Preprint ANL/MCS-P864-1200

ON SOLVING MATHEMATICAL PROGRAMS WITH COMPLEMENTARITY CONSTRAINTS AS NONLINEAR PROGRAMS

MIHAI ANITESCU*

Abstract. We investigate the possibility of solving mathematical programs with complementarity constraints (MPCCs) using algorithms and procedures of smooth nonlinear programming. Although MPCCs do not satisfy a constraint qualification, we establish sufficient conditions for their Lagrange multiplier set to be nonempty. MPCCs that have nonempty Lagrange multiplier sets and that satisfy the quadratic growth condition can be approached by the elastic mode with a bounded penalty parameter. In this context, the elastic mode transforms MPCC into a nonlinear program with additional variables that has an isolated stationary point and local minimum at the solution of the original problem, which in turn makes it approachable by sequential quadratic programming algorithms. We also prove that a modified version of the elastic mode exhibits global convergence to C-stationary points when applied to the optimization of parametric mixed P variational inequalities. The robustness of the elastic mode when applied to MPCCs is demonstrated by several numerical examples.

1. Introduction. Complementarity constraints can be used to model numerous economics or engineering applications [24, 31]. Solving optimization problems with complementarity constraints may prove difficult for classical nonlinear optimization, however, given that, at a solution x^*, such problems cannot satisfy a constraint qualification [24]. As a result, algorithms based on the linearization of the feasible set, such as sequential quadratic programming (SQP) algorithms, may fail because feasibility of the linearization can no longer be guaranteed in a neighborhood of the solution [24].

Several methods have been recently proposed to accommodate such problems. For example, a nondifferentiable penalty term in the objective function can be used to replace the complementarity constraints [25], while maintaining the same solution set. Although the new problem may now satisfy the constraint qualification the nondifferentiability of the objective function is an obstacle to the efficient computation of an optimal point. Another method is the disjunctive nonlinear programming (disjunctive NLP) approach [24], though this may lead to a large number of subcases to account for the alternatives involving degenerate complementarity constraints. If all constraint functions, with the exception of the complementarity constraints, are linear, then efficient active set approaches can be defined, if the linear independence constraint qualification holds [16]. Still other approaches have been defined for problems whose complementarity constraints originate in equilibrium conditions [24].

A nonsmooth approach has been proposed in [31] for MPCCs in which the underlying complementarity constraints originate in a variational inequality with strong regularity properties. A bundle trust-region algorithm is defined in which each element of the bundle is generated from the generalized gradient of the reduced objective

* Thackeray 301, Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260 (anitescu@math.pitt.edu). This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38. This work was also supported by award DMS-9973071 of the National Science Foundation.

function. The key step is to produce an element of the generalized gradient [31, Equations (7.24), (7.25)], which may be quite costly for general cases at points where there are a substantial number of degenerate complementarity constraints.

In this work we investigate the possibility of solving MPCCs by applying certain SQP algorithms to their nonlinear programming formulation. This endeavor is important because it allows one to extend the considerable body of analytical and computational expertise of smooth nonlinear programming to this new class of problems. The advantage of such an approach over disjunctive programming, for example, is that it considers simultaneously all the alternatives involving degenerate complementarity constraints. The disadvantage is that the description of the constraint set is considerably less well behaved.

Recognizing that the potential infeasibility of the subproblems with linearized constraints may prevent normal termination of SQP algorithms, we discuss their use in conjunction with the elastic mode [18]. The elastic mode is a standard technique of approaching infeasible subproblems by relaxing the constraints and introducing a differentiable penalty term in the objective function. To show that such an approach can accommodate a large class of MPCCs, we use the framework from [33] to determine sufficient conditions for MPCCs to have nonempty Lagrange multiplier sets.

As in [33], the first- and second-order optimality properties of an MPCC are compared with the similar properties of two nonlinear programs that involve no complementarity constraints and may thus satisfy a constraint qualification. Here, however, we consider the optimality properties of an MPCC formulated as a nonlinear program with differentiable data. In [33] MPCC is equivalently described with the complementarity constraints replaced by an equality involving the nondifferentiable function $\min\{x_1, x_2\}$. The two formulations will ultimately have similar properties, but the smooth description is important in anticipation of the use of a standard nonlinear programming algorithm to solve MPCCs.

The elastic mode approach we present here is different from other nonlinear programming approaches for MPCC in the following important respect. Virtually all smooth nonlinear programming approaches currently described in the literature for finding a solution x^* of MPCC consist of transforming it into another nonlinear program depending on a parameter p, MPCC(p) and then finding the solution x^p of the modified problem [21, 24, 34]. The problem MPCC(p) will have enough constraint regularity for x^p to be found reasonably efficiently. The solution x^* is then obtained in the limit as $p \to 0$, and $x^p \neq x^*$ for any p. The program MPCC(0) is undefined, or does not satisfy a constraint qualification (if the parameter is a penalty parameter c, the same observation is valid by choosing $p = \frac{1}{c}$).

For the elastic mode, under conditions to be specified in the body of this work, MPCC is transformed into a problem MPCC(c) that satisfies a constraint qualification and has x^* as a local solution for all c sufficiently large but finite. So MPCC is transformed by a finite procedure in a nonlinear program with the same solution that satisfies a constraint qualification, which does not happen for the other approaches. To our knowledge, the developments presented here are the first systematic approach of this type that is valid for a generic instance of mathematical programs with complementarity constrains.

The paper is structured as follows. In the remainder of Section 1 we review the relevant nonlinear programming concepts. In Section 2 we discuss sufficient conditions for MPCC to have a nonempty Lagrange multiplier set, in spite of not satisfying a constraint qualification at any point. This allows us to argue in Section 3 that the elastic mode applied to an instance of the MPCC class will retrieve a local solution of the problem for a finite value of the penalty parameter, a point which is supported by several numerical examples. In Section 4 we show that an adaptive elastic mode approach can be guaranteed to retrieve a feasible C-stationary point of an optimization problem whose complementarity constraints originate in a mixed P variational inequality. To achieve this global convergence result we will allow the penalty parameter to grow to ∞, if necessary.

1.1. Optimality Conditions for General Nonlinear Programming. We review the optimality conditions for a general nonlinear program

$$\min_x \tilde{f}(x) \quad \text{subject to } \tilde{g}(x) \leq 0, \; \tilde{h}(x) = 0. \tag{1.1}$$

Here $\tilde{g} : \mathcal{R}^n \to \mathcal{R}^m$, $\tilde{h} : \mathcal{R}^n \to \mathcal{R}^r$. We assume that $\tilde{f}$, $\tilde{g}$, and $\tilde{h}$ are twice continuously differentiable.

In this work we will denote quantities connected to nonlinear programs such as (1.1) by the superscript $\tilde{}$, since f, g, and h will later denote the objective value and constraints of MPCC.

We call x a stationary point of (1.1) if the Fritz-John condition holds: There exist multipliers $0 \neq \tilde{\lambda} = (\tilde{\lambda}_0, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_{m+r}) \in \mathcal{R}^{m+r+1}$, such that

$$\nabla_x \mathcal{L}(x, \tilde{\lambda}) = 0, \;\; \tilde{h}(x) = 0; \;\; \tilde{\lambda}_i \geq 0, \;\; \tilde{g}_i(x) \leq 0, \quad \text{for } i = 1, 2, \ldots, m; \;\; \sum_{i=1}^m \tilde{\lambda}_i \tilde{g}_i(x) = 0. \tag{1.2}$$

Here $\mathcal{L}$ is the Lagrangian function

$$\mathcal{L}(x, \tilde{\lambda}) = \tilde{\lambda}_0 \tilde{f}(x) + \sum_{i=1}^m \tilde{\lambda}_i \tilde{g}_i(x) + \sum_{j=1}^r \tilde{\lambda}_{m+j} \tilde{h}_j(x). \tag{1.3}$$

A local solution x^* of (1.1) is a stationary point [30]. We introduce the sets of generalized Lagrange multipliers

$$\Lambda^g(x) = \left\{ 0 \neq \tilde{\lambda} \in \mathcal{R}^{m+r+1} \mid \; \tilde{\lambda} \text{ satisfies (1.2) at } x \right\}, \tag{1.4}$$

$$\Lambda_1^g(x) = \left\{ \tilde{\lambda} \in \Lambda^g(x) \mid \tilde{\lambda}_0 = 1 \right\}. \tag{1.5}$$

The active set at a stationary point x is

$$\tilde{\mathcal{A}}(x) = \left\{ i \in \{1, 2, \ldots, m\} \mid \tilde{g}_i(x) = 0 \right\}. \tag{1.6}$$

The inactive set at x is the complement of $\tilde{\mathcal{A}}(x)$:

$$\tilde{\mathcal{A}}^c(x) = \{1, 2, \ldots, m\} - \tilde{\mathcal{A}}(x). \tag{1.7}$$

With this notation, the complementarity condition from (1.2), $\sum_{i=1}^m \tilde{\lambda}_i g_i(x) = 0$, becomes $\tilde{\lambda}_{\tilde{\mathcal{A}}^c(x)} = 0$.

If certain regularity conditions hold at a stationary point x (discussed below), there exist $\tilde{\mu} = (\tilde{\mu}_1, \tilde{\mu}_2, \ldots, \tilde{\mu}_{m+r}) \in \mathcal{R}^{m+r}$ that satisfy the Karush-Kuhn-Tucker (KKT) conditions [4, 5, 12]:

$$\begin{aligned} &\nabla_x \tilde{f}(x) + \textstyle\sum_{i=1}^m \tilde{\mu}_i \nabla_x \tilde{g}_i(x) + \sum_{j=1}^r \tilde{\mu}_{m+j} \nabla_x \tilde{h}_j(x) = 0, \;\; \tilde{h}(x) = 0; \\ &\tilde{\mu}_i \geq 0, \;\; \tilde{g}_i(x) \leq 0, \;\; \tilde{\mu}_i \tilde{g}_i(x) = 0, \quad \text{for } i = 1, 2, \ldots, m. \end{aligned} \tag{1.8}$$

In this case, $\tilde{\mu}$ are referred to as the Lagrange multipliers, and x is called a Karush-Kuhn-Tucker (KKT) point. We denote the set of Lagrange multipliers by

$$\Lambda(x) = \left\{ \tilde{\mu} \in \mathcal{R}^{m+r} \mid \; \tilde{\mu} \text{ satisfies (1.8) at } x \; \right\}. \tag{1.9}$$

A simple inspection of the definitions of $\Lambda(x)$ and $\Lambda_1^g(x)$ reveals that

$$\tilde{\mu} \in \Lambda(x) \Leftrightarrow (1, \tilde{\mu}) \in \Lambda_1^g(x).$$

Also, because of the first-order homogeneity of the conditions (1.2), and from (1.8), it immediately follows that

$$\Lambda(x) \neq \emptyset \Leftrightarrow \Lambda_1^g(x) \neq \emptyset \Leftrightarrow \exists \tilde{\lambda} \in \Lambda^g(x), \text{ such that } \tilde{\lambda}_0 \neq 0. \tag{1.10}$$

The regularity condition, or constraint qualification, ensures that a linear approximation of the feasible set in the neighborhood of a stationary point x captures the geometry of the feasible set. The regularity condition that we will use at times at a stationary point x is the Mangasarian-Fromovitz constraint qualification (MFCQ) [27, 26]:

$$\text{(MFCQ)} \qquad \begin{array}{l} 1.\ \nabla_x \tilde{h}_j(x),\ j = 1, 2, \ldots, r, \text{ are linearly independent and} \\ 2.\ \exists p \neq 0 \ \text{ such that } \ \nabla_x \tilde{h}_j(x)^T p = 0, \ \ j = 1, 2, \ldots, r \\ \text{and } \ \nabla_x \tilde{g}_i(x)^T p < 0, \ \ i \in \tilde{\mathcal{A}}(x). \end{array}$$

It is well known [17] that MFCQ is equivalent to the fact that the set $\Lambda(x)$ of Lagrange multipliers of (1.1) is not empty and bounded at a stationary point x of (1.1). Note that $\Lambda(x)$ is certainly polyhedral in any case.

It is useful to extend this constraint qualification for points that are not stationary or even feasible. We say that the Mangasarian-Fromovitz constraint qualification holds at a possibly infeasible point x of (1.1) if MFCQ holds at x where the active set is now defined as

$$\tilde{\mathcal{A}}(x) = \{ i \in \{1, 2, \ldots, m\} \mid \tilde{g}_i(x) \geq 0 \}. \tag{1.11}$$

We call x a generalized (possibly infeasible) Fritz-John point if

$$\begin{array}{c} \exists 0 \neq \tilde{\mu} \in \mathcal{R}^{m+r}, \ \tilde{\mu}_i \geq 0, \ i = 1, 2, \ldots, m \ \text{ such that} \\ \sum_{i \in \tilde{\mathcal{A}}(x)} \tilde{\mu}_i \nabla_x \tilde{g}_i(x) + \sum_{j=1}^{r} \tilde{\mu}_{m+j} \nabla_x \tilde{h}_j(x) = 0. \end{array} \tag{1.12}$$

By the alternative theorem,

$$\text{(1.12) holds} \Leftrightarrow \text{MFCQ does not hold at } x. \tag{1.13}$$

Another condition that we will use on occasion is the strict Mangasarian-Fromovitz constraint qualification (SMFCQ). We say that this condition is satisfied by (1.1) at a KKT point x if

$$\text{(SMFCQ)} \qquad \begin{array}{l} \text{1) MFCQ is satisfied at } x \text{ and} \\ \text{2) the Lagrange multiplier set } \Lambda(x) \text{ contains exactly one element.} \end{array}$$

The critical cone at a stationary point x is [10, 35]

$$\begin{aligned}\mathcal{C}(x) \;=\; & \Big\{u \in \mathcal{R}^n \mid \nabla_x \tilde{h}_j(x)^T u = 0,\ j = 1, 2, \ldots, r, \\ & \nabla_x \tilde{g}_i(x)^T u \le 0,\ i \in \tilde{\mathcal{A}}(x);\ \nabla_x \tilde{f}(x)^T u \le 0\Big\}.\end{aligned} \tag{1.14}$$

We now review the conditions for a point x^* to be a solution of (1.1). The second-order necessary conditions for x^* to be a local minimum are that $\Lambda^g(x^*) \neq \emptyset$ and [22]

$$\forall u \in \mathcal{C}(x^*),\ \ \exists \tilde{\lambda}^* \in \Lambda^g(x^*), \text{ such that } u^T \nabla^2_{xx} \mathcal{L}(x^*, \tilde{\lambda}^*) u \ge 0. \tag{1.15}$$

The second-order sufficient conditions for x^* to be a local minimum are that $\Lambda^g(x^*) \neq \emptyset$ and [22]

$$\forall u \in \mathcal{C}(x^*),\ \ u \neq 0,\ \ \exists \tilde{\lambda}^* \in \Lambda^g(x^*), \text{ such that } u^T \nabla^2_{xx} \mathcal{L}(x^*, \tilde{\lambda}^*) u > 0. \tag{1.16}$$

1.2. Notation. For a mapping $q : \mathcal{R}^n \to \mathcal{R}^l$, we define

$$q^+(x) = \begin{pmatrix} \max\{q_1(x), 0\} \\ \max\{q_2(x), 0\} \\ \vdots \\ \max\{q_l(x), 0\} \end{pmatrix} \text{ and } q^-(x) = \begin{pmatrix} \max\{-q_1(x), 0\} \\ \max\{-q_2(x), 0\} \\ \vdots \\ \max\{-q_l(x), 0\} \end{pmatrix}.$$

With this definition, it immediately follows that $q(x) = q^+(x) - q^-(x)$ and that $|q_i(x)| = q_i^+(x) + q_i^-(x)$, $i = 1, 2, \ldots, l$.

We denote the L_∞ nondifferentiable penalty function by

$$\tilde{P}_\infty(x) = \max\left\{\tilde{g}_1(x), \tilde{g}_2(x), ..., \tilde{g}_m(x), \left|\tilde{h}_1(x)\right|, \left|\tilde{h}_2(x)\right|, \ldots, \left|\tilde{h}_r(x)\right|, 0\right\}. \tag{1.17}$$

We also define the L_1 penalty function as

$$\tilde{P}_1(x) = \sum_{i=1}^{m} \tilde{g}_i^+(x) + \sum_{j=1}^{r} \left|\tilde{h}_j(x)\right|. \tag{1.18}$$

It is immediate that

$$0 \le \tilde{P}_\infty(x) \le \tilde{P}_1(x) \le (m + r)\tilde{P}_\infty(x).$$

An obvious consequence of (1.18) and (1.17) is that x is a feasible point of (1.1) if and only if $\tilde{P}_1(x) = \tilde{P}_\infty(x) = 0$.

We say that the nonlinear program (1.1) satisfies the quadratic growth condition with a parameter $\tilde{\sigma}$ at x^* if

$$\max\left\{\tilde{f}(x) - \tilde{f}(x^*), \tilde{P}_\infty(x)\right\} \ge \tilde{\sigma}\, ||x - x^*||^2 \tag{1.19}$$

holds for some $\tilde{\sigma} > 0$ and all x in a neighborhood of x^*. The quadratic growth condition is equivalent to the second-order sufficient conditions (1.16) [7, 8, 22, 23, 35].

For the case in which MFCQ holds at a solution x^* of (1.1), the quadratic growth condition at x^* is equivalent to [7]

$$\tilde{f}(x) - \tilde{f}(x^*) \geq \tilde{\sigma}_{\tilde{f}} \, ||x - x^*||^2 \tag{1.20}$$

for some $\tilde{\sigma}_{\tilde{f}} > 0$ and all x feasible in a neighborhood of x^*.

If $\tilde{F} : \mathcal{R}^n \to \mathcal{R}^m$ is a differentiable mapping we denote its Jacobian by

$$\nabla_x \tilde{F}(x) = \begin{bmatrix} \frac{\partial \tilde{F}_1}{\partial x_1} & \frac{\partial \tilde{F}_2}{\partial x_1} & \cdots & \frac{\partial \tilde{F}_m}{\partial x_1} \\ \frac{\partial \tilde{F}_1}{\partial x_2} & \frac{\partial \tilde{F}_2}{\partial x_2} & \cdots & \frac{\partial \tilde{F}_m}{\partial x_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \tilde{F}_1}{\partial x_n} & \frac{\partial \tilde{F}_2}{\partial x_n} & \cdots & \frac{\partial \tilde{F}_m}{\partial x_n} \end{bmatrix}.$$

We use this convention since now gradients of scalar valued maps become column vectors, which is the usual setup in the optimization literature. For vector valued maps $\tilde{F}$, this definition of the Jacobian is the transpose of the one used, for example, in [24].

1.3. Exact Penalty Conditions for Degenerate Nonlinear Programming. We now assume that at a solution x^* of the nonlinear program (1.1) the following conditions hold:

1. The Lagrange multiplier set at x^*, $\Lambda(x^*)$, is not empty.
2. The quadratic growth condition (1.19) is satisfied.

Then there exists a neighborhood $\mathcal{V}(x^*)$, some penalty parameters $\tilde{c}_1 \geq 0$, $\tilde{c}_\infty \geq 0$ and some growth parameters $\sigma_1 > 0$ and $\sigma_\infty > 0$ such that [8, Theorem 3.113]

$$\begin{aligned} \forall x \in \mathcal{V}(x^*), \psi_1(x) = \tilde{f}(x) + \tilde{c}_1 \tilde{P}_1(x) &\geq \tilde{f}(x^*) + \sigma_1 \, ||x - x^*||^2 \\ &= \psi_1(x^*) + \sigma_1 \, ||x - x^*||^2 , \end{aligned} \tag{1.21}$$

$$\begin{aligned} \forall x \in \mathcal{V}(x^*), \psi_\infty(x) = \tilde{f}(x) + \tilde{c}_\infty \tilde{P}_\infty(x) &\geq \tilde{f}(x^*) + \sigma_\infty \, ||x - x^*||^2 \\ &= \psi_\infty(x^*) + \sigma_\infty \, ||x - x^*||^2 . \end{aligned} \tag{1.22}$$

Therefore, x^* becomes an unconstrained strict local minimum for the nondifferentiable functions $\psi_1(x)$ and $\psi_\infty(x)$. Such functions are called nondifferentiable exact merit functions for the nonlinear program (1.1) [4, 5, 12]. If (1.21) and (1.22) are satisfied then we say that the functions $\psi_1(x)$ and $\psi_\infty(x)$ satisfy a quadratic growth condition near x^*.

The minimal values of $\tilde{c}_1$ and $\tilde{c}_\infty$ that result in (1.21) and (1.22) holding depend on both first and second-order properties of the nonlinear program (1.1) at x^*. However, in order for x^* to be a stationary point for $\psi_1(x)$ and $\psi_\infty(x)$ the parameters $\tilde{c}_1$ and $\tilde{c}_\infty$ must satisfy [2, 4, 12]

$$\tilde{c}_1 \geq \min_{\tilde{\mu} \in \Lambda(x^*)} ||\tilde{\mu}||_\infty, \quad \tilde{c}_\infty \geq \min_{\tilde{\mu} \in \Lambda(x^*)} ||\tilde{\mu}||_1. \tag{1.23}$$

Therefore the size of the penalty parameters $\tilde{c}_1$ and $\tilde{c}_\infty$ that makes the corresponding merit function exact is connected to the minimal size of the Lagrange multipliers of (1.1).

1.4. Nonlinear programming algorithms with global convergence safeguards. A desirable feature of nonlinear programming algorithms is that any of their accumulation points should be meaningful for the original nonlinear program (1.1). Since one possible outcome is that the nonlinear program is infeasible, one should define meaningful in this context.

We construct the following nonlinear program which is associated with the constraints of (1.1)

$$
\begin{array}{lrll}
\min_{x,u,v,w} & e_m^T u + e_r^T\,(v+w) & & \\
\text{subject to} & \tilde{g}_i(x) \leq u_i, & i = 1,2,\ldots,m, & \\
& -v_j \leq \tilde{h}_j(x) \leq w_j, & j = 1,2,\ldots,r & \\
& u,v,w \geq 0, & &
\end{array}
\tag{1.24}
$$

where e_m and e_r are vectors whose entries are all ones, of dimension m and r, respectively. The nonlinear program (1.24) is closely related to the unconstrained minimization of $\tilde{P}_1(x)$, and is also related to the feasibility restoration phase of the Filter SQP algorithm [15]. The nonlinear program (1.24) satisfies MFCQ everywhere and has two types of stationary points:

1. Stationary points (x,u,v,w) at which the objective function is 0. The component x of such stationary points is a feasible point for (1.1).
2. Stationary points (x,u,v,w) at which the objective function is not 0. The component x of such points is an infeasible point for (1.1). From the stationarity conditions (1.8) applied to (1.24) it is immediate that such points x must be infeasible Fritz-John points (1.12) of (1.1) at which, from (1.13), MFCQ cannot hold.

If, while attempting to solve (1.1), an algorithm encounters a point of the second type then, based on first-order information alone, the objective function of (1.24) and $\tilde{P}_1(x)$ cannot be locally reduced.

We say that an algorithm is globally convergent to a local stationary point of (1.1), or that it has a global convergence safeguard, if any accumulation point of the algorithm is either

A) an infeasible Fritz-John point (MFCQ does not hold),
B) a feasible Fritz-John point (MFCQ does not hold), or
C) a KKT point.

An algorithm with this property is FilterSQP [13, 14]. The algorithm does not use a merit function, and when a linearized version of (1.1) becomes infeasible, the algorithm enters a feasibility restoration phase which is based on a nonlinear program that is essentially (1.24) [15]. If the algorithm does not exit the feasibility restoration phase, then it will end up in either case A) or B).

1.5. Formulation of Mathematical Programs with Complementarity Constraints. We use notation similar to the one in [33] to define a mathematical program with complementarity constraints (MPCC).

$$
\text{(MPCC)} \quad
\begin{array}{lrll}
\min_x & f(x) & & \\
\text{subject to} & g_i(x) & \leq 0, & i = 1,2,\ldots,n_i \\
& h_j(x) & = 0, & j = 1,2,\ldots,n_e \\
& F_{k,1}(x) & \leq 0, & k = 1,2,\ldots,n_c \\
& F_{k,2}(x) & \leq 0, & k = 1,2,\ldots,n_c \\
& F_{k,1}(x)F_{k,2}(x) & \leq 0, & k = 1,2,\ldots,n_c.
\end{array}
$$

In this work we assume that the data of (MPCC) ($f(x), h(x), g(x)$ and $F_{k,i}(x)$, for $k = 1, 2, \ldots, n_c$, and $i = 1, 2$) are twice continuously differentiable.

For a given k, the constraints $F_{k,1}(x) \le 0$, $F_{k,2}(x) \le 0$ imply that $F_{k,1}(x)F_{k,2}(x) \le 0$ is equivalent to $F_{k,1}(x)F_{k,2}(x) = 0$. The constraints $F_{k,1}(x)F_{k,2}(x) \le 0$ are therefore called complementarity constraints and are active at any feasible point of (MPCC).

Since we cannot have $F_{k,1}(x) < 0$, $F_{k,2}(x) < 0$, and $F_{k,1}(x)F_{k,2}(x) < 0$ simultaneously, it follows that MFCQ cannot hold at any feasible point x [24, 33].

1.6. MPCC Notation. If i is one of $1, 2$ we define $\bar{i} = 2 - i + 1$. Therefore $i = 1 \Rightarrow \bar{i} = 2$, and $i = 2 \Rightarrow \bar{i} = 1$. The complementarity constraints can thus be written as $F_{k,i}(x)F_{k,\bar{i}}(x) \le 0$, $k = 1, 2, \ldots, n_c$. We use the notation

$$F(x) = \left(F_{11}(x), F_{12}(x), F_{21}(x), F_{22}(x), \ldots, F_{n_c1}(x), F_{n_c2}(x)\right)^T. \tag{1.25}$$

The active set of the inequality constraints $g_i(x) \le 0$, $1 \le i \le m$, at a feasible point x is

$$\mathcal{A}(x) = \{i \in \{1, 2, \ldots, n_i\} \mid g_i(x) = 0\}. \tag{1.26}$$

We use the following notation:

$$\mathcal{I}(x) = \left\{(k, i) \in \{1, 2, \ldots, n_c\} \times \{1, 2\} \mid F_{k,\bar{i}}(x) < 0\right\}, \tag{1.27}$$

$$\overline{\mathcal{I}}(x) = \{(k, i) \in \{1, 2, \ldots, n_c\} \times \{1, 2\} \mid F_{k,i}(x) < 0\}, \tag{1.28}$$

$$\mathcal{D}(x) = \left\{(k, i) \in \{1, 2, \ldots, n_c\} \times \{1, 2\} \mid F_{k,i}(x) = F_{k,\bar{i}}(x) = 0\right\}, \tag{1.29}$$

$$\mathcal{I}^c(x) = \{1, 2, \ldots, n_c\} \times \{1, 2\} - \mathcal{I}(x), \tag{1.30}$$

$$\mathcal{K}(x) = \{k \in \{1, 2, \ldots, n_c\} \mid (k, 1) \in \mathcal{I}(x) \ \text{ or } \ (k, 2) \in \mathcal{I}(x)\}, \tag{1.31}$$

$$\overline{\mathcal{K}}(x) = \{k \in \{1, 2, \ldots, n_c\} \mid F_{k,1}(x) = F_{k,2}(x) = 0\} = \{1, 2, \ldots, n_c\} - \mathcal{K}(x). \tag{1.32}$$

There are two cases for the constraints involved in the complementarity constraints at a feasible point x.

1. $F_{k,1}(x) + F_{k,2}(x) < 0$. In this case there is an $i(k) \in \{1, 2\}$ such that $F_{k,i(k)} = 0$ and $F_{k,\bar{i}(k)} < 0$. Therefore, with our notation $k \in \mathcal{K}(x)$, $(k, i(k)) \in \mathcal{I}(x)$ and $(k, \bar{i}(k)) \in \overline{\mathcal{I}}(x)$. We call $F_{k,1}(x), F_{k,2}(x)$ a nondegenerate (or strictly complementary) pair. In the rest of the paper $i(k)$ and $\bar{i}(k)$ will have the meaning defined in this paragraph, whenever $k \in \mathcal{K}$.
2. $F_{k,1}(x) + F_{k,2}(x) = 0$, or $F_{k,1}(x) = F_{k,2}(x) = 0$. In this case $k \in \overline{\mathcal{K}}(x)$, $(k, 1) \in \mathcal{D}(x)$ and $(k, 2) \in \mathcal{D}(x)$. We call $F_{k,1}(x), F_{k,2}(x)$ a degenerate pair.

Therefore $\mathcal{I}(x)$ contains the indices of the active constraints at which strict complementarity occurs, whereas $\mathcal{D}(x)$ contains the indices of the constraints that are degenerate at x from the point of view of complementarity. The set $\mathcal{K}(x)$ represents the indices k at which strict complementarity occurs and $\overline{\mathcal{K}}(x)$ the indices k at which complementarity degeneracy occurs.

Since we are interested in the behavior of (MPCC) at a solution point x^*, we may avoid the dependence of these index sets on x. Therefore we denote $\mathcal{I} = \mathcal{I}(x^*)$, $\mathcal{D} = \mathcal{D}(x^*)$, $\mathcal{K} = \mathcal{K}(x^*)$, and $\mathcal{A} = \mathcal{A}(x^*)$. At x^* we denote by $n_{\mathcal{I}}$ and $n_{\mathcal{D}}$ the number of elements in $\mathcal{I}$ and $\mathcal{D}$, respectively.

For a set of pairs $\mathcal{J} \subset \{1, 2, \ldots, n_c\} \times \{1, 2\}$ we denote by $F_{\mathcal{J}}$ a map whose components are $F_{k,i}$ with $(k, i) \in \mathcal{J}$.

1.7. Associated Nonlinear Programs at x^*. In this section we associate two nonlinear programs to (MPCC). This will help with characterizing the stationarity conditions for (MPCC). The notation is from [33].

At x^* we associate the relaxed nonlinear program (RNLP) to (MPCC).

$$\text{(RNLP)} \qquad \begin{array}{rlll} \min_x & f(x) & & \\ \text{subject to} & g_i(x) & \leq 0, & i = 1, 2, \ldots, n_i \\ & h_j(x) & = 0, & j = 1, 2, \ldots, n_e \\ & F_{\mathcal{D}}(x) & \leq 0, & \\ & F_{\mathcal{I}}(x) & = 0. & \end{array}$$

As it can be seen, (RNLP) is obtained from (MPCC) by dropping the elements from $F(x)$ that are inactive at x^*, as well as the complementarity constraints, but enforcing the complements of inactive constraints as equality constraints.

We also associate at x^* the tightened nonlinear program (TNLP), in which all the complementarity constraints in (MPCC) are dropped and all active constraints at x^* connected to complementarity constraints are replaced by equality constraints.

$$\text{(TNLP)} \qquad \begin{array}{rlll} \min_x & f(x) & & \\ \text{subject to} & g_i(x) & \leq 0, & i = 1, 2, \ldots, n_i \\ & h_j(x) & = 0, & j = 1, 2, \ldots, n_e \\ & F_{\mathcal{D}}(x) & = 0, & \\ & F_{\mathcal{I}}(x) & = 0. & \end{array}$$

We immediately see that, near x^*, (TNLP) is a more constrained problem than (MPCC), which in turn is more constrained than (RNLP), and all three programs have the same objective function. As a result, if x^* is a local solution of (RNLP), then it must be a local solution of (MPCC). Also, if x^* is a local solution of (MPCC), then it will be a local solution of (TNLP). None of the reverse implications hold in general for either local solutions or stationary points.

However, if (TNLP) satisfies SMFCQ at a solution x^* of (MPCC), then x^* is a Karush-Kuhn-Tucker point of (TNLP) and (RNLP) [33].

2. The Lagrange Multiplier Set of (MPCC). In this section we analyze the relationship between the relevant mathematical objects of (MPCC) and (RNLP) at a solution x^*. The (RNLP) formulation does not immediately violate MFCQ, the way (MPCC) does. By establishing a correspondence between the Lagrange multiplier sets of (RNLP) and (MPCC) we ensure that, under certain conditions, (MPCC) has a nonempty Lagrange multiplier set, although it does not satisfy a constraint qualification.

2.1. Critical Cones. In this section we compare the critical cones of (MPCC) and (RNLP). The active sets play a structural part in the definition of the critical cones. We have that

$$\nabla_x \left(F_{k,1}F_{k,2}\right)(x^*) = F_{k,1}(x^*)\nabla_x F_{k,2}(x^*) + F_{k,2}(x^*)\nabla_x F_{k,1}(x^*).$$

We distinguish two cases.

1. If $k \in \overline{\mathcal{K}}$, we have that $F_{k,1}(x^*) = F_{k,2}(x^*) = 0$, and, as a result,

$$k \in \overline{\mathcal{K}} \Rightarrow \nabla_x \left(F_{k,1}F_{k,2}\right)(x^*) = 0. \tag{2.1}$$

   Therefore, if $k \in \overline{\mathcal{K}}$, the constraint $F_{k,1}(x)F_{k,2}(x) \leq 0$, which is active at x^*, has no bearing on the definition (1.14) of the critical cone (it would just add the constraint $0 \leq 0$).

2. If $k \in \mathcal{K}$, then there exist an $i(k)$ such that $(k, i(k)) \in \mathcal{I}$ and $(k, \bar{i}(k)) \in \overline{\mathcal{I}}$. The constraints $F_{k,i(k)}(x) \leq 0$ and $F_{k,i(k)}(x) F_{k,\bar{i}(k)}(x) \leq 0$ are active at x^*, whereas $F_{k,\bar{i}(k)}(x^*) < 0$ and the corresponding constraint is inactive at x^*. Therefore we have that

$$\nabla_x \left(F_{k,i(k)} F_{k,\bar{i}(k)} \right) (x^*) = F_{k,\bar{i}(k)}(x^*) \nabla_x F_{k,i(k)}(x^*), \tag{2.2}$$

and thus the constraints connected to k that enter the definition of the critical cone (1.14) are

$$\nabla_x F_{k,i(k)}(x^*)^T u \leq 0, \quad F_{k,\bar{i}(k)}(x^*) \left(\nabla_x F_{k,i(k)}(x^*) \right)^T u \leq 0$$

for u an element of the critical cone.

Using the definition (1.14) we get that the critical cone of (MPCC) is

$$\mathcal{C}_{\mathsf{MPCC}} = \{ u \in R^n \mid \begin{array}{lcll} \nabla_x f(x^*)^T u & \leq & 0, & \\ \nabla_x g_i(x^*)^T u & \leq & 0, & i \in \mathcal{A} \\ \nabla_x h_j(x^*)^T u & = & 0, & j \in 1, 2, \ldots, n_e \\ \nabla_x F_{k,1}(x^*)^T u & \leq & 0, & k \in \overline{\mathcal{K}} \\ \nabla_x F_{k,2}(x^*)^T u & \leq & 0, & k \in \overline{\mathcal{K}} \\ \nabla_x F_{k,i(k)}(x^*)^T u & \leq & 0, & (k, i(k)) \in \mathcal{I} \\ F_{k,\bar{i}(k)}(x^*) \nabla_x F_{k,i(k)}(x^*)^T u & \leq & 0, & (k, i(k)) \in \mathcal{I} \} . \end{array} \tag{2.3}$$

We use (1.14) again to determine the critical cone of the relaxed nonlinear program. It is immediate from the definition of the index sets $\mathcal{I}, \mathcal{K}$, and $\mathcal{D}$ that all constraints involving components of $F(x)$ are active at x^* for (RNLP). It thus follows that the critical cone of (RNLP) is

$$\mathcal{C}_{\mathsf{RNLP}} = \{ u \in R^n \mid \begin{array}{lcll} \nabla_x f(x^*)^T u & \leq & 0, & \\ \nabla_x g_i(x^*)^T u & \leq & 0, & i \in \mathcal{A} \\ \nabla_x h_j(x^*)^T u & = & 0, & j \in 1, 2, \ldots, n_e \\ \nabla_x F_{k,1}(x^*)^T u & \leq & 0, & k \in \overline{\mathcal{K}} \\ \nabla_x F_{k,2}(x^*)^T u & \leq & 0, & k \in \overline{\mathcal{K}} \\ \nabla_x F_{k,i(k)}(x^*)^T u & = & 0, & (k, i(k)) \in \mathcal{I} \} . \end{array} \tag{2.4}$$

LEMMA 2.1. $\mathcal{C}_{\mathsf{MPCC}} = \mathcal{C}_{\mathsf{RNLP}}$.

Proof The conclusion is immediate, by noting that all the constraints involving the critical cones are the same with the exception of the ones involving indices k for which $(k, i(k)) \in \mathcal{I}$. For these k, from the definition (1.27) of the index sets it follows that $F_{k,\bar{i}(k)}(x^*) < 0$. We therefore have that

$$\begin{array}{lclcrcl} \nabla_x F_{k,i(k)}(x^*)^T u & \leq & 0 & \text{and} & F_{k,\bar{i}(k)}(x^*) \nabla_x F_{k,i(k)}(x^*)^T u & \leq & 0 \Leftrightarrow \\ \nabla_x F_{k,i(k)}(x^*)^T u & \leq & 0 & \text{and} & \nabla_x F_{k,i(k)}(x^*)^T u & \geq & 0 \Leftrightarrow \\ & & & & \nabla_x F_{k,i(k)}(x^*)^T u & = & 0. \end{array}$$

Since the remaining constraints of (RNLP) and (MPCC) are the same this equivalence proves the claim. $\diamond$

2.2. Generalized Lagrange Multipliers. The set of generalized Lagrange multipliers of (MPCC) at x^* is a set of multiples

$$0 \neq (\alpha, \nu, \pi, \mu, \eta) \in \mathcal{R} \times \mathcal{R}^{n_i} \times \mathcal{R}^{n_e} \times \mathcal{R}^{2n_c} \times \mathcal{R}^{n_c}$$

that satisfies the Fritz-John conditions (1.2). Since μ are the multipliers corresponding to the components of $F(x)$, we will index them by elements in $(1,2,\dots,n_c)\times(1,2)$. The Fritz-John conditions for (MPCC) at x^* are that x^* is feasible for (MPCC) and that

$$
\begin{gathered}
\alpha\nabla_x f(x^*) + \sum_{i=1}^{n_i}\nu_i\nabla_x g_i(x^*) + \sum_{j=1}^{n_e}\pi_j\nabla_x h_j(x^*) + \\
\sum_{k=1}^{n_c}\left[\mu_{k,1}\nabla_x F_{k,1}(x^*) + \mu_{k,2}\nabla_x F_{k,2}(x^*) + \eta_k\nabla_x\left(F_{k,1}F_{k,2}\right)(x^*)\right] = 0
\end{gathered}
\tag{2.5}
$$

$$
\begin{array}{rrrl}
F_{k,i}(x^*) \le 0, & \mu_{k,i} \ge 0, & \mu_{k,i}F_{k,i}(x^*) = 0, & k = 1,2,\dots,n_c, \\
 & & & i = 1,2 \\
g_i(x^*) \le 0, & \nu_i \ge 0, & \nu_i g_i(x^*) = 0, & i = 1,2,\dots,n_i \\
F_{k,1}(x^*)F_{k,2}(x^*) \le 0, & \eta_k \ge 0, & \eta_k F_{k,1}(x^*)F_{k,2}(x^*) = 0, & k = 1,2,\dots,n_c.
\end{array}
\tag{2.6}
$$

From our definition of the index sets it follows that $F_{\overline{\mathcal{I}}}(x^*) < 0$ and $g_{\mathcal{A}^c}(x^*) < 0$. Therefore, from the complementarity conditions (2.6), it follows that $\mu_{\overline{\mathcal{I}}} = 0$ and $\nu_{\mathcal{A}^c} = 0$.

We can also determine the relations satisfied by the generalized Lagrange multipliers of (RNLP). As discussed above, the index sets that define (RNLP) have been chosen such that all constraints involving components of $F(x)$ are active. Therefore the generalized Lagrange multipliers are

$$
0 \ne (\tilde{\alpha}, \tilde{\nu}, \tilde{\pi}, \tilde{\mu}, \tilde{\eta}) \in \mathcal{R}\times\mathcal{R}^{n_i}\times\mathcal{R}^{n_e}\times\mathcal{R}^{n_{\mathcal{D}}}\times\mathcal{R}^{n_{\mathcal{I}}}
$$

that satisfy the Fritz-John conditions:

$$
\begin{gathered}
\tilde{\alpha}\nabla_x f(x^*) + \sum_{i=1}^{n_i}\tilde{\nu}_i\nabla_x g_i(x^*) + \sum_{j=1}^{n_e}\tilde{\pi}_j\nabla_x h_j(x^*) + \\
\sum_{k\in\overline{\mathcal{K}}}\left[\tilde{\mu}_{k,1}\nabla_x F_{k,1}(x^*) + \tilde{\mu}_{k,2}\nabla_x F_{k,2}(x^*)\right] + \sum_{k\in\mathcal{K}}\tilde{\eta}_{k,i(k)}\nabla_x F_{k,i(k)}(x^*) = 0
\end{gathered}
\tag{2.7}
$$

$$
\begin{array}{rrrl}
g_i(x^*) \le 0, & \tilde{\nu}_i \ge 0, & \tilde{\nu}_i g_i(x^*) = 0, & i = 1,2,\dots,n_i \\
 & \tilde{\mu}_{k,1} \ge 0, & \tilde{\mu}_{k,2} \ge 0, & k \in \overline{\mathcal{K}}.
\end{array}
\tag{2.8}
$$

Here $\tilde{\mu}$ is a vector that is indexed by elements of $\mathcal{D}$, and $\tilde{\eta}$ is indexed by elements of $\mathcal{I}$.

2.2.1. Other types of stationary points of (MPCC). In the analysis of (MPCC), other useful types of stationarity at a solution x^* can be defined, based on the interpretation of (MPCC) as a problem with nonsmooth constraints [21, 33]:

- *C-stationary points* of (MPCC) are points x^* that, together with an appropriate set of multipliers $(\alpha,\nu,\pi,\mu,\eta)$, with $\alpha = 1$, satisfy (2.5) and (2.6), except for the conditions $\mu_{k,1} \ge 0$ and $\mu_{k,2} \ge 0$, for $k \in \overline{K}$, which are now relaxed to $\mu_{k,1}\mu_{k,2} \ge 0$, for $k \in \overline{K}$.

- *M-stationary points* of (MPCC) are points x^* that, together with an appropriate set of multipliers $(\alpha, \nu, \pi, \mu, \eta)$, with $\alpha = 1$, satisfy (2.5) and (2.6), except for the conditions $\mu_{k,1} \geq 0$ and $\mu_{k,2} \geq 0$, for $k \in \overline{K}$, which are now relaxed to

$$\text{for } k \in \overline{K}, \text{ either } \mu_{k,1} > 0, \ \mu_{k,2} > 0, \text{ or } \mu_{k,1}\mu_{k,2} = 0.$$

- *B-stationary points* of (MPCC) are points x^* that, together with an appropriate set of multipliers $(\alpha, \nu, \pi, \mu, \eta)$, with $\alpha = 1$, satisfy (2.5) and (2.6). The last type of points coincide with the notion of a KKT point.

2.3. Relations between the generalized Lagrange Multiplier Sets of (MPCC) and (RNLP). Take $\tilde{\lambda} = (\tilde{\alpha}, \tilde{\nu}, \tilde{\pi}, \tilde{\mu}, \tilde{\eta}) \in \Lambda^g{}_{\mathsf{RNLP}}$. We construct from the generalized multiplier $\tilde{\lambda}$ of (RNLP) a generalized multiplier λ° of (MPCC). We define the following types of components of λ°.

1. Components that correspond to the objective function or the inequality constraints $g_i(x) \leq 0$ and equality constraints $h_j(x) = 0$

$$\alpha^\circ = \tilde{\alpha}; \ \ \nu^\circ = \tilde{\nu}; \ \ \pi^\circ = \tilde{\pi}. \tag{2.9}$$

2. Components connected to the pairwise degenerate constraints. For these we have $k \in \overline{\mathcal{K}}$ and $(k,1),(k,2) \in \mathcal{D}$ or $F_{k,1}(x^*) = F_{k,2}(x^*) = 0$. We define

$$\mu^\circ_{k,i} = \tilde{\mu}_{k,i}, \ (k,i) \in \mathcal{D}; \quad \eta^\circ_k = 0, \ k \in \overline{\mathcal{K}}. \tag{2.10}$$

Similar to the equation (2.1) we have that

$$\nabla_x \left(F_{k,1}F_{k,2}\right)(x^*) = 0,$$

and therefore

$$\begin{aligned} \tilde{\mu}_{k,1}\nabla_x F_{k,1}(x^*) + \tilde{\mu}_{k,2}\nabla_x F_{k,2}(x^*) = \mu^\circ_{k,1}\nabla_x F_{k,1}(x^*) \quad + \\ \mu^\circ_{k,2}\nabla_x F_{k,2}(x^*) + \eta^\circ_k \nabla_x \left(F_{k,1}F_{k,2}\right)(x^*). \end{aligned} \tag{2.11}$$

3. Components connected to pairwise strictly complementary constraints. In this case we have $k \in \mathcal{K}$, $(k,i(k)) \in \mathcal{I}$, and $(k,\bar{i}(k)) \in \overline{\mathcal{I}}$. Therefore $F_{k,\bar{i}(k)}(x^*) < 0$, $F_{k,i(k)}(x^*) = 0$, and we thus define the multipliers

$$\begin{array}{rcll} \mu^\circ_{k,i(k)} & = & \max\left\{\tilde{\eta}_{k,i(k)}, 0\right\}, & (k,i(k)) \in \mathcal{I} \\ \mu^\circ_{k,\bar{i}(k)} & = & 0, & (k,\bar{i}(k)) \in \overline{\mathcal{I}} \\ \eta^\circ_k & = & \frac{1}{F_{k,\bar{i}(k)}(x^*)}\min\left\{\tilde{\eta}_{k,i(k)}, 0\right\}, & k \in \mathcal{K}. \end{array} \tag{2.12}$$

It is immediate from these definitions that $\mu^\circ_{k,i(k)} \geq 0$ and $\eta^\circ_k \geq 0$. Since, for fixed k, $\tilde{\eta}_{k,i(k)}$ is the only multiplier of (RNLP) involved in definition (2.12), we obtain using (2.2) that

$$\begin{aligned} \tilde{\eta}_{k,i(k)}\nabla_x F_{k,i(k)}(x^*) &= \left[\max\left\{\tilde{\eta}_{k,i(k)}, 0\right\} + \min\left\{\tilde{\eta}_{k,i(k)}, 0\right\}\right]\nabla_x F_{k,i(k)}(x^*) \\ &= \mu^\circ_{k,i(k)}\nabla_x F_{k,i(k)}(x^*) + \eta^\circ_k F_{k,\bar{i}(k)}(x^*)\nabla_x F_{k,i(k)}(x^*) \quad = \\ \mu^\circ_{k,i(k)}\nabla_x F_{k,i(k)}(x^*) &+ \mu^\circ_{k,\bar{i}(k)}\nabla_x F_{k,\bar{i}(k)}(x^*) + \eta^\circ_k \nabla_x \left(F_{k,i(k)}F_{k,\bar{i}(k)}\right)(x^*). \end{aligned} \tag{2.13}$$

After we compare the terms that, following (2.11) and (2.13), are equal in (2.7) and (2.5), we get that $\lambda^\circ = (\alpha^\circ, \nu^\circ, \pi^\circ, \mu^\circ, \eta^\circ)$ satisfies (2.5) as well as (2.6). By tracing the definition of λ° we also have that $\tilde{\lambda} \neq 0 \Rightarrow \lambda^\circ \neq 0$. Therefore λ° is a generalized Lagrange multiplier of (MPCC) or

$$\lambda^\circ = (\alpha^\circ, \nu^\circ, \pi^\circ, \mu^\circ, \eta^\circ) \in \Lambda^g_{\mathsf{MPCC}}, \tag{2.14}$$

where $\alpha^\circ = \tilde{\alpha}$ from (2.9).

THEOREM 2.2. *If the set of Lagrange multipliers of (RNLP) is not empty, then the set of Lagrange multipliers of (MPCC) is not empty.*

Proof Since the Lagrange multiplier set of (RNLP) is not empty, we can choose $\tilde{\lambda} = (1, \tilde{\nu}, \tilde{\pi}, \tilde{\mu}, \tilde{\eta}) \in \Lambda^g_{1,\ \mathsf{RNLP}}$. From (2.14) it follows that $\lambda^\circ = (1, \nu^\circ, \pi^\circ, \mu^\circ, \eta^\circ) \in \Lambda^g_{1,\mathsf{MPCC}}$ is a generalized multiplier of (MPCC). From (1.10) it follows that the Lagrange multiplier set of (MPCC) is not empty. $\diamond$

COROLLARY 2.3. *Assume that (TNLP) satisfies SMFCQ at a solution x^* of (MPCC), i.e.*

1. *$\nabla_x F_{\mathcal{D}}(x^*)$, $\nabla_x F_{\mathcal{I}}(x^*)$, and $\nabla_x h(x^*)$ are linearly independent.*
2. *There exists $p \neq 0$ such that $\nabla_x F^T_{\mathcal{D}}(x^*)p = 0$, $\nabla_x F^T_{\mathcal{I}}(x^*)p = 0$, $\nabla_x h^T(x^*)p = 0$, $\nabla_x g^T_i(x^*)p < 0$, for $i \in \mathcal{A}(x^*)$.*
3. *The Lagrange multiplier set of (TNLP) at x^* has a unique element.*

Then the Lagrange multiplier set of (MPCC) is not empty.

Proof From [33, Theorem 2], since (TNLP) satisfies SMFCQ at x^*, the Lagrange multiplier set of (RNLP) is not empty. Following Theorem 2.2, we obtain that the Lagrange multiplier set of (MPCC) is not empty, which proves the claim. $\diamond$

Unfortunately, the reverse statement of Theorem 2.2 does not hold in the absence of SMFCQ, as is shown in [33]. Indeed, consider the following example:

$$\begin{array}{lll} \min_{y,x} & y - x & \\ & y & \leq 0 \\ & y + x & \leq 0 \\ & y(y+x) & \leq 0 \\ & x & \leq 0. \end{array} \tag{2.15}$$

The unique minimum of this problem is $(0,0)$. However, if we construct the associated (RNLP) formulation, we obtain

$$\begin{array}{lll} \min_{y,x} & y - x & \\ & y & \leq 0 \\ & y + x & \leq 0 \\ & x & \leq 0. \end{array} \tag{2.16}$$

The point $(y, 0)$ is feasible for $y < 0$ for the now-linear program (2.16). Thus (2.16) is unbounded and cannot have $(0,0)$ as a stationary point. Therefore Theorem 2.2 cannot be applied, since the Lagrange multiplier set of (2.16) is empty. In this situation (TNLP) associated to (2.15) of (2.15) does not satisfy either MFCQ or SMFCQ.

2.4. An alternative formulation. We also investigate the following equivalent formulation of (MPCC), where the complementarity constraints have been replaced by one constraint:

$$\begin{array}{rrll} \min_x & f(x) & & \\ \text{subject to} & g_i(x) & \leq 0, & i = 1, 2, \ldots, n_i \\ & h_j(x) & = 0, & j = 1, 2, \ldots, n_e \\ & F_{k,1}(x) & \leq 0, & k = 1, 2, \ldots, n_c \\ & F_{k,2}(x) & \leq 0, & k = 1, 2, \ldots, n_c \\ & \sum_{k=1}^{n_c} F_{k,1}(x) F_{k,2}(x) & \leq 0. & \end{array} \tag{2.17}$$

At a feasible point of the above program, we must have that $\sum_{k=1}^{n_c} F_{k,1}(x)F_{k,2}(x) = 0$ and the equivalence between (2.17) and (MPCC) follows immediately. This formulation is of interest in computations because it has less constraints than (MPCC).

LEMMA 2.4. *If the Lagrange multiplier set of (MPCC) is not empty, there exists a Lagrange multiplier* $(1, \nu, \pi, \mu, \eta) \in \Lambda^g_{\text{MPCC}}$ *such that* $\eta_k = \eta_1,\ k = 2, 3, \ldots, n_c$.

Proof Let $\lambda^\circ = (1, \nu^\circ, \pi^\circ, \mu^\circ, \eta^\circ) \in \Lambda^g_{\text{MPCC}}$ be a Lagrange multiplier of (MPCC). Now let $d \in \mathcal{R}^{n_c}$ such that $d \geq 0$.

If k corresponds to degenerate complementarity constraints, $k \in \overline{K}$, we have, as argued above, that $F_{k,1}(x^*) = F_{k,2}(x^*) = 0$, and thus

$$d_k \nabla_x \left(F_{k,1} F_{k,2}\right)(x^*) = 0.$$

For this case, define

$$\eta^*_k = \eta^\circ_k + d_k, \quad \mu^*_{k,1} = \mu^\circ_{k,1}, \quad \mu^*_{k,2} = \mu^\circ_{k,2},$$

which results in

$$\begin{aligned}(2.18)\qquad & \mu^*_{k,1}\nabla_x F_{k,1}(x^*) + \mu^*_{k,2}\nabla_x F_{k,2}(x^*) + \eta^*_k \nabla_x \left(F_{k,1}F_{k,2}\right)(x^*) = \\ & \mu^\circ_{k,1}\nabla_x F_{k,1}(x^*) + \mu^\circ_{k,2}\nabla_x F_{k,2}(x^*) + \eta^\circ_k \nabla_x \left(F_{k,1}F_{k,2}\right)(x^*).\end{aligned}$$

If k corresponds to strict complementarity constraints, $k \in \mathcal{K}$, we have that $F_{k,i(k)}(x^*) = 0$, $F_{k,\bar{i}(k)}(x^*) < 0$ and thus $\mu_{k,\bar{i}(k)} = 0$. Define

$$\eta^*_k = \eta^\circ_k + d_k, \quad \mu^*_{k,i(k)} = \mu^\circ_{k,i(k)} - d_k F_{k,\bar{i}(k)}(x^*) \geq 0, \quad \mu^*_{k,\bar{i}(k)} = \mu^\circ_{k,\bar{i}(k)} = 0.$$

Since $F_{k,i(k)}(x^*) = 0$ we have that

$$\nabla_x \left(F_{k,1}F_{k,2}\right)(x^*) = F_{k,\bar{i}(k)}(x^*)\nabla_x F_{k,i(k)}(x^*).$$

In turn, the last equation implies that

$$\begin{aligned}&\mu^*_{k,i(k)}\nabla_x F_{k,i(k)}(x^*) + \mu^*_{k,\bar{i}(k)}\nabla_x F_{k,\bar{i}(k)}(x^*) + \eta^*_k \nabla_x \left(F_{k,1}F_{k,2}\right)(x^*) = \\ &\left(\mu^\circ_{k,i(k)} - d_k F_{k,\bar{i}}(x^*)\right)\nabla_x F_{k,i(k)}(x^*) + \mu^\circ_{k,\bar{i}(k)}\nabla_x F_{k,\bar{i}(k)}(x^*) + \\ &\left(\eta^\circ_k + d_k\right)\nabla_x \left(F_{k,1}F_{k,2}\right)(x^*) = \mu^\circ_{k,i(k)}\nabla_x F_{k,i(k)}(x^*) + \\ &\mu^\circ_{k,\bar{i}(k)}\nabla_x F_{k,\bar{i}(k)}(x^*) + \eta^\circ_k \nabla_x \left(F_{k,1}F_{k,2}\right)(x^*).\end{aligned}$$

Since λ° satisfies (2.5) and (2.6), it follows from the preceding equation and (2.18), in a manner similar to the proof of the Theorem 2.2, that $\lambda^* = (1, \nu^\circ, \pi^\circ, \mu^*, \eta^*)$ satisfies (2.5) and (2.6) and thus $\lambda^* \in \Lambda^g_{\text{MPCC}}$ for any $0 \leq d \in \mathcal{R}^{n_c}$ where $\eta^* = \eta^\circ + d$. The conclusion is immediate, since we can always choose a vector $d \geq 0$ such that $\eta^*_k = \eta^*_1$, $k = 1, 2, \ldots, n_c$. One such choice, for example, is $d = ||\eta^\circ||_\infty (1, 1, \ldots, 1)^T - \eta^\circ$. $\diamond$

We now describe the Lagrange multiplier set of the alternative formulation (2.17). We denote mathematical objects connected to (2.17) by the subscript ${}_{MPCC1}$. We write the Fritz-John conditions (1.2) for (2.17) at the point x^*, and we obtain

$$\begin{aligned}(2.19)\qquad & \alpha^\diamond \nabla_x f(x^*) + \sum_{i=1}^{n_i} \nu^\diamond_i \nabla_x g_i(x^*) + \sum_{j=1}^{n_e} \pi^\diamond_j \nabla_x h_j(x^*) + \\ & \sum_{k=1}^{n_c}\sum_{i=1}^{2} \mu^\diamond_{k,i}\nabla_x F_{k,i}(x^*) + \eta^\diamond_1 \sum_{k=1}^{n_c} \nabla_x \left(F_{k,1}F_{k,2}\right)(x^*) = 0\end{aligned}$$

$$
(2.20)\qquad \begin{array}{llll} F_{k,i}(x^*) \leq 0, & \mu^\diamond_{k,i} \geq 0, & \mu^\diamond_{k,i} F_{k,i}(x^*) = 0, & k = 1,2,\ldots,n_c, \\ & & & i = 1,2 \\ g_i(x^*) \leq 0, & \nu^\diamond_i \geq 0, & \nu^\diamond_i g_i(x^*) = 0, & i = 1,2,\ldots,n_i \end{array}
$$

and $\eta_1^\diamond \geq 0 \in \mathcal{R}$. A generalized multiplier of (2.17) is thus

$$
\lambda^\diamond = (\alpha^\diamond, \nu^\diamond, \pi^\diamond, \mu^\diamond, \eta_1^\diamond) \in \Lambda^g_{MPCC1} \subset \mathcal{R} \times \mathcal{R}^{n_i} \times \mathcal{R}^{n_e} \times \mathcal{R}^{2n_c} \times \mathcal{R},
$$

where $\lambda^\diamond$ satisfies the Fritz-John conditions (2.19), (2.20).

THEOREM 2.5. *The formulation (2.17) has a nonempty Lagrange multiplier set if and only if (MPCC) has a nonempty Lagrange multiplier set.*

Proof If the Lagrange multiplier set of (2.17) is not empty, then there exists $\lambda^\diamond = (1, \nu^\diamond, \pi^\diamond, \mu^\diamond, \eta_1^\diamond) \in \mathcal{R} \times \mathcal{R}^{n_i} \times \mathcal{R}^{n_e} \times \mathcal{R}^{2n_c} \times \mathcal{R}$ that satisfies (2.19–2.20). Define $\eta^* = (\eta_1^\diamond, \eta_1^\diamond, \ldots, \eta_1^\diamond)^T \in \mathcal{R}^{n_c}$ and $\lambda^* = (1, \nu^\diamond, \pi^\diamond, \mu^\diamond, \eta^*)$. It follows by inspection that λ^* satisfies (2.5), (2.6) at x^*. Therefore λ^* is a generalized Lagrange multiplier of (MPCC), which means that $(\nu^\diamond, \pi^\diamond, \mu^\diamond, \eta^*)$ is a Lagrange multiplier of (MPCC). Thus the Lagrange multiplier set of (MPCC) is not empty. Conversely, applying Lemma 2.4, if the Lagrange multiplier set of (MPCC) is not empty, there exists the generalized Lagrange multiplier $\lambda = (1, \nu, \pi, \mu, \eta)$ of (MPCC) that satisfies $\eta_k = \eta_1$, for $k = 1, 2, \ldots, n_c$. It immediately follows that, since λ satisfies (2.5) and (2.6), $(1, \nu, \pi, \mu, \eta_1)$ satisfies (2.19) and (2.20) and is thus a generalized Lagrange multiplier of (2.17). Therefore (ν, π, μ, η_1) is a Lagrange multiplier of (2.17) at x^*. The proof is complete. $\diamond$

Theorems 2.2 and 2.5 give sufficient conditions for (MPCC) and (2.17) to have a nonempty Lagrange multiplier set in spite of the fact that neither satisfy a constraint qualification at any point in the usual sense of nonlinear programming. In Section 3 these conditions will imply that a relaxed version of either (MPCC) or (2.17) will have the same solution as (MPCC) and will satisfy MFCQ, which makes either approachable by SQP algorithms.

3. The Elastic Mode. An important class of techniques for solving nonlinear programs (1.1) is sequential quadratic programming. The main step in an algorithm of this type is solving the quadratic program

$$
(3.1)\qquad \begin{array}{lll} \min_d & \nabla_x \tilde{f}(x)^T d + d^T W d, & \\ \text{subject to} & \tilde{g}_i(x) + \nabla_x \tilde{g}_i(x)^T d \leq 0, & i = 1,2,\ldots,m \\ & \tilde{h}_j(x) + \nabla_x \tilde{h}_j(x)^T d = 0, & j = 1,2,\ldots,r. \end{array}
$$

The matrix W can be the Hessian of the Lagrangian (1.1) at x [13], or a positive definite matrix that approximates the Hessian of the Lagrangian on a certain subspace [12, 18, 28]. A trust-region type constraint may be added to (3.1) for the case in which W is not positive definite [13]. The solution $\overline{d}$ of (3.1) is then used in conjunction with a merit function and/or line search to determine a new iterate. We give here only a brief description of SQP algorithms, since our interest is solely in showing how the difficulties regarding the potential infeasibility of (3.1) when applied to (MPCC) can be circumvented. For more details about SQP methods see [12, 13, 18, 28].

If a nonlinear program satisfies MFCQ at x^* then the quadratic program will be feasible in a neighborhood of x^*. If MFCQ does not hold at x^*, however, the possibility exists that (3.1) is infeasible, no matter how close to x^* [24, 31, 33]. This

If *the Quadratic Program (3.1) is infeasible or its Lagrange multipliers are too large* then

NLPC: *Find the solution* $(x^{c_1}, u^{c_1}, v^{c_1}, w^{c_1})$ *of the relaxed NLP (3.3) by SQP.*

If $||(u^{c_1}, v^{c_1}, w^{c_1})|| = 0$, then x^{c_1} *solves (1.1).* Stop.

otherwise *update* c_1*:* $c_1 = c_1 + K$ *and return to* NLPC.

TABLE 3.1

An adaptive L_1 elastic mode approach

is an issue in the context of this paper because (MPCC) does not satisfy the MFCQ at a solution x^*.

In the case of infeasible subproblems some of the SQP algorithms initiate the *elastic mode* [18]. This consists of modifying the nonlinear program (1.1) by relaxing the constraints and adding a penalty term to the objective function. First we consider the case in which the added penalty term is of the L_∞ type:

$$(3.2)\qquad \begin{array}{rrl} \min_{x,\zeta} & \tilde{f}(x) + c_\infty \zeta & \\ \text{subject to} & \tilde{g}_i(x) \le \zeta, & i = 1, 2, \ldots, m, \\ & -\zeta \le \tilde{h}_j(x) \le \zeta, & j = 1, 2, \ldots, r \\ & \zeta \ge 0. & \end{array}$$

An alternative elastic mode strategy consists of using an L_1 approach. The modified nonlinear program becomes

$$(3.3)\qquad \begin{array}{lrl} \min_{x,u,v,w} & \tilde{f}(x) + c_1\left(e_m^T u + e_r^T (v + w)\right) & \\ \text{subject to} & \tilde{g}_i(x) \le u_i, & i = 1, 2, \ldots, m, \\ & -v_j \le \tilde{h}_j(x) \le w_j, & j = 1, 2, \ldots, r \\ & u, v, w \ge 0, & \end{array}$$

where e_m and e_r are vectors whose entries are all ones, of dimension m and r, respectively. We call c_∞ and c_1 the penalty parameters. Note that a point x is a stationary point of (3.2) and (3.3) if and only if it is a stationary point of $\psi_\infty(x)$ and, respectively, $\psi_1(x)$, for $\tilde{c}_\infty = c_\infty$ and $\tilde{c}_1 = c_1$ [4, 5].

All the constraints are now inequality constraints. A quadratic program analogous to (3.1) is constructed for (3.2) or (3.3), which now satisfies MFCQ at any feasible point. A feasible point of (3.2) or (3.3), respectively, can be immediately obtained by choosing ζ and u, v, w, respectively, to be sufficiently large.

An adaptive elastic mode strategy is presented in Table 3.1 when an L_1 approach is used. The elastic mode subproblem (3.3) is solved successively for increasing values of the penalty parameter, in an attempt to find its appropriate value. An equivalent strategy exists when the L_∞ approach (3.2) is used. The quantity K is a fixed positive parameter. We present here only the essential characteristics of the elastic mode. For more details about an adaptive elastic mode setup and the way it can be incorporated in an SQP framework, see [18].

For fixed penalty parameter c_1, the problem (3.3) can be solved by SQP as argued above. The possibility exists, however, that c_1 may have to be increased indefinitely in Table 3.1 before a solution of (1.1) is obtained. In the following theorem we discuss sufficient conditions that ensure that the elastic mode relaxations (3.2) and (3.3) have x^* as a component of the solution for sufficiently large but finite penalty parameter.

THEOREM 3.1. *Assume that, at a solution x^* of (1.1), we have that*

- *the Lagrange multiplier set of (1.1) is not empty,*
- *the quadratic growth condition (1.19) is satisfied at x^*, and*
- *the data of (1.1) are twice continuously differentiable.*

Then,

1. *For sufficiently large but finite values of the penalty parameter c_∞ and, respectively, c_1, we have that the points $(x^*, 0)$ and, respectively, $(x^*, 0_m, 0_n, 0_n)$, are local minima of (3.2) and (3.3) at which both MFCQ and the quadratic growth condition (1.19) are satisfied.*
2. *For the same values c_∞ and, respectively, c_1 we have that the points $(x^*, 0)$ and $(x^*, 0_m, 0_n, 0_n)$ are isolated stationary points of (3.2) and (3.3). Therefore, any SQP algorithm with global convergence safeguards that does not leave a sufficiently small neighborhood of these points will in fact converge to them.*
3. *If initialized sufficiently close to x^* and with sufficiently large penalty parameter, the adaptive elastic mode strategy from Table 3.1 will recover x^* for a finite value of the penalty parameter.*

Proof We will prove part 1 of the Theorem only for the L_∞ case, the L_1 case following in the same manner. If (x, ζ) is a feasible point of (3.2), it immediately follows from the definition (1.17) of the L_∞ penalty function, $\tilde{P}_\infty(x)$, that $\zeta \geq \tilde{P}_\infty(x)$. From (1.22), under the assumptions stated in this Theorem, we have that there exists $\tilde{c}_\infty > 0$ such that the penalty function $\psi_\infty(x)$ satisfies a quadratic growth condition at x^*. Choose now

$$c_\infty = \tilde{c}_\infty + 1.$$

Using (1.22), we obtain that, in a sufficiently small neighborhood of x^*, we must have

$$\tilde{f}(x) + \tilde{c}_\infty \zeta \geq \tilde{f}(x) + \tilde{c}_\infty \tilde{P}_\infty(x) \geq \sigma_1 \left|\left|x - x^*\right|\right|^2.$$

Whenever $\zeta \leq \frac{1}{\sigma_1}$, we will have that $\sigma_1 \zeta^2 \leq \zeta$. Therefore, in a sufficiently small neighborhood of $(x^*, 0)$, for all (x, ζ) feasible, we will have that

$$\tilde{f}(x) + c_\infty \zeta = \tilde{f}(x) + \tilde{c}_\infty \zeta + \zeta \geq \sigma_1 \left(\left|\left|x - x^*\right|\right|^2 + \zeta^2 \right).$$

Therefore, for our choice of c_∞ we have that (3.2) satisfies the quadratic growth condition for feasible points (x, ζ). Since (3.2) clearly satisfies MFCQ everywhere, this is equivalent to the quadratic growth condition (1.19) holding for all (x, ζ) in a neighborhood of $(x^*, 0)$ [7, 8]. The proof of part 1 of the theorem is complete.

From the conclusion of part 1 we have that, since MFCQ and the quadratic growth condition holds for (3.2) and, respectively, (3.3), at $(x^*, 0)$ and, respectively, $(x^*, 0_m, 0_n, 0_n)$, these points must be isolated stationary points of the respective nonlinear programs [1]. Therefore any algorithm with global convergence safeguards that does not leave their neighborhood, will converge to them. This concludes the proof of part 2.

Part 3 immediately follows since, if the initial c_1 is chosen larger than the c_1 obtained in part 1, the update rule will not even need to be triggered, provided that we start in a sufficiently small neighborhood of $(x^*, 0_m, 0_n, 0_m)$. $\diamond$

Discussion

- Note that the conditions used in the Theorem are fairly weak. The quadratic growth is the weakest possible second-order sufficient condition. Relaxing our Lagrange multiplier requirement would result in a problem with an empty Lagrange multiplier set, for which few regularity results are known that could be algorithmically useful.

- The proof of part 1 will work for any choice $c_\infty > \tilde{c}_\infty$, for a possibly different neighborhood of $(x^*, 0)$. When we discuss the stationarity conditions of (3.2) and (3.3) in connection to c_∞ and c_1 we will still use the lower bound (1.23).
- For part 2, an SQP algorithm with a global convergence safeguard is, for example, FilterSQP [13]. For (3.2) and (3.3) other SQP algorithms using the merit functions $\psi_\infty(x)$ and $\psi_1(x)$ will always accumulate at KKT stationary points [4, 5]. Obtaining that an algorithm will not leave a neighborhood of a solution point depends on the properties of the merit function used. Near a point that satisfies the quadratic growth condition and MFCQ, this is achievable for certain SQP algorithms that use nondifferentiable exact merit functions, such as the one described in [1].
- The adaptive elastic mode presented in Table 3.1 has the potential of choosing an appropriate value of the penalty parameter c_1 without input from the user. However, determining the appropriate initial range of the penalty parameter is a function of the problem. In particular, it depends on the behavior of the solution of (3.3) when c_1 is smaller than the critical value that makes $(x^*, 0_m, 0_n, 0_n)$ a local solution. It may be possible that an excessively low initial choice of the penalty parameter will push the solution in a region where subsequent increases of the penalty parameter may not even result in a feasible point. In this work we do not discuss the appropriate initial range, we merely prove that it exists.

We now apply Theorem 3.1 for the case of interest in this work, MPCC. The following corollary is a simple restatement of Theorem 3.1 for (MPCC).

Corollary 3.2. *Assume that (MPCC) satisfies the following conditions, at a solution x^*:*

- *The Lagrange multiplier set of (MPCC) not empty. From Theorem 2.2, SMFCQ holding for (TNLP) is a sufficient condition for this assumption to hold.*
- *The quadratic growth condition (1.19) is satisfied at x^*.*
- *The data of (MPCC) are twice continuously differentiable.*

Then the conclusions of Theorem 3.1 hold for (MPCC).

Consequently, when started sufficiently close to a solution and with a sufficiently large penalty parameter, the adaptive elastic mode strategy presented in Table 3.1 applied to (MPCC) will end with a finite c_1 as soon as (MPCC) satisfies the quadratic growth condition and has a nonempty Lagrange multiplier set at a solution x^*. Since SMFCQ is a generic condition for (MPCC) and holds with probability 1 for instances of problems in the MPCC class [33] and the quadratic growth condition is the weakest second-order sufficient condition, the elastic mode can be expected to locally solve (MPCC) for a finite value of the parameter c_1.

As for the elastic mode nonlinear program (3.3) itself, Theorem 3.1 part 2, gives sufficient conditions for an algorithm to converge to its solution for a finite value of the penalty parameter c_1. The fact that, for the conditions stated, the solution is an isolated stationary point, makes it very likely for most algorithms of nonlinear programming to converge to x^*.

Some rate of convergence results can also be extended to the class of problems discussed here. If the matrix W of (3.1) is positive definite, then an SQP algorithm using an Armijo search in the direction $\overline{d}$ applied to either (3.2) or (3.3) will induce at least R-linear convergence of the iterates to $(x^*, 0)$ and $(x^*, 0_m, 0_n, 0_n)$, when used in conjunction with an L_∞ penalty function under the assumptions stated above [1]. An algorithm that is superlinearly convergent under the conditions stated here can

TABLE 3.2
Results obtained with MINOS

Problem	Var-Con-CC	Value	Status	Feval	Infeas
gnash14	21-13-1	-0.17904	Optimal	80	0.0
gnash15	21-13-1	-354.699	Infeasible	236	7.1E0
gnash16	21-13-1	-241.441	Infeasible	272	1.0E1
gnash17	21-13-1	-90.7491	Infeasible	439	5.3E0
gne	16-17-10	0	Infeasible	259	2.6E1
pack-rig1-8	89-76-1	0.721818	Optimal	220	0.0E0
pack-rig1-16	401-326-1	0.742102	Optimal	1460	0.0E0
pack-rig1-32	1697-1354-1	N/A	Interrupted	N/A	N/A

TABLE 3.3
Results obtained with SNOPT

Problem	Var-Con-CC	Value	Status	Feval	Elastic
gnash14	21-13-1	-0.17904	Optimal	27	Yes
gnash15	21-13-1	-354.699	Optimal	12	None
gnash16	21-13-1	-241.441	Optimal	7	None
gnash17	21-13-1	-90.7491	Optimal	9	None
gne	16-17-10	0	Optimal	10	Yes
pack-rig1-8	89-76-1	0.721818	Optimal	15	None
pack-rig1-16	401-326-1	0.742102	Optimal	21	None
pack-rig1-32	1697-1354-1	0.751564	Optimal	19	None

be used to solve (3.3). The algorithm solves, at each step, a quadratically constrained quadratic subproblem [3]. Under stronger second-order assumptions, a superlinear rate of convergence is achievable for SQP algorithms [6, 11, 19, 20, 32, 36].

3.1. Numerical Experiments. We conducted some numerical experiments on MPCCs from the collection MacMPEC of Sven Leyffer. To validate the conclusions of this work, we used two widely employed nonlinear solvers MINOS [28] and SNOPT [18]. SNOPT implements an adaptive L_1 elastic mode approach.

We considered three types of problem, all of which appear in [31]

1. Stackelberg games [31, Section 12.1], which characterize market complementarity problems in which one of the players has a temporal advantage over the others. In our table these are the `gnash` problems.
2. Generalized Nash complementarity points [31, Section 12.2]. In our table this is the `gne` problem, an instance of the problem 12.34 in [31]. This is a restricted market complementarity problem.
3. Optimum packaging problem. The problem involves designing the support of a membrane such that the area of contact between the membrane and a specified rigid obstacle is minimized, subject to the constraint that a certain region must be in contact [31, Chapter 10]. The underlying variational inequality is defined by a two-dimensional elliptic operator, which is discretized on a grid of 8×8, 16×16, and 32×32 elements, which are the problems `pack-rig` followed by the discretization index in our table.

With the exception of `gne`, all the problems have the complementarity constraints lumped together as one inequality, as in the formulation (2.17).

In the tables showing the results for MINOS and SNOPT, we indicate the number of variables, constraints, and complementarity constraints ("Var-Con-CC" in the first column), the final value of the objective function, the number of function evaluations and the final status of the run. The runs for both MINOS and SNOPT were done on the NEOS server [29] at Argonne National Laboratory. All the runs except one completed: the exception was `pack-rigid-32`, in MINOS, which we were forced to interrupt after it had been running on the World Wide Web interface of NEOS for about 8 hours.

The fact that (MPCC) does not satisfy MFCQ does not immediately result in the algorithm's running into an infeasible QP and failure. But it suggests a significant expectation that this would occur. Indeed, it can be seen that MINOS fails in more than half of the instances of MPCCs with an "infeasible" message and a large value of the measure of infeasibility. SNOPT, by contrast, solves all the problems presented in a reasonable number of iterations, needing to initiate the elastic mode for two problems as shown in the table.

We have not determined any immediate correspondence between initiating the elastic mode in SNOPT and final infeasibility of MINOS, but that is to be expected because the two algorithms are not completely equivalent in the absence of the elastic mode. However, the use of the elastic mode considerably increases the robustness of sequential quadratic programs and is guaranteed to succeed for a finite penalty parameter under the conditions discussed in this paper.

3.2. Convergence effects of the penalty parameter. An important issue when using a penalty approach is the choice of the penalty parameter. In this subsection we investigate the effect of large penalty parameters in the relaxed formulations (3.2) and (3.3), associated with (MPCC), over the region of convergence of SQP algorithms.

Consider a local minimum x^* of (MPCC). We make the following assumptions (refer to Section 2 for a description of the notation)

(B1) The first pair of constraints is strictly complementary, or $F_{1,1}(x^*) = 0$ and $F_{1,2}(x^*) < 0$ (the first pair can eventually be relabeled to make the second constraint inactive at x^*). To continue using the aggregate notation, while referring separately to the first pair of constraints, we define the following sets of indices at x^*, derived from the ones introduced in Subsection 1.6:

$$\begin{aligned} \widehat{\mathcal{I}} &= \mathcal{I} - \{(1,1)\} \\ \widehat{\overline{\mathcal{I}}} &= \overline{\mathcal{I}} - \{(1,2)\} \\ \widehat{\mathcal{K}} &= \mathcal{K} - \{1\}\,. \end{aligned} \tag{3.4}$$

The other index sets defined in Subsection 1.6 do not change their meaning and we will therefore use the same notation.

(B2) We assume that $F_\beta(x^*) \leq -c_F$, for $\beta \in \widehat{\overline{\mathcal{I}}}$, where $c_F > 0$. Therefore, with the exception of the first pair of constraints, all inactive constraints are "strongly" inactive.

(B3) The minimal singular value of

$$J = \left[\nabla_x g(x)\ \nabla_x h(x)\ \nabla_x F_{\mathcal{D}}(x)\ \nabla_x F_{\widehat{\mathcal{I}}}(x)\ \nabla_x F_{1,1}(x)\ \nabla_x F_{1,2}(x)\right],$$

is bounded bellow by σ_m, for any x. This implies that, for vectors u,v of appropriate dimensions we must have

$$Ju = v \Rightarrow ||u|| \leq \frac{1}{\sigma_m} ||v|| .$$

In particular, the same property must hold for any matrix made of a subset of the columns of J.

(B4) The values of the norms of the first and second derivatives of the data of (MPCC) are bounded above by c_D for any x.

(B5) Define the set

$$\begin{array}{rcl} \mathcal{P} = \{x \,|g(x) & \leq & 0, \\ h(x) & = & 0, \\ F_{\mathcal{D}}(x) & = & 0, \\ F_{\widehat{\mathcal{I}}}(x) & = & 0, \\ F_{1,1}(x) & = & 0, \\ F_{1,2}(x) & = & 0, \\ F_{\beta}(x) & \leq & -\frac{c_F}{2},\ \beta \in \widehat{\widehat{\mathcal{I}}}\Big\} . \end{array}$$

We assume that $\mathcal{P}$ is feasible and that there exists $\tilde{x} \in \mathcal{P}$ such that

$$(3.5) \qquad ||x^* - \tilde{x}|| \leq -c_{\mathcal{P}} F_{1,2}(x^*),$$

for some parameter $c_{\mathcal{P}} > 0$.

(B6) The linear independence of the columns of J implies that there exists $\tilde{u}$ that satisfies the following constraints at $\tilde{x}$,

$$\begin{array}{rcrl} \nabla_x g_i(\tilde{x})^T \tilde{u} & = & 0, & i \in \mathcal{A}_g(\tilde{x}) \\ \nabla_x g_i(\tilde{x})^T \tilde{u} & = & -1, & i \notin \mathcal{A}_g(\tilde{x}) \\ \nabla_x h(\tilde{x})^T \tilde{u} & = & 0, & \\ \nabla_x F_{\mathcal{D}}(\tilde{x})^T \tilde{u} & = & 0, & \\ \nabla_x F_{\widehat{\mathcal{I}}}(\tilde{x})^T \tilde{u} & = & 0, & \\ \nabla_x F_{1,1}(\tilde{x})^T \tilde{u} & = & -1, & \\ \nabla_x F_{1,2}(\tilde{x})^T \tilde{u} & = & 0. & \end{array}$$

Here $\mathcal{A}_g(\tilde{x})$ contains the indices of components of g that are active at $\tilde{x}$. From Assumption (B3) we must have that

$$||\tilde{u}|| \leq \frac{\sqrt{n_i + 1}}{\sigma_m}.$$

We assume that there exists $\epsilon_\Gamma > 0$ and a twice continuously differentiable arc $x(t)$ that is feasible for (MPCC) for $0 \leq t \leq \epsilon_\Gamma$ and that satisfies $x(0) = \tilde{x}$ and $\left.\frac{dx(t)}{dt}\right|_{t=0} = \tilde{u}$. We denote by

$$c_\Gamma = \max_{t \in [0, \epsilon_\Gamma]} \max \left\{ \left|\left| \frac{dx(t)}{dt} \right|\right|, \left|\left| \frac{d^2 x(t)}{dt^2} \right|\right| \right\}$$

We will now analyze the subclass of (MPCC) problems that satisfies the above assumptions for fixed values of $c_F, c_D, \sigma_m, c_{\mathcal{P}}, c_\Gamma, \epsilon_\Gamma$. We investigate the dependence of the relevant optimality quantities with respect to $F_{1,2}(x^*)$. The case of interest in this analysis is $F_{1,2}(x^*) \approx 0$ (almost degeneracy).

The assumptions above can be easily relaxed by considering them valid only on suitable neighborhoods of x^*. Assumptions (B1)–(B2) describe the complementarity situation at x^* for (MPCC), and state that, with exception of $F_{1,2}(x)$, all other inactive constraints are strongly inactive: for all feasible points near x^* only $F_{1,2}(x)$ can switch from inactive to active. Assumption (B3) is essentially a uniform linear independence property for the (TNLP) program associated to (MPCC). It can be relaxed to include only the gradients of the components of $g(x)$ that can be active in a certain neighborhood of x^*. Assumption (B5) immediately follows from Assumption (B3) for $|F_{1,2}(x^*)|$ sufficiently small and $c_{\mathcal{P}}$ then depends on c_D, σ_m and c_F. Note that, if all involved mappings were linear then Assumption (B5) is a simple corollary to Hoffman's Lemma [8, Theorem 2.200]. Assumption (B6) is also implied by Assumptions (B2), (B3), (B4) and (B5) and ϵ_Γ and c_Γ depend on c_D, σ_m and c_F. However, to simplify the presentation we will consider c_F, c_D, σ_m, $c_{\mathcal{P}}$, ϵ_Γ and c_Γ to be the primary parameters and we will ignore their interdependence.

Since from our assumptions, the linear independence constraint qualification holds for (TNLP) at x^*, then, from Corollary 2.3, there will exist the Lagrange multipliers $\nu \geq 0$, π, $\mu \geq 0$ and $\eta \geq 0$ of (MPCC), such that $\nu^T g(x^*) = 0$ and

$$0 = \nabla_x f(x^*) + \nabla_x g(x^*)\nu + \nabla_x h(x^*)\pi + \nabla_x F_{\mathcal{D}}(x^*)\mu_{\mathcal{D}} +$$

$$\sum_{k\in\widehat{\mathcal{K}}} \nabla_x F_{k,i(k)}(x^*)\overbrace{\left(\mu_{k,i(k)} + \eta_k F_{k,\bar{i}(k)}(x^*)\right)}^{\tilde{\mu}_{k,i(k)}} + \nabla_x F_{1,1}(x^*)\overbrace{\left(\mu_{1,1} + \eta_1 F_{1,2}(x^*)\right)}^{\tilde{\mu}_{1,1}}. \tag{3.6}$$

In this relation, we ignore multipliers η_k, for $k \in \overline{K}$, because in the previous equation they would multiply $\nabla_x \left(F_{k,1}F_{k,2}\right)(x^*)$ which is 0 for degenerate pairs.

We can immediately see that the above equation implies that $(\nu, \pi, \mu_{\mathcal{D}}, \tilde{\mu})$ are Lagrange multipliers of (TNLP) and (RNLP). Here we define $\tilde{\mu} = \left(\tilde{\mu}_{k,i(k)} \,|k \in \mathcal{K}\right)$. The first component of $\tilde{\mu}$ is $\tilde{\mu}_{1,1}$. Due to our linear independence assumption (B3), the associated nonlinear programs (RNLP) and (TNLP) have a unique Lagrange multiplier at x^* which is the same for both nonlinear programs [33]. From Assumptions (B2) and (B4) we must have that

$$||(\nu, \pi, \mu_{\mathcal{D}}, \tilde{\mu})|| \leq \frac{1}{\sigma_m} ||\nabla_x f(x^*)|| \leq \frac{c_D}{\sigma_m}. \tag{3.7}$$

We now analyze the way in which the Lagrange multipliers of (MPCC) of minimum 1 or ∞ norm can be obtained from the Lagrange multipliers of (TNLP) in this particular situation.

If $k \in \overline{\mathcal{K}}$ then $(k,1), (k,2) \in \mathcal{D}$, and we can choose, same as in Section 2, $\eta_k = 0$. The multipliers $\mu_{k,1}$ and $\mu_{k,2}$ are components of $\mu_{\mathcal{D}}$ and are the same for (TNLP), (RNLP) and (MPCC).

If $k \in \mathcal{K}$, then we have that either $\tilde{\mu}_{k,i(k)} \geq 0$ and then we can take $\mu_{k,i(k)} = \tilde{\mu}_{k,i(k)} \geq 0$, $\mu_{k,\bar{i}(k)} = 0$ and $\eta_k = 0$; or $\tilde{\mu}_{k,i(k)} < 0$ and then we take $\mu_{k,i(k)} = 0$, $\mu_{k,\bar{i}(k)} = 0$ and $0 < \eta_k = \frac{\tilde{\mu}_{k,i(k)}}{F_{k,\bar{i}(k)}(x^*)} \leq \frac{\left|\tilde{\mu}_{k,i(k)}\right|}{c_F}$. The multiplier $\mu_{k,\bar{i}(k)}$ is always 0 because it corresponds to an inactive constraint of MPCC, since $F_{k,\bar{i}(k)}(x^*) < 0$.

Either way, we obtain

$$0 \;\leq\; \eta_k, \mu_{k,1}, \mu_{k,2} \;\leq \left|\tilde{\mu}_{k,i(k)}\right| \max\left\{\frac{1}{c_F}, 1\right\}, \quad k = 2, \ldots, n_c. \tag{3.8}$$

It can be immediately seen from (3.6) that our choices for $\mu_{k,i(k)} \geq 0$ and $\eta_k \geq 0$ are not unique in order to satisfy $\tilde{\mu}_{k,i(k)} = \mu_{k,i(k)} + F_{k,\bar{i}(k)}(x^*)\eta_k$. However, it is obvious that the multiplier vector (ν, π, μ, η) has a minimal 1,2, or ∞ norm only if one of $\mu_{k,i(k)}$ and η_k are 0.

We now have two different cases, according to the value of $\tilde{\mu}_{1,1}$, the multiplier corresponding to the almost degenerate pair.

1. $\tilde{\mu}_{1,1} \geq 0$. In this case we can choose $\mu_{1,1} = \tilde{\mu}_{1,1} > 0$ and $\eta_1 = 0$. Using (3.8) and (3.7) we obtain that

   $$||(\nu, \pi, \mu, \eta)|| \leq \max\left\{\tfrac{1}{c_F}, 1\right\} ||(\nu, \pi, \mu_{\mathcal{D}}, \tilde{\mu})|| \leq \tfrac{c_D}{\sigma_m} \max\left\{\tfrac{1}{c_F}, 1\right\}. \tag{3.9}$$

   Using the inequality (1.23) and the inequalities between the ∞, 1 and 2 norms we get that whenever

   $$\tilde{c}_1, \tilde{c}_\infty \geq \frac{(n_i + n_e + 3n_c)\, c_D}{\sigma_m} \max\left\{\frac{1}{c_F}, 1\right\} \geq ||(\nu, \pi, \mu, \eta)||_1\,, \tag{3.10}$$

   the solution x^* of (MPCC) will be a stationary point of the penalty functions $\psi_1(x)$ and $\psi_\infty(x)$. In light of the discussion following Theorem 3.1 concerning the connection between $\tilde{c}_\infty$ and $\tilde{c}_1$ and the penalty parameters c_∞ and c_1 of the elastic mode, if c_F and σ_m are large, then the penalty parameter of the elastic mode need not take very large values before convergence to x^* is observed.
2. $\tilde{\mu}_{1,1} < 0$. In this case we must have that $\mu_{1,1} = 0$ and $\eta_1 = \frac{\tilde{\mu}_{1,1}}{F_{1,2}(x^*)}$. Using the equation (1.23) we get that in order for the solution x^* of (MPCC) to be a stationary point of (3.2) and (3.3), or equivalently, of the penalty functions $\psi_\infty(x)$ and $\psi_1(x)$ with $\tilde{c}_\infty = c_\infty$ and $\tilde{c}_1 = c_1$ [4, 5], we must have at least that

   $$c_1, c_\infty \geq \eta_1 = \frac{\tilde{\mu}_{1,1}}{F_{1,2}(x^*)}. \tag{3.11}$$

   If, in addition, we have that

   $$c_1, c_\infty \geq \frac{\tilde{\mu}_{1,1}}{F_{1,2}(x^*)} + \frac{c_D\,(n_i + n_e + 3n_c)}{\sigma_m} \max\left\{\frac{1}{c_F}, 1\right\} \geq ||\nu, \pi, \mu, \eta||_1\,,$$

   then the solution x^* of (MPCC) will be a stationary point of the penalty functions $\psi_1(x)$ and $\psi_\infty(x)$, and thus of the nonlinear programs (3.2) and (3.3).

Note that here we discuss only necessary bounds on c_∞ and c_1 for stationarity, because they will also be necessary for optimality. We then see that, in the second case, the minimal value of either penalty parameters necessary for the elastic mode to end with x^* must exceed $\frac{\tilde{\mu}_{1,1}}{F_{1,2}(x^*)}$. If $|F_{1,2}(x^*)|$ is very small, this means that c_1 and c_∞ must have exceedingly large values before the elastic mode approach ends with a solution. This is an undesirable effect because problems with large penalty parameters may take longer to solve. It is also perhaps surprising because, due to Assumption (B3), any (TNLP) associated to (MPCC) at a feasible point x in a neighborhood of x^* will be in fact very well conditioned, at least as far as constraints are concerned. In this sense we could talk of a well conditioned (MPCC).

However, in this case we argue that, although x^* is a local minimum, it cannot be a minimum in a sufficiently large neighborhood.

THEOREM 3.3. *Assume that at a local minimum x^* of (MPCC) assumptions (B1)-(B6) hold. Further, assume that $\tilde{\mu}_{1,1} < 0$. Define*

$$a_1 = c_{\mathcal{P}} c_D > 0 \text{ and } a_2 = c_{\mathcal{P}} c_D \left(\sqrt{n_e + n_i + 2n_c} \frac{c_D}{\sigma_m} + 1 \right) \frac{\sqrt{n_i + 1}}{\sigma_m} > 0.$$

Also define

$$t_\Gamma = \min \left\{ \epsilon_\Gamma, \ \frac{-\tilde{\mu}_{1,1}}{2 c_\Gamma c_D \left(c_\Gamma + 1 \right)} \right\}, \quad t_b = \frac{a_1}{\frac{1}{2}\eta_1 - a_2}.$$

Here η_1 is the minimal multiplier corresponding to the first complementarity constraint of (MPCC), $\eta_1 = \frac{\tilde{\mu}_{1,1}}{F_{1,2}(x^*)}$. If $0 < t_b < t_\Gamma$, then we must have that

$$f(x(t^*)) - f(x^*) < 0,$$

where $t^* = \frac{t_b + t_\Gamma}{2}$.

Proof This follows from investigating the behavior of the objective function $f(x)$ along $x(t)$. We therefore estimate $\left.\frac{df(x(t))}{dt}\right|_{t=0} = \nabla_x f(\tilde{x})^T \tilde{u}$. We get

$$\begin{array}{rl} \nabla_x f(\tilde{x})^T \tilde{u} \leq \nabla_x f(x^*)^T \tilde{u} + c_D \, ||\tilde{u}|| \, ||x^* - \tilde{x}|| & \leq \\ \nabla_x f(x^*)^T \tilde{u} - c_D c_{\mathcal{P}} \frac{\sqrt{n_i+1}}{\sigma_m} F_{1,2}(x^*). & \end{array} \tag{3.12}$$

Using (3.6) and the definition of $\tilde{u}$ we obtain that

$$\begin{array}{rl} \tilde{u}^T \nabla_x f(x^*) = -\tilde{u}^T \nabla_x g(x^*) \nu - \tilde{u}^T \nabla_x h(x^*) \pi - \tilde{u}^T \nabla_x F_{\mathcal{D}}(x^*) \mu_{\mathcal{D}} & - \\ \tilde{u}^T \nabla_x F_{\widehat{\mathcal{I}}}(x^*) \tilde{\mu}_{\widehat{\mathcal{I}}} - \tilde{u}^T \nabla_x F_{1,1}(x^*) \tilde{\mu}_{1,1} & \leq \\ -\tilde{u}^T \nabla_x g(\tilde{x}) \nu - \tilde{u}^T \nabla_x h(\tilde{x}) \pi - \tilde{u}^T \nabla_x F_{\mathcal{D}}(\tilde{x}) \mu_{\mathcal{D}} & - \\ \tilde{u}^T \nabla_x F_{\widehat{\mathcal{I}}}(\tilde{x}) \tilde{\mu}_{\widehat{\mathcal{I}}} - \tilde{u}^T \nabla_x F_{1,1}(\tilde{x}) \tilde{\mu}_{1,1} + c_D \, ||(\nu, \pi, \mu_{\mathcal{D}}, \tilde{\mu})||_1 \, ||\tilde{u}|| \, ||x^* - \tilde{x}|| & = \\ \tilde{\mu}_{1,1} + c_D \, ||(\nu, \pi, \mu_{\mathcal{D}}, \tilde{\mu})||_1 \, ||\tilde{u}|| \, ||x^* - \tilde{x}|| & \leq \\ \tilde{\mu}_{1,1} - \frac{\sqrt{n_i+1}}{\sigma_m} c_{\mathcal{P}} c_D \, ||(\nu, \pi, \mu_{\mathcal{D}}, \tilde{\mu})||_1 \, F_{1,2}(x^*). & \end{array} \tag{3.13}$$

On the other hand, from the intermediate value theorem we have that

$$f(\tilde{x}) - f(x^*) \leq c_D \, ||x^* - \tilde{x}|| \leq -c_D c_{\mathcal{P}} F_{1,2}(x^*).$$

Using now the last inequality, as well as (3.7), (3.5), (3.13) and (3.12) we obtain that

$$\begin{array}{rcl} f(\tilde{x}) - f(x^*) & \leq & -a_1 F_{1,2}(x^*) \\ \left(\nabla_x f(\tilde{x})\right)^T \tilde{u} & \leq & \tilde{\mu}_{1,1} - a_2 F_{1,2}(x^*), \end{array}$$

where $a_1 > 0$ and $a_2 > 0$ are defined in the body of the theorem. Here we used that $||\nu, \pi, \mu_{\mathcal{D}}, \tilde{\mu}||_1 \leq \sqrt{n_e + n_i + 2n_c} \frac{c_D}{\sigma_m}$, which follows from applying the inequality between the 1 and 2 norms in (3.7).

Therefore, we obtain that for $0 \leq t \leq \epsilon_\Gamma$ and for some $\hat{t} \in [0, \epsilon_\Gamma]$ we must have

$$\begin{array}{rcl} f(\tilde{x}) - f(x^*) & \leq & -a_1 F_{1,2}(x^*) \\ f(x(t)) - f(\tilde{x}) & \leq & \left(\nabla_x f(x(t))^T \left. \frac{d(x(t))}{dt} \right|_{t=0} \right) t + \\ & & \frac{1}{2} \left(\frac{d(x(t)^T)}{dt} \nabla^2_{xx} f(x(t)) \frac{d(x(t))}{dt} + \frac{d^2(x(t)^T)}{dt^2} \nabla_x f(x(t)) \right) \Big|_{t=\hat{t}} t^2 \\ & \leq & t \left(\tilde{\mu}_{1,1} - a_2 F_{1,2}(x^*) \right) + t^2 c_\Gamma c_D \left(c_\Gamma + 1 \right), \end{array}$$

which, by adding the two inequalities, implies that

$$
\begin{aligned}
f(x(t)) - f(x^*) &\leq \tfrac{1}{2}t\tilde{\mu}_{1,1} - F_{1,2}(x^*)\left(a_1 + ta_2\right) + \tfrac{1}{2}t\tilde{\mu}_{1,1} + t^2 c_\Gamma c_D \left(c_\Gamma + 1\right) \quad = \\
&= -F_{1,2}(x^*)\left(-\tfrac{1}{2}t\eta_1 + a_1 + ta_2\right) + t\left(\tfrac{1}{2}\tilde{\mu}_{1,1} + tc_\Gamma c_D\left(c_\Gamma + 1\right)\right).
\end{aligned} \tag{3.14}
$$

Recall, we work under the assumption that $\tilde{\mu}_{1,1} < 0$ and $F_{1,2}(x^*) < 0$. Using the definitions of t_Γ and t_b from the statement of the Theorem, it follows that, if $0 < t_b < t < t_\Gamma$, then we have that both terms on the right of (3.14) are negative and thus

$$f(x(t)) - f(x^*) < 0.$$

Since $t^* = \frac{t_b + t_\Gamma}{2}$ satisfies $t_b < t^* < t_\Gamma$, the conclusion follows. $\diamond$

As we argued in (3.11), we need at least that $c_1, c_\infty \geq \eta_1$ in order for x^* to be a stationary point of $\psi_1(x^*)$ and $\psi_\infty(x^*)$, and thus of (3.3) and (3.2), and the elastic mode to converge to x^* locally. This may lead to an exceedingly large value of c_1 and c_∞ when η_1 is large. However, if η_1 is large and $\tilde{\mu}_{1,1}$ is bounded away from 0, then we will have that t_b is positive and small and $0 < t_b < t_\Gamma$. In turn Theorem 3.3 implies that there exists t^* such that $x(t^*)$ is feasible and for which $f(x(t^*)) < f(x^*)$, in spite of the fact that x^* is a local minimum of (MPCC).

We therefore get that, if the penalty parameters c_1 and c_∞ need to be large under the assumptions set forth at the beginning of this section, then there exist feasible points (MPCC) of lower value than $f(x^*)$ in a neighborhood of x^* whose size is about the order of $|F_{1,2}(x^*)|$ (3.5).

This shows that choosing very large parameters c_1 and c_∞ of the elastic mode may result in convergence towards an otherwise shallow local minimum x^* in nearly degenerate cases. By strictly local standards x^* need not be a shallow minimum, as measured by the quadratic growth parameter. However, the nonsmooth nature of the complementarity constraint allows for a feasible arc $x(t)$ that starts at $\tilde{x}$ close to x^* and on which a significant decrease of the objective function can be obtained.

If (MPCC) is otherwise well conditioned, in the sense that all associated (TNLP) are well conditioned in a neighborhood of x^*, then smaller values of c_1 and c_∞ will avoid x^* (which cannot be a stationary point of the relaxed NLP for small values of the penalty parameters) and will instead converge to better minima. The equations (3.9) and (3.10) suggest that appropriate initial values of the penalty parameters, which will avoid convergence to such shallow minima x^*, should be of the order of the norm of the Lagrange multiplier of (TNLP).

3.2.1. Example of shallow minimum convergence for large penalty parameter. Consider the following mathematical program with complementarity constraints

$$
\begin{array}{rlcl}
\min & z & & \\
\text{subject to} & F_1(y,z) = z - y^2 & \leq & 0 \\
 & F_2(y,z) = z + 1 - \epsilon - (y-1)^2 & \leq & 0 \\
 & F_1(y,z)F_2(y,z) & \leq & 0
\end{array}
$$

The feasible region consists of one piece of each of the curves $F_1(y,z) = 0$ and $F_2(y,z) = 0$ and is presented in Figure 3.1 for $\epsilon = 0.3$ in the (y,z) space. The two curves intersect at the point $\tilde{x} = (0.5\epsilon, 0.25\epsilon^2)$. The problem has a local shallow minimum $x^* = (0,0)$ and a global minimum $(1, \epsilon - 1)$. The point x^* can be brought closer to degeneracy, thus needing larger penalty parameters for the elastic mode to converge to it, by taking $\epsilon > 0$ closer to 0. The Figure 3.1 also illustrates the point

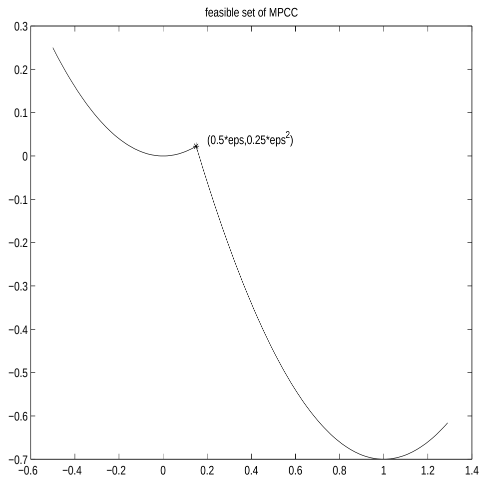

FIG. 3.1. *Example of a shallow minimum that needs a large penalty parameter*

made by and after equation (3.14) that the increase in the objective function that is encountered by going from x^* to $\tilde{x}$ on $F_1(y, z) = 0$ is followed by a sharp decrease as we switch to $F_2(y, z) = 0$. Note that the quadratic growth parameter at x^* is 1, independent of $\epsilon > 0$.

We solve this example with SNOPT, but instead of using the elastic mode, we use directly the relaxed nonlinear program (3.2). We use $\epsilon = 0.1$ and the starting point $(-0.01, 0.0001)$. For $c_\infty = 1000$ we observed that the algorithm converges to the shallow minimum x^*, whereas for $c_\infty = 10$ the algorithm converges to $(1, \epsilon - 1)$. This validates the above observation that using a smaller penalty parameter will avoid convergence to the shallow minimum.

4. Global convergence of an elastic mode approach applied to optimization of parametric mixed P nonlinear complementarity problems. In Section 3 we proved that the adaptive elastic mode approach, described in Table 3.1, applied to (MPCC), will retrieve a solution x^*, provided that it is started sufficiently close to x^* and with a sufficiently large penalty parameter c_1. As we argued before, the latter requirement cannot be relaxed in general since starting with an exceedingly low c_1 may induce the drift of the algorithm to a point from which feasibility cannot be recovered once c_1 is increased. In a special though quite important case we now show that a variation of the elastic mode can be guaranteed to retrieve a feasible C-stationary point of (MPCC).

4.1. The mixed P property. The key notion used in this section is *the mixed P property* [24, pg.277]. Before describing the special class of MPCCs to be solved, we define and prove some useful properties of partitions with the mixed P property. These will allow us, in turn, to prove a global convergence result for a special type of adaptive elastic mode approach applied to the optimization of parametric mixed P nonlinear complementarity problems.

Let $A \in \mathcal{R}^{(m+l)\times m}$, $B \in \mathcal{R}^{(m+l)\times m}$ and $C \in \mathcal{R}^{(m+l)\times l}$. We say that the partition $[A\ B\ C]$ satisfies the mixed P property if

$$\begin{aligned} 0 \neq (y, w, z) \in \mathcal{R}^{2m+l},\ Ay + Bw + Cz = 0 \Rightarrow \\ \exists i,\ 1 \leq i \leq m,\ \text{such that } y_i w_i > 0. \end{aligned} \tag{4.1}$$

Lemma 4.1. *Assume $[A\ B\ C]$ satisfies the mixed P property. Let $D \in \mathcal{R}^{m\times m}$ be a diagonal matrix such that all its diagonal entries satisfy $d_i \neq 0$, $i = 1, 2, \ldots, m$. Then the partition $[AD\ BD\ C]$ also satisfies the mixed P property.*

Proof Let $0 \neq (y, w, z) \in \mathcal{R}^{2m+l}$ such that $ADy + BDw + Cz = 0$. Let $\tilde{y} = Dy$ and $\tilde{w} = Dw$. We then have that $0 \neq (\tilde{y}, \tilde{w}, z)$ and $A\tilde{y} + B\tilde{w} + Cz = 0$. From (4.1) we obtain that $\exists i$, $1 \leq i \leq m$, such that $0 < \tilde{y}_i \tilde{w}_i = d_i^2 y_i w_i$, which in turns implies that $y_i w_i > 0$. The proof is complete. $\diamond$

Lemma 4.2. *Assume that $[A\ B\ C]$ satisfies the mixed P property. The system of linear constraints*

$$A^T\theta \leq 0, \quad B^T\theta \leq 0, \quad C^T\theta = 0$$

has the unique feasible point $\theta = 0$.

Proof Let $0 \neq y \in \mathcal{R}^m$. An immediate consequence of the fact that $[A\ B\ C]$ satisfies the mixed P property is that the matrix $[B\ C]$ is invertible [24, Prop.6.1.5]. We define $w \in \mathcal{R}^m$ and $z \in \mathcal{R}^l$ by

$$\begin{aligned} w &= -[I_m\ 0][B\ C]^{-1}Ay \\ z &= -[0\ I_l][B\ C]^{-1}Ay. \end{aligned}$$

Here we denote by I_k the $k \times k$ identity block. It can immediately be seen that (y, w, z) satisfies $Ay + Bw + Cz = 0$. Using the mixed P property of $[A\ B\ C]$ we obtain that $\exists i$, $1 \leq i \leq m$, such that $y_i w_i > 0$. Let $Q = -[I_m\ 0][B\ C]^{-1}A$. Since $w = Qy$ this means that $\forall y \neq 0$, there $\exists i$, $1 \leq i \leq m$, such that $y_i\,(Qy)_i > 0$ and thus Q is a P matrix [9, Thm.3.3.4(b),(c)]. Therefore Q^T is also a P matrix [9, Thm.3.3.4(a)(c)], where

$$Q^T = -A^T \begin{bmatrix} B^T \\ C^T \end{bmatrix}^{-1} \begin{bmatrix} I_m \\ 0 \end{bmatrix}.$$

Let now θ be a feasible point of the linear constraints in the statement of the theorem. There exist $\eta_1, \eta_2 \in \mathcal{R}^m$, $\eta_1 \geq 0$, $\eta_2 \geq 0$ such that

$$A^T\theta + \eta_1 = 0, \quad B^T\theta + \eta_2 = 0, \quad C^T\theta = 0.$$

We can solve for θ from the last 2 equations to obtain that

$$\theta = -\begin{bmatrix} B^T \\ C^T \end{bmatrix}^{-1} \begin{bmatrix} I_m \\ 0 \end{bmatrix} \eta_2. \tag{4.2}$$

Substituting in the remaining equation we get that

$$0 = \eta_1 - A^T \begin{bmatrix} B^T \\ C^T \end{bmatrix}^{-1} \begin{bmatrix} I_m \\ 0 \end{bmatrix} \eta_2,$$

which, using our definition for Q and Q^T, can be rewritten as

$$-\eta_1 = Q^T\eta_2.$$

From the definition of a P matrix, it follows that, if $\eta_2 \neq 0$ there exists i, where $1 \leq i \leq m$, such that $-\eta_{1,i}\eta_{2,i} > 0$, or $\eta_{1,i}\eta_{2,i} < 0$. This would contradict the fact that $\eta_1 \geq 0$ and $\eta_2 \geq 0$. The only alternative remains $\eta_1 = \eta_2 = 0$. From (4.2) this results in $\theta = 0$ which proves our claim. $\diamond$

4.2. Optimization of parameterized mixed P variational inequalities. We now define the following mathematical program with complementarity constraints together with its relaxed version

$$
\tag{4.3}
\begin{array}{rcl}
 & (MPEC) & \\
\min_{x,y,w,z} & f(x,y,w,z) & \\
\text{sbj. to} & g(x) & \leq 0 \\
 & h(x) & = 0 \\
 & F(x,y,w,z) & = 0 \\
 & y, w & \leq 0 \\
 & (y^T w = 0) \;\; y^T w & \leq 0
\end{array}
\qquad
\begin{array}{rcl}
 & (MPEC(c)) & \\
\min_{x,y,w,z,\zeta} & f(x,y,w,z) + c\zeta & \\
\text{sbj. to} & g(x) & \leq 0 \\
 & h(x) & = 0 \\
 & F(x,y,w,z) & = 0 \\
 & y, w & \leq 0 \\
 & y^T w & \leq \zeta \\
 & \zeta & \geq 0.
\end{array}
$$

The last constraint of (MPEC) can be formulated as either an equality or an inequality constraint without altering the feasible set.

Here $x \in \mathcal{R}^n$, $y, w \in \mathcal{R}^m$, $z \in \mathcal{R}^l$, $f : \mathcal{R}^{n+2m+l} \to \mathcal{R}$, $h : \mathcal{R}^n \to \mathcal{R}^{n_h}$, $g : \mathcal{R}^n \to \mathcal{R}^{n_g}$, $F : \mathcal{R}^{n+2m+l} \to \mathcal{R}^{m+l}$. In (MPEC(c)) we relax only the complementarity constraints $y^T w \leq 0$. This approach is different from the one in Section 3 where all constraints are relaxed.

For fixed x, the system of generalized equations

$$
\tag{4.4}
F(x,y,w,z) = 0, \;\; y \leq 0, \;\; w \leq 0, \;\; w^T y = 0
$$

defines a mixed nonlinear complementarity problem. We can therefore interpret y, w, z as the state variables and x as the parameters of the parameterized nonlinear complementarity problem (4.4). Due to this particular structure of the constraints the first problem from (4.3) is called mathematical problem with equilibrium constraints or MPEC [24].

For the remainder of this section we make the following assumptions:

(A1) The mappings f, g, h, F are twice continuously differentiable.

(A2) The constraints involving only the parameters x satisfy an MFCQ type condition:

$$
\begin{array}{l}
\nabla_x h(x) \text{ has full column rank and } \exists p \in \mathcal{R}^n \text{ such that} \\
\nabla_x h(x)^T p = 0 \text{ and } \nabla g_i(x)^T p < 0, \; \forall i \text{ such that } g_i(x) \geq 0.
\end{array}
$$

(A3) The partitioned matrix $\left[\nabla_y F^T, \nabla_w F^T, \nabla_z F^T\right]$ satisfies the mixed P property (4.1).

An instance of the problem (4.3) which satisfies this assumptions consists of the packaging problems with rigid or flexible obstacles after the additional state constraints have been replaced by a penalty term [31, Section 10].

Note that (MPEC) is identical with the problem studied in [24] except for

1. The marginally weaker mixed P_0 property is assumed in [24]. In that case that was possible because strict feasibility was maintained at all times in a penalty interior point algorithm. Here we discuss the behavior of sequential quadratic programming algorithms which may approach the solution through the infeasible region.
2. The stronger assumption is made in [24] that the set of feasible parameters x can be described by a finite set of linear equalities and inequalities.

THEOREM 4.3. *The nonlinear program (MPEC(c)) satisfies MFCQ at any point* (x, y, z, w, ζ) *and for any value* c *of the penalty parameter.*

Proof We denote by

$$\mathcal{A}(x) = \{i \in \{1, 2, \ldots, n_g\} \ | g_i(x) \geq 0\}.$$

If MFCQ doesn't hold for (MPEC(c)) at (x, y, w, z, ζ) then, by (1.13), there exist the multipliers $\lambda \in \mathcal{R}^{n_h}$, $\mu_i \in \mathcal{R}$ for $i \in \mathcal{A}(x)$, $\theta \in \mathcal{R}^{(m+l)}$, $\eta_y \in \mathcal{R}^m$, $\eta_w \in \mathcal{R}^m$, $\eta_0 \in \mathcal{R}$, $\eta_\zeta \in \mathcal{R}$ not all of them 0 such that $\mu_i \geq 0$ for $i \in \mathcal{A}(x)$, $\eta_y \geq 0$, $\eta_w \geq 0$, $\eta_0 \geq 0$ and $\eta_\zeta \geq 0$ which satisfy

$$(4.5) \qquad \begin{array}{rcl} \nabla_x h(x)\lambda + \sum_{i\in\mathcal{A}(x)} \mu_i \nabla_x g_i(x) + \nabla_x F(x, y, w, z)\theta & = & 0 \\ \eta_y + \eta_0 w + \nabla_y F(x, y, w, z)\theta & = & 0 \\ \eta_w + \eta_0 y + \nabla_w F(x, y, w, z)\theta & = & 0 \\ \nabla_z F(x, y, w, z)\theta & = & 0 \\ \eta_0 + \eta_\zeta & = & 0. \end{array}$$

Since $\eta_0, \eta_\zeta \geq 0$, the last equation implies that $\eta_0 = \eta_\zeta = 0$. Replacing $\eta_0 = 0$ in the other equations, and using that $\eta_w, \eta_y \geq 0$, we obtain that θ must satisfy

$$\begin{array}{rcl} \nabla_y F(x, y, w, z)\theta & \leq & 0 \\ \nabla_w F(x, y, w, z)\theta & \leq & 0 \\ \nabla_z F(x, y, w, z)\theta & = & 0. \end{array}$$

Using assumption (A3) and Lemma 4.2 we obtain that $\theta = 0$. Replacing $\theta = 0$ in (4.5) we get that $\eta_w = \eta_y = 0$ and that

$$\nabla_x h(x)\lambda + \sum_{i\in\mathcal{A}(x)} \mu_i \nabla_x g_i(x) = 0$$

where λ and μ_i for $i \in \mathcal{A}(x)$ cannot all be equal to 0, which contradicts assumption (A2), by (1.13). This completes the proof of the result. $\diamond$

THEOREM 4.4. *Let* $(x_n, y_n, w_n, z_n, \zeta_n)$ *be a stationary point of (MPEC(c_n)). If* $\lim_{n\to\infty} c_n = \infty$ *then any accumulation point* $(x^*, y^*, w^*, z^*, \zeta^*)$ *of the sequence* $(x_n, y_n, w_n, z_n, \zeta_n)$ *must satisfy* $\zeta^* = 0$ *and* (x^*, y^*, w^*, z^*) *is a feasible C-stationary point of (MPCC).*

Proof: Feasibility. From our assumption, $(x_n, y_n, z_n, w_n, \zeta_n)$ is a stationary point of (MPEC(c_n)), which, by Theorem 4.3, satisfies MFCQ everywhere. There must exist the Lagrange multipliers $\lambda^n \in \mathcal{R}^{n_h}$, $\mu^n \in \mathcal{R}^{n_g}$, $\theta^n \in \mathcal{R}^{(m+l)}$, $\eta_y^n \in \mathcal{R}^m$, $\eta_w^n \in \mathcal{R}^m$, $\alpha_1^n, \alpha_2^n \in \mathcal{R}$ such that $\mu^n \geq 0$, $\eta_y^n \geq 0$, $\eta_w^n \geq 0$ and $\alpha_1^n, \alpha_2^n \geq 0$, that, together with $(x_n, y_n, w_n, z_n, \zeta_n)$, satisfy the KKT conditions (1.8), part of which include the following equations

$$(4.6) \qquad \begin{array}{rcl} \nabla_x f(x_n, y_n, w_n, z_n) + \nabla_x h(x_n)\lambda^n + & & \\ \nabla_x g(x_n)\mu^n + \nabla_x F(x_n, y_n, w_n, z_n)\theta^n & = & 0 \\ \nabla_y f(x_n, y_n, w_n, z_n) + \alpha_1^n w_n + \eta_y^n + \nabla_y F(x_n, y_n, w_n, z_n)\theta^n & = & 0 \\ \nabla_w f(x_n, y_n, w_n, z_n) + \alpha_1^n y_n + \eta_w^n + \nabla_w F(x_n, y_n, w_n, z_n)\theta^n & = & 0 \\ \nabla_z f(x_n, y_n, w_n, z_n) + \nabla_z F(x_n, y_n, w_n, z_n)\theta^n & = & 0 \\ \alpha_2^n + \alpha_1^n & = & c_n \\ g(x_n) \leq 0, \ \ y_n \leq 0, \ \ w_n \leq 0, \ \ (w_n^T y_n - \zeta_n) \leq 0, \ \ \zeta_n & \geq & 0 \\ g(x_n)^T \mu^n = 0, \ \ y_n^T \eta_y^n = 0, \ \ w_n^T \eta_w^n = 0, \ \ \alpha_1^n (w_n^T y_n - \zeta_n) = 0, \ \ \alpha_2^n \zeta_n & = & 0. \end{array}$$

Let now

$$\widetilde{\lambda}^n = \left(\lambda^n, \mu^n, \theta^n, \eta_y^n, \eta_w^n, \alpha_1^n, \alpha_2^n\right)$$

Since $\alpha_1^n + \alpha_2^n = c^n$, and $c^n \to \infty$ we must have that $\left\|\widetilde{\lambda}^n\right\|_\infty \to \infty$ as $n \to \infty$. Therefore the sequence $\frac{\widetilde{\lambda}^n}{\left\|\widetilde{\lambda}^n\right\|_\infty}$, admits an accumulation point

$$\widetilde{\lambda}^* = \left(\lambda^*, \mu^*, \theta^*, \eta_y^*, \eta_w^*, \alpha_1^*, \alpha_2^*\right)$$

that satisfies $\left\|\widetilde{\lambda}^*\right\|_\infty = 1$ and $\mu^* \geq 0$, $\eta_y^* \geq 0$, $\eta_w^* \geq 0$, $\alpha_1^* \geq 0$ and $\alpha_2^* \geq 0$. We can assume without loss of generality (after eventually restricting the respective sequences to subsequences) that

$$\frac{\widetilde{\lambda}^n}{\left\|\widetilde{\lambda}^n\right\|_\infty} \to \widetilde{\lambda}^* \text{ and } (x_n, y_n, w_n, z_n, \zeta_n) \to (x^*, y^*, w^*, z^*, \zeta^*).$$

We now divide (4.6) by $\left\|\tilde{\lambda}^n\right\|_\infty$ and take the limit as $n \to \infty$, to obtain

$$\begin{array}{rcl}
\nabla_x h(x^*)\lambda^* + \nabla_x g(x^*)\mu^* + \nabla_x F(x^*, y^*, w^*, z^*)\theta^* & = & 0 \\
\alpha_1^* w^* + \eta_y^* + \nabla_y F(x^*, y^*, w^*, z^*)\theta^* & = & 0 \\
\alpha_1^* y^* + \eta_w^* + \nabla_w F(x^*, y^*, w^*, z^*)\theta^* & = & 0 \\
\nabla_z F(x^*, y^*, w^*, z^*)\theta^* & = & 0 \\
g(x^*) \leq 0, \;\; y^* \leq 0, \;\; w^* \leq 0, \;\; ({w^*}^T y^* - \zeta^*) \leq 0, \;\; \zeta^* & \geq & 0 \\
g(x^*)^T \mu^* = 0, \;\; {y^*}^T \eta_y^* = 0, \;\; {w^*}^T \eta_w^* = 0, \;\; \alpha_1^*({w^*}^T y^* - \zeta^*) = 0, \;\; \alpha_2^* \zeta^* & = & 0.
\end{array} \tag{4.7}$$

Obviously, $\mu^* \geq 0$, $g(x^*) \leq 0$ and $g(x^*)^T \mu^* = 0$ imply that $\mu_i^* = 0$ whenever $1 \leq i \leq n_g$ and $i \notin \mathcal{A}(x^*)$.

Take now an index j such that $1 \leq j \leq m$. Since $\alpha_1^* \geq 0$, $w_j^* \leq 0$, $y_j^* \leq 0$, $\eta_{w,j}^* \geq 0$, and $\eta_{y,j}^* \geq 0$, we must have that

$$\eta_{y,j}^* + \alpha_1^* w_j^* > 0 \Rightarrow \eta_{y,j}^* > 0 \overset{(4.7)}{\Longrightarrow} y_j^* = 0 \Rightarrow \eta_{w,j}^* + \alpha_1^* y_j^* \geq 0.$$

Similarly $\eta_{w,j}^* + \alpha_1^* y_j^* > 0 \Rightarrow \eta_{y,j}^* + \alpha_1^* w_j^* \geq 0$. We therefore conclude that for $j = 1, 2, \ldots, m$ we must have that

$$\left(\eta_{w,j}^* + \alpha_1^* y_j^*\right)\left(\eta_{y,j}^* + \alpha_1^* w_j^*\right) \geq 0. \tag{4.8}$$

We can therefore define for $j = 1, 2, \ldots, m$ the quantities

$$d_j = \begin{cases} 1 & \text{if} \quad \left(\eta_{w,j}^* + \alpha_1^* y_j^*\right) > 0 \text{ or } \left(\eta_{y,j}^* + \alpha_1^* w_j^*\right) > 0 \\ -1 & \text{if} \quad \left(\eta_{w,j}^* + \alpha_1^* y_j^*\right) < 0 \text{ or } \left(\eta_{y,j}^* + \alpha_1^* w_j^*\right) < 0 \\ 1 & \text{if} \quad \left(\eta_{y,j}^* + \alpha_1^* w_j^*\right) = \left(\eta_{w,j}^* + \alpha_1^* y_j^*\right) = 0. \end{cases}$$

From our observation and the definition of d_j we must have

$$d_j\left(\eta_{y,j}^* + \alpha_1^* w_j^*\right) \geq 0, \quad d_j\left(\eta_{w,j}^* + \alpha_1^* y_j^*\right) \geq 0, \quad j = 1, 2, \ldots, m.$$

Denote now by $D \in \mathcal{R}^{m\times m}$ the matrix whose diagonal elements are $d_j, j = 1, 2, \ldots, m$. The middle equations from (4.7) and our definition of D imply that

$$\begin{aligned} D\nabla_y F(x^*, y^*, w^*, z^*)\theta^* &\le 0,\\ D\nabla_w F(x^*, y^*, w^*, z^*)\theta^* &\le 0,\\ \nabla_z F(x^*, y^*, w^*, z^*)\theta^* &= 0. \end{aligned}$$

Using now assumption (A3), and Lemmas 4.1 and 4.2 this implies that $\theta^* = 0$. Replacing this in (4.7), we obtain that

$$\nabla_x h(x^*)\lambda^* + \sum_{i\in\mathcal{A}(x^*)} \mu_i^* \nabla_x g_i(x^*) = 0,$$

which, using assumption (A2) implies that $\lambda^* = 0$ and $\mu^* = 0$. The fact that $\theta^* = 0$ also implies from (4.7) that

$$\eta_y^* + \alpha_1^* w^* = 0, \;\; \eta_w^* + \alpha_1^* y^* = 0. \tag{4.9}$$

Multiplying the first relation with ${y^*}^T$ and the second one with ${w^*}^T$ and using the complementarity relations ${y^*}^T \eta_y^* = 0$ and ${w^*}^T \eta_w^* = 0$ from (4.7) we obtain that

$$\alpha_1^* {y^*}^T w^* = 0. \tag{4.10}$$

We have the following cases.

1. $\alpha_1^* > 0$. Then (4.10) implies that ${y^*}^T w^* = 0$. From the equation $\alpha_1^*({w^*}^T y^* - \zeta^*) = 0$ of (4.7) we get that $\zeta^* = {y^*}^T w^* = 0$.
2. $\alpha_1^* = 0$. Then from (4.9) we get that $\eta_y^* = \eta_w^* = 0$. It then follows that the only nonzero component of $\tilde{\lambda}^*$ is α_2^* which must then satisfy $\alpha_2^* = \left\|\tilde{\lambda}^*\right\|_\infty = 1$. The complementarity condition $\alpha_2^*\zeta^* = 0$ from (4.7) now implies $\zeta^* = 0$.

In either case we obtain $\zeta^* = 0$ which shows that the limit point (x^*, y^*, w^*, z^*) must be feasible.

Proof: C-stationarity. We return to the equation (4.6). We define

$$\widehat{\eta}_y^n = \eta_y^n + \alpha_1^n w_n, \; \widehat{\eta}_w^n = \eta_w^n + \alpha_1^n y_n. \tag{4.11}$$

Following the same argument that led to (4.8), we obtain that

$$\widehat{\eta}_{y,j}^n \widehat{\eta}_{w,j}^n \ge 0, \; j = 1, 2, \ldots, m. \tag{4.12}$$

Define now

$$\widehat{\lambda}^n = \left(\lambda^n, \mu^n, \theta^n, \widehat{\eta}_y^n, \widehat{\eta}_w^n\right).$$

The components of $\widehat{\lambda}^n$ satisfy a set of equations derived from (4.6):

$$\begin{aligned} \nabla_x f(x_n, y_n, w_n, z_n) + \nabla_x h(x_n)\lambda^n + &\\ \nabla_x g(x_n)\mu^n + \nabla_x F(x_n, y_n, w_n, z_n)\theta^n &= 0\\ \nabla_y f(x_n, y_n, w_n, z_n) + \widehat{\eta}_y^n + \nabla_y F(x_n, y_n, w_n, z_n)\theta^n &= 0\\ \nabla_w f(x_n, y_n, w_n, z_n) + \widehat{\eta}_w^n + \nabla_w F(x_n, y_n, w_n, z_n)\theta^n &= 0\\ \nabla_z f(x_n, y_n, w_n, z_n) + \nabla_z F(x_n, y_n, w_n, z_n)\theta^n &= 0\\ g(x_n) \le 0, \;\; g(x_n)^T\mu^n = 0, \;\; y_n \le 0, \;\; w_n \le 0.& \end{aligned} \tag{4.13}$$

Assume now that $\widehat{\lambda}^n$ admits a subsequence that diverges to ∞. We can assume without loss of generality that the entire sequence itself diverges to ∞. Define now the sequence

$$\widetilde{\widehat{\lambda}}^n = \frac{\widehat{\lambda}^n}{\left\|\widehat{\lambda}^n\right\|_\infty},$$

which, being bounded, must admit a convergent subsequence. We assume, again without loss of generality, that the sequence $\widetilde{\widehat{\lambda}}^n$ is itself convergent to

$$\widetilde{\widehat{\lambda}}^* = \left(\widetilde{\lambda}^*, \widetilde{\mu}^*, \widetilde{\theta}^*, \widetilde{\eta}_y^*, \widetilde{\eta}_w^*\right),$$

with $\left\|\widetilde{\widehat{\lambda}}^*\right\|_\infty = 1$. From the construction of $\widehat{\lambda}^n$ we must have that $\widetilde{\mu}^* \geq 0$, whereas from (4.12) we must have that

$$\widetilde{\eta}_{y,j}^* \widetilde{\eta}_{w,j}^* \geq 0,\ j = 1, 2, \ldots, m. \tag{4.14}$$

Dividing now all equations involving multipliers of (4.13) by $\left\|\widehat{\lambda}^n\right\|_\infty$ and taking the limit as $n \to \infty$, we obtain that

$$\begin{array}{rcl} \nabla_x h(x^*)\widetilde{\lambda}^* + \nabla_x g(x^*)\widetilde{\mu}^* + \nabla_x F(x^*, y^*, w^*, z^*)\widetilde{\theta}^* & = & 0 \\ \widetilde{\eta}_y^* + \nabla_y F(x^*, y^*, w^*, z^*)\widetilde{\theta}^* & = & 0 \\ \widetilde{\eta}_w^* + \nabla_w F(x^*, y^*, w^*, z^*)\widetilde{\theta}^* & = & 0 \\ \nabla_z F(x^*, y^*, w^*, z^*)\widetilde{\theta}^* & = & 0 \\ g(x^*) \leq 0,\ \ g(x^*)^T \widetilde{\mu}^* = 0,\ \ y^* \leq 0,\ \ w^* \leq 0. & & \end{array} \tag{4.15}$$

Using now the exact same argument that we applied to (4.7), and which led to the conclusion that $\theta^* = 0$ and, subsequently, $\zeta^* = 0$, we get that (4.15), (4.14) and Assumption (A3) imply that $\widetilde{\theta}^* = 0$. In turn, this implies that $\widetilde{\eta}_y^* = \widetilde{\eta}_w^* = 0$ and, from Assumption (A2) and using the complementarity relation on the last line of (4.15), that $\widetilde{\lambda}^* = 0$, $\widetilde{\mu}^* = 0$ and thus $\widetilde{\widehat{\lambda}}^* = 0$, which is a contradiction with $\left\|\widetilde{\widehat{\lambda}}^*\right\|_\infty = 1$. This implies that the sequence $\widehat{\lambda}^n$ must be bounded. Let

$$\widehat{\lambda}^* = \left(\lambda^*, \mu^*, \theta^*, \widehat{\eta}_y^*, \widehat{\eta}_w^*\right)$$

be a limit point of this sequence. We assume without loss of generality that it is the unique limit point. From (4.12) we must have that

$$\widehat{\eta}_{y,j}^* \widehat{\eta}_{w,j}^* \geq 0,\ \ j = 1, 2, \ldots, m. \tag{4.16}$$

From our definition of $\widehat{\eta}_w^n$ and $\widehat{\eta}_y^n$ (4.11), it does not immediately follow that the corresponding limit point satisfy a complementarity relation with w^* and, respectively, y^*. Although we have that $\eta_{w,j}^n w_{n,j} = 0$ and $\eta_{y,j}^n y_{n,j} = 0$, for $j = 1, 2, \ldots, m$ from (4.6), the additional terms $\alpha_1 y_{n,j}$ and $\alpha_1 w_{n,j}$ may potentially prevent a corresponding complementarity relation from holding for $\widehat{\eta}_w^n$ and $\widehat{\eta}_y^n$, or the respective limits, since α_1^n may diverge to ∞. In the following we prove that that is not the case.

We are going to show that, if $y_j^* < 0$, for some j among $1, 2, \ldots, m$, then $\widehat{\eta}_{y,j}^* = 0$. Due to the fact that (x^*, y^*, w^*, z^*) is feasible for (MPEC), as we proved in the first part of this Theorem, we must have that $w_j^* y_j^* = 0$ and thus $w_j^* = 0$. We also must have that $y_{n,j} < 0$ for all n sufficiently large, and thus, from the complementarity constraints in (4.6) we also have that $\eta_{y,j}^n = 0$. We have the following cases:

1. The sequence $w_{n,j}$ has a nonzero sequence $w_{n_k,j} < 0$, $k = 1, 2, \ldots$. The complementarity constraints in (4.6) imply that $\eta_{w,j}^{n_k} = 0$, and thus, from (4.11) we get that $\widehat{\eta}_{w,j}^{n_k} = \alpha_1^{n_k} y_{n_k,j}$. Since, per our assumption, $\widehat{\eta}_w^{n_k}$ is convergent, it follows that the subsequence $\widehat{\eta}_{w,j}^{n_k} = \alpha_1^{n_k} y_{n_k,j}$ must be bounded. We therefore get that
$$\left|\widehat{\eta}_{y,j}^{n_k}\right| = \left|\alpha_1^{n_k} w_{n_k,j}\right| = \left|\alpha_1^{n_k} y_{n_k,j}\right| \frac{w_{n_k,j}}{y_{n_k,j}} = \left|\widehat{\eta}_{w,j}^{n_k}\right| \frac{w_{n_k,j}}{y_{n_k,j}} \overset{k\to\infty}{\longrightarrow} \left|\widehat{\eta}_{w,j}^{*}\right| \frac{w_j^*}{y_j^*} = 0.$$
Since $\widehat{\eta}_y^n$ is a convergent sequence this means that $\widehat{\eta}_{y,j}^* = 0$.
2. For all n sufficiently large we must have $w_{n,j} = 0$. From (4.11) this implies that $\widehat{\eta}_{y,j}^n = 0$, which in turn implies that $\widehat{\eta}_{y,j}^* = 0$.

Either way, we see that the conclusion becomes that $\widehat{\eta}_{y,j}^*$=0. We reach a similar conclusion that if $w_j^* < 0$ for some $j = 1, 2, \ldots, m$, then $\widehat{\eta}_{w,j}^* = 0$. We therefore obtain that for any j among $1, 2, \ldots, m$ we must have that

$$w_j^* < 0 \Rightarrow \widehat{\eta}_{w,j}^* = 0; \quad y_j^* < 0 \Rightarrow \widehat{\eta}_{y,j}^* = 0. \tag{4.17}$$

Taking now the limit in (4.13) as $n \to \infty$ we obtain that (x^*, y^*, w^*, z^*) is feasible from the first part of the proof and that, together with $\widehat{\lambda}^*$, it satisfies the equations

$$\begin{array}{rcl}
\nabla_x f(x^*, y^*, w^*, z^*) + \nabla_x h(x^*)\lambda^* + & & \\
\nabla_x g(x^*)\mu^* + \nabla_x F(x^*, y^*, w^*, z^*)\theta^* & = & 0 \\
\nabla_y f(x^*, y^*, w^*, z^*) + \widehat{\eta}_y^* + \nabla_y F(x^*, y^*, w^*, z^*)\theta^* & = & 0 \\
\nabla_w f(x^*, y^*, w^*, z^*) + \widehat{\eta}_w^* + \nabla_w F(x^*, y^*, w^*, z^*)\theta^* & = & 0 \\
\nabla_z f(x^*, y^*, w^*, z^*) + \nabla_z F(x^*, y^*, w^*, z^*)\theta^* & = & 0, \\
g(x^*) \leq 0, \ \ \mu^* \geq 0, \ \ g(x^*)^T \mu^* & = & 0.
\end{array} \tag{4.18}$$

The last line of equations and inequalities implies that $\mu_i^* = 0$ whenever $i \notin \mathcal{A}(x^*)$.

From equations (4.16), (4.17) and (4.18), and using the conclusion of the feasibility part of the proof, we get that the point (x^*, y^*, w^*, z^*) is a C-stationary point [21] with associated multiplier $\widehat{\lambda}^* = \left(\lambda^*, \mu^*, \theta^*, \widehat{\eta}_y^*, \widehat{\eta}_w^*\right)$: It satisfies (2.5) and (2.6) with $\alpha = 1$ and where the requirements $\mu_{k,1} \geq 0$, $\mu_{k,2} \geq 0$, for $k \in \overline{K}$ have been relaxed to $\mu_{k,1}\mu_{k,2} \geq 0$, for $k \in \overline{K}$. ⋄

The preceding result also allows us to characterize all local solutions of MPEC.

Corollary 4.5. *Assume that (MPEC) satisfies assumptions (A1), (A2) and (A3) everywhere and that (x^*, y^*, w^*, z^*) is a strict local minimum of (MPEC). Then (x^*, y^*, w^*, z^*) is a C-stationary point of (MPEC).*

Proof It is immediate from the definition of (MPEC(c)) that $(x^c, y^c, w^c, z^c, \zeta^c)$ is a local solution of (MPEC(c)) if and only if (x^c, y^c, w^c, z^c) is a local solution of

$$(MPEC1(c)) \qquad \begin{array}{rll}
\min_{x,y,w,z} & f(x,y,w,z) + cy^T w & \\
\text{sbj. to} & g(x) & \leq 0 \\
& h(x) & = 0 \\
& F(x,y,w,z) & = 0 \\
& y, w & \leq 0.
\end{array}$$

Choose some $c_0 > 0$, $n = 0$

MPEC1: *Find a solution (stationary point)* $(x^{c_n}, y^{c_n}, w^{c_n}, z^{c_n}, \zeta^{c_n})$ *of (MPEC(*c_n*)).*

If $\zeta^{c_n} = 0$, then $(x^{c_n}, y^{c_n}, w^{c_n}, z^{c_n})$ *solves (MPEC).* Stop.

otherwise *update c:* $c_{n+1} = c_n + K$ *and n:* $n = n + 1$ *and return to* MPEC1.

TABLE 4.1

An adaptive L_1 *modified elastic mode approach*

If $\widehat{x} = (x^*, y^*, w^*, z^*)$ is a strict local minimum of (MPEC), then there exist $\delta > 0$ and a ball $B(\widehat{x}, \delta)$, whose boundary we denote by Γ, such that for any $(x, y, w, z) \in \Gamma$, a feasible point of (MPEC1(c)), we must have that

$$\max\{f(x, y, w, z) - f(x^*, y^*, w^*, z^*),\ y^T w\} > 0.$$

This implies that there exists $\widehat{c}$ such that, for all $\gamma > \widehat{c}$ we have that for any (x, y, w, z), a feasible point of (MPEC1(c)) on the boundary Γ of $B(\widehat{x}, \delta)$, we must have that

$$f(x, y, w, z) - f(x^*, y^*, w^*, z^*) + \gamma y^T w > 0.$$

If this were not true, then for any n there exists $\gamma_n > n$ such that, for some $(x^{\circ,n}, y^{\circ,n}, w^{\circ,n}, z^{\circ,n}) \in \Gamma$, a feasible point of (MPEC1(c)), we have that

$$f(x^{\circ,n}, y^{\circ,n}, w^{\circ,n}, z^{\circ,n}) - f(x^*, y^*, w^*, z^*) + \gamma_n {y^{\circ,n}}^T w^{\circ,n} \leq 0. \tag{4.19}$$

Since Γ is compact, the sequence $(x^{\circ,n}, y^{\circ,n}, w^{\circ,n}, z^{\circ,n})$ has an accumulation point $(x^\circ, y^\circ, w^\circ, z^\circ) \in \Gamma$ that must be feasible for (MPEC1(c)). Dividing (4.19) by γ_n and taking the limit as $n \to \infty$, we get that ${y^\circ}^T w^\circ = 0$, or that $(x^\circ, y^\circ, w^\circ, z^\circ)$ is in effect feasible for (MPEC). But (4.19) also implies that, for all n,

$$f(x^{\circ,n}, y^{\circ,n}, w^{\circ,n}, z^{\circ,n}) - f(x^*, y^*, w^*, z^*) \leq 0.$$

Taking the limit in the last inequality we obtain that

$$f(x^\circ, y^\circ, w^\circ, z^\circ) - f(x^*, y^*, w^*, z^*) \leq 0.$$

which contradict our choice of δ.

Therefore, $\widehat{c}$ with the properties specified above must exist. This shows that, for $c > \widehat{c}$, (MPEC1(c)) will have a local solution inside of $B(\widehat{x}, \delta)$. For all $n > \widehat{c}$ let (x^n, y^n, w^n, z^n) be the local solution of (MPEC1(n)) in $B(\widehat{x}, \delta)$ with the lowest value. By an argument similar to the one that lead to the existence of $\widehat{c}$ it follows that $(x^n, y^n, w^n, z^n) \to (x^*, y^*, w^*, z^*)$. It also follows from the observation at the beginning of the proof that $(x^n, y^n, w^n, z^n, {y^n}^T w^n)$ is a local solution, and thus a stationary point, of (MPEC(n)). From Theorem 4.4 it thus follows that (x^*, y^*, w^*, z^*) is a C-stationary point of (MPEC). The proof is complete. $\diamond$

4.3. A globally convergent modified elastic mode for the optimization of parameterized mixed P variational inequalities. The results from Subsection 4.2 allow us to define a modified elastic mode approach with good global convergence properties for the optimization of parameterized mixed P variational inequalities. The modification is described in Table 4.1.

THEOREM 4.6. *Consider the algorithm described in Table 4.1. Assume that, for a fixed c_n, the subproblem (MPEC(c_n)) is solved with a nonlinear programming algorithm with global convergence safeguards which does not diverge to ∞ and produces $(x^{c_n}, y^{c_n}, w^{c_n}, z^{c_n}, \zeta^{c_n})$. Then either*

1. *the algorithm stops at a finite n with $\zeta^{c_n} = 0$ and $(x^{c_n}, y^{c_n}, w^{c_n}, z^{c_n})$ is a B-stationary (KKT) point of (MPEC); or*
2. *$\zeta^{c_n} > 0$, $\forall n$ and any accumulation point (x^*, y^*, w^*, z^*) of $(x^{c_n}, y^{c_n}, w^{c_n}, z^{c_n})$ is a C-stationary point of (MPEC).*

Proof Since, for a fixed value c_n, the NLP algorithm does not diverge, it will have an accumulation point $(x^{c_n}, y^{c_n}, w^{c_n}, z^{c_n}, \zeta^{c_n})$. Since, from Theorem 4.3, MFCQ holds at every point of (MPEC(c_n)) and the NLP algorithm has global convergence safeguards, it follows that cases A) and B) in Subsection 1.4 cannot occur, and thus $(x^{c_n}, y^{c_n}, w^{c_n}, z^{c_n}, \zeta^{c_n})$ is a KKT point of (MPEC(c_n)).

If the algorithm ends with a finite n and, thus, a finite value of the penalty c_n and $\zeta^{c_n} = 0$, from the fact that $(x^{c_n}, y^{c_n}, w^{c_n}, z^{c_n}, 0)$ is a KKT point of (MPEC(c_n)) it follows that $(x^{c_n}, y^{c_n}, w^{c_n}, z^{c_n})$ is in effect a KKT point of MPEC, which proves part 1.

If $\zeta^{c_n} > 0$ for any n, then c_n is increased to ∞, and, by applying Theorem 4.4, we get that any accumulation point (x^*, y^*, w^*, z^*) of $(x^{c_n}, y^{c_n}, w^{c_n}, z^{c_n})$ is a C-stationary point of (MPEC). The proof is complete.

$\diamond$

We can therefore claim that the adaptive modified elastic mode in Table 4.1 is globally convergent: Any accumulation point produced by this algorithm is at least a C-stationary point of the problem (MPEC). It is true that the desirable global convergence result would require that any limit point be a B-stationary point. Note, however, that the example (2.15) satisfies the assumptions (A1), (A2) and (A3) at $(0, 0, 0)$ (after introducing the slack variable $w = y + x$), but $(0, 0, 0)$ cannot be a KKT and thus a B-stationary point. So global convergence to B-stationary points cannot be guaranteed, in general.

5. Conclusions. A class of mathematical models, mathematical programs with complementarity constraints (MPCCs), which describe a wide variety of problems in economics and engineering cannot satisfy the Mangasarian-Fromovitz constraint qualification (MFCQ) at a solution point. Therefore, sequential quadratic programming algorithms may encounter infeasible quadratic program subproblems arbitrarily close to a solution x^*.

In this work we determine sufficient conditions for (MPCC) to have a nonempty Lagrange multiplier set. Based on results from [33], we establish that having the strict MFCQ hold for the tightened nonlinear program (TNLP) results in the MPCC's having a nonempty Lagrange multiplier set.

As a result, the strict Mangasarian-Fromovitz constraint qualification (SMFCQ) for (TNLP) and the quadratic growth condition near a solution x^* of (MPCC) imply that the adaptive elastic mode strategy as presented in Table 3.1 will terminate with a solution of the original problem for a sufficiently large but finite value of the penalty parameter, c_1, provided that it is initiated sufficiently close to the solution. For this value of c_1 the modified nonlinear program (3.3) satisfies MFCQ at all feasible points and the quadratic growth at the solution and can thus be locally solved by certain sequential quadratic programming algorithms [1]. In addition (3.3) has now an isolated stationary point at the point corresponding to the solution of (MPCC), which means that any sequential quadratic programming algorithm with global convergence safeguards that does not leave a neighborhood of the solution will in effect converge to it. We demonstrate this point by applying the adaptive elastic mode implemented in SNOPT [18] to several problems. As has been argued elsewhere [33], SMFCQ can be expected to hold at the solution of almost all MPCCs (in a measure theoretic sense). Also, the quadratic growth condition is the weakest sufficient second-order condition, and is thus the most general possible. We can therefore claim that the elastic mode approach will locally solve a generic instance of the MPCC class for a finite value of the penalty parameter.

An important issue, whenever a relaxation-penalty approach is applied, concerns the choice of the penalty parameter. In particular, the penalty parameter, though finite, may have to be very large for the relaxed problem to have the same solution as the original problem. If this happens near a local solution x^* of an MPCC we show that, under certain assumptions, there will be feasible points of MPCC that have lower objective values than x^*. This makes x^* an undesirable end point which can be avoided if the penalty parameter is not aggressively increased.

We also show that any accumulation point of an adaptive elastic mode type approach, presented in Table 4.1, is a C-stationary point of an optimization problem whose complementarity constraints originate in a parameterized mixed P variational inequality, if the penalty parameter of the elastic mode is allowed to increase to ∞. A corollary to this observation is that any strict local minimum of such problem must be a C-stationary point. Although the desirable result should involve B-stationarity, we show by an example that there exist such optimization problems that do not have B-stationary local minima.

The elastic mode therefore provides a framework for solving mathematical programs with complementarity constraints by using sequential quadratic programming algorithms. The benefit of this perspective is that a large class of algorithms, whose behavior and properties have been amply analyzed and for which sophisticated finely tuned implementations already exist, can be used to solve mathematical programs with complementarity constraints.

Acknowledgments. Thanks to Sven Leyffer for providing the AMPL examples and guidance for the numerical results. Thanks to Danny Ralph and Stefan Scholtes for the comments made on the presentation of partial results of this work at ISMP 2000 in Atlanta, and to Danny Ralph for comments on an earlier version of this paper. Thanks to Alexander Shapiro for countless discussions concerning degenerate nonlinear programming.

REFERENCES

[1] M. Anitescu, *Degenerate nonlinear programming with a quadratic growth condition.* SIAM Journal in Optimization, 10:4 (2000), pp. 1116-1135, 2000.

[2] M. Anitescu, *On the rate of convergence of sequential quadratic programming with nondifferentiable exact penalty function in the presence of constraint degeneracy* To appear in Mathematical Programming.

[3] M. Anitescu, *A superlinearly convergent sequential quadratically constrained quadratic programming algorithm for degenerate nonlinear programming.* To appear in SIAM Journal of Optimization.

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

[5] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 1995.

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

[7] J. F. Bonnans and A. Ioffe, *Second-order sufficiency and quadratic growth for nonisolated minima*, Mathematics of Operations Research , 20:4 (1995), pp. 801–819.

[8] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[9] R. W. Cottle, J.-S. Pang and R. E. Stone, *The Linear Complementarity Problem*, Academic Press, Boston, 1992.

[10] A. V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York, 1983.

[11] A. Fischer, *Modified Wilson's method for nonlinear programs with nonunique multipliers*, Mathematics of Operations Research 24:3 (1999), pp. 699–727.

[12] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Chichester, 1987.

[13] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Numerical Analysis Report, NA/171, Department of Mathematics, University of Dundee, UK.

[14] R. Fletcher, S. Leyffer and P.L. Toint *On the global convergence of a filter-SQP type algorithm*, Numerical Analysis Report, NA/183, Department of Mathematics, University of Dundee, UK.

[15] R. Fletcher and S. Leyffer, *Filter-type algorithms for solving systems of algebraic equations and inequalities* Numerical Analysis Report, NA/204, Department of Mathematics, University of Dundee, UK.

[16] M. Fukushima and P. Tseng, *An implementable active-set algorithm for computing a B-stationary point of a mathematical program with linear complementarity constraints*, to appear in SIAM J. of Optimization.

[17] J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Mathematical Programming 12 (1977), pp. 136–138.

[18] P. E. Gill, W. Murray and M. A. Saunders, *User's guide for SNOPT 5.3: A Fortran package for large-scale nonlinear programming*, Report NA 97-5, Department of Mathematics, University of California, San Diego, 1997.

[19] W. W. Hager, *Stabilized sequential quadratic programming*, To appear in Computational Optimization and Applications .

[20] W. W. Hager and M. S. Gowda, *Stability in the presence of degeneracy and error estimation*, Technical Report, Department of Mathematics, University of Florida, 1997.

[21] X. Hu and D. Ralph, *Convergence of a penalty method for mathematical programming with complementarity constraints.* Preprint.

[22] A. Ioffe, *Necessary and sufficient conditions for a local minimum.3: Second order conditions and augmented duality*, SIAM Journal of Control and Optimization, 17:2 (1979), pp. 266–288.

[23] A. Ioffe, *On Sensitivity analysis of nonlinear programs in Banach spaces: the approach via composite unconstrained optimization*, SIAM Journal of Optimization, 4:1 (1994), pp. 1–43.

[24] Z.-Q. Luo, J.-S. Pang and D. Ralph, *Mathematical Programs with Complementarity Constraints* Cambridge University Press, 1996.

[25] Z.-Q. Luo, J.-S. Pang, D. Ralph and S.-Q. Wu, *Exact penalization and stationarity conditions of mathematical programs with complementarity constraints.* Mathematical Programming, 34(2):142-162, 1986.

[26] O. L. Mangasarian, *Nonlinear Programming*, McGraw-Hill, New York 1969.

[27] O. L. Mangasarian and S. Fromovitz, *The Fritz John necessary optimality conditions in the presence of equality constraints*, J. Math. Anal. and Appl. 17 (1967), pp. 34-47.

[28] B. Murtagh and M. A. Saunders, *MINOS 5.0 user's guide*, Technical Report SOL 83-20, Stanford University, Stanford, 1983.

[29] *The NEOS Guide*. Available online at *http://www.mcs.anl.gov/otc/Guide.*

[30] E. Polak, *Optimization*, Springer, New York, 1997.

[31] J. Outrata, M. Kocvara and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Kluwer Academic Publishers, Dordrecht, 1998.

[32] D. Ralph and S. J. Wright, *Superlinear convergence of an interior-point method despite dependent constraints*, Preprint ANL/MCS-P622-1196, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., 1996.

[33] H. Scheel and S. Scholtes, *Mathematical Programs with Equilibrium Constraints: Stationarity, Optimality and Sensitivity*, Mathematics of Operations Research, 25:1 (2000), pp. 1–25.

[34] S. Scholtes, *Convergence Properties of a Regularization Scheme for Mathematical Programs with Complementarity Constraints*, SIAM J. of Optimization, 11:4 (2001), pp. 918–936.

[35] A. Shapiro, *Sensitivity analysis of nonlinear programs and differentiability properties of metric projections*, SIAM J. Control and Optimization , 26:3 (1988), pp. 628–645.

[36] S. J. WRIGHT, *Modifying SQP for degenerate problems*, Preprint ANL/MCS-P699-1097, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill, 1997.