Newton-Type Methods for Optimization Problems without Constraint Qualifications
— Source link ↗

Alexey F. Izmailov, Mikhail V. Solodov

Topics: Constraint (information theory) and Optimization problem

Related papers:

- An Algorithm for Degenerate Nonlinear Programming with Rapid Local Convergence

- Superlinear Convergence of a Stabilized SQP Method to a Degenerate Solution

- Local behavior of an iterative framework for generalized equations with nonisolated solutions

- Stabilized Sequential Quadratic Programming

- Modifying SQP for Degenerate Problems

Share this paper:

NEWTON-TYPE METHODS FOR OPTIMIZATION PROBLEMS WITHOUT CONSTRAINT QUALIFICATIONS*

A. F. IZMAILOV† AND M. V. SOLODOV‡

Abstract. We consider equality-constrained optimization problems, where a given solution may not satisfy any constraint qualification but satisfies the standard second-order sufficient condition for optimality. Based on local identification of the rank of the constraints degeneracy via the singular-value decomposition, we derive a modified primal-dual optimality system whose solution is locally unique, nondegenerate, and thus can be found by standard Newton-type techniques. Using identification of active constraints, we further extend our approach to mixed equality- and inequality-constrained problems, and to mathematical programs with complementarity constraints (MPCC). In particular, for MPCC we obtain a local algorithm with quadratic convergence under the second-order sufficient condition only, without any constraint qualifications, not even the special MPCC constraint qualifications.

Key words. constraints degeneracy, regularity, second-order sufficiency, singular-value decomposition, Newton method, mathematical programs with complementarity constraints (MPCC)

AMS subject classifications. 90C30, 90C33, 90C55, 65K05

DOI. 10.1137/S1052623403427264

1. Introduction. We consider the mathematical programming problem

$$
(1.1) \qquad
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & F(x) = 0, \ G(x) \le 0,
\end{aligned}
$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is a smooth function, and $F : \mathbf{R}^n \to \mathbf{R}^l$ and $G : \mathbf{R}^n \to \mathbf{R}^m$ are smooth mappings. Let $\bar{x} \in \mathbf{R}^n$ be a local solution of (1.1). For the superlinear convergence of the method presented below, we shall need the problem data to be twice differentiable near $\bar{x}$, with the second derivatives being continuous at $\bar{x}$. For quadratic convergence, second derivatives should be Lipschitz continuous near $\bar{x}$.

The stationary points of problem (1.1) and the associated (normal) Lagrange multipliers are characterized by the Karush–Kuhn–Tucker (KKT) optimality system

$$
(1.2) \qquad \frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad F(x) = 0, \quad \mu \ge 0, \quad G(x) \le 0, \quad \langle \mu, G(x) \rangle = 0,
$$

where

$$
L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}, \quad L(x, \lambda, \mu) = f(x) + \langle \lambda, F(x) \rangle + \langle \mu, G(x) \rangle
$$

is the standard Lagrangian function of problem (1.1).

For a local solution $\bar{x}$ of (1.1), let $\bar{Y}$ be the set of Lagrange multipliers associated with $\bar{x}$; that is,

$$\bar{Y} = \bar{Y}(\bar{x}) = \left\{ (\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m \,\middle|\, \frac{\partial L}{\partial x}(\bar{x}, \lambda, \mu) = 0, \ \mu \geq 0, \ \langle \mu, G(\bar{x}) \rangle = 0 \right\}.$$

If nonempty, $\bar{Y}$ is a convex polyhedral set. As is well known, $\bar{Y}$ is nonempty and bounded if and only if the Mangasarian–Fromovitz constraint qualification (MFCQ) holds:

(1.3) $\operatorname{rank} F'(\bar{x}) = l$

and

$$\exists \bar{\xi} \in \ker F'(\bar{x}) \text{ such that } G'_{\bar{I}}(\bar{x})\bar{\xi} < 0,$$

where $\bar{I} = \bar{I}(\bar{x}) = \{i = 1, \ldots, m \mid G_i(\bar{x}) = 0\}$ is the set of indices of inequality constraints active at $\bar{x}$. If the stronger linear independence constraint qualification (LICQ) holds, i.e.,

the rows of $F'(\bar{x})$, $G'_{\bar{I}}(\bar{x})$ are linearly independent,

then $\bar{Y}$ is a singleton. At the same time, as is well known, violation of MFCQ can give rise to the situation when $\bar{Y} = \emptyset$.

In this paper, we are concerned with problem (1.1) (or system (1.2)) in the following setting: We do not assume *any* constraint qualification (CQ) to hold at the solution $\bar{x}$; at the same time, we assume the existence of a pair of Lagrange multipliers $(\bar{\lambda}, \bar{\mu}) \in \bar{Y}$ and the standard second-order sufficient condition (SOSC) with this pair of multipliers:

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

where $\bar{K}$ is the critical cone of problem (1.1) at $\bar{x}$; that is,

$$\bar{K} = \bar{K}(\bar{x}) = \{\xi \in \ker F'(\bar{x}) \mid \langle f'(\bar{x}), \xi \rangle \leq 0, \ G'_{\bar{I}}(\bar{x})\xi \leq 0\}.$$

The case of possible violation of standard CQs has recently received much attention in the literature [40, 20, 21, 16, 35, 1, 2, 3, 23, 17, 5, 4, 24, 41, 42, 43]. As discussed in some of those references, the motivation for considering such *irregular* cases comes from various problems, where either standard CQs are inherently violated (e.g., mathematical programs with equilibrium constraints (MPEC) [34, 37, 3, 36, 38, 18]) or constraints tend to be degenerate or "nearly" (numerically) degenerate (e.g., some classes of large-scale problems). Of interest are both theoretical properties of irregular problems as well as convergence of optimization algorithms applied to such problems (and possible modification of the algorithms to improve robustness). We note that most of the cited non-MPEC literature (actually, all except for [21, 20, 2, 17]) deal with the case when LICQ can be violated (or, more generally, the multiplier can be nonunique), but MFCQ still holds. We comment on those issues in more detail in section 5, where some comparisons are also given.

We emphasize that in this paper, we do not assume *any* CQ. Hence, the existence of multipliers is not automatic. At the same time, our development relies on the

existence of the pair of multipliers $(\bar{\lambda}, \bar{\mu})$. We point out that the situation when CQs are violated but the multipliers exist is not unusual and deserves special consideration. For instance, it is known that for purely equality-constrained problems, multipliers exist generically, provided the number of variables is large enough with respect to the number of constraints [6]. Also, much of the MPEC literature (see, for example, [36, 18]) is concerned with the case of normal Lagrange multipliers. Even though their existence cannot be guaranteed, it turns out that in many cases of MPEC interest the multipliers do exist.

The rest of the paper is organized as follows. In section 2, we introduce the main ideas for the purely equality-constrained problem. The local algorithm we propose is well defined and attains superlinear/quadratic convergence under the sole assumption of SOSC. It is based on the local analysis of the constraints degeneracy via the singular-value decomposition, and on using this information to construct a modified (optimality) system of nonlinear equations whose solution is locally unique and nondegenerate. In section 3, we extend the method to the case of mixed equality and inequality constraints via the identification of constraints active at a solution. We note that SOSC is again sufficient for this construction. An application to mathematical programs with complementarity constraints (MPCC) is discussed in section 4, together with two examples to illustrate this application. In particular, we show that our approach leads to a local quadratically convergent algorithm even if the special MPCC constraint qualifications do not hold (classical CQs are automatically violated in MPCC). In section 5, we compare our approach with some other algorithms, which either do not require CQs or use some relaxed CQs.

2. Equality-constrained problems. We start our discussion with the special case of the purely equality-constrained problem

$$\begin{array}{ll} \text{(2.1)} & \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & F(x) = 0. \end{array} \end{array}$$

In this case,

$$L(x, \lambda) = f(x) + \langle \lambda, F(x) \rangle, \quad x \in \mathbf{R}^n, \lambda \in \mathbf{R}^l,$$

and the KKT system (1.2) reduces to the Lagrange system of nonlinear equations

$$\text{(2.2)} \qquad \qquad \Phi(x, \lambda) = 0,$$

where

$$\Phi : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^n \times \mathbf{R}^l, \quad \Phi(x, \lambda) = \left(\frac{\partial L}{\partial x}(x, \lambda), F(x) \right).$$

The set of Lagrange multipliers associated with a local solution $\bar{x}$ of (2.1) is the solution set of the system of linear equations:

$$\text{(2.3)} \qquad \qquad \bar{Y} = \left\{ \lambda \in \mathbf{R}^l \,\middle|\, (F'(\bar{x}))^{\mathrm{T}} \lambda = -f'(\bar{x}) \right\}.$$

For (2.1), MFCQ and LICQ are the same: They both reduce to (1.3). Furthermore, we have that $\bar{Y}$ is a singleton if and only if (1.3) holds. Thus, without (1.3), $\bar{Y}$ cannot be a singleton (in which case $\bar{Y}$ is a plane parallel to $\ker(F'(\bar{x}))^{\mathrm{T}} = (\mathrm{im}\, F'(\bar{x}))^{\perp}$). It follows that without (1.3), a solution $(\bar{x}, \bar{\lambda})$ of (2.2) cannot be isolated, $\Phi'(\bar{x}, \bar{\lambda})$ is necessarily singular, and one cannot use standard Newton-type methods for solving (2.2) (or

(2.1)). For future discussion, recall that a Newton method for (2.2) is equivalent to the sequential quadratic programming (SQP) algorithm applied to (2.1); see, e.g., [8].

Thus, we face the situation where any solution of (2.2) of the form $(\bar{x}, \lambda)$, $\lambda \in \mathbf{R}^l$, is expected to be degenerate. It is important to comment that a class of methods developed for computing degenerate solutions of general systems of nonlinear equations (see [27, 9]; see also [26, section 2] and the references therein) is not applicable in this context for some principal reasons. Specifically, it can be shown that the so-called nonsingularity of the second differential [9] is automatically violated for the Lagrange optimality system (2.2) in the situation where SOSC holds but (1.3) does not. Thus, while this class of methods for solving degenerate equations can be useful elsewhere, it does not help in solving (2.2) (or (2.1)) when LICQ is violated. Another possibility is to use the regularization methods, e.g., of the Levenberg–Marquardt type, such as those in [17]. This class of methods can attain the superlinear convergence under the stated assumptions. However, the regularization parameter has to be driven to zero, thus making the subproblems locally numerically ill-conditioned whenever standard regularity conditions do not hold. In this paper, we propose a different approach, one not based on the Levenberg–Marquardt type of regularization.

Suppose, for a moment, that $\bar{x}$ and $\bar{\lambda} \in \bar{Y}$ are known. Let $\bar{\Pi}$ be the orthogonal projector in $\mathbf{R}^l$ onto $(\operatorname{im} F'(\bar{x}))^\perp = \ker(F'(\bar{x}))^\mathrm{T}$, i.e.,

$$(2.4) \qquad \bar{\Pi}^2 = \bar{\Pi}, \quad \operatorname{im} \bar{\Pi} = \ker(F'(\bar{x}))^\mathrm{T}, \quad \ker \bar{\Pi} = \operatorname{im} F'(\bar{x}),$$

and consider the following modification of system (2.2):

$$(2.5) \qquad \begin{aligned} (F'(x))^\mathrm{T} \lambda &= -f'(x), \\ F(x) + \bar{\Pi}(\lambda - \bar{\lambda}) &= 0. \end{aligned}$$

Obviously, $(\bar{x}, \bar{\lambda})$ is a solution of (2.5). Let $(\bar{x}, \lambda)$ be any other solution. Then it holds that

$$\left. \begin{aligned} \lambda - \bar{\lambda} &\in \ker(F'(x))^\mathrm{T} = (\operatorname{im} F'(\bar{x}))^\perp \\ \lambda - \bar{\lambda} &\in \ker \bar{\Pi} = \operatorname{im} F'(\bar{x}) \end{aligned} \right\} \Rightarrow \lambda - \bar{\lambda} = 0.$$

Hence, $\bar{\lambda}$ is the unique solution of (2.5) associated with $\bar{x}$.

Of course, $\bar{x}$ and $\bar{Y}$ are unknown. The idea is to construct a suitable approximation of the system (2.5) using the information available at some $(\tilde{x}, \tilde{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$, which is close enough to $\bar{x}$ and some $\bar{\lambda} \in \bar{Y}$. We shall modify the system (2.2) in such a way that the new system will have an isolated solution of the form $(\bar{x}, \hat{\lambda})$ with some $\hat{\lambda} \in \bar{Y}$ close to $\bar{\lambda}$. This general idea goes back to [22, 27]. However, the proposals in [22, 27] have some drawbacks when compared to the approach developed here. First, the regularized system proposed in the cited work is different and, in particular, it is overdetermined. Second, the initial guess close enough to the specific multiplier is required. And even more importantly, no procedure for approximating $\bar{\Pi}$ was suggested in [22, 27], nor were conditions that are necessary for a suitable approximation discussed. Here, we give these questions a thorough consideration. Furthermore, we extend the approach to mixed constraints, including the important case of complementarity constraints.

Let a pair $(\tilde{x}, \tilde{\lambda})$ close enough to $(\bar{x}, \bar{\lambda})$ be given, and consider the "regularized" system

$$(2.6) \qquad \Phi_{\tilde{x}, \tilde{\lambda}}(x, \lambda) = 0,$$

where

$$\Phi_{\tilde{x},\tilde{\lambda}} : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^n \times \mathbf{R}^l, \quad \Phi_{\tilde{x},\tilde{\lambda}}(x,\lambda) = \left(\frac{\partial L}{\partial x}(x,\lambda),\, F(x) + \Pi(\tilde{x},\tilde{\lambda})(\lambda - \tilde{\lambda}) \right),$$

and $\Pi(\tilde{x},\tilde{\lambda})$ is a projector in $\mathbf{R}^l$; that is,

$$(2.7) \qquad (\Pi(\tilde{x},\tilde{\lambda}))^2 = \Pi(\tilde{x},\tilde{\lambda}),$$

satisfying

$$(2.8) \qquad \Pi(\tilde{x},\tilde{\lambda}) \to \bar{\Pi} \ \text{ as } \tilde{x} \to \bar{x},\, \tilde{\lambda} \to \bar{\lambda}.$$

In what follows, we shall show how an appropriate $\Pi(\tilde{x},\tilde{\lambda})$ can be constructed using the singular-value decomposition of $F'(\tilde{x})$. But first, we show that the general idea outlined above indeed resolves the problem of nonisolated multipliers.

PROPOSITION 2.1. *Let $(\bar{x},\bar{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$ be a solution of (2.2).*

Then for every $(\tilde{x},\tilde{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x},\bar{\lambda})$, there exists a unique element $\hat{\lambda} = \hat{\lambda}(\tilde{x},\tilde{\lambda}) \in \mathbf{R}^l$ such that $(\bar{x},\hat{\lambda})$ is a solution of (2.6) (that is, $\hat{\lambda}$ is the unique element in $\bar{Y}$ satisfying $\Pi(\tilde{x},\tilde{\lambda})(\hat{\lambda} - \tilde{\lambda}) = 0$). Moreover,

$$(2.9) \qquad \hat{\lambda} \to \bar{\lambda} \ \text{ as } \tilde{x} \to \bar{x},\, \tilde{\lambda} \to \bar{\lambda}.$$

Proof. Consider the linear system of equations

$$(2.10) \qquad (F'(\bar{x}))^{\mathrm{T}}\lambda = -f'(\bar{x}), \quad \bar{\Pi}\tilde{\Pi}\lambda = \bar{\Pi}\tilde{\Pi}\tilde{\lambda},$$

where $\lambda \in \mathbf{R}^l$ is the unknown, and $\tilde{\lambda} \in \mathbf{R}^l$ and the projector $\tilde{\Pi} : \mathbf{R}^l \to \mathbf{R}^l$ are parameters. One can think of the image space of the operator associated with (2.10) as the l-dimensional space $\operatorname{im}(F'(\bar{x}))^{\mathrm{T}} \times (\operatorname{im} F'(\bar{x}))^{\perp}$ (recall that $(\operatorname{im} F'(\bar{x}))^{\perp} = \ker(F'(\bar{x}))^{\mathrm{T}} = \operatorname{im}\bar{\Pi}$). Note also that the right-hand side of (2.10) certainly belongs to $\operatorname{im}(F'(\bar{x}))^{\mathrm{T}} \times (\operatorname{im} F'(\bar{x}))^{\perp}$, since $-f'(\bar{x}) = (F'(\bar{x}))^{\mathrm{T}}\bar{\lambda}$.

We first show that for the choice of $\tilde{\Pi} = \bar{\Pi}$, the operator of the system (2.10) is nonsingular. Take an arbitrary $\eta \in \mathbf{R}^l$ satisfying

$$(F'(\bar{x}))^{\mathrm{T}}\eta = 0, \quad \bar{\Pi}^2\eta = 0.$$

Taking into account the definition of $\bar{\Pi}$ (and, in particular, the equality $\bar{\Pi}^2 = \bar{\Pi}$), from the second equation it follows that $\eta \in \ker\bar{\Pi} = \operatorname{im} F'(\bar{x})$. On the other hand, the first equation can be written in the form $\eta \in \ker(F'(\bar{x}))^{\mathrm{T}} = (\operatorname{im} F'(\bar{x}))^{\perp}$. We conclude that $\eta = 0$, which establishes the assertion.

Next, note that the system (2.10) with $\tilde{\Pi} = \bar{\Pi}$ and $\tilde{\lambda} = \bar{\lambda}$ has a solution $\bar{\lambda}$, and according to what we have already proved, this solution is unique. Now, according to the standard stability theorem for linear systems with nonsingular matrices, we conclude that for every $(\tilde{\Pi},\tilde{\lambda})$ close enough to $(\bar{\Pi},\bar{\lambda})$ (in the appropriate space), (2.10) has a unique solution $\hat{\lambda}$ and, moreover,

$$(2.11) \qquad \hat{\lambda} \to \bar{\lambda} \ \text{ as } \tilde{\Pi} \to \bar{\Pi},\, \tilde{\lambda} \to \bar{\lambda}.$$

Furthermore, taking into account the equality $\tilde{\Pi}^2 = \tilde{\Pi}$, from the second equation in (2.10) we have

$$\begin{aligned} 0 &= \bar{\Pi}\tilde{\Pi}(\hat{\lambda} - \tilde{\lambda}) \\ &= \tilde{\Pi}^2(\hat{\lambda} - \tilde{\lambda}) + (\bar{\Pi} - \tilde{\Pi})\tilde{\Pi}(\hat{\lambda} - \tilde{\lambda}) \\ &= (E + (\bar{\Pi} - \tilde{\Pi}))\tilde{\Pi}(\hat{\lambda} - \tilde{\lambda}), \end{aligned}$$

where $E + (\bar{\Pi} - \tilde{\Pi})$ is nonsingular provided that $\tilde{\Pi}$ is close enough to $\bar{\Pi}$ (E denotes the identity operator). Hence, for such $\tilde{\Pi}$, it must hold that $\tilde{\Pi}(\hat{\lambda} - \tilde{\lambda}) = 0$. Combined with the first equation in (2.10), this equality means that $(\bar{x}, \hat{\lambda})$ is a solution of the system

$$(2.12) \qquad \frac{\partial L}{\partial x}(x, \lambda) = 0, \quad F(x) + \tilde{\Pi}(\lambda - \tilde{\lambda}) = 0.$$

Suppose now that there exists some other $\lambda \in \mathbf{R}^l$ such that $(\bar{x}, \lambda)$ also solves (2.12). Then $\lambda \in \bar{Y}$ and $\tilde{\Pi}(\lambda - \tilde{\lambda}) = 0$. As is easy to see, it follows that $(\bar{x}, \lambda)$ satisfies (2.10). But the solution of the latter system is unique, as established above. Hence, $(\bar{x}, \hat{\lambda})$ is the unique solution of (2.12).

To complete the proof, it remains to recall (2.8). In particular, the relation (2.9) follows from (2.11).   □

We next show how the linear operator satisfying (2.7), (2.8) can be constructed using the singular-value decomposition of $F'(\tilde{x})$ for any $\tilde{x}$ close enough to $\bar{x}$. The construction is based on the identification of the rank of $F'(\bar{x})$, using information at $(\tilde{x}, \tilde{\lambda})$ only, and a certain *error bound*.

Recall that the singular-value decomposition of $F'(x)$ is

$$F'(x) = U(x)\Sigma(x)(V(x))^{\mathrm{T}},$$

where $\Sigma(x)$ is the diagonal $l \times n$ matrix with singular values $\sigma_j(x)$ as the diagonal entries, $j = 1, \ldots, s$, $s = \min\{n, l\}$, $\sigma_1(x) \geq \sigma_2(x) \geq \cdots \geq \sigma_s(x) \geq 0$, while $U(x)$ and $V(x)$ are the orthogonal $l \times l$ and $n \times n$ matrices, respectively. Evidently, rank $F'(x)$ coincides with the number of positive singular values, and im $F'(x)$ is spanned by the columns of $U(x)$ corresponding to these values (the columns of $U(x)$ are called left singular vectors).

The error bound to be used below holds under the SOSC. For (2.1), the SOSC (1.4) takes the form

$$(2.13) \qquad \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle > 0 \quad \forall \xi \in \ker F'(\bar{x}) \setminus \{0\}.$$

By [21, Lemma 2], (2.13) implies that there exists a constant $M > 0$ such that

$$(2.14) \qquad \|x - \bar{x}\| + \mathrm{dist}(\lambda, \bar{Y}) \leq M\|\Phi(x, \lambda)\|$$

for all $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$.

PROPOSITION 2.2. *Let* $(\bar{x}, \bar{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$ *be a solution of* (2.2), *and assume that the error bound* (2.14) *holds with some* $M > 0$ *for all* $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ *close enough to* $(\bar{x}, \bar{\lambda})$. *Let* $\theta \in (0, 1)$.

Then for every $(\tilde{x}, \tilde{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l$ *close enough to* $(\bar{x}, \bar{\lambda})$, *the integer*

$$(2.15) \qquad r(\tilde{x}, \tilde{\lambda}) = \max\{0, \max\{j = 1, \ldots, s \mid \sigma_j(\tilde{x}) > \|\Phi(\tilde{x}, \tilde{\lambda})\|^{\theta}\}\}$$

coincides with rank $F'(\bar{x})$.

Proof. Let $\theta \in (0, 1)$ be fixed, and let $\bar{r} = \bar{r}(\bar{x}) = \mathrm{rank}\, F'(\bar{x})$. By the well-known results of the perturbation theory for linear operators [30, 31], if F is sufficiently smooth, then it holds that

$$|\sigma_j(\tilde{x}) - \sigma_j(\bar{x})| \leq \|F'(\tilde{x}) - F'(\bar{x})\|$$

for all $j = 1, \ldots, s$. Then for $(\tilde{x}, \tilde{\lambda})$ close enough to $(\bar{x}, \bar{\lambda})$, and for every $j = \bar{r} + 1, \ldots, s$, we have that

$$
\begin{aligned}
\sigma_j(\tilde{x}) &= |\sigma_j(\tilde{x}) - \sigma_j(\bar{x})| \\
&\leq \|F'(\tilde{x}) - F'(\bar{x})\| \\
&\leq N\|\tilde{x} - \bar{x}\| \\
&\leq MN\|\Phi(\tilde{x}, \tilde{\lambda})\| \\
&\leq \|\Phi(\tilde{x}, \tilde{\lambda})\|^\theta,
\end{aligned}
$$

where the second inequality follows from the local Lipschitz-continuity (with some modulus $N > 0$) of F', and the third inequality is by (2.14).

The right-hand side of the last inequality tends to zero as $(\tilde{x}, \tilde{\lambda}) \to (\bar{x}, \bar{\lambda})$ and hence, separates the first $\bar{r}$ singular values of $F'(\tilde{x})$ from the other singular values. In particular, it is easily seen that $r(\tilde{x}, \tilde{\lambda}) = \bar{r}$, where $r(\tilde{x}, \tilde{\lambda})$ is defined by (2.15). □

Proposition 2.2 provides all the information necessary for defining the operator $\Pi(\tilde{x}, \tilde{\lambda})$.

PROPOSITION 2.3. *Suppose that the assumptions of Proposition 2.2 hold. Let $\Pi(\tilde{x}, \tilde{\lambda})$ be defined as the orthogonal projector onto the linear subspace in $\mathbf{R}^l$ spanned by the last $l - r(\tilde{x}, \tilde{\lambda})$ left singular vectors of $F'(\tilde{x})$.*

Then conditions (2.7) and (2.8) hold.

Proof. Condition (2.7) follows immediately by the definition of $\Pi(\tilde{x}, \tilde{\lambda})$.

Condition (2.8) follows from the fact that for each $\tilde{\lambda}$ close enough to $\bar{\lambda}$, $\Pi(\cdot, \tilde{\lambda})$ is the so-called total projector of the 0-group of the symmetric operator $F'(\cdot)(F'(\cdot))^\mathrm{T}$ near the point $\bar{x}$, and from the continuity properties of total projectors [30]. (Singular vectors can behave quite chaotically, while total projectors are continuous, and even smooth, under natural assumptions.) □

Throughout the rest of the paper, we assume that $\Pi(\tilde{x}, \tilde{\lambda})$ is chosen according to the presented procedure. A linear operator $\Pi(\tilde{x}, \tilde{\lambda})$ possessing the required properties can be defined differently from the choice presented above [9]. We do not consider those other possibilities here in order to simplify the presentation. We also note that other identification functions could be used in (2.15), similar to techniques for identification of active constraints in [13].

We are now ready to state our local algorithm. The starting point $(\tilde{x}, \tilde{\lambda})$ is supplied as the input information for the algorithm, together with the problem data.

ALGORITHM 2.1. *Preliminary step.* Fix $\theta \in (0, 1)$.

Initialization step. Compute $r(\tilde{x}, \tilde{\lambda})$ according to (2.15). Define $\Pi(\tilde{x}, \tilde{\lambda})$ as the orthogonal projector onto the linear subspace in $\mathbf{R}^l$ spanned by the last $l - r(\tilde{x}, \tilde{\lambda})$ left singular vectors of $F'(\tilde{x})$.

Main step. Generate the sequence $\{(x^k, \lambda^k)\}$ by the Newton method applied to the system (2.6), starting from the initial point $(x^0, \lambda^0) = (\tilde{x}, \tilde{\lambda})$:

$$
(x^{k+1}, \lambda^{k+1}) = (x^k, \lambda^k) - (\Phi'_{\tilde{x}, \tilde{\lambda}}(x^k, \lambda^k))^{-1}\Phi_{\tilde{x}, \tilde{\lambda}}(x^k, \lambda^k), \quad k = 0, 1, \ldots.
$$

We note that although Algorithm 2.1 is a completely implementable local scheme, some issues still need to be resolved before its successful numerical realization can be possible. The most important question is how to couple it with some globally convergent method so that the switch to the fast locally convergent iterations can be done efficiently (for example, it is important to avoid an excessive number of singular-value decompositions when we are far from a solution, and so the local Newton steps are not

yet acceptable). Also, it should be verified in practice how large or small is the neighborhood of a solution where (2.15) gives correct identification, and how this depends on the choice of the identification function. We next provide an example showing that the region where (2.15) gives correct identification and where the resulting Newton method applied to (2.6) attains quadratic convergence can be quite large. Of course, this example is very simple and does not warrant any general conclusions.

Example 2.1. Let $n = 1$, $l = 1$, $f(x) = x^2/2$, $F(x) = x^2/2$.

Obviously, $\bar{x} = 0$ is a solution of (2.1) with $\bar{Y} = \mathbf{R}$. LICQ is violated, while SOSC (2.13) holds for any $\bar{\lambda} > -1$.

If we apply the standard Newton method to the Lagrangian system (2.2) starting at some $(x^0, \lambda^0) \in \mathbf{R} \times \mathbf{R}$, then the resulting sequence is given by

$$x^{k+1} = x^k/2, \quad \lambda^{k+1} = (\lambda^k - 1)/2,$$

which evidently converges to $(\bar{x}, \bar{\lambda}) = (0, -1)$, but only linearly. Note also that $\{\lambda^k\}$ is attracted by the multiplier for which SOSC does not hold.

Take any $\tilde{x} \in (0, 1)$ and, for example, $\tilde{\lambda} = 0$. Applying Algorithm 2.1, singular-value decomposition of $F'(\tilde{x})$ gives $\Sigma(\tilde{x}) = \tilde{x}$, $U(\tilde{x}) = 1$, $V(\tilde{x}) = 1$. Furthermore, for any $\theta \in (0, 1)$,

$$\|\Phi(\tilde{x}, \tilde{\lambda})\|^\theta = \left((1 + \tilde{\lambda})^2 \tilde{x}^2 + \tilde{x}^4/4\right)^{\theta/2} > \tilde{x}^\theta > \tilde{x} = \sigma_1(\tilde{x}).$$

Hence, (2.15) implies that

$$r(\tilde{x}, \tilde{\lambda}) = 0, \quad \Pi(\tilde{x}, \tilde{\lambda}) = 1,$$

which results in the modified optimality system

$$0 = \Phi_{\tilde{x}, \tilde{\lambda}}(x, \lambda) = (x + \lambda x, \, x^2/2 + \lambda).$$

This system has the unique solution $\hat{\lambda} = 0$ associated with $\bar{x}$, and the Jacobian of this system at $(\bar{x}, \hat{\lambda})$ is the identity matrix. Thus, Algorithm 2.1 generates a quadratically convergent sequence.

Note that in this example, (2.15) defines the correct modified optimality system for any $\tilde{x} \in (0, 1)$, which is the largest region for which it makes sense to speak about quadratic convergence to $(\bar{x}, \hat{\lambda})$.

We proceed to establish the convergence properties of Algorithm 2.1.

THEOREM 2.4. *Let $\bar{x} \in \mathbf{R}^n$ be a local solution of (2.1), and assume that $\bar{\lambda} \in \bar{Y}$ is such that the SOSC (2.13) is satisfied.*

Then for every neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{\lambda})$, and for every $(\tilde{x}, \tilde{\lambda}) \in \mathcal{U}$ close enough to $(\bar{x}, \bar{\lambda})$, Algorithm 2.1 is well defined and generates the sequence $\{(x^k, \lambda^k)\}$ such that

(a) $\{(x^k, \lambda^k)\}$ *remains in $\mathcal{U}$;*

(b) $\{(x^k, \lambda^k)\}$ *converges superlinearly/quadratically (depending on the smoothness of f and F) to $(\bar{x}, \hat{\lambda})$, where $\hat{\lambda}$ is defined in Proposition 2.1.*

Proof. Essentially, we have to show that the matrix

$$\Phi'_{\bar{x}, \bar{\lambda}}(\bar{x}, \bar{\lambda}) = \begin{pmatrix} \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) & (F'(\bar{x}))^{\mathrm{T}} \\ F'(\bar{x}) & \bar{\Pi} \end{pmatrix}$$

is nonsingular. Once this fact is established, the assertions follow from the theorem on small perturbations of nonsingular matrices and from the standard argument for

proving local convergence of the Newton method (taking into account Propositions 2.1 and 2.2, and [21, Lemma 2]).

Take an arbitrary pair $(\xi, \eta) \in \ker \Phi'_{\bar{x}, \bar{\lambda}}(\bar{x}, \bar{\lambda})$, i.e.,

$$(2.16) \qquad \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi + (F'(\bar{x}))^{\mathrm{T}}\eta = 0, \quad F'(\bar{x})\xi + \bar{\Pi}\eta = 0.$$

Recall that $F'(\bar{x})\xi$ and $\bar{\Pi}\eta$ belong to the complementary subspaces. Hence, the second equality can be decoupled into two equations:

$$(2.17) \qquad F'(\bar{x})\xi = 0, \quad \bar{\Pi}\eta = 0.$$

By the first equalities in (2.16) and (2.17),

$$\begin{aligned}
\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \, \xi \right\rangle &= -\langle (F'(\bar{x}))^{\mathrm{T}}\eta, \, \xi \rangle \\
&= -\langle \eta, \, F'(\bar{x})\xi \rangle \\
&= 0.
\end{aligned}$$

Taking into account again the first equality in (2.17), the SOSC (2.13) implies that $\xi = 0$. Now, from the first equality in (2.16), it follows that $\eta \in \ker(F'(\bar{x}))^{\mathrm{T}} = (\operatorname{im} F'(\bar{x}))^{\perp}$. On the other hand, the second equality in (2.17) implies the inclusion $\eta \in \ker \bar{\Pi} = \operatorname{im} F'(\bar{x})$. We conclude that $\eta = 0$. We thus established that $\Phi'_{\bar{x}, \bar{\lambda}}(\bar{x}, \bar{\lambda})$ is nonsingular. □

Remark 2.1. A more accurate investigation of the iteration of Algorithm 2.1 leads to the following estimate of the rate of convergence (under the appropriate smoothness assumptions):

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

3. Mixed constraints. We now turn our attention to the general problem (1.1). As before, we assume that we have good initial approximations to a local solution $\bar{x}$ of (1.1) and some $(\bar{\lambda}, \bar{\mu}) \in \bar{Y}$ satisfying the SOSC (1.4).

3.1. Using identification of active constraints. As is well known, the KKT system (1.2) can be written in the form

$$(3.1) \qquad \Psi(x, \, \lambda, \, \mu) = 0,$$

where $\Psi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$,

$$\Psi(x, \, \lambda, \, \mu) = \left(\frac{\partial L}{\partial x}(x, \, \lambda, \, \mu), \, F(x), \, \min\{\mu, \, -G(x)\} \right),$$

with the minimum operation taken componentwise. According to [21, Lemma 2] (see also [17, Lemma 5]), the residual of (3.1) can be used to estimate the distance from a given $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ to the set $\{\bar{x}\} \times \bar{Y}$. Specifically, under our assumptions there exists a constant $M > 0$ such that

$$(3.2) \qquad \|x - \bar{x}\| + \operatorname{dist}((\lambda, \, \mu), \, \bar{Y}) \leq M\|\Psi(x, \, \lambda, \, \mu)\|$$

for all $(x, \, \lambda, \, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. This error bound can be further used to identify the constraints active at $\bar{x}$, following the approach of [13].

PROPOSITION 3.1. *Let $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be a solution of (1.2), and assume that the error bound (3.2) holds with some $M > 0$ for all $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. Let $\tau \in (0, 1)$.*

Then for every $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, the index set

$$(3.3) \qquad I(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) = \{i = 1, \ldots, m \mid -G_i(\tilde{x}) \leq \|\Psi(\tilde{x}, \tilde{\lambda}, \tilde{\mu})\|^\tau\}$$

coincides with $\bar{I}$.

Remark 3.1. As remarked by a referee, there exist some conflicting opinions about the value of the identification technique proposed in [13] (the possible criticism is that the neighborhood of correct identification is difficult to characterize, and in some cases can be rather small). That said, satisfactory computational experience with some algorithms which employ this technique is reported in [29, 10], and it is also used, for example, in [14, 43] (although there are no numerical results for the latter two references).

Once the set $I = \bar{I}$ is identified, we can consider instead of (1.1) the purely equality-constrained problem

$$(3.4) \qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & F(x) = 0, \ G_I(x) = 0. \end{array}$$

Obviously, violation at $\bar{x}$ of LICQ for (1.1) means that (3.4) also does not satisfy LICQ. To apply the approach developed in section 2, we need to make sure that SOSC for (3.4) holds whenever we have SOSC for the original problem (1.1). It is easy to see that this is indeed the case.

PROPOSITION 3.2. *Let $\bar{x} \in \mathbf{R}^n$ be a local solution of (1.1), and assume that $(\bar{\lambda}, \bar{\mu}) \in \bar{Y}$ is such that the SOSC (1.4) is satisfied.*

Then $\bar{x}$ is a local solution of problem (3.4) with $I = \bar{I}$, and $(\bar{\lambda}, \bar{\mu}_{\bar{I}})$ is a Lagrange multiplier associated with this solution, i.e.,

$$(3.5) \qquad f'(\bar{x}) + (F'(\bar{x}))^{\mathrm{T}} \bar{\lambda} + (G'_{\bar{I}}(\bar{x}))^{\mathrm{T}} \bar{\mu}_{\bar{I}} = 0.$$

Furthermore, SOSC for (3.4) holds with this multiplier:

$$(3.6) \qquad \begin{array}{l} \langle f''(\bar{x})\xi, \xi \rangle + \langle \bar{\lambda}, F''(\bar{x})[\xi, \xi] \rangle + \langle \bar{\mu}_{\bar{I}}, G''_{\bar{I}}(\bar{x})[\xi, \xi] \rangle > 0 \\[4pt] \forall \, \xi \in (\ker F'(\bar{x}) \cap \ker G'_{\bar{I}}(\bar{x})) \setminus \{0\}. \end{array}$$

Proof. The equality (3.5) obviously follows from the definition of $\bar{Y}$. Furthermore, since $\bar{\mu}_i = 0$ for all $i \notin \bar{I}$, the left-most expression in (3.6) equals $\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \rangle$. Since we also have that $\ker F'(\bar{x}) \cap \ker G'_{\bar{I}}(\bar{x}) \subset \bar{K}$, it follows that (1.4) implies (3.6). □

We complete this section with a formal statement of the local algorithm for mixed-constrained problems. The starting point $(\tilde{x}, \tilde{\lambda}, \tilde{\mu})$ is supplied as the input information for the algorithm.

ALGORITHM 3.1. *Preliminary step.* Fix $\tau \in (0, 1)$.

Initialization step. Define the index set $I(\tilde{x}, \tilde{\lambda}, \tilde{\mu})$ according to (3.3).

Main step. Apply Algorithm 2.1 to problem (3.4) with $I = I(\tilde{x}, \tilde{\lambda}, \tilde{\mu})$, using $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}_I)$ as the starting point.

The following theorem is a direct consequence of Theorem 2.4 and Propositions 3.1 and 3.2.

THEOREM 3.3. *Let $\bar{x} \in \mathbf{R}^n$ be a local solution of (1.1), and assume that $(\bar{\lambda}, \bar{\mu}) \in \bar{Y}$ is such that the SOSC (1.4) is satisfied.*

Then for every $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *close enough to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$, *Algorithm* 3.1 *is well defined and generates a sequence* $\{(x^k, \lambda^k, \mu_{\tilde{I}}^k)\}$, *which converges superlinearly/ quadratically (depending on the smoothness of* f, F, *and* G*) to* $(\bar{x}, \hat{\lambda}, \hat{\mu}_{\bar{I}})$, *with some* $\hat{\lambda} \in \mathbf{R}^l$ *and* $\hat{\mu}_{\bar{I}} \in \mathbf{R}^{|\bar{I}|}$ *such that*

$$(\hat{\lambda}, \hat{\mu}_{\bar{I}}) \to (\bar{\lambda}, \bar{\mu}_{\bar{I}}) \quad as \ \tilde{x} \to \bar{x}, \ \tilde{\lambda} \to \bar{\lambda}, \ \tilde{\mu} \to \bar{\mu}.$$

3.2. Reformulation with slack variables. Instead of identifying active constraints, one could transform (1.1) into the purely equality-constrained problem using slack variables:

(3.7)
$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & F(x) = 0, \ g_i(x) + \sigma_i^2 = 0, \ i = 1, \ldots, m, \end{array}$$

where $g_1, \ldots, g_m$ are the components of G. As is well known, local solutions of (1.1) and (3.7) are in correspondence with each other: $\bar{x}$ is a local solution of (1.1) if and only if $(\bar{x}, \bar{\sigma})$ with $\bar{\sigma} = (\sqrt{-g_1(\bar{x})}, \ldots, \sqrt{-g_m(\bar{x})})$ is a well-defined local solution of (3.7).

This approach is quite attractive because of its universality and simplicity. However, it requires stronger assumptions. Specifically, it is known that to guarantee SOSC for (3.7) at $(\bar{x}, \bar{\sigma})$ with multiplier $(\bar{\lambda}, \bar{\mu})$, one needs not only SOSC for the original problem (1.1), but also the strict complementarity condition (SCC), i.e., $\bar{\mu}_i > 0$ for all $i \in \bar{I}$ (see [7, Theorem 1.32]). We note that the SCC is regarded as a somewhat restrictive (and hence, undesirable) assumption.

4. Application to MPCC. An interesting application of the approach developed in this paper is concerned with MPCC, the mathematical programs with complementarity constraints. Here, we consider the problem of the form

(4.1)
$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & F(x) = 0, \ G(x) \leq 0, \\ & H_1(x) \geq 0, \ H_2(x) \geq 0, \ \langle H_1(x), H_2(x) \rangle = 0, \end{array}$$

where $H_1, H_2 : \mathbf{R}^n \to \mathbf{R}^s$ are smooth mappings. Of course, (4.1) can be considered as a usual mixed-constrained optimization problem. Recall, however, that MFCQ is violated at every feasible point of this problem (since there exists no feasible point strictly satisfying the inequality constraints). Therefore, this class of problems is of particular interest in the context of our development.

The Lagrangian for (4.1) is given by

$$\begin{aligned} L(x, \lambda, \mu, \mu^1, \mu^2, \nu) = {} & f(x) + \langle \lambda, F(x) \rangle + \langle \mu, G(x) \rangle \\ & - \langle \mu^1, H_1(x) \rangle - \langle \mu^2, H_2(x) \rangle + \nu \langle H_1(x), H_2(x) \rangle, \\ & x \in \mathbf{R}^n, \ \lambda \in \mathbf{R}^l, \ \mu \in \mathbf{R}^m, \ \mu^1, \mu^2 \in \mathbf{R}^s, \ \nu \in \mathbf{R}, \end{aligned}$$

and the set $\bar{Y}$ associated with a local solution $\bar{x}$ of (4.1) consists of tuples $(\lambda, \mu, \mu^1, \mu^2, \nu) \in \mathbf{R}^l \times \mathbf{R}^m \times \mathbf{R}^s \times \mathbf{R}^s \times \mathbf{R}$ such that

$$(F'(\bar{x}))^{\mathrm{T}}\lambda + (G'(\bar{x}))^{\mathrm{T}}\mu - (H_1'(\bar{x}))^{\mathrm{T}}(\mu^1 - \nu H_2(\bar{x})) - (H_2'(\bar{x}))^{\mathrm{T}}(\mu^2 - \nu H_1(\bar{x}))) = -f'(\bar{x}),$$
$$\mu \geq 0, \ \langle \mu, G(\bar{x}) \rangle = 0,$$
$$\mu^1 \geq 0, \ \langle \mu^1, H_1(\bar{x}) \rangle = 0, \ \mu^2 \geq 0, \ \langle \mu^2, H_2(\bar{x}) \rangle = 0.$$

It is easy to see that $\bar{x}$ is a stationary point in (4.1) in the sense that $\bar{Y} \neq \emptyset$ if and only if it is strongly stationary in the terminology of [36]. Moreover, the multipliers corresponding to these stationarity concepts are related by explicit formulas (see [18, Proposition 4.1]). We note that strong stationarity is regarded as a reasonable stationarity concept in the context of Newton-type methods for MPCC.

As is easy to see,

$$(4.2) \qquad \bar{K} = \left\{ \xi \in \ker F'(\bar{x}) \; \middle| \; \begin{array}{l} \langle f'(\bar{x}), \xi \rangle \leq 0, \; G'_I(\bar{x})\xi \leq 0, \\ (H_1)'_{I_1 \setminus I_2}(\bar{x})\xi = 0, \; (H_2)'_{I_2 \setminus I_1}(\bar{x})\xi = 0, \\ (H_1)'_{I_0}(\bar{x})\xi \geq 0, \; (H_2)'_{I_0}(\bar{x})\xi \geq 0 \end{array} \right\},$$

where

$$I = \{i = 1, \ldots, m \mid G_i(\bar{x}) = 0\},$$

$$I_1 = \{i = 1, \ldots, s \mid (H_1)_i(\bar{x}) = 0\}, \quad I_2 = \{i = 1, \ldots, s \mid (H_2)_i(\bar{x}) = 0\}, \quad I_0 = I_1 \cap I_2.$$

Note that necessarily $I_1 \cup I_2 = \{1, \ldots, s\}$. For (4.1), the SOSC (1.4) takes the form

$$(4.3) \qquad \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu}, \bar{\mu}^1, \bar{\mu}^2, \bar{\nu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in \bar{K} \setminus \{0\},$$

with $\bar{K}$ defined in (4.2) and a tuple of multipliers $(\bar{\lambda}, \bar{\mu}, \bar{\mu}^1, \bar{\mu}^2, \bar{\nu}) \in \bar{Y}$. This condition is exactly the same as SOSC for MPCC in [18, Definition 3.4] if we take into account the relation between multipliers corresponding to the two different (but equivalent) stationarity concepts. Recall also that the MPCC-LICQ [36, 38, 18] consists of saying that

$$(4.4) \qquad \text{the rows of } F'(\bar{x}), \; G'_I(\bar{x}), \; (H_1)'_{I_1}(\bar{x}), \; (H_2)'_{I_2}(\bar{x}) \text{ are linearly independent.}$$

If we have a good approximation to $\bar{x}$ and $(\bar{\lambda}, \bar{\mu}, \bar{\mu}^1, \bar{\mu}^2, \bar{\nu}) \in \bar{Y}$ satisfying (4.3), then Algorithm 3.1 can be directly applied to (4.1), and its local convergence is characterized by Theorem 3.3. Note that no CQ is needed, as compared to other methods for MPECs which typically need special CQs, such as MPCC-LICQ (4.4). We comment more on this issue in section 5.

We complete this section with two examples that illustrate Algorithm 3.1 when applied to MPCC. The first example is rather simple, and, in particular, it satisfies the MPCC-LICQ (4.4). Together with the SOSC (4.3), this means that other methods (such as slightly modified SQP) can also be used in that case. We include this example mainly because it is standard in the MPEC literature, e.g., [36], and because it is easier to follow its development. The second example is a modification of [25, Example 4.2]. Here, MPCC-LICQ is not satisfied, and so this better serves to illustrate the power of the approach presented in this paper.

Example 4.1. Let $n = 2$, $l = m = 0$, $s = 1$, $f(x) = a_1 x_1 + a_2 x_2$, where $a_1, a_2 > 0$, $H_1(x) = x_1$, $H_2(x) = x_2$. Obviously, $\bar{x} = 0$ is a solution of (4.1), and

$$\bar{Y} = \{(\bar{\mu}^1, \bar{\mu}^2, \nu) \mid \bar{\mu}^1 = a_1, \bar{\mu}^2 = a_2, \nu \in \mathbf{R}\}.$$

We further have that $I_0 = I_1 = I_2 = \{1\}$, and

$$\bar{K} = \{\xi \in \mathbf{R}^2 \mid a_1\xi_1 + a_2\xi_2 \leq 0, \xi_1 \geq 0, \xi_2 \geq 0\} = \{0\}.$$

Hence, the SOSC (4.3) is satisfied automatically.

On the initialization step of Algorithm 3.1, for $(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}^2, \tilde{\nu})$ close enough to $(\bar{x}, a_1, a_2, \bar{\nu})$ with some $\bar{\nu} \in \mathbf{R}$, it is correctly identified that both inequality constraints are active at the solution; that is, $\bar{I} = I(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}^2, \tilde{\nu}) = \{1, 2\}$. Hence, problem (3.4) takes the form

$$\begin{aligned} \text{minimize} \quad & a_1 x_1 + a_2 x_2 \\ \text{subject to} \quad & \tilde{F}(x) = (x_1,\, x_2,\, x_1 x_2) = 0. \end{aligned}$$

As is easy to see, $\bar{\Pi}$ is the orthogonal projector onto

$$\ker(\tilde{F}'(\bar{x}))^{\mathrm{T}} = \ker \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \{y \in \mathbf{R}^3 \mid y_1 = y_2 = 0\},$$

i.e.,

$$\bar{\Pi} y = (0,\, 0,\, y_3), \quad y \in \mathbf{R}^3.$$

On the initialization step of Algorithm 2.1, it is correctly identified that $r(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}^2, \tilde{\nu}) = 2$. Then $\Pi(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}^2, \tilde{\nu})$ is taken as some operator approximating $\bar{\Pi}$.

Suppose, for a moment, that $\Pi(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}^2, \tilde{\nu}) = \bar{\Pi}$. Then the corresponding regularized system (2.6) takes the form

$$\begin{aligned} a_1 - \mu^1 + \nu x_2 &= 0, \\ a_2 - \mu^2 + \nu x_1 &= 0, \\ x_1 = 0, \; x_2 &= 0, \\ x_1 x_2 + \nu - \tilde{\nu} &= 0. \end{aligned}$$

Evidently, this system has a solution of the form $(\bar{x}, a_1, a_2, \tilde{\nu})$. The Jacobian of this system at this solution is the matrix

$$\begin{pmatrix} 0 & \tilde{\nu} & -1 & 0 & 0 \\ \tilde{\nu} & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

which is nonsingular for any choice of $\tilde{\nu}$. Hence, the Newton method applied to this system is locally quadratically convergent to this solution.

Obviously, if $\Pi(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}^2, \tilde{\nu})$ is close enough to $\bar{\Pi}$, then we obtain a small perturbation of the above system of equations, and thus the same conclusions hold.

We next present an example where MPCC-LICQ (4.4) does not hold. The authors are not aware of any approaches which lead to a local algorithm with superlinear convergence for this example, except for the one presented here and in [43].

Example 4.2. Let $n = 3$, $l = m = 0$, $s = 2$, $f(x) = x_1 + x_2 + x_1^2 + x_2^2 + (x_3 - 1)^2$, $H_1(x) = (x_1 + x_2,\, x_1^2 - (x_3 - 1)^2)$, $H_2(x) = (x_2,\, x_3)$.

As is easy to see, $\bar{x} = (0, 0, 1)$ is the solution of (4.1). We also have that $I_1 = \{1, 2\}$, $I_2 = \{1\}$, $I_0 = \{1\}$.

Note that MPCC-LICQ (4.4) is clearly violated, since

$$H_1'(\bar{x}) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In fact, any other "standard" MPEC CQ is violated in this example. On the other hand, the constraints of this problem satisfy the special 2-regularity condition [25], and thus the corresponding (directional) optimality conditions hold. For some critical directions, those conditions can be reduced to the form considered in this section. (Of course, one can just verify the latter directly.) In particular, we have that

$$\bar{Y} = \{(\mu^1, \bar{\mu}^2, \nu) \mid \mu_1^1 = 1, \mu_2^1 \geq 0, \bar{\mu}^2 = 0, \nu \in \mathbf{R}\}.$$

Take $\bar{\mu}^1 = (1, 0)$, $\bar{\nu} = 0$. It can be verified that

$$\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\mu}^1, \bar{\mu}^2, \bar{\nu}) = 2E,$$

where E is the identity matrix. It follows that the SOSC (4.3) holds for this solution.

Suppose we have a point $(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}^2, \tilde{\nu})$ close enough to this $(\bar{x}, \bar{\mu}^1, \bar{\mu}^2, \bar{\nu})$. Then on the initialization step of Algorithm 3.1, it is correctly identified that $\bar{I} = I(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}^2, \tilde{\nu}) = \{1, 2, 3\}$. Therefore, problem (3.4) takes the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \tilde{F}(x) = 0, \end{array}$$

where

$$\tilde{F}(x) = \begin{pmatrix} x_1 + x_2 \\ x_1^2 - (x_3 - 1)^2 \\ x_2 \\ x_2(x_1 + x_2) + x_3(x_1^2 - (x_3 - 1)^2) \end{pmatrix}.$$

We further have that $\bar{\Pi}$ is the orthogonal projector onto

$$\ker(\tilde{F}'(\bar{x}))^{\mathrm{T}} = \ker \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \{y \in \mathbf{R}^4 \mid y_1 = y_3 = 0\},$$

i.e.,

$$\bar{\Pi}y = (0, y_2, 0, y_4), \quad y \in \mathbf{R}^4.$$

On the initialization step of Algorithm 2.1, it is correctly identified that $r(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}_1^2, \tilde{\nu}) = 2$. Then $\Pi(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}_1^2, \tilde{\nu})$ is taken as some operator approximating $\bar{\Pi}$. Suppose for a moment that $\Pi(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}_1^2, \tilde{\nu}) = \bar{\Pi}$. Then the corresponding regularized system (2.6) takes the form

$$1 + 2x_1 - \mu_1^1 - 2\mu_2^1 x_1 + \nu(x_2 + 2x_1 x_3) = 0,$$

$$1 + 2x_2 - \mu_1^1 - \mu_1^2 + \nu(x_1 + 2x_2) = 0,$$

$$2(x_3 - 1) + 2\mu_2^1(x_3 - 1) + \nu(x_1^2 - 3x_3^2 + 4x_3 - 1) = 0,$$

$$x_1 + x_2 = 0,$$

$$x_1^2 - (x_3 - 1)^2 + \mu_2^1 - \tilde{\mu}_2^1 = 0,$$

$$x_2 = 0,$$

$$x_2(x_1 + x_2) + x_3(x_1^2 - (x_3 - 1)^2) + \nu - \tilde{\nu} = 0.$$

It can be easily seen that this system has a solution $(\bar{x}, (1, \tilde{\mu}_2^1), 0, \tilde{\nu})$, and the Jacobian of this system at this solution is

$$
\begin{pmatrix}
2(1 - \tilde{\mu}_2^1 + \tilde{\nu}) & \tilde{\nu} & 0 & -1 & 0 & 0 & 0 \\
\tilde{\nu} & 2(1 + \tilde{\nu}) & 0 & -1 & 0 & -1 & 0 \\
0 & 0 & 2(1 + \tilde{\mu}_2^1 - \tilde{\nu}) & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},
$$

which is nonsingular (in fact, even well conditioned) for all $\tilde{\mu}_2^1$ and $\tilde{\nu}$ close enough to $\bar{\mu}_2^1$ and $\bar{\nu}$, respectively. Hence, the Newton method applied to this system is locally quadratically convergent to this solution.

If $\Pi(\tilde{x}, \tilde{\mu}^1, \tilde{\mu}_1^2, \tilde{\nu})$ is close enough to $\bar{\bar{\Pi}}$, then we obtain a small perturbation of the above system of equations, and thus the same conclusions hold.

5. Comparisons with other approaches. In this section, we briefly review some recent work on Newton-type methods for optimization problems, where CQs are relaxed or removed. We note that a direct complete comparison of different approaches is in most cases impossible. Below, we discuss some advantages and disadvantages, making comparisons where possible.

First, we comment on the standard SQP algorithm. As already mentioned above, in the case of equality constraints, the superlinear convergence of SQP requires LICQ and SOSC. By contrast, our approach needs SOSC only. Of course, this does not come without a computational price—singular-value decomposition of the Jacobian of the constraints is used to set up the modified optimality system. Similarly, for mixed constraints, the SQP method needs the strict MFCQ (which is equivalent to the uniqueness of the multiplier) and SOSC, while for our proposal based on the identification of active constraints, SOSC is sufficient. Additionally, subproblems in our method are systems of linear equations, which are simpler than quadratic programs in SQP. On the other hand, it should be kept in mind that SQP is certainly easier to globalize, as some well-developed globalization strategies are readily available.

SQP with nonunique multipliers has been analyzed in [1, 4], which assumes MFCQ and some second-order sufficiency conditions. Global convergence based on linesearch using a nonsmooth penalty function is established. The algorithm is primal: No approximation of the Hessian of the Lagrangian function is used. Thus, the rate of convergence is only linear, of course.

The so-called stabilized SQP and related algorithms [40, 20, 16, 17, 41, 42, 33] employ regularization techniques, with smart choices of regularization parameters based on error bounds for the primal-dual solution of the problem. Most of the cited literature deals with the case of inequality constraints only, but perhaps the analysis could be extended to mixed constraints. In [42, Theorem 8], local superlinear convergence of an algorithm of this kind to the set $\{\bar{x}\} \times \bar{Y}$ is established under MFCQ and the SOSC (2.13) (the latter is supposed to be satisfied for all $\bar{\mu} \in \bar{Y}$). In [17, Theorem 8], local superlinear convergence to $(\bar{x}, \hat{\mu})$ for some $\hat{\mu} \in \bar{Y}$ close to $\bar{\mu}$ is established under the strong SOSC, that is, (2.13) with

$$
\bar{K} = \{\xi \in \mathbf{R}^n \mid G'_{\bar{I}_+}(\bar{x})\xi = 0\}, \quad \bar{I}_+ = \{i \in I \mid \bar{\mu}_i > 0\}.
$$

In [33], the stabilized version of SQP [40] is combined with active-set strategies, and convergence is established under a certain relaxed CQ assumption and the strong

SOSC. To our knowledge, [20, 17, 33] are the only papers on stabilized SQP, where MFCQ is not assumed to be satisfied. Compared, e.g., to [42], we do not assume any CQ. Compared to [20, 17, 33], we need a weaker SOSC. Thus, conditions for local superlinear convergence in our approach are weaker than those for stabilized versions of SQP. Additionally, each iteration of a stabilized SQP method typically requires more than just solving a QP (in [33] linear equations are solved instead of QPs, but the number of equations to be solved per iteration can be more than one). It should also be noted that the stabilized SQP shares the common (at least conceptual) drawback of regularization methods: Its subproblems become more ill-conditioned (and more difficult to solve) as the method approaches an irregular solution and the regularization parameter tends to zero. Our approach is free from this drawback. On the other hand, comments above concerning globalization still apply.

Another approach which has recently received some attention is the so-called sequential quadratically constrained quadratic programming (SQCQP), which employs subproblems based on the quadratic approximation of both the objective function and the constraints. The overall behavior may be quite favorable, provided an effective method to solve subproblems is available. In [5], for purely inequality-constrained problems, it is shown that the SQCQP algorithm is locally superlinearly convergent to a local solution satisfying MFCQ and the quadratic growth condition. (Under MFCQ, the latter is equivalent to a weaker form of SOSC, that is, (2.13) with $\bar{\mu}$ replaced by some $\mu(\xi) \in \bar{Y}$.) Globalization issues are considered in [19, 39]. Of course, SQCQP subproblems are more complex than quadratic programs (as in SQP) or linear systems of equations (as in our approach). In the convex case, SQCQP subproblems can be solved rather efficiently (after formulating them as second-order cone programs) by the interior points techniques. But in the nonconvex case, the situation is not so clear. Also, we note that an extension to the nonlinear equality-constrained problems seems problematic, as the subproblems would very likely be infeasible (even locally).

We next mention [2], which suggests an approach for transforming a problem with violated MFCQ into a problem satisfying MFCQ. It should be noted, though, that this approach relies on some strong "quadratic growth condition," i.e., an estimate of the distance to the given solution from *all* (not only feasible) points close enough to this solution. Under MFCQ, this is equivalent to the standard quadratic growth condition, but without MFCQ, it is much stronger, and some regularity of constraints is implicitly subsumed in this condition.

Very recently, another method for solving degenerate problems has been proposed in [43]. In the case of inequality or mixed constraints, it reduces the problem to the purely equality-constrained case using the same technique to identify active constraints that is employed in our approach. Local superlinear rate of convergence is obtained under essentially the same assumptions as ours (but unlike in our method, the quadratic rate is not achieved). One system of linear equations is solved at each iteration, but the system itself is different from ours. The advantage of [43] when compared to our approach is that singular-value decomposition is not needed and the linear system to be solved is closer in structure to systems that arise in the SQP or primal-dual interior point methods (in particular, it preserves sparsity). The disadvantage is that it involves a regularization parameter which tends to zero with convergence, which means that close to the solution subproblems are ill-conditioned (in the absence of CQs).

We now turn our attention to Newton-type methods for MPEC. Inevitable violation of MFCQ in (4.1) leads to numerous difficulties in theoretical justification of those methods. One of the most significant difficulties is the following: SQP subproblems

can be simply infeasible arbitrarily close to a solution. In [3], the special elastic mode is assumed to be used by a QP solver in order to overcome the infeasibility problem. On the other hand, it was recognized recently that even the behavior of rather basic SQP algorithms (without complex modifications) turns out to be quite favorable in many examples of MPCC. In [18], the authors give a theoretical explanation for this phenomenon. They prove local superlinear convergence under the MPCC-LICQ (4.4), the SOSC (4.3) with $\bar{K}$ defined in (4.2), and some additional assumptions on the multipliers and on the properties of a QP solver. The problem is supposed to be reformulated using slacks, and the authors *assume* that SQP subproblems are feasible (unless the starting point satisfies an additional complementarity condition). It should be pointed out, however, that under the assumptions of [18], SQP subproblems are consistent for some important special cases. For example, this is the case when there are only complementarity constraints in (4.1), i.e., there are no additional equality and inequality constraints (see [18, Lemma 6.1]). In the approach described in section 4, we need only strong stationarity and the SOSC (4.3). We do not assume MPCC-LICQ, and our approach is QP-free. That said, comments on globalization again apply.

To complete this discussion, we briefly examine the examples presented in [18]. All these examples are taken from *MacMPEC*, a collection of MPECs [32]. In `jr*` and `s14`, the SOSC (4.3) holds with any choice of the multipliers. In `scholtes4`, strong stationarity does not hold, i.e., $\bar{Y} = \emptyset$. Finally, in `ralph2`, the SOSC (4.3) holds with the appropriate choice of multipliers. Similarly to section 4, it can be shown that our approach is applicable to all those examples, except for `scholtes4` (which also cannot be efficiently handled by any of the other known methods).

Summarizing, our approach requires weaker assumptions for local superlinear convergence than the alternatives, *where any exist* (the exception to this comment is the subsequent paper [43], which needs the same assumptions). Of course, this does not come for free. The computational price to pay is the singular-value decomposition of the Jacobian of the constraints. Arguably, this is not a very serious drawback, because (in theory) this computation has to be done only once, at the initialization stage of the local method. And it should be kept in mind that once it is done, the subproblems to be solved are just linear equations, which are simpler than in (most) alternatives. A more serious drawback is that the presented local algorithm does not come with a *natural* globalization procedure. The subject of globalization is an issue of current research. At this time, we can prove that the local algorithm can be globalized using the so-called hybrid strategy, in the spirit of methods in the complementarity literature, e.g., [11, 15, 12, 28]. Of course, within the hybrid strategy in general, the singular-value decomposition would need to be computed more than once, until the identification in (2.15) and (3.3) becomes correct. This is a clear numerical disadvantage. We are looking for other (better) alternatives. As already mentioned above, a number of questions still need to be resolved before the proposed approach can be implemented and its numerical potential evaluated.

REFERENCES

[1] M. ANITESCU, *Degenerate nonlinear programming with a quadratic growth condition*, SIAM J. Optim., 10 (2000), pp. 1116–1135.

[2] M. ANITESCU, *Nonlinear Programs with Unbounded Lagrange Multiplier Sets*, Preprint ANL/MCS-P796-0200, Mathematics and Computer Science Division, Argonne National

Laboratory, Argonne, IL, 2000.

[3] M. Anitescu, *On Solving Mathematical Programs with Complementarity Constraints as Nonlinear Programs*, Preprint ANL/MCS-P864-1200, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, 2000.

[4] M. Anitescu, *On the rate of convergence of sequential quadratic programming with nondifferentiable exact penalty function in the presence of constraint degeneracy*, Math. Program., 92 (2002), pp. 359–386.

[5] M. Anitescu, *A superlinearly convergent sequential quadratically constrained quadratic programming algorithm for degenerate nonlinear programming*, SIAM J. Optim., 12 (2002), pp. 949–978.

[6] A. V. Arutyunov, *Optimality Conditions: Abnormal and Degenerate Problems*, Kluwer Academic Publishers, Dordrecht, 2000.

[7] D. P. Bertsekas, *Constraint Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[8] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, Belmont, MA, 1999.

[9] O. A. Brezhneva and A. F. Izmailov, *Construction of defining systems for finding singular solutions of nonlinear equations*, Comput. Math. Math. Phys., 42 (2002) pp. 8–19.

[10] A. N. Daryina, A. F. Izmailov, and M. V. Solodov, *A class of active-set Newton methods for mixed complementarity problems*, SIAM J. Optim., to appear.

[11] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Program., 75 (1996), pp. 407–439.

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

[13] F. Facchinei, A. Fischer, and C. Kanzow, *On the accurate identification of active constraints*, SIAM J. Optim., 9 (1999), pp. 14–32.

[14] F. Facchinei and C. Lazzari, *Local feasible QP-free algorithms for the constrained minimization of SC^1 functions*, J. Optim. Theory Appl., 119 (2003), pp. 218–316.

[15] M. C. Ferris, C. Kanzow, and T. S. Munson, *Feasible descent algorithms for mixed complementarity problems*, Math. Program., 86 (1999), pp. 475–497.

[16] A. Fischer, *Modified Wilson's method for nonlinear programs with nonunique multipliers*, Math. Oper. Res., 24 (1999), pp. 699–727.

[17] A. Fischer, *Local behaviour of an iterative framework for generalized equations with nonisolated solutions*, Math. Program., 94 (2002), pp. 91–124.

[18] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes, *Local Convergence of SQP Methods for Mathematical Programs with Equilibrium Constraints*, Numerical Analysis Report NA/209, Department of Mathematics, University of Dundee, Scotland, UK, 2002.

[19] M. Fukushima, Z.-Q. Luo, and P. Tseng, *A sequential quadratically constrained quadratic programming method for differentiable convex minimization*, SIAM J. Optim., 13 (2003), pp. 1098–1119.

[20] W. W. Hager, *Stabilized sequential quadratic programming*, Comput. Optim. Appl., 12 (1999), pp. 253–273.

[21] W. W. Hager and M. S. Gowda, *Stability in the presence of degeneracy and error estimation*, Math. Program., 85 (1999), pp. 181–192.

[22] A. F. Izmailov, *Lagrange methods for finding degenerate solutions of conditional extremum problems*, Comput. Math. Math. Phys., 36 (1996), pp. 423–429.

[23] A. F. Izmailov and M. V. Solodov, *Optimality conditions for irregular inequality-constrained problems*, SIAM J. Control Optim., 40 (2001), pp. 1280–1295.

[24] 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.

[25] A. F. Izmailov and M. V. Solodov, *Complementarity constraint qualification via the theory of 2-regularity*, SIAM J. Optim., 13 (2002), pp. 368–385.

[26] 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.

[27] A. F. Izmailov and A. A. Tretyakov, *2-Regular Solutions of Nonlinear Problems, Theory and Numerical Methods*, Fizmatlit, Moscow, 1999 (in Russian).

[28] C. Kanzow, *Strictly feasible equation-based methods for mixed complementarity problems*, Numer. Math., 89 (2001), pp. 135–160.

[29] C. Kanzow and H. D. Qi, *A QP-free constrained Newton-type method for variational inequality problems*, Math. Program., 85 (1999), pp. 81–106.

[30] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, 1984.
[31] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice–Hall, Englewood Cliffs, NJ, 1974.
[32] S. Leyffer, *MacMPEC: AMPL Collection of MPECs*, Technical report, 2000.
[33] D.-H. Li and L. Qi, *A Stabilized SQP Method via Linear Equations*, Applied mathematics technical report AMR00/5, The University of New South Wales, 2000.
[34] Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical programs with equilibrium constraints*, Cambridge University Press, Cambridge, UK, 1996.
[35] D. Ralph and S. J. Wright, *Superlinear convergence of an interior-point method despite dependent constraints*, Math. Oper. Res., 25 (2000), pp. 179–194.
[36] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1–22.
[37] S. Scholtes and M. Stöhr, *Exact penalization of mathematical programs with equilibrium constraints*, SIAM J. Control Optim., 37 (1999), pp. 617–652.
[38] S. Scholtes and M. Stöhr, *How stringent is the linear independence assumption for mathematical programs with complementarity constraints?*, Math. Oper. Res., 26 (2001), pp. 851–863.
[39] M. V. Solodov, *On the sequential quadratically constrained quadratic programming methods*, Math. Oper. Res., 29 (2004), pp. 64–79.
[40] S. J. Wright, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Comput. Optim. Appl., 11 (1998), pp. 253–275.
[41] S. J. Wright, *Modifying SQP for degenerate problems*, SIAM J. Optim., 13 (2002), pp. 470–497.
[42] S. J. Wright, *Constraint identification and algorithm stabilization for degenerate nonlinear programs*, Math. Program., 95 (2003), pp. 137–160.
[43] S. J. Wright, *An Algorithm for Degenerate Nonlinear Programming with Rapid Local Convergence*, Optimization technical report 03-02, Computer Sciences Department, University of Wisconsin, Madison, 2003.