Smooth SQP Methods for Mathematical Programs with Nonlinear Complementarity Constraints*

Houyuan Jiang
CSIRO Mathematical and Information Science
Private Bag 2
Glen Osmond, SA 5064, Australia
Email: Houyuan.Jiang@cmis.csiro.au

Daniel Ralph
The University of Melbourne
Department of Mathematics and Statistics
Parkville, Vic. 3052, Australia
Email: danny@ms.unimelb.edu.au

December 29, 1997; Revised August, 1999 and January 2000.

The final version of this paper appears in
Computational Optimization and Applications, 25 (2002), pp. 123–150

Abstract

Mathematical programs with nonlinear complementarity constraints are reformulated using better-posed but nonsmooth constraints. We introduce a class of functions, parameterized by a real scalar, to approximate these nonsmooth problems by smooth nonlinear programs. This smoothing procedure has the extra benefits that it often improves the prospect of feasibility and stability of the constraints of the associated nonlinear programs and their quadratic approximations. We present two globally convergent algorithms based on sequential quadratic programming, SQP, as applied in exact penalty methods for nonlinear programs. Global convergence of the `implicit smooth SQP` method depends on existence of a lower-level nondegenerate (strictly complementary) limit point of the iteration sequence. Global convergence of the `explicit smooth SQP` method depends on a weaker property, i.e. existence of a limit point at which a generalized constraint qualification holds. We also discuss some practical matters relating to computer implementations.

Key Words. Mathematical programs with equilibrium constraints, bilevel optimization, complementarity problems, sequential quadratic programming, exact penalty, generalized constraint qualification, global convergence, smoothing method.

AMS (MOS) Subject Classifications. 90C, 90D65.

*This work is supported by the Australian Research Council.

1 Introduction

Mathematical programs with equilibrium constraints (MPEC for short) form a relatively new and interesting class of optimization problems. The roots of MPEC lie in game theory and especially bilevel optimization. MPEC include a number of significant applications in economics and engineering. See the monograph [28] for comprehensive theoretical treatment, applications and references.

The MPEC considered in this paper is a mathematical program with nonlinear complementarity problem (NCP) constraints:

$$\begin{array}{rl} \min\limits_{x,y} & f(x,y) \\ \text{subject to} & g(x,y) \geq 0 \\ & 0 \leq F(x,y) \ \perp y \ \geq 0 \end{array} \tag{1}$$

where $f : \Re^{n+m} \to \Re$, $g : \Re^{n+m} \to \Re^l$, $F : \Re^{n+m} \to \Re^m$ are continuously differentiable, and $w \perp y$ indicates orthogonality of any vectors $w, y \in \Re^m$. The constraints $g(x,y) \geq 0$ are called the upper-level constraints. By lower-level or equilibrium constraints we mean the system $0 \leq F(x,y) \perp y \geq 0$, which constitutes a nonlinear complementarity problem in y for each fixed x.

We omit equality constraints in the upper-level for simplicity, but these can easily be handled and would be useful for the following case. Lower-level mixed complementarity constraints [7] can be dealt with quite easily by moving equations and their associated variables to the upper level. For example, consider the following lower-level mixed complementarity constraints

$$\begin{array}{l} F_1(x,y,z) = 0 \\ 0 \leq F_2(x,y,z) \ \perp \ z \geq 0, \end{array}$$

where $F_1 : \Re^{n+m_1+m_2} \to \Re^{m_1}$, $F_2 : \Re^{n+m_1+m_2} \to \Re^{m_2}$. By renaming the tuple (x,y) as the upper-level vector and z as the lower-level vector, and moving the equations $F_1(x,y,z) = 0$ to the upper level, we obtain an MPEC with upper-level constraints that are specified by nonlinear equalities and inequalities, and lower-level nonlinear complementarity constraints.

Clearly, the MPEC (1) is equivalent to the smooth nonlinear program (NLP) obtained by writing the complementarity condition $F(x,y) \perp y$ as an inner product $F(x,y)^T y = 0$. Unfortunately, it has been proved [4] that the Mangasarian-Fromovitz Constraint Qualification does not hold at any feasible point of this smooth NLP even if the usual inequality constraints $g(x,y) \geq 0$ are omitted and the lower-level NCP problem has very fine properties such as strong monotonicity with respect to y. Since this constraint qualification is almost synonymous with numerical stability of the feasible set, its failure to hold suggests that well-developed nonlinear programming theory and numerical methods are not readily applicable for solving this form of MPEC: the feasible set of the smooth NLP is numerically ill posed. See [19, 28] for more discussions and numerical examples.

Instead we let $w = F(x,y)$ and substitute a nonsmooth equation $\Phi(y,w) = 0 \in \Re^m$, constructed using the Fischer-Burmeister functional [9] for example, for the complementarity problem $y, w \geq 0$, $y^T w = 0$:

$$\begin{array}{ll} \min\limits_{x,y,w} & f(x,y) \\ \text{subject to} & g(x,y) \geq 0 \\ & F(x,y) - w = 0 \\ & \Phi(y,w) = 0. \end{array} \tag{2}$$

The mapping Φ is then "smoothed" by introducing a parameterization $\Psi(y, w, \mu)$ that is differentiable if the scalar μ is nonzero but coincides with $\Phi(y, w)$ when $\mu = 0$. By a smoothing method we mean an algorithm that solves (1) either by solving an augmented problem like

$$\begin{array}{ll} \min\limits_{x,y,w,\mu} & f(x,y) \\ \text{subject to} & g(x,y) \geq 0 \\ & F(x,y) - w = 0 \\ & \Psi(y,w,\mu) = 0, \\ & e^{\mu} - 1 = 0, \end{array} \tag{3}$$

where e is Euler's constant, so that the last constraint requires $\mu = 0$ (cf. [18] for complementarity problems); or by approximately solving the following problem for a sequence of values $\mu = \mu_k \to 0$,

$$\begin{array}{ll} \min\limits_{x,y,w} & f(x,y) \\ \text{subject to} & g(x,y) \geq 0 \\ & F(x,y) - w = 0 \\ & \Psi(y,w,\mu) = 0. \end{array} \tag{4}$$

The introduction of the smoothing parameter μ has three consequences: Nonsmooth problems are transformed into smooth problems, except when $\mu = 0$; well-posedness can be improved in the sense that feasibility and constraint qualifications, hence stability, are often more likely to be satisfied for all values of μ; and solvability of quadratic approximation problems is improved. This opens the way to use sequential quadratic programming (SQP) methods from classical nonlinear programming.

The methods presented in this paper follow some ideas from [8, 12] which try to use well-developed numerical methods for the solution of smooth nonlinear programs. In [8], smooth nonlinear programs of the type (4) are formed and assumed to be solvable by an unspecified (black box) method. Under further conditions — that will be relaxed in the explicit smoothing method to be presented in Section 6 — it is shown that limit points of the sequence of approximate solutions of the parametric nonlinear programs satisfy generalized Karush-Kuhn-Tucker (KKT) conditions [16] given in terms of the Clarke generalized derivatives [5]. We call this an `explicit` smoothing method because the smoothing parameter is updated separately from the direction-finding process. In [12] another explicit smoothing method is proposed, which is an SQP-based method for MPEC with linear complementarity constraints and upper-level constraints only on x, and limit points satisfying a lower-level nondegeneracy (strict complementarity) condition are shown to be piecewise stationary points for (1).

This paper details methods for solving the problems (2) and (3) using SQP in an ℓ_1-exact penalty framework. The first method, `implicit smooth SQP`, applies to (3); Theorem 5.10 assumes lower-level nondegeneracy at limit points amongst other conditions to ensure that limit points of the iteration sequence are piecewise stationary points of (1). The term implicit means that the smoothing parameter is included as one of the variables of the problem formulation and updated at each iteration using the QP solution, like the other variables. Neither of these convergence results is surprising given that lower-level nondegeneracy at a feasible point of (1) implies locally that the problem is a smooth nonlinear program.

To move beyond the realm of standard nonlinear programming, we present the `explicit smooth SQP` method that is aimed at solving the problem (2) by solving a sequence of problems (4) where we expect $\mu = \mu_k \to 0$ and limit points of the iteration sequence to satisfy generalized Karush-Kuhn-Tucker (KKT) conditions of (2). The result of perhaps the most novelty is Theorem 6.4 which extends global convergence theory for exact penalty methods in nonlinear programming to MPEC by using a generalized constraint qualification at limit points of the iteration sequence. Explicit smooth SQP can be viewed as an implementation of the smoothing method of [8] though the convergence analysis of the new method is more demanding.

Our main goal is to explore convergence conditions and analysis for smooth SQP methods. Given this and the length of the paper, a numerical investigation will be pursued in a future publication.

We mention that the development of numerical methods for the solution of MPEC is at a less advanced stage than optimality theory [4, 26, 27, 28, 29, 30, 32, 33, 34, 40, 43]. When the upper-level constraints exclude y, i.e. take the form $g(x) \geq 0$, the implicit function approach may be possible. In this approach it is assumed that y can be found as a function of x by solving the NCP appearing in the constraints, and the MPEC is collapsed to the problem of minimizing the nondifferentiable objective function $f(x, y(x))$ subject to $g(x) \geq 0$. This nonsmooth problem can be tackled by bundle methods as proposed in [23, 24, 33, 34] or using another nonsmooth optimization method such as Shor's R-algorithm as implemented in SolvOpt [25]; see [7] for some computational comparisons. However with mixed upper-level constraints, i.e. involving y and possibly x, the implicit programming approach transforms an MPEC into a problem with nondifferentiable constraints in addition to a nonsmooth objective, a format which has not been seriously studied with regard to computational methods.

Some methods which can be extended to handle mixed upper-level constraints include the penalty interior-point algorithm (PIPA for short) [28], smoothing methods [8, 12], which are related to the interior-point approach, and piecewise sequential quadratic programming (PSQP) [28, 29, 37]. Apart from this paper, the only implementations of these algorithms we know of that handle joint upper-level constraints are discussed in [19]. Penalty interior-point algorithms converge globally under suitable conditions, at least in the implicit case [28], while the piecewise SQP method exhibits local superlinear convergence under the uniqueness of multipliers and some second-order sufficient conditions, but surprisingly without requiring a strict complementarity condition. Some preliminary numerical experiments have been carried out for the PIPA and PSQP [28, 29, 19], and smoothing methods [8, 12]. See also [7] for a comparison of PIPA with implicit programming methods. The theoretical results and numerical experience show some promise for these methods. We also refer the reader to [24, 33, 34] for other numerical methods and applications of MPEC.

The rest of the paper is organized as follows. In the next section, we review first-order optimality theory for nonsmooth programs using the Clarke calculus. In Section 3, we reformulate the MPEC (1) into equivalent — in the sense of global optima, local optima, generalized stationarity, or piecewise stationarity as the case may be — but generally better-posed nonsmooth programs by means of functions introduced there. Constraint qualifications for the reformulated nonsmooth programs are studied in Section 4. In Section 5, we present implicit smooth SQP for solving the reformulation (3) and give details of global convergence under lower-level nondegeneracy at limit points. In Section 6, we propose explicit smooth SQP and establish its global convergence to generalized KKT points under generalized constraint qualifications; the analogs of the various results developed in Section 5 are given here. Section 7 briefly gives concrete examples of smoothing functions from the literature.

A word about notation: For a locally Lipschitz real-valued function f and a vector-valued locally Lipschitz function H, ∂f and ∂H denote the Clarke generalized subgradient and the Clarke generalized Jacobian respectively, see [5]. For a continuously differentiable real-valued function f and a vector-valued continuously differentiable function H, we use ∇f and F' to indicate the gradient of f and the Jacobian of H. If x_1 and x_2 are two vectors with the same dimension, then $x_1^T x_2$ denotes the inner products of these two vectors. By $\|\cdot\|$, we mean the Euclidean norm. $\Re^n$ denotes the real Euclidean space of column vectors of length n; for $u \in \Re^n$ and $v \in \Re^m$, (u, v) denotes the column vector $[u^T \ v^T]^T$ in $\Re^{n+m}$.

2 Preliminaries on Nonsmooth Programming

Consider the nonsmooth program or NSP:

$$\begin{array}{rl} \min\limits_{u} & f(u) \\ \text{subject to} & g(u) \geq 0 \\ & h(u) = 0, \end{array} \tag{5}$$

where $f : \Re^n \to \Re$, $g : \Re^n \to \Re^l$ and $h : \Re^n \to \Re^m$ are locally Lipschitz.

Definition 2.1 *The point u^* is said to be a generalized stationary point of (5) if there exists a Karush-Kuhn-Tucker (KKT) multiplier vector $(\lambda_g, \lambda_h) \in \Re^{l+m}$ such that the following generalized Karush-Kuhn-Tucker (GKKT) conditions hold:*

$$\begin{array}{l} \partial f(u^*) - \partial g(u^*)^T \lambda_g + \partial h(u^*)^T \lambda_h \ni 0 \\ 0 \leq g(u^*) \ \perp \ \lambda_g \geq 0 \\ h(u^*) = 0, \end{array}$$

where ∂ denotes the Clarke generalized gradient for a scalar function and the Clarke generalized Jacobian for a vector-valued function [5].

If f, g and h happen to be smooth at u^*, then the GKKT conditions reduce to the usual Karush-Kuhn-Tucker (KKT) condition:

$$\begin{array}{l} \nabla f(u^*) - g'(u^*)^T \lambda_g + h'(u^*)^T \lambda_h = 0 \\ 0 \leq g(u^*) \ \perp \ \lambda_g \geq 0 \\ h(u^*) = 0. \end{array}$$

In this case, u^* is called a stationary point or a KKT point of (5).

For convenience, we may assume that in the above NSP, the first l_1 ($l_1 \leq l$) inequality constraints are active and the rest are inactive at u^*, i.e.,

$$\begin{array}{l} g_i(u^*) = 0, \quad 1 \leq i \leq l_1 \\ g_i(u^*) > 0, \quad i > l_1. \end{array}$$

Let

$$G(u) \;=\; \begin{pmatrix} g_1(u) \\ \vdots \\ g_{l_1}(u) \\ h_1(u) \\ \vdots \\ h_m(u) \end{pmatrix}.$$

Associated with the above NSP, we recall some well-known regularity conditions under which a local solution is a generalized stationary point [16].

`Generalized Linear Independence Constraint Qualification` (GLICQ): Each element of the generalized Jacobian $\partial G(u^*)$ [5] has full row rank.

`Generalized Mangasarian-Fromovitz Constraint Qualification` (GMFCQ): (i) there exists $d \in \Re^n$ such that for all element $(A_1, \ldots, A_{l_1}, B_1, \ldots, B_m) \in \partial G(u^*)$,

$$\begin{array}{l} A_i^T d > 0, \text{ for } i = 1, \ldots, l_1 \\ B_j^T d = 0, \text{ for } j = 1, \ldots, m; \end{array}$$

(ii) for any element of $(A_1, \ldots, A_{l_1}, B_1, \ldots, B_m) \in \partial G(u^*)$, $(B_1, \ldots, B_m)$ has full row rank.

`Generalized Constant Rank Constraint Qualification` (GCRCQ): There is a neighborhood of u^* such that for any u in this neighborhood, rank of each element of the generalized Jacobian $\partial G(u)$ is invariant.

We mention that the above three constraint qualifications are slightly stronger than those given in [16] to keep notation simple. When (5) is defined by smooth (continuously differentiable) functions, the GLICQ and GMFCQ reduce to the classical LICQ and MFCQ, see [28, 31]; and, as in the smooth case, the GLICQ implies the GMFCQ. However the CRCQ usually used in the smooth case [17] is stronger than the smooth version of the GCRCQ in that the former requires constant rank of submatrices of rows of the Jacobian $G'(u)$ for u near u^*. (We mention an example to distinguish these two CRCQs: let $g(u_1, u_2) = (u_1 + u_2, \, (u_1 + u_2)^2)$, and observe that the rank of $g'(u_1, u_2)$ is always one, whereas the rank of $g_2'(u_1, u_2)$ is either zero if $(u_1, u_2) = (0, 0)$, or one otherwise.)

Under these generalized CQs, Hiriart-Urruty [16] proved the optimality conditions in the following proposition. These optimality conditions also hold under the next assumption:

`Piecewise Affine Constraint Condition` (PACC): Both g and h are piecewise affine.

See [41] for a proof of generalized stationarity under the PACC.

Proposition 2.2 *Suppose u^* is a local minimizer of the nonsmooth program (5) and one of the GCRCQ, GLICQ, GMFCQ or PACC holds at u^*. Then u^* is a generalized stationary point of (5). Furthermore, if f, g and h are smooth at u^*, then u^* is a stationary point or a KKT point of (5).*

3 Equivalent Reformulations of MPEC

As explained in Section 1, the smooth nonlinear programming reformulation of the MPEC (1) is numerically ill posed. The strategy we use in this article is to approximate the MPEC by well-behaved nonlinear programming problems. To this end, we introduce a class of smoothing functions on which some properties are imposed as we proceed. Suppose the function $\psi : \Re^3 \to \Re$ satisfies the following assumptions:

(A1) ψ is locally Lipschitz and directionally differentiable on $\Re^3$, and ψ is continuously differentiable at every point (a, b, c) with $c \neq 0$.

(A2) $\psi(a, b, 0) = 0$ if and only if $a \geq 0,\ b \geq 0,\ ab = 0$.

Section 7 contains standard examples all of which satisfy the assumptions (A1)–(A2) and the assumptions (A3)–(A5) to be introduced in the sequel.

Let $\phi : \Re^2 \to \Re$ and the parametric function $\phi_c : \Re^2 \to \Re$ be defined for any $(a, b) \in \Re^2$ and $c \in \Re$ by

$$\phi(a,b) \;=\; \psi(a,b,0),$$

and

$$\phi_c(a,b) \;=\; \psi(a,b,c).$$

Clearly, $\phi_0 \equiv \phi$. By means of the functions ϕ and ψ, we define two nonsmooth programs:

$$\begin{array}{ll} \min\limits_{x,y,w} & f(x,y) \\ \text{subject to} & g(x,y) \geq 0 \\ & F(x,y) - w = 0 \\ & \phi(y_i, w_i) = 0, \quad i = 1, \ldots, m \end{array} \tag{6}$$

and

$$\begin{array}{ll} \min\limits_{x,y,w,\mu} & f(x,y) \\ \text{subject to} & g(x,y) \geq 0 \\ & F(x,y) - w = 0 \\ & \psi(y_i, w_i, \mu) = 0, \quad i = 1, \ldots, m \\ & e^{\mu} - 1 = 0. \end{array} \tag{7}$$

It is easy to see that (6) and (7) are (2) and (3) respectively with

$$\Psi(y,w,\mu) \;=\; \begin{pmatrix} \psi(y_1, w_1, \mu) \\ \vdots \\ \psi(y_m, w_m, \mu) \end{pmatrix}$$

and

$$\Phi(y,w) \;=\; \begin{pmatrix} \phi(y_1, w_1) \\ \vdots \\ \phi(y_m, w_m) \end{pmatrix} \;=\; \Psi(y,w,0).$$

Since differentiability of ϕ and ψ is not assumed at (a,b) and (a,b,c) respectively, (6) and (7) are nonsmooth programs in general. On the other hand, by the assumption (A1), when $\mu \neq 0$, the functions involved in (7) are smooth at (x,y,w,μ), which is a nice property to be used in the subsequent analysis. Next we give some relationships between the MPEC (1) and the nonsmooth programs (6) and (7).

Proposition 3.1 *Under the assumptions (A1) and (A2), the following statements are equivalent.*

(i) (x,y) *is a feasible point (local solution, global solution) of (1).*

(ii) (x,y,w) *with* $w = F(x,y)$ *is a feasible point (local solution, global solution) of (6).*

(iii) (x,y,w,μ) *with* $w = F(x,y)$ *and* $\mu = 0$ *is a feasible point (local solution, global solution) of (7).*

Proof. Given that $e^{\mu} - 1 = 0$ has a unique solution $\mu = 0$ and the assumption (A2) is satisfied, it is clear that all three statements are equivalent regarding feasible points. The equivalence with respect to local solutions or global solutions is an obvious consequence. ∎

Since f, g and F are smooth, it can be shown, by Proposition 2.3.3 of [5] and its Corollary 1, that (x^*, y^*, w^*) is a generalized stationary point of (6) if and only if the following holds: there exists a KKT multiplier vector $(\lambda_g, \lambda_F, \lambda_\Phi) \in \Re^{l+2m}$ such that the following GKKT conditions hold:

$$\begin{pmatrix} \nabla f(x^*, y^*) \\ 0 \end{pmatrix} - \begin{pmatrix} g'(x^*, y^*)^T \\ 0 \end{pmatrix} \lambda_g + \begin{pmatrix} F'(x^*, y^*)^T \\ -I \end{pmatrix} \lambda_F$$

$$+ \begin{pmatrix} 0 \\ \partial\Phi(y^*, w^*) \end{pmatrix} \lambda_\Phi \ni 0 \tag{8}$$

$$\begin{array}{l} 0 \leq g(x^*, y^*) \ \perp \ \lambda_g \geq 0 \\ F(x^*, y^*) - w^* = 0 \\ \Phi(y^*, w^*) = 0, \end{array}$$

where 0 denotes appropriate zero vectors or matrices, and $I \in \Re^{m \times m}$ is the identity matrix. Similarly, (x^*, y^*, w^*, μ^*) is a generalized stationary point of (7) if and only if there exists a KKT multiplier vector $(\lambda_g, \lambda_F, \lambda_\Psi, \lambda_\mu) \in \Re^{l+2m+1}$ such that the following

GKKT conditions hold:

$$\begin{pmatrix} \nabla f(x^*, y^*) \\ 0 \end{pmatrix} - \begin{pmatrix} g'(x^*, y^*)^T \\ 0 \end{pmatrix} \lambda_g + \begin{pmatrix} F'(x^*, y^*)^T \\ -I \\ 0 \end{pmatrix} \lambda_F$$

$$+ \begin{pmatrix} 0 \\ \partial\Psi(y^*, w^*, \mu^*) \end{pmatrix} \lambda_\Psi + \begin{pmatrix} 0 \\ e^{\mu^*} \end{pmatrix} \lambda_\mu \ni 0 \tag{9}$$

$$\begin{array}{l} 0 \le g(x^*, y^*) \;\perp\; \lambda_g \ge 0 \\ F(x^*, y^*) - w^* = 0 \\ \Psi(y^*, w^*, \mu^*) = 0 \\ e^{\mu^*} - 1 = 0. \end{array}$$

Note in (9) that $\mu^* = 0$.

The assumption (A1) ensures the following inclusion

$$\Pi_{ab} \partial\psi(a, b, 0) \;\supseteq\; \partial\phi(a, b)$$

for any $(a, b) \in \Re^2$, where Π_{ab} denotes the projection operator on $\Re^3$: $\Pi_{ab}(\alpha, \beta, 0) = (\alpha, \beta)$; see Proposition 2.3.16 in [5]. We introduce another assumption to ensure that these sets are in fact identical.

(A3) $\Pi_{ab}\partial\psi(a, b, 0) = \partial\phi(a, b)$ for any $(a, b) \in \Re^2$, where $\phi(a, b)$ is defined as $\psi(a, b, 0)$.

A direct consequence of the assumption (A3) is that

$$\Pi_{ab}\partial\Psi(y, w, 0) \;=\; \partial\Phi(y, w), \quad \forall (y, w) \in \Re^{2m}.$$

Proposition 3.2 *Under the assumptions (A1)–(A2), if (x^*, y^*, w^*) is a generalized stationary point of (6), then $(x^*, y^*, w^*, 0)$ is a generalized stationary point of (7). Conversely, if (A3) holds as well as (A1)–(A2), and if $(x^*, y^*, w^*, 0)$ is a generalized stationary point of (7), then (x^*, y^*, w^*) is a generalized stationary point of (6).*

Proof. Suppose (x^*, y^*, w^*) is a generalized stationary point of (6), then there exists a KKT multiplier vector $(\lambda_g, \lambda_F, \lambda_\Phi)$ such that (8) holds. Let λ_μ be an element of $-\partial_\mu\Psi(y^*, w^*, \mu^*)\lambda_\Phi$ with $\mu^* = 0$. It follows from the remark before the assumption (A3) that $(\lambda_g, \lambda_F, \lambda_\Phi, \lambda_\mu)$ is a KKT multiplier satisfying (9); i.e., $(x^*, y^*, w^*, 0)$ is a generalized stationary point of (7). Conversely, if $(x^*, y^*, w^*, 0)$ is a generalized stationary point of (7), it is easy to see from the assumption (A3) and the GKKT conditions (8) and (9) that (x^*, y^*, w^*) is a generalized stationary point of (6). ■

By Propositions 3.1 and 3.2, (6) and (7) are completely equivalent in the sense that global solutions, local solutions, generalized stationary points and feasible points correspond one another. However, it is not yet clear what relationships the optimality condition of the MPEC (1) and the nonlinear programming problems (6) and (7) have.

Let $z^* = (x^*, y^*)$ be a feasible point of the MPEC (1). Let $\mathcal{F}$ be the feasible set of (1), i.e.,

$$\mathcal{F} \;=\; \{z = (x, y) \;:\; g(z) \ge 0,\, 0 \le F(z) \;\perp\; y \ge 0\}.$$

Denote by $\mathcal{T}(z^*, \mathcal{F})$ the tangent cone to $\mathcal{F}$ at z^*: $\mathcal{T}(z^*, \mathcal{F})$ is the set of limit points of sequences $\{\frac{z^k - z^*}{\tau_k}\}$ where $\{z^k\} \subseteq \mathcal{F}$ converges to z^* and $\tau_k \downarrow 0$.

Definition 3.3 *A point $z^* \in \mathcal{F}$ is said to be a primal stationary [28] or B-stationary [40] point of the MPEC (1) if the following condition holds:*

$$\nabla f(z^*)^T d \geq 0, \ \forall d \in \mathcal{T}(z^*, \mathcal{F}).$$

A decomposition or disjunction technique was very useful in establishing optimality conditions for MPEC in [28]. For any feasible point $z^* \in \mathcal{F}$, define

$$\begin{array}{rcl} \alpha(z^*) & = & \{1 \leq i \leq m : F_i(z^*) = 0 < y_i^*\} \\ \beta(z^*) & = & \{1 \leq i \leq m : F_i(z^*) = 0 = y_i^*\} \\ \gamma(z^*) & = & \{1 \leq i \leq m : F_i(z^*) > 0 = y_i^*\} \end{array} \tag{10}$$

and the family of index sets

$$\mathcal{A}(z^*) \ = \ \{(\mathcal{J}, \mathcal{K}) \, : \, \mathcal{J} \supseteq \alpha(z^*), \, \mathcal{K} \supseteq \gamma(z^*), \, \mathcal{J} \cap \mathcal{K} = \emptyset, \, \mathcal{J} \cup \mathcal{K} = \{1, 2, \ldots, m\}\}.$$

For each partition $\mathcal{J} \cup \mathcal{K}$ of $\{i \, : \, 1 \leq i \leq m\}$, let

$$\begin{array}{rcll} \mathcal{F}_{(\mathcal{J},\mathcal{K})} & = & \{z : & g(z) \geq 0, \\ & & & F_i(z) = 0 \leq y_i, \quad \forall i \in \mathcal{J}, \\ & & & F_i(z) \geq 0 = y_i, \quad \forall i \in \mathcal{K} \ \}. \end{array}$$

Using the family of sets $\{\mathcal{F}_{(\mathcal{J},\mathcal{K})} \, : \, (\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^*)\}$, the feasible set $\mathcal{F}$ of (1) can be locally decomposed at any feasible point $z^* \in \mathcal{F}$, and hence stationarity conditions for (1) defined in [28] can be characterized in terms of traditional nonlinear programs associated with each $\mathcal{F}_{(\mathcal{J},\mathcal{K})}$, which has the form of a standard nonlinear programming feasible region. The disjunctive approach can be carried over to constraint stability.

`Piecewise Constraint Qualification` at a point $z^* \in \mathcal{F}$: For each $(\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^*)$, the above representation of $\mathcal{F}_{(\mathcal{J},\mathcal{K})}$ satisfies a standard smooth constraint qualification at z^* (for example, the MFCQ, LICQ, or CRCQ).

We now state a disjunctive first-order optimality condition studied in [28] where it was called "primal-dual stationarity"; we call it "piecewise stationarity" to distinguish it from generalized stationarity which also has a primal-dual flavor.

Definition 3.4 *A point $z^* = (x^*, y^*)$ is a piecewise stationary point of the MPEC (1) if it is feasible and, for each $(\mathcal{J}, \mathcal{K}) \in \mathcal{A}(z^*)$, there exist KKT multipliers $\xi \in \Re^l$, $\eta \in \Re^m$ and $\pi \in \Re^m$ such that*

$$\begin{array}{l} \nabla_x f(x^*, y^*) - g_x'(x^*, y^*)^T \xi - F_x'(x^*, y^*)^T \eta = 0, \\ \nabla_y f(x^*, y^*) - g_y'(x^*, y^*)^T \xi - F_y'(x^*, y^*)^T \eta - \pi = 0, \\ 0 \leq g(x^*, y^*) \ \perp \ \xi \geq 0, \\ F_i(x^*, y^*) = 0, \ 0 \leq y_i^* \ \perp \ \pi_i \geq 0, \quad \forall i \in \mathcal{J}, \\ 0 \leq F_i(x^*, y^*) \ \perp \ \eta_i \geq 0, \ y_i^* = 0, \quad \forall i \in \mathcal{K}. \end{array} \tag{11}$$

The next result is essentially due to [28].

Proposition 3.5 *Let $z^* = (x^*, y^*)$. If z^* is a piecewise stationary point of the MPEC (1) then it is primal stationary for (1). Conversely, if z^* is primal stationary for (1) and a piecewise constraint qualification holds at z^*, then z^* is piecewise stationary for (1).*

The idea of strict complementarity of a solution of a complementarity problem is adapted to nonfeasible points of the MPEC (1).

Definition 3.6 *A point $(x, y) \in \Re^{n+m}$ is said to be* `lower-level nondegenerate` *if $y_i \neq F_i(x, y)$ for $i = 1, \dots, m$. A point $(x, y, w) \in \Re^{n+2m}$ is said to be* `lower-level nondegenerate` *if $y_i \neq w_i$ for $i = 1, \dots, m$.*

Suppose $z^* = (x^*, y^*)$ is feasible for the MPEC (1). Then lower-level nondegeneracy of (x^*, y^*) is equivalent to the strict complementarity condition: for any i ($1 \leq i \leq m$), either $y_i^* > 0 = F_i(x^*, y^*)$ or $y_i^* = 0 < F_i(x^*, y^*)$. Lower-level nondegeneracy of (x^*, y^*) is also equivalent to lower-level nondegeneracy of (x^*, y^*, w^*) with $w^* = F(x^*, y^*)$; and to the family of index sets $\mathcal{A}(z^*)$ reducing to a singleton, i.e. $\mathcal{A}(z^*) = \{(\alpha(z^*), \gamma(z^*))\}$. If the function Φ is continuously differentiable at such a feasible lower-level nondegenerate point (x^*, y^*, w^*), then piecewise stationarity of (6) at (x^*, y^*, w^*) coincides with the classical KKT conditions.

The next result shows that stationarity conditions on the MPEC (1), (6) and (7) coincide at lower-level nondegenerate points. To this end, we impose another condition on the function ψ.

(A4) If $\psi(a, b, 0) = 0$ and $(p, q, r) \in \partial\psi(a, b, 0)$, then

$$p^2 + q^2 > 0, \ pq \geq 0,$$

and

$$\begin{array}{l} p = 0, \ q \neq 0, \ \text{if } a > 0 \\ p \neq 0, \ q = 0, \ \text{if } b > 0. \end{array}$$

Proposition 3.7 *Suppose (x^*, y^*) is a lower-level nondegenerate feasible point of the MPEC (1). Assume that the assumptions (A1)–(A4) are satisfied. Then the following statements are equivalent.*

(i) (x^*, y^*) is a piecewise stationary point of the MPEC (1).

(ii) (x^*, y^*, w^*) is a (generalized) stationary point of (6), where $w^* = F(x^*, y^*)$.

(iii) (x^*, y^*, w^*, μ^*) is a (generalized) stationary point of (7), where $w^* = F(x^*, y^*)$ and $\mu^* = 0$.

Proof. (i) $\Longrightarrow$ (ii). Let $\mathcal{J} = \alpha(z^*)$ and $\mathcal{K} = \gamma(z^*)$. $(\mathcal{J}, \mathcal{K})$ is the unique element of $\mathcal{A}(z^*)$ since $\beta(z^*) = \emptyset$. It follows that there exist multipliers $\xi \in \Re^l$, $\eta \in \Re^m$ and $\pi \in \Re^p$ such that (11) holds.

Let $\lambda_g = \xi$, $\lambda_F = -\eta$. We now define a vector λ_Φ. For $i \in \mathcal{J} = \alpha(z^*)$ and $(A_i, B_i) \in \partial\phi(y_i^*, w_i^*)$, the assumption (A4) implies that $A_i = 0, B_i \neq 0$. Therefore, $(\lambda_\Phi)_i = \frac{(\lambda_F)_i}{B_i}$ is well-defined for any $i \in \alpha(z^*)$. Similarly, $(\lambda_\Phi)_i = \frac{(-\pi)_i}{A_i}$ is well-defined with $(A_i, B_i) \in \partial\phi(y_i^*, w_i^*)$ for any $i \in \gamma(z^*)$ by the assumption (A4).

By the assumption (A4), it is easy to verify that $(\lambda_g, \lambda_F, \lambda_\Phi)$ is a KKT multiplier such that the GKKT conditions (8) holds, i.e., (x^*, y^*, w^*) is a generalized stationary point of (6).

(ii) $\Longrightarrow$ (i). Suppose there exists a KKT multiplier $(\lambda_g, \lambda_F, \lambda_\Phi) \in \Re^{l+2m}$ such that (8) holds at (x^*, y^*, w^*). Let

$$\begin{pmatrix} A \\ B \end{pmatrix} \in \ \partial\Phi(y^*, w^*).$$

Clearly $A = \text{diag}(A_1, \ldots, A_n)$ and $B = \text{diag}(B_1, \ldots, B_n)$ are diagonal matrices. Since $z^* = (x^*, y^*)$ is lower-level nondegenerate, $y_i^* \neq w_i^*$ for $i = 1, \ldots, m$ and $\beta(z^*) = \emptyset$. By the assumptions (A3) and (A4), $A_i = 0$ for $i \in \alpha(z^*)$ and $B_i = 0$ for $i \in \gamma(z^*)$. Moreover, it can be shown from (8) that $-\lambda_F + B\lambda_\Phi = 0$, i.e., $\lambda_F = B\lambda_\Phi$. Let $\xi = \lambda_g$, $\eta = -\lambda_F = -B\lambda_\Phi$ and $\pi = -A\lambda_\Phi$. We immediately obtain that (ξ, η, π) is a KKT multiplier such that (11) holds for the given $(\mathcal{J}, \mathcal{K})$ such that $\mathcal{J} = \alpha(z^*)$ and $\mathcal{K} = \gamma(z^*)$. Since z^* is a lower-level nondegenerate feasible point, $(\mathcal{J}, \mathcal{K})$ is the only element in $\mathcal{A}(z^*)$. This proves that (i) holds.

The desired results follow from Proposition 3.2. ■

We next study the relationship between the piecewise stationary point and the generalized stationary point of (6) or (7) under the assumptions (A1)–(A3).

Suppose (x^*, y^*) is a piecewise stationary point of the MPEC (1). It turns out that $(dx, dy) = 0 \in \Re^{n+m}$ is a local solution of the following MPEC:

$$\begin{array}{ll} \min\limits_{dx,dy} & \nabla f(x^*, y^*)^T (dx, dy) \\ \text{subject to} & g'(x^*, y^*)(dx, dy) + g(x^*, y^*) \geq 0 \\ & 0 \leq y^* + dy \ \perp \ F(x^*, y^*) + F'(x^*, y^*)(dx, dy) \geq 0, \end{array}$$

or that $(dx, dy, dw) = 0 \in \Re^{n+2m}$ is a local solution of the following MPEC:

$$\begin{array}{ll} \min\limits_{dx,dy,dw} & \nabla f(x^*, y^*)^T (dx, dy) \\ \text{subject to} & g'(x^*, y^*)(dx, dy) + g(x^*, y^*) \geq 0 \\ & F(x^*, y^*) + F'(x^*, y^*)(dx, dy) - (w^* + dw) = 0 \\ & 0 \leq y^* + dy \ \perp \ w^* + dw \geq 0, \end{array}$$

or, by Proposition 3.2, that $(dx, dy, dw) = 0 \in \Re^{n+2m}$ is a local solution of the following nonsmooth program:

$$\begin{array}{ll} \min\limits_{dx,dy,dw} & \nabla f(x^*, y^*)^T (dx, dy) \\ \text{subject to} & g'(x^*, y^*)(dx, dy) + g(x^*, y^*) \geq 0 \\ & F(x^*, y^*) + F'(x^*, y^*)(dx, dy) - (w^* + dw) = 0 \\ & \Phi(y^* + dy, w^* + dw) = 0. \end{array} \tag{12}$$

Then under the GLICQ or GMFCQ at 0 on the last nonsmooth problem (12), which is equivalent to the GLICQ or GMFCQ at (x^*, y^*, w^*) on the problem (6), we have that $0 \in \Re^{n+2m}$ is a generalized stationary point of (12), which is equivalent to saying that (x^*, y^*, w^*) is a generalized stationary point of (6). Similarly if the PACC holds for (6), so that g, F and ϕ are affine functions, then piecewise stationarity of (1) implies generalized stationarity.

The following result summarizes the above discussion.

Proposition 3.8 *Suppose the assumptions (A1)–(A3) hold. Suppose (x^*, y^*) is a piecewise stationary point of the MPEC (1). Assume that the GLICQ, GMFCQ, or PACC is satisfied at (x^*, y^*, w^*) with $w^* = F(x^*, y^*)$ for the problem (6). Then (x^*, y^*, w^*) is a generalized stationary point of the problem (6) and (x^*, y^*, w^*, μ^*) with $\mu^* = 0$ is a generalized stationary point of the problem (7).*

Remark. The PACC applies in particular when g and F are affine, and $\phi(a,b) = \min\{a,b\}$; see also Section 7.

We point out that the converse of the above proposition does not hold in general. This can be demonstrated by the following example, which also shows that the definition of generalized stationary points is much weaker than that of piecewise stationary points.

Example 3.9 *Consider the following MPEC problem:*

$$\begin{array}{rl} \min\limits_{x,y} & 0.5x^2 + 0.5y^2 + x - y \\ \text{subject to} & 0 \leq (y-x) \ \perp \ y \geq 0. \end{array}$$

This MPEC has a unique piecewise stationary point $(-1,0)$*. Let*

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

and $\phi(a,b) = \psi(a,b,0)$*, which is the Fischer-Burmeister functional [9], see Section 7. Clearly,* ψ *and* ϕ *satisfy the assumptions (A1)–(A4). However,* $(0,0,0)$ *is a generalized stationary point of the problem (6). Note that the feasible point* $(0,0)$ *of this MPEC is lower-level degenerate (strict complementarity fails).*

4 MPEC Constraint Qualifications

For a given feasible point $z^* = (x^*,y^*)$ of the MPEC (1), let α, β and γ be the respective index sets $\alpha(z^*)$, $\beta(x^*)$ and $\gamma(z^*)$ defined in (10). The MPEC is said to be R-regular in y at the feasible point (x^*,y^*) if the submatrix $F'_y(x^*,y^*)_{\alpha\alpha}$ of $F'_y(x^*,y^*)$ is nonsingular and the Schur-complement

$$F'_y(x^*,y^*)_{\beta\beta} - F'_y(x^*,y^*)_{\beta\alpha} F'_y(x^*,y^*)^{-1}_{\alpha\alpha} F'_y(x^*,y^*)_{\alpha\beta}$$

is a P-matrix.

Consider the constraint mapping

$$H(x,y,w,\mu) = \begin{pmatrix} g(x,y) \\ F(x,y) - w \\ \Psi(y,w,\mu) \\ e^{\mu} - 1 \end{pmatrix} \tag{13}$$

associated with (7). Let V be an element of the generalized Jacobian of this mapping at (x^*,y^*,w^*,μ^*) with $\mu^* = 0$:

$$V = \begin{pmatrix} g'_x(x^*,y^*) & g'_y(x^*,y^*) & 0 & 0 \\ F'_x(x^*,y^*) & F'_y(x^*,y^*) & -I & 0 \\ 0 & A & B & C \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where $(A,B,C) \in \partial\Psi(y^*,w^*,\mu^*)$, A and B are appropriate diagonal matrices. Then the submatrix $\bar{V}$ of V corresponding to the equilibrium constraints, i.e., the equality constraint functions $F(x,y) - w$, $\Psi(y,w,\mu)$ and $e^{\mu} - 1$, is of the following form:

$$\bar{V} = \begin{pmatrix} F'_x(x^*,y^*) & F'_y(x^*,y^*) & -I & 0 \\ 0 & A & B & C \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{14}$$

Clearly, the matrix V has full row rank if the following matrix is nonsingular:

$$U = \begin{pmatrix} F'_y(x^*, y^*) & -I \\ A & B \end{pmatrix}. \tag{15}$$

The following lemma and proposition can be proved in a standard way in the literature of nonlinear complementarity problems. For example [42, Proposition 2.1] and also [10, Theorem 9].

Lemma 4.1 *Suppose for $M, N, E \in \Re^{m\times m}$, M and N are diagonal matrices such that MN positive semidefinite and M^2+N^2 is positive definite, and E is a P-matrix. Then $M + NE$ is nonsingular.*

Proposition 4.2 *Suppose the MPEC (1) is R-regular in y at a feasible point (x^*, y^*) of the MPEC (1). Then under the assumptions (A1)–(A4), the matrix U defined in (15) is nonsingular.*

We now study generalized constraint qualifications for the problems (6) and (7) (or (2) and (3) respectively) at the feasible point (x^*, y^*, w^*) and (x^*, y^*, w^*, μ^*) under R-regularity respectively. Recall that $V \in \partial H(x^*, y^*, w^*, \mu^*)$. By Proposition 4.2, an equivalent reduction of the matrix V is the following matrix (reduction of a matrix under nonsingular transformation)

$$\begin{pmatrix} g'_x(x^*, y^*) - g'_y(x^*, y^*)(U^{-1})_{yy}F'_x(x^*, y^*) & 0 & 0 & 0 \\ F'_x(x^*, y^*) & F'_y(x^*, y^*) & -I & 0 \\ 0 & A & B & C \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where A, B, C and U are defined in (14) and (15), $(U^{-1})_{yy}$ is a submatrix of the matrix U^{-1}:

$$U^{-1} = \begin{pmatrix} (U^{-1})_{yy} & (U^{-1})_{yw} \\ (U^{-1})_{wy} & (U^{-1})_{ww} \end{pmatrix}.$$

By this observation, the following results can easily be verified.

Proposition 4.3 *Suppose (x^*, y^*) is a feasible point of the MPEC (1), the MPEC is R-regular at this point, and the assumptions (A1)–(A4) are satisfied. Let $w^* = F(x^*, y^*)$ and $\mu^* = 0$ and define $\mathcal{I}_g = \{i : \ g_i(x^*, y^*) = 0\}$ and*

$$\Gamma = g'_x(x^*, y^*) - g'_y(x^*, y^*)(U^{-1})_{yy}F'_x(x^*, y^*).$$

Then the following conclusions hold.

(i) The GCRCQ holds for (6) at (x^*, y^*, w^*) and for (7) at (x^*, y^*, w^*, μ^*) if the row submatrix $\Gamma_{\mathcal{I}_g}$ corresponding to the active indexes of g at (x^*, y^*), for any U defined in (15), has constant rank around (x^*, y^*). In particular, the GCRCQ holds for (6) at (x^*, y^*, w^*) and for (7) at (x^*, y^*, w^*, μ^*) if $g(x, y) = g(x)$ and the matrix $g'(x^*)_{\mathcal{I}_g}$ has constant rank around x^*.

(ii) The GLICQ holds for (6) at (x^*, y^*, w^*) and for (7) at (x^*, y^*, w^*, μ^*) if the row submatrix $\Gamma_{\mathcal{I}_g}$, for any U defined in (15), has full row rank. In particular, the GLICQ holds for (6) at (x^*, y^*, w^*) and for (7) at (x^*, y^*, w^*, μ^*) if $g(x, y) = g(x)$ and the matrix $g'(x^*)_{\mathcal{I}_g}$ has full row rank.

(iii) The GMFCQ holds for (6) at (x^*, y^*, w^*) and for (7) at (x^*, y^*, w^*, μ^*) if there exists a vector $d \in \Re^n$ such that for any U defined in (15)

$$\Gamma_{\mathcal{I}_g} d > 0.$$

In particular, the GMFCQ holds for (6) at (x^*, y^*, w^*) and for (7) at (x^*, y^*, w^*, μ^*) if $g(x, y) = g(x)$ and there exists a vector $d \in \Re^n$ such that

$$(g_i)'(x^*)d > 0, \text{ for } i \in \mathcal{I}_g.$$

5 Implicit Smooth SQP

5.1 Background and the Algorithm

By the assumption (A1), the nonsmooth programming problem (7) (or (3)) is smooth for any $\mu \neq 0$. This important feature allows us to use traditional nonlinear programming approaches such as sequential quadratic programming (SQP) methods for solving MPEC. To this end, we introduce a quadratic program (QP) that approximates (7). For any given (x, y, w, μ) with $\mu \neq 0$ and $d = (dx, dy, dw, d\mu)$,

$$\begin{array}{ll} \min\limits_{d \in \Re^{n+2m+1}} & \nabla f(x,y)^T \begin{pmatrix} dx \\ dy \end{pmatrix} + \frac{1}{2} d^T W d \\ \\ \text{subject to} & g'(x,y) \begin{pmatrix} dx \\ dy \end{pmatrix} + g(x,y) \geq 0 \\ \\ & F'(x,y) \begin{pmatrix} dx \\ dy \end{pmatrix} - dw + (F(x,y) - w) = 0 \\ \\ & Ady + Bdw + Cd\mu + \Psi(y, w, \mu) = 0 \\ & e^{\mu} d\mu + e^{\mu} - 1 = 0, \end{array} \tag{16}$$

where $\{(A, B, C)\} = \partial\Psi(x, y, \mu)$ which is singleton by the assumption (A1) since $\mu \neq 0$. An exact penalty merit function of (7) is defined by

$$\begin{aligned} \theta_\rho(x, y, w, \mu) \quad = \quad f(x, y) \quad + \quad \rho \Bigg[& \sum_{i=1}^{l} \max\{-g_i(x, y), 0\} \\ & + \sum_{j=1}^{m} (|F_j(x, y) - w_j| + |\psi(y_j, w_j, \mu)|) + |e^{\mu} - 1| \Bigg] \end{aligned}$$

where ρ is a positive number. If two penalty parameters are used, then we may define another penalty function

$$\begin{aligned} \Theta_{(\rho^g, \rho^{\mathrm{NCP}})}(x, y, w, \mu) \quad = \quad f(x, y) \quad & + \quad \rho^g \sum_{i=1}^{l} \max\{-g_i(x, y), 0\} \\ & + \quad \rho^{\mathrm{NCP}} \left[\sum_{j=1}^{m} (|F_j(x, y) - w_j| + |\psi(y_j, w_j, \mu)|) + |e^{\mu} - 1| \right], \end{aligned}$$

where ρ^g and $\rho^{\rm NCP}$ are two positive numbers. When $\rho^g = \rho^{\rm NCP} = \rho$, $\Theta_{(\rho^g,\rho^{\rm NCP})}$ reduces to the penalty function θ_ρ. It is easy to see that $\Theta_{(\rho^g,\rho^{\rm NCP})}$ is not differentiable in general, but directionally differentiable if ψ is directionally differentiable.

If the QP (16) has a solution d, then its KKT condition can be written as follows:

$$\begin{pmatrix} \nabla_x f \\ \nabla_y f \\ 0 \\ 0 \end{pmatrix} + Wd - \begin{pmatrix} g_x'(x,y)^T \\ g_y'(x,y)^T \\ 0 \\ 0 \end{pmatrix} \lambda_g + \begin{pmatrix} F_x'(x,y)^T \\ F_y'(x,y)^T \\ -I \\ 0 \end{pmatrix} \lambda_F + \begin{pmatrix} 0 \\ A^T \\ B^T \\ C^T \end{pmatrix} \lambda_\Psi + \begin{pmatrix} 0 \\ 0 \\ 0 \\ e^\mu \end{pmatrix} \lambda_\mu = 0$$

$$\begin{array}{l} 0 \le g'(x,y)(dx,dy) + g(x,y) \;\perp\; \lambda_g \ge 0 \\ F'(x,y)(dx,dy) - dw + (F(x,y) - w) = 0 \\ Ady + Bdw + Cd\mu + \Psi(y,w,\mu) = 0 \\ e^\mu d\mu + e^\mu - 1 = 0, \end{array} \tag{17}$$

where $(\lambda_g, \lambda_F, \lambda_\Psi, \lambda_\xi)$ is the corresponding KKT multiplier.

The existence of solutions to quadratic programs generated in traditional SQP methods play a critical role, in particular SQP fails if one of the associated quadratic programs is infeasible. In order to overcome QP infeasibility, some modifications have been introduced; see [1, 2]. Our strategy below is similar to that proposed in [1, 2] but with several notable differences. A modified quadratic program of (16) is defined as follows:

$$\begin{array}{ll} \displaystyle\min_{d\in\Re^{n+2m+1},\xi\in\Re^l} & \nabla f(x,y)^T(dx,dy) + \frac{1}{2}d^TWd + \rho\sum_{i=1}^{l}\xi_i \\ \text{subject to} & g'(x,y)(dx,dy) + g(x,y) \ge -\xi \\ & F'(x,y)(dx,dy) - dw + (F(x,y) - w) = 0 \\ & Ady + Bdw + Cd\mu + \Psi(y,w,\mu) = 0 \\ & e^\mu d\mu + e^\mu - 1 = 0 \\ & \xi \ge 0, \end{array} \tag{18}$$

where ρ is a positive penalty parameter. Note that if the constraint submatrix U given in (15) is invertible then the second, third and fourth block-rows of constraints can be solved for $(dy, dw, d\mu)$ in terms of dx. This means that by choice of ξ with sufficiently large components, the QP (18) is a feasible problem, an observation which is put to use in the next subsection to show that the modified SQP method is well defined.

If the modified QP (18) has a solution (d,ξ), then its KKT condition is a modification of the KKT condition (17):

$$\begin{pmatrix} \nabla_x f \\ \nabla_y f \\ 0 \\ 0 \end{pmatrix} + Wd - \begin{pmatrix} g_x'(x,y)^T \\ g_y'(x,y)^T \\ 0 \\ 0 \end{pmatrix} \lambda_g + \begin{pmatrix} F_x'(x,y)^T \\ F_y'(x,y)^T \\ -I \\ 0 \end{pmatrix} \lambda_F + \begin{pmatrix} 0 \\ A^T \\ B^T \\ C^T \end{pmatrix} \lambda_\Psi + \begin{pmatrix} 0 \\ 0 \\ 0 \\ e^\mu \end{pmatrix} \lambda_\mu = 0$$

$$\begin{array}{l} \rho\ \tilde{e} = \lambda_g + \lambda_\xi \\ 0 \le g'(x,y)(dx,dy) + g(x,y) + \xi \ \perp\ \lambda_g \ge 0 \\ F'(x,y)(dx,dy) - dw + (F(x,y) - w) = 0 \\ Ady + Bdw + Cd\mu + \Psi(y,w,\mu) = 0 \\ e^\mu d\mu + e^\mu - 1 = 0 \\ 0 \le \xi \ \perp\ \lambda_\xi \ge 0, \end{array} \tag{19}$$

where $\tilde{e}$ is the vector of all ones in $\Re^l$, $(\lambda_g, \lambda_F, \lambda_\Psi, \lambda_\mu, \lambda_\xi)$ is the corresponding KKT multiplier.

The inequality constraints are perturbed by introducing a vector of artificial variables $\xi \in \Re^l$. This modification improves the prospect of the feasibility of the modified QP (18). One may also relax the equality constraints in the QP (16) by introducing further artificial variables. However, because of the special structure of the MPEC, we do not change the equality constraints. As shall be seen later, the modified QP (18) is always feasible under assumptions that are considered mild in the context of nonlinear complementarity problems.

Let $u = (x, y, w, \mu)$. We propose our first modified SQP method.

Algorithm: `Implicit Smooth SQP`

Step 0. (**Initialization**) Let $\rho_{-1} > 0$, $\delta_1 > 0$, $\delta_2 > 0$, $\sigma \in (0,1)$, $\tau \in (0,1)$. Choose $(x^0, y^0, w^0, \mu^0) \in \Re^{n+2m+1}$ such that $\mu^0 > 0$, and a symmetric positive definite matrix W_0 in $\Re^{(n+2m+1)\times(n+2m+1)}$. Set $k := 0$.

Step 1. (**Search direction**) Solve the modified QP (18) with $(x,y,w,\mu) = (x^k, y^k, w^k, \mu^k)$, $W = W_k$, and $\rho = \rho_{k-1}$. Let (d^k, ξ^k) be a solution of this modified QP and $\lambda^k = (\lambda_g^k, \lambda_F^k, \lambda_\Psi^k, \lambda_\mu^k, \lambda_\xi^k)$ be its corresponding multiplier.

Step 2. (**Termination check**) If a stopping rule is satisfied, terminate. Otherwise, go to Step 3.

Step 3. (**Penalty update**) Let

$$\tilde{\rho}_k = \begin{cases} \rho_{k-1} & \text{if } \rho_{k-1} \ge \max\limits_{1\le i\le l+2m+1} |\lambda_i^k| \\ \delta_1 + \max\limits_{1\le i\le l+2m+1} |\lambda_i^k| & \text{otherwise.} \end{cases}$$

Define $\rho_k^g = \rho_{k-1}$, $\rho_k^{\mathrm{NCP}} = \tilde{\rho}_k$, and

$$\rho_k \ = \begin{cases} \tilde{\rho}_k & \text{if } \sum\limits_{1\le i\le l} \xi_i^k = 0 \\ \tilde{\rho}_k + \delta_2 & \text{otherwise.} \end{cases}$$

Step 4. (**Line search**) Let $t_k = (\tau)^{i_k}$ where i_k is the smallest nonnegative integer such that $i = i_k$ satisfies

$$\Theta_{(\rho_k^g,\rho_k^{\rm NCP})}(u^k + (\tau)^i d^k) - \Theta_{(\rho_k^g,\rho_k^{\rm NCP})}(u^k) \le -\sigma(\tau)^i (d^k)^T W_k d^k.$$

Step 5. (**Update**) Let $u^{k+1} = u^k + t_k d^k$. Choose a symmetric positive definite matrix $W_{k+1} \in \Re^{(n+2m+1)\times(n+2m+1)}$. Set $k := k+1$. Go to Step 1.

Remarks.

(i) If the modified QP (18) is replaced by the QP (16) to generate the search direction in the above algorithm, then our SQP method is very similar to classical SQP methods for smooth nonlinear programming [15, 36]. The difference is that here we anticipate nonsmoothness of ψ. If further μ is treated as a parameter rather than a variable, namely the last equation in (16) is omitted at each iteration, then the above modified SQP method begins to look like the explicit smoothing SQP method proposed in Fukushima, Luo and Pang [12]. See Section 6 for an explicit SQP method that has the convergence properties of the explicit smoothing method of Facchinei, Jiang and Qi [8].

(ii) Since only inequality constraints are relaxed, we use the merit function $\Theta_{(\rho^g,\rho^{\rm NCP})}$ which has two (likely different) penalty parameters, unlike θ_ρ used in the classical SQP methods. The updates for $\tilde{\rho}$, ρ^g and $\rho^{\rm NCP}$ are to ensure that the solution of the modified QP (18) is a descent direction of the merit function $\Theta_{(\rho^g,\rho^{\rm NCP})}$. In the update of ρ, we increase it by a positive constant δ_2 in the case that $\sum \xi_i > 0$ in an attempt to force a decrease in the feasibility gap of the QP (16) at the next iteration.

5.2 QP Subproblems and the Penalty Function

Definition 5.1 *F is said to be a P_0-function with respect to y if for each $x \in \Re^n$, $F(x,\cdot)$ is a P_0-function; i.e., for any $y, \bar{y} \in \Re^m$ with $y \neq \bar{y}$, there exists an index i such that $y_i \neq \bar{y}_i$ and*

$$(y_i - \bar{y}_i)(F_i(x,y) - F_i(x,\bar{y}) \ge 0.$$

We introduce a new condition on ψ to extend invertibility of the matrix U in (15). to infeasible points.

(A5) For $c \neq 0$, if $(p,q,r) \in \partial\psi(a,b,c)$, then $pq > 0$.

Proposition 5.2 *Suppose F is a P_0-function with respect to y. If the assumptions (A1)–(A5) hold, then the matrix given by (15),*

$$U = \begin{pmatrix} F'_y(x,y) & -I \\ A & B \end{pmatrix},$$

is nonsingular for any (x,y,w,μ) with $\mu \neq 0$, where $(A,B,C) \in \partial\Psi(y,w,\mu) = \{\Psi'(y,w,\mu)\}$.

Proof. Since $(A, B, C) \in \partial\Psi(y, w, \mu)$, both A and B are diagonal matrices with nonzero diagonal elements. It turns out that nonsingularity of the matrix U is equivalent to nonsingularity of the matrix $A + BF'_y(x, y)$, or $B^{-1}A + F'_y(x, y)$. Note that $B^{-1}A$ is a diagonal positive definite matrix, and $F'_y(x, y)$ is a P_0-matrix. Therefore, nonsingularity of $B^{-1}A + F'_y(x, y)$ follows; see [6]. This completes the proof. ∎

The following result concerns the feasibility of the quadratic programs (16) and (18).

Proposition 5.3 *Suppose $F'_y(x, y)$ is a P_0-matrix, the assumptions (A1)–(A5) hold and $\mu \neq 0$. Let U be defined in Proposition 5.2. Then*

(i) The modified QP (18) is always feasible.

(ii) The QP (16) has a nonempty feasible set if and only if the following system is consistent with respect to dx:

$$[g'_x(x, y) - g'_y(x, y)(U^{-1})_{yy}F'_x(x, y)]dx$$

$$-g'_y(x, y)[(U^{-1})_{yy}(F(x, y) - w) + (U^{-1})_{yw}(\Psi(y, w, \mu) + Cd\mu)] + g(x, y) \geq 0.$$

(iii) If d is a solution of (16) or (d, ξ) is a solution of (18), then

$$d\mu = -\frac{e^{\mu} - 1}{e^{\mu}}$$

and (dy, dw) is uniquely determined by dx and $d\mu$, i.e.,

$$(dy, dw) = U^{-1} \begin{pmatrix} -F'_x(x, y)dx - F(x, y) + w \\ -\Psi(y, w, \mu) - Cd\mu \end{pmatrix}.$$

The above proposition gives not only a characterization for nonemptiness of the feasible set of (16), but also shows that a solution of (16) can be found by solving a reduced QP in the variable x-space, and a system of linear equation (the similar argument also applies to the modified QP (18)). This fact can be computationally significant as n is often much smaller than m.

The feasibility of the QP (16) is a serious issue in the context of MPEC. Fukushima and Pang [13] discussed it from a different angle, namely for mathematical programs with linear complementarity constraints. We remark that the P_0 property assumed in our paper is not necessarily required in [13].

The following is a simple yet important consequence of the above proposition for the case when there are no joint (upper-level) constraints on (x, y).

Corollary 5.4 *Suppose $F'_y(x, y)$ is a P_0-matrix, (A1)–(A5) hold and $\mu \neq 0$. Assume that $g(x, y) = g(x)$. Then (16) has a nonempty feasible set if and only if $g'(x)dx + g(x) \geq 0$ is consistent with respect to dx.*

The next result on $d\mu$ is proved in [18]. It basically says that $\{\mu^k\}$ is a positive and monotonically decreasing sequence.

Lemma 5.5 *Suppose* $(x, y, w, \mu) \in \Re^{n+2m+1}$, $\mu > 0$, *and* $d = (dx, dy, dw, d\mu)$ *solves the QP (16) or* (d, ξ) *solves the modified QP (18). Then* $d\mu \in (-\mu, 0)$ *so that*

$$\mu + td\mu \in (0, \mu)$$

for any $t \in (0, 1]$.

We now study some properties of the exact penalty function $\Theta_{(\rho^g,\rho^{\mathrm{NCP}})}$. The vector $u^* = (x^*, y^*, w^*, \mu^*)$ is said to be a critical point of (or stationary for) the penalty function $\Theta_{(\rho^g,\rho^{\mathrm{NCP}})}$ for the given positive parameters ρ^g and ρ^{NCP} if for any direction $d \in \Re^{n+2m+1}$,

$$\Theta'_{(\rho^g,\rho^{\mathrm{NCP}})}(u^*; d) \geq 0.$$

In the rest of this subsection, we collect some useful properties which shall be used for proving global convergence of the (modified) implicit smooth SQP in the next subsection. Since all functions are smooth when $\mu \neq 0$, these properties follow directly from the nonlinear programming results presented in the Appendix of the manuscript of this paper [20]. Moreover, we are mainly concerned with the properties associated with the modified QP.

Proposition 5.6 *Let* $\mu \neq 0$.

(i) For $d \in \Re^{n+2m+1}$, $\Theta'_{(\rho^g,\rho^{\mathrm{NCP}})}$ *is directionally differentiable at* u *along the direction* d *and* $\Theta'_{(\rho^g,\rho^{\mathrm{NCP}})}(x, y, w, \mu; d)$ *can be easily evaluated.*

(ii) If (d, ξ) *is a solution of the modified QP (18),* $\rho^g = \rho$ *and* $\rho^{\mathrm{NCP}} \geq \max_{1\leq i\leq l+2m+1} |\lambda_i|$ *with* λ *the KKT multiplier of the modified QP (18), then*

$$\begin{aligned}\Theta'_{(\rho^g,\rho^{\mathrm{NCP}})}(x, y, w, \mu; d) \leq\ & \nabla f(x, y)^T(dx, dy) - (\lambda_g)^T g'(x, y)(dx, dy) \\ & +(\lambda_F)^T(F'(x, y)(dx, dy) - dw) \\ & +(\lambda_\Psi)^T\Psi'(y, w, \mu)(dy, dw, d\mu) + \lambda_\mu e^\mu d\mu\end{aligned}$$

and

$$\Theta'_{(\rho^g,\rho^{\mathrm{NCP}})}(x, y, w, \mu; d) \leq -d^T W d.$$

Proof. When $\mu \neq 0$, $g(x, , y)$, $F(x, y) - w$, $\Psi(y, w, \mu)$, and $e^\mu - 1$ are all continuously differentiable at (x, y, w, μ). Hence the results follow from [20, Proposition A.1] and Lemma 5.5. ■

Proposition 5.7 *Let* $u^* = (x^*, y^*, w^*, \mu^*)$ *be given. Suppose the matrix* W^* *is symmetric positive definite,* $F'_y(x^*, y^*)$ *is a* P_0 *matrix and* Ψ *is smooth at* (y^*, w^*, μ^*).

(i) For given $\rho^g > 0$ *and all large* $\rho^{\mathrm{NCP}} > 0$, u^* *is a critical point of* $\Theta_{(\rho^g,\rho\mathrm{NCP})}$ *if and only if there exists* (d, ξ) *with* $d = 0$ *that is a solution of the modified QP (18) with* $u = u^*$, $W = W^*$ *and* $\rho = \rho^g$.

(ii) If u^* *is a KKT point of (7) and* λ *is its KKT multiplier, then* u^* *is a critical point of* $\Theta_{(\rho^g,\rho^{\mathrm{NCP}})}$ *with* $\min\{\rho^g, \rho^{\mathrm{NCP}}) \geq \max_{1\leq i\leq l+2m+1} |\lambda_i|$.

(iii) If u^* *is a critical point of* $\Theta_{(\rho^g,\rho^{\mathrm{NCP}})}$ *for some* $\rho^g > 0$ *and all sufficiently large* $\rho^{\mathrm{NCP}} > 0$, *and* u^* *is feasible for (7), then* u^* *is a KKT point of (7).*

Proof. The desired results can be proved from Propositions 5.3, Propositions A.3 and A.5 of [20]. ■

5.3 Global Convergence under Lower-Level Nondegeneracy

Lemma 5.8 *Suppose* $(x, y, w, \mu) \in \Re^{n+2m+1}$ *with* $\mu \neq 0$*, and* $W \in \Re^{(n+2m+1)\times(n+2m+1)}$ *is symmetric positive definite. Suppose* (d, ξ) *is a solution of the modified QP (18) and* λ *is its corresponding KKT multiplier. Then* d *is a descent direction of the merit function* $\Theta_{(\rho^g,\rho^{\rm NCP})}$ *if* $d \neq 0$*,* $\rho^g = \rho$ *and* $\rho^{\rm NCP} \geq \max\limits_{1\leq i\leq l+2m+1} |\lambda_i|$.

Proof. The lemma follows from the second inequality of (ii) in Proposition 5.6. ∎

Lemma 5.8 shows that solving the modified QP (18) generates a descent direction of the merit function $\Theta_{(\rho^g,\rho^{\rm NCP})}$ for sufficiently large $\rho^{\rm NCP}$ when W is symmetric positive definite and $\mu \neq 0$. Furthermore the line search in Step 4 is well-defined, i.e., t_k can be determined in a finitely many steps. Therefore, the implicit smooth SQP method is well-defined when $\mu \neq 0$ and W is symmetric positive definite at each step. Moreover, since the line search chooses the step size $t_k \in (0, 1)$, Lemma 5.5 can be used to show that $\mu^{k+1} > 0$ if $\mu^k > 0$; hence $\mu^k \neq 0$ for each k if $\mu^0 > 0$.

To present the global convergence of implicit smooth SQP, we assume the following standard conditions:

(B1) There exist two positive numbers $\alpha < \beta$ such that each of the symmetric matrices W_k used in implicit smooth SQP satisfies the following condition for all vectors u of appropriate dimension:

$$\alpha\|u\|^2 \leq u^T W_k u \leq \beta\|u\|^2.$$

(B2) For all large k, $\rho_k = \rho_* > 0$.

Under the assumption (B1) and the feasibility of the modified QP (18) at each iteration, implicit smooth SQP is well-defined. The assumption (B2) can be shown to hold under some further conditions. As a consequence of the condition (B2), we obtain that for all sufficiently large k,

$$\rho_k^g = \rho_*,\ \rho_k^{\rm NCP} = \rho_*^{\rm NCP},\ \xi^k = 0,$$

where $\rho_*^{\rm NCP}$ is a positive constant. We assume that implicit smooth SQP does not terminate in Step 2. Let $\{u^k\} = \{(x^k, y^k, w^k, \mu^k)\}$ be generated by implicit smooth SQP.

Lemma 5.9 *Suppose (A1)–(A5) hold, (B1)–(B2) hold, and* F *is a* P_0*-function with respect to* y*. Suppose* $\{u^k\}$ *is the sequence generated by the algorithm,* $\{(d^k, \xi^k)\}$ *is the sequence of solutions of the modified QP (18), and* $\lim_{k\to\infty, k\in K} u^k = u^*$ *for a subset* $K \subseteq \{1, 2, \ldots, \}$*. Then the following conclusions hold.*

(i) $\{d^k\}_{k\in K}$ *and* $\{\xi^k\}_{k\in K}$ *are bounded.*

(ii) Assume Ψ *is continuously differentiable near* u^**. If* d^**,* ξ^* *and* W^* *are accumulation points of the sequences* $\{d^k\}_{k\in K}$*,* $\{\xi^k\}_{k\in K}$ *and* $\{W_k\}_{k\in K}$ *respectively, then* (d^*, ξ^*) *is a solution of the modified QP (18) with* $u = u^*$*,* $W = W^*$ *and* $\rho = \rho_*$*. Furthermore,* $\Theta_{(\rho_*,\rho_*^{\rm NCP})}$ *is directionally differentiable at* u^* *and it holds that*

$$\Theta'_{(\rho_*,\rho_*^{\rm NCP})}(u^*; d^*) \leq -(d^*)^T W^* d^*.$$

(iii) Assume the step size sequence $\{t_k\}$ determined by the Armijo line search satisfies $\lim_{k\to\infty,k\in K} t_k = 0$. Under the smoothness assumption in (ii), we have

$$\limsup_{k\to\infty,k\in K} \frac{\Theta_{(\rho_*,\rho_*^{\mathrm{NCP}})}(u^k + t_k d^k) - \Theta_{(\rho_*,\rho_*^{\mathrm{NCP}})}(u^k)}{t_k} \le \Theta'_{(\rho_*,\rho_*^{\mathrm{NCP}})}(u^*; d^*).$$

Proof. Since $g(x,y)$, $F(x,y) - w$, $\Psi(y,w,\mu)$ and $e^\mu - 1$ are smooth at u^*, The desired results follows from Lemmas A.2 and A.3 of [20]. ■

The condition (B2) may not hold in general. The following additional conditions ensure that (B2) is satisfied, as shown below. Let H be the function representing the equality constraints of (7), i.e. $H(u) = (F(x,y) - w,\ \Psi(y,w,\mu))$ with $u = (x,y,w,\mu)$.

(B3) $\{u^k\}$ is bounded.

(B4) The generalized Jacobian $\partial H(u^*)$ has full row rank at any accumulation point u^* of $\{u^k\}$.

(B5) For any accumulation point u^* of $\{u^k\}$ and any $V \in \partial H(u^*)$, there exists $d = (dx, dy, dw, d\mu)$ such that $g'(x^*,y^*)(dx,dy) + g(x^*,y^*) > 0$ and $Vd + H(u^*) = 0$.

Note that the conditions (B4) and (B5) together are equivalent to the GMFCQ if u^* is a feasible point of (7).

We are now ready to establish global convergence of implicit smooth SQP under the assumption that Ψ is smooth at the accumulation point. We remark that the smoothness of Ψ at a limit point means that the problem has essentially (asymptotically) been reduced to smooth nonlinear programming.

Theorem 5.10 *Suppose the assumptions (A1)–(A5) hold, the standing assumptions (B1) holds, and F is a P_0-function with respect to y. Suppose $\mu^0 > 0$, $\{u^k\}$ is the sequence generated by the algorithm. We obtain the following conclusions.*

(i) If (a) the condition (B2) holds, (b) u^* is an accumulation point of $\{u^k\}$, (c) Ψ is continuously differentiable at u^*, then u^* is both a critical point of $\Theta_{(\rho_*,\rho_*^{\mathrm{NCP}})}$ and a (classical or primal or piecewise) stationary point of the MPEC (7).

(ii) If conditions (B3), (B4) and (B5) hold, then so does (B2).

Proof. (i) Since Ψ is smooth at u^*, all results follow from Theorems A.1 of [20] and Proposition 3.7.

(ii) This follows from Theorem A.2 in [20] but with some suitable modifications given that Ψ may be nonsmooth at some accumulation points of the sequence $\{u^k\}$. ■

Remark. In the above theorem, global convergence of implicit smooth SQP requires smoothness of the function Ψ at $u^* = (x^*, y^*, w^*, \mu^*)$. As shall be seen in Section 7, if ϕ is the Fischer-Burmeister function in Example 7.1, the min function in Example 7.2, or the Kanzow-Kleinmichel function in Example 7.3, then the smoothness condition on Ψ is satisfied at any lower-level nondegenerate point, i.e., Ψ is smooth, in fact twice continuously differentiable.

As already noted, lower-level nondegeneracy at a limit point u^* often results in smoothness of the function Ψ at this point, which means that we can apply classical theory and obtain classical results. Hence superlinear convergence under the lower-level nondegeneracy condition and the assumption that the stepsize takes the value 1 for all large k would be no surprise though our merit function $\Theta_{(\rho^g,\rho^{\mathrm{NCP}})}$ has two penalty parameters. The unit stepsize assumption is needed in nonlinear programming due to the well-known Maratos effect, which can prevent superlinear convergence of an SQP method that uses an exact penalty merit function unless a second-order correction to the feasibility of the iterate is performed at each iteration. See [11, 36].

In order to study the rate of convergence of implicit smooth SQP, further conditions such as the LICQ, the second order sufficient condition, careful update rules of the matrix sequence $\{W_k\}$, etc. are needed. We conjecture that superlinear convergence results similar to those of [35, 36] can be obtained.

6 Explicit Smooth SQP

Global convergence of the implicit smooth SQP method requires the lower-level nondegeneracy condition at an accumulation point. This assumption is not unusual for convergence of MPEC algorithms such as PIPA [28] and also the explicit smoothing SQP method of Fukushima, Luo and Pang [12], but is still rather strong in that it essentially reduces the problem to one of nonlinear programming, which is not tenable in general.

As an alternative we propose an explicit smooth SQP algorithm for which global convergence can be established without assuming lower-level nondegeneracy. This method has a similar computational form to the SQP method of [12], although our smoothing parameter update has to be carried out more carefully, like the original smoothing method for MPEC of Facchinei, Jiang and Qi [8]. Moreover the method given here weakens the assumptions needed in [8] as explained in Remark (ii) following Theorem 6.4.

Note that the term explicit refers to the fact that the smoothing parameter μ is not treated as a variable in the QP subproblem at each iteration, nor is it updated in the line search which determines the next iterate $(x^{k+1}, y^{k+1}, w^{k+1})$. In our explicit smooth SQP method, the smoothing parameter tends to be updated less often than once per QP-solve, unlike the the implicit smooth SQP method of the last section and the explicit smoothing algorithm of [12].

Recall definitions of ϕ_μ and Φ_μ in Section 3. We approximate the MPEC (1) by the following nonlinear programming problem with $\mu \neq 0$:

$$\begin{array}{rl} \min\limits_{x,y,w} & f(x,y) \\ \text{subject to} & g(x,y) \geq 0 \\ & F(x,y) - w = 0 \\ & \Phi_\mu(y,w) = 0 \end{array} \tag{20}$$

which is (4) with

$$\Phi_\mu(y,w) \quad = \quad \begin{pmatrix} \phi_\mu(y_1,w_1) \\ \vdots \\ \phi_\mu(y_m,w_m) \end{pmatrix} \quad = \quad \Psi(y,w,\mu).$$

Obviously, when $\mu = 0$, the above problem reduces to (6). Therefore, our goal is to find approximate solutions of (6) for each $\mu \neq 0$ and then locate a solution or a generalized stationary point of (6) by driving μ to zero.

Similar to implicit smooth SQP, we want to find an approximate solution of (20) by solving a sequence of quadratic programs. More precisely, for any given (x, y, w), $\mu \neq 0$ and $d = (dx, dy, dw)$, we define a modified quadratic program (which is a modified quadratic model of (20) at (x, y) for the fixed $\mu \neq 0$ and $\rho > 0$) as follows:

$$\begin{array}{rl} \displaystyle\min_{d \in \Re^{n+2m}, \xi \in \Re^l} & \nabla f(x,y)^T(dx,dy) + \frac{1}{2} d^T W d + \rho \displaystyle\sum_{i=1}^{l} \xi_i \\ \\ \text{subject to} & g'(x,y)(dx,dy) + g(x,y) \geq -\xi \\ & F'(x,y)(dx,dy) - dw + (F(x,y) - w) = 0 \\ & A dy + B dw + \Phi_\mu(y,w) = 0 \\ & \xi \geq 0, \end{array} \tag{21}$$

where $[A \ B] = \Phi'_\mu(x, y)$ and the matrix $W \in \Re^{(n+2m)\times(n+2m)}$ is symmetric positive definite.

If the modified QP (21) has a solution (d, ξ), then its KKT condition has the following form:

$$\begin{pmatrix} \nabla_x f \\ \nabla_y f \\ 0 \end{pmatrix} + Wd - \begin{pmatrix} g'_x(x,y)^T \\ g'_y(x,y)^T \\ 0 \end{pmatrix} \lambda_g + \begin{pmatrix} F'_x(x,y)^T \\ F'_y(x,y)^T \\ -I \end{pmatrix} \lambda_F + \begin{pmatrix} 0 \\ A^T \\ B^T \end{pmatrix} \lambda_{\Phi_\mu} = 0$$

$$\begin{array}{l} \rho\, \tilde{e} = \lambda_g + \lambda_\xi \\ 0 \leq g'(x,y)(dx,dy) + g(x,y) + \xi \ \perp \ \lambda_g \geq 0 \\ F'(x,y)(dx,dy) - dw + (F(x,y) - w) = 0 \\ A dy + B dw + \Phi_\mu(y,w) = 0 \\ 0 \leq \xi \ \perp \ \lambda_\xi \geq 0, \end{array} \tag{22}$$

where $\tilde{e}$ is the vector of all ones in $\Re^l$.

We can immediately write down a quadratic model of (20), that is without the artificial variable ξ:

$$\begin{array}{rl} \displaystyle\min_{d \in \Re^{n+2m}, \xi \in \Re^l} & \nabla f(x,y)^T(dx,dy) + \frac{1}{2} d^T W d \\ \text{subject to} & g'(x,y)(dx,dy) + g(x,y) \geq 0 \\ & F'(x,y)(dx,dy) - dw + (F(x,y) - w) = 0 \\ & A dy + B dw + \Phi_\mu(y,w) = 0. \end{array} \tag{23}$$

A penalty merit function of (20) is defined by

$$\Theta_{(\rho^g, \rho^{\mathrm{NCP}}, \mu)}(x,y,w) \ = \ f(x,y) \ + \ \rho^g \sum_{i=1}^{l} \max\{-g_i(x,y), 0\} + \rho^{\mathrm{NCP}} \sum_{j=1}^{m} [(|F_j(x,y) - w_j| + |\phi_\mu(y_j, w_j)|)]$$

where ρ^g and ρ^{NCP} are positive numbers.

Before presenting our second method, we give a result for the case $\mu \neq 0$, when the problem (20) is a smooth NLP, that is parallel to Proposition 4.3 which deals with the case $\mu = 0$. This result will not be used directly in the proof of convergence of the explicit smooth SQP method but gives some idea of when the constraints of the nonlinear problem (20) are numerically stable.

Proposition 6.1 *Suppose (x, y, w) is a feasible point of (20) with $\mu \neq 0$, F is a P_0-function with respect to y, and the assumptions (A1)–(A2) and (A5) are satisfied. Define $\mathcal{I}_g = \{i : \ g_i(x, y) = 0\}$,*

$$\begin{aligned}
\Gamma &= g'_x(x, y) - g'_y(x, y)(U^{-1})_{yy}F'_x(x, y) \\
U &= \begin{pmatrix} F'_y(x, y) & -I \\ A & B \end{pmatrix}, \quad [A\ B] = \Phi'_\mu(y, w).
\end{aligned}$$

Then the following conclusions hold.

(i) The CRCQ holds for (20) at (x, y, w) if the row submatrix $\Gamma_{\mathcal{I}_g}$ corresponding to the active indexes of g at (x, y) has a constant rank around (x, y, w), In particular, the CRCQ holds for (20) at (x, y, w) if $g(x, y) = g(x)$ and if the matrix $g'(x)_{\mathcal{I}_g}$ has constant rank around x.

(ii) The LICQ holds for (20) at (x, y, w) if the row submatrix $\Gamma_{\mathcal{I}_g}$ has full row rank. In particular, the LICQ holds for (20) at (x, y, w) if $g(x, y) = g(x)$ and if the row submatrix $g'(x)_{\mathcal{I}_g}$ has full row rank.

(iii) The MFCQ holds for (20) at (x, y, w) if there exists a vector $d \in \Re^n$ such that

$$\Gamma_{\mathcal{I}_g} d > 0.$$

In particular, the MFCQ holds for (20) at (x, y, w) if $g(x, y) = g(x)$ and there exists a vector $d \in \Re^n$ such that

$$g'_i(x)d > 0, \text{ for } i \in \mathcal{I}_g.$$

Unlike in Section 3, in this section we let $u = (x, y, w)$ and $d = (dx, dy, dw)$ because μ is regarded as a parameter, but not a variable. For the same reason, we use the subscript k, i.e. μ_k, to denote the value of the parameter μ at the kth iteration.

Algorithm: `Explicit Smooth SQP`

Step 0. (**Initialization**) Let $\rho_{-1} > 0$, $\delta_1 > 0$, $\delta_2 > 0$, $\beta_\mu \in (0, 1)$, $\beta_\varepsilon \in (0, 1)$, $\sigma \in (0, 1)$, $\tau \in (0, 1)$. Choose $u^0 = (x^0, y^0, w^0) \in \Re^{n+2m}$, and choose $\mu_0 > 0$, $\varepsilon_0 > 0$, and a symmetric positive definite matrix $W_0 \in \Re^{(n+2m)\times(n+2m)}$. Set $k := 0$.

Step 1. (**Search direction**) Solve the modified QP (21) with $(x, y, w) = (x^k, y^k, w^k)$, $\mu = \mu_k$ and $W = W_k$ and $\rho = \rho_{k-1}$. Let (d^k, ξ^k) be a solution of this QP and $\lambda^k = (\lambda_g, \lambda_F, \lambda_\Phi, \lambda_\xi)$ be its corresponding KKT multiplier.

Step 2. (**Termination check**) If a stopping rule is satisfied, terminate. Otherwise, go to Step 3.

Step 3. (**Penalty update**) Let

$$\tilde{\rho}_k = \begin{cases} \rho_{k-1} & \text{if } \rho_{k-1} \geq \max\limits_{1 \leq i \leq l+2m+1} |\lambda_i^k| \\ \delta_1 + \max\limits_{1 \leq i \leq l+2m+1} |\lambda_i^k| & \text{otherwise.} \end{cases}$$

Define $\rho_k^g = \rho_{k-1}$ and $\rho_k^{\mathrm{NCP}} = \tilde{\rho}_k$ and

$$\rho_k = \begin{cases} \tilde{\rho}_k & \text{if } \sum\limits_{1 \leq i \leq l} \xi_i^k = 0 \\ \tilde{\rho}_k + \delta_2 & \text{otherwise.} \end{cases}$$

Step 4. (**Line search**) Let $t_k = (\tau)^{i_k}$ where i_k is the smallest nonnegative integer such that $i = i_k$ satisfies

$$\Theta_{(\rho_k^g, \rho_k^{\mathrm{NCP}}, \mu_k)}(u^k + (\tau)^i d^k) - \Theta_{(\rho_k^g, \rho_k^{\mathrm{NCP}}, \mu_k)}(u^k) \leq -\sigma(\tau)^i (d^k)^T W_k d^k.$$

Step 5. (**Update**) Let

$$\begin{aligned} u^{k+1} &= u^k + t_k d^k \\ \mu_{k+1} &= \begin{cases} \beta_\mu \, \mu_k & \text{if } \|d^k\| \leq \varepsilon_k \\ \mu_k & \text{otherwise} \end{cases} \\ \varepsilon_{k+1} &= \begin{cases} \beta_\varepsilon \, \varepsilon_k & \text{if } \|d^k\| \leq \varepsilon_k \\ \varepsilon_k & \text{otherwise.} \end{cases} \end{aligned}$$

Choose a symmetric positive definite matrix $W_{k+1} \in \Re^{(n+2m)\times(n+2m)}$. Set $k := k+1$, and go to Step 1.

We next present some results analogous to those in Subsection 5.2. The proofs are very similar, so we omit all proofs.

Proposition 6.2 *Suppose $F'_y(x, y)$ is a P_0-matrix, the assumptions (A1)–(A5) hold and $\mu \neq 0$. Then (21) has a nonempty feasible set. Moreover (23) has a nonempty feasible set if and only if the following system is consistent with respect to dx:*

$$\Gamma dx + g(x, y, w, \mu) \ \geq \ 0$$

where Γ and U are given by Proposition 6.1 and $g(x, y, w, \mu)$ is the vector

$$g(x, y, w, \mu) \ = \ g(x, y) \ - \ g'_y(x, y)[(U^{-1})_{yy}(F(x, y) - w) + (U^{-1})_{yw}\Phi_\mu(y, w)].$$

Furthermore, (dy, dw) is uniquely determined by dx, i.e.,

$$(dy, dw) = U^{-1} \begin{pmatrix} -F'_x(x, y)dx - F(x, y) + w \\ -\Phi_\mu(y, w) \end{pmatrix}.$$

In the case where $g(x, y) = g(x)$, the above consistency condition becomes consistency with respect to dx:

$$g'(x)dx + g(x) \geq 0.$$

Proposition 6.3 *Let $\mu \neq 0$.*

(i) $\Theta'_{(\rho^g,\rho^{\mathrm{NCP}},\mu)}$ *is directionally differentiable at u. Furthermore, if (d,ξ) is a solution of the modified QP (21), $\rho^g = \rho$ and $\rho^{\mathrm{NCP}} \geq \max_{1\leq i\leq l+2m} |\lambda_i|$ with λ its KKT multiplier, then*

$$\begin{aligned}\Theta'_{(\rho^g,\rho^{\mathrm{NCP}},\mu)}(x,y,w;d) \;\leq\; & \nabla f(x,y)^T(dx,dy) - (\lambda_g)^T g'(x,y)(dx,dy)\\ & +(\lambda_F)^T(F'(x,y)(dx,dy) - dw)\\ & +(\lambda_{\Phi_\mu})^T \Phi'_\mu(y,w)(dy,dw)\end{aligned}$$

and

$$\Theta'_{(\rho^g,\rho^{\mathrm{NCP}},\mu)}(x,y,w;d) \leq -d^T W d.$$

(ii) Suppose W is symmetric positive definite. If (d,ξ) is a solution of the modified QP (21) with $d \neq 0$, then d is a descent direction of the penalty function $\Theta_{(\rho^g,\rho^{\mathrm{NCP}},\mu)}$ for $\rho^g = \rho$ and any ρ^{NCP} satisfying the condition in (i).

We need assumptions (B1) and (B2) as Subsection 5.3, although here $u = (x,y,w)$, i.e. μ is omitted, hence the order of each matrix W_k is $n+2m$ rather than $n+2m+1$ as in Section 5. As before, we can ensure (B2) by assuming conditions (B3)–(B5). The function H in the conditions (B4) and (B5) now corresponds to the equality constraints of (6), that is $H(u) = (F(x,y) - w, \Phi(y,w)) = (F(x,y) - w, \Psi(y,w,0))$.

Theorem 6.4 *Assume the assumptions (A1)–(A5) hold, the assumption (B1) holds, and F is a P_0-function with respect to y. Let $\mu_0 > 0$, $\{u^k\}$, $\{\mu_k\}$ and $\{\varepsilon_k\}$ be the sequences generated by the algorithm.*

(i) If the assumption (B2) holds and $\{u^k\}$ has a limit point, then

$$\lim_{k\to\infty} \mu_k = 0, \lim_{k\to\infty} \varepsilon_k = 0.$$

(ii) Let $K = \{k : \|d^k\| \leq \varepsilon_k\}$. If we assume that the assumption (B2) holds and $\{u^k\}_{k\in K}$ has an accumulation point $u^* = (x^*,y^*,w^*)$, then u^* is a generalized stationary point of (6). Furthermore, if (x^*,y^*) is lower-level nondegenerate, then (x^*,y^*) is a (classical or primal or piecewise) stationary point of the MPEC

(iii) If conditions (B1), (B3), (B4) and (B5) hold, then so does (B2).

Proof. (i) Obviously $\{\mu_k\}$ is bounded. Suppose μ_* is an accumulation point of $\{\mu_k\}$. If $\mu_* > 0$, then $\|d^k\| \leq \varepsilon_k$ occurs only finitely many times. This means that after finitely many iterations, μ_k and ε_k remain unchanged, i.e., for some k_0 and all $k \geq k_0$, $\mu_k = \mu_{k_0} > 0$ and $\varepsilon_k = \varepsilon_{k_0} > 0$. In this case, our smoothing method reduces to the modified SQP method presented in [20, Appendix] for a smooth nonlinear program (20). By [20, Theorem A.1] and its proof, it follows that some subsequence of $\{d^k\}$ approaches to 0 as $k \to \infty$, which implies that $\|d^k\| \leq \varepsilon_{k_0}$ will eventually happen, a contradiction. Therefore, $\lim_{k\to\infty} \mu_k = 0$. By the update rule in Step 5, it is also true that $\lim_{k\to\infty} \varepsilon_k = 0$.

(ii) By the assumption (B2) and the update rule of the penalty parameter, the KKT multiplier sequence $\{\lambda^k\}_{k\in K}$ is bounded and $\xi^k = 0$ for all large enough k since

$\rho_k = \rho_*$ for all sufficiently large k. Note that for each $k \in K$, $\|d^k\| \leq \varepsilon_k$. Hence $\lim_{k\to\infty, k\in K} d^k = 0$. By passing to the limit for $k \in K$, it follows from the KKT condition (22) and the assumption (A3) that u^* is a generalized stationary point of (6). From Proposition 3.7, (x^*, y^*) is a piecewise stationary point of the MPEC if (x^*, y^*) is lower-level nondegenerate.

(iii) This follows, as does part (ii) of Theorem 5.10, by a straightforward extension of a similar result for the smooth case [20, Theorem A.2]. ∎

Remarks.

(i) As discussed in Subsection 5.5, we may find a solution of the modified QP (21) by solving a reduced QP and some systems of linear equations, which may reduce the computational cost significantly especially if the matrices defining the QP are dense.

(ii) Loosely speaking, the first part of Theorem 6.1 under the assumption (iii) can be viewed as a generalization of Theorem 5.14(b) of [8] when the assumptions (A1)-(A5) in [8] are valid, and the smoothing function ψ used in (20) has the form of that defined in Example 7.2 below. To explain further, in [8]: (a) The upper-level constraints have the form $g(x, y) \equiv g(x) \leq 0$, i.e. the MPEC is an implicit program. (b) The upper-level and lower-level feasible sets are assumed to be compact, while in our case compactness is not assumed in either the upper or lower levels (the lower-level feasible set is $\Re^m_+$, corresponding to an NCP). (c) The lower-level objective function F is assumed to be uniformly strongly monotone in [8]; we assume that F is at most a uniform P_0-function in y.

Regarding (b), we should say that nonlinear constraints are allowed in the definition of lower-level feasible set in [8]. However, the KKT conditions of the lower-level variational inequality problem are a parametric mixed complementarity problem. As mentioned in Section 1, this case can be treated as an MPEC of the form (1) with some additional upper-level equality constraints.

7 Special Examples of Smoothing Functions

In this section we give examples of the function ψ satisfying the assumptions (A1)–(A5). Hence these special forms of ψ correspond to particular implementations of smooth SQP methods for MPEC.

Example 7.1

$$\psi(a, b, c) = \sqrt{a^2 + b^2 + c^2} - (a + b).$$

This function is used for proposing an SQP method in [12]. Corresponding to ψ is the function $\phi(a, b) = \sqrt{a^2 + b^2} - (a + b)$, which is now known as the Fischer-Burmeister function [9]. The introduction of ψ originates from [21] for handling linear complementarity problems.

If $(a, b, c) \neq (0, 0, 0)$, then ψ is smooth at (a, b, c) with $\nabla\psi(a, b, c) = (p, q, r)$ such that

$$p = \frac{a}{\sqrt{a^2 + b^2 + c^2}} - 1, \; q = \frac{b}{\sqrt{a^2 + b^2 + c^2}} - 1, \; r = \frac{c}{\sqrt{a^2 + b^2 + c^2}}.$$

If $(a,b,c) = (0,0,0)$, then ψ is locally Lipschitz at (a,b,c) and its generalized Jacobian is the ball [18]:

$$\partial\psi(a,b,c) = \{(p,q,r) :\ (p+1)^2 + (q+1)^2 + r^2 \le 1\}.$$

Example 7.2

$$\psi(a,b,c) = \sqrt{(a-b)^2 + c^2} - (a+b).$$

This function is used for proposing a smoothing method in [8]. Corresponding to ψ is the function $\phi(a,b) = |a-b| - (a+b) = -2\min\{a,b\}$. The introduction of ψ also originates from [21].

If either $a \ne b$ or $c \ne 0$, then ψ is smooth at (a,b,c) with $\nabla\psi(a,b,c) = (p,q,r)$ such that

$$p = \frac{a-b}{\sqrt{(a-b)^2+c^2}} - 1,\ q = \frac{b-a}{\sqrt{(a-b)^2+c^2}} - 1,\ r = \frac{c}{\sqrt{(a-b)^2+c^2}}.$$

If $a = b$, and $c = 0$, then ψ is locally Lipschitz at (a,b,c) and its generalized Jacobian is the intersection of a plane with a box:

$$\partial\psi(a,b,c) = \{(p,q,r) :\ p+q = -2,\ p \in [-2,0],\ q \in [-2,0],\ r \in [-1,1]\}.$$

Example 7.3

$$\begin{aligned}\psi(a,b,c) &= \sqrt{a^2+b^2+\lambda ab + c^2} - (a+b)\\ \phi(a,b) &= \sqrt{a^2+b^2+\lambda ab} - (a+b)\end{aligned}$$

where $\lambda \in [-2,2)$ is a parameter. The function ϕ is introduced in [22] for solving nonlinear complementarity problems. Apparently, when $\lambda = 0$, ϕ reduces to the Fischer-Burmeister function (Example 7.1), and when $\lambda = -2$, ϕ reduces to the min function (Example 7.2).

So we may assume that $\lambda \in (-2,2)$. If $(a,b,c) \ne (0,0,0)$, then ψ is smooth at (a,b,c) and $\nabla\psi(a,b,c) = (p,q,r)$ with

$$p = \frac{a+\lambda b/2}{\sqrt{a^2+b^2+\lambda ab+c^2}} - 1,\ q = \frac{b+\lambda a/2}{\sqrt{a^2+b^2+\lambda ab+c^2}} - 1,\ r = \frac{c}{\sqrt{a^2+b^2+\lambda ab+c^2}}.$$

If $(a,b,c) = (0,0,0)$, then ψ is locally Lipschitz at (a,b,c) and its generalized Jacobian is an ellipsoid:

$$\partial\psi(a,b,c) = \{(p,q,r) :\ \alpha(p+1)^2 + \alpha(q+1)^2 + \beta(p-q)^2 + r^2 \le 1\},$$

where $\alpha = \frac{2}{2+\lambda}$, $\beta = \frac{2\lambda}{4-\lambda^2}$.

Example 7.4

$$\begin{aligned}\psi(a,b,c) &= \lambda[\sqrt{a^2+b^2+c^2} - (a+b)] - \tfrac{(1-\lambda)}{4}(\sqrt{a^2+c^2}+a)(\sqrt{b^2+c^2}+b)\\ \phi(a,b) &= \lambda[\sqrt{a^2+b^2} - (a+b)] - (1-\lambda)\max\{a,0\}\ \max\{b,0\}\end{aligned}$$

where $\lambda \in (0,1]$ is a parameter. The function ϕ is introduced in [3] for solving nonlinear complementarity problems. When $\lambda = 1$, ϕ reduces to the Fischer-Burmeister function in Example 7.1.

If $c \neq 0$, then ψ is smooth at (a,b,c) and $\nabla\psi(a,b,c) = (p,q,r)$ with

$$p = \lambda(\tfrac{a}{\sqrt{a^2+b^2+c^2}} - 1) - \tfrac{1-\lambda}{4}(\tfrac{a}{\sqrt{a^2+c^2}} + 1)(\sqrt{b^2+c^2} + b)$$

$$q = \lambda(\tfrac{b}{\sqrt{a^2+b^2+c^2}} - 1) - \tfrac{1-\lambda}{4}(\tfrac{b}{\sqrt{b^2+c^2}} + 1)(\sqrt{a^2+c^2} + a)$$

$$r = \lambda\tfrac{c}{\sqrt{a^2+b^2+c^2}} - \tfrac{1-\lambda}{4}[\tfrac{c}{\sqrt{a^2+c^2}}(\sqrt{b^2+c^2} + b) + \tfrac{c}{\sqrt{b^2+c^2}}(\sqrt{a^2+c^2} + a)].$$

If $(a,b,c) = (0,0,0)$, then ψ is locally Lipschitz at (a,b,c) and its generalized Jacobian is the ball:

$$\partial\psi(a,b,c) = \{(p,q,r): \ (p+\lambda)^2 + (q+\lambda)^2 + r^2 \leq \lambda^2\}.$$

If $(a,b) \neq (0,0)$ and $c = 0$, then ψ is locally Lipschitz at $(a,b,0)$ and its Jacobian or generalized Jacobian is of the form:

$$\partial\psi(a,b,0) = \left\{ (p,q,r): \begin{array}{c} p = \lambda(\frac{a}{\sqrt{a^2+b^2}} - 1) - \frac{1-\lambda}{4}\alpha(|b|+b) \\[2ex] q = \lambda(\frac{b}{\sqrt{a^2+b^2}} - 1) - \frac{1-\lambda}{4}\beta(|a|+a) \\[2ex] r = -\frac{1-\lambda}{4}[\gamma_a(|b|+b) + \gamma_b(|a|+a)] \\[2ex] \begin{array}{ll} \alpha \in [0,2] & \text{if } a = 0 \\ \alpha = \frac{a}{|a|} + 1 & \text{if } a \neq 0 \end{array} \\[2ex] \begin{array}{ll} \beta \in [0,2] & \text{if } b = 0 \\ \beta = \frac{b}{|b|} + 1 & \text{if } b \neq 0 \end{array} \\[2ex] \begin{array}{ll} \gamma_a \in [-1,1] & \text{if } a = 0 \\ \gamma_a = 0 & \text{if } a \neq 0 \end{array} \\[2ex] \begin{array}{ll} \gamma_b \in [-1,1] & \text{if } b = 0 \\ \gamma_b = 0 & \text{if } b \neq 0 \end{array} \end{array} \right\}.$$

Note that $a/|a|$ is equal to the sign of a for $a \neq 0$.

Part (i) of the next proposition demonstrates that special explicit smooth SQP methods can be proposed based on these smoothing functions. Its proof is an easy consequence of the above formulae for $\partial\psi(a,b,c)$. Part (ii), which is evident, says that the smoothing functions in Examples 7.1, 7.2 and 7.3 satisfy the smoothness assumption needed for global convergence in Theorems 5.10.

Proposition 7.5

(i) Each function ψ defined in Examples 7.1, 7.2, 7.3 and 7.4 satisfies the assumptions (A1)–(A5).

(ii) Each function ϕ defined in Examples 7.1, 7.2 and 7.3 is twice continuously differentiable at any nondegenerate point (a,b), i.e., $a \neq b$.

8 Concluding Remarks

In this article, mathematical programs with equilibrium constraints are reformulated as better-posed nonsmooth programs and then, by means of so-called smoothing functions, approximated by (smooth) nonlinear programs. Consequently, some techniques that are well known in the context of nonlinear programming can be used for solving MPEC. In particular, we have developed two classes of SQP methods. Some global convergence results of these methods have been established. Numerical experience is yet to be established.

The extent of these convergence results depends critically on the convergence theory available for the corresponding nonlinear programming algorithm. So we expect the future application of different nonlinear programming methods in the context of smoothing for MPEC and other nonsmooth optimization problems will give rise to different global convergence results.

We have also given concrete examples of smoothing functions motivated by the literature on complementarity problems. It would be interesting to find other smoothing functions to satisfy the assumptions (A1)–(A5), and other smoothing functions which may not satisfy those assumptions but may play similar roles in other algorithms.

Acknowledgement. The authors thank Francisco Facchinei for suggesting the reference [1] and for subsequent discussions on nonlinear programming; Jong-Shi Pang for comments on the meaning of "well-posedness"; and two anonymous referees for their insightful comments.

References

[1] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, Sydney, 1982.

[2] J.V. Burke and S.P. Han, *A robust sequential quadratic programming method*, Math. Programming, 43 (1989), pp. 277–303.

[3] B. Chen, X. Chen and C. Kanzow, *A penalized Fischer-Burmeister NCP-function*, Math. Programming, to appear.

[4] Y. Chen and M. Florian, *The nonlinear bilevel programming problem: formulations, regularity and optimality conditions*, Optimization, 32 (1995), pp. 193–209.

[5] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[6] R.W. Cottle, J.S. Pang and R.E. Stone, *The Linear Complementarity Problem*, Academic Press, New York, 1992.

[7] S. Dirkse and M.C. Ferris, *Modeling and solution environments for MPEC: GAMS & MATLAB*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Norwell, MA, 1999, pp. 127–148.

[8] F. Facchinei, H. Jiang and L. Qi, *A smoothing method for mathematical programs with equilibrium constraints*, Math. Programming, 85 (1999), pp. 107–134.

[9] A. Fischer, *A special Newton-type optimization method*, Optimization, 24 (1992), pp. 269–284.

[10] A. Fischer and C. Kanzow, *On finite termination of an iterative method for linear complementarity problems*, Math. Programming, 74 (1996), pp. 279–292.

[11] R. Fletcher, *Practical Methods of Optimization*, John Wiley, 2nd Edition, 1987.

[12] M. Fukushima, Z.-Q. Luo and J.S. Pang, *A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints*, Comput. Optim. Appl., 10 (1998), pp. 5–34.

[13] M. Fukushima and J.S. Pang, *Some feasibility issues in mathematical programs with equilibrium constraints*, SIAM J. Optim., 8 (1998), pp. 673–681.

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

[15] S.P. Han, *A globally convergent method for nonlinear programming*, to appear in J. Optim. Theory Appl., 22 (1977), pp. 297–309.

[16] J.-B. Hiriart-Urruty, *Refinements of necessary optimality conditions in nondifferentiable programming I*, Appl. Math. Optim., 5 (1979), pp. 63–82.

[17] R. Janin, *Directional derivative of the marginal function in nonlinear programming*, Math. Programming Study, 21 (1984), pp. 110–126.

[18] H. Jiang, *Smoothed Fischer-Burmeister equation methods for the complementarity problem*, Department of Mathematics and Statistics, The University of Melbourne, June 1997.

[19] H. Jiang and D. Ralph, *QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints*, Comput. Optim. Appl., 13 (1999), pp. 25–59.

[20] H. Jiang and D. Ralph, *Smooth SQP methods for mathematical programs with nonlinear complementarity constraints*, Department of Mathematics and Statistics, The University of Melbourne, June, 1999 (revised), *http://www.ms.unimelb.edu.au/~danny/smooth-mpec.ps.*

[21] C. Kanzow, *Some noninterior continuation methods for linear complementarity problems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 851–868.

[22] C. Kanzow and H. Kleinmichel, *A new class of semismooth Newton-type methods for nonlinear complementarity problems*, Comput. Optim. Appl., 11 (1998), pp. 227–251.

[23] M. Kočvara and J.V. Outrata, *On optimization of systems governed by implicit complementarity problems*, Numer. Funct. Anal. Optim. 15 (1994), pp. 869–887.

[24] M. Kočvara and J.V. Outrata, *On the solution of optimum design problems with variational inequalities*, in Recent Advances in Nonsmooth Optimization, D.Z. Du, L. Qi and R.S. Womersley, eds., World Scientific, Singapore, 1995, pp. 172–192.

[25] A.V. Kuntsevich and F. Kappel, *SolvOpt: The solver for local nonlinear optimization problems*, Institute for Mathematics, Karl-Franzens University of Graz, Austria, 1997.

[26] M.B. Lignola and J. Morgan, *Stability of regularized bilevel programming problem*, J. Optim. Theory Appl., 93 (1997), pp. 575–596.

[27] M.B. Lignola and J. Morgan, *Existence and solutions to generalized bilevel programming problems*, in Multilevel Optimization: Algorithms, Complexity and Applications, A. Migdalas, P.M. Pardalos, P. Värbrand, eds., Kluwer Academic Publishers, Norwell, MA, 1998, pp. 315–332.

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

[29] Z.-Q. Luo, J.S. Pang and D. Ralph, *Piecewise sequential quadratic programming for mathematical Programs with nonlinear complementarity constraints*, in Multilevel Optimization: Algorithms, Complexity and Applications, A. Migdalas, P.M. Pardalos, P. Värbrand, eds., Kluwer Academic Publishers, Norwell, MA, 1998, pp. 209–229.

[30] Z.-Q. Luo, J.S. Pang, D. Ralph and S.-Q. Wu, *Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints*, Math. Programming, 75 (1996), pp. 19–76.

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

[32] O.L. Mangasarian, *Mathematical programming in machine learning*, in Nonlinear Optimization and Applications, G. Di Pillo and F. Giannessi, eds., Plenum Publishing Corporation, New York, 1996, pp. 283–295.

[33] J.V. Outrata, *On optimization problems with variational inequality constraints*, SIAM J. Optim., 4 (1994), pp. 340–357.

[34] J.V. Outrata and J. Zowe, *A numerical approach to optimization problems with variational inequality constraints*, Math. Programming, 68 (1995), pp. 105–130.

[35] M.J.D. Powell, *The convergence of variable metric methods for nonlinearly constrained optimization calculations*, in Nonlinear Programming 3, O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds., Academic Press, New York, 1978, pp. 27–63.

[36] M.J.D. Powell, *Variable metric methods for constrained optimization*, in Mathematical Programming: The State of the Art, A. Bachem, M. Grötschel and B. Korte, eds., Springer-Verlag, Berlin, 1983, pp. 288–311.

[37] D. Ralph, *Sequential quadratic programming for mathematical programs with linear complementarity constraints*, in CTAC95 Computational Techniques and Applications, R.L. May and A.K. Easton, eds., World Scientific, Singapore, 1996, pp. 663–668.

[38] S.M. Robinson, *Stability theory for systems of inequalities, part II: Differentiable nonlinear systems*, SIAM J. Numer. Anal., 13 (1976), pp. 497–513.

[39] R.T. Rockafellar, *Convex Analysis*, Princeton, New Jersey, 1970.

[40] H. Scheel and S. Scholtes, *Mathematical programs with equilibrium constraints: Stationarity, optimality, and sensitivity*, Math. Oper. Res., to appear.

[41] S. Scholtes and M. Stöhr, *Exact penalization of mathematical programs with equilibrium constraints*, SIAM J. Control Optim. 37 (1999), pp. 617–652.

[42] N. Yamashita and M. Fukushima, *Modified Newton methods for solving a semismooth reformulation of monotone complementarity problems*, Math. Programming, 76 (1997), pp. 469–497.

[43] J.J. Ye and D.L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization, 33 (1995), pp. 9–27.

A Appendix: Global Convergence of a Modified SQP Method for Smooth Nonlinear Programs

Consider the nonlinear programming problem:

$$\begin{array}{rl} \min\limits_{u} & f(u) \\ \text{subject to} & g(u) \geq 0 \\ & h(u) = 0, \end{array} \tag{24}$$

where $f : \Re^n \to \Re$, $g : \Re^n \to \Re^l$ and $h : \Re^n \to \Re^m$ are continuously differentiable. The standard ℓ_1-exact penalty function for this problem is defined as

$$\theta_\rho(u) \;=\; f(u) + \rho \left[\sum_{1 \leq i \leq l} \max\{-g_i(u), 0\} + \sum_{1 \leq j \leq m} |h_j(u)| \right]$$

where $\rho > 0$ is a parameter. Associated with the above penalty function, we define another penalty function using two different penalty parameters:

$$\Theta_{(\rho^g, \rho^h)}(u) \;=\; f(u) + \rho^g \sum_{1 \leq i \leq l} \max\{-g_i(u), 0\} + \rho^h \sum_{1 \leq j \leq m} |h_j(u)|.$$

When $\rho^g = \rho^h = \rho$, $\theta_\rho = \Theta_{(\rho^g, \rho^h)}$. We will give references and comments after presenting a sequential quadratic programming (SQP) method which uses this exact penalty function and its associated penalty function as a merit function.

Algorithm: `Modified SQP Method`

Step 0. (**Initialization**) Let $\rho_{-1} > 0$, $\delta_1 > 0$, $\delta_2 > 0$, $\sigma \in (0, 1)$, $\tau \in (0, 1)$. Choose $u^0 \in \Re^n$ and a symmetric positive definite matrix $W_0 \in \Re^{n \times n}$. Set $k := 0$.

Step 1. (**Search direction**) Solve the following modified QP problem (which is a modified quadratic model of (24)) with $u = u^k$, $W = W_k$, and $\rho = \rho_{k-1}$:

$$\begin{array}{rl} \min\limits_{d \in \Re^n, \xi \in \Re^l} & \nabla f(u)^T d + \frac{1}{2} d^T W d + \rho \sum\limits_{i=1}^{l} \xi_i \\ \text{subject to} & g'(u) d + g(u) \geq -\xi \\ & h'(u) d + h(u) = 0 \\ & \xi \geq 0. \end{array} \tag{25}$$

Let (d^k, ξ^k) be a solution of this QP and $\lambda^k = (\lambda_g^k, \lambda_h^k, \lambda_\xi^k)$ be its corresponding KKT multiplier.

Step 2. (**Termination check**) If a stopping rule is satisfied, terminate. Otherwise, go to Step 3.

Step 3. (**Penalty update**) Let

$$\tilde{\rho}_k = \begin{cases} \rho_{k-1} & \text{if } \rho_{k-1} \geq \max\limits_{1 \leq i \leq l+m} |\lambda_i^k| \\ \delta_1 + \max\limits_{1 \leq i \leq l+m} |\lambda_i^k| & \text{otherwise.} \end{cases}$$

Define $\rho_k^g = \rho_{k-1}$, $\rho_k^h = \tilde{\rho}_k$, and

$$\rho_k = \begin{cases} \tilde{\rho}_k & \text{if } \sum_{1 \le i \le l} \xi_i^k = 0 \\ \tilde{\rho}_k + \delta_2 & \text{otherwise.} \end{cases}$$

Step 4. (**Line search**) Let $t_k = (\tau)^{i_k}$ where i_k is the smallest nonnegative integer such that $i = i_k$ satisfies

$$\Theta_{(\rho_k^g, \rho_k^h)}(u^k + (\tau)^i d^k) - \Theta_{(\rho_k^g, \rho_k^h)}(u^k) \le -\sigma(\tau)^i (d^k)^T W_k d^k.$$

Step 5. (**Update**) Let $u^{k+1} = u^k + t_k d^k$. Choose a symmetric positive definite matrix $W_{k+1} \in \Re^{n \times n}$. Set $k := k+1$. Go to Step 1.

SQP methods and variants are amongst the most important and most popular methods for general nonlinear programs. In classical SQP methods such as [15, 36], the search direction is usually obtained by solving a quadratic model of (24), i.e., the following QP problem with $u = u^k$ and $W = W_k$:

$$\begin{array}{rl} \min\limits_{d \in \Re^n} & \nabla f(u)^T d + \frac{1}{2} d^T W d \\ \text{subject to} & g'(u)d + g(u) \ge 0 \\ & h'(u)d + h(u) = 0. \end{array} \tag{26}$$

Global convergence of the SQP methods can be established under standard assumptions. See for example, [15, 36]. See also [1, 2, 11] for variants.

Our modified SQP method is close to the modified SQP methods proposed in [1, 2]; in particular we use [1] as a kind of template for the development of the subsequent results. However unlike [1], we use the ℓ_1- rather than the ℓ_∞-exact penalty function, which leads to modified QP (25) different from the QP used in [1], and our penalty parameter update must also take into account the way the equality constraints are treated differently from the inequality constraints (the former are not relaxed).

Remarks.

(i) The introduction of the modified QP (25) aims to improve the prospect of feasibility of the QP (26). Certainly, one may further improve the prospect of feasibility of the modified QP (25) by relaxing the equality constraints. However, we do not proceed in this way since the relaxation of the inequality constraints is enough as far as this paper is concerned. It can be easily seen that our modified SQP method is exactly the same as the classical SQP methods of [15, 36] if the modified QP (25) is replaced by the QP (26) at Step 3 and $\rho_k^g = \rho_k^h = \tilde{\rho}_k$ at Step 4.

(ii) In the merit function $\Theta_{(\rho^g, \rho^h)}$, there are two different penalty parameters ρ^g and ρ^h. In order to ensure a solution of the modified QP (25) to be a descent direction of the merit function $\Theta_{(\rho^g, \rho^h)}$ at its noncritical point, we must carefully choose values for penalty parameters ρ^g and ρ^h. It turns out that ρ^g should be equal to ρ used in the objective function of (25), and that ρ^h should be no less than the minimum of the absolute values of all multipliers corresponding to equality constraints in (25). The inequality constraints are relaxed through the introduction of the nonnegative variable ξ, which measures the feasibility gap of inequality constraints. The ultimate goal of the algorithm is to find a solution for (24), which must be at least feasible. Therefore, The penalty term $\rho \sum xi_i$ in the objective function of (25) is to force the feasibility gap as small as possible. As can be seen from Step 3, the penalty parameter ρ is increased by a positive constant δ_2 if such a gap is not zero.

If the QP (26) has a solution d and $\lambda = (\lambda_g, \lambda_h)$ is its corresponding KKT multiplier, then the KKT condition of this QP can be written as follows:

$$\begin{array}{l} \nabla f(u) + Wd - g'(u)^T \lambda_g + h'(u)^T \lambda_h = 0 \\ 0 \le g'(u)d + g(u) \ \perp \ \lambda_g \ge 0 \\ h'(u)d + h(u) = 0. \end{array} \tag{27}$$

Similarly, if the modified QP (25) has a solution (d, ξ) and $\lambda = (\lambda_g, \lambda_h, \lambda_\xi)$ is its corresponding KKT multiplier, then the KKT condition of the modified QP is read as:

$$\begin{array}{l} \nabla f(u) + Wd - g'(u)^T \lambda_g + h'(u)^T \lambda_h = 0 \\ \rho \ \tilde{e} = \lambda_g + \lambda_\xi \\ 0 \le g'(u)d + g(u) + \xi \ \perp \ \lambda_g \ge 0 \\ h'(u)d + h(u) = 0 \\ 0 \le \xi \ \perp \ \lambda_\xi \ge 0, \end{array} \tag{28}$$

where $\tilde{e}$ is the vector of all ones in $\Re^l$.

Before giving a proof of the global convergence of the modified SQP method, we next study some useful properties associated with KKT points of the NLP (24) and critical points of the penalty function $\Theta_{(\rho^g,\rho^h)}$.

Proposition A.1

(i) For any $u \in \Re^n$ and $d \in \Re^n$, $\rho^g \in \Re$ and $\rho^h \in \Re$, $\Theta'_{(\rho^g,\rho^h)}$ is directionally differentiable at u along the direction d and it holds that

$$\begin{array}{rcl} \Theta'_{(\rho^g,\rho^h)}(u;d) & = & \nabla f(u)^T d \\ & & + \rho^g \left[\sum_{i:\ g_i<0} -g_i'(u)d + \sum_{i:\ g_i=0} \max\{-g_i'(u)d, 0\} + \sum_{i:\ g_i>0} 0 \right] \\ & & + \rho^h \left[\sum_{j:\ h_j>0} h_j'(u)d + \sum_{j:\ h_j=0} |h_j'(u)d| + \sum_{j:\ h_j<0} -h_j'(u)d \right]. \end{array}$$

(ii) If d is a solution of the QP (26), and if $\min\{\rho^g, \rho^h\} \ge \max_{1\le i\le l+m} |\lambda_i|$ where $\lambda = (\lambda_g, \lambda_h)$ is the KKT multiplier of the QP (26), then

$$\Theta'_{(\rho^g,\rho^h)}(u;d) \le \ \nabla f(u)^T d - (\lambda_g)^T g'(u)d + (\lambda_h)^T h'(u)d$$

and

$$\Theta'_{(\rho^g,\rho^h)}(u;d) \le -d^T W d.$$

(iii) If (d, ξ) is a solution of the modified QP (25), $\rho^g = \rho$ and $\rho^h \geq \max_{1\leq i\leq l+m} |\lambda_i|$ where $\lambda = (\lambda_g, \lambda_h, \lambda_\xi)$ is the KKT multiplier of the modified QP (25), then

$$\Theta'_{(\rho^g,\rho^h)}(u; d) \leq \nabla f(u)^T d - (\lambda_g)^T g'(u)d + (\lambda_h)^T h'(u)d$$

and

$$\Theta'_{(\rho^g,\rho^h)}(u; d) \leq -d^T W d.$$

Proof. (i) The directional differentiability of $\Theta_{(\rho^g,\rho^h)}$ follows from the continuous differentiability of f, g and h. The directional derivative of $\Theta_{(\rho^g,\rho^h)}$ is straightforward to calculate.

(ii) By direct calculation, one may show that

$$\Theta'_{(\rho^g,\rho^h)}(u; d) \leq \nabla f(u)^T d - \rho^g \|\max\{-g(u), 0\}\|_1 - \rho^h \|h(u)\|_1.$$

Since $\lambda = (\lambda_g, \lambda_h)$ is the KKT multiplier of the QP (26), it follows from the KKT condition (27) of this QP that

$$-(\lambda_g)^T g'(u)d = (\lambda_g)^T g(u) \geq -\sum_{1\leq i\leq l} (\lambda_g)_i \; |\max\{-g_i(u), 0\}|$$

$$(\lambda_h)^T h'(u)d = -(\lambda_h)^T h(u) \geq -\sum_{1\leq j\leq m} |(\lambda_h)_j| \; |h_j(u)|.$$

Therefore, the first inequality in (ii) follows from $\min\{\rho^g, \rho^h\} \geq \max_{1\leq i\leq l+m} |\lambda_i|$, and the second inequality follows from the KKT condition (27).

(iii) Let the directional derivative of $\max\{-g_i(u), 0\}$ at u along the direction d be denoted by $(\max\{-g_i(u), 0\})'(u; d)$ for each i $(1 \leq i \leq l)$. It suffices to show the following inequality holds for each i:

$$\rho^g (\max\{-g_i(u), 0\})'(u; d) \leq -(\lambda_g)_i g_i'(u)d. \tag{29}$$

Case I: $\xi_i = 0$. If $g_i(u) < 0$, then

$$(\max\{-g_i(u), 0\})'(u; d) = -g_i'(u)d \leq g_i(u) + \xi_i < 0,$$

which implies (29) by $(\lambda_g)_i \leq \rho^g$. If $g_i(u) = 0$, then

$$(\max\{-g_i(u), 0\})'(u; d) = \max\{-g_i'(u)d, 0\} = 0$$

and

$$-(\lambda_g)_i g_i'(u)d = (\lambda_g)_i (g_i(u) + \xi_i) = 0,$$

where the second equation follows from the complementarity condition $(\lambda_g)_i(g_i'(u)d + g_i(u) + \xi_i) = 0$. Therefore (29) follows. If $g_i(u) > 0$, then the complementarity condition shows that

$$-(\lambda_g)_i g_i'(u)d = (\lambda_g)_i (g_i(u) + \xi_i) \geq 0.$$

So the fact that $(\max\{-g_i(u), 0\})'(u; d) = 0$ shows (29).

Case II: $\xi_i > 0$. In this case, we have $\rho^g = (\lambda_g)_i$. If $g_i(u) < 0$, it is straightforward. If $g_i(u) = 0$, then the complementarity condition implies that

$$\begin{array}{rcl}\rho^g(\max\{-g_i(u),0\})'(u;d) & = & \rho^g \max\{-g_i'(u)d,0\} \\ & = & \max\{-(\lambda_g)_i g_i'(u)d,0\} \\ & = & \max\{(\lambda_g)_i \xi_i,0\} \\ & = & (\lambda_g)_i \xi_i \\ & = & -(\lambda_g)_i g_i'(u)d.\end{array}$$

If $g_i(u) > 0$, the complementarity condition shows that

$$\begin{array}{rcl}\rho^g(\max\{-g_i(u),0\})'(u;d) & = & 0 \\ & \leq & (\lambda_g)_i(g_i(u)+\xi_i) \\ & = & -(\lambda_g)_i g_i'(u)d.\end{array}$$

■

Proposition A.2 *Suppose the matrix W is symmetric positive semi-definite.*

(i) If d is a solution of the QP (26) and $\lambda = (\lambda_g, \lambda_h)$ is its KKT multiplier, then $(d,0)$ is a solution of the modified QP (25) for any $\rho \geq \max_{1\leq i\leq l+m} |\lambda_i|$.

(ii) If $(d,0)$ is a solution of the modified QP (25) for any $\rho > 0$, then d is a solution of the QP (26).

Proof. (i) If d is a solution of the QP (26) and $\lambda = (\lambda_g, \lambda_h)$ is its KKT multiplier, then $(d,0)$ is a KKT point of the modified QP (25) with a KKT multiplier $(\lambda_g, \lambda_h, \lambda_\xi)$ such that $\lambda_\xi = \rho\, \tilde{e} - \lambda_g$ for any $\rho \geq \max_{1\leq i\leq l+m} |\lambda_i|$. Since W is symmetric positive semi-definite, the modified QP (25) is a convex quadratic program. This shows that $(d,0)$ is a solution of the modified QP (25).

(ii) Similar to (i). ■

Proposition A.3 *Suppose the matrix W is symmetric positive definite.*

(i) Assume the modified QP (25) is feasible at u^*. If for a given $\rho^g > 0$ and all sufficiently large $\rho^h > 0$, u^* is a critical point of $\Theta_{(\rho^g,\rho^h)}$, then there exists a vector (d,ξ) with $d = 0$ such that (d,ξ) is a solution of the modified QP (25) with $u = u^*$ and $\rho = \rho^g$.

(ii) If there exists a vector (d,ξ) with $d = 0$ such that (d,ξ) is a solution of the modified QP (25) with $u = u^*$ and $\rho = \rho^g$ and $\lambda = (\lambda_g, \lambda_h)$ is its associated multiplier, then u^* is a critical point of $\Theta_{(\rho^g,\rho^h)}$ for any $\rho^h \geq \max_{1\leq i\leq m+l} |\lambda_i|$.

Proof. (i) Suppose u^* is a critical point of $\Theta_{(\rho^g,\rho^h)}$. Since the modified QP (25) is feasible and convex, it has an optimal solution (d,ξ). If $d \neq 0$, then the second inequality of (iii) in Proposition A.1 and the positive definiteness of the matrix W imply that $\Theta'_{(\rho^g,\rho^h)}(u^*;d) < 0$, which is a contradiction. Therefore $d = 0$.

(ii) Suppose $(0, \xi)$ is a solution of the modified QP (25). Then the KKT condition of (25) is read as follows: For some KKT multiplier $(\lambda_g, \lambda_h, \lambda_\xi)$,

$$\begin{array}{l} \nabla f(u^*) - g'(u^*)\lambda_g + h'(u^*)\lambda_h = 0 \\ \rho^g \ \tilde{e} = \lambda_g + \lambda_\xi \\ 0 \leq g(u^*) + \xi \ \perp \ \lambda_g \geq 0 \\ h(u^*) = 0 \\ 0 \leq \xi \ \perp \ \lambda_\xi \geq 0. \end{array}$$

Assume that $d \neq 0$ is any given vector in $\Re^n$. We now calculate $\Theta'_{(\rho^g,\rho^h)}(u^*; d)$ term by term. If $g_i(u^*) < 0$, then the above KKT condition shows that $\xi_i > 0$, $(\lambda_\xi)_i = 0$ and $(\lambda_g)_i = \rho^g$. It follows that

$$\rho^g(-g_i'(u^*)d) = -(\lambda_g)_i g_i'(u^*)d.$$

If $g_i(u^*) = 0$, then it is easy to show that

$$\rho^g \max\{-g_i'(u^*)d, 0\} \geq -(\lambda_g)_i g_i'(u^*)d.$$

If $g_i(u^*) > 0$, then $(\lambda_g)_i = 0$ and

$$\rho^g \ 0 = -(\lambda_g)_i g_i'(u^*)d.$$

The KKT condition shows that $h(u^*) = 0$. Clearly, we have

$$\rho^h |h_j'(u^*)d| \geq |(\lambda_h)_j \ h_j'(u^*)d| \geq (\lambda_h)_j h_j'(u^*)d.$$

From the above argument, we have the following inequality

$$\begin{array}{rcl} \Theta'_{(\rho^g,\rho^h)}(u^*; d) & \geq & \nabla f(u^*)^T d - \lambda_g^T g'(u^*)d + \lambda_h^T h'(u^*)d \\ & = & (\nabla f(u^*) - g'(u^*)^T \lambda_g + h'(u^*)^T \lambda_h)^T d \\ & = & 0, \end{array}$$

where the last equality follows from the first equation of the above KKT condition. Therefore, u^* is a critical point of $\Theta_{(\rho^g,\rho^h)}$. ■

Proposition A.4 *Suppose the matrix W is symmetric positive definite. Then u^* is a KKT point of the nonlinear program (24) if and only if $d = 0$ is the unique solution of the QP (26) at $u = u^*$.*

Proof. This result is classical and we omit the proof. ■

Proposition A.5

(i) If u^* is a KKT point of the nonlinear program (24) and λ is its KKT multiplier, and if $\min\{\rho^g, \rho^h\} \geq \max_{1 \leq i \leq l+m} |\lambda_i|$, then u^* is a critical point of $\Theta_{(\rho^g,\rho^h)}$.

(ii) If u^* is a critical point of $\Theta_{(\rho^g,\rho^h)}$ for some $\rho^g > 0$ and some sufficiently large ρ^h, and u^* is a feasible point of the nonlinear program (24), then u^* is a KKT point of (24).

Proof. (i) If u^* is a KKT point of the nonlinear program (24) and λ is its KKT multiplier, then Proposition A.4 implies that $d = 0$ is a unique solution of the QP (26) with any positive definite matrix W and $u = u^*$ and λ its multiplier. It follows from (i) of Proposition A.2 that $(0, 0)$ is a solution of the modified QP (25) with the matrix W, $u = u^*$ and $\rho \geq \max\limits_{1 \leq i \leq l+m} |\lambda_i|$. Hence (ii) of Proposition A.3 shows that u^* is a critical point of $\Theta_{(\rho^g,\rho^h)}$ if ρ^h is sufficiently large.

(ii) If u^* is a critical point of the penalty function $\Theta_{(\rho^g,\rho^h)}$ for some $\rho^g > 0$ and $\rho^h > 0$, then (i) of Proposition A.3 implies that (d, ξ) with $d = 0$ is a solution of the modified QP with the matrix W, the penalty parameter $\rho = \rho^h$ and $u = u^*$. Consequently, the KKT condition (28) holds for some KKT multiplier $\lambda = (\lambda_g, \lambda_h, \lambda_\xi)$ with $u = u^*$. Here we show that $\xi = 0$. If $\xi_i \neq 0$, then $\xi_i > 0$. The complementarity condition $(g_i(u^*) + \xi_i)(\lambda_g)_i = 0$ and the feasibility condition that $g_i(u^*) \geq 0$ show that $(\lambda_g)_i = 0$. This means that $(\lambda_\xi)_i = \rho^g > 0$. By the complementarity condition $\xi_i(\lambda_\xi)_i = 0$, we have that $\xi_i = 0$, a contradiction. Therefore, $\xi = 0$. It follows from (ii) of Proposition A.2 that $d = 0$ is a (unique) solution of the QP (26) with the matrix W and $u = u^*$. Then the desired result follows from Proposition A.4. ■

Proposition A.6 *Suppose the matrix W is symmetric positive definite.*

(i) If the QP (26) has nonempty feasible set, then it has a unique solution d. Suppose $d \neq 0$ and $\lambda \in \Re^{l+m}$ is its KKT multiplier. Then d is a descent direction of the merit function $\Theta_{(\rho^g,\rho^h)}$ if $\min\{\rho^g, \rho^h\} \geq \max\limits_{1 \leq i \leq l+m} |\lambda_i|$.

(ii) If the modified QP (25) has nonempty feasible set, then it has a solution (d, ξ). Suppose $d \neq 0$ and $\lambda = (\lambda_g, \lambda_h, \lambda_\xi) \in \Re^{l+2m}$ is its KKT multiplier. Then d is a descent direction of the merit function $\Theta_{(\rho^g,\rho^h)}$ for $\rho^g = \rho$ and $\rho^h \geq \max\limits_{1 \leq i \leq l+m} |\lambda_i|$.

Proof. (i) The first part follows from the symmetric positive definiteness of the matrix W. The second part follows from (ii) of Proposition A.1.

(ii) The first part follows from the fact that the modified QP (25) is convex. The second part follows from (iii) of Proposition A.1. ■

To study global convergence of our modified SQP algorithm, in the sequel, we impose the following standard standing conditions:

(C1) For each k, the modified QP (25) has a nonempty feasible set.

(C2) There exist two positive constants $\alpha < \beta$ such that for each k, the following condition holds:

$$\alpha \|d\|^2 \leq d^T W_k d \leq \beta \|d\|^2, \quad \forall d \in \Re^n.$$

(C3) After finitely many iterations, the value of the penalty parameter ρ_k does not change, i.e, for sufficient large k, $\rho_k = \rho_*$, $\rho_k^g = \rho_k = \rho_*$, $\rho_k^h = \rho_*^h$.

Remarks. The condition (C3) is usually not assumed in the literature. However, it is implied by appropriate constraint qualifications on all accumulation points of the sequence $\{u^k\}$. We first investigate global convergence results under the assumption (C3). Later on we shall study global convergence without it.

Lemma A.7 *Under the standing assumptions (C1)–(C3), the modified SQP method is well-defined, and the algorithm either terminates in Step 2 after finitely many iterations, or generates infinite sequences* $\{u^k\}$, $\{d^k\}$ *and* $\{\lambda^k\}$.

Proof. By the standing assumption (C1), the search direction and the penalty parameter update is well-defined at Step 1 and Step 3 respectively. By (iii) of Proposition A.1, Proposition A.6 and the standing assumption (C2), it can be easily shown that the line search at Step 4 is well-defined. Therefore, the well-definedness of the algorithm follows. The proof is complete. ■

At Step 2 of the algorithm, we did not specify a stopping rule. If at some iteration k, (d^k, ξ^k) with $d^k = 0$ is a solution of the modified QP (25), then Proposition A.3 shows that a critical point of the penalty function $\Theta_{(\rho_k^g, \rho_k^h)}$ is obtained if ρ_k^h is sufficiently large. By Proposition A.5, if u^k is feasible for (24), this critical point is also a KKT point of the nonlinear program (24). Therefore, in the sequel, we assume that the algorithm does not terminate and generates the sequences $\{u^k\}$, $\{d^k\}$ and $\{\lambda^k\}$ with $d^k \neq 0$.

Lemma A.8 *Assume the standing conditions (C1)–(C3) hold. Suppose* u^* *is an accumulation point of* $\{u^k\}$, *i.e., for some subset* K, $\lim_{k\to\infty, k\in K} u^k = u^*$. *Then the following conclusions hold.*

(i) For all sufficiently large k, $\sum_{1\le i\le l} \xi_i^k = 0$, *i.e., the QP (26) is feasible, and furthermore,* $\rho_k^g = \rho_k = \rho_*$, $\rho_k^h = \rho_*^h$.

(ii) $\{\lambda^k\}$ *and* $\{d^k\}_{k\in K}$ *are bounded.*

(iii) If d^*, ξ^* *and* W^* *are accumulation points of the sequences* $\{d^k\}_{k\in K}$, $\{\xi^k\}_{k\in K}$ *and* $\{W_k\}_{k\in K}$ *respectively, then* (d^*, ξ^*) *is a solution of the modified QP (25) with* $u = u^*$, $W = W^*$ *and* $\rho = \rho_*$. *Furthermore,*

$$\Theta'_{(\rho_*^g, \rho_*^h)}(u^*; d^*) \le -(d^*)^T W^* d^*.$$

Proof. (i) It follows from the penalty update rule in Step 3 and the standing condition (C3).

(ii) The penalty update rule in Step 3 implies that the boundedness of $\{\lambda^k\}$. By the first equation of the KKT condition, we can show that the set $\{W_k d^k\}_{k\in K}$ is bounded. Then the boundedness of the subsequence $\{d^k\}_{k\in K}$ is implied by the standing assumption (C2).

(iii) In view of (i), (ii), and the KKT condition (28) for each $k \in K$, by taking the limit, we obtain for some vector $\xi^* = 0$ that (d^*, ξ^*) is a KKT point of the modified QP (25) with $u = u^*$, $W = W^*$ and $\rho = \rho_*$. Since this modified QP is convex, (d^*, ξ^*) is a solution of this QP. Furthermore, the assumption in (iii) of Proposition A.1 is satisfied for this modified QP. Therefore the last inequality in (iv) follows. ■

The following lemma is a technical result.

Lemma A.9 *Let* $\rho^g > 0$ *and* $\rho^h > 0$. *Suppose* $\lim_{k\to\infty, k\in K} u^k = u^*$, $\lim_{k\to\infty, k\in K} t_k = 0$, *and* $\lim_{k\to\infty, k\in K} d^k = d^*$. *Then it holds that*

$$\limsup_{k\to\infty, k\in K} \frac{\Theta_{(\rho^g, \rho^h)}(u^k + t_k d^k) - \Theta_{(\rho^g, \rho^h)}(u^k)}{t_k} \le \Theta'_{(\rho^g, \rho^h)}(u^*; d^*).$$

Proof. Let $G : \Re^n \to \Re^{1+l+m}$ be defined by $G(u) = (f(u), g(u), h(u))$. Then the function $\Theta_{(\rho^g,\rho^h)}(u)$ is a composite function of some piecewise affine convex function φ and the function G, i.e., $\Theta_{(\rho^g,\rho^h)}(u) = \varphi(G(u))$. For any k, convexity of the function φ implies that there exists a vector $\eta^k \in \partial\varphi(G(u^k + t_k d^k))$ such that

$$\varphi(G(u^k + t_k d^k)) - \varphi(G(u^k)) \leq (\eta^k)^T (G(u^k + t_k d^k) - G(u^k)).$$

Therefore, because of boundedness of $\{\eta^k\}_{k\in K}$, we obtain

$$\begin{aligned}
\limsup_{k\to\infty,k\in K} \frac{\varphi(G(u^k + t_k d^k)) - \varphi(G(u^k))}{t_k} &\leq \limsup_{k\to\infty,k\in K} (\eta^k)^T \frac{G(u^k+t_kd^k)-G(u^k)}{t_k} \\
&\leq \max_{\eta^*\in\partial\varphi(G(u^*))} (\eta^*)^T \nabla G(u^*)^T d^* \\
&= \varphi'(G(u^*); \nabla G(u^*)^T d^*),
\end{aligned}$$

where the last inequality follows from the classical convex analysis [39]. Again, convexity of φ and continuous differentiability of G imply that

$$\varphi'(G(u^*); \nabla G(u^*)^T d^*) = \Theta'_{(\rho^g,\rho^h)}(u^*; d^*).$$

The desired result is proved. ■

We now establish the main global convergence result.

Theorem A.10 *Assume* $\lim_{k\to\infty,k\in K} u^k = u^*$. *Suppose the standing assumptions (C1)–(C3) are satisfied. Then* u^* *is both a critical point of the penalty function* $\Theta_{(\rho^g_*,\rho^h_*)}$ *and a KKT point of the nonlinear program (24).*

Proof. By the standing assumption (C3), $\rho_k = \rho_*$ for all sufficiently large k. This means that $\{\Theta_{(\rho^g_k,\rho^h_k)}(u^k)\}$ is a monotonically decreasing sequence when $k \geq k^*$ for some k^*. By Lemma A.8, passing to the subsequence, we may assume that

$$\lim_{k\to\infty,k\in K} d^k = d^*$$
$$\lim_{k\to\infty,k\in K} W_k = W^*.$$

We next aim to prove the following equality:

$$d^* = 0. \tag{30}$$

If the step length sequence $\{t_k\}_{k\in K}$ is bounded away from zero, by the line search rule, the monotonically decreasing property of $\{\Theta_{(\rho^g_k,\rho^h_k)}(u^k)\}$, then

$$\lim_{k\to\infty,k\in K} (d^k)^T W_k d^k = 0,$$

i.e.,

$$(d^*)^T W^* d^* = 0,$$

which implies (30) by the standing assumption (C2). Otherwise, by passing to the subsequence, we may assume that

$$\lim_{k\to\infty, k\in K} t_k = 0.$$

The line search rule implies that for any $k \in K$

$$-\sigma(d^k)^T W_k d^k \le \frac{\Theta_{(\rho_k^g,\rho_k^h)}(u^k + \frac{t_k}{\tau} d^k) - \Theta_{(\rho_k^g,\rho_k^h)}(u^k)}{t_k/\tau}.$$

Taking the super limit, we obtain

$$\begin{aligned}
&-\sigma(d^*)^T W^* d^* \\
&\quad \le \limsup_{k\to\infty, k\in K} \frac{\Theta_{(\rho_*^g,\rho_*^h)}(u^k + \frac{t_k}{\tau} d^k) - \Theta_{(\rho_*^g,\rho_*^h)}(u^k)}{t_k/\tau} \\
&\quad \le \Theta'_{(\rho_*^g,\rho_*^h)}(u^*; d^*) \\
&\quad \le -(d^*)^T W^* d^*,
\end{aligned}$$

where the second last inequality follows from Lemma A.9 and the last inequality follows from (iv) of Lemma A.8. The above inequality shows that

$$(1-\sigma)(d^*)^T W^* d^* \le 0,$$

which implies (30) by the standing assumption (C2). Taking the limit in the KKT conditions (28) with $u = u^k$ $(k \in K)$, we claim that (d^*, ξ^*) with $d^* = 0$ and some $\xi^* \in \Re^l$, is a KKT point of the modified QP (25) with $u = u^*$, $W = W^*$, and $\rho = \rho_*$. Since the modified QP is convex by the positive definiteness of W^*, (d^*, ξ^*) with $d^* = 0$ and some $\xi^* \in \Re^l$ is a solution of the modified QP (25) with $u = u^*$, $W = W^*$, and $\rho = \rho_*$. In view of (ii) of Proposition A.3 u^* is a critical point of the penalty function $\Theta_{(\rho_*^g,\rho_*^h)}$.

Since $\xi^* = 0$, u^* must also be a KKT point of (24). ■

We now turn to the case when the condition (C3) is not assumed, i.e., $\rho_k \to \infty$. To this end, we need to impose new conditions in addition to (C1) and (C2).

(C4) $\{u^k\}$ is bounded.

(C5) $h'(u^*)$ is of full row rank for any accumulation point u^* of $\{u^k\}$.

Lemma A.11 *Assume the standing conditions (C1), (C2), (C4) and (C5) hold. Then* $\{(d^k, \lambda_g^k, \lambda_h^k, \lambda_\xi^k)/\rho_{k-1}\}$ *is bounded.*

Proof. Since $\rho_{k-1}\tilde{e} = \lambda_g^k + \lambda_\xi^k$ and $\lambda_g^k \ge 0$, $\lambda_\xi^k \ge 0$ for any k, we only need to prove that $\{r_k\}$ is bounded where $r_k = \|(d^k, \lambda_h^k)/\rho_{k-1}\|$. Assume for contradiction that $\{r_k\}$ is unbounded, i.e., for some subset K, $\lim_{k\to\infty,k\in K} r_k = \infty$. We may assume that $\lim_{k\to\infty,k\in K} u^k = u^*$, $\lim_{k\to\infty,k\in K} W_k = W^*$ without loss of generality. Dividing the KKT conditions (28) at the kth iteration by $\rho_{k-1} r_k$ and passing to the limit, we obtain

$$\begin{aligned}
&W^*\tilde{d} + h'(u^*)^T \tilde{\lambda}_h = 0 \\
&h'(u^*)\tilde{d} = 0,
\end{aligned}$$

for some vectors $\tilde{d}$ and $\tilde{\lambda}_h$. These equations and the Conditions (C2) and (C5) show that $\tilde{d} = 0$ and $\tilde{\lambda}_h = 0$, which contradicts the assumption that

$$\|(\tilde{d}, \tilde{\lambda}_h)\| = \lim_{k\to\infty, k\in K} r_k/r_k = 1.$$

Therefore, $\{r_k\}$ is bounded and the desired result holds. ■

(C6) There exists d such that $h'(u^*)d + h(u^*) = 0$ and $g'(u^*)d + g(u^*) > 0$ for any accumulation point u^* of $\{u^k\}$.

For a feasible point u^*, assumptions (C5) and (C6) together are equivalent to the usual Mangasarian Fromovitz constraint qualification (MFCQ).

Lemma A.12 *Assume the standing conditions (C1), (C2), (C4), (C5) and (C6) hold. If $\{\rho_k\} \to \infty$ then*

(i) $\{(d^k, \xi^k)\}$ is bounded;

(ii) $\{\lambda^k/\rho_{k-1}\} \to 0$; and

(iii) $\xi^k = 0$ for all sufficiently large k.

Proof. (i) For a contradiction let $\{(u^k, W_k)\}_{k\in K}$ be a convergent subsequence with limit (u^*, W^*) such that $\{\|(d^k, \xi^k)\|\} \to \infty$. Let $\hat{d}$ be the vector such that $d = \hat{d}$ satisfies (C6). ¿From (C5) and (C6) respectively, $h'(u^*)$ has full rank and $h'(u^*)\hat{d} + h(u^*) = 0$; thus the classical calculus results of Lyusternik and Graves assure the existence of a convergent sequence $\{\hat{d}^k\}_{k\in K}$ with limit $\hat{d}$ satisfying $h'(u^k)\hat{d}^k + h(u^k) = 0$. By (C6), for large enough $k \in K$, $g'(u^k)\hat{d}^k + g(u^k) > 0$. Note that $\{(\hat{d}_k, 0)\}_{k\in K}$ is bounded by boundedness of $\{\hat{d}^k\}_{k\in K}$, and the pair $(\hat{d}^k, 0)$ is feasible for the QP at iteration $k \in K$, where 0 is the origin in $\Re^l$.

We now have

$$\begin{aligned} & \nabla f(u^k)^T \hat{d}^k + \frac{1}{2}(\hat{d}^k)^T W^k \hat{d}^k \\ \geq\ & \nabla f(u^k)^T d^k + \frac{1}{2}(d^k)^T W^k d^k + \rho_{k-1} \sum \xi_i^k \\ \geq\ & \nabla f(u^k)^T d^k + \frac{1}{2}(d^k)^T W^k d^k \end{aligned} \tag{31}$$

because first, (d^k, ξ^k) is optimal for the QP at iteration k, and second, $\xi^k \geq 0$. Since the quantity (31) is bounded for $k \in K$, it follows, by uniform positive definiteness of $\{W_k\}_{k\in K}$ and boundedness of $\{u_k\}_{k\in K}$, that $\{d^k\}_{k\in K}$ is bounded. Finally, it is obvious from optimality of (d^k, ξ^k) that $\xi^k = [-g'(u^k)d^k + g(u^k)]_+$ for each k, hence $\{\xi^k\}_{k\in K}$ is also bounded. This contradicts the unboundedness condition.

(ii) We have $\{d^k/\rho_{k-1}\} \to 0$ because $\{d^k\}$ is bounded from (i), and $\{\rho_k\} \to \infty$ by hypothesis. Adapting the proof of Lemma A.11 we get, for any limit point $(u^*, d^*, \xi^*, \hat{\lambda}_g, \hat{\lambda}_h)$ of $\{(u^k, d^k, \xi^k, \lambda_g^k/\rho_{k-1}, \lambda_h^k/\rho_{k-1})\}$, that

$$\begin{aligned} & 0 = -g'(u^*)^T \hat{\lambda}_g + h'(u^*)^T \hat{\lambda}_h \\ & 0 \leq g'(u^*)d^* + g(u^*) + \xi^* \quad \perp \quad \hat{\lambda}_g \geq 0 \\ & 0 = h'(u^*)d^* + h(u^*) \\ & 0 \leq \xi^*. \end{aligned}$$

Let $I = \{i : (\hat{\lambda}_g)_i > 0\}$ and observe that for $i \in I$, $0 = g_i'(u^*)d^* + g_i(u^*) + \xi_i^*$. Thus the first equation above gives

$$0 = -\sum_{i \in I} g_i'(u^*)(\hat{\lambda}_g)_i + h'(u^*)\hat{\lambda}_h \tag{32}$$

Now for $i \in I$, $g_i'(u^*)d^* + g_i(u^*) = -\xi_i^* \leq 0$. Also for the vector $\hat{d}$ satisfying (C6), as above, we have $g_i'(u^*)\hat{d} + g_i(u^*) > 0$ for any i, and $h'(u^*)\hat{d} + h(u^*) = 0$. Subtracting inequalities involving g_i and h respectively gives: $g_i'(u^*)(\hat{d} - d^*) > 0$ for $i \in I$, and $h'(u^*)(\hat{d} - d^*) = 0$, where $h'(u^*)$ has full rank by (C5). We have just verified that the MFCQ holds for the system $g_I'(u^*)(u - u^*) \geq 0$ and $h'(u^*)(u - u^*) = 0$ at $u = u^*$. It follows that the set of multipliers $(\hat{\lambda}_g)_I \geq 0$ and $\hat{\lambda}_h$ satisfying the Lagrangian equation (32) is bounded [14]; hence $((\hat{\lambda}_g)_I, \hat{\lambda}_h) = (0, 0)$ is the unique solution of this equation with $(\hat{\lambda}_g)_I \geq 0$.

Since $\{(\lambda_g^k/\rho_{k-1}, \lambda_h^k/\rho_{k-1})\}$ is already known to be bounded from Lemma A.11, it follows that this sequence converges to 0.

(iii) The KKT conditions for the QP solved at iteration k include $\rho_{k-1}\hat{e} = \lambda_g^k + \lambda_\xi^k$, hence using part (ii),

$$\frac{\lambda_\xi^k}{\rho_{k-1}} = \hat{e} - \frac{\lambda_g^k}{\rho_{k-1}} \to \hat{e}.$$

Thus λ_ξ^k is strictly positive for all large k. ¿From the previously mentioned KKT conditions, λ_ξ^k is complementary (orthogonal) to the nonnegative vector ξ^k, hence $\xi^k = 0$ for sufficiently large k. ∎

Theorem A.13 *Assume the standing conditions (C1), (C2), (C4), (C5) and (C6) hold. Then (C3) also holds.*

Proof. Assume (C3) fails, in which case we have $\{\rho_k\} \to \infty$. Applying Lemma A.12, part (ii) yields that $\tilde{\rho}_k = \rho_{k-1}$ for large enough k; and part (iii) yields that $\rho_k = \tilde{\rho}_k$ for large enough k. Thus $\rho_{k-1} = \rho_k$ for all large k, a contradiction. ∎

Remark. Theorem A.13 remains to hold under suitable conditions when the function h is smooth at all u^k but possibly nonsmooth at some accumulation points of $\{u^k\}$. To this end, the Conditions (C5) and (C6) need to be replaced by the Conditions (C5') and (C6') below:

(C5') V is of full row rank for any $V \in \partial h(u^*)$ and for any accumulation point u^* of $\{u^k\}$.

(C6') For any accumulation point u^* of $\{u^k\}$ and for any $V \in \partial h(u^*)$, there exists d such that

$$\begin{aligned} &Vd + h(u^*) = 0 \\ &g'(u^*)d + g(u^*) > 0. \end{aligned}$$

This generalized version of Theorem A.13 is useful in the context of MPEC.