A MOVING ASYMPTOTES ALGORITHM USING NEW LOCAL
CONVEX APPROXIMATION METHODS WITH EXPLICIT
SOLUTIONS

MOSTAFA BACHAR†, THIERRY ESTEBENET ‡, AND ALLAL GUESSAB §

Abstract. In this paper we propose new local convex approximations for solving unconstrained non-linear optimization problems, based on a moving asymptotes algorithm. This method provides the second order information for the moving asymptotes location. As a consequence, at each step of the iterative process, a strictly convex approximation sub-problem is generated and solved. All sub-problems will have explicit global optima, which considerably reduce the computational cost of our optimization method and generate an iteration sequence. For this method, we prove the convergence of the optimization algorithm under basic assumptions. In addition, we present an industrial problem to illustrate a practical application and numerical test of our method.

Key words. Geometric convergence, Nonlinear programming, Method of moving asymptotes, Multivariate convex approximations.

AMS subject classifications. 65K05, 65K10, 65L10, 90C30, 46N10

1. Motivation and theoretical justification. The early work on the so-called method of moving asymptotes (MMA) was introduced for the first time, without any global convergence analysis, by Svanberg [24] in 1987. This method can be seen as a generalization of the CONvex LINearization method (CONLIN), see [5] for instance. Later on, Svanberg [25] proposed a globally convergent, but in reality slow to converge, new method. Since then many different versions have been suggested. For more details on this topic, see the references [6, 30, 33, 34]. For reasons of simplicity, we consider the unconstrained optimization problem: Find \( x^* = (x_{*,1}, x_{*,2}, \ldots, x_{*,d})^\top \in \mathbb{R}^d \) such that

\[
(1.1) \quad f(x^*) = \min_{x \in \mathbb{R}^d} f(x),
\]

where \( x = (x_1, x_2, \ldots, x_d)^\top \in \mathbb{R}^d \) and \( f \) is a given non-linear real-valued objective function, typically twice continuously differentiable. In order to introduce our extension of the early method more clearly, we will first present the most important facet of the original approach. The MMA method generates a sequence of convex and separable sub-problems, which can be solved by any available algorithm taking their special structures into account. The idea behind MMA is the segmentation of the \( d \)-dimensional space into \((d)-one-dimensional \) (1D) spaces.

Given the iteration points \( x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_d^{(k)})^\top \in \mathbb{R}^d \), at the iteration \( k \), \( L_j^{(k)} \) and \( U_j^{(k)} \), are respectively the lower and upper asymptotes, that are adapted at each iteration step, such that, for \( j = 1, \ldots, d \),

\[
L_j^{(k)} < x_j < U_j^{(k)}. \]

†Department of Mathematics, College of Sciences, King Saud University, Riyadh, Saudi Arabia (mbachar@ksu.edu.sa).
‡ALSTOM Ltd., Brown Boveri Strasse 10, CH-5401 Baden, Switzerland (thierry.estebenet@power.alstom.com).
§Laboratoire de Mathématiques et de leurs Applications, UMR CNRS 4152, Université de Pau et des Pays de l’Adour, 64000 Pau, France (allal.guessab@univ-pau.fr).
During the MMA process, the objective function $f$ is iteratively approximated at the $k-$th iteration as follows:

$$\tilde{f}^{(k)}(x) = r^{(k)} + \sum_{j=1}^{d} \frac{p_j^{(k)}}{U_j^{(k)} - x_j} + \frac{q_j^{(k)}}{x_j - L_j^{(k)}}.$$ 

The parameters $r^{(k)}, p_j^{(k)}$ and $q_j^{(k)}$ are adjusted such that the first order approximation is satisfied, i.e.,

$$\tilde{f}^{(k)}(x^{(k)}) = f(x^{(k)})$$
$$\nabla \tilde{f}^{(k)}(x^{(k)}) = \nabla f(x^{(k)}),$$

where $\nabla f(x)$ is the gradient of the objective function $f$ at $x$. The parameter $p_j^{(k)}$ is set to zero when $\frac{\partial \tilde{f}^{(k)}}{\partial x_j}(x^{(k)}) < 0$, and $q_j^{(k)}$ is set to zero when $\frac{\partial \tilde{f}^{(k)}}{\partial x_j}(x^{(k)}) > 0$, such that $\tilde{f}^{(k)}$ is a monotonous increasing or decreasing function of $x_j$. The coefficients $p_j^{(k)}$ and $q_j^{(k)}$ are then given respectively by:

$$p_j^{(k)} = (U_j^{(k)} - x_j^{(k)})^2 \max \left\{ 0, \frac{\partial \tilde{f}^{(k)}}{\partial x_j}(x^{(k)}) \right\}$$
$$q_j^{(k)} = (x_j^{(k)} - L_j^{(k)})^2 \max \left\{ 0, -\frac{\partial \tilde{f}^{(k)}}{\partial x_j}(x^{(k)}) \right\}.$$ 

These parameters are strictly positive, such that all approximating functions $\tilde{f}^{(k)}$ are strictly convex and hence each sub-problem has a single global optimum.

With this technique, the form of each approximated function is derived by the selected values of the parameters $L_j^{(k)}$ and $U_j^{(k)}$, which are chosen by the MMA method.

Several rules for choosing these values are discussed in detail in [24]. Svanberg also shows how the parameters $L_j^{(k)}$ and $U_j^{(k)}$ can be used to control the general process. If the convergence process tends to oscillate, it may be stabilized by moving the asymptotes closer to the current iteration point and if the convergence process is slow and monotonic, it may be relaxed by moving the asymptotes at a limited distance away from their position at the current iteration. Several heuristic rules were also given for an adaptation process for automatic adjustment of these asymptotes at each iteration, see [24, 25]. The most important features of MMA can be summarized as:

- The MMA approximation is a first order approximation at $x^{(k)}$, i.e.,

$$\tilde{f}^{(k)}(x^{(k)}) = f(x^{(k)})$$
$$\nabla \tilde{f}^{(k)}(x^{(k)}) = \nabla f(x^{(k)}).$$

- It is an explicit rational, strictly convex function for all $x$ such that $L_j^{(k)} < x_j < U_j^{(k)}$, with poles (asymptotes in $L_j^{(k)}$ or in $U_j^{(k)}$), and they are monotonic (increasing if $\frac{\partial F(x^{(k)})}{\partial x_j} > 0$ and decreasing if $\frac{\partial F(x^{(k)})}{\partial x_j} < 0$).

- The MMA approximation is separable, which means that the approximation function $F : \mathbb{R}^d \to \mathbb{R}$ can be expressed as the sum of functions of the individual variables, i.e., there exist real functions $F_1, F_2, \cdots, F_d$ such that:

$$F(x) = F_1(x_1) + F_2(x_2) + \cdots + F_d(x_d).$$

Such a property is crucial in practice, because the Hessian matrices of the approximations will be diagonal, and this allows us to address large-scale problems.
New local convex approximations methods with explicit solutions

- It is smooth, i.e., functions \( \tilde{f}^{(k)} \) are twice continuously differentiable in the interval \( L_j^{(k)} < x_j < U_j^{(k)} \), \( j = 1, \ldots, d \).
- At each outer iteration, given the current point \( x^{(k)} \), a sub-problem is generated and solved, and its solution defines the next iteration \( x^{(k+1)} \), so only a single inner iteration is performed.

However, it should be mentioned that this method does not perform well in some cases, and can even fail when the curvature of the approximation is not correctly assigned [23]. Indeed, it is important to realize that all convex approximations including MMA, which are based on first order approximations, do not provide any information about the curvature. The second derivative information is contained in the Hessian matrix of the objective function \( H[f] \), whose \((i,j)\) component is \( \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \). Updating of the moving asymptotes remains a difficult problem. One possible approach is to use the diagonal second derivatives of the objective function in order to define the ideal values of these parameters in the MMA method.

In fact, MMA was extended in order to include the first and second order derivatives of the objective function. For instance, the simple example of MMA that uses a second order approximation at iterate \( x^{(k)} \) was proposed by Fleury [5]:

\[
(1.5) \quad \tilde{f}^{(k)}(x) = f(x^{(k)}) + 
\sum_{j=1}^{d} \left( \frac{1}{x_j^{(k)} - a_j^{(k)}} - \frac{1}{x_j - a_j^{(k)}} \right) (x_j^{(k)} - a_j^{(k)}) \frac{\partial f(x)}{\partial x_j}(x^{(k)}),
\]

where, for each \( j = 1, \ldots, d \), the moving asymptote \( a_j^{(k)} \) determined from the first and second derivatives is defined by:

\[
a_j^{(k)} = x_j^{(k)} + 2 \left( \frac{\partial f(x^{(k)})}{\partial x_j}(x^{(k)}) \right) \frac{\partial^2 f(x^{(k)})}{\partial x_i \partial x_j}(x^{(k)}).
\]

Several versions have been suggested in recent literature to obtain a practical implementation of MMA that takes full advantage of the second order information, e.g., Smaoui et al. [23], Chickermane et al. [4] and the papers cited therein provide additional reading on this topic. The limitations of the asymptote analysis method for first order convex approximations are discussed by Smaoui et al. [23], where an approximation based on second order information is compared with one based on only first order. The second order approximation is shown to achieve the best compromise between robustness and accuracy.

In contrast to the traditional approach, our method replaces the implicit problem (1.1) with a sequence of convex explicit sub-problems having a simple algebraic form, that can be solved explicitly. More precisely, in our method, an outer iteration starts from the current iterate \( x^{(k)} \) and ends up with a new iterate \( x^{(k+1)} \). At each inner iteration, within an explicit outer iteration, a convex sub-problem is generated and solved. In this sub-problem, the original objective function is replaced by a linear function plus a rational function, which approximates the original functions around \( x^{(k)} \). The optimal solution of the sub-problem becomes \( x^{(k+1)} \) and the outer iteration is completed. As in MMA, we will show that our approximation schemes share all of the features listed above. In addition, our explicit iteration method is extremely simple to implement and is easy to use. Furthermore, MMA is very convenient to
use in practice, but its theoretical convergence properties have not been exhaustively studied. This paper presents a detailed study of the convergence properties of the proposed method.

The major motivation of this paper was to propose an approximation scheme which will also be shown to meet all well-known properties of convexity and separability of the MMA. In particular, our proposed scheme provides the following major advantages:

1. An important aspect of our approximation scheme is that all its associated sub-problems have explicit solutions.
2. It generates an iteration sequence, that is bounded and converges to a stationary point of the objective function.
3. It converges geometrically.

The rest of the paper is organized as follows: For clarity of discussion, the one-dimensional case is first considered. To this end, due to the separability of the approximations that we will consider later for multivariate setting, we present our methodology for a single real variable in Section 2. In the follow up to this paper we have shown that the formulation extends to the multidimensional case. Indeed, Section 3 describes the extensions to more general settings of our univariate approach, where an explicit description of the proposed method will be derived and the corresponding algorithm presented. We also show that the proposed method has some favorable convergence properties. In order to avoid the evaluation of second derivatives, we will use a sequence of diagonal Hessian estimations where only the first and zeroth order information are accumulated during the previous iterations. We conclude Section 3 by giving a simple one-dimensional example, which illustrates the performance of our method by showing that it has a wider convergence domain than the classical Newton method. As an illustration, a realistic industrial inverse problem of multistage turbines using a through-flow code will be presented in section 4. Finally, concluding remarks are offered in Section 5.

2. One variable objective function. Since the simplicity of the one-dimensional case allows to detail all the necessary steps with very simple computations, let us first consider the general optimization problem (1.1) of a single real variable. To this end, we first list the necessary notation and terminology.

Let \( d := 1 \) and \( \Omega \subset \mathbb{R} \) be an open subset and \( f : \Omega \mapsto \mathbb{R} \) be a given twice differentiable function in \( \Omega \). Throughout, we assume that \( f' \) does not vanish at a given suitable initial point \( x^{(0)} \in \Omega \), that is \( f'(x^{(0)}) \neq 0 \), since if this is not the case we have nothing to solve. Starting from the initial design point \( x^{(0)} \) the iterates \( x^{(k)} \) are computed successively by solving sub-problems of the form: Find \( x^{(k+1)} \) such that

\[
(2.1) \quad f(x^{(k+1)}) = \min_{x \in \Omega} \tilde{f}^{(k)}(x),
\]

where the approximating function \( \tilde{f}^{(k)} \) of the objective function \( f \) at the \( k \)-th iteration has the following form:

\[
(2.2) \quad \tilde{f}^{(k)}(x) = b^{(k)} + c^{(k)}(x - x^{(k)}) + d^{(k)} \left( \frac{1}{2} \frac{(x^{(k)} - a^{(k)})^3}{x - a^{(k)}} + \frac{1}{2} (x^{(k)} - a^{(k)})(x - 2x^{(k)} + a^{(k)}) \right) 
\]
with

\[ a^{(k)} = \begin{cases} 
  L^{(k)} & \text{if } f'(x^{(k)}) < 0, \text{ and } L^{(k)} < x^{(k)} \\
  U^{(k)} & \text{if } f'(x^{(k)}) > 0, \text{ and } U^{(k)} > x^{(k)},
\end{cases} \]

where, the asymptotes \( U^{(k)} \) and \( L^{(k)} \) are adjusted heuristically as the optimization progresses, or guided by a proposed given function where the first and second derivative are evaluated at the current iteration point \( x^{(k)} \). Also, the approximate parameters \( b^{(k)}, c^{(k)} \) and \( d^{(k)} \) will be determined for each iterations. To evaluate them, we use the objective function value, its first-derivatives as well as its second-derivatives at \( x^{(k)} \). The parameters \( b^{(k)}, c^{(k)} \) and \( d^{(k)} \) are determined in such a way that the following set of interpolation conditions are satisfied:

\[
\begin{align*}
\tilde{f}^{(k)}(x^{(k)}) &= f(x^{(k)}), \\
(\tilde{f}^{(k)})'(x^{(k)}) &= f'(x^{(k)}), \\
(\tilde{f}^{(k)})''(x^{(k)}) &= f''(x^{(k)}).
\end{align*}
\]

Therefore, it is easy to verify that \( b^{(k)}, c^{(k)} \) and \( d^{(k)} \) are explicitly given by

\[
\begin{align*}
b^{(k)} &= f(x^{(k)}), \\
c^{(k)} &= f'(x^{(k)}), \\
d^{(k)} &= f''(x^{(k)}).
\end{align*}
\]

Throughout this section we will assume that

\[ f''(x^{(k)}) > 0, (\forall k \geq 0). \]

Let us now define the notion of feasibility for a sequence of asymptotes \( \{a^{(k)}\} := \{a^{(k)}\}_{k^*} \), which we shall need in the follow up discussion.

**Definition 2.1.** A sequence of asymptotes \( \{a^{(k)}\} \) is said feasible, if for all \( k \geq 0 \), there exist two real numbers \( L^{(k)} \) and \( U^{(k)} \) satisfying the following:

\[
\begin{align*}
a^{(k)} &= \begin{cases} 
  L^{(k)} & \text{if } f'(x^{(k)}) < 0, \text{ and } L^{(k)} < x^{(k)} + 2 \frac{f'(x^{(k)})}{f''(x^{(k)})}, \\
  U^{(k)} & \text{if } f'(x^{(k)}) > 0, \text{ and } U^{(k)} > x^{(k)} + 2 \frac{f'(x^{(k)})}{f''(x^{(k)})}.
\end{cases}
\end{align*}
\]

It is clear from the above definition that every feasible sequence of asymptotes \( \{a^{(k)}\} \) automatically satisfies all the constraints of type (2.3).

The following proposition, which is easily obtained by a simple algebraic manipulation, shows that the lower bound of the difference between the asymptotes and the current iterate \( x^{(k)} \), can be estimated as in (2.7).

**Proposition 2.2.** Let \( \{a^{(k)}\} \) be a sequence of asymptotes and let the assumptions (2.3) valid. Then, \( \{a^{(k)}\} \) is feasible if and only if

\[
2 \left| \frac{f'(x^{(k)})}{f''(x^{(k)})} \right| < \left| x^{(k)} - a^{(k)} \right|.
\]

It is interesting to note that our approximation scheme can be seen as an extension of Fleury’s method [17]. Indeed, we have the following:
Remark 2.3. Back to the approximations \( \hat{f}^{(k)} \), given in (2.2), if we write

\[
\hat{a}^{(k)} = x^{(k)} + \frac{2f'(x^{(k)})}{f''(x^{(k)})},
\]

using the values of the parameters given in (2.5), the approximating functions \( \tilde{f}^{(k)} \) can also be rewritten as:

\[
\tilde{f}^{(k)}(x) = f(x^{(k)}) + \frac{f''(x^{(k)})}{2} (\hat{a}^{(k)} - a^{(k)}) (x - x^{(k)}) + \frac{f''(x^{(k)})}{2} (x^{(k)} - a^{(k)})^3 r^{(k)}(x),
\]

(2.8)

with

\[
\ r^{(k)}(x) = \left( \frac{1}{x - a^{(k)}} - \frac{1}{x^{(k)} - a^{(k)}} \right).
\]

If we choose \( \hat{a}^{(k)} = a^{(k)} \) then the approximating functions become:

\[
\tilde{f}^{(k)}(x) = f(x^{(k)}) + \left( \frac{1}{x^{(k)} - a^{(k)}} - \frac{1}{x - a^{(k)}} \right) (x^{(k)} - a^{(k)})^2 f'(x^{(k)}).
\]

(2.9)

This is exactly the one-dimensional version of the approximation functions of Fleury given by equations (1.5). Hence, our approximation can be seen as a natural extension of Fleury’s method [17].

The following lemma summarizes the basic properties of feasible sequences of asymptotes. In what follows, we denote \( \text{sign}(\cdot) \) the usual sign function.

Lemma 2.4. If \( \{a^{(k)}\} \) is a feasible sequence of asymptotes, then for all \( k \) the following statements are true:

i) \( \frac{\text{sign}(f'(x^{(k)}))}{x^{(k)} - a^{(k)}} = -\frac{1}{|x^{(k)} - a^{(k)}|}. \)

ii) \( \frac{x^{(k)} - a^{(k)} + 2f'(x^{(k)})}{f''(x^{(k)})} = \frac{|x^{(k)} - a^{(k)}|}{|x^{(k)} - a^{(k)}|}. \)

iii) At each iteration, the first derivative of the approximating function \( \tilde{f}^{(k)} \) is given by:

\[
(\tilde{f}^{(k)})'(x) = \frac{f''(x^{(k)})}{2} (x^{(k)} - a^{(k)}) (e[f](x^{(k)}) - \frac{x^{(k)} - a^{(k)}}{x - a^{(k)}})^2,
\]

(2.10)

with

\[
\ e[f](x^{(k)}) := \frac{|x^{(k)} - a^{(k)}| - 2|f'(x^{(k)})|}{f''(x^{(k)})}.
\]

Proof. The proof of i) is straightforward, since it is an immediate consequence of the fact that the sequence of asymptotes \( \{a^{(k)}\} \) is feasible. We will only give a sketch of the proof of parts ii) and iii). By i) and the obvious fact that:

\[
(2.12)
\]

\[ f'(x^{(k)}) = \text{sign}(f'(x^{(k)})) \frac{2|f'(x^{(k)})|}{f''(x^{(k)})}, \]
we have
\[
\frac{x^{(k)} - a^{(k)}}{x^{(k)} - a^{(k)}} + 2 \frac{f'(x^{(k)})}{f''(x^{(k)})} = 1 + 2 \frac{f'(x^{(k)})}{f''(x^{(k)})} \left| \frac{\text{sign}(f'(x^{(k)}))}{x^{(k)} - a^{(k)}} \right|
\]
\[(2.13)\]
\[
= 1 - \frac{2f'(x^{(k)})}{f''(x^{(k)})} \frac{1}{|x^{(k)} - a^{(k)}|}
\]
\[
= \frac{|x^{(k)} - a^{(k)}| - 2f'(x^{(k)})}{|x^{(k)} - a^{(k)}|}
\]

Finally, part iii) is a consequence of part ii) and the new expression of \( \tilde{f}^{(k)} \) given in (2.8).

By defining the suitable index set
\[
\mathcal{I}^{(k)} = \{ [L^{(k)}, +\infty) \mid \text{if } f'(x^{(k)}) < 0 \}, \quad \mathcal{I}^{(k)} = \{ -\infty, U^{(k)} \mid \text{if } f'(x^{(k)}) > 0 \}
\]

we now are able to define our iterative sequence \( \{ x^{(k)} \} \). We still assume that \( f \) is a twice differentiable function in \( \Omega \) satisfying \( f''(x^{(k)}) > 0 \), \( \forall k \geq 0 \).

**Theorem 2.5.** In the above notation, let \( \Omega \subset \mathbb{R} \) be an open subset of the real line, \( x_0 \in \Omega \) and \( x^{(k)} \) being respectively the initial and the current point of the sequence \( \{x^{(k)}\} \). Let the choice of the sequence of asymptotes \( \{a^{(k)}\} \) be feasible. Then, for each \( k \geq 0 \) the approximated function \( \tilde{f}^{(k)} \) defined by (2.2) is a strictly convex function in \( \mathcal{I}^{(k)} \). Furthermore, for each \( k \geq 0 \), the function \( \tilde{f}^{(k)} \) attains its minimum at
\[
x^{(k+1)} = a^{(k)} - \text{sign} \left(f'(x^{(k)})\right) \sqrt{g^{(k)}},
\]
where
\[
g^{(k)} := \frac{|x^{(k)} - a^{(k)}|^3}{|x^{(k)} - a^{(k)}| - 2f'(x^{(k)})/f''(x^{(k)})}
\]

**Proof.** An important characteristic of our approximate problem obtained via the approximation function \( \tilde{f}^{(k)} \) is its strict convexity in \( \mathcal{I}^{(k)} \). To prove the strict convexity of the approximation, we have to show that \( (\tilde{f}^{(k)})'' \) is non-negative in \( \mathcal{I}^{(k)} \). Indeed, by a simple calculation of the second derivatives of \( \tilde{f}^{(k)} \), we have:
\[
(\tilde{f}^{(k)})''(x) = f''(x^{(k)}) \left( \frac{x^{(k)} - a^{(k)}}{x - a^{(k)}} \right)^3.
\]
Hence, to prove the convexity of \( \tilde{f}^{(k)} \), we have to show that
\[
f''(x^{(k)}) \left( \frac{x^{(k)} - a^{(k)}}{x - a^{(k)}} \right)^3 > 0, \quad \forall x \in \mathcal{I}^{(k)}.
\]
But \( f''(x^{(k)}) > 0 \) and so, according to the definition of the set \( \mathcal{I}^{(k)} \), it follows that \( x^{(k)} - a^{(k)} \) and \( x - a^{(k)} \) have the same sign in the interval \( \mathcal{I}^{(k)} \), then we immediately get the strict convexity of \( \tilde{f}^{(k)} \) on \( \mathcal{I}^{(k)} \). Furthermore, according to (2.10) if \( \tilde{f}^{(k)} \) attains its minimum at \( x^{(k)}_* \), then it is easy to see that \( x^{(k)}_* \) is solution of the equation
\[
\left( \frac{x^{(k)} - a^{(k)}}{x - a^{(k)}} \right)^2 = \frac{|x^{(k)} - a^{(k)}| - 2f'(x^{(k)})/f''(x^{(k)})}{|x^{(k)} - a^{(k)}|}.
\]
Note that Proposition 2.2 ensures that the numerator of the last term on the right hand side is strictly positive. Now, by taking the square root and using a simple transformation we can see that the unique solution $x^*_k$ belonging to $I^k$ is given by (2.15). This completes the proof of the theorem.

**Remark 2.6.** At this point, we should remark that the notion of feasibility for a sequence of moving asymptotes, as defined in Definition 2.1, plays an important role for the existence of the explicit minimum, given by (2.15) of the approximate function $f^k$, related to each sub-problem belonging to $I^k$. More precisely, it guarantees the positivity of the numerator of the fraction in the right hand side of (2.18) and hence ensures the existence of a single global optimum for the approximate function at each iteration.

We now give a short discussion about an extension to the above approach. Our study in this section has been in a framework that at each iteration the second derivative needs to be evaluated exactly. We will now focus our analysis by examining what happens when the second derivative of the objective function $f$ may not be known or expensive to evaluate. Thus, in order to reduce the computational effort, we suggest to approximate at each iteration the second derivative $f''(x^k)$ by some positive real value $s^k$. In this situation, we shall propose the following procedure for selecting moving asymptotes:

\[
\hat{a}^k = \begin{cases} 
L^k & \text{if } f'(x^k) < 0, \text{ and } L^k < x^k + 2\frac{f(x^k)}{s^k}, \\
U^k & \text{if } f'(x^k) > 0, \text{ and } U^k > x^k + 2\frac{f(x^k)}{s^k}.
\end{cases}
\]

It is clear that all the previous results are easily carried over in the case where in the interpolation conditions (2.4), $f''(x^k)$ is replaced by an approximate (strictly) positive value $s^k$, according to constraints (2.19). Indeed, the statements of Theorem 2.5 apply with straightforward changes.

In Section 3 for the multivariate case, we will discuss a strategy to determine at each iteration a reasonably good numerical approximation to the second derivative. We will also establish a multivariate version of Theorem 2.5 and show in this setting a general convergence result.

3. The Multivariate Setting. To develop our methods for the multivariate case, we need to replace the approximating functions (2.2) of the univariate objective function by using suitable strictly convex multivariate approximating functions. The practical implementation of this method is considerably more complex than in the univariate case, due to the fact that, at each iteration, the approximating function in the multivariate setting generates a sequence of diagonal Hessian estimates. In this Section, as well as in the univariate objective approximating function presented in Section 2, the function value $f(x^k)$, the first-order derivatives $\frac{\partial f(x^k)}{\partial x_j}$, $j = 1 \ldots d$, as well as the second-order information and the moving asymptotes at the design point $x^k$, were used to build up our approximation. To reduce the computational cost, the Hessian of the objective function at each iteration will be replaced by a sequence of diagonal Hessian estimates. These approximate matrices use only zero and first order information accumulated during the previous iterations. However, in view of practical difficulties of evaluating the second-order derivatives, a fitting algorithmic scheme is proposed in order to adjust the curvature of the approximation.
The purpose of the first part of this section is to give a complete discussion on the theoretical aspects, concerning the multivariate setting of the convergence result established in Theorem 3.4, and to expose the computational difficulties that may be incurred. We will first describe the setup and notations for our approach. Below, we comment on the relationships between the new method and several of the most closely-related ideas. Our approximation scheme leaves, as in one-dimensional case, all well-known properties of convexity and separability of the MMA unchanged, with the following major advantages:

1. All our sub-problems have explicit solutions.
2. It generates an iteration sequence, that is bounded and converges to a local solution.
3. It converges geometrically.

To simplify the notation, for every \(j = 1, \ldots, d\), we use the following notation \(f_j\) to denote the first order partial derivative of \(f\) with respect to each variable \(x_j\). We also use the notation \(f_{i,j}\) for the second order partial derivatives with respect to \(x_i\) first and then \(x_j\). For any \(x, y \in \mathbb{R}^d\), we will denote the standard inner product of \(x\) and \(y\) by \((x, y)\), and \(\|x\| := \sqrt{(x, x)}\), the Euclidean vector norm of \(x \in \mathbb{R}^d\).

### 3.1. The convex approximation in \(\Omega \subset \mathbb{R}^d\).

To build up the approximate optimization sub-problems \(P[k]\), taken into account the approximate optimization problem as a solution strategy of the optimization problem (1.1), we will seek to construct a successive approximate of sub-problems \(P[k], k \in \mathbb{N}\), at successive iteration points \(x^{(k)}\). That is, at each iteration \(k\), we shall seek a suitable explicit rational approximating function \(\tilde{f}^{(k)}\), strictly convex and relatively easy to implement. The solution of the sub-problems \(P[k]\) is denoted by \(x^{*(k)}\), and will be obtained explicitly. The optimum \(x^{*(k)}\) of the sub-problems \(P[k]\), will be considered as the starting point \(x^{(k+1)} := x^{*(k)}\) for the next subsequent approximate sub-problems \(P[k+1]\).

Therefore, for a given suitable initial approximation \(x^{(0)} \in \Omega\), the approximate optimization sub-problems \(P[k], k \in \mathbb{N}\), of a successive iteration points \(x^{(k)} \in \mathbb{R}^d\) can be written as: Find \(x^{*(k)}\) such that

\[
\tilde{f}^{(k)}(x^{*(k)}) := \min_{x \in \Omega} \tilde{f}^{(k)}(x),
\]

where the approximating function is defined by:

\[
(3.1) \quad \tilde{f}^{(k)}(x) = \sum_{j=1}^{d} \left( \frac{(\alpha^{(k)})_j}{x_j - L^{(k)}_j} + \frac{(\alpha^{(k)})_j}{U^{(k)}_j - x_j} \right) + \langle \beta^{(k)}_{-}, x - L^{(k)} \rangle + \langle \beta^{(k)}_{+}, U^{(k)} - x \rangle + \gamma(k),
\]

the coefficients \(\beta^{(k)}_{-}, \beta^{(k)}_{+}, L^{(k)}, U^{(k)}\) are given by:

\[
(3.2) \quad \beta^{(k)}_{-} = \left( (\beta^{(k)}_{-})_1, \ldots, (\beta^{(k)}_{-})_d \right)^T,
\]

\[
\beta^{(k)}_{+} = \left( (\beta^{(k)}_{+})_1, \ldots, (\beta^{(k)}_{+})_d \right)^T,
\]

\[
L^{(k)} = \left( L^{(k)}_1, \ldots, L^{(k)}_d \right)^T,
\]

\[
U^{(k)} = \left( U^{(k)}_1, \ldots, U^{(k)}_d \right)^T.
\]
and $\gamma^{(k)} \in \mathbb{R}$. They represent the unknown parameters that need to be computed based on the available information. In order to ensure that the functions $\tilde{f}^{(k)}$ have suitable properties discussed earlier, we will assume that the following conditions (3.3) are satisfied for all $k$:

$$
\begin{align*}
(\alpha_-^{(k)})_j &= (\beta_-^{(k)})_j = 0 \quad \text{if } f_{,j}(x^{(k)}) > 0, \\
(\alpha_+^{(k)})_j &= (\beta_+^{(k)})_j = 0 \quad \text{if } f_{,j}(x^{(k)}) < 0, \quad j = 1, \ldots, d.
\end{align*}
$$

Our approximation can be viewed as a generalization of the univariate approximation to the multivariate case since the approximation functions $\tilde{f}^{(k)}$ are of the form of a linear function plus a rational function. It can be easily checked that the first and the second order derivatives of $\tilde{f}^{(k)}$ have the following forms:

$$
\begin{align*}
(\alpha_-^{(k)})_j &= (\beta_-^{(k)})_j = 0 \quad \text{if } f_{,j}(x^{(k)}) > 0, \\
(\alpha_+^{(k)})_j &= (\beta_+^{(k)})_j = 0 \quad \text{if } f_{,j}(x^{(k)}) < 0, \quad j = 1, \ldots, d.
\end{align*}
$$

Now, making use of (3.3), these observations imply that if $f_{,j}(x^{(k)}) > 0$ then

$$
\tilde{f}^{(k)}_{,jj}(x) = \frac{2 (\alpha_+^{(k)})_j}{(x_j - L_j^{(k)})^3} + \frac{2 (\alpha_-^{(k)})_j}{(U_j^{(k)} - x_j)^3}, \quad j = 1, \ldots, d.
$$

and if $f_{,j}(x^{(k)}) < 0$ then,

$$
\tilde{f}^{(k)}_{,jj}(x) = \frac{2 (\alpha_-^{(k)})_j}{(x_j - L_j^{(k)})^3}.
$$

Since the approximations $\tilde{f}^{(k)}$ are separable functions, then all the mixed second derivatives of $f$ are identically zero. Therefore, if $i \neq j$ then we have

$$
\tilde{f}^{(k)}_{,ij}(x) = 0, \quad i, j = 1, \ldots, d.
$$

Also, the approximating functions $\tilde{f}^{(k)}$ need to be identically equal to the first order approximations of the objective functions $f$ at the current iteration point $x = x^{(k)}$, i.e.,

$$
\begin{align*}
\tilde{f}^{(k)}(x^{(k)}) &= f(x^{(k)}), \\
\tilde{f}^{(k)}_{,j}(x^{(k)}) &= f_{,j}(x^{(k)}), \quad \forall j = 1, \ldots, d.
\end{align*}
$$

In addition to the above first order approximations, the approximating function $\tilde{f}^{(k)}$ should include the information of the second-order derivatives $f$. Indeed, the proposed approximation will be improved, if we impose that

$$
\tilde{f}^{(k)}_{,jj}(x) = f_{,jj}(x), \quad \forall j = 1, \ldots, d.
$$
Since the second derivatives of the original functions $f$ may not be known or expensive to evaluate, the above interpolation conditions (3.10) are not in general satisfied. However, it makes sense to use second order derivatives information, to improve the convergence speed.

The strategy of using the second order information without excessive effort consists of approximating at each iteration the Hessian $H^{(k)}[f] := [f_{ij} (x^{(k)})]$ by a simple structure of an easy calculated matrix.

Our choice for approximating the derivatives is based on the spectral parameters as detailed in [12], where the Hessian of the function $f$ is approximated by the diagonal matrix $S^{(k)} I$ (i.e. $n^{(k)} I$ in [12]), with $I$ the $d$-by-$d$ identity matrix, and the coefficients $S^{(k)}_{jj}$ are simply chosen such that

$$S^{(k)}_{jj} := \frac{d^{(k)}}{\|x^{(k)} - x^{(k-1)}\|^2} \approx f_{,jj} (x^{(k)}),$$

where

$$d^{(k)} := \langle \nabla f(x^{(k)}) - \nabla f(x^{(k-1)}), x^{(k)} - x^{(k-1)} \rangle > 0.$$

The last conditions (3.12) ensure that the approximations $\tilde{f}^{(k)}$ are strictly convex for all iterates $x^{(k)}$, since the parameters $S^{(k)}_{jj}$ are chosen strictly positive.

Thus, if we use the three identities (3.6), (3.7), (3.8) and the above approximation conditions, we get immediately, after some manipulations,

\begin{equation}
\alpha^{(k)}_j := \begin{cases} \frac{1}{2} S^{(k)}_{jj} (x^{(k)}_j - L^{(k)}_j)^3 & \text{if } f_{,j} (x^{(k)}) < 0 \\ 0 & \text{otherwise} \end{cases}
\end{equation}

\begin{equation}
\beta^{(k)}_j := \begin{cases} \frac{1}{2} S^{(k)}_{jj} (U^{(k)}_j - x^{(k)}_j)^3 & \text{if } f_{,j} (x^{(k)}) > 0 \\ 0 & \text{otherwise} \end{cases}
\end{equation}

and

\begin{equation}
\gamma^{(k)} := f(x^{(k)}) - \sum_{j=1}^d \left( \frac{\alpha^{(k)}_j}{x^{(k)}_j - L^{(k)}_j} + \frac{\alpha^{(k)}_j}{U^{(k)}_j - x^{(k)}_j} \right) - \langle \beta^{(k)}_-, x^{(k)} - L^{(k)} \rangle - \langle \beta^{(k)}_+, U^{(k)} - x^{(k)} \rangle.
\end{equation}

Our strategy will be to update the lower and upper moving asymptotes, $L^{(k)}_j$ and $U^{(k)}_j$ at each iteration based on second order information, by generalizing the definition 2.1 from
Furthermore, \( f(x) \) defined by (3.20). Then the objective function equations (3.18), where \( S_{jj}^{(k)} \) given in (3.11) as follows:

\[
A_j^{(k)} = \begin{cases} 
L_j^{(k)} & \text{if } f_j(x^{(k)}) < 0, \text{ and } L_j^{(k)} < x_j^{(k)} + \frac{2f_j(x^{(k)})}{S_{jj}^{(k)}}, \\
U_j^{(k)} & \text{if } f_j(x^{(k)}) > 0, \text{ and } U_j^{(k)} > x_j^{(k)} + \frac{2f_j(x^{(k)})}{S_{jj}^{(k)}}.
\end{cases}
\]

Note that, as it was done in the univariate case, see Proposition 2.2, we have the following result.

**Proposition 3.1.** Let \( A^{(k)} = (A_1^{(k)}, A_2^{(k)}, \ldots, A_d^{(k)})^T \in \mathbb{R}^d \) be the moving asymptotes with the components given by (3.18). Then, for all \( j = 1, \ldots, d \) and for all \( k \) we have:

\[
\frac{2|f_j(x^{(k)})|}{S_{jj}^{(k)}} < |x_j^{(k)} - A_j^{(k)}|.
\]

To define our multivariate iterative scheme, we start from some given suitable initial approximation \( x^{(0)} \in \Omega \), and let \( \{x^{(k)}\} = \{x^{(k)}\}_k \) be the iterative sequence defined by:

\[
x^{(k+1)} = (x_1^{(k+1)}, \ldots, x_d^{(k+1)})^T, \text{ for all } k \geq 0, \text{ and } j = 1 \ldots, d,
\]

\[
x_j^{(k+1)} = A_j^{(k)} + \text{sign} \left( f_j(x^{(k)}) \right) \sqrt{g_j^{(k)}}, \ (j = 1, \ldots, d),
\]

where

\[
g_j^{(k)} = \frac{2f_j(x^{(k)})}{S_{jj}^{(k)}} - |x_j^{(k)} - A_j^{(k)}|^3 = \begin{cases} 
\frac{(\alpha^{(k)})_j}{(\beta^{(k)})_j} & \text{if } f_j(x^{(k)}) < 0, \\
\frac{(\alpha^{(k)})_j}{(\beta^{(k)})_j} & \text{if } f_j(x^{(k)}) > 0.
\end{cases}
\]

It should be pointed out that the sequence \( \{x^{(k)}\} \) is well-defined for all \( k \), since the denominators of (3.20) never vanish and it is straightforward to see that \( g_j^{(k)} \) in (3.21) are positive real numbers.

It would be more precise to use the set notation and write: \( T^{(k)} = T_1^{(k)} \times T_2^{(k)} \times \cdots \times T_d^{(k)}, \)

with

\[
T_j^{(k)} = \begin{cases} 
L_j^{(k)} & , +\infty \\
-\infty, U_j^{(k)} & , j = 1, \ldots, d.
\end{cases}
\]

Now we are in a position to present one main result of this paper.

**Theorem 3.2.** Let \( \Omega \) be a given open subset of \( \mathbb{R}^d \), \( f : \Omega \to \mathbb{R} \) twice-differentiable objective function in \( \Omega \). We assume that the moving asymptotes \( A^{(k)} \in \mathbb{R}^d \) are defined by the equations (3.18), where \( S_{jj}^{(k)} > 0, k \geq 0, j = 1, \ldots, d \), and let \( \{x^{(k)}\} \) be the iterative sequence defined by (3.20). Then the objective function \( \tilde{f}^{(k)} \) defined by the equation (3.1), with the coefficients (3.13-3.16), is a first order strictly convex approximation of \( f \) that satisfies:

\[
f_{j,j}^{(k)}(x^{(k)}) = S_{jj}^{(k)}, \ j = 1, \ldots, d.
\]

Furthermore, \( f^{(k)} \) attains its minimum at \( x^{(k+1)} \).

**Proof.** By construction, the approximation \( \tilde{f}^{(k)} \) is a first order approximation of \( f \) at \( x = x^{(k)} \) and satisfies:

\[
f_{j,j}^{(k)}(x^{(k)}) = S_{jj}^{(k)}, \forall j = 1, \ldots, d.
\]
As \((\alpha^{(k)})_j\) (respectively \((\alpha^{(k)})_j\)) have the same sign than \(x_j - I_j^{(k)}\) (respectively \(U_j^{(k)} - x_j\)) in \(I^{(k)}\), we can easily deduce from (3.5) that the approximation is strictly convex in \(I^{(k)}\). In addition, by using (3.4), we may verify that \(x_j^{(k+1)}\) given by (3.20) is the unique solution in \(I^{(k)}\) of the equations

\[(3.25) \quad f_j^{(k)}(x) = 0, \forall j = 1, \ldots, d,\]

and thereby completes the proof of the Theorem. \(\square\)

The sequence of sub-problems generated by (3.20) is computed by the Algorithm 3.3.

**Algorithm 3.3. Method of the moving asymptotes with spectral updating**

**Step 1. Initialization**

Define \(x^{(0)}\)
Set \(k \leftarrow 0\)

**Step 2. Stopping criterion**

If \(x^{(k)}\) verifies convergence criteria conditions of the problem (1.1),
Stop and take \(x^{(k)}\) as the solution.

**Step 3. Computation of the spectral parameters \(S_j^{(k)}\), the moving asymptotes \(A_j^{(k)}\) and the intermediate parameter \(g_j^{(k)}\)**

Compute
\[
d^{(k)} = \langle \nabla f(x^{(k)}), x^{(k)} - x^{(k-1)} \rangle,
\]
For \(j = 0, 1, \ldots, d\)
\[
S_j^{(k)} = \|x^{(k)} - x^{(k-1)}\|^2,
\]
\[
A_j^{(k)} = x_j^{(k)} + 2\alpha \frac{f_j^{(k)}(x^{(k)})}{S_j^{(k)}}, \quad \alpha > 1,
\]
\[
g_j^{(k)} = \frac{\|x_j^{(k)} - A_j^{(k)}\|^3}{\|x_j^{(k)} - A_j^{(k)}\|^2 - 2\frac{f_j^{(k)}(x^{(k)})}{S_j^{(k)}}}.
\]

**Step 4. Computation of the solution of the sub-problem**

\[
x_j^{(k+1)} = A_j^{(k)} - \text{sign} \left( f_j^{(k)}(x^{(k)}) \right) \sqrt{d^{(k)}} g_j^{(k)} \text{ for } j = 0, 1, \ldots, d,
\]
Set \(k \leftarrow k + 1\)
Go to Step 2.

**3.2. A multivariate convergence result.** This subsection aims to show that the proposed method is convergent, in the sense that the optimal iterative sequence \(\{x^{(k)}\}\) generated by Algorithm 3.3 converges geometrically to \(x^*\). That is, there exists an \(\xi \in (0, 1]\) such that

\[(3.26) \quad \|x^{(k)} - x^*\| \leq \frac{\xi^k}{1 - \xi} \|x^{(1)} - x^{(0)}\|.
\]

To this end, the following assumptions are required. Let us suppose there exist positive constants \(r, M, C\) and \(\xi < 1\) such that the following assumptions hold.

**Assumption M1:**

\[B_r := \{x \in \mathbb{R} : \|x - x^{(0)}\| \leq r\} \subset \Omega.\]
Assumption M2: We assume that the sequence of moving asymptotes \( \{A^{(k)}\} \) defined by (3.18) satisfies
\[
\sup_{k \geq 0} \|x^{(k)} - A^{(k)}\| \leq C.
\]
and for all \( j = 1, \ldots, d \),
\[
\frac{2C\sqrt{d}}{MS_j^{(k)}} \leq \left| x_j^{(k)} - A_j^{(k)} \right| - \frac{2f_j(x^{(k)})}{S_j^{(k)}}.
\]

Assumption M3: We require that for all \( k > 0 \), and for all \( j \in \{1, \ldots, d\} \) such that \( x_j^{(k-1)} \neq x_j^{(k)} \), we have
\[
\sup_{k > 0} \sup_{x \in B} \|\nabla f_j(x) - \frac{f_j(x^{(k-1)})}{x_j^{(k-1)} - x_j^{(k)}} e^{(j)}\| \leq \frac{\xi}{M},
\]
where \( e^{(j)} \) is the vector of \( \mathbb{R}^d \) with \( j \)-th component equal to 1 and all other components equal to 0.

Assumption M4: For all \( j = 1, \ldots, d \), the initial iterate \( x^0 \) satisfies:
\[
0 < |f_j(x^0)| \leq \frac{r}{M} (1 - \xi).
\]

Let us briefly make some comments on these assumptions:

- First, in order to control the feasibility of moving asymptotes we need to find a (strictly) positive lower bound of
\[
\left| x_j^{(k)} - A_j^{(k)} \right| - \frac{2f_j(x^{(k)})}{S_j^{(k)}}.
\]
which needs to be large according to some predetermined tolerance, see Proposition 3.1. So, when equations (3.28) hold, then the sequence of the moving asymptotes \( \{A^{(k)}\} \) is automatically feasible. Also note that, when we evaluate the approximate function \( \tilde{f}^{(k)} \) if the difference between the asymptotes and the current iteration point is eventually small enough, then imposing condition (3.28) avoids the possibility of (3.31) to become negative or close to zero. In the assumption M2, the equation (3.27) enforces the quite natural conditions that we have: At each iteration \( k \), the distance between \( x^{(k)} \) and the asymptote \( A^{(k)} \) is bounded above by some constant.

- The assumption M3 tells us that \( \nabla f_j(x) \) is sufficiently close to \( \frac{f_j(x^{(k-1)})}{x_j^{(k-1)} - x_j^{(k)}} e^{(j)} \).

- The assumption M4, as we will see, is only used to obtain uniqueness of the limit of the iteration sequence generated by Theorem 3.2. The convergence result is established without this assumption. It also says that \( |f_j(x^0)| \) is small enough and that \( f_j(x^0) \) is not equal to 0. This assumption will also play an important role to show that \( \nabla f \) has a unique zero in \( B_r \).

The assumptions M2 and M3 will be used in conjunction with the assumption M4 to prove that the sequence of iteration points \( \{x^{(k)}\} \) defined by (3.20) has various nice properties and converges geometrically to the unique zero of \( \nabla f \) in \( B_r \). In addition, note that the constant \( C \) ensures that the distances between the current points \( x^k \) and moving asymptotes are finite, and the constant \( M \) ensures that the process starts reasonably close to the solution.
We are now prepared to state and to show our main convergence result.

**Theorem 3.4.** Given the assumptions M1-M4, the sequence \( \{x^{(k)}\} \) defined in (3.20) is completely contained in the closed ball \( B_r \), and converges geometrically to the unique stationary point of \( f \) belonging to the ball \( B_r \).

Before we show Theorem 3.4, we present some preparatory lemmas. The first key ingredient is the following simple observation:

**Lemma 3.5.** Let \( k \) be a fixed positive integer. Assume that there exists \( j \in \{1, \ldots, d\} \) such that \( f_{j}(x^{(k-1)}) \neq 0 \), then the \( j \)-th components of the two successive iterates \( x^{(k)} \) and \( x^{(k-1)} \) are distinct.

**Proof.** Indeed, assume the contrary, that is \( x^{(k)}_j = x^{(k-1)}_j \). Then from equation (3.20), we have

\[
(x^{(k-1)} - A^{(k-1)}_j)^2 = (x^{(k)} - A^{(k-1)}_j)^2 = g^{(k-1)}_j
\]

or equivalently \( f_{j}(x^{(k-1)}) = 0 \), which gives a contradiction and proves the lemma. \( \Box \)

**Remark 3.6.** The above Lemma says that if the \( j \)-th partial derivative of \( f \) does not vanish at the iterate \( x^{(k-1)} \), then the required condition in assumption M4 is satisfied.

We will also need to prove a useful lemma, which bounds the distance between two consecutive iterates \( x^{(k-1)} \) and \( x^{(k)} \).

**Lemma 3.7.** Let assumptions M2-M4 be satisfied and let the sequence \( \{x^{(k)}\} \) defined as in equation (3.20). Then, the following inequalities hold for all positive integers \( k \) and \( j = 1, \ldots, d \),

\[
| x^{(k)}_j - x^{(k-1)}_j | \leq \frac{M}{\sqrt{d}} | f_{j}(x^{(k-1)}) |
\]

\[
\| x^{(k)} - x^{(k-1)} \| \leq M \max_{1 \leq j \leq d} | f_{j}(x^{(k-1)}) |
\]

**Proof.** Let us fix an integer \( k \) such that \( k > 0 \). Then using (3.20), \( x^{(k)}_j - x^{(k-1)}_j \) can be written in the form:

\[
x^{(k)}_j - x^{(k-1)}_j = A^{(k-1)}_j - \text{sign} \left( f_{j}(x^{(k-1)}) \right) \sqrt{g^{(k-1)}_j} - x^{(k-1)}_j
\]

\[
= (x^{(k-1)}_j - A^{(k-1)}_j) (-1 + \Delta),
\]

where, in the last equality, we have denoted

\[
\Delta = \frac{-\text{sign}(f_{j}(x^{(k-1)}))}{x^{(k-1)}_j - A^{(k-1)}_j} \sqrt{g^{(k-1)}_j}.
\]

Now, as in one dimension, see Lemma 2.4, it is easy to check that

\[
\frac{\text{sign}(f_{j}(x^{(k-1)}))}{x^{(k-1)}_j - A^{(k-1)}_j} = -\frac{1}{x^{(k-1)}_j - A^{(k-1)}_j},
\]

consequently, \( \Delta \) can also be expressed in fraction form:

\[
\Delta = \frac{\sqrt{g^{(k-1)}_j}}{x^{(k-1)}_j - A^{(k-1)}_j}.
\]
Since

\[(3.38) \quad g_{j}^{(k-1)} := \frac{x_{j}^{(k-1)} - A_{j}^{(k-1)}}{|x_{j}^{(k-1)} - A_{j}^{(k-1)}|^{3}}\]

it follows from (3.37) that

\[(3.39) \quad \left| \frac{x_{j}^{(k)} - x_{j}^{(k-1)}}{|x_{j}^{(k-1)} - A_{j}^{(k-1)}|^{2}} \right| = \frac{2\|f_{j}(x^{(k-1)})\|}{S_{jj}^{(k-1)}}.\]

Taking into account that \(\tilde{g}_{j}^{(k-1)} > 1\), then, using square root property, we get

\[\sqrt{\tilde{g}_{j}^{(k-1)}} < \tilde{g}_{j}^{(k-1)}.\]

Therefore, by (3.39), we conclude that

\[(3.40) \quad \left| \frac{x_{j}^{(k)} - x_{j}^{(k-1)}}{|x_{j}^{(k-1)} - A_{j}^{(k-1)}|} \right| = \frac{2\|f_{j}(x^{(k-1)})\|}{S_{jj}^{(k-1)}}.\]

We then obtain the desired conclusion by using assumptions M2. The second inequality in Lemma 3.7 is now an immediate consequence of the definition of the Euclidean norm. 

Now, we are ready to prove Theorem 3.4.

**Proof of Theorem 3.4:** Given a fixed positive integer \(k\), let us pick up any integer \(j\) between 1 and \(d\). We start by showing the following inequality

\[(3.41) \quad \left| \frac{x_{j}^{(k)} - x_{j}^{(k-1)}}{|x_{j}^{(k-1)} - A_{j}^{(k-1)}|} \right| \leq \frac{2\|f_{j}(x^{(k-1)})\|}{S_{jj}^{(k-1)}}.\]

To see this, we may distinguish two cases:

- **Case I:** \(x_{j}^{(k-1)} \neq x_{j}^{(k)}\). Let us set

\[(3.42) \quad \tilde{\beta}_{j}^{(k-1)} = \frac{1}{2} \left[ \frac{x_{j}^{(k-1)} - A_{j}^{(k-1)}}{S_{jj}^{(k-1)}} \right] - f_{j}(x^{(k-1)}),\]

and let us introduce the auxiliary function \(\varphi : B_{r} \rightarrow \mathbb{R}\) as follows:

\[(3.43) \quad \varphi(x) = f_{j}(x) - \frac{f_{j}(x^{(k-1)})}{2S_{jj}^{(k-1)}} h(x_{j}),\]

where

\[h(x_{j}) := -\frac{1}{2} \left[ \frac{x_{j}^{(k-1)} - A_{j}^{(k-1)}}{S_{jj}^{(k-1)}} \right] - f_{j}(x^{(k-1)}) - \tilde{\beta}_{j}^{(k-1)}.\]

Using equation (3.43), it is easy to verify that

\[h(x_{j}^{(k-1)}) = \frac{1}{2} S_{jj}^{(k-1)} \left[ x_{j}^{(k)} - x_{j}^{(k-1)} \right],\]

\[h(x_{j}^{(k)}) = 0.\]
consequently \( \varphi \) satisfies
\[
\varphi(x^{(k-1)}) = 0 \\
\varphi(x^{(k)}) = f_j(x^{(k)}).
\]

Also it is easy to see that
\[
\nabla \varphi(x) = \nabla f_j(x) - \frac{f_j(x^{(k-1)})}{x_j^{(k-1)} - x_j^{(k)}} e_j.
\]

Hence, taking into account assumption M3 and the mean-value theorem applied to \( \varphi \) we get:
\[
\left| f_j(x) \right| = \left| \varphi(x^{(k)}) - \varphi(x^{(k-1)}) \right| \\
\leq \sup_{x \in B} \left\| \nabla \varphi(x) \right\| \left\| x^{(k)} - x^{(k-1)} \right\| \\
= \sup_{x \geq 1} \sup_{x \in B} \left\| \nabla f_j(x) - \frac{f_j(x^{(k-1)})}{x_j^{(k-1)} - x_j^{(k)}} e_j \right\| \left\| x^{(k)} - x^{(k-1)} \right\| \\
\leq \frac{\xi}{m} \left\| x^{(k)} - x^{(k-1)} \right\|.
\]

Finally, the above inequality (3.45) together with Lemma 3.7 imply that (3.42) holds true for the case \( x_j^{(k-1)} \neq x_j^{(k)} \).

• Case II: \( x_j^{(k-1)} = x_j^{(k)} \). Then, inequality (3.42) obviously holds true in this case also.

Now, combining inequality (3.42) and employing Lemma 3.7 again we deduce immediately that
\[
\left\| x^{(k)} - x^{(k-1)} \right\| \leq \xi \left\| x^{(k-1)} - x^{(k-2)} \right\|.
\]

Consequently, we have
\[
\left\| x^{(k)} - x^{(0)} \right\| = \left\| \sum_{i=1}^{k-1} \left( x^{(i)} - x^{(i-1)} \right) \right\| \\
\leq \sum_{i=1}^{k-1} \left\| x^{(i)} - x^{(i-1)} \right\| \\
\leq \left( \sum_{i=1}^{k-1} \xi^{i-1} \right) \left\| x^{(1)} - x^{(0)} \right\| \\
\leq \frac{1}{1-\xi} \left\| x^{(1)} - x^{(0)} \right\|.
\]

Applying Lemma 3.7 with \( k = 1 \) and using assumption M4, we conclude that
\[
\left\| x^{(1)} - x^{(0)} \right\| \leq r(1 - \xi).
\]

Combining this with the previous inequality leads to:
\[
\left\| x^{(k)} - x^{(0)} \right\| \leq r,
\]

which shows that each iterate \( x^{(k)} \) belongs to the ball \( B_r \). Next, we prove that \( \{ x^{(k)} \} \) is a Cauchy sequence, and since \( \mathbb{R}^d \) is complete, it has a limit, say \( x_* \), in \( B_r \). Indeed, for any integer such that \( k \geq 0 \) and \( l > 0 \), we have
\[
\left\| x^{(k+l)} - x^{(k)} \right\| = \left\| \sum_{i=0}^{l-1} \left( x^{(k+i+1)} - x^{(k+i)} \right) \right\| \\
\leq \sum_{i=0}^{l-1} \left\| x^{(k+i+1)} - x^{(k+i)} \right\| \\
\leq \xi^k \left\| x^{(1)} - x^{(0)} \right\| \sum_{i=0}^{l-1} \xi^i \\
\leq \frac{\xi^k}{1-\xi} \left\| x^{(1)} - x^{(0)} \right\|.
\]
As \( t \) goes to infinity in (3.49), we can get more precise estimates than those obtained in (3.46):

\[
\|x^{(k)} - x_*\| \leq \frac{\xi^k}{1 - \xi} \|x^{(1)} - x^{(0)}\|,
\]

thus proving that \( \{x^{(k)}\} \) converges geometrically to a limit \( x_* \). Recalling equation (3.48), we obviously have \( x_* \in B_r \). Now, if the sequence \( \{x^{(k)}\} \) is convergent to a limit \( x_* \), passing to the limit in equation (3.45), then we immediately deduce from the continuity of \( \nabla f \) that \( \nabla f(x_*) = 0 \). To complete the proof we show that, under assumption M3, \( x_* \) is the unique stationary point of \( f \) in \( B_r \). To this end, assume that there is another point \( \tilde{x} \in B_r \) with \( \tilde{x} \neq x_* \) and which solves \( \nabla f(x) = 0 \). We will show that this gives a contradiction. Since by assumption M4, we have \( f_j(x_0) \neq 0 \), then Lemma 3.5 with \( k = 1 \) ensures that \( x_0 \neq x^{(1)}_j \), for all \( j = 1, \ldots, d \). Hence, we may define for each \( j = 1, \ldots, d \), the auxiliary function:

\[
\lambda_j(x) = \frac{x^{(1)}_j - x^{(0)}_j}{f'_j(x_0)} f_j(x) - f_j(x_0) \left( x_j - x^{(1)}_j \right).
\]

Obviously \( \lambda_j \) simultaneously satisfies \( \lambda_j(x_*) = 0 \) and \( \lambda_j(\tilde{x}) = x_* - \tilde{x}_j \). Therefore, applying again Lemma 3.7 for \( k = 1 \), we get from the mean value theorem and assumption 3.29

\[
|x^{(1)}_j - \tilde{x}_j| = |\lambda_j(x_*) - \lambda_j(\tilde{x})| \\
\leq \sup_{x \in B_r} \|\nabla \lambda_j(x)\| \|\tilde{x} - x_*\| \\
= \frac{x^{(1)}_j - x^{(0)}_j}{f'_j(x_0)} \sup_{x \in B_r} \|\nabla f_j(x) - f_j(x_0) e^{(j)}(x^{(0)}_j - x^{(1)}_j)\| \|\tilde{x} - x_*\| \\
\leq \frac{\xi}{\sqrt{2}} \|\tilde{x} - x_*\|.
\]

Then, we immediately obtain that

\[
0 < \|\tilde{x} - x_*\| \leq \xi \|\tilde{x} - x_*\|,
\]

but \( \xi \in (0, 1) \), and therefore the last equality holds only if \( \tilde{x} = x_* \), which is clearly a contradiction. Hence, we can conclude that \( f \) has a unique stationary point. Thus, the Theorem is proved.

We conclude this section by giving a simple one-dimensional example, which illustrates the performance of our method by showing that it has a wider convergence domain than the classical Newton’s method.

Example 3.8. Consider the function \( f : \mathbb{R} \to \mathbb{R} \) defined by

\[
f(x) = -e^{-x^2},
\]

its first and second order derivatives are given, respectively, by

\[
f'(x) = 2xe^{-x^2}
\]

\[
f''(x) = 2 \left(1 - 2x^2\right) e^{-x^2}.
\]

Since the second derivative of \( f \) is positive over the interval \( \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right] \) then the Newton’s method shall converge to the minimum of \( f \).

Let us recall that the famous Newton’s method for finding \( x_* \) uses the iterative scheme \( \{x^{(k)}\} \) defined by

\[
x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}, \quad (\forall k \geq 0),
\]

(3.55)
starting from some initial value \( x^{(0)} \). It converges geometrically in some neighborhood of \( x^\ast \) for a simple root \( x^\ast \).

In our example, the Newton iteration becomes

\[
x^{(k+1)} = x^{(k)} \left( 1 - \frac{1}{1 - 2(x^{(k)})^2} \right), \quad k \geq 0.
\]

Starting from the initial approximation \( x^{(0)} = \frac{1}{2} \) (respectively \( x^{(0)} = -\frac{1}{2} \)) the sequence given by \( x^{(k)} = \frac{1}{2} (-1)^{k} \) (respectively \( x^{(k)} = \frac{1}{2} (-1)^{k+1} \)) and then the sequence \( \{x^{(k)}\} \) does not converge. Also for initial values belonging to the interval \([1/\sqrt{2}, 1/2]\) after some iteration the sequence lies outside the interval \([1/\sqrt{2}, 1/2]\) and diverges. The domain of convergence of the Newton’s method is only the interval \([-1/2, 1/2]\).

Contrary to the Newton’s method, it is observed that our MMA method converges for any initial value taken into the larger interval \([-1/\sqrt{2}, 1/\sqrt{2}]\). Convergence results are summarized in Table (3.1).

4. A multistage turbines using a through-flow code. The investigation of through-flow have been used for many years in the analysis and the design of turbomachines, especially in the 70–th by many authors, see for example [15, 21, 31]. The main idea of these investigations are based on the numerical analysis of the streamline curvatures and the matrix through flow. More details can be found in [14, 19, 16]. The streamline curvature method offers flexible method of determining an Euler solution of axisymmetric flow through a turbomachine. The theory of streamline curvature through flow calculations has been described by many authors, particularly by John Denton [14]. From the assumption of axial symmetry it is possible to define a series of meridional stream surfaces. A surface of revolution along which particles are assumed to move through the machine. The principal of streamline curvature is to write the equation of motion along lines roughly perpendicular to these stream surfaces (quasi orthogonal lines) in terms of the curvature of the surfaces in the meridional plane, as shown in the left panel of the Figure 4.1. The two unknowns states that we are interested are the meridional fluid component of the velocity \( V_m (m/s) \) in the direction of the streamlines and the mass flow rate \( \dot{m} (kg/s) \).

The mass flow rate is evaluated at each location point at the intersections of the stream lines and the quasi orthogonal lines, and also depends on the variation of the meridional fluid velocity \( V_m \). The continuity equation takes the form:

\[
\dot{m} = 2\pi \int_{r_{hub}}^{r_{tip}} r \rho V_m (q,m) \sin \alpha (1 - b) dq
\]
Fig. 4.1. *Meridional view of the flow path (left panel), and Steam path design geometry (right panel)*

where $0 \leq b < 1$ is the blade blockage factor, $r$ the radius of the rotating machine axis (m) and $\rho$ the fluid density (${kg}/m^3$). The inlet mass flow rate is the mass flow rate calculated along the first quasi orthogonal line. Knowing the geometrical lean angle of the blades i.e. inclination of the blades in the tangential direction $\varepsilon$ (rad), the total enthalpy $H$ (N.m), the static temperature $T$ (K), and the entropy $S$ ($J/K$), as input data functions evaluated by empirical rules, we can find the variation of the meridional fluid velocity $V_m$ in function of the distance $q$ (m) along the quasi orthogonal lines, and the meridional direction $m$ by solving the equilibrium equation:

\[
\frac{1}{2} \frac{dV_m^2(q,m)}{dq} = \frac{V_m^2(q,m)}{2r} \sin \alpha + \frac{V_m \partial V_m(q,m)}{\partial m} \cos \alpha - \frac{1}{2r} \frac{d\left(r^2V_m^2(q,m)\right)}{dq} - \tan \varepsilon \frac{V_m}{r} \frac{\partial (r \theta)}{\partial m},
\]

(4.2)

where $\theta$ represents the direction of rotation and the values of $r \theta$ are specified while operating the design mode. The angle $\alpha$ (rad) between the quasi orthogonal and the stream surface and the radius of curvature $r_c$ (m) are updated with respect to the mass flow rate $\dot{m}$ ($kg/s$) distribution. The enthalpy is updated according to IAPWS-IF97 steam function as described in [29]. The entropy is calculated by fundamental thermodynamic relations between the internal energy of the system and external parameters (e.g. friction losses).

The computational parameters of the streamlines are drawn in a meridional view of the flow path in the left panel of the Figure 4.1, with one of the quasi-normal stations that are strategically located in the flow, between the tip and hub contours. Several stations are generally placed in the inlet duct upstream of the turbo-machine, the minimum number of quasi-orthogonal is simply one between the adjacent pair of blade rows, which would then represent both outlet conditions from the previous row and inlet conditions to the next. In our streamline curvature calculation tool, there is one quasi orthogonal at each edge of each blade row. Given these equations, and a step-by-step procedure we obtain a solution as described in [22].

In the left panel of the Figure 4.2 the contour of the turbo machine is limited by, on the top, the line that follows the tip contour at the casing and on the bottom, a line that follows the geometry of the hub contour at the rotor. Intermediate lines are additional streamlines, distributed according to the mass flow rate that goes through the stream tubes. Vertical inclined lines are the quasi orthogonal stations mainly located at the inlet and outlet of moving and fixed blade rows.
The possibility to impose a target mass flow rate at the inlet of the turbo-machine is very important for its final design as it is driven by downstream conditions. Equation 4.1 shows that the mass flow rate depends explicitly on the shape of the turbo-machine through the position of the extreme points $r_{hub}$ and $r_{tip}$ of the quasi orthogonal. The purpose of our inverse problem is to identify both hub and tip contours of the turbo-machine to achieve an expected mass flow rate at the inlet of the turbo-machine.

The geometry of the contours of the turbo-machine is defined by an uni-variate interpolation of $n$ points along the $r$-axis. The interpolation is based on the improved method developed by Hiroshi Akima [10]. In this method, the interpolating function is a piecewise function composed of a set of polynomials applicable to successive intervals of the given data points. We use the third-degree polynomials default option as it is not required to reduce any undulations in resulting curves.

In this realistic example that we have, we use five points on each curve describing respectively the hub and the tip contours, see the right panel of the Figure 4.2. The initial ten data points are extracted from an existing geometry, and are chosen arbitrary equidistant along the axial direction. Their radial position is linearly interpolated using the two closest points.

The unconstrained optimization will be to find $r^* = (r^{*1}, r^{*2}, \ldots, r^{*10})^T \in \mathbb{R}^{10}$ such that:

\[
(4.3) \quad f(r^*) = \min_{r \in \mathbb{R}^{10}} f(r),
\]

where $f(r) := \left( \hat{\dot{m}}(r) - \dot{m}(r) \right)^2$, $\dot{m}(r)$ is the mass flow rate that depends on the design parameters and $\hat{\dot{m}}$ is the imposed inlet mass flow rate.

In our example, the target inlet mass flow rate $\dot{m} = 200$ kg/s, and the initial realistic practical geometry gives an initial mass flow rate $\dot{m}_0 = 161.20$ kg/s with

\[
r_0 = (0.828, 0.836, 0.853, 0.853, 0.853, 0.962, 1.05, 1.337, 1.701, 2.124)^T.
\]

The difference between the target and the initial inlet mass flow value is about 20% which is consider to be very significant in practice. The initial shape is shown in the left panel of the Figure 4.2.

The moving asymptotes are chosen such that the condition (3.18) is automatically satisfied, and their numerical implementations are

\[
A_j^{(k)} = \begin{cases} 
L_j^{(k)} = r_j^{(k)} + 4 \frac{f_j(r^{(k)})}{S_{rj}} & \text{if } f_j(r^{(k)}) < 0, \\
U_j^{(k)} = r_j^{(k)} + 4 \frac{f_j(r^{(k)})}{S_{rj}} & \text{if } f_j(r^{(k)}) > 0.
\end{cases}
\]
It is important to note the simplest form which is used here for the selection of the moving asymptotes. The first-order partial derivatives are numerically calculated using a two-point formula that computes the slope

\[ f(r_1, \ldots, r_j + h, \ldots, r_{10}) - f(r_1, \ldots, r_j - h, \ldots, r_{10}) \]

\[ 2h, j = 1, \ldots, 10, \]

with an error of order \( h^2 \). For our numerical study, \( h \) has been chosen equal to \( 5.10^{-4} \) that corresponds to about \( 5.10^{-2} \% \) of the size of the design parameters, which gives an approximation largely enough accurate. To avoid computing second-order derivatives of the objective function \( f \), we use the spectral parameter as defined in (3.11). We observe a good convergence to the target inlet mass flow rate summed up in Table 4.1. The final steam path geometry is compared with the initial geometry in the right panel of the Figure 4.2, where the optimized hub and tip contours values are

\[ r_\ast = (0.824, 0.821, 0.857, 0.851, 0.853, 0.966, 1.074, 1.331, 1.703, 2.124)^T. \]

It appears that the hub contour of the optimized shape is largely more deformed than the tip contour, and the shape is more sensitive with respect to the design parameters of the hub than the tip contours.

5. Concluding remarks. In this paper we have developed and analyze new local convex approximation methods with explicit solutions of non-linear problems for unconstrained optimization, for large scale systems and framework the problems that evolve structural mechanical optimization of multi-scale model based on moving asymptotes algorithm (MMA) developed by Svanberg. We have shown that the problem leads to use the second derivative information in order to solve more efficiently structural optimization problems without constraint. The basic idea MMA methods can be interpreted as techniques that approximate a priori the curvature of the object function. In order to avoid the second derivatives evaluations of our algorithm a sequence of diagonal Hessian estimates, where only the first and zero order information are accumulated during the previous iterations was used. As consequence, at each step of the iterative process, a strictly convex approximation sub-problem is generated and solved. A convergence result with taken into account the second order derivatives information for our optimization algorithm under fairly mild assumptions, was presented in details.

It has been shown that the approximation scheme meet all well-known properties of convexity and separability of the MMA. In particular, we will have the following major advantages:
• All subproblems will have an explicit solutions. This reduce considerably the computational cost of the proposed method.

• It generates an iteration sequence, that, under mild technical assumptions, will be bounded and converges geometrically to a stationary point of the objective function with one or several variables from any "good" staring point.

The numerical results and the theoretical analysis of the convergence are very promising and indicate that the MMA method may be further developed for solving general large scale optimization problems, the methods proposed here can be also extended to the more realistic problem with constraints. We are now working to extend our approach to constrained optimization problems and investigate the stability of the algorithm against reference cases described in [32].

Acknowledgments. The authors are grateful to Roland Simmen for his valuable support to implement the method with the through flow streamline curvature algorithm. The authors thank financial support from the Volubilis Hubert Curien Program (Grant No. MA/13/286). We would also like to thank the two anonymous referees for providing us with constructive comments and suggestions.

REFERENCES


1 ALSTOM Ltd., Brown Boveri Strasse 10, CH-5401 Baden, Switzerland (roland.simmen@power.alstom.com)