MULTIPLE USE OF BACKTRACKING LINE SEARCH IN UNCONSTRAINED OPTIMIZATION

Branislav Ivanov, Bilall I. Shaini and Predrag S. Stanimirović

Abstract. The class of gradient methods is a very efficient iterative technique for solving unconstrained optimization problems. Motivated by recent modifications of some variants of the SM method, this study proposed two methods that are globally convergent as well as computationally efficient. Each of the methods is globally convergent under the influence of a backtracking line search. Results obtained from the numerical implementation of these methods and performance profiling show that the methods are very competitive with respect to well-known traditional methods.

Keywords: unconstrained optimization; gradient methods; line search.

1. Introduction

The following unconstrained optimization problem

\[(1.1) \min_{x \in \mathbb{R}^n} f(x),\]

is ubiquitous in all areas of science and practical engineering applications. In (1.1), the function \(f : \mathbb{R}^n \to \mathbb{R}\) is uniformly convex (UC) and twice continuously differentiable (TCD).

The most frequent iterations for solving (1.1) is the gradient descent (GD) iterative scheme

\[(1.2) \quad x_{k+1}^{GD} = x_k^{GD} - t_k g_k,\]

where \(t_k > 0\) is the stepsize and \(g_k := \nabla f(x_k)\) corresponds to the gradient of \(f\). The step length \(t_k\) is mainly calculated using the backtracking line search (BLS).

The Newton iterations stabilized by the line search are defined as

\[(1.3) \quad x_{k+1} = x_k - t_k G_k^{-1} g_k,\]
wherein $G_k^{-1}$ means the inverse of the Hessian matrix $G_k := \nabla^2 f(x_k)$. In order to avoid time consuming computation of the Hessian and its inverse, practical numerical methods for solving unconstrained optimization problem are derived from the usage of appropriate approximations $H_k$ of $G_k^{-1}$. The general scheme of quasi-Newton type with line search [16] is given by

$$x_{k+1} = x_k - t_k H_k g_k.$$  

(1.4)

In order to define efficient class of quasi-Newton methods, we use the simplest scalar approximation of the Hessian with respect to known classifications from [5, 8]:

$$B_k := \gamma_k I, \gamma_k > 0,$$

(1.5)

where $I$ is an identity matrix of appropriate order and $\gamma_k > 0$ is a real parameter. The choice (1.5) leads to the iterative prototype

$$x_{k+1} = x_k - \gamma_k^{-1} t_k g_k,$$

(1.6)

where $t_k$ denotes the basic step size and $\gamma_k^{-1}$ is an additional step size which should be defined appropriately. Clearly, the value $\gamma_k^{-1} t_k$ can be considered as a composite step size, so that iterations (1.6) are GD methods. The iterations (1.6) are known as improved gradient descent (IGD) methods.

Andrei in [1, 3] originated so called Accelerated Gradient Descent (AGD) iterations in the form

$$x_{AGD}^{k+1} = x_{AGD}^k - \theta_{AGD} t_k g_k.$$  

(1.7)

The AGD process (1.7) was improved into the Modified AGD (MAGD) method [7] as

$$x_{MAGD}^{k+1} = x_{MAGD}^k - \theta_k (t_k + t_k^2 - t_k^3) g_k.$$  

(1.8)

A few variants of the IGD class (1.6) were proposed in [7, 10, 11, 14, 15]. The $SM$ method belongs to the class IGD methods. It was originated in [14] by the iterative process

$$x_{SM}^{k+1} = x_{SM}^k - t_k \left( \gamma_k^{SM} \right)^{-1} g_k,$$

(1.9)

where $t_k > 0$ is the basic step size and $\gamma_k^{SM} > 0$ is the gain parameter determined as in

$$
\gamma_k^{SM} = 2 \gamma_k \frac{\gamma_k^{SM} \left[ f(x_{SM}^{k+1}) - f(x_{SM}^k) \right] + t_k \| g_k \|_2^2}{t_k^3 \| g_k \|_2^2}.
$$

The ADSS model from [10] is defined as

$$x_{ADSS}^{k+1} = x_{ADSS}^k - \left( t_k \left( \gamma_k^{ADSS} \right)^{-1} + l_k \right) g_k.$$  

(1.10)
where \( t_k \) and \( l_k \) are determined by BLSs. The TADSS method [15] is defined by the iterative rule

\[
x_{k+1}^{TADSS} = x_k^{TADSS} - (t_k \left( \gamma_k^{TADSS} \right)^{-1} - 1) g_k.
\]

The next scheme was proposed as the modified SM (MSM) method in [7]:

\[
x_{k+1}^{MSM} = x_k^{MSM} - (t_k + t_k^2 - t_k^3) \left( \gamma_k^{MSM} \right)^{-1} g_k.
\]

The acceleration parameters in ADD, ADSS, TADSS and MSM methods are summarized in Table 1.1.

| Method | Acceleration parameter \( \gamma_{k+1} \) | Reference |
|--------|---------------------------------|-----------|
| ADD    | \( \gamma_{ADD}^{TADSS} = 2^\frac{2}{\left( (x_{k+1}^{ADD})^2 - f(x_k^{ADD})^2 \right) - (t_k \left( \gamma_k^{ADD} \right)^{-1} - 1) g_k} \) | (2014) [11] |
| ADSS   | \( \gamma_{ADSS}^{TADSS} = 2^\frac{2}{\left( (x_{k+1}^{ADSS})^2 - f(x_k^{ADSS})^2 \right) + \left( (t_k \left( \gamma_k^{ADSS} \right)^{-1} - 1) g_k \right) g_k} \) | (2015) [10] |
| TADSS  | \( \gamma_{TADSS}^{TADSS} = \psi_k = t_k \left( \gamma_k^{TADSS} \right)^{-1} - 1 + 1 \) | (2015) [15] |
| MSM    | \( \gamma_{MSM}^{MSM} = 2^\gamma_k \left( (x_{k+1}^{MSM})^2 - f(x_k^{MSM})^2 \right) + (t_k + t_k^2 - t_k^3) \left( \gamma_k^{MSM} \right)^{-1} g_k \) | (2019) [7] |

The main goal of this research is to study the impact of multiple usage of backtracking line search in modified SM method [7] and practical computational performance of two new methods. Our intention is to propose and investigate improvements of the MSM method. Globally, we investigate possibility to multiple use backtracking line search in the modified MSM method.

Main results of the present study can be highlighted as follows:

1. A novel iterative scheme is proposed using the idea of computing the step parameters \( t_k \), \( t_k^2 \) and \( t_k^3 \) in the MSM method by means of multiple BLS procedures. The resulting iterations will be denoted as TMSM and DMSM.

2. Convergence behavior of the proposed iterations are investigated on appropriate quadratic functions.

3. Numerical experiments compare introduced methods with existing iterations and analyze three main performances: number of iterative steps and function evaluations and CPU time.

The remainder of the paper is developed according to the following hierarchy of sections. Two modifications of the MSM methods, termed as TMSM and DMSM methods, are introduced in Section 2. Section 3. investigates the convergence of the presented TMSM and DMSM methods. In Section 4., we perform a number of numerical experiments and compare main performances of the novel methods with similar available methods. Final remarks are presented in Section 5.
2. Multiple use of backtracking line search in modified SM method

The MSM method is based on the iteration

\begin{equation}
\mathbf{x}^{\text{MSM}}_{k+1} = \mathbf{x}^{\text{MSM}}_k - t^{\text{MSM}}_k \left( \gamma^{\text{MSM}}_k \right)^{-1} \mathbf{g}_k,
\end{equation}

where \( t^{\text{MSM}}_k = t_k + t^3_k - t^2_k \). The leading idea in defining \( t^{\text{MSM}}_k \) arises from the observation \( t_k + t^2_k > t^{\text{MSM}}_k > t_k \), which means that the MSM method proposes a slightly greater step size with respect to the SM iterations. Since \( t_k \) arises from the BLS procedure, which ensures \( t_k \in (0, 1) \), it implies \( t_k \leq t^{\text{MSM}}_k \leq t_k + t^2_k \).

Our intention in current research is to improve behaviour of iterations (2.1) using two or three appropriately defined step-parameters. Following this idea, a method based on triple usage of the BLS in the MSM method is obtained when \( t^2_k \) is substituted with \( l^2_k \) and \( t^3_k \) is substituted with \( j^3_k \) in (2.1), where \( t_k, l_k \) and \( j_k \) are defined by independent LS procedures: the first BLS (Algorithm 1) calculates \( t_k \), another BLS (Algorithm 2) calculates \( l_k \), while the third BLS (Algorithm 3) determines \( j_k \).

Replacing the above changes gives the expression of the TMSM iteration:

\begin{equation}
\mathbf{x}^{\text{TMSM}}_{k+1} = \mathbf{x}^{\text{TMSM}}_k - t^{\text{TMSM}}_k \left( \gamma^{\text{TMSM}}_k \right)^{-1} \mathbf{g}_k,
\end{equation}

where

\begin{equation}
t^{\text{TMSM}}_k = \begin{cases} 
  t_k + l^2_k - j^3_k, & t_k + l^2_k - j^3_k > t_k \\
  t_k, & t_k + l^2_k - j^3_k \leq t_k.
\end{cases}
\end{equation}

The second order Taylor development of \( f(\mathbf{x}^{\text{TMSM}}_{k+1}) \) gives

\begin{equation}
f(\mathbf{x}^{\text{TMSM}}_{k+1}) \approx f(\mathbf{x}^{\text{TMSM}}_k) - t^{\text{TMSM}}_k \left( \gamma^{\text{TMSM}}_k \right)^{-1} \mathbf{g}_k^T \mathbf{g}_k + \frac{1}{2} \left( t^{\text{TMSM}}_k \right)^2 \left( \left( \gamma^{\text{TMSM}}_k \right)^{-1} \mathbf{g}_k \right)^T \nabla^2 f(\xi) \left( \gamma^{\text{TMSM}}_k \right)^{-1} \mathbf{g}_k.
\end{equation}

The parameter \( \xi \) in (2.4) fulfills the condition \( \xi \in [\mathbf{x}^{\text{TMSM}}_k, \mathbf{x}^{\text{TMSM}}_{k+1}] \). One possible choice is

\begin{equation}
\xi = \mathbf{x}^{\text{TMSM}}_k + \delta(\mathbf{x}^{\text{TMSM}}_{k+1} - \mathbf{x}^{\text{TMSM}}_k) = \mathbf{x}^{\text{TMSM}}_k - \varphi t^{\text{TMSM}}_k \left( \gamma^{\text{TMSM}}_k \right)^{-1} \mathbf{g}_k, \quad 0 \leq \varphi \leq 1.
\end{equation}

According to [14], \( \nabla^2 f(\xi) \) is approximated as \( \gamma^{\text{TMSM}}_k \mathbf{I} \). So, (2.4) reduces to

\begin{equation}
f(\mathbf{x}^{\text{TMSM}}_{k+1}) \approx f(\mathbf{x}^{\text{TMSM}}_k) - t^{\text{TMSM}}_k \left( \gamma^{\text{TMSM}}_k \right)^{-1} \| \mathbf{g}_k \|^2 + \frac{1}{2} \left( t^{\text{TMSM}}_k \right)^2 \gamma^{\text{TMSM}}_{k+1} \left( \gamma^{\text{TMSM}}_k \right)^{-2} \| \mathbf{g}_k \|^2.
\end{equation}
Then $\gamma_{k+1}^{TMSM}$ can be obtained from (2.6) as

$\gamma_{k+1}^{TMSM} = 2\gamma_k^{TMSM} \frac{f(x_{k+1}^{TMSM}) - f(x_k^{TMSM}) + t_k^{TMSM} \|g_k\|^2}{(t_k^{TMSM})^2 \|g_k\|^2}$.

The improper situation $\gamma_{k+1}^{TMSM} < 0$ can be resolved by taking $\gamma_{k+1}^{TMSM} = 1$.

The BLS method is implemented in the Algorithm 1 from [14]. Algorithm 1 defines $t_k$ starting from $t = 1$ and subsequently decreases values of $t$ so that it reduces the value of the objective $f$ enough.

**Algorithm 1** The backtracking line search calculates $t_k$.

**Require:** A real function $f(x)$, appropriate search direction $d_k$ at the point $x_k$ and the positive real numbers $0 < \sigma < 0.5$ and $\beta \in (0, 1)$.

1: $t = 1$.
2: While $f(x_k + td_k) > f(x_k) + \sigma t g_k^T d_k$, do $t := t \beta$.
3: Output $t_k := t$.

**Algorithm 2** The second backtracking line search calculates $l_k$.

**Require:** Objective function $f(x)$, the search direction $d_k$ at the point $x_k$ and positive real numbers $0 < \sigma_l < 0.5$ and $\beta_l \in (0, 1)$.

1: $l = 1$.
2: While $f(x_k + ld_k) > f(x_k) + \sigma_l l g_k^T d_k$, take $l := l \beta_l$.
3: Return $l_k = l$.

**Algorithm 3** The third backtracking line search calculates $j_k$.

**Require:** Objective function $f(x)$, the search direction $d_k$ at the point $x_k$ and positive real numbers $0 < \sigma_j < 0.5$ and $\beta_j \in (0, 1)$.

1: $j = 1$.
2: While $f(x_k + jd_k) > f(x_k) + \sigma_j j g_k^T d_k$, take $j := j \beta_j$.
3: Return $j_k = j$.

Finally, the TMSM method is described in Algorithm 4.

It is expectable that the total number of iterations (NoFI) required by the TMSM method will be smaller than the number of iterations of the MSM method, but an increase in the number of function evaluations (NoFE) and the CPU time (CPUT) is expectable. Based on these indicators, we came up with the idea to omit one line search in the TMSM method. This would drastically reduce the CPUT and the NoFE. Following this idea, a method of double use backtracking line search in modified SM method is obtained. In this way, we get a new expression of the DMSM iteration:

$\gamma_{k+1}^{DMSM} = \gamma_k^{DMSM} - t_k^{DMSM} (\gamma_k^{DMSM})^{-1} g_k$. 

(2.8) $x_{k+1}^{DMSM} = x_k^{DMSM} - t_k^{DMSM} (\gamma_k^{DMSM})^{-1} g_k$, 

$\gamma_{k+1}^{TMSM} = 2\gamma_k^{TMSM} \frac{f(x_{k+1}^{TMSM}) - f(x_k^{TMSM}) + t_k^{TMSM} \|g_k\|^2}{(t_k^{TMSM})^2 \|g_k\|^2}$.
**Algorithm 4** Triple use of backtracking line search in the MSM method (the TMSM method)

**Require:** Objective function $f(x)$, initial point $x_0^{TSM} \in \text{dom}(f)$ and parameters $0 < \lambda < 1$, $0 < \nu < 1$.

1. Put $k = 0$, evaluate $f(x_0^{TSM})$, $g_0 = \nabla f(x_0^{TSM})$, and put $\gamma_0^{TSM} = 1$.

2. If $\|g_k\| \leq \lambda$ and $|f(x_{k+1}^{TSM}) - f(x_k^{TSM})| \leq \nu$, STOP; else go to Step 3.

3. (The first backtracking) Compute $t_k^{TSM}$ using Algorithm 1.

4. (The second backtracking) Compute $l_k^{TSM}$ using Algorithm 2.

5. (The third backtracking) Compute $j_k^{TSM}$ using Algorithm 3.

6. Determine $t_k^{TSM}$ using (2.3).

7. Compute $x_{k+1}^{TSM} = x_k^{TSM} - (\gamma_k^{TSM})^{-1}t_k^{TSM}g_k$.

8. Compute $f(x_{k+1}^{TSM})$ and $g_{k+1} = \nabla f(x_{k+1}^{TSM})$.

9. Determine $\gamma_{k+1}^{TSM}$ using (2.7).

10. If $\gamma_{k+1}^{TSM} < 0$, then take $\gamma_{k+1}^{TSM} = 1$.

11. Set $k := k + 1$, go to the step 2.

12. Return $x_{k+1}^{TSM}$ and $f(x_{k+1}^{TSM})$.

where

$$t_k^{DMSM} = \begin{cases} t_k + t_k^2 - j_k^3, & t_k + t_k^2 - j_k^3 > t_k \\ t_k, & t_k + t_k^2 - j_k^3 \leq t_k. \end{cases}$$

In exactly the same way as for the TMSM method, we arrive at

$$\gamma_{k+1}^{DMSM} = 2t_k^{DMSM} \frac{\gamma_k^{DMSM} [f(x_{k+1}^{DMSM}) - f(x_k^{DMSM})] + t_k^{DMSM} \|g_k\|^2}{(t_k^{DMSM})^2 \|g_k\|^2}. \tag{2.10}$$

The difficulty $\gamma_{k+1}^{DMSM} < 0$ can be resolved using $\gamma_{k+1}^{DMSM} = 1$.

The DMSM method is presented in Algorithm 5:
Algorithm 5 Double use backtracking line search in the MSM method (the DMSM method)

**Require:** Function \( f(x) \), chosen initial point \( x_0^{DMSM} \in \text{dom}(f) \) and parameters \( 0 < \lambda < 1, 0 < \nu < 1 \).

1. Put \( k = 0 \), evaluate \( f(x_0^{DMSM}) \), \( g_0 = \nabla f(x_0^{DMSM}) \) and take \( \gamma_0^{DMSM} = 1 \).

2. If \[ \|g_k\| \leq \lambda \quad \text{and} \quad \frac{|f(x_{k+1}^{DMSM}) - f(x_k^{DMSM})|}{1 + |f(x_k^{DMSM})|} \leq \nu, \]
   STOP; else go to Step 3.

3. (The first backtracking) Compute \( t_k^{DMSM} \in (0, 1] \) using Algorithm 1.

4. (The second backtracking) Compute \( j_k \in (0, 1] \) using Algorithm 3.

5. Determine \( t_{DMSM}^k \) using (2.9).

6. Compute \( x_{k+1}^{DMSM} = x_k^{DMSM} - (\gamma_k^{DMSM})^{-1} t_k^{DMSM} g_k \).

7. Compute \( f(x_{k+1}^{DMSM}) \) and \( g_{k+1} = \nabla f(x_{k+1}^{DMSM}) \).

8. Determine the scalar approximation \( \gamma_{k+1}^{DMSM} I \) of the Hessian of \( f \) at the point \( x_{k+1}^{DMSM} \) using (2.10).

9. If \( \gamma_{k+1}^{DMSM} < 0 \), then take \( \gamma_{k+1}^{DMSM} = 1 \).

10. Put \( k := k + 1 \), go to the step 2.

11. Return \( x_{k+1}^{DMSM} \) and \( f(x_{k+1}^{DMSM}) \).

3. Convergence analysis

The content of this section is the convergence analysis of the TMSM and DMSM methods. In the following part, we restate and derive some basic statements which will be used in the convergence analysis of Algorithms 4 and 5. The proofs can be found in [1, 9, 12, 13, 14] and have been omitted:

\( (H_1) \) the function \( f \) is bounded below on \( B_0 = \{ x \in \mathbb{R}^n \mid f(x) \leq f(x_0) \} \);

\( (H_2) \) the gradient \( g \) is Lipschitz continuous in an open convex set \( B \supseteq B_0 \):

\[ \|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in B, \quad L > 0. \]

**Proposition 3.1.** [1, 13] Let \( d_k \) be a descent direction and the gradient \( g_k \) satisfies the Lipschitz condition (3.1). If \( t_k \) is determined by the BLS in Algorithm 1, then

\[ t_k \geq \min \left\{ 1, \frac{\beta(1 - \sigma)}{L} \frac{\|g_T d_k\|}{\|d_k\|^2} \right\}. \]

**Lemma 3.1.** If the function \( f \) is UC and TCD on \( \mathbb{R}^n \) then there exist \( m, M \) such that

\[ 0 < m \leq 1 \leq M, \]
then \( f(x) \) possesses a minimizer \( x^* \) and
\[
m\|y\|^2 \leq y^T \nabla^2 f(x)y \leq M \|y\|^2, \quad \forall \ x, y \in \mathbb{R}^n;
\]
\[
\frac{1}{2} m\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2} M\|x - x^*\|^2, \quad \forall \ x \in \mathbb{R}^n;
\]
\[
m\|x - y\|^2 \leq (g(x) - g(y))^T (x - y) \leq M\|x - y\|^2, \quad \forall \ x, y \in \mathbb{R}^n.
\]

**Lemma 3.2.** [14] The following inequality holds for a TCD and UC function \( f \) and for the IGD sequence \( \{x_k\} \) generated by (1.6):
\[
f(x_k) - f(x_{k+1}) \geq \mu \|g_k\|^2,
\]
with
\[
\mu = \min \left\{ \frac{\sigma}{M}, \frac{\sigma(1 - \sigma)}{L} \beta \right\}.
\]

In further, it is assumed in this section that \( d_k \) is a descent direction. Further, the scalar approximation of Hessian is TCD. Moreover, instead of (3.4) and (3.3) it is sufficient to assume:
\[
m \leq \gamma_k \leq M, \quad 0 < m \leq 1 \leq M, \quad m, M \in \mathbb{R}.
\]
So, all values \( \gamma_k < 0 \) will be replaced by \( \gamma_k = 1 \), while the cases \( \gamma_k > M \) will be resolved by \( \gamma_k = M \).

**Theorem 3.1.** Let (\( H_1 \)) and (\( H_2 \)) and (3.9) be true and the mapping \( f \) is UC. Then the sequence \( \{x_k^{DMSM}\} \) fulfills (3.7)–(3.8).

**Proof.** From (2.8), it can be concluded
\[
x_{k+1}^{DMSM} = x_k^{DMSM} - t_k^{DMSM} (\gamma_k^{DMSM})^{-1} g_k
\]
\[
= x_k^{DMSM} - t_k^{DMSM} \frac{1}{t_k} (\gamma_k^{DMSM})^{-1} g_k
\]
\[
= x_k^{DMSM} + t_k d_k,
\]
where \( d_k = -\frac{t_k^{DMSM}}{t_k} (\gamma_k^{DMSM})^{-1} g_k \).

Based on the stopping condition of the backtracking line search (Algorithm 1), we conclude
\[
f(x_k^{DMSM}) - f(x_{k+1}^{DMSM}) \geq -\sigma t_k g_k^T d_k, \quad \forall \ k \in \mathbb{N}.
\]

In the situation \( t_k < 1 \), by putting expression for \( d_k \) into (3.10), the following inequalities can be derived:
\[
f(x_k^{DMSM}) - f(x_{k+1}^{DMSM}) \geq -\sigma t_k g_k^T d_k
\]
\[
= -\sigma t_k g_k^T \left( -\frac{t_k^{DMSM}}{t_k} (\gamma_k^{DMSM})^{-1} g_k \right)
\]
\[
= \sigma t_k \frac{t_k^{DMSM}}{t_k} (\gamma_k^{DMSM})^{-1} \|g_k\|^2.
\]
Now, from (3.2), it follows that

\begin{equation}
\begin{aligned}
t_k \geq & -\frac{\beta(1-\sigma)}{L} \cdot \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{d}_k\|^2} \\
= & -\frac{\beta(1-\sigma)}{L} \cdot \frac{\mathbf{g}_k^T}{\|\mathbf{g}_k\|^2} \cdot \left( -\frac{t_{DMSM}^{k,1}}{t_k} (\gamma_k^{DMSM})^{-1} \mathbf{g}_k \right)
\end{aligned}
\end{equation}

By applying inequality (3.12) to (3.11), we obtain

\begin{equation}
\begin{aligned}
f(\mathbf{x}_{DMSM}^k) - f(\mathbf{x}_{DMSM}^{k+1}) & \geq \sigma \gamma_k^{DMSM} \frac{t_{k,1}}{t_k} (\gamma_k^{DMSM})^{-1} \|\mathbf{g}_k\|^2 \\
\geq & \sigma \frac{(1-\sigma)\beta}{L} \|\mathbf{g}_k\|^2.
\end{aligned}
\end{equation}

In the case \( t_k = 1 \), based on (3.9) and (3.10) the following inequality holds

\begin{equation}
\begin{aligned}
f(\mathbf{x}_{DMSM}^k) - f(\mathbf{x}_{DMSM}^{k+1}) & \geq -\sigma \mathbf{g}_k^T \mathbf{d}_k \\
= & -\sigma \mathbf{g}_k^T \left( -\frac{t_{DMSM}^{k,1}}{t_k} (\gamma_k^{DMSM})^{-1} \mathbf{g}_k \right) \\
= & \frac{\sigma}{\gamma_k^{DMSM}} \frac{t_{DMSM}^{k,1}}{t_k} \|\mathbf{g}_k\|^2.
\end{aligned}
\end{equation}

According to (2.9), it follows that \( t_{k,1}^{DMSM} \geq t_k \), which implies

\begin{equation}
\begin{aligned}
f(\mathbf{x}_{DMSM}^k) - f(\mathbf{x}_{DMSM}^{k+1}) & \geq \frac{\sigma}{\gamma_k^{DMSM}} \|\mathbf{g}_k\|^2 \\
\geq & \frac{\sigma}{M} \|\mathbf{g}_k\|^2.
\end{aligned}
\end{equation}

Finally, from (3.13) and (3.15) we get (3.8). \( \square \)
Theorem 3.2. Let \((H_1)\) and \((H_2)\) are valid in conjunction with (3.9) and \(f\) be a UC function.

(a) The sequence \(\{x^{DMSM}_k\}\) satisfies \(\lim_{k \to \infty} \|g_k\| = 0\), and \(\{x^{DMSM}_k\}\) converges to \(x^*\).

(b) The sequence \(\{x^{TMSM}_k\}\) satisfies \(\lim_{k \to \infty} \|g_k\| = 0\), and \(\{x^{TMSM}_k\}\) converges to \(x^*\).

Proof. Analogously as the proof of [14, Theorem 4.1]. \(\square\)

Lemma 3.3 confirms the convergence of the DMSM method on the strictly convex quadratic (SCQ) functions

\(f(x) = \frac{1}{2}x^T Ax - b^T x\),

where \(A \in \mathbb{R}^{n \times n}\) is a symmetric positive definite matrix and \(b \in \mathbb{R}^n\). The eigenvalues of \(A\) are ordered as \(\lambda_1 \leq \cdots \leq \lambda_n\).

Lemma 3.3. The DMSM iterations (2.8) applied on a SCQ function \(f\) given by the expression (3.16) satisfy the inequality

\(\lambda_1 \leq \frac{\gamma_{DMSM}^{k+1}}{t_{k+1}} \leq \frac{2\lambda_{\infty}}{\sigma},\ k \in \mathbb{N}\).

Proof. Simple verification gives

\[
\begin{align*}
\frac{1}{2}(x_{DMSM}^{k+1})^T A x_{DMSM}^{k+1} - b^T x_{DMSM}^{k+1} &\geq \frac{1}{2}(x_{DMSM}^{k+1})^T A x_{DMSM}^{k+1} + b^T x_{DMSM}^{k+1} \\
- \frac{1}{2}(x_{DMSM}^{k+1})^T A x_{DMSM}^{k+1} + b^T x_{DMSM}^{k+1} &\geq - \frac{1}{2}(x_{DMSM}^{k+1})^T A x_{DMSM}^{k+1} + b^T x_{DMSM}^{k+1}.
\end{align*}
\]

The substitute of (2.8) in (3.18) gives

\[
\begin{align*}
\frac{1}{2} \left[ x_{DMSM}^{k} - t_{DMSM}^{k} (\gamma_{DMSM}^{k})^{-1} g_k \right]^T \\
	imes A \left[ x_{DMSM}^{k} - t_{DMSM}^{k} (\gamma_{DMSM}^{k})^{-1} g_k \right] \\
- b^T \left[ x_{DMSM}^{k} - t_{DMSM}^{k} (\gamma_{DMSM}^{k})^{-1} g_k \right] \\
- \frac{1}{2} (x_{DMSM}^{k})^T A x_{DMSM}^{k} + b^T x_{DMSM}^{k} \\
= - \frac{1}{2} t_{DMSM}^{k} (\gamma_{DMSM}^{k})^{-1} (x_{DMSM}^{k})^T A g_k \\
- \frac{1}{2} t_{DMSM}^{k} (\gamma_{DMSM}^{k})^{-1} g_k^T A x_{DMSM}^{k} \\
+ \frac{1}{2} t_{DMSM}^{k} (\gamma_{DMSM}^{k})^{-1} g_k^T A g_k \\
+ t_{DMSM}^{k} (\gamma_{DMSM}^{k})^{-1} b^T g_k.
\end{align*}
\]
Since the gradient of the function (3.16) corresponding to DMSM is equal to
\[ g_k = A x_k^{DMSM} - b, \]
one can verify
\[
\begin{align*}
  f(x_{k+1}^{DMSM}) - f(x_k^{DMSM}) &= t_k^{DMSM} (\gamma_k^{DMSM})^{-1} [b^T g_k - (x_k^{DMSM})^T A g_k] \\
  &+ \frac{1}{2} (t_k^{DMSM})^2 (\gamma_k^{DMSM})^{-2} g_k^T A g_k \\
  &= t_k^{DMSM} (\gamma_k^{DMSM})^{-1} [b^T - (x_k^{DMSM})^T A] g_k \\
  &+ \frac{1}{2} (t_k^{DMSM})^2 (\gamma_k^{DMSM})^{-2} g_k^T A g_k \\
  &= t_k^{DMSM} (\gamma_k^{DMSM})^{-1} g_k^T g_k \\
  &+ \frac{1}{2} (t_k^{DMSM})^2 (\gamma_k^{DMSM})^{-2} g_k^T A g_k. 
\end{align*}
\]
\[ \tag{3.21} \]
After substitute (3.21) into (2.10), the parameter \( \gamma_{k+1}^{DMSM} \) becomes
\[
\begin{align*}
  \gamma_{k+1}^{DMSM} &= 2 \gamma_k^{DMSM} \\
  &= \frac{\gamma_k^{DMSM}}{L} \left[ f(x_{k+1}^{DMSM}) - f(x_k^{DMSM}) + t_k^{DMSM} \|g_k\|^2 \right] \\
  &= \frac{g_k^T A g_k}{L} \frac{\|g_k\|^2}{\|g_k\|^2}. 
\end{align*}
\[ \tag{3.22} \]
Therefore, the following inequalities are valid:
\[ \tag{3.23} \lambda_1 \leq \gamma_{k+1}^{DMSM} \leq \lambda_n, \quad k \in \mathbb{N}. \]
The inequality in (3.17) follows from (3.23) in conjunction with \( 0 < t_{k+1} \leq 1 \). In order to verify the right hand side inequality in (3.17), it suffices to observe the upper bound caused by the BLS
\[ t_k \geq \frac{\beta(1 - \sigma)\gamma_k}{L}, \]
which implies
\[ \tag{3.24} \frac{\gamma_{k+1}^{DMSM}}{t_{k+1}} < \frac{L}{\beta(1 - \sigma)} . \]
Using \( g(x) = Ax - b \) in common with the fact that \( A \) symmetric, it follows that
\[ \|g(x) - g(y)\| = \|Ax - Ay\| = \|A(x - y)\| \leq \|A\|\|x - y\| = \lambda_n\|x - y\|. \]
The Lipschitz constant \( L \) in (3.24) can be equal to the largest eigenvalue \( \lambda_n \). Using \( \sigma \in (0, 0.5) \), \( \beta \in (\sigma, 1) \) one obtains
\[ \tag{3.25} \frac{\gamma_{k+1}^{DMSM}}{t_{k+1}} < \frac{L}{\beta(1 - \sigma)} = \frac{\lambda_n}{\beta(1 - \sigma)} < \frac{2\lambda_n}{\sigma}. \]
So, the right inequality in (3.17) is verified. \( \Box \)
Theorem 3.3. Let \( f \) be a SCQ function defined in (3.16). In the case \( \lambda_n < 2\lambda_1 \), the DMSM method (2.8) satisfies

\[
(d_i^{k+1})^2 \leq \delta^2 (d_i^k)^2,
\]

where

\[
\delta = \max \left\{ 1 - \frac{\sigma \lambda_1}{2\lambda_n}, \frac{\lambda_n}{\lambda_1} - 1 \right\}.
\]

In addition,

\[
\lim_{k \to \infty} \|g_k\| = 0.
\]

Proof. Let \( \{v_1, \ldots, v_n\} \) be orthonormal eigenvectors of \( A \). On the basis of (3.20), there exist real quantities \( d_1^k, d_2^k, \ldots, d_n^k \) satisfying

\[
g_k = \sum_{i=1}^{n} d_i^k v_i.
\]

Now, using (2.8) one can simply deduce

\[
g_{k+1} = Ax_{k+1}^{DMSM} - b = A(x_k^{DMSM} - t_k^{DMSM} (\gamma_k^{DMSM})^{-1} g_k) - b = g_k - t_k^{DMSM} (\gamma_k^{DMSM})^{-1} Ag_k = (I - t_k^{DMSM} (\gamma_k^{DMSM})^{-1} A) g_k.
\]

Using the simple linear approximation of \( g_{k+1} \) as in (3.30), we get

\[
g_{k+1} = \sum_{i=1}^{n} d_i^{k+1} v_i = \sum_{i=1}^{n} (1 - t_k^{DMSM} (\gamma_k^{DMSM})^{-1} \lambda_i) d_i^k v_i.
\]

To prove (3.27), it is enough to show that \( \left| 1 - \frac{\lambda_i}{\gamma_k^{DMSM} (t_k^{DMSM})^{-1}} \right| \leq \delta \). Two cases can be observed. First, if \( \lambda_i \leq \frac{\gamma_k^{DMSM}}{t_k^{DMSM}} \), implying (3.17), we can conclude the following:

\[
1 > \frac{\lambda_i}{\gamma_k^{DMSM} (t_k^{DMSM})^{-1}} \geq \frac{\sigma \lambda_1}{2\lambda_n} \implies 1 - \frac{\lambda_i}{\gamma_k^{DMSM} (t_k^{DMSM})^{-1}} \leq 1 - \frac{\sigma \lambda_1}{2\lambda_n} \leq \delta.
\]

Now, let us examine another case \( \frac{\gamma_k^{DMSM}}{t_k^{DMSM}} < \lambda_i \). Since

\[
1 < \frac{\lambda_i}{\gamma_k^{DMSM} (t_k^{DMSM})^{-1}} \leq \frac{\lambda_n}{\lambda_1},
\]
it follows that

\[
(3.34) \quad \left| 1 - \frac{\lambda_i}{\lambda_1} \right| \leq \frac{\lambda_n}{\lambda_1} - 1 \leq \delta.
\]

Now, in order to prove \( \lim_{k \to \infty} \| g_k \| = 0 \), it suffices to use the orthonormality of \( \{v_1, \ldots, v_n\} \) in common with (3.30) and conclude

\[
(3.35) \quad \| g_k \|^2 = \sum_{i=1}^{n} (d_i^k)^2.
\]

Since (3.27) is valid and \( 0 < \delta < 1 \) holds, (3.35) initiates that (3.30).

4. Numerical results

All the considered methods are coded in Matlab R2017a programming language and executed on the notebook with Intel Core i3 2.0 GHz CPU, 8 GB RAM and Windows 10 operating system. The number of iterations (NofI), number of function evaluations (NofFE) and the CPU time (CPUT) are analyzed in numerical experiments.

Numerical testing is based on 24 test functions from [2], where a lot of the problems are taken over from CUTEr collection [4]. For each of tested functions in Tables 4.1, 4.2 and 4.3, 12 numerical testings are performed with 100, 200, 300, 500, 1000, 2000, 3000, 5000, 7000, 8000, 10000 and 15000 unknowns. Tables 4.1, 4.2 and 4.3 arrange summary numerical results for AGD, MAGD, MSM, SM, DMSM and TMSM, tested on 24 functions.

For each of six tested methods (AGD, MAGD, SM, MSM, DMSM and TMSM), the same stopping criteria are used:

\[
\| g_k \| \leq 10^{-6} \quad \text{and} \quad \frac{|f(x_{k+1}) - f(x_k)|}{1 + |f(x_k)|} \leq 10^{-16}.
\]

The BLS parameters for AGD, MAGD, MSM and SM methods are \( \sigma = 0.0001 \) and \( \beta = 0.8 \). The backtracking procedures in the DMSM method are implemented using \( \sigma = 0.0001 \) and \( \beta = 0.8 \) for Algorithm 1 and \( \sigma_j = 0.00015 \) and \( \beta_j = 0.85 \) for Algorithm 3.

The backtracking procedures in the TMSM method are developed using \( \sigma = 0.0001 \) and \( \beta = 0.8 \) for Algorithm 1, \( \sigma_l = 0.0002 \) and \( \beta_l = 0.9 \) for Algorithm 2 and \( \sigma_j = 0.00015 \) and \( \beta_j = 0.85 \) for Algorithm 3.

Table 4.4 contains average values of NofI, the NofFE and the CPUT for all 288 numerical experiments.

Based on the values for NofI given in Table 4.4, it can be concluded that the DMSM and TMSM methods gives superior results with respect to MAGD, AGD, MSM and SM methods.
Table 4.1: Numerical results of the AGD, MAGD, MSM, SM, DMSM and TMSM methods for the NofI.

| Test function                          | MAGD | TMSM | MSM | DMSM | SM  | AGD |
|----------------------------------------|------|------|-----|------|-----|-----|
| Perturbed Quadratic                    | 352325 | 31269 | 34828 | 31386 | 59908 | 353897 |
| Raydan 1                               | 58504 | 30148 | 26046 | 17238 | 14918 | 22620 |
| Diagonal 3                             | 119719 | 6767 | 7030 | 7077 | 12827 | 120416 |
| Generalized Tridiagonal 1              | 647 | 332 | 346 | 350 | 325 | 670 |
| Extended Tridiagonal 1                 | 692219 | 685 | 1370 | 728 | 4206 | 3564 |
| Extended TET                           | 455 | 191 | 156 | 156 | 156 | 443 |
| Diagonal 4                             | 8084 | 96 | 96 | 96 | 96 | 120 |
| Diagonal 5                             | 48 | 72 | 72 | 72 | 72 | 48 |
| Extended Himmelblau                    | 302 | 312 | 260 | 264 | 196 | 396 |
| Perturbed quadratic diagonal           | 1060824 | 36640 | 37454 | 31662 | 44903 | 2542050 |
| Quadratic QF1                          | 362896 | 32099 | 36169 | 33138 | 62927 | 366183 |
| Extended quadratic penalty QP1         | 229 | 338 | 369 | 208 | 271 | 210 |
| Extended quadratic penalty QP2         | 556357 | 1735 | 1674 | 990 | 3489 | 395887 |
| Quadratic QP2                          | 71647 | 31745 | 32727 | 30642 | 64076 | 100286 |
| Extended Tridiagonal 2                 | 1665 | 994 | 959 | 983 | 943 | 1657 |
| ARWHEAD (CUTE)                         | 12834 | 328 | 430 | 502 | 270 | 5667 |
| Almost Perturbed Quadratic             | 354369 | 30790 | 33652 | 32902 | 67890 | 356094 |
| LIARWHD (CUTE)                         | 925138 | 1257 | 3029 | 1726 | 18691 | 1054019 |
| ENGVAL1 (CUTE)                         | 822 | 623 | 461 | 434 | 375 | 743 |
| QUARTC (CUTE)                          | 177 | 302 | 217 | 220 | 290 | 171 |
| Generalized Quartic                    | 229 | 191 | 181 | 186 | 189 | 187 |
| Diagonal 7                             | 159 | 144 | 147 | 111 | 108 | 72 |
| Diagonal 8                             | 154 | 120 | 120 | 109 | 118 | 60 |
| Full Hessian FH3                       | 63 | 63 | 63 | 63 | 63 | 45 |

Performance profiles from [6] are used in comparing the selected methods. As usual, the NofI, NoFfe and CPUT profiles are used. All numerical results are represented in Figures 4.1 and 4.2. Figure 4.1 (left) shows the performances of compared methods related to NofI. Figure 4.1 (right) illustrates the performance of these methods relative to NoFfe. Graphs in Figure 4.2 illustrate the behavior of considered methods with respect to CPUT.

From the results displayed in Tables 4.1, 4.2 and 4.3 and according to graphs in Figures 4.1 and Figure 4.2, the following can be observed.

1. The DMSM and TMSM methods give better results compared to other methods when we compare the number of iterations.

2. The SM, MSM, DMSM and TMSM exhibit better performances than the AGD and MAGD methods.

From Figure 4.1 (left), it is observable that the graph of the DMSM method comes first to the top, which signifies that the DMSM outperforms other considered methods with respect to the NofI.
Table 4.2: Numerical results of the AGD, MAGD, MSM, SM, DMSM and TMSM methods for the NofFE.

| Test function                     | MAGD  | TMSM  | MSM   | DMSM  | SM    | AGD   |
|-----------------------------------|-------|-------|-------|-------|-------|-------|
| Perturbed Quadratic               | 13855 | 45704 | 20010 | 37059 | 33791 | 13916 |
| Raydan 1                          | 12821 | 13059 | 31260 | 32676 | 81412 | 43180 |
| Diagonal 3                        | 12444 | 13130 | 38158 | 80193 | 69906 | 42647 |
| Generalized Tridiagonal 1         | 9057  | 2934  | 1911  | 2061  | 1094  | 9334  |
| Extended Tridiagonal 1            | 20773 | 14797 | 10989 | 9147  | 3621  | 14292 |
| Extended TET                      | 1430  | 1689  | 528   | 948   | 528   | 3794  |
| Diagonal 4                        | 13344 | 2316  | 636   | 1320  | 636   | 1332  |
| Diagonal 5                        | 108   | 300   | 156   | 228   | 156   | 108   |
| Extended Himmelblau               | 5192  | 3636  | 976   | 1908  | 686   | 6897  |
| Perturbed quadratic diagonal      | 38728 | 139974| 341299| 629088| 460028| 91421 |
| Quadratic QF1                     | 13192 | 61661 | 208286| 392426| 352975| 133101 |
| Extended quadratic penalty QP1    | 2939  | 6400  | 2196  | 5121  | 2332  | 2613  |
| Extended quadratic penalty QP2    | 88461 | 44962 | 11491 | 14058 | 25905 | 98520 |
| Quadratic QF2                     | 281096| 642829| 183142| 364257| 359333| 398923 |
| Extended Tridiagonal 2            | 9613  | 9779  | 2866  | 4951  | 2728  | 8166  |
| ARWHEAD (CUTE)                    | 468970| 15416 | 5322  | 8503  | 3919  | 21428 |
| Almost Perturbed Quadratic        | 13936 | 63912 | 194876| 393591| 338797| 140033 |
| LIARWHD (CUTE)                    | 41619197 | 397888 | 27974 | 33271 | 180457 | 4747666 |
| ENGVAL1 (CUTE)                    | 8332  | 10120 | 2285  | 4319  | 2702  | 6882  |
| QUARTC (CUTE)                     | 414   | 414   | 494   | 780   | 640   | 402   |
| Generalized Quartic               | 1244  | 1311  | 493   | 836   | 507   | 849   |
| Diagonal 7                        | 745   | 930   | 504   | 696   | 335   | 333   |
| Diagonal 8                        | 740   | 805   | 383   | 546   | 711   | 304   |
| Full Hessian FH3                  | 1955  | 2160  | 566   | 1263  | 631   | 1352  |

Figure 4.1 (right) confirms that all six methods are able to solve all test cases. Further, the MSM method is superior in 58.33% of all tests with respect to MAGD (41.17%), TMSM(0%), DMSM(4.17%), SM(29.17%) and AGD(16.67%).

Graphs in Figure 4.2 again confirm that all the methods are able to solve test problems, and the MSM is winer in 54.17% of the tests with respect to MAGD (41.17%), TMSM(0%), DMSM(4.17%), SM(37.50%) and AGD(4.17%).

According to individual data arranged in the tables 4.1-4.3, generated average values as well as the presented graphs, the conclusion is that the DMSM method is winer concerning the NofI.

Compared to the previous numerical results obtained during the testing of AGD, MAGD, MSM, SM, DMSM and TMSM methods, in the next test for parameter values in the second and third backtracking line search we take the values that are less than the values in primary backtracking. The aim of this test is to answer the question: Does the choice of higher or lower parameter values in the second and third backtracking line search in relation to the primary backtracking line search...
Table 4.3: Numerical results of the AGD, MAGD, MSM, SM, DMSM and TMSM methods for the CPUT.

| Test function                  | MAGD  | TMSM  | MSM   | DMSM  | SM    | AGD   |
|--------------------------------|-------|-------|-------|-------|-------|-------|
| Perturbed Quadratic            | 6049.53 | 344.172 | 116.281 | 198.328 | 185.641 | 6756.047 |
| Raydan 1                       | 334.266 | 388.156 | 31.906 | 67.344 | 36.078 | 158.359 |
| Diagonal 3                     | 6401.969 | 199.547 | 52.609 | 120.406 | 102.875 | 5527.844 |
| Generalized Tridiagonal 1      | 7.781 | 4.641 | 1.469 | 3.625 | 1.203 | 11.344 |
| Extended Tridiagonal 1         | 853.172 | 26.828 | 29.047 | 17.297 | 90.281 | 55.891 |
| Diagonal 4                     | 16.172 | 0.719 | 0.263 | 0.359 | 0.141 | 0.781 |
| Diagonal 5                     | 0.313 | 0.750 | 0.344 | 0.734 | 0.328 | 0.391 |
| Extended Himmelblau            | 1.031 | 1.094 | 0.297 | 0.703 | 0.188 | 1.953 |
| Perturbed quadratic diagonal   | 22820.172 | 534.750 | 139.625 | 273.188 | 185.266 | 44978.750 |
| Quadratic QF1                  | 846.453 | 258.938 | 81.531 | 168.453 | 138.172 | 12602.563 |
| Extended quadratic penalty QP1  | 0.063 | 2.234 | 1.000 | 3.516 | 0.797 | 1.266 |
| Extended quadratic penalty QP2  | 1872.797 | 12.578 | 3.516 | 8.063 | 6.547 | 3558.734 |
| Quadratic QF2                  | 798.563 | 243.938 | 73.438 | 153.109 | 132.703 | 1582.766 |
| Extended Tridiagonal 2         | 2.531 | 4.938 | 1.047 | 2.375 | 1.031 | 3.719 |
| ARWHEAD (CUTE)                 | 138.000 | 6.422 | 1.969 | 4.609 | 1.359 | 95.641 |
| Almost Perturbed Quadratic      | 7086.563 | 285.563 | 73.047 | 153.891 | 133.516 | 13337.125 |
| LIARWHD (CUTE)                 | 15372.625 | 10.203 | 9.250 | 12.641 | 82.016 | 27221.516 |
| ENGVAL1 (CUTE)                 | 2.641 | 4.328 | 1.047 | 2.375 | 1.031 | 3.719 |
| QUARTC (CUTE)                  | 2.078 | 4.531 | 1.844 | 3.297 | 2.313 | 2.469 |
| Generalized Quartic            | 0.500 | 0.734 | 0.281 | 0.375 | 0.188 | 0.797 |
| Diagonal 7                     | 0.688 | 0.953 | 0.547 | 1.469 | 0.375 | 0.625 |
| Diagonal 8                     | 0.656 | 0.781 | 0.469 | 1.078 | 0.797 | 0.438 |
| Full Hessian FH3               | 1.188 | 1.672 | 0.391 | 1.234 | 0.391 | 1.438 |

Table 4.4: Average numerical outcomes for 24 test functions tested on 12 numerical experiments.

| Average performances          | MAGD  | TMSM  | MSM   | DMSM  | SM    | AGD   |
|-------------------------------|-------|-------|-------|-------|-------|-------|
| Number of iterations          | 182494.42 | 8622.54 | 9064.83 | 7947.21 | 14575.25 | 221896.04 |
| No. of fun.evaluation         | 5885007.25 | 228961.54 | 64424.04 | 11029 8.83 | 93938.63 | 8434868.21 |
| CPU time (sec)                | 3190.98 | 97.51 | 25.90 | 49.99 | 46.00 | 4829.4 |

directly affect the numerical results of DMSM and TMSM methods?

The primary BLS uses the same parameters $\sigma = 0.0001$ and $\beta = 0.8$ as in the first test for AGD, MAGD, MSM and SM methods. The BLS procedures in the DMSM method are implemented using $\sigma = 0.0001$ and $\beta = 0.8$ for Algorithm 1 and $\sigma_j = 0.00005$ and $\beta_j = 0.7$ for Algorithm 3. Also, the BLS in the TMSM method are implemented using $\sigma = 0.0001$ and $\beta = 0.8$ for Algorithm 1, $\sigma_l = 0.00001$ and $\beta_l = 0.6$
for Algorithm 2 and $\sigma_j = 0.00005$ and $\beta_j = 0.7$ for Algorithm 3.

All other conditions (stop criteria and number of variables) remain the same as in the first numerical experiment.

The obtained numerical results are shown in the Tables 4.5, 4.6 and 4.7.

Table 4.8 includes the average values of NofI, the NofFE and the CPUT in a second numerical experiment.

According to the NofI values given in Table 4.8, it can be notified that the DMSM method gives better results and in the second numerical experiment compared to MAGD, AGD, MSM, SM and TMSM methods.

All numerical results from Tables 4.5, 4.6 and 4.7 are represented in Figures 4.3 and 4.4. Figure 4.3 (left) shows the NofI performances of compared methods. Figure 4.3 (right) demonstrates the NofFE profile of these methods. Figure 4.4
Table 4.5: Numerical results of the AGD, MAGD, MSM, SM, DMSM and TMSM methods for the NofI.

| Test function                  | MAGD | MSM | SM  | AGD | TMSM | DMSM |
|-------------------------------|------|-----|-----|-----|------|------|
| Perturbed Quadratic           | 3523 | 3482| 3990| 3538| 3569 | 2848 |
| Raydan 1                      | 5856 | 2604| 1401| 2262| 9801 | 1759 |
| Diagonal 3                    | 1197 | 703 | 1204| 1204| 8372 | 6409 |
| Generalized Tridiagonal 1     | 647  | 346 | 325 | 670 | 342  | 348  |
| Extended Tridiagonal 1        | 6922 | 137| 420 | 356 | 907  | 760  |
| Extended TET                  | 455  | 156 | 156 | 443 | 156  | 156  |
| Diagonal 4                    | 8084 | 96  | 96  | 120 | 96   | 96   |
| Diagonal 5                    | 48   | 72  | 72  | 48  | 72   | 72   |
| Extended Himmelblau           | 302  | 260 | 196 | 396 | 288  | 294  |
| Perturbed quadratic diagonal  | 1060 | 3754| 4490| 2542| 3103 | 3733 |
| Quadratic QF1                 | 3628 | 3616| 271 | 396 | 2658 | 2658 |
| Extended quadratic penalty QP1| 229  | 369 | 271 | 303 | 362  | 362  |
| Extended quadratic penalty QP2| 3565 | 3489| 3958 | 204 | 1908 | 1908 |
| Quadratic QF2                 | 7164 | 3272| 6407| 1002| 3945 | 2865 |
| Extended quadratic exponential EP1| 67 | 100 | 73  | 48  | 107  | 107  |
| Extended Tridiagonal 2        | 1665 | 659 | 543 | 165 | 528  | 615  |
| ARWHEAD (CUTE)                | 1283 | 430 | 270 | 567 | 304  | 281  |
| Almost Perturbed Quadratic    | 3543 | 3365| 6078 | 356 | 2674 | 2674 |
| LIARWHD (CUTE)                | 9251 | 3029| 186 | 1054| 3543 | 3543 |
| ENGVAL1 (CUTE)                | 822  | 461 | 375 | 743 | 418  | 482  |
| QUARTC (CUTE)                 | 177  | 217 | 290 | 171 | 289  | 275  |
| Generalized Quartic           | 229  | 181 | 189 | 187 | 197  | 195  |
| Full Hessian FH3              | 63   | 63  | 45  | 63  | 63   | 63   |
| Diagonal 9                    | 3256 | 1054| 1361| 329 | 1021 | 1122 |

shows the performance CPUT.

Fig. 4.3: Performance profiles based on the NofI (left) and NoFE (right).
Table 4.6: Numerical results of the AGD, MAGD, MSM, SM, DMSM and TMSM methods for the NofFE.

| Test function                     | MAGD  | MSM | SM  | AGD  | TMSM | DMSM |
|-----------------------------------|-------|-----|-----|------|------|------|
| Perturbed Quadratic               | 13855459 | 200106 | 637910 | 13916515 | 423496 | 260678 |
| Raydan 1                          | 1282162  | 311260 | 81412  | 431804  | 124905 | 280011 |
| Diagonal 3                        | 4244404  | 38158  | 69006  | 1264718 | 95962  | 54865  |
| Generalized Tridiagonal 1         | 9057   | 1191  | 1094   | 9334   | 2408   | 2153   |
| Extended Tridiagonal 1            | 2077341 | 10989 | 35621  | 14292  | 13562  | 6800   |
| Extended TET                      | 4130   | 528   | 528    | 3794   | 1080   | 828    |
| Diagonal 4                        | 133440  | 636   | 636    | 1332   | 1284   | 996    |
| Diagonal 5                        | 108    | 156   | 156    | 108    | 300    | 228    |
| Extended Himmelblau               | 5192   | 976   | 668    | 6897   | 2136   | 2418   |
| Perturbed quadratic diagonal      | 38728371 | 341299 | 460028 | 94921578 | 619938 | 529154 |
| Quadratic QF1                     | 13192789 | 208286 | 352975 | 13310016 | 1472273 | 243573 |
| Extended quadratic penalty QP1    | 2039   | 2196   | 2326   | 2613   | 5073   | 3895   |
| Extended quadratic penalty QP2    | 8846145 | 11491  | 25905  | 9852040 | 29847  | 21345  |
| Quadratic QF2                     | 2810965 | 183142 | 333035 | 3998239 | 444580 | 257674 |
| Extended quadratic exponential EFT| 1513   | 894    | 961    | 990    | 2083   | 1617   |
| Extended Tridiagonal 2            | 9613   | 2866   | 2728   | 8166   | 4446   | 4456   |
| ARWHEAD (CUTE)                    | 468970  | 5322   | 3919   | 214284 | 9038   | 6761   |
| Almost Perturbed Quadratic        | 13936462 | 194876 | 338797 | 14003318 | 424470 | 237534 |
| LIARWHD (CUTE)                    | 41619179 | 27974  | 180457 | 14747667 | 22254  | 53306  |
| ENGVAL1 (CUTE)                    | 8332   | 2285   | 2702   | 6882   | 6064   | 4442   |
| QUARTC (CUTE)                     | 414    | 494    | 640    | 402    | 1264   | 990    |
| Generalized Quartic               | 1244   | 493    | 507    | 849    | 1043   | 798    |
| Full Hessian FH3                  | 1955   | 566    | 631    | 1352   | 1152   | 957    |
| Diagonal 9                        | 12984028 | 68189  | 89287  | 13114711 | 131327 | 125119 |

Fig. 4.4: Performance profiles arising from CPUT.
Table 4.7: Numerical results of the AGD, MAGD, MSM, SM, DMSM and TMSM methods for the CPUT.

| Test function                  | MAGD | MSM | SM  | AGD | TMSM | DMSM |
|-------------------------------|------|-----|-----|-----|------|------|
| Perturbed Quadratic           | 6049.531 | 116.281 | 116.281 | 6756.047 | 219.328 | 134.781 |
| Raydan 1                      | 2114.206 | 64.906 | 198.129 | 173.357 | 114.384 | 16.184 |
| Generalized Tridiagonal 1     | 13410.969 | 82.669 | 102.579 | 8277.844 | 129.734 | 86.688 |
| Extended Tridiagonal 1        | 7.781 | 1.469 | 1.203 | 11.344 | 2.989 | 2.969 |
| Extended TET                  | 2.766 | 0.616 | 0.694 | 4.219 | 1.234 | 0.938 |
| Diagonal 4                    | 16.172 | 0.203 | 0.141 | 0.781 | 0.344 | 0.172 |
| Diagonal 5                    | 0.313 | 0.344 | 0.328 | 0.391 | 0.594 | 0.516 |
| Extended Himmelblau           | 1.031 | 0.297 | 0.188 | 1.053 | 0.688 | 0.875 |
| Quadratic QF1                 | 6846.453 | 81.531 | 138.172 | 12602.563 | 173.953 | 91.047 |
| Extended quadratic penalty QP | 1.063 | 1.000 | 0.797 | 1.206 | 2.781 | 1.813 |
| Quadratic QP2                 | 1827.297 | 3.516 | 6.547 | 3558.734 | 8.750 | 2.906 |
| Quadratic exponential EPI     | 768.563 | 73.438 | 132.703 | 1582.766 | 169.266 | 80.141 |
| Extended quadratic exponential EPI | 1.844 | 0.888 | 0.438 | 0.769 | 1.080 | 0.859 |
| ARWHEAD (CUTE)                | 0.844 | 0.688 | 0.438 | 0.769 | 1.080 | 0.859 |
| Almost Perturbed Quadratic    | 138.000 | 1.969 | 1.359 | 25.641 | 2.813 | 2.625 |
| LIARWHD (CUTE)                | 0.844 | 0.688 | 0.438 | 0.769 | 1.080 | 0.859 |
| ENGVALI (CUTE)                | 2.461 | 1.047 | 1.188 | 3.906 | 2.578 | 2.391 |
| QUARTC (CUTE)                 | 2.078 | 1.844 | 2.313 | 2.469 | 1.625 | 3.203 |
| Generalized Quartic           | 0.500 | 0.281 | 0.188 | 0.797 | 1.422 | 0.800 |
| Full Hessian FH3              | 1.188 | 0.391 | 0.391 | 1.438 | 1.063 | 0.891 |
| Diagonal 9                    | 6662.984 | 43.609 | 38.672 | 6353.172 | 61.984 | 114.704 |

Table 4.8: Average numerical results in the second numerical experiment.

| Average performances | MAGD | MSM | SM  | AGD | TMSM | DMSM |
|----------------------|------|-----|-----|-----|------|------|
| Number of iterations | 106051.21 | 9497.04 | 151.36.33 | 235632.88 | 90.58.46 | 8004.88 |
| No. of fun.evaluation | 6426009.58 | 67265.54 | 97642.88 | 8982579.21 | 118332.71 | 87521.54 |
| CPU time (sec)       | 3468.58 | 27.71 | 47.58 | 5094.18 | 53.49 | 40.45 |

In accordance with obtained numerical data generated in the second numerical experiment, we can give an answer to the question, that independently of the choice of parameter values in the second and third backtracking line search, the DMSM iterations has the best results in relation to NoI. Also, if we compare the average results obtained in Tables 4.4 and 4.8, we can see that there is a slight percentage decrease in the average numerical results of the NoFE and CPUT, the DMSM method compared to the MSM method in the second numerical experiment.

5. Conclusion

Multiple usage of the backtracking line search in the modified SM (MSM)
method lead to two improvements of the MSM scheme, denoted as the TMSM and DMSM methods. Proposed iterations are investigated both theoretically and numerically. The linear convergence of the defined model is proved for UC and for a subset of SCQ functions. Numerical experiments confirm that the derived TMSM and DMSM methods outperform the SM, AGD, MAGD and the MSM with respect to the number of iterations. Numerical values arranged in Tables 4.1-4.8 confirm the better performance of presented accelerated gradient descent method. Finally, the obtained TMSM and DMSM methods can be used as a motivation for different possibilities of deriving new efficient schemes for unconstrained optimization.

REFERENCES

1. N. Andrei: An acceleration of gradient descent algorithm with backtracking for unconstrained optimization. Numer. Algor. 42 (2006), 63–73.
2. N. Andrei: An unconstrained optimization test functions collection. Adv. Model. Optim. 10 (2008), 147—161.
3. N. Andrei: Relaxed gradient descent and a new gradient descent methods for unconstrained optimization. https://camo.ici.ro/neculai/newgrad.pdf, Visited October 29, 2020.
4. I. Bongartz, A. R. Conn, N. I. M. Gould and Ph. L. Toint: CUTEr: constrained and unconstrained testing environments. ACM Trans. Math. Softw. 21 (1995), 123-160.
5. C. Brezinski: A classification of quasi-Newton methods. Numer. Algor. 33 (2003), 123–135.
6. E. D. Dolan and J. J. Moré: Benchmarking optimization software with performance profiles. Math. Program. 91 (2002), 201—213.
7. B. Ivanov, P. S. Stanimirović, G. V. Milovanović, S. Djordjević and I. Brajević: Accelerated multiple step-size methods for solving constrained optimization problems. Optimization Methods and Software (2019), https://doi.org/10.1080/10556788.2019.1653868.
8. J. Nocedal and S. J. Wright: Numerical Optimization. Springer-Verlag New York, Inc, 1999.
9. J. M. Ortega and W. C. Rheinboldt: Iterative Solution of Nonlinear Equation in Several Variables. Academic Press, New York, London, 1970.
10. M. J. Petrović: An accelerated Double Step Size method in unconstrained optimization. Applied Math. Comput. 250 (2015), 309–319.
11. M. J. Petrović and P. S. Stanimirović: Accelerated Double Direction method for solving unconstrained optimization problems. Mathematical Problems in Engineering 2014 (2014), Article ID 965104, 8 pages.
12. R. T. Rockafellar: Convex Analysis. Princeton University Press, Princeton, 1970.
13. Z. -J. Shi: Convergence of line search methods for unconstrained optimization. App. Math. Comput. 157 (2004), 393–405.
14. P. S. Stanimirović and M. B. Miladinović: Accelerated gradient descent methods with line search. Numer. Algor. 54 (2010), 503–520.
15. P. S. Stani\v{m}irovi\'c, G. V. Milovanovi\'c and M. J. Petrovi\'c: A transformation of accelerated double step size method for unconstrained optimization. mathematical Problems in Engineering 2015 (2015), Article ID 283679, 8 pages.

16. W. Sun and Y. -X. Yuan: Optimization Theory and Methods: Nonlinear Programming. Springer, Berlin, 2006.

Branislav Ivanov  
University of Belgrade, Technical Faculty in Bor  
Department of Management  
Vojске Jugoslavije 12, 19210 Bor, Serbia  
ivanov.branislav@gmail.com

Bilall I. Shaini  
State University of Tetova  
Rr. e Ilindenit, p.n., Tetovo  
R. Macedonia  
bilall.shaini@unite.edu.mk

Predrag S. Stani\v{m}irovi\'c  
University of Ni\v{s}, Faculty of Sciences and Mathematics  
Department of Computer Science  
Vi\v{s}egradska 33, 18000 Ni\v{s}, Serbia  
pecko@pmf.ni.ac.rs