Bol. Soc. Paran. Mat. (3s.) **v. 38** 6 (2020): 127–140.
©SPM –ISSN-2175-1188 on line ISSN-00378712 in press
SPM: www.spm.uem.br/bspm doi:10.5269/bspm.v38i6.35641

Convergence of a Two-parameter Family of Conjugate Gradient Methods with a Fixed Formula of Stepsize

KHELIFA Bouaziz and YAMINA Laskri

abstract: We prove the global convergence of a two-parameter family of conjugate gradient methods that use a new and different formula of stepsize from Wu [14]. Numerical results are presented to confirm the effectiveness of the proposed stepsizes by comparing with the stepsizes suggested by Sun and his colleagues [2,12].

Key Words: Unconstrained optimization, Conjugate gradient methods, Global convergence, Stepsize, Line search.

Contents

1. Introduction

Let us consider the following unconstrained minimization problem:

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

where f is a differentiable objective function, has the following form

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

where

$$d_k = \begin{cases} -g_k & \text{for } k = 1, \\ -g_k + \beta_k^{\mu_k,\omega_k} d_{k-1} & \text{for } k \geq 2, \end{cases} \tag{1.3}$$

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k with the following formulas

$$\beta_k^{\mu_k,\omega_k} = \frac{g_k^T y_{k-1}}{D_k}, \tag{1.4}$$

where

$$D_k = (1 - \mu_k - \omega_k) \|g_{k-1}\|^2 + \mu_k d_{k-1}^T y_{k-1} - \omega_k d_{k-1}^T g_{k-1}, \tag{1.5}$$

2010 *Mathematics Subject Classification:* 35B40, 35L70.

Submitted February 26, 2017. Published January 16, 2018

where $\|.\|$ means the Euclidean norm, $y_{k-1} = g_k - g_{k-1}$, and D_k depends on parameters $\mu_k \in [0, 1[$ and $\omega_k \in [0, 1 - \mu_k[$. Let us remark that the descent direction d_k is defined such that

$$g_k^T d_k = -c \|g_k\|^2 , \tag{1.6}$$

where $0 < c < 1 - \mu_k - \omega_k$.
The parametrized expression (1.4) is taken from [2]. It only covers a subset of a larger family introduced by Dai and Yan [3]. Three classical versions of nonlinear CG are particular cases of formula (1.4):

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \text{Hestenes-Stiefel [7]}$$

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \qquad \text{Polak-Ribière-Polyak [10]}$$

$$\beta_k^{LS} = \frac{-g_k^T y_{k-1}}{d_{k-1}^T g_{k-1}}. \qquad \text{Liu-Storey [8]}$$

Other important cases are not covered by the present study, such as the Fletcher-Reeves method [6], the Conjugate Descent method [5], and the Dai-Yuan method [4]. In addition to studying the convergence of method in other conditions by Sellami and all [11]. The efficiency of the conjugate gradient method depends majorly on the stepsize. Line search technique has been used in various literatures to obtain the stepsize. A very recent development is to obtain the stepsize with a unified formula which is referred to as stepsize without line search. In the implementation of any conjugate gradient (CG) method, the stepsize is often determined by certain line search conditions such as the Wolfe conditions [13]. These types of line search involve extensive computation of function values and gradients, which often becomes a significant burden for large-scale problems, which spurred Sun and Zhang [12] to pursue the conjugate gradient method where they calculated the stepsize instead of the line search according to the following formula

$$\alpha_k = -\delta g_k^T d_k / \|d_k\|_{Q_k}^2 , \tag{1.7}$$

where $\|d_k\|_{Q_k} = \sqrt{d_k^T Q_k d_k}$, $\delta \in (0, \nu_{min}/\tau)$, τ is a Lipschitz constant of f, and $\{Q_k\}$ is a sequence of positive definite matrices satisfying for positive constants ν_{min} and ν_{max} that

$$\nu_{min} d^T d \leq d^T Q_k d \leq \nu_{max} d^T d, \ \forall k, \ \forall d \in R^n.$$

But the formula for the stepsize above involve a positive matrix. For large scale optimization problems, this may cost additional memory space and execution time during the computations.
The aim of the paper is to employ a formula for α_k without a matrix which uses both available function value and gradient information. Under suitable assumptions,

we prove the global convergence of a two-parameter family of conjugate gradient methods.
Lately Wu [14] succeeded to obtain the derive formula from the stepsize, this formula is

$$\alpha_k = \frac{-\delta g_k^T d_k}{(\bar{g}_{k+1} - g_k)^T d_k + \gamma\theta_k}, \tag{1.8}$$

where

$$\theta_k = 6(f_k - \bar{f}_{k+1}) + 3(g_k + \bar{g}_{k+1})^T d_k,$$

f_k, g_k, $\bar{f}_{k+1}$ and $\bar{g}_{k+1}$ denote $f(x_k)$, $\bigtriangledown f(x_k)$, $f(x_k + d_k)$, and $\bigtriangledown f(x_k + d_k)$, respectively, δ and γ are parameters satisfying

$$\delta \in (0, \kappa/\tau), \tag{1.9}$$

and

$$\gamma \geq 0 \text{ if } \tau = \kappa, \quad \text{or } \gamma \in (0, \frac{\kappa - \delta\tau}{3(\tau - \kappa)}) \text{ if } \tau > \kappa, \tag{1.10}$$

κ and τ are defined in Assumption 2.1 below.
He proved that the above formula for α_k can ensure global convergence for CD, FR and PR methods.
In this paper, our goal is to employ the step-formula (1.8) to prove the Two-parameter family of conjugate gradient method, which was expounded by Chen and Sun [2] using the formula (1.7).
This paper is organized as follows. Some preliminary results on the family of CG methods with the fixed-form stepsize formula (1.8) are given in Section 2. Section 3 includes the main convergence properties of the two-parameter family of conjugate gradient methods without line search. Numerical experiments and discussions are given in Section 4.

2. Properties of the stepsize

The present section gathers technical results concerning the stepsize α_k generated by (1.8), which will be useful to derive the global convergence properties of the next section.
Assumption 2.1 *The function f is LC^1 and strongly convex in R^n, i.e, there exists constants $\tau > 0$ and $\kappa \geq 0$ such that*

$$\|\bigtriangledown f(u) - \bigtriangledown f(v)\| \leq \tau \|u - v\|, \forall u, v \in R^n, \tag{2.1}$$

and

$$[\bigtriangledown f(u) - \bigtriangledown f(v)]^T (u - v) \geq \kappa \|u - v\|^2, \forall u, v \in R^n, \tag{2.2}$$

or equivalently,

$$f(u) - f(v) \geq \bigtriangledown f(v)^T (u - v) + \frac{\kappa}{2} \|u - v\|^2, \forall u, v \in R^n. \tag{2.3}$$

Note that Assumption 2.1 implies that the level set

$$L = \{x \in R^n | f(x) \leq f(x_1)\} \text{ is bounded.}$$

Lemma 2.2 *Suppose that x_k is given by (1.2), (1.3) and (1.8). Then*

$$g_{k+1}^T d_k = \rho_k g_k^T d_k, \tag{2.4}$$

holds for all k, where

$$0 < \rho_k = 1 - \delta \Phi_k \left\| d_k \right\|^2 / [(\bar{g}_{k+1} - g_k)^T d_k + \gamma \theta_k], \tag{2.5}$$

and

$$\Phi_k = \begin{cases} 0 & \text{for } \alpha_k = 0, \\ (g_{k+1} - g_k)^T (x_{k+1} - x_k) / \left\| x_{k+1} - x_k \right\|^2 & \text{for } \alpha_k \neq 0. \end{cases} \tag{2.6}$$

Proof [14] Lemma 2.

Lemma 2.3 *Suppose that Assumption 2.1 holds. Then the following inequalities*

$$\kappa \leq \Phi_k \leq \tau, \tag{2.7}$$

and

$$[\kappa + 3\gamma(\kappa - \tau)] \left\| d_k \right\|^2 \leq (\bar{g}_{k+1} - g_k)^T d_k + \gamma \theta_k \leq (1 + 3\gamma)\tau \left\| d_k \right\|^2, \tag{2.8}$$

hold for all k.

Proof The lemma can be proved by the same way of the proof of lemma 1 in [14].

Corollary 2.4 *Suppose that Assumption 2.1 holds. Then*

$$\frac{\delta \kappa}{(1 + 3\gamma)\tau} \leq 1 - \rho_k \leq \frac{\delta \tau}{\kappa + 3\gamma(\kappa - \tau)}, \tag{2.9}$$

holds for all k.

Lemma 2.5 *Suppose that Assumption 2.1 holds and $\{x_k\}$ is generated by (1.2), (1.3) and (1.8). Then*

$$\sum_{d_k \neq 0} \frac{(g_k^T d_k)^2}{\left\| d_k \right\|^2} < \infty. \tag{2.10}$$

Proof [14] Lemma 3.

Lemma 2.6 *Suppose that Assumption 2.1 holds, then we have*

$$\sum_k \alpha_k^2 \left\| d_k \right\|^2 < \infty. \tag{2.11}$$

Proof By (1.8), (2.8) and (2.10) we have

$$\sum_k \alpha_k^2 \left\| d_k \right\|^2 = \sum_k \frac{(\delta g_k^T d_k)^2}{[(\bar{g}_{k+1} - g_k)^T d_k + \gamma \theta_k]^2} \left\| d_k \right\|^2 \tag{2.12}$$

$$\leq [\frac{\delta}{\kappa + 3\gamma(\kappa - \tau)}]^2 \sum_{d_k \neq 0} \frac{(g_k^T d_k)^2}{\left\| d_k \right\|^2} \tag{2.13}$$

$$< \infty.$$

3. Global convergence of the two-parameter family

In this section, we discuss the convergence properties of a two-parameter family of conjugate gradient methods, in which $\beta_k^{\mu_k,\omega_k}$ is given by (1.4) and (1.5).
We give the following algorithm firstly.
Algorithm 3.1 Step 0: Given $x_1 \in R^n$, set $d_1 = -g_1$, $k = 1$.
Step 1: If $\|g_k\| = 0$ then stop else go to Step 2.
Step 2: Set $x_{k+1} = x_k + \alpha_k d_k$ where d_k is defined by (1.3), and α_k is defined by (1.8).
Step 3: Compute $\beta_{k+1}^{\mu_{k+1},\omega_{k+1}}$ using formula (1.4).
Step 4: Set $k := k + 1$, go to Step 1.
Lemma 3.2 *Under Assumption 2.1, the method defined by (1.2), (1.3), (1.8) and (1.4) will generate a sequence $\{x_k\}$ such that $f(x_{k+1}) \leq f(x_k)$.*
Proof The lemma can be proved by the same way of the proof of lemma 3 in [14], with slightly modification.
Lemma 3.3 *Suppose that Assumption 2.1 holds. Then.*

$$D_k \geq (1 - \mu_k - \omega_k)\|g_{k-1}\|^2 - (\omega_k + \frac{\delta\kappa\mu_k}{(1+3\gamma)\tau})d_{k-1}^T g_{k-1} \geq 0. \tag{3.1}$$

Proof Since $y_{k-1} = g_k - g_{k-1}$, (2.4) also reads

$$\begin{aligned} D_k &= (1 - \mu_k - \omega_k)\|g_{k-1}\|^2 + (\mu_k\rho_{k-1} - \mu_k - \omega_k)d_{k-1}^T g_{k-1} \\ &= (1 - \mu_k - \omega_k)\|g_{k-1}\|^2 - (\mu_k(1-\rho_{k-1}) + \omega_k)d_{k-1}^T g_{k-1}. \end{aligned}$$

According to (2.9) and $d_{k-1}^T g_{k-1} \leq 0$ the conclusion is immediate □
Remark 3.4 *Let us suppose first that Assumption 2.1 is valid. If D_k cancels, then (3.1) implies*

$$(1 - \mu_k - \omega_k)\|g_{k-1}\|^2 - (\omega_k + \frac{\delta\kappa\mu_k}{(1+3\gamma)\tau})d_{k-1}^T g_{k-1} = 0.$$

Since the left-hand side is the sum of two nonnegative terms, we obtain

$$\begin{cases} (1 - \mu_k - \omega_k)\|g_{k-1}\|^2 = 0 & (a), \\ (\omega_k + \dfrac{\delta\kappa\mu_k}{(1+3\gamma)\tau})d_{k-1}^T g_{k-1} = 0 & (b), \end{cases}$$

- Case 1: If $\mu_k + \omega_k < 1$, (a) boils down to $\|g_{k-1}\|^2 = 0$, which means that convergence is reached at iteration $k-1$.
- Case 2: If $\mu_k = 0$: then D_k is the sum of two nonnegative terms, so $D_k = 0$ implies that both cancel:

$$\begin{cases} (1 - \omega_k)\|g_{k-1}\|^2 = 0, \\ \omega_k d_{k-1}^T g_{k-1} = 0. \end{cases}$$

If $\omega_k < 1$, the conclusion is the same as in Case 1.
Lemma 3.5 *Under Assumption 2.1 we have*

$$\lim_{k\longrightarrow\infty} \inf \|g_k\| > 0 \Longrightarrow \lim_{k\longrightarrow\infty} \beta_k^{\mu_k,\omega_k} = 0,$$

where $\beta_k^{\mu_k,\omega_k}$ is defined by (1.4).
Proof According to (2.11), we conclude that

$$\lim_{k\longrightarrow\infty} \|x_{k+1} - x_k\|^2 = \lim_{k\longrightarrow\infty} \|\alpha_k d_k\|^2 = 0.$$

Because f is continuously differentiable, $\|g_k\|$ is bounded according to Assumption 2.1 and the boundedness of L, we have also $\lim_{k\longrightarrow\infty} y_{k-1} = 0$ and

$$\lim_{k\longrightarrow\infty} g_k^T y_{k-1} = 0. \tag{3.2}$$

If $\liminf_{k\longrightarrow\infty} \|g_k\| > 0$, there exists a positive constant $\psi > 0$ such that

$$\|g_k\| \geq \psi \quad \text{for all } k. \tag{3.3}$$

According to (1.4), we have

$$\mid g_k^T y_{k-1} \mid = \mid \beta_k^{\mu_k,\omega_k} \mid\mid D_k \mid . \tag{3.4}$$

Let us consider the iteration indices k such that $\mu_k + \omega_k \in [0, 1/2]$. According to (3.1) and $d_{k-1}^T g_{k-1} \leq 0$, (3.4) implies that

$$\mid g_k^T y_{k-1} \mid \geq \mid \beta_k^{\mu_k,\omega_k} \mid (1 - \mu_k - \omega_k) \|g_{k-1}\|^2 ,$$

which leads to

$$\mid g_k^T y_{k-1} \mid \geq \mid \beta_k^{\mu_k,\omega_k} \mid \psi^2/2, \tag{3.5}$$

given (3.3).
Let us establish a similar result in the more complex case $\mu_k + \omega_k \in (1/2, 1[$. As a preliminary step, let us show that

$$g_k^T d_k \leq -\psi^2/2, \tag{3.6}$$

for all sufficiently large values of k.
In the case $g_{k-1}^T d_{k-1} = 0$, we have $\beta_k^{\mu_k,\omega_k} = 0$, so $d_k = -g_k$ and (3.6) is valid according to (3.3).
Now let us consider the case where $g_{k-1}^T d_{k-1} < 0$. Given (1.3), (1.4) and (2.4) we have

$$\begin{aligned} g_k^T d_k &= g_k^T(-g_k + \beta_k^{\mu_k,\omega_k} d_{k-1}) \\ &= -\|g_k\|^2 + \frac{g_k^T y_{k-1}}{(1-\mu_k-\omega_k)\|g_{k-1}\|^2 + \mu_k d_{k-1}^T y_{k-1} - \omega_k d_{k-1}^T g_{k-1}} g_k^T d_{k-1}, \\ &= -\|g_k\|^2 + \frac{\rho_{k-1} g_k^T y_{k-1}}{(1-\mu_k-\omega_k)\|g_{k-1}\|^2 + (\mu_k\rho_{k-1} - \mu_k - \omega_k) d_{k-1}^T g_{k-1}} g_{k-1}^T d_{k-1} \\ &\leq -\psi^2 + \frac{|\rho_{k-1} g_k^T y_{k-1}|}{|\mu_k\rho_{k-1} - \mu_k - \omega_k|}, \end{aligned} \tag{3.7}$$

thus for sufficiently large k, (3.6) is also true.
From (1.4) and (3.6) we have

$$|g_k^T y_{k-1}| \geq |\beta_k^{\mu_k,\omega_k}|[(1-\mu_k-\omega_k)\psi^2 + (\mu_k(1-\rho_{k-1})+\omega_k)\psi^2/2].$$

From (2.9), we have

$$\begin{aligned} |g_k^T y_{k-1}| &\geq |\beta_k^{\mu_k,\omega_k}|\{(1-\mu_k-\omega_k)\psi^2 + [\frac{\delta\kappa\mu_k}{(1+3\gamma)\tau}+\omega_k]\psi^2/2\}. \\ &= |\beta_k^{\mu_k,\omega_k}|[1-\omega_k/2-(1-\frac{\delta\kappa}{2(1+3\gamma)\tau})\mu_k]\psi^2, \end{aligned} \tag{3.8}$$

for all sufficiently large values of k. Given $\mu_k+\omega_k \in (1/2, 1[$, the latter inequality implies

$$| g_k^T y_{k-1} | \geq | \beta_k^{\mu_k,\omega_k} | S\psi^2. \tag{3.9}$$

Since $S \leq 1/2$, (3.9) is implied by (3.5), so that (3.9) holds in the whole domain $\mu_k \in [0, 1[$, $\omega_k \in [0, 1-\mu_k[$.
Finally, (3.2) and (3.9) jointly imply $\lim_{k\longrightarrow\infty} |\beta_k^{\mu_k,\omega_k}| = 0$.
On the other hand, consider the case where Assumption 2.1 is not necessarily valid. If $\mu_k = 0$, then we have

$$| g_k^T y_{k-1} | \geq | \beta_k^{0,\omega_k} | \psi^2/2.$$

The proof is similar to that of (3.9), where the two cases to examine are $\omega_k \in [0, 1/2]$ and $\omega_k \in (1/2, 1[$.
Finally, according (3.2) we have $\lim_{k\longrightarrow\infty} |\beta_k^{0,\omega_k}| = 0$. □
Theorem 3.6 *Under Assumption 2.1, the method defined by (1.2), (1.3), (1.8) and (1.4) will generate a sequence $\{x_k\}$ such that*

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

Proof Suppose on the contrary that $\|g_k\| \geq \psi$ for all k.
Since $\liminf_{k\longrightarrow\infty} \|g_k\| \neq 0$, by lemma 3.5 we have $\beta_k^{\mu_k,\omega_k} \longrightarrow 0$, as $k \longrightarrow 0$. Since L is bounded, both $\{x_k\}$ and $\{g_k\}$ are bounded. By using

$$\|d_k\| \leq \|g_k\| + |\beta_k^{\mu_k,\omega_k}| \, \|d_{k-1}\|, \tag{3.10}$$

one can show that $\{\|d_k\|\}$ is uniformly bounded. Definition (1.3) implies the following relation

$$\begin{aligned} |g_k^T d_k| &= |g_k^T(-g_k + \beta_k^{\mu_k,\omega_k} d_{k-1})| & (3.11) \\ &\geq \|g_k\|^2 - |\beta_k^{\mu_k,\omega_k}| \, \|g_k\| \, \|d_{k-1}\|. & (3.12) \end{aligned}$$

From (1.4), (2.4) and using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} |\beta_k^{\mu_k,\omega_k}| &= |\frac{g_k^T(g_k-g_{k-1})}{(1-\mu_k-\omega_k)\|g_{k-1}\|^2 + \mu_k d_{k-1}^T y_{k-1} - \omega_k d_{k-1}^T g_{k-1}}| \\ &\leq \frac{\|g_k\| \, \|g_k-g_{k-1}\|}{|(1-\mu_k-\omega_k)\|g_{k-1}\|^2 - [(1-\rho_{k-1})\mu_k+\omega_k]d_{k-1}^T g_{k-1}|}. \end{aligned} \tag{3.13}$$

From (2.1) and (2.13) we have

$$\begin{aligned} \|g_k - g_{k-1}\| &\le \tau\alpha_{k-1}\|d_{k-1}\| \\ &\le \left(\frac{\tau\delta}{\kappa + 3\gamma(\kappa-\tau)}\right)\frac{|g_{k-1}^T d_{k-1}|}{\|d_{k-1}\|}. \end{aligned} \tag{3.14}$$

According (1.9) and (1.10) we deduce that

$$\|g_k - g_{k-1}\| \le \frac{|g_{k-1}^T d_{k-1}|}{\|d_{k-1}\|}. \tag{3.15}$$

From (1.6) we have

$$\begin{aligned} |(1-\mu_k-\omega_k)\|g_{k-1}\|^2 &- [(1-\rho_{k-1})\mu_k + \omega_k]d_{k-1}^T g_{k-1}| \\ &\ge \frac{|-1+\mu_k+\omega_k|}{c}|g_{k-1}^T d_{k-1}| \\ &= m|g_{k-1}^T d_{k-1}|, (m \ge 1). \end{aligned} \tag{3.16}$$

By (3.13), (3.15), and (3.16) we have

$$|\beta_k^{\mu_k,\omega_k}|\,\|d_{k-1}\| \le \frac{\|g_k\|}{m}. \tag{3.17}$$

Hence by substituting (3.17) in (3.12), we have

$$|g_k^T d_k| \ge A\|g_k\|^2, A = \frac{m-1}{m}, \tag{3.18}$$

for large k. Thus we have

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2\|g_k\|^2} \ge A^2\frac{\|g_k\|^2}{\|d_k\|^2}. \tag{3.19}$$

Since $\|g_k\| \ge \psi$ and $\|d_k\|$ is bounded above, we conclude that there is $\varepsilon > 0$ such that

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2\|g_k\|^2} \ge \varepsilon,$$

which implies

$$\sum_{d_k\neq 0}\frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty. \tag{3.20}$$

This is a contradiction to lemma 2.5. □

4. Numerical experiments and discussions

In this part, we present the numerical experiments of the new formula (1.8) and apply it using (1.4), computer (Processor: Intel(R)core(TM)i3-3110M cpu@2.40 GHZ, Ram 4.00 GB) through the Matlab programme.
20 testing problems have been taken from [1]. This will lead us to test for the global convergence properties of our method. Stopping criteria is set to $\|g_k\| \leq \varepsilon$ where $\varepsilon = 10^{-6}$. Taking into consideration the following parameters: $\delta = 0.75$, $\gamma = 0.01$, $\mu_k = \mu = 0.5$ and $\omega_k = \omega = 0.4$.

Table 1 list numerical results. The meaning of each column is as follows :
"Problem "the name of the test problem.
"N "the dimension of the test problem.
"k "the number of iterations.
"$\|g_k\|$"the norm of the gradient.

On conducting the numerical experiments we reached the following conclusions: The results of the Table 1 indicate that the expression (1.8) provides time and memory better than the expression (1.7) that uses the matrices $\{Q_k\}$ that may become a burden at some times to show the converge.
The value of δ is set too large in the experience of the line search method may generate x_{k+1} such that $f(x_k) < f(x_{k+1})$ and this should not occur theoretically, therefore δ value should be sufficiently small, see lemma 3.2. The more the value δ is small the bigger the iterations are, this fact allows for the convergence to occur in the end.
If the value γ is small, the number of iterations diminishes and $\|g_k\|$ converges rapidly and that is because it is bound to the parameters τ and κ. Furthermore, the selection of the parameters values μ and ω determines the value of the $\|g_k\|$ and the number of iterations.

There is a number of 20 large-scale unconstrained optimization test problems in generalized or extended from CUTE [1] collection. For each test function we have taken six numerical experiments with the number of variables increasing as n = 1000, 2000, 4000, 6000, 8000, 10000.
We adopt the performance profiles by Delan and Moré [9] to compare the performance between the following tow conjugate gradient algorithms
T-PF1: two-parameter family of conjugate gradient methods by using a formula of (1.7).
T-PF2: two-parameter family of conjugate gradient methods by using a formula of (1.8).
Figure 1,2,3 and 4 give performance profiles of the two methods for the number of iterations, CPU time, function evaluations, and gradient evaluations respectively.
From the above three figures, we can see that all the methods are efficient. The new method T-PF2 performs better than the T-PF1 method, for the given test problems. These obtained preliminary results are indeed encouraging.

Table 1

	Problem	N	k	$\|g_k\|$
1	Raydan 1	10	19	1.120790293916461e-07
		50	51	1.389923874174295e-07
2	Diagonal 5	500	342	1.026183259422179e-07
		1000	631	1.003760103090495e-07
3	Perturbed quadratic digonal	100	23	2.170295100224989e-07
		500	29	1.155266677665186e-07
4	Extended quadratic penalty 1	100	18	1.884874483160104e-07
		500	19	1.016400612612853e-07
5	QUARTC	100	15	1.643581275426157e-07
		500	16	1.550430128841257e-07
6	Diagonal 7	100	15	1.279002965759446e-07
		500	16	1.099362564544676e-07
7	Extended Maratos	100	173	1.594462458701618e-07
		500	286	1.024825161025436e-07
8	Diagonal 8	50	20	3.247344971185945e-07
		200	21	1.653875889473304e-07
9	Freudenstein and Roth	100	21	2.457891545157802e-07
		500	24	1.623516986314701e-07
10	Rosenbrock	100	19	1.002347534497171e-07
		500	21	1.001365348958891e-07
11	White and Holst	100	24	2.070125493427497e-07
		500	27	1.810192036538758e-07
12	Beale	100	145	1.244578200376349e-07
		500	221	1.067324585387475e-07
13	Penalty	100	71	1.532487962140017e-07
		500	89	1.360078512369044e-07
14	Cliff	100	34	2.005665322082197e-07
		500	39	1.897012587633221e-07
15	LIARWHD	100	248	1.185263304998455e-07
		500	409	1.058770638716538e-07
16	Almost Perturbed Quadratic	2	21	1.148131830118244e-07
		500	29	1.017343210015877e-07
17	Diagonal 4	2	30	1.319452728458628e-07
		500	31	1.001055919601577e-07
18	Staircase 1	2	21	2.292660111210135e-07
		500	25	1.025016344917332e-07
19	Power	2	21	2.251830705880973e-07
		500	27	1.095143872286877e-07
20	Full Hessian 2	2	27	1.516469135063661e-07
		500	29	1.100214389422758e-07

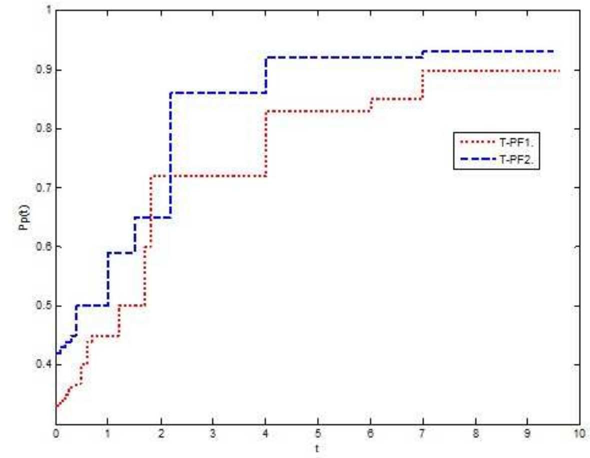

Figure 1: Performance files based on Iterations

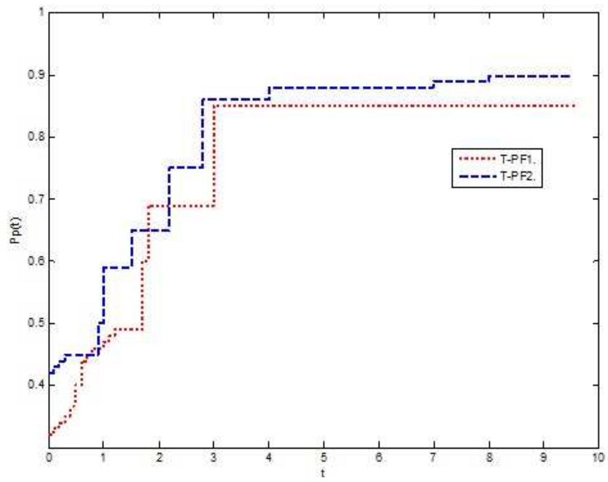

Figure 2: Performance files based on CPU Time

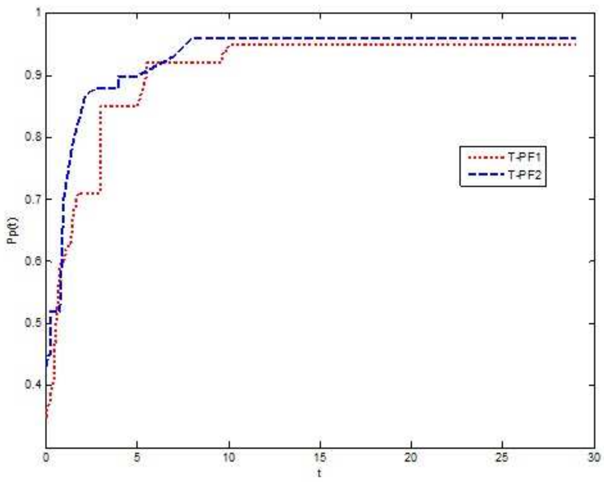

Figure 3: Performance files based on Function Evaluations

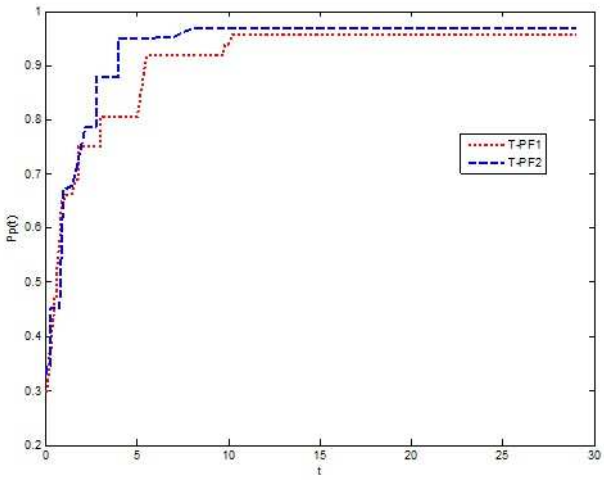

Figure 4: Performance files based on Gradient Evaluations

In the performance profile plot, the top curve corresponds to the method that solved the most problems in time that is within a factor t of the best time. The percentage of the test problems for which a method is reported as 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 each of the methods. In essence, the right side is a measure of the algorithm's robustness.
The performance results of the number of iterations and CPU time are shown in Figs. 1 and 2, respectively, to compare the performance based on the CPU time between the T-PF1 method in (1.4) and T-PF2 method. That is, for each method. In particular, T-PF2 is fastest for about 83%. Also, it is interesting to observe in Figure 1 that the T-PF2 codes are the top performer, relative to the iteration metric, for values of $t \geq 6.5$. Figure 2 indicates that, relative to the CPU time metric, T-PF2 is faster (for 79%), than T-PF1. Hence, T-PF2 code is the top performers, relative to the CPU time metric, for values of $t \geq 3$.
In Figure 3, we compare performance based on the number of function evaluations. Since the top curve in Figure 3 corresponds to T-PF2, this algorithm is clearly the fastest for this set of 20 test problems. Also, it is interesting to observe in Figure 3 that the T-PF2 code is the top performers, relative to the number of function evaluations, for values of $t \geq 6$.
Finally in Figure 4, we compare performance based on the number of gradient evaluations. T-PF2 is faster (for 72%), than T-PF1. Hence, T-PF2 code is the top performers, relative to the the number of gradient evaluations, for values of $t \geq 4.5$.
In conclusion, Figs. 1 - 4 suggest that our proposed method T-PF2 exhibits the best overall performance since it illustrates the highest probability of being the optimal solver, followed by the T-PF1 conjugate gradient method relative to all performance metrics.
Acknowledgements. The authors would like to thank the referees and the editors for their careful reading and some useful comments on improving the presentation of this paper.

References

1. N. Andrei, *An unconstrained optimization test functions collection*, Adv. Model. Optim, 10(1), 147–161, (2008).
2. X. Chen and J. Sun, *Global convergence of a two-parameter family of conjugate gradient methods without line search*, Journal of Computational and Applied Mathematics, 146(1),37–45, (2002).
3. Y. Dai and Y. Yuan, *A three-parameter family of nonlinear conjugate gradient methods*, Mathematics of Computation, 70(235), 1155–1167, (2001).
4. Y.H. Dai and Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM Journal on Optimization, 10(1), 177–182, (1999).
5. R. Fletcher, *Practical methods of optimization*, John Wiley & Sons, (1987).
6. R. Fletcher and C. M. Reeves, *Function minimization by conjugate gradients*, The computer journal, 7(2), 149–154, (1964).
7. M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, volume 49. NBS, (1952).

8. Y. Liu and C. Storey, *Efficient generalized conjugate gradient algorithms*, part 1: theory, Journal of Optimization Theory and Applications, 69(1), 129–137, (1991).

9. J. Moré and E. D. Dolan, *Benchmarking optimization software with performance files*, Math. Program, 91(2), 2001-2013, (2002).

10. B. T. Polyak, *The conjugate gradient method in extremal problems*, USSR Computational Mathematics and Mathematical Physics, 9(4), 94–112, (1969).

11. B. Sellami Y. Laskri and R. Benzine, *A new two-parameter family of nonlinear conjugate gradient methods*, Optimization, 64 (4) , 993–1009, (2015).

12. J. Sun and J. Zhang, *Global convergence of conjugate gradient methods without line search*, Annals of Operations Research, 103(1-4), 161–173, (2001).

13. P. Wolfe, *Convergence conditions for ascent methods*, ii: Some corrections. SIAM review, 13(2), 185–188, (1971).

14. Q.-j. Wu, *A nonlinear conjugate gradient method without line search and its global convergence*, In Computational and Information Sciences (ICCIS), 2011 International Conference on. IEEE, 1148–1152, (2011).

Khelifa BOUAZIZ,
Y. Laskri,
Department of mathematics
Badji mokhtar University
Annaba, Algeria.
E-mail address: `Elkhalifamaths@gmail.com`
E-mail address: `Yamina.laskri@univ-annaba.org`