COMPUTATIONAL
&APPLIED
MATHEMATICS

Solving the dual subproblem of the Method of Moving Asymptotes using a trust-region scheme

MÁRCIA A. GOMES-RUGGIERO,

MAEL SACHINE* and SANDRA A. SANTOS

Department of Applied Mathematics, IMECC, UNICAMP, 13081-970, Campinas, SP, Brazil

E-mails: marcia@ime.unicamp.br / mael@ime.unicamp.br / sandra@ime.unicamp.br

Abstract. An alternative strategy to solve the subproblems of the Method of Moving Asymptotes (MMA) is presented, based on a trust-region scheme applied to the dual of the MMA subproblem. At each iteration, the objective function of the dual problem is approximated by a regularized spectral model. A globally convergent modification to the MMA is also suggested, in which the conservative condition is relaxed by means of a summable controlled forcing sequence. Another modification to the MMA previously proposed by the authors [*Optim. Methods Softw.*, 25 (2010), pp. 883–893] is recalled to be used in the numerical tests. This modification is based on the spectral parameter for updating the MMA models, so as to improve their quality. The performed numerical experiments confirm the efficiency of the indicated modifications, especially when jointly combined.

Mathematical subject classification: 49M37, 49M29, 65K05, 90C30.

Key words: nonlinear programming, Method of Moving Asymptotes, spectral parameter, global convergence, dual problem.

1 Introduction

This study proposes a new strategy for solving the subproblems of the Method of Moving Asymptotes (MMA) by means of its dual formulation, using a trust-region technique. The MMA is a very popular method within the structural

---

optimization community and applies to the inequality constrained nonlinear programming problem with simple bounds as follows:

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \qquad\qquad i = 1, \ldots, m, \\
& x_j^{\min} \leq x_j \leq x_j^{\max}, \quad j = 1, \ldots, n,
\end{aligned}
\tag{1}
$$

where $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ is the vector of the variables, $x_j^{\min}$ and $x_j^{\max}$ are given real numbers for each j and $f_0, f_1, \ldots, f_m$ are real-valued twice continuously differentiable functions.

The original version of the MMA [16] was introduced in 1987 by Svanberg, as a generalization of the convex linearization method (CONLIN) [9], without global convergence. In 1995, Svanberg [17] proposed a globally convergent version. Several other MMA versions have appeared since then, see for instance [5, 20, 22, 23] and references therein. In 1998, Svanberg [18] developed a primal-dual interior-point method for solving the subproblems, in which a sequence of relaxed Karush-Kuhn-Tucker (KKT) conditions are solved by Newton's method. In 2003, Ni [15] proposed a globally convergent algorithm that combines the method of moving asymptotes with a trust-region technique, in order to solve bound-constrained problems. In its more recent version [19], the MMA was merged into the Conservative Convex and Separable Approximation (CCSA) class of methods, which are globally convergent.

In the current work, the dual problem associated with the MMA subproblem is stated and analyzed. The explicit expression of the dual objective function is accessible due to the separability of the rational models of the MMA. The intrinsic features of such a function are highlighted, namely being concave and continuously differentiable. The discontinuities of the second-order derivatives are discussed as well. Motivated by such features, we have proposed a trust-region approach for solving the dual of the MMA subproblem by means of a quadratic model that has a spectral regularization term. The solution of the trust-region subproblem has a closed form.

Another contribution of this work is related to the conservative condition responsible for defining the current outer iterate and ensuring the global convergence of the method. A relaxed conservative condition is proposed, based

on a summable controlled forcing sequence [13], so that the maintenance of global convergence of the MMA with this modification is proved. The modified MMA algorithm together with its convergence results are presented in [11].

In the numerical experiments, a third modification of the MMA previously proposed by the authors [10] is incorporated. This modification, also considered in the algorithm and theoretical results of [11], is based on the spectral parameter for the updating of a key parameter of the method, that ensures strict convexity of the model functions. The second-order information provided by the spectral parameter is included in the model functions that define the rational approximations of both the objective function and the nonlinear constraints at the beginning of each iteration, so as to improve the quality of the models. The computational results corroborate the proposed modifications, especially when jointly combined.

The structure of this paper is as follows. In Section 2, the basic ideas of the MMA are presented, and the relaxed conservative condition is described. In Section 3, a discussion of the dual problem associated with the MMA subproblem is provided together with details of our trust-region approach applied to the dual of the MMA subproblem. The numerical results are given in Section 4, and final remarks, in Section 5, conclude the text.

2   The Method of Moving Asymptotes and our modifications

Following Svanberg's approach [16], artificial variables $y = (y_1, \ldots, y_m)^T$ are introduced in problem (1), so that the following enlarged problem is addressed:

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) + \sum_{i=1}^{m}\left(c_i y_i + \tfrac{1}{2} d_i y_i^2\right) \\
\text{subject to} \quad & f_i(x) - y_i \leq 0, \quad i = 1, \ldots, m, \\
& x \in X, \quad y \geq 0,
\end{aligned}
\tag{2}
$$

where $X = \{x \in \mathbb{R}^n; x_j^{\min} \leq x_j \leq x_j^{\max}, j = 1, \ldots, n\}$ and c_i and d_i are real numbers such that $c_i \geq 0$ and $d_i > 0$ for $i = 1, \ldots, m$. The constants c_i must be chosen large enough so that the variables y_i are zero at the optimal solution, in case the original problem has a nonempty feasible set and fulfills a constraint qualification (e.g. Mangasarian-Fromovitz [14]).

The 2002 version of MMA for solving problem (2) performs outer and inner iterations. The indices (k, ℓ) are used to denote the ℓ-th inner iteration within the k-th outer iteration.

To start, it is necessary to choose $x^{(1)} \in X$, and then to compute $y^{(1)}$, obtaining an initial feasible estimate $(x^{(1)}, y^{(1)})$ for problem (2).

Thus, given $(x^{(k)}, y^{(k)})$, a subproblem is generated and solved. This subproblem is obtained from (2), replacing the objective function and the functions that define the inequality constraints by separable strictly convex models $g_i^{(k,\ell)}$. Moreover, the original box is reduced, being defined around the current point with the aid of the parameter $\sigma^{(k)}$. This subproblem is given next

$$
\begin{aligned}
\text{minimize} \quad & g_0^{(k,\ell)}(x) + \sum_{i=1}^{m}\left(c_i y_i + \tfrac{1}{2} d_i y_i^2\right) \\
\text{subject to} \quad & g_i^{(k,\ell)}(x) - y_i \le 0, \quad i = 1, \ldots, m, \\
& x \in X^{(k)}, \ y \ge 0,
\end{aligned}
\tag{3}
$$

for $k \in \{1, 2, 3, \ldots\}$ and $\ell \in \{0, 1, 2, \ldots\}$, where

$$
X^{(k)} = \left\{ x \in X \mid x_j \in \left[x_j^{(k)} - 0.9\sigma_j^{(k)}, \ x_j^{(k)} + 0.9\sigma_j^{(k)}\right], \ j = 1, \ldots, n \right\}.
$$

The vector $\sigma^{(k)} = (\sigma_1^{(k)}, \ldots, \sigma_n^{(k)})^T$ contains strictly positive parameters and its updating is done as in [19].

Denoting the optimal solution of subproblem (3) by $(\hat{x}^{(k,\ell)}, \hat{y}^{(k,\ell)})$, at the ℓ-th inner iteration, if the *conservative condition* holds at $\hat{x}^{(k,\ell)}$ for all functions of the problem, that is,

$$
f_i(\hat{x}^{(k,\ell)}) \le g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}), \quad \forall i \in \{0, 1, \ldots, m\},
\tag{4}
$$

then we set $(x^{(k+1)}, y^{(k+1)}) = (\hat{x}^{(k,\ell)}, \hat{y}^{(k,\ell)})$, and the k-th outer iteration is completed, after ℓ inner iterations. Otherwise, if $g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) < f_i(\hat{x}^{(k,\ell)})$ for at least an index $i \in \{0, 1, \ldots, m\}$, another inner iteration must be performed. The model for the function f_i is maintained the same for the index i such that the approximation is conservative in $\hat{x}^{(k,\ell)}$, that is, $g_i^{(k,\ell)}(x) \equiv g_i^{(k,\ell+1)}(x)$. For the indices for which the approximation $g_i^{(k,\ell)}$ does not fulfill the conservative condition (4) in $\hat{x}^{(k,\ell)}$, the model is modified so that the new approximation $g_i^{(k,\ell+1)}$ may satisfy the conservative condition in $\hat{x}^{(k,\ell+1)}$.

It is worth mentioning that the conservative condition is demanded for both the objective function and the constraints, producing, with regards to problem (2), strict reduction of the objective function value and feasible iterates, respectively.

In the MMA, the approximating functions are stated as

$$g_i^{(k,\ell)}(x) = \sum_{j=1}^n \left(\frac{p_{ij}^{(k,\ell)}}{u_j^{(k)} - x_j} + \frac{q_{ij}^{(k,\ell)}}{x_j - l_j^{(k)}} \right) + r_i^{(k,\ell)},$$

where the poles of the moving asymptotes $l_j^{(k)}$ and $u_j^{(k)}$ are

$$l_j^{(k)} = x_j^{(k)} - \sigma_j^{(k)} \qquad \text{and} \qquad u_j^{(k)} = x_j^{(k)} + \sigma_j^{(k)},$$

and the coefficients $p_{ij}^{(k,\ell)}$, $q_{ij}^{(k,\ell)}$ and $r_i^{(k,\ell)}$ are given by

$$p_{ij}^{(k,\ell)} = (\sigma_j^{(k)})^2 \max\left\{ 0, \frac{\partial f_i}{\partial x_j}(x^{(k)}) \right\} + \frac{\rho_i^{(k,\ell)} \sigma_j^{(k)}}{4},$$

$$q_{ij}^{(k,\ell)} = (\sigma_j^{(k)})^2 \max\left\{ 0, -\frac{\partial f_i}{\partial x_j}(x^{(k)}) \right\} + \frac{\rho_i^{(k,\ell)} \sigma_j^{(k)}}{4},$$

$$r_i^{(k,\ell)} = f_i(x^{(k)}) - \sum_{j=1}^n \left(\frac{p_{ij}^{(k,\ell)} + q_{ij}^{(k,\ell)}}{\sigma_j^{(k)}} \right).$$

Within an outer iteration k, the only difference between two inner iterations are the values of the parameters $\rho_i^{(k,\ell)}$. These parameters are strictly positive, so that all the approximating functions $g_i^{(k,\ell)}$ are strictly convex and every subproblem has a single global optimum. The updating of parameters $\rho_i^{(k,\ell)}$ is the one suggested in [19].

The model functions $g_i^{(k,\ell)}$ are first-order approximations to the original functions f_i at the current estimate, that is, conditions

$$g_i^{(k,\ell)}(x^{(k)}) = f_i(x^{(k)}) \qquad \text{and} \qquad \nabla g_i^{(k,\ell)}(x^{(k)}) = \nabla f_i(x^{(k)})$$

must hold for all $i = 0, 1, \ldots, m$. Another condition that must be satisfied by the approximating functions is separability, that is,

$$g_i^{(k,\ell)}(x) = g_{i0}^{(k,\ell)} + \sum_{j=1}^n g_{ij}^{(k,\ell)}(x_j).$$

Such a property is crucial in practice, because the Hessian matrices of the approximations are diagonal ones, allowing us to address large-scale problems. The model proposed in [19] satisfies such a condition with $g_{i0}^{(k,\ell)} = r_i^{(k,\ell)}$ and

$$
g_{ij}^{(k,\ell)}(x_j) = \left(\frac{p_{ij}^{(k,\ell)}}{u_j^{(k)} - x_j} + \frac{q_{ij}^{(k,\ell)}}{x_j - l_j^{(k)}} \right).
$$

In [10], a modified version of the MMA is proposed based on the spectral parameter, used in the updating of the parameters $\rho_i^{(k,\ell)}$. The second-order information provided by the spectral parameter is included in the model functions $g_i^{(k,\ell)}$ that define the rational approximations of the objective function and of the nonlinear constraints at the beginning of each outer iteration. The motivation for proposing this idea came from the numerical observation that in many cases the algorithm with Svanberg's original updating stops making significant progress in the solution of the sequence of solved subproblems. By improving the quality of the approximations this drawback was overcome. Moreover, the idea preserves the global convergence property of the CCSA class, as proved in [11].

We have devised another modification for the MMA, based on relaxing the conservative condition, by means of a summable controlled forcing sequence. In this sense, we say that the *relaxed conservative condition* holds at the iterate $\hat{x}^{(k,\ell)}$ if

$$
f_i(\hat{x}^{(k,\ell)}) \leq g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) + \mu_k \max\left\{ 1, \left| g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) \right| \right\}, \tag{5}
$$

for all $i \in \{0, 1, \ldots, m\}$, where

$$
\sum_{k=1}^{\infty} \mu_k \leq \mu < \infty. \tag{6}
$$

Therefore, the conservative condition is more relaxed at the beginning of the generated sequence, and ultimately achieved in the end. The original conservative condition is recovered if $\mu_k \equiv 0, \forall k$.

When it comes to the global convergence analysis, the arguments follow the outline of the presented in [19], with the main modifications briefly described bellow. The whole sequence of steps of this procedure is described in detail, together with the convergence results, and presented in [11].

Let the functions F_i be defined, for $x \in X$ and $y \in \mathbb{R}^m$, by

$$F_0(x, y) = f_0(x) + \sum_{i=1}^{m} (c_i y_i + \tfrac{1}{2} d_i y_i^2),$$

and

$$F_i(x, y) = f_i(x) - y_i, \quad \text{for all } i = 1, \ldots, m.$$

The original conservative condition (4) provides strict reduction in the objective function value F_0 and feasible iterates for problem (2). In Lemma 5 of [11] we prove that, adopting the relaxed conservative condition (5), the outer iterates of problem (2) might be infeasible and the values F_0 might increase. However, the most important remark of Lemma 5 is that this occurs in a controlled way, i.e.: $F_i(x^{(k)}, y^{(k)}) \leq \bar{\mu}_{k-1,i}$ for $i \geq 1$ and $k \geq 2$ and $F_0(x^{(k+1)}, y^{(k+1)}) < F_0(x^{(k)}, y^{(k)}) + \bar{\mu}_{k,0}$ for $k \geq 1$, where $\bar{\mu}_{k,i} = \mu_k \max\{1, |g_i^{(k,\ell)}(\hat{x}^{(k,\ell)})|\}$. After this lemma we have also proved that

$$\sum_{k=1}^{\infty} \bar{\mu}_{k,i} < \infty, \quad \text{for all } i = 0, 1, \ldots, m.$$

Due to these results, in Lemma 7 of [11], we prove that the sequence $\{F_0(x^{(k)}, y^{(k)})\}$ is convergent. This result is the same one stated by Lemma 7.8 of [19], in which the monotonicity had a crucial role on the proof. Considering this, the whole reasoning of [19] remains valid based on the fact that $\mu_k \to 0$ as $k \to \infty$, so that the global convergence of the MMA modified algorithm is maintained. The actual choice for the sequence $\{\mu_k\}_{k=1}^{\infty}$ is provided in Section 4, within the description of the numerical results.

In the section that follows, a brief analysis of the dual of the MMA subproblem and its properties will motivate a new strategy for solving the MMA subproblems.

3   Solving the MMA subproblems: interior-point methods *versus* a trust-region strategy

In this section, we propose a new strategy for solving the MMA subproblems by means of the associated dual problem, using a trust-region technique. This

new strategy is an alternative for both approaches already devised by Svanberg: the dual and the primal-dual interior-point ones. The dual approach is based on Lagrangian relaxation duality and was implemented with a linesearch technique [16]. In the primal-dual interior-point approach, a sequence of relaxed KKT conditions are solved by Newton's method [18]. We have devised a regularized quadratic model for the dual subproblem with the solution expressed in a closed form.

To simplify the notation, we omit the indices k and ℓ of the outer and inner iterations, respectively. We denote the bounds of the variables by the values α_j and β_j, i.e., $\alpha_j = \max\{x_j^{\min}, x_j^{(k)} - 0.9\sigma_j^{(k)}\}$ and $\beta_j = \min\{x_j^{\max}, x_j^{(k)} + 0.9\sigma_j^{(k)}\}$, for $j = 1, \ldots, n$, so that the box constraints of the MMA subproblems are: $\alpha_j \le x_j \le \beta_j$ for $j = 1, \ldots, n$ and $y_i \ge 0$ for $i = 1, \ldots, m$.

Initially, we show how to obtain an explicit expression of the dual objective function, thereby generating the dual problem corresponding to the MMA subproblem. We highlight some properties associated with the dual function. Then, we propose a trust-region scheme and present the algorithm.

3.1   The dual problem associated with the MMA subproblem

Considering only the main constraints, since the simple box and the non-negativity constraints will be incorporated in the minimization process, the Lagrangian corresponding to subproblem (3) is given by:

$$
\begin{aligned}
\mathcal{L}(x, y, \lambda) &= g_0(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{m}\left(c_i y_i + \tfrac{1}{2}d_i y_i^2\right) - \sum_{i=1}^{m}\lambda_i y_i \\
&= \sum_{j=1}^{n} \mathcal{L}_j(x_j, \lambda) + r_0 + \lambda^T r + \sum_{i=1}^{m}\left(c_i y_i + \tfrac{1}{2}d_i y_i^2 - \lambda_i y_i\right)
\end{aligned}
$$

where

$$
r = (r_1, \ldots, r_m)^T, \quad p_j = (p_{1j}, \ldots, p_{mj})^T, \quad q_j = (q_{1j}, \ldots, q_{mj})^T,
$$

$$
\mathcal{L}_j(x_j, \lambda) = \frac{p_{0j} + \lambda^T p_j}{u_j - x_j} + \frac{q_{0j} + \lambda^T q_j}{x_j - l_j},
$$

and $\lambda = (\lambda_1, \ldots, \lambda_m)^T$ is the vector of Lagrange multipliers.

The dual objective function $\mathcal{W}$ is defined, for $\lambda \geq 0$, as follows:

$$
\begin{aligned}
\mathcal{W}(\lambda) \;&=\; \min_{x,y} \left\{ \mathcal{L}(x, y, \lambda); \; \alpha_j \leq x_j \leq \beta_j, \; y_i \geq 0, \; \forall i, \forall j \right\} \\
&=\; r_0 + \lambda^T r + \sum_{j=1}^{n} \widehat{W}_j(\lambda) + \sum_{i=1}^{m} \widetilde{W}_i(\lambda)
\end{aligned}
\tag{7}
$$

where

$$
\widehat{W}_j(\lambda) \;=\; \min_{x_j} \left\{ \mathcal{L}_j(x_j, \lambda); \; \alpha_j \leq x_j \leq \beta_j \right\}, \quad j = 1, \ldots, n,
\tag{8}
$$

$$
\widetilde{W}_i(\lambda) \;=\; \min_{y_i} \left\{ c_i y_i + \tfrac{1}{2} d_i y_i^2 - \lambda_i y_i; \; y_i \geq 0 \right\}, \quad i = 1, \ldots, m.
\tag{9}
$$

Note that the separability of the MMA primal approximations allows the Lagrangian function $\mathcal{L}(x, y, \lambda)$ to be written as the sum of $n + m$ individual functions and therefore, the $(n + m)$-dimensional minimization problem (7) can be split into the $n + m$ minimization problems (8) and (9). The use of the minimum instead of the infimum in expressions (7)-(9) is justified by the existence of the minimizers of (8) and (9). The expressions of these minimizers, which we denote by $x_j(\lambda)$ and $y_i(\lambda_i)$, respectively, are:

$$
x_j(\lambda) = \max \left\{ \alpha_j, \min \left\{ \beta_j, x_j^*(\lambda) \right\} \right\}, \quad \text{for } j = 1, \ldots, n,
\tag{10}
$$

where

$$
x_j^*(\lambda) = \frac{(p_{0j} + \lambda^T p_j)^{1/2} l_j + (q_{0j} + \lambda^T q_j)^{1/2} u_j}{(p_{0j} + \lambda^T p_j)^{1/2} + (q_{0j} + \lambda^T q_j)^{1/2}}
\tag{11}
$$

and

$$
y_i(\lambda_i) = \max \left\{ 0, \frac{\lambda_i - c_i}{d_i} \right\}, \quad \text{for } i = 1, \ldots, m.
\tag{12}
$$

Note that $x_j : \mathbb{R}^m \to \mathbb{R}$ and $y_i : \mathbb{R} \to \mathbb{R}$ are continuous functions of λ, but not differentiable at the points λ such that $x_j(\lambda) = \alpha_j$ and $x_j(\lambda) = \beta_j$, for all $j = 1, \ldots, n$, and $\lambda_i = c_i$, for all $i = 1, \ldots, m$, respectively. Because there are explicit expressions for the minimizers $x_j(\lambda)$ of (8) and $y_i(\lambda_i)$ of (9), there is also an explicit expression for the dual objective function (7), which is:

$$
\begin{aligned}
\mathcal{W}(\lambda) \;=\; & r_0 + \lambda^T r + \sum_{j=1}^{n} \left(\frac{p_{0j} + \lambda^T p_j}{u_j - x_j(\lambda)} + \frac{q_{0j} + \lambda^T q_j}{x_j(\lambda) - l_j} \right) \\
& + \sum_{i=1}^{m} \left(c_i y_i(\lambda_i) + \tfrac{1}{2} d_i y_i^2(\lambda_i) - \lambda_i y_i(\lambda_i) \right).
\end{aligned}
$$

Thus, the dual problem corresponding to the MMA subproblem (3) is given by

$$\begin{aligned} \text{maximize} \quad & \mathcal{W}(\lambda) \\ \text{subject to} \quad & \lambda \geq 0. \end{aligned} \tag{13}$$

Once the dual problem (13) is solved, the optimal solution of the MMA (primal) subproblem (3) is obtained by replacing the dual optimal solution in the expressions of $x_j(\lambda)$ and $y_i(\lambda_i)$.

3.2   Properties of the dual function

Before proposing our approach to solve the dual problem (13) corresponding to the MMA subproblem (3), we comment on some properties associated with the dual function $\mathcal{W}$.

First, note that the function $\mathcal{W} : \mathbb{R}^m \to \mathbb{R}$ is concave, since it is the pointwise minimum of a collection of functions which are linear in λ. Moreover, it is continuous because $x_j(\lambda)$ and $y_i(\lambda_i)$ depend continuously on λ and $l_j < \alpha_j \leq x_j(\lambda) \leq \beta_j < u_j$. More than that, the function $\mathcal{W}$ is continuously differentiable and its first-order partial derivatives with respect to the dual variables λ_i are given by the constraints of the primal subproblem evaluated at $x_j(\lambda)$ and $y_i(\lambda_i)$, i.e.,

$$\begin{aligned} \frac{\partial \mathcal{W}}{\partial \lambda_i}(\lambda) &= g_i(x(\lambda)) - y_i(\lambda) \\ &= \sum_{j=1}^{n} \left(\frac{p_{ij}}{u_j - x_j(\lambda)} + \frac{q_{ij}}{x_j(\lambda) - l_j} \right) + r_i - y_i(\lambda_i), \end{aligned}$$

for all $i = 1, \dots, m$ and $\lambda \in \mathbb{R}^m$, as stated in Proposition 6.1.1 of [2]. Note that, although y does not belong to a compact set, the existence of the minimizer (12) justifies the usage of such a proposition.

Since the dual problem can be written explicitly, and the associated primal problem displays a relatively simple algebraic form, the second-order partial derivatives of the dual function can be written in a closed form:

$$\frac{\partial^2 \mathcal{W}}{\partial \lambda_i \partial \lambda_k}(\lambda) = \sum_{j=1}^{n} \left[\left(\frac{p_{ij}}{(u_j - x_j(\lambda))^2} - \frac{q_{ij}}{(x_j(\lambda) - l_j)^2} \right) \left(\frac{\partial x_j}{\partial \lambda_k}(\lambda) \right) \right] - y_i'(\lambda_i),$$

where we have abused on the notation by referring to $\partial x_j / \partial \lambda_k(\lambda)$, as $x_j(\lambda)$ is not differentiable at all points. The value of such a derivative assumed by a free

variable $x_j(\lambda)$, i.e., $\alpha_j < x_j(\lambda) < \beta_j$, may be different from the value of this derivative when the variable $x_j(\lambda)$ is fixed, i.e., $x_j(\lambda) = \alpha_j$ or $x_j(\lambda) = \beta_j$, which is obviously zero. This means that the second derivatives of the dual function are discontinuous whenever a free primal variable becomes fixed, or vice versa. From the primal-dual relationships (10), we see that the dual space is partitioned in several regions separated by second-order hypersurfaces of discontinuity. These surfaces are defined by $x_j^*(\lambda) = \alpha_j$ and $x_j^*(\lambda) = \beta_j$, where $x_j^*(\lambda)$ is given by (11).

3.3   Trust-region method

In this subsection we present a strategy to solve the dual subproblems of the MMA, using a trust-region scheme. Consider then the dual problem corresponding to the MMA subproblem as the minimization of function $W(\lambda)$ subject to no other constraints than non-negativity requirements on the dual variables:

$$\begin{aligned} \text{minimize} \quad & W(\lambda) \\ \text{subject to} \quad & \lambda \geq 0, \end{aligned} \tag{14}$$

where $W(\lambda) = -\mathcal{W}(\lambda)$. The quadratic model for the function W, adopted at each iteration $\bar{k}$ of the trust-region algorithm is:

$$m_{\bar{k}}(\lambda) = W(\lambda^{(\bar{k})}) + \nabla W(\lambda^{(\bar{k})})^T (\lambda - \lambda^{(\bar{k})}) + \tfrac{1}{2}\eta^{(\bar{k})}\|\lambda - \lambda^{(\bar{k})}\|_2^2, \tag{15}$$

where $\eta^{(\bar{k})}$ is the spectral parameter associated with the function W at the current iterate, that is,

$$\eta^{(\bar{k})} = \frac{(s^{(\bar{k})})^T t^{(\bar{k})}}{(s^{(\bar{k})})^T s^{(\bar{k})}} \tag{16}$$

with

$$s^{(\bar{k})} = \lambda^{(\bar{k})} - \lambda^{(\bar{k}-1)} \qquad \text{and} \qquad t^{(\bar{k})} = \nabla W(\lambda^{(\bar{k})}) - \nabla W(\lambda^{(\bar{k}-1)}).$$

The second-order term in the quadratic model $m_{\bar{k}}$ can be interpreted as a quadratic regularization term of a linear model of the function W in the proximal sense, where the spectral parameter $\eta^{(\bar{k})}$ has the flavour of an adaptive regularization parameter [12]. This interpretation justifies the second-order term of the model, since the Hessian matrix $\nabla^2 W$ is discontinuous. Furthermore, models

similar to (15) have been considered, as in [1] where the quadratic term of the model includes the spectral parameter in order to speed up a procedure based on the projected gradient, and in [21] where spherical quadratic convex approximations are employed in gradient-only optimization methods.

At each iteration $\bar{k}$, we should minimize the model $m_{\bar{k}}$ subject to a trust region and to the nonnegativity of the dual variables. Any norm may be used to define the trust region, but since the feasible set of (14) is an orthant, the choice $|| \cdot ||_\infty$ fits better in the sense that the constraints of the trust-region subproblem are simple-bounded ones.

Therefore, we obtain the problem

$$\begin{aligned} \text{minimize} \quad & m_{\bar{k}}(\lambda) \\ \text{subject to} \quad & \underline{\lambda}^{(\bar{k})} \leq \lambda \leq \overline{\lambda}^{(\bar{k})}, \end{aligned} \tag{17}$$

where $\underline{\lambda}_i^{(\bar{k})} = \max\{0, \lambda_i^{(\bar{k})} - \Delta^{(\bar{k})}\}$, $\overline{\lambda}_i^{(\bar{k})} = \lambda_i^{(\bar{k})} + \Delta^{(\bar{k})}$ and $\Delta^{(\bar{k})} > 0$ is the trust-region radius. The solution $\hat{\lambda} \in \mathbb{R}^m$ of problem (17) is given by the closed form

$$\hat{\lambda} = \min\left\{\overline{\lambda}^{(\bar{k})}, \max\left\{\underline{\lambda}^{(\bar{k})}, \lambda^{(\bar{k})} - \frac{1}{\eta^{(\bar{k})}}\nabla W(\lambda^{(\bar{k})})\right\}\right\}. \tag{18}$$

A model algorithm based on the trust-region framework is given next for completeness.

Algorithm: A trust-region approach applied to the dual of the MMA subproblem

Given $\lambda^{(1)}$, $\Delta^{(1)} > 0$, $0 < \upsilon < \omega < 1$, $0 < \gamma_0 \leq \gamma_1 < 1 \leq \gamma_2$, for $\bar{k} = 1, 2, \ldots$ until convergence

1. Compute $\eta^{(\bar{k})}$ using (16) and $\hat{\lambda}$ as in (18).

2. Compute $W(\hat{\lambda})$ and

$$\theta_{\bar{k}} = \frac{W(\lambda^{(\bar{k})}) - W(\hat{\lambda})}{m_{\bar{k}}(\lambda^{(\bar{k})}) - m_{\bar{k}}(\hat{\lambda})}.$$

3. Set

$$\lambda^{(\bar{k}+1)} = \begin{cases} \hat{\lambda}, & \text{if } \theta_{\bar{k}} > \upsilon \\ \lambda^{(\bar{k})}, & \text{otherwise.} \end{cases}$$

4. Set

$$
\Delta^{(\bar{k}+1)} \in
\begin{cases}
[\Delta^{(\bar{k})}, \gamma_2 \Delta^{(\bar{k})}], & \text{if } \theta_{\bar{k}} \geq \omega & \text{[very successful iteration]} \\
\Delta^{(\bar{k})}, & \text{if } \upsilon < \theta_{\bar{k}} < \omega & \text{[successful iteration]} \\
[\gamma_0 \Delta^{(\bar{k})}, \gamma_1 \Delta^{(\bar{k})}], & \text{otherwise.} & \text{[unsuccessful iteration]}
\end{cases}
$$

In the first iteration of our algorithm, to compute the spectral parameter $\eta^{(1)}$, we need another estimate $\lambda^{(0)}$ distinct from the initial estimate $\lambda^{(1)}$. This estimate $\lambda^{(0)}$ is computed by perturbing $\lambda^{(1)}$, i.e., $\lambda^{(0)} = \lambda^{(1)} + \varepsilon$. In the numerical tests, we have used $\varepsilon = 10^{-3}$.

We have stated a summary of the algorithm used in the numerical experiments. The choice adopted in this work for updating the trust-region radius is based on [6, 7]. For further algorithmic details, see [11].

Despite not being our primary motivation, it is worth mentioning that (18) coincides with the first Spectral Projected Gradient (SPG) trial point (cf. [3]) for problem (14) within the bound constraints of (17). For further details on the SPG, see also [4] and references therein. Instead of adopting the linesearch procedure of the SPG algorithm, we use a trust-region scheme. As usual in methods that employ spectral gradients, better practical results are obtained by not imposing sufficient functional decrease at every iteration. In this sense, the acceptance condition of the Step 3 provides a nonmonotone decrease for the function W because $\theta_{\bar{k}} > \upsilon$ may be seen as the relaxed Armijo-like condition

$$
W(\hat{\lambda}) < W(\lambda^{(\bar{k})}) + \upsilon \nabla W(\lambda^{(\bar{k})})^T (\hat{\lambda} - \lambda^{(\bar{k})}) + \frac{\upsilon}{2} \eta^{(\bar{k})} \|\hat{\lambda} - \lambda^{(\bar{k})}\|_2^2,
$$

whenever $m_{\bar{k}}(\lambda^{(\bar{k})}) > m_{\bar{k}}(\hat{\lambda})$.

4   Numerical results

This section is concerned with the description of the computational tests of modified versions of the MMA, based on the spectral updating, the relaxed conservative condition, and our trust-region approach applied to the dual of the MMA subproblem. The code was implemented in `Matlab` and the experiments were run in a Mac Pro with two Xeon E5462 processors of 2.8 Ghz and 12 GB of RAM memory (without multiprocessing).

Two families of academic problems were addressed, parameterized by the number of variables $n > 1$, and suggested in [19]. Their general structure resembles that of topology optimization problems, namely nonconvex problems with a large number of variables, upper and lower bounds on all variables, and a relatively small number of general inequality constraints.

Problem 1 has a strictly convex objective function and nonlinear constraints defined by means of strictly concave functions, so that the feasible region is nonconvex. Problem 2, on the other hand, has a strictly concave objective function and the functions that define the feasible region are strictly convex. They are stated as

$$\textbf{Academic problem 1:} \qquad\qquad \textbf{Academic problem 2:}$$

$$\text{minimize} \quad x^T S x \qquad\qquad\qquad \text{minimize} \quad -x^T S x$$

$$\text{s.t.} \quad \frac{n}{2} - x^T P x \le 0, \qquad\qquad \text{s.t.} \quad x^T P x - \frac{n}{2} \le 0,$$

$$\frac{n}{2} - x^T Q x \le 0, \qquad\qquad\qquad x^T Q x - \frac{n}{2} \le 0,$$

$$-1 \le x_j \le 1, \ \forall j, \qquad\qquad\qquad -1 \le x_j \le 1, \ \forall j,$$

where the square matrices S, P and Q of dimension n are symmetric and positive definite. Their elements are given by

$$S_{ij} = \frac{2 + \text{sen}(4\pi\alpha_{ij})}{(1 + |i - j|) \ln n}, \quad P_{ij} = \frac{1 + 2\alpha_{ij}}{(1 + |i - j|) \ln n}, \quad Q_{ij} = \frac{3 - 2\alpha_{ij}}{(1 + |i - j|) \ln n},$$

where

$$\alpha_{ij} = \frac{(i + j - 2)}{(2n - 2)} \in [0, 1]$$

for all i and j. The feasible starting points for Problems 1 and 2 are $x^{(1)} = (0.5, \ldots, 0.5)^T \in \mathbb{R}^n$ and $x^{(1)} = (0.25, \ldots, 0.25)^T \in \mathbb{R}^n$, respectively.

The problem dimension n varied in $\{100, 500, 1000, 2000\}$ for both problems. Problems 1 and 2 are formulated as in (1), so they were initially written in the format (2) with $d_i = 1$ and $c_i = 1000$, for $i = 1, \ldots, m$. These choices have produced $y \equiv 0$ for each outer iterate.

To establish the stopping criteria, note that the KKT conditions of the considered problems may be stated as follows, using the notation $a^+ = \max\{0, a\}$

and $a^- = \max\{0, -a\}$:

$$(1 + x_j) \left(\frac{\partial f_0}{\partial x_j}(x) + \lambda_1 \frac{\partial f_1}{\partial x_j}(x) + \lambda_2 \frac{\partial f_2}{\partial x_j}(x) \right)^+ = 0, \quad j = 1, \ldots, n,$$

$$(1 - x_j) \left(\frac{\partial f_0}{\partial x_j}(x) + \lambda_1 \frac{\partial f_1}{\partial x_j}(x) + \lambda_2 \frac{\partial f_2}{\partial x_j}(x) \right)^- = 0, \quad j = 1, \ldots, n,$$

$$f_i(x)^+ = 0, \quad i = 1, 2,$$

$$\lambda_i f_i(x)^- = 0, \quad i = 1, 2,$$

$$\lambda_i \geq 0, \quad i = 1, 2,$$

$$-1 \leq x_j \leq 1, \quad j = 1, \ldots, n.$$

The $2n + 4$ equalities displayed previously may be concisely stated as $r_\varphi(x, \lambda) = 0$, $\varphi = 1, \ldots, 2n + 4$. As a by-product of the strategies employed to solve the problems, the inequalities of the KKT system are always fulfilled by the primal and dual variables, x_j and λ_i, respectively. The outer loop finishes successfully whenever x and λ are such that

$$\frac{1}{n} \sum_{\varphi=1}^{2n+4} \left(r_\varphi(x, \lambda) \right)^2 \leq 10^{-10}.$$

The sequence $\{\mu_k\}_{k=1}^\infty$ used in (5) to relax the conservative condition was chosen as follows:

$$\mu_k = \frac{N_k}{(k + 1)^{1.1}},$$

with

$$N_1 = ||r_\varphi^{(1)}||_2, \qquad N_2 = \min \left\{ ||r_\varphi^{(1)}||_2, ||r_\varphi^{(2)}||_2 \right\},$$

and for $k \geq 3$

$$N_k = \min \left\{ ||r_\varphi^{(k-2)}||_2, ||r_\varphi^{(k-1)}||_2, ||r_\varphi^{(k)}||_2 \right\},$$

where $r_\varphi^{(k)} \equiv r_\varphi(x^{(k)}, \lambda^{(k)})$ is the residue of the KKT conditions of problem (2) at the k-th outer iteration. To ensure that the sequence N_k is bounded, we take $N_k = \min\{N_k, N_{\max}\}$. However, the value $N_{\max}$, set at 10^{12} in the numerical tests, was never reached. In this way, the sequence $\{\mu_k\}_{k=1}^\infty$ naturally fulfills the assumption (6).

Three strategies were adopted to solve the problems: in **Strategy 1** the spectral parameter was used to update the parameters $\rho_i^{(k,\ell)}$ at the beginning of each outer iteration; in **Strategy 2** the relaxed conservative condition (5) was employed as the acceptance criterion so that the solution of the MMA subproblem becomes the next outer iterate; in **Strategy 3** both strategies 1 and 2 are combined.

We have compared eight distinct instances: Svanberg's primal-dual approach (PD), our dual trust-region approach (TR), and the combination of these approaches with each of the three strategies described above.

From the whole set of comparative results involving all these combinations, and thoroughly described in [11], we have noticed that the strategies that used the dual trust-region approach are competitive in terms of the demanded number of iterations, and more efficient when it comes to the CPU time spent, in comparison with those that rely upon the primal-dual approach.

Among the instances that used the dual trust-region approach, we have observed that in most of the cases Strategy 1 usually needs slightly more outer iterations to reach convergence than the pure algorithm without any modification. However, the amount of additional inner iterations decreases in a larger proportion, so that for both problems, the total number of solved subproblems is smaller for the spectral strategy than for the method without further modifications.

Analyzing Strategy 2, for Problem 1, we have noticed that despite the increase in the number of outer iterations, the additional inner iterations demanded were so few that the total number of solved subproblems is even smaller than in Strategy 1. For Problem 2, although the additional amount of inner iterations performed is not so small, all in all, the total effort decreases when compared with Strategy 1 and with the original algorithm.

Focusing now on Strategy 3, the results obtained were excellent. The number of both outer and additional inner iteration decreased by a large amount, and consequently the CPU time spent is the least among the four instances that used the dual trust-region approach for solving the MMA subproblem.

Additional tests were produced by randomly generating ten initial points within the simple bounds of the problems, for each of the dimensions under con-

sideration. Thus, eighty tests were solved, forty for Problem 1 and forty for Problem 2. Each strategy was used for solving these tests, and the results corroborate the previous ones. It is important to note that for each test, including the aforementioned, with the initial point from the literature, all the strategies achieved the same optimal solution.

In Figure 1 we depict the performance profiles [8] of the results corresponding to the eighty generated tests. Figure 1(a) is concerned with the number of solved subproblems, whereas Figure 1(b), with the CPU time spent. In both profiles we notice that Strategies 3 PD and 3 TR are the most efficient. From Figure 1(a) it is evident that Svanberg's PD is more efficient in terms of the number of solved subproblems. Nevertheless, both approaches (PD and TR) for solving the subproblems are competitive, when compared pairwise with each of the three strategies. When it comes to the CPU time spent, the strategies that rely upon the dual trust-region approach were more efficient, as can be seen in Figure 1(b).

5   Conclusions

We have proposed a new strategy for solving the MMA subproblems by means of its dual formulation, using a trust-region technique. This alternative approach deals with the dual problem associated with the MMA subproblem, that is a maximization problem of a concave function under nonnegativity constraints. We have taken advantage of the dual objective function properties, such as being concave and continuously differentiable up to first-order, together with the existence of a closed form for the solution of the subproblem obtained with a regularized spectral model within a trust-region scheme. Such a globalization strategy was the key point in recasting, in a simpler way, the dual approach originally adopted by Svanberg [16], and replaced by the primal-dual approach [18]. We have also presented a modification for the MMA, based on relaxing the conservative condition by means of a summable controlled forcing sequence, so that the maintenance of global convergence is proved [11]. Another modification for the MMA, previously proposed by the authors, was recalled to be used in the numerical tests. It is based on the spectral parameter for updating the parameters $\rho_i^{(k,\ell)}$, so as to improve the quality of the MMA models.
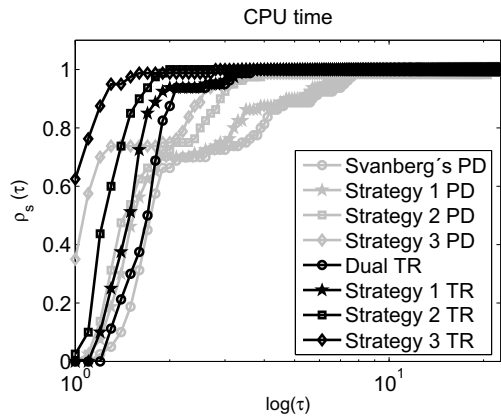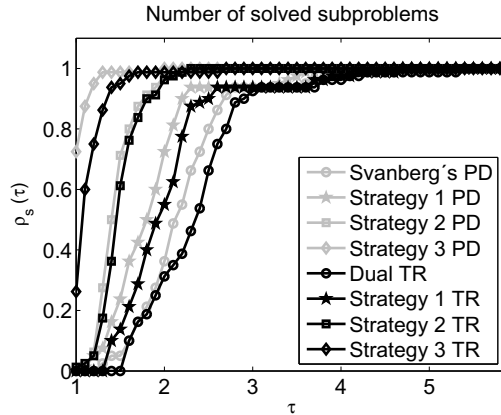
(a)

(b)

Figure 1 – Performance profiles for Problems 1 and 2.

The numerical experiments revealed that the suggested dual approach is simpler and more efficient than Svanberg's primal-dual strategy for solving the family of test problems under consideration. Indeed, we have noticed that the performance of our dual trust-region approach was quite similar to the one of Svanberg's primal-dual approach in terms of the employed number of iterations, but when it comes to the CPU time demanded, our approach was by far superior. Additionally, the performances of both the trust-region dual and the primal-dual approaches were improved in an increasing pattern with the addition of each suggested modification, namely using the spectral updating (Strat-

egy 1), the relaxed conservative condition (Strategy 2) and the combination of these two ideas (Strategy 3), pointing out the potential contribution of such modifications for the original algorithm.

REFERENCES

[1]   A. Auslender, P.J.S. Silva and M. Teboulle, *Nonmonotone projected gradient methods based on barrier and Euclidean distances*. Comput. Optim. Appl., **38** (2007), 305–327.

[2]   D.P. Bertsekas, *Nonlinear Programming: Second Edition*. Athena Scientific, Belmont (2003). (Second Printing).

[3]   E.G. Birgin, J.M. Martínez and M. Raydan, *Nonmonotone spectral projected gradient methods on convex sets*. SIAM J. Optim., **10** (2000), 1196–1211.

[4]   E.G. Birgin, J.M. Martínez and M. Raydan, *Spectral Projected Gradient Methods*. In *Encyclopedia of Optimization*, C.A. Floudas and P.M. Pardalos, editors. Second Edition. Springer (2009), 3652–3659.

[5]   M. Bruyneel, P. Duysinx and C. Fleury, *A family of MMA approximations for structural optimization*. Struct. Multidiscip. Optim., **24** (2002), 263–276.

[6]   A R. Conn, N.I.M. Gould and Ph.L. Toint, *LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization (Release A)*. Springer-Verlag, Berlin, Heidelberg (1992).

[7]   A.R. Conn, N.I.M. Gould and Ph.L. Toint, *Trust-Region Methods*. SIAM, Philadelphia (2000).

[8]   E.D. Dolan and J.J. Moré, *Benchmarking optimization software with performance profiles*. Math. Program., **91** (2002), 201–213.

[9]   C. Fleury and V. Braibant, *Structural optimization: A new dual method using mixed variables*. Internat. J. Numer. Methods Engrg., **23** (1986), 409–428.

[10] M.A. Gomes-Ruggiero, M. Sachine and S.A. Santos, *A spectral updating for the method of moving asymptotes*. Optim. Methods Softw., **25**(6) (2010), 883–893.

[11] M.A. Gomes-Ruggiero, M. Sachine and S.A. Santos, *Globally convergent modifications to the Method of Moving Asymptotes and the solution of the subproblems using trust regions: theoretical and numerical results*. Technical Report RP 15/10, IMECC, Unicamp, revised in November (2010), 44 p. Available at http://www.ime.unicamp.br/rel_pesq/relatorio.html.

[12] O. Güler, *New proximal point algorithms for convex minimization*. SIAM J. Optim., **2**(4) (1992), 649–664.

[13] D.H. Li and M. Fukushima, *A derivative-free line search and global convergence of Broyden-like method for nonlinear equations*. Optim. Methods Softw., **13**(3) (2000), 181–201.

[14] O.L. Mangasarian and S. Fromovitz, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*. J. Math. Anal. Appl., **17** (1967), 37–47.

[15] Q. Ni, *A globally convergent method of moving asymptotes with trust region technique*. Optim. Methods Softw., **18**(3) (2003), 283–297.

[16] K. Svanberg, *The method of moving asymptotes – a new method for structural optimization*. Internat. J. Numer. Methods Engrg., **24** (1987), 359–373.

[17] K. Svanberg, *A Globally Convergent Version of MMA without Linesearch*. In: G.I.N. Rozvany and N. Olhoff (eds). Proceedings of the First World Congress of Structural and Multidisciplinary Optimization, (1995), 9–16.

[18] K. Svanberg, *The Method of Moving Asymptotes – Modelling aspects and solution schemes*. Lecture Notes for the DCAMM course Advanced Topics in Structural Optimization, (1998), 24 p.

[19] K. Svanberg, *A class of globally convergent optimization methods based on conservative convex separable approximations*. SIAM J. Optim., **12** (2002), 555–573.

[20] H. Wang and Q. Ni, *A new method of moving asymptotes for large-scale unconstrained optimization*. Appl. Math. Comput., **203** (2008), 62–71.

[21] D.N. Wilke, S. Kok and A.A. Groenwold, *The application of gradient-only optimization methods for problems discretized using non-constant methods*. Struct. Multidiscip. Optim., **40** (2010), 433–451.

[22] W.H. Zhang and C. Fleury, *A modification of convex approximation methods for structural optimization*. Comput. & Structures, **64** (1997), 89–95.

[23] C. Zillober, *Global convergence of a nonlinear programming method using convex approximations*. Numer. Algorithms, **27** (2001), 256–289.