SEMISMOOTH NEWTON METHOD FOR THE LIFTED REFORMULATION OF MATHEMATICAL PROGRAMS WITH COMPLEMENTARITY CONSTRAINTS

A. F. Izmailov[†], A. L. Pogosyan[‡], and M. V. Solodov[§]

December 2, 2009 (Revised June 7, 2010)

ABSTRACT

We consider a reformulation of mathematical programs with complementarity constraints, where by introducing an artificial variable the constraints are converted into equalities which are once but not twice differentiable. We show that the Lagrange optimality system of such a reformulation is semismooth and BD-regular at the solution under reasonable assumptions. Thus, fast local convergence can be obtained by applying the semismooth Newton method. Moreover, it turns out that the squared residual of the Lagrange system is continuously differentiable (even though the system itself is not), which opens the way for a natural globalization of the local algorithm. Preliminary numerical results are also reported.

Key words: mathematical program with complementarity constraints, semismooth Newton method, BD-regularity, second-order sufficiency, merit function.
AMS subject classifications. 90C30, 90C33, 90C55, 65K05.

---

[*] Research of the first two authors is supported by the Russian Foundation for Basic Research Grant 10-01-00251. The third author is supported in part by CNPq Grants 300513/2008-9 and 471267/2007-4, by PRONEX–Optimization, and by FAPERJ.
[†] Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: `izmaf@ccas.ru`
[‡] Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: `pogosyan85@gmail.com`
[§] IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email: `solodov@impa.br`

1 Introduction

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

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function and $G,\, H : \mathbb{R}^n \to \mathbb{R}^m$ are smooth mappings (precise smoothness requirements would be specified when needed). Additional equality and inequality constraints can be added to our problem setting without any principal difficulties. We shall consider the form of (1.1) for simplicity.

MPCC is an important example of a *mathematical program with equilibrium constraints* [17, 20]. As is well known, feasible points of MPCC inevitably violate standard constraint qualifications, and thus the problem often requires special analysis and special algorithmic developments. Concerning the latter, it should be noted that there exists some numerical evidence of good performance of the usual sequential quadratic programming (SQP) algorithms when applied to MPCC [6]. Also, [7] gives some partial theoretical justification for local superlinear convergence of SQP when applied to MPCC. However, it is very easy to provide examples satisfying all *natural* in MPCC context requirements and such that SQP does not possess superlinear convergence; see, e.g., the example in [7, Sec. 7.3], discussed also in detail in [10, Sec. 6]. Therefore, developing special algorithms, which take into account MPCC structure and have guaranteed attractive convergence properties, is still worthwhile.

The following idea, called "lifting MPCC", had been proposed in [28]. Consider the set in the space $\mathbb{R}^2$ of variables $(a,\, b)$ defined by the basic complementarity condition:

$$D = \{(a,\, b) \in \mathbb{R}^2 \mid a \geq 0,\ b \geq 0,\ ab = 0\}.$$

This set is "nonsmooth", in the sense that it has a kink at the origin. Introducing an artificial variable $c \in \mathbb{R}$, consider a smooth curve S in the space $\mathbb{R}^3$ of variables $(a,\, b,\, c)$ such that the projection of S onto the plane $(a,\, b)$ coincides with the set D. This can be done, for example, as follows:

$$S = \{(a,\, b,\, c) \in \mathbb{R}^3 \mid a = (-\min\{0,\, c\})^s,\ b = (\max\{0,\, c\})^s\}, \quad s > 1.$$

In [28], it is suggested to use the power $s = 3$. This leads to a reformulation of the original problem (1.1) given by

$$\min\ f(x) \quad \text{s.t.} \quad -(\min\{0,\, y\})^3 - G(x) = 0, \quad (\max\{0,\, y\})^3 - H(x) = 0, \tag{1.2}$$

where the operations of taking minimum, maximum, and applying power are understood component-wise. As is easy to see, a point $\bar{x} \in \mathbb{R}^n$ is a (local) solution of (1.1) if, and only if, the point $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ is a (local) solution of (1.2) with $\bar{y}$ uniquely defined by $\bar{x}$. At the same time, the constraints in the reformulation (1.2) are twice differentiable equalities, which appear much simpler than the original complementarity constraints in (1.1). Of course, this does not come for free – a closer look reveals that the Jacobian of the Lagrange optimality system of the reformulation (1.2) is inevitably degenerate whenever the lower-level strict complementarity does not hold at $\bar{x}$, i.e., if $G_i(\bar{x}) = H_i(\bar{x}) = 0$ for some $i \in$

$\{1, \ldots, m\}$. In fact, at any feasible point of (1.2) the derivatives of the Lagrangian with respect to y_i for indices i that fail strict complementarity are zero, and thus the Jacobian matrix has zero rows. To deal with degeneracy, in [28] an approach somewhat similar in spirit to relaxation/regularization schemes for MPCC [27, 25] is suggested. Geometrically, degeneracy arising in the lifted reformulation means that the tangent of the smooth curve S (regardless of the power s used in its definition) at the point (0, 0, 0) is always vertical. It is proposed in [28] to add to the objective function of (1.2) a certain regularization term (linear in y) which shifts the point corresponding to the solution from (0, 0, 0) to some other nearby point on S where the tangent to S is not vertical. This point is obtained by solving the regularized problem and is then used to initialize the method for solving (1.2) itself, the hope being that such "pre-processing" may prevent the method from getting stuck at nonoptimal weakly stationary points. However, one still needs to solve problem (1.2), whose Lagrange optimality system is likely degenerate. Also, the numerical experience reported in [28] indicates that the starting point used to initialize the method for the regularized version of (1.2) must already be feasible, which is a limitation in general.

The issue of degeneracy of smoothed lifted MPCC is similar in nature to degeneracy of smooth equation-based reformulations of nonlinear complementarity problems (NCP) in the absence of strict complementarity [12]; see also [8] for a detailed discussion. At the same time, it is known that nonsmooth equation-based reformulations of NCP, such as based on the Fischer-Burmeister function [5, 1] or the min-function, can have appropriate regularity properties without strict complementarity and are thus generally preferred for constructing Newton-type methods for NCP. Drawing on this experience for NCP, in this work we propose to use, instead of power $s = 3$ which leads to degeneracy of the Lagrange optimality system of the lifted MPCC reformulation, the power $s = 2$ which leads to its nonsmoothness but can be expected to have better regularity properties. Specifically, consider the reformulation of (1.1) given by

$$\min\ f(x) \quad \text{s.t.} \quad (\min\{0,\, y\})^2 - G(x) = 0, \quad (\max\{0,\, y\})^2 - H(x) = 0. \tag{1.3}$$

As we shall show, nonsmoothness of the Lagrange optimality system of (1.3) is *structured* and allows application of the semismooth Newton method [14, 15, 22, 23] under reasonable assumptions. Moreover, it turns out that the squared residual of the Lagrange optimality system of (1.3) is actually continuously differentiable, even though the system itself is not. This opens the way to a natural globalization of the local semismooth Newton method. The latter is again similar to the nonsmooth Fischer-Burmeister equality-based reformulation of NCP, for which the squared residual becomes smooth and can be used for globalization [5, 1, 11, 2].

The rest of the paper is organized as follows. In Section 2 we collect some preliminary material and terminology from MPCC literature, needed for further developments. Section 3 contains the statement of the semismooth Newton method for the lifted reformulation of MPCC and the proof of its local superlinear/quadratic convergence. A comparison with some alternatives is also given in this section. Globalization issues are discussed in Section 4 and numerical results are reported in Section 5.

Some final words about our notation. For u, $v \in \mathbb{R}^q$, by $\langle u, v\rangle$ we denote the Euclidean inner product between u and v, and by $\|\cdot\|$ the associated norm. If $u \in \mathbb{R}^q$ and $I \subset \{1, \ldots, q\}$, then u_I stands for the subvector of u with components u_i, $i \in I$. By $\mathrm{diag}(u)$ we define the quadratic $q \times q$-matrix with the components of the vector $u \in \mathbb{R}^q$ on the diagonal and zeroes elsewhere. For an arbitrary matrix A, we denote by A^{T} its transpose. Finally, we say that $\Phi : \mathbb{R}^q \to \mathbb{R}^r$ is locally Lipschitz-continuous with respect to the point $\bar{u} \in \mathbb{R}^q$ if $\|\Phi(u) - \Phi(\bar{u})\| \leq \ell\|u - \bar{u}\|$ for some $\ell > 0$ and all $u \in \mathbb{R}^q$ close enough to $\bar{u}$.

2 Some basic facts about MPCC and lifted MPCC

Our notation and definitions are standard in MPCC literature, e.g., [26, 7, 9]. Let $\bar{x} \in \mathbb{R}^n$ be a feasible point of problem (1.1). We define the sets of indices

$$\begin{array}{c} I_G = I_G(\bar{x}) = \{i = 1, \ldots, m \mid G_i(\bar{x}) = 0\}, \\ I_H = I_H(\bar{x}) = \{i = 1, \ldots, m \mid H_i(\bar{x}) = 0\}, \\ I_0 = I_G \cap I_H, \end{array} \tag{2.1}$$

where $I_G \cup I_H = \{1, \ldots, m\}$, and I_0 is called the set of *degenerate indices*. If $I_0 = \emptyset$, we say that *lower-level strict complementarity* holds at $\bar{x}$. This condition, however, is considered as restrictive in MPCC literature. We emphasize that it would not be assumed anywhere in our developments.

The special *MPCC-Lagrangian* for problem (1.1) is given by

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

where $x \in \mathbb{R}^n$ and $\mu = (\mu^G, \mu^H) \in \mathbb{R}^m \times \mathbb{R}^m$. A point $\bar{x}$ which is feasible in (1.1) is called *weakly stationary* if there exists $\bar{\mu} = (\bar{\mu}^G, \bar{\mu}^H) \in \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\frac{\partial \mathcal{L}}{\partial x}(\bar{x}, \bar{\mu}) = 0, \quad (\bar{\mu}_G)_{I_H \setminus I_G} = 0, \quad (\bar{\mu}_H)_{I_G \setminus I_H} = 0. \tag{2.2}$$

The point $\bar{x}$ is called *strongly stationary* if, in addition to (2.2),

$$(\bar{\mu}_G)_{I_0} \geq 0, \quad (\bar{\mu}_H)_{I_0} \geq 0. \tag{2.3}$$

When conditions (2.2) and (2.3) hold, $\bar{\mu}$ is called an *MPCC-multiplier* associated to the strongly stationary point $\bar{x}$ of problem (1.1).

It is said that the special *MPCC linear independence constraint qualification* (MPCC-LICQ) holds at a feasible point $\bar{x}$ if

$$G_i'(\bar{x}),\ i \in I_G, \quad H_i'(\bar{x}),\ i \in I_H \quad \text{are linearly independent.} \tag{2.4}$$

If a local solution $\bar{x}$ of problem (1.1) satisfies MPCC-LICQ, then $\bar{x}$ is a strongly stationary point and the associated MPCC-multiplier $\bar{\mu}$ is unique [26, Theorem 2].

The usual critical cone for problem (1.1) at the point $\bar{x}$ is given by

$$C(\bar{x}) = \left\{ \xi \in \mathbb{R}^n \;\middle|\; \begin{array}{l} G'_{I_G \setminus I_H}(\bar{x})\xi = 0,\ H'_{I_H \setminus I_G}(\bar{x})\xi = 0,\ G'_{I_0}(\bar{x})\xi \geq 0,\ H'_{I_0}(\bar{x})\xi \geq 0, \\ \langle f'(\bar{x}), \xi \rangle \leq 0 \end{array} \right\}. \tag{2.5}$$

The following cone takes also into account the second-order information about the last constraint in (1.1):

$$C_2(\bar{x}) = \left\{ \xi \in \mathbb{R}^n \;\middle|\; \begin{array}{l} G'_{I_G \setminus I_H}(\bar{x})\xi = 0,\; H'_{I_H \setminus I_G}(\bar{x})\xi = 0,\; G'_{I_0}(\bar{x})\xi \geq 0,\; H'_{I_0}(\bar{x})\xi \geq 0, \\ \langle f'(\bar{x}),\, \xi\rangle \leq 0,\; \langle G'_i(\bar{x}),\, \xi\rangle \langle H'_i(\bar{x}),\, \xi\rangle = 0,\; i \in I_0 \end{array} \right\}. \tag{2.6}$$

Evidently, it holds that

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

If $\bar{x}$ is a strongly stationary point, then for any associated MPCC-multiplier $\bar{\mu} = (\bar{\mu}_G,\, \bar{\mu}_H)$ it holds that

$$C(\bar{x}) = \left\{ \xi \in \mathbb{R}^n \;\middle|\; \begin{array}{l} G'_{I_G \setminus I_H}(\bar{x})\xi = 0,\; H'_{I_H \setminus I_G}(\bar{x})\xi = 0,\; G'_{I_0}(\bar{x})\xi \geq 0,\; H'_{I_0}(\bar{x})\xi \geq 0, \\ (\bar{\mu}_G)_i \langle G'_i(\bar{x}),\, \xi\rangle = 0,\; (\bar{\mu}_H)_i \langle H'_i(\bar{x}),\, \xi\rangle = 0,\; i \in I_0 \end{array} \right\} \tag{2.7}$$

and

$$C_2(\bar{x}) = \left\{ \xi \in \mathbb{R}^n \;\middle|\; \begin{array}{l} G'_{I_G \setminus I_H}(\bar{x})\xi = 0,\; H'_{I_H \setminus I_G}(\bar{x})\xi = 0,\; G'_{I_0}(\bar{x})\xi \geq 0,\; H'_{I_0}(\bar{x})\xi \geq 0, \\ (\bar{\mu}_G)_i \langle G'_i(\bar{x}),\, \xi\rangle = 0,\; (\bar{\mu}_H)_i \langle H'_i(\bar{x}),\, \xi\rangle = 0, \\ \langle G'_i(\bar{x}),\, \xi\rangle \langle H'_i(\bar{x}),\, \xi\rangle = 0,\; i \in I_0 \end{array} \right\}, \tag{2.8}$$

respectively.

The special *MPCC second-order sufficient condition* (MPCC-SOSC) is said to hold at a strongly stationary point $\bar{x}$ with an associated MPCC-multiplier $\bar{\mu}$, if

$$\left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}. \tag{2.9}$$

The weaker *piece-wise second-order sufficient condition* consists of saying that

$$\left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C_2(\bar{x}) \setminus \{0\}. \tag{2.10}$$

Condition (2.10) (and, consequently, (2.9)) is indeed sufficient for local optimality of $\bar{x}$, see [9]. In general, (2.10) is a more natural condition than (2.9). This is because condition (2.10) with the strict inequality replaced by non-strict is a necessary condition for optimality under MPCC-LICQ [26, Theorem 7], while (2.9) does not have any necessary counterpart.

We say that the *upper-level strict complementarity condition* (ULSCC) holds for some MPCC-multiplier $\bar{\mu}$ associated to $\bar{x}$ if

$$(\bar{\mu}_G)_{I_0} > 0, \quad (\bar{\mu}_H)_{I_0} > 0. \tag{2.11}$$

Under ULSCC, it holds that $C(\bar{x}) = C_2(\bar{x}) = K(\bar{x})$, where

$$K(\bar{x}) = \{\xi \in \mathbb{R}^n \mid G'_{I_G}(\bar{x})\xi = 0,\; H'_{I_H}(\bar{x})\xi = 0\}. \tag{2.12}$$

In that case, conditions (2.9) and (2.10) become equivalent and can be stated as

$$\left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in K(\bar{x}) \setminus \{0\}. \tag{2.13}$$

Let us now turn our attention to the lifted MPCC reformulation (1.3). Note first that the value $\bar{y}$ of the artificial variable y that corresponds to any given feasible point $\bar{x}$ of the original problem (1.1) is uniquely defined: the point $(\bar{x}, \bar{y})$ is feasible in (1.3) if, and only if,

$$\bar{y}_{I_H \setminus I_G} = -(G_{I_H \setminus I_G}(\bar{x}))^{1/2}, \quad \bar{y}_{I_G \setminus I_H} = (H_{I_G \setminus I_H}(\bar{x}))^{1/2}, \quad \bar{y}_{I_0} = 0. \tag{2.14}$$

Furthermore, it is immediate that $\bar{x}$ is a (local) solution of the original problem (1.1) if, and only if, $(\bar{x}, \bar{y})$ is a (local) solution of the lifted MPCC reformulation (1.3). In addition, it is also easy to see (as in [28]) that MPCC-LICQ at a point $\bar{x}$ feasible in (1.1) is equivalent to linear independence of constraints gradients of (1.3) at the point $(\bar{x}, \bar{y})$.

Let us define the usual Lagrangian of the lifted problem (1.3):

$$L(x, y, \lambda) = f(x) + \langle \lambda_G, (\min\{0, y\})^2 - G(x)\rangle + \langle \lambda_H, (\max\{0, y\})^2 - H(x)\rangle,$$

where $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ and $\lambda = (\lambda_G, \lambda_H) \in \mathbb{R}^m \times \mathbb{R}^m$. The Lagrange optimality system characterizing stationary points of (1.3) and the associated multipliers is given by

$$\frac{\partial L}{\partial x}(x, y, \lambda) = 0, \quad \frac{\partial L}{\partial y}(x, y, \lambda) = 0, \quad (\min\{0, y\})^2 - G(x) = 0, \quad (\max\{0, y\})^2 - H(x) = 0,$$

where

$$\frac{\partial L}{\partial x}(x, y, \lambda) = \frac{\partial \mathcal{L}}{\partial x}(x, \lambda), \tag{2.15}$$

$$\frac{\partial L}{\partial y_i}(x, y, \lambda) = 2(\lambda_G)_i \min\{0, y_i\} + 2(\lambda_H)_i \max\{0, y_i\}, \quad i = 1, \ldots, m. \tag{2.16}$$

Observe that for any $i = 1, \ldots, m$, the right-hand side in (2.16) is not differentiable at points (x, y, λ) such that $y_i = 0$.

The following correspondence between stationary points and multipliers for the original problem (1.1) and its lifted reformulation (1.3) can be checked by direct verification, as in [28].

Proposition 2.1 *Let f, G and H be differentiable at a point $\bar{x} \in \mathbb{R}^n$ which is feasible in problem* (1.1).

If $\bar{x}$ is a strongly stationary point of (1.1) *and $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H)$ is an associated MPCC-multiplier, then the point $(\bar{x}, \bar{y})$ with $\bar{y}$ given by* (2.14) *is a stationary point of problem* (1.3) *and $\bar{\lambda} = \bar{\mu}$ is an associated Lagrange multiplier.*

Conversely, if $(\bar{x}, \bar{y})$ is a stationary point of problem (1.3)*, then $\bar{x}$ is a weakly stationary point of problem* (1.1). *In addition, if there exists a Lagrange multiplier $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H)$ associated to $(\bar{x}, \bar{y})$ and such that $(\bar{\lambda}_G)_{I_0} \geq 0$ and $(\bar{\lambda}_H)_{I_0} \geq 0$, then $\bar{x}$ is a strongly stationary point of problem* (1.1) *and $\bar{\mu} = \bar{\lambda}$ is an associated MPCC-multiplier.*

3 Semismooth Newton method for lifted MPCC

We start with reminding the reader some basic facts of nonsmooth analysis and the semismooth Newton method (SNM), see [14, 15, 22, 23] and [4, Chapter 7].

Consider a mapping $\Phi: \mathbb{R}^q \to \mathbb{R}^r$ which is locally Lipschitz-continuous around a point $u \in \mathbb{R}^q$. The *B-differential* of Φ at $u \in \mathbb{R}^q$ is the set

$$\partial_B \Phi(u) = \{\Lambda \in \mathbb{R}^{r \times q} \mid \exists \{u^k\} \subset \mathcal{D}_\Phi: \{u^k\} \to u,\ \{\Phi'(u^k)\} \to \Lambda\ (k \to \infty)\},$$

where $\mathcal{D}_\Phi$ is the set of points at which Φ is differentiable (under the stated assumptions Φ is differentiable almost everywhere around u). Then the *Clarke generalized Jacobian* is given by

$$\partial \Phi(u) = \operatorname{conv} \partial_B \Phi(u),$$

where $\operatorname{conv} X$ stands for the convex hull of the set X.

Furthermore, Φ is said to be *semismooth* [19] at $u \in \mathbb{R}^q$ if it is locally Lipschitz-continuous around u, directionally differentiable at u in every direction, and satisfies the condition

$$\sup_{\Lambda \in \partial \Phi(u+v)} \|\Phi(u+v) - \Phi(u) - \Lambda v\| = o(\|v\|).$$

If the stronger condition

$$\sup_{\Lambda \in \partial \Phi(u+v)} \|\Phi(u+v) - \Phi(u) - \Lambda v\| = O(\|v\|^2)$$

holds, then Φ is said to be *strongly semismooth* at u.

For our purposes, the following basic facts would be needed. If $\Phi_1,\ \Phi_2: \mathbb{R}^q \to \mathbb{R}^r$ are (strongly) semismooth at $u \in \mathbb{R}^q$, then so are the mappings $(\Phi_1(\cdot) + \Phi_2(\cdot))$, $\langle \Phi_1(\cdot),\ \Phi_2(\cdot)\rangle$, $\min\{\Phi_1(\cdot),\ \Phi_2(\cdot)\}$, $\max\{\Phi_1(\cdot),\ \Phi_2(\cdot)\}$, and $(\Phi_1(\cdot),\ \Phi_2(\cdot))$.

Recall finally that $\Phi: \mathbb{R}^q \to \mathbb{R}^q$ is said to be *BD-regular* at a point $\bar{u} \in \mathbb{R}^q$ if all the matrices $\Lambda \in \partial_B \Phi(\bar{u})$ are nonsingular.

The SNM iterative scheme for the equation

$$\Phi(u) = 0, \tag{3.1}$$

where $\Phi: \mathbb{R}^q \to \mathbb{R}^q$, is then given by

$$\Lambda_k(u^{k+1} - u^k) = -\Phi(u^k), \quad \Lambda_k \in \partial_B \Phi(u^k), \quad k = 0,\ 1,\ \ldots. \tag{3.2}$$

The following is the basic local convergence result for SNM.

Theorem 3.1 *Let $\Phi: \mathbb{R}^q \to \mathbb{R}^q$ be semismooth at $\bar{u} \in \mathbb{R}^q$. Let $\bar{u}$ be a solution of the equation* (3.1) *such that Φ is BD-regular at $\bar{u}$.*

Then any starting point $u^0 \in \mathbb{R}^q$ sufficiently close to $\bar{u}$ uniquely defines SNM iterative sequence $\{u^k\}$ satisfying (3.2)*, and this sequence converges to $\bar{u}$. The rate of convergence is superlinear, and it is quadratic if Φ is strongly semismooth at $\bar{u}$.*

We shall next consider SNM applied to the Lagrange optimality system of the lifted MPCC reformulation (1.3). Taking into account the relation (2.15), this system takes the form of the equation (3.1) if we define $\Phi: \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m) \to \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ by

$$\Phi(u) = \begin{pmatrix} \frac{\partial \mathcal{L}}{\partial x}(x, \lambda) \\ \frac{\partial \mathcal{L}}{\partial y}(x, y, \lambda) \\ (\min\{0, y\})^2 - G(x) \\ (\max\{0, y\})^2 - H(x) \end{pmatrix}, \tag{3.3}$$

where $u = (x, y, \lambda)$, $\lambda = (\lambda_G, \lambda_H)$.

Using (2.16) and (3.3), by direct calculations it follows that the B-differential of Φ at an arbitrary point $u = (x, y, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ consists of all matrices of the form

$$\Lambda = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial x^2}(x, \lambda) & 0 & -(G'(x))^{\mathrm{T}} & -(H'(x))^{\mathrm{T}} \\ 0 & 2A(y, \lambda) & 2B_{\min}(y) & 2B_{\max}(y) \\ -G'(x) & 2B_{\min}(y) & 0 & 0 \\ -H'(x) & 2B_{\max}(y) & 0 & 0 \end{pmatrix}, \tag{3.4}$$

where

$$A(y, \lambda) = \mathrm{diag}(a(y, \lambda)), \quad B_{\min}(y) = \mathrm{diag}(\min\{0, y\}), \quad B_{\max}(y) = \mathrm{diag}(\max\{0, y\}), \tag{3.5}$$

and the vector $a(y, \lambda) \in \mathbb{R}^m$ is defined by

$$a_i = \begin{cases} (\lambda_G)_i, & \text{if } y_i < 0, \\ (\lambda_G)_i \text{ or } (\lambda_H)_i, & \text{if } y_i = 0, \\ (\lambda_H)_i, & \text{if } y_i > 0, \end{cases} \quad i = 1, \ldots, m. \tag{3.6}$$

SNM for the Lagrange optimality system of the lifted MPCC reformulation (1.3) is therefore given by the following

Algorithm 3.1 Choose $u^0 = (x^0, y^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ and set $k = 0$.

1. Compute a matrix $\Lambda_k = \Lambda$ according to (3.4)–(3.6) with $(x, y, \lambda) = (x^k, y^k, \lambda^k)$. Compute $u^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ as a solution of the linear system
$$\Lambda_k u = \Lambda_k u^k - \Phi(u^k), \tag{3.7}$$
where Φ is defined in (3.3).

2. Increase k by 1 and go to Step 1.

Let $\bar{x}$ be a strongly stationary point of the original problem (1.1) and let $\bar{\mu}$ be an associated MPCC-multiplier. Let $\bar{y}$ be given by (2.14). According to Theorem 3.1, local superlinear (quadratic) convergence of Algorithm 3.1 to the solution $\bar{u} = (\bar{x}, \bar{y}, \bar{\mu})$ of the equation (3.1) would be established if we show (strong) semismoothness and BD-regularity of Φ at $\bar{u}$. The semismoothness properties easily follow from calculus of (strongly) semismooth mappings summarized above. We thus omit the proof.

Proposition 3.1 *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $G,\, H : \mathbb{R}^n \to \mathbb{R}^m$ *be differentiable around the point* $\bar{x} \in \mathbb{R}^n$*, with their derivatives being semismooth at this point.*

Then for any $\lambda = (\lambda_G,\, \lambda_H) \in \mathbb{R}^m \times \mathbb{R}^m$ *and* $y \in \mathbb{R}^m$*, the mapping* Φ *defined in* (3.3) *is semismooth at the point* $(\bar{x},\, y,\, \lambda)$*. Moreover, if* f*,* G *and* H *are twice differentiable around* $\bar{x}$*, with their derivatives being locally Lipschitz-continuous with respect to* $\bar{x}$*, then* Φ *is strongly semismooth at* $(\bar{x},\, y,\, \lambda)$*.*

Now, according to (3.4)–(3.6), and taking also into account (2.3) and (2.14), we obtain that the B-differential of Φ at the point $\bar{u} = (\bar{x},\, \bar{y},\, \bar{\mu})$ consists of all matrices of the form

$$\Lambda = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\mu}) & 0 & -(G'(\bar{x}))^{\mathrm{T}} & -(H'(\bar{x}))^{\mathrm{T}} \\ 0 & 2A & 2B_{\min} & 2B_{\max} \\ -G'(\bar{x}) & 2B_{\min} & 0 & 0 \\ -H'(\bar{x}) & 2B_{\max} & 0 & 0 \end{pmatrix}, \tag{3.8}$$

where

$$A = \mathrm{diag}(a), \quad B_{\min} = \mathrm{diag}(b_{\min}), \quad B_{\max} = \mathrm{diag}(b_{\max}), \tag{3.9}$$

with the vector $a \in \mathbb{R}^m$ given by

$$a_i = \begin{cases} 0, & \text{if } i \in \{1,\, \ldots,\, m\} \setminus I_0, \\ (\bar{\mu}_G)_i \text{ or } (\bar{\mu}_H)_i, & \text{if } i \in I_0, \end{cases} \tag{3.10}$$

and the vectors $b_{\min}$ and $b_{\max}$ given by

$$(b_{\min})_{I_H \setminus I_G} = -(G_{I_H \setminus I_G}(\bar{x}))^{1/2}, \quad (b_{\min})_{I_G} = 0, \tag{3.11}$$

$$(b_{\max})_{I_G \setminus I_H} = (H_{I_G \setminus I_H}(\bar{x}))^{1/2},\ (b_{\max})_{I_H} = 0. \tag{3.12}$$

We next show that the mapping Φ is BD-regular under reasonable assumptions.

Proposition 3.2 *Let* f*,* G *and* H *be twice differentiable at the point* $\bar{x} \in \mathbb{R}^n$ *which is strongly stationary for problem* (1.1) *and satisfies MPCC-LICQ* (2.4)*. Let* $\bar{\mu}$ *be the (unique) associated MPCC-multiplier. Assume finally that ULSCC* (2.11) *and the second-order sufficient condition* (2.13) *hold.*

Then Φ *defined by* (3.3) *is BD-regular at* $\bar{u} = (\bar{x},\, \bar{y},\, \bar{\mu})$*, where* $\bar{y}$ *is given by* (2.14)*.*

Proof. Suppose that for some matrix $\Lambda \in \partial\Phi_B(\bar{u})$ and some $\xi \in \mathbb{R}^n$, $\eta \in \mathbb{R}^m$, $\zeta_G \in \mathbb{R}^m$ and $\zeta_H \in \mathbb{R}^m$ it holds that $\Lambda v = 0$, where $v = (\xi,\, \eta,\, \zeta_G,\, \zeta_H)$. According to (3.8)–(3.12), taking into account also (2.11), we then conclude that

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\mu})\xi - (G'(\bar{x}))^{\mathrm{T}}\zeta_G - (H'(\bar{x}))^{\mathrm{T}}\zeta_H = 0, \tag{3.13}$$

$$(\zeta_G)_{I_H \setminus I_G} = 0, \quad (\zeta_H)_{I_G \setminus I_H} = 0, \quad \eta_{I_0} = 0, \tag{3.14}$$

$$-\langle G_i'(\bar{x}),\, \xi \rangle - 2(G_i(\bar{x}))^{1/2}\eta_i = 0,\ i \in I_H \setminus I_G, \quad G_{I_G}'(\bar{x})\xi = 0, \tag{3.15}$$

$$-\langle H_i'(\bar{x}),\, \xi\rangle + 2(H_i(\bar{x}))^{1/2}\eta_i = 0,\ i \in I_G \setminus I_H, \quad H_{I_H}'(\bar{x})\xi = 0. \tag{3.16}$$

The second relations in (3.15) and (3.16) mean that $\xi \in K(\bar{x})$, see (2.12). Moreover, multiplying both sides of the equality (3.13) by ξ and using the first two relations in (3.14) and the second relations in (3.15) and (3.16), we obtain that

$$\begin{aligned} 0 &= \left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle - \langle \zeta_G,\, G'(\bar{x})\xi\rangle - \langle \zeta_H,\, H'(\bar{x})\xi\rangle \\ &= \left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle. \end{aligned}$$

Hence, by (2.13), we conclude that $\xi = 0$.

Substituting now $\xi = 0$ into (3.13) and using the first two relations in (3.14), we have that

$$(G_{I_G}'(\bar{x}))^{\mathrm{T}}(\zeta_G)_{I_G} + (H_{I_H}'(\bar{x}))^{\mathrm{T}}(\zeta_H)_{I_H} = 0.$$

From the latter and (2.4), it follows that

$$(\zeta_G)_{I_G} = 0, \quad (\zeta_H)_{I_H} = 0. \tag{3.17}$$

In addition, using $\xi = 0$ in the first relations of (3.15) and (3.16), we have that

$$\eta_{I_H \setminus I_G} = 0, \quad \eta_{I_G \setminus I_H} = 0. \tag{3.18}$$

Combining (3.14), (3.17) and (3.18) gives that $\eta = 0$, $\zeta_G = 0$, $\zeta_H = 0$, i.e., $v = 0$.

We have thus shown that any matrix Λ in question has only zero in its null space. It follows that all the matrices are nonsingular. ∎

Given Theorem 3.1 and Propositions 3.1 and 3.2, we immediately obtain local convergence and rate of convergence for Algorithm 3.1.

Theorem 3.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $G,\, H : \mathbb{R}^n \to \mathbb{R}^m$ be twice differentiable in a neighbourhood of a strongly stationary point $\bar{x} \in \mathbb{R}^n$ of problem* (1.1)*, and let their second derivatives be continuous at $\bar{x}$. Let MPCC-LICQ* (2.4) *be satisfied at $\bar{x}$, and let $\bar{\mu}$ be the (unique) MPCC-multiplier associated to $\bar{x}$. Assume, finally, that ULSCC* (2.11) *and the second-order sufficient condition* (2.13) *are satisfied.*

Then any point $(x^0,\, y^0,\, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ close enough to $(\bar{x},\, \bar{y},\, \bar{\mu})$ uniquely defines SNM iterative sequence of Algorithm 3.1*, and this sequence converges to $(\bar{x},\, \bar{y},\, \bar{\mu})$. The rate of convergence is superlinear, and if the second derivatives of f, G and H are locally Lipschitz-continuous with respect to $\bar{x}$ then the rate is quadratic.*

We next comment on how our local convergence result in Theorem 3.2 compares to some alternatives in MPCC literature.

As already discussed above, using the lifted reformulation (1.2), suggested in [28], gives a Lagrange optimality system which is smooth but inherently degenerate (at least in the absence of lower-level strict complementarity). Nevertheless, this degeneracy is again structured and can be tackled, in principle, by the tools for solving degenerate equations and reformulations of complementarity problems developed in [8]. But careful consideration of the approach of [8] applied to the lifted reformulation (1.2) shows that it would require the same assumptions as those for Algorithm 3.1 in Theorem 3.2, while the approach itself is quite a bit more complicated. In addition, methods in [8] do not come with natural globalization strategies. In Section 4 below we shall show that Algorithm 3.1 allows natural globalization by linesearch for the squared residual of the Lagrange optimality system for (1.3).

Another possibility is to apply the usual SQP directly to the original problem (1.1), perhaps introducing slacks so that the complementarity condition is stated for simpler bound constraints. In practice, this approach appears to work rather well [6]. There is also some local analysis showing superlinear convergence of SQP applied to MPCC [7], although it is not quite complete. In any case, the assumptions used in [7] are stronger than those for Algorithm 3.1 in Theorem 3.2. Specifically, in addition to the hypotheses of Theorem 3.2, in [7, Theorem 5.7] it is assumed that all MPCC multipliers are non-zero (Assumption [A4]), that the active-set QP solver applied to SQP subproblems always picks a linearly independent basis (Assumption [A5]) and, perhaps most importantly, that the exact complementarity always holds from some iteration on (Assumption [A6]). Strict complementarity is dropped in [7, Theorem 5.14] but it is additionally assumed that the constraints of SQP subproblems remain consistent along iterations (Assumption [A7]). (Also, it is not clear which dual solution μ^* [7, Theorem 5.14] refers to.) Note also that, unlike SQP, Algorithm 3.1 solves linear systems of equations rather than quadratic subproblems. On the other hand, SQP comes with well-developed globalization strategies which may be preferable to the one based on linesearch for the squared residual of the optimality system suggested in Section 4.

Finally, local superlinear convergence of piecewise-SQP in [17, 24], and of an active-set Newton method in [9], had been shown under assumptions weaker than those for Algorithm 3.1 in Theorem 3.2. Specifically, those methods do not require ULSSC (2.11). But, just as in the case of [8], the methods in question come without ready-to-use globalization strategies, while a reasonable globalization scheme for Algorithm 3.1 would be proposed next.

We complete this section with an observation that the mapping Φ defined in (3.3) is piecewise smooth, and SNM specified in Algorithm 3.1 can be interpreted as the piecewise Newton method developed in [13]. However, the semismooth approach is more general and could possibly be used for other (semismooth but not piecewise smooth) lifting MPCC reformulations.

4 Globalization of the local method

In this section, we propose to globalize the local SNM of Algorithm 3.1 by introducing linesearch for the *merit function* $\varphi\colon \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m) \to \mathbb{R}$,

$$\varphi(u) = \frac{1}{2}\|\Phi(u)\|^2, \tag{4.1}$$

where Φ is defined in (3.3). It turns out that this merit function φ is continuously differentiable, even though Φ itself is not. Moreover, the gradient of φ is explicitly computable using any element of the B-differential of Φ. It is interesting to point out that those properties are similar to the popular equation-based reformulation of NCP based on the Fischer–Burmeister function and its squared residual [5, 1, 11, 2].

Proposition 4.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $G, H : \mathbb{R}^n \to \mathbb{R}^m$ be twice differentiable at a point $x \in \mathbb{R}^n$.*

Then for any $y \in \mathbb{R}^m$ and $\lambda = (\lambda_G, \lambda_H) \in \mathbb{R}^m \times \mathbb{R}^m$, the function φ defined by (4.1) and (3.3) is differentiable at the point $u = (x, y, \lambda)$ and it holds that

$$\varphi'(u) = \Lambda\Phi(u) \quad \forall\, \Lambda \in \partial_B \Phi(u). \tag{4.2}$$

Moreover, if f, G and H are twice continuously differentiable on $\mathbb{R}^n$, then the function φ is continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$.

Proof. Nonsmoothness of φ can only be induced by the components of Φ that correspond to partial derivatives of L with respect to y (see (3.3)); all the other components of Φ are sufficiently smooth under the stated assumptions.

Observe that for any $t \in \mathbb{R}$ it holds that

$$\min\{0, t\} \max\{0, t\} = 0.$$

Therefore, for each $i = 1, \ldots, m$, from (2.16) it follows that

$$\left(\frac{\partial L}{\partial y_i}(x, y, \lambda)\right)^2 = 4((\lambda_G)_i^2(\min\{0, y_i\})^2 + (\lambda_H)_i^2(\max\{0, y_i\})^2), \tag{4.3}$$

where the right-hand side is a differentiable function in the variables $y \in \mathbb{R}^m$ and $\lambda = (\lambda_G, \lambda_H) \in \mathbb{R}^m \times \mathbb{R}^m$. This shows that φ has the announced differentiability properties.

Furthermore, from (4.3) it follows that

$$\begin{aligned}\left(\frac{1}{2}\left\|\frac{\partial L}{\partial y}(x, y, \lambda)\right\|^2\right)' &= \begin{pmatrix} 0 \\ 4((\lambda_G)_1^2(\min\{0, y_1\}) + (\lambda_H)_1^2(\max\{0, y_1\})) \\ \cdots \\ 4((\lambda_G)_1^2(\min\{0, y_m\}) + (\lambda_H)_m^2(\max\{0, y_m\})) \\ 4(\lambda_G)_1(\min\{0, y_1\})^2 \\ \cdots \\ 4(\lambda_G)_m(\min\{0, y_m\})^2 \\ 4(\lambda_H)_1(\max\{0, y_1\})^2 \\ \cdots \\ 4(\lambda_H)_m(\max\{0, y_m\})^2 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 2A(y, \lambda) \\ 2B_{\min}(y) \\ 2B_{\max}(y) \end{pmatrix} \frac{\partial L}{\partial y}(x, y, \lambda), \end{aligned} \tag{4.4}$$

where the last equality is by (2.16), (3.5) and (3.6). Differentiating the other parts of φ and combining the result with (4.4) and with (3.3)–(3.6), gives the equality (4.2). ∎

Recall that according to (3.4)–(3.6), all the matrices $\Lambda \in \partial_B \Phi(u)$ are symmetric at any $u \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$. Using this fact, as well as (4.2), for any such matrix and for any direction $v \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ computed as a solution of the linear system

$$\Lambda v = -\Phi(u),$$

it holds that

$$\langle \varphi'(u), v \rangle = \langle \Lambda \Phi(u), v \rangle = \langle \Phi(u), \Lambda v \rangle = -\langle \Phi(u), \Phi(u) \rangle = -\|\Phi(u)\|^2 = -2\varphi(u). \tag{4.5}$$

In particular, if the point u is not a solution of the equation (3.1) or, equivalently, is not a global minimizer of the function φ, then v is a descent direction for φ at the point u. This immediately suggests a natural globalization strategy for the local SNM in Algorithm 3.1. We next state our globalized algorithm, which is similar to the method in [11] for the reformulation of NCP based on the Fischer–Burmeister function. The latter, however, assumes existence and boundedness of Newton directions along the iterative process.

Algorithm 4.1 Choose parameters $\varepsilon \in (0, 1/2)$, $\tau \in (0, 1)$, $M > 0$, $\theta > 0$, and a starting point $u^0 = (x^0, y^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$. Set $k = 0$.

1. If $\Phi(u^k) = 0$, stop. Otherwise, compute some matrix $\Lambda_k = \Lambda$ according to the formulas (3.4)–(3.6) with $(x, y, \lambda) = (x^k, y^k, \lambda^k)$. Compute $\tilde{u}^{k+1} \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ as a solution of the linear system (3.7), where Φ is defined in (3.3). If $\tilde{u}^{k+1}$ exists and
$$\|\tilde{u}^{k+1} - u^k\| \leq \max\{M, 1/(\varphi(u^k))^\theta\}, \tag{4.6}$$
then set $v^k = \tilde{u}^{k+1} - u^k$; otherwise set $v^k = -\Lambda_k \Phi(u^k)$. If $v^k = 0$, stop.
2. Compute the stepsize value α_k by the Armijo rule: $\alpha_k = \tau^j$, where j is the smallest nonnegative integer which satisfies the inequality
$$\varphi(u^k + \tau^j v^k) \leq \varphi(u^k) + \varepsilon \tau^j \langle \varphi'(u^k), v^k \rangle. \tag{4.7}$$
Set $u^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) = u^k + \alpha_k v^k$.
3. Increase k by 1 and go to Step 1.

If $\Phi(u^k) = 0$ for some iterate, we stop since the equation is solved. Note that if we do not stop according to this test, then neither $v^k = 0$ nor $\varphi'(u^k) = 0$ can happen when the Newton direction exists, because of (4.5). If $v^k = 0$ is the gradient direction at the point where the Newton direction does not exist, then u^k is a stationary point of φ and we stop since no further progress is possible. We assume, from now on, that Algorithm 4.1 does not stop, which means that $v^k \neq 0$ and $\varphi'(u^k) \neq 0$ for all k.

The test (4.6) for the size of the Newton direction (where $M > 0$ and $\theta > 0$ should be chosen large to allow more Newton directions) can be checked within the inner solver for the Newton system, and plays the same role as detecting its inconsistency – i.e., that something is wrong. In such cases, we resort to the back-up gradient direction and proceed.

Theorem 4.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $G, H : \mathbb{R}^n \to \mathbb{R}^m$ be twice continuously differentiable on $\mathbb{R}^n$.*

Then Algorithm 4.1 is well-defined, and any accumulation point $\bar{u}$ of any sequence $\{u^k\}$ generated by this algorithm is a stationary point of the function φ, i.e.,

$$\varphi'(\bar{u}) = 0. \tag{4.8}$$

Proof. First note that linesearch is always in a direction of descent, because the Newton direction satisfies (4.5) and otherwise the direction is the negative gradient, and thus, in any case,

$$\langle \varphi'(u^k),\, v^k \rangle < 0 \tag{4.9}$$

for all k. By the standard facts concerning the Armijo linesearch, it follows that the sequences $\{\alpha_k\}$ and $\{u^k\}$ are well-defined. Furthermore, by (4.7) we have that the sequence $\{\varphi(u^k)\}$ is monotonically nonincreasing. Since it is bounded below (by zero), it converges. Then, by (4.7), we have that

$$\lim_{k\to\infty} \alpha_k \langle \varphi'(u^k),\, v^k \rangle = 0. \tag{4.10}$$

Let $\bar{u}$ be an accumulation point of the sequence $\{u^k\}$, and let $\{u^{k_j}\}$ be a subsequence convergent to $\bar{u}$. Consider the two possible cases:

$$\limsup_{j\to\infty} \alpha_{k_j} > 0 \quad \text{or} \quad \lim_{j\to\infty} \alpha_{k_j} = 0. \tag{4.11}$$

In the first case, passing onto a further subsequence if necessary, we can assume that the entire $\{\alpha_{k_j}\}$ is separated from zero:

$$\liminf_{j\to\infty} \alpha_{k_j} > 0.$$

Then (4.10) implies that

$$\lim_{j\to\infty} \langle \varphi'(u^{k_j}),\, v^{k_j} \rangle = 0. \tag{4.12}$$

If within this subsequence the Newton direction is used for infinitely many indices j, by (4.5) we have in (4.12) that

$$\langle \varphi'(u^{k_j}),\, v^{k_j} \rangle = -2\varphi(u^{k_j})$$

for infinitely many j. Then (4.12) implies that $\varphi(\bar{u}) = 0$, i.e., $\bar{u}$ is a global minimizer of φ, which certainly implies (4.8). On the other hand, if Newton directions are used only for finitely many indices j, then

$$\langle \varphi'(u^{k_j}),\, v^{k_j} \rangle = -\langle \varphi'(u^{k_j}),\, \varphi'(u^{k_j}) \rangle = -\|\varphi'(u^{k_j})\|^2$$

for all j large enough, and we conclude by (4.12) that (4.8) holds.

It remains to consider the second case in (4.11). Suppose first that the sequence $\{v^{k_j}\}$ is unbounded. Note that this can only happen when the Newton directions are used infinitely often (because otherwise $v^{k_j} = -\varphi'(u^{k_j})$ for all j large enough, and hence, $\{v^{k_j}\}$ converges to $-\varphi'(\bar{u})$). But then the condition $\|v^{k_j}\| \leq \max\{M, 1/(\varphi(u^{k_j}))^\theta\}$ implies that $\varphi(u^{k_j}) \to 0$ so that $\varphi(\bar{u}) = 0$, and hence, $\bar{u}$ is again a global minimizer of φ, and (4.8) follows.

Let finally $\{v^{k_j}\}$ be bounded. Taking a further subsequence, if necessary, assume that $\{v^{k_j}\}$ converges to some $\bar{v}$. Since in the second case in (4.11) for each j large enough the initial stepsize value had been reduced at the current point u^{k_j} at least once, the value $\alpha_{k_j}/\tau > \alpha_{k_j}$ does not satisfy (4.7), i.e.,

$$\frac{\varphi(u^{k_j} + (\alpha_{k_j}/\tau)v^{k_j}) - \varphi(u^{k_j})}{\alpha_{k_j}/\tau} > \varepsilon\langle\varphi'(u^{k_j}), v^{k_j}\rangle.$$

Employing the Mean-Value Theorem and the fact that $\alpha_{k_j} \to 0$ as $j \to \infty$, and passing onto the limit as $j \to \infty$, we obtain that

$$\langle\varphi'(\bar{u}), \bar{v}\rangle \geq \varepsilon\langle\varphi'(\bar{u}), \bar{v}\rangle,$$

which may only hold when

$$\langle\varphi'(\bar{u}), \bar{v}\rangle \geq 0.$$

Combining this with (4.9), we obtain that

$$\langle\varphi'(\bar{u}), \bar{v}\rangle = 0.$$

Considering, as above, the two cases when the number of times the Newton direction had been used is infinite or finite, the latter relation implies that (4.8) holds. ■

According to the proof of Theorem 4.1, if along a subsequence convergent to $\bar{u} = (\bar{x}, \bar{y}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ the Newton direction had been used infinitely many times, then $\Phi(\bar{u}) = 0$, i.e., $(\bar{x}, \bar{y})$ is a stationary point of problem (1.3) and $\bar{\lambda}$ is an associated Lagrange multiplier. By Proposition 2.1, it then follows that $\bar{x}$ is a weakly stationary point of (1.1). Convergence to a stationary point of φ which is not its global minimizer can only happen when Newton directions are not used along the corresponding subsequence from some point on at all. But even in that case, since for any stationary point $\bar{u}$ of the function φ it holds that $\Lambda\Phi(\bar{u}) = 0$ for each matrix $\Lambda \in \partial_B\Phi(\bar{u})$ (see (4.2)), if among those matrices *at least one* is nonsingular we immediately obtain that $\Phi(\bar{u}) = 0$. Thus we can expect global convergence of Algorithm 4.1 to weakly stationary points of (1.1).

Finally, we show that Algorithm 4.1 preserves fast local convergence of Algorithm 3.1 under the relevant assumptions.

Theorem 4.2 *Let f, G and H be twice continuously differentiable on $\mathbb{R}^n$, and let a sequence $\{u^k\}$ generated by Algorithm 4.1 have an accumulation point $\bar{u} = (\bar{x}, \bar{y}, \bar{\mu})$, where $\bar{x}$ is a strongly stationary point of the problem* (1.1)*, $\bar{y}$ is given by* (2.14)*, and $\bar{\mu}$ is an MPCC-multiplier associated to $\bar{x}$. Assume that MPCC-LICQ* (2.4)*, ULSCC* (2.11)*, and the second-order sufficient condition* (2.13) *hold.*

Then the whole sequence $\{u^k\}$ converges to $(\bar{x}, \bar{y}, \bar{\mu})$. The rate of convergence is superlinear, and if the second derivatives of f, G and H are locally Lipschitz-continuous with respect to $\bar{x}$ then it is quadratic.

Proof. Let u^k be close to $\bar{u}$, and let $\tilde{u}^{k+1}$ be computed as in Algorithm 4.1. According to Theorem 3.2, for u^k be close to $\bar{u}$, this point is well-defined and

$$\|\tilde{u}^{k+1} - \bar{u}\| = o(\|u^k - \bar{u}\|). \tag{4.13}$$

As a consequence, $\tilde{u}^{k+1}$ would be accepted by the test (4.6).

Furthermore, according to Proposition 3.2, under the stated assumptions the mapping Φ is BD-regular at $\bar{u}$. It is well known [21] that BD-regularity implies the error bound of the form

$$\|u - \bar{u}\| = O(\|\Phi(u)\|).$$

Employing this error bound and (4.13), and also taking into account Lipschitz-continuity of Φ near $\bar{u}$ (see Proposition 3.1), we obtain that

$$\begin{aligned}
\varphi(\tilde{u}^{k+1}) &= \frac{1}{2}\|\Phi(\tilde{u}^{k+1}) - \Phi(\bar{u})\|^2 \\
&= O(\|\tilde{u}^{k+1} - \bar{u}\|^2) \\
&= o(\|u^k - \bar{u}\|^2) \\
&= o(\|\Phi(u^k)\|^2) \\
&= o(\varphi(u^k)).
\end{aligned}$$

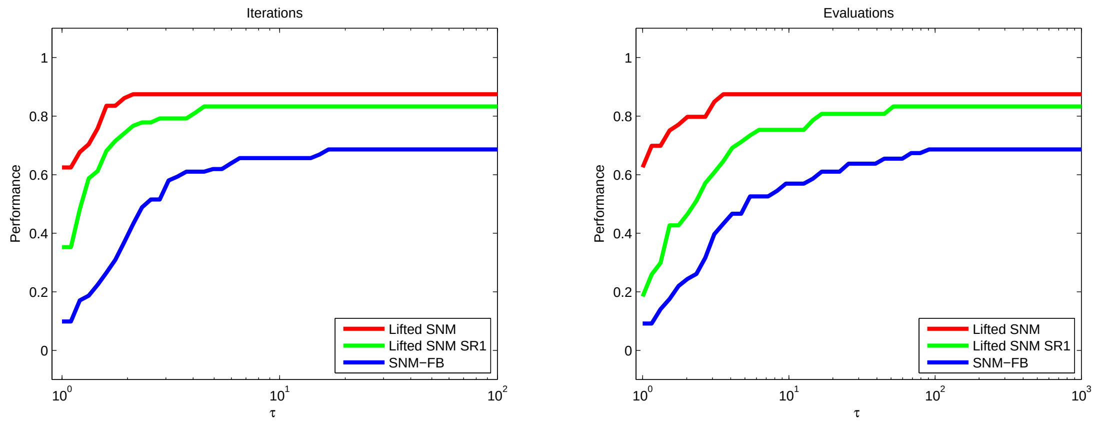
Setting $v^k = \tilde{u}^{k+1} - u^k$, the above relation implies that if u^k is close enough to $\bar{u}$ then

$$\begin{aligned}
\varphi(u^k + v^k) &= \varphi(\tilde{u}^{k+1}) \\
&\leq (1 - 2\varepsilon)\varphi(u^k) \\
&= \varphi(u^k) - \varepsilon\|\Phi(u^k)\|^2 \\
&= \varphi(u^k) + \varepsilon\langle\varphi'(u^k),\, v^k\rangle,
\end{aligned}$$

where the last equality is by (4.5) (recall also that $\varepsilon \in (0,\, 1/2)$). Therefore, $\alpha_k = 1$ will be accepted by the Armijo rule: inequality (4.7) holds with $j = 0$. This shows that iterations of Algorithm 4.1 reduce to the (local) Algorithm 3.1. The assertions now follow from Theorem 3.2. ∎

5 Numerical results

In this section, we present some preliminary numerical experience with two versions of SNM applied to the lifted MPCC, SNM applied to the optimality conditions of the original MPCC, and SQP with linesearch for the original MPCC. We use small test problems derived from MacMPEC [16]. Our selection of test problems is the same as in [9]. Specifically, 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. We also ignore simple bounds when there are any (which sometimes affects the solutions and stationary points of these problems). We end up with 38 problems.



(a) Iterations

(b) Evaluations

Figure 1: Lifted SNMs vs SNM-FB.

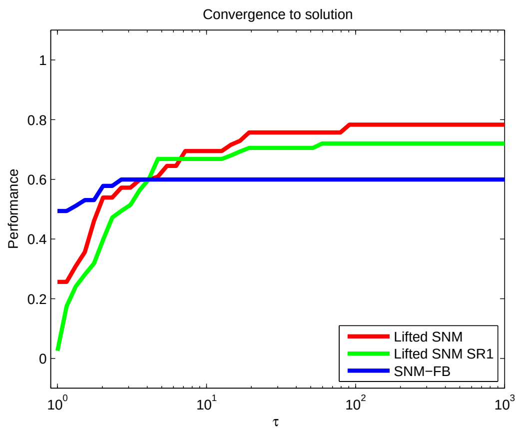


Figure 2: Lifted SNMs vs SNM-FB: convergence to solution.

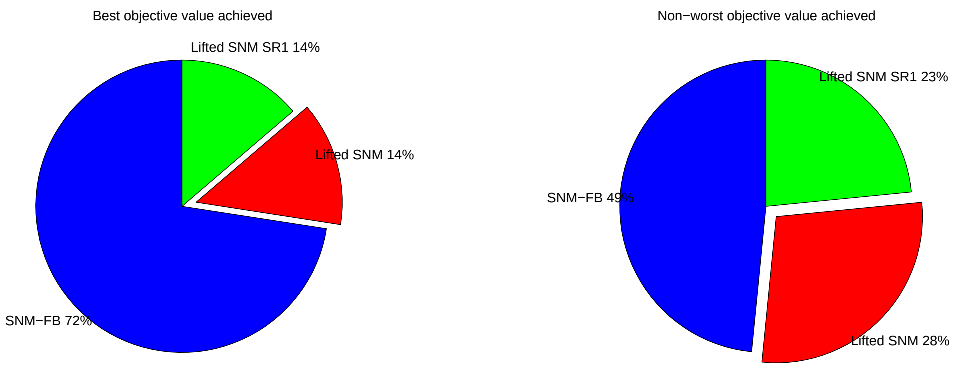
We consider two versions of SNM for lifted MPCC. One is Algorithm 4.1 as stated, and we refer to it as `Lifted SNM`. The other is a quasi-Newton version of Algorithm 4.1, with the Hessian in (3.4) replaced by its SR1 approximations [18, (6.24), (18.13)], and we call it `Lifted SNM SR1`. SR1 updates were chosen because in the context of SNM we do not need to care about positive definiteness of approximations of the Hessian, but we want to keep them symmetric.

We first compare both methods with `SNM-FB`, which is the natural linesearch version of the semismooth Newton method applied to the Fischer-Burmeister reformulation of the first-order optimality conditions of the original MPCC (1.1); see [9] for details of the specific implementation. Another algorithm chosen for comparison is `SQP BFGS`, which is the quasi-Newton version of the SQP algorithm, with BFGS approximations of the Hessian complemented by Powell's modification, and with linesearch for the l_1 penalty function (e.g., see [18, Section 18], and also [9] for details of our implementation). The latter method was implemented without any tools for tackling possible infeasibility of subproblems and for avoiding the Maratos effect.

The parameters of Algorithm 4.1 were chosen as follows: $\varepsilon = 10^{-4}$, $\tau = 0.5$, $M = 10^5$, $\theta = 1$. All computations were performed in Matlab environment, with the QP-subproblems of `SQP BFGS` solved by the built-in Matlab QP-solver. For `Lifted SNM` and `Lifted SNM SR1`, we used the stopping criterion of the form

$$\|\Phi(x^k,\, y^k,\, \lambda^k)\| < 10^{-6},$$

where Φ is defined in (3.3). `SNM-FB` and `SQP BFGS` were stopped when the Fischer-Burmeister residual of the the first-order optimality conditions of the original MPCC (1.1) becomes smaller than 10^{-6}. Failures were declared when the needed accuracy was not achieved after 500 iterations or when the method in question failed to complete an iteration, for whatever reason.



(a) Best objective value achieved

(b) Non-worst objective value achieved

Figure 3: Lifted SNMs vs SNM-FB.

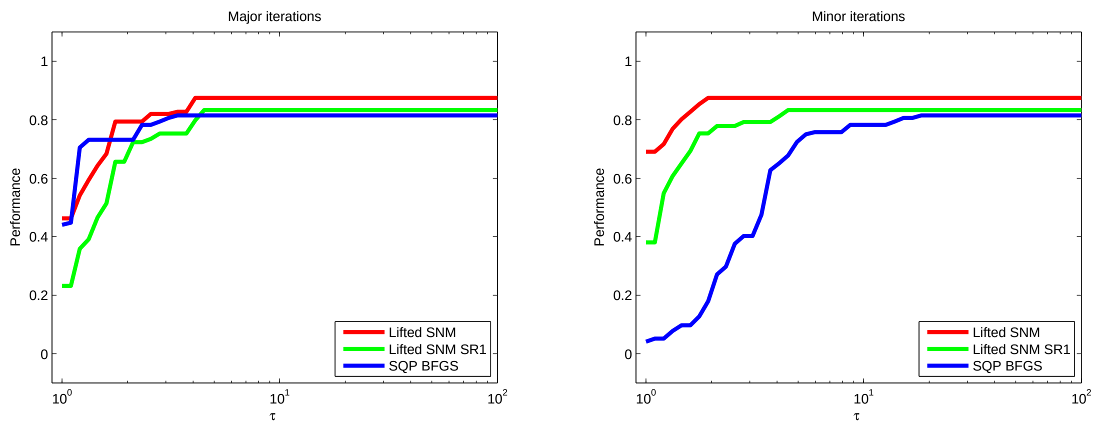
We performed 100 runs for each algorithm under consideration from the same sample of randomly generated starting points. Primal starting points for each algorithm were generated in a cubic neighborhood around the solution (the “solutions” were found in the course of experiments), with the edge of the cube equal to 20. For the lifted reformulation, we defined the starting value y^0 of the auxiliary variable as follows:

$$y_i^0 = \begin{cases} |H_i(x^0)|^{1/2}, & \text{if } H_i(x^0) \geq G_i(x^0), \\ -|G_i(x^0)|^{1/2}, & \text{if } H_i(x^0) < G_i(x^0), \end{cases} \qquad i = 1,\, \ldots,\, m,$$

where x^0 is the primal starting point. Dual starting points for all algorithms were generated the same way as primal ones but around 0, and with additional nonnegativity restrictions for their components corresponding to inequality constraints (this concerns `SNM-FB` and `SQP BFGS` which are applied to the original problem (1.1)). In the case of a successful run, “convergence to solution” was declared when the distance from the last primal iterate to the solution was smaller than 10^{-3}.

Fig. 1 reports on the average numbers of iterations and evaluations of constraint functions and derivatives values for `Lifted SNM`, `Lifted SNM SR1` and `SNM-FB` over successful runs, in the form of a performance profile [3]. (Note that these methods do not require objective function values). For each algorithm, the value of the plotted function at $\tau \in [1, +\infty)$ corresponds to the part of the problems in the test set for which the achieved result (the average iteration count, or the evaluation count) was no more than τ times worse (bigger) than the best result among the three algorithms. Failure is regarded as infinitely many times worse than the best result. Thus, the value at $\tau = 1$ characterizes "pure efficiency" of the algorithm (that is, the part of problems for which the given algorithm demonstrated the best result), while the value at $\tau = +\infty$ characterizes robustness of the algorithm (that is, the part of problems which were successfully solved by the given algorithm). It is evident that both `Lifted SNM` and `Lifted SNM SR1` seriously outperform `SNM-FB` both in robustness and efficiency. `Lifted SNM SR1` is less efficient (and somewhat less robust) than `Lifted SNM`, which is a natural price for not computing the true Hessian. Of course, the former has its usual advantages when computing second derivatives is too costly or simply impossible.

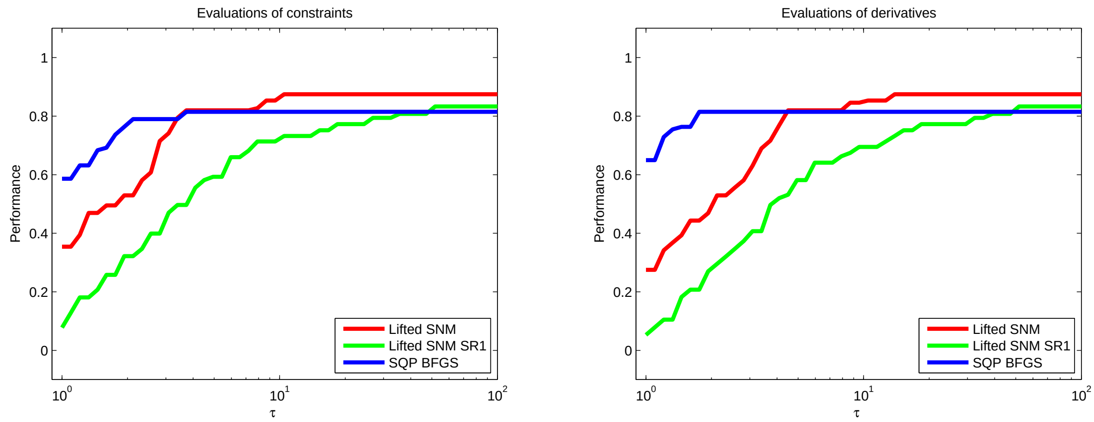
Apart from robustness and efficiency, another important characteristic of any algorithm is the quality of the output produced, i.e., the percentage of those cases when the algorithm converges to a true solution rather than to some nonoptimal stationary point. Fig. 2 reports on this aspect of behavior of `Lifted SNM`, `Lifted SNM SR1` and `SNM-FB`. Here, for each algorithm we look at the inverse of the number of convergences to solution. Note that this result equals to $+\infty$ when the given algorithm gave no convergences to the solution for a given problem, and this adds to the cases of failure. That is why the values on the right end are smaller than in Fig. 1. One can see that `SNM-FB` has in principle a stronger tendency of convergence to the solution than `Lifted SNM` and `Lifted SNM SR1`, but the picture becomes different when this data is combined with robustness.



(a) Major iterations (b) Minor iterations

Figure 4: Lifted SNMs vs SQP BFGS.

Diagrams in Fig. 3 are intended to give some impression of the ability of `Lifted SNM`, `Lifted SNM SR1` and `SNM-FB` to achieve smaller values of the objective function in the cases of successful runs when, in particular, the last produced iterate is (nearly) feasible. We report on percentage of those problems for which each algorithm demonstrated the best (smallest) and the non-worst average of the achieved objective function values over successful



(a) Evaluations of constraints (b) Evaluations of derivatives

Figure 5: Lifted SNMs vs SQP BFGS.

runs among the three SNM-based algorithms. The results were regarded as equal when the difference was less than 10^{-3}. Note that for some particular problems the algorithms can fall into both "best" and "worst" categories, if all three algorithms give the same result. `SNM-FB` demonstrates better ability in decreasing the objective function. Note, however, that comparative robustness of the algorithms is not reflected in Fig. 3

We proceed with comparisons of `Lifted SNM` and `Lifted SNM SR1` with `SQP BFGS`. The information in Figs. 4–7 is produced similarly to Figs. 1–3. Some special features are the following.

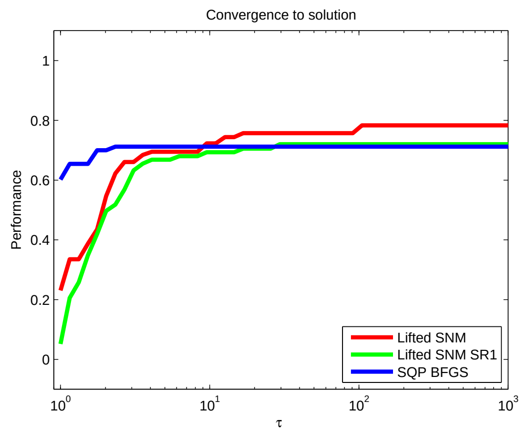
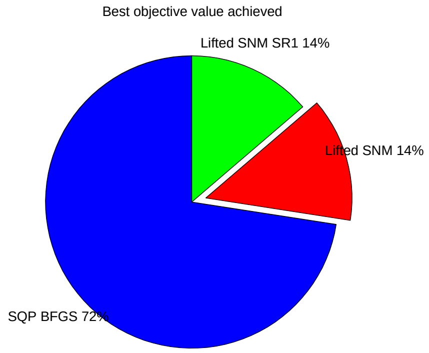
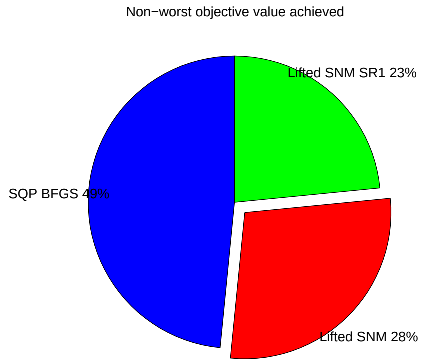


Figure 6: Lifted SNMs vs SQP BFGS: convergence to solution.

First, Fig. 4 reports separately on major and minor iteration counts. For `Lifted SNM` and `Lifted SNM SR1` these two counts are the same, since these algorithms are QP-free and each major iteration consists of solving one linear system, followed by linesearch. `SQP BFGS` subproblems are general QPs with inequality constraints. Solving each of these subproblems by the active set QP-solver usually requires more than one minor (inner) iteration, and each minor iteration includes solving a certain linear system. One can see from Fig. 4 that `Lifted SNM` and `Lifted SNM SR1` are comparable with `SQP BFGS` in terms of major iterations, but outperform the latter in terms of minor iterations. Moreover, lifted SNMs are even somewhat more robust than our simple implementation of `SQP BFGS`.



(a) Best objective value achieved



(b) Non-worst objective value achieved

Figure 7: Lifted SNMs vs SQP.

Fig. 5 reports separately on the numbers of evaluation of constraint functions and of derivatives, since these two counts are not the same for `SQP BFGS`. This method requires less evaluations than both `Lifted SNM` and `Lifted SNM SR1`. Note, however, that `SQP BFGS` requires also evaluations of the objective function, not needed in SNMs.

Figs. 6 and 7 demonstrate that `SQP BFGS` has better properties of convergence to solution and of reducing the objective function value. This is quite natural, since SQP is clearly more optimization-related. Somewhat surprisingly, Fig. 3 and Fig. 7 turn out to be exactly the same. Most likely, this happened accidentally (recall that we round off the averages of achieved objective function values up to the accuracy 10^{-3}).

We do not report here on the comparisons of `Lifted SNM` and `Lifted SNM SR1` with the active-set Newton methods developed in [9] (these combine SNM and SQP with active-set strategies). The reason is that numerical results in [9] on the same test set indicate that the active-set phase can be beneficial in certain situations but does not seriously change the behavior "on average".

Numerical results presented in this section allow for the following (very preliminary) conclusions. The idea of lifting can be useful. For example, lifted SNM algorithms outperform in efficiency and robustness SNM applied to the original MPCC. They also compare quite favorably in the same indicators even with SQP (at least in its simple implementations). On the other hand, the quality of the output of lifted SNMs with respect to optimality is generally lower than that of more traditional (especially SQP-based) algorithms.

6 Concluding Remarks

We have shown that the Lagrange optimality system of the lifted reformulation (with power $s = 2$) of MPCC, although nonsmooth, can be regular in an appropriate sense. Sufficient conditions for its regularity are reasonable: MPCC-LICQ, upper-level strict complementarity (ULSCC), and the second-order sufficiency. Under these assumptions, the Lagrange system can be solved by a fast semismooth Newton method. Moreover, it turns out that the squared residual of this system is actually smooth, which allows for a natural globalization strategy. The resulting globalized algorithm preserves fast local convergence under the relevant assumptions. Preliminary numerical experience suggest that the approach developed in this work has some potential.

We note, in passing, that Proposition 3.2 and Theorems 3.2 and 4.2 remain valid with the strong stationarity assumption replaced by weak stationarity, and with ULSCC replaced by the assumption that the I_0-components of the corresponding multiplier are not equal to zero (allowed to be negative).

References

[1] Billups S. C. Algorithms for complementarity problems and generalized equations. Tech. report 95–14 and Ph.D. thesis. Computer Sciences Department, University of Wisconsin. Madison, 1995.

[2] De Luca T., Facchinei F., and Kanzow C. A theoretical and numerical comparison of some semismooth algorithms for complementarity problems. *Comput. Optim. Appl.* 16 (2000), 173–205.

[3] E. Dolan and J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91:201–213, 2002.

[4] Facchinei F. and Pang J.-S. *Finite-Dimensional Variational Inequalities and Complementarity Problems.* New York: Springer-Verlag, 2003.

[5] Facchinei F. and Soares J. A new merit function for nonlinear complementarity problems and a related algorithm. *SIAM J. Optim.* 7 (1997), 225–247.

[6] Fletcher R. and Leyffer S. Solving mathematical programs with equilibrium constraints as nonlinear programs. *Optim. Methods Softw.* 19 (2004), 15–40.

[7] Fletcher R., Leyffer S., Ralph D., and Scholtes S. Local convergence of SQP methods for mathematical programs with equilibrium constraints. *SIAM J. Optim.* 17 (2006), 259–286.

[8] Izmailov A.F. and Solodov M.V. Superlinearly convergent algorithms for solving singular equations and smooth reformulations of complementarity problems. *SIAM J. Optim.* 13 (2002), 386–405.

[9] Izmailov A.F. and Solodov M.V. An active-set Newton method for mathematical programs with complementarity constraints. *SIAM J. Optim.* 19 (2008), 1003–1027.

[10] Izmailov A.F. and Solodov M.V. On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. *Math. Program.* 117 (2009), 271–304.

[11] Jiang H. Global convergence analysis of the generalized Newton and Gauss–Newton methods of the Fischer–Burmeister equation of the complementarity problem. *Math. Oper. Res.*, 24 (1999), 529–543.

[12] Kanzow C. and Kleinmichel H. A new class of semismooth Newton-type methods for nonlinear complementarity problems. *Comput. Optim. Appl.* 11 (1998), 227–251.

[13] Kojima M. and Shindo S. Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. *J. Oper. Res. Soc. Japan* 29 (1986), 352–374.

[14] Kummer B. Newton's method for nondifferentiable functions. *Advances in Mathematical Optimization*, J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommalzsch, L. Tammer, M. Vlach, and K. Zimmerman, eds., V. 45, pp. 114–125. Berlin: Akademie-Verlag, 1988.

[15] Kummer B. Newton's method based on generalized derivatives for nonsmooth functions. *Advances in Optimization*, W. Oettli and D. Pallaschke, eds., pp. 171–194. Berlin: Springer-Verlag, 1992.

[16] Leyffer S. MacMPEC: AMPL collection of Mathematical Programs with Equilibrium Constraints. http://wiki.mcs.anl.gov/leyffer/index.php/MacMPEC.

[17] Luo Z.-Q., Pang J.-S., and Ralph D. *Mathematical programs with equilibrium constraints.* Cambridge: Cambridge Univ. Press, 1996.

[18] J. Nocedal and S.J. Wright. *Numerical Optimization. Second edition.* Springer, New York, 2006.

[19] Mifflin R. Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control Optim.* 15 (1977), 959–972.

[20] Outrata J.V., Kocvara M., and Zowe J. *Nonsmooth approach to optimization problems with equilibrium constraints: Theory, applications and numerical results.* Boston: Kluwer Acad. Publ., 1998.

[21] Pang J. S. and Qi L. Nonsmooth equations: Motivation and algorithms. *SIAM J. Optim.* 3 (1993), 443–465.

[22] Qi L. Convergence analysis of some algorithms for solving nonsmooth equations. *Math. Oper. Res.* 18 (1993), 227–244.

[23] Qi L. and Sun J. A nonsmooth version of Newton's method. *Math. Program.* 58 (1993), 353–367.

[24] Ralph D. Sequential quadratic programming for mathematical programs with linear complementarity constraints. *Computational Techniques and Applications: CTAC95.*, pp. 663–668. Singapore: World Scientific, 1996.

[25] Ralph D. and Wright S. J. Some properties of regularization and penalization schemes for MPECs. *Optim. Methods Softw.* 19 (2004), 527–556.

[26] Scheel H. and Scholtes S. Mathematical programs with complementarity constraints: stationarity, optimality and sensitivity. *Math. Oper. Res.* 25 (2000), 1–22.

[27] Scholtes S. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM J. Optim.* 11 (2001), 918–936.

[28] Stein O. Lifting mathematical programs with complementarity constraints. *Mathematical Programming*, 2010. DOI 10.1007/s10107-010-0345-y.