Ann. Acad. Rom. Sci.
Ser. Math. Appl.
ISSN 2066-5997 Vol. 16, No. 1/2024

THE BEST SPECTRAL CORRECTION OF DMDY CONJUGATE GRADIENT METHOD*

Khaoula Meansri† Noureddine Benrabia ‡ Mourad Ghiat§
Hamza Guebbai¶ Imane Hafaidia‖

Abstract

In this paper, we present an enhanced spectral correction for the DMDY conjugate gradient method. Our approach involves integrating a third term and determining its parameter through three different approaches. The primary objective is to ensure the sufficient descent condition. By applying the Wolfe line search conditions, we establish the global convergence property for all three proposed algorithms. Numerical tests conclusively demonstrate the superior efficiency of our algorithms, surpassing that of existing methods.

*Accepted for publication on January, 5-th, 2023

†khaoulameansri@gmail.com, k.meansri@univ-soukahras.dz Laboratory Informatics and Mathematics (LIM), University of Mohamed Cherif Messaadia Souk Ahras, B.P.1553, Souk Ahras, 41000, Algeria; Paper written with financial support of: Not supported

‡noureddinebenrabia@yahoo.fr University of Mohamed Cherif Messaadia Souk Ahras, B.P.1553, Souk Ahras, 41000, Algeria, Laboratory of Applied Mathematics and Modeling

§mourad.ghi24@gmail.com, ghiat.mourad@univ-guelma.dz Laboratory of Applied Mathematics and Modeling, University of 8 Mai 1945 Guelma, B.P.401, Guelma, 24000, Algeria

¶guebaihamza@yahoo.fr, guebbai.hamza@univ-guelma.dz Laboratory of Applied Mathematics and Modeling, University of 8 Mai 1945 Guelma, B.P.401, Guelma, 24000, Algeria

‖hafaidia.imane@yahoo.com Constantine 1 - Frères Mentouri University, Laboratory of Applied Mathematics and Modeling, University of 8 Mai 1945 Guelma, B.P.401, Guelma, 24000, Algeria

MSC: 90C06, 90C30, 65K05

keywords: Spectral correction, Conjugate gradient methods, Sufficient descent condition and global convergence, Numerical tests.

1 Introduction

The purpose of utilizing Nonlinear Conjugate Gradient (NCG) methods, extensively studied in [1, 2], is to minimize unconstrained optimization problems formulated in the following manner:

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

where, $n \in \mathbb{N}^*$ is supposed to be very large and $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable function.
To solve the problem (1) starting form an initial point $x_0 \in \mathbb{R}^n$, the NCG method generates a sequence of points $\{x_k\}_{k\in\mathbb{N}}$ defined by

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

where, the stepsizes $\alpha_k \in \mathbb{R}_+^*$ are determined by some line search and are very important for global convergence of conjugate gradient methods. In our work, we use line search to satisfying the Wolfe conditions [3, 4]

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \rho\alpha_k g_k^t d_k, \tag{3}$$

$$g_k^t d_{k-1} \geq \sigma g_{k-1}^t d_{k-1}, \tag{4}$$

where, $0 < \rho < \sigma < 1$, and $\delta < \sigma < 1$. $d_k \in \mathbb{R}^n$ are search directions given by

$$\begin{cases} d_0 = -g_0, \\ d_k = -g_k + \beta_k d_{k-1}, \quad k \geq 1, \end{cases} \tag{5}$$

where, $g_k = g(x_k) = \nabla f(x_k)$ is the gradient of the function f in the point x_k and $\beta_k \in \mathbb{R}^*$ is a scalar called the conjugate gradient parameter. In the following table, we recall some famous formulas of this parameter

The Formula	Authors
$\beta_k^{HS} = \dfrac{g_k^t y_{k-1}}{d_{k-1}^t y_{k-1}}$	Hestenes-Stiefel (1952), [5].
$\beta_k^{FR} = \dfrac{\|g_k\|^2}{\|g_{k-1}\|^2}$	Fletcher Reeves(1964), [6].
$\beta_k^{PRP} = \dfrac{g_k^t y_{k-1}}{\|g_{k-1}\|^2}$	Polak-Ribiére-Polyak (1969), [7, 8].
$\beta_k^{CD} = \dfrac{\|g_k\|^2}{-g_{k-1}^t d_{k-1}}$	Conjugate Descent (1987), [9].
$\beta_k^{LS} = \dfrac{g_k^t y_{k-1}}{-g_{k-1}^t d_{k-1}}$	Liu -Storey (1991), [10].
$\beta_k^{DY} = \dfrac{\|g_k\|^2}{d_{k-1}^t y_{k-1}}$	Dai-Yuan (1999), [11].

Where, $y_{k-1} = g_k - g_{k-1}$ and $\|.\|$ is the Euclidean norm.
To enhance the aforementioned classical NCG methods, several alternative approaches have been suggested. Among them, a notable method proposed by Hager and Zhang [12] is a modified version of the HS method known as the CG-DESCENT method. This method introduces improvements in the following aspects:

$$\beta_k^{N+} = max\left\{\beta_k^N, \eta_k\right\}, \tag{6}$$

where,

$$\beta_k^N = \beta_k^{HS} - 2\frac{\|y_{k-1}\|^2}{(y_{k-1}^t d_{k-1})^2} g_k^t d_{k-1}, \quad \eta_k = \frac{-1}{\|d_k\|^2 \, min\left\{\|g_k\|, \eta\right\}}$$

and $\eta > 0$ is a constant. The modification demonstrates that the resulting descent vector exhibits enhanced efficiency, particularly when employed in conjunction with an inexact line search.

Another modification was introduced by Hailin Liu, Sui Sun, and Xiaoyong Li [13], who adapted the classical DY method to obtain

$$\beta_k^{MDY} = \frac{\|g_k\|^2}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}}, \quad \mu > 1. \tag{7}$$

This modification demonstrates that the obtained descent direction is more efficient, leading to a more effective and convergent method compared to the classical DY method.

Numerous researchers have put forth various methods employing different techniques to address the problem (1). Notably, the three-term conjugate gradient method (TTCG) has emerged as a reliable and efficient alternative to classical conjugate gradient algorithms. This superiority has been demonstrated in several papers [14, 15]. It should be highlighted that the most efficient formulation of the TTCG method is as follows:

$$d_k = -g_k + \beta_k d_k - \theta_k g_k \tag{8}$$

The various three-term conjugate gradient algorithms are distinguished by their parameter choices, such as β_k and θ_k. For instance, Zhang et al. [16, 17] proposed the three-term PRP conjugate gradient method (TTPRP) and the three-term FR conjugate gradient method (TTFR). These modifications ensure that a descent direction is obtained and when combined with Armijo line search, they exhibit global convergence. Building upon these concepts, Zhang, Li, Weijun Zhou and Donghui Li [18] introduced the three-term HS conjugate gradient method (TTHS), which guarantees a descent direction and global convergence when using standard Wolfe line search.
Zoltan and Sanja [19] modified the classical FR conjugate gradient direction by incorporating the term $\theta_k g_k$, where θ_k is defined in three different ways (refer to [19]). Similarly, Habibu Abdullahi and al [20] modified the classical DY conjugate gradient direction into a three-term conjugate gradient algorithm by adding the term $\nu_k g_k$, where ν_k is defined in three distinct ways as well.
On the other hand, the equation (8) can be expressed in the following alternative form:

$$d_k = -(1 + \theta_k) g_k + \beta_k d_{k-1}.$$

This alternative form represents the spectral conjugate gradient (SCG) method. The SCG method, known for its straightforward implementation, is highly effective and efficient in solving the problem (1). Notable references supporting its efficacy include [21, 22, 23].
The main objective of this paper is to introduce a novel spectral method that exhibits improved numerical performance for large-scale optimization problems. Our goal is to determine the appropriate values for the parameters θ_k and β_k in order to construct an efficient and coherent method. In terms of efficiency, we choose the conjugate parameter β_k from (7) to be the conjugate parameter in our spectral method. Building upon the idea presented by Zoltan and Sanja in [19], we propose a modification to the classical descent

of the DMDY conjugate gradient method (1) by defining the search direction

$$d_k = -(1+\psi_k)g_k + \beta_k^{MDY} d_{k-1}, \quad k \geq 1.$$

In order to improve the efficiency and robustness of the spectral conjugate gradient (SCG) method, our proposed approach involves defining the search direction using a spectral parameter $(1+\psi_k)$. Our primary goal is to determine an optimal and effective selection of the parameter ψ_k that will result in a more efficient and reliable SCG method. By carefully tuning this parameter, we aim to enhance the overall performance of the SCG method in terms of both computational efficiency and the ability to handle complex optimization problems effectively.

2 New corrections

In this section, we present a novel spectral conjugate gradient algorithm, which serves as an enhancement of the DMDY conjugate gradient algorithm proposed by Hailin Liu, Sui Sun and Xiaoyong Li [13]. The primary objective of this algorithm is to ensure the fulfillment of the sufficient descent condition. The algorithm, denoted as (2), involves the computation of the direction d_k as follows:

$$d_k = -(1+\psi_k)g_k + \beta_k^{MDY} d_{k-1}, \quad k \geq 1. \tag{9}$$

To account for the coefficient ψ_k in three different forms, denoted as $\psi_{k,1}$, $\psi_{k,2}$ and $\psi_{k,3}$, we establish three distinct conjugate gradient directions. These directions are named as MDMDY1, MDMDY2 and MDMDY3 respectively.

Three different forms of ψ_k

The search direction is defined by the formula (9), where the parameter β_k^{MDY} given by (7) with $\mu > 1$.

- **MDMDY1 direction**

  We have

  $$d_k = -(1+\psi_{k,1})g_k + \beta_k^{MDY} d_{k-1}.$$

  By using (7) and multiplying by g_k^t we get

  $$g_k^t d_k = -(1+\psi_{k,1})\|g_k\|^2 + \frac{\|g_k\|^2}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}} g_k^t d_{k-1}.$$

For the sufficient descent direction we find

$$\psi_{k,1} = \frac{g_k^t d_{k-1}}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}}, \qquad \forall k \in \mathbb{N}. \tag{10}$$

So we get

$$g_k^t d_k = -\left\| g_k \right\|^2. \tag{11}$$

- **MDMDY2 direction**

We have

$$d_k = -(1+\psi_{k,2})g_k + \beta_k^{MDY} d_{k-1}.$$

By using (7) and multiplying by g_k^t we get

$$g_k^t d_k = -(1+\psi_{k,2}) \left\| g_k \right\|^2 + \frac{\left\| g_k \right\|^2}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}} g_k^t d_{k-1}.$$

If we find

$$\vartheta_k = \frac{g_k^t \left(\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}\right)}{\sqrt{2}} \quad \textit{and} \quad \varphi_k = \sqrt{2} \left\| g_k \right\|^2 d_{k-1}.$$

By the formula

$$\vartheta_k^t \varphi_k \leq \frac{1}{2}(\left\| \vartheta_k \right\|^2 + \left\| \varphi_k \right\|^2).$$

Therefore

$$\begin{aligned} g_k^t d_k \quad & \leq -(1+\psi_{k,2}) \left\| g_k \right\|^2 + \tfrac{1}{2\left(\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}\right)^2} \\ & \left(\tfrac{1}{2} \left\| g_k \right\|^2 \left(\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}\right)^2 + 2 \left\| g_k \right\|^4 \left\| d_{k-1} \right\|^2\right), \\ & = -\tfrac{3}{4} \left\| g_k \right\|^2 + \tfrac{\left\| g_k \right\|^4 \left\| d_{k-1} \right\|^2}{\left(\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}\right)^2} - \psi_{k,2} \left\| g_k \right\|^2. \end{aligned}$$

For the sufficient descent direction we find

$$\psi_{k,2} = \frac{\left\| g_k \right\|^2 \left\| d_{k-1} \right\|^2}{\left(\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}\right)^2}, \qquad \forall k \in \mathbb{N}. \tag{12}$$

So we get

$$g_k^t d_k \leq -\frac{3}{4} \left\| g_k \right\|^2. \tag{13}$$

- **MDMDY3 direction**
  We define the third $\psi_{k,3}$ by two parts such that, the first part is $\psi_{k,1}$ and the second part is properly chosen such that the sufficient descent direction is satisfied, therefore

$$\psi_{k,3} = \frac{g_k^t d_{k-1}}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}} + \frac{\|g_k\|^2}{\left(\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}\right)^2}, \quad \forall k \in \mathbb{N}. \tag{14}$$

  So

$$g_k^t d_k \leq -\|g_k\|^2. \tag{15}$$

We will now provide a proof that for $\mu > 1$, the three different directions MDMDY1, MDMDY2 and MDMDY3 satisfy the sufficient descent condition. This is stated formally in the following theorem:

Theorem 1 *If $\mu > 1$, then the direction MDMDY1, MDMDY2 and MDMDY3 are a sufficient descent direction for all $k \in \mathbb{N}$, i.e.*

$$g_k^t d_k \leq -C\|g_k\|^2, \quad \forall k \geqslant 0. \tag{16}$$

Proof 1 *For $k = 0$, we have for all the three directions $d_0 = -g_0$, then*

$$g_0^t d_0 = -\|g_0\|^2, \quad for \quad C = 1.$$

For $k \geq 1$, considering the conditions (11) and (15), we can establish that the directions MDMDY1 and MDMDY3 satisfy the sufficient descent condition with $C = 1$. Additionally, based on condition (13), we can conclude that the direction MDMDY2 also satisfies the sufficient descent condition with $C = \frac{3}{4}$.

Now, we will introduce our three different algorithms, each consisting of the following steps:

Algorithms 1 *(MDMDY1, MDMDY2 and MDMDY3)*
Step0: *Choosing the initial point $x_0 \in \mathbb{R}^n$ and the parameter $\mu > 1$, $\varepsilon > 0$ and $d_0 = -g_0$, such as $k = 0$.*
Step1:

- *If $\|g_k\| \leq \varepsilon$ stop.*
- *Else go to step2.*

Step2: *Calculate step length* α_k *with Wolfe line search condition (3), (4) for MDMDY1, MDMDY2 and MDMDY3.*
Step3: *Calculate the direction (9) with* β_k^{MDY} *formula (7) and*

$$\psi_{k,1} = \frac{g_k^t d_{k-1}}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}},$$

$$\psi_{k,2} = \frac{\|g_k\|^2 \|d_{k-1}\|^2}{\left(\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}\right)^2},$$

$$\psi_{k,3} = \frac{g_k^t d_{k-1}}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}} + \frac{\|g_k\|^2}{\left(\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}\right)^2},$$

formulas for $MDMDY1$, $MDMDY2$ *and* $MDMDY3$ *respectively.*
Step4: *Set* $x_{k+1} = x_k + \alpha_k d_k$.
Step5: *Set* $k = k + 1$*, then go to step1.*

3 Global convergence result

In this section, we present the global convergence analysis for our three different algorithms: MDMDY1, MDMDY2 and MDMDY3. To establish the global convergence, certain basic assumptions are required.

Assumption 1

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. The level set $\mathit{\Gamma} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ is bounded, where $x_0 \in \mathbb{R}^n$ is the starting point of the iteration and f is a continuously differentiable function in a neighborhood $\aleph$ of $\mathit{\Gamma}$.

Namely, there exists a constant $D > 0$, such that

$$\|x\| \leq D, \qquad \forall x \in \aleph. \tag{17}$$

Assumption 2

The gradient $g(x)$ of f is Lipschitz continuous in $\aleph$. Namely, there exists a constant $L > 0$, such that

$$\| g(x_1) - g(x_2) \| \leq L \| x_1 - x_2 \|, \qquad \forall x_1, x_2 \in \aleph. \tag{18}$$

By using Assumption 1 and Assumption 2, we deduce that $\forall x \in \aleph$ there exists a positive constant $\varrho > 0$, such that

$$\| g(x) \| \leq \varrho, \quad \forall x \in \aleph. \tag{19}$$

Also, in order to prove the global convergence of the new methods, we need the following two results.

Lemma 1 *[24] Suppose that the Assumption 1 and Assumption 2 are satisfied. Let the sequence $\{x_k\}_{k\in\mathbb{N}}$ be generated by the three different algorithms (MDMDY1, MDMDY2 and MDMDY3) and $d_k \in \mathbb{R}^n$ satisfied the condition (16). α_k is determined from Wolfe line search (3), (4). If*

$$\sum_{k=0}^{\infty} \frac{1}{\| d_k \|^2} = \infty. \tag{20}$$

Then

$$\liminf_{k\to\infty} \| g_k \| = 0.$$

Lemma 2 *[25] Suppose that the Assumption 1 and Assumption 2 hold and the sequence $\{x_k\}_{k\in\mathbb{N}}$ be generated by the three different algorithms (MDMDY1, MDMDY2 and MDMDY3) and $d_k \in \mathbb{R}^n$ satisfy the condition(16). α_k is determined by Wolfe line search (3), (4). Then*

$$\alpha_{k-1} \geq \frac{(1-\sigma) \mid g_{k-1}^t d_{k-1} \mid}{L \| d_{k-1} \|^2}. \tag{21}$$

Proof 2 *With the Wolfe conditions (3) and (4), we get*

$$d_{k-1}^t (g_k - g_{k-1}) \geq (1-\sigma) \mid g_{k-1}^t d_{k-1} \mid. \tag{22}$$

From condition (18) and using the Cauchy Schwarz inequality, we have

$$d_{k-1}^t (g_k - g_{k-1}) \leq L\alpha_{k-1} \|d_{k-1}\|^2. \tag{23}$$

By condition (22), therefore

$$(1-\sigma) \mid g_{k-1}^t d_{k-1} \mid \leq L\alpha_{k-1} \|d_{k-1}\|^2.$$

So we have proved (21).

This indicates that α_k obtained by our method is different to zero, hence there exists a constant $\gamma > 0$, such that

$$\alpha_k \geq \gamma, \quad \forall k \geq 0. \tag{24}$$

We need also the theorem bellow to prove the global convergence of the three algorithms (MDMDY1, MDMDY2 and MDMDY3).

Theorem 2 *Suppose that the Assumption 1 and Assumption 2 are satisfied and the vector sequence $\{x_k\}_{k\in\mathbb{N}}$ is generated by the three different algorithms (MDMDY1, MDMDY2 and MDMDY3). Then α_k is determined from Wolfe line search (3), (4), then*

$$\lim_{k\to\infty} inf \mid\mid g_k \mid\mid = 0. \tag{25}$$

Proof 3 *We prove by contradiction i.e, assume that there exists $\varepsilon > 0$, such that*

$$\mid\mid g_k \mid\mid > \varepsilon, \quad \forall k \geq 0. \tag{26}$$

We have

$$\begin{aligned} \left|\beta_k^{MDY}\right| &= \left|\frac{\|g_k\|^2}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}}\right|, \\ &\leq \frac{\|g_k\|^2}{d_{k-1}^t y_{k-1}}. \end{aligned} \tag{27}$$

On the other hand, this proof is divided into three parts corresponding to three algorithms (MDMDY1, MDMDY2 and MDMDY3).

Part 1: MDMDY1 direction

We have

$$\begin{aligned} |\psi_{k,1}| &= \left|\frac{g_k^t d_{k-1}}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}}\right|, \\ &\leq \frac{\left|g_k^t d_{k-1}\right|}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}}, \\ &\leq \frac{1}{\mu}. \end{aligned} \tag{28}$$

From (9), we get

$$d_k = -(1 + \psi_{k,1})g_k + \beta_k^{MDY} d_{k-1}.$$

This implies

$$\|d_k\| \leq \|g_k\| + \left|\beta_k^{MDY}\right| \|d_{k-1}\| + |\psi_{k,1}| \|g_k\|.$$

From (27) and (28), we have

$$\|d_k\| \leq \|g_k\| + \frac{\|g_k\|^2}{d_{k-1}^t y_{k-1}} \|d_{k-1}\| + \frac{1}{\mu} \|g_k\|.$$

From (4), (11) and (26), we get

$$\|d_k\| \leq \|g_k\| + \frac{\|g_k\|^2}{(1-\sigma)\varepsilon^2} \|d_{k-1}\| + \frac{1}{\mu} \|g_k\|.$$

By (2), (24), (17) and (19), we have

$$\|d_k\| \leq M_1. \tag{29}$$

Where $M_1 = (1+\frac{1}{\mu})\varrho + \frac{\varrho^2 D}{\gamma(1-\sigma)\varepsilon^2}$.

Part 2: MDMDY2 direction

We have

$$\begin{aligned} |\psi_{k,2}| &= \left| \frac{\|g_k\|^2\|d_{k-1}\|^2}{\left(\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}\right)^2} \right|, \\ &\leq \frac{\|g_k\|^2\|d_{k-1}\|^2}{\left(d_{k-1}^t y_{k-1}\right)^2}. \end{aligned} \tag{30}$$

By (9), we get

$$d_k = -(1+\psi_{k,2})g_k + \beta_k^{MDY} d_{k-1}.$$

This implies

$$\|d_k\| \leq \|g_k\| + \left|\beta_k^{MDY}\right| \|d_{k-1}\| + |\psi_{k,2}| \|g_k\|.$$

From(27) and (30), therefore

$$\|d_k\| \leq \|g_k\| + \frac{\|g_k\|^2}{d_{k-1}^t y_{k-1}} \|d_{k-1}\| + \frac{\|g_k\|^2 \|d_{k-1}\|^2}{\left(d_{k-1}^t y_{k-1}\right)^2} \|g_k\|.$$

By using (4), (13) and (26), we obtain

$$\|d_k\| \leq \|g_k\| + \frac{4\|g_k\|^2}{3(1-\sigma)\varepsilon^2} \|d_{k-1}\| + \frac{16\|g_k\|^2 \|d_{k-1}\|^2}{9(1-\sigma)^2\varepsilon^4} \|g_k\|.$$

From (2), (24), (17) and (19), thus

$$\|d_k\| \leq M_2. \tag{31}$$

Where $M_2 = \varrho + \frac{4\varrho^2 D}{3\gamma(1-\sigma)\varepsilon^2} + \frac{16\varrho^3 D^2}{9\gamma^2(1-\sigma)^2\varepsilon^4}$.

Part 3: MDMDY3 direction

We have

$$\begin{aligned} |\psi_{k,3}| &= \left| \frac{g_k^t d_{k-1}}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}} + \frac{\|g_k\|^2}{\left(\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}\right)^2} \right|, \\ &\leq \frac{\left|g_k^t d_{k-1}\right|}{\mu|d_{k-1}^t g_k| + d_{k-1}^t y_{k-1}} + \frac{\|g_k\|^2}{\left(d_{k-1}^t y_{k-1}\right)^2}, \\ &\leq \frac{1}{\mu} + \frac{\|g_k\|^2}{\left(d_{k-1}^t y_{k-1}\right)^2}. \end{aligned} \tag{32}$$

From (9), we get

$$d_k = -(1 + \psi_{k,3})g_k + \beta_k^{MDY} d_{k-1}.$$

This implies

$$\|d_k\| \leq \|g_k\| + \left|\beta_k^{MDY}\right| \|d_{k-1}\| + |\psi_{k,3}| \|g_k\|.$$

By (27) and (32), we have

$$\|d_k\| \leq \|g_k\| + \frac{\|g_k\|^2}{d_{k-1}^t y_{k-1}} \|d_{k-1}\| + (\frac{1}{\mu} + \frac{\|g_k\|^2}{\left(d_{k-1}^t y_{k-1}\right)^2}) \|g_k\|.$$

From (4), (15) and (26), therefore

$$\|d_k\| \leq \|g_k\| + \frac{\|g_k\|^2}{(1-\sigma)\varepsilon^2} \|d_{k-1}\| + \frac{1}{\mu} \|g_k\| + \frac{\|g_k\|^3}{(1-\sigma)^2 \varepsilon^4}.$$

By using (2), (24), (17) and (19), thus

$$\|d_k\| \leq M_3. \tag{33}$$

With $M_3 = (1 + \frac{1}{\mu})\varrho + \frac{\varrho^2 D}{\gamma(1-\sigma)\varepsilon^2} + \frac{\varrho^3}{(1-\sigma)^2\varepsilon^4}$.
So, by using (29), (31) and (33) and applying (20) this is a contradiction with (26), thus we have proved (25).

4 Numerical results

In this section, we present the results of numerical tests conducted to compare the performance of our three algorithms, namely MDMDY1, MDMDY2 and MDMDY3. The tests were conducted using the strong Wolfe line search conditions with $\rho = 0.0001$ and $\sigma = 0.1$. The parameter μ was set to 1.1 and the three different forms $\psi_{k,1}$, $\psi_{k,2}$ and $\psi_{k,3}$ were employed, as given in (10), (12) and (14) respectively.
The comparison was made against the following three conjugate gradient methods:

- DMDY: defined in (7) with the parameter $\mu = 1.1$.
- TTFR: is presented in [17].
- CG-DESCENT: is presented in [12].

For that we selected 85 unconstrained optimization test problems from [26], this problem was tested for a number of variables:
$n = 2, 10, 20, 25, 100, 200, ..., 10000$. The completion criterion for all algorithms is $\|g_k\|^2 \leq 10^{-7}$ or number of iterations exceeded 50000.
Running on the PC machine ($Intel^R Core^{TM} i3{-}2348M\ CPU$ @ 2.30 GHz, 4.00 *Go RAM*). Our use of performance profiles given by Dolan and Moré [27] to compare the performance according to CPU time, the number of iterations and the number of gradient evaluations. Define S the set of solvers its number is denoted by n_s, and P is the assortment of test issues, with n_p representing the count of test problems. For every problem $p \in P$ and solver $s \in S$, let $\tau_{p,s}$ signify CPU time or the number of iterations or the number of gradient needed to address problem $p \in P$ using solver $s \in S$. Consequently, an assessment of diverse solvers is established on the performance ratio, as follows

$$r_{p,s} = \frac{\tau_{p,s}}{min\{\tau_{p,s} : s \in S\}},$$

Assume a parameter r_M where $r_M \geq r_{p,s}$ holds true for all selected problems and solvers and $r_{p,s} = r_M$, if and only if solver S fails to resolve problem P. The comprehensive assessment of solver performance is subsequently determined by the performance profile function, articulated as follows

$$\rho_s(\tau) = \frac{size\{p \in P : log_2(r_{p,s}) \leq \tau\}}{n_p},$$

here, where τ is greater than or equal to 1, and $size\{p \in P : log_2(r_{p,s}) \leq \tau\}$ is the count of elements in the set $\{p \in P : log_2(r_{p,s}) \leq \tau\}$, then $\rho_s(\tau)$ represents the probability of the solver $s \in S$ that a performance ratio $r_{p,s}$ is within a factor $\tau \in \mathbb{R}$. The ρ_s is the distribution function for the performance ratio. The value of $\rho_s(1)$ represents the probability of the solver outperforming the remaining solvers. In essence, we illustrate, for each method, the proportion P of problems for which the method achieves a time within a certain factor of the best time. The left segment of the figure indicates the percentage of test problems in which a method is the quickest, while the right segment reveals the percentage of test problems successfully addressed by each method. The top curve represents the method that effectively resolved the highest number of problems within a time frame close to the best time.
Figure 1, 2 and **3** represent the performance profile measured by CPU time, the number of iterations and the number of gradient evaluations respectively.

Based on the figures presented, it is evident that our three different algorithms, namely MDMDY1, MDMDY2 and MDMDY3, exhibit superior efficiency in terms of computation time, number of iterations and error reduction. Notably, the $MDMDY1$ method stands out as the most efficient, outperforming the DMDY, TTFR and CG-DESCENT methods. These results confirm the effectiveness of our proposed algorithms and their superiority over existing methods in terms of optimization efficiency and accuracy.

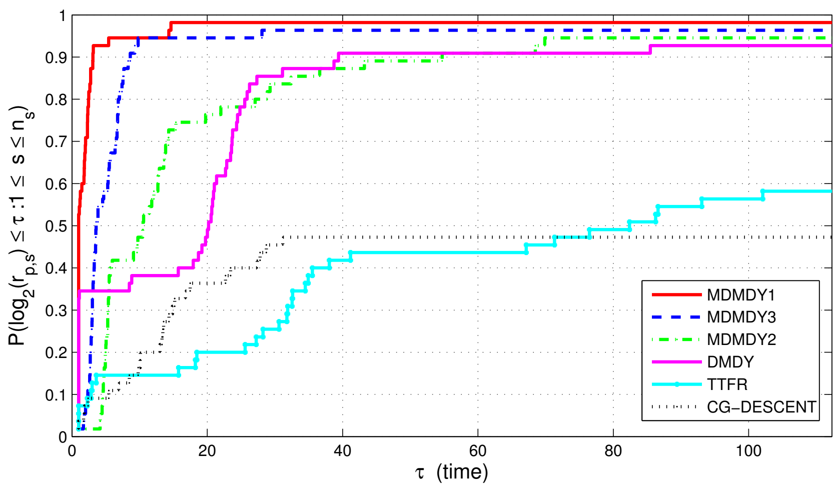

Figure 1: Performance profile for CPU time

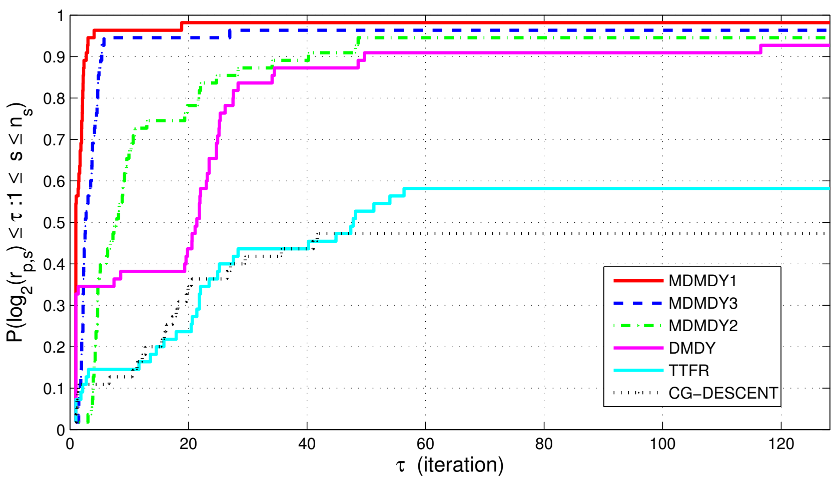

Figure 2: Performance profile for the number of iterations

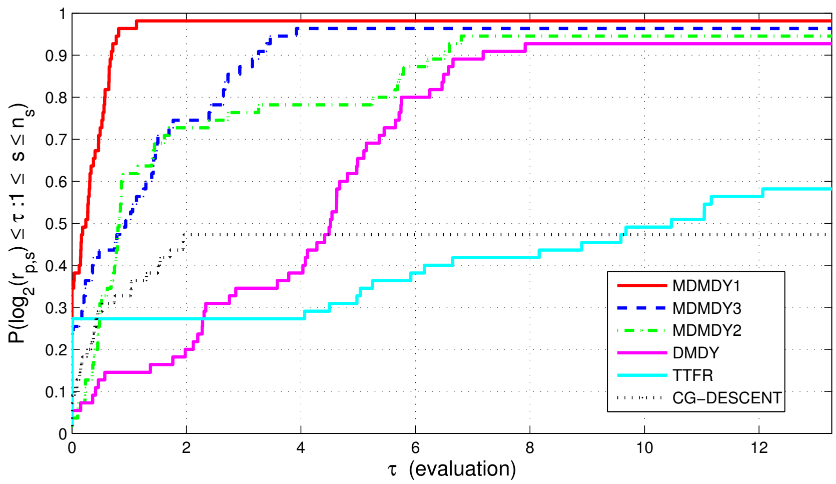

Figure 3: Performance profile for the number of gradient evaluations.

5 Conclusion

In this paper, we have introduced a novel spectral conjugate gradient method that incorporates three different directions based on the DMDY direction. These directions serve as a correction to the classical DY conjugate gradient method. The selection of these directions is made by verifying the sufficient descent condition, which ensures their effectiveness. Moreover, we have established a more efficient global convergence for our proposed method.
To validate the effectiveness of our approach, we conducted numerical tests. The results showed significant improvements in terms of computation time, number of iterations and number of gradient evaluations. Our method outperformed several well-known conjugate gradient methods, further confirming its efficiency and superiority in practical optimization problems.

References

[1] Abdullahi, Habibu and Awasthi, AK and Waziri, Mohammed Yusuf and Halilu, Abubakar Sani. Descent three-term DY-type conjugate gradient methods for constrained monotone equations with application. Computational and Applied Mathematics. 41:1-32, 2022.

[2] Aminifard, Zohre and Babaie-Kafaki, Saman. A modified descent Polak–Ribiére-Polyak conjugate gradient method with global convergence property for nonconvex functions. Calcolo. 56:1-11, 2019.

[3] P. Wolfe. Convergence conditions for ascent methods. SIAM Rev. 11:226-235, 1969.

[4] P. Wolfe. Convergence conditions for ascent methods. II: some corrections. SIAM Rev. 13:185-188, 1971.

[5] Hestenes, M.R., Stiefel, E. Method of conjugate gradient for solving linear equations. J, Res. Nat. Bur. Stand. 49:409-436, 1952.

[6] Fletcher,R.,Reeves,C. Function minimization bu conjugate gradients.Comput.J. 7:149-154, 1964.

[7] Polak, E., Ribiere, G. Note sur la convergence de directions conjugées. Rev. Francaise Informat Recherche Operatinelle, 3e Année. 16:35-43, 1969.

[8] Polyak,B.T. The conjugate gradient method in extreme problems. USSR Comp.Math.Math.Phys. 9:94-112, 1969.

[9] Fletcher, R. Practical method of optimization, vol I: unconstrained optimization, 2nd edn. Wiley, New York, 1997.

[10] Liu,Y.,Storey,C. Efficient generalized conjugate gradient algorithms part1:theory.J. Comput.Appl. Math. 69:17-41, 1992.

[11] Dai, Y., Yuan, Y. A nonlinear conjugate gradient with a strong global convergence properties. SIAM J. Optim. 10:177-182, 2000.

[12] W.W.Hager, H.Zhang. A new conjugate gradient method with guaranteed descent and an efficient line search. SIAMJ. Optim. 16:170-192, 2005.

[13] Liu, Hailin, Sui Sun Cheng, and Xiaoyong Li. A conjugate gradient method with sufficient descent and global convergence for unconstrained nonlinear optimization. Applied Mathematics E-Notes [electronic only]. 11: 139-147, 2011.

[14] Al-Baali, Mehiddin, Yasushi Narushima, and Hiroshi Yabe. A family of three-term conjugate gradient methods with sufficient descent property for unconstrained optimization. Computational Optimization and Applications. 60:89-110, 2015.

[15] Liu, J. K., Y. M. Feng, and L. M. Zou. Some three-term conjugate gradient methods with the inexact line search condition. Calcolo. 55:1-16, 2018.

[16] Zhang, L., Zhou, W. and Li, D. A descent modified Polak-Ribiére-Polyak conjugate gradient method and its global convergence. IMA Journal of Numerical Analysis. 26:629-640, 2006.

[17] Zhang, L., Zhou, W. and Li, D. Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search. Numerische Mathematik. 104:561-572, 2006.

[18] Zhang, Li, Weijun Zhou, and Donghui Li. Some descent three-term conjugate gradient methods and their global convergence. Optimisation Methods and Software. 22.4:697-711, 2007.

[19] Zoltan P, Sanja R. FR type methods for systems of large-scale nonlinear monotone equations. Appl Math Comput. 269: 816-823, 2015.

[20] Abdullahi, Habibu, et al. Descent three-term DY-type conjugate gradient methods for constrained monotone equations with application. Computational and Applied Mathematics. 41.1:32, 2022.

[21] Birgin, Ernesto G., and Jesus Manuel Martinez. A spectral conjugate gradient method for unconstrained optimization. Applied Mathematics and optimization. 43:117-128, 2001.

[22] Liu, J. K., Y. M. Feng, and L. M. Zou. A spectral conjugate gradient method for solving large-scale unconstrained optimization. Computers and Mathematics with Applications. 77.3 :731-739, 2019.

[23] Faramarzi, Parvaneh, and Keyvan Amini. A modified spectral conjugate gradient method with global convergence. Journal of Optimization Theory and Applications 182:667-690, 2019.

[24] Sugiki, K., Narushima, Y., Yabe, H. Globally convergent three-term conjugate gradient methods that use secant conditions and generate descent search directions for unconstrained optimization. J.Optim. Theory Appl. 153.3:733–757, 2012.

[25] J.K. Liu and S.J. Li. New hybrid conjugate gradient method for unconstrained optimization. Appl. Math. Comput. 245:36-43, 2014.

[26] Andrei, Neculai. An unconstrained optimization test functions collection. Adv. Model. Optim 10.1:147-161, 2008.

[27] Dolan, E. D. , Moré, J. J. Benchmarking optimization software with performance profiles. Math. Program. Ser. A. 91:201-213, 2002.