SMOOTHING IMPLICIT PROGRAMMING APPROACHES FOR STOCHASTIC MATHEMATICAL PROGRAMS WITH LINEAR COMPLEMENTARITY CONSTRAINTS *

GUI-HUA LIN[†], XIAOJUN CHEN[‡] AND MASAO FUKUSHIMA[§]

May 23, 2003

Abstract In this paper, we consider the stochastic mathematical program with equilibrium constraints (SMPEC), which can be thought as a generalization of the mathematical program with equilibrium constraints. Many decision problems can be formulated as SMPECs in practice. We discuss both lower-level wait-and-see and here-and-now decision problems. 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.

Key words. stochastic mathematical program with equilibrium constraints, two-stage stochastic programming problem, implicit programming, complementarity problem, penalty function, wait-and-see, here-and-now, stationarity, nondegeneracy, convergence.

1. Introduction. A mathematical program with equilibrium constraints (MPEC) is a constrained optimization problem in which the essential constraints are defined by a parametric variational inequality:

$$\begin{array}{lll} & \text{minimize} & f(x,y) \\ (1) & \text{subject to} & x \in X, \\ & & y \text{ solves } \mathrm{VI}(F(x,\cdot), C(x)). \end{array}$$

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

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

Problem (1) can be regarded as a generalization of a bilevel programming problem and it therefore plays an important role in many fields such as engineering design, economic equilibrium, multilevel game, and mathematical programming itself. In the recent optimization study, MPECs have been receiving much attention, see the monograph of Luo et al. (1996) and the attached references, cf. Fukushima et al. (1998, 1999), Jiang and Ralph (2000), Lin and Fukushima (2003), etc.

Stochastic programming is another important branch of mathematical programming that deals with problems in which optimal decisions are sought under uncertainty. Modelling the uncertainty

*This work was supported in part by the Scientific Research Grant-in-Aid from the Ministry of Education, Science, Sports, and Culture of Japan.

[†]Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan. /Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China. E-mail: ghlin@amp.i.kyoto-u.ac.jp.

[‡]Department of Mathematical System Science, Hirosaki University, Hirosaki 036-8560, Japan. E-mail: chen@cc.hirosaki-u.ac.jp.

[§]Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan. E-mail: fuku@amp.i.kyoto-u.ac.jp.

by random objects may lead to diverse stochastic programming problems, a special case of which is the so-called two-stage stochastic program with recourse:

$$\text{(2)} \qquad \begin{array}{ll} \text{minimize} & p(x) + E_\omega[Q(x,\omega)] \\ \text{subject to} & x \in X, \end{array}$$

where $p : \Re^n \to \Re$, E_ω means expectation with respect to the random variable $\omega \in \Omega$, and

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

where $Y : \Re^n \times \Omega \to 2^{\Re^m}$ and $g : \Re^m \times \Omega \to \Re$. Problem (2) minimizes the cost of the master decision plus the expected cost of the recourse decision, where "recourse" means the ability to take corrective action after random events have taken place. Many applications of such problems can be found in practice, especially in financial planning. For further details, see Birge and Louveaux (1997), Chen (2000), and Schultz et al. (1996).

In this paper, we discuss MPECs under uncertainty. That is, we consider the stochastic mathematical program with equilibrium constraints (SMPEC):

$$\text{(3)} \qquad \begin{array}{ll} \text{minimize} & E_\omega[f(x,y,\omega)] \\ \text{subject to} & x \in X, \quad \omega \in \Omega, \\ & y \text{ solves } \mathrm{VI}(F(x,\cdot,\omega), C(x,\omega)), \end{array}$$

where $X \subseteq \Re^n$, $f : \Re^{n+m} \times \Omega \to \Re$, $F : \Re^{n+m} \times \Omega \to \Re^m$, and $C : \Re^n \times \Omega \to 2^{\Re^m}$. Obviously, if Ω is a singleton, then problem (3) 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 (Chen and Florian 1995). 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 (3), the variational inequality constraints reduce to the complementarity constraints and problem (3) is equivalent to the following stochastic mathematical program with complementarity constraints (SMPCC):

$$\text{(4)} \qquad \begin{array}{ll} \text{minimize} & E_\omega[f(x,y,\omega)] \\ \text{subject to} & x \in X, \quad \omega \in \Omega, \\ & y \geq 0, \;\; F(x,y,\omega) \geq 0, \\ & y^T F(x,y,\omega) = 0. \end{array}$$

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 $\mathrm{VI}(F(x,\cdot,\omega), C(x,\omega))$ 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 (Pang 1994). As a result, problem (3) can be reformulated as a program like (4) under some conditions, see the monograph of Luo et al. (1996)

for details. Hence, problem (4) constitutes an important subclass of SMPECs. In this paper, we concentrate on this kind of SMPECs.

Problem (4) looks like a standard MPEC. However, the existence of the random variable ω means that (4) involves multiple complementarity-type constraints and it is therefore more difficult to solve than an ordinary MPEC generally.

We call (4) 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 Patriksson and Wynter (1999), 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 paper, we will propose an implicit programming approach for (4). The problems dealt with in this paper 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:

$$(5) \qquad \begin{array}{ll} \text{minimize} & E_\omega[f(x,y,\omega) + d^T z(\omega)] \\ \text{subject to} & x \in X, \quad \omega \in \Omega, \\ & y \geq 0, \;\; F(x,y,\omega) + z(\omega) \geq 0, \\ & y^T(F(x,y,\omega) + z(\omega)) = 0, \\ & z(\omega) \geq 0, \end{array}$$

where $z(\omega)$ is called a recourse variable and $d \in \Re^m$ is a vector with positive elements. We call (5) 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 (Kall and Wallace 1994 and Vajda 1972). The following example illustrates the two models.

EXAMPLE 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

$$\begin{pmatrix} x \\ -2x \end{pmatrix} - u\begin{pmatrix} 1 \\ 0 \end{pmatrix} - v\begin{pmatrix} 1 \\ -1 \end{pmatrix} - w\begin{pmatrix} 0 \\ -1 \end{pmatrix} = 0, \tag{6}$$

$$0 \leq u \;\perp\; (y-c) \geq 0,$$
$$0 \leq v \;\perp\; (y-t) \geq 0,$$
$$0 \leq w \;\perp\; (D(\omega) - d(\omega)x - t) \geq 0. \tag{7}$$

Here, $\lambda \perp \mu$ means $\lambda\mu = 0$. It follows from (6) that

$$u = x - v, \quad w = 2x - v.$$

This implies that $w = x + u \geq a > 0$, which together with (7) yields $t = D(\omega) - d(\omega)x$. Thus the above optimality conditions may further be rewritten as

$$\begin{array}{l} 0 \leq (x-v) \;\perp\; (y-c) \geq 0, \\ 0 \leq v \;\perp\; (y - D(\omega) + d(\omega)x) \geq 0. \end{array} \tag{8}$$

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 (8) for all $\omega \in \Omega$ in general. So, by introducing the recourse variables, the company's problem is represented as the following model:

$$\begin{array}{ll} \text{minimize} & -xy + \beta E_\omega[\, z(\omega) \,] \\ \text{subject to} & a \leq x \leq b, \\ & 0 \leq (x-v) \;\perp\; (y-c) \geq 0, \\ & 0 \leq v \;\perp\; (y - D(\omega) + d(\omega)x + z(\omega)) \geq 0, \\ & z(\omega) \geq 0, \quad \omega \in \Omega, \end{array}$$

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

$$\begin{aligned} &0 \le (x - v(\omega)) \ \perp \ (y(\omega) - c) \ge 0, \\ &0 \le v(\omega) \ \perp \ (y(\omega) - D(\omega) + d(\omega)x) \ge 0 \end{aligned}$$

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 the paper: 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 is also included. In Section 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.

Notation used in the paper: Throughout, 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 a given function $F : \Re^n \to \Re^m$ and a vector $x \in \Re^n$, $\nabla F(x)$ is the transposed Jacobian of F at x and $\mathcal{I}_F(x) := \{i \mid F_i(x) = 0\}$ stands for the active index set of F at x. 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.

2. A 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):

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{\ell=1}^{L} p_\ell f(x, y_\ell) \\ \text{subject to} & g(x) \le 0, \ h(x) = 0, \\ & y_\ell \ge 0, \ N_\ell x + M_\ell y_\ell + q_\ell \ge 0, \\ & y_\ell^T (N_\ell x + M_\ell y_\ell + q_\ell) = 0, \ \ell = 1, 2, \cdots, L, \end{array} \tag{9}$$

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.,

$$\textstyle\sum_{\ell=1}^{L} p_\ell = 1, \quad p_\ell \ge 0, \qquad \ell = 1, \cdots, L,$$

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 (9) 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 $\ell = 1, 2, \cdots, L$.

By letting

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_L \end{pmatrix}, \quad \mathbf{N} = \begin{pmatrix} N_1 \\ \vdots \\ N_L \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} M_1 & & O \\ & \ddots & \\ O & & M_L \end{pmatrix}, \quad \mathbf{q} = \begin{pmatrix} q_1 \\ \vdots \\ q_L \end{pmatrix},$$

and $\mathrm{f}(x, \mathbf{y}) = \sum_{\ell=1}^{L} p_\ell f(x, y_\ell)$, problem (9) can be rewritten as

$$\begin{array}{lll} & \text{minimize} & \mathrm{f}(x, \mathbf{y}) \\ (10) & \text{subject to} & g(x) \leq 0,\ h(x) = 0, \\ & & \mathbf{y} \geq 0,\ \mathbf{N}x + \mathbf{M}\mathbf{y} + \mathbf{q} \geq 0, \\ & & \mathbf{y}^T(\mathbf{N}x + \mathbf{M}\mathbf{y} + \mathbf{q}) = 0, \end{array}$$

which is an ordinary MPEC. However, since L is usually very large in practice, problem (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 (9) as a mathematical program with multiple complementarity-type constraints.

Recently, Chen and Fukushima (2003) have suggested a smoothing method for an MPEC with P-matrix linear complementarity constraints. We will develop a similar smoothing method for solving problem (9), or equivalently (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.

2.1. Preliminaries. We first recall some basic concepts and properties that will be used later on.

Definition 1 (Cottle et al. 1992) 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 2 (Cottle et al. 1992) 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

$$(11) \qquad \Phi_{\epsilon,\mu}(x, y, w; N, M, q) = \begin{pmatrix} Nx + (M + \epsilon I)y + q - w \\ \phi_\mu(y[1], w[1]) \\ \vdots \\ \phi_\mu(y[m], w[m]) \end{pmatrix},$$

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 (Chen and Fukushima 2003 and Kanzow 1996).

THEOREM 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:*

(i) *The function $\Phi_{\epsilon,\mu}$ defined by (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;*

(ii) *The equation $\Phi_{\epsilon,\mu}(x,y,w;N,M,q) = 0$ has a unique solution $(y(x,\epsilon,\mu), w(x,\epsilon,\mu))$, which is continuously differentiable with respect to x and satisfies*

$$\begin{aligned} & y(x,\epsilon,\mu) > 0, \quad w(x,\epsilon,\mu) > 0, \\ & y(x,\epsilon,\mu)[i]w(x,\epsilon,\mu)[i] = \mu^2, \quad i = 1,\cdots,m. \end{aligned} \tag{12}$$

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 (9) and $\mu > 0$. Theorem 1 indicates that, for any fixed ℓ and $\mu > 0$, the smooth equation

$$\Phi_\mu(x, y_\ell, w_\ell; N_\ell, M_\ell, q_\ell) = 0 \tag{13}$$

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 (13) is equivalent to the system

$$\begin{aligned} & y_\ell \ge 0, \ N_\ell x + (M_\ell + \mu I)y_\ell + q_\ell \ge 0, \\ & y_\ell[i]\Big(N_\ell x + (M_\ell + \mu I)y_\ell + q_\ell\Big)[i] = \mu^2, \quad i = 1,\cdots,m \end{aligned} \tag{14}$$

in the sense that $y_\ell(x,\mu)$ solves (14) if and only if

$$\Phi_\mu(x, y_\ell(x,\mu), w_\ell(x,\mu); N_\ell, M_\ell, q_\ell) = 0 \tag{15}$$

with

$$w_\ell(x,\mu) = N_\ell x + (M_\ell + \mu I)y_\ell(x,\mu) + q_\ell \ . \tag{16}$$

Since (14) with $\mu = 0$ reduces to the linear complementarity problem LCP$(x; N_\ell, M_\ell, q_\ell)$:

$$\begin{aligned} & y_\ell \ge 0, \ N_\ell x + M_\ell y_\ell + q_\ell \ge 0, \\ & y_\ell^T\Big(N_\ell x + M_\ell y_\ell + q_\ell\Big) = 0, \end{aligned} \tag{17}$$

we see that $y_\ell(x,\mu)$ tends to a solution of (17) 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 (17) 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 (14), see Chen and Fukushima (2003).

The following lemma will be used later on.

LEMMA 1 (Fukushima et al. 1998) *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 $\hat{\mathcal{K}} = \{\ i \mid\ \hat{d}_i = 1\}$, then*

$$\begin{pmatrix} M & -I \\ I - \hat{D} & \hat{D} \end{pmatrix}$$

is nonsingular.

2.2. Method. Let $\{\mu_k\}$ be a sequence of positive numbers converging to 0. A smoothing implicit programming method for problem (9), which we call SIP-I, generates a sequence $\{x^{(k)}\}$ by solving the problems

$$\begin{array}{ll} \text{minimize} & \theta_{\mu_k}(x) \\ \text{subject to} & g(x) \le 0,\ h(x) = 0, \end{array} \tag{18}$$

where

$$\theta_{\mu_k}(x) = \sum_{\ell=1}^{L} p_\ell f(x, y_\ell(x, \mu_k))$$

and $y_\ell(x, \mu_k)$ satisfies the system (15)–(16) with $\mu = \mu_k$. Let

$$y_\ell^{(k)} = y_\ell(x^{(k)}, \mu_k), \quad \ell = 1, \cdots, L.$$

Throughout, we denote by $\mathcal{F}_1$ and $\mathcal{X}$ the feasible regions of problems (9) and (18), 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. In addition, we let the bold letter $\mathbf{y}$ stand for the vector $(y_1^T, \cdots, y_L^T)^T$ for simplicity.

Note that, by Theorem 1, problem (18) is a smooth mathematical program. In particular, if $\mathcal{X} = \Re^n$, problem (18) reduces to a smooth unconstrained optimization problem. Moreover, under some suitable conditions, (18) is a convex program, see Chen and Fukushima (2003) for details. Therefore, we may expect that problem (18) may be relatively easy to deal with, provided the evaluation of the function $y_\ell(x, \mu_k)$ is not very expensive.

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 a sequence of local optimal solutions of problems (18). To this end, the following assumptions are made throughout this subsection:

A1: SIP-I produces a bounded sequence $\{x^{(k)}\}$ of local optimal solutions of (18).

A2: For any bounded sequence $\{x^k\}$ in $\mathcal{X}$, the sequence $\{y_\ell(x^k, \mu_k)\}$ is bounded for each ℓ.

LEMMA 2 *Suppose that all M_ℓ, $\ell = 1, \cdots, L$, are P_0-matrices in problem (9) and $(x^*, \mathbf{y}^*) \in \mathcal{F}_1$. Assume that, for each ℓ, the submatrix $M_\ell[\mathcal{K}_\ell^*]$ is nondegenerate, where*

$$\mathcal{K}_\ell^* \ = \ \{\ i \mid\ (N_\ell x^* + M_\ell y_\ell^* + q_\ell)[i] = 0\}.$$

Then there exist a neighborhood U^* of $(x^*, \mathbf{y}^*)$ and a positive constant π^* such that

$$\|y_\ell(x, \mu_k) - y_\ell\| \le \mu_k \pi^* (\|y_\ell\| + \sqrt{m}), \quad \ell = 1, \cdots, L \tag{19}$$

holds for any $(x, \mathbf{y}) \in U^* \cap \mathcal{F}_1$ *and every* k.

PROOF: For any $(x, \mathbf{y}) \in \Re^{n+mL}$, we let

$$w_\ell = N_\ell x + M_\ell y_\ell + q_\ell, \quad \ell = 1, \cdots, L.$$

Since $(x, \mathbf{y}) \in \mathcal{F}_1$ implies

$$\Phi_0(x, y_\ell, w_\ell; N_\ell, M_\ell, q_\ell) = 0, \quad \ell = 1, \cdots, L, \tag{20}$$

(15) and (20) yield

$$\begin{aligned} 0 &= \Phi_{\mu_k}(x, y_\ell(x, \mu_k), w_\ell(x, \mu_k); N_\ell, M_\ell, q_\ell) - \Phi_0(x, y_\ell, w_\ell; N_\ell, M_\ell, q_\ell) \\ &= \begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell(x, \mu_k) & D_\ell(x, \mu_k) \end{pmatrix} \begin{pmatrix} y_\ell(x, \mu_k) - y_\ell \\ w_\ell(x, \mu_k) - w_\ell \end{pmatrix} - \mu_k \begin{pmatrix} -y_\ell \\ 2a_\ell^k[1]\mu_k \\ \vdots \\ 2a_\ell^k[m]\mu_k \end{pmatrix}. \end{aligned} \tag{21}$$

Here, $D_\ell(x, \mu_k) = \mathrm{diag}\Big(a_\ell^k[i](y_\ell(x, \mu_k)[i] + y_\ell[i])\Big)$ and

$$\begin{aligned} a_\ell^k[i] &= \frac{1}{\sqrt{(y_\ell(x, \mu_k)[i])^2 + (w_\ell(x, \mu_k)[i])^2 + 2\mu_k^2} + \sqrt{(y_\ell[i])^2 + (w_\ell[i])^2}} \\ &= \frac{1}{y_\ell(x, \mu_k)[i] + w_\ell(x, \mu_k)[i] + y_\ell[i] + w_\ell[i]}, \end{aligned} \tag{22}$$

where the last equality follows from (15) and (20). From (12), we see that, for any ℓ, i, k,

$$0 < a_\ell^k[i](y_\ell(x, \mu_k)[i] + y_\ell[i]) < 1. \tag{23}$$

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 (Fukushima et al. 1998). We next prove that there exists a neighborhood U^* of $(x^*, \mathbf{y}^*)$ and a positive constant π^* such that

$$\left\| \begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell(x, \mu_k) & D_\ell(x, \mu_k) \end{pmatrix}^{-1} \right\| \le \pi^*, \quad \ell = 1, \cdots, L \tag{24}$$

holds for any $(x, \mathbf{y}) \in U^* \cap \mathcal{F}_1$ and any k. Otherwise, there must be an index ℓ, a subsequence $\{k_j\}$ of $\{k\}$, and a sequence $\{(x^j, \mathbf{y}^j)\} \subset \mathcal{F}_1$ such that

$$\lim_{j\to\infty} (x^j, \mathbf{y}^j) = (x^*, \mathbf{y}^*), \tag{25}$$

$$\lim_{j\to\infty} \begin{pmatrix} M_\ell + \mu_{k_j} I & -I \\ I - D_\ell(x^j, \mu_{k_j}) & D_\ell(x^j, \mu_{k_j}) \end{pmatrix} = \begin{pmatrix} M_\ell & -I \\ I - \hat{D}_\ell & \hat{D}_\ell \end{pmatrix} := H, \tag{26}$$

where $\hat{D}_\ell := \mathrm{diag}(\hat{d}_\ell[1], \cdots, \hat{d}_\ell[m])$ with $0 \le \hat{d}_\ell[i] \le 1$ for each i and the matrix H is singular. Note that, by (22),

$$\hat{d}_\ell[i] = \lim_{j\to\infty} \frac{y_\ell(x^j, \mu_{k_j})[i] + y_\ell^j[i]}{y_\ell(x^j, \mu_{k_j})[i] + w_\ell(x^j, \mu_{k_j})[i] + y_\ell^j[i] + w_\ell^j[i]}.$$

By Assumption A2, the sequence $\{y_\ell(x^j, \mu_{k_j})\}$ is bounded and then so is the sequence $\{w_\ell(x^j, \mu_{k_j})\}$. It is not difficult to see that

$$\hat{d}_\ell[i] = 1 \quad \Rightarrow \quad \lim_{j\to\infty} \Big(w_\ell(x^j, \mu_{k_j})[i] + w^j_\ell[i] \Big) = 0 \quad \Rightarrow \quad \lim_{j\to\infty} w^j_\ell[i] = 0.$$

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 1, H is nonsingular. This is a contradiction and hence we have (24).

Moreover, we have from (12) that, for every k, $\ell = 1, \cdots, L$, and $i = 1, \cdots, m$,

$$\begin{aligned} 2a^k_\ell[i]\mu_k &\le \frac{2\mu_k}{\sqrt{(y_\ell(x,\mu_k)[i])^2 + (w_\ell(x,\mu_k)[i])^2 + 2\mu_k^2}} \\ &\le \frac{2\mu_k}{\sqrt{2y_\ell(x,\mu_k)[i]w_\ell(x,\mu_k)[i] + 2\mu_k^2}} \\ &= 1. \end{aligned} \tag{27}$$

Thus, it follows from (21), (24), and (27) that

$$\begin{aligned} \|y_\ell(x,\mu_k) - y_\ell\| &\le \left\| \begin{pmatrix} y_\ell(x,\mu_k) - y_\ell \\ w_\ell(x,\mu_k) - w_\ell \end{pmatrix} \right\| \\ &= \mu_k \left\| \begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell(x,\mu_k) & D_\ell(x,\mu_k) \end{pmatrix}^{-1} \begin{pmatrix} -y_\ell \\ 2a^k_\ell[1]\mu_k \\ \vdots \\ 2a^k_\ell[m]\mu_k \end{pmatrix} \right\| \\ &\le \mu_k \pi^* (\|y_\ell\| + \sqrt{m}). \end{aligned} \tag{28}$$

Note that the above inequalities hold for all ℓ and then the proof is completed. □

Under Assumptions A1 and A2, the sequence $\{(x^{(k)}, \mathbf{y}^{(k)})\}$ generated by SIP-I is bounded. In the rest of this subsection, we further make the following assumptions:

A3: The sequence $\{(x^{(k)}, \mathbf{y}^{(k)})\}$ generated by SIP-I is convergent to a point $(x^*, \mathbf{y}^*)$.

A4: There exists a neighborhood V^* of $(x^*, \mathbf{y}^*)$ such that $x^{(k)}$ minimizes θ_{μ_k} over $V^*|_\mathcal{X}$ for all k large enough, where $V^*|_\mathcal{X} = \{x \in \mathcal{X} \mid (x, \mathbf{y}) \in V^*\}$.

THEOREM 2 *Suppose that all matrices M_ℓ, $\ell = 1, \cdots, L$, are P_0-matrices in problem (9). Let $\{(x^{(k)}, \mathbf{y}^{(k)})\}$ and $(x^*, \mathbf{y}^*)$ 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 2, then $(x^*, \mathbf{y}^*)$ is a local optimal solution of problem (9).*

PROOF: First we have $(x^*, \mathbf{y}^*) \in \mathcal{F}_1$ immediately from (14). By Lemma 2, there exist a closed sphere $\mathcal{B}_1 \subseteq V^*$ centered at the point $(x^*, \mathbf{y}^*)$ with positive radius and a positive number π^* such that (19) holds for any $(x, \mathbf{y}) \in \mathcal{F}_1 \cap \mathcal{B}_1$ and every k. Since $\mathcal{F}_1 \cap \mathcal{B}_1$ is a nonempty compact set, the continuity of f ensures that the problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{\ell=1}^{L} p_\ell f(x, y_\ell) \\ \text{subject to} & (x, \mathbf{y}) \in \mathcal{F}_1 \cap \mathcal{B}_1 \end{array} \tag{29}$$

has a nonempty solution set. Let $(\bar{x}, \bar{\mathbf{y}})$ be an arbitrary solution of (29).

For any $(x, \mathbf{y}) \in \mathcal{F}_1 \cap \mathcal{B}_1$, by the mean-value theorem, we have

$$\begin{aligned} \theta_{\mu_k}(x) &= \sum_{\ell=1}^{L} p_\ell f(x, y_\ell(x, \mu_k)) \\ &= \sum_{\ell=1}^{L} p_\ell \Big(f(x, y_\ell) + (y_\ell(x, \mu_k) - y_\ell)^T \nabla_y f(x, t_\ell y_\ell + (1 - t_\ell) y_\ell(x, \mu_k)) \Big), \end{aligned} \tag{30}$$

where $t_\ell \in [0, 1]$ for each ℓ. By (19), we have

$$\begin{aligned} \|t_\ell y_\ell + (1 - t_\ell) y_\ell(x, \mu_k)\| &= \|(1 - t_\ell)(y_\ell(x, \mu_k) - y_\ell) + y_\ell\| \\ &\le \|y_\ell(x, \mu_k) - y_\ell\| + \|y_\ell\| \\ &\le \mu_k \pi^* (\|y_\ell\| + \sqrt{m}) + \|y_\ell\|. \end{aligned} \tag{31}$$

Since $\mathcal{F}_1 \cap \mathcal{B}_1$ is bounded, it follows from (31) that the set

$$\{(x, t_\ell y_\ell + (1 - t_\ell) y_\ell(x, \mu_k)) \;\Big|\; (x, \mathbf{y}) \in \mathcal{F}_1 \cap \mathcal{B}_1, k = 1, 2, \cdots, t_\ell \in [0, 1]\}$$

is bounded for each $\ell = 1, \cdots, L$, and so there exists a constant $\tau > 0$ such that

$$\|\nabla_y f(x, t_\ell y_\ell + (1 - t_\ell) y_\ell(x, \mu_k))\| \le \tau, \quad \ell = 1, \cdots, L$$

holds for any $(x, \mathbf{y}) \in \mathcal{F}_1 \cap \mathcal{B}_1, t_\ell \in [0, 1]$, and every k. It then follows from (30) and (19) 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 $(x, \mathbf{y}) \in \mathcal{F}_1 \cap \mathcal{B}_1$ and k. In particular,

$$\Big|\theta_{\mu_k}(\bar{x}) - \sum_{\ell=1}^{L} p_\ell f(\bar{x}, \bar{y}_\ell)\Big| \le \tau \mu_k \pi^* \sum_{\ell=1}^{L} \Big(\|\bar{y}_\ell\| + \sqrt{m}\Big), \quad \forall k. \tag{32}$$

On the one hand, the continuity of f yields

$$\lim_{k \to \infty} \theta_{\mu_k}(x^{(k)}) = \sum_{\ell=1}^{L} p_\ell \lim_{k \to \infty} f(x^{(k)}, y_\ell^{(k)}) = \sum_{\ell=1}^{L} p_\ell f(x^*, y_\ell^*). \tag{33}$$

Note that, by Assumption A4 and the fact that $\mathcal{F}_1 \cap \mathcal{B}_1 \subseteq V^*$, $x^{(k)}$ is an optimal solution of the problem

$$\begin{array}{ll} \text{minimize} & \theta_{\mu_k}(x) \\ \text{subject to} & x \in \mathcal{X}_1 := \{\, x \in \mathcal{X} \mid \; (x, \mathbf{y}) \in \mathcal{F}_1 \cap \mathcal{B}_1 \,\}, \end{array}$$

provided k is large enough, and $\bar{x}$ is a feasible point of this problem. We then have from (32) that, for every sufficiently large k,

$$\theta_{\mu_k}(x^{(k)}) \le \theta_{\mu_k}(\bar{x}) \le \sum_{\ell=1}^{L} p_\ell f(\bar{x}, \bar{y}_\ell) + \tau \mu_k \pi^* \sum_{\ell=1}^{L} \Big(\|\bar{y}_\ell\| + \sqrt{m}\Big). \tag{34}$$

Therefore, taking into account the equality (33) and the fact that $\mu_k \to 0$ as $k \to \infty$, we have by letting $k \to \infty$ in (34) that

$$\sum_{\ell=1}^{L} p_\ell f(x^*, y_\ell^*) \leq \sum_{\ell=1}^{L} p_\ell f(\bar{x}, \bar{y}_\ell).$$

On the other hand, since $(\bar{x}, \bar{\mathbf{y}})$ is a solution of problem (29), we have

$$\sum_{\ell=1}^{L} p_\ell f(x^*, y_\ell^*) \geq \sum_{\ell=1}^{L} p_\ell f(\bar{x}, \bar{y}_\ell).$$

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 $(x^*, \mathbf{y}^*)$ is an optimal solution of problem (29), in other words, $(x^*, \mathbf{y}^*)$ is a local optimal solution of problem (9). This completes the proof. $\Box$

THEOREM 3 *Suppose that all M_ℓ, $\ell = 1, \cdots, L$, are P-matrices in problem (9) and, for each k, $x^{(k)}$ is a global optimal solution of problem (18). Then the limit point $(x^*, \mathbf{y}^*)$ is a global optimal solution of problem (9).*

PROOF: Recall that $(x^*, \mathbf{y}^*) \in \mathcal{F}_1$. Since all M_ℓ are P-matrices (and hence nondegenerate), the conditions of Lemma 2 are satisfied at every point $(x, \mathbf{y}) \in \mathcal{F}_1$. Thus, for any point $(\hat{x}, \hat{\mathbf{y}}) \in \mathcal{F}_1$, there exist a neighborhood $\hat{U} = U(\hat{x}, \hat{\mathbf{y}})$ of $(\hat{x}, \hat{\mathbf{y}})$ and a positive constant $\hat{\pi} = \pi(\hat{x}, \hat{\mathbf{y}})$ such that for any $(x, \mathbf{y}) \in \hat{U} \cap \mathcal{F}_1$,

$$\|y_\ell(x, \mu_k) - y_\ell\| \leq \mu_k \hat{\pi}(\|y_\ell\| + \sqrt{m})$$

holds for each ℓ and k.

For an arbitrary positive number η, we define the set $\mathcal{F}_{1,\eta}$ by

$$\mathcal{F}_{1,\eta} = \{(x, \mathbf{y}) \in \mathcal{F}_1 \;\Big|\; \|(x, \mathbf{y}) - (x^*, \mathbf{y}^*)\| \leq \eta\}. \tag{35}$$

It is clear that $\mathcal{F}_{1,\eta}$ is a nonempty compact set. Since the set of neighborhoods

$$\mathcal{U} = \{U(\hat{x}, \hat{\mathbf{y}}) \;\Big|\; (\hat{x}, \hat{\mathbf{y}}) \in \mathcal{F}_{1,\eta}\}$$

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 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 (19) for any $(x, \mathbf{y}) \in \mathcal{F}_{1,\eta}$ and every k.

In a similar way to the proof of Theorem 2, we can show that $(x^*, \mathbf{y}^*)$ is an optimal solution of the problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{\ell=1}^{L} p_\ell f(x, y_\ell) \\ \text{subject to} & (x, \mathbf{y}) \in \mathcal{F}_{1,\eta}. \end{array}$$

Noting that $\eta > 0$ is arbitrary, we see that $(x^*, \mathbf{y}^*)$ is actually a global optimal solution of problem (9) and so the proof is completed. □

2.4. Limiting behavior of stationary points. In the last subsection, we have discussed the convergence of optimal solutions of problems (18). In practice, it may not be easy to obtain an 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 (18).

For problem (18), we will use the standard definition of stationarity, i.e., $x \in \mathcal{X}$ is said to be *stationary* to (18) if there exist Lagrange multiplier vectors λ and ν satisfying the Karush-Kuhn-Tucker conditions

$$\begin{aligned}&\nabla\theta_{\mu_k}(x) + \nabla g(x)\lambda + \nabla h(x)\nu = 0,\\ &\lambda \geq 0, \quad g(x)^T\lambda = 0.\end{aligned}$$

Recalling that problem (9) is equivalent to an ordinary MPEC (10), we employ the same terminologies for (9) as in the study of MPECs. Denote

$$F(x, \mathbf{y}) = \mathbf{N}x + \mathbf{M}\mathbf{y} + \mathbf{q}, \quad G(x, \mathbf{y}) = \mathbf{y}$$

in problem (10) and suppose that $(x^*, \mathbf{y}^*)$ is a feasible point of (10).

Definition 3 The MPEC-*linear independence constraint qualification* (MPEC-LICQ) is said to hold at $(x^*, \mathbf{y}^*)$ if

$$\left\{ \begin{pmatrix} \nabla g_l(x^*) \\ 0 \end{pmatrix}, \begin{pmatrix} \nabla h_r(x^*) \\ 0 \end{pmatrix}, \nabla F_i(x^*, \mathbf{y}^*), \nabla G_j(x^*, \mathbf{y}^*) \;\middle|\; l \in \mathcal{I}_g^* := \mathcal{I}_g(x^*),\ 1 \leq r \leq s_2,\ i \in \mathcal{I}_F(x^*, \mathbf{y}^*),\ j \in \mathcal{I}_G(x^*, \mathbf{y}^*) \right\}$$

is linearly independent.

This condition is not particularly stringent and has been assumed often in the literature on MPECs (cf. Fukushima and Pang 1999, Lin and Fukushima 2002, and Scholtes 2001). Note that this definition is different from the standard definition of LICQ in nonlinear programming theory that would require the gradient of the function $F(x, \mathbf{y})^T G(x, \mathbf{y})$ be linearly independent of the above vectors, which cannot happen in any case actually.

Definition 4 Assume that the MPEC-LICQ holds at $(x^*, \mathbf{y}^*)$. We say $(x^*, \mathbf{y}^*)$ is a *Clarke or C-stationary* point if there exist multiplier vectors u^*, v^*, λ^*, and ν^* such that

$$\begin{aligned}&\nabla \mathrm{f}(x^*, \mathbf{y}^*) - \sum_{i \in \mathcal{I}_F(x^*, \mathbf{y}^*)} u^*[i] \nabla F_i(x^*, \mathbf{y}^*) - \sum_{i \in \mathcal{I}_G(x^*, \mathbf{y}^*)} v^*[i] \nabla G_i(x^*, \mathbf{y}^*)\\ &\quad + \sum_{i \in \mathcal{I}_g^*} \lambda^*[i] \begin{pmatrix} \nabla g_i(x^*) \\ 0 \end{pmatrix} + \sum_{i=1}^{s_2} \nu^*[i] \begin{pmatrix} \nabla h_i(x^*) \\ 0 \end{pmatrix} = 0,\end{aligned} \tag{36}$$

$$\lambda^* \geq 0, \qquad u^*[i] v^*[i] \geq 0 \quad \text{for} \;\; i \in \mathcal{I}_F(x^*, \mathbf{y}^*) \cap \mathcal{I}_G(x^*, \mathbf{y}^*). \tag{37}$$

If furthermore, there holds

$$u^*[i] \geq 0, \quad v^*[i] \geq 0, \qquad \forall i \in \mathcal{I}_F(x^*, \mathbf{y}^*) \cap \mathcal{I}_G(x^*, \mathbf{y}^*), \tag{38}$$

we call $(x^*, \mathbf{y}^*)$ a *Bouligand or B-stationary* point.

Definition 5 A solution y_ℓ^* of problem LCP$(x^*; N_\ell, M_\ell, q_\ell)$ is said to satisfy the *strict complementarity condition* if $\mathcal{I}^*_{W_\ell} \cap \mathcal{I}^*_{Y_\ell} = \emptyset$, where

$$\begin{aligned} \mathcal{I}^*_{W_\ell} &= \{i \mid (N_\ell x^* + M_\ell y_\ell^* + q_\ell)[i] = 0\}, \\ \mathcal{I}^*_{Y_\ell} &= \{i \mid y_\ell^*[i] = 0\}. \end{aligned}$$

Theorem 4 *Suppose that all M_ℓ, $\ell = 1, \cdots, L$, are P_0-matrices in problem (9) and for each k, $x^{(k)}$ is a stationary point of (18). Let $(x^*, \mathbf{y}^*)$ be an accumulation point of the sequence $\{(x^{(k)}, \mathbf{y}^{(k)})\}$ generated by SIP-I. If the MPEC-LICQ is satisfied at $(x^*, \mathbf{y}^*)$, then $(x^*, \mathbf{y}^*)$ is a C-stationary point of problem (9). In particular, if y_ℓ^* satisfies the strict complementarity condition for each ℓ, then $(x^*, \mathbf{y}^*)$ is a B-stationary point.*

Proof: Assume without loss of generality that the sequence $\{(x^{(k)}, \mathbf{y}^{(k)})\}$ converges to $(x^*, \mathbf{y}^*)$. Since the MPEC-LICQ holds at $(x^*, \mathbf{y}^*)$, it is obvious that, for every k sufficiently large, problem (18) satisfies the standard LICQ at $x^{(k)}$ and then, by the stationarity of $x^{(k)}$, there exist unique Lagrange multiplier vectors λ^k and ν^k such that

$$\nabla \theta_{\mu_k}(x^{(k)}) + \nabla g(x^{(k)})\lambda^k + \nabla h(x^{(k)})\nu^k = 0, \tag{39}$$

$$g(x^{(k)}) \le 0, \qquad h(x^{(k)}) = 0, \tag{40}$$

$$\lambda^k \ge 0, \qquad g(x^{(k)})^T \lambda^k = 0. \tag{41}$$

Moreover, it follows from the continuity of g that, for any k sufficiently large,

$$\mathcal{I}_g(x^{(k)}) \subseteq \mathcal{I}_g^*. \tag{42}$$

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 (Ortega and Rheinboldt 1970), we have

$$\begin{aligned} &\begin{pmatrix} \nabla y_\ell(x^{(k)}, \mu_k)^T \\ \nabla w_\ell(x^{(k)}, \mu_k)^T \end{pmatrix} \\ = & -\begin{pmatrix} \nabla_{y_\ell} \Phi_{\mu_k}(x^{(k)}, y^{(k)}, w^{(k)}; N_\ell, M_\ell, q_\ell) \\ \nabla_{w_\ell} \Phi_{\mu_k}(x^{(k)}, y^{(k)}, w^{(k)}; N_\ell, M_\ell, q_\ell) \end{pmatrix}^{-T} \nabla_x \Phi_{\mu_k}(x^{(k)}, y^{(k)}, w^{(k)}; N_\ell, M_\ell, q_\ell)^T \\ = & -\begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell^k & D_\ell^k \end{pmatrix}^{-1} \begin{pmatrix} N_\ell \\ O \end{pmatrix}, \end{aligned} \tag{43}$$

where $D_\ell^k = \mathrm{diag}\left(\frac{y_\ell^{(k)}[1]}{y_\ell^{(k)}[1] + w_\ell^{(k)}[1]}, \cdots, \frac{y_\ell^{(k)}[m]}{y_\ell^{(k)}[m] + w_\ell^{(k)}[m]} \right)$ and the existence of the inverse matrix follows from Theorem 1. Furthermore, since

$$\begin{pmatrix} M_\ell + \mu_k I & -I \\ I - D_\ell^k & D_\ell^k \end{pmatrix}^{-1} = \begin{pmatrix} E_\ell^k D_\ell^k & E_\ell^k \\ -I + (M_\ell + \mu_k I) E_\ell^k D_\ell^k & (M_\ell + \mu_k I) E_\ell^k \end{pmatrix} \tag{44}$$

with

$$E_\ell^k = \left(D_\ell^k M_\ell + I - (1 - \mu_k) D_\ell^k \right)^{-1} \tag{45}$$

(see Fukushima et al. 1998), it follows from (43) that

$$\nabla y_\ell(x^{(k)}, \mu_k) = -N_\ell^T D_\ell^k (E_\ell^k)^T, \quad \ell = 1, \cdots, L. \tag{46}$$

Thus, (39) becomes

$$
\begin{aligned}
0 &= \nabla\theta_{\mu_k}(x^{(k)}) + \nabla g(x^{(k)})\lambda^k + \nabla h(x^{(k)})\nu^k \\
&= \sum_{\ell=1}^{L} p_\ell\Big(\nabla_x f(x^{(k)}, y_\ell^{(k)}) + \nabla y_\ell(x^{(k)}, \mu_k)\nabla_y f(x^{(k)}, y_\ell^{(k)})\Big) + \nabla g(x^{(k)})\lambda^k + \nabla h(x^{(k)})\nu^k \\
&= \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y_\ell^{(k)}) - \sum_{\ell=1}^{L} p_\ell N_\ell^T D_\ell^k (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}) + \nabla g(x^{(k)})\lambda^k + \nabla h(x^{(k)})\nu^k \\
&= \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y_\ell^{(k)}) - \sum_{\ell=1}^{L} N_\ell^T u_\ell^k + \nabla g(x^{(k)})\lambda^k + \nabla h(x^{(k)})\nu^k, \qquad (47)
\end{aligned}
$$

where u_ℓ^k is defined by

$$
u_\ell^k = p_\ell D_\ell^k (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}), \tag{48}
$$

that is,

$$
u_\ell^k[i] = \frac{p_\ell\ y_\ell^{(k)}[i]}{y_\ell^{(k)}[i] + w_\ell^{(k)}[i]}\ e_i^T (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}), \quad i = 1, \cdots, m. \tag{49}
$$

For each ℓ, we let v_ℓ^k satisfy

$$
p_\ell \nabla_y f(x^{(k)}, y_\ell^{(k)}) - M_\ell^T u_\ell^k - v_\ell^k = 0. \tag{50}
$$

It then follows from (48) and (45) that, for every $\ell = 1, \cdots, L$,

$$
\begin{aligned}
v_\ell^k &= p_\ell \nabla_y f(x^{(k)}, y_\ell^{(k)}) - M_\ell^T u_\ell^k \\
&= p_\ell \nabla_y f(x^{(k)}, y_\ell^{(k)}) - p_\ell M_\ell^T D_\ell^k (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}) \\
&= p_\ell\Big((E_\ell^k)^{-T} - M_\ell^T D_\ell^k\Big)(E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}) \\
&= p_\ell\Big(I - (1-\mu_k) D_\ell^k\Big)(E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)})
\end{aligned}
$$

and hence

$$
v_\ell^k[i] = \frac{p_\ell(\mu_k y_\ell^{(k)}[i] + w_\ell^{(k)}[i])}{y_\ell^{(k)}[i] + w_\ell^{(k)}[i]}\ e_i^T (E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)}), \quad i = 1, \cdots, m. \tag{51}
$$

Taking into account (40)–(42), we obtain from (47) and (50) that

$$
\begin{aligned}
0 &= \begin{pmatrix} \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y_\ell^{(k)}) \\ p_1 \nabla_y f(x^{(k)}, y_1^{(k)}) \\ \vdots \\ p_L \nabla_y f(x^{(k)}, y_L^{(k)}) \end{pmatrix} - \begin{pmatrix} N_1^T & \cdots & N_L^T \\ M_1^T & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & M_L^T \end{pmatrix} u^k - \begin{pmatrix} O & \cdots & O \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix} v^k \\
&\quad + \sum_{i \in \mathcal{I}_g^*} \lambda^k[i] \begin{pmatrix} \nabla g_i(x^{(k)}) \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \nu^k, \qquad (52)
\end{aligned}
$$

where $u^k = \left((u_1^k)^T, \cdots, (u_L^k)^T\right)^T$ and $v^k = \left((v_1^k)^T, \cdots, (v_L^k)^T\right)^T$. We can further rewrite (52) as

$$\begin{pmatrix} \sum_{\ell=1}^L p_\ell \nabla_x f(x^{(k)}, y_\ell^{(k)}) \\ p_1 \nabla_y f(x^{(k)}, y_1^{(k)}) \\ \vdots \\ p_L \nabla_y f(x^{(k)}, y_L^{(k)}) \end{pmatrix} - \sum_{\ell=1}^L \sum_{i \notin \mathcal{I}_{W_\ell}^*} u_\ell^k[i] \begin{pmatrix} N_\ell[i] \\ \vdots \\ M_\ell[i] \\ \vdots \\ 0 \end{pmatrix} - \sum_{\ell=1}^L \sum_{i \notin \mathcal{I}_{Y_\ell}^*} v_\ell^k[i] \begin{pmatrix} 0 \\ \vdots \\ e_i \\ \vdots \\ 0 \end{pmatrix}$$

$$= \sum_{\ell=1}^L \sum_{i \in \mathcal{I}_{W_\ell}^*} u_\ell^k[i] \begin{pmatrix} N_\ell[i] \\ \vdots \\ M_\ell[i] \\ \vdots \\ 0 \end{pmatrix} + \sum_{\ell=1}^L \sum_{i \in \mathcal{I}_{Y_\ell}^*} v_\ell^k[i] \begin{pmatrix} 0 \\ \vdots \\ e_i \\ \vdots \\ 0 \end{pmatrix}$$

$$- \sum_{i \in \mathcal{I}_g^*} \lambda^k[i] \begin{pmatrix} \nabla g_i(x^{(k)}) \\ 0 \\ \vdots \\ 0 \end{pmatrix} - \sum_{i=1}^{s_2} \nu^k[i] \begin{pmatrix} \nabla h_i(x^{(k)}) \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{53}$$

We next prove that, for each ℓ,

$$i \notin \mathcal{I}_{W_\ell}^* \quad \Rightarrow \quad \lim_{k \to \infty} u_\ell^k[i] = 0, \tag{54}$$

$$i \notin \mathcal{I}_{Y_\ell}^* \quad \Rightarrow \quad \lim_{k \to \infty} v_\ell^k[i] = 0. \tag{55}$$

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

$$\left\|(E_{\hat{\ell}}^k)^T \nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\right\| = \max_{1 \le \ell \le L} \left\|(E_\ell^k)^T \nabla_y f(x^{(k)}, y_\ell^{(k)})\right\|, \quad \forall k \tag{56}$$

and

$$\lim_{k \to \infty} \left\|(E_{\hat{\ell}}^k)^T \nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\right\| = +\infty. \tag{57}$$

Note that, from (49) and (56), we have

$$|u_\ell^k[i]| \le \frac{p_\ell y_\ell^{(k)}[i]}{y_\ell^{(k)}[i] + w_\ell^{(k)}[i]} \|(E_{\hat{\ell}}^k)^T \nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|$$

for each ℓ, i, and k. In consequence, for each ℓ, we have

$$i \notin \mathcal{I}_{W_\ell}^* \quad \Rightarrow \quad \frac{y_\ell^{(k)}[i]}{y_\ell^{(k)}[i] + w_\ell^{(k)}[i]} \to 0 \quad \Rightarrow \quad \frac{|u_\ell^k[i]|}{\|(E_{\hat{\ell}}^k)^T \nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|} \to 0. \tag{58}$$

Similarly, we can show that, for all $\ell = 1, \cdots, L$,

$$i \notin \mathcal{I}_{Y_\ell}^* \quad \Rightarrow \quad \frac{|v_\ell^k[i]|}{\|(E_{\hat{\ell}}^k)^T \nabla_y f(x^{(k)}, y_{\hat{\ell}}^{(k)})\|} \to 0. \tag{59}$$

Let d^k denote the vector on the right-hand side of equality (53). It then follows from (53) and (57)–(59) that

$$\lim_{k\to\infty} \frac{d^k}{\|(E^k_{\hat\ell})^T \nabla_y f(x^{(k)}, y^{(k)}_{\hat\ell})\|} = 0. \tag{60}$$

Since the MPEC-LICQ holds at $(x^*, \mathbf{y}^*)$, the vectors on the right-hand side of (53) are linearly independent when k is sufficiently large and so, by (60), all the sequences generated by dividing the multipliers that appear on the right-hand side of (53) by the number $\|(E^k_{\hat\ell})^T \nabla_y f(x^{(k)}, y^{(k)}_{\hat\ell})\|$ are convergent to 0 as $k \to \infty$. This fact, together with (58) and (59), implies that, for any i,

$$\lim_{k\to\infty} \frac{u^k_{\hat\ell}[i]}{\|(E^k_{\hat\ell})^T \nabla_y f(x^{(k)}, y^{(k)}_{\hat\ell})\|} = 0, \tag{61}$$

$$\lim_{k\to\infty} \frac{v^k_{\hat\ell}[i]}{\|(E^k_{\hat\ell})^T \nabla_y f(x^{(k)}, y^{(k)}_{\hat\ell})\|} = 0. \tag{62}$$

However, noticing that

$$\frac{u^k_{\hat\ell}[i] + v^k_{\hat\ell}[i]}{\|(E^k_{\hat\ell})^T \nabla_y f(x^{(k)}, y^{(k)}_{\hat\ell})\|} = p_{\hat\ell}\Big(1 + \frac{\mu_k y^{(k)}_{\hat\ell}[i]}{y^{(k)}_{\hat\ell}[i] + w^{(k)}_{\hat\ell}[i]}\Big) \frac{(E^k_{\hat\ell})^T \nabla_y f(x^{(k)}, y^{(k)}_{\hat\ell})[i]}{\|(E^k_{\hat\ell})^T \nabla_y f(x^{(k)}, y^{(k)}_{\hat\ell})\|}$$

holds for any i and k, there exists an index $\hat\imath$ such that

$$\lim_{k\to\infty} \frac{|u^k_{\hat\ell}[\hat\imath] + v^k_{\hat\ell}[\hat\imath]|}{\|(E^k_{\hat\ell})^T \nabla_y f(x^{(k)}, y^{(k)}_{\hat\ell})\|} \ge \frac{1}{\sqrt{m}} \lim_{k\to\infty} p_{\hat\ell}\Big(1 + \frac{\mu_k y^{(k)}_{\hat\ell}[\hat\imath]}{y^{(k)}_{\hat\ell}[\hat\imath] + w^{(k)}_{\hat\ell}[\hat\imath]}\Big) = \frac{p_{\hat\ell}}{\sqrt{m}} > 0.$$

This contradicts (61) and (62). As a result, the implications (54) and (55) are true.

Consider equality (53) again. By (54) and (55), the left-hand side of (53) is convergent as $k \to \infty$. Recall that the vectors on the right-hand side of (53) 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 (53) are convergent. In consequence, by letting $k \to \infty$ in (53), we obtain the equality corresponding to (36).

In addition, since both $y^{(k)}_\ell[i]$ and $w^{(k)}_\ell[i]$ are positive, we have from (51) and (49) that

$$u^k_\ell[i] v^k_\ell[i] \ge 0, \quad i = 1, \cdots, m.$$

This together with (41) yields (37). Therefore, $(x^*, \mathbf{y}^*)$ is a C-stationary point of problem (9). 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. □

3. A smoothing implicit programming method for here-and-now problems. In this section, we consider the following discrete here-and-now problem:

$$\begin{array}{ll}
\text{minimize} & \displaystyle\sum_{\ell=1}^{L} p_\ell\Big(f(x, y, \omega_\ell) + d^T z_\ell\Big) \\
\text{subject to} & g(x) \le 0,\ h(x) = 0, \\
& y \ge 0,\ N_\ell x + M_\ell y + q_\ell + z_\ell \ge 0, \\
& y^T(N_\ell x + M_\ell y + q_\ell + z_\ell) = 0, \\
& z_\ell \ge 0,\ \ell = 1, \cdots, L,
\end{array} \tag{63}$$

where d is a constant vector with positive elements. As mentioned in Section 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.

Analogous to the previous section, we let the bold letter $\mathbf{z}$ stand for the vector $(z_1^T, \cdots, z_L^T)^T$ for simplicity and, in addition, we let Λ denote the matrix $(I, \cdots, I) \in \Re^{m \times mL}$ in the following. It is easy to see that problem (63) can be rewritten as

$$\begin{array}{lll} & \text{minimize} & \sum_{\ell=1}^{L} p_\ell \Big(f(x, y, \omega_\ell) + d^T z_\ell \Big) \\ & \text{subject to} & g(x) \le 0, \;\; h(x) = 0, \;\; \mathbf{z} \ge 0, \\ (64) & & N_\ell x + M_\ell y + q_\ell + z_\ell \ge 0, \;\; \ell = 1, \cdots, L, \\ & & y \ge 0, \;\; Nx + My + q + \Lambda \mathbf{z} \ge 0, \\ & & y^T (Nx + My + q + \Lambda \mathbf{z}) = 0, \end{array}$$

where $N = \sum_{\ell=1}^{L} N_\ell$, $M = \sum_{\ell=1}^{L} M_\ell$, and $q = \sum_{\ell=1}^{L} q_\ell$. Problem (64) is actually an ordinary MPEC and hence we may solve it using some methods designed for solving MPECs, see Fukushima et al. (1998, 1999), Jiang and Ralph (2000), Lin and Fukushima (2003), etc. In this section, we propose a smoothing implicit programming method akin to SIP-I for solving problem (64) with the help of a penalty technique. Note that SIP-I cannot be applied to (64) directly because of the constraints

$$N_\ell x + M_\ell y + q_\ell + z_\ell \ge 0, \qquad \ell = 1, \cdots, L.$$

Suppose that M is a P_0-matrix. For each $(x, \mathbf{z})$ and $\mu_k > 0$, let $y(x, \Lambda \mathbf{z}, \mu_k)$ and $w(x, \Lambda \mathbf{z}, \mu_k)$ solve

$$\Phi_{\mu_k} \Big(x, y(x, \Lambda \mathbf{z}, \mu_k), w(x, \Lambda \mathbf{z}, \mu_k); N, M, q + \Lambda \mathbf{z} \Big) = 0. \tag{65}$$

The existence and differentiability of the above implicit functions follow from Theorem 1. Note that the implicit functions are denoted by $y(x, \Lambda \mathbf{z}, \mu_k)$ and $w(x, \Lambda \mathbf{z}, \mu_k)$, rather than $y(x, \mathbf{z}, \mu_k)$ and $w(x, \mathbf{z}, \mu_k)$, respectively. As in Subsection 2.1, we may consider the smooth optimization problem

$$\begin{array}{lll} & \text{minimize} & \sum_{\ell=1}^{L} p_\ell \Big(f(x, y(x, \Lambda \mathbf{z}, \mu_k), \omega_\ell) + d^T z_\ell \Big) \\ (66) & \text{subject to} & g(x) \le 0, \;\; h(x) = 0, \;\; \mathbf{z} \ge 0, \\ & & N_\ell x + M_\ell y(x, \Lambda \mathbf{z}, \mu_k) + q_\ell + z_\ell \ge 0, \;\; \ell = 1, \cdots, L, \end{array}$$

as an approximation of problem (64). Since the feasible region of problem (66) is dependent on μ_k, (66) may not be easy to solve. Therefore, we apply a penalty technique to this problem and obtain the following approximation of problem (64):

$$\begin{array}{lll} (67) & \text{minimize} & \theta_k(x, \mathbf{z}) \\ & \text{subject to} & g(x) \le 0, \;\; h(x) = 0, \;\; \mathbf{z} \ge 0, \end{array}$$

where

$$\theta_k(x, \mathbf{z}) = \sum_{\ell=1}^{L} p_\ell \Big(f(x, y(x, \Lambda \mathbf{z}, \mu_k), \omega_\ell) + d^T z_\ell \Big) + \rho_k \sum_{\ell=1}^{L} \psi \Big(-(N_\ell x + M_\ell y(x, \Lambda \mathbf{z}, \mu_k) + q_\ell + z_\ell) \Big),$$

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 (67) is common for all k.

Now we present our method, called SIP-II, for problem (63): Choose two sequences $\{\mu_k\}$ and $\{\rho_k\}$ of positive numbers. We then solve problems (67) to get a sequence $\{(x^{(k)}, \mathbf{z}^{(k)})\}$ and, for each k, let $y^{(k)} = y(x^{(k)}, \Lambda\mathbf{z}^{(k)}, \mu_k)$.

In what follows, we let $\mathcal{F}_2$ and $\mathcal{Z}$ stand for the feasible regions of problems (64) and (67), respectively. Also, 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

$$\lim_{k\to\infty} \mu_k = 0, \quad \lim_{k\to\infty} \rho_k = +\infty, \quad \lim_{k\to\infty} \mu_k\rho_k = 0. \tag{68}$$

The following lemma is helpful in establishing convergence theory for SIP-II. We omit its proof because it is similar to that of Lemma 2.

Lemma 3 *Suppose that M is a P_0-matrix in problem (63) and, for any bounded sequence $\{(x^k, \mathbf{z}^k)\}$ in $\mathcal{Z}$, $\{y(x^k, \Lambda\mathbf{z}^k, \mu_k)\}$ is bounded. If $(x^*, y^*, \mathbf{z}^*) \in \mathcal{F}_2$ and the submatrix $M[\mathcal{K}^*]$ is nondegenerate, where $\mathcal{K}^* = \{\ i \mid\ (Nx^* + My^* + q + \Lambda\mathbf{z}^*)[i] = 0\}$, there exist a neighborhood U^* of $(x^*, y^*, \mathbf{z}^*)$ and a positive constant π^* such that*

$$\|y(x, \Lambda\mathbf{z}, \mu_k) - y\| \le \mu_k\pi^*(\|y\| + \sqrt{m}) \tag{69}$$

for any $(x, y, \mathbf{z}) \in U^* \cap \mathcal{F}_2$ 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 5 *Let M be a P_0-matrix and $\psi : \Re^m \to [0, +\infty)$ be a continuously differentiable function satisfying $\psi(0) = 0$ and*

$$\psi(y) \le \psi(y'), \quad \forall y' \ge y \in \Re^m, \tag{70}$$

and, for any bounded sequence $\{(x^k, \mathbf{z}^k)\}$ in $\mathcal{Z}$, $\{y(x^k, \Lambda\mathbf{z}^k, \mu_k)\}$ be bounded. Suppose that the sequence $\{(x^{(k)}, y^{(k)}, \mathbf{z}^{(k)})\}$ generated by SIP-II with $(x^{(k)}, \mathbf{z}^{(k)})$ being a local optimal solution of (67) is convergent to $(x^*, y^*, \mathbf{z}^*) \in \mathcal{F}_2$. If there exists a neighborhood V^* of $(x^*, y^*, \mathbf{z}^*)$ such that $(x^{(k)}, \mathbf{z}^{(k)})$ 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 3, then $(x^*, y^*, \mathbf{z}^*)$ is a local optimal solution of (63).

Proof: By Lemma 3, there exist a closed sphere $\mathcal{B}_2 \subseteq V^*$ centered at the point $(x^*, y^*, \mathbf{z}^*)$ with positive radius and a positive number π^* such that (69) holds for any $(x, y, \mathbf{z}) \in \mathcal{F}_2 \cap \mathcal{B}_2$ and every k. Since $\mathcal{F}_2 \cap \mathcal{B}_2$ is a nonempty compact set, the problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{\ell=1}^{L} p_\ell\Big(f(x, y, \omega_\ell) + d^T z_\ell\Big) \\ \text{subject to} & (x, y, \mathbf{z}) \in \mathcal{F}_2 \cap \mathcal{B}_2 \end{array} \tag{71}$$

has an optimal solution, say $(\bar{x}, \bar{y}, \bar{\mathbf{z}})$.

Consider an arbitrary $(x, y, \mathbf{z}) \in \mathcal{F}_2 \cap \mathcal{B}_2$. We then have from the mean-value theorem that

$$\begin{aligned} \theta_k(x, \mathbf{z}) \;=\; & \sum_{\ell=1}^{L} p_\ell\Big(f(x, y, \omega_\ell) + d^T z_\ell \\ & +(y(x, \Lambda\mathbf{z}, \mu_k) - y)^T \nabla_y f(x, (1-t)y(x, \Lambda\mathbf{z}, \mu_k) + ty, \omega_\ell)\Big) \\ & +\rho_k \sum_{\ell=1}^{L} \psi\Big(-(N_\ell x + M_\ell y(x, \Lambda\mathbf{z}, \mu_k) + q_\ell + z_\ell)\Big), \end{aligned} \tag{72}$$

where $t \in [0,1]$. Note that, by (69),

$$\begin{aligned}\|(1-t)y(x,\Lambda\mathbf{z},\mu_k)+ty\| &= \|(1-t)(y(x,\Lambda\mathbf{z},\mu_k)-y)+y\|\\ &\le \|y(x,\Lambda\mathbf{z},\mu_k)-y\|+\|y\|\\ &\le \mu_k\pi^*(\|y\|+\sqrt{m})+\|y\|.\end{aligned}$$

This indicates that the set

$$\{(x,(1-t)y(x,\Lambda\mathbf{z},\mu_k)+ty)\;\big|\;(x,y,\mathbf{z})\in\mathcal{F}_2\cap\mathcal{B}_2, t\in[0,1], k=1,2,\cdots\}$$

is bounded. Similarly, we see that

$$\{(x,tM_\ell(y-y(x,\Lambda\mathbf{z},\mu_k)))\;\big|\;(x,y,\mathbf{z})\in\mathcal{F}_2\cap\mathcal{B}_2, \ell=1,\cdots,L, t\in[0,1], k=1,2,\cdots\}$$

is also bounded. Then, by the continuous differentiability of both f and ψ, there exists a constant $\tau>0$ such that, for $\ell=1,\cdots,L$,

$$\|\nabla_y f(x,(1-t)y(x,\Lambda\mathbf{z},\mu_k)+ty,\omega_\ell)\| \le \tau, \tag{73}$$

$$\|\nabla\psi\Big(tM_\ell(y-y(x,\Lambda\mathbf{z},\mu_k))\Big)\| \le \tau \tag{74}$$

hold for any $(x,y,\mathbf{z})\in\mathcal{F}_2\cap\mathcal{B}_2$, $t\in[0,1]$, and every k. Noticing that $(x,y,\mathbf{z})\in\mathcal{F}_2\cap\mathcal{B}_2$ implies $N_\ell x+M_\ell y+q_\ell+z_\ell\ge 0$ for each ℓ, we have from (70) and (74) that

$$\begin{aligned}\psi\Big(-(N_\ell x+M_\ell y(x,\Lambda\mathbf{z},\mu_k)+q_\ell+z_\ell)\Big) &\le \psi\Big(M_\ell(y-y(x,\Lambda\mathbf{z},\mu_k))\Big)\\ &= \psi\Big(M_\ell(y-y(x,\Lambda\mathbf{z},\mu_k))\Big)-\psi(0)\\ &= \nabla\psi\Big(t'M_\ell(y-y(x,\Lambda\mathbf{z},\mu_k))\Big)^T M_\ell\Big(y-y(x,\Lambda\mathbf{z},\mu_k)\Big)\\ &\le \tau\|M_\ell\|\,\|y-y(x,\Lambda\mathbf{z},\mu_k)\|,\end{aligned}$$

where $t'\in[0,1]$ and the second equality follows from the mean-value theorem. This, together with (72)–(73) and (69), yields

$$\begin{aligned}\Big|\theta_k(x,\mathbf{z})-\sum_{\ell=1}^L p_\ell\Big(f(x,y,\omega_\ell)+d^Tz_\ell\Big)\Big| &\le \tau\|y(x,\Lambda\mathbf{z},\mu_k)-y\|+\Big(\tau\rho_k\sum_{\ell=1}^L\|M_\ell\|\Big)\|y-y(x,\Lambda\mathbf{z},\mu_k)\|\\ &\le \pi^*\tau\Big(\mu_k+\mu_k\rho_k\sum_{\ell=1}^L\|M_\ell\|\Big)(\|y\|+\sqrt{m})\end{aligned}$$

for any $(x,y,\mathbf{z})\in\mathcal{F}_2\cap\mathcal{B}_2$ and k. In particular,

$$\Big|\theta_k(\bar{x},\bar{\mathbf{z}})-\sum_{\ell=1}^L p_\ell\Big(f(\bar{x},\bar{y},\omega_\ell)+d^T\bar{z}_\ell\Big)\Big|\le\pi^*\tau\Big(\mu_k+\mu_k\rho_k\sum_{\ell=1}^L\|M_\ell\|\Big)(\|\bar{y}\|+\sqrt{m}). \tag{75}$$

Moreover, since ψ is always nonnegative, we have from the continuity of f that

$$\begin{aligned}\lim_{k\to\infty}\theta_k(x^{(k)},\mathbf{z}^{(k)}) &\ge \sum_{\ell=1}^L p_\ell\lim_{k\to\infty}\Big(f(x^{(k)},y^{(k)},\omega_\ell)+d^Tz_\ell^{(k)}\Big)\\ &= \sum_{\ell=1}^L p_\ell\Big(f(x^*,y^*,\omega_\ell)+d^Tz_\ell^*\Big).\end{aligned} \tag{76}$$

Note that, by the fact that $\mathcal{F}_2 \cap \mathcal{B}_2 \subseteq V^*$, $(x^{(k)}, \mathbf{z}^{(k)})$ is an optimal solution of the problem

$$\begin{array}{ll} \text{minimize} & \theta_k(x, \mathbf{z}) \\ \text{subject to} & (x, \mathbf{z}) \in \mathcal{Z}_2 := \{\, (x, \mathbf{z}) \in \mathcal{Z} \mid \ (x, y, \mathbf{z}) \in \mathcal{F}_2 \cap \mathcal{B}_2 \,\}, \end{array}$$

provided k is large enough, and $(\bar{x}, \bar{\mathbf{z}})$ is a feasible point of this problem. We then have from (75) that, for every k sufficiently large,

$$\begin{aligned} \theta_k(x^{(k)}, \mathbf{z}^{(k)}) &\leq \theta_k(\bar{x}, \bar{\mathbf{z}}) \\ &\leq \sum_{\ell=1}^{L} p_\ell\Big(f(\bar{x}, \bar{y}, \omega_\ell) + d^T \bar{z}_\ell\Big) + \pi^* \tau\Big(\mu_k + \mu_k \rho_k \sum_{\ell=1}^{L} \|M_\ell\|\Big)(\|\bar{y}\| + \sqrt{m}). \end{aligned} \tag{77}$$

Therefore, taking into account the equality (76) and Assumption A5, we have by letting $k \to \infty$ in (77) that

$$\sum_{\ell=1}^{L} p_\ell\Big(f(x^*, y^*, \omega_\ell) + d^T z_\ell^*\Big) \leq \sum_{\ell=1}^{L} p_\ell\Big(f(\bar{x}, \bar{y}, \omega_\ell) + d^T \bar{z}_\ell\Big),$$

while the converse inequality follows from the fact that $(\bar{x}, \bar{y}, \bar{\mathbf{z}})$ is an optimal solution of problem (71) immediately. As a result, we have

$$\sum_{\ell=1}^{L} p_\ell\Big(f(x^*, y^*, \omega_\ell) + d^T z_\ell^*\Big) = \sum_{\ell=1}^{L} p_\ell\Big(f(\bar{x}, \bar{y}, \omega_\ell) + d^T \bar{z}_\ell\Big),$$

namely, $(x^*, y^*, \mathbf{z}^*)$ is an optimal solution of problem (71) and hence it is a local optimal solution of problem (63). This completes the proof. □

It is not difficult to see that the function $\psi(y) = \sum_{i=1}^{m} (\max(y[i], 0))^\sigma$, where $\sigma \geq 2$ is a positive integer, satisfies the conditions assumed in Theorem 5. This function is often employed for solving constrained optimization problems. For more details, see Bertsekas (1982).

THEOREM 6 *Let M be a P-matrix, the function ψ be the same as in Theorem 5 and, for any bounded sequence $\{(x^k, \mathbf{z}^k)\}$ in $\mathcal{Z}$, $\{y(x^k, \Lambda \mathbf{z}^k, \mu_k)\}$ be bounded. Assume that $(x^{(k)}, \mathbf{z}^{(k)})$ is a global optimal solution of (67) for each k. Then any accumulation point $(x^*, y^*, \mathbf{z}^*) \in \mathcal{F}_2$ of the sequence $\{(x^{(k)}, y^{(k)}, \mathbf{z}^{(k)})\}$ generated by SIP-II is a global optimal solution of problem (63).*

We omit the proof of the above theorem since it is similar to that of Theorem 3. Next, we discuss the limiting behavior of stationary points of problems (67).

THEOREM 7 *Suppose that M is a P_0-matrix in problem (63), the function $\psi : \Re^m \to [0, +\infty)$ is a continuously differentiable function such that $\psi(0) = 0$, and $(x^{(k)}, \mathbf{z}^{(k)})$ is a stationary point of (67) for each k. Let $(x^*, y^*, \mathbf{z}^*) \in \mathcal{F}_2$ be an accumulation point of the sequence $\{(x^{(k)}, y^{(k)}, \mathbf{z}^{(k)})\}$ generated by SIP-II. If the MPEC-LICQ is satisfied at $(x^*, y^*, \mathbf{z}^*)$, then $(x^*, y^*, \mathbf{z}^*)$ is a C-stationary point of (63). In particular, if y^* satisfies the strict complementarity condition, then $(x^*, y^*, \mathbf{z}^*)$ is B-stationary.*

A proof of this result is given in the appendix. Note that, in Theorem 7, the conditions on the penalty function are quite weak. We can easily find such a penalty function (Bertsekas 1982).

In addition, similar discussion is valid for the following generalized problem:

$$
\begin{array}{lll}
 & \text{minimize} & \displaystyle\sum_{\ell=1}^{L} p_\ell\Big(f(x,y,\omega_\ell) + d_\ell^T z_\ell\Big) \\
 & \text{subject to} & g(x) \le 0,\ h(x) = 0, \\
(78) & & y \ge 0,\ N_\ell x + M_\ell y + q_\ell + A_\ell z_\ell \ge 0, \\
 & & y^T(N_\ell x + M_\ell y + q_\ell + A_\ell z_\ell) = 0, \\
 & & z_\ell \ge 0,\ \ell = 1, \cdots, L,
\end{array}
$$

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.

4. Conclusions. In this paper, 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.

References

Bertsekas, D. P. 1982. *Constrained optimization and Lagrange multiplier methods.* Academic Press, New York.

Birge, J. R., F. Louveaux. 1997. *Introduction to stochastic programming.* Springer, New York.

Chen, X. 2000. Newton-type methods for stochastic programming. *Mathematical and Computer Modelling* **31** 89–98.

Chen, X., M. Fukushima. 2003. A smoothing method for a mathematical program with P-matrix linear complementarity constraints. *Computational Optimization and Applications,* to appear.

Chen, X., L. Qi, R. S. Womersley. 1995. Newton's method for quadratic stochastic programs with recourse. *Journal of Computational and Applied Mathematics* **60** 29–46.

Chen, X., R. S. Womersley. 1996. A parallel Newton method for quadratic stochastic programs with recourse. *Annals of Operations Research* **64** 113–141.

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

Clarke, F. H. 1990. *Optimization and nonsmooth analysis.* SIAM, Philadelphia.

Cottle, R. W., J. S. Pang, R. E. Stone. 1992. *The linear complementarity problem.* Academic Press, New York.

Fischer, A. 1992. A special Newton-type optimization method. *Optimization* **24** 269–284.

Fukushima, M., Z. Q. Luo, J. S. Pang. 1998. A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints. *Computational Optimization and Applications* **10** 5–34.

Fukushima, M., J. S. Pang. 1999. 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* **477** 105–116. M. Théra and R. Tichatschke (eds.), Springer-Verlag, Berlin/Heidelberg.

Gabriel , S. A., J. J. Moré. 1997. Smoothing of mixed complementarity problems. *Complementarity and Variational Problems: State of the Art* 105–116. M.C. Ferris and J.S. Pang (eds.), SIAM, Philadelphia, Pennsylvania.

Jiang, H., D. Ralph. 2000. Smooth SQP methods for mathematical programs with nonlinear complementarity constraints. *SIAM Journal on Optimization* **10** 779–808.

Kall, P., S.W. Wallace. 1994. *Stochastic programming.* John Wiley & Sons, Chichester.

Kanzow, C. 1996. Some noninterior continuation methods for linear complementarity problems. *SIAM Journal on Matrix Analysis and Applications* **17** 851–868.

Lin, G. H., M. Fukushima. A modified relaxation scheme for mathematical programs with complementarity constraints. *Annals of Operations Research*, to appear.

Lin, G. H., M. Fukushima. 2003. A new relaxation method for mathematical programs with complementarity constraints. *Journal of Optimization Theory and Applications* **118** 81–116.

Lin, G. H., M. Fukushima. 2002. 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.

Luo, Z. Q., J. S. Pang, D. Ralph. 1996. *Mathematical programs with equilibrium constraints*. Cambridge University Press, Cambridge, United Kingdom

Ortega, J. M., W.C. Rheinboldt. 1970. *Iterative solution of nonlinear equations in several variables*. Academic Press.

Pang, J. S. 1994. Complementarity problems. *Handbook on Global Optimization* 271–338. R. Horst and P. Pardalos (eds.), Kluwer Academic Publishers, B.V. Dordrecht.

Pang, J. S., M. Fukushima. 1999. Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints. *Computational Optimization and Applications* **13** 111–136.

Patriksson, M., L. Wynter. 1999. Stochastic mathematical programs with equilibrium constraints. *Operations Research Letters* **25** 159–167.

Scheel, H., S. Scholtes. 2000. Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. *Mathematics of Operations Research* **25** 1–22.

Scholtes, S. 2001. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization* **11** 918–936.

Schultz, R., L. Stougie, M. H. van der Vlerk. 1996. Two-stage stochastic integer programming: a survey. *Statistica Neerlandica* **50** 404–416.

Shimizu, K., Y. Ishizuka, J. Bard. 1997. *Nondifferentiable and two-level mathematical programming.* Kluwer Academic Publishers, Boston.

Vajda, S. 1972. *Probabilistic programming.* Academic Press, New York.

Appendix: proof of Theorem 7. Assume without loss of generality that the sequence $\{(x^{(k)}, y^{(k)}, \mathbf{z}^{(k)})\}$ converges to $(x^*, y^*, \mathbf{z}^*)$. Since the MPEC-LICQ holds at $(x^*, y^*, \mathbf{z}^*)$, problem (67) satisfies the standard LICQ at $(x^{(k)}, \mathbf{z}^{(k)})$ for all k sufficiently large and so, by the stationarity of $(x^{(k)}, \mathbf{z}^{(k)})$, there exist unique Lagrange multiplier vectors λ^k, ν^k, and $\gamma^k = \left((\gamma_1^k)^T, \cdots, (\gamma_L^k)^T\right)^T$ such that

$$\nabla\theta_k(x^{(k)}, \mathbf{z}^{(k)}) + \begin{pmatrix} \nabla g(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \lambda^k + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \nu^k - \begin{pmatrix} O & \cdots & O \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix} \gamma^k = 0, \tag{A.1}$$

$$h(x^{(k)}) = 0, \quad 0 \le \lambda^k \perp g(x^{(k)}) \ge 0, \qquad 0 \le \gamma_\ell^k \perp z_\ell^{(k)} \ge 0, \quad \ell = 1, \cdots, L. \tag{A.2}$$

In the rest of the proof, we suppose k is large enough so that (A.1)–(A.2) and (42) are satisfied and furthermore, for each $\ell = 1, \cdots, L$, there hold

$$\begin{aligned} \mathcal{I}_{W_\ell}^k &:= \{i \mid \left(N_\ell x^{(k)} + M_\ell y^{(k)} + q_\ell + z_\ell^{(k)}\right)[i] = 0\} \\ &\subseteq \ \mathcal{I}_{W_\ell}^* := \{i \mid (N_\ell x^* + M_\ell y^* + q_\ell + z_\ell^*)[i] = 0\}, \end{aligned} \tag{A.3}$$

$$\begin{aligned} \mathcal{I}_{W}^k &:= \{i \mid \left(N x^{(k)} + M y^{(k)} + q + \Lambda\mathbf{z}^{(k)}\right)[i] = 0\} \\ &\subseteq \ \mathcal{I}_{W}^* := \{i \mid (N x^* + M y^* + q + \Lambda\mathbf{z}^*)[i] = 0\}, \end{aligned} \tag{A.4}$$

$$\mathcal{I}_Y^k := \{i \mid y^{(k)}[i] = 0\} \subseteq \mathcal{I}_Y^* := \{i \mid y^*[i] = 0\}, \tag{A.5}$$

$$\mathcal{I}_{Z_\ell}^k := \{i \mid z_\ell^{(k)}[i] = 0\} \subseteq \mathcal{I}_{Z_\ell}^* := \{i \mid z_\ell^*[i] = 0\}. \tag{A.6}$$

Consider the implicit functions $y(\cdot,\cdot,\cdot)$ and $w(\cdot,\cdot,\cdot)$ defined by (65). For these functions, let ∇_x, ∇_z, and $\nabla_{(x,z)}$ denote the derivatives with respect to the first argument, the second argument, and the first and the second arguments, respectively. Analogous to (43), we have from the implicit function theorem that

$$\begin{pmatrix} \nabla_{(x,z)} y(x^{(k)}, \Lambda\mathbf{z}^{(k)}, \mu_k)^T \\ \nabla_{(x,z)} w(x^{(k)}, \Lambda\mathbf{z}^{(k)}, \mu_k)^T \end{pmatrix} = - \begin{pmatrix} M+\mu_k I & -I \\ I - D^k & D^k \end{pmatrix}^{-1} \begin{pmatrix} N & I \\ O & O \end{pmatrix}, \tag{A.7}$$

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 (A.7) that

$$\nabla_x y(x^{(k)}, \Lambda\mathbf{z}^{(k)}, \mu_k) = -N^T D^k (E^k)^T, \tag{A.8}$$

$$\nabla_z y(x^{(k)}, \Lambda\mathbf{z}^{(k)}, \mu_k) = -D^k (E^k)^T. \tag{A.9}$$

Thus, from the definition of θ_k, (A.8)–(A.9), and by a straightforward calculus, (A.1) becomes

$$\begin{aligned} 0 \;=\; & \begin{pmatrix} \sum_{\ell=1}^L p_\ell (\nabla_x f(x^{(k)}, y^{(k)}, \omega_\ell) - N^T D^k (E^k)^T \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell)) \\ p_1 d - \sum_{\ell=1}^L p_\ell D^k (E^k)^T \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) \\ \vdots \\ p_L d - \sum_{\ell=1}^L p_\ell D^k (E^k)^T \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) \end{pmatrix} \\ & - \begin{pmatrix} N_1^T - N^T D^k (E^k)^T M_1^T & \cdots & N_L^T - N^T D^k (E^k)^T M_L^T \\ I - D^k (E^k)^T M_1^T & \cdots & -D^k (E^k)^T M_L^T \\ -D^k (E^k)^T M_1^T & \cdots & -D^k (E^k)^T M_L^T \\ \vdots & \vdots & \vdots \\ -D^k (E^k)^T M_1^T & \cdots & I - D^k (E^k)^T M_L^T \end{pmatrix} \beta^k \\ & + \begin{pmatrix} \nabla g(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \lambda^k + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \nu^k - \begin{pmatrix} O & \cdots & O \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix} \gamma^k. \end{aligned}$$

Here, $\beta^k \in \Re^{n+mL}$ is given by $\beta^k = \left((\beta_1^k)^T, \cdots, (\beta_L^k)^T \right)^T$ with

$$\beta_\ell^k = \rho_k \nabla\psi(-(N_\ell x^{(k)} + M_\ell y(x^{(k)}, \Lambda\mathbf{z}^{(k)}, \mu_k) + q_\ell + z_\ell^{(k)}))$$

for each ℓ. We then have

$$0 = \begin{pmatrix} \sum_{\ell=1}^L p_\ell \nabla_x f(x^{(k)}, y^{(k)}, \omega_\ell) \\ p_1 d \\ \vdots \\ p_L d \end{pmatrix} + \begin{pmatrix} \nabla g(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \lambda^k + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ \vdots \\ O \end{pmatrix} \nu^k$$

$$
-\begin{pmatrix} N_1^T & \cdots & N_L^T \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix}\beta^k - \begin{pmatrix} O & \cdots & O \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix}\gamma^k - \begin{pmatrix} N^T \\ I \\ \vdots \\ I \end{pmatrix}u^k, \tag{A.10}
$$

where u^k is defined by

$$
u^k = D^k(E^k)^T \sum_{\ell=1}^{L} \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k\Big). \tag{A.11}
$$

Furthermore, by letting

$$
\sum_{\ell=1}^{L} p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - \sum_{\ell=1}^{L} M_\ell^T \beta_\ell^k - M^T u^k - v^k = 0, \tag{A.12}
$$

we have

$$
\begin{aligned}
v^k &= \sum_{\ell=1}^{L} \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k\Big) - M^T u^k \\
&= \Big((E^k)^{-T} - M^T D^k\Big)(E^k)^T \sum_{\ell=1}^{L} \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k\Big) \\
&= \Big(I - (1-\mu_k)D^k\Big)(E^k)^T \sum_{\ell=1}^{L} \Big(p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) - M_\ell^T \beta_\ell^k\Big),
\end{aligned} \tag{A.13}
$$

where the second equality follows from (A.11) and the third equality follows from $E^k = \Big(D^k M + I - (1-\mu_k)D^k\Big)^{-1}$ by definition. Combining (A.10) and (A.12) yields

$$
\begin{aligned}
0 \; = \; & \begin{pmatrix} \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y^{(k)}, \omega_\ell) \\ \sum_{\ell=1}^{L} p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) \\ p_1 d \\ \vdots \\ p_L d \end{pmatrix} + \begin{pmatrix} \nabla g(x^{(k)}) \\ O \\ O \\ \vdots \\ O \end{pmatrix}\lambda^k + \begin{pmatrix} \nabla h(x^{(k)}) \\ O \\ O \\ \vdots \\ O \end{pmatrix}\nu^k \\
& -\begin{pmatrix} N_1^T & \cdots & N_L^T \\ M_1^T & \cdots & M_L^T \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix}\beta^k - \begin{pmatrix} O & \cdots & O \\ O & \cdots & O \\ I & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & I \end{pmatrix}\gamma^k - \begin{pmatrix} N^T \\ M^T \\ I \\ \vdots \\ I \end{pmatrix}u^k - \begin{pmatrix} O \\ I \\ O \\ \vdots \\ O \end{pmatrix}v^k.
\end{aligned} \tag{A.14}
$$

Taking into account (42), and (A.2)–(A.6), we can rewrite (A.14) as

$$
\begin{pmatrix} \sum_{\ell=1}^{L} p_\ell \nabla_x f(x^{(k)}, y^{(k)}, \omega_\ell) \\ \sum_{\ell=1}^{L} p_\ell \nabla_y f(x^{(k)}, y^{(k)}, \omega_\ell) \\ p_1 d \\ \vdots \\ p_L d \end{pmatrix} - \sum_{i \notin \mathcal{I}_W^*} u^k[i] \begin{pmatrix} N[i] \\ M[i] \\ e_i \\ \vdots \\ e_i \end{pmatrix} - \sum_{i \notin \mathcal{I}_Y^*} v^k[i] \begin{pmatrix} 0 \\ e_i \\ 0 \\ \vdots \\ 0 \end{pmatrix}
$$

$$= \; -\sum_{i\in\mathcal{I}_g^*}\lambda^k[i]\begin{pmatrix}\nabla g_i(x^{(k)})\\ 0\\ \vdots\\ 0\end{pmatrix} - \sum_{i=1}^{s_2}\nu^k[i]\begin{pmatrix}\nabla h_i(x^{(k)})\\ 0\\ \vdots\\ 0\end{pmatrix} + \sum_{\ell=1}^{L}\sum_{i\in\mathcal{I}_{W_\ell}^*}\beta_\ell^k[i]\begin{pmatrix}N_\ell[i]\\ M_\ell[i]\\ \vdots\\ e_i\\ \vdots\\ 0\end{pmatrix}$$

$$\text{(A.15)} \qquad + \sum_{\ell=1}^{L}\sum_{i\in\mathcal{I}_{Z_\ell}^*}\gamma_\ell^k[i]\begin{pmatrix}0\\ \vdots\\ e_i\\ \vdots\\ 0\end{pmatrix} + \sum_{i\in\mathcal{I}_W^*}u^k[i]\begin{pmatrix}N[i]\\ M[i]\\ e_i\\ \vdots\\ e_i\end{pmatrix} + \sum_{i\in\mathcal{I}_Y^*}v^k[i]\begin{pmatrix}0\\ e_i\\ 0\\ \vdots\\ 0\end{pmatrix}.$$

We next prove

$$\text{(A.16)} \qquad i\notin\mathcal{I}_W^* \;\Rightarrow\; \lim_{k\to\infty}u^k[i]=0,$$

$$\text{(A.17)} \qquad i\notin\mathcal{I}_Y^* \;\Rightarrow\; \lim_{k\to\infty}v^k[i]=0.$$

Note that, by (A.13) and (A.11),

$$\text{(A.18)} \qquad u^k[i] \;=\; \frac{y^{(k)}[i]}{y^{(k)}[i]+w^{(k)}[i]}\,e_i^T(E^k)^T\sum_{\ell=1}^{L}\Big(p_\ell\nabla_y f(x^{(k)},y^{(k)},\omega_\ell)-M_\ell^T\beta_\ell^k\Big),$$

$$\text{(A.19)} \qquad v^k[i] \;=\; \frac{\mu_k y^{(k)}[i]+w^{(k)}[i]}{y^{(k)}[i]+w^{(k)}[i]}\,e_i^T(E^k)^T\sum_{\ell=1}^{L}\Big(p_\ell\nabla_y f(x^{(k)},y^{(k)},\omega_\ell)-M_\ell^T\beta_\ell^k\Big).$$

In a similar way to the proof of Theorem 4, we can show that

$$\Big\{\Big\|(E^k)^T\sum_{\ell=1}^{L}\Big(p_\ell\nabla_y f(x^{(k)},y^{(k)},\omega_\ell)-M_\ell^T\beta_\ell^k\Big)\Big\|\Big\}$$

is bounded. Then, it follows from (A.18) and (A.19) that

$$i\notin\mathcal{I}_W^*\Rightarrow \quad \lim_{k\to\infty}\frac{y^{(k)}[i]}{y^{(k)}[i]+w^{(k)}[i]}=0 \quad \Rightarrow \lim_{k\to\infty}u^k[i]=0,$$

$$i\notin\mathcal{I}_Y^*\Rightarrow \quad \lim_{k\to\infty}\frac{\mu_k y^{(k)}[i]+w^{(k)}[i]}{y^{(k)}[i]+w^{(k)}[i]}=0 \quad \Rightarrow \lim_{k\to\infty}v^k[i]=0.$$

By (A.16) and (A.17), the left-hand side of equality (A.15) is convergent. Moreover, from the assumption that the MPEC-LICQ holds at $(x^*,y^*,\mathbf{z}^*)$, the vectors on the right-hand side of equality (A.15) are linearly independent for all k sufficiently large. Thus, all the sequences of the multipliers that appear on the right-hand side of (A.15) are convergent. In consequence, by letting $k\to\infty$ in (A.15), we obtain the equality corresponding to (36).

In addition, since both $y^{(k)}[i]$ and $w^{(k)}[i]$ are positive, we have from (A.13) and (A.11) that

$$u^k[i]v^k[i]\ge 0, \quad i=1,\cdots,m.$$

This together with (A.2) yields the results corresponding to (37). Therefore, $(x^*,y^*,\mathbf{z}^*)$ is a C-stationary point of problem (63). If, in addition, y^* satisfies the strict complementarity condition, then $(x^*,y^*,\mathbf{z}^*)$ is a B-stationary point by the definitions of C-stationarity and B-stationarity immediately. This completes the proof of Theorem 7. ∎