Globally convergent modifications to the Method of Moving Asymptotes and the solution of the subproblems using trust regions: theoretical and numerical results*

Márcia A. Gomes-Ruggiero Mael Sachine† Sandra A. Santos

December 3, 2010

Revised version

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. This report contains all the global convergence results and the complete set of numerical and graphical elements that sustain our performance analysis.

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

AMS Classification: 49 M29, 49M37, 65K05, 90C30.

*This work was partially supported by FAPESP (Grants 06/52846-7, 06/53768-0 and 10/09773-4), CNPq (Grants 303465/2007-7 and 306220/2009-1), and PRONEX-Optimization. It was developed at the State University of Campinas, Department of Applied Mathematics, Campinas, SP, Brazil.

†Corresponding author: mael@ime.unicamp.br.

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:

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

where $x = (x_1, \ldots, x_n)^T \in I\!R^n$ 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 [14] was introduced in 1987 by Svanberg, as a generalization of the convex linearization method (CONLIN) [8], without global convergence. In 1995, Svanberg [15] proposed a globally convergent version. Several other MMA versions have appeared since then, see for instance [4, 18, 20, 21] and references therein. In 1998, Svanberg [16] 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 [13] 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 [17], 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 [11], so that the maintenance of global convergence of the MMA algorithm with this modification is proved.

In the numerical experiments, a third modification of the MMA previously proposed by the authors [9] is incorporated, 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 report is as follows. In Section 2, the basic ideas of the MMA are presented. The proposed modifications to the MMA - the spectral updating and the relaxed conservative condition - are described and the modified algorithm is summarized in Section 3. In Section 4, the complete theoretical analysis of global convergence of the proposed algorithm is given. A discussion of the dual problem associated with the MMA subproblem is provided in Section 5, together with details of our trust-region approach applied to the dual of the MMA subproblem. The numerical results are presented in Section 6, and final remarks, in Section 7, conclude the text.

2 The Method of Moving Asymptotes

Following Svanberg's approach [14], artificial variables $y = (y_1, \dots, y_m)^T$ are introduced in problem (1), so that the following enlarged problem is addressed:

$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^{m} (c_i y_i + \frac{1}{2} d_i y_i^2) \\ \text{subject to} & f_i(x) - y_i \le 0, \quad i = 1, \dots, m, \\ & x \in X, \ \ y \ge 0, \end{array} \tag{2}$$

where $X = \{x \in I\!R^n; x_j^{min} \le x_j \le x_j^{max}, j = 1, \dots, n\}$ and c_i and d_i are real numbers such that $c_i \ge 0$ and $d_i > 0$ for $i = 1, \dots, m$. 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 [12]).

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

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

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

$$X^{(k)} = \{x \in X \mid x_j \in [x_j^{(k)} - 0.9\sigma_j^{(k)},\ x_j^{(k)} + 0.9\sigma_j^{(k)}],\ j = 1, \ldots, n\}. \tag{4}$$

The vector $\sigma^{(k)} = (\sigma_1^{(k)}, \ldots, \sigma_n^{(k)})^T$ contains strictly positive parameters and its updating is done as in [17], where each vector $\sigma^{(k)}$ belongs to a compact set S as follows:

$$S = \left\{ \sigma \in I\!R^n \mid \sigma_j^{\min} \leq \sigma_j \leq \sigma_j^{\max},\ j = 1, \ldots, n \right\}, \tag{5}$$

where $\sigma_j^{\min}$ and $\sigma_j^{\max}$ are given real numbers such that $0 < \sigma_j^{\min} < \sigma_j^{\max} < \infty$.

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,

$$f_i(\hat{x}^{(k,\ell)}) \leq g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}), \ \ \forall\, i \in \{0, 1, \ldots, m\}, \tag{6}$$

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 (6) 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

$$g_i^{(k,\ell)}(x) = \sum_{j=1}^{n} \left(\frac{p_{ij}^{(k,\ell)}}{u_j^{(k)} - x_j} + \frac{q_{ij}^{(k,\ell)}}{x_j - l_j^{(k)}} \right) + r_i^{(k,\ell)}, \tag{7}$$

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

$$p_{ij}^{(k,\ell)} = (\sigma_j^{(k)})^2 \max\left\{0, \frac{\partial f_i}{\partial x_j}(x^{(k)})\right\} + \frac{\rho_i^{(k,\ell)}\sigma_j^{(k)}}{4}, \tag{8}$$

$$\tag{9}$$

$$q_{ij}^{(k,\ell)} = (\sigma_j^{(k)})^2 \max\left\{0, -\frac{\partial f_i}{\partial x_j}(x^{(k)})\right\} + \frac{\rho_i^{(k,\ell)}\sigma_j^{(k)}}{4}, \tag{10}$$

$$\tag{11}$$

$$r_i^{(k,\ell)} = f_i(x^{(k)}) - \sum_{j=1}^{n} \left(\frac{p_{ij}^{(k,\ell)} + q_{ij}^{(k,\ell)}}{\sigma_j^{(k)}} \right). \tag{12}$$

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 [17], and included here for further reference. For $\ell = 0$, the following values are used:

$$\rho_i^{(1,0)} = 1, \tag{13a}$$

$$\rho_i^{(k+1,0)} = \max\{0.1\, \rho_i^{(k,\hat{\ell}(k))},\ \rho_i^{min}\}, \tag{13b}$$

where $\hat{\ell}(k)$ is the number of inner iterations needed within the k-th outer iteration, and ρ_i^{min} is a fixed strictly positive number, such as, 10^{-5}.

At each inner iteration, the updating of $\rho_i^{(k,\ell)}$ is based on the solution of the most recent subproblem. If $g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) < f_i(\hat{x}^{(k,\ell)})$ the value $\rho_i^{(k,\ell+1)}$ is chosen so that

$$g_i^{(k,\ell+1)}(\hat{x}^{(k,\ell)}) = f_i(\hat{x}^{(k,\ell)}),$$

which provides $\rho_i^{(k,\ell+1)} = \rho_i^{(k,\ell)} + \delta_i^{(k,\ell)}$ where

$$\delta_i^{(k,\ell)} = \frac{f_i(\hat{x}^{(k,\ell)}) - g_i^{(k,\ell)}(\hat{x}^{(k,\ell)})}{w(\hat{x}^{(k,\ell)}, x^{(k)}, \sigma^{(k)})}. \tag{14}$$

Thus we obtain

$$\begin{array}{rcll} \rho_i^{(k,\ell+1)} & = & \min\{10\,\rho_i^{(k,\ell)},\ 1.1\,(\rho_i^{(k,\ell)} + \delta_i^{(k,\ell)})\} & \text{if} \quad \delta_i^{(k,\ell)} > 0, \\ \rho_i^{(k,\ell+1)} & = & \rho_i^{(k,\ell)} & \text{if} \quad \delta_i^{(k,\ell)} \le 0. \end{array} \tag{15}$$

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

$$g_i^{(k,\ell)}(x^{(k)}) = f_i(x^{(k)}) \qquad \text{and} \qquad \nabla g_i^{(k,\ell)}(x^{(k)}) = \nabla f_i(x^{(k)}) \tag{16}$$

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 [17] satisfies such a condition with $g_{i0}^{(k,\ell)} = r_i^{(k,\ell)}$ and

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

What is more, the approximating functions $g_i^{(k,\ell)}$ of the MMA subproblem satisfy (cf. [17])

$$g_i^{(k,\ell)}(x) = v_i(x, x^{(k)}, \sigma^{(k)}) + \rho_i^{(k,\ell)} w(x, x^{(k)}, \sigma^{(k)}), \quad i = 0, 1, \ldots, m, \tag{17}$$

where $v_i(x, \xi, \sigma)$ and $w(x, \xi, \sigma)$ are real-valued functions defined on the set D given by

$$D = \{(x, \xi, \sigma) \mid \xi \in X, \sigma \in S, x \in X(\xi, \sigma)\}, \tag{18}$$

with S is defined in (5) and $X(\xi, \sigma)$ is a subset of X given by

$$X(\xi, \sigma) = \{x \in X \mid x_j \in [\xi_j - 0.9\sigma_j,\ \xi_j + 0.9\sigma_j],\ j = 1, \ldots, n\}. \tag{19}$$

The aforementioned definitions and notation will be used in the convergence analysis of the algorithm (Section 4).

In the section that follows, a brief analysis of the second-order information of the model functions and its connection with the spectral parameter will motivate a strategy for updating the parameters $\rho_i^{(k,0)}$, $i = 0, 1, \dots, m$ (cf. [9]). The relaxed conservative condition is presented together with the summarized algorithm.

3 Modifications to the MMA

In this section, we propose modifications to the MMA, based on the spectral parameter, used in the updating of the parameter $\rho_i^{(k,\ell)}$, and on relaxing the conservative condition. 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 iteration. The motivation for proposing such an idea came from the numerical observation that Svanberg's original updating stalls for some problems, not making significant progress in the solution of the sequence of solved subproblems. By improving the quality of the approximations such a drawback was overcome. Moreover, the idea preserves the global convergence property of the CCSA class, as proved in Section 4. This proposal was previously presented in [9]. The conservative condition is relaxed by means of a summable controlled forcing sequence, so that the maintenance of global convergence is proved. The algorithm that combines both of these ideas is explicitly presented.

3.1 Second-order information and the spectral parameter

Evaluating the non-mixed second-order partial derivatives of the model function $g_i^{(k,\ell)}$ given in (7) at the current iteration point $x^{(k)}$, we obtain the following expression:

$$\frac{\partial^2 g_i^{(k,\ell)}}{\partial x_j^2}(x^{(k)}) \;=\; \frac{2\left|\frac{\partial f_i}{\partial x_j}(x^{(k)})\right|}{\sigma_j^{(k)}} \;+\; \frac{\rho_i^{(k,\ell)}}{(\sigma_j^{(k)})^2}. \tag{20}$$

Recalling that the functions $g_i^{(k,\ell)}$ are already first-order approximations of the original functions f_i at the current iteration point $x^{(k)}$, as stated in (16), the model will improve if we demand that

$$\frac{\partial^2 g_i^{(k,\ell)}}{\partial x_j^2}(x^{(k)}) = \frac{\partial^2 f_i}{\partial x_j^2}(x^{(k)}) \tag{21}$$

for all $j = 1, \dots, n$. If we somehow approximate the non-mixed second-order derivatives of the functions f_i, and use the relationship (20), we may devise a new strategy for updating the parameters ρ_i. Our choice for approximating the derivatives is based on the spectral parameters, as detailed next.

From the Mean Value Theorem of the Integral Calculus we know that, given a continuously differentiable function $f : I\!R^n \to I\!R$ we have that $f(y) = f(x) + \nabla f(x + \alpha(y - x))^T (y - x)$, for some $\alpha \in (0, 1)$. Moreover, if f is twice continuously differentiable, then $\nabla f(y) = \nabla f(x) + \int_0^1 \nabla^2 f(x + \alpha\,(y - x))\, d\alpha\ (y - x)$. By setting $s = y - x$, the scalar

$$\eta = \frac{s^T t}{s^T s}, \tag{22}$$

where $t = \nabla f(y) - \nabla f(x) = (\int_0^1 \nabla^2 f(x + \alpha s)\, d\alpha) s$ defines a Rayleigh quotient with respect to the average Hessian matrix $(\int_0^1 \nabla^2 f(x + \alpha s)\, d\alpha)$. Such quotient has its value between the smallest and the largest eigenvalue of the average Hessian matrix, what motivates the terminology *spectral* parameter for (22). Thus, if we require that the Hessian of the functions f_i are approximated by scalar matrices, we might say that ηI, with η given in (22) is the matrix of such type that best approximates the average Hessian.

In the MMA context, the idea is to use the spectral parameter (22) as an approximation to the non-mixed second-order derivatives, that is

$$\frac{\partial^2 f_i}{\partial x_j^2}(x) \approx \eta_i,\ \forall j. \tag{23}$$

Such approximation will be used in the beginning of each outer iteration to obtain $\rho_i^{(k,0)}$, thus improving the quality of the model functions.

The points used to compute the direction s are the current estimate and the previous one, that is, $s^{(k)} = x^{(k)} - x^{(k-1)}$. It is also necessary to compute vectors $t_i^{(k)} = \nabla f_i(x^{(k)}) - \nabla f_i(x^{(k-1)})$, so that

$$\eta_i^{(k)} = \frac{(s^{(k)})^T t_i^{(k)}}{(s^{(k)})^T s^{(k)}},\ \ i = 0, 1, \ldots, m. \tag{24}$$

The motivation for using the spectral parameter in the MMA comes from the need of avoiding computing second-order derivatives of the original functions $f_i,\ i = 0, 1, \ldots, m$, due to the intrinsic involved cost. Therefore, a possible way to use second-order information without an excessive overload is to approximate the Hessian by cheap matrices with a simple structure, scalar matrices in this case.

Combining (20), (21) and (23) for $\ell = 0$ we obtain

$$\eta_i^{(k)} \ \approx\ \frac{2\left|\frac{\partial f_i}{\partial x_j}(x^{(k)})\right|}{\sigma_j^{(k)}} \ +\ \frac{\rho_i^{(k,0)}}{(\sigma_j^{(k)})^2}, \qquad \forall j = 1, \ldots, n.$$

Now, $\rho_i^{(k,0)}$ may be computed based on the solution of the following least-squares problem:

$$\underset{\rho_i}{\text{minimize}} \quad \sum_{j=1}^{n} \left(\rho_i \; + \; b_{ij} \; - \; \eta_i^{(k)} \, (\sigma_j^{(k)})^2\right)^2 \tag{25}$$

where

$$b_{ij} \; = \; 2 \, \sigma_j^{(k)} \left| \frac{\partial f_i}{\partial x_j}(x^{(k)}) \right| . \tag{26}$$

As the objective function of problem (25) is a quadratic in ρ_i, it follows that its minimizer ρ_i^* is given by

$$\rho_i^* = \frac{1}{n} \sum_{j=1}^{n} \left(\eta_i^{(k)} (\sigma_j^{(k)})^2 - b_{ij} \right), \tag{27}$$

and it will be conveniently used in our algorithm.

3.2 Relaxing the conservative condition

We have noticed in the numerical investigation that the strict conservative condition, demanded by Svanberg [17] and defined in (6), many times generates assembling and solution of additional subproblems, without significative progress towards the solution, especially in the initial iterations. To overcome this drawback, we decided to relax the conservative condition as follows: we say that the *relaxed conservative condition* holds at the iterate $\hat{x}^{(k,\ell)}$ if

$$f_i(\hat{x}^{(k,\ell)}) \; \leq \; g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) \; + \; \mu_k \, \max\left\{1, \left| g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) \right| \right\}, \tag{28}$$

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

$$\sum_{k=1}^{\infty} \mu_k \leq \mu < \infty. \tag{29}$$

Therefore, the sequence $\{\mu_k\}_{k=1}^{\infty}$ goes to zero as the external iterations progress so that the conservative condition (6) is more relaxed in the beginning of the generated sequence, and ultimately achieved in the end. Naturally, the original conservative condition is obtained if $\mu_k \equiv 0, \forall k$.

In Figure 1 we illustrate an example where the original conservative condition (6) is violated, and for μ_k sufficiently large the relaxed conservative condition (28) is verified. To simplify the notation, we set

$$\bar{\mu}_{k,i} = \mu_k \, \max\left\{1, \left| g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) \right| \right\}. \tag{30}$$

A simple way to relax the conservative condition would be $f_i(\hat{x}^{(k,\ell)}) \; \leq \; g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) \; +$

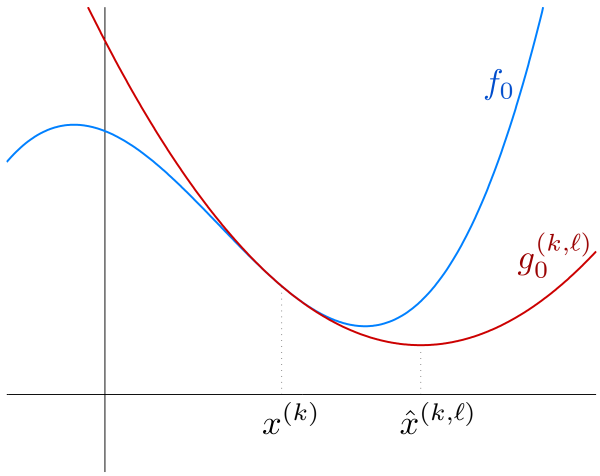

Figure 1: Example in which the original conservative condition is violated and the relaxed conservative condition is verified if μ_k is large enough.

$\mu_k\, g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) = (1+\mu_k)g_i^{(k,\ell)}(\hat{x}^{(k,\ell)})$. However, if $g_i^{(k,\ell)}(\hat{x}^{(k,\ell)})$ is zero or negative, the conservative condition is not relaxed. Therefore, the term $\bar{\mu}_{k,i}$ ensures that there is a relaxation, regardless the value or the signal of the approximation $g_i^{(k,\ell)}$ at $\hat{x}^{(k,\ell)}$. The absolute value in (30) is necessary for the case where the value $g_i^{(k,\ell)}(\hat{x}^{(k,\ell)})$ is negative, whereas the maximum is essential for the case where $g_i^{(k,\ell)}(\hat{x}^{(k,\ell)})$ is zero. The chosen value 1 does not affect the magnitude of the relaxation since if $|g_i^{(k,\ell)}(\hat{x}^{(k,\ell)})|$ is less than 1 in the early iterations, where μ_k can still be large, so we have a large relaxation, which is desired; and if $|g_i^{(k,\ell)}(\hat{x}^{(k,\ell)})|$ is less than 1 in the final iterations, the value of μ_k will be small, so that the relaxation will be small.

The conservative condition (6) provides strict reduction of the objective function value and feasible iterates for problem (2). Hence, by relaxing it, the outer iterates may be infeasible with respect to problem (2), specially in the beginning of the sequence, and its objective function value may increase, but in a controlled way.

When it comes to the global convergence analysis, it is worth mentioning that in [17], the sequence $\{F_0(x^{(k)}, y^{(k)})\}_{k=1}^{\infty}$, where $F_0(x,y) = f_0(x) + \sum(c_i y_i + \frac{1}{2} d_i y_i^2)$, is monotonically decreasing and bounded below, being thus convergent. Adopting the relaxed conservative condition (28), despite losing the monotonic pattern, such a sequence is still convergent due to (29). Moreover, the reasoning of [17] remains valid based on the fact that $\mu_k \to 0$ as $k \to \infty$, as detailed in Section 4, so that the global convergence is maintained. The actual choice for the sequence $\{\mu_k\}_{k=1}^{\infty}$ is provided in Section 6, together with the description of the numerical results.

3.3 The modified MMA algorithm

The MMA with the proposed modifications is presented next, for a given initial estimate $x^{(1)} \in X$. The artificial variables y were omitted on purpose, for simplicity, so the whole process is presented just with the original variables $x \in I\!R^n$. Moreover, we have highlighted our modifications; by removing them, Svanberg's original algorithm is obtained.

Algorithm 1: Method of the Moving Asymptotes with spectral updating and relaxed conservative condition

Step 1. *Initialization*

Define $\rho_i^{(1,0)} = 1$ as in (13a) for $i = 0, 1, \ldots, m$.

Set $k = 1$.

Step 2. *Stopping criterion*

If $x^{(k)}$ verifies the KKT conditions of the problem (2),

stop and take $x^{(k)}$ as the solution.

Step 3. *Computation of the parameters $\rho_i^{(k,0)}$ and $\sigma_j^{(k)}$ and the asymptotes $l_j^{(k)}$ and $u_j^{(k)}$*

Set $\ell = 0$.

If $k > 1$, compute

$$s^{(k)} = x^{(k)} - x^{(k-1)},$$

$$t_i^{(k)} = \nabla f_i(x^{(k)}) - \nabla f_i(x^{(k-1)}) \text{ for } i = 0, 1, \ldots, m,$$

$$\eta_i^{(k)} = \min\left\{\eta_i^{max}, \max\left\{\eta_i^{min}, \frac{(s^{(k)})^T t_i^{(k)}}{(s^{(k)})^T s^{(k)}}\right\}\right\} \text{ for } i = 0, 1, \ldots, m,$$

b_{ij} as in (26), for $i = 0, 1, \ldots, m$ and $j = 1, \ldots, n$,

ρ_i^* as in (27) for $i = 0, 1, \ldots, m$.

For each $i = 0, 1, \ldots, m$, if $\rho_i^* > 0$, set $\rho_i^{(k,0)} = \rho_i^*$.

Otherwise, compute $\rho_i^{(k,0)}$ as in (13b).

Compute $\sigma_j^{(k)}$ for $j = 1, \ldots, n$ as in [17].

Compute $l_j^{(k)} = x_j^{(k)} - \sigma_j^{(k)}$ and $u_j^{(k)} = x_j^{(k)} + \sigma_j^{(k)}$, for $j = 1, \ldots, n$.

Step 4. *Generation and solution of the subproblem*

Compute the coefficients $p_{ij}^{(k,\ell)}$, $q_{ij}^{(k,\ell)}$ and $r_i^{(k,\ell)}$, for $i = 0, 1, \ldots, m$ and $j = 1, \ldots, n$ as in (8), (10) and (12), respectively.

Define the approximating functions $g_i^{(k,\ell)}$, for $i = 0, 1, \ldots, m$ as in (7).

Solve the subproblem (3), obtaining $\hat{x}^{(k,\ell)}$.

Step 5. *Test the relaxed conservative condition*

Compute μ_k such that $\{\mu_k\}_{k=1}^{\infty}$ satisfies (29).

If the relaxed (28) conservative condition is not verified for some index $i \in \{0, 1, \ldots, m\}$ then, for such an index

Update the parameters $\rho_i^{(k,\ell)}$ as in [17], set $\ell = \ell + 1$ and go to **Step 4**.

Otherwise, set $x^{(k+1)} = \hat{x}^{(k,\ell)}$, $k = k + 1$ and go to **Step 2**.

The main differences between our approach and the MMA version of 2002 [17] are the computation of the parameters $\rho_i^{(k,0)}$ in the beginning of each outer iteration and the relaxed conservative condition of Step 5. The spectral parameters used in Step 3, namely $\eta_i^{(k)}$, $i = 0, 1, \ldots, m$, are the projections of $\left((s^{(k)})^T t_i^{(k)}\right) / \left((s^{(k)})^T s^{(k)}\right)$ in the interval $[\eta_i^{min}, \eta_i^{max}]$, with $0 < \eta_i^{min} < \eta_i^{max} < +\infty$. This safeguard is adopted to avoid negative or positive, but either too small or too large values for $\eta_i^{(k)}$. The actual choice for the sequence $\{\mu_k\}_{k=1}^{\infty}$ is provided in Section 6, together with the description of the numerical results.

The global convergence of Algorithm 1 is analyzed in the next section.

4 Theoretical analysis of global convergence

This section contains the convergence properties of the method, combining the spectral updating and the relaxed conservative condition, and described in Algorithm 1.

A given point $(x, y) \in I\!R^n \times I\!R^m$ is a KKT point of the problem (2) if and only if there are Lagrange multipliers which together with (x, y) satisfy the KKT conditions of the problem.

To characterize the solutions of the problem (2), we will show that it is equivalent to the following nonsmooth problem:

$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^{m} \left(c_i f_i^+(x) + \frac{1}{2} d_i \left(f_i^+(x) \right)^2 \right) \\ \text{subject to} & x \in X, \end{array} \tag{31}$$

where $f_i^+(x) = \max\{0, f_i(x)\}$ and X is the compact set:

$$X = \{x \in I\!R^n; x_j^{min} \le x_j \le x_j^{max}, j = 1, \ldots, n\}.$$

The formulation (31) will not be used for solving problem (2), but solely to show that the problem (2) admits a global solution. The artificial variables y_i may be eliminated from the problem (2). The resulting problem is exactly (31). The following proposition establishes the equivalence between the solutions of the problems (2) and (31).

Proposition 1. *The pair $(\hat{x}, \hat{y})$ is a global optimal solution of problem (2) if and only if $\hat{x}$ is a global optimal solution of problem (31), where $\hat{y}_i = f_i^+(\hat{x})$, $i = 1, \ldots, m$.*

Proof: It is easy to see that if $x \in X$ then $(x, f^+(x))$ is feasible for (2). Moreover, if (x, y) is feasible for (2) then $y_i \ge f_i^+(x)$ for all $i = 1, \ldots, m$, and consequently

$$f_0(x) + \sum_{i=1}^{m} \left(c_i f_i^+(x) + \frac{d_i}{2}(f_i^+(x))^2 \right) \le f_0(x) + \sum_{i=1}^{m} \left(c_i y_i + \frac{d_i}{2} y_i^2 \right). \tag{32}$$

Now, let $(\hat{x}, \hat{y})$ be the global optimal solution of problem (2) and $x \in X$ be an arbitrary feasible point for (31). Since $(x, f^+(x))$ is feasible for (2), we have that

$$\begin{aligned} f_0(\hat{x}) + \sum_{i=1}^{m} \left(c_i f_i^+(\hat{x}) + \frac{d_i}{2}(f_i^+(\hat{x}))^2 \right) &\le f_0(\hat{x}) + \sum_{i=1}^{m} \left(c_i \hat{y}_i + \frac{d_i}{2} \hat{y}_i^2 \right) \\ &\le f_0(x) + \sum_{i=1}^{m} \left(c_i f_i^+(x) + \frac{d_i}{2}(f_i^+(x))^2 \right). \end{aligned}$$

Conversely, if $\hat{x}$ is a global optimal solution of problem (31), then $\hat{x} \in X$, and consequently $(\hat{x}, f^+(\hat{x}))$ is feasible for (2). On the other hand, if (x, y) is feasible for (2), we have that x is feasible for (31), and, by (32), we have that

$$\begin{aligned} f_0(\hat{x}) + \sum_{i=1}^{m} \left(c_i f_i^+(\hat{x}) + \frac{d_i}{2}(f_i^+(\hat{x}))^2 \right) &\le f_0(x) + \sum_{i=1}^{m} \left(c_i f_i^+(x)_i + \frac{d_i}{2}(f_i^+(x))^2 \right) \\ &\le f_0(x) + \sum_{i=1}^{m} \left(c_i y_i + \frac{d_i}{2} y_i^2 \right). \quad \blacksquare \end{aligned}$$

The next proposition ensures that the problem under consideration is well formulated.

Proposition 2. *There is at least one global optimal solution of problem (2).*

Proof: From the continuity of the objective function of problem (31) in the compact set X, it follows that the problem (31) admits at least one global optimal solution. Therefore, from Proposition 1, the problem (2) also admits at least one global optimal solution. ■

The following result establishes that the feasible set of problem (2) verifies a constraint qualification.

Proposition 3. *If $(\hat{x}, \hat{y})$ is an optimal solution, local or global, of problem (2), then there are Lagrange multipliers which together with $(\hat{x}, \hat{y})$ satisfy the KKT conditions for such a problem.*

Proof: Consider the Arrow, Hurwicz and Uzawa constraint ([2], see also [3, p.329]): let x^* be a local minimizer of problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_1(x) \leq 0, \ldots, g_m(x) \leq 0, \end{array} \tag{33}$$

where f and g_i are continuously differentiable real-valued functions. If there is a vector $d \in I\!R^n$ such that

$$\nabla g_i(x^*)^T d < 0, \quad \forall\, i \in A(x^*)$$

where $A(x)$ is the index set of the active constraints of the problem (33) in x, for any $x \in I\!R^n$, then there are Lagrange multipliers which together with x^*, satisfy the KKT conditions of the problem (33).

Such a constraint qualification is a particular case of the well-known Mangasarian-Fromovitz constraint qualification ([12], see also [3, p.329]) for the inequality constrained problem, and it holds for problem (2). Indeed, assume that $(\hat{x}, \hat{y})$ is an optimal solution of problem (2) and construct a corresponding vector $d = (d_x, d_y) \in I\!R^n \times I\!R^m$ as follows:

$$(d_x)_j = \left\{ \begin{array}{rl} 1, & \text{if } \hat{x}_j = x_j^{min} \\ -1, & \text{if } \hat{x}_j = x_j^{max}, \\ 0, & \text{otherwise} \end{array} \right. \quad j = 1, \ldots, n, \qquad (d_y)_i = 1 + \sum_{j=1}^{n} \left| \frac{\partial f_i}{\partial x_j}(\hat{x}) \right|, \quad i = 1, \ldots, m.$$

As a result, the inner product of d and the gradient vector, calculated at $(\hat{x}, \hat{y})$, of any active constraint in problem (2) is strictly negative. Hence there exists a Lagrange multiplier vector that together with $(\hat{x}, \hat{y})$ satisfies the KKT conditions of problem (2). ■

Let Ω be the set of KKT points of the problem (2). Such a set is nonempty, as established in Propositions 2 and 3. Let $||\Omega - (x^{(k)}, y^{(k)})||$ denote the Euclidean distance from the point

$(x^{(k)}, y^{(k)})$ to the set Ω, i.e.,

$$||\Omega - (x^{(k)}, y^{(k)})|| = \inf_{(x,y)\in\Omega} \{||(x, y) - (x^{(k)}, y^{(k)})||\}.$$

Theorem 1. *If the Algorithm 1 is applied to a problem of the form (2), then* $||\Omega - (x^{(k)}, y^{(k)})|| \to 0$ *as* $k \to \infty$.

In other words, every limit point of the sequence generated by the Algorithm 1 is a KKT point of the problem (2). To prove Theorem 1, some preliminary results are needed.

In the first lemma we prove that Algorithm 1 is well-defined, that is, the relaxed conservative condition is verified after a finite number of inner iterations.

Lemma 1. *In each outer iteration* k*, only a finite number* ℓ *of inner iterations are needed until* $f_i(\hat{x}^{(k,\ell)}) \le g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) + \bar{\mu}_{k,i}$ *for all* i*, where* $\bar{\mu}_{k,i}$ *is given by (30).*

Proof: By Taylor's theorem we have

$$g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) \;=\; g_i^{(k,\ell)}(x^{(k)}) \;+\; \nabla g_i^{(k,\ell)}(x^{(k)})^T p \;+\; \frac{1}{2} p^T \nabla^2 g_i^{(k,\ell)}(x^{(k)} + t_1 p) p$$

and

$$f_i(\hat{x}^{(k,\ell)}) \;=\; f_i(x^{(k)}) \;+\; \nabla f_i(x^{(k)})^T p \;+\; \frac{1}{2} p^T \nabla^2 f_i(x^{(k)} + t_2 p) p$$

for some $t_1 \in (0, 1)$, some $t_2 \in (0, 1)$ and $p = \hat{x}^{(k,\ell)} - x^{(k)}$. From (17) we obtain

$$\nabla^2 g_i^{(k,\ell)}(x) = \nabla^2_{xx} v_i(x, x^{(k)}, \sigma^{(k)}) + \rho_i^{(k,\ell)} \nabla^2_{xx} w(x, x^{(k)}, \sigma^{(k)})$$

and using (16) we get

$$\begin{aligned} g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) - f_i(\hat{x}^{(k,\ell)}) + \bar{\mu}_{k,i} &= \frac{1}{2} p^T \nabla^2 g_i^{(k,\ell)}(x^{(k)} + t_1 p) p - \frac{1}{2} p^T \nabla^2 f_i(x^{(k)} + t_2 p) p + \bar{\mu}_{k,i} \\ &= \frac{1}{2} p^T \nabla^2_{xx} v_i(x^{(k)} + t_1 p, x^{(k)}, \sigma^{(k)}) p \\ &+ \frac{\rho_i^{(k,\ell)}}{2} p^T \nabla^2_{xx} w(x^{(k)} + t_1 p, x^{(k)}, \sigma^{(k)}) p \\ &- \frac{1}{2} p^T \nabla^2 f_i(x^{(k)} + t_2 p) p \;+\; \bar{\mu}_{k,i}. \end{aligned}$$

But $\nabla^2_{xx} v_i(x, \xi, \sigma) \ge 0, \; \forall (x, \xi, \sigma) \in D$, and so

$$\begin{aligned} g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) \;-\; f_i(\hat{x}^{(k,\ell)}) + \bar{\mu}_{k,i} &\ge \frac{\rho_i^{(k,\ell)}}{2} p^T \nabla^2_{xx} w(x^{(k)} + t_1 p, x^{(k)}, \sigma^{(k)}) p \\ &- \frac{1}{2} p^T \nabla^2 f_i(x^{(k)} + t_2 p) p \;+\; \bar{\mu}_{k,i}. \end{aligned}$$

Dividing both sides by $\|p\|^2$ we have

$$\begin{aligned}\frac{1}{\|p\|^2}\left(g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) \;-\; f_i(\hat{x}^{(k,\ell)}) + \bar{\mu}_{k,i}\right) \;&\geq\; \frac{\rho_i^{(k,\ell)}}{2\|p\|^2} p^T \nabla^2_{xx} w(x^{(k)} + t_1 p, x^{(k)}, \sigma^{(k)}) p \\ &-\; \frac{1}{2\|p\|^2} p^T \nabla^2 f_i(x^{(k)} + t_2 p) p \;+\; \frac{\bar{\mu}_{k,i}}{\|p\|^2}.\end{aligned}$$

Defining

$$\delta(X) \equiv \sup_{x,y \in X} \|x - y\|,$$

it holds that $\|p\| \leq \delta(X)$ and so

$$\begin{aligned}\frac{1}{\|p\|^2}\left(g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) \;-\; f_i(\hat{x}^{(k,\ell)}) + \bar{\mu}_{k,i}\right) \;&\geq\; \frac{\rho_i^{(k,\ell)}}{2\|p\|^2} p^T \nabla^2_{xx} w(x^{(k)} + t_1 p, x^{(k)}, \sigma^{(k)}) p \\ &-\; \frac{1}{2\|p\|^2} p^T \nabla^2 f_i(x^{(k)} + t_2 p) p \;+\; \frac{\bar{\mu}_{k,i}}{\delta(X)^2}.\end{aligned}$$

Considering the worst case for which the relaxed conservative condition is satisfied, we define the scalars τ and κ_i such that

$$\tau \;=\; \min_{x,\xi,\sigma,h} \{\, h^T \nabla^2_{xx} w(x,\xi,\sigma) h \mid (x,\xi,\sigma) \in D,\, h \in I\!R^n,\, h^T h = 1\}, \tag{34}$$

$$\kappa_i \;=\; \max_{x,h} \{\, h^T \nabla^2 f_i(x) h \mid x \in X,\, h \in I\!R^n,\, h^T h = 1\}, \tag{35}$$

with the set D given in (18).

As all the sets under consideration are compact, the minimum and maximum values in expressions (34) and (35), respectively, are well defined. The scalar κ_i is finite because the Hessian matrix $\nabla^2 f_i(x)$ is continuous in X. The scalar τ is finite and strictly positive because the Hessian matrix $\nabla^2_{xx} w(x,\xi,\sigma)$ is positive definite and continuous in all its arguments.

Hence, a sufficient condition for the inequality $g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) + \bar{\mu}_{k,i} \geq f_i(\hat{x}^{(k,\ell)})$ to hold is that

$$\rho_i^{(k,\ell)} \tau + \frac{2\bar{\mu}_{k,i}}{\delta(X)^2} \geq \kappa_i.$$

But each time that $g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) + \bar{\mu}_{k,i} < f_i(\hat{x}^{(k,\ell)})$, the corresponding $\rho_i^{(k,\ell)}$ is increased by at least a factor 1.1 (see (15)). This can be done only a finite number of times, for each i, before $\rho_i^{(k,\ell)} \tau + 2\bar{\mu}_{k,i}/\delta(X)^2 \geq \kappa_i$ is satisfied. (Note that, for a fixed k, the values $\rho_i^{(k,\ell)}$ are nondecreasing in ℓ). ■

As a consequence of Lemma 1, only outer iterations need to be considered in the analysis of global convergence. Therefore, the following shorter notations will be used:

$\hat{\ell}(k)$ = the number of inner iterations needed within the k-th outer iteration,

$\rho_i^{(k)} = \rho_i^{(k,\hat{\ell}(k))}$ and $g_i^{(k)}(x) = g_i^{(k,\hat{\ell}(k))}(x)$.

This means that the subproblem used at the k-th (outer) iteration to calculate the next iteration point (that is, whose optimal solution should satisfy the relaxed conservative condition) is the following:

$$\begin{array}{lll} \text{minimize} & g_0^{(k)}(x) + \displaystyle\sum_{i=1}^{m}(c_i y_i + \tfrac{1}{2} d_i y_i^2) & \\ \text{subject to} & g_i^{(k)}(x) - y_i \leq 0, & i = 1, \ldots, m, \\ & x \in X^{(k)},\ y \geq 0, & \end{array} \tag{36}$$

with $X^{(k)}$ given in (4). The optimal solution of the subproblem (36) is the new iterate $(x^{(k+1)}, y^{(k+1)})$. Note that $f_i(x^{(k)}) = g_i^{(k)}(x^{(k)})$ and $f_i(x^{(k+1)}) \leq g_i^{(k)}(x^{(k+1)}) + \bar{\mu}_{k,i}$ for all $i = 0, 1, \ldots, m$.

In the next lemma, an upper bound for the parameters $\rho_i^{(k)},\ i = 0, 1, \ldots, m$ is established. The spectral parameter is employed in the proof.

Lemma 2. *For each $i = 0, 1, \ldots, m$, there is a finite number $\rho_i^{\max}$ such that $\rho_i^{(k)} \leq \rho_i^{\max}$ for all outer iterations k.*

Proof: From the proof of Lemma 1, if $\rho_i^{(k,0)}\tau + 2\bar{\mu}_{k,i}/\delta(X)^2 \geq \kappa_i$ for all k, then the relaxed conservative condition always hold. Thus, for the outer iterations k for which the values $\rho_i^{(k,0)}$ generated by Algorithm 1 are computed from (13b), an upper bound for $\rho_i^{(k)}$ is $\rho_i^{\max} = \rho_i^{(1,0)} = 1$. For the outer iterations k for which the values $\rho_i^{(k,0)}$ are computed by the expression $\rho_i^* = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n}\left(\eta_i^{(k)}(\sigma_j^{(k)})^2 - b_{ij}\right)$ for $i = 0, 1, \ldots, m$, that rests upon the spectral parameter, as $\eta_i^{\min} \leq \eta_i^{(k)} \leq \eta_i^{\max}$, $\sigma_j^{\min} \leq \sigma_j^{(k)} \leq \sigma_j^{\max}$ and $b_{ij} \geq 0$ for all $i = 0, 1, \ldots, m$ and $j = 1, \ldots, n$, an upper bound for $\rho_i^{(k)}$is $\rho_i^{\max} = \max\{1, \eta_i^{\max}\hat{\sigma}^2\}$ where $\hat{\sigma} = \max\limits_j\{\sigma_j^{\max}\}$.

Now, as $\hat{\ell}(k)$ is the number of inner iterations needed within the k-th outer iteration, the parameter $\rho_i^{(k,\hat{\ell}(k))} = \rho_i^{(k)}$ was in charge for the fulfillment of the relaxed conservative condition, whereas such a condition was not verified for the parameter $\rho_i^{(k,\hat{\ell}(k)-1)}$. Thus

$$\rho_i^{(k,\hat{\ell}(k)-1)}\tau + 2\bar{\mu}_{k,i}/\delta(X)^2 < \kappa_i.$$

From (15) we conclude that $\rho_i^{(k,\ell+1)} \leq 10\rho_i^{(k,\ell)}$ for each inner iteration ℓ inside the outer iteration k, and so:

$$\rho_i^{(k)} \;=\; \rho_i^{(k,\hat{\ell}(k))} \;\leq\; 10\,\rho_i^{(k,\hat{\ell}(k)-1)} \;<\; 10\left(\frac{\kappa_i}{\tau} - \frac{2\bar{\mu}_{k,i}}{\tau\delta(X)^2}\right) \;<\; 10\frac{\kappa_i}{\tau}. \tag{37}$$

Thus, the effective upper bound for $\rho_i^{(k)}$ is

$$\rho_i^{max} = \max\left\{1,\ \eta_i^{max}\hat{\sigma}^2,\ 10\frac{\kappa_i}{\tau}\right\}. \quad \blacksquare$$

It is worth noting that the upper bound for $\rho_i^{(k)}$ computed in (37) is the same obtained by Svanberg in [17]. Although the theoretical upper bound for $\rho_i^{(k)}$ did not decrease by relaxing the conservative condition, in practice, as $\bar{\mu}_{k,i} > 0, \forall k$ and $\tau > 0$, smaller values for $\rho_i^{(k)}$ are used in each outer iterations k, so that fewer inner iterations are necessary inside each outer iteration. The impact of this modification will be clear in the section of the numerical results.

Let the set Q be defined by

$$Q = \{\rho \in I\!R^{m+1} \mid 0 < \rho_i^{\min} \leq \rho_i \leq \rho_i^{\max},\ i = 0, 1, \ldots, m\}.$$

Note that the lower bounds $\rho_i^{\min}$ for the parameters ρ_i must be strictly positive, because it is the nonnegativity of such parameters that provides strictly positive coefficients p_{ij} and q_{ij} simultaneously, so that both asymptotes are active in the MMA approximation.

Let the functions F_i be defined, for $x \in X$ and $y \in I\!R^m$, by

$$\begin{aligned} F_0(x,y) &= f_0(x) + \sum_{i=1}^{m}(c_i y_i + \tfrac{1}{2} d_i y_i^2), \\ F_i(x,y) &= f_i(x) - y_i, \quad i = 1, \ldots, m. \end{aligned}$$

Then the problem (2) can be written as

$$\begin{array}{ll} \text{minimize} & F_0(x,y) \\ \text{subject to} & F_i(x,y) \leq 0, \quad\; i = 1, \ldots, m, \\ & x \in X,\ y \geq 0. \end{array} \tag{38}$$

Let the functions G_i be defined, for $(x, \xi, \sigma) \in D, \rho \in Q$ and $y \in I\!R^m$, by

$$\begin{aligned} G_0(x,y,\xi,\sigma,\rho) &= v_0(x,\xi,\sigma) + \rho_0\, w(x,\xi,\sigma) + \sum_{i=1}^{m}(c_i y_i + \tfrac{1}{2} d_i y_i^2), \\ G_i(x,y,\xi,\sigma,\rho) &= v_i(x,\xi,\sigma) + \rho_i\, w(x,\xi,\sigma) - y_i, \quad i = 1, \ldots, m. \end{aligned}$$

Note that each function G_i is continuous on the set on which it is defined.

Let the problem PSUB(ξ, σ, ρ) be defined, for given $(\xi, \sigma, \rho) \in X \times S \times Q$, as the following problem in the variables (x, y):

$$\begin{aligned} \text{minimize} \quad & G_0(x, y, \xi, \sigma, \rho) \\ \text{subject to} \quad & G_i(x, y, \xi, \sigma, \rho) \leq 0, \quad i = 1, \ldots, m, \\ & x \in X(\xi, \sigma),\ y \geq 0, \end{aligned} \tag{39}$$

with $X(\xi, \sigma)$ defined in (19). Then, the subproblem (36) is equivalent to the problem PSUB$(x^{(k)}, \sigma^{(k)}, \rho^{(k)})$.

The next two lemmas are presented as in [17]. They are concerned with the solution of problem PSUB(ξ, σ, ρ), labeled by (39), and the relaxed conservative condition is not used in their proofs.

Lemma 3. *For each given $\xi \in X, \sigma \in S$ and $\rho \in Q$, there is a unique optimal solution of PSUB(ξ, σ, ρ). This solution is also the only KKT point of PSUB(ξ, σ, ρ).*

Proof: The existence of an optimal solution follows by arguments similar to those in the proof of Proposition 2. The uniqueness follows from the fact that the problem obtained by eliminating y is strictly convex in x. Consider the Slater constraint qualification (see e.g. [3, p.331]): let x^* be a local minimizer of the problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_1(x) \leq 0, \ldots, g_m(x) \leq 0, \end{aligned} \tag{40}$$

where$f$ and g_i are continuously differentiable real-valued functions. Assume that the functions f and g_i are convex and there is a feasible point $\bar{x}$ such that

$$g_i(\bar{x}) < 0, \quad \forall\, i \in A(x^*)$$

where $A(x)$ defines the index set of the active constraints of the problem (40) in x, for any point $x \in I\!R^n$. Then there exist Lagrange multipliers such that, together with x^*, satisfy the KKT conditions of the problem (40).

Now, given a vector d such that $\nabla g_i(x^*)^T d < 0$, $\forall i \in A(x^*)$ then there exists a point $\bar{x}$ such that $g_i(\bar{x}) < 0$, $\forall i \in A(x^*)$. Thus, setting d as in the proof of Proposition 3, the Slater's constraint qualifications are fulfilled for the convex problem PSUB(ξ, σ, ρ). Therefore, the KKT conditions are both necessary and sufficient conditions for a global optimum. ■

For the unique optimal solution of PSUB(ξ, σ, ρ) to be the unique KKT point of PSUB(ξ, σ, ρ), it would be enough that such a solution verified some constraint qualification, like the one of Arrow, Hurwicz and Uzawa (used in the proof of Proposition 3). The Slater's constraint qualifications suits better because the functions G_i, $i = 1, \ldots, m$ are convex. Hence, $(x^{(k+1)}, y^{(k+1)})$ is the only KKT point of PSUB$(x^{(k)}, \sigma^{(k)}, \rho^{(k)})$.

Lemma 4. *For each given $\sigma \in S$ and $\rho \in Q$ the following holds: a given point $(\hat{x}, \hat{y})$ is a KKT point of the problem (2) if and only if $(\hat{x}, \hat{y})$ is a KKT point of the subproblem PSUB$(\hat{x}, \sigma, \rho)$.*

Proof: For a given $\hat{x} \in X$, let $B(\hat{x}, \epsilon) = \{x \in I\!R^n \mid ||x - \hat{x}|| < \epsilon\}$, and note that there is an $\epsilon > 0$ such that $X \cap B(\hat{x}, \epsilon) = X(\hat{x}, \sigma) \cap B(\hat{x}, \epsilon)$. This implies that $(\hat{x}, \hat{y})$ is the optimal solution of (the strictly convex problem) PSUB$(\hat{x}, \sigma, \rho)$ if and only if $(\hat{x}, \hat{y})$ is the optimal solution of PSUB$(\hat{x}, \sigma, \rho)$ with the simple bound constraints $x \in X(\hat{x}, \sigma)$ replaced by the (looser) simple bound constraints $x \in X$. Further, the following holds for $i = 0, 1, \ldots, m$:

$$\begin{aligned} G_i(\hat{x}, \hat{y}, \hat{x}, \sigma, \rho) &= F_i(\hat{x}, \hat{y}), \\ \frac{\partial G_i}{\partial x_j}(\hat{x}, \hat{y}, \hat{x}, \sigma, \rho) &= \frac{\partial F_i}{\partial x_j}(\hat{x}, \hat{y}), \\ \frac{\partial G_i}{\partial y_i}(\hat{x}, \hat{y}, \hat{x}, \sigma, \rho) &= \frac{\partial F_i}{\partial y_i}(\hat{x}, \hat{y},). \end{aligned}$$

These observations imply that $(\hat{x}, \hat{y})$ is a KKT point of the subproblem PSUB$(\hat{x}, \sigma, \rho)$ if and only if $(\hat{x}, \hat{y})$ is a KKT point of the problem (38). ■

In particular, if $(x^{(k+1)}, y^{(k+1)}) = (x^{(k)}, y^{(k)})$, then $(x^{(k)}, y^{(k)})$ is a KKT point of the problem (2), and then the algorithm should be stopped. From now on, it is therefore assumed that $(x^{(k+1)}, y^{(k+1)}) \neq (x^{(k)}, y^{(k)})$ for all k.

Whenever the conservative condition (6) is verified for the problem (2) and the solution of a subproblem is accepted as the next outer iterate, one ensures the strict reduction of the objective function values, that is,

$$\begin{aligned} & f_0(x^{(k+1)}) + \sum_{i=1}^{m} \left(c_i\, y_i^{(k+1)} + \tfrac{1}{2}\, d_i\, (y_i^{(k+1)})^2 \right) \\ \leq\ & g_0^{(k,\ell)}(x^{(k+1)}) + \sum_{i=1}^{m} \left(c_i\, y_i^{(k+1)} + \tfrac{1}{2}\, d_i\, (y_i^{(k+1)})^2 \right) \\ <\ & g_0^{(k,\ell)}(x^{(k)}) + \sum_{i=1}^{m} \left(c_i\, y_i^{(k)} + \tfrac{1}{2}\, d_i\, (y_i^{(k)})^2 \right) \\ =\ & f_0(x^{(k)}) + \sum_{i=1}^{m} \left(c_i\, y_i^{(k)} + \tfrac{1}{2}\, d_i\, (y_i^{(k)})^2 \right). \end{aligned} \tag{41}$$

Moreover, it guarantees feasible iterates with respect to problem (2), that is, for $i = 1, \ldots, m$,

$$f_i(x^{(k+1)}) \; - \; y_i^{(k+1)} \; \leq \; g_i^{(k,\ell)}(x^{(k+1)}) \; - \; y_i^{(k+1)} \; \leq \; 0. \tag{42}$$

Thus, the original conservative condition (6) provides strict reduction in the objective function value F_0 and feasible iterates for the problem (38). In the next lemma we prove that, when it comes to the relaxed conservative condition (28), the outer iterates of problem (38) might be infeasible and the values F_0 might increase, but in a controlled way.

Lemma 5. *The infeasibility of each generated iteration point with respect to problem (38) is controlled as follows:* $F_i(x^{(k)}, y^{(k)}) \leq \bar{\mu}_{k-1,i}$ *for* $i \geq 1$ *and* $k \geq 2$. *Furthermore,* $F_0(x^{(k+1)}, y^{(k+1)}) < F_0(x^{(k)}, y^{(k)}) + \bar{\mu}_{k,0}$ *for* $k \geq 1$.

Proof: The initial estimate $(x^{(1)}, y^{(1)})$ is feasible by construction. Thus, for $i \geq 1$ we have

$$F_i(x^{(k+1)}, y^{(k+1)}) \; \leq \; G_i(x^{(k+1)}, y^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) \; + \; \bar{\mu}_{k,i} \; \leq \; \bar{\mu}_{k,i},$$

respectively from the relaxed conservative condition, and the feasibility of problem PSUB$(x^{(k)}, \sigma^{(k)}, \rho^{(k)})$. Besides,

$$\begin{aligned} F_0(x^{(k+1)}, y^{(k+1)}) \; &\leq \; G_0(x^{(k+1)}, y^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) + \bar{\mu}_{k,0} \\ &< \; G_0(x^{(k)}, y^{(k)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) + \bar{\mu}_{k,0} \\ &= \; F_0(x^{(k)}, y^{(k)}) + \bar{\mu}_{k,0}, \end{aligned}$$

where the first inequality holds due to the relaxed conservative condition, the second due to Lemma 3 and the assumption that $(x^{(k+1)}, y^{(k+1)}) \neq (x^{(k)}, y^{(k)})$ for all k, and the equality comes from the fact that the models are first order approximations to the original functions at the current estimate. ■

Lemma 5 is distinct from Lemma 7.6 of [17] because by relaxing the conservative condition the outer iterates are no longer feasible with respect to the original augmented problem. Moreover, the objective function values of problem (38) for the generated sequence does not strictly decrease any longer. In spite of that, as we will show, since the elements of the sequence $\bar{\mu}_{k,i}$, defined in (30), go to zero as $k \to \infty$, the feasibility and strict reduction of the objective function are both asymptotically fulfilled.

Indeed, let the functions $\tilde{g}_i$, for $(x, \xi, \sigma) \in D$ and $\rho \in Q$, be defined by

$$\tilde{g}_i(x, \xi, \sigma, \rho) = v_i(x, \xi, \sigma) + \rho_i \, w(x, \xi, \sigma).$$

Each function $\tilde{g}_i$ is continuous on the compact set on which it is defined. Therefore,

$$|g_i^{(k,\ell)}(\hat{x}^{(k,\ell)})| \le \max\{\tilde{g}_i(x,\xi,\sigma,\rho)|(x,\xi,\sigma) \in D, \rho \in Q\} \le M < \infty,$$

for all k and i, and consequently $\{\bar{\mu}_{k,i}\}$ is summable and thus convergent to zero.

Lemma 6. *All the iteration points $(x^{(k)}, y^{(k)})$ remain in a compact set.*

Proof: Each iterate $x^{(k)}$ belongs to the compact set X. If $x \in X$ is held fixed in problem (36), the following problem is obtained in the variable $y \in I\!R^m$

$$\begin{array}{lll} \underset{y}{\text{minimize}} & M + \sum_{i=1}^{m} \left(c_i y_i + \frac{d_i}{2} y_i^2\right) & \\ \text{subject to} & N_i - y_i \le 0, & i = 1,\ldots,m, \\ & y \ge 0, & \end{array} \tag{43}$$

where $M = g_0^{(k)}(x)$ and $N_i = g_i^{(k)}(x)$, for $i = 1,\ldots,m$.

Note that problem (43) is separable, being thus equivalent to m unidimensional problems

$$\begin{array}{ll} \underset{y_i}{\text{minimize}} & c_i y_i + \frac{d_i}{2} y_i^2 \\ \text{subject to} & N_i - y_i \le 0, \\ & y_i \ge 0. \end{array} \tag{44}$$

Observe that (44) has a unique optimal solution $y_i = \max\{0, N_i\} = \max\{0, g_i^{(k)}(x)\} = g_i^+(x)$, because $c_i > 0$ and $d_i > 0$, $\forall i \in \{1,\ldots,m\}$.

Holding $x^{(k+1)} \in X^{(k)}$ fixed, we obtain $y_i^{(k+1)} = \max\{0, g_i^{(k)}(x^{(k+1)})\}$. Hence,

$$\begin{aligned} 0 \le y_i^{(k+1)} &\le g_i^{(k)}(x^{(k+1)}) = \tilde{g}_i(x^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) \\ &\le \max\{\tilde{g}_i(x,\xi,\sigma,\rho) \,|\, (x,\xi,\sigma) \in D, \rho \in Q\}. \quad \blacksquare \end{aligned}$$

As a consequence of Lemma 6, the sequence $\{(x^{(k)}, y^{(k)})\}_{k=1}^{\infty}$ has at least one convergent subsequence. Thus, there is a point (x^*, y^*) and an infinite subset $\mathcal{K}$ of the positive integers such that $\{(x^{(k)}, y^{(k)})\} \to (x^*, y^*)$ as $k \in \mathcal{K}$ and $k \to \infty$.

Further, since the sequence $\{(\sigma^{(k)}, \rho^{(k)})\}_{k\in\mathcal{K}}$ (with $\mathcal{K}$ from above) stays in the compact set $S \times Q$, there is a point $(\sigma^*, \rho^*) \in S \times Q$ and an infinite subset $\tilde{\mathcal{K}} \subseteq \mathcal{K}$ such that $(\sigma^{(k)}, \rho^{(k)}) \to (\sigma^*, \rho^*)$ as $k \in \tilde{\mathcal{K}}$ and $k \to \infty$.

Next, from Lemma 6, the sequence $\{(x^{(k+1)}, y^{(k+1)})\}_{k\in\tilde{\mathcal{K}}}$ (with $\tilde{\mathcal{K}}$ from above) also has at least one convergent subsequence. Thus, there is a point $(\bar{x}, \bar{y})$ and an infinite subset

$\bar{\mathcal{K}} \subseteq \tilde{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{(x^{(k+1)}, y^{(k+1)})\} \to (\bar{x}, \bar{y})$ as $k \in \bar{\mathcal{K}}$ and $k \to \infty$.

In [17], the whole sequence of objective function values of problem (2), namely $\{F_0(x^{(k)}, y^{(k)})\}_{k=1}^{\infty}$, is convergent because it is monotonically decreasing and bounded below. By adopting the relaxed conservative condition, we have lost the strict decreasing of the values F_0, as established in Lemma (5). Despite that, the sequence $\{F_0(x^{(k)}, y^{(k)})\}_{k=1}^{\infty}$ generated by Algorithm 1 is convergent as well, as proved in the next result.

Lemma 7. *The sequence* $\{F_0(x^{(k)}, y^{(k)})\}_{k=1}^{\infty}$ *is convergent.*

Proof: Given an arbitrary $\epsilon > 0$, since $\{\bar{\mu}_{k,i}\}$ is summable, there is a $k_0 \in I\!N$ such that

$$\sum_{\kappa=k_0}^{\infty} \bar{\mu}_{\kappa,0} \leq \frac{\epsilon}{2}.$$

Since $\{(x^{(k)}, y^{(k)})\}$ remains in a compact set, and F_0 is a continuous function, we have that $\{F_0(x^{(k)}, y^{(k)})\}$ is bounded. Let F^* be the smallest limit point of $\{F_0(x^{(k)}, y^{(k)})\}$, i.e.

$$F^* = \liminf F_0(x^{(k)}, y^{(k)}) \in (-\infty, \infty).$$

By definition of lim inf there exists some $k_1 \geq k_0$ such that $F_0(x^{(k)}, y^{(k)}) > F^* - \frac{\epsilon}{2}$ for all $k \geq k_1$. Moreover, there is $k_2 \geq k_1$ such that $|F_0(x^{(k_2)}, y^{(k_2)}) - F^*| \leq \frac{\epsilon}{2}$.

For an arbitrary $k \geq k_2$ we have two possibilities: either $F_0(x^{(k)}, y^{(k)}) \leq F_0(x^{(k_2)}, y^{(k_2)})$ or $F_0(x^{(k)}, y^{(k)}) > F_0(x^{(k_2)}, y^{(k_2)})$. In the first case it is clear that $|F_0(x^{(k)}, y^{(k)}) - F^*| \leq \frac{\epsilon}{2}$. In the second one, we use Lemma 5 to ensure that

$$\begin{aligned} F_0(x^{(k)}, y^{(k)}) &\leq F_0(x^{(k-1)}, y^{(k-1)}) + \bar{\mu}_{k-1,0} \\ &\leq F_0(x^{(k-2)}, y^{(k-2)}) + \bar{\mu}_{k-1,0} + \bar{\mu}_{k-2,0} \\ &\vdots \\ &\leq F_0(x^{(k_2)}, y^{(k_2)}) + \sum_{\kappa=k_2}^{k-1} \bar{\mu}_{\kappa,0} \\ &\leq F_0(x^{(k_2)}, y^{(k_2)}) + \sum_{\kappa=k_2}^{\infty} \bar{\mu}_{\kappa,0}. \end{aligned}$$

Since, by hypothesis, $F_0(x^{(k)}, y^{(k)}) - F_0(x^{(k_2)}, y^{(k_2)}) > 0$, we have that $|F_0(x^{(k)}, y^{(k)}) - F_0(x^{(k_2)}, y^{(k_2)})| \leq \frac{\epsilon}{2}$, and then

$$|F_0(x^{(k)}, y^{(k)}) - F^*| \leq |F_0(x^{(k)}, y^{(k)}) - F_0(x^{(k_2)}, y^{(k_2)})| + |F_0(x^{(k_2)}, y^{(k_2)}) - F^*| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2}. \quad \blacksquare$$

In the next two lemmas we prove that the objective function values at the limit points coincide.

Lemma 8. *$F_0(x^{(k)}, y^{(k)}) \to F_0(x^*, y^*)$ as $k \to \infty$ (not only for $k \in \mathcal{K}$).*

Proof: From Lemma 7 the sequence $\{F_0(x^{(k)}, y^{(k)})\}_{k=1}^{\infty}$ is convergent, that it, $F_0(x^{(k)}, y^{(k)}) \to F_0^*$ as $k \to \infty$, for some real number F_0^*. But since $F_0(x^{(k)}, y^{(k)}) \to F_0(x^*, y^*)$ as $k \in \mathcal{K}$ and $k \to \infty$, it follows that $F_0^* = F_0(x^*, y^*)$. ■

Lemma 9. *$F_0(\bar{x}, \bar{y}) = F_0(x^*, y^*)$.*

Proof: From Lemma 8 it follows that $F_0(x^{(k+1)}, y^{(k+1)}) \to F_0(x^*, y^*)$ as $k \in \bar{\mathcal{K}}$ and $k \to \infty$. But since $(x^{(k+1)}, y^{(k+1)}) \to (\bar{x}, \bar{y})$ as $k \in \bar{\mathcal{K}}$ and $k \to \infty$, from the continuity of F_0 it also holds that $F_0(x^{(k+1)}, y^{(k+1)}) \to F_0(\bar{x}, \bar{y})$ as $k \in \bar{\mathcal{K}}$ and $k \to \infty$. ■

In the next lemma we show that the unique optimal solution of the strictly convex problem (39) with parameters $(\xi, \sigma, \rho) = (x^*, \sigma^*, \rho^*)$ is given by the limit point $(\bar{x}, \bar{y})$.

Lemma 10. *$(\bar{x}, \bar{y})$ is the unique optimal solution of the problem PSUB(x^*, σ^*, ρ^*).*

Proof: Since $x^{(k+1)} \in X(x^{(k)}, \sigma^{(k)})$ and $G_i(x^{(k+1)}, y^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) \leq 0$, it follows, by letting $k \in \bar{\mathcal{K}}$ and $k \to \infty$, that $\bar{x} \in X(x^*, \sigma^*)$ and $G_i(\bar{x}, \bar{y}, x^*, \sigma^*, \rho^*) \leq 0$ for $i \geq 1$. Thus, $(\bar{x}, \bar{y})$ is a feasible solution of PSUB(x^*, σ^*, ρ^*). Let $(\bar{\bar{x}}, \bar{\bar{y}})$ be an arbitrary feasible solution of PSUB(x^*, σ^*, ρ^*), so that $\bar{\bar{x}} \in X(x^*, \sigma^*)$ and $G_i(\bar{\bar{x}}, \bar{\bar{y}}, x^*, \sigma^*, \rho^*) \leq 0$ for $i \geq 1$. We must show that $G_0(\bar{x}, \bar{y}, x^*, \sigma^*, \rho^*) \leq G_0(\bar{\bar{x}}, \bar{\bar{y}}, x^*, \sigma^*, \rho^*)$.

For $\tau = 1, 2, 3, \ldots$, let

$$\bar{\bar{x}}^{(\tau)} = \bar{\bar{x}} + \alpha^{(\tau)}(x^* - \bar{\bar{x}}), \quad (45a)$$

$$\bar{\bar{y}}^{(\tau)} = \bar{\bar{y}} + \frac{1}{\tau}(1, \ldots, 1)^T. \quad (45b)$$

If $\alpha^{(\tau)} = 0$ then, for $i \geq 1$,

$$\begin{aligned} G_i(\bar{\bar{x}}^{(\tau)}, \bar{\bar{y}}^{(\tau)}, x^*, \sigma^*, \rho^*) &= v_i(\bar{\bar{x}}^{(\tau)}, x^*, \sigma^*) + \rho^* w(\bar{\bar{x}}^{(\tau)}, x^*, \sigma^*) - \bar{\bar{y}}_i^{(\tau)} \\ &= G_i(\bar{\bar{x}}, \bar{\bar{y}}, x^*, \sigma^*, \rho^*) - \tfrac{1}{\tau} \leq -\tfrac{1}{\tau}. \end{aligned}$$

It is therefore possible to choose the scalar $\alpha^{(\tau)}$ such that $0 < \alpha^{(\tau)} < 1/\tau$ and

$$G_i(\bar{\bar{x}}^{(\tau)}, \bar{\bar{y}}^{(\tau)}, x^*, \sigma^*, \rho^*) \leq -\frac{1}{2\tau}$$

for $i \geq 1$. Then $(\bar{\bar{x}}^{(\tau)}, \bar{\bar{y}}^{(\tau)})$ is in the interior of the feasible set of PSUB(x^*, σ^*, ρ^*). In particular, $\bar{\bar{x}}^{(\tau)}$ is in the interior of $X(x^*, \sigma^*)$. This implies that for each τ, there exists an integer $K(\tau)$ such that, for all $k \in \bar{\mathcal{K}}$ with $k > K(\tau)$, $\bar{\bar{x}}^{(\tau)} \in X(x^{(k)}, \sigma^{(k)})$ and $G_i(\bar{\bar{x}}^{(\tau)}, \bar{\bar{y}}^{(\tau)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) \leq 0$ for $i \geq 1$. For all these $k \in \bar{\mathcal{K}}$ with $k > K(\tau)$ it then holds that $G_0(\bar{\bar{x}}^{(\tau)}, \bar{\bar{y}}^{(\tau)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) \geq G_0(x^{(k+1)}, y^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)})$, because $(x^{(k+1)}, y^{(k+1)})$ is the optimal solution of PSUB$(x^{(k)}, \sigma^{(k)}, \rho^{(k)})$.

Now, for each τ, let the integer $k(\tau) \in \bar{\mathcal{K}}$ satisfy $k(\tau) > \max\{\tau, K(\tau)\}$ and let $\tau \to \infty$. Then, by the construction (45), we have $(\bar{\bar{x}}^{(\tau)}, \bar{\bar{y}}^{(\tau)}) \to (\bar{\bar{x}}, \bar{\bar{y}})$ and by the taken subsequence $(x^{(k(\tau)+1)}, y^{(k(\tau)+1)}) \to (\bar{x}, \bar{y})$ and $(x^{(k(\tau))}, \sigma^{(k(\tau))}, \rho^{(k(\tau))}) \to (x^*, \sigma^*, \rho^*)$. Thus, $G_0(\bar{x}, \bar{y}, x^*, \sigma^*, \rho^*) \leq G_0(\bar{\bar{x}}, \bar{\bar{y}}, x^*, \sigma^*, \rho^*)$. ■

The lemma that comes next relates the limit points of the subsequences $\{(x^{(k)}, y^{(k)})\}_{k \in \mathcal{K}}$ and $\{(x^{(k+1)}, y^{(k+1)})\}_{k \in \bar{\mathcal{K}}}$, with $\bar{\mathcal{K}} \subset \mathcal{K}$.

Lemma 11. $(\bar{x}, \bar{y}) = (x^*, y^*)$.

Proof: From $G_i(x^{(k)}, y^{(k)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) = F_i(x^{(k)}, y^{(k)}) \leq \bar{\mu}_{k-1,i}$ for $i \geq 1$ and $k \geq 2$, it follows that $G_i(x^*, y^*, x^*, \sigma^*, \rho^*) \leq 0$ for $i \geq 1$, by letting $k \in \bar{\mathcal{K}}$ and $k \to \infty$. Further, by definition, $x^* \in X(x^*, \sigma^*)$. Thus, (x^*, y^*) is a feasible solution of PSUB(x^*, σ^*, ρ^*).

From $F_0(x^{(k+1)}, y^{(k+1)}) \leq G_0(x^{(k+1)}, y^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) + \bar{\mu}_{k,0}$, it follows, again by letting $k \in \bar{\mathcal{K}}$ and $k \to \infty$, that $F_0(\bar{x}, \bar{y}) \leq G_0(\bar{x}, \bar{y}, x^*, \sigma^*, \rho^*)$.

By definition, $F_0(x^*, y^*) = G_0(x^*, y^*, x^*, \sigma^*, \rho^*)$. From Lemma 9 it then follows that $G_0(x^*, y^*, x^*, \sigma^*, \rho^*) \leq G_0(\bar{x}, \bar{y}, x^*, \sigma^*, \rho^*)$. But since $(\bar{x}, \bar{y})$ is the unique global optimal solution of PSUB(x^*, σ^*, ρ^*), it then follows that $(x^*, y^*) = (\bar{x}, \bar{y})$. ■

By combining the previous results, the next lemma characterizes the stationarity of the limit point generated by Algorithm 1.

Lemma 12. *(x^*, y^*) is a KKT point of the problem (2).*

Proof: It follows from Lemmas 3, 4, 10 and 11. ■

Finally, in the following we prove the global convergence result of Algorithm 1.

Proof of Theorem 1: Assume that the statement in Theorem 1 is false. Then there is an $\epsilon > 0$ and an infinite subset $\mathcal{K}_0$ of the integers such that

$$||(x, y) - (x^{(k)}, y^{(k)})|| \geq \epsilon \text{ for all } (x, y) \in \Omega \text{ and every } k \in \mathcal{K}_0. \quad (46)$$

Then, as a consequence of Lemma 6, the sequence $\{(x^{(k)}, y^{(k)})\}_{k\in\mathcal{K}_0}$ has at least one convergent subsequence. Thus, there is a point $(\hat{x}, \hat{y})$ and an infinite subset $\bar{\mathcal{K}_0} \subseteq \mathcal{K}_0$ such that $(x^{(k)}, y^{(k)}) \to (\hat{x}, \hat{y})$ as $k \in \bar{\mathcal{K}_0}$ and $k \to \infty$.

But then, by letting $(\hat{x}, \hat{y})$ play the role of (x^*, y^*) in the above lemmas, in particular from Lemma 12, it follows that $(\hat{x}, \hat{y})$ is a KKT point of the problem (38) and thus it is also a KKT point of the original problem (2). As a result, $(\hat{x}, \hat{y}) \in \Omega$. By letting $(x, y) = (\hat{x}, \hat{y})$ in (46), a contradiction has been established. Therefore, the statement in Theorem 1 cannot be false, it must be true. ■

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.

5 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 [14]. In the primal-dual interior-point approach, a sequence of relaxed KKT conditions are solved by Newton's method [16]. 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.

5.1 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:

$$\mathcal{L}(x,y,\lambda) = g_0(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{m} \left(c_i y_i + \tfrac{1}{2} d_i y_i^2\right) - \sum_{i=1}^{m} \lambda_i y_i$$

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

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

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

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:

$$\begin{aligned} \mathcal{W}(\lambda) &= \min_{x,y} \{\mathcal{L}(x,y,\lambda);\ \alpha_j \leq x_j \leq \beta_j,\ y_i \geq 0,\ \forall i,\ \forall j\} \qquad (47) \\ &= r_0 + \lambda^T r + \sum_{j=1}^{n} \widehat{\mathcal{W}}_j(\lambda) + \sum_{i=1}^{m} \widetilde{\mathcal{W}}_i(\lambda) \end{aligned}$$

where

$$\widehat{\mathcal{W}}_j(\lambda) = \min_{x_j} \{\mathcal{L}_j(x_j,\lambda);\ \alpha_j \leq x_j \leq \beta_j\},\ j = 1, \ldots, n, \qquad (48)$$

$$\widetilde{\mathcal{W}}_i(\lambda) = \min_{y_i} \{c_i y_i + \tfrac{1}{2} d_i y_i^2 - \lambda_i y_i;\ y_i \geq 0\},\ i = 1, \ldots, m. \qquad (49)$$

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 (47) can be split into the $n+m$ minimization problems (48) and (49). The use of the minimum instead of the infimum in expressions (47)-(49) is justified by the existence of the minimizers of (48) and (49). The expressions of these minimizers, which we denote by $x_j(\lambda)$ and $y_i(\lambda_i)$, respectively, are:

$$x_j(\lambda) = \max\left\{\alpha_j, \min\left\{\beta_j, x_j^*(\lambda)\right\}\right\}, \quad \text{for } j = 1, \ldots, n, \qquad (50)$$

where

$$x_j^*(\lambda) = \frac{(p_{0j} + \lambda^T p_j)^{1/2} l_j + (q_{0j} + \lambda^T q_j)^{1/2} u_j}{(p_{0j} + \lambda^T p_j)^{1/2} + (q_{0j} + \lambda^T q_j)^{1/2}} \qquad (51)$$

and

$$y_i(\lambda_i) = \max\left\{0, \frac{\lambda_i - c_i}{d_i}\right\}, \quad \text{for } i = 1, \ldots, m. \qquad (52)$$

Note that $x_j : I\!R^m \to I\!R$ and $y_i : I\!R \to I\!R$ 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 (48) and $y_i(\lambda_i)$ of (49), there is also an explicit expression for the dual objective function (47), which is:

$$\mathcal{W}(\lambda) = r_0 \; + \; \lambda^T r \; + \; \sum_{j=1}^{n} \left(\frac{p_{0j} + \lambda^T p_j}{u_j - x_j(\lambda)} + \frac{q_{0j} + \lambda^T q_j}{x_j(\lambda) - l_j} \right)$$

$$+ \sum_{i=1}^{m} \left(c_i y_i(\lambda_i) \; + \; \tfrac{1}{2} \, d_i \, y_i^2(\lambda_i) \; - \; \lambda_i \, y_i(\lambda_i) \right).$$

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

$$\begin{array}{rl} \text{maximize} & \mathcal{W}(\lambda) \\ \text{subject to} & \lambda \geq 0. \end{array} \tag{53}$$

Once the dual problem (53) 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 $y_i(\lambda_i)$.

5.2 Properties of the dual function

Before proposing our approach to solve the dual problem (53) corresponding to the MMA subproblem (3), we comment on some properties associated with the dual function $\mathcal{W}$.

First, note that the function $\mathcal{W} : I\!R^m \to I\!R$ 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.,

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

for all $i = 1, \ldots, m$ and $\lambda \in I\!R^m$, as stated in Proposition 6.1.1 of [3]. Note that, although y does not belong to a compact set, the existence of the minimizer (52) 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 refereing 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 (50), 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 (51).

5.3 Trust-region method

In this subsection we present a strategy to solve the dual subproblem 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:

$$\begin{array}{ll} \text{minimize} & W(\lambda) \\ \text{subject to} & \lambda \geq 0, \end{array} \tag{54}$$

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:

$$m_{\bar{k}}(\lambda) = W(\lambda^{(\bar{k})}) + \nabla W(\lambda^{(\bar{k})})^T (\lambda - \lambda^{(\bar{k})}) + \tfrac{1}{2} \eta^{(\bar{k})} \|\lambda - \lambda^{(\bar{k})}\|_2^2, \tag{55}$$

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

$$\eta^{(\bar{k})} = \frac{(s^{(\bar{k})})^T t^{(\bar{k})}}{(s^{(\bar{k})})^T s^{(\bar{k})}} \tag{56}$$

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 [10]. This interpretation justifies the second-order term of the model, since the Hessian matrix $\nabla^2 W$ is discontinuous. Furthermore, models similar to (55) 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 [19] 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 (54) 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

$$\begin{aligned} \text{minimize} \quad & m_{\bar{k}}(\lambda) \\ \text{subject to} \quad & \underline{\lambda}^{(\bar{k})} \le \lambda \le \overline{\lambda}^{(\bar{k})}, \end{aligned} \tag{57}$$

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 $\hat{\lambda} \in I\!R^m$ of problem (57) is given by the closed form

$$\hat{\lambda} = \min\left\{\overline{\lambda}^{(\bar{k})}, \max\left\{\underline{\lambda}^{(\bar{k})}, \lambda^{(\bar{k})} - \frac{1}{\eta^{(\bar{k})}} \nabla W(\lambda^{(\bar{k})})\right\}\right\}. \tag{58}$$

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

Algorithm 2: A trust-region approach applied to the dual of the MMA subproblem

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

1. Compute $\eta^{(\bar{k})}$ using (56) and $\hat{\lambda}$ as in (58).

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

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

Concerning the implementation of this model algorithm, some comments are in order:

1) The expression defined by $\theta_{\bar{k}}$ in Step 2 measures the agreement between the model function $m_{\bar{k}}$ and the objective function W. More precisely, it is the ratio between the actual reduction of the function and the reduction predicted by the model.

2) 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 $\varepsilon = 10^{-3}$.

3) To obtain $\hat{\lambda}$ we must take a step, from $\lambda^{(\bar{k})}$, of length $1/\eta^{(\bar{k})}$ in the direction $-\nabla W_{\bar{k}}$. It is true that the value of the spectral parameter $\eta^{(\bar{k})}$ will never be negative, since the function W is convex (we have already seen that $\mathcal{W}$ is concave). However, to avoid very small (positive) or too large values for $\eta^{(\bar{k})}$, we project it in the interval (η^{min}, η^{max}), where $0 < \eta^{min} < \eta^{max} < +\infty$. The adopted values were $\eta^{min} = 10^{-3}$ and $\eta^{max} = 10^{3}$.

4) In our algorithm, whenever the current estimate changes, we take $W_{\bar{k}+1} = W(\hat{\lambda})$, but it is necessary to evaluate the gradient at the new point.

5) In the numerical implementation we have used $\lambda^{(1)} = (0, \ldots, 0)^T$ and the initial trust-region radius was set as $\Delta^{(1)} = 0.1\|\nabla W(\lambda^{(1)})\|$. The trust-region updating rule presented in the Algorithm 2 is based on Sections 3.2.4 and 3.3.4 of [5], as well as the choices of the parameters γ_0, γ_2, υ and ω, among others that have not appeared here. All these choices also fit the recommendations of [6].

Despite not being our primary motivation, it is worth mentioning that (58) coincides with the first Spectral Projected Gradient (SPG) trial point (cf. [3]) for problem (54) within the bound constraints of (57). 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 $m_{\bar{k}}(\lambda^{(\bar{k})}) > m_{\bar{k}}(\hat{\lambda})$.

6 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 [17]. 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

Academic Problem 1:

$$\begin{aligned} \text{minimize} \quad & f_0(x) = x^T S x \\ \text{s.t.} \quad & f_1(x) = \frac{n}{2} - x^T P x \leq 0, \\ & f_2(x) = \frac{n}{2} - x^T Q x \leq 0, \\ & -1 \leq x_j \leq 1, \; j = 1, \ldots, n. \end{aligned} \tag{59}$$

Academic Problem 2:

$$\begin{aligned} \text{minimize} \quad & f_0(x) = -x^T S x \\ \text{s.t.} \quad & f_1(x) = x^T P x - \frac{n}{2} \leq 0, \\ & f_2(x) = x^T Q x - \frac{n}{2} \leq 0, \\ & -1 \leq x_j \leq 1, \; j = 1, \ldots, n. \end{aligned} \tag{60}$$

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 $\alpha_{ij} = (i + j - 2)/(2n - 2) \in [0, 1]$ for all i and j. The feasible starting points for Problems 1 and 2 are $x^{(1)} = (0.5, \ldots, 0.5)^T \in I\!R^n$ and $x^{(1)} = (0.25, \ldots, 0.25)^T \in I\!R^n$, 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 $a^- = \max\{0, -a\}$:

$$
\begin{array}{rcll}
(1 + x_j)\left(\dfrac{\partial f_0}{\partial x_j}(x) + \lambda_1 \dfrac{\partial f_1}{\partial x_j}(x) + \lambda_2 \dfrac{\partial f_2}{\partial x_j}(x)\right)^+ & = & 0, & j = 1, \ldots, n, \\
(1 - x_j)\left(\dfrac{\partial f_0}{\partial x_j}(x) + \lambda_1 \dfrac{\partial f_1}{\partial x_j}(x) + \lambda_2 \dfrac{\partial f_2}{\partial x_j}(x)\right)^- & = & 0, & j = 1, \ldots, n, \\
f_i(x)^+ & = & 0, & i = 1, 2, \\
\lambda_i f_i(x)^- & = & 0, & i = 1, 2, \\
\lambda_i & \geq & 0, & i = 1, 2, \\
-1 \;\leq\; x_j & \leq & 1, & j = 1, \ldots, n.
\end{array}
$$

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

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

The sequence $\{\mu_k\}_{k=1}^{\infty}$ used in (28) to relax the conservative condition was chosen as follows:

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

with

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

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 (29).

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 (28) 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.

The numerical results corresponding to the Academic Problems 1 and 2 are given in Tables 1 and 2, respectively. In each entry of the table we report:

Outer iterations (Additional inner iterations); CPU time.

The total number of solved subproblems is the sum of the outer and the additional inner iterations. The reported CPU time is the average of ten runs of the algorithm. Such results are schematically depicted in Figures 2 and 3 for Problem 1 and in Figures 4 and 5 for Problem 2.

Analyzing these results, 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.

n	Svanberg's PD	Strategy 1 PD	Strategy 2 PD	Strategy 3 PD
100	104(135); 4.1267	108(101); 3.8254	132(59); 3.8818	97(11); 2.1315
500	147(185); 49.666	153(138); 45.977	158(36); 36.454	115(0); 23.287
1000	174(222); 214.74	179(162); 197.36	223(40); 184.98	128(0); 96.354
2000	185(229); 868.98	189(185); 819.99	368(82); 1206.3	138(0); 404.32

n	Dual TR	Strategy 1 TR	Strategy 2 TR	Strategy 3 TR
100	106(134); 2.0159	103(96); 1.7122	121(63); 1.9412	99(9); 1.1947
500	151(184); 45.821	156(147); 42.304	150(39); 32.617	105(0); 19.804
1000	177(214); 204.46	180(161); 186.98	223(28); 172.34	124(0); 90.175
2000	186(232); 857.63	190(186); 802.26	274(60); 866.71	123(0); 353.22

Table 1: Numerical results for Problem 1.

n	Svanberg's PD	Strategy 1 PD	Strategy 2 PD	Strategy 3 PD
100	218(265); 11.553	222(198); 12.949	189(158); 6.9761	199(60); 6.3205
500	392(415); 130.99	392(317); 121.86	353(280); 105.12	357(97); 86.954
1000	438(437); 566.14	443(337); 475.37	416(350); 459.24	418(142); 378.66
2000	479(503); 2208.1	477(379); 2084.9	442(423); 1947.3	452(185); 1678.3

n	Dual TR	Strategy 1 TR	Strategy 2 TR	Strategy 3 TR
100	223(268); 3.8960	224(204); 3.4818	189(154); 2.9401	201(89); 2.6735
500	389(430); 109.26	390(339); 100.49	355(284); 90.385	353(123); 74.973
1000	441(434); 469.01	445(379); 450.50	417(351); 423.04	410(153); 346.09
2000	481(505); 2081.4	487(450); 2008.4	445(422); 1861.9	443(241); 1603.6

Table 2: Numerical results for Problem 2.

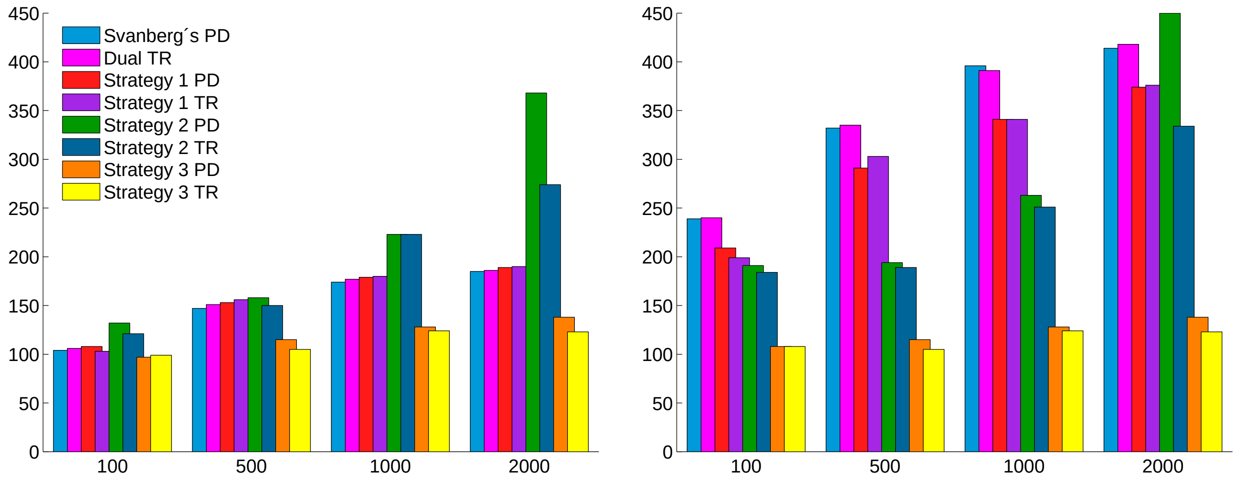

Figure 2: Dimension × Outer iterations (left) and Dimension × Inner iterations (right), for Problem 1.

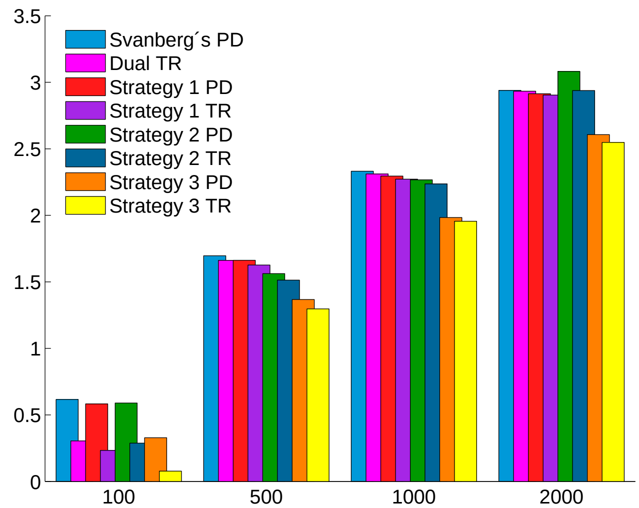

Figure 3: Dimension × $\log_{10}$(CPU time), for Problem 1.

In relative terms, denoting respectively by sub_S and sub_M the total number of subproblems solved using Svanberg's algorithm and using the modifications of Strategy J PD (J $= 1, 2, 3$), Dual TR, and Strategy J TR (J $= 1, 2, 3$), we have computed the ratios

$$\text{sub}_\text{M}/\text{sub}_\text{S}. \tag{61}$$

Similarly, using cpu_S and cpu_M to denote, respectively, the total CPU time demanded by Svanberg's algorithm and by the one with the modifications of Strategy J PD (J $= 1, 2, 3$), Dual TR, and Strategy J TR (J $= 1, 2, 3$), we obtain

$$\text{cpu}_\text{M}/\text{cpu}_\text{S}. \tag{62}$$

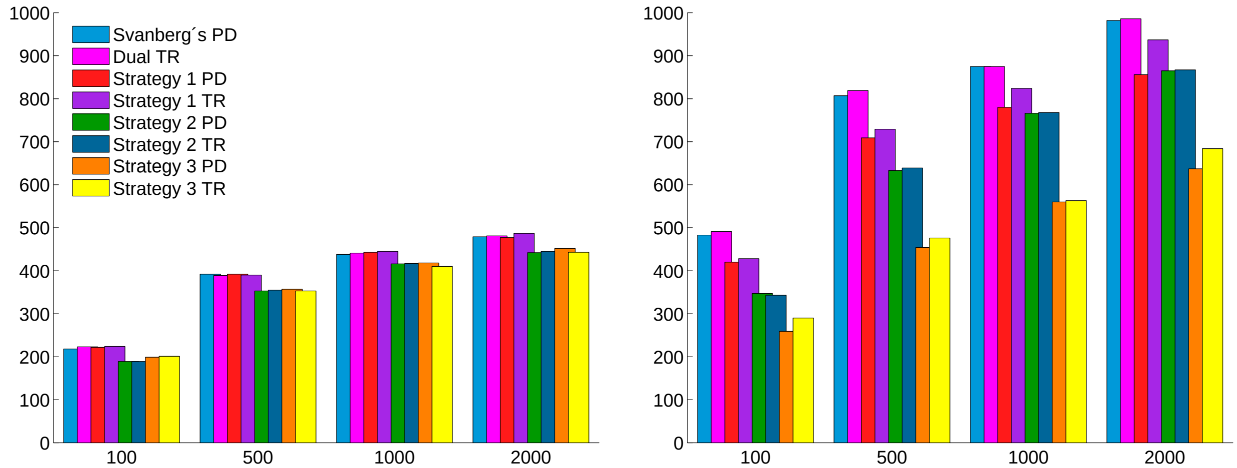

Figure 4: Dimension × Outer iterations (left) and Dimension × Inner iterations (right), for Problem 2.

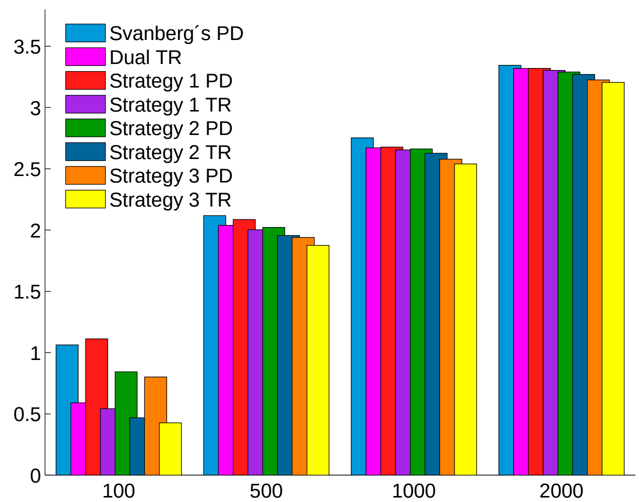

Figure 5: Dimension × $\log_{10}$(CPU time), for Problem 2.

The relative measures (61) and (62) were calculated with the values provided in Tables 1 and 2, and the results are presented in Tables 3 and 4. The obtained percentages put each strategy in perspective with respect to the performance of Svanberg's original algorithm, both in terms of the number of solved subproblems, as well as the CPU time demanded.

The results present in Tables 5 and 6 were produced by randomly generating ten initial points within the simple bounds of the problems, for each of the dimensions under consideration. Thus, forty tests for Problem 1 and forty tests for Problem 2 were solved. Each strategy was used for solving these tests, and the three values between round brackets are the minimum, average and maximum number of solved subproblems, respectively, followed by the average of the CPU time demanded, given in seconds. The results corroborate the previous ones. It is important to note that for each test, including the aforementioned, with

	Strat. 1 PD		Strat. 2 PD		Strat. 3 PD		Dual TR		Strat. 1 TR		Strat. 2 TR		Strat. 3 TR	
n	Sub	CPU	Sub	CPU	Sub	CPU	Sub	CPU	Sub	CPU	Sub	CPU	Sub	CPU
100	87%	93%	80%	94%	45%	52%	100%	49%	83%	41%	77%	47%	45%	29%
500	88%	93%	58%	76%	35%	47%	101%	92%	91%	85%	57%	66%	32%	40%
1000	86%	92%	66%	86%	32%	45%	99%	95%	86%	87%	63%	80%	31%	42%
2000	90%	94%	109%	139%	33%	47%	101%	99%	91%	92%	81%	100%	30%	41%

Table 3: Relative measures for Problem 1.

	Strat. 1 PD		Strat. 2 PD		Strat. 3 PD		Dual TR		Strat. 1 TR		Strat. 2 TR		Strat. 3 TR	
n	Sub	CPU	Sub	CPU	Sub	CPU	Sub	CPU	Sub	CPU	Sub	CPU	Sub	CPU
100	87%	112%	72%	60%	54%	55%	102%	34%	89%	30%	71%	25%	60%	23%
500	88%	93%	78%	80%	56%	66%	101%	83%	90%	77%	79%	69%	59%	57%
1000	89%	84%	88%	81%	64%	67%	100%	83%	94%	80%	88%	75%	64%	61%
2000	87%	94%	88%	88%	65%	76%	100%	94%	95%	91%	88%	84%	70%	73%

Table 4: Relative measures for Problem 2.

the initial point from the literature, all the strategies achieved the same optimal solution.

In Figures 6 and 7 we depict the performance profiles [7] of the results corresponding to the generated tests: forty for Problem 1 and forty for Problem 2. For each figure, the graph on the left is concerned with the number of solved subproblems, whereas the graph on the right, with the CPU time spent. From Figure 6, we notice that Strategies 3 PD and 3 TR are the most efficient. We can also observe that both approaches for solving the subproblems are competitive, when compared pairwise with each of the three strategies. From Figure 7 it

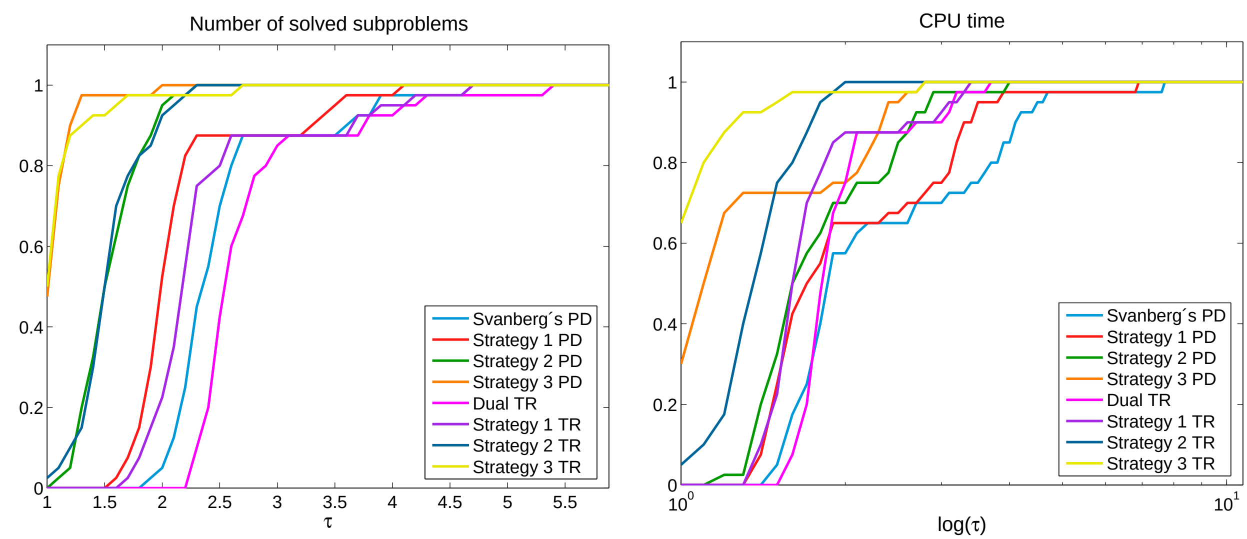

Figure 6: Performance profile for Problem 1. Vertical axis: $\rho_s(\tau)$

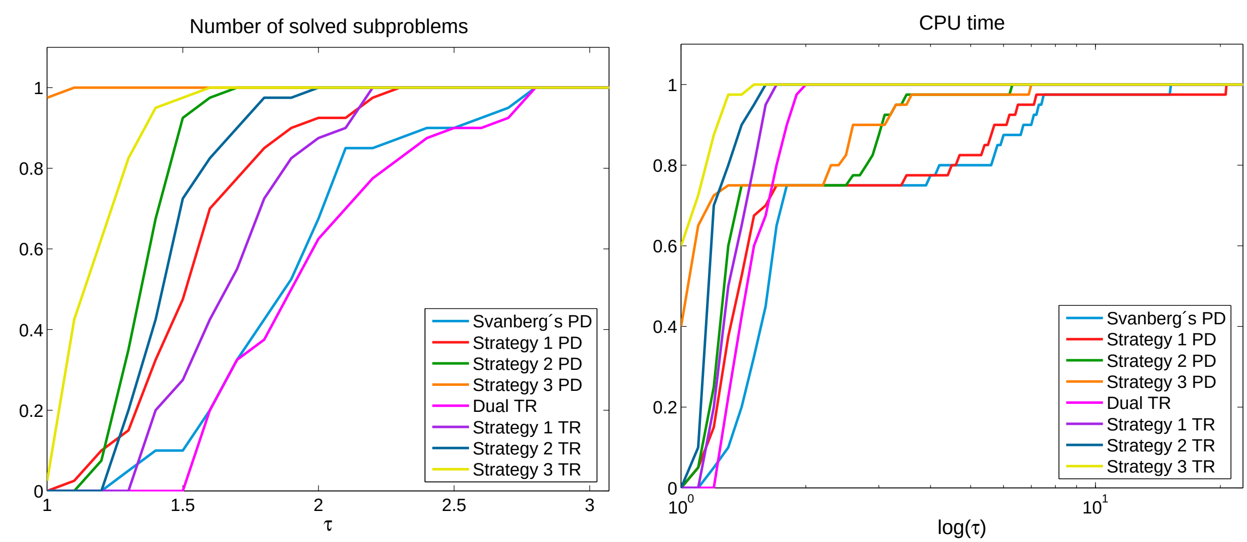

Figure 7: Performance profile for Problem 2. Vertical axis: $\rho_s(\tau)$

n	Svanberg's PD	Strategy 1 PD	Strategy 2 PD	Strategy 3 PD
100	(231, 279.2, 413); 5.1826	(161, 228.6, 368); 4.3418	(123, 147.9, 183); 3.3030	(111, 130.1, 161); 2.8220
500	(326, 446.5, 1226); 67.241	(274, 382.4, 1053); 60.529	(192, 232.2, 370); 44.728	(145, 174.9, 274); 36.003
1000	(370, 507, 1422); 279.46	(316, 445.6, 1305); 255.54	(237, 332.1, 736); 226.88	(165, 231.9, 717); 171.81
2000	(437, 583.8, 1271); 1233.5	(372, 511, 1116); 1124.1	(336, 387.5, 478); 1058.2	(188, 222.7, 346); 653.87

n	Dual TR	Strategy 1 TR	Strategy 2 TR	Strategy 3 TR
100	(235, 293.4, 434); 2.3654	(172, 232.6, 397); 1.9872	(125, 158.9, 215); 1.6097	(96, 113.1, 139); 1.2376
500	(364, 475.2, 1113); 63.320	(278, 412.6, 1081); 57.844	(179, 216.8, 317); 37.463	(142, 171.1, 263); 31.956
1000	(388, 524.8, 1413); 275.56	(348, 479.4, 1423); 259.92	(246, 291.1, 372); 191.83	(168, 265.4, 987); 186.42
2000	(456, 701.2, 1296); 1445.0	(397, 620.2, 1250); 1324.6	(347, 407.2, 551); 1060.4	(193, 272.9, 450); 782.24

Table 5: Further results for Problem 1.

n	Svanberg's PD	Strategy 1 PD	Strategy 2 PD	Strategy 3 PD
100	(309, 426.2, 889); 15.254	(260, 360.6, 837); 16.674	(175, 271.9, 630); 7.0938	(120, 199.7, 457); 7.1492
500	(486, 648.4, 892); 103.42	(372, 535.3, 779); 90.732	(327, 481, 736); 81.545	(235, 355.1, 578); 68.591
1000	(634, 950, 2109); 557.76	(513, 792.2, 1862); 498.71	(483, 743.5, 1470); 455.17	(316, 567.3, 1098); 389.21
2000	(716, 880.8, 1045); 1918.1	(613, 731.1, 899); 1703.4	(564, 712.9, 884); 1649.3	(407, 539.3, 744); 1396.0

n	Dual TR	Strategy 1 TR	Strategy 2 TR	Strategy 3 TR
100	(332, 438.8, 884); 3.2094	(255, 370.1, 835); 2.8255	(175, 273.7, 652); 2.2557	(131, 214.6, 487); 1.9409
500	(506, 651.4, 882); 87.161	(410, 566.5, 788); 79.360	(337, 492.7, 749); 71.515	(266, 402, 629); 63.489
1000	(634, 1021.6, 2155); 537.43	(560, 879.2, 1911); 484.79	(503, 822, 1509); 459.16	(402, 682.4, 1312); 409.51
2000	(690, 941.8, 1196); 1984.1	(617, 828.6, 1016); 1822.4	(578, 831.7, 1425); 1828.0	(476, 683.1, 931); 1618.7

Table 6: Further results for Problem 2.

is evident that Svanberg's PD is more efficient in terms of the number of solved subproblems. Nevertheless, when it comes to the CPU time spent, the strategies that rely upon the dual trust-region approach were more efficient. Figure 8 contains the results of both Problems 1 and 2.

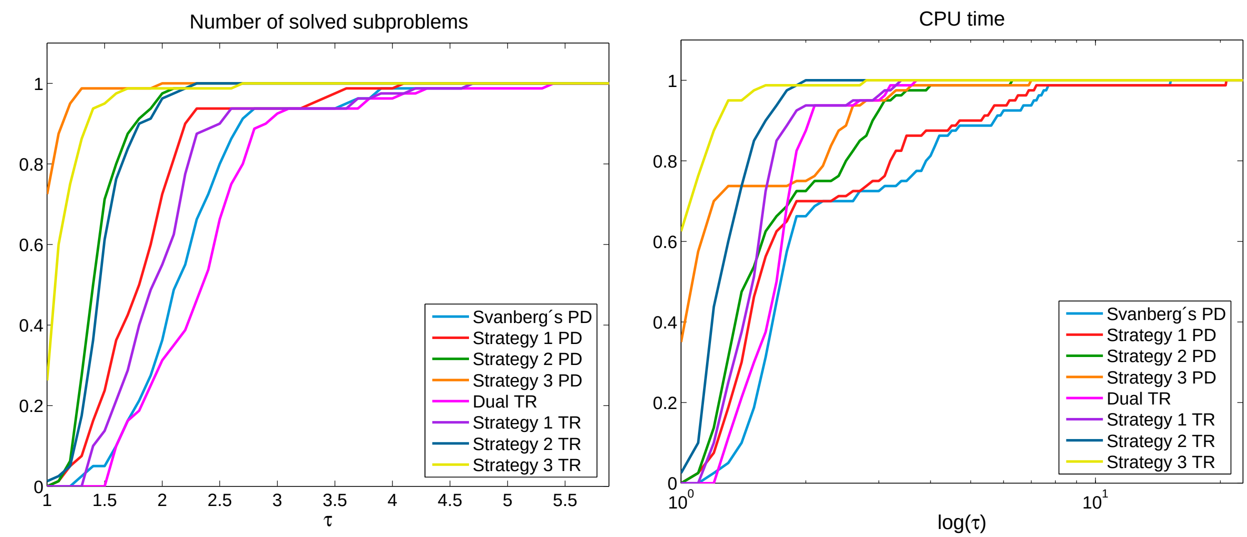

Figure 8: Performance profile for Problems 1 and 2. Vertical axis: $\rho_s(\tau)$

7 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 [14], and replaced by the primal-dual approach [16]. 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. 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.

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 (Strategy 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.

Acknowledgements. The authors are very grateful for the peer review of the anonymous referee of the companion paper, which helped to improve the presentation of this manuscript as well.

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] K. J. Arrow, L. Hurwicz and H. Uzawa *Constraint qualifications in maximization problems*, Naval Research Logistics Quarterly, **8** (1961), 175–191.

[3] 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, pp. 3652–3659.

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

[5] 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.

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

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

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

[9] 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.

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

[11] 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.

[12] 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.

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

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

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

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

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

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

[19] 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.

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

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