGS update positive definite whether the curvature condition holds or not. It is essentially proposed by Powell (1978), for modifying the BFGS update when applied to a Lagrange function for constrained optimization, and used for the first time by Al-Baali (2004) for unconstrained optimization. Al-Baali (2014) has shown that this damped technique maintains the useful properties that the BFGS method has for convex functions and Al-Baali and Purnama (2012) have shown the usefulness of this damped technique when applied to a simple quadratic and ill conditioned problem. One of the authors recommended choices for φ_k is given by

$$\varphi_k = \begin{cases} \dfrac{\sigma_2}{1-\rho_k} & \text{if } \rho_k \le 0 \text{ or } \rho_k < \min(1-\sigma_2, \dfrac{h_k}{1+\sigma_4}) \\ \dfrac{\sigma_3}{\rho_k - 1} & \text{if } \rho_k > 1+\sigma_3 \\ 1 & \text{otherwise,} \end{cases} \quad (9)$$

where

$$\rho_k = \frac{s_k^T y_k}{s_k^T B_k s_k}, \qquad h_k = \frac{y_k^T B_k^{-1} y_k}{s_k^T y_k}, \quad (10)$$

$0 < \sigma_2 \le 1$, $\sigma_3 \ge 3$ and $\sigma_4 \ge 0$. Note that the condition $s_k^T \hat{y}_k > 0$ always holds and the values of $\sigma_2 = 0.8$, $\sigma_3 = \infty$, $\sigma_4 = 0$ reduce the above choice of Al-Baali for φ_k to that of Powell (1978), except when $\rho_k = h_k$ which reduces the Broyden family of updates to a single update.

The aim of this paper is to apply the damped quasi-Newton technique to certain combined CG and quasi-Newton methods as described in the next section. In Section 3, we obtain the convergence property of the methods. In Section 4, we describe some numerical results based on applications to a set of standard test problems. It is shown that the proposed combined methods are competitive with the robust BFGS method.

Combining Conjugate Gradient and Damped Quasi-Newton Methods

In this section, we propose an algorithm on the basis of combining the CG and damped BFGS methods in the following way. We first apply the damped technique to the search direction of Al-Bayati and Latif (2011) to obtain

$$d_k = -\gamma_k g_k + \frac{1}{r}\sum_{i=1}^{r} \beta_{k-i} d_{k-i}, \quad (11)$$

where γ_k is a suitable positive scalar, $r \ge 2$ is a preset integer and β_k is a CG parameter (eg, see (4)) or given by

$$\beta_k = \max\{\frac{s_k^T \hat{y}_k}{s_k^T B_k s_k}, 0\}, \quad (12)$$

where B_k is a quasi-Newton Hessian approximation. The search direction (11) is reduced to that of Al-Bayati and Latif (2011) if $\hat{y}_k = y_k$ (ie, $\varphi_k = 1$) for all values of k and to that of conjugate gradient direction if $\gamma_k = 1$, $r = 1$ and β_{k-1} is given by a conjugate gradient parameter (eg, by one choice in (4)).

Similarly, applying the damped technique to the search direction of Al-Bayati and Latif (2011) for the combined CG and quasi-Newton methods, it follows that

$$d_k = \begin{cases} -B_k^{-1} g_k & \text{if} \quad k \le r-1 \\ -B_k^{-1}\{\beta_k g_k + \sum_{i=2}^{r} \beta_{k-i+1} d_{k-i+1}\} & \text{if} \quad k > r-1, \end{cases} \quad (13)$$

Here, we consider the damped BFGS Hessian which is defined by (7) with y_k replaced by $\hat{y}_k$ (given by (11)) that is

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\hat{y}_k \hat{y}_k^T}{s_k^T \hat{y}_k}. \quad (14)$$

We note that search direction (13) is reduced to the damped $BFGS$ direction for sufficiently large value of r and hence to the $BFGS$ direction when $\hat{y}_k = y_k$. Thus for a fixed value of r, the corresponding method is reduced in the limit to that of the damped $BFGS$ method so that the global and superlinear convergence property on convex functions is obtained (for details, see Al-Baali, 2014). It is assumed that the steplength α_k is chosen such that the following Wolfe-Powell conditions hold:

$$f(x_k + \alpha_k d_k) \le f_k + \sigma_0 \alpha_k d_k^T g_k, \qquad (15)$$

$$d_k^T g_{k+1} \ge \sigma_1 d_k^T g_k, \qquad (16)$$

where $0 < \sigma_0 < 0.5$ and $\sigma_0 < \sigma_1 < 1$ and $\alpha_k = 1$ is tried first. Note that condition (16) ensures the curvature condition so that the BFGS update is maintained positive definite.

For simplicity, the Armijo (1966) rule is sometimes used for defining α_k as follows. Setting scalars $\rho \in (0,1)$ and a starting value α, we choose α_k be the largest value in the set $\{\alpha, \alpha\rho, \alpha\rho^2, \alpha\rho^3, ...\}$ such that

$$f(x_k + \alpha_k d_k) \le f_k - \sigma_0 \alpha_k g_k^T g_k \qquad (17)$$

which is like condition (15) with $d_k^T g_k$ replaced by $-g_k^T g_k$. Although this line search condition does not guarantee the curvature condition, the damped BFGS update is maintained positive definite.

We now outline the proposed algorithm which combines CG with the damped quasi-Newton methods as follows.

Algorithm 2.1

Step 1: Given a starting point x_0, a symmetric and positive definite matrix B_0, $2 \le r \le n$, and values for σ_0, σ_1, σ_2, σ_3, σ_4 (such that $0 < \sigma_0 < 0.5$, $\sigma_0 < \sigma_1 < 1$, $0 < \sigma_2 \le 1$, $\sigma_3 \ge 3$, $\sigma_4 \ge 0$), a tolerance $\varepsilon > 0$, and set $k = 0$.

Step 2: If the Euclidean norm $\|g_k\| \le \varepsilon$ holds, then stop.

Step 3: Compute the search direction as defined by (13).

Step 4: Find a step length α_k such that condition (17) holds and compute a new point x_{k+1} by (2) and hence the gradient g_{k+1} at this point.

Step 5: Compute s_k, y_k and $\hat{y}_k$**,** using (5), (6) and (8), with (9)-(10), respectively.

Step 6: Update B_k to B_{k+1}, using formula (14).

Step 7: Set $k = k + 1$ and go to Step 2.

Although the line search condition in Step 1 does not guarantee the curvature condition, the damped technique in Step 5 ensures that $s_k^T \hat{y}_k > 0$ which is required in Step 6 to maintain the damped BFGS update positive definite. However, if α_k is computed such that the Wolfe-Powell conditions (15) and (16) hold, then the condition $s_k^T y_k > 0$ holds which yields the condition $s_k^T \hat{y}_k > 0$ for any value of $0 < \varphi_k \le 1$. Thus, we consider the following algorithm.

Algorithm 2.2

It is defined by Algorithm 2.1 with Step 4 is replaced by

Step 4: Find α_k such that the Wolfe conditions (15) and (16) hold and compute a new point x_{k+1} by (2) and hence the gradient g_{k+1} at this point.

We note that the above two algorithms are reduced in the limit to that of the damped BFGS method of Al-Baali (2014). They satisfy the descent property as shown in the next section. Since Algorithm 2.2 uses the Wolfe-Powell conditions on α_k, it has the global and superlinear convergence property that the BFGS method has for uniformly convex functions (see Al-Baali, 2014, for details). However, since Algorithm 2.1 may not satisfy this property, we study its convergence properties in the next section.

Convergence Properties

To ensure that Algorithm 2.1 converges globally, consider the following standard assumptions:

H_1**:** The level set $\Omega = \{x \in \Re^n \,|\, f(x) \le f(x_0)\}$ is bounded.

H_2: In some neighborhood N of Ω, f is continuously differentiable and its gradient satisfies the Lipschitz condition

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad \forall \mathrm{x}, \mathrm{y} \in N \tag{18}$$

for some constant $L > 0$.

Theorem 3.1

Let d_k be defined by (2)and (13) for some $\gamma_k > 0$ and β_k be given by one formula in (4).Then the descent property $d_k^T g_k < 0$ holds for all k.

Proof.

From (2) and (3) we get

$$\beta_{k-i} = \|g_k\|^2 \psi_{ki} \tag{19}$$

where

$$\psi_{ki} = \max_{2 \le i \le r}\left(\frac{g_k^{\mathrm{T}} d_k}{\gamma_k}, 0\right). \tag{20}$$

If CG direction is used, then

$$\begin{aligned}
-\gamma_k \|g_k\|^2 + \beta_{k-i} g_k^{\mathrm{T}} d_{k-i} &\le -\gamma_k \|g_k\|^2 + \beta_{k-i} \max\{d_{k-i}^{\mathrm{T}} \hat{y}_k, 0\} \\
&\le -\gamma_k \|g_k\|^2 + \|g_k\|^2 \psi_{ki}^{\mathrm{T}} \max\{d_{k-i}^{\mathrm{T}} g_{k-i}, 0\} \\
&\le -\gamma_k \|g_k\|^2 + \|g_k\|^2 \psi_{ki}^{\mathrm{T}} \psi_{ki} \\
&\le 0
\end{aligned}$$

Then we obtain the following the descent property

$$\begin{aligned}
g_k^{\mathrm{T}} d_k &= -\gamma_k \|g_k\|^2 + \frac{1}{r}\sum_{i=1}^{r} \beta_{k-i+1} g_k^{\mathrm{T}} d_k \\
&\le \frac{1}{r}\sum_{i=1}^{r} \{-\gamma_k \|g_k\|^2 + \beta_{k-i+1} g_k^{\mathrm{T}} g_{k-i}\} \\
&\le 0.
\end{aligned}$$

Since the damped methods guarantee that the Hessian approximations are maintained positive definite and bounded (see Al-Baali, 2014), so that $\| B_k^{-1} \| \leq c,$ where c is a positive constant, it follows from (13) and the above result that Algorithms 2.1 and 2.2 satisfy the descent property.

Theorem 3.2

The Algorithm 2.1 generates an infinite sequence $\{x_k\}$ if (H_1) and (H_2) hold, then

$$\sum_{k=r}^{\infty} \frac{\|g_k\|^4}{\gamma_k} < +\infty, \tag{21}$$

where

$$\gamma_k = \max_{2\leq i\leq r}(\|g_k\|^2, \|d_{k-i+1}\|^2). \tag{22}$$

Proof.

Since $\{f_k\}$ is a decreasing sequence and satisfies Assumption (H_1) and Assumption (H_2) it is a convergent sequence .In particular, we have by (H_1) and (16) that

$$\sum_{k=1}^{\infty} (f_k - f_{k+1}) < +\infty \tag{23}$$

Using modified Armijo line search rule (15), it follows that there exist $\eta > 0$ such that

$$f_k - f_{k+1} \geq \eta \frac{\|g_k\|^4}{\gamma_k}$$

Since $\gamma_k = \max_{2\leq i\leq r}(\|g_k\|^2, \|d_{k-i+1}\|^2)$

Let $K_1 = \{k \| \alpha_k = s\}, \qquad K_2 = \{k \| \alpha_k < s\}$

We have

$$f_k - f_{k+1} \geq s\mu_1 g_k^{\mathrm{T}} g_k$$

Because of $\|g_k\|^2 \leq \gamma_k$, we have

$$f_k - f_{k+1} < -2\mu_1\alpha_k g_k^{\mathrm{T}} d_k$$

This together with (23) implies that (21) holds.

We now show that the class of combined conjugate gradient and quasi-Newton methods satisfies the descent property.

Theorem 3.3

If the conditions in Theorem 4.2 hold, then either $\lim_{k\to\infty} \|g_k\| = 0$ or $\{x_k\}$ has no bound.

Proof.

If $\lim_{k\to\infty} \|g_k\| \neq 0$, then there exists an infinite subset $\Omega_0 \subset \{r, r+1, \ldots\}$ and $\varepsilon > 0$ such that:

$$\|g_k\| > \varepsilon, \ \mathrm{k} \in \Omega_0. \tag{24}$$

Thus

$$\frac{\varepsilon^4}{\gamma_k} \leq \frac{\|g_k\|^4}{\gamma_k} \qquad \forall \ k \in \Omega_0$$

(25)

By Theorem 3.2 and for $k \geq 1$, we obtain

$$\|d_k\|^2 \leq \max_{1\leq i\leq r}\{\|g_i\|^2\}\ . \tag{26}$$

Now if $k \leq r$, then the conclusion is obvious. Otherwise, $k > r$, by induction process, we obtain the following conclusion:

$$\sum_{k\in\Omega_0}\frac{\varepsilon^4}{\gamma_k} \leq \sum_{k=r}^{+\infty}\frac{\|g_k\|^4}{\gamma_k} < +\infty \tag{27}$$

Then there exists at least one $i : 2 \leq i \leq r$ such that: $\lim_{k\in\Omega_0, k\to\infty}\|d_{k-i+1}\| = +\infty$

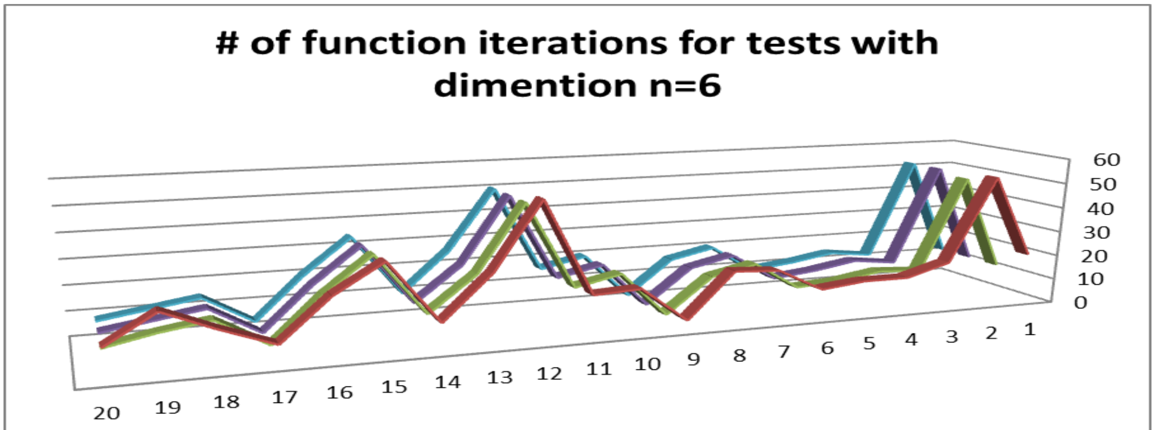
Therefor $\{x_k\}$ has no bound.

Numerical Results

We now test the behavior of Algorithm 2.1, using $B_0 = I$, the identity matrix, and the three values of *r=2, 4* and *10.* For comparison, we also consider the value of $r = \infty$ which yields the standard BFGS method. The values for the other parameters, required in Step 1 of Algorithm 2.1, which we used are $\sigma_0 = 0.01$, $\sigma_1 = 0.5$, $\sigma_2 = 0.9$, $\sigma_3 = 9$, $\sigma_4 = 10^{-4}$, $\varepsilon = 10^{-6}$. The line search subroutine, which is required in Step 4 to compute the step length α_k always tried the value of $\alpha_k = 1$ first.

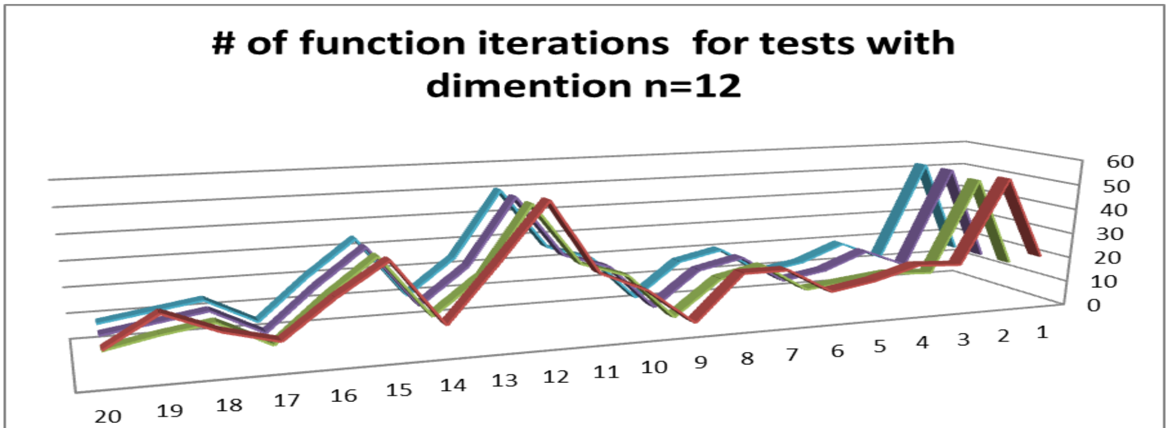
We applied the above four algorithms to a set of 80 test problems obtained as follows. We used 20 different types of standard minimization problems (their functions are described by Andrei, 2005, and listed in the Appendix). We used these problems with four values of dimension *n=6, 12, 120* and 36*0* so that the total number of test problems is 80. The number of iterations (referred to as NIT) which are required to solve each problem in the set for these values of *n* are given in Figures 1, 2, 3 and 4, respectively. For a clarity comparison of these algorithms, we also used these numbers to obtain the plots in these figures. Note that the number of gradient evaluations is equal to NIT.
Similarly, a comparison of the number of function evaluations (referred to as NFE) which are required to solve the test problems is illustrated in Figures 5, 6, 7 and 8.

The eight figures clearly show that the proposed algorithms solved all problems successfully. They also show that the performance of the proposed algorithms is competitive with that of the BFGS method for all values of *r* and *n*. In some cases, for the problems 10, 11, 12, 13 and 19, and *n*=120 and 360 (see Figures 3, 4, 7 and 8), the proposed algorithms work substantially better than the $BFGS$ methods.



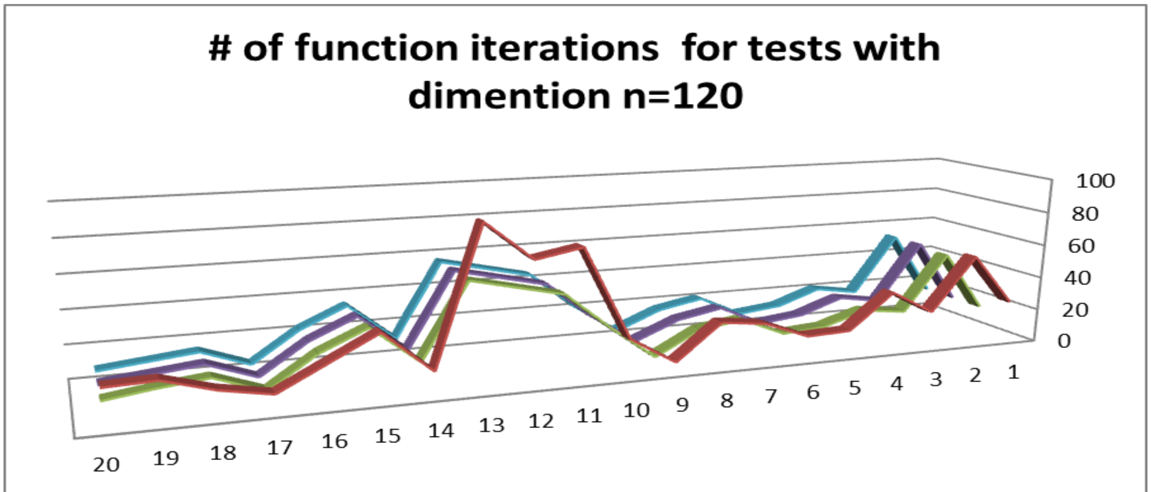
	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
BFGS (r=infinity)	12	23	15	8	24	35	11	28	54	17	18	4	22	21	12	14	14	19	53	20
r=2	7	11	14	3	20	34	10	24	50	16	20	2	16	20	9	10	13	12	50	12
r=4	8	11	14	3	20	34	10	24	50	16	20	2	16	20	9	11	14	12	52	12
r=10	8	11	14	3	20	34	10	26	50	16	20	2	16	20	9	11	14	12	52	12

Figure 1: NIT required to solve problems with dimension n =6
by Algorithm 2.1 for *r=2, 4, 10* and $r = \infty$ (the *BFGS* method)



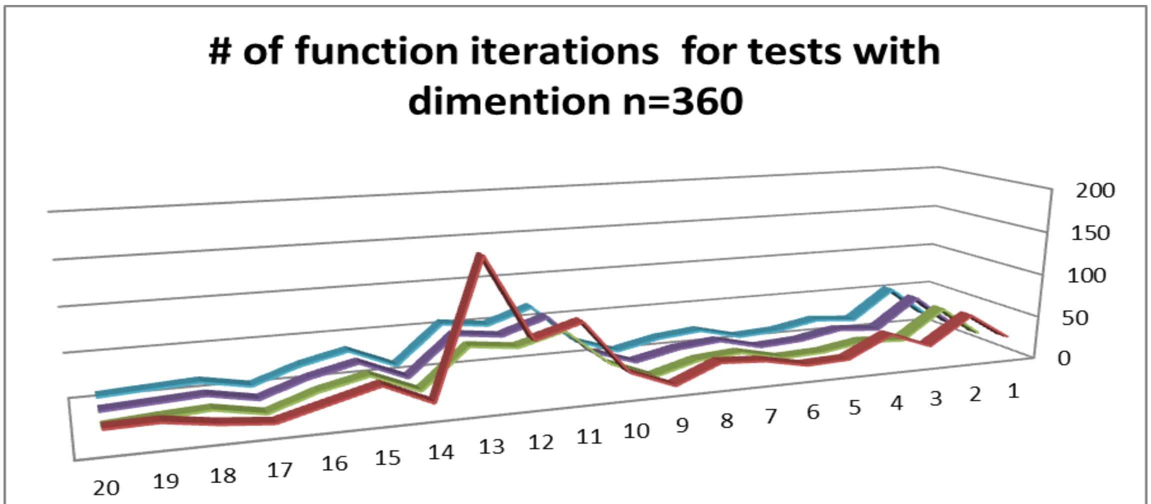
	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
BFGS (r=infinity)	12	23	15	10	24	36	11	32	54	26	18	4	22	22	12	15	20	19	53	20
r=2	7	11	14	4	20	34	10	24	50	26	20	2	16	20	9	11	13	12	50	14
r=4	8	11	14	4	20	34	10	24	50	26	20	2	16	20	9	12	19	12	52	14
r=10	8	11	14	4	20	34	10	24	50	26	20	2	16	20	9	12	19	12	52	14

Figure 2: NIT required to solve problems with dimension n =12
by Algorithm 2.1 for *r=2, 4, 10* and $r = \infty$ (the *BFGS* method)



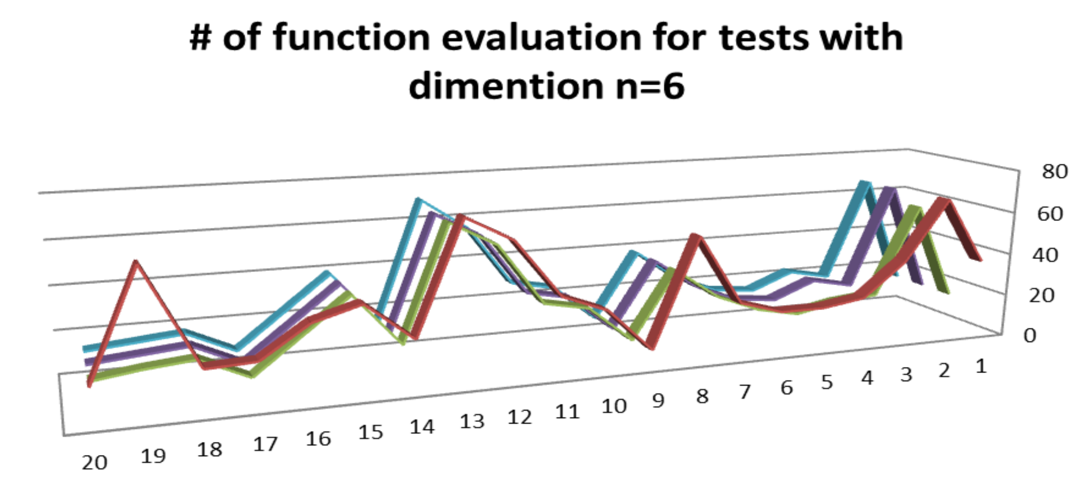
	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
BFGS (r=infinity)	22	23	15	10	24	38	14	93	71	76	22	6	27	24	14	15	36	22	54	24
r=2	7	11	14	4	22	34	12	56	50	44	22	2	16	20	9	11	19	16	50	15
r=4	8	11	14	4	22	34	12	56	50	44	22	5	16	21	9	12	21	17	52	15
r=10	8	11	14	4	22	34	12	56	50	44	22	5	16	21	9	12	21	17	52	15

Figure 3: NIT required to solve problems with dimension n =120
by Algorithm 2.1 for *r=2, 4, 10* and $r=\infty$ (the *BFGS* method)



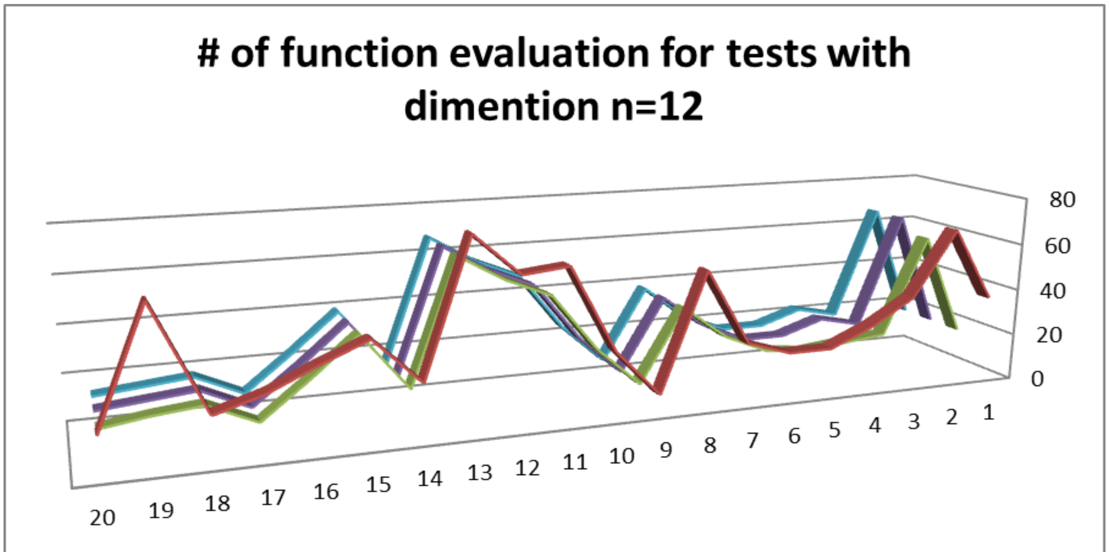
	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
BFGS (r=infinity)	22	23	15	10	24	38	14	163	71	86	26	6	27	24	14	16	40	22	54	24
r=2	7	11	14	4	22	34	12	56	50	64	22	2	16	20	9	11	19	16	50	16
r=4	8	11	14	4	22	34	12	56	50	67	22	5	16	21	9	12	21	17	52	16
r=10	8	11	14	4	22	34	12	56	50	67	22	5	16	21	9	12	21	17	52	16

Figure 4: NIT required to solve problems with dimension n =360
by Algorithm 2.1 for *r=2, 4, 10* and $r=\infty$ (the *BFGS* method)



	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
BFGS (r=infinity)	16	63	19	20	34	40	23	73	62	36	28	8	58	26	20	20	23	39	67	36
r=2	12	15	17	7	23	39	15	67	56	28	24	7	38	24	16	12	16	18	60	15
r=4	13	15	17	7	23	39	15	67	56	28	24	7	38	24	16	14	22	18	66	15
r=10	13	15	17	7	23	39	15	70	56	28	24	7	38	24	16	14	22	18	66	15

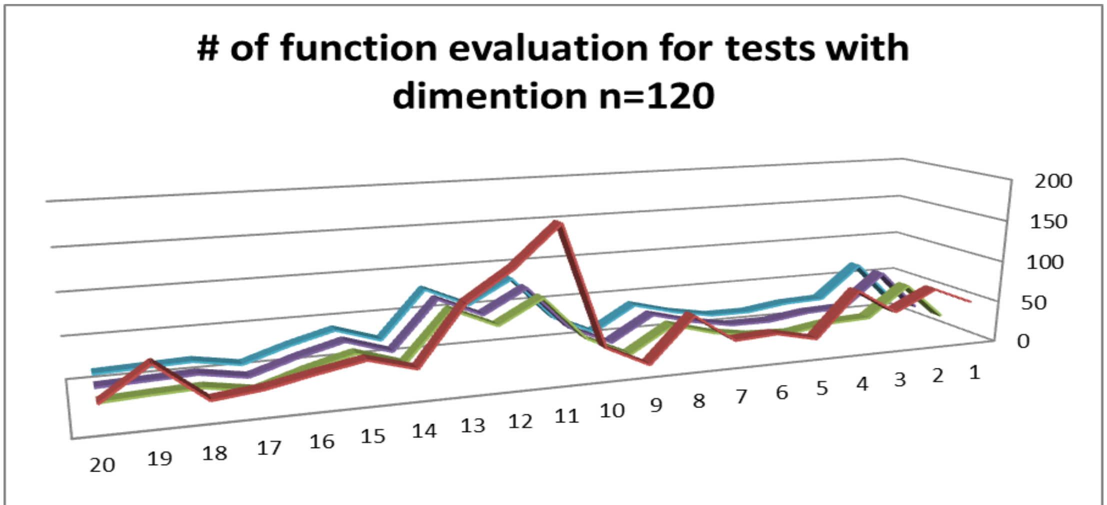
Figure 5: NFE required to solve problems with dimension $n=6$ by Algorithm 2.1 for *r=2, 4, 10* and $r=\infty$ (the *BFGS* method)



	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
BFGS (r=infinity)	16	63	19	25	34	42	23	79	62	64	28	8	58	26	20	20	28	39	67	36
r=2	12	15	17	8	23	39	15	67	56	48	24	7	38	24	16	14	16	18	60	17
r=4	13	15	17	8	23	39	15	67	56	48	24	7	38	24	16	16	22	18	66	17
r=10	13	15	17	8	23	39	15	67	56	48	24	7	38	24	16	16	22	18	66	17

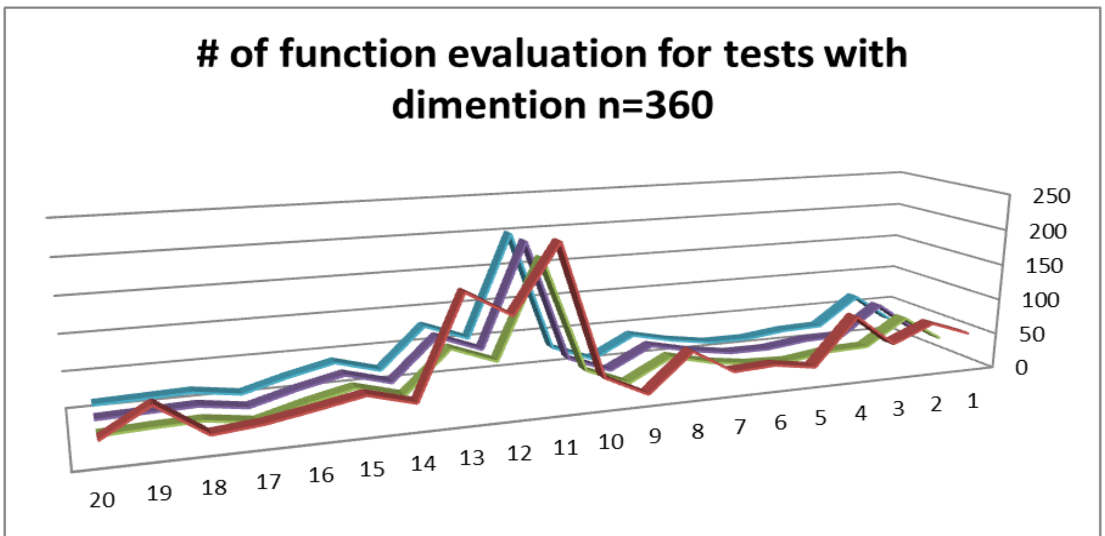
Figure 6: NFE required to solve problems with dimension $n=12$

by Algorithm 2.1 for r=2, 4, 10 and $r = \infty$ (the $BFGS$ method)



	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
BFGS (r=infinity)	27	63	19	25	36	46	31	98	132	178	34	9	62	28	31	20	74	44	68	48
r=2	12	15	17	8	25	39	22	79	56	84	31	7	38	24	16	14	22	24	61	18
r=4	13	15	17	8	25	39	22	79	56	84	31	8	38	25	16	16	23	25	66	18
r=10	13	15	17	8	25	39	22	79	56	84	31	8	38	25	16	16	23	25	66	18

Figure 7: NFE required to solve problems with dimension n =120
by Algorithm 2.1 for r=2, 4, 10 and $r = \infty$ (the $BFGS$ method)



	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
BFGS (r=infinity)	27	63	19	25	36	48	31	167	132	222	38	9	62	28	31	22	88	44	68	48
r=2	12	15	17	8	25	39	22	79	56	189	31	7	38	24	16	14	22	24	61	26
r=4	13	15	17	8	25	39	22	79	56	200	31	8	38	25	16	16	23	25	66	26
r=10	13	15	17	8	25	39	22	79	56	200	31	8	38	25	16	16	23	25	66	26

Figure 8: NFE required to solve problems with dimension n =360

by Algorithm 2.1 for *r=2, 4, 10* and $r = \infty$ (the $BFGS$ method)

Conclusion

We proposed an algorithm which combine the damped quasi-Newton methods and conjugate gradient method. For simplicity, we used the inexact Armijo line search strategy and show that the number r of terms in the search direction affects the numerical performance and proposed the algorithm. It is competitive with the robust BFGS method in terms of iterations and function and gradient evaluations. However, further improvements are expected.

الملخص

نقترح في هذا البحث خوارزمية لحل مسائل أمثليات غير مقيدة وغير خطيه وذلك بتركيب طريقة التدرج المترافقة مع نموذج البعلي–باول في تخامد طريقة ب.ف.ج.س. في حل مسائل في الامثليات غير المقيدة .وقد تم اثبات ان هذه الطريقة تمتلك خاصية التقارب الشامل وقد اظهرته النتائج العددية.

References

1. Aini N, Hajar N, Rivaie M, Mamat M. (2018), A modified conjugate gradient coefficient under exact line search for unconstrained optimization. *Int J Eng Technol*;7(3.28 Special Issue 28):84-88.
2. Aini N, Mamat M, Rivaie M, Sulaiman IM. (2019), A Hybrid of Quasi-Newton Method with CG Method for Unconstrained Optimization. *J Phys Conf Ser*;1366(1). doi:10.1088/1742-6596/1366/1/012079
3. Al-Baali, M. (2004), Quasi-Wolfe conditions for quasi-Newton methods on large scaleoptimization problems," *40th Workshop on "Large Scale Nonlinear Optimization*, Erice, Italy, June 22 - July 1.
4. Al-Baali, M. (2014), Damped techniques for enforcing convergence of quasi-Newton methods, *Optimization Methods and Software*, 29, 919-936.
5. Al-Baali, M. and Grandintti, L. (2009), On practical modification of quasi-Newton BFGS method, *Advanced Modeling and Optimization*, 11, 63-76.
6. Al-Baali, M. and Purnama, A. (2012), Numerical experience with damped quasi-Newton optimization methods when the objective function is quadratic, *Sultan Qaboos University Journal for Science*, 17, 1-11.
7. Al-Bayati, A. and Latif, I. (2011), A new CG-algorithm with self–scaling vm-update for unconstraint optimization, *AAM Journal for Application and Applied Mathematics*, 6, 1-12.
8. Andrei, N. (2005), A new gradient descent method for unconstrained optimization, *Research institute for informatics 8-10 Averscu Avenu,* Bucharest, Romania.
9. Armijo, L. (1966), Minimization of function having Lipschitz continuous first partial derivatives, *Pacific Journal of Mathematics*, 16, 1-13.
10. Dennis, J.E. and Schnabel, R.B. (1996), Numerical Methods for Unconstrained Optimization and Nonlinear Equations, *SIAM publications*, Philadelphia.
11. Fletcher, R. (1987), Practical Methods of Optimization, *Willey*, Chester, England.
12. Fletcher, R. and Reeves, C. (1964), Function minimization by conjugate gradients, *Computer J.*, 7, 149-154.
13. Forsgren A, Odland T. (2015), On the connection between the conjugate gradient method and quasi-Newton methods on quadratic problems. *Comput Optim Appl*;60(2):377-392
14. Hestenes, M. and Stiefel, E. (1952), Methods of conjugate gradients for solving linear systems, J. *Res. Nat Bureau Stand.*, 29, 409-430.
15. Ibrahim MAH Bin, Mamat M, June LW. (2014), BFGS method: A new search direction. *Sains Malaysiana*. ;43(10):1591-1597.
16. Ibrahim MAH, Mamat M, Leong WJ. (2014),The hybrid BFGS-CG method in solving unconstrained optimization problems. *Abstr Appl Anal*
17. Li M. (2018),A modified Hestense–Stiefel conjugate gradient method close to the memoryless BFGS quasi-Newton method. *Optim Methods Softw*;33(2):336-353. doi:10.1080/10556788.2017.1325885
18. Li X, Shi J, Dong X, Yu J. (2019),A new conjugate gradient method based on Quasi-Newton equation for unconstrained optimization. *J Comput Appl Math.*;350:372-379. doi:10.1016/j.cam.2018.10.035
19. Nocedal, J. (1980), Updating quasi-Newton matrices with limited storage, *Mathematics of Computation*, 35, 773-782.
20. Nocedal, J. and Wright, S. (1999), Numerical Optimization, *Spring Series in Operations Research, Springer Verlag*, New York, USA.

21. Polak, E. and Ribière, G. (1969), Note sur la convergence des methods de direction conjuguees. revue ferue francaised, *Inform. Rec. Oper.*, 16, 35-43.
22. Powell, M.J.D. (1978), Algorithms for nonlinear constraints that use Lagrange functions, *Math. Programming*, 14, 224-248.
23. Shi, Z. and Shen, J. (2004), A gradient-related algorithm with inexact line searcher, *J. Computational and Applied Mathematics*, 170, 349-370.