A CLASS OF ACTIVE-SET NEWTON METHODS FOR MIXED COMPLEMENTARITY PROBLEMS*

A. N. Daryina[1], A. F. Izmailov[2] and M. V. Solodov[3]

October 3, 2003 (Revised April 5, 2004)

ABSTRACT

Based on the identification of indices active at a solution of the mixed complementarity problem (MCP), we propose a class of Newton methods for which local superlinear convergence holds under extremely mild assumptions. In particular, the error bound condition needed for the identification procedure and the nondegeneracy condition needed for the convergence of the resulting Newton method are individually and collectively strictly weaker than the property of semistability of a solution. Thus the local superlinear convergence conditions of the presented method are weaker than conditions required for the semismooth (generalized) Newton methods applied to MCP reformulations. Moreover, they are also weaker than convergence conditions of the linearization (Josephy–Newton) method. For the special case of optimality systems with primal-dual structure, we further consider the question of superlinear convergence of primal variables. We illustrate our theoretical results with numerical experiments on some specially constructed MCPs whose solutions do not satisfy the usual regularity assumptions.

Key words. Mixed complementarity problem, semistability, 2-regularity, weak regularity, error bound, Newton method.
AMS subject classifications. 90C30, 65K05.

---

* Research of the first two authors is supported by the Russian Foundation for Basic Research Grant 01-01-00810 and the Council for the State Support of Leading Scientific Schools Grant 00-15-96141. The third author is supported in part by CNPq Grants 300734/95-6(RN) and 471780/2003-0, and by FAPERJ.

[1] Russian Peoples Friendship University, Miklukho-Maklaya Str. 6, 117198 Moscow, Russia.
Email: anna.daryina@mail.ru

[2] Moscow State University, Faculty of Computational Mathematics and Cybernetics, Department of Operations Research, Leninskiye Gori, GSP-2, 119992 Moscow, Russia.
Email: izmaf@ccas.ru

[3] 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

The *mixed complementarity problem* (MCP) [8] is the variational inequality on a generalized box, that is

$$\text{find } x \in B \text{ such that } \langle F(x), y - x \rangle \geq 0 \text{ for all } y \in B, \tag{1.1}$$

where $F : \mathbf{R}^n \to \mathbf{R}^n$ and

$$B = \{x \in \mathbf{R}^n \mid l_i \leq x_i \leq u_i,\ i = 1, \ldots, n\},$$

$l_i \in \mathbf{R} \cup \{-\infty\}$, $u_i \in \mathbf{R} \cup \{+\infty\}$, $l_i < u_i$ for all $i = 1, \ldots, n$. Equivalently, it can be stated as

$$\text{find } x \in B \text{ such that } F_i(x) \begin{cases} \geq 0, & \text{if } x_i = l_i, \\ = 0, & \text{if } x_i \in (l_i, u_i), \quad i = 1, \ldots, n. \\ \leq 0, & \text{if } x_i = u_i, \end{cases}$$

As is well known, many important problems can be cast in the format of MCP [10, 8]. As a special case of MCP, we mention the nonlinear complementarity problem (NCP), which corresponds to setting $l_i = 0$, $u_i = +\infty$, $i = 1, \ldots, n$. The systems of nonlinear equations are obtained by choosing $l_i = -\infty$, $u_i = +\infty$, $i = 1, \ldots, n$. Another important example is the primal-dual Karush-Kuhn-Tucker (KKT) optimality system: find $z \in \mathbf{R}^p$ and $\mu \in \mathbf{R}^m$ such that

$$\begin{array}{c} g(z) - (G'(z))^T \mu = 0, \\ \mu \geq 0, \quad G(z) \geq 0, \quad \langle \mu, G(z) \rangle = 0, \end{array} \tag{1.2}$$

where $g : \mathbf{R}^p \to \mathbf{R}^p$ and $G : \mathbf{R}^p \to \mathbf{R}^m$. The KKT system (1.2) can be written as an MCP if we set $n = p + m$ and

$$F(x) = \begin{pmatrix} g(z) - (G'(z))^T \mu \\ G(z) \end{pmatrix}, \quad x = (z, \mu) \in \mathbf{R}^p \times \mathbf{R}^m,$$

$l_i = -\infty$, $i = 1, \ldots, p$, $l_i = 0$, $i = p + 1, \ldots, n$, $u_i = +\infty$, $i = 1, \ldots, n$. Under well-known assumptions, (1.2) represents the first-order primal-dual necessary conditions characterizing solutions in variational inequality or constrained optimization problems. We note that the inclusion of pure equality constraints in the KKT system does not add anything conceptually important in the setting of this paper. For this reason, when talking about the KKT systems we shall only consider the format of (1.2).

This paper follows the development of Newton methods based on the identification of active constraints for KKT systems, presented in [15]. Apart from extending the ideas to MCP, this paper contains a number of improvements and refinements, as will be pointed out in the sequel. In particular, the MCP regularity condition introduced here, even when reduced to the special case of KKT systems, is strictly weaker than the one in [15]. Also some numerical experiments will be reported to illustrate the local behavior of the proposed method under weak assumptions.

As another somewhat related recent work on active-set methods, we mention [4]. In that reference, the special case of NCP is considered. We note that our assumptions for local superlinear convergence are neither stronger nor weaker than those for the method of [4]. On the one hand, [4] can deal with non-isolated solutions, while our assumptions do imply that the given solution is locally unique. On the other hand, monotonicity of F is essential in all of the constructions in [4], while no assumptions of this type are being made in this paper. Also, it seems that the error bound assumed in [4] cannot be directly compared with more standard type of bounds, such as ours. Finally, we consider a more general class of Newton methods, not restricted to the Gauss-Newton method for the (over-determined) system of nonlinear equations obtained via some identification procedure.

We start in Section 2 with deriving a new error bound for MCP (an upper estimate for the distance from a given point to a solution of MCP), based on a smooth reformulation of MCP and a 2-regularity condition [13, 14]. It has been shown previously that in the context of NCP [13] and KKT [15], this construction leads to error bounds which hold under weaker conditions than the alternatives for the corresponding problems (such as b-regularity, semistability, R_0-property, quasi-regularity, etc.). Here, we extend the analysis and the comparisons to MCP. In addition, we further prove that 2-regularity is not only a sufficient, but also a necessary condition for the associated error bound to hold.

Error bounds have many applications [24], among which is identifying active constraints in constrained optimization [7] (see also [8, Ch. 6.7]). In the context of MCP, those ideas correspond to identifying the sets of indices

$$\begin{gathered} A = A(\bar{x}) = \{i = 1, \dots, n \mid F_i(\bar{x}) = 0\}, \\ N = N(\bar{x}) = \{i = 1, \dots, n \mid F_i(\bar{x}) \neq 0\}, \\ N_l = N_l(\bar{x}) = \{i \in N \mid \bar{x}_i = l_i\}, \\ N_u = N_u(\bar{x}) = \{i \in N \mid \bar{x}_i = u_i\}, \end{gathered}$$

where $\bar{x}$ is some solution of MCP. If the specified sets can be correctly identified using information available at a point x close enough to the solution $\bar{x}$, then locally MCP can be reduced to a system of nonlinear equations (which is structurally a much simpler problem to solve). In the sequel, we shall also use the following partitioning of the set of active indices:

$$\begin{gathered} A_0 = A_0(\bar{x}) = \{i \in A \mid \bar{x}_i = l_i \text{ or } \bar{x}_i = u_i\}, \\ A_+ = A_+(\bar{x}) = \{i \in A \mid \bar{x}_i \in (l_i, u_i)\}, \\ A_{0l} = A_{0l}(\bar{x}) = \{i \in A_0 \mid \bar{x}_i = l_i\}, \\ A_{0u} = A_{0u}(\bar{x}) = \{i \in A_0 \mid \bar{x}_i = u_i\}. \end{gathered}$$

The analog of the strict complementarity condition in NCP (or KKT) corresponds, in the setting of MCP, to saying that $A_0 = \emptyset$. Under this assumption, locally MCP trivially reduces to a system of nonlinear equations, which simplifies the local structure of MCP significantly. The condition of strict complementarity, however, is restrictive. We emphasize that this condition is not assumed anywhere in this paper.

In Section 3, we propose a new class of active-set Newton methods for solving MCP. Each iteration of the method consists of solving one system of linear equations. We note that when specified to the setting of KKT, this class is different from what has been discussed in [15]. Moreover, the nondegeneracy condition that we introduce here is weaker than the corresponding condition in [15]. Also, the new condition permits specific deterministic choice of parameters involved in reducing the MCP to a system of equations, while in [15] in general a generic choice of parameters had to be made (at least without strengthening somewhat the regularity assumptions). We show that the conditions needed for the identification of active sets and for convergence of the proposed local Newton method are weaker than semistability of the MCP solution [3, 8] (equivalently, the R_0-property of the natural residual). This implies, in particular, that the proposed method attains local superlinear/quadratic convergence under assumptions considerably weaker than what is needed for semismooth Newton methods (SNM) for MCP [1, 19, 9, 16, 18] (BD-regularity of the reformulation being used). Even more remarkably, our assumptions are also strictly weaker than those needed for the linearization (Josephy-Newton) method [17, 3, 12] (which are semistability *and* hemistability of the solution). It should be also noted that in the latter methods subproblems are linearized MCPs, which are computationally more complex than systems of linear equations in our methods.

In Section 4, we turn our attention to the specific case of KKT, and in particular consider the issue of superlinear convergence of primal variables. Some comments on the comparison of convergence conditions for various Newton-type methods for MCP constitute Section 5. Numerical experiments are presented and discussed in Section 6.

A few words about our notation. Given a finite set I, $|I|$ stands for its cardinality. By $\mathbf{R}(m,\, n)$ we denote the space of $m \times n$ matrices with real entries. By E we shall denote the identity matrix whose dimension would be always clear from the context. For $x \in \mathbf{R}^n$ and an index set $I \subset \{1,\, \ldots,\, n\}$, x_I stands for the vector with components x_i, $i \in I$. For a linear operator Λ, $\operatorname{im}\Lambda$ is its range (image space), and $\ker\Lambda$ is its kernel (null space). For a directionally differentiable mapping $\phi : \mathbf{R}^n \to \mathbf{R}^m$, by $\phi'(x;\, d)$ we denote the usual directional derivative of ϕ at $x \in \mathbf{R}^n$ in the direction $d \in \mathbf{R}^n$. If $\{z^k\}$ is a sequence in $\mathbf{R}^p$ and $\{t_k\}$ is a sequence in $\mathbf{R}$ such that $t_k \to 0+$ as $k \to \infty$, by $z^k = o(t_k)$ we mean that $\lim_{k\to\infty} \|z^k\|/t_k = 0$.

2 A New Error Bound for MCP

In this section, we are interested in estimating the distance to a solution of MCP in terms of some *computable* quantity. As is well-known [1, 9], MCP can be equivalently reformulated as a system of nonlinear equations via the following transformation.

Let $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ be a *complementarity function*, i.e., a function such that

$$\psi(a, b) = 0 \quad \Leftrightarrow \quad a \geq 0,\ b \geq 0,\ ab = 0.$$

Assuming that ψ also satisfies the following additional assumptions:

$$a > 0,\ b < 0 \quad \Rightarrow \quad \psi(a,\, b) < 0,$$

$$a > 0,\ b > 0 \quad \Rightarrow \quad \psi(a,\, b) > 0,$$

solutions of MCP coincide with solutions of the following system of nonlinear equations:

$$\Psi(x) = 0, \tag{2.1}$$

where

$$\Psi : \mathbf{R}^n \to \mathbf{R}^n, \quad \Psi_i(x) = \begin{cases} F_i(x), & \text{if } i \in I_F, \\ \psi(x_i - l_i, F_i(x)), & \text{if } i \in I_l, \\ -\psi(u_i - x_i, -F_i(x)), & \text{if } i \in I_u, \\ \psi(x_i - l_i, -\psi(u_i - x_i, -F_i(x))), & \text{if } i \in I_{lu}, \end{cases}$$

$$\begin{aligned} I_F &= \{i = 1, \ldots, n \mid -\infty = l_i,\ u_i = +\infty\}, \\ I_l &= \{i = 1, \ldots, n \mid -\infty < l_i,\ u_i = +\infty\}, \\ I_u &= \{i = 1, \ldots, n \mid -\infty = l_i,\ u_i < +\infty\}, \\ I_{lu} &= \{i = 1, \ldots, n \mid -\infty < l_i,\ u_i < +\infty\}. \end{aligned}$$

Complementarity functions to be mentioned in the sequel are the natural residual $\psi_{NR}(a, b) = \min\{a, b\}$, the Fischer-Burmeister function $\psi_{FB}(a, b) = a + b - \sqrt{a^2 + b^2}$, and $\psi_S(a, b) = 2ab - (\min\{0, a + b\})^2$ (where S stands for "smooth"). All these functions satisfy the assumptions above. The corresponding reformulations of MCP would be denoted by Ψ_{NR}, Ψ_{FB} and Ψ_S, respectively. For the purposes of this paper, the signs of the components of Ψ could be chosen differently from the above (see, however, [2, 9] for the justification of the choice adopted here).

Let $\bar{x}$ be a solution of MCP. It is known [8, Ch. 6.2] that the natural residual error bound

$$\|x - \bar{x}\| \leq M \|\Psi_{NR}(x)\| \quad \forall\, x \in U \tag{2.2}$$

holds for some neighborhood U of $\bar{x}$ and some constant $M > 0$ if, and only if, $\bar{x}$ is a *semistable* [3, 8] solution of MCP. It can be seen that this is also equivalent to an error bound in terms of Ψ_{FB}. Furthermore, for MCP semistability is equivalent to the R_0-property of Ψ_{NR} at $\bar{x}$, which is $\{\xi \in \mathbf{R}^n \mid \Psi'_{NR}(\bar{x}; \xi) = 0\} = \{0\}$, and the latter is also equivalent to the corresponding property for Ψ_{FB}. As can be easily checked, the R_0-property means that

$$L = \{0\},$$

where L denotes the solution set of the "linearized" MCP:

$$L = L(\bar{x}) = \left\{ \xi \in \mathbf{R}^n \;\middle|\; \begin{array}{c} \xi_i \geq 0,\ \langle F'_i(\bar{x}),\, \xi\rangle \geq 0,\ \xi_i \langle F'_i(\bar{x}),\, \xi\rangle = 0,\ i \in A_{0l}, \\ \xi_i \leq 0,\ \langle F'_i(\bar{x}),\, \xi\rangle \leq 0,\ \xi_i \langle F'_i(\bar{x}),\, \xi\rangle = 0,\ i \in A_{0u}, \\ F'_{A_+}(\bar{x})\xi = 0,\ \xi_N = 0 \end{array} \right\}. \tag{2.3}$$

Semistability is one of the weakest conditions under which a computable error bound for MCP had been exhibited up to now. To our knowledge, alternative conditions are either stronger or different in nature and not comparable to semistability (e.g., analyticity [23] or subanalyticity of F). In what follows, we provide an error bound under a condition which we show to be strictly weaker than semistability.

Definition 2.1 Let $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighbourhood of $\bar{x} \in \mathbf{R}^n$ and $\Psi' : \mathbf{R}^n \to \mathbf{R}(n,\, n)$ be directionally differentiable at $\bar{x}$. Then Ψ is 2-*regular* at $\bar{x}$ if

$$T = \{0\},$$

where

$$T = T(\bar{x}) = \{\xi \in \ker \Psi'(\bar{x}) \mid (\Psi')'(\bar{x};\, \xi)\xi \in \operatorname{im} \Psi'(\bar{x})\} \tag{2.4}$$
$$= \{\xi \in \ker \Psi'(\bar{x}) \mid P(\Psi')'(\bar{x};\, \xi)\xi = 0\},$$

with P being the orthogonal projector onto $(\operatorname{im} \Psi'(\bar{x}))^{\perp}$.

The above is a special case of 2-regularity of a nonlinear mapping [14, 13], corresponding to the case when the mapping acts from some space into itself.

We next give an error bound result based on the smooth MCP reformulation Ψ_S. A comparison with semistability will be made later. We note that the fact that the error bound below is actually equivalent to 2-regularity is new.

Theorem 2.1 *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be sufficiently smooth near a point $\bar{x} \in \mathbf{R}^n$, which is a solution of MCP.*

The mapping Ψ_S is 2-regular at $\bar{x}$ if, and only if, there exist a neighborhood U of $\bar{x}$ and a constant $M > 0$ such that

$$\|x - \bar{x}\| \le M(\|(E - P)\Psi_S(x)\| + \|P\Psi_S(x)\|^{1/2}) \quad \forall\, x \in U. \tag{2.5}$$

Proof. Obviously, Ψ_S satisfies the smoothness assumptions in the Definition 2.1. The sufficiency part of the assertion is a direct consequence of [13, Theorem 4]. We prove the necessity part.

Take any $\xi \in T$. For any $t \ge 0$, we have that

$$\|\Psi_S(\bar{x} + t\xi)\| = \|\Psi_S(\bar{x} + t\xi) - \Psi_S(\bar{x})\| = \|t\Psi_S'(\bar{x})\xi\| + o(t) = o(t),$$

where the last equality follows from $\xi \in \ker \Psi_S'(\bar{x})$. We further have that

$$\begin{aligned}
\|P\Psi_S(\bar{x} + t\xi)\| &= \|P(\Psi_S(\bar{x} + t\xi) - \Psi_S(\bar{x}))\| \\
&\le t \sup_{\tau \in [0,1]} \|P\Psi_S'(\bar{x} + \tau t\xi)\| \|\xi\| \\
&= t \sup_{\tau \in [0,1]} \|P(\Psi_S'(\bar{x} + \tau t\xi) - \Psi_S'(\bar{x}))\| \|\xi\| \\
&= t \sup_{\tau \in [0,1]} \|\tau t P(\Psi_S')'(\bar{x}; \xi)\xi\| \|\xi\| + o(t^2) \\
&= o(t^2),
\end{aligned}$$

where we have used the Mean-Value Theorem, the fact that $P\Psi_S'(\bar{x}) = 0$, the positive homogenuity of the mapping $P(\Psi_S')'(\bar{x}; \cdot)$ and the fact that $P(\Psi_S')'(\bar{x}; \xi)\xi = 0$.

Therefore, (2.5) implies that

$$t\|\xi\| = \|\bar{x} + t\xi - \bar{x}\| \le M(\|(E - P)\Psi_S(\bar{x} + t\xi)\| + \|P\Psi_S(\bar{x} + t\xi)\|^{1/2}) = o(t),$$

which means that $\xi = 0$. We have thus established that $T = \{0\}$, i.e., Ψ_S is 2-regular at $\bar{x}$. ∎

Adjusting M and U, if necessary, the error bound (2.5) can be simplified into the following relation (less accurate, but possibly easier to use):

$$\|x - \bar{x}\| \leq M\|\Psi_S(x)\|^{1/2} \quad \forall\, x \in U. \tag{2.6}$$

Note that for NCP or KKT the error bound (2.6) is implied by error bound (2.2) (this follows from a comparison of growth rates for ψ_{NR}, ψ_{FB} and ψ_S, given in [28]). However, for MCP it is not clear whether one can use the same comparison, as the definition of Ψ involves a superposition of the functions ψ. In any case, the more accurate estimate (2.5) does not follow from (2.2). Moreover, (2.5) and (2.6) can hold when (2.2) does not, as shown next.

Proposition 2.1 *Semistability of a solution $\bar{x}$ of MCP (equivalently, error bound* (2.2)*) implies 2-regularity of Ψ_S at $\bar{x}$ (equivalently, error bound* (2.5)*), but not vice versa.*

Proof. Let $\bar{x}$ be a solution of MCP. The fact that 2-regularity of Ψ_S at $\bar{x}$ can hold when the error bound (2.2) (semistability of $\bar{x}$) does not, has been already shown for two special cases of MCP: namely, NCP [13, Example 1] and KKT [15, Example 2]. Thus no further justification is needed for this assertion.

Let $\bar{x}$ be a semistable solution. To prove that this implies 2-regularity of Ψ_S at $\bar{x}$, it suffices to show that $T \subset L$, where T and L are defined in (2.4) and (2.3), respectively.

By direct computations, we have that

$$(\Psi'_S)_i(\bar{x}) = \begin{cases} 0, & \text{if } i \in A_0, \\ \alpha_i F'_i(\bar{x}), & \text{if } i \in A_+, \\ \beta_i e^i, & \text{if } i \in N, \end{cases}$$

where $e^1, \ldots, e^n$ is the canonical basis in $\mathbf{R}^n$, and

$$\alpha_i = \begin{cases} 2(\bar{x}_i - l_i), & \text{if } i \in I_l, \\ 2(u_i - \bar{x}_i), & \text{if } i \in I_u, \\ 1, & \text{if } i \in I_F, \\ 4(\bar{x}_i - l_i)(u_i - \bar{x}_i), & \text{if } i \in I_{lu}, \end{cases}$$

$$\beta_i = \begin{cases} 2F_i(\bar{x}), & \text{if } i \in I_l, \\ -2F_i(\bar{x}), & \text{if } i \in I_u, \\ 4F_i(\bar{x})(u_i - l_i) + 2(\min\{0, u_i - l_i - F_i(\bar{x})\})^2, & \text{if } i \in I_{lu} \cap N_l, \\ -4F_i(\bar{x})(u_i - l_i), & \text{if } i \in I_{lu} \cap N_u. \end{cases}$$

Observe that $\alpha_i \neq 0\ \forall\, i \in A_+$ and $\beta_i \neq 0\ \forall\, i \in N$. Hence,

$$\ker \Psi'_S(\bar{x}) = \{\xi \in \mathbf{R}^n \mid F'_{A_+}(\bar{x})\xi = 0,\ \xi_N = 0\}.$$

Then P, the orthogonal projector onto $(\mathrm{im}\, \Psi'_S(\bar{x}))^\perp$ in $\mathbf{R}^n$, satisfies

$$(Py)_i = y_i \quad \forall\, y \in \mathbf{R}^n,\ \forall\, i \in A_0.$$

Since $P\Psi_S' = (P\Psi_S)'$, it now easily follows that

$$T \subset \{\xi \in \mathbf{R}^n \mid (\Psi_S')_i'(\bar{x};\xi)\xi = 0,\ i \in A_0,\ F_{A_+}'(\bar{x})\xi = 0,\ \xi_N = 0\}. \tag{2.7}$$

For any $i \in A_0$ and $\xi \in \mathbf{R}^n$, we further obtain that

$$(\Psi_S')_i'(\bar{x};\xi)\xi = 2\begin{cases} \psi_S(\xi_i, \langle F_i'(\bar{x}), \xi\rangle), & \text{if } i \in I_l, \\ -\psi_S(-\xi_i, -\langle F_i'(\bar{x}), \xi\rangle), & \text{if } i \in I_u, \\ \psi_S(\xi_i, (u_i - l_i)\langle F_i'(\bar{x}), \xi\rangle), & \text{if } i \in I_{lu} \cup A_{0l}, \\ -(u_i - l_i)\psi_S(-\xi_i, -\langle F_i'(\bar{x}), \xi\rangle), & \text{if } i \in I_{lu} \cup A_{0u}. \end{cases}$$

Since ψ_S is a complementarity function, the right-hand side of the latter equality is zero if, and only if,

$$\begin{array}{llll} \xi_i \geq 0, & \langle F_i'(\bar{x}), \xi\rangle \geq 0, & \xi_i\langle F_i'(\bar{x}), \xi\rangle = 0, & i \in A_{0l}, \\ \xi_i \leq 0, & \langle F_i'(\bar{x}), \xi\rangle \leq 0, & \xi_i\langle F_i'(\bar{x}), \xi\rangle = 0, & i \in A_{0u}. \end{array}$$

Hence, the right-hand side in (2.7) coincides with L defined in (2.3). In particular, we have established that $T \subset L$, which completes the proof. ∎

3 A Class of Active-Set Newton Methods

The following technique for identifying the relevant index sets is based on the ideas of [7], see also [8, Ch. 6.7]. Define the *identification function*

$$\rho : \mathbf{R}_+ \to \mathbf{R}, \quad \rho(t) = \begin{cases} \bar{\rho}, & \text{if } t \geq \bar{t}, \\ -1/\log t, & \text{if } t \in (0,\, \bar{t}), \\ 0, & \text{if } t = 0, \end{cases}$$

where $\bar{t} \in (0,\, 1)$ and $\bar{\rho} > 0$ are fixed numbers (the choice of $\bar{t}$ and $\bar{\rho}$ does not affect theoretical analysis; in our numerical experiments reported in Section 6, we use $\bar{t} = 0.9$ and $\bar{\rho} = -1/\log \bar{t}$, as suggested in [7]). For any $x \in \mathbf{R}^n$, define further the index sets

$$A(x) = \{i = 1, \ldots, n \mid |F_i(x)| \leq \rho(\|\Psi_S(x)\|)\}, \tag{3.1}$$

$$N(x) = \{1, \ldots, n\} \setminus A(x), \tag{3.2}$$

$$N_l(x) = \{i \in N(x) \mid x_i - l_i \leq u_i - x_i\}, \quad N_u(x) = N(x) \setminus N_l(x), \tag{3.3}$$

$$A_0(x) = \{i \in A(x) \mid \min\{|x_i - l_i|, |u_i - x_i|\} \leq \rho(\|\Psi_S(x)\|)\}, \quad A_+(x) = A(x) \setminus A_0(x), \tag{3.4}$$

$$A_{0l}(x) = \{i \in A_0(x) \mid x_i - l_i \leq u_i - x_i\}, \quad A_{0u}(x) = A_0(x) \setminus A_{0l}(x). \tag{3.5}$$

Proposition 3.1 *If Ψ_S is 2-regular at a solution $\bar{x}$ of MCP (equivalently, the error bound* (2.5) *holds), then for any $x \in \mathbf{R}^n$ sufficiently close to $\bar{x}$, it holds that*

$$A(x) = A, \quad N(x) = N, \quad N_l(x) = N_l, \quad N_u(x) = N_u, \tag{3.6}$$

$$A_{0l}(x) = A_{0l}, \quad A_{0u}(x) = A_{0u}, \quad A_0(x) = A_0, \quad A_+(x) = A_+. \tag{3.7}$$

Proof. Take any $i \in A$. Then for any x close enough to $\bar{x}$, we have that

$$|F_i(x)| = |F_i(x) - F_i(\bar{x})| \le K\|x - \bar{x}\| \le KM\|\Psi_S(x)\|^{1/2} \le \rho(\|\Psi_S(x)\|),$$

where the second inequality is by the local Lipschitz-continuity of F (with some modulus $K > 0$), the third inequality is by (2.5), and the last follows from the fact that $\lim_{t\to 0+} t^\nu \log t = 0$ for any $\nu > 0$. The above shows that $i \in A(x)$. Hence, $A \subset A(x)$.

Take any $i \in \{1, \ldots, n\} \setminus A$. In that case, there exists some $\gamma > 0$ such that for any x close enough to $\bar{x}$ it holds that

$$|F_i(x)| \ge \gamma, \quad \rho(\|\Psi_S(x)\|) < \gamma.$$

It follows that $i \notin A(x)$, which shows that $A(x) \subset A$.

We have therefore established the first (and hence, the second) equality in (3.6).

The other relations either hold trivially (e.g., (3.3)) or can be verified by considerations similar to the above. ∎

We note that any other MCP reformulation Ψ with a corresponding valid error bound can be used in the identification procedure. But since it has been established above that Ψ_S requires the weakest assumptions for the error bound to hold, it is fair to say that this is the function which should be used for this purpose. However, different choices of the function ρ are possible under the same assumptions. For example,

$$\rho : \mathbf{R}_+ \to \mathbf{R}, \quad \rho(t) = t^\theta, \quad \theta \in (0, 1/2).$$

Observe also that in the implementation of the identification procedure, the following obvious relations can be taken into account: $I_F \subset A_+$, $I_l \subset (A_{0l} \cup A_+ \cup N_l)$ and $I_u \subset (A_{0u} \cup A_+ \cup N_u)$.

Once the index sets are identified, we have the following relations which are guaranteed to be satisfied at a solution $\bar{x}$ of MCP:

$$F_A(x) = 0, \quad x_{A_{0l}\cup N_l} = l_{A_{0l}\cup N_l}, \quad x_{A_{0u}\cup N_u} = u_{A_{0u}\cup N_u}.$$

For simplicity of notation, suppose that the components of $x \in \mathbf{R}^n$ are ordered in such a way that $x = (x_{A_+}, x_{A_{0l}\cup N_l}, x_{A_{0u}\cup N_u})$. Then MCP locally reduces to the following system of nonlinear equations:

$$F_A(x_{A_+}, l_{A_{0l}\cup N_l}, u_{A_{0u}\cup N_u}) = 0. \tag{3.8}$$

Observe that in the absence of strict complementarity (when $A_0 \ne \emptyset$, i.e., $|A| > |A_+|$), the system is over-determined (the number of equations is larger than the number of unknowns). This opens up a number of options. Of course, one can just solve the system by the Gauss–Newton method (GNM). This possibility will be considered. However, we prefer not to limit ourselves to GNM for the following reason: the Gauss–Newton approach can destroy structure present in F_A (for example, sparsity or the primal-dual structure in the case of KKT).

Our proposal is to consider the following system of nonlinear equations:

$$\Phi_C(x_{A_+}) = 0, \tag{3.9}$$

where

$$\Phi_C : \mathbf{R}^{|A_+|} \to \mathbf{R}^{|A_+|}, \quad \Phi_C(x_{A_+}) = C(x_{A_+})F_A(x_{A_+}, l_{A_{0l}\cup N_l}, u_{A_{0u}\cup N_u}),$$

with $C : \mathbf{R}^{|A_+|} \to \mathbf{R}(|A_+|, A)$ being a smooth mapping (possibly constant). Clearly, $\bar{x}_{A_+}$ is a solution of (3.9) for any choice of C. The Jacobian of (3.9) at this solution is given by

$$\Phi'_C(\bar{x}_{A_+}) = C(\bar{x}_{A_+})\frac{\partial F_A}{\partial x_{A_+}}(\bar{x}), \tag{3.10}$$

where we have taken into account that $F_A(\bar{x}) = 0$. Thus $\bar{x}_{A_+}$ can be found by applying Newton-type methods to (3.9) whenever the matrix in (3.10) is nonsingular.

Note that GNM for (3.8) would essentially correspond to choosing in (3.9)

$$C(x_{A_+}) = \left(\frac{\partial F_A}{\partial x_{A_+}}(x_{A_+}, l_{A_{0l}\cup N_l}, u_{A_{0u}\cup N_u})\right)^T, \tag{3.11}$$

and applying to the resulting system an approximate version of the pure Newton method. Indeed, with the notation of (3.11), the Gauss–Newton iteration for (3.8) has the form

$$x_{A_+}^{k+1} = x_{A_+}^k - \left(C(x_{A_+}^k)\frac{\partial F_A}{\partial x_{A_+}}(x_{A_+}^k, l_{A_{0l}\cup N_l}, u_{A_{0u}\cup N_u})\right)^{-1} \Phi_C(x_{A_+}^k). \tag{3.12}$$

Observe that the above formula is just an approximation of the standard Newton iteration for (3.9), where the Jacobian $\Phi'_C(x_{A_+}^k)$ is replaced by $C(x_{A_+}^k)\frac{\partial F_A}{\partial x_{A_+}}(x_{A_+}^k, l_{A_{0l}\cup N_l}, u_{A_{0u}\cup N_u})$. Due to (3.10), this change preserves the superlinear convergence of the pure Newton iteration for (3.9). Note finally that with the choice of (3.11), we have

$$\Phi'_C(\bar{x}_{A_+}) = \left(\frac{\partial F_A}{\partial x_{A_+}}(\bar{x})\right)^T \frac{\partial F_A}{\partial x_{A_+}}(\bar{x}). \tag{3.13}$$

This immediately motivates the following definition.

Definition 3.1 A solution $\bar{x}$ of MCP is referred to as *weakly regular* if

$$\operatorname{rank} \frac{\partial F_A}{\partial x_{A_+}}(\bar{x}) = |A_+|.$$

Clearly, weak regularity is a necessary and sufficient condition for the matrix in (3.13) to be nonsingular, and hence, for the superlinear convergence of GNM applied to (3.8) (or the approximate Newton method applied to (3.9) with the choice of (3.11)). It is also clear that weak regularity is necessary for the nonsingularity of the matrix in (3.10) corresponding to the more general scheme, and this is regardless of the choice of C.

We next show that weak regularity is implied by semistability, but not vice versa. Moreover, 2-regularity of Ψ_S at $\bar{x}$ and weak regularity, when combined, are still a weaker condition than semistability.

Proposition 3.2 *Let $\bar{x}$ be a solution of MCP. Then semistability of $\bar{x}$ implies weak regularity of $\bar{x}$, but not vice versa.*

Proof. Suppose that $\bar{x}$ is a semistable solution. If it is not weakly regular, then there exists $\xi_{A_+} \in \ker \frac{\partial F_A}{\partial x_{A_+}}(\bar{x}) \setminus \{0\}$. But then setting $\xi_{A_0 \cup N} = 0$, we obtain $\xi \neq 0$ such that $\xi \in L$, where L is defined in (2.3). This contradicts semistability.

The lack of the reverse implication is established in Example 3.1 below. ∎

The following result is of special importance.

Proposition 3.3 *Let $\bar{x}$ be a solution of MCP. Then semistability of $\bar{x}$ implies the combination of 2-regularity of Ψ_S at $\bar{x}$ and weak regularity of $\bar{x}$, but not vice versa.*

Proof. The forward assertion is by Propositions 2.1 and 3.2. The lack of the reverse implication is shown in Example 3.1. ∎

Example 3.1 Let $n = 2$, $l_i = 0$, $u_i = +\infty$, $i = 1, 2$, and let $F(x) = ((x_1 - 1)^2, x_1 + x_2 - 1)$.

The point $\bar{x} = (1, 0) \in \mathbf{R}^2$ is the solution of this NCP, and we have $A = \{1, 2\}$, $A_0 = A_{0l} = \{2\}$, $A_+ = \{1\}$, with all the other index sets being empty.

We first verify that semistability is violated. Noting that $F'_{A_+}(\bar{x}) = 0$, it can be seen that the cone L defined in (2.3) is

$$L = \{\xi \in \mathbf{R}^2 \mid \xi_2 \geq 0,\ \xi_1 + \xi_2 \geq 0,\ \xi_2(\xi_1 + \xi_2) = 0\} \neq \{0\}.$$

Thus $\bar{x}$ is not semistable.

Weak regularity certainly holds, as

$$\frac{\partial F_A}{\partial x_{A_+}}(\bar{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \operatorname{rank} \frac{\partial F_A}{\partial x_{A_+}}(\bar{x}) = 1 = |A_+|.$$

We proceed to show 2-regularity of Ψ_S at $\bar{x}$. It can be seen that $\Psi'_S(\bar{x}) = 0$. Hence,

$$\ker \Psi'_S(\bar{x}) = \mathbf{R}^2 = (\operatorname{im} \Psi'_S(\bar{x}))^\perp, \quad P = E,$$

and further

$$T = \{\xi \in \mathbf{R}^2 \mid (\Psi'_S)'(\bar{x};\, \xi)\xi = 0\}.$$

We obtain that

$$(\Psi'_S)'(\bar{x};\, \xi)\xi = \begin{pmatrix} 4\xi_1^2 \\ 2\xi_2(\xi_1 + 2\xi_2) - \min\{0,\, \xi_2 + (\xi_1 + \xi_2)\} \end{pmatrix}.$$

Hence,

$$T = \{\xi \in L \mid \xi_1 = 0\} = \{0\},$$

and Ψ_S is 2-regular at $\bar{x}$.

We have thus constructed a local algorithm with superlinear convergence under assumptions weaker than semistability of the MCP solution. Specifically, we have the following.

Theorem 3.1 *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be sufficiently smooth near a point $\bar{x} \in \mathbf{R}^n$, which is a solution of MCP. Suppose that this solution is weakly regular and Ψ_S is 2-regular at $\bar{x}$.*

For any $x^0 \in \mathbf{R}^n$ sufficiently close to $\bar{x}$, if the index sets $A = A(x^0)$, $A_+ = A_+(x^0)$, $A_{0l} = A_{0l}(x^0)$, $A_{0u} = A_{0u}(x^0)$, $N_l = N_l(x^0)$ and $N_u = N_u(x^0)$ are defined according to (3.1)-(3.5)*, then GNM applied to the system* (3.8) *(with $x^0_{A_+}$ as a starting point) is well-defined and superlinearly convergent to $\bar{x}_{A_+}$.*

Proof. By 2-regularity of Ψ_S at $\bar{x}$ and Proposition 3.1, for any $x = x^0$ close to $\bar{x}$ the index sets defined according to (3.1)-(3.5) correctly identify the index sets at the solution $\bar{x}$. Then $\bar{x}_{A_+}$ is the solution of (3.8). By weak regularity, we have that the matrix in (3.13) is nonsingular. Hence, GNM applied to (3.8) is locally superlinearly convergent to $\bar{x}_{A_+}$. ■

As already mentioned above (see also Section 4), it sometimes can be useful to choose the mapping C differently from the Gauss–Newton option of (3.11). For example, we might want to take $C(\cdot) = C \in \mathbf{R}(|A_+|, |A|)$, a fixed matrix, in order to preserve in the matrix $C\frac{\partial F_A}{\partial x_{A_+}}(x_{A_+}, l_{A_{0l}\cup N_l}, u_{A_{0u}\cup N_u})$ the structure (primal-dual, sparsity, etc.) of the matrix $\frac{\partial F_A}{\partial x_{A_+}}(x_{A_+}, l_{A_{0l}\cup N_l}, u_{A_{0u}\cup N_u})$. This motivates the following considerations.

Proposition 3.4 *Suppose that a solution $\bar{x}$ of MCP is weakly regular.*

Then the set of matrices $C \in \mathbf{R}(|A_+|, |A|)$ such that $\Phi'_C(\bar{x}_{A_+})$ is nonsingular is open and dense in $\mathbf{R}(|A_+|, |A|)$.

Proof. The determinant $\det \Phi'_C(\bar{x}_{A_+})$ is a polynomial with respect to the elements of the matrix $C \in \mathbf{R}(|A_+|, |A|)$. By weak regularity (see Definition 3.1), this polynomial is not everywhere zero, because it is not zero for the choice $C = (\frac{\partial F_A}{\partial x_{A_+}}(\bar{x}))^T$. Hence, the set where this polynomial is not zero is obviously open and dense in $\mathbf{R}(|A_+|, |A|)$. ■

Proposition 3.4 justifies choosing C in any desirable way, as the chance that the resulting system would be degenerate is negligible (the set of matrices for which this would happen is of the Lebesgue measure zero). Of course, one should make reasonable choices. For example, it should hold that $\operatorname{rank} C = |A_+|$.

4 Karush-Kuhn-Tucker Systems

In the case of the KKT system (1.2), the developments of Section 3 give

$$F_A(x_{A_+}, l_{A_{0l}\cup N_l}, u_{A_{0u}\cup N_u}) = \begin{pmatrix} g(z) - (G'_{I_+}(z))^T \mu_{I_+} \\ G_I(z) \end{pmatrix}, \quad x_{A_+} = (z, \mu_{I_+}) \in \mathbf{R}^p \times \mathbf{R}^{|I_+|},$$

where

$$I = I(\bar{z}) = \{i = 1, \ldots, m \mid G_i(\bar{z}) = 0\}, \quad I_+ = I_+(\bar{z}) = \{i \in I \mid \bar{\mu}_i > 0\},$$
$$A = \{1, \ldots, p\} \cup \{p + j \mid j \in I\}, \quad A_+ = \{1, \ldots, p\} \cup \{p + j \mid j \in I_+\}.$$

Defining

$$\phi : \mathbf{R}^p \times \mathbf{R}^{|I_+|} \to \mathbf{R}^p, \quad \phi(z, \mu_{I_+}) = g(z) - (G'_{I_+}(z))^T \mu_{I_+},$$

the Definition 3.1 of weak regularity consists of saying that

$$\text{rank} \begin{pmatrix} \frac{\partial \phi}{\partial z}(\bar{z}, \bar{\mu}_{I_+}) & -(G'_{I_+}(\bar{z}))^T \\ G'_I(\bar{z}) & 0 \end{pmatrix} = p + |I_+|. \tag{4.1}$$

In the case of optimization (g is a gradient mapping), $\frac{\partial \phi}{\partial z}(\bar{z}, \bar{\mu}_{I_+})$ coincides with the Hessian of the standard Lagrangian function at a KKT point $(\bar{z}, \bar{\mu})$.

We first show that (4.1) is weaker than the regularity condition for KKT systems introduced in [15, Definition 2]. Defining $I_0 = I \setminus I_+$, the latter states that

$$\exists\, D_1, D_2 \in \mathbf{R}(|I_0|, |I_0|) \text{ such that } \det \begin{pmatrix} \frac{\partial \phi}{\partial z}(\bar{z}, \bar{\mu}_{I_+}) & -(G'_{I_0}(\bar{z}))^T & -(G'_{I_+}(\bar{z}))^T \\ D_1 G'_{I_0}(\bar{z}) & D_2 & 0 \\ G'_{I_+}(\bar{z}) & 0 & 0 \end{pmatrix} \neq 0. \tag{4.2}$$

Proposition 4.1 *Let $(\bar{z}, \bar{\mu})$ be a solution of the KKT system. Then* (4.2) *implies* (4.1)*, but not vice versa.*

Proof. Assume (4.2). If (4.1) does not hold, then there exists $(\zeta, \nu_{I_+}) \neq 0$ in the kernel of the matrix in (4.1). But then $(\zeta, 0, \nu_{I_+}) \neq 0$ will be in the kernel of the matrix in (4.2) for any choice of $D_1, D_2 \in \mathbf{R}(|I_0|, |I_0|)$.

The lack of the reverse implication is shown in Example 4.1. ∎

Example 4.1 Let $p = 2$, $m = 1$, $g(z) = (z_2, 0)$, $G(z) = z_1$.

The point $(\bar{z}, \bar{\mu}) = (0, 0)$ is a solution of the KKT system (1.2), and we have $I_+ = \emptyset$, $I_0 = \{1\}$.

The matrix in (4.1) takes the form

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix},$$

and it has full column rank.

The matrix in (4.2) takes the form

$$\begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 0 \\ D_1 & 0 & D_2 \end{pmatrix},$$

which is always singular.

Note that in Example 4.1, $\frac{\partial\phi}{\partial z}(\bar{z}, \bar{\mu}_{I_+})$ is asymmetric. It is an open question at this time whether the conditions (4.2) and (4.1) are different also in the symmetric (optimization) case.

We next provide a characterization of weak regularity for KKT systems in terms of a constraint qualification and a second-order condition. We say that the weak linear independence constraint qualification (WLICQ) holds at $\bar{z}$, if

$$\operatorname{rank} G'_{I_+}(\bar{z}) = |I_+|.$$

We say that the second-order condition (SOC) holds if

$$\left\langle \frac{\partial\phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+})\zeta,\, \zeta \right\rangle \neq 0 \quad \forall\, \zeta \in \ker G'_I(\bar{z}) \setminus \{0\}.$$

Obviously, WLICQ is implied by the standard linear independence constraint qualification, while SOC is implied by the standard second-order sufficiency condition.

Proposition 4.2 *Let $(\bar{z},\, \bar{\mu})$ be a solution of the KKT system. Then WLICQ and SOC imply weak regularity of $(\bar{z},\, \bar{\mu})$. Weak regularity implies WLICQ.*

Proof. The fact that weak regularity subsumes WLICQ is obvious.

In view of (4.1), it suffices to prove that under WLICQ and SOC, the equality

$$\begin{pmatrix} \frac{\partial\phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+})\zeta - (G'_{I_+}(\bar{z}))^T \nu_{I_+} \\ G'_I(\bar{z})\zeta \end{pmatrix} = 0 \tag{4.3}$$

implies that $(\zeta,\, \nu_{I_+}) = 0$. Indeed, from (4.3) we obtain that

$$\left\langle \frac{\partial\phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+})\zeta,\, \zeta \right\rangle = \langle (G'_{I_+}(\bar{z}))^T \nu_{I_+},\, \zeta\rangle = \langle \nu_{I_+},\, G'_{I_+}(\bar{z})\zeta\rangle = 0.$$

Since $\zeta \in \ker G'_I(\bar{z})$, SOC implies that $\zeta = 0$. Thus $(G'_{I_+}(\bar{z}))^T \nu_{I_+} = 0$. By WLICQ, we have that $\nu_{I_+} = 0$, which concludes the proof. ∎

Under the weak regularity condition (4.1) and 2-regularity of Ψ_S at $(\bar{z}, \bar{\mu})$, we can solve the reduced KKT system (3.8) by GNM, with the convergence result given by Theorem 3.1. We next show that for KKT systems with symmetric $\frac{\partial\phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+})$, the latter two assumptions, when combined, are equivalent to semistability of $(\bar{z}, \bar{\mu})$ (under the second-order necessity condition, the latter is further equivalent to the uniqueness of $\bar{\mu}$ associated to $\bar{z}$ and the second-order sufficiency condition, see [3, Proposition 6.2], [15, Proposition 1]).

Proposition 4.3 *If $\frac{\partial\phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+})$ is symmetric, then semistability of a solution $(\bar{z}, \bar{\mu})$ of the KKT system is equivalent to the combination of 2-regularity of Ψ_S at this solution and weak regularity of this solution.*

Proof. The fact that semistability implies the other two properties is given by Proposition 3.3.

We next re-examine the proof of Proposition 2.1 under the new assumptions. In addition to the sets of indices defined above, let $I_N = \{1, \ldots, m\} \setminus I$ and $I_0 = I \setminus I_+$. Suppose, for simplicity of notation, that the ordering is such that in the set $\{1, \ldots, m\}$ first come the indices in I_+, then in I_0, then in I_N. With this convention, the matrix whose rows are comprised by $F'(\bar{x}), i \in A_+$, $e^i, i \in N$, is given by

$$\begin{pmatrix} \frac{\partial \phi}{\partial z}(\bar{z}, \bar{\mu}_{I_+}) & -(G'_{I_+}(\bar{z}))^T & -(G'_{I_0}(\bar{z}))^T & -(G'_{I_N}(\bar{z}))^T \\ G'_{I_+}(\bar{z}) & 0 & 0 & 0 \\ 0 & 0 & 0 & E_{I_N} \end{pmatrix}.$$

Under the assumption that $\frac{\partial \phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+})$ is symmetric, the weak regularity condition (4.1) implies that the rows of this matrix are linearly independent. Revisiting the proof of Proposition 2.1, it is easily seen that in this case

$$\operatorname{im} \Psi'_S(\bar{x}) = \{y \in \mathbf{R}^n \mid y_i = 0,\ i \in A_0\},$$

and

$$(Py)_i = \begin{cases} y_i, & \text{if } i \in A_0, \\ 0, & \text{if } i \in A_+ \cup N, \end{cases} \qquad y \in \mathbf{R}^n,$$

where

$$A_0 = A \setminus A_+ = \{p + j \mid j \in I_0\}, \quad N = \{i = 1,\, \ldots,\, n\} \setminus A = \{p + j \mid j \in I_N\}.$$

It follows that the inclusion (2.7) holds as equality, in which case the proof of Proposition 2.1 establishes that $T = L$, where T and L are defined in (2.4) and (2.3), respectively. Semistability and 2-regularity are therefore equivalent in that case. ∎

It is not difficult to observe that the Gauss–Newton iteration in a certain sense destroys the primal-dual structure present in a KKT system. For example, it does not seem possible to analyze the superlinear convergence of the primal variables separately from the convergence of the primal-dual pair. Proposition 3.4, on the other hand, allows us to make other choices of C in (3.9), with the expectation that they would still do the job. We next make one specific choice, and analyze conditions for the superlinear convergence of primal variables. We refer also to [15] (the discussion following Proposition 6) for some possibilities of how further assumptions about the problem can be taken into account in the framework of that paper. Similar options under similar assumptions could be analyzed here, but we shall not pursue the details.

Assuming again that the active constraints are ordered in such a way that the strongly active are first (i.e., for the first $|I_+|$ active constraints the corresponding multiplier is positive), let

$$C = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & C_2 & 0 \end{pmatrix}, \tag{4.4}$$

where $C_1 \in \mathbf{R}(p,p)$ and $C_2 \in \mathbf{R}(|I_+|,|I_+|)$ are arbitrary nonsingular matrices. Note that $\operatorname{rank} C = p + |I_+| = |A_+|$ holds. We shall consider further the case where in the implementation of the Newton method for (3.9) the matrix $\frac{\partial\phi}{\partial z}(z^k,\, \mu^k_{I_+})$ at iteration k is replaced by some (e.g., quasi-Newton) approximation H_k. The resulting iteration is then given by

$$C \begin{pmatrix} g(z^k) - (G'_{I_+}(z^k))^T \mu^k_{I_+} + H_k(z^{k+1} - z^k) - (G'_{I_+}(z^k))^T(\mu^{k+1}_{I_+} - \mu^k_{I_+}) \\ G_I(z^k) + G'_I(z^k)(z^{k+1} - z^k) \end{pmatrix} = 0, \tag{4.5}$$

and taking into account the chosen structure of C, it holds that

$$g(z^k) + H_k(z^{k+1} - z^k) - (G'_{I_+}(z^k))^T \mu^{k+1}_{I_+} = 0, \tag{4.6}$$

$$G_{I_+}(z^k) + G'_{I_+}(z^k)(z^{k+1} - z^k) = 0. \tag{4.7}$$

To establish a sufficient condition for the superlinear rate of convergence of primal variables, we shall need assumptions somewhat stronger than weak regularity. We say that the strong second-order sufficiency condition (SSOSC) holds, if

$$\left\langle \frac{\partial\phi}{\partial z}(\bar z,\, \bar\mu_{I_+})\zeta,\, \zeta \right\rangle \neq 0 \quad \forall\, \zeta \in \ker G'_{I_+}(\bar z) \setminus \{0\}.$$

Theorem 4.1 *Suppose that the sequence* $\{(z^k,\, \mu^k_{I_+})\}$ *generated according to* (4.5) *with* C *given by* (4.4) *converges to* $(\bar z,\, \bar\mu_{I_+})$*, a solution of the KKT system.*

If $\{z^k\}$ *converges to* $\bar z$ *superlinearly, then*

$$\Pi\left(\left(\frac{\partial\phi}{\partial z}(\bar z,\, \bar\mu_{I_+}) - H_k\right)(z^{k+1} - z^k)\right) = o(\|z^{k+1} - z^k\|), \tag{4.8}$$

where Π *is the orthogonal projector onto* $\ker G'_{I_+}(\bar z)$ *in* $\mathbf{R}^p$*.*

Under WLICQ and SSOSC, the condition (4.8) *is also sufficient for* $\{z^k\}$ *to converge to* $\bar z$ *at superlinear rate.*

Proof. By (4.6), we have that

$$\begin{aligned} -H_k(z^{k+1} - z^k) &= g(z^k) - (G'_{I_+}(z^k))^T \mu^{k+1}_{I_+} \\ &= g(\bar z) - (G'_{I_+}(\bar z))^T \mu^{k+1}_{I_+} + \frac{\partial\phi}{\partial z}(\bar z, \mu^{k+1}_{I_+})(z^k - \bar z) + o(\|z^k - \bar z\|) \\ &= (G'_{I_+}(\bar z))^T(\bar\mu_{I_+} - \mu^{k+1}_{I_+}) + \frac{\partial\phi}{\partial z}(\bar z, \bar\mu_{I_+})(z^k - \bar z) + o(\|z^k - \bar z\|), \end{aligned}$$

where in the third equality we have used the fact that $g(\bar z) = (G'_{I_+}(\bar z))^T \bar\mu_{I_+}$ and the assumption that $\{\mu^k_{I_+}\} \to \bar\mu_{I_+}$. Hence,

$$\left(\frac{\partial\phi}{\partial z}(\bar z, \bar\mu_{I_+}) - H_k\right)(z^{k+1} - z^k) = (G'_{I_+}(\bar z))^T(\bar\mu_{I_+} - \mu^{k+1}_{I_+}) + \frac{\partial\phi}{\partial z}(\bar z, \bar\mu_{I_+})(z^{k+1} - \bar z) + o(\|z^k - \bar z\|). \tag{4.9}$$

Suppose first that $z^{k+1} - \bar{z} = o(\|z^k - \bar{z}\|)$. Then we obtain from (4.9) that

$$\left(\frac{\partial \phi}{\partial z}(\bar{z}, \bar{\mu}_{I_+}) - H_k\right)(z^{k+1} - z^k) = (G'_{I_+}(\bar{z}))^T(\bar{\mu}_{I_+} - \mu_{I_+}^{k+1}) + o(\|z^k - \bar{z}\|). \tag{4.10}$$

For any $\zeta \in \ker G'_{I_+}(\bar{z})$, it holds that

$$\langle (G'_{I_+}(\bar{z}))^T(\bar{\mu}_{I_+} - \mu_{I_+}^{k+1}),\, \zeta\rangle = \langle \bar{\mu}_{I_+} - \mu_{I_+}^{k+1},\, G'_{I_+}(\bar{z})\zeta\rangle = 0,$$

which implies that

$$\Pi(G'_{I_+}(\bar{z}))^T(\bar{\mu}_{I_+} - \mu_{I_+}^{k+1}) = 0.$$

By the Lipschitz-continuity of the projection operator, we then have from (4.10) that

$$\Pi\left(\left(\frac{\partial \phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+}) - H_k\right)(z^{k+1} - z^k)\right) = o(\|z^k - \bar{z}\|),$$

and (4.8) follows by noting that

$$\|z^{k+1} - z^k\| \geq \|z^k - \bar{z}\| - \|z^{k+1} - \bar{z}\| = \|z^k - \bar{z}\| - o(\|z^k - \bar{z}\|) \geq \|z^k - \bar{z}\|/2$$

for all k large enough.

We proceed to prove the second assertion. From (4.7), we have that

$$\begin{aligned} 0 &= G_{I_+}(z^k) + G'_{I_+}(z^k)(z^{k+1} - z^k) \\ &= G_{I_+}(\bar{z}) + G'_{I_+}(\bar{z})(z^k - \bar{z}) + G'_{I_+}(z^k)(z^{k+1} - z^k) + o(\|z^k - \bar{z}\|) \\ &= G'_{I_+}(\bar{z})(z^{k+1} - \bar{z}) + \eta^k, \end{aligned} \tag{4.11}$$

where taking into account that $\{z^k\} \to \bar{z}$,

$$\eta^k = (G'_{I_+}(z^k) - G'_{I_+}(\bar{z}))(z^{k+1} - z^k) + o(\|z^k - \bar{z}\|) = o(\|z^{k+1} - z^k\|) + o(\|z^k - \bar{z}\|).$$

By WLICQ, for every k there exists $v^k \in \mathbf{R}^p$ such that

$$G'_{I_+}(\bar{z})v^k = \eta^k, \quad v^k = o(\|z^{k+1} - z^k\|) + o(\|z^k - \bar{z}\|). \tag{4.12}$$

Denoting $\zeta^k = z^{k+1} - \bar{z} + v^k$, from (4.11) we then have that $G'_{I_+}(\bar{z})\zeta^k = 0$, i.e., $\zeta^k \in \ker G'_{I_+}(\bar{z})$. By (4.9), we have that

$$\begin{aligned} \left\langle \left(\frac{\partial \phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+}) - H_k\right)(z^{k+1} - z^k),\, \zeta^k\right\rangle &= \left\langle \frac{\partial \phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+})(z^{k+1} - \bar{z}),\, \zeta^k\right\rangle \\ &\quad + \langle \bar{\mu}_{I_+} - \mu_{I_+}^{k+1},\, G'_{I_+}(\bar{z})\zeta^k\rangle + o(\|\zeta^k\|\|z^k - \bar{z}\|). \end{aligned}$$

And using SSOSC, we further have that there exists $\gamma > 0$ such that

$$\gamma\|\zeta^k\|^2 \leq \left|\left\langle \frac{\partial \phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+})\zeta^k,\, \zeta^k\right\rangle\right|$$

$$
\begin{aligned}
&= \left|\left\langle \frac{\partial \phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+})(z^{k+1} - \bar{z} + v^k),\, \zeta^k \right\rangle\right| \\
&= \left|\left\langle \left(\frac{\partial \phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+}) - H_k\right)(z^{k+1} - z^k),\, \zeta^k \right\rangle + \left\langle \frac{\partial \phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+}) v^k,\, \zeta^k \right\rangle\right| \\
&\quad + o(\|\zeta^k\|\|z^k - \bar{z}\|) \\
&= \left|\left\langle \left(\frac{\partial \phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+}) - H_k\right)(z^{k+1} - z^k),\, \zeta^k \right\rangle\right| \\
&\quad + o(\|\zeta^k\|\|z^{k+1} - z^k\|) + o(\|\zeta^k\|\|z^k - \bar{z}\|) \\
&= \left|\left\langle \Pi \left(\frac{\partial \phi}{\partial z}(\bar{z},\, \bar{\mu}_{I_+}) - H_k\right)(z^{k+1} - z^k),\, \zeta^k \right\rangle\right| \\
&\quad + o(\|\zeta^k\|\|z^{k+1} - z^k\|) + o(\|\zeta^k\|\|z^k - \bar{z}\|) \\
&= o(\|\zeta^k\|\|z^{k+1} - z^k\|) + o(\|\zeta^k\|\|z^k - \bar{z}\|),
\end{aligned}
$$

where the third equality is by (4.12), and the last equality is by (4.8). Dividing both sides in the relation above by $\|\zeta^k\|$, we have that

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

And finally,

$$
\|z^{k+1} - \bar{z}\| \le \|\zeta^k\| + \|v^k\| = o(\|z^{k+1} - z^k\|) + o(\|z^k - \bar{z}\|) = o(\|z^{k+1} - \bar{z}\|) + o(\|z^k - \bar{z}\|),
$$

from which it follows that $z^{k+1} - \bar{z} = o(\|z^k - \bar{z}\|)$. ■

5 Comparison with Other Newton-Type Methods

In this section, we very briefly compare convergence conditions needed for the active-set Newton method described above with conditions required by other Newton-type methods for MCP.

We start with some comments on the special case of KKT. In that case, weak regularity and 2-regularity of Ψ_S are equivalent to semistability of the solution. In the case when this solution is a local minimizer in an optimization problem, this is further equivalent to the uniqueness of the multiplier and the standard second-order sufficient condition for optimality [3]. A detailed comparison with other constraint qualifications and regularity conditions for KKT systems is given in [15] (even though the weak regularity condition itself is different in [15], and the resulting class of active-set methods is different, most of the comparison comments in [15] still apply here, due to the equivalence with semistability). Summarizing, conditions for the superlinear convergence of the proposed method are strictly weaker than those of SNM for KKT systems, and are the same as conditions required by the sequential quadratic programming method (SQP) in its basic form. Note that SQP subproblems are quadratic programs, while the subproblems of the method proposed here are just systems of linear equations. On the other hand, convergence conditions of SQP can be somewhat weakened, but at the expense of nontrivial modifications of the basic iteration, which come with a computational price [29, 11, 12]. We next focus on the more general case of MCP.

In the case of general MCP, convergence conditions for our method are, of course, again significantly weaker than conditions for SNM [20, 21, 26, 27] applied to Ψ_{FB} or Ψ_{NR} [1, 19, 9] (see also [8, Ch. 9]). Indeed, the latter need BD-regularity of the corresponding function at the solution, which certainly implies the corresponding error bound [25], and thus semistability, but not vice versa. We note, in the passing, that BD-regularity of Ψ_{FB} and Ψ_{NR} are not related, i.e., none is weaker or stronger than the other. The fact that BD-regularity of Ψ_{NR} does not imply this property for Ψ_{FB} has been exhibited in [22, Example 2.1]. We have not been able to find in the literature an example showing that BD-regularity of Ψ_{FB} does not imply this property for Ψ_{NR}. So we provide such an example below.

Example 5.1 Let $n = 2$, $F(x) = (x_2, -x_1 + x_2)$, $l_i = 0$, $u_i = +\infty$, $i = 1, 2$. Then $\bar{x} = 0$ is a solution of this NCP, and it can be seen that all matrices in the B-differential of Ψ_{FB} at $\bar{x}$ are nonsingular, while the B-differential of Ψ_{NR} at $\bar{x}$ contains the singular matrix $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$.

Most remarkably, convergence conditions for our method are also strictly weaker than those for the linearization (Josephy-Newton) method [17, 3], which consists of solving the following subproblems: Given the current iterate $x^k \in \mathbf{R}^n$,

$$\text{find } x^{k+1} \in B \text{ such that } \langle F(x^k) + F'(x^k)(x^{k+1} - x^k), y - x^{k+1} \rangle \geq 0 \text{ for all } y \in B.$$

Those type of methods are sometimes referred to as point-based approximation methods [12]. In the case of optimization, they are closely related to SQP. Conditions for the local superlinear convergence of the linearization method are semistability *and* hemistability [3] of the solution. And in general [3, Remark 2.4], hemistability does not follow from semistability (it does in the case of KKT). Since our method requires something even less than semistability, we conclude that its local convergence properties are significantly stronger.

6 Local Numerical Experiments

In this section we report numerical experiments on some examples designed to highlight the case where various standard regularity conditions do not hold, and thus SNM-based methods have trouble or converge slowly. This is precisely the case where the switch to our local algorithm can be useful. We have also implemented a globalization of our local algorithm, in the spirit of hybrid versions of SNM/FB methods for complementarity problems based on the Ψ_{FB} reformulation (e.g., "General Line Search Algorithm" of [22]). Details of this implementation and numerical results on the MCPLIB test problems collection (the newer version of [6]) can be found in the technical report [5]. Before describing our local experiments, we give some general conclusions for the globalized version of [5]. The option of switching to our active-set step never harms too much the global algorithm, though we certainly have to pay some extra price for computing this step at those iterations at which it is eventually rejected (there are safeguards and heuristics that can be used to avoid computing the active-set step too early, see [5]). We emphasize that the overall goal of the presented approach is not in improving the SNM (or any other algorithm) when it works efficiently, but rather in safeguarding *fast local convergence* in irregular cases for which other methods do not work well. The results of [5] indicate that this can be achieved without paying a high price over iterations of the globally convergent hybrid SNM/FB method.

We now describe some possible scenarios of the *local* behavior of SNM/FB and GNM/AS by applying them to some small test problems with various combinations of satisfied and violated regularity properties of the solution that have been discussed above. Constructing artificial examples allows us to obtain a rather complete selection of irregular MCPs with "various degrees of irregularity", and to make reliable conclusions about the reasons for the observed performance of the algorithms. This is not possible, for example, for MCPLIB problems, since precise regularity properties of solutions are typically not known. On the other hand, we cannot draw far reaching conclusions based on small artificial examples. What follows is intended merely to illustrate the theoretical results and comparisons obtained above.

By GNM/AS we mean here the algorithm specified by (3.11), (3.12), with the index sets identified at the starting point (thus, the local properties of this algorithm are given by Theorem 3.1).

By SNM/FB we mean the iterative procedure

$$x^{k+1} = x^k - \Lambda_k^{-1}\Psi_{FB}(x^k), \quad \Lambda_k \in \partial_B \Psi_{FB}(x^k), \quad k = 0, 1, \ldots,$$

where the element Λ_k of the B-differential of Ψ_{FB} at x^k is computed by the procedure suggested in [1] (with $z_i = 1\ \forall\, i = 1, \ldots, n$, in the notation of [1], and the "computer zero" parameter set to 10^{-10}).

The stopping criterion for both methods is

$$\|\Psi_{FB}(x^k)\| < 10^{-9}. \tag{6.1}$$

The first problem is a slight modification of Example 3.1.

Example 6.1 Let $n = 2$, $l_i = 0$, $u_i = +\infty$, $i = 1, 2$, and let $F(x) = ((x_1 - 1)^2, x_1 + x_2 + x_2^2 - 1)$. The point $\bar{x} = (1, 0)$ is the solution of this NCP. Semistability is violated here (and hence, BD-regularity for Ψ_{NR} and Ψ_{FB} is violated), while 2-regularity of Ψ_S and weak regularity hold at $\bar{x}$. The starting point is $x^0 = (1.5, -0.5)$, with $\|x^0 - \bar{x}\| \approx$ 7.1e–01, $\|\Psi_{FB}(x^0)\| \approx$ 8.2e–01, $\det \Lambda_0 \approx$ 1.4e+00.

SNM/FB converges in 13 steps. At termination, $\|x^{13} - \bar{x}\| \approx$ 3.0e–05, $\|\Psi_{FB}(x^{13})\| \approx$ 6.2e–10, $\det \Lambda_{13} \approx$ –8.3e–05 (note that the latter indicates degeneracy). The rate of convergence is linear, with the ratio approaching 1/2.

The behavior of GNM/AS is reported in Table 1, and it clearly shows fast quadratic convergence.

Table 1: *GNM/AS for Example 6.1*

k	0	1	2	3
$\|x^k - \bar{x}\|$	7.1e–01	1.3e–01	3.7e–03	9.9e–08
$\|\Psi_{FB}(x^k)\|$	8.2e–01	1.6e–02	1.4e–05	9.9e–15
$\frac{\|x^k - \bar{x}\|}{\|x^{k-1} - \bar{x}\|}$		1.8e–01	2.9e–02	2.7e–05

The next problem is a modification of [15, Example 1].

Example 6.2 Let $p = m = 2$, $f(z) = (z_1 + z_2)^2/2 + (z_1 + z_2)^3/3$, $G(z) = (z_1, z_2)$, $z \in \mathbf{R}^2$, $\bar{z} = 0$, $\bar{\mu} = 0$. Semistability holds here, but for Ψ_{NR} (and hence, for Ψ_{FB}), BD-regularity is violated. The starting point is $z^0 = (1, 2)$, $\mu^0 = (0.01, 0.01)$, with $\|x^0 - \bar{x}\| \approx$ 2.2e+00, $\|\Psi_{FB}(x^0)\| \approx$ 1.7e+01, $\det \Lambda_0 \approx$ 4.3e–04.

The behavior of SNM/FB is as follows: $\det \Lambda_1 \approx$ 7.3e–10, $\det \Lambda_2 \approx 0$, but the corresponding linear system is solvable, and the method manages to escape the "bad" region. Specifically, $\det \Lambda_3 \approx$ 4.3e–16, while $\det \Lambda_4 \approx$ 2.5e–00, and the algorithm converges in 7 iterations. At the final step, $\|x^7 - \bar{x}\| \approx$ 1.0e–16, $\|\Psi_{FB}(x^7)\| \approx$ 1.4e–16, $\det \Lambda_7 \approx 2$. The rate of convergence is superlinear. The behavior of GNM/AS is reported in Table 2, and it also shows the superlinear rate.

Note that while SNM/FB and GNM/AS exhibit similar convergence for this problem, the performance of SNM/FB depends on the specific implementation. In particular, the solution which is produced for a given *degenerate* linear system clearly depends on the linear solver being used. The choice of this solution can affect the overall convergence. Also, in general (this is not the case for this example), when BD-regularity is violated different procedures to compute Λ_k could result in different linear systems some of which can be ill-conditioned close to the solution, preventing fast convergence of SNM/FB.

Table 2: *GNM/AS for Example 6.2*

k	0	1	2	3	4	5	6	7
$\|x^k - \bar{x}\|$	2.2e+00	9.0e–01	3.2e–01	7.1e–02	5.0e–03	2.8e–05	9.1e–10	9.3e–19
$\|\Psi_{FB}(x^k)\|$	1.7e+01	4.1e+00	9.3e–01	1.6e–01	1.0e–02	5.7e–05	1.8e–09	1.8e–18
$\frac{\|x^k - \bar{x}\|}{\|x^{k-1} - \bar{x}\|}$		4.0e–01	3.5e–01	2.2e–01	7.1e–02	5.6e–03	3.2e–05	1.0e–09

The next problem is [15, Example 4]. Combined with Example 6.5 below, it seems to indicate that weak regularity is somewhat more important for the success of GNM/AS than 2-regularity.

Example 6.3 Let $p = m = 1$, $f(z) = z^4/4$, $G(z) = z$, $z \in \mathbf{R}$, $\bar{z} = 0$, $\bar{\mu} = 0$. Weak regularity holds here but 2-regularity of Ψ_S does not, and thus, semistability is violated (and hence, BD-regularity for Ψ_{NR} and Ψ_{FB} is also violated). The starting point is $z^0 = 1$, $\mu^0 = 0.1$, with $\|x^0 - \bar{x}\| \approx$ 1.0e+00, $\|\Psi_{FB}(x^0)\| \approx$ 9.1e–01, $\det \Lambda_0 \approx$ 2.7e+00.

SNM/FB converges in 18 steps. At termination, $\|x^{18} - \bar{x}\| \approx$ 6.8e–04, $\|\Psi_{FB}(x^{18})\| \approx$ 3.1e–10, $\det \Lambda_{13} \approx$ 3.1e–06. The rate of convergence is linear with ratio approaching 2/3.

The behavior of GNM/AS is reported in Table 3, and it shows fast quadratic convergence.

Table 3: *GNM/AS for Example 6.3*

k	0	1	2	3	4
$\|x^k - \bar{x}\|$	1.0e+00	6.0e–01	2.2e–01	2.7e–03	9.0e-13
$\|\Psi_{FB}(x^k)\|$	9.1e–01	2.2e–01	1.0e–03	2.0e–08	7.4e-37
$\frac{\|x^k - \bar{x}\|}{\|x^{k-1} - \bar{x}\|}$		6.0e–01	3.6e–01	1.3e–02	3.3e-10

The next example is borrowed from [22, Example 2.1].

Example 6.4 Let $n = 2$, $F(x) = (-x_1 + x_2,\, -x_2)$, $x \in \mathbf{R}^2$, $\bar{x} = 0$. BD-regularity holds for Ψ_{NR} (and hence, semistability also holds), but not for Ψ_{FB}. The starting point is $x^0 = (2,\, 4)$, with $\|x^0 - \bar{x}\| \approx$ 4.5e+00, $\|\Psi_{FB}(x^0)\| \approx$ 5.8e+00, $\det \Lambda_0 = 0$.

Here, SNM/FB fails to make a step. At the same time, GNM/AS terminates after 1 step at the exact solution. The reason for this is that $A_0 = \{1,\, 2\}$. Thus, the iteration of GNM/AS reduces to identifying the index sets.

Note that the problem in Example 6.4 is actually a linear complementarity problem, that is, NCP with affine F. We point out that in the case of affine F, just one step of GNM/AS gives the exact solution, whenever the index sets are correctly identified. For example, this behavior is observed also for the problem `badfree` from the MCPLIB collection: once x^k is close to the solution $\bar{x} = (0,\, 0,\, 0.5,\, 0.5,\, 1)$, GNM/AS produces $x^{k+1} = \bar{x}$. At the same time, for x^k close to $\bar{x}$, a degenerate Λ_k is computed, and SNM/FB fails, see [5].

The next example is [15, Example 2] and it shows that both 2-regularity of Ψ_S and weak regularity are important for fast convergence of GNM/AS in general (recall, however, Example 6.3).

Example 6.5 Let $p = m = 2$, $f(z) = z_1^2/2 + z_2^3/3$, $G(z) = (z_1 - z_2^2/2,\, z_1 + z_2^2/2)$, $z \in \mathbf{R}^2$, $\bar{z} = 0$, $\bar{\mu} = 0$. Semistability is violated (and hence, BD-regularity for Ψ_{NR} and Ψ_{FB} is violated). For Ψ_S, 2-regularity holds, but weak regularity does not. The starting point is $z^0 = (0.1,\, 0.1)$, $\mu^0 = (0.1,\, 0.1)$, with $\|x^0 - \bar{x}\| \approx$ 2.0e–01, $\|\Psi_{FB}(x^0)\| \approx$ 1.3e–01, $\det \Lambda_0 \approx$ 2.0e–01.

Both SNM/FB and GNM/AS converge in 12 steps, and $\|x^{12} - \bar{x}\| \approx$ 2.4e–05, $\|\Psi_{FB}(x^{12})\| \approx$ 8.4e–10, $\det \Lambda_{12} \approx$ 2.9e–04. The rate of convergence is linear with ratio 1/2.

Acknowledgment. We thank the two anonymous referees for constructive suggestions. In particular, Proposition 4.1 and Example 4.1 are in response to a question posed by one of the referees.

References

[1] S.C. Billups. Algorithms for complementarity problems and generalized equations. Technical Report 95-14, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, August 1995. PhD thesis.

[2] S.C. Billups. A homotopy-based algorithm for mixed complementarity problems. *SIAM Journal on Optimization*, 12:583–605, 2002.

[3] J. F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29:161–186, 1994.

[4] H. Dan, N. Yamashita, and M. Fukushima. A superlinearly convergent algorithm for the monotone nonlinear complementarity problem without uniqueness and nondegeneracy conditions. *Mathematics of Operations Research*, 27:743–754, 2002.

[5] A.N. Daryina, A.F. Izmailov, and M.V. Solodov. Numerical results for a globalized active-set Newton method for mixed complementarity problems. IMPA technical report A282, April 2004.

[6] S.P. Dirkse and M.C. Ferris. MCPLIB: A collection of nonlinear mixed complementarity problems. *Optimization Methods and Software*, 5:319–345, 1995.

[7] F. Facchinei, A. Fischer, and C. Kanzow. On the accurate identification of active constraints. *SIAM Journal on Optimization*, 9:14–32, 1999.

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

[9] M.C. Ferris, C. Kanzow, and T.S. Munson. Feasible descent algorithms for mixed complementarity problems. *Mathematical Programming*, 86:475–497, 1999.

[10] M.C. Ferris and J.-S. Pang. Engineering and economic applications of complementarity problems. *SIAM Review*, 39:669–713, 1997.

[11] A. Fischer. Modified Wilson's method for nonlinear programs with nonunique multipliers. *Mathematics of Operations Research*, 24:699–727, 1999.

[12] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Mathematical Programming*, 94:91–124, 2002.

[13] A.F. Izmailov and M.V. Solodov. Error bounds for 2-regular mappings with Lipschitzian derivatives and their applications. *Mathematical Programming*, 89:413–435, 2001.

[14] A.F. Izmailov and M.V. Solodov. The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions. *Mathematics of Operations Research*, 27:614–635, 2002.

[15] A.F. Izmailov and M.V. Solodov. Karush-Kuhn-Tucker systems: regularity conditions, error bounds and a class of Newton-type methods. *Mathematical Programming*, 95:631–650, 2003.

[16] H. Jiang and D. Ralph. Global and local superlinear convergence analysis of Newton-type methods for semismooth equations with smooth least squares. In M. Fukushima and L. Qi, editors, *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pages 181–210. Kluwer Academic Publishers, 1999.

[17] N.H. Josephy. *Newton's Method for Generalized Equations and the PIES Energy Model*. PhD thesis, Department of Industrial Engineering, University of Wisconsin–Madison, 1979.

[18] C. Kanzow. Strictly feasible equation-based methods for mixed complementarity problems. *Numerische Mathematik*, 89:135–160, 2001.

[19] C. Kanzow and M. Fukushima. Solving box constrained variational inequality problems by using the natural residual with D-gap function globalization. *Operations Research Letters*, 23:45–51, 1998.

[20] B. Kummer. Newton's method for nondifferentiable functions. In J. Guddat et al., editor, *Advances in Mathematical Optimization*, volume 45, pages 114–125. Akademie-Verlag, Berlin, 1988.

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

[22] T. De Luca, F. Facchinei, and C. Kanzow. A theoretical and numerical comparison of some semismooth algorithms for complementarity problems. *Computational Optimization and Applications*, 16:173–205, 2000.

[23] Z.-Q. Luo and J.-S. Pang. Error bounds for analytic systems and their applications. *Mathematical Programming*, 67:1–28, 1994.

[24] J.-S. Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79:299–332, 1997.

[25] J.-S. Pang and L. Qi. Nonsmooth equations: Motivation and algorithms. *SIAM Journal on Optimization*, 3:443–465, 1993.

[26] L. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18:227–244, 1993.

[27] L. Qi and J. Sun. A nonsmooth version of Newton's method. *Mathematical Programming*, 58:353–367, 1993.

[28] P. Tseng. Growth behavior of a class of merit functions for the nonlinear complementarity problem. *Journal of Optimization Theory and Applications*, 89:17–37, 1996.

[29] S.J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Computational Optimization and Applications*, 11:253–275, 1998.