

## A CLASS OF GLOBALLY CONVERGENT OPTIMIZATION METHODS BASED ON CONSERVATIVE CONVEX SEPARABLE APPROXIMATIONS\*

KRISTER SVANBERG<sup>†</sup>

**Abstract.** This paper deals with a certain class of optimization methods, based on conservative convex separable approximations (CCSA), for solving inequality-constrained nonlinear programming problems. Each generated iteration point is a feasible solution with lower objective value than the previous one, and it is proved that the sequence of iteration points converges toward the set of Karush–Kuhn–Tucker points. A major advantage of CCSA methods is that they can be applied to problems with a very large number of variables (say  $10^4$ – $10^5$ ) even if the Hessian matrices of the objective and constraint functions are dense.

**Key words.** nonlinear programming, constrained minimization, convex approximations, method of moving asymptotes,

**AMS subject classifications.** 49M37, 65K05, 90C30

**PII.** S1052623499362822

**1. Introduction.** The purpose of this paper is to present and investigate a new class of optimization methods which we call *conservative convex separable approximation* (CCSA) *methods*. These methods are intended for inequality-constrained nonlinear programming problems, which are assumed to be written as *minimization* problems with *less than or equal to* constraints. There are *outer* and *inner* iterations in the methods. An *outer* iteration starts from the current iterate  $x^{(k)}$  and ends up with a new iterate  $x^{(k+1)}$ . In each *inner* iteration, within a given outer iteration, a convex subproblem is generated and solved. In this subproblem, the original objective and constraint functions are replaced by certain convex separable functions which approximate the original functions around  $x^{(k)}$ . The optimal solution of the subproblem is either accepted or rejected. If accepted, it becomes  $x^{(k+1)}$  and the outer iteration is completed. If rejected, a new inner iteration is made, with a modified subproblem based on somewhat modified approximating functions. These inner iterations are repeated until the approximating objective and constraint functions become greater than or equal to the original functions at the optimal solution of the subproblem. When this happens, we say that the approximating functions are *conservative*. This does not imply that the feasible set of the subproblem is completely contained in the original feasible set, but it does imply that the optimal solution of the subproblem is a feasible solution of the original problem, with lower objective value than the previous iterate. Each new outer iteration requires function values and first order derivatives of the original objective and constraint functions, calculated at the current iterate  $x^{(k)}$ . Each new inner iteration requires function values, but no derivatives, calculated at the optimal solution of the most recent subproblem.

To use an approach based on solving a sequence of convex subproblems is not a new idea. It is used also in, e.g., *sequential quadratic programming* (SQP) where, at each iteration, a convex quadratic programming (QP) problem is solved and a

---

\*Received by the editors October 20, 1999; accepted for publication (in revised form) December 5, 2000; published electronically January 4, 2002. This research was supported by the Swedish Research Council for the Engineering Sciences (TFR).

<http://www.siam.org/journals/siopt/12-2/36282.html>

<sup>†</sup>Optimization and Systems Theory, KTH, SE-10044 Stockholm, Sweden (krille@math.kth.se).

linesearch on a merit function is performed; see, e.g., [5] and [2]. However, the (linear) constraints in the QP subproblems do not in general force the iteration points to be feasible with respect to the original constraints, and thus they are not conservative in the above meaning. In contrast to SQP methods, CCSA methods introduce curvature both in the objective function and in the constraint functions of the subproblem. This curvature is updated in the inner iterations until the approximating functions become conservative, and then there is no need for any linesearch. Another class of methods which generate feasible iteration points is *interior point methods*, see, e.g., [1], [3], and [4]. But in these methods feasibility is typically preserved by adding a logarithmic barrier function to the objective function, and *not* by using conservative approximations of the constraint functions as in CCSA methods.

It should be emphasized that a major benefit of CCSA methods is that they can be successfully applied to problems with a very large number of variables, even if the Hessian matrices of the objective and constraint functions are dense. This property is to a large extent due to the usage of separable approximations.

One of the CCSA methods presented here, namely the *method of moving asymptotes* (MMA), has a background in the structural optimization field, where function and gradient evaluations are very time-consuming (involving huge finite element calculations), and where the users often consider it important that the generated iteration points are feasible. The original version of MMA, presented in [7], usually worked quite well in practice but was not globally convergent and sometimes failed on certain problems. A later version, presented in [8], was globally convergent but turned out to be too slow in practice. The version of MMA presented in this paper apparently outperforms both of these earlier versions, in theory as well as in practice. Moreover, MMA is now just one of several alternative methods within the concept of CCSA methods, which is introduced here and for which a convergent proof has not appeared before.

The paper is organized as follows. In section 2, a convenient formulation of inequality-constrained optimization problems is suggested and shown to have some important properties. In particular, the set of Karush–Kuhn–Tucker (KKT) points is nonempty. In sections 3 and 4, a general description of CCSA methods is given, and then some specific CCSA methods are described in sections 5 and 6. In section 7, it is proved that CCSA methods are globally convergent in the following sense: From any starting point, the sequence of generated iteration points converges towards the set of KKT points. In section 8, finally, numerical results on some large scale problems are presented.

**2. Considered problem and some basic properties.** Inequality-constrained nonlinear programming problems are often written in the following form, where  $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$  is the vector of variables,  $x_j^{\min}$  and  $x_j^{\max}$  are given real numbers, and  $f_0, f_1, \dots, f_m$  are given, typically twice continuously differentiable, real-valued functions:

$$(2.1) \quad \begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & && x_j^{\min} \leq x_j \leq x_j^{\max}, \quad j = 1, \dots, n. \end{aligned}$$

In this paper, however, any problem of this type is transformed into a closely related problem of the following extended form where, in addition to the variables  $x \in \mathbb{R}^n$ ,

there also appear “artificial” variables  $y = (y_1, \dots, y_m)^T \in \mathbb{R}^m$ :

$$\begin{aligned}
 (2.2) \quad & \text{minimize} && f_0(x) + \sum_{i=1}^m c_i y_i \\
 & \text{subject to} && f_i(x) - y_i \leq 0, \quad i = 1, \dots, m, \\
 & && x_j^{\min} \leq x_j \leq x_j^{\max}, \quad j = 1, \dots, n, \\
 & && y_i \geq 0, \quad i = 1, \dots, m.
 \end{aligned}$$

If the constants  $c_i$  are chosen as very large numbers, then typically  $\hat{y} = 0$  in any optimal solution  $(\hat{x}, \hat{y})$  of (2.2), and then the corresponding  $\hat{x}$  is an optimal solution of (2.1).

We prefer to work with (2.2) instead of (2.1) for several reasons. First, there always exist feasible solutions of (2.2), and also at least one optimal solution. Further, each optimal solution (local or global) of (2.2) always satisfies the KKT conditions. There is also a reason from a modelling point of view: In many applications, the user should be able to give a rough overestimate (possibly very large) of how much he would require in improved objective value in order to accept a unit increase of the right-hand side of a certain constraint in (2.1). Such an overestimate could then be used as the corresponding coefficient  $c_i$  in (2.2), and then problem (2.2) would be at least as relevant as problem (2.1).

As mentioned above, and as will be proved below, there always exists at least one KKT point of (2.2), i.e., a point which satisfies the KKT conditions of the problem. The following relations between KKT points of (2.1) and (2.2) can be readily seen by comparing the KKT conditions for the two problems. First, assume that  $\hat{x}$  is a KKT point of (2.1) with Lagrange multipliers  $\lambda_i$  for the constraints  $f_i(x) \leq 0$ , and assume that  $c_i \geq \lambda_i$  for all  $i$ . Then  $(x, y) = (\hat{x}, 0)$  is a KKT point of (2.2) with precisely these values  $\lambda_i$  on the Lagrange multipliers for the constraints  $f_i(x) - y_i \leq 0$ . Next, assume that  $(x, y) = (\hat{x}, 0)$  is a KKT point of (2.2) with Lagrange multipliers  $\lambda_i$  for the constraints  $f_i(x) - y_i \leq 0$  (which will of necessity satisfy  $\lambda_i \leq c_i$  for all  $i$ ). Then  $\hat{x}$  is a KKT point of (2.1) with precisely these values  $\lambda_i$  on the Lagrange multipliers for the constraints  $f_i(x) \leq 0$ . If there happens to be no KKT point of (2.1), then there is no KKT point of (2.2) with  $y = 0$ , no matter how large the coefficients  $c_i$  are chosen. In this case, however, there is always at least one KKT point of (2.2) with  $y \neq 0$ .

For the remainder of this paper, we will in fact consider a further extended problem formulation, with one more “artificial” variable  $z \in \mathbb{R}$ . This formulation contains (2.2) as a special case, but also some other important problem classes such as least squares problems and minimax problems. The (small) price we have to pay for this generality is that the formulation of the problem may look a bit messy at first sight, namely as follows:

$$\begin{aligned}
 (2.3) \quad & \text{minimize} && f_0(x) + a_0 z + \sum_{i=1}^m (c_i y_i + \frac{1}{2} d_i y_i^2) \\
 & \text{subject to} && f_i(x) - a_i z - y_i \leq 0, \quad i = 1, \dots, m, \\
 & && x_j^{\min} \leq x_j \leq x_j^{\max}, \quad j = 1, \dots, n, \\
 & && z \geq 0 \text{ and } y_i \geq 0, \quad i = 1, \dots, m.
 \end{aligned}$$

Here,  $f_0, f_1, \dots, f_m$  are given, twice continuously differentiable, real-valued functions, while  $a_0, a_i, c_i$ , and  $d_i$  are given real numbers such that  $a_0 > 0, a_i \geq 0, c_i \geq 0, d_i \geq 0$ ,

and  $c_i + d_i > 0$  for  $i = 1, \dots, m$ . Further,  $a_i c_i > a_0$  for all  $i$  such that  $a_i > 0$ . Finally,  $x_j^{\min}$  and  $x_j^{\max}$  are given real numbers such that  $x_j^{\min} < x_j^{\max}$  for  $j = 1, \dots, n$ .

Problem (2.2) is obtained as a special case of (2.3) by letting  $a_i = d_i = 0$  for  $i = 1, \dots, m$  and  $a_0 = 1$ , since then  $z = 0$  in any optimal solution of (2.3).

As will be shown below, the considered problem (2.3) is equivalent to the following, typically nonsmooth, problem (2.4) in the variables  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ :

$$(2.4) \quad \begin{aligned} &\text{minimize} && f_0(x) + a_0 \max_{i \in \mathcal{A}_1} \left\{ \frac{f_i^+(x)}{a_i} \right\} + \sum_{i \in \mathcal{A}_0} (c_i f_i^+(x) + \frac{1}{2} d_i (f_i^+(x))^2) \\ &\text{subject to} && x \in X, \end{aligned}$$

where we have used the notation

$$\begin{aligned} X &= \{x \in \mathbb{R}^n \mid x_j^{\min} \leq x_j \leq x_j^{\max}, j = 1, \dots, n\}, \\ \mathcal{A}_1 &= \{i \in \{1, \dots, m\} \mid a_i > 0\}, \\ \mathcal{A}_0 &= \{i \in \{1, \dots, m\} \mid a_i = 0\}, \text{ and} \\ f_i^+(x) &= \max\{0, f_i(x)\}. \end{aligned}$$

This formulation (2.4) will *not* be used for solving problem (2.3), but it shows that least squares problems, minimum 1-norm problems, and minimax problems are all special cases of problem (2.3). It is also used in the proof of Proposition 2.3 below.

**PROPOSITION 2.1.** *If  $x \in X$  is held fixed in problem (2.3), the corresponding optimal values of the variables  $y$  and  $z$  are unique. These unique optimal values are as follows: If  $\mathcal{A}_1 = \emptyset$ , then  $z = 0$  and  $y_i = f_i^+(x)$  for  $i \in \{1, \dots, m\}$ . If  $\mathcal{A}_1 \neq \emptyset$ , then  $z = \max_{i \in \mathcal{A}_1} \{ \frac{f_i^+(x)}{a_i} \}$ ,  $y_i = 0$  for  $i \in \mathcal{A}_1$ , and  $y_i = f_i^+(x)$  for  $i \in \mathcal{A}_0$ .*

*Proof.* If  $\mathcal{A}_1 = \emptyset$ , the result follows from the assumptions that  $a_0 > 0$  and  $c_i + d_i > 0$  for all  $i$ . If  $\mathcal{A}_1 \neq \emptyset$ , one also has to use the assumptions that  $a_i c_i > a_0$  for all  $i \in \mathcal{A}_1$ .  $\square$

This implies that the variables  $y_i$  and  $z$  can formally be eliminated from problem (2.3). The resulting problem is precisely (2.4). This gives our next proposition.

**PROPOSITION 2.2.** *The vector  $(\hat{x}, \hat{y}, \hat{z})$  is a global optimal solution of problem (2.3) if and only if  $\hat{x}$  is a global optimal solution of problem (2.4) while  $\hat{y}$  and  $\hat{z}$  are as in Proposition 2.1.*

*Proof.* The proof follows from Proposition 2.1.  $\square$

**PROPOSITION 2.3.** *There is at least one global optimal solution of problem (2.3).*

*Proof.* In problem (2.4), the objective function is continuous on the compact set  $X$ . Thus, there is at least one global optimal solution of problem (2.4). But then Proposition 2.2 implies that there is at least one global optimal solution of problem (2.3).  $\square$

**PROPOSITION 2.4.** *If  $(\hat{x}, \hat{y}, \hat{z})$  is an optimal solution, local or global, of problem (2.3), then there are Lagrange multipliers which together with  $(\hat{x}, \hat{y}, \hat{z})$  satisfy the KKT conditions.*

*Proof.* It is well known (see, e.g., section 9.4 in [6]) that if  $\hat{x}$  is an optimal solution of a problem of the form

$$(2.5) \quad \begin{aligned} &\text{minimize} && h_0(x) \\ &\text{subject to} && h_i(x) \leq 0, \quad i = 1, \dots, m, \\ &&& x \in \mathbb{R}^n, \end{aligned}$$

and if there is a vector  $\Delta x$  such that  $\nabla h_i(\hat{x})\Delta x < 0$  for all  $i > 0$  such that  $h_i(\hat{x}) = 0$  (i.e., the inner product of  $\Delta x$  and the gradient vector of any active constraint is strictly negative), then there are Lagrange multipliers  $\lambda_i, i = 1, \dots, m$ , which together with  $\hat{x}$  satisfy the KKT conditions, which in this case are

$$\begin{aligned} \frac{\partial h_0}{\partial x_j}(\hat{x}) + \sum_{i=1}^m \lambda_i \frac{\partial h_i}{\partial x_j}(\hat{x}) &= 0, \quad j = 1, \dots, n \quad (\partial L / \partial x_j = 0), \\ h_i(\hat{x}) &\leq 0, \quad i = 1, \dots, m \quad (\text{primal feasibility}), \\ \lambda_i &\geq 0, \quad i = 1, \dots, m \quad (\text{dual feasibility}), \\ \lambda_i h_i(\hat{x}) &= 0, \quad i = 1, \dots, m \quad (\text{compl slackness}). \end{aligned}$$

This result shall now be applied to problem (2.3). Assume that  $(\hat{x}, \hat{y}, \hat{z})$  is an optimal solution of problem (2.3) and construct a corresponding vector  $(\Delta x, \Delta y, \Delta z)$  as follows. For  $j = 1, \dots, n$ , let  $\Delta x_j = 1$  if  $\hat{x}_j = x_j^{\min}$ ,  $\Delta x_j = -1$  if  $\hat{x}_j = x_j^{\max}$ ,  $\Delta x_j = 0$ , otherwise. For  $i = 1, \dots, m$ , let  $\Delta y_i = 1 + \sum_{j=1}^n |\frac{\partial f_i}{\partial x_j}(\hat{x})|$ . Finally, let  $\Delta z = 1$ .

Then it is easily checked that the inner product of  $(\Delta x, \Delta y, \Delta z)$  and the gradient vector, calculated at  $(\hat{x}, \hat{y}, \hat{z})$ , of any active constraint in problem (2.3) is strictly negative.  $\square$

**3. General description of a CCSA method.** A CCSA method for solving problems of the form (2.3) consists of “outer” and “inner” iterations. The index  $k$  is used to denote the outer iteration number, while the index  $\ell$  is used to denote the inner iteration number. Within each outer iteration, there may be zero, one, or several inner iterations. The double index  $(k, \ell)$  is used to denote the  $\ell$ th inner iteration within the  $k$ th outer iteration.

The first iteration point  $(x^{(1)}, y^{(1)}, z^{(1)})$  is obtained by first choosing an  $x^{(1)} \in X$ , and then calculating  $y^{(1)}$  and  $z^{(1)}$  in accordance with Proposition 2.1.

An outer iteration, going from the  $k$ th iteration point  $(x^{(k)}, y^{(k)}, z^{(k)})$  to the  $(k + 1)$ th iteration point  $(x^{(k+1)}, y^{(k+1)}, z^{(k+1)})$ , can be described as follows.

Given  $(x^{(k)}, y^{(k)}, z^{(k)})$ , an approximating subproblem is generated and solved. This subproblem is obtained from (2.3) by replacing  $X$  with a certain convex subset  $X^{(k)}$  and by replacing the functions  $f_i(x)$  with certain strictly convex separable functions  $g_i^{(k,0)}(x)$  satisfying  $g_i^{(k,0)}(x^{(k)}) = f_i(x^{(k)})$ . The optimal solution of this subproblem is denoted  $(\hat{x}^{(k,0)}, \hat{y}^{(k,0)}, \hat{z}^{(k,0)})$ .

If  $g_i^{(k,0)}(\hat{x}^{(k,0)}) \geq f_i(\hat{x}^{(k,0)})$  for all  $i = 0, 1, \dots, m$ , the next iteration point becomes  $(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) = (\hat{x}^{(k,0)}, \hat{y}^{(k,0)}, \hat{z}^{(k,0)})$ , and the outer iteration is completed (without any inner iterations needed).

Otherwise, an inner iteration is made, which means that a new subproblem is generated and solved at  $x^{(k)}$ , with new approximating functions  $g_i^{(k,1)}(x)$ , still satisfying  $g_i^{(k,1)}(x^{(k)}) = f_i(x^{(k)})$  but more conservative than  $g_i^{(k,0)}(x)$  for those indices  $i$  for which the above inequality was violated. The optimal solution of this new subproblem is denoted  $(\hat{x}^{(k,1)}, \hat{y}^{(k,1)}, \hat{z}^{(k,1)})$ .

If  $g_i^{(k,1)}(\hat{x}^{(k,1)}) \geq f_i(\hat{x}^{(k,1)})$  for all  $i = 0, 1, \dots, m$ , the next iteration point becomes  $(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) = (\hat{x}^{(k,1)}, \hat{y}^{(k,1)}, \hat{z}^{(k,1)})$ , and the outer iteration is completed. Otherwise, another inner iteration is made, with new approximating functions  $g_i^{(k,2)}(x)$ , etc.

These inner iterations are repeated until  $g_i^{(k,\ell)}(\hat{x}^{(k,\ell)}) \geq f_i(\hat{x}^{(k,\ell)})$  for all  $i = 0, 1, \dots, m$ , which always happens after a finite number of inner iterations. Then the

next iteration point becomes  $(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) = (\hat{x}^{(k,\ell)}, \hat{y}^{(k,\ell)}, \hat{z}^{(k,\ell)})$ , and the outer iteration is completed (with  $\ell$  inner iterations needed).

**4. Requirements on the approximating functions.** The CCSA subproblem looks as follows, for  $k \in \{1, 2, 3, \dots\}$  and  $\ell \in \{0, 1, 2, \dots\}$ :

$$\begin{aligned} & \text{minimize} && g_0^{(k,\ell)}(x) + a_0 z + \sum_{i=1}^m (c_i y_i + \frac{1}{2} d_i y_i^2) \\ & \text{subject to} && g_i^{(k,\ell)}(x) - a_i z - y_i \leq 0, && i = 1, \dots, m, \\ & && x \in X^{(k)}, \quad y \geq 0, \quad z \geq 0, \end{aligned}$$

where the set  $X^{(k)}$  and the approximating functions  $g_i^{(k,\ell)}(x)$  will be specified below.

The set  $X^{(k)}$  is chosen as  $X^{(k)} = X(x^{(k)}, \sigma^{(k)})$ , where  $\sigma^{(k)} = (\sigma_1^{(k)}, \dots, \sigma_n^{(k)})^T$  is a vector of strictly positive parameters, and  $X(\xi, \sigma)$  is the subset of  $X$  defined by

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

Thus,

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

How to choose values on the parameters  $\sigma_j^{(k)}$  will be discussed later. For the moment, it is sufficient to know that each vector  $\sigma^{(k)}$  belongs to a given compact set  $S$  of the form

$$(4.1) \quad S = \{\sigma \in \mathbb{R}^n \mid \sigma_j^{\min} \leq \sigma_j \leq \sigma_j^{\max}, j = 1, \dots, n\},$$

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

The approximating functions  $g_i^{(k,\ell)}(x)$  in the CCSA subproblem are chosen as

$$(4.2) \quad g_i^{(k,\ell)}(x) = v_i(x, x^{(k)}, \sigma^{(k)}) + \rho_i^{(k,\ell)} w_i(x, x^{(k)}, \sigma^{(k)}), \quad i = 0, 1, \dots, m,$$

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

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

In order to ensure that the functions  $g_i^{(k,\ell)}(x)$  in (4.2) have suitable properties, the following conditions (4.3a)–(4.3k) must be satisfied for  $i = 0, 1, \dots, m$ :

(4.3a)  $v_i$  and  $w_i$  are continuous functions on the set  $D$ ,

(4.3b)  $\nabla_x v_i = \left( \frac{\partial v_i}{\partial x_1}, \dots, \frac{\partial v_i}{\partial x_n} \right)$  exists and is continuous on  $D$ ,

(4.3c)  $\nabla_x w_i = \left( \frac{\partial w_i}{\partial x_1}, \dots, \frac{\partial w_i}{\partial x_n} \right)$  exists and is continuous on  $D$ ,

(4.3d) the  $n \times n$  Hessian matrix  $\nabla_{xx}^2 v_i$  exists and is continuous on  $D$ ,

(4.3e) the  $n \times n$  Hessian matrix  $\nabla_{xx}^2 w_i$  exists and is continuous on  $D$ ,

(4.3f)  $v_i(x, \xi, \sigma) = f_i(x)$  if  $x = \xi \in X$ ,

- (4.3g)  $w_i(x, \xi, \sigma) = 0$  if  $x = \xi \in X$ ,
- (4.3h)  $\nabla_x v_i(x, \xi, \sigma) = \nabla f_i(x)$  if  $x = \xi \in X$ ,
- (4.3i)  $\nabla_x w_i(x, \xi, \sigma) = (0, \dots, 0)$  if  $x = \xi \in X$ ,
- (4.3j)  $\nabla_{xx}^2 v_i(x, \xi, \sigma)$  is positive semidefinite for all  $(x, \xi, \sigma) \in D$ ,
- (4.3k)  $\nabla_{xx}^2 w_i(x, \xi, \sigma)$  is positive definite for all  $(x, \xi, \sigma) \in D$ .

Some choices of appropriate function  $v_i$  and  $w_i$  will be suggested in section 5. The parameters  $\rho_i^{(k,\ell)}$  are strictly positive. The larger the  $\rho_i^{(k,\ell)}$ , the more conservative the approximation will be. Within a given outer iteration  $k$ , the only differences between two inner iterations are the values of these  $\rho_i^{(k,\ell)}$ . How to choose values on these parameters will be described below.

It follows from the above conditions that the functions  $g_i^{(k,\ell)}$  are first order approximations of the original functions  $f_i$  at the current iteration point, i.e.,

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

Further, the approximating functions  $g_i^{(k,\ell)}$  are strictly convex since  $\rho_i^{(k,\ell)} > 0$ . In addition to the above conditions (4.3a)–(4.3k), the approximating functions should be separable, i.e., on the form

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

This property is not used in the forthcoming theoretical analysis of global convergence, but it is essential in practice when attacking large scale problems.

**5. Four examples of CCSA functions.** In this section we give four different examples of CCSA functions  $v_i$  and  $w_i$ . In each of these four examples, and for each fixed vector  $\lambda \geq 0 \in \mathbb{R}^m$ , the Lagrange function  $L(x, y, z, \lambda)$  corresponding to the CCSA subproblem can easily be minimized analytically with respect to  $x \in X^{(k)}$ ,  $y \geq 0$ , and  $z \geq 0$ . If all  $d_i > 0$  and a term  $\varepsilon z^2$  is added to the objective function, this analytical minimization gives a unique point  $(\hat{x}(\lambda), \hat{y}(\lambda), \hat{z}(\lambda))$ . The concave dual function  $\varphi(\lambda) = L(\hat{x}(\lambda), \hat{y}(\lambda), \hat{z}(\lambda), \lambda)$  then becomes an explicit function, and the dual problem of maximizing  $\varphi(\lambda)$  subject to the simple bounds  $\lambda_i \geq 0, i = 1, \dots, m$ , can be solved by, e.g., a conjugate gradient or a Newton-type method, combined with an active set strategy to take care of the nonnegativity constraints on the dual variables. If  $\hat{\lambda}$  is an optimal solution of this dual problem, then  $(x, y, z) = (\hat{x}(\hat{\lambda}), \hat{y}(\hat{\lambda}), \hat{z}(\hat{\lambda}))$  is the unique optimal solution of the CCSA subproblem.

*Example 5.1.* Linear and separable quadratic approximations:

$$\begin{aligned} v_i(x, \xi, \sigma) &= f_i(\xi) + \nabla f_i(\xi)(x - \xi), \quad \text{and} \\ w_i(x, \xi, \sigma) &= \frac{1}{2} \sum_{j=1}^n \left( \frac{x_j - \xi_j}{\sigma_j} \right)^2, \quad \text{so that} \\ g_i^{(k,\ell)}(x) &= f_i(x^{(k)}) + \nabla f_i(x^{(k)})(x - x^{(k)}) + \frac{\rho_i^{(k,\ell)}}{2} \sum_{j=1}^n \left( \frac{x_j - x_j^{(k)}}{\sigma_j^{(k)}} \right)^2. \end{aligned}$$

*Example 5.2.* Linear and separable logarithm approximations:

$$v_i(x, \xi, \sigma) = f_i(\xi) + \nabla f_i(\xi)(x - \xi), \quad \text{and}$$

$$w_i(x, \xi, \sigma) = -\frac{1}{2} \sum_{j=1}^n \ln(1 - (x_j - \xi_j)^2 / \sigma_j^2), \text{ so that}$$

$$g_i^{(k, \ell)}(x) = f_i(x^{(k)}) + \nabla f_i(x^{(k)})(x - x^{(k)})$$

$$- \frac{\rho_i^{(k, \ell)}}{2} \sum_{j=1}^n \ln(1 - (x_j - x_j^{(k)})^2 / (\sigma_j^{(k)})^2).$$

*Example 5.3.* Linear and separable square root approximations:

$$v_i(x, \xi, \sigma) = f_i(\xi) + \nabla f_i(\xi)(x - \xi), \text{ and}$$

$$w_i(x, \xi, \sigma) = \sum_{j=1}^n \left(1 - \sqrt{1 - (x_j - \xi_j)^2 / \sigma_j^2}\right), \text{ so that}$$

$$g_i^{(k, \ell)}(x) = f_i(x^{(k)}) + \nabla f_i(x^{(k)})(x - x^{(k)})$$

$$+ \rho_i^{(k, \ell)} \sum_{j=1}^n \left(1 - \sqrt{1 - (x_j - x_j^{(k)})^2 / (\sigma_j^{(k)})^2}\right).$$

*Example 5.4.* MMA approximations: Here, the approximating functions are chosen as

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

where the “moving asymptotes”  $l_j^{(k)}$  and  $u_j^{(k)}$  are given by

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

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

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

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

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

This means that

$$g_i^{(k, \ell)}(x) = v_i(x, x^{(k)}, \sigma^{(k)}) + \rho_i^{(k, \ell)} w_i(x, x^{(k)}, \sigma^{(k)}),$$

where, after some manipulations,

$$v_i(x, \xi, \sigma) = f_i(\xi) + \sum_{j=1}^n \frac{\sigma_j^2 \frac{\partial f_i}{\partial x_j}(\xi)(x_j - \xi_j) + \sigma_j \left| \frac{\partial f_i}{\partial x_j}(\xi) \right| (x_j - \xi_j)^2}{\sigma_j^2 - (x_j - \xi_j)^2}, \text{ and}$$

$$w_i(x, \xi, \sigma) = \frac{1}{2} \sum_{j=1}^n \frac{(x_j - \xi_j)^2}{\sigma_j^2 - (x_j - \xi_j)^2}.$$

Example 5.4 defines the new globally convergent version of MMA, which is a further development of [8]. The original MMA, [7], can be considered as a special case of the above by letting all  $\rho_i^{(k,\ell)} = 0$ . Consequently, no inner iterations were performed in the original MMA, and global convergence could not be proved.

**6. Rules for updating the parameters  $\rho_i^{(k,\ell)}$  and  $\sigma_j^{(k)}$ .** We begin with the parameters  $\rho_i^{(k,\ell)}$ . For  $\ell = 0$ , the following values are used, where  $\rho_i^{\min}$  is a fixed, strictly positive “small” number, e.g.,  $10^{-5}$ :

$$(6.1a) \quad \rho_i^{(1,0)} = 1,$$

$$(6.1b) \quad \rho_i^{(k+1,0)} = \max\{0.1\rho_i^{(k,\hat{\ell}(k))}, \rho_i^{\min}\},$$

where  $\hat{\ell}(k)$  is the number of inner iterations needed within the  $k$ th outer iteration, so that  $\rho_i^{(k,\hat{\ell}(k))}$  is the latest value of  $\rho_i^{(k,\ell)}$ .

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

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

which in view of (4.2) gives that  $\rho_i^{(k,\ell+1)} = \rho_i^{(k,\ell)} + \delta_i^{(k,\ell)}$ , where

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

In order to get a globally convergent method, this natural value is modified as follows:

$$(6.2) \quad \begin{aligned} \rho_i^{(k,\ell+1)} &= \min\{10\rho_i^{(k,\ell)}, 1.1(\rho_i^{(k,\ell)} + \delta_i^{(k,\ell)})\} & \text{if } \delta_i^{(k,\ell)} > 0, \\ \rho_i^{(k,\ell+1)} &= \rho_i^{(k,\ell)} & \text{if } \delta_i^{(k,\ell)} \leq 0. \end{aligned}$$

This means that in the beginning of each new inner iteration, the parameters  $\rho_i$  are increased or unaltered but never decreased. Therefore, it is important that they can be decreased again in the beginning of each new outer iteration, as they are in (6.1b), since otherwise the method could be too conservative.

Now to the values of the parameters  $\sigma_j^{(k)}$ . Updating rules for these parameters depend on the specific functions  $v_i$  and  $w_i$ . In each of the four examples in the previous section, the  $n \times n$  Hessian matrix  $\nabla_{xx}^2 w_i(x, \xi, \sigma)$  is diagonal with  $\frac{\partial^2 w_i}{\partial x_j^2}(x, \xi, \sigma) \geq \frac{1}{\sigma_j^2}$  for all  $j$  and every  $(x, \xi, \sigma) \in D$ , with equality if  $x_j = \xi_j$ . The curvature of the function  $w_i$  in the “ $x_j$ -direction” thus increases with decreasing values of  $\sigma_j$ . This makes the following heuristic rule for updating these parameters reasonable. If a certain variable  $x_j$  is oscillating, it should be stabilized by a decreased value of the corresponding  $\sigma_j$ . If the variable  $x_j$  is monotonically increasing, or monotonically decreasing, it should be released by an increased value of the corresponding  $\sigma_j$ . One possible way of implementing this rule is as follows.

In the first two outer iterations, when  $k = 1$  and  $k = 2$ ,

$$\sigma_j^{(k)} = 0.5(x_j^{\max} - x_j^{\min}),$$

while in later outer iterations, when  $k \geq 3$ ,

$$\sigma_j^{(k)} = \gamma_j^{(k)} \sigma_j^{(k-1)},$$

where

$$\gamma_j^{(k)} = \begin{cases} 0.7 & \text{if } (x_j^{(k)} - x_j^{(k-1)})(x_j^{(k-1)} - x_j^{(k-2)}) < 0, \\ 1.2 & \text{if } (x_j^{(k)} - x_j^{(k-1)})(x_j^{(k-1)} - x_j^{(k-2)}) > 0, \\ 1 & \text{if } (x_j^{(k)} - x_j^{(k-1)})(x_j^{(k-1)} - x_j^{(k-2)}) = 0, \end{cases}$$

provided that this leads to values that satisfy

$$0.01(x_j^{\max} - x_j^{\min}) \leq \sigma_j^{(k)} \leq 10(x_j^{\max} - x_j^{\min}).$$

If any of these bounds is violated, the corresponding  $\sigma_j^{(k)}$  is set to the violated bound. Thus,  $\sigma_j^{\min} = 0.01(x_j^{\max} - x_j^{\min})$  and  $\sigma_j^{\max} = 10(x_j^{\max} - x_j^{\min})$  in the set  $S$  defined in (4.1) above.

**7. Theoretical analysis of global convergence.** A given point  $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}$  is a KKT point of the problem (2.3) if and only if there are Lagrange multipliers which together with  $(x, y, z)$  satisfy the KKT conditions of the problem.

Let  $\Omega$  be the set of KKT points of the original problem (2.3).  $\Omega$  is nonempty by Propositions 2.3 and 2.4. Then let  $\|\Omega - (x^{(k)}, y^{(k)}, z^{(k)})\|$  denote the Euclidean distance from the point  $(x^{(k)}, y^{(k)}, z^{(k)})$  to the set  $\Omega$ , i.e.,

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

**THEOREM 7.1.** *If any of the CCSA methods described above is applied to a problem of the form (2.3), then  $\|\Omega - (x^{(k)}, y^{(k)}, z^{(k)})\| \rightarrow 0$  as  $k \rightarrow \infty$ .*

Before the proof of this main theorem, some preparations are needed.

**LEMMA 7.2.** *In each outer iteration  $k$ , only a finite number  $\ell$  of inner iterations are needed until  $g_i^{(k, \ell)}(\hat{x}^{(k, \ell)}) \geq f_i(\hat{x}^{(k, \ell)})$  for all  $i$ .*

*Proof.* A sufficient condition for the inequality  $g_i^{(k, \ell)}(\hat{x}^{(k, \ell)}) \geq f_i(\hat{x}^{(k, \ell)})$  to hold is that  $\rho_i^{(k, \ell)} \tau_i \geq \kappa_i$ , where

$$\begin{aligned} \kappa_i &= \max_{x, h} \{ h^T \nabla^2 f_i(x) h \mid x \in X, h \in \mathbb{R}^n, h^T h = 1 \}, \text{ and} \\ \tau_i &= \min_{x, \xi, \sigma, h} \{ h^T \nabla_{xx}^2 w_i(x, \xi, \sigma) h \mid (x, \xi, \sigma) \in D, h \in \mathbb{R}^n, h^T h = 1 \}. \end{aligned}$$

The number  $\kappa_i$  is finite since the Hessian matrix  $\nabla^2 f_i(x)$  is continuous on  $X$ . The number  $\tau_i$  is finite and strictly positive since the Hessian matrix  $\nabla_{xx}^2 w_i(x, \xi, \sigma)$  is positive definite and continuous in all its arguments. But each time that  $g_i^{(k, \ell)}(\hat{x}^{(k, \ell)}) < f_i(\hat{x}^{(k, \ell)})$ , the corresponding  $\rho_i^{(k, \ell)}$  is increased by at least a factor 1.1; see (6.2). This can be done only a finite number of times, for each  $i$ , before  $\rho_i^{(k, \ell)} \tau_i \geq \kappa_i$  is satisfied. (Note that, for a fixed  $k$ ,  $\rho_i^{(k, \ell)}$  is nondecreasing in  $\ell$ .)  $\square$

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

$$\begin{aligned} \hat{\ell}(k) &= \text{the number of inner iterations needed within the } k\text{th outer iteration,} \\ \rho_i^{(k)} &= \rho_i^{(k, \hat{\ell}(k))}, \text{ and } g_i^{(k)}(x) = g_i^{(k, \hat{\ell}(k))}(x). \end{aligned}$$

This means that the subproblem used at the  $k$ th (outer) iteration to calculate the next iteration point is the following:

$$\begin{aligned}
 (7.1) \quad & \text{minimize} && g_0^{(k)}(x) + a_0z + \sum_{i=1}^m (c_i y_i + \frac{1}{2}d_i y_i^2) \\
 & \text{subject to} && g_i^{(k)}(x) - a_i z - y_i \leq 0, && i = 1, \dots, m, \\
 & && x \in X^{(k)}, \quad y \geq 0, \quad z \geq 0.
 \end{aligned}$$

The optimal solution of (7.1) is the new iteration point  $(x^{(k+1)}, y^{(k+1)}, z^{(k+1)})$ . Note that  $g_i^{(k)}(x^{(k)}) = f_i(x^{(k)})$  and  $g_i^{(k)}(x^{(k+1)}) \geq f_i(x^{(k+1)})$  for all  $i = 0, 1, \dots, m$ .

LEMMA 7.3. *For each  $i = 0, 1, \dots, m$ , 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 updating rules (6.1b) and (6.2) and the proof of Lemma 7.2, it follows that  $\rho_i^{(k)} \leq 10(1 + \kappa_i/\tau_i)$  will always hold.  $\square$

Let the set  $Q$  be defined by

$$Q = \{\rho \in \mathbb{R}^{m+1} \mid \rho_i^{\min} \leq \rho_i \leq \rho_i^{\max}, \quad i = 0, 1, \dots, m\}.$$

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

$$\begin{aligned}
 F_0(x, y, z) &= f_0(x) + a_0z + \sum_{i=1}^m (c_i y_i + \frac{1}{2}d_i y_i^2), \\
 F_i(x, y, z) &= f_i(x) - a_i z - y_i, \quad i = 1, \dots, m.
 \end{aligned}$$

Then the original problem (2.3) can be written

$$\begin{aligned}
 (7.2) \quad & \text{minimize} && F_0(x, y, z) \\
 & \text{subject to} && F_i(x, y, z) \leq 0, && i = 1, \dots, m, \\
 & && x \in X, \quad y \geq 0, \quad z \geq 0.
 \end{aligned}$$

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

$$\begin{aligned}
 G_0(x, y, z, \xi, \sigma, \rho) &= v_0(x, \xi, \sigma) + \rho_0 w_0(x, \xi, \sigma) + a_0z + \sum_{i=1}^m (c_i y_i + \frac{1}{2}d_i y_i^2), \\
 G_i(x, y, z, \xi, \sigma, \rho) &= v_i(x, \xi, \sigma) + \rho_i w_i(x, \xi, \sigma) - a_i z - y_i, \quad i = 1, \dots, m.
 \end{aligned}$$

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

Let the problem  $\text{PSUB}(\xi, \sigma, \rho)$  be defined, for given  $(\xi, \sigma, \rho) \in X \times S \times Q$ , as the following problem in the variables  $(x, y, z)$ :

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

Then the CCSA subproblem (7.1) is equivalent to the problem  $\text{PSUB}(x^{(k)}, \sigma^{(k)}, \rho^{(k)})$ , i.e., the problem (7.3) with  $\xi = x^{(k)}$ ,  $\sigma = \sigma^{(k)}$ , and  $\rho = \rho^{(k)}$ .

LEMMA 7.4. *For each given  $\xi \in X$ ,  $\sigma \in S$ , and  $\rho \in Q$ , there is a unique optimal solution of  $\text{PSUB}(\xi, \sigma, \rho)$ . This solution is also the only KKT point of  $\text{PSUB}(\xi, \sigma, \rho)$ .*

*Proof.* The existence of an optimal solution follows by arguments similar to those in the proof of Proposition 2.3. The uniqueness follows from the fact that the problem obtained by eliminating  $y$  and  $z$  is strictly convex in  $x$ . Finally,  $\text{PSUB}(\xi, \sigma, \rho)$  is a convex problem for which the Slater's constraint qualifications are fulfilled. Therefore, the KKT conditions are both necessary and sufficient conditions for a global optimum.  $\square$

Thus,  $(x^{(k+1)}, y^{(k+1)}, z^{(k+1)})$  is the only KKT point of  $\text{PSUB}(x^{(k)}, \sigma^{(k)}, \rho^{(k)})$ .

LEMMA 7.5. *For each given  $\sigma \in S$  and  $\rho \in Q$  the following holds: A given point  $(\hat{x}, \hat{y}, \hat{z})$  is a KKT point of the original problem (2.3) if and only if  $(\hat{x}, \hat{y}, \hat{z})$  is a KKT point of the subproblem  $\text{PSUB}(\hat{x}, \sigma, \rho)$ .*

*Proof.* For a given  $\hat{x} \in X$ , let  $B(\hat{x}, \varepsilon) = \{x \in \mathbb{R}^n ; \|x - \hat{x}\| < \varepsilon\}$ , and note that there is an  $\varepsilon > 0$  such that  $X \cap B(\hat{x}, \varepsilon) = X(\hat{x}, \sigma) \cap B(\hat{x}, \varepsilon)$ . This implies that  $(\hat{x}, \hat{y}, \hat{z})$  is the optimal solution of (the strictly convex problem)  $\text{PSUB}(\hat{x}, \sigma, \rho)$  if and only if  $(\hat{x}, \hat{y}, \hat{z})$  is the optimal solution of  $\text{PSUB}(\hat{x}, \sigma, \rho)$  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  $i = 0, 1, \dots, m$ :

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

These observations imply that  $(\hat{x}, \hat{y}, \hat{z})$  is a KKT point of the subproblem  $\text{PSUB}(\hat{x}, \sigma, \rho)$  if and only if  $(\hat{x}, \hat{y}, \hat{z})$  is a KKT point of the problem (7.2).  $\square$

In particular, if  $(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) = (x^{(k)}, y^{(k)}, z^{(k)})$ , then  $(x^{(k)}, y^{(k)}, z^{(k)})$  is a KKT point of the original problem (2.3), and then the algorithm should be stopped. From now on, it is therefore assumed that  $(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) \neq (x^{(k)}, y^{(k)}, z^{(k)})$  for all  $k$ .

LEMMA 7.6. *Each generated iteration point is a feasible solution of the problem (7.2), i.e.,  $F_i(x^{(k)}, y^{(k)}, z^{(k)}) \leq 0$  for  $i \geq 1$  and  $k \geq 1$ . Further, each generated iteration point has a strictly lower objective value than the previous one, i.e.,  $F_0(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) < F_0(x^{(k)}, y^{(k)}, z^{(k)})$  for  $k \geq 1$ .*

*Proof.* The starting point  $(x^{(1)}, y^{(1)}, z^{(1)})$  is feasible by construction. After that,  $F_i(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) \leq G_i(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) \leq 0$  for  $i \geq 1$ . Further,  $F_0(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) \leq G_0(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) < G_0(x^{(k)}, y^{(k)}, z^{(k)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) = F_0(x^{(k)}, y^{(k)}, z^{(k)})$ .  $\square$

LEMMA 7.7. *All the iteration points  $(x^{(k)}, y^{(k)}, z^{(k)})$  remain in a compact set.*

*Proof.* First,  $x^{(k)} \in X$ , which is a compact set. Next, let the functions  $\tilde{g}_i$  be defined, for  $(x, \xi, \sigma) \in D$  and  $\rho \in Q$ , by

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

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

By the same arguments as in Proposition 2.1, it follows that  $y_i^{(k+1)} \leq g_i^{(k)}(x^{(k+1)})$ . But since  $g_i^{(k)}(x^{(k+1)}) = \tilde{g}_i(x^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)})$ , it then follows that  $y_i^{(k+1)} \leq \tilde{g}_i(x^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) \leq \max\{\tilde{g}_i(x, \xi, \sigma, \rho) \mid (x, \xi, \sigma) \in D, \rho \in Q\}$ .

The existence of an upper bound on  $z^{(k)}$  is proved in a similar way.  $\square$

As a consequence of Lemma 7.7, the sequence  $\{(x^{(k)}, y^{(k)}, z^{(k)})\}_{k=1}^\infty$  has at least one convergent subsequence. Thus, there is a point  $(x^*, y^*, z^*)$  and an infinite subset  $\mathcal{K}$  of the positive integers such that  $(x^{(k)}, y^{(k)}, z^{(k)}) \rightarrow (x^*, y^*, z^*)$  as  $k \in \mathcal{K}$  and  $k \rightarrow \infty$ .

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  $(\sigma^{(k)}, \rho^{(k)}) \rightarrow (\sigma^*, \rho^*)$  as  $k \in \tilde{\mathcal{K}}$  and  $k \rightarrow \infty$ .

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

In the following,  $(x^*, y^*, z^*)$ ,  $(\sigma^*, \rho^*)$ , and  $(\bar{x}, \bar{y}, \bar{z})$  are these just-described limit points.

LEMMA 7.8.  $F_0(x^{(k)}, y^{(k)}, z^{(k)}) \rightarrow F_0(x^*, y^*, z^*)$  as  $k \rightarrow \infty$  (not only for  $k \in \mathcal{K}$ ).

*Proof.* The sequence  $\{F_0(x^{(k)}, y^{(k)}, z^{(k)})\}_{k=1}^\infty$  is monotonically decreasing and bounded below by the global optimal value of the problem (2.3) (which exists and is finite according to Proposition 2.3). Thus,  $F_0(x^{(k)}, y^{(k)}, z^{(k)}) \rightarrow F_0^*$  as  $k \rightarrow \infty$  for some real number  $F_0^*$ . But since  $F_0(x^{(k)}, y^{(k)}, z^{(k)}) \rightarrow F_0(x^*, y^*, z^*)$  as  $k \in \mathcal{K}$  and  $k \rightarrow \infty$ , it follows that  $F_0^* = F_0(x^*, y^*, z^*)$ .  $\square$

LEMMA 7.9.  $F_0(\bar{x}, \bar{y}, \bar{z}) = F_0(x^*, y^*, z^*)$ .

*Proof.* From Lemma 7.8, it follows that  $F_0(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) \rightarrow F_0(x^*, y^*, z^*)$  as  $k \in \bar{\mathcal{K}}$  and  $k \rightarrow \infty$ . But since  $(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) \rightarrow (\bar{x}, \bar{y}, \bar{z})$  as  $k \in \bar{\mathcal{K}}$  and  $k \rightarrow \infty$ , it also holds that  $F_0(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}) \rightarrow F_0(\bar{x}, \bar{y}, \bar{z})$  as  $k \in \bar{\mathcal{K}}$  and  $k \rightarrow \infty$ .  $\square$

LEMMA 7.10.  $(\bar{x}, \bar{y}, \bar{z})$  is the unique optimal solution of  $\text{PSUB}(x^*, \sigma^*, \rho^*)$ .

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

For  $\nu = 1, 2, 3, \dots$ , let  $\tilde{x}^{(\nu)} = \tilde{x} + \alpha^{(\nu)}(x^* - \tilde{x})$ ,  $\tilde{y}^{(\nu)} = \tilde{y} + \frac{1}{\nu}(1, \dots, 1)^T$ , and  $\tilde{z}^{(\nu)} = \tilde{z} + \frac{1}{\nu}$ . If  $\alpha^{(\nu)} = 0$ , then  $G_i(\tilde{x}^{(\nu)}, \tilde{y}^{(\nu)}, \tilde{z}^{(\nu)}, x^*, \sigma^*, \rho^*) \leq -\frac{1}{\nu}$  for  $i \geq 1$ . It is therefore possible to choose the scalar  $\alpha^{(\nu)}$  such that  $0 < \alpha^{(\nu)} < 1/\nu$  and  $G_i(\tilde{x}^{(\nu)}, \tilde{y}^{(\nu)}, \tilde{z}^{(\nu)}, x^*, \sigma^*, \rho^*) \leq -\frac{1}{2\nu}$  for  $i \geq 1$ . Then  $(\tilde{x}^{(\nu)}, \tilde{y}^{(\nu)}, \tilde{z}^{(\nu)})$  is in the interior of the feasible set of  $\text{PSUB}(x^*, \sigma^*, \rho^*)$ . In particular,  $\tilde{x}^{(\nu)}$  is in the interior of  $X(x^*, \sigma^*)$ . This implies that for each  $\nu$ , there is an integer  $K(\nu)$  such that, for all  $k \in \bar{\mathcal{K}}$  with  $k > K(\nu)$ ,  $\tilde{x}^{(\nu)} \in X(x^{(k)}, \sigma^{(k)})$  and  $G_i(\tilde{x}^{(\nu)}, \tilde{y}^{(\nu)}, \tilde{z}^{(\nu)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) \leq 0$  for  $i \geq 1$ . For all these  $k \in \bar{\mathcal{K}}$  with  $k > K(\nu)$  it then holds that  $G_0(\tilde{x}^{(\nu)}, \tilde{y}^{(\nu)}, \tilde{z}^{(\nu)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)}) \geq G_0(x^{(k+1)}, y^{(k+1)}, z^{(k+1)}, x^{(k)}, \sigma^{(k)}, \rho^{(k)})$  (because  $(x^{(k+1)}, y^{(k+1)}, z^{(k+1)})$  is the optimal solution of  $\text{PSUB}(x^{(k)}, \sigma^{(k)}, \rho^{(k)})$ ).

Now, for each  $\nu$ , let the integer  $k(\nu) \in \bar{\mathcal{K}}$  satisfy  $k(\nu) > \max\{\nu, K(\nu)\}$ , and let  $\nu \rightarrow \infty$ . Then  $(\tilde{x}^{(\nu)}, \tilde{y}^{(\nu)}, \tilde{z}^{(\nu)}) \rightarrow (\tilde{x}, \tilde{y}, \tilde{z})$ ,  $(x^{(k(\nu)+1)}, y^{(k(\nu)+1)}, z^{(k(\nu)+1)}) \rightarrow (x^*, y^*, z^*)$ , and  $(x^{(k(\nu))}, \sigma^{(k(\nu))}, \rho^{(k(\nu))}) \rightarrow (x^*, \sigma^*, \rho^*)$ . Thus,  $G_0(\tilde{x}, \tilde{y}, \tilde{z}, x^*, \sigma^*, \rho^*) \geq G_0(\bar{x}, \bar{y}, \bar{z}, x^*, \sigma^*, \rho^*)$ .  $\square$

LEMMA 7.11.  $(\bar{x}, \bar{y}, \bar{z}) = (x^*, y^*, z^*)$ .

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

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

By definition,  $F_0(x^*, y^*, z^*) = G_0(x^*, y^*, z^*, x^*, \sigma^*, \rho^*)$ . From Lemma 7.9 it then follows that  $G_0(x^*, y^*, z^*, x^*, \sigma^*, \rho^*) \leq G_0(\bar{x}, \bar{y}, \bar{z}, x^*, \sigma^*, \rho^*)$ . But since  $(\bar{x}, \bar{y}, \bar{z})$  is the unique global optimal solution of  $\text{PSUB}(x^*, \sigma^*, \rho^*)$ , it then follows that  $(x^*, y^*, z^*) = (\bar{x}, \bar{y}, \bar{z})$ .  $\square$

LEMMA 7.12.  $(x^*, y^*, z^*)$  is a KKT point of the problem (7.2).

*Proof.* Follows from Lemmas 7.4, 7.5, 7.10, and 7.11.  $\square$

*Proof of Theorem 7.1.* Assume that the statement in Theorem 7.1 is false. Then there is an  $\varepsilon > 0$  and an infinite subset  $\mathcal{K}_0$  of the integers such that

$$(7.4) \quad \|(x, y, z) - (x^{(k)}, y^{(k)}, z^{(k)})\| \geq \varepsilon \text{ for all } (x, y, z) \in \Omega \text{ and every } k \in \mathcal{K}_0.$$

Then, as a consequence of Lemma 7.7, the sequence  $\{(x^{(k)}, y^{(k)}, z^{(k)})\}_{k \in \mathcal{K}_0}$  has at least one convergent subsequence. Thus, there is a point  $(\hat{x}, \hat{y}, \hat{z})$  and an infinite subset  $\bar{\mathcal{K}}_0 \subseteq \mathcal{K}_0$  such that  $(x^{(k)}, y^{(k)}, z^{(k)}) \rightarrow (\hat{x}, \hat{y}, \hat{z})$  as  $k \in \bar{\mathcal{K}}_0$  and  $k \rightarrow \infty$ .

But then, by letting  $(\hat{x}, \hat{y}, \hat{z})$  play the role of  $(x^*, y^*, z^*)$  in the above lemmas, in particular Lemma 7.12, it follows that  $(\hat{x}, \hat{y}, \hat{z})$  is a KKT point of the problem (7.2), and thus also a KKT point of the original problem (2.3). Thus,  $(\hat{x}, \hat{y}, \hat{z}) \in \Omega$ . By letting  $(x, y, z) = (\hat{x}, \hat{y}, \hat{z})$  in (7.4), a contradiction has then been established. Therefore, the statement in Theorem 7.1 can not be false, but must be true.  $\square$

**8. Test problems and numerical results.** As mentioned in the introduction, a major benefit of CCSA methods is that they can be successfully applied to problems with a very large number of variables, even if the Hessian matrices of the objective and constraint functions are dense. Such problems often appear in, e.g., structural optimization, in particular in the subfield dealing with topology optimization. To illustrate this, we present two problems which are parameterized by the integer  $n$  = the number of variables  $x_j$ . The general structure of these problems resembles the corresponding structure of topology optimization problems (nonconvex problems with a large number of variables, upper and lower bounds on all variables, and a relatively small number of general inequality constraints); but in order to facilitate the reader’s making her own numerical tests, the problems are not genuine structural optimization problems (which would require a finite element package) but are instead explicitly stated “academic” problems.

**8.1. Three matrices which are used in the test problems.** Let  $n$  be a given positive integer  $> 1$  and let  $S$ ,  $P$ , and  $Q$  be symmetric  $n \times n$  matrices with elements given by

$$s_{ij} = \frac{2 + \sin(4\pi\alpha_{ij})}{(1 + |i - j|) \ln n}, \quad p_{ij} = \frac{1 + 2\alpha_{ij}}{(1 + |i - j|) \ln n}, \quad q_{ij} = \frac{3 - 2\alpha_{ij}}{(1 + |i - j|) \ln n},$$

where  $\alpha_{ij} = \frac{i+j-2}{2n-2} \in [0, 1]$  for all  $i$  and  $j$ .

The matrices  $S$ ,  $P$ , and  $Q$  are positive definite, and for  $n = 9$  they look as follows:

$$S = \frac{1}{\ln 9} \begin{pmatrix} 2.0000 & 1.3536 & 1.0000 & 0.6768 & 0.4000 & 0.2155 & 0.1429 & 0.1616 & 0.2222 \\ 1.3536 & 3.0000 & 1.3536 & 0.6667 & 0.3232 & 0.2000 & 0.2155 & 0.2857 & 0.3384 \\ 1.0000 & 1.3536 & 2.0000 & 0.6464 & 0.3333 & 0.3232 & 0.4000 & 0.4512 & 0.4286 \\ 0.6768 & 0.6667 & 0.6464 & 1.0000 & 0.6464 & 0.6667 & 0.6768 & 0.6000 & 0.4512 \\ 0.4000 & 0.3232 & 0.3333 & 0.6464 & 2.0000 & 1.3536 & 1.0000 & 0.6768 & 0.4000 \\ 0.2155 & 0.2000 & 0.3232 & 0.6667 & 1.3536 & 3.0000 & 1.3536 & 0.6667 & 0.3232 \\ 0.1429 & 0.2155 & 0.4000 & 0.6768 & 1.0000 & 1.3536 & 2.0000 & 0.6464 & 0.3333 \\ 0.1616 & 0.2857 & 0.4512 & 0.6000 & 0.6768 & 0.6667 & 0.6464 & 1.0000 & 0.6464 \\ 0.2222 & 0.3384 & 0.4286 & 0.4512 & 0.4000 & 0.3232 & 0.3333 & 0.6464 & 2.0000 \end{pmatrix},$$

$$P = \frac{1}{\ln 9} \begin{pmatrix} 1.0000 & 0.5625 & 0.4167 & 0.3438 & 0.3000 & 0.2708 & 0.2500 & 0.2344 & 0.2222 \\ 0.5625 & 1.2500 & 0.6875 & 0.5000 & 0.4062 & 0.3500 & 0.3125 & 0.2857 & 0.2656 \\ 0.4167 & 0.6875 & 1.5000 & 0.8125 & 0.5833 & 0.4688 & 0.4000 & 0.3542 & 0.3214 \\ 0.3438 & 0.5000 & 0.8125 & 1.7500 & 0.9375 & 0.6667 & 0.5312 & 0.4500 & 0.3958 \\ 0.3000 & 0.4062 & 0.5833 & 0.9375 & 2.0000 & 1.0625 & 0.7500 & 0.5938 & 0.5000 \\ 0.2708 & 0.3500 & 0.4688 & 0.6667 & 1.0625 & 2.2500 & 1.1875 & 0.8333 & 0.6562 \\ 0.2500 & 0.3125 & 0.4000 & 0.5312 & 0.7500 & 1.1875 & 2.5000 & 1.3125 & 0.9167 \\ 0.2344 & 0.2857 & 0.3542 & 0.4500 & 0.5938 & 0.8333 & 1.3125 & 2.7500 & 1.4375 \\ 0.2222 & 0.2656 & 0.3214 & 0.3958 & 0.5000 & 0.6562 & 0.9167 & 1.4375 & 3.0000 \end{pmatrix},$$

$$Q = \frac{1}{\ln 9} \begin{pmatrix} 3.0000 & 1.4375 & 0.9167 & 0.6562 & 0.5000 & 0.3958 & 0.3214 & 0.2656 & 0.2222 \\ 1.4375 & 2.7500 & 1.3125 & 0.8333 & 0.5938 & 0.4500 & 0.3542 & 0.2857 & 0.2344 \\ 0.9167 & 1.3125 & 2.5000 & 1.1875 & 0.7500 & 0.5312 & 0.4000 & 0.3125 & 0.2500 \\ 0.6562 & 0.8333 & 1.1875 & 2.2500 & 1.0625 & 0.6667 & 0.4688 & 0.3500 & 0.2708 \\ 0.5000 & 0.5938 & 0.7500 & 1.0625 & 2.0000 & 0.9375 & 0.5833 & 0.4062 & 0.3000 \\ 0.3958 & 0.4500 & 0.5312 & 0.6667 & 0.9375 & 1.7500 & 0.8125 & 0.5000 & 0.3438 \\ 0.3214 & 0.3542 & 0.4000 & 0.4688 & 0.5833 & 0.8125 & 1.5000 & 0.6875 & 0.4167 \\ 0.2656 & 0.2857 & 0.3125 & 0.3500 & 0.4062 & 0.5000 & 0.6875 & 1.2500 & 0.5625 \\ 0.2222 & 0.2344 & 0.2500 & 0.2708 & 0.3000 & 0.3438 & 0.4167 & 0.5625 & 1.0000 \end{pmatrix}.$$

**8.2. Problem 1.** In the first considered problem, called Problem 1, the objective function is strictly convex, but the nonlinear constraint functions are strictly concave so the set of feasible solutions is nonconvex. The formulation of the problem is as follows:

$$(8.1) \quad \begin{aligned} & \text{minimize} && f_0(x) = x^T S x \\ & \text{subject to} && f_1(x) = \frac{n}{2} - x^T P x \leq 0, \\ & && f_2(x) = \frac{n}{2} - x^T Q x \leq 0, \\ & && -1 \leq x_j \leq 1, \quad j = 1, \dots, n, \end{aligned}$$

with starting point  $x^{(0)} = (0.5, 0.5, \dots, 0.5)^T$ .

**8.3. Problem 2.** In the second considered problem, called Problem 2, the non-linear constraint functions are strictly convex, but the objective function is strictly concave and thus nonconvex. The formulation of the problem is as follows.

$$\begin{aligned}
 & \text{minimize} && f_0(x) = -x^T S x \\
 & \text{subject to} && f_1(x) = x^T P x - \frac{n}{2} \leq 0, \\
 & && f_2(x) = x^T Q x - \frac{n}{2} \leq 0, \\
 & && -1 \leq x_j \leq 1, \quad j = 1, \dots, n,
 \end{aligned}
 \tag{8.2}$$

with starting point  $x^{(0)} = (0.25, 0.25, \dots, 0.25)^T$ .

**8.4. Numerical results.** We have used the CCSA method based on MMA approximations (see Example 5.4 in section 5) to solve the above two problems with  $n = 1000, 2000, 5000, 10000,$  and  $20000$ . Both problems are of the form (2.1), and they were first transformed to the form (2.3) with  $a_0 = 1, a_1 = a_2 = 0, d_1 = d_2 = 1,$  and  $c_1 = c_2 = 1000$ . It then turned out that  $y = 0$  and  $z = 0$  in the optimal solution of each generated CCSA subproblem.

Concerning the termination criterion that we used, first note that the KKT conditions of the considered problems (8.1) and (8.2) can be written as follows, using the notations  $a^+ = \max\{0, a\}$  and  $a^- = \max\{0, -a\}$ :

$$(1 + x_j) \left( \frac{\partial f_0}{\partial x_j} + \lambda_1 \frac{\partial f_1}{\partial x_j} + \lambda_2 \frac{\partial f_2}{\partial x_j} \right)^+ = 0, \quad j = 1, \dots, n,
 \tag{8.3a}$$

$$(1 - x_j) \left( \frac{\partial f_0}{\partial x_j} + \lambda_1 \frac{\partial f_1}{\partial x_j} + \lambda_2 \frac{\partial f_2}{\partial x_j} \right)^- = 0, \quad j = 1, \dots, n,
 \tag{8.3b}$$

$$f_i(x)^+ = 0, \quad i = 1, 2,
 \tag{8.3c}$$

$$\lambda_i f_i(x)^- = 0, \quad i = 1, 2,
 \tag{8.3d}$$

$$\lambda_i \geq 0, \quad i = 1, 2,
 \tag{8.3e}$$

$$-1 \leq x_j \leq 1, \quad j = 1, \dots, n.
 \tag{8.3f}$$

Equations (8.3a)–(8.3d) can be written more concisely as  $r_k(x, \lambda) = 0, k = 1, \dots, 2n + 4$ . The inequalities (8.3e) and (8.3f) are always satisfied by the primal variables  $x_j$  and the dual variables  $\lambda_i$  obtained from the solution of the CCSA subproblem. The outer iterations were terminated when these  $x$  and  $\lambda$  also satisfied

$$\frac{1}{n} \sum_{k=1}^{2n+4} (r_k(x, \lambda))^2 \leq 10^{-10}.
 \tag{8.4}$$

A similar, but harder, termination criterion was used when solving the CCSA subproblems; a subproblem was considered as solved when a condition corresponding to (8.4) was satisfied with the right-hand side equal to  $10^{-16}$ .

The optimal solutions obtained for the case  $n = 1000$  are plotted in Figures 8.1 and 8.2, with the index  $j$  on the horizontal axis and  $x_j$  on the vertical axis.

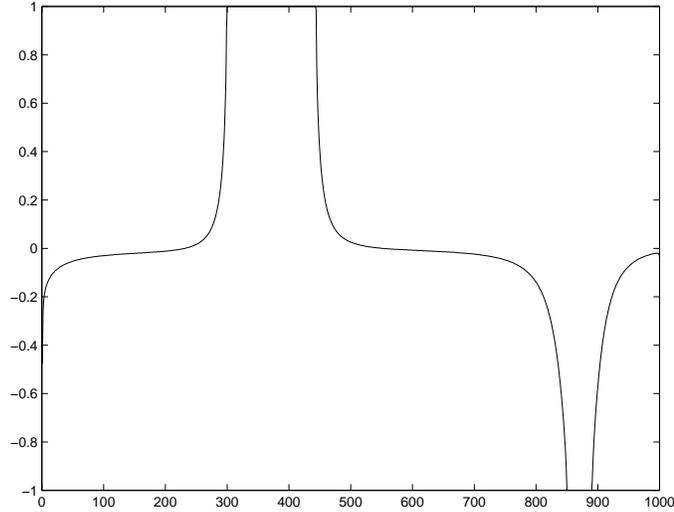


FIG. 8.1. Obtained  $x_j$  for Problem 1 with  $n = 1000$ .

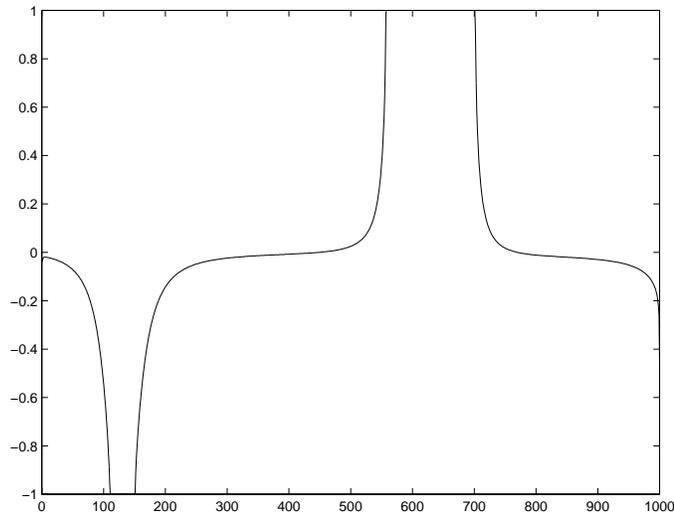


FIG. 8.2. Obtained  $x_j$  for Problem 2 with  $n = 1000$ .

Additional results are presented in Tables 8.1 and 8.2, where for each problem we present the number of variables ( $n$ ), the objective value of the obtained optimal solution, the number of variables which are at the upper or lower bound in the obtained optimal solution, the obtained values of the two Lagrange multipliers  $\lambda_1$  and  $\lambda_2$ , the total number of required outer iterations, and the total number of additionally required inner iterations.

The method was implemented in Fortran 77 on a Sun Enterprise 4000 (using only one of the four processors). The total required CPU-time was approximately  $2n^2/10^6$  CPU-minutes for Problem 1 and approximately  $5n^2/10^6$  CPU-minutes for Problem 2. Most of this CPU-time was spent calculating function values and gradients of  $f_0(x)$ ,

TABLE 8.1  
*Results for Problem 1.*

Number of variables	Objective value	Variables at bounds	$\lambda_1$	$\lambda_2$	Total number of outer iter.	Total number of inner iter.
1000	260.85	184	0.138	0.451	177	209
2000	523.51	353	0.147	0.442	190	224
5000	1312.05	840	0.156	0.431	221	263
10000	2626.76	1629	0.161	0.425	251	296
20000	5256.56	3184	0.165	0.420	286	316

TABLE 8.2  
*Results for Problem 2.*

Number of variables	Objective value	Variables at bounds	$\lambda_1$	$\lambda_2$	Total number of outer iter.	Total number of inner iter.
1000	-739.15	184	0.549	0.862	436	415
2000	-1476.49	353	0.558	0.853	465	471
5000	-3687.95	840	0.569	0.844	584	606
10000	-7373.24	1629	0.575	0.839	682	704
20000	-14743.44	3182	0.580	0.835	793	816

$f_1(x)$ , and  $f_2(x)$ , while only a minor part was spent solving the CCSA subproblems. It should be noted that the matrices  $S$ ,  $P$ , and  $Q$  are never stored. Instead, the elements  $s_{ij}$ ,  $p_{ij}$ , and  $q_{ij}$  are generated as needed when calculating function values and gradients of  $f_0(x)$ ,  $f_1(x)$ , and  $f_2(x)$  at a given iteration point.

It could finally be mentioned that it is virtually impossible to solve the considered problems by, e.g., an SQP method. The approximate Hessian matrix (of the Lagrange function) simply becomes too big.

**9. Conclusions.** A class of optimization methods based on the concept of conservative convex separable approximations has been presented. Global convergence has been theoretically proved, and it has been demonstrated that the methods work numerically.

We do not claim that a CCSA method is always the natural choice, but for certain problems it is certainly a competitive alternative. This is typically the case for problems with a very large number of variables and a relatively small number of general inequality constraints, in particular if it is desirable that the iteration points remain feasible.

Finally, it could be noted that if the considered problem also contains some linear constraints, these can simply be included as (exactly the same) linear constraints in the CCSA subproblems. Since exact approximations are conservative approximations, the global convergence properties of the methods will not be altered.

**Acknowledgments.** I am most grateful to my colleague Anders Forsgren for numerous valuable discussions. I also want to thank the two anonymous referees for their constructive criticism of an early version of the paper and for valuable suggestions on how to improve the presentation.

#### REFERENCES

- [1] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley, New York, 1968.

- [2] R. FLETCHER, *Practical Methods of Optimization*, 2nd ed., John Wiley, Chichester, England, 1987.
- [3] A. FORSGREN AND P. E. GILL, *Primal-dual interior methods for nonconvex nonlinear programming*, SIAM J. Optim., 8 (1998), pp. 1132–1152.
- [4] D. GAY, M. OVERTON, AND M. WRIGHT, *A primal-dual interior method for nonconvex nonlinearly constrained optimization*, in *Advances in Nonlinear Programming*, Y. Yuan, ed., Kluwer, Dordrecht, The Netherlands, 1998, pp. 31–56.
- [5] P. GILL, W. MURRAY, M. SAUNDERS, AND M. WRIGHT, *Constrained nonlinear programming*, in *Optimization, Handbooks Oper. Res. Management Sci. 1*, G. Nemhauser, A. R. Kan, and M. Todd, eds., North-Holland, Amsterdam, 1989, pp. 171–210.
- [6] D. G. LUENBERGER, *Optimization by Vector Space Methods*, John Wiley, New York, 1968.
- [7] K. SVANBERG, *The method of moving asymptotes—a new method for structural optimization*, Internat. J. Numer. Methods Engrg., 24 (1987), pp. 359–373.
- [8] K. SVANBERG, *A globally convergent version of MMA without linesearch*, in *Proceedings of the First World Congress of Structural and Multidisciplinary Optimization*, N. Olhoff and G. Rozvany, eds., Pergamon Press, Elmsford, NY, 1995, pp. 9–16.