Studies on Methods for Mathematical Programs with Equilibrium Constraints

Gui-Hua LIN

Submitted in partial fulfillment of
the requirement for the degree of
DOCTOR OF INFORMATICS
(Applied Mathematics and Physics)

Kyoto University
Kyoto 606-8501, Japan
December, 2003

Preface

Mathematical program with equilibrium constraints, abbreviated as MPEC, is a constrained optimization problem in which the essential constraints are defined by some parametric variational inequalities or a parametric complementarity system. This problem can be thought as a generalization of the so-called bilevel programming problem that is a mathematical program with optimization constraints. MPEC is also closely related to the well-known Stackelberg game. As a result, MPEC plays a very important role in many fields such as engineering design, economic equilibrium, multilevel game, and mathematical programming theory itself, and it has been receiving much attention in the recent optimization world.

On the other hand, MPEC is very difficult to deal with because, from the geometric point of view, its feasible region is not convex and not connected even in general, and in theory, its constraints fail to satisfy a standard constraint qualification such as the linear independence constraint qualification or the Mangasarian-Fromovitz constraint qualification at any feasible point. Therefore, the well-developed nonlinear programming theory cannot be applied to MPECs directly. There have been proposed several approaches such as sequential quadratic programming approach, implicit programming approach, penalty function approach, interior point method approach, and reformulation approach in the literature on MPECs.

Our main purpose is to develop more efficient methods for solving MPECs. Moreover, we notice that, in many practical problems, some elements may involve uncertain data, and hence we also pay great attention to the stochastic mathematical programs with equilibrium constraints (SMPECs). Thus, the thesis may be divided into two parts:

The first part consists of Chapters 2 to 5, in which we focus on the study of MPECs, particularly a special and important subclass — the mathematical programs with complementarity constraints (MPCCs). We first give some modified exact penalty results for nonlinear programs and MPECs in Chapter 2 and then, in Chapters 3 and 4, we propose two relaxation methods for MPCCs, one of which uses an expansive simplex instead of the nonnegative orthant involved in the complementarity constraints and the other suggests a scheme with bi-hyperbola approximation strategy. Some convergence results have been given for the proposed methods. In Chapter 5, we consider a hybrid approach with active index set identification for MPCCs. It has been shown that, unlike most existing methods, the hybrid approach may solve an MPCC in a finite number of iterations.

The second part includes Chapters 6–8, in which we deal with the SMPECs. We discuss two kinds of SMPECs: One is the lower-level wait-and-see model, in which the upper-level decision is made at once and a lower-level decision may be made after a random event is observed, and the other is the here-and-now model that requires us to make all decisions before a random event is observed. It has been shown that many decision problems can be formulated as SMPECs in practice. Several methods have been proposed in Chapters 6, 7, and 8, respectively. In particular, in Chapter 6, we suggest a smoothing implicit programming approach for both the lower-level wait-and-see decision model and the here-and-now decision model. Subsequently, we consider a special here-and-now problem in Chapters 7 and 8. We first give some equivalent reformulations of the problem and then, based on the reformulations, we propose some penalty methods and a regularization method in Chapters 7 and 8, respectively.

The importance of MPECs and SMPECs has been known in the optimization world and, particularly, the study of SMPECs is still in its initial stages. We hope that the results obtained in this thesis will be helpful to advance the study in the field.

Gui-Hua Lin
December, 2003

Contents

Chapter 1

Introduction

Mathematical program with equilibrium constraints (MPEC) is a constrained optimization problem whose constraints include some parametric variational inequalities or a parametric complementarity system. This problem plays a very important role in many fields such as engineering design, economic equilibrium, multilevel game, and mathematical programming theory itself, and it has been receiving much attention in the recent optimization world. In this chapter, we give a brief overview on MPECs. We also introduce some knowledge about the stochastic mathematical programs with equilibrium constraints (SMPECs).

1.1 Background on MPECs

MPEC is generally an optimization problem with two types of variables, an upper-level variable $x \in \Re^n$ and a lower-level variable $y \in \Re^m$:

$$\begin{aligned} \text{minimize} \quad & f(x, y) \\ \text{subject to} \quad & (x, y) \in Z, \\ & y \text{ solves } \mathrm{VI}(F(x, \cdot), C(x)). \end{aligned} \tag{1.1}$$

Here, Z is a subset of $\Re^{n+m}$, $f : \Re^{n+m} \to \Re, F : \Re^{n+m} \to \Re^m$, $C : \Re^n \to 2^{\Re^m}$ are mappings, and $\mathrm{VI}(F(x, \cdot), C(x))$ denotes the *variational inequality* problem defined by the pair $(F(x, \cdot), C(x))$; that is, y solves $\mathrm{VI}(F(x, \cdot), C(x))$ if and only if $y \in C(x)$ and

$$(v - y)^T F(x, y) \geq 0, \qquad \forall v \in C(x).$$

It is well-known [36] that, for a given variational inequality problem $\mathrm{VI}(G, Y)$, if the function G is the gradient mapping of a differentiable function $g : \Re^m \to \Re$ and the set Y is convex in $\Re^m$, then $\mathrm{VI}(G, Y)$ is just a restatement of the first-order necessary conditions of optimality for the optimization problem

$$\begin{array}{ll} \text{minimize} & g(y) \\ \text{subject to} & y \in Y. \end{array}$$

Therefore, MPEC (1.1) can be regarded as a generalization of the so-called bilevel programming problem that is a mathematical program with optimization constraints. Moreover, MPEC is also closely related to the well-known Stackelberg game, see [62, 66].

When $C(x) \equiv \Re^m_+$ for each x in problem (1.1), the parametric variational inequality constraints reduce to a parametric complementarity system and then problem (1.1) is equivalent to the following mathematical program with complementarity constraints (MPCC):

$$\begin{array}{ll} \text{minimize} & f(x, y) \\ \text{subject to} & (x, y) \in Z, \\ & y \geq 0, \;\; F(x, y) \geq 0, \\ & y^T F(x, y) = 0. \end{array} \tag{1.2}$$

On the other hand, if the set-valued function C in problem (1.1) is defined by

$$C(x) := \{y \in \Re^m | \;\; c(x, y) \leq 0\},$$

where $c : \Re^{n+m} \to \Re^s$ is continuously differentiable, then, under some suitable conditions, the variational inequality problem $\mathrm{VI}(F(x, \cdot), C(x))$ has an equivalent Karush-Kuhn-Tucker representation [68]:

$$\begin{array}{l} F(x, y) + \nabla_y c(x, y)\lambda = 0, \\ \lambda \geq 0, \quad c(x, y) \leq 0, \quad \lambda^T c(x, y) = 0, \end{array}$$

where λ is the Lagrange multiplier vector, and hence, problem (1.1) can be reformulated as a program like (1.2) under some conditions, see the monograph [62] for details. Hence, problem (1.2) constitutes an important subclass of MPECs. In this thesis, we particularly concentrate on this kind of MPECs.

As mentioned above, MPEC plays a very important role in many fields and it has been receiving more and more attention in the optimization world. For more details, we refer to the monographs [62, 66] and the references attached in the end of the thesis.

On the other hand, MPEC is very difficult to deal with because, from the geometric point of view, its feasible region is not convex and not connected even in general, and in theory, its constraints fail to satisfy a standard constraint qualification such as the linear independence constraint qualification or the Mangasarian-Fromovitz constraint qualification at any feasible point [17]. Therefore, the developed nonlinear programming theory may not be applied to MPEC class directly. At present, a natural and popular approach is try to find some suitable approximations of an MPEC so that the MPEC can be solved by solving a sequence of subproblems. Along this way, many methods have been proposed in the literature.

One family of the methods employs some kinds of smoothing or relaxation techniques. In particular, Facchinei et al. [24] and Fukushima and Pang [31] make use of the Fischer-Burmeister function to generate some smooth approximations of MPECs and subsequently, a similar scheme was presented by Scholtes [76].

Another family of the methods uses a penalty technique. For example, Huang et al [41] utilized the Fischer-Burmeister function to penalize the whole complementarity system, whereas Hu and Ralph [38] suggested a method to penalize the complementarity constraints only. Moreover, the works [39, 40, 63, 77] also belong to this family.

The other methods in the literature on MPECs include the sequential quadratic programming methods [29, 44, 62], implicit programming methods [12, 62], interior point methods [60, 62], and implementable active-set method [33]. In addition, nonsmooth methods for MPECs can be found in [66]. One purpose of the thesis is to develop more efficient methods for solving MPECs.

1.2 Background on SMPECs

Stochastic mathematical program with equilibrium constraints (SMPEC) can be formulated as follows:

$$\begin{array}{ll} \text{minimize} & E_\omega[f(x,y,\omega)] \\ \text{subject to} & (x,y)\in Z, \quad \omega\in\Omega, \\ & y \text{ solves VI}(F(x,\cdot,\omega),C(x,\omega)), \end{array} \tag{1.3}$$

where Z is a subset of $\Re^{n+m}$, Ω denotes the underlying sample space, E_ω means expectation with respect to the random variable $\omega\in\Omega$, and $f:\Re^{n+m}\times\Omega\to\Re, F:\Re^{n+m}\times\Omega\to\Re^m$, $C:\Re^n\times\Omega\to 2^{\Re^m}$ are mappings. Obviously, if Ω is a singleton, then problem (1.3) reduces to an ordinary MPEC, and so the SMPEC (1.3) can be thought as a generalization of the MPEC (1.1). Since an MPEC is already very hard to handle, so SMPECs may be more difficult to deal with because the number of random events is usually very large in practice.

The SMPEC (1.3) is also closely related to the so-called two-stage stochastic program with recourse [72]:

$$\begin{array}{ll} \text{minimize} & p(x) + E_\omega[Q(x,\omega)] \\ \text{subject to} & x \in X, \end{array} \tag{1.4}$$

where $p : \Re^n \to \Re$, $X \subseteq \Re^n$, and $Q : \Re^n \times \Omega \to \Re$ is defined by

$$Q(x,\omega) := \inf_{y \in Y(x,\omega)} g(y,\omega)$$

with $Y : \Re^n \times \Omega \to 2^{\Re^m}$ and $g : \Re^m \times \Omega \to \Re$. Many applications of problem (1.4) can be found in practice, especially in financial planning. For further details, see [3, 11, 15, 16, 79].

SMPEC was first discussed in [72], the main results of which are concerned with the existence of solutions, the convexity and directional differentiability of an implicit objective function, and links between SMPEC and bilevel models. Actually, there has been no effective algorithms suggested for solving SMPECs so far. In the second half of the thesis, we study SMPECs systematically. We discuss two kinds of SMPECs: One is the lower-level wait-and-see model, in which the upper-level decision is made at once and a lower-level decision may be made after a random event is observed, and the other is the here-and-now model that requires us to make all decisions before a random event is observed. See Chapter 6 for details.

1.3 Main Contributions

The purpose of the thesis is to develop efficient methods for solving MPECs and SMPECs. We may divide the thesis into two parts: The first part includes Chapters 2–5 in which we deal with the MPECs and the second part consists of Chapters 6–8 in which we study the SMPECs. Our main results can be summarized as follows.

In Chapter 2, we give some modified exact penalty results for nonlinear programs and MPECs. In particular, instead of the abstract subanalytic property and error bounds employed in [62], some of our results use a kind of convexity that is discussed in detail as well.

In Chapter 3, we propose a new relaxation method for MPCCs. Our method replaces the complementarity constraints by a variational inequality defined on an expansive simplex. It is well known that such a variational inequality problem can be represented by a finite number of inequalities. We remove some inequalities and obtain a standard smooth nonlinear program. We investigate the limiting behavior of the relaxed problems and obtain some exciting convergence results. In particular, some conditions assumed in the convergence theory are new and can be verified easily in practice.

In Chapter 4, we suggest another relaxation method with bi-hyperbola approximation for MPCCs. This method possesses similar properties to the regularization method [76] and the subproblems in the new method have less constraints.

In Chapter 5, we are devoted to develop some methods that enable us to compute a solution or a point with some kind of stationarity by solving a finite number of nonlinear programs. To this end, we apply an active set identification technique to some existing methods and present some hybrid algorithms. We show that, under some suitable assumptions, the algorithms indeed possess a finite termination property, unlike most existing methods that require to solve an infinite sequence of nonlinear programs. Further discussions and extensions are also included.

We study the SMPECs in the rest of the thesis.

In Chapter 6, we first introduce the problems and then, we show that many decision problems can be formulated as SMPECs in practice. We discuss both the lower-level wait-and-see decision model and the here-and-now decision model. For the lower-level wait-and-see model, we propose a smoothing implicit programming method and establish a comprehensive convergence theory. For the here-and-now decision problem, we apply a penalty technique and suggest a similar method. We show that the two methods possess similar convergence properties.

In Chapters 7 and 8, we consider a special here-and-now problem. We show that the stochastic linear complementarity problem may be formulated as this kind of SMPECs. We give some equivalent reformulations of the problem and then propose some penalty methods and a regularization method in Chapters 7 and 8, respectively.

1.4 Preliminaries

We will use the following notations and terminologies in this thesis.

1.4.1 Notations

Throughout, all vectors are thought as column vectors and ^{T} means the transpose operation. For $u \in \Re^s$, let $\|u\|$ and $\|u\|_1$ denote the norms defined by

$$\|u\| := \Big(\sum_{i=1}^{s} u_i^2\Big)^{1/2} \quad \text{and} \quad \|u\|_1 := \sum_{i=1}^{s} |u_i|,$$

respectively. For a nonempty closed set $V \subseteq \Re^s$, we denote

$$\mathrm{dist}(u, V) := \min_{v \in V} \|u - v\|$$

and

$$\Pi_V(u) := \Big\{ v \in V \;\Big|\; \|u - v\| = \mathrm{dist}(u, V) \Big\}.$$

Moreover, $B(u, \delta)$ stands for the closed ball $\{v \in \Re^s \mid \ \|u - v\| \leq \delta\}$ and $\Re^s_+$ denotes the nonnegative orthant in $\Re^s$. For a real scalar a, we denote $(a)_+ := \max\{0, a\}$. For two vectors u and v in $\Re^s$, $u \perp v$ means $u^T v = 0$ and both $\min(u, v)$ and $\max(u, v)$ are understood to be taken componentwise, i.e.,

$$\begin{aligned} \min(u, v) &:= (\min\{u_1, v_1\}, \cdots, \min\{u_s, v_s\})^T, \\ \max(u, v) &:= (\max\{u_1, v_1\}, \cdots, \max\{u_s, v_s\})^T \end{aligned}$$

For a given function $G : \Re^s \to \Re^{s'}$ and a vector $u \in \Re^s$, $\nabla G(u)$ is the transposed Jacobian of G at u, whereas for a real valued function $g : \Re^s \to \Re$, $\nabla g(u)$ denotes the gradient vector of g at u. Moreover,

$$\mathcal{I}_G(u) := \{i \mid \ G_i(u) = 0\}$$

stands for the active index set of G at u. In addition, e_i denotes the unit vector with the ith element to be 1; I and O denote the identity matrix and the zero matrix with suitable dimension, respectively.

1.4.2 Terminologies

We first recall some basic concepts for the standard nonlinear programming problem

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & c_i(z) \leq 0, \quad i = 1, \cdots, l, \\ & c_i(z) = 0, \quad i = l+1, \cdots, s, \end{array} \tag{1.5}$$

where $f : \Re^n \to \Re$ and $c : \Re^n \to \Re^s$ are twice continuously differentiable.

Definition 1.1 The *linear independence constraint qualification* (LICQ) is said to hold at a feasible point z of problem (1.5) if the set of vectors $\{\nabla c_i(z) \mid \ i \in \mathcal{I}_c(z)\}$ is linearly independent.

Definition 1.2 We say z to be a *stationary* point of problem (1.5) if it is feasible to (1.5) and there exists a Lagrange multiplier vector $\lambda \in \Re^s$ such that

$$\begin{aligned} &\nabla f(z) + \nabla c(z)\lambda = 0, \\ &\lambda_i \geq 0,\ \lambda_i c_i(z) = 0, \quad i = 1, \cdots, l. \end{aligned}$$

Definition 1.3 Let z be a stationary point of problem (1.5) and λ be a Lagrange multiplier vector corresponding to z. We say the *weak second-order necessary condition* (WSONC) holds at z if we have

$$d^T\Big(\nabla^2 f(z) + \sum_{i=1}^{s} \lambda_i \nabla^2 c_i(z)\Big)d \geq 0$$

for any $d \in \mathcal{T}(z) := \Big\{ d \in \Re^n \ \Big|\ d^T \nabla c_i(z) = 0,\ \ \forall\ i \in \mathcal{I}_c(z) \Big\}$.

We next consider the mathematical program with complementarity constraints (MPCC)

$$\begin{aligned} \text{minimize} \quad & f(z) \\ \text{subject to} \quad & g(z) \leq 0,\ h(z) = 0 \\ & G(z) \geq 0,\ H(z) \geq 0 \\ & G(z)^T H(z) = 0, \end{aligned} \tag{1.6}$$

where $f : \Re^n \to \Re, g : \Re^n \to \Re^p, h : \Re^n \to \Re^q$, and $G, H : \Re^n \to \Re^m$ are all twice continuously differentiable functions. Let $\mathcal{F}$ denote the feasible region of the above problem.

Definition 1.4 The MPEC-*linear independence constraint qualification* (MPEC-LICQ) is said to hold at $\bar{z} \in \mathcal{F}$ if the set of vectors

$$\Big\{ \nabla g_l(\bar{z}), \nabla h_r(\bar{z}), \nabla G_i(\bar{z}), \nabla H_j(\bar{z}) \mid \\ l \in \mathcal{I}_g(\bar{z}),\ r = 1, \cdots, q,\ i \in \mathcal{I}_G(\bar{z}),\ j \in \mathcal{I}_H(\bar{z}) \Big\}$$

is linearly independent.

This condition is not particularly stringent [78] and has been assumed often in the literature on MPCCs [31, 38, 41, 54, 76]. Note that this definition is different from the standard definition of LICQ in nonlinear programming theory that would require the gradient of the function $G(z)^T H(z)$ be linearly independent of the above vectors, which cannot happen in any case actually.

In the study of MPCCs, there are several kinds of stationarity defined for problem (1.6) [75].

Definition 1.5 We say $\bar{z} \in \mathcal{F}$ to be a *Bouligand or B-stationary* point of problem (1.6) if it satisfies

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

where

$$\mathcal{T}(\bar{z}, \mathcal{F}) := \left\{ d \in \Re^n \;\middle|\; t_k(z^k - \bar{z}) \to d, z^k \to \bar{z}, z^k \in \mathcal{F}, t_k \geq 0, k = 1, 2, \cdots \right\}$$

stands for the tangent cone of $\mathcal{F}$ at $\bar{z}$.

Definition 1.6 (1) $\bar{z} \in \mathcal{F}$ is called *weakly stationary* to problem (1.6) if there exist multiplier vectors $\bar{\lambda} \in \Re^p, \bar{\mu} \in \Re^q$, and $\bar{u}, \bar{v} \in \Re^m$ such that

$$\nabla f(\bar{z}) + \nabla g(\bar{z})\bar{\lambda} + \nabla h(\bar{z})\bar{\mu} - \nabla G(\bar{z})\bar{u} - \nabla H(\bar{z})\bar{v} = 0, \quad (1.7)$$

$$\bar{\lambda} \geq 0, \qquad \bar{\lambda}^T g(\bar{z}) = 0, \quad (1.8)$$

$$\bar{u}_i = 0, \quad i \notin \mathcal{I}_G(\bar{z}), \quad (1.9)$$

$$\bar{v}_i = 0, \quad i \notin \mathcal{I}_H(\bar{z}). \quad (1.10)$$

(2) $\bar{z} \in \mathcal{F}$ is called a *Clarke or C-stationary* point of problem (1.6) if there exist multiplier vectors $\bar{\lambda} \in \Re^p, \bar{\mu} \in \Re^q$, and $\bar{u}, \bar{v} \in \Re^m$ such that (1.7)–(1.10) hold with

$$\bar{u}_i \bar{v}_i \geq 0, \quad i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})$$

and we say $\bar{z}$ is *Mordukhovich or M-stationary* to problem (1.6) if, furthermore, either $\bar{u}_i > 0,\ \bar{v}_i > 0$ or $\bar{u}_i \bar{v}_i = 0$ for all $i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})$.

(3) $\bar{z} \in \mathcal{F}$ is called a *strongly or S-stationary* point of problem (1.6) if there exist multiplier vectors $\bar{\lambda}, \bar{\mu}, \bar{u}$, and $\bar{v}$ such that (1.7)–(1.10) hold with

$$\bar{u}_i \geq 0, \quad \bar{v}_i \geq 0, \quad i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}).$$

It is well-known [75] that, if the MPEC-LICQ holds at $\bar{z}$, B-stationarity is equivalent to S-stationarity.

Definition 1.7 A weakly stationary point $\bar{z} \in \mathcal{F}$ of problem (1.6) is said to satisfy the *upper level strict complementarity* (ULSC) condition if there exist multiplier vectors $\bar{\lambda}, \bar{\mu}, \bar{u}$, and $\bar{v}$ satisfying (1.7)–(1.10) and

$$\bar{u}_i \bar{v}_i \neq 0, \quad i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}).$$

The ULSC condition is clearly weaker than the so-called *lower level strict complementarity* (LLSC) condition (which means $\mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}) = \emptyset$ and in this case, $\bar{z}$ is also said to be *nondegenerate*). Moreover, it is obvious that any M-stationary point of problem (1.6) satisfying the upper level strict complementarity condition must be a B-stationary point.

Chapter 2

Exact Penalty Results for Nonlinear Programs and MPECs

Because of the presence of variational inequality or complementarity constraints, MPEC has such an intrinsic feature that its feasible region is nonconvex or nonsmooth in general and hence it is very difficult to handle. At present, a popular approach is to reformulate an MPEC as a standard nonlinear program. In this respect, penalty functions have provided a powerful approach, both as a theoretical tool and as a computational vehicle. Recently, based on the study of subanalytic optimization problems and with the help of the theory of error bounds, some exact penalty results for nonlinear programs and MPECs were proved by Luo, Pang, and Ralph [62]. In this chapter, we will show that those results remain valid under some other mild conditions. In particular, instead of the subanalytic property and error bounds, which are somewhat abstract and difficult to verify in practice, some of our results use a property called strong convexity with order σ, which is a generalization of the ordinary strong convexity [49] and will be discussed in detail.

2.1 Preliminaries

The following definitions will be used later on.

Definition 2.1 (See [62]) A set $X \subseteq \Re^n$ is said to be *subanalytic* if for any $u \in \Re^n$, there exist a neighborhood U of u and a bounded set $Z \subseteq \Re^{n+p}$ with some nonnegative integer p such that

(a) for any $v \in \Re^{n+p}$, there exist a neighborhood V of v and a finite family $\{Z_{ij} \mid 1 \leq$

$i \le l, 1 \le j \le q\}$ of sets $Z_{ij} = \{z \in V \mid f_{ij}(z) = 0\}$ or $\{z \in V \mid f_{ij}(z) < 0\}$ defined for some real analytic functions f_{ij} on V such that

$$Z \cap V = \bigcup_{i=1}^{l} \bigcap_{j=1}^{q} Z_{ij};$$

(b) $X \cap U = \{x \in \Re^n \mid (x, y) \in Z \text{ for some } y \in \Re^p\}$.

A function $f : \Re^n \to \Re$ is said to be *subanalytic* if its graph is subanalytic.

The class of subanalytic functions is broader than the class of analytic functions and is employed by many papers, although it is somewhat abstract. For more details, we refer the reader to [2, 20, 61, 62].

Definition 2.2 (See [64]) Let $0 < p \le 1$ be a constant and $G : \Re^n \to \Re^m$ be a mapping. We say G to be *Hölder continuous* with order p on $X \subseteq \Re^n$, if there exists a constant L such that

$$\|G(x) - G(y)\| \le L\|x - y\|^p, \qquad \forall x, y \in X. \tag{2.1}$$

This concept is a generalization of the Lipschitz continuity, which is, by definition, Hölder continuity with order $p = 1$. Note that Hölder continuity makes sense only when $0 < p \le 1$. In fact, when $p > 1$, condition (2.1) implies that all directional derivatives of G at any interior point are zero and so G is quite trivial. In addition, for $0 < p \ne p' \le 1$, Hölder continuous functions with order p and those with order p' constitute different classes of functions. For example, the function

$$G(x) := \|x\|^{\frac{1}{2}}, \qquad \forall x \in \Re^n$$

is Hölder continuous with order $p = \frac{1}{2}$ on $\Re^n$ and not Lipschitz continuous on $\Re^n$.

Definition 2.3 A function $f : \Re^n \to \Re$ is said to be *strongly convex* with order $\sigma > 0$ on a convex set $X \subseteq \Re^n$, if there exists a constant $c > 0$ such that

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y) - ct(1-t)\|x - y\|^\sigma \tag{2.2}$$

for any $x, y \in X$ and any $t \in [0, 1]$.

When $\sigma = 2$, this property reduces to the strong convexity in the ordinary sense [49]. But if $\sigma \ne 2$, they are different. For example, we can see from the results given in Section 2.4 that the function $f(x) = x^4$ is strongly convex with order 4 and not strongly convex (with order 2) on $\Re$.

2.2 Penalty Results for Nonlinear Programs

Consider the following nonlinear program:

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & x \in X, \\ & g(x) \leq 0, \ \ h(x) = 0, \end{array} \tag{2.3}$$

where $\theta : \Re^n \to \Re, g : \Re^n \to \Re^m$, and $h : \Re^n \to \Re^l$ are all continuous functions and $X \subseteq \Re^n$ is a nonempty closed set. Let W denote the feasible region of (2.3) and

$$r(x) := \sum_{i=1}^{m} (g_i(x))_+ + \sum_{j=1}^{l} |h_j(x)|$$

be the residual for the constraints in (2.3) at $x \in X$. Then the function r may be used as a penalty function for problem (2.3). The following theorem is shown in [62]:

Theorem 2.1 *Let $X \subseteq \Re^n$ be a compact subanalytic set, θ be Lipschitz continuous on X, and g_i, h_j be continuous subanalytic. Suppose problem (2.3) is feasible. Then there exist positive constants α^* and γ^* such that for $\alpha \geq \alpha^*$ and $\gamma \geq \gamma^*$, problem (2.3) is equivalent to*

$$\begin{array}{ll} \text{minimize} & \theta(x) + \alpha r(x)^{1/\gamma} \\ \text{subject to} & x \in X \end{array} \tag{2.4}$$

in the sense that x^* solves (2.3) if and only if it solves (2.4).

Moreover, the following result will be used:

Theorem 2.2 (Lojasiewicz Inequality [2]) *Let $\phi, \psi : S \to \Re$ be continuous subanalytic and $S \subseteq \Re^n$ be compact subanalytic. If $\phi^{-1}(0) \subseteq \psi^{-1}(0)$, then there exist constants $\rho > 0$ and $N^* > 0$ such that*

$$\rho|\psi(x)|^{N^*} \leq |\phi(x)|, \qquad \forall x \in S. \tag{2.5}$$

Now we give our penalty results for problem (2.3). First of all, we define a new function. Suppose that problem (2.3) is feasible, i.e., $W \neq \emptyset$. Then we can take a vector $d \in W$ and define a function θ_d on X by

$$\theta_d(x) := \theta(x_d), \qquad \forall x \in X,$$

where $x_d := (1-t_d)x + t_d d$ and t_d is the smallest number $t \in [0, 1]$ such that $(1-t)x + td \in W$.

Theorem 2.3 *Suppose that X, g, h are the same as in Theorem 2.1, problem (2.3) is feasible, and the function $(\theta - \theta_d)$ is continuous subanalytic for some $d \in W$. Then the conclusion of Theorem 2.1 remains valid.*

Proof: Let $r|_X$ denote the restriction of r on X. Noticing that both r and $(\theta - \theta_d)$ are continuous subanalytic and

$$(r|_X)^{-1}(0) = W \subseteq (\theta - \theta_d)^{-1}(0),$$

we have from Theorem 2.2 that there exist constants $\rho > 0$ and $N^* > 0$ such that

$$\rho|\theta(x) - \theta(x_d)|^{N^*} \leq r(x), \qquad \forall x \in X. \tag{2.6}$$

Let

$$\mu = \max\{1, \max_{x \in X} r(x)\}, \quad \alpha^* > (\mu/\rho)^{1/N^*}, \quad \gamma^* = N^*,$$

and

$$\alpha \geq \alpha^*, \quad \gamma \geq \gamma^*.$$

(a) Assume that $\bar{x}$ solves problem (2.3). Then for any $x \in X$, we have

$$\begin{aligned} \theta(x) + \alpha r(x)^{1/\gamma} &= \theta(x_d) + (\theta(x) - \theta(x_d)) + \alpha r(x)^{1/N^*} r(x)^{1/\gamma - 1/N^*} \\ &\geq \theta(\bar{x}) - |\theta(x) - \theta(x_d)| + \alpha\rho^{1/N^*}|\theta(x) - \theta(x_d)|\mu^{1/\gamma - 1/N^*} \\ &\geq \theta(\bar{x}) + (\alpha\rho^{1/N^*}\mu^{-1/N^*} - 1)|\theta(x) - \theta(x_d)| \\ &\geq \theta(\bar{x}) \\ &= \theta(\bar{x}) + \alpha r(\bar{x})^{1/\gamma}. \end{aligned}$$

Therefore, $\bar{x}$ is a global optimal solution of problem (2.4).

(b) If $\bar{x}$ solves (2.4), we can claim that $\bar{x}$ is an optimal solution of problem (2.3). In fact, since W is compact and problem (2.3) is feasible, it has an optimal solution, denoted by $\tilde{x}$. In a way similar to (a), we have

$$\begin{aligned} \theta(\tilde{x}) &= \theta(\tilde{x}) + \alpha r(\tilde{x})^{1/\gamma} \\ &\geq \theta(\bar{x}) + \alpha r(\bar{x})^{1/\gamma} \\ &\geq \theta(\tilde{x}) + (\alpha\rho^{1/N^*}\mu^{-1/N^*} - 1)|\theta(\bar{x}) - \theta(\bar{x}_d)| \\ &\geq \theta(\tilde{x}). \end{aligned}$$

This implies $\theta(\bar{x}) = \theta(\bar{x}_d)$ and then

$$\theta(\bar{x}) = \theta(\bar{x}_d) \geq \theta(\tilde{x}) \geq \theta(\bar{x}) + \alpha r(\bar{x})^{1/\gamma},$$

where the first inequality holds because $\tilde{x}$ solves (2.3) and $\bar{x}_d$ is feasible to (2.3). Hence, we have $r(\bar{x}) = 0$ and $\theta(\bar{x}) = \theta(\tilde{x})$. The former implies $\bar{x} \in W$ and so $\bar{x}$ is an optimal solution to problem (2.3). This completes the proof. ∎

The new condition given in Theorem 2.3 may be satisfied by choosing d appropriately even if W is not convex and θ is not Lipschitz continuous, as the next example shows.

Example 2.1 Consider the following problem:

$$\begin{array}{ll} \text{minimize} & \theta(x) := \sin^2(3x^{\frac{1}{3}}) \\ \text{subject to} & x \in [0, \pi^3],\ \cos(3x^{\frac{1}{3}}) \leq 0. \end{array}$$

Then the feasible region is given by

$$W = \left[\frac{\pi^3}{216}, \frac{\pi^3}{8}\right] \bigcup \left[\frac{125\pi^3}{216}, \pi^3\right],$$

which is nonconvex. We note that θ is not Lipschitz continuous on $[0, \pi^3]$, which means the conditions of Theorem 2.1 are not satisfied for this problem. However, we can show that the assumptions of Theorem 2.3 hold. In fact, for $d = \frac{\pi^3}{10}$, the function

$$\theta_d(x) = \begin{cases} 1, & x \in [0, \frac{\pi^3}{216}) \\ \sin^2(3x^{\frac{1}{3}}), & x \in [\frac{\pi^3}{216}, \frac{\pi^3}{8}] \\ 1, & x \in (\frac{\pi^3}{8}, \frac{125\pi^3}{216}) \\ \sin^2(3x^{\frac{1}{3}}), & x \in [\frac{125\pi^3}{216}, \pi^3] \end{cases}$$

is continuous and piecewise smooth and so it is continuous subanalytic on $[0, \pi^3]$.

We next consider another kind of error bounds for problem (2.3), which is different from (2.6). We say that a function $u : X \to [0, \infty)$ provides an *error bound* of order $\nu > 0$ on W, if there exists a positive constant β such that

$$u(x) \geq \beta\Big(\text{dist}(x, W)\Big)^{\nu}, \quad \forall x \in X.$$

For more details of error bounds, we refer the reader to [64, 69] and the references therein.

Theorem 2.4 *Let X be a closed subset of $\Re^n$, g and h be continuous on X, and θ be Hölder continuous with order $p > 0$ and Hölder constant L on X. Assume that $r(x)$ provides an error bound of order $\nu > 0$ on W with the corresponding constant β and* *suppose that problem (2.3) is feasible. Then problem (2.3) has the same solution set as the problem*

$$\begin{array}{ll} \text{minimize} & \theta(x) + \alpha r(x)^{N^*} \\ \text{subject to} & x \in X, \end{array} \tag{2.7}$$

where $N^* := \frac{p}{\nu}$ *and* $\alpha > L\beta^{-N^*}$.

Proof: By the assumption of the theorem, we have

$$r(x) \geq \beta\Big(\text{dist}(x, W)\Big)^{\nu}, \qquad \forall x \in X. \tag{2.8}$$

(a) If $\bar{x}$ solves problem (2.3), then for any $x \in X$, we have from (2.8) and the Hölder continuity of θ that

$$\begin{aligned} \theta(x) + \alpha r(x)^{N^*} &= \theta(z) + (\theta(x) - \theta(z)) + \alpha r(x)^{p/\nu} \\ &\geq \theta(\bar{x}) + (\alpha\beta^{p/\nu} - L)\Big(\text{dist}(x, W)\Big)^{p} \\ &\geq \theta(\bar{x}) \\ &= \theta(\bar{x}) + \alpha r(\bar{x})^{N^*}, \end{aligned}$$

where $z \in \Pi_W(x)$. Therefore, $\bar{x}$ is a global optimal solution of problem (2.7).

(b) Let $\bar{x} \in X$ be a solution of problem (2.7). Then for any $x \in W$,

$$\theta(\bar{x}) + \alpha r(\bar{x})^{N^*} \leq \theta(x) + \alpha r(x)^{N^*} = \theta(x). \tag{2.9}$$

Let $t := \inf_{x \in W} \theta(x)$. Then for any $\varepsilon > 0$, we can find an $x_\varepsilon \in W$ such that $\theta(x_\varepsilon) \leq t + \varepsilon$. By (2.8), (2.9), and the Hölder continuity of θ, we have

$$\begin{aligned} t + \varepsilon &\geq \theta(x_\varepsilon) \\ &\geq \theta(\bar{x}) + \alpha r(\bar{x})^{N^*} \\ &= \theta(\bar{z}) + (\theta(\bar{x}) - \theta(\bar{z})) + \alpha r(\bar{x})^{N^*} \\ &\geq t + (\alpha\beta^{p/\nu} - L)\|\bar{x} - \bar{z}\|^{p}, \end{aligned}$$

where $\bar{z} \in \Pi_W(\bar{x})$. Therefore,

$$\|\bar{x} - \bar{z}\|^{p} \leq (\alpha\beta^{p/\nu} - L)^{-1}\varepsilon$$

for any $\varepsilon > 0$ and so $\bar{x} = \bar{z} \in W$. Therefore, (2.9) becomes

$$\theta(\bar{x}) \leq \theta(x), \qquad \forall x \in W,$$

i.e., $\bar{x}$ solves problem (2.3). This completes the proof. ∎

The set X need not be compact and the functions g and h need not be subanalytic in the last theorem, which are in contrast with Theorems 2.1 and 2.3. If X is compact and g, h are subanalytic, as in Theorems 2.1 and 2.3, the exponent of the penalty term can be chosen elastically. This result is stated in the following theorem, whose proof is omitted here.

Theorem 2.5 *Assume that X, g, and h are the same as in Theorem 2.1, θ is Hölder continuous on X, and problem (2.3) is feasible. Then the conclusion of Theorem 2.1 remains true.*

Now we consider the special case of problem (2.3):

$$\begin{array}{ll} \text{minimize} & \theta(x) \\ \text{subject to} & x \in X, \\ & g(x) \leq 0. \end{array} \tag{2.10}$$

We will show some new penalty results for problem (2.10) which will be applied to the mathematical program with a nonlinear complementarity system in the next section. In the rest of this section, we let φ denote the function defined by

$$\varphi(x) := \max_{1 \leq i \leq m} g_i(x).$$

In general, condition (2.8) is difficult to verify in practice. The proof of the following theorem indicates that it holds when X is convex and φ is strongly convex with order σ on X.

Theorem 2.6 *Assume that $X \subseteq \Re^n$ is a closed convex set, θ is Hölder continuous with order $p > 0$ and Hölder constant L on X, and φ is strongly convex with order $\sigma > 0$ and the corresponding constant c on X. Suppose that problem (2.10) is feasible. Then problem (2.10) has the same solution set as problem (2.7) with*

$$r(x) := \sum_{i=1}^{m} \Big(g_i(x)\Big)_+, \quad N^* := \frac{p}{\sigma}, \quad \alpha > L\Big(\frac{c}{2}\Big)^{-N^*}.$$

Proof: By Theorem 2.4 and its proof, it is enough to prove that (2.8) holds with $\beta := \frac{c}{2}$ and $\nu := \sigma$ for any $x \in X$. In fact, assume that $\varphi(x) > 0$ and $\varphi(z) = 0$, where $z \in \Pi_S(x)$ with $S := \{x \in X \mid \varphi(x) \leq 0\}$. Since φ is strongly convex with order σ and constant c on X, it follows from (2.2) that

$$\varphi\Big(\frac{x+z}{2}\Big) \leq \frac{1}{2}\varphi(x) - \frac{c}{4}\|x - z\|^\sigma.$$

Note that $\varphi(\frac{x+z}{2}) > 0$. (Otherwise, since $\frac{x+z}{2} \in X$, this will contradict $z \in \Pi_S(x)$.) In consequence,

$$\frac{c}{2}\|x - z\|^\sigma \le \varphi(x) = (\varphi(x))_+ \le r(x),$$

i.e., (2.8) holds with $\beta = \frac{c}{2}$ and $\nu = \sigma$. This completes the proof. ∎

Note that it is easy to verify that if each g_i is strongly convex with order σ, then the function φ is also strongly convex with order σ. We also have the following result.

Theorem 2.7 *Assume that $X \subseteq \Re^n$ is compact and convex and the other conditions are the same as in Theorem 2.6. Let $\gamma \ge \frac{\sigma}{p}$ and $\alpha > L\left(\frac{c}{2}\right)^{-p/\sigma}$. Then problem (2.10) has the same solution set as the problem*

$$\begin{array}{ll} \text{minimize} & \theta(x) + \alpha r(x)^{1/\gamma} \\ \text{subject to} & x \in X. \end{array}$$

2.3 Penalty Results for MPECs

Consider the following mathematical program with equilibrium constraints (MPEC):

$$\begin{array}{ll} \text{minimize} & f(x, y) \\ \text{subject to} & (x, y) \in Z, \\ & y \text{ solves VI}(F(x, \cdot), C(x)), \end{array} \tag{2.11}$$

where $f : \Re^{n+m} \to \Re, F : \Re^{n+m} \to \Re^m, Z \subseteq \Re^{n+m}$, $C : \Re^n \to 2^{\Re^m}$ is defined by a continuously differentiable function $g : \Re^{n+m} \to \Re^l$ as

$$C(x) := \{y \in \Re^m \mid \ g(x, y) \le 0\}.$$

Let $\mathcal{F}$ denote the feasible region of problem (2.11), which is assumed to be nonempty. If F is continuous, $g_i(x, \cdot)$ is convex for all $x \in X$, where

$$X := \{x \in \Re^n \mid \ (x, y) \in Z \ \text{ for some } \ y \in \Re^m\},$$

$\nabla_y g_i(x, y)$ exists and is continuous at every (x, y) in an open set containing $\mathcal{F}$ for each $i = 1, \cdots, l$, Z is compact, and the constraint qualification called SBCQ [62] holds on

$\mathcal{F}$, then problem (11) is equivalent to the following mathematical program for some $\delta > 0$ ([62], Theorem 1.3.5):

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & (x,y,\lambda) \in Z \times (B(0,\delta) \cap \Re^l_+), \\ & F(x,y) + \sum_{i=1}^{l} \lambda_i \nabla_y g_i(x,y) = 0, \\ & g(x,y) \le 0,\ \lambda^T g(x,y) = 0. \end{array} \tag{2.12}$$

Roughly speaking, the SBCQ means that for any $(x,y) \in \mathcal{F}$, problem (2.12) is feasible and for a bounded subset of $\mathcal{F}$, the corresponding set of Lagrange multipliers is also bounded. Let

$$W := \{(x,y) \in \Re^{n+m} \mid\ (x,y,\lambda) \text{ satisfies the constraints of (2.12) for some } \lambda\}.$$

This set is nonempty if, under the SBCQ, $\mathcal{F}$ is nonempty. We choose some $d \in W$ and define the function f_d in a way similar to the definition of θ_d in the previous section. Then, comparing (2.12) with (2.3) and applying Theorems 2.3 and 2.5, we obtain the following result directly:

Theorem 2.8 *Let $F, g_i, \nabla_y g_i$ be continuous subanalytic and Z be compact subanalytic. Let f be Hölder continuous with order p on Z or $(f - f_d)$ be continuous subanalytic for some $d \in W$. Furthermore, assume that each $g_i(x,\cdot)$ is convex for all $x \in X$ and the SBCQ holds on $\mathcal{F}$. Then there exist constants $\delta > 0, \alpha^* > 0$, and $\gamma^* > 0$ such that for any $\alpha \ge \alpha^*$ and $\gamma \ge \gamma^*$, problem (2.11) is equivalent to the problem*

$$\begin{array}{ll} \text{minimize} & f(x,y) + \alpha r(x,y,\lambda)^{1/\gamma} \\ \text{subject to} & (x,y,\lambda) \in Z \times (B(0,\delta) \cap \Re^l_+), \end{array} \tag{2.13}$$

where

$$r(x,y,\lambda) := \|F(x,y) + \sum_{i=1}^{l} \lambda_i \nabla_y g_i(x,y)\|_1 + \sum_{i=1}^{l} ((g_i(x,y))_+ + \lambda_i |g_i(x,y)|),$$

in the sense that (x^*,y^*) solves (2.11) if and only if (x^*,y^*,λ^*) solves (2.13) for some $\lambda^* \in \Re^l_+$.

Now we consider a special class of MPECs:

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & (x,y) \in Z, \\ & y \ge 0,\ F(x,y) \ge 0, \\ & y^T F(x,y) = 0, \end{array} \tag{2.14}$$

i.e., the mathematical programs with complementarity constraints. Let S denote the feasible region of problem (2.14), $Z_1 := Z \cap (\Re^n \times \Re^m_+)$,

$$r(x,y) := \sum_{i=1}^{m} (-F_i(x,y))_+ + |y^T F(x,y)|, \tag{2.15}$$

and

$$\psi(x,y) := \min\Big\{ \min_{1\le i\le m} F_i(x,y), -y^T F(x,y) \Big\}.$$

In a way similar to Theorems 2.4 and 2.6, we can show the following results.

Theorem 2.9 *Assume that Z is a closed subset of $\Re^{n+m}$, f is Hölder continuous with order p and Hölder constant L on Z_1, and F is continuous on Z_1. Assume that $r(x,y)$ defined by (2.15) provides an error bound of order $\nu > 0$ with the corresponding constant β on S and problem (2.14) is feasible. Then problem (2.14) has the same solution set as the problem*

$$\begin{array}{ll} \text{minimize} & f(x,y) + \alpha r(x,y)^{N^*} \\ \text{subject to} & (x,y) \in Z_1, \end{array} \tag{2.16}$$

where $N^* := \frac{p}{\nu}$ and $\alpha > L\beta^{-N^*}$.

Theorem 2.10 *Assume that F, f are the same as in Theorem 2.9 and Z is closed and convex. Suppose that problem (2.14) is feasible. If the function $(-\psi)$ is strongly convex with order σ and the corresponding constant c on Z, then problem (2.14) is equivalent to problem (2.16) with $N^* := \frac{p}{\sigma}$ and $\alpha > L\left(\frac{c}{2}\right)^{-N^*}$ in the sense that (x^*, y^*) solves (2.14) if and only if it solves (2.16).*

2.4 Some Properties Related to Strong Convexity

For the strong convexity employed in Theorems 2.6–2.7 and 2.10, we have the following results:

Theorem 2.11 *If each $f_i, i = 1, \cdots, m$, is strongly convex with order σ on a convex set X, then $\sum_{i=1}^m t_i f_i$ and $\max_{1\le i\le m} f_i$ are also strongly convex with order σ on X, where $t_i > 0$, $i = 1, \cdots, m$.*

Proof: Immediate from Definition 2.3. ∎

Theorem 2.12 *Suppose that* $X \subseteq \Re^n$ *is convex and* $f : \Re^n \to \Re$ *is continuously differentiable on an open set containing* X*. Then* f *is strongly convex with order* σ *on* X *if and only if there exists a constant* $c > 0$ *such that*

$$f(y) \geq f(x) + (y-x)^T \nabla f(x) + c\|x-y\|^\sigma, \qquad \forall x, y \in X. \tag{2.17}$$

Proof: Assume that f is strongly convex with order σ on X and c is a constant that appears in (2.2). Then for any x, $y \in X$ and $t \in (0,1)$, we have

$$\begin{aligned} f(y) - f(x) &\geq \frac{1}{t}(f(ty + (1-t)x) - f(x)) + c(1-t)\|x-y\|^\sigma \\ &= (y-x)^T \nabla f(x + \xi(y-x)) + c(1-t)\|x-y\|^\sigma \end{aligned}$$

for some $\xi \in (0,t)$. Letting $t \to 0$, we have (2.17) from the continuity of ∇f.

Conversely, suppose (2.17) holds for some $c > 0$. For any $x, y \in X$ and $t \in (0,1)$, we have

$$f(x) - f(tx + (1-t)y) \geq (1-t)(x-y)^T \nabla f(tx + (1-t)y) + c(1-t)^\sigma \|x-y\|^\sigma$$

and

$$f(y) - f(tx + (1-t)y) \geq t(y-x)^T \nabla f(tx + (1-t)y) + ct^\sigma \|x-y\|^\sigma.$$

In consequence, we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(x) - ct(1-t)((1-t)^{\sigma-1} + t^{\sigma-1})\|x-y\|^\sigma. \tag{2.18}$$

If $0 < \sigma \leq 2$, then

$$(1-t)^{\sigma-1} + t^{\sigma-1} \geq (1-t) + t = 1.$$

If $\sigma > 2$, since the real function $\phi(t) = t^{\sigma-1}$ is convex on $(0,1)$, then

$$(1-t)^{\sigma-1} + t^{\sigma-1} \geq \left(\frac{1}{2}\right)^{\sigma-2}.$$

It follows from (2.18) that there exists some constant $c' > 0$ independent of x, y and t such that

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - c't(1-t)\|x-y\|^\sigma,$$

i.e., f is strongly convex with order σ on X. ∎

For a given concept of convexity, there usually exists some kind of monotonicity relevant to it, see [49] and the references therein. Now we define the strong monotonicity with order σ and discuss its relation to the strong convexity with order σ.

Definition 2.4 A mapping $G : \Re^n \to \Re^n$ is said to be *strongly monotone* with order σ on a convex set X if there exists a constant $\beta > 0$ such that

$$(y-x)^T(G(y) - G(x)) \geq \beta\|y-x\|^\sigma, \qquad \forall x, y \in X. \tag{2.19}$$

Theorem 2.13 *Let $X \subseteq \Re^n$ be convex and $f : \Re^n \to \Re$ be continuously differentiable on an open set containing X. Then f is strongly convex with order σ on X if and only if ∇f is strongly monotone with order σ on X.*

Proof: Suppose that f is strongly convex with order σ on X. By Theorem 2.12, there exists a constant $c > 0$ such that (2.17) holds. Then for any $x, y \in X$, one has

$$f(y) - f(x) \geq (y-x)^T \nabla f(x) + c\|x-y\|^\sigma$$

and

$$f(x) - f(y) \geq (x-y)^T \nabla f(y) + c\|x-y\|^\sigma.$$

Therefore,

$$(y-x)^T(\nabla f(y) - \nabla f(x)) \geq 2c\|x-y\|^\sigma,$$

i.e., ∇f is strongly monotone with order σ on X with $\beta = 2c$.

Conversely, assume that (2.19) holds for some $\beta > 0$ and $F = \nabla f$. Set

$$t_i := \frac{i}{m+1}, \qquad i = 0, 1, \cdots, m+1,$$

where m is a positive integer. By the mean-value theorem, there exist $\xi_i \in (t_i, t_{i+1}), 0 \leq i \leq m$, such that

$$f(x + t_{i+1}(y-x)) - f(x + t_i(y-x)) = (t_{i+1} - t_i)(y-x)^T \nabla f(x + \xi_i(y-x)).$$

Hence, it follows from (2.19) that

$$\begin{aligned} f(y) - f(x) &= \sum_{i=0}^{m}(f(x + t_{i+1}(y-x)) - f(x + t_i(y-x))) \\ &= \sum_{i=0}^{m}(t_{i+1} - t_i)(y-x)^T(\nabla f(x + \xi_i(y-x)) - \nabla f(x)) + (y-x)^T \nabla f(x) \\ &\geq \beta\|y-x\|^\sigma \sum_{i=0}^{m} \xi_i^{\sigma-1}(t_{i+1} - t_i) + (y-x)^T \nabla f(x). \end{aligned}$$

Letting $m \to +\infty$ and noticing that

$$\lim_{m \to +\infty} \sum_{i=0}^{m} \xi_i^{\sigma-1}(t_{i+1} - t_i) = \int_0^1 t^{\sigma-1} dt = \frac{1}{\sigma},$$

we have

$$f(y) - f(x) \geq \frac{\beta}{\sigma}\|y - x\|^{\sigma} + (y - x)^T \nabla f(x).$$

By Theorem 2.12, f is strongly convex with order σ on X. ∎

Chapter 3

New Relaxation Method for MPECs

Consider the following mathematical program with complementarity constraints:

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & g(x,y) \le 0,\ h(x,y) = 0, \\ & F(x,y) \ge 0,\ y \ge 0, \\ & y^T F(x,y) = 0, \end{array} \tag{3.1}$$

where $f : \Re^{n+m} \to \Re, g : \Re^{n+m} \to \Re^p, h : \Re^{n+m} \to \Re^q$, and $F : \Re^{n+m} \to \Re^m$ are all twice continuously differentiable functions. As mentioned in Chapter 1, the major difficulty in solving (3.1) is that its constraints fail to satisfy a standard constraint qualification at any feasible point [17], which is necessary for the regularity of a nonlinear program, so that standard methods are likely to fail for this problem. There have been proposed several approaches such as sequential quadratic programming (SQP) approach, implicit programming approach, penalty function approach, and reformulation approach. In this chapter, we study problem (3.1) from another point of view. We use an expansive simplex instead of the nonnegative orthant involved in the complementarity constraints. In other words, our method replaces the complementarity constraints by a variational inequality defined on an expansive simplex. It is well known that such a variational inequality problem can be represented by a finite number of inequalities. We remove some inequalities and obtain a standard nonlinear program. We will show that the linear independence constraint qualification (LICQ) or the Mangasarian-Fromovitz constraint qualification (MFCQ) holds for the relaxed problem under some mild conditions. We also consider a limiting behavior of the relaxed problem. In particular, some sufficient conditions of B-stationarity for a feasible point of the original problem are new and can be verified easily in practice.

3.1 Relaxed Problem

Let $e := (1, 1, \cdots, 1)^T$. For $i = 1, 2, \cdots, m$, $e_i \in \Re^m$ denotes the i-th column of the $m \times m$ identity matrix. Also we let $e_0 \in \Re^m$ denote the zero vector, i.e., $e_0 := (0, 0, \cdots, 0)^T$. Then the expansive simplex mentioned above is defined by

$$\Omega_k := \text{co}\{e_0^k, e_1^k, \cdots, e_m^k\},$$

where co stands for the convex hull, k is a positive integer, and

$$e_i^k := \frac{1}{k}e + ke_i, \quad i = 0, 1, \cdots, m. \tag{3.2}$$

For a fixed $x \in \Re^n$, the problem VI($F(x, \cdot), \Omega_k$) is obviously equivalent to finding a $y \in \Re^m$ such that

$$y \in \Omega_k, \quad (e_i^k - y)^T F(x, y) \geq 0, \quad i = 0, 1, \cdots, m$$

or equivalently,

$$\sum_{j=1}^m y_j \leq \frac{m}{k} + k, \quad y_i \geq \frac{1}{k}, \quad (e_i^k - y)^T F(x, y) \geq 0, \quad i = 0, 1, \cdots, m.$$

In order to simplify the relaxed problem, we replace the condition $y \in \Omega_k$ by $y \geq 0$ and consider the following problem as an approximation of problem (3.1):

$$\begin{array}{ll} \text{minimize} & f(x, y) \\ \text{subject to} & g(x, y) \leq 0, \;\; h(x, y) = 0, \;\; y \geq 0, \\ & (e_i^k - y)^T F(x, y) \geq 0, \;\; i = 0, 1, \cdots, m. \end{array} \tag{3.3}$$

Let $\mathcal{F}$ and $\mathcal{F}_k$ denote the feasible sets of problems (3.1) and (3.3), respectively, and let

$$\phi_i^k(x, y) := (e_i^k - y)^T F(x, y), \quad i = 0, 1, \cdots, m. \tag{3.4}$$

By (3.2), we have

$$\phi_i^k(x, y) = \phi_0^k(x, y) + kF_i(x, y), \quad i = 1, 2, \cdots, m. \tag{3.5}$$

Then we have the following results.

Theorem 3.1 *For problems (3.1) and (3.3), we have*

(i) *for any* k, $\mathcal{F} \subseteq \mathcal{F}_{k+1} \subseteq \mathcal{F}_k$;

(ii) $\mathcal{F} = \bigcap_{k=1}^{\infty} \mathcal{F}_k$, *which, together with the continuity of the functions involved, implies that any accumulation point of a sequence* $\{(x^k, y^k) \mid (x^k, y^k) \in \mathcal{F}_k\}$ *belongs to* $\mathcal{F}$.

Proof: (i) $\mathcal{F} \subseteq \mathcal{F}_{k+1}$ is clear. Let $(x, y) \in \mathcal{F}_{k+1}$. Then, since for each $i = 0, 1, \cdots, m$, e_i^k can be represented as

$$e_i^k = \sum_{j=0}^{m} t_{ij} e_j^{k+1}, \quad \sum_{j=0}^{m} t_{ij} = 1, \quad t_{ij} \geq 0, \quad j = 0, 1, \cdots, m,$$

we have

$$(e_i^k - y)^T F(x, y) = \sum_{j=0}^{m} t_{ij} (e_j^{k+1} - y)^T F(x, y) \geq 0, \quad i = 0, 1, \cdots, m,$$

i.e., $(x, y) \in \mathcal{F}_k$. Hence $\mathcal{F}_{k+1} \subseteq \mathcal{F}_k$.

(ii) From (i), we only need to prove $\bigcap_{k=1}^{\infty} \mathcal{F}_k \subseteq \mathcal{F}$. Let $(x, y) \in \bigcap_{k=1}^{\infty} \mathcal{F}_k$. Then we have

$$g(x, y) \leq 0, \;\; h(x, y) = 0, \;\; y \geq 0$$

and, for every $i = 1, 2, \cdots, m$,

$$\phi_i^k(x, y) = (\frac{1}{k} e + k e_i - y)^T F(x, y) \geq 0, \quad \forall k,$$

which implies

$$(\frac{1}{k^2} e + e_i - \frac{1}{k} y)^T F(x, y) \geq 0, \quad \forall k. \tag{3.6}$$

Letting $k \to \infty$ in (3.6), we have

$$F_i(x, y) = e_i^T F(x, y) \geq 0$$

and hence $F(x, y) \geq 0$. On the other hand,

$$(e_0^k - y)^T F(x, y) \geq 0, \quad \forall k$$

implies $-y^T F(x, y) \geq 0$. So, we have $y^T F(x, y) = 0$. Therefore, $(x, y) \in \mathcal{F}$ and so $\bigcap_{k=1}^{\infty} \mathcal{F}_k \subseteq \mathcal{F}$. This completes the proof. ∎

The following result shows that problem (3.3) may satisfy some constraint qualification at its feasible points. This is in contrast with problem (3.1), for which a standard contraint qualification fails to hold at any feasible point.

Theorem 3.2 *For any* $(\bar{x}, \bar{y}) \in \mathcal{F}$ *with* $F(\bar{x}, \bar{y}) \neq 0$*, we have*

$$\phi_i^k(\bar{x}, \bar{y}) > 0, \quad i = 0, 1, \cdots, m, \quad \forall k$$

and so they are inactive constraints at $(\bar{x}, \bar{y})$ *in problem (3.3). In this case, if the system*

$$g(x, y) \leq 0, \quad h(x, y) = 0, \quad y \geq 0$$

satisfies some constraint qualification such as LICQ or MFCQ at $(\bar{x}, \bar{y})$*, then, for any fixed* k*, there exists a neighborhood* $U_k(\bar{x}, \bar{y})$ *of* $(\bar{x}, \bar{y})$ *such that problem (3.3) satisfies the same constraint qualification at any point* $(x, y) \in U_k(\bar{x}, \bar{y})$*.*

Proof: We obtain the first part from the definition (3.4) of ϕ_i^k and

$$e_i^k > 0, \quad 0 \neq F(\bar{x}, \bar{y}) \geq 0, \quad \bar{y}^T F(\bar{x}, \bar{y}) = 0,$$

immediately. The second part follows from the continuity of $g, h, \phi_i^k, i = 0, 1, \cdots, m$, and their gradients directly. ∎

In what follows, we let $G(x, y) := y$ and

$$\phi^k(x, y) := (\phi_0^k(x, y), \phi_1^k(x, y), \cdots, \phi_m^k(x, y))^T,$$

where ϕ_i^k are defined by (3.4). Note that the gradients of $G_j, j = 1, \cdots, m$, are constant vectors. Nevertheless, we will often write $\nabla G_j(x, y)$, etc., to specify the point under consideration.

Theorem 3.3 *For any* $(\bar{x}, \bar{y}) \in \mathcal{F}$*, if the set of vectors*

$$\Big\{\nabla F_i(\bar{x}, \bar{y}), \nabla G_i(\bar{x}, \bar{y}), \nabla g_l(\bar{x}, \bar{y}), \nabla h_r(\bar{x}, \bar{y}) \;\Big|\; i = 1, \cdots, m, \; l \in \mathcal{I}_g(\bar{x}, \bar{y}), \; r = 1, \cdots, q\Big\}$$

is linearly independent, then there exist a neighborhood $U(\bar{x}, \bar{y})$ *of* $(\bar{x}, \bar{y})$ *and a positive integer* K *such that, for any* $(x, y) \in U(\bar{x}, \bar{y})$ *and any* $k \geq K$*, the following conditions hold:*

(i) $\mathcal{I}_F(x, y) \subseteq \mathcal{I}_F(\bar{x}, \bar{y}), \; \mathcal{I}_G(x, y) \subseteq \mathcal{I}_G(\bar{x}, \bar{y}), \; \mathcal{I}_g(x, y) \subseteq \mathcal{I}_g(\bar{x}, \bar{y})$*;*

(ii) $\Big\{\nabla F_i(x, y), \nabla G_i(x, y), \nabla g_l(x, y), \nabla h_r(x, y) \;|\; i = 1, \cdots, m, l \in \mathcal{I}_g(\bar{x}, \bar{y}), r = 1, \cdots, q\Big\}$ *is linearly independent;*

(iii) $\mathcal{I}_{\phi^k}(x, y) \subseteq \{0\} \cup \mathcal{I}_F(\bar{x}, \bar{y})$*.*

Proof: It is obvious that there exists a neighborhood $U_1(\bar{x}, \bar{y})$ such that conditions (i) and (ii) hold for any $(x, y) \in U_1(\bar{x}, \bar{y})$ by the continuity of $F, G, g, \nabla F, \nabla g$, and ∇h. Now we show that there exist a neighborhood $U_2(\bar{x}, \bar{y})$ and a positive integer K

satisfying condition (iii) for any $(x, y) \in U_2(\bar{x}, \bar{y})$ and any $k \geq K$. Otherwise, there must be an $i_0 \notin \{0\} \cup \mathcal{I}_F(\bar{x}, \bar{y})$, a subsequence $\{k_j\}$ of $\{k\}$, and a sequence $\{(x^j, y^j)\}$ converging to $(\bar{x}, \bar{y})$ such that

$$\phi_{i_0}^{k_j}(x^j, y^j) = 0, \quad \forall j.$$

Since

$$\frac{1}{k_j}\phi_{i_0}^{k_j}(x^j, y^j) = F_{i_0}(x^j, y^j) + \frac{1}{k_j^2}\sum_{i=1}^{m} F_i(x^j, y^j) - \frac{1}{k_j}(y^j)^T F(x^j, y^j),$$

we have

$$\lim_{j\to\infty} \frac{1}{k_j}\phi_{i_0}^{k_j}(x^j, y^j) = F_{i_0}(\bar{x}, \bar{y}) > 0.$$

This implies that

$$\lim_{j\to\infty} \phi_{i_0}^{k_j}(x^j, y^j) = +\infty.$$

This is a contradiction and so the neighborhood $U_2(\bar{x}, \bar{y})$ and positive integer K mentioned above exist. Let

$$U(\bar{x}, \bar{y}) = U_1(\bar{x}, \bar{y}) \cap U_2(\bar{x}, \bar{y}).$$

Then conditions (i)–(iii) hold for any $(x, y) \in U(\bar{x}, \bar{y})$ and any $k \geq K$. ∎

We further have the following result about constraint qualifications.

Theorem 3.4 *Let $(\bar{x}, \bar{y}) \in \mathcal{F}$ be nondegenerate, i.e., $\mathcal{I}_F(\bar{x}, \bar{y}) \cap \mathcal{I}_G(\bar{x}, \bar{y}) = \emptyset$, and assume*

$$F(\bar{x}, \bar{y}) \neq 0. \tag{3.7}$$

If the set of vectors

$$\Big\{\nabla F_i(\bar{x}, \bar{y}), \nabla G_i(\bar{x}, \bar{y}), \nabla g_l(\bar{x}, \bar{y}), \nabla h_r(\bar{x}, \bar{y}) \;\Big|\; i = 1, \cdots, m, l \in \mathcal{I}_g(\bar{x}, \bar{y}), r = 1, \cdots, q\Big\}$$

is linearly independent, then there exists a neighborhood $U(\bar{x}, \bar{y})$ of $(\bar{x}, \bar{y})$ such that, for any sufficiently large k, problem (3.3) satisfies the standard LICQ at any point $(x, y) \in U(\bar{x}, \bar{y}) \cap \mathcal{F}_k$.

Proof: By Theorem 3.3, there exist a neighborhood $U(\bar{x}, \bar{y})$ and a positive integer K such that Theorem 3.3 (i)–(iii) hold for any $(x, y) \in U(\bar{x}, \bar{y})$ and any $k \geq K$. Now we let $k \geq K$ and choose an arbitrary point $(x, y) \in U(\bar{x}, \bar{y}) \cap \mathcal{F}_k$.

Suppose that the LICQ does not hold at (x, y) for problem (3.3). This means the set of vectors

$$\Big\{\nabla \phi_i^k(x, y), \nabla G_j(x, y), \nabla g_l(x, y), \nabla h_r(x, y)\Big| \\ i \in \mathcal{I}_{\phi^k}(x, y), j \in \mathcal{I}_G(x, y), l \in \mathcal{I}_g(x, y), r = 1, \cdots, q\Big\},$$

which is, by (3.5),

$$\Big\{\nabla\phi_0^k(x,y), k\nabla F_i(x,y)+\nabla\phi_0^k(x,y), \nabla G_j(x,y), \nabla g_l(x,y), \nabla h_r(x,y) \ \Big| \ 0\neq i\in\mathcal{I}_{\phi^k}(x,y), j\in\mathcal{I}_G(x,y), l\in\mathcal{I}_g(x,y), r=1,2,\cdots,q\Big\}$$

in the case where $0\in\mathcal{I}_{\phi^k}(x,y)$ or

$$\Big\{k\nabla F_i(x,y)+\nabla\phi_0^k(x,y), \nabla G_j(x,y), \nabla g_l(x,y), \nabla h_r(x,y) \ \Big| \ i\in\mathcal{I}_{\phi^k}(x,y), j\in\mathcal{I}_G(x,y), l\in\mathcal{I}_g(x,y), r=1,2,\cdots,q\Big\}$$

in the case where $0\notin\mathcal{I}_{\phi^k}(x,y)$, is linearly dependent. Hence, by Theorem 3.3 (ii), $\nabla\phi_0^k(x,y)$ can be represented as a linear combination of the vectors

$$\Big\{\nabla F_i(x,y), \nabla G_j(x,y), \nabla g_l(x,y), \nabla h_r(x,y) \ \Big| \ i\in\mathcal{I}_{\phi^k}(x,y)\setminus\{0\}, j\in\mathcal{I}_G(x,y), l\in\mathcal{I}_g(x,y), r=1,2,\cdots,q\Big\}.$$

Therefore, there exist numbers

$$\Big\{\lambda_i, \mu_j, u_l, v_r \ \Big| \ i\in\mathcal{I}_{\phi^k}(x,y)\setminus\{0\}, j\in\mathcal{I}_G(x,y), l\in\mathcal{I}_g(x,y), r=1,2,\cdots,q\Big\}$$

such that

$$\begin{aligned}\nabla\phi_0^k(x,y) &= \sum_{i\in\mathcal{I}_{\phi^k}(x,y)\setminus\{0\}}\lambda_i\nabla F_i(x,y)+\sum_{j\in\mathcal{I}_G(x,y)}\mu_j\nabla G_j(x,y)\\ &\quad+\sum_{l\in\mathcal{I}_g(x,y)}u_l\nabla g_l(x,y)+\sum_{r=1}^q v_r\nabla h_r(x,y).\end{aligned}$$

Since $\phi_0^k(x,y)=(\frac{1}{k}e-y)^TF(x,y)$, we then have

$$\begin{aligned}&\sum_{i\in\mathcal{I}_{\phi^k}(x,y)\setminus\{0\}}(y_i-\frac{1}{k}+\lambda_i)\nabla F_i(x,y)+\sum_{i\notin\mathcal{I}_{\phi^k}(x,y)\setminus\{0\}}(y_i-\frac{1}{k})\nabla F_i(x,y)\\ &+\sum_{j\in\mathcal{I}_G(x,y)}(\mu_j+F_j(x,y))\nabla G_j(x,y)+\sum_{j\notin\mathcal{I}_G(x,y)}F_j(x,y)\nabla G_j(x,y)\\ &+\sum_{l\in\mathcal{I}_g(x,y)}u_l\nabla g_l(x,y)+\sum_{r=1}^q v_r\nabla h_r(x,y)=0.\end{aligned}$$

By Theorem 3.3 (ii), we have

$$y_i=\begin{cases}\frac{1}{k}-\lambda_i, & i\in\mathcal{I}_{\phi^k}(x,y)\setminus\{0\}\\ \frac{1}{k}, & i\notin\mathcal{I}_{\phi^k}(x,y)\setminus\{0\}\end{cases} \tag{3.8}$$

and

$$F_i(x,y)=0, \quad i \notin \mathcal{I}_G(x,y). \tag{3.9}$$

Suppose that $\phi_0^k(x,y)=0$. We then have by (3.9) that

$$\phi_i^k(x,y)=kF_i(x,y)+\phi_0^k(x,y)=0, \quad i \notin \mathcal{I}_G(x,y). \tag{3.10}$$

On the other hand, we have

$$\mathcal{I}_G(x,y)=\mathcal{I}_G(\bar{x},\bar{y}). \tag{3.11}$$

Otherwise, since $\mathcal{I}_G(x,y) \subseteq \mathcal{I}_G(\bar{x},\bar{y})$, there exists an $i_0 \in \mathcal{I}_G(\bar{x},\bar{y}) \setminus \mathcal{I}_G(x,y)$. Then we must have $\phi_{i_0}^k(x,y)=0$ by (3.10). But, by the nondegenerate property of $(\bar{x},\bar{y})$, $i_0 \in \mathcal{I}_G(\bar{x},\bar{y})$ means $i_0 \notin \mathcal{I}_F(\bar{x},\bar{y})$. So, by Theorem 3.3 (iii), we have $\phi_{i_0}^k(x,y)>0$. This is a contradiction and so (3.11) holds. This means that

$$y_i=0, \quad i \in \mathcal{I}_G(\bar{x},\bar{y}). \tag{3.12}$$

Note that $\mathcal{I}_G(\bar{x},\bar{y}) \neq \emptyset$ by (3.7). So, if $i \in \mathcal{I}_G(\bar{x},\bar{y})$, then $i \notin \mathcal{I}_F(\bar{x},\bar{y})$ by the nondegeneracy assumption and so $i \notin \mathcal{I}_{\phi^k}(x,y) \setminus \{0\}$ by Theorem 3.1 (iii). Hence, from (3.8), we have $y_i=\frac{1}{k}$. This contradicts (3.12). Therefore, we have $\phi_0^k(x,y) \neq 0$ and so, by (3.9),

$$\phi_i^k(x,y)=kF_i(x,y)+\phi_0^k(x,y) \neq 0, \quad i \notin \mathcal{I}_G(x,y). \tag{3.13}$$

By Theorem 3.3 (iii) and the fact that $0 \notin \mathcal{I}_{\phi^k}(x,y)$, $\phi_i^k(x,y)=0$ means that $i \in \mathcal{I}_F(\bar{x},\bar{y})$. On the other hand, for any $i \in \mathcal{I}_F(\bar{x},\bar{y})$, by the nondegeneracy of $(\bar{x},\bar{y})$, we have $i \notin \mathcal{I}_G(\bar{x},\bar{y})$ and so $i \notin \mathcal{I}_G(x,y)$ by Theorem 3.3 (i), which implies $\phi_i^k(x,y) \neq 0$ by (3.13). Hence $\mathcal{I}_{\phi^k}(x,y)=\emptyset$, i.e., the last $(m+1)$ inequality constraints in problem (3.3) are all inactive at (x,y) and so, by Theorem 3.3 (ii), the LICQ holds at (x,y). This also contradicts our assumption. Therefore, the LICQ holds at (x,y) for problem (3.3). This completes the proof. ∎

3.2 Convergence Analysis

In this section, we investigate the behavior of problem (3.3) as $k \to \infty$. We first consider the convergence of global optimal solutions.

Theorem 3.5 *Suppose that (x^k,y^k) is a global optimal solution of problem (3.3) and (x^*,y^*) is an accumulation point of the sequence $\{(x^k,y^k)\}$ as $k \to \infty$. Then (x^*,y^*) is a global optimal solution of problem (3.1).*

Proof: Taking a subsequence if necessary, we assume without loss of generality that

$$\lim_{k\to\infty}(x^k, y^k) = (x^*, y^*).$$

By Theorem 3.1, $(x^*, y^*) \in \mathcal{F}$. Since $\mathcal{F} \subseteq \mathcal{F}_k$ for all k, then

$$f(x^k, y^k) \le f(x, y), \quad \forall (x, y) \in \mathcal{F}, \quad \forall k.$$

Letting $k \to \infty$, we have from the continuity of f that

$$f(x^*, y^*) \le f(x, y), \quad \forall (x, y) \in \mathcal{F},$$

i.e., (x^*, y^*) is a global optimal solution of problem (3.1). ∎

In a similar way, we can prove the next theorem.

Theorem 3.6 *Let $\{\varepsilon_k\} \subseteq (0, +\infty)$ be convergent to 0 and $(x^k, y^k) \in \mathcal{F}_k$ be an approximate solution of problem (3.3) satisfying*

$$f(x^k, y^k) - \varepsilon_k \le f(x, y), \quad \forall (x, y) \in \mathcal{F}_k.$$

Then any accumulation point of $\{(x^k, y^k)\}$ is a global optimal solution of problem (3.1).

Now we consider the limiting behavior of stationary points of problem (3.3).

Theorem 3.7 *Let $(x^k, y^k) \in \mathcal{F}_k$ be a stationary point of problem (3.3) with Lagrange multiplier vectors $\lambda^k, \mu^k, \delta^k$, and γ^k satisfying (3.15)–(3.16), and $(\bar{x}, \bar{y})$ be an accumulation point of the sequence $\{(x^k, y^k)\}$. Suppose the set of vectors*

$$\Big\{\nabla F_i(\bar{x}, \bar{y}), \nabla G_i(\bar{x}, \bar{y}), \nabla g_l(\bar{x}, \bar{y}), \nabla h_r(\bar{x}, \bar{y}) \;\Big|\; i = 1, \cdots, m,\ l \in \mathcal{I}_g(\bar{x}, \bar{y}),\ r = 1, \cdots, q\Big\}$$

is linearly independent. Then we have the following statements.

(a) *$(\bar{x}, \bar{y})$ is a weakly stationary point of problem (3.1) and, if $F(\bar{x}, \bar{y}) \ne 0$, $(\bar{x}, \bar{y})$ is C-stationary. Especially, if $(\bar{x}, \bar{y})$ is nondegenerate, it is B-stationary;*

(b) *If $(x^k, y^k) \in \mathcal{F}$ for some k, then (x^k, y^k) is B-stationary to problem (3.1) and if $(x^k, y^k) \in \mathcal{F}$ for infinitely many k, $(\bar{x}, \bar{y})$ is B-stationary;*

(c) *If $0 \notin \mathcal{I}_{\phi^k}(x^k, y^k)$ for infinitely many k, then $(\bar{x}, \bar{y})$ is a B-stationary point to (3.1).*

Proof: Without loss of generality, we assume that

$$\lim_{k\to\infty}(x^k, y^k) = (\bar{x}, \bar{y}). \tag{3.14}$$

Then by Theorem 3.1, we have $(\bar{x}, \bar{y}) \in \mathcal{F}$. By Theorem 3.3, for any sufficiently large k, we have

$$\mathcal{I}_F(x^k, y^k) \subseteq \mathcal{I}_F(\bar{x}, \bar{y}), \quad \mathcal{I}_G(x^k, y^k) \subseteq \mathcal{I}_G(\bar{x}, \bar{y}),$$
$$\mathcal{I}_g(x^k, y^k) \subseteq \mathcal{I}_g(\bar{x}, \bar{y}), \quad \mathcal{I}_{\phi^k}(x^k, y^k) \subseteq \{0\} \cup \mathcal{I}_F(\bar{x}, \bar{y}),$$

and

$$\Big\{\nabla F_i(x^k, y^k), \nabla G_i(x^k, y^k), \nabla g_l(x^k, y^k), \nabla h_r(x^k, y^k) \ \Big| \ i = 1, \cdots, m, \ l \in \mathcal{I}_g(\bar{x}, \bar{y}), \ r = 1, \cdots, q\Big\}$$

is linearly independent. Note that the MPEC-LICQ holds at $(\bar{x}, \bar{y})$ for problem (3.1).

By the stationarity of (x^k, y^k) for problem (3.3), there exist Lagrange multiplier vectors $\lambda^k, \mu^k, \delta^k$, and γ^k such that

$$\nabla f(x^k, y^k) - \sum_{i \in \mathcal{I}_{\phi^k}(x^k, y^k)} \lambda_i^k \nabla \phi_i^k(x^k, y^k) - \sum_{j \in \mathcal{I}_G(x^k, y^k)} \mu_j^k \nabla G_j(x^k, y^k) + \sum_{l \in \mathcal{I}_g(x^k, y^k)} \delta_l^k \nabla g_l(x^k, y^k) + \sum_{r=1}^{q} \gamma_r^k \nabla h_r(x^k, y^k) = 0 \quad (3.15)$$

and

$$\lambda^k \geq 0, \quad \mu^k \geq 0, \quad \delta^k \geq 0. \quad (3.16)$$

Since

$$\nabla \phi_i^k(x^k, y^k) = \nabla \phi_0^k(x^k, y^k) + k \nabla F_i(x^k, y^k), \quad i = 1, 2, \cdots, m$$

and

$$\nabla \phi_0^k(x^k, y^k) = \sum_{i=1}^{m} (\frac{1}{k} - y_i^k) \nabla F_i(x^k, y^k) - \sum_{j=1}^{m} F_j(x^k, y^k) \nabla G_j(x^k, y^k), \quad (3.17)$$

it follows from (3.15) that

$$\begin{aligned} \nabla f(x^k, y^k) &= \sum_{i \in \mathcal{I}_{\phi^k}(x^k, y^k)} \lambda_i^k \nabla \phi_i^k(x^k, y^k) + \sum_{j \in \mathcal{I}_G(x^k, y^k)} \mu_j^k \nabla G_j(x^k, y^k) \\ &\quad - \sum_{l \in \mathcal{I}_g(x^k, y^k)} \delta_l^k \nabla g_l(x^k, y^k) - \sum_{r=1}^{q} \gamma_r^k \nabla h_r(x^k, y^k) \\ &= \sum_{0 \neq i \in \mathcal{I}_{\phi^k}(x^k, y^k)} k \lambda_i^k \nabla F_i(x^k, y^k) + a_k \nabla \phi_0^k(x^k, y^k) \end{aligned}$$

$$
\begin{aligned}
& + \sum_{j \in \mathcal{I}_G(x^k,y^k)} \mu_j^k \nabla G_j(x^k,y^k) \\
& - \sum_{l \in \mathcal{I}_g(x^k,y^k)} \delta_l^k \nabla g_l(x^k,y^k) - \sum_{r=1}^{q} \gamma_r^k \nabla h_r(x^k,y^k) \qquad (3.18) \\
= & \sum_{i \in \mathcal{I}_F(\bar{x},\bar{y})} u_i^k \nabla F_i(x^k,y^k) + \sum_{i \notin \mathcal{I}_F(\bar{x},\bar{y})} a_k(\frac{1}{k} - y_i^k) \nabla F_i(x^k,y^k) \\
& + \sum_{j \in \mathcal{I}_G(\bar{x},\bar{y})} v_j^k \nabla G_j(x^k,y^k) - \sum_{j \notin \mathcal{I}_G(\bar{x},\bar{y})} a_k F_j(x^k,y^k) \nabla G_j(x^k,y^k) \\
& - \sum_{l \in \mathcal{I}_g(\bar{x},\bar{y})} w_l^k \nabla g_l(x^k,y^k) - \sum_{r=1}^{q} \gamma_r^k \nabla h_r(x^k,y^k), \qquad (3.19)
\end{aligned}
$$

where

$$
a_k := \begin{cases} \sum_{i \in \mathcal{I}_{\phi^k}(x^k,y^k)} \lambda_i^k, & \mathcal{I}_{\phi^k}(x^k,y^k) \neq \emptyset \\ 0, & \mathcal{I}_{\phi^k}(x^k,y^k) = \emptyset, \end{cases} \qquad (3.20)
$$

$$
u_i^k := \begin{cases} k\lambda_i^k + a_k(\frac{1}{k} - y_i^k), & 0 \neq i \in \mathcal{I}_{\phi^k}(x^k,y^k) \\ a_k(\frac{1}{k} - y_i^k), & i \in \mathcal{I}_F(\bar{x},\bar{y}) \setminus \mathcal{I}_{\phi^k}(x^k,y^k), \end{cases} \qquad (3.21)
$$

$$
v_j^k := \begin{cases} \mu_j^k - a_k F_j(x^k,y^k), & j \in \mathcal{I}_G(x^k,y^k) \\ -a_k F_j(x^k,y^k), & j \in \mathcal{I}_G(\bar{x},\bar{y}) \setminus \mathcal{I}_G(x^k,y^k), \end{cases} \qquad (3.22)
$$

$$
w_l^k := \begin{cases} \delta_l^k, & l \in \mathcal{I}_g(x^k,y^k) \\ 0, & l \in \mathcal{I}_g(\bar{x},\bar{y}) \setminus \mathcal{I}_g(x^k,y^k). \end{cases}
$$

Since

$$
\Big\{ \nabla F_i(\bar{x},\bar{y}), \nabla G_i(\bar{x},\bar{y}), \nabla g_l(\bar{x},\bar{y}), \nabla h_r(\bar{x},\bar{y}) \ \Big| \ i = 1,\cdots,m,\ l \in \mathcal{I}_g(\bar{x},\bar{y}),\ r = 1,\cdots,q \Big\}
$$

is linearly independent, it follows from (3.14) and (3.19) that the multiplier sequences

$$
\begin{aligned}
& \Big\{ u_i^k \ \Big| \ i \in \mathcal{I}_F(\bar{x},\bar{y}) \Big\}, && \Big\{ v_j^k \ \Big| \ j \in \mathcal{I}_G(\bar{x},\bar{y}) \Big\}, \\
& \Big\{ a_k(\frac{1}{k} - y_i^k) \ \Big| \ i \notin \mathcal{I}_F(\bar{x},\bar{y}) \Big\}, && \Big\{ w_l^k \ \Big| \ l \in \mathcal{I}_g(\bar{x},\bar{y}) \Big\}, \qquad (3.23) \\
& \Big\{ -a_k F_j(x^k,y^k) \ \Big| \ j \notin \mathcal{I}_G(\bar{x},\bar{y}) \Big\}, && \Big\{ \gamma_r^k \ \Big| \ r = 1,\cdots,q \Big\}
\end{aligned}
$$

are convergent. Next we will consider several cases to prove statements (a)–(c).

(I) First we show that if $(x^k, y^k) \in \mathcal{F}$ for some k, then it is a B-stationary point of problem (3.1), namely, there exist multiplier vectors $\bar{\lambda}, \bar{\mu}, \bar{\gamma}$, and $\bar{\delta} \geq 0$ such that

$$
\begin{aligned}
\nabla f(\bar{x},\bar{y}) - \sum_{i \in \mathcal{I}_F(\bar{x},\bar{y})} \bar{\lambda}_i \nabla F_i(\bar{x},\bar{y}) - \sum_{j \in \mathcal{I}_G(\bar{x},\bar{y})} \bar{\mu}_j \nabla G_j(\bar{x},\bar{y}) & \\
+ \sum_{l \in \mathcal{I}_g(\bar{x},\bar{y})} \bar{\delta}_l \nabla g_l(\bar{x},\bar{y}) + \sum_{r=1}^{q} \bar{\gamma}_r \nabla h_r(\bar{x},\bar{y}) & = 0 \qquad (3.24)
\end{aligned}
$$

holds with

$$\bar{\lambda}_i \geq 0, \quad \bar{\mu}_i \geq 0, \quad i \in \mathcal{I}_F(\bar{x},\bar{y}) \cap \mathcal{I}_G(\bar{x},\bar{y}). \tag{3.25}$$

In fact, if $F(x^k,y^k) \neq 0$, then, from Theorem 3.2, $\mathcal{I}_{\phi^k}(x^k,y^k) = \emptyset$ and so (3.15)–(3.16) mean that (x^k,y^k) is a B-stationary point of problem (3.1). If $F(x^k,y^k) = 0$, then $\mathcal{I}_{\phi^k}(x^k,y^k) = \{0,1,\cdots,m\}$ and so, it follows from (3.17) and (3.18) that

$$\begin{aligned} 0 &= \nabla f(x^k,y^k) - \sum_{i=1}^{m} k\lambda_i^k \nabla F_i(x^k,y^k) - a_k \nabla \phi_0^k(x^k,y^k) \\ &\quad - \sum_{j\in\mathcal{I}_G(x^k,y^k)} \mu_j^k \nabla G_j(x^k,y^k) + \sum_{l\in\mathcal{I}_g(x^k,y^k)} \delta_l^k \nabla g_l(x^k,y^k) + \sum_{r=1}^{q} \gamma_r^k \nabla h_r(x^k,y^k) \\ &= \nabla f(x^k,y^k) - \sum_{i\in\mathcal{I}_F(x^k,y^k)} (a_k(\frac{1}{k} - y_i^k) + k\lambda_i^k) \nabla F_i(x^k,y^k) \\ &\quad - \sum_{j\in\mathcal{I}_G(x^k,y^k)} \mu_j^k \nabla G_j(x^k,y^k) + \sum_{l\in\mathcal{I}_g(x^k,y^k)} \delta_l^k \nabla g_l(x^k,y^k) + \sum_{r=1}^{q} \gamma_r^k \nabla h_r(x^k,y^k). \end{aligned}$$

For $i \in \mathcal{I}_F(x^k,y^k) \cap \mathcal{I}_G(x^k,y^k)$, we have from (3.16) and (3.20) that

$$a_k(\frac{1}{k} - y_i^k) + k\lambda_i^k = \frac{1}{k}a_k + k\lambda_i^k \geq 0, \quad \mu_i^k \geq 0$$

and hence, comparing with (3.24) and (3.25), we see that (x^k,y^k) is a B-stationary point of problem (3.1). This shows the first half of statement (b). Next we suppose $(x^{k'},y^{k'}) \in \mathcal{F}$ for infinitely many k' and show $(\bar{x},\bar{y})$ is a B-stationary point of problem (3.1). In fact, since for any sufficiently large k',

$$\mathcal{I}_F(x^{k'},y^{k'}) \subseteq \mathcal{I}_F(\bar{x},\bar{y}), \quad \mathcal{I}_G(x^{k'},y^{k'}) \subseteq \mathcal{I}_G(\bar{x},\bar{y}), \quad \mathcal{I}_g(x^{k'},y^{k'}) \subseteq \mathcal{I}_g(\bar{x},\bar{y}),$$

we have

$$\begin{aligned} \nabla f(x^{k'},y^{k'}) &= \sum_{i\in\mathcal{I}_F(\bar{x},\bar{y})} \hat{u}_i^{k'} \nabla F_i(x^{k'},y^{k'}) + \sum_{j\in\mathcal{I}_G(\bar{x},\bar{y})} \hat{v}_j^{k'} \nabla G_j(x^{k'},y^{k'}) \\ &\quad - \sum_{l\in\mathcal{I}_g(\bar{x},\bar{y})} \hat{w}_l^{k'} \nabla g_l(x^{k'},y^{k'}) - \sum_{r=1}^{q} \hat{\gamma}_r^{k'} \nabla h_r(x^{k'},y^{k'}). \end{aligned}$$

By the assumptions of the theorem, the multiplier sequences converge. Letting $k' \to \infty$, we have the B-stationarity of $(\bar{x},\bar{y})$. This shows the second half of statement (b).

(II) Next we assume that $(x^k,y^k) \notin \mathcal{F}$ for all sufficiently large k.

(IIa) We consider the case where $\mathcal{I}_F(\bar{x},\bar{y}) \neq \emptyset$.

(i) We first prove statement (c), i.e., if there is a subsequence $\{k_l\}$ of $\{k\}$ such that $0 \notin \mathcal{I}_{\phi^{k_l}}(x^{k_l}, y^{k_l})$ for all l, then $(\bar{x}, \bar{y})$ is a B-stationary point of problem (3.1). In fact, noting that, by (3.20) and (3.21),

$$\begin{aligned}\sum_{i\in\mathcal{I}_F(\bar{x},\bar{y})} u_i^{k_l} &= \sum_{i\in\mathcal{I}_{\phi^{k_l}}(x^{k_l},y^{k_l})} \Big(k_l\lambda_i^{k_l} + a_{k_l}(\frac{1}{k_l} - y_i^{k_l})\Big) + \sum_{i\in\mathcal{I}_F(\bar{x},\bar{y})\setminus\mathcal{I}_{\phi^{k_l}}(x^{k_l},y^{k_l})} a_{k_l}(\frac{1}{k_l} - y_i^{k_l})\\ &= a_{k_l}\Big(k_l + \sum_{i\in\mathcal{I}_F(\bar{x},\bar{y})}(\frac{1}{k_l} - y_i^{k_l})\Big)\end{aligned}$$

and

$$\begin{aligned}&\lim_{l\to\infty}\sum_{i\in\mathcal{I}_F(\bar{x},\bar{y})} u_i^{k_l} \quad \text{exists},\\ &\lim_{l\to\infty}\Big(k_l + \sum_{i\in\mathcal{I}_F(\bar{x},\bar{y})}(\frac{1}{k_l} - y_i^{k_l})\Big) = +\infty,\end{aligned}$$

we have

$$\lim_{l\to\infty} a_{k_l} = 0. \tag{3.26}$$

Therefore, we obtain

$$\lim_{k\to\infty} a_k(\frac{1}{k} - y_i^k) = \lim_{l\to\infty} a_{k_l}(\frac{1}{k_l} - y_i^{k_l}) = 0, \quad i \notin \mathcal{I}_F(\bar{x}, \bar{y}) \tag{3.27}$$

and

$$\lim_{k\to\infty} a_k F_j(x^k, y^k) = \lim_{l\to\infty} a_{k_l} F_j(x^{k_l}, y^{k_l}) = 0, \quad j \notin \mathcal{I}_G(\bar{x}, \bar{y}). \tag{3.28}$$

On the other hand, by (3.16), (3.26), and (3.21)–(3.22), we have

$$\lim_{k\to\infty} u_i^k \geq 0, \quad \lim_{k\to\infty} v_i^k \geq 0, \quad i \in \mathcal{I}_G(\bar{x}, \bar{y}) \cap \mathcal{I}_F(\bar{x}, \bar{y}). \tag{3.29}$$

It then follows from (3.19) and (3.27)–(3.29) that conditions (3.24) and (3.25) hold. Therefore, $(\bar{x}, \bar{y})$ is a B-stationary point of problem (3.1). This shows statement (c). The rest of the proof will be devoted to showing statement (a).

(ii) Suppose that $0 \in \mathcal{I}_{\phi^k}(x^k, y^k)$ for all sufficiently large k. Then it follows from (3.5) that

$$\mathcal{I}_{\phi^k}(x^k, y^k) = \{0\} \cup \mathcal{I}_F(x^k, y^k). \tag{3.30}$$

(iia) If there exist a subsequence $\{k_l\}$ of $\{k\}$ and an index i_0 such that

$$i_0 \notin \mathcal{I}_G(\bar{x}, \bar{y}), \quad i_0 \in \mathcal{I}_F(\bar{x}, \bar{y}) \setminus \mathcal{I}_F(x^{k_l}, y^{k_l}), \quad \forall l$$

or

$$i_0 \notin \mathcal{I}_F(\bar{x},\bar{y}), \quad i_0 \in \mathcal{I}_G(\bar{x},\bar{y}) \setminus \mathcal{I}_G(x^{k_l},y^{k_l}), \quad \forall l,$$

then, by (3.21) and (3.22),

$$u_{i_0}^{k_l} = a_{k_l}(\frac{1}{k_l} - y_{i_0}^{k_l}), \quad \forall l \tag{3.31}$$

or

$$v_{i_0}^{k_l} = -a_{k_l}F_{i_0}(x^{k_l},y^{k_l}), \quad \forall l \tag{3.32}$$

holds. Since

$$\lim_{l\to\infty}(\frac{1}{k_l} - y_{i_0}^{k_l}) = -\bar{y}_{i_0} < 0$$

in the former case, or

$$\lim_{l\to\infty} F_{i_0}(x^{k_l},y^{k_l}) = F_{i_0}(\bar{x},\bar{y}) > 0$$

in the latter case, it follows from (3.31) or (3.32) that $\{a_{k_l}\}$ converges. Then we also have (3.27)–(3.29) and hence $(\bar{x},\bar{y})$ is a B-stationary point of problem (3.1).

(iib) Now suppose that

$$\{1,\cdots,m\} \setminus \mathcal{I}_F(\bar{x},\bar{y}) \subseteq \mathcal{I}_G(x^k,y^k) \tag{3.33}$$

and

$$\{1,\cdots,m\} \setminus \mathcal{I}_G(\bar{x},\bar{y}) \subseteq \mathcal{I}_F(x^k,y^k) \tag{3.34}$$

for all sufficiently large k. Then, since $F_j(x^k,y^k)=0$ for any $j \notin \mathcal{I}_G(\bar{x},\bar{y})$ and $y_i^k = 0$ for any $i \notin \mathcal{I}_F(\bar{x},\bar{y})$, (3.19) yields

$$\begin{aligned}\nabla f(x^k,y^k) \quad = \quad & \sum_{i\in\mathcal{I}_F(\bar{x},\bar{y})} u_i^k \nabla F_i(x^k,y^k) + \sum_{j\in\mathcal{I}_G(\bar{x},\bar{y})} v_j^k \nabla G_j(x^k,y^k) \\ & + \sum_{i\notin\mathcal{I}_F(\bar{x},\bar{y})} \frac{a_k}{k}\nabla F_i(x^k,y^k) - \sum_{l\in\mathcal{I}_g(\bar{x},\bar{y})} w_l^k \nabla g_l(x^k,y^k) - \sum_{r=1}^{q} \gamma_r^k \nabla h_r(x^k,y^k)\end{aligned} \tag{3.35}$$

for all sufficiently large k. If $F(\bar{x},\bar{y}) = 0$, i.e., $\mathcal{I}_F(\bar{x},\bar{y}) = \{1,\cdots,m\}$, then (3.35) implies that the limit $(\bar{x},\bar{y})$ of $\{(x^k,y^k)\}$ satisfies the weak stationarity condition (3.24) for problem (3.1). If $F(\bar{x},\bar{y}) \neq 0$, then there exists an index i such that $F_i(\bar{x},\bar{y}) > 0$ and $\bar{y}_i = 0$, which implies

$$\mathcal{I}_G(\bar{x},\bar{y}) \setminus \mathcal{I}_F(\bar{x},\bar{y}) \neq \emptyset$$

and

$$\sum_{i\in\mathcal{I}_G(\bar{x},\bar{y})\setminus\mathcal{I}_F(\bar{x},\bar{y})} F_i(\bar{x},\bar{y}) > 0. \tag{3.36}$$

By (3.33) and (3.34), for all sufficiently large k, we have

$$\begin{aligned} 0 &= \phi_0^k(x^k,y^k) \\ &= \sum_{i=1}^m (\frac{1}{k} - y_i^k)F_i(x^k,y^k) \\ &= \sum_{i\in\mathcal{I}_G(\bar{x},\bar{y})} (\frac{1}{k} - y_i^k)F_i(x^k,y^k) \\ &= \sum_{i\in\mathcal{I}_G(\bar{x},\bar{y})\cap\mathcal{I}_F(\bar{x},\bar{y})} (\frac{1}{k} - y_i^k)F_i(x^k,y^k) + \sum_{i\in\mathcal{I}_G(\bar{x},\bar{y})\setminus\mathcal{I}_F(\bar{x},\bar{y})} \frac{1}{k}F_i(x^k,y^k). \end{aligned} \tag{3.37}$$

For any $i \in \mathcal{I}_F(\bar{x},\bar{y}) \cap \mathcal{I}_G(\bar{x},\bar{y})$, it follows from (3.21) and (3.30) that

$$\begin{aligned} |a_k(\frac{1}{k} - y_i^k)F_i(x^k,y^k)| &= \begin{cases} 0, & i \in \mathcal{I}_F(x^k,y^k) \\ |u_i^k F_i(x^k,y^k)|, & i \in \mathcal{I}_F(\bar{x},\bar{y}) \setminus \mathcal{I}_F(x^k,y^k) \end{cases} \\ &\le |u_i^k F_i(x^k,y^k)| \end{aligned}$$

and so

$$\lim_{k\to\infty} a_k(\frac{1}{k} - y_i^k)F_i(x^k,y^k) = 0. \tag{3.38}$$

Hence, by (3.36), (3.37), and (3.38), we have

$$\lim_{k\to\infty} \frac{a_k}{k} = -\lim_{k\to\infty} \frac{N_k}{D_k} = 0, \tag{3.39}$$

where

$$N_k = \sum_{i\in\mathcal{I}_G(\bar{x},\bar{y})\cap\mathcal{I}_F(\bar{x},\bar{y})} a_k(\frac{1}{k} - y_i^k)F_i(x^k,y^k)$$

and

$$D_k = \sum_{i\in\mathcal{I}_G(\bar{x},\bar{y})\setminus\mathcal{I}_F(\bar{x},\bar{y})} F_i(x^k,y^k).$$

Therefore, taking a limit in (3.35), we obtain (3.24) from (3.39). Now we proceed to showing

$$\bar{\lambda}_i\bar{\mu}_i \ge 0, \quad i \in \mathcal{I}_F(\bar{x},\bar{y}) \cap \mathcal{I}_G(\bar{x},\bar{y}), \tag{3.40}$$

i.e., $(\bar{x},\bar{y})$ is C-stationary. Let $i \in \mathcal{I}_F(\bar{x},\bar{y}) \cap \mathcal{I}_G(\bar{x},\bar{y})$. Note that, by the assumption of (ii),

$$kF_i(x^k,y^k) = \phi_i^k(x^k,y^k) - \phi_0^k(x^k,y^k) = \phi_i^k(x^k,y^k) \ge 0, \quad i = 1,\cdots,m,$$

i.e.,

$$F(x^k, y^k) \geq 0 \tag{3.41}$$

for all sufficiently large k. Suppose that there exists a subsequence $\{k_l\}$ of $\{k\}$ such that

$$y_i^{k_l} F_i(x^{k_l}, y^{k_l}) \neq 0, \quad \forall l.$$

It follows from (3.21) and (3.22) that

$$u_i^{k_l} = a_{k_l}(\frac{1}{k_l} - y_i^{k_l}), \quad v_i^{k_l} = -a_{k_l} F_i(x^{k_l}, y^{k_l}).$$

By (3.39) and (3.41), we have

$$\lim_{k\to\infty} u_i^k v_i^k = \lim_{l\to\infty} u_i^{k_l} v_i^{k_l} = \lim_{l\to\infty} a_{k_l}^2 y_i^{k_l} F_i(x^{k_l}, y^{k_l}) \geq 0. \tag{3.42}$$

Next we suppose that

$$y_i^k F_i(x^k, y^k) = 0 \tag{3.43}$$

for all sufficiently large k. First consider the case where there exists a subsequence $\{k_l\}$ of $\{k\}$ such that

$$y_i^{k_l} \neq 0, \quad \forall l.$$

Then, by (3.43) and (3.22), $F_i(x^{k_l}, y^{k_l}) = 0$ and hence $v_i^{k_l} = 0$ for any sufficiently large l. So we obtain

$$\lim_{k\to\infty} u_i^k v_i^k = \lim_{l\to\infty} u_i^{k_l} v_i^{k_l} = 0. \tag{3.44}$$

Next consider the case where $y_i^k = 0$ for all sufficiently large k. If there exists a subsequence $\{k_l\}$ of $\{k\}$ such that

$$F_i(x^{k_l}, y^{k_l}) \neq 0, \quad \forall l,$$

then, by (3.21) and (3.39), we have

$$\lim_{l\to\infty} u_i^{k_l} = \lim_{l\to\infty} \frac{a_{k_l}}{k_l} = 0$$

and so (3.44) also holds. If for any sufficiently large k,

$$y_i^k = 0, \quad F_i(x^k, y^k) = 0,$$

then, by (3.16), (3.21)–(3.22), and (3.39),

$$\lim_{k\to\infty} u_i^k v_i^k = \lim_{k\to\infty} \mu_i^k(k\lambda_i^k + \frac{a_k}{k}) = \lim_{k\to\infty} k\lambda_i^k\mu_i^k \geq 0.$$

Therefore, we always have $\lim_{k\to\infty} u_i^k v_i^k \geq 0$, i.e., $(\bar{x}, \bar{y})$ is a C-stationary point of problem (3.1). Moreover, if $(\bar{x}, \bar{y})$ is nondegenerate, then it readily follows from the definitions of the weak stationarity and nondegeneracy that $(\bar{x}, \bar{y})$ is B-stationary to problem (3.1).

(IIb) Consider the case where $\mathcal{I}_F(\bar{x}, \bar{y}) = \emptyset$. Then $\mathcal{I}_G(\bar{x}, \bar{y}) = \{1, \cdots, m\}$ and so $(\bar{x}, \bar{y})$ is nondegenerate. Moreover, (3.19) becomes

$$\begin{aligned}\nabla f(x^k, y^k) &= \sum_{i=1}^m a_k(\frac{1}{k} - y_i^k)\nabla F_i(x^k, y^k) + \sum_{j=1}^m v_j^k \nabla G_j(x^k, y^k) \\ &\quad - \sum_{l\in\mathcal{I}_g(\bar{x},\bar{y})} w_l^k \nabla g_l(x^k, y^k) - \sum_{r=1}^q \gamma_r^k \nabla h_r(x^k, y^k). \qquad (3.45)\end{aligned}$$

For any sufficiently large k, since $(x^k, y^k) \notin \mathcal{F}$, there exists an index $j \in \mathcal{I}_G(\bar{x}, \bar{y}) \setminus \mathcal{I}_G(x^k, y^k)$. Therefore, we can choose an index j_0 and a subsequence $\{k_l\}$ of $\{k\}$ such that

$$j_0 \in \mathcal{I}_G(\bar{x}, \bar{y}) \setminus \mathcal{I}_G(x^{k_l}, y^{k_l}), \quad \forall l,$$

i.e., by (3.22),

$$v_{j_0}^{k_l} = -a_{k_l} F_{j_0}(x^{k_l}, y^{k_l}), \quad \forall l.$$

Since $\{v_{j_0}^{k_l}\}$ converges and, by $\mathcal{I}_F(\bar{x}, \bar{y}) = \emptyset$,

$$\lim_{l\to\infty} F_{j_0}(x^{k_l}, y^{k_l}) > 0,$$

it follows that the sequence $\{a_{k_l}\}$ is convergent. Noticing that $\{y^k\}$ tends to $\bar{y} = 0$ as $k \to \infty$, we have that, for each j,

$$\lim_{k\to\infty} a_k(\frac{1}{k} - y_j^k) = \lim_{l\to\infty} a_{k_l}(\frac{1}{k_l} - y_j^{k_l}) = 0.$$

Letting $k \to \infty$ in (3.45) and denoting

$$\bar{v}_j = \lim_{k\to\infty} v_j^k, \quad \bar{w}_l = \lim_{k\to\infty} w_l^k, \quad \bar{\gamma}_r = \lim_{k\to\infty} \gamma_r^k,$$

we obtain

$$\nabla f(\bar{x}, \bar{y}) = \sum_{j=1}^m \bar{v}_j \nabla G_j(x^k, y^k) - \sum_{l\in\mathcal{I}_g(\bar{x},\bar{y})} \bar{w}_l \nabla g_l(\bar{x}, \bar{y}) - \sum_{r=1}^q \bar{\gamma}_r \nabla h_r(\bar{x}, \bar{y}).$$

This, together with

$$\mathcal{I}_F(\bar{x},\bar{y}) \cap \mathcal{I}_G(\bar{x},\bar{y}) = \emptyset,$$

implies that $(\bar{x},\bar{y})$ is a B-stationary point of problem (3.1).

Combining case IIa(ii) and case IIb shows that statement (a) holds. This completes the proof. ∎

For a sequence $\{(x^k,y^k)\}$ of stationary points of problem (3.3), let us define

$$\begin{aligned} \mathcal{I}_1 &:= \{i \mid \ y_i^k > 0 \ \text{ for infinitely many } k\}, \\ \mathcal{I}_2 &:= \{i \mid \ F_i(x^k,y^k) \neq 0 \ \text{ for infinitely many } k\}. \end{aligned}$$

Then we have

$$\begin{aligned} \{1,\cdots,m\} \setminus \mathcal{I}_G(\bar{x},\bar{y}) &\subseteq \mathcal{I}_1 \cap \mathcal{I}_F(\bar{x},\bar{y}), \\ \{1,\cdots,m\} \setminus \mathcal{I}_F(\bar{x},\bar{y}) &\subseteq \mathcal{I}_2 \cap \mathcal{I}_G(\bar{x},\bar{y}). \end{aligned}$$

From the proof of Theorem 3.7, we have the next corollary immediately.

Corollary 3.1 *Let the assumptions in Theorem 3.7 be satisfied. If $\mathcal{I}_1 \setminus \mathcal{I}_F(\bar{x},\bar{y}) \neq \emptyset$ or $\mathcal{I}_2 \setminus \mathcal{I}_G(\bar{x},\bar{y}) \neq \emptyset$, then $(\bar{x},\bar{y})$ is a B-stationary point of problem (3.1).*

Next we consider some other sufficient conditions on M- and B-stationarity for problem (3.1). We say $(x^k,y^k) \in \mathcal{F}_k$ satisfies the *second-order necessary conditions* if there exist multiplier vectors $\lambda^k \in \Re^{m+1}, \mu^k \in \Re^m, \gamma^k \in \Re^q$, and $\delta^k \in \Re^p$ such that

$$\lambda^k \geq 0, \quad \mu^k \geq 0, \quad \delta^k \geq 0, \tag{3.46}$$

$$(\lambda^k)^T \phi^k(x^k,y^k) = 0, \ (\mu^k)^T G(x^k,y^k) = 0, \ (\delta^k)^T g(x^k,y^k) = 0, \tag{3.47}$$

$$\nabla_{(x,y)} L_k(x^k,y^k,\lambda^k,\mu^k,\delta^k,\gamma^k) = 0, \tag{3.48}$$

and

$$d^T \nabla^2_{(x,y)} L_k(x^k,y^k,\lambda^k,\mu^k,\delta^k,\gamma^k) d \geq 0, \quad \forall d \in \mathcal{T}_k(x^k,y^k), \tag{3.49}$$

where

$$L_k(x,y,\lambda,\mu,\delta,\gamma) := f(x,y) - \lambda^T \phi^k(x,y) - \mu^T G(x,y) + \delta^T g(x,y) + \gamma^T h(x,y)$$

stands for the Lagrangian of problem (3.3) and for $(x,y) \in \mathcal{F}_k$,

$$\begin{aligned} \mathcal{T}_k(x,y) := \Big\{ d \in \Re^{n+m} \ \Big| \quad & d^T \nabla \phi_i^k(x,y) = 0, \ i \in \mathcal{I}_{\phi^k}(x,y); \\ & d^T \nabla G_j(x,y) = 0, \ j \in \mathcal{I}_G(x,y); \\ & d^T \nabla g_l(x,y) = 0, \ l \in \mathcal{I}_g(x,y); \\ & d^T \nabla h_r(x,y) = 0, \ r = 1,2,\cdots,q \Big\}. \end{aligned}$$

We next introduce a new kind of conditions weaker than the second-order necessary conditions for problem (3.3). Suppose that α_k is a nonnegative number. We say that, at a stationary point (x^k, y^k) of problem (3.3), the matrix $\nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k)$ is *bounded below* with constant α_k on the corresponding tangent space $\mathcal{T}_k(x^k, y^k)$ if

$$d^T \nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k) d \geq -\alpha_k ||d||^2, \quad \forall d \in \mathcal{T}_k(x^k, y^k). \tag{3.50}$$

Condition (3.50) is clearly weaker than (3.49). In fact, for the matrix $\nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k)$, there must exist a number α_k such that (3.50) hold. For example, any nonnegative number α such that $-\alpha$ is less than the smallest eigenvalue of $\nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k)$ must satisfy (3.50). However, condition (3.49) means that the matrix $\nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k)$ should have some kind of semi-definiteness on the tangent space $\mathcal{T}_k(x^k, y^k)$. Note that, in (3.50), the constant $-\alpha_k$ may be larger than the smallest eigenvalue mentioned above.

Theorem 3.8 *Let $(x^k, y^k) \in \mathcal{F}_k$ be a stationary point of problem (3.3) with multiplier vectors $\lambda^k, \mu^k, \delta^k$, and γ^k satisfying conditions (3.46)–(3.48) and, for each k, $\nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k)$ be bounded below with constant α_k on the corresponding tangent space $\mathcal{T}_k(x^k, y^k)$. Suppose that $(\bar{x}, \bar{y})$ is an accumulation point of the sequence $\{(x^k, y^k)\}$ with $F(\bar{x}, \bar{y}) \neq 0$, the sequence $\{\alpha_k\}$ is bounded, and the set of vectors*

$$\Big\{ \nabla F_i(\bar{x}, \bar{y}), \nabla G_i(\bar{x}, \bar{y}), \nabla g_l(\bar{x}, \bar{y}), \nabla h_r(\bar{x}, \bar{y}) \ \Big| \ i = 1, \cdots, m, \ l \in \mathcal{I}_g(\bar{x}, \bar{y}), \ r = 1, \cdots, q \Big\}$$

is linearly independent. Then $(\bar{x}, \bar{y})$ is an M-stationary point of problem (3.1). Furthermore, if $(\bar{x}, \bar{y})$ satisfies the upper level strict complementarity condition, it is B-stationary to problem (3.1).

Proof: Since (3.46)–(3.48) are equivalent to (3.15) and (3.16), it follows from Theorem 3.7 (a) that $(\bar{x}, \bar{y})$ is a C-stationary point of problem (3.1). By the proof of Theorem 3.7, $(\bar{x}, \bar{y})$ is not B-stationary only in the case IIa(iib), i.e., for all sufficiently large k,

$$(x^k, y^k) \notin \mathcal{F}, \ \mathcal{I}_F(\bar{x}, \bar{y}) \neq \emptyset, \tag{3.51}$$
$$\mathcal{I}_{\phi^k}(x^k, y^k) = \{0\} \cup \mathcal{I}_F(x^k, y^k), \tag{3.52}$$
$$\{1, 2, \cdots, m\} \setminus \mathcal{I}_F(\bar{x}, \bar{y}) \subseteq \mathcal{I}_G(x^k, y^k), \tag{3.53}$$
$$\{1, 2, \cdots, m\} \setminus \mathcal{I}_G(\bar{x}, \bar{y}) \subseteq \mathcal{I}_F(x^k, y^k). \tag{3.54}$$

In the rest of the proof, we therefore assume (3.51)–(3.54) and use the same setting as in the proof of Theorem 3.7. Then (3.19) holds with (3.20)–(3.22). Suppose that $(\bar{x}, \bar{y})$ is not an M-stationary point of problem (3.1). Then, by the definitions of C-stationarity and M-stationarity, there exists an $i_0 \in \mathcal{I}_F(\bar{x}, \bar{y}) \cap \mathcal{I}_G(\bar{x}, \bar{y})$ such that

$$\bar{u}_{i_0} = \lim_{k\to\infty} u_{i_0}^k < 0, \quad \bar{v}_{i_0} = \lim_{k\to\infty} v_{i_0}^k < 0, \tag{3.55}$$

where we use the fact that both the sequences $\{u_{i_0}^k\}$ and $\{v_{i_0}^k\}$ are convergent.

We claim that

$$y_{i_0}^k F_{i_0}(x^k, y^k) \neq 0 \tag{3.56}$$

for all sufficiently large k. In fact, if there exists a subsequence $\{k_l\}$ of $\{k\}$ such that

$$y_{i_0}^{k_l} F_{i_0}(x^{k_l}, y^{k_l}) = 0, \quad \forall l,$$

namely,

$$y_{i_0}^{k_l} = 0 \quad \text{or} \quad F_{i_0}(x^{k_l}, y^{k_l}) = 0, \quad \forall l, \tag{3.57}$$

then we have from (3.21)–(3.22) and (3.57) that

$$u_{i_0}^{k_l} \geq 0 \quad \text{or} \quad v_{i_0}^{k_l} \geq 0, \quad \forall l.$$

This contradicts (3.55), and so (3.56) holds for all sufficiently large k. Then (3.55) becomes

$$\bar{u}_{i_0} = \lim_{k\to\infty} u_{i_0}^k = \lim_{k\to\infty} a_k(\frac{1}{k} - y_{i_0}^k) < 0, \tag{3.58}$$

$$\bar{v}_{i_0} = \lim_{k\to\infty} v_{i_0}^k = -\lim_{k\to\infty} a_k F_{i_0}(x^k, y^k) < 0 \tag{3.59}$$

by (3.21) and (3.22). By Theorem 3.3 (ii), we may suppose that k is sufficiently large so that for any k,

$$\Big\{\nabla F_i(x^k, y^k), \nabla G_i(x^k, y^k), \nabla g_l(x^k, y^k), \nabla h_r(x^k, y^k) \;\Big|\; i = 1, \cdots, m,\ l \in \mathcal{I}_g(\bar{x}, \bar{y}),\ r = 1, \cdots, q\Big\}$$

is linearly independent. Note that

$$\lim_{k\to\infty} \frac{\frac{1}{k} - y_{i_0}^k}{F_{i_0}(x^k, y^k)} = -\frac{\bar{u}_{i_0}}{\bar{v}_{i_0}} < 0 \tag{3.60}$$

by (3.58) and (3.59). Therefore, we can choose a bounded sequence $\{d^k\} \subseteq \Re^{n+m}$ such that for all sufficiently large k,

$$(d^k)^T \nabla F_i(x^k, y^k) = 0, \quad i = 1, \cdots, m,\ i \neq i_0; \tag{3.61}$$

$$(d^k)^T\nabla G_j(x^k,y^k)=0,\ j=1,\cdots,m,\ j\neq i_0; \tag{3.62}$$

$$(d^k)^T\nabla F_{i_0}(x^k,y^k)=1; \tag{3.63}$$

$$(d^k)^T\nabla G_{i_0}(x^k,y^k)=\frac{\frac{1}{k}-y^k_{i_0}}{F_{i_0}(x^k,y^k)}; \tag{3.64}$$

$$(d^k)^T\nabla g_l(x^k,y^k)=0,\ l\in\mathcal{I}_g(\bar{x},\bar{y}); \tag{3.65}$$

$$(d^k)^T\nabla h_r(x^k,y^k)=0,\ r=1,2,\cdots,q. \tag{3.66}$$

Since

$$\nabla\phi^k_0(x^k,y^k)=\sum_{i=1}^m(\frac{1}{k}-y^k_i)\nabla F_i(x^k,y^k)-\sum_{j=1}^m F_j(x^k,y^k)\nabla G_j(x^k,y^k), \tag{3.67}$$

we have from (3.61)–(3.64) that

$$\begin{aligned}&(d^k)^T\nabla\phi^k_0(x^k,y^k)\\&=\sum_{i=1}^m(\frac{1}{k}-y^k_i)(d^k)^T\nabla F_i(x^k,y^k)-\sum_{j=1}^m F_j(x^k,y^k)(d^k)^T\nabla G_j(x^k,y^k)=0.\end{aligned} \tag{3.68}$$

On the other hand, noting that $i_0\notin\mathcal{I}_F(x^k,y^k)\cup\mathcal{I}_G(x^k,y^k)$ for all sufficiently large k by (3.56), we have from (3.5), (3.61), and (3.68) that

$$\begin{aligned}(d^k)^T\nabla\phi^k_i(x^k,y^k)&=(d^k)^T\nabla\phi^k_0(x^k,y^k)+(d^k)^T\nabla F_i(x^k,y^k)\\&=0,\qquad\qquad 0\neq i\in\mathcal{I}_{\phi^k}(x^k,y^k).\end{aligned} \tag{3.69}$$

It follows from (3.62), (3.64)–(3.66), and (3.68)–(3.69) that

$$d^k\in\mathcal{T}_k(x^k,y^k) \tag{3.70}$$

for all sufficiently large k. By (3.67), we have

$$\begin{aligned}\nabla^2\phi^k_0(x^k,y^k)&=\sum_{i=1}^m(\frac{1}{k}-y^k_i)\nabla^2F_i(x^k,y^k)-\sum_{i=1}^m\nabla F_i(x^k,y^k)G_i(x^k,y^k)^T\\&\quad-\sum_{j=1}^m\nabla G_j(x^k,y^k)\nabla F_j(x^k,y^k)^T,\end{aligned} \tag{3.71}$$

where we use the fact that $\nabla G_j(x^k,y^k), j=1,\cdots,m$, are constant vectors. On the other hand, we can write

$$\begin{aligned}\nabla_{(x,y)}L_k(x^k,y^k,\lambda^k,\mu^k,\delta^k,\gamma^k)&=\nabla f(x^k,y^k)-\sum_{i=0}^m\lambda^k_i\nabla\phi^k_i(x^k,y^k)-\sum_{j=1}^m\mu^k_j\nabla G_j(x^k,y^k)\\&\quad+\sum_{l=1}^p\delta^k_l\nabla g_l(x^k,y^k)+\sum_{r=1}^q\gamma^k_r\nabla h_r(x^k,y^k)\end{aligned}$$

$$\begin{aligned} &= \nabla f(x^k, y^k) - a_k \nabla \phi_0^k(x^k, y^k) \\ &\quad - \sum_{i=1}^m k\lambda_i^k \nabla F_i(x^k, y^k) - \sum_{j=1}^m \mu_j^k \nabla G_j(x^k, y^k) \\ &\quad + \sum_{l=1}^p \delta_l^k \nabla g_l(x^k, y^k) + \sum_{r=1}^q \gamma_r^k \nabla h_r(x^k, y^k), \end{aligned}$$

where $a_k = \sum_{i=0}^m \lambda_i^k$ is the same as that in the proof of Theorem 3.7, and so we have from (3.71) that

$$\begin{aligned} \nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k) &= \nabla^2 f(x^k, y^k) - a_k \nabla^2 \phi_0^k(x^k, y^k) - \sum_{i=1}^m k\lambda_i^k \nabla^2 F_i(x^k, y^k) \\ &\quad + \sum_{l=1}^p \delta_l^k \nabla^2 g_l(x^k, y^k) + \sum_{r=1}^q \gamma_r^k \nabla^2 h_r(x^k, y^k) \\ &= \nabla^2 f(x^k, y^k) + a_k \sum_{i=1}^m \nabla F_i(x^k, y^k) G_i(x^k, y^k)^T \\ &\quad + a_k \sum_{j=1}^m \nabla G_j(x^k, y^k) \nabla F_j(x^k, y^k)^T \\ &\quad - \sum_{i=1}^m (k\lambda_i^k + a_k(\frac{1}{k} - y_i^k)) \nabla^2 F_i(x^k, y^k) \\ &\quad + \sum_{l=1}^p \delta_l^k \nabla^2 g_l(x^k, y^k) + \sum_{r=1}^q \gamma_r^k \nabla^2 h_r(x^k, y^k). \end{aligned}$$

Since $\nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k)$ is bounded below with constant α_k on the corresponding tangent space $\mathcal{T}_k(x^k, y^k)$, we have from (3.50) and (3.70) that there exists a constant C such that

$$(d^k)^T \nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k) d^k \geq -\alpha_k ||d^k||^2 \geq C, \tag{3.72}$$

where the last inequality follows from the boundedness of the sequences $\{\alpha_k\}$ and $\{d^k\}$. Note that

$$\begin{aligned} &(d^k)^T \nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k) d^k \\ &\quad = (d^k)^T \nabla^2 f(x^k, y^k) d^k + a_k \sum_{i=1}^m (d^k)^T \nabla F_i(x^k, y^k) G_i(x^k, y^k)^T d^k \\ &\quad + a_k \sum_{j=1}^m (d^k)^T \nabla G_j(x^k, y^k) \nabla F_j(x^k, y^k)^T d^k \\ &\quad - \sum_{i=1}^m (k\lambda_i^k + a_k(\frac{1}{k} - y_i^k))(d^k)^T \nabla^2 F_i(x^k, y^k) d^k \end{aligned}$$

$$
\begin{aligned}
&+\sum_{l=1}^{p}\delta_l^k(d^k)^T\nabla^2 g_l(x^k,y^k)d^k+\sum_{r=1}^{q}\gamma_r^k(d^k)^T\nabla^2 h_r(x^k,y^k)d^k\\
=\;&(d^k)^T\nabla^2 f(x^k,y^k)d^k+\frac{2a_k(\frac{1}{k}-y_{i_0}^k)}{F_{i_0}(x^k,y^k)}\\
&-\sum_{i=1}^{m}(k\lambda_i^k+a_k(\frac{1}{k}-y_i^k))(d^k)^T\nabla^2 F_i(x^k,y^k)d^k\\
&+\sum_{l=1}^{p}\delta_l^k(d^k)^T\nabla^2 g_l(x^k,y^k)d^k+\sum_{r=1}^{q}\gamma_r^k(d^k)^T\nabla^2 h_r(x^k,y^k)d^k. \qquad (3.73)
\end{aligned}
$$

By the twice continuous differentiability of the functions involved, the boundedness of the sequence $\{d^k\}$, and the convergence of the sequences $\{(x^k,y^k)\}$, $\{\delta_l^k\}$, and $\{\gamma_r^k\}$, the terms

$$
(d^k)^T\nabla^2 f(x^k,y^k)d^k,\quad \sum_{l=1}^{p}\delta_l^k(d^k)^T\nabla^2 g_l(x^k,y^k)d^k,\quad \sum_{r=1}^{q}\gamma_r^k(d^k)^T\nabla^2 h_r(x^k,y^k)d^k
$$

are all bounded. Noticing that, for all sufficiently large k, $i\notin\mathcal{I}_F(x^k,y^k)$ implies $i\notin\mathcal{I}_{\phi^k}(x^k,y^k)$ by (3.52) and so $\lambda_i^k=0$ by (3.46) and (3.47), we have from the convergence of the sequences in (3.23) and the definition (3.21) of u_i^k that the sequence $\{k\lambda_i^k+a_k(\frac{1}{k}-y_i^k)\}$ is bounded for any $i=1,2,\cdots,m$. Hence, the term

$$
\sum_{i=1}^{m}(k\lambda_i^k+a_k(\frac{1}{k}-y_i^k))(d^k)^T\nabla^2 F_i(x^k,y^k)d^k
$$

is also bounded. However, since $\lim_{k\to\infty}y_{i_0}^k=\bar{y}_{i_0}=0$, we have $a_k\to+\infty$ by (3.58) and so

$$
\frac{2a_k(\frac{1}{k}-y_{i_0}^k)}{F_{i_0}(x^k,y^k)}\to-\infty
$$

as $k\to\infty$ by (3.60). Therefore, it follows from (3.73) that

$$
(d^k)^T\nabla^2_{(x,y)}L_k(x^k,y^k,\lambda^k,\mu^k,\delta^k,\gamma^k)d^k\to-\infty
$$

as $k\to\infty$. This contradicts (3.72) and hence $(\bar{x},\bar{y})$ is M-stationary to problem (3.1). This completes the proof of the first part of the theorem. The second part of the theorem follows from the definitions of M-stationarity and the upper level strict complementarity immediately. ∎

Corollary 3.2 *Let $\{(x^k,y^k)\}$ and $(\bar{x},\bar{y})$ be the same as in Theorem 3.8. If (x^k,y^k) together with the corresponding multiplier vectors λ^k,μ^k,δ^k, and γ^k satisfies the second-order necessary conditions (3.46)–(3.49) and the set of vectors*

$$
\Big\{\nabla F_i(\bar{x},\bar{y}),\nabla G_i(\bar{x},\bar{y}),\nabla g_l(\bar{x},\bar{y}),\nabla h_r(\bar{x},\bar{y})\;\Big|\;i=1,\cdots,m,l\in\mathcal{I}_g(\bar{x},\bar{y}),r=1,\cdots,q\Big\}
$$

is linearly independent, then the conclusion of Theorem 3.8 remains true.

Corollary 3.2 establishes convergence to a B-stationary point under the second-order necessary conditions and the upper level strict complementarity. These or similar conditions have also been assumed in [31, 62], but they are somewhat restrictive and may be difficult to verify in practice. The next theorem provides a new condition for convergence to a B-stationary point, which can be dealt with more easily. We note that, unlike [31, 62], it relies on neither the upper level strict complementarity nor the asymptotic weak nondegeneracy.

Theorem 3.9 *Let $\{(x^k, y^k)\}$ and $(\bar{x}, \bar{y})$ be the same as in Theorem 3.8 and $\lambda^k, \mu^k, \delta^k$, and γ^k be the multiplier vectors corresponding to (x^k, y^k) with (3.46)–(3.49). Let β_k be the smallest eigenvalue of the matrix $\nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k)$. If the sequence $\{\beta_k\}$ is bounded below and*

$$\Big\{\nabla F_i(\bar{x}, \bar{y}), \nabla G_i(\bar{x}, \bar{y}), \nabla g_l(\bar{x}, \bar{y}), \nabla h_r(\bar{x}, \bar{y}) \;\Big|\; i = 1, 2, \cdots, m, l \in \mathcal{I}_g(\bar{x}, \bar{y}), r = 1, 2, \cdots, q\Big\}$$

is linearly independent, then $(\bar{x}, \bar{y})$ is a B-stationary point of problem (3.1).

Proof: It is easy to see that the assumptions of Theorem 3.8 are satisfied with $\alpha_k = \max\{-\beta_k, 0\}$ and so $(\bar{x}, \bar{y})$ is an M-stationary point of problem (3.1). Suppose that $(\bar{x}, \bar{y})$ is not B-stationary to problem (3.1). As mentioned at the beginning of the proof of Theorem 3.8, this occurs only in the case where (3.51)–(3.54) hold for all sufficiently large k. By the definitions of B- and M-stationarity, there exists an $i_0 \in \mathcal{I}_F(\bar{x}, \bar{y}) \cap \mathcal{I}_G(\bar{x}, \bar{y})$ such that

$$\bar{u}_{i_0} = \lim_{k\to\infty} u^k_{i_0} < 0, \quad \bar{v}_{i_0} = \lim_{k\to\infty} v^k_{i_0} = 0 \tag{3.74}$$

or

$$\bar{u}_{i_0} = \lim_{k\to\infty} u^k_{i_0} = 0, \quad \bar{v}_{i_0} = \lim_{k\to\infty} v^k_{i_0} < 0. \tag{3.75}$$

From (3.20)–(3.22) and (3.46), we know that either of (3.74) and (3.75) implies

$$\lim_{k\to\infty} a_k = +\infty. \tag{3.76}$$

By Theorem 3.3, we may suppose that k is large enough so that (3.51)–(3.54) hold,

$$\mathcal{I}_F(x^k, y^k) \subseteq \mathcal{I}_F(\bar{x}, \bar{y}), \quad \mathcal{I}_G(x^k, y^k) \subseteq \mathcal{I}_G(\bar{x}, \bar{y}), \quad \mathcal{I}_g(x^k, y^k) \subseteq \mathcal{I}_g(\bar{x}, \bar{y}),$$

and

$$\Big\{\nabla F_i(x^k, y^k), \nabla G_i(x^k, y^k), \nabla g_l(x^k, y^k), \nabla h_r(x^k, y^k) \;\Big|\; i = 1, \cdots, m,\ l \in \mathcal{I}_g(\bar{x}, \bar{y}),\ r = 1, \cdots, q\Big\}$$

is linearly independent. Therefore, we can choose a vector $d^k \in \Re^{n+m}$ such that (3.61)–(3.62) and (3.65)–(3.66) hold and

$$(d^k)^T \nabla F_{i_0}(x^k, y^k) = 1, \quad (d^k)^T \nabla G_{i_0}(x^k, y^k) = -1.$$

Furthermore, we can choose the sequence $\{d^k\}$ to be bounded. By the assumptions of the theorem, there exists a constant C such that

$$(d^k)^T \nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k) d^k \geq \beta_k ||d^k||^2 \geq C \tag{3.77}$$

holds for all sufficiently large k. Note that, by the definition of d^k and (3.73),

$$\begin{aligned} &(d^k)^T \nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k) d^k \\ &\quad = (d^k)^T \nabla^2 f(x^k, y^k) d^k - 2a_k - \sum_{i=1}^m (k\lambda_i^k + a_k(\frac{1}{k} - y_i^k))(d^k)^T \nabla^2 F_i(x^k, y^k) d^k \\ &\qquad + \sum_{l=1}^p \delta_l^k (d^k)^T \nabla^2 g_l(x^k, y^k) d^k + \sum_{r=1}^q \gamma_r^k (d^k)^T \nabla^2 h_r(x^k, y^k) d^k. \end{aligned} \tag{3.78}$$

In a way similar to Theorem 3.8, we can show that all the terms on the right-hand side of (3.78) except the term $(-2a_k)$ are bounded. This, together with (3.76), implies that

$$(d^k)^T \nabla^2_{(x,y)} L_k(x^k, y^k, \lambda^k, \mu^k, \delta^k, \gamma^k) d^k \to -\infty$$

as $k \to \infty$. This contradicts (3.77) and hence $(\bar{x}, \bar{y})$ is B-stationary to problem (3.1). This completes the proof. ∎

3.3 Computational Results

We have tested the method on various small scale examples of MPECs, which have been used to test other methods in the literature. We applied the MATLAB 6.0 built-in solver function $fmincon$ to problem (3.3) with various values of k. The computational results are summarized in Tables 3.1–3.4, which indicate the proposed method produces good approximate solutions of (3.1) in a small number of iterations. In the tables, (x^k, y^k) is the (approximate) solution of (3.1) produced by solving (3.3), Ite stands for the number of iterations spent by $fmincon$, and $r(x^k, y^k)$ denotes the residual for the constraints in problem (3.3) at (x^k, y^k), i.e.,

$$\begin{aligned} r(x^k, y^k) \quad := \quad & \sum_{l=1}^p (g_l(x^k, y^k))_+ + \sum_{r=1}^q |h_r(x^k, y^k)| + \sum_{j=1}^m (-y_j^k)_+ \\ & + \sum_{i=1}^m (-F_i(x^k, y^k))_+ + |(y^k)^T F(x^k, y^k)|. \end{aligned}$$

Table 3.1: Computational results for Problem 3.1

Size (p, m, n)		(1, 1, 2)
Initial point		(3, 0, 0)
$k = 10, 10^2$	(x^k, y^k)	(2.7101, 0.5365, 0)
	Ite	7
	$f(x^k, y^k)$	10.4925
	$r(x^k, y^k)$	0

Table 3.2: Computational results for Problem 3.2

Size (m, n, p, q)		(2, 4, 4, 2)
Initial point		(1, 1, 1, 1, 0, 0)
$k = 10, 10^2, 10^4$	(x^k, y^k)	(0.5000, 0.5000, 0.5000, 0.5000, 0, 0)
	Ite	2
	$f(x^k, y^k)$	-1.0000
	$r(x^k, y^k)$	0

Problem 3.1. This problem is given in [77], which has two upper-level variables $(x_1, x_2) \in \Re^2$ and one lower-level variable $y \in \Re$:

$$\begin{array}{ll} \text{minimize} & x_1^2 + 10(x_2 - 1)^2 + (y + 1)^2 \\ \text{subject to} & x_2 \geq 0,\ x_1 - e^{x_2} - e^y \geq 0 \\ & y \geq 0,\ y(x_1 - e^{x_2} - e^y) = 0. \end{array}$$

Problem 3.2. This is equivalent to Problem 5 in [24] and goes back to [67].

$$\begin{array}{ll} \text{minimize} & x_1^2 - 2x_1 + x_2^2 - 2x_2 + x_3^2 + x_4^2 \\ \text{subject to} & 0 \leq x_1 \leq 2,\ 0 \leq x_2 \leq 2 \\ & x_3 - x_1 + x_3 y_1 - y_1 = 0 \\ & x_4 - x_2 + x_4 y_2 - y_2 = 0 \\ & y_1 \geq 0,\ y_2 \geq 0 \\ & F(x, y) \geq 0,\ y^T F(x, y) = 0, \end{array}$$

where

$$F(x, y) = \begin{pmatrix} 0.25 - (x_3 - 1)^2 \\ 0.25 - (x_4 - 1)^2 \end{pmatrix}.$$

Table 3.3: Computational results for Problem 3.3

Size (p, m, n)		(3, 6, 2)
Initial point		(0, 0, 1.60, 0.20, 0.44, 1.36, 0, 0)
$k = 10^2$	(x^k, y^k)	(0, 2, 1.9034, 0.9276, 0, 1.2689, 0, 0)
	Ite	7
	$f(x^k, y^k)$	-12.7533
	$r(x^k, y^k)$	0.0703
$k = 10^4$	(x^k, y^k)	(0, 2, 1.8753, 0.9065, 0, 1.2502, 0, 0)
	Ite	6
	$f(x^k, y^k)$	-12.6795
	$r(x^k, y^k)$	0.0007
$k = 10^6, 10^8$	(x^k, y^k)	(0, 2, 1.8750, 0.9063, 0, 1.2500, 0, 0)
	Ite	6
	$f(x^k, y^k)$	-12.6787
	$r(x^k, y^k)$	0.0004

Problem 3.3. This is Problem 11 in [24], which is equivalent to the following MPEC:

$$\begin{array}{ll} \text{minimize} & -x_1^2 - 3x_2 - 4y_1 + y_2^2 \\ \text{subject to} & y \geq 0,\ F(x,y) \geq 0,\ y^T F(x,y) = 0 \\ & x_1^2 + 2x_2 \leq 4,\ x_1 \geq 0,\ x_2 \geq 0, \end{array}$$

where

$$F(x,y) = \begin{pmatrix} 2y_1 + 2y_3 - 3y_4 - y_5 \\ -5 - y_3 + 4y_4 - y_6 \\ x_1^2 - 2x_1 + x_2^2 - 2y_1 + y_2 + 3 \\ x_2 + 3y_1 - 4y_2 - 4 \\ y_1 \\ y_2 \end{pmatrix}.$$

Problem 3.4. This is equivalent to Problem 10 in [24].

$$\begin{array}{ll} \text{minimize} & (x_5 + x_7 - 200)(x_5 + x_7) + (x_6 + x_8 - 160)(x_6 + x_8) \\ \text{subject to} & 0 \leq x_1 \leq 10,\ 0 \leq x_2 \leq 5 \\ & 0 \leq x_3 \leq 15,\ 0 \leq x_4 \leq 20 \\ & x_1 + x_2 + x_3 + x_4 \leq 40 \end{array}$$

$$x_5 - 4 + 0.4y_1 + 0.6y_2 - y_3 + y_4 = 0$$
$$x_6 - 13 + 0.7y_1 + 0.3y_2 - y_5 + y_6 = 0$$
$$x_7 - 35 + 0.4y_7 + 0.6y_8 - y_9 + y_{10} = 0$$
$$x_8 - 2 + 0.7y_7 + 0.3y_8 - y_{11} + y_{12} = 0$$
$$y \geq 0,\ F(x,y) \geq 0,\ y^T F(x,y) = 0,$$

where

$$F(x,y) = \begin{pmatrix} x_1 - 0.4x_5 - 0.7x_6 \\ x_2 - 0.6x_5 - 0.3x_6 \\ x_5 \\ -x_5 + 20 \\ x_6 \\ -x_6 + 20 \\ x_3 - 0.4x_7 - 0.7x_8 \\ x_4 - 0.6x_7 - 0.3x_8 \\ x_7 \\ -x_7 + 40 \\ x_8 \\ -x_8 + 40 \end{pmatrix}.$$

Since y stands for the Lagrangian multiplier vector in the original problem [24], we only list the values of x in Table 3.4.

3.4 Concluding Remarks

Suppose that the condition $\sum_{j=1}^m y_j \leq \frac{m}{k} + k$ is retained in the constraints of problem (3.3), i.e., problem (3.3) is replaced by the problem

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & g(x,y) \leq 0,\ h(x,y) = 0, \\ & y \geq 0,\ \sum_{j=1}^m y_j \leq \dfrac{m}{k} + k, \\ & (e_i^k - y)^T F(x,y) \geq 0,\ i = 0, 1, \cdots, m. \end{array} \tag{3.79}$$

Then, since the constraint

$$\sum_{j=1}^m y_j \leq \frac{m}{k} + k$$

Table 3.4: Computational results for Problem 3.4

Size (m, n, p, q)		(12, 8, 9, 4)
x^0		(0, 0, 0, 0, 0, 0, 0, 0)
$k = 10^2$	x^k	(6.9858, 2.9766, 12.0064, 18.0312, -0.0173, 10.0143, 30.0896, -0.0173)
	Ite	19
	$f(x^k, y^k)$	-6.6097e+003
	$r(x^k, y^k)$	0.5500
$k = 10^4$	x^k	(7.0369, 3.0553, 11.9632, 17.9447, 0.0921, 10.0000, 29.9079, 0)
	Ite	20
	$f(x^k, y^k)$	-6.6000e+003
	$r(x^k, y^k)$	0.1953e-004
$k = 10^6$	x^k	(6.4449, 2.7621, 12.3172, 18.4758, 0, 9.2069, 30.7930, 0.0001)
	Ite	7
	$f(x^k, y^k)$	-6.5987e+003
	$r(x^k, y^k)$	0.1200e-003
$k = 10^8$	x^k	(6.4447, 2.7620, 12.3173, 18.4758, 0, 9.2068, 30.7932, 0)
	Ite	7
	$f(x^k, y^k)$	-6.5987e+003
	$r(x^k, y^k)$	0.1200e-005

eventually becomes inactive at any fixed point as k tends to ∞, all the results established in the previous sections remain true except that the results in Theorem 3.1 are replaced by $\mathcal{F} = \lim_{k\to\infty} \mathcal{F}_k$. When the set $Z := \{z \in \Re^{n+m} \mid \ g(z) \le 0, h(z) = 0\}$ is bounded, problem (3.79) has a compact feasible region and so it is solvable for any k as long as it is feasible.

In addition, we remark that the term $\frac{1}{k}e$ in (3.2) is necessary for problem (3.3) to have desirable properties. In fact, the problem

$$\begin{array}{ll} \text{minimize} & f(x, y) \\ \text{subject to} & g(x, y) \le 0,\ h(x, y) = 0,\ y \ge 0, \\ & (ke_i - y)^T F(x, y) \ge 0,\ i = 0, 1, \cdots, m \end{array} \tag{3.80}$$

is difficult to handle because problem (3.80) does not satisfy the MFCQ at any point $(\bar{x}, \bar{y}) \in \mathcal{F}$ for all sufficiently large k. For simplicity, we assume that the constraints

$g(x,y) \le 0$ and $h(x,y) = 0$ are absent and let

$$G(x,y) = y, \quad \psi_i^k(x,y) = (ke_i - y)^T F(x,y), \quad i = 0, 1, \cdots, m. \tag{3.81}$$

Note that

$$\psi_i^k(x,y) = kF_i(x,y) + \psi_0^k(x,y), \quad i = 1, 2, \cdots, m. \tag{3.82}$$

At $(\bar{x}, \bar{y}) \in \mathcal{F}$, the set of active constraints is

$$\Big\{\psi_0^k,\ \psi_i^k,\ G_j \ \Big|\ i \in \mathcal{I}_F(\bar{x},\bar{y}),\ j \in \mathcal{I}_G(\bar{x},\bar{y})\Big\}.$$

Suppose the MFCQ holds at $(\bar{x}, \bar{y})$ for problem (3.80). Then there exists a vector $(x, y) \in \Re^{n+m}$ such that

$$\nabla \psi_0^k(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} > 0, \tag{3.83}$$

$$\nabla \psi_i^k(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} > 0, \quad i \in \mathcal{I}_F(\bar{x},\bar{y}), \tag{3.84}$$

$$y_j = \nabla G_j(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} > 0, \quad j \in \mathcal{I}_G(\bar{x},\bar{y}). \tag{3.85}$$

(i) Assume that $\mathcal{I}_F(\bar{x},\bar{y}) \ne \emptyset$ and k is large enough to satisfy

$$1 - \frac{1}{k} \sum_{i \in \mathcal{I}_F(\bar{x},\bar{y})} \bar{y}_i > 0 \tag{3.86}$$

By (3.82) and (3.84), we have

$$-\nabla F_i(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} < \frac{1}{k} \nabla \psi_0^k(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix}, \quad i \in \mathcal{I}_F(\bar{x},\bar{y}). \tag{3.87}$$

It then follows from (3.81), (3.85), and (3.87) that

$$\begin{aligned}
\nabla \psi_0^k(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} &= -\sum_{i=1}^m \bar{y}_i \nabla F_i(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} - \sum_{j=1}^m F_j(\bar{x},\bar{y}) y_j \\
&= -\sum_{i \in \mathcal{I}_F(\bar{x},\bar{y})} \bar{y}_i \nabla F_i(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} - \sum_{j \in \mathcal{I}_G(\bar{x},\bar{y})} F_j(\bar{x},\bar{y}) y_j \\
&\le \Big(\frac{1}{k} \sum_{i \in \mathcal{I}_F(\bar{x},\bar{y})} \bar{y}_i\Big) \nabla \psi_0^k(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix},
\end{aligned}$$

i.e.,

$$\Big(1 - \frac{1}{k} \sum_{i \in \mathcal{I}_F(\bar{x},\bar{y})} \bar{y}_i\Big) \nabla \psi_0^k(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} \le 0.$$

By (3.86), we have

$$\nabla\psi_0^k(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} \leq 0.$$

This contradicts (3.83) and hence the MFCQ does not hold at $(\bar{x},\bar{y})$ for problem (3.80) when k is sufficiently large.

(ii) Suppose that $\mathcal{I}_F(\bar{x},\bar{y}) = \emptyset$. Then we have

$$\bar{y} = 0, \quad F(\bar{x},\bar{y}) > 0$$

and, by (3.85), $y > 0$. It follows that

$$\begin{aligned} \nabla\psi_0^k(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} &= -\sum_{i=1}^m \bar{y}_i \nabla F_i(\bar{x},\bar{y})^T \begin{pmatrix} x \\ y \end{pmatrix} - \sum_{j=1}^m F_j(\bar{x},\bar{y}) y_j \\ &= -\sum_{j=1}^m F_j(\bar{x},\bar{y}) y_j < 0, \end{aligned}$$

which also contradicts (3.83) and then the MFCQ does not hold at $(\bar{x},\bar{y})$ for problem (3.80).

Chapter 4

A Modified Scheme for MPECs

Consider the following mathematical program with complementarity constraints:

$$\begin{array}{rl} \text{minimize} & f(z) \\ \text{subject to} & g(z) \le 0,\ h(z) = 0 \\ & G(z) \ge 0,\ H(z) \ge 0 \\ & G(z)^T H(z) = 0, \end{array} \tag{4.1}$$

where $f : \Re^n \to \Re, g : \Re^n \to \Re^p, h : \Re^n \to \Re^q$, and $G, H : \Re^n \to \Re^m$ are all twice continuously differentiable functions. Recently, Scholtes [76] presented a regularization scheme

$$\begin{array}{rl} \text{minimize} & f(z) \\ \text{subject to} & g(z) \le 0,\ h(z) = 0 \\ & G(z) \ge 0,\ H(z) \ge 0 \\ & G_i(z)H_i(z) \le \epsilon,\ i = 1, 2, \cdots, m, \end{array} \tag{4.2}$$

where ϵ is a positive parameter, as an approximation of problem (4.1) and proved, under the MPEC-LICQ and the upper level strict complementarity condition, that an accumulation point of stationary points satisfying the weak second-order necessary conditions for the relaxed problems is a B-stationary point of the original problem. In this chapter, we employ the following scheme as an approximation of problem (4.1):

$$\begin{array}{rl} \text{minimize} & f(z) \\ \text{subject to} & g(z) \le 0,\ h(z) = 0 \\ & G_i(z)H_i(z) \le \epsilon^2 \\ & (G_i(z) + \epsilon)(H_i(z) + \epsilon) \ge \epsilon^2 \end{array} \tag{4.3}$$

$$i = 1, 2, \cdots, m,$$

in which there are less constraints than problem (4.2). We will show that the standard linear independence constraint qualification (LICQ) holds for the new relaxed problem under some mild conditions. Furthermore, we will give some sufficient conditions of B-stationarity for a feasible point of the original problem.

4.1 Some Results on Constraint Qualifications

In this section, we discuss constraint qualifications for problem (4.3). We let $\mathcal{F}$ and $\mathcal{F}_\epsilon$ denote the feasible sets of problems (4.1) and (4.3), respectively, and let, for $i = 1, \cdots, m$ and $z \in \Re^n$,

$$\begin{aligned}
\phi_{\epsilon,i}(z) &:= (G_i(z) + \epsilon)(H_i(z) + \epsilon) - \epsilon^2, \\
\psi_{\epsilon,i}(z) &:= G_i(z)H_i(z) - \epsilon^2,
\end{aligned}$$

and

$$\begin{aligned}
\Phi_\epsilon(z) &:= (\phi_{\epsilon,1}(z), \phi_{\epsilon,2}(z), \cdots, \phi_{\epsilon,m}(z))^T, \\
\Psi_\epsilon(z) &:= (\psi_{\epsilon,1}(z), \psi_{\epsilon,2}(z), \cdots, \psi_{\epsilon,m}(z))^T.
\end{aligned}$$

Then we have

$$\nabla \phi_{\epsilon,i}(z) = (G_i(z) + \epsilon)\nabla H_i(z) + (H_i(z) + \epsilon)\nabla G_i(z), \tag{4.4}$$

$$\nabla \psi_{\epsilon,i}(z) = H_i(z)\nabla G_i(z) + G_i(z)\nabla H_i(z) \tag{4.5}$$

for $i = 1, \cdots, m$ and

$$\begin{aligned}
\nabla \Phi_\epsilon(z) &= (\nabla \phi_{\epsilon,1}(z), \cdots, \nabla \phi_{\epsilon,m}(z)), \\
\nabla \Psi_\epsilon(z) &= (\nabla \psi_{\epsilon,1}(z), \cdots, \nabla \psi_{\epsilon,m}(z)).
\end{aligned}$$

Theorem 4.1 *We have $\mathcal{F} = \bigcap_{\epsilon>0} \mathcal{F}_\epsilon$ and, for any $\epsilon > 0$,*

$$\mathcal{I}_{\Phi_\epsilon}(z) \cap \mathcal{I}_{\Psi_\epsilon}(z) = \emptyset. \tag{4.6}$$

Proof: First of all, $\mathcal{F} \subseteq \bigcap_{\epsilon>0} \mathcal{F}_\epsilon$ is evident. Let $z \in \bigcap_{\epsilon>0} \mathcal{F}_\epsilon$. Then for any $\epsilon > 0$,

$$\begin{aligned}
&G_i(z)H_i(z) \leq \epsilon^2, \\
&G_i(z)H_i(z) + \epsilon(G_i(z) + H_i(z)) \geq 0,
\end{aligned}$$

and so

$$\epsilon + (G_i(z) + H_i(z)) \geq 0.$$

Letting $\epsilon \to 0$, we have

$$G_i(z)H_i(z) = 0,\ G_i(z) + H_i(z) \geq 0, \quad i = 1, 2, \cdots, m.$$

This means that $z \in \mathcal{F}$ and hence $\mathcal{F} = \bigcap_{\epsilon>0} \mathcal{F}_\epsilon$.

Next we prove (4.6). Suppose that for some $\epsilon > 0$ and some $z \in \mathcal{F}_\epsilon$, $i \in \mathcal{I}_{\Phi_\epsilon}(z) \cap \mathcal{I}_{\Psi_\epsilon}(z)$. Then

$$G_i(z)H_i(z) = \epsilon^2,$$
$$G_i(z)H_i(z) + \epsilon(G_i(z) + H_i(z)) = 0.$$

Combining these equalities, we have

$$G_i(z) + H_i(z) + \epsilon = 0.$$

It then follows that

$$0 = \epsilon^2 - G_i(z)H_i(z) = \epsilon^2 + H_i(z)^2 + \epsilon H_i(z) = \left(H_i(z) + \frac{\epsilon}{2}\right)^2 + \frac{3}{4}\epsilon^2,$$

which is a contradiction and so (4.6) holds. ∎

Next we show that, in contrast with problem (4.1), problem (4.3) satisfies the standard LICQ at a feasible point under some conditions.

Theorem 4.2 *For any $\bar{z} \in \mathcal{F}$, if the set of vectors*

$$\left\{\nabla g_l(\bar{z}), \nabla h_r(\bar{z}), \nabla G_i(\bar{z}), \nabla H_i(\bar{z}) \;\middle|\; l \in \mathcal{I}_g(\bar{z}),\ r = 1, \cdots, q,\ i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})\right\}$$

is linearly independent, then, for any fixed $\epsilon > 0$, there exists a neighborhood $U_\epsilon(\bar{z})$ of $\bar{z}$ such that problem (4.3) satisfies the LICQ at any point $z \in U_\epsilon(\bar{z}) \cap \mathcal{F}_\epsilon$.

Proof: For any $\bar{z} \in \mathcal{F}$, it is obvious that

$$\psi_{\epsilon,i}(\bar{z}) < 0, \quad i = 1, 2, \cdots, m$$

and

$$\phi_{\epsilon,i}(\bar{z}) = 0 \iff i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}).$$

On the other hand, it follows from the continuity of the functions g, Φ_ϵ, and Ψ_ϵ that, for any fixed $\epsilon > 0$, there exists a neighborhood $U_\epsilon(\bar{z})$ of $\bar{z}$ such that, for any point $z \in U_\epsilon(\bar{z}) \cap \mathcal{F}_\epsilon$,

$$\mathcal{I}_g(z) \subseteq \mathcal{I}_g(\bar{z}), \quad \mathcal{I}_{\Phi_\epsilon}(z) \subseteq \mathcal{I}_{\Phi_\epsilon}(\bar{z}), \quad \mathcal{I}_{\Psi_\epsilon}(z) \subseteq \mathcal{I}_{\Psi_\epsilon}(\bar{z}).$$

This means that all the functions

$$\phi_{\epsilon,i}, \quad \psi_{\epsilon,j}, \quad i \notin \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}), \quad j = 1, 2, \cdots, m$$

are inactive at z in problem (4.3). In addition, we have that

$$H_i(z) + \epsilon \neq 0, \quad G_i(z) + \epsilon \neq 0, \quad i \in \mathcal{I}_{\Phi_\epsilon}(z).$$

From (4.4), we obtain the conclusion immediately. ∎

Note that, if $\bar{z} \in \mathcal{F}$ is nondegenerate or lower level strictly complementary, namely,

$$\mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}) = \emptyset,$$

then the condition in Theorem 4.2 becomes very simple. Furthermore, under the MPEC-LICQ, we have the following stronger result in which the neighborhood is independent of the parameter ϵ.

Theorem 4.3 *For any $\bar{z} \in \mathcal{F}$, if the MPEC-LICQ holds at $\bar{z}$, which means*

$$\left\{ \nabla g_l(\bar{z}), \nabla h_r(\bar{z}), \nabla G_i(\bar{z}), \nabla H_j(\bar{z}) \;\middle|\; l \in \mathcal{I}_g(\bar{z}),\ r = 1, \cdots, q,\ i \in \mathcal{I}_G(\bar{z}),\ j \in \mathcal{I}_H(\bar{z}) \right\}$$

is linearly independent, then there exist a neighborhood $U(\bar{z})$ of $\bar{z}$ and a positive constant $\bar{\epsilon}$ such that problem (4.3) satisfies the LICQ at any point $z \in U(\bar{z}) \cap \mathcal{F}_\epsilon$ for any $\epsilon \in (0, \bar{\epsilon})$.

Proof: We first consider matrix functions whose columns consist of the vectors

$$\begin{aligned}
&\nabla g_l(z): \quad l \in \mathcal{I}_g(\bar{z}), \\
&\nabla h_r(z): \quad r = 1, \cdots, q, \\
&\nabla G_i(z): \quad i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}), \\
&\nabla G_i(z) + \frac{G_i(z) + \epsilon}{H_i(z) + \epsilon} \nabla H_i(z) \ \text{ or } \ \nabla G_i(z) + \frac{G_i(z)}{H_i(z)} \nabla H_i(z): \quad i \in \mathcal{I}_G(\bar{z}) \setminus \mathcal{I}_H(\bar{z}), \\
&\nabla H_j(z): \quad j \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}), \\
&\nabla H_j(z) + \frac{H_j(z) + \epsilon}{G_j(z) + \epsilon} \nabla G_j(z) \ \text{ or } \ \nabla H_j(z) + \frac{H_j(z)}{G_j(z)} \nabla G_j(z): \quad j \in \mathcal{I}_H(\bar{z}) \setminus \mathcal{I}_G(\bar{z}).
\end{aligned}$$

Note that there are finitely many such matrix functions, which are denoted by

$$A_1(z, \epsilon),\ A_2(z, \epsilon), \cdots,\ A_N(z, \epsilon). \tag{4.7}$$

Rearranging components if necessary, we may suppose that all these matrices are convergent to the same matrix $A(\bar{z})$ with columns

$$\nabla g_l(\bar{z}): \quad l \in \mathcal{I}_g(\bar{z}), \tag{4.8}$$

$$\nabla h_r(\bar{z}): \quad r = 1, \cdots, q, \tag{4.9}$$

$$\nabla G_i(\bar{z}): \quad i \in \mathcal{I}_G(\bar{z}), \tag{4.10}$$

$$\nabla H_j(\bar{z}): \quad j \in \mathcal{I}_H(\bar{z}), \tag{4.11}$$

respectively, as $z \to \bar{z}$ and $\epsilon \to 0$. It follows from the MPEC-LICQ assumption of the theorem that $A(\bar{z})$ has full column rank. Since the functions G, H, and g are continuous, there exist a neighborhood $U(\bar{z})$ of $\bar{z}$ and a positive constant $\bar{\epsilon}$ such that for any $\epsilon \in (0, \bar{\epsilon})$ and any point $z \in U(\bar{z}) \cap \mathcal{F}_\epsilon$, all the matrices in (4.7) have full column rank and

$$\mathcal{I}_G(z) \subseteq \mathcal{I}_G(\bar{z}), \quad \mathcal{I}_H(z) \subseteq \mathcal{I}_H(\bar{z}), \quad \mathcal{I}_g(z) \subseteq \mathcal{I}_g(\bar{z}). \tag{4.12}$$

Now we let $\epsilon \in (0, \bar{\epsilon})$ and $z \in U(\bar{z}) \cap \mathcal{F}_\epsilon$ and show that problem (4.3) satisfies the LICQ at z. We suppose that the multiplier vectors λ, μ, δ, and γ satisfy

$$\sum_{l \in \mathcal{I}_g(z)} \lambda_l \nabla g_l(z) + \sum_{r=1}^{q} \mu_r \nabla h_r(z) + \sum_{i \in \mathcal{I}_{\Phi_\epsilon}(z)} \delta_i \nabla \phi_{\epsilon,i}(z) + \sum_{j \in \mathcal{I}_{\Psi_\epsilon}(z)} \gamma_j \nabla \psi_{\epsilon,j}(z) = 0. \tag{4.13}$$

By (4.4) and (4.5), we have

$$\begin{aligned}
\sum_{i \in \mathcal{I}_{\Phi_\epsilon}(z)} \delta_i \nabla \phi_{\epsilon,i}(z) \quad = \quad & \sum_{i \in \mathcal{I}_{\Phi_\epsilon}(z) \cap \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})} \delta_i \Big((H_i(z) + \epsilon) \nabla G_i(z) + (G_i(z) + \epsilon) \nabla H_i(z) \Big) \\
& + \sum_{i \in \mathcal{I}_{\Phi_\epsilon}(z) \setminus \mathcal{I}_H(\bar{z})} \delta_i (H_i(z) + \epsilon) \Big(\nabla G_i(z) + \frac{G_i(z) + \epsilon}{H_i(z) + \epsilon} \nabla H_i(z) \Big) \\
& + \sum_{i \in \mathcal{I}_{\Phi_\epsilon}(z) \setminus \mathcal{I}_G(\bar{z})} \delta_i (G_i(z) + \epsilon) \Big(\nabla H_i(z) + \frac{H_i(z) + \epsilon}{G_i(z) + \epsilon} \nabla G_i(z) \Big)
\end{aligned}$$

and

$$\begin{aligned}
\sum_{j \in \mathcal{I}_{\Psi_\epsilon}(z)} \gamma_j \nabla \psi_{\epsilon,j}(z) \quad = \quad & \sum_{j \in \mathcal{I}_{\Psi_\epsilon}(z) \cap \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})} \gamma_j \Big(H_j(z) \nabla G_j(z) + G_j(z) \nabla H_j(z) \Big) \\
& + \sum_{j \in \mathcal{I}_{\Psi_\epsilon}(z) \setminus \mathcal{I}_H(\bar{z})} \gamma_j H_j(z) \Big(\nabla G_j(z) + \frac{G_j(z)}{H_j(z)} \nabla H_j(z) \Big) \\
& + \sum_{j \in \mathcal{I}_{\Psi_\epsilon}(z) \setminus \mathcal{I}_G(\bar{z})} \gamma_j G_j(z) \Big(\nabla H_j(z) + \frac{H_j(z)}{G_j(z)} \nabla G_j(z) \Big).
\end{aligned}$$

Note that (4.6) and (4.12) hold. Then, renumbering terms if necessary, we can choose a matrix $A_k(z,\epsilon)$ from (4.7) so that (4.13) can be rewritten as

$$A_k(z,\epsilon)\begin{pmatrix} \lambda \\ 0 \\ \mu \\ \delta_I(H_I(z)+\epsilon e_I) \\ \gamma_{II}H_{II}(z) \\ 0 \\ \delta_{III}(H_{III}(z)+\epsilon e_{III}) \\ \gamma_{IV}H_{IV}(z) \\ 0 \\ \delta_I(G_I(z)+\epsilon e_I) \\ \gamma_{II}G_{II}(z) \\ 0 \\ \delta_V(G_V(z)+\epsilon e_V) \\ \gamma_{VI}G_{VI}(z) \\ 0 \end{pmatrix} = 0, \tag{4.14}$$

where

$$\begin{aligned} I &:= \mathcal{I}_{\Phi_\epsilon}(z)\cap\mathcal{I}_G(\bar z)\cap\mathcal{I}_H(\bar z), \\ II &:= \mathcal{I}_{\Psi_\epsilon}(z)\cap\mathcal{I}_G(\bar z)\cap\mathcal{I}_H(\bar z), \\ III &:= \mathcal{I}_{\Phi_\epsilon}(z)\setminus\mathcal{I}_H(\bar z), \\ IV &:= \mathcal{I}_{\Psi_\epsilon}(z)\setminus\mathcal{I}_H(\bar z), \\ V &:= \mathcal{I}_{\Phi_\epsilon}(z)\setminus\mathcal{I}_G(\bar z), \\ VI &:= \mathcal{I}_{\Psi_\epsilon}(z)\setminus\mathcal{I}_G(\bar z), \end{aligned}$$

and $e_{\mathcal{I}} := (1,1,\cdots,1)^T \in \Re^{|\mathcal{I}|}$. Since $A_k(z,\epsilon)$ has full column rank, it follows from (4.14) that the multiplier vector in (4.14) is zero. Noticing that

$$\begin{aligned} H_i(z)+\epsilon \neq 0, \quad G_i(z)+\epsilon \neq 0, &\qquad i\in\mathcal{I}_{\Phi_\epsilon}(z), \\ H_i(z)\neq 0, \quad G_i(z)\neq 0, &\qquad i\in\mathcal{I}_{\Psi_\epsilon}(z), \end{aligned}$$

and

$$\delta = \begin{pmatrix} \delta_I \\ \delta_{III} \\ \delta_V \end{pmatrix}, \quad \gamma = \begin{pmatrix} \gamma_{II} \\ \gamma_{IV} \\ \delta_{VI} \end{pmatrix},$$

we have from (4.14) that

$$(\lambda^T, \mu^T, \delta^T, \gamma^T) = 0,$$

which implies that problem (4.3) satisfies the LICQ at z. This completes the proof. ∎

4.2 Convergence Analysis

In this section, we consider the limiting behavior of problem (4.3) as $\epsilon \to 0$. First we give the convergence of global optimal solutions.

Theorem 4.4 *Let $\{\epsilon_k\} \subseteq (0, +\infty)$ be convergent to 0 and suppose that z^k is a global optimal solution of problem (4.3) with $\epsilon = \epsilon_k$. If z^* is an accumulation point of the sequence $\{z^k\}$ as $k \to \infty$, then z^* is a global optimal solution of problem (4.1).*

Proof: Taking a subsequence if necessary, we assume without loss of generality that $\lim_{k\to\infty} z^k = z^*$. By Theorem 4.1, $z^* \in \mathcal{F}$. Since $\mathcal{F} \subseteq \mathcal{F}_{\epsilon_k}$ for all k, we have

$$f(z^k) \le f(z), \quad \forall z \in \mathcal{F}, \quad \forall k.$$

Letting $k \to \infty$, we have from the continuity of f that

$$f(z^*) \le f(z), \quad \forall z \in \mathcal{F},$$

i.e., z^* is a global optimal solution of problem (4.1). ∎

In a similar way, we can prove the next theorem.

Theorem 4.5 *Let both $\{\epsilon_k\} \subseteq (0, +\infty)$ and $\{\bar{\epsilon}_k\} \subseteq (0, +\infty)$ be convergent to 0 and $z^k \in \mathcal{F}_{\epsilon_k}$ be an $\bar{\epsilon}_k$-approximate solution of problem (4.3) with $\epsilon = \epsilon_k$, i.e.,*

$$f(z^k) - \bar{\epsilon}_k \le f(z), \quad \forall z \in \mathcal{F}_{\epsilon_k}.$$

Then any accumulation point of $\{z^k\}$ is a global optimal solution of problem (4.1).

Now we consider the limiting behavior of stationary points of problem (4.3). Recall that $\bar{z} \in \mathcal{F}$ is a C-stationary point if and only if there exist multiplier vectors $\bar{\lambda} \in \Re^p, \bar{\mu} \in \Re^q$, and $\bar{u}, \bar{v} \in \Re^m$ such that

$$\nabla f(\bar{z}) + \nabla g(\bar{z})\bar{\lambda} + \nabla h(\bar{z})\bar{\mu} - \nabla G(\bar{z})\bar{u} - \nabla H(\bar{z})\bar{v} = 0, \tag{4.15}$$

$$\bar{\lambda} \ge 0, \quad \bar{z} \in \mathcal{F}, \quad \bar{\lambda}^T g(\bar{z}) = 0, \tag{4.16}$$

$$\bar{u}_i = 0, \quad i \notin \mathcal{I}_G(\bar{z}), \tag{4.17}$$

$$\bar{v}_i = 0, \quad i \notin \mathcal{I}_H(\bar{z}), \tag{4.18}$$

and

$$\bar{u}_i \bar{v}_i \geq 0, \quad i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}) \tag{4.19}$$

and $\bar{z} \in \mathcal{F}$ is M-stationary to problem (4.1) if, furthermore, either $\bar{u}_i > 0,\ \bar{v}_i > 0$ or $\bar{u}_i \bar{v}_i = 0$ for all $i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})$. Moreover, under the MPEC-LICQ, $\bar{z} \in \mathcal{F}$ is a B-stationary point if and only if there exist multiplier vectors $\bar{\lambda} \in \Re^p, \bar{\mu} \in \Re^q$, and $\bar{u}, \bar{v} \in \Re^m$ such that (4.15)–(4.18) hold and

$$\bar{u}_i \geq 0, \quad \bar{v}_i \geq 0, \quad i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}). \tag{4.20}$$

Then we have the following convergence results.

Theorem 4.6 *Let $\{\epsilon_k\} \subseteq (0, +\infty)$ be convergent to 0 and $z^k \in \mathcal{F}_{\epsilon_k}$ be a stationary point of problem (4.3) with $\epsilon = \epsilon_k$ for each k. Suppose that $\bar{z}$ is an accumulation point of the sequence $\{z^k\}$. Then, if the MPEC-LICQ holds at $\bar{z}$, $\bar{z}$ is a C-stationary point of problem (4.1).*

Proof: Without loss of generality, we assume that

$$\lim_{k \to \infty} z^k = \bar{z}. \tag{4.21}$$

Since all the functions involved in problem (4.1) are continuous, $\mathcal{F}$ is closed and hence $\bar{z} \in \mathcal{F}$ by Theorem 4.1. It follows from the MPEC-LICQ assumption, (4.21), and Theorem 4.3 that, for any sufficiently large k, problem (4.3) with $\epsilon = \epsilon_k$ satisfies the LICQ at z^k and hence, by the stationarity of z^k, there exist unique Lagrange multiplier vectors $\lambda^k \in \Re^p, \mu^k \in \Re^q$, and $\delta^k, \gamma^k \in \Re^m$ such that

$$\nabla f(z^k) + \nabla g(z^k)\lambda^k + \nabla h(z^k)\mu^k - \nabla \Phi_{\epsilon_k}(z^k)\delta^k + \nabla \Psi_{\epsilon_k}(z^k)\gamma^k = 0, \tag{4.22}$$

$$\lambda^k \geq 0, \quad \delta^k \geq 0, \quad \gamma^k \geq 0, \tag{4.23}$$

$$g(z^k) \leq 0, \quad h(z^k) = 0, \quad \Phi_{\epsilon_k}(z^k) \geq 0, \quad \Psi_{\epsilon_k}(z^k) \leq 0, \tag{4.24}$$

$$g(z^k)^T \lambda^k = 0, \quad \Phi_{\epsilon_k}(z^k)^T \delta^k = 0, \quad \Psi_{\epsilon_k}(z^k)^T \gamma^k = 0. \tag{4.25}$$

It follows from (4.23)–(4.25) that

$$\lambda_i^k = 0, \quad i \notin \mathcal{I}_g(z^k), \tag{4.26}$$

$$\delta_i^k = 0, \quad i \notin \mathcal{I}_{\Phi_{\epsilon_k}}(z^k), \tag{4.27}$$

$$\gamma_i^k = 0, \quad i \notin \mathcal{I}_{\Psi_{\epsilon_k}}(z^k). \tag{4.28}$$

Now suppose that, for all sufficiently large k, (4.22)–(4.25) hold and, in addition, the conditions

$$\mathcal{I}_G(z^k) \subseteq \mathcal{I}_G(\bar{z}), \quad \mathcal{I}_H(z^k) \subseteq \mathcal{I}_H(\bar{z}), \quad \mathcal{I}_g(z^k) \subseteq \mathcal{I}_g(\bar{z}) \tag{4.29}$$

hold and all the matrix functions $A_i(z,\epsilon),\ i=1,\cdots,N$, in (4.7) defined in the proof of Theorem 4.3 have full column rank at (z^k,ϵ_k). By (4.4) and (4.5), we have

$$\begin{aligned}\nabla\Phi_{\epsilon_k}(z^k)\delta^k &= \sum_{i\in\mathcal{I}_G(\bar z)\cap\mathcal{I}_H(\bar z)} \delta_i^k\Big((H_i(z^k)+\epsilon_k)\nabla G_i(z^k)+(G_i(z^k)+\epsilon_k)\nabla H_i(z^k)\Big)\\ &\quad+\sum_{i\in\mathcal{I}_G(\bar z)\setminus\mathcal{I}_H(\bar z)} \delta_i^k(H_i(z^k)+\epsilon_k)\Big(\nabla G_i(z^k)+\frac{G_i(z^k)+\epsilon_k}{H_i(z^k)+\epsilon_k}\nabla H_i(z^k)\Big)\\ &\quad+\sum_{i\in\mathcal{I}_H(\bar z)\setminus\mathcal{I}_G(\bar z)} \delta_i^k(G_i(z^k)+\epsilon_k)\Big(\nabla H_i(z^k)+\frac{H_i(z^k)+\epsilon_k}{G_i(z^k)+\epsilon_k}\nabla G_i(z^k)\Big)\end{aligned} \tag{4.30}$$

and

$$\begin{aligned}\nabla\Psi_{\epsilon_k}(z^k)\gamma^k &= \sum_{j\in\mathcal{I}_G(\bar z)\cap\mathcal{I}_H(\bar z)} \gamma_j^k\Big(H_j(z^k)\nabla G_j(z^k)+G_j(z^k)\nabla H_j(z^k)\Big)\\ &\quad+\sum_{j\in\mathcal{I}_G(\bar z)\setminus\mathcal{I}_H(\bar z)} \gamma_j^k H_j(z^k)\Big(\nabla G_j(z^k)+\frac{G_j(z^k)}{H_j(z^k)}\nabla H_j(z^k)\Big)\\ &\quad+\sum_{j\in\mathcal{I}_H(\bar z)\setminus\mathcal{I}_G(\bar z)} \gamma_j^k G_j(z^k)\Big(\nabla H_j(z^k)+\frac{H_j(z^k)}{G_j(z^k)}\nabla G_j(z^k)\Big).\end{aligned} \tag{4.31}$$

Then, taking into account (4.6), we have from (4.22) and (4.26)–(4.31) that

$$\begin{aligned}-\nabla f(z^k) &= \nabla g(z^k)\lambda^k+\nabla h(z^k)\mu^k\\ &\quad-\sum_{i\in\mathcal{I}_G(\bar z)\cap\mathcal{I}_H(\bar z)}\Big(\delta_i^k(H_i(z^k)+\epsilon_k)-\gamma_i^k H_i(z^k)\Big)\nabla G_i(z^k)\\ &\quad-\sum_{i\in\mathcal{I}_{\Phi_{\epsilon_k}}(z^k)\setminus\mathcal{I}_H(\bar z)} \delta_i^k(H_i(z^k)+\epsilon_k)\Big(\nabla G_i(z^k)+\frac{G_i(z^k)+\epsilon_k}{H_i(z^k)+\epsilon_k}\nabla H_i(z^k)\Big)\\ &\quad-\sum_{i\in\mathcal{I}_{\Psi_{\epsilon_k}}(z^k)\setminus\mathcal{I}_H(\bar z)}\Big(-\gamma_i^k H_i(z^k)\Big)\Big(\nabla G_i(z^k)+\frac{G_i(z^k)}{H_i(z^k)}\nabla H_i(z^k)\Big)\\ &\quad-\sum_{i\in\mathcal{I}_G(\bar z)\cap\mathcal{I}_H(\bar z)}\Big(\delta_i^k(G_i(z^k)+\epsilon_k)-\gamma_i^k G_i(z^k)\Big)\nabla H_i(z^k)\\ &\quad-\sum_{i\in\mathcal{I}_{\Phi_{\epsilon_k}}(z^k)\setminus\mathcal{I}_G(\bar z)} \delta_i^k(G_i(z^k)+\epsilon_k)\Big(\nabla H_i(z^k)+\frac{H_i(z^k)+\epsilon_k}{G_i(z^k)+\epsilon_k}\nabla G_i(z^k)\Big)\\ &\quad-\sum_{i\in\mathcal{I}_{\Psi_{\epsilon_k}}(z^k)\setminus\mathcal{I}_H(\bar z)}\Big(-\gamma_i^k G_i(z^k)\Big)\Big(\nabla H_i(z^k)+\frac{H_i(z^k)}{G_i(z^k)}\nabla G_i(z^k)\Big)\end{aligned}$$

$$= \quad A_{N_k}(z^k, \epsilon_k) \begin{pmatrix} \lambda^k_{\mathcal{I}_g(\bar{z})} \\ \mu^k \\ u^k \\ v^k \end{pmatrix}, \tag{4.32}$$

where u^k, v^k are given by

$$u_i^k := \begin{cases} \delta_i^k (H_i(z^k) + \epsilon_k), & i \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cap \mathcal{I}_G(\bar{z}) \\ -\gamma_i^k H_i(z^k), & i \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k) \cap \mathcal{I}_G(\bar{z}) \\ 0, & i \in \mathcal{I}_G(\bar{z}) \setminus \Big(\mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cup \mathcal{I}_{\Psi_{\epsilon_k}}(z^k)\Big), \end{cases} \tag{4.33}$$

$$v_i^k := \begin{cases} \delta_i^k (G_i(z^k) + \epsilon_k), & i \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cap \mathcal{I}_H(\bar{z}) \\ -\gamma_i^k G_i(z^k), & i \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k) \cap \mathcal{I}_H(\bar{z}) \\ 0, & i \in \mathcal{I}_H(\bar{z}) \setminus \Big(\mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cup \mathcal{I}_{\Psi_{\epsilon_k}}(z^k)\Big), \end{cases} \tag{4.34}$$

respectively, and $A_{N_k}(z, \epsilon)$ is one of the matrix functions in (4.7). As we assumed above, $A_{N_k}(z^k, \epsilon_k)$ has full column rank for all sufficiently large k. In consequence, it follows from (4.21) and (4.32) that all the multiplier sequences

$$\{\lambda_l^k \mid l \in \mathcal{I}_g(\bar{z})\}, \quad \{\mu_r^k \mid r = 1, \cdots, q\}, \tag{4.35}$$

$$\{u_i^k \mid i \in \mathcal{I}_G(\bar{z})\}, \quad \{v_j^k \mid j \in \mathcal{I}_H(\bar{z})\} \tag{4.36}$$

are convergent. Let $\bar{\lambda} \in \Re^p, \bar{\mu} \in \Re^q$, and $\bar{u}, \bar{v} \in \Re^m$ be as follows:

$$\bar{\lambda}_l := \begin{cases} \lim_{k\to\infty} \lambda_l^k &, \ l \in \mathcal{I}_g(\bar{z}) \\ 0 &, \ l \notin \mathcal{I}_g(\bar{z}) \end{cases}, \tag{4.37}$$

$$\bar{\mu}_r := \lim_{k\to\infty} \mu_r^k, \quad r = 1, 2, \cdots, q, \tag{4.38}$$

$$\bar{u}_i := \begin{cases} \lim_{k\to\infty} u_i^k &, \ i \in \mathcal{I}_G(\bar{z}) \\ 0 &, \ i \notin \mathcal{I}_G(\bar{z}) \end{cases}, \tag{4.39}$$

$$\bar{v}_j := \begin{cases} \lim_{k\to\infty} v_j^k &, \ j \in \mathcal{I}_H(\bar{z}) \\ 0 &, \ j \notin \mathcal{I}_H(\bar{z}) \end{cases}. \tag{4.40}$$

Letting $k \to \infty$ in (4.32) and noticing that

$$\lim_{k\to\infty} A_{N_k}(z^k, \epsilon_k) = A(\bar{z}),$$

where $A(\bar{z})$ is the matrix with the columns (4.8)–(4.11), we have from (4.37)–(4.40) that

$$-\nabla f(\bar{z}) = \nabla g(\bar{z})\bar{\lambda} + \nabla h(\bar{z})\bar{\mu} - \nabla G(\bar{z})\bar{u} - \nabla H(\bar{z})\bar{v},$$

i.e., (4.15) holds for the multiplier vectors $\bar{\lambda}, \bar{\mu}, \bar{u}, \bar{v}$. On the other hand, we have (4.16)–(4.18) immediately from (4.23), (4.24), (4.37), (4.39), and (4.40). Then the rest of the proof is to show (4.19). In fact, for each $i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})$, we have from (4.6) and (4.33)–(4.34) that

$$u_i^k v_i^k = \begin{cases} (\delta_i^k)^2 (H_i(z^k) + \epsilon_k)(G_i(z^k) + \epsilon_k) = (\delta_i^k \epsilon_k)^2, & i \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \\ (\gamma_i^k)^2 H_i(z^k) G_i(z^k) = (\gamma_i^k \epsilon_k)^2, & i \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k) \\ 0, & i \notin \mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cup \mathcal{I}_{\Psi_{\epsilon_k}}(z^k). \end{cases}$$

Letting $k \to \infty$, we obtain (4.19) since the sequences $\{u_i^k\}$ and $\{v_i^k\}$ in (4.36) are convergent. Hence $\bar{z}$ is a C-stationary point of problem (4.1). This completes the proof. ∎

From the definitions of B- and C-stationarity, we have the following result immediately.

Corollary 4.1 *Let the assumptions in Theorem 4.6 be satisfied. If, in addition, $\bar{z}$ is nondegenerate, then it is a B-stationary point of problem (4.1).*

On the other hand, we can prove some similar convergence results as in Chapter 3. Let

$$L_\epsilon(z, \lambda, \mu, \delta, \gamma) := f(z) + \lambda^T g(z) + \mu^T h(z) - \delta^T \Phi_\epsilon(z) + \gamma^T \Psi_\epsilon(z)$$

stands for the Lagrangian of problem (4.3) and

$$\mathcal{T}_\epsilon(z) := \Big\{ d \in \Re^n : \quad \begin{array}{l} d^T \nabla \phi_{\epsilon,i}(z) = 0, \ i \in \mathcal{I}_{\Phi_\epsilon}(z); \\ d^T \nabla \psi_{\epsilon,j}(z) = 0, \ j \in \mathcal{I}_{\Psi_\epsilon}(z); \\ d^T \nabla g_l(z) = 0, \ l \in \mathcal{I}_g(z); \\ d^T \nabla h_r(z) = 0, \ r = 1, 2, \cdots, q \Big\}. \end{array}$$

Suppose that α is a nonnegative number. Recall that, at a stationary point z of problem (4.3), the matrix $\nabla_z^2 L_\epsilon(z, \lambda, \mu, \delta, \gamma)$ is bounded below with constant α on the corresponding tangent space $\mathcal{T}_\epsilon(z)$ means

$$d^T \nabla_z^2 L_\epsilon(z, \lambda, \mu, \delta, \gamma) d \geq -\alpha ||d||^2, \quad \forall d \in \mathcal{T}_\epsilon(z). \tag{4.41}$$

See Section 4.2 for more details about this property.

Theorem 4.7 *Let $\{\epsilon_k\} \subseteq (0, +\infty)$ be convergent to 0 and $z^k \in \mathcal{F}_{\epsilon_k}$ be a stationary point of problem (4.3) with $\epsilon = \epsilon_k$ and multiplier vectors $\lambda^k, \mu^k, \delta^k$, and γ^k. Suppose* *that, for each k, $\nabla_z^2 L_{\epsilon_k}(z^k, \lambda^k, \mu^k, \delta^k, \gamma^k)$ is bounded below with constant α_k on the corresponding tangent space $\mathcal{T}_{\epsilon_k}(z^k)$. Let $\bar{z}$ be an accumulation point of the sequence $\{z^k\}$. If the sequence $\{\alpha_k\}$ is bounded and the MPEC-LICQ holds at $\bar{z}$, then $\bar{z}$ is an M-stationary point of problem (4.1).*

Proof: Assume that $\lim_{k\to\infty} z^k = \bar{z}$ without loss of generality. First of all, we note from Theorem 4.6 that $\bar{z}$ is a C-stationary point of problem (4.1). To prove the theorem, we assume to the contrary that $\bar{z}$ is not M-stationary to problem (4.1). Then, it follows from the definitions of C-stationarity and M-stationarity that there must exist an $i_0 \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})$ such that

$$\bar{u}_{i_0} < 0, \quad \bar{v}_{i_0} < 0. \tag{4.42}$$

By (4.33)–(4.34) and (4.39)–(4.40), we have

$$i_0 \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cup \mathcal{I}_{\Psi_{\epsilon_k}}(z^k)$$

for every sufficiently large k. First we consider the case where $i_0 \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k)$ for infinitely many k. Furthermore, taking a subsequence if necessary, we may assume without loss of generality that

$$i_0 \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k) \tag{4.43}$$

for all sufficiently large k. Then, by (4.33) and (4.34),

$$\bar{u}_{i_0} = -\lim_{k\to\infty} \gamma_{i_0}^k H_{i_0}(z^k) < 0, \tag{4.44}$$

$$\bar{v}_{i_0} = -\lim_{k\to\infty} \gamma_{i_0}^k G_{i_0}(z^k) < 0, \tag{4.45}$$

and so

$$\lim_{k\to\infty} \frac{H_{i_0}(z^k)}{G_{i_0}(z^k)} = \frac{\bar{u}_{i_0}}{\bar{v}_{i_0}} > 0. \tag{4.46}$$

In what follows, we suppose that, for all sufficiently large k, (4.22)–(4.25), (4.29), and

$$\frac{H_{i_0}(z^k)}{G_{i_0}(z^k)} > 0$$

hold and all the matrix functions $A_i(z, \epsilon),\ i = 1, \cdots, N$, in (4.7) have full column rank at (z^k, ϵ_k). For such k, the matrix $A_{N_k}(z^k, \epsilon_k)$ whose columns consist of the vectors

$$\begin{aligned} &\nabla g_l(z^k): \quad l \in \mathcal{I}_g(\bar{z}), \\ &\nabla h_r(z^k): \quad r = 1, \cdots, q, \end{aligned}$$

$$
\begin{array}{ll}
\nabla G_i(z^k): & i \in \Big(\mathcal{I}_G(\bar z) \cap \mathcal{I}_H(\bar z)\Big) \cup \Big(\mathcal{I}_G(\bar z) \setminus (\mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cup \mathcal{I}_{\Psi_{\epsilon_k}}(z^k))\Big), \\
\nabla G_i(z^k) + \dfrac{G_i(z^k)+\epsilon_k}{H_i(z^k)+\epsilon_k}\nabla H_i(z^k): & i \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \setminus \mathcal{I}_H(\bar z), \\
\nabla G_i(z^k) + \dfrac{G_i(z^k)}{H_i(z^k)}\nabla H_i(z^k): & i \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k) \setminus \mathcal{I}_H(\bar z), \\
\nabla H_j(z^k): & j \in \Big(\mathcal{I}_G(\bar z) \cap \mathcal{I}_H(\bar z)\Big) \cup \Big(\mathcal{I}_H(\bar z) \setminus (\mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cup \mathcal{I}_{\Psi_{\epsilon_k}}(z^k))\Big), \\
\nabla H_j(z^k) + \dfrac{H_j(z^k)+\epsilon_k}{G_j(z^k)+\epsilon_k}\nabla G_j(z^k): & j \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \setminus \mathcal{I}_G(\bar z), \\
\nabla H_j(z^k) + \dfrac{H_j(z^k)}{G_j(z^k)}\nabla G_j(z^k): & j \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k) \setminus \mathcal{I}_G(\bar z)
\end{array}
$$

has full column rank. Therefore, we can choose a vector $d^k \in \Re^n$ such that

$$(d^k)^T \nabla g_l(z^k) = 0, \qquad l \in \mathcal{I}_g(\bar z); \tag{4.47}$$

$$(d^k)^T \nabla h_r(z^k) = 0, \qquad r = 1, \cdots, q; \tag{4.48}$$

$$(d^k)^T \nabla G_i(z^k) = 0, \quad i \in \Big(\mathcal{I}_G(\bar z) \cap \mathcal{I}_H(\bar z)\Big) \cup \Big(\mathcal{I}_G(\bar z) \setminus (\mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cup \mathcal{I}_{\Psi_{\epsilon_k}}(z^k))\Big),\ i \neq i_0; \tag{4.49}$$

$$(d^k)^T \Big(\nabla G_i(z^k) + \frac{G_i(z^k)+\epsilon_k}{H_i(z^k)+\epsilon_k}\nabla H_i(z^k)\Big) = 0, \quad i \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \setminus \mathcal{I}_H(\bar z); \tag{4.50}$$

$$(d^k)^T \Big(\nabla G_i(z^k) + \frac{G_i(z^k)}{H_i(z^k)}\nabla H_i(z^k)\Big) = 0, \quad i \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k) \setminus \mathcal{I}_H(\bar z); \tag{4.51}$$

$$(d^k)^T \nabla H_j(z^k) = 0, \quad j \in \Big(\mathcal{I}_G(\bar z) \cap \mathcal{I}_H(\bar z)\Big) \cup \Big(\mathcal{I}_H(\bar z) \setminus (\mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cup \mathcal{I}_{\Psi_{\epsilon_k}}(z^k))\Big),\ j \neq i_0; \tag{4.52}$$

$$(d^k)^T \Big(\nabla H_j(z^k) + \frac{H_j(z^k)+\epsilon_k}{G_j(z^k)+\epsilon_k}\nabla G_j(z^k)\Big) = 0, \quad j \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \setminus \mathcal{I}_G(\bar z); \tag{4.53}$$

$$(d^k)^T \Big(\nabla H_j(z^k) + \frac{H_j(z^k)}{G_j(z^k)}\nabla G_j(z^k)\Big) = 0, \quad j \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k) \setminus \mathcal{I}_G(\bar z); \tag{4.54}$$

$$(d^k)^T \nabla G_{i_0}(z^k) = 1; \tag{4.55}$$

$$(d^k)^T \nabla H_{i_0}(z^k) = -\frac{H_{i_0}(z^k)}{G_{i_0}(z^k)}.$$

Then for any $i \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k)$ and any $j \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k)$, since

$$
\begin{aligned}
\nabla \phi_{\epsilon_k,i}(z^k) &= (G_i(z^k)+\epsilon_k)\nabla H_i(z^k) + (H_i(z^k)+\epsilon_k)\nabla G_i(z^k), \\
\nabla \psi_{\epsilon_k,j}(z^k) &= H_j(z^k)\nabla G_j(z^k) + G_j(z^k)\nabla H_j(z^k),
\end{aligned}
$$

we have

$$
\begin{aligned}
(d^k)^T \nabla \phi_{\epsilon_k,i}(z^k) = 0, &\quad i \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k), \\
(d^k)^T \nabla \psi_{\epsilon_k,j}(z^k) = 0, &\quad j \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k),
\end{aligned}
$$

and so $d^k \in \mathcal{T}_{\epsilon_k}(z^k)$. Furthermore, we can choose the sequence $\{d^k\}$ to be bounded. Since $\nabla_z^2 L_{\epsilon_k}(z^k, \lambda^k, \mu^k, \delta^k, \gamma^k)$ is bounded below with constant α_k on the corresponding tangent space $\mathcal{T}_{\epsilon_k}(z^k)$, we have from (4.41) that there exists a constant C such that

$$(d^k)^T \nabla_z^2 L_{\epsilon_k}(z^k, \lambda^k, \mu^k, \delta^k, \gamma^k) d^k \geq -\alpha_k ||d^k||^2 \geq C, \tag{4.56}$$

where the last inequality follows from the boundedness of the sequences $\{\alpha_k\}$ and $\{d^k\}$. Note that, by (4.26)–(4.28) and

$$\begin{aligned}
\nabla^2 \phi_{\epsilon_k,i}(z^k) &= \nabla G_i(z^k) \nabla H_i(z^k)^T + \nabla H_i(z^k) \nabla G_i(z^k)^T \\
&\quad + (G_i(z^k) + \epsilon_k) \nabla^2 H_i(z^k) + (H_i(z^k) + \epsilon_k) \nabla^2 G_i(z^k), \\
\nabla^2 \psi_{\epsilon_k,j}(z^k) &= \nabla G_j(z^k) \nabla H_j(z^k)^T + \nabla H_j(z^k) \nabla G_j(z^k)^T \\
&\quad + G_j(z^k) \nabla^2 H_j(z^k) + H_j(z^k) \nabla^2 G_j(z^k),
\end{aligned}$$

there holds

$$\begin{aligned}
\nabla_z^2 L_{\epsilon_k}(z^k, \lambda^k, \mu^k, \delta^k, \gamma^k) &= \nabla^2 f(z^k) + \sum_{l=1}^{p} \lambda_l^k \nabla^2 g_l(z^k) + \sum_{r=1}^{q} \mu_r^k \nabla^2 h_r(z^k) \\
&\quad - \sum_{i=1}^{m} \delta_i^k \nabla^2 \phi_{\epsilon_k,i}(z^k) + \sum_{j=1}^{m} \gamma_j^k \nabla^2 \psi_{\epsilon_k,j}(z^k) \\
&= \nabla^2 f(z^k) + \sum_{l \in \mathcal{I}_g(\bar{z})} \lambda_l^k \nabla^2 g_l(z^k) + \sum_{r=1}^{q} \mu_r^k \nabla^2 h_r(z^k) \\
&\quad - \sum_{i \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k)} \delta_i^k \nabla^2 \phi_{\epsilon_k,i}(z^k) + \sum_{j \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k)} \gamma_j^k \nabla^2 \psi_{\epsilon_k,j}(z^k).
\end{aligned}$$

We then have

$$\begin{aligned}
&(d^k)^T \nabla_z^2 L_{\epsilon_k}(z^k, \lambda^k, \mu^k, \delta^k, \gamma^k) d^k \\
&= (d^k)^T \nabla^2 f(z^k) d^k + \sum_{l \in \mathcal{I}_g(\bar{z})} \lambda_l^k (d^k)^T \nabla^2 g_l(z^k) d^k + \sum_{r=1}^{q} \mu_r^k (d^k)^T \nabla^2 h_r(z^k) d^k \\
&\quad - \sum_{i \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k)} \delta_i^k \Big((d^k)^T \nabla G_i(z^k) \nabla H_i(z^k)^T d^k + (d^k)^T \nabla H_i(z^k) \nabla G_i(z^k)^T d^k \\
&\qquad + (G_i(z^k) + \epsilon_k)(d^k)^T \nabla^2 H_i(z^k) d^k + (H_i(z^k) + \epsilon_k)(d^k)^T \nabla^2 G_i(z^k) d^k \Big) \\
&\quad + \sum_{j \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k)} \gamma_j^k \Big((d^k)^T \nabla G_j(z^k) \nabla H_j(z^k)^T d^k + (d^k)^T \nabla H_j(z^k) \nabla G_j(z^k)^T d^k \\
&\qquad + G_j(z^k)(d^k)^T \nabla^2 H_j(z^k) d^k + H_j(z^k)(d^k)^T \nabla^2 G_j(z^k) d^k \Big).
\end{aligned} \tag{4.57}$$

By the twice continuous differentiability of the functions, the boundness of the sequence $\{d^k\}$, and the convergence of the sequences $\{z^k\}$, $\{\lambda_l^k\}$ and $\{\mu_r^k\}$ (by (4.37)–(4.38)), the terms

$$(d^k)^T\nabla^2 f(z^k)d^k, \quad \sum_{l\in\mathcal{I}_g(\bar{z})} \lambda_l^k (d^k)^T\nabla^2 g_l(z^k)d^k, \quad \sum_{r=1}^{q} \mu_r^k (d^k)^T\nabla^2 h_r(z^k)d^k$$

are all bounded. Consider arbitrary indices i and j such that $i \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k)$ for infinitely many k and $j \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k) \setminus \{i_0\}$ for infinitely many k, respectively. If

$$i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}) \quad \text{or} \quad j \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}),$$

then

$$(d^k)^T\nabla G_i(z^k) = 0 \quad \text{or} \quad (d^k)^T\nabla H_j(z^k) = 0$$

and, by (4.33)–(4.34) and (4.39)–(4.40), the sequences

$$\Big\{\delta_i^k (G_i(z^k) + \epsilon_k)\Big\}, \quad \Big\{\delta_i^k (H_i(z^k) + \epsilon_k)\Big\},$$

and

$$\Big\{\gamma_j^k G_j(z^k)\Big\}, \quad \Big\{\gamma_j^k H_j(z^k)\Big\}$$

are all convergent. If

$$i, j \notin \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}),$$

then, also by (4.33)–(4.34) and (4.39)–(4.40), the sequences $\{\delta_i^k\}$ and $\{\gamma_j^k\}$ are convergent. Therefore, we have that the terms

$$\sum_{i\in\mathcal{I}_{\Phi_{\epsilon_k}}(z^k)} \delta_i^k \Big((d^k)^T\nabla G_i(z^k)\nabla H_i(z^k)^T d^k + (d^k)^T\nabla H_i(z^k)\nabla G_i(z^k)^T d^k + (G_i(z^k)+\epsilon_k)(d^k)^T\nabla^2 H_i(z^k)d^k + (H_i(z^k)+\epsilon_k)(d^k)^T\nabla^2 G_i(z^k)d^k\Big)$$

and

$$\sum_{j\in\mathcal{I}_{\Psi_{\epsilon_k}}(z^k)\setminus\{i_0\}} \gamma_j^k \Big((d^k)^T\nabla G_j(z^k)\nabla H_j(z^k)^T d^k + (d^k)^T\nabla H_j(z^k)\nabla G_j(z^k)^T d^k + G_j(z^k)(d^k)^T\nabla^2 H_j(z^k)d^k + H_j(z^k)(d^k)^T\nabla^2 G_j(z^k)d^k\Big)$$

are bounded. On the other hand, however, we have (4.43) for all sufficiently large k and

$$\begin{aligned} &\gamma_{i_0}^k \Big((d^k)^T\nabla G_{i_0}(z^k)\nabla H_{i_0}(z^k)^T d^k + (d^k)^T\nabla H_{i_0}(z^k)\nabla G_{i_0}(z^k)^T d^k \\ &+G_{i_0}(z^k)(d^k)^T\nabla^2 H_{i_0}(z^k)d^k + H_{i_0}(z^k)(d^k)^T\nabla^2 G_{i_0}(z^k)d^k\Big) \qquad (4.58) \\ = \; & -\frac{2\gamma_{i_0}^k H_{i_0}(z^k)}{G_{i_0}(z^k)} + \gamma_{i_0}^k\Big(G_{i_0}(z^k)(d^k)^T\nabla^2 H_{i_0}(z^k)d^k + H_{i_0}(z^k)(d^k)^T\nabla^2 G_{i_0}(z^k)d^k\Big). \end{aligned}$$

Since (4.46) holds and $\gamma_{i_0}^k \to +\infty$ as $k \to \infty$ by (4.23) and (4.44), we have

$$-\frac{2\gamma_{i_0}^k H_{i_0}(z^k)}{G_{i_0}(z^k)} \to -\infty$$

as $k \to \infty$. Note that, by (4.44) and (4.45), the sequences

$$\left\{\gamma_{i_0}^k G_{i_0}(z^k)\right\}, \quad \left\{\gamma_{i_0}^k H_{i_0}(z^k)\right\}$$

are also convergent. We then have that the term (4.58) tends to $-\infty$ as $k \to \infty$. Therefore, it follows from (4.57) that

$$(d^k)^T \nabla_z^2 L_{\epsilon_k}(z^k, \lambda^k, \mu^k, \delta^k, \gamma^k) d^k \to -\infty$$

as $k \to \infty$. This contradicts (4.56) and hence $\bar{z}$ is M-stationary to problem (4.1).

Finally we consider the case where $i_0 \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k)$ for infinitely many k. By (4.33) and (4.34), we have from (4.42) that

$$\begin{aligned} \bar{u}_{i_0} &= \lim_{k\to\infty} \delta_{i_0}^k (H_{i_0}(z^k) + \epsilon_k) < 0, \\ \bar{v}_{i_0} &= \lim_{k\to\infty} \delta_{i_0}^k (G_{i_0}(z^k) + \epsilon_k) < 0, \end{aligned}$$

and so

$$\lim_{k\to\infty} \frac{H_{i_0}(z^k) + \epsilon_k}{G_{i_0}(z^k) + \epsilon_k} = \frac{\bar{u}_{i_0}}{\bar{v}_{i_0}} > 0.$$

Therefore, we can also choose a bounded sequence $\{d^k\}$ such that (4.47)–(4.55) and

$$(d^k)^T \nabla H_{i_0}(z^k) = -\frac{H_{i_0}(z^k) + \epsilon_k}{G_{i_0}(z^k) + \epsilon_k}$$

hold for each k. In a similar way, we then obtain a contradiction and so $\bar{z}$ is M-stationary to problem (4.1). This completes the proof. ∎

Corollary 4.2 *Let $\{\epsilon_k\}, \{z^k\}$, and $\bar{z}$ be the same as in Theorem 4.7. If z^k together with the corresponding multiplier vectors $\lambda^k, \mu^k, \delta^k$, and γ^k satisfies the weak second-order necessary conditions for problem (4.3) with $\epsilon = \epsilon_k$ and the MPEC-LICQ holds at $\bar{z}$, then $\bar{z}$ is an M-stationary point of problem (4.1).*

Corollary 4.3 *Let the assumptions in Theorem 4.7 be satisfied. If, in addition, $\bar{z}$ satisfies the upper level strict complementarity condition, then it is a B-stationary point of problem (4.1).*

Theorem 4.8 *Let* $\{\epsilon_k\}, \{z^k\}$, *and* $\bar{z}$ *be the same as in Theorem 4.7 and* $\lambda^k, \mu^k, \delta^k$, *and* γ^k *be the multiplier vectors corresponding to* z^k. *Let* β_k *be the smallest eigenvalue of the matrix* $\nabla_z^2 L_{\epsilon_k}(z^k, \lambda^k, \mu^k, \delta^k, \gamma^k)$. *If the sequence* $\{\beta_k\}$ *is bounded below and the MPEC-LICQ holds at* $\bar{z}$, *then* $\bar{z}$ *is a B-stationary point of problem (4.1).*

Proof: It is easy to see that the assumptions of Theorem 4.7 are satisfied with $\alpha_k = \max\{-\beta_k, 0\}$ and so $\bar{z}$ is an M-stationary point of problem (4.1). Suppose that $\bar{z}$ is not B-stationary to problem (4.1). Then, by the definitions of B- and M-stationarity, there exists an $i_0 \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})$ such that

$$\bar{u}_{i_0} < 0, \quad \bar{v}_{i_0} = 0 \tag{4.59}$$

or

$$\bar{u}_{i_0} = 0, \quad \bar{v}_{i_0} < 0.$$

By (4.33)–(4.34) and (4.39)–(4.40), we have

$$i_0 \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cup \mathcal{I}_{\Psi_{\epsilon_k}}(z^k)$$

for every sufficiently large k. Without loss of generality, we assume that (4.59) holds.

First we consider the case where $i_0 \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k)$ for infinitely many k. By taking a subsequence if necessary, we assume

$$i_0 \in \mathcal{I}_{\Psi_{\epsilon_k}}(z^k) \tag{4.60}$$

for all sufficiently large k. Then, it follows from (4.33), (4.34), and (4.59) that

$$\bar{u}_{i_0} = -\lim_{k\to\infty} \gamma_{i_0}^k H_{i_0}(z^k) < 0$$

and so, by (4.23), we have

$$\lim_{k\to\infty} \gamma_{i_0}^k = +\infty. \tag{4.61}$$

Now we suppose that, for all sufficiently large k, (4.22)–(4.25) and (4.29) hold and the matrix $A_{N_k}(z^k, \epsilon_k)$ defined in the proof of Theorem 4.7 has full column rank. Therefore, we can choose a vector $d^k \in \Re^n$ such that

$$\begin{aligned}
&(d^k)^T \nabla g_l(z^k) = 0, && l \in \mathcal{I}_g(\bar{z});\\
&(d^k)^T \nabla h_r(z^k) = 0, && r = 1, 2, \cdots, q;\\
&(d^k)^T \nabla G_i(z^k) = 0, && i \in \Big(\mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})\Big) \cup \Big(\mathcal{I}_G(\bar{z}) \setminus (\mathcal{I}_{\Phi_{\epsilon_k}}(z^k) \cup \mathcal{I}_{\Psi_{\epsilon_k}}(z^k))\Big),\ i \neq i_0;
\end{aligned}$$

$$
\begin{aligned}
&(d^k)^T\Big(\nabla G_i(z^k)+\frac{G_i(z^k)+\epsilon_k}{H_i(z^k)+\epsilon_k}\nabla H_i(z^k)\Big)=0, \quad i\in\mathcal{I}_{\Phi_{\epsilon_k}}(z^k)\setminus\mathcal{I}_H(\bar z);\\
&(d^k)^T\Big(\nabla G_i(z^k)+\frac{G_i(z^k)}{H_i(z^k)}\nabla H_i(z^k)\Big)=0, \qquad i\in\mathcal{I}_{\Psi_{\epsilon_k}}(z^k)\setminus\mathcal{I}_H(\bar z);\\
&(d^k)^T\nabla H_j(z^k)=0, \quad j\in\Big(\mathcal{I}_G(\bar z)\cap\mathcal{I}_H(\bar z)\Big)\cup\Big(\mathcal{I}_H(\bar z)\setminus(\mathcal{I}_{\Phi_{\epsilon_k}}(z^k)\cup\mathcal{I}_{\Psi_{\epsilon_k}}(z^k))\Big),\ j\neq i_0;\\
&(d^k)^T\Big(\nabla H_j(z^k)+\frac{H_j(z^k)+\epsilon_k}{G_j(z^k)+\epsilon_k}\nabla G_j(z^k)\Big)=0, \quad j\in\mathcal{I}_{\Phi_{\epsilon_k}}(z^k)\setminus\mathcal{I}_G(\bar z);\\
&(d^k)^T\Big(\nabla H_j(z^k)+\frac{H_j(z^k)}{G_j(z^k)}\nabla G_j(z^k)\Big)=0, \qquad j\in\mathcal{I}_{\Psi_{\epsilon_k}}(z^k)\setminus\mathcal{I}_G(\bar z);\\
&(d^k)^T\nabla G_{i_0}(z^k)=1;\\
&(d^k)^T\nabla H_{i_0}(z^k)=-1.
\end{aligned}
$$

Furthermore, we can choose the sequence $\{d^k\}$ to be bounded. By the assumptions of the theorem, there exists a constant C such that

$$
(d^k)^T\nabla_z^2 L_{\epsilon_k}(z^k,\lambda^k,\mu^k,\delta^k,\gamma^k)d^k \geq \beta_k||d^k||^2 \geq C \tag{4.62}
$$

holds for all k. In a similar way to the proof of Theorem 4.7, we can show that all the terms on the right-hand side of (4.57) except

$$
\begin{aligned}
&\gamma_{i_0}^k\Big((d^k)^T\nabla G_{i_0}(z^k)\nabla H_{i_0}(z^k)^T d^k+(d^k)^T\nabla H_{i_0}(z^k)\nabla G_{i_0}(z^k)^T d^k\\
&+G_{i_0}(z^k)(d^k)^T\nabla^2 H_{i_0}(z^k)d^k+H_{i_0}(z^k)(d^k)^T\nabla^2 G_{i_0}(z^k)d^k\Big)
\end{aligned}
$$

are bounded. On the other hand,

$$
\gamma_{i_0}^k\Big((d^k)^T\nabla G_{i_0}(z^k)\nabla H_{i_0}(z^k)^T d^k+(d^k)^T\nabla H_{i_0}(z^k)\nabla G_{i_0}(z^k)^T d^k\Big)=-2\gamma_{i_0}^k\to-\infty
$$

by the definition of $\{d^k\}$ and (4.61), and

$$
\gamma_{i_0}^k\Big(G_{i_0}(z^k)(d^k)^T\nabla^2 H_{i_0}(z^k)d^k+H_{i_0}(z^k)(d^k)^T\nabla^2 G_{i_0}(z^k)d^k\Big)
$$

is bounded by the convergence of the sequences

$$
\Big\{\gamma_{i_0}^k G_{i_0}(z^k)\Big\},\quad \Big\{\gamma_{i_0}^k H_{i_0}(z^k)\Big\}.
$$

In consequence, we have

$$
(d^k)^T\nabla_z^2 L_{\epsilon_k}(z^k,\lambda^k,\mu^k,\delta^k,\gamma^k)d^k\to-\infty
$$

as $k\to\infty$. This contradicts (4.62) and hence $\bar z$ is B-stationary to problem (4.1).

For the case where $i_0 \in \mathcal{I}_{\Phi_{\epsilon_k}}(z^k)$ for infinitely many k, we can show that $\bar{z}$ is B-stationary to problem (4.1) in a similar way as in the proof of Theorem 4.7. This completes the proof. ∎

The next example shows that the new condition in Theorem 4.8 is actually independent of the upper level strict complementarity condition employed in Corollary 4.3 and [38, 41, 57, 76].

Example 4.1 Consider the following problem:

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & z_1 + z_2 \geq 0, \quad z_2 \geq 0, \\ & z_2(z_1 + z_2) = 0. \end{array} \tag{4.63}$$

Then the modified relaxation scheme (4.3) for (4.63) can be written as

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & z_2(z_1 + z_2) + \epsilon(z_1 + 2z_2) \geq 0, \\ & z_2(z_1 + z_2) - \epsilon^2 \leq 0. \end{array} \tag{4.64}$$

Let

$$\begin{aligned} G(z) &:= z_1 + z_2, \\ H(z) &:= z_2, \\ \phi_\epsilon(z) &:= z_2(z_1 + z_2) + \epsilon(z_1 + 2z_2), \\ \psi_\epsilon(z) &:= z_2(z_1 + z_2) - \epsilon^2, \\ L_\epsilon(z, \delta, \gamma) &:= f(z) - \delta\phi_\epsilon(z) + \gamma\psi_\epsilon(z) \end{aligned}$$

and denote by $\bar{z}$ a solution of (4.63) with multipliers $(\bar{u}, \bar{v})$ satisfying (4.15)–(4.20) and by z_ϵ a solution of (4.64) with Lagrange multipliers $(\delta_\epsilon, \gamma_\epsilon)$. In order to show that the two conditions do not imply each other, we consider the following two cases.

(I) Let $f(z) := (z_1 + z_2)^2 + z_2^2$. Then we have $\bar{z} = z_\epsilon = (0, 0)$ for any $\epsilon > 0$. We can show that the smallest eigenvalue β_ϵ of $\nabla_z^2 L_\epsilon(z_\epsilon, \delta_\epsilon, \gamma_\epsilon)$ is independent of ϵ, whereas the upper level strict complementarity condition does not hold at $\bar{z}$. In fact, since $\delta_\epsilon = \gamma_\epsilon = 0$ for any $\epsilon > 0$, we have

$$\nabla_z^2 L_\epsilon(z_\epsilon, \delta_\epsilon, \gamma_\epsilon) = \begin{pmatrix} 2 & 2 - \delta_\epsilon + \gamma_\epsilon \\ 2 - \delta_\epsilon + \gamma_\epsilon & 4 - 2\delta_\epsilon + 2\gamma_\epsilon \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$$

and hence $\beta_\epsilon = 3 - \sqrt{5}$ for any $\epsilon > 0$. On the other hand, we have $\bar{u} = \bar{v} = 0$. This, together with $\mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}) = \{1\}$, implies that the upper level strict complementarity condition does not hold at $\bar{z}$.

(II) Let $f(z) := (z_1 + 1)^2 + (z_2 + 2)^2$. Then, for any $\epsilon > 0$, we have $\bar{z} = z_\epsilon = (0, 0)$. On the one hand, since $\bar{u} = \bar{v} = 2$ and $\mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z}) = \{1\}$, the upper level strict complementarity condition holds at $\bar{z}$. On the other hand, we have

$$\delta_\epsilon = 2\epsilon^{-1}, \quad \gamma_\epsilon = 0$$

and

$$\nabla_z^2 L_\epsilon(z_\epsilon, \delta_\epsilon, \gamma_\epsilon) = \begin{pmatrix} 2 & -\delta_\epsilon + \gamma_\epsilon \\ -\delta_\epsilon + \gamma_\epsilon & 2 - 2\delta_\epsilon + 2\gamma_\epsilon \end{pmatrix} = \begin{pmatrix} 2 & -2\epsilon^{-1} \\ -2\epsilon^{-1} & 2 - 4\epsilon^{-1} \end{pmatrix}.$$

It then follows that $\beta_\epsilon = 2 - 2(1 + \sqrt{2})\epsilon^{-1}$, which tends to $-\infty$ as $\epsilon \to 0^+$. This means that the condition given in Theorem 4.8 does not hold.

4.3 Concluding Remarks

In this chapter, we have proposed a modified relaxation scheme for a mathematical program with complementarity constraints. The new relaxed problem involves less constraints than the one considered by Scholtes [76]. All desirable properties established in [76] remain valid for the new relaxed problem. In addition, we obtain some new sufficient conditions for B-stationarity described by the eigenvalues of the Hessian matrix of the Lagrangian of the relaxed problem. From the proof, it is easy to see that, even if the matrix mentioned above is replaced by the Hessian matrix of the simpler function

$$\tilde{L}_\epsilon(z, \gamma, \delta) := \gamma^T \Psi_\epsilon(z) - \delta^T \Phi_\epsilon(z),$$

all the results remain true. Similar extension is possible for the relaxation schemes presented by Scholtes [76] and Lin and Fukushima [57] as well.

Finally, we remark that, if $\bar{z}$ is degenerate, the gradient of $\phi_{\epsilon,i}$ or $\psi_{\epsilon,i}$ at z_ϵ tends to 0 as $\epsilon \to 0^+$ for $i \in \mathcal{I}_G(\bar{z}) \cap \mathcal{I}_H(\bar{z})$. In fact, the method presented in [76] also has a similar problem. This is a possible deficiency of these methods, which may cause some numerical instability in practical calculations.

Chapter 5

Hybrid Approach with Active Set Identification for MPECs

Consider the following mathematical program with complementarity constraints (MPCC):

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & g(z) \leq 0,\ h(z) = 0 \\ & G(z) \geq 0,\ H(z) \geq 0 \\ & G(z)^T H(z) = 0, \end{array} \tag{5.1}$$

where $f : \Re^n \to \Re, g : \Re^n \to \Re^p, h : \Re^n \to \Re^q$, and $G, H : \Re^n \to \Re^m$ are all twice continuously differentiable functions. There have been proposed several approaches such as sequential quadratic programming approach [29, 44, 62], implicit programming approach [12, 62], penalty function approach [38, 41, 59, 62, 63, 77], active-set approach [33], and reformulation approach [24, 31, 54, 57, 76]. However, these methods require to solve an infinite sequence of nonlinear programs. The purpose of this chapter is to develop some methods that enable us to compute a solution or a point with some kind of stationarity to problem (5.1) by solving a finite number of nonlinear programs. To this end, we will apply an active set identification technique to a smoothing continuation method [31] and present some hybrid algorithms. Further discussions and some extensions will also be included.

5.1 Preliminaries

Consider the smoothing continuation method [31] that uses the perturbed Fischer-Burmeister function

$$\phi_\epsilon(a,b) := a + b - \sqrt{a^2 + b^2 + 2\epsilon^2},$$

where $\epsilon \geq 0$. Define the function $\Phi_\epsilon : \Re^n \to \Re^m$ by

$$\Phi_\epsilon(z) := \Big(\phi_\epsilon(G_1(z), H_1(z)), \cdots, \phi_\epsilon(G_m(z), H_m(z))\Big)^T$$

and consider the nonlinear programming problem

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & \Phi_\epsilon(z) = 0, \\ & g(z) \leq 0, \;\; h(z) = 0. \end{array} \qquad (P_\epsilon)$$

Note that (P_0) is equivalent to problem (5.1) and Φ_ϵ is differentiable everywhere for any $\epsilon > 0$. We assume that problem (P_ϵ) has a solution (or a stationary point) z^ϵ for each small scalar $\epsilon > 0$. We may expect to find a solution or a point with some kind of stationarity to problem (5.1) by tracing the trajectory $\{z^\epsilon\}$ as $\epsilon \to 0$. Suppose that $\{\epsilon_k\}$ is a positive sequence converging to zero. The following convergence result is given in [31].

Theorem 5.1 *Let z^k be a stationary point of problem (P_{ϵ_k}) and the sequence $\{z^k\}$ converge to z^* as $\epsilon_k \to 0$. Suppose that the WSONC holds at each z^k, the MPCC-LICQ holds at z^*, and $\{z^k\}$ is asymptotically weakly nondegenerate. Then z^* is B-stationary to problem (5.1).*

We now recall the asymptotically weak nondegeneracy [31], which is assumed in the above theorem and will also be employed in the subsequent analysis. Suppose $\{z^k\}$ converges to z^* as $\epsilon_k \to 0$. Then $z^* \in \mathcal{F}$. It can be shown [31] that, for each $i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*)$,

$$\nabla\Phi_{\epsilon_k,i}(z^k) = \frac{H_i(z^k)}{G_i(z^k) + H_i(z^k)}\nabla G_i(z^k) + \frac{G_i(z^k)}{G_i(z^k) + H_i(z^k)}\nabla H_i(z^k). \qquad (5.2)$$

Therefore, every accumulation point r of $\Big\{\nabla\Phi_{\epsilon_k,i}(z^k)\Big\}$ can be represented as

$$r = \xi_i(r)\nabla G_i(z^*) + \eta_i(r)\nabla H_i(z^*) \qquad (5.3)$$

for some $(\xi_i(r), \eta_i(r))$ with $(1-\xi_i(r))^2 + (1-\eta_i(r))^2 \le 1$. We say that $\{z^k\}$ is *asymptotically weakly nondegenerate* if, for each $i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*)$, neither $\xi_i(r)$ nor $\eta_i(r)$ vanishes for any accumulation point r of $\left\{\nabla\Phi_{\epsilon_k,i}(z^k)\right\}$.

Roughly speaking, the asymptotically weak nondegeneracy of $\{z^k\}$ means that, for each $i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*)$, $G_i(z^k)$ and $H_i(z^k)$ approach zero in the same order of magnitude. This property is obviously weaker than the nondegeneracy (lower-level strict complementarity), because it vacuously holds when z^* is nondegenerate. We will show in Section 5.5 that it is even weaker than the upper-level strict complementarity (ULSC) condition, which is often employed in the literature on MPCC, see [38, 41, 54, 57, 76].

In addition, we can prove that the smoothing continuation method possesses similar convergence properties to the methods proposed in Chapters 3 and 4. Here, we state one of such results.

Theorem 5.2 *Let z^k be a stationary point of problem (P_{ϵ_k}) and, for each k, $(\lambda_g^k, \lambda_h^k, \lambda_\Phi^k)$ be a multiplier vector corresponding to z^k. Suppose that the sequence $\{z^k\}$ converges to z^* as $\epsilon_k \to 0$ and, for each k, $\nabla_z^2 L_{\epsilon_k}(z^k, \lambda_g^k, \lambda_h^k, \lambda_\Phi^k)$ is bounded below with constant $\alpha_k \ge 0$ on the corresponding tangent space $\mathcal{T}_{\varepsilon_k}(z^k)$, which means*

$$d^T \nabla_z^2 L_{\epsilon_k}(z^k, \lambda_g^k, \lambda_h^k, \lambda_\Phi^k) d \ge -\alpha_k ||d||^2, \quad \forall d \in \mathcal{T}_{\epsilon_k}(z^k), \tag{5.4}$$

where

$$\begin{aligned} L_\epsilon(z, \lambda_g, \lambda_h, \lambda_\Phi) &:= f(z) + \lambda_g^T g(z) + \lambda_h^T h(z) + \lambda_\Phi^T \Phi_\epsilon(z), \\ \mathcal{T}_\epsilon(z) &:= \Big\{ d \in \Re^n \mid \; d^T \nabla\Phi_{\epsilon,i}(z) = 0, \; i = 1, 2, \cdots, m; \\ &\qquad\qquad d^T \nabla g_l(z) = 0, \quad l \in \mathcal{I}_g(z); \\ &\qquad\qquad d^T \nabla h_r(z) = 0, \quad r = 1, 2, \cdots, q \Big\}. \end{aligned}$$

If the sequence $\{\alpha_k\}$ is bounded, $\{z^k\}$ is asymptotically weakly nondegenerate, and the MPCC-LICQ holds at z^*, then z^* is a B-stationary point of problem (5.1).

Actually, the condition that problem (P_{ϵ_k}) satisfies the WSONC at z^k for each k means that $\nabla_z^2 L_{\epsilon_k}(z^k, \lambda_g^k, \lambda_h^k, \lambda_\Phi^k)$ is bounded below with constant 0. In consequence, Theorem 5.1 is actually a corollary of Theorem 5.2. Note that, for the matrix $\nabla_z^2 L_{\epsilon_k}(z^k, \lambda_g^k, \lambda_h^k, \lambda_\Phi^k)$, there must exist a number α_k such that (5.4) holds. For example, any nonnegative scalar α_k such that $(-\alpha_k)$ is less than the smallest eigenvalue of $\nabla_z^2 L_{\epsilon_k}(z^k, \lambda_g^k, \lambda_h^k, \lambda_\Phi^k)$ must satisfy (5.4). However, the WSONC means that the matrix should have some kind of semi-definiteness on the tangent space $\mathcal{T}_{\epsilon_k}(z^k)$. Therefore, the assumptions in Theorem 5.2 are weaker than the conditions of Theorem 5.1. Since the proof of Theorem 5.2 is similar to that of Theorem 5.1, it is omitted here, see [31].

5.2 A Hybrid Algorithm for MPCC

We first introduce an active set identification technique for MPCC. Active set identification plays an important role in optimization theory [6, 8, 22, 23, 82]. Accurate identification of active constraints is important from both theoretical and practical points of view. For problem (5.1), by means of active set identification, the combinatorial constraints

$$G(z) \geq 0, \quad H(z) \geq 0, \quad G(z)^T H(z) = 0 \tag{5.5}$$

may be replaced by some equality and/or inequality constraints that are easier to deal with.

For a point $\bar{z} \in \mathcal{F}$, let $\alpha(\bar{z}), \beta(\bar{z})$ and $\gamma(\bar{z})$ be the index sets defined by

$$\begin{aligned} \alpha(\bar{z}) &:= \{i \mid G_i(\bar{z}) > 0,\ H_i(\bar{z}) = 0\}, \\ \beta(\bar{z}) &:= \{i \mid G_i(\bar{z}) = 0,\ H_i(\bar{z}) = 0\}, \\ \gamma(\bar{z}) &:= \{i \mid G_i(\bar{z}) = 0,\ H_i(\bar{z}) > 0\}, \end{aligned}$$

respectively. Obviously, $\alpha(\bar{z}) \cup \beta(\bar{z}) \cup \gamma(\bar{z}) = \{1, \ldots, m\}$ and these index sets are mutually disjoint. Let a sequence $\{z^k\}$ be generated so that it converges to $z^* \in \mathcal{F}$. If z^* is nondegenerate, namely, if $\beta(z^*) = \emptyset$, it is generally not difficult to identify the correct index sets finitely. However, when z^* is degenerate, it is not necessarily easy to identify the active index sets. In the following, we will particularly be interested in the case where $\{z^k\}$ is convergent to a degenerate point $z^* \in \mathcal{F}$.

Let $\{\epsilon_k\}$ be a positive sequence converging to zero and let z^k stand for a solution or a stationary point of problem (P_{ϵ_k}) for each k. Suppose that $\{z^k\}$ converges to some z^* throughout this chapter. Note that

$$G_i(z^k) > 0,\ H_i(z^k) > 0,\ G_i(z^k)H_i(z^k) = \epsilon_k^2 \tag{5.6}$$

for each k and each i. We try to estimate the index sets $\alpha(z^*)$, $\beta(z^*)$ and $\gamma(z^*)$ by some index sets α^k, β^k and γ^k, respectively, which are obtained from z^k and satisfy $\alpha^k \cup \beta^k \cup \gamma^k = \{1, \ldots, m\}$ with $\alpha^k, \beta^k, \gamma^k$ being mutually disjoint. Given the index sets α^k, β^k and γ^k, we then solve the nonlinear programming problem

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & G_i(z) \geq 0,\ H_i(z) = 0, \quad i \in \alpha^k, \\ & G_i(z) = 0,\ H_i(z) = 0, \quad i \in \beta^k, \\ & G_i(z) = 0,\ H_i(z) \geq 0, \quad i \in \gamma^k, \\ & g(z) \leq 0,\ h(z) = 0. \end{array} \tag{5.7}$$

This problem is no longer an MPCC and hence easier to deal with. Denote by $\hat{z}^k$ a stationary point of problem (5.84). Obviously, we always have $\beta^k \subseteq \beta(\hat{z}^k)$, and $\beta(\hat{z}^k)$ may contain some $i \in \alpha^k$ with $G_i(\hat{z}^k) = 0$ or some $i \in \gamma^k$ with $H_i(\hat{z}^k) = 0$. If the Lagrange multipliers corresponding to the constraints

$$\begin{aligned} &G_i(z) \geq 0,\ H_i(z) = 0, \quad i \in \alpha^k \cap \beta(\hat{z}^k), \\ &G_i(z) = 0,\ H_i(z) = 0, \quad i \in \beta^k, \\ &G_i(z) = 0,\ H_i(z) \geq 0, \quad i \in \gamma^k \cap \beta(\hat{z}^k) \end{aligned} \tag{5.8}$$

are all nonnegative, then $\hat{z}^k$ is a B-stationary point of problem (5.1) under the MPCC-LICQ assumption at the point [33]. Therefore, assuming that (5.84) can be solved exactly, we may terminate the method in finite steps, unlike the method in [31], which needs to solve an infinite sequence of nonlinear programs.

The key to success is to define the index sets α^k, β^k and γ^k such that

$$\alpha^k = \alpha(z^*),\ \beta^k = \beta(z^*),\ \gamma^k = \gamma(z^*) \tag{5.9}$$

for all k large enough. To this end, we may use an *identification function* $\rho : \Re^n \to [0, +\infty)$ satisfying

$$\lim_{k\to\infty} \rho(z^k) \ = \ 0 \tag{5.10}$$

and, for all k large enough,

$$\max_{i\in\beta(z^*)} \Big\{G_i(z^k), H_i(z^k)\Big\} \ \leq \ \rho(z^k), \tag{5.11}$$

$$\max_{i\in\alpha(z^*)\cup\gamma(z^*)} \Big\{\min\{G_i(z^k), H_i(z^k)\}\Big\} \ \leq \ \rho(z^k), \tag{5.12}$$

and consider the following hybrid algorithm that combines the smoothing continuation method with an active set identification technique.

Algorithm H:

Step 0: Choose $\epsilon_0 > 0$ and set $k := 0$.

Step 1: Solve problem (P_{ϵ_k}) and denote by z^k one of its stationary points. Set

$$\alpha^k \ := \ \Big\{ i \,\Big|\, G_i(z^k) > \rho(z^k), \quad H_i(z^k) \leq \rho(z^k) \Big\}, \tag{5.13}$$

$$\beta^k \ := \ \Big\{ i \,\Big|\, G_i(z^k) \leq \rho(z^k), \quad H_i(z^k) \leq \rho(z^k) \Big\}, \tag{5.14}$$

$$\gamma^k \ := \ \Big\{ i \,\Big|\, G_i(z^k) \leq \rho(z^k), \quad H_i(z^k) > \rho(z^k) \Big\}. \tag{5.15}$$

If $\alpha^k \cup \beta^k \cup \gamma^k = \{1, \ldots, m\}$, go to Step 2. Otherwise, go to Step 4.

Step 2: Solve problem (5.84) to get a stationary point $\hat{z}^k$ and go to Step 3.

Step 3: If the Lagrange multipliers corresponding to the constraints (5.85) are all nonnegative, then terminate. Else, go to Step 4.

Step 4: Choose an $\epsilon_{k+1} \in (0, \epsilon_k)$ and let $k := k+1$. Go to Step 1.

Next, we make some remarks on the identification function ρ and Algorithm H.

First of all, we have that, if $\beta(z^*) \neq \emptyset$,

$$\max_{i \in \alpha(z^*) \cup \gamma(z^*)} \Big\{ \min\{G_i(z^k), H_i(z^k)\} \Big\} < \min_{i \in \beta(z^*)} \Big\{ G_i(z^k), H_i(z^k) \Big\} \tag{5.16}$$

holds for all k large enough as long as $\alpha(z^*) \cup \gamma(z^*) \neq \emptyset$. In fact, we have from (5.6) that

$$\begin{aligned} \lim_{k\to\infty} \frac{\epsilon_k^2}{\min\{G_i(z^k), H_i(z^k)\}} &= \lim_{k\to\infty} \max\{G_i(z^k), H_i(z^k)\} \\ &= 0, \qquad\qquad i \in \beta(z^*) \end{aligned}$$

and

$$\begin{aligned} \lim_{k\to\infty} \frac{\epsilon_k^2}{\min\{G_j(z^k), H_j(z^k)\}} &= \lim_{k\to\infty} \max\{G_j(z^k), H_j(z^k)\} \\ &= G_j(z^*) + H_j(z^*) > 0, \qquad j \in \alpha(z^*) \cup \gamma(z^*). \end{aligned}$$

In consequence,

$$\lim_{k\to\infty} \frac{\min\{G_j(z^k), H_j(z^k)\}}{\min\{G_i(z^k), H_i(z^k)\}} = 0 \tag{5.17}$$

holds for each $i \in \beta(z^*)$ and each $j \in \alpha(z^*) \cup \gamma(z^*)$ and hence we have (5.16). This inequality means that condition (5.11) implies condition (5.12) for all k sufficiently large if $\beta(z^*) \neq \emptyset$ and $\alpha(z^*) \cup \gamma(z^*) \neq \emptyset$.

Moreover, it is obvious that $\{\alpha^k, \beta^k, \gamma^k\}$ defined by (5.13)–(5.15) is mutually disjoint for each k. On the other hand, conditions (5.11) and (5.12) ensure that, when k is sufficiently large,

$$\min\{G_i(z^k), H_i(z^k)\} \leq \rho(z^k), \quad \forall i.$$

This means $\alpha^k \cup \beta^k \cup \gamma^k = \{1, \ldots, m\}$ and so $\{\alpha^k, \beta^k, \gamma^k\}$ defined in Step 1 is a partition of $\{1, \ldots, m\}$ for all k sufficiently large.

Furthermore, we have from (5.10) that, when k is sufficiently large,

$$\alpha^k \supseteq \alpha(z^*), \;\; \beta^k \subseteq \beta(z^*), \;\; \gamma^k \supseteq \gamma(z^*).$$

In addition, it follows from (5.11) and (5.14) that $\beta(z^*) \subseteq \beta^k$ for all k sufficiently large. Note that both $\{\alpha^k, \beta^k, \gamma^k\}$ and $\{\alpha(z^*), \beta(z^*), \gamma(z^*)\}$ are partitions of $\{1, \ldots, m\}$. Therefore, (5.9) holds when k is large enough.

The above analysis indicates that Algorithm H may possess a finite termination property. The key question is, of course, how to define the identification function ρ. This is not a trivial task because the function may depend on the unknown point z^* generally. Next, we consider the case where $\{z^k\}$ is asymptotically weakly nondegenerate.

Theorem 5.3 *Suppose the sequence $\{z^k\}$ generated by Algorithm H is asymptotically weakly nondegenerate. Let*

$$\rho_1(z) := \tau || \min(G(z), H(z))||^\sigma, \tag{5.18}$$

where $\tau > 0$ and $\sigma \in (0,1)$ are constants and

$$\min(G(z), H(z)) := \Big(\min\{G_1(z), H_1(z)\}, \cdots, \min\{G_m(z), H_m(z)\}\Big)^T. \tag{5.19}$$

Then (5.10)–(5.12) hold with $\rho = \rho_1$, i.e., the function ρ_1 can serve as an identification function.

Proof: Note that the complementarity constraints (5.5) are equivalent to $\rho_1(z) = 0$. By the fact that $\{z^k\}$ is convergent to $z^* \in \mathcal{F}$ and the continuity of ρ_1, we have

$$\lim_{k\to\infty} \rho_1(z^k) = \rho_1(z^*) = 0.$$

Therefore, condition (5.10) holds with $\rho = \rho_1$. On the other hand, for any $i \in \alpha(z^*) \cup \gamma(z^*)$, we have

$$\min\Big\{G_i(z^k), H_i(z^k)\Big\} \le || \min(G(z^k), H(z^k))|| \le \rho_1(z^k)$$

for all k sufficiently large, where the first inequality follows from (5.19) and the second inequality follows from the fact that $\{|| \min(G(z^k), H(z^k))||\}$ is convergent to 0 and the constant σ lies in the interval $(0, 1)$. This means that (5.12) with $\rho = \rho_1$ holds when k is sufficiently large.

We next prove (5.11) with $\rho = \rho_1$ holds when k is sufficiently large. We may assume $\beta(z^*) \neq \emptyset$, because (5.11) holds vacuously if $\beta(z^*)$ is empty. Let $i \in \beta(z^*)$. Since the set of accumulation points of the sequence $\Big\{\nabla\Phi_{\epsilon_k,i}(z^k)\Big\}$ is compact, by the asymptotically weak nondegeneracy of $\{z^k\}$, the set of the coefficient pairs in (5.3) is a compact subset of

$$\Re^2_{++} := \Big\{(\xi, \eta)^T \in \Re^2 \mid \xi > 0, \eta > 0\Big\}.$$

Then, by (5.76) and (5.3), there exist positive constants $a_i < b_i$ such that

$$a_i \leq \frac{G_i(z^k)}{H_i(z^k)} \leq b_i, \qquad \forall k.$$

Let $a := \min_{i \in \beta(z^*)} a_i$ and $b := \max_{i \in \beta(z^*)} b_i$. It follows that

$$0 < a \leq \frac{G_i(z^k)}{H_i(z^k)} \leq b \tag{5.20}$$

for each k and each $i \in \beta(z^*)$. We then have from (5.20) that

$$G_i(z^k) \leq bH_i(z^k), \quad H_i(z^k) \leq a^{-1}G_i(z^k)$$

and so

$$\max\{G_i(z^k), H_i(z^k)\} \leq (a^{-1} + b) \min\{G_i(z^k), H_i(z^k)\} \tag{5.21}$$

for each k and each $i \in \beta(z^*)$. By (5.21) and (5.19), we have

$$\begin{aligned} \max\Big\{G_i(z^k), H_i(z^k)\Big\} &\leq (a^{-1} + b)|| \min(G(z^k), H(z^k))|| \\ &\leq \rho_1(z^k) \end{aligned}$$

for all k sufficiently large, where the last inequality follows from the same facts as above. This completes the proof of (5.11) and so the function ρ_1 given by (5.18) can serve as an identification function. □

Theorem 5.4 *Suppose that the sequence $\{z^k\}$ generated by Algorithm H is asymptotically weakly nondegenerate. Then the function*

$$\rho_2(z) := \tau||\Phi_0(z)||^\sigma, \qquad \tau > 0, \ \sigma \in (0, 1) \tag{5.22}$$

is an identification function.

Proof: Noting that, for each i and each k,

$$\begin{aligned} \frac{2}{2+\sqrt{2}}|| \min(G(z^k), H(z^k))|| &\leq ||\Phi_0(z^k)|| \\ &\leq (2+\sqrt{2})|| \min(G(z^k), H(z^k))|| \end{aligned}$$

(see [81]), we have

$$\begin{aligned} \tau\Big(\frac{2}{2+\sqrt{2}}\Big)^\sigma || \min(G(z^k), H(z^k))||^\sigma &\leq \rho_2(z^k) \\ &\leq \tau(2+\sqrt{2})^\sigma || \min(G(z^k), H(z^k))||^\sigma. \end{aligned} \tag{5.23}$$

Let

$$\begin{aligned}\hat{\rho}_1(z) &:= \tau\Big(\frac{2}{2+\sqrt{2}}\Big)^{\sigma} \|\min(G(z),H(z))\|^{\sigma},\\ \bar{\rho}_1(z) &:= \tau(2+\sqrt{2})^{\sigma}\|\min(G(z),H(z))\|^{\sigma}.\end{aligned}$$

By Theorem 5.3, both $\hat{\rho}_1$ and $\bar{\rho}_1$ are identification functions for Algorithm H. As a result, condition (5.10) holds with $\bar{\rho}_1$. This, together with the second inequality in (5.23), implies that condition (5.10) holds with $\rho = \rho_2$. On the other hand, the first inequality in (5.23), together with the fact that conditions (5.11) and (5.12) with $\rho = \hat{\rho}_1$ hold for all k sufficiently large, means that (5.11) and (5.12) hold with $\rho = \rho_2$ when k is sufficiently large. In consequence, the function ρ_2 satisfies conditions (5.10)–(5.12). This completes the proof. □

The next lemma indicates that problem (P_ϵ) satisfies the standard LICQ under some appropriate assumptions, unlike problem (5.1), which fails to satisfy any constraint qualification at any feasible point. A proof of the lemma may be found in [31].

Lemma 5.1 *If the MPCC-LICQ holds at z^*, then there exist a neighborhood $U(z^*)$ of z^* and a positive constant ϵ^* such that, for any $\epsilon \in (0, \epsilon^*)$, problem (P_ϵ) satisfies the standard LICQ at any feasible point in $U(z^*)$.*

Summarizing the above arguments, we obtain the following concluding result.

Theorem 5.5 *Let $\sigma \in (0,1), \tau > 0$, and ρ be ρ_1 defined by (5.18) or ρ_2 defined by (5.22). Suppose that the sequence $\{z^k\}$ of stationary points of problems (P_{ϵ_k}) converges to z^* as $\epsilon_k \to 0$, the MPCC-LICQ holds at z^*, and the sequence $\{z^k\}$ is asymptotically weakly nondegenerate. Then, for all k sufficiently large, we have*

(i) *problem (P_{ϵ_k}) satisfies the standard LICQ at z^k;*

(ii) *the sets α^k, β^k, and γ^k defined by (5.13), (5.14), and (5.15) satisfy (5.9).*

This theorem indicates that, by means of the technique introduced above, we can identify the active sets $\alpha(z^*), \beta(z^*)$, and $\gamma(z^*)$ in a finite number of iterations under some mild conditions. As a result, if the problem

$$\begin{aligned}&\text{minimize} && f(z)\\ &\text{subject to} && G_i(z) \ge 0,\ H_i(z) = 0, \quad i \in \alpha(z^*),\\ &&& G_i(z) = 0,\ H_i(z) = 0, \quad i \in \beta(z^*),\\ &&& G_i(z) = 0,\ H_i(z) \ge 0, \quad i \in \gamma(z^*),\\ &&& g(z) \le 0,\ h(z) = 0\end{aligned} \tag{5.24}$$

can be solved exactly, we may expect that Algorithm H terminates in finite steps. In particular, if z^* is a stationary point obtained by solving (5.24), then, under the assumptions of either Theorem 5.1 or Theorem 5.2, the algorithm terminates in a finite number of iterations by producing a B-stationary point of problem (5.1). Note that it is possible that the algorithm terminates by producing a B-stationary point of (5.1) before we solve problem (5.24).

Remark 5.1 In order to ensure $\alpha^k \cup \beta^k \cup \gamma^k = \{1, 2, \ldots, m\}$ in Step 1 for each k, we may define

$$\beta^k := \{1, \ldots, m\} \setminus (\alpha^k \cup \gamma^k)$$

instead of (5.14) in Algorithm H.

Remark 5.2 Under the assumption of asymptotically weak nondegeneracy of $\{z^k\}$, we may define $\{\alpha^k, \beta^k, \gamma^k\}$ in a different manner in Step 1. For example, we may put

$$\alpha^k := \left\{ i \,\middle|\, G_i(z^k) > \sqrt{\epsilon_k}, \quad H_i(z^k) \le \sqrt{\epsilon_k} \right\}, \tag{5.25}$$

$$\beta^k := \left\{ i \,\middle|\, G_i(z^k) \le \sqrt{\epsilon_k}, \quad H_i(z^k) \le \sqrt{\epsilon_k} \right\}, \tag{5.26}$$

$$\gamma^k := \left\{ i \,\middle|\, G_i(z^k) \le \sqrt{\epsilon_k}, \quad H_i(z^k) > \sqrt{\epsilon_k} \right\} \tag{5.27}$$

instead of (5.13)–(5.15). In fact, the proof of Theorem 5.3 indicates that there exist two positive numbers a and b such that (5.20) holds for each $i \in \beta(z^*)$ and each k. Therefore, by the equality in (5.6) and the asymptotically weak nondegeneracy, we deduce

$$G_i(z^k) = O(\epsilon_k), \quad H_i(z^k) = O(\epsilon_k), \qquad i \in \beta(z^*), \tag{5.28}$$

$$\min\{G_i(z^k), H_i(z^k)\} = O(\epsilon_k^2), \qquad i \in \alpha(z^*) \cup \gamma(z^*). \tag{5.29}$$

Since $\{\sqrt{\epsilon_k}\}$ approaches zero slower than both $\{\epsilon_k\}$ and $\{\epsilon_k^2\}$, we have from (5.28)–(5.29) that

$$\min\{G_i(z^k), H_i(z^k)\} \le \sqrt{\epsilon_k}$$

for each i and each k sufficiently large. This means $\alpha^k \cup \beta^k \cup \gamma^k = \{1, 2, \ldots, m\}$ when k is sufficiently large. Noticing that $\{\alpha^k, \beta^k, \gamma^k\}$ defined by (5.25)–(5.27) is mutually disjoint for each k, we have that $\{\alpha^k, \beta^k, \gamma^k\}$ given by (5.25)–(5.27) is a partition of $\{1, \ldots, m\}$ for all k sufficiently large. On the other hand, it is obvious that, when k is sufficiently large,

$$\alpha^k \supseteq \alpha(z^*), \ \beta^k \subseteq \beta(z^*), \ \gamma^k \supseteq \gamma(z^*).$$

Moreover, it follows from (5.26) and (5.28) that $\beta(z^*) \subseteq \beta^k$ for all k sufficiently large. Note that both $\{\alpha^k, \beta^k, \gamma^k\}$ and $\{\alpha(z^*), \beta(z^*), \gamma(z^*)\}$ are partitions of $\{1, \ldots, m\}$. Therefore, (5.9) holds when k is large enough. So, we may employ (5.25)–(5.27) instead of (5.13)–(5.15) in Step 1 of Algorithm H. From the computational point of view, (5.25)–(5.27) are simpler than (5.13)–(5.15).

5.3 Modified Hybrid Method with Index Addition Strategy

For Algorithm H, the asymptotically weak nondegeneracy condition is a key assumption. Although this condition is not excessively stringent because it is implied by the ULSC condition (see Section 5.5), nevertheless, it is certainly desirable to lessen the required assumptions. In this section, we introduce a modified method with index addition strategy and in the next section, we describe another method with the converse strategy. Both of these two methods do not require the assumption of asymptotically weak nondegeneracy.

Let ρ be a function defined by (5.18) or (5.22) with $\tau > 0$ and $\sigma \in (0, 1)$.

Algorithm HIA:

Step 0: Choose $\theta_0 > 0$ and $\epsilon_0 > 0$. Set $k := 0$.

Step 1: Solve problem (P_{ϵ_k}) to obtain a stationary point z^k and set

$$\alpha_0^k := \left\{ i \;\middle|\; G_i(z^k) > \rho(z^k), \quad H_i(z^k) \le \rho(z^k) \right\}, \tag{5.30}$$

$$\beta_0^k := \left\{ i \;\middle|\; G_i(z^k) \le \rho(z^k), \quad H_i(z^k) \le \rho(z^k) \right\}, \tag{5.31}$$

$$\gamma_0^k := \left\{ 1, 2, \cdots, m \right\} \setminus \left(\alpha_0^k \cup \beta_0^k \right), \tag{5.32}$$

$$\delta_0^k := +\infty,$$

and $j := 0$. Go to Step 2.

Step 2: Solve the problem

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & G_i(z) \ge 0, \; H_i(z) = 0, \quad i \in \alpha_j^k, \\ & G_i(z) = 0, \; H_i(z) = 0, \quad i \in \beta_j^k, \\ & G_i(z) = 0, \; H_i(z) \ge 0, \quad i \in \gamma_j^k, \\ & g(z) \le 0, \; h(z) = 0 \end{array} \tag{5.33}$$

to get a stationary point $\hat{z}_j^k$.

Step 3: If, in (5.86), the Lagrange multipliers corresponding to the constraints

$$\begin{aligned} &G_i(z) \geq 0,\ H_i(z) = 0, \quad i \in \alpha_j^k \cap \beta(\hat{z}_j^k), \\ &G_i(z) = 0,\ H_i(z) = 0, \quad i \in \beta_j^k, \\ &G_i(z) = 0,\ H_i(z) \geq 0, \quad i \in \gamma_j^k \cap \beta(\hat{z}_j^k) \end{aligned} \tag{5.34}$$

are all nonnegative, then terminate. Else, if there is an $\hat{i} \in \alpha_j^k \cup \gamma_j^k$ such that

$$\min_{i \in \alpha_j^k \cup \gamma_j^k} \Big\{ \max \Big\{ G_i(z^k), H_i(z^k) \Big\} \Big\} = \max \Big\{ G_{\hat{i}}(z^k), H_{\hat{i}}(z^k) \Big\} < \theta_k, \tag{5.35}$$

then set

$$\delta_{j+1}^k := \min \Big\{ \delta_j^k,\ \frac{1}{2} \min \Big\{ G_i(\hat{z}_j^k) + H_i(\hat{z}_j^k) \mid \quad i \in \alpha(\hat{z}_j^k) \cup \gamma(\hat{z}_j^k) \Big\} \Big\} \tag{5.36}$$

and

$$\alpha_{j+1}^k := \alpha_j^k \setminus \{\hat{i}\}, \tag{5.37}$$

$$\beta_{j+1}^k := \beta_j^k \cup \{\hat{i}\}, \tag{5.38}$$

$$\gamma_{j+1}^k := \gamma_j^k \setminus \{\hat{i}\}, \tag{5.39}$$

$$j := j + 1$$

and go to Step 2. Otherwise, let $j_k := j$ and $\delta_k := \delta_j^k$ and go to Step 4.

Step 4: Choose $\epsilon_{k+1} \in (0, \epsilon_k)$ and set $\theta_{k+1} := \min \Big\{ \theta_k,\ \delta_k \Big\}$. Go to Step 1 with $k := k + 1$.

The next theorem describes the relations between the sets $\{\alpha_j^k, \beta_j^k, \gamma_j^k\}$, $j = 0, 1, \cdots, j_k$, and $\{\alpha(z^*), \beta(z^*), \gamma(z^*)\}$.

Theorem 5.6 *Suppose that the sequence $\{z^k\}$ generated by Algorithm HIA converges to z^* as $\epsilon_k \to 0$. Then there is an integer $k_0 \geq 0$ such that, for any $k \geq k_0$,*

(i) $\alpha_0^k \supseteq \alpha(z^*)$, $\beta_0^k \subseteq \beta(z^*)$, $\gamma_0^k \supseteq \gamma(z^*)$;

(ii) *if $\beta(z^*) = \emptyset$, namely, z^* is nondegenerate, then we have*

$$\alpha_0^k = \alpha(z^*), \quad \beta_0^k = \emptyset, \quad \gamma_0^k = \gamma(z^*); \tag{5.40}$$

(iii) *if* $\beta(z^*) \neq \emptyset$ *and, for some* j, β_j^k *is a proper subset of* $\beta(z^*)$*, then any index* $\hat{i} \in \alpha_j^k \cup \gamma_j^k$ *satisfying (5.35) belongs to* $\beta(z^*)$ *and hence we have*

$$\alpha_0^k \supseteq \alpha_1^k \supseteq \cdots \supseteq \alpha_j^k \supseteq \alpha_{j+1}^k \supseteq \alpha(z^*), \tag{5.41}$$

$$\beta_0^k \subseteq \beta_1^k \subseteq \cdots \subseteq \beta_j^k \subseteq \beta_{j+1}^k \subseteq \beta(z^*), \tag{5.42}$$

$$\gamma_0^k \supseteq \gamma_1^k \supseteq \cdots \supseteq \gamma_j^k \supseteq \gamma_{j+1}^k \supseteq \gamma(z^*). \tag{5.43}$$

Proof: First of all, we note that

$$\alpha_0^k \supseteq \alpha_1^k \supseteq \cdots \supseteq \alpha_j^k \supseteq \alpha_{j+1}^k,$$
$$\beta_0^k \subseteq \beta_1^k \subseteq \cdots \subseteq \beta_j^k \subseteq \beta_{j+1}^k,$$
$$\gamma_0^k \supseteq \gamma_1^k \supseteq \cdots \supseteq \gamma_j^k \supseteq \gamma_{j+1}^k$$

by (5.37)–(5.39). Next we show the existence of the integer k_0. We only consider the case where $\rho = \rho_2$. We may deal with the other case similarly.

Since $\{z^k\}$ converges to $z^* \in \mathcal{F}$, we have from the continuity of the function ρ that

$$\lim_{k\to\infty} \rho(z^k) = 0 \tag{5.44}$$

and hence, for each $i \in \alpha(z^*) \cup \gamma(z^*)$,

$$\begin{aligned}
\max\Big\{G_i(z^k), H_i(z^k)\Big\} &> \rho(z^k) \quad \Big(:= \tau\|\Phi_0(z^k)\|^\sigma\Big) \qquad (5.45)\\
&\geq 2\,\|\Phi_0(z^k)\|\\
&\geq 2\phi_0(G_i(z^k), H_i(z^k))\\
&= 2\Big(G_i(z^k) + H_i(z^k) - \sqrt{(G_i(z^k))^2 + (H_i(z^k))^2}\Big)\\
&= \frac{4G_i(z^k)H_i(z^k)}{G_i(z^k) + H_i(z^k) + \sqrt{(G_i(z^k))^2 + (H_i(z^k))^2}}\\
&\geq \min\Big\{G_i(z^k), H_i(z^k)\Big\} \qquad (5.46)
\end{aligned}$$

when k is large enough, where the second inequality follows from the fact that $\{\|\Phi_0(z^k)\|\}$ is convergent to 0 and the constant σ lies in the interval $(0, 1)$, and the last inequality follows from (5.6) and the fact that

$$\begin{aligned}
G_i(z^k) + H_i(z^k) + \sqrt{(G_i(z^k))^2 + (H_i(z^k))^2} &\leq 2(G_i(z^k) + H_i(z^k))\\
&\leq 4\max\Big\{G_i(z^k), H_i(z^k)\Big\}.
\end{aligned}$$

Thus, we have from (5.45)–(5.46) and the continuity of the functions G and H that, for any k sufficiently large,

$$G_i(z^k) > \rho(z^k), \quad H_i(z^k) \le \rho(z^k), \qquad i \in \alpha(z^*), \tag{5.47}$$

$$G_i(z^k) \le \rho(z^k), \quad H_i(z^k) > \rho(z^k), \qquad i \in \gamma(z^*). \tag{5.48}$$

Note that (5.30)–(5.32) imply

$$\gamma_0^k \supseteq \left\{ i \mid \quad G_i(z^k) \le \rho(z^k), \quad H_i(z^k) > \rho(z^k) \right\} \tag{5.49}$$

for each k. Moreover, it is obvious from (5.44) that $\beta_0^k \subseteq \beta(z^*)$ when k is sufficiently large. It then follows from (5.30)–(5.32) and (5.47)–(5.49) that there exists an integer $k_1 \ge 0$ such that, for any $k \ge k_1$,

$$\alpha_0^k \supseteq \alpha(z^*), \ \beta_0^k \subseteq \beta(z^*), \ \gamma_0^k \supseteq \gamma(z^*). \tag{5.50}$$

If z^* is nondegenerate, then (5.40) follows from (5.50) and the fact that both $\{\alpha_0^k, \beta_0^k, \gamma_0^k\}$ and $\{\alpha(z^*), \beta(z^*), \gamma(z^*)\}$ are partitions of the set $\{1, 2, \cdots, m\}$. Therefore, (i) and (ii) hold for all $k \ge k_1$.

Suppose $\beta(z^*) \ne \emptyset$. If $\beta(z^*) = \{1, 2, \cdots, m\}$, then we have $\beta_{j+1}^k \subseteq \beta(z^*)$ for any k. Next we suppose that $\beta(z^*)$ is a proper subset of $\{1, 2, \cdots, m\}$. We will show that, when k is large enough, if there is an $\hat{i} \in \alpha_j^k \cup \gamma_j^k$ satisfying (5.35) and β_j^k is a proper subset of $\beta(z^*)$, then $\hat{i} \in \beta(z^*)$. Note that

$$\lim_{k\to\infty} \max\left\{G_i(z^k), H_i(z^k)\right\} = G_i(z^*) + H_i(z^*) > 0, \qquad i \notin \beta(z^*), \tag{5.51}$$

$$\lim_{k\to\infty} \max\left\{G_i(z^k), H_i(z^k)\right\} = 0, \qquad i \in \beta(z^*). \tag{5.52}$$

We have from (5.51) and (5.52) that there exists an integer $k_0 \ge k_1$ such that, for any $k \ge k_0$,

$$\max_{i\in\beta(z^*)} \left\{ \max\left\{G_i(z^k), H_i(z^k)\right\}\right\} < \min_{i\notin\beta(z^*)} \left\{ \max\left\{G_i(z^k), H_i(z^k)\right\}\right\}. \tag{5.53}$$

This inequality means that the index $\hat{i}$ satisfying (5.35) must be in $\beta(z^*)$ as long as β_j^k is a proper subset of $\beta(z^*)$, namely,

$$\left(\alpha_j^k \cup \gamma_j^k\right) \cap \beta(z^*) = \beta(z^*) \setminus \beta_j^k \ne \emptyset.$$

By (5.38), we have $\beta_{j+1}^k \subseteq \beta(z^*)$ and therefore, (5.41)–(5.43) hold for all $k \ge k_0$. Note that, since $k_0 \ge k_1$, (i) and (ii) also hold for all $k \ge k_0$. This completes the proof. $\square$

We proceed to analyzing convergence properties of Algorithm HIA in detail. First, we make the following assumption:

A1: Even if the identical subproblem appears in Step 2 at infinitely many iterations and this problem may have an infinite number of solutions, we always obtain the same solution, or at most finitely many different solutions.

This assumption seems reasonable in practice, since an iterative method applied to solve a subproblem will generate an identical sequence as long as the same starting point is chosen.

In Algorithm HIA, for each k, the parameter θ_k is expectantly used as a positive lower bound of

$$\min_{i \notin \beta(z^*)} \max \Big\{ G_i(z^*), H_i(z^*) \Big\} = \min_{i \notin \beta(z^*)} \Big(G_i(z^*) + H_i(z^*) \Big) > 0. \tag{5.54}$$

In fact, a key technique for obtaining a finite termination property of Algorithm HIA is to choose the index sets α_j^k, β_j^k and γ_j^k so that

$$\alpha_{j'_k}^k = \alpha(z^*), \ \beta_{j'_k}^k = \beta(z^*), \ \gamma_{j'_k}^k = \gamma(z^*) \tag{5.55}$$

hold for some index $j'_k \in \{0, 1, \cdots, j_k\}$ when k is large enough. In order to ensure this, the number θ_k needs to be small enough to exclude all the indices in $\alpha(z^*) \cup \gamma(z^*)$ from $\beta_{j'_k}^k$ for all k sufficiently large. Another requirement is that all the indices in $\beta(z^*)$ remain in $\beta_{j'_k}^k$ when k is sufficiently large.

Since the index set $\{1, 2, \cdots, m\}$ has a finite number of partitions, there are a finite number of subproblems (5.86). By Assumption A1, the set

$$S := \Big\{ \hat{z}_j^k \mid \quad 0 \le j \le j_k, \ k = 0, 1, \cdots \Big\} \tag{5.56}$$

is a finite set. Recall that $\hat{z}_j^k \in \mathcal{F}$ for any k and j. We consider the following two cases.

Case I: $\bigcup_{\hat{z}_j^k \in S} \Big(\alpha(\hat{z}_j^k) \cup \gamma(\hat{z}_j^k) \Big) \ne \emptyset$. In this case, we have

$$\min_{\hat{z}_j^k \in S} \Big\{ G_i(\hat{z}_j^k) + H_i(\hat{z}_j^k) \mid \quad i \in \alpha(\hat{z}_j^k) \cup \gamma(\hat{z}_j^k) \Big\} > 0,$$

since S is a finite set. It then follows from (5.36) and the way of updating δ_k in Step 3 that the parameter δ_k stays at a positive constant when k is sufficiently large. So, by the updating rule of θ_k, there exists an integer $\bar{k} > 0$ such that

$$\theta_k = \theta_{\bar{k}} > 0, \quad k \ge \bar{k}. \tag{5.57}$$

Since

$$\lim_{k \to \infty} \max \Big\{ G_i(z^k), H_i(z^k) \Big\} = 0, \quad \forall i \in \beta(z^*),$$

we have

$$\max\left\{G_i(z^k), H_i(z^k)\right\} < \theta_{\bar{k}}, \quad \forall i \in \beta(z^*) \tag{5.58}$$

for all sufficiently large k. Moreover, by the definition of $\beta(z^*)$, we have

$$\max_{i\in\beta(z^*)}\left\{\max\left\{G_i(z^k), H_i(z^k)\right\}\right\} < \min_{i\notin\beta(z^*)}\left\{\max\left\{G_i(z^k), H_i(z^k)\right\}\right\} \tag{5.59}$$

for any k large enough. Taking into account (5.35), we deduce from (5.58) and (5.59) that the indices in $\beta(z^*)$ are inevitably included by some β_j^k and, in Step 3, these indices must be chosen earlier than the indices in $\alpha(z^*) \cup \gamma(z^*)$. As a result, there is some $j \in \{0, 1, \cdots, j_k\}$ such that $\beta(z^*) \subseteq \beta_j^k$ whenever k is sufficiently large. This together with Theorem 5.6 means that, when k is large sufficiently, there must be some index $j'_k \in \{0, 1, \cdots, j_k\}$ satisfying (5.55), and then problem (5.86) with $j = j'_k$ is actually equivalent to problem (5.24). As long as a solution of problem (5.86) with $j = j'_k$ yields a B-stationary point of problem (5.1), Algorithm HIA may terminate in a finite number of iterations. Of course, it may happen that the algorithm stops by getting another B-stationary point of (5.1) before we identify the correct index sets.

Furthermore, we make the following assumption, which is most likely to hold when problem (5.1) has finitely many B-stationary points:

A2: The limit point z^* of $\{z^k\}$, which is a sequence of stationary points of (P_{ϵ_k}), is a B-stationary point of problem (5.1) and it belongs to the set S given by (5.56).

Then, it follows from (5.36) and the updating rule of θ_k that the number $\theta_{\bar{k}}$ in (5.57) is actually a positive lower bound of (5.54). Thus, our requirements are fulfilled: That is, in Step 3,

(a) all the indices in $\alpha(z^*) \cup \gamma(z^*)$ are excluded from $\beta_{j_k}^k$ eventually;

(b) all the indices in $\beta(z^*)$ remain in $\beta_{j_k}^k$ eventually.

Therefore, we are able to identify the correct index sets $\alpha(z^*), \beta(z^*)$, and $\gamma(z^*)$ in a finite number of iterations and furthermore, we may terminate the algorithm finitely by getting z^*. Note that Algorithm HIA may stop prematurely by producing another B-stationary point of problem (5.1).

Case II: $\alpha(\hat{z}_j^k) \cup \gamma(\hat{z}_j^k) = \emptyset, \ \forall \hat{z}_j^k \in S$. In this case, we have from the updating rules of δ_k and θ_k that δ_k remains to be $+\infty$ and so

$$\theta_k \equiv \theta_0, \qquad \forall k.$$

Since the strategy in Algorithm HIA is to add some indices to β_j^k one after another, by Theorem 5.6, we have the same conclusion as in Case I: When k is sufficiently large, there exists some index $j'_k \in \{0, 1, \cdots, j_k\}$ satisfying (5.55). Furthermore, suppose that Assumption A2 holds. In the present case, this means

$$\beta(z^*) = \{1, 2, \cdots, m\}.$$

Recall that $\theta_k \equiv \theta_0$ for all k. As a result, all indices should be chosen to be in $\beta_{j_k}^k$ eventually when k becomes large enough, i.e., we also can identify the index sets $\alpha(z^*), \beta(z^*)$, and $\gamma(z^*)$ in a finite number of steps.

The preceding analysis together with Lemma 5.1 yields the following concluding result.

Theorem 5.7 *Suppose that the sequence $\{z^k\}$ generated by Algorithm HIA converges to z^* as $\epsilon_k \to 0$ and the MPCC-LICQ holds at z^*. Then, under Assumption A1, we have that, for any sufficiently large k,*

(i) *problem (P_{ϵ_k}) satisfies the standard LICQ at z^k;*

(ii) *there exists $j'_k \in \{0, 1, \cdots, j_k\}$ such that $\alpha_{j'_k}^k, \beta_{j'_k}^k$, and $\gamma_{j'_k}^k$ satisfy condition (5.55).*

If furthermore, Assumption A2 also holds, then

$$\alpha_{j_k}^k = \alpha(z^*), \; \beta_{j_k}^k = \beta(z^*), \; \gamma_{j_k}^k = \gamma(z^*) \tag{5.60}$$

hold for all k sufficiently large.

In consequence, without the assumption of asymptotically weak nondegeneracy, we have attained the same target as Algorithm H, which is to identify the index sets $\alpha(z^*), \beta(z^*)$, and $\gamma(z^*)$ finitely. Thus, Algorithm HIA may terminate in a finite number of iterations by producing a B-stationary point of problem (5.1).

On the other hand, if the sequence $\{z^k\}$ is asymptotically weakly nondegenerate, the sets $\alpha_0^k, \beta_0^k, \gamma_0^k$ given in Step 1 of Algorithm HIA are the same as the sets $\alpha^k, \beta^k, \gamma^k$ given in Step 1 of Algorithm H when k is large enough. Theorem 5.5(ii) immediately yields the following result.

Theorem 5.8 *Suppose the sequence $\{z^k\}$ generated by Algorithm HIA converges to z^* as $\epsilon_k \to 0$, the MPCC-LICQ holds at z^*, and $\{z^k\}$ is asymptotically weakly non-degenerate. Then the sets α_0^k, β_0^k, and γ_0^k satisfy (5.55) with $j'_k = 0$ when k is large enough.*

The main strategy in Algorithm HIA is to add some indices, which are chosen from $\alpha_j^k \cup \gamma_j^k$, to β_j^k. Since $\alpha_j^k \cup \gamma_j^k$ contains $\beta(z^*)$ for any k large enough, condition (5.55) holds for some $j'_k \in \{0, 1, \cdots, j_k\}$ when k is large enough. In order to ensure this, the inclusions in (5.50) are necessary. In the above discussion, we suppose that $\tau > 0$ and $\sigma \in (0, 1)$, just as in Algorithm H. Actually, from the proof of Theorem 5.6, (5.50) remains true for the case where $\tau \geq 2$ and $\sigma = 1$ and furthermore, so do Theorems 5.6 and 5.7. Moreover, since the functions G and H play a symmetric role in problem (5.1), we may exchange the definitions of α_0^k and γ_0^k in Step 1 of Algorithm HIA, namely, let

$$\begin{aligned} \gamma_0^k &:= \Big\{ i \mid \quad G_i(z^k) \leq \rho(z^k), \quad H_i(z^k) > \rho(z^k) \Big\}, \\ \alpha_0^k &:= \Big\{ 1, 2, \cdots, m \Big\} \setminus \Big(\beta_0^k \cup \gamma_0^k \Big) \end{aligned}$$

instead of (5.32) and (5.30), respectively.

5.4 Modified Hybrid Method with Index Subtraction Strategy

In this section, we consider another hybrid algorithm that adopts an index subtraction strategy. One advantage of this algorithm is that the function ρ employed in the last section can be replaced by a sequence of positive numbers.

Algorithm HIS:

Step 0: Choose $\eta > 0$, $\theta_0 > 0$, $\xi_0 > 0$, and $\epsilon_0 > 0$. Set $k := 0$.

Step 1: Solve problem (P_{ϵ_k}) and let z^k denote one of its stationary points. Set

$$\alpha_0^k := \Big\{ i \mid \quad G_i(z^k) > \eta, \quad H_i(z^k) \leq \xi_k \Big\}, \tag{5.61}$$

$$\gamma_0^k := \Big\{ i \mid \quad G_i(z^k) \leq \xi_k, \quad H_i(z^k) > \eta \Big\} \setminus \alpha_0^k, \tag{5.62}$$

$$\beta_0^k := \Big\{ 1, 2, \cdots, m \Big\} \setminus \Big(\alpha_0^k \cup \gamma_0^k \Big), \tag{5.63}$$

$$\delta_0^k := +\infty,$$

and $j := 0$. Go to Step 2.

Step 2: If problem (5.86) is solvable, let $\hat{z}_j^k$ denote one of its stationary points and go to Step 3. Otherwise, go to Step 4.

Step 3: If, in (5.86), the Lagrange multipliers corresponding to the constraints (5.34) are all nonnegative, then terminate. Else, if there is an $\hat{i} \in \beta_j^k$ such that

$$\max_{i \in \beta_j^k} \Big\{ \max \Big\{ G_i(z^k), H_i(z^k) \Big\} \Big\} = \max \Big\{ G_{\hat{i}}(z^k), H_{\hat{i}}(z^k) \Big\} > \theta_k, \tag{5.64}$$

then set

$$\delta_{j+1}^k := \min\left\{\delta_j^k,\ \frac{1}{2}\min\left\{G_i(\hat{z}_j^k) + H_i(\hat{z}_j^k) \mid \quad i \in \alpha(\hat{z}_j^k) \cup \gamma(\hat{z}_j^k)\right\}\right\} \tag{5.65}$$

and

$$\begin{aligned}
\alpha_{j+1}^k &:= \begin{cases} \alpha_j^k \cup \{\hat{i}\}, & \text{if } G_{\hat{i}}(z^k) \geq H_{\hat{i}}(z^k) \\ \alpha_j^k, & \text{otherwise,} \end{cases} \\
\beta_{j+1}^k &:= \beta_j^k \setminus \{\hat{i}\}, \\
\gamma_{j+1}^k &:= \begin{cases} \gamma_j^k, & \text{if } G_{\hat{i}}(z^k) \geq H_{\hat{i}}(z^k) \\ \gamma_j^k \cup \{\hat{i}\}, & \text{otherwise,} \end{cases} \\
j &:= j+1
\end{aligned}$$

and go to Step 2. Otherwise, let $j_k := j$ and $\delta_k := \delta_j^k$ and go to Step 4.

Step 4: Choose $\epsilon_{k+1} \in (0, \epsilon_k)$, $\xi_{k+1} \in (0, \xi_k]$, and set $\theta_{k+1} := \min\left\{\theta_k,\ \delta_k\right\}$. Let $k := k+1$ and go to Step 1.

Note that, since $G_i(z^k) \to 0$ for each $i \in \beta(z^*) \cup \gamma(z^*)$, the index set α_0^k determined in Step 1 will eventually consist of indices in $\alpha(z^*)$ only. Similarly, γ_0^k will eventually consist of indices in $\gamma(z^*)$ only. Therefore, we must have

$$\alpha_0^k \subseteq \alpha(z^*),\ \beta_0^k \supseteq \beta(z^*),\ \gamma_0^k \subseteq \gamma(z^*) \tag{5.66}$$

for all k sufficiently large. This is the key requirement for Algorithm HIS, which subtracts some indices from β_j^k so that condition (5.55) holds for some $j_k' \in \{0, 1, \cdots, j_k\}$ when k is large enough.

In Algorithm HIS, the strict positivity of η is essential, and the sequence $\{\xi_k\}$ plays only a subsidiary role: The sets α_0^k and γ_0^k should be chosen not too large so that the key condition (5.66) holds as early as possible. To this end, we may choose $\{\xi_k\}$ to be a null sequence. On the other hand, from the computational viewpoint, it is desirable that the set β_0^k is as small as possible, i.e., the sets α_0^k and γ_0^k are as large as possible. In consequence, it would be important to choose the constant $\eta > 0$ and the sequence $\{\xi_k\}$ appropriately.

Another practical choice is simply to remove $\{\xi_k\}$ from Algorithm HIS. For example, we may define

$$\begin{aligned}
\alpha_0^k &:= \left\{\, i \mid \quad G_i(z^k) > \eta\right\}, \\
\gamma_0^k &:= \left\{\, i \mid \quad H_i(z^k) > \eta\right\} \setminus \alpha_0^k, \\
\beta_0^k &:= \left\{\, 1, 2, \cdots, m\right\} \setminus \left(\alpha_0^k \cup \gamma_0^k\right)
\end{aligned}$$

instead of (5.61)–(5.63), or preferably, let

$$\alpha_0^k := \Big\{ i \mid \ G_i(z^k) > \eta,\ H_i(z^k) \le \eta \Big\},$$
$$\gamma_0^k := \Big\{ i \mid \ G_i(z^k) \le \eta,\ H_i(z^k) > \eta \Big\},$$
$$\beta_0^k := \Big\{ 1, 2, \cdots, m \Big\} \setminus \Big(\alpha_0^k \cup \gamma_0^k \Big),$$

which is equivalent to letting $\xi_k \equiv \eta$ $(\forall k)$ in (5.61)–(5.63).

Moreover, the parameter θ_k is also expected to be a positive lower bound of (5.54) so that the indices outside $\beta(z^*)$ can be removed from $\beta_{j_k}^k$ eventually in Step 3. As analyzed in the last subsection, Assumptions A1 and A2 guarantee that the parameter θ_k satisfies this requirement, i.e., θ_k can serve as a positive lower bound of (5.54) when k is large enough.

In a similar way to the last section, comprehensive and detailed analysis can be given for Algorithm HIS. In particular, under Assumption A1, θ_k stays at a positive constant when k is sufficiently large in both Cases I and II considered in the last subsection. It then follows from (5.64) that all the indices in $\beta(z^*)$ cannot be selected in Step 3 for all k large enough, i.e., they remain in $\beta_{j_k}^k$ eventually. Thus, we have the following result.

Theorem 5.9 *Suppose that the sequence $\{z^k\}$ generated by Algorithm HIS converges to z^* as $\epsilon_k \to 0$ and Assumption A1 holds. Then, for any sufficiently large k, we have*

$$\alpha_0^k \subseteq \alpha_1^k \subseteq \cdots \subseteq \alpha_{j_k}^k \subseteq \alpha(z^*),$$
$$\beta_0^k \supseteq \beta_1^k \supseteq \cdots \supseteq \beta_{j_k}^k \supseteq \beta(z^*),$$
$$\gamma_0^k \subseteq \gamma_1^k \subseteq \cdots \subseteq \gamma_{j_k}^k \subseteq \gamma(z^*).$$

If furthermore, Assumption A2 holds, then we have (5.60) for all k sufficiently large.

This theorem indicates that, under similar assumptions to the previous subsection, Algorithm HIS also has a finite termination property.

5.5 Further Discussions

We have mainly considered B-stationarity for MPCC in the previous sections. In this section, we make some remarks on the algorithms, including the employed assumptions, the stopping criteria, and some extensions.

5.5.1 Remarks on the assumptions

We first investigate the connection between the ULSC condition and the asymptotically weakly nondegeneracy condition [31]. Our result can be stated as follows.

Theorem 5.10 *Let z^k be stationary to problem (P_{ϵ_k}) for each k and the sequence $\{z^k\}$ converge to z^* as $\epsilon_k \to 0$. Suppose that the MPCC-LICQ holds at z^*. If the ULSC condition holds at z^*, then $\{z^k\}$ is asymptotically weakly nondegenerate.*

Proof: The theorem trivially holds when $\mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*) = \emptyset$. So we assume $\mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*) \neq \emptyset$. First, we show that z^* is a weakly stationary point, i.e., there exist multiplier vectors $\bar{\lambda}_g, \bar{\lambda}_h, \bar{\lambda}_G$, and $\bar{\lambda}_H$ satisfying

$$\nabla f(z^*) + \nabla g(z^*)\bar{\lambda}_g + \nabla h(z^*)\bar{\lambda}_h - \nabla G(z^*)\bar{\lambda}_G - \nabla H(z^*)\bar{\lambda}_H = 0, \tag{5.67}$$

$$\bar{\lambda}_g \geq 0, \quad \bar{\lambda}_g^T g(z^*) = 0, \tag{5.68}$$

$$\bar{\lambda}_{G,i} = 0, \quad i \notin \mathcal{I}_G(z^*), \tag{5.69}$$

$$\bar{\lambda}_{H,i} = 0, \quad i \notin \mathcal{I}_H(z^*). \tag{5.70}$$

Note that the ULSC condition means

$$\bar{\lambda}_{G,i}\bar{\lambda}_{H,i} \neq 0, \quad i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*). \tag{5.71}$$

By the continuity of the functions involved and the assumptions that $\{z^k\}$ converges to z^* and the MPCC-LICQ holds at z^*, we have that, for all k large enough,

$$\mathcal{I}_G(z^k) \subseteq \mathcal{I}_G(z^*), \quad \mathcal{I}_H(z^k) \subseteq \mathcal{I}_H(z^*), \quad \mathcal{I}_g(z^k) \subseteq \mathcal{I}_g(z^*) \tag{5.72}$$

and the set

$$\begin{aligned}
&\Big\{\nabla g_l(z^k), \nabla h_r(z^k): \quad l \in \mathcal{I}_g(z^*),\ r = 1, \cdots, q\Big\} \\
\bigcup\ &\Big\{\nabla G_i(z^k), \nabla H_i(z^k): \quad i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*)\Big\} \\
\bigcup\ &\Big\{\nabla G_i(z^k) + \frac{G_i(z^k)}{H_i(z^k)}\nabla H_i(z^k): \quad i \in \mathcal{I}_G(z^*) \setminus \mathcal{I}_H(z^*)\Big\} \\
\bigcup\ &\Big\{\nabla H_i(z^k) + \frac{H_i(z^k)}{G_i(z^k)}\nabla G_i(z^k): \quad i \in \mathcal{I}_H(z^*) \setminus \mathcal{I}_G(z^*)\Big\}
\end{aligned} \tag{5.73}$$

is linearly independent. It follows from the stationarity of z^k that there exist Lagrange multiplier vectors $\lambda_g^k \in \Re^p, \lambda_h^k \in \Re^q$, and $\lambda_\Phi^k \in \Re^m$ such that

$$\nabla f(z^k) + \nabla g(z^k)\lambda_g^k + \nabla h(z^k)\lambda_h^k - \nabla\Phi_{\epsilon_k}(z^k)\lambda_\Phi^k = 0, \tag{5.74}$$

$$\lambda_g^k \geq 0, \quad g(z^k)^T\lambda_g^k = 0. \tag{5.75}$$

Note that

$$\nabla\Phi_{\epsilon_k,i}(z^k) = \frac{H_i(z^k)}{G_i(z^k)+H_i(z^k)}\nabla G_i(z^k) + \frac{G_i(z^k)}{G_i(z^k)+H_i(z^k)}\nabla H_i(z^k) \tag{5.76}$$

holds for each i and k [31]. Moreover, by (5.72) and (5.75), we have $\lambda_{g,l}^k = 0$ for every $l \notin \mathcal{I}_g(z^*)$. Therefore, the equation (5.74) becomes

$$\begin{aligned} 0 \;=\; & \nabla f(z^k) + \sum_{l\in\mathcal{I}_g(z^*)} \lambda_{g,l}^k \nabla g_l(z^k) + \nabla h(z^k)\lambda_h^k \\ & - \sum_{i\in\mathcal{I}_G(z^*)\cap\mathcal{I}_H(z^*)} \left(\frac{\lambda_{\Phi,i}^k H_i(z^k)}{G_i(z^k)+H_i(z^k)}\nabla G_i(z^k) + \frac{\lambda_{\Phi,i}^k G_i(z^k)}{G_i(z^k)+H_i(z^k)}\nabla H_i(z^k)\right) \\ & - \sum_{i\in\mathcal{I}_G(z^*)\setminus\mathcal{I}_H(z^*)} \frac{\lambda_{\Phi,i}^k H_i(z^k)}{G_i(z^k)+H_i(z^k)} \left(\nabla G_i(z^k) + \frac{G_i(z^k)}{H_i(z^k)}\nabla H_i(z^k)\right) \\ & - \sum_{i\in\mathcal{I}_H(z^*)\setminus\mathcal{I}_G(z^*)} \frac{\lambda_{\Phi,i}^k G_i(z^k)}{G_i(z^k)+H_i(z^k)} \left(\nabla H_i(z^k) + \frac{H_i(z^k)}{G_i(z^k)}\nabla G_i(z^k)\right). \end{aligned} \tag{5.77}$$

By the linear independence of the set (5.73), we have that all the multiplier sequences

$$\left\{\lambda_{g,l}^k : \; l\in\mathcal{I}_g(z^*)\right\}, \quad \left\{\lambda_{h,r}^k : \; r = 1,2,\cdots,q\right\},$$
$$\left\{\frac{\lambda_{\Phi,i}^k H_i(z^k)}{G_i(z^k)+H_i(z^k)} : \; i\in\mathcal{I}_G(z^*)\right\}, \quad \left\{\frac{\lambda_{\Phi,j}^k G_j(z^k)}{G_j(z^k)+H_j(z^k)} : \; j\in\mathcal{I}_H(z^*)\right\}$$

are convergent. Define $\bar\lambda_g, \bar\lambda_h, \bar\lambda_G$, and $\bar\lambda_H$ as follows:

$$\bar\lambda_{g,l} = \begin{cases} \lim_{k\to\infty} \lambda_{g,l}^k & , \quad l\in\mathcal{I}_g(z^*) \\ 0 & , \quad l\notin\mathcal{I}_g(z^*) \end{cases}; \tag{5.78}$$

$$\bar\lambda_{h,r} = \lim_{k\to\infty} \lambda_{h,r}^k, \quad r = 1,2,\cdots,q; \tag{5.79}$$

$$\bar\lambda_{G,i} = \begin{cases} \lim_{k\to\infty} \frac{\lambda_{\Phi,i}^k H_i(z^k)}{G_i(z^k)+H_i(z^k)} & , \quad i\in\mathcal{I}_G(z^*) \\ 0 & , \quad i\notin\mathcal{I}_G(z^*) \end{cases}; \tag{5.80}$$

$$\bar\lambda_{H,j} = \begin{cases} \lim_{k\to\infty} \frac{\lambda_{\Phi,j}^k G_j(z^k)}{G_j(z^k)+H_j(z^k)} & , \quad j\in\mathcal{I}_H(z^*) \\ 0 & , \quad j\notin\mathcal{I}_H(z^*) \end{cases}. \tag{5.81}$$

Letting $k\to\infty$ in (5.77) and taking into account (5.75) and (5.78)–(5.81), we have that (5.67)–(5.70) hold. Next we suppose that the ULSC holds at z^*. It then follows from (5.71) and (5.80)–(5.81) that, for each $i\in\mathcal{I}_G(z^*)\cap\mathcal{I}_H(z^*)$,

$$\bar\lambda_{G,i} = \lim_{k\to\infty} \frac{\lambda_{\Phi,i}^k H_i(z^k)}{G_i(z^k)+H_i(z^k)} \neq 0, \tag{5.82}$$

$$\bar\lambda_{H,i} = \lim_{k\to\infty} \frac{\lambda_{\Phi,i}^k G_i(z^k)}{G_i(z^k)+H_i(z^k)} \neq 0. \tag{5.83}$$

Since both $G_i(z^k)$ and $H_i(z^k)$ are positive for each i and k, it follows from (5.82) and (5.83) that

$$0 < |\bar{\lambda}_{G,i}| \leq \liminf_{k\to\infty} \; |\lambda^k_{\Phi,i}|,$$
$$\lim_{k\to\infty} \lambda^k_{\Phi,i} = \bar{\lambda}_{G,i} + \bar{\lambda}_{H,i}$$

for each $i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*)$. Therefore, $\left\{\lambda^k_{\Phi,i}\right\}$ is convergent to a nonzero number for each $i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*)$. We then have from (5.82) and (5.83) that, for each $i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*)$,

$$\lim_{k\to\infty} \frac{H_i(z^k)}{G_i(z^k) + H_i(z^k)} = \frac{\bar{\lambda}_{G,i}}{\bar{\lambda}_{G,i} + \bar{\lambda}_{H,i}} \neq 0,$$
$$\lim_{k\to\infty} \frac{G_i(z^k)}{G_i(z^k) + H_i(z^k)} = \frac{\bar{\lambda}_{H,i}}{\bar{\lambda}_{G,i} + \bar{\lambda}_{H,i}} \neq 0.$$

This, together with (5.76), implies that the sequence $\{z^k\}$ is asymptotically weakly nondegenerate. This completes the proof of the theorem. □

Let us make a remark on Assumption A2 employed by Algorithms HIA and HIS. Recall that we assume the sequence $\{z^k\}$ converges to z^*. Actually, the sequence $\{z^k\}$ may have multiple limit points in general. In this case, it is easy to see that, as long as one of the limit points satisfies the assumptions made for z^* in Algorithms HIA and HIS, we may obtain similar conclusions. Thus, Assumption A2 can be restated as follows :

A2 ′ : The set S (defined in Section 5.3) contains an accumulation point of the sequence $\{z^k\}$ that is a B-stationary point of problem (5.1).

This indicates that our assumptions for Algorithms HIA and HIS are really not very restrictive.

5.5.2 Stopping criteria

The stopping criterion in Step 3 of Algorithms H, HIA, and HIS is used to check the B-stationarity of the point $\hat{z}^k$, which is based on the fact that, for given $(\alpha^k, \beta^k, \gamma^k)$, if $\hat{z}^k$ is a stationary point of the problem

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & G_i(z) \geq 0, \;\; H_i(z) = 0, \quad i \in \alpha^k, \end{array}$$

$$
\begin{aligned}
&G_i(z) = 0,\ H_i(z) = 0, \quad i \in \beta^k, \\
&G_i(z) = 0,\ H_i(z) \geq 0, \quad i \in \gamma^k, \\
&g(z) \leq 0,\ h(z) = 0
\end{aligned}
\tag{5.84}
$$

with nonnegative Lagrange multipliers related to the constraints

$$
\begin{aligned}
&G_i(z) \geq 0,\ H_i(z) = 0, \quad i \in \alpha^k \cap \beta(\hat{z}^k), \\
&G_i(z) = 0,\ H_i(z) = 0, \quad i \in \beta^k, \\
&G_i(z) = 0,\ H_i(z) \geq 0, \quad i \in \gamma^k \cap \beta(\hat{z}^k)
\end{aligned}
\tag{5.85}
$$

and the MPCC-LICQ holds at $\hat{z}^k$, then $\hat{z}^k$ is a B-stationary point of problem (5.1) [33]. From the meaning of B-stationarity, we see that B-stationary points are obviously good candidates for local minimizers of problem (5.1). However, it may be difficult to obtain such a point for MPCCs in general. Actually, we may use some other conditions to check M-stationarity or C-stationarity in view of the fact that, if the MPCC-LICQ holds at $\hat{z}^k$ and the Lagrange multipliers $\lambda_{G,i}^k$ and $\lambda_{H,i}^k$ corresponding to (5.85) satisfy some suitable conditions, then $\hat{z}^k$ is C- or M-stationary to problem (5.1). Hence, the algorithms may terminate finitely by producing a C- or M-stationary point of problem (5.1) under some weaker conditions.

5.5.3 Comparison of the algorithms

Comparing with Algorithm H, Algorithms HIA and HIS may need to solve more subproblems

$$
\begin{aligned}
\text{minimize} \quad & f(z) \\
\text{subject to} \quad & G_i(z) \geq 0,\ H_i(z) = 0, \quad i \in \alpha_j^k, \\
& G_i(z) = 0,\ H_i(z) = 0, \quad i \in \beta_j^k, \\
& G_i(z) = 0,\ H_i(z) \geq 0, \quad i \in \gamma_j^k, \\
& g(z) \leq 0,\ h(z) = 0.
\end{aligned}
\tag{5.86}
$$

On the other hand, Algorithm H may have to solve more subproblems (P_ϵ) than the other two algorithms in general. From both theoretical and computational points of view, problem (5.86) is easier to deal with than problem (P_ϵ). For example, under the condition that the functions G, H, h are all affine and each g_i is convex, the feasible region of problem (5.86) is convex, but that of problem (P_ϵ) is not convex.

5.5.4 Extensions

We have presented Algorithms H, HIA and HIS by applying an active set identification technique to the smoothing continuation method. Actually, the proposed approaches may be extended by using other subproblems instead of (P_ϵ) in Step 1 of the algorithms.

(I) Since problem (P_ϵ) is equivalent to

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & g(z) \leq 0,\ h(z) = 0 \\ & G(z) + H(z) \geq 0 \\ & G_i(z)H_i(z) = \epsilon^2,\ i = 1, \cdots, m, \end{array} \tag{5.87}$$

we may use problem (5.87) instead of (P_ϵ) in Step 1 of the algorithms at each iteration. It is obvious that all analysis and conclusions remain valid. Note that, for any $\epsilon > 0$, the constraints

$$G_i(z) + H_i(z) \geq 0, \quad i = 1, 2, \cdots, m$$

are always inactive and so problem (5.87) seems simpler than (P_ϵ).

(II) The regularization scheme

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & g(z) \leq 0,\ h(z) = 0 \\ & G(z) \geq 0,\ H(z) \geq 0 \\ & G_i(z)H_i(z) \leq \epsilon,\ i = 1, \cdots, m \end{array} \tag{5.88}$$

and the penalty scheme

$$\begin{array}{ll} \text{minimize} & f(z) + \epsilon^{-1}G(z)^T H(z) \\ \text{subject to} & g(z) \leq 0,\ h(z) = 0 \\ & G(z) \geq 0,\ H(z) \geq 0, \end{array} \tag{5.89}$$

where ϵ is a positive parameter, have been proposed as approximate problems of (5.1) in [76] and [38], respectively. These two methods share similar properties to the smoothing continuation method. We may replace (P_ϵ) by (5.88) or (5.89) in Step 1 of the algorithms at each iteration and we can obtain similar results.

5.6 Computational Results

We have tested the proposed algorithms on various instances of MPCCs. In our experiments, we employed the MATLAB 6.0 built-in solver function *fmincon* to solve the subproblems at each iteration. The computational results indicate that the proposed approach can find an optimal solution of an MPCC in a small number of iterations. We report the details below.

Table 5.1: Computational results for Problem 5.1

Smoothing Continuation Method	z	$\epsilon_0 = 10^{-2}$	(0.5211,0.5211,0.5232,0.5232,0.0044,0.0044)
		$\epsilon_1 = 10^{-4}$	(0.5011,0.5011,0.5011,0.5011,0.0000,0.0000)
		$\epsilon_2 = 10^{-6}$	(0.5000,0.5000,0.5000,0.5000,-0.0000,-0.0000)
	Ite		27
Algorithm H	$\hat{z}$	$\epsilon_0 = 10^{-2}$	(0.5000,0.5000,0.5000,0.5000,0,0)
	$\beta(\hat{z})$	$\epsilon_0 = 10^{-2}$	{1, 2}
	Ite		15
Regularization Method	z	$\epsilon_0 = 10^{-2}$	(0.5000,0.5000,0.5000,0.5000,0.0000,0)
	Ite		5
Algorithm H	$\hat{z}$	$\epsilon_0 = 10^{-2}$	(0.5000,0.5000,0.5000,0.5000,0,0)
	$\beta(\hat{z})$	$\epsilon_0 = 10^{-2}$	{1, 2}
	Ite		6
Penalty Method	z	$\epsilon_0 = 10^{-2}$	(0.5000,0.5000,0.5000,0.5000,0.0000,-0.0000)
	Ite		6
Algorithm H	$\hat{z}$	$\epsilon_0 = 10^{-2}$	(0.5000,0.5000,0.5000,0.5000,0,0)
	$\beta(\hat{z})$	$\epsilon_0 = 10^{-2}$	{1, 2}
	Ite		7

5.6.1 Computational results for Algorithm H

In order to get a comprehensive computational experience, we have investigated Algorithm H incorporating various methods mentioned in the last section. In our testing, we set $\epsilon_0 = 10^{-2}$ and updated this parameter by $\epsilon_{k+1} = 10^{-2}\epsilon_k$. The point z^k is used as the starting point for the next step in Algorithm H and the related methods. We employed

$$\rho_1(z) = ||\min(G(z), H(z))||^{\frac{1}{2}}$$

Table 5.2: Computational results for Problem 5.2

Smoothing Continuation Method	$(z_1, \cdots, z_5)$	$\epsilon_0 = 10^{-2}$	(4.6978,3.9999,4.1277,0.0000,0.0000)
		$\epsilon_1 = 10^{-4}$	(4.8294,4.0000,4.0569,0.0000,-0.0000)
		$\epsilon_2 = 10^{-6}$	(5.0408,4.0000,2.1732,0.0000,-0.0000)
		$\epsilon_3 = 10^{-8}$	(5.0000,4.0000,2.0000,0.0000,0.0000)
	Ite		117
Algorithm H	$(\hat{z}_1, \cdots, \hat{z}_5)$	$\epsilon_0 = 10^{-2}$	(5.0000,4.0000,2.0000,0.0000,-0.0000)
	$\beta(\hat{z})$	$\epsilon_0 = 10^{-2}$	$\emptyset$
	Ite		30
Regularization Method	$(z_1, \cdots, z_5)$	$\epsilon_0 = 10^{-2}$	(5.0000,4.0000,2.0000,0.0001,0.0002)
		$\epsilon_1 = 10^{-4}$	(5.0000,4.0000,2.0000,-0.0000,-0.0000)
	Ite		15
Algorithm H	$(\hat{z}_1, \cdots, \hat{z}_5)$	$\epsilon_0 = 10^{-2}$	(5.0000,4.0000,2.0000,-0.0000,-0.0000)
	$\beta(\hat{z})$	$\epsilon_0 = 10^{-2}$	{2, 3}
	Ite		15
Penalty Method	$(z_1, \cdots, z_5)$	$\epsilon_0 = 10^{-2}$	(5.0000,4.0000,2.0000,0.0000,-0.0000)
	Ite		5
Algorithm H	$(\hat{z}_1, \cdots, \hat{z}_5)$	$\epsilon_0 = 10^{-2}$	(5,4,2,0,0)
	$\beta(\hat{z})$	$\epsilon_0 = 10^{-2}$	{2, 3}
	Ite		6

and

$$\rho_2(z) = ||\Phi_0(z)||^{\frac{1}{2}}$$

as the identification function ρ in Step 1 of Algorithm H, and we found that the two functions yielded almost the same numerical results for all examples solved. This is not surprising because $\{\rho_1(z^k)\}$ and $\{\rho_2(z^k)\}$ tend to 0 in the same order as $z^k \to z^*$.

First, we report on numerical results for problems from an AMPL collection of MPECs called `MacMPEC` [50]. We notice that, since most problems are small-scale and, especially, the cardinalities of the lower-level degenerate index sets at the solutions are quite low, the proposed approach always solved the problems in only one iteration. Here we only show detailed results for three problems, Problem 5.1, Problem 5.2, and Problem 5.3, which are coded as `desilva.mod`, `ex.9.1.1.mod`, and `bilevel1.mod`, respectively, in `MacMPEC`. The computational results with the identification function ρ_1 are reported in Tables 5.1–5.3, where $\hat{z}$ and z denote the points obtained by the hybrid algorithms and the related methods, respectively, and *Ite* stands for the number of total iterations spent by the solver *fmincon*. In addition, $\beta(\hat{z})$ denotes the lower-level degenerate index set estimated at the point $\hat{z}$. Recall that the objective of the hybrid approach is to identify the set $\beta(z^*)$.

Table 5.3: Computational results for Problem 5.3

Smoothing Continuation Method	$(z_1,\cdots,z_5)$	$\epsilon_0=10^{-2}$	(24.9959,30.0163,4.9959,10.0041,0.0000)
		$\epsilon_1=10^{-4}$	(25.0000,30.0002,5.0000,10.0000,0.0000)
		$\epsilon_2=10^{-6}$	(25.0000,30.0000,5.0000,10.0000,0.0000)
	Ite		57
Algorithm H	$(\hat{z}_1,\cdots,\hat{z}_5)$	$\epsilon_0=10^{-2}$	(25.0000,30.0000,5.0000,10.0000,0.0000)
	$\beta(\hat{z})$	$\epsilon_0=10^{-2}$	{6}
	Ite		22
Regularization Method	$(z_1,\cdots,z_5)$	$\epsilon_0=10^{-2}$	(24.9998,29.9995,5.0002,9.9997,0.0007)
		$\epsilon_1=10^{-4}$	(25.0000,30.0000,5.0000,10.0000,0.0000)
	Ite		13
Algorithm H	$(\hat{z}_1,\cdots,\hat{z}_5)$	$\epsilon_0=10^{-2}$	(25.0000,30.0000,5.0000,10.0000,0.0000)
	$\beta(\hat{z})$	$\epsilon_0=10^{-2}$	{6}
	Ite		13
Penalty Method	$(z_1,\cdots,z_5)$	$\epsilon_0=10^{-2}$	(25.0000,30.0000,5.0000,10.0000,-0.0000)
	Ite		7
Algorithm H	$(\hat{z}_1,\cdots,\hat{z}_5)$	$\epsilon_0=10^{-2}$	(25,30,5,10,0)
	$\beta(\hat{z})$	$\epsilon_0=10^{-2}$	{6}
	Ite		8

Problem 5.1 This is Problem 5 in [24]. In `MacMPEC`, it is called `desilva.mod`.

$$\begin{aligned}
\text{minimize} \quad & z_1^2-2z_1+z_2^2-2z_2+z_3^2+z_4^2 \\
\text{subject to} \quad & 0\le z_1\le 2,\ 0\le z_2\le 2 \\
& z_3-z_1+z_3z_5-z_5=0 \\
& z_4-z_2+z_4z_6-z_6=0 \\
& G(z)\ge 0,\ H(z)\ge 0,\ G(z)^TH(z)=0,
\end{aligned}$$

where

$$G(z)=\begin{pmatrix} z_5 \\ z_6 \end{pmatrix}, \quad H(z)=\begin{pmatrix} 0.25-(z_3-1)^2 \\ 0.25-(z_4-1)^2 \end{pmatrix}.$$

For this problem, $z^*=(0.5,0.5,0.5,0.5,0,0)$ and $\beta(z^*)=\{1,2\}$. In our testing, we used $z=(1,1,\cdots,1)$ as the initial point for all methods.

Problem 5.2 This problem is equivalent to the bilevel programming problem 9.2.2 in [28]. It corresponds to `ex.9.1.1.mod` in `MacMPEC`.

$$\begin{aligned}
\text{minimize} \quad & -z_1-3z_2+2z_3 \\
\text{subject to} \quad & 0\le z_1\le 8 \\
& z_4+3z_5+z_6-z_7+z_8=1 \\
& 4z_4-2z_5-3z_6=0 \\
& G(z)\ge 0,\ H(z)\ge 0,\ G(z)^TH(z)=0,
\end{aligned}$$

Table 5.4: Problems 5.4–5.7 generated by QPECgen

Parameters in QPECgen	Input Data			
	Problem 5.4	Problem 5.5	Problem 5.6	Problem 5.7
`qpec_type`	300	300	300	300
`(n, m)`	(8, 5)	(6, 10)	(10, 10)	(6, 12)
`(l, p)`	(4, 5)	(4, 10)	(5, 10)	(4, 12)
`cond_P`	100	100	100	100
`scale_P`	100	100	100	100
`convex_f`	1	1	1	1
`symm_M`	1	1	1	1
`mono_M`	1	1	1	1
`cond_M`	200	200	200	200
`scale_M`	200	200	200	200
`second_deg`	2	6	5	6
`first_deg`	2	2	3	2
`mix_deg`	2	2	3	4
`tol_deg`	1.0e-6	1.0e-6	1.0e-6	1.0e-6
`implicit`	1	1	0	1
`rand_seed`	0	0	0	0
`output`	3	3	3	3

where

$$G(z) = \begin{pmatrix} z_4 \\ z_5 \\ z_6 \\ z_7 \\ z_8 \end{pmatrix}, \quad H(z) = \begin{pmatrix} 2z_1 - z_2 - 4z_3 + 16 \\ -8z_1 - 3z_2 + 2z_3 + 48 \\ 2z_1 - z_2 + 3z_3 - 12 \\ z_2 \\ -z_2 + 4 \end{pmatrix}.$$

For this problem, $z^* = (5, 4, 2, 0, 0, 0, 0, 1)$ and $\beta(z^*) = \{2, 3\}$. We employed $z = (0.5, \cdots, 0.5)$ as the initial point for all methods.

Problem 5.3 This is `bilevel1.mod` in `MacMPEC` and goes back to [24].

$$\begin{array}{ll} \text{minimize} & 2z_1 + 2z_2 - 3z_3 - 3z_4 - 60 \\ \text{subject to} & 0 \le z_1 \le 50, \ 0 \le z_2 \le 50 \\ & z_1 + z_2 + z_3 - 2z_4 \le 40 \\ & 2z_3 - 2z_1 - z_5 + z_6 + 2z_9 = -40 \\ & 2z_4 - 2z_2 - z_7 + z_8 + 2z_{10} = -40 \\ & G(z) \ge 0, \ H(z) \ge 0, \ G(z)^T H(z) = 0, \end{array}$$

Table 5.5: Some data obtained by QPECgen for Problems 5.4–5.7

Problem 5.4	x^*	(0.0723, 0.4345, 0.2387, -0.4003, -0.2870, -0.5857, -0.4421, 0.2286)
	y^*	(0, 0, 0, 0, 0)
	$\beta(x^*, y^*)$	{1, 2}
Problem 5.5	x^*	(0.0872, 0.2576, -0.1181, 0.2958, -0.1939, -0.0858)
	y^*	(0, 0, 0, 0, 0, 0, 0, 0.7559, 0.6660, 0.0115)
	$\beta(x^*, y^*)$	{1, 2, 3, 4, 5, 6}
Problem 5.6	x^*	(0.6369, 0.6371, 0.1739, -0.7158, -0.8703, -0.7478, -0.4383, 0.1886, 0.0741, 0.1494)
	y^*	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
	$\beta(x^*, y^*)$	{1, 2, 3, 4, 5}
Problem 5.7	x^*	(-0.1018, 0.0166, -0.3092, 0.1948, -0.4273, 0.0296)
	y^*	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.8546, 0.3146)
	$\beta(x^*, y^*)$	{1, 2, 3, 4, 5, 6}

where

$$G(z) = \begin{pmatrix} z_5 \\ z_6 \\ z_7 \\ z_8 \\ z_9 \\ z_{10} \end{pmatrix}, \quad H(z) = \begin{pmatrix} z_3 + 10 \\ -z_3 + 20 \\ z_4 + 10 \\ -z_4 + 20 \\ z_1 - 2z_3 - 10 \\ z_2 - 2z_4 - 10 \end{pmatrix}.$$

We have $z^* = (25, 30, 5, 10, 0, 0, 0, 0, 0, 0)$ and $\beta(z^*) = \{6\}$ for this problem and we used $z = (10, \cdots, 10)$ as the initial point for all methods.

Tables 5.1–5.3 show that the proposed methods were able to find the solutions of Problems 5.1, 5.2, and 5.3 very quickly. This, as we mentioned above, may be due to the small scale of the problems and the low cardinality of the lower-level degenerate index sets at the solutions.

Next we report on numerical results for somewhat larger test problems generated by QPECgen of Jiang and Ralph [43]. The QPECgen generator is a MATLAB program that uses a set of parameters, see Table 5.4 or [43] for detail. Once these parameters are specified, the program can randomly generate a quadratic program with linear complementarity constraints

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}(x^T, y^T)P\begin{pmatrix} x \\ y \end{pmatrix} + c^T x + d^T y \\ \text{subject to} \quad & A\begin{pmatrix} x \\ y \end{pmatrix} + a \leq 0, \\ & y \geq 0, \ Nx + My + q \geq 0, \end{aligned}$$

Table 5.6: Computational results for Problem 5.4 [a]

Smoothing Continuation Method	$(x_1, \cdots, x_5)$	$\epsilon_0 = 10^{-2}$	(0.0730,0.4335,0.2367,-0.4022,-0.2890)
		$\epsilon_1 = 10^{-4}$	(0.0724,0.4345,0.2386,-0.4003,-0.2870)
	Ite		43
Algorithm H	$(\hat{x}_1, \cdots, \hat{x}_5)$	$\epsilon_0 = 10^{-2}$	(0.0723,0.4345,0.2386,-0.4003,-0.2870)
	$\beta(\hat{x}, \hat{y})$	$\epsilon_0 = 10^{-2}$	{1, 2}
	Ite		32
Regularization Method	$(x_1, \cdots, x_5)$	$\epsilon_0 = 10^{-2}$	(0.0727,0.4341,0.2386,-0.4002,-0.2886)
		$\epsilon_1 = 10^{-4}$	(0.0723,0.4345,0.2386,-0.4003,-0.2871)
		$\epsilon_2 = 10^{-6}$	(0.0723,0.4345,0.2386,-0.4004,-0.2871)
	Ite		52
Algorithm H	$(\hat{x}_1, \cdots, \hat{x}_5)$	$\epsilon_0 = 10^{-2}$	(0.0731,0.4353,0.2377,-0.4006,-0.2883)
		$\epsilon_1 = 10^{-4}$	(0.0723,0.4346,0.2386,-0.4003,-0.2870)
	$\beta(\hat{x}, \hat{y})$	$\epsilon_0 = 10^{-2}$	{1, 2}
		$\epsilon_1 = 10^{-4}$	{1, 2}
	Ite		51
Penalty Method	$(x_1, \cdots, x_5)$	$\epsilon_0 = 10^{-2}$	(0.0723,0.4345,0.2387,-0.4003,-0.2870)
	Ite		31
Algorithm H	$(\hat{x}_1, \cdots, \hat{x}_5)$	$\epsilon_0 = 10^{-2}$	(0.0723,0.4345,0.2386,-0.4003,-0.2870)
	$\beta(\hat{x}, \hat{y})$	$\epsilon_0 = 10^{-2}$	{1, 2}
	Ite		35

[a] We used $(0.5, 0.5, \cdots, 0.5)$ as the initial point for all methods.

$$y^T(Nx + My + q) = 0,$$

where P, A, N, M are constant matrices and c, d, a, q are constant vectors with appropriate dimensions. QPECgen also outputs an approximate solution of a generated problem.

We set the QPECgen parameters as in Table 5.4 to generate Problems 5.4–5.7. In particular, the parameters `n` and `m` denote the dimensions of the variables x and y, respectively, and `second_deg` stands for the cardinality of the lower-level degenerate index set at a solution.

Some data obtained by the QPECgen generator are summarized in Table 5.5, in which (x^*, y^*) denotes the solution given by QPECgen and $\beta(x^*, y^*)$ stands for the lower-level degenerate index set at (x^*, y^*), i.e., $\beta(x^*, y^*) := \{i \mid (Nx^* + My^* + q)_i = 0 = y_i^*\}$.

The computational results with the identification function ρ_1 for Problems 5.4–5.7 are reported in Tables 5.6–5.9. In the tables, we only list the values of the first five components of variable x, i.e., $(x_1, \cdots, x_5)$, because that would be sufficient to illustrate the behavior of the tested algorithms. Similarly, $(\hat{x}^k, \hat{y}^k)$ and (x^k, y^k) denote the points obtained by Algorithm H and the related methods at iteration k, respectively.

The results shown in the tables reveal that it was not difficult to identify the active

Table 5.7: Computational results for Problem 5.5 [a]

Smoothing Continuation Method	$(x_1,\cdots,x_5)$	$\epsilon_0=10^{-2}$	(0.0800,0.2520,-0.1084,0.3054,-0.1941)
		$\epsilon_1=10^{-4}$	(0.0867,0.2586,-0.1176,0.2971,-0.1944)
		$\epsilon_2=10^{-6}$	(0.0872,0.2576,-0.1181,0.2958,-0.1940)
	Ite		103
Algorithm H	$(\hat{x}_1,\cdots,\hat{x}_5)$	$\epsilon_0=10^{-2}$	(0.0800,0.2520,-0.1084,0.3054,-0.1941)
		$\epsilon_1=10^{-4}$	(0.0871,0.2577,-0.1181,0.2957,-0.1940)
	$\beta(\hat{x},\hat{y})$	$\epsilon_0=10^{-2}$	{1, 2, 3, 4, 5, 6, 10}
		$\epsilon_1=10^{-4}$	{1, 2, 3, 4, 5, 6}
	Ite		59
Regularization Method	$(x_1,\cdots,x_5)$	$\epsilon_0=10^{-2}$	(0.0864,0.2614,-0.1160,0.2941,-0.1931)
		$\epsilon_1=10^{-4}$	(0.0869,0.2577,-0.1179,0.2958,-0.1938)
		$\epsilon_2=10^{-6}$	(0.0872,0.2576,-0.1181,0.2958,-0.1940)
	Ite		30
Algorithm H	$(\hat{x}_1,\cdots,\hat{x}_5)$	$\epsilon_0=10^{-2}$	(0.0864,0.2614,-0.1160,0.2941,-0.1931)
		$\epsilon_1=10^{-4}$	(0.0871,0.2577,-0.1181,0.2957,-0.1940)
	$\beta(\hat{x},\hat{y})$	$\epsilon_0=10^{-2}$	{1, 2, 3, 4, 5, 6, 10}
		$\epsilon_1=10^{-4}$	{1, 2, 3, 4, 5, 6}
	Ite		30
Penalty Method	$(x_1,\cdots,x_5)$	$\epsilon_0=10^{-2}$	(0.0872,0.2576,-0.1181,0.2958,-0.1940)
		$\epsilon_1=10^{-4}$	(0.0871,0.2576,-0.1181,0.2957,-0.1940)
		$\epsilon_2=10^{-6}$	(0.0871,0.2593,-0.1169,0.2969,-0.1965)
		$\epsilon_3=10^{-8}$	(-0.1005,0.2845,-0.0061,0.2642,-0.1371)
	Ite		59
Algorithm H	$(\hat{x}_1,\cdots,\hat{x}_5)$	$\epsilon_0=10^{-2}$	(0.0871,0.2577,-0.1181,0.2957,-0.1940)
		$\epsilon_1=10^{-4}$	(0.0871,0.2577,-0.1181,0.2957,-0.1940)
		$\epsilon_2=10^{-6}$	(0.0871,0.2577,-0.1181,0.2957,-0.1940)
		$\epsilon_3=10^{-8}$	(0.0871,0.2577,-0.1181,0.2957,-0.1940)
	$\beta(\hat{x},\hat{y})$	$\epsilon_0=10^{-2}$	{1, 2, 3, 4, 5, 6}
		$\epsilon_1=10^{-4}$	{1, 2, 3, 4, 5, 6}
		$\epsilon_2=10^{-6}$	{1, 2, 3, 4, 5, 6}
		$\epsilon_3=10^{-8}$	{1}
	Ite		71

[a] We used $(2,2,\cdots,2)$ as the initial point for all methods.

sets by Algorithm H, at least for the test problems used in our numerical experiments, although we have observed that the penalty method may not be very stable when the parameter ϵ becomes small. In fact, we got the correct active sets in no more than three steps in all cases and, since the computed points satisfy the B-stationarity conditions, Algorithm H terminated. Especially, as mentioned in the last section, Algorithm H may terminate by finding a solution before the correct index sets are obtained, see Table 5.2. Moreover, we notice that, since the number of active indices is larger than the dimension of z, Problems 5.2, 5.5, and 5.7 do not satisfy the MPCC-LICQ at the solutions. Nevertheless, we were able to obtain the solutions successfully, which shows the robustness of the proposed approach.

We have also solved these test problems by some solvers from NEOS [21]. The results indicates that the hybrid approach is comparable to those methods. For example, the total iterations spent by the solvers DONLP2 / MINOS / SNOPT for Problems 5.1, 5.2, and 5.3 are 8 / 4 / 8, 12 / 3 / 6, and 9 / 11 / 9, respectively. Moreover, it should be emphasized that the proposed algorithms are theoretically guaranteed to solve MPCCs under mild conditions, while the methods in NEOS are in general not, even if they could solve the test problems successfully.

5.6.2 Computational results for Algorithms HIA and HIS

In this subsection, we examine the effectiveness of Algorithms HIA and HIS on some examples of MPCC. Since the numerical results shown in the last subsection have revealed that Algorithm H is comparable to some existing methods such as the smoothing continuation method [31], the penalty function method [38], and the regularization method [76], we only compare Algorithms HIA and HIS with Algorithm H.

We show the QPECgen parameters used to generate Problems 5.8 and 5.9 in Table 5.10, and summarize some data output by the QPECgen generator in Table 5.11. In our experiments, we set $\epsilon_0 = 10^{-2}$ and updated this parameter by $\epsilon_{k+1} = 10^{-1}\epsilon_k$. Moreover, we employ

$$\rho(x,y) = \|\min(Nx + My + q, y)\|^{\frac{4}{5}}$$

as the identification function in Step 1 of both Algorithm H and Algorithm HIA. In Algorithm HIS, we use the sequence $\{\xi_k\}$ given by

$$\xi_k = \rho(x^k, y^k), \qquad k = 0, 1, 2, \cdots,$$

which is a reasonable choice to compare with the other two methods. See the tables for the setting of the other parameters involved.

The computational results for Problems 5.8 and 5.9 are reported in Tables 5.12 and 5.13, respectively. In the tables, `distance` denotes the distance between the obtained point and the solution (x^*, y^*) measured by the infinity norm. In the `Ite` column, a sum $\nu_1 + \nu_2$ means that ν_1 is the number of iterations spent by *fmincon* for solving problem (P_ϵ) and ν_2 denotes the number of iterations spent by solving subproblem (5.86), whereas a single number ν stands for the number of iterations spent by solving subproblem (5.86) (as there is no need to solve problem (P_ϵ) in these cases).

The results shown in the tables reveal that both Algorithms HIA and HIS were able to identify the active sets successfully. As mentioned in the previous section, Algorithms HIA and HIS need to solve more problems of the form (5.86) than Algorithm H, whereas the latter has to solve more problems of the form (P_ϵ). Our experiments show that problem (5.86) can be solved in fewer iterations than problem (P_ϵ) generally.

Table 5.8: Computational results for Problem 5.6 [a]

Smoothing Continuation Method	$(x_1, \cdots, x_5)$	$\epsilon_0 = 10^{-2}$	(0.6053,0.7516,0.1275,-0.6123,-0.6650)
		$\epsilon_1 = 10^{-4}$	(0.6358,0.6367,0.1739,-0.7155,-0.8697)
		$\epsilon_2 = 10^{-6}$	(0.6366,0.6368,0.1736,-0.7157,-0.8698)
		$\epsilon_3 = 10^{-8}$	(0.6370,0.6366,0.1738,-0.7158,-0.8703)
	Ite		185
Algorithm H	$(\hat{x}_1, \cdots, \hat{x}_5)$	$\epsilon_0 = 10^{-2}$	(0.7361,0.4854,0.3201,-0.6325,-0.6232)
		$\epsilon_1 = 10^{-4}$	(0.6368,0.6371,0.1739,-0.7158,-0.8704)
	$\beta(\hat{x}, \hat{y})$	$\epsilon_0 = 10^{-2}$	{2, 4, 7}
		$\epsilon_1 = 10^{-4}$	{1, 2, 3, 4, 5}
	Ite		132
Regularization Method	$(x_1, \cdots, x_5)$	$\epsilon_0 = 10^{-2}$	(0.6368,0.6321,0.1734,-0.7145,-0.8746)
		$\epsilon_1 = 10^{-4}$	(0.6368,0.6368,0.1737,-0.7158,-0.8704)
		$\epsilon_2 = 10^{-6}$	(0.6367,0.6371,0.1739,-0.7158,-0.8702)
		$\epsilon_3 = 10^{-8}$	(0.6369,0.6370,0.1739,-0.7158,-0.8703)
	Ite		105
Algorithm H	$(\hat{x}_1, \cdots, \hat{x}_5)$	$\epsilon_0 = 10^{-2}$	(0.6369,0.6371,0.1739,-0.7159,-0.8703)
		$\epsilon_1 = 10^{-4}$	(0.6368,0.6370,0.1739,-0.7158,-0.8704)
		$\epsilon_2 = 10^{-6}$	(0.6368,0.6370,0.1739,-0.7158,-0.8704)
	$\beta(\hat{x}, \hat{y})$	$\epsilon_0 = 10^{-2}$	{1, 4, 5}
		$\epsilon_1 = 10^{-4}$	{2, 3, 4, 5}
		$\epsilon_2 = 10^{-6}$	{1, 2, 3, 4, 5}
	Ite		102
Penalty Method	$(x_1, \cdots, x_5)$	$\epsilon_0 = 10^{-2}$	(0.6369,0.6370,0.1739,-0.7158,-0.8703)
		$\epsilon_1 = 10^{-4}$	(0.6368,0.6370,0.1739,-0.7158,-0.8702)
		$\epsilon_2 = 10^{-6}$	(0.6365,0.6368,0.1742,-0.7158,-0.8709)
		$\epsilon_3 = 10^{-8}$	(0.6360,0.6343,0.1735,-0.7148,-0.8683)
	Ite		104
Algorithm H	$(\hat{x}_1, \cdots, \hat{x}_5)$	$\epsilon_0 = 10^{-2}$	(0.6368,0.6370,0.1739,-0.7158,-0.8703)
		$\epsilon_1 = 10^{-4}$	(0.6368,0.6371,0.1739,-0.7158,-0.8704)
		$\epsilon_2 = 10^{-6}$	(0.6369,0.6370,0.1739,-0.7158,-0.8703)
		$\epsilon_3 = 10^{-8}$	(0.6368,0.6371,0.1739,-0.7158,-0.8703)
	$\beta(\hat{x}, \hat{y})$	$\epsilon_0 = 10^{-2}$	{1, 2, 3, 4, 5}
		$\epsilon_1 = 10^{-4}$	{1, 2, 3, 4, 5}
		$\epsilon_2 = 10^{-6}$	{3, 4, 5}
		$\epsilon_3 = 10^{-8}$	{2, 3, 4, 5}
	Ite		132

[a] $(0.5, 0.5, \cdots, 0.5)$ was employed as the initial point for all methods.

Table 5.9: Computational results for Problem 5.7 [a]

Method			Result
Smoothing Continuation Method	$(x_1, \cdots, x_5)$	$\epsilon_0 = 10^{-2}$	(-0.0915,0.0196,-0.3204,0.2065,-0.4307)
		$\epsilon_1 = 10^{-4}$	(-0.0991,0.0163,-0.3107,0.1972,-0.4279)
		$\epsilon_2 = 10^{-6}$	(-0.1030,0.0150,-0.3080,0.1958,-0.4340)
		$\epsilon_3 = 10^{-8}$	(-0.1002,0.0159,-0.3091,0.1947,-0.4276)
	Ite		104
Algorithm H	$(\hat{x}_1, \cdots, \hat{x}_5)$	$\epsilon_0 = 10^{-2}$	(-0.1018,0.0167,-0.3091,0.1947,-0.4275)
		$\epsilon_1 = 10^{-4}$	(-0.1018,0.0167,-0.3091,0.1947,-0.4275)
	$\beta(\hat{x}, \hat{y})$	$\epsilon_0 = 10^{-2}$	$\emptyset$
		$\epsilon_1 = 10^{-4}$	{1, 2, 3, 4, 5, 6}
	Ite		41
Regularization Method	$(x_1, \cdots, x_5)$	$\epsilon_0 = 10^{-2}$	(-0.1007,0.0160,-0.3093,0.1952,-0.4274)
		$\epsilon_1 = 10^{-4}$	(-0.1007,0.0162,-0.3095,0.1954,-0.4274)
		$\epsilon_2 = 10^{-6}$	(-0.1016,0.0166,-0.3092,0.1948,-0.4273)
		$\epsilon_3 = 10^{-8}$	(-0.1016,0.0167,-0.3093,0.1948,-0.4271)
	Ite		35
Algorithm H	$(\hat{x}_1, \cdots, \hat{x}_5)$	$\epsilon_0 = 10^{-2}$	(-0.1007,0.0160,-0.3093,0.1952,-0.4274)
		$\epsilon_1 = 10^{-4}$	(-0.1018,0.0167,-0.3091,0.1947,-0.4275)
	$\beta(\hat{x}, \hat{y})$	$\epsilon_0 = 10^{-2}$	{1, 2, 3, 4, 5, 6, 8, 10}
		$\epsilon_1 = 10^{-4}$	{1, 2, 3, 4, 5, 6}
	Ite		32
Penalty Method	$(x_1, \cdots, x_5)$	$\epsilon_0 = 10^{-2}$	(-0.1017,0.0166,-0.3092,0.1947,-0.4273)
		$\epsilon_1 = 10^{-4}$	(-0.1018,0.0166,-0.3092,0.1947,-0.4273)
		$\epsilon_2 = 10^{-6}$	(-0.0885,0.0124,-0.3119,0.1926,-0.4274)
		$\epsilon_3 = 10^{-8}$	(0.2274,0.4623,-0.1441,-0.1743,-1.2077)
	Ite		65
Algorithm H	$(\hat{x}_1, \cdots, \hat{x}_5)$	$\epsilon_0 = 10^{-2}$	(-0.1018,0.0167,-0.3091,0.1947,-0.4275)
		$\epsilon_1 = 10^{-4}$	(-0.1018,0.0167,-0.3092,0.1947,-0.4275)
		$\epsilon_2 = 10^{-6}$	(-0.1017,0.0167,-0.3091,0.1947,-0.4275)
		$\epsilon_3 = 10^{-8}$	(-0.0645,-0.0118,-0.2910,0.1588,-0.4824)
	$\beta(\hat{x}, \hat{y})$	$\epsilon_0 = 10^{-2}$	{1, 2, 3, 4, 5, 6}
		$\epsilon_1 = 10^{-4}$	{1, 3, 4, 5, 6}
		$\epsilon_2 = 10^{-6}$	{2, 5, 6}
		$\epsilon_3 = 10^{-8}$	{8}
	Ite		79

[a] We used $(0.5, 0.5, \cdots, 0.5)$ as the initial point for all methods.

Table 5.10: Parameters in QPECgen for Problems 5.8-5.9

Problem	`qpec_type`	`(n, m)`	`(l, p)`	`cond_P`	`scale_P`	`convex_f`
# 5.8	300	(10, 10)	(5, 10)	100	100	1
# 5.9	300	(8, 14)	(4, 14)	100	100	1
Problem	`symm_M`	`mono_M`	`cond_M`	`scale_M`	`second_deg`	`first_deg`
# 5.8	1	1	200	200	4	3
# 5.9	1	1	200	200	4	2
Problem	`mix_deg`	`tol_deg`	`implicit`	`rand_seed`	`output`	
# 5.8	3	1.0e-6	0	0	3	
# 5.9	1	1.0e-6	0	0	3	

Table 5.11: Some data obtained by QPECgen for Problems 5.8-5.9

Problem	x^*	y^*	$\beta(x^*, y^*)$
# 5.8	(0.6369, 0.6371, 0.1739, -0.7158, -0.8703, -0.7478, -0.4383, 0.1886, 0.0741, 0.1494)	(0, 0, ···, 0)	{1, 2, 3, 4}
# 5.9	(-0.7330, -0.2090, -0.4140, 0.0168, -0.7084, -0.1104, 0.0030, -0.4658)	(0,···,0,0.4681,0.4739,0.2088)	{1, 2, 3, 4}

Table 5.12: Computational results for Problem 5.8 [a]

	ϵ_k	degenerate set	distance	Ite
Algorithm H	10^{-2}	$\beta^k = \emptyset$	0.0007	25+10
	10^{-3}	$\beta^k = \emptyset$	0.0010	29+10
	10^{-4}	$\beta^k = \{4\}$	0.0005	23+11
Algorithm HIA	10^{-2}	$\beta_0^k = \emptyset$	0.0007	25+10
		$\beta_1^k = \{4\}$	0.0007	10
		$\beta_2^k = \{2,4\}$	0.0005	11
		$\beta_3^k = \{1,2,4\}$	0.0004	10
		$\beta_4^k = \{1,2,3,4\}$	0.0004	10
Algorithm HIS	10^{-2}	$\beta_0^k = \{1,2,3,4,7,9,10\}$	1.5944	25+6
		$\beta_1^k = \{1,2,3,4,9,10\}$	1.2922	5
		$\beta_2^k = \{1,2,3,4,10\}$	0.4131	9
		$\beta_3^k = \{1,2,3,4\}$	0.0004	10

[a] We used $(1, 1, \cdots, 1)$ as the initial point for all methods and we set $\theta_0 = 0.2$ in Algorithms HIA and HIS. In addition, the parameter η in Algorithm HIS is set to be 0.5.

Table 5.13: Computational results for Problem 5.9 [a]

	ϵ_k	degenerate set	distance	Ite
Algorithm H	10^{-2}	$\beta^k = \emptyset$	0.0003	27+6
	10^{-3}	$\beta^k = \{1,2,3\}$	0.0004	19+5
	10^{-4}	$\beta^k = \{1,2,3,4\}$	0.0003	35+7
Algorithm HIA	10^{-2}	$\beta_0^k = \emptyset$	0.0003	27+6
		$\beta_1^k = \{2\}$	0.0003	6
		$\beta_2^k = \{1,2\}$	0.0009	3
		$\beta_3^k = \{1,2,3\}$	0.0009	3
		$\beta_4^k = \{1,2,3,4\}$	0.0009	3
Algorithm HIS	10^{-2}	$\beta_0^k = \{1,2,3,4,11\}$	0.8554	27+6
		$\beta_1^k = \{1,2,3,4\}$	0.0009	3

[a] We employed $(1, 1, \cdots, 1)$ as the initial point for all methods. We set $\theta_0 = 0.2$ in Algorithms HIA and HIS and, $\eta = 0.2$ in Algorithm HIS.

Chapter 6

Smoothing Implicit Programming Approach for SMPECs

From this chapter, we begin to study the stochastic mathematical programs with equilibrium constraints (SMPECs), which can be thought as generalizations of the mathematical programs with equilibrium constraints. We first introduce the problems and then, we show that many decision problems can be formulated as SMPECs in practice. We discuss two kinds of models: the lower-level wait-and-see decision model and the here-and-now decision model. For the lower-level wait-and-see model, we propose a smoothing implicit programming method and establish a comprehensive convergence theory. For the here-and-now decision problem, we apply a penalty technique and suggest a similar method. We show that the two methods possess similar convergence properties.

6.1 Introduction

In this chapter, we discuss MPECs under uncertainty. That is, we consider the stochastic mathematical program with equilibrium constraints (SMPEC):

$$\begin{array}{ll} \text{minimize} & E_\omega[f(x,y,\omega)] \\ \text{subject to} & x \in X, \quad \omega \in \Omega, \\ & y \text{ solves } \mathrm{VI}(F(x,\cdot,\omega), C(x,\omega)), \end{array} \tag{6.1}$$

where X is a subset of $\Re^n$, Ω stands for the underlying sample space, E_ω means expectation with respect to the random variable $\omega \in \Omega$, and $f : \Re^{n+m} \times \Omega \to \Re, F : \Re^{n+m} \times \Omega \to \Re^m$, $C : \Re^n \times \Omega \to 2^{\Re^m}$ are mappings. Obviously, if Ω is a singleton, then problem (6.1) reduces to an ordinary MPEC, and so SMPECs can be thought as generalized MPECs. It is well known that an MPEC is a hard problem because its constraints fail to satisfy a standard constraint qualification at any feasible point [17]. This suggests that SMPECs may be more difficult to deal with because the number of random events is usually very large in practice.

When $C(x,\omega) \equiv \Re^m_+$ for any $x \in X$ and any $\omega \in \Omega$ in problem (6.1), the variational inequality constraints reduce to the complementarity constraints and problem (6.1) is equivalent to the following stochastic mathematical program with complementarity constraints (SMPCC):

$$\begin{array}{ll} \text{minimize} & E_\omega[f(x,y,\omega)] \\ \text{subject to} & x \in X, \quad \omega \in \Omega, \\ & y \geq 0, \ F(x,y,\omega) \geq 0, \\ & y^T F(x,y,\omega) = 0. \end{array} \tag{6.2}$$

On the other hand, if the set $C(x,\omega)$ is defined by

$$C(x,\omega) = \{y \in \Re^m | \ \ c(x,y,\omega) \leq 0\},$$

where $c(\cdot,\cdot,\omega)$ is continuously differentiable, then, under some suitable conditions, the variational inequality problem VI$(F(x,\cdot,\omega), C(x,\omega))$ has an equivalent Karush-Kuhn-Tucker representation

$$\begin{array}{l} F(x,y,\omega) + \nabla_y c(x,y,\omega)\lambda(x,\omega) = 0, \\ \lambda(x,\omega) \geq 0, \quad c(x,y,\omega) \leq 0, \quad \lambda(x,\omega)^T c(x,y,\omega) = 0, \end{array}$$

where $\lambda(x,\omega)$ is the Lagrange multiplier vector [68]. As a result, problem (6.1) can be reformulated as a program like (6.2) under some conditions, see the monograph [62] for details. Hence, problem (6.2) constitutes an important subclass of SMPECs. In this chapter, we concentrate on this kind of SMPECs.

Problem (6.2) looks like a standard MPEC. However, the existence of the random variable ω means that (6.2) involves multiple complementarity-type constraints and it is therefore more difficult to solve than an ordinary MPEC generally.

We call (6.2) a lower-level *wait-and-see* model with an upper-level decision x and a lower-level decision y. In this kind of decision problems, we wait until an observation is made on the random events, and then we make an optimal lower-level decision $y = y(\omega)$ based on the observed information. This kind of problems has been discussed in [72], the main results of which are concerned with the existence of solutions, the convexity and directional differentiability of an implicit objective function, and links between SMPEC and bilevel models. Actually, there have been no effective algorithms suggested for solving SMPECs so far. In this chapter, we will propose an implicit programming approach for (6.2). The problems dealt with in this chapter include not only the lower-level wait-and-see model, but a more practical problem as well that requires us to make all decisions at once, before ω is observed:

$$\begin{array}{ll} \text{minimize} & E_\omega[f(x,y,\omega)+d^T z(\omega)] \\ \text{subject to} & x \in X, \quad \omega \in \Omega, \\ & y \geq 0, \ F(x,y,\omega)+z(\omega) \geq 0, \\ & y^T(F(x,y,\omega)+z(\omega)) = 0, \\ & z(\omega) \geq 0, \end{array} \tag{6.3}$$

where $z(\omega)$ is called a recourse variable and $d \in \Re^m$ is a vector with positive elements. We call (7.1) a *here-and-now* model.

We note that, in either of the lower-level wait-and-see and here-and-now models, the upper-level decision is made 'here-and-now'. Therefore, both models yield decision problems unlike single-level stochastic programming, in which the wait-and-see model is not a decision problem [46, 83]. The following example illustrates the two models.

Example 6.1 There are a food company who makes picnic lunches and a vendor who sells lunches to hikers on every Sunday. The company and the vendor have the following contract:

C1: The vendor buys lunches from the company at the price $x \in [a,b]$ determined by the company, where a and b are two positive constants.

C2: The vendor decides the amount y of lunches that he buys from the company, where y must be no less than the minimum amount $c > 0$.

C3: The vendor pays the company for the whole lunches he buys, i.e., the vendor pays xy to the company.

C4: The vendor sells lunches to hikers at the price $2x$ and get the proceeds for the total number of lunches actually sold.

C5: Even if there are any unsold lunches, the vendor cannot return them to the company but he can dispose of the unsold lunches with no cost.

We suppose that the demand of lunches depends on the price and the weather on that day. Since the weather is uncertain, we may treat it as a random variable. More specifically, we suppose that the demand is given by the function

$$\phi(x,\omega) := D(\omega) - d(\omega)x, \quad \omega \in \Omega,$$

where $D(\omega) \geq 0$ and $d(\omega) \geq 0$ are random variables. Therefore, the actual amount of lunches sold is given by $\min(y, \phi(x,\omega))$, which also depends on the weather on that day.

The decisions by the company and the vendor are x and y, respectively. The company's objective is to maximize its total profit xy, while the vendor's objective is to maximize its total profit $2x\min(y, \phi(x,\omega)) - xy$. The latter problem may be written as

$$\begin{array}{ll} \text{maximize}_{y,t} & x(2t-y) \\ \text{subject to} & y \geq c, \; y - t \geq 0, \\ & D(\omega) - d(\omega)x - t \geq 0, \end{array}$$

whose optimality conditions are stated as

$$\begin{pmatrix} x \\ -2x \end{pmatrix} - u\begin{pmatrix} 1 \\ 0 \end{pmatrix} - v\begin{pmatrix} 1 \\ -1 \end{pmatrix} - w\begin{pmatrix} 0 \\ -1 \end{pmatrix} = 0, \tag{6.4}$$

$$\begin{aligned} &0 \leq u \perp (y-c) \geq 0, \\ &0 \leq v \perp (y-t) \geq 0, \\ &0 \leq w \perp (D(\omega) - d(\omega)x - t) \geq 0. \end{aligned} \tag{6.5}$$

Here, $\lambda \perp \mu$ means $\lambda\mu = 0$. It follows from (6.4) that

$$u = x - v, \quad w = 2x - v.$$

This implies that $w = x + u \geq a > 0$, which together with (6.5) yields $t = D(\omega) - d(\omega)x$. Thus the above optimality conditions may further be rewritten as

$$\begin{aligned} &0 \leq (x-v) \perp (y-c) \geq 0, \\ &0 \leq v \perp (y - D(\omega) + d(\omega)x) \geq 0. \end{aligned} \tag{6.6}$$

Then the company's problem may be written as the following stochastic MPEC:

$$\begin{array}{ll} \text{minimize}_{x,y} & -xy \\ \text{subject to} & a \leq x \leq b, \\ & 0 \leq (x-v) \perp (y-c) \geq 0, \\ & 0 \leq v \perp (y - D(\omega) + d(\omega)x) \geq 0. \end{array}$$

Now there are two cases.

Here-and-now model: Suppose that both the company and the vendor have to make decision on Saturday, without knowing the weather of Sunday. In this case, there is no (x, v) satisfying (6.6) for all $\omega \in \Omega$ in general. So, by introducing the recourse variables, the company's problem is represented as the following model:

$$\begin{array}{ll} \text{minimize} & -xy + \beta E_\omega[\, z(\omega)\,] \\ \text{subject to} & a \le x \le b, \\ & 0 \le (x - v) \;\perp\; (y - c) \ge 0, \\ & 0 \le v \;\perp\; (y - D(\omega) + d(\omega)x + z(\omega)) \ge 0, \\ & z(\omega) \ge 0, \quad \omega \in \Omega, \end{array}$$

where $\beta > 0$ is a constant.

Lower-level wait-and-see model: Suppose that the company makes a decision on Saturday, but the vendor can make a decision on Sunday morning after knowing the weather of that day. In this case, the vendor's decision may depend on the observation of ω, which is given by $(y(\omega), v(\omega))$ that satisfies

$$\begin{array}{l} 0 \le (x - v(\omega)) \;\perp\; (y(\omega) - c) \ge 0, \\ 0 \le v(\omega) \;\perp\; (y(\omega) - D(\omega) + d(\omega)x) \ge 0 \end{array}$$

for each $\omega \in \Omega$. Therefore the company's problem is represented as the following model:

$$\begin{array}{ll} \text{minimize} & E_\Omega[\, -y(\omega)x\,] \\ \text{subject to} & a \le x \le b, \quad \omega \in \Omega, \\ & 0 \le (x - v(\omega)) \;\perp\; (y(\omega) - c) \ge 0, \\ & 0 \le v(\omega) \;\perp\; (y(\omega) - D(\omega) + d(\omega)x) \ge 0. \end{array}$$

Organization of this chapter: In the next section, we will present a smoothing implicit programming method for the lower-level wait-and-see problem with linear complementarity constraints. A comprehensive convergence theory will also be included. In Section 6.3, we deal with the here-and-now problem and, by means of a penalty technique, we suggest a similar method for solving this kind of problems. Concluding remarks are made in the final section.

We will use the following notations in this and next two chapters: All vectors are thought as column vectors and $x[i]$ stands for the ith coordinate of the vector $x \in \Re^n$, whereas for a matrix M, we denote by $M[i]$ the vector whose elements consist of the ith row of M. If $\mathcal{K}$ is an index set, we let $M[\mathcal{K}]$ be the principal submatrix of M whose elements consist of those of M indexed by $\mathcal{K}$. For any vectors u and v of the same dimension, we denote $u \perp v$ to mean $u^T v = 0$. For two vectors $u, v \in \Re^s$, $\min(u, v)$ is understood to be taken componentwise, i.e.,

$$\min(u, v) = (\min\{u[1], v[1]\}, \cdots, \min\{u[s], v[s]\})^T.$$

In addition, e_i denotes the unit vector with $e_i[i] = 1$; I and O denote the identity matrix and the zero matrix with suitable dimension, respectively.

6.2 Smoothing Implicit Programming Method for Lower-Level Wait-And-See Problems

In this section, we consider the following stochastic mathematical program with linear complementarity constraints (SMPLCC):

$$\begin{array}{ll} \text{minimize} & \sum_{\ell=1}^{L} p_\ell f(x, y_\ell) \\ \text{subject to} & g(x) \leq 0, \;\; h(x) = 0, \\ & y_\ell \geq 0, \;\; N_\ell x + M_\ell y_\ell + q_\ell \geq 0, \\ & y_\ell^T (N_\ell x + M_\ell y_\ell + q_\ell) = 0, \;\; \ell = 1, \cdots, L, \end{array} \tag{6.7}$$

which corresponds to the discrete case where $\Omega = \{\omega_1, \omega_2, \cdots, \omega_L\}$. Here, p_ℓ denotes the probability of the random event $\omega_\ell \in \Omega$, i.e.,

$$\sum_{\ell=1}^{L} p_\ell = 1, \quad p_\ell \geq 0, \qquad \ell = 1, \cdots, L,$$

the functions $f : \Re^{n+m} \to \Re, g : \Re^n \to \Re^{s_1}$, and $h : \Re^n \to \Re^{s_2}$ are all continuously differentiable, and $N_\ell \in \Re^{m \times n}$, $M_\ell \in \Re^{m \times m}$, $q_\ell \in \Re^m$ are realizations of the random coefficients. Problem (6.7) represents a lower-level wait-and-see model, since the lower-level decisions y_ℓ are associated with possible outcomes ω_ℓ of the random variable ω, which means that, unlike the upper-level decision, the lower-level decisions are made after a random event is observed. Throughout we assume $p_\ell > 0$ for all $\ell = 1, \cdots, L$.

By letting

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_L \end{pmatrix}, \quad \mathbf{N} = \begin{pmatrix} N_1 \\ \vdots \\ N_L \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} M_1 & & O \\ & \ddots & \\ O & & M_L \end{pmatrix}, \quad \mathbf{q} = \begin{pmatrix} q_1 \\ \vdots \\ q_L \end{pmatrix}, \tag{6.8}$$

and

$$\mathrm{f}(x, \mathbf{y}) = \sum_{\ell=1}^{L} p_\ell f(x, y_\ell), \tag{6.9}$$

problem (6.7) can be rewritten as

$$\begin{array}{ll} \text{minimize} & \mathrm{f}(x, \mathbf{y}) \\ \text{subject to} & g(x) \leq 0,\ h(x) = 0, \\ & \mathbf{y} \geq 0,\ \mathbf{N}x + \mathbf{M}\mathbf{y} + \mathbf{q} \geq 0, \\ & \mathbf{y}^T(\mathbf{N}x + \mathbf{M}\mathbf{y} + \mathbf{q}) = 0, \end{array} \tag{6.10}$$

which is an ordinary MPEC. However, since L is usually very large in practice, problem (6.10) is a large-scale program with variables $(x, \mathbf{y}) \in \Re^{n+mL}$ so that some methods for MPECs may cause more computational difficulties. Here, we treat problem (6.7) as a mathematical program with multiple complementarity-type constraints.

Recently, Chen and Fukushima [12] have suggested a smoothing method for an MPEC with P-matrix linear complementarity constraints. We will develop a similar smoothing method for solving problem (6.7), or equivalently (6.10), with $\mathbf{M}$ being a P_0-matrix. Specifically, in addition to smoothing, we will employ a regularization technique to make an implicit programming approach applicable. We will investigate the limiting behavior of the method under appropriate assumptions.

6.2.1 Preliminaries

We first recall some basic concepts and properties that will be used later on.

Definition 6.1 [19] Suppose that M is an $m \times m$ matrix. We call M a *P-matrix* if all the principal minors of M are positive, or equivalently,

$$\max_{1 \leq i \leq m} y[i](My)[i] > 0, \qquad 0 \neq \forall y \in \Re^m,$$

and we call M a *P_0-matrix* if all the principal minors of M are nonnegative, or equivalently,

$$\max_{1 \leq i \leq m} y[i](My)[i] \geq 0, \qquad \forall y \in \Re^m.$$

It is obvious that a P-matrix must be a P_0-matrix and, if M is a P_0-matrix and μ is a positive number, then the matrix $M + \mu I$ is a P-matrix.

Definition 6.2 [19] A square matrix is said to be *nondegenerate* if all of its principal submatrices are nonsingular.

It is easy to see that a P-matrix is nondegenerate.

For given $N \in \Re^{m\times n}$, $M \in \Re^{m\times m}$, $q \in \Re^m$, and two positive numbers ϵ and μ, we define the function

$$\Phi_{\epsilon,\mu}(x,y,w;N,M,q) = \begin{pmatrix} Nx + (M+\epsilon I)y + q - w \\ \phi_\mu(y[1],w[1]) \\ \vdots \\ \phi_\mu(y[m],w[m]) \end{pmatrix}, \tag{6.11}$$

where $\phi_\mu : \Re^2 \to \Re$ is the perturbed Fischer-Burmeister function

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

Then we have the following well-known result [12, 47].

Theorem 6.1 *Suppose that M is a P_0-matrix. Then, for given $x \in \Re^n$, $\epsilon > 0$, and $\mu > 0$, we have the following statements:*

(1) *The function $\Phi_{\epsilon,\mu}$ defined by (6.11) is continuously differentiable with respect to (y,w) and the Jacobian matrix $\nabla_{(y,w)}\Phi_{\epsilon,\mu}(x,y,w;N,M,q)$ is nonsingular everywhere;*

(2) *The equation*

$$\Phi_{\epsilon,\mu}(x,y,w;N,M,q) = 0 \tag{6.12}$$

has a unique solution $(y(x,\epsilon,\mu), w(x,\epsilon,\mu))$, which is continuously differentiable with respect to x and satisfies

$$\begin{aligned} &y(x,\epsilon,\mu) > 0, \quad w(x,\epsilon,\mu) > 0, \\ &y(x,\epsilon,\mu)[i]w(x,\epsilon,\mu)[i] = \mu^2, \quad i = 1,\cdots,m. \end{aligned} \tag{6.13}$$

In the following, to mitigate the notational complication, we assume $\epsilon = \mu$ and denote $\Phi_{\epsilon,\mu}$, $y(x,\epsilon,\mu)$, and $w(x,\epsilon,\mu)$ by Φ_μ, $y(x,\mu)$, and $w(x,\mu)$, respectively. Our analysis will remain valid, however, even though the two parameters are treated independently.

Suppose that each M_ℓ is a P_0-matrix in problem (6.7) and $\mu > 0$. Theorem 6.1 indicates that, for any fixed ℓ and $\mu > 0$, the smooth equation

$$\Phi_\mu(x,y_\ell,w_\ell;N_\ell,M_\ell,q_\ell) = 0 \tag{6.14}$$

gives two functions $y_\ell(\cdot,\mu)$ and $w_\ell(\cdot,\mu)$ that are well-defined and continuously differentiable. Note that

$$\phi_\mu(a,b)=0 \quad \Longleftrightarrow \quad a\ge 0, b\ge 0, ab=\mu^2.$$

As a result, the equation (6.14) is equivalent to the system

$$\begin{aligned} &y_\ell \ge 0,\ \ N_\ell x+(M_\ell+\mu I)y_\ell+q_\ell \ge 0,\\ &y_\ell[i]\Big(N_\ell x+(M_\ell+\mu I)y_\ell+q_\ell\Big)[i]=\mu^2, \quad i=1,\cdots,m \end{aligned} \tag{6.15}$$

in the sense that $y_\ell(x,\mu)$ solves (6.15) if and only if

$$\Phi_\mu(x,y_\ell(x,\mu),w_\ell(x,\mu);N_\ell,M_\ell,q_\ell)=0 \tag{6.16}$$

with

$$w_\ell(x,\mu)=N_\ell x+(M_\ell+\mu I)y_\ell(x,\mu)+q_\ell\ . \tag{6.17}$$

Since the system (6.15) with $\mu=0$ reduces to the linear complementarity problem LCP$(x;N_\ell,M_\ell,\,q_\ell)$:

$$\begin{aligned} &y_\ell \ge 0,\ \ N_\ell x+M_\ell y_\ell+q_\ell \ge 0,\\ &y_\ell^T\Big(N_\ell x+M_\ell y_\ell+q_\ell\Big)=0, \end{aligned} \tag{6.18}$$

we see that $y_\ell(x,\mu)$ tends to a solution of (6.18) as $\mu\to 0$, provided that it is convergent.

In our analysis, we will simply assume that $y_\ell(x,\mu)$ is bounded as $\mu\to 0$. In particular, if M_ℓ is a P-matrix, then (6.18) has a unique solution for any x and it can be shown that $y_\ell(x,\mu)$ actually converges to it as $\mu\to 0$, even without using the regularization term μI in (6.15), see [12].

The following lemma will be used later on.

Lemma 6.1 [29] *Let M be a P_0-matrix and $\hat D=\mathrm{diag}(\hat d_1,\cdots,\hat d_m)$ with $0\le \hat d_i\le 1$ for each i. If the principal submatrix $M[\hat{\mathcal K}]$ is nonsingular, where $\hat{\mathcal K}=\{\ i\ |\ \ \hat d_i=1\}$, then the matrix*

$$\begin{pmatrix} M & -I\\ I-\hat D & \hat D \end{pmatrix}$$

is nonsingular.

6.2.2 Method

Let $\{\mu_k\}$ be a sequence of positive numbers converging to 0. A smoothing implicit programming method for problem (6.7), which we call SIP-I, generates a sequence $\{x^{(k)}\}$ by solving the problems

$$\begin{aligned} \text{minimize} \quad & \theta_{\mu_k}(x) \\ \text{subject to} \quad & g(x) \le 0, \;\; h(x) = 0, \end{aligned} \tag{6.19}$$

where

$$\theta_{\mu_k}(x) = \sum_{\ell=1}^{L} p_\ell f(x, y_\ell(x, \mu_k)) \tag{6.20}$$

and $y_\ell(x, \mu_k)$ satisfies the system (6.16)–(6.17) with $\mu = \mu_k$. Let

$$y_\ell^{(k)} = y_\ell(x^{(k)}, \mu_k), \quad \ell = 1, \cdots, L. \tag{6.21}$$

Throughout, we denote by $\mathcal{F}_1$ and $\mathcal{X}$ the feasible regions of problems (6.7) and (6.19), respectively. Moreover, particular sequences generated by the method will be denoted by $\{x^{(k)}\}, \{y^{(k)}\}$, etc., while general sequences will be denoted by $\{x^k\}, \{y^k\}$, etc.

Note that, by Theorem 6.1, problem (6.19) is a smooth mathematical program. In particular, if $\mathcal{X} = \Re^n$, problem (6.19) reduces to a smooth unconstrained optimization problem. Moreover, under some suitable conditions, (6.19) is a convex program, see [12] for details. Therefore, we may expect that problem (6.19) may be relatively easy to deal with, provided the evaluation of the function $y_\ell(x, \mu_k)$ is not very expensive.

6.2.3 Limiting behavior of local optimal solutions

In this section, we first give some properties of the functions $y_\ell(\cdot, \mu)$ and then investigate the limiting behavior of sequences of local optimal solutions of problem (6.19). To this end, the following assumptions are assumed throughout this subsection:

A1: SIP-I produces a bounded sequence $\{x^{(k)}\}$ of local optimal solutions of (6.19).

A2: For any bounded sequence $\{x^k\}$ in $\mathcal{X}$, the sequence $\{y_\ell(x^k, \mu_k)\}$ is bounded for each ℓ.

Lemma 6.2 *Suppose that all M_ℓ, $\ell = 1, \cdots, L$, are P_0-matrices in problem (6.7) and $(x^*, y_1^*, \cdots, y_L^*) \in \mathcal{F}_1$. Assume that, for each ℓ, the submatrix $M_\ell[\mathcal{K}_\ell^*]$ is nondegenerate, where*

$$\mathcal{K}_\ell^* \ = \ \{ \ i \ | \ \ (N_\ell x^* + M_\ell y_\ell^* + q_\ell)[i] = 0 \}.$$

Then there exist a neighborhood U^* of $(x^*, y_1^*, \cdots, y_L^*)$ and a positive constant π^* such that

$$\|y_\ell(x, \mu_k) - y_\ell\| \le \mu_k \pi^* (\|y_\ell\| + \sqrt{m}), \quad \ell = 1, \cdots, L \tag{6.22}$$

holds for any $(x, y_1, \cdots, y_L) \in U^* \cap \mathcal{F}_1$ and every k.

Proof: For any $(x, y_1, \cdots, y_L) \in \Re^{n+mL}$, we let

$$w_\ell = N_\ell x + M_\ell y_\ell + q_\ell, \quad \ell = 1, \cdots, L.$$

Since $(x, y_1, \cdots, y_L) \in \mathcal{F}_1$ implies

$$\Phi_0(x, y_\ell, w_\ell; N_\ell, M_\ell, q_\ell) = 0, \quad \ell = 1, \cdots, L, \tag{6.23}$$

(6.16) and (6.23) yield

$$\begin{aligned} 0 \ &= \ \Phi_{\mu_k}(x, y_\ell(x, \mu_k), w_\ell(x, \mu_k); N_\ell, M_\ell, q_\ell) - \Phi_0(x, y_\ell, w_\ell; N_\ell, M_\ell, q_\ell) \\ &= \ \begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell(x, \mu_k) & D_\ell(x, \mu_k) \end{pmatrix} \begin{pmatrix} y_\ell(x, \mu_k) - y_\ell \\ w_\ell(x, \mu_k) - w_\ell \end{pmatrix} - \mu_k \begin{pmatrix} -y_\ell \\ 2a_\ell^k[1]\mu_k \\ \vdots \\ 2a_\ell^k[m]\mu_k \end{pmatrix}. \end{aligned} \tag{6.24}$$

Here, $D_\ell(x, \mu_k) = \mathrm{diag}\Big(a_\ell^k[i](y_\ell(x, \mu_k)[i] + y_\ell[i])\Big)$ and

$$\begin{aligned} a_\ell^k[i] \ &= \ \frac{1}{\sqrt{(y_\ell(x, \mu_k)[i])^2 + (w_\ell(x, \mu_k)[i])^2 + 2\mu_k^2} + \sqrt{(y_\ell[i])^2 + (w_\ell[i])^2}} \\ &= \ \frac{1}{y_\ell(x, \mu_k)[i] + w_\ell(x, \mu_k)[i] + y_\ell[i] + w_\ell[i]}, \end{aligned} \tag{6.25}$$

where the last equality follows from (6.16) and (6.23). From (6.13), we see that, for any ℓ, i, and any k,

$$0 < a_\ell^k[i](y_\ell(x, \mu_k)[i] + y_\ell[i]) < 1. \tag{6.26}$$

Note that the matrix

$$\begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell(x, \mu_k) & D_\ell(x, \mu_k) \end{pmatrix}$$

is nonsingular [29]. We next prove that there exists a neighborhood U^* of $(x^*, y_1^*, \cdots, y_L^*)$ and a positive constant π^* such that

$$\left\| \begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell(x, \mu_k) & D_\ell(x, \mu_k) \end{pmatrix}^{-1} \right\| \leq \pi^*, \quad \ell = 1, \cdots, L \tag{6.27}$$

holds for any $(x, y_1, \cdots, y_L) \in U^* \cap \mathcal{F}_1$ and any k. Otherwise, there must be an index ℓ, a subsequence $\{k_j\}$ of $\{k\}$, and a sequence $\{(x^j, y_1^j, \cdots, y_L^j)\} \subset \mathcal{F}_1$ such that

$$\lim_{j \to \infty} (x^j, y_1^j, \cdots, y_L^j) = (x^*, y_1^*, \cdots, y_L^*) \tag{6.28}$$

and

$$\lim_{j \to \infty} \left\| \begin{pmatrix} M_\ell + \mu_{k_j} I & -I \\ I - D_\ell(x^j, \mu_{k_j}) & D_\ell(x^j, \mu_{k_j}) \end{pmatrix}^{-1} \right\| = +\infty. \tag{6.29}$$

By (6.26), the sequence $\{D_\ell(x^j, \mu_{k_j})\}$ is bounded. Hence, passing to a further subsequence if necessary, we may assume that

$$\lim_{j \to \infty} D_\ell(x^j, \mu_{k_j}) = \hat{D}_\ell := \operatorname{diag}(\hat{d}_\ell[1], \cdots, \hat{d}_\ell[m]).$$

Note that, by (6.25),

$$\hat{d}_\ell[i] = \lim_{j \to \infty} \frac{y_\ell(x^j, \mu_{k_j})[i] + y_\ell^j[i]}{y_\ell(x^j, \mu_{k_j})[i] + w_\ell(x^j, \mu_{k_j})[i] + y_\ell^j[i] + w_\ell^j[i]}.$$

By Assumption A2, the sequence $\{y_\ell(x^j, \mu_{k_j})\}$ is bounded and then so is the sequence $\{w_\ell(x^j, \mu_{k_j})\}$. If $\hat{d}_\ell[i] = 1$, we have

$$\lim_{j \to \infty} \frac{w_\ell(x^j, \mu_{k_j})[i] + w_\ell^j[i]}{y_\ell(x^j, \mu_{k_j})[i] + w_\ell(x^j, \mu_{k_j})[i] + y_\ell^j[i] + w_\ell^j[i]} = 0. \tag{6.30}$$

We claim that

$$\lim_{j \to \infty} \Big(w_\ell(x^j, \mu_{k_j})[i] + w_\ell^j[i] \Big) = 0. \tag{6.31}$$

In fact, suppose that (6.31) does not hold and, without loss of generality, assume

$$\lim_{j \to \infty} \Big(w_\ell(x^j, \mu_{k_j})[i] + w_\ell^j[i] \Big) = \tau > 0.$$

Since all numbers involved are nonnegative and bounded, it follows that

$$\lim_{j \to \infty} \frac{w_\ell(x^j, \mu_{k_j})[i] + w_\ell^j[i]}{y_\ell(x^j, \mu_{k_j})[i] + w_\ell(x^j, \mu_{k_j})[i] + y_\ell^j[i] + w_\ell^j[i]} \neq 0,$$

which contradicts (6.30). Therefore, we must have (6.31) and hence

$$w_\ell^*[i] = \lim_{j\to\infty} w_\ell^j[i] = 0.$$

This implies that

$$\mathcal{K}_\ell^* \ \supseteq \ \hat{\mathcal{K}}_\ell = \{\ i \mid \ \hat{d}_\ell[i] = 1\}.$$

Taking into account the assumption that the submatrix $M_\ell[\mathcal{K}_\ell^*]$ is nondegenerate, we deduce that the submatrix $M_\ell[\hat{\mathcal{K}}_\ell]$ is nonsingular and then, by Lemma 6.1, the limit matrix

$$\begin{pmatrix} M_\ell & -I \\ I - \hat{D}_\ell & \hat{D}_\ell \end{pmatrix}$$

is nonsingular. We therefore have

$$\lim_{j\to\infty} \begin{pmatrix} M_\ell + \mu_{k_j} I & -I \\ I - D_\ell(x^j, \mu_{k_j}) & D_\ell(x^j, \mu_{k_j}) \end{pmatrix}^{-1} = \begin{pmatrix} M_\ell & -I \\ I - \hat{D}_\ell & \hat{D}_\ell \end{pmatrix}^{-1}.$$

However, this contradicts (6.29) and hence we obtain (6.27). In addition, we have from (6.13) that, for every k, $\ell = 1, \cdots, L$, and $i = 1, \cdots, m$,

$$\begin{aligned} 2a_\ell^k[i]\mu_k &\le \frac{2\mu_k}{\sqrt{(y_\ell(x,\mu_k)[i])^2 + (w_\ell(x,\mu_k)[i])^2 + 2\mu_k^2}} \\ &\le \frac{2\mu_k}{\sqrt{2y_\ell(x,\mu_k)[i]w_\ell(x,\mu_k)[i] + 2\mu_k^2}} \\ &= 1. \end{aligned} \tag{6.32}$$

Thus, it follows from (6.24), (6.27), and (6.32) that

$$\begin{aligned} \|y_\ell(x,\mu_k) - y_\ell\| &\le \left\| \begin{pmatrix} y_\ell(x,\mu_k) - y_\ell \\ w_\ell(x,\mu_k) - w_\ell \end{pmatrix} \right\| \\ &= \mu_k \left\| \begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell(x,\mu_k) & D_\ell(x,\mu_k) \end{pmatrix}^{-1} \begin{pmatrix} -y_\ell \\ 2a_\ell^k[1]\mu_k \\ \vdots \\ 2a_\ell^k[m]\mu_k \end{pmatrix} \right\| \\ &\le \mu_k \pi^*(\|y_\ell\| + \sqrt{m}). \end{aligned} \tag{6.33}$$

Note that the above inequalities hold for all ℓ and then the proof is completed. ∎

Under Assumptions A1 and A2, the sequence $\left\{(x^{(k)}, y_1^{(k)}, \cdots, y_L^{(k)})\right\}$ generated by SIP-I is bounded. In the rest of this subsection, we further make the following assumptions:

A3: The sequence $\left\{(x^{(k)}, y_1^{(k)}, \cdots, y_L^{(k)})\right\}$ generated by SIP-I is convergent to a point $(x^*, y_1^*, \cdots, y_L^*)$.

A4: There exists a neighborhood V^* of $(x^*, y_1^*, \cdots, y_L^*)$ such that $x^{(k)}$ minimizes θ_{μ_k} over $V^*|_{\mathcal{X}}$ for all k large enough, where $V^*|_{\mathcal{X}} = \{x \in \mathcal{X} \mid (x, y_1, \cdots, y_L) \in U^*\}$.

Theorem 6.2 *Suppose that all matrices M_ℓ, $\ell = 1, \cdots, L$, are P_0-matrices in problem (6.7). Let $\left\{(x^{(k)}, y_1^{(k)}, \cdots, y_L^{(k)})\right\}$ and $(x^*, y_1^*, \cdots, y_L^*)$ be a sequence generated by SIP-I and its limit point, respectively. If the submatrix $M_\ell[\mathcal{K}_\ell^*]$ is nondegenerate for each ℓ, where $\mathcal{K}_\ell^*$ is the same as in Lemma 6.2, then $(x^*, y_1^*, \cdots, y_L^*)$ is a local optimal solution of problem (6.7).*

Proof: First we have $(x^*, y_1^*, \cdots, y_L^*) \in \mathcal{F}_1$ immediately from (6.15). By Lemma 6.2, there exist a neighborhood $U^* \subseteq V^*$ of the point $(x^*, y_1^*, \cdots, y_L^*)$ and a positive number π^* such that (6.22) holds for any $(x, y_1, \cdots, y_L) \in U^* \cap \mathcal{F}_1$ and every k.

Choose a positive number η and let

$$\mathcal{F}_{1,\eta} = \left\{(x, y_1, \cdots, y_L) \in U^* \cap \mathcal{F}_1 \;\middle|\; \|(x, y_1, \cdots, y_L) - (x^*, y_1^*, \cdots, y_L^*)\| \le \eta\right\}.$$

Since $\mathcal{F}_{1,\eta}$ is a nonempty compact set, the continuity of f ensures that the problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{\ell=1}^{L} p_\ell f(x, y_\ell) \\ \text{subject to} & (x, y_1, \cdots, y_L) \in \mathcal{F}_{1,\eta} \end{array} \tag{6.34}$$

has a nonempty solution set. Let $(\bar{x}, \bar{y}_1, \cdots, \bar{y}_L)$ be an arbitrary solution of (6.34).

For any $(x, y_1, \cdots, y_L) \in \mathcal{F}_{1,\eta}$, by the mean-value theorem, we have

$$\begin{aligned} \theta_{\mu_k}(x) &= \sum_{\ell=1}^{L} p_\ell f(x, y_\ell(x, \mu_k)) \\ &= \sum_{\ell=1}^{L} p_\ell \Big(f(x, y_\ell) + (y_\ell(x, \mu_k) - y_\ell)^T \nabla_y f(x, \alpha_\ell y_\ell + (1 - \alpha_\ell) y_\ell(x, \mu_k))\Big), \end{aligned} \tag{6.35}$$

where $\alpha_\ell \in [0, 1]$ for each ℓ. By (6.22), we have

$$\begin{aligned} \|\alpha_\ell y_\ell + (1 - \alpha_\ell) y_\ell(x, \mu_k)\| &= \|(1 - \alpha_\ell)(y_\ell(x, \mu_k) - y_\ell) + y_\ell\| \\ &\le \|y_\ell(x, \mu_k) - y_\ell\| + \|y_\ell\| \\ &\le \mu_k \pi^* (\|y_\ell\| + \sqrt{m}) + \|y_\ell\|. \end{aligned} \tag{6.36}$$

Since $\mathcal{F}_{1,\eta}$ is bounded, it follows from (6.36) that the set

$$\left\{(x, \alpha_\ell y_\ell + (1 - \alpha_\ell) y_\ell(x, \mu_k)) \;\middle|\; (x, y_1, \cdots, y_L) \in \mathcal{F}_{1,\eta}, k = 1, 2, \cdots, \alpha_\ell \in [0, 1]\right\}$$

is bounded for each $\ell = 1, \cdots, L$, and so there exists a constant $\tau > 0$ such that

$$\|\nabla_y f(x, \alpha_\ell y_\ell + (1-\alpha_\ell) y_\ell(x, \mu_k))\| \le \tau, \quad \ell = 1, \cdots, L$$

holds for any $(x, y_1, \cdots, y_L) \in \mathcal{F}_{1,\eta}, \alpha_\ell \in [0,1]$, and every k. It then follows from (6.35) and (6.22) that

$$\Big|\theta_{\mu_k}(x) - \sum_{\ell=1}^{L} p_\ell f(x, y_\ell)\Big| \le \tau \sum_{\ell=1}^{L} \|y_\ell(x, \mu_k) - y_\ell\| \le \tau \mu_k \pi^* \sum_{\ell=1}^{L} \Big(\|y_\ell\| + \sqrt{m}\Big)$$

for any $(x, y_1, \cdots, y_L) \in \mathcal{F}_{1,\eta}$ and k. In particular,

$$\Big|\theta_{\mu_k}(\bar{x}) - \sum_{\ell=1}^{L} p_\ell f(\bar{x}, \bar{y}_\ell)\Big| \le \tau \mu_k \pi^* \sum_{\ell=1}^{L} \Big(\|\bar{y}_\ell\| + \sqrt{m}\Big), \quad \forall k. \tag{6.37}$$

On the one hand, the continuity of f yields

$$\lim_{k\to\infty} \theta_{\mu_k}(x^{(k)}) = \sum_{\ell=1}^{L} p_\ell \lim_{k\to\infty} f(x^{(k)}, y_\ell^{(k)}) = \sum_{\ell=1}^{L} p_\ell f(x^*, y_\ell^*). \tag{6.38}$$

Note that, by Assumption A4 and the fact that $U^* \subseteq V^*$, $x^{(k)}$ is a global optimal solution of problem (6.19) when k is large enough, and $\bar{x}$ is a feasible point of (6.19). We then have from (6.37) that, for every sufficiently large k,

$$\theta_{\mu_k}(x^{(k)}) \le \theta_{\mu_k}(\bar{x}) \le \sum_{\ell=1}^{L} p_\ell f(\bar{x}, \bar{y}_\ell) + \tau \mu_k \pi^* \sum_{\ell=1}^{L} \Big(\|\bar{y}_\ell\| + \sqrt{m}\Big). \tag{6.39}$$

Therefore, taking into account the equality (6.38) and the fact that $\mu_k \to 0$ as $k \to \infty$, we have by letting $k \to \infty$ in (6.39) that

$$\sum_{\ell=1}^{L} p_\ell f(x^*, y_\ell^*) \le \sum_{\ell=1}^{L} p_\ell f(\bar{x}, \bar{y}_\ell).$$

On the other hand, since $(\bar{x}, \bar{y}_1, \cdots, \bar{y}_L)$ is a solution of problem (6.34), we have

$$\sum_{\ell=1}^{L} p_\ell f(x^*, y_\ell^*) \ge \sum_{\ell=1}^{L} p_\ell f(\bar{x}, \bar{y}_\ell).$$

It then follows that

$$\sum_{\ell=1}^{L} p_\ell f(x^*, y_\ell^*) = \sum_{\ell=1}^{L} p_\ell f(\bar{x}, \bar{y}_\ell).$$

This means that $(x^*, y_1^*, \cdots, y_L^*)$ is a global optimal solution of problem (6.34), in other words, $(x^*, y_1^*, \cdots, y_L^*)$ is a local optimal solution of problem (6.7). This completes the proof. ∎

Theorem 6.3 *Suppose that all M_ℓ, $\ell = 1, \cdots, L$, are P-matrices in problem (6.7) and, for each k, $(x^{(k)}, y_1^{(k)}, \cdots, y_L^{(k)})$ is a global optimal solution of problem (6.19). Then the limit point $(x^*, y_1^*, \cdots, y_L^*)$ is a global optimal solution of problem (6.7).*

Proof: Recall that $(x^*, y_1^*, \cdots, y_L^*) \in \mathcal{F}_1$. Since all M_ℓ are P-matrices (and hence nondegenerate), the conditions of Lemma 6.2 are satisfied at every point $(x, y_1, \cdots, y_L) \in \mathcal{F}_1$. Thus, for any point $(\hat{x}, \hat{y}_1, \cdots, \hat{y}_L) \in \mathcal{F}_1$, there exist a neighborhood $\hat{U} = U(\hat{x}, \hat{y}_1, \cdots, \hat{y}_L)$ of $(\hat{x}, \hat{y}_1, \cdots, \hat{y}_L)$ and a positive constant $\hat{\pi} = \pi(\hat{x}, \hat{y}_1, \cdots, \hat{y}_L)$ such that for any $(x, y_1, \cdots, y_L) \in \hat{U} \cap \mathcal{F}_1$,

$$\|y_\ell(x, \mu_k) - y_\ell\| \le \mu_k \hat{\pi}(\|y_\ell\| + \sqrt{m})$$

holds for each ℓ and k.

For an arbitrary positive number η, we define the set $\mathcal{F}_{1,\eta}$ by

$$\mathcal{F}_{1,\eta} = \Big\{(x, y_1, \cdots, y_L) \in \mathcal{F}_1 \;\Big|\; \|(x, y_1, \cdots, y_L) - (x^*, y_1^*, \cdots, y_L^*)\| \le \eta\Big\}. \tag{6.40}$$

It is clear that $\mathcal{F}_{1,\eta}$ is a nonempty compact set. Since the set of neighborhoods

$$\mathcal{U} = \Big\{U(\hat{x}, \hat{y}_1, \cdots, \hat{y}_L) \;\Big|\; (\hat{x}, \hat{y}_1, \cdots, \hat{y}_L) \in \mathcal{F}_{1,\eta}\Big\}$$

is an open covering of $\mathcal{F}_{1,\eta}$, there is a finite number of neighborhoods, say $U_1, U_2, \cdots, U_s$, in $\mathcal{U}$ such that $\{U_1, U_2, \cdots, U_s\}$ constitutes a covering of $\mathcal{F}_{1,\eta}$. By Lemma 6.2, there exist constants $\pi_1, \cdots, \pi_s$ corresponding to the sets $U_1, U_2, \cdots, U_s$. Then, by setting

$$\pi^* = \max\{\pi_1, \pi_2, \cdots, \pi_s\},$$

we have (6.22) for any $(x, y_1, \cdots, y_L) \in \mathcal{F}_{1,\eta}$ and every k.

Consider problem (6.34) with $\mathcal{F}_{1,\eta}$ defined by (6.40). In a similar way to the proof of Theorem 6.2, we can show that $(x^*, y_1^*, \cdots, y_L^*)$ is a global optimal solution of (6.34). Noting that $\eta > 0$ is arbitrary, we see that $(x^*, y_1^*, \cdots, y_L^*)$ is actually a global optimal solution of problem (6.7) and so the proof is completed. ∎

From a computational viewpoint, it is in general difficult or even impossible to get an exact optimal solution of an optimization problem. For this reason, the following result is more interesting in practice.

Theorem 6.4 *Suppose that $\{\epsilon_k\} \subseteq (0, +\infty)$ is convergent to 0 and, for every k, $x^{(k)} \in \mathcal{X}$ is an approximate local (or global) optimal solution of the problem (6.19) satisfying*

$$\theta_{\mu_k}(x^{(k)}) - \epsilon_k \le \theta_{\mu_k}(x), \quad \forall x \in \mathcal{X} \cap U(x^{(k)}) \text{ (or } x \in \mathcal{X}),$$

where $U(x^{(k)})$ is a neighborhood of the point $x^{(k)}$. Let $y_\ell^{(k)} = y_\ell(x^{(k)}, \mu_k)$ for each $\ell = 1, \cdots, L$. Then, under the same conditions as Theorem 6.2 or Theorem 6.3, the corresponding conclusion remains true.

6.2.4 Limiting behavior of stationary points

In the last subsection, we have discussed the convergence of optimal solutions of problems (6.19). In practice, it may not be easy to obtain an optimal solution, even an approximate optimal solution, whereas computation of stationary points may be relatively easy. Therefore, it is necessary to study the limiting behavior of stationary points of subproblems (6.19).

For problem (6.19), we will use the standard definition of stationarity. Moreover, recalling that problem (6.7) is equivalent to an ordinary MPEC (6.10), we employ the same terminologies for (6.7) as in the study of MPECs. Suppose $(x^*, \mathbf{y}^*)$ is a feasible point of (6.10).

Definition 6.3 Assume that the MPEC-LICQ holds at $(x^*, \mathbf{y}^*)$ in problem (6.10). We say $(x^*, \mathbf{y}^*)$ is a C-stationary point if there exist multiplier vectors u^*, v^*, λ^*, and γ^* such that

$$\nabla \mathrm{f}(x^*, \mathbf{y}^*) - \sum_{i \in \mathcal{I}_G(x^*, \mathbf{y}^*)} u^*[i] \nabla G_i(x^*, \mathbf{y}^*) - \sum_{i \in \mathcal{I}_F(x^*, \mathbf{y}^*)} v^*[i] \nabla F_i(x^*, \mathbf{y}^*)$$
$$+ \sum_{i \in \mathcal{I}_g^*} \lambda^*[i] \begin{pmatrix} \nabla g_i(x^*) \\ 0 \end{pmatrix} + \sum_{i=1}^{s_2} \gamma^*[i] \begin{pmatrix} \nabla h_i(x^*) \\ 0 \end{pmatrix} = 0, \qquad (6.41)$$
$$\lambda^* \geq 0, \qquad u^*[i] v^*[i] \geq 0 \quad \text{for} \quad i \in \mathcal{I}_G(x^*, \mathbf{y}^*) \cap \mathcal{I}_F(x^*, \mathbf{y}^*). \qquad (6.42)$$

If furthermore, there holds

$$u^*[i] \geq 0, \quad v^*[i] \geq 0, \qquad \forall i \in \mathcal{I}_G(x^*, \mathbf{y}^*) \cap \mathcal{I}_F(x^*, \mathbf{y}^*), \qquad (6.43)$$

we call $(x^*, \mathbf{y}^*)$ a B-stationary point.

Definition 6.4 A solution y_ℓ^* of problem LCP$(x^*; N_\ell, M_\ell, q_\ell)$ is said to satisfy the *strict complementarity condition* if $\mathcal{I}_{Y_\ell}^* \cap \mathcal{I}_{W_\ell}^* = \emptyset$, where

$$\begin{aligned} \mathcal{I}_{Y_\ell}^* &= \{i \mid y_\ell^*[i] = 0\}, \\ \mathcal{I}_{W_\ell}^* &= \{i \mid (N_\ell x^* + M_\ell y_\ell^* + q_\ell)[i] = 0\}. \end{aligned}$$

Theorem 6.5 *Suppose that all M_ℓ, $\ell = 1, \cdots, L$, are P_0-matrices in problem (6.7) and for each k, $x^{(k)}$ is a stationary point of (6.19). Let $(x^*, y_1^*, \cdots, y_L^*)$ be an accumulation point of the sequence $\left\{(x^{(k)}, y_1^{(k)}, \cdots, y_L^{(k)})\right\}$ generated by SIP-I. If the MPEC-LICQ* *is satisfied at* $(x^*, y_1^*, \cdots, y_L^*)$*, then* $(x^*, y_1^*, \cdots, y_L^*)$ *is a C-stationary point of problem (6.7). In particular, if* y_ℓ^* *satisfies the strict complementarity condition for each* ℓ*, then* $(x^*, y_1^*, \cdots, y_L^*)$ *is B-stationary.*

Proof: Assume without loss of generality that the sequence $\left\{(x^{(k)}, y_1^{(k)}, \cdots, y_L^{(k)})\right\}$ converges to $(x^*, y_1^*, \cdots, y_L^*)$. Since the MPEC-LICQ holds at $(x^*, y_1^*, \cdots, y_L^*)$, it is obvious that, for every k sufficiently large, problem (6.19) satisfies the standard LICQ at $x^{(k)}$ and then, by the stationarity of $x^{(k)}$, there exist unique Lagrange multiplier vectors λ^k and γ^k such that

$$\nabla\theta_{\mu_k}(x^{(k)}) + \nabla g(x^{(k)})\lambda^k + \nabla h(x^{(k)})\gamma^k = 0, \tag{6.44}$$

$$g(x^{(k)}) \le 0, \qquad h(x^{(k)}) = 0, \tag{6.45}$$

$$\lambda^k \ge 0, \qquad g(x^{(k)})^T\lambda^k = 0. \tag{6.46}$$

In the remainder of the proof, we suppose k is large enough so that (6.44)–(6.46) hold and, in addition,

$$\mathcal{I}_g(x^{(k)}) \subseteq \mathcal{I}_g^*, \tag{6.47}$$

which follows from the continuity of g.

Note that $\Phi_{\mu_k}(x, y_\ell(x,\mu_k), w_\ell(x,\mu_k); N_\ell, M_\ell, q_\ell) = 0$ is satisfied for each ℓ. By the implicit function theorem [64], we have

$$\begin{aligned}
&\begin{pmatrix} \nabla y_\ell(x^{(k)},\mu_k)^T \\ \nabla w_\ell(x^{(k)},\mu_k)^T \end{pmatrix} \\
= \; & -\begin{pmatrix} \nabla_{y_\ell}\Phi_{\mu_k}(x^{(k)}, y^{(k)}, w^{(k)}; N_\ell, M_\ell, q_\ell) \\ \nabla_{w_\ell}\Phi_{\mu_k}(x^{(k)}, y^{(k)}, w^{(k)}; N_\ell, M_\ell, q_\ell) \end{pmatrix}^{-T} \nabla_x\Phi_{\mu_k}(x^{(k)}, y^{(k)}, w^{(k)}; N_\ell, M_\ell, q_\ell)^T \\
= \; & -\begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell^k & D_\ell^k \end{pmatrix}^{-1} \begin{pmatrix} N_\ell \\ O \end{pmatrix},
\end{aligned} \tag{6.48}$$

where $D_\ell^k = \operatorname{diag}\left(\frac{y_\ell^{(k)}[1]}{y_\ell^{(k)}[1]+w_\ell^{(k)}[1]}, \cdots, \frac{y_\ell^{(k)}[m]}{y_\ell^{(k)}[m]+w_\ell^{(k)}[m]}\right)$ and the existence of the inverse matrix follows from Theorem 6.1. Furthermore, since

$$\begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell^k & D_\ell^k \end{pmatrix}^{-1} = \begin{pmatrix} E_\ell^k D_\ell^k & E_\ell^k \\ -I + (M_\ell + \mu_k I)E_\ell^k D_\ell^k & (M_\ell + \mu_k I)E_\ell^k \end{pmatrix} \tag{6.49}$$

with

$$E_\ell^k = \left(D_\ell^k M_\ell + I - (1-\mu_k)D_\ell^k\right)^{-1} \tag{6.50}$$

(see [29]), it follows from (6.48) that

$$\nabla y_\ell(x^{(k)}, \mu_k) = -N_\ell^T D_\ell^k (E_\ell^k)^T, \quad \ell = 1, \cdots, L. \tag{6.51}$$

Thus, (6.44) becomes

$$\begin{aligned}
0 &= \nabla\theta_{\mu_k}(x^{(k)}) + \nabla g(x^{(k)})\lambda^k + \nabla h(x^{(k)})\gamma^k \\
&= \sum_{\ell=1}^{L} p_\ell \Big(\nabla_x f(x^{(k)}, y_\ell^{(k)}) + \nabla y_\ell(x^{(k)}, \mu_k) \nabla_y f(x^{(k)}, y_\ell^{(k)}) \Big) \\
&\qquad + \nabla g(x^{(k)})\lambda^k + \nabla h(x^{(k)})\gamma^k \\
&= \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y_\ell^{(k)}) - \sum_{\ell=1}^{L} p_\ell N_\ell^T D_\ell^k (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}) \\
&\qquad + \nabla g(x^{(k)})\lambda^k + \nabla h(x^{(k)})\gamma^k \\
&= \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y_\ell^{(k)}) - \sum_{\ell=1}^{L} N_\ell^T v_\ell^k + \nabla g(x^{(k)})\lambda^k + \nabla h(x^{(k)})\gamma^k,
\end{aligned} \tag{6.52}$$

where v_ℓ^k is defined by

$$v_\ell^k = p_\ell D_\ell^k (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}), \tag{6.53}$$

that is,

$$v_\ell^k[i] = \frac{p_\ell \; y_\ell^{(k)}[i]}{y_\ell^{(k)}[i] + w_\ell^{(k)}[i]} \; e_i^T (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}), \quad i = 1, \cdots, m. \tag{6.54}$$

For each ℓ, we let

$$u_\ell^k = p_\ell \nabla_y f(x^{(k)}, y_\ell^{(k)}) - M_\ell^T v_\ell^k,$$

that is,

$$p_\ell \nabla_y f(x^{(k)}, y_\ell^{(k)}) - u_\ell^k - M_\ell^T v_\ell^k = 0. \tag{6.55}$$

It then follows from (6.53) and (6.50) that, for every $\ell = 1, \cdots, L$,

$$\begin{aligned}
u_\ell^k &= p_\ell \nabla_y f(x^{(k)}, y_\ell^{(k)}) - M_\ell^T v_\ell^k \\
&= p_\ell \nabla_y f(x^{(k)}, y_\ell^{(k)}) - p_\ell M_\ell^T D_\ell^k (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}) \\
&= p_\ell \Big((E_\ell^k)^{-T} - M_\ell^T D_\ell^k \Big) (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}) \\
&= p_\ell \Big(I - (1 - \mu_k) D_\ell^k \Big) (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)})
\end{aligned}$$

and hence

$$u_\ell^k[i] = \frac{p_\ell(\mu_k y_\ell^{(k)}[i] + w_\ell^{(k)}[i])}{y_\ell^{(k)}[i] + w_\ell^{(k)}[i]} \; e_i^T (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}), \quad i = 1, \cdots, m. \tag{6.56}$$

Taking into account (6.45)–(6.47), we obtain from (6.52) and (6.55) that

$$0 = \begin{pmatrix} \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y_\ell^{(k)}) \\ p_1 \nabla_y f(x^{(k)}, y_1^{(k)}) \\ \vdots \\ p_L \nabla_y f(x^{(k)}, y_L^{(k)}) \end{pmatrix} - \begin{pmatrix} O & \cdots & O \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix} u^k - \begin{pmatrix} N_1^T & \cdots & N_L^T \\ M_1^T & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & M_L^T \end{pmatrix} v^k$$
$$+ \sum_{i \in \mathcal{I}_g^*} \lambda^k[i] \begin{pmatrix} \nabla g_i(x^{(k)}) \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \gamma^k, \tag{6.57}$$

where

$$u^k = \begin{pmatrix} u_1^k \\ \vdots \\ u_L^k \end{pmatrix}, \qquad v^k = \begin{pmatrix} v_1^k \\ \vdots \\ v_L^k \end{pmatrix}.$$

We can further rewrite (6.57) as

$$\begin{pmatrix} \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y_\ell^{(k)}) \\ p_1 \nabla_y f(x^{(k)}, y_1^{(k)}) \\ \vdots \\ p_L \nabla_y f(x^{(k)}, y_L^{(k)}) \end{pmatrix} - \sum_{\ell=1}^{L} \sum_{i \notin \mathcal{I}_{Y_\ell}^*} u_\ell^k[i] \begin{pmatrix} 0 \\ \vdots \\ e_i \\ \vdots \\ 0 \end{pmatrix} - \sum_{\ell=1}^{L} \sum_{i \notin \mathcal{I}_{W_\ell}^*} v_\ell^k[i] \begin{pmatrix} N_\ell[i] \\ \vdots \\ M_\ell[i] \\ \vdots \\ 0 \end{pmatrix}$$
$$= \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}_{Y_\ell}^*} u_\ell^k[i] \begin{pmatrix} 0 \\ \vdots \\ e_i \\ \vdots \\ 0 \end{pmatrix} + \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}_{W_\ell}^*} v_\ell^k[i] \begin{pmatrix} N_\ell[i] \\ \vdots \\ M_\ell[i] \\ \vdots \\ 0 \end{pmatrix}$$
$$- \sum_{i \in \mathcal{I}_g^*} \lambda^k[i] \begin{pmatrix} \nabla g_i(x^{(k)}) \\ 0 \\ \vdots \\ 0 \end{pmatrix} - \sum_{i=1}^{s_2} \gamma^k[i] \begin{pmatrix} \nabla h_i(x^{(k)}) \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{6.58}$$

We next prove that, for each ℓ,

$$i \notin \mathcal{I}_{Y_\ell}^* \quad \Rightarrow \quad \lim_{k \to \infty} u_\ell^k[i] = 0, \tag{6.59}$$
$$i \notin \mathcal{I}_{W_\ell}^* \quad \Rightarrow \quad \lim_{k \to \infty} v_\ell^k[i] = 0. \tag{6.60}$$

To this end, it is enough to show that $\{\|(E_\ell^k)^T\nabla_y f(x^{(k)}, y_\ell^{(k)})\|\}$ is bounded for every ℓ. Otherwise, taking a further subsequence if necessary, there is an index $\hat{\ell}$ satisfying

$$\left\|(E_{\hat{\ell}}^k)^T\nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\right\| = \max_{1\le\ell\le L}\left\|(E_\ell^k)^T\nabla_y f(x^{(k)}, y_\ell^{(k)})\right\|, \quad \forall k \tag{6.61}$$

and

$$\lim_{k\to\infty}\left\|(E_{\hat{\ell}}^k)^T\nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\right\| = +\infty. \tag{6.62}$$

Note that, from (6.56) and (6.61), we have

$$\begin{aligned} |u_\ell^k[i]| &\le \frac{p_\ell(\mu_k y_\ell^{(k)}[i] + w_\ell^{(k)}[i])}{y_\ell^{(k)}[i] + w_\ell^{(k)}[i]}\|(E_\ell^k)^T\nabla_y f(x^{(k)}, y_\ell^{(k)})\| \\ &\le \frac{p_\ell(\mu_k y_\ell^{(k)}[i] + w_\ell^{(k)}[i])}{y_\ell^{(k)}[i] + w_\ell^{(k)}[i]}\|(E_{\hat{\ell}}^k)^T\nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\| \end{aligned}$$

for each ℓ, i, and k. In consequence, for each ℓ, we have

$$i \notin \mathcal{I}_{Y_\ell}^* \quad \Rightarrow \quad \frac{\mu_k y_\ell^{(k)}[i] + w_\ell^{(k)}[i]}{y_\ell^{(k)}[i] + w_\ell^{(k)}[i]} \to 0 \quad \Rightarrow \quad \frac{|u_\ell^k[i]|}{\|(E_{\hat{\ell}}^k)^T\nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|} \to 0. \tag{6.63}$$

Similarly, we can show that, for all $\ell = 1, \cdots, L$,

$$i \notin \mathcal{I}_{W_\ell}^* \quad \Rightarrow \quad \frac{|v_\ell^k[i]|}{\|(E_{\hat{\ell}}^k)^T\nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|} \to 0. \tag{6.64}$$

Let d^k denote the vector on the right-hand side of equality (6.58). It then follows from (6.58) and (6.62)–(6.64) that

$$\lim_{k\to\infty}\frac{d^k}{\|(E_{\hat{\ell}}^k)^T\nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|} = 0. \tag{6.65}$$

Since the MPEC-LICQ holds at $(x^*, y_1^*, \cdots, y_L^*)$, the vectors on the right-hand side of (6.58) are linearly independent when k is sufficiently large and so, by (6.65), all the sequences generated by dividing the multipliers that appear on the right-hand side of (6.58) by the number $\|(E_{\hat{\ell}}^k)^T\nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|$ are convergent to 0 as $k\to\infty$. This fact, together with (6.63) and (6.64), implies that, for any i,

$$\lim_{k\to\infty}\frac{u_{\hat{\ell}}^k[i]}{\|(E_{\hat{\ell}}^k)^T\nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|} = 0, \tag{6.66}$$

$$\lim_{k\to\infty}\frac{v_{\hat{\ell}}^k[i]}{\|(E_{\hat{\ell}}^k)^T\nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|} = 0. \tag{6.67}$$

However, noticing that

$$\frac{u_{\hat{\ell}}^k[i] + v_{\hat{\ell}}^k[i]}{\|(E_{\hat{\ell}}^k)^T \nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|} = p_{\hat{\ell}}\Big(1 + \frac{\mu_k y_{\hat{\ell}}^{(k)}[i]}{y_{\hat{\ell}}^{(k)}[i] + w_{\hat{\ell}}^{(k)}[i]}\Big)\frac{(E_{\hat{\ell}}^k)^T \nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})[i]}{\|(E_{\hat{\ell}}^k)^T \nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|}$$

holds for any i and k, there exists an index $\hat{i}$ such that

$$\lim_{k\to\infty} \frac{|u_{\hat{\ell}}^k[\hat{i}] + u_{\hat{\ell}}^k[\hat{i}]|}{\|(E_{\hat{\ell}}^k)^T \nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|} \geq \frac{1}{\sqrt{m}} \lim_{k\to\infty} p_{\hat{\ell}}\Big(1 + \frac{\mu_k y_{\hat{\ell}}^{(k)}[\hat{i}]}{y_{\hat{\ell}}^{(k)}[\hat{i}] + w_{\hat{\ell}}^{(k)}[\hat{i}]}\Big) = \frac{p_{\hat{\ell}}}{\sqrt{m}} > 0.$$

This contradicts (6.66) and (6.67). As a result, the implications (6.59) and (6.60) are true.

Consider equality (6.58) again. By (6.59) and (6.60), the left-hand side of (6.58) is convergent as $k \to \infty$. Recall that the vectors on the right-hand side of (6.58) are linearly independent when k is sufficiently large. These facts imply that all the sequences of the multipliers that appear on the right-hand side of (6.58) are convergent. In consequence, by letting $k \to \infty$ in (6.58), we obtain the equality corresponding to (6.41).

In addition, since both $y_\ell^{(k)}[i]$ and $w_\ell^{(k)}[i]$ are positive, we have from (6.56) and (6.54) that

$$u_\ell^k[i] v_\ell^k[i] \geq 0, \quad i = 1, \cdots, m.$$

This together with (6.46) yields (6.42). Therefore, $(x^*, y_1^*, \cdots, y_L^*)$ is a C-stationary point of problem (6.7). This completes the proof of the first part. The second half of the theorem follows from the definitions of C-stationarity and B-stationarity immediately. ∎

6.3 Smoothing Implicit Programming Method for Here-And-Now Problems

In this section, we consider the following discrete here-and-now problem:

$$\begin{aligned}
\text{minimize} \quad & \sum_{\ell=1}^{L} p_\ell\Big(f(x, y, \omega_\ell) + d^T z_\ell\Big) \\
\text{subject to} \quad & g(x) \leq 0,\ h(x) = 0, \\
& y \geq 0,\ N_\ell x + M_\ell y + q_\ell + z_\ell \geq 0, \\
& y^T(N_\ell x + M_\ell y + q_\ell + z_\ell) = 0, \\
& z_\ell \geq 0,\ \ell = 1, \cdots, L,
\end{aligned} \tag{6.68}$$

where d is a vector with positive constants. As mentioned in Section 6.1, x and y represent the upper-level and the lower-level decisions, respectively, that we have to make at once, before $\omega_\ell, \ell = 1, \cdots, L$, are observed, whereas z_ℓ is the recourse variable corresponding to ω_ℓ.

It is easy to see that problem (6.68) can be rewritten as

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{l=1}^{L} p_l\Big(f(x, y, \omega_l) + d^T z_l\Big) \\ \text{subject to} & g(x) \leq 0,\ h(x) = 0, \\ & N_\ell x + M_\ell y + q_\ell + z_\ell \geq 0, \\ & z_\ell \geq 0,\ \ell = 1, \cdots, L, \\ & y \geq 0,\ Nx + My + q + \sum_{l=1}^{L} z_l \geq 0, \\ & y^T(Nx + My + q + \sum_{l=1}^{L} z_l) = 0, \end{array} \tag{6.69}$$

where $N = \sum_{l=1}^{L} N_l, M = \sum_{l=1}^{L} M_l$, and $q = \sum_{l=1}^{L} q_l$, or equivalently,

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{\ell=1}^{L} p_\ell f(x, y, \omega_\ell) + \mathbf{d}^T \mathbf{z} \\ \text{subject to} & g(x) \leq 0,\ h(x) = 0, \\ & \mathbf{y} - \mathbf{D} y = 0,\ \mathbf{z} \geq 0, \\ & \mathbf{y} \geq 0,\ \mathbf{N}x + \mathbf{M}\mathbf{y} + \mathbf{q} + \mathbf{z} \geq 0, \\ & \mathbf{y}^T(\mathbf{N}x + \mathbf{M}\mathbf{y} + \mathbf{q} + \mathbf{z}) = 0 \end{array} \tag{6.70}$$

with $\mathbf{y}, \mathbf{N}, \mathbf{M}, \mathbf{q}$ defined by (6.8) and

$$\mathbf{z} := \begin{pmatrix} z_1 \\ \vdots \\ z_L \end{pmatrix}, \quad \mathbf{d} := \begin{pmatrix} p_1 d \\ \vdots \\ p_L d \end{pmatrix}, \quad \mathbf{D} := \begin{pmatrix} I \\ \vdots \\ I \end{pmatrix}. \tag{6.71}$$

Both problems (6.69) and (6.70) are actually ordinary MPECs, which are large scale problems in practice. In this section, we propose a smoothing implicit programming method akin to SIP-I for solving problem (6.68) with the help of a penalty technique. Note that SIP-I cannot be applied to (6.69) or (6.70) directly because of the existence of some non-complementarity constraints involving the variable y.

On the one hand, for any feasible point $(x, y, z_1, \cdots, z_L)$ of problem (6.69), $(Nx + My + q + \sum_{l=1}^{L} z_l)[i] = 0$ implies that $(N_\ell x + M_\ell y + q_\ell + z_\ell)[i] = 0$ holds for every ℓ. This indicates that the MPEC-LICQ does not hold for problem (6.69) in general. Therefore, in this section, the MPEC-LICQ means the one for problem (6.70). On the other hand, because the complementarity constraints in problem (6.69) are lower dimensional, we use them to generate the subproblems.

Suppose that M is a P_0-matrix. For each $(x, z_1, \cdots, z_L)$ and $\mu_k > 0$, let $y(x, \sum_{l=1}^L z_l, \mu_k)$ and $w(x, \sum_{l=1}^L z_l, \mu_k)$ solve

$$\Phi_{\mu_k}\Big(x, y(x, \textstyle\sum_{l=1}^L z_l, \mu_k), w(x, \sum_{l=1}^L z_l, \mu_k); N, M, q + \sum_{l=1}^L z_l\Big) = 0. \tag{6.72}$$

The existence and differentiability of the above implicit functions follow from Theorem 6.1. Note that the implicit functions are denoted by $y(x, \sum_{l=1}^L z_l, \mu_k)$ and $w(x, \sum_{l=1}^L z_l, \mu_k)$, rather than $y(x, z_1, \cdots, z_L, \mu_k)$ and $w(x, z_1, \cdots, z_L, \mu_k)$, respectively. In the following, we use ∇_z to denote the derivative with respect to the second argument. As mentioned in Section 6.1, the smooth optimization problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{\ell=1}^L p_\ell\Big(f(x, y(x, \textstyle\sum_{l=1}^L z_l, \mu_k), \omega_\ell) + d^T z_\ell\Big) \\ \text{subject to} & g(x) \le 0,\ h(x) = 0, \\ & N_\ell x + M_\ell y(x, \sum_{l=1}^L z_l, \mu_k) + q_\ell + z_\ell \ge 0, \\ & z_\ell \ge 0,\ \ell = 1, \cdots, L \end{array} \tag{6.73}$$

is an approximation of problem (6.69). Since the feasible region of problem (6.73) is dependent on μ_k, (6.73) may not be easy to solve. Therefore, we apply a penalty technique to this problem and obtain the following approximation of problem (6.69):

$$\begin{array}{ll} \text{minimize} & \theta_k(x, z_1, \cdots, z_L) \\ \text{subject to} & g(x) \le 0,\ h(x) = 0, \\ & z_\ell \ge 0,\ \ell = 1, \cdots, L, \end{array} \tag{6.74}$$

where

$$\begin{aligned} \theta_k(x, z_1, \cdots, z_L) \;=\; & \sum_{\ell=1}^L p_\ell\Big(f(x, y(x, \textstyle\sum_{l=1}^L z_l, \mu_k), \omega_\ell) + d^T z_\ell\Big) \\ & + \rho_k \sum_{\ell=1}^L \psi\Big(-(N_\ell x + M_\ell y(x, \textstyle\sum_{l=1}^L z_l, \mu_k) + q_\ell + z_\ell)\Big), \end{aligned}$$

ρ_k is a positive parameter, and $\psi : \Re^m \to [0, +\infty)$ is a smooth penalty function. Some specific penalty functions will be given later. Note that the feasible region of problem (6.74) is common for all k.

Now we present our method, called SIP-II, for problem (6.68): Choose two sequences $\{\mu_k\}$ and $\{\rho_k\}$ of positive numbers. We then solve the problems (6.74) to get a sequence

$\{(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})\}$ and let

$$y^{(k)} = y(x^{(k)}, \textstyle\sum_{l=1}^{L} z_l^{(k)}, \mu_k).$$

In what follows, we let $\mathcal{F}_2$ and $\mathcal{Z}$ stand for the feasible regions of problems (6.69) and (6.74), respectively. Also, we use (6.8) and (6.71) to generate some related vectors such as $\mathbf{y}^{(k)}, \mathbf{y}^*, \mathbf{z}^{(k)}, \mathbf{z}^*$, and so on. In addition, we make the following assumption on SIP-II throughout this section:

A5: In SIP-II, the parameters μ_k and ρ_k are selected to satisfy

$$\lim_{k\to\infty} \mu_k = 0, \quad \lim_{k\to\infty} \rho_k = +\infty, \quad \lim_{k\to\infty} \mu_k \rho_k = 0. \tag{6.75}$$

The following lemma is helpful in establishing convergence theory for SIP-II. We omit its proof because it is similar to that of Lemma 6.2.

Lemma 6.3 *Suppose that M is a P_0-matrix in problem (6.68) and, for any bounded sequence $\{(x^k, z_1^k, \cdots, z_L^k)\}$ in $\mathcal{Z}$, $\{y(x^k, \sum_{l=1}^{L} z_l^k, \mu_k)\}$ is bounded. If $(x^*, y^*, z_1^*, \cdots, z_L^*) \in \mathcal{F}_2$ and the submatrix $M[\mathcal{K}^*]$ is nondegenerate, where $\mathcal{K}^* = \{\ i \mid\ (Nx^* + My^* + q + \sum_{l=1}^{L} z_l^*)[i] = 0\}$, there exist a neighborhood U^* of $(x^*, y^*, z_1^*, \cdots, z_L^*)$ and a positive constant π^* such that*

$$\|y(x, \textstyle\sum_{l=1}^{L} z_l, \mu_k) - y\| \le \mu_k \pi^* (\|y\| + \sqrt{m}) \tag{6.76}$$

for any $(x, y, z_1, \cdots, z_L) \in U^* \cap \mathcal{F}_2$ and any k.

We next investigate the limiting behavior of the sequence generated by SIP-II. We will show that SIP-II possesses similar convergence properties to SIP-I.

Theorem 6.6 *Let M be a P_0-matrix and $\psi : \Re^m \to [0, +\infty)$ be a continuously differentiable function satisfying $\psi(0) = 0$ and*

$$\psi(t) \le \psi(t'), \quad \forall t' \ge t \in \Re^m, \tag{6.77}$$

and, for any bounded sequence $\{(x^k, z_1^k, \cdots, z_L^k)\}$ in $\mathcal{Z}$, $\{y(x^k, \sum_{l=1}^{L} z_l^k, \mu_k)\}$ be bounded. Suppose that the sequence $\left\{(x^{(k)}, y^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})\right\}$ generated by SIP-II with $(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})$ being a local optimal solution of (6.74) is convergent to $(x^*, y^*, z_1^*, \cdots, z_L^*) \in \mathcal{F}_2$. If there exists a neighborhood V^* of $(x^*, y^*, z_1^*, \cdots, z_L^*)$ such that $(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})$ minimizes θ_k over $V^*|_{\mathcal{Z}}$ for all k large enough and the submatrix $M[\mathcal{K}^*]$ is nondegenerate, where $\mathcal{K}^*$ is the same as in Lemma 6.3, then $(x^*, y^*, z_1^*, \cdots, z_L^*)$ is a local optimal solution of (6.68).

Proof: By Lemma 6.3, there exist a neighborhood $U^* \subseteq V^*$ of $(x^*, y^*, z_1^*, \cdots, z_L^*)$ and a positive number π^* such that (6.76) holds for any $(x, y, z_1, \cdots, z_L) \in U^* \cap \mathcal{F}_2$ and every k. Let

$$\mathcal{F}_{2,\eta} = \Big\{(x, y, z_1, \cdots, z_L) \in U^* \cap \mathcal{F}_2 \;\Big|\; \|(x, y, z_1, \cdots, z_L) - (x^*, y^*, z_1^*, \cdots, z_L^*)\| \leq \eta\Big\},$$

where $\eta > 0$ is a constant. Since $\mathcal{F}_{2,\eta}$ is a nonempty compact set, the problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{\ell=1}^{L} p_\ell\Big(f(x, y, \omega_\ell) + d^T z_\ell\Big) \\ \text{subject to} & (x, y, z_1, \cdots, z_L) \in \mathcal{F}_{2,\eta} \end{array} \tag{6.78}$$

has a global optimal solution, say $(\bar{x}, \bar{y}, \bar{z}_1, \cdots, \bar{z}_L)$.

Suppose $(x, y, z_1, \cdots, z_L) \in \mathcal{F}_{2,\eta}$. We then have from the mean-value theorem that

$$\begin{aligned} \theta_k(x, z_1, \cdots, z_L) \;=\; & \sum_{\ell=1}^{L} p_\ell\Big(f(x, y, \omega_\ell) + d^T z_\ell \\ & +(y(x, \textstyle\sum_{l=1}^{L} z_l, \mu_k) - y)^T \nabla_y f(x, (1-\alpha) y(x, \textstyle\sum_{l=1}^{L} z_l, \mu_k) + \alpha y, \omega_\ell)\Big) \\ & +\rho_k \sum_{\ell=1}^{L} \psi\Big(- (N_\ell x + M_\ell y(x, \textstyle\sum_{l=1}^{L} z_l, \mu_k) + q_\ell + z_\ell)\Big), \end{aligned} \tag{6.79}$$

where $\alpha \in [0, 1]$. Note that, by (6.76),

$$\begin{aligned} \|(1-\alpha) y(x, \textstyle\sum_{l=1}^{L} z_l, \mu_k) + \alpha y\| &= \|(1-\alpha)(y(x, \textstyle\sum_{l=1}^{L} z_l, \mu_k) - y) + y\| \\ &\leq \|y(x, \textstyle\sum_{l=1}^{L} z_l, \mu_k) - y\| + \|y\| \\ &\leq \mu_k \pi^*(\|y\| + \sqrt{m}) + \|y\|. \end{aligned}$$

This indicates that the set

$$\Big\{(x, (1-\alpha) y(x, \textstyle\sum_{l=1}^{L} z_l, \mu_k) + \alpha y) \;\Big|\; (x, y, z_1, \cdots, z_L) \in \mathcal{F}_{2,\eta}, \alpha \in [0, 1], k = 1, 2, \cdots\Big\}$$

is bounded. Similarly, we see

$$\begin{aligned} &\Big\{(x, \alpha M_\ell(y - y(x, \textstyle\sum_{l=1}^{L} z_l, \mu_k))) \;\Big| \\ &\quad (x, y, z_1, \cdots, z_L) \in \mathcal{F}_{2,\eta}, \ell = 1, \cdots, L, \alpha \in [0, 1], k = 1, 2, \cdots\Big\} \end{aligned}$$

is also bounded. Then, by the continuous differentiability of both f and ψ, there exists a constant $\tau > 0$ such that

$$\|\nabla_y f(x, (1-\alpha) y(x, \textstyle\sum_{l=1}^{L} z_l, \mu_k) + \alpha y, \omega_\ell)\| \leq \tau \tag{6.80}$$

and

$$\|\nabla\psi\Big(\alpha M_\ell(y - y(x, \textstyle\sum_{l=1}^L z_l, \mu_k))\Big)\| \leq \tau, \qquad \ell = 1, \cdots, L \tag{6.81}$$

hold for any $(x, y, z_1, \cdots, z_L) \in \mathcal{F}_{2,\eta}, \alpha \in [0,1]$, and every k. Noticing that $(x, y, z_1, \cdots, z_L) \in \mathcal{F}_{2,\eta}$ implies $N_\ell x + M_\ell y + q_\ell + z_\ell \geq 0$ for each ℓ, we have from (6.77) and (6.81) that

$$\begin{aligned}
& \psi\Big(-(N_\ell x + M_\ell y(x, \textstyle\sum_{l=1}^L z_l, \mu_k) + q_\ell + z_\ell)\Big) \\
\leq\ & \psi\Big(M_\ell(y - y(x, \textstyle\sum_{l=1}^L z_l, \mu_k))\Big) \\
=\ & \psi\Big(M_\ell(y - y(x, \textstyle\sum_{l=1}^L z_l, \mu_k))\Big) - \psi(0) \\
=\ & \nabla\psi\Big(\alpha' M_\ell(y - y(x, \textstyle\sum_{l=1}^L z_l, \mu_k))\Big)^T M_\ell\Big(y - y(x, \textstyle\sum_{l=1}^L z_l, \mu_k)\Big) \\
\leq\ & \tau\|M_\ell\|\ \|y - y(x, \textstyle\sum_{l=1}^L z_l, \mu_k)\|,
\end{aligned} \tag{6.82}$$

where $\alpha' \in [0,1]$ and the second equality follows from the mean-value theorem. This, together with (6.79)–(6.80) and (6.76), yields

$$\begin{aligned}
& \Big|\theta_k(x, z_1, \cdots, z_L) - \sum_{\ell=1}^L p_\ell\Big(f(x, y, \omega_\ell) + d^T z_\ell\Big)\Big| \\
\leq\ & \tau\|y(x, \textstyle\sum_{l=1}^L z_l, \mu_k) - y\| + \Big(\tau\rho_k \displaystyle\sum_{\ell=1}^L \|M_\ell\|\Big)\|y - y(x, \textstyle\sum_{l=1}^L z_l, \mu_k)\| \\
\leq\ & \pi^*\tau\Big(\mu_k + \mu_k\rho_k \sum_{\ell=1}^L \|M_\ell\|\Big)(\|y\| + \sqrt{m})
\end{aligned}$$

for any $(x, y, z_1, \cdots, z_L) \in \mathcal{F}_{2,\eta}$ and k. In particular,

$$\begin{aligned}
& \Big|\theta_k(\bar{x}, \bar{z}_1, \cdots, \bar{z}_L) - \sum_{\ell=1}^L p_\ell\Big(f(\bar{x}, \bar{y}, \omega_\ell) + e^T \bar{z}_\ell\Big)\Big| \\
\leq\ & \pi^*\tau\Big(\mu_k + \mu_k\rho_k \sum_{\ell=1}^L \|M_\ell\|\Big)(\|\bar{y}\| + \sqrt{m}).
\end{aligned} \tag{6.83}$$

Moreover, since ψ is always nonnegative, we have from the continuity of f that

$$\begin{aligned}
\lim_{k\to\infty} \theta_k(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)}) \ &\geq\ \sum_{\ell=1}^L p_\ell \lim_{k\to\infty}\Big(f(x^{(k)}, y^{(k)}, \omega_\ell) + d^T z^{(k)}\Big) \\
&=\ \sum_{\ell=1}^L p_\ell\Big(f(x^*, y^*, \omega_\ell) + d^T z^*\Big).
\end{aligned} \tag{6.84}$$

Note that, by the fact that $U^* \subseteq V^*$, $(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})$ is a global optimal solution of problem (6.74) when k is large enough, and $(\bar{x}, \bar{z}_1, \cdots, \bar{z}_L)$ is a feasible point of (6.74). We then have from (6.83) that, for every k sufficiently large,

$$\begin{aligned} & \theta_k(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)}) \\ \leq\ & \theta_k(\bar{x}, \bar{z}_1, \cdots, \bar{z}_L) \\ \leq\ & \sum_{\ell=1}^{L} p_\ell\Big(f(\bar{x}, \bar{y}, \omega_\ell) + e^T \bar{z}_\ell\Big) + \pi^* \tau\Big(\mu_k + \mu_k \rho_k \sum_{\ell=1}^{L} \|M_\ell\|\Big)(\|\bar{y}\| + \sqrt{m}). \end{aligned} \tag{6.85}$$

Therefore, taking into account the equality (6.84) and Assumption A5, we have by letting $k \to \infty$ in (6.85) that

$$\sum_{\ell=1}^{L} p_\ell\Big(f(x^*, y^*, \omega_\ell) + d^T z_\ell^*\Big) \leq \sum_{\ell=1}^{L} p_\ell\Big(f(\bar{x}, \bar{y}, \omega_\ell) + e^T \bar{z}_\ell\Big).$$

On the other hand, since $(\bar{x}, \bar{y}, \bar{z}_1, \cdots, \bar{z}_L)$ is a solution of problem (6.78), we have

$$\sum_{\ell=1}^{L} p_\ell\Big(f(x^*, y^*, \omega_\ell) + d^T z_\ell^*\Big) \geq \sum_{\ell=1}^{L} p_\ell\Big(f(\bar{x}, \bar{y}, \omega_\ell) + e^T \bar{z}_\ell\Big).$$

It then follows that

$$\sum_{\ell=1}^{L} p_\ell\Big(f(x^*, y^*, \omega_\ell) + d^T z_\ell^*\Big) = \sum_{\ell=1}^{L} p_\ell\Big(f(\bar{x}, \bar{y}, \omega_\ell) + e^T \bar{z}_\ell\Big),$$

namely, $(x^*, y^*, z_1^*, \cdots, z_L^*)$ is a global optimal solution of problem (6.78) and hence it is a local optimal solution of problem (6.68). This completes the proof. ∎

It is not difficult to see that the function

$$\psi(y) := \sum_{i=1}^{m} \Big(\max\{y[i], 0\}\Big)^\nu, \quad y \in \Re^m,$$

where $\nu \geq 2$ is a positive integer, satisfies the conditions assumed in Theorem 6.6. This function is often employed for solving constrained optimization problems. For more details, see [1].

Theorem 6.7 *Let M be a P-matrix, the function ψ be the same as in Theorem 6.6 and, for any bounded sequence $\{(x^k, z_1^k, \cdots, z_L^k)\}$ in $\mathcal{Z}$, $\{y(x^k, \sum_{l=1}^{L} z_l^k, \mu_k)\}$ be bounded. Assume that $(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})$ is a global optimal solution of (6.74) for each k. Then any accumulation point $(x^*, y^*, z_1^*, \cdots, z_L^*) \in \mathcal{F}_2$ of the sequence $\Big\{(x^{(k)}, y^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})\Big\}$ generated by SIP-II is a global optimal solution of problem (6.68).*

We omit the proof of the above theorem since it is similar to that of Theorem 6.3. Next, we discuss the limiting behavior of stationary points of problems (6.74).

Theorem 6.8 *Suppose that M is a P_0-matrix in problem (6.68), the function $\psi : \Re^m \to [0, +\infty)$ is given by $\psi(y) := \|\max(y, 0)\|^2$, and $(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})$ is a stationary point of (6.74) for each k. Let $(x^*, y^*, z_1^*, \cdots, z_L^*) \in \mathcal{F}_2$ be an accumulation point of the sequence $\left\{(x^{(k)}, y^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})\right\}$ generated by SIP-II. If the MPEC-LICQ is satisfied at $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$, then $(x^*, y^*, z_1^*, \cdots, z_L^*)$ is a C-stationary point of problem (6.68). In particular, if y^* satisfies the strict complementarity condition, then $(x^*, y^*, z_1^*, \cdots, z_L^*)$ is S-stationary to (6.68).*

Proof: Assume without loss of generality that the sequence $\left\{(x^{(k)}, y^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})\right\}$ converges to $(x^*, y^*, z_1^*, \cdots, z_L^*)$. By the MPEC-LICQ assumption, problem (6.74) satisfies the standard LICQ at $(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})$ for all k sufficiently large and so, by the stationarity of $(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)})$, there exist unique Lagrange multiplier vectors λ^k, γ^k, and

$$\alpha^k := \begin{pmatrix} \alpha_1^k \\ \vdots \\ \alpha_L^k \end{pmatrix}$$

such that

$$\nabla\theta_k(x^{(k)}, z_1^{(k)}, \cdots, z_L^{(k)}) + \begin{pmatrix} \nabla g(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \lambda^k + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \gamma^k - \begin{pmatrix} O & \cdots & O \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix} \alpha^k = 0, \quad (6.86)$$

$$h(x^{(k)}) = 0, \qquad 0 \le \lambda^k \perp g(x^{(k)}) \ge 0, \quad (6.87)$$

$$0 \le \alpha_\ell^k \perp z_\ell^{(k)} \ge 0, \qquad \ell = 1, \cdots, L. \quad (6.88)$$

In the rest of the proof, we suppose k is large enough so that (6.86)–(6.88) and (6.47) are satisfied and furthermore, for each $\ell = 1, \cdots, L$, there hold

$$\mathcal{I}_{Z_\ell}^k := \left\{i \;\middle|\; z_\ell^{(k)}[i] = 0\right\} \subseteq \mathcal{I}_{Z_\ell}^* := \left\{i \;\middle|\; z_\ell^*[i] = 0\right\}, \quad (6.89)$$

$$\mathcal{I}_{W_\ell}^k := \left\{i \;\middle|\; \left(N_\ell x^{(k)} + M_\ell y^{(k)} + q_\ell + z_\ell^{(k)}\right)[i] = 0\right\} \subseteq \mathcal{I}_{W_\ell}^* := \left\{i \;\middle|\; (N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i] = 0\right\}, \quad (6.90)$$

$$\mathcal{I}_Y^k := \left\{i \;\middle|\; y^{(k)}[i] = 0\right\} \subseteq \mathcal{I}_Y^* := \left\{i \;\middle|\; y^*[i] = 0\right\}, \quad (6.91)$$

$$\mathcal{I}_W^k := \left\{i \;\middle|\; \left(Nx^{(k)} + My^{(k)} + q + \textstyle\sum_{l=1}^L z_l^{(k)}\right)[i] = 0\right\} \subseteq \mathcal{I}_W^* := \left\{i \;\middle|\; (Nx^* + My^* + q + \textstyle\sum_{l=1}^L z_l^*)[i] = 0\right\}. \quad (6.92)$$

It is clear that

$$\mathcal{I}_W^* = \cap_{\ell=1}^L \mathcal{I}_{W_\ell}^*. \tag{6.93}$$

Analogous to (6.48), we have from the implicit function theorem that

$$\begin{pmatrix} \nabla_{(x,z)} y(x^{(k)}, \sum_{l=1}^L z_l^{(k)}, \mu_k)^T \\ \nabla_{(x,z)} w(x^{(k)}, \sum_{l=1}^L z_l^{(k)}, \mu_k)^T \end{pmatrix} = - \begin{pmatrix} M + \mu_k I & -I \\ I - D^k & D^k \end{pmatrix}^{-1} \begin{pmatrix} N & I \\ O & O \end{pmatrix}, \tag{6.94}$$

where $D^k := \mathrm{diag}\left(\frac{y^{(k)}[1]}{y^{(k)}[1] + w^{(k)}[1]}, \cdots, \frac{y^{(k)}[m]}{y^{(k)}[m] + w^{(k)}[m]} \right)$. Since

$$\begin{pmatrix} M + \mu_k I & -I \\ I - D^k & D^k \end{pmatrix}^{-1} = \begin{pmatrix} E^k D^k & E^k \\ -I + (M + \mu_k I) E^k D^k & (M + \mu_k I) E^k \end{pmatrix}$$

with $E^k := \left(D^k M + I - (1 - \mu_k) D^k \right)^{-1}$, it follows from (6.94) that

$$\nabla_{(x,z)} y(x^{(k)}, \sum_{l=1}^L z_l^{(k)}, \mu_k) = - \begin{pmatrix} N^T D^k (E^k)^T \\ D^k (E^k)^T \end{pmatrix}.$$

As a result, we have

$$\nabla_x y(x^{(k)}, \sum_{l=1}^L z_l^{(k)}, \mu_k) = -N^T D^k (E^k)^T, \tag{6.95}$$

$$\nabla_z y(x^{(k)}, \sum_{l=1}^L z_l^{(k)}, \mu_k) = -D^k (E^k)^T. \tag{6.96}$$

Thus, from the definition of θ_k, (6.95)–(6.96), and by a straightforward calculus, (6.86) becomes

$$\begin{aligned} 0 \;=\; & \begin{pmatrix} \sum_{\ell=1}^L p_\ell (\nabla_x f(x^{(k)}, y^{(k)}, \omega_\ell) - N^T D^k (E^k)^T \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell)) \\ p_1 d - \sum_{\ell=1}^L p_\ell D^k (E^k)^T \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) \\ \vdots \\ p_L d - \sum_{\ell=1}^L p_\ell D^k (E^k)^T \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) \end{pmatrix} \\ & - \begin{pmatrix} N_1^T - N^T D^k (E^k)^T M_1^T & \cdots & N_L^T - N^T D^k (E^k)^T M_L^T \\ I - D^k (E^k)^T M_1^T & \cdots & -D^k (E^k)^T M_L^T \\ -D^k (E^k)^T M_1^T & \cdots & -D^k (E^k)^T M_L^T \\ \vdots & \vdots & \vdots \\ -D^k (E^k)^T M_1^T & \cdots & I - D^k (E^k)^T M_L^T \end{pmatrix} \beta^k \\ & + \begin{pmatrix} \nabla g(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \lambda^k + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \gamma^k - \begin{pmatrix} O & \cdots & O \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix} \alpha^k. \end{aligned}$$

Here, $\beta^k \in \Re^{n+mL}$ is given by

$$\beta^k := \begin{pmatrix} \beta_1^k \\ \vdots \\ \beta_L^k \end{pmatrix}$$

with

$$\beta_\ell^k := 2\rho_k \max\Big(-(N_\ell x^{(k)} + M_\ell y^{(k)} + q_\ell + z_\ell^{(k)}),\ 0\Big) \tag{6.97}$$

for each ℓ. We then have

$$\begin{aligned} 0 \;=\; & \begin{pmatrix} \sum_{\ell=1}^L p_\ell \nabla_x f(x^{(k)}, y^{(k)}, \omega_\ell) \\ p_1 d \\ \vdots \\ p_L d \end{pmatrix} + \begin{pmatrix} \nabla g(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \lambda^k + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \gamma^k \\ & - \begin{pmatrix} O & \cdots & O \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix} \alpha^k - \begin{pmatrix} N_1^T & \cdots & N_L^T \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix} \beta^k - \begin{pmatrix} N^T \\ I \\ \vdots \\ I \end{pmatrix} v^k, \end{aligned} \tag{6.98}$$

where v^k is defined by

$$v^k := D^k (E^k)^T \sum_{\ell=1}^L \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k \Big). \tag{6.99}$$

Furthermore, by letting

$$\sum_{\ell=1}^L p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - \sum_{\ell=1}^L M_\ell^T \beta_\ell^k - u^k - M^T v^k = 0, \tag{6.100}$$

we have

$$\begin{aligned} u^k &= \sum_{\ell=1}^L \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k \Big) - M^T v^k \\ &= \Big((E^k)^{-T} - M^T D^k \Big) (E^k)^T \sum_{\ell=1}^L \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k \Big) \\ &= \Big(I - (1-\mu_k) D^k \Big) (E^k)^T \sum_{\ell=1}^L \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k \Big), \end{aligned} \tag{6.101}$$

where the second equality follows from (6.99). Combining (6.98) and (6.100) yields

$$
\begin{aligned}
0 \;=\; & \begin{pmatrix} \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y^{(k)}, \omega_\ell) \\ \sum_{\ell=1}^{L} p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) \\ p_1 d \\ \vdots \\ p_L d \end{pmatrix} + \begin{pmatrix} \nabla g(x^{(k)}) \\ O \\ O \\ \vdots \\ O \end{pmatrix} \lambda^k + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ O \\ \vdots \\ O \end{pmatrix} \gamma^k \\
& - \begin{pmatrix} O & \cdots & O \\ O & \cdots & O \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix} \alpha^k - \begin{pmatrix} N_1^T & \cdots & N_L^T \\ M_1^T & \cdots & M_L^T \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix} \beta^k - \begin{pmatrix} O \\ I \\ O \\ \vdots \\ O \end{pmatrix} u^k - \begin{pmatrix} N^T \\ M^T \\ I \\ \vdots \\ I \end{pmatrix} v^k . \qquad (6.102)
\end{aligned}
$$

We can show that, when k is large sufficiently,

$$\beta_\ell^k[i] = 0 \quad \text{as long as} \quad i \notin \mathcal{I}_{W_\ell}^*. \qquad (6.103)$$

In fact, if $i \notin \mathcal{I}_{W_\ell}^*$, namely, $(N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i] > 0$, then, when k is large enough, there must hold $(N_\ell x^{(k)} + M_\ell y^{(k)} + q_\ell + z_\ell^{(k)})[i] > 0$ and hence

$$\beta_\ell^k[i] = 2\rho_k \max\Big\{ -(N_\ell x^{(k)} + M_\ell y^{(k)} + q_\ell + z_\ell^{(k)})[i],\ 0 \Big\} = 0.$$

Taking into account (6.47) and (6.87)–(6.92), we can rewrite (6.102) as

$$
\begin{aligned}
& \begin{pmatrix} \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y^{(k)}, \omega_\ell) \\ \sum_{\ell=1}^{L} p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) \\ p_1 d \\ \vdots \\ p_L d \end{pmatrix} - \sum_{i \notin \mathcal{I}_Y^*} u^k[i] \begin{pmatrix} 0 \\ e_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} - \sum_{i \notin \mathcal{I}_W^*} v^k[i] \begin{pmatrix} N[i] \\ M[i] \\ e_i \\ \vdots \\ e_i \end{pmatrix} \\
= & -\sum_{i \in \mathcal{I}_g^*} \lambda^k[i] \begin{pmatrix} \nabla g_i(x^{(k)}) \\ 0 \\ \vdots \\ 0 \end{pmatrix} - \sum_{i=1}^{s_2} \gamma^k[i] \begin{pmatrix} \nabla h_i(x^{(k)}) \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}_{Z_\ell}^*} \alpha_\ell^k[i] \begin{pmatrix} 0 \\ \vdots \\ e_i \\ \vdots \\ 0 \end{pmatrix} \\
& + \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}_{W_\ell}^*} \beta_\ell^k[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \\ \vdots \\ e_i \\ \vdots \\ 0 \end{pmatrix} + \sum_{i \in \mathcal{I}_Y^*} u^k[i] \begin{pmatrix} 0 \\ e_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \sum_{i \in \mathcal{I}_W^*} v^k[i] \begin{pmatrix} N[i] \\ M[i] \\ e_i \\ \vdots \\ e_i \end{pmatrix}. \qquad (6.104)
\end{aligned}
$$

We next prove

$$i \notin \mathcal{I}_Y^* \quad \Rightarrow \quad \lim_{k\to\infty} u^k[i] = 0, \tag{6.105}$$

$$i \notin \mathcal{I}_W^* \quad \Rightarrow \quad \lim_{k\to\infty} v^k[i] = 0. \tag{6.106}$$

Note that, by (6.101) and (6.99),

$$u^k[i] \;=\; \frac{\mu_k y^{(k)}[i] + w^{(k)}[i]}{y^{(k)}[i] + w^{(k)}[i]} e_i^T (E^k)^T \sum_{\ell=1}^{L} \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k \Big), \tag{6.107}$$

$$v^k[i] \;=\; \frac{y^{(k)}[i]}{y^{(k)}[i] + w^{(k)}[i]} e_i^T (E^k)^T \sum_{\ell=1}^{L} \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k \Big). \tag{6.108}$$

In a similar way to the proof of Theorem 6.5, we can show that

$$\left\{ \left\| (E^k)^T \sum_{\ell=1}^{L} \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k \Big) \right\| \right\}$$

is bounded. Then, it follows from (6.107) and (6.108) that

$$i \notin \mathcal{I}_Y^* \Rightarrow \quad \lim_{k\to\infty} \frac{\mu_k y^{(k)}[i] + w^{(k)}[i]}{y^{(k)}[i] + w^{(k)}[i]} = 0 \quad \Rightarrow \lim_{k\to\infty} u^k[i] = 0,$$

$$i \notin \mathcal{I}_W^* \Rightarrow \quad \lim_{k\to\infty} \frac{y^{(k)}[i]}{y^{(k)}[i] + w^{(k)}[i]} = 0 \quad \Rightarrow \lim_{k\to\infty} v^k[i] = 0.$$

By (6.105) and (6.106), the left-hand side of equality (6.104) is convergent. Moreover, from the assumption that the MPEC-LICQ holds at $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$, we can prove that all the sequences of the multipliers that appear on the right-hand side of (6.104) are bounded. In face, by letting $u_\ell^k := \Big(I - (1-\mu_k) D^k\Big)(E^k)^T \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k\Big)$ for $\ell = 1, \cdots, L$,

$$\mathbf{v}^k := \begin{pmatrix} v^k \\ \vdots \\ v^k \end{pmatrix} \in \Re^{mL}, \quad \mathbf{u}^k := \begin{pmatrix} u_1^k \\ \vdots \\ u_L^k \end{pmatrix},$$

and

$$\begin{aligned} \mathbf{a}^k &:= \mathbf{u}^k + \mathbf{M}^T(\beta^k + \mathbf{v}^k), \\ \mathbf{b}^k &:= \mathbf{u}^k, \\ \mathbf{c}^k &:= \beta^k + \mathbf{v}^k, \end{aligned}$$

we can obtain from (6.102) that

$$0 \;=\; \begin{pmatrix} \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y^{(k)}, \omega_\ell) \\ \sum_{\ell=1}^{L} p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) \\ 0 \\ \mathbf{d} \end{pmatrix} + \begin{pmatrix} \nabla g(x^{(k)}) \\ O \\ O \\ O \end{pmatrix} \lambda^k + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ O \\ O \end{pmatrix} \gamma^k$$

$$
- \begin{pmatrix} O \\ O \\ O \\ I \end{pmatrix} \alpha^k + \begin{pmatrix} O \\ -\mathbf{D}^T \\ I \\ O \end{pmatrix} \mathbf{a}^k - \begin{pmatrix} O \\ O \\ I \\ O \end{pmatrix} \mathbf{b}^k - \begin{pmatrix} \mathbf{N}^T \\ O \\ \mathbf{M}^T \\ I \end{pmatrix} \mathbf{c}^k. \quad (6.109)
$$

Applying (6.105) and (6.106), it is not difficult to show that

$$
\begin{aligned}
\mathbf{y}^*[i] > 0 \quad &\Rightarrow \quad \lim_{k\to\infty} \mathbf{u}^k[i] = 0, \\
(\mathbf{N}x^* + \mathbf{M}\mathbf{y}^* + \mathbf{q} + \mathbf{z}^*)[i] > 0 \quad &\Rightarrow \quad \lim_{k\to\infty} \mathbf{v}^k[i] = 0.
\end{aligned}
$$

From (6.93), (6.103), and the MPEC-LICQ assumption, we can see that all the sequences of the multiplier vectors in (6.109) are bounded, which implies the boundedness of the multiplier sequences that appear on the right-hand side of (6.104). In consequence, assuming these vector sequences are all convergent without loss of generality and letting $k \to \infty$ in (6.104), we may obtain the equality corresponding to (6.41).

In addition, since both $y^{(k)}[i]$ and $w^{(k)}[i]$ are positive, we have from (6.107) and (6.108) that

$$
u^k[i]v^k[i] \geq 0, \quad i = 1, \cdots, m.
$$

This together with (6.87)–(6.88) yields the results corresponding to (6.42). Therefore, $(x^*, y_1^*, \cdots, y_L^*)$ is a C-stationary point of problem (6.68). If, in addition, y^* satisfies the strict complementarity condition, then $(x^*, y^*, z_1^*, \cdots, z_L^*)$ is a S-stationary point by the definitions of C-stationarity and S-stationarity immediately. This completes the proof of Theorem 6.8. ∎

Note that similar discussion is valid for the following generalized problem:

$$
\begin{array}{ll}
\text{minimize} & \displaystyle\sum_{\ell=1}^{L} p_\ell \Big(f(x, y, \omega_\ell) + d_\ell^T z_\ell \Big) \\
\text{subject to} & g(x) \leq 0, \ h(x) = 0, \\
& y \geq 0, \ N_\ell x + M_\ell y + q_\ell + A_\ell z_\ell \geq 0, \\
& y^T (N_\ell x + M_\ell y + q_\ell + A_\ell z_\ell) = 0, \\
& z_\ell \geq 0, \ \ell = 1, \cdots, L,
\end{array} \quad (6.110)
$$

where, for each ℓ, d_ℓ is a positive vector and $A_\ell \in \Re^{m \times m}$ is a recourse matrix corresponding to the recourse variable z_ℓ.

6.4 Conclusions

In this chapter, we have presented smoothing implicit programming methods for stochastic mathematical programs with linear complementarity constraints, including both the lower-level wait-and-see and here-and-now cases. Comprehensive convergence theory has also been established. Recall that, as we mentioned in the first section, SMPECs contain the ordinary MPECs as a special subclass. As a result, all the conclusions remain true for MPECs.

Chapter 7

Some Reformulations and Algorithms for SMPECs

In this chapter, we consider the following here-and-now problem:

$$
\begin{aligned}
\text{minimize} \quad & f(x,y) + E_\omega[d^T z(\omega)] \\
\text{subject to} \quad & g(x,y) \le 0,\ h(x,y) = 0, \\
& y \ge 0,\ N(\omega)x + M(\omega)y + q(\omega) + z(\omega) \ge 0, \\
& y^T(N(\omega)x + M(\omega)y + q(\omega) + z(\omega)) = 0, \\
& z(\omega) \ge 0, \quad \omega \in \Omega,
\end{aligned}
\tag{7.1}
$$

where $f : \Re^{n+m} \to \Re, g : \Re^{n+m} \to \Re^{s_1}$, and $h : \Re^{n+m} \to \Re^{s_2}$ are all continuously differentiable, the data $N(\omega) \in \Re^{m\times n}$, $M(\omega) \in \Re^{m\times m}$, and $q(\omega) \in \Re^m$ are random variables, $z(\omega)$ is the corresponding recourse variable, and the vector $d \in \Re^m$ is a constant vector with positive elements. Note that g and h are functions of both the upper-level and the lower-level variables, unlike the problem dealt with in Chapter 6.3 where g and h are functions of the variable x only. We will give some equivalent reformulations and propose some penalty methods for solving problem (7.1). The notations employed in this chapter are the same as the previous chapter.

7.1 Examples

Many problems can be formulated as this kind of models: The first example can be found in Section 6.1. Now we describe another example.

Example 7.1 An ordinary linear complementarity problem is to find a vector $y \in \Re^m$ such that

$$y \geq 0, \quad My + q \geq 0, \quad y^T(My + q) = 0,$$

where $M \in \Re^{m \times m}$ and $q \in \Re^m$. In many practical problems, some elements may involve uncertain data, which can be characterized by random variables. Therefore, it is meaningful to consider the following stochastic linear complementarity problem:

$$y \geq 0,\ M(\omega)y + q(\omega) \geq 0,\ y^T(M(\omega)y + q(\omega)) = 0, \quad \forall \omega \in \Omega. \tag{7.2}$$

In general, there may not exist a vector y satisfying these complementarity conditions for all $\omega \in \Omega$. In order to get a reasonable resolution, we may introduce recourse variables $z(\omega) \geq 0$ to the inequality $M(\omega)y + q(\omega) \geq 0$ and try to find a vector $y \geq 0$ that minimizes the total recourse. Thus, we obtain the following problem

$$\begin{array}{ll} \text{minimize} & E_\omega[d^T z(\omega)] \\ \text{subject to} & 0 \leq y \perp (M(\omega)y + q(\omega) + z(\omega)) \geq 0, \\ & z(\omega) \geq 0,\ \omega \in \Omega, \end{array} \tag{7.3}$$

where d is a constant vector with positive elements. Problem (7.3) is obviously a special case of problem (7.1).

7.2 Reformulations

Because of the existence of the complementarity constraints, problem (7.1) does not satisfy a standard constraint qualification such as the linear independence constraint qualification or the Mangasarian-Fromovitz constraint qualification at any feasible point [17]. Therefore, the conventional theory and algorithms in nonlinear programming cannot be applied to this problem directly. In order to develop some effective methods for solving problem (7.1), we give some reformulations of (7.1) in the following.

At first, we define the function $Q : \Re^n \times \Re^m \times \Omega \to [0, +\infty]$ by

$$Q(x, y, \omega) := \sup \Big\{ -(u + ty)^T(N(\omega)x + M(\omega)y + q(\omega)) \ \Big|\ u + ty \leq d, \quad u \geq 0 \Big\}$$

and consider the problem

$$\begin{array}{ll} \text{minimize} & f(x, y) + E_\omega[Q(x, y, \omega)] \\ \text{subject to} & g(x, y) \leq 0,\ h(x, y) = 0, \quad y \geq 0, \end{array} \tag{7.4}$$

which will be shown to be equivalent to problem (7.1). In what follows, we let $\mathcal{F}_1$ and $\mathcal{F}_2$ denote the feasible regions of problems (7.1) and (7.4), respectively. We assume both problems (7.1) and (7.4) have an optimal solution. Moreover, we suppose f is bounded from below on $\mathcal{F}_2$.

7.2.1 Properties of the function Q

In order to show the equivalence between problems (7.4) and (7.1), we first give some properties of the function Q.

Theorem 7.1 *For any $x \in \Re^n, y \in \Re^m$, and $\omega \in \Omega$, $Q(x,y,\omega) < +\infty$ if and only if the set*

$$Z(x,y,\omega) := \left\{ z(\omega) \;\middle|\; \begin{array}{l} y^T(N(\omega)x + M(\omega)y + q(\omega) + z(\omega)) = 0 \\ N(\omega)x + M(\omega)y + q(\omega) + z(\omega) \geq 0, \;\; z(\omega) \geq 0 \end{array} \right\}$$

is nonempty; moreover, we have

$$Q(x,y,\omega) = \min \left\{ d^T z(\omega) \;\middle|\; z(\omega) \in Z(x,y,\omega) \right\}. \tag{7.5}$$

We can prove the above theorem using the duality theorem in linear programming immediately. On the other hand, we may write

$$\begin{aligned} Z(x,y,\omega) &= \left\{ z(\omega) \;\middle|\; \begin{array}{l} y^T(N(\omega)x + M(\omega)y + q(\omega) + z(\omega)) \leq 0 \\ N(\omega)x + M(\omega)y + q(\omega) + z(\omega) \geq 0, \;\; z(\omega) \geq 0 \end{array} \right\} \\ &= \left\{ z(\omega) \;\middle|\; \begin{array}{l} y^T z(\omega) \leq -y^T(N(\omega)x + M(\omega)y + q(\omega)) \\ z(\omega) \geq -(N(\omega)x + M(\omega)y + q(\omega)), \;\; z(\omega) \geq 0 \end{array} \right\}. \end{aligned}$$

Therefore, when $Q(x,y,\omega) < +\infty$, we have from Theorem 7.1 and the duality theorem that

$$Q(x,y,\omega) = \max \left\{ -(u+ty)^T(N(\omega)x + M(\omega)y + q(\omega)) \;\middle|\; u + ty \leq d, \;\; u \geq 0, \;\; t \leq 0 \right\}.$$

Furthermore, we have the the following result, which will be used in the subsequent analysis.

Theorem 7.2 *Let $g(x,y) \leq 0, h(x,y) = 0, y \geq 0$, and $\omega \in \Omega$. Then $Q(x,y,\omega) = +\infty$ if and only if there exists an index i such that*

$$\Big(N(\omega)x + M(\omega)y + q(\omega)\Big)[i] > 0, \quad y[i] > 0. \tag{7.6}$$

Proof: " if ": Suppose that there exists an index i satisfying (7.6). Let t be a real number and $u(t) \in \Re^m$ be defined by

$$u(t) := ty[i]e_i - ty.$$

Then, for any $t \leq 0$, we have

$$u(t) \geq 0, \quad u(t) + ty \leq d.$$

It follows from the definition of Q and (7.6) that

$$\begin{aligned} Q(x,y,\omega) &\geq \sup\Big\{ -(u(t)+ty)^T(N(\omega)x + M(\omega)y + q(\omega)) \ \Big| \ t \leq 0 \Big\} \\ &= \sup\Big\{ -ty[i](N(\omega)x + M(\omega)y + q(\omega))[i] \ \Big| \ t \leq 0 \Big\} \\ &= +\infty, \end{aligned}$$

from which we obtain $Q(x,y,\omega) = +\infty$ immediately.

" only if ": First of all, we define

$$\begin{aligned} \mathcal{I}_1 &:= \Big\{ i \ \Big| \ y[i] = 0 \Big\}, \\ \mathcal{I}_2 &:= \Big\{ i \ \Big| \ y[i] > 0, \quad (N(\omega)x + M(\omega)y + q(\omega))[i] = 0 \Big\}, \\ \mathcal{I}_3 &:= \Big\{ i \ \Big| \ y[i] > 0, \quad (N(\omega)x + M(\omega)y + q(\omega))[i] < 0 \Big\}, \\ \mathcal{I}_4 &:= \Big\{ i \ \Big| \ y[i] > 0, \quad (N(\omega)x + M(\omega)y + q(\omega))[i] > 0 \Big\}. \end{aligned}$$

These index sets are mutually disjoint and, since $y \geq 0$, we see that

$$\mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3 \cup \mathcal{I}_4 = \{1, 2, \cdots, n\}. \tag{7.7}$$

Denote $C := \{ \ (u,t) \ | \ \ u + ty \leq d, \ u \geq 0\}$. Then, we have

$$Q(x,y,\omega) = \sup\Big\{ -\sum_{i=1}^{m}(u+ty)[i](N(\omega)x + M(\omega)y + q(\omega))[i] \ \Big| \ (u,t) \in C \Big\}. \tag{7.8}$$

(a) For any $i \in \mathcal{I}_1$ and any $(u,t) \in C$, we have $0 \leq u[i] \leq d[i]$. It follows that the term

$$\begin{aligned} & -\sum_{i \in \mathcal{I}_1}(u+ty)[i](N(\omega)x + M(\omega)y + q(\omega))[i] \\ = & -\sum_{i \in \mathcal{I}_1} u[i](N(\omega)x + M(\omega)y + q(\omega))[i] \end{aligned}$$

is bounded on C.

(b) It is obvious that

$$-\sum_{i\in\mathcal{I}_2}(u+ty)[i](N(\omega)x+M(\omega)y+q(\omega))[i]=0.$$

(c) Let $i\in\mathcal{I}_3$. For any $(u,t)\in C$, we have

$$-\sum_{i\in\mathcal{I}_3}(u+ty)[i](N(\omega)x+M(\omega)y+q(\omega))[i]\le-\sum_{i\in\mathcal{I}_3}d[i](N(\omega)x+M(\omega)y+q(\omega))[i]$$

and so the term on the left-hand side is bounded above on C.

The facts (a)–(c) together with (7.7) and (7.8) indicate that, if $Q(x,y,\omega)=+\infty$, we must have $\mathcal{I}_4\neq\emptyset$, from which we see that (7.6) must hold for some index i and hence the proof is completed. ∎

From Theorem 7.2, we have the following result immediately.

Corollary 7.1 *Let $g(x,y)\le 0, h(x,y)=0, y\ge 0$, and $\omega\in\Omega$. Then $Q(x,y,\omega)<+\infty$ if and only if*

$$y[i](N(\omega)x+M(\omega)y+q(\omega))[i]\le 0 \tag{7.9}$$

holds for every i.

This corollary, together with Theorem 7.1, yields the next result.

Corollary 7.2 *Let $g(x,y)\le 0, h(x,y)=0, y\ge 0$, and $\omega\in\Omega$. If $Q(x,y,\omega)<+\infty$, then*

$$Q(x,y,\omega)=d^Tz(x,y,\omega), \tag{7.10}$$

where

$$z(x,y,\omega):=\max\{-(N(\omega)x+M(\omega)y+q(\omega)),\ 0\}. \tag{7.11}$$

Proof: Noticing that (7.9) together with $y\ge 0$ holds for each i, we can easily prove that $z(x,y,\omega)\in Z(x,y,\omega)$. On the other hand, for any $z(\omega)\in Z(x,y,\omega)$, since

$$z(\omega)\ge-(N(\omega)x+M(\omega)y+q(\omega)),\qquad z(\omega)\ge 0,$$

we have from the definition (7.11) that

$$z(\omega)-z(x,y,\omega)\ge 0,$$

which implies that

$$d^T z(\omega) \geq d^T z(x, y, \omega).$$

Since $z(\omega) \in Z(x, y, \omega)$ is arbitrary, (7.10) follows form (7.5) at once. ∎

We next show the equivalence between problems (7.1) and (7.4).

7.2.2 Continuous case

Let ω be a continuous random variable and $p(\omega)$ represent the probability density function of ω. Suppose that the probability of any nonempty open set in Ω is positive and $(N(\omega), M(\omega), q(\omega))$ is continuous with respect to ω. Analogous to the discrete case, we have the following result:

Theorem 7.3 *If (x^*, y^*) solves problem (7.4), then there exist $z^*(\omega) \in Z(x^*, y^*, \omega)$, $\omega \in \Omega$, such that $(x^*, y^*, z^*(\omega))_{\omega\in\Omega}$ solves problem (7.1). Conversely, if $(x^*, y^*, z^*(\omega))_{\omega\in\Omega}$ solves problem (7.1), then (x^*, y^*) solves (7.4).*

Proof: (a) Suppose that (x^*, y^*) solves (7.4). Then we claim that

$$Q(x^*, y^*, \omega) < +\infty, \qquad \forall \omega \in \Omega. \tag{7.12}$$

In fact, if $Q(x^*, y^*, \bar{\omega}) = +\infty$ for some $\bar{\omega} \in \Omega$, then, by Theorem 7.2, there exists an index i such that

$$\Big(N(\bar{\omega})x^* + M(\bar{\omega})y^* + q(\bar{\omega})\Big)[i] > 0, \quad y^*[i] > 0.$$

It follows from the continuity of $(N(\omega), M(\omega), q(\omega))$ that there is a neighborhood $U(\bar{\omega})$ of $\bar{\omega}$ such that

$$\Big(N(\omega)x^* + M(\omega)y^* + q(\omega)\Big)[i] > 0, \quad y^*[i] > 0$$

hold for any $\omega \in U(\bar{\omega})$. We then have from Theorem 7.2 that

$$Q(x^*, y^*, \omega) = +\infty, \qquad \forall \omega \in U(\bar{\omega}).$$

Therefore,

$$f(x^*, y^*) + E_\omega[Q(x^*, y^*, \omega)] \geq f(x^*, y^*) + \int_{U(\bar{\omega})} Q(x^*, y^*, \omega)p(\omega)d\omega = +\infty.$$

This is a contradiction and hence (7.12) must hold. As a result, we have from Theorem 7.1 and Corollary 7.2 that there exists $z^*(\omega) \in Z(x^*, y^*, \omega)$ such that

$$Q(x^*, y^*, \omega) = \min\Big\{d^T z(\omega) \;\Big|\; z(\omega) \in Z(x^*, y^*, \omega)\Big\} = d^T z^*(\omega).$$

It then follows that, for any $(x, y, z(\omega))_{\omega\in\Omega} \in \mathcal{F}_1$,

$$\begin{aligned} f(x^*, y^*) + E_\omega[d^T z^*(\omega)] &= f(x^*, y^*) + E_\omega[Q(x^*, y^*, \omega)] \\ &\leq f(x, y) + E_\omega[Q(x, y, \omega)] \\ &\leq f(x, y) + E_\omega[d^T z(\omega)], \end{aligned} \tag{7.13}$$

where the first inequality follows from the optimality of (x^*, y^*) and the last inequality follows from Theorem 7.1 and the fact that $(x, y, z(\omega))_{\omega\in\Omega} \in \mathcal{F}_1$ implies $Z(x, y, \omega)$ is nonempty (and hence there holds (7.5)) for any $\omega \in \Omega$. The inequality (7.13) means that $(x^*, y^*, z^*(\omega))_{\omega\in\Omega}$ is an optimal solution of problem (7.1).

(b) Let $(x^*, y^*, z^*(\omega))_{\omega\in\Omega}$ be an optimal solution of problem (7.1). Note that $z^*(\omega) \in Z(x^*, y^*, \omega)$ for any $\omega \in \Omega$. It then follows from Theorem 7.1 that

$$Q(x^*, y^*, \omega) \leq d^T z^*(\omega), \qquad \omega \in \Omega. \tag{7.14}$$

We next show that (x^*, y^*) is a global optimal solution of problem (7.4), namely, for any $(x, y) \in \mathcal{F}_2$,

$$f(x^*, y^*) + E_\omega[Q(x^*, y^*, \omega)] \leq f(x, y) + E_\omega[Q(x, y, \omega)]. \tag{7.15}$$

Let $(x, y) \in \mathcal{F}_2$.

(b1) Suppose that $Q(x, y, \omega) < +\infty$ for every $\omega \in \Omega$ and let $z(x, y, \omega)$ be the vector defined in Corollary 7.2. By the same corollary, we see that $z(x, y, \omega) \in Z(x, y, \omega_\ell)$ and $Q(x, y, \omega) = d^T z(x, y, \omega)$. It is not difficult to see that $(x, y, z(x, y, \omega))_{\omega\in\Omega} \in \mathcal{F}_1$ and

$$\begin{aligned} f(x^*, y^*) + E_\omega[Q(x^*, y^*, \omega)] &\leq f(x^*, y^*) + E_\omega[d^T z^*(\omega)] \\ &\leq f(x, y) + E_\omega[d^T z(x, y, \omega)] \\ &= f(x, y) + E_\omega[Q(x, y, \omega)], \end{aligned}$$

where the first inequality follows from (7.14) and the second inequality follows from the optimality of $(x^*, y^*, z^*(\omega))_{\omega\in\Omega}$ to problem (7.1). So, (7.15) is valid in this case.

(b2) If $Q(x, y, \hat{\omega}) = +\infty$ for some $\hat{\omega} \in \Omega$, in a similar way to (a), we can show that there exists a neighborhood $U(\hat{\omega})$ of $\hat{\omega}$ such that

$$Q(x, y, \omega) = +\infty, \qquad \forall \omega \in U(\hat{\omega}).$$

It follows that $E_\omega[Q(x, y, \omega)] = +\infty$, which implies that (7.15) remains true.

Therefore, (x^*, y^*) is a global optimal solution of problem (7.4) and hence the proof of the theorem is completed. ∎

7.2.3 Discrete case

Suppose that $\Omega = \{\omega_1, \omega_2, \cdots, \omega_L\}$ and, for each $\ell = 1, 2, \cdots, L$, the probability p_ℓ of the random event ω_ℓ is positive. Then, problems (7.1) and (7.4) reduce to

$$\begin{aligned} \text{minimize} \quad & f(x,y) + \sum_{\ell=1}^{L} p_\ell d^T z_\ell \\ \text{subject to} \quad & g(x,y) \le 0,\ h(x,y) = 0, \\ & y \ge 0,\ N_\ell x + M_\ell y + q_\ell + z_\ell \ge 0, \\ & y^T(N_\ell x + M_\ell y + q_\ell + z_\ell) = 0, \\ & z_\ell \ge 0, \quad \ell = 1, 2, \cdots, L \end{aligned} \tag{7.16}$$

and

$$\begin{aligned} \text{minimize} \quad & f(x,y) + \sum_{\ell=1}^{L} p_\ell Q(x, y, \omega_\ell) \\ \text{subject to} \quad & g(x,y) \le 0,\ h(x,y) = 0, \quad y \ge 0, \end{aligned} \tag{7.17}$$

respectively. The following result shows that problem (7.17) is equivalent to the MPEC (7.16).

Theorem 7.4 *If (x^*, y^*) solves problem (7.17), then there exist $z_\ell^* \in Z(x^*, y^*, \omega_\ell), \ell = 1, 2, \cdots, L$, such that $(x^*, y^*, z_1^*, \cdots, z_L^*)$ solves the MPEC (7.16). Conversely, if $(x^*, y^*, z_1^*, \cdots, z_L^*)$ solves the MPEC (7.16), then (x^*, y^*) solves problem (7.17).*

Proof: (a) Suppose that (x^*, y^*) solves (7.17). By assumption, problem (7.17) has a finite optimal value and furthermore, taking into account that each p_ℓ is positive, we see that

$$Q(x^*, y^*, \omega_\ell) < +\infty, \qquad \ell = 1, 2, \cdots, L.$$

Then, by Corollary 7.2, there exist $z_\ell^* \in Z(x^*, y^*, \omega_\ell), \ell = 1, 2, \cdots, L$, such that $(x^*, y^*, z_1^*, \cdots, z_L^*)$ is feasible to problem (7.16) and

$$Q(x^*, y^*, \omega_\ell) = d^T z_\ell^*, \qquad \ell = 1, 2, \cdots, L.$$

Here, we use the lower boundedness of the function $d^T z_\ell$ on $Z(x^*, y^*, \omega_\ell)$ and the polyhedral convexity of the set $Z(x^*, y^*, \omega_\ell)$. We next prove that $(x^*, y^*, z_1^*, \cdots, z_L^*)$ solves problem (7.16).

Let $(x, y, z_1, \cdots, z_L) \in \mathcal{F}_1$. This implies that $z_\ell \in Z(x, y, \omega_\ell)$ for each ℓ and hence, by Theorem 7.1, we have

$$Q(x, y, \omega_\ell) \le d^T z_\ell, \qquad \ell = 1, 2, \cdots, L.$$

It then follows that

$$\begin{aligned} f(x^*,y^*)+\sum_{\ell=1}^{L}p_\ell d^T z_\ell^* &= f(x^*,y^*)+\sum_{\ell=1}^{L}p_\ell Q(x^*,y^*,\omega_\ell) \\ &\le f(x,y)+\sum_{\ell=1}^{L}p_\ell Q(x,y,\omega_\ell) \\ &\le f(x,y)+\sum_{\ell=1}^{L}p_\ell d^T z_\ell, \end{aligned}$$

which means that $(x^*,y^*,z_1^*,\cdots,z_L^*)$ is an optimal solution of problem (7.16).

(b) Suppose that $(x^*,y^*,z_1^*,\cdots,z_L^*)$ solves problem (7.16). Note that, $z_\ell^*\in Z(x^*,y^*,\omega_\ell)$, $\ell=1,\cdots,L$. It then follows from Theorem 7.1 that

$$Q(x^*,y^*,\omega_\ell)\le d^T z_\ell^*, \qquad \ell=1,2,\cdots,L. \tag{7.18}$$

Let $(x,y)\in\mathcal{F}_2$. First we suppose that $Q(x,y,\omega_\ell)<+\infty$ for every ℓ. This means that $Z(x,y,\omega_\ell)$ is nonempty for each ℓ. For any $z_\ell\in Z(x,y,\omega_\ell)$, $\ell=1,2,\cdots,L$, we have $(x,y,z_1,\cdots,z_L)\in\mathcal{F}_1$ and then, by (7.18) and the optimality of $(x^*,y^*,z_1^*,\cdots,z_L^*)$,

$$\begin{aligned} f(x^*,y^*)+\sum_{\ell=1}^{L}p_\ell Q(x^*,y^*,\omega_\ell) &\le f(x^*,y^*)+\sum_{\ell=1}^{L}p_\ell d^T z_\ell^* \\ &\le f(x,y)+\sum_{\ell=1}^{L}p_\ell d^T z_\ell. \end{aligned}$$

Noticing that $z_\ell\in Z(x,y,\omega_\ell)$ is arbitrary for every ℓ and taking (7.5) into account, we obtain

$$f(x^*,y^*)+\sum_{\ell=1}^{L}p_\ell Q(x^*,y^*,\omega_\ell)\le f(x,y)+\sum_{\ell=1}^{L}p_\ell Q(x,y,\omega_\ell) \tag{7.19}$$

for any $(x,y)\in\mathcal{F}_2$. If $Q(x,y,\omega_\ell)=+\infty$ for some index ℓ, (7.19) holds immediately from the positivity of p_l. This means that (x^*,y^*) is an optimal solution of problem (7.17). ∎

7.3 Further Discussions on Discrete Problems

We continue to discuss the discrete problem (7.16) in this section. Theorem 7.4 indicates that problem (7.16) is equivalent to problem (7.17) and furthermore, by Corollary 7.1

and the positivity of every p_ℓ, problem (7.17) is equivalent to the problem

$$\begin{aligned} \text{minimize} \quad & f(x,y) + \sum_{\ell=1}^{L} p_\ell Q(x,y,\omega_\ell) \\ \text{subject to} \quad & g(x,y) \le 0,\ \ h(x,y) = 0, \quad y \ge 0, \\ & y[i](N_\ell x + M_\ell y + q_\ell)[i] \le 0, \\ & i = 1, \cdots, m,\ \ \ell = 1, \cdots, L. \end{aligned} \tag{7.20}$$

Let $\mathcal{F}_3$ denote the feasible region of problem (7.20). We then have from Corollary 7.1 that

$$Q(x,y,\omega_\ell) < +\infty, \qquad \forall (x,y) \in \mathcal{F}_3,\ \forall \ell.$$

By Corollary 7.2, we see that, for any $(x,y) \in \mathcal{F}_3$ and any ℓ,

$$Q(x,y,\omega_\ell) = d^T z_\ell(x,y)$$

with $z_\ell(x,y) = \max(-(N_\ell x + M_\ell y + q_\ell),\ 0)$. Let

$$\theta(x,y) := \sum_{\ell=1}^{L} p_\ell d^T \max\Big(- (N_\ell x + M_\ell y + q_\ell),\ 0\Big). \tag{7.21}$$

Then problem (7.20) may be rewritten as

$$\begin{aligned} \text{minimize} \quad & f(x,y) + \theta(x,y) \\ \text{subject to} \quad & g(x,y) \le 0,\ \ h(x,y) = 0, \quad y \ge 0, \\ & y[i](N_\ell x + M_\ell y + q_\ell)[i] \le 0, \\ & i = 1, \cdots, m,\ \ \ell = 1, \cdots, L, \end{aligned} \tag{7.22}$$

which is no longer an SMPEC.

Example 7.2 Consider the SMPEC

$$\begin{aligned} \text{minimize} \quad & p(z_1[1] + z_1[2]) + (1-p)(z_2[1] + z_2[2]) \\ \text{subject to} \quad & z_1[1] \ge 0,\ z_1[2] \ge 0,\ z_2[1] \ge 0,\ z_2[2] \ge 0, \\ & 0 \le \begin{pmatrix} y[1] \\ y[2] \end{pmatrix} \perp \begin{pmatrix} 2y[1] - 3 + z_1[1] \\ y[2] - 5 + z_1[2] \end{pmatrix} \ge 0, \\ & 0 \le \begin{pmatrix} y[1] \\ y[2] \end{pmatrix} \perp \begin{pmatrix} 2y[2] - 7 + z_2[1] \\ y[1] - 2 + z_2[2] \end{pmatrix} \ge 0, \end{aligned} \tag{7.23}$$

where p is a constant such that $0 < p < 1$. For this problem, problem (7.22) becomes

$$\begin{array}{ll} \text{minimize} & p\Big(\max\{3-2y[1],\ 0\}+\max\{5-y[2],\ 0\}\Big) \\ & \quad +(1-p)\Big(\max\{7-2y[2],\ 0\}+\max\{2-y[1],\ 0\}\Big) \\ \text{subject to} & y[1]\geq 0,\quad y[2]\geq 0, \\ & y[1](2y[1]-3)\leq 0,\quad y[2](y[2]-5)\leq 0, \\ & y[1](2y[2]-7)\leq 0,\quad y[2](y[1]-2)\leq 0. \end{array} \tag{7.24}$$

Problem (7.24) has a unique solution $y^* = (\frac{3}{2}, \frac{7}{2})$ for any $p \in (0,1)$ and it is easy to verify that the linear independence constraint qualification holds at y^*. This indicates that problem (7.24) has ordinary constraints, unlike the original SMPEC (7.23) that does not satisfy any standard constraint qualification at any feasible point.

Problem (7.22) possesses the same optimal solution set as problem (7.16) in the sense of Theorem 7.4. However, it may not be easy to obtain a global optimal solution of an optimization problem in practice, whereas computation of stationary points may be relatively easy. Therefore, it is necessary to study the relation between the stationary points of problems (7.22) and (7.16). Recall that, for each ℓ, $N_\ell[i]$ and $M_\ell[i]$ denote the column vectors whose elements comprise the ith row of the matrices N_ℓ and M_ℓ, respectively.

Since θ is a nonsmooth convex function, we will use the following standard definition of stationarity for problem (7.22):

Definition 7.1 A point $(x^*, y^*) \in \Re^{n+m}$ is said to be *stationary* to problem (7.22) if it is feasible to (7.22) and there exist Lagrange multiplier vectors λ^*, μ^*, ν^*, and ξ_ℓ^*, $\ell = 1, \cdots, L$, such that

$$0 \in \nabla f(x^*,y^*) + \partial\theta(x^*,y^*) + \nabla g(x^*,y^*)\lambda^* + \nabla h(x^*,y^*)\mu^* - \begin{pmatrix} O \\ I \end{pmatrix}\nu^*$$
$$+\sum_{\ell=1}^{L}\sum_{i=1}^{m}\xi_\ell^*[i]\Big((N_\ell x^* + M_\ell y^* + q_\ell)[i]\begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i]\begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big), \tag{7.25}$$
$$0 \leq \lambda^* \perp (-g(x^*,y^*)) \geq 0, \tag{7.26}$$
$$0 \leq \nu^* \perp \quad y^* \geq 0, \tag{7.27}$$
$$0 \leq \xi_\ell^*[i] \perp (-y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i]) \geq 0. \tag{7.28}$$

Here, $\partial\theta(x^*, y^*)$ stands for the subdifferential [74] of θ at the point (x^*, y^*).

For each ℓ and i, we let $\theta_{\ell,i}(x, y) := \max\{-(N_\ell x + M_\ell y + q_\ell)[i], 0\}$. It then follows that

$$\partial\theta_{\ell,i}(x^*,y^*) = \begin{cases} \mathrm{co}\left\{ \begin{pmatrix} -N_\ell[i] \\ -M_\ell[i] \end{pmatrix}, 0 \right\}, & (N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0 \\ \left\{ \begin{pmatrix} -N_\ell[i] \\ -M_\ell[i] \end{pmatrix} \right\}, & (N_\ell x^* + M_\ell y^* + q_\ell)[i] < 0 \\ \left\{ 0 \right\}, & (N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0 \end{cases} \tag{7.29}$$

and

$$\partial\theta(x^*,y^*) = \sum_{\ell=1}^{L}\sum_{i=1}^{m} p_\ell \; d[i] \; \partial\theta_{\ell,i}(x^*,y^*), \tag{7.30}$$

where co stands for the convex hull.

On the other hand, it is easy to see that problem (7.16) is equivalent to the following ordinary MPEC:

$$\begin{aligned} \text{minimize} \quad & f(x,y) + \mathbf{d}^T\mathbf{z} \\ \text{subject to} \quad & g(x,y) \le 0,\ h(x,y) = 0, \\ & \mathbf{y} - Dy = 0,\ \mathbf{z} \ge 0, \\ & \mathbf{y} \ge 0,\ Nx + M\mathbf{y} + q + \mathbf{z} \ge 0, \\ & \mathbf{y}^T(Nx + M\mathbf{y} + q + \mathbf{z}) = 0, \end{aligned} \tag{7.31}$$

where $\mathbf{y} \in \Re^{mL}, \mathbf{z} \in \Re^{mL}$, and

$$\mathbf{d} = \begin{pmatrix} p_1 d \\ \vdots \\ p_L d \end{pmatrix}, \quad D = \begin{pmatrix} I \\ \vdots \\ I \end{pmatrix}, \quad N = \begin{pmatrix} N_1 \\ \vdots \\ N_L \end{pmatrix}, \quad M = \begin{pmatrix} M_1 & & O \\ & \ddots & \\ O & & M_L \end{pmatrix}, \quad q = \begin{pmatrix} q_1 \\ \vdots \\ q_L \end{pmatrix}.$$

Therefore, we may employ the same terminologies as in the literature on MPECs. Suppose that $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ is a feasible point of problem (7.31).

Definition 7.2 We say the MPEC-*linear independence constraint qualification* (MPEC-LICQ) to hold at $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ if the system

$$\begin{aligned} & g(x,y) \le 0,\ h(x,y) = 0, \\ & \mathbf{y} - Dy = 0,\ \mathbf{z} \ge 0, \\ & \mathbf{y} \ge 0,\ Nx + M\mathbf{y} + q + \mathbf{z} \ge 0 \end{aligned}$$

satisfies the linear independence constraint qualification (LICQ) at $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$.

Definition 7.3 We say $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ to be a *Bouligand or B-stationary* point of the MPEC (7.31) if

$$\mathbf{v}^T \begin{pmatrix} \nabla f(x^*, y^*) \\ 0 \\ \mathbf{d} \end{pmatrix} \geq 0, \qquad \forall \mathbf{v} \in \mathcal{T}(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*),$$

where $\mathcal{T}(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ stands for the tangent cone of the feasible region of problem (7.31) at $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$.

It is well-known [75] that, under the MPEC-LICQ, $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ is a B-stationary point of (7.31) if and only if there exist multiplier vectors $\lambda, \mu, \nu, \alpha, \beta$, and γ such that

$$\begin{pmatrix} \nabla_x f(x^*, y^*) \\ \nabla_y f(x^*, y^*) \\ 0 \\ \mathbf{d} \end{pmatrix} + \begin{pmatrix} \nabla_x g(x^*, y^*) \\ \nabla_y g(x^*, y^*) \\ O \\ O \end{pmatrix} \lambda + \begin{pmatrix} \nabla_x h(x^*, y^*) \\ \nabla_y h(x^*, y^*) \\ O \\ O \end{pmatrix} \mu$$
$$+ \begin{pmatrix} O \\ -D^T \\ I \\ O \end{pmatrix} \nu - \begin{pmatrix} O \\ O \\ O \\ I \end{pmatrix} \alpha - \begin{pmatrix} O \\ O \\ I \\ O \end{pmatrix} \beta - \begin{pmatrix} N^T \\ O \\ M^T \\ I \end{pmatrix} \gamma = 0, \tag{7.32}$$

$$0 \leq \lambda \perp (-g(x^*, y^*)) \geq 0, \tag{7.33}$$
$$0 \leq \alpha \perp \mathbf{z}^* \geq 0, \tag{7.34}$$
$$\mathbf{y}^* \geq 0 \ \text{ and } \ \mathbf{y}^*[i] > 0 \Rightarrow \beta[i] = 0, \tag{7.35}$$
$$(Nx^* + M\mathbf{y}^* + q + \mathbf{z}^*) \geq 0 \ \text{ and } \ (Nx^* + M\mathbf{y}^* + q + \mathbf{z}^*)[i] > 0 \Rightarrow \gamma[i] = 0, \tag{7.36}$$
$$\beta[i] \geq 0, \ \gamma[i] \geq 0, \quad \forall i \in \mathcal{I}^* := \{i \mid \ \mathbf{y}^*[i] = (Nx^* + M\mathbf{y}^* + q + \mathbf{z}^*)[i] = 0\}. \tag{7.37}$$

Moreover, for any feasible point (x, y) of problem (7.22), the point $(x, y, z_1, \cdots, z_L)$ with

$$z_\ell = \max(-(N_\ell x + M_\ell y + q_\ell), \ 0), \qquad \ell = 1, \cdots, L$$

is a feasible point of problem (7.31) and this point is eligible in the sense that, if $(x, y, \hat{z}_1, \cdots, \hat{z}_L)$ is another feasible point of (7.31), then

$$f(x, y) + \sum_{\ell=1}^{L} p_\ell d^T z_\ell \leq f(x, y) + \sum_{\ell=1}^{L} p_\ell d^T \hat{z}_\ell.$$

We further have the following theorem.

Theorem 7.5 *Suppose that (x^*, y^*) is a feasible point of problem (7.22). Let*

$$z_\ell^* = \max(-(N_\ell x^* + M_\ell y^* + q_\ell), \ 0), \qquad \ell = 1, \cdots, L \tag{7.38}$$

and

$$\mathbf{y}^* = ((y^*)^T, \cdots, (y^*)^T)^T, \qquad \mathbf{z}^* = ((z_1^*)^T, \cdots, (z_L^*)^T)^T. \tag{7.39}$$

If $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ *is a B-stationary point of the MPEC (7.31) and the MPEC-LICQ holds at* $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$*, then* (x^*, y^*) *is a stationary point of problem (7.22). Conversely, if* (x^*, y^*) *is a stationary point of problem (7.22) and the MPEC-LICQ holds at* $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$*, then* $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ *is a B-stationary point of the MPEC (7.31).*

Proof: (a) Suppose that $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ is a B-stationary point of problem (7.31) and the MPEC-LICQ holds at $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$. Then there exist multiplier vectors $\lambda, \mu, \nu, \alpha, \beta$, and γ satisfying (7.32)–(7.37). We will show that there exist multiplier vectors λ^*, μ^*, ν^*, and ξ_ℓ^*, $\ell = 1, \cdots, L$, such that conditions (7.25)–(7.28) hold. To this end, we denote

$$\nu := \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_L \end{pmatrix}, \quad \alpha := \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_L \end{pmatrix}, \quad \beta := \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_L \end{pmatrix}, \quad \gamma := \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_L \end{pmatrix},$$

and let

$$\lambda^* := \lambda, \qquad \mu^* := \mu, \tag{7.40}$$

and, for each ℓ and i,

$$\xi_\ell^*[i] \;:=\; \begin{cases} \frac{\alpha_\ell[i]}{y^*[i]}, & y^*[i] > 0,\ (N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0, \\ \frac{|\sum_{\ell=1}^L \beta_\ell[i]|}{(N_\ell x^* + M_\ell y^* + q_\ell)[i]}, & y^*[i] = 0,\ (N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{7.41}$$

$$\sigma_\ell^*[i] \;:=\; \begin{cases} 1 - \frac{\alpha_\ell[i]}{p_\ell d[i]}, & y^*[i] = 0, \\ 1, & y^*[i] > 0, \end{cases} \tag{7.42}$$

$$\nu^* \;:=\; \sum_{\ell=1}^L \beta_\ell + \begin{pmatrix} \sum_{\ell=1}^L \xi_\ell^*[1](N_\ell x^* + M_\ell y^* + q_\ell)[1] \\ \vdots \\ \sum_{\ell=1}^L \xi_\ell^*[m](N_\ell x^* + M_\ell y^* + q_\ell)[m] \end{pmatrix}. \tag{7.43}$$

Note that (7.32) can be rewritten as

$$\nabla f(x^*, y^*) + \nabla g(x^*, y^*)\lambda + \nabla h(x^*, y^*)\mu - \sum_{\ell=1}^L \begin{pmatrix} O \\ I \end{pmatrix} \nu_\ell - \sum_{\ell=1}^L \begin{pmatrix} N_\ell^T \\ O \end{pmatrix} \gamma_\ell = 0, \tag{7.44}$$

$$\nu_\ell - \beta_\ell - M_\ell^T \gamma_\ell = 0, \qquad \ell = 1, \cdots, L, \tag{7.45}$$

$$p_\ell d - \alpha_\ell - \gamma_\ell = 0, \qquad \ell = 1, \cdots, L. \tag{7.46}$$

Substituting (7.45) into (7.44), we obtain

$$\nabla f(x^*, y^*) + \nabla g(x^*, y^*)\lambda + \nabla h(x^*, y^*)\mu - \begin{pmatrix} O \\ I \end{pmatrix} \sum_{\ell=1}^{L} \beta_\ell - \sum_{\ell=1}^{L} \begin{pmatrix} N_\ell^T \\ M_\ell^T \end{pmatrix} \gamma_\ell = 0,$$

or equivalently,

$$\begin{aligned} &\nabla f(x^*, y^*) + \nabla g(x^*, y^*)\lambda + \nabla h(x^*, y^*)\mu \\ &\quad - \begin{pmatrix} O \\ I \end{pmatrix} \sum_{\ell=1}^{L} \beta_\ell - \sum_{\ell=1}^{L} \sum_{i=1}^{m} \gamma_\ell[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix} = 0. \end{aligned} \tag{7.47}$$

We next prove that

$$\gamma_\ell[i] = p_\ell d[i]\sigma_\ell^*[i] - \xi_\ell^*[i]y^*[i], \qquad \forall \ell,\ \forall i. \tag{7.48}$$

Since (x^*, y^*) is feasible to problem (7.22), we only need to consider three cases (a1)–(a3):

(a1) Suppose $y^*[i] > 0$ and $(N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0$. It then follows from (7.41), (7.42), and (7.46) that

$$p_\ell d[i]\sigma_\ell^*[i] - \xi_\ell^*[i]y^*[i] = p_\ell d[i] - \alpha_\ell[i] = \gamma_\ell[i].$$

Namely, (7.48) holds in this case.

(a2) Suppose that $y^*[i] > 0$ and $(N_\ell x^* + M_\ell y^* + q_\ell)[i] < 0$. We then have from (7.38) that $z_\ell^*[i] \geq -(N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0$. Since $\mathbf{z}^*[(\ell-1)m + i] = z_\ell^*[i] > 0$, by (7.34), we see that $\alpha_\ell[i] = \alpha[(\ell-1)m + i] = 0$. This, together with (7.41)–(7.42) and (7.46) yields

$$p_\ell d[i]\sigma_\ell^*[i] - \xi_\ell^*[i]y^*[i] = p_\ell d[i] = p_\ell d[i] - \alpha_\ell[i] = \gamma_\ell[i].$$

(a3) Suppose that $y^*[i] = 0$. It then follows from (7.41), (7.42), and (7.46) that

$$p_\ell d[i]\sigma_\ell^*[i] - \xi_\ell^*[i]y^*[i] = p_\ell d[i] - \alpha_\ell[i] = \gamma_\ell[i].$$

Therefore, (7.48) holds in any case. Substituting (7.40), (7.43), and (7.48) into (7.47), we obtain

$$\begin{aligned} 0 \;=\;& \nabla f(x^*, y^*) + \nabla g(x^*, y^*)\lambda^* + \nabla h(x^*, y^*)\mu^* - \begin{pmatrix} O \\ I \end{pmatrix} \nu^* \\ &+ \sum_{\ell=1}^{L} \sum_{i=1}^{m} \xi_\ell^*[i]\Big((N_\ell x^* + M_\ell y^* + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big) \\ &+ \sum_{\ell=1}^{L} \sum_{i=1}^{m} p_\ell d[i]\sigma_\ell^*[i] \begin{pmatrix} -N_\ell[i] \\ -M_\ell[i] \end{pmatrix}. \end{aligned}$$

So, in order to prove (7.25), we only need to show

$$\sum_{\ell=1}^{L}\sum_{i=1}^{m} p_\ell d[i]\sigma_\ell^*[i]\begin{pmatrix}-N_\ell[i]\\ -M_\ell[i]\end{pmatrix} \in \partial\theta(x^*, y^*),$$

which, by (7.29) and (7.30), is equivalent to

$$\forall \ell,\ \forall i,\quad \begin{cases}\sigma_\ell^*[i] \in [0,1], & (N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0,\\ \sigma_\ell^*[i] = 1\ , & (N_\ell x^* + M_\ell y^* + q_\ell)[i] < 0,\\ \sigma_\ell^*[i] = 0\ , & (N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0.\end{cases} \tag{7.49}$$

(a4) Suppose $(N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0$. Then, we have from (7.38) that

$$(N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i] = 0.$$

If $y^*[i] > 0$, then $\sigma_\ell^*[i] = 1 \in [0,1]$ by (7.42). If $y^*[i] = 0$, we have from (7.39) that

$$\mathbf{y}^*[(\ell-1)m + i] = y^*[i] = 0$$

and

$$(Nx^* + M\mathbf{y}^* + q + \mathbf{z}^*)[(\ell-1)m + i] = (N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i] = 0.$$

Thus, by (7.37), $\gamma_\ell[i] = \gamma[(\ell-1)m + i] \geq 0$ and hence, by (7.34) and (7.46),

$$0 \leq \alpha_\ell[i] = p_\ell d[i] - \gamma_\ell[i] \leq p_\ell d[i].$$

As a result, by (7.42), we obtain

$$\sigma_\ell^*[i] = 1 - \frac{\alpha_\ell[i]}{p_\ell d[i]} \in [0,1].$$

(a5) Suppose $(N_\ell x^* + M_\ell y^* + q_\ell)[i] < 0$. In a similar way to (a2), we can show that $\alpha_\ell[i] = 0$ and hence, by the definition (7.42) of $\sigma_\ell^*[i]$, we see that $\sigma_\ell^*[i] = 1$.

(a6) Suppose $(N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0$. It then follows from the feasibility of (x^*, y^*) to problem (7.22) that $y^*[i] = 0$ and furthermore, by (7.38),

$$\begin{aligned}(Nx^* + M\mathbf{y}^* + q + \mathbf{z}^*)[(\ell-1)m + i] &= (N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i]\\ &= (N_\ell x^* + M_\ell y^* + q_\ell)[i]\\ &> 0.\end{aligned}$$

From (7.36), we see that $\gamma_\ell[i] = \gamma[(\ell-1)m + i] = 0$. Thus, by (7.46), we obtain $\alpha_\ell[i] = p_\ell d[i]$ and hence, by the definition (7.42) of $\sigma_\ell^*[i]$,

$$\sigma_\ell^*[i] = 1 - \frac{\alpha_\ell[i]}{p_\ell d[i]} = 0.$$

This completes the proof of (7.49) and hence (7.25) holds. Note that condition (7.26) follows from (7.33) and (7.40) immediately. In addition, the definition of ξ_ℓ^* implies

$$\xi_\ell^*[i] \geq 0, \quad \xi_\ell^*[i] y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0, \qquad \forall \ell,\ \forall i. \tag{7.50}$$

Since $(x^*, y^*, z_1^*, \cdots, z_L^*)$ is feasible to problem (7.16), it then follows that, for any ℓ and i,

$$-y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] = y^*[i] z_\ell^*[i] \geq 0,$$

which, together with (7.50) indicates that (7.28) holds. We next show (7.27).

At first, for every i, we have from (7.35), (7.50), and the definition (7.43) of ν^* that

$$\begin{aligned} \nu^*[i] y^*[i] &= \sum_{\ell=1}^{L} \beta[(\ell-1)m+i] \mathbf{y}^*[(\ell-1)m+i] \\ &\quad + \sum_{\ell=1}^{L} \xi_\ell^*[i] y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] \\ &= 0. \end{aligned} \tag{7.51}$$

The rest is to prove $\nu^* \geq 0$. If $y^*[i] > 0$, we see from (7.51) that $\nu^*[i] = 0$. If $y^*[i] = 0$, we let

$$\mathcal{I}(i) := \{\ell \mid (N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0\}.$$

It follows from (7.41) and (7.43) that

$$\begin{aligned} \nu^*[i] &= \sum_{\ell=1}^{L} \beta_\ell[i] + \sum_{\ell \in \mathcal{I}(i)} \xi_\ell^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] \\ &= \sum_{\ell=1}^{L} \beta_\ell[i] + n_1 \Big| \sum_{\ell=1}^{L} \beta_\ell[i] \Big|, \end{aligned} \tag{7.52}$$

where n_1 denotes the cardinality of the index set $\mathcal{I}(i)$. If $\mathcal{I}(i) \neq \emptyset$, then $\nu^*[i] \geq 0$ by (7.52). If $\mathcal{I}(i) = \emptyset$, it follows that

$$(N_\ell x^* + M_\ell y^* + q_\ell)[i] \leq 0, \qquad \forall \ell$$

and then, by (7.38) and (7.39), there hold

$$\mathbf{y}^*[(\ell-1)m+i] = y^*[i] = 0$$

and

$$(Nx^* + M\mathbf{y}^* + q + \mathbf{z}^*)[(\ell-1)m+i] = (N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i] = 0$$

for every ℓ. From (7.37) and (7.52), we obtain $\nu^*[i] = \sum_{\ell=1}^L \beta_\ell[i] \geq 0$. Thus, we have $\nu^* \geq 0$ in any case. This completes the proof of (7.27) and hence (x^*, y^*) is a stationary point of problem (7.22).

(b) Suppose that (x^*, y^*) is a stationary point of problem (7.22) and the MPEC-LICQ holds at the point $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$. By the stationarity of (x^*, y^*), there exist multiplier vectors λ^*, μ^*, ν^*, and $\xi_\ell^*,\ \ell = 1, \cdots, L$, satisfying conditions (7.25)–(7.28). We will prove that $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ is a B-stationary point of problem (7.31), i.e., there exist multiplier vectors $\lambda, \mu, \nu, \alpha, \beta$, and γ such that (7.32)–(7.37) hold.

By (7.29) and (7.30), condition (7.25) means that there exist multiplier vectors $\sigma_\ell^*,\ \ell = 1, \cdots, L$, satisfying (7.49) and

$$\begin{aligned} 0 \;=\; & \nabla f(x^*, y^*) + \nabla g(x^*, y^*)\lambda^* + \nabla h(x^*, y^*)\mu^* - \begin{pmatrix} O \\ I \end{pmatrix} \nu^* \\ & + \sum_{\ell=1}^{L}\sum_{i=1}^{m} \xi_\ell^*[i]\Big((N_\ell x^* + M_\ell y^* + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big) \\ & + \sum_{\ell=1}^{L}\sum_{i=1}^{m} p_\ell d[i]\sigma_\ell^*[i] \begin{pmatrix} -N_\ell[i] \\ -M_\ell[i] \end{pmatrix}. \end{aligned} \tag{7.53}$$

Let

$$\lambda := \lambda^*, \tag{7.54}$$

$$\mu := \mu^*, \tag{7.55}$$

$$\alpha_\ell[i] := p_\ell d[i](1 - \sigma_\ell^*[i]) + \xi_\ell^*[i]y^*[i], \tag{7.56}$$

$$\beta_\ell[i] := \frac{1}{L}\nu^*[i] - \xi_\ell^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i], \tag{7.57}$$

$$\gamma_\ell[i] := p_\ell d[i]\sigma_\ell^*[i] - \xi_\ell^*[i]y^*[i], \tag{7.58}$$

$$\nu_\ell := \beta_\ell + M_\ell^T \gamma_\ell, \tag{7.59}$$

and

$$\alpha := \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_L \end{pmatrix}, \quad \beta := \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_L \end{pmatrix}, \quad \gamma := \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_L \end{pmatrix}, \quad \nu := \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_L \end{pmatrix}.$$

It is easy to verify (7.45)–(7.47) from (7.53)–(7.58) and so we obtain (7.32). In addition, condition (7.33) follows from (7.26) and (7.54) immediately. We next show (7.34)–(7.37). Note that, by (7.28) and (7.49), we have $\alpha \geq 0$. Moreover, by (7.38) and (7.39), we have $\mathbf{y}^* \geq 0$, $\mathbf{z}^* \geq 0$, and $N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^* \geq 0$ for any ℓ, which in turn implies $Nx^* + M\mathbf{y}^* + q + \mathbf{z}^* \geq 0$. Take an arbitrary index j with $1 \leq j \leq mL$. Then, there exist ℓ and i such that

$$1 \leq \ell \leq L, \quad 1 \leq i \leq m, \quad j = (\ell - 1)m + i.$$

(b1) Suppose $\mathbf{z}^*[j] > 0$. This means $z_\ell^*[i] > 0$. It then follows from (7.38) that $(N_\ell x^* + M_\ell y^* + q_\ell)[i] < 0$ and hence we have

$$\sigma_\ell^*[i] = 1, \qquad \xi_\ell^*[i] y^*[i] = 0$$

from (7.49) and (7.28), respectively. This indicates $\alpha[j] = \alpha_\ell[i] = 0$ and so (7.34) holds.

(b2) It is easy to see from (7.27) and (7.28) that

$$\beta[j]\mathbf{y}[j] = \beta_\ell[i] y^*[i] = \frac{1}{L}\nu^*[i] y^*[i] - \xi_\ell^*[i] y^*[i] (N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0,$$

which indicates that (7.35) holds.

(b3) Suppose $(Nx^* + M\mathbf{y}^* + q + \mathbf{z}^*)[j] > 0$. This means that $(N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i] > 0$. It then follows from (7.38) that $(N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0$ and hence $y^*[i] = 0$. This indicates $\gamma[j] = \gamma_\ell[i] = 0$ and therefore (7.36) holds.

(b4) Let $\mathcal{I}^*$ be defined as in (7.37) and suppose $j \in \mathcal{I}^*$. It is obvious that

$$\gamma[j] = p_\ell d[i] \sigma_\ell^*[i] \geq 0.$$

On the other hand, since $(Nx^* + M\mathbf{y}^* + q + \mathbf{z}^*)[j] = 0$ implies that $(N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i] = 0$, we see from (7.38) that $(N_\ell x^* + M_\ell y^* + q_\ell)[i] \leq 0$. Therefore,

$$\beta[j] = \beta_\ell[i] = \frac{1}{L}\nu^*[i] - \xi_\ell^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] \geq \frac{1}{L}\nu^*[i] \geq 0.$$

This indicates that (7.37) holds.

Thus, the multiplier vectors defined by (7.54)–(7.58) satisfy conditions (7.32)–(7.37) and hence $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ is a B-stationary point of problem (7.31). ∎

7.4 Smoothed Penalty Methods for Discrete Problems

We have established the equivalence between problems (7.16) and (7.22). Note that, although (7.22) is no longer an SMPEC and the function θ is convex, this problem may not be easy to deal with because firstly, the objective function is not differentiable everywhere, and secondly, since L is usually very large in practice, problem (7.22) has a great many constraints. We will propose a smoothed penalty method for solving problem (7.22) in this section.

The following functions will be used later on: Let ϵ be a nonnegative constant. The functions $\phi_\epsilon : \Re \to [0, +\infty)$ and $\psi_\epsilon : \Re \to [0, +\infty)$ are defined by

$$\phi_\epsilon(t) := \frac{\sqrt{t^2+\epsilon^2}+t}{2}$$

and

$$\psi_\epsilon(t) := \sqrt{t^2+\epsilon^2},$$

respectively. It is obvious that, for any fixed t,

$$\lim_{\epsilon \downarrow 0} \phi_\epsilon(t) = \phi_0(t) = \max\{t, 0\}, \qquad \lim_{\epsilon \downarrow 0} \psi_\epsilon(t) = \psi_0(t) = |t|.$$

However, unlike the limit functions ϕ_0 and ψ_0, both ϕ_ϵ and ψ_ϵ are differentiable everywhere for each $\epsilon > 0$.

7.4.1 Smoothed penalty method (I)

By means of the differentiable function ϕ_ϵ and with the help of a smoothed penalty technique, we obtain the following smooth approximation of problem (7.22):

$$\begin{array}{ll} \text{minimize} & f(x,y) + \vartheta_\epsilon(x,y) + \rho\delta_\epsilon(x,y) \\ \text{subject to} & g(x,y) \le 0, \;\; h(x,y) = 0, \quad y \ge 0, \end{array} \tag{7.60}$$

where $\rho > 0$ is a penalty parameter and

$$\vartheta_\epsilon(x,y) := \sum_{\ell=1}^{L}\sum_{i=1}^{m} p_\ell d[i]\phi_\epsilon(-(N_\ell x + M_\ell y + q_\ell)[i]), \tag{7.61}$$

$$\delta_\epsilon(x,y) := \sum_{\ell=1}^{L}\sum_{i=1}^{m} \phi_\epsilon(y[i](N_\ell x + M_\ell y + q_\ell)[i]). \tag{7.62}$$

Note that, for any (x,y),

$$\begin{aligned} \vartheta_0(x,y) &= \theta(x,y), \\ \delta_0(x,y) &= \sum_{\ell=1}^{L}\sum_{i=1}^{m} \max\{y[i](N_\ell x + M_\ell y + q_\ell)[i],\ 0\}. \end{aligned}$$

Hence, when $\epsilon = 0$ and $\rho = \bar{\rho} > 0$, problem (7.60) reduces to

$$\begin{array}{ll} \text{minimize} & f(x,y) + \theta(x,y) + \bar{\rho}\delta_0(x,y) \\ \text{subject to} & g(x,y) \le 0,\ h(x,y) = 0,\ \ y \ge 0. \end{array} \tag{7.63}$$

For problem (7.63), we will use a similar definition of stationarity to problem (7.22): We say $(x^*, y^*) \in \Re^{n+m}$ to be *stationary* to problem (7.63) if it is feasible and there exist Lagrange multiplier vectors λ, μ, and ν such that

$$\begin{aligned} 0 \in\ & \nabla f(x^*, y^*) + \partial\theta(x^*, y^*) + \bar{\rho}\partial\delta_0(x^*, y^*) \\ & +\nabla g(x^*, y^*)\lambda + \nabla h(x^*, y^*)\mu - \begin{pmatrix} O \\ I \end{pmatrix} \nu, \end{aligned} \tag{7.64}$$

$$0 \le \lambda \perp (-g(x^*, y^*)) \ge 0, \tag{7.65}$$

$$0 \le \nu \perp \quad y^* \ge 0, \tag{7.66}$$

where $\partial\delta_0$ denotes Clarke subdifferential operator [18]. Let

$$\delta_{\ell,i}(x,y) := \max\Big\{ y[i](N_\ell x + M_\ell y + q_\ell)[i],\ 0 \Big\}, \qquad \forall \ell,\ \forall i.$$

Then, since the functions $\delta_{\ell,i}$ are Clarke regular [18], we have

$$\begin{aligned} &\partial\delta_{\ell,i}(x,y) \\ &= \begin{cases} \text{co}\Big\{ (N_\ell x + M_\ell y + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}, 0 \Big\}, & y[i](N_\ell x + M_\ell y + q_\ell)[i] = 0 \\ \Big\{ (N_\ell x + M_\ell y + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix} \Big\}, & y[i](N_\ell x + M_\ell y + q_\ell)[i] > 0 \\ \Big\{ 0 \Big\}, & y[i](N_\ell x + M_\ell y + q_\ell)[i] < 0 \end{cases} \end{aligned} \tag{7.67}$$

and

$$\partial\delta_0(x,y) = \sum_{\ell=1}^{L} \sum_{i=1}^{m} \partial\delta_{\ell,i}(x,y), \tag{7.68}$$

where $\partial\delta_{\ell,i}$ denotes Clarke subdifferential operator of $\delta_{\ell,i}$.

Theorem 7.6 *Let (x^*, y^*) be a stationary point of problem (7.22). Then, (x^*, y^*) is a stationary point of problem (7.63) for any $\bar{\rho}$ sufficiently large. Conversely, if (x^*, y^*) is a stationary point of problem (7.63), and $\delta_0(x^*, y^*) = 0$, i.e., $(x^*, y^*) \in \mathcal{F}_3$, then (x^*, y^*) is stationary to (7.22).*

Proof: (a) Suppose (x^*, y^*) is a stationary point of problem (7.22). We will show that, when $\bar{\rho}$ is sufficiently large, (x^*, y^*) is a stationary point of problem (7.63). By the stationarity of (x^*, y^*) to problem (7.22), there exist multiplier vectors λ^*, μ^*, ν^*, and $\xi_\ell^*,\ \ell = 1, \cdots, L$, satisfying conditions (7.25)–(7.28). Let

$$\lambda := \lambda^*, \qquad \mu := \mu^*, \qquad \nu := \nu^*.$$

Then (7.65) and (7.66) follow from (7.26) and (7.27) immediately. Comparing (7.64) with (7.25), in order to finish the proof, we only need to show that, when $\bar{\rho}$ is sufficiently large,

$$\sum_{\ell=1}^{L}\sum_{i=1}^{m} \xi_\ell^*[i]\Big((N_\ell x^* + M_\ell y^* + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big) \in \bar{\rho}\partial\delta_0(x^*, y^*).$$

By (7.67) and (7.68), it is sufficient to show that, when $\bar{\rho}$ is sufficiently large,

$$\forall \ell,\ \forall i, \quad \begin{cases} \xi_\ell^*[i] \in [0, \bar{\rho}], & y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0, \\ \xi_\ell^*[i] = \bar{\rho}\ , & y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0, \\ \xi_\ell^*[i] = 0\ , & y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] < 0. \end{cases}$$

Indeed, this follows from (7.28) and the fact that $y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] \le 0$ for each ℓ and i.

(b) Suppose (x^*, y^*) is a stationary point of problem (7.63) and $\delta_0(x^*, y^*) = 0$. Then, there exist multiplier vectors λ, μ, and ν satisfying (7.64)–(7.66). Note that $(x^*, y^*) \in \mathcal{F}_3$ implies

$$y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] \le 0, \qquad \forall \ell,\ \forall i. \tag{7.69}$$

It then follows that

$$\partial\delta_{\ell,i}(x^*, y^*) = \begin{cases} \mathrm{co}\Big\{(N_\ell x^* + M_\ell y^* + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}, 0\Big\}, & y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0 \\ \Big\{0\Big\}, & y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] < 0 \end{cases}$$

for any ℓ and i, and

$$\partial\delta_0(x^*, y^*) = \sum_{\ell=1}^{L}\sum_{i=1}^{m} \partial\delta_{\ell,i}(x^*, y^*).$$

Condition (7.64) means that there exist multiplier vectors $\xi_\ell,\ \ell = 1, \cdots, L$, such that

$$\forall \ell,\ \forall i, \quad \begin{cases} \xi_\ell[i] \in [0, \bar{\rho}], & y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0 \\ \xi_\ell[i] = 0\ , & y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] < 0 \end{cases} \tag{7.70}$$

and

$$0 \in \nabla f(x^*, y^*) + \partial\theta(x^*, y^*) + \nabla g(x^*, y^*)\lambda + \nabla h(x^*, y^*)\mu - \begin{pmatrix} O \\ I \end{pmatrix} \nu$$
$$+ \sum_{\ell=1}^{L} \sum_{i=1}^{m} \xi_\ell[i] \Big((N_\ell x^* + M_\ell y^* + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix} \Big). \qquad (7.71)$$

Let

$$\lambda^* := \lambda, \quad \mu^* := \mu, \quad \nu^* := \nu$$

and

$$\xi_\ell^* := \xi_\ell, \qquad \ell = 1, \cdots, L.$$

Then (7.25)–(7.27) follow from (7.71) and (7.65)–(7.66), and (7.28) follows from (7.69)–(7.70). Therefore, (x^*, y^*) is a stationary point of problem (7.22). ∎

We then have the following algorithm for problem (7.22).

Algorithm SP-I:

Step 1: Choose $\epsilon_0 > 0$ and $\rho_0 > 0$. Set $k := 0$.

Step 2: Solve problem (7.60) with $\epsilon = \epsilon_k$ and $\rho = \rho_k$ to get a stationary point (x^k, y^k) and go to Step 3.

Step 3: If a stopping rule is satisfied, then terminate. Otherwise, choose $\epsilon_{k+1} \in (0, \epsilon_k)$ and $\rho_{k+1} \geq \rho_k$. Go to Step 2 with $k := k + 1$.

In what follows, we suppose that the sequences $\{\epsilon_k\}$ and $\{\rho_k\}$ satisfy

$$\lim_{k\to\infty} \epsilon_k = 0, \qquad \lim_{k\to\infty} \rho_k = \bar{\rho}, \qquad (7.72)$$

where $\bar{\rho}$ is a sufficiently large constant. Recall that $\mathcal{F}_3$ denotes the feasible region of problem (7.20), which is the same as that of problem (7.22).

We next investigate the limiting behavior of the sequence $\{(x^k, y^k)\}$ generated by Algorithm SP-I. The convergence result can be stated as follows.

Theorem 7.7 *Suppose that Algorithm SP-I generates a sequence $\{(x^k, y^k)\}$ of stationary points of (7.60) with $\epsilon = \epsilon_k$ and $\rho = \rho_k$. For any accumulation point (x^*, y^*) of the sequence $\{(x^k, y^k)\}$, if the system*

$$g(x, y) \leq 0, \quad h(x, y) = 0, \quad y \geq 0 \qquad (7.73)$$

satisfies the Mangasarian-Fromovitz constraint qualification (MFCQ) at (x^*, y^*)*, then* (x^*, y^*) *is a stationary point of problem (7.63). Furthermore, if* $\delta_0(x^*, y^*) = 0$*, then* (x^*, y^*) *is a stationary point of problem (7.22).*

Proof: Assume without loss of generality that $\lim_{k\to\infty}(x^k, y^k) = (x^*, y^*)$. We will show that (x^*, y^*) is stationary to problem (7.63), i.e., there exist multiplier vectors λ, μ, and ν such that (7.64)–(7.66) hold.

First of all, by the stationarity of (x^k, y^k) for problem (7.60) with $\epsilon = \epsilon_k$ and $\rho = \rho_k$, there exist Lagrange multiplier vectors λ^k, μ^k, and ν^k such that

$$\begin{aligned} &\nabla f(x^k, y^k) + \nabla \vartheta_{\epsilon_k}(x^k, y^k) + \rho_k \nabla \delta_{\epsilon_k}(x^k, y^k) \\ &\qquad + \nabla g(x^k, y^k)\lambda^k + \nabla h(x^k, y^k)\mu^k - \begin{pmatrix} O \\ I \end{pmatrix} \nu^k = 0, \end{aligned} \tag{7.74}$$

$$0 \le \lambda^k \perp (-g(x^k, y^k)) \ge 0, \tag{7.75}$$

$$0 \le \nu^k \perp \quad y^k \ge 0. \tag{7.76}$$

Note that, by (7.61) and (7.62),

$$\nabla \vartheta_{\epsilon_k}(x^k, y^k) = \sum_{\ell=1}^{L}\sum_{i=1}^{m} p_\ell d[i] \phi'_{\epsilon_k}(-(N_\ell x^k + M_\ell y^k + q_\ell)[i]) \begin{pmatrix} -N_\ell[i] \\ -M_\ell[i] \end{pmatrix}, \tag{7.77}$$

$$\begin{aligned} \nabla \delta_{\epsilon_k}(x^k, y^k) = &\sum_{\ell=1}^{L}\sum_{i=1}^{m} \phi'_{\epsilon_k}(y^k[i](N_\ell x^k + M_\ell y^k + q_\ell)[i]) \\ &\Big((N_\ell x^k + M_\ell y^k + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^k[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big), \end{aligned} \tag{7.78}$$

where

$$\phi'_{\epsilon_k}(t) = \frac{1}{2}\Big(\frac{t}{\sqrt{t^2 + \epsilon_k^2}} + 1\Big), \qquad \forall\, t \in \Re. \tag{7.79}$$

We can then rewrite (7.74) as

$$\begin{aligned} &-\nabla f(x^k, y^k) - \nabla g(x^k, y^k)\lambda^k - \nabla h(x^k, y^k)\mu^k + \begin{pmatrix} O \\ I \end{pmatrix} \nu^k \\ = &\sum_{\ell=1}^{L}\sum_{i=1}^{m} \xi_\ell^k[i]\Big((N_\ell x^k + M_\ell y^k + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^k[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big) \\ &+ \sum_{\ell=1}^{L}\sum_{i=1}^{m} p_\ell d[i] \sigma_\ell^k[i] \begin{pmatrix} -N_\ell[i] \\ -M_\ell[i] \end{pmatrix}, \end{aligned} \tag{7.80}$$

where

$$\xi_\ell^k[i] := \rho_k \phi'_{\epsilon_k}(y^k[i](N_\ell x^k + M_\ell y^k + q_\ell)[i]), \tag{7.81}$$

$$\sigma_\ell^k[i] := \phi'_{\epsilon_k}(-(N_\ell x^k + M_\ell y^k + q_\ell)[i]). \tag{7.82}$$

We next prove that the sequences $\{\lambda^k\}$, $\{\mu^k\}$, and $\{\nu^k\}$ are bounded. For the contradiction purpose, let

$$\tau_k := \sum_{i=1}^{m}(\lambda^k[i] + |\mu^k[i]| + \nu^k[i]), \tag{7.83}$$

and suppose $\lim_{k\to\infty}\tau_k = +\infty$. Taking a subsequence if necessary, we may assume that the limits

$$\lambda'[i] := \lim_{k\to\infty}\frac{\lambda^k[i]}{\tau_k}, \quad \mu'[i] := \lim_{k\to\infty}\frac{\mu^k[i]}{\tau_k}, \quad \nu'[i] := \lim_{k\to\infty}\frac{\nu^k[i]}{\tau_k}, \qquad i = 1,\cdots,m$$

exist. It is clear from (7.83) that

$$\sum_{i=1}^{m}(\lambda'[i] + |\mu'[i]| + \nu'[i]) = 1.$$

It follows from (7.72), (7.79), and (7.81)–(7.82) that both $\{\xi_\ell^k[i]\}$ and $\{\sigma_\ell^k[i]\}$ are bounded for each ℓ and each i. Thus, dividing (7.80) by τ_k and taking a limit, we get

$$-\nabla g(x^*,y^*)\lambda' - \nabla h(x^*,y^*)\mu' + \begin{pmatrix} O \\ I \end{pmatrix}\nu' = 0.$$

Furthermore, taking (7.75) and (7.76) into account, we obtain $\lambda' \ge 0, \nu' \ge 0$, and

$$\begin{aligned} \lambda'[i] &= 0, \qquad i \notin \mathcal{I}_g(x^*,y^*), \\ \nu'[i] &= 0, \qquad i \notin \mathcal{I}_\pi(x^*,y^*), \end{aligned}$$

where $\pi : \Re^{n+m} \to \Re^m$ is given by $\pi(x,y) = y$. Thus we have

$$\begin{aligned} \sum_{i\in\mathcal{I}_g(x^*,y^*)}\lambda'[i] + \sum_{i=1}^{s_2}\mu'[i] + \sum_{i\in\mathcal{I}_\pi(x^*,y^*)}\nu'[i] &= 1, \\ \lambda'[i] \ge 0, \qquad i \in \mathcal{I}_g(x^*,y^*)&, \\ \nu'[i] \ge 0, \qquad i \in \mathcal{I}_\pi(x^*,y^*)&, \end{aligned}$$

and

$$-\sum_{i\in\mathcal{I}_g(x^*,y^*)}\lambda'[i]\nabla g_i(x^*,y^*) - \sum_{i=1}^{s_2}\mu'[i]\nabla h_i(x^*,y^*) + \sum_{i\in\mathcal{I}_\pi(x^*,y^*)}\nu'[i]\begin{pmatrix} 0 \\ e_i \end{pmatrix} = 0.$$

This contradicts the assumption that the system (7.73) satisfies the MFCQ at (x^*,y^*). Hence all the sequences $\{\lambda^k\}$, $\{\mu^k\}$, and $\{\nu^k\}$ are bounded. Since $\{\xi_\ell^k[i]\}$ and $\{\sigma_\ell^k[i]\}$ are bounded for any ℓ and i, we may assume, without loss of generality, that the following limits exist:

$$\lambda := \lim_{k\to\infty} \lambda^k, \quad \mu := \lim_{k\to\infty} \mu^k, \quad \nu := \lim_{k\to\infty} \nu^k,$$

and

$$\xi_\ell := \lim_{k\to\infty} \xi_\ell^k, \qquad \sigma_\ell := \lim_{k\to\infty} \sigma_\ell^k, \qquad \forall \ell.$$

Taking a limit in (7.75), (7.76), and (7.80), we obtain (7.65), (7.66), and

$$\begin{aligned} & -\nabla f(x^*, y^*) - \nabla g(x^*, y^*)\lambda - \nabla h(x^*, y^*)\mu + \begin{pmatrix} O \\ I \end{pmatrix} \nu \\ = \; & \sum_{\ell=1}^{L}\sum_{i=1}^{m} \xi_\ell[i]\Big((N_\ell x^* + M_\ell y^* + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big) \\ & + \sum_{\ell=1}^{L}\sum_{i=1}^{m} p_\ell d[i] \sigma_\ell[i] \begin{pmatrix} -N_\ell[i] \\ -M_\ell[i] \end{pmatrix}. \end{aligned} \tag{7.84}$$

Thus, in order to show that (x^*, y^*) is a stationary point of problem (7.63), we only need to prove that the vector on the right-hand side of (7.84) belongs to the set $\bar\rho\partial\delta_0(x^*, y^*) + \partial\theta(x^*, y^*)$.

(i) We first prove that

$$\sum_{\ell=1}^{L}\sum_{i=1}^{m} \xi_\ell[i]\Big((N_\ell x^* + M_\ell y^* + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big) \in \bar\rho\partial\delta_0(x^*, y^*). \tag{7.85}$$

By (7.68), it is sufficient to show that, for any ℓ and i,

$$\xi_\ell[i]\Big((N_\ell x^* + M_\ell y^* + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big) \in \bar\rho\partial\delta_{\ell,i}(x^*, y^*),$$

which, by (7.67), is equivalent to

$$\begin{cases} \xi_\ell[i] \in [0, \bar\rho], & y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0, \\ \xi_\ell[i] = \bar\rho\;, & y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0, \\ \xi_\ell[i] = 0\;, & y^*[i](N_\ell x^* + M_\ell y^* + q_\ell)[i] < 0. \end{cases} \tag{7.86}$$

In fact, we can obtain (7.86) immediately from (7.72) and the facts that

$$\lim_{k\to\infty}(x^k, y^k) = (x^*, y^*)$$

and

$$\xi_\ell^k[i] = \frac{\rho_k}{2}\Bigg(\frac{y^k[i](N_\ell x^k + M_\ell y^k + q_\ell)[i]}{\sqrt{(y^k[i](N_\ell x^k + M_\ell y^k + q_\ell)[i])^2 + \epsilon_k^2}} + 1\Bigg),$$

and hence (7.85) must hold.

(ii) We next prove that

$$\sum_{\ell=1}^{L}\sum_{i=1}^{m} p_\ell d[i]\sigma_\ell[i]\begin{pmatrix}-N_\ell[i]\\-M_\ell[i]\end{pmatrix}\in\partial\theta(x^*,y^*).$$

By (7.30), it is enough to show

$$\sigma_\ell[i]\begin{pmatrix}-N_\ell[i]\\-M_\ell[i]\end{pmatrix}\in\partial\theta_{\ell,i}(x^*,y^*),\qquad\forall\ell,\ \forall i. \tag{7.87}$$

There are three cases:

(iia) Suppose $(N_\ell x^*+M_\ell y^*+q_\ell)[i]=0$. We then have from (7.79) and (7.82) that

$$0\le\sigma_\ell^k[i]\le 1,\qquad\forall k.$$

Passing to the limit yields $0\le\sigma_\ell[i]\le 1$ and hence

$$\sigma_\ell[i]\begin{pmatrix}-N_\ell[i]\\-M_\ell[i]\end{pmatrix}\in\mathrm{co}\Big\{\begin{pmatrix}-N_\ell[i]\\-M_\ell[i]\end{pmatrix},0\Big\}=\partial\theta_{\ell,i}(x^*,y^*),$$

where the equality follows from (7.29).

(iib) Suppose $(N_\ell x^*+M_\ell y^*+q_\ell)[i]<0$. Note that

$$\sigma_\ell^k[i]=\frac{1}{2}\Big(\frac{-(N_\ell x^k+M_\ell y^k+q_\ell)[i]}{\sqrt{((N_\ell x^k+M_\ell y^k+q_\ell)[i])^2+\epsilon_k^2}}+1\Big),\qquad\forall k.$$

Taking a limit in the above equality, we obtain $\sigma_\ell[i]=1$ immediately and so

$$\sigma_\ell[i]\begin{pmatrix}-N_\ell[i]\\-M_\ell[i]\end{pmatrix}=\begin{pmatrix}-N_\ell[i]\\-M_\ell[i]\end{pmatrix}\in\partial\theta_{\ell,i}(x^*,y^*).$$

(iic) Suppose $(N_\ell x^*+M_\ell y^*+q_\ell)[i]>0$. It is easy to show that, for any k,

$$0\le\sigma_\ell^k[i]\le\frac{\epsilon_k}{2\Big(\sqrt{((N_\ell x^k+M_\ell y^k+q_\ell)[i])^2+\epsilon_k^2}+(N_\ell x^k+M_\ell y^k+q_\ell)[i]\Big)}.$$

Letting $k\to\infty$, we see that $\sigma_\ell[i]=0$ and so

$$\sigma_\ell[i]\begin{pmatrix}-N_\ell[i]\\-M_\ell[i]\end{pmatrix}=0\in\partial\theta_{\ell,i}(x^*,y^*).$$

Consequently, (7.87) holds in each case. (i) and (ii) indicate that the vector on the right-hand side of (7.84) belongs to the set $\bar\rho\partial\delta_0(x^*,y^*)+\partial\theta(x^*,y^*)$. This completes the proof of the first part of the theorem. The second half readily follows from Theorem 7.6. ∎

7.4.2 Smoothed penalty method (II)

In the last subsection, making use of the function ϕ_ϵ, we obtained the smooth problem (7.60). By applying a similar smoothed penalty technique to the constraints of problem (7.60), we may further get the following unconstrained smooth problem:

$$\text{minimize} \qquad f(x,y) + \vartheta_{\epsilon_k}(x,y) + \rho_k \delta_{\epsilon_k}(x,y) + \rho_k r_{\epsilon_k}(x,y), \tag{7.88}$$

where

$$r_{\epsilon_k}(x,y) := \sum_{i=1}^{s_1} \phi_{\epsilon_k}(g_i(x,y)) + \sum_{i=1}^{s_2} \psi_{\epsilon_k}(h_i(x,y)) + \sum_{i=1}^{m} \phi_{\epsilon_k}(-y[i]). \tag{7.89}$$

Corresponding to problem (7.63), we have a nonsmooth approximation of problem (7.22):

$$\text{minimize} \qquad f(x,y) + \theta(x,y) + \bar{\rho}\delta_0(x,y) + \bar{\rho} r_0(x,y), \tag{7.90}$$

where

$$r_0(x,y) := \sum_{i=1}^{s_1} \max\{(g_i(x,y)),\ 0\} + \sum_{i=1}^{s_2} |h_i(x,y)| + \sum_{i=1}^{m} \max\{-y[i],\ 0\}.$$

We say $(x^*, y^*) \in \Re^{n+m}$ to be *stationary* to problem (7.90) if

$$0 \in \nabla f(x^*,y^*) + \partial\theta(x^*,y^*) + \bar{\rho}\partial\delta_0(x^*,y^*) + \bar{\rho}\partial r_0(x^*,y^*). \tag{7.91}$$

Theorem 7.8 *Let (x^*, y^*) be a stationary point of problem (7.22). Then, (x^*, y^*) is a stationary point of problem (7.90) for any $\bar{\rho}$ sufficiently large. Conversely, if (x^*, y^*) is a stationary point of problem (7.90), and $\delta_0(x^*,y^*) + r_0(x^*,y^*) = 0$, i.e., $(x^*,y^*) \in \mathcal{F}_3$, then (x^*, y^*) is stationary to problem (7.22).*

The proof of this theorem is similar to that of Theorem 7.6 and so it is omitted here.

A new algorithm for solving problem (7.22) can be stated as follows.

Algorithm SP-II:

Step 1: Choose $\epsilon_0 > 0$ and $\rho_0 > 0$. Set $k := 0$.

Step 2: Solve problem (7.88) to get a stationary point, say (x^k, y^k), and go to Step 3.

Step 3: If a stopping rule is satisfied, then terminate. Otherwise, choose $\epsilon_{k+1} \in (0, \epsilon_k)$ and $\rho_{k+1} \geq \rho_k$. Go to Step 2 with $k := k+1$.

We suppose that the sequences $\{\epsilon_k\}$ and $\{\rho_k\}$ also satisfy condition (7.72). Then, we have the following convergence result for Algorithm SP-II.

Theorem 7.9 *Suppose that Algorithm SP-II generates a sequence $\{(x^k, y^k)\}$ of stationary points of (7.88). Then any accumulation point (x^*, y^*) of the sequence $\{(x^k, y^k)\}$ is a stationary point of problem (7.90). Furthermore, if $\delta_0(x^*, y^*) + r_0(x^*, y^*) = 0$, then (x^*, y^*) is a stationary point of problem (7.22).*

Proof: Assume without loss of generality that $\lim_{k\to\infty}(x^k, y^k) = (x^*, y^*)$. Since the second part of the theorem follows from the first part and Theorem 7.8 directly, we only prove the first part of the theorem, namely, we will show (7.91).

Firstly, it follows from the stationarity of (x^k, y^k) that

$$\nabla f(x^k, y^k) + \nabla\vartheta_{\epsilon_k}(x^k, y^k) + \rho_k\nabla\delta_{\epsilon_k}(x^k, y^k) + \rho_k\nabla r_{\epsilon_k}(x^k, y^k) = 0. \tag{7.92}$$

Note that, by (7.89),

$$\begin{aligned}\nabla r_{\epsilon_k}(x^k, y^k) \;&=\; \sum_{i=1}^{s_1}\phi'_{\epsilon_k}(g_i(x^k, y^k))\nabla g_i(x^k, y^k)\\ &\quad+\sum_{i=1}^{s_2}\psi'_{\epsilon_k}(h_i(x^k, y^k))\nabla h_i(x^k, y^k) - \sum_{i=1}^{m}\phi'_{\epsilon_k}(-y^k[i])\begin{pmatrix}0\\ e_i\end{pmatrix},\end{aligned} \tag{7.93}$$

where ϕ'_{ϵ_k} is given by (7.79) and

$$\psi'_{\epsilon_k}(t) = \frac{t}{\sqrt{t^2+\epsilon_k^2}}, \qquad \forall\ t\in\Re. \tag{7.94}$$

Substituting (7.77), (7.78), and (7.93) into (7.92), we obtain

$$\begin{aligned}-\nabla f(x^k, y^k) \;&=\; \nabla g(x^k, y^k)\lambda^k + \nabla h(x^k, y^k)\mu^k - \begin{pmatrix}O\\ I\end{pmatrix}\nu^k\\ &\quad+\sum_{\ell=1}^{L}\sum_{i=1}^{m}\xi_\ell^k[i]\Big((N_\ell x^k + M_\ell y^k + q_\ell)[i]\begin{pmatrix}0\\ e_i\end{pmatrix} + y^k[i]\begin{pmatrix}N_\ell[i]\\ M_\ell[i]\end{pmatrix}\Big)\\ &\quad+\sum_{\ell=1}^{L}\sum_{i=1}^{m}p_\ell d[i]\sigma_\ell^k[i]\begin{pmatrix}-N_\ell[i]\\ -M_\ell[i]\end{pmatrix},\end{aligned} \tag{7.95}$$

where the multiplier vectors $\lambda^k, \mu^k, \nu^k, \xi_\ell^k$, and σ_ℓ^k are given by

$$\begin{aligned}\lambda^k[i] \;&:=\; \rho_k\phi'_{\epsilon_k}(g_i(x^k, y^k)), && i = 1,\cdots,s_1,\\ \mu^k[i] \;&:=\; \rho_k\psi'_{\epsilon_k}(h_i(x^k, y^k)), && i = 1,\cdots,s_2,\\ \nu^k[i] \;&:=\; \rho_k\phi'_{\epsilon_k}(-y^k[i]), && i = 1,\cdots,m,\\ \xi_\ell^k[i] \;&:=\; \rho_k\phi'_{\epsilon_k}(y^k[i](N_\ell x^k + M_\ell y^k + q_\ell)[i]), && i = 1,\cdots,m,\ \ell = 1,\cdots,L,\\ \sigma_\ell^k[i] \;&:=\; \phi'_{\epsilon_k}(-(N_\ell x^k + M_\ell y^k + q_\ell)[i]), && i = 1,\cdots,m,\ \ell = 1,\cdots,L,\end{aligned}$$

respectively. Since $\{\rho_k\}$ is bounded, it follows from (7.79) and (7.94) that all the multiplier vectors are bounded. Therefore, we may assume without loss of generality that the following limits exist:

$$\lambda := \lim_{k\to\infty} \lambda^k, \quad \mu := \lim_{k\to\infty} \mu^k, \quad \nu := \lim_{k\to\infty} \nu^k,$$

and

$$\xi_\ell := \lim_{k\to\infty} \xi_\ell^k, \qquad \sigma_\ell := \lim_{k\to\infty} \sigma_\ell^k, \qquad \forall \ell.$$

Taking a limit in (7.95) yields

$$\begin{aligned} -\nabla f(x^*, y^*) \;\; = \;\; & \nabla g(x^*, y^*)\lambda + \nabla h(x^*, y^*)\mu - \begin{pmatrix} O \\ I \end{pmatrix} \nu \\ & + \sum_{\ell=1}^{L}\sum_{i=1}^{m} \xi_\ell[i]\Big((N_\ell x^* + M_\ell y^* + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big) \\ & + \sum_{\ell=1}^{L}\sum_{i=1}^{m} p_\ell d[i]\sigma_\ell[i] \begin{pmatrix} -N_\ell[i] \\ -M_\ell[i] \end{pmatrix}. \end{aligned} \tag{7.96}$$

Thus, in order to show that (x^*, y^*) is a stationary point of problem (7.90), we only need to prove that the vector on the right-hand side of (7.96) belongs to the set

$$\bar\rho\partial r_0(x^*, y^*) + \bar\rho\partial\delta_0(x^*, y^*) + \partial\theta(x^*, y^*).$$

In a similar way to the proof of Theorem 7.7, we can show that

$$\sum_{\ell=1}^{L}\sum_{i=1}^{m} \xi_\ell[i]\Big((N_\ell x^* + M_\ell y^* + q_\ell)[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} + y^*[i] \begin{pmatrix} N_\ell[i] \\ M_\ell[i] \end{pmatrix}\Big) \in \bar\rho\partial\delta_0(x^*, y^*)$$

and

$$\sum_{\ell=1}^{L}\sum_{i=1}^{m} p_\ell d[i]\sigma_\ell[i] \begin{pmatrix} -N_\ell[i] \\ -M_\ell[i] \end{pmatrix} \in \partial\theta(x^*, y^*).$$

Now let us prove that

$$\nabla g(x^*, y^*)\lambda + \nabla h(x^*, y^*)\mu - \begin{pmatrix} O \\ I \end{pmatrix} \nu \in \bar\rho\partial r_0(x^*, y^*). \tag{7.97}$$

To this end, it is enough to show

$$i = 1, \cdots, s_1, \quad \begin{cases} \lambda[i] \in [0, \bar\rho], & g_i(x^*, y^*) = 0 \\ \lambda[i] = \bar\rho\ , & g_i(x^*, y^*) > 0 \\ \lambda[i] = 0\ , & g_i(x^*, y^*) < 0, \end{cases} \tag{7.98}$$

$$i = 1, \cdots, s_2, \quad \begin{cases} \mu[i] \in [-\bar{\rho}, \bar{\rho}], & h_i(x^*, y^*) = 0 \\ \mu[i] = \bar{\rho} \ , & h_i(x^*, y^*) > 0 \\ \mu[i] = -\bar{\rho} \ , & h_i(x^*, y^*) < 0, \end{cases} \tag{7.99}$$

and

$$i = 1, \cdots, m, \quad \begin{cases} \nu[i] \in [0, \bar{\rho}], & y^*[i] = 0 \\ \nu[i] = \bar{\rho} \ , & y^*[i] < 0 \\ \nu[i] = 0 \ , & y^*[i] > 0. \end{cases} \tag{7.100}$$

In fact, (7.98)–(7.100) follow from (7.72) and the facts that

$$\lim_{k \to \infty} (x^k, y^k) = (x^*, y^*)$$

and

$$\begin{aligned} \lambda^k[i] &= \frac{\rho_k}{2}\Big(\frac{g_i(x^k, y^k)}{\sqrt{(g_i(x^k, y^k))^2 + \epsilon_k^2}} + 1\Big), \\ \mu^k[i] &= \frac{\rho_k h_i(x^k, y^k)}{\sqrt{(h_i(x^k, y^k))^2 + \epsilon_k^2}}, \\ \nu^k[i] &= \frac{\rho_k}{2}\Big(\frac{-y^k[i]}{\sqrt{(-y^k[i])^2 + \epsilon_k^2}} + 1\Big), \end{aligned}$$

and hence (7.97) must hold. This completes the proof. ∎

7.4.3 Numerical results

We have tested the proposed methods on Example 7.2. In our experiments, we employed the MATLAB 6.5 built-in solver function *fmincon* to solve the constrained subproblems (7.60) and used *fminunc* to solve the unconstrained subproblems (7.88). For both SP-I and SP-II, we set $\epsilon_0 = 10^{-2}$, $\rho_0 = 10^3$, and updated these parameters by $\epsilon_{k+1} = 10^{-1}\epsilon_k$ and $\rho_{k+1} = \min\{10\rho_k, \ \bar{\rho}\}$, respectively. Moreover, the constants p and $\bar{\rho}$ are set to be 0.25 and 10^5, respectively. In addition, the initial point is chosen to be $y^0 = (0, 0)$ and the computed solution y^k at the kth iteration is used as the starting point for the next iteration.

The computational results for Example 7.2 by SP-I and SP-II are reported in Tables 7.1 and 7.2, respectively. In the tables, *Ite* stands for the number of iterations spent by *fmincon* or *fminunc* to solve the subproblems. The results shown in the tables reveal that the proposed methods are able to solve Example 7.1 successfully.

Table 7.1: Computational Results by SP-I

ϵ_k	ρ_k	y^k	Ite
10^{-2}	10^3	(1.4206,3.4292)	22
10^{-3}	10^4	(1.4744,3.4780)	6
10^{-4}	10^5	(1.4919,3.4931)	7
10^{-5}	10^5	(1.4992,3.4993)	12
10^{-6}	10^5	(1.4999,3.4999)	6
10^{-7}	10^5	(1.5000,3.5000)	6

Table 7.2: Computational Results by SP-II

ϵ_k	ρ_k	y^k	Ite
10^{-2}	10^3	(1.4171,3.4339)	34
10^{-3}	10^4	(1.4744,3.4780)	9
10^{-4}	10^5	(1.4919,3.4931)	11
10^{-5}	10^5	(1.4992,3.4993)	12
10^{-6}	10^5	(1.4999,3.4999)	7
10^{-7}	10^5	(1.5000,3.5000)	20

7.5 Conclusions

A class of stochastic mathematical programs with linear complementarity constraints, called the here-and-now model, has been dealt with in this chapter. We have presented a number of reformulations of the problem and then, based on these reformulations, proposed two smoothed penalty methods. Comprehensive convergence theory for the two methods have been established.

Chapter 8

Regularization Method for SMPECs

In this chapter, assuming that the underlying sample space Ω is discrete and finite, i.e., $\Omega = \{\omega_1, \omega_2, \cdots, \omega_L\}$ for some integer $L > 0$, we consider the here-and-now model [52, 53] of SMPECs:

$$\begin{array}{ll} \text{minimize} & f(x,y) + \displaystyle\sum_{\ell=1}^{L} p_\ell d^T z_\ell \\ \text{subject to} & g(x,y) \le 0, \; h(x,y) = 0, \\ & y \ge 0, \; N_\ell x + M_\ell y + q_\ell + z_\ell \ge 0, \\ & y^T(N_\ell x + M_\ell y + q_\ell + z_\ell) = 0, \\ & z_\ell \ge 0, \quad \ell = 1, 2, \cdots, L. \end{array} \tag{8.1}$$

Here, the functions $f : \Re^{n+m} \to \Re, g : \Re^{n+m} \to \Re^{s_1}$, $h : \Re^{n+m} \to \Re^{s_2}$ are all continuously differentiable, $d \in \Re^m$ is a constant vector with positive elements and, for each ℓ, $N_\ell := N(\omega_\ell) \in \Re^{m\times n}$, $M_\ell := M(\omega_\ell) \in \Re^{m\times m}$, and $q_\ell := q(\omega_\ell) \in \Re^m$ are given matrices and vectors associated with the random event ω_ℓ, $z_\ell \in \Re^m$ is a recourse variable, p_ℓ denotes the probability of ω_ℓ and is assumed to be positive throughout.

Based on some reformulations, two penalty methods have been proposed for solving problem (8.1) in Chapter 7. In addition, a smoothing implicit programming method incorporating a penalty technique has been suggested for solving a similar problem in Chapter 6. However, like the penalty methods in standard nonlinear programming, the methods suggested in the previous chapters cannot ensure the feasibility of a limit point of a generated sequence in general. In this chapter, we will present a regularization method for problem (8.1) and show that, under a quite weak condition, an accumulation point of the generated sequence is a feasible point of the original problem. We will also establish global convergence to an S-stationary point of the problem under additional assumptions.

8.1 Preliminaries and Regularization Method

In this section, we propose a regularization method for problem (8.1). We first recall some basic concepts. Since problem (8.1) is equivalent to the following ordinary MPEC (8.2), we will employ the same stationary concepts as in the literature on MPECs:

$$\begin{array}{ll} \text{minimize} & f(x,y) + \mathbf{d}^T\mathbf{z} \\ \text{subject to} & g(x,y) \leq 0,\ h(x,y) = 0, \\ & \mathbf{y} - Dy = 0,\ \mathbf{z} \geq 0, \\ & \mathbf{y} \geq 0,\ Nx + M\mathbf{y} + \mathbf{q} + \mathbf{z} \geq 0, \\ & \mathbf{y}^T(Nx + M\mathbf{y} + \mathbf{q} + \mathbf{z}) = 0, \end{array} \tag{8.2}$$

where $\mathbf{y} := (y^T, \cdots, y^T)^T \in \Re^{mL}$, $\mathbf{z} := (z_1^T, \cdots, z_L^T)^T \in \Re^{mL}$, and

$$\mathbf{d} := \begin{pmatrix} p_1 d \\ \vdots \\ p_L d \end{pmatrix}, D := \begin{pmatrix} I \\ \vdots \\ I \end{pmatrix}, N := \begin{pmatrix} N_1 \\ \vdots \\ N_L \end{pmatrix}, M := \begin{pmatrix} M_1 & & O \\ & \ddots & \\ O & & M_L \end{pmatrix}, \mathbf{q} := \begin{pmatrix} q_1 \\ \vdots \\ q_L \end{pmatrix}. \tag{8.3}$$

Suppose that $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ is a feasible point of problem (8.2).

Definition 8.1 We say $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ to be a *Bouligand or B-stationary* point of the MPEC (8.2) if

$$\mathbf{v}^T \begin{pmatrix} \nabla f(x^*, y^*) \\ 0 \\ \mathbf{d} \end{pmatrix} \geq 0, \qquad \forall \mathbf{v} \in \mathcal{T}(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*),$$

where $\mathcal{T}(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ stands for the tangent cone of the feasible region of problem (8.2) at $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$.

Definition 8.2 We say $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ to be a *strongly or S-stationary* point of (8.2) if there exist multiplier vectors λ, μ, $\boldsymbol{\nu}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ such that

$$\begin{pmatrix} \nabla_x f(x^*, y^*) \\ \nabla_y f(x^*, y^*) \\ 0 \\ \mathbf{d} \end{pmatrix} + \begin{pmatrix} \nabla_x g(x^*, y^*) \\ \nabla_y g(x^*, y^*) \\ O \\ O \end{pmatrix} \lambda + \begin{pmatrix} \nabla_x h(x^*, y^*) \\ \nabla_y h(x^*, y^*) \\ O \\ O \end{pmatrix} \mu$$

$$+\begin{pmatrix} O \\ -D^T \\ I \\ O \end{pmatrix} \boldsymbol{\nu} - \begin{pmatrix} O \\ O \\ O \\ I \end{pmatrix} \boldsymbol{\alpha} - \begin{pmatrix} O \\ O \\ I \\ O \end{pmatrix} \boldsymbol{\beta} - \begin{pmatrix} N^T \\ O \\ M^T \\ I \end{pmatrix} \boldsymbol{\gamma} = 0, \tag{8.4}$$

$$0 \leq \lambda \perp (-g(x^*, y^*)) \geq 0, \tag{8.5}$$

$$0 \leq \boldsymbol{\alpha} \perp \mathbf{z}^* \geq 0, \tag{8.6}$$

$$\mathbf{y}^* \geq 0, \tag{8.7}$$

$$\mathbf{y}^*[i] > 0 \Rightarrow \boldsymbol{\beta}[i] = 0, \tag{8.8}$$

$$(Nx^* + M\mathbf{y}^* + \mathbf{q} + \mathbf{z}^*) \geq 0, \tag{8.9}$$

$$(Nx^* + M\mathbf{y}^* + \mathbf{q} + \mathbf{z}^*)[i] > 0 \Rightarrow \boldsymbol{\gamma}[i] = 0, \tag{8.10}$$

$$\boldsymbol{\beta}[i] \geq 0,\ \boldsymbol{\gamma}[i] \geq 0, \quad \forall i \in \mathcal{I}^* := \{i \mid \ \mathbf{y}^*[i] = (Nx^* + M\mathbf{y}^* + \mathbf{q} + \mathbf{z}^*)[i] = 0\}. \tag{8.11}$$

It is well-known [75] that any S-stationary point of (8.2) must be a B-stationary point of problem (8.2). In order to look for some effective methods for solving problem (8.1), some equivalent reformulations of problem (8.1) have been introduced recently [53]. In particular, for any $x \in \Re^n, y \in \Re^m$, and each ℓ, it has been shown that the set

$$Z_\ell(x,y) := \left\{ z_\ell \;\middle|\; \begin{array}{l} y^T(N_\ell x + M_\ell y + q_\ell + z_\ell) = 0 \\ N_\ell x + M_\ell y + q_\ell + z_\ell \geq 0,\ z_\ell \geq 0 \end{array} \right\}$$

is nonempty if and only if

$$Q_\ell(x,y) := \sup\Big\{ -(u+ty)^T(N_\ell x + M_\ell y + q_\ell) \;\Big|\; u + ty \leq d, \quad u \geq 0, \quad t \leq 0 \Big\}$$

is finite. Based on this observation, we obtain the model

$$\begin{array}{ll} \text{minimize} & f(x,y) + \displaystyle\sum_{\ell=1}^{L} p_\ell Q_\ell(x,y) \\ \text{subject to} & g(x,y) \leq 0,\ h(x,y) = 0, \quad y \geq 0. \end{array} \tag{8.12}$$

Furthermore, we have the following result.

Theorem 8.1 [53] *If (x^*, y^*) solves problem (8.12), then there exist $z_\ell^*, \ell = 1, 2, \cdots, L$, such that $(x^*, y^*, z_1^*, \cdots, z_L^*)$ solves problem (8.1). Conversely, if $(x^*, y^*, z_1^*, \cdots, z_L^*)$ solves problem (8.1), then the point (x^*, y^*) solves problem (8.12).*

In what follows, we denote by $\mathcal{F}_1$ and $\mathcal{F}_2$ the feasible regions of problems (8.1) and (8.12), respectively. The next result will be used later on.

Theorem 8.2 [53] *Let $g(x,y) \leq 0, h(x,y) = 0$, and $y \geq 0$. Then the following statements are equivalent:*

(i) $Q_\ell(x,y) < +\infty$ *for every* $\ell = 1, 2, \cdots, L$.

(ii) *For any i and any ℓ, there holds*

$$y[i](N_\ell x + M_\ell y + q_\ell)[i] \leq 0. \tag{8.13}$$

(iii) *The point $(x, y, z_1, \cdots, z_L)$ with $z_\ell := \max(-(N_\ell x + M_\ell y + q_\ell),\ 0), \ell = 1, \cdots, L$, is a feasible point of problem (8.1).*

On the other hand, we note that, for every ℓ, the function Q_ℓ may not be finite-valued and not differentiable everywhere in general. We next introduce a smooth approximation of this function: Let ϵ be a positive parameter. For each ℓ, we define the function $Q_\ell^\epsilon : \Re^n \times \Re^m \to [0, +\infty)$ as follows:

$$\begin{aligned} Q_\ell^\epsilon(x,y) \quad := \quad \max\Big\{ & -(u+ty)^T(N_\ell x + M_\ell y + q_\ell) - \frac{\epsilon}{2}(t^2 + \|u\|^2) \ \Big| \\ & u + ty \leq d, \quad u \geq 0, \quad t \leq 0 \Big\}. \end{aligned} \tag{8.14}$$

By the convex programming theory, we see that any Karush-Kuhn-Tucker point of the problem

$$\begin{array}{ll} \text{maximize} & -(u+ty)^T(N_\ell x + M_\ell y + q_\ell) - \dfrac{\epsilon}{2}(t^2 + \|u\|^2) \\ \text{subject to} & u + ty \leq d, \quad u \geq 0, \quad t \leq 0 \end{array} \tag{8.15}$$

must be an optimal solution and, since $\epsilon > 0$, problem (8.15) indeed has a unique optimal solution. This implies that the function Q_ℓ^ϵ is well-defined for each ℓ. We next show that Q_ℓ^ϵ is differentiable everywhere. To this end, let

$$\begin{aligned} g(x,y,u,t) &:= u + ty - d, \\ h(x,y,u,t) &:= -u, \\ c(x,y,u,t) &:= t, \end{aligned}$$

and

$$\begin{aligned} L_\ell^\epsilon(x,y,u,t,\zeta,\eta,\xi) \quad := \quad & (u+ty)^T(N_\ell x + M_\ell y + q_\ell) + \frac{\epsilon}{2}(t^2 + \|u\|^2) \\ & + \zeta^T g(x,y,u,t) + \eta^T h(x,y,u,t) + \xi c(x,y,u,t). \end{aligned}$$

We then have that $\nabla^2_{(u,t)} L_\ell^\epsilon(x,y,u,t,\zeta,\eta,\xi) = \epsilon I$.

Lemma 8.1 *For any* $(x, y) \in \Re^{n+m}$, *let* $u_\ell := u(x, y)$ *and* $t_\ell := t(x, y)$ *be the unique optimal solution of problem (8.15) and* $\zeta_\ell := \zeta(x, y), \eta_\ell := \eta(x, y), \xi_\ell := \xi(x, y)$ *be the corresponding Lagrangian multiplier vectors. Then,*

(a) *for any* $(u, t) \in \Re^{m+1}$, *there holds*

$$(u, t)^T \nabla^2_{(u,t)} L^\epsilon_\ell(x, y, u_\ell, t_\ell, \zeta_\ell, \eta_\ell, \xi_\ell)(u, t) \geq \epsilon \|(u, t)\|^2.$$

(b) *the linear independence constraint qualification is satisfied at* (x, y, u_ℓ, t_ℓ), *that is, the set of vectors*

$$\Big\{\nabla_{(u,t)} g_i(x, y, u_\ell, t_\ell), \nabla_{(u,t)} h_j(x, y, u_\ell, t_\ell), \nabla_{(u,t)} c(x, y, u_\ell, t_\ell) \;\Big|\; i \in \mathcal{I}_g(x, y, u_\ell, t_\ell), j \in \mathcal{I}_h(x, y, u_\ell, t_\ell)\Big\}$$

when $t_\ell = 0$, *or*

$$\Big\{\nabla_{(u,t)} g_i(x, y, u_\ell, t_\ell), \nabla_{(u,t)} h_j(x, y, u_\ell, t_\ell) \;\Big|\; i \in \mathcal{I}_g(x, y, u_\ell, t_\ell), j \in \mathcal{I}_h(x, y, u_\ell, t_\ell)\Big\}$$

when $t_\ell \neq 0$, *is linearly independent.*

Proof: It is obvious that (a) holds. Moreover, since for any index set $\mathcal{I} \subseteq \{1, \cdots, m\}$, the set of vectors

$$\Big\{\nabla_{(u,t)} g_i(x, y, u_\ell, t_\ell), \nabla_{(u,t)} h_j(x, y, u_\ell, t_\ell), \nabla_{(u,t)} c(x, y, u_\ell, t_\ell) \;\Big|\; i \in \mathcal{I},\ j \notin \mathcal{I}\Big\}$$

must be linearly independent and, in addition, there always holds

$$\mathcal{I}_g(x, y, u_\ell, t_\ell) \cap \mathcal{I}_h(x, y, u_\ell, t_\ell) = \emptyset,$$

we then see that (b) is true. ∎

Thus, by Theorem 2 [4, Page 130], we have the following result immediately.

Theorem 8.3 *The functions* $u(x, y), t(x, y), \zeta(x, y), \eta(x, y)$, *and* $\xi(x, y)$ *given in Lemma 8.1 are well-defined and continuous. Furthermore, the function* Q^ϵ_ℓ *defined by (8.14) is differentiable everywhere and*

$$\nabla Q^\epsilon_\ell(x, y) = \begin{pmatrix} -N^T_\ell(u_\ell + t_\ell y) \\ -M^T_\ell(u_\ell + t_\ell y) - t_\ell(N_\ell x + M_\ell y + q_\ell) \end{pmatrix} + \begin{pmatrix} 0 \\ t_\ell \zeta_\ell \end{pmatrix}, \tag{8.16}$$

where u_ℓ, t_ℓ, *and* ζ_ℓ *are the same as in Lemma 8.1.*

As a result, the problem

$$\begin{aligned} \text{minimize} \quad & f(x,y) + \sum_{\ell=1}^{L} p_\ell Q_\ell^\epsilon(x,y) \\ \text{subject to} \quad & g(x,y) \le 0,\ h(x,y) = 0, \quad y \ge 0 \end{aligned} \tag{8.17}$$

is a smooth approximation of problem (8.12). We then have the following algorithm.

Algorithm RA:

Step 1: Choose $\epsilon_0 > 0$ and set $k := 0$.

Step 2: Solve problem (8.17) with $\epsilon = \epsilon_k$ to get a stationary point (x^k, y^k) and go to Step 3.

Step 3: If a stopping rule is satisfied, then terminate. Otherwise, choose an $\epsilon_{k+1} \in (0, \epsilon_k)$ and return to Step 2 with $k := k + 1$.

In what follows, we suppose that the sequence $\{\epsilon_k\}$ is convergent to 0 and, for simplicity, we denote $Q_\ell^{\epsilon_k}$ by Q_ℓ^k for each k and ℓ. Recall that $\mathcal{F}_2$ denotes the feasible region of problem (8.12), which is the same as that of problem (8.17).

8.2 Convergence Analysis

We will investigate the limiting behavior of the sequence generated by Algorithm RA in this section. Our first result is concerned with the feasibility of the limit point of the generated sequence, which can be stated as follows.

Theorem 8.4 *Let $\{(x^k, y^k)\}$ be a sequence generated by Algorithm RA and suppose that $\{Q_\ell^k(x^k, y^k)\}$ is bounded for each ℓ. Then, for any accumulation point (x^*, y^*) of the sequence $\{(x^k, y^k)\}$, the vector $(x^*, y^*, z_1^*, \cdots, z_L^*)$ is feasible to problem (8.1), where*

$$z_\ell^* := \max(-(N_\ell x^* + M_\ell y^* + q_\ell),\ 0), \qquad \ell = 1, \cdots, L.$$

Proof: Assume without loss of generality that $\lim_{k\to\infty}(x^k, y^k) = (x^*, y^*)$. It is obvious from the continuity of the functions g and h that

$$g(x^*, y^*) \le 0, \quad h(x^*, y^*) = 0, \quad y^* \ge 0.$$

Suppose that the assertion of the theorem does not hold. Then, by Theorem 8.2, there exist some ℓ and i such that

$$y^*[i] > 0, \quad (N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0.$$

Therefore, we can find a constant $\eta > 0$ and an integer $k_0 > 0$ such that

$$y^k[i] > \eta, \ \ (N_\ell x^k + M_\ell y^k + q_\ell)[i] > \eta, \qquad \forall k \geq k_0. \tag{8.18}$$

For any $t \leq 0$, we define $u(t) := ty^k[i]e_i - ty^k$. It is easy to see that

$$u(t) \geq 0, \quad u(t) + ty^k \leq d, \qquad \forall t \leq 0.$$

It then follows from the definition of Q_ℓ^k that

$$\begin{aligned} Q_\ell^k(x^k, y^k) &\geq \sup\Big\{ -(u(t)+ty^k)^T(N_\ell x^k + M_\ell y^k + q_\ell) - \frac{\epsilon_k}{2}(t^2 + \|u(t)\|^2) \ \Big| \ t \leq 0 \Big\} \\ &= \sup\Big\{ -ty^k[i](N_\ell x^k + M_\ell y^k + q_\ell)[i] - \frac{\epsilon_k}{2} t^2 (1 + \|y^k[i]e_i - y^k\|^2) \ \Big| \ t \leq 0 \Big\}. \end{aligned}$$

By straightforward calculus, we can show that, for any $k \geq k_0$,

$$\begin{aligned} Q_\ell^k(x^k, y^k) &\geq \frac{(y^k[i](N_\ell x^k + M_\ell y^k + q_\ell)[i])^2}{2\epsilon_k(1 + \|y^k[i]e_i - y^k\|^2)} \\ &\geq \frac{\eta^4}{2\epsilon_k(1 + \|y^k[i]e_i - y^k\|^2)}, \end{aligned} \tag{8.19}$$

where the second inequality follows from (8.18). Taking into account the fact that

$$\lim_{k\to\infty} \|y^k[i]e_i - y^k\| = \|y^*[i]e_i - y^*\|, \quad \lim_{k\to\infty} \epsilon_k = 0,$$

we see from (8.19) that the sequence $\{Q_\ell^k(x^k, y^k)\}$ is unbounded. This is a contradiction and hence there must be some vectors $z_\ell^*, \ell = 1, 2, \cdots, L$, such that $(x^*, y^*, z_1^*, \cdots, z_L^*)$ is feasible to problem (8.1). This completes the proof. ∎

The main convergence result can be stated as follows.

Theorem 8.5 *Suppose that Algorithm RA generates a sequence $\{(x^k, y^k)\}$ of stationary points of problems (8.17) and, for each k, (u_ℓ^k, t_ℓ^k) is the corresponding unique optimal solution of problem (8.15) with $\epsilon := \epsilon_k$. Assume that, for each ℓ, both $\{Q_\ell^k(x^k, y^k)\}$ and $\{t_\ell^k\}$ are bounded. Moreover, suppose that (x^*, y^*) is an accumulation point of the sequence $\{(x^k, y^k)\}$ such that the system*

$$g(x, y) \leq 0, \quad h(x, y) = 0, \quad y \geq 0 \tag{8.20}$$

satisfies the MFCQ at (x^*, y^*), and let

$$z_\ell^* := \max(-(N_\ell x^* + M_\ell y^* + q_\ell),\ 0), \qquad \ell = 1, \cdots, L \tag{8.21}$$

and

$$\mathbf{y}^* := ((y^*)^T, \cdots, (y^*)^T)^T, \qquad \mathbf{z}^* := ((z_1^*)^T, \cdots, (z_L^*)^T)^T. \tag{8.22}$$

Then $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ is an S-stationary point of problem (8.2).

Proof: Assume without loss of generality that $\lim_{k\to\infty}(x^k, y^k) = (x^*, y^*)$. From Theorem 8.4, we see $(x^*, y^*, z_1^*, \cdots, z_L^*) \in \mathcal{F}_1$ and hence $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ is a feasible point of (8.2). We next show that $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ is S-stationary to problem (8.2), that is, there exist multiplier vectors $\lambda, \mu, \boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ such that (8.4)–(8.11) hold.

First of all, by the stationarity of (x^k, y^k) to (8.17), there exist Lagrange multiplier vectors $a^k \in \Re^{s_1}, b^k \in \Re^{s_2}$, and $c^k \in \Re^m$ such that

$$\nabla f(x^k, y^k) + \sum_{\ell=1}^{L} p_\ell \nabla Q_\ell^k(x^k, y^k) + \nabla g(x^k, y^k)a^k + \nabla h(x^k, y^k)b^k - \begin{pmatrix} O \\ I \end{pmatrix} c^k = 0, \tag{8.23}$$

$$0 \le a^k \perp (-g(x^k, y^k)) \ge 0, \tag{8.24}$$

$$0 \le c^k \perp \quad y^k \ge 0. \tag{8.25}$$

From (8.16), we can rewrite (8.23) as

$$\nabla f(x^k, y^k) + \sum_{\ell=1}^{L} p_\ell \begin{pmatrix} -N_\ell^T(u_\ell^k + t_\ell^k y^k) \\ -M_\ell^T(u_\ell^k + t_\ell^k y^k) - t_\ell^k(N_\ell x^k + M_\ell y^k + q_\ell - \zeta_\ell^k) \end{pmatrix} + \nabla g(x^k, y^k)a^k + \nabla h(x^k, y^k)b^k - \begin{pmatrix} O \\ I \end{pmatrix} c^k = 0,$$

where $\zeta_\ell^k := \zeta(x^k, y^k)$ and the function $\zeta(x, y)$ is defined as in Lemma 8.1. This condition is further equivalent to

$$\begin{aligned} 0 \;=\; & \nabla f(x^k, y^k) + \nabla g(x^k, y^k)a^k + \nabla h(x^k, y^k)b^k - \begin{pmatrix} O \\ I \end{pmatrix} c^k \\ & - \sum_{\ell=1}^{L} p_\ell t_\ell^k \begin{pmatrix} O \\ I \end{pmatrix} (N_\ell x^k + M_\ell y^k + q_\ell - \zeta_\ell^k) - \sum_{\ell=1}^{L} p_\ell \begin{pmatrix} N_\ell^T \\ M_\ell^T \end{pmatrix} (u_\ell^k + t_\ell^k y^k). \end{aligned} \tag{8.26}$$

We next prove that the sequences $\{a^k\}$, $\{b^k\}$ and $\{c^k\}$ are bounded. To this end, let

$$\rho_k := \sum_{i=1}^{s_1} a^k[i] + \sum_{i=1}^{s_2} |b^k[i]| + \sum_{i=1}^{m} c^k[i]. \tag{8.27}$$

Suppose that at least one of the sequences $\{a^k\}$, $\{b^k\}$ and $\{c^k\}$ is unbounded. We then have $\lim_{k\to\infty} \rho_k = +\infty$ and, taking a subsequence if necessary, we may assume that the limits

$$\begin{aligned} \bar{a}[i] &:= \lim_{k\to\infty} \frac{a^k[i]}{\rho_k}, \qquad i = 1, \cdots, s_1, \\ \bar{b}[i] &:= \lim_{k\to\infty} \frac{b^k[i]}{\rho_k}, \qquad i = 1, \cdots, s_2, \\ \bar{c}[i] &:= \lim_{k\to\infty} \frac{c^k[i]}{\rho_k}, \qquad i = 1, \cdots, m \end{aligned}$$

exist. It is clear from (8.27) that

$$\sum_{i=1}^{s_1} \bar{a}[i] + \sum_{i=1}^{s_2} |\bar{b}[i]| + \sum_{i=1}^{m} \bar{c}[i] = 1.$$

For each ℓ, since $\{t_\ell^k\}$ is bounded and

$$0 \le u_\ell^k \le d - t_\ell^k y^k, \qquad \forall k,$$

we see that $\{u_\ell^k\}$ is bounded. Moreover, by the continuity of the functions given in Theorem 8.3, $\{\zeta_\ell^k\}$ is also bounded. Thus, dividing (8.26) by ρ_k and taking a limit, we get

$$\nabla g(x^*, y^*)\bar{a} + \nabla h(x^*, y^*)\bar{b} - \begin{pmatrix} O \\ I \end{pmatrix} \bar{c} = 0.$$

Furthermore, taking (8.24) and (8.25) into account, we obtain $\bar{a} \ge 0, \bar{c} \ge 0$, and

$$\begin{aligned} \bar{a}[i] = 0, &\quad i \notin \mathcal{I}_g(x^*, y^*), \\ \bar{c}[i] = 0, &\quad i \notin \mathcal{I}_\pi(x^*, y^*), \end{aligned}$$

where $\pi : \Re^{n+m} \to \Re^m$ is given by $\pi(x, y) := y$. It follows that

$$\begin{aligned} \sum_{i\in\mathcal{I}_g(x^*,y^*)} \bar{a}[i] + \sum_{i=1}^{s_2} |\bar{b}[i]| + \sum_{i\in\mathcal{I}_\pi(x^*,y^*)} \bar{c}[i] = 1, \\ \bar{a}[i] \ge 0, \qquad i \in \mathcal{I}_g(x^*, y^*), \\ \bar{c}[i] \ge 0, \qquad i \in \mathcal{I}_\pi(x^*, y^*), \end{aligned}$$

and

$$\sum_{i\in\mathcal{I}_g(x^*,y^*)} \bar{a}[i]\nabla g_i(x^*, y^*) + \sum_{i=1}^{s_2} \bar{b}[i]\nabla h_i(x^*, y^*) - \sum_{i\in\mathcal{I}_\pi(x^*,y^*)} \bar{c}[i] \begin{pmatrix} 0 \\ e_i \end{pmatrix} = 0.$$

This contradicts the assumption that the system (8.20) satisfies the MFCQ at (x^*, y^*) and hence all the sequences $\{a^k\}$, $\{b^k\}$, and $\{c^k\}$ are bounded. Recall that $\{t_\ell^k\}$, $\{u_\ell^k\}$, and $\{\zeta_\ell^k\}$ are also bounded for each ℓ.

Now let us proceed to showing (8.4)–(8.11) step by step. First we show (8.4) and (8.5). Let

$$\begin{aligned}
\lambda^k &:= a^k, && (8.28)\\
\mu^k &:= b^k, && (8.29)\\
\alpha_\ell^k[i] &:= p_\ell(d[i] - u_\ell^k[i] - t_\ell^k y^k[i]), && (8.30)\\
\beta_\ell^k[i] &:= \frac{1}{L} c^k[i] + p_\ell t_\ell^k (N_\ell x^k + M_\ell y^k + q_\ell - \zeta_\ell^k)[i], && (8.31)\\
\gamma_\ell^k[i] &:= p_\ell(u_\ell^k[i] + t_\ell^k y^k[i]), && (8.32)\\
\nu_\ell^k &:= \beta_\ell^k + M_\ell^T \gamma_\ell^k, && (8.33)
\end{aligned}$$

and

$$\boldsymbol{\alpha}^k := \begin{pmatrix} \alpha_1^k \\ \vdots \\ \alpha_L^k \end{pmatrix}, \quad \boldsymbol{\beta}^k := \begin{pmatrix} \beta_1^k \\ \vdots \\ \beta_L^k \end{pmatrix}, \quad \boldsymbol{\gamma}^k := \begin{pmatrix} \gamma_1^k \\ \vdots \\ \gamma_L^k \end{pmatrix}, \quad \boldsymbol{\nu}^k := \begin{pmatrix} \nu_1^k \\ \vdots \\ \nu_L^k \end{pmatrix}.$$

Then (8.26) can be rewritten as

$$\begin{aligned}
0 \;=\; & \begin{pmatrix} \nabla_x f(x^k, y^k) \\ \nabla_y f(x^k, y^k) \\ 0 \\ \mathbf{d} \end{pmatrix} + \begin{pmatrix} \nabla_x g(x^k, y^k) \\ \nabla_y g(x^k, y^k) \\ O \\ O \end{pmatrix} \lambda^k + \begin{pmatrix} \nabla_x h(x^k, y^k) \\ \nabla_y h(x^k, y^k) \\ O \\ O \end{pmatrix} \mu^k \\
& + \begin{pmatrix} O \\ -D^T \\ I \\ O \end{pmatrix} \boldsymbol{\nu}^k - \begin{pmatrix} O \\ O \\ O \\ I \end{pmatrix} \boldsymbol{\alpha}^k - \begin{pmatrix} O \\ O \\ I \\ O \end{pmatrix} \boldsymbol{\beta}^k - \begin{pmatrix} N^T \\ O \\ M^T \\ I \end{pmatrix} \boldsymbol{\gamma}^k, \qquad (8.34)
\end{aligned}$$

where $\mathbf{d}, D, N$, and M are defined as in (8.3). Since all the multiplier vectors are bounded, without loss of generality, we may assume that

$$\lambda := \lim_{k\to\infty} \lambda^k, \qquad \mu := \lim_{k\to\infty} \mu^k, \qquad \boldsymbol{\nu} := \lim_{k\to\infty} \boldsymbol{\nu}^k,$$

and

$$\boldsymbol{\alpha} := \lim_{k\to\infty} \boldsymbol{\alpha}^k, \qquad \boldsymbol{\beta} := \lim_{k\to\infty} \boldsymbol{\beta}^k, \qquad \boldsymbol{\gamma} := \lim_{k\to\infty} \boldsymbol{\gamma}^k.$$

Taking a limit in (8.34), we obtain (8.4) immediately. Moreover, we have (8.5) from (8.24) and (8.28) by letting $k \to \infty$.

We next prove (8.6)–(8.11). To this end, we let

$$z_\ell^k := \max(-(N_\ell x^k + M_\ell y^k + q_\ell),\ 0), \qquad \ell = 1, \cdots, L \tag{8.35}$$

and

$$\mathbf{y}^k := ((y^k)^T, \cdots, (y^k)^T)^T, \qquad \mathbf{z}^k := ((z_1^k)^T, \cdots, (z_L^k)^T)^T. \tag{8.36}$$

Recall that, for each ℓ and each k, (u_ℓ^k, t_ℓ^k) is a Karush-Kuhn-Tucker point of problem (8.15) with $\epsilon = \epsilon_k$. Thus, the Lagrange multiplier vectors $\zeta_\ell^k, \eta_\ell^k$, and ξ_ℓ^k satisfy

$$(N_\ell x^k + M_\ell y^k + q_\ell) + \epsilon_k u_\ell^k + \zeta_\ell^k - \eta_\ell^k = 0, \tag{8.37}$$

$$(y^k)^T(N_\ell x^k + M_\ell y^k + q_\ell) + \epsilon_k t_\ell^k + (y^k)^T \zeta_\ell^k + \xi_\ell^k = 0, \tag{8.38}$$

$$0 \le \zeta_\ell^k \perp (d - u_\ell^k - t_\ell^k y^k) \ge 0, \tag{8.39}$$

$$0 \le \eta_\ell^k \perp u_\ell^k \ge 0, \tag{8.40}$$

$$0 \le \xi_\ell^k \perp (-t_\ell^k) \ge 0. \tag{8.41}$$

Moreover, for any index j with $1 \le j \le mL$, there exist ℓ and i such that

$$1 \le \ell \le L, \quad 1 \le i \le m, \quad j = (\ell - 1)m + i. \tag{8.42}$$

It is obvious that $\boldsymbol{\alpha}^k \ge 0$ and $\mathbf{z}^k \ge 0$ from the definitions (8.30), (8.35) and (8.36) for every k. Taking a limit, we obtain $\boldsymbol{\alpha} \ge 0$ and $\mathbf{z}^* \ge 0$. Suppose $\mathbf{z}^*[j] > 0$ and let ℓ and i satisfy (8.42). Then, since $z_\ell^*[i] > 0$, it follows from (8.21) that $(N_\ell x^* + M_\ell y^* + q_\ell)[i] < 0$. We then have from (8.37) and (8.40) that

$$\begin{aligned} \zeta_\ell^k[i] &\ge \zeta_\ell^k[i] - \eta_\ell^k[i] \\ &= -(N_\ell x^k + M_\ell y^k + q_\ell)[i] - \epsilon_k u_\ell^k[i] \\ &\to -(N_\ell x^* + M_\ell y^* + q_\ell)[i] \\ &> 0. \end{aligned}$$

This implies that $\zeta_\ell^k[i] > 0$ when k is large sufficiently and so, by (8.39),

$$u_\ell^k[i] + t_\ell^k y^k[i] = d[i].$$

Therefore, we have from the definition (8.30) that $\boldsymbol{\alpha}^k[j] = \alpha_\ell^k[i] = 0$ for all k sufficiently large. By taking a limit, we have $\boldsymbol{\alpha}[j] = 0$ and hence (8.6) holds.

It is easy to see that $\mathbf{y}^* \ge 0$. Suppose $\mathbf{y}^*[j] > 0$ and let ℓ and i satisfy (8.42). Then, since $\mathbf{y}^*[j] = y^*[i] > 0$, by Theorem 8.2, there must hold $(N_\ell x^* + M_\ell y^* + q_\ell)[i] = 0$. Moreover, we have $y^k[i] > 0$ when k is large sufficiently. Thus, it follows from (8.25) that

$c^k[i] = 0$ for all k sufficiently large. In addition, it follows from Theorem 8.3 that the sequences $\{\zeta_\ell^k[i]\}$ and $\{\eta_\ell^k[i]\}$ are bounded. Taking a subsequence if necessary, we may assume that both the sequences are convergent. We claim that $\{\zeta_\ell^k[i]\}$ is convergent to 0. Otherwise, we have from (8.37) that the limit of $\{\eta_\ell^k[i]\}$ must be positive. This means that, when k is large sufficiently, there holds $\eta_\ell^k[i] > 0$ and so, by (8.40), $u_\ell^k[i] = 0$. Therefore, we have that, when k is large sufficiently,

$$d[i] - u_\ell^k[i] - t_\ell^k y^k[i] = d[i] - t_\ell^k y^k[i] \geq d[i] > 0$$

and hence, by (8.39), $\zeta_\ell^k[i] = 0$. This is a contradiction. As a result, the sequence $\{\zeta_\ell^k[i]\}$ must be convergent to 0. Thus, taking a limit in (8.31) and noting that $\{t_\ell^k\}$ is bounded, we obtain

$$\boldsymbol{\beta}[j] = \lim_{k\to\infty} \boldsymbol{\beta}^k[j] = \lim_{k\to\infty} \beta_\ell^k[i] = 0.$$

This shows (8.7) and (8.8).

It is easy to see that $Nx^* + M\mathbf{y}^* + \mathbf{q} + \mathbf{z}^* \geq 0$. Suppose $(Nx^* + M\mathbf{y}^* + \mathbf{q} + \mathbf{z}^*)[j] > 0$ and let ℓ and i satisfy (8.42). Then, since $(N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i] > 0$, it follows from (8.21) that $(N_\ell x^* + M_\ell y^* + q_\ell)[i] > 0$ and hence, by Theorem 8.2, $y^*[i] = 0$. Moreover, we have from (8.37) and (8.39)–(8.40) that

$$\begin{aligned} \eta_\ell^k[i] &= (N_\ell x^k + M_\ell y^k + q_\ell)[i] + \epsilon_k u_\ell^k[i] + \zeta_\ell^k[i] \\ &\geq (N_\ell x^k + M_\ell y^k + q_\ell)[i] \\ &\to (N_\ell x^* + M_\ell y^* + q_\ell)[i] \\ &> 0. \end{aligned}$$

In consequence, $\eta_\ell^k[i] > 0$ when k is large sufficiently and then, by (8.40), we have $u_\ell^k[i] = 0$. Taking a limit in (8.32) and noting that $\{t_\ell^k\}$ is bounded, we obtain

$$\boldsymbol{\gamma}[j] = \lim_{k\to\infty} \boldsymbol{\gamma}^k[j] = \lim_{k\to\infty} \gamma_\ell^k[i] = 0$$

and hence (8.9) and (8.10) hold.

Let $\mathcal{I}^*$ be defined as in (8.11) and suppose $j \in \mathcal{I}^*$. Note that

$$u_\ell^k[i] \geq 0, \quad t_\ell^k \leq 0, \quad c^k[i] \geq 0, \qquad \forall k$$

and $(Nx^* + M\mathbf{y}^* + \mathbf{q} + \mathbf{z}^*)[j] = 0$ implies $(N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i] = 0$, where ℓ and i satisfy (8.42), and hence $(N_\ell x^* + M_\ell y^* + q_\ell)[i] \leq 0$ by (8.21). We then have from (8.31)–(8.32) that

$$\boldsymbol{\beta}[j] = \lim_{k\to\infty} \beta_\ell^k[i] \geq \lim_{k\to\infty} p_\ell t_\ell^k (N_\ell x^k + M_\ell y^k + q_\ell)[i] \geq 0$$

and

$$\boldsymbol{\gamma}[j] = \lim_{k\to\infty} \gamma_\ell^k[i] \geq \lim_{k\to\infty} p_\ell t_\ell^k y^k[i] = 0.$$

This indicates that (8.11) holds.

Therefore, the multiplier vectors $\lambda, \mu, \boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ indeed satisfy conditions (8.4)–(8.11) and so $(x^*, y^*, \mathbf{y}^*, \mathbf{z}^*)$ is an S-stationary point of problem (8.2). This completes the proof of the theorem. ∎

8.3 Conclusions

The SMPEC (8.1) has been discussed in the previous chapters and two penalty methods have been proposed there. The main difficulty with the two methods consists in the feasibility of a limit point of the generated sequence, which has not been addressed completely. In this chapter, based on a reformulation given in Chapter 7, we propose a regularization method for solving the SMPEC (8.1). It has been shown that, under a weak condition, an accumulation point of the generated sequence is a feasible point of the original problem. Global convergence to an S-stationary point of the problem has also been established.

Bibliography

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

[2] E. Bierstone and P.D. Milman, *Semianalytic and subanalytic sets*, Institut des Hautes Etudes Scientifiques, Publications Mathématiques, 67 (1988), 5–42.

[3] J.R. Birge and F. Louveaux, Introduction to Stochastic Programming, Springer, New York, 1997.

[4] J.F. Bonnans and A. Shapiro, *Optimization problems with perturbations: A guided tour*, SIAM Review, 40 (1998), 228–264.

[5] J.V. Burke, *An exact penalization viewpoint of constrained optimization*, SIAM Journal on Control and Optimization, 29 (1991), 968–998.

[6] J.V. Burke, *On the identification of active constraints II: The nonconvex case*, SIAM Journal on Numerical Analysis, 27 (1990), 1081–1102.

[7] J.V. Burke and M.C. Ferris, *Weak sharp minima in mathematical programming,* SIAM Journal on Control and Optimization, 31 (1993), 1340–1359.

[8] J.V. Burke and J.J. Moré, *On the identification of active constraints*, SIAM Journal on Numerical Analysis, 25 (1988), 1197–1211.

[9] B. Chen and P.T. Harker, *Smooth approximations to nonlinear complementarity problems*, SIAM Journal on Optimization, 7 (1997), 403–420.

[10] C. Chen and O.L. Mangasarian, *A class of smoothing functions for nonlinear and mixed complementarity problems*, omputational Optimization and Applications, 5 (1996), 97–138.

[11] X. Chen, *Newton-type methods for stochastic programming*, Mathematical and Computer Modelling, 31 (2000), 89–98.

[12] X. Chen and M. Fukushima, *A smoothing method for a mathematical program with P-matrix linear complementarity constraints*, Computational Optimization and Applications, to appear.

[13] X. Chen and M. Fukushima, *Approximation and convergence in stochastic linear complementarity problems*, Technical Report 2003-009, Department of Applied Mathematics and Physics, Kyoto University (July, 2003).

[14] X. Chen, L. Qi, and D. Sun, *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities*, Mathematics of Computation, 67 (1998), 519–540.

[15] X. Chen, L. Qi, and R.S. Womersley, *Newton's method for quadratic stochastic programs with recourse*, Journal of Computational and Applied Mathematics, 60 (1995), 29–46.

[16] X. Chen and R.S. Womersley, *A parallel Newton method for quadratic stochastic programs with recourse*, Annals of Operations Research, 64 (1996), 113–141.

[17] Y. Chen and M. Florian, *The Nonlinear Bilevel Programming Problem: Formulations, Regularity and Optimality Conditions*, Optimization, 32 (1995), 193–209.

[18] F.H. Clarke, Optimization and Nonsmooth Analysis, SIAM, Philadelphia, 1990.

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

[20] J.P. Dedieu, *Penalty functions in subanalytic optimization,* Optimization, 26 (1992), 27–32.

[21] E.D. Dolan, R. Fourer, J.J. Moré, and T.S. Munson, *The NEOS server for optimization: Version 4 and beyond*, Technical Report, `http://www-neos.mcs.anl.gov/neos/ftp/v4.pdf`, 2001.

[22] F. Facchinei, A. Fischer, and C. Kanzow, *On the accurate identification of active constraints*, SIAM Journal on Optimization, 9 (1998), 14–32.

[23] F. Facchinei, A. Fischer, and C. Kanzow, *On the identification of zero variables in an interior-point framework*, SIAM Journal on Optimization, 10 (2000), 1058–1078.

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

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

[26] M.L. Flegel and C. Kanzow, *An Abadie-type constraint qualification for mathematical programs with equilibrium constraints*, manuscript, University of Würzburg, Würzburg, 2002.

[27] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes, *Local Convergence of SQP Methods for Mathematical Programs with Equilibrium Constraints*, Numerical Analysis Report, Department of Mathematics, University of Dundee, Dundee, Scotland, 2001.

[28] C.A. Floudas, P.M. Pardalos, C.S. Adjiman, W.R. Esposito, Z.H. Gümüs, S.T. Harding, J.L. Klepeis, C.A. Meyer, C.A. Schweiger, Handbook of Test Problems in Local and Global Optimization, Nonconvex Optimization and Its Applications, Volume 33, Dordrecht, Kluwer Academic Publishers, 1999.

[29] M. Fukushima, Z.Q. Luo, and J.S. Pang, *A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints*, Computational Optimization and Applications, 10 (1998), 5–34.

[30] M. Fukushima, Z.Q. Luo, and P. Tseng, *A sequential quadratically constrained quadratic programming method for differentiable convex minimization*, SIAM Journal on Optimization, 13 (2003), 1098–1119.

[31] M. Fukushima and J.S. Pang, *Convergence of a smoothing continuation method for mathematical problems with complementarity constraints*, Ill-posed Variational Problems and Regularization Techniques, Lecture Notes in Economics and Mathematical Systems, Vol. 477, M. Théra and R. Tichatschke (eds.), Springer-Verlag, Berlin/Heidelberg, 1999, 105–116.

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

[33] M. Fukushima and P. Tseng, *An implementable active-set algorithm for computing a B-stationary point of the mathematical program with linear complementarity constraints*, SIAM Journal on Optimization, 12 (2002), 724–739.

[34] S.A. Gabriel and J.J. Moré, *Smoothing of mixed complementarity problems,* Complementarity and Variational Problems: State of the Art, M.C. Ferris and J.S. Pang (eds.), SIAM, Philadelphia, Pennsylvania (1997), 105–116.

[35] G. Gürkan, A.Y. Özge and S.M. Robinson, *Sample-path solution of stochastic variational inequalities*, Mathematical Programming, 84 (1999), 313–333.

[36] P.T. Harker and J.S. Pang, *Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications*, Mathematical Programming, 48 (1990), 161–220.

[37] A. Haurie and F. Moresino, *S-adapted oligopoly equilibria and approximations in stochastic variational inequalities, with application to option pricing*, Annals of Operations Research, 114 (2002), 183–201.

[38] X. Hu and D. Ralph, *Convergence of a penalty method for mathematical programming with equilibrium constraints*, Journal of Optimization Theory and Applications, to appear.

[39] X.X. Huang, X.Q. Yang, and K.L. Teo, *A smoothed l_1 penalty method for mathematical programs with complementarity constraints*, manuscript, Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, 2003.

[40] X.X. Huang, X.Q. Yang, and K.L. Teo, *Partial augmented Lagrangian method and mathematical programs with complementarity constraints*, manuscript, Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, 2002.

[41] X.X. Huang, X.Q. Yang, and D.L. Zhu, *A sequential smooth penalization approach to mathematical programs with complementarity constraints*, manuscript, Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, 2001.

[42] H. Jiang and D. Ralph, *Extension of quasi-Newton methods to mathematical programs with complementarity constraints*, Computational Optimization and Applications, 25 (2003), 123–150.

[43] H. Jiang and D. Ralph, *QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints*, Computational Optimization and Applications, 13 (1999), 25–59.

[44] H. Jiang and D. Ralph, *Smooth SQP methods for mathematical programs with nonlinear complementarity constraints*, SIAM Journal on Optimization, 10 (2000), 779–808.

[45] K. Jittorntrum, *Solution point differentiability without strict complementarity in nonlinear programming*, Mathematical Programming Study, 21 (1984), 127–138.

[46] P. Kall and S.W. Wallace, Stochastic Programming, John Wiley & Sons, Chichester, 1994.

[47] C. Kanzow, *Some noninterior continuation methods for linear complementarity problems*, SIAM Journal on Matrix Analysis and Applications, 17 (1996), 851–868.

[48] C. Kanzow and H. Jiang, *A continuation method for (strongly) monotone variational inequalities*, Mathematical Programming, 81 (1998), 103–125.

[49] S. Karamardian and S. Schaible, *Seven kinds of monotone maps*, Journal of Optimization Theory and Applications, 66 (1990), 37–46.

[50] S. Leyffer, *MacMPEC: AMPL collection of MPECs*, Technical Report, `http://www-unix.mcs.anl.gov/~leyffer/MacMPEC/`, 2000.

[51] S. Leyffer, *The penalty interior point method fails to converge for mathematical programs with equilibrium constraints*, Numerical Analysis Report NA/208, Department of Mathematics, University of Dundee, Dundee, Scotland, 2002.

[52] G.H. Lin, X. Chen and M. Fukushima, *Smoothing implicit programming approaches for stochastic mathematical programs with linear complementarity constraints*, Technical Report 2003-006, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan (2003).

[53] G.H. Lin and M. Fukushima, *A class of stochastic mathematical programs with complementarity constraints: Reformulations and algorithms*, Technical Report 2003-010, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan, 2003.

[54] G.H. Lin and M. Fukushima, *A modified relaxation scheme for mathematical programs with complementarity constraints*, Annals of Operations Research, to appear.

[55] G.H. Lin and M. Fukushima, *Hybrid algorithms with active set identification for mathematical programs with complementarity constraints*, Technical Report 2002-008, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan, 2002.

[56] G.H. Lin and M. Fukushima, *Hybrid algorithms with index addition and subtraction strategies for solving mathematical programs with complementarity constraints*, Technical Report 2003-003, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan, 2003.

[57] G.H. Lin and M. Fukushima, *New relaxation method for mathematical programs with complementarity constraints*, Journal of Optimization Theory and Applications, 118 (2003), 81–116.

[58] G.H. Lin and M. Fukushima, *Regularization method for stochastic mathematical programs with complementarity constraints*, Technical Report 2003-012, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan, 2003.

[59] G.H. Lin and M. Fukushima, *Some exact penalty results for nonlinear programs and their applications to mathematical programs with equilibrium constraints*, Journal of Optimization Theory and Applications, 118 (2003), 67–80.

[60] X. Liu and J. Sun, *Generalized stationary pointd and an interior point method for mathematical programs with equilibrium constraints*, Preprint, National University of Singapore, Singapore, 2002.

[61] M.S. Lojasiewicz, Ensembles semi-analytiques, Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, 1964.

[62] Z.Q. Luo, J.S. Pang, and D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, United Kingdom, 1996.

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

[64] J.M. Ortega and W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, 1970.

[65] J.V. Outrata, *On optimization problems with variational inequality constraints*, SIAM Journal on Optimization, 4 (1994), 340–357.

[66] J.V. Outrata, M. Kocvara and J. Zowe, Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results, Kluwer Academic Publishers, Boston, MA, 1998.

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

[68] J.S. Pang, *Complementarity problems*, Handbook on Global Optimization, R. Horst and P. Pardalos (eds.), Kluwer Academic Publishers, B.V. Dordrecht, 1994, 271–338.

[69] J.S. Pang, *Error bounds in mathematical programming,* Mathematical Programming, 79 (1997), 299–332.

[70] J.S. Pang and M. Fukushima, *Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints*, Computational Optimization and Applications, 13 (1999), 111–136.

[71] J.S. Pang and M. Fukushima, *Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games*, Technical Report 2002-009, Department of Applied Mathematics and Physics, Kyoto University, 2002.

[72] M. Patriksson and L. Wynter, *Stochastic mathematical programs with equilibrium constraints*, Operations Research Letters, 25 (1999), 159–167.

[73] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming, 58 (1993), 353–367.

[74] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.

[75] H.S. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, Mathematics of Operations Research, 25 (2000), 1–22.

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

[77] S. Scholtes and M. Stöhr, *Exact penalization of mathematical programs with complementarity constraints*, SIAM Journal on Control and Optimization, 37 (1999), 617–652.

[78] S. Scholtes and M. Stöhr, *How stringent is the linear independence assumption for mathematical programs with complementarity constraints*, Mathematics of Operations Research, 26 (2001), 851–863.

[79] R. Schultz, L. Stougie, M.H. van der Vlerk, *Two-stage stochastic integer programming: A survey*, Statistica Neerlandica, 50 (1996), 404–416.

[80] K. Shimizu, Y. Ishizuka, and J. Bard, Nondifferentiable and Two-Level Mathematical Programming, Kluwer Academic Publishers, Boston, 1997.

[81] P. Tseng, *Growth behavior of a class of merit functions for the nonlinear complementarity problem*, Journal of Optimization Theory and Applications, 89 (1996), 17–37.

[82] N. Yamashita, H. Dan, and M. Fukushima, *On the identification of degenerate indices in the nonlinear complementarity problem with the proximal point algorithm*, Mathematical Programming, to appear.

[83] S. Vajda, Probabilistic Programming, Academic Press, New York, 1972.

[84] H. Xu, *An MPCC approach for stochastic Stackelberg-Nash-Cournot equilibrium*, manuscript, University of Southampton, Highfield Southampton, 2003.

Acknowledgements

First of all, I would like to express my sincere appreciation to Professor Masao Fukushima of Kyoto University for his supervising this thesis. He kindly gave me lots of suggestions and continual guidance. His excellent work and profound knowledge related to optimization theory and many other fields are very valuable to my research. Without his help, I could not take any progress on my research.

I am deeply grateful to Professor Xiaojun Chen of Hirosaki University and Professor Paul Tseng of University of Washington for their earnest guidance and helpful suggestions. I am also thankful to Professor Tetsuya Takine and Professor Nobuo Yamashita of Kyoto University for their friendly advices. Professors Zun-Quan Xia and Sining Zheng of Dalian University of Technology gave me continual support and encouragement. I would like to thank all of them very much.

I am particularly thankful to the Ministry of Education, Science, Sports and Culture of Japan and the Ministry of Education of China for their financially supporting me to study and research in Kyoto University.

In addition, I would like to thank all the members in Fukushima research group. While I studied in Kyoto University, I received much help from them that enriched my life in Kyoto.

Finally, I would like to pay my particular thanks to my parents, my wife, and my daughter.