

AN ACTIVE-SET NEWTON METHOD FOR MATHEMATICAL PROGRAMS WITH COMPLEMENTARITY CONSTRAINTS*

A. F. IZMAILOV[†] AND M. V. SOLODOV[‡]

Abstract. For a mathematical program with complementarity constraints (MPCC), we propose an active-set Newton method, which has the property of local quadratic convergence under the MPCC linear independence constraint qualification (MPCC-LICQ) and the standard second-order sufficient condition (SOSC) for optimality. Under MPCC-LICQ, this SOSC is equivalent to the piecewise SOSC on branches of MPCC, which is weaker than the special MPCC-SOSC often employed in the literature. The piecewise SOSC is also more natural than MPCC-SOSC because, unlike the latter, it has an appropriate second-order necessary condition as its counterpart. In particular, our assumptions for local quadratic convergence are weaker than those required by standard SQP when applied to MPCC and are equivalent to assumptions required by piecewise SQP for MPCC. Moreover, each iteration of our method consists of solving a linear system of equations instead of a quadratic program. Some globalization issues of the local scheme are also discussed, and illustrative examples and numerical experiments are presented.

Key words. mathematical program with complementarity constraints, active set method, Newton method, SQP, second-order sufficiency, quadratic convergence

AMS subject classifications. 90C30, 90C33, 49M37, 65K10

DOI. 10.1137/070690882

1. Introduction. We consider a *mathematical program with complementarity constraints* (MPCC)

$$(1.1) \quad \min f(x) \quad \text{s.t.} \quad G(x) \geq 0, H(x) \geq 0, \langle G(x), H(x) \rangle \leq 0,$$

where $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is a smooth function and $G, H : \mathbf{R}^n \rightarrow \mathbf{R}^m$ are smooth mappings (twice differentiable and possessing Lipschitzian second derivatives in a neighborhood of the solution of interest). We note that “usual” equality and inequality constraints can be added to our problem setting without any substantial difficulties. We shall consider the case when the problem has only complementarity constraints for the sake of simplicity. Note also that the last constraint in (1.1) could be written as an equality, which is more standard in the complementarity literature. However, it is known that in the context of MPCC, there are good numerical reasons to use the inequality formulation for this constraint. Also, this makes the associated set of Lagrange multipliers smaller, which has both numerical and theoretical advantages. MPCC is perhaps one of the most important instances of a *mathematical program*

*Received by the editors May 8, 2007; accepted for publication (in revised form) May 15, 2008; published electronically October 22, 2008.

<http://www.siam.org/journals/siopt/19-3/69088.html>.

[†]Moscow State University, Faculty of Computational Mathematics and Cybernetics, Department of Operations Research, Leninskiye Gori, GSP-2, 119899 Moscow, Russia (izmaf@ccas.ru). Research of this author is supported by the Russian Foundation for Basic Research grants 06-01-00530, 07-01-00270, 07-01-00416, and 07-01-90102-Mongm, and by Russian Federation President’s Grant NS-693.2008.1 for the support of leading scientific schools. The author also thanks IMPA, where he was a visitor when this work was initiated.

[‡]IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil (solodov@impa.br). Research of this author is supported in part by CNPq grants 301508/2005-4, 490200/2005-2, and 550317/2005-8, by PRONEX–Optimization, and by FAPERJ grant E-26/151.942/2004.

with *equilibrium constraints*, which has recently attracted considerable attention in the optimization literature; see [15, 16].

In order to explain the contribution of this work, some preliminaries from MPCC theory will be needed. To this end, let

$$L(x, \lambda) = f(x) - \langle \lambda_G, G(x) \rangle - \langle \lambda_H, H(x) \rangle + \lambda_0 \langle G(x), H(x) \rangle$$

be the standard Lagrangian of problem (1.1), where $x \in \mathbf{R}^n$ and $\lambda = (\lambda_G, \lambda_H, \lambda_0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$. As for any other mathematical program (MP), stationary points of (1.1) and the associated Lagrange multipliers are characterized by the Karush–Kuhn–Tucker (KKT) optimality system:

$$(1.2) \quad \begin{aligned} \frac{\partial L}{\partial x}(x, \lambda) &= 0, \quad \lambda_0 \geq 0, \quad \langle G(x), H(x) \rangle \leq 0, \\ \lambda_G &\geq 0, \quad G(x) \geq 0, \quad \langle \lambda_G, G(x) \rangle = 0, \quad \lambda_H \geq 0, \quad H(x) \geq 0, \quad \langle \lambda_H, H(x) \rangle = 0. \end{aligned}$$

In the above, we omit the condition $\lambda_0 \langle G(x), H(x) \rangle = 0$, because it is redundant (it follows from $\langle G(x), H(x) \rangle = 0$, which is implied by feasibility of x in (1.1)). For $\bar{x} \in \mathbf{R}^n$, let $\Lambda(\bar{x})$ stand for the set of Lagrange multipliers associated with \bar{x} , that is, the set of $\lambda = \bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H, \bar{\lambda}_0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$ satisfying (1.2) for $x = \bar{x}$. As is well known and can be easily checked, MPCC constraints violate the Mangasarian–Fromovitz constraint qualification and, even more so, the linear independence constraint qualification (LICQ), at every feasible point. Therefore, in general, \bar{x} being a local solution of (1.1) does not guarantee that the set of Lagrange multipliers $\Lambda(\bar{x})$ is nonempty. Nevertheless, $\Lambda(\bar{x})$ happens to be nonempty in many cases of interest, and this became one of the common settings in MPCC literature.

Define further the so-called *MPCC-Lagrangian* of problem (1.1):

$$\mathcal{L}(x, \mu) = f(x) - \langle \mu_G, G(x) \rangle - \langle \mu_H, H(x) \rangle,$$

where $x \in \mathbf{R}^n$ and $\mu = (\mu_G, \mu_H) \in \mathbf{R}^m \times \mathbf{R}^m$. To a feasible point \bar{x} we associate the index sets

$$(1.3) \quad \begin{aligned} I_G &= I_G(\bar{x}) = \{i = 1, \dots, m \mid G_i(\bar{x}) = 0\}, \quad I_H = I_H(\bar{x}) = \{i = 1, \dots, m \mid H_i(\bar{x}) = 0\}, \\ I_0 &= I_G \cap I_H. \end{aligned}$$

A feasible point \bar{x} of (1.1) is said to be a *strongly stationary point* of this problem if there exists an *MPCC-multiplier* $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H) \in \mathbf{R}^m \times \mathbf{R}^m$ satisfying

$$(1.4) \quad \frac{\partial \mathcal{L}}{\partial x}(\bar{x}, \bar{\mu}) = 0, \quad (\bar{\mu}_G)_{I_H \setminus I_G} = 0, \quad (\bar{\mu}_H)_{I_G \setminus I_H} = 0, \quad (\bar{\mu}_G)_{I_0} \geq 0, \quad (\bar{\mu}_H)_{I_0} \geq 0,$$

where y_I stands for the subvector of the vector y , with components $y_i, i \in I$. Without nonnegativity conditions in (1.4), \bar{x} is called a *weakly stationary point* of (1.1).

We say that *MPCC linear independence constraint qualification* (MPCC-LICQ) holds at \bar{x} if the gradients

$$(1.5) \quad G'_i(\bar{x}), \quad i \in I_G, \quad H'_i(\bar{x}), \quad i \in I_H \quad \text{are linearly independent.}$$

It was shown in [18, Theorem 2] that if MPCC-LICQ holds at a local solution \bar{x} of (1.1), then this point is strongly stationary, and the associated MPCC-multiplier $\bar{\mu}$ is unique.

The following proposition summarizes some results obtained in [7, Proposition 4.1] and [9, Proposition 1], which will be used in what follows. Its proof can be obtained by a direct computation. Let

$$(1.6) \quad \bar{\nu} = \max \left\{ 0, \max_{i \in I_G \setminus I_H} \left(-\frac{(\bar{\mu}_G)_i}{H_i(\bar{x})} \right), \max_{i \in I_H \setminus I_G} \left(-\frac{(\bar{\mu}_H)_i}{G_i(\bar{x})} \right) \right\}.$$

PROPOSITION 1.1. *A feasible point \bar{x} of problem (1.1) is a stationary point of this problem if and only if it is a strongly stationary point of this problem. Moreover, if $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H, \bar{\lambda}_0)$ is a Lagrange multiplier associated with \bar{x} , then $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H)$ defined by*

$$(1.7) \quad (\bar{\mu}_G)_i = (\bar{\lambda}_G)_i - \bar{\lambda}_0 H_i(\bar{x}), \quad i \in I_G \setminus I_H, \quad (\bar{\mu}_G)_i = (\bar{\lambda}_G)_i, \quad i \in I_H,$$

$$(1.8) \quad (\bar{\mu}_H)_i = (\bar{\lambda}_H)_i - \bar{\lambda}_0 G_i(\bar{x}), \quad i \in I_H \setminus I_G, \quad (\bar{\mu}_H)_i = (\bar{\lambda}_H)_i, \quad i \in I_G,$$

is an MPCC-multiplier associated with \bar{x} . Conversely, if $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H)$ is an MPCC-multiplier associated with \bar{x} , then any $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H, \bar{\lambda}_0)$ satisfying (1.7)–(1.8) and

$$(1.9) \quad \bar{\lambda}_0 \geq \bar{\nu},$$

with $\bar{\nu}$ defined in (1.6), is a Lagrange multiplier associated with \bar{x} .

Furthermore, for any $\xi \in \mathbf{R}^n$ and any $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H, \bar{\lambda}_0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$ and $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H) \in \mathbf{R}^m \times \mathbf{R}^m$ satisfying (1.7)–(1.8), it holds that

$$(1.10) \quad \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})[\xi, \xi] = \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})[\xi, \xi] + 2\bar{\lambda}_0 \sum_{i=1}^m \langle G'_i(\bar{x}), \xi \rangle \langle H'_i(\bar{x}), \xi \rangle.$$

In particular, if $\bar{\mu}$ is the unique MPCC-multiplier associated with \bar{x} (e.g., under MPCC-LICQ (1.5)), then $\Lambda(\bar{x})$ is the ray defined by (1.7)–(1.9), with its origin corresponding to $\bar{\lambda}_0 = \bar{\nu}$.

It can be easily checked that the standard critical cone of problem (1.1) at \bar{x} is given by

$$(1.11) \quad C(\bar{x}) = \left\{ \xi \in \mathbf{R}^n \mid \begin{array}{l} G'_{I_G \setminus I_H}(\bar{x})\xi = 0, \quad H'_{I_H \setminus I_G}(\bar{x})\xi = 0, \quad G'_{I_0}(\bar{x})\xi \geq 0, \quad H'_{I_0}(\bar{x})\xi \geq 0, \\ \langle f'(\bar{x}), \xi \rangle \leq 0 \end{array} \right\}.$$

We say that *MPCC-second-order sufficient condition* (MPCC-SOSC) holds at a strongly stationary point \bar{x} of problem (1.1), with the associated MPCC-multiplier $\bar{\mu}$, if

$$(1.12) \quad \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})[\xi, \xi] > 0 \quad \forall \xi \in C(\bar{x}) \setminus \{0\}.$$

Note that, for every $\xi \in C(\bar{x})$, we obtain from (1.11) that (1.10) takes the form

$$(1.13) \quad \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})[\xi, \xi] = \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})[\xi, \xi] + 2\bar{\lambda}_0 \sum_{i \in I_0} \langle G'_i(\bar{x}), \xi \rangle \langle H'_i(\bar{x}), \xi \rangle,$$

where the last term in the right-hand side is nonnegative. Thus, according to Proposition 1.1, MPCC-SOSC implies the usual SOSC

$$(1.14) \quad \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})[\xi, \xi] > 0 \quad \forall \xi \in C(\bar{x}) \setminus \{0\}$$

for any $\bar{\lambda}$ satisfying (1.7)–(1.9). In particular, under MPCC-LICQ (1.5), MPCC-SOSC (1.12) (with the unique MPCC-multiplier $\bar{\mu}$) implies SOSC (1.14), with any $\bar{\lambda}$ in the ray $\Lambda(\bar{x})$, including the origin of this ray.

It is important to point out that MPCC-SOSC is a rather strong condition. In particular, it cannot be linked to any second-order necessary condition for (1.1). By this we mean that a solution of (1.1) that satisfies MPCC-LICQ (1.5) (and thus is strongly stationary) does not have to satisfy the condition obtained from (1.12) by replacing the strict inequality by nonstrict. In our developments, we shall be making use of a SOSC weaker than (1.12), which also happens to be much more natural, because it is related to an appropriate second-order necessary condition for (1.1), as explained below.

For each partition (I_1, I_2) of I_0 (i.e., a pair of index sets such that $I_1 \cup I_2 = I_0$, $I_1 \cap I_2 = \emptyset$), define the *branch* (or *piece*) *MP* at \bar{x} by

$$(1.15) \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & G_{(I_G \setminus I_H) \cup I_1}(x) = 0, H_{(I_H \setminus I_G) \cup I_2}(x) = 0, G_{I_2}(x) \geq 0, H_{I_1}(x) \geq 0. \end{array}$$

There is a finite number of such branch MPs, \bar{x} is feasible for each of them, and in a neighborhood of \bar{x} the feasible set of (1.1) is a union of feasible sets of all branch MPs. It is not difficult to see that the union of the critical cones of all branch MPs at \bar{x} is given by

$$(1.16) \quad C_2(\bar{x}) = \left\{ \xi \in \mathbf{R}^n \mid \begin{array}{l} \langle f'(\bar{x}), \xi \rangle \leq 0, G'_{I_G \setminus I_H}(\bar{x})\xi = 0, H'_{I_H \setminus I_G}(\bar{x})\xi = 0, \\ G'_{I_0}(\bar{x})\xi \geq 0, H'_{I_0}(\bar{x})\xi \geq 0, \langle G'_i(\bar{x}), \xi \rangle \langle H'_i(\bar{x}), \xi \rangle = 0, i \in I_0 \end{array} \right\},$$

where the subscript “2” indicates that, unlike $C(\bar{x})$, this set takes into account the second-order information about the last constraint in (1.1). By direct comparison of (1.11) and (1.16), we have that

$$(1.17) \quad C_2(\bar{x}) \subset C(\bar{x}).$$

We say that *piecewise SOSC* holds at a strongly stationary point \bar{x} of problem (1.1), with an associated MPCC-multiplier $\bar{\mu}$, if

$$(1.18) \quad \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})[\xi, \xi] > 0 \quad \forall \xi \in C_2(\bar{x}) \setminus \{0\}.$$

From (1.4), it evidently follows that if $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H)$ is an MPCC-multiplier associated with \bar{x} , then the pair $((\bar{\mu}_G)_{I_G}, (\bar{\mu}_H)_{I_H})$ is a Lagrange multiplier associated with \bar{x} for the branch MP (1.15). It follows that piecewise SOSC (1.18) implies SOSC for each branch at \bar{x} . This, in turn, guarantees that \bar{x} is a strict local solution of (1.1). Thus, piecewise SOSC is indeed sufficient for optimality, even though it is evidently weaker than MPCC-SOSC (see (1.17)).

It is important to emphasize that under MPCC-LICQ (1.5), the condition obtained from (1.18) by replacing the strict inequality by nonstrict is necessary for optimality [18, Theorem 7]. In this sense, piecewise SOSC (1.18) is a more natural assumption than MPCC-SOSC (1.12), as the latter has no relation to any second-order necessary optimality condition.

Suppose that MPCC-LICQ (1.5) and piecewise SOSC (1.18) (with the unique MPCC-multiplier $\bar{\mu}$) hold at a strongly stationary point \bar{x} of problem (1.1). From (1.13) and [9, Proposition 2], it follows that in this case either SOSC (1.14) holds

with all $\bar{\lambda}$ in the ray $\Lambda(\bar{x})$, or possibly there exists $\hat{\nu} \geq \bar{\nu}$ such that SOSC (1.14) does not hold for all $\bar{\lambda}$ corresponding to $\bar{\lambda}_0 \in [\bar{\nu}, \hat{\nu}]$, and holds for all $\bar{\lambda}$ corresponding to $\bar{\lambda}_0 > \hat{\nu}$. Conversely, if SOSC (1.14) holds for some $\bar{\lambda} \in \Lambda(\bar{x})$, from (1.13) and (1.16), taking also into account (1.17), it is easy to see that piecewise SOSC (1.18) holds as well. Thus, under MPCC-LICQ, SOSC (with some multiplier) is equivalent to piecewise SOSC.

Despite the inevitable violation of standard constraint qualifications, there exists some numerical evidence of good performance of sequential quadratic programming (SQP) algorithms for MPCCs (see [6]). Moreover, [7] gives some theoretical justification for local superlinear convergence of the SQP algorithm for MPCC under a set of assumptions that includes MPCC-LICQ and MPCC-SOSC, among other things. However, it is very easy to provide examples satisfying all *natural* in MPCC context requirements (say, MPCC-LICQ and piecewise SOSC), and such that SQP does not possess superlinear convergence; see, e.g., the example in [7, section 7.3], discussed also in detail in [11, section 6]. This means that the existing evidence supporting the use of standard optimization algorithms (say, SQP) for MPCC cannot be regarded as completely satisfactory, and it still makes sense to develop special algorithms which take into account special structure of MPCC, and which are guaranteed to achieve quadratic convergence under more natural assumptions.

Let us recall now the main idea of the piecewise SQP algorithm, suggested originally in [17] for MPs with linear complementarity constraints and then extended in [15] to the nonlinear case. An iteration of piecewise SQP is organized as follows: identify *any* branch MP valid at the solution \bar{x} that is being approximated, and perform a step of standard SQP for this branch. In order to identify a valid branch MP, it suffices to (over)estimate the sets $I_G \setminus I_H$ and $I_H \setminus I_G$ (see (1.15)). Locally, this comes for free, with no significant computational cost and with no assumptions needed. However, in order to justify the overall superlinear convergence of piecewise SQP, one needs to guarantee superlinear convergence of SQP for each branch, and dual convergence to the same multiplier for all branches. This results in the following set of assumptions: MPCC-LICQ (1.5) and piecewise SOSC (1.18) at the solution \bar{x} .

In this paper, we suggest a local algorithm based on the following idea (to some extent motivated by the development in [13]). Instead of an arbitrary valid branch, we identify the index sets I_G and I_H and perform the Newton–Lagrange steps for the following purely equality-constrained *tightened MP*:

$$(1.19) \quad \min f(x) \quad \text{s.t.} \quad G_{I_G}(x) = 0, \quad H_{I_H}(x) = 0.$$

Note that this problem is not a branch MP, in general, but its feasible set is contained in the feasible sets of all branch MPs.

For quadratic convergence of the Newton–Lagrange method for (1.19), we need to assume MPCC-LICQ (1.5) and SOSC for this problem, the latter being evidently guaranteed by piecewise SOSC (1.18). Local identification of I_G and I_H uses the procedure suggested in [4] and the error bound following from [8, Lemma 2] and [5, Theorem 2] (see (2.6)). The identification technique based on this combination of tools (first used for problems without any regularity assumptions on constraints in [10]) still costs nothing computationally. The error bound requires some $\bar{\lambda} \in \Lambda(\bar{x})$ satisfying SOSC (1.14). According to our discussion above, the existence of such $\bar{\lambda}$ can again be guaranteed under MPCC-LICQ (1.5) and piecewise SOSC (1.18). Hence, we obtain local quadratic convergence of our algorithm under the same set of assumptions as for piecewise SQP: MPCC-LICQ (1.5) and piecewise SOSC (1.18) at \bar{x} . At the same

time, our local algorithm enjoys the advantage of being quadratic program (QP)-free: it requires solving only one linear system per iteration. Of course, within a local framework, this may not always be a big advantage. Note, however, that globalized Algorithm 3.2 in section 3.1 is QP-free globally.

2. Local algorithms. As is well known, the KKT system (1.2) can be written in the form

$$\Phi(x, \lambda) = 0,$$

where $\Phi : \mathbf{R}^n \times (\mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}) \rightarrow \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$,

$$\Phi(x, \lambda) = \left(\frac{\partial L}{\partial x}(x, \lambda), \rho(\lambda_G, G(x)), \rho(\lambda_H, H(x)), \rho(\lambda_0, -\langle G(x), H(x) \rangle) \right),$$

and $\rho : \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$ is a complementarity function (that is, a function such that $\rho(a, b) = 0 \iff a \geq 0, b \geq 0, ab = 0$) applied componentwise. In what follows, we shall make use of two complementarity functions, namely, the natural residual $\rho(a, b) = \min\{a, b\}$ and the Fischer–Burmeister function $\rho(a, b) = \sqrt{a^2 + b^2} - a - b$. The corresponding version of Φ will be denoted by Φ_{NR} and Φ_{FB} , respectively. As is well known, both these mappings are semismooth (and in particular, locally Lipschitz). Moreover, according to [19], these two complementarity functions are equivalent in terms of their growth rates. This means that, throughout the paper, Φ_{NR} can actually be replaced by Φ_{FB} without any changes in the analysis or results.

ALGORITHM 2.1. Preliminary step. Fix $\theta \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n, \lambda^0 = (\lambda_G^0, \lambda_H^0, \lambda_0^0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$.

Identification step. Compute the index sets

$$(2.1) \quad I_G = I_G(x^0, \lambda^0) = \{i = 1, \dots, m \mid G_i(x^0) \leq \|\Phi_{NR}(x^0, \lambda^0)\|^\theta\},$$

$$(2.2) \quad I_H = I_H(x^0, \lambda^0) = \{i = 1, \dots, m \mid H_i(x^0) \leq \|\Phi_{NR}(x^0, \lambda^0)\|^\theta\}.$$

Main step. Generate the sequence $\{(x^k, \mu^k)\}$, with $\mu^k = (\mu_G^k, \mu_H^k) \in \mathbf{R}^m \times \mathbf{R}^m$, as follows.

- Generate the sequence $\{(x^k, (\mu_G^k)_{I_G}, (\mu_H^k)_{I_H})\}$ by the Newton–Lagrange method for tightened MP (1.19) (that is, the Newton method applied to the Lagrange optimality system of this problem) starting from $(x^0, (\mu_G^0)_{I_G}, (\mu_H^0)_{I_H})$, with $(\mu_G^0)_{I_G}$ and $(\mu_H^0)_{I_H}$ defined by

$$(2.3) \quad (\mu_G^0)_i = (\lambda_G^0)_i - \lambda_0^0 H_i(x^0), \quad i \in I_G \setminus I_H, \quad (\mu_G^0)_i = (\lambda_G^0)_i, \quad i \in I_G \cap I_H,$$

$$(2.4) \quad (\mu_H^0)_i = (\lambda_H^0)_i - \lambda_0^0 G_i(x^0), \quad i \in I_H \setminus I_G, \quad (\mu_H^0)_i = (\lambda_H^0)_i, \quad i \in I_G \cap I_H.$$

- Set

$$(2.5) \quad (\mu_G^k)_{I_H \setminus I_G} = 0, \quad (\mu_H^k)_{I_G \setminus I_H} = 0 \quad \forall k = 0, 1, \dots$$

THEOREM 2.1. *Let \bar{x} be a local solution of MPCC (1.1), and assume that MPCC-LICQ (1.5) holds at \bar{x} . Furthermore, let $\bar{\mu}$ be the (unique) MPCC-multiplier associated with \bar{x} , and suppose that (x^0, λ^0) is close enough to $(\bar{x}, \bar{\lambda})$, with some $\bar{\lambda} \in \Lambda(\bar{x})$ satisfying SOSC (1.14).*

Then Algorithm 2.1 correctly generates the sequence $\{(x^k, \mu^k)\}$, which converges quadratically to $(\bar{x}, \bar{\mu})$.

Proof. According to [8, Lemma 2] and [5, Theorem 2], SOSC (1.14) implies the existence of $c > 0$ such that the error bound

$$(2.6) \quad \text{dist}((x, \lambda), \{\bar{x}\} \times \Lambda(\bar{x})) \leq c \|\Phi_{NR}(x, \lambda)\|$$

holds for all $(x, \lambda) \in \mathbf{R}^n \times (\mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R})$ close enough to $(\bar{x}, \bar{\lambda})$. Since (x^0, λ^0) is close enough to $(\bar{x}, \bar{\lambda})$, from [4, Theorem 2.2] it follows that the index sets $I_G = I_G(x^0, \lambda^0)$ and $I_H = I_H(x^0, \lambda^0)$, computed according to (2.1) and (2.2), coincide with $I_G = I_G(\bar{x})$ and $I_H = I_H(\bar{x})$, respectively, defined in (1.3).

Furthermore, the point \bar{x} is a local solution of tightened MP (1.19), and MPCC-LICQ (1.5) means that LICQ holds at \bar{x} for the constraints of (1.19). In particular, \bar{x} is a stationary point of (1.19), and from (1.4) it evidently follows that $((\bar{\mu}_G)_{I_G}, (\bar{\mu}_H)_{I_H})$ is the unique Lagrange multiplier associated with this stationary point.

Stationarity of \bar{x} in (1.19) evidently implies that

$$\langle f'(\bar{x}), \xi \rangle = 0 \quad \forall \xi \in \ker G'_{I_G}(\bar{x}) \cap \ker H'_{I_H}(\bar{x}),$$

where $\ker A$ stands for the kernel (null space) of a linear operator A . Hence, by (1.11),

$$\ker G'_{I_G}(\bar{x}) \cap \ker H'_{I_H}(\bar{x}) \subset C(\bar{x}).$$

From Proposition 1.1 (see (1.10)) and SOSC (1.14), it now follows that

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})[\xi, \xi] = \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})[\xi, \xi] > 0 \quad \forall \xi \in (\ker G'_{I_G}(\bar{x}) \cap \ker H'_{I_H}(\bar{x})) \setminus \{0\},$$

and according to the equalities $(\bar{\mu}_G)_{I_H \setminus I_G} = 0$ and $(\bar{\mu}_H)_{I_G \setminus I_H} = 0$ in (1.4), the latter means that SOSC holds at \bar{x} for tightened MP (1.19) (with the unique associated multiplier $((\bar{\mu}_G)_{I_G}, (\bar{\mu}_H)_{I_H})$).

Finally, since (x^0, λ^0) is close enough to $(\bar{x}, \bar{\lambda})$, the pair $((\mu_G^0)_{I_G}, (\mu_H^0)_{I_H})$ defined by (2.3)–(2.4) will be close enough to $((\bar{\mu}_G)_{I_G}, (\bar{\mu}_H)_{I_H})$ (recall that, according to Proposition 1.1, the latter satisfies (1.7)–(1.8)). From the standard convergence result for the Newton–Lagrange method, it now follows that the sequence $\{(x^k, (\mu_G^k)_{I_G}, (\mu_H^k)_{I_H})\}$ is correctly defined and converges quadratically to the point $(\bar{x}, (\bar{\mu}_G)_{I_G}, (\bar{\mu}_H)_{I_H})$. At the same time, according to (1.4) and (2.5), it holds that

$$(\mu_G^k)_{I_H \setminus I_G} = (\bar{\mu}_G)_{I_H \setminus I_G} = 0, \quad (\mu_H^k)_{I_G \setminus I_H} = (\bar{\mu}_H)_{I_G \setminus I_H} = 0 \quad \forall k = 0, 1, \dots$$

This completes the proof. \square

Let us discuss briefly the assumptions of Theorem 2.1. These assumptions are, in a sense, “minimal.” In particular, none of them can be removed, as illustrated next.

MPCC-LICQ (1.5) is needed for nondegeneracy of constraints of tightened MP (1.19), which is clearly necessary for the approach to be valid; otherwise, the linearized constraints of tightened MP can be inconsistent arbitrarily close to a solution. To this end, consider, e.g., $n = 2, m = 1, f(x) = x_2^2/2, G(x) = x_1 + x_1^2/2$, and $H(x) = x_1 + x_1^2$. Then $\bar{x} = 0$ is a strongly stationary point of (1.1) satisfying SOSC (1.14) but violating MPCC-LICQ (1.5). It is easily seen that linearization of tightened MP (1.19) at any $x \in \mathbf{R}^2$ with $x_1 \neq 0$ is inconsistent.

The role of SOSC (1.14) is twofold. First, it is used for identification of the index sets I_G and I_H . To see that without SOSC identification can be incorrect, consider $n = 2, m = 1, f(x) = x_1, G(x) = x_1$, and $H(x) = x_2$. Then $\bar{x} = 0$ is a solution of (1.1) satisfying MPCC-LICQ (1.5) but violating SOSC (1.14) (one can

even add, e.g., $|x_2|^3$ to the objective function in order to make this solution strict). Take $\lambda^0 = (1, 0, 0) \in \Lambda(\bar{x})$, and let $x_1^0 \geq 0, x_2^0 \geq 0$. Then $\|\Phi_{NR}(x^0, \lambda^0)\| = x_1^0$, and for any fixed $\theta \in (0, 1)$, by taking $x_2^0 = (x_1^0)^{\theta/2}$, we obtain a point $x^0 = (x_1^0, x_2^0)$ which can be arbitrarily close to \bar{x} , while (2.2) will always (incorrectly) identify I_H as empty at such point.

Finally, even if the identification is correct, SOSC (1.14) is still needed as it guarantees SOSC for tightened MP (1.19). Let, e.g., $n = 2, m = 1, f(x) = x_1 + |x_2|^3, G(x) = x_1, \text{ and } H(x) \equiv 1$. Then $\bar{x} = 0$ is a solution satisfying MPCC-LICQ (1.5) but violating SOSC (1.14). Moreover, tightened MP is also violating SOSC. It can be checked directly that the convergence rate of the Newton–Lagrange method for tightened MP (1.19) is only linear.

Note that the presented algorithm appears more suitable for globalization than, say, piecewise SQP. This is because Algorithm 2.1 uses as a dual starting point an approximation of Lagrange multiplier rather than an approximation of MPCC-multiplier. The proximity to points satisfying KKT system (1.2) (and hence, to Lagrange multipliers) can be controlled via some globally defined merit functions (like the norm of Φ_{NR} or Φ_{FB}). By contrast, the definition (1.4) of MPCC-multipliers involves the index sets I_G and I_H depending on a specific \bar{x} , and it seems difficult to suggest a reasonable globally defined merit function characterizing MPCC-multipliers.

Furthermore, having in mind globalization of convergence, it can be useful to consider a modified algorithm, with **Identification step** being performed not only once (at the beginning of the process) but before each iteration of **Main step**. Identification is a very cheap procedure and, therefore, this modification will not increase computational costs significantly. However, in this case we will need to generate not only the sequence $\{(x^k, \mu^k)\}$ but also an appropriate sequence $\{\lambda^k\} \subset \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$, and redefine I_G and I_H accordingly:

$$(2.7) \quad I_G = I_G(x^k, \lambda^k) = \{i = 1, \dots, m \mid G_i(x^k) \leq \|\Phi_{NR}(x^k, \lambda^k)\|^\theta\},$$

$$(2.8) \quad I_H = I_H(x^k, \lambda^k) = \{i = 1, \dots, m \mid H_i(x^k) \leq \|\Phi_{NR}(x^k, \lambda^k)\|^\theta\}$$

for each $k = 0, 1, \dots$. Clearly, for all the conclusions of Theorem 2.1 to remain valid for this modified algorithm, it suffices to show that $\{\lambda^k\}$ stays close to $\bar{\lambda}$, which can be achieved by keeping it close to λ^0 . In particular, one can just take $\lambda^k = \lambda^0 \forall k = 1, 2, \dots$. Another option, which seems more suitable for globalization purposes (and which is more in the spirit of SQP methods), is realized in the following method.

ALGORITHM 2.2. Preliminary step. Fix $\theta \in (0, 1)$. Set $k = 0$ and choose $x^0 \in \mathbf{R}^n$ and $\lambda^0 = (\lambda_G^0, \lambda_H^0, \lambda_0^0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$.

Identification step. Define the index sets I_G and I_H according to (2.7) and (2.8). If $k = 0$, define $(\mu_G^0)_{I_G}$ and $(\mu_H^0)_{I_H}$ according to (2.3)–(2.4).

Main step. Compute (x^{k+1}, μ^{k+1}) as follows.

- The triple $(x^{k+1}, (\mu_G^{k+1})_{I_G}, (\mu_H^{k+1})_{I_H})$ is generated by the step of Newton–Lagrange method for tightened MP (1.19) from the point $(x^k, (\mu_G^k)_{I_G}, (\mu_H^k)_{I_H})$.
- Set $(\mu_G^{k+1})_{I_H \setminus I_G} = 0, (\mu_H^{k+1})_{I_G \setminus I_H} = 0$.

Set

$$(2.9) \quad \nu_{k+1} = \max \left\{ 0, \max_{i \in I_G \setminus I_H} \left(-\frac{(\mu_G^{k+1})_i}{H_i(x^{k+1})} \right), \max_{i \in I_H \setminus I_G} \left(-\frac{(\mu_H^{k+1})_i}{G_i(x^{k+1})} \right) \right\},$$

and define $\lambda^{k+1} = (\lambda_G^{k+1}, \lambda_H^{k+1}, \lambda_0^{k+1})$ as follows:

$$(2.10) \quad \lambda_0^{k+1} = \max \{ \nu_{k+1}, \lambda_0^k \},$$

$$(2.11) \quad (\lambda_G^{k+1})_i = (\mu_G^{k+1})_i + \lambda_0^{k+1} H_i(x^{k+1}), i \in I_G \setminus I_H, \quad (\lambda_G^{k+1})_i = (\mu_G^{k+1})_i, i \in I_H,$$

$$(2.12) \quad (\lambda_H^{k+1})_i = (\mu_H^{k+1})_i + \lambda_0^{k+1} G_i(x^{k+1}), i \in I_H \setminus I_G, \quad (\lambda_H^{k+1})_i = (\mu_H^{k+1})_i, i \in I_G.$$

Adjust k by 1 and go to **Identification step**.

For purposes of convergence analysis, we need to introduce some auxiliary dual estimates. Suppose that, for some $k = 0, 1, \dots$, Algorithm 2.2 correctly defined x^k , λ^k , and μ^k . Define $\hat{\lambda}^k = (\hat{\lambda}_G^k, \hat{\lambda}_H^k, \hat{\lambda}_0^k) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$ as follows:

$$(2.13) \quad \hat{\lambda}_0^k = \lambda_0^k,$$

$$(2.14) \quad (\hat{\lambda}_G^k)_i = (\bar{\mu}_G)_i + \lambda_0^k H_i(\bar{x}), i \in I_G \setminus I_H, \quad (\hat{\lambda}_G^k)_i = (\bar{\mu}_G)_i, i \in I_H,$$

$$(2.15) \quad (\hat{\lambda}_H^k)_i = (\bar{\mu}_H)_i + \lambda_0^k G_i(\bar{x}), i \in I_H \setminus I_G, \quad (\hat{\lambda}_H^k)_i = (\bar{\mu}_H)_i, i \in I_G.$$

According to (1.7)–(1.8) and (2.13)–(2.15), it holds that

$$(2.16) \quad \|\hat{\lambda}^k - \bar{\lambda}\| \leq \left(1 + \max \left\{ \max_{i \in I_G \setminus I_H} H_i(\bar{x}), \max_{i \in I_H \setminus I_G} G_i(\bar{x}) \right\}\right) |\lambda_0^k - \bar{\lambda}_0|,$$

and hence, $\hat{\lambda}^k$ is close to $\bar{\lambda}$ provided (x^k, λ^k) is close enough to $(\bar{x}, \bar{\lambda})$.

From Theorem 2.1, we obtain that if (x^k, λ^k) is close enough to $(\bar{x}, \bar{\lambda})$, then the points x^{k+1} and μ^{k+1} will be correctly defined by Algorithm 2.2, and

$$(2.17) \quad \|x^{k+1} - \bar{x}\| = O\left(\|(x^k - \bar{x}, \mu^k - \bar{\mu})\|^2\right) = O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right),$$

$$(2.18) \quad \|\mu^{k+1} - \bar{\mu}\| = O\left(\|(x^k - \bar{x}, \mu^k - \bar{\mu})\|^2\right) = O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right).$$

Furthermore, according to (1.6)–(1.8), (2.9), (2.11)–(2.15), (2.17), and (2.18), we obtain the estimates

$$(2.19) \quad |\nu_{k+1} - \bar{\nu}| = O\left(\|x^{k+1} - \bar{x}\| + O\left(\|\mu^{k+1} - \bar{\mu}\|\right)\right) = O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right),$$

$$(2.20) \quad \begin{aligned} \|\lambda^{k+1} - \hat{\lambda}^k\| &= O\left(\|x^{k+1} - \bar{x}\| + O\left(\|\mu^{k+1} - \bar{\mu}\|\right) + O\left(|\lambda_0^{k+1} - \lambda_0^k|\right)\right) \\ &= O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right) + O\left(|\lambda_0^{k+1} - \lambda_0^k|\right), \end{aligned}$$

$$(2.21) \quad \begin{aligned} \|\lambda^{k+1} - \bar{\lambda}\| &= O\left(\|x^{k+1} - \bar{x}\| + O\left(\|\mu^{k+1} - \bar{\mu}\|\right) + O\left(|\lambda_0^{k+1} - \bar{\lambda}_0|\right)\right) \\ &= O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right) + O\left(|\lambda_0^{k+1} - \bar{\lambda}_0|\right). \end{aligned}$$

Let us consider separately the two cases

$$(2.22) \quad \bar{\lambda}_0 > \bar{\nu} \quad \text{or} \quad \bar{\lambda}_0 = \bar{\nu}$$

(see (1.9)).

LEMMA 2.2. *Let \bar{x} be a local solution of MPCC (1.1), and assume that MPCC-LICQ (1.5) holds at \bar{x} . Furthermore, suppose that, for some $k = 0, 1, \dots$, Algorithm 2.2 generated points x^k, λ^k , and μ^k such that (x^k, λ^k) is close enough to $(\bar{x}, \bar{\lambda})$, with some $\bar{\lambda} \in \Lambda(\bar{x})$ satisfying SOSC (1.14).*

Then the points x^{k+1} , λ^{k+1} , and μ^{k+1} will be correctly generated by Algorithm 2.2, and if the first relation in (2.22) holds, with $\bar{\nu}$ defined according to (1.6), then

$$(2.23) \quad \lambda_0^{k+1} = \lambda_0^k,$$

$$(2.24) \quad \|\lambda^{k+1} - \hat{\lambda}^k\| = O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right).$$

Proof. By the first relation in (2.22), we obtain that if (x^k, λ^k) is close enough to $(\bar{x}, \bar{\lambda})$, then

$$\lambda_0^k > \bar{\lambda}_0 - (\bar{\lambda}_0 - \bar{\nu})/2 = (\bar{\nu} + \bar{\lambda}_0)/2, \quad \nu_{k+1} < \bar{\nu} + (\bar{\lambda}_0 - \bar{\nu})/2 = (\bar{\nu} + \bar{\lambda}_0)/2,$$

with inequality in the last relation being implied by (2.16) and (2.19). Thus $\lambda_0^k > \nu_{k+1}$, and by (2.10), we obtain (2.23). Estimate (2.24) follows from (2.20) and (2.23). \square

LEMMA 2.3. *Under the assumptions of Lemma 2.2, if the second relation in (2.22) holds, with $\bar{\nu}$ defined according to (1.6), then the following estimates are valid.*

If $\lambda_0^{k+1} = \nu_{k+1}$, then

$$(2.25) \quad \|\lambda^{k+1} - \bar{\lambda}\| = O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right),$$

$$(2.26) \quad \|\hat{\lambda}^{k+1} - \bar{\lambda}\| = O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right).$$

If $\lambda_0^{k+1} = \lambda_0^k$, then (2.24) holds.

Proof. If $\lambda_0^{k+1} = \nu_{k+1}$, then by (2.16), (2.19), (2.21), and the second relation in (2.22), we obtain the estimates

$$\begin{aligned} \|\lambda^{k+1} - \bar{\lambda}\| &= O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right) + O(|\lambda_0^{k+1} - \bar{\lambda}_0|) \\ &= O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right) + O(|\nu_{k+1} - \bar{\nu}|) = O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right), \\ \|\hat{\lambda}^{k+1} - \bar{\lambda}\| &= O(|\lambda_0^{k+1} - \bar{\lambda}_0|) = O(|\nu_{k+1} - \bar{\nu}|) = O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^2\right). \end{aligned}$$

This proves (2.25) and (2.26).

If $\lambda_0^{k+1} = \lambda_0^k$, then estimate (2.24) follows immediately from (2.20). \square

We are now in position to prove convergence of Algorithm 2.2.

THEOREM 2.4. *Under the assumptions of Theorem 2.1, Algorithm 2.2 correctly generates the sequence $\{(x^k, \lambda^k, \mu^k)\}$ such that $\{(x^k, \mu^k)\}$ converges quadratically to $(\bar{x}, \bar{\mu})$. Moreover, if the first relation in (2.22) holds, then the sequence $\{(x^k, \lambda^k)\}$ converges quadratically to $(\bar{x}, \hat{\lambda}^0)$, with $\hat{\lambda}^0$ defined according to (2.13)–(2.15), and $\hat{\lambda}^0 \in \Lambda(\bar{x})$.*

Proof. If the first relation in (2.22) holds, then employing (2.13)–(2.15), (2.16), (2.17), and Lemma 2.2 (see (2.23) and (2.24)), it can be shown (by standard argument) that if (x^0, λ^0) is close enough to $(\bar{x}, \bar{\lambda})$, then each further step of Algorithm 2.2 will produce a pair (x^{k+1}, λ^{k+1}) , with $\lambda_0^{k+1} = \lambda_0^k = \lambda_0^0$, and this new pair will be close to $(\bar{x}, \hat{\lambda}^k) = (\bar{x}, \hat{\lambda}^0)$, which in turn is close to $(\bar{x}, \bar{\lambda})$. Then by the same argument as in the proof of Theorem 2.1, for any k , the index sets $I_G(x^k, \lambda^k)$ and $I_H(x^k, \lambda^k)$ computed according to (2.7) and (2.8) will coincide with $I_G = I_G(\bar{x})$ and $I_H = I_H(\bar{x})$ defined in (1.3), respectively. This means that Algorithm 2.2 generates

exactly the same trajectory $\{(x^k, \mu^k)\}$ as Algorithm 2.1, and quadratic convergence follows now from Theorem 2.1. Furthermore, quadratic convergence of $\{(x^k, \lambda^k)\}$ to $(\bar{x}, \hat{\lambda}^0)$ follows from (2.17), (2.24), and the above-established equality $\hat{\lambda}^k = \hat{\lambda}^0 \forall k$. Finally, $\hat{\lambda}^0 \in \Lambda(\bar{x})$ according to (2.13)–(2.15), the first relation in (2.22), and Proposition 1.1.

We proceed with the case when the second relation in (2.22) holds. Again we need to show that if (x^0, λ^0) is close enough to $(\bar{x}, \bar{\lambda})$, then $\{(x^k, \lambda^k)\}$ stays close to $(\bar{x}, \bar{\lambda})$. Then the needed assertion will follow the same way as for the previous case.

From (2.16)–(2.17) and (2.24)–(2.26), it follows that, for any $q \in (0, 1/2]$, there exists $\varepsilon > 0$ such that for all (x^k, λ^k) satisfying $\|x^k - \bar{x}\| < \varepsilon$ and $\|\lambda^k - \bar{\lambda}\| < \varepsilon$ the following estimates are valid.

If $\lambda_0^{k+1} = \nu_{k+1}$, then

$$(2.27) \quad \|(x^{k+1} - \bar{x}, \lambda^{k+1} - \bar{\lambda})\| \leq q \|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|,$$

$$(2.28) \quad \|(x^{k+1} - \bar{x}, \hat{\lambda}^{k+1} - \bar{\lambda})\| \leq q \|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|.$$

If $\lambda_0^{k+1} = \lambda_0^k$, then

$$(2.29) \quad \|(x^{k+1} - \bar{x}, \lambda^{k+1} - \hat{\lambda}^k)\| \leq q \|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|.$$

Let (x^0, λ^0) be close enough to $(\bar{x}, \bar{\lambda})$, so that

$$(2.30) \quad \|(x^0 - \bar{x}, \lambda^0 - \hat{\lambda}^0)\| < \delta, \quad \|(x^0 - \bar{x}, \lambda^0 - \bar{\lambda})\| < \delta,$$

where $\delta > 0$ satisfies the inequality

$$(2.31) \quad (q + 1)\delta \leq \varepsilon$$

(see (2.16)). We now prove by induction that $\forall k = 1, 2, \dots$, it holds that

$$(2.32) \quad \|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\| < \delta,$$

$$(2.33) \quad \|(x^k - \bar{x}, \hat{\lambda}^k - \bar{\lambda})\| < \delta,$$

$$(2.34) \quad \|(x^k - \bar{x}, \lambda^k - \bar{\lambda})\| < \varepsilon.$$

Let $k = 1$. If $\lambda_0^1 = \nu_1$, then by (2.27), (2.30), and (2.31), we obtain

$$(2.35) \quad \|(x^1 - \bar{x}, \lambda^1 - \bar{\lambda})\| \leq q \|(x^0 - \bar{x}, \lambda^0 - \hat{\lambda}^0)\| < q\delta < \varepsilon,$$

i.e., (2.34) holds for $k = 1$. Furthermore, by (2.28), (2.30), and by the inequality $q < 1$,

$$(2.36) \quad \|(x^1 - \bar{x}, \hat{\lambda}^1 - \bar{\lambda})\| \leq q \|(x^0 - \bar{x}, \lambda^0 - \hat{\lambda}^0)\| < q\delta < \delta,$$

i.e., (2.33) holds for $k = 1$. Finally, by (2.35), (2.36), and by the inequality $q \leq 1/2$,

$$\|(x^1 - \bar{x}, \lambda^1 - \hat{\lambda}^1)\| \leq \|(x^1 - \bar{x}, \lambda^1 - \bar{\lambda})\| + \|\hat{\lambda}^1 - \bar{\lambda}\| < 2q\delta \leq \delta,$$

i.e., (2.32) holds for $k = 1$.

On the other hand, if $\lambda_0^1 = \lambda_0^0$, then by (2.13)–(2.15) $\hat{\lambda}^1 = \hat{\lambda}^0$, and by (2.29), (2.30),

$$\left\| \left(x^1 - \bar{x}, \lambda^1 - \hat{\lambda}^0 \right) \right\| \leq q \left\| \left(x^0 - \bar{x}, \lambda^0 - \hat{\lambda}^0 \right) \right\| < q\delta,$$

and hence, by the inequality $q < 1$, we have that

$$(2.37) \quad \left\| \left(x^1 - \bar{x}, \lambda^1 - \hat{\lambda}^1 \right) \right\| = \left\| \left(x^1 - \bar{x}, \lambda^1 - \hat{\lambda}^0 \right) \right\| < q\delta < \delta,$$

i.e., (2.32) holds for $k = 1$. Furthermore, by (2.30), we have that

$$\left\| \left(x^1 - \bar{x}, \hat{\lambda}^1 - \bar{\lambda} \right) \right\| = \left\| \left(x^1 - \bar{x}, \hat{\lambda}^0 - \bar{\lambda} \right) \right\| < \delta,$$

i.e., (2.33) holds for $k = 1$. Finally, by (2.31), (2.33) for $k = 1$, and (2.37), we obtain

$$\left\| \left(x^1 - \bar{x}, \lambda^1 - \bar{\lambda} \right) \right\| \leq \left\| \left(x^1 - \bar{x}, \lambda^1 - \hat{\lambda}^1 \right) \right\| + \left\| \hat{\lambda}^1 - \bar{\lambda} \right\| < q\delta + \delta \leq \varepsilon,$$

i.e., (2.34) holds for $k = 1$.

Now suppose that the hypothesis is valid for $k = s$. If $\lambda_0^{s+1} = \nu_{s+1}$, then by (2.27), (2.30), and (2.31), we obtain that

$$(2.38) \quad \left\| \left(x^{s+1} - \bar{x}, \lambda^{s+1} - \bar{\lambda} \right) \right\| \leq q \left\| \left(x^s - \bar{x}, \lambda^s - \hat{\lambda}^s \right) \right\| < q\delta < \varepsilon,$$

i.e., (2.34) holds for $k = s + 1$. Furthermore, by (2.28), (2.30), and by the inequality $q < 1$,

$$(2.39) \quad \left\| \left(x^{s+1} - \bar{x}, \hat{\lambda}^{s+1} - \bar{\lambda} \right) \right\| \leq q \left\| \left(x^s - \bar{x}, \lambda^s - \hat{\lambda}^s \right) \right\| < q\delta < \delta,$$

i.e., (2.33) holds for $k = s + 1$. Finally, by (2.38), (2.39), and by the inequality $q \leq 1/2$,

$$\left\| \left(x^{s+1} - \bar{x}, \lambda^{s+1} - \hat{\lambda}^{s+1} \right) \right\| \leq \left\| \left(x^{s+1} - \bar{x}, \lambda^{s+1} - \bar{\lambda} \right) \right\| + \left\| \hat{\lambda}^{s+1} - \bar{\lambda} \right\| < 2q\delta \leq \delta,$$

i.e., (2.32) holds for $k = s + 1$.

On the other hand, if $\lambda_0^{s+1} = \lambda_0^s$, then by (2.13)–(2.15) $\hat{\lambda}^{s+1} = \hat{\lambda}^s$, and by (2.29), (2.30), we have that

$$\left\| \left(x^{s+1} - \bar{x}, \lambda^{s+1} - \hat{\lambda}^s \right) \right\| \leq q \left\| \left(x^s - \bar{x}, \lambda^s - \hat{\lambda}^s \right) \right\| < q\delta,$$

and hence, by the inequality $q < 1$,

$$(2.40) \quad \left\| \left(x^{s+1} - \bar{x}, \lambda^{s+1} - \hat{\lambda}^{s+1} \right) \right\| = \left\| \left(x^{s+1} - \bar{x}, \lambda^{s+1} - \hat{\lambda}^s \right) \right\| < q\delta < \delta,$$

i.e., (2.32) holds for $k = s + 1$. Furthermore, by (2.30), we obtain

$$\left\| \left(x^{s+1} - \bar{x}, \hat{\lambda}^{s+1} - \bar{\lambda} \right) \right\| = \left\| \left(x^{s+1} - \bar{x}, \hat{\lambda}^s - \bar{\lambda} \right) \right\| < \delta,$$

i.e., (2.33) holds for $k = s + 1$. Finally, by (2.31), (2.33) for $k = s + 1$, and (2.40), we derive that

$$\left\| \left(x^{s+1} - \bar{x}, \lambda^{s+1} - \bar{\lambda} \right) \right\| \leq \left\| \left(x^{s+1} - \bar{x}, \lambda^{s+1} - \hat{\lambda}^{s+1} \right) \right\| + \left\| \hat{\lambda}^{s+1} - \bar{\lambda} \right\| < q\delta + \delta \leq \varepsilon,$$

i.e., (2.34) holds for $k = s + 1$. This completes the proof by induction. \square

3. Globalization issues. In this section, we discuss some possible ways of globalizing the local scheme presented above. The first approach is based on a generic outer phase steering the iterates toward stationary points. This globalization uses a test of linear decrease for the KKT residual to decide when active-set steps are successful. We also give a specific implementation of this approach along the lines of hybrid semismooth Newton methods for mixed complementarity problems, for which both global convergence and superlinear rate of convergence can be formally proved under reasonable assumptions. The second approach below is based on SQP. It is therefore quite close in spirit to existing algorithms, and can be easily incorporated into them. However, this method may converge to weakly (i.e., not only strongly) stationary points. We do not provide a formal convergence analysis for this method. The reason is that such analysis would primarily concern the study of global convergence properties of standard linesearch SQP algorithms for MPCCs, which is a general issue not related specifically to local algorithms suggested above.

3.1. Hybrid globalization. We next show how our local algorithm can be embedded into any globally convergent scheme. By this we mean that having chosen and fixed some outer-phase global strategy which is guaranteed to produce primal-dual iterates converging to stationary (in some sense) points of MPCC (1.1), the role of our local method is to force quadratic convergence rate under natural assumptions stated above. The key to this construction is the proof that close to a solution with stated properties, the Newton–Lagrange step for (1.19) provides quadratic (hence, also arbitrarily fast linear) decrease for the Fischer–Burmeister residual Φ_{FB} of the KKT system (1.2) for MPCC (1.1).

ALGORITHM 3.1. Preliminary step. Fix $\theta, q \in (0, 1)$. Set $k = 0$, and choose $x^0 \in \mathbf{R}^n$ and $\lambda^0 = (\lambda_G^0, \lambda_H^0, \lambda_0^0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$.

Identification step. Define the index sets I_G and I_H according to (2.7) and (2.8). If $k = 0$, or if I_G or I_H does not coincide with its counterpart computed at the previous iteration, or if $I_G \cup I_H \neq \{1, \dots, m\}$, go to **Outer-phase step**.

Active-set step. If the current point (x^k, λ^k) was generated by **Outer-phase step**, set $\tilde{k} = k$, store (x^k, λ^k) , and define $(\mu_G^k)_{I_G}$ and $(\mu_H^k)_{I_H}$ by

$$(3.1) \quad (\mu_G^k)_i = (\lambda_G^k)_i - \lambda_0^k H_i(x^k), \quad i \in I_G \setminus I_H,$$

$$(3.2) \quad (\mu_H^k)_i = (\lambda_H^k)_i - \lambda_0^k G_i(x^k), \quad i \in I_H \setminus I_G,$$

$$(3.3) \quad (\mu_G^k)_i = (\lambda_G^k)_i, \quad (\mu_H^k)_i = (\lambda_H^k)_i, \quad i \in I_G \cap I_H.$$

Compute (x^{k+1}, μ^{k+1}) as follows.

- The triple $(x^{k+1}, (\mu_G^{k+1})_{I_G}, (\mu_H^{k+1})_{I_H})$ is generated by the step of Newton–Lagrange method for tightened MP (1.19) from the point $(x^k, (\mu_G^k)_{I_G}, (\mu_H^k)_{I_H})$.
- $(\mu_G^{k+1})_{I_H \setminus I_G} = 0, (\mu_H^{k+1})_{I_G \setminus I_H} = 0$.

If there exists $i \in I_G \setminus I_H$ such that $H_i(x^{k+1}) = 0$, or there exists $i \in I_H \setminus I_G$ such that $G_i(x^{k+1}) = 0$, go to **Outer-phase step**. Otherwise, define $\lambda^{k+1} = (\lambda_G^{k+1}, \lambda_H^{k+1}, \lambda_0^{k+1})$ according to (2.9)–(2.12). If the point (x^{k+1}, λ^{k+1}) is well defined and satisfies the condition

$$(3.4) \quad \|\Phi_{FB}(x^{k+1}, \lambda^{k+1})\| \leq q \|\Phi_{FB}(x^k, \lambda^k)\|,$$

adjust k by 1, and go to **Identification step**.

Outer-phase step. If the current point (x^k, λ^k) was generated by **Active-set step**, set $k = \tilde{k}$ and $(x^k, \lambda^k) = (x^{\tilde{k}}, \lambda^{\tilde{k}})$.

Compute (x^{k+1}, λ^{k+1}) according to the outer-phase strategy. Adjust k by 1, and go to **Identification step**.

Global convergence properties of Algorithm 3.1 are quite transparent. By (3.4), we immediately obtain the following result.

THEOREM 3.1. *Let $\{(x^k, \lambda^k)\}$ be a trajectory generated by Algorithm 3.1, and suppose that all the iterates in this trajectory with k large enough are generated by **Active-set step** of the algorithm. Then*

$$(3.5) \quad \Phi_{FB}(x^k, \lambda^k) \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

In particular, the primal part of any accumulation point of $\{(x^k, \lambda^k)\}$ is strongly stationary for (1.1), while the dual part is an associated Lagrange multiplier.

Except for the case considered in Theorem 3.1, the only other possibility is that all the iterates are generated by the outer-phase strategy (because unsuccessful active-set iterates are eventually discarded). In this case, the method inherits global convergence of the outer strategy. Possible choices of outer strategies will be discussed below.

To prove quadratic convergence of Algorithm 3.1, some work is required. We start with the following dual estimate.

LEMMA 3.2. *Let \bar{x} be a strongly stationary point of MPCC (1.1), and assume that MPCC-LICQ (1.5) holds at \bar{x} . Let $\bar{\lambda} \in \Lambda(\bar{x})$.*

Then there exists $c > 0$ such that, for each (x^k, λ^k) close enough to $(\bar{x}, \bar{\lambda})$, it holds that

$$(3.6) \quad \|\lambda^k - \hat{\lambda}^k\| \leq c \operatorname{dist}(\lambda^k, \Lambda(\bar{x})),$$

where $\hat{\lambda}^k$ is defined according to (2.13)–(2.15).

Proof. We argue by contradiction. If $\lambda^k \in \Lambda(\bar{x})$, then by Proposition 1.1 and by (2.13)–(2.15), we have that $\lambda^k = \hat{\lambda}^k$, and (3.6) holds with any $c \geq 0$. Suppose that there exists a sequence $\{(x^k, \lambda^k)\}$ convergent to $(\bar{x}, \bar{\lambda})$ such that $\lambda^k \notin \Lambda(\bar{x}) \forall k$, and

$$(3.7) \quad \|\lambda^k - \hat{\lambda}^k\| / \operatorname{dist}(\lambda^k, \Lambda(\bar{x})) \rightarrow \infty \quad \text{as } k \rightarrow \infty.$$

Let $\bar{\lambda}^k$ be the orthogonal projection of λ^k onto $\Lambda(\bar{x})$. Then (3.7) is equivalent to

$$(3.8) \quad \|\lambda^k - \bar{\lambda}^k\| / \|\lambda^k - \hat{\lambda}^k\| \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

For each k , we have that

$$\left(\lambda^k - \hat{\lambda}^k \right) / \|\lambda^k - \hat{\lambda}^k\| = (\lambda^k - \bar{\lambda}^k) / \|\lambda^k - \hat{\lambda}^k\| + (\bar{\lambda}^k - \hat{\lambda}^k) / \|\lambda^k - \hat{\lambda}^k\|.$$

Observe that in this equality the left-hand side has unit norm and belongs to the “vertical” hyperplane $\lambda_0 = 0$; the first term in the right-hand side tends to 0 as $k \rightarrow \infty$, by (3.8); while the second term in the right-hand side belongs to the straight line containing the ray $\Lambda(\bar{x})$ (see Proposition 1.1 and (2.13)–(2.15)), which does not belong to the “vertical” hyperplane. The contradiction is now evident. \square

THEOREM 3.3. *Let $\{(x^k, \lambda^k)\}$ be a trajectory generated by Algorithm 3.1, and suppose that this trajectory has an accumulation point $(\bar{x}, \bar{\lambda})$, with \bar{x} being a strongly stationary point of problem (1.1) and $\bar{\lambda}$ being an associated Lagrange multiplier, satisfying MPCC-LICQ (1.5) and SOSC (1.14).*

Then the entire trajectory $\{(x^k, \lambda^k)\}$ converges to $(\bar{x}, \bar{\lambda})$, and the rate of convergence is quadratic.

Proof. Let (x^k, λ^k) be close to $(\bar{x}, \bar{\lambda})$. Furthermore, let (x^{k+1}, λ^{k+1}) be computed by the **Active-set step** of Algorithm 3.1 (this point is correctly defined, according to Theorem 2.4).

We next construct $\bar{\lambda}^{k+1} \in \Lambda(\bar{x})$ satisfying the estimate

$$(3.9) \quad \|\lambda^{k+1} - \bar{\lambda}^{k+1}\| = O\left(\left\|\left(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k\right)\right\|^2\right).$$

This is done separately for the two possible cases in (2.22).

If the first relation in (2.22) holds, we define $\bar{\lambda}^{k+1} = \hat{\lambda}^k$. In this case, by Proposition 1.1, by (2.13)–(2.15), and by the proximity of λ_0^k to $\bar{\lambda}_0$, we have that $\hat{\lambda}^k \in \Lambda(\bar{x})$. The estimate (3.9) now follows from (2.24). Let the second relation in (2.22) hold. If $\lambda_0^{k+1} = \nu_{k+1}$, then set $\bar{\lambda}^{k+1} = \bar{\lambda}$. In this case, estimate (3.9) follows from (2.25). If $\lambda_0^{k+1} = \lambda_0^k$, then $\lambda_0^k > \nu_{k+1}$, and we define $\bar{\lambda}^{k+1}$ as follows.

If $\lambda_0^k \geq \bar{\nu}$, then set $\bar{\lambda}^{k+1} = \hat{\lambda}^k$. In this case, $\hat{\lambda}^k \in \Lambda(\bar{x})$ according to Proposition 1.1 and (2.13)–(2.15), and estimate (3.9) follows from (2.24).

If $\lambda_0^k < \bar{\nu}$, then $\nu_{k+1} < \lambda_0^k < \bar{\nu}$, and by (2.16), (2.19), and the second relation in (2.22), we have that

$$(3.10) \quad \|\hat{\lambda}^k - \bar{\lambda}\| = O(|\lambda_0^k - \bar{\lambda}_0|) = O(|\nu_{k+1} - \bar{\nu}|) = O\left(\left\|\left(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k\right)\right\|^2\right).$$

Set $\bar{\lambda}^{k+1} = \bar{\lambda}$. Then estimate (3.9) follows from (2.24), (3.10), and from the inequality

$$\|\lambda^{k+1} - \bar{\lambda}^{k+1}\| \leq \|\lambda^{k+1} - \hat{\lambda}^k\| + \|\hat{\lambda}^k - \bar{\lambda}\|.$$

Set $\varphi_{FB}(x, \lambda) = \|\Phi_{FB}(x, \lambda)\|^2$, $x \in \mathbf{R}^n$, $\lambda \in \mathbf{R} \times \mathbf{R}^m \times \mathbf{R}^m$. As is well known, the function φ_{FB} is smooth, and since $(\bar{x}, \bar{\lambda}^{k+1})$ is a global unconstrained minimizer of this function, we obtain the equalities

$$(3.11) \quad \varphi_{FB}(\bar{x}, \bar{\lambda}^{k+1}) = 0, \quad \varphi'_{FB}(\bar{x}, \bar{\lambda}^{k+1}) = 0.$$

Recall that, under our assumptions, the error bound (2.6) holds for all (x, λ) close enough to $(\bar{x}, \bar{\lambda})$. Then, by (2.17), (3.9), (3.11), and by Lemma 3.2, we obtain that

$$\begin{aligned} & \|\Phi_{FB}(x^{k+1}, \lambda^{k+1})\|^2 = \varphi_{FB}(x^{k+1}, \lambda^{k+1}) \\ &= \varphi_{FB}(x^{k+1}, \lambda^{k+1}) - \varphi_{FB}(\bar{x}, \bar{\lambda}^{k+1}) \\ &= \langle \varphi'_{FB}(\bar{x}, \bar{\lambda}^{k+1}), (x^{k+1} - \bar{x}, \lambda^{k+1} - \bar{\lambda}^{k+1}) \rangle + O\left(\|(x^{k+1} - \bar{x}, \lambda^{k+1} - \bar{\lambda}^{k+1})\|^2\right) \\ &= O\left(\|(x^{k+1} - \bar{x}, \lambda^{k+1} - \bar{\lambda}^{k+1})\|^2\right) = O\left(\|(x^k - \bar{x}, \lambda^k - \hat{\lambda}^k)\|^4\right) \\ &= O\left(\left(\|x^k - \bar{x}\| + \text{dist}(\lambda^k, \Lambda(\bar{x}))\right)^4\right) = O\left(\|\Phi_{NR}(x^k, \lambda^k)\|^4\right) = O\left(\|\Phi_{FB}(x^k, \lambda^k)\|^4\right), \end{aligned}$$

where the last relation follows from the equivalence of $\|\Phi_{NR}(\cdot)\|$ and $\|\Phi_{FB}(\cdot)\|$ in terms of their growth rates [19].

Evidently, the above relation implies (3.4) for any fixed $q \in (0, 1)$, if (x^k, λ^k) is close enough to $(\bar{x}, \bar{\lambda})$. This implies that the **Active-set step** will be accepted, and Algorithm 3.1 will be further working identically to the (local) Algorithm 2.2. The result now follows from Theorem 2.4. \square

One possible choice of the outer-phase algorithm is the elastic mode SQP method discussed in section 3.2 below. Another possibility is to use the merit function φ_{FB} in order to organize the outer phase as well, by means of globalizing the semismooth Newton method applied to the equation $\Phi_{FB}(x, \lambda) = 0$. The resulting algorithm is in the spirit of the method for complementarity problems in [14], and its extension to globalization of an active-set method for mixed complementarity problems in [3, section 3]. One advantage of such a scheme is that one can guarantee the overall monotonicity of the sequence $\{\|\Phi_{FB}(x^k, \lambda^k)\|\}$, and thus no backup safeguards are needed when entering the active-set phase (i.e., global convergence can be proved without such safeguards). That is why we present this scheme as a separate algorithm.

ALGORITHM 3.2. Preliminary step. Fix $\theta, q, \varepsilon, \tau \in (0, 1)$, $\delta, \gamma > 0$. Set $k = 0$, and choose $x^0 \in \mathbf{R}^n$ and $\lambda^0 = (\lambda_G^0, \lambda_H^0, \lambda_0^0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$.

Identification step. Define the index sets I_G and I_H according to (2.7) and (2.8). If $k = 0$ or if I_G or I_H does not coincide with its counterpart computed at the previous iteration or if $I_G \cup I_H \neq \{1, \dots, m\}$, go to **SNM–FB step**.

Active-set step. If the current point (x^k, λ^k) was generated by **SNM–FB step**, define $(\mu_G^k)_{I_G}$ and $(\mu_H^k)_{I_H}$ by (3.1)–(3.3). Compute (x^{k+1}, μ^{k+1}) as follows.

- The triple $(x^{k+1}, (\mu_G^{k+1})_{I_G}, (\mu_H^{k+1})_{I_H})$ is generated by the step of Newton–Lagrange method for tightened MP (1.19) from the point $(x^k, (\mu_G^k)_{I_G}, (\mu_H^k)_{I_H})$.
- $(\mu_G^{k+1})_{I_H \setminus I_G} = 0, (\mu_H^{k+1})_{I_G \setminus I_H} = 0$.

If there exists $i \in I_G \setminus I_H$ such that $H_i(x^{k+1}) = 0$ or there exists $i \in I_H \setminus I_G$ such that $G_i(x^{k+1}) = 0$, go to **SNM–FB step**. Otherwise, define $\lambda^{k+1} = (\lambda_G^{k+1}, \lambda_H^{k+1}, \lambda_0^{k+1})$ according to (2.9)–(2.12). If the point (x^{k+1}, λ^{k+1}) is well-defined and satisfies the condition (3.4), adjust k by 1, and go to **Identification step**.

SNM–FB step. Compute $\Lambda_k \in \partial_B \Phi_{FB}(x^k, \lambda^k)$ and

$$(x^{k+1}, \lambda^{k+1}) = (x^k, \lambda^k) - \Lambda_k^{-1} \Phi_{FB}(x^k, \lambda^k).$$

If this point is well-defined and (3.4) holds, and satisfies the condition, adjust k by 1, and go to **Identification step**.

If x^{k+1} is well-defined but (3.4) does not hold, set $d^k = x^{k+1} - x^k$. If

$$\langle \varphi'_{FB}(x^k, \lambda^k), d^k \rangle \leq -\gamma \|d^k\|^\delta,$$

go to **Linesearch step**.

Gradient step. Set $d^k = -\varphi'_{FB}(x^k, \lambda^k)$.

Linesearch step. Compute the stepsize parameter α_k according to the Armijo rule: $\alpha_k = \tau^s$, where s is the smallest nonnegative integer satisfying

$$\varphi_{FB}((x^k, \lambda^k) + \tau^s d^k) \leq \varphi_{FB}(x^k, \lambda^k) + \varepsilon \tau^s \langle \varphi'_{FB}(x^k, \lambda^k), d^k \rangle.$$

Set $(x^{k+1}, \lambda^{k+1}) = (x^k, \lambda^k) + \alpha_k d^k$, adjust k by 1, and go to **Identification step**.

THEOREM 3.4. Let $\{(x^k, \lambda^k)\}$ be a trajectory generated by Algorithm 3.2.

Then any accumulation point $(\bar{x}, \bar{\lambda})$ of $\{(x^k, \lambda^k)\}$ satisfies $\varphi'_{FB}(\bar{x}, \bar{\lambda}) = 0$.

Furthermore, if there exists an infinite subsequence of $\{(x^k, \lambda^k)\}$ such that all the iterates in this subsequence are generated by **Active-set step**, then (3.5) holds. In that case, the primal part of any accumulation point of $\{(x^k, \lambda^k)\}$ is strongly stationary in (1.1), while the dual part is an associated Lagrange multiplier.

Proof. If there exists an infinite subsequence of $\{(x^k, \lambda^k)\}$ such that all the iterates in this subsequence are generated by **Active-set step** of the algorithm, then (3.5) follows immediately from (3.4) and the fact that the values of φ_{FB} are nonincreasing

along the trajectories of the algorithm. The only other possibility is that the “tail” of the trajectory is generated by the outer-phase algorithm, in which case the result can be obtained extending [14, Theorem 3.1] to the setting of mixed complementarity problems. \square

Finally, to obtain the rate of convergence result, one should just repeat the proof of Theorem 3.3, with Algorithm 3.1 replaced by Algorithm 3.2.

THEOREM 3.5. *Let $\{(x^k, \lambda^k)\}$ be a trajectory generated by Algorithm 3.2, and suppose that this trajectory has an accumulation point $(\bar{x}, \bar{\lambda})$, with \bar{x} being a strongly stationary point of problem (1.1) and $\bar{\lambda}$ being an associated Lagrange multiplier, satisfying MPCC-LICQ (1.5) and SOSC (1.14).*

Then the entire trajectory $\{(x^k, \lambda^k)\}$ converges to $(\bar{x}, \bar{\lambda})$, and the rate of convergence is quadratic.

We have thus developed a QP-free algorithm for MPCC, with justified global convergence and quadratic rate of convergence under MPCC-LICQ and the usual SOSC (1.14).

3.2. Globalization based on SQP with linesearch. Introducing slack variables, MPCC (1.1) can be equivalently written in the form

$$(3.12) \quad \min_{(x, y, z)} f(x) \quad \text{s.t.} \quad G(x) = y, H(x) = z, y \geq 0, z \geq 0, \langle y, z \rangle \leq 0.$$

As is well known, this reformulated MPCC has the same properties (MPCC constraint qualifications and SOSC) as (1.1), while being preferable for numerical solution by SQP [6, 7].

We first discuss the outer (elastic mode SQP) phase of the algorithm stated below. When SQP is applied to MPCC, under natural assumptions SQP subproblems can be infeasible, even arbitrarily close to a solution. Thus some kind of constraints relaxation (known as *elastic mode*; see, e.g., [1]) has to be used. Let $u^k = (x^k, y^k, z^k) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m$ be the current primal iterate, and let $\lambda_0^k \geq 0$ be the current estimate of the Lagrange multiplier corresponding to the last constraint in (3.12). We suggest partial relaxation of SQP constraints, which gives the following subproblems:

$$(3.13) \quad \begin{aligned} \min_{(d, t)} \quad & \langle f'(x^k), \xi \rangle + \frac{1}{2} \langle \mathcal{H}_k \xi, \xi \rangle + \lambda_0^k \langle \eta, \zeta \rangle + ct \\ \text{s.t.} \quad & -te \leq y^k - G(x^k) + \eta - G'(x^k) \xi \leq te, \\ & -te \leq z^k - H(x^k) + \zeta - H'(x^k) \xi \leq te, \\ & y^k + \eta \geq 0, z^k + \zeta \geq 0, \langle y^k, z^k \rangle + \langle z^k, \eta \rangle + \langle y^k, \zeta \rangle \leq 0, \end{aligned}$$

where $d = (\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m$, $t \in \mathbf{R}$, \mathcal{H}_k is an $n \times n$ positive definite symmetric matrix, $c > 0$ is the (penalty) parameter, and $e \in \mathbf{R}^m$ is the vector of ones.

If (d^k, t_k) is a solution of (3.13), then the next iterate is defined by $u^{k+1} = u^k + \alpha_k d^k$, where $\alpha_k \in (0, 1]$ is the stepsize parameter. Choosing $y^0 \geq 0$ and $z^0 \geq 0$, by the first two constraints in the last line of (3.13), it evidently holds that $y^k \geq 0$ and $z^k \geq 0$ for all k . The last three constraints in (3.13) are then always consistent (for example, $\eta = -y^k$ and $\zeta = 0$ satisfies this part of constraints), while the other constraints in (3.13) are consistent due to the elastic mode. It follows that subproblems (3.13) are always feasible. Furthermore, the objective function in (3.13) is bounded below on the nonempty feasible set. Hence, by the Frank–Wolfe theorem [2, Theorem 2.8.1], the subproblem (3.13) has a solution.

Taking into account that $y^k \geq 0$ and $z^k \geq 0$ for all k , the following penalty function can be used in the linesearch procedure for choosing the stepsize parameter:

for $u = (x, y, z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m$,

$$(3.14) \quad \varphi_c(u) = f(x) + c\psi(u) + c\langle y, z \rangle, \quad \psi(u) = \|(y, z) - (G(x), H(x))\|_\infty.$$

If (d^k, t^k) is a solution of SQP subproblem (3.13), then, by direct computation of directional derivative and by standard argument, it can be seen that d^k is a direction of descent for φ_c , provided c is large enough. This justifies the linesearch procedure along the direction obtained from (3.13).

Now let $\lambda^k = (\lambda_G^k, \lambda_H^k, \lambda_0^k)$ be the current estimate of the Lagrange multipliers corresponding to inequality constraints in (3.12). It can be easily seen that such λ^k is a natural approximation of Lagrange multipliers of the original MPCC (1.1). Define the index sets I_G and I_H according to (2.7) and (2.8), respectively. Once we have reasons to believe that the index sets I_G and I_H give a correct identification, we shall set the corresponding slacks to zero ($y_{I_G}^k = 0, z_{I_H}^k = 0$) and switch to the inner (active-set) phase. We note that identification cannot be correct if $I_G \cup I_H \neq \{1, \dots, m\}$. Another sign of incorrect identification is when the sets I_G and I_H are not yet stable (i.e., change from one iteration to the next). The inner phase consists in applying SQP to the tightened MP

$$(3.15) \quad \min_{(x, y, z)} f(x) \quad \text{s.t.} \quad G(x) = y, \quad H(x) = z, \quad y_{I_G} = 0, \quad z_{I_H} = 0,$$

i.e., we find a solution d^k of

$$(3.16) \quad \begin{aligned} \min_d \quad & \langle f'(x^k), \xi \rangle + \frac{1}{2} \langle \mathcal{H}_k \xi, \xi \rangle \\ \text{s.t.} \quad & y^k - G(x^k) + \eta - G'(x^k) \xi = 0, \quad y_{I_G}^k + \eta_{I_G} = 0, \\ & z^k - H(x^k) + \zeta - H'(x^k) \xi = 0, \quad z_{I_H}^k + \zeta_{I_H} = 0, \end{aligned}$$

and set $u^{k+1} = u^k + \alpha_k d^k$, with some $\alpha_k \in (0, 1]$. Infeasibility of the active-set subproblem (3.16) is again one of the signs of incorrect identification, in which case we go back to the outer phase. We shall show below that if the subproblem (3.16) is feasible, its solution provides a direction of descent for the same penalty function (3.14) that is used in the outer phase. This justifies incorporating the active-set phase into the global SQP framework.

Having in mind fast local convergence, the matrices \mathcal{H}_k in (3.13) and (3.16) should in some specific sense (i.e., not necessarily on the whole space) “approximate” the Hessians with respect to x of the Lagrangians of (3.12) and (3.15), respectively, at the limiting primal-dual solution. It can be easily checked that both these Hessians coincide with $\frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})$, where \bar{x} is the primal limiting solution, while $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H)$ is the part of dual limiting solution, corresponding to the first two constraints in (3.12) and (3.15). (For problem (3.12), $\bar{\mu}$ is an MPCC-multiplier associated with \bar{x} , by necessity. For problem (3.15), this is the case as well, if the index sets I_G and I_H are correctly identified and provided \bar{x} is a strongly stationary point of MPCC (1.1) with unique associated MPCC-multiplier.) In order to approximate $\frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})$, one might need to compute an approximation $\mu^k = (\mu_G^k, \mu_H^k)$ of $\bar{\mu}$. Within the inner phase, these estimates can be computed directly as Lagrange multipliers corresponding to the first two constraints in (3.16). Within the outer phase, they can be derived from λ^k by the equalities

$$(3.17) \quad \mu_G^k = \lambda_G^k - \lambda_0^k H(x^k), \quad \mu_H^k = \lambda_H^k - \lambda_0^k G(x^k)$$

(note that these formulas do not use identification of active indices).

We proceed to formally state the proposed algorithm.

ALGORITHM 3.3. Preliminary step. Fix $\theta, \varepsilon, \tau \in (0, 1)$, and $c > 0$. Set $k = 0$, and choose $u^0 = (x^0, y^0, z^0) \in \mathbf{R}^n \times \mathbf{R}_+^m \times \mathbf{R}_+^m$ and $\lambda^0 = (\lambda_G^0, \lambda_H^0, \lambda_0^0) \in \mathbf{R}_+ \times \mathbf{R}_+^m \times \mathbf{R}_+^m$.

Identification step. Define the index sets I_G and I_H according to (2.7) and (2.8). If $k = 0$ or if I_G or I_H does not coincide with its counterpart computed at the previous iteration or if $I_G \cup I_H \neq \{1, \dots, m\}$, go to **Elastic mode SQP step**.

Active-set step. If d^{k-1} was generated by **Elastic mode SQP step**, set $\tilde{k} = k$, store $u^{\tilde{k}}$ and $\lambda^{\tilde{k}}$, redefine $u^k = (x^k, y^k, z^k)$ by setting $y_{I_G}^k = 0$, $z_{I_H}^k = 0$, and define $\mu^k = (\mu_G^k, \mu_H^k)$ by (3.1)–(3.3). and

$$(3.18) \quad (\mu_G^k)_i = 0, \quad i \in I_H \setminus I_G, \quad (\mu_H^k)_i = 0, \quad i \in I_G \setminus I_H.$$

Using μ^k , choose an $n \times n$ positive definite symmetric matrix \mathcal{H}_k . If (3.16) is infeasible, go to **Elastic mode SQP step**.

Compute $d^k = (\xi^k, \eta^k, \zeta^k)$ as a solution of (3.16) and $\mu^{k+1} = (\mu_G^{k+1}, \mu_H^{k+1})$ as an associated Lagrange multiplier corresponding to the first two constraints in (3.16). Set $\tilde{x}^{k+1} = x^k + \xi^k$. If there exists $i \in I_G \setminus I_H$ such that $H_i(\tilde{x}^{k+1}) = 0$ or there exists $i \in I_H \setminus I_G$ such that $G_i(\tilde{x}^{k+1}) = 0$, go to **Elastic mode SQP step**. Otherwise, define $\lambda^{k+1} = (\lambda_G^{k+1}, \lambda_H^{k+1}, \lambda_0^{k+1})$ according to (2.9)–(2.12), with x^{k+1} replaced by \tilde{x}^{k+1} , and go to **Linesearch step**.

Elastic mode SQP step. If d^{k-1} was generated by **Active-set step**, redefine $u^k = (x^k, y^k, z^k)$ by setting

$$(3.19) \quad y_i^k = 0 \quad \forall i = 1, \dots, m \text{ such that } y_i^k < 0,$$

$$(3.20) \quad z_i^k = 0 \quad \forall i = 1, \dots, m \text{ such that } z_i^k < 0$$

and if

$$(3.21) \quad \varphi_c(u^k) > \varphi_c(u^{\tilde{k}}),$$

then set $k = \tilde{k}$, $u^k = u^{\tilde{k}}$, and $\lambda^k = \lambda^{\tilde{k}}$.

Using $\mu^k = (\mu_G^k, \mu_H^k)$ computed according to (3.17), choose an $n \times n$ positive definite symmetric matrix \mathcal{H}_k .

Compute (d^k, t^k) as a solution of (3.13) and $\lambda^{k+1} = (\lambda_G^{k+1}, \lambda_H^{k+1}, \lambda_0^{k+1})$ as an associated Lagrange multiplier corresponding to inequality constraints in (3.13).

Linesearch step. Compute the stepsize parameter α_k according to the Armijo rule: $\alpha_k = \tau^s$, where s is the smallest nonnegative integer satisfying

$$(3.22) \quad \varphi_c(u^k + \tau^s d^k) \leq \varphi_c(u^k) + \varepsilon \tau^s \varphi'_c(u^k; d^k).$$

Set $u^{k+1} = u^k + \alpha_k d^k$, adjust k by 1, and go to **Identification step**.

Observe that the active-set iterations always start with $u^k = (x^k, y^k, z^k)$ satisfying complementarity. Indeed, the SQP iterations in the elastic mode start with $y^k \geq 0$, $z^k \geq 0$ and maintain nonnegativity. Furthermore, active-set iterations start with $(y^k)_{I_G} = 0$, $(z^k)_{I_H} = 0$, where $I_G \cup I_H = \{1, \dots, m\}$. The only way complementarity can be violated during a sequence of active-set steps is when some component of y or z becomes negative. Obviously, this can happen only for indices which are not in I_G in the case of y and not in I_H in the case of z . Once a component becomes negative, this index is immediately added to the corresponding set (see (2.7), (2.8)), which makes the sets change. In such a case, we get out of the active-set phase, restore nonnegativity (see (3.19), (3.20)), and if such a point breaks monotonicity

of the sequence of the penalty function values (that is, if (3.21) happens), we go back to the last iterate preceding the active-set phase (which was determined to be premature).

We next show that when within **Active-set step** of Algorithm 3.3 the subproblem (3.16) is feasible, the generated direction d^k is of descent for the penalty function (3.14) at u^k , and hence the linesearch procedure along this direction is well-defined.

LEMMA 3.6. *Let $d^k = (\xi^k, \eta^k, \zeta^k)$ and $\mu^{k+1} = (\mu_G^{k+1}, \mu_H^{k+1})$ be computed within **Active-set step** of Algorithm 3.3 from the primal-dual solution of (3.16).*

Then

$$(3.23) \quad \varphi'_c(u^k; d^k) \leq -\langle \mathcal{H}_k \xi^k, \xi^k \rangle - (c - \|\mu^{k+1}\|_1) \psi(u^k).$$

In particular, d^k is a direction of descent for φ_c , provided either $\xi^k \neq 0$ or $c > \|\mu^{k+1}\|_1$ and $\psi(u^k) > 0$.

Proof. First note that, as observed above, $y^k \geq 0$, $z^k \geq 0$, because otherwise the index sets would have changed, and we would not be solving (3.16). Furthermore, recall that we set

$$(3.24) \quad y_{I_G}^k = 0, \quad z_{I_H}^k = 0$$

when the algorithm enters the active-set phase. Moreover, these equalities are preserved within this phase, because the last line in (3.16) implies

$$(3.25) \quad \eta_{I_G}^k = 0, \quad \zeta_{I_H}^k = 0.$$

We thus have that whenever $z_i^k > 0$, it holds that $i \notin I_H$. Since the algorithm can enter the active-set phase only with $I_G \cup I_H = \{1, \dots, m\}$, we have that $i \in I_G$. Therefore, $y_i^k = 0$ by the first equality in (3.24), and hence $\eta_i^k = 0$ by the first equality in (3.25). This shows that $\langle z^k, \eta^k \rangle = 0$. Analogously, $\langle y^k, \zeta^k \rangle = 0$. It follows that the directional derivative of the term in the definition (3.14) of φ_c that penalizes complementarity violation is equal to zero.

By direct computation of the directional derivative and by standard argument,

$$(3.26) \quad \psi'(u^k; d^k) = -\psi(u^k).$$

Furthermore, by the Lagrange optimality conditions for (3.16), it holds that

$$(3.27) \quad f'(x^k) + \mathcal{H}_k \xi^k - (G'(x^k))^T \mu_G^{k+1} - (H'(x^k))^T \mu_H^{k+1} = 0,$$

$$(3.28) \quad (\mu_G^{k+1})_{I_H \setminus I_G} = 0, \quad (\mu_H^{k+1})_{I_G \setminus I_H} = 0.$$

Taking again into account the structure of constraints in (3.16), from (3.24)–(3.28) we obtain

$$\begin{aligned} \langle f'(x^k), \xi^k \rangle &= -\langle \mathcal{H}_k \xi^k, \xi^k \rangle + \langle \mu_G^{k+1}, G'(x^k) \xi^k \rangle + \langle \mu_H^{k+1}, H'(x^k) \xi^k \rangle \\ &= -\langle \mathcal{H}_k \xi^k, \xi^k \rangle + \sum_{i \in I_G} (\mu_G^{k+1})_i (y_i^k - G_i(x^k) + \eta_i^k) \\ &\quad + \sum_{i \in I_H} (\mu_H^{k+1})_i (z_i^k - H_i(x^k) + \zeta_i^k) \\ &= -\langle \mathcal{H}_k \xi^k, \xi^k \rangle + \sum_{i \in I_G} (\mu_G^{k+1})_i (y_i^k - G_i(x^k)) + \sum_{i \in I_H} (\mu_H^{k+1})_i (z_i^k - H_i(x^k)) \\ &\leq -\langle \mathcal{H}_k \xi^k, \xi^k \rangle + \|\mu^{k+1}\|_1 \psi(u^k), \end{aligned}$$

where $\mu^{k+1} = (\mu_G^{k+1}, \mu_H^{k+1})$. Combining the latter with (3.26), we obtain (3.23). \square

If $\xi^k = 0$ and $\psi(u^k) = 0$, we obtain from (3.14) and from the constraints of (3.16) that $d^k = 0$ and the point u^k is feasible in (3.12). Furthermore, (3.27) and (3.28) show that x^k is a weakly stationary point of (1.1). Otherwise, the linesearch procedure is well-defined and results in the decrease of the penalty function value with respect to $\varphi_c(u^k)$. Overall, the method generates iterates such that the sequence $\{\varphi_c(u^k)\}$ is nonincreasing, as in standard SQP framework.

4. Numerical examples. In this section, we illustrate behavior of the algorithms discussed above by some numerical examples. In what follows, **Linearization** is the linesearch SQP method with \mathcal{H}_k being the identity matrix, applied to the original problem formulation (without slacks), while **SQP-slacks** is the linesearch SQP method with $\mathcal{H}_k = \frac{\partial^2 \mathcal{L}}{\partial x^2}(x^k, \mu^k)$ applied to the problem formulation with slacks. The first simple choice of \mathcal{H}_k is motivated by robustness (if \mathcal{H}_k is not positive definite, then the subproblems sometimes do not have a solution, while more sophisticated choices of positive definite matrices require complex quasi-Newton implementations). The second choice of \mathcal{H}_k is motivated by its efficiency (when subproblems are solvable). SQP-type methods were all implemented in their basic form, without elastic mode (which corresponds to setting $t = 0$), without any attempts to modify \mathcal{H}_k with respect to the two alternative choices above, and without any tools for avoiding the Maratos effect. While without a doubt important for any professional implementation, all those details have no real bearing for illustrating our proposal for forcing fast local convergence by the active-set phase. Linesearch parameters were chosen as follows: $\varepsilon = 0.1$ and $\tau = 0.5$. We used the simplest update rule for penalty parameters: $c_0 = \|\lambda^1\|_\infty + 1$, and then for each $k = 1, 2, \dots$, we set $c_k = c_{k-1}$ if $c_{k-1} \geq \|\lambda^{k+1}\|_\infty$, and $c_k = \|\lambda^{k+1}\|_\infty + 1$ otherwise. The other implemented methods are the following. **SNM-FB** is Algorithm 3.2 without **Active-set step** and with parameters $\delta = 2.1$, $\gamma = 10^{-9}$, $\varepsilon = 10^{-4}$, and $\tau = 0.5$. **Linearization+AS** and **SQP-slacks+AS** are the modifications of algorithms **Linearization** and **SQP-slacks**, respectively, supplied with the option of switching to **Active-set step**, implemented as specified in Algorithm 3.1. Finally, **SNM-FB+AS** is precisely Algorithm 3.2. The identification test parameter and the linear decrease parameter were chosen as follows: $\theta = 0.5$, $q = 0.9$. All computations were performed in Matlab environment, with the QP-subproblems solved by the built-in Matlab QP-solver. We used the stopping criterion of the form

$$(4.1) \quad \|\Phi_{FB}(x^k, \lambda^k)\| < 10^{-7}.$$

We start with reporting some local runs of the algorithms discussed above for the following example, which is a modified version of **ralph2** in MacMPEC [12]. A separate consideration of this example is due to the fact that it is known to violate MPCC-SOSC, and so we expect that our method may behave better than SQP. The problem **ralph2** is modified by introducing higher-order nonlinear terms, in order to prevent the tendency for finite termination, which is quite common for SQP in the cases of “simple” (affine) constraints.

Example 4.1. The problem

$$\min x_1^2 + x_2^2 - 4x_1x_2 + x_2^3 \text{ s.t. } x_1 + x_2^2/2 \geq 0, x_2 - x_1^2 \geq 0, (x_1 + x_2^2/2)(x_2 - x_1^2) \leq 0,$$

has two local solutions $\bar{x}^1 = (0, 0)$ and $\bar{x}^2 = (1, 1)$, the latter being global, both satisfying MPCC-LICQ (1.5). The first solution satisfies piecewise SOSC (1.18) but violates MPCC-SOSC (1.12), while the second satisfies MPCC-SOSC (1.12).

We use the primal starting points close to \bar{x}^1 to facilitate convergence to this solution. Selected results for $\lambda_G^0 = 0.01$, $\lambda_H^0 = 0.02$, $\lambda_0^0 = 5$ are presented in Table 4.1.

TABLE 4.1
Example 4.1, local runs.

| Algorithm | $x^0 = 10^{-3} \times$ | | | | |
|---------------|------------------------|----------|----------|----------|----------|
| | (10, 1) | (7, 3) | (5, 5) | (3, 7) | (1, 10) |
| SQP-slacks | 11 | 10 | 10 | 10 | 4 |
| SNM-FB | 9 | 9 | 10 | 10 | 5 |
| SQP-slacks+AS | 3 | 3 | 3 | 3 | 3 |
| SNM-FB+AS | 4 | 3 | 3 | 3 | 3 |

For each run, we report the number of iterations before convergence was declared. Bold-faced numbers mean that convergence was achieved by active-set steps.

These results evidently demonstrate that, in this case, the active-set phase is useful. And this is precisely our message—we do not claim that it should always result in faster convergence than that for some nonmodified method, but it is easy to incorporate, is useful at least *sometimes*, and works as it is supposed to. To give some more validation of our claim, in the rest of this section we present numerical results for global convergence of our algorithms on some small test problems derived from MacMPEC [12]. The set of test problems was obtained as follows. We select all the problems in MacMPEC satisfying the following criteria: they have no more than 10 variables, and they do not have any inequality constraints apart from complementarity constraints (to be consistent with the problem setting of the paper). This makes 37 problems. Furthermore, we ignore the simple bounds (again in order to be consistent with the problem setting of the paper), which of course may sometimes affect the solutions/stationary points of these problems. Finally, `ralph1` suggests two different objective functions, and we use both, labeling the corresponding problems `ralph11` and `ralph12`. Thus, we end up with 38 problems.

We performed the runs of each algorithm from the same randomly generated starting points. Primal starting points were generated in a cubic neighborhood around the solution (a feasible point with the objective function value equal to the optimal value reported in MacMPEC; these points were found in the course of our experiments), with the edge of the cube equal to 20. Dual starting points for equality constraints were generated the same way, but around 0, while for dual starting points corresponding to the complementarity constraints multipliers there was the additional nonnegativity restriction. In the process of collecting information, we disregard the runs when at least one of the QP-employing algorithms fails because of a failure of the QP solver (such failures must be avoided in professional implementations by using elastic mode and modifications of the Hessian, or quasi-Newton updates with appropriate line-search, etc.; in any case these failures are concerned with the outer phase, rather than the use of the active-set step). Thus, we keep generating random starting points until we have 100 that do not cause QP solver failures.

When reporting the results, we count the cases of failure (when convergence was not achieved after 50 steps), the cases of convergence (to KKT points), and provide some details about convergence. We are not concerned whether the obtained KKT point is a local/global solution or not (this, once again, has to do mostly with behavior of the outer phase).

Columns of Tables 4.2 and 4.3 contain average/summarized information on the performance of each algorithm for 100 runs from random starting points. First row of each cell contains average characteristics over successful runs: iteration count, last active-set steps, overall count of active-set steps. Thus the average number of useless

TABLE 4.2
Results on MacMPEC problems.

| Problem | Algorithm | | | | | |
|---------|---------------|-----------------------|------------|----------------------|------------|-----------------------|
| | Linearization | Linearization+AS | SQP-slacks | SQP-slacks+AS | SNM-FB | SNM-FB+AS |
| bard1 | 15.2 1 | 6.4/0.9/0.9 0/87 | 3.4 0 | 3.0/0.3/0.3 0/27 | 14.0 24 | 7.5/1.0/1.0 19/79 |
| bard1m | 14.2 17 | 11.8/0.2/0.3 5/18 | 1.3 1 | 1.6/0.03/0.2 0/3 | 15.4 49 | 8.9/0.8/1.2 43/45 |
| dempe | 5.5 41 | 22/5.7/6.6 1/72 | 8.1 0 | 8.4/0.6/3.2 0/36 | 22.9 66 | 21.9/6.5/7.2 56/44 |
| desilva | 13.2 0 | 10.6/3.0/6.0 0/99 | 7.0 0 | 7.8/3.4/5.0 5/67 | 16.2 41 | 11.4/5.0/6.6 24/77 |
| ex9.2.1 | 2.0 4 | 2.0/0/0 4/0 | 2.5 8 | 2.5/0/0 8/0 | 9.6 26 | 8.4/0.9/1.0 26/72 |
| ex9.1.4 | 5.9 0 | 5.8/0.02/0.02 0/2 | 2.0 0 | 2.0/0/0 0/0 | 30.5 58 | 30.2/0.4/0.4 59/18 |
| ex9.2.1 | 4.4 20 | 4.4/0.07/0.07 19/6 | 1.7 4 | 1.7/0.01/0.06 4/1 | 13.3 36 | 10.6/0.5/0.9 34/35 |
| ex9.2.4 | 20.3 0 | 10.8/0.4/0.4 0/43 | 2.6 0 | 2.6/0.02/0.02 0/2 | 13.1 17 | 10.7/0.3/0.3 17/22 |
| ex9.2.5 | 13.9 0 | 5.8/0.3/0.6 0/32 | 3.1 0 | 3.1/0.1/0.3 0/12 | 19.4 26 | 16.6/1.0/1.0 17/79 |
| ex9.2.7 | 4.2 14 | 4.3/0.06/0.06 13/5 | 1.8 3 | 1.7/0.06/0.07 3/6 | 13.2 36 | 10.7/0.6/0.9 36/38 |
| ex9.2.8 | 14.4 0 | 14.4/0.4/0.4 0/37 | 2.5 0 | 2.5/0/0 0/0 | 10.8 66 | 7.3/1.0/1.0 55/45 |
| ex9.2.9 | 8.9 4 | 8.7/0.6/0.6 4/59 | 2.5 0 | 2.5/0.03/0.03 0/3 | 11.9 28 | 7.5/1.0/1.0 23/77 |
| flp2 | 17.6 11 | 10.2/0.7/1.8 0/69 | 2.3 0 | 2.4/0.01/0.5 0/1 | 11.3 26 | 12.1/1.0/1.6 20/78 |
| gauvin | 4.3 0 | 3.6/0.3/0.3 0/30 | 3.2 0 | 3.0/0.1/0.2 0/13 | 12.0 18 | 8.6/1.0/1.1 22/75 |
| jr1 | 2.9 0 | 2.9/0.01/0.5 0/1 | 2.5 0 | 3.0/0.01/0.7 0/1 | 8.7 14 | 7.9/1.0/1.7 5/93 |
| jr2 | 4.1 0 | 4.1/0.03/0.2 0/3 | 4.3 3 | 4.2/0.1/0.8 8/11 | 9.1 14 | 6.2/0.9/1.5 15/79 |
| kth1 | 5.7 0 | 5.6/0.09/0.09 0/9 | 1.8 0 | 1.8/0.03/0.03 0/3 | 11.4 48 | 3.4/1.0/1.0 20/80 |
| kth2 | 5.8 0 | 5.7/0.09/0.2 0/9 | 2.5 0 | 2.5/0.1/0.3 0/10 | 9.8 39 | 8.1/1.0/1.6 25/73 |
| kth3 | 4.0 0 | 3.0/0.4/0.4 0/38 | 3.0 0 | 2.8/0.4/0.6 0/44 | 9 12 | 6.8/1.0/1.4 11/85 |

active-set steps (eventually disregarded by backup safeguards) for **Linearization+AS** and **SQP-slack+AS** equals the difference between the third and the second number. Second row of each cell contains the overall number of failures and those cases when convergence was achieved by active-set steps. Note that what should be compared is the behavior of a given outer-phase algorithm with and without using the AS step. For **Linearization** and **SNM-FB**, in many cases using active-set step helps in terms of either robustness, efficiency, or both. **SQP-slacks** is very efficient by itself, and

TABLE 4.3
Results on MacMPEC problems.

| Problem | Algorithm | | | | | |
|--------------|---------------|-----------------------|------------|-----------------------|------------|-----------------------|
| | Linearization | Linearization+AS | SQP-slacks | SQP-slacks+AS | SNM-FB | SNM-FB+AS |
| nash1 | 7.1 27 | 8.6/0.3/0.7 12/29 | 2.4 0 | 2.3/0.06/0.4 1/5 | 9.8 32 | 11.3/1.0/1.7 32/67 |
| outrata31 | 20.8 3 | 20.7/0.3/0.3 3/63 | 11.7 19 | 13.7/0.2/1.1 23/13 | 16.5 44 | 15.6/0.6/1.8 44/28 |
| outrata32 | 35.3 26 | 30.3/0.9/1.8 14/43 | 10.7 30 | 14.0/0.3/0.9 34/18 | 17.3 37 | 16.7/0.4/1.9 33/29 |
| outrata33 | 25 8 | 22.4/0.7/6.6 4/48 | 11.3 10 | 15.7/0.4/0.7 16/31 | 17.0 36 | 17.3/0.1/1.1 32/5 |
| outrata34 | 39.7 12 | 23.1/1.6/1.6 9/91 | 11.0 38 | 18.2/0.6/0.7 42/35 | 14.2 29 | 14.2/0.6/0.6 29/42 |
| ralph11 | 3.5 26 | 3.4/0.3/0.3 15/26 | 1 0 | 1/0/0 0/0 | 10 57 | 2.7/0.9/0.9 7/83 |
| ralph12 | 6.6 39 | 6.3/0.5/0.5 21/37 | 1.4 0 | 14/0/0 0/0 | 100 | 2.1/1.0/1.0 17/83 |
| ralph2 | 2.0 0 | 1.9/0.1/0.1 0/10 | 3.8 0 | 2.0/0.3/0.3 0/25 | 9.2 35 | 2.7/1.0/1.0 1/99 |
| scholtes1 | 3.7 57 | 2.8/0.9/1.0 57/37 | 7.1 0 | 7.7/0.1/4.6 1/1 | 11.9 27 | 10.7/0.9/3.1 8/87 |
| scholtes2 | 100 | 9.1/7.1/7.4 33/67 | 11.2 2 | 10.1/7.1/7.2 0/100 | 12.9 19 | 10.1/7.5/7.5 2/98 |
| scholtes3 | 14.9 0 | 14.7/0.3/0.9 0/34 | 3.2 0 | 3.1/0.3/0.7 0/34 | 10.5 29 | 4.3/1.0/1.0 66/33 |
| scholtes5 | 2.6 0 | 2.6/0/0.01 0/0 | 2.3 6 | 2.6/0/0 6/0 | 10.0 14 | 9.8/0/0.08 13/0 |
| scale1 | 4 88 | 3.9/0.2/1.7 87/2 | 3.7 27 | 3.5/0.1/0.9 21/10 | 7.9 11 | 5.1/0.9/1.4 52/44 |
| scale2 | 35.8 0 | 11.2/1.0/2.5 0/95 | 3.6 0 | 3.6/0.04/0.2 0/4 | 6.6 1 | 5.2/1.0/1.0 0/96 |
| scale3 | 28.5 9 | 11.6/1.0/2.6 9/91 | 3.0 13 | 2.5/0.09/0.5 13/8 | 7.1 1 | 4.9/1.0/1.1 0/99 |
| scale4 | 22.2 95 | 20.2/0.2/3.2 95/1 | 2.4 46 | 2.4/0/1.0 46/0 | 9.1 69 | 26.6/0.02/0.9 57/1 |
| scale5 | 26.4 95 | 11.1/1.0/2.8 91/9 | 4.1 0 | 4.1/0/0.4 0/0 | 16.7 82 | 16.0/0.8/0.8 80/16 |
| sl1 | 13.1 49 | 13.1/0/0.08 49/0 | 1.2 0 | 1.2/0/0 0/0 | 16.8 66 | 14.0/0.4/0.4 60/17 |
| stackelberg1 | 17.2 4 | 2/1/1 0/100 | 2.2 0 | 2/1/1 0/100 | 6.0 0 | 2/1/1 0/100 |

in our implementation, active-set steps sometimes improve or harm it just slightly, being overall comparable. Recall, however, Example 4.1, which puts in evidence that active-set steps do outperform SQP in the case where MPCC-SOSC does not hold. Also, it should be kept in mind that SQP-slacks (just as standard SQP) does not possess fully justified superlinear convergence, unlike active-set steps. One can see from Tables 4.2 and 4.3 that, apart from being intended for the cases of weaker SOSC, active-set step usually does not harm at all, neither robustness nor efficiency. The

number of disregarded active-set steps remains very low. In many cases, active-set step is used just once, on the last iteration, which means that the corresponding outer-phase algorithm without active-set step could not possibly converge faster (usually it converges slower, at least in the cases of **Linearization** and **SNM-FB**). In some cases, the active-set strategy helps seriously.

REFERENCES

- [1] M. ANITESCU, *On using the elastic mode in nonlinear programming approaches to mathematical programs with complementarity constraints*, SIAM J. Optim., 15 (2005), pp. 1203–1236.
- [2] R. W. COTTLE, J.-S. PANG, AND R. E. STONE, *The Linear Complementarity Problem*, Academic Press, Boston, 1992.
- [3] A. N. DARYINA, A. F. IZMAILOV, AND M. V. SOLODOV, *Numerical results for a globalized active-set Newton method for mixed complementarity problems*, Comput. Appl. Math., 24 (2005), pp. 293–316.
- [4] F. FACCHINEI, A. FISCHER, AND C. KANZOW, *On the accurate identification of active constraints*, SIAM J. Optim., 9 (1999), pp. 14–32.
- [5] A. FISCHER, *Local behavior of an iterative framework for generalized equations with nonisolated solutions*, Math. Program., 94 (2002), pp. 91–124.
- [6] R. FLETCHER AND S. LEYFFER, *Solving mathematical programs with equilibrium constraints as nonlinear programs*, Optim. Methods Softw., 19 (2004), pp. 15–40.
- [7] R. FLETCHER, S. LEYFFER, D. RALPH, AND S. SCHOLTES, *Local convergence of SQP methods for mathematical programs with equilibrium constraints*, SIAM J. Optim., 17 (2006), pp. 259–286.
- [8] W. W. HAGER AND M. S. GOWDA, *Stability in the presence of degeneracy and error estimation*, Math. Program., 85 (1999), pp. 181–192.
- [9] A. F. IZMAILOV, *Mathematical programs with complementarity constraints: Regularity, optimality conditions, and sensitivity*, Comput. Math. Math. Phys., 44 (2004), pp. 1145–1164.
- [10] A. F. IZMAILOV AND M. V. SOLODOV, *Newton-type methods for optimization problems without constraint qualifications*, SIAM J. Optim., 15 (2004), pp. 210–228.
- [11] A. F. IZMAILOV AND M. V. SOLODOV, *On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions*, Math. Program., 117 (2009), pp. 271–304.
- [12] S. LEYFFER, *MacMPEC: AMPL collection of MPECs*, www.mcs.anl.gov/~leyffer/MacMPEC/.
- [13] G. H. LIN AND M. FUKUSHIMA, *Hybrid approach with active set identification for mathematical programs with complementarity constraints*, J. Optim. Theory Appl., 128 (2006), pp. 1–28.
- [14] T. DE LUCA, F. FACCHINEI, AND C. KANZOW, *A theoretical and numerical comparison of some semismooth algorithms for complementarity problems*, Comput. Optim. Appl., 16 (2000), pp. 173–205.
- [15] Z.-Q. LUO, J.-S. PANG, AND D. RALPH, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK, 1996.
- [16] J. V. OUTFRATA, M. KOCVARA, AND J. ZOWE, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results*, Kluwer Academic Publishers, Dordrecht, 1998.
- [17] D. RALPH, *Sequential quadratic programming for mathematical programs with linear complementarity constraints*, in Computational Techniques and Applications: CTAC95, R. L. May and A. K. Easton, eds., World Scientific, Singapore, 1996, pp. 663–668.
- [18] H. SCHEEL AND S. SCHOLTES, *Mathematical programs with complementarity constraints: Stationarity, optimality and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1–22.
- [19] P. TSENG, *Growth behavior of a class of merit functions for the nonlinear complementarity problem*, J. Optim. Theory Appl., 89 (1996), pp. 17–37.