Convergence of a Penalty Method for Mathematical Programming with Complementarity Constraints[1]

X. M. HU[2]   AND   D. RALPH[3]

Abstract.   We adapt the convergence analysis of smoothing (Ref. 1) and regularization (Ref. 2) methods to a penalty framework for mathematical programs with complementarity constraints (MPCC), and show that the penalty framework shares similar convergence properties to these methods. Moreover, we give sufficient conditions for a sequence generated by the penalty framework to be attracted to a B-stationary point of the MPCC.

Key Words. Complementarity constraints, penalty methods, B-stationarity, linear independence constraint qualification.

[2] Research Scientist, CSIRO Manufacturing and Infrastructure Technology, Australia.
[3] University Lecturer, University of Cambridge, Cambridge, England.

1 Introduction

Mathematical programs with complementarity constraints (MPCC) are receiving increasing attention in optimization and some communities interested in applied science such as engineering and economics. There are many practical problems in engineering and economics that are modelled using the MPCC formulation. MPCC also finds applications in mathematical programming itself, due to reformulation of bilevel programming. For further discussion, see Refs. 3–4.

Consider the following MPCC:

$$\min \quad f(z) \tag{1a}$$

$$\text{s.t.} \quad g(z) \leq 0, \qquad h(z) = 0, \tag{1b}$$

$$G(z) \geq 0, \qquad H(z) \geq 0, \tag{1c}$$

$$G(z)^T H(z) = 0, \tag{1d}$$

where $f : \mathbb{R}^n \to \mathbb{R}, G : \mathbb{R}^n \to \mathbb{R}^m, H : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p, h : \mathbb{R}^n \to \mathbb{R}^q$ are smooth. The ith component of g will be denoted by g_i in the rest of the paper and similarly for other vector-valued functions. In the rest of the paper, we denote by $\mathcal{F}^0$ the set of points satisfying all constraints of (1) except $G(z)^T H(z) = 0$, that is, $\mathcal{F}^0 = \{z \in \mathbb{R}^n : g(z) \leq 0, h(z) = 0, G(z) \geq 0, H(z) \geq 0\}$.

The complementarity constraint $G(z)^T H(z) = 0$ of MPCCs can be blamed for the failure of the standard Mangasarian-Fromovitz constraint qualification in nonlinear programming (Refs. 5-6). As a consequence, most of the well-developed theory for nonlinear programming cannot be directly applied to MPCCs. For example, since the constraint system of an MPCC is not regular, algorithms based on linearizing the constraint system may have unstable numerical behavior; see Robinson (Ref. 7, Corollary 3). Therefore, designing algorithms for MPCCs is of great interest.

Another difficulty in dealing with MPCCs is their combinatorial nature due to the complementarity constraints. In general, the local and global structure of the feasible set is complicated [there exist exponentially many branches (Ref. 3)]. Due to this, the optimality conditions for MPCCs are complex and are not easy to verify, which makes it more difficult to develop efficient algorithms.

Recently, Fukushima and Pang (Ref. 1) and Scholtes (Ref. 2) study convergence of a smoothing method and a regularization method, respectively, each of which determines stationary points, in the usual sense, of a sequence of nonlinear programs that converges to the MPCC. In each case, they show that the sequence of iterates has B-stationary (Ref. 8) limit points under a linear independence constraint qualification (MPCC-LICQ) and additional conditions such as weak second order conditions imposed at each iterate and either asymptotic weak nondegeneracy for the smoothing method or upper level strict complementarity for the regularization method.

We adapt the analysis of Refs. 1-2 to a penalty framework for MPCCs that removes the complementarity constraint from the formulation (1) by adding a general C^2 penalty term $\rho \, \psi(G(z), H(z))$ to the objective function, where ρ is a positive parameter. We give

conditions on ψ such that the penalty method has similar convergence properties to the smoothing and regularization methods, see Theorem 2.1. An immediate consequence is convergence of the standard penalty method in which $\psi(G(z), H(z)) = G(z)^T H(z)$. Tin-Loi and his collaborators have applied the standard penalty method as well as the smoothing and regularization to various problems in mechanics; see Refs. 9-13, which include computational comparisons of these methods.

In revision, we became aware of a related penalty method by Huang, Yang and Zhu (Ref. 14) that penalizes the complementarity conditions $G(z) \geq 0$, $H(z) \geq 0$, $G(z)^T H(z) = 0$ using the square of the Fischer-Burmeister function; c.f. Refs. 15-16. The main convergence result (Ref.14, Theorem 3.2) is similar to our Theorem 2.1 and is also based on the technique of Ref. 1, though the penalty term of Ref. 14 is $C^{1,1}$ rather than C^2. Coincidentally, Lemma 2.3 shows that our penalty framework applies to the cube of the Fischer-Burmeister function, which is C^2. See further discussion following Theorem 2.1.

Note that the feasible set of the nonlinear program solved at each iteration in a penalty method is relaxed compared to the smoothing and regularization methods, an advantage at each iteration, but this is countered by the potential for the method to converge to a point that is infeasible for the MPCC (1). We investigate sufficient conditions for feasibility in Section 3.

There are several reasons for people to be interested in the types of methods described above. First of all, it is very difficult to find a B-stationary point for an MPCC in general, but it is much easier (relatively speaking) to find a stationary point for an NLP under some CQ, so these three methods provide practical ways of finding a B-stationary point of an MPCC. Second, B-stationary points are good candidates for locally optimal solutions. Third, the importance or meaning of the MPCC-LICQ becomes clear when the MPCC is written as a limit of a family of smooth NLPs: the LICQ used for the MPCC is closely related to the usual LICQ for each of the NLPs.

The paper is developed as follows. In Section 1.1, we given various stationarity concepts for the MPCC (1). In Section 1.2, we will give a brief description of each of the smoothing, regularization, and penalty methods. Section 2 discusses the convergence properties of the general penalty method and includes our main result, Theorem 2.1. As mentioned already, Section 3 deals with feasibility of limit points $\bar{z}$ of the sequence generated by the penalty method; Section 3.1 gives sufficient conditions for feasibility of $\bar{z}$ while Section 3.2 investigates properties of B-stationary points $\bar{z}$ such that penalty iterates exist nearby (Corollary 3.1) and indeed are attracted to $\bar{z}$ as $\rho \to \infty$ (Theorem 3.1).

1.1   Stationarity Conditions

The MPCC-LICQ for (1) at a feasible point $\bar{z}$ says that the following vectors

$$\begin{aligned}
\nabla G_i(\bar{z}), &\quad i \in I_G(\bar{z}) \\
\nabla H_j(\bar{z}), &\quad j \in I_H(\bar{z}) \\
\nabla g_r(\bar{z}), &\quad r \in I_g(\bar{z}) \\
\nabla h_l(\bar{z}), &\quad l = 1, \dots, q
\end{aligned}$$

are linearly independent, where

$$I_g(\bar{z}) = \{i : g_i(\bar{z}) = 0\}, I_G(\bar{z}) = \{i : G_i(\bar{z}) = 0\}, I_H(\bar{z}) = \{i : H_i(\bar{z}) = 0\}.$$

The usual LICQ for the nonlinear program (1) would require, in addition, that the gradient of $G(z)^T H(z)$ be linearly independent of the above gradients, which cannot happen in any case for MPCCs since the Mangasarian-Fromovitz constraint qualification, which is necessary for the usual LICQ, fails (Ref. 5). But obviously, the MPCC-LICQ at $\bar{z}$ is the usual NLP LICQ for the NLP problem formed from (1) by dropping the complementarity constraint $G(z)^T H(z) = 0$.

We also use the following notation of index sets in this paper:

$$I_g^+(z) = \{i : g_i(z) > 0\}, I_G^+(z) = \{i : G_i(z) > 0\}, I_H^+(z) = \{i : H_i(z) > 0\}.$$

A feasible point $\bar{z}$ of (1) is called critical (Ref. 8) if there exist MPCC multipliers, $\bar{\lambda}_i \geq 0, \bar{\mu}_j, \bar{\gamma}_k, \bar{v}_l$ (the complementarity between the multipliers and the corresponding constraints is assumed here) satisfying

$$\nabla f(\bar{z}) \,+\, \sum_{i \,\in\, I_g(\bar{z})} \bar{\lambda}_i \nabla g_i(\bar{z}) \,+\, \sum_{j=1}^q \bar{\mu}_j \nabla h_j(\bar{z})$$
$$-\, \sum_{k \,\in\, I_G(\bar{z})} \bar{\gamma}_k \nabla G_k(\bar{z}) \,-\, \sum_{l \,\in\, I_H(\bar{z})} \bar{v}_l \nabla H_l(\bar{z}) = 0.$$

The expression on the left above is actually the gradient of the so-called MPCC Lagrangian $L(z; \lambda, \mu, \xi, \eta)$ at $(\bar{z}, \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta})$ with respect to z, where

$$L(z; \lambda, \mu, \xi, \eta) \quad = \quad f(z) + g(z)^T \lambda + h(z)^T \mu - G(z)^T \xi - H(z)^T \mu. \tag{2}$$

Scheel and Scholtes (Ref. 8) give the following definitions of stationarity of $\bar{z}$.

C-stationarity: $\bar{\lambda}_i \geq 0$ and $\bar{\gamma}_k \bar{v}_k \geq 0$ for all $k \in I_G(\bar{z}) \cap I_H(\bar{z})$.

M-stationarity: $\bar{\lambda}_i \geq 0$ and for all $k \in I_G(\bar{z}) \cap I_H(\bar{z})$ either $\bar{\gamma}_k, \bar{v}_k > 0$ or $\bar{\gamma}_k \bar{v}_k = 0$.

B-stationarity: $\bar{\lambda}_i \geq 0$ and $\bar{\gamma}_k \geq 0, \bar{v}_k \geq 0$ for all $k \in I_G(\bar{z}) \cap I_H(\bar{z})$.

A critical point $\bar{z}$ is said to satisfy the upper level strict complementarity (ULSC) if there exist MPCC multipliers with $\bar{\gamma}_k \bar{v}_k \neq 0, \forall k \in I_G(\bar{z}) \cap I_H(\bar{z})$.

See Ref. 8 for a discussion of these various stationarity conditions and their relations to others in the literature such as the Clarke generalized stationarity.

1.2 Smoothing, Regularization and Penalty Methods

We present two nonlinear programming families approximating the MPCCs used in Refs. 1-2 and a third family used in a simple penalty method for the MPCCs used in Refs. 9-13.

1.2.1. Smoothing Family. See Refs. 1, 17. Let $\varepsilon > 0$ denote a parameter where $\varepsilon \to 0^+$ and consider the following family of NLPs:

$$\min_{z} \quad f(z) \tag{3a}$$

$$\text{s.t.} \quad g(z) \le 0, \quad h(z) = 0, \tag{3b}$$

$$\Phi^{\mathcal{E}}(z) = 0, \tag{3c}$$

where

$$\Phi^{\mathcal{E}}(z) = (\phi(G_1(z), H_1(z), \varepsilon), \cdots, \phi(G_m(z), H_m(z), \varepsilon))^T$$

and $\phi(a, b, \varepsilon) = a + b - \sqrt{a^2 + b^2 + \varepsilon}$ is the perturbed Fischer-Burmeister function (Refs. 15-16). Note that for $\varepsilon = 0$, $\Phi^0(z) = 0$ if and only if $G(z) \ge 0, H(z) \ge 0$ and $G(z)^T H(z) = 0$.

1.2.2. Regularization Family. See Ref. 2. Let $t > 0$ be a parameter where $t \to 0^+$. Scholtes (Ref. 2) considers the following family of NLPs:

$$\min_{z \in \mathcal{F}^0} \quad f(z) \tag{4a}$$

$$\text{s.t.} \quad G_i(z) H_i(z) \le t, \qquad i = 1, \ldots, m \tag{4b}$$

This is just one of the two regularization methods used by Scholtes (Ref. 2).

1.2.3. Penalty Family. Let $\rho > 0$ be an additional parameter where $\rho \uparrow \infty$. Consider the penalty formulation

$$\min_{z \in \mathcal{F}^0} \quad f(z) + \rho \sum_{i=1}^{m} G_i(z) H_i(z) \tag{5}$$

This penalty formulation is used in Refs. 9-13. We shall analyze the convergence property of the penalty method (7) which includes (5) as a special case for more general penalty functions in Section 2.

The three methods associated with these families of NLPs all assume (i) the existence of stationary points that also satisfy a weak second order necessary condition of (3), or (4), or (7), respectively; (ii) the existence of a limit point $\bar{z}$ of the stationary points as the corresponding parameter converges (or diverges); (iii) the MPCC-LICQ holds at this limit point; and (iv) some kind of strict complementarity, in the hope that $\bar{z}$ is some kind of stationary point of (1). In fact, under the MPCC-LICQ and other technical conditions, the convergence properties of these three methods are very similar. See Section 2 for details of the penalty approach.

Let $\bar{z}$ be a limit point of a sequence of stationary points $\{z^k\}$ of (3) with $\varepsilon = \varepsilon^k \to 0^+$ or of (4) with $t = t^k \to 0^+$. The following developments are given in Refs. 1-2. First, it is clear that $\bar{z}$ is feasible for the MPCC (1). Second, if the MPCC-LICQ holds at $\bar{z}$, it is possible to show that $\bar{z}$ is C-stationary. Third, under the weak second order necessary condition at each stationary point of the corresponding NLPs, $\bar{z}$ is an M-stationary point. Further conditions are needed to ensure that $\bar{z}$ is B-stationary.

For comparison, recall the standard and rather straightforward convergence theory of penalty methods for nonlinear programs (see, for example, Ref. 18) which says that any limit point of global minimizers of the penalized problem of a constrained optimization problem

as the penalty parameter goes to infinity, is a global minimizer of the original problem. In practice, however, it is very hard to find a global minimizer for a nonconvex problem. So this standard theory does not contribute much unless we have efficient algorithm(s) to find global minimizers of NLPs. Tin-Loi and his collaborators (Refs. 9-13) use the penalty method (5) to solve a special class of MPCCs and report good numerical performance of the method but do not consider convergence analysis.

2   Penalty Methods for MPCC

We review some standard ideas from nonlinear programming.

Consider the nonlinear program NLP of general form

$$\min \quad f(z) \tag{6a}$$

$$\text{s.t.} \quad g(z) \leq 0, \quad h(z) = 0, \tag{6b}$$

where $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^p, h : \mathbb{R}^n \to \mathbb{R}^q$ are C^2.

We recall that $\bar{z}$ is stationary for (6) if it is feasible and there exist Karush-Kuhn-Tucker (KKT) multiplier vectors $\lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^q$ satisfying

$$\nabla_z L(\bar{z}; \lambda, \mu) = 0$$
$$g(\bar{z}) \leq 0, \quad \lambda \geq 0, \quad \lambda^T g(\bar{z}) = 0$$
$$h(\bar{z}) = 0,$$

where the Lagrangian function is $L(z; \lambda, \mu) = f(z) + \sum_{i=1}^{p} \lambda_i g_i(z) + \sum_{i=1}^{q} \mu_i h_i(z)$.

The standard LICQ at $\bar{z}$ says that the gradients of binding constraints,

$$\nabla g_i(\bar{z}), \quad i : g_i(\bar{z}) = 0,$$
$$\nabla h_j(\bar{z}), \quad j = 1, \ldots, q,$$

are linearly independent. LICQ guarantees existence and uniqueness of multipliers if $\bar{z}$ is a local minimizer of (6).

Assuming the uniqueness of the KKT multipliers, the standard second order necessary condition (SONC) says that the matrix

$$\bar{M} = \nabla_{zz}^2 L(\bar{z}; \lambda, \mu)$$

is positive semidefinite on the critical cone

$$C(\bar{z}, \lambda) = \left\{ d : \begin{array}{ll} \nabla g_i(\bar{z})^T d = 0 & i : g_i(\bar{z}) = 0, \lambda_i > 0 \\ \nabla g_i(\bar{z})^T d \leq 0 & i : g_i(\bar{z}) = 0, \lambda_i = 0 \\ \nabla h_j(\bar{z})^T d = 0 & j = 1, \ldots, q \end{array} \right\}.$$

That is, $d^T \bar{M} d \geq 0$ for $d \in C(\bar{z}, \lambda)$. We will need a slightly weaker condition that we call the weak second order necessary condition (WSONC), which requires the positive semidefiniteness of the matrix $\bar{M}$ on the critical subspace

$$C'(\bar{z}, \lambda) \;=\; C(\bar{z}, \lambda) \cap \{d \,:\, \nabla g_i(\bar{z})^T d = 0, \; i \,:\, g_i(\bar{z}) = 0\}.$$

Recall that any local minimizer $\bar{z}$ of (6) is stationary and satisfies the (W)SONC provided the standard LICQ holds at $\bar{z}$.

Example 2.1 We take the example from Ref. 2 for $a > 0, b > 0$:

$$\begin{aligned} \min \quad & (1/2)\,(x - a)^2 + (1/2)\,(y - b)^2, \\ \text{s.t.} \quad & x \geq 0, \quad y \geq 0, \quad xy = 0. \end{aligned}$$

The KKT points for the corresponding penalty problem (5), given $\rho > 0$, are discussed below.

Case 1. $(x, y; \lambda, \mu) = ((\rho b - a)/(\rho^2 - 1), (\rho a - b)/(\rho^2 - 1); 0, 0)$, if $xy \neq 0$. This is a saddle point of the corresponding penalty problem (5) [in fact, WSONC fails to hold for the corresponding penalty problem (5) at the point] when $\rho > 1$.

Case 2. $(x, y; \lambda, \mu) = (0, b; \rho b - a, 0)$, if $x = 0$ but $y \neq 0$. The multiplier corresponding to $x \geq 0$ is positive for large ρ and WSONC holds at the point for the corresponding penalty problem (5). The point $(x, y) = (0, b)$ is B-stationarity of the original MPCC and in fact is also a local minimizer of the MPCC.

Case 3. $(x, y; \lambda, \mu) = (a, 0; 0, \rho a - b)$, if $x \neq 0, y = 0$. This case is similar to Case 2.

Note that $(x, y) = (0, 0)$ is not a stationary point for the penalty problem, since otherwise $(\lambda, \mu) = (-a, -b) < 0$.

The weak second order necessary condition can be used in algorithms to rule out the bad case, namely Case 1. So if we find stationary points (x^k, y^k) of the penalty problem (5) for this example satisfying the WSONC, then (x^k, y^k) will be equal to either $(0, b)$ or $(a, 0)$, one of the two local optimal solutions of the original problem.

We will present a general penalty method [see (7) below,] and give conditions on penalty functions ψ that yield reasonable convergence results. Let $\psi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ be a function which is at least twice continuously differentiable and satisfies the following conditions:

Condition A. If $x, y \in \mathbb{R}_+^m$, $\psi(x, y) \geq 0$ with equality if and only if $x^T y = 0$.

Condition B. If $x, y \in \mathbb{R}_+^m$, then:

(a) $0 = \dfrac{\partial \psi}{\partial y_i}(x, y)$ if $x_i = 0$, and $0 = \dfrac{\partial \psi}{\partial x_i}(x, y)$ if $y_i = 0$;

(b) $\dfrac{\partial \psi}{\partial x_i} > 0, \dfrac{\partial \psi}{\partial y_i} > 0$ at $(x, y) \geq 0$ with $x_i > 0, y_i > 0$.

6

Condition C. Let $x^k, y^k \in \mathbb{R}_+^m$ and $z^k := (x^k, y^k) \rightarrow (\bar{x}, \bar{y})$.

    (a) Evaluated at (x^k, y^k), $(\partial \psi / \partial x_i)/(\partial \psi / \partial y_i) \rightarrow 0$ if $\bar{x}_i > 0, \bar{y}_i = 0$; $(\partial \psi / \partial y_i)/(\partial \psi / \partial x_i) \rightarrow 0$ if $\bar{x}_i = 0, \bar{y}_i > 0$.

    (b) If $(\bar{x}, \bar{y}) = (0, 0)$ and $\dfrac{\partial \psi}{\partial x_i}(z^k) / \dfrac{\partial \psi}{\partial y_i}(z^k) \rightarrow \gamma > 0$, then there exists $\sigma > 0$ such that

$$(v^{ki})^T V^{ki} v^{ki} / [\frac{\partial \psi}{\partial x_i}(z^k)]^3 \rightarrow -\infty$$

where $v^{ki} = (\dfrac{\partial \psi}{\partial x_i}(z^k), -\sigma \dfrac{\partial \psi}{\partial y_i}(z^k))^T \in \mathbb{R}^2$, and $V^{ki} \in \mathbb{R}^{2 \times 2}$ is the Hessian of ψ with respect to (x_i, y_i) at z^k.

Note that, when $(x^k, y^k) = (G(z^k), H(z^k))$, Condition C(a) corresponds to strictly complementarity constraint indices at $\bar{z}$ and Condition C(b) is for the biactive or degenerate indices at $\bar{z}$.

Now, we give two functions which satisfy all the above conditions.

Example 2.2 Let $\psi(x, y) = x^T y$. It is easy to check that $\psi(\cdot, \cdot)$ satisfies all conditions above with $\sigma = 1$ in Condition C(b). This penalty function is used in (5), which appears in Refs 9-13.

Example 2.3 Let $\phi(a, b)$ denote the Fischer-Burmeister function, that is,

$$\phi(a, b) = a + b - \sqrt{a^2 + b^2}$$

and

$$\psi(x_1, \ldots, x_m, y_1, \ldots, y_m) = \sum_{i=1}^{m} \phi(x_i, y_i)^3.$$

Then, ψ satisfies Conditions A-C. The proof of this claim involves only technical details of calculating partial derivatives of $\phi(\cdot)$ and is omitted here.

We formulate the following penalty problem for (1)

$$\min_{z \in \mathcal{F}^0} \quad f(z) + \rho \, \psi(G(z), H(z)) \tag{7}$$

The smoothing method and the regularization method need not address the issue of feasibility of the limit point of these stationary points because feasibility in the limit is automatic. Feasibility of the limit is an issue for the penalty method however. We will discuss this in Section 3. In this section, we simply assume that the limit point $\bar{z}$ is a feasible point of (1).

Now, we state a convergence result for the penalty method (7). Recall the definitions of C-stationary, M-stationary, and B-stationary points and ULSC from Section 1.1.

Theorem 2.1 Let $\psi(\cdot, \cdot)$ satisfy Conditions A–C and let z^k be a stationary point of (7) for each $\rho = \rho^k$, where $\rho^k \uparrow \infty$. Suppose that $\bar{z}$ is a limit point of $\{z^k\}$ and $\bar{z}$ is a feasible point of (1). Assume that the MPCC-LICQ holds at $\bar{z}$ for (1). Then

(i) $\bar{z}$ is a C-stationary point of (1);

(ii) if WSONC holds for (7) at each z^k, then $\bar{z}$ is a M-stationary point of (1);

(iii) moreover, if the ULSC assumption holds at $\bar{z}$, then $\bar{z}$ is a B-stationary point of (1).

Proof.   By taking a subsequence if necessary (we also assume the same for other sequences), we assume that $z^k \to \bar{z}$. To simplify notation, the partial derivatives of ψ (with respect to x_i, y_i) will be taken at $(G(z^k), H(z^k))$ without specifying this argument. Let $\lambda^k, \mu^k, \xi^k, \eta^k$ be the Lagrange multipliers of (7) at z^k for given ρ^k, that is, let

$$\nabla f(z^k) + \nabla g(z^k)^T \lambda^k + \nabla h(z^k)^T \mu^k$$

$$-\nabla G(z^k)^T (\xi^k - \rho^k \frac{\partial \psi}{\partial x}) - \nabla H(z^k)^T (\eta^k - \rho^k \frac{\partial \psi}{\partial y}) = 0, \tag{8a}$$

$$G(z^k) \geq 0, \quad \xi^k \geq 0, \quad (\xi^k)^T G(z^k) = 0, \tag{8b}$$

$$H(z^k) \geq 0, \quad \eta^k \geq 0, \quad (\eta^k)^T H(z^k) = 0, \tag{8c}$$

$$g(z^k) \leq 0, \quad \lambda^k \geq 0, \quad (\lambda^k)^T g(z^k) = 0, \tag{8d}$$

$$h(z^k) = 0. \tag{8e}$$

Let $\gamma_i^k = \xi_i^k - \rho^k \frac{\partial \psi}{\partial x_i}$ and $v_i^k = \eta_i^k - \rho^k \frac{\partial \psi}{\partial y_i}$.

(i)   First, we show that the limit $\lim_{k \to \infty} \gamma_i^k$ exists and $\bar{\gamma}_i := \lim_{k \to \infty} \gamma_i^k$ is zero if $i \in I_G^+(\bar{z})$. Similarly, $0 = \bar{v}_i := \lim_{k \to \infty} v_i^k$ if $i \in I_H^+(\bar{z})$.

Let $i \in I_G^+(\bar{z})$ then, $i \in I_H(\bar{z})$ by the feasibility assumption of $\bar{z}$ and $\xi_i^k = 0$ for sufficiently large k. To this end, assume that there exist a positive number $\alpha > 0$ and a subsequence (we denote the subsequence by the sequence itself for the sake of notational simplicity) such that $|\gamma_i^k| \geq \alpha$ for sufficiently large k, which implies that $H_i(z^k) > 0$ by Condition B(a) on $\psi(\cdot, \cdot)$ and, as a consequence, $\eta_i^k = 0$. Then,

$$|\gamma_i^k|/|v_i^k| = \left(\rho^k |\frac{\partial \psi}{\partial x_i}|\right)/\left(\rho^k |\frac{\partial \psi}{\partial y_i}|\right) = |\frac{\partial \psi}{\partial x_i}/\frac{\partial \psi}{\partial y_i}| \to 0, \tag{9}$$

by Condition C(a) on $\psi(\cdot, \cdot)$. So, $\lim |v_i^k| = +\infty$. Similarly, for $i \in I_H^+(\bar{z}) \cap I_G(\bar{z})$ and $|v_i^k| \geq \alpha$ for some $\alpha > 0$ and infinitely many k, we have

$$|v_i^k|/|\gamma_i^k| = \left(\rho^k |\frac{\partial \psi}{\partial y_i}|\right)/\left(\rho^k |\frac{\partial \psi}{\partial x_i}|\right) = |\frac{\partial \psi}{\partial y_i}/\frac{\partial \psi}{\partial x_i}| \to 0, \tag{10}$$

which implies that $\lim |\gamma_i^k| = +\infty$.

8

Let $\beta^k := \|(\lambda^k, \mu^k, \gamma^k, v^k)\|$. If there exists an index $i \in I_G^+(\bar{z}) \cap I_H(\bar{z})$ such that $\lim |\gamma_i^k| \neq 0$ or $i \in I_G(\bar{z}) \cap I_H^+(\bar{z})$ such that $\lim |v_i^k| \neq 0$, then $\beta^k \to +\infty$ as shown by (9) and (10). To this end, dividing (8a) by β^k and taking any limit point $(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}, \tilde{v})$ of $(\lambda^k, \mu^k, \gamma^k, v^k)/\beta^k$ yields $(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}, \tilde{v}) \neq 0$ and

$$\sum_{i \in I_g(\bar{z})} \tilde{\lambda}_i \nabla g_i(\bar{z}) + \sum_{i=1}^q \tilde{\mu}_i \nabla h_i(\bar{z}) - \sum_{i \in I_G(\bar{z})} \tilde{\gamma}_i \nabla G_i(\bar{z}) - \sum_{i \in I_H(\bar{z})} \tilde{v}_i \nabla H_i(\bar{z}) = 0, \quad (11)$$

by (9) and (10). Equation (11) contradicts the MPCC-LICQ at $\bar{z}$.

Therefore, $\lim \gamma_i^k = \bar{\gamma}_i = 0$ for $i \in I_G^+(\bar{z})$, and $\lim v_i^k = \bar{v}_i = 0$ for $i \in I_H^+(\bar{z})$. Moreover,

$$\{\gamma_i^k\}_{i \in I_G(\bar{z})}, \quad \{v_i^k\}_{i \in I_H(\bar{z})}, \quad \{\lambda_i^k\}_{i \in I_g(\bar{z})}, \quad \{\mu_i^k\}_{i=1}^q$$

are bounded. Otherwise, dividing (8a) by β^k and taking limit will lead to a contradiction to the MPCC-LICQ at $\bar{z}$ as done above.

In particular, MPCC-LICQ implies that $\gamma^k, v^k, \lambda^k, \mu^k$ have unique limits as $k \to \infty$, denoted by $\bar{\gamma}, \bar{v}, \bar{\lambda}, \bar{\mu}$, respectively. So, $\bar{z}$ is a critical point of (1) by noting the limit of (8a) and $\bar{\lambda} \geq 0$.

Note that, by Condition B on ψ, one has

$$\bar{\gamma}_i \bar{v}_i = \lim_{k \to \infty} \gamma_i^k v_i^k \geq 0,$$

because

$$\gamma_i^k v_i^k = \begin{cases} 0, & \text{if } G_i(z^k) > 0 \text{ and } H_i(z^k) = 0, \\ 0, & \text{if } G_i(z^k) = 0 \text{ and } H_i(z^k) > 0, \\ \xi_i^k \eta_i^k, & \text{if } G_i(z^k) = 0 \text{ and } H_i(z^k) = 0, \\ (\rho^k)^2 (\partial\psi/\partial x_i)(\partial\psi/\partial y_i), & \text{otherwise.} \end{cases}$$

So, $\bar{z}$ is a C-stationary point of (1) by definition.

(ii) Suppose that $\bar{z}$ is not an M-stationary point of (1); that is, there is at least one index $\bar{\imath}$ with $G_{\bar{\imath}}(\bar{z}) = H_{\bar{\imath}}(\bar{z}) = 0$, but $\bar{\gamma}_{\bar{\imath}} < 0, \bar{v}_{\bar{\imath}} < 0$ [this implies that $H_{\bar{\imath}}(z^k) > 0, G_{\bar{\imath}}(z^k) > 0$ for sufficiently large k]. We will see that this is impossible due to the weak second order necessary condition for (7) at z^k for given ρ^k. This part of the proof of is identical in spirit to that in Refs. 1-2.

First, it is easy to see that

$$\lim_{k \to \infty} (\partial\psi/\partial x_{\bar{\imath}})/(\partial\psi/\partial y_{\bar{\imath}}) = \lim_{k \to \infty} (\rho^k \partial\psi/\partial x_{\bar{\imath}})/(\rho^k \partial\psi/\partial y_{\bar{\imath}}) = \bar{\gamma}_{\bar{\imath}}/\bar{v}_{\bar{\imath}} > 0.$$

Second, by the MPCC-LICQ at $\bar{z}$, the following system has full row rank,

$$\begin{aligned}
\nabla g_i(\bar{z})^T d &= 0, && \text{for } i \in I_g(\bar{z}), \\
\nabla h_i(\bar{z})^T d &= 0, && \text{for } i = 1, \ldots, q, \\
\nabla G_i(\bar{z})^T d &= 0, && \text{for } i \in I_G(\bar{z}), i \neq \bar{\imath}, \\
\nabla H_i(\bar{z})^T d &= 0, && \text{for } i \in I_H(\bar{z}), i \neq \bar{\imath}, \\
\nabla G_{\bar{\imath}}(\bar{z})^T d &= \bar{\gamma}_{\bar{\imath}}, \\
\nabla H_{\bar{\imath}}(\bar{z})^T d &= -\sigma \bar{v}_{\bar{\imath}},
\end{aligned}$$

9

where $\sigma > 0$ is the one in Condition C(b); and therefore, all small perturbation of this system will have solutions that are bounded. Therefore, there exists d^k solving the following system:

$$
\begin{aligned}
\nabla g_i(z^k)^T d^k &= 0, &\quad &\text{for } i \in I_g(\bar{z}), \\
\nabla h_i(z^k)^T d^k &= 0, &\quad &\text{for } i = 1, \ldots, q, \\
\nabla G_i(z^k)^T d^k &= 0, &\quad &\text{for } i \in I_G(\bar{z}), i \neq \bar{\imath}, \\
\nabla H_i(z^k)^T d^k &= 0, &\quad &\text{for } i \in I_H(\bar{z}), i \neq \bar{\imath}, \\
\nabla G_{\bar{\imath}}(z^k)^T d^k &= \rho^k \frac{\partial \psi}{\partial x_{\bar{\imath}}}(z^k) = \gamma_{\bar{\imath}}^k, \\
\nabla H_{\bar{\imath}}(z^k)^T d^k &= -\sigma \rho^k \frac{\partial \psi}{\partial y_{\bar{\imath}}}(z^k) = -\sigma v_{\bar{\imath}}^k,
\end{aligned}
$$

and the sequence $\{d^k\}$ is bounded.

It is easy to see that d^k is in the critical subspace of problem (5) at z^k, since only the first four groups of equations are used in the definition of the critical subspace, and

$$(d^k)^T \nabla^2_{zz} L_{\rho^k}(z^k; \xi^k, \eta^k, \lambda^k, \mu^k) d^k$$

$$= (d^k)^T \Big(\nabla^2 f + \sum_{i=1}^p \lambda_i^k \nabla^2 g_i + \sum_{i=1}^q \mu_i^k \nabla^2 h_i \Big) d^k \tag{12a}$$

$$- (d^k)^T \Big(\sum_{i=1}^m \gamma_i^k \nabla^2 G_i(z^k) + \sum_{i=1}^m v_i^k \nabla^2 H_i(z^k) \Big) d^k \tag{12b}$$

$$+ (d^k)^T \rho^k \Big(\frac{\partial^2 \psi}{\partial x_{\bar{\imath}}^2} \nabla G_{\bar{\imath}}(z^k) \nabla G_{\bar{\imath}}(z^k)^T + 2 \frac{\partial^2 \psi}{\partial x_{\bar{\imath}} \partial y_{\bar{\imath}}} \nabla G_{\bar{\imath}}(z^k) \nabla H_{\bar{\imath}}(z^k)^T \tag{12c}$$

$$+ \frac{\partial^2 \psi}{\partial y_{\bar{\imath}}^2} \nabla H_{\bar{\imath}}(z^k) \nabla H_{\bar{\imath}}(z^k)^T \Big) d^k \tag{12d}$$

$$+ (d^k)^T \rho^k \sum_{(i,j) \neq (\bar{\imath}, \bar{\imath})} \Big(\frac{\partial^2 \psi}{\partial x_i \partial x_j} \nabla G_i(z^k) \nabla G_j(z^k)^T \tag{12e}$$

$$+ 2 \frac{\partial^2 \psi}{\partial x_i \partial y_j} \nabla G_i(z^k) \nabla H_j(z^k)^T + \frac{\partial^2 \psi}{\partial y_i \partial y_j} \nabla H_i(z^k) \nabla H_j(z^k)^T \Big) d^k \tag{12f}$$

where

$$
\begin{aligned}
L_{\rho^k}(z; \xi, \eta, \lambda, \mu) &= f(z) + g(z)^T \lambda + h(z)^T \mu \\
&\quad + \rho^k \psi(G(z), H(z)) - G(z)^T \xi - H(z)^T \eta.
\end{aligned}
$$

Since $z^k \to \bar{z}$ and $\gamma^k, v^k, \lambda^k, \mu^k$ have limits, so terms in (12a-b) are bounded and terms in (12e-f) are zero for sufficiently large k. The sum of terms in (12c-d) is

$$(\rho^k)^3 \cdot \Big(\frac{\partial^2 \psi}{\partial x_{\bar{\imath}}^2} \cdot \Big(\frac{\partial \psi}{\partial x_{\bar{\imath}}} \Big)^2 - 2\sigma \frac{\partial^2 \psi}{\partial x_{\bar{\imath}} \partial y_{\bar{\imath}}} \cdot \frac{\partial \psi}{\partial x_{\bar{\imath}}} \cdot \frac{\partial \psi}{\partial y_{\bar{\imath}}} + \sigma^2 \frac{\partial^2 \psi}{\partial y_{\bar{\imath}}^2} \cdot \Big(\frac{\partial \psi}{\partial y_{\bar{\imath}}} \Big)^2 \Big)$$

$$= (v^{ki})^T V^{ki} v^{ki} / \Big(\frac{\partial \psi}{\partial x_i} \Big)^3,$$

10

which tends to $-\infty$ by Condition C(b) on $\psi(\cdot, \cdot)$. This is impossible, since the weak second order necessary condition for (7) at z^k requires

$$(d^k)^T \nabla^2 L_{\rho^k}(z^k; \xi^k, \eta^k, \lambda^k, \mu^k) d^k \geq 0.$$

Hence $\bar{z}$ is a M-stationary point for the MPCC (1).

(iii)   Since $\bar{z}$ is a M-stationary point by (ii), by definition of B-stationarity, under the ULSC assumption, $\bar{z}$ is a B-stationary point. This completes the proof of the theorem.   □

We compare Theorem 2.1 with the recent convergence result (Ref. 14, Theorem 3.2) of Huang, Yang, and Zhu, who apply a penalty method in which all constraints are converted to penalty terms: $\sum \max\{g_i(z), 0\}^2$ for $g(z) \leq 0$, $h(z)^T h(z)$ for $h(z) = 0$, and the square of the Fischer-Burmeister function, that is $\Phi^0(z)^T \Phi^0(z)$ in the notation of Section 1.2, for the complementarity conditions $G(z) \geq 0$, $H(z) \geq 0$, $G(z)^T H(z) = 0$. Showing B-stationarity of limit points of the iteration sequence is the main objective of both Theorem 3.2 in Ref. 14 and Theorem 2.1. There are two main differences between these results. First, the penalty terms $\sum \max\{g_i(z), 0\}^2$ and $\Phi^0(z)^T \Phi^0(z)$ used in Ref. 14 are C^1 but not C^2; their derivatives are locally Lipschitz and semismooth however, which allows nonsmooth second order conditions to be used (c.f. weak second order necessary condition above). Allowing such C^{1,1} penalty terms improves the generality of the analysis and likewise increases the computational difficulty of finding a stationary point of the penalty problem that also satisfies the appropriate second order conditions. Although our penalty framework uses C^2 functions, an extension along the lines of Ref. 14 would be a promising research direction because in each case the convergence analysis relies heavily on that of Ref. 1. Second, Theorem 3.2 in Ref. 14 establishes a weak second order condition in addition to the B-stationarity at (feasible) limit points of the method.

In the rest of this section, we will show that a limit point $\bar{z}$ from Theorem 2.1 satisfies the weak second order necessary condition given in Ref. 14, Theorem 3.2(ii), provided the general penalty function ψ has a further property in addition to Conditions A, B, C.

Condition D.

(a) Let $0 \leq (x, y) \to (\bar{x}, \bar{y})$. Then exists a constant $\Gamma > 0$ such that

$$|\lim \frac{\partial^2 \psi}{\partial x_i^2}(x, y)| \leq \Gamma \, |\lim \frac{\partial \psi}{\partial x_i}(x, y)|, \quad \text{if } \bar{x}_i > 0, \bar{y}_i = 0,$$

$$|\lim \frac{\partial^2 \psi}{\partial y_i^2}(x, y)| \leq \Gamma \, |\lim \frac{\partial \psi}{\partial y_i}(x, y)|, \quad \text{if } \bar{x}_i = 0, \bar{y}_i > 0.$$

(b) Let $(x, y) \geq 0$ and $i \neq j$. Then,

$$\frac{\partial^2 \psi}{\partial x_i \partial x_j}(x, y) = \frac{\partial^2 \psi}{\partial x_i \partial y_j}(x, y) = \frac{\partial^2 \psi}{\partial y_i \partial y_j}(x, y) = 0.$$

It is immediate that Condition D holds for $\psi(x, y) = \sum x_i y_i$, i.e., the standard penalty method. It also easy to see that it holds for $\psi(x, y) = \sum \phi(x_i, y_i)^3$ by calculating its second order derivatives.

Let $\bar{z}$ be a feasible point of (1) at which the MPCC-LICQ holds. Assume that $\bar{z}$ is a critical point of (1), that is, $\nabla_z L(\bar{z}; \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta}) = 0$ for some (unique) multipliers $\bar{\xi}, \bar{\eta}, \bar{\lambda}, \bar{\mu}$, where L is the MPCC Lagrangian given by (2). The MPCC Lagrangian should not be confused with the Lagrangian L_ρ of the penalty problem (7). We say the MPCC (1) satisfies a weak second order necessary condition (MPCC-WSONC) at $\bar{z}$ if $\nabla^2_{zz} L(\bar{z}; \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta})$ is positive semidefinite on the following linear space

$$C''(\bar{z}) = \{d \in \mathbb{R}^n : \nabla g_i(\bar{z})^T d = 0, i \in I_g(\bar{z}), \nabla h_i(\bar{z})^T d = 0, i = 1, \ldots, q,$$
$$\nabla G_i(\bar{z})^T d = 0, i \in I_G(\bar{z}), \nabla H_i(\bar{z})^T d = 0, i \in I_H(\bar{z})\}.$$

which is a subspace of the cone consisting of the critical directions of (1). The usual SONC for MPCCs requires that $\nabla^2_{zz} L(\bar{z}; \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta})$ is positive semidefinite on the cone of critical directions; see Ref. 3, Theorem 5.2.1 or Ref. 8.

Proposition 2.1 Let $\psi(\cdot, \cdot)$ satisfy Conditions A-D. Let z^k be a stationary point of (7) at which WSONC holds, for each $\rho = \rho^k$, where $\rho^k \uparrow \infty$. Suppose that $\bar{z}$ is a limit point of $\{z^k\}$, $\bar{z}$ is a feasible point of (1), and that the MPCC-LICQ holds at $\bar{z}$ for (1). Then, MPCC-WSONC holds at $\bar{z}$.

Proof. We assume, by taking a subsequence if necessary, that $z^k \to \bar{z}$. Also, we omit the argument $(G(z^k), H(z^k))$ when writing the partial derivatives of ψ.

Let $d \in C''(\bar{z})$. As seen from the proof of Theorem 2.1, due to the MPCC-LICQ at $\bar{z}$, there exists d^k, with $\lim d^k = d$, and

$$\nabla g_i(z^k)^T d^k = 0, i \in I_g(\bar{z}), \quad \nabla h_i(z^k)^T d^k = 0, i = 1, \ldots, q, \tag{13a}$$

$$\nabla G_i(z^k)^T d^k = 0, i \in I_G(\bar{z}), \quad \nabla H_i(z^k)^T d^k = 0, i \in I_H(\bar{z}). \tag{13b}$$

Obviously, d^k is in the critical subspace of the constraints of (7); hence, the WSONC gives $0 \le d^{kT} \nabla^2_{zz} L_{\rho^k}(z^k; \xi^k, \eta^k, \lambda^k, \mu^k) d^k$.

Now, we examine the nonnegative scalars $d^{kT} \nabla^2_{zz} L_{\rho^k}(z^k; \xi^k, \eta^k, \lambda^k, \mu^k) d^k$ in light of (12a)-(12f), as $k \to \infty$. From Condition D(a), if $G_i(\bar{z}) > 0, H_i(\bar{z}) = 0$, then $|\rho^k \frac{\partial^2 \psi}{\partial x_i^2}| \le \Gamma |\rho^k \frac{\partial \psi}{\partial x_i}| = \Gamma \gamma_i^k \to 0$, as in the proof of Theorem 2.1. Similarly, $\rho^k \frac{\partial^2 \psi}{\partial y_i^2} \to 0$ for i such that $G_i(\bar{z}) = 0 < H_i(\bar{z})$. Also, recall that $I_G(\bar{z}) \cup I_H(\bar{z}) = \{1, \ldots, m\}$ by the feasibility of $\bar{z}$ for (1). Therefore, by Condition D and (13), the terms in (12c)–(12f) vanish as $k \to \infty$:

$$\rho^k (d^k)^T \frac{\partial^2 \psi}{\partial x_i \partial x_j} \nabla G_i(z^k) \nabla G_j(z^k)^T d^k \begin{cases} = 0, & i = j \text{ and } i \in I_G(\bar{z}), \text{ or } i \ne j, \\ \to 0, & i = j \text{ and } G_i(\bar{z}) > 0, \end{cases}$$

$$\rho^k (d^k)^T \frac{\partial^2 \psi}{\partial y_i \partial y_j^2} \nabla H_i(z^k) \nabla H_j(z^k)^T d^k \begin{cases} = 0, & i = j \text{ and } i \in I_H(\bar{z}), \text{ or } i \ne j, \\ \to 0, & i = j \text{ and } H_i(\bar{z}) > 0, \end{cases}$$

$$\rho^k (d^k)^T \frac{\partial^2 \psi}{\partial x_i \partial y_j} \nabla G_i(z^k) \nabla H_j(z^k)^T d^k = 0, \text{ for } i, j = 1, \ldots, m.$$

This only leaves the terms in (12a)-(12b).

As in the proof of Theorem 2.1, there exist the following limits as $k \to \infty$: $\lambda^k \to \bar{\lambda}$, $\mu^k \to \bar{\mu}$, $\gamma^k = \xi^k - \rho^k \dfrac{\partial \psi}{\partial x} \to \bar{\xi}$, $v^k = \eta^k - \rho^k \dfrac{\partial \psi}{\partial y} \to \bar{\mu}$, and $\nabla_z L(\bar{z}; \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta}) = 0$. Therefore using (12a)-(12b),

$$0 \ \leq \ \lim d^{k^T} \nabla^2_{zz} L_{\rho^k}(z^k; \xi^k, \eta^k, \lambda^k, \mu^k) d^k \ = \ d^T \nabla^2_{zz} L(\bar{z}; \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta}) d. \qquad \square$$

3 Feasibility of Limit Points of the Penalty Method

As we have mentioned, the limit point of stationary points generated by smoothing method (3) or by regularization method (4) is feasible for the original MPCC (1), but this is not automatically true for the penalty method (7). However, this is a consequence of the tradeoff between a tight feasible set $\{z : g(z) \leq 0, h(z) = 0, \Phi^{\varepsilon}(z) = 0\}$ for the smoothing method (3) or $\{z : g(z) \leq 0, h(z) = 0, G(z) \geq 0, H(z) \geq 0, G_i(z)H_i(z) \leq t, i = 1, \ldots, m\}$ for the regularization method (4), and a relaxed feasible set $\{z : g(z) \leq 0, h(z) = 0, G(z) \geq 0, H(z) \geq 0\}$ for the penalty method (7). That is, it is easier to implement the penalty method than the smoothing or regularization methods in the sense that feasible solutions of (7) are easier to find; indeed the feasible set of (7) remains unchanged for all ρ.

As an aside, we note for the smoothing method that it is not immediately clear whether the smoothing problem (3) is feasible. Even under MPCC-LICQ at a feasible point $\bar{z}$ for (1), feasibility of (3) near $\bar{z}$, for small $\varepsilon > 0$, has not been discussed in Ref. 1. However, under MPCC-LICQ, the local feasibility follows some Clarke nonsmooth stability result as developed in Ref. 17. For the regularization method, however, it is clear that the subproblem (4) is feasible if the original MPCC is feasible. Moreover, some nice local relations between solutions of the original MPCC and those of the regularization (4) are established in Section 4 of Ref. 2, under some conditions.

In this section, we give conditions to ensure that the limit points of the sequence $\{z^k\}$ generated by the penalty method (7) are feasible for (1). We investigate also conditions under which $\{z^k\}$ is attracted to a B-stationary point of (1).

3.1 Constraint Qualifications to Ensure Feasibility

In this section we consider a sequence of stationary points z^k of (7) with $\rho = \rho^k \uparrow \infty$ and suppose that $\bar{z}$ is the limit point of $\{z^k\}$. We analyze the properties of $\bar{z}$ for

$$\min_{z \in \mathcal{F}^0} \quad \psi(G(z), H(z)) \tag{14}$$

which can be thought of as the limiting problem as $\rho \to \infty$ of the penalty formulation (7).

We will make use of the standard Mangasarian Fromovitz constraint qualifications (MFCQ) for a smooth nonlinear programs of the general form

$$\begin{aligned} \min \quad & f(z) \\ \text{s.t.} \quad & g(z) \leq 0, \quad h(z) = 0. \end{aligned}$$

The MFCQ holds at a feasible point z of this NLP if first $\nabla h(z)$ has full row rank (all gradients $\nabla h_j(z)$ are linearly independent) and, second, there exists a direction d such that $\nabla h(z)d = 0$ and $\nabla g_i(z)^T d < 0$ for i in the active set $I_g(z) = \{i\,;\, g_i(z) = 0\}$. This is, of course, weaker than the standard LICQ which requires linear independence of the set of all gradients $\nabla g_i(z)$, $i \in I_g(z)$, together with all gradients $\nabla h_j(z)$.

Lemma 3.1 Let $\bar{z}$ be as above. This point is feasible for (14). It is also stationary for (14) if the constraints of (14) satisfy the standard Mangasarian-Fromovitz CQ or linear independence CQ at $\bar{z}$.

Proof. If $\bar{z}$ is feasible for (1), then $\bar{z}$ is a global minimizer of the above problem. So it is a stationary point of the above problem since LICQ holds for the problem.

Now, suppose $\bar{z}$ is not a feasible point of (1). Assume, by taking a subsequence if necessary, that $z^k \to \bar{z}$. Since the LICQ implies the MFCQ, we only give the proof under the MFCQ.

The first equation in (8) is

$$0 = \nabla f(z^k) + \rho^k \nabla \big(\psi \circ (G, H)\big)(z^k) + \nabla g(z^k)^T \lambda^k + \nabla h(z^k)^T \mu^k$$
$$- \nabla G(z^k)^T \xi^k - \nabla H(z^k)^T \eta^k. \tag{15}$$

If $\{(\lambda^k, \mu^k, \xi^k, \eta^k)/\rho^k)\}$ is bounded then, by taking a limit point $(\bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta})$ of this sequence, it is easy to see that

$$0 = \nabla\big(\psi \circ (G, H)\big)(\bar{z}) + \nabla g(\bar{z})^T \bar{\lambda} + \nabla h(\bar{z})^T \bar{\mu} - \nabla G(\bar{z})^T \bar{\xi} - \nabla H(\bar{z})^T \bar{\eta}$$

and, furthermore, that $(\bar{z}, \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta})$ satisfies all the KKT conditions of (14).

Otherwise, we derive a contradiction by dividing both sides of (15) by $\gamma^k = \|(\rho^k, \lambda^k, \mu^k, \xi^k, \eta^k)\|$, and taking a limit point $(\hat{\rho}, \hat{\lambda}, \hat{\mu}, \hat{\xi}, \hat{\eta})$ of the (bounded) sequence $\{(\rho^k, \lambda^k, \mu^k, \xi^k, \eta^k)/\gamma^k\}$ for which $\hat{\rho} = 0$. Also, the vector $(\hat{\lambda}, \hat{\mu}, \hat{\xi}, \hat{\eta})$ has unit length hence is nonzero, and satisfies

$$0 = \nabla g(\bar{z})^T \hat{\lambda} + \nabla h(\bar{z})^T \hat{\mu} - \nabla G(\bar{z})^T \hat{\xi} - \nabla H(\bar{z})^T \hat{\eta}.$$

It can be easily shown that each of $\hat{\lambda}$, $\hat{\xi}$, $\hat{\eta}$ is nonnegative and orthogonal to its respective counterpart $g(\bar{z})$, $G(\bar{z})$, $H(\bar{z})$. This yields a contradiction to the MFCQ for (14) at $\bar{z}$ by the Motzkin theorem of the alternative (see, for example, Ref. 19, pp.28–29). $\qquad\square$

Under some checkable conditions, it is possible to establish the feasibility of the limit point. The next result uses an infeasible-point MPCC-LICQ for this purpose.

Let $\tilde{z}$ satisfy all the constraints of the MPCC (1) with the possible exception of complementarity, that is, $g(\tilde{z}) \le 0, h(\tilde{z}) = 0, G(\tilde{z}) \ge 0, H(\tilde{z}) \ge 0$. We say the infeasible-point MPCC-LICQ holds at $\tilde{z}$ if the gradients

$$\begin{array}{ll}
\nabla g_i(\tilde{z}), & i \in I_g(\tilde{z}), \qquad\qquad \nabla h_j(\tilde{z}), \quad j = 1, \ldots, q, \\
\nabla G_i(\tilde{z}), & i : G_i(\tilde{z}) = 0, \qquad\qquad \nabla H_i(\tilde{z}), \quad i : H_i(\tilde{z}) = 0, \\
\nabla G_i(\tilde{z}), & i : G_i(\tilde{z}), H_i(\tilde{z}) > 0, \qquad \nabla H_i(\tilde{z}), \quad i : G_i(\tilde{z}), H_i(\tilde{z}) > 0
\end{array}$$

are linearly independent. Note that if $\tilde{z}$ is feasible for the MPCC (1), this CQ reduces to the usual MPCC-LICQ.

Lemma 3.2 Let $\bar{z}$ be as given in the beginning of this section. If the infeasible-point MPCC-LICQ holds at $\bar{z}$, then $\bar{z}$ is feasible for the MPCC (1).

Proof. As usual, we assume $z^k \to \bar{z}$, and we write the partial derivatives of ψ without specifying the argument $(G(z^k), H(z^k))$.

If there exists an index i with $G_i(\bar{z}) > 0, H_i(\bar{z}) > 0$, one has $\gamma^k := \xi^k - \rho^k \dfrac{\partial \psi}{\partial x_i} = -\rho^k \dfrac{\partial \psi}{\partial x_i} \to -\infty$ and $v^k := \eta^k - \rho^k \dfrac{\partial \psi}{\partial y_i} = -\rho^k \dfrac{\partial \psi}{\partial y_i} \to -\infty$ by Condition B(b).

Dividing (8a) by $\|(\lambda^k, \mu^k, \gamma^k, v^k)\|$ and taking the limit as $k \to \infty$ yields a contradiction to the linear independence hypothesis. $\qquad\square$

Now, we look at the penalty function $\psi(G(z), H(z)) = \sum_{i=1}^{m} G_i(z)H_i(z)$. Here, we can replace the infeasible-point MPCC-LICQ in Lemma 3.2 by either of the following two weaker constraint qualifications, in which $J(\bar{z}) = \{i : G_i(\bar{z}) > 0, H_i(\bar{z}) > 0\}$:

(CQ1) MFCQ holds at $\bar{z}$ for the following system:

$$
\begin{array}{ll}
g(z) \leq 0, & h(z) = 0, \\
G_i(z) = 0, & i : G_i(\bar{z}) = 0, \\
H_i(z) = 0, & i : H_i(\bar{z}) = 0, \\
G_i(z) \leq G_i(\bar{z}), & H_i(z) \leq H_i(\bar{z}), \quad i \in J(\bar{z}).
\end{array}
$$

(CQ2) MFCQ holds at $\bar{z}$ for the following system:

$$
\begin{array}{ll}
g(z) \leq 0, & h(z) = 0, \\
G_i(z) = 0, & i : G_i(\bar{z}) = 0, \\
H_i(z) = 0, & i : H_i(\bar{z}) = 0, \\
\multicolumn{2}{l}{\displaystyle\sum_{i \in J(\bar{z})} G_i(z)H_i(z) \leq \sum_{i \in J(\bar{z})} G_i(\bar{z})H_i(\bar{z}).}
\end{array}
$$

If $J(\bar{z})$ is empty then the inequalities corresponding to $J(\bar{z})$ in each of (CQ1) and (CQ2) are vacuously omitted.

Lemma 3.3 Let $\psi(G(z), H(z)) = \sum_{i=1}^{m} G_i(z)H_i(z)$ and $\bar{z}$ be as above. Then, (CQ1) implies (CQ2), and (CQ2) implies that $\bar{z}$ is a feasible for (14).

Proof. It is clear that (CQ1) implies (CQ2). Let (CQ2) hold and, for a contradiction, assume that $\bar{z}$ is infeasible, hence $J(\bar{z}) \neq \emptyset$. From (CQ2), there exists a direction d such that

$$
\begin{array}{ll}
\displaystyle\sum_{i \in J(\bar{z})} [G_i(\bar{z})\nabla H_i(\bar{z}) + H_i(\bar{z})\nabla G_i(\bar{z})]^T d < 0, & \\
\nabla g_i(\bar{z})^T d < 0, & i \in I_g(\bar{z}), \\
\nabla h(\bar{z}) d = 0, & \\
\nabla G_i(\bar{z})^T d = 0, & i \in I_G(\bar{z}), \\
\nabla H_i(\bar{z})^T d = 0, & i \in I_H(\bar{z}).
\end{array}
$$

15

Since

$$\sum_{i \in J(\bar{z})} [G_i(\bar{z}) \nabla H_i(\bar{z}) + H_i(\bar{z}) \nabla G_i(\bar{z})]^T d = \nabla \big(\psi \circ (G, H)\big)(\bar{z})^T d,$$

we see that d is a descent direction for the linearized version of (14) at $\bar{z}$. This is equivalent, by classical arguments using by Motzkin's theorem of the alternative (see, for example, Ref. 19), to saying that there do not exist KKT multipliers at $\bar{z}$ for the problem (14) when $\psi(G(z), H(z)) = \sum_{i=1}^m G_i(z) H_i(z)$. This contradicts the conclusion of Lemma 3.1.   □

It is easy to see that Lemma 3.2 is a corollary of Lemma 3.3 when $\psi(G(z), H(z)) = \sum_{i=1}^m G_i(z) H_i(z)$.

3.2 Local Structure of Solutions of the Penalty Problem

One gap in the main convergence result Theorem 2.1 is that the existence of suitable stationary points z^k of (7) with $\rho = \rho^k$ is not addressed. Another gap in the theorem is that the feasibility of the limit point is not established. In this section, we fill in these two gaps at least locally. We show in Corollary 3.1 that a strict local minimizer z^* of the MPCC (1) induces local solutions of the penalty problem (7) for all sufficiently large ρ. Then, in Theorem 3.1, we will describe a situation in which the sequence generated by the penalty method is attracted to a local minimum of the MPCC (1).

Let $B(\tilde{z}, \varepsilon)$ denote the closed ball centred at $\tilde{z}$ with radius ε.

Lemma 3.4 If z^* is a strict local minimum of the MPCC (1), then:

   (i) for any $r > 0$ such that z^* is the only minimum of the MPCC (1) in $B(z^*, r)$, there exists $\sigma(r) > 0$ such that, for any feasible point z of the MPCC (1) with $\|z - z^*\| = r$,

$$f(z) \geq f(z^*) + \sigma(r);$$

and

   (ii) for any $\beta(r) \in (0, \sigma(r))$ where $r, \sigma(r)$ are given in (i), there exists $\rho(r) > 0$ such that

$$f(z) + \rho\psi(z) \geq f(z^*) + \beta(r)$$

holds for any $\rho \geq \rho(r)$ and any feasible point z for the penalty problem (7) with $\|z - z^*\| = r$.

Proof.   The first claim follows from compactness of the sphere $\{z : \|z - a^*\| = r\}$ and the fact that z^* is a strict local minimum of the MPCC (1). We give a proof of the second claim. Let $\sigma > 0, \rho^k \uparrow \infty$ and z^k be feasible for (7) with $\rho = \rho^k$ such that $\|z^k - z^*\| = r > 0$ and $f(z^k) + \rho^k\psi(z^k) \leq f(z^*) + \sigma$, for all k. Since $\{z^k\}$ is bounded, it has a limit point $\bar{z}$ with $\|\bar{z} - z^*\| = r$. Thus

$$0 \leq \psi(\bar{z}) = \lim_{k \to \infty} \psi(z^k) \leq \limsup_{k \to \infty} [f(z^*) - f(z^k) + \sigma]/\rho^k \to 0,$$

16

which means $\bar{z}$ is feasible for (1). Moreover,

$$f(\bar{z}) \leq \limsup_{k \to \infty}(f(z^k) + \rho^k \psi(z^k)) \leq f(z^*) + \sigma.$$

From (i), $\sigma \geq \sigma(r)$. $\qquad\square$

Corollary 3.1 Let z^* be a strict local minimum of (1). Then, for each $r > 0$ and $\rho(r) > 0$ as in Lemma 3.4, and each $\rho \geq \rho(r)$, there exists a local minimum of (7) in the ball of radius r about z^*.

Proof. We denote the problem (7) with $\|z - z^*\| \leq r$ added to its constraints by $(P(\rho, r))$. From (ii) in Lemma 3.4, the global minimum of $(P(\rho, r))$ occurs at a point $\bar{z}(\rho, r)$ in the interior of $B(z^*, r)$, because

(a) z^* lies in the interior of $B(z^*, r)$;

(b) z^* is feasible for $(P(\rho, r))$;

(c) $f(z^*) + \rho\psi(z^*) = f(z^*) < f(z) + \rho\psi(z)$ for any z feasible for $(P(\rho, r))$ that lies on the boundary of $B(z^*, r)$.

A consequence of this interiority property of $\bar{z}(\rho, r)$ is that $\bar{z}(\rho, r)$ is a local minimum of (7). $\qquad\square$

Corollary 3.1 says that the penalty method may generate a convergent sequence $\{z^k\} \to z^*$ if z^* is a strict local minimum of the MPCC (1).

We need one more technical result.

Lemma 3.5 For each feasible point $\tilde{z}$ of the MPCC (1) at which MPCC-LICQ holds, there exists $\varepsilon > 0$ such that the following statement is valid. If $\{z^k\}$ is a sequence of stationary points of (7) corresponding to $\rho = \rho^k \uparrow \infty$, and $\bar{z}$ is a limit point of $\{z^k\}$ in $B(\tilde{z}, \varepsilon)$, then $\bar{z}$ is feasible for (1).

Proof. Let the MPCC-LICQ hold at a feasible point $\bar{z}$ of (1). This implies that the infeasible-point MPCC-LICQ of Lemma 3.2 holds at each z near $\bar{z}$ that satisfies the penalty constraints $g(z) \leq 0$, $h(z) = 0$, $G(z) \geq 0$, $H(z) \geq 0$. The result therefore follows from Lemma 3.2. $\qquad\square$

We come to the main result of this section, which gives sufficient conditions for the sequence generated by the penalty method (7) to be attracted to a B-stationary point z^*. We assume the MPCC-LICQ holds at z^*; hence, there are unique MPCC multipliers λ, μ, ξ, η that satisfy $\nabla_z L(z^*; \lambda, \mu, \xi, \eta) = 0$, where $L(z; \lambda, \mu, \xi, \eta)$ is the MPCC Lagrangian (2). We also require the MPCC to satisfy the strong second order sufficient condition (MPCC-SSOSC) at z^*, as used in Ref. 2:

$$d^T \nabla_{zz}^2 L(z^*; \lambda, \mu, \xi, \eta)d > 0,$$

17

for all directions $d \neq 0$ satisfying

$$
\begin{aligned}
\nabla g_i(z^*)^T d &= 0, & i &: \lambda_i > 0, \\
\nabla h(z^*)d &= 0, & & \\
\nabla_z G_j(z^*)^T d &= 0, & j &: \xi_j \neq 0, \\
\nabla_z H_k(z^*)^T d &= 0, & k &: \eta_k \neq 0.
\end{aligned}
$$

Theorem 3.1 Assume that z^* is a B-stationary point of the MPCC (1) such that MPCC-LICQ, MPCC-SSOSC, and ULSC hold at z^*. Let $S(\rho)$ denote the set of stationary points z of (7) at which WSONC holds. Then, there exists $r^* > 0$ such that

$$
\emptyset \neq S(\rho) \cap B(z^*, r^*) \to z^*, \qquad \text{as } \rho \uparrow \infty.
$$

Proof. Under the assumptions of the theorem, z^* is a strict local minimum and locally unique B-stationary point of the MPCC (1); see Ref. 8; that is, there exists $r_1 > 0$ such that z^* is the unique strict minimum and the unique B-stationary point of (1) in $B(z^*, r_1)$. Given $0 < r \leq r_1$ and $\rho(r)$ from part (ii) of Lemma 3.4, it follows from Corollary 3.1 that $S(\rho)$ is nonempty when $\rho \geq \rho(r)$. It is not hard to deduce from the definitions of MPCC-LICQ and ULSC that, for some $r_2 > 0$ and any $z \in B(z^*, r_2)$, the MPCC-LICQ holds z if it is feasible for (1) and the ULSC holds at z if it is C-stationary for (1). Let $\epsilon > 0$ be given by Lemma 3.5 and $r^* = \min\{r_1, r_2, \varepsilon\}$. It follows from Lemma 3.5 that every limit point $\bar{z}$ of $S(\rho) \cap B(z^*, r^*)$ as $\rho \to +\infty$ is feasible for the MPCC (1). Also, Theorem 2.1 says that $\bar{z}$ is a B-stationary point of (1), thus $\bar{z} = z^*$. $\qquad \square$

That is, for z^* as in Theorem 3.1 and the penalty method as in Theorem 2.1, if any iterate z^k comes within a certain radius of z^* and the penalty parameter ρ^k is sufficiently large, then by increasing the penalty parameter we have reason to expect that $z^k \to z^*$.

References

1. FUKUSHIMA, M., and PANG, J. S., *Convergence of a Smoothing Continuation Method for Mathematical Programs with Complementarity Constraints*, Ill-posed Variational Problems and Regularization Techniques, Edited by M. THÉRA and R. TICHATSCHKE, Springer, New York, NY, pp.99–110, 2000.

2. SCHOLTES, S., *Convergence Properties of a Regularization Scheme for Mathematical Programs with Complementarity Constraints*, SIAM Journal on Optimization, Vol. 11, pp. 918–936, 2001.

3. LUO, Z. Q., PANG, J. S., and RALPH, D., *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, England, 1996.

4. OUTRATA, J., KOČVARA, M., and ZOWE, J., *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Nonconvex Optimization and Its Applications, Vol. 28, Kluwer Academic Publishers, Dordrecht, Holland, 1998.

5. CHEN, Y., and FLORIAN, M., *The Nonlinear Bilevel Programming Problem: Formulations, Regularity and Optimality Conditions*, Optimization, Vol. 32, pp.193–209, 1995.

6. YE, J. J., ZHU, D. L., and ZHU, Q. J., *Exact Penalization and Necessary Optimality Conditions for Generalized Bilevel Programming Problems*, SIAM Journal on Optimization, Vol. 7, pp.481–507, 1997.

7. ROBINSON, S. M., *Stability Theory for Systems of Inequalities, Part II: Differentiable Nonlinear Systems*, SIAM Journal on Numerical Analysis, Vol. 13, pp.497–513, 1976.

8. SCHEEL, H., and SCHOLTES, S., *Mathematical Programs with Complementarity Constraints: Stationarity, Optimality, and Sensitivity*, Mathematics of Operations Research, Vol. 25, pp.1–22, 2000.

9. FERRIS, M. C., and TIN-LOI, F., *Nonlinear Programming Approach for a Class of Inverse Problems in Elastoplasticity*, Structural Engineering and Mechanics, Vol. 6, pp.857–870, 1998.

10. FERRIS, M. C., and TIN-LOI, F., *On the Solution of a Minimum Weight Elastoplastic Problem Involving Displacement and Complementarity Constraints*, Computer Methods in Applied Mechanics and Engineering, Vol. 174, pp.107–120, 1999.

11. TIN-LOI, F., *On the Numerical Solution of a Class of Unilateral Contact Structural Optimization Problems*, Structural Optimization, Vol. 17, pp.155–161, 1999.

12. TIN-LOI, F., and QUE, N. S., *On a Class of Parameter Identification Problems*, Computational Mechanics for the Next Millennium, Edited by C. M. WANG, K. H. LEE and K. K. ANG, Proceedings of APCOM 99, Fourth Asia-Pacific Conference on Computational Mechanics, Elsevier Science, Oxford, England, Vol. 1, pp.311–316, 1999.

13. TIN-LOI, F., and QUE, N. S., *Nonlinear Programming Approaches for an Inverse Problem in Quasibrittle Fracture*, International Journal of Mechanical Sciences, Vol. 44, pp. 843–858, 2002.

14. HUANG, X. X., YANG, X. Q., and ZHU, D. L., *A Sequential Smooth Penalization Approach to Mathematical Programs with Complementarity Constraints*, Manuscript, Department of Applied Mathematics, Hong Kong Polytechnic University, Kowloon, Hong Kong, China, 2001.

15. FISCHER, A., *A Special Newton-Type Optimization Method*, Optimization, Vol. 24, pp.269–284, 1992.

16. KANZOW, C., *Some Noninterior Continuation Methods for Linear Complementarity Problems*, SIAM Journal on Matrix Analysis and Applications, Vol. 17, pp.851–868, 1996.

17. FACCHINEI, F., JIANG, H., and QI, L., *A Smoothing Method for Mathematical Programs with Equilibrium Constraints*, Mathematical Programming, Vol. 85, pp.107–134, 1999.

18. BERTSEKAS, D. P., *Constrained Optimization and Lagrange Multiplier Methods* , Academic Press, New York, NY, 1982.

19. MANGASARIAN, O. L., *Nonlinear Programming*, McGraw-Hill, New York, NY, 1969.