Thai Journal of Mathematics
Vol. 18, No. 1 (2020),
Pages 501 - 517

ISSN 1686-0209

SPECTRAL THREE-TERM CONJUGATE DESCENT METHOD FOR SOLVING NONLINEAR MONOTONE EQUATIONS WITH CONVEX CONSTRAINTS

Auwal Bala Abubakar, [1*], Jewaidu Rilwan[1], Seifu Endris Yimer,[2], Abdulkarim Hassan Ibrahim [4], Idris Ahmed [3]

[1] *Department of Mathematical Sciences, Faculty of Physical Sciences, Bayero University, Kano. Kano, Nigeria.*
[2] *Department of Mathematics, College of Computation and Natural Science , Debre Berhan University, P.O. Box 445, Debre Berhan , Ethiopia.*
[3] *Department of Mathematics and Computer Science, Faculty of Natural and Applied Sciences , Sule Lamido University Kafin Hausa, P.M.B 048, Jigawa State, Nigeria.*
[4] *Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha Uthit Road, Bang Mod,Thung Khru, Bangkok 10140, Thailand.*
Email addresses: ababubakar.mth@buk.edu.ng (A. B. Abubakar), jrilwan.mth@buk.edu.ng (J. Rilwan), seifuendris@gmail.com (S. E. Yimer), ibrahimkarym@gmail.com (A. H. Ibrahim), idrisahamedgml1988@gmail.com (I. Ahmed)

Abstract This paper proposes three (3) three term conjugate gradient (CG) methods based on the well known conjugate descent (CD) CG parameter. The two directions were obtained by adding a term to the CD direction such that the sufficient descent property is satisfied. Under some assumptions, we establish the convergence of the proposed methods. In addition, numerical examples were given to show the capability of the two methods in solving nonlinear monotone equations with convex constraints.

MSC: 65K05; 90C52; 90C56; 92C55

Keywords: Nonlinear monotone equations, Conjugate gradient method, Projection method

Submission date: 30.11.2019 / Acceptance date: 18.02.2020

1. Introduction

A nonlinear monotone equation with convex constraints is an equation of the form

$$F(x) = 0, \quad \text{subject to} \quad x \in \Omega, \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and monotone. The set $\Omega \subset \mathbb{R}^n$ is assumed to be nonempty, closed and convex.

The above problem feature in many applications, such as the subproblems in the generalized proximal algorithms with Bregman distance [1], reformulation of ℓ_1-norm regularized problems [2], and conversionn of variational inequality problems into nonlinear monotone equations via fixed point maps [3].

*Corresponding author.

Published by The Mathematical Association of Thailand.

Among famous and efficient methods for solving (1.1) are the conjugate gradient (CG) methods. They are iterative methods for solving unconstrained optimization problem and nonlinear systems, due to low memory requirement (see [4–10] and references their in). This and other reasons inspired many authors to propose CG methods combined with the projection method by Solodov and Svaiter [11].

Given an initial guess x_k, the conjugate gradient algorithm combined with the projection method computes the next guess x_{k+1} as

$$x_{k+1} = P_\Omega[x_k - \zeta_k F(z_k)] \tag{1.2}$$

where

$$P_\Omega(x) = \arg\min\{\|x-y\| \ : y \in \Omega.\},$$

$\zeta_k = \frac{F(z_k)^T(x_k - z_k)}{\|F(z_k)\|^2}$ and $z_k = x_k + \alpha_k d_k$. α_k is the step size obtained using a suitable line search procedure and the search direction

$$d_k = \begin{cases} -F(x_k), & \text{if } k = 0, \\ -F(x_k) + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{1.3}$$

with β_k being the conjugate gradient parameter. An important condition needed for the direction to satisfy is

$$F(x_k)^T d_k \leq -c\|F(x_k)\|^2, \ c > 0. \tag{1.4}$$

This condition ensures a decrease in the norm of the residual at each iteration. A nice property of the projection map is it's nonexpansiveness, that is

$$\|P_\Omega(x) - P_\Omega(y)\| \leq \|x-y\|, \quad \forall x, y \in \mathbb{R}^n. \tag{1.5}$$

Among such methods are the two directions proposed by Ahookhosh et al. [12]. The search directions are defined as

$$d_k = \begin{cases} -F(x_k), & \text{if } k = 0, \\ -F(x_k) + \beta_k^{PRP} w_{k-1} - \theta_k^i y_{k-1}, & \text{if } k \geq 1 \text{ and i=1,2}, \end{cases} \tag{1.6}$$

where

$$\beta_k^{PRP} = \frac{F(x_k)^T y_{k-1}}{\|F(x_{k-1})\|^2},$$

$$\theta_k^1 = \frac{F(x_k)^T y_{k-1}\|w_{k-1}\|^2}{\|F(x_{k-1})\|^4},$$

$$\theta_k^2 = \frac{F(x_k)^T w_{k-1}}{\|F(x_{k-1})\|^2} + \frac{F(x_k)^T y_{k-1}\|y_{k-1}\|^2}{\|F(x_{k-1})\|^4},$$

$$y_{k-1} = F(x_k) - F(x_{k-1}), \ w_{k-1} = z_{k-1} - x_{k-1} = \alpha_{k-1} d_{k-1}.$$

The global convergence was established under the line search

$$-F(z_k)^T d_k \geq \sigma \alpha_k \|F(z_k)\|\|d_k\|^2, \ \sigma > 0. \tag{1.7}$$

Another is that proposed by Papp and Rapajiċ [13] which is modified Fletcher-Reeves (FR) CG method. The directions are defined as

$$d_k = \begin{cases} -F(x_k), & \text{if } k = 0, \\ -F(x_k) + \beta_k^{FR} w_{k-1} - \theta_k^i F(x_k), & \text{if } k \geq 1 \text{ and i=1,2,3}, \end{cases} \tag{1.8}$$

where

$$\beta_k^{FR} = \frac{\|F(x_k)\|^2}{\|F(x_{k-1})\|^2},$$

$$\theta_k^1 = \frac{F(x_k)^T w_{k-1}}{\|F_{k-1}\|^2},$$

$$\theta_k^2 = \frac{\|F(x_k)\|^2 \|w_{k-1}\|^2}{\|\|F(x_{k-1})\|\|^4},$$

$$\theta_k^3 = \frac{F(x_k)^T w_{k-1}}{\|\|F(x_{k-1})\|\|^2} + \frac{\|F(x_k)\|^2}{\|\|F(x_{k-1})\|\|^4}.$$

The three directions based on θ_k^1, θ_k^2, and θ_k^3 are called M3TFR1, M3TFR2 and M3TFR3 respectively. Numerical results presented reveals that overall, M3TFR1 has the least number of iterations. However, in terms of number of function evaluations and CPU time, M3TFR2 was the most efficient and robust.
Likewise, Feng et al. [14] proposed a search direction given by

$$d_k = \begin{cases} -F(x_k), & \text{if } k = 0, \\ -\left(1 + \beta_k \frac{F(x_k)^T d_{k-1}}{\|F(x_k)\|^2}\right) F(x_k) + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{1.9}$$

where $\beta_k = \frac{\|F(x_k)\|}{\|d_{k-1}\|}$.
The global convergence was proved using the line search

$$-F(z_k)^T d_k \geq \sigma \alpha_k \|d_k\|^2, \ \sigma > 0. \tag{1.10}$$

. Just recently, Abubakar et. al [15] proposed a modified version of the method proposed by [13] defined as

$$d_k = \begin{cases} -F(x_k), & \text{if } k = 0, \\ -F(x_k) + \frac{\|F(x_k)\|^2 w_{k-1} - F(x_k)^T w_{k-1} F(x_k)}{\max\{\mu \|w_{k-1}\| \|F(x_k)\|, \|F(x_{k-1})\|^2\}}, & \text{if } k \geq 1, \end{cases} \tag{1.11}$$

where $\mu > 0$ is a positive constant. The difference between the M3TFR1 direction and the above direction is the scaling term appearing in the denominator of Equation (1.11). To have more incite on conjugate gradient projection methods, the reader is referred to [16–22].

Motivated by the above methods and in particular the method proposed in [12, 13], we present three spectral conjugate descent projection methods. Section 2 gives detail of the proposed methods. In Section 3, the global convergence of the proposed methods is discussed. Numerical experiments are carried out in section 4. Finally section 5 has the conclusion.

2. Algorithm: Inspiration and convergence analysis

We begin this section by defining the projection map, then introduce our proposed method together with is convergence analysis.

Definition 2.1. Let $\Omega \subset \mathbb{R}^n$ be a nonempty closed convex set. Then for any $x \in \mathbb{R}^n$, its projection onto Ω, denoted by $P_\Omega(x)$, is defined by

$$P_\Omega(x) = \arg\min\{\|x - y\| \ : y \in \Omega\}.$$

Indeed, P_Ω is nonexpansive, That is,

$$\|P_\Omega(x) - P_\Omega(y)\| \leq \|x - y\|, \quad \forall x, y \in \mathbb{R}^n. \tag{2.1}$$

All through, we assume the followings

(K_1) The function F is monotone, that is,

$$(F(x) - F(y))^T(x - y) \geq 0, \quad \forall x, y \in \mathbb{R}^n.$$

(K_2) The function F is Lipschitz continuous, that is there exists a positive constant L such that

$$\|F(x) - F(y)\| \leq L\|x - y\|, \ \forall x, y \in \mathbb{R}^n.$$

(K_3) The solution set of (1.1), denoted by $\overline{\Omega}$, is nonempty.

An important property that is required of a method for solving equation (1.1) to possess is that the direction d_k satisfy

$$F(x_k)^T d_k \leq -\delta\|F(x_k)\|^2, \tag{2.2}$$

where $\delta > 0$. The above relation (2.2) is called the sufficient descent condition if $F(x)$ is the gradient vector of a real valued function $f : \mathbb{R}^n \to \mathbb{R}$.

Inspired by the directions proposed in [12] and [13], we propose the following search direction

$$d_k = \begin{cases} -F(x_k), & \text{if } k = 0, \\ -F(x_k) + \beta_k^{CD} w_{k-1} - \lambda_k F(x_k), & \text{if } k \geq 1, \end{cases} \tag{2.3}$$

where

$$\beta_k^{CD} = \frac{\|F(x_k)\|^2}{-d_{k-1}^T F(x_{k-1})}, \tag{2.4}$$

$w_{k-1} = z_{k-1} - x_{k-1} = \alpha_{k-1} d_{k-1}$. The parameter λ_k will be derived in three different approaches such that (2.2) is satisfied.

M3TCD1 direction: Multiplying (2.3) by $F(x_k)^T$ and substituting (2.4) in (2.3), we have

$$F(x_k)^T d_k = -\|F(x_k)\|^2 + \frac{\|F(x_k)\|^2}{-d_{k-1}^T F(x_{k-1})} F(x_k)^T w_{k-1} - \lambda_k \|F(x_k)\|^2.$$

By choosing

$$\lambda_k = \frac{F(x_k)^T w_{k-1}}{-d_{k-1}^T F(x_{k-1})}, \tag{2.5}$$

then (2.2) is satisfied with $\delta = 1$, that is $F(x_k)^T d_k = -\|F(x_k)\|^2$.

we call the first modified three term CD like direction M3TCD1 which is defined by (2.3),(2.4) and (2.5).

M3TCD2 direction: Applying similar approach used in [12] and [13], we have

$$\begin{aligned} &F(x_k)^T d_k \\ &= \big[- \|F(x_k)\|^2 (d_{k-1}^T F(x_{k-1}))^2 - d_{k-1}^T F(x_{k-1}) \|F(x_k)\|^2 \\ &\quad \times F(x_k)^T w_{k-1} - \lambda_k \|F(x_k)\|^2 (d_{k-1}^T F(x_{k-1}))^2 \big] / (d_{k-1}^T F(x_{k-1}))^2. \end{aligned}$$

Using the relation $u^T v \leq \frac{1}{2}(\|u\|^2 + \|v\|^2)$ and letting $u = -\frac{1}{\sqrt{2}}(d_{k-1}^T F(x_{k-1}))F(x_k)$ and $v = \sqrt{2}\|F(x_k)\|^2 w_{k-1}$, then

$$
\begin{aligned}
&F(x_k)^T d_k \\
&= \big[- \|F(x_k)\|^2 (d_{k-1}^T F(x_{k-1}))^2 + \frac{1}{4}(d_{k-1}^T F(x_{k-1}))^2 \|F(x_k)\|^2 \\
&\quad + \|F(x_k)\|^4 \|w_{k-1}\|^2 - \lambda_k \|F(x_k)\|^2 (d_{k-1}^T F(x_{k-1}))^2 \big] / (d_{k-1}^T F(x_{k-1}))^2 \\
&= -\|F(x_k)\|^2 + \frac{1}{4}\|F(x_k)\|^2 + \frac{\|F(x_k)\|^4 \|w_{k-1}\|^2}{(d_{k-1}^T F(x_{k-1}))^2} - \lambda_k \|F(x_k)\|^2 \\
&= -\frac{3}{4}\|F(x_k)\|^2 + \frac{\|F(x_k)\|^4 \|w_{k-1}\|^2}{(d_{k-1}^T F(x_{k-1}))^2} - \lambda_k \|F(x_k)\|^2.
\end{aligned}
\tag{2.6}
$$

Choosing

$$
\lambda_k = \frac{\|F(x_k)\|^2 \|w_{k-1}\|^2}{(d_{k-1}^T F(x_{k-1}))^2}, \tag{2.7}
$$

then then (2.2) is satisfied with $\delta = \frac{3}{4}$, that is $F(x_k)^T d_k = -\frac{3}{4}\|F(x_k)\|^2$. The direction defined by (2.3),(2.4) and (2.7) is called M3TCD2.

M3TCD3 direction: This third parameter λ_k is chosen as

$$
\lambda_k = \frac{F(x_k)^T w_{k-1}}{-d_{k-1}^T F(x_{k-1})} + \frac{\|F(x_k)\|^2}{(d_{k-1}^T F(x_{k-1}))^2}. \tag{2.8}
$$

The first term is (2.5) and the second term is chosen such that (2.2) is satisfied. Now, multiplying (2.3) by $F(x_k)^T$ and substituting (2.4) and (2.8) in (2.3), we have

$$
\begin{aligned}
F(x_k)^T d_k &= -\|F(x_k)\|^2 + \frac{\|F(x_k)\|^2}{-d_{k-1}^T F(x_{k-1})} F(x_k)^T w_{k-1} \\
&\quad - \left(\frac{F(x_k)^T w_{k-1}}{-d_{k-1}^T F(x_{k-1})} + \frac{\|F(x_k)\|^2}{(d_{k-1}^T F(x_{k-1}))^2} \right) \|F(x_k)\|^2 \\
&= -\|F(x_k)\|^2 - \frac{\|F(x_k)\|^4}{(d_{k-1}^T F(x_{k-1}))^2} \\
&\leq -\|F(x_k)\|^2.
\end{aligned}
$$

The above relation shows that (2.2) is satisfied with $\delta = 1$. We call the direction defined by (2.3), (2.4) and (2.8) M3TCD3.

To prove the global convergence of Algorithm 1, the following lemmas are needed.

Lemma 2.2. *The directions M3TCD1, M3TCD2 and M3TCD3 by satisfy the sufficient descent condition (2.2).*

Remark 2.3. Since M3TCD1, M3TCD2 and M3TCD3 by satisfy the sufficient descent condition (2.2) $\forall k \in \mathbb{N} \bigcup \{0\}$, then

$$
F(x_{k-1})^T d_{k-1} \leq -\delta \|F(x_{k-1})\|^2,
$$

which implies

$$
\frac{1}{-F(x_{k-1})^T d_{k-1}} \leq \frac{1}{\delta \|F(x_{k-1})\|^2} \tag{2.10}
$$

Algorithm 1

Step 0. Given an arbitrary initial point $x_0 \in \mathbb{R}^n$, parameters $\sigma > 0$, $0 < \rho < 1$, $Tol > 0$ and set $k := 0$.

Step 1. If $\|F(x_k)\| \leq Tol$, stop, otherwise go to **Step 2**.

Step 2. Compute d_k using Equation (2.3).

Step 3. Compute the step size $\alpha_k = \max\{\gamma\rho^i \ : i = 0, 1, 2, \cdots\}$ such that

$$-F(x_k + \alpha_k d_k)^T d_k \geq \sigma\alpha_k \|F(x_k + \alpha_k d_k)\| \|d_k\|^2. \tag{2.9}$$

Step 4. Set $z_k = x_k + \alpha_k d_k$. If $z_k \in \Omega$ and $\|F(z_k)\| \leq Tol$, stop. Else compute

$$x_{k+1} = P_\Omega[x_k - \zeta_k F(z_k)]$$

where

$$\zeta_k = \frac{F(z_k)^T(x_k - z_k)}{\|F(z_k)\|^2}.$$

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

Lemma 2.4. *Suppose assumptions (K_1)-(K_3) holds, $\{x_k\}$ and $\{z_k\}$ defined by Algorithm 1. Then*

$$\alpha_k \geq \max\left\{\gamma, \frac{\rho\delta\|F(x_k)\|}{(L + \sigma\|F(x_k + \alpha_k d_k)\|)\|d_k\|^2}\right\}. \tag{2.11}$$

Proof. By the line search (2.9), if $\alpha_k \neq \gamma$, the $\alpha_k^{'} = \alpha_k \rho^{-1}$ does not satisfy (2.9), that is

$$-F(x_k + \alpha_k^{'} d_k)^T d_k < \sigma\alpha_k^{'} \|F(x_k + \alpha_k^{'} d_k)\| \|d_k\|^2.$$

Now from (2.2) and assumption (K_2), we have

$$\begin{aligned}\delta\|F(x_k)\|^2 &\leq -F(x_k)^T d_k \\ &= (F(x_k + \alpha_k^{'} d_k) - F(x_k))^T d_k - F(x_k + \alpha_k^{'} d_k)^T d_k \\ &\leq \alpha_k^{'}(L + \sigma\|F(x_k + \alpha_k d_k)\|)\|d_k\|^2.\end{aligned}$$

The desired result is obtained after solving for $\alpha_k^{'}$. ∎

Lemma 2.5. *Suppose assumptions (K_1)-(K_3) holds, then $\{x_k\}$ and $\{z_k\}$ defined by Algorithm 1 are bounded. In addition, we have*

$$\lim_{k\to\infty} \|x_k - z_k\| = 0 \tag{2.12}$$

and

$$\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0. \tag{2.13}$$

Proof. We will start by showing that the sequences $\{x_k\}$ and $\{z_k\}$ are bounded. Suppose $\bar{x} \in \overline{\Omega}$, then by monotonicity of F, we get

$$F(z_k)^T(x_k - \bar{x}) \geq F(z_k)^T(x_k - z_k). \tag{2.14}$$

Also by definition of z_k and the line search (2.9), we have

$$F(z_k)^T(x_k - z_k) \geq \sigma\alpha_k^2\|F(z_k)\|\|d_k\|^2 \geq 0. \tag{2.15}$$

So, we have

$$
\begin{aligned}
\|x_{k+1}-\bar{x}\|^2 &= \|P_\Omega[x_k-\zeta_k F(z_k)]-\bar{x}\|^2 \le \|x_k-\zeta_k F(z_k)-\bar{x}\|^2 \\
&= \|x_k-\bar{x}\|^2 - 2\zeta_k F(z_k)^T(x_k-\bar{x}) + \|\zeta F(z_k)\|^2 \\
&\le \|x_k-\bar{x}\|^2 - 2\zeta_k F(z_k)^T(x_k-z_k) + \|\zeta F(z_k)\|^2 \\
&= \|x_k-\bar{x}\|^2 - \left(\frac{F(z_k)^T(x_k-z_k)}{\|F(z_k)\|}\right)^2 \\
&\le \|x_k-\bar{x}\|^2
\end{aligned} \tag{2.16}
$$

Thus the sequence $\{\|x_k-\bar{x}\|\}$ is decreasing and convergent, and hence $\{x_k\}$ is bounded. Furthermore, from equation (2.16), we have

$$\|x_{k+1}-\bar{x}\|^2 \le \|x_k-\bar{x}\|^2, \tag{2.17}$$

and we can conclude recursively that

$$\|x_k-\bar{x}\|^2 \le \|x_0-\bar{x}\|^2, \quad \forall k \ge 0.$$

Then from Assumption (G_2), we obtain

$$\|F(x_k)\| = \|F(x_k)-F(\bar{x})\| \le L\|x_k-\bar{x}\| \le L\|x_0-\bar{x}\|.$$

If we let $L\|x_0-\bar{x}\| = M$, then the sequence $\{F(x_k)\}$ is bounded, that is,

$$\|F(x_k)\| \le M, \quad \forall k \ge 0. \tag{2.18}$$

By the definition of z_k, equation (2.15), monotonicity of F and the Cauchy-Schwatz inequality, we get

$$\sigma\|x_k-z_k\| = \frac{\sigma\|\alpha_k d_k\|^2}{\|x_k-z_k\|} \le \frac{F(z_k)^T(x_k-z_k)}{\|x_k-z_k\|} \le \frac{F(z_k)^T(x_k-z_k)}{\|x_k-z_k\|} \le \|F(x_k)\|. \tag{2.19}$$

The boundedness of the sequence $\{x_k\}$ together with equation (2.18)-(2.19), implies the sequence $\{z_k\}$ is bounded.

Since $\{z_k\}$ is bounded, then for any $\bar{x} \in \Omega$, the sequence $\{z_k-\bar{x}\}$ is also bounded, that is, there exists a positive constant $\nu > 0$ such that

$$\|z_k-\bar{x}\| \le \nu.$$

This together with Assumption (G_2) yields

$$\|F(z_k)\| = \|F(z_k)-F(\bar{x})\| \le L\|z_k-\bar{x}\| \le L\nu.$$

Therefore, using equation (2.16), we have

$$\frac{\sigma^2}{(L\nu)^2}\|x_k-z_k\|^4 \le \|x_k-\bar{x}\|^2 - \|x_{k+1}-\bar{x}\|^2,$$

which implies

$$\frac{\sigma^2}{(L\nu)^2}\sum_{k=0}^{\infty}\|x_k - z_k\|^4 \le \sum_{k=0}^{\infty}(\|x_k - \bar{x}\|^2 - \|x_{k+1} - \bar{x}\|^2) \le \|x_0 - \bar{x}\| < \infty. \tag{2.20}$$

Equation (2.20) implies

$$\lim_{k\to\infty}\|x_k - z_k\| = 0.$$

However, using Equation 2.1, the definition of ζ_k and the Cauchy-Schwartz inequality, we have

$$\begin{aligned}\|x_{k+1} - x_k\| &= \|P_\Omega[x_k - \zeta_k F(z_k)] - x_k\| \\ &\le \|x_k - \zeta_k F(z_k) - x_k\| \\ &= \|\zeta_k F(z_k)\| \\ &= \|x_k - z_k\|,\end{aligned} \tag{2.21}$$

which yields

$$\lim_{k\to\infty}\|x_{k+1} - x_k\| = 0.$$

■

Equation (2.12) and definition of z_k implies that

$$\lim_{k\to\infty}\alpha_k\|d_k\| = 0. \tag{2.22}$$

Theorem 2.6. *Suppose that assumptions (K_1)-(K_3) hold and let the sequence $\{x_k\}$ be generated by Algorithm 1, then*

$$\liminf_{k\to\infty}\|F(x_k)\| = 0, \tag{2.23}$$

Proof. Suppose by contradiction that (2.23) is not true, then there exist $r_0 > 0$ such that $\forall k \ge 0$

$$\|F(x_k)\| \ge r_0. \tag{2.24}$$

This combined with (2.2) yields

$$\|d_k\| \ge \delta r_0 \ \forall k \ge 0. \tag{2.25}$$

Next we will show that the directions M3TCD1, M3TCD2 and M3TCD3 are bounded.
M3TCD1: Using (2.3), (2.4), (2.5), (2.10), (2.18) and (2.24), we have

$$\begin{aligned} \|d_k\| &= \left\| -F(x_k) + \frac{\|F(x_k)\|^2}{-d_{k-1}^T F(x_{k-1})} w_{k-1} - \frac{F(x_k)^T w_{k-1}}{-d_{k-1}^T F(x_{k-1})} F(x_k) \right\| \\ &\leq \|F(x_k)\| + 2\frac{\|F(x_k)\|^2 \|w_{k-1}\|}{\delta \|F(x_{k-1})\|^2} \\ &= \|F(x_k)\| + 2\frac{\|F(x_k)\|^2 \alpha_{k-1} \|d_{k-1}\|}{\delta \|F(x_{k-1})\|^2} \\ &\leq M + \frac{2M^2}{\delta r_0^2} \alpha_{k-1} \|d_{k-1}\|. \end{aligned} \tag{2.26}$$

Equation (2.22) implies that for all $\epsilon_0 > 0$ there exist k_0 such that $\alpha_{k-1}\|d_{k-1}\| < \epsilon_0$ for all $k > k_0$. So choosing $\epsilon_0 = r_0^2$ and $\kappa = \max\{\|d_0\|, \|d_1\|, \cdots, \|d_{k_0}\|, M_1\}$ where $M_1 = M(1 + \frac{2M}{\delta})$. Therefore $\|d_k\| \leq \kappa$.
M3TCD2 direction: Also by (2.3), (2.4), (2.7), (2.10), (2.18) and (2.24), we have

$$\begin{aligned} \|d_k\| &= \left\| -F(x_k) + \frac{\|F(x_k)\|^2}{-d_{k-1}^T F(x_{k-1})} w_{k-1} - \frac{\|F(x_k)\|^2 \|w_{k-1}\|^2}{(d_{k-1}^T F(x_{k-1}))^2} F(x_k) \right\| \\ &\leq \|F(x_k)\| + \frac{\|F(x_k)\|^2 \|w_{k-1}\|}{\delta \|F(x_{k-1})\|^2} + \frac{\|F(x_k)\|^3 \|w_{k-1}\|^2}{\delta^2 \|F(x_{k-1})\|^4} \\ &= \|F(x_k)\| + \frac{\|F(x_k)\|^2 \alpha_{k-1} \|d_{k-1}\|}{\delta \|F(x_{k-1})\|^2} + \frac{\|F(x_k)\|^3 (\alpha_{k-1} \|d_{k-1}\|)^2}{\delta^2 \|F(x_{k-1})\|^4} \\ &\leq M + \frac{M^2}{\delta r_0^2} \alpha_{k-1} \|d_{k-1}\| + \frac{M^3 (\alpha_{k-1} \|d_{k-1}\|)^2}{\delta^2 r_0^4}. \end{aligned} \tag{2.27}$$

In a similar way as above, letting $\kappa = \max\{\|d_0\|, \|d_1\|, \cdots, \|d_{k_0}\|, M_1\}$ where $M_1 = M(1 + \frac{M}{\delta} + (\frac{M}{\delta})^2)$, we get that $\|d_k\| \leq \kappa$.

M3TCD3 direction: Again by (2.3), (2.4), (2.8), (2.10), (2.18) and (2.24), we have

$$\begin{aligned}\|d_k\| &= \left\| -F(x_k) + \frac{\|F(x_k)\|^2}{-d_{k-1}^T F(x_{k-1})} w_{k-1} - \left(\frac{F(x_k)^T w_{k-1}}{-d_{k-1}^T F(x_{k-1})} + \frac{\|F(x_k)\|^2}{(d_{k-1}^T F(x_{k-1}))^2} \right) F(x_k) \right\| \\ &\leq \|F(x_k)\| + 2\frac{\|F(x_k)\|^2 \|w_{k-1}\|}{\delta \|F(x_{k-1})\|^2} + \frac{\|F(x_k)\|^3}{\delta^2 \|F(x_{k-1})\|^4} \\ &= \|F(x_k)\| + 2\frac{\|F(x_k)\|^2 \alpha_{k-1}\|d_{k-1}\|}{\delta \|F(x_{k-1})\|^2} + \frac{\|F(x_k)\|^3}{\delta^2 \|F(x_{k-1})\|^4} \\ &\leq M + \frac{M^2}{\delta r_0^2}\alpha_{k-1}\|d_{k-1}\| + \frac{M^3}{\delta^2 r_0^4}.\end{aligned} \tag{2.28}$$

Using same argument and letting $\kappa = \max\{\|d_0\|, \|d_1\|, \cdots, \|d_{k_0}\|, M_1\}$ where $M_1 = M(1 + \frac{2M}{\delta} + \frac{M^2}{\delta^2 r_0^4})$. Hence $\|d_k\| \leq \kappa$.
Now multiplying both sides of (2.11) with $\|d_k\|$, we have

$$\begin{aligned}\alpha_k \|d_k\| &\geq \max\left\{ \gamma, \frac{\rho\delta \|F(x_k)\|}{(L + \sigma\|F(x_k + \alpha_k d_k)\|)\|d_k\|^2} \right\} \|d_k\| \\ &\geq \max\left\{ \gamma\delta r_0 \ , \ \frac{\rho\delta r_0^2}{L(1+\sigma\nu)\kappa} \right\}.\end{aligned}$$

The above relation contradicts (2.22) and therefore (2.23) must hold. ∎

3. Numerical Experiments

This section investigates the numerical performance of the proposed algorithms with other conjugate gradient algorithms.
We tested the following algorithms:

CGD: the algorithm proposed by Xiao and Zhu [23]

PCG: the algorithm proposed by Liu and Li [24]

M3TCD1: Algorithm 1 with the choice of λ_k using (2.5)

M3TCD2: Algorithm 1 with the choice of λ_k using (2.7)

M3TCD3: Algorithm 1 with the choice of λ_k using (2.8)

All algorithms were coded in MATLAB using a windows 10 operating system of 2.4GHz Intel(R) Core(TM) i3-7100U CPU with 8GB RAM. The experiments were carried out on eight benchmark test problems using seven initial points with dimension ranging from

n =5000 to 100000. Note that one of the initial points was randomly chosen. For the implementation of M3TCD1, M3TCD2 and M3TCD3, we choose the following parameters: $\sigma = 10^{-4}$, $\rho = 0.9$ and $\gamma = 1$. We implemented CGD and PCG as in [23, 24].

The chosen stopping condition was

$$\|F_k\| \leq 10^{-6}$$

The algorithm is also terminated if the iteration exceeds 1000. To this end, we give a list of the test problem utilized in this experiment.

Problem 1. This problem is the Exponential function [25] with constraint set $C = R^n_+$, that is,

$$f_1(x) = e^{x_1} - 1,$$
$$f_i(x) = e^{x_i} + x_i - 1, \text{ for } i = 2, 3, ..., n.$$

Problem 2. Modified Logarithmic function [26] with constraint set $C = \{x \in R^n : \sum_{i=1}^n x_i \leq n, x_i > -1, i = 1, 2, \dots, n\}$, that is,

$$f_i(x) = \ln(x_i + 1) - \frac{x_i}{n}, \; i = 2, 3, ..., n.$$

Problem 3. The Nonsmooth Function [27] with constraint set $C = R^n_+$.

$$f_i(x) = 2x_i - \sin|x_i|, \; i = 1, 2, 3, ..., n.$$

Problem 4. [28] The function with constraint set $C = R^n_+$, that is,

$$f_i(x) = \min\left(\min(|x_i|, x_i^2), \max(|x_i|, x_i^3)\right) \text{ for } i = 2, 3, ..., n$$

Problem 5. The Strictly convex function [29], with constraint set $C = R^n_+$, that is,

$$f_i(x) = e^{x_i} - 1, \; i = 2, 3, \cdots, n.$$

Problem 6. Tridiagonal Exponential function [30] with constraint set $C = R^n_+$, that is,

$$f_1(x) = x_1 - e^{\cos(h(x_1 + x_2))},$$
$$f_i(x) = x_i - e^{\cos(h(x_{i-1} + x_i + x_{i+1}))}, \text{ for } 2 \leq i \leq n-1,$$
$$f_n(x) = x_n - e^{\cos(h(x_{n-1} + x_n))}, \text{ where } h = \frac{1}{n+1}$$

Problem 7. Nonsmooth function [31] with with constraint set $C = \{x \in R^n : \sum_{i=1}^n x_i \leq n, x_i \geq -1, \; 1 \leq i \leq n\}$.

$$f_i(x) = x_i - \sin|x_i - 1|, \; i = 2, 3, \cdots, n.$$

Problem 8. The Trig exp function [25] with constraint set $C = R^n_+$, that is,

$$f_1(x) = 3x_1^3 + 2x_2 - 5 + \sin(x_1 - x_2)\sin(x_1 + x_2)$$
$$f_i(x) = 3x_i^3 + 2x_{i+1} - 5 + \sin(x_i - x_{i+1})\sin(x_i + x_{i+1}) + 4x_i - x_{i-1}e^{x_{i-1} - x_i} - 3 \text{ for } i = 2, 3, .$$
$$f_n(x) = x_{n-1}e^{x_{n-1} - x_n} - 4x_n - 3, \text{ where h} = \frac{1}{n+1}.$$

In addition, we employ the performance profile developed in [32] where the performance metric is based on number of iterations, CPU time (in seconds) and number of function evaluations which are used to obtain Figures 1-3. These figures present a wealth of information including efficiency and robustness of the methods. For instance, Fig. 1 shows that the the three proposed method (M3TCD1, M3TCD2, M3TCD3) exhibits the best overall performance since it illustrates the best probability of being the optimal solver, outperforming CGD and PCG.

Analytically, the performance profile with respect to number of function evaluations shows that M3TCD2 solves and wins 51% of the test problems with the least number of function evaluations while M3TCD1, M3TCD3, CGD and PCG solves and wins about 28%, 19%, 8% and 19% of the test problems, respectively. On the overall, it is worth noticing that one our proposed method (M3TCD2) outperform CGD and PCG which implies that the proposed method is computationally efficient.

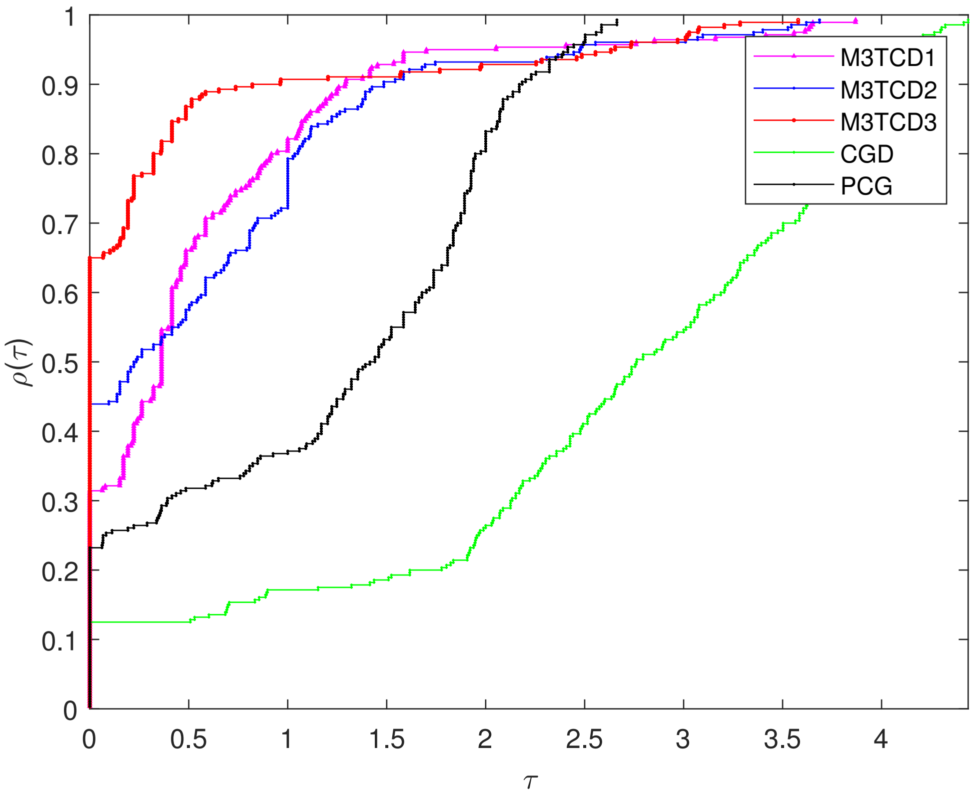

FIGURE 1. Performance based on the number of iterations.

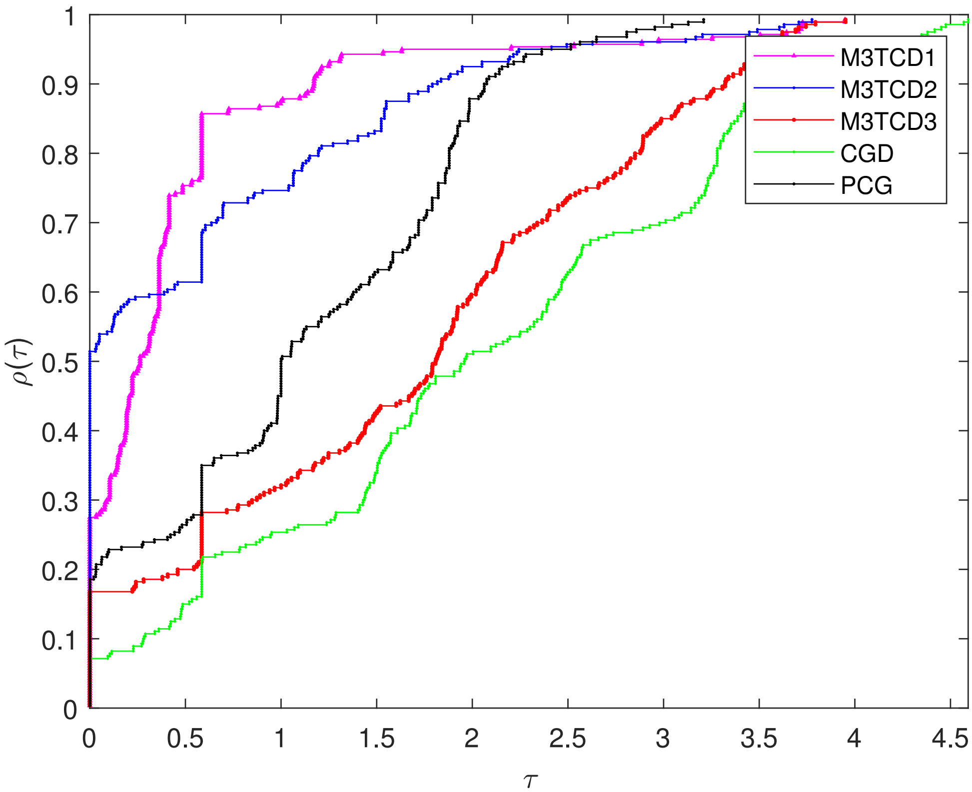

FIGURE 2. Performance based function evaluation.

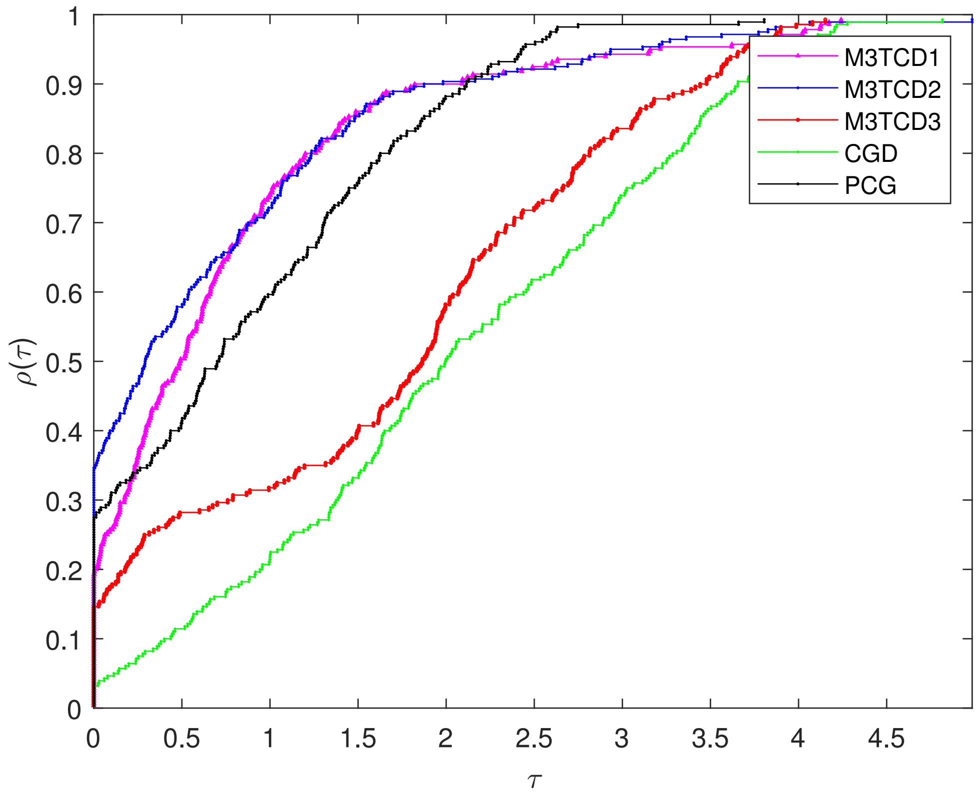

Figure 3. Performance based on CPU time.

4. Conclusions

In this article, we modified the well known conjugate descent (CD) direction and proposed three distinct spectral conjugate gradient algorithms for solving (1.1). The modification was achieved by adding the term $-\lambda_k F(x_k)$ to the CD direction making it three-term. Using three different approaches as in [33], we obtained three distinct definition of λ_k corresponding to the three directions **M3TCD1**, **M3TCD2** and **M3TCD3** respectively. The proposed directions are bounded and satisfy the sufficient descent property. The convergence of the proposed algorithms was established under suitable assumptions. Finally, we give some numerical experiments to show the efficiency of the algorithms compared with two existing algorithms namely; CGD and PCG.

Acknowledgements

This project was supported by Theoretical and Computational Science (TaCS) Center under Computational and Applied Science for Smart research Innovation Cluster (CLASSIC), Faculty of Science, KMUTT. The first author was supported by the Petchra

Pra Jom Klao Doctoral Scholarship, Academic for Ph.D. Program at KMUTT (Grant No.26/2560).

References

[1] S.V. M. Iusem, N Alfredo, Newton-type methods with generalized distances for constrained optimization, Optimization 41(3)(1997) 257–278.

[2] M. AT Figueiredo, R.D. Nowak, S.J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, IEEE Journal of selected topics in signal processing 1(4)(2007) 586–597.

[3] T.L. Magnanti, G. Perakis, Solving variational inequality and fixed point problems by line searches and potential optimization, Mathematical programming 101(3)(2004) 435–461.

[4] M. Al-Baali, Y. Narushima, H. Yabe, A family of three-term conjugate gradient methods with sufficient descent property for unconstrained optimization, Computational Optimization and Applications 60(1)(2015) 89–110.

[5] W.W. Hager H. Zhang, A survey of nonlinear conjugate gradient methods, Pacific journal of Optimization, 2(1)(2006) 35–58.

[6] Y. Narushima, A smoothing conjugate gradient method for solving systems of nonsmooth equations, Applied Mathematics and Computation, 219(16)(2013) 8646–8655.

[7] Y. Narushima, H. Yabe, J. Ford, A three-term conjugate gradient method with sufficient descent property for unconstrained optimization, SIAM Journal on Optimization, 21(1)(2011) 212–230.

[8] L. Zhang, W. Zhou, Spectral gradient projection method for solving nonlinear monotone equations, Journal of Computational and Applied Mathematics, 196(2)(2006) 478–484.

[9] H. Mohammad, A.B. Abubakar, A positive spectral gradient-like method for nonlinear monotone equations, Bulletin of Computational and Applied Mathematics 5(1)(2017) 99–115.

[10] A.B. Abubakar P. Kumam, An improved three-term derivative-free method for solving nonlinear equations, Computational and Applied Mathematics, 37(5)(2018) 6760–6773.

[11] M.V. Solodov B.F. Svaiter, A globally convergent inexact newton method for systems of monotone equations, In Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods (1998) 355–369.

[12] M. Ahookhosh, K. Amini, S, Bahrami, Two derivative-free projection approaches for systems of large-scale nonlinear monotone equations, Numerical Algorithms, 64(1)(2013) 21–42.

[13] Zoltan Papp and Sanja Rapajiȁ, Fr type methods for systems of large-scale nonlinear monotone equations, Applied Mathematics and Computation, 269(2015) 816 – 823.

[14] D. Feng, M. Sun, X. Wang, A family of conjugate gradient methods for large-scale nonlinear equations, Journal of Inequalities and Applications, 2017 (2017) 236.

[15] A.B. Abubakar, P. Kumam, H. Mohammad, A.M. Awwal, K. Sitthithakerngkiet, A modified fletcher–reeves conjugate gradient method for monotone nonlinear equations with some applications, Mathematics 7(8)(2019) 745.

[16] A.B. Abubakar, P. Kumam, A.M. Awwal, A descent dai-liao projection method for convex constrained nonlinear monotone equations with applications, Thai Journal of Mathematics 17(1)(2018).
[17] A.B. Abubakar, P. Kumam, A.M. Awwal, P. Thounthong, A modified self-adaptive conjugate gradient method for solving convex constrained monotone nonlinear equations for signal recovery problems, Mathematics 7(8)(2019) 693.
[18] A.A. Muhammed, P. Kumam, A.B. Abubakar, A. Wakili, N. Pakkaranang, A new hybrid spectral gradient projection method for monotone system of nonlinear equations with convex constraints, Thai Journal of Mathematics, 16(4)(2018).
[19] A.B. Abubakar, P. Kumam, A descent dai-liao conjugate gradient method for nonlinear equations, Numerical Algorithms, 81(1)(2019) 197–210.
[20] A.M. Awwal, P. Kumam, A.B. Abubakar, A modified conjugate gradient method for monotone nonlinear equations with convex constraints, Applied Numerical Mathematics 145(2019) 507 – 520.
[21] A.M. Awwal, P. Kumam, A.B. Abubakar, Spectral modified polak–ribière–polyak projection conjugate gradient method for solving monotone systems of nonlinear equations, Applied Mathematics and Computation 362(2019) 124514.
[22] A.B. Abubakar, P. Kumam, H. Mohammad, A.M. Awwal, An efficient conjugate gradient method for convex constrained monotone nonlinear equations with applications, Mathematics, 7(9)(2019) 767.
[23] Y. Xiao, H. Zhu, A conjugate gradient method to solve convex constrained monotone equations with applications in compressive sensing, Journal of Mathematical Analysis and Applications, 405(1)(2013) 310–319.
[24] J.K. Liu, S.J. Li, A projection method for convex constrained monotone nonlinear equations with applications, Computers & Mathematics with Applications 70(10)(2015) 2442–2453.
[25] W.L. Cruz, J. Martínez, M. Raydan, Spectral residual method without gradient information for solving large-scale nonlinear systems of equations, Mathematics of Computation 75(255)(2006) 1429–1448.
[26] W.L. Cruz, J.M. Martínez, M. Raydan, Spectral residual method without gradient information for solving large-scale nonlinear systems of equations, Mathematics of Computation 75(255)(2006) 1429–1448.
[27] W. Zhou, D. Li, Limited memory bfgs method for nonlinear monotone equations, Journal of Computational Mathematics 25(1)(2007).
[28] W.L. Cruz, A spectral algorithm for large-scale systems of nonlinear monotone equations, Numerical Algorithms 76(4)(2017) 1109–1130.
[29] C. Wang, Y. Wang, and C. Xu, A projection method for a system of nonlinear monotone equations with convex constraints, Mathematical Methods of Operations Research 66(1)(2007) 33–46.
[30] Y. Bing, G. Lin, An efficient implementation of merrills method for sparse or partially separable systems of nonlinear equations, SIAM Journal on Optimization, 1(2)(1991) 206–221.
[31] G. Yu, S. Niu, J. Ma, Multivariate spectral gradient projection method for nonlinear monotone equations with convex constraints, Journal of Industrial and Management Optimization 9(1)(2013) 117–129.
[32] E.D. Dolan, J.J. Moré, Benchmarking optimization software with performance profiles, Mathematical Programming 91(2)(2002) 201–213.

[33] Z. Papp, S. Rapajić, Fr type methods for systems of large-scale nonlinear monotone equations, Applied Mathematics and Computation (2015) 269:816–823.