Hindawi
Journal of Mathematics
Volume 2017, Article ID 2715854, 12 pages
https://doi.org/10.1155/2017/2715854

*Research Article*

# A New Modified Three-Term Conjugate Gradient Method with Sufficient Descent Property and Its Global Convergence

**Bakhtawar Baluch,[1] Zabidin Salleh,[1] Ahmad Alhawarat,[2] and U. A. M. Roslan[1]**

[1] *School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia*
[2] *Department of Mathematics, College of Science, Al-Isra University, Amman, Jordan*

Correspondence should be addressed to Zabidin Salleh; zabidin@umt.edu.my

Received 4 April 2017; Revised 30 June 2017; Accepted 3 August 2017; Published 13 September 2017

Academic Editor: Liwei Zhang



A new modified three-term conjugate gradient (CG) method is shown for solving the large scale optimization problems. The idea relates to the famous Polak-Ribière-Polyak (PRP) formula. As the numerator of PRP plays a vital role in numerical result and not having the jamming issue, PRP method is not globally convergent. So, for the new three-term CG method, the idea is to use the PRP numerator and combine it with any good CG formula's denominator that performs well. The new modification of three-term CG method possesses the sufficient descent condition independent of any line search. The novelty is that by using the Wolfe Powell line search the new modification possesses global convergence properties with convex and nonconvex functions. Numerical computation with the Wolfe Powell line search by using the standard test function of optimization shows the efficiency and robustness of the new modification.

## 1. Introduction

The conjugate gradient method is an efficient and organized tool for solving the large-scale nonlinear optimization problem, due to its simplicity, easiness, and low memory requirements. This method is very popular for mathematician and engineers and those who are interested in solving the large-scale optimization problems [1–3].

Consider the unconstrained optimization problem

$$\min \quad f(x), \quad x \in \mathbb{R}^n, \tag{1}$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable and its gradient is available $g(x)$. Generally CG method generates an iterative sequence $(x_k)$ defined by

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, \ldots, \tag{2}$$

where $\alpha_k > 0$ is a line search and $d_k$ is a search direction defined by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{3}$$

where the term $\beta_k$ is a scalar. There are six essential formulas for $\beta_k$, which are stated as

$$\beta_k^{\mathrm{HS}} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \tag{4}$$

(Hestenes and Stiefel [4], 1952),

$$\beta_k^{\mathrm{FR}} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \tag{5}$$

(Fletcher and Reeves [5], 1964),

$$\beta_k^{\mathrm{PRP}} = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}} \tag{6}$$

(Polak et al. [6, 7], 1969),

$$\beta_k^{\mathrm{CD}} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}} \tag{7}$$

(Conjugate-Descent [8], 1997),

$$\beta_k^{\mathrm{LS}} = -\frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T g_{k-1}} \tag{8}$$

(Liu and Storey [9], 1991),

$$\beta_k^{\mathrm{DY}} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}} \quad (9)$$

(Dai and Yuan [10], 2000).

Generally inexact line search is used in order to get the global convergence of conjugate gradient method, such as Wolfe line search or strong Wolfe line search; the Wolfe line search is given as

$$\begin{gathered} f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k g_k^T d_k, \\ g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \end{gathered} \quad (10)$$

where $0 < \rho < \sigma < 1$. The strong Wolfe line search is computing $\alpha_k$ such that

$$\begin{gathered} f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k g_k^T d_k, \\ \left| g(x_k + \alpha_k d_k)^T d_k \right| \leq \sigma \left| g_k^T d_k \right|. \end{gathered} \quad (11)$$

Recently Alhawarat et al. [11, 12] and Alhawarat and Salleh [13, 14] have proposed an efficient and hybrid conjugate gradient method that satisfies the global convergence properties. To enhance the effectiveness of two-term conjugate gradient method, the three-term conjugate gradient has been widely studied and given much importance. The three-term conjugate gradient method attains different numerical outcomes, depending on how the scalar parameter is being selected. The papers by Beale [15], McGuire and Wolfe [16], Nazareth [17], Deng and Li [18], Dai and Yuan [19], Zhang et al. [20, 21], Cheng [22], Zhang et al. [23], Al-Bayati and Sharif [24], Narushima et al. [25], Andrei [26–28], Sugiki et al. [29], Al-Baali et al. [30], Babaie-Kafaki and Ghanbari [31], and Sun and Liu [32] presented different types of three-term conjugate gradient method along with their numerical performance and efficiency and proved their global convergence properties. As a comparison with classical conjugate gradient algorithms, the proposed three-term conjugate gradient algorithms are numerically strong, efficient, reliable, and robust compared to the classical conjugate gradient algorithms and Beale [15] was the first to propose the three-term conjugate gradient method.

In the new three-term modification, we put our attention on the numerator of PRP method, in which the parameter $\beta_k$ is given as

$$\beta_k^{\mathrm{PRP}} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}. \quad (12)$$

The PRP method is among one of the most efficient and reliable conjugate gradient method due to good numerical performance. The global convergence of PRP is established when the objective function $f$ is strongly convex and the line search is exact [6]. On the other hand Powell [33] through his analysis expressed that there exist nonconvex functions for which PRP method does not converge globally. Gilbert and Nocedal [34] established the so-called $\mathrm{PRP}^+$ method; in this method $\beta_k^{\mathrm{PRP}^+}$ is restricted to be nonnegative denoted as $\beta_k^{\mathrm{PRP}^+} = \max\{0, \mathrm{PRP}\}$. If the standard Wolfe line search (10) is used, then $\mathrm{PRP}^+$method attains the global convergence and also sufficient descent conditions are being satisfied.

Recently Sun and Liu [32] proposed a new conjugate gradient method called TMPRP 1 method by using the VFR formula from Wei et al. [35], in which the search direction is stated as

$$d_k^{(\mathrm{TMPRP1})} = \begin{cases} -g_k, & \text{if } k = 0, \\ -\left(1 + \beta_k^{\mathrm{MPRP}} \dfrac{g_k^T d_{k-1}}{\|g_k\|^2}\right) g_k + \beta_k^{\mathrm{MPRP}} d_{k-1}, & \text{if } k \geq 1, \end{cases} \quad (13)$$

where $\beta_k^{\mathrm{MPRP}} = g_k^T(g_k - g_{k-1})/(\mu|g_k^T d_{k-1}| + \|g_{k-1}\|^2)$ or $\beta_k^{\mathrm{MPRP}+} = \max\{g_k^T(g_k - g_{k-1})/(\mu|g_k^T d_{k-1}| + \|g_{k-1}\|^2), 0\}$.

This method has attractive property of satisfying the sufficient descent condition $g_k^T d_k = -\|g_k\|^2$ independent of any line search and attains global convergence if standard Wolfe line is used. As compared with the strong Wolfe line search, the standard Wolfe line search takes less computation in order to get an acceptable step size at each iteration. Hence the standard Wolfe line search increases the effectiveness of the conjugate gradient method [32].

The rest of the paper is organized as follows. In Section 2, the motivation and formula for construction of three-term conjugate gradient method are given. In Section 3 we have presented Algorithm 1.1 in which the general form of three-term conjugate gradient method is shown. In Section 4 the sufficient descent condition and the global convergence properties for convex and nonconvex function are proven. In Section 5, the detailed numerical results to test the proposed method are reported.

## 2. Motivation and Formula

Wei et al. [35] proposed three new formulas which are given in the following:

$$\begin{gathered} \beta_k^{2*}(\mu, \eta) = \frac{\eta \|g_k\|^2}{\mu |g_k^T d_{k-1}| - \eta g_{k-1}^T d_{k-1}}, \\ \text{for } \eta \in [0, +\infty),\ \mu \in (\eta, +\infty), \\ \beta_k^{**}(\mu_1, \mu_2, \mu_3) = \frac{\mu_1 \|g_k\|^2}{\mu_2 |g_k^T d_{k-1}| - \mu_3 g_{k-1}^T d_{k-1}}, \\ \text{for } \mu_1 \in (0, +\infty),\ \mu_2 \in [\mu_1 + \varepsilon_1, +\infty),\ \mu_3 \in (0, +\infty), \\ \beta_k^{\mathrm{VFR}}(\mu) = \frac{\mu_1 \|g_k\|^2}{\mu_2 |g_k^T d_{k-1}| + \mu_3 \|g_{k-1}\|^2}, \\ \text{for } \mu_1 \in (0, +\infty),\ \mu_2 \in [\mu_1 + \varepsilon_1, +\infty),\ \mu_3 \in (0, +\infty). \end{gathered} \quad (14)$$

There is an efficient conjugate gradient method named $\beta_k^{2*}(\mu, \eta)$. In this formula, the denominator plays an important role in satisfying the sufficient descent condition and performs well in terms of global convergence and numerical result. This motivated us to take the denominator from this formula. Secondly, the PRP [6, 7] method is considered to be one of the most proficient CG parameters due to the properties of its numerator $(g_k - g_{k-1})$. If the step taken becomes very small, then $(g_k - g_{k-1})$ reaches zero such as $(g_k - g_{k-1}) \approx 0$. Afterwards $\beta_k^{PRP} \approx 0$; then the search direction continued as the steepest descent method. So the numerator of PRP method worked efficiently and does not jam.

This motivated us to construct a new modified three-term conjugate gradient method, such as

$$\beta_k^{BZAU} = \frac{g_k^T (g_k - g_{k-1})}{-\eta g_{k-1}^T d_{k-1} + \mu |g_k^T d_{k-1}|}, \quad \text{for } \eta \in [1, +\infty), \ \mu \in (\eta, +\infty). \tag{15}$$

Further, Powell [36] showed that the PRP method can cycle infinitely without approaching a minimum point, even if the step size $\alpha_k$ is chosen to the least positive minimizer. To overcome this, Gilbert and Nocedal [34] showed their analysis

$$\beta_k^{PRP+} = \max\left\{\beta_k^{PRP}, 0\right\}. \tag{16}$$

So

$$\beta_k^{BZAU+} = \max\left\{\frac{g_k^T (g_k - g_{k-1})}{-\eta g_{k-1}^T d_{k-1} + \mu |g_k^T d_{k-1}|}, 0\right\} \quad \text{for } \eta \in [1, +\infty), \ \mu \in (\eta, +\infty), \tag{17}$$

$$\theta_k^{BZAU} = \frac{g_k^T d_{k-1}}{-\eta g_{k-1}^T d_{k-1} + \mu |g_k^T d_{k-1}|} \quad \text{for } \eta \in [1, +\infty), \ \mu \in (\eta, +\infty). \tag{18}$$

In $\beta_k^{BZAU}$, $\beta_k^{BZAU+}$, and $\theta_k^{BZAU}$, the parameters $\eta$ and $\mu$ have an important role in the sense that when $\eta$ is getting smaller, the numbers of iteration, function evaluation, and gradient evaluation are decreased and when $\mu$ is getting larger, the numbers of iteration, function evaluation, and gradient evaluation are also decreased. So we observe that the best value for the parameters is $(\eta, \mu) = (1, 2)$.

## 3. Algorithm 1.1

*Step 0.* Given an initial point $x_0 \in \mathbb{R}^n$, $\mu = 2$, $\eta = 1$, $0 < \rho < \sigma < 1$, and set $d_0 = -g_0$, $k := 0$.

*Step 1.* If $\|g_k\| \le \varepsilon$, where $\varepsilon = 10^{-6}$, then the algorithm stops; otherwise, go to Step 2.

(Note: all the norm $\|\cdot\|$ we use in this paper means $\|\cdot\|_2$).

*Step 2.* Compute the search direction (19) by using $\beta_k^{BZAU}$ and $\theta_k^{BZAU}$ where

$$\beta_k^{BZAU} = \frac{g_k^T (g_k - g_{k-1})}{-\eta g_{k-1}^T d_{k-1} + \mu |g_k^T d_{k-1}|},$$
$$\theta_k^{BZAU} = \frac{g_k^T d_{k-1}}{-\eta g_{k-1}^T d_{k-1} + \mu |g_k^T d_{k-1}|}$$
$$d_k = \begin{cases} -g_k & \text{if } k = 0, \\ -g_k + \beta_k^{BZAU} d_{k-1} - \theta_k^{BZAU} y_{k-1} & \text{if } k \ge 1, \end{cases} \tag{19}$$

where $y_{k-1} = g_k - g_{k-1}$.

*Step 3.* Determine the step size $\alpha_k > 0$ by the Wolfe line search (10).

*Step 4.* Compute $x_{k+1} = x_k + \alpha_k d_k$ where $\alpha_k$ is given in Step 3 and $d_k$ is given in Step 2.

*Step 5.* Set $k = k + 1$ and go to Step 1.

## 4. Global Convergence of Modified Three Term

*Assumptions 1.* (A1) The level set $\mathbb{R}_0 = \{x \mid f(x) \le f(x_0)\}$ is bounded.

(A2) In some neighborhood $\mathcal{N}$ of $\mathbb{R}_0$, the gradient $g(x)$ is Lipschitz continuous on an open convex set $E$ that contains $\mathbb{R}_0$; that is, there exists a positive constant $L > 0$ such that

$$\|g(x_k) - g(x_{k-1})\| \le L \|x_k - x_{k-1}\| \quad \text{for any } x_k, x_{k-1} \in E; \tag{20}$$

then Assumptions (A1) and (A2) and [32, 34] imply that there exist positive constants $\gamma$ and $b$ such that

$$\|g(x_k)\| \le \gamma \quad \forall x_k \in \mathbb{R}_0, \tag{21}$$

$$\|x_k - x_{k-1}\| \le b \quad \forall x_k, \ x_{k-1} \in \mathbb{R}_0. \tag{22}$$

Since $f(x_k)$ is decreasing as $k \to +\infty$, from Assumption (A1) it is shown that the sequence $(x_k)$ created by Algorithm 1.1 will be contained in a bounded region. Then the sequence $(x_k)$ is convergent.

Now we will prove the sufficient descent condition independent of line search $g_k^T d_k = -\|g_k\|^2$ and also $\|g_k\| \le \|d_k\|$. From (19),

$$d_k = -g_k + \beta_k^{BZAU} d_{k-1} - \theta_k^{BZAU} y_{k-1}. \tag{23}$$

Multiplying by $g_k^T$, we obtain

$$g_k^T d_k = -\|g_k\|^2 + \frac{g_k^T (g_k - g_{k-1}) (g_k^T d_{k-1})}{-\eta g_{k-1}^T d_{k-1} + \mu |g_k^T d_{k-1}|} - \frac{(g_k^T d_{k-1}) g_k^T (g_k - g_{k-1})}{-\eta g_{k-1}^T d_{k-1} + \mu |g_k^T d_{k-1}|}; \tag{24}$$

that is,

$$g_k^T d_k = -\|g_k\|^2. \tag{25}$$

Hence the sufficient descent condition independent of line search holds.

Now we prove that $\|g_k\| \le \|d_k\|$. As we have $g_k^T d_k = -\|g_k\|^2$, by taking modulus on both sides

$$\left|g_k^T d_k\right| = \left|-\|g_k\|^2\right| = \|g_k\|^2. \tag{26}$$

By Schwarz inequality we have

$$\left|g_k^T d_k\right| \le \|g_k\| \|d_k\|. \tag{27}$$

So,

$$\begin{gathered}\|g_k\|^2 \le \|g_k\| \|d_k\| \\ \text{or } \|g_k\| \le \|d_k\|.\end{gathered} \tag{28}$$

Thus we have

$$\begin{gathered}g_k^T d_k = -\|g_k\|^2, \\ \|g_k\| \le \|d_k\|.\end{gathered} \tag{29}$$

**Lemma 1.** *Assumptions (A1) and (A2) hold if $x_0$ is supposed to be an initial point. Now consider any method in the form of (2), in which $d_k$ is a descent direction and $\alpha_k$ satisfies the Wolfe condition (10) or the strong Wolfe line search condition (11). Then we have the Zoutendijk condition:*

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < +\infty, \tag{30}$$

*which is normally used to prove the global convergence of CG method. From (29) the Zoutendijk condition is equivalent to the following inequality:*

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{31}$$

*Definition 2* (see [32]). The function $f$ is called uniformly convex on $\mathbb{R}^n$, if there exists a positive constant $m$ such that

$$m\|d_k\|^2 \le d^T \nabla^2 f(x_k) d \quad \forall x, d \in \mathbb{R}^n, \tag{32}$$

where the function $f$ has the Hessian matrix $\nabla^2 f(x_k)$.

We now show the global convergence of Algorithm 1.1 for uniformly convex functions.

**Lemma 3.** *Let both sequences $(x_k)$ and $(d_k)$ be generated by Algorithm 1.1 and suppose that (32) holds; then*

$$z_1 \alpha_k \|d_k\|^2 \le -g_k^T d_k, \tag{33}$$

*where $z_1 = (1-\rho)^{-1}(m/2)$, $m$ is a positive constant, and $\rho$ is a positive number whose range is $0 < \rho < \sigma < 1$, from the Wolfe line search (10).*

*Proof.* Detail of proof can be seen in Lemma 2.1 of [37]. □

**Theorem 4.** *Suppose that Assumptions (A1) and (A2) hold and the function $f$ is uniformly convex; then*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{34}$$

*Proof.* From (15), (33), and (A2), we have

$$\begin{aligned}\left|\beta_k^{\mathrm{BZAU}}\right| &\le \left|\frac{g_k^T(g_k - g_{k-1})}{\eta\left(-g_{k-1}^T d_{k-1}\right)}\right| \\ &\le \frac{\|g_k\| L \|x_k - x_{k-1}\|}{\eta\left(z_1 \alpha_{k-1} \|d_{k-1}\|^2\right)} \\ &= \frac{\|g_k\| L \alpha_{k-1} \|d_{k-1}\|}{\eta\left(z_1 \alpha_{k-1} \|d_{k-1}\|^2\right)}, \\ \left|\beta_k^{\mathrm{BZAU}}\right| \|d_{k-1}\| &\le \left(\frac{L\|g_k\|}{\eta\left(z_1\|d_{k-1}\|\right)}\right)\|d_{k-1}\| = \frac{L\|g_k\|}{\eta z_1}.\end{aligned} \tag{35}$$

Now

$$\left|\theta_k^{\mathrm{BZAU}}\right| \|y_{k-1}\| \le \left|\frac{g_k^T d_{k-1}}{\eta\left(-g_{k-1}^T d_{k-1}\right)}\right| \|y_{k-1}\|. \tag{36}$$

From (18), (33), and (A2), we have

$$\begin{aligned}\left|\theta_k^{\mathrm{BZAU}}\right| \|y_{k-1}\| &\le \left(\frac{\|g_k\| L \|x_k - x_{k-1}\|}{\eta\left(z_1 \alpha_{k-1} \|d_{k-1}\|^2\right)}\right)\|d_{k-1}\| \\ &= \left(\frac{\|g_k\| L \alpha_{k-1} \|d_{k-1}\|}{\eta\left(z_1 \alpha_{k-1} \|d_{k-1}\|^2\right)}\right)\|d_{k-1}\| \\ &= \frac{L\|g_k\|}{\eta z_1}.\end{aligned} \tag{37}$$

Combining (35) and (37) with (19),

$$\begin{aligned}\|d_k\| &\le \|g_k\| + \left|\beta_k^{\mathrm{BZAU}}\right| \|d_{k-1}\| + \left|\theta_k^{\mathrm{BZAU}}\right| \|y_{k-1}\| \\ &\le \|g_k\| + \frac{L\|g_k\|}{\eta z_1} + \frac{L\|g_k\|}{\eta z_1} = \left(1 + \frac{2L}{\eta z_1}\right)\|g_k\|.\end{aligned} \tag{38}$$

Now letting $\sqrt{B} = (1 + 2L/\eta z_1)$, we get $\|d_k\|^2 \le B\|g_k\|^2$,

$$\begin{gathered}\frac{1}{\|d_k\|^2} \ge \frac{1}{B\|g_k\|^2}, \\ \frac{B\|g_k\|^4}{\|d_k\|^2} \ge \frac{\|g_k\|^4}{\|g_k\|^2} = \|g_k\|^2.\end{gathered} \tag{39}$$

So by (31), we get

$$\lim_{k\to\infty} \|g_k\|^2 \le B \lim_{k\to\infty} \frac{\|g_k\|^4}{\|d_k\|^2} = 0. \tag{40}$$

□

We now show the global convergence for nonconvex functions.

**Lemma 5.** *Let Assumptions (A1) and (A2) hold. Consider the sequence $(x_k)$ to be generated by Algorithm 1.1. If there exists a positive constant $\epsilon > 0$ in such a way that $\|g_k\| \geq \epsilon$ for every $k \geq 0$,*

$$\sum_{k=0}^{\infty} \left\|u_{k+1} - u_k\right\|^2 < +\infty, \tag{41}$$

*where $u_k = d_k/\|d_k\|$.*

*Proof.* Since $g_k^T d_k = -\|g_k\|^2$, $\|g_k\| \leq \|d_k\|$, and $\|g_k\| \geq \epsilon$ for every $k$, then we get $\|d_k\| > 0$ for every $k$, so that $u_k$ is well defined. If

$$\begin{aligned} r_k &= -\frac{\left(1 + \theta_k^{\mathrm{BZAU}} y_{k-1} g_k^T / \left\|g_k\right\|^2\right) g_k}{\left\|d_k\right\|}, \\ \delta_k &= \beta_k^{\mathrm{BZAU}+} \frac{\left\|d_{k-1}\right\|}{\left\|d_k\right\|}, \end{aligned} \tag{42}$$

then we get $u_k = r_k + \delta_k u_{k-1}$. Also $u_k$ and $u_{k-1}$ are unit vectors, so

$$r_k = \left\|u_k - \delta_k u_{k-1}\right\| = \left\|\delta_k u_k - u_{k-1}\right\| \tag{43}$$

as we know $\delta_k \geq 0$,

$$\begin{aligned} \left\|u_k - u_{k-1}\right\| &\leq \left\|\left(1 + \delta_k\right)\left(u_k - u_{k-1}\right)\right\| \\ &\leq \left\|u_k - \delta_k u_{k-1}\right\| + \left\|\delta_k u_k - u_{k-1}\right\| \\ &= 2\left\|r_k\right\|. \end{aligned} \tag{44}$$

Now from (18), (22), and (A2)

$$\begin{aligned} \left|\theta_k^{\mathrm{BZAU}}\right| \frac{\left\|y_{k-1}\right\| \left\|g_k^T\right\|}{\left\|g_k\right\|^2} &\leq \left|\frac{g_k^T d_{k-1}}{\mu \left|g_k^T d_{k-1}\right|}\right| \frac{\left\|y_{k-1}\right\| \left\|g_k^T\right\|}{\left\|g_k\right\|^2} \\ &\leq \frac{L\left\|x_k - x_{k-1}\right\|}{\mu\left\|g_k\right\|} \leq \frac{Lb}{\epsilon\mu}. \end{aligned} \tag{45}$$

Now from (21), (22), and (45), there is a constant $N_1 \geq 0$ as follows:

$$\begin{aligned} &\left\|-\left(1 + \frac{\theta_k^{\mathrm{BZAU}} y_{k-1} g_k^T}{\left\|g_k\right\|^2}\right) g_k\right\| \\ &\leq \left\|g_k\right\| + \left(\left|\theta_k^{\mathrm{BZAU}}\right| \frac{\left\|y_{k-1}\right\| \left\|g_k^T\right\|}{\left\|g_k\right\|^2}\right) \left\|g_k\right\| \leq \gamma + \frac{Lb}{\epsilon\mu}\gamma \\ &= N_1. \end{aligned} \tag{46}$$

Therefore, from (31) and (46), we have

$$\begin{aligned} \sum_{k=0}^{\infty} \left\|r_k\right\|^2 &\leq \sum_{k=0}^{\infty} \frac{N_1^2}{\left\|d_k\right\|^2} \leq \sum_{k=0}^{\infty} \frac{N_1^2}{\left\|g_k\right\|^4} \frac{\left\|g_k\right\|^4}{\left\|d_k\right\|^2} \\ &= \frac{N_1^2}{\epsilon^4} \sum_{k=0}^{\infty} \frac{\left\|g_k\right\|^4}{\left\|d_k\right\|^2} < +\infty. \end{aligned} \tag{47}$$

This along with (44) completes the proof. □

**Theorem 6.** *Suppose that Assumptions (A1) and (A2) hold. Then the sequence $(x_k)$ generated by Algorithm 1.1 satisfies*

$$\lim_{k \to \infty} \inf \left\|g_k\right\| = 0. \tag{48}$$

*Proof.* Suppose that the conclusion (48) is not true. Then there exists a positive constant $\epsilon > 0$ in such a way that $\|g_k\| \geq \epsilon\ \forall k \geq 0$.

The proof has the following two parts.

*Part 1.* We noticed that for any $l$ and $k$ we have $l \geq k$, such that

$$x_l - x_k = \sum_{i=k}^{l-1} \left(x_{i+1} - x_i\right) = \sum_{i=k}^{l-1} \left\|s_i\right\| u_i. \tag{49}$$

Proceeding the same proof of Theorem 2.2, step 1 from [32], we have

$$\sum_{i=k}^{l-1} \left\|s_i\right\| \leq 2b. \tag{50}$$

*Part 2.* Taking a bound on the direction $d_k$. Now from (19) and (46) we have

$$\begin{aligned} \left\|d_k\right\|^2 &\leq \left(\left\|-\left(1 + \frac{\theta_k^{\mathrm{BZAU}} g_k^T y_{k-1}}{\left\|g_k\right\|^2}\right) g_k\right\| \right. \\ &\left. + \left|\beta_k^{\mathrm{BZAU}+}\right| \left\|d_{k-1}\right\|\right)^2 \leq \left(N_1 + \left|\beta_k^{\mathrm{BZAU}+}\right| \left\|d_{k-1}\right\|\right)^2 \\ &\leq 2N_1^2 + 2\left(\left|\beta_k^{\mathrm{BZAU}+}\right| \left\|d_{k-1}\right\|\right)^2 \leq 2N_1^2 + 2 \\ &\cdot \frac{\gamma^2 L^2 \left\|s_{k-1}\right\|^2}{\varepsilon^2} \left\|d_{k-1}\right\|^2. \end{aligned} \tag{51}$$

At the beginning of proof we assume that $\lim_{k\to\infty} \inf \|g_k\| \neq 0$; then there exists a constant $\epsilon > 0$ and also there exists $\gamma > 0$ such that $\|g_k\| > \gamma > 0$. Then

$$\begin{gathered} \sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\left\|d_k\right\|^2} > +\infty \Longrightarrow \\ \frac{1}{\left\|d_k\right\|^2} > +\infty, \end{gathered} \tag{52}$$

which contradicts with (A2), (31), and (51). Hence it is proved that $\lim_{k\to\infty} \inf \|g_k\| = 0$. □

## 5. Numerical Results

In this part we compare the numerical results of proposed three-term BZAU (Bakhtawar, Zabidin, Ahmad and Ummu) method with recently developed TMPRP1 method and also compare their performance. The Wolfe line search (10) is used and the values of the parameters for BZAU and TMPRP1 method are $\mu = 2, 10^{(-4)}$; $\eta = 1, 0$; $\rho = 0.1, 0.1$; and $\sigma = 0.5, 0.5$,

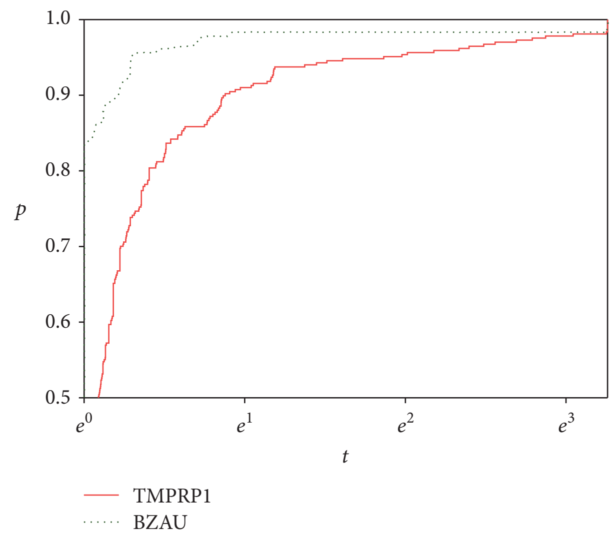


FIGURE 1: Performance profiles based on the number of iterations.

respectively. The code was written in Matlab 7.1 and run on an i5 computer with 2.40 GHz CPU processor, 2.0 GB RAM memory. We test the functions taken from [38] with dimension ranges $[2, 5000]$. The main purpose in optimization for the selection of large number of test functions is to test the unconstrained optimization algorithms properly. Dantzig (1914–2005) said the final test of a theory is its capacity to solve the problems which originated it. This is one of the main reasons we select the large-scale unconstrained optimization problems to test the theoretical progress in numerical form through mathematical programming [38].

Moré et al. [39] claimed the efficiency of a method and that algorithm for a small number of test functions is not suitable because this will lead to the choice of an algorithm that is not favorable. Testing a method or algorithm for a large number of test functions would lead to large amount of data and from that data we can interpret which method or algorithm is more efficient and robust. But the number of test functions should not be very large nor very small, so there is a benchmark of 75 numbers of test functions which are chosen to test the efficiency of any method.

Practically, optimizers need to evaluate nonlinear optimization method. To prove the global convergence properties of any method, the theory is not enough to determine the reliability and efficiency of a method. As a result, the robustness of any method is established by testing the large number of test problems [38].

In global convergence property $\beta_k^{\mathrm{BZAU}}$ is used in case of proving convex function and property $\beta_k^{\mathrm{BZAU+}}$ is used for proving nonconvex function. But in the numerical part $\beta_k^{\mathrm{BZAU+}}$ is used for a comparison with TMPRP1 method. The TMPRP1 possesses the sufficient descent property without any line searches. Theoretically, TMPRP1 method established well and converges globally. When it comes to numerical computation, the TMPRP1 method is tested by a benchmark of 75 numbers of test functions and shows a promising result.

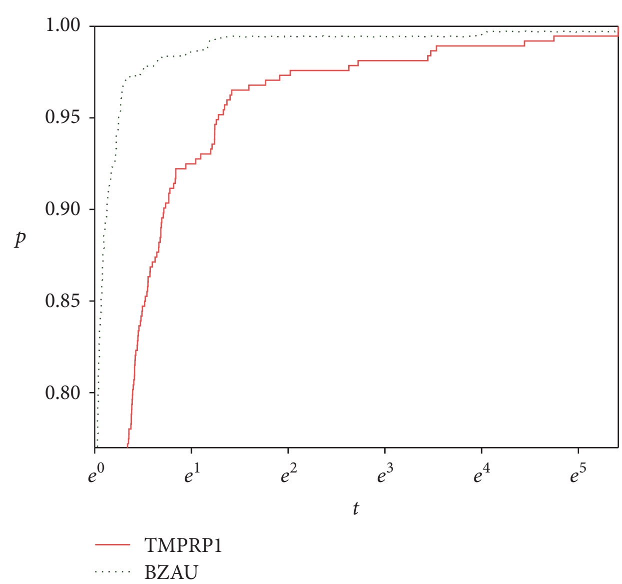


FIGURE 2: Performance profiles based on the CPU time.

Hence the TMPRP1 method is then compared with our BZAU method.

In Table 1 number of iterations, number of function evaluations, number of gradient evaluations, and CPU time are represented by NI/NF/GE/CT. If the CT exceeds 500 seconds and the NI is more than 10000 iterations, then the function is given the name of Fail F. This standard is followed by every paper. For most of the function we can get the result within this limit and the function that does not come in this limit is named Fail F.

The performance profiles are adopted by Dolan and Moré [40]. In Figures 1–4 we compare the performance based on the NI/NF/GE/CT. For every method, we plot fraction $P$ of problems for which the method is within a factor $t$ of

Table 1: A list of test problem functions.

| Problem | $n$ | BZAU<br>NI/GE/FE/CT | TMPRP1<br>NI/GE/FE/CT |
|---|---|---|---|
| Extended white and holst function | 500 | 31/223/191/1.4647 | 32/193/160/1.1246 |
| | 1000 | 31/223/191/1.845 | 32/193/160/1.4398 |
| Extended Rosenbrock function | 500 | 28/170/141/1.1573 | 21/110/88/0.8869 |
| | 1000 | 28/170/141/1.152 | 21/110/88/0.9232 |
| Extended Himmelblau function | 500 | 7/34/26/0.6446 | 10/47/36/0.6638 |
| | 1000 | 7/34/26/0.7544 | 11/51/39/0.697 |
| | 5000 | 7/34/26/0.9163 | 12/55/42/1.0252 |
| Extended DENSCHNB function | 2 | 5/22/16/0.6135 | 6/26/19/0.5955 |
| | 1000 | 6/26/16/0.68 | 6/26/19/0.7011 |
| Shallow function | 500 | 9/41/31/0.6188 | 8/39/30/0.6341 |
| | 1000 | 9/41/31/0.7291 | 8/39/30/0.764 |
| DIXMAANA function | 6000 | 8/35/26/1.65 | 10/43/32/1.7494 |
| | 60 | 8/35/26/2.1339 | 10/67/56/2.7067 |
| NONDIA | 500 | 8/59/50/0.6976 | 11/66/54/0.7215 |
| | 5000 | 14/84/69/1.4958 | 16/90/73/1.5903 |
| DQDRTIC | 100 | 5/21/15/0.6202 | 5/21/15/0.658 |
| | 1000 | 5/21/15/0.5997 | 5/21/15/0.6591 |
| Extended Block Diagonal BD 1 | 500 | 32/37000/28750/1.9822 | 35/73750/64750/2.6201 |
| | 1000 | 32/75500/59000/2.4557 | 36/147000/128500/4.261 |
| Extended Tridiagonal 1 | 2 | 4/18/13/0.5789 | 1694/5096/3401/18.6813 |
| | 5000 | 2072/7767/5694/130.172 | F |
| Diagonal 4 | 100 | 2/9/6/0.5781 | 2/9/6/0.589 |
| | 5000 | 2/9/6/0.6138 | 2/9/6/0.623 |
| Extended Powell | 100 | 400/41100/31075/7.6346 | 92/9925/7600/2.4096 |
| | 500 | 400/205500/155375/7.58 | 92/49625/38000/2.3136 |
| Diagonal 2 | 5 | 44/135/90/1.008 | 48/147/98/1.2072 |
| | 100 | 964/2895/1930/12.4522 | 968/2907/1938/12.5155 |
| Sum Squares | 100 | 57/229/171/1.4846 | 57/229/171/1.5077 |
| | 1000 | 165/661/495/3.1385 | 165/661/495/3.2041 |
| SINQUAD | 70 | 75/376/300/2.0587 | 138/653/514/3.5782 |
| | 100 | 131/630/498/3.9387 | 594/2252/1657/11.1859 |
| DIXMAANC | 66 | 6/29/22/0.7458 | 7/33/25/0.7998 |
| | 120 | 6/29/22/0.7667 | 7/33/25/0.7857 |
| DIXMAANE | 150 | 60/300/239/2.1251 | 434/1374/939/7.1568 |
| | 3000 | 244/1220/975/25.8693 | F |
| DIXMAANG | 150 | 66/330/263/2.3892 | 681/2073/1391/9.6909 |
| | 3000 | 2961/13363/10401/281.4836 | F |
| DIXMAANI | 3 | 20/90/69/0.8965 | 43/136/92/1.0161 |
| | 12 | 287/1269/961/6.297 | 342/1375/1032/6.5883 |
| DIXMAANK | 3 | 21/94/72/0.87 | 52/163/110/0.99 |
| | 12 | 684/2069/1384/9.0948 | 695/2904/1398/9.2842 |
| DENSCHNA | 100 | 8/40/31/0.6462 | 8/37/28/0.7929 |
| | 500 | 8/40/31/0.8236 | 8/37/28/0.8122 |
| DENSCHNB | 2 | 5/22/16/0.674 | 6/26/19/0.997 |
| | 20 | 5/22/16/0.6559 | 6/26/19/0.6583 |

Table 1: Continued.

| Problem | $n$ | BZAU NI/GE/FE/CT | TMPRP1 NI/GE/FE/CT |
|---|---|---|---|
| DENSCHNC | 100 | 7/38/30/0.7599 | 8/42/33/0.6927 |
| | 500 | 7/38/30/0.8327 | 8/42/33/0.8875 |
| ARGLINB | 500 | 2/9/6/0.5714 | 2/9/6/0.5631 |
| | 1000 | 2/9/6/0.5777 | 2/9/6/0.6239 |
| Power | 2 | 2/9/6/0.5587 | 2/9/6/0.5702 |
| | 50 | 66/265/198/1.7672 | 66/265/198/1.7369 |
| Fletcher | 5 | 22/120/97/0.964 | 26/143/116/1.0727 |
| | 50 | 242/1333/1090/8.6331 | 278/1495/1216/9.7661 |
| HIMMELBH | 100 | 9/77/67/0.8398 | 8/74/65/1.1388 |
| | 5000 | 9/77/67/1.1324 | 8/74/65/1.7183 |
| NONDQUAR | 5 | 660/2649/1988/14.3065 | 739/2964/2224/16.1369 |
| | 30 | 7170/28703/21532/110.70 | F |
| Generalized quartic GQ1 | 1000 | 11/47/35/0.706 | 13/55/41/0.6912 |
| | 5000 | 11/47/35/0.987 | 12/52/39/1.219 |
| Generalized quartic GQ2 | 500 | 24/106/81/0.9497 | 26/115/88/1.0062 |
| | 1000 | 23/102/78/0.9412 | 24/107/82/0.9761 |
| Extended Quadratic Penalty | 100 | 12/61/48/0.7067 | 10/43/32/0.642 |
| | 5000 | 15/92/76/1.235 | 18/119/100/1.2217 |
| ARWHEAD | 2 | 8/36/27/0.6249 | 7/32/24/0.5877 |
| | 100 | 7/39/31/0.6106 | 6/35/28/0.6174 |
| Quartic | 2 | 3941/11826/7884/56.888 | 3945/11838/7892/59.0699 |
| | 5 | 5358/16077/10718/80.029 | 5362/16089/10726/78.05 |
| Extrosnb | 100 | 0/1/0/0.5027 | 0/1/0/0.5208 |
| | 5000 | 0/496/334/0.5188 | 0/1/0/0.5168 |
| LIARWHD | 2 | 11/56/44/0.7354 | 12/66/53/0.794 |
| | 100 | 16/94/77/0.8861 | 20/134/113/1.3092 |
| ENGVAL8 | 2 | 7/36/26/0.6582 | 8/37/28/0.6912 |
| | 100 | F | F |
| Hager | 2 | 4/17/12/0.6103 | 4/17/12/0.5686 |
| | 100 | 25/151/125/1.2086 | 25/102/76/0.9215 |
| Extended Wood | 100 | 91/509/417/3.0717 | 116/627/510/3.6229 |
| | 500 | 68/423/354/2.7678 | 87/535/447/3.3762 |
| BIGGSB1 | 500 | 1579/6316/4736/31.3271 | 1627/6508/4880/32.1275 |
| | 1000 | 2463/9852/7388/48.4366 | 3509/14036/10526/68.93 |
| Extended Beale | 100 | 15/71/55/0.7541 | 12/59/46/0.7065 |
| | 500 | 15/71/55/0.8143 | 12/59/46/0.8638 |
| Extended Penalty | 2 | 9/43/33/0.6341 | 10/46/35/0.6735 |
| | 1000 | 37/252/214/1.7596 | F |
| Extended Maratos Function | 500 | 33/223/189/2.1234 | 31/213/181/1.5054 |
| | 1000 | F | 31/213/181/1.5017 |
| Quadratic QF2 Function | 2 | 6/26/19/0.6122 | 6/26/19/0.6209 |
| | 500 | 293/1269/975/6.6934 | 277/996/768/5.1248 |
| SINCOS | 1000 | 8/39/30/0.8623 | 12/51/38/0.9807 |
| | 5000 | 8/39/30/1.1289 | 12/51/38/1.2905 |
| STAIRCASE S1 | 10 | 31/125/93/1.0555 | 32/129/96/1.0384 |
| | 50 | 135/540/404/2.7473 | 197/790/592/4.0499 |

Table 1: Continued.

| Problem | $n$ | BZAU NI/GE/FE/CT | TMPRP1 NI/GE/FE/CT |
|---|---|---|---|
| GENROSEN-2 | 2 | 25/159/133/1.3379 | 21/110/88/0.9879 |
| | 500 | 6889/36000/29110/185.5 | 74/488/413/3.4692 |
| TRIDIAGONAL DOUBLE BORDED | 50 | 395/1585/1189/7.8008 | 540/2165/1624/11.11 |
| | 100 | 773/3099/2325/15.481 | 1016/4073/3056/20.09 |
| EDENSCH FUNCTION | 500 | 3090/9276/6185/40.9297 | 3094/9288/6193/34.4141 |
| | 1000 | 3090/9276/6185/135.232 | 3094/9288/6193/131.676 |
| DENSCHNF | 500 | 10/59/48/0.9205 | 12/73/60/0.9032 |
| | 1000 | 10/59/48/0.7546 | 12/73/60/0.8977 |
| Generalized Triagonal-1 | 100 | 23/95/71/0.9284 | F |
| | 500 | 25/124/98/1.5822 | 24/100/75/0.9883 |
| Generalized Triagonal-2 | 50 | 43/205/161/1.225 | 35/165/129/1.5369 |
| | 100 | 39/188/148/1.2548 | 39/191/151/1.2668 |
| Extended Trigonomenctric | 5 | 313/965/651/5.1384 | 324/990/665/5.2108 |
| | 100 | 97/437/339/2.6942 | 318/997/678/5.3339 |
| DIAGONAL-1 | 2 | 5/22/16/0.6467 | 6/27/20/0.6174 |
| | 10 | 23/95/71/0.889 | 25/103/77/0.9587 |
| EG3 | 50 | 32/138/105/1.1712 | 33/141/107/1.1336 |
| | 200 | 35/180/144/1.4381 | F |
| Extended Three Exponential | 100 | 8/1700/1250/0.6441 | 6/1300/950/0.6111 |
| | 1000 | 8/17000/12500/0.999 | 6/13000/9500/0.8019 |
| Extended Freudenstein and Roth | 30 | 16/91/74/0.7806 | F |
| | 50 | 16/141/124/1.0795 | F |
| Dixon3DQ | 1000 | 3049/12196/9146/49.19 | 3409/13636/10226/55.20 |
| | 2000 | 4959/19836/14876/80.41 | 6473/25892/19418/117.9 |
| Quadratic Qf1 | 1000 | 187/749/561/4.2752 | 187/749/561/4.3368 |
| | 5000 | 426/1705/1278/10.7051 | 426/1705/1278/11.6042 |
| Extended Quadratic penalty Q2 | 500 | 46/466/419/2.9755 | 51/576/524/4.0679 |
| | 5000 | 61/740/678/6.627 | 86/1052/970/9.3083 |
| Sphere | 2 | 1/5/3/0.5581 | 1/5/3/0.549 |
| | 500 | 1/5/3/0.5498 | 1/5/3/0.5447 |
| Raydan1 | 2 | 101/304/202/1.6957 | 105/316/210/2.0696 |
| | 100 | 66/266/199/1.5276 | 69/278/208/2.0739 |
| BDQRTIC | 10 | 37/179/141/1.1653 | 52/240/187/1.4611 |
| | 500 | F | F |
| DIXMAANJ | 3 | 23/88/64/0.809 | 52/162/109/1.0504 |
| | 45 | 3015/13079/10063/63.498 | 7074/21256/14181/93.00 |
| DIXMAANB | 3 | 5/23/17/0.5848 | 6/27/20/0.652 |
| | 900 | 8/34/25/0.945 | 8/34/25/0.954 |
| DIXMAAND | 300 | 8/38/29/0.932 | 9/42/32/0.9416 |
| | 900 | 7/34/26/1.0641 | 10/46/35/1.215 |
| DIXMAANF | 300 | 90/451/360/3.386 | 1327/3998/2670/19.805 |
| | 900 | 1071/4832/3760/29.3661 | 3488/10489/7000/103.27 |
| DIXMAANH | 300 | 81/408/326/2.6333 | 1319/3974/2654/17.820 |
| | 900 | 1476/6413/4936/76.60 | 3462/10409/6946/121.09 |
| DIXMAANL | 3 | 16/76/59/0.7726 | 52/164/111/1.1245 |
| | 45 | 3023/13115/10091/57.84 | 7031/21130/14098/85.66 |
| Raydan 2 | 500 | 4/17/12/0.5985 | 2/9/6/0.5584 |
| | 1000 | 5/21/15/0.6352 | 2/9/6/0.5593 |

Table 1: Continued.

| Problem | $n$ | BZAU NI/GE/FE/CT | TMPRP1 NI/GE/FE/CT |
|---|---|---|---|
| Perturbed Quadratic | 1000 | 187/749/561/3.9731 | 187/749/561/3.7987 |
| | 5000 | 425/1701/1275/11.039 | 425/1701/1275/10.9626 |
| A Quadratic Function | 2 | 6/26/19/0.6334 | 6/26/19/0.6101 |
| | 100 | 91/396/304/2.2252 | 96/417/320/2.421 |
| TRIDIA | 500 | 240/961/720/5.146 | 240/961/720/5.5388 |
| | 1000 | 349/1397/1047/7.1665 | 349/1397/1047/7.090 |
| Zett1 | 2 | 12/52/39/0.6578 | 14/61/46/0.6748 |
| Six hump | 2 | 7/31/23/0.6482 | 8/36/27/0.6714 |
| Booth | 2 | 2/9/6/0.5722 | 2/9/6/0.5513 |
| Three hump | 2 | 6/28/21/0.5959 | 7/31/23/0.6031 |
| EG2 | 2 | 4/20/15/0.6011 | 5/26/20/1.3894 |

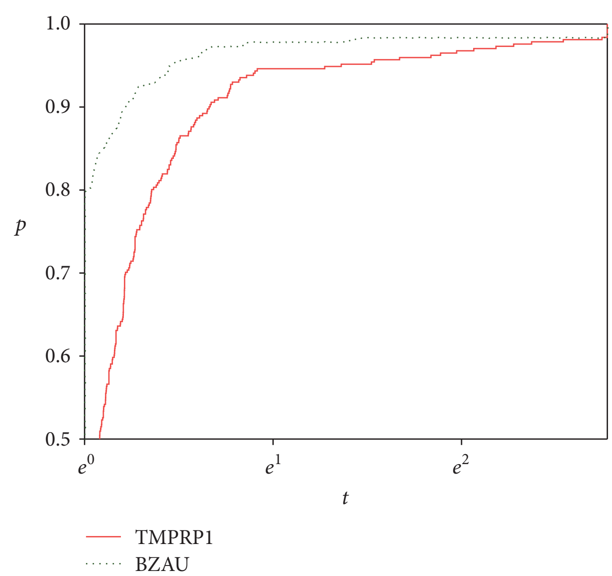


Figure 3: Performance profiles based on number of gradient evaluations.

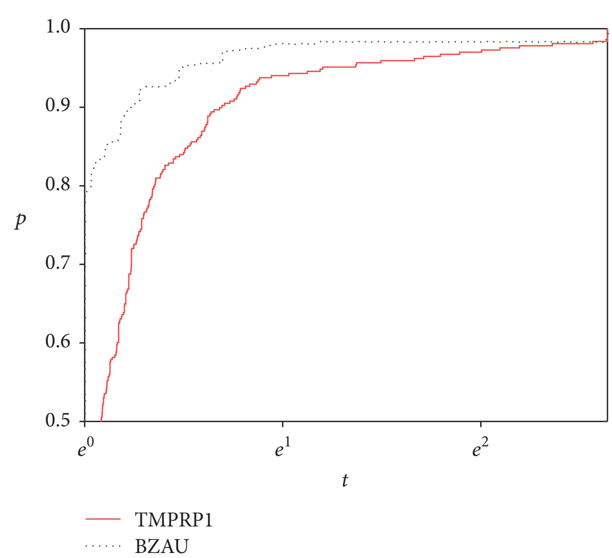


Figure 4: Performance profiles based on number of function evaluations.

best time. The left hand side of the figure represents the percentage of the test problem of which method is robust and the fastest; the right hand side of the figure shows the percentage of test problems that are solved successfully by either the BZAU or TMPRP1 method. In the graph there are two axes $x$ and $y$, as there are much values of $x$ which creates difficulty in understanding the graph. The value of $x$-axis is then converted in natural log of $t$ so it shows $x$-axis values in exponent like $e^0, e^1, e^2, e^3$ and the values of $y$-axis are taken in linear form of $P$. Comparing Figures 1–4 shows that the BZAU method outperforms the TMPRP1 method in every case. The top curve is the most efficient method, so the new modified three-term CG method is also efficient in terms of numerical result.

## 6. Conclusion

In this paper, we have proposed a new modified three-term conjugate gradient method for unconstrained optimization. The new modified three-term BZAU method possesses the sufficient descent property independent of any line search. Global convergence is shown for both convex and nonconvex functions using the Wolfe line search. In numerical result we compare the three-term BZAU method with TMPRP1 method [32]. As in [32] the TMPRP1 method is shown to be numerically efficient when it comes to comparison with other two robust methods such as CG_Descent method [41] and DTPRP method [42]. That is the reason for comparing our BZAU method with TMPRP1 method.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

[1] J. Nocedal, "Conjugate gradient methods and nonlinear optimization," in *Linear and Nonlinear Conjugate Gradient Related Methods*, J. L. Nazareth, Ed., pp. 9–23, SIAM, Philadelphia, PA, USA, 1995.

[2] E. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer, New York, NY, USA, 1997.

[3] R. Ziadi, R. Ellaia, and A. Bencherif-Madani, "Global optimization through a stochastic perturbation of the Polak-Ribière conjugate gradient method," *Journal of Computational and Applied Mathematics*, vol. 317, pp. 672–684, 2017.

[4] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, vol. 49, pp. 409–436 (1953), 1952.

[5] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The Computer Journal*, vol. 7, pp. 149–154, 1964.

[6] E. Polak and G. Ribière, "Note Sur la convergence de directions conjugeès, Rev," *Revue Française d'Informatique et de Recherche Opérationnelle*, vol. 3, no. 16, pp. 35–43, 1969.

[7] B. T. Polyak, "The conjugate gradient method in extreme problems," *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 4, pp. 94–112, 1969.

[8] R. Fletcher, *Practical Method of Optimization, Vol. I: Unconstrained Optimization*, Wiley, New York, NY, USA, 2nd edition, 1997.

[9] Y. Liu and C. Storey, "Efficient generalized conjugate gradient algorithms. I. Theory," *Journal of Optimization Theory and Applications*, vol. 69, no. 1, pp. 129–137, 1991.

[10] Y. H. Dai and Y. Yuan, "A nonlinear conjugate gradient with a strong global convergence properties," *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177–182, 2000.

[11] A. Alhawarat, M. Mamat, M. Rivaie, and Z. Salleh, "An efficient hybrid conjugate gradient method with the strong Wolfe-Powell line search," *Mathematical Problems in Engineering*, vol. 2015, Article ID 103517, 7 pages, 2015.

[12] A. Alhawarat, Z. Salleh, M. Mamat, and M. Rivaie, "An efficient modified Polak–Ribière–Polyak conjugate gradient method with global convergence properties," *Optimization Methods and Software*, pp. 1–14, 2016.

[13] A. Alhawarat and Z. Salleh, "Modification of nonlinear conjugate gradient method with weak Wolfe-Powell line search," *Abstract and Applied Analysis*, Article ID 7238134, Art. ID 7238134, 6 pages, 2017.

[14] Z. Salleh and A. Alhawarat, "An efficient modification of the Hestenes-Stiefel nonlinear conjugate gradient method with restart property," *Journal of Inequalities and Applications*, vol. 2016, no. 1, article no. 110, 2016.

[15] E. M. Beale, "A derivative of conjugate gradients," in *Numerical Methods for Nonlinear Optimization*, F. A. Lootsma, Ed., pp. 39–43, Academic Press, London, 1972.

[16] M. F. McGuire and P. Wolfe, "Evaluating a Restart Procedure for Conjugate Gradients," Tech. Rep., IBM Research Center, Yorktown Heights, 1973.

[17] L. Nazareth, "A conjugate direction algorithm without line searches," *Journal of Optimization Theory and Applications*, vol. 23, no. 3, pp. 373–387, 1977.

[18] N. Y. Deng and Z. Li, "Global convergence of three terms conjugate gradient methods," *Optim. Method Softw*, vol. 4, pp. 273–282, 1995.

[19] Y. H. Dai and Y. Yuan, "A nonlinear conjugate gradient method with a strong global convergence property," *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177–182, 1999.

[20] L. Zhang, W. Zhou, and D.-H. Li, "A descent modified Polak-Ribiére-Polyak conjugate gradient method and its global convergence," *IMA Journal of Numerical Analysis*, vol. 26, no. 4, pp. 629–640, 2006.

[21] L. Zhang, W. Zhou, and D. Li, "Some descent three-term conjugate gradient methods and their global convergence," *Optimization Methods and Software*, vol. 22, no. 4, pp. 697–711, 2007.

[22] W. Cheng, "A two-term PRP-based descent method," *Numerical Functional Analysis and Optimization*, vol. 28, no. 11-12, pp. 1217–1230, 2007.

[23] J. Zhang, Y. Xiao, and Z. Wei, "Nonlinear conjugate gradient methods with sufficient descent condition for large-scale unconstrained optimization," *Mathematical Problems in Engineering*, Article ID 243290, Art. ID 243290, 16 pages, 2009.

[24] A. Y. Al-Bayati and W. H. Sharif, "A new three-term conjugate gradient method for unconstrained optimization," *Canadian Journal on Science & Engineering Mathematics*, vol. 1, no. 5, pp. 108–124, 2010.

[25] Y. Narushima, H. Yabe, and J. A. Ford, "A three-term conjugate gradient method with sufficient descent property for unconstrained optimization," *SIAM Journal on Optimization*, vol. 21, no. 1, pp. 212–230, 2011.

[26] N. Andrei, "A modified Polak-Ribière-Polyak conjugate gradient algorithm for unconstrained optimization," *Optimization. A Journal of Mathematical Programming and Operations Research*, vol. 60, no. 12, pp. 1457–1471, 2011.

[27] N. Andrei, "On three-term conjugate gradient algorithms for unconstrained optimization," *Applied Mathematics and Computation*, vol. 219, no. 11, pp. 6316–6327, 2013.

[28] N. Andrei, "A simple three-term conjugate gradient algorithm for unconstrained optimization," *Journal of Computational and Applied Mathematics*, vol. 241, pp. 19–29, 2013.

[29] K. Sugiki, Y. Narushima, and H. Yabe, "Globally convergent three-term conjugate gradient methods that use secant conditions and generate descent search directions for unconstrained optimization," *Journal of Optimization Theory and Applications*, vol. 153, no. 3, pp. 733–757, 2012.

[30] M. Al-Baali, Y. Narushima, and H. Yabe, "A family of three-term conjugate gradient methods with sufficient descent property for unconstrained optimization," *Computational Optimization and Applications. An International Journal*, vol. 60, no. 1, pp. 89–110, 2015.

[31] S. Babaie-Kafaki and R. Ghanbari, "Two modified three-term conjugate gradient methods with sufficient descent property," *Optimization Letters*, vol. 8, no. 8, pp. 2285–2297, 2014.

[32] M. Sun and J. Liu, "Three modified Polak-Ribière-Polyak conjugate gradient methods with sufficient descent property," *Journal of Inequalities and Applications*, vol. 2015, no. 1, 2015.

[33] M. J. Powell, "Restart procedures for the conjugate gradient method," *Mathematical Programming*, vol. 12, no. 2, pp. 241–254, 1977.

[34] J. C. Gilbert and J. Nocedal, "Global convergence properties of conjugate gradient methods for optimization," *SIAM Journal on Optimization*, vol. 2, no. 1, pp. 21–42, 1992.

[35] Z. Wei, G. Li, and L. Qi, "New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems," *Applied Mathematics and Computation*, vol. 179, no. 2, pp. 407–430, 2006.

[36] M. J. D. Powell, "Nonconvex minimization calculations and the conjugate gradient method," in *Numerical Analysis*, vol. 1066 of *Lecture Notes in Mathematics*, pp. 122–141, Springer, Berlin, Germany, 1984.

[37] Z.-f. Dai and B.-S. Tian, "Global convergence of some modified PRP nonlinear conjugate gradient methods," *Optimization Letters*, vol. 5, no. 4, pp. 615–630, 2011.

[38] N. Andrei, "An unconstrained optimization test functions collection," *Advanced Modeling and Optimization*, vol. 10, no. 1, pp. 147–161, 2008.

[39] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, "Testing unconstrained optimization software," *ACM Transactions on Mathematical Software*, vol. 7, no. 1, pp. 17–41, 1981.

[40] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming. A Publication of the Mathematical Programming Society*, vol. 91, no. 2, Ser. A, pp. 201–213, 2002.

[41] W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," *Pacific Journal of Optimization. An International Journal*, vol. 2, no. 1, pp. 35–58, 2006.

[42] Z. Dai and F. Wen, "Another improved Wei-Yao-Liu nonlinear conjugate gradient method with sufficient descent property," *Applied Mathematics and Computation*, vol. 218, no. 14, pp. 7421–7430, 2012.