DESCENT PROPERTY AND GLOBAL CONVERGENCE OF A NEW SEARCH DIRECTION METHOD FOR UNCONSTRAINED OPTIMIZATION

MOHAMMED BELLOUFI AND RACHID BENZINE

ABSTRACT. Conjugate gradient methods are probably the most famous iterative methods for solving large scale optimization problems in scientific and engineering computation, characterized by the simplicity of their iteration and their low memory requirements. It is well known that the search direction plays a main role in the line search method. In this paper, we propose a new search direction with the Wolfe line search technique for solving unconstrained optimization problems. Under the above line searches and some assumptions, the global convergence properties of the given methods are discussed. Numerical results and comparisons with other CG methods are given.

1. Introduction

Consider the unconstrained optimization problem

$$\{\min f(x), \quad x \in \mathbb{R}^n\}, \tag{1.1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is continuously differentiable. The line search method usually takes the following iterative formula

$$x_{k+1} = x_k + \alpha_k d_k \tag{1.2}$$

where x_k is the current iterate point, $\alpha_k > 0$ is a steplength and d_k is a search direction. Different choices of d_k and α_k will determine different line search methods([14,15,16]).

We denote $f(x_k)$ by f_k, $\nabla f(x_k)$ by g_k and $\nabla f(x_{k+1})$ by g_{k+1}, respectively. $\|.\|$ denotes the Euclidian norm of vectors and define $y_k = g_{k+1} - g_k$. In this article, as in other algorithms and convergence analysis, the steplength $\alpha_k > 0$ is computed by carrying out certain line searches. The strong Wolfe search is to find a positive steplength α_k such that:

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k \tag{1.3}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le -\sigma g_k^T d_k \tag{1.4}$$

where $\delta \in \left]0, \frac{1}{2}\right[$ and $\sigma \in]\delta, 1[$.

The steepest descent method is one of the simplest and the most fundamental minimization methods for unconstrained optimization. Since it uses the negative gradient as its descent direction, it is also called the gradient method.

For many problems, the steepest descent method is very slow. Although the method usually works well in the early steps, as a stationnary point is approached, it descends very slowly with

---

Date: October 06, 2013
AMS 2000 subject classifications: 90C26; 65H10 .
Key words and phrases. Unconstrained optimization, Conjugate gradient method, strong Wolfe line search, Global convergence.
This paper is in final form and no version of it will be submitted for publication elsewhere.

zigzaguing phenomena. There are some ways to overcome these difficulties of zigzagging by defleting the gradient. Rather then moving along $d_k = -\nabla f(x_k) = -g_k$, we can move along

$$d_k = -D_k \nabla f(x_k),$$

or along

$$d_k = -g_k + h_k, \tag{1.5}$$

where D_k is an appropriate matrix and h_k is an appropriate vector.

In this work we try to accelerate the convergence of the gradient method by introducing a new direction d_k^{BB} defined as follows:

$$d_k^{BB} = \begin{cases} -\frac{g_k}{\|g_k\|^2} & \text{if } k = 1 \\ -\frac{1}{\|g_k\|^2} g_k + d_{k-1} & \text{if } k \geq 2 \end{cases} \tag{1.6}$$

Note that our directions d_k^{BB} and those of different versions of conjugate gradient methods are of the form (1.5). By using (1.2), (1.3), (1.4) and (1.6), we get a new algorithm noted $CGBB$. The main aim of this note is to show that the descent property holds for all k and the global convergence is achieved for the strong Wolfe search (1.3), (1.4).

On the other hand, we can consider that our algorithm is very close to one of the versions of the conjugate gradient method.

The iterative formula of the conjugate gradient method is given by (1.2), where α_k is a steplength which is computed by carrying out a line search, and d_k is the search direction defined by

$$d_{k+1} = \begin{cases} -g_k & \text{si } k = 1 \\ -g_{k+1} + \beta_k d_k & \text{si } k \geq 2 \end{cases} \tag{1.6}$$

where β_k is a scalar and $g(x)$ denotes $\nabla f(x)$. If f is a strictly convex quadratic function, namely,

$$f(x) = \frac{1}{2} x^T H x + b^T x, \tag{1.7}$$

where H is a positive definite matrix and if α_k is the exact one-dimensional minimizer along the direction d_k, i.e.,

$$\alpha_k = \arg\min_{\alpha > 0} \{ f(x + \alpha d_k \} \tag{1.8}$$

then (1.2)–(1.6) is called the linear conjugate gradient method. Otherwise, (1.2)–(1.6) is called the nonlinear conjugate gradient method.

Conjugate gradient methods differ in their way of defining the scalar parameter β_k. In the literature, there have been proposed several choices for β_k which give rise to distinct conjugate gradient methods. The most well known conjugate gradient methods are the Hestenes–Stiefel (HS) method [08], the Fletcher–Reeves (FR) method [07], the Polak-Ribière-Polyak (PR) method [11,113], the Conjugate Descent method(CD) [06], the Liu-Storey (LS) method [10], the Dai-Yuan (DY) method [05], and Hager and Zhang (HZ) method [09]. The update parameters of these methods are respectively specified as follows:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k},\ \beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2},\ \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2},\ \beta_k^{CD} = -\frac{\|g_{k+1}\|^2}{d_k^T g_k},$$

$$\beta_k^{LS} = -\frac{g_{k+1}^T y_k}{d_k^T g_k},\ \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k},\ \beta_k^{HZ} = \left(y_k - 2 d_k \frac{\|y_k\|^2}{d_k^T y_k} \right)^T \frac{g_{k+1}}{d_k^T y_k}$$

The convergence behavior of the above formulas with some line search conditions has been studied by many authors for many years. The FR method with an exact line search was proved to globally convergent on general functions by Zoutendijk [18]. However, the PRP method and the HS method with exact and inexact line searchs are not globally convergent, see Powell's counterexample

[12]. Compared to the PRP and HS conjugate gradient method, our new algorithm is globally convergent. Numerical tests show that our algorithm accelerates the convergence of the gradient method and is at least as efficient as the other conjugate gradient methods.

This paper is organized as follows. In the next section, the New algorithms are stated and descent property is presented. The global convergence of the new methods are established in Section 3. Numerical results and a conclusion are presented in Section 4 and in Section 5, respectively.

2. CGBB Algorithm

In this section, we give the specific form of the proposed new conjugate gradient method. As reported before our search directions d_k^{BB} are defined as follows:

$$d_k^{BB} = \begin{cases} -\frac{g_k}{\|g_k\|^2} & \text{if } k = 1 \\ -\frac{1}{\|g_k\|^2} g_k + d_{k-1} & \text{if } k \geq 2 \end{cases} \tag{2.1}$$

CGBB Algorithm

The algorithm is given as follows:

Algorithm 1. *Step 0: Given $x_1 \in \mathbb{R}^n$, set $d_1^{BB} = -\frac{g_1}{\|g_1\|^2}$, $k := 1$.*

Step 1: If $\|g_k\| = 0$ then stop else go to Step 2.

Step 2: Set $x_{k+1} = x_k + \alpha_k d_k^{BB}$ where d_k^{BB} is defined by (2.1), and α_k is defined by (1.3) and (1.4).

Step 3. Set $k := k+1$ and go to Step 1.

The following theorem indicates that, if α_k is computed by the Wolfe line search (1.3) and (1.4) , then the search direction d_k^{BB} satisfies the descent property.

Theorem 1. *If the steplength α_k is computed by the Wolfe line search (1.3) and (1.4) with $\delta < \sigma < \frac{1}{2}$, , then for the proposed conjugate gradient method, the inequality*

$$-\sum_{j=0}^{k-1} \sigma^j \leq g_k^T d_k \leq -2 + \sum_{j=0}^{k-1} \sigma^j \tag{2.2}$$

holds for all k, and hence the descent property

$$g_k^T d_k < 0, \forall k \tag{2.3}$$

holds, as long as $g_k \neq 0$.

Proof. The proof is by induction. For $k = 1$ Equations (2.2) and (2.3) is clearly satisfied.

Now we suppose that (2.2) and (2.3) hold for any $k \geq 1$.

It follows from the definition (2.1) of d_{k+1} that

$$g_{k+1}^T d_{k+1} = -1 + g_{k+1}^T d_k \tag{2.4}$$

and hence from (1.4) and (2.3) that

$$-1 + \sigma g_k^T d_k \leq g_{k+1}^T d_{k+1} \leq -1 - \sigma g_k^T d_k \tag{2.5}$$

Also, by induction assumption (2.2), we have

$$-\sum_{j=0}^{k} \sigma^j = -1 - \sigma \sum_{j=0}^{k-1} \sigma^j \leq g_{k+1}^T d_{k+1}$$

$$\leq -1 + \sigma \sum_{j=0}^{k-1} \sigma^j = -2 + \sum_{j=0}^{k} \sigma^j$$

Then, (2.2) holds for $k+1$.

Since

$$g_{k+1}^T d_{k+1} \leq -2 + \sum_{j=0}^{k} \sigma^j \tag{2.6}$$

and

$$\sum_{j=0}^{k} \sigma^j < \sum_{j=0}^{\infty} \sigma^j = \frac{1}{1-\sigma} \tag{2.7}$$

where $\sigma \in \left]0, \frac{1}{2}\right]$, it follows from $1-\sigma > \frac{1}{2}$ that $-2 + \sum_{j=0}^{k} \sigma^j < 0$. Hence, from (2.6), we obtain $g_{k+1}^T d_{k+1} < 0$. We complete the proof by induction. □

3. Global convergence

In order to establish the global convergence of the proposed method, we assume that the following assumption always holds, i.e. Assumption 3.1 :

Assumption 3.1 :

Let f be twice continuously differentiable, and the level set $L = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_1)\}$ be bounded

Theorem 2. *Suppose that x_1 is a starting point for which Assumption 3.1 holds. Consider the New method (1.2) and (2.1). If the steplength α_k is computed by the strong Wolfe line search (1.3) and (1.4) with $\delta < \sigma < \frac{1}{2}$, then the method is globally convergent, i.e.,*

$$\liminf_{k \longrightarrow \infty} \|g_k\| = 0 \tag{3.1}$$

Proof. It is shown in theorem 1 that the descent property (2.3) holds for $\sigma \in \left]0, \frac{1}{2}\right]$, so from (1.4), (2.2), and (2.7) it follows that

$$\left|g_k^T d_{k-1}\right| \leq -\sigma g_{k-1}^T d_{k-1} \leq \sigma \sum_{j=0}^{k-2} \sigma^j = \sum_{j=0}^{k-1} \sigma^j \leq \frac{\sigma}{1-\sigma} \tag{3.2}$$

Thus from the definition of d_k and using (3.2) we deduce that

$$\begin{aligned} \|d_k\|^2 &= \frac{1}{\|g_k\|^2} - \frac{2}{\|g_k\|^2} g_k^T d_{k-1} + \|d_{k-1}\|^2 \\ &\leq \frac{1}{\|g_k\|^2} + \frac{2\sigma}{1-\sigma} \frac{1}{\|g_k\|^2} + \|d_{k-1}\|^2 \\ &= \left(\frac{1+\sigma}{1-\sigma}\right) \frac{1}{\|g_k\|^2} + \|d_{k-1}\|^2 \qquad (3.3) \end{aligned}$$

By applying this relation repeatedly, it follows that

$$\begin{aligned} \|d_k\|^2 &\leq \left(\frac{1+\sigma}{1-\sigma}\right) \sum_{j=2}^{k} \frac{1}{\|g_j\|^2} + \frac{1}{\|g_1\|^2} \\ &\leq \left(\frac{1+\sigma}{1-\sigma}\right) \sum_{j=1}^{k} \frac{1}{\|g_j\|^2} \qquad (3.4) \end{aligned}$$

where we used the facts that

$$\frac{1}{\|g_1\|^2} \leq \left(\frac{1+\sigma}{1-\sigma}\right)\frac{1}{\|g_1\|^2}$$

Now we prove (3.1) by contradiction. It assumes that (3.1) does not hold, then there exists a constant $\varepsilon > 0$ such that

$$\|g_k\| \geq \varepsilon > 0 \tag{3.5}$$

holds for all k sufficiently large. Since g_k is bounded above on the level set L, it follows from (3.4) that

$$\|d_k\|^2 \leq c_1 k \tag{3.6}$$

where c_1 is a positive constant. From (2.2) and (2.7), we have

$$\cos\theta_k = -\frac{g_k^T d_k}{\|g_k\|\,\|d_k\|} \geq \left(2 - \sum_{j=0}^{k-1}\sigma^j\right)\frac{1}{\|g_k\|\,\|d_k\|}$$

$$\geq \left(\frac{1-2\sigma}{1-\sigma}\right)\frac{1}{\|g_k\|\,\|d_k\|} \tag{3.7}$$

Since $\sigma < \frac{1}{2}$, substituting (3.6) and (3.5) into (3.7) gives

$$\sum_k \cos^2\theta_k \geq \left(\frac{1-2\sigma}{1-\sigma}\right)^2 \sum_k \frac{1}{\|g_k\|^2\,\|d_k\|^2} \geq c_2 \sum_k \frac{1}{k} \tag{3.8}$$

where c_2 is a positive constant. Therefore, the series $\sum_k \cos^2\theta_k$ is divergent.

Let M be an upper bound of $\left\|\nabla^2 f(x)\right\|$ on the level set L, then

$$g_{k+1}^T d_k = \left(g_k + a_k \nabla^2 f(x)\right)^T d_k \leq g_k^T d_k + M a_k \|d_k\|^2$$

Thus by using (1.4) we obtain

$$a_k \geq -\frac{(1-\sigma)}{M\,\|d_k\|^2} g_k^T d_k \tag{3.9}$$

Substituting a_k of (3.9) into (1.3) gives

$$f_{k+1} \leq f_k - \frac{(1-\sigma)\,\delta}{M}\left(\frac{g_k^T d_k}{\|d_k\|}\right)^2$$

$$= f_k - c_3\,\|g_k\|^2 \cos^2\theta_k,$$

where $c_3 = \frac{(1-\sigma)\delta}{M} > 0$. Since $f\,(x)$ is bounded below, $\sum_k \|g_k\|^2 \cos^2\theta_k$ converges, which indicates that $\sum_k \cos^2\theta_k$ converges by use of (3.5). This fact contradicts (3.8). We complete the proof. □

4. Numerical results and discussions

In this section we report some numerical results obtained with an implementation of the $CGBB$ algorithm. For our numerical tests, we used test functions and Fortran programs from ([01],[03]). Considering the same criterias as in ([02]), the code is written in Fortran and compiled with f90 on a Workstation Intel Pentium 4 with 2 GHz. We selected a number of 105 unconstrained optimization test functions in generalized or extended form [17] (some from CUTE library [03]). For each test function we have taken twenty (20) numerical experiments with the number of variables increasing as $n = 2, 10, 30, 50, 70, 100,\ 300, 500, 700, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000$. The algorithm implements the Wolfe line search conditions (1.3) and (1.4), and the same stopping criterion $\|\nabla f(x_k)\| < 10^{-6}$. In all the algorithms we considered in this numerical study the maximum number of iterations is limited to 100000.

The comparisons of algorithms are given in the following context. Let f_i^{ALG1} and f_i^{ALG2} be the optimal value found by ALG1 and ALG2, for problem $i = 1, ..., 962$, respectively. We say that, in the particular problem i, the performance of ALG1 was better than the performance of ALG2 if:

$$\left| f_i^{ALG1} - f_i^{ALG2} \right| < 10^{-3}$$

and the number of iterations, or the number of function-gradient evaluations, or the CPU time of ALG1 was less than the number of iterations, or the number of function-gradient evaluations, or the CPU time corresponding to ALG2, respectively.

In a performance profile plot, the top curve corresponds to the method that solved the most problems in a time that was within a factor τ of the best time. The percentage of the test problems for which a method is the fastest is given on the left axis of the plot. The right side of the plot gives the percentage of the test problems that were successfully solved by these algorithms, respectively. Mainly, the right side is a measure of the robustness of an algorithm.

In the set of numerical experiments we compare $CGBB$ algorithm to *Steepest descent algorithm*, $CG\_DESCNET$, PRP and FR conjugate gradient methods.

In Fig. 1, we consider the CPU time to compare the performance of $CGBB$ algorithm to *Steepest descent algorithm* by using profiles of Dolan and Moré ([04]).

Figs. 2 – 4 list the performance of the $CGBB$, $CG\_DESCNET$, PRP and FR conjugate gradient methods. relative to CPU time, the number of iterations and the number of gradient evaluations, respectively, which were evaluated using the profiles of Dolan and Moré.

From fig. 1, when comparing $CGBB$ algorithm with *Steepest descent algorithm* subject to CPU time metric, we see that $CGBB$ algorithm is top performer.

Clearly, Figs. 2 – 4 present that our proposed method $CGBB$ exhibits the best overall performance since it illustrates the highest probability of being the optimal solver, followed by the $CG\_DESCENT$, PRP and FR conjugate gradient methods relative to all performance metrics.

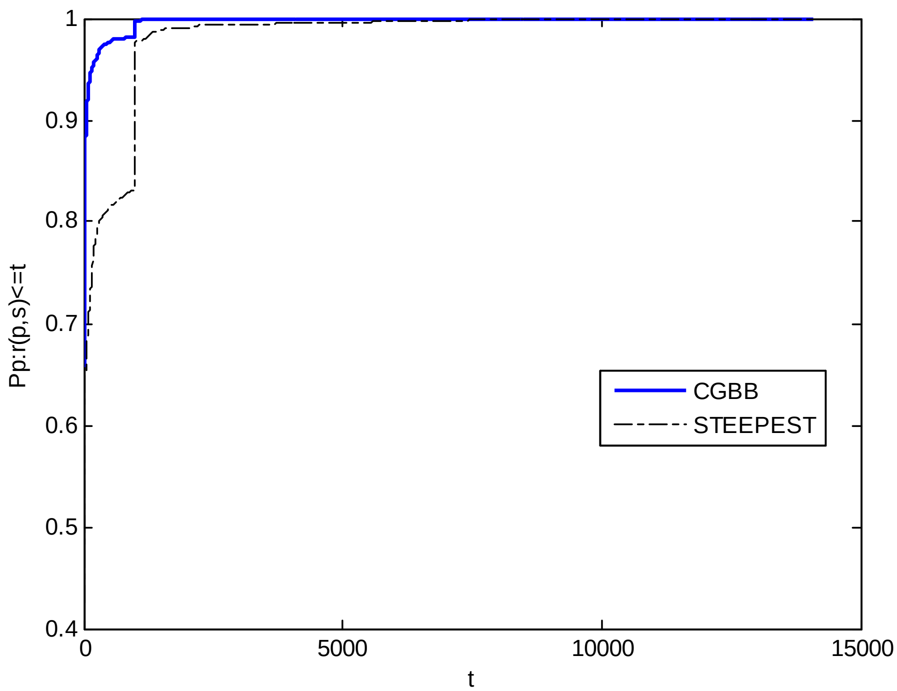

FIGURE 1. Performance profiles of conjugate gradient methods CGBB and STEEPEST based on CPU time.

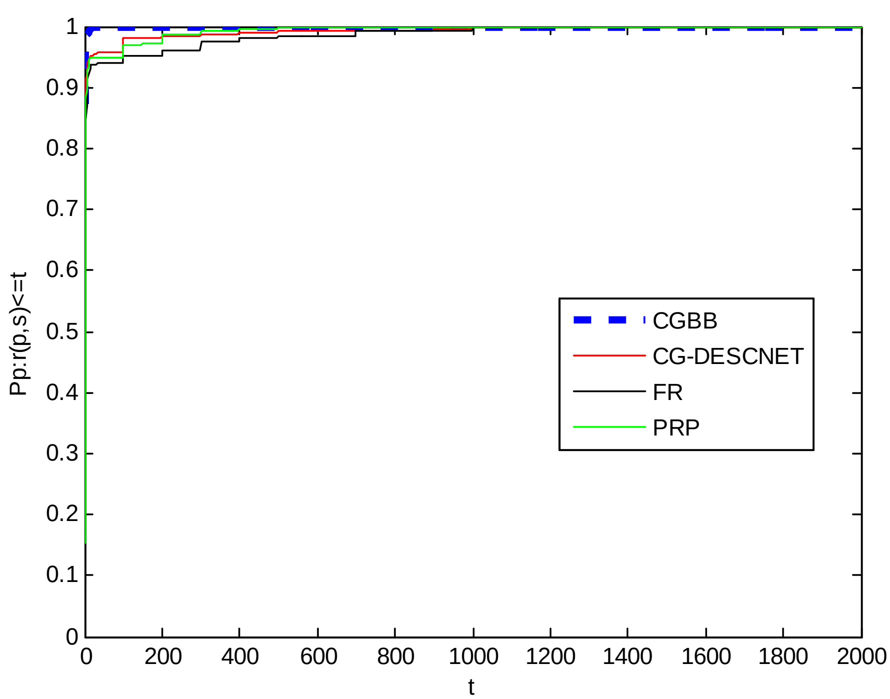

FIGURE 2. Performance based on CPU time.

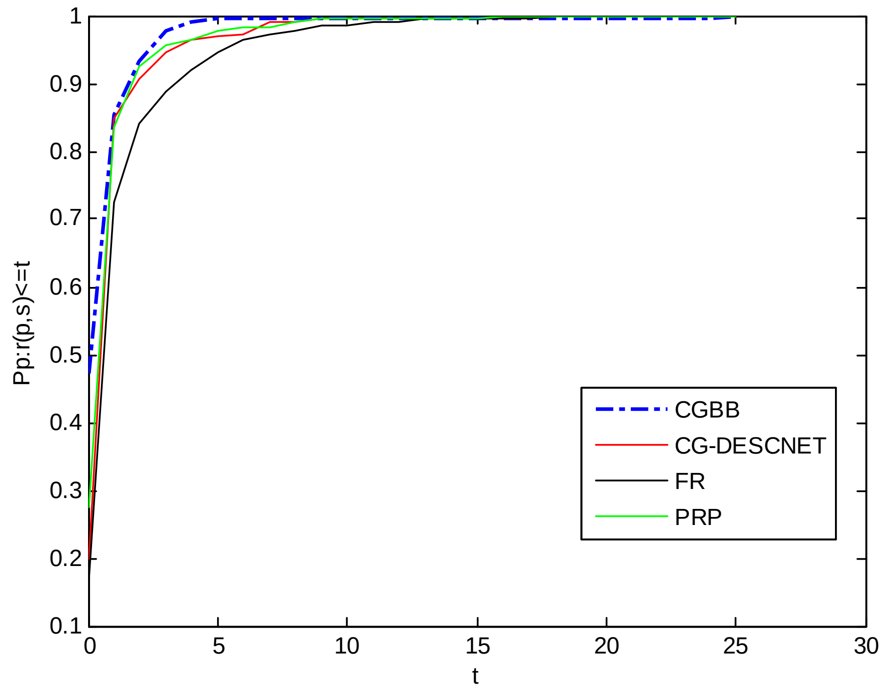

Figure 3. Performance based on the number of iterations.

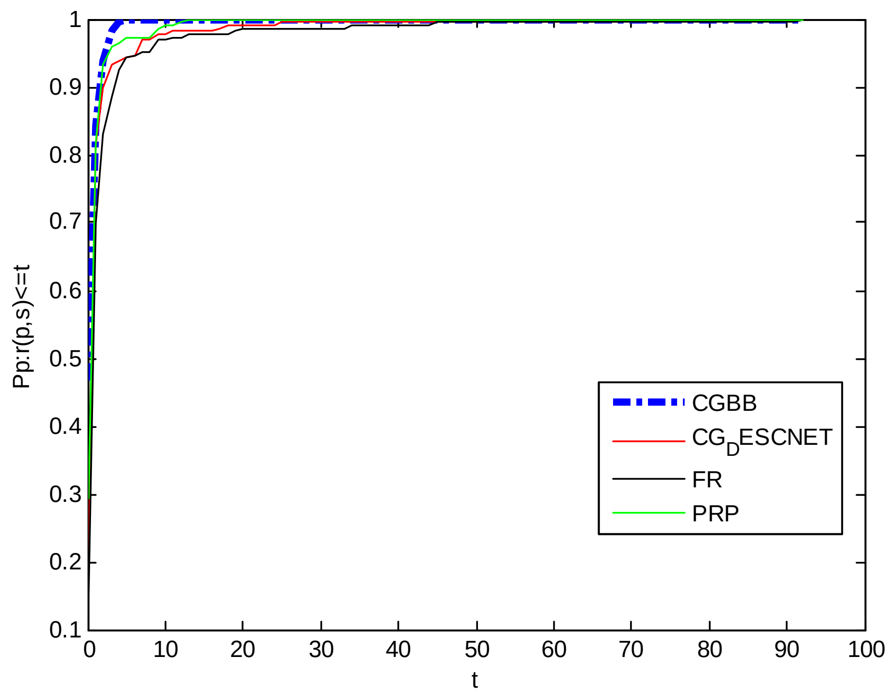

Figure 4. Performance based on the number of gradient evaluations.

5. CONCLUSION

In this paper, we have proposed a new and simple d_k that is easy to implement. We have also provided proof that this method converges globally with strong Wolfe line search.The presented numerical results illustrated the efficiency and robustness of our proposed method.

Our future work is concentrated on studying the convergence properties and numerical performance of our proposed method using different inexact line searches

REFERENCES

[1] N. Andrei, An unconstrained optimization test functions collection, Adv. Modell. Optim. 10 (2008) 147–161.
[2] N. Andrei, Another conjugate gradient algorithm for unconstrained optimization. Annals of Academy of Romanian Scientists, Series on Science and Technology of Information, vol. 1, nr.1, 2008, pp.7-20
[3] I. Bongartz, A. Conn, N. Gould, P. Toint, Cute: constrained and unconstrained testing environments, ACM Transaction on Mathematical Software 21 (1995) 123–160.
[4] E. Dolan, J.J. Moré, Benchmarking optimization software with performance profiles, Mathematical Programming 91 (2002) 201–213.
[5] Y. Dai, Y. Yuan, A nonlinear conjugate gradient with a strong global convergence properties, SIAM J. Optimiz. 10 (2000) 177–182.
[6] R. Fletcher, Practical Method of Optimization, second ed., Unconstrained Optimization, vol. I, Wiley, New York, 1997.
[7] R. Fletcher, C. Reeves, Function minimization by conjugate gradients, Comput. J. 7 (1964) 149–154.
[8] M.R. Hestenes, E. Stiefel, Method of conjugate gradient for solving linear equations, J. Res. Nat. Bur. Stand. 49 (1952) 409–436.
[9] W.W. Hager, H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM Journal on Optimization 16 (2005) 170–192.
[10] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms. Part 1: Theory, J. Optimiz. Theory Appl. 69 (1992) 129–137.
[11] B.T. Polyak, The conjugate gradient method in extreme problems, USSR Comp. Math. Math. Phys. 9 (1969) 94–112.
[12] M.J.D. Powell, Nonconvex minimization calculations and the conjugate gradient method, Lecture Notes in Mathematics, 1066, Springer-Verlag, Berlin, 1984, pp. 122–141.
[13] E. Polak, G. Ribière, Note Sur la convergence de directions conjugèes, Rev. Francaise Informat Recherche Operationelle 3e Annèe 16(1969) 35–43.
[14] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM Journal of Optimization, 7 (1997) 26–33.
[15] J. Schropp, A note on minimization problems and multistep methods. Numeric Mathematic, 78 (1997) 87–101.
[16] J. Schropp, One-step and multistep procedures for constrained minimization problems. IMA Journal of Numerical Analysis, 20 (2000) 135–152.
[17] P.Wolfe, Convergence conditions for ascent méthods. *Siam Review* ,11,pp.226-235 ,(1969) .
[18] G. Zoutendijk, Nonlinear programming computational methods, in: J. Abadie (Ed.), Integer and Nonlinear Programming, North-Holland, Amsterdam, 1970, pp. 37–86.

LABORATORY INFORMATICS AND MATHEMATICS (LIM), FACULTY OF SCIENCE AND TECHNOLOGY, DEPARTMENT OF MATHEMATICS AND INFORMATICS, MOHAMED CHERIF MESSAADIA UNIVERSITY, SOUK AHRAS, ALGERIA.
E-mail address: belloufi_mohammed@yahoo.fr

LANOS, BADJI MOKHTAR UNIVERSITY, ANNABA, ALGERIA.
E-mail address: rabenzine@yahoo.fr