GLOBAL CONVERGENCE OF AN ELASTIC MODE APPROACH FOR A CLASS OF MATHEMATICAL PROGRAMS WITH COMPLEMENTARITY CONSTRAINTS

MIHAI ANITESCU

Abstract. We prove that any accumulation point of an elastic mode approach, applied to the optimization of a mixed P variational inequality, that approximately solves the relaxed subproblems is a C-stationary point of the problem of optimizing a parametric mixed P variational inequality. If, in addition, the accumulation point satisfies the MPCC-LICQ constraint qualification and if the solutions of the subproblem satisfy approximate second-order sufficient conditions, then the limiting point is an M-stationary point. Moreover, if the accumulation point satisfies the upper-level strict complementarity condition, the accumulation point will be a strongly stationary point. If we assume that the penalty function associated with the feasible set of the mathematical program with complementarity constraints has bounded level sets and if the objective function is bounded below, we show that the algorithm will produce bounded iterates and will therefore have at least one accumulation point. We prove that the obstacle problem satisfies our assumptions for both a rigid and a deformable obstacle. The theoretical conclusions are validated by several numerical examples.

Key words. MPCC, global convergence, complementarity constraints, nonlinear programming, obstacle problem.

AMS subject classifications. 65K10, 90C33.

1. Introduction. Complementarity constraints are used to model numerous economics or engineering applications [18, 22]. 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 [18]. Nevertheless, recently there has been substantial interest in solving mathematical programs with complementarity constraints (MPCCs) by using classical nonlinear programming techniques. It has been shown that an SQP elastic mode approach can be expected to locally solve the generic case of MPCC [1]. Several positive results in the same direction have been proved for FilterSQP [10]. These results have been validated by an extensive numerical investigation of SNOPT and of FilterSQP [9]. SNOPT is an algorithm that implements a version of the elastic mode considered here [15]. That success has also been empirically extended to interior point approaches coupled with a relaxation strategy much like the elastic mode approach [2].

Classical nonlinear programming techniques are not the only ones developed for solving MPCC. Other types of techniques have been developed, most notably bundle nonsmooth trust region methods for implicit programming [22] and disjunctive programming [18]. We believe that further investigation in the behavior of classical nonlinear programming techniques is warranted, however, given the success of such techniques in solving large-scale problems, as well as the level of maturity of their software implementations.

Most of the convergence results presented so far in the literature are of a local nature [1, 24, 9]. In this work, we investigate the global convergence of an elastic-mode approach for a special class of MPCCs: optimization of parametric variational inequalities that satisfy the mixed P property [18, Definition 6.1.4]. To accommodate the fact that the penalty parameter may need to be driven to infinity and to avoid the possibility that an insufficiently penalized relaxation will require an infinite number

of steps to solve, we consider the possibility where, for a given penalty parameter, a subproblem is solved only inexactly.

Some of our techniques are related to the ones in [18, Section 6.1]. That reference was the original inspiration for the class of problems considered here. A major difficulty with that work is that unless lower level strict complementarity is satisfied at the point towards which the algorithm defined in [18, Section 6.1] converges, nothing can be said about the quality of that point [18, Page 312]. In this work we refined the range of outcomes that was provided in [18]. We show that under assumptions similar to the ones in [18, Section 6], the outcome of an elastic mode approach can be connected to weaker, though broadly employed in the literature [23, 24, 16], stationarity concepts, without requiring lower-level strict complementarity.

We note that global convergence to a B-stationary point was proved for an active set method that assumes that the MPCC linear independence constraint qualification (MPCC-LICQ) holds everywhere [14]. The assumptions of this work neither imply nor are implied from the assumptions of [14]. In particular, we do not assume uniform MPCC-LICQ, but we do assume that the problem has a structure that is more restrictive than the one in [14].

2. Accumulation Points Are C-Stationary Points. In this section, we discuss the mixed P property, we state our model problem and we present and prove our global convergence result.

2.1. The Mixed P Property. The key notion used in this section is the *mixed P partition* [18, Definition 6.1.4].

Definition (mixed P partition). Let $A \in \mathbb{R}^{(n_c+l) \times n_c}$, $B \in \mathbb{R}^{(n_c+l) \times n_c}$, and $C \in \mathbb{R}^{(n_c+l) \times l}$. We say that the partition $[A \ B \ C]$ is mixed P partition if

$$0 \neq (y, w, z) \in \mathbb{R}^{2n_c+l}, \ Ay + Bw + Cz = 0 \Rightarrow \exists i, \ 1 \leqslant i \leqslant n_c, \text{ such that } y_i w_i > 0. \tag{2.1}$$

LEMMA 2.1. *Assume $[A \ B \ C]$ is a mixed P partition. Let $D \in \mathbb{R}^{n_c \times n_c}$ be a diagonal matrix such that all its diagonal entries satisfy $d_i \neq 0$, $i = 1, 2, \ldots, n_c$. Then $[AD \ BD \ C]$ is also a mixed P partition.*

Proof Let $0 \neq (y, w, z) \in \mathbb{R}^{2n_c+l}$ 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 (2.1) we obtain that $\exists i, \ 1 \leqslant i \leqslant n_c$, 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. $\diamond$

THEOREM 2.2. *Assume that $[A \ B \ C]$ is a mixed P partition. The system of linear constraints*

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

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

Proof Let $0 \neq y \in R^{n_c}$. An immediate consequence of the fact that $[A \ B \ C]$ is a mixed P partition is that the matrix $[B \ C]$ is invertible [18]. We define $w \in \mathbb{R}^{n_c}$ and $z \in \mathbb{R}^l$ by

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

Here we denote by I_k the $k \times k$ identity block. One can immediately see that (y, w, z) satisfies $Ay + Bw + Cz = 0$. Using that $[A \ B \ C]$ is a mixed P partition, we obtain that $\exists i, \ 1 \leqslant i \leqslant n_c$, such that $y_i w_i > 0$. Let $Q = -[I_{n_c} \ 0][B \ C]^{-1} A$. Since $w = Qy$,

this means that $\forall y \neq 0$, $\exists i$, $1 \leqslant i \leqslant n_c$, such that $y_i (Qy)_i > 0$, and thus Q is a P matrix [6]. Therefore, Q^T is also a P matrix [6], where

$$Q^T = -A^T \left[\begin{array}{c} B^T \\ C^T \end{array} \right]^{-1} \left[\begin{array}{c} I_{n_c} \\ 0 \end{array} \right].$$

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

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

We solve for θ from the last two equations to obtain that

$$\theta = - \left[\begin{array}{c} B^T \\ C^T \end{array} \right]^{-1} \left[\begin{array}{c} I_{n_c} \\ 0 \end{array} \right] \eta_2. \tag{2.2}$$

Substituting in the remaining equation, we get that

$$0 = \eta_1 - A^T \left[\begin{array}{c} B^T \\ C^T \end{array} \right]^{-1} \left[\begin{array}{c} I_{n_c} \\ 0 \end{array} \right] \eta_2,$$

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 $1 \leqslant i \leqslant n_c$, such that $-\eta_{1,i}\eta_{2,i} > 0$, or $\eta_{1,i}\eta_{2,i} < 0$. This would contradict the fact that $\eta_1 \geqslant 0$ and $\eta_2 \geqslant 0$. The only alternative remains $\eta_1 = \eta_2 = 0$. From (2.2) this results in $\theta = 0$, which proves our claim. $\diamond$

2.2. Optimization of Parameterized Mixed P Variational Inequalities. We now define the following mathematical program with complementarity constraints that we study in this work together with its relaxed version.

$$
\begin{array}{ll}
\text{(OMPV)} \\
\min_{x,y,w,z} \quad f(x,y,w,z) \\
\text{sbj.to} \quad
\begin{array}{lll}
g(x) & \leqslant 0 & (\mu) \\
h(x) & = 0 & (\lambda) \\
F(x,y,w,z) & = 0 & (\theta) \\
y, w & \leqslant 0 & (\eta_{y,w}) \\
y^T w & \leqslant 0 & (\alpha_c)
\end{array}
\end{array}
\qquad
\begin{array}{ll}
\text{(OMPV(c))} \\
\min_{x,y,w,z,\zeta_1,\zeta_2} \quad f(x,y,w,z) + \quad c(\zeta_1 + \zeta_2) \\
\text{sbj.to} \quad
\begin{array}{lll}
g(x) & \leqslant 0 & (\mu) \\
h(x) & = 0 & (\lambda) \\
-\zeta_1 e_{n_c+l} \leqslant F(x,y,w,z) & \leqslant \zeta_1 e_{n_c+l}(\theta^{-,+}) \\
y, w & \leqslant 0 & (\eta_{y,w}) \\
y^T w & \leqslant \zeta_2 & (\alpha_c) \\
\zeta_1, \zeta_2 & \geqslant 0 & (\alpha_{1,2})
\end{array}
\end{array}
$$

Here we have shown in parentheses the symbols we will use for the Lagrange multipliers. We have denoted here by e_{n_c+l} a vector of all ones, of dimension $n_c + l$.

The last constraint, which together with the bound constraints $y, w \leqslant 0$ form the complementarity constraints of (OMPV), can be formulated as either an equality or an inequality constraint without altering the feasible set [10]. This constraint makes the problem a member of the class of MPCCs. Many of the issues and properties we will discuss in this work are relevant to MPCCs and we will use the MPCC acronym to identify them. We note that there exists a more general class of related problems, mathematical programs with equilibrium constraints (MPEC) [18]. However, when

the convex set over which the equilibrium constraints are defined can be represented by a finite number of inequalities, which is the case in most applications, an MPEC problem can be reduced to an MPCC.

Here $x \in \mathbb{R}^n$, $y, w \in \mathbb{R}^{n_c}$, $z \in \mathbb{R}^l$, $f : \mathbb{R}^{n+2n_c+l} \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^{n_e}$, $g : \mathbb{R}^n \to \mathbb{R}^{n_i}$, $F : \mathbb{R}^{n+2n_c+l} \to \mathbb{R}^{n_c+l}$. In (OMPV(c)) we relax the complementarity constraints $y^T w \leqslant 0$ as well as the nonlinear equation $F(x, y, w, z) = 0$. A large variety of other relaxations, connected to various nondifferentiable exact penalty functions, lead to similar results. For example, all nonlinear constraints can be relaxed [1]. We do not further pursue that avenue.

2.2.1. Stationary Points of (OMPV). The difficult nature of mathematical programs with complementarity constraints has led to the definition of stationary points of (OMPV) different from the ones that correspond to the nonlinear programming interpretation of (OMPV). Formally, we consider multipliers of (OMPV) that satisfy $\alpha_c = 0$, though we relax the requirement that $\eta_y, \eta_w \geqslant 0$, and we denote the new multipliers by $\widehat{\eta}_y, \widehat{\eta}_w$. We call such multipliers $(\lambda, \mu, \theta, \widehat{\eta}_y, \widehat{\eta}_w)$ MPCC multipliers. The corresponding stationary points (x, y, z, w), together with the MPCC multipliers satisfy the following relations:

$$
\begin{aligned}
\nabla_x f(x,y,w,z)^T + \nabla_x h(x)^T \lambda + & \\
\nabla_x g(x)^T \mu + \nabla_x F(x,y,w,z)^T \theta &= 0 \\
\nabla_y f(x,y,w,z)^T + \widehat{\eta}_y + \nabla_y F(x,y,w,z)^T \theta &= 0 \\
\nabla_w f(x,y,w,z)^T + \widehat{\eta}_w + \nabla_w F(x,y,w,z)^T \theta &= 0 \\
\nabla_z f(x,y,w,z)^T + \nabla_z F(x,y,w,z)^T \theta &= 0 \\
g(x) \leqslant 0, \mu \geqslant 0, h(x) = 0, g(x)^T \mu = 0 & \\
F(x,y,z,w) = 0, y \leqslant 0,\ w \leqslant 0,\ y^T w = 0, & \\
\sum_{k=1}^{n_c} y_k |\widehat{\eta}_{y,k}| = 0,\ \sum_{k=1}^{n_c} w_k |\widehat{\eta}_{w,k}| = 0 &
\end{aligned}
\tag{2.3}
$$

We distinguish the following types of stationarity [23, 16].
- Weakly stationary points where no sign requirements are made on $\widehat{\eta}_y, \widehat{\eta}_w$
- C-stationary points, where we require that $\widehat{\eta}_{y,k} \widehat{\eta}_{w,k} \geqslant 0$, $k = 1, 2, \ldots, n_c$
- M-stationary points, that are C-stationary points with the additional requirement that either $\widehat{\eta}_{y,k} \geqslant 0$ or $\widehat{\eta}_{w,k} \geqslant 0$, $k = 1, 2, \ldots, n_c$
- B-stationary points, for which $d = 0$ is a solution of the problem obtained by linearizing all the data of (OMPV) with the exception of the complementarity constraint $y^T w \leqslant 0$
- Strongly stationary points, which satisfy

$$y_k = 0,\ w_k = 0 \Rightarrow \widehat{\eta}_{y,k} \geqslant 0 \text{ and } \widehat{\eta}_{w,k} \geqslant 0,\ \ k = 1, 2, \ldots, n_c$$

  It is immediate that such points are also KKT points in the nonlinear programming sense of (OMPV) [23].

If a point is a strongly stationary point, then it is also a stationary point of any other type [23]. Also, a stationary point of any type is a weakly stationary point [23]. In addition, an M-stationary point is also a C-stationary point. No other relation holds in general between these stationarity concepts. For an approach that uses the linearization of the data, B-stationary points seem to be the desirable outcome. However, the amount of work necessary to recognize B-stationary points may be exponential in the dimension of the problem [22].

Definition (ULSC)[13, 24]. A weakly stationary point (x, y, z, w) of (OMPV) satisfies the upper level strict complementarity (ULSC) property if there exists an

MPCC multiplier that satisfies

$$y_k + w_k = 0 \;\Rightarrow\; \widehat{\eta}_{y,k}\widehat{\eta}_{w,k} \neq 0, \; k = 1, 2, \ldots, n_c.$$

2.2.2. Parametric Mixed P Variational Inequalities. For fixed x, the system of generalized equations

$$F(x, y, w, z) = 0, \; y \leqslant 0, \; w \leqslant 0, \; w^T y = 0 \tag{2.4}$$

defines a mixed nonlinear complementarity problem, that is an instance of a variational inequality. We can therefore interpret y, w, z as the state variables and x as the parameters of the parameterized variational inequality (2.4).

For the remainder of this work 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, for any x,
    i) $\nabla_x h(x)$ has full column rank.
    ii) $\exists p \in \mathbb{R}^n$ such that $\nabla_x h(x)p = 0$ and $\nabla g_i(x)p < 0$ whenever $g_i(x) \geqslant 0$.
    iii) The linearization $h(x) + \nabla_x h(x)d = 0$, $g(x) + \nabla_x g(x)d \leqslant 0$ is feasible.
A3 The partition $[\nabla_y F, \nabla_w F, \nabla_z F]$ is a mixed P partition (2.1).
Note that **A2i** and **A2ii** do not imply **A2iii**, since we allow for the point x to be infeasible for the constraints $g(x) \leqslant 0$, $h(x) = 0$. The Assumption **A2** holds when $g \leqslant 0, h = 0$ represent a polyhedral set in a minimal representation.

We call the problem (OMPV), that contains (2.4) as a constraint, under the assumptions **A1**–**A3** *optimization of parameterized mixed P variational inequalities*. Note that (OMPV) under the assumptions **A1**–**A3** was also studied in [18] but with a different algorithmic approach and outcome.

THEOREM 2.3. *The nonlinear program (OMPV(c)) has a feasible linearization and satisfies the Mangasarian-Fromovitz constraint qualification (MFCQ) [4, Page 441] [19] at any point* $(x, y, w, z, \zeta_1, \zeta_2)$.

Proof. Consider the constraints $g(x) \leqslant 0$, $h(x) = 0$, $y \leqslant 0$, $w \leqslant 0$. Using Assumption **A2**, we obtain that there exist d_x, d_y, d_w such that

$$
\begin{array}{rcl}
g(x) + \nabla_x g(x)d_x & < & 0 \\
h(x) + \nabla_x h(x)d_x & = & 0 \\
y + d_y & < & 0 \\
w + d_w & < & 0.
\end{array}
$$

Choose also $d_w = 0_{n_c}$, $d_y = 0_{n_c}$, $d_z = 0_l$, and consider the linearization of the feasible set of the program (OMPV(c)) at the point $(x, y, w, z, \zeta_1, \zeta_2)$:

$$
\begin{array}{rcl}
g(x) + \nabla_x g(x)d_x & \leqslant & 0 \\
h(x) + \nabla_x h(x)d_x & = & 0 \\
-\zeta_1 e_{n_c+l} - d_{\zeta_1} e_{n_c+l} \leqslant F(x, y, w, z) + & & \\
\nabla_{(x,y,w,z)} F(x, y, w, z) \left(d_x, d_y, d_w, d_z\right)^T & \leqslant & \zeta_1 e_{n_c+l} + d_{\zeta_1} e_{n_c+l} \\
y + d_y, w + d_w & \leqslant & 0 \\
y^T w + w^T d_y + y^T d_w & \leqslant & \zeta_2 + d_{\zeta_2} \\
\zeta_1 + d_{\zeta_1} & \geqslant & 0 \\
\zeta_2 + d_{\zeta_2} & \geqslant & 0.
\end{array}
$$

If we choose d_{ζ_1}, d_{ζ_2} sufficiently large, then we obtain a $d = (d_x, d_y, d_w, d_z, d_{\zeta_1}, d_{\zeta_2})$ such that the inequalities in the above system are strictly satisfied, which shows that (OMPV(c)) satisfies has a feasible linearization and satisfies MFCQ everywhere.   ⋄

In the following, we introduce the notion of ε stationary point. A variant of it has been used before for Lagrange multiplier algorithms [3, Proposition 4.2.2]. This will help define an algorithm that needs to spend only a finite number of steps to approximately solve (OMPV(c)), before having to increase the penalty parameter. This property is unlike the regularization approach in [24], where the relaxed problem must be solved exactly.

Definition (ε stationary point). We say that $(x, y, w, z, \zeta_1, \zeta_2)$ is an ε stationary point of (OMPV(c)) if there exists $(\lambda, \mu, \theta, \eta_y, \eta_w, \alpha_c, \alpha_1, \alpha_2)$ such that the following conditions are satisfied:

$$
\text{Dual} \begin{cases}
\begin{aligned}
\nabla_x f(x,y,w,z)^T + \nabla_x h(x)^T \lambda + & \\
\nabla_x g(x)^T \mu + \nabla_x F(x,y,w,z)^T (\theta^+ - \theta^-) &= t_x \\
\nabla_y f(x,y,w,z)^T + \eta_y + \alpha_c w + \nabla_y F(x,y,w,z)^T (\theta^+ - \theta^-) &= t_y \\
\nabla_w f(x,y,w,z)^T + \eta_w + \alpha_c y + \nabla_w F(x,y,w,z)^T (\theta^+ - \theta^-) &= t_w \\
\nabla_z f(x,y,w,z)^T + \nabla_z F(x,y,w,z)^T (\theta^+ - \theta^-) &= t_z \\
\|\theta^+\|_1 + \|\theta^-\|_1 + \alpha_1 = c + t_{\alpha 1}; \ \alpha_c + \alpha_2 &= c + t_{\alpha 2} \\
\mu \geqslant 0; \ \eta_y, \eta_w \geqslant 0; \theta^+, \theta^- \geqslant 0; \ \alpha_c, \alpha_1, \alpha_2 &\geqslant 0,
\end{aligned}
\end{cases}
$$

$$
\text{Primal} \begin{cases}
\begin{aligned}
g(x) &\leqslant t_g \\
h(x) &= t_h \\
-\zeta_1 e_{n_c + l} - t_{1F} \leqslant F(x,y,w,z) &\leqslant \zeta_1 e_{n_c + l} + t_{2F} \\
y, w &\leqslant 0 \\
y^T w &\leqslant \zeta_2 + t_c \\
\zeta_1, \zeta_2 &\geqslant 0,
\end{aligned}
\end{cases}
$$

$$
\text{Compl.} \begin{cases}
(\zeta_1 e_{n_c + l} - F)^T \theta^+ + (F + \zeta_1 e_{n_c + l})^T \theta^- = t_{cF} \\
\alpha_c (\zeta_2 - w^T y) = t_{cc}; \ g(x)^T \mu = t_{cg}; \\
|\alpha_2 \zeta_2| \leqslant t_{cp}; \ |\alpha_1 \zeta_1| \leqslant t_{cp}; \ |y^T \eta_y| \leqslant t_{cp}; \ |w^T \eta_w| \leqslant t_{cp},
\end{cases}
$$

where the size of the inexactness is bounded above by ε, that is,

$$
\|t_g, t_h, t_{1F}, t_{2F}, t_c, t_x, t_y, t_w, t_z, t_{\alpha 1}, t_{\alpha 2}, t_{cc}, t_{cF}, t_{cg}, t_{cp}\|_\infty \leqslant \varepsilon.
$$

When defining our approximate stationary point, we assumed that bound constraints that involve a zero bound and the sign condition for the resulting multiplier can be exactly enforced. For interior point methods this assumption is readily satisfied (even in finite-precision arithmetic).

Definition (Global Convergence Safeguard). We say that a nonlinear programming algorithm has a global convergence safeguard if any accumulation point of the algorithm is one of the following

1. An infeasible point of the nonlinear program at which the linearization of the constraints is infeasible.
2. A feasible point of the nonlinear program that does not satisfy MFCQ.
3. A feasible point of the nonlinear program that satisfies MFCQ and that is a KKT point of the nonlinear program.

An example of such an algorithm is FilterSQP [11]. We use the following assumption.

Alg1 The nonlinear programming algorithm has a global convergence safeguard.

LEMMA 2.4. *Any accumulation point of a nonlinear programming algorithm that satisfies Assumption **Alg1** and is applied to (OMPV(c)) is a KKT point.*

Proof Since the nonlinear programming algorithm has a global convergence safeguard, it cannot end in case 1 or case 2, following Theorem 2.3. The conclusion follows, and the proof is complete.   ◇

Based on this Lemma, we will assume in the rest of this work that we have access to an algorithm that satisfies the Assumption **Alg1** and that can provide, for any $\varepsilon > 0$, an ϵ stationary point of (OMPV(c)). We can now state the following theorem, that will be our main convergence tool.

THEOREM 2.5. *Assume that (OMPV) satisfies assumptions **A1,A2** and **A3** and that the relaxed problems (OMPV(c)) are solved with an algorithm that satisfies the Assumption **Alg1**. Let $(x^n, y^n, w^n, z^n, \zeta_1^n, \zeta_2^n)$ be an ε^n stationary point of (OMPV(c^n)). Assume that $\lim_{n\to\infty} c^n = \infty$, $\lim_{n\to\infty} \varepsilon^n = 0$, and $\lim_{n\to\infty} c^n\varepsilon^n = 0$. Then any accumulation point $(x^*, y^*, w^*, z^*, \zeta_1^*, \zeta_2^*)$ of $(x^n, y^n, w^n, z^n, \zeta_1^n, \zeta_2^n)$ must satisfy $\zeta_1^* = 0$, $\zeta_2^* = 0$, and (x^*, y^*, w^*, z^*) is a feasible C-stationary point of (OMPV).*

Proof: Feasibility. From our assumption, $(x^n, y^n, w^n, z^n, \zeta_1^n, \zeta_2^n)$ is a ε^n stationary point of (OMPV(c^n)), which, by Theorem 2.3, satisfies MFCQ everywhere. There must exist the Lagrange multipliers $\lambda^n \in \mathbb{R}^{n_e}$, $\mu^n \in \mathbb{R}^{n_i}_+$, $\theta^{+n}, \theta^{-n} \in \mathbb{R}^{n_c+l}_+$, $\eta_y^n \in \mathbb{R}^{n_c}_+$, $\eta_w^n \in \mathbb{R}^{n_c}_+$, $\alpha_1^n, \alpha_2^n, \alpha_c^n \in \mathbb{R}_+$ that, together with $(x^n, y^n, w^n, z^n, \zeta_1^n, \zeta_2^n)$, satisfy the ε^n approximate KKT conditions, which include the following equations

$$
\begin{aligned}
\nabla_x f(x^n, y^n, w^n, z^n)^T + \nabla_x h(x^n)^T \lambda^n + & \\
\nabla_x g(x^n)^T \mu^n + \nabla_x F(x^n, y^n, w^n, z^n)^T(\theta^{+n} - \theta^{-n}) &= t_x^n \\
\nabla_y f(x^n, y^n, w^n, z^n)^T + \alpha_c^n w^n + \eta_y^n + \nabla_y F(x^n, y^n, w^n, z^n)^T(\theta^{+n} - \theta^{-n}) &= t_y^n \\
\nabla_w f(x^n, y^n, w^n, z^n)^T + \alpha_c^n y^n + \eta_w^n + \nabla_w F(x^n, y^n, w^n, z^n)^T(\theta^{+n} - \theta^{-n}) &= t_w^n \\
\nabla_z f(x^n, y^n, w^n, z^n)^T + \nabla_z F(x^n, y^n, w^n, z^n)^T(\theta^{+n} - \theta^{-n}) &= t_z^n \\
(\zeta_1^n e_{n_c+l} - F(x^n, y^n, w^n, z^n))^T\theta^{+n} + (F(x^n, y^n, w^n, z^n) + \zeta_1^n e_{n_c+l})^T\theta^{-n} &= t_{cF}^n \\
\alpha_1^n + \|\theta^{+n}\|_1 + \|\theta^{-n}\|_1 = c^n + t_{\alpha_1}^n; \; g(x^n) \leqslant t_g^n, \;\; y^n \leqslant 0, \;\; w^n &\leqslant 0 \\
\alpha_c^n(w^{nT}y^n - \zeta_2^n) = t_{cc}^n, \;\; \alpha_2^n + \alpha_c^n = c^n + t_{\alpha_2}^n; \; (w^{nT}y^n - \zeta_2^n) \leqslant t_c^n, \;\; \zeta_2^n &\geqslant 0 \\
g(x_n)^T\mu^n = t_{cg}^n, \;\; |\alpha_1^n\zeta_1^n| \leqslant t_{cp}^n; \;\; |y_n^T\eta_y^n| \leqslant t_{cp}^n, \;\; |w^{nT}\eta_w^n| \leqslant t_{cp}^n, \;\; |\alpha_2^n\zeta_2^n| &\leqslant t_{cp}^n
\end{aligned}
\tag{2.5}
$$

$$
\left\| t_x^n, t_y^n, t_w^n, t_z^n, t_{cc}^n, t_{cF}^n, t_{cg}^n, t_{cy}^n, t_{cw}^n, t_{\alpha_1}^n, t_{\alpha_2}^n, t_{cp}^n \right\|_\infty \leqslant \varepsilon^n
$$

We ignore for the time being the effect of the variable ζ_1 over the optimality conditions. We also denote $\theta^n = \theta^{+n} - \theta^{-n}$ and $\widetilde{\lambda}^n = \left(\lambda^n, \mu^n, \theta^n, \eta_y^n, \eta_w^n, \alpha_c^n, \alpha_2^n\right)$.

Since $\alpha_c^n + \alpha_2^n = c^n + t_{\alpha_2}^n$, $c^n \to \infty$ and $\varepsilon^n \to 0$, we must have that $\left\|\widetilde{\lambda}^n\right\|_\infty \to \infty$ as $n \to \infty$. Therefore, the sequence $\frac{\widetilde{\lambda}^n}{\|\widetilde{\lambda}^n\|_\infty}$, admits an accumulation point

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

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

$$
\frac{\widetilde{\lambda}^n}{\left\|\widetilde{\lambda}^n\right\|_\infty} \to \widetilde{\lambda}^* \text{ and } (x^n, y^n, w^n, z^n, \zeta_2^n) \to (x^*, y^*, w^*, z^*, \zeta_2^*).
$$

We now divide (2.5) by $\left\|\tilde{\lambda}^n\right\|_\infty$ and take the limit as $n \to \infty$, to obtain

$$
\begin{aligned}
\nabla_x h(x^*)^T \lambda^* + \nabla_x g(x^*)^T \mu^* + \nabla_x F(x^*, y^*, w^*, z^*)^T \theta^* &= 0 \\
\alpha_c^* w^* + \eta_y^* + \nabla_y F(x^*, y^*, w^*, z^*)^T \theta^* &= 0 \\
\alpha_c^* y^* + \eta_w^* + \nabla_w F(x^*, y^*, w^*, z^*)^T \theta^* &= 0 \\
\nabla_z F(x^*, y^*, w^*, z^*)^T \theta^* &= 0 \\
g(x^*) \leqslant 0, \ y^* \leqslant 0, \ w^* \leqslant 0, \ (w^{*^T} y^* - \zeta_2^*) \leqslant 0, \ \zeta_2^* &\geqslant 0 \\
g(x^*)^T \mu^* = 0, \ y^{*^T} \eta_y^* = 0, \ w^{*^T} \eta_w^* = 0, \ \alpha_c^*(w^{*^T} y^* - \zeta_2^*) = 0, \ \alpha_2^* \zeta^* &= 0.
\end{aligned}
$$
$$(2.6)$$

Take now an index k such that $1 \leqslant k \leqslant n_c$. Since $\alpha_c^* \geqslant 0$, $w_k^* \leqslant 0$, $y_k^* \leqslant 0$, $\eta_{w,k}^* \geqslant 0$, and $\eta_{y,k}^* \geqslant 0$, we must have that

$$
\eta_{y,k}^* + \alpha_c^* w_k^* > 0 \Rightarrow \eta_{y,k}^* > 0 \overset{(2.6)}{\Rightarrow} y_k^* = 0 \Rightarrow \eta_{w,k}^* + \alpha_c^* y_k^* \geqslant 0.
$$

Similarly, we have that

$$
\eta_{w,k}^* + \alpha_c^* y_k^* > 0 \Rightarrow \eta_{w,k}^* > 0 \overset{(2.6)}{\Rightarrow} w_k^* = 0 \Rightarrow \eta_{y,k}^* + \alpha_c^* w_k^* \geqslant 0.
$$

We therefore conclude that, for $k = 1, 2, \ldots, n_c$, we must have that

$$
(\eta_{w,k}^* + \alpha_c^* y_k^*)(\eta_{y,k}^* + \alpha_c^* w_k^*) \geqslant 0.
$$

We can therefore define for $k = 1, 2, \ldots, n_c$ the quantities

$$
d_k = \begin{cases}
1 & \text{if} \ (\eta_{w,k}^* + \alpha_c^* y_k^*) > 0 \text{ or } (\eta_{y,k}^* + \alpha_c^* w_k^*) > 0 \\
-1 & \text{if} \ (\eta_{w,k}^* + \alpha_c^* y_k^*) < 0 \text{ or } (\eta_{y,k}^* + \alpha_c^* w_k^*) < 0 \\
1 & \text{if} \ (\eta_{w,k}^* + \alpha_c^* y_k^*) = (\eta_{y,k}^* + \alpha_c^* w_k^*) = 0
\end{cases}.
$$

From our observation and the definition of d_k, $k = 1, 2, \ldots, n_c$, we must have that

$$
d_k(\eta_{w,k}^* + \alpha_c^* y_k^*) \geqslant 0 \quad \text{and} \quad d_k(\eta_{y,k}^* + \alpha_c^* w_k^*) \geqslant 0, \ k = 1, 2, \ldots, n_c.
$$

We denote by $D \in \mathbb{R}^{n_c \times n_c}$ the matrix whose diagonal elements are d_k, $k = 1, 2, \ldots, n_c$. The middle equations from (2.6) and our definition of D imply that

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

From Assumption **(A3)**, Lemma 2.1, and Theorem 2.2, the preceding equation implies that $\theta^* = 0$. Replacing this in (2.6), we obtain that

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

which, from Assumption **(A2)**, implies that $\lambda^* = 0$ and $\mu^* = 0$. The fact that $\theta^* = 0$ also implies from (2.6) that

$$
\eta_y^* + \alpha_c^* w^* = 0, \ \eta_w^* + \alpha_c^* y^* = 0. \tag{2.7}
$$

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 (2.6), we obtain that

$$
\alpha_c^* y^{*^T} w^* = 0. \tag{2.8}
$$

We have the following cases.

1. $\alpha_c^* > 0$. Then (2.8) implies that $y^{*^T} w^* = 0$. From the equation $\alpha_c^*(w^{*^T} y^* - \zeta_2^*) = 0$ of (2.6) we get that $\zeta_2^* = y^{*^T} w^* = 0$.

2. $\alpha_c^* = 0$. Then from (2.7) 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 $\alpha_2^* \zeta_2^* = 0$ from (2.6) now implies $\zeta_2^* = 0$.

In either case we obtain $\zeta_2^* = 0$.

Assume now that $\zeta_1^n \xrightarrow{n \to \infty} \zeta_1^* > 0$. Define $\tilde{\lambda}^n = (\lambda^n, \mu^n, \theta^{+n}, \theta^{-n}, \eta_y^n, \eta_w^n, \alpha_c^n, \alpha_1^n, \alpha_2^n)$, that by a similar observation as the one in the previous paragraphs must satisfy that $\left\| \tilde{\lambda}^n \right\|_\infty \to \infty$ as $n \to \infty$. Consider an accumulation point $\tilde{\lambda}^* = \left(\lambda^*, \mu^*, \theta^{+*}, \theta^{-*}, \eta_y^*, \eta_w^*, \alpha_c^*, \alpha_1^*, \alpha_2^*\right)$ of the sequence $\frac{\tilde{\lambda}^n}{\|\tilde{\lambda}^n\|_\infty}$. Using the complementarity relationships from equation (2.5), we obtain that $\theta^{+*^T} \theta^{-*} = 0$. Indeed, if we assume that $\theta_k^{+*} > 0$, for some $k = 1, 2, \ldots, n_c$, the complementarity relations from equation (2.5) in which we take the limit after dividing with $\left\| \tilde{\lambda}^n \right\|_\infty \to \infty$, result in $0 < \zeta_1^* = F_k(x^*, y^*, w^*, z^*)$, which in turn implies that $\theta_k^{-*} = 0$.

Repeating the argument that led us to the conclusion that $\zeta_2^* = 0$, we obtain that $0 = \theta^* = \theta^{+*} - \theta^{-*}$, which in conjunction with $\theta^{+*^T} \theta^{-*} = 0$, implies that $\theta^{+*} = \theta^{-*} = 0$ and that $(\lambda^*, \mu^*, \theta^{+*}, \theta^{-*}) = 0$ as well as $\eta_y^* + \alpha_c^* w^* = 0$, $\eta_w^* + \alpha_c^* y^* = 0$. The last two equations imply that $\eta_y^n, \eta_w^n = O(\alpha_c^n)$. Using the definition of $\tilde{\lambda}$ we obtain that $\left\| \tilde{\lambda}^n \right\|_\infty \leqslant \Gamma \left\| (\alpha_c^n, \alpha_1^n, \alpha_2^n) \right\|_\infty \leqslant \Gamma c^n$ for all n sufficiently large, for some positive Γ, which in turn implies that $\frac{\theta^{+n}}{c^n} \xrightarrow{n \to \infty} 0, \frac{\theta^{-n}}{c^n} \xrightarrow{n \to \infty} 0$. Using these relations together with $\alpha_1^n + \|\theta^{+n}\|_1 + \|\theta^{-n}\|_1 = c^n + t_{\alpha_1}^n$, we obtain that $\alpha_1^* > 0$. However, from the limit of complementarity relationships from equation (2.5), which are obtained after dividing with $\left\| \tilde{\lambda}^n \right\|_\infty \to \infty$, we obtain that $\alpha_1^* \zeta_1^* = 0$. This is a contradiction with the initial assumption that $\zeta_1^* > 0$. We must therefore have that $\zeta_1^* = 0$ in addition to $\zeta_2^* = 0$, which shows that the limit point (x^*, y^*, w^*, z^*) must be feasible.

Proof: C-stationarity. We return to the equation (2.5). We define

$$\hat{\eta}_y^n = \eta_y^n + \alpha_c^n w_n, \ \hat{\eta}_w^n = \eta_w^n + \alpha_c^n y_n. \tag{2.9}$$

Following our definition of an ε^n stationary point, we have that, $\forall \, k = 1, 2, \ldots, n_c$,

$$\hat{\eta}_{y,k}^n \hat{\eta}_{w,k}^n = \eta_{y,k}^n \eta_{w,k}^n + (\alpha_c^n)^2 y_k^n w_k^n + \alpha_c^n \left(\eta_{y,k}^n y_k^n + \eta_{w,k}^n w_k^n \right) \geqslant -2c^n \varepsilon^n - 2(\varepsilon^n)^2 \xrightarrow{n \to \infty} 0. \tag{2.10}$$

Define

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

The components of $\hat{\lambda}^n$ satisfy a set of equations derived from (2.5):

$$
\begin{aligned}
\nabla_x f(x_n, y_n, w_n, z_n)^T + \nabla_x h(x_n)^T \lambda^n + \\
\nabla_x g(x_n)^T \mu^n + \nabla_x F(x_n, y_n, w_n, z_n)^T \theta^n &= t_x^n \\
\nabla_y f(x_n, y_n, w_n, z_n)^T + \hat{\eta}_y^n + \nabla_y F(x_n, y_n, w_n, z_n)^T \theta^n &= t_y^n \\
\nabla_w f(x_n, y_n, w_n, z_n)^T + \hat{\eta}_w^n + \nabla_w F(x_n, y_n, w_n, z_n)^T \theta^n &= t_w^n \\
\nabla_z f(x_n, y_n, w_n, z_n)^T + \nabla_z F(x_n, y_n, w_n, z_n)^T \theta^n &= t_z^n \\
h(x_n) = t_h^n, \ g(x_n) \leqslant t_g^n, \ g(x_n)^T \mu^n = 0, \ y_n \leqslant 0, \ w_n \leqslant 0, &
\end{aligned}
\tag{2.11}
$$

where $\left\| t_g^n, t_h^n, t_x^n, t_y^n, t_w^n, t_z^n, t_y^n \right\|_\infty \leqslant \varepsilon^n$. Assume 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 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 $\widetilde{\mu}^* \geqslant 0$, whereas from (2.10) we must have that

$$\widetilde{\eta}_{y,k}^* \, \widetilde{\eta}_{w,k}^* \geqslant 0, \ k = 1, 2, \ldots, n_c. \tag{2.12}$$

Dividing now all equations involving multipliers of (2.11) by $\left\| \widehat{\lambda}^n \right\|_\infty$ and taking the limit as $n \to \infty$, we obtain that

$$
\begin{aligned}
\nabla_x h(x^*)^T \widetilde{\lambda}^* + \nabla_x g(x^*)^T \widetilde{\mu}^* + \nabla_x F(x^*, y^*, w^*, z^*)^T \widetilde{\theta}^* &= 0 \\
\widetilde{\eta}_y^* + \nabla_y F(x^*, y^*, w^*, z^*)^T \widetilde{\theta}^* &= 0 \\
\widetilde{\eta}_w^* + \nabla_w F(x^*, y^*, w^*, z^*)^T \widetilde{\theta}^* &= 0 \\
\nabla_z F(x^*, y^*, w^*, z^*)^T \widetilde{\theta}^* &= 0 \\
h(x^*) = 0 \ g(x^*) \leqslant 0, \ g(x^*)^T \widetilde{\mu}^* = 0, \ y^* \leqslant 0, \ w^* \leqslant 0.
\end{aligned}
\tag{2.13}
$$

Using the same argument that we applied to (2.6) and that led to the conclusion that $\theta^* = 0$ and, subsequently, $\zeta^* = 0$, we get that (2.13), (2.12), 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 (2.13), 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 (2.10) we must have that

$$\widehat{\eta}_{y,k}^* \, \widehat{\eta}_{w,k}^* \geqslant 0, \ \ k = 1, 2, \ldots, n_c. \tag{2.14}$$

From our definition of $\widehat{\eta}_w^n$ and $\widehat{\eta}_y^n$ (2.9), it does not immediately follow that the corresponding limit point satisfy a complementarity relation with w^* and, respectively, y^*. Although we have that $\eta_{w,k}^n w_k^n \to 0$ and $\eta_{y,k}^n y_k^n \to 0$, for $k = 1, 2, \ldots, n_c$ from (2.5), the additional terms $\alpha_c y_k^n$ and $\alpha_c w_k^n$ 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_c^n may diverge to ∞.

In the following we show that that is not the case. We prove that $\widehat{\eta}^*_{y,k} y^*_k = 0$. Since $\widehat{\lambda}^n$ is bounded we must have that

$$O(1) = \widehat{\eta}^n_{y,k} = \eta^n_{y,k} + \alpha^n_c w^n_k, \quad O(1) = \widehat{\eta}^n_{w,k} = \eta^n_{w,k} + \alpha^n_c y^n_k, \qquad (2.15)$$

and that $y^*_k = 0 \Rightarrow y^n_k \overset{n\to\infty}{\longrightarrow} 0 \Rightarrow \widehat{\eta}^*_{y,k} y^*_k = 0$, which would complete the proof.

Assume then that $y^*_k < 0$. Since the limit point is feasible for (OMPV), we must have that $w^*_k = 0$, and therefore that $w^n_k \overset{n\to\infty}{\longrightarrow} 0$. Multiplying the second equation in (2.15) by w^n_k, we obtain that $\lim_{n\to\infty} \eta^n_{w,k} w^n_k + \alpha^n_c y^n_k w^n_k = 0$.

Since, from the definition of ε^n stationary points, we have that $\left| \eta^n_{w,k} w^n_k \right| \leq \varepsilon^n$, this implies that $\alpha^n_c y^n_k w^n_k \overset{n\to\infty}{\longrightarrow} 0$. Using the first equation in (2.15) and the fact that $(x^n, y^n, w^n, z^n, \zeta^n_1, \zeta^n_2)$ is a ε^n stationary point, we obtain that

$$\widehat{\eta}^n_{y,k} y^n_k = \eta^n_{y,k} y^n_k + \alpha^n_c y^n_k w^n_k \overset{n\to\infty}{\longrightarrow} 0.$$

The last equation proves that $\widehat{\eta}^*_{y,k} y^*_k = 0$, $k = 1, 2, \ldots n_c$. Similarly, it also follows that $\widehat{\eta}^*_{w,k} w^*_k = 0$, $k = 1, 2, \ldots n_c$.

Taking now the limit in (2.11) as $n \to \infty$, and using (2.14), and $\widehat{\lambda}^n \overset{n\to\infty}{\longrightarrow} \widehat{\lambda}^*$, we obtain that (x^*, y^*, w^*, z^*) is a C-stationary point with with MPCC multiplier $\widehat{\lambda}^* = (\lambda^*, \mu^*, \theta^*, \widehat{\eta}^*_y, \widehat{\eta}^*_w)$. The proof is complete. $\diamond$

Note that, in order to obtain a similar result, MPCC-LICQ was needed in [24]. The preceding result also allows us to characterize all local solutions of (OMPV).

COROLLARY 2.6. *Assume that (OMPV) satisfies assumptions **A1**, **A2**, and **A3** everywhere and that (x^*, y^*, w^*, z^*) is a strict local minimum of (OMPV). Then (x^*, y^*, w^*, z^*) is a C-stationary point of (OMPV).*

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

$$(OMPV1(c)) \quad \begin{array}{c} \min\limits_{x,y,w,z} \quad f(x,y,w,z) + cy^T w + c\|F(x,y,w,z)\|_\infty \\ \text{sbj.to} \quad g(x) \leqslant 0, \ h(x) = 0, \ y, w \leqslant 0. \end{array}$$

If $\widehat{x} = (x^*, y^*, w^*, z^*)$ is a strict local minimum of (OMPV), 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 (OMPV1(c)), we must have that

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

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 (OMPV1(c)) on the boundary Γ of $B(\widehat{x}, \delta)$, we must have that

$$f(x,y,w,z) - f(x^*,y^*,w^*,z^*) + \gamma\left(y^T w + \|F(x,y,w,z)\|_\infty\right) > 0.$$

If this is not true, then for any n there exists $\gamma_n > n$ such that, for some $(x^n, y^n, w^n, z^n) \in \Gamma$, a feasible point of (OMPV1(c)), we have that

$$f(x^n,y^n,w^n,z^n) - f(x^*,y^*,w^*,z^*) + \gamma_n\left(y^{n^T} w^n + \|F(x^n,y^n,w^n,z^n)\|_\infty\right) \leqslant 0. \tag{2.16}$$

Since Γ is compact, the sequence (x^n, y^n, w^n, z^n) has an accumulation point $(x^\circ, y^\circ, w^\circ, z^\circ) \in \Gamma$ that must be feasible for (OMPV1(c)). Dividing (2.16) by γ_n and taking

Choose $c_0 > 0$, $n = 0$, $K > 1$, an integer $q \geqslant 1$ and a sequence $\varepsilon^n \to 0$.
MPCC Find   an ε^n solution $\left(x^{c^n}, y^{c^n}, w^{c^n}, z^{c^n}, \zeta_1^{c^n}, \zeta_2^{c^n} \right)$ of (OMPV(c^n)).
          If $\zeta_1^{c^n} + \zeta_2^{c^n} > (\varepsilon^n)^{\frac{1}{q}}$,
                    update c : $c^{n+1} = Kc^n$ and n : $n = n + 1$.
                    return to **MPCC**

TABLE 2.1
Elastic-mode algorithm

the limit as $n \to \infty$, we get that $y^{\circ^T} w^\circ = 0$ and that $F(x^\circ, y^\circ, w^\circ, z^\circ) = 0$, that is that $(x^\circ, y^\circ, w^\circ, z^\circ)$ is in effect feasible for (OMPV). But (2.16) also implies that, for all n,

$$f(x^n, y^n, w^n, z^n) - f(x^*, y^*, w^*, z^*) \leqslant 0.$$

Taking the limit in the last inequality, we obtain that

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

which contradicts our choice of δ.

Therefore, $\widehat{c}$ with the properties specified above must exist. This shows that, for $c > \widehat{c}$, (OMPV1(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 (OMPV1(n)) in $B(\widehat{x}, \delta)$ with the lowest value. By an argument similar to the one that led 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 $(x^n, y^n, w^n, z^n, F(x^n, y^n, w^n, z^n), y^{n^T} w^n)$ is a local solution, and thus a stationary point, of (OMPV(n)). From Theorem 2 it thus follows that (x^*, y^*, w^*, z^*) is a C-stationary point of (OMPV). The proof is complete.   ◇

We note that the above result could also be proven using [23, Theorem 2], after one proves that, under assumptions **A1**, **A2** and **A3**, (OMPV) satisfies MPCC-MFCQ at any solution. The proof follows once we use the dual form of MPCC-MFCQ in conjunction with Theorem 2.2. However, the resulting proof is not shorter than the one we just provided.

2.3. A globally convergent modified elastic mode for the optimization of parameterized mixed P variational inequalities. We now describe our adaptive elastic-mode approach. Although the global convergence result we have presented in the preceding subsection deals with the situation where $c^n \to \infty$, we are interested in also allowing the penalty parameter to stay bounded, because in that case we can recover a strongly stationary point [1]. This is a major advantage over regularization methods and smoothing methods, which even under the strongest assumptions recover a solution of the original problem only in the limit of the range of the smoothing/regularization parameter. [13, 24]. An important issue in that case is how should the penalty parameter c^n be chosen. Since MPCC does not have bounded Lagrange multipliers at a solution, one cannot apply the update that takes into account the local size of the Lagrange multipliers [3]. Here, we select c^n based on a comparison with $(\varepsilon^n)^{\frac{1}{q}}$, where $q \geq 1$ is an integer. In fact, when testing for size of ζ_1^n, ζ_2^n, one may want to compare with the size of the solution of the quadratic subproblem of an SQP method [1]. The equivalent test in that case would require $q = 2$, and one can show that it does not locally interfere with superlinear convergence for a method that uses exact second-order derivatives when it converges to a strongly stationary point,

much as in [1]. That argument is tenuous and beyond the scope of this paper. Here we simply assume that the parameter q is provided by the user, and the test takes the form in Table 2.1.

THEOREM 2.7. *Consider the algorithm described in Table 2.1. Assume that the problem (OMPV) satisfies the assumptions* ***A1***, ***A2*** *and* ***A3***. *Assume that, for a fixed c^n, the subproblem (OMPV(c^n)) is solved with a nonlinear programming algorithm that satisfies Assumption* ***Alg1***. *Assume that $\lim_{n \to \infty} c^n \varepsilon^n = 0$. Assume that the algorithm does not diverge to ∞ and produces $\left(x^{c^n}, y^{c^n}, w^{c^n}, z^{c^n}, \zeta_1^{c^n}, \zeta_2^{c^n} \right)$. Then either*

1. *the penalty parameter update rule is activated a finite number of times and then any accumulation point of $\left(x^{c^n}, y^{c^n}, w^{c^n}, z^{c^n} \right)$ is a strongly stationary point of (OMPV) or*
2. *the penalty parameter update rule is activated an infinite number of times, and then any accumulation point of $\left(x^{c^n}, y^{c^n}, w^{c^n}, z^{c^n} \right)$ is a C-stationary point of (OMPV).*

Proof   Part 1. Since the penalty parameter update rule is activated only a finite number of times, it follows that there exist a $c^* > 0$ and an n_0 such that the penalty parameter satisfies $c^n = c^*$, $\forall n \geqslant n_0$. Therefore any accumulation point $(x^*, y^*, w^*, z^*, \zeta_1^*, \zeta_2^*)$ of $\left(x^{c^n}, y^{c^n}, w^{c^n}, z^{c^n}, \zeta_1^{c^n}, \zeta_2^{c^n} \right)$ is a stationary point of (OMPV(c^*)) that verifies, from the test of the update rule, $\zeta_1^* = \zeta_2^* = 0$. It can be immediately verified that such points are strongly stationary (KKT) points of (OMPV), much as in [1, 3]. Since such proof is fairly straightforward from the above references, it is omitted here.

Part 2. If the penalty parameter is updated an infinite number of times, it follows that c^n is increased to ∞, and, by applying Theorem 2, we get that any accumulation point (x^*, y^*, w^*, z^*) of $\left(x^{c^n}, y^{c^n}, w^{c^n}, z^{c^n} \right)$ is a C-stationary point of (OMPV). $\diamond$

3. When MPCC-LICQ Holds Accumulation Points Are M-stationary Points. The convergence result can be improved when we consider stronger stationary conditions. In the rest of this section, we assume that the linear independence constraint qualification in an MPCC sense (MPCC-LICQ) holds at the solution of the mathematical program with complementary constraints under consideration.

Definition (MPCC-LICQ). We say that MPCC-LICQ holds at a feasible (x, y, z, w), point of (OMPV) if the gradients of all active constraints of (OMPV) at (x, y, z, w), with the exception of the complementary constraint $y^T w \leqslant 0$, are linearly independent.

To accommodate the fact that a solution of a nonlinear program is never exactly determined, we will work again with approximate optimality conditions.

Definition (χ active constraints). We say that a constraint $\tilde{g}(\tilde{x}) \leqslant 0 \ (= 0)$ of a nonlinear program is χ active at a point $\tilde{x}^*$ if we have that $|\tilde{g}(\tilde{x}^*)| \leqslant \chi$.

Definition (ϵ, χ second-order stationary point). We say that the point $\tilde{x} = (x, y, z, w, \zeta_1, \zeta_2)$, together with a Lagrange multiplier $\tilde{\lambda} = (\lambda, \mu, \theta^{+n}, \theta^{-n}, \eta_y, \eta_w, \alpha_c, \alpha_1, \alpha_2)$ is an ε, χ second-order point of (OMPV(c)) if

1. $\tilde{x} = (x, y, z, w, \zeta_1, \zeta_2)$, is an ε stationary point of (OMPV(c)), that satisfies *exactly* the primal-dual complementarity involving the slack variables $\eta_{y,k}^T y = 0$, $\eta_{w,k}^T w = 0$.
2. $u^T \Lambda_{xx}^c (\tilde{x}, \tilde{\lambda}) u > 0$ for any u that is at the same time in the null space of the gradients of the active bound constraints of (OMPV(c)) and null space of a subset of the χ-active non-bound constraints of (OMPV(c)).

Here, $\Lambda^c(\tilde{x}, \tilde{\lambda})$ is the Lagrangian function of (OMPV(c)). For a general nonlinear program that satisfies linear independence and strict complementarity at a solution, the above condition is equivalent to second-order sufficient conditions. In most other cases, it is weaker than second-order sufficient conditions.

We need two sequences ε, χ to define our approximate second-order sufficient point. If one uses ε to define the almost active constraints, and if all constraints that are active at $\tilde{x}^* = (x^*, y^*, z^*, w^*, \zeta_1^*, \zeta_2^*)$ are infeasible at $\tilde{x} = (x, y, z, w, \zeta_1, \zeta_2)$, and if ε is too small, then no constraint will be ε active, and then the second condition of our definition becomes too strong.

We denote by $A(\chi, \tilde{x})$ the matrix that is formed by the gradients of the active bound constraints and χ active nonbound constraints of (OMPV(c)). The second condition of the above definition can be verified in the process, commonly encountered in active-set methods, of solving a linear system with the following matrix:

$$\left[\begin{array}{cc} \Lambda_{xx}(\tilde{x}, \tilde{\lambda}) & A^T(\chi, \tilde{x}) \\ A(\chi, \tilde{x}) & 0 \end{array} \right].$$

Special versions of the symmetric indefinite factorization will reveal whether this matrix has the inertia that is compatible with the second-order condition [5].

It thus seems possible to define an active-set approach that, for given $\epsilon > 0$ and $\chi > 0$ determines an ϵ, χ second-order stationary point. Nevertheless, we are not aware of any software package that is guaranteed to provide such points at this time. Moreover, our definition is unlikely to work for algorithms such as interior point approaches. We are currently investigating (1) ways of defining alternate approximate second-order stationary points that can accommodate approaches other than active-set approaches and (2) whether currently used active-set approaches satisfy our assumptions.

THEOREM 3.1. *Assume that the problem (OMPV) satisfies assumptions **A1**, **A2** and **A3** and that every instance of the problem (OMPV(c^n)) is solved with an algorithm that satisfies Assumption **Alg1**. Assume that $\tilde{x}^n = (x^n, y^n, z^n, w^n, \zeta_1^n, \zeta_2^n)$ is a ε^n, χ^n second-order stationary point of (OMPV(c^n)), for all $n = 1, 2, \ldots, \infty$ and for sequences $\{c^n\}, \{\varepsilon^n\}, \{\chi^n\}$ that satisfy $\lim_{n \to \infty} c^n = \infty$, $\lim_{n \to \infty} \varepsilon^n = 0$, $\lim_{n \to \infty} \chi^n = 0$ and $\lim_{n \to \infty} c^n \varepsilon^n = 0$. Let $(x^*, y^*, z^*, w^*, \zeta_1^*, \zeta_2^*)$ be an accumulation point of this sequence. If (x^*, y^*, z^*, w^*) satisfies MPCC-LICQ, then (x^*, y^*, z^*, w^*) must be an M-stationary point of (OMPV).*

Proof. The argument from [23, Theorem 3.3], that was used to prove the similar result for the case where $\epsilon^n = 0$, $\chi^n = 0$, applies here. We outline the main elements and the way the argument continues to apply for approximate stationary points.

The limit point of the sequence $\tilde{x}^n = (x^n, y^n, z^n, w^n, \zeta_1^n, \zeta_2^n)$ is $(x^*, y^*, z^*, w^*, \zeta_1^*, \zeta_2^*)$. Using Theorem 2.5, we obtain that $0 = \zeta_1^* = \zeta_2^*$ and that (x^*, y^*, w^*, z^*), with the associated MPCC multiplier, is a C-stationary point. Note that since MPCC-LICQ holds, the MPCC multiplier must be unique. Assume now that the limit point $(x^*, y^*, z^*, w^*, 0, 0)$ is not an M-stationary point. This means that for the unique MPCC multiplier $\left(\lambda^*, \mu^*, \theta^*, \widehat{\eta}_y^*, \widehat{\eta}_w^*\right)$ there exists an index k^C, $k^C \in \{1, 2, \ldots, n_c\}$ such that $\widehat{\eta}_{w,k^C}^* < 0$ and $\widehat{\eta}_{y,k^C}^* < 0$.

If we follow the logic we went through to obtain equation (2.10) starting from equation (2.5), as well as the definition of an ε stationary point, this means that for any n, there exists multiplier $\widetilde{\lambda}^n = \left(\lambda^n, \mu^n, \theta^{+n}, \theta^{-n}, \eta_y^n, \eta_w^n, \alpha_c^n, \alpha_1^n, \alpha_2^n\right)$ of (OMPV(c^n)), such that $y_{k^C}^n < 0, w_{k^C}^n < 0, \eta_{y,k^C}^n = 0, \eta_{w,k^C}^n = 0, \zeta_2^n > 0, \alpha_2^n = 0$. Therefore,

the bound constraints $y_{k^C} \leqslant 0, w_{k^C} \leqslant 0$ *must be inactive* at the current n, and we must also have that $\alpha_c^n = c^n + t_{\alpha 2}^n = c^n + O(\varepsilon^n)$. Moreover, any nonbound constraint of (OMPV(c^n)) that is χ^n active must be a relaxed active constraint of (OMPV) at (x^*, y^*, z^*, w^*). In addition, following again the proof of Theorem 2.5, that we followed to obtain equation (2.10) starting from equation (2.5), we obtain that the MPCC multipliers corresponding to the index k^C must satisfy $\widehat{\eta}_{y,k^C}^* = \lim\limits_{n \to \infty} \alpha_c^n w_{k^c}^n < 0, \widehat{\eta}_{w,k^C}^* = \lim\limits_{n \to \infty} \alpha_c^n y_{k^c}^n < 0$. In particular, this means that the sequence $\frac{w_{k^c}^n}{y_{k^c}^n}$ is lower bounded away from 0.

Choose now a vector $d^n = (d_x^n, d_y^n, d_w^n, d_z^n)$ that satisfies the following constraints:

$$
\begin{array}{lll}
\nabla_x g_i(x^n) d_x^n & = \ 0 & i : g_i(x^*) = 0 \\
\nabla_x h(x^n) d_x^n & = \ 0 & \\
\nabla_{(x,y,w,z)} F(x^n, y^n, w^n, z^n) \left(d_x^n, d_y^n, d_w^n, d_z^n \right)^T & = \ 0 & \\
d_{y,k}^n & = \ 0 & k : y_k^* = 0, \ k \neq k^C \\
d_{w,k}^n & = \ 0 & k : w_k^* = 0, \ k \neq k^C \\
d_{y,k^C}^n & = \ 1 & \\
d_{w,k^C}^n & = \ -\frac{w_{k^C}^n}{y_{k^C}^n}. &
\end{array}
$$

Since the limit point (x^*, y^*, z^*, w^*) satisfies MPCC-LICQ, such a vector $d^n = (d_x^n, d_y^n, d_w^n, d_z^n)$ must exist for n sufficiently large. In addition, from (a) our observation that, any non bound constraint of (OMPV(c^n)), that is χ^n active must be a relaxed active constraint of (OMPV) at (x^*, y^*, z^*, w^*) and (b) since the constraints $y_{k^C} \leqslant 0, w_{k^C} \leqslant 0$ are inactive at $\tilde{x}^n = (x^n, y^n, z^n, w^n, \zeta_1^n, \zeta_2^n)$, it follows that $\tilde{d}^n = (d_x^n, d_y^n, d_w^n, d_z^n, 0, 0)$ is in the null space of the gradients of the active bound constraints and χ^n active non-bound constraints. Consider the Lagrangian of (OMPV(c))

$$
\begin{aligned}
\Lambda^c(\tilde{x}, \tilde{\lambda}) \ = \ & f(x, y, z, w) + g(x)^T \lambda + h(x)^T \mu + F(x, y, w, z)^T (\theta^+ - \theta^-) + y^T \eta_y \\
& w^T \eta_w + \alpha_c y^T w + (c - e_{n_c+l}^T \theta^+ - e_{n_c+l}^T \theta^- - \alpha_1) \zeta_1 + (c - \alpha_c - \alpha_2) \zeta_2.
\end{aligned}
$$

Following the assumption that $\tilde{x}^n = (x^n, y^n, z^n, w^n, \zeta_1^n, \zeta_2^n)$, together with the multiplier $\widetilde{\lambda}^n = \left(\lambda^n, \mu^n, \theta^{+n}, \theta^{-n}, \eta_y^n, \eta_w^n, \alpha_c^n, \alpha_1^n, \alpha_2^n \right)$, is a ε^n, χ^n second-order stationary point, we must have that $\tilde{d}^{n,T} \nabla_{\tilde{x}\tilde{x}} \Lambda^{c^n}(\tilde{x}^n, \tilde{\lambda}^n) \tilde{d}^n \geqslant 0$.

Following the expression of the Lagrangian of (OMPV(c)) and the definition of d^n, and since $\lambda^n, \mu^n, \theta^{+,n}, \theta^{-,n}$ are bounded, we obtain that

$$
\tilde{d}^{n,T} \nabla_{\tilde{x}\tilde{x}} \Lambda^{c^n}(\tilde{x}^n, \tilde{\lambda}^n) \tilde{d}^n = -\alpha_c^n \frac{w_{k^c}^n}{y_{k^c}^n} + O(1).
$$

However, as we argued in the beginning of this proof, the fraction in the equation before is lower bounded away from 0, whereas we have that $\alpha_c^n \to \infty$, which means that $\limsup \tilde{d}^{n,T} \nabla_{\tilde{x}\tilde{x}} \Lambda^{c^n}(\tilde{x}^n, \tilde{\lambda}^n) \tilde{d}^n < 0$.

This contradicts the assumption that $\tilde{x}^n$ is an ε^n, χ^n second-order stationary point, which in turn contradicts our assumption that (x^*, y^*, z^*, w^*) is not an M-stationary point. Therefore, (x^*, y^*, z^*, w^*) is an M-stationary point, and the proof is complete.   ◇

An important question is whether the result can be improved to show convergence to a strongly stationary point when MPCC-LICQ holds. A situation where the M-stationarity result can be enhanced when ULSC holds at the convergence point. The following theorem goes back to a result from [13] for a smoothing method, and has

also been stated in [24] for the regularization method, when each subproblem is solved exactly (though, in practice, this may take an infinite number of steps). We formally state this result in the following paragraph.

THEOREM 3.2. *If, in addition to the assumptions of Theorem 3.1, we have that ULSC holds at the accumulation point (x^*, y^*, z^*, w^*), then (x^*, y^*, z^*, w^*) is a strongly stationary point and, as a result, a B-stationary point.*

Proof Since MPCC-LICQ holds at the solution, there exists a unique MPCC multiplier that satisfies the M-stationarity conditions. From the uniqueness property, the same multiplier must also ULSC, which, together with the M-stationarity conditions, implies that (x^*, y^*, z^*, w^*) is a strongly stationary point of (OMPV). Following [23, Theorem 4], we get that the point must also be a B-stationary point.   ◇

3.1. M-Stationary Points in Finite Arithmetic. In this subsection we discuss whether M-stationary points and strongly stationary points can be distinguished in finite arithmetic.

The following result is in the vein of backward error analysis, and it shows that in any neighborhood of an M-stationary point there is a strongly stationary point of a perturbed problem. The result is stated for (OMPV) though it can be immediately extended to the entire MPCC class. Note that we do use no other property of (OMPV) except the twice continuous differentiability of the data.

THEOREM 3.3. *Assume that (x^*, y^*, z^*, w^*) is an M-stationary point of (OMPV). Then, for any $\delta > 0$, the following exist*

1. *A perturbation $f^\delta(x, y, w, z)$ of the objective function $f(x, y, w, z)$ that satisfies $\left\| \nabla_{\tilde{x}} f^\delta(x, y, w, z) - \nabla_{\tilde{x}} f(x, y, z, w) \right\| \leqslant \delta$ for all $\tilde{x} = (x, y, z, w)$ in a neighborhood of (x^*, y^*, z^*, w^*).*
2. *A vector l_F^δ that satisfies $\left\| l_F^\delta \right\| \leqslant \delta$.*
3. *A point $(x^\delta, y^\delta, z^\delta, w^\delta)$ that satisfies $\left\| (x^\delta, y^\delta, z^\delta, w^\delta) - (x^*, y^*, z^*, w^*) \right\| \leqslant \delta$ and that is a strongly stationary point for the perturbed problem*

$$(\delta OMPV) \quad \begin{array}{ll} \min\limits_{x,y,w,z} & f^\delta(x, y, w, z) \\ sbj.to & g(x) \quad\quad\quad\quad \leqslant 0 \\ & h(x) \quad\quad\quad\quad = 0 \\ & F(x, y, w, z) \quad = l_F^\delta \\ & y, w \quad\quad\quad\quad \leqslant 0 \\ & (y^T w = 0) \ y^T w \ \leqslant 0 \end{array} .$$

Note A pure backward error result would require that the M-stationary point at hand is the strongly stationary point of a nearby problem, which is generally not true because of the structure of the problem.

Proof By the definition of an M-stationary point it follows that there exists an MPCC multiplier $(\lambda, \mu \geqslant 0, \theta, \widehat{\eta}_y, \widehat{\eta}_w)$ at $\tilde{x} = (x^*, y^*, w^*, z^*)$ that satisfies

$$
\begin{aligned}
\nabla_x f(x^*, y^*, w^*, z^*)^T + \nabla_x h(x^*)^T \lambda + & \\
\nabla_x g(x^*)^T \mu + \nabla_x F(x^*, y^*, w^*, z^*)^T \theta &= 0 \\
\nabla_y f(x^*, y^*, w^*, z^*)^T + \hat{\eta}_y + \nabla_y F(x^*, y^*, w^*, z^*)^T \theta &= 0 \\
\nabla_w f(x^*, y^*, w^*, z^*)^T + \hat{\eta}_w + \nabla_w F(x^*, y^*, w^*, z^*)^T \theta &= 0 \\
\nabla_z f(x^*, y^*, w^*, z^*)^T + \nabla_z F(x^*, y^*, w^*, z^*)^T \theta &= 0 \\
\end{aligned}
$$
$$
g(x^*) \leqslant 0, \ h(x^*) = 0, \ F(x, y, z, w) = 0,
$$
$$
\mu \geqslant 0, \ y \leqslant 0, \ w \leqslant 0, \ y^T w = 0,
$$
$$
y^T w = 0, \ g(x)^T \mu = 0, \sum_{k=1}^{n_c} y_k \left| \widehat{\eta}_{y,k} \right| = 0, \ \ \sum_{k=1}^{n_c} w_k \left| \widehat{\eta}_{w,k} \right| = 0,
$$

in addition to the sign condition on the multipliers $\widehat{\eta}_y, \widehat{\eta}_w$ associated with the variables involved in the complementarity constraints. These conditions are the following:

$$\forall k = 1, 2, \ldots, n_C, \quad \widehat{\eta}_{y,k}\widehat{\eta}_{w,k} \geqslant 0 \text{ and } \left\{ \begin{array}{l} \widehat{\eta}_{w,k} < 0 \Rightarrow \widehat{\eta}_{y,k} = 0, \\ \widehat{\eta}_{y,k} < 0 \Rightarrow \widehat{\eta}_{w,k} = 0, \end{array} \right.$$

To simplify the notation, we assume, without loss of generality that the negative multipliers appear only in the y variables. In turn, this assumption implies that there is a partition $\tilde{K} \cup \tilde{K}^c = 1, 2, \ldots, n_C$, that satisfies the following:

$$k \in \tilde{K} \Rightarrow \widehat{\eta}_{y,k} < 0, \widehat{\eta}_{w,k} = 0;$$
$$k \in \tilde{K}^c \Rightarrow \widehat{\eta}_{y,k} \geqslant 0, \widehat{\eta}_{w,k} \geqslant 0.$$

We now construct the following family of points,

$$\tilde{x}(t) = (x(t), y(t), w(t), z(t))$$
$$x(t) = x^*, y(t) = y^*, z(t) = z^*;$$
$$w_k(t) = w_k^*, \ k \in \tilde{K}^c;$$
$$w_k(t) = w_k^* - t, \ k \in \tilde{K}.$$

We have that the point $\tilde{x}(t)$ satisfies

$$\begin{array}{rcl}
\nabla_x f(x^*, y^*, w(t), z^*) + \nabla_x h(x^*)\lambda + \\
\nabla_x g(x^*)^T\mu + \nabla_x F(x^*, y^*, w(t), z^*)^T\theta & = & l_x(t) \\
\nabla_y f(x^*, y^*, w(t), z^*)^T + \widehat{\eta}_y + \nabla_y F(x^*, y^*, w(t), z^*)^T\theta & = & l_y(t) \\
\nabla_w f(x^*, y^*, w(t), z^*)^T + \widehat{\eta}_w + \nabla_w F(x^*, y^*, w(t), z^*)^T\theta & = & l_w(t) \\
\nabla_z f(x^*, y^*, w(t), z^*)^T + \nabla_z F(x^*, y^*, w(t), z^*)^T\theta & = & l_z(t) \\
g(x^*) \leqslant 0, h(x^*) = 0, \ F(x^*, y^*, w(t), z^*) = l_F(t), \\
\mu \geqslant 0, y^* \leqslant 0, \ w(t) \leqslant 0, \ y^{*,T}w(t) = 0, \\
g(x^*)^T\mu = 0, \sum_{k=1}^{n_c} y_k^* |\widehat{\eta}_{y,k}| = 0, \sum_{k=1}^{n_c} w_k^*(t) |\widehat{\eta}_{w,k}| = 0.
\end{array}$$

Since, from Assumption **A1**, the data of (OMPV) is twice continuously differentiable, we have that there exists $c_l > 0$ that depends only on the point (x^*, y^*, w^*, z^*) such that the residuals satisfy $\|l_x(t), l_y(t), l_w(t), l_z(t), l_F(t)\| \leqslant c_l t$ for all t sufficiently small. There exists a t^δ such that, for all, $t \leqslant t^\delta$ we have at the same time that

$$\|(x(t), y(t), z(t), w(t)) - (x^*, y^*, z^*, w^*)\| \leqslant \delta \text{ and } \|l_x(t), l_y(t), l_w(t), l_z(t), l_F(t)\| \leqslant \delta.$$

After defining $l_F^\delta = l_F(t^\delta)$, $f^\delta(x, y, w, z) = f(x, y, w, z) + x^T l_x(t^\delta) + y^T l_y(t^\delta) + w^T l_w(t^\delta) + z^T l_z(t^\delta)$ and $(x^\delta, y^\delta, z^\delta, w^\delta) = (x(t^\delta), y(t^\delta), z(t^\delta), w(t^\delta))$, the conclusion of the theorem follows. ⋄

If the point towards which we are converging is an M-stationary point that satisfies MPCC-LICQ and that is not a strongly stationary point, then a descent direction can be found [14]. In that sense, the Theorem 3.1 is still weaker than the ideal result, which is that if MPCC-LICQ holds at the point towards which we are converging, then that point is a strongly stationary point and a B-stationary point [14, 23].

However, the preceding theorem shows that in finite arithmetic, one may not be able use only the signs of the multipliers to predict whether we are converging to a strongly stationary point $(x^\delta, y^\delta, w^\delta, z^\delta)$ or to a proper M-stationary point (x^*, y^*, w^*, z^*) where descent is still possible. This point will be demonstrated later with a numerical example. To guarantee that one can escape a proper M-stationary point, at least when MPCC-LICQ holds, then one has to combine a nonlinear programming algorithm with an active-set method of the type studied in [14]. How to robustly switch between the two is the subject of future research.

4. Conditions for Global Convergence. To obtain a global convergence result, we need to insure that the iterates do not drift away to ∞. To achieve such a result, we make two more assumptions about the problem and one about the algorithm that is used to solve the relaxed problem (OMPV(c)).

    A4 The penalty function $\psi(x, y, w, z) = ||F(x, y, w, z)||_\infty + y^T w$ has bounded level sets over the set defined by the constraints $g(x) \leq 0$, $h(x) = 0$, $y \leq 0$, $w \leq 0$.

    A5 The objective function $f(x, y, w, z)$ is bounded below over the same set.

    Alg2 For any fixed value of c, the algorithm that is applied for solving the problem (OMPV(c)) decreases the merit function $f(x, y, z, w) + c\psi(x, y, z, w)$.

Assumption **Alg2** is quite natural in connection with the subproblem (OMPV(c)). If the constraints $g(x)$ and $h(x)$ are linear and if the algorithm applied is a sequential quadratic programming algorithm that uses a positive definite matrix in the quadratic program (from a BFGS type approximation for example), then one can show that for a fixed c the sequential quadratic programming algorithm produces a sequence of decreasing values of $f(x, y, z, w) + c\psi(x, y, w, z)$, much as in the case of the L_∞ penalty function [3]. Assumption **A5** is standard in most global convergence results [8]. Assumption **A4**, however, seems to be quite restrictive for general nonlinear programming, unless the feasible set is compact. Nevertheless, we will show that for the obstacle problem to be presented in Section 5, the assumption does hold. A similar condition has been used to enforce boundedness of the iterates in [18], for a different merit function, that did not enjoy the exactness property and could not lead to the outcome of part 1 of Theorem 2.7.

THEOREM 4.1. *Assume that (OMPV) satisfies assumptions **A1**–**A5** and that the algorithm that is used to solve the subproblems satisfies assumptions **Alg1** and **Alg2**. Then the solution sequence produced by the algorithm in Table 2.1 is bounded, and any accumulation point is a C-stationary point.*

Proof Let B_f denote the lower bound of the objective function $f(x, y, w, z)$, which exists from Assumption **A5**. It then follows that the merit function $\Psi(x, y, w, z, c) = \frac{1}{c}(f(x, y, w, z) - B_f) + \psi(x, y, w, z)$ is decreased at any step of the algorithm in Table 2.1. When c is fixed, the decrease follows from Assumption **Alg2**. When c is decreased, $\Psi(x, y, w, z, c)$ must decrease since at that point $f(x, y, w, z) - B_f > 0$. Therefore, all the iterates (x^n, y^n, w^n, z^n) will satisfy $\psi(x^n, y^n, w^n, z^n) < \Psi(x^n, y^n, w^n, z^n, c^n) \leq \Psi(x^0, y^0, w^0, z^0, c^0)$. The conclusion follows from Assumption **A4** and from Theorem 2.5. $\diamond$

We emphasize that the value of the lower bound B_f does not need to be known for the decrease in the (unknown) merit function $\Psi(x, y, w, z)$ to occur.

We note that global convergence results for methods that use a penalty term are generally restricted to the case where the global solution of the subproblem is obtained [8, Theorem 12.1.1]. In our case, local solutions of the relaxed/penalized subproblems under the assumptions described here are sufficient.

5. The Obstacle Problem. As examples of problems that satisfy these conclusions we present several instances of the obstacle problem from [22]. This problem concerns the optimization of an elastic membrane that can be in contact with a rigid or elastic obstacle. The design parameters quantify the shape of suport of the mem-

brane. The problem is the following:

$$
\begin{array}{lll}
\min_{x,y,w,z} & f(x,z) & \\
\text{sbj.to} & g(x) & \leqslant 0 \\
\text{(OBST)} & -A(x)z + \phi(x) & = y \\
& k(\phi(x) - A(x)z) + \chi(x) - z & = w \\
& y, w & \leqslant 0 \\
& (y^T w = 0) \ \ y^T w & \leqslant 0
\end{array}
$$

Here all functions are differentiable with respect to their parameter. In addition, for any x we have that the matrix $A(x)$ is positive definite. The parameter k satisfies $k \geqslant 0$, and if $k = 0$, then the obstacle is rigid. The case $k > 0$ models the situation when the obstacle is flexible. The inequality constraints $g(x) \leqslant 0$ are box constraints on the design parameters x.

LEMMA 5.1. *The problem (OBST) satisfies assumptions **A1,A2,A3**.*

Proof Assumption **A1** is satisfied immediately from the differentiability properties of all functions involved in the definition of the obstacle problem. The parameters x satisfy box constraints, and therefore the functions $g(x)$ are linear and always satisfy Assumption **A2**.

To verify Assumption **A3** for (OBST) in the (OMPV) framework, we have

$$
\begin{aligned}
F(x,y,z,w) &= \left(\begin{array}{c} y + A(x)z - \phi(x) \\ w - k(\phi(x) - A(x)z) - \chi(x) + z \end{array} \right) \\
\nabla_{(y,z,w)} F(x,y,w,z) &= \left(\begin{array}{ccc} I & 0 & A(x) \\ 0 & I & I + kA(x) \end{array} \right).
\end{aligned}
$$

We prove that the partition of $\nabla_{(y,z,w)} F(x,y,w,z)$ in blocks corresponding to the variables y, z, w is a mixed P partition and thus Assumption **A3** is satisfied. Take a vector $(\bar{y}, \bar{w}, \bar{z})^T$ that satisfies $\nabla_{(y,z,w)} F(x,y,w,z)(\bar{y}, \bar{w}, \bar{z})^T = 0$, that is

$$
\bar{y} + A(x)\bar{z} = 0, \quad \bar{w} + (I + kA(x))\bar{z} = 0,
$$

as well as $\bar{y}_k \bar{w}_k \leqslant 0$, $k = 1, 2, \ldots, n_c$. In turn, this implies that $\bar{y}^T \bar{w} \leqslant 0$. Solving for $\bar{y}, \bar{w}$ from the displayed equations, this implies that $\bar{z}^T A(x)^T (I + kA(x))\bar{z} \leqslant 0$ which, in turn, implies that $\bar{z}^T A(x)^T \bar{z} + k\bar{z}^T A(x)^T A(x)\bar{z} \leqslant 0$. Since the matrix $A(x)$ is positive definite and $k \geqslant 0$, we obtain that $\bar{z} = 0$, and subsequently $\bar{y} = 0$ and $\bar{w} = 0$. This proves that Assumption **A3** holds for (OBST) and completes the proof of the theorem. ◇

Concerning the level sets of $\psi(x)$, we have the following lemma.

LEMMA 5.2. *For any $\beta > 0$, we have that the set*

$$
\mathcal{L}_\beta = \{(x,y,w,z) \in \mathbb{R}^{n+2n_c+l} | y \leq 0, \ w \leq 0, \ g(x) \leq 0, \ \psi(x,y,w,z) < \beta\}
$$

*is bounded. Therefore, problem (OBST) satisfies Assumption **A4**.*

Proof We start assume the conclusion is false. This means that there is a $\beta > 0$ for which the level set $\mathcal{L}_\beta$ is unbounded.

We first note that we have box constraints on the variables x, which means that they can never get unbounded. Therefore, only the (y, w, z) part can get unbounded. Thus, there exists a sequence $\tilde{x}^n = (x^n, y^n, w^n, z^n)$ such that $\Gamma^n = \|(y^n, w^n, z^n)\| \to \infty$ and $\psi(x^n, y^n, w^n, z^n) < \beta$. In turn, the last statement implies that we have

$$
\|F(x^n, y^n, w^n, z^n)\|_\infty \leq \beta, \quad y^{n^T} w^n \leq \beta. \tag{5.1}
$$

Since the sequence

$$\widehat{x}^n = \left(x^n, \frac{y^n}{\Gamma^n}, \frac{w^n}{\Gamma^n}, \frac{z^n}{\Gamma^n} \right)$$

is bounded, it admits a convergent subsequence. To simplify notation, we assume that the whole sequence $\widehat{x}^n$ is convergent to a point $(\overline{x}, \overline{y}, \overline{w}, \overline{z})$, that satisfies $||(\overline{y}, \overline{w}, \overline{z})|| = 1$. We divide the first equation in (5.1) by Γ^n, and the second equation by $(\Gamma^n)^2$, and we take the limit in both as $n \to \infty$. Since $F(x, y, w, z)$ is linear in y, w, z, and since the mappings $\chi(\cdot)$ and $\phi(\cdot)$ are continuous we obtain that the limit point $(\overline{x}, \overline{y}, \overline{w}, \overline{z})$ satisfies the following relations:

$$\overline{y} + A(\overline{x})\overline{z} = 0, \quad \overline{w} + kA(x)\overline{z} + \overline{z} = 0, \quad \overline{w}^T \overline{y} = 0. \qquad (5.2)$$

Solving for $\overline{y}$ and $\overline{w}$ from the first two equations, and replacing the results in the third equation, we obtained that $\overline{z}^T A(\overline{x})^T (kA(x) + I)\overline{z} = 0$. In turn, the fact that $A(\overline{x})$ is a positive definite matrix implies that $\overline{z} = 0$. Subsequently, from (5.2) we obtain that $\overline{w} = \overline{y} = 0$. This contradicts $||(\overline{y}, \overline{w}, \overline{z})|| = 1$ and proves the claim.   ⋄.

We are now ready to state our main result for the obstacle problem.

THEOREM 5.3. *Assume that $f(x, z)$, the objective function of (OBST) is bounded below on the set $g(x) \leq 0$, $y \leq 0$, $w \leq 0$. If the algorithm from Table 2.1 is applied to (OBST) and the (OMPV(c)) subproblems are solved with an algorithm that satisfies Assumptions **Alg1** and **Alg2**, then the sequence of iterates is bounded and any accumulation point is a C-stationary point.*

Proof The result follows from Lemmas 5.1 and 5.2 and Theorem 4.1.   ⋄

6. Numerical Results. In this section we apply an algorithm like the one in Table 2.1 to three instances of the obstacle problem. All problems have an objective function that is nonnegative over the set defined by the parameter constraints and the bound constraints on the variables y, w. Therefore Theorem 5.3 applies and, if the algorithm we use satisfies **Alg1** and **Alg2**, then any accumulation point of the algorithm is a C-stationary point (and even a strongly stationary point if the penalty parameter stays bounded, from Theorem 2.7).

In their original form, the problems are like (OMPV) except that the constraints $y, w \leq 0$ are replaced by $y, w \geq 0$ [22]. One can immediately be see that all of the results in the preceding sections apply if we change the signs of the variables y, w, and correspondingly switching the signs of the multipliers $\eta_w, \eta_y, \widehat{\eta}_y$, and $\widehat{\eta}_w$. In all problems, we deal with an elastic membrane that is hanging on top of an obstacle. The membrane is attached to a support whose shape can change as a part of the optimization process.

- **The incidence set identification problem** [22, Section 9.4]. In this problem, the shape of the support must be changed in such fashion that the shape of the contact region is as close as possible to a prescribed shape. The objective function here is the discrepancy between some measure of the difference between the contact region for current design and the sought after contact region. Therefore in this problem the final objective function should be as close as possible to zero. These are the `is` problems, which in reference [9] are called the `incid-set` problems.
- **The packaging problem with compliant obstacle** [22, Section 9.3]. In this problem we try to find the shape of the support that will minimized the area of the membrane, while keeping the membrane in contact with the

obstacle over at least a prescribed region. The obstacle here is compliant (it can deform under pressure from the membrane). The objective function is the area of the membrane. These are the `pc` problems, which in reference [9] are the `pack-comp` problems.

- **The packaging problem with rigid obstacle** [22, Section 9.2]. This is the same as the preceding problem except that now the obstacle is constrained to be rigid. These are the `pg` problems, which in reference [9] are the `pack-rig` problems. The shape of the optimal membrane for the problem `pg-2-32` is displayed in Figure 6.1 both in a transparent fashion on top of the parabolic obstacle and by itself with the final mesh projected on the bottom plane.

The additional constraints of the problems that would not fit the assumptions **A3** are treated by means of a penalty function. We emphasize that this was not a choice we made in order to have the problems fit our framework. The use of a penalty function was the modeling choice from [22], and it was necessary there for the computation of the generalized gradient. Therefore, the problems solved have the same formulation as [22].

For each problem we have six variants. We consider three different grid sizes, all related to a finite element discretization: 8×8, 16×16, and 32×32. The names of the associated problems contain `8`, `16`, or `32` in their name. We also consider two types of obstacle. The first obstacle is a linear obstacle [22, Example 9.1]. The corresponding problems have in their name the digit `1`. The second obstacle is a parabolic obstacle [22, Example 9.2]. The corresponding problems have in their name the digit `2`.

The problems have been modeled using the AMPL modeling language [12], starting from the AMPL model files from the MacMPEC Library of Sven Leyffer [17, 9]. We have implemented the algorithm in Table 2.1 as an AMPL script. We chose the following parameters: $q = 2$, $K = 10$, $c_0 = 10$, and $\epsilon^n = 10^{-3} 12^{-n}$. Note that $c^n \leqslant 10^{n+1}$, which means that $c^n \epsilon^n \to 0$, as required by our results. In addition, we stop the algorithm in Table 2.1 when $\zeta_1^n + \zeta_2^n \leq 1e - 7$.

To solve the nonlinear programming problem for fixed penalty parameter c that corresponds to the section following the label **OMPV** in Table 2.1, we have used the interior-point solver `knitro` [21]. To produce an ϵ^n stationary point as required in Table 2.1, we have set parameters `opttol=feastol=`ϵ^n. We have set the maximum number of iterations of `knitro` to 4000.

Does `knitro` satisfy our assumptions? Since `knitro` is an interior point algorithm, it satisfies the bound constraints in the definition of ϵ stationary point exactly. We note that `knitro` satisfies a weaker version of the property **Alg1**, where MFCQ is replaced by LICQ [21]. Note, however, that Assumption **Alg1** is merely a way to ensure that and approximate KKT point of (OMPV(c)) can be found. In all our experiments, `knitro` always returned an ϵ stationary point with a prescribed $\epsilon > 0$. We cannot a priori guarantee that Assumption **Alg2** holds for `knitro` for any problem, because the algorithm uses a completely different technique to approach the optimal point from that used by sequential quadratic programming algorithms with an exact penalty function, for which **Alg2** can be shown to hold [3]. Nevertheless, **Alg2** did hold for the examples we have tried (at least with respect to the first and last iterate for a fixed penalty parameter c). As a result, Theorem 5.3 applies to give global convergence to C-stationary points.

The problem of verifying the assumptions of Theorem 3.1, by either an a priori guarantee or an a posteriori test, is more difficult. We need to guarantee that the outcome of a given algorithm satisfies the approximate second-order conditions. Our
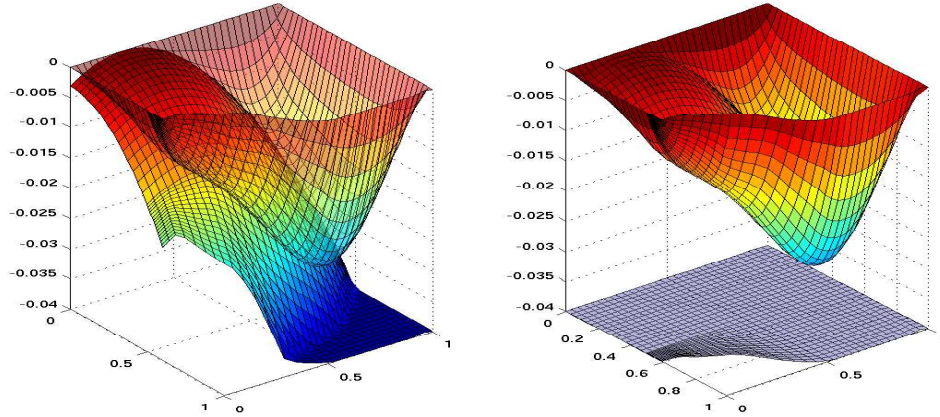
Fig. 6.1. *Solution of the obstacle problem with a rigid parabolic obstacle on a 32 × 32 mesh.*

definition of second order stationary points is oriented towards active-set methods, and cannot be guaranteed for `knitro`. Unfortunately, it also involves the use of derivatives that are not interactively provided in AMPL, and we could not test for it, or a variant of it that may have been appropriate for interior point methods. In addition, in order to guarantee is that the assumptions of Theorem 3.1 holds, we anyway need to verify that MPCC-LICQ holds, which is also difficult to do while using AMPL. Therefore the only test that we did on the outcome, with respect to convergence to M-stationary points, was to see whether the solution point and multipliers at hand satisfy the M-stationarity condition.

We did not choose an active-set software that was available to us, though it would seem appropriate from the discussion following the definition of ϵ, χ second-order stationary points, to solve the subproblem (OMPV(c)) for the following reasons. (1) Our attempt to solve the subproblems (OMPV(c)) by either `lancelot` or `minos` were unsuccessful on problems for grids of size 16 and above (at least in a reasonable amount of time). (2) For the package `SNOPT` it has already been demonstrated that its elastic mode approach works on problems like the one described here [9]. Using our elastic mode approach in a loop where the `SNOPT` elastic mode approach could be triggered when solving the inner problem (OMPV(c)) seemed pointless.

We study whether we are approaching either a C-stationary point or an M-stationary point. Following the proof of Theorem 2.5, once we have obtained the Lagrange multipliers η_w, η_y of the constraints $y, w \leq 0$ in (OMPV(c)), we can construct the following approximation to the MPCC multipliers $\widehat{\eta}_w, \widehat{\eta}_y$:

$$\widehat{\eta}_{w,k} = \eta_{w,k} + cy_k, \quad \widehat{\eta}_{y,k} = \eta_{y,k} + cw_k, \quad k = 1, 2, \ldots, n_C.$$

We now define the following parameters:

$$\text{Cstat} = \min_{k=1,2,\ldots,n_C} \widehat{\eta}_{w,k}\widehat{\eta}_{y,k}, \quad \text{Mstat} = \max_{k=1,2,\ldots,n_C} \min\{\widehat{\eta}_{w,k}, \widehat{\eta}_{y,k}\}.$$

To preserve the clarity of the presentation, we ignore as this point the superscript [n]. Following the proof of the Theorem 2.5, we get that, if $c \to \infty$ and $\liminf \text{Cstat} \geq 0$, then any accumulation point of the algorithm in Table 2.1 is a C-stationary point. In addition, when (OMPV) is formulated with the constraints $y, w \geq 0$ (as we have

Problem	Obj	Uc	Ut	Cstat	Mstat	Feval	KFeval
is-1-8	2.352e-08	0	5	4.10e-11	2.89e-09	204	390
is-1-16	8.639e-06	1	6	9.38e-08	7.85e-06	451	4001
is-1-32	5.904e-06	2	7	3.36e-08	5.52e-05	2906	1097
is-2-8	4.517e-03	1	6	5.12e-08	2.84e-07	302	1712
is-2-16	3.006e-03	1	6	1.27e-06	1.02e-03	434	4001
is-2-32	1.774e-03	2	5	1.01e-05	3.54e-03	2083	4001
pc-1-8	6.000e-01	1	5	6.32e-14	1.40e-03	75	4001
pc-1-16	6.169e-01	1	7	3.82e-21	5.65e-07	297	4001
pc-1-32	6.529e-01	2	6	9.60e-18	8.93e-05	4999	3081
pc-2-8	6.731e-01	1	5	1.01e-19	3.03e-06	78	1421
pc-2-16	7.271e-01	2	5	3.60e-16	1.77e-03	289	1358
pc-2-32	7.826e-01	2	6	1.84e-16	1.22e-04	645	1350
pr-1-8	7.879e-01	1	6	9.28e-18	1.03e-06	193	81
pr-1-16	8.260e-01	2	5	1.68e-16	1.14e-05	218	54
pr-1-32	8.508e-01	2	5	1.95e-17	1.17e-03	644	3040
pr-2-8	7.804e-01	1	6	3.20e-18	1.46e-06	183	33
pr-2-16	1.085e+00	3	6	2.32e-15	1.73e-05	342	208
pr-2-32	1.135e+00	3	6	1.36e-14	1.59e-04	661	2792

TABLE 6.1
Numerical results

done in our AMPL files) and when $\limsup \text{Mstat} \leq 0$, then the limiting point is an M-stationary point.

The numerical results are organized in Table 6.1. We have displayed the name of the problem (Problem), the final values of the objective function (Obj), the number of penalty updates (Uc), the number of tolerance updates (Ut), the C-stationarity indicator (Cstat), the M-stationarity indicator (Mstat), the number of function evaluations (Feval) needed by our implementation of the algorithm in Table 2.1 and the number of function evaluations (KFeval) needed by `knitro` in order to solve (OMPV) directly. The final tolerance parameter is $\epsilon = 10^{-3}12^{-Ut}$ and the final penalty parameter is $c = 10^{-Uc-1}$.

From Table 6.1 it cannot be claimed that one solver is constantly better than the other (in terms of the number of function evaluations), though the elastic mode approach solves 2/3 of the problems faster than `knitro`. We see that `knitro` applied directly to the chosen instances of the obstacle problem stops with a maximum number of iterations reached in five instances, which is a sign of lack of convergence. Note that we went substantially beyond the 1,000 maximum iterations that represent the default settings of `knitro`. However, our elastic mode approach leads to resolution of all the instances of these problems, as predicted by Theorem 4.1, although it also uses `knitro` in the inner loop! This conclusion is in line with previous work that has shown that the elastic mode results in a substantially more robust behavior of nonlinear programming solvers when applied to MPCC [1, 9].

In verifying the conclusion of Theorems 2.7 and 3.1 for the numerical outcome at hand, we have to decide whether we are in the situation where $c \to \infty$. In absence of any additional information, such as whether the point toward which we are converging satisfies MPCC-LICQ and whether strict complementarity holds, which cannot be extracted from those solvers in a simple and robust manner, a robust test for this

condition is a difficult to design. On the one hand, we can consider the values of the penalty parameter derived from the value of Uc in Table 6.1, $c = 10^{-Uc-1}$, to be sufficiently small as to indicate that we have converged in all cases to strongly stationary points. This conclusion would be in line with the fact that, in a statistical sense, this is the expected outcome [23], as well as with preceding numerical investigations [9]. This outcome would confirm the first case of Theorem 2.7. We note that the values of the objective function are consistent with previous numerical experiments [9].

On the other hand, we see that Cstat $\geqslant 0$ and Mstat, though positive, is very small towards convergence. This is consistent with approaching an M-stationary point, which is an outcome allowed by Theorem 2.7, as well as Theorem 3.1. Either way, the result of Theorem 2.7 is verified.

On the issue of distinguishing between M-stationary points and strongly stationary points in finite arithmetic, we have observed what was predicted by Theorem 3.3. For example, for problem pr-1-16, that was formulated with $y \geq 0$ and $w \geq 0$, the numerical results for index $k = 19$ were $y_{19} = 1.039e - 05$, $w_{19} = 1.42e - 04$, $\widehat{\eta}_{y,19} = 0.14$, $\widehat{\eta}_{w,19} = 1.03e - 02$. In absence of any additional information (such as whether MPCC-LICQ holds, which cannot be tested for in AMPL), it is difficult to decide whether the algorithm converges to an M-stationary point at which descent is still possible, or whether it converges to a strongly stationary point. The same situation, in general lines, was noticed for all problems.

7. Conclusions. We have shown that any accumulation point of the elastic mode approach that solves subproblems inexactly, but with increasing accuracy, is a C-stationary point of an optimization problem of parameterized mixed P variational inequalities (Theorem 2.5). If, in addition, the accumulation point satisfies MPCC-LICQ and the solver used provides a point that approximately satisfies the second-order conditions, then the resulting point is M-stationary (Theorem 3.1). We have also shown that any M-stationary point of such problems is in the neighborhood of a strongly stationary point of a perturbed problem, for arbitrarily small perturbations (Theorem 3.3). In practical terms, this means that strongly stationary points and M-stationary points are difficult to distinguish in finite arithmetic.

In the process of guaranteeing that the iterates do not drift to infinity, we construct a merit function with bounded level sets that is decreased at every step, even when the penalty parameter is updated (Theorem 4.1). In turn, this ensures the boundedness of the iterates of the algorithm. We have shown that optimization problems built around the obstacle problem [22, Section 9] satisfy the problem assumptions **A1–A5** that we have used in proving our convergence results (Theorem 5.3). We have implemented our algorithm and applied it to 18 instances of the obstacle problem from the MacMpec [17] library. The numerical results demonstrate our theoretical findings as well as the significant increase in robustness if the elastic mode is used with a nonlinear programming solver.

REFERENCES

[1] M. ANITESCU, *On solving mathematical programs with complementarity constraints as nonlinear programs*, Preprint ANL/MCS-P864-1200, Argonne National Laboratory, Argonne, Illinois, 2000.

[2] H. Y. BENSON, A. SEN, D. F. SHANNO, AND R. J. VANDERBEI, *Interior-point algorithms, penalty methods, and equilibrium constraints*, Preprint ORFE 03-02, Princeton University, Department of Operations Research and Financial Engineering, Princeton, New Jersey, 2003.

[3] D. P. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 1995.

[4] F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[5] J. R. BUNCH AND L. KAUFMAN, *Some stable methods for calculating inertia and solving symmetric linear systems*, Mathematics of Computation, 31 (1977), pp. 163–179.

[6] R. W. COTTLE, J.-S. PAN, AND R. E. STONE, *The Linear Complementarity Problem*, Academic Press, Boston, 1992.

[7] E. D. DOLAN, R. FOURER, J.-P. GOUX, AND T. S. MUNSON, *Kestrel: An interface from modeling systems to the NEOS server*, Preprint ANL/MCS-P986-0802, Mathematics and Computer Science Division, Argonne National Laboratory, 2002.

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

[9] R. FLETCHER AND S. LEYFFER, *Solving mathematical programs with complementarity constraints as nonlinear programs*, Optimization Methods And Software, 19 (2004), pp. 15–40.

[10] R. FLETCHER, S. LEYFFER, S. SHOLTES, AND D. RALPH, *Local convergence of sqp methods for mathematical programs with equilibrium constraints*, Tech. Rep. NA 210, University of Dundee, Dundee, UK, 2002.

[11] R. FLETCHER, S. LEYFFER, AND P. L. TOINT, *On the global convergence of a filter-sqp type algorithm*, Numerical Analysis Report NA/197, Department of Mathematics, University of Dundee, UK, 1999.

[12] R. FOURER, D. M. GAY, AND B. W. KERNIGHAN, *AMPL: A Modeling Language for Mathematical Programming*, Brooks/Cole Publishing Company, Pacific Grove, 2002.

[13] M. FUKUSHIMA AND J.-S. PANG, *Convergence of a smoothing continuation method for mathematical programs with complementarity constraints*, in Ill-Posed Variational Problems and Regularization Techniques, M. Thera and R. Tichatschke, eds., vol. 477 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1999, pp. 99–110.

[14] M. FUKUSHIMA AND P. TSENG, *An implementable active-set algorithm for computing a B-stationary point of a mathematical program with linear complementarity constraints*, SIAM Journal of Optimization, (To appear).

[15] 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.

[16] H. JIANG AND D. RALPH, *Smooth SQP methods for mathematical programs with nonlinear complementarity constraints*, SIAM Journal of Optimization, 10 (2000), pp. 779–808.

[17] S. LEYFFER, *Ampl collection of mathematical programs with equilibrium constraints*. Available on line at *http://www-unix.mcs.anl.gov/leyffer/MacMPEC/*.

[18] Z.-Q. LUO, J.-S. PANG, AND D. RALPH, *Mathematical Programs with Complementarity Constraints*, Cambridge University Press, Cambridge, UK, 1996.

[19] O. L. MANGASARIAN AND S. FROMOVITZ, *The Fritz John necessary optimality conditions in the presence of equality constraints*, Journal of Mathematical Analysis and Applications, 17 (1967), pp. 34–47.

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

[21] J. NOCEDAL, R. H. BYRD, AND M. E. HRIBAR, *An interior point algorithm for large scale nonlinear programming*, SIAM J. Optimization, 9(4) (1999), pp. 877–900.

[22] J. OUTRATA, M. KOCVARA, AND J. ZOWE, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results*, Kluwer Academic Publishers, 1998.

[23] H. SCHEEL AND S. SCHOLTES, *Mathematical programs with equilibrium constraints: Stationarity, optimality and sensitivity*, Mathematics of Operations Research, 25(1) (2000), pp. 1–25.

[24] S. SCHOLTES, *Convergence properties of a regularisation scheme for mathematical programs with complementarity constraints*, SIAM Journal on Optimization, 11 (2001), pp. 918–936.