Applicable Analysis and Discrete Mathematics
available online at http://pefmath.etf.rs

Appl. Anal. Discrete Math. **15** (2021), 519–535.
https://doi.org/10.2298/AADM190325033G

A GLOBALLY CONVERGENT MODIFIED MULTIVARIATE VERSION OF THE METHOD OF MOVING ASYMPTOTES

Allal Guessab and Abderrazak Driouch*

In this paper, we introduce an extension of our previous paper, *A globally convergent version to the Method of Moving Asymptotes*, in a multivariate setting. The proposed multivariate version is a globally convergent result for a new method, which consists iteratively of the solution of a modified version of the method of moving asymptotes. It is shown that the algorithm generated has some desirable properties. We state the conditions under which the present method is guaranteed to converge geometrically. The resulting algorithms are tested numerically and compared with several well-known methods.

1. MOTIVATION AND THEORETICAL JUSTIFICATION

We consider some new iterative methods for solving the unconstrained optimization problem: Find $\boldsymbol{x}_* \in \Omega$ such that

$$f(\boldsymbol{x}_*) = \min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x}), \tag{1}$$

where Ω is an open subset of $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is a given non-linear real-valued objective function, typically twice continuously differentiable, which could be non-convex. In order to evaluate the merit of using second order information an extension of the method of moving asymptotes, that accounts for the curvatures,

*Corresponding author. Allal Guessab
2020 Mathematics Subject Classification. 49M29, 49M37, 65K05, 90C30.
Keywords and Phrases: Non-convex, Non Linear Optimization, Global Convergence, Method of Moving Asymptotes.

was proposed in [**1**]. The authors have assumed in [**1**] that the objective function should fulfill for each iteration k the following conditions:

$$f_{,,jj}(\boldsymbol{x}^{(k)}) > 0, \quad j = 1, \dots, n. \tag{2}$$

This is a real weakness of this approach, which drastically limits its application. Hence, this method is very restrictive and also has the following disadvantages :

- It needs good initial solution $\boldsymbol{x}_*$ close to the exact solution.
- It converges slowly, in many cases, to the optimum $\boldsymbol{x}_*$.
- It does not always converge.
- Its performance degrades when it applied to non convex functions.
- Incapable of handling non-separability.

The purpose of this paper is three-fold. First, we develop an extension of our previous paper [**4**] in a multivariate setting. Second, we propose a modified version of the [**1**] method by removing the restrictive condition (2) on the objective functions. We will concentrate on the method of moving asymptotes since it is considered to belong to the most efficient methods. We believe that similar ideas can be developed for the other members of several methods for solving minimization problems. Third, we show how the proposed algorithm can be modified in order that the technique can be applied to a fairly large class of objective functions. Moreover, it is shown, as will be proved below, that the new method converges geometrically. Comparative numerical simulations are conducted to show the success of the proposed extension for various kinds of different test functions. The results suggest that this latter is significantly faster compared to the [**1**] method, Newton's method or the BFGS method on all test functions and it can succeed where these latter diverge simultaneously. It also has the advantage that, under appropriate conditions, global convergence of the algorithm is guaranteed.

The remainder of this paper is organized as follows. In Section 2, we develop a modified moving asymptotes method. Section 3 contains all technical results that are essential to show that our method is guaranteed to converge geometrically. In Section 4, we will concentrate on how our algorithm can be extended to the general setting where the objective function is nonconvex and non-separable. The point is not to give a definite answer to this problem but, rather, to clarify and understand how our algorithm can be used for this general situation. The test problems are Wood's, Powell's and Branin's functions. They are all documented in [**3**]. In section 5, various numerical experiments conduct to confirm our theoretical finding. In addition, the comparison with the models considered is also made and we conclude in Section 6.

2. A SPECIAL MODIFIED MULTIVARIATE VERSION OF MOVING ASYMPTOTES METHOD

Throughout this section we assume that w is a function satisfying the following conditions:

$$w \text{ is a real-valued function, defined and continuous on } \mathbb{R}^n, \tag{3}$$

$$\lim_{|\boldsymbol{x}|\to+\infty} w(\boldsymbol{x}) = 0. \tag{4}$$

The general modification of moving asymptotes method that we examine herein may be described as follows: Given the iteration point $\tilde{\boldsymbol{x}}^{(k)}$ (at iteration k).

- Our approach is to iteratively approximate at the k-th iteration the objective function by the approximating function $\tilde{f}_w^{(k)}$ where:

$$f(\boldsymbol{x}) \approx \tilde{f}_w^{(k)}(\boldsymbol{x}) \;=\; \sum_{j=1}^{n} \frac{\tilde{c}_j^{(k)}}{x_j - \tilde{d}_j^{(k)}} + \left\langle \tilde{\boldsymbol{b}}^{(k)}, \boldsymbol{x} - \tilde{\boldsymbol{d}}^{(k)} \right\rangle + \tilde{a}^{(k)} \tag{5}$$

  and the coefficients $\tilde{\boldsymbol{b}}^{(k)}, \tilde{\boldsymbol{c}}^{(k)}, \tilde{\boldsymbol{d}}^{(k)}$ are some chosen parameters given by

$$\tilde{\boldsymbol{b}}^{(k)} = \left(\tilde{b}_1^{(k)}, \ldots, \tilde{b}_n^{(k)}\right), \tilde{\boldsymbol{c}}^{(k)} = \left(\tilde{c}_1^{(k)}, \ldots, \tilde{c}_n^{(k)}\right), \tilde{\boldsymbol{d}}^{(k)} = \left(\tilde{d}_1^{(k)}, \ldots, \tilde{d}_n^{(k)}\right),$$

  and $\tilde{a}^{(k)} \in \mathbb{R}$. A straightforward calculation shows that the first and the second-order partial derivatives of $\tilde{f}_w^{(k)}$ have the following expressions:

$$(\tilde{f}_w^{(k)})_{,j}(\boldsymbol{x}) = \tilde{\boldsymbol{b}}_j^{(k)} - \frac{\tilde{c}_j^{(k)}}{(x_j - \tilde{d}_j^{(k)})^2}, \quad j = 1, \ldots, n, \tag{6}$$

$$(\tilde{f}_w^{(k)})_{,,jj}(\boldsymbol{x}) = \frac{2\tilde{c}_j^{(k)}}{(x_j - \tilde{d}_j^{(k)})^3}, \quad j = 1, \ldots, n, \tag{7}$$

  and since the function $\tilde{f}_w^{(k)}$ is separable, therefore if $i \neq j$, we have:

$$(\tilde{f}_w^{(k)})_{,,ij} = 0, \quad i, j = 1, \ldots, n. \tag{8}$$

- The approximating function $\tilde{f}_w^{(k)}$ is first order approximations of the original function f at the current iteration point $\tilde{\boldsymbol{x}}^{(k)}$, i.e.,

$$\tilde{f}_w^{(k)}(\tilde{\boldsymbol{x}}^{(k)}) = f(\tilde{\boldsymbol{x}}^{(k)}), (\tilde{f}_w^{(k)})_{,j}(\tilde{\boldsymbol{x}}^{(k)}) = f_{,j}(\tilde{\boldsymbol{x}}^{(k)}), \quad j = 1, \ldots, n \tag{9}$$

  In addition to the above conditions, the approximating function should satisfy the following more general condition (10):

$$(\tilde{f}_w^{(k)})_{,,jj}(\tilde{\boldsymbol{x}}^{(k)}) \;=\; \left| f_{,,jj}(\tilde{\boldsymbol{x}}^{(k)}) + w(\tilde{\boldsymbol{x}}^{(k)}) f_{,j}(\tilde{\boldsymbol{x}}^{(k)}) \right|. \tag{10}$$

Consequently, in the present situation, the approximate parameters $\tilde{a}^{(k)}, \tilde{\boldsymbol{b}}^{(k)}$ and $\tilde{\boldsymbol{c}}^{(k)}$ can be expressed in the following forms:

$$\tilde{a}^{(k)} = f(\tilde{\boldsymbol{x}}^{(k)}) - \sum_{j=1}^{n} \frac{\tilde{x}_j^{(k)}}{\tilde{c}_j^{(k)} - \tilde{d}_j^{(k)}} - \left\langle \tilde{\boldsymbol{b}}^{(k)}, \tilde{\boldsymbol{x}}^{(k)} - \tilde{\boldsymbol{d}}^{(k)} \right\rangle, \tag{11}$$

$$\tilde{b}_j^{(k)} = f_{,j}(\tilde{\boldsymbol{x}}^{(k)}) + \frac{\tilde{c}_j^{(k)}}{(\tilde{x}_j^{(k)} - \tilde{d}_j^{(k)})^2}, \tag{12}$$

$$\tilde{c}_j^{(k)} = \frac{\left| f_{,,jj}(\tilde{\boldsymbol{x}}^{(k)}) + w(\tilde{\boldsymbol{x}}^{(k)}) f_{,j}(\tilde{\boldsymbol{x}}^{(k)}) \right|}{2} \left(\tilde{x}_j^{(k)} - \tilde{d}_j^{(k)} \right)^3. \tag{13}$$

Furthermore, in order to fully determine an explicit expression for the approximating function $\tilde{f}_w^{(k)}$, the parameter $\tilde{d}^{(k)}$ is chosen such that

$$\tilde{d}_j^{(k)} = \tilde{x}_j^{(k)} + 2\tilde{\alpha}_j^{(k)} \frac{f_{,j}(\tilde{\boldsymbol{x}}^{(k)})}{\left| f_{,,jj}(\tilde{\boldsymbol{x}}^{(k)}) + w(\tilde{\boldsymbol{x}}^{(k)}) f_{,j}(\tilde{\boldsymbol{x}}^{(k)}) \right|}, \tag{14}$$

for simplicity, we put

$$\gamma_j^{(k)} = \left| f_{,,jj}(\tilde{\boldsymbol{x}}^{(k)}) + w(\tilde{\boldsymbol{x}}^{(k)}) f_{,j}(\tilde{\boldsymbol{x}}^{(k)}) \right| > 0. \tag{15}$$

So we can rewrite

$$\tilde{d}_j^{(k)} = \tilde{x}_j^{(k)} + 2\tilde{\alpha}_j^{(k)} \frac{f_{,j}(\tilde{\boldsymbol{x}}^{(k)})}{\gamma_j^{(k)}}, \tag{16}$$

where $\{\tilde{\boldsymbol{\alpha}}^{(k)}\}_k := \left\{ (\tilde{\alpha}_1^{(k)}, \ldots, \tilde{\alpha}_n^{(k)}) \right\}_k$ is a sequence of $\mathbb{R}^n$ with

$$\tilde{\alpha}_j^{(k)} > 1, k \in \mathbb{N} \quad \textit{and} \quad j = 1, \ldots, n. \tag{17}$$

Different rules for how to choose these values (and possible weight functions in (10)) will be provided later. We note that our method does not use the interpolation condition (19), but instead we have incorporated a first- and second-order information, as given in (13). Moreover, in particular, if you take $w = 0$ and at each iteration condition (2) is fulfilled, then our iterative scheme obviously reduces to the one introduced in [**1**]. Hence, subsequent iterations of the [**1**] method are essentially the same as the approximating function (5), except that in the proposed approximating function $\tilde{f}_w^{(k)}$ the parameters $\boldsymbol{c}^{(k)}$ and $\boldsymbol{d}^{(k)}$ are replaced by those given in (13) and (16) respectively. Thus our scheme starts with some guess point $\tilde{\boldsymbol{x}}^{(0)}$ and generates successive iterates by

$$f(\tilde{\boldsymbol{x}}^{(k+1)}) \leftarrow \tilde{f}_w^{(k)}(\tilde{\boldsymbol{x}}^{(k+1)}) = \min_{\boldsymbol{x} \in \Omega} \tilde{f}_w^{(k)}(\boldsymbol{x}). \tag{18}$$

For the sake of notation simplicity, we have removed the index w in the iterative sequence $\tilde{\boldsymbol{x}}_w^{(k)}$.

One of the key ingredient of the proposed approach is to work with (13) instead of

$$f_{,,jj}^{(k)}(\boldsymbol{x}^{(k)}) \quad = \quad f_{,,jj}(\boldsymbol{x}^{(k)}), \tag{19}$$

This modification will play an important role in the analysis of the proposed modified algorithm. Indeed, there are several good reasons for this choice. First, as mentioned above, the main reason is that this allows us to apply our method to a large class of objective functions. Second, there are also important advantages from the numerical point of view: many experimental results will confirm that the iterative scheme based on our modification version (13) can yield significantly fewer iterations than the [**1**] method, Newton's method or the BFGS method. Furthermore, in contrast to these three approaches, our method converges even if the starting point is very far from the true solution. In addition, as we will see, the key features of the present method are:

- It does not require us to build a good initial solution close to the exact solution.
- It converges geometrically for a large class of functions w that satisfy conditions (3) and (4).
- It will succeed where the [**1**] method, Newton's method and the BFGS Method break down.

Newton's method and the BFGS Method have a well-studied convergence theory that guarantees the convergence to a solution under a standard set of assumptions. For these and other their variants, the interested reader should consult one of the many excellent books on this subject [**2**, pp. 48–75] and [**5**, pp. 75–89]. We refer the readers to [**1**] and the references therein for the method of moving asymptotes.

3. CONVERGENCE ANALYSIS

We start this section with a result concerning an explicit expression for the iterative sequence $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_k$ generated by the approximating function $\tilde{f}_w^{(k)}$, as given in (5). Here, we continue to denote by $\tilde{\boldsymbol{c}}^{(k)}, \tilde{\boldsymbol{d}}^{(k)}$ and $\tilde{\boldsymbol{\alpha}}^{(k)}$ the coefficients given by (13), (16) and (17) respectively. Note that condition (17), imposed on the parameters $\tilde{\alpha}^{(k)}$, is crucial since it will guarantee strict convexity of the approximating function $\tilde{f}_w^{(k)}$. For brevity, in the following we use the notation:

$$\mathcal{I}^{(k)} = \mathcal{I}_1^{(k)} \times \ldots \times \mathcal{I}_n^{(k)}, \text{ with } \mathcal{I}_i^{(k)} = \Big]-\infty, \tilde{d}_i^{(k)}\Big[\cup \Big]\tilde{d}_i^{(k)}, +\infty\Big[, \quad i = 1, \ldots, n. \tag{20}$$

Now we are in position to state the first main result.

Theorem 1. *With the above notation, let $\Omega \subset \mathbb{R}^n$ be an open subset, a given twice continuously differentiable function f in Ω, $\tilde{\boldsymbol{x}}^{(0)} \in \Omega$ and $\tilde{\boldsymbol{x}}^{(k)}$ being respectively the initial and a current point of the sequence $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$. Then, for each $k > 0$ the approximating function $\tilde{f}_w^{(k)}$ defined by (5) is a strictly convex function on $\mathcal{I}^{(k)}$. In addition, it has an unique minimum at*

$$\tilde{\boldsymbol{x}}^{(k+1)} \leftarrow \tilde{\boldsymbol{x}}_*^{(k)} = \tilde{\boldsymbol{d}}^{(k)} + \boldsymbol{G}^{(k)} \tag{21}$$

where $\boldsymbol{G}^{(k)} = \left(G_1^{(k)}, \ldots, G_n^{(k)}\right)$ with

$$G_j^{(k)} = (\tilde{x}_j^{(k)} - \tilde{d}_j^{(k)})\sqrt{\tilde{s}_j^{(k)}} \tag{22}$$

$$\tilde{s}_j^{(k)} = \frac{\tilde{\alpha}_j^{(k)}}{\tilde{\alpha}_j^{(k)} - 1}. \tag{23}$$

Proof. The main ingredient here is a suitable application of the condition (17) imposed on the coefficient $\tilde{\boldsymbol{d}}^{(k)}$. We first start by showing that the approximating function $\tilde{f}_w^{(k)}$ is well defined and strictly convex in $\mathcal{I}^{(k)}$. To this end we prove that $\nabla^2 \tilde{f}_w^{(k)}$ is positive semidefinite in $\mathcal{I}^{(k)}$. Indeed, a simple calculation reveals that

$$\nabla^2 \tilde{f}_w^{(k)}(\boldsymbol{x}) = \begin{pmatrix} \gamma_1^{(k)} \left(\frac{\tilde{x}_1^{(k)} - \tilde{d}_1^{(k)}}{x_1 - \tilde{d}_1^{(k)}}\right)^3 & 0 & \cdots & 0 \\ 0 & \gamma_2^{(k)} \left(\frac{\tilde{x}_2^{(k)} - \tilde{d}_2^{(k)}}{x_2 - \tilde{d}_2^{(k)}}\right)^3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \gamma_n^{(k)} \left(\frac{\tilde{x}_n^{(k)} - \tilde{d}_n^{(k)}}{x_n - \tilde{d}_n^{(k)}}\right)^3 \end{pmatrix} \tag{24}$$

In view of (24), it remains to show that the Hessian matrix on the right-hand side of (24) is positive semidefinite for all $x \in \mathcal{I}^{(k)}$. Since $\gamma^{(k)}$ is nonnegative (as can be seen in (15)), and the two terms $\tilde{x}_j^{(k)} - \tilde{d}_j^{(k)}$ and $x_j - \tilde{d}_j^{(k)}$ have the same sign in the interval $\mathcal{I}_j^{(k)}$, and for all $\boldsymbol{x}$ in $\mathcal{I}^{(k)}$, $\boldsymbol{x}^T \nabla^2 \tilde{f}_w^{(k)}(\boldsymbol{x})\boldsymbol{x} \geq 0$ then $\tilde{f}_w^{(k)}$ is a convex function on $\mathcal{I}^{(k)}$. Furthermore, the function $\tilde{f}_w^{(k)}$ being well definite and continuous in $\mathcal{I}^{(k)}$, this implies the existence of a minimum, which by convexity is the unique critical point $\tilde{\boldsymbol{x}}_*^{(k)}$. Now, looking for $\nabla \tilde{f}_w^{(k)}(\boldsymbol{x}) = \mathbf{0}_{\mathbb{R}^n}$ we conclude that the optimum

$\tilde{\boldsymbol{x}}_*^{(k)}$ is one solution of the system

$$
(25)\qquad \begin{pmatrix} f_{,1}\left(\tilde{\boldsymbol{x}}^{(k)}\right)+\frac{1}{2}\gamma_1^{(k)}\left(\tilde{x}_1^{(k)}-\tilde{d}_1^{(k)}\right)\left(1-\frac{\left(\tilde{x}_1^{(k)}-\tilde{d}_1^{(k)}\right)^2}{\left(x_1-\tilde{d}_1^{(k)}\right)^2}\right)\\ \vdots \\ f_{,j}\left(\tilde{\boldsymbol{x}}^{(k)}\right)+\frac{1}{2}\gamma_j^{(k)}\left(\tilde{x}_j^{(k)}-\tilde{d}_j^{(k)}\right)\left(1-\frac{\left(\tilde{x}_j^{(k)}-\tilde{d}_j^{(k)}\right)^2}{\left(x_j-\tilde{d}_j^{(k)}\right)^2}\right)\\ \vdots \\ f_{,n}\left(\tilde{\boldsymbol{x}}^{(k)}\right)+\frac{1}{2}\gamma_n^{(k)}\left(\tilde{x}_n^{(k)}-\tilde{d}_n^{(k)}\right)\left(1-\frac{\left(\tilde{x}_n^{(k)}-\tilde{d}_n^{(k)}\right)^2}{\left(x_n-\tilde{d}_n^{(k)}\right)^2}\right) \end{pmatrix} = \begin{pmatrix} 0\\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix},
$$

or, equivalently,

$$
(26)\quad f_{,j}\left(\tilde{\boldsymbol{x}}^{(k)}\right)+\frac{1}{2}\gamma_j^{(k)}\left(\tilde{x}_j^{(k)}-\tilde{d}_j^{(k)}\right)\left(1-\frac{\left(\tilde{x}_j^{(k)}-\tilde{d}_j^{(k)}\right)^2}{\left(x_j-\tilde{d}_j^{(k)}\right)^2}\right)=0,\quad j=1,\ldots,n,
$$

which, after trivial calculations, implies

$$
(27)\qquad \left(\frac{\tilde{x}_j^{(k)}-\tilde{d}_j^{(k)}}{x_j-\tilde{d}_j^{(k)}}\right)^2 = 1+\frac{2f_{,j}\left(\tilde{\boldsymbol{x}}^{(k)}\right)}{\gamma_j^{(k)}\left(\tilde{x}_j^{(k)}-\tilde{d}_j^{(k)}\right)}.
$$

Now using (16), we can write $\frac{2f_{,j}\left(\tilde{\boldsymbol{x}}^{(k)}\right)}{\gamma_j^{(k)}\left(\tilde{x}_j^{(k)}-\tilde{d}_j^{(k)}\right)} = -\frac{1}{\tilde{\alpha}_j^{(k)}}$. Therefore, after some simplification, equation (27) becomes

$$
(28)\qquad \left(\frac{\tilde{x}_j^{(k)}-\tilde{d}_j^{(k)}}{x_j-\tilde{d}_j^{(k)}}\right)^2 = 1-\frac{1}{\tilde{\alpha}_j^{(k)}}.
$$

By condition (17) imposed on the parameter $\boldsymbol{\alpha}^{(k)}$, it can be deduced from (28) that the solvability of our subproblem can always be guaranteed. Indeed, under this condition, the required identity (21) immediately follows from (28) and the above mentioned fact that $\tilde{x}_j^k-\tilde{d}_j^{(k)}$ and $x_j-\tilde{d}_j^{(k)}$ have the same sign in the interval $\mathcal{I}_j^{(k)}$. This completes the proof of the theorem. □

3.1 Convergence study

In this Section, we give our second main result of this paper, that is sufficient conditions on the data (the point $\tilde{\boldsymbol{x}}^{(0)}$, the gradient ∇f in a neighborhood of $\tilde{\boldsymbol{x}}^{(0)}$, the family $diag(H_f\left(\tilde{\boldsymbol{x}}^{(k)}\right)), k\geq 0)$, which guarantee that first derivative of f vanishes in a neighborhood of $\boldsymbol{x}^*$, first, and secondly, the convergence of the method

to this zero.
To establish our convergence results, we need the following assumptions. We assume that there exist positive constants r, M and $\xi < 1$ such that the following assumptions hold: Here where $\|.\|$ is the standard Euclidean norm on $\mathbb{R}^n$.

Assumption 1.

$$B_r := \left\{ \boldsymbol{x} \in \mathbb{R}^n : \left\| \boldsymbol{x} - \tilde{\boldsymbol{x}}^{(0)} \right\| \leq r \right\} \subset \Omega.$$

Assumption 2.

$$0 < \frac{\tilde{\alpha}_j^{(k)}}{\tilde{\alpha}_j^{(k)} - 1} \leq \frac{M}{2} \gamma_j^k, \quad (k > 0), \quad j = 1, \ldots, n. \tag{29}$$

Assumption 3.

$$\sup_{k>0} \sup_{\boldsymbol{x} \in B} \left\| \nabla f_{,j}(\boldsymbol{x}) - \frac{f_{,j}(\boldsymbol{x}^{(k-1)})}{x_j^{(k-1)} - x_j^{(k)}} \boldsymbol{e}^{(j)} \right\| \leq \frac{\xi}{M}, \tag{30}$$

where $\mathbf{e}^{(j)}$ is the vector of $\mathbb{R}^n$ with $j-th$ component equal to 1 and all other components equal to 0.

Assumption 4.

$$0 < \left| f_{,j} \left(\tilde{\boldsymbol{x}}^{(0)} \right) \right| \leq \frac{r}{M\sqrt{n}} (1 - \xi).$$

Assumption 2 enforces the quite natural conditions (17). Indeed, if condition 2 holds, then (17) is also satisfied. Furthermore, Assumption 3 tells us that the coefficient $\nabla f_{,j}(\tilde{\boldsymbol{x}}^{(k)})$ does not change too much in a neighborhood of $\tilde{\boldsymbol{x}}^{(0)}$, and finally for any k the functions $\nabla f_{,j}(\tilde{\boldsymbol{x}}^{(k)})$ and $\frac{f_{,j}\left(\boldsymbol{x}^{(k-1)}\right)}{\boldsymbol{x}^{(k-1)} - \tilde{x}^{(k)}}$ have not to change too much in a neighborhood of $\tilde{\boldsymbol{x}}^{(k)}$. Assumption 4 only says that $\left| f_{,j} \left(\tilde{\boldsymbol{x}}^{(0)} \right) \right|$ is small enough and that $f_{,j} \left(\tilde{\boldsymbol{x}}^{(0)} \right)$ is non zero.

Throughout this subsection, we assume that Assumptions 1-4 hold. The constants r, M and $\xi < 1$ that appear in the subsequent analysis are always the constants from Assumptions 1-4. Our aim is to show that the sequence $\left\{ \tilde{\boldsymbol{x}}^{(k)} \right\}_{k \geq 0}$ defined in Theorem 1 converges geometrically to a point $\boldsymbol{x}^*$ in the sense that

$$\left\| \tilde{\boldsymbol{x}}^{(k)} - \boldsymbol{x}^* \right\| \leq \frac{\xi^k}{1 - \xi} \left\| \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(0)} \right\|.$$

Theorem 2. *Assume Assumptions 1-4 hold. Let the assumptions of theorem 1 be valid and let $\boldsymbol{G}^{(k)}$ be defined by (22). Then the sequence $\left\{ \tilde{\boldsymbol{x}}^{(k)} \right\}_{k \geq 0}$ given by*

$$\tilde{\boldsymbol{x}}^{(k+1)} = \tilde{\boldsymbol{d}}^{(k)} + \boldsymbol{G}^{(k)} \tag{31}$$

is completely contained in the ball B_r, and converges to the unique zero of ∇f in B_r.

We first state some auxiliary lemmas, which will be needed in our investigation.

Lemma 3. *Let Assumption 2 be satisfied and let the sequence* $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$ *be as defined in Theorem 2. Then, for any positive integer k the following inequality holds.*

$$\left\|\tilde{\boldsymbol{x}}^{(k)}-\tilde{\boldsymbol{x}}^{(k-1)}\right\| \leq M\left\|\nabla f(\tilde{\boldsymbol{x}}^{(k-1)})\right\|. \tag{32}$$

Proof. Let us fix a positive integer k. Using (21) and (23) we may write

$$\begin{aligned}\tilde{x}_j^{(k)}-\tilde{x}_j^{(k-1)} &= \tilde{d}_j^{(k-1)}+(\tilde{x}_j^{(k-1)}-\tilde{d}_j^{(k-1)})\sqrt{\tilde{s}_j^{(k-1)}}-\tilde{x}_j^{(k-1)} \\ &= (\tilde{x}_j^{(k-1)}-\tilde{d}_j^{(k-1)})\left(\sqrt{\tilde{s}_j^{(k-1)}}-1\right) \qquad (33)\end{aligned}$$

Now, from (17) we have $\tilde{s}_j^{(k-1)}>1,\ j=1,\ldots,n,\quad (k\geq 1)$, we then immediately deduce $\sqrt{\tilde{s}_j^{(k-1)}}<\tilde{s}_j^{(k-1)}$. Therefore, by (16), (23) and (33), we arrive at

$$\begin{aligned}\left|\tilde{x}_j^{(k)}-\tilde{x}_j^{(k-1)}\right| &\leq \left|\tilde{x}_j^{(k-1)}-\tilde{d}_j^{(k-1)}\right|\left(\tilde{s}_j^{(k-1)}-1\right), \qquad (34)\\ &\leq \frac{2\tilde{\alpha}_j^{(k-1)}}{(\tilde{\alpha}_j^{(k-1)}-1)\gamma_j^{(k-1)}}\left|f_{,j}(\tilde{\boldsymbol{x}}^{(k-1)})\right|. \qquad (35)\end{aligned}$$

Finally, combing Assumption 2 and this last inequality, we get the required inequality (32). □

In order to prove that the sequence $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$ converges geometrically, we need some further preparatory results.

Lemma 4. *Let Assumption 3 be satisfied and let the sequence* $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$ *be defined as in Theorem 2. Then, for any positive k the following inequality holds.*

$$\left|f_{,j}(\tilde{\boldsymbol{x}}^{(k)})\right| \leq \frac{\xi}{M}\left|\tilde{x}_j^{(k)}-\tilde{x}_j^{(k-1)}\right|,\ j=1,\ldots,n. \tag{36}$$

Proof. Fix a positive integer k. Let us define $\tilde{t}_j^{(k-1)}$ by

$$\tilde{t}_j^{(k-1)} := \frac{\gamma_j^{(k-1)}}{2}(\tilde{x}_j^{(k-1)}-\tilde{d}_j^{(k-1)})-f_{,j}\left(\tilde{\boldsymbol{x}}^{(k-1)}\right), \tag{37}$$

and the auxiliary function $\varphi : B \to \mathbb{R}$ as follows:

$$\varphi(\boldsymbol{x}) := f_{,j}(\boldsymbol{x})-\frac{f_{,j}(\tilde{\boldsymbol{x}}^{(k-1)})}{\frac{1}{2}\gamma_j^{(k-1)}(\tilde{x}_j^{(k)}-\tilde{x}_j^{(k-1)})}h_j(\boldsymbol{x}),$$

where

$$h_j(\boldsymbol{x}) := -f_{,j}\left(\tilde{\boldsymbol{x}}^{(k-1)}\right) + \frac{\gamma_j^{(k-1)}}{2}\left(x_j - \tilde{x}_j^{(k)} + \tilde{x}_j^{(k-1)} - \tilde{d}_j^{(k-1)}\right) - \tilde{t}_j^{(k-1)}.$$

Using (37), it is easily checked that φ satisfies

$$\begin{aligned} \varphi(\tilde{\boldsymbol{x}}^{(k-1)}) &= 0, \\ \varphi(\tilde{\boldsymbol{x}}^{(k)}) &= f_{,j}(\tilde{\boldsymbol{x}}^{(k)}). \end{aligned}$$

Also it is easy to see that

$$\nabla\varphi(\boldsymbol{x}) = \nabla f_{,j}(\boldsymbol{x}) - \frac{f_{,j}(\tilde{\boldsymbol{x}}^{(k-1)})}{\tilde{x}_j^{(k-1)} - \tilde{x}_j^{(k)}}\boldsymbol{e}^{(j)}. \tag{38}$$

Then, from the mean-value theorem and Assumption 3 we get

$$\begin{aligned} \left|f_{,j}\left(\tilde{\boldsymbol{x}}^{(k)}\right)\right| &= \left|\varphi(\tilde{\boldsymbol{x}}^{(k)}) - \varphi(\tilde{\boldsymbol{x}}^{(k-1)})\right| \\ &\leq \sup_{\boldsymbol{x}\in B_r} ||\nabla\varphi(\boldsymbol{x})|| \left|\left|\tilde{\boldsymbol{x}}^{(k)} - \tilde{\boldsymbol{x}}^{(k-1)}\right|\right| \\ &\leq \sup_{\boldsymbol{x}\in B_r} \left|\left|\nabla f_{,j}(\boldsymbol{x}) - \tfrac{f_{,j}(\tilde{\boldsymbol{x}}^{(k-1)})}{\tilde{x}_j^{(k-1)} - \tilde{x}_j^{(k)}}\boldsymbol{e}^{(j)}\right|\right| \left|\left|\tilde{\boldsymbol{x}}^{(k)} - \tilde{\boldsymbol{x}}^{(k-1)}\right|\right| \\ &\leq \tfrac{\xi}{M}\left|\left|\tilde{\boldsymbol{x}}^{(k)} - \tilde{\boldsymbol{x}}^{(k-1)}\right|\right|. \end{aligned} \tag{39}$$

This shows that the required inequality (36) holds true for any positive integer k. □

The next result shows that for all k the iterate $\tilde{\boldsymbol{x}}^{(k)}$ remains in the interval B_r.

Lemma 5. *Let Assumption 2-3 be satisfied and let the sequence $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$ be as defined in Theorem 2. Assume that the starting point $\tilde{\mathbf{x}}_0$ belongs to the interval B_r, where r is defined in Assumption 1. Then, all terms of the sequence $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$ lie inside the ball B_r.*

Proof. Indeed, combining inequalities (32), (36) of Lemmas 3 and 4 respectively, we immediately obtain

$$\left|\left|\tilde{\boldsymbol{x}}^{(k)} - \tilde{\boldsymbol{x}}^{(k-1)}\right|\right| \leq \xi\left|\left|\tilde{\boldsymbol{x}}^{(k-1)} - \tilde{\boldsymbol{x}}^{(k-2)}\right|\right| \leq \ldots \leq \xi^{k-1}\left|\left|\tilde{\boldsymbol{x}}^{(1)} - \tilde{\boldsymbol{x}}^{(0)}\right|\right|, \tag{40}$$

and therefore we have

$$\begin{aligned} \left|\left|\tilde{\boldsymbol{x}}^{(k)} - \tilde{\boldsymbol{x}}^{(0)}\right|\right| &\leq \textstyle\sum_{l=1}^{k}\left|\left|\tilde{\boldsymbol{x}}^{(l)} - \tilde{\boldsymbol{x}}^{(l-1)}\right|\right| \\ &\leq \left(\textstyle\sum_{l=1}^{k}\xi^{l-1}\right)\left|\left|\tilde{\boldsymbol{x}}^{(1)} - \tilde{\boldsymbol{x}}^{(0)}\right|\right| \\ &\leq \tfrac{\left|\left|\tilde{\boldsymbol{x}}^{(1)} - \tilde{\boldsymbol{x}}^{(0)}\right|\right|}{1-\xi}. \end{aligned} \tag{41}$$

Finally, applying inequality (32) for $k = 1$ and using Assumption 4 we immediately get

$$\left|\left|\tilde{\boldsymbol{x}}^{(k)} - \tilde{\boldsymbol{x}}^{(0)}\right|\right| \quad \leq \quad r, \tag{42}$$

which shows that each $\tilde{\boldsymbol{x}}^{(k)}$ belongs to B_r. □

As a consequence of the previous three lemmas, we are now in a position to prove Theorem 2.

Proof of Theorem 2. Since the entire sequence $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$ remains in the (closed) ball B_r by Lemma 5, every limit point of this sequence belongs to this set, too. Hence, it remains to show that the sequence $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$ converges. To this end, we first note that, for $k \geq 0$ and $l \geq 0$, we have

$$\begin{aligned} \left|\left|\tilde{\boldsymbol{x}}^{(k+l)} - \tilde{\boldsymbol{x}}^{(k)}\right|\right| &= \left|\left|\textstyle\sum_{v=0}^{l-1}(\tilde{\boldsymbol{x}}^{(k+v+1)} - \tilde{\boldsymbol{x}}^{(k+v)})\right|\right| \\ &\leq \textstyle\sum_{v=0}^{l-1}\left|\left|\tilde{\boldsymbol{x}}^{(k+v+1)} - \tilde{\boldsymbol{x}}^{(k+v)}\right|\right| \\ &\leq \xi^k \textstyle\sum_{v=0}^{l-1}\xi^v \left|\left|\tilde{\boldsymbol{x}}^{(1)} - \tilde{\boldsymbol{x}}^{(0)}\right|\right| \\ &\leq \tfrac{\xi^k}{1-\xi}\left|\left|\tilde{\boldsymbol{x}}^{(1)} - \tilde{\boldsymbol{x}}^{(0)}\right|\right|, \end{aligned} \tag{43}$$

then the sequence $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$ is a Cauchy sequence. Being Cauchy in B_r, it has a limit, $\boldsymbol{x}^*$, in B_r. Now, thanks to the continuity of $f_{,j}$ on B_r and the convergence of the sequence $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$ imply

$$|f_{,j}\left(\tilde{\boldsymbol{x}}^*\right)| = \lim_{k\to+\infty}\left|f_{,j}\left(\tilde{\boldsymbol{x}}^{(k)}\right)\right| \leq \frac{\xi}{M}\lim_{k\to+\infty}\left|\tilde{x}_j^{(k)} - \tilde{x}_j^{(k-1)}\right| = 0,\ j = 1, \ldots, n, \tag{44}$$

and then $\nabla f\left(\tilde{x}^*\right) = 0$. Passing to the limit for l tending to ∞ in (43), we deduce that

$$\left|\left|\tilde{\boldsymbol{x}}^{(k)} - \tilde{\boldsymbol{x}}^*\right|\right| \leq \frac{\xi^k}{1-\xi}\left|\left|\tilde{\boldsymbol{x}}^{(1)} - \tilde{\boldsymbol{x}}^{(0)}\right|\right|, \tag{45}$$

which shows the geometric convergence of the sequence $\left\{\tilde{\boldsymbol{x}}^{(k)}\right\}_{k\geq 0}$ to $\tilde{\boldsymbol{x}}^*$.

We are now in a position to prove that ∇f has an unique zero in B_r. To this end, we proceed by contradiction, assuming that ∇f has another zero $\tilde{\boldsymbol{y}}^* \in B_r$. Let us introduce the auxiliary function

$$\lambda_j(\boldsymbol{x}) = \frac{\tilde{x}_j^{(1)} - \tilde{x}_j^{(0)}}{f_{,j}(\tilde{\boldsymbol{x}}_j^{(0)})}\left(f_{,j}\left(\boldsymbol{x}\right) - \frac{f_{,j}(\tilde{\boldsymbol{x}}_j^{(0)})}{\tilde{x}_j^{(0)} - \tilde{x}_j^{(1)}}\left(x_j - \tilde{x}_j^*\right)\right),$$

which satisfies $\lambda_j(\tilde{\boldsymbol{x}}^*) = 0$ and $\lambda_j(\tilde{\boldsymbol{y}}^*) = \tilde{y}_j^* - \tilde{x}_j^*$. Therefore, applying (32) for $k = 1$, it follows from the mean value theorem and Lemma 3, inequality (32) for $k = 1$,

$$\begin{aligned} \left|\tilde{x}_j^* - \tilde{y}_j^*\right| &= \left|\lambda_j(\tilde{\boldsymbol{x}}^*) - \lambda_j(\tilde{\boldsymbol{y}}^*)\right| \\ &\leq \sup_{\boldsymbol{x}\in B_r} \left|\left|\nabla\lambda_j(\boldsymbol{x})\right|\right| \left|\left|\tilde{\boldsymbol{x}}^* - \tilde{\boldsymbol{y}}^*\right|\right| \\ &\leq \left|\frac{\tilde{x}_j^{(1)} - \tilde{x}_j^{(0)}}{f_{,j}(\tilde{\boldsymbol{x}}^0)}\right| \sup_{\boldsymbol{x}\in B} \left|\left|\nabla f_{,j}(\boldsymbol{x}) - \frac{f_{,j}(\tilde{\boldsymbol{x}}_0)}{\tilde{x}_j^0 - \tilde{x}_j^{(1)}} \boldsymbol{e}^{(j)}\right|\right| \left|\left|\tilde{\boldsymbol{x}}^* - \tilde{\boldsymbol{y}}^*\right|\right| \\ &\leq M \frac{\xi}{M} \left|\left|\tilde{\boldsymbol{x}}^* - \tilde{\boldsymbol{y}}^*\right|\right| \\ &= \xi \left|\left|\tilde{\boldsymbol{x}}^* - \tilde{\boldsymbol{y}}^*\right|\right|. \end{aligned} \tag{46}$$

This yields $\tilde{\boldsymbol{x}}^* = \tilde{\boldsymbol{y}}^*$ since $\xi < 1$. Thus, the theorem is proved.

3.2 Description of algorithm

The results of the previous section may be used to construct the following algorithm.

Algorithm 1 Modified Method of Moving Asymptotes

Input: $\tilde{\boldsymbol{x}}^{(0)} \in \mathbb{R}^n$, w, $M_1, M_2 \geq 1$, (and an optional error tolerance $\varepsilon > 0$).
$k = 0$
REPEAT
For $j = 0, 1, \ldots, n$
$\gamma_j^{(k)} = \left|f_{,,jj}(\tilde{\boldsymbol{x}}^{(k)}) + w(\tilde{\boldsymbol{x}}^{(k)}) f_{,j}(\tilde{\boldsymbol{x}}^{(k)})\right|$,
$\tilde{\alpha}_j^{(k)} = M_1 \left(1 + \frac{2}{M_2 \gamma_j^{(k)}}\right)$,
$\tilde{d}_j^{(k)} = \tilde{x}_j^{(k)} + 2\tilde{\alpha}_j^{(k)} \frac{f_{,j}(\tilde{\boldsymbol{x}}^{(k)})}{\gamma_j^{(k)}}$,
$\tilde{s}_j^{(k)} = \frac{\tilde{\alpha}_j^{(k)}}{\tilde{\alpha}_j^{(k)} - 1}$,
$\tilde{x}_j^{(k+1)} = \tilde{d}_j^{(k)} + (\tilde{x}_j^{(k)} - \tilde{d}_j^{(k)}) \sqrt{\tilde{s}_j^{(k)}}$,
while $\left|\left|\nabla f(\tilde{\boldsymbol{x}}^{(k)})\right|\right| > \epsilon$.
$k \longleftarrow k + 1$

4. MINIMIZING NONCONVEX NON-SEPARABLE FUNCTIONS

The separability is a measure of difficulty of different objective functions. In general the separable functions are relatively easy to solve, when compared with their inseparable counterpart, because each variable of a separable function is independent of the other variables. If all the parameters or variables are independent, then a sequence of n independent optimization processes can be performed. As a result, each design variable or parameter can be optimized independently, and if the objective function f is separable, then, for $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$

$$\arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) = \arg\min_{x_1,x_2,\ldots,x_n} f(x_1, x_2, \ldots, x_n) = \arg\min_{x_1,x_2,\ldots,x_n} \sum_{i=1}^{n} f_i(x_i) \tag{47}$$

$$= \sum_{i=1}^{n} \arg\min_{x_i} f_i(x_i) \tag{48}$$

where $f(\boldsymbol{x}) = \sum_{i=1}^{n} f_i(x_i)$. On the other hand, a function is called non separable, if its variables show inter-relation among themselves or are not independent. If the objective function variables are independent of each other, then the objective functions can be decomposed into sub-objective functions. Then, each of these sub-objectives involves only one decision variable, and Algorithm 1 solves the problem (1) for each variable independently of others, and then is like we minimize n functions of dimension 1. According to [**6**], the general condition of separability to see if the function is easy to optimize or not is given as

$$\frac{\partial f}{\partial x_i}(\bar{\boldsymbol{x}}) = g(x_i)h(\bar{\boldsymbol{x}}). \tag{49}$$

In this section, we extend Algorithm 1 for the non separable non convex optimization problem, which consists of a cyclic update of the variables x_i, let's consider the problem (1), where here the objective function is non separable: we can formulate this problem as: At each iteration of this method, the function is minimized with respect to a single of variables while the rest of the variables are held fixed. More specifically, at iteration k of the algorithm, the variable x_i is updated by solving the following subproblems

(50)

$$\left\{\begin{array}{lll} x_{\sigma(1)}^{(k+1)} & := & \arg\min_{x_{\sigma(1)}} f(x_{\sigma(1)}, x_{\sigma(2)}^{(k)}, \ldots, x_{\sigma(i)}^{(k)}, \ldots, x_{\sigma(n)}^{(k)}) \\ x_{\sigma(2)}^{(k+1)} & := & \arg\min_{x_{\sigma(2)}} f(x_{\sigma(1)}^{(k)}, x_{\sigma(2)}, \ldots, x_{\sigma(i)}^{(k)}, \ldots, x_{\sigma(n)}^{(k)}) \\ & \vdots & \\ x_{\sigma(i)}^{(k+1)} & := & \arg\min_{x_{\sigma(i)}} f(x_{\sigma(1)}^{(k+1)}, x_{\sigma(2)}^{(k+1)}, \ldots, x_{\sigma(i-1)}^{(k+1)}, x_{\sigma(i)}, \ldots, x_{\sigma(n)}^{(k)}) \\ & \vdots & \\ x_{\sigma(n)}^{(k+1)} & := & \arg\min_{x_{\sigma(n)}} f(x_{\sigma(1)}^{(k+1)}, x_{\sigma(2)}^{(k+1)}, \ldots, x_{\sigma(i)}^{(k+1)}, \ldots, x_{\sigma(n-1)}^{(k+1)}, x_{\sigma(n)}), \end{array}\right.$$

where σ is a uniformly random permutation of $\{1, \ldots, n\}$. Let us use $\boldsymbol{x}^{(k)}$ to denote the sequence of iterates generated by this algorithm, where

$$\boldsymbol{x}^{(k)} \triangleq (x_{\sigma(1)}^{(k)}, x_{\sigma(2)}^{(k)}, \ldots, x_{\sigma(n)}^{(k)})$$

at iteration k, the selected variable (say variable i) is computed by solving the n following subproblems

$$\left\{\begin{array}{l} \arg\min f_{\sigma(i)}(x_{\sigma(i)}) \\ \text{s.t} \quad x_{\sigma(i)} \in \mathbb{R} \end{array}\right. \tag{51}$$

where $f_{\sigma(i)}(x_{\sigma(i)}) := f(x^{(k+1)}_{\sigma(1)}, x^{(k+1)}_{\sigma(2)}, \dots, x^{(k+1)}_{\sigma(i-1)}, x_{\sigma(i)}, \dots, x^{(k)}_{\sigma(n)})$ is an approximation of the original objective function at the point $\boldsymbol{x}^{(k-1)}$. Algorithm 2 summarizes the process of solving a non convex optimization problem for a non separable objective function:

Algorithm 2 Modified version for non separable function

1: Input a feasible point $\boldsymbol{x}^{(0)}$, and set $k = 0$
2: **Repeat**
3: $k \Leftarrow k + 1$, choose a permutation σ,
4: If $\sigma := \{1, \dots, n\}$
5: For $i = 1, \dots, n$
6: Solve by the **algorithm** 1 the n one-dimensional problems

$$\begin{cases} \arg\min f_i(x_i) \\ \text{s.t} \quad x_i \in \mathbb{R} \end{cases} \qquad i = 1, \dots, n$$

7: $\boldsymbol{x}^{(k+1)} = (\arg\min_{x_1} f_1(x_1), \dots, \arg\min_{x_n} f_n(x_n))$
8: **Until** some convergence criterion is met

Note that Algorithm 2 reduces to Algorithm 1 for the case where condition of separability (49) is valid and then all of the variables are independent of each other.

5. NUMERICAL RESULTS

We employ the present method (designated as present) to solve some nonlinear, non-convex optimization problems. The purpose of this testing is to show that the algorithm works better than other algorithms in the same problem area. More precisely, we will compare it first with the [**1**] method, Newton's method and the BFGS method using the test function:

$$\begin{aligned} f_4(x, y, z) &= \frac{1}{2}\left(\exp(x^2) + 2\exp(y) + \frac{1}{2}(z-3)^4\right) + 3\left(\sin(x) - \frac{1}{6}\sin(2x)\right) \\ &- \left(\frac{1}{3}y^3 + \frac{5}{2}y^2 + 3(y+z) - 6\right) \end{aligned}$$

It is assumed that all methods use the finite difference method to compute first and second derivatives.

Numerical results are summarized in Table 1, where for each weight function w, we present the objective functions, the starting points, the methods used, the number of iterations N to obtain the objective value of the obtained optimal solution $\tilde{\boldsymbol{x}}^{(N)}$ and $f(\boldsymbol{x})$ at $\tilde{\boldsymbol{x}}^{(N)}$. We use the following stopping criteria $\left\|\nabla f(\tilde{\boldsymbol{x}}^{(N)})\right\| \leq \epsilon$, (the absolute value of the derivative of the function is less than or equal to the tolerance). For numerical illustrations we used different values of ϵ. Therefore, when

Initial point	Method	N	$X_N=(x_N,y_N,z_N)$	$f(X_N)\simeq\min f(x)$	ϵ
$\widetilde{x}^{(0)}=$ (10,100,200)	Newton BFGS [**1**] Present	div. div. div. 110	$\begin{pmatrix} -957.55090028105\text{e-}3 \\ 3.48246759967065\text{e}{+}0 \\ 2.44224957030741\text{e}{+}0 \end{pmatrix}^T$	$-23.99550098-$ $15414e{+}0$	10^{-16}
$\widetilde{x}^{(0)}=$ $(2,5,3)$	Newton BFGS [**1**] Present	19 19 22 12	$\begin{pmatrix} -957.55090028105e-3 \\ 3.48246759967065e+0 \\ 2.44224957030741e+0 \end{pmatrix}^T$	$-23.99550098-$ $15414e{+}0$	10^{-15}

Table 1: Numerical comparisons of the [**1**] method. Newton's method, the BFGS Method and the present method in three dimension. Here $\omega(x) = (1+|x|)^{\frac{1}{4}}\exp(-20|x|)$ and $\widetilde{\alpha}^{(k)} = \left(5\left(1+\frac{1}{7\gamma_1^{(k)}}\right), 2\left(1+\frac{1}{4\gamma_2^{(k)}}\right), 4\left(1+\frac{1}{3\gamma_3^{(k)}}\right)\right)$.

the stopping criterion is satisfied, $\tilde{\boldsymbol{x}}_* = \tilde{\boldsymbol{x}}^{(N)}$ is taken as the optimal solution. In Table 1 div. means that the stopping criteria is not satisfied.

The process described in section 4 has also been implemented in *Matlab*. The algorithm was terminated when the norm of the gradient $\left\|\nabla f(x^{(k)})\right\|$ was less than a specified tolerance ϵ. For all of test functions, the tolerance ϵ was taken to be very close to 0, in order to find the most exact solution. In these examples we want to use our algorithm to help compute the minimum of Wood's and Powell's and Branin's function, see [**3**]. For all these test functions in the table 2, the results include the number of variable of the objective function, respectively the number of iterations, evaluations functions N_{iter} and N_{eval} required to achieve convergence, the final function value and the CPU time required by the present algorithm.

6. CONCLUSION

The proposed modified method of moving asymptotes has been computationally shown that it needs only a small number of iterations to converge to the exact solution up to the specified error tolerance. The algorithm is easy to use since all tuning parameters are automatically chosen. Furthermore, test examples on non-convex and non-separable functions confirm that the algorithm is expected to be more efficient than the [**1**] method, Newton's method and the BFGS method. It also has the advantage that, under appropriate conditions, global convergence of the algorithm is guaranteed.

Functions	Number of variables	N_{iter}	N_{eval}	$f(X_N) \simeq \min f(x)$	CPU time
Wood's function	2	7	28	$452.7579 \cdot 10^{-30}$	0.004
	3	12	35	-8.776523408	0.025
	4	15	51	$12.7068 \cdot 10^{-30}$	0.067
Powell's function	2	13	42	$8.54987 \cdot 10^{-18}$	0.003
	3	18	60	$36.0706 \cdot 10^{-21}$	0.006
	4	34	73	$14.3158 \cdot 10^{-24}$	0.077
Branins's function	2	30	120	$397.887357 \cdot 10^{-3}$	0.016

Table 2: Results of applying algorithm 2 on the Wood, the Powell, and the Branin's function.

Acknowledgements. This article is dedicated to Professor Gradimir V. Milovanovic on the Occasion of his 70th anniversary.

REFERENCES

1. M. Bachar,T. Estebenet and A. Guessab: *A moving asymptotes algorithm using new local convex approximation methods with explicit solutions.* Electron. Trans. Numer. Anal., **43** (2014), 21–44.
2. R. L. Burden and J. D. Faires: *Numerical Analysis.* Cengange Learning, Boston, 2011.
3. W. C. Davidon and L. Nazareth: *DRVOCR-A FORTRAN Implementation of Davidon's Optimally Conditioned Method*, TM-306, Argonne National Lab., Argonne, 111., 1977.
4. A. Guessab, A. Driouch and O. Nouisser: *A globally convergent version to the Method of Moving Asymptotes.* Appl. Anal.Discrete Math. 13 (2019), 905-917.
5. A. Greenbaum and T. P. Chartier: Numerical Methods. Princeton University Press, Princeton, 2012.
6. M. Jamil and X-S. Yang: *A literature survey of benchmark functions for global optimization problems*, Int. Journal of Mathematical Modelling and Numerical Optimisation, (2013, Vol. 4, No. 2, 150–194.

Allal Guessab
Université de Pau et des Pays de l'Adour,
E2S UPPA, CNRS, LMAP, 64000, Pau, France
E-mail: *allal.guessab@univ-pau.fr*

(Received 25. 03. 2019.)
(Revised 12. 10. 2021.)

Abderrazak Driouch
Université de Pau et des Pays de l'Adour,
E2S UPPA, CNRS, LMAP, 64000, Pau, France
E-mail: *driouchabderrazak@gmail.com*