Taylor & Francis
Taylor & Francis Group

Semismooth SQP method for equality-constrained optimization problems with an application to the lifted reformulation of mathematical programs with complementarity constraints

A.F. Izmailov[a], A.L. Pogosyan[a] and M.V. Solodov[b]*

[a]*VMK Faculty, OR Department, Leninskiye Gory, Moscow State University, MSU, Uchebniy Korpus 2, 119991 Moscow, Russia;* [b]*IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil*

Dedicated to the 60-th birthday of Prof. Florian Potra

We consider the sequential quadratic programming (SQP) algorithm applied to equality-constrained optimization problems, where the problem data is differentiable with Lipschitz-continuous first derivatives. For this setting, Dennis–Moré-type analysis of primal superlinear convergence is presented. Our main motivation is a special modification of SQP tailored to the structure of the lifted reformulation of mathematical programs with complementarity constraints (MPCC). For this problem, we propose a special positive definite modification of the matrices in the generalized Hessian, which is suitable for globalization of SQP based on the penalty function, and at the same time can be expected to satisfy our general Dennis–Moré-type conditions, thus preserving local superlinear convergence. (Standard quasi-Newton updates in the SQP framework require twice differentiability of the problem data at the solution for superlinear convergence.) Preliminary numerical results comparing a number of quasi-Newton versions of semismooth SQP applied to MPCC are also reported.

Keywords: sequential quadratic programming; semismoothness; *B*-differential; *BD*-regularity; semismooth Newton method; second-order sufficiency; mathematical programs with complementarity constraints

AMS Subject Classification: 90C30; 90C33; 90C55; 65K05

1. Introduction

We consider the equality-constrained optimization problem

$$\min f(x) \quad \text{s.t.} \quad h(x) = 0, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ are differentiable with locally Lipschitz-continuous first derivatives, but not necessarily twice differentiable. Problems with these smoothness properties arise in stochastic programming and optimal control (the so-called extended linear-quadratic problems [26,29,30]), in semi-infinite programming and in primal decomposition procedures

---

*Corresponding author. Email: solodov@impa.br

(see [16,24] and references therein). Once but not twice differentiable functions arise also when reformulating complementarity constraints as in [13] or in the lifting approach [15,32]. Other possible sources are subproblems in augmented Lagrangian methods with lower-level equality constraints (treated directly) and upper-level inequality constraints (treated via Lagrangian penalization, which gives certain terms that are not twice differentiable in general); see, e.g. [1].

Our main application in this paper, considered in Section 3, is concerned with reformulations of complementarity conditions via the lifting approach [15,32]. To this end, consider the mathematical program with complementarity constraints (MPCC)

$$\min f(x) \quad \text{s.t.} \quad G(x) \geq 0, \quad H(x) \geq 0, \quad \langle G(x), H(x) \rangle = 0, \tag{2}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $G, H : \mathbb{R}^n \to \mathbb{R}^m$ are twice differentiable, with locally bounded second derivatives. We refer the reader to [7,8,13,14,20,23,28,31] for motivations and theoretical and computational developments for this problem class. As suggested in [32], the constraints of (2) can be written as smooth equalities by introducing an auxiliary variable $y \in \mathbb{R}^m$ and using the lifted reformulation:

$$(-\min\{0, y\})^s - G(x) = 0, \quad (\max\{0, y\})^s - H(x) = 0, \quad s > 1,$$

where the operations of taking minimum, maximum and applying power are understood component-wise. In [32], the value $s = 3$ is employed, which gives a problem with twice continuously differentiable data but leads to its degeneracy. The approach of [15] uses $s = 2$, which converts (2) to

$$\min f(x) \quad \text{s.t.} \quad (\min\{0, y\})^2 - G(x) = 0, \quad (\max\{0, y\})^2 - H(x) = 0. \tag{3}$$

Under our assumptions, the constraints in (3) are differentiable, with locally Lipschitz-continuous derivatives, but generally not twice differentiable. The advantage is that the problem (3) has better regularity properties than when the power $s = 3$ is used; see the discussion in [15] and also Remark 4.

In this paper, we consider the semismooth version of the sequential quadratic programming (SQP) method [2] for problem (1), and its modification tailored specifically to the structure of lifted MPCC (3). It should be noted that, in local analysis, semismooth SQP for (1) is just the semismooth Newton method (SNM) applied to the Lagrange optimality system of (1). For generic systems of nonlinear equations, such methods were developed in [17,18,25,27], with local superlinear convergence properties established under reasonable regularity assumptions (see the discussion in Section 2.1). Moreover, local convergence and rate of convergence properties of semismooth SQP for more general optimization problems with additional inequality constraints were studied in [9,26]. SNM for the Lagrange system of the lifted reformulation (3) of MPCC (2) was proposed in [15], with globalization strategy based on linesearch for the squared residual of the Lagrange system.

In this paper, instead of globalization based on the residual of the Lagrange system of (3), we shall employ the l_1-penalty function in a way that takes into account special structure of the lifted MPCC (3). This strategy is more in the spirit of classical SQP algorithms [2] and can be advantageous compared to the use of the residual of the Lagrange system. In particular, it is natural to expect that globalization using the penalty function, being more optimization-oriented, should outperform globalization based on the residual in terms of the 'quality' of the output: the former should be much less attracted by nonoptimal stationary points than the latter. This is also confirmed by our numerical results is Section 4.

It should be emphasized that we do not just apply generic SQP with l_1-penalty globalization to the lifted MPCC (3), and thus our development requires some theoretical analysis as well.

There are several reasons why we believe SQP should be modified from its generic version to take into account special structure of lifted MPCC. First, globalization based on linesearch for a penalty function requires positive definite modifications of the matrices in the generalized Hessian of the Lagrangian of problem (3). In principle, this can be done by standard quasi-Newton updates, such as BFGS. However, it is known that for systems of nonlinear equations quasi-Newton methods with standard updates are guaranteed to preserve superlinear convergence only when the equation mapping is actually differentiable at the solution, while the Lagrange optimality system of (3) is not differentiable. For this reason, instead of using standard quasi-Newton updates to make the matrices positive definite, we suggest a special modification directly linked to the structure of the problem in hand (the two approaches are also compared numerically in Section 4, and our results confirm that the proposed special approach does work better). To show that this modification can be expected to preserve high convergence rate, we need some general quasi-Newton-type results for problem (1). Since there appears to be no quasi-Newton theory for semismooth problems that suits our specific needs, our first contribution is developing such theory. In particular, we give Dennis–Moré-type analysis of primal superlinear convergence rate for a generic quasi-Newton semismooth SQP method for problem (1), assuming primal–dual convergence. It should be noted that generic quasi-Newton versions of semismooth SQP methods for optimization problems with Lipschitzian derivatives were previously discussed in [9,26]. But these results concern local primal–dual superlinear convergence, assuming that the matrices being used are close enough to true generalized Hessians. Within the globalization context of this paper, our *a posteriori* analysis is more appropriate. We assume primal–dual convergence as induced by the globalization strategy, but do not require the matrices to be close to the true generalized Hessians (instead, natural extensions of classical Dennis–Moré conditions are considered).

We next introduce our notation. The Lagrangian of problem (1) is given by

$$L(x, \lambda) = f(x) + \langle \lambda, h(x) \rangle,$$

where $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^l$. Stationary points and the Lagrange multipliers of problem (1) are characterized by the Lagrange optimality system

$$\frac{\partial L}{\partial x}(x, \lambda) = 0, \quad h(x) = 0. \tag{4}$$

As is well known, if $\bar{x} \in \mathbb{R}^n$ is a local solution of problem (1) satisfying the regularity condition

$$\operatorname{rank} h'(\bar{x}) = l, \tag{5}$$

then $\bar{x}$ is a stationary point of problem (1) in the sense that there exists the (unique) associated Lagrange multiplier $\bar{\lambda} \in \mathbb{R}^l$ such that $(\bar{x}, \bar{\lambda})$ satisfies (4).

Let a mapping $\Phi \colon \mathbb{R}^q \to \mathbb{R}^r$ be Lipschitz-continuous around a point $u \in \mathbb{R}^q$. The B-differential [6, Section 7.4] of Φ at u is the set

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

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

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

where conv S stands for the convex hull of the set S. Furthermore, for a mapping $\Phi \colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^r$ which is Lipschitz-continuous around a point $(u, v) \in \mathbb{R}^p \times \mathbb{R}^q$, we define the partial B-differential of Φ at (u, v) with respect to u as the B-differential of the mapping $\Phi(\cdot, v)$, and

we denote it by $(\partial_B)_u \Phi(u, v)$. Similarly, the partial Clarke generalized Jacobian of Φ at (u, v) with respect to u is the Clarke generalized Jacobian of the mapping $\Phi(\cdot, v)$, which we denote by $\partial_u \Phi(u, v)$.

The mapping $\Phi \colon \mathbb{R}^q \to \mathbb{R}^r$ is said to be *semismooth* [6, Section 7.4] 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.

We will be saying that a mapping $\Phi \colon \mathbb{R}^q \to \mathbb{R}^r$ is locally Lipschitz-continuous with respect to a given point $\bar{u} \in \mathbb{R}^q$ if

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

For a given set $S \subset \mathbb{R}^q$ and any $u \in \mathbb{R}^q$, we define

$$\mathrm{dist}(u, S) = \inf_{w \in S} \|u - w\|.$$

By π_S we denote the Euclidian projector onto S. For a linear operator Λ, $\ker \Lambda$ is its kernel (the null space) and $\mathrm{im}\, \Lambda$ is its image (the range space). For a given $u \in \mathbb{R}^q$, by $\mathrm{diag}(u)$ we mean the $q \times q$-matrix with the components of the vector u on the diagonal and zeroes elsewhere. The derivative of a function $\varphi \colon \mathbb{R}^q \to \mathbb{R}$ at $u \in \mathbb{R}^q$ in a direction $v \in \mathbb{R}^q$ will be denoted by $\varphi'(u; v)$.

2. Semismooth SQP

In this section, we start with discussing the known primal–dual convergence properties of SQP applied to the general equality-constrained problem (1) with the stated smoothness assumptions. After that, we consider the quasi-Newton modifications and formulate Dennis–Moré-type conditions for primal superlinear convergence rate.

2.1 *The basic algorithm and its local convergence*

Given the current iterate $(x^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^l$, an iteration of the SQP algorithm [2] for problem (1) consists of computing a stationary point x^{k+1} and an associated Lagrange multiplier λ^{k+1} of the subproblem

$$\min \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle H_k(x - x^k), x - x^k \rangle \quad \text{s.t.} \quad h(x^k) + h'(x^k)(x - x^k) = 0, \quad (6)$$

where H_k is a symmetric $n \times n$-matrix. Equivalently, this iteration can be stated in the form of solving the linear system

$$H_k(x - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) = -\frac{\partial L}{\partial x}(x^k, \lambda^k), \quad h'(x^k)(x - x^k) = -h(x^k). \quad (7)$$

In the case of twice differentiable data, the basic form of the method corresponds to taking H_k as the Hessian of the Lagrangian with respect to the primal variable x. Accordingly, under our

smoothness assumptions the natural choice is

$$H_k \in \partial_x \frac{\partial L}{\partial x}(x^k, \lambda^k). \tag{8}$$

We shall refer to the iterative process defined by (6) (or (7)) with the choice of H_k given by (8) as *semismooth SQP*. As mentioned above, methods of this kind were considered, e.g. in [9,26].

As is well known, semismooth SQP can be interpreted as the SNM (see [17,18,25,27] and [6, Chapter 7]) applied to the Lagrange optimality system (4). Specifically, define $\Phi : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}^n \times \mathbb{R}^l$ by

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x, \lambda), h(x) \right), \quad u = (x, \lambda), \tag{9}$$

and consider the equation

$$\Phi(u) = 0. \tag{10}$$

An iteration of the SNM for (10) consists of solving the equation

$$\Phi(u^k) + \Lambda_k(u - u^k) = 0,$$

for the current iterate $u^k \in \mathbb{R}^n \times \mathbb{R}^l$ and some matrix $\Lambda_k \in \partial\Phi(u^k)$. It can be seen [11] that for any $u = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ it holds that

$$\partial\Phi(u) = \left\{ \begin{pmatrix} H & (h'(x))^{\mathrm{T}} \\ h'(x) & 0 \end{pmatrix} \middle| H \in \partial_x \frac{\partial L}{\partial x}(x, \lambda) \right\}. \tag{11}$$

Hence, if the matrix H_k is chosen as in (8), the matrix of the iteration system (7) satisfies the inclusion

$$\begin{pmatrix} H_k & (h'(x^k))^{\mathrm{T}} \\ h'(x^k) & 0 \end{pmatrix} \in \partial\Phi(u^k).$$

Therefore, local convergence properties of semismooth SQP follow from the corresponding general results for SNM (see [17,18,25,27] and [6, Chapter 7]). Specifically, primal–dual local superlinear convergence of semismooth SQP is obtained under the following assumptions:

(i) The derivatives of f and h are semismooth at a stationary point $\bar{x}$ of problem (1) (hence, the mapping Φ defined in (9) is semismooth at the solution $\bar{u} = (\bar{x}, \bar{\lambda})$ of Equation (10), where $\bar{\lambda}$ is the Lagrange multiplier associated with $\bar{x}$).
(ii) Φ is CD-regular at $\bar{u}$, i.e.

$$\text{all } \Lambda \in \partial\Phi(\bar{u}) \text{ are nonsingular.} \tag{12}$$

From (11) it easily follows that (12) holds if the regularity condition (5) is satisfied at $\bar{x}$ and

$$\langle H\xi, \xi \rangle > 0 \quad \forall \xi \in \ker h'(\bar{x}) \setminus \{0\}, \forall H \in \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}). \tag{13}$$

Condition (13) can be regarded as a natural counterpart of the standard second-order sufficient optimality condition for problem (1) with Lipschitzian derivatives, originally established in [16].

Remark 1 Condition (13) can be replaced by a seemingly weaker but actually equivalent condition

$$\langle H\xi, \xi \rangle > 0 \quad \forall \xi \in \ker h'(\bar{x}) \setminus \{0\}, \quad \forall H \in (\partial_B)_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}). \tag{14}$$

(The equivalence holds because the set of matrices involved in (13) is the convex hull of the set of matrices involved in (14).)

The CD-regularity condition (12) can be relaxed when more specific choices of H_k are employed, as is often the case in applications (when there is more than one element in the Clarke generalized Jacobian, the choice is usually not arbitrary but follows some rule tailored to the given problem structure). To this end, let $\Delta : \mathbb{R}^n \times \mathbb{R}^l \to 2^{\mathbb{R}^{n \times n}}$ be a multifunction such that

$$\Delta(x, \lambda) \subset \partial_x \frac{\partial L}{\partial x}(x, \lambda) \quad \forall x \in \mathbb{R}^n, \quad \forall \lambda \in \mathbb{R}^l, \tag{15}$$

and consider the process with iteration systems of the form (7), where

$$H_k \in \Delta(x^k, \lambda^k) \tag{16}$$

for all k. The multifunction Δ accounts for the specific rules to choose the matrix H_k when there is more than one possibility. Set

$$\bar{\Delta} = \bar{\Delta}(\bar{x}, \bar{\lambda}) = \left\{ H \in \mathbb{R}^{n \times n} \left| \begin{array}{c} \exists \{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l, \{H_k\} \subset \mathbb{R}^{n \times n} \text{ such that} \\ H_k \in \Delta(x^k, \lambda^k) \forall k, \{(x^k, \lambda^k)\} \to (\bar{x}, \bar{\lambda}), \\ \{H_k\} \to H \end{array} \right. \right\}. \tag{17}$$

From (15), by the upper semicontinuity of the generalized Jacobian and the results in [11], it follows that

$$\bar{\Delta} \subset \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}). \tag{18}$$

In the local convergence analysis we can then replace the CD-regularity condition (12) by the assumption that

$$\begin{pmatrix} H & (h'(\bar{x}))^{\mathrm{T}} \\ h'(\bar{x}) & 0 \end{pmatrix} \text{ is nonsingular } \forall H \in \bar{\Delta}. \tag{19}$$

According to (11) and (18), (19), the set of such matrices is generally smaller than $\partial \Phi(\bar{u})$. Therefore, (19) is generally weaker than CD-regularity.

One obvious general possibility is to take

$$\Delta(x, \lambda) = (\partial_B)_x \frac{\partial L}{\partial x}(x, \lambda), \quad x \in \mathbb{R}^n, \quad \lambda \in \mathbb{R}^l. \tag{20}$$

Different (for example, problem-related) choices of $\Delta(\cdot)$ can also be useful; see Section 3. Similar in spirit structure-related constructions in the context of SNM for complementarity problems were employed in [12].

According to the discussion above, in the following theorem on primal–dual local superlinear convergence we replace the condition (19) by the cruder but 'optimization-related' conditions that are sufficient for it (cf. (13)).

THEOREM 2.1   *Let the derivatives of f and h be semismooth at a stationary point $\bar{x}$ of problem (1). Let $\bar{\lambda}$ be Lagrange multiplier associated to $\bar{x}$. Assume that the regularity condition (5) is satisfied at $\bar{x}$, and*

$$\langle H\xi, \xi \rangle > 0 \quad \forall \xi \in \ker h'(\bar{x}) \setminus \{0\}, \quad \forall H \in \bar{\Delta}, \tag{21}$$

where $\bar{\Delta}$ is defined in (17) for some fixed multifunction $\Delta : \mathbb{R}^n \times \mathbb{R}^l \to 2^{\mathbb{R}^{n \times n}}$ satisfying (15).

Then for any rule of choosing H_k satisfying (16), any starting point $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$ uniquely defines the sequence $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ such that for all k the point (x^{k+1}, λ^{k+1}) satisfies (7), and this sequence converges to $(\bar{x}, \bar{\lambda})$ at a superlinear rate. Moreover, the rate of convergence is quadratic provided the derivatives of f and h are strongly semismooth at $\bar{x}$.

2.2   *Quasi-Newton versions and primal superlinear convergence*

While in some applications the problem structure admits calculus rules that allow to compute an element H_k of the generalized Hessian exactly, this may not be the case in general. It is therefore meaningful to ask 'how much' the inclusion (8) can be violated without destroying the primal superlinear rate (assuming convergence, i.e. in *a posteriori* analysis). Another motivation for considering possible violation of the inclusion (8) is related to globalization of SQP via linesearch for a penalty function. Indeed, as is well known, the matrices in $\partial_x(\partial L/\partial x)(x^k, \lambda^k)$ need not be positive definite under any natural assumptions, while to obtain a descent direction for the penalty function H_k has to be positive definite. This again motivates considering H_k that possibly does not satisfy (8).

In this section, we give an exact characterization of how (8) can be violated while still preserving the superlinear rate of primal convergence. Specifically, we introduce the following Dennis–Moré-type condition:

$$\max_{W \in \partial_x \frac{\partial L}{\partial x}(x^k, \lambda^k)} \| \pi_{\ker h'(\bar{x})}((H_k - W)(x^{k+1} - x^k)) \| = o(\|x^{k+1} - x^k\|). \tag{22}$$

Note that in the case of twice continuously differentiable data, this condition reduces to the standard Dennis–Moré condition for quasi-Newton SQP; see, e.g. [22, (18.63)].

It should be mentioned that for systems of equations, quasi-Newton versions of SNM based on classical matrix updates (e.g. BFGS, Broyden, etc.) actually require differentiability at the solution for superlinear convergence (see, e.g. the discussions in [10,21]). In other words, without differentiability at the solution, the relevant Dennis–Moré condition cannot be expected to hold, and superlinear convergence rate can be lost. This fact is well known as folklore, of course. But we could not find in the literature a formal justification we could cite. We thus provide the following example.

Example 2.2   Let $\Phi : \mathbb{R} \to \mathbb{R}$ be given by

$$\Phi(u) = \begin{cases} \dfrac{u}{2} + u^2 & \text{if } u \leq 0, \\ u - u^2 & \text{if } u > 0. \end{cases}$$

The solution of interest of the equation $\Phi(u) = 0$ is $\bar{u} = 0$; Φ is locally Lipschitz-continuous at $\bar{u}$ and differentiable everywhere except at $\bar{u}$. Evidently, $\partial_B \Phi(0) = \{1/2, 1\}$, $\partial \Phi(0) = [1/2, 1]$. In particular, Φ is CD-regular at 0.

The secant method (which is the one-dimensional instance of Broyden's quasi-Newton method; see, e.g. [22, (Section A.2)]) works in this example as follows: started from $u^0 > 0$ small enough, it generates $u^1 > 0$, then $u^2 < 0$ and $u^3 < 0$, then $u^4 > 0$, then $u^5 < 0$ and $u^6 < 0$, etc. (see Figure 1). After computing two subsequent iterates corresponding to the same smooth piece, namely, $u^{k-2} < 0$ and $u^{k-1} < 0$ for some $k \geq 4$, the method generates good secant approximation of the function, and the next iterate $u^k > 0$ is 'superlinearly closer' to the solution compared to u^{k-1}, i.e. $u^k = o(u^{k-1})$. Then, according to [22, (A.60)], we derive the estimate

$$\begin{aligned} u^{k+1} &= u^k - \frac{\Phi(u^k)(u^k - u^{k-1})}{\Phi(u^k) - \Phi(u^{k-1})} \\ &= u^k - \frac{(u^k - (u^k)^2)(u^k - u^{k-1})}{u^k - (u^k)^2 - u^{k-1}/2 - (u^{k-1})^2} \\ &= u^k - \frac{-u^k u^{k-1} + o(u^k u^{k-1})}{-u^{k-1}/2 + o(u^{k-1})} \end{aligned}$$
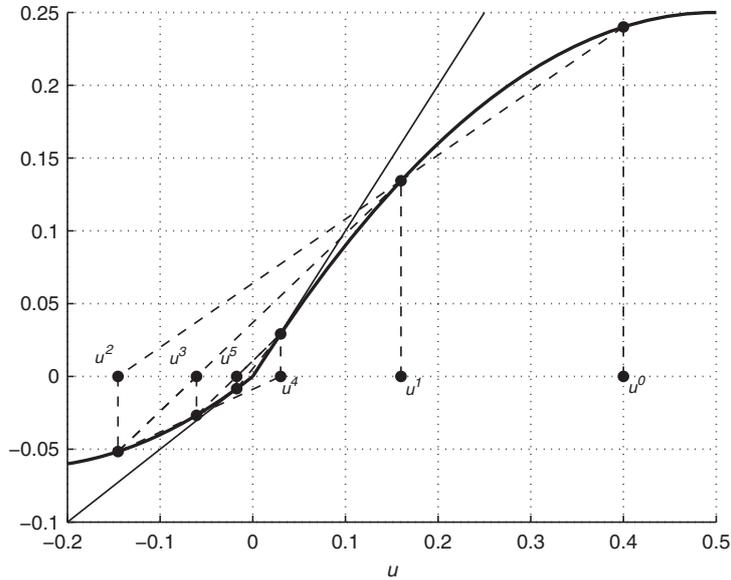
Figure 1.　Broyden's iterates in Example 2.2.

$$= u^k - 2u^k + o(u^k)$$
$$= -u^k + o(u^k),$$

and superlinear decrease of the distance to the solution is lost.

It is natural to say that the situation is likely the same for semismooth SQP if classical matrix updates were to be used. For this reason, when talking about quasi-Newton versions of semismooth SQP, we have in mind not so much standard matrix updates but rather other possible types of approximations. One example would be given in Section 3.2, where the structure of lifted MPCC is used to modify an element of the generalized Hessian of the Lagrangian to force the resulting matrix to be positive definite, and at the same time we have reasons to expect that this modification would satisfy the Dennis–Moré condition.

We start with necessary conditions for primal superlinear convergence of a general perturbed version of semismooth SQP. After that, we apply the obtained result to perturbations associated to quasi-Newton modifications and establish sufficient conditions for primal superlinear rate of convergence.

PROPOSITION 1　*Let the derivatives of f and h be semismooth at a stationary point $\bar{x}$ of problem (1). Let $\bar{\lambda}$ be a Lagrange multiplier associated with $\bar{x}$. Suppose the sequence $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ is such that for all k the relations*

$$\frac{\partial L}{\partial x}(x^k, \lambda^k) + W_k(x^{k+1} - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + \omega_1^k = 0,$$
$$h(x^k) + h'(x^k)(x^{k+1} - x^k) + \omega_2^k = 0$$

(23)

hold with some $W_k \in \partial_x(\partial L/\partial x)(x^k, \lambda^k)$ and some $\omega_1^k \in \mathbb{R}^n$ and $\omega_2^k \in \mathbb{R}^l$. Assume that $\{(x^k, \lambda^k)\}$ converges to $(\bar{x}, \bar{\lambda})$.

Then if the rate of convergence of $\{x^k\}$ is superlinear, it holds that

$$\pi_{\ker h'(\bar{x})}(\omega_1^k) = o(\|x^k - \bar{x}\|), \tag{24}$$

$$\omega_2^k = o(\|x^k - \bar{x}\|). \tag{25}$$

Proof   First, from the calculus of the Clarke generalized Jacobians [4] and from semismoothness of the derivatives of f and h at $\bar{x}$, one can easily derive the estimate

$$\frac{\partial L}{\partial x}(x^k, \lambda^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k) - W_k(x^k - \bar{x}) = o(\|x^k - \bar{x}\|). \tag{26}$$

From (23), employing (26), convergence of $\{\lambda^k\}$ to $\bar{\lambda}$ and superlinear convergence of $\{x^k\}$ to $\bar{x}$, we obtain that

$$
\begin{aligned}
\omega_1^k &= -\frac{\partial L}{\partial x}(x^k, \lambda^k) - W_k(x^{k+1} - x^k) - (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) \\
&= -\left(\frac{\partial L}{\partial x}(x^k, \lambda^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k) - W_k(x^k - \bar{x})\right) - \left(\frac{\partial L}{\partial x}(\bar{x}, \lambda^k) - \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda})\right) \\
&\quad - W_k(x^{k+1} - \bar{x}) - (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) \\
&= -(h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) - (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + o(\|x^k - \bar{x}\|) \\
&= -(h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) - ((h'(x^k))^{\mathrm{T}} - (h'(\bar{x}))^{\mathrm{T}})(\lambda^{k+1} - \lambda^k) + o(\|x^k - \bar{x}\|) \\
&= -(h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + o(\|x^k - \bar{x}\|),
\end{aligned}
$$

where it was also taken into account that by local uniform boundedness of generalized Jacobians of Lipschitzian mappings, the sequence $\{W_k\}$ is bounded. Applying the projector onto $\ker h'(\bar{x})$ to both sides of the equality above, using linearity of this projector and the fact that $\pi_{\ker h'(\bar{x})}(x) = 0$ for any $x \in \operatorname{im}(h'(\bar{x}))^{\mathrm{T}}$, we obtain (24).

Similarly,

$$
\begin{aligned}
\omega_2^k &= -h(x^k) - h'(x^k)(x^{k+1} - x^k) \\
&= -(h(x^k) - h(\bar{x}) - h'(\bar{x})(x^k - \bar{x})) \\
&\quad + (h'(x^k) - h'(\bar{x}))(x^k - \bar{x}) - h'(x^k)(x^{k+1} - \bar{x}) \\
&= o(\|x^k - \bar{x}\|),
\end{aligned}
$$

which gives (25). ∎

We now apply the above result to perturbations of semismooth SQP induced specifically by the quasi-Newton modifications. We show that the Dennis–Moré-type condition (22) is necessary for primal superlinear convergence, and it is also sufficient under the second-order condition (27). Simpler sufficient conditions for (27) are discussed in Remark 3, after the proof of the theorem.

THEOREM 2.3   *Let the derivatives of f and h be semismooth at a stationary point $\bar{x}$ of problem (1). Let $\bar{\lambda}$ be a Lagrange multiplier associated with $\bar{x}$. Suppose the sequence $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ is such that for all k the point (x^{k+1}, λ^{k+1}) satisfies (7) with some symmetric $n \times n$-matrix H_k. Assume that $\{(x^k, \lambda^k)\}$ converges to $(\bar{x}, \bar{\lambda})$.*

If the rate of convergence of $\{x^k\}$ is superlinear then the Dennis–Moré-type condition (22) holds.

Conversely, if

$$\liminf_{k \to \infty} \max_{W \in \partial_x \frac{\partial L}{\partial x}(x^k, \lambda^k)} \langle W\xi, \xi \rangle > 0 \quad \forall \xi \in \ker h'(\bar{x}) \setminus \{0\}, \tag{27}$$

and the Dennis–Moré-type condition (22) *holds, then the rate of convergence of* $\{x^k\}$ *is superlinear.*

Proof The first assertion follows immediately from Proposition 1. Indeed, $\{(x^k, \lambda^k)\}$ satisfies (23) with

$$\omega_1^k = (H_k - W_k)(x^{k+1} - x^k), \quad \omega_2^k = 0,$$

and with *any* choice of $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k, \lambda^k)$. Hence, by (24), it holds that

$$\max_{W \in \partial_x \frac{\partial L}{\partial x}(x^k, \lambda^k)} \|\pi_{\ker h'(\bar{x})}((H_k - W)(x^{k+1} - x^k))\| = o(\|x^k - \bar{x}\|).$$

Taking into account superlinear convergence of $\{x^k\}$ to $\bar{x}$, the last estimate is equivalent to (22).

We next prove the second assertion. By the second equation in (7), by Hoffman's error bound for linear systems [6, Lemma 3.2.3], and by the Mean Value Theorem, it holds that

$$\begin{aligned}
\text{dist}(x^{k+1} - \bar{x}, \ker h'(\bar{x})) &= O(\|h'(\bar{x})(x^{k+1} - \bar{x})\|) \\
&= O(\|h(x^k) + h'(x^k)(x^{k+1} - x^k) - h'(\bar{x})(x^{k+1} - \bar{x})\|) \\
&= O(\|h(x^k) - h(\bar{x}) - h'(\bar{x})(x^k - \bar{x})\|) \\
&\quad + O(\|h'(x^k) - h'(\bar{x})\| \|x^{k+1} - x^k\|) \\
&= o(\|x^k - \bar{x}\|).
\end{aligned}$$

Hence, for each k there exists $\xi^k \in \ker h'(\bar{x})$ such that

$$x^{k+1} - \bar{x} = \xi^k + o(\|x^k - \bar{x}\|). \tag{28}$$

From the first equality in (7), employing (26) and convergence of $\{\lambda^k\}$ to $\bar{\lambda}$, we obtain that for any choice of $W_k \in \partial_x(\partial L/\partial x)(x^k, \lambda^k)$ it holds that

$$\begin{aligned}
-H_k(x^{k+1} - x^k) &= \frac{\partial L}{\partial x}(x^k, \lambda^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) \\
&= \frac{\partial L}{\partial x}(x^k, \lambda^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k) - W_k(x^k - \bar{x}) \\
&\quad + \frac{\partial L}{\partial x}(\bar{x}, \lambda^k) - \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}) + W_k(x^k - \bar{x}) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) \\
&= W_k(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) \\
&\quad + o(\|x^k - \bar{x}\|) \\
&= W_k(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + o(\|x^k - \bar{x}\|).
\end{aligned}$$

Hence,

$$W_k(x^{k+1} - \bar{x}) = -(H_k - W_k)(x^{k+1} - x^k) - (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + o(\|x^k - \bar{x}\|).$$

By (28), by the inclusion $\xi^k \in \ker h'(\bar{x})$, by the fact that $\langle x, \xi \rangle = \langle \pi_{\ker h'(\bar{x})}(x), \xi \rangle$ for any $x \in \mathbb{R}^n$ and any $\xi \in \ker h'(\bar{x})$, and by (22), we then further obtain

$$
\begin{aligned}
\langle W_k \xi^k, \xi^k \rangle &= \langle W_k(x^{k+1} - \bar{x}), \xi^k \rangle + o(\|x^k - \bar{x}\| \|\xi^k\|) \\
&= -\langle (H_k - W_k)(x^{k+1} - x^k), \xi^k \rangle - \langle \lambda^{k+1} - \bar{\lambda}, h'(\bar{x})\xi^k \rangle + o(\|x^k - \bar{x}\| \|\xi^k\|) \\
&= -\langle (H_k - W_k)(x^{k+1} - x^k), \xi^k \rangle + o(\|x^k - \bar{x}\| \|\xi^k\|) \\
&= -\langle \pi_{\ker h'(\bar{x})}((H_k - W_k)(x^{k+1} - x^k)), \xi^k \rangle + o(\|x^k - \bar{x}\| \|\xi^k\|) \\
&= o(\|x^{k+1} - x^k\| \|\xi^k\|) + o(\|x^k - \bar{x}\| \|\xi^k\|).
\end{aligned}
\tag{29}
$$

From (27) it follows that for any k, one can choose $W_k \in \partial_x(\partial L/\partial x)(x^k, \lambda^k)$ in such a way that there exists $\gamma > 0$ satisfying

$$
\langle W_k \xi^k, \xi^k \rangle \geq \gamma \|\xi^k\|^2
$$

for all k large enough. Then (29) implies

$$
\gamma \|\xi^k\|^2 = o(\|x^{k+1} - x^k\| \|\xi^k\|) + o(\|x^k - \bar{x}\| \|\xi^k\|),
$$

and hence,

$$
\|\xi^k\| = o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|).
$$

Recalling (28), we then obtain the estimate

$$
\|x^{k+1} - \bar{x}\| = o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|),
$$

which by the standard argument implies that $\|x^{k+1} - \bar{x}\| = o(\|x^k - \bar{x}\|)$. ∎

Remark 2 From the proof of the sufficiency part of Theorem 2.3 it is evident that for any sequence $\{\Delta_k\}$ of sets such that $\Delta_k \subset \partial_x(\partial L/\partial x)(x^k, \lambda^k)$ for all k, the condition (22) can be replaced by the formally weaker condition

$$
\sup_{W \in \Delta_k} \|\pi_{\ker h'(\bar{x})}((H_k - W)(x^{k+1} - x^k))\| = o(\|x^{k+1} - x^k\|),
\tag{30}
$$

provided the condition (27) is replaced by the formally stronger condition

$$
\liminf_{k \to \infty} \max_{W \in \Delta_k} \langle W\xi, \xi \rangle > 0 \quad \forall \xi \in \ker h'(\bar{x}) \setminus \{0\}.
\tag{31}
$$

Combined with the other assumptions of Theorem 2.3, these two conditions still imply superlinear rate of convergence of $\{x^k\}$ to $\bar{x}$. But by the necessity part of Theorem 2.3, the stronger condition (22) must hold anyway. Thus, such a modification actually would not make Theorem 2.3 any sharper.

Remark 3 The condition (27) is automatically satisfied provided SOSC (13) holds (or (14), which is equivalent to (13), according to Remark 1).

Another example when (27) admits a more 'verifiable' sufficient counterpart is the following. Suppose that

$$\inf_{W \in \Delta(x^k, \lambda^k)} \|\pi_{\ker h'(\bar{x})}((H_k - W)(x^{k+1} - x^k))\| = o(\|x^{k+1} - x^k\|),$$

where $\Delta : \mathbb{R}^n \times \mathbb{R}^l \to 2^{\mathbb{R}^{n \times n}}$ is a given multifunction satisfying (15), and that (21) holds with $\bar{\Delta}$ defined according to (17). Then for each k there exists $W_k \in \Delta(x^k, \lambda^k)$ such that

$$\|\pi_{\ker h'(\bar{x})}((H_k - W_k)(x^{k+1} - x^k))\| = o(\|x^{k+1} - x^k\|),$$

which further implies (30) with $\Delta_k = \{W_k\}$. On the other hand, taking into account (15), local uniform boundedness of generalized Jacobians of Lipschitzian mappings, (17), and the inclusion $W_k \in \Delta(x^k, \lambda^k)$, condition (21) implies (31). Therefore, superlinear rate of convergence of $\{x^k\}$ follows from Remark 2.

2.3 *Globalization of convergence*

Since the usual globalization strategies for SQP do not require twice differentiability of the data, the local semismooth SQP discussed above can be globalized by standard techniques. For example, by linesearch for the l_1-penalty function (e.g. [3, Section 17]). This approach requires the matrices H_k in (7) to be bounded and uniformly positive definite, i.e. there must exist $\gamma > 0$ such that for all k

$$\langle H_k \xi, \xi \rangle \geq \gamma \|\xi\|^2 \quad \forall \xi \in \mathbb{R}^n. \tag{32}$$

This is enough for reasonable global convergence properties of the linesearch algorithm.

For specific problem classes, it may make sense to consider special choices of H_k that take into account structure. In Section 3, this would be done in the context of lifted MPCC. Furthermore, as already mentioned, in the semismooth case special choices of H_k may have better chances to satisfy the Dennis–Moré conditions than standard quasi-Newton updates.

3. Semismooth SQP for lifted MPCC

In this section, we first review relevant properties of MPCC and its lifted reformulation. We proceed to consider applying semismooth SQP to the lifted MPCC, and propose some special modification to the generalized Hessian of the Lagrangian which is attractive both for global convergence (it is positive definite) and for local convergence (it can be expected to satisfy the Dennis–Moré-type condition).

3.1 *Preliminaries*

In this subsection, we collect some facts concerning MPCC (2) and its lifted reformulation (3). Our notation and definitions are standard in MPCC literature, e.g. [8,14,31].

Let $\bar{x} \in \mathbb{R}^n$ be a feasible point of problem (2). We define the sets of indices

$$I_G = I_G(\bar{x}) = \{i = 1, \ldots, m \mid G_i(\bar{x}) = 0\},$$
$$I_H = I_H(\bar{x}) = \{i = 1, \ldots, m \mid H_i(\bar{x}) = 0\},$$
$$I_0 = I_G \cap I_H.$$

The special MPCC-Lagrangian for problem (2) is defined as

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

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 (2) 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

$$\frac{\partial \mathcal{L}}{\partial x}(\bar{x}, \bar{\mu}) = 0, \quad (\bar{\mu}_G)_{I_H \setminus I_G} = 0, \quad (\bar{\mu}_H)_{I_G \setminus I_H} = 0. \tag{33}$$

The point $\bar{x}$ is called strongly stationary if, in addition to (33),

$$(\bar{\mu}_G)_{I_0} \geq 0, \quad (\bar{\mu}_H)_{I_0} \geq 0. \tag{34}$$

When conditions (33) and (34) hold, $\bar{\mu}$ is called an MPCC-multiplier associated to the strongly stationary point $\bar{x}$ of problem (2).

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

$$G'_i(\bar{x}), \ i \in I_G, \quad H'_i(\bar{x}), \ i \in I_H \text{ are linearly independent.} \tag{35}$$

If a local solution $\bar{x}$ of problem (2) satisfies MPCC-LICQ, then $\bar{x}$ is a strongly stationary point and the associated MPCC-multiplier $\bar{\mu}$ is unique [31, Theorem 2].

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

$$(\bar{\mu}_G)_{I_0} > 0, \quad (\bar{\mu}_H)_{I_0} > 0. \tag{36}$$

Under ULSCC, various relevant second-order sufficient optimality conditions (see [14]) at a strongly stationary point $\bar{x}$ of problem (2) for an associated MPCC-multiplier $\bar{\mu}$ reduce to the following:

$$\left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in K(\bar{x}) \setminus \{0\}, \tag{37}$$

where

$$K(\bar{x}) = \{\xi \in \mathbb{R}^n \mid G'_{I_G}(\bar{x})\xi = 0, \quad H'_{I_H}(\bar{x})\xi = 0\}. \tag{38}$$

Let us now turn our attention to the lifted MPCC reformulation (3). Note first that the value $\bar{y}$ of the auxiliary variable y that corresponds to any given feasible point $\bar{x}$ of the original problem (2) is uniquely defined: the point $(\bar{x}, \bar{y})$ is feasible in (3) if and only if

$$\bar{y}_{I_H \setminus I_G} = -(G_{I_H \setminus I_G}(\bar{x}))^{1/2}, \quad \bar{y}_{I_G \setminus I_H} = (H_{I_G \setminus I_H}(\bar{x}))^{1/2}, \quad \bar{y}_{I_0} = 0. \tag{39}$$

The usual Lagrangian of the lifted problem (3) is given by

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

where $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ and $\lambda = (\lambda_G, \lambda_H) \in \mathbb{R}^m \times \mathbb{R}^m$. Then the Lagrange optimality system characterizing stationary points of (3) and the associated Lagrange multipliers is given by

$$\frac{\partial L}{\partial x}(x, y, \lambda) = 0, \quad \frac{\partial L}{\partial y}(x, y, \lambda) = 0,$$

$$(\min\{0, y\})^2 - G(x) = 0, \quad (\max\{0, y\})^2 - H(x) = 0. \tag{40}$$

Observe that

$$\frac{\partial L}{\partial x}(x, y, \lambda) = \frac{\partial \mathcal{L}}{\partial x}(x, \lambda), \tag{41}$$

$$\frac{\partial L}{\partial y_i}(x, y, \lambda) = 2(\lambda_G)_i \min\{0, y_i\} + 2(\lambda_H)_i \max\{0, y_i\}, \quad i = 1, \dots, m. \tag{42}$$

At any point $(x, y, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$, the right-hand side in (42) is not differentiable for any $i \in \{1, \dots, m\}$ such that $y_i = 0$. Therefore, the system (40) is nonsmooth. But under our assumptions, it is locally Lipschitz-continuous.

Furthermore, the following facts are known [15,32] or are easy to verify:

(i) The point $\bar{x}$ is a (local) solution of the original problem (2) if, and only if, $(\bar{x}, \bar{y})$ with $\bar{y}$ given by (39) is a (local) solution of the lifted MPCC reformulation (3).

(ii) If $\bar{x}$ is a strongly stationary point of (2) 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 (39) is a stationary point of problem (3) and $\bar{\lambda} = \bar{\mu}$ is an associated Lagrange multiplier.

(iii) Conversely, if $(\bar{x}, \bar{y})$ is a stationary point of problem (3), then $\bar{x}$ is a weakly stationary point of problem (2). 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 (2) and $\bar{\mu} = \bar{\lambda}$ is an associated MPCC-multiplier.

(iv) If the derivatives of G and H are semismooth at $\bar{x}$ (in particular, if the second derivatives of these mappings are continuous at $\bar{x}$), then the derivatives of the constraints functions in (3) are semismooth at $(\bar{x}, \bar{y})$.

(v) MPCC-LICQ at a point $\bar{x}$ feasible in (2) is equivalent to the regularity condition for the constraints of (3) at the point $(\bar{x}, \bar{y})$.

(vi) For a feasible point $\bar{x}$ of (2), if (ξ, η) belongs to the null space of the constraints' Jacobian of (3) at $(\bar{x}, \bar{y})$ then necessarily $\xi \in K(\bar{x})$.

(vii) For a strongly stationary point $\bar{x}$ of (2), for an associated MPCC-multiplier $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H)$, and for $\bar{y}$ given by (39), the set $(\partial_B)_{(x,y)}(\partial L/\partial(x, y))(\bar{x}, \bar{y}, \bar{\mu})$ is comprised by all the matrices of the form

$$\mathcal{H} = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu}) & 0 \\ 0 & 2\,\mathrm{diag}(a) \end{pmatrix},$$

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

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

(This follows from (33), (39) and (42).) The representation above then implies that the counterpart of the second-order sufficient optimality condition (14) (or equivalently, (13)) holds for problem (3) if, and only if, ULSCC (36) and second-order sufficient condition (37) hold.

We complete this section with the following observation.

Remark 4 The discussion in [32, Example 4.2] puts in evidence the following deficiency of the lifted MPCC reformulation with $s > 2$. In such reformulation, the objective function is constant along the 'vertical' tangent subspace to the feasible set at $(\bar{x}, \bar{y})$, i.e. the subspace consisting of $(\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying $\xi = 0$, $\eta_{I_H \setminus I_G} = 0$, $\eta_{I_G \setminus I_H} = 0$. Moreover, when $s > 2$, the constraints of the lifted problem cannot contribute to the Hessian of its Lagrangian along this subspace.

These two facts mean that no reasonable second-order sufficient optimality condition can hold for this problem at $(\bar{x}, \bar{y})$. However, in the lifted MPCC reformulation with $s = 2$, the constraints can appropriately contribute to (generalized) Hessians, and (14) (or equivalently, (13)) can hold in this case.

3.2 Local convergence of semismooth SQP for lifted MPCC

For the current iterate $(x^k, y^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$, with $\lambda^k = (\lambda_G^k, \lambda_H^k)$, an iteration of semismooth SQP (6) applied to the lifted MPCC (3) consists of solving the subproblem

$$
\begin{aligned}
\min \quad & \langle f'(x^k), x - x^k \rangle + \frac{1}{2}\left\langle \mathcal{H}_k \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix}, \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix} \right\rangle \\
\text{s.t.} \quad & (\min\{0, y^k\})^2 - G(x^k) - G'(x^k)(x - x^k) + 2B_{\min}(y^k)(y - y^k) = 0, \\
& (\max\{0, y^k\})^2 - H(x^k) - H'(x^k)(x - x^k) + 2B_{\max}(y^k)(y - y^k) = 0.
\end{aligned}
\tag{43}
$$

In the above, $\mathcal{H}_k$ is a symmetric $(n + m) \times (n + m)$-matrix, and for $y \in \mathbb{R}^m$ we denote

$$
B_{\min}(y) = \mathrm{diag}(\min\{0, y\}), \quad B_{\max}(y) = \mathrm{diag}(\max\{0, y\}).
\tag{44}
$$

Furthermore, the counterpart of the iteration system (7), characterizing stationary points of problem (43) and the associated Lagrange multipliers, has the form

$$
\begin{aligned}
\mathcal{H}_k \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix} + \begin{pmatrix} -(G'(x^k))^T \\ 2B_{\min}(y^k) \end{pmatrix}(\lambda_G - \lambda_G^k) + \begin{pmatrix} -(H'(x^k))^T \\ 2B_{\max}(y^k) \end{pmatrix}(\lambda_H - \lambda_H^k) &= -\begin{pmatrix} \dfrac{\partial \mathcal{L}}{\partial x}(x^k, \lambda^k) \\ \dfrac{\partial L}{\partial y}(x^k, y^k, \lambda^k) \end{pmatrix}, \\
-G'(x^k)(x - x^k) + 2B_{\min}(y^k)(y - y^k) &= -((\min\{0, y^k\})^2 - G(x^k)), \\
-H'(x^k)(x - x^k) + 2B_{\max}(y^k)(y - y^k) &= -((\max\{0, y^k\})^2 - H(x^k)).
\end{aligned}
\tag{45}
$$

By direct calculations, employing (41) and (42), it can be seen that the basic choice of $\mathcal{H}_k$ (corresponding to (16) and (20)), is of the form

$$
\mathcal{H}_k = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial x^2}(x^k, \lambda^k) & 0 \\ 0 & 2\,\mathrm{diag}(a(y^k, \lambda^k)) \end{pmatrix},
\tag{46}
$$

where for $y \in \mathbb{R}^m$ and $\lambda = (\lambda_G, \lambda_H) \in \mathbb{R}^m \times \mathbb{R}^m$ the vector $a(y, \lambda) \in \mathbb{R}^m$ is defined by

$$
a_i(y, \lambda) = \begin{cases} (\lambda_G)_i & \text{if } y_i < 0, \\ (\lambda_G)_i \text{ or } (\lambda_H)_i & \text{if } y_i = 0, \quad i = 1, \dots, m. \\ (\lambda_H)_i & \text{if } y_i > 0, \end{cases}
\tag{47}
$$

Specifically, $(\partial_B)_{(x,y)}(\partial L/\partial(x, y))(x^k, y^k, \lambda^k)$ is comprised by all such matrices $\mathcal{H}_k$. Local convergence properties of semismooth SQP for lifted MPCC follow directly from [15], where SNM for lifted MPCC was considered (recall that for the purposes of local analysis SQP for equality-constrained problems is just a Newton method). The same result can, of course, be derived from Theorem 2.1 by figuring out what is needed for semismoothness of the derivatives of the constraints in (3), and what are the counterparts for problem (3) of the regularity condition (5) and of the second-order condition (13) (or equivalently (14)). Specifically, the convergence statement is the following (see [15] for a proof).

THEOREM 3.1 *Let f, G and H be twice differentiable, with their second derivatives continuous at a strongly stationary point $\bar{x}$ of problem* (2). *Let MPCC-LICQ* (35) *be satisfied at $\bar{x}$, and let $\bar{\mu}$ be the (unique) MPCC-multiplier associated to $\bar{x}$. Assume, finally, that ULSCC* (36) *and the second-order sufficient optimality condition* (37) *are satisfied.*

Then for any rule for choosing $\mathcal{H}_k$ satisfying (46) *and* (47), *any starting 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 the sequence $\{(x^k, y^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ such that for all k the point $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ satisfies* (45), *and this sequence converges to $(\bar{x}, \bar{y}, \bar{\mu})$ at a superlinear rate. Moreover, if the second derivatives of f, G and H are locally Lipschitz-continuous with respect to $\bar{x}$ then the rate of convergence is quadratic.*

A discussion of relations of the local convergence result in Theorem 3.1 with local convergence of other Newton-type methods for MPCC (2) can be found in [15].

Globalization strategy suggested in [15] relies on linesearch for the squared residual of the Lagrange system (40), which happens to be differentiable even though the Lagrange system itself is not. This strategy has some theoretical advantages, but it is aimed at solving the Lagrange system rather than the original optimization problem. From the optimization point of view, it may be more promising to employ the linesearch procedure for the l_1-penalty function, as is common when solving general optimization problems (see the discussion in Section 2.3). However, matrices $\mathcal{H}_k$ defined according to (46) and (47) are not necessarily positive definite, and their direct use in the globalization scheme is hardly possible: the direction $p^k = (\tilde{x}^{k+1} - x^k, \tilde{y}^{k+1} - y^k)$, where $(\tilde{x}^{k+1}, \tilde{y}^{k+1})$ is a stationary point of problem (43) with this $\mathcal{H}_k$, is not necessarily a direction of descent for the l_1-penalty function. Thus, $\mathcal{H}_k$ must be appropriately modified. One possibility is to use standard quasi-Newton approximations of $\mathcal{H}_k$; see [3, Section 18.2], [22, p. 538]. But this would destroy the diagonal structure in (46). Moreover, standard quasi-Newton schemes do not achieve local superlinear convergence in the nonsmooth case, in general. It thus seems more reasonable to preserve and employ the special lifted MPCC structure. Observe that the upper-left block in (46) corresponds to twice continuously differentiable data, while all the nonsmoothness is contained in the lower-right diagonal block. The idea is to use some usual quasi-Newton updates on the smooth part only, at the same time modifying the nonsmooth part to make the matrix positive definite. Specifically, we re-define $\mathcal{H}_k$ as follows:

$$\mathcal{H}_k = \begin{pmatrix} H_k & 0 \\ 0 & 2\,\mathrm{diag}(a^k) \end{pmatrix}, \tag{48}$$

where H_k is some symmetric positive definite approximation of $(\partial^2 \mathcal{L}/\partial x^2)(x^k, \lambda^k)$ (say, built by BFGS updates with Powell's modification; see, e.g. [22, pp. 536, 537]), and for $y \in \mathbb{R}^m$ and $\lambda = (\lambda_G, \lambda_H) \in \mathbb{R}^m \times \mathbb{R}^m$ the vector $a^k \in \mathbb{R}^m$ is defined by

$$a_i^k = \max\{a_i(y^k, \lambda^k), \rho(\sigma(x^k, y^k, \lambda^k))\}, \quad i = 1, \ldots, m, \tag{49}$$

where a_i is defined by (47), $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ is some continuous function such that $\rho(0) = 0$ and $\rho(t)$ is separated from 0 when t is separated from 0, and $\sigma : \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ is the residual of the Lagrange system (40) of problem (3):

$$\sigma(x, y, \lambda) = \left\| \begin{pmatrix} \dfrac{\partial \mathcal{L}}{\partial x}(x, \lambda) \\ \dfrac{\partial L}{\partial y}(x, y, \lambda) \\ (\min\{0, y\})^2 - G(x) \\ (\max\{0, y\})^2 - H(x) \end{pmatrix} \right\|. \tag{50}$$

As is easy to see, if H_k is positive definite and $\sigma(x^k, y^k, \lambda^k) > 0$, then it follows from (32), (48) and (49) that $\mathcal{H}_k$ constructed above is also positive definite. The obtained SQP direction is then a direction of descent for the l_1-penalty function and can be used in globalization via linesearch (see Section 3.3). At the same time, as we show next, primal superlinear convergence of this specific version of quasi-Newton semismooth SQP holds under what appears to be a natural Dennis–Moré-type condition *on the smooth part of the problem*. We thus have reasons to expect this condition to hold.

THEOREM 3.2 *Let the derivatives of f, G and H be semismooth at a strongly stationary point $\bar{x}$ of problem (2), with an associated MPCC-multiplier $\bar{\mu}$. Let the sequence $\{(x^k, y^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times (\mathbb{R}^m \times \mathbb{R}^m)$ be such that for each k the point $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ satisfies (45), where the matrix $\mathcal{H}_k$ is of the form (48) with some symmetric $n \times n$-matrix H_k, and with $a^k \in \mathbb{R}^m$ defined by (49), (47) and (50) for some function $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\rho(t) \to 0$ as $t \to 0+$. Assume that $\{(x^k, y^k, \lambda^k)\}$ converges to $(\bar{x}, \bar{y}, \bar{\mu})$.*

If the rate of convergence of $\{(x^k, y^k)\}$ is superlinear then the Dennis–Moré-type condition

$$\left\| \pi_{K(\bar{x})} \left(\left(H_k - \frac{\partial^2 \mathcal{L}}{\partial x^2}(x^k, \lambda^k) \right)(x^{k+1} - x^k) \right) \right\| = o \left(\left\| \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\| \right) \quad (51)$$

holds.

Conversely, if ULSCC (36), the second-order sufficient optimality condition (37) and the Dennis–Moré-type condition (51) hold, then the rate of convergence of $\{(x^k, y^k)\}$ is superlinear.

Proof Since $\{y^k\}$ converges to $\bar{y}$, from (39) it follows that for all k large enough

$$y^k_{I_H \setminus I_G} < 0, \quad y^k_{I_G \setminus I_H} > 0,$$

and hence, by (47),

$$a_{I_H \setminus I_G}(y^k, \lambda^k) = (\lambda^k_G)_{I_H \setminus I_G}, \quad a_{I_G \setminus I_H}(y^k, \lambda^k) = (\lambda^k_H)_{I_G \setminus I_H}.$$

By convergence of $\{\lambda^k\}$ to $\bar{\mu}$ and by (33), this implies that

$$\{a_{I_H \setminus I_G}(y^k, \lambda^k)\} \to (\bar{\mu}_G)_{I_H \setminus I_G} = 0, \quad \{a_{I_G \setminus I_H}(y^k, \lambda^k)\} \to (\bar{\mu}_H)_{I_G \setminus I_H} = 0. \quad (52)$$

On the other hand, for $i \in I_0$ from (47) it follows that, for each k, $a_i(y^k, \lambda^k)$ equals either $(\lambda^k_G)_i$ or $(\lambda^k_H)_i$. Note that according to (34),

$$\{(\lambda^k_G)_{I_0}\} \to (\bar{\mu}_G)_{I_0} \geq 0, \quad \{(\lambda^k_H)_{I_0}\} \to (\bar{\mu}_H)_{I_0} \geq 0,$$

and hence, we conclude that

$$\liminf_{k \to \infty} a_i(y^k, \lambda^k) \geq 0 \quad \forall i \in I_0.$$

Combining the last relation with (52), we get

$$\liminf_{k \to \infty} a_i(y^k, \lambda^k) \geq 0 \quad \forall i = 1, \dots, m. \quad (53)$$

We next show that

$$\lim_{k \to \infty} (a^k_i - a_i(y^k, \lambda^k)) = 0 \quad \forall i = 1, \dots, m. \quad (54)$$

Fix any $i \in \{1, \dots, m\}$. If $a^k_i = a_i(y^k, \lambda^k)$ for all k large enough, (54) is obvious. Otherwise, define the subsequence $\{k_j\}$ collecting all the indices such that $a^{k_j}_i \neq a_i(y^{k_j}, \lambda^{k_j})$. By (49), it

evidently holds that

$$\liminf_{k\to\infty}(a_i^k - a_i(y^k, \lambda^k)) \geq 0, \tag{55}$$

and

$$\begin{aligned}
\limsup_{k\to\infty}(a_i^k - a_i(y^k, \lambda^k)) &= \limsup_{j\to\infty}(a_i^{k_j} - a_i(y^{k_j}, \lambda^{k_j})) \\
&= \limsup_{j\to\infty}(\rho(\sigma(x^{k_j}, y^{k_j}, \lambda^{k_j})) - a_i(y^{k_j}, \lambda^{k_j})) \\
&= \limsup_{j\to\infty}(-a_i(y^{k_j}, \lambda^{k_j})) \\
&= -\liminf_{j\to\infty} a_i(y^{k_j}, \lambda^{k_j}) \\
&\leq 0,
\end{aligned}$$

where the third equality holds since $\sigma(x^k, y^k, \lambda^k) \to 0$, and the inequality is by (53). Combining the latter relation with (55), we obtain (54).

With (54) in hand, according to (46) condition (51) is equivalent to the following:

$$\left\|\pi_{K(\bar{x})\times\mathbb{R}^m}\left((\mathcal{H}_k - \mathcal{W}_k)\begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix}\right)\right\| = o\left(\left\|\begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix}\right\|\right),$$

where

$$\mathcal{W}_k = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial x^2}(x^k, \lambda^k) & 0 \\ 0 & 2\,\mathrm{diag}(a(y^k, \lambda^k)) \end{pmatrix} \in (\partial_B)_{(x,y)}\frac{\partial L}{\partial(x, y)}(x^k, y^k, \lambda^k).$$

The result now follows from Theorem 2.3 and Remark 3, and from the above-mentioned fact that ULSCC (36) and second-order sufficient condition (37) imply the counterpart of the second-order sufficient optimality condition (14) for problem (3). ∎

3.3 Globalization of convergence of SQP for lifted MPCC

Let $\psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+$ be the l_1-penalty for the constraints in (3), i.e.

$$\psi(x, y) = \left\|\begin{pmatrix} (\min\{0, y\})^2 - G(x) \\ (\max\{0, y\})^2 - H(x) \end{pmatrix}\right\|_1.$$

Define the corresponding family of penalty functions: $\varphi_c : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$,

$$\varphi_c(x, y) = f(x) + c\psi(x, y),$$

where $c > 0$ is the penalty parameter.

The following is the usual linesearch SQP method, but with the specific choice of the matrix in the objective function of subproblems, tailored to the lifted MPCC structure.

ALGORITHM 1 *Choose the parameters $\bar{c} > 0$ and $\varepsilon, \theta \in (0, 1)$, and a continuous function $\rho :$ $\mathbb{R}_+ \to \mathbb{R}_+$ such that $\rho(0) = 0$ and $\rho(t)$ is separated from 0 when t is separated from 0. Choose $(x^0, y^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^m \times \mathbb{R}^m)$ and set $k = 0$.*

(1) *If $\sigma(x^k, y^k, \lambda^k) = 0$, stop. Otherwise, choose a symmetric positive definite matrix $H_k \in \mathbb{R}^{n \times n}$ and define $\mathcal{H}_k$ according to (47)–(50). Compute $(\tilde{x}^{k+1}, \tilde{y}^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^m$ as a stationary point of problem (43), and an associated Lagrange multiplier $\lambda^{k+1} \in \mathbb{R}^m \times \mathbb{R}^m$. Set $\xi^k = \tilde{x}^{k+1} - x^k$, $\eta^k = \tilde{y}^{k+1} - y^k$, $p^k = (\xi^k, \eta^k)$.*

(2) *Choose*

$$c_k \geq \|\lambda^{k+1}\|_\infty + \bar{c}$$

and compute

$$\varphi'_{c_k}((x^k, y^k); p^k) = \langle f'(x^k), \xi^k \rangle - c_k \psi(x^k, y^k)$$

(see, e.g. [22, (18.29)] for the formula of this directional derivative).

(3) *Set $\alpha = 1$. If the inequality*

$$\varphi_{c_k}((x^k, y^k) + \alpha p^k) \leq \varphi_{c_k}(x^k, y^k) + \varepsilon\alpha\varphi'_{c_k}((x^k, y^k); p^k) \tag{56}$$

is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (56) again, etc., until (56) becomes valid.

(4) *Set*

$$x^{k+1} = x^k + \alpha_k\xi^k, \quad y^{k+1} = y^k + \alpha_k\eta^k, \tag{57}$$

adjust k by 1, and go to step 1.

It is interesting to observe that constraints in the SQP subproblems (43) for the lifted MPCC have a certain 'elastic mode' feature, which allows us to expect that the subproblems may be feasible without any special modifications. Specifically, if we define the index sets

$$J^k_+ = \{i = 1, \ldots, m \mid y^k_i > 0\}, \quad J^k_- = \{i = 1, \ldots, m \mid y^k_i < 0\},$$
$$J^k_0 = \{1, \ldots, m\} \setminus (J^k_+ \cup J^k_-),$$

then the constraints in (43) have the form

$$G_i(x^k) + G'_i(x^k)(x - x^k) = 0, \quad i \in J^k_+ \cup J^k_0,$$
$$H_i(x^k) + H'_i(x^k)(x - x^k) = 0, \quad i \in J^k_- \cup J^k_0,$$
$$G_i(x^k) + G'_i(x^k)(x - x^k) + (y^k_i)^2 - 2y^k_i y_i = 0, \quad i \in J^k_-,$$
$$H_i(x^k) + H'_i(x^k)(x - x^k) + (y^k_i)^2 - 2y^k_i y_i = 0, \quad i \in J^k_+.$$

Note that if the first group of constraints (the one not involving the variable y) is consistent then all the constraints of the subproblem are consistent (for each x satisfying the first group of constraints there are the uniquely defined y_i, $i \in J^k_+ \cup J^k_-$, such that (x, y) satisfies the second group of constraints). In particular, if the $m + |J^k_0|$ gradients in the first group of constraints are linearly independent, the SQP subproblem (43) is feasible. The latter can be expected to happen, since usually m is smaller than n.

We next show that Algorithm 1 has reasonable global convergence properties. Even though the algorithm looks rather standard, some special analysis is needed. The reason is that while the matrices $\mathcal{H}_k$ (tailored to the lifted MPCC structure) are positive definite, they are not uniformly positive definite as required in standard convergence analysis of SQP. As already commented, our motivation for constructing special matrices is two-fold. First, usual quasi-Newton updates would destroy the diagonal structure in the lifted MPCC formulation. Second, usual quasi-Newton updates are unlikely to yield superlinear convergence, because of nonsmoothness.

THEOREM 3.3 *Let the derivatives of f, G and H be Lipschitz-continuous on $\mathbb{R}^n$. Assume that the matrices H_k are chosen in Algorithm 1 in such a way that the sequence $\{H_k\}$ is bounded, and there exists $\gamma > 0$ such that (32) holds for all k. Let $\{(x^k, y^k, \lambda^k)\}$ be a sequence generated by Algorithm 1, and suppose that*

$$c_k = c > 0 \tag{58}$$

holds for all k large enough.

Then either the algorithm stops on some iteration k with (x^k, y^k, λ^k) satisfying the Lagrange optimality system (40), or at least one of the following occurs:

(1) *It holds that*

$$\varphi_c(x^k, y^k) \to -\infty \quad \text{as } k \to \infty. \tag{59}$$

(2) *There exists a subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ such that*

$$\sigma(x^{k_j}, y^{k_j}, \lambda^{k_j}) \to 0 \quad \text{as } j \to \infty, \tag{60}$$

and in particular, every accumulation point of $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ satisfies the Lagrange optimality system (40). Moreover, for any subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ such that

$$\liminf_{j \to \infty} \sigma(x^{k_j}, y^{k_j}, \lambda^{k_j}) > 0, \tag{61}$$

it holds that

$$\{\xi^{k_j}\} \to 0, \quad \{\eta^{k_j}\} \to 0, \quad \sigma(x^{k_j}, y^{k_j}, \lambda^{k_j+1}) \to 0 \quad \text{as } j \to \infty,$$

and in particular, for every accumulation point $(\bar{x}, \bar{y}, \bar{\lambda})$ of $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ it holds that $\{(x^{k_j+1}, y^{k_j+1})\}$ converges to $(\bar{x}, \bar{y})$, and every accumulation point of the subsequence $\{(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1})\}$ satisfies the Lagrange optimality system (40).

Proof We consider the case when $\sigma(x^k, y^k, \lambda^k) \neq 0$ for all k. In this case, it follows from (32), (48) and (49) that $\mathcal{H}_k$ is positive definite. Then, by the standard analysis of linesearch SQP (e.g. [3, Theorem 17.2]), it follows that $(\tilde{x}^{k+1}, \tilde{y}^{k+1}, \lambda^{k+1})$ (and hence, p^k) is well defined,

$$\varphi_{c_k}'((x^k, y^k); p^k) \leq -\langle \mathcal{H}_k p^k, p^k \rangle - \bar{c}\psi(x^k, y^k) < 0, \tag{62}$$

and the linesearch procedure terminates with some $\alpha_k > 0$.

If there is no subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ satisfying (60) then there exists $\delta > 0$ such that

$$\sigma(x^k, y^k, \lambda^k) \geq \delta \tag{63}$$

for all k. Hence, there exists $\tilde{\delta} > 0$ such that $\rho(\sigma(x^k, y^k, \lambda^k)) \geq \tilde{\delta}$ for all k. From (32), (48) and (49) it then follows that

$$\forall (\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^m \quad \langle \mathcal{H}_k(\xi, \eta), (\xi, \eta) \rangle \geq \langle H_k\xi, \xi \rangle + 2\rho(\sigma(x^k, y^k, \lambda^k))\|\eta\|^2$$

$$\geq \min\{\gamma, 2\tilde{\delta}\}\|(\xi, \eta)\|^2.$$

In this case, the matrices $\mathcal{H}_k$ are uniformly positive definite, and by the standard analysis (e.g. [3, Theorem 17.2]), there exists $\bar{\alpha} > 0$ such that $\alpha_k \geq \bar{\alpha}$ for all k, and

$$\varphi_c(x^{k+1}, y^{k+1}) \leq \varphi_c(x^k, y^k) + \bar{\alpha}\varepsilon\varphi_c'((x^k, y^k); p^k).$$

Then either $\{\varphi_c(x^k, y^k)\}$ is unbounded below (i.e. (59) holds) or it follows that $\{\varphi_c(x^k, y^k)\}$ converges and hence,

$$\varphi_c'((x^k, y^k); p^k) \to 0 \quad \text{as } k \to \infty,$$

in which case (62) implies that

$$\{\xi^k\} \to 0, \quad \{\eta^k\} \to 0, \quad \sigma(x^k, y^k, \lambda^{k+1}) \to 0 \quad \text{as } k \to \infty.$$

(The last relation is passing onto the limit in the first relation in (45) and in (62).)

Note that the fact that $c_k = c$ for all k large enough implies that the sequence $\{\lambda^k\}$ is bounded. Then taking into account (57) and uniform Lipschitz-continuity of $\sigma(\cdot, \cdot, \lambda^k)$, we conclude that

$$\|x^{k+1} - x^k\| \to 0, \quad \|y^{k+1} - y^k\| \to 0, \quad \sigma(x^{k+1}, y^{k+1}, \lambda^{k+1}) \to 0 \quad \text{as } k \to \infty,$$

where the last asymptotic relation contradicts (63). We thus established that if (59) does not hold then there exists a subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ satisfying (60).

Finally, for any subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ satisfying (61), the assertions of the theorem follow by repeating the argument above for the subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$. ∎

The convergence result stated above gives reasons to expect that an accumulation point $(\bar{x}, \bar{y})$ of the sequence $\{(x^k, y^k)\}$ would be stationary for problem (3), which further implies that $\bar{x}$ would be a weakly stationary point of (2).

4. Numerical results

This section presents some preliminary numerical results comparing the algorithm proposed above with some alternative quasi-Newton approaches on a set of MPCC test problems derived from MacMPEC [19]. Our selection of test problems is the same as in [14,15]. Specifically, we select all the problems in MacMPEC which have no more than 10 variables and do not have any inequality constraints apart from complementarity constraints. We ignore simple bounds, which of course sometimes changes solutions and stationary points of these problems. We end up with the set of 38 problems that fit the format considered in this paper.

In what follows, we use the abbreviation `Lifted ssSQP BFGS` for Algorithm 1 with H_k computed according to the BFGS rule with Powell's modification (see, e.g. [22, pp. 536, 537]). We compare this method with its direct alternative, which can be formally described as Algorithm 1 but with the full matrix $\mathcal{H}_k$ (including the part corresponding to nonsmoothness) computed by BFGS with Powell's modification, instead of our proposal (47)–(50). We call this algorithm `Lifted ssSQP full BFGS`. Another alternative we implement is called `Lifted SNM SR1`, which is a quasi-Newton version of SNM for lifted MPCC suggested in [15, Algorithm 4.1], where true Hessians are replaced by SR1 approximations [22, p. 538]. The fourth algorithm chosen for comparison is `SQP slacks BFGS`, which is also the quasi-Newton SQP (BFGS with Powell's modification) with linesearch for the l_1-penalty function, but applied to the following (non-lifted) slack reformulation of the original MPCC (2):

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

It is known that introducing slacks can be advantageous for numerical solution by SQP [7,8] (see [14] for details of our implementation of this algorithm). The SQP methods were implemented without any tools for tackling possible infeasibility of subproblems, and without any tools for

avoiding the Maratos effect. These are issues not related to our proposal, and they should be dealt with as in the usual (but more sophisticated) SQP implementations.

The parameters of Algorithm 1 were chosen as follows: $\varepsilon = 10^{-4}, \theta = 0.5$, and $\bar{c} = 1$. At each iteration, we define c_k by the simplest rule:

$$c_k = \|\lambda^{k+1}\|_\infty + \bar{c}.$$

Note that this choice may violate condition (58) in Theorem 3.3 on global convergence, but it is attractive because of its simplicity. It also works well in our experiments, as it allows for a decrease of the penalty parameter. The latter is particularly useful when large values of this parameter arise on early iterations. Allowing to decrease the penalty parameter can be combined with the need for (58) using some more involved updating rules for c_k (see, e.g. [22, Section 18.3] and [3, Section 17.1]).

The function ρ in (49) was defined as follows:

$$\rho(t) = \begin{cases} t & \text{if } t < 0.1, \\ 0.1 & \text{if } t \geq 0.1. \end{cases}$$

All computations were performed in Matlab environment. For SQP slacks BFGS, the QP-subproblems were solved by the built-in Matlab QP-solver. For Lifted ssSQP BFGS, Lifted ssSQP full BFGS and Lifted SNM SR1, we used the same stopping criterion of the form

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

where σ is defined in (50). SQP slacks BFGS was stopped when the residual of the first-order optimality conditions of the original MPCC (2) becomes smaller than 10^{-6}.

Failure was declared when the needed accuracy was not achieved after 500 iterations, or when the method in question failed to make the current step, for whatever reason.

We performed 100 runs of each algorithm from the same sample of randomly generated starting points. Primal starting points were generated in a cubic neighborhood around the solution (solutions were found in the course of our experiments), with the edge of the cube equal to 20. In the cases of Lifted ssSQP BFGS, Lifted ssSQP full BFGS and Lifted SNM SR1 algorithms, we defined the starting value y^0 of the auxiliary variable as follows:

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

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 (for SQP slacks BFGS, with additional nonnegativity restriction for the components corresponding to inequality constraints). Convergence to solution was declared when the distance from the last primal iterate to the solution was smaller than 10^{-3}.

Figure 2 reports on the average numbers of major and minor iteration counts for all the algorithms over successful runs, in the form of a performance profile [5]. For Lifted ssSQP BFGS, Lifted ssSQP full BFGS and Lifted SNM SR1, minor and major iterations counts are the same, since these algorithms are QP-free: each major iteration consists of solving one linear system, followed by linesearch. At the same time, SQP slacks 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 iteration, and each minor iteration includes solving a linear system. 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 in
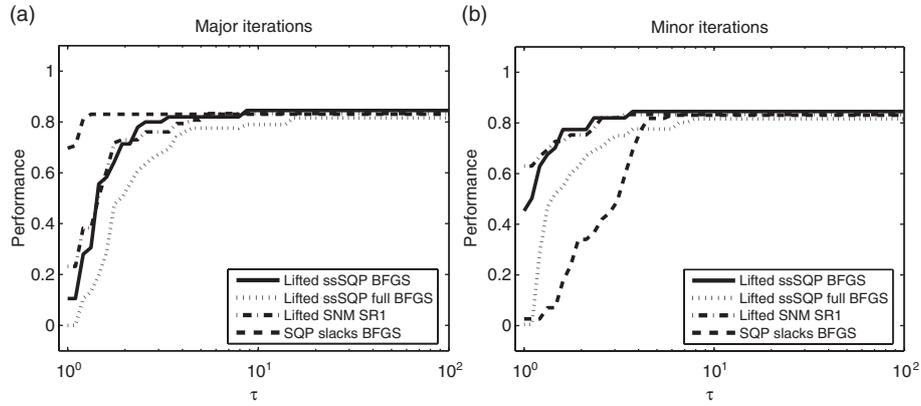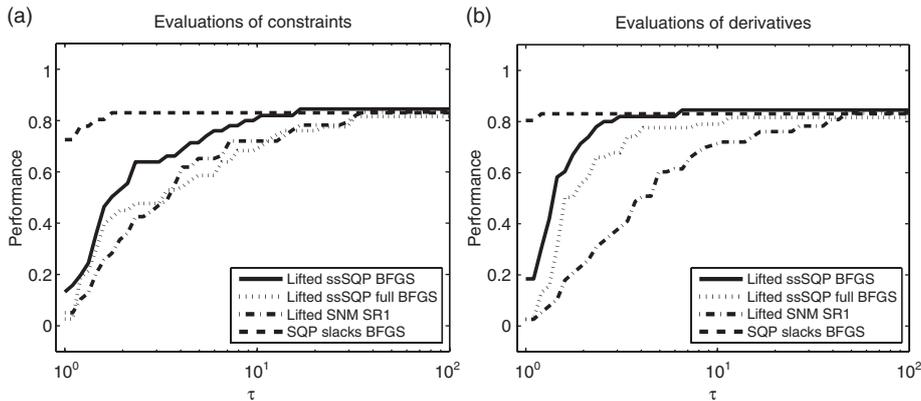
Figure 2.   Number of iterations.

Figure 3.   Evaluations of constraints/derivatives values.

Figure 2; the constraint/derivative evaluation count in Figure 3) was no more than τ times worse (bigger) than the best result among the four 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 (i.e. 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).

One can see from Figure 2 that in terms of efficiency, `Lifted ssSQP BFGS` and `Lifted SNM SR1` are outperformed by `SQP slacks BFGS` by major iteration count, but for minor iterations the picture is opposite. Also, `Lifted ssSQP full BFGS` is an evident loser in these comparisons, which supports the intuition behind our special construction. At the same time, all methods demonstrate similar robustness. For `Lifted ssSQP BFGS`, many failures (more than 33% of the runs) were observed for problems `ex9.1.2`, `ex9.1.4`, `ralph1`, `scholtes5`; in most cases, failures were due to infeasibility of subproblems.

Figure 3 reports on the average numbers of evaluations of constraint functions and their derivatives, and the picture is quite similar to the one with major iterations. However, here we can see that `Lifted ssSQP BFGS` outperforms `Lifted SNM SR1`.

Figure 4(a) reports on the average numbers of evaluations of the objective function (`Lifted SNM SR1` is not presented in this figure since it does not compute objective function values).
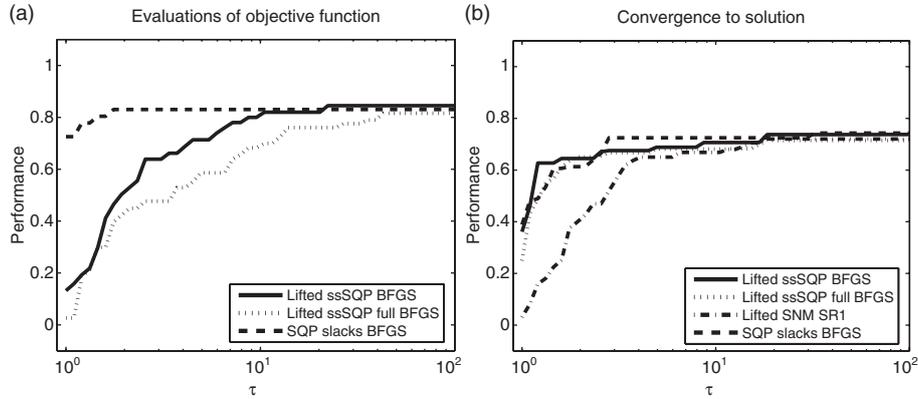
Figure 4. Evaluations of the objective function and convergence to solution.

Apart from robustness and efficiency, another important characteristic of any algorithm is the quality of its outcome, i.e. the percentage of the cases when the algorithm converges to a true solution rather than to some nonoptimal stationary point. Figure 4(b) reports on this aspect of behavior of the algorithms. Here, the 'result' of each algorithm is the inverse number of convergences to the solution. Note that this result equals to $+\infty$ when the given algorithm never converged to the solution for a given problem, and this adds to the cases of failure. This is why the values on the right end are smaller than they are in the other figures. One can see that `Lifted ssSQP BFGS`, `Lifted ssSQP full BFGS` and `SQP slacks BFGS` behave comparably and they all have a stronger tendency of convergence to the solution than `Lifted SNM SR1`.

Figure 5 gives information about the ability of the algorithms to achieve smaller values of the objective function in the cases of successful runs (in these cases the last iterate is feasible or nearly feasible; therefore, it makes sense to look at the objective function values as another performance indicator). These diagrams were obtained as follows. For each algorithm and each problem, we compute the average objective function value achieved over successful runs. The given algorithm was classified into 'best' category when this value was the smallest among the four algorithms, and into 'worst' category when it was the largest. The average objective function value was regarded as equal to the smallest or to the largest when the difference was less than 10^{-3}. Note that for some particular problems an algorithm can fall into both 'best' and 'worst'
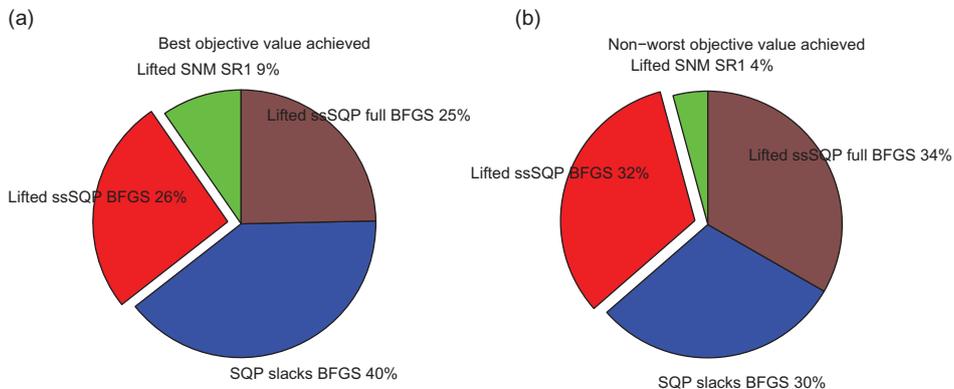

Figure 5. Objective value achieved.

categories, if all the four algorithms give the same average objective function value. Once we know how many times each algorithm falls into each category, we sum up these numbers for each category, and show the portion of this sum corresponding to each algorithm in Figure 5. Somewhat surprisingly, `Lifted ssSQP BFGS` and `Lifted ssSQP full BFGS` are comparable with `SQP slacks BFGS` by these characteristics. And as expected, they all seriously outperform `Lifted SNM SR1`.

For test problems `dempe`, `desilva`, `outrata31`, `outrata32`, `outrata33`, `outrata34`, `scholtes1` and `scholtes2`, the version `Lifted ssSQP BFGS` significantly outperforms `SQP slacks BFGS` in terms of robustness and in terms of the minor iterations count. For `dempe`, `desilva`, `scholtes1`, `scholtes2`, the same tendency is observed in terms of convergence to solution. The opposite tendency is observed for test problems `ex9.1.2`, `ex9.1.4`, `ex9.2.1`, `ex9.2.7`, `ex9.2.8`, `ex9.2.9`, `scholtes5`, for which `SQP slacks BFGS` significantly outperforms `Lifted ssSQP BFGS` in terms of robustness and in terms of convergence to solution.

We complete our numerical results by considering [32, Example 6.1]. Correcting some misprints, this problem is of the form (2) with $n = 3$, $m = 2$, $f(x) = 0.1x_1 + 0.1x_2 + 0.8x_3$, $G(x) = (x_1, -x_1 - x_2 - x_3 + 1)$, $H(x) = (x_2, -x_1 - x_2 + x_3 + 1)$. Then $\bar{x} = (0, 0, -1)$ is the unique global minimizer (and a strongly stationary point), while $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ are nonoptimal weakly stationary points.

When started from $x^0 = (0, 0, 1)$, all three lifted methods are not able to leave this nonoptimal weakly stationary point. Another 'difficult' feasible starting point suggested in [32] is $x^0 = (0.5, 0, 0.5)$. Starting from this point, `Lifted SNM SR1` converges to the nonoptimal weakly stationary point $(1, 0, 0)$, while `Lifted ssSQP BFGS` successfully converges to the solution $\bar{x}$. Finally, when started from random starting points generated as described above, `Lifted SNM SR1` converges to $\bar{x}$ for 45% of runs, while `Lifted ssSQP BFGS` converges to $\bar{x}$ for 98% of runs. Therefore, the latter algorithm does not experience any serious difficulties with getting stuck at nonoptimal weakly stationary points. `Lifted ssSQP full BFGS` behaves similarly to `Lifted ssSQP BFGS` on this problem.

Acknowledgements

References

[1] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt, *On Augmented Lagrangian methods with general lower-level constraints*, SIAM J. Optim. 18 (2007), pp. 1286–1309.

[2] B.T. Boggs and J.W. Tolle, *Sequential quadratic programming*, Acta Numer. 4 (1996), pp. 1–51.

[3] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects*, 2nd ed., Springer, Berlin, Germany, 2006.

[4] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

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

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

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

[8] 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.

[9] J. Han and D. Sun, *Superlinear convergence of approximate Newton methods for LC^1 optimization problems without strict complementarity*, in *Recent Advances in Nonsmooth Optimization*, D.-Z. Du, L. Qi, and R.S. Womersley, eds., vol. 58, World Scientific, Singapore, 1993, pp. 353–367.

[10] C.-M. Ip and J. Kyparisis, *Local convergence of quasi-Newton methods for B-differentiable equations*, *Math. Program.* 56 (1992), pp. 71—89.

[11] A.F. Izmailov and A.S. Kurennoy, *Partial Clarke generalized Jacobian and other generalized differentials*, *Theoretical and Applied Problems of Nonlinear Analysis*, Computing Center RAS, Moscow, 2010 (in Russian).

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

[13] A.F. Izmailov and M.V. Solodov, *The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions*, Math. Oper. Res. 27 (2002), pp. 614–635.

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

[15] A.F. Izmailov, A.L. Pogosyan, and M.V. Solodov, *Semismooth Newton method for the lifted reformulation of mathematical programs with complementarity constraints*, Comput. Optim. Appl. 2010. DOI 10.1007/s10589-010-9341-7.

[16] D. Klatte and K. Tammer, *On the second order sufficient conditions to perturbed $C^{1,1}$ optimization problems*, Optimization 19 (1988), pp. 169–180.

[17] B. Kummer, *Newton's method for nondifferentiable functions*, in *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, Akademie-Verlag, Berlin, 1988, pp. 114–125.

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

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

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

[21] J.M. Martínez and L. Qi, *Inexact Newton methods for solving nonsmooth equations*, J. Comput. Appl. Math. 60 (1995), pp. 127–145.

[22] J. Nocedal and S.J. Wright, *Numerical Optimization*, 2nd ed., Springer, New York, 2006.

[23] J.V. Outrata, M. Kocvara, and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results*, Kluwer Academic Publishers, Boston, 1998.

[24] L. Qi, *LC^1 functions and LC^1 optimization problems*, Technical Report AMR 91/21, School of Mathematics, The University of New South Wales, Sydney, Australia, 1991.

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

[26] L. Qi, *Superlinearly convergent approximate Newton methods for LC^1 optimization problems*, Math. Program. 64 (1994), pp. 277–294.

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

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

[29] R.T. Rockafellar, *Computational schemes for solving large-scale problems in extended linear-quadratic programming*, Math. Program. 48 (1990), pp. 447–474.

[30] R.T. Rockafellar and J.B. Wets, *Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time*, SIAM J. Control Optim. 28 (1990), pp. 810–922.

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

[32] O. Stein, *Lifting mathematical programs with complementarity constraints*, Math. Program. 2010. DOI 10.1007/s10107-010-0345-y.