

Solutions of Two-stage Stochastic Minimax Problems

Hailin Sun · Xiaojun Chen

Received: date / Accepted: date

Abstract This paper introduces a class of two-stage stochastic minimax problems where the first-stage objective function is nonconvex-concave while the second-stage objective function is strongly convex-concave. We establish properties of the second-stage minimax value function and solution functions, and characterize the existence and relationships among saddle points, minimax points, and KKT points. We apply the sample average approximation (SAA) to the class of two-stage stochastic minimax problems and prove the convergence of the KKT points as the sample size tends to infinity. An inexact parallel proximal gradient descent ascent algorithm is proposed to solve this class of problems with the SAA. Numerical experiments demonstrate the effectiveness of the proposed algorithm and validate the convergence properties of the SAA approach.

Keywords Two-stage stochastic minimax problem · Nonconvex-nonsmooth · Saddle point · Sample average approximation · Proximal gradient method

Mathematics Subject Classification (2020) 90C15 · 49K35 · 90C47

1 Introduction

In this paper, we consider the following two-stage stochastic minimax problem

$$\min_{x_1 \in X_1} \max_{y_1 \in Y_1} \psi(x_1, y_1) := F_1(x_1, y_1) + \mathbb{E}[\psi_2(x_1, y_1, \xi)], \quad (1)$$

H. Sun

School of Mathematical Sciences, Nanjing Normal University, Nanjing, 210023, China

E-mail: hlsun@njnu.edu.cn

X. Chen

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China

Corresponding author. E-mail: xiaojun.chen@polyu.edu.hk

where

$$\psi_2(x_1, y_1, \xi) := \min_{x_2 \in X_2(x_1, \xi)} \max_{y_2 \in Y_2(y_1, \xi)} F_2(x_2, y_2, \xi), \quad (2)$$

$\mathbb{E}$ denotes the expectation, $\xi : \Omega \rightarrow \Xi \subset \mathbb{R}^l$ is a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$, $X_1 \subset \mathbb{R}^{n_1}$, $Y_1 \subset \mathbb{R}^{m_1}$ are convex compact sets, $F_1 : \mathbb{R}^{n_1} \times \mathbb{R}^{m_1} \rightarrow \mathbb{R}$, $F_2 : \mathbb{R}^{n_2} \times \mathbb{R}^{m_2} \times \mathbb{R}^l \rightarrow \mathbb{R}$,

$$X_2(x_1, \xi) := \{x_2 \in \mathbb{R}^{n_2} : T(\xi)x_1 + W(\xi)x_2 \leq h(\xi)\},$$

$$Y_2(y_1, \xi) := \{y_2 \in \mathbb{R}^{m_2} : A(\xi)y_1 + B(\xi)y_2 \leq c(\xi)\},$$

and $T(\xi) \in \mathbb{R}^{l_2 \times n_1}$, $W(\xi) \in \mathbb{R}^{l_2 \times n_2}$, $A(\xi) \in \mathbb{R}^{s_2 \times m_1}$, $B(\xi) \in \mathbb{R}^{s_2 \times m_2}$, $h(\xi) \in \mathbb{R}^{l_2}$, $c(\xi) \in \mathbb{R}^{s_2}$ for all $\xi \in \Xi$. The objective functions in (1) and (2) have the following structure.

- $F_1(x_1, y_1) := f(x_1) + \psi_1(x_1, y_1) - g(y_1)$, where $f : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$ is a proper and lower semicontinuous (lsc) function, $\psi_1 : \mathbb{R}^{n_1} \times \mathbb{R}^{m_1} \rightarrow \mathbb{R}$ is a Lipschitz continuously differentiable function, $g : \mathbb{R}^{m_1} \rightarrow \mathbb{R}$ is a proper and lsc function, and for every $x_1 \in X_1$, $g(\cdot) - \psi_1(x_1, \cdot)$ is σ -strongly convex with $\sigma > 0$.
- For almost every (a.e.) $\xi \in \Xi$, $F_2(\cdot, \cdot, \xi)$ is a σ -strongly convex-strongly concave function.

The two-stage stochastic minimax problem (1)-(2) represents an extension of the classical two-stage stochastic minimization model originally developed by Dantzig [13] and Beale [1]. Two-stage stochastic minimization problem [3, 34] is a fundamental framework for sequential decision-making under uncertainty, where first-stage “here-and-now” decisions (e.g., infrastructure investments) must be made before observing random outcomes, followed by second-stage “wait-and-see” recourse actions (e.g., operational adjustments) that adapt to the realized uncertainty. Due to its modeling power, the two-stage stochastic minimization problem has been widely used in many important engineering and scientific applications, such as wireless resource optimization [25], transportation network design [26], and machine/deep learning [23].

In contrast to stochastic optimization’s expectation-based framework, the minimax approach explicitly considers worst-case scenarios to achieve robust solutions. The study of minimax problems can be traced back to von Neumann’s seminal work [27] in 1928 on convex-concave deterministic minimax problems. In recent years, nonconvex-nonsmooth minimax problems [2, 12, 20, 21, 42] and stochastic minimax problems [6, 22, 35] have garnered significant attention due to their applications in data science, machine learning, game theory, and robust decision-making. Jin et al. [21] gave the definitions of global minimax points and local minimax points by considering the minimax problem as a two-player sequential game. Cohen and Teboulle [12] analyzed proximal gradient methods for nonconvex and strongly concave minimax problems. Chen and Luo [6] proposed a recursive anchored iteration method for smooth minimax problems and proved that their algorithm achieves near-stationarity.

To the best of our knowledge, the two-stage stochastic minimax problem has not been investigated, although two-stage stochastic minimization and minimax problem have been extensively studied as two separate classes of mathematical models. However, many practical scenarios involve stochastic environments that combine both risk-neutral and risk-averse components, requiring decision-makers to simultaneously address: (i) sequential decision-making, (ii) stochastic uncertainties, and (iii) adversarial considerations (extreme events or opponent behaviors). This constitutes the primary motivation for studying two-stage stochastic minimax problems (see Example 1).

Another motivation for studying the two-stage stochastic minimax problem (1)-(2) lies in its capacity to model two-stage stochastic two-player zero-sum games (see Example 2), which constitute a specialized subclass of two-stage stochastic Nash equilibrium problems (SNEPs). Pang et al. [29] examined two-stage SNEPs involving risk-averse players under uncertainty, developing an iterative best-response solution framework. Zhang et al. [43] subsequently investigated a mixed non-cooperative game formulation for two-stage decision-making in uncertain environments. Further advancing this line of research, Lei et al. [24] introduced synchronous, asynchronous, and randomized best-response schemes for SNEPs, with specific applications to two-stage SNEPs featuring both linear and quadratic recourse structures.

The KKT conditions for two-stage SNEPs can be characterized through two-stage stochastic variational inequalities (SVIs). Rockafellar and Wets [33] and Chen et al. [7] extended SVIs from single-stage to multi-stage and two-stage SVIs, respectively. Rockafellar and Sun [31] studied the progressive hedging algorithm (PHA) for solving multi-stage SVIs when the random variable follows a discrete distribution. Discrete approximation methods have been proposed to approximate two-stage SVIs [9, 8], allowing the approximated SVIs to be solved using PHA. In addition to PHA, dynamic stochastic approximation-type algorithms effectively solve two-stage SVIs [5].

However, most existing research on two-stage SNEPs and two-stage SVIs has been confined to smooth and monotone cases, leaving further developments for nonsmooth and nonconvex problems unexplored.

The main contributions of this paper are summarized as follows.

- We introduce a two-stage stochastic minimax model (1)-(2) and investigate the properties of the second-stage minimax value functions and solution functions. Based on these properties, we examine the existence and relationships among saddle points, minimax points, and KKT points for the nonconvex-nonsmooth two-stage stochastic minimax problem (1)-(2).
- We apply the sample average approximation (SAA) method to problem (1)-(2) and prove that the divergence between the KKT point sets of the SAA problem and the true problem (1)-(2) converges to zero almost surely.
- We propose an Inexact Parallel Proximal Gradient Descent Ascent (IPPGDA) algorithm for solving problem (1)-(2), with both subsequence and global convergence analyses. Unlike the single-stage minimax problems solved by the parallel proximal gradient descent-ascent (PPGDA) algorithm in [12],

the inexactness in our approach arises not only from solving the inner maximization problem but also from solving the second-stage minimax problem.

The remainder of this paper is organized as follows. Section 2 introduces the motivating examples, necessary assumptions, and fundamental concepts. Section 3 investigates the properties of problem (1)-(2), analyzing the existence and relationships among saddle points, minimax points and KKT points. Section 4 examines the SAA method for problem (1)-(2), including its convergence properties. In section 5, we present IPPGDA algorithm for solving problem (1)-(2), along with its subsequence and global convergence analysis, and demonstrate the effectiveness of our proposed algorithm and the convergence behavior of the SAA method through numerical experiments. Section 6 concludes the paper.

Notation For a locally Lipschitz continuous function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\partial f(\bar{x})$ denotes the Clarke subdifferential [11] of f at point $\bar{x}$. For a Lipschitz continuously differentiable function $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $\partial^2 g(\bar{x}, \bar{y})$ represents the Clarke generalized Hessian of g at point $(\bar{x}, \bar{y})$, while $\partial_{xx}^2 g(\bar{x}, \bar{y})$ and $\partial_{yy}^2 g(\bar{x}, \bar{y})$ denote the Clarke generalized Hessians of g with respect to (w.r.t.) x and y at point $(\bar{x}, \bar{y})$. For a vector-valued function $H : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, $\partial_x H(\bar{x}, \bar{y})$ denotes the Clarke generalized Jacobian of H w.r.t. x at point $(\bar{x}, \bar{y})$. A function $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is said (strongly) convex-(strongly) concave if $g(\cdot, y)$ is (strongly) convex in $\mathbb{R}^n$ for any fixed $y \in \mathbb{R}^m$, and $g(x, \cdot)$ is (strongly) concave in $\mathbb{R}^m$ for any fixed $x \in \mathbb{R}^n$. For $a \in \mathbb{R}^n$, $\|a_+\|_0 := \sum_{i=1}^n (\max\{a_i, 0\})^0$ with $0^0 = 0$, where a_i is the i th element of a . For $x \in \mathbb{R}^n$ and $X, Y \subseteq \mathbb{R}^n$, $\mathbb{D}(x, Y) := \inf_{y \in Y} \|x - y\|$ and $\mathbb{D}(X, Y) := \sup_{x \in X} \inf_{y \in Y} \|x - y\|$. Let $\mathcal{L}_p^n$ denote the Lebesgue space of measurable functions from a measure space to $\mathbb{R}^n$ with finite L^p -norm, where $p \geq 2$ and for any measurable function $\mathbf{x} \in \mathcal{L}_p^n$, the L^p -norm is defined as $\|\mathbf{x}\|_p := (\int_{\Omega} \sum_{i=1}^n |(\mathbf{x}(\xi(\omega)))_i|^p P(d\omega))^{1/p}$. For a convex set $X \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$, $\mathcal{N}_X(x)$ denotes the normal cone to X at x .

2 Examples, assumptions and concepts

In this section, we provide two motivating examples for the study of the two-stage stochastic minimax problem (1)-(2). Additionally, we introduce several necessary assumptions and concepts required for the theoretical analysis in this paper.

The first example is a renewable energy storage scheduling problem in a stochastic environment that combines both risk-neutral and risk-averse criteria.

Example 1 [Renewable energy storage scheduling]: A microgrid operator selects energy storage capacity $x \geq 0$ in the first stage to minimize the total cost, which comprises two components: (1) The investment cost $C_{\text{inv}}(x, z)$, subject to an adversarial uncertainty $z \in Z$ representing risks like volatile material prices or uncertain government subsidies; (2) The expected future losses under renewable generation uncertainty ξ and adversarial electricity prices

$p \in \mathcal{P}(\xi) \subset \mathbb{R}_+^{24}$, where the uncertainty set expands price bounds proportionally to the renewable forecast error ξ .

In the second stage, after observing ξ , the operator adjusts charge/discharge decisions y_t^{ch} and y_t^{dis} ($t = 1, \dots, 24$) subject to storage constraints: power limits $0 \leq y_t^{\text{ch}}, y_t^{\text{dis}} \leq 0.2x$, state-of-charge dynamics $\text{SOC}_t = \text{SOC}_{t-1} + 0.9y_t^{\text{ch}} - y_t^{\text{dis}}/0.9$ with $0 \leq \text{SOC}_t \leq x$, and boundary conditions $\text{SOC}_0 = \text{SOC}_{24} = 0$. We denote by $Y(x)$ the feasible region of $y := (y_t^{\text{ch}}, y_t^{\text{dis}})_{t=1}^{24}$ satisfying above constraints. Moreover, the second-stage objective function is

$$Q(y, p, \xi) := \sum_t \left[p_t(y_t^{\text{dis}} - y_t^{\text{ch}}) + 0.1(y_t^{\text{ch}} + y_t^{\text{dis}}) \right] + 100 \left(\xi + \sum_t (y_t^{\text{dis}} - y_t^{\text{ch}}) \right)^2,$$

which captures real-time market costs, battery degradation penalties, and renewable energy imbalance penalties. The complete formulation, which integrates both stages, is a two-stage stochastic minimax problem as follows:

$$\min_{x \geq 0} \max_{z \in Z} \left(C_{\text{inv}}(x, z) + \mathbb{E}_\xi \left[\min_{y \in Y(x)} \max_{p \in \mathcal{P}(\xi)} Q(y, p, \xi) \right] \right).$$

The second example concerns a two-stage stochastic two-player zero-sum game.

Example 2 [Two-stage stochastic two-player zero-sum game]: The two-player zero-sum game is a basic model in game theory [40]. There are two players, each with an associated set of strategies. While one player aims to maximize her payoff, the other player attempts to take action to minimize this payoff.

The two-stage stochastic two-player zero-sum game is a two-stage stochastic extension of the two-player zero-sum game. There are two players in the game, player 1's problem is

$$\min_{x_1 \in X_1} F_1(x_1, y_1) + \mathbb{E} \left[\min_{x_2 \in X_2(x_1, \xi)} F_2(x_2, y_2, \xi) \right] \quad (3)$$

and player 2's problem is

$$\min_{y_1 \in Y_1} -F_1(x_1, y_1) + \mathbb{E} \left[\min_{y_2 \in Y_2(y_1, \xi)} -F_2(x_2, y_2, \xi) \right]. \quad (4)$$

When F_1 and F_2 are convex w.r.t. x_1 and x_2 , respectively, and concave w.r.t. y_1 and y_2 , respectively, and when X_1 , Y_1 , $X_2(x_1, \xi)$ and $Y_2(y_1, \xi)$ are convex and compact sets, then (3)-(4) is equivalent to problem (1)-(2).

To investigate the two-stage stochastic minimax problem (1)-(2), we need the following assumptions.

Assumption 1 For every $(x_1, y_1) \in X_1 \times Y_1$ and $\xi \in \Xi$,

- (i) F_2 is continuous, and $F_2(\cdot, \cdot, \xi)$ is Lipschitz continuously differentiable. Moreover, $\nabla_{x_2} F_2$ and $\nabla_{y_2} F_2$ are Lipschitz continuous;
- (ii) $X_2(x_1, \xi)$ and $Y_2(y_1, \xi)$ are nonempty;

(iii) $B(\xi)$ and $W(\xi)$ are of full row rank.

Remark 1 Assumption 1 (iii) and the strong convexity-strong concavity of $F_2(\cdot, \cdot, \xi)$ play different roles in our analysis. The full row rank condition on $W(\xi)$ and $B(\xi)$ guarantees the uniqueness of Lagrange multipliers in the smooth second-stage KKT systems associated with fixed primal solutions. The strong convexity-strong concavity assumption further ensures the uniqueness of the second-stage saddle point and the single-valuedness and stability of the corresponding primal-dual solution mapping. These properties will be used below to derive the differentiability of the second-stage value functions, which forms the basis of the subsequent algorithmic analyses.

The full row rank condition is also related to standard recourse regularity conditions in two-stage stochastic programming. In particular, under fixed recourse, it reduces to checking the rank of a deterministic recourse matrix; under the complete recourse condition, the recourse matrix has full row rank; see, e.g., [34, Section 2.1].

Assumption 2 The support set $\Xi \subset \mathbb{R}^l$ of the random vector ξ is compact. The random matrices $A(\cdot)$, $B(\cdot)$, $T(\cdot)$, $W(\cdot)$ and random vectors $c(\cdot)$, $h(\cdot)$ are continuous.

Remark 2 In the case when Ξ is unbounded, under the tightness of random variable ξ , for any $\epsilon \in (0, 1)$ there exists a compact subset $\bar{\Xi} \subset \Xi$ such that $\Pr\{\xi \in \bar{\Xi}\} \geq 1 - \epsilon$. Then we may omit the $\xi \in \Xi \setminus \bar{\Xi}$ and consider the compact support $\bar{\Xi}$ in our problem.

We provide the definitions of the following concepts for the two-stage stochastic minimax problem (1)-(2): saddle point, local saddle point, global minimax point, and local minimax point.

Definition 1 A point $(x_1^*, y_1^*) \in X_1 \times Y_1$ is called a saddle point of problem (1)-(2) if for all $(x_1, y_1) \in X_1 \times Y_1$, it holds

$$\psi(x_1^*, y_1) \leq \psi(x_1^*, y_1^*) \leq \psi(x_1, y_1^*). \quad (5)$$

We call $(x_1^*, y_1^*) \in X_1 \times Y_1$ a local saddle point of problem (1)-(2), if there exists a $\delta > 0$ such that (5) holds for all $(x_1, y_1) \in (X_1 \times Y_1) \cap \mathcal{B}((x_1^*, y_1^*), \delta)$, where $\mathcal{B}((x_1^*, y_1^*), \delta)$ denotes the neighborhood of (x_1^*, y_1^*) with radius δ .

Definition 2 A pair of mappings $(\bar{x}_2 : X_1 \times Y_1 \times \Xi \rightarrow \mathbb{R}^{n_2}, \bar{y}_2 : X_1 \times Y_1 \times \Xi \rightarrow \mathbb{R}^{m_2})$ is called a saddle point mapping of the second-stage minimax problem in (2), if $(x_2^*, y_2^*) := (\bar{x}_2(x_1, y_1, \xi), \bar{y}_2(x_1, y_1, \xi))$ is a saddle point of the minimax problem in (2), namely, for any $y_2 \in Y_2(y_1, \xi)$ and $x_2 \in X_2(x_1, \xi)$,

$$F_2(x_2^*, y_2, \xi) \leq F_2(x_2^*, y_2^*, \xi) \leq F_2(x_2, y_2^*, \xi).$$

Definition 3 A point $(\bar{x}_1, \bar{y}_1) \in X_1 \times Y_1$ is a global minimax point of problem (1)-(2), if

$$\psi(\bar{x}_1, y_1) \leq \psi(\bar{x}_1, \bar{y}_1) \leq \max_{y_1' \in Y_1} \psi(x_1, y_1') \quad (6)$$

holds for any $(x_1, y_1) \in X_1 \times Y_1$.

Definition 4 A point $(\bar{x}_1, \bar{y}_1) \in X_1 \times Y_1$ is a local minimax point of problem (1)-(2), if there exists $\delta_0 > 0$, such that for any $\delta \in (0, \delta_0]$ and any $(x_1, y_1) \in X_1 \times Y_1$ satisfying $\|x_1 - \bar{x}_1\| \leq \delta$ and $\|y_1 - \bar{y}_1\| \leq \delta$, we have

$$\psi(\bar{x}_1, y_1) \leq \psi(\bar{x}_1, \bar{y}_1) \leq \max_{y'_1 \in Y_1} \psi(x_1, y'_1).$$

Let $\bar{\psi}(x_1) := \max_{y_1 \in Y_1} \psi(x_1, y_1)$. Consider the first-stage minimization problem

$$\min_{x_1 \in X_1} \bar{\psi}(x_1). \quad (7)$$

Remark 3 Note that the definition of a local minimax point in Definition 4 implies that $\bar{y}_1$ is a local maximum point of $\psi(\bar{x}_1, \cdot)$ (since $\psi(\bar{x}_1, \cdot)$ is strongly concave, $\bar{y}_1$ is also a global maximum point of $\psi(\bar{x}_1, \cdot)$), and $\bar{x}_1$ is a local minimum of $\bar{\psi}(\cdot)$.

Moreover, compared with [21, Definition 14], Definition 4 uses the global maximization term $\max_{y'_1 \in Y_1} \psi(x_1, y'_1)$, whereas [21, Definition 14] uses a localized maximization term over $\{y'_1 \in Y_1 : \|y'_1 - \bar{y}_1\| \leq h(\delta)\}$ for some function $h(\delta) \rightarrow 0$ as $\delta \downarrow 0$. In general nonconcave settings, these two definitions are not equivalent.

Under our assumptions, $\psi(x_1, \cdot)$ is strongly concave for every $x_1 \in X_1$. Hence

$$y_1^*(x_1) := \arg \max_{y_1 \in Y_1} \psi(x_1, y_1)$$

is unique and continuous in x_1 . Therefore, for

$$h(\delta) := \sup_{\|x_1 - \bar{x}_1\| \leq \delta} \|y_1^*(x_1) - \bar{y}_1\|,$$

we have $h(\delta) \rightarrow 0$ as $\delta \downarrow 0$, and the global maximizer $y_1^*(x_1)$ is contained in the localized neighborhood used in [21, Definition 14]. Thus, for this admissible choice of $h(\delta)$, Definition 4 is equivalent to [21, Definition 14].

We also consider the following minimax problem

$$\min_{(x_1, \mathbf{x}_2) \in \mathbf{X}} \max_{(y_1, \mathbf{y}_2) \in \mathbf{Y}} F_1(x_1, y_1) + \mathbb{E}[F_2(\mathbf{x}_2(\xi), \mathbf{y}_2(\xi), \xi)], \quad (8)$$

where

$$\mathbf{X} := \{(x_1, \mathbf{x}_2) \in \mathbb{R}^{n_1} \times \mathcal{L}_p^{n_2} : x_1 \in X_1, T(\xi)x_1 + W(\xi)\mathbf{x}_2(\xi) \leq h(\xi), \text{ a.e. } \xi \in \Xi\}$$

and

$$\mathbf{Y} := \{(y_1, \mathbf{y}_2) \in \mathbb{R}^{m_1} \times \mathcal{L}_p^{m_2} : y_1 \in Y_1, A(\xi)y_1 + B(\xi)\mathbf{y}_2(\xi) \leq c(\xi), \text{ a.e. } \xi \in \Xi\}.$$

We will consider the relationship between (8) and the two-stage stochastic minimax problem (1)-(2) in Section 3.1. Moreover, the definition of a saddle point, local saddle point, global minimax point and local minimax point for the minimax problem (8) is as follows.

Definition 5 A point $(x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*) \in \mathbf{X} \times \mathbf{Y}$ is called a saddle point of the minimax problem (8) if for any $(x_1, \mathbf{x}_2, y_1, \mathbf{y}_2) \in \mathbf{X} \times \mathbf{Y}$, we have

$$\begin{aligned} F_1(x_1^*, y_1) + \mathbb{E}[F_2(\mathbf{x}_2^*(\xi), \mathbf{y}_2(\xi), \xi)] &\leq F_1(x_1^*, y_1^*) + \mathbb{E}[F_2(\mathbf{x}_2^*(\xi), \mathbf{y}_2^*(\xi), \xi)] \\ &\leq F_1(x_1, y_1^*) + \mathbb{E}[F_2(\mathbf{x}_2(\xi), \mathbf{y}_2^*(\xi), \xi)]. \end{aligned} \quad (9)$$

We call $(x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*) \in \mathbf{X} \times \mathbf{Y}$ a local saddle point of the minimax problem (8) if there exists a $\delta > 0$ such that (9) holds for all $(x_1, \mathbf{x}_2, y_1, \mathbf{y}_2) \in (\mathbf{X} \times \mathbf{Y}) \cap \mathcal{B}((x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*), \delta)$, where $\mathcal{B}((x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*), \delta)$ denotes the neighborhood of $(x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*)$ with radius δ , that is $\mathcal{B}((x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*), \delta) = \{(x_1, \mathbf{x}_2, y_1, \mathbf{y}_2) \in \mathbb{R}^{n_1} \times \mathcal{L}_p^{n_2} \times \mathbb{R}^{m_1} \times \mathcal{L}_p^{m_2} : \|x_1 - x_1^*\|_2 + \|\mathbf{x}_2 - \mathbf{x}_2^*\|_p + \|y_1 - y_1^*\|_2 + \|\mathbf{y}_2 - \mathbf{y}_2^*\|_p \leq \delta\}$.

Definition 6 A point $(\tilde{x}_1, \tilde{\mathbf{x}}_2, \tilde{y}_1, \tilde{\mathbf{y}}_2) \in \mathbf{X} \times \mathbf{Y}$ is called a global minimax point of the minimax problem (8) if

$$\begin{aligned} F_1(\tilde{x}_1, y_1) + \mathbb{E}[F_2(\tilde{\mathbf{x}}_2(\xi), \mathbf{y}_2(\xi), \xi)] &\leq F_1(\tilde{x}_1, \tilde{y}_1) + \mathbb{E}[F_2(\tilde{\mathbf{x}}_2(\xi), \tilde{\mathbf{y}}_2(\xi), \xi)] \\ &\leq \max_{(y_1', \mathbf{y}_2'(\xi)) \in \mathbf{Y}} F_1(x_1, y_1') + \mathbb{E}[F_2(\mathbf{x}_2(\xi), \mathbf{y}_2'(\xi), \xi)] \end{aligned} \quad (10)$$

holds for any $(x_1, \mathbf{x}_2, y_1, \mathbf{y}_2) \in \mathbf{X} \times \mathbf{Y}$.

Moreover, a point $(\tilde{x}_1, \tilde{\mathbf{x}}_2, \tilde{y}_1, \tilde{\mathbf{y}}_2) \in \mathbf{X} \times \mathbf{Y}$ is called a local minimax point of the minimax problem (8), if there exists $\delta_0 > 0$, such that for any $\delta \in (0, \delta_0]$ and any $(x_1, \mathbf{x}_2, y_1, \mathbf{y}_2) \in \mathbf{X} \times \mathbf{Y}$ satisfying $\|x_1 - \tilde{x}_1\| + \|\mathbf{x}_2 - \tilde{\mathbf{x}}_2\|_p \leq \delta$ and $\|y_1 - \tilde{y}_1\| + \|\mathbf{y}_2 - \tilde{\mathbf{y}}_2\|_p \leq \delta$, we have

$$\begin{aligned} F_1(\tilde{x}_1, y_1) + \mathbb{E}[F_2(\tilde{\mathbf{x}}_2(\xi), \mathbf{y}_2(\xi), \xi)] &\leq F_1(\tilde{x}_1, \tilde{y}_1) + \mathbb{E}[F_2(\tilde{\mathbf{x}}_2(\xi), \tilde{\mathbf{y}}_2(\xi), \xi)] \\ &\leq \max_{(y_1', \mathbf{y}_2'(\xi)) \in \mathbf{Y}} F_1(x_1, y_1') + \mathbb{E}[F_2(\mathbf{x}_2(\xi), \mathbf{y}_2'(\xi), \xi)]. \end{aligned}$$

3 Properties of two-stage stochastic minimax problems

In this section, we investigate the properties of minimax value functions, saddle points and minimax points of problem (1)-(2).

For given $\xi \in \Xi$ and $(x_1, y_1) \in X_1 \times Y_1$, we first investigate the properties of optimal value functions

$$f_{21}(x_2, y_1, \xi) := \max_{y_2 \in Y_2(y_1, \xi)} F_2(x_2, y_2, \xi), \quad (11)$$

$$f_{22}(x_1, y_2, \xi) := \min_{x_2 \in X_2(x_1, \xi)} F_2(x_2, y_2, \xi), \quad (12)$$

and minimax value function $\psi_2(x_1, y_1, \xi)$. Note that the minimax problem in (2) is a strongly convex-strongly concave minimax problem with linear constraints for any (x_1, y_1, ξ) . Then the KKT condition of the minimax problem

in (2) with given $(x_1, y_1) \in X_1 \times Y_1$ can be stated as the following system of nonsmooth equations in the variable $\mu = (x_2, y_2, \pi_{x_2}, \pi_{y_2})$:

$$H(\mu, \xi) = \begin{pmatrix} \nabla_{x_2} F_2(x_2, y_2, \xi) + W(\xi)^\top \pi_{x_2} \\ -\nabla_{y_2} F_2(x_2, y_2, \xi) + B(\xi)^\top \pi_{y_2} \\ \min(\pi_{x_2}, h(\xi) - T(\xi)x_1 - W(\xi)x_2) \\ \min(\pi_{y_2}, c(\xi) - A(\xi)y_1 - B(\xi)y_2) \end{pmatrix} = 0, \quad (13)$$

where $\pi_{x_2} \in \mathbb{R}^{l_2}$ and $\pi_{y_2} \in \mathbb{R}^{s_2}$ are corresponding Lagrange multipliers, and “min” denotes the component-wise minimum operator on a pair of vectors.

Lemma 1 *Under Assumptions 1-2, the following statements hold.*

- (i) *For any given $\xi \in \Xi$, $f_{21}(\cdot, \cdot, \xi)$ is continuously differentiable and strongly convex-concave; moreover, $\nabla_{y_1} f_{21}$ and $\nabla_{x_2} f_{21}$ are Lipschitz continuous w.r.t. (x_2, y_1) and continuous w.r.t. ξ .*
- (ii) *For any given $\xi \in \Xi$, $f_{22}(\cdot, \cdot, \xi)$ is continuously differentiable and convex-strongly concave; moreover, $\nabla_{x_1} f_{22}$ and $\nabla_{y_2} f_{22}$ are Lipschitz continuous w.r.t. (x_1, y_2) and continuous w.r.t. ξ .*
- (iii) *For any given $\xi \in \Xi$, $\psi_2(\cdot, \cdot, \xi)$ is convex-concave and continuously differentiable. And there exist $\pi_{x_2} : X_1 \times Y_1 \times \Xi \rightarrow \mathbb{R}^{l_2}$ and $\pi_{y_2} : X_1 \times Y_1 \times \Xi \rightarrow \mathbb{R}^{s_2}$ such that*

$$(\bar{x}_2(x_1, y_1, \xi), \bar{y}_2(x_1, y_1, \xi), \pi_{x_2}(x_1, y_1, \xi), \pi_{y_2}(x_1, y_1, \xi))$$

satisfies the KKT condition of the minimax problem in (2), where

$$\nabla_{x_1} \psi_2(\cdot, \cdot, \xi) = T(\xi)^\top \pi_{x_2}(\cdot, \cdot, \xi), \quad \nabla_{y_1} \psi_2(\cdot, \cdot, \xi) = -A(\xi)^\top \pi_{y_2}(\cdot, \cdot, \xi),$$

and $(\bar{x}_2(x_1, y_1, \xi), \bar{y}_2(x_1, y_1, \xi))$ is the unique saddle point of the minimax problem in (2). Moreover, $(\bar{x}_2, \bar{y}_2, \pi_{x_2}, \pi_{y_2})$ is Lipschitz continuous w.r.t. (x_1, y_1) over $X_1 \times Y_1$ and continuous w.r.t. ξ .

- (iv) *$(\bar{x}_2(x_1, y_1, \xi), \bar{y}_2(x_1, y_1, \xi))$ is contained in a convex and compact set $\hat{X}_2 \times \hat{Y}_2$ for all $(x_1, y_1, \xi) \in X_1 \times Y_1 \times \Xi$. Moreover, there exist convex and compact sets $\tilde{X}_2$ and $\tilde{Y}_2$ such that $\hat{X}_2 \subset \tilde{X}_2$, $\hat{Y}_2 \subset \tilde{Y}_2$,*

$$\max_{y_2 \in Y_2(y_1, \xi) \cap \tilde{Y}_2} F_2(x_2, y_2, \xi) = f_{21}(x_2, y_1, \xi), \quad \forall (x_2, y_1, \xi) \in \hat{X}_2 \times Y_1 \times \Xi \quad (14)$$

and

$$\min_{x_2 \in X_2(x_1, \xi) \cap \tilde{X}_2} F_2(x_2, y_2, \xi) = f_{22}(x_1, y_2, \xi), \quad \forall (x_1, y_2, \xi) \in X_1 \times \hat{Y}_2 \times \Xi. \quad (15)$$

Proof (i) For any given $(x_2, \xi) \in \mathbb{R}^{n_2} \times \Xi$, the concavity of f_{21} w.r.t. y_1 is established in [34, Proposition 2.21 (i)]. Moreover, for any $\lambda \in (0, 1)$ and $x_2^1, x_2^2 \in \mathbb{R}^{n_2}$, we have

$$\begin{aligned} & f_{21}(\lambda x_2^1 + (1 - \lambda)x_2^2, y_1, \xi) \\ &= \max_{y_2 \in Y_2(y_1, \xi)} F_2(\lambda x_2^1 + (1 - \lambda)x_2^2, y_2, \xi) \\ &\leq \max_{y_2 \in Y_2(y_1, \xi)} \lambda F_2(x_2^1, y_2, \xi) + (1 - \lambda) F_2(x_2^2, y_2, \xi) - \frac{\sigma}{2} \lambda (1 - \lambda) \|x_2^1 - x_2^2\|^2 \\ &\leq \lambda \max_{y_2 \in Y_2(y_1, \xi)} F_2(x_2^1, y_2, \xi) + (1 - \lambda) \max_{y_2 \in Y_2(y_1, \xi)} F_2(x_2^2, y_2, \xi) - \frac{\sigma}{2} \lambda (1 - \lambda) \|x_2^1 - x_2^2\|^2 \\ &= \lambda f_{21}(x_2^1, y_1, \xi) + (1 - \lambda) f_{21}(x_2^2, y_1, \xi) - \frac{\sigma}{2} \lambda (1 - \lambda) \|x_2^1 - x_2^2\|^2, \end{aligned}$$

which implies the strong convexity of f_{21} w.r.t. x_2 .

Note that the KKT condition of problem in (11) is

$$\begin{aligned} 0 &= -\nabla_{y_2} F_2(x_2, y_2, \xi) + B(\xi)^\top \pi_{y_2}, \\ 0 &\leq \pi_{y_2}^\top c(\xi) - A(\xi)y_1 - B(\xi)y_2 \geq 0, \end{aligned} \quad (16)$$

where $\pi_{y_2} \in \mathbb{R}^{s_2}$ is the corresponding Lagrange multiplier. Let $(\hat{y}_2, \hat{\pi}_{y_2})$ be a KKT pair of (16). By [39, Theorem 2], under full row rank of $B(\xi)$ in $Y_2(y_1, \xi)$, the set of Lagrange multipliers $\{\hat{\pi}_{y_2}\}$ is a singleton. Then by [34, Corollary 2.23], $\nabla_{y_1} f_{21}(x_2, y_1, \xi) = \{-A(\xi)^\top \hat{\pi}_{y_2}\}$ is a singleton and $f_{21}(x_2, \cdot, \xi)$ is differentiable. Moreover, by the Danskin theorem, $\nabla_{x_2} f_{21}(x_2, y_1, \xi) = \nabla_{x_2} F_2(x_2, \hat{y}_2, \xi)$, where $\hat{y}_2$ is the unique solution of (11).

Let $(\hat{y}_2 : \mathbb{R}^{n_2} \times Y_1 \times \Xi \rightarrow \mathbb{R}^{m_2}, \hat{\pi}_{y_2} : \mathbb{R}^{n_2} \times Y_1 \times \Xi \rightarrow \mathbb{R}^{s_2})$ be the KKT pair mapping of (16), such that $(\hat{y}_2(x_2, y_1, \xi), \hat{\pi}_{y_2}(x_2, y_1, \xi)) = (\hat{y}_2, \hat{\pi}_{y_2})$, the KKT pair of (16) with corresponding (x_2, y_1, ξ) . Since $F_2(\cdot, \cdot, \xi)$ is σ -strongly convex-strongly concave, every element in $\partial_{y_2 y_2}^2 F_2(x_2, y_2, \xi)$ is a negative definite matrix. Then under Assumption 1 (i) and (iii), and Assumption 2, applying Theorem 4 (in the Appendix) to the problem in (11) yields that $(\hat{y}_2, \hat{\pi}_{y_2})$ is Lipschitz continuous w.r.t. (x_2, y_1) , and continuous w.r.t. ξ , which implies $f_{21}(\cdot, \cdot, \xi)$ is continuously differentiable for any given $\xi \in \Xi$, $\nabla_{y_1} f_{21}(x_2, y_1, \xi) = -A(\xi)^\top \hat{\pi}_{y_2}(x_2, y_1, \xi)$, $\nabla_{x_2} f_{21}(x_2, y_1, \xi) = \nabla_{x_2} F_2(x_2, \hat{y}_2(x_2, y_1, \xi), \xi)$, $\nabla_{y_1} f_{21}$ and $\nabla_{x_2} f_{21}$ are Lipschitz continuous w.r.t. (x_2, y_1) , and continuous w.r.t. ξ .

The proof of part (ii) follows a similar argument as part (i) and will be omitted here.

(iii) Note that

$$\psi_2(x_1, y_1, \xi) := \min_{x_2 \in X_2(x_1, \xi)} f_{21}(x_2, y_1, \xi) \quad (17)$$

and $f_{21}(x_2, \cdot, \xi)$ is concave. For any $\lambda \in [0, 1]$, and $y_1^1, y_1^2 \in Y_1$, let

$$x_2^\lambda = \arg \min_{x_2 \in X_2(x_1, \xi)} f_{21}(x_2, \lambda y_1^1 + (1 - \lambda)y_1^2, \xi),$$

$x_2^1 = \arg \min_{x_2 \in X_2(x_1, \xi)} f_{21}(x_2, y_1^1, \xi)$ and $x_2^2 = \arg \min_{x_2 \in X_2(x_1, \xi)} f_{21}(x_2, y_1^2, \xi)$. Then for any given x_1 and ξ , we have

$$\begin{aligned} \psi_2(x_1, \lambda y_1^1 + (1 - \lambda)y_1^2, \xi) &= f_{21}(x_2^\lambda, \lambda y_1^1 + (1 - \lambda)y_1^2, \xi) \\ &\geq \lambda f_{21}(x_2^\lambda, y_1^1, \xi) + (1 - \lambda) f_{21}(x_2^\lambda, y_1^2, \xi) \\ &\geq \lambda f_{21}(x_2^1, y_1^1, \xi) + (1 - \lambda) f_{21}(x_2^2, y_1^2, \xi) \\ &= \lambda \psi_2(x_1, y_1^1, \xi) + (1 - \lambda) \psi_2(x_1, y_1^2, \xi), \end{aligned}$$

which implies the concavity of ψ_2 w.r.t. y_1 . Moreover, applying [34, Proposition 2.21] to (17), ψ_2 is convex w.r.t. x_1 .

Moreover, by continuous differentiability and strong convexity of $f_{21}(\cdot, y_1, \xi)$ from part (i), every element in $\partial_{x_2 x_2}^2 f_{21}(x_2, y_1, \xi)$ is a positive definite matrix. Then under Assumption 1 (i) and (iii), Assumption 2 and applying Theorem 4 (in the Appendix) to the problem in (17), similar as in part

(i), $(\bar{x}_2, \pi_{x_2})$, the KKT pair of the problem in (17), is Lipschitz continuous w.r.t. (x_1, y_1) and continuous w.r.t. ξ . Combining with [34, Corollary 2.23], $\nabla_{x_1} \psi_2(x_1, y_1, \xi) = \{T(\xi)^\top \pi_{x_2}(x_1, y_1, \xi)\}$ is a singleton, $\nabla_{x_1} \psi_2$ is Lipschitz continuous w.r.t. (x_1, y_1) , and continuous w.r.t. ξ . Obviously, $\psi_2(\cdot, y_1, \xi)$ is continuously differentiable.

Since F_2 is σ -strongly convex-strongly concave,

$$\psi_2(x_1, y_1, \xi) = \max_{y_2 \in Y_2(y_1, \xi)} f_{22}(x_1, y_2, \xi). \quad (18)$$

Then similar as above argument, $(\bar{y}_2, \pi_{y_2})$, the KKT pair of (18), is Lipschitz continuous w.r.t. (x_1, y_1) , and continuous w.r.t. ξ . Then $\nabla_{y_1} \psi_2(x_1, y_1, \xi) = \{-A(\xi)^\top \pi_{y_2}(x_1, y_1, \xi)\}$ is a singleton, $\nabla_{y_1} \psi_2$ is Lipschitz continuous w.r.t. (x_1, y_1) and continuous w.r.t. ξ , and $\psi_2(x_1, \cdot, \xi)$ is continuously differentiable.

Note also that $(\bar{x}_2(x_1, y_1, \xi), \bar{y}_2(x_1, y_1, \xi), \pi_{x_2}(x_1, y_1, \xi), \pi_{y_2}(x_1, y_1, \xi))$ is the unique solution of the KKT condition of the minimax problem in (2), then $(\bar{x}_2(x_1, y_1, \xi), \bar{y}_2(x_1, y_1, \xi))$ is the unique saddle point of the minimax problem in (2).

(iv) Since $X_1 \times Y_1 \times \Xi$ is compact and $(\bar{x}_2, \bar{y}_2)$ is continuous, for all $(x_1, y_1, \xi) \in X_1 \times Y_1 \times \Xi$, there exists convex and compact set $\hat{X}_2 \times \hat{Y}_2 \subset \mathbb{R}^{n_2} \times \mathbb{R}^{m_2}$ such that $(\bar{x}_2(x_1, y_1, \xi), \bar{y}_2(x_1, y_1, \xi)) \in \hat{X}_2 \times \hat{Y}_2$.

By part (i), the solution function $\hat{y}_2$ of the problem in (11) is continuous over $\hat{X}_2 \times Y_1 \times \Xi$. By the boundedness of $\hat{X}_2 \times Y_1 \times \Xi$, there exists convex and compact set $\check{Y}_2$ such that $\hat{Y}_2 \subset \check{Y}_2$, $\hat{y}_2(x_2, y_1, \xi) \in \check{Y}_2$ over $\hat{X}_2 \times Y_1 \times \Xi$ and (14) holds.

Similarly, we can prove the existence of $\check{X}_2$ such that $\hat{X}_2 \subset \check{X}_2$ and (15) holds. $\square$

To study the two-stage stochastic minimax problem (1)-(2), we need the following definitions.

Definition 7 [34, Section 9.2.4] It is said that functions $h_1 : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}$ and $h_2 : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}$ are random lsc and random upper semicontinuous (usc) respectively, if the epi-graphical multifunction $\xi \rightarrow \text{epi } h_1(\cdot, \xi)$ and the hypo-graphical multifunction $\xi \rightarrow \text{hypo } h_2(\cdot, \xi)$ are closed valued and measurable respectively.

Definition 8 A function $h_3 : \mathbb{R}^n \times \mathbb{R}^m \times \Xi \rightarrow \mathbb{R}$ is random lower-upper semicontinuous w.r.t. (x, y) if for any given y , $h_3(\cdot, y, \cdot)$ is random lsc, and for any given x , $h_3(x, \cdot, \cdot)$ is random usc.

Proposition 1 Under Assumptions 1-2, the following statements hold.

- (i) f_{21} is random lower-upper semicontinuous on $\hat{X}_2 \times Y_1$, f_{22} is random lower-upper semicontinuous on $X_1 \times \hat{Y}_2$, where $\hat{X}_2 \times \hat{Y}_2$ is a compact set containing all saddle points of problem (2).
- (ii) ψ_2 is random lower-upper semicontinuous on $X_1 \times Y_1$.

(iii) If $\psi(\cdot, y_1)$ is quasi-convex on X_1 for every $y_1 \in Y_1$, $\lambda > 0$ and $\tilde{f} : \mathbb{R}^{n_1} \rightarrow \mathbb{R}^{n_1}$ is a continuously differentiable vector-valued function, then there exists a local saddle point (x_1^*, y_1^*) of problem

$$\min_{x_1 \in X_1} \max_{y_1 \in Y_1} \psi_\lambda(x_1, y_1) := F_1(x_1, y_1) + \lambda \|\tilde{f}(x_1)_+\|_0 + \mathbb{E}[\psi_2(x_1, y_1, \xi)]. \quad (19)$$

Proof (i) Let $\tilde{Y}_2(\xi) := \{(y_1, y_2) \in Y_1 \times \mathbb{R}^{m_2} : A(\xi)y_1 + B(\xi)y_2 \leq c(\xi)\}$. From Assumption 2, if $\{\xi^k\} \subset \Xi$, $\xi^k \rightarrow \bar{\xi} \in \Xi$, $(y_1^k, y_2^k) \in \tilde{Y}_2(\xi^k)$ and $(y_1^k, y_2^k) \rightarrow (\bar{y}_1, \bar{y}_2)$, then $(\bar{y}_1, \bar{y}_2) \in \tilde{Y}_2(\bar{\xi})$, which implies $\tilde{Y}_2$ is closed. By [34, Remark 62], $\tilde{Y}_2$ is a closed valued measurable multifunction. Moreover, since F_2 is a Carathéodory function, for given x_2 , by [34, Comments after Theorem 9.49], $F_2(x_2, \cdot, \cdot) + \delta_{\tilde{Y}_2(\xi)}(\cdot, \cdot)$ is a random usc function, where $\delta_{\tilde{Y}_2(\xi)}(y_1, y_2) = 0$ if $(y_1, y_2) \in \tilde{Y}_2(\xi)$ and $\delta_{\tilde{Y}_2(\xi)}(y_1, y_2) = -\infty$ otherwise. Then by [34, Theorem 9.49], $f_{21}(x_2, \cdot, \cdot)$ is measurable.

Note that by Lemma 1 (iv), $\text{dom}(F_2(x_2, \cdot, \xi) + \delta_{Y_2(y_1, \xi) \cap \tilde{Y}_2(\cdot)}(\cdot))$ is nonempty and bounded for a.e. $\xi \in \Xi$, according to [34, Theorem 9.50], it follows that $f_{21}(x_2, \cdot, \cdot)$ is random usc for all $x_2 \in \tilde{X}_2$.

Then, we show that for any given $y_1 \in Y_1$, $f_{21}(\cdot, y_1, \cdot)$ is random lsc. Since $F_2(\cdot, y_2, \cdot)$ is random lsc for given y_2 . Let $\bar{x}_2 \in \mathbb{R}^{n_2}$, $\{x_2^k\} \subset \mathbb{R}^{n_2}$ with $x_2^k \rightarrow \bar{x}_2$ as $k \rightarrow \infty$ and $y_2^1 = \arg \max_{y_2 \in Y_2(y_1, \xi)} F_2(\bar{x}_2, y_2, \xi)$, then we have

$$\begin{aligned} \liminf_{k \rightarrow \infty} f_{21}(x_2^k, y_1, \xi) &= \liminf_{k \rightarrow \infty} \max_{y_2 \in Y_2(y_1, \xi)} F_2(x_2^k, y_2, \xi) \\ &\geq \liminf_{k \rightarrow \infty} F_2(x_2^k, y_2^1, \xi) \\ &\geq F_2(\bar{x}_2, y_2^1, \xi) \\ &= f_{21}(\bar{x}_2, y_1, \xi), \end{aligned}$$

which implies f_{21} is lsc w.r.t. x_2 .

Applying [34, Theorem 9.49] to $\max_{y_2 \in Y_2(y_1, \xi)} F_2(x_2, y_2, \xi)$, the optimal value function $f_{21}(\cdot, y_1, \cdot)$ is jointly measurable. Combining the measurability of $f_{21}(\cdot, y_1, \cdot)$ and the lower semicontinuity of f_{21} w.r.t. x_2 , by [34, Theorem 9.48], $f_{21}(\cdot, y_1, \cdot)$ is random lsc for any given $y_1 \in Y_1$. Then f_{21} is random lower-upper semicontinuous on $\tilde{X}_2 \times Y_1$.

The proof for f_{22} follows a similar argument as for f_{21} and will be omitted here.

(ii) Note that $f_{21}(x_2, \cdot, \cdot)$ is random usc. Let $\bar{y}_1 \in Y_1$, $\{y_1^k\} \subset Y_1$ with $y_1^k \rightarrow \bar{y}_1$ as $k \rightarrow \infty$ and $x_2^1 \in \arg \min_{x_2 \in X_2(x_1, \xi)} f_{21}(x_2, \bar{y}_1, \xi)$. Then

$$\begin{aligned} \limsup_{k \rightarrow \infty} \psi_2(x_1, y_1^k, \xi) &= \limsup_{k \rightarrow \infty} \min_{x_2 \in X_2(x_1, \xi)} f_{21}(x_2, y_1^k, \xi) \\ &\leq \limsup_{k \rightarrow \infty} f_{21}(x_2^1, y_1^k, \xi) \\ &\leq f_{21}(x_2^1, \bar{y}_1, \xi) \\ &= \psi_2(x_1, \bar{y}_1, \xi), \end{aligned}$$

which implies that ψ_2 is usc w.r.t. y_1 for given x_1 and ξ .

Applying [34, Theorem 9.49] to problem (17), the minimax value function $\psi_2(x_1, \cdot, \cdot)$ is jointly measurable. Combining the measurability of $\psi_2(x_1, \cdot, \cdot)$ and

the upper semicontinuity of ψ_2 w.r.t. y_1 , by [34, Theorem 9.48], $\psi_2(x_1, \cdot, \cdot)$ is random usc for any given $x_1 \in X_1$.

Similar to the random upper semicontinuity of ψ_2 w.r.t. y_1 for any given $x_1 \in X_1$, we can prove the random lower semicontinuity of ψ_2 w.r.t. x_1 for any given $y_1 \in Y_1$. Then ψ_2 is random lower-upper semicontinuous w.r.t. (x_1, y_1) on $X_1 \times Y_1$. We omit the details here.

Finally, part (iii) is a corollary of [2, Proposition 2.2]. $\square$

By [33, Page 439], we note that, for the random lower-upper semicontinuity conclusion in this proposition, the continuity of the random matrices and vectors w.r.t. ξ can be weakened to measurability.

Remark 4 Note that the second-stage minimax problem in (2) is a parametric minimax problem. In Lemma 1, we investigate the continuity and the boundedness of the second-stage saddle point mapping $(\bar{x}_2, \bar{y}_2)$. In Proposition 1, we consider the random lower-upper semicontinuity of the minimax value function ψ_2 of the second-stage minimax problem in (2).

Let

$$\min_{x_1 \in X_1} \max_{y_1 \in Y_1} F_1(x_1, y_1) + \mathbb{E} \left[\hat{\psi}_2(x_1, y_1, \xi) \right] \quad (20)$$

and

$$\hat{\psi}_2(x_1, y_1, \xi) := \min_{x_2 \in X_2(x_1, \xi) \cap \tilde{X}_2} \max_{y_2 \in Y_2(y_1, \xi) \cap \tilde{Y}_2} F_2(x_2, y_2, \xi), \quad (21)$$

where $\tilde{X}_2$ and $\tilde{Y}_2$ are given in (14)-(15). By Lemma 1 and Proposition 1, it is easy to show the equivalence between (1)-(2) and (20)-(21) in the sense that: (i) for every $(x_1, y_1) \in X_1 \times Y_1$ they have the same and unique saddle point $(\bar{x}_2(x_1, y_1, \xi), \bar{y}_2(x_1, y_1, \xi))$, and (14)-(15) hold; (ii) they have same sets of points (x_1^*, y_1^*) satisfying (5).

3.1 Interchange of the expectation and the minimax operator

Two-stage stochastic minimization problem has been considered as an infinite-dimensional large-scale optimization problem [34, Section 2.3.1]. Can the two-stage stochastic minimax problem (1)-(2) also be treated as an infinite-dimensional large-scale minimax problem (8)? To answer this question, we examine whether the expectation and the minimax operator in the second-stage problem can be interchanged.

Proposition 2 *Under Assumptions 1-2, the two-stage stochastic minimax problem (1)-(2) is equivalent to the minimax problem (8). Moreover, the following statements hold.*

- (i) *If $(x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*)$ is a (local) saddle point of (8), then (x_1^*, y_1^*) is a (local) saddle point of (1).*

- (ii) Suppose that $F_1(\cdot, y_1) + \mathbb{E}[f_{22}(\cdot, \mathbf{y}_2(\xi), \xi)]$ is quasi-convex on X_1 for every $y_1 \in Y_1$ and every essentially bounded measurable selection $\mathbf{y}_2(\xi) \in Y_2(y_1, \xi)$ a.e. Then (8) has a saddle point. Moreover, every (local) saddle point (x_1^*, y_1^*) of (1) lifts to a (local) saddle point $(x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*)$ of (8), where $(\mathbf{x}_2^*, \mathbf{y}_2^*) := (\bar{\mathbf{x}}_2(x_1^*, y_1^*, \cdot), \bar{\mathbf{y}}_2(x_1^*, y_1^*, \cdot))$.
- (iii) In addition to (ii), if $\lambda > 0$ and f is a continuously differentiable vector-valued function, then there exists a local saddle point $(x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*)$ of

$$\min_{(x_1, \mathbf{x}_2) \in \mathbf{X}} \max_{(y_1, \mathbf{y}_2) \in \mathbf{Y}} F_1(x_1, y_1) + \lambda \|\tilde{f}(x_1)_+\|_0 + \mathbb{E}[F_2(\mathbf{x}_2(\xi), \mathbf{y}_2(\xi), \xi)].$$

Proof For any given $(x_1, y_1) \in X_1 \times Y_1$, we consider the expectation of the second-stage problem

$$\mathbb{E} \left[\min_{x_2 \in X_2(x_1, \xi)} \max_{y_2 \in Y_2(y_1, \xi)} F_2(x_2, y_2, \xi) \right] \quad (22)$$

firstly. Note that by Remark 4, (22) is equivalent to

$$\mathbb{E} \left[\min_{x_2 \in X_2(x_1, \xi) \cap \tilde{X}_2} \max_{y_2 \in Y_2(y_1, \xi) \cap \tilde{Y}_2} F_2(x_2, y_2, \xi) \right]. \quad (23)$$

By Proposition 1 (i), $\max_{y_2 \in Y_2(y_1, \xi) \cap \tilde{Y}_2} F_2(\cdot, y_2, \xi)$ is random lsc over $\hat{X}_2$. Then by [34, Theorem 9.108], (23) is equivalent to

$$\min_{\mathbf{x}_2 \in \tilde{X}_2(x_1)} \mathbb{E} \left[\max_{y_2 \in Y_2(y_1, \xi) \cap \tilde{Y}_2} F_2(\mathbf{x}_2(\xi), y_2, \xi) \right], \quad (24)$$

where

$$\tilde{X}_2(x_1) := \left\{ \mathbf{x}_2 \in \mathcal{L}_p^{n_2} : \mathbf{x}_2(\xi) \in \hat{X}_2, T(\xi)x_1 + W(\xi)\mathbf{x}_2(\xi) \leq h(\xi), \text{ a.e. } \xi \in \Xi \right\}.$$

Moreover, for any fixed measurable selection $\mathbf{x}_2 \in \tilde{X}_2(x_1)$, the mapping $(y_2, \xi) \mapsto F_2(\mathbf{x}_2(\xi), y_2, \xi)$ is random usc, since $\mathbf{x}_2(\cdot)$ is measurable and F_2 is continuous. Therefore, by [34, Theorem 9.108], we have

$$\mathbb{E} \left[\max_{y_2 \in Y_2(y_1, \xi) \cap \tilde{Y}_2} F_2(\mathbf{x}_2(\xi), y_2, \xi) \right] = \max_{\mathbf{y}_2 \in \tilde{Y}_2(y_1)} \mathbb{E}[F_2(\mathbf{x}_2(\xi), \mathbf{y}_2(\xi), \xi)],$$

where

$$\tilde{Y}_2(y_1) := \left\{ \mathbf{y}_2 \in \mathcal{L}_p^{m_2} : \mathbf{y}_2(\xi) \in \tilde{Y}_2, A(\xi)y_1 + B(\xi)\mathbf{y}_2(\xi) \leq c(\xi), \text{ a.e. } \xi \in \Xi \right\}.$$

Taking the minimum over $\tilde{X}_2(x_1)$ on both sides yields that (24) is equivalent to

$$\min_{\mathbf{x}_2 \in \tilde{X}_2(x_1)} \max_{\mathbf{y}_2 \in \tilde{Y}_2(y_1)} \mathbb{E}[F_2(\mathbf{x}_2(\xi), \mathbf{y}_2(\xi), \xi)]. \quad (25)$$

Then for any $(x_1, y_1) \in X_1 \times Y_1$,

$$\begin{aligned} & \mathbb{E} \left[\min_{x_2 \in X_2(x_1, \xi)} \max_{y_2 \in Y_2(y_1, \xi)} F_2(x_2, y_2, \xi) \right] \\ &= \min_{\mathbf{x}_2 \in \tilde{X}_2(x_1)} \max_{\mathbf{y}_2 \in \tilde{Y}_2(y_1)} \mathbb{E} [F_2(\mathbf{x}_2(\xi), \mathbf{y}_2(\xi), \xi)]. \end{aligned}$$

Given that Lemma 1(iv) ensures $(\bar{\mathbf{x}}_2(x_1, y_1, \xi), \bar{\mathbf{y}}_2(x_1, y_1, \xi)) \in \hat{X}_2 \times \hat{Y}_2 \subset \tilde{X}_2 \times \tilde{Y}_2$ for all $(x_1, y_1, \xi) \in X_1 \times Y_1 \times \Xi$, it follows, by Sion's theorem [37] for fixed x_1 , that (1)-(2) is equivalent to (8).

In (i), the second-stage components of a (local) saddle point $(x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*)$ of (8) form the unique saddle point of (2) at (x_1^*, y_1^*) a.e. Taking second-stage best responses in the saddle inequalities gives the saddle inequalities for ψ . In the local case, continuity of these responses by Lemma 1(i)-(ii) keeps them in the original neighborhood for (x_1, y_1) sufficiently close to (x_1^*, y_1^*) , proving the local assertion.

Then we consider (ii). Its additional assumption implies that $\psi(\cdot, y_1)$ is quasi-convex for every $y_1 \in Y_1$. We first prove the lifting assertion. Since (x_1^*, y_1^*) is a saddle point of (1),

$$\begin{aligned} & \min_{(x_1, \mathbf{x}_2) \in \mathbf{X}} \max_{(y_1, \mathbf{y}_2) \in \mathbf{Y}} F_1(x_1, y_1) + \mathbb{E} [F_2(\mathbf{x}_2(\xi), \mathbf{y}_2(\xi), \xi)] \\ &= \min_{x_1 \in X_1} \max_{y_1 \in Y_1} F_1(x_1, y_1) + \mathbb{E} [\psi_2(x_1, y_1, \xi)] \\ &= F_1(x_1^*, y_1^*) + \mathbb{E} [\psi_2(x_1^*, y_1^*, \xi)] \\ &= \max_{y_1 \in Y_1} \min_{x_1 \in X_1} F_1(x_1, y_1) + \mathbb{E} [\psi_2(x_1, y_1, \xi)] \\ &= \max_{(y_1, \mathbf{y}_2) \in \mathbf{Y}} \min_{(x_1, \mathbf{x}_2) \in \mathbf{X}} F_1(x_1, y_1) + \mathbb{E} [F_2(\mathbf{x}_2(\xi), \mathbf{y}_2(\xi), \xi)], \end{aligned} \tag{26}$$

where the second and third equalities are from [15, Theorem 1.4.1]. The last equality follows by applying Sion's theorem to $F_1(x_1, y_1) + \mathbb{E} [f_{22}(x_1, \mathbf{y}_2(\xi), \xi)]$ in $(x_1, \mathbf{y}_2)$ for fixed y_1 , together with the expectation-minimization interchange.

By Lemma 1 (iv)

$$\begin{aligned} z_1(x_1, \mathbf{x}_2) &:= \max_{y_1 \in Y_1} F_1(x_1, y_1) + \mathbb{E} [f_{21}(\mathbf{x}_2(\xi), y_1, \xi)] \\ &= \max_{(y_1, \mathbf{y}_2) \in \mathbf{Y}} F_1(x_1, y_1) + \mathbb{E} [F_2(\mathbf{x}_2(\xi), \mathbf{y}_2(\xi), \xi)]. \end{aligned}$$

By the compactness of Y_1 and [4, Proposition 4.4], applied to the maximization defining $z_1(x_1, \mathbf{x}_2) - f(x_1)$, this difference is continuous w.r.t. $(x_1, \mathbf{x}_2)$ in the product norm topology. Since f is lsc, z_1 is lsc. By weak compactness of the bounded selection sets and lower-upper semicontinuity, the extrema are attained. For the given x_1^* , there exists $\mathbf{x}_2^*$ such that

$$\min_{(x_1, \mathbf{x}_2) \in \mathbf{X}} z_1(x_1, \mathbf{x}_2) = z_1(x_1^*, \mathbf{x}_2^*) = \max_{(y_1, \mathbf{y}_2) \in \mathbf{Y}} F_1(x_1^*, y_1) + \mathbb{E} [F_2(\mathbf{x}_2^*(\xi), \mathbf{y}_2(\xi), \xi)]. \tag{27}$$

Similarly, for the given y_1^* , there exists $\mathbf{y}_2^*$ such that

$$\max_{(y_1, \mathbf{y}_2) \in \mathbf{Y}} z_2(y_1, \mathbf{y}_2) = z_2(y_1^*, \mathbf{y}_2^*) = \min_{(x_1, \mathbf{x}_2) \in \mathbf{X}} F_1(x_1, y_1^*) + \mathbb{E} [F_2(\mathbf{x}_2(\xi), \mathbf{y}_2^*(\xi), \xi)], \tag{28}$$

where $z_2(y_1, \mathbf{y}_2) := \min_{x_1 \in X_1} F_1(x_1, y_1) + \mathbb{E}[f_{22}(x_1, \mathbf{y}_2(\xi), \xi)]$. Combining (27)-(28), we have

$$\begin{aligned} & \max_{(y_1, \mathbf{y}_2) \in \mathbf{Y}} F_1(x_1^*, y_1) + \mathbb{E}[F_2(\mathbf{x}_2^*(\xi), \mathbf{y}_2(\xi), \xi)] \\ &= F_1(x_1^*, y_1^*) + \mathbb{E}[F_2(\mathbf{x}_2^*(\xi), \mathbf{y}_2^*(\xi), \xi)] \\ &= \min_{(x_1, \mathbf{x}_2) \in \mathbf{X}} F_1(x_1, y_1^*) + \mathbb{E}[F_2(\mathbf{x}_2(\xi), \mathbf{y}_2^*(\xi), \xi)], \end{aligned}$$

which implies that $(x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*)$ is a saddle point.

For a local saddle point (x_1^*, y_1^*) of (1), the strong convexity-strong concavity of F_2 ensures that the corresponding second-stage local saddle point is also its unique global saddle point. Take $(\mathbf{x}_2^*, \mathbf{y}_2^*)$ to be the corresponding saddle point mappings in Lemma 1. Restricting X_1 and Y_1 to sufficiently small closed convex neighborhoods of x_1^* and y_1^* preserves the additional quasi-convexity assumption, so the preceding argument shows that $(x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*)$ is a local saddle point of (8).

Finally, Sion's theorem gives a saddle point of (1), which lifts to a saddle point of (8) by the above argument. This completes (ii).

Part (iii) follows from part (ii) and [2, Proposition 2.2] directly. $\square$

Note that under the additional assumption in Proposition 2(ii), problem (1)-(2) is equivalent to two-stage stochastic two-player zero-sum game (3)-(4).

Remark 5 We consider three kinds of saddle points for two-stage stochastic minimax problem (1)-(2): the saddle point (x_1^*, y_1^*) of (1), the saddle point mapping $(\bar{\mathbf{x}}_2(x_1, y_1, \xi), \bar{\mathbf{y}}_2(x_1, y_1, \xi))$ of the second-stage minimax problem in (2) (see Definition 2), and the saddle point $(x_1^*, \mathbf{x}_2^*, y_1^*, \mathbf{y}_2^*)$ of (8).

Under Assumptions 1-2 and the additional assumption in Proposition 2(ii), parts (i)-(ii) show that (x_1^*, y_1^*) is a saddle point of (1) if and only if its lift is a saddle point of (8). By Lemma 1,

$$(\bar{\mathbf{x}}_2(x_1^*, y_1^*, \cdot), \bar{\mathbf{y}}_2(x_1^*, y_1^*, \cdot)) = (\mathbf{x}_2^*(\cdot), \mathbf{y}_2^*(\cdot))$$

and $(\bar{\mathbf{x}}_2(\cdot, \cdot, \cdot), \bar{\mathbf{y}}_2(\cdot, \cdot, \cdot))$ is continuous.

Remark 6 Proposition 2(i) also applies to the two-stage stochastic minimax problem with cardinality penalties (19). Since the cardinality penalty is generally discontinuous, we may consider its continuous relaxation $\lambda r(\tilde{f}(\cdot), \mu)$, where $r(\cdot, \cdot)$ is defined in [2, Formulation (3.3)], leading to the following problem:

$$\min_{x_1 \in X_1} \max_{y_1 \in Y_1} \tilde{\psi}_\lambda(x_1, y_1) := F_1(x_1, y_1) + \lambda r(\tilde{f}(x_1), \mu) + \mathbb{E}[\psi_2(x_1, y_1, \xi)]. \quad (29)$$

By [2, Theorem 4.1], under suitable conditions (e.g. [2, Assumption 4.1]), (x_1^*, y_1^*) is a saddle point of (19) if and only if it is a saddle point of (29). Moreover, (x_1^*, y_1^*) is a local saddle point of (19) if it is a local saddle point of (29).

3.2 Global and local minimax points

Since (1) is a nonconvex-nonsmooth minimax problem, the sets of saddle points and local saddle points of (1) may be empty. We therefore consider the sets of global minimax points and local minimax points of (1), denoted by S_g and S_l , respectively.

Throughout the remainder of this paper, we additionally assume that f is locally Lipschitz continuous on an open neighborhood of X_1 . Under Assumptions 1–2, Lemma 1 and the compactness of Y_1 ensure that $\bar{\psi}$ is locally Lipschitz continuous near X_1 . A point $x_1^* \in X_1$ is called a stationary point of (7) if $0 \in \partial\bar{\psi}(x_1^*) + \mathcal{N}_{X_1}(x_1^*)$.

Proposition 3 *Suppose that Assumptions 1–2 hold. Then (a) x_1^* is a stationary point of problem (7) if and only if (b) there exists a corresponding y_1^* such that (x_1^*, y_1^*) is a solution of the following variational inequality:*

$$\begin{cases} 0 \in \partial f(x_1) + \nabla_{x_1} \psi_1(x_1, y_1) + \nabla_{x_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] + \mathcal{N}_{X_1}(x_1), \\ 0 \in -\nabla_{y_1} \psi_1(x_1, y_1) - \nabla_{y_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] + \partial g(y_1) + \mathcal{N}_{Y_1}(y_1). \end{cases} \quad (30)$$

Moreover, (c) $S_g \subset S_l \subset S_{kkt}$ are nonempty, where S_{kkt} is the solution set of (30).

Proof We prove (a) $\Rightarrow$ (b) firstly. When x_1^* is a stationary point of problem (7), by Lemma 1 (iii) and Proposition 1 (ii), $\mathbb{E}[\psi_2]$ is convex-concave and continuous, combined with strong convexity of $g(\cdot) - \psi_1(x_1, \cdot)$, there must exist a unique y_1^* such that

$$0 \in -\nabla_{y_1} \psi_1(x_1^*, y_1^*) - \nabla_{y_1} \mathbb{E}[\psi_2(x_1^*, y_1^*, \xi)] + \partial g(y_1^*) + \mathcal{N}_{Y_1}(y_1^*),$$

and $\bar{\psi}(x_1^*) = \psi(x_1^*, y_1^*)$. Moreover, by Danskin's theorem,

$$\partial\bar{\psi}(x_1^*) = \partial_{x_1} \psi(x_1^*, y_1^*) = \partial f(x_1^*) + \nabla_{x_1} \psi_1(x_1^*, y_1^*) + \nabla_{x_1} \mathbb{E}[\psi_2(x_1^*, y_1^*, \xi)]. \quad (31)$$

Then $0 \in \partial\bar{\psi}(x_1^*) + \mathcal{N}_{X_1}(x_1^*)$ implies that (x_1^*, y_1^*) is a solution of (30).

The proof of (b) $\Rightarrow$ (a) is similar and we omit the details.

Now we prove (c). By (31), $\bar{\psi}(\cdot)$ is Lipschitz continuous over X_1 . Then, by the compactness of X_1 , the global optimal solution set of (7), denoted by $S_{\bar{\psi}}^*$, is nonempty. Then it is easy to check that

$$S_g := \left\{ (\bar{x}_1^*, \bar{y}_1^*(\bar{x}_1^*)) : \bar{x}_1^* \in S_{\bar{\psi}}^* \text{ and } \bar{y}_1^*(\bar{x}_1^*) := \arg \max_{y_1 \in Y_1} \psi(\bar{x}_1^*, y_1) \right\}.$$

By Remark 3, if $\bar{y}_1$ is a global maximum point of $\psi(\bar{x}_1, \cdot)$, and $\bar{x}_1$ is a local minimum point of $\bar{\psi}(\cdot)$, then $(\bar{x}_1, \bar{y}_1)$ is a local minimax point of problem (1). It is trivial to observe that all global minimax points satisfy the above conditions and are therefore local minimax points, that is, $S_g \subset S_l$.

Moreover, $S_l \subset S_{kkt}$ follows directly from [20, Theorem 3.11]. $\square$

Note that problem (8) is also a nonconvex-nonsmooth minimax problem; consequently, the set of saddle points may be empty. In this case, we consider global minimax points and local minimax points of (8).

Proposition 4 *Under Assumptions 1-2, the following statements hold.*

(i) $(\tilde{x}_1, \tilde{y}_1)$ is a global minimax point of (1) if and only if $(\tilde{x}_1, \tilde{\mathbf{x}}_2, \tilde{y}_1, \tilde{\mathbf{y}}_2)$ is a global minimax point of (8), where

$$(\tilde{\mathbf{x}}_2(\xi), \tilde{\mathbf{y}}_2(\xi)) = (\mathbf{x}_2^*(\tilde{x}_1, \tilde{y}_1, \xi), \mathbf{y}_2^*(\tilde{x}_1, \tilde{y}_1, \xi))$$

is the unique saddle point of the minimax problem in (2) with $(\tilde{x}_1, \tilde{y}_1)$ for a.e. $\xi \in \Xi$;

(ii) $(\tilde{x}_1, \tilde{y}_1)$ is a local minimax point of (1) if and only if $(\tilde{x}_1, \tilde{\mathbf{x}}_2, \tilde{y}_1, \tilde{\mathbf{y}}_2)$ is a local minimax point of (8), where $(\tilde{\mathbf{x}}_2(\xi), \tilde{\mathbf{y}}_2(\xi)) = (\mathbf{x}_2^*(\tilde{x}_1, \tilde{y}_1, \xi), \mathbf{y}_2^*(\tilde{x}_1, \tilde{y}_1, \xi))$ is the unique saddle point of the minimax problem in (2) with $(\tilde{x}_1, \tilde{y}_1)$ for a.e. $\xi \in \Xi$.

Proof (i) We only show the backward implication, since the forward implication follows from the strong convexity-concavity of the second-stage problem immediately. For any global minimax point $(\tilde{x}_1, \tilde{\mathbf{x}}_2, \tilde{y}_1, \tilde{\mathbf{y}}_2) \in \mathbf{X} \times \mathbf{Y}$ of the minimax problem (8), we have

$$\begin{aligned} & F_1(\tilde{x}_1, \tilde{y}_1) + \mathbb{E}[F_2(\tilde{\mathbf{x}}_2(\xi), \mathbf{y}_2(\xi), \xi)] \\ & \leq F_1(\tilde{x}_1, \tilde{y}_1) + \mathbb{E}[F_2(\tilde{\mathbf{x}}_2(\xi), \tilde{\mathbf{y}}_2(\xi), \xi)] \\ & \leq \max_{\mathbf{y}_2'(\xi) \in \mathbf{Y}_2(\tilde{y}_1)} F_1(\tilde{x}_1, \tilde{y}_1) + \mathbb{E}[F_2(\mathbf{x}_2(\xi), \mathbf{y}_2'(\xi), \xi)] \end{aligned}$$

for any $(\mathbf{x}_2, \mathbf{y}_2) \in \mathbf{X}_2(\tilde{x}_1) \times \mathbf{Y}_2(\tilde{y}_1)$ with

$$\mathbf{X}_2(\tilde{x}_1) := \{\mathbf{x}_2 \in \mathcal{L}_p^{n_2} : \mathbf{x}_2(\xi) \in \tilde{X}_2, T(\xi)\tilde{x}_1 + W(\xi)\mathbf{x}_2(\xi) \leq h(\xi), \text{ a.e. } \xi \in \Xi\}$$

and

$$\mathbf{Y}_2(\tilde{y}_1) := \{\mathbf{y}_2 \in \mathcal{L}_p^{m_2} : \mathbf{y}_2(\xi) \in \tilde{Y}_2, A(\xi)\tilde{y}_1 + B(\xi)\mathbf{y}_2(\xi) \leq c(\xi), \text{ a.e. } \xi \in \Xi\},$$

which implies $(\tilde{\mathbf{x}}_2(\xi), \tilde{\mathbf{y}}_2(\xi))$ is the global minimax point of the second-stage minimax problem in (2) with $(\tilde{x}_1, \tilde{y}_1)$ for a.e. $\xi \in \Xi$. Since by Lemma 1, the minimax problem in (2) is strongly convex-strongly concave, $(\tilde{\mathbf{x}}_2(\xi), \tilde{\mathbf{y}}_2(\xi)) = (\mathbf{x}_2^*(\tilde{x}_1, \tilde{y}_1, \xi), \mathbf{y}_2^*(\tilde{x}_1, \tilde{y}_1, \xi))$ is the unique saddle point of the second-stage minimax problem in (2), and then (10) is equivalent to (6).

The proof of (ii) is similar as (i), we omit the details. $\square$

Remark 7 If F_1 is continuously differentiable, then the KKT condition of the two-stage minimax problem (1)-(2) is

$$\begin{cases} 0 \in \nabla_{x_1} F_1(x_1, y_1) + \mathbb{E}[T(\xi)^\top \boldsymbol{\pi}_x(\xi)] + \mathcal{N}_{X_1}(x_1), \\ 0 \in -\nabla_{y_1} F_1(x_1, y_1) + \mathbb{E}[A(\xi)^\top \boldsymbol{\pi}_y(\xi)] + \mathcal{N}_{Y_1}(y_1), \\ 0 = \nabla_{\mathbf{x}_2(\xi)} F_2(\mathbf{x}_2(\xi), \mathbf{y}_2(\xi), \xi) + W(\xi)^\top \boldsymbol{\pi}_x(\xi), \\ 0 \leq \boldsymbol{\pi}_x(\xi) \perp h(\xi) - T(\xi)x_1 - W(\xi)\mathbf{x}_2(\xi) \geq 0, \\ 0 = -\nabla_{\mathbf{y}_2(\xi)} F_2(\mathbf{x}_2(\xi), \mathbf{y}_2(\xi), \xi) + B(\xi)^\top \boldsymbol{\pi}_y(\xi), \\ 0 \leq \boldsymbol{\pi}_y(\xi) \perp c(\xi) - A(\xi)y_1 - B(\xi)\mathbf{y}_2(\xi) \geq 0, \quad \text{a.e. } \xi \in \Xi, \end{cases} \quad (32)$$

where $\mathbf{x}_2 \in \mathcal{L}_p^{n_2}$, $\mathbf{y}_2 \in \mathcal{L}_p^{m_2}$, $\boldsymbol{\pi}_x \in \mathcal{L}_q^{l_2}$ and $\boldsymbol{\pi}_y \in \mathcal{L}_q^{s_2}$, with $1/p + 1/q = 1$. Clearly, (32) is a two-stage SVI.

4 Sample average approximation of the two-stage stochastic minimax problem

In this section, we consider SAA of two-stage stochastic minimax problem (1)-(2). Let $\xi^1, \dots, \xi^N$ be independent and identically distributed (i.i.d.) samples of random variable ξ . We consider the following SAA problem:

$$\min_{x_1 \in X_1} \max_{y_1 \in Y_1} \psi_N(x_1, y_1) := f(x_1) + \psi_1(x_1, y_1) + \Phi_N(x_1, y_1) - g(y_1), \quad (33)$$

where $\Phi_N(x_1, y_1) := \frac{1}{N} \sum_{i=1}^N \psi_2(x_1, y_1, \xi^i)$, $\psi_2(x_1, y_1, \xi^i)$ is defined in (2).

Let $\bar{\psi}_N(x_1) := \max_{y_1 \in Y_1} \psi_N(x_1, y_1)$. Then $\bar{x}_1^N$ is a stationary point of

$$\min_{x_1 \in X_1} \bar{\psi}_N(x_1), \quad (34)$$

if $0 \in \partial \bar{\psi}_N(\bar{x}_1^N) + \mathcal{N}_{X_1}(\bar{x}_1^N)$.

Proposition 5 *Suppose that Assumptions 1-2 hold. Then $\bar{x}_1^N$ is a stationary point of problem (34) if and only if there exists a corresponding $\bar{y}_1^N$ such that $(\bar{x}_1^N, \bar{y}_1^N)$ is a solution of*

$$\begin{cases} 0 \in \partial f(x_1) + \nabla_{x_1} \psi_1(x_1, y_1) + \nabla_{x_1} \Phi_N(x_1, y_1) + \mathcal{N}_{X_1}(x_1), \\ 0 \in -\nabla_{y_1} \psi_1(x_1, y_1) - \nabla_{y_1} \Phi_N(x_1, y_1) + \partial g(y_1) + \mathcal{N}_{Y_1}(y_1). \end{cases} \quad (35)$$

Note that $\nabla_{x_1} \Phi_N(x_1, y_1) = \frac{1}{N} \sum_{i=1}^N \nabla_{x_1} \psi_2(x_1, y_1, \xi^i)$ and $\nabla_{y_1} \Phi_N(x_1, y_1) = \frac{1}{N} \sum_{i=1}^N \nabla_{y_1} \psi_2(x_1, y_1, \xi^i)$. The proof of Proposition 5 is similar as the proof of Proposition 3, here we omit the details.

Then we consider the convergence between (1) and (33) in the sense of stationary points of first-stage minimization problem (7) and its SAA problem (34). By Propositions 3 and 5, that is equivalent to the convergence of solutions between (30) and (35). We need the following notations.

For simplicity of notation, we set

$$\begin{pmatrix} G_x(x_1, y_1) \\ G_y(x_1, y_1) \end{pmatrix} := \begin{pmatrix} \partial f(x_1) + \nabla_{x_1} \psi_1(x_1, y_1) + \nabla_{x_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] \\ -\nabla_{y_1} \psi_1(x_1, y_1) - \nabla_{y_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] + \partial g(y_1) \end{pmatrix} \quad (36)$$

and

$$\begin{pmatrix} G_x^N(x_1, y_1) \\ G_y^N(x_1, y_1) \end{pmatrix} := \begin{pmatrix} \partial f(x_1) + \nabla_{x_1} \psi_1(x_1, y_1) + \nabla_{x_1} \Phi_N(x_1, y_1) \\ -\nabla_{y_1} \psi_1(x_1, y_1) - \nabla_{y_1} \Phi_N(x_1, y_1) + \partial g(y_1) \end{pmatrix}. \quad (37)$$

Theorem 1 *Suppose that Assumptions 1-2 hold and $\{\xi^1, \dots, \xi^N\}$ is a set of i.i.d. samples of random variable ξ . Let S^* be the solution set of SVI (30) and $\{(\bar{x}_1^N, \bar{y}_1^N)\}$ be a sequence of solutions of its SAA problem (35). Then almost surely (a.s.)*

$$\lim_{N \rightarrow \infty} \mathbb{D}((\bar{x}_1^N, \bar{y}_1^N), S^*) = 0. \quad (38)$$

Proof Note that, by using notation (36) and (37), SVI (30) and its SAA problem (35) can be written as

$$0 \in \begin{pmatrix} G_x(x_1, y_1) \\ G_y(x_1, y_1) \end{pmatrix} + \mathcal{N}_{X_1 \times Y_1}((x_1, y_1)), \text{ and } 0 \in \begin{pmatrix} G_x^N(x_1, y_1) \\ G_y^N(x_1, y_1) \end{pmatrix} + \mathcal{N}_{X_1 \times Y_1}((x_1, y_1)).$$

Moreover, since $\partial f(\cdot)$ and $\partial g(\cdot)$ are outer semicontinuous (osc)¹ [11, Proposition 2.1.5], $\begin{pmatrix} G_x(\cdot, \cdot) \\ G_y(\cdot, \cdot) \end{pmatrix}$ and $\begin{pmatrix} G_x^N(\cdot, \cdot) \\ G_y^N(\cdot, \cdot) \end{pmatrix}$ are osc. By [41, Lemma 4.2 (i)], for any $\epsilon > 0$, there exists a $\delta > 0$ such that if

$$\sup_{(x_1, y_1) \in X_1 \times Y_1} \mathbb{D} \left(\begin{pmatrix} G_x^N(x_1, y_1) \\ G_y^N(x_1, y_1) \end{pmatrix}, \begin{pmatrix} G_x(x_1, y_1) \\ G_y(x_1, y_1) \end{pmatrix} \right) \leq \delta, \quad (39)$$

then $\mathbb{D}(S_N, S^*) \leq \epsilon$, where S_N is the solution set of the SAA problem (35).

Then we only need to show that, for any $\delta > 0$, (39) holds for N sufficiently large. Note that for any set $A, B, C \subset \mathbb{R}^{n_1+m_1}$, $\mathbb{D}(A + C, B + C) \leq \mathbb{D}(A, B)$, then

$$\begin{aligned} \mathbb{D} \left(\begin{pmatrix} G_x^N(x_1, y_1) \\ G_y^N(x_1, y_1) \end{pmatrix}, \begin{pmatrix} G_x(x_1, y_1) \\ G_y(x_1, y_1) \end{pmatrix} \right) &\leq \left\| \begin{pmatrix} \nabla_{x_1} \Phi_N(x_1, y_1) - \nabla_{x_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] \\ \nabla_{y_1} \Phi_N(x_1, y_1) - \nabla_{y_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] \end{pmatrix} \right\| \\ &\leq \left\| \begin{pmatrix} \nabla_{x_1} \Phi_N(x_1, y_1) - \nabla_{x_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] \\ \nabla_{y_1} \Phi_N(x_1, y_1) - \nabla_{y_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] \end{pmatrix} \right\|_1. \end{aligned}$$

Moreover, by Lemma 1, $\psi_2(\cdot, \cdot, \xi)$ is continuously differentiable, then we have $\nabla_{x_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] = \mathbb{E}[\nabla_{x_1} \psi_2(x_1, y_1, \xi)]$, $\nabla_{y_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] = \mathbb{E}[\nabla_{y_1} \psi_2(x_1, y_1, \xi)]$, $\nabla_{x_1} \psi_2(\cdot, \cdot, \xi) = T(\xi)^\top \pi_{x_2}(\cdot, \cdot, \xi)$ and $\nabla_{y_1} \psi_2(\cdot, \cdot, \xi) = -A(\xi)^\top \pi_{y_2}(\cdot, \cdot, \xi)$ are Lipschitz continuous over the compact set $X_1 \times Y_1$. By the uniform law of large numbers [34, Theorem 9.60],

$$\lim_{N \rightarrow \infty} \sup_{\substack{(x_1, y_1) \in X_1 \times Y_1 \\ i=1, \dots, n_1+m_1}} \left| \begin{pmatrix} \nabla_{x_1} \Phi_N(x_1, y_1) - \nabla_{x_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] \\ \nabla_{y_1} \Phi_N(x_1, y_1) - \nabla_{y_1} \mathbb{E}[\psi_2(x_1, y_1, \xi)] \end{pmatrix}_i \right| = 0, \text{ a.s.}$$

and then

$$\lim_{N \rightarrow \infty} \sup_{(x_1, y_1) \in X_1 \times Y_1} \mathbb{D} \left(\begin{pmatrix} G_x^N(x_1, y_1) \\ G_y^N(x_1, y_1) \end{pmatrix}, \begin{pmatrix} G_x(x_1, y_1) \\ G_y(x_1, y_1) \end{pmatrix} \right) = 0, \text{ a.s.,}$$

which implies (38). $\square$

¹ A multifunction $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is osc, if for all $\epsilon > 0$, there exists $\delta > 0$ such that $S(x') \subset S(x) + \epsilon B_n$ for all $x' \in x + \delta B_n$, where B_n denotes the unit ball in $\mathbb{R}^n$ [11, 32].

5 The inexact parallel proximal gradient descent ascent algorithm for two-stage stochastic minimax problem

This section presents IPPGDA algorithm for the SAA problem (33) of the two-stage stochastic minimax problem (1)-(2). Let $\bar{f} := f + \mathbf{1}_{X_1}$ and $\bar{g} := g + \mathbf{1}_{Y_1}$, where $\mathbf{1}$ denotes the indicator function. Throughout this section, assume that g is convex and $\psi_1(x_1, \cdot)$ is concave for every $x_1 \in X_1$, and set $\beta_1^y = 1/L_{yy}$, with $L_{yy} > 0$ specified below.

Algorithm 1 IPPGDA algorithm for the two-stage stochastic minimax problem (1)-(2)

Require: initial point (x_1^1, y_1^1) , stepsizes $\beta_1^y = 1/L_{yy}$, β_1^x and sequence $\{\epsilon^k\}$

- 1: **for** $k = 1, 2, \dots$ **do**
- 2: **for** $i = 1, 2, \dots, N$ **do**
- 3: Solve (13) with given (x_1^k, y_1^k) and ξ^i to obtain $\tilde{\mu}^{k,i} := (\tilde{x}_2^{k,i}, \tilde{y}_2^{k,i}, \tilde{\pi}_{x_2}^i, \tilde{\pi}_{y_2}^i)$, such that $\|H(\tilde{\mu}^{k,i}, \xi^i)\| < \epsilon^k$.
- 4: **end for**
- 5: $\tilde{v}_x^k = \frac{1}{N} \sum_{i=1}^N T(\xi^i)^\top \tilde{\pi}_{x_2}^i$
- 6: $\tilde{v}_y^k = -\frac{1}{N} \sum_{i=1}^N A(\xi^i)^\top \tilde{\pi}_{y_2}^i$
- 7: $y_1^{k+1} = \arg \max_{y_1 \in \mathbb{R}^{m_1}} \{\langle \nabla_{y_1} \psi_1(x_1^k, y_1^k) + \tilde{v}_y^k, y_1 - y_1^k \rangle - \bar{g}(y_1) - \frac{1}{2\beta_1^y} \|y_1 - y_1^k\|^2\}$
- 8: $x_1^{k+1} = \arg \min_{x_1 \in \mathbb{R}^{n_1}} \{\bar{f}(x_1) + \langle \nabla_{x_1} \psi_1(x_1^k, y_1^k) + \tilde{v}_x^k, x_1 - x_1^k \rangle + \frac{1}{2\beta_1^x} \|x_1 - x_1^k\|^2\}$
- 9: **end for**

Remark 8 IPPGDA algorithm is an inexact version of PPGDA in [12]. Since the gradient of the second-stage minimax value function $\Phi_N(x_1, y_1)$ cannot be computed exactly, we cannot apply PPGDA. Instead, IPPGDA algorithm finds an inexact solution of problem (13), which defines an inexact gradient of $\Phi_N(x_1, y_1)$ at each step. In Section 5.4, we will show that under suitable conditions, $\|H(\tilde{\mu}^{k,i}, \xi^i)\| < \frac{\delta^k \sqrt{\lambda}}{\max\{\bar{a}, \bar{t}\}}$ implies $\|\tilde{\pi}_{x_2}^i - \pi_{x_2}^i\| \leq \frac{\delta^k}{\bar{t}}$ and $\|\tilde{\pi}_{y_2}^i - \pi_{y_2}^i\| \leq \frac{\delta^k}{\bar{a}}$, where $\bar{a} := \max_{\xi \in \Xi} \|A(\xi)\|$, $\bar{t} := \max_{\xi \in \Xi} \|T(\xi)\|$, $(\pi_{x_2}^i, \pi_{y_2}^i)$ is the π -component of the unique solution to (13). Let

$$\delta_x^k := \tilde{v}_x^k - v_x^k, \quad \delta_y^k := \tilde{v}_y^k - v_y^k,$$

where v_x^k and v_y^k are the true gradients of $\Phi_N(x_1, y_1)$ w.r.t. x_1 and y_1 . Consequently, $\|\delta_x^k\| = \|\tilde{v}_x^k - v_x^k\| \leq \delta^k$ and $\|\delta_y^k\| = \|\tilde{v}_y^k - v_y^k\| \leq \delta^k$. The requirement of $\{\delta^k\}$ for convergence of IPPGDA algorithm will be presented in Lemma 5.

We consider the SAA two-stage stochastic minimax problem (33) as the following minimization problem

$$\min_{x_1} \Psi_N(x_1) := \bar{f}(x_1) + \theta_N(x_1), \quad (40)$$

where

$$\theta_N(x_1) := \max_{y_1} \psi_1(x_1, y_1) + \Phi_N(x_1, y_1) - \bar{g}(y_1),$$

and prove that the subsequence and global convergence of IPPGDA algorithm to a critical point of problem (40). As in [12], we consider the perturbed gradient-like descent sequence as follows:

Definition 9 A sequence $\{(x_1^k, \nu_k)\}_{k \in \mathbb{N}} \subset \text{dom } \Psi_N \times \mathbb{R}_+$ is called a perturbed gradient-like descent sequence if the following conditions hold.

- Condition 1 (Perturbed sufficient decrease property). There exists $c_1 > 0$ such that for every $k \in \mathbb{N}$,

$$c_1(\|x_1^{k+1} - x_1^k\|^2 + \nu_k^2) \leq \left(\Psi_N(x_1^k) + \frac{1}{2}\nu_k^2 \right) - \left(\Psi_N(x_1^{k+1}) + \frac{1}{2}\nu_{k+1}^2 \right). \quad (41)$$

- Condition 2 (Perturbed subgradient lower bound on iterates gap). There exists $c_2 > 0$ such that for every $k \in \mathbb{N}$, one can find $\zeta^{k+1} \in \partial \Psi_N(x_1^{k+1})$, which satisfies

$$\|\zeta^{k+1}\| \leq c_2(\|x_1^k - x_1^{k+1}\| + \nu_k). \quad (42)$$

- Condition 3. Let $\{x_1^k\}_{k \in \mathcal{K} \subset \mathbb{N}}$ be a subsequence that converges to a point $\bar{x}_1^N$. Then

$$\limsup_{k \rightarrow \infty, k \in \mathcal{K}} \Psi_N(x_1^k) \leq \Psi_N(\bar{x}_1^N).$$

The focus of this section is on how to incorporate the error terms $\|\tilde{v}_x^k - v_x^k\|$ and $\|\tilde{v}_y^k - v_y^k\|$ into the perturbed gradient-like descent sequence, satisfying the three conditions in Definition 9, without increasing the computational burden.

In Section 5.1, we establish key estimates for the sequences $\{(x_1^k, y_1^k)\}$. In Section 5.2, we verify that the generated sequence $\{(x_1^k, \nu_k)\}_{k \in \mathbb{N}}$ is a perturbed gradient-like descent sequence. Section 5.3 investigates the Kurdyka–Łojasiewicz (KL) properties of the problem and the subsequence and global convergence of IPPGDA algorithm to a critical point of (40). In Section 5.4, we address the second-stage problem with an inexact solution. Finally, Section 5.5 presents numerical experiments to illustrate the performance of the proposed IPPGDA algorithm and the convergence behavior of the SAA approach.

5.1 The key estimates of IPPGDA algorithm

We consider Lipschitz continuity properties of problem (40) in the following lemma. To this end, in addition to the Lipschitz continuity of $\nabla \psi_1$, we further assume that there exist $L_{xx}^1 > 0$, $L_{xy}^1 > 0$, $L_{yx}^1 > 0$, $L_{yy}^1 > 0$ such that for any $x_1, \bar{x}_1 \in \mathbb{R}^{n_1}$ and $y_1, \bar{y}_1 \in \mathbb{R}^{m_1}$,

$$\begin{aligned} \|\nabla_{x_1} \psi_1(x_1, y_1) - \nabla_{x_1} \psi_1(\bar{x}_1, y_1)\| &\leq L_{xx}^1 \|x_1 - \bar{x}_1\|, \\ \|\nabla_{x_1} \psi_1(x_1, y_1) - \nabla_{x_1} \psi_1(x_1, \bar{y}_1)\| &\leq L_{xy}^1 \|y_1 - \bar{y}_1\|, \\ \|\nabla_{y_1} \psi_1(x_1, y_1) - \nabla_{y_1} \psi_1(\bar{x}_1, y_1)\| &\leq L_{yx}^1 \|x_1 - \bar{x}_1\|, \\ \|\nabla_{y_1} \psi_1(x_1, y_1) - \nabla_{y_1} \psi_1(x_1, \bar{y}_1)\| &\leq L_{yy}^1 \|y_1 - \bar{y}_1\|. \end{aligned}$$

Lemma 2 *Let*

$$\tilde{y}_1^N(x_1) := \arg \max_{y_1 \in Y_1} \psi_1(x_1, y_1) + \Phi_N(x_1, y_1) - \bar{g}(y_1). \quad (43)$$

Under Assumption 1, the following statements hold.

(i) *There exists $L_{yx} > 0$ such that $\tilde{y}_1^N : \mathbb{R}^{n_1} \rightarrow \mathbb{R}^{m_1}$ is (L_{yx}/σ) -Lipschitz continuous.*

(ii) $\nabla \theta_N(x_1) = \nabla_{x_1} \psi_1(x_1, \tilde{y}_1^N(x_1)) + \nabla_{x_1} \Phi_N(x_1, \tilde{y}_1^N(x_1))$.

(iii) *There exists $L_\theta > 0$ such that $\nabla \theta_N(\cdot)$ is L_θ -Lipschitz continuous.*

Proof By Lemma 1 (iii), both

$$\nabla_{x_1} \Phi_N(\cdot, \cdot) = \frac{1}{N} \sum_{i=1}^N T(\xi^i)^\top \pi_{x_2}(\cdot, \cdot, \xi^i)$$

and

$$\nabla_{y_1} \Phi_N(\cdot, \cdot) = \frac{1}{N} \sum_{i=1}^N -A(\xi^i)^\top \pi_{y_2}(\cdot, \cdot, \xi^i)$$

are Lipschitz continuous over $X_1 \times Y_1$. Then there exist $L_{xx}^2 > 0$, $L_{xy}^2 > 0$, $L_{yx}^2 > 0$ and $L_{yy}^2 > 0$ such that for any $x_1, \bar{x}_1 \in X_1$ and $y_1, \bar{y}_1 \in Y_1$,

$$\begin{aligned} \|\nabla_{x_1} \Phi_N(x_1, y_1) - \nabla_{x_1} \Phi_N(\bar{x}_1, y_1)\| &\leq L_{xx}^2 \|x_1 - \bar{x}_1\|, \\ \|\nabla_{x_1} \Phi_N(x_1, y_1) - \nabla_{x_1} \Phi_N(x_1, \bar{y}_1)\| &\leq L_{xy}^2 \|y_1 - \bar{y}_1\|, \\ \|\nabla_{y_1} \Phi_N(x_1, y_1) - \nabla_{y_1} \Phi_N(\bar{x}_1, y_1)\| &\leq L_{yx}^2 \|x_1 - \bar{x}_1\|, \\ \|\nabla_{y_1} \Phi_N(x_1, y_1) - \nabla_{y_1} \Phi_N(x_1, \bar{y}_1)\| &\leq L_{yy}^2 \|y_1 - \bar{y}_1\|. \end{aligned}$$

Let $L_{xx} := L_{xx}^1 + L_{xx}^2$, $L_{xy} := L_{xy}^1 + L_{xy}^2$, $L_{yx} := L_{yx}^1 + L_{yx}^2$, $L_{yy} := L_{yy}^1 + L_{yy}^2$ and $L_\theta = L_{xx} + \frac{L_{xy} L_{yx}}{\sigma}$. Then (i) is from [12, Lemma 1], (ii) is from [12, Proposition 1] and (iii) is from [12, Lemma 2]. $\square$

Then we analyze the y -step (step 7) and establish the inexact relationship between y_1^{k+1} and y_1^k , and the corresponding maximizers $\tilde{y}_1^N(x_1^k)$ and $\tilde{y}_1^N(x_1^{k+1})$.

Lemma 3 *Suppose that Assumption 1 holds. Let $\kappa = L_{yy}/\sigma$ and $\epsilon_y^k = \frac{2\|\delta_y^k\|}{\sigma + L_{yy}}$. Then for any $\beta > 0$ and every $k \in \mathbb{N}$ (with $k \geq 2$ in (45) and (47)), we have*

$$\|y_1^{k+1} - \tilde{y}_1^N(x_1^{k+1})\| \leq \sqrt{\kappa/(\kappa + 1)} \|y_1^k - \tilde{y}_1^N(x_1^k)\| + \frac{L_{yx}}{\sigma} \|x_1^{k+1} - x_1^k\| + \epsilon_y^k, \quad (44)$$

$$\|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\| \leq \sqrt{\kappa/(\kappa + 1)} \left(\|y_1^k - \tilde{y}_1^N(x_1^{k-1})\| + \frac{L_{yx}}{\sigma} \|x_1^k - x_1^{k-1}\| \right) + \epsilon_y^k, \quad (45)$$

$$\begin{aligned} \|y_1^{k+1} - \tilde{y}_1^N(x_1^{k+1})\|^2 &\leq (1 + \beta) \left(\frac{\kappa + 1/2}{\kappa + 1} \|y_1^k - \tilde{y}_1^N(x_1^k)\|^2 \right. \\ &\quad \left. + \frac{(2\kappa + 1)L_{yx}^2}{\sigma^2} \|x_1^{k+1} - x_1^k\|^2 \right) + (1 + \beta^{-1})(\epsilon_y^k)^2, \end{aligned} \quad (46)$$

$$\begin{aligned} \|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\|^2 &\leq (1 + \beta) \frac{\kappa + 1/2}{\kappa + 1} (\|y_1^k - \tilde{y}_1^N(x_1^{k-1})\|^2 + \frac{2\kappa L_{yx}^2}{\sigma^2} \|x_1^k - x_1^{k-1}\|^2) \\ &\quad + (1 + \beta^{-1})(\epsilon_y^k)^2 \end{aligned} \quad (47)$$

Proof Consider the following problem:

$$\min_{y_1 \in \mathbb{R}^{m_1}} \Gamma_k(y_1) := \bar{g}(y_1) - \psi_1(x_1^k, y_1) - \Phi_N(x_1^k, y_1). \quad (48)$$

Then $\tilde{y}_1^N(x_1^k) = \arg \min_{y_1 \in \mathbb{R}^{m_1}} \Gamma_k(y_1)$ is unique. By σ -strong convexity of Γ_k , it follows that

$$\Gamma_k(y_1^{k+1}) - \Gamma_k(\tilde{y}_1^N(x_1^k)) \geq \langle 0, y_1^{k+1} - \tilde{y}_1^N(x_1^k) \rangle + \frac{\sigma}{2} \|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\|^2.$$

Note that $y_1^{k+1} \in Y_1$ and $\tilde{y}_1^N(x_1^k) \in Y_1$. Applying [38, Lemma 3.1] to $\bar{g}(y_1) - \langle \nabla_{y_1} \psi_1(x_1^k, y_1^k) + v_y^k + \delta_y^k, y_1 \rangle$ yields:

$$\begin{aligned} &\frac{2}{L_{yy}} [\bar{g}(y_1^{k+1}) - \bar{g}(\tilde{y}_1^N(x_1^k)) - \langle \nabla_{y_1} \psi_1(x_1^k, y_1^k) + v_y^k + \delta_y^k, y_1^{k+1} - \tilde{y}_1^N(x_1^k) \rangle] \\ &\leq \|y_1^k - \tilde{y}_1^N(x_1^k)\|^2 - \|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\|^2 - \|y_1^{k+1} - y_1^k\|^2. \end{aligned} \quad (49)$$

Applying [38, Lemma 3.3] to $\psi_1(x_1^k, y_1) + \Phi_N(x_1^k, y_1)$ gives:

$$\begin{aligned} &\psi_1(x_1^k, \tilde{y}_1^N(x_1^k)) + \Phi_N(x_1^k, \tilde{y}_1^N(x_1^k)) - (\psi_1(x_1^k, y_1^{k+1}) + \Phi_N(x_1^k, y_1^{k+1})) \\ &\leq \langle -\nabla_{y_1} \psi_1(x_1^k, y_1^k) - v_y^k, y_1^{k+1} - \tilde{y}_1^N(x_1^k) \rangle + \frac{L_{yy}}{2} \|y_1^{k+1} - y_1^k\|^2. \end{aligned} \quad (50)$$

Combining (48)-(50), we obtain:

$$\begin{aligned} &\frac{L_{yy}}{2} (\|y_1^k - \tilde{y}_1^N(x_1^k)\|^2 - \|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\|^2) + \|\delta_y^k\| \|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\| \\ &\geq \Gamma_k(y_1^{k+1}) - \Gamma_k(\tilde{y}_1^N(x_1^k)) \\ &\geq \frac{\sigma}{2} \|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\|^2. \end{aligned}$$

Consequently,

$$\|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\|^2 \leq \frac{L_{yy}}{\sigma + L_{yy}} \|y_1^k - \tilde{y}_1^N(x_1^k)\|^2 + \epsilon_y^k \|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\|,$$

which implies

$$\left(\|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\| - \frac{1}{2}\epsilon_y^k \right)^2 \leq \left(\frac{\sqrt{L_{yy}}}{\sqrt{\sigma} + L_{yy}} \|y_1^k - \tilde{y}_1^N(x_1^k)\| + \frac{1}{2}\epsilon_y^k \right)^2.$$

Then the above inequality implies

$$\|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\| \leq \sqrt{\kappa/(\kappa+1)} \|y_1^k - \tilde{y}_1^N(x_1^k)\| + \epsilon_y^k.$$

Combining the triangle inequality, the Lipschitz continuity of $\tilde{y}_1^N(\cdot)$ (Lemma 2 (i)) and the above inequality, we have (44) and (45) as follows:

$$\begin{aligned} \|y_1^{k+1} - \tilde{y}_1^N(x_1^{k+1})\| &\leq \|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\| + \|\tilde{y}_1^N(x_1^k) - \tilde{y}_1^N(x_1^{k+1})\| \\ &\leq \sqrt{\frac{\kappa}{\kappa+1}} \|y_1^k - \tilde{y}_1^N(x_1^k)\| + \frac{L_{yx}}{\sigma} \|x_1^{k+1} - x_1^k\| + \epsilon_y^k \end{aligned}$$

and

$$\begin{aligned} \|y_1^{k+1} - \tilde{y}_1^N(x_1^k)\| &\leq \sqrt{\frac{\kappa}{\kappa+1}} \|y_1^k - \tilde{y}_1^N(x_1^k)\| + \epsilon_y^k \\ &\leq \sqrt{\frac{\kappa}{\kappa+1}} \left(\|y_1^k - \tilde{y}_1^N(x_1^{k-1})\| + \|\tilde{y}_1^N(x_1^k) - \tilde{y}_1^N(x_1^{k-1})\| \right) + \epsilon_y^k \\ &\leq \sqrt{\frac{\kappa}{\kappa+1}} \left(\|y_1^k - \tilde{y}_1^N(x_1^{k-1})\| + \frac{L_{yx}}{\sigma} \|x_1^k - x_1^{k-1}\| \right) + \epsilon_y^k. \end{aligned}$$

Finally, squaring both sides of (44) and (45), and applying $(a+b)^2 \leq (1+\gamma)a^2 + (1+\gamma^{-1})b^2$ twice with $\gamma = 1/(2\kappa)$ and β , respectively, we obtain (46) and (47). $\square$

Next we consider the x -step (step 8) from view of function value gap and subgradient bound in the following lemma.

Lemma 4 *Suppose that Assumption 1 holds. Let $\{(x_1^k, y_1^k)\}_{k \in \mathbb{N}}$ be the sequence generated by IPPGDA algorithm. Then*

(i) *for every $k \geq 1$, we have*

$$\begin{aligned} \Psi_N(x_1^{k+1}) - \Psi_N(x_1^k) &\leq -\frac{1}{2} \left(\frac{1}{\beta_1^x} - L_{xy}^2 - L_\theta - 1 \right) \|x_1^{k+1} - x_1^k\|^2 \\ &\quad + \frac{1}{2} \|y_1^k - \tilde{y}_1^N(x_1^k)\|^2 + \frac{1}{2} \|\delta_x^k\|^2; \end{aligned} \quad (51)$$

(ii) *there exist $M > 0$ and $\zeta^{k+1} \in \partial\Psi_N(x_1^{k+1})$, which satisfy*

$$\|\zeta^{k+1}\| \leq M(\|x_1^{k+1} - x_1^k\| + \|y_1^k - \tilde{y}_1^N(x_1^k)\|) + \|\delta_x^k\|. \quad (52)$$

Proof The proof of the lemma is similar as [12, Lemmas 7 and 8]. Note that the key difference is that there is a δ_x^k in x -step (step 8) as follows

$$x_1^{k+1} = \arg \min_{x_1 \in \mathbb{R}^{n_1}} \left\{ \bar{f}(x_1) + \langle \nabla_{x_1} \psi_1(x_1^k, y_1^k) + v_x^k + \delta_x^k, x_1 - x_1^k \rangle + \frac{1}{2\beta_x} \|x_1 - x_1^k\|^2 \right\}.$$

This leads to the $\frac{1}{2}\|\delta_x^k\|^2$ in (51) and $\|\delta_x^k\|$ in (52). Since the remainder of the proof is essentially the same as in [12, Lemmas 7 and 8], we omit the details of the proof. $\square$

Remark 9 Lemmas 3 and 4 are extensions of [12, Lemmas 6, 7, and 8]. In [12], the authors analyzed the divergence caused by using an inexact proximal gradient step:

$$y_1^{k+1} = \arg \max_{y_1 \in \mathbb{R}^{m_1}} \left\{ \langle \nabla_{y_1} \psi_1(x_1^k, y_1^k) + v_y^k, y_1 - y_1^k \rangle - \bar{g}(y_1) - \frac{1}{2\beta_y} \|y_1 - y_1^k\|^2 \right\}$$

to approximate the inner maximization (43). Different from the PPGDA in [12], we can not calculate v_y^k in step 7 and v_x^k in step 8 of Algorithm 1 exactly. Instead, we do inexact proximal gradient steps with inexact terms $\tilde{v}_x^k$ and $\tilde{v}_y^k$.

5.2 Perturbed gradient-like descent sequence

In this section, we show that the sequence $\{(x_1^k, \nu_k)\}$ is a perturbed gradient-like descent sequence, where ν_k is defined as

$$\nu_k := \sqrt{s\|y_1^k - \tilde{y}_1^N(x_1^k)\|^2 + (\delta^{k-1})^2}, \quad (53)$$

where $s > 0$, $\delta^k \geq \max\{\|\delta_x^k\|, \|\delta_y^k\|\}$ and $\{\delta^k\} \downarrow 0$. We consider the perturbed sufficient descent property firstly.

Lemma 5 *Let $\{(x_1^k, y_1^k)\}_{k \in \mathbb{N}}$ be the sequence generated by Algorithm 1. Suppose that Assumption 1 holds and there is $\beta \in (0, \frac{1}{2\kappa+1})$ such that*

$$\frac{1}{\beta_x^x} > L_{xy}^2 + L_\theta + 1 + \max \left\{ \bar{\eta}, \frac{(2\kappa+2)\bar{\eta}}{(1-\beta-2\beta\kappa)} \right\} > 0,$$

$$s \in \left(\frac{2\kappa+2}{1-\beta-2\beta\kappa}, \frac{(\frac{1}{\beta_x^x} - L_{xy}^2 - L_\theta - 1)}{\bar{\eta}} \right),$$

and

$$(\delta^{k-1})^2 \geq \frac{\kappa+1}{(1+\beta)(\kappa+1/2)} \left(\frac{4s(1+\beta)}{\beta(\sigma+L_{yy})^2} + 2 \right) (\delta^k)^2, \quad (54)$$

where $\bar{\eta} := \frac{(1+\beta)(2\kappa+1)L_{yx}^2}{\sigma^2}$. Then there exists $c_1 > 0$ such that for every $k \in \mathbb{N}$, (41) (Condition 1 in Definition 9) holds.

Proof For every $k \in \mathbb{N}$, let

$$\Delta_k := \left(\Psi_N(x_1^{k+1}) + \frac{1}{2}\nu_{k+1}^2 \right) - \left(\Psi_N(x_1^k) + \frac{1}{2}\nu_k^2 \right).$$

Then by Lemma 4 (i), we have

$$\Delta_k \leq \frac{1}{2} \left(L_{xy}^2 + L_\theta + 1 - \frac{1}{\beta^x} \right) \|x_1^{k+1} - x_1^k\|^2 + \frac{1}{2}\|\delta_x^k\|^2 + \frac{1}{2s}\nu_k^2 + \frac{1}{2}(\nu_{k+1}^2 - \nu_k^2). \quad (55)$$

Moreover, by applying (46) in Lemma 3, we have

$$\begin{aligned} \nu_{k+1}^2 + \|\delta_x^k\|^2 &= s\|y_1^{k+1} - \tilde{y}_1^N(x_1^{k+1})\|^2 + (\delta^k)^2 + \|\delta_x^k\|^2 \\ &\leq \frac{(1+\beta)(\kappa+1/2)}{\kappa+1} s\|y_1^k - \tilde{y}_1^N(x_1^k)\|^2 + \left(\frac{4s(1+\beta)}{\beta(\sigma+L_{yy})^2} + 2 \right) (\delta^k)^2 \\ &\quad + \bar{\eta}s\|x_1^k - x_1^{k+1}\|^2. \end{aligned} \quad (56)$$

Substituting $(\delta^{k-1})^2$ (in (54)) into (56), we have

$$\begin{aligned} \nu_{k+1}^2 + \|\delta_x^k\|^2 &\leq \frac{(1+\beta)(\kappa+1/2)}{\kappa+1} (s\|y_1^k - \tilde{y}_1^N(x_1^k)\|^2 + (\delta^{k-1})^2) + \bar{\eta}s\|x_1^k - x_1^{k+1}\|^2 \\ &= \frac{(1+\beta)(\kappa+1/2)}{\kappa+1} \nu_k^2 + \bar{\eta}s\|x_1^k - x_1^{k+1}\|^2. \end{aligned} \quad (57)$$

Combining (55) and (57), we obtain

$$\Delta_k \leq -\frac{t_1}{2}\|x_1^{k+1} - x_1^k\|^2 - \frac{t_2}{2}\nu_k^2$$

with

$$t_1 = \frac{1}{\beta^x} - L_{xy}^2 - L_\theta - 1 - \bar{\eta}s$$

and

$$t_2 = 1 - \frac{1}{s} - \frac{(1+\beta)(\kappa+1/2)}{\kappa+1} = \frac{s(\frac{1-\beta}{2} - \beta\kappa) - \kappa - 1}{s(\kappa+1)}.$$

Since $\beta < \frac{1}{1+2\kappa}$ and $s > \frac{2\kappa+2}{1-\beta-2\beta\kappa} > 0$, we have $t_2 > 0$. Moreover, by the assumed stepsize condition

$$\frac{1}{\beta_1^x} > L_{xy}^2 + L_\theta + 1 + \max \left\{ \bar{\eta}, \frac{(2\kappa+2)\bar{\eta}}{(1-\beta-2\beta\kappa)} \right\} > 0,$$

we have $t_1 > 0$, $\frac{2\kappa+2}{1-\beta-2\beta\kappa} < \frac{(\frac{1}{\beta_1^x} - L_{xy}^2 - L_\theta - 1)}{\bar{\eta}}$ and $s \in (\frac{2\kappa+2}{1-\beta-2\beta\kappa}, \frac{(\frac{1}{\beta_1^x} - L_{xy}^2 - L_\theta - 1)}{\bar{\eta}})$. Then we obtain (41) (Condition 1) with $c_1 = \min\{t_1, t_2\}/2$. $\square$

We then consider perturbed subgradient lower bound property (Condition 2 in Definition 9).

Lemma 6 (Perturbed subgradient lower bound) *Suppose that Assumption 1 holds. Let $\{(x_1^k, y_1^k)\}_{k \in \mathbb{N}}$ be the sequence generated by Algorithm 1, and $\{\nu_k\}_{k \in \mathbb{N}}$ be defined in (53). Then, there exists $c_2 > 0$ such that for every $k \in \mathbb{N}$, one can find $\zeta^{k+1} \in \partial\Psi_N(x_1^{k+1})$, which satisfies (42) (Condition 2 in Definition 9).*

Proof By Lemma 4 (ii), there exists $M > 0$ such that for every $k > 0$ and $c_2 = \max\left\{M, \sqrt{\max\{2M^2/s, 2\}}\right\}$, one can find $\zeta^{k+1} \in \partial\Psi_N(x_1^{k+1})$ such that

$$\begin{aligned} \|\zeta^{k+1}\| &\leq M(\|x_1^{k+1} - x_1^k\| + \|y_1^k - \tilde{y}_1^N(x_1^k)\|) + \|\delta_x^k\| \\ &\leq M(\|x_1^{k+1} - x_1^k\|) + \sqrt{2M^2\|y_1^k - \tilde{y}_1^N(x_1^k)\|^2 + 2(\delta^{k-1})^2} \\ &\leq M(\|x_1^{k+1} - x_1^k\|) + \sqrt{\max\left\{\frac{2M^2}{s}, 2\right\}}\nu_k \\ &\leq c_2(\|x_1^{k+1} - x_1^k\| + \nu_k), \end{aligned}$$

where the second inequality follows from $(a+b)^2 \leq (1+\gamma)a^2 + (1+\gamma^{-1})b^2$ and $\|\delta_x^k\| \leq \delta^k \leq \delta^{k-1}$ in (54). $\square$

Then we consider Condition 3 in Definition 9.

Lemma 7 *Suppose that Assumption 1 holds. Let $\{(x_1^k, y_1^k)\}_{k \in \mathbb{N}}$ be the sequence generated by Algorithm 1, and $\{\nu_k\}_{k \in \mathbb{N}}$ be as defined in (53), such that $\{(x_1^k, \nu_k)\}_{k \in \mathbb{N}}$ satisfies Condition 1 in Definition 9. Let $\{x_1^k\}_{k \in \mathcal{K} \subset \mathbb{N}}$ be a subsequence that converges to a point $\bar{x}_1$. Then*

$$\limsup_{k \rightarrow \infty, k \in \mathcal{K}} \Psi_N(x_1^k) \leq \Psi_N(\bar{x}_1).$$

Proof The proof of the lemma is similar to that of [12, Lemma 13]; the only difference is that we need to add a noise term δ_x^k after $\nabla_{x_1}\psi_1(x_1^k, y_1^k) + v_x^k$ (that is $\nabla_u c(u^l, w^l)$ in the proof of [12, Lemma 13]). Note that this change does not alter the proof process, we omit the details of the proof here. $\square$

5.3 Subsequence and global convergence

In the above section, we have verified three conditions in Definition 9, which implies the sequence $\{x_1^k, \nu_k\}$ is a perturbed gradient-like descent sequence. Now we investigate the subsequence and global convergence of IPPGDA algorithm. We start from the KL property of our problem.

Assumption 3 *$f, \psi_1, F_2(\cdot, \cdot, \xi^i) (i = 1, \dots, N)$, and g are semialgebraic, and X_1, Y_1 are semialgebraic sets.*

Lemma 8 *Let $\phi : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ be a semialgebraic function and Q be a semialgebraic set in y -space. Suppose that, for every $x \in \mathbb{R}^n$, the minimum of $\phi(x, \cdot)$ over Q is attained. Then the function $\varphi(\cdot) := \min\{\phi(\cdot, y) : y \in Q\}$ is semialgebraic.*

Proof Since ϕ and Q are semialgebraic, the set

$$\text{epi } \phi = \{(x, y, \alpha) : \alpha \geq \phi(x, y), x \in \mathbb{R}^n, y \in Q\}$$

is semialgebraic. Since the minimum defining $\varphi(x)$ is attained for every x , we have

$$\text{epi } \varphi = \{(x, \alpha) : \exists y \in Q \text{ such that } \alpha \geq \phi(x, y)\},$$

which is the projection of $\text{epi } \phi$ onto the (x, α) -space. By the Tarski–Seidenberg theorem, $\text{epi } \varphi$ is semialgebraic [18, Page 395]. Therefore, φ is a semialgebraic function.

Proposition 6 *Suppose that Assumptions 1 and 3 hold. Then Ψ_N is semialgebraic.*

Proof Note that $\mathbf{1}_{Y_2(y_1, \xi)}(y_2)$ and $\mathbf{1}_{X_2(x_1, \xi)}(x_2)$ are semialgebraic functions w.r.t. (y_1, y_2) and (x_1, x_2) respectively. Moreover, by (11) and (2), for any ξ^i , $i = 1, \dots, N$,

$$f_{21}(x_2, y_1, \xi^i) = \max_{y_2} F_2(x_2, y_2, \xi^i) - \mathbf{1}_{Y_2(y_1, \xi^i)}(y_2)$$

and

$$\psi_2(x_1, y_1, \xi^i) = \min_{x_2} f_{21}(x_2, y_1, \xi^i) + \mathbf{1}_{X_2(x_1, \xi^i)}(x_2).$$

Then by Assumption 3 and Lemma 8, $f_{21}(x_2, y_1, \xi^i)$ is semialgebraic, and then $\psi_2(x_1, y_1, \xi^i)$ is semialgebraic for every ξ^i , $i = 1, \dots, N$.

Since ψ_1 , $\psi_2(\cdot, \cdot, \xi^i)$, $i = 1, \dots, N$, g , and Y_1 are semialgebraic, Lemma 8 implies that θ_N is semialgebraic. Since f and X_1 are semialgebraic, so are $\bar{f} = f + \mathbf{1}_{X_1}$ and Ψ_N . $\square$

Unlike the direct assumption of semialgebraicity for related functions in [12], we must rigorously establish the semialgebraic property of function Ψ_N in our framework.

Finally, we arrive the subsequence and global convergence here.

Theorem 2 *Suppose that Assumption 1 holds. Let $\{(x_1^k, y_1^k)\}_{k \in \mathbb{N}}$ be the sequence generated by Algorithm 1 under the conditions of Lemma 5. Then the following statements hold.*

- (i) *There exists a nonempty and bounded set of cluster points of the sequence $\{x_1^k\}_{k \in \mathbb{N}}$, such that there exists $\Omega \subset \text{crit } \Psi_N$, $\lim_{k \rightarrow \infty} \text{dist}(x_1^k, \Omega) = 0$, and Ψ_N is finite and constant on Ω . Moreover, let $\{x_1^k\}_{k \in \mathcal{K} \subset \mathbb{N}}$ be a subsequence converging to point $\bar{x}_1^N \in X_1$. Then the subsequence $\{y_1^k\}_{k \in \mathcal{K} \subset \mathbb{N}}$ converges to $\bar{y}_1^N(\bar{x}_1^N)$;*
- (ii) *In addition, suppose Assumption 3 holds. Then $\sum_{k=1}^{\infty} \|x_1^{k+1} - x_1^k\| < \infty$ and $\{x_1^k\}_{k \in \mathbb{N}}$ converges to a critical point $\bar{x}_1^N \in \text{crit } \Psi_N$. Moreover, $\{y_1^k\}_{k \in \mathbb{N}}$ converges to $\bar{y}_1^N(\bar{x}_1^N)$.*

Proof Since X_1 is compact, by Lemmas 5-7, the sequence $\{(x_1^k, \nu_k)\}_{k \in \mathbb{N}}$ is a gradient-like descent sequence and the subsequence convergence of $\{x_1^k\}_{k \in \mathbb{N}}$ comes from [12, Lemma 4]. Moreover, under Assumption 3, by Proposition 6, Ψ_N is semialgebraic. The estimates also hold for the limiting subdifferential, so [12, Theorem 1] gives global convergence. Its inclusion in the Clarke subdifferential yields Clarke stationarity.

Moreover, since the sequence $\{(x_1^k, \nu_k)\}_{k \in \mathbb{N}}$ satisfies Condition 1 in Definition 9 and X_1 is compact, by [12, Lemma 3], we have $\nu_k \rightarrow 0$ which implies $s\|y_1^k - \tilde{y}_1^N(x_1^k)\| \rightarrow 0$. Note from Lemma 2 that $\tilde{y}_1^N(\cdot)$ is continuous, we have in part (i),

$$\lim_{k \rightarrow \infty, k \in \mathcal{K}} \|y_1^k - \tilde{y}_1^N(\bar{x}_1^N)\| \leq \lim_{k \rightarrow \infty, k \in \mathcal{K}} (\|y_1^k - \tilde{y}_1^N(x_1^k)\| + \|\tilde{y}_1^N(x_1^k) - \tilde{y}_1^N(\bar{x}_1^N)\|) = 0,$$

the subsequence convergence of $\{y_1^k\}$; and in part (ii),

$$\lim_{k \rightarrow \infty} \|y_1^k - \tilde{y}_1^N(\bar{x}_1^N)\| \leq \lim_{k \rightarrow \infty} (\|y_1^k - \tilde{y}_1^N(x_1^k)\| + \|\tilde{y}_1^N(x_1^k) - \tilde{y}_1^N(\bar{x}_1^N)\|) = 0,$$

the global convergence of $\{y_1^k\}$. $\square$

5.4 A semi-smooth Newton method for problem (13)

By Remark 8, in step 3 of Algorithm 1, for $i = 1, \dots, N$ and every iteration k , our purpose is to find a π -component of an inexact solution of (13), denoted by $(\tilde{\pi}_{x_2^k}^i, \tilde{\pi}_{y_2^k}^i)$, such that $\|\tilde{\pi}_{x_2^k}^i - \pi_{x_2^k}^i\| \leq \frac{\delta^k}{t}$ and $\|\tilde{\pi}_{y_2^k}^i - \pi_{y_2^k}^i\| \leq \frac{\delta^k}{a}$. In this section, we apply a semi-smooth Newton method with the residual tolerance $\epsilon^k := \frac{\delta^k \sqrt{\lambda}}{\max\{t, a\}}$, which, by Theorem 3, guarantees the error bounds required in Remark 8.

For a fixed (x_2, y_2, ξ) , let

$$M_1 \in \begin{pmatrix} \partial_{x_2 x_2}^2 F_2(x_2, y_2, \xi) & \partial_{x_2 y_2}^2 F_2(x_2, y_2, \xi) \\ -\partial_{y_2 x_2}^2 F_2(x_2, y_2, \xi) & -\partial_{y_2 y_2}^2 F_2(x_2, y_2, \xi) \end{pmatrix}, \quad M_2 := \begin{pmatrix} W(\xi) & 0 \\ 0 & B(\xi) \end{pmatrix},$$

and let $U \in \mathbb{R}^{(l_2+s_2) \times (l_2+s_2)}$ be a diagonal matrix with $u_{ii} \in [0, 1]$, $i = 1, \dots, l_2 + s_2$. Note that M_2 is of full row rank. Then every matrix in $\partial_\mu H(\mu, \xi)$ is of the form

$$JH(\mu, \xi) = \begin{pmatrix} M_1 & M_2^\top \\ (U - I)M_2 & U \end{pmatrix}.$$

Note that M_1 and M_2 depend on (x_2, y_2, ξ) and ξ , respectively, but their arguments are omitted for notational simplicity.

For any $\xi \in \Xi$, define

$$\mathcal{J}_c(\xi) := \text{conv} \left(\bigcup_{\mu \in \tilde{X}_2 \times \tilde{Y}_2 \times \mathbb{R}^{l_2+s_2}} \partial_\mu H(\mu, \xi) \right),$$

where “conv” denotes the convex hull. Every $\hat{J} \in \mathcal{J}_c(\xi)$ can therefore be written as

$$\hat{J} = \begin{pmatrix} M_1 & M_2^\top \\ (U - I)M_2 & U \end{pmatrix},$$

where M_1 is a convex combination of the corresponding matrices specified above, and U is diagonal with diagonal entries in $[0, 1]$.

Proposition 7 *Suppose that Assumptions 1–2 hold. There exists $\underline{\lambda} > 0$ such that, for every $\xi \in \Xi$ and every $\hat{J} \in \mathcal{J}_c(\xi)$, $\hat{J}$ is nonsingular and $\|\hat{J}^{-1}\| \leq \frac{1}{\sqrt{\underline{\lambda}}}$.*

Proof Since $F_2(\cdot, \cdot, \xi)$ is σ -strongly convex-strongly concave, its saddle-gradient mapping is σ -strongly monotone, so every M_1 occurring in the representation of $\hat{J} \in \mathcal{J}_c(\xi)$ is nonsingular. By the Schur complement, the determinant of $\hat{J}$ is given by

$$\det(\hat{J}) = \det(M_1) \det(U - (U - I)M_2M_1^{-1}M_2^\top).$$

Furthermore, since M_2 is of full row rank, the strong monotonicity of M_1 implies that $M_2M_1^{-1}M_2^\top$ is positive definite. It then follows from [17, 10] that

$$U - (U - I)M_2M_1^{-1}M_2^\top$$

is nonsingular. Therefore, $\hat{J}$ is nonsingular.

Moreover, by the compactness of $\check{X}_2 \times \check{Y}_2 \times \Xi$, the outer semicontinuity and local boundedness of the Clarke generalized Jacobian, and the above block representation, the matrices in $\mathcal{J}_c(\xi)$ form a compact set as ξ ranges over Ξ . Since every such matrix is nonsingular,

$$0 < \underline{\lambda} := \min_{\substack{\xi \in \Xi \\ \hat{J} \in \mathcal{J}_c(\xi)}} \lambda_{\min}(\hat{J}^\top \hat{J}).$$

Consequently,

$$\|\hat{J}^{-1}\| \leq \frac{1}{\sqrt{\underline{\lambda}}}, \quad \xi \in \Xi, \quad \hat{J} \in \mathcal{J}_c(\xi).$$

This completes the proof. $\square$

Definition 10 [28, 30] The nonlinear function $H(\cdot, \xi)$ is semi-smooth, if $H(\cdot, \xi)$ is locally Lipschitz and for all $d \in \mathbb{R}^{n_2+m_2+l_2+s_2}$ such that the following limit exists:

$$\lim_{G \in \partial_\mu H(\mu+t\tilde{d}, \xi), \tilde{d} \rightarrow d, t \downarrow 0} G\tilde{d}.$$

Assumption 4 *For any $\xi \in \Xi$, $\nabla_{x_2} F_2(\cdot, \cdot, \xi)$ and $\nabla_{y_2} F_2(\cdot, \cdot, \xi)$ are semi-smooth over $\check{X}_2 \times \check{Y}_2$.*

Proposition 8 *Under Assumption 4, for any $\xi \in \Xi$, $H(\cdot, \xi)$ is semi-smooth over $\check{X}_2 \times \check{Y}_2 \times \mathbb{R}^{l_2+s_2}$.*

Proof Since the “min” operator is a semi-smooth operator [28,30], and the corresponding functions in “min” operator are linear functions, $\min(\cdot, h(\xi) - T(\xi)x_1 - W(\xi)\cdot)$ and $\min(\cdot, c(\xi) - A(\xi)y_1 - B(\xi)\cdot)$ are semi-smooth. Combining with Assumption 4, $H(\cdot, \xi)$ is semi-smooth over $\tilde{X}_2 \times \tilde{Y}_2 \times \mathbb{R}^{l_2+s_2}$. $\square$

Then we apply the following generalized-Jacobian-based semismooth Newton iteration [28,30] to solve the system of nonlinear equations (13).

Algorithm 2 Semi-smooth Newton method for nonlinear equation (13)

Require: initial point μ^0 and $t := 0$

- 1: **while** $\|H(\mu^t, \xi)\| \geq \epsilon$ **do**
- 2: Choose $G^t \in \partial_\mu H(\mu^t, \xi)$ and solve $G^t d^t = -H(\mu^t, \xi)$ for d^t .
- 3: Set $\mu^{t+1} := \mu^t + d^t$ and $t := t + 1$.
- 4: **end while**

Ensure: μ^t

The following theorem gives the local convergence and residual control required by IPPGDA for the unit-step iteration in Algorithm 2. In the numerical implementation, damping or backtracking may be used as a globalization device before the local regime is reached.

Theorem 3 *Suppose that Assumptions 1, 2 and 4 hold. Then, for any $\xi \in \Xi$, let μ^* be the unique solution of (13). There exists a neighborhood $\mathcal{U}$ of μ^* such that, if $\mu^0 \in \mathcal{U}$, the following statements hold:*

- (i) *The unit-step sequence generated by Algorithm 2 is well defined, remains in $\mathcal{U}$, and converges superlinearly to μ^* . Moreover, $\{\|H(\mu^t, \xi)\|\}$ converges to 0 superlinearly;*
- (ii) *the entire sequence $\{\mu^t\}$ satisfies*

$$\|\mu^t - \mu^*\| \leq \frac{1}{\sqrt{\lambda}} \|H(\mu^t, \xi)\|,$$

where $\lambda > 0$ is the constant in Proposition 7.

Proof Under Assumptions 1–2, by Lemma 1, for any $\xi \in \Xi$, there exists a unique $\mu^* \in \mathbb{R}^{n_2+m_2+l_2+s_2}$ such that $H(\mu^*, \xi) = 0$.

By the definition of $\mathcal{J}_c(\xi)$, every generalized Jacobian used in Algorithm 2 belongs to $\mathcal{J}_c(\xi)$. Hence, by Proposition 7, it is nonsingular and satisfies

$$\|G(\mu, \xi)^{-1}\| \leq \frac{1}{\sqrt{\lambda}}.$$

Moreover, under Assumption 4, Proposition 8 shows that $H(\cdot, \xi)$ is semi-smooth at μ^* . Therefore, the standard local convergence result for the generalized-Jacobian-based semi-smooth Newton method [30, Corollary 3.4] implies that there exists a neighborhood $\mathcal{U}$ of μ^* such that, whenever $\mu^0 \in \mathcal{U}$, the unit-step

sequence generated by Algorithm 2 is well defined, remains in $\mathcal{U}$, and converges superlinearly to μ^* . The corresponding residual sequence $\{\|H(\mu^t, \xi)\|\}$ also converges to 0 superlinearly. This proves part (i).

We next prove part (ii). By the generalized mean-value theorem for locally Lipschitz vector-valued mappings [19, Remark 2, equations (8)–(9)] and the definition of $\mathcal{J}_c(\xi)$, there exists $\hat{J}^t \in \mathcal{J}_c(\xi)$ such that

$$H(\mu^t, \xi) - H(\mu^*, \xi) = \hat{J}^t(\mu^t - \mu^*).$$

Since $H(\mu^*, \xi) = 0$, Proposition 7 yields

$$\|\mu^t - \mu^*\| \leq \|(\hat{J}^t)^{-1}\| \|H(\mu^t, \xi)\| \leq \frac{1}{\sqrt{\Delta}} \|H(\mu^t, \xi)\|.$$

This proves part (ii). $\square$

Remark 10 In Algorithm 1 and Remark 8, it is sufficient to ensure that

$$\|\tilde{v}_x^k - v_x^k\| \leq \delta^k, \quad \|\tilde{v}_y^k - v_y^k\| \leq \delta^k.$$

Let

$$\tilde{\mu}^{k,i} := (\tilde{x}_2^{k,i}, \tilde{y}_2^{k,i}, \tilde{\pi}_{x_2^k}^i, \tilde{\pi}_{y_2^k}^i)$$

be the approximate solution returned by Algorithm 2 for problem (13) with given (x_1^k, y_1^k) and ξ^i (step 3 in Algorithm 1), and let

$$\mu^{k,i} := (x_2^{k,i}, y_2^{k,i}, \pi_{x_2^k}^i, \pi_{y_2^k}^i)$$

be the true solution.

Suppose that Algorithm 2 is initialized in the local neighborhood specified in Theorem 3. By Theorem 3(ii), if the residual tolerance in Algorithm 2 is chosen as

$$\epsilon^k := \frac{\delta^k \sqrt{\Delta}}{\max\{\bar{t}, \bar{a}\}}, \quad \bar{t} := \max_{\xi \in \Xi} \|T(\xi)\|, \quad \bar{a} := \max_{\xi \in \Xi} \|A(\xi)\|,$$

then, at termination,

$$\|\tilde{\mu}^{k,i} - \mu^{k,i}\| \leq \frac{1}{\sqrt{\Delta}} \|H(\tilde{\mu}^{k,i}, \xi^i)\| < \frac{\delta^k}{\max\{\bar{t}, \bar{a}\}}.$$

Consequently,

$$\|\tilde{\pi}_{x_2^k}^i - \pi_{x_2^k}^i\| \leq \frac{\delta^k}{\max\{\bar{t}, \bar{a}\}}, \quad \|\tilde{\pi}_{y_2^k}^i - \pi_{y_2^k}^i\| \leq \frac{\delta^k}{\max\{\bar{t}, \bar{a}\}}.$$

Therefore,

$$\begin{aligned} \|\tilde{v}_x^k - v_x^k\| &= \left\| \frac{1}{N} \sum_{i=1}^N T(\xi^i)^\top (\tilde{\pi}_{x_2^k}^i - \pi_{x_2^k}^i) \right\| \\ &\leq \frac{1}{N} \sum_{i=1}^N \|T(\xi^i)\| \|\tilde{\pi}_{x_2^k}^i - \pi_{x_2^k}^i\| \leq \frac{\bar{t}}{\max\{\bar{t}, \bar{a}\}} \delta^k \leq \delta^k, \end{aligned}$$

and similarly,

$$\begin{aligned}\|\tilde{v}_y^k - v_y^k\| &= \left\| \frac{1}{N} \sum_{i=1}^N A(\xi^i)^\top (\tilde{\pi}_{y_2^k}^i - \pi_{y_2^k}^i) \right\| \\ &\leq \frac{\bar{a}}{\max\{\bar{t}, \bar{a}\}} \delta^k \leq \delta^k.\end{aligned}$$

Thus, when Algorithm 2 is initialized in the local neighborhood specified in Theorem 3, the stopping rule $\|H(\tilde{\mu}^{k,i}, \xi^i)\| < \epsilon^k$ guarantees the error bounds required in Remark 8.

With Algorithm 2 and Theorem 3, we can replace steps 2-4 in Algorithm 1 with the following procedure:

for $i = 1, 2, \dots, N$ **do**
 solve (13) with given (x_1^k, y_1^k) and ξ^i by Algorithm 2 (semi-smooth Newton method) with $\|H(\mu^t, \xi^i)\| < \epsilon^k$, where $\epsilon^k := \frac{\delta^k \sqrt{\bar{\lambda}}}{\max\{\bar{a}, \bar{t}\}}$, and obtain $\tilde{\mu}^{k,i} = \mu^t$.
end for

The convergence result (Theorem 2) still holds for Algorithm 1 under the conditions of Theorems 2 and 3.

5.5 Numerical experiments

In this section, we present numerical experiments on a two-stage stochastic two-player zero-sum game to validate both our theoretical framework and proposed algorithm. All the numerical experiments are conducted using MATLAB 2022b on a PC with 12th Gen Intel(R) Core(TM) i7-12700 running at 2.10 GHz and 32 GB of RAM.

Two-player two-stage stochastic zero-sum game: Two-stage quadratic SNEPs are investigated under the monotone condition [29, 43, 24]. Here we consider a nonconvex-nonsmooth extension of the two-stage SNEPs, as the two-stage stochastic minimax problem (1)-(2), where

$$F_1(x_1, y_1) := \|x_1\|_1 - \frac{1}{2} x_1^\top Q_1 x_1 + (d_1)^\top x_1 + x_1^\top O_1 y_1 - \frac{1}{2} y_1^\top S_1 y_1 - (t_1)^\top y_1,$$

$$F_2(x_2, y_2, \xi) := \frac{1}{2} x_2^\top Q_2(\xi) x_2 + (d_2(\xi))^\top x_2 + x_2^\top O_2(\xi) y_2 - \frac{1}{2} y_2^\top S_2(\xi) y_2 - (t_2(\xi))^\top y_2,$$

$\xi : \Omega \rightarrow \Xi \subset \mathbb{R}^l$ is a random variable, $X_1 := [\text{lb}, \text{ub}]^{n_1} \subset \mathbb{R}^{n_1}$, $Y_1 = \mathbb{R}^{m_1}$, $Q_1 \in \mathbb{R}^{n_1 \times n_1}$, $O_1 \in \mathbb{R}^{n_1 \times m_1}$, $S_1 \in \mathbb{R}^{m_1 \times m_1}$, $d_1 \in \mathbb{R}^{n_1}$, $t_1 \in \mathbb{R}^{m_1}$,

$$X_2(x_1, \xi) := \{x_2 \in \mathbb{R}^{n_2} : T(\xi)x_1 + W(\xi)x_2 \leq h(\xi)\},$$

$$Y_2(y_1, \xi) := \{y_2 \in \mathbb{R}^{m_2} : A(\xi)y_1 + B(\xi)y_2 \leq c(\xi)\},$$

$Q_2(\xi) \in \mathbb{R}^{n_2 \times n_2}$, $S_2(\xi) \in \mathbb{R}^{m_2 \times m_2}$ are symmetric positive definite matrices, $O_2(\xi) \in \mathbb{R}^{n_2 \times m_2}$, $d_2(\xi) \in \mathbb{R}^{n_2}$, $t_2(\xi) \in \mathbb{R}^{m_2}$, $T(\xi) \in \mathbb{R}^{l_2 \times n_1}$, $W(\xi) \in \mathbb{R}^{l_2 \times n_2}$, $A(\xi) \in \mathbb{R}^{s_2 \times m_1}$, $B(\xi) \in \mathbb{R}^{s_2 \times m_2}$, $h(\xi) \in \mathbb{R}^{l_2}$, $c(\xi) \in \mathbb{R}^{s_2}$ for a.e. $\xi \in \Xi$.

The setting of this two-stage stochastic minimax problem is as follows.

- First stage setting: $n_1 = 3, m_1 = 2, Q_1 = 0.1I_{n_1}, S_1 = I_{m_1}, O_1 \in \mathbb{R}^{n_1 \times m_1}, d_1 \in \mathbb{R}^{n_1}, t_1 \in \mathbb{R}^{m_1}$ with all elements randomly generated from a uniform distribution over $[0, 1]$.
- Fixed matrices and vectors in second-stage problems: $n_2 = 4, m_2 = 3, l_2 = s_2 = 2, \bar{Q}_2 = \text{diag}(1, 2, 3, 4), \bar{S}_2 = I_{m_2}, \bar{h} = (0.1, 0.1)^\top, \bar{c} = (0.1, 0.1)^\top, \bar{O}_2, \bar{T}, \bar{A}, \bar{d}_2, \bar{t}_2$ are randomly generated matrices or vectors with corresponding dimensions, where each element is drawn from a uniform distribution over $[0, 1]$.
- Random matrices and vectors in second-stage problems: A 49-dimensional random vector ξ , where each component is independently sampled from a uniform distribution over $[-1, 1]$. The first $\frac{n_2 \times (n_2 + 1)}{2}$ elements form the upper triangular part of the random symmetric matrix $\tilde{Q}_2(\xi) \in \mathbb{R}^{n_2 \times n_2}$; the next $\frac{m_2 \times (m_2 + 1)}{2}$ elements form the upper triangular part of the random symmetric matrix $\tilde{S}_2(\xi) \in \mathbb{R}^{m_2 \times m_2}$; the remaining elements construct random matrices and vectors $\tilde{O}_2(\xi), \tilde{T}(\xi), \tilde{A}(\xi), \tilde{d}_2(\xi), \tilde{t}_2(\xi), \tilde{h}(\xi), \tilde{c}(\xi)$. For each experimental setup, we generate N i.i.d. samples.
- Second-stage setting: $Q_2(\xi) = \tau \bar{Q}_2 + 0.1 \tilde{Q}_2(\xi), S_2(\xi) = \tau \bar{S}_2 + 0.1 \tilde{S}_2(\xi), O_2(\xi) = \bar{O}_2 + 0.1 \tilde{O}_2(\xi), W(\xi) = (I_{l_2}, \mathbf{0}), B(\xi) = (I_{s_2}, \mathbf{0}), h(\xi) = \bar{h} + 0.1 \tilde{h}(\xi), c(\xi) = \bar{c} + 0.1 \tilde{c}(\xi), T(\xi) = \bar{T} + 0.1 \tilde{T}(\xi), A(\xi) = \bar{A} + 0.1 \tilde{A}(\xi), d_2(\xi) = \bar{d}_2 + 0.1 \tilde{d}_2(\xi), t_2(\xi) = \bar{t}_2 + 0.1 \tilde{t}_2(\xi)$, where $\mathbf{0}$ denotes the zero matrix of appropriate dimension as specified by the context.
- The residual value (Res.val) is defined as follows:

$$\begin{aligned} \text{Res.val}(x_1^k, y_1^k) := & \| -S_1 y_1^k - t_1 + O_1^\top x_1^k + \tilde{v}_y^k \| \\ & + \min_{\eta \in \partial \|x_1^k\|_1} \|x_1^k - \text{mid}(x_1^k - \eta - w^k, \text{ub}, \text{lb})\|, \end{aligned}$$

where $w^k = (d_1 + O_1 y_1^k + \tilde{v}_x^k - Q_1 x_1^k)$ and “mid” denotes the component-wise median operator on three vectors. Since both the ℓ_1 -subdifferential and the box projection are separable across coordinates, the minimization over $\eta \in \partial \|x_1^k\|_1$ in the residual can be computed componentwise. In our computation, we use a minimizer η^k in the following form:

$$(\eta^k)_i = \begin{cases} -1, & \text{if } (x_1^k)_i < 0 \\ 1, & \text{if } (x_1^k)_i > 0 \\ \text{mid}(-1, -w_i^k, 1), & \text{if } (x_1^k)_i = 0. \end{cases}$$

- The termination criterion is set as $\text{Res.val} \leq 10^{-4}$.

Although $Y_1 = \mathbb{R}^{m_1}$ in the numerical model, the first-stage objective is strongly concave in y_1 and tends to $-\infty$ as $\|y_1\| \rightarrow \infty$, since $S_1 = I_{m_1}$. Hence, all relevant maximizers lie in a sufficiently large compact set, and the numerical problem can equivalently be viewed on this compact restriction of Y_1 .

Experiment 1: We set the sample size to $N = 50$, with lower bound $\text{lb} = -10$ and upper bound $\text{ub} = 10$. Five initial points (x_1^0, y_1^0) are randomly generated with x_1^0 components uniform on $[7, 10]$ and y_1^0 components uniform

on $[0, 1]$. For $\tau = 0.1$ and 0.5 , we randomly generated five problems with five different initial points per τ -value. The convergence behavior of IPPGDA algorithm regarding the residual values is shown in Fig. 1.

Note that $Q_2(\xi) = \tau\bar{Q}_2 + 0.1\tilde{Q}_2(\xi)$, $S_2(\xi) = \tau\bar{S}_2 + 0.1\tilde{S}_2(\xi)$. Fig. 1 demonstrates that smaller τ values lead to more divergent convergence paths, attributable to stronger stochastic influences at lower τ values. Moreover, when $\tau \leq 0.1$, $Q_2(\xi)$ and $S_2(\xi)$ may lose positive definiteness, leading to potential algorithmic non-convergence. Nevertheless, the algorithm performs well in the reported experiments with $\tau = 0.1$, although positive definiteness is not guaranteed by the random matrix construction.

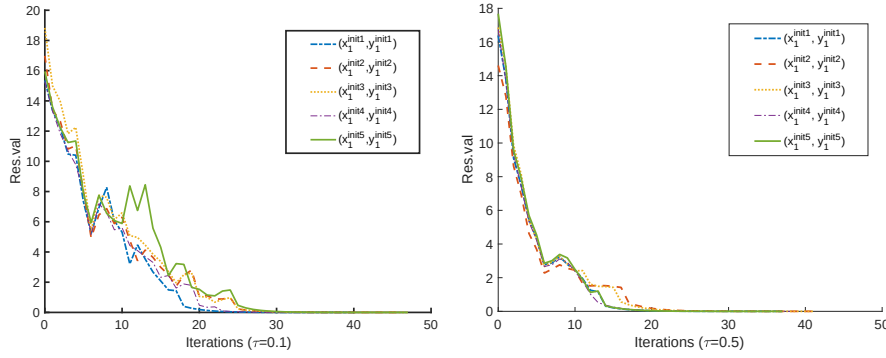


Fig. 1: Residual values versus iteration numbers with different starting points and random matrices for $\tau = 0.1$ and 0.5 , respectively

Experiment 2: We set $\tau = 0.5$ throughout this experiment and conduct numerical experiments with sample sizes $N = 10, 50, 200, 500, 1000, 3000$. The feasible set X_1 is defined by the box constraints $[-10, 10]^3$ and $[-20, 20]^3$. To make the effect of the two feasible sets more visible, we generate d_1 by $d_1 = 15 \text{rand}(n_1, 1)$ in this experiment. For each combination of sample sizes and constraint sets, we randomly generate 30 instances of the two-stage stochastic minimax problem and solve them using IPPGDA algorithm. The resulting minimax values are summarized by box plots in Fig. 2. For each sample size, the box plot is based on 30 random instances: the red line denotes the median, the lower and upper edges of the box represent the 25th and 75th percentiles, respectively, the whiskers indicate the non-outlier range, and the red crosses denote outliers.

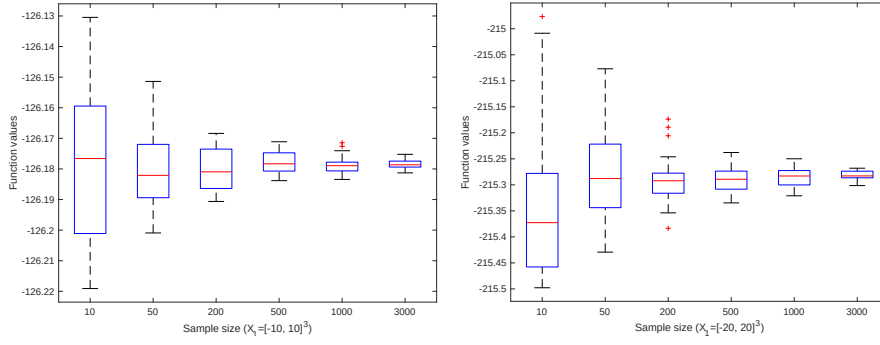


Fig. 2: Box plots of the SAA minimax values over 30 random instances for different sample sizes when $X_1 = [-10, 10]^3$ and $X_1 = [-20, 20]^3$, respectively

Fig. 2 shows that the SAA problems exhibit convergent behavior as the sample size increases, which is consistent with theoretical expectations. Moreover, a larger feasible set X_1 leads to a smaller objective function value, which also aligns with our expectations.

To quantify the convergence behavior of the SAA solutions, we compute a reference minimax value using 10,000 samples and treat it as a surrogate true value. For each sample size N , we calculate the absolute difference between the minimax value obtained from each of the 30 random instances and this reference value, and then report the average of these 30 absolute differences. The resulting average absolute errors are shown in Fig. 3. Since the sample sizes range from 10 to 3000, we use a logarithmic scale on the horizontal axis and explicitly label each sample size. The decreasing trend in both panels indicates that the SAA minimax values become closer to the reference value as the sample size increases.

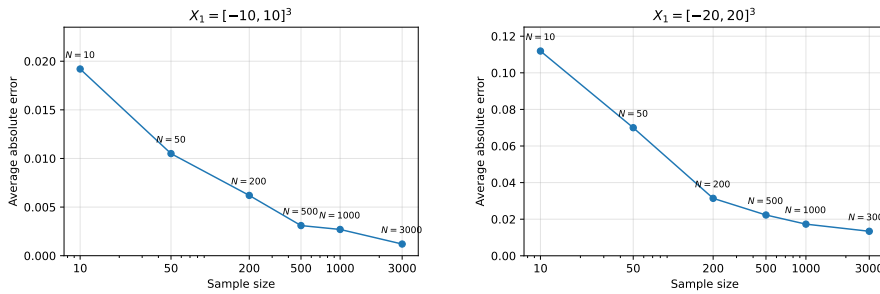


Fig. 3: Average absolute errors of the SAA minimax values for different sample sizes.

To further illustrate the computational effort for different sample sizes, we report the means and standard deviations of the outer iteration counts and CPU times of IPPGDA over the 30 random instances in Table 1. The CPU time is measured in seconds on the computing platform described above.

Table 1: Mean $\pm$ standard deviation of the outer iteration counts and CPU times of IPPGDA over 30 random instances.

Sample size N	$X_1 = [-10, 10]^3$		$X_1 = [-20, 20]^3$	
	Time (s)		Time (s)	
	Iter. (mean $\pm$ std.)	(mean $\pm$ std.)	Iter. (mean $\pm$ std.)	(mean $\pm$ std.)
10	42.8333 $\pm$ 0.3790	0.0246 $\pm$ 0.0018	78.1667 $\pm$ 5.5466	0.0456 $\pm$ 0.0031
50	42.8000 $\pm$ 0.4068	0.1080 $\pm$ 0.0019	78.7333 $\pm$ 2.0833	0.2160 $\pm$ 0.0159
200	42.9000 $\pm$ 0.3051	0.4269 $\pm$ 0.0048	78.5667 $\pm$ 1.5013	0.7892 $\pm$ 0.0338
500	42.9333 $\pm$ 0.2537	1.0744 $\pm$ 0.0298	78.3333 $\pm$ 0.7112	1.9554 $\pm$ 0.0694
1000	42.8333 $\pm$ 0.3790	2.1618 $\pm$ 0.0672	78.5000 $\pm$ 0.6297	3.9399 $\pm$ 0.1395
3000	42.8000 $\pm$ 0.4068	6.4793 $\pm$ 0.1493	78.4000 $\pm$ 0.5632	11.5274 $\pm$ 0.3218

As expected, the average CPU time increases with the sample size, since each outer iteration requires solving more second-stage minimax subproblems. The means and standard deviations of both the CPU time and the outer iteration count over the 30 instances are reported in Table 1.

6 Conclusion

In this paper, we introduce the two-stage stochastic minimax problem (1)-(2), analyze the Lipschitz continuity of second-stage minimax value function and solution functions, along with the properties and relationships among saddle points, minimax points and KKT points. We further prove the convergence of the SAA method for problem (1)-(2). To solve the SAA problem, we propose an IPPGDA algorithm. The algorithm utilizes a semi-smooth Newton approach to solve second-stage subproblems, obtaining approximate gradients of the second-stage minimax value function which are subsequently integrated into an inexact first-stage proximal gradient scheme for the minimax problem. Numerical experiments demonstrate the effectiveness of the IPPGDA algorithm and validate the convergence properties of the SAA approach.

Several broader research directions remain open. One is to develop a theoretical framework beyond the strongly convex–strongly concave setting, allowing nonunique second-stage solutions, nonsmooth value functions, and more general nonconvex–nonconcave structures. Another is, building on the large-deviation analysis for SAA in [36], to establish a nonasymptotic statistical theory for SAA, including exponential deviation bounds, confidence guarantees, and sample-complexity estimates under suitable tail conditions. On the algorithmic side, correction-step semismooth Newton methods, such as that of Feng, Ding and Li [16], may allow weaker generalized-Jacobian regularity, while scalable methods permitting inexact solution of second-stage problems

also deserve further study. Further extensions to multistage and distributionally robust minimax models with nonlinear or coupled recourse constraints also constitute important research directions.

Acknowledgements The authors are grateful to the editor and the two anonymous reviewers for their constructive comments and valuable suggestions, which have helped improve the quality and presentation of this paper. This research is supported by the National Key R&D Program of China (No.2023YFA1009300), NSFC/RGC N_PolyU507/22, Hong Kong Research Grant Council project PolyU15300124 and the National Natural Science Foundation of China (No. 12261160365).

Conflict of interest

The authors declare that they have no conflict of interest.

7 Appendix: A generalization of Theorem 2G.8 in [14] (implicit function theorem for stationary points)

In [14], Dontchev and Rockafellar gave Theorem 2G.8, an implicit function theorem for stationary points for the parametric nonlinear programming problem in the form

$$\begin{aligned} \min_z \quad & h_0(p, z) \\ \text{s.t.} \quad & h_i(p, z) \leq 0, \quad i \in [1, s], \\ & h_i(p, z) = 0, \quad i \in [s+1, m] \end{aligned} \quad (58)$$

with parameter p , where $h_i : \mathbb{R}^k \times \mathbb{R}^n \rightarrow \mathbb{R}$ is twice continuously differentiable, $i = 0, \dots, m$. Let

$$L(p, z, \tau) = h_0(p, z) + \tau_1 h_1(p, z) + \dots + \tau_m h_m(p, z).$$

For a fixed p , the variational inequality capturing the associated first-order conditions of problem (58) is

$$(0, 0) \in H(p, z, \tau) + N_E(z, \tau), \quad (59)$$

where

$$H(p, z, \tau) = (\nabla_z L(p, z, \tau)^\top, -\nabla_\tau L(p, z, \tau)^\top)^\top, \quad E = \mathbb{R}^n \times \mathbb{R}_+^s \times \mathbb{R}^{m-s}.$$

Let

$$S(p) := \{(z, \tau) | (0, 0) \in H(p, z, \tau) + N_E(z, \tau)\} \quad (60)$$

be the solution mapping of the generalized equation (59) and assume that $S(p)$ is nonempty in a neighborhood of $\bar{p}$.

To study the two-stage stochastic minimax problem (1)-(2), we need an implicit function theorem for problems (11) and (17), where the objective functions F_2 and f_{21} are continuously differentiable but not twice continuously differentiable. Now we give implicit function theorem for (58) where h_i is

continuously differentiable and $\nabla_z h_i$ is Lipschitz continuous, $i = 0, \dots, m$. The theorem is extended from [14, Theorem 2G.8], and weakens the twice continuous differentiability of h_i to the continuous differentiability of h_i .

Throughout this appendix, $\partial_{zz}^2 L$ and the partial generalized Jacobian of H are understood as the corresponding column-block projections of the joint Clarke generalized Jacobians of $\nabla_z L$ and H , respectively, with the parameter p included among the joint variables.

For any $(\bar{z}, \bar{\tau}) \in S(\bar{p})$, we give an auxiliary problem of (58). Let $\bar{G} \in \partial_{zz}^2 L(\bar{p}, \bar{z}, \bar{\tau})$,

$$\bar{h}_0(w, \bar{G}) := L(\bar{p}, \bar{z}, \bar{\tau}) + \langle \nabla_z L(\bar{p}, \bar{z}, \bar{\tau}), w \rangle + \frac{1}{2} \langle w, \bar{G} w \rangle$$

and

$$\bar{h}_i(w) := h_i(\bar{p}, \bar{z}) + \langle \nabla_z h_i(\bar{p}, \bar{z}), w \rangle, \quad \forall i = 1, \dots, m.$$

Then the auxiliary problem of (58) with parameters ν and $u_1, \dots, u_m$ is

$$\begin{aligned} \min_w \quad & \bar{h}_0(w, \bar{G}) - \langle \nu, w \rangle \\ \text{s.t.} \quad & \bar{h}_i(w) + u_i \leq 0, \quad i \in I_0, \\ & \bar{h}_i(w) + u_i = 0, \quad i \in I/I_0, \\ & \bar{h}_i(w) + u_i \text{ free}, \quad i \in I_1, \end{aligned}$$

where

$$\begin{aligned} I &= \{i \in [1, m] : h_i(\bar{p}, \bar{z}) = 0\} \supset \{s+1, \dots, m\}, \\ I_0 &= \{i \in [1, s] : h_i(\bar{p}, \bar{z}) = 0 \text{ and } \bar{\tau}_i = 0\}, \\ I_1 &= \{i \in [1, s] : h_i(\bar{p}, \bar{z}) < 0\}. \end{aligned}$$

For any $\bar{G} \in \partial_{zz}^2 L(\bar{p}, \bar{z}, \bar{\tau})$, let

$$\bar{L}(w, \varsigma, \bar{G}) - \langle \nu, w \rangle + \langle \varsigma, u \rangle := \bar{h}_0(w, \bar{G}) - \langle \nu, w \rangle + \varsigma_1(\bar{h}_1(w) + u_1) + \dots + \varsigma_m(\bar{h}_m(w) + u_m).$$

The corresponding first-order conditions are given by the variational inequality

$$(0, 0) \in (\nabla_w \bar{L}(w, \varsigma, \bar{G}), -\nabla_\varsigma \bar{L}(w, \varsigma, \bar{G})) - (\nu, u) + N_{\bar{E}}(w, \varsigma),$$

where $\bar{E} = \mathbb{R}^n \times \{\varsigma \in \mathbb{R}^m : \varsigma_i \geq 0, \forall i \in I_0 \text{ and } \varsigma_i = 0, \forall i \in I_1\}$. Let

$$\begin{aligned} M^+ &= \{w \in \mathbb{R}^n : w \perp \nabla_z h_i(\bar{p}, \bar{z}) \text{ for all } i \in I \setminus I_0\}, \\ M^- &= \{w \in \mathbb{R}^n : w \perp \nabla_z h_i(\bar{p}, \bar{z}) \text{ for all } i \in I\}, \end{aligned} \tag{61}$$

$$\bar{S}(\nu, u, \bar{G}) := \{(w, \varsigma) : (0, 0) \in (\nabla_w \bar{L}(w, \varsigma, \bar{G}), -\nabla_\varsigma \bar{L}(w, \varsigma, \bar{G})) - (\nu, u) + N_{\bar{E}}(w, \varsigma)\}.$$

Now we present the implicit function theorem for stationary points of (58).

Theorem 4 *Let $(\bar{z}, \bar{\tau}) \in S(\bar{p})$ for the mapping S in (60), constructed from continuously differentiable functions h_i such that $\nabla_z h_i$ is Lipschitz continuous. Assume*

(A1): *For any $\bar{G} \in \partial_{zz}^2 L(\bar{p}, \bar{z}, \bar{\tau})$, $\bar{S}(\cdot, \cdot, \bar{G})$ has a Lipschitz continuous single-valued localization $\bar{s}$ around $(0, 0)$ for $(0, 0)$.*

Then S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $(\bar{z}, \bar{\tau})$.

In particular, condition (A1) holds when the following conditions are both fulfilled:

- (a) the gradients $\nabla_z h_i(\bar{p}, \bar{z})$, $i \in I$, are linearly independent;
 (b) for every $\bar{G} \in \partial_{zz}^2 L(\bar{p}, \bar{z}, \bar{\tau})$,

$$\langle w, \bar{G}w \rangle > 0$$

for every nonzero $w \in M^+$ with $\bar{G}w \perp M^-$, where M^+ and M^- are defined in (61).

On the other hand, condition (A1) always entails condition (a).

Proof Let $h = (h_1, \dots, h_m)$. Since $\nabla_\tau L(p, z, \tau) = h(p, z)$, the partial Clarke generalized Jacobian of H with respect to (z, τ) at $(\bar{p}, \bar{z}, \bar{\tau})$ consists of matrices

$$J\bar{H}(\bar{G}) := \begin{pmatrix} \bar{G} & \nabla_z h(\bar{p}, \bar{z})^\top \\ -\nabla_z h(\bar{p}, \bar{z}) & 0 \end{pmatrix}, \quad \bar{G} \in \partial_{zz}^2 L(\bar{p}, \bar{z}, \bar{\tau}).$$

Moreover, the critical cone of E at $(\bar{z}, \bar{\tau})$ relative to $-H(\bar{p}, \bar{z}, \bar{\tau})$ is

$$K_E(\bar{z}, \bar{\tau}, -H(\bar{p}, \bar{z}, \bar{\tau})) = \bar{E}.$$

By the reduction result for polyhedral variational inequalities in [14, Theorem 2E.6], after translating $(z, \tau) = (\bar{z}, \bar{\tau}) + (w, \varsigma)$, the partially linearized generalized equation corresponding to $J\bar{H}(\bar{G})$ is precisely the variational inequality defining $\bar{S}(\nu, u, \bar{G})$. Hence condition (A1) means that every generalized-Jacobian linearization of (59) is strongly regular. Thus, (59) is parametrically CD-regular in the sense of [19, Definition 4]. Since the Lipschitz continuity of $\nabla_z h_i$ implies that H is locally Lipschitz continuous, the first assertion follows from [19, Theorem 4].

It remains to verify the last two assertions. Fix any $\bar{G} \in \partial_{zz}^2 L(\bar{p}, \bar{z}, \bar{\tau})$. Once $\bar{G}$ is fixed, the auxiliary problem is a quadratic program with affine constraints, and its first-order system is an affine-polyhedral variational inequality. Therefore, the argument in [14, Theorem 2G.8], based on [14, Theorem 2E.8], applies with $\nabla_{zz}^2 L(\bar{p}, \bar{z}, \bar{\tau})$ replaced by $\bar{G}$. Consequently, conditions (a) and (b) imply that $\bar{S}(\cdot, \cdot, \bar{G})$ has a Lipschitz continuous single-valued localization around $(0, 0)$ for $(0, 0)$. Since $\bar{G}$ was arbitrary, condition (A1) follows.

Finally, the proof that condition (A1) implies condition (a) in [14, Theorem 2G.8] uses only the local single-valuedness of the auxiliary affine variational inequality and the active constraint gradients. It does not use the existence or continuity of a classical Hessian and therefore applies without modification. $\square$

Unlike [14, Theorem 2G.8], we do not claim the necessity of condition (A1), since the additional full-row-rank assumption on the parameter Jacobian required for the ample parameterization is not imposed here.

References

1. Beale, E.M.: On minimizing a convex function subject to linear inequalities. J. R. Stat. Soc. Ser. B. Stat. Methodol. **17**, 173–184 (1955)

2. Bian, W., Chen, X.: Nonsmooth convex–concave saddle point problems with cardinality penalties. *Math. Program.* **217**, 221–267 (2026)
3. Birge, J.R., Louveaux, F.: *Introduction to Stochastic Programming*. Springer (1997)
4. Bonnans, J.F., Shapiro, A.: *Perturbation Analysis of Optimization Problems*. Springer Science & Business Media (2013)
5. Chen, L., Liu, Y., Yang, X., Zhang, J.: Stochastic approximation methods for the two-stage stochastic linear complementarity problem. *SIAM J. Optim.* **32**, 2129–2155 (2022)
6. Chen, L., Luo, L.: Near-optimal algorithms for making the gradient small in stochastic minimax optimization. *J. Mach. Learn. Res.* **25**, 1–44 (2024)
7. Chen, X., Pong, T.K., Wets, R.J.: Two-stage stochastic variational inequalities: an ERM-solution procedure. *Math. Program.* **165**, 71–111 (2017)
8. Chen, X., Shapiro, A., Sun, H.: Convergence analysis of sample average approximation of two-stage stochastic generalized equations. *SIAM J. Optim.* **29**, 135–161 (2019)
9. Chen, X., Sun, H., Xu, H.: Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems. *Math. Program.* **177**, 255–289 (2019)
10. Chen, X., Xiang, S.: Computation of error bounds for P-matrix linear complementarity problems. *Math. Program.* **106**, 513–525 (2006)
11. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. SIAM (1990)
12. Cohen, E., Teboulle, M.: Alternating and parallel proximal gradient methods for non-smooth, nonconvex minimax: a unified convergence analysis. *Math. Oper. Res.* **50**, 141–168 (2025)
13. Dantzig, G.B.: Linear programming under uncertainty. *Manag. Sci.* **1**, 197–206 (1955)
14. Dontchev, A.L., Rockafellar, R.T.: *Implicit Functions and Solution Mappings*, vol. 543. Springer (2009)
15. Facchinei, F., Pang, J.S.: *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer (2003)
16. Feng, F., Ding, C., Li, X.: A quadratically convergent semismooth newton method for nonlinear semidefinite programming without generalized jacobian regularity. *Mathematical Programming* pp. 1–41 (2025)
17. Gabriel, S.A., Moré, J.J.: Smoothing of mixed complementarity problems. *Complementarity and Variational Problems: State of the Art* **92**, 105–116 (1997)
18. Ioffe, A.D.: *Variational Analysis of Regular Mappings*. Springer (2017)
19. Izmailov, A.F.: Strongly regular nonsmooth generalized equations. *Math. Program.* **147**, 581–590 (2014)
20. Jiang, J., Chen, X.: Optimality conditions for nonsmooth nonconvex-nonconcave min-max problems and generative adversarial networks. *SIAM J. Math. Data Sci.* **5**, 693–722 (2023)
21. Jin, C., Netrapalli, P., Jordan, M.: What is local optimality in nonconvex-nonconcave minimax optimization? In: *International Conference on Machine Learning*, pp. 4880–4889. PMLR (2020)
22. Lan, G., Li, Y.: A novel catalyst scheme for stochastic minimax optimization. *arXiv preprint arXiv:2311.02814* (2023)
23. Lee, W.: Resource allocation for multi-channel underlay cognitive radio network based on deep neural network. *IEEE Commun. Lett.* **22**, 1942–1945 (2018)
24. Lei, J., Shanbhag, U.V., Pang, J.S., Sen, S.: On synchronous, asynchronous, and randomized best-response schemes for stochastic Nash games. *Math. Oper. Res.* **45**, 157–190 (2020)
25. Liu, A., Yang, R., Quek, T.Q., Zhao, M.J.: Two-stage stochastic optimization via primal-dual decomposition and deep unrolling. *IEEE Trans. Signal Process.* **69**, 3000–3015 (2021)
26. Liu, C., Fan, Y., Ordóñez, F.: A two-stage stochastic programming model for transportation network protection. *Comput. Oper. Res.* **36**, 1582–1590 (2009)
27. v. Neumann, J.: Zur theorie der gesellschaftsspiele. *Math. Ann.* **100**, 295–320 (1928)
28. Pang, J.S.: Newton’s method for B-differentiable equations. *Math. Oper. Res.* **15**, 311–341 (1990)
29. Pang, J.S., Sen, S., Shanbhag, U.V.: Two-stage non-cooperative games with risk-averse players. *Math. Program.* **165**, 235–290 (2017)

30. Qi, L.: Convergence analysis of some algorithms for solving nonsmooth equations. *Math. Oper. Res.* **18**, 227–244 (1993)
31. Rockafellar, R.T., Sun, J.: Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging. *Math. Program.* **174**, 453–471 (2019)
32. Rockafellar, R.T., Wets, R.J.: *Variational Analysis*. Springer (1998)
33. Rockafellar, R.T., Wets, R.J.: Stochastic variational inequalities: single-stage to multi-stage. *Math. Program.* **165**, 331–360 (2017)
34. Shapiro, A., Dentcheva, D., Ruszczyński, A.: *Lectures on Stochastic Programming: Modeling and Theory*. SIAM (2021)
35. Shapiro, A., Kleywegt, A.: Minimax analysis of stochastic problems. *Optim. Methods Softw.* **17**, 523–542 (2002)
36. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. *Optimization* **57**, 395–418 (2008)
37. Sion, M.: On general minimax theorems. *Pacific J. Math.* **8**, 171–176 (1958)
38. Teboulle, M.: A simplified view of first order methods for optimization. *Math. Program.* **170**, 67–96 (2018)
39. Wachsmuth, G.: On LICQ and the uniqueness of Lagrange multipliers. *Oper. Res. Lett.* **41**, 78–80 (2013)
40. Washburn, A.R.: *Two-person Zero-sum Games*. Springer (2014)
41. Xu, H.: Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. *J. Math. Anal. Appl.* **368**, 692–710 (2010)
42. Xu, Z., Zhang, H., Xu, Y., Lan, G.: A unified single-loop alternating gradient projection algorithm for nonconvex-concave and convex-nonconcave minimax problems. *Math. Program.* **201**, 635–706 (2023)
43. Zhang, M., Sun, J., Xu, H.: Two-stage quadratic games under uncertainty and their solution by progressive hedging algorithms. *SIAM J. Optim.* **29**, 1799–1818 (2019)