Hybridization Rule Applied on Accelerated Double Step Size Optimization Scheme

Milena J. Petrović[a]

[a]*University of Priština, Faculty of Sciences and Mathematics, Kosovska Mitrovica, Serbia*

Abstract. A hybrid accelerated model with two step length parameters for solving unconstrained optimization problems is presented. Applied hybridization process involves an efficient three term hybrid method. The accelerated double step size model is taken as guiding operator in this hybridization process. Defined method is convergent on the set of uniformly convex functions as well as on the set on strictly convex quadratics. We display a Dolan Moré performance profiles of derived iteration and of some other comparative hybrid and accelerated methods regarding the number of iterations and the number of function evaluations metrics. Displayed numerical test results confirm that derived model keeps a good properties of its forerunner method and outperform other comparative hybrid accelerated schemes.

1. Introduction

We present an accelerated hybrid gradient descent model with two step length parameters for solving unconstrained optimization problems. Developed method belongs to the class of accelerated gradient descent methods which is introduced in [23]. Also, it is based on three term hybrid relations from [8], and therewith this method can be classified as a hybrid method. In paper [8] Khan proposed an efficient hybrid set of three equations which presents an improved version of Ishikawa's, Mann's and Picard's hybrid models [7, 9, 18].

In [23], the authors presented an accelerated gradient SM method and showed that this iterative model outperforms classical gradient descent GD method, as well as Andrei's accelerated gradient AGD method from [1]. Using the hybridization principle from [8], the authors in [15] derived a hybrid version of the SM method and numerically proved that this method upgrades the SM model. Later on in [11] an initial correction of the HSM iteration was taken. In the same paper, some improved performance characteristics of modified HSM scheme, i.e. the MHSM method, was noticed.

Based on the Khans' hybridization rule and on accelerated gradient method with two search direction defined in [16], the authors in [17] determined hybrid accelerated double direction HADD model.

Still, the most important idea on which the determination of presented method is based on, can be found in [12]. In this paper, an accelerated double step size unconstrained optimization method is presented and denoted as ADSS method. Numerical tests confirmed that the ADSS method has better performance

characteristics then the accelerated gradient SM method from [23]. The method presented in this paper is generated as a hybrid version of the ADSS iteration, where the hybrid rule, with the adequate operator, is defined as proposed in [8].

This paper is organized as follows. In the second Section we give an overview of some relevant accelerated and hybrid optimization methods. The main idea is elaborated in the third Section wherein we define a hybrid accelerated double step size model and determine the accelerated parameter value of so derived iteration. In the fourth Section we prove convergence properties of our hybrid accelerated method on the sets of uniformly convex and strictly convex quadratic functions. Results of numerical experiments and comparative performance analysis are the contents of the last Section.

2. Optimization methods with accelerated and hybrid features

The research in this paper considers optimization models for solving unconstrained optimization problems. General formulation of these methods is given by the expression:

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

where, x_{k+1} is the next iterative value of the objective function f which is to be minimized (or maximized), x_k is the current function value, $t_k > 0$ is iterative step length and d_k is an iterative search direction. For the goal function $f : \mathbb{R}^n \to \mathbb{R}$ we suppose that it is twice differentiable uniformly convex function. Further on we use the following notations for the gradient and for the Hessian at the $k - th$ iterative point of the objective function respectively:

$$g_k = \nabla f(x_k), \quad G_k = \nabla^2 f(x_k). \tag{2}$$

From the expression (1) we can clearly see that the iterative step length value t_k and iterative search direction vector d_k are the two main parameters which indicate the efficiency and convergence properties of a certain optimization model.

The step size value can be calculated by the exact or inexact line search procedures. Practical researches confirm the benefits of the inexact line search techniques when compared to the exact line search procedure. That is so, mostly because, the inexact line search algorithms require less computational time. Among some known inexact line search procedures, such as procedures of Wolfe, Powell, Goldstein, Armijo [3, 6, 21, 24], within this paper we use Armijo's Backtracking line search algorithm to determine the step length value of each iteration. To be more precise, in this research we will use two different Backtracking line search procedures, presented in [12] in order to determine two iterative step length values denoted as t_k and p_k:

Algorithm 1 The backtracking line search starting from $t = 1$. Calculation of the step size t_k

---

Require: Objective function $f(x)$, the direction d_k of the search at the point x_k and numbers
    $0 < \sigma_t < 0.5$ and $\eta_1 \in (\sigma_t, 1)$.
1: Set $t = 1$.
2: While $f(x_k + td_k) > f(x_k) + \sigma_t t g_k^T d_k$ take $t := \eta_1 t$.
3: Return $t_k = t$.

Algorithm 2 The backtracking line search starting from $p = 1$. Calculation of the step size p_k

---

Require: Objective function $f(x)$, the direction d_k of the search at the point x_k and numbers
    $0 < \sigma_p < 0.5$ and $\eta_2 \in (\sigma_p, 1)$.
1: Set $p = 1$.
2: While $f(x_k + pd_k) > f(x_k) + \sigma_p p g_k^T d_k$ take $p := \eta_2 p$.
3: Return $p_k = p$.

Regarding the search direction, the basic models are expressed through the Cauchy's gradient descent direction

$$d_k = -g_k \tag{3}$$

and the Newton's direction generated by the Hessian of the goal function

$$d_k = -G_k^{-1} g_k. \tag{4}$$

There are many search directions induced by these previous two. Herein we mention only few. For example in some conjugate gradient methods the search direction is determined in the next way:

$$l_k = \begin{cases} -g_k & \text{if } k = 0 \\ -g_k + \beta_k l_{k-1} & \text{if } k \geq 1, \end{cases}$$

Here, β_k is a scalar parameter which can be differently defined due to relevant method. In [5], β_k is calculated as

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

Therewith, in [19, 20], the value of the parameter β is obtained as next:

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

During the development of gradient models, accelerated gradient descent methods were segregate as a subclass of the class of gradient methods [23]. The essential fact which characterizes these schemes is determination of the acceleration parameter. Usual way of computing this very important element is through the features of Taylor's expansion taken on the posed scheme. This way of accelerated parameter determination is confirmed as a good choice [13]. We highlight here three accelerated parameter expressions used in efficient accelerated double step size model, i.e. the ADSS method, in hybrid accelerated gradient method, the HSM method, and in hybrid accelerated double direction method, the HADD scheme. These three models are used in this paper as comparative methods.

$$\gamma_{k+1}^{ADSS} = 2 \frac{f(x_{k+1}) - f(x_k) + \left(\alpha_k \gamma_k^{-1} + \beta_k\right) \|g_k\|^2}{\left(\alpha_k \gamma_k^{-1} + \beta_k\right)^2 \|g_k\|^2}, \qquad ([12])$$

$$\gamma_{k+1}^{HSM} = 2\gamma_k \frac{\gamma_k \left[f(x_{k+1}) - f(x_k)\right] + (\alpha_k + 1) t_k \|g_k\|^2}{(\alpha_k + 1)^2 t_k^2 \|g_k\|^2}, \quad ([15])$$

$$\gamma_{k+1}^{HADD} = 2 \frac{f(x_{k+1}) - f(x_k) - \alpha g_k^T \left(t_k^2 d_k - t_k \gamma_k^{-1} g_k\right)}{\alpha^2 t_k^2 \left(t_k d_k - \gamma_k^{-1} g_k\right)^T \left(t_k d_k - \gamma_k^{-1} g_k\right)}, \qquad ([17]).$$

3. Hybridization of accelerated double step size model

A hybrid model, defined as a set of relations, was first introduced by Picard [18]. Thereafter, Mann exposed his hybrid scheme in [9]. Some others hybrid schemes are presented in [7, 8, 11, 15, 17].

In [15] a hybridization over accelerated gradient SM scheme, which is introduced in [23], was explained and defined. In the same paper the authors proved convergence properties of so defined hybrid model which they denoted as the HSM method. Displayed numerical outcomes confirm the improvement towards the starting SM method.

Herein, we use the same idea in order to define a hybrid version of accelerated double step size model defined in [12]. The ADSS method is defined by the next relation:

$$x_{k+1} = x_k - t_k \gamma_k^{-1} g_k - p_k g_k, \tag{7}$$

where x_{k+1} is the next iterative point, x_k is the current one. Parameters t_k and p_k are two step lengths computed by two deferentially initialized Backtracking procedures. Variable $\gamma_k \equiv \gamma_{k+1}^{ADSS}$ stays for an acceleration factor and it is obtained trough the features of Taylor's series exposed on the scheme (7).

Using the hybridization process proposed in [15] and the expression of the ADSS method (7), we get the following set of relations:

$$
\begin{aligned}
&x_1 = x \in \mathbb{R}, \\
&x_{k+1} = Ty_k = y_k - t_k\gamma_k^{-1}g_k - p_kg_k, \\
&y_k = (1-\alpha_k)x_k + \alpha_k Tx_k = (1-\alpha_k)x_k + \alpha_k(x_k - t_k\gamma_k^{-1}g_k - p_kg_k) = x_k - \alpha_k\left(t_k\gamma_k^{-1} + p_k\right)g_k. \quad k \in \mathbb{N}.
\end{aligned}
\tag{8}
$$

In the previous three term relations, $T : \mathbb{C} \to \mathbb{C}$ is the mapping on nonempty convex subset C of a normed space $\mathbb{E}$. Parameter $\{\alpha_k\} \in (0,1)$ is so called correction hybrid parameter. Now, we substitute y_k from the third equation of (8) into the second equation, precisely in Ty_k expression. After this replacement, we obtain

$$
x_{k+1} = x_k - g_k(\alpha_k + 1)(t_k\gamma_k^{-1} + p_k),
\tag{9}
$$

which presents the hybrid version of the *ADSS*. We denote this process as the *HADSS* method. It is obvious that the *HADSS* scheme is accelerated gradient descent scheme with accelerated parameter $\gamma_k \equiv \gamma_{k+1}^{HADSS}$ which is to be derived. Before we calculate an iterative value of the accelerated factor, let us simplify the relation (9) in next way:

$$
x_{k+1} = x_k - g_k\alpha(t_k\gamma_k^{-1} + p_k).
\tag{10}
$$

In (10) we do the following substitution: $\alpha \equiv \alpha_k + 1 \in (1,2) \quad \forall k$. This way, we make practical computations easier and theoretical analysis more concise.

Now, in order to determine the accelerated parameter γ_{k+1}^{HADSS}, we exposed a second order Taylor series of the scheme (10) :

$$
f(x_{k+1}) \approx f(x_k) + \alpha\left(t_k\gamma_k^{-1} + p_k\right)\|g_k\|^2 + \frac{1}{2}\alpha^2\left(t_k\gamma_k^{-1} + p_k\right)\nabla^2 f(\xi)\left(t_k\gamma_k^{-1} + p_k\right)\|g_k\|^2.
\tag{11}
$$

Here, for the parameter ξ the following is valid

$$
\xi \in [x_k, x_{k+1}], \quad \xi = x_k + \kappa(x_{k+1} - x_k) = x_k - \kappa\alpha g_k\left(t_k\gamma_k^{-1} + p_k\right), \quad 0 \le \kappa \le 1.
$$

Instead of using the Hessian of posed iteration, $\nabla^2 f(\xi)$, we put in the expression (11) appropriate scalar matrix approximation $\gamma_{k+1}I$. Then the relation (11) is turned to:

$$
f(x_{k+1}) \approx f(x_k) + \alpha\left(t_k\gamma_k^{-1} + p_k\right)\|g_k\|^2 + \frac{1}{2}\alpha^2\left(t_k\gamma_k^{-1} + p_k\right)^2\nabla^2 f(\xi)\|g_k\|^2.
\tag{12}
$$

Directly, from expression (12) we can calculate the value of the iterative acceleration factor of the HADSS scheme:

$$
\gamma_{k+1}^{HADSS} = \gamma_{k+1} = 2\frac{f(x_{k+1}) - f(x_k) + \alpha\left(t_k\gamma_k^{-1} + p_k\right)\|g_k\|^2}{\alpha^2\left(t_k\gamma_k^{-1} + p_k\right)^2\|g_k\|^2}.
\tag{13}
$$

It's important to consider the Second-Order Necessary Condition and Second-Order Sufficient Condition. Certainly, these two assumptions must be fulfilled. For that purpose we add the positivity condition for the acceleration parameter: $\gamma_{k+1}^{HADSS} > 0$. However, it may happen that iterative value of γ_{k+1}^{HADSS} is a negative one. In this case, we simply choose $\gamma_{k+1}^{HADSS} = 1$. The next iterative point of the HADSS scheme will then be calculated as:

$$
x_{k+2} = x_{k+1} - \alpha g_{k+1}(t_{k+1} + p_{k+1}).
$$

Finally, we restate the Algorithm 3 which describes the main algorithm, termed as the HADSS algorithm.

Algorithm 3 HADSS method

Require: $0 < \rho < 1, \quad 0 < \tau < 1, x_0 \in \mathbb{R}^n$ and $\alpha \in (1, 2)$.
1: Set $k = 0$. For given x_0, take $\gamma_0 = 1$ and calculate $f(x_0)$ and $g(x_0)$.
2: If the condition $\|g_k\| < \epsilon$ is fulfilled, then return x_k, $f(x_k)$ and stop the algorithm, else continue by the next step.
3: Compute the iterative step lengths, t_k and p_k by Algorithms 1 and 2 respectively.
4: Compute x_{k+1} using (10), then calculate $f(x_{k+1})$ and g_{k+1}.
5: Calculate the approximation parameter γ_{k+1} using (13).
6: If $\gamma_{k+1} < 0$ take $\gamma_{k+1} = 1$.
7: Set $k := k + 1$ and go to Step 2.
8: Return x_{k+1} and $f(x_{k+1})$.

---

4. Convergence properties of the HADSS iteration

We analyze herein the convergence properties of derived hybrid accelerated scheme with two iterative step length parameters. In this regard, we consider uniformly convex and strictly convex quadratic functions.

On the set of uniformly convex functions the next statements are valid (related proofs can be found in [10, 22]).

Proposition 4.1. [10, 22] *If the function $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and uniformly convex on $\mathbb{R}^n$ then:*

1. *the function f has a lower bound on $L_0 = \{x \in \mathbb{R}^n | f(x) \le f(x_0)\}$, where $x_0 \in \mathbb{R}^n$ is available;*
2. *the gradient g is Lipschitz continuous in an open convex set B which contains L_0, i.e. there exists $L > 0$ such that*

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad \forall x, y \in B. \tag{14}$$

Lemma 4.1. [10, 22] *Under the assumptions of Proposition 4.1 the real numbers m, M exist therein satisfying*

$$0 < m \le 1 \le M, \tag{15}$$

such that $f(x)$ has an unique minimizer x^* and

$$m\|y\|^2 \le y^T \nabla^2 f(x)y \le M\|y\|^2, \quad \forall x, y \in \mathbb{R}^n; \tag{16}$$

$$\frac{1}{2}m\|x - x^*\|^2 \le f(x) - f(x^*) \le \frac{1}{2}M\|x - x^*\|^2, \quad \forall x \in \mathbb{R}^n; \tag{17}$$

$$m\|x - y\|^2 \le (g(x) - g(y))^T(x - y) \le M\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n. \tag{18}$$

Now, based on the previous two statements we are able to reveal what is the value of the iterative decreasing of the objective function when the HADSS scheme is applied. This is the contents of the following lemma.

Lemma 4.2. *Let the function f be a twice continuously differentiable and uniformly convex on $\mathbb{R}^n$ and let the sequence $\{x_k\}$ be generated by Algorithm 3. Then*

$$f(x_k) - f(x_{k+1}) \ge \mu \left\|g_k\right\|^2, \tag{19}$$

where

$$\mu = \min\left\{\frac{\sigma_t}{M}, \frac{\sigma_t(1 - \sigma_t)}{L}\eta_1, \sigma_p, \frac{\sigma_p(1 - \sigma_p)}{L}\eta_2\right\}. \tag{20}$$

Proof. The proof of this Lemma is similar to the proof of the relevant Lemma from [12] after we substitute parameters σ_α and σ_β by σ_t and σ_p respectively. $\square$

In the next theorem we confirm that the HADSS method is at least linearly convergent on the set of uniformly convex and twice continuously differentiable functions. This theorem can be proved the same way as the adequate theorem (4.1) in [23].

Theorem 4.1. *For the twice continuously differentiable and uniform convex function f on $\mathbb{R}^n$ and the sequence $\{x_k\}$ generated by Algorithm 3 the next holds:*

$$\lim_{k\to\infty}\left\|g_k\right\| = 0. \tag{21}$$

In this regard, the sequence $\{x_k\}$ converges to the optimal solution at least linearly.

Convergence properties of defined HADSS scheme can be proved as well on the set of strictly convex quadratics. Strictly convex quadratic functions are expressed by the following equation

$$f(x) = \frac{1}{2}x^T A x - b^T x. \tag{22}$$

In previous relation (22) A is a real $n \times n$ symmetric positive definite matrix and $b \in \mathbb{R}^n$ is a given vector of real numbers. We denote by λ_1 and λ_n the smallest and the largest eigenvalues of the matrix A respectively.

Lemma 4.3. *Assume that f is the strictly convex quadratic function defined by (22), where $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite matrix. Then, for the smallest and the largest eigenvalues of A, λ_1 and λ_n, the following estimations are true when the hybrid accelerated double step size model (10) is applied:*

$$\frac{1}{\lambda_1} + 1 \geq \alpha\left(t_{k+1}\gamma_{k+1}^{-1} + p_{k+1}\right) \geq \frac{1}{2\lambda_n}\left(\sigma_t + \sigma_p\right). \tag{23}$$

Proof. We start with evaluating the value of function decreasing in two successive iterative points:

$$f(x_{k+1}) - f(x_k) = \frac{1}{2}x_{k+1}^T A x_{k+1} - b^T x_{k+1} - \frac{1}{2}x_k^T A x_k + b^T x_k. \tag{24}$$

Knowing that the gradient of the function (22) is $g_k = A x_k - b$, after applying the relation of the HADSS scheme and the symmetry condition ($g_k^T A g_k = g_k A g_k^T$), we obtain the next calculations:

$$\begin{aligned}
f(x_{k+1}) - f(x_k) &= \frac{1}{2}(x_k - \alpha g_k(t_k\gamma_k^{-1} + p_k))^T A (x_k - \alpha g_k(t_k\gamma_k^{-1} + p_k)) \\
&\quad - b^T(x_k - \alpha g_k(t_k\gamma_k^{-1} + p_k)) - \frac{1}{2}x_k^T A x_k + b^T x_k \\
&= -\alpha t_k\gamma_k^{-1} x_k^T A g_k - \alpha p_k x_k^T A g_k - \alpha^2 t_k p_k \gamma_k^{-1} g_k^T A g_k \\
&\quad + \frac{1}{2}\alpha^2 t_k^2 \gamma_k^{-2} g_k^T A g_k + \frac{1}{2}\alpha^2 p_k^2 g_k^T A g_k + \alpha t_k\gamma_k^{-1} b^T g_k + \alpha p_k b^T g_k \\
&= -\alpha t_k\gamma_k^{-1}(x_k^T A - b^T)g_k - \alpha p_k(x_k^T A - b^T)g_k + \frac{\alpha^2}{2}(t_k^2\gamma_k^{-2} + 2t_k p_k\gamma_k^{-1} + p_k^2)g_k^T A g_k \\
&= -\alpha t_k\gamma_k^{-1}g^T g_k - \alpha p_k g^T g_k + \frac{\alpha^2}{2}(t_k\gamma_k^{-1} + p_k)^2 g^T A g_k \\
&= -\alpha(t_k\gamma_k^{-1} + p_k)g^T g_k + \frac{\alpha^2}{2}(t_k\gamma_k^{-1} + p_k)^2 g^T A g_k.
\end{aligned}$$

Now, we replace derived equality

$$f(x_{k+1}) - f(x_k) = -\alpha(t_k\gamma_k^{-1} + p_k)g^T g_k + \frac{\alpha^2}{2}(t_k\gamma_k^{-1} + p_k)^2 g^T A g_k$$

into the relation (13):

$$\gamma_{k+1} = 2 \frac{\left(-\alpha(t_k\gamma_k^{-1} + p_k)g^T g_k + \frac{\alpha^2}{2}(t_k\gamma_k^{-1} + p_k)^2 g^T A g_k\right) + \alpha\left(t_k\gamma_k^{-1} + p_k\right)g^T g_k}{\alpha^2\left(t_k\gamma_k^{-1} + p_k\right)^2 g^T g_k}.$$

Elementary calculations lead us to

$$\gamma_{k+1} = \frac{\alpha^2\left(t_k\gamma_k^{-1} + p_k\right)^2 g^T A g_k}{\alpha^2\left(t_k\gamma_k^{-1} + p_k\right)^2 g^T g_k} = \frac{g^T A g_k}{g^T g_k}. \tag{25}$$

Last obtained relation proves that γ_{k+1} is the Rayleigh quotient of the real symmetric matrix A at the vector $t_k\gamma_k^{-1} + p_k$. Therewith, we can conclude the following:

$$\lambda_1 \le \gamma_{k+1} \le \lambda_n, \quad k \in \mathbb{N}. \tag{26}$$

Knowing that $0 \le t_k, p_k \le 1$ and considering the previous (26) we prove the left hand side in inequalities (23)

$$\frac{1}{\lambda_1} + 1 \ge \alpha\left(t_{k+1}\gamma_{k+1}^{-1} + p_{k+1}\right).$$

In order to prove the right hand side of (23), we use the inequalities [12, eq. (4.13)] and [23, eq. (4.8)]. Transformed through notation used in this paper, these two inequalities are:

$$t_k > \frac{\eta_1(1 - \sigma_t)\gamma_k}{L}, \tag{27}$$

$$p_k > \frac{\eta_2(1 - \sigma_t)}{L}. \tag{28}$$

We are using the fact that the eigenvalue λ_n of matrix A has the property of Lipschitz constant L, as well. This is truly so since the matrix A is symmetric and $g(x) = A(x) - b$. All of these give:

$$\|g(x) - g(y)\| = \|Ax - Ay\| = \|A(x - y)\| \le \|A\|\|x - y\| = \lambda_n\|x - y\|. \tag{29}$$

To finish the proof of the right hand side of the inequalities (23) now we can apply the following estimations

$$\begin{aligned}
\alpha\left(t_{k+1}\gamma_{k+1}^{-1} + p_{k+1}\right) &\ge \alpha\left(\frac{\eta_1(1 - \sigma_t)}{L}\gamma_k\gamma_k^{-1} + \eta_2(1 - \sigma_p)L\right) \\
&= \frac{\alpha}{L}\left(\eta_1(1 - \sigma_t) + \eta_2(1 - \sigma_p)\right) \\
&> \frac{\alpha}{\lambda_n}(\sigma_t \cdot \frac{1}{2} + \sigma_p \cdot \frac{1}{2}) \\
&= \frac{\alpha}{2\lambda_n}\left(\sigma_t + \sigma_p\right).
\end{aligned}$$

which completes the proof of the lemma. $\quad\square$

Theorem 4.2. *Assume that for the the largest and the smallest eigenvalues of symmetric positive definite matrix A the following condition is fulfilled*

$$\lambda_n < \frac{2\lambda_1}{1 + \lambda_1}. \tag{30}$$

Therewith, let the iterations (10) *be applied on strictly convex quadratic function f which is defined by the relation* (22). *Then, the next holds:*

$$(d_i^{k+1})^2 \le \delta^2(d_i^k)^2 \tag{31}$$

where

$$\delta = \max\left\{1 - \frac{\lambda_1}{2\lambda_n}\left(\sigma_t + \sigma_p\right), \lambda_n\left(\frac{1}{\lambda_1} + 1\right) - 1\right\}, \tag{32}$$

and $d_i^k \in \mathbb{R}$, $k, i, n \in \mathbb{N}$. Furthermore,

$$\lim_{k\to\infty}\|\mathbf{g}_k\| = 0. \tag{33}$$

Proof. Assume that $\{v_1, v_2, \ldots, v_n\}$ is the set of the orthonormal eigenvectors of symmetric positive definite matrix A. Applying the Algorithm 3 we can generate the sequence of values $\{\mathbf{x}_k\}$. We are familiar with the fact that $\mathbf{g}_k = A\mathbf{x}_k - \mathbf{b}$ for some k and chosen iterative value $\mathbf{x}_k$. Than again, the next expression is also true

$$\mathbf{g}_k = \sum_{i=1}^{n} d_i^k \mathbf{v}_i, \tag{34}$$

for some constants $d_1^k, d_2^k, \ldots, d_n^k \in \mathbf{R}$.

Applying (10) there we get

$$\begin{aligned}
\mathbf{g}_{k+1} &= A\mathbf{x}_{k+1} - \mathbf{b} = A(\mathbf{x}_k - \alpha t_k \gamma_k^{-1} \mathbf{g}_k - \alpha p_k \mathbf{g}_k) - \mathbf{b} \\
&= A\mathbf{x}_k - \alpha t_k \gamma_k^{-1} A\mathbf{g}_k - \alpha p_k A\mathbf{g}_k - \mathbf{b} = \mathbf{g}_k - \alpha t_k \gamma_k^{-1} A\mathbf{g}_k - \alpha p_k A\mathbf{g}_k \\
&= \left(I - \alpha t_k \gamma_k^{-1} A - \alpha p_k A\right)\mathbf{g}_k.
\end{aligned} \tag{35}$$

Taking (34) produces

$$\mathbf{g}_{k+1} = \sum_{i=1}^{n} d_i^{k+1} \mathbf{v}_i = \sum_{i=1}^{n} \left(1 - \alpha t_k \gamma_k^{-1} \lambda_i - \alpha p_k \lambda_i\right) d_i^k \mathbf{v}_i. \tag{36}$$

Now, in order to prove (31), we need to confirm that $|1 - \alpha\lambda_i\left(t_k\gamma_k^{-1} + p_k\right)| \le \delta$. Practically, we analyze two cases:

1.
$$\alpha\lambda_i\left(t_k\gamma_k^{-1} + p_k\right) \le 1.$$

Previous assumption implies following estimations:

$$\begin{aligned}
1 &\ge \alpha\lambda_i\left(t_k\gamma_k^{-1} + p_k\right) \ge \frac{\lambda_1}{2\lambda_n}\left(\sigma_t + \sigma_p\right) \\
\implies 1 &- \alpha\lambda_i\left(t_k\gamma_k^{-1} + p_k\right) \le 1 - \frac{\lambda_1}{2\lambda_n}\left(\sigma_t + \sigma_p\right) \le \delta.
\end{aligned} \tag{37}$$

2.
$$\alpha\lambda_i\left(t_k\gamma_k^{-1} + p_k\right) > 1.$$

From this presumption we have:

$$\begin{aligned}
1 &< \alpha\lambda_i\left(t_k\gamma_k^{-1} + p_k\right) \le \lambda_n\left(\frac{1}{\lambda_1} + 1\right) \\
\implies |1 &- \alpha\lambda_i\left(t_k\gamma_k^{-1} + p_k\right)| \le \lambda_n\left(\frac{1}{\lambda_1} + 1\right) - 1 < \delta.
\end{aligned} \tag{38}$$

Under the posed condition (30) the following two inequalities are valid

-

$$0 < \frac{\lambda_1}{2\lambda_n}\left(\sigma_t + \sigma_p\right) < 1 \Longrightarrow 0 < 1 - \frac{\lambda_1}{2\lambda_n}\left(\sigma_t + \sigma_p\right) < 1 \tag{39}$$

-

$$\lambda_n\left(\frac{1}{\lambda_1} + 1\right) > \frac{\lambda_1}{\lambda_1} + \lambda_1 > 1 \Longrightarrow 0 < \lambda_n\left(\frac{1}{\lambda_1} + 1\right) - 1 < \frac{2\lambda_1}{1 + \lambda_1}\frac{1 + \lambda_1}{\lambda_1} - 1 = 1 \tag{40}$$

which confirm that parameter $\delta \in (0, 1)$.

Finally, to prove (33) we use representation (34) and the fact that $\{v_1, v_2, \dots, v_n\}$ is an orthonormal system of eigenvectors. These facts lead us to the following relation

$$\|g_k\|^2 = \sum_{i=1}^{n}\left(d_i^k\right)^2. \tag{41}$$

Since we previously showed that the parameter δ under condition (30) satisfies $0 < \delta < 1$, the final conclusion

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

is evident. □

5. Numerical comparisons

In this section, a performance profiles regarding tested metrics of four comparative models are presented. We tested the performance of following four methods: HADSS, ADSS, HSM and HADD. We naturally chose the ADSS method as the comparatione model since derived HADSS iteration originates from it. We also selected the HSM process for comparatione because it is determined on the same hybrid basis as the HADSS model. In [12] the author numerically confirmed that the ADSS method outperforms the SM method as well as the accelerated double direction ADD method presented in [16]. Taking the similar analysis in [15] the dominance of the HSM iteration among the other two comparative methods, which are its forerunner SM method and the Nestorov's accelerated gradient method with line search i.e. NLS method, is proved. In order to complete this numerical research we included one more hybrid accelerated scheme, the HADD method, presented in [17].

We chose to analyze the next two characteristics of all tested methods: needed number of iterations and number of function evaluations. Numerical tests are based on total 1200 test outcomes, which involves 30 test functions from [2]. On each chosen test function all four comparative methods were applied for the next 10 different number of parameters:1000, 2000, 3000, 5000, 7000, 8000, 10000, 15000, 20000, 30000. All tests are implemented on a Workstation Intel Celeron 1.6 GHz. The usual stopping criteria were taken:

$$\|\mathbf{g}_k\| \le 10^{-6} \quad \text{and} \quad \frac{|f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)|}{1 + |f(\mathbf{x}_k)|} \le 10^{-16}.$$

In backtracking procedures we used the next values for required parameters $\sigma_1 = 0.0001$, $\eta_1 = 0.8$, $\sigma_1 = 0.0002$, $\eta_1 = 0.9$.

For displaying the performance features of analyzed comparative models, i.e. the efficiency of derived HADSS algorithm versus ADSS, HSM and HADD iterations, we use Dolan and Moré's performance profiles of tested metrics [4]. We present the Dolan-Moré's performance profiles subject to the number of iterations and number of function evaluations in Figure 5. From both displays in Figure 5, (left) and (right), we see that the HADSS and the ADSS methods are evidently more robust and more efficient and that these two models convincingly upgrade the HSM and the HADD algorithms. Therewith, an interesting similar behaviors

Figure 5.1.(Left) Performance profiles regarding the number of iterations metric
(Right) Performance profiles regarding the function evaluations metric

of the HADSS and ADSS iterations regarding both analyzed metrics can be spotted. This fact points that the hybrid version of the ADSS model at least keeps the same good characteristics as its forerunner. Still, the HADSS has significant better performance profiles subject to the number of iterations and number of function evaluations metrics than hybrid HSM and HADD methods which are defined based on the same hybridization rule.
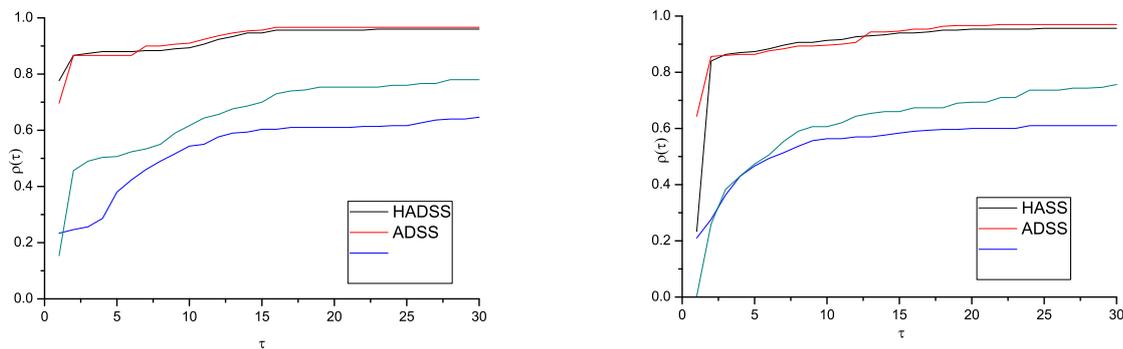
6. Conclusion

We present a hybrid accelerated double step size method, equipped with two Backtracking line search procedures, for solving unconstrained optimization problems. This method is generated using three term hybrid rule where for the guiding operator we take the accelerated double step size iteration. Defined hybrid accelerated process is convergent on the sets of uniformly convex and strictly convex quadratic functions. Numerical analysis confirms that derived method is effective for the large scale of test functions. Presented HADSS schemes shows efficient and robust performance subject to the number of iterations and the number of function evaluations metrics, which is very similar to behavior of its forerunner ADSS method. Intensive numerical comparisons, taken on 1200 optimization problems of different dimensions, show that derived method convincingly outperforms hybrid accelerated HSM and HADD methods which are defined based on the same hybridization rule as the HADSS scheme. Dolan-Moré's performance profiles subject to analyzed metrics are used to illustrate the efficiency and the robustness of proposed optimization model.

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(1)**(2008) 147–161
[3] L. Armijo, *Minimization of functions having Lipschitz first partial derivatives*, Pac. J. Math, **6** (1966) 1–3.
[4] E.D. Dolan, J.J. Moré, J.J. *Benchmarking optimization software with performance profiles*, Math. Program. **91** (2002) 201–213
[5] R. Fletcher, C.M. Reeves, *Function minimization by conjugate gradients*, Comput. J. **7** (1964) 149–154.
[6] A.A. Goldstein, *On steepest descent*, SIAM J. Control **3** (1965) 147–151
[7] S. Ishikawa, *Fixed points by a new iteration method*, Proc. Am. Math. Soc. **44** (1974) 147-150
[8] S.H. Khan, *A Picard-Mann hybrid iterative process*, Fixed Point Theory and Applications, **69** (2013)
[9] W.R. Mann, *Mean value methods in iterations*, Proc. Am. Math. Soc. **4** (1953) 506-510
[10] J.M. Ortega, W.C. Rheinboldt, *Iterative Solution Of Nonlinear Equation in Several Variables*, Academic Press, 1970.

[11] M. Petrović, S. Panić, M.M. Carević,*Initial improvement of the hybrid accelerated gradient descent process*, Bull. Aust. Math. Soc., **98 (2)** (2018) 331–338

[12] M.J. Petrović, *An accelerated Double Step Size method in unconstrained optimization*, Applied Math. Comput. **250**, (2015) 309–319

[13] M. J. Petrović, M. Ivanović, M. Djordjević, *Comparative performance analysis of some accelerated and hybrid accelerated gradient models*, University thought, Nat. Sci. **9(1)** doi: 10.5937/univtho9-18174 (2019)

[14] M. J. Petrović, N. Kontrec, S.Panić, *Determination of accelerated factors in gradient decent iterations based on Taylor's series*, University thought, Nat. Sci. **7(1)** (2017) 41–45

[15] M. Petrović, V. Rakocević, N. Kontrec, S. Panić, D. Ilić, *Hybridization of accelerated gradient descent method*, Numer. Algor. **79(3)** (2018) 769–786

[16] M.J. Petrović, P.S. Stanimirović, *Accelerated Double Direction method for solving unconstrained optimization problems*, Mathematical Problems in Engineering, Volume 2014 (2014), Article ID 965104, 8 pages.

[17] M. Petrović, P.S. Stanimirović, N. Kontrec, J. Mladenović, *Initial improvement of the hybrid accelerated gradient descent process*, Mathematical Problems in Engineering, Article ID 1523267, 8 pages (2018)

[18] E. Picard, *Memoire sur la theorie des equations aux derivees partielles et la methode des approximations successives*, J. Math. Pures Appl. **6** (1890)145-210

[19] E.Polak, G. Ribière, *Note sur la convergence de directions conjugees*, Rev. Fran. Inf. Rech. Opérat. **3** (1969) 35-43

[20] B.T. Polyak, *The conjugate gradient method in extreme problems*, Comput. Math. Mathem. Phy. **9** (1969) 94-112

[21] M.J.D. Powell, *Some global convergence properties of a variable-metric algorithm for minimization without exact line search*, AIAM-AMS Proc., Philadelphia **9** (1976) 53–72

[22] Rockafellar, R.T, *Convex Analysis*, Princeton University Press, Princeton (1970)

[23] Stanimirovic, P.S., Miladinovic, M.B., *Accelerated gradient descent methods with line search*, Numer. Algor. **54** (2010) 503–520.

[24] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Rev. **11** (1968) 226–235