SIAM J. OPTIM.
Vol. 16, No. 3, pp. 670–696

AN IMPLICIT PROGRAMMING APPROACH FOR A CLASS OF STOCHASTIC MATHEMATICAL PROGRAMS WITH COMPLEMENTARITY CONSTRAINTS*

HUIFU XU†

Abstract. In this paper, we consider a class of stochastic mathematical programs in which the complementarity constraints are subject to random factors and the objective function is the mathematical expectation of a smooth function which depends on both upper and lower level variables and random factors. We investigate the existence, uniqueness, and differentiability of the lower level equilibrium defined by the complementarity constraints using a nonsmooth version of implicit function theorem. We also study the differentiability and convexity of the objective function which implicitly depends upon the lower level equilibrium. We propose numerical methods to deal with difficulties due to the continuous distribution of the random variables and intrinsic nonsmoothness of lower level equilibrium solutions due to the complementarity constraints in order that the treated programs can be readily solved by available numerical methods for deterministic mathematical programs with complementarity constraints.

Key words. stochastic mathematical programs with complementarity constraints, lower level equilibrium, discretization, implicit smoothing

AMS subject classifications. 90C15, 90C30, 90C31, 90C33

DOI. 10.1137/040608544

1. Introduction.

Mathematical programs with equilibrium constraints (MPEC) are a class of optimization problems with two sets of variables: upper level variables and lower level variables and an equilibrium constraint defined by a parametric variational inequality or a complementarity system with lower variables being its prime variables and upper level variables being its parameters.

Over the past few years, MPEC has developed as a new area in optimization; see [23, 25] for an overview. One of the driving forces of the rapid development is that MPEC has found useful applications in many areas such as economics, management, and engineering. A particularly interesting example for MPEC is a Stackelberg–Nash leader-follower model for competition in an oligopoly market where a number of firms compete to supply homogeneous goods into a market in a noncooperative manner [24, 32]. Suppose that a distinct strategic firm (called leader), may anticipate the reaction of the remaining nonstrategic firms (called followers) to his decision and use this knowledge to select his optimal supply by minimizing the objective function. The followers' reaction to the leader's decision can be described by a Nash equilibrium which can be mathematically formulated as a variational inequality (VI). In structural optimization, the objective is often to optimize the performance of a structure, or its construction cost or weight by selecting design parameters, such as the shape of structure, or the choice of material under the constraints of the behavior of the structure, where the values of the state variables such as displacements, stresses, and contact forces are described by an equilibrium of minimal energy. The problem can be modeled as MPEC similarly [23, 25].

*Received by the editors May 18, 2004; accepted for publication (in revised form) June 2, 2005; published electronically January 6, 2006. This research is supported by the United Kingdom Engineering and Physical Sciences Research Council grant GR/S90850/01.
http://www.siam.org/journals/siopt/16-3/60854.html

†School of Mathematics, University of Southampton, Southampton SO17 1BJ, UK (h.xu@maths.soton.ac.uk).

In MPEC models, the underlying data are deterministic. However, in some important practical instances, there may be some stochastic (uncertain) factors involved in MPEC models. For instance, in a Stackelberg leader-follower equilibrium model, the leader's decision may be subject to some uncertainty in market demand. This is particularly so when the decision is made now for future output. Ignoring such an uncertainty may result in a decision being made on the basis of a particular market realization which occurs at a very low probability. De Wolf and Smeers [4] first considered this kind of stochastic leader-follower problem and applied it to model the European gas market. Xu [36] considered a more general model and investigated it with an MPEC approach.

In mechanical optimization, a structural equilibrium may be subject to the random properties of materials and randomly varying conditions such as weather and external forces [2]. The distribution of these random factors may be obtained from experience or through observation. It might be undesirable to base the optimal choice of design parameters on the expected values of the random data.

Patriksson and Wynter [26] first considered a general class of stochastic mathematical programs with equilibrium constraints (SMPEC) as follows:

$$\text{SMPEC}\quad \begin{array}{ll} \min & E(x) := \mathbb{E}[f(x, y(x, \xi(\omega)))] \\ \text{s.t.} & x \in \mathcal{X}, \end{array}$$

where $\xi : \Omega \to \mathbb{R}^l$ denotes a vector of random variables defined on a sample space Ω, $y(x, \xi(\omega))$ denotes a measurable selection from $\mathcal{S}(x, \xi(\omega))$, the set of solutions for the lower level VI problem parameterized by the upper level variable x and random vector $\xi(\omega)$; $\mathbb{E}$ denotes the expected value. They investigated the existence of an optimal solution and the directional differentiability of the objective function.

In this paper we consider a less complicated SMPEC model as follows:

$$\text{SMPCC}\quad \begin{array}{ll} \min & \mathbb{E}\,[f(x, y, \xi(\omega))] \\ \text{s.t.} & x \in \mathcal{X}, \\ & 0 \leq y \perp F(x, y, \xi(\omega)) \geq 0, \end{array} \tag{1}$$

where, by a slight abuse of notation, $f : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ denotes a continuously differentiable function, $F : \mathbb{R}^m \times \mathbb{R}^n_+ \times \mathbb{R}^l \to \mathbb{R}^m$ denotes a continuously differentiable vector valued function, $\xi : \Omega \to \mathbb{R}^l$ denotes a vector of random variables defined on sample space Ω, $\mathbb{E}$ denotes the expected value, and $\mathcal{X}$ denotes a closed subset of $\mathbb{R}^m$.

In this model, we *implicitly assume* that the lower level vector of variables y uniquely solves a stochastic complementarity problem for every x and the realization of $\xi(\omega)$. The uniqueness can be guaranteed by the uniform strong monotonicity of F in y. Therefore in this model y is essentially a function of x and $\xi(\omega)$, not an independent decision vector. This is significantly different from an SMPEC model recently considered by Shapiro [31] where y is regarded as a second decision vector. The optimal upper level variable x is chosen to minimize the expected value of the objective function since the random factors are not realized at the time a decision is made. Model (1) is first investigated by Lin, Chen, and Fukushima [20] with a focus on the case when $\xi(\omega)$ is a random variable with a finite discrete distribution. It is shown that such a program can be transformed into a standard deterministic MPCC. Subsequently, a smoothing method is proposed for solving the transferred program. Lin, Chen, and Fukushima [20] also considered a variation of the model where the complementarity constraint may not necessarily have a solution for every realization of $\xi(\omega)$ and, consequently, a recourse is considered. In a revised version of the paper [21] (with a different title), Lin, Chen, and Fukushima proposed a Monte Carlo method for solving this type of recourse SMPEC model. The work is extended by Lin and Fukushima [22].

In this paper, we focus on the case when $\xi(\omega)$ is a vector of random variables with a known continuous distribution. We find that the case is more challenging in that the resulting SMPCC is no longer equivalent to a standard deterministic MPCC. To be more specific, let $\rho(t)$ denote the joint density function of $\xi(\omega)$ and $\mathcal{T}$ denote the support set of $\rho(t)$. Then (1) can be rewritten as

$$\begin{array}{ll} \min & E(x) := \int_{\mathcal{T}} f(x,y,t)\rho(t)dt \\ \text{s.t.} & x \in \mathcal{X}, \\ & 0 \le y \perp F(x,y,t) \ge 0,\ t \in \mathcal{T}. \end{array} \tag{2}$$

Note that here t is a vector when $l > 1$ and hence the integration is multiple in general. As a result of this reformulation, we have transformed the stochastic program (1) into a deterministic program. Of course, there is a fundamental difference between a standard deterministic mathematical program with complementarity constraints and (2) since here the complementarity constraint contains a vector of parameters t and the objective function involves an integration with respect to t.

We need to investigate the properties of lower level equilibrium solution $y(x,t)$ defined by the complementarity problem in the constraint of (2) before proposing numerical methods to solve the problem. By using a nonlinear complementarity problem (NCP) function, we reformulate the complementarity constraint as an underdetermined system of nonsmooth equations and then investigate the dependence of the lower level prime variable y on the upper level vector of variables x and parametric vector t using a nonsmooth version of the implicit function theorem. We discuss Lipschitz continuity, and piecewise smoothness of $y(x,t)$ on space $\mathcal{X} \times \mathcal{T}$. The discussion is extended to upper level expected value function $E(x)$.

With the nice properties of lower level equilibrium solution and upper level objective function, we propose some numerical methods for solving (2). The methods are focused on addressing two fundamental issues in the problem. One is that since $\mathcal{T}$ is a set of positive Lebesgue measures, $y(x,t)$ is an infinite dimensional variable. This is significantly different from the case when $\mathcal{T}$ is a finite set and (2) can be easily reformulated as a standard deterministic mathematical program with a complementarity constraint. We deal with this issue by discretizing the support set $\mathcal{T}$ and replacing the integration in the objective function with a numerical integration. This kind of deterministic discretization approach is not necessarily efficient when $l > 1$ and/or $\mathcal{T}$ is large, but it is rather stable and suitable for $l = 1$ and/or a small $\mathcal{T}$. The other issue is the nonsmoothness in the constraint caused by the complementarity structure. This is similar to the deterministic MPCC case. We deal with this problem with a popular implicit NCP smoothing method as in the deterministic MPEC case.

During the revision of this paper, a new work on SMPEC by Shaprio has come up. In [31], Shapiro considered a slightly different model from (1) by choosing $y(x,\xi(\omega))$ in such a way that $f(x,y,\xi(\omega))$ is minimized for given x and $\xi(\omega)$, and in doing so he described his model as a two stage stochastic decision making problem. Moreover, he proposed a sample average approximation method to solve the problem and presented a probabilistic estimate of sample size for an ϵ-global optimizer of the original SMPEC to be a δ-optimizer of a sample average approximation program. The sample average approximation approach provides an effective alternative to the deterministic discretization approach that we will discuss in this paper in either case when: (a) $l \ge 2$, (b) the support set $\mathcal{T}$ is large, (c) the distribution of $\xi(\omega)$ is unknown.

The rest of this paper is organized as follows. In section 2, we investigate properties of lower level equilibrium solution using a nonsmooth version of implicit function theorem under the assumption that $F(x, y, t)$ is uniformly strongly monotone with respect to y. We show global Lipschitzness and piecewise smoothness for the lower level equilibrium solution $y(x, t)$. We then move on to discuss properties of upper level expected value function $E(x)$ and show that $E(x)$ is differentiable under some moderate conditions. We also use a stochastic Stackelberg–Nash–Cournot equilibrium problem as an example to discuss the differentiability and convexity of $E(x)$. In section 3, we propose a deterministic discretization approach to approximating (2) and obtain an error bound for an approximate global minimum. Note that our discussion is based on the case when $\xi(\omega)$ is a random variable ($l = 1$), but the results can be easily extended to $l > 1$ case. In section 4, we discuss an implicit smoothing approach for solving (2) and obtain error bounds for a global optimal solution of the smoothed program. Finally, in section 5, we investigate the limiting behavior of the Clarke stationary points of both discretized and smoothed programs.

2. Reformulation and characterization. It is well known that a complementarity problem can be transformed into a system of nonsmooth equations and consequently a deterministic mathematical program with complementarity constraint can be transformed into a program with nonsmooth equality constraints. In this section, we will use the same idea to deal with the complementarity constraints in SMPCC.

2.1. Reformulation of the complementarity constraints. Let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be an NCP function [35], that is, it satisfies at least the following two properties:

$$\phi(a, b) = 0 \Longleftrightarrow a, b \geq 0 \quad \text{and} \quad ab = 0.$$

Then the complementarity constraints in (2) can be reformulated as

$$\Phi(x, y, t) := \begin{pmatrix} \phi(y_1, F_1(x, y, t)) \\ \vdots \\ \phi(y_n, F_n(x, y, t)) \end{pmatrix} = 0. \tag{3}$$

The reformulation is well known; see for instance [18, 16]. There are many NCP functions available in literature; see [35] for a review. Here we only consider the most popular two NCP functions.

One is the "min" function which is defined as

$$\phi(a, b) = \min(a, b).$$

The function is globally Lipschitz continuous and is continuously differentiable everywhere except at the line $a = b$.

The other is the Fischer–Burmeister function [9] which is defined as

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

The function is also globally Lipschitz continuous and is continuously differentiable everywhere except at $(0, 0)$.

With an NCP function, program (2) can be reformulated as

$$\begin{array}{ll} \min & \int_{\mathcal{T}} f(x, y, t)\rho(t)dt \\ \text{s.t.} & x \in \mathcal{X}, \\ & \Phi(x, y, t) = 0. \end{array} \tag{4}$$

Unfortunately, due to the presence of the integral with respect to t in the objective function, there is no available algorithm that can be directly applied to solve program (4). Our purpose here is to properly treat (4) so that it can be solved by existing algorithms for deterministic MPCC. From here on, we will focus on (4) rather than (2).

2.2. Properties of lower level equilibrium. For given x and t, the lower level equilibrium y is defined as a solution of (3). We are interested in the existence, uniqueness of such a solution and its dependence on x and t. For this purpose, we need to make some basic preparations.

Let $F : \mathbb{R}^m \times \mathbb{R}^n_+ \times \mathbb{R}^l \to \mathbb{R}^n$. $F(x, y, t)$ is said to be *uniformly strongly monotone* with respect to y if there exists a constant $\alpha > 0$ such that

$$(F(x, y', t) - F(x, y'', t))^T(y' - y'') \geq \alpha\|y' - y''\|^2 \forall y', y'' \in \mathbb{R}^n_+,\ x \in \mathcal{X},\ t \in \mathcal{T}. \tag{5}$$

Here and later the superscript T denotes the transpose of a vector and matrix.

Let $H : \mathbb{R}^j \to \mathbb{R}^l$ be a locally Lipschitz continuous function. The Clarke generalized Jacobian [3] of H at $x \in \mathbb{R}^j$ is defined as

$$\partial H(x) := \text{conv}\left\{ \lim_{\substack{y \in D_H \\ y \to x}} \nabla H(y) \right\},$$

where D_H denotes the set of points near x at which H is Frechét differentiable, $\nabla H(y)$ denotes the usual Jacobian of H which is a $l \times j$ matrix, "conv" denotes the convex hull of a set. When $l = 1$ or $j = 1$, ∂H reduces to the Clarke subdifferential.

Let $D_a = diag(d_1^a, \dots, d_n^a) \in \mathbb{R}^{n\times n}$ denote the diagonal matrix with the (i, i)th entry being d_i^a, for $i = 1, \dots, n$. Let $D_b = diag(d_1^b, \dots, d_n^b) \in \mathbb{R}^{n\times n}$ denote the diagonal matrix with the (i, i)th entry being d_i^b, for $i = 1, \dots, n$. Let I denote the identity matrix in $\mathbb{R}^{n\times n}$. The function Φ defined by (3) is locally Lipschitz continuous and the Clarke generalized Jacobian of Φ with respect to y can be expressed as

$$\partial_y\Phi(x, y, t) = \left\{ (D_a, D_b) \begin{pmatrix} I \\ \nabla_y F(x, y, t) \end{pmatrix} : (d_i^a, d_i^b) \in \partial\phi(y_i, F_i(x, y, t)), i = 1, \dots, n \right\}. \tag{6}$$

Moreover,

$$\partial_y\Phi(x, y, t) \subset \partial_y\Phi_1(x, y, t) \times \cdots \times \partial_y\Phi_n(x, y, t),$$

where

$$\partial_y\Phi_i(x, y, t) = \{d_i^a e_i + d_i^b \nabla_y F(x, y, t) :\ (d_i^a, d_i^b) \in \partial\phi(y_i, F_i(x, y, t))\},\ i = 1, \dots, n.$$

The following proposition shows that under some proper conditions, the Clarke generalized Jacobian $\partial_y\Phi(x, y, t)$ is uniformly nonsingular.

PROPOSITION 2.1. *Suppose that $F(x, y, t)$ is uniformly strongly monotone with respect to y, and $\phi(a, b)$ is either the min-function or the Fischer–Burmeister function. Then there exists a constant $C > 0$ such that for all $x \in \mathcal{X}$, $y \geq 0$ and $t \in \mathcal{T}$*

$$\|(D_a + D_b\nabla_y F(x, y, t))^{-1}\| \leq C\ \forall (d_i^a, d_i^b) \in \partial\phi(y_i, F_i(x, y, t)),\ i = 1, \dots, n.$$

Here and later on $\|\cdot\|$ denotes the 2-norm of a matrix and a vector.

We will not provide a proof since the result follows straightforwardly from [17, Proposition 3.2] where a similar conclusion is proved with ϕ being the Fischer–Burmeister function and F a P-function (of y). The case when ϕ is a min-function can be dealt with similarly.

Since y is implicitly dependent of x and t through the nonsmooth system of equations (3), the classical implicit function theorem cannot be used to study (3). We need the following generalized implicit function theorem which is established in [36] and is essentially due to Theorem 7.1.1 and the subsequent corollary of [3].

LEMMA 2.2 [36, Lemma 3.2]. *Consider an underdetermined system of nonsmooth equations*

$$H(y, z) = 0,$$

where $H : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ *is locally Lipschitz. Let* $(\bar{y}, \bar{z}) \in \mathbb{R}^m \times \mathbb{R}^n$ *be such that* $H(\bar{y}, \bar{z}) = 0$. *Suppose that* $\partial_y H(\bar{y}, \bar{z})$ *is nonsingular. Then*

(i) *there exist neighborhoods* Z *of* $\bar{z}$, Y *of* $\bar{y}$, *and a locally Lipschitz function* $y : Z \to Y$ *such that* $y(\bar{z}) = \bar{y}$ *and, for every* $z \in Z$, $y = y(z)$ *is the unique solution of the problem* $H(y, z) = 0$, $y \in Y$;

(ii) *for* $z \in Z$,

$$\partial y(z) \subset \{-R^{-1}U : (R, U) \in \partial H(y(z), z), R \in \mathbb{R}^{m\times m}, U \in \mathbb{R}^{m\times n}\}. \tag{7}$$

With Proposition 2.1 and Lemma 2.2, we are ready to give our main results on the lower level equilibrium solution.

THEOREM 2.3. *Let* $\Phi(x, y, t)$ *be defined as in* (3). *Suppose that* F *is uniformly strongly monotone in* y *and uniformly locally Lipschitz continuous in* x. *Then*

(i) *there exists a unique locally Lipschitz continuous function* $y(x, t)$ *such that*

$$\Phi(x, y(x, t), t) = 0 \tag{8}$$

for every $x \in \mathcal{X}$ *and* $t \in \mathcal{T}$;

(ii) *for every* $t \in \mathcal{T}$, $y(\cdot, t)$ *is piecewise smooth in* $\mathcal{X}$; *moreover, if* $\mathcal{T}$ *is a set of positive Lebesgue measure, then* $y(\cdot, \cdot)$ *is piecewise smooth in* $\mathcal{X} \times \mathcal{T}$, *and for fixed* x, $y(x, \cdot)$ *is piecewise smooth in* $\mathcal{T}$;

(iii) *the Clarke generalized Jacobian of* $y(x, t)$ *with respect to* x *can be estimated as follows:*

$$\begin{aligned}\partial_x y(x, t) &\subset \{-R^{-1}U : (U, R, V) \in \partial\Phi(x, y(x, t), t), U \in \mathbb{R}^{n\times m}, R \in \mathbb{R}^{n\times n},\\ &\qquad V \in \mathbb{R}^{n\times l}\}\\ &\subset \{-R^{-1}U : (U, R, V) \in \partial_C\Phi(x, y(x, t), t), U \in \mathbb{R}^{n\times m}, R \in \mathbb{R}^{n\times n},\\ &\qquad V \in \mathbb{R}^{n\times l}\},\end{aligned}$$

where $\partial_C\Phi = \partial\Phi_1 \times \cdots \times \partial\Phi_n$; *moreover, if* F *is uniformly globally Lipschitz continuous in* x, *then* $y(x, t)$ *is also uniformly globally Lipschitz continuous in* x;

(iv) *the Clarke generalized Jacobian of* $y(x, t)$ *with respect to* t *can be estimated as follows:*

$$\begin{aligned}\partial_t y(x, t) &\subset \{-R^{-1}V : (U, R, V) \in \partial\Phi(x, y(x, t), t), U \in \mathbb{R}^{n\times m}, R \in \mathbb{R}^{n\times n},\\ &\qquad V \in \mathbb{R}^{n\times l}\}\\ &\subset \{-R^{-1}V : (U, R, V) \in \partial_C\Phi(x, y(x, t), t), U \in \mathbb{R}^{n\times m}, R \in \mathbb{R}^{n\times n},\\ &\qquad V \in \mathbb{R}^{n\times l}\};\end{aligned}$$

if F is uniformly globally Lipschitz continuous in t, then $y(x,t)$ is also uniformly globally Lipschitz continuous in t.

The results are expected. In particular, similar results to parts (i) and (ii) are established by Facchinei and Pang in the context of sensitivity and stability analysis in [7]. For completeness, we attach a proof which utilizes the nonsmooth implicit function theorem in the appendix.

In practice, $y(x,t)$ represents an equilibrium at scenario t of the uncertainty. The piecewise smoothness of a component $y_i(x,t)$ at a point (x,t) implies that the value of ith lower level decision variable at the equilibrium may change at different rates at the point. In what follows, we investigate the piecewise structure of $y(x,t)$ and its differentiability.

Let

$$\mathcal{D} = \{(x,t) : x \in \mathcal{X}, t \in \mathcal{T}, y_i(x,t) + F_i(x, y(x,t), t) > 0, i = 1, \dots, n\}.$$

Obviously $y(x,t)$ is continuously differentiable on $\mathcal{D}$, and

$$\nabla_x y(x,t) = -\nabla_y \Phi(x, y(x,t), t)^{-1} \nabla_x \Phi(x, y(x,t), t)$$

and

$$\nabla_t y(x,t) = -\nabla_y \Phi(x, y(x,t), t)^{-1} \nabla_t \Phi(x, y(x,t), t) \ \forall x \in \mathcal{D}.$$

In general the structure of set $\mathcal{D}$ is complex even when x is a single variable.

2.3. Properties of the objective function. Let $y(x,t)$ be the solution of (3). We consider the objective function of the SMPCC

$$E(x) := \int_{\mathcal{T}} f(x, y(x,t), t) \rho(t) dt.$$

For simplicity of discussion, we make a blanket assumption that $E(\cdot)$ takes finite value on $\mathcal{X}$. We also assume throughout this subsection that $\mathcal{T}$ is a set of positive Lebesgue measure. We are interested in the properties of $E(x)$ such as Lipschitz continuity, differentiability, and convexity which are related to the development of numerical methods and the uniqueness of optimal solution. Note that in the general context of SMPEC, Patriksson and Wynter [26] investigated Lipschitz continuity and directional differentiability of the objective function. Our approach and results here are more specifically utilizing Clarke subdifferential.

THEOREM 2.4. *Let $\Phi(x,y,t)$ be defined as in* (3). *Suppose that F is uniformly strongly monotone in y and uniformly globally Lipschitz continuous in x. Suppose also that f is globally Lipschitz continuous with respect to (x,y), that is, for every $t \in \mathcal{T}$, there exists $L(t) > 0$ such that*

$$|f(x', y', t) - f(x'', y'', t)| \le L(t)(\|x' - x''\| + \|y' - y''\|) \ \forall x', x'' \in \mathcal{X}, \text{ and } y', y'' \in \mathbb{R}^n_+.$$

Suppose also that

$$\int_{\mathcal{T}} L(t) \rho(t) dt < \infty. \tag{9}$$

Then $E(x)$ is globally Lipschitz continuous and piecewise smooth. Moreover,

$$\partial E(x) \subset \int_{\mathcal{T}} [\nabla_x f(x, y(x,t), t) + \nabla_y f(x, y(x,t), t) \partial_x y(x,t)] \rho(t) dt. \tag{10}$$

Proof. Let $x', x'' \in \mathcal{X}$. Then

$$\begin{aligned}|E(x') - E(x'')| &\le \int_{\mathcal{T}} |f(x', y(x', t), t) - f(x'', y(x'', t), t)|\rho(t)dt \\ &\le \int_{\mathcal{T}} L(t)(\|x' - x''\| + \|y(x', t) - y(x'', t)\|)\rho(t)dt.\end{aligned}$$

By Part (iii) of Theorem 2.3, $y(x, t)$ is uniformly Lipschitz in x. Thus, there exists a constant $C > 0$ such that

$$\|y(x', t) - y(x'', t)\| \le C\|x' - x''\|.$$

Consequently,

$$|E(x') - E(x'')| \le \left[(1 + C)\int_{\mathcal{T}} L(t)\rho(t)dt\right] \|x' - x''\|.$$

The global Lipschitz continuity of $E(x)$ follows from this and (9). Given the Lipschitz continuity, we can obtain (10) by applying [3, Theorem 2.7.2] to $E(x)$. The piecewise smoothness of $E(x)$ is obvious given the piecewise smoothness of $y(x, t)$ and the smoothness of $f(x, y, t)$. ∎

The above theorem shows the global Lipschitzness and subdifferentiability of $E(x)$. In what follows we investigate its differentiability. Let $y(x, t)$ be the solution of (3). For $x \in \mathcal{X}$, let

$$\mathcal{T}_i(x) := \{t \in \mathcal{T} : y_i(x, t) \ge 0, F_i(x, y(x, t), t) \ge 0, y_i(x, t) + F_i(x, y(x, t), t) = 0\} \tag{11}$$

for $i = 1, \dots, n$. This is a set of points in $\mathcal{T}$ where the ith complementarity constraint degenerates for fixed x.

LEMMA 2.5. *$\mathcal{T}_i(x)$ is Lebesgue measurable.*

Proof. By Theorem 2.3, for each fixed x, $y_i(x, \cdot)$ is continuous. Thus $\mathcal{T}_i(x)$ is Lebesgue measurable. ∎

In general, the Lebesgue measure of $\mathcal{T}_i(x)$ in space $\mathcal{T}$ may not be zero.

ASSUMPTION 2.6. *For $i = 1, \dots, n$, the Lebesgue measure of $\mathcal{T}_i(x)$ relative to that of $\mathcal{T}$ is zero.*

In subsection 2.4, we will show that Assumption 2.6 holds in a practical instance.

PROPOSITION 2.7. *Suppose that for any x, t at which $y(\cdot, t)$ is continuously differentiable at x, the following holds:*

$$y(x', t) - y(x, t) - \nabla_x y(x, t)(x' - x) = o(\|x - x'\|). \tag{12}$$

Under Assumption 2.6,

(i) *$E(x)$ is differentiable and*

$$\nabla E(x) = \int_{\mathcal{T}\setminus\mathcal{T}(x)} [\nabla_x f(x, y(x, t), t) + \nabla_y f(x, y(x, t), t)\nabla_x y(x, t)]\rho(t)dt, \tag{13}$$

where

$$\mathcal{T}(x) = \bigcup_{i=1}^{n} \mathcal{T}_i(x);$$

(ii) *if, in addition, ∇f is uniformly Lipschitz continuous in x and y, then $E(\cdot)$ is continuously differentiable on $\mathcal{X}$.*

See a proof in the appendix.

The proposition above shows that the only possibility that $E(\cdot)$ is not differentiable at x is when the Lebesgue measure of $\mathcal{T}(x)$ is not zero.

Finally, we discuss the convexity of $E(x)$. We assume that for every $t \in T$, $f(x, y, t)$ is convex in x, y. Since the density function $\rho(t)$ is nonnegative, the integral of f with respect t gives a convex function of x, y. Unfortunately, these conditions are not adequate to ensure the convexity of $E(x)$ because $E(x)$ involves the integration of the $y(x, t)$. It is obvious that if each component function $y_i(x, t)$ is convex in x for every $t \in \mathcal{T}$, then $E(x)$ is convex. So a sufficient condition to ensure the convexity of $E(x)$ is when $y_i(x, t)$ becomes convex in x.

In general, $y_i(\cdot, t)$ is not necessarily convex. However, under some particular circumstances, we may obtain the convexity of $y_i(\cdot, t)$. We will discuss all these through an example in the next subsection.

2.4. An example. Consider a stochastic Stackelberg–Nash–Cournot equilibrium problem in an oligopoly market where $n + 1$ firms compete to supply homogeneous goods into a market in a noncooperative manner. A strategic firm, called the leader, needs to make a decision for its future output now. Assume that the leader has perfect knowledge of how other firms, called followers, react to his output and the future market distribution. Then the leader's decision problem can be formulated as

$$\max_{x\geq 0}\mathbb{E}\left[xp\left(x + \sum_{i=1}^{n} y_i(x, \xi(\omega)), \xi(\omega)\right)\right] - c_0(x), \tag{14}$$

where

$$y_i(x, \xi(\omega)) \in \arg\max_{y_i\geq 0}\left(y_i p\left(x + y_i + \sum_{k=1, k\neq i}^{n} y_k(x, \xi(\omega)), \xi(\omega)\right) - c_i(y_i)\right), i = 1, \ldots, n. \tag{15}$$

Here x denotes the leader's decision variable, y_i denotes the ith follower's decision variable, and $p(q, \xi(\omega))$ denotes the inverse market demand function which is subject to a random shock $\xi(\omega)$, that is, if the total supply to the market by all firms is q, then market price at scenario $\xi(\omega)$ is $p(q, \xi(\omega))$; $c_0(q)$ denotes the leader's cost function and $c_i(q)$, $i = 1, \ldots, n$, denotes follower i's cost function. We assume that both demand function $p(\cdot, \cdot)$ and cost functions $c_i(q)$, $i = 0, \ldots, n$, are sufficiently smooth.

In this problem, the followers are assumed to play a Nash–Cournot game after the leader's output is known and the market demand is realized and the leader needs to make a decision to maximize its expected profit before the realization of market demand. The problem was initially considered by De Wolf and Smeers [4] in the study of competition in the European gas market where the random variable $\xi(\omega)$ has only a finite discrete distribution. Recently Xu [36] extended the model to the case when the random variable $\xi(\omega)$ has a continuous distribution and reformulated (16) as a stochastic mathematical program with complementarity constraints. Assuming the leader knows the distribution of $\xi(\omega)$, Xu further reformulated the program as follows:

$$\begin{array}{ll} \max & E(x) := \int_0^u \left[xp\left(x + e^T y(x, t), t\right)\right] \rho(t)dt - c_0(x) \\ \text{s.t.} & x \geq 0, \\ & y(x, t) \text{ solves } 0 \leq y \perp F(x, y, t) \geq 0,\ t \in [0, u], \end{array} \tag{16}$$

where

$$F_i(x, y, t) = -p(x + y^T e, t) - p'_x(x + y^T e, t)y_i + c'_i(y_i),\ i = 1, \dots, n,$$

$e = (1, \dots, 1)^T$, and ρ is the density function of the random variable $\xi(\omega)$ with an interval support set $[0, u]$. Note that here and later on we use $a'(x)$ rather than $\nabla a(x)$ to denote the derivative of a real-valued function $a(x)$ with a single variable.

Obviously program (16) is an example of program (2). In what follows, we will investigate the differentiability and convexity of $E(x)$. For simplicity of discussion, we assume that the demand function is linear, that is,

$$p(q, t) = \alpha - \beta q + \gamma t,\ \text{for } t \in [0, u],$$

where $\alpha, \beta, \gamma > 0$.

PROPOSITION 2.8. *$E(x)$ is differentiable for $x > 0$.*

Proof. We use Proposition 2.7 to prove the result. Since $p(q, t) = \alpha - \beta q + \gamma t$,

$$F_i(x, y, t) = -\alpha + \beta(x + y^T e) - \gamma t + \beta y_i + c'_i(y_i),$$

and

$$\frac{dF_i(x, y, t)}{dy_j} = \begin{cases} 2\beta + c''_i(y_i), & j = i, \\ \beta, & j \neq i. \end{cases}$$

Since $\beta > 0$ and $c''_i(q) \geq 0$, it is easy to verify that $\nabla_y F(x, y, t)$ is uniformly positive definite. Therefore by Theorem 2.3, the complementarity problem

$$0 \leq y \perp F(x, y, t) \geq 0$$

has a unique solution $y(x, t)$ for every $x \geq 0$ and $t \in [0, u]$. Note that $y(x, t)$ is followers' Nash–Cournot equilibrium at demand scenario $p(\cdot, t)$. At the Nash equilibrium, the aggregate supply by followers is $y(x, t)^T e$. By Theorem 2.3, $y(x, t)^T e$ is a piecewise smooth function of x and t. Moreover, by [36, Proposition 3.4]

$$\partial_x y(x, t)e \in (-1, 0)\ \forall t \in [0, u]. \tag{17}$$

In what follows, we investigate the monotonicity of $y_i(\cdot, t)$, $i = 1, \dots, n$, for fixed $t \in [0, u]$. By the complementarity condition, we have

$$\min\left(y_i(x, t), -\alpha + \beta\left(x + y(x, t)^T e\right) - \gamma t + \beta y_i(x, t) + c'_i(y_i(x, t))\right) = 0, \text{ for } i = 1, \dots, n.$$

If $y_i(x, t) > 0$, then

$$-\alpha + \beta(x + y(x, t)^T e) - \gamma t + \beta y_i(x, t) + c'_i(y_i(x, t)) = 0. \tag{18}$$

Consequently, we have

$$\partial_x y_i(x, t) \subset -(\beta + c''_i(y_i(x, t)))^{-1}\beta(1 + \partial_x y(x, t)e).$$

Since $\beta > 0$, $c''_i(y_i(x, t)) \geq 0$, by (17), the relation above implies that every element of $\partial_x y_i(x, t)$ is negative. This shows $y_i(\cdot, t)$ is strictly decreasing at a point x where $y_i(x, t) > 0$. Furthermore, from this and the continuity of $y_i(\cdot, t)$, we can easily show that if there exists $x_i(t)$ at which $y_i(x_i(t), t) = 0$, then $y_i(x, t) = 0$ for all $x > x_i(t)$.

Let $x_i(t)$ denote the smallest x at which $y_i(x,t) = 0$ (being $+\infty$ if it does not exist). Then $y_i(\cdot,t)$ is strictly decreasing on $[0, x_i(t)]$ and $y_i(x,t) = 0$ for $x \geq x_i(t)$. The economic interpretation of $x_i(t)$ is that in certain demand scenarios, when the leader's supply reaches $x_i(t)$, follower i drops out of the market since it is no longer making a profit.

We now verify (12) for $y(x,t)$. We consider $y_i(x,t)$. Obviously, (12) is satisfied for $x > x_i(t)$. Let $x \in (0, x_i(t))$ and suppose that $y(\cdot,t)$ is differentiable at x. Then by differentiating both sides of (18) with respect to x, we obtain

$$(y_i)'_x(x,t) = -(2\beta + c_i''(y_i(x,t)))^{-1}\beta \left(1 + \sum_{\substack{j \pm 1 \\ y_j > 0}} (y_j)'_x e \right).$$

Since $c_i(\cdot)$ is assumed to be sufficiently smooth, we can show from the relation above that $y_i(x,t)$ is also twice continuously differentiable. This shows that $y(\cdot,t)$ is twice continuously differentiable at the considered point; consequently, (12) holds.

Now we prove that Assumption 2.6 holds. First let $t \in [0,u]$ be fixed. For $x > x_i(t)$, we have $y_i(x,t) = 0$. Thus

$$\partial_x F_i(x, y(x,t), t) = \beta \left(1 + \partial_x y(x,t) e\right).$$

By (17), every element of $\partial_x F_i(x, y(x,t), t)$ is positive. This shows that $F_i(x, y(x,t), t) > 0$ for $x > x_i(t)$. Therefore $x_i(t)$ is the only point satisfying the following:

$$y_i(x,t) + F_i(x, y(x,t), t) = 0. \tag{19}$$

This shows that for each t there exist at most n degenerate points.

In what follows, we investigate the behavior of $x_i(t)$ as t varies. For this purpose, we need to consider $(y_i)'_t(x,t)$. Consider

$$y_i(x,t) - \alpha + \beta(x + y(x,t)^T e) - \gamma t + \beta y_i(x,t) + c_i'(y_i(x,t)) = 0. \tag{20}$$

By differentiating both sides with respect to t, we obtain that

$$\partial_t y_i(x,t)(1 + c_i''(y_i(x,t)) + \beta) = -\beta \partial_t y(x,t) e + \gamma. \tag{21}$$

Since by part (iii) of [36, Proposition 3.4],

$$\partial_t y(x,t) e \subset \left(0, \frac{\gamma}{\beta}\right],$$

and $c_i''(y_i(x,t)) \geq 0$, we know from (21) that $y_i(x,\cdot)$ is strictly increasing at a considered t where $y_i(x,t) = x_i(t)$. This shows that $x_i(t)$ is strictly increasing as t increases, which means for any x, there exists at most one $t \in [0,u]$ such that $x_i(t) = x$. This show that $\mathcal{T}_i(x)$ defined by (11) contains at most n points. Hence Assumption 2.6 holds. By Proposition 2.7, $E(x)$ is differentiable. ∎

Note that this result significantly strengthens the previous result on lower level equilibrium $y(x,t)$ in [36].

We now investigate the concavity of $E(x)$. For this purpose, we look at the convexity of $y_i(\cdot,t)$. Suppose that $y_i(\cdot,t)$, $i = 1,\dots,n$, is differentiable at x where $y_i(x,t) > 0$. Differentiating (20) with respect to x (ignoring the first term $y_i(x,t)$), we obtain

$$(\beta + c_i''(y_i(x,t))(y_i)'_x(x,t) = -\beta(1 + y'_x e). \tag{22}$$

Since $y_i(\cdot,t)$ is strictly decreasing on $(0, x_i(t))$, for each i and fixed t, there exists at most one point at which (19) is satisfied. Let $x_i(t)$ denote such a point (being $+\infty$

if it does not exist). We are interested in the details of the structure of $y_i(\cdot, t)$. For simplicity of discussion, we further assume that the marginal cost of follower i, c_i, is affine. Then $F_i(x, y, t)$ is affine in x and y, and $y_i(x, t)$ is piecewise affine.

Let

$$\mathcal{X}(t) = \{x_i(t), i = 1, \dots, n\}.$$

For $x \in \mathcal{X} \backslash \mathcal{X}(t)$, the strict complementarity is satisfied for each i, $i = 1, \dots, n$, which means $y_i(\cdot, t)$, $i = 1, \dots, n$, is continuously differentiable at x. Therefore the only possibility that $y_i(\cdot, t)$ becomes nonsmooth (where it switches from one smooth piece to another) is at $x_j(t) \in \mathcal{X}(t)$, where $x_j(t) < x_i(t)$. From (22) we see that

$$\lim_{x \uparrow x_j(t)} (y_i)'_x(x, t) > \lim_{x \downarrow x_j(t)} (y_i)'_x(x, t), \tag{23}$$

which shows that $y_i(\cdot, t)$ is not differentiable at $x_j(t)$. Moreover, (23) indicates a decrease of the derivative of $y_i(\cdot, t)$ at the point $x_j(t)$. The market interpretation is that at the point where follower j drops out of the market, the unit increase on the leader's supply will more significantly reduce the remaining follower's optimal supply. Obviously, (23) indicates the local concavity of $y_i(\cdot, t)$ at point $x_j(t)$. We can summarize the main properties of $y_i(x, t)$ as follows:

- $y_i(x, t)$ is continuous and piecewise affine;
- for $x < x_i(t)$, $y_i(\cdot, t)$ is concave, and at $x_i(t)$, $y_i(\cdot, t)$ is locally convex;
- if the followers are identical, then $y_i(\cdot, t)$ is convex;
- if $X(t) = \{+\infty\}$, that is, no follower drops out as the leader increases supply up to its capacity, then $y_i(\cdot, t)$ is convex;
- $y_i(x, t)$ is not differentiable at $x_j(t) < x_i(t)$, $j = 1, \dots, n$ and $x_i(t)$, but it is differentiable elsewhere.

Note that, at this stage we are not ready to assert whether or not $E(x)$ is concave. Observe first that $E(x)$ is a function of $Q(x, t)$, where $Q(x, t) = \sum_{i=1}^n y_i(x, t)$. If we can show that $Q(x, t)$ is convex, then it is not difficult to see that $E(x)$ is concave under usual assumptions [36]. For this purpose we look at the convexity of $Q(\cdot, t)$. Let $\mathcal{I}(x, t) = \{i : y_i(x, t) > 0\}$. Since $c_i'' = 0$, we have from (22) that

$$Q'_x(x, t) = -\frac{|\mathcal{I}(x, t)|}{1 + |\mathcal{I}(x, t)|},$$

where $|\mathcal{I}(x, t)|$ denotes the cardinality of set $|\mathcal{I}(x, t)|$. Obviously, as the value of x changes from $x_i(t) - \delta$ to $x_i(t) + \delta$, where δ is sufficiently small, $|\mathcal{I}(x, t)|$ decreases and $Q'_x(\cdot, t)$ increases. This shows the convexity of $Q(\cdot, t)$. Note that Sherali [33] obtained a similar conclusion in a deterministic Stackelberg model. Based on the discussion above, we have the following.

PROPOSITION 2.9. *If $p(q, t)$ is affine and $c_i(q)$, $i = 1, \dots, n$, is also affine, then $E(x)$ is concave for $x \geq 0$.*

3. Discretization methods. In this section, we discuss numerical methods for solving programs (2) through (4). The main obstacle that prevents direct application of many recently developed numerical methods for deterministic MPEC to (2) is the presence of an integral in the objective function which requires lower level variables to be solved from constraint before the objective function can be evaluated. In general, it is difficult to obtain an explicit expression of $y(x, t)$ even when F is an affine function. Our idea here is to discretize the integral and replace it with a numerical integration.

The resulting discretized program can then be solved by available numerical methods for deterministic MPECs.

To simplify the discussion, we focus on the case when $\xi(\omega)$ is a random variable. It is not difficult to see that the methods and results established in this section can be easily extended to the case when $\xi(\omega)$ consists of several random variables.

First, we deal with possible unboundedness of the support set $\mathcal{T}$. The following result is a special case of Berge's well-known stability theorem and is needed in several places in later discussion.

LEMMA 3.1. *Consider a general constrained minimization problem*

$$\begin{array}{ll} \min & p(x) \\ \text{s.t.} & x \in \mathcal{C}, \end{array}$$

where $p : \mathbb{R}^n \to \mathbb{R}$ *is continuous and* $\mathcal{C}$ *is a subset of* $\mathbb{R}^n$*, and a perturbed program*

$$\begin{array}{ll} \min & \tilde{p}(x) \\ \text{s.t.} & x \in \mathcal{C}, \end{array}$$

where $\tilde{p} : \mathbb{R}^n \to \mathbb{R}$ *is continuous and*

$$|\tilde{p}(x) - p(x)| \leq \delta \ \forall x \in \mathcal{C}.$$

Suppose that x^* *is a global minimizer of* $p(x)$ *over* $\mathcal{C}$*, and* $\tilde{x}^*$ *is a global minimizer of* $\tilde{p}(x)$ *over* $\mathcal{C}$*. Then*

$$|p(x^*) - \tilde{p}(\tilde{x}^*)| \leq \delta.$$

PROPOSITION 3.2. *Suppose that the support set* $\mathcal{T}$ *is unbounded. Let*

$$\mathcal{T}_N := \{t \in \mathcal{T} : \|t\|_\infty \leq N\},$$

where N *is a positive number and* $\|\cdot\|_\infty$ *denotes the infinity norm. Let* x_N *be a global minimizer of the following program:*

$$\begin{array}{ll} \min & E_N(x) := \int_{\mathcal{T}_N} f(x,y,t)\rho(t)dt \\ \text{s.t.} & x \in \mathcal{X}, \\ & \Phi(x,y,t) = 0. \end{array} \tag{24}$$

Then for every $\delta > 0$*, there exists* $N_0 > 0$ *such that for all* $N > N_0$*,*

$$|E(x) - E_N(x)| \leq \delta \ \forall x \in \mathcal{X},$$

and

$$|E(x^*) - E_N(x_N)| \leq \delta,$$

where x^* *denotes a global minimizer of program* (4).

Proof. The first inequality is obvious. The second inequality follows from the first one and Lemma 3.1. ∎

The proposition shows that we can approximate program (4) with (24). To simplify the discussion, we assume, from here on, that the support set $\mathcal{T}$ is bounded. Since $\xi(\omega)$ is a random variable, $\mathcal{T}$ is a bounded interval. We normalize it to $[0, u]$. Let $\mathcal{T}_K$ denote a set of grid points of $\mathcal{T}$ where

$$\mathcal{T}_K = \left\{ t : t_0 = 0, t_l = t_{l-1} + \frac{u}{K}, \text{ for } l = 1, \ldots, K \right\}.$$

Note that we can discretize the program (2) directly by considering

$$
(25)\qquad \begin{aligned} \min \quad & E_K(x) := \frac{u}{K}\sum_{l=1}^{K} f(x, y(x, t_l), t_l)\rho(t_l) \\ \text{s.t.} \quad & x \in \mathcal{X}, \\ & y(x, t_l) \text{ solves } 0 \le y \perp F(x, y, t_l) \ge 0,\ l = 1, \dots, K. \end{aligned}
$$

PROPOSITION 3.3. *Let $E_K(x)$ be defined as in* (25). *Suppose that $f(x, y(x,t), t)$ is uniformly locally Lipschitz with respect to t, that is, for every $t \in \mathcal{T}$, there exists a constant $A(t) > 0$ such that*

$$
(26)\qquad |f(x, y(x, t''), t'') - f(x, y(x, t'), t')| \le A(t)|t'' - t'|
$$

for t', t'' near t, where $A : \mathcal{T} \to \mathbb{R}^+$ is continuous and

$$
(27)\qquad \int_0^u A(t)\rho(t)dt < \infty.
$$

Suppose also that the density function $\rho(t)$ is differentiable on $\mathcal{T}$ and

$$
(28)\qquad \int_0^u |f(x, y(x, t), t)\rho'(t)|dt < \infty.
$$

Then

(i) *there exists a constant $\tilde{C}$ such that*

$$
(29)\qquad |E_K(x) - E(x)| \le \frac{\tilde{C}u}{K}\ \forall x \in \mathcal{X};
$$

(ii)

$$
|E_K(x_K) - E(x^*)| \le \frac{\tilde{C}u}{K},
$$

where x_K denotes a global minimizer of $E_K(\cdot)$ and x^* denotes a global minimizer of $E(\cdot)$.

Proof. Part (i). Let

$$
\Delta_l(x, t) := f(x, y(x, t_l), t_l)\rho(t_l) - f(x, y(x, t), t)\rho(t), \text{ for } t \in (t_{l-1}, t_l).
$$

By definition,

$$
E_K(x) - E(x) = \frac{u}{K}\sum_{l=1}^{K}\int_{t_{l-1}}^{t_l} \Delta_l(x, t)dt.
$$

Since

$$
|\Delta_l(x, t)| \le \frac{u}{K}\left(\rho(t_l)\sup_{t\in[t_{l-1},t_l]} A(t) + |f(x, y(x, t), t)|\sup_{t\in[t_{l-1},t_l]}|\rho'(t)|\right), \text{ for } t \in (t_{l-1}, t_l),
$$

by (27) and (28), there exists a constant $\delta > 0$ such that for K sufficiently large

$$
\begin{aligned} |E_K(x) - E(x)| &\le \frac{u}{K}\sum_{l=1}^{K}\int_{t_{l-1}}^{t_l}\left(\rho(t_l)\sup_{t\in[t_{l-1},t_l]} A(t) + |f(x, y(x, t), t)|\sup_{t\in[t_{l-1},t_l]}|\rho'(t)|\right)dt \\ &\le \left(\int_0^u A(t)\rho(t)dt + \int_0^u |f(x, y(x, t), t)||\rho'(t)|dt + \delta\right)\frac{u}{K}. \end{aligned}
$$

The conclusion follows by letting $\tilde{C} = \left(\int_0^u A(t)\rho(t)dt + \int_0^u |f(x, y(x,t), t)||\rho'(t)|dt + \delta\right)$.

Part (ii) follows from Part (i) and Lemma 3.1. ∎

Remark 1. We make a few comments on the assumptions made in Proposition 3.3. First, the condition on the continuous differentiability of $\rho(t)$ can be relaxed to piecewise smoothness. Second, (26) and (27) holds under the following conditions:

(a) $f(x, y, t)$ is globally Lipschitz continuous in x and y and (9) holds,
(b) $f(x, y, t)$ is uniformly locally Lipschitz continuous in t with rank $A_1(t)$ where $\int_0^u A_1(t)\rho(t) < \infty$,
(c) F is uniformly monotone in y and it is uniformly globally Lipschitz continuous in t.

Note that both (a) and (c) are assumed in Theorem 2.4.

The advantage of (25) is that we can rewrite it as

$$
\begin{array}{ll}
\min & E_K(x) := \dfrac{u}{K}\displaystyle\sum_{l=1}^{K} f(x, y_l, t_l)\rho(t_l) \\
\text{s.t.} & x \in \mathcal{X}, \\
 & 0 \le y_l \perp F(x, y_l, t_l) \ge 0,\ l = 1, \dots, K,
\end{array}
\tag{30}
$$

where we can treat x and $y_1, \dots, y_K$ equally in the sense that there is no need to solve y_l from constraints in advance. Note that (30) may be viewed as a discrete stochastic mathematical program with complementarity constraints if we regard $\rho(t_l), l = 1, \dots, K$ as probability distribution. See [26, 20] for some research on discrete stochastic mathematical program with complementarity constraints.

It is easy to see that (30) is a deterministic mathematical program with complementarity constraint, therefore a number of numerical methods proposed in [10, 11, 6, 12, 13, 15, 16, 19, 34] can be applied to this program.

The disadvantage is that to obtain a better approximation, K may be large and, consequently, a large number of variables are introduced in (30). The discretized scheme is useful only when the support set of the density function is small and/or the random variable is relatively evenly distributed over the support set.

An alternative approach to solving (2) is the sample average approximation (SAA) method. SAA is well known in stochastic programming and is effective when a problem involves several random variables. The basic idea is to generate a sample $\xi^1, \dots, \xi^N$ with independent identical distribution as ξ and to solve the following SAA program:

$$
\begin{array}{ll}
\displaystyle\min_{x \in \mathcal{X}} & \frac{1}{N}\sum_{i=1}^{N} \left[f(x, y^i, \xi^i)\right] \\
\text{s.t.} & 0 \le y^i \perp F(x, y^i, \xi^i) \ge 0,\ i = 1, \dots, N,
\end{array}
\tag{31}
$$

to obtain an approximate solution of the original problem (1). In comparison with (30), the SAA scheme generates less evenly spread grid points which usually concentrate in areas where the density function take relatively larger values; see [31] for details.

4. An implicit smoothing method. In this section, we deal with the nonsmoothness of lower equilibrium solution $y(x, t)$. It is well known that such nonsmoothness arises from the nature of complementarity. The issue has been extensively discussed in deterministic MPCC and many methods have been proposed to deal with it. It is beyond the scope of this paper to give a comprehensive review on this topic. Here we just mention two types of methods.

One is called the smoothing NCP function method. The idea of this kind of method is to find a smooth approximation of an NCP function and replace the NCP reformulation with such a smoothed approximate NCP reformulation; see recent work in [6, 16] and references therein.

The other kind of method is called the regularization method which reformulates complementarity constraints as a system of inequalities. Such a system is often ill-posed because some constraint qualifications may not hold at any feasible point. A small perturbation at the right-hand side of the system may effectively overcome this problem; see [34] for details.

Both methods will generate a smooth approximation of the solution of a complementarity problem. Here we will use the smoothing NCP function methods.

Recall that the smoothing of an NCP function $\phi(a,b)$ is a function $\psi(a,b,c)$ satisfying the following:

(A1) $\psi(a,b,0)=\phi(a,b)$;

(A2) $\psi(a,b,c)$ is Lipschitz continuous and is continuously differentiable everywhere except at $c=0$;

(A3) (Strong Jacobian Consistency [1]) for $(a,b)\in\mathbb{R}^2$,

$$\partial_{a,b}\psi(a,b,0)=\partial\phi(a,b).$$

A smoothing function of $\min(a,b)$ is

$$\psi(a,b,c)=-\frac{1}{2}\left(\sqrt{(a-b)^2+c^2}-(a+b)\right)$$

and a smoothing function of the Fischer–Burmeister function is

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

see, for instance, [18].

Let $H:\mathbb{R}^n\to\mathbb{R}^m$ be a locally Lipschitz function. The ϵ-generalized Jacobian is defined as

$$\partial^{\epsilon}H(x)=\operatorname{conv}\bigcup_{y\in B(x,\epsilon)}\partial H(y),$$

where $B(x,\epsilon)$ denotes the unit ball in $\mathbb{R}^n$ centered at x with radius ϵ. The notion was introduced in [37] as a generalization of ϵ-subdifferential [30] for the purpose of the approximation of the Clarke generalized Jacobian in solving nonsmooth equations.

LEMMA 4.1. *Let $\psi(a,b,c)$ be a smoothing of an NCP function $\phi(a,b)$ satisfying properties* A1–A3. *Then there exists continuous function $\epsilon:\mathbb{R}_+\to\mathbb{R}_+$ such that a*

$$\nabla_{a,b}\psi(a,b,c)\in\partial^{\epsilon(c)}\phi(a,b) \tag{32}$$

for c close to 0*, where*

$$\lim_{c\to 0}\epsilon(c)=0.$$

Proof. The conclusion follows from the upper semicontinuity of the Clarke generalized Jacobian and the strong Jacobian consistency of ψ. ∎

Let ψ be either the smoothing of the min-function or the smoothing of the Fischer–Burmeister function. Let

$$\Psi(x,y,t,\mu) := \begin{pmatrix} \psi(y_1, F_1(x,y,t),\mu) \\ \vdots \\ \psi(y_n, F_n(x,y,t),\mu) \end{pmatrix}. \tag{33}$$

Then

$$\Psi(x,y,t,0) = \begin{pmatrix} \psi(y_1, F_1(x,y,t),0) \\ \vdots \\ \psi(y_n, F_n(x,y,t),0) \end{pmatrix} = \begin{pmatrix} \phi(y_1, F_1(x,y,t)) \\ \vdots \\ \phi(y_n, F_n(x,y,t)) \end{pmatrix}.$$

We consider the following program which is a smoothing of (4):

$$\begin{array}{ll} \min & \int_{\mathcal{T}} f(x,y,t)\rho(t)dt \\ \text{s.t.} & x \in \mathcal{X}, \\ & \Psi(x,y,t,\mu) = 0. \end{array} \tag{34}$$

We regard the approach as implicit smoothing in the sense that by replacing Φ with Ψ, we achieve the smoothing of the implicit function $y(x,t)$. Note that Lin, Chen, and Fukushima [20] considered a similar approach for a class of discrete stochastic mathematical programs with complementarity constraint. Here we rely more heavily on the implicit function approach in dealing with (34).

Recall that a vector-valued function $g : \mathbb{R}^n \to \mathbb{R}^m$ is called *calm* at point $\bar{x}$ if there exists a $\kappa > 0$ such that

$$\|g(x) - g(\bar{x})\| \leq \kappa \|x - \bar{x}\|$$

for all x near $\bar{x}$; see page 351 in [28].

PROPOSITION 4.2. *Suppose that $\mathcal{T}$ is a set of positive Lebesgue measures. Suppose also that F is uniformly strongly monotone with respect to y and it is uniformly locally Lipschitz continuous with respect to x. Suppose that ϕ is either min-function or the Fischer–Burmeister function. Then*

(i) $\partial_y \Psi(x,y,t,\mu)$ *is uniformly nonsingular and there exists $\mu_0 > 0$ such that the system of equations*

$$\Psi(x,y,t,\mu) = 0$$

defines a unique implicit function $\tilde{y}(x,t,\mu)$ which satisfies

$$\Psi(x,\tilde{y}(x,t,\mu),t,\mu) = 0, \text{ for } x \in \mathcal{X}, t \in \mathcal{T}, \ |\mu| \in (0,\mu_0];$$

(ii) $\tilde{y}(x,t,\mu)$ *is continuously differentiable with respect to (x,t,μ) on $\mathcal{X} \times \mathcal{T} \times [-\mu_0,\mu_0]\backslash\{0\}$; it is locally Lipschitz continuous with respect to x and t if F is so;*

(iii) $\tilde{y}(x,t,\mu)$ *is uniformly calm in μ at 0, that is, there exists $\hat{C} > 0$ such that*

$$\|\tilde{y}(x,t,\mu) - \tilde{y}(x,t,0)\| \leq \hat{C}|\mu|, \ \textit{for } |\mu| \in [0,\mu_0]; \tag{35}$$

(iv) *there exists a real valued function $\epsilon : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$\nabla_x \tilde{y}(x,t,\mu) \in \partial_x^{\epsilon(\mu)} y(x,t), \ |\mu| \in [0,\mu_0], \tag{36}$$

where $\lim_{\mu\to 0} \epsilon(\mu) = 0$.

From part (iv), we see that any accumulation matrix of $\nabla_x \tilde{y}(x,t,\mu)$ as $\mu \to 0$ is contained in the Clarke generalized Jacobian $\partial y(x,t)$.

Proof of Proposition 4.2. Part (i). The uniform nonsingularity of $\partial_y \Psi(x,y,t,\mu)$ follows essentially from [17, Proposition 3.2]. The existence and uniqueness of $\tilde{y}(x,t,\mu)$ for some $\mu_0 > 0$ follows from part (i) of Theorem 2.3.

Part (ii). The continuous differentiability of $\tilde{y}(x,t,\mu)$ follows from Part (i) and the classical implicit function theorem. We can prove the local Lipschitz continuity by applying Lemma 2.2.

Since $\partial_y \Psi(x,y,t,\mu)$ is uniformly nonsingular, $\partial_{(x,t,\mu)}\Psi(x,y,t,\mu)$ is bounded in a closed neighborhood of $(x,t,\mu) \in \mathcal{X} \times \mathcal{T} \times [-\mu_0,\mu_0]\backslash\{0\}$. It is evident by part (ii) Lemma 2.2 that $\partial\tilde{y}(x,t,\mu)$ is bounded, hence $\tilde{y}(x,t,\mu)$ is locally Lipschitz continuous.

Part (iii). Since $\nabla\tilde{y}(x,t,\mu)$ is continuous for $\mu \neq 0$,

$$\tilde{y}(x,t,\mu) - \tilde{y}(x,t,0) = \int_0^1 \tilde{y}'_\mu(x,t,\mu\nu)\mu d\nu.$$

Thus

$$\begin{aligned}\|\tilde{y}(x,t,\mu) - \tilde{y}(x,t,0)\| &\le |\mu| \int_0^1 \|\tilde{y}'_\mu(x,t,\mu\nu)\| d\nu \\ &= |\mu| \int_0^1 \|\nabla_y \Psi(x,\tilde{y}(x,t,\mu),t,\mu\nu)^{-1}\Psi'_\mu(x,\tilde{y}(x,t,\mu\nu),t,\mu)\| d\nu \\ &\le \hat{C}|\mu|,\end{aligned}$$

where $\hat{C}$ is a constant.

Part (iv). By definition,

$$\Psi(x,\tilde{y}(x,t,\mu),t,\mu) = 0.$$

By the classical implicit function theorem,

$$\nabla_x \tilde{y}(x,t,\mu) = -\nabla_y \Psi(x,\tilde{y}(x,t,\mu),t,\mu)^{-1}\nabla_x \Psi(x,\tilde{y}(x,t,\mu),t,\mu).$$

Consider $\nabla\Psi(x,\tilde{y}(x,t,\mu),t,\mu)$. Since $\tilde{y}(x,t,\cdot)$ is uniformly calm at $\mu = 0$ as we proved in part (iii), by Lemma 4.1, we know that there exists $\epsilon_1(\mu) > 0$ such that

$$\nabla\Psi(x,\tilde{y}(x,t,\mu),t,\mu) \in \partial^{\epsilon_1(\mu)}\Psi(x,y,t).$$

By the definition of ϵ-generalized Jacobian and part (ii) of Lemma 2.2, there exists $\epsilon : \mathbb{R} \to \mathbb{R}_+$, $\epsilon(\mu) \to 0$ as $\mu \to 0$, such that (36) holds. This completes the proof. □

COROLLARY 4.3. *Suppose that $\mathcal{T}$ is a set of positive Lebesgue measure. Suppose also that F is uniformly strongly monotone with respect to y and is uniformly Lipschitz continuous with respect to x. Let*

$$\tilde{E}(x,\mu) = \int_{\mathcal{T}} f(x,\tilde{y}(x,t,\mu),t)\rho(t)dt. \tag{37}$$

Then there exists $\mu_0 > 0$ such that

(i) *$\tilde{E}(x,\mu)$ is uniformly calm with respect to μ at 0, that is, there exists $\hat{C} > 0$ such that*

$$\|\tilde{E}(x,\mu) - E(x)\| \le \hat{C}|\mu|, \ \textit{for } |\mu| \in [0,\mu_0]; \tag{38}$$

(ii) *there exists a real valued function* $\epsilon : \mathbb{R} \to \mathbb{R}_+$ *such that*

$$\nabla_x \tilde{E}(x,\mu) \in \partial^{\epsilon(\mu)} E(x), \ |\mu| \in [0, \mu_0], \tag{39}$$

where $\lim_{\mu \to 0} \epsilon(\mu) = 0$.

Proof. Part (i). By the assumption on f and part (iii) of Proposition 4.2,

$$\begin{aligned} |f(x,\tilde{y}(x,t,\mu),t) - f(x,\tilde{y}(x,t,0),t)| &\le L(t)\|\tilde{y}(x,t,\mu) - \tilde{y}(x,t,0)\| \\ &\le \hat{C}L(t)|\mu|. \end{aligned}$$

Hence

$$|\tilde{E}(x,\mu) - E(x)| \le \hat{C}|\mu| \int_{\mathcal{T}} L(t)\rho(t)dt.$$

Part (ii) follows from Part (iv) of Proposition 4.2. $\square$

Corollary 4.3 shows that $\tilde{E}(x,\mu)$ approximates $E(x)$ uniformly. It also implies that any accumulation vector of $\nabla_x \tilde{E}(x,\mu)$ as $\mu \to 0$ is an element of $\partial E(x)$. Therefore $\nabla_x \tilde{E}(x,\mu)$ can be used to calculate an element of the Clarke subdifferential of $E(x)$.

THEOREM 4.4. *Let* $\{\mu_k\}$ *be a strictly decreasing sequence such that* $\mu_k \downarrow 0$ *as* $k \to \infty$. *Let* $\{(x_k, \tilde{y}(x_k,\cdot,\mu_k))\}$ *be a sequence of solutions of* (34). *Under the conditions of Proposition* 4.2*,*

(i) *any accumulation point of* $\{(x_k, \tilde{y}(x_k,\cdot,\mu_k))\}$ *is a solution of* (4)*;*

(ii) *there exists a constant* $C > 0$ *such that*

$$|\tilde{E}(x_k,\mu_k) - E^*| \le C\mu_k,$$

where E^* *denotes the minimum of* (2)*;*

(iii) *if* x *is an accumulation point of* $\{x_k\}$ *and* M *is an accumulation matrix of* $\{\nabla_x \tilde{y}(x_k,t,\mu_k)\}$ *and* ξ *is an accumulation vector of* $\{\nabla_x \tilde{E}(x_k,u_k)\}$*, then* $M \in \partial_x y(x,t)$ *and* $\xi \in \partial E(x)$.

Proof. Parts (i) and (ii) follow from part (i) of Corollary 4.3 and Lemma 3.1. Part (iii) follows from part (iv) of Proposition 4.2 and part (ii) of Corollary 4.3. $\square$

Theorem 4.4 ensures a smooth approximation of (2) by (34). There exist at least two ways to solve the latter. One is to consider

$$\begin{array}{ll} \min & \tilde{E}(x,\mu) \\ \text{s.t.} & x \in \mathcal{X} \end{array} \tag{40}$$

and solve it with a smooth nonlinear programming method which depends only on the function and gradient values of $\tilde{E}(x,\mu)$. In this way, we only treat x as a variable. The other is to discretize the smoothed program (34). In what follows, we consider the latter.

We consider the discretized smoothed program

$$\begin{array}{ll} \min & \tilde{E}_K(x,\mu) := \dfrac{u}{K}\displaystyle\sum_{l=0}^{K} f(x,\tilde{y}(x,t_l,\mu),t_l)\rho(t_l) \\ \text{s.t.} & x \in \mathcal{X}, \\ & \tilde{y}(x,t_l,\mu) \text{ solves } \Psi(x,y,t_l,\mu) = 0, \ l = 1,\dots,K. \end{array} \tag{41}$$

PROPOSITION 4.5. *Let* $\tilde{E}_K(x,\mu)$ *be defined as in* (41)*. Suppose that* $f(x,\tilde{y}(x,t,\mu),t)$ *is uniformly locally Lipschitz with respect to* t*, that is, for every* $t \in \mathcal{T}$*, there exists a positive constant* $\tilde{A}(t)$ *(depends on* t*) such that*

$$|f(x,\tilde{y}(x,t'',\mu),t'') - f(x,\tilde{y}(x,t',\mu),t')| \le \tilde{A}(t)|t'' - t'| \tag{42}$$

for all t', t'' near t, where $\tilde{A} : \mathcal{T} \to \mathbb{R}^+$ is continuous. Moreover, there exists $\mu_0 > 0$, such that for $\mu \in [0, \mu_0]$,

$$\int_0^u \tilde{A}(t)\rho(t)dt < \infty. \tag{43}$$

Suppose also that the density function $\rho(t)$ is differentiable on $\mathcal{T}$ and

$$\int_0^u |f(x, \tilde{y}(x, t, \mu), t)\rho'(t)|dt < \infty, \tag{44}$$

where $\rho'(t)$ denotes the derivative of $\rho(t)$. Then

(i) *there exists a constant $\tilde{C}$ such that*

$$|\tilde{E}_K(x, \mu) - \tilde{E}(x, \mu)| \leq \frac{\tilde{C}u}{K} \ \forall x \in \mathcal{X}; \tag{45}$$

(ii)

$$|\tilde{E}_K(x_K^\mu, \mu) - \tilde{E}(x_\mu, \mu)| \leq \frac{\tilde{C}u}{K}, \tag{46}$$

where x_K^μ denotes a global minimizer of $\tilde{E}_K(x, \mu)$ and x_μ denotes a global minimizer of $\tilde{E}(x, \mu)$.

We omit the proof as it is similar to that of Proposition 3.3.

It might be helpful to make a few comments about conditions (42)–(44). It is not difficult to verify that (42)–(44) hold under the conditions (a)–(c) in Remark 1 and (28). Indeed, under the condition (c), both $\|\nabla_y \Psi(x, \tilde{y}(x, t, \mu), t, \mu)^{-1}\|$ and $\|\nabla_t \Psi(x, \tilde{y}(x, t, \mu), t, \mu)\|$ are uniformly bounded for μ sufficiently small. Since

$$\nabla_t \tilde{y}(x, t, \mu) = -\nabla_y \Psi(x, \tilde{y}(x, t, \mu), t, \mu)^{-1} \nabla_t \Psi(x, \tilde{y}(x, t, \mu), t, \mu),$$

then $\nabla_t \tilde{y}(x, t, \mu)$ is uniformly bounded which implies that $\tilde{y}(x, \cdot, \mu)$ is uniformly globally Lipschitz continuous in set $\mathcal{T}$. Combining this with conditions (a) and (c) in the remark, we can prove (42) and (43). Finally, (44) follows the uniform calmness of $\tilde{y}(x, t, \cdot)$ at $\mu = 0$ and (28).

Based on Proposition 4.5, we can solve the smoothed program (34) by solving the discretized program (41). Since the latter is a typical deterministic smooth mathematical program with complementarity constraint, it can be solved by a number of existing algorithms such as those proposed by Jiang and Ralph [16]. Note that if we choose μ to be a proportion of T/K, we can easily obtain an estimation of $|\tilde{E}_K(x_K^\mu, \mu) - E^*|$ using Theorem 4.4 and Proposition 4.5.

Note also that in order to reduce the error bound in (46), we need to increase the number of grid points K, which means increasing the number of lower level variables and equality constraints in (41). This may increase problem size and reduce the computational efficiency. In contrast, the first way may avoid the increase of problem size although it also requires discretization of $\mathcal{T}$ to compute numerically the function and gradient values of $\tilde{E}(x, \mu)$.

5. Optimality conditions. In the preceding sections, we outlined three ways to solve (2): (a) solving discretized program (25) and increase K if necessary; (b) solving smoothed program (40); (c) solving smoothed discretized program (41). The error bounds obtained in Proposition 3.3, Theorem 4.4, and Proposition 4.5 are based on *global* minimizers of the relevant programs although these results would also apply to local minimizers after localizing the set $\mathcal{C}$ in Lemma 3.1. In practice, finding a global minimizer might be difficult and in some cases we might just find a stationary point. Consequently, we want to know whether or not an accumulation point of the sequence of stationary points is a stationary point of program (4). For this purpose, we need to investigate the optimality conditions of program (4), the discretized program (25), the smoothed program (40), and smoothed discretized program (41) and their relationship.

Program (4). The generalized Karush–Kuhn–Tucker (KKT) condition [14] of the program (4) is

$$0 \in \int_{\mathcal{T}} [\nabla_x f(x, y(x,t), t)^T + \partial_x y(x,t)^T \nabla_y f(x, y(x,t), t)^T] \rho(t) dt + \mathcal{N}_{\mathcal{X}}(x), \tag{47}$$

where $\mathcal{N}_{\mathcal{X}}(x)$ denotes the normal cone of $\mathcal{X}$ at x, that is,

$$\mathcal{N}_{\mathcal{X}}(x) = \{d : d^T(x' - x) \leq 0 \ \forall x' \in \mathcal{X}\}.$$

A point x^* satisfying the KKT condition is known as a Clarke stationary point. Using the estimation of $\partial_x y(x,t)$ in Part (iii) of Theorem 2.3, we obtain

$$0 \in \int_{\mathcal{T}} [\nabla_x f(x, y(x,t), t)^T + \Im_\Phi(x,t)^T \nabla_y f(x, y(x,t), t)^T] \rho(t) dt + \mathcal{N}_{\mathcal{X}}(x) \tag{48}$$

where

$$\Im_\Phi(x,t) = \{-R^{-1}U : (U, R, V) \in \partial\Phi(x, y(x,t), t), U \in \mathbb{R}^{n\times m}, R \in \mathbb{R}^{n\times n}, V \in \mathbb{R}^{n\times l}\}.$$

Discretized program. Consider the discretized program (30) which is equivalent to

$$\begin{array}{ll} \min & E_K(x) := \dfrac{u}{K} \displaystyle\sum_{l=1}^{K} f(x, y_l, t_l)\rho(t_l) \\ \text{s.t.} & x \in \mathcal{X}, \\ & \Phi(x, y_l, t_l) = 0, \ l = 1, \ldots, K. \end{array} \tag{49}$$

Note that (49) can be viewed as a discretized program of (4). The generalized KKT condition of (49) is

$$\left\{ \begin{array}{l} 0 \in \dfrac{u}{K} \displaystyle\sum_{l=1}^{K} \nabla_x f(x, y_l, t_l)^T \rho(t_l) + \sum_{l=1}^{K} \partial_x \Phi(x, y_l, t_l)^T \lambda_l + \mathcal{N}_{\mathcal{X}}(x), \\ 0 \in \dfrac{u}{K} \nabla_{y_l} f(x, y_l, t_l)^T \rho(t_l) + \partial_{y_l} \Phi(x, y_l, t_l)^T \lambda_l, l = 1, \ldots, K, \end{array} \right.$$

which can be equivalently written as

$$0 \in \frac{u}{K} \sum_{l=1}^{K} \left[\nabla_x f(x, y_l, t_l)^T - \partial_x \Phi(x, y_l, t_l) \partial_{y_l} \Phi(x, y_l, t_l)^{-T} \nabla_{y_l} f(x, y_l, t_l)^T\right] \rho(t_l) + \mathcal{N}_{\mathcal{X}}(x).$$

Solving $y_l, l = 1, \dots, K$, from $\Phi(x, y_l, t_l) = 0$ and writing $\partial_{y_l}\Phi(x, y_l, t_l)$ and $\nabla_{y_l} f(x, y_l, t_l)$ as $\partial_y\Phi(x, y(x, t_l), t_l)$ and $\nabla_y f(x, y(x, t_l), t_l)$, we can rewrite the KKT condition as

$$0 \in \frac{u}{K}\sum_{l=1}^{K}[\nabla_x f(x, y(x, t_l), t_l)^T - \partial_x\Phi(x, y(x, t_l), t_l)^T \partial_y\Phi(x, y(x, t_l), t_l)^{-T}\nabla_y f(x, y(x, t_l), t_l)^T]\rho(t_l) + \mathcal{N}_{\mathcal{X}}(x). \tag{50}$$

Naturally, we would like to link (50) to the following condition:

$$0 \in \int_0^u [\nabla_x f(x, y(x, t), t)^T - \partial_x\Phi(x, y(x, t), t)^T \partial_y\Phi(x, y(x, t), t)^{-T}\nabla_y f(x, y(x, t), t)^T]\rho(t)dt + \mathcal{N}_{\mathcal{X}}(x) \tag{51}$$

and view (51) as a limit of (50). It is not difficult to prove that when $\mathcal{T}(x)$ is a finite set and

$$\int_0^u \left|d(\nabla_x\Phi(x, y(x, t), t)^T\nabla_y\Phi(x, y(x, t), t)^{-T}\nabla_y f(x, y(x, t), t)^T]\rho(t)/dt\right| dt < \infty,$$

any accumulation point of sequence $\{x_K\}$, where x_K satisfies (50), is a KKT point satisfying (51). It seems, however, difficult to extend the conclusion to the general case. Observe also that the KKT condition (48) is sharper than that of (51), which means even if an accumulation point of sequence $\{x_K\}$ satisfies (51), it is not necessarily a KKT point of (48).

Smoothed program. Consider the smoothed program (40). Let x_μ be a KKT point of the program. We are interested in the convergence of sequence $\{x_\mu\}$ as $\mu \to 0$.

PROPOSITION 5.1. *Suppose that x_μ is a KKT point of* (34) *and x^* is an accumulation point of sequence $\{x_\mu\}$ as $\mu \to 0$. Then x^* is a KKT point of* (4).

Proof. By definition

$$0 \in \nabla_x\tilde{E}(x_\mu, \mu)^T + \mathcal{N}_{\mathcal{X}}(x_\mu).$$

By upper semicontinuity of the normal cone,

$$\overline{\lim_{\mu\to 0}}\ \mathcal{N}_{\mathcal{X}}(x_\mu) \subset \mathcal{N}_{\mathcal{X}}(x^*),$$

where $\overline{\lim}$ denotes the outer limit.

Note that if we treat μ as a variable, then $\tilde{E}(\cdot,\cdot)$ is continuously differentiable at any point (x, μ), where $\mu > 0$ and is locally Lipschitz continuous near point $(x, 0)$. For a set valued mapping $\mathcal{A} : \mathbb{R}^m \times \mathbb{R} \to 2^{\mathbb{R}^{(m+1)\times m}}$, we use $\Pi_x\mathcal{A}(x, \epsilon)$ to denote the set of all $m \times m$ matrices U such that, for some vector $V \in \mathbb{R}^m$, the $(m + 1) \times m$ matrix $[U^T, V]^T$ belongs to $\mathcal{A}(x, \epsilon)$. Using this notation, we have

$$\nabla_x\tilde{E}(x_\mu, \mu) = \Pi_x\nabla\tilde{E}(x_\mu, \mu).$$

By the definition of the Clarke generalized Jacobian

$$\overline{\lim_{\mu\to 0}}\ \nabla\tilde{E}(x_\mu, \mu) \subset \partial\tilde{E}(x^*, 0).$$

Hence

$$\overline{\lim_{\mu\to 0}}\ \nabla_x \tilde{E}(x_\mu,\mu) = \overline{\lim_{\mu\to 0}}\ \Pi_x \nabla \tilde{E}(x_\mu,\mu) \subset \Pi_x \partial \tilde{E}(x^*,0) = \partial \tilde{E}(x^*).$$

The last equality is due to the Jacobian consistency. This shows

$$0 \in \partial E(x^*)^T + \mathcal{N}_{\mathcal{X}}(x^*).$$

The proof is complete. ∎

Discretized smoothed program. Finally, we consider the discretized smoothed program (41) with $\mathcal{X} \subset \mathbb{R}^n$

$$\begin{array}{ll} \min & E_K(x,\mu) := \dfrac{u}{K}\displaystyle\sum_{l=0}^{K} f(x,y_l,t_l)\rho(t_l) \\ \text{s.t.} & x \in \mathcal{X}, \\ & \Psi(x,y_l,t_l,\mu) = 0,\ l = 1,\dots,K. \end{array} \tag{52}$$

The KKT condition of this program is

$$\begin{cases} 0 & \in \dfrac{u}{K}\displaystyle\sum_{l=1}^{K} \nabla_x f(x,y_l,t_l)^T \rho(t_l) + \sum_{l=1}^{K} \nabla_x \Phi(x,y_l,t_l,\mu)^T \lambda_l + \mathcal{N}_{\mathcal{X}}(x), \\ 0 & = \dfrac{u}{K} \nabla_{y_l} f(x,y_l,t_l)^T \rho(t_l) + \nabla_{y_l} \Phi(x,y_l,t_l,\mu)^T \lambda_l, l = 1,\dots,K, \end{cases} \tag{53}$$

equivalently,

$$0 \in \frac{u}{K}\sum_{l=1}^{K} \left[\nabla_x f(x,y_l,t_l)^T - \nabla_x \Phi(x,y_l,t_l,\mu)^T \nabla_{y_l}\Phi(x,y_l,t_l,\mu)^{-T} \nabla_{y_l} f(x,y_l,t_l,\mu)^T\right] \rho(t_l) + \mathcal{N}_{\mathcal{X}}(x)$$

Since y_l can be solved from $\Phi(x,y_l,t_l,\mu) = 0$, we can express y_l as $\tilde{y}(x,t_l,\mu)$. Thus we have

$$\begin{aligned} 0 \in \frac{u}{K}\sum_{l=1}^{K} &[\nabla_x f(x,\tilde{y}(x,t_l,\mu),t_l)^T \\ &- \nabla_x \Phi(x,\tilde{y}(x,t_l,\mu),t_l,\mu)^T \nabla_y \Phi(x,\tilde{y}(x,t_l,\mu),t_l,\mu)^{-T} \nabla_y f(x,\tilde{y}(x,t_l,\mu),t_l,\mu)^T]\rho(t_l) \\ &+ \mathcal{N}_{\mathcal{X}}(x). \end{aligned}$$

Driving K to ∞, we obtain

$$\begin{aligned} 0 \in \int_0^u &[\nabla_x f(x,\tilde{y}(x,t,\mu),t)^T \\ &- \nabla_x \Phi(x,\tilde{y}(x,t,\mu),t,\mu)^T \nabla_y \Phi(x,\tilde{y}(x,t,\mu),t,\mu)^{-T} \nabla_y f(x,\tilde{y}(x,t,\mu),t,\mu)^T]\rho(t)dt \\ &+ \mathcal{N}_{\mathcal{X}}(x). \end{aligned}$$

Driving μ to 0 and considering the strong Jacobian consistency of ψ, we obtain

$$0 \in \int_0^u \left[\nabla_x f(x,y(x,t),t)^T + \Im_\Phi(x,y(x,t),t)^T \nabla_y f(x,y(x,t),t)^T\right] \rho(t)dt + \mathcal{N}_{\mathcal{X}}(x).$$

From the discussion above, we can conclude that, from a KKT perspective, numerical methods based on the smoothed program (40) and the discretized smoothed program (41) may be more preferable.

Appendix.

Proof of Theorem 2.3. Part (i). Since F is uniformly strongly monotone in y, it is well known that the complementarity problem in (2) has a unique solution for all $t \in \mathcal{T}$ and $x \in \mathcal{X}$; see, for instance, [8, Corollary 3.2]. Thus, (3) has a unique solution for each $x \in \mathcal{X}$ and $t \in [0, T]$. Here we use Lemma 2.2. Under the assumption on F, the Clarke generalized Jacobian $\partial_y \Phi(x, y, t)$ is uniformly nonsingular. By Lemma 2.2, for $(\bar{x}, \bar{y}, \bar{u}) \in \mathcal{X} \times \mathbb{R}^n_+ \times \mathcal{T}$, there exists a Lipschitz continuous function $y(x, t)$ such that $y(\bar{x}, \bar{t}) = \bar{y}$, and (8) holds for (x, t) in a neighborhood of $(\bar{x}, \bar{t})$. The uniform monotonicity of F with respect to y allows the implicit function to be extended to the whole area $\mathcal{X} \times \mathcal{T}$.

Part (ii). Since Φ is piecewise smooth, by [29, Lemma 4.11], the implicit function $y(x, t)$ which is defined in part (ii) is piecewise smooth with respect to either x for fixed t or t for fixed x or both.

Part (iii). By part (ii) of Lemma 2.2,

$$\partial_x y(x, t) \subset \{-R^{-1}U : (U, R, V) \in \partial\Phi(x, y(x, t), t), U \in \mathbb{R}^{n\times m}, R \in \mathbb{R}^{n\times n}, V \in \mathbb{R}^{n\times l}\}.$$

This shows the first differential inclusion. The second inclusion is well known; see, for example, [3]. To show the uniform boundedness of $\partial_x y(x, t)$, we use the first differential inclusion. Thus it suffices to show the uniform boundedness of R^{-1} and U. Since F is uniformly strongly monotone, by Proposition 2.1, R^{-1} is uniformly bounded, and since F is uniformly Lipschitz continuous in x, U is uniformly bounded. The uniform global Lipschitz continuity of $y(x, t)$ in x follows subsequently.

Part (iv). We can show the differential inclusions as in Part (iii) by using Part (ii) of Lemma 2.2 with respect to y and t. To show the uniform boundedness of $\partial_t y(x, t)$, it suffices to show the uniform boundedness of R^{-1} and V. Since F is uniformly strongly monotone, by Proposition 2.1, R^{-1} is uniformly bounded, and since F is uniformly Lipschitz continuous in t, V is uniformly bounded. The uniform global Lipschitz continuity of $y(x, t)$ in t follows subsequently. □

Proof of Proposition 2.7. By Lemma 2.5, $\mathcal{T}(x)$ is Lebesgue measurable, and by Assumption 2.6, the Lebesgue measure of $\mathcal{T}(x)$ is zero.

Let $x' \in \mathcal{X}$ be any point close to x, let

$$\xi^T = \int_{\mathcal{T}\setminus\mathcal{T}(x)} [\nabla_x f(x, y(x, t), t) + \nabla_y f(x, y(x, t), t)\nabla_x y(x, t)]\rho(t)dt.$$

Let

$$R(x', x) = (E(x') - E(x) - \xi^T(x' - x))/\|x' - x\|.$$

Then

$$R(x', x) = R_1(x', x) + R_2(x', x) + R_3(x', x),$$

where

$$R_1(x', x) = \int_{\mathcal{T}\setminus\mathcal{T}(x)} \left[\frac{f(x', y(x', t), t) - f(x, y(x', t), t) - \nabla_x f(x, y(x, t), t)(x' - x)}{\|x' - x\|}\right]\rho(t)dt$$

and

$$R_2(x', x) = \int_{\mathcal{T}\setminus\mathcal{T}(x)} \left[\frac{f(x, y(x', t), t) - f(x, y(x, t), t) - \nabla_y f(x, y(x, t), t)\nabla_x y(x, t)(x' - x)}{\|x' - x\|}\right]\rho(t)dt$$

and

$$R_3(x', x) = \int_{\mathcal{T}(x)} \left[\frac{f(x', y(x', t), t) - f(x, y(x, t), t)}{\|x' - x\|} \right] \rho(t) dt.$$

Since f is continuously differentiable in x, it is obvious that $R_1(x', x) \to 0$ as $x' \to x$.

We now estimate R_3.

$$\begin{aligned} |R_3(x', x)| &\leq \int_{\mathcal{T}(x)} \frac{|f(x', y(x', t), t) - f(x, y(x, t), t)|}{\|x' - x\|} \rho(t) dt \\ &\leq \int_{\mathcal{T}(x)} \frac{L(t)(\|x' - x\| + \|y(x', t) - y(x, t)\|)}{\|x' - x\|} \rho(t) dt \\ &\leq (1 + C) \int_{\mathcal{T}(x)} L(t) \rho(t) dt \\ &= 0. \end{aligned}$$

The last equality is due to the fact that the Lebseque measure of $\mathcal{T}(x)$ is zero.

Finally, we estimate $R_2(x', x)$. By (12) and twice continuous differentiability of f, we have

$$f(x, y(x', t), t) - f(x, y(x, t), t) - \nabla_y f(x, y(x, t), t) \nabla_x y(x, t)(x' - x) = o(\|x' - x\|)$$

which implies

$$R_2(x', x) \to 0, \text{ as } x' \to x.$$

This shows

$$R(x', x) \to 0, \text{ as } x' \to x,$$

and hence (13).

Now we show the continuity of $\nabla E(\cdot)$,

$$\begin{aligned} \nabla_x E(x') - \nabla_x E(x) &= \int_{\mathcal{T} \setminus \mathcal{T}(x')} [\nabla_x f(x', y(x', t), t) + \nabla_y f(x', y(x', t), t) \nabla_x y(x', t)] \rho(t) dt \\ &\quad - \int_{\mathcal{T} \setminus \mathcal{T}(x)} [\nabla_x f(x, y(x, t), t) + \nabla_y f(x, y(x, t), t) \nabla_x y(x, t)] \rho(t) dt \\ &= \int_{\mathcal{T} \setminus \mathcal{T}(x') \cup \mathcal{T}(x)} [\nabla_x f(x', y(x', t), t) + \nabla_y f(x', y(x', t), t) \nabla_x y(x', t) \\ &\quad - \nabla_x f(x, y(x, t), t) + \nabla_y f(x, y(x, t), t) \nabla_x y(x, t)] \rho(t) dt \\ &\quad + \int_{\mathcal{T}(x) \setminus \mathcal{T}(x') \cap \mathcal{T}(x)} [\nabla_x f(x', y(x', t), t) \\ &\quad + \nabla_y f(x', y(x', t), t) \nabla_x y(x', t)] \rho(t) dt \\ &\quad - \int_{\mathcal{T}(x') \setminus \mathcal{T}(x') \cap \mathcal{T}(x)} [\nabla_x f(x, y(x, t), t) \\ &\quad + \nabla_y f(x, y(x, t), t) \nabla_x y(x, t)] \rho(t) dt \end{aligned}$$

We show that the three terms at the right-hand side of the last equality tends to zero as $x' \to x$. Since $\mathcal{T}(x') \to \mathcal{T}(x)$ as $x' \to x$, the Lebesgue measure of $\mathcal{T} \setminus \mathcal{T}(x') \cup \mathcal{T}(x)$ tends to that of $\mathcal{T}$. Moreover, ∇f is uniformly continuous in x, y by assumption, $y(x', t)$ uniformly approximates $y(x, t)$ by part (iv) of Theorem 2.3, and $\nabla y(x', t)$ uniformly approximates $\nabla y(x, t)$ by (12). This shows the first term tends to zero. The proofs for the second and third terms are similar. This completes the proof. ∎

Acknowledgments. I would like to thank Daniel Ralph for helpful comments and suggestions particularly regarding the optimality conditions. I would also like to thank Houyuan Jiang, Gui-Hua Lin, and two anonymous referees for careful reading of the paper and helpful comments which lead to a significant improvement of the paper.

REFERENCES

[1] X. Chen, L. Qi, and D. Sun, *Global and superlinear convergence of the smoothing Newton's method and its application to general box constrained variational inequalities*, Math. Comp., 67 (1998), pp. 519–540.

[2] S. Christiansen, M. Patriksson, and L. Wynter, *Stochastic Bilevel Programming in Structral Optimization*, preprint, PRISM, Université de Versailles, Versailles, France, 1999.

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

[4] D. De Wolf and Y. Smeers, *A Stochastic version of a Stackelberg–Nash–Cournot equilibrium model*, Management Sciences, 43 (1997), pp. 190–197.

[5] A. Evgrafov and M. Patriksson, *On the existence of solutions to stochastic mathematical programs with equilibrium constraints*, J. Optim. Theory Appl., 121 (2004), pp. 65–76.

[6] F. Facchinei, H. Jiang, and L. Qi, *A smoothing method for mathematical programs with equilibrium constraints*, Math. Program., 85 (1999), pp. 81–106.

[7] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, 2003.

[8] M. Ferris and J. S. Pang, *Engineering and economic applications of complementarity problems*, SIAM Rev., 39 (1997), pp. 669–713.

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

[10] R. Fletcher and S. Leyffer, *Numerical experience with solving MPECs as NLPs*, University of Dundee Report NA210, Dundee, UK, 2002.

[11] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes, *Local Convergence of SQP Methods for Mathematical Programs with Equilibrium Constraints*, University of Dundee Report NA209, Dundee, UK, 2002.

[12] M. Fukushima and J.-S. Pang, *Convergence of a smoothing continuation method for mathematical programs with complementarity constraints*, Ill-posed Variational Problems and Regularization Techniques (Trier 1998), Lecture Notes in Econom. Math. Systems, 477, Springer-Verlag, Berlin, 1999, pp. 99–110.

[13] M. Fukushima and P. Tseng, *An implementable active-set algorithm for computing a B-stationary point of a mathematical program with linear complementarity constraints*, SIAM J. Optim., 12 (2002), pp. 724–739.

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

[15] X. Hu and D. Ralph, *A note on sensitivity of value functions of mathematical programs with complementarity constraints*, Math. Program., 93 (2002), pp. 265–279.

[16] H. Jiang and D. Ralph, *Smooth SQP methods for mathematical programs with nonlinear complementarity constraints*, SIAM J. Optim., 10 (2000), pp. 779–808.

[17] H. Jiang and L. Qi, *A new nonsmooth equations approach to nonlinear complementarity problems*, SIAM J. Control Optim., 35 (1997), pp. 178–193.

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

[19] S. Leyffer, *Mathematical Programs with Complementarity Constraints*, SIAG/OPT Views-and-News, 14 (2003), pp. 15–18.

[20] G.-H. Lin, X. Chen, and M. Fukushima, *Smoothing Implicit Programming Approaches for Stochastic Mathematical Programs with Linear Complementarity Constraints*, http://www.amp.i.kyoto-u.ac.jp/tecrep/index-e.html, 2003.

[21] G.-H. Lin, X. Chen, and M. Fukushima, *Combined smoothing implicit programming and penalty method for stochastic method for stochastic mathematical programs with equilibrium constraints*, Kyoto University, Japan, 2004, preprint.

[22] G.-H Lin and M. Fukushima, *A class of stochastic mathematical programs with complementarity constraints: reformulations and algorithms*, J. Ind. Manag. Optim., 1 (2005), pp. 99–122.

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

[24] F. H. Murphy, H. D. Sherali, and A. L. Soyster, *A mathematical programming approach for determining oligopolistic market equilibruium*, Math. Program., 24 (1982), pp. 92–106.
[25] J. Outrata, M. Kocvara, and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Theory, Applications, and Numerical Constraints, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
[26] M. Patriksson and L. Wynter, *Stochastic mathematical programs with equilibrium constraints*, Oper. Res. Letters, 25 (1999), pp. 159–167.
[27] M. Patriksson and L. Wynter, *Stochastic mathematical programs with equilibrium constraints*, Oper. Res. Lett., 25(1999), pp. 159–167.
[28] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
[29] D. Ralph and H. Xu, *Implicit smoothing and its application to optimization with piecewise smooth equality constraints*, J. Optim. Theory Appl., 124 (2005), pp. 673–699.
[30] E. Polak, D. Q. Mayne, and Y. Wardi, *On the extension of constrained optimization algorithms from differentiable to nondifferentiable problems*, SIAM J. Control Optim., 21 (1983), pp. 179–203.
[31] A. Shapiro, *Stochastic mathematical programs with equilibrium constraints*, , School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 2004, preprint. To appear in J. Optim. Theory Appl.
[32] H. D. Sherali, A. L. Soyster, and F. H. Murphy, *Stackelberg–Nash–Cournot equilibria: Characterizations and computations*, Oper. Res., 31 (1983), pp. 253–276.
[33] H. D. Sherali, *A multiple leader Stackelberg model and analysis*, Oper. Res., 32 (1984), pp. 390–404.
[34] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM J. Optim., 11 (2001), pp. 918–936.
[35] D. Sun and L. Qi, *On NCP functions*, Comput. Optim. Appl., 13 (1999), pp. 201–220.
[36] H. Xu, *An MPCC approach for stochastic Stackelberg–Nash–Cournot equilibrium*, Optimization, 54 (2005), pp. 27–57.
[37] H. Xu and X. W. Chang, *Approximate Newton methods for nonsmooth equations*, J. Optim. Theory Appl., 93 (1997), pp. 373–394.