Distributionally Robust Stochastic Variational Inequalities

Hailin Sun[*] Alexander Shapiro[†] Xiaojun Chen[‡]

October 31, 2020

Abstract: We propose a formulation of Distributionally Robust Variational Inequalities (DRVI) to deal with uncertainties of distributions of the involved random variables in variational inequalities. Examples of the DRVI are provided, including the optimality conditions for distributionally robust optimization and distributionally robust games. The existence of solutions and monotonicity of the DRVI are discussed. Moreover, we propose a sample average approximation (SAA) approach to the DRVI and study its convergence properties. Numerical examples of distributionally robust games are presented to illustrate solutions of the DRVI and convergence properties of the SAA approach.

Keywords: Distributional robustness, variational inequalities, monotonicity, sample average approximation, stochastic games

AMS Classification: 90C33, 90C15

1 Introduction

Let $X \subseteq \mathbb{R}^n$ be a nonempty closed convex set and $\mathcal{N}_X(x)$ be the normal cone to X at $x \in \mathbb{R}^n$ (note that $\mathcal{N}_X(x) = \emptyset$ if $x \notin X$). Let $\xi \in \mathbb{R}^\ell$ be a random vector with support set $\Xi \subset \mathbb{R}^\ell$ equipped with its Borel sigma algebra $\mathcal{B}$ and probability distribution P. We consider the stochastic variational inequalities (SVI)

$$0 \in \mathbb{E}_P[\Phi(x, \xi)] + \mathcal{N}_X(x), \tag{1.1}$$

where $\Phi : X \times \Xi \to \mathbb{R}^n$ is such that the corresponding expectation is well defined. By writing $\mathbb{E}_P$ we emphasize that the expectation is taken with respect to a considered probability measure (distribution) P on $(\Xi, \mathcal{B})$. With some abuse of the notation we use ξ to denote random vector

[*]School of Mathematical Sciences, Jiangsu Key Laboratory for NSLSCS, Nanjing Normal University, Nanjing, China, 210023 (mathhlsun@163.com). This author's work was partly supported by the National Natural Science Foundation of China (Grant No. 11401308).

[†]School of Industrial and Systems Engineering, Georgia Institute of Technology, USA (alex.shapiro@isye.gatech.edu). This author's work was partly supported by NSF grant 1633196.

[‡]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China (maxjchen@polyu.edu.hk). This author's work was partly supported by Hong Kong Research Grant Council PolyU15300219.

whose probability distribution is supported on the set Ξ, and also a point (an element) of the set Ξ, specific meaning will be clear from the context.

The SVI provide a unified form of the first order optimality conditions of stochastic optimization and model numerous equilibrium problems in economic, finance, management and engineering [25, 27, 29]. In the recent two decades, the SVI have been studied extensively and many new algorithms for solving the SVI have been developed [5, 7, 11]. Moreover, the two-stage SVI and multi-stage SVI have been introduced and investigated actively in the last few years [6, 8, 9, 22, 23, 29]. In the SVI, the probability distribution of ξ is supposed to be known (specified) exactly. However, unlike well-studied distributionally robust optimization (DRO), the theory and algorithms of distributionally robust variational inequalities (DRVI) are very limited. In practice the "true" distribution $\mathbb{P}$ of random variables is not known and could be estimated at best from historical data. The uncertainty of the "true" distribution in itself motivates the distributionally robust approach. We suggest the following formulation of the DRVI as a counterpart of (1.1):

$$0 \in \mathbb{E}_P[\Phi(x, \xi)] + \mathcal{N}_X(x), \tag{1.2}$$

$$P \in \arg\max_{Q \in \mathfrak{M}} \mathbb{E}_Q[\phi(x, \xi)], \tag{1.3}$$

where $\phi : X \times \Xi \to \mathbb{R}$ and $\mathfrak{M}$ is a specified set of probability measures (distributions) on $(\Xi, \mathcal{B})$. Note that by solving the above DRVI we mean to find a pair $\bar{x} \in X$ and $\bar{P} \in \mathfrak{M}$ satisfying (1.2)-(1.3). We give examples of such DRVI in section 2.

Our main contributions in this paper are threefold.

- Based on (1.2)-(1.3), we propose a comprehensive formulation of the DRVI to deal with the uncertain distribution in the SVI. We show that the first order optimality conditions of distributionally robust optimization and distributionally robust games are special cases of this formulation of the DRVI.

- We define the monotonicity of the DRVI and show that there is a pair of a decision vector $\bar{x} \in X$ and a distribution $\bar{P} \in \mathfrak{M}$ such that $(\bar{x}, \bar{P})$ solves the DRVI under certain conditions.

- We propose a SAA approach to the DRVI and investigate its convergence properties. Moreover, we use numerical examples of distributionally robust games to illustrate the formulation of the DRVI and the convergence of the SAA approach.

In section 2, we review three fundamental examples that are special cases of the DRVI. The first two examples are the first order optimality conditions of two types of DRO problems. The last example is an equivalent formulation of distributionally robust games (DRG) with convex objective functions of players and share constraints among players. In section 3, we define the monotonicity of the DRVI and prove the existence of solutions to the DRVI. In section 4, we propose a SAA approach for the DRVI with the corresponding convergence analysis. In section 5, we use numerical examples to illustrate the DRVI and the convergence of the SAA.

2 Formulation of the DRVI

In this section we give an extended (multivariate) definition of the DRVI and consider three relevant examples.

Definition 2.1 *(DRVI) Let $\mathfrak{M}_i$, $i = 1, ..., r$, be sets of probability measures on the sample space $(\Xi, \mathcal{B})$, $X \subseteq \mathbb{R}^n$ be a nonempty closed convex set, $\Phi : X \times \Xi \to \mathbb{R}^n$, $\phi_i : X \times \Xi \to \mathbb{R}$, $i = 1, \dots, r$, be continuous functions in $x \in X$, $\Phi(x, \cdot)$ and $\phi_i(x, \cdot)$ are measurable. The DRVI is to find a pair $(x, P) \in X \times \mathfrak{M}$ satisfying*

$$0 \in \mathbb{E}_P[\Phi(x, \xi)] + \mathcal{N}_X(x), \tag{2.1}$$

$$P_i \in \arg\max_{Q \in \mathfrak{M}_i} \mathbb{E}_Q[\phi_i(x, \xi)], \quad i = 1, \dots, r, \tag{2.2}$$

where $\mathfrak{M} := \{P_1 \times \ldots \times P_r : P_i \in \mathfrak{M}_i,\ i = 1, ..., r\}$,

$$\mathbb{E}_P[\Phi(x, \xi)] := \left(\mathbb{E}_{P_1}[\Phi_1(x, \xi)]^\top, \cdots, \mathbb{E}_{P_r}[\Phi_r(x, \xi)]^\top\right)^\top$$

with $\Phi(x, \xi) = (\Phi_1^\top(x, \xi), \cdots, \Phi_r^\top(x, \xi))^\top$, $\Phi_i(x, \xi) \in \mathbb{R}^{n_i}$ and $\sum_{i=1}^r n_i = n$.

Example 2.1 *Consider the following distributionally robust stochastic program*

$$\min_{x \in X} \sup_{P \in \mathfrak{M}} \mathbb{E}_P[\phi(x, \xi)], \tag{2.3}$$

where $\phi : X \times \Xi \to \mathbb{R}$. A point $(\bar{x}, \bar{P}) \in X \times \mathfrak{M}$ is a saddle point of the minimax problem (2.3) *if and only if*

$$\bar{x} \in \arg\min_{x \in X} \mathbb{E}_{\bar{P}}[\phi(x, \xi)] \text{ and } \bar{P} \in \arg\max_{P \in \mathfrak{M}} \mathbb{E}_P[\phi(\bar{x}, \xi)]. \tag{2.4}$$

Assuming that ϕ is differentiable in x and the differentiation and expectation operators can be interchanged, we can write the optimality conditions for the first problem in (2.4) *in the form* (2.1) *with $\Phi(x, \xi) := \nabla_x \phi(x, \xi)$. This leads to the DRVI of the form* (2.1) - (2.2) *with $r = 1$.*

Example 2.2 *Consider the following distributionally robust stochastic program*

$$\begin{aligned} \min_{x \in X} \quad & \sup_{P_0 \in \mathfrak{M}} \mathbb{E}_{P_0}[\phi_0(x, \xi)] \\ \text{s.t.} \quad & \sup_{P_1 \in \mathfrak{M}} \mathbb{E}_{P_1}[\phi_1(x, \xi)] \leq 0, \end{aligned} \tag{2.5}$$

where $\phi_i(x, \xi)$, $i = 0, 1$, are convex and twice continuously differentiable w.r.t. x.

The corresponding Largrange function is

$$L(x, \lambda) := \sup_{P_0 \in \mathfrak{M}} \mathbb{E}_{P_0}[\phi_0(x, \xi)] + \lambda \sup_{P_1 \in \mathfrak{M}} \mathbb{E}_{P_1}[\phi_1(x, \xi)],$$

where $\lambda \geq 0$. Suppose that the supremum in (2.5) *is finite valued for every $x \in X$ and the Slater constraint qualification holds [19], then DRO* (2.5) *is equivalent to*

$$\min_{x\in X}\max_{\lambda\geq 0}\sup_{P_0\in\mathfrak{M}}\mathbb{E}_{P_0}[\phi_0(x,\xi)] + \lambda\sup_{P_1\in\mathfrak{M}}\mathbb{E}_{P_1}[\phi_1(x,\xi)].$$

Since $\phi_i, i = 0, 1$ are convex, the above problem is equivalent to

$$\min_{x\in X}\max_{\lambda\geq 0, P_0\in\mathfrak{M}, P_1\in\mathfrak{M}}\mathbb{E}_{P_0}[\phi_0(x,\xi)] + \lambda\mathbb{E}_{P_1}[\phi_1(x,\xi)].$$

Then the corresponding DRVI is

$$\begin{aligned}
&0 \in \mathbb{E}_{P_0}[\nabla_x\phi_0(x,\xi^i)] + \lambda\mathbb{E}_{P_1}[\nabla_x\phi_1(x,\xi^i)] + \mathcal{N}_X(x),\\
&0 \in -\mathbb{E}_{P_1}[\phi_1(x,\xi^i)] + \mathcal{N}_{\mathbb{R}_+}(\lambda),\\
&P_i \in \arg\max_{Q\in\mathfrak{M}}\mathbb{E}_Q[\phi_i(x,\xi)], \quad i = 0, 1.
\end{aligned}$$

Example 2.3 *Consider the following distributionally robust formulation of Nash equilibrium with r players: find $(x_1^*, ..., x_r^*) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_r}$ such that*

$$x_i^* \in \arg\min_{x_i\in X_i}\max_{P_i\in\mathfrak{M}_i}\mathbb{E}_{P_i}[\phi_i(x_i, x_{-i}^*, \xi)],\ i = 1, ..., r. \tag{2.6}$$

Here $X_i \subset \mathbb{R}^{n_i}$ is a nonempty convex closed set, $\mathfrak{M}_i$ is a set of probability measures on $(\Xi_i, \mathcal{B}_i)$, $\Xi_i \subset \mathbb{R}^{\ell_i}$, and $\phi_i : \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_r} \times \Xi_i \to \mathbb{R}$, $i = 1, ..., r$. Similar to (2.4), *problem* (2.6) *leads to the following DRVI formulation (under appropriate differentiability assumptions)*

$$\begin{aligned}
&0 \in \mathbb{E}_{P_i}[\Phi_i(x_1, ..., x_r, \xi)] + \mathcal{N}_{X_i}(x_i),\ i = 1, ..., r,\\
&P_i \in \arg\max_{Q_i\in\mathfrak{M}_i}\mathbb{E}_{Q_i}[\phi_i(x_1, ..., x_r, \xi)],\ i = 1, ..., r,
\end{aligned}$$

with $\Phi_i(x_1, ..., x_r, \xi) := \nabla_{x_i}\phi_i(x_1, ..., x_r, \xi)$, $i = 1, ..., r$.

Remark 2.1 *If $X = \mathbb{R}^n_+$, then (1.2)-(1.3) reduces to the formulation of the distributionally robust complementarity problem (DRCP)*

$$0 \leq x\ \perp \mathbb{E}_P[\Phi(x,\xi)] \geq 0, \qquad P \in \arg\max_{Q\in\mathfrak{M}}\mathbb{E}_Q[\phi(x,\xi)]. \tag{2.7}$$

Other formulation of the DRCP from [8] can be written as follows

$$0 \leq x, \quad \max_{P\in\mathfrak{M}}\mathbb{E}_P[-\Phi_i(x,\xi)] \leq 0,\ i = 1, \ldots, n \tag{2.8}$$

$$\max_{P\in\mathfrak{M}}\mathbb{E}_P[x^\top\Phi(x,\xi)] = 0. \tag{2.9}$$

Obviously, if (x^*, P^*) is a solution of (2.8)-(2.9), then it is a solution of (2.7) with $\phi(x,\xi) = x^\top\Phi(x,\xi)$. Hence, (2.8)-(2.9) is a special case of (2.7).

3 Existence of solutions of the DRVI

In this section we investigate existence of solutions of the DRVI in three cases: discrete distributions, continuous distributions and monotone setting.

3.1 Finite dimensional setting

Suppose that the random vector ξ has a discrete distribution with a finite support $\Xi := \{\xi^1, ..., \xi^m\}$ of cardinality m. Then a probability distribution on Ξ can be identified with probability vector $q \in \Delta_m$, where

$$\Delta_m := \left\{q \in \mathbb{R}^m_+ : q_1 + ... + q_m = 1\right\}.$$

That is, each set $\mathfrak{M}_i$, $i = 1, ..., r$, can be viewed as a subset of Δ_m, and can be assumed to be convex and closed. Condition (2.2) can be written then as $0 \in -\phi^x + \mathcal{N}_{\mathfrak{M}}(p)$, where

$$\phi^x := (\phi_1(x, \xi^1), ..., \phi_1(x, \xi^m), ..., \phi_r(x, \xi^1), ..., \phi_r(x, \xi^m))^\top \in \mathbb{R}^{rm}$$

and $\mathcal{N}_{\mathfrak{M}}(p)$ is the normal cone to the set $\mathfrak{M} := \mathfrak{M}_1 \times ... \times \mathfrak{M}_r \subset \mathbb{R}^{rm}$ at $p := (p^1, \cdots, p^r)$. Thus in that case the corresponding DRVI can be written as the following finite dimensional VI:

$$0 \in \sum_{i=1}^m p^i \Phi(x, \xi^i) + \mathcal{N}_X(x), \tag{3.1}$$

$$0 \in -\phi^x + \mathcal{N}_{\mathfrak{M}}(p), \tag{3.2}$$

in variables $(x, p) \in X \times \Delta^r_m$ and $p^i\Phi(x, \xi^i) = (p^i_1\Phi_1(x, \xi^i)^\top, ..., p^i_r\Phi_r(x, \xi^i)^\top)^\top$.

In that setting existence of solution follows by the standard results, e.g. [15, Corollary 2.2.5].

Proposition 3.1 *Suppose $\Phi(x, \xi)$ and $\phi(x, \xi)$ are continuous in x and the set X is bounded (and hence the set $X \times \mathfrak{M} \subset \mathbb{R}^n \times \mathbb{R}^{rm}$ is convex compact). Then finite dimensional VI* (3.1)-(3.2) *has a nonempty and compact solution set.*

3.2 Continuous distributions setting

Let us consider now settings with continuous distributions of the random vector ξ. We assume existence of a reference probability measure $\mathbb{P}$ on $(\Xi, \mathcal{B})$ and that the ambiguity set consists of probability measures in some sense close to the reference measure $\mathbb{P}$. To proceed consider the space[1] $\mathcal{Z} := L_p(\Xi, \mathcal{B}, \mathbb{P})$, $p \in [1, \infty)$, and its dual space $\mathcal{Z}^* := L_q(\Xi, \mathcal{B}, \mathbb{P})$, $q \in (1, \infty]$,

[1] Banach spaces $\mathcal{Z}$ and $\mathcal{Z}^*$, equipped with the respective weak and weak* topologies, are paired topological vector spaces with respect to the bilinear form (scalar product) $\langle \zeta, Z\rangle = \int_\Xi \zeta Z d\mathbb{P}$, $Z \in \mathcal{Z}$, $\zeta \in \mathcal{Z}^*$. Note that the weak topology of $\mathcal{Z}$ and weak* topology of $\mathcal{Z}^*$, restricted to respective bounded sets, are metrizable and hence can be described in terms of convergent sequences. The weak convergence $Z_k \stackrel{w}{\to} \bar{Z}$ means that $\langle \zeta, Z_k\rangle$ converges to $\langle \zeta, \bar{Z}\rangle$ for any $\zeta \in \mathcal{Z}^*$. The weak* convergence $\zeta_k \stackrel{w^*}{\to} \bar{\zeta}$ means that $\langle \zeta_k, Z\rangle$ converges to $\langle \bar{\zeta}, Z\rangle$ for any $Z \in \mathcal{Z}$.

$1/p + 1/q = 1$. We assume $p = q = 2$ in this section (Section 3). We also use notation $\Phi_i^x(\cdot) := \Phi_i(x, \cdot)$ and $\phi_i^x(\cdot) := \phi_i(x, \cdot)$.

Assumption 3.1 *Suppose that, for $i = 1, ..., r$, the set $\mathfrak{M}_i$ in (2.2) consists of probability measures that are absolutely continuous with respect to $\mathbb{P}$ and consider the set $\mathfrak{A}_i := \{\zeta = dQ/d\mathbb{P} : Q \in \mathfrak{M}_i\}$ of the corresponding density functions. Suppose further that $\mathfrak{A}_i$ is a bounded, convex and weakly* closed subset of $\mathcal{Z}^*$, and that $\phi_i^x \in \mathcal{Z}$ for every $x \in X$ and $i = 1, ..., r$.*

Assumption 3.1 will hold for several setting of ambiguity sets, e.g., law invariant coherent risk measure [26], ϕ-divergence ball [20] and so on.

Since $\phi_i^x \in \mathcal{Z}$, it follows that for any $\zeta \in \mathfrak{A}_i$ and $dQ = \zeta d\mathbb{P}$ the integral

$$\mathbb{E}_Q[\phi_i^x] = \int_\Xi \phi_i^x \, \zeta d\mathbb{P}$$

is well defined and finitely valued. In what follows, we consider the ambiguity sets that satisfy Assumption 3.1 and hence (2.1)-(2.2) can be rewritten as

$$0 \in \int_\Xi \Phi_i^x \zeta_i d\mathbb{P} + \mathcal{N}_{X_i}(x_i), \quad i = 1, \ldots, r, \tag{3.3}$$

$$\zeta_i \in \arg\max_{\eta \in \mathfrak{A}_i} \int_\Xi \phi_i^x \eta d\mathbb{P}, \quad i = 1, \ldots, r. \tag{3.4}$$

Under Assumption 3.1, for $i = 1, \cdots, r$, the set $\mathfrak{A}_i$ is convex and closed in the weak* topology of $\mathcal{Z}_i^*$, and hence is weakly* compact. It follows that the set

$$\bar{\mathfrak{A}}_i^x := \arg\max_{\eta \in \mathfrak{A}_i} \int_\Xi \phi_i^x \, \eta d\mathbb{P}$$

is nonempty for any $x \in X$ (note that the set $\bar{\mathfrak{A}}_i^x$ represents the set of densities of the "arg max" probability measures in the right hand side of (2.2)) and $\mathfrak{A}_i$. Consider the mapping Φ and denote $\Phi^x(\cdot) := \Phi(x, \cdot)$. Suppose that for every $x \in X$, every component of Φ^x belongs to the space $\mathcal{Z} := \mathcal{Z}^1 \times \cdots \times \mathcal{Z}^r$. Consider the multifunction $\mathfrak{F} : X \rightrightarrows \mathbb{R}^n$ defined as

$$\mathfrak{F}(x) := \left\{ y = \int_\Xi \Phi^x \, \zeta d\mathbb{P} : \zeta \in \bar{\mathfrak{A}}^x \right\},$$

where $\zeta = (\zeta_1, ..., \zeta_r)$ with $\zeta_i \in \bar{\mathfrak{A}}_i^x$, $i = 1, ..., r$ and $\Phi^x \, \zeta := (\Phi_1(x, \cdot) \, \zeta_1, ..., \Phi_r(x, \cdot) \, \zeta_r)^\top$. In order to show the existence of solutions of the DRVI we need to verify that the following generalized equations have a solution

$$0 \in \mathfrak{F}(x) + \mathcal{N}_X(x). \tag{3.5}$$

Proposition 3.2 *Suppose that the set X is nonempty convex closed and bounded, and the mappings ϕ^x and Φ^x are weakly continuous with respect to $x \in X$. Then generalized equations (3.5) have a solution.*

Proof. It suffices to verify that the multifunction $\mathfrak{F}$ is closed, that is, for any sequences $x_k \in X$ converging to $\bar{x}$ and $y_k \in \mathfrak{F}(x_k)$ converging to $\bar{y}$, we have $\bar{y} \in \mathfrak{F}(\bar{x})$. Indeed, consider the multifunction $\mathfrak{G} : X \rightrightarrows X$ defined as

$$\mathfrak{G}(x) := \arg\min_{v \in X} \{\operatorname{dist}(v, \mathfrak{F}(x))\},$$

where $\operatorname{dist}(x, A)$ denotes the Euclidean distance from x to a set $A \subset \mathbb{R}^n$. Note that if the set A is convex, then $\operatorname{dist}(\cdot, A)$ is a convex function. We have that for every $x \in X$, the set $\bar{\mathfrak{A}}^x$ is convex and hence the set $\mathfrak{F}(x)$ is convex, and thus $\mathfrak{G}(x)$ is convex. Also if $\mathfrak{F}$ is closed, then $\mathfrak{G}$ is closed. It follows by Kakutani's fixed-point theorem that the multifunction $\mathfrak{G}$ has a fixed point $\bar{x} \in X$. Let $\bar{y}$ be the closest point of $\mathfrak{F}(\bar{x})$ to $\bar{x}$. Then $\bar{y} - \bar{x} \in \mathcal{N}_X(\bar{x})$.

In order to verify that $\mathfrak{F}$ is closed we can proceed as follows. By the weak* compactness of $\mathfrak{A}$ and the weak continuity of ϕ^x, we have that the multifunction $X \ni x \mapsto \bar{\mathfrak{A}}^x$ is weakly* closed[2] (e.g., [4, Propsition 4.4 and discussion on page 264]). By the weak continuity of Φ^x it follows that $\mathfrak{F}$ is closed. This follows from the fact that if $Z_k \xrightarrow{w} \bar{Z}$ and $\zeta_k \xrightarrow{w^*} \bar{\zeta}$, then $\langle \zeta_k, Z_k \rangle \to \langle \bar{\zeta}, \bar{Z} \rangle$ (e.g., [4, Theorem 2.23(iv)]). □

Remark 3.1 Recall that it is assumed that $\phi^x \in \mathcal{Z}$ for every $x \in X$. The mapping ϕ^x is weakly continuous if $\phi(x, \xi)$ is continuous in x and there is $\eta \in \mathcal{Z}$ such that $|\phi(x, \xi)| \le \eta(\xi)$ for all $x \in X$ and $\xi \in \Xi$. Indeed then for any $\zeta \in \mathcal{Z}^*$ we have that $|\phi^x \zeta| \le \eta |\zeta|$ and $\int \eta |\zeta| d\mathbb{P} < \infty$. Thus for a sequence $\{x_k\} \subset X$ converging to $\bar{x}$ it follows by the Lebesgue dominated convergence theorem that

$$\lim_{k \to \infty} \int_\Xi \phi^{x_k}(z)\, \zeta(z) d\mathbb{P}(z) = \int_\Xi \lim_{k \to \infty} \phi^{x_k}(z)\, \zeta(z) d\mathbb{P}(z) = \int_\Xi \phi^{\bar{x}}(z)\, \zeta(z) d\mathbb{P}(z).$$

This shows that ϕ^x is weakly continuous. Similar conditions can be applied to every component of the mapping Φ^x to guarantee its weak continuity.

Remark 3.2 For the set $\mathfrak{A}_i \subset \mathcal{Z}_i^*$ of density functions, we consider functional $\mathcal{R}_i : \mathcal{Z}_i \to \mathbb{R}$ defined as

$$\mathcal{R}_i(Z) := \sup_{\zeta_i \in \mathfrak{A}_i} \int_\Xi Z \zeta^i d\mathbb{P}. \tag{3.6}$$

Since $\mathfrak{A}_i$ is a bounded subset of $\mathcal{Z}_i^*$, the value $\mathcal{R}_i(Z)$ is finite for any $Z \in \mathcal{Z}_i$. This functional can be viewed as the dual representation of the corresponding so-called coherent risk measure. Various examples of coherent risk measures, their dual representations and closed forms for the corresponding sets $\bar{\mathfrak{A}}^x$ are given in (e.g., [25, Section 6.3.2]).

The optimality condition (2.2) can be written in the VI form as follows. For each player i, $i = 1, ..., r$, recall that $\mathcal{Z}_i$ and $\mathcal{Z}_i^*$ can be viewed as paired spaces with respect to the bilinear form $\langle \zeta_i, Z \rangle = \int_\Xi \zeta_i Z d\mathbb{P}$. Consider the indicator function $\mathbb{I}_{\mathfrak{A}_i}(\cdot)$ of the set $\mathfrak{A}_i \subset \mathcal{Z}_i^*$, that is $\mathbb{I}_{\mathfrak{A}_i}(\zeta_i) = 0$

[2] That is, if $x_k \in X$ converges to $\bar{x}$ and $\zeta_k \in \bar{\mathfrak{A}}_{x_k}$ is such that $\zeta_k \xrightarrow{w^*} \bar{\zeta}$, then $\bar{\zeta} \in \bar{\mathfrak{A}}_{\bar{x}}$.

for $\zeta_i \in \mathfrak{A}_i$ and $\mathbb{I}_{\mathfrak{A}_i}(\zeta_i) = +\infty$ for $\zeta_i \notin \mathfrak{A}_i$. At a point $\zeta_i \in \mathfrak{A}_i$ the subdifferential $\partial \mathbb{I}_{\mathfrak{A}_i}(\zeta)$ is equal to the normal cone

$$\mathcal{N}_{\mathfrak{A}_i}(\zeta_i) = \{Z \in \mathcal{Z}_i : \langle \eta - \zeta_i, Z\rangle \leq 0, \ \forall \eta \in \mathfrak{A}_i\}.$$

For $\zeta_i \notin \mathfrak{A}_i$ the normal cone $\mathcal{N}_{\mathfrak{A}_i}(\zeta_i) = \emptyset$. For $Z \in \mathcal{Z}_i$ we have that $\bar{\zeta}_i \in \arg\min_{\zeta_i \in \mathfrak{A}_i} \langle \zeta_i, -Z\rangle$ iff $\bar{\zeta}_i \in \arg\min_{\zeta_i \in \mathcal{Z}_i^*} \langle \zeta_i, -Z\rangle + \mathbb{I}_{\mathfrak{A}_i}(\zeta_i)$. Since the subdifferential of $\langle \zeta_i, -Z\rangle + \mathbb{I}_{\mathfrak{A}_i}(\zeta_i)$ at $\bar{\zeta}_i$ is equal to $-Z + \partial \mathbb{I}_{\mathfrak{A}_i}(\bar{\zeta}_i)$, it follows that $\bar{\zeta}_i \in \arg\min_{\zeta_i \in \mathfrak{A}_i} \langle \zeta_i, -Z\rangle$ iff $0 \in -Z + \mathcal{N}_{\mathfrak{A}_i}(\bar{\zeta}_i)$, that is, $0 \in -\phi(x, \cdot) + \mathcal{N}_{\mathfrak{A}_i}(\zeta_i)$. Therefore the optimality condition (2.2) can be written here as

$$0 \in -\phi^x + \mathcal{N}_{\mathfrak{A}}(\zeta), \tag{3.7}$$

where $\mathfrak{A} := \mathfrak{A}^1 \times \cdots \times \mathfrak{A}^r$ and $\zeta = (\zeta^1, ..., \zeta^r)$. Note that by pairing $\mathcal{Z}$ and $\mathcal{Z}^*$, the normal cone $\mathcal{N}_{\mathfrak{A}}(\zeta)$ is a subset of the space $\mathcal{Z}$.

This can be compared with the finite dimensional setting discussed in Section 3.1. Let $\mathbb{P}$ be the probability measure on the corresponding finite set $\Xi = \{\xi^1, ..., \xi^m\}$ assigning equal probability $1/m$ to each elementary event. Then any probability measure Q on Ξ is absolutely continuous with respect to $\mathbb{P}$ and its density $dQ/d\mathbb{P}$ is given by mq where $q \in \Delta_m$ is the respective probability vector.

3.3 Monotonicity property

By (3.7) in section 3.2, we can write DRVI (3.3)-(3.4) as follows:

$$0 \in \int_\Xi \Phi^x \zeta d\mathbb{P} + \mathcal{N}_X(x), \tag{3.8}$$

$$\mathbf{0} \in -\phi^x + \mathcal{N}_{\mathfrak{A}}(\zeta), \tag{3.9}$$

where $\mathbf{0} : \Xi \to \mathbb{R}^r$ is a constant function with value 0, $\phi^x = (\phi_1(x, \cdot), \cdots, \phi_r(x, \cdot))^\top$ is a vector-valued random function, $\mathcal{N}_{\mathfrak{A}}(\zeta) := \mathcal{N}_{\mathfrak{A}_1}(\zeta_1) \times \cdots \times \mathcal{N}_{\mathfrak{A}_r}(\zeta_r)$.

Note that $\mathbb{R}^n \times \mathcal{Z}$ and $\mathbb{R}^n \times \mathcal{Z}^*$ are paired by the bilinear form (scalar product), that is, for $x, z \in \mathbb{R}^n$, $u \in \mathcal{Z}$ and $\zeta \in \mathcal{Z}^*$,

$$\langle (x, u), (z, \zeta)\rangle := x^\top z + \sum_{i=1}^r \int_\Xi u_i \zeta_i d\mathbb{P}.$$

Consider mapping $\mathcal{G} : \mathbb{R}^n \times \mathcal{Z}^* \to \mathbb{R}^n \times \mathcal{Z}$ defined as

$$\mathcal{G}(x, \zeta) := \begin{pmatrix} \int_\Xi \Phi^x \zeta d\mathbb{P} \\ -\phi^x \end{pmatrix},$$

and denote the DRVI (2.1)-(2.2) by DRVI($\mathcal{G}, (X, \mathfrak{A})$). Monotonicity properties of this mapping are defined in the usual way. In particular, the mapping $\mathcal{G}$ is said to be monotone if for any $(x, \zeta), (z, \eta) \in \mathbb{R}^n \times \mathfrak{A}$, we have

$$\left\langle \mathcal{G}(x, \zeta) - \mathcal{G}(z, \eta), \begin{pmatrix} x - z \\ \zeta - \eta \end{pmatrix} \right\rangle \geq 0, \tag{3.10}$$

and $\mathcal{G}$ is said to be strongly monotone if there is $\alpha > 0$ such that

$$\left\langle \mathcal{G}(x,\zeta) - \mathcal{G}(z,\eta), \begin{pmatrix} x-z \\ \zeta-\eta \end{pmatrix} \right\rangle \geq \alpha(\|x-z\|^2 + \|\zeta-\eta\|_{L_2}^2), \tag{3.11}$$

where $\|\zeta-\eta\|_{L_2}$ is defined by function metric in L_2 space. Moreover, it is easy to observe that, $\mathcal{G}(x,\zeta)-\mathcal{G}(x,\eta) = \left((\int_\Xi \Phi^x \zeta d\mathbb{P})^\top - (\int_\Xi \Phi^x \eta d\mathbb{P})^\top, \mathbf{0}\right)^\top$, $\mathcal{G}$ cannot be strongly monotone. However, $\mathcal{G}$ can be monotone under some reasonable conditions.

To investigate the monotonicity of $\mathcal{G}$, we use $\nabla_x f(x,\cdot)$ and $J_x f(x,\cdot)$ to denote the partial derivative and partial Jacobian of f with respect to x. For $j = 1,\cdots,r$, $\tilde{\xi}_j : \Omega \to \Xi$ is a random vector with continuous distribution Q_j such that $dQ_j = \zeta_j d\mathbb{P}$, let $S_j(\xi)^\top, S_j(\tilde{\xi}_j)^\top \in \mathbb{R}^\ell$, $\tilde{\xi} := (\tilde{\xi}_1,\cdots,\tilde{\xi}_r) \in \Xi_r \subseteq \mathbb{R}^{r\ell}$,

$$S(\tilde{\xi}) := (S_1(\tilde{\xi}_1)^\top,\cdots,S_r(\tilde{\xi}_r)^\top)^\top \quad \text{and} \quad S(\xi) := (S_1(\xi)^\top,\cdots,S_r(\xi)^\top)^\top.$$

Lemma 3.1 *Suppose for any $\tilde{\xi} \in \Xi_r$, $S(\tilde{\xi})$ is a positive semidefinite matrix. Then $\int_\Xi S(\xi)\zeta d\mathbb{P}$ is positive semidefinite.*

Proof. We consider a discrete approximation of $\int_\Xi S(\xi)\zeta d\mathbb{P}$ firstly. Let $\Xi^N := \{\xi^1,\cdots,\xi^N\}$ be a discrete approximation of Ξ with the weight vectors $\{p_1^1,\cdots,p_1^N\},\cdots,\{p_r^1,\cdots,p_r^N\}$ such that for $i = 1,\cdots,N$, $j = 1,\cdots,r$, $p_j^i \geq 0$, $\sum_{i=1}^N p_j^i = 1$ and w.p.1

$$\lim_{N\to\infty} \sum_{i=1}^N p_j^i S_j(\xi^i) = \int_\Xi S_j(\xi)\zeta_j d\mathbb{P}. \tag{3.12}$$

There are several ways to construct an approximation above. One way is construct i.i.d. samples $\Xi_j^N := \{\xi_j^1,\cdots,\xi_j^{N_j}\}$ of continuous distribution Q_j such that $dQ_j = \zeta_j d\mathbb{P}$ for $j = 1,\cdots,r$. Then $\Xi^N = \cup_{j=1}^r \Xi_j^N$, $N = |\Xi^N| \leq \sum_{j=1}^r N^j$ and $p_j^i = \frac{1}{N_j}$ if $\xi^i \in \Xi_j^N$ and $p_j^i = 0$ otherwise.

Let $P^i := \text{diag}(p_1^i,\cdots,p_r^i)$ for $i = 1,\cdots,N$, then

$$\sum_{i=1}^N P^i S(\xi^i) := \left(\textstyle\sum_{i=1}^N p_1^i S_1(\xi^i)^\top,\cdots,\sum_{i=1}^N p_r^i S_r(\xi^i)^\top\right)^\top$$

is an approximation of $\int_\Xi S(\xi)\zeta d\mathbb{P}$. We then prove $\sum_{i=1}^N P^i S(\xi^i)$ is positive semidefinite.

To this end, we do the following procedure.

Step 0. Let $k = 1$. We reorder the weight vectors $\{p_1^1,\cdots,p_1^N\},\cdots,\{p_r^1,\cdots,p_r^N\}$ to $\{p_1^{(1)},\cdots,p_1^{(N)}\}$, $\cdots$, $\{p_r^{(1)},\cdots,p_r^{(N)}\}$ such that $p_j^{(1)} \geq p_j^{(2)} \cdots \geq p_j^{(N)}$ for $j = 1,\cdots,r$. Let $\tilde{\xi}^k = (\tilde{\xi}_1^{(1)},\cdots,\tilde{\xi}_r^{(1)})$ and $\tilde{p}_k = \min\{p_j^{(1)}, j = 1,\cdots,r\}$. We construct $\tilde{p}_k J_x \Phi(x,\tilde{\xi}^k)$ and reduce $\tilde{p}_k$ from $p_j^{(1)}$, that is new $p_j^{(1)} := p_j^{(1)} - \tilde{p}_k$, for $j = 1,\cdots,r$. Note that by the condition of the lemma, $\tilde{p}_k S(\tilde{\xi}^k)$ is positive semidefinite. Let $k = k+1$.

Step 1. For $j = 1, \cdots, r$, since we have reduced $\tilde{p}_{k-1}$ from $p_j^{(1)}$ and $p_j^{(1)}$ may not be the largest one of $\{p_j^{(1)}, \cdots, p_j^{(N)}\}$, we reorder the the weight vectors again. To easy notation, we still denote the newly reordered weight vectors as $\{p_1^{(1)}, \cdots, p_1^{(N)}\}, \cdots, \{p_r^{(1)}, \cdots, p_r^{(N)}\}$. Note that now $\sum_{i=1}^N p_j^{(i)} = 1 - \sum_{i=1}^{k-1} \tilde{p}_i$. If $\sum_{i=1}^N p_j^{(i)} = 0$, for $j = 1, \cdots, r$, stop. Note that $\sum_{i=1}^N p_j^{(i)} = \sum_{i=1}^N p_l^{(i)}$ for all $j, l \in \{1, \cdots, r\}$. Otherwise, go to Step 2.

Step 2. Let $\tilde{\xi}^k = (\tilde{\xi}_1^{(1)}, \cdots, \tilde{\xi}_r^{(1)})$ and $\tilde{p}_k = \min\{p_j^{(1)}, j = 1, \cdots, r\}$. We construct $\tilde{p}_k S(\tilde{\xi}^k)$ and reduce $\tilde{p}_k$ from $p_j^{(1)}$, that is new $p_j^{(1)} := p_j^{(1)} - \tilde{p}_k$, for $j = 1, \cdots, r$. Note that by the condition of the lemma, $\tilde{p}_k S(\tilde{\xi}^k)$ is positive semidefinite. Let $k = k + 1$. Go to Step 1.

Note that $\tilde{\xi}^k \in \Xi_r^N$ and $|\Xi_r^N| = N^r$, the procedure above will stop at finite iterations, denote by K, $K \leq N^r$. Since $S(\tilde{\xi}^k)$ is positive semidefinite for $k = 1, \cdots, K$ and $\sum_{k=1}^K \tilde{p}_k = 1$, then

$$\sum_{i=1}^N P^i S(\xi^i) = \sum_{k=1}^K \tilde{p}_k S(\tilde{\xi}^k)$$

is positive semidefinite, and by (3.12), $\int_\Xi S(\xi)\zeta d\mathbb{P}$ is positive semidefinite w.p.1. $\square$

Let $\tilde{\xi} := (\tilde{\xi}_1, \cdots, \tilde{\xi}_r) \in \Xi_r \subset \mathbb{R}^{r\ell}$, and

$$\int_\Xi J_x \Phi(x, \xi)\zeta d\mathbb{P} := \begin{pmatrix} \int_\Xi J_x \Phi_1^x \zeta_1 d\mathbb{P} \\ \vdots \\ \int_\Xi J_x \Phi_r^x \zeta_r d\mathbb{P} \end{pmatrix}.$$

Proposition 3.3 *Consider DRVI* (3.8)-(3.9). *Suppose (a) for $i = 1, \cdots, r$ and $\xi \in \Xi$, $\Phi_i(\cdot, \xi)$ and $\phi_i(\cdot, \xi)$ are continuously differentiable, (b) for any $\tilde{\xi} \in \Xi^r$ and $x \in X$,*

$$\left((J_x \Phi_1(x, \tilde{\xi}_1))^\top, \cdots, (J_x \Phi_r(x, \tilde{\xi}_r))^\top\right)^\top$$

is a positive semidefinite matrix, (c) for $\mathbb{P}$-a.e. $\xi \in \Xi$, $\zeta^i(\xi)(\tilde{x}_i^\top \Phi_i(x, \xi) - J_x \phi_i(x, \xi)\tilde{x}) \geq 0$, for all $x \in X, \tilde{x} \in \mathbb{R}^n$ and $\tilde{\zeta} \in \mathcal{Z}$, $i = 1, \cdots, r$, then $\mathcal{G}$ is monotone over $X \times \mathfrak{A}$.

Proof. It is easy to observe that

$$J\mathcal{G}_{(x,\zeta)} = \begin{pmatrix} \int_\Xi J_x \Phi_1^x \zeta_1 d\mathbb{P} & \Phi_1(x, \cdot) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \int_\Xi J_x \Phi_r^x \zeta_r d\mathbb{P} & \mathbf{0} & \cdots & \Phi_r(x, \cdot) \\ -J_x \phi_1(x, \cdot) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -J_x \phi_r(x, \cdot) & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}, \quad J\mathcal{G}_{(x,\zeta)} \begin{pmatrix} \tilde{x} \\ \tilde{\zeta} \end{pmatrix} = \begin{pmatrix} \int_\Xi J_x \Phi_1^x \zeta_1 d\mathbb{P}\tilde{x} + \int_\Xi \Phi_1^x \tilde{\zeta}_1 d\mathbb{P} \\ \vdots \\ \int_\Xi J_x \Phi_r^x \zeta_r d\mathbb{P}\tilde{x} + \int_\Xi \Phi_r^x \tilde{\zeta}_r d\mathbb{P} \\ -J_x \phi_1(x, \cdot)\tilde{x} \\ \vdots \\ -J_x \phi_r(x, \cdot)\tilde{x} \end{pmatrix}$$

and

$$\left\langle (\tilde{x}, \tilde{\zeta}), J\mathcal{G}_{(x,\zeta)} \begin{pmatrix} \tilde{x} \\ \tilde{\zeta} \end{pmatrix} \right\rangle = \sum_{i=1}^r \left(\tilde{x}_i^\top \int_\Xi J_x \Phi_i^x \zeta_i d\mathbb{P}\tilde{x} + \tilde{x}_i^\top \int_\Xi \Phi_i^x \tilde{\zeta}_i d\mathbb{P} - \int_\Xi J_x \phi_i^x \tilde{\zeta}_i d\mathbb{P}\tilde{x}] \right).$$

By condition (b) and Lemma 3.1, $\int_\Xi J_x\Phi(x,\xi)\zeta d\mathbb{P}$ is positive semidefinite. By condition (c), for any $\tilde{x} \in X$ and $\tilde{\zeta} \in \mathcal{Z}$,

$$\sum_{i=1}^r \left(\tilde{x}_i^\top \int_\Xi \Phi_i^x \tilde{\zeta}_i d\mathbb{P} - \int_\Xi J_x\phi_i^x \tilde{\zeta}_i d\mathbb{P}\tilde{x} \right) \geq 0. \tag{3.13}$$

Then $\left\langle (\tilde{x},\tilde{\zeta}), J\mathcal{G}_{x,\zeta}\begin{pmatrix}\tilde{x}\\ \tilde{\zeta}\end{pmatrix}\right\rangle \geq 0$ holds for any $\tilde{x} \in \mathbb{R}^n$ and $\tilde{\zeta} \in \mathcal{Z}$, and then by [12, Theorem 3.1], $\mathcal{G}(x,\zeta)$ is monotone over $X \times \mathfrak{A}$. □

The above proposition shows the monotone properties of the DRVI in the continuous distributions setting (section 3.2). Note that in the case of finite dimensional setting (section 3.1), we can simply rewrite Proposition 3.3 as follows.

Corollary 3.1 *Consider DRVI* (3.1)-(3.2). *Suppose (a) for $i = 1,\cdots,r$ and $\xi \in \Xi^m$, $\Phi_i(\cdot,\xi)$ and $\phi_i(\cdot,\xi)$ are continuously differentiable, (b) for any $\tilde{\xi} \in (\Xi^m)^r$ and $x \in X$,*

$$\left((J_x\Phi_1(x,\tilde{\xi}_1))^\top, \cdots, (J_x\Phi_r(x,\tilde{\xi}_r))^\top \right)^\top$$

is a positive semidefinite matrix, (c) for all $\xi \in \Xi^m$, $x \in X$, $\tilde{x} \in \mathbb{R}^n$, $i = 1,...,r$, $\tilde{x}_i^\top \Phi_i(x,\xi) - J_x\phi_i(x,\xi)\tilde{x} = 0$. Then $\mathcal{G}$ corresponding to DRVI (3.1)-(3.2) *is monotone over $X \times \mathfrak{M}$.*

In what follows, we give the existence of solutions of the DRVI based on the monotone properties.

Definition 3.1 ([16, Definition 12.1]) *The mapping $\mathcal{G} : \mathbb{R}^n \times \mathfrak{A} \to \mathbb{R}^n \times \mathcal{Z}$ is hemicontinuous on $\mathbb{R}^n \times \mathcal{Z}^*$ if $\mathcal{G}$ is continuous on line segments in $\mathbb{R}^n \times \mathcal{Z}^*$, i.e., for every pair of points $(x,\zeta),(z,\eta) \in \mathbb{R}^n \times \mathcal{Z}^*$, the following function is continuous*

$$t \mapsto \left\langle \mathcal{G}(tx + (1-t)z, t\zeta + (1-t)\eta), \begin{pmatrix} x - z\\ \zeta - \eta\end{pmatrix}\right\rangle, \ \ 0 \leq t \leq 1.$$

Definition 3.2 ([16, Definition 12.3 (i)]) *The mapping $\mathcal{G} : X \times \mathcal{Z}^* \to X \times \mathcal{Z}$ is weakly coercive if there exists $(x_0,\zeta^0) \in \mathbb{R}^n \times \mathcal{Z}^*$ such that*

$$\left\langle \mathcal{G}(x,\zeta), \begin{pmatrix} x - x_0\\ \zeta - \zeta^0\end{pmatrix}\right\rangle \to \infty \ \text{ as } \ \|x - x_0\| + \|\zeta - \zeta^0\| \to \infty \ \text{ and } (x,\zeta) \in X \times \mathcal{Z}^*.$$

Theorem 3.1 *Suppose the conditions of Proposition 3.3 hold. If $X \subseteq \mathbb{R}^n$ is a closed and convex set, $\mathfrak{A}$ is convex and weakly* compact in $\mathcal{Z}$ and $\mathcal{G}$ is weakly coercive, then DRVI($\mathcal{G},(X,\mathfrak{A})$) has a solution.*

By Proposition 3.3, it is obvious that $\mathcal{G}$ is hemicontinuous and monotone on $\mathbb{R}^n \times \mathscr{P}$. Then Theorem 3.1 is from [16, Theorem 12.1 and Corollary 12.2] directly.

Moreover, we can also have a finite dimension version of Theorem 3.1 as follows.

Corollary 3.2 *Suppose the conditions of Corollary 3.1 hold. If $X \subseteq \mathbb{R}^n$ is a closed and convex set, $\mathfrak{M}$ is compact in $\mathbb{R}^{rm}$ and $\mathcal{G}$ is coercive, then DRVI($\mathcal{G},(X,\mathfrak{M})$)* (3.1)-(3.2) *has a solution.*

3.4 Examples of monotone DRVI

We illustrate the monotone property and coerciveness of $\mathcal{G}$ in the DRVI by two examples from the DRO and distributionally robust generalized Nash equilibrium.

Example 3.1 *Consider the DRO* (2.3)*, where ϕ is convex and twice continuously differentiable, $\Xi := \{\xi^1, \xi^2\}$, $\mathfrak{M} \subset \{(p^1, p^2) : p^i \geq 0, p^1 + p^2 = 1, i = 1, 2\}$ is convex and compact.*

Then the corresponding DRVI is

$$0 \in p^1 \nabla_x \phi(x, \xi^1) + p^2 \nabla_x \phi(x, \xi^2) + \mathcal{N}_X(x), \tag{3.14}$$

$$\mathbf{0} \in \begin{pmatrix} -\phi(x, \xi^1) \\ -\phi(x, \xi^2) \end{pmatrix} + \mathcal{N}_{\mathfrak{M}}((p^1, p^2)). \tag{3.15}$$

And the corresponding function is

$$\mathcal{G}(x, P) = \begin{pmatrix} p^1 \nabla_x \phi(x, \xi^1) + p^2 \nabla_x \phi(x, \xi^2) \\ -\phi(x, \xi^1) \\ -\phi(x, \xi^2) \end{pmatrix}.$$

Moreover,

$$J\mathcal{G}_{(x,P)} = \begin{pmatrix} p^1 \nabla_{xx} \phi(x, \xi^1) + p^2 \nabla_{xx} \phi(x, \xi^2) & \nabla_x \phi(x, \xi^1) & \nabla_x \phi(x, \xi^2) \\ -\nabla_x \phi(x, \xi^1)^\top & 0 & 0 \\ -\nabla_x \phi(x, \xi^2)^\top & 0 & 0 \end{pmatrix}$$

is positive semidefinite over $X \times \mathfrak{M}$ and then $\mathcal{G}$ is monotone.

Then we prove the coercive of $\mathcal{G}$. Let $X = \mathbb{R}^2_+$ and $x_0 = (0, 0)$, $P^0 = (1, 0)$. Suppose for any $\xi \in \Xi$, $\phi(x, \xi)$ is a strongly convex function of x with parameter $m(\xi^i) > 0$, $i = 1, 2$, we have when x sufficiently large, $\phi(x, \xi^i) \geq 0$, $i = 1, 2$ and

$$\phi(0, \xi^i) \geq \phi(x, \xi^i) - \nabla_x \phi(x, \xi^i)^\top x + \frac{m(\xi^i)}{2} \|x\|_2^2.$$

Then

$$\begin{aligned}
&\liminf_{x \geq 0, \|x\| \to \infty} \frac{\left\langle \mathcal{G}(x, P), \begin{pmatrix} x - x_0 \\ P - P^0 \end{pmatrix} \right\rangle}{\|x\|} \\
= &\liminf_{x \geq 0, \|x\| \to \infty} \frac{x^\top (p^1 \nabla_x \phi(x, \xi^1) + p^2 \nabla_x \phi(x, \xi^2)) + (P - P^0)^\top \begin{pmatrix} -\phi(x, \xi^1) \\ -\phi(x, \xi^2) \end{pmatrix}}{\|x\|} \\
= &\liminf_{x \geq 0, \|x\| \to \infty} \frac{p_1 (x^\top \nabla_x \phi(x, \xi^1) - \phi(x, \xi^1)) + p_2 (x^\top \nabla_x \phi(x, \xi^2) - \phi(x, \xi^2)) + \phi(x, \xi^1)}{\|x\|} \\
\geq &\liminf_{x \geq 0, \|x\| \to \infty} \frac{p_1 (x^\top \nabla_x \phi(x, \xi^1) - \phi(x, \xi^1)) + p_2 (x^\top \nabla_x \phi(x, \xi^2) - \phi(x, \xi^2))}{\|x\|} \\
\geq &\liminf_{x \geq 0, \|x\| \to \infty} \frac{\sum_{i=1}^2 p^i (\frac{m(\xi^i)}{2} \|x\|^2 - \phi(0, \xi^i))}{\|x\|} > 0.
\end{aligned}$$

Combining the monotonicity and coerciveness of $\mathcal{G}$, by Corollary 3.2, the DRVI has a solution.

We can also consider a distributionally robust generalized Nash equilibrium problem as follows.

Example 3.2 *Consider the distributionally robust generalized Nash equilibrium problem as follows:*

$$\min_{x_i \in X_i} \max_{P_i \in \mathfrak{M}_i} \mathbb{E}_{P_i}[f_i(x,\xi)] + g_i(x), \quad \text{s.t. } b_1 x_1 + b_2 x_2 \le c, \quad i = 1,2, \tag{3.16}$$

where $x = (x_1, x_2)$, $x_i \in \mathbb{R}^{n_i}$, $X_i \subseteq \mathbb{R}^{n_i}$ *is a convex set, for* P_i*-a.e.* ω, $\forall P_i \in \mathfrak{M}_i$ $f_i(\cdot, \xi)$ *is convex and twice continuously differentiable with respect to* x*, and* g_i *is convex and twice continuously differentiable,* $\Xi := \{\xi^1, \xi^2\}$, $\mathfrak{M}_i \subset \{(p_i^1, p_i^2) : p_i^j \ge 0, p_i^1 + p_i^2 = 1, j = 1,2\}$ *is convex and compact,* $i = 1,2$*. Suppose* $J\left(\nabla_{x_1} g_1(x), \nabla_{x_2} g_2(x)\right)^\top$ *is positive semi-definite.*

Then the corresponding DRVI is

$$0 \in p_i^1 \nabla_{x_i} f_i(x,\xi^1) + p_i^2 \nabla_{x_i} f_i(x,\xi^2) + \nabla_{x_i} g_i(x) + b_i \mu + \mathcal{N}_{X_i}(x_i), \qquad i = 1,2 \tag{3.17}$$

$$0 \in c - b_1 x_1 - b_2 x_2 + \mathcal{N}_{\mathbb{R}_+}(\mu) \tag{3.18}$$

$$\mathbf{0} \in \begin{pmatrix} -f_i(x,\xi^1) \\ -f_i(x,\xi^2) \end{pmatrix} + \mathcal{N}_{\mathfrak{M}_i}((p_i^1, p_i^2)), \quad i = 1,2. \tag{3.19}$$

Let

$$\Phi(x,\xi) = \begin{pmatrix} \nabla_{x_1} f_1(x_1,\xi) + \nabla_{x_1} g_1(x) + b_1 \mu \\ \nabla_{x_2} f_2(x_2,\xi) + \nabla_{x_2} g_2(x) + b_2 \mu \end{pmatrix} \quad \text{and} \quad \phi(x,\xi) = \begin{pmatrix} -f_1(x,\xi) \\ -f_2(x,\xi) \end{pmatrix}.$$

Then the DRVI (3.17)-(3.19) *is corresponding to* (3.1)-(3.2) *with* (3.18)*. Moreover,*

$$\mathcal{G}(x,\mu,P) = \begin{pmatrix} p_1^1 \nabla_{x_1} f_1(x,\xi^1) + p_1^2 \nabla_{x_1} f_1(x,\xi^2) + \nabla_{x_1} g_1(x) + b_1 \mu \\ p_2^1 \nabla_{x_2} f_2(x,\xi^1) + p_2^2 \nabla_{x_2} f_2(x,\xi^2) + \nabla_{x_2} g_2(x) + b_2 \mu \\ c - b_1 x_1 - b_2 x_2 \\ -f_1(x,\xi^1) \\ -f_1(x,\xi^2) \\ -f_2(x,\xi^1) \\ -f_2(x,\xi^2) \end{pmatrix}.$$

For $i,j = 1,2$*, let*

$$a_{ij} = p_i^1 \nabla_{x_i x_j} f_i(x_i,\xi^1) + p_i^2 \nabla_{x_i x_j} f_i(x_i,\xi^2) + \nabla_{x_i x_j} g_i(x),$$

then $J\mathcal{G}_{(x,\mu,P)}$ *is*

$$\begin{pmatrix} a_{11} & a_{12} & b_1 & \nabla_{x_1} f_1(x,\xi^1) & \nabla_{x_1} f_1(x,\xi^2) & 0 & 0 \\ a_{21} & a_{22} & b_2 & 0 & 0 & \nabla_{x_2} f_2(x,\xi^1) & \nabla_{x_2} f_2(x,\xi^2) \\ -b_1 & -b_2 & 0 & 0 & 0 & 0 & 0 \\ -\nabla_{x_1} f_1(x,\xi^1)^\top & -\nabla_{x_2} f_1(x,\xi^1)^\top & 0 & 0 & 0 & 0 & 0 \\ -\nabla_{x_1} f_1(x,\xi^2)^\top & -\nabla_{x_2} f_1(x,\xi^2)^\top & 0 & 0 & 0 & 0 & 0 \\ -\nabla_{x_1} f_2(x,\xi^1)^\top & -\nabla_{x_2} f_2(x,\xi^1)^\top & 0 & 0 & 0 & 0 & 0 \\ -\nabla_{x_1} f_2(x,\xi^2)^\top & -\nabla_{x_2} f_2(x,\xi^2)^\top & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

It is obvious that in general, $\mathcal{G}$ is nonmonotone. Moreover, $\mathcal{G}$ can be monotone if f_1 is only w.r.t. (x_1, ξ) and f_2 is only w.r.t. (x_2, ξ), that is $f_1(x_1, \xi)$ and $f_2(x_2, \xi)$, then

$$\begin{pmatrix} \nabla_{x_1x_1} f_1(x_1, \xi^i) + \nabla_{x_1} g(x) & \nabla_{x_1x_2} g(x) \\ \nabla_{x_2x_1} g(x) & \nabla_{x_2x_2} f_2(x_2, \xi^j) + \nabla_{x_2x_2} g(x) \end{pmatrix}$$

is positive semidefinite for $i, j = 1, 2$.

Then we show the coerciveness of $\mathcal{G}$. Similar as in Example 3.1, let $X_j = \mathbb{R}_+$ and $x_{j0} = 0$, $\mu_0 = 0$, $P_j^0 = (1, 0)$ for $j = 1, 2$. Suppose $f_j(x, \xi^i)$ and $g(x)$ are strongly convex with parameter $m_j(\xi^i) > 0$ and $m_j^g > 0$ respectively, for $i = 1, 2$ and $j = 1, 2$, $c > 0$, we have

$$\begin{aligned}
&\liminf_{(x,\mu)\geq 0, \|(x,\mu)\|\to\infty} \frac{\left\langle \mathcal{G}(x, \mu, P), \left(x - x_0, \mu - \mu_0, P - P^0\right)^\top \right\rangle}{\|(x, \mu, P)\|} \\
\geq &\liminf_{(x,\mu)\geq 0, \|(x,\mu)\|\to\infty} \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} \left(p_j^i(x_j \nabla_{x_j} f_j(x_j, \xi^i) + x_j \nabla_{x_j} g(x) - f_j(x_j, \xi^i)) + p_{j0}^i f_j(x_j, \xi^i)\right) + \mu c}{\|(x, \mu)\|} \\
\geq &\liminf_{(x,\mu)\geq 0, \|(x,\mu)\|\to\infty} \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} p_j^i(x_j \nabla_{x_j} f_j(x_j, \xi^i) - f_j(x_j, \xi^i) + x_j \nabla_{x_j} g_j(x)) + \mu c}{\|(x, \mu)\|} \\
\geq &\liminf_{(x,\mu)\geq 0, \|(x,\mu)\|\to\infty} \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} p_j^i(\frac{m_j(\xi^i)}{2}\|x_j\|^2 - f_j(0, \xi^i) + \frac{m_j^g}{2}\|x_j\|^2 - g_j(0)) + \mu c}{\|(x, \mu)\|} > 0.
\end{aligned}$$

Combining the monotonicity and coerciveness of $\mathcal{G}$, by Corollary 3.2, the DRVI has a solution.

4 Discretization of Probability Distributions

In this section, we consider the discretization of DRVI with the ambiguity sets formed from continuous distributions in the setting specified in Assumption 3.1. There are several ways to discretize the ambiguity set [10, 26, 30]. We propose a SAA approach to the DRVI. For the sake of simplicity we assume here that $r = 1$ and drop the subscript i in Φ_i^x and ϕ_i^x, etc. An extension for $r > 1$ will be straightforward. Recall that $\mathbb{P}$ is the reference probability measure (distribution) on $(\Xi, \mathcal{B})$, $\mathcal{Z} = L_p(\Xi, \mathcal{B}, \mathbb{P})$, $\mathcal{Z}^* = L_q(\Xi, \mathcal{B}, \mathbb{P})$, $\mathfrak{A}$ is a convex bounded weakly* closed subset of $\mathcal{Z}^*$ of densities associated with the ambiguity set $\mathfrak{M}$, and

$$\mathcal{R}(Z) = \sup_{\zeta \in \mathfrak{A}} \int_\Xi Z(s)\zeta(s) d\mathbb{P}(s),\ Z \in \mathcal{Z}. \tag{4.1}$$

Let us introduce some definitions.

It is said that random variables $Y, Y' : \Xi \to \mathbb{R}$ are distributionally equivalent (with respect to $\mathbb{P}$), denoted $Y \overset{\mathcal{D}}{\sim} Y'$, if $\mathbb{P}(Y \leq y) = \mathbb{P}(Y' \leq y)$ for all $y \in \mathbb{R}$. It is said that a functional $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ is *law invariant* if $\mathcal{R}(Z) = \mathcal{R}(Z')$ for any distributionally equivalent $Z, Z' \in \mathcal{Z}$. The set $\mathfrak{A} \subset \mathcal{Z}^*$ is said to be law invariant if $\zeta \in \mathfrak{A}$ and $\zeta' \overset{\mathcal{D}}{\sim} \zeta$, then $\zeta' \in \mathfrak{A}$. It is known that the functional $\mathcal{R}$ is law invariant iff the corresponding set $\mathfrak{A}$ is law invariant (cf., [26, Theorem 2.3]).

Assumption 4.1 *The set $\mathfrak{A}$ is law invariant.*

By Assumption 4.1 we have that the functional $\mathcal{R}(Z)$ is law invariant, and hence can be viewed as a function of the respective cumulative distribution function (CDF) of Z. It is possible to proceed with the required discretization by making discretization of the corresponding CDF of $\phi^x(\xi)$. However, such approach is indirect and inconvenient for applications. Therefore we discuss below several important cases where this can be performed in a rather straightforward way.

Consider an iid sample $\xi^j \in \Xi$, $j = 1, ..., N$, from the reference distribution $\mathbb{P}$. With the law invariant risk measure $\mathcal{R}$ is associated the corresponding empirical functional[3] $\hat{\mathcal{R}}_N : \mathbb{R}^N \to \mathbb{R}$. The functional $\mathcal{R}_N$ has the dual representation

$$\mathcal{R}_N(Z) := \sup_{\zeta \in \mathfrak{A}^N} N^{-1} \sum_{j=1}^{N} \zeta_j Z(\xi^j), \tag{4.2}$$

where $\mathfrak{A}^N$ is the respective convex closed set of densities[4] $\zeta = (\zeta_1, ..., \zeta_N)$. In the next section we give examples of how the empirical functional can be constructed.

For the generated sample,

$$\hat{\mathcal{R}}_N(\phi^x) = \sup_{\zeta \in \mathfrak{A}^N} N^{-1} \sum_{j=1}^{N} \zeta_j \phi^x(\xi^j) \tag{4.3}$$

can be considered as an empirical estimate of $\mathcal{R}(\phi^x)$. We have that under mild regularity conditions, $\hat{\mathcal{R}}_N(\phi^x)$ epiconverges w.p.1 to $\mathcal{R}(\phi^x)$ on X (cf. [24]). This suggests the following discretization of problem (3.3) - (3.4):

$$0 \in \sum_{j=1}^{N} \zeta_j \Phi(x, \xi^j) + \mathcal{N}_X(x), \tag{4.4}$$

$$\zeta \in \arg\max_{\eta \in \mathfrak{A}^N} \sum_{j=1}^{N} \eta_j \phi(x, \xi^j). \tag{4.5}$$

4.1 Construction of the empirical estimates

Here we discuss construction of the empirical estimates of the risk measure $\mathcal{R}$ defined in (4.1). For a random variable Z we denote by $H_Z(z) := \mathbb{P}(Z \le z)$ its cumulative distribution function (CDF) and by $H_Z^{-1}(t) := \inf\{\tau : H_Z(\tau) \ge t\}$ the corresponding quantile function (also called Value-at-Risk). Note that the subdifferential of $\mathcal{R}(Z)$ is given by

$$\partial\mathcal{R}(Z) = \arg\max_{\zeta \in \mathfrak{A}} \int_{\Xi} Z(s)\zeta(s)d\mathbb{P}(s) \tag{4.6}$$

(e.g., [25, eq. (6.49), page 284]). Let us consider the following example.

[3] Any $Z : \{\xi^1, ..., \xi^N\} \to \mathbb{R}$ can be identified with N-dimentional vector $(Z(\xi_1), ..., Z(\xi_N))$, and hence the empirical risk measure can be viewed as defined on $\mathbb{R}^N$.

[4] Note that ζ is a density on $\{\xi^1, ..., \xi^N\}$ if $\zeta \ge 0$ and $N^{-1}\sum_{i=1}^N \zeta_i = 1$, i.e., $N^{-1}\zeta \in \Delta_N$.

Example 4.1 Consider the Average Value-at-Risk,

$$\mathsf{AV@R}_{1-\alpha}(Z) := \frac{1}{1-\alpha}\int_{\alpha}^{1} H_Z^{-1}(t)dt = \inf_{\tau\in\mathbb{R}}\left\{\tau + (1-\alpha)^{-1}\mathbb{E}_{\mathbb{P}}[Z-\tau]_+\right\},\ \alpha\in(0,1). \tag{4.7}$$

Here $\mathcal{Z} = L_1(\Xi, \mathcal{B}, \mathbb{P})$ and a minimizer in the right hand side of (4.7) is $\bar{\tau} = H_Z^{-1}(\alpha)$. The empirical estimate of $\mathsf{AV@R}_{1-\alpha}(\phi^x)$ is then

$$\widehat{\mathsf{AV@R}}_{(1-\alpha)N}(\phi^x) = \inf_{\tau\in\mathbb{R}}\left\{\tau + \frac{1}{(1-\alpha)N}\sum_{j=1}^{N}\left[\phi^x(\xi^j)-\tau\right]_+\right\}. \tag{4.8}$$

We have that $\partial\mathsf{AV@R}_{1-\alpha}(Z)$ is a singleton iff $\mathbb{P}\{Z=\kappa\}=0$, where $\kappa_\alpha := H_Z^{-1}(\alpha)$. Suppose that $\partial\mathsf{AV@R}_{1-\alpha}(Z) = \{\bar{\zeta}\}$ is a singleton. Then

$$\bar{\zeta}(s) = \begin{cases} (1-\alpha)^{-1} & if \quad Z(s) > \kappa,\ s\in\Xi, \\ \quad 0 & if \quad Z(s) < \kappa,\ s\in\Xi, \end{cases} \tag{4.9}$$

(cf. [25, eq. (6.80), page 292]). For $x \in X$ and $Z := \phi^x$ let $\{\bar{\zeta}^x\}$ be the corresponding subdifferential. The subdifferential $\hat{\zeta}^x = (\zeta_1^x, ..., \zeta_N^x)$ of the corresponding empirical estimate is obtained by replacing κ_α with their empirical estimates. That is $\zeta_j^x = (1-\alpha)^{-1}$ if $\phi^x(\xi^j) > \kappa_{\alpha,N}$ and $\zeta_j^x = 0$ if $\phi^x(\xi^j) < \kappa_{\alpha,N}$, where $\kappa_{\alpha,N}$ is the empirical estimate of κ_α. Note that because of the assumption $\mathbb{P}\{Z=\kappa\}=0$, the empirical estimate $\kappa_{\alpha,N}$ converges w.p.1 to κ_α.

Consider the probability distribution P_N^x on $\{\xi^1, ..., \xi^N\}$ associated with density $\hat{\zeta}^x$, i.e., with ξ^j is assigned probability $1/((1-\alpha)N)$ if $\phi^x(\xi^j) > \kappa_{\alpha,N}^x$, and 0 otherwise. We view P_N^x as the empirical counterpart of P^x, where P^x is the probability measure absolutely continuous with respect to $\mathbb{P}$ and having density $\bar{\zeta}^x$, i.e.,

$$dP^x = \bar{\zeta}^x d\mathbb{P}. \tag{4.10}$$

Consider a continuous bounded function $g : \Xi \to \mathbb{R}$. Since $g(\cdot)$ is bounded and continuous, $\kappa_{\alpha,N}^x \to \kappa_\alpha^x$ w.p.1 and $\mathbb{P}\{\phi^x(\xi) = \kappa_\alpha^x\} = 0$, we have that

$$\int_\Xi g(s)dP_N^x(s) = \frac{1}{(1-\alpha)N}\sum_{\phi^x(\xi^j)>\kappa_{\alpha,N}^x} g(\xi^j)$$

converges w.p.1 to

$$\int_\Xi g(s)\bar{\zeta}^x(s)d\mathbb{P}(s) = \frac{1}{1-\alpha}\int_{\phi^x(\xi)>\kappa_\alpha^x} g(s)d\mathbb{P}(s).$$

That is P_N^x converges weakly[5] to P^x. Moreover, by Proposition 7.1 in the Appendix, we have if $\{x_N\}$ is a sequence in X converging to x, then $\int_\Xi g(s)dP_N^{x_N}(s)$ converges to $\int_{\phi^x(\xi)>\kappa} g(s)d\mathbb{P}(s)$ w.p.1, and hence $P_N^{x_N}$ converges weakly to P^x.

[5] Recall that a sequence P_N of probability measures converges weakly to a probability measure P if $\int g dP_N \to \int g dP$ for any bounded continuous function $g : \Xi \to \mathbb{R}$, see e.g., Billingsley [3] for a discussion of weak convergence of probability measures.

Spectral risk measure. This can be extended to a general setting. Let us first consider spectral risk measure $\mathcal{R}$. That is[6]

$$\mathcal{R}(H_Z^{-1}) := \int_0^1 \sigma(t) H_Z^{-1}(t) dt, \tag{4.11}$$

where $\sigma : [0, 1) \to \mathbb{R}_+$ is monotonically nondecreasing, left side continuous function such that $\int_0^1 \sigma(t)dt = 1$. The Average Value-at-Risk $\mathsf{AV@R}_\alpha$ is a spectral risk measure with spectral function $\sigma(t) = 0$ for $t \in [0, 1-\alpha)$, and $\sigma(t) = 1/\alpha$ for $t \in [1-\alpha, 1]$.

Let $H_{\phi^x}(z) := \mathbb{P}\{\phi^x(\xi) \le z\}$ be the cumulative distribution function (CDF) of $\phi^x(\xi)$ and $H_{\phi^x,N}$ be the CDF of $\phi^x(\xi^j)$, $j = 1, .., N$. That is, function $H_{\phi^x,N}(\cdot)$ is stepwise constant with jumps $1/N$ at points $\phi^x_{(1)}, ..., \phi^x_{(N)}$, where $\phi^x_{(1)}, ..., \phi^x_{(N)}$ are values $\phi^x(\xi^1), ..., \phi^x(\xi^N)$ arranged in the increasing order, i.e.,

$$H_{\phi^x,N}(\cdot) = N^{-1} \sum_{j=1}^N \mathbf{1}_{(-\infty, \phi^x_{(j)}]}(\cdot). \tag{4.12}$$

Then

$$\mathcal{R}(H^{-1}_{\phi^x,N}) = \int_0^1 \sigma(t) H^{-1}_{\phi^x,N}(t) dt = \sum_{j=1}^N q_j \phi^x_{(j)}, \tag{4.13}$$

where

$$q_j := \int_{(j-1)/N}^{j/N} \sigma(t)dt, \; j = 1, ..., N. \tag{4.14}$$

Note that $q_j \ge 0$, $\sum_{j=1}^N q_j = \int_0^1 \sigma(t)dt = 1$.

Remark 4.1 The corresponding set $\mathfrak{A}^N$ is the convex hull of vectors $(q_{\pi(1)}, ..., q_{\pi(N)})$, $\pi \in \Pi$, where Π is the set of permutations of the set $\{1, ..., N\}$. By Hardy - Littlewood inequality, we have here

$$\sup_{\zeta \in \mathfrak{A}^N} \sum_{j=1}^N \zeta_j \phi^x(\xi^j) = \sum_{j=1}^N q_j \phi^x_{(j)}, \tag{4.15}$$

and the corresponding maximizer $\bar{\zeta} \in \arg\max_{\zeta \in \mathfrak{A}^N} \sum_{j=1}^N \zeta_j \phi^x(\xi^j)$ is given by $\bar{\zeta} = (q_{\pi(1)}, ..., q_{\pi(N)})$ with the permutation $\pi \in \Pi$ corresponding to the order $\phi^x_{(1)} \le \cdots \le \phi^x_{(N)}$. Note that this permutation and hence the maximizer $\bar{\zeta}$ depend on x.

We can also write this spectral risk measure in the form

$$\mathcal{R}(\phi^x) = \int_0^1 \mathsf{AV@R}_{1-\alpha}(\phi^x) d\mu(\alpha), \tag{4.16}$$

where μ is the probability measure on the interval $[0, 1)$ associated with the spectral function $\sigma(\cdot)$, given by

$$\mu(\alpha) = (1-\alpha)\sigma(\alpha) + \int_0^\alpha \sigma(t)dt.$$

[6]By the law invariance of $\mathcal{R}(Z)$ it can be considered as a function of H_Z.

This is the so-called Kusuoka representation of the spectral risk measure (e.g., [25, p. 307]). That is

$$\mathcal{R}(\phi^x) = \int_0^1 \mathbb{E}_{\mathbb{P}}\{\tau(\alpha) + (1-\alpha)^{-1}[\phi^x - \tau(\alpha)]_+\}d\mu(\alpha), \tag{4.17}$$

where $\tau(\alpha) := H_{\phi^x}^{-1}(\alpha)$. The empirical estimate $\mathcal{R}(H_{\phi^x,N}^{-1})$ can be written then as

$$\mathcal{R}(H_{\phi^x,N}^{-1}) = \frac{1}{N}\int_0^1 \sum_{j=1}^N \Big\{\hat{\tau}_N(\alpha) + (1-\alpha)^{-1}[\phi^x(\xi^j) - \hat{\tau}_N(\alpha)]_+\Big\}d\mu(\alpha), \tag{4.18}$$

where $\hat{\tau}_N(\alpha)$ is the empirical estimate of $H_{\phi^x}^{-1}(\alpha)$.

The subdifferential of $\mathcal{R}(\phi^x)$ can be taken inside the integral in (4.16), i.e.,

$$\partial\mathcal{R}(\phi^x) = \int_0^1 \partial\mathsf{AV@R}_{1-\alpha}(\phi^x)d\mu(\alpha). \tag{4.19}$$

We have that $\partial\mathcal{R}(\phi^x) = \{\bar{\zeta}^x\}$ is a singleton iff $\partial\mathsf{AV@R}_{1-\alpha}(\phi^x)$ is a singleton for μ-almost every $\alpha \in [0,1)$, i.e., iff $\mathbb{P}\{\phi^x = \kappa_\alpha\} = 0$ for μ-almost every $\alpha \in [0,1)$, where $\kappa_\alpha = H_{\phi^x}^{-1}(\alpha)$. Then we have by Example 4.1 that the subdifferential $\partial\mathsf{AV@R}_{1-\alpha}(\phi^x) = \{\bar{\zeta}_\alpha\}$ is given by

$$\bar{\zeta}_\alpha(s) = \begin{cases} (1-\alpha)^{-1} & if \quad \phi^x(s) > \kappa_\alpha,\ s \in \Xi, \\ 0 & if \quad \phi^x(s) < \kappa_\alpha,\ s \in \Xi. \end{cases} \tag{4.20}$$

The subdifferential $\hat{\zeta}_\alpha^x = (\zeta_{\alpha 1}^x, ..., \zeta_{\alpha N}^x)$ of the corresponding empirical estimate is obtained by replacing $\kappa_\alpha = H_{\phi^x}^{-1}(\alpha)$ with their empirical estimates.

For a continuous and bounded function $g : \Xi \to \mathbb{R}$ we have that

$$\int_\Xi g(s)dP_N^x(s) = \int_{\alpha\in[0,1)} \frac{1}{N}\sum_{j=1}^N g(\xi^j)\zeta_{\alpha j}^x d\mu(\alpha) = \int_{\alpha\in[0,1)} \frac{1}{(1-\alpha)N} \sum_{\phi^x(\xi^j)>\kappa_N^\alpha} g(\xi^j)d\mu(\alpha)$$

converges w.p.1 to

$$\int_\Xi g(s)\bar{\zeta}^x(s)d\mathbb{P}(s) = \int_{\alpha\in[0,1)}\int_\Xi g(s)\bar{\zeta}_\alpha^x(s)d\mathbb{P}(s)d\mu(\alpha) = \int_{\alpha\in[0,1)} \frac{1}{1-\alpha}\int_{\phi^x(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\mu(\alpha).$$

Then we have that P_N^x converges weakly to P^x, where P^x has density $\bar{\zeta}^x$ (see (4.10)). Moreover, by Proposition 7.1 in the Appendix, we have if $\{x_N\}$ is a sequence in X converging to x, then $\int_\Xi g(s)dP_N^{x_N}(s)$ converges to $\int_{\phi^x(\xi)>\kappa} g(s)d\mathbb{P}(s)$ w.p.1, and hence $P_N^{x_N}$ converges weakly to P^x.

Law invariant coherent risk measure. By dual representation, any law invariant coherent risk measure can be represented as follows

$$\max_{\zeta\in\mathfrak{A}} \int_\Xi Z(s)\zeta(s)d\mathbb{P}(s) = \mathcal{R}(Z) = \mathcal{R}(H_Z^{-1}) = \sup_{\sigma\in\mathfrak{S}}\int_0^1 \sigma(t)H_{\phi^x}^{-1}(t)dt \tag{4.21}$$

with

$$\mathfrak{S} := \{\sigma = H_\zeta^{-1} : \zeta \in \mathfrak{A}\} \tag{4.22}$$

being a set of spectral functions.

Let $H_{\phi^x,N}$ denote the CDF of the empirical distribution corresponding to the i.i.d. samples $\{\phi^x(\xi^1), \cdots, \phi^x(\xi^N)\}$. Note that $H_{\phi^x,N}$ is a function of the random sample, and hence is random. We have

$$\partial\mathcal{R}(H_{\phi^x}^{-1}) = \arg\max_{\sigma\in\mathfrak{S}} \int_0^1 \sigma(t)H_{\phi^x}^{-1}(t)dt \text{ and } \partial\mathcal{R}\left(H_{\phi^x,N}^{-1}\right) = \arg\max_{\sigma\in\mathfrak{S}} \int_0^1 \sigma(t)H_{\phi^x,N}^{-1}(t)dt. \quad (4.23)$$

Lemma 4.1 *Consider a point $\bar{x} \in X$ and a sequence $\{x_N\} \subset X$ converging to $\bar{x}$. Suppose that Assumptions 3.1 and 4.1 hold, $\phi(\cdot, \xi)$ is continuous and $\partial\mathcal{R}(H_{\phi^{\bar{x}}}^{-1}) = \{\bar{\sigma}\}$ is a singleton. Then any sequence $\sigma_N \in \partial\mathcal{R}\left(H_{\phi^{x_N},N}^{-1}\right)$ weakly* converges to $\bar{\sigma}$ w.p.1.*

Proof. We first note that $H_{\phi^x}^{-1}$ and $H_{\phi^{x_N},N}^{-1}$ belong to the space $\mathcal{L}_p$. We can apply a general theory of sensitivity analysis applied to the optimization problem (4.23) with viewing $H_{\phi^x}^{-1}$ as parameter in the space $\mathcal{L}_p$. We have that $H_{\phi^{x_N},N}^{-1}$ converges w.p.1 to $H_{\phi^x}^{-1}$ in the norm topology of $\mathcal{L}_p$ as $N \to \infty$. This can be proved by an extension of [24, Theorem 2.1] (see Theorem 7.1 in the Appendix). Since the set $\mathfrak{S}$ is weakly* compact and the maximizer $\bar{\sigma}$ of the right hand side of (4.23) is unique, it follows by [4, Lemma 4.3 and example 4.5] that if

$$\sigma_N \in \arg\max_{\sigma\in\mathfrak{S}} \int_0^1 \sigma(t)H_{\phi^{x_N},N}^{-1}(t)dt, \quad (4.24)$$

then $\{\sigma_N\}$ is weak* convergent w.p.1 to $\bar{\sigma}$. □

For law invariant coherent risk measure, by the Kusuoka representation, (4.21) can also be presented as

$$\mathcal{R}(H_{\phi^x}^{-1}) = \sup_{\sigma\in\mathfrak{S}} \int_0^1 \sigma(t)H_{\phi^x}^{-1}(t)dt = \sup_{\mu\in\mathfrak{V}} \int_0^1 \mathrm{AVaR}_{1-\alpha}(\phi^x)d\mu(\alpha) \quad (4.25)$$

and its SAA can be written as

$$\mathcal{R}(H_{\phi^x,N}^{-1}) = \sup_{\mu\in\mathfrak{V}} \frac{1}{N} \int_0^1 \sum_{j=1}^N \left\{\hat{\tau}_N(\alpha) + (1-\alpha)^{-1}[\phi^x(\xi^j) - \hat{\tau}_N(\alpha)]_+\right\}d\mu(\alpha), \quad (4.26)$$

where $\mathfrak{V} := \{\mu : \mu(\alpha) = (1-\alpha)\sigma(\alpha) + \int_0^\alpha \sigma(t)dt, \sigma \in \mathfrak{S}\}$. Then we have that $\partial\mathcal{R}(\phi^x) = \{\bar{\zeta}^x\}$ is a singleton (which implies $\partial\mathcal{R}(H_{\phi^{\bar{x}}}^{-1})$ is a singleton) if and only if $\partial\mathsf{AV@R}_{1-\alpha}(\phi^x)$ is a singleton for μ-almost every $\alpha \in [0, 1)$, i.e., if and only if $\mathbb{P}\{\phi^x = \kappa_\alpha\} = 0$ for μ-almost every $\alpha \in [0, 1)$, where $\kappa_\alpha = H_{\phi^x}^{-1}(\alpha)$, $\mu \in \mathfrak{V}$. Note that if $\mathbb{P}\{\phi^x = \kappa_\alpha\} = 0$ for every $\alpha \in [0, 1)$, then the condition that $\partial\mathcal{R}(H_{\phi^{\bar{x}}}^{-1}) = \{\bar{\sigma}\}$ is a singleton holds.

Then we consider the convergence analysis between

$$\max_{\eta\in\mathfrak{A}^N} \sum_{j=1}^N \eta_j\phi(x_N, \xi^j) = \sup_{\mu\in\mathfrak{V}} \frac{1}{N} \int_0^1 \sum_{j=1}^N \left\{\hat{\tau}_N(\alpha) + (1-\alpha)^{-1}[\phi^{x_N}(\xi^j) - \hat{\tau}_N(\alpha)]_+\right\}d\mu(\alpha) \quad (4.27)$$

and

$$\max_{\eta\in\mathfrak{A}} \int_\Xi \phi^{\bar{x}}\eta d\mathbb{P} = \sup_{\mu\in\mathfrak{V}} \int_0^1 \mathrm{AVaR}_{1-\alpha}(\phi^{\bar{x}})d\mu(\alpha), \quad (4.28)$$

where $x_N \to \bar{x}$. Let $(\zeta_N^{x_N},\ \mu_N)$ and $(\zeta^*, \bar{\mu})$ denote the optimal solutions of (4.27) and (4.28), respectively. Note that $\zeta_N^{x_N}$ corresponds to a discrete distribution $P_N^{x_N}$ and ζ^* corresponds to a continuous distribution $P^* = \zeta^*\mathbb{P}$.

Proposition 4.1 *Consider a point $\bar{x} \in X$ and a sequence $\{x_N\} \subset X$ converging to $\bar{x}$ and the ambiguity set corresponding to a law invariant coherent risk measure. Suppose (i) Assumptions 3.1 and 4.1 hold, $\phi(\cdot, \xi)$ is Lipschitz continuous, (ii) $\partial\mathcal{R}(H^{-1}_{\phi^{\bar{x}}}) = \{\bar{\sigma}\}$ is a singleton, (iii) the CDF of $\phi^{\bar{x}}$ is strictly monotone, and (iv) there exists positive measure $\hat{\mu}$ such that for all N sufficiently large, $\int_{[0,1]} h(t)\hat{\mu}(t) \geq \int_{[0,1]} h(t)\mu_N(t)$ for all bounded function $h(t)$. Then $P_N^{x_N}$ converges weakly to $P^{\bar{x}}$.*

The proof of Proposition 4.1 is in the Appendix.

Example 4.2 *Consider the ψ-divergence approach to construction of the uncertainty sets. The concept of ψ-divergence is originated in Csiszár [13] and Morimoto [17], and was extensively discussed in Ben-Tal and Teboulle [2]. We also can refer to Bayraksan and Love [1] for a recent survey of this approach. That is, consider a convex lower semicontinuous function $\psi : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ such that $\psi(1) = 0$. For $x < 0$ we set $\psi(x) = +\infty$. For $c > 0$ consider*

$$\mathfrak{A} := \left\{\zeta \in \mathfrak{D} : \textstyle\int_\Xi \psi(\zeta(s))d\mathbb{P}(s) \leq c\right\}, \tag{4.29}$$

where $\mathfrak{D} := \{\zeta \in \mathcal{Z}^* : \int \zeta d\mathbb{P} = 1,\ \zeta \succeq 0\}$ denotes the set of densities. If $\zeta \overset{\mathcal{D}}{\sim} \zeta'$, then $\int_\Xi \psi(\zeta(s))d\mathbb{P}(s) = \int_\Xi \psi(\zeta'(s))d\mathbb{P}(s)$. Hence the set $\mathfrak{A}$ and the corresponding functional $\mathcal{R}$ are law invariant. Since ψ-divergence is a law invariant coherent risk measure, it has Kusuoka representation (Note that the representation is only for constructing the SAA and proving the weak convergence).

Proposition 4.2 *[14, Proposition 5.6] A ψ-divergence risk measure can be written in the form*

$$\mathcal{R}(Z) = \sup_{\sigma \in \mathfrak{S}} \int_0^1 \sigma(t) H_Z^{-1}(t)dt, \tag{4.30}$$

where $\mathfrak{S} := \{\sigma : [0,1] \to [0,\infty] : \sigma \text{ is non-decreasing }, \int_0^1 \sigma(t)dt = 1, \int_0^1 \psi(\sigma(t))dt \leq c\}$.

Moreover, let $\mu(\alpha) = (1-\alpha)\sigma(\alpha) + \int_0^\alpha \sigma(t)dt$ (that is $\sigma(t) = \int_0^t \frac{1}{1-\alpha}d\mu(\alpha)$). By Kusuoka representation, we have

$$\mathcal{R}(Z) = \sup_{\mu \in \mathfrak{V}} \int_0^1 \mathsf{AV@R}_{1-\alpha}(Z)d\mu(\alpha) = \sup_{\mu \in \mathfrak{V}} \int_0^1 \mathbb{E}_\mathbb{P}\{\tau(\alpha) + (1-\alpha)^{-1}[Z - \tau(\alpha)]_+\}d\mu(\alpha), \tag{4.31}$$

where $\mathfrak{V} := \{\mu : [0,1) \to [0,\infty] : \int_0^1 d\mu(\alpha) = 1, \int_0^1 \psi(\int_0^t \frac{1}{1-\alpha}d\mu(\alpha))dt \leq c\}$. Although the structure of $\mathfrak{V}$ looks complicated, the discretization way is exactly SAA and same as in the paper. Then with the conditions of Proposition 4.1, we can show $P_N^{x_N}$ converges weakly to $P^{\bar{x}}$.

By discussion above, we have shown that for law invariant coherent risk measure and under mild conditions, $P_N^{x_N}$ converges weakly to $P^{\bar{x}}$.

However, to prove the convergence between (3.3)-(3.4) and (4.4)-(4.5), we need stronger convergence results between $P_N^{x_N}$ and $P^{\bar{x}}$. To this end, we need the following assumption.

Assumption 4.2 *Let $\mathfrak{M}$ and $\mathfrak{M}^N$ be nonempty and closed. Suppose*

(a) there exists a weakly compact set $\hat{\mathfrak{M}} \subset \mathscr{M}$ such that $\mathfrak{M}, \mathfrak{M}^N \subset \hat{\mathfrak{M}}$ holds for N sufficiently large;

(b) $\sup_{P \in \hat{\mathfrak{M}}} \mathbb{E}_P[\|\xi\|]$ is bounded.

Assumption 4.2 is used and discussed in [28]. One sufficient condition for Assumption 4.2 is the compactness of support set Ξ. Then we prove the main convergence result for the case when $r = 1$. It is straightforward to extend the result to the case when $r > 1$.

Theorem 4.1 *Let $(\hat{x}_N, \hat{\zeta}_N) \in X \times \mathfrak{A}^N$ be a solution of the SAA variational inequalities* (4.4) - (4.5)*. Suppose (a) Assumptions 3.1, 4.1 and 4.2 hold, (b) $\phi(x, \cdot)$ is Lipschitz continuous and bounded on Ξ, $\Phi(\cdot, \xi)$ and $\Phi(x, \cdot)$ are Lipschitz continuous with Lipschitz modulus $\kappa(\xi)$ and $\bar{\kappa}$ over X and Ξ respectively, and $\sup_{P \in \hat{\mathfrak{M}}} \mathbb{E}_P[\kappa(\xi)] < \infty$, (c) $\hat{x}_N$ converges w.p.1 to a point $\bar{x}$, (d) $\partial \mathcal{R}(\phi^{\bar{x}}) = \{\bar{\zeta}\}$ is a singleton, (e) $P^{\bar{x}}$ is probability measure on $(\Xi, \mathcal{B})$ with density $\bar{\zeta}$, and $P_N^{\hat{x}_N}$ is the empirical measure associated with $\hat{\zeta}_N$, and $P_N^{\hat{x}_N}$ weakly converges to $P^{\bar{x}}$. Then $(\bar{x}, \bar{\zeta})$ is a solution of the DRVI* (3.3) - (3.4)*.*

The proof of Theorem 4.1 is in the Appendix. Note that the sufficient conditions for assumption (e) are given in Proposition 4.1.

5 Numerical examples

In this section, we use a continuous version of Example 3.2 of the distributionally robust generalized Nash equilibrium problem to illustrate the SAA approach and its convergence, where f_i and g_i are quadratic convex functions, $\mathfrak{M}_i$ is constructed by modified χ^2-distance, $i = 1, 2$. Particularly, let

$$f_i(x, \xi) := \frac{1}{2} x_i^\top \tilde{M}_i(\xi) x_i + \tilde{c}_i(\xi)^\top x_i, \qquad g_i(x) := \frac{1}{2} x_i^\top M_i x_i + c_i^\top x_i + x_i^\top R_i x_{-i},$$

$X_i = \mathbb{R}_+^2$, $b_1 = b_2 = (1, 1)^\top$ and $c = 10$. Let $\mathbb{P}$ follow the uniform distribution over $[-1, 1]$, $\xi : \Omega \to [-1, 1]$, then the density function of $\mathbb{P}$ is a constant function with value $\frac{1}{2}$ over $[-1, 1]$ and the ambiguity set

$$\mathfrak{M}_i := \left\{ P \in \mathscr{P} : \int_{\xi \in [-1,1]} 2(p(\xi) - 1)^2 d\xi \le 0.05 \right\},$$

where $\mathscr{P}$ denotes all probability measures over $[-1,1]$, $p(\xi)$ is the density function of P, $i=1,2$. Note that this is a particular case of ψ-divergence. It is obvious that $\mathfrak{M}_i$ is a weakly compact subset in $\mathcal{L}_2$ over $[-1,1]$. Let E be the 2×2 matrix with all elements 1, $R_i = E$, $\tilde{M}_i(\xi) = 5I + \xi I$ and $M_i = I$. Then for any $\xi^i, \xi^j \in [0,1]$,

$$\begin{pmatrix} \nabla_{x_1x_1} f_1(x,\xi^i) + \nabla_{x_1x_1} g_1(x) & \nabla_{x_1x_2} f_1(x,\xi^i) + \nabla_{x_1x_2} g_1(x) \\ \nabla_{x_2x_1} f_2(x,\xi^j) + \nabla_{x_2x_1} g_2(x) & \nabla_{x_2x_2} f_2(x,\xi^j) + \nabla_{x_2x_2} g_2(x) \end{pmatrix} = \begin{pmatrix} \tilde{M}_1(\xi^i) + M_1 & R_1 \\ R_2 & \tilde{M}_2(\xi^j) + M_2 \end{pmatrix}$$

is positive definite and then for any $\xi^i, \xi^j \in [0,1]$,

$$\begin{pmatrix} \tilde{M}_1(\xi^i) + M_1 & R_1 & b_1 \\ R_2 & \tilde{M}_2(\xi^j) + M_2 & b_2 \\ -b_1^\top & -b_2^\top & 0 \end{pmatrix}$$

is positive semi-definite. Since random variable $\phi^x(\xi)$ in (4.26) is $f_1(x,\xi)$ and $f_2(x,\xi)$ and $\mathbb{P}$ follows the uniform distribution over $[-1,1]$, $f_1(x,\xi)$ and $f_2(x,\xi)$ follow continuous distribution and their α-quantiles, denote by κ_α^1 and κ_α^2, are unique for all $\alpha \in [0,1]$ when $x \neq 0$. In this case, $\mathbb{P}(f_1(x,\xi) = \kappa_\alpha^1) = 0$ and $\mathbb{P}(f_2(x,\xi) = \kappa_\alpha^2) = 0$ for all $\alpha \in [0,1]$, and then $\partial_{f_1}\mathcal{R}(f_1(x,\xi))$ and $\partial_{f_2}\mathcal{R}(f_2(x,\xi))$ are singleton.

Let $\{\xi^1, \cdots, \xi^N\}$ be the i.i.d. sample of ξ generated from $\mathbb{P}$. Then we solve

$$0 \in \sum_{j=1}^N p_i^j \nabla_{x_i} f_i(x,\xi^j) + \nabla_{x_i} g_i(x) + b_i\mu + \mathcal{N}_{X_i}(x_i), \qquad i = 1,2 \tag{5.1}$$

$$0 \in c - b_1x_1 - b_2x_2 + \mathcal{N}_{\mathbb{R}_+}(\mu), \tag{5.2}$$

where for $i=1,2$, $P_i = (p_i^1, \cdots, p_i^N)$ is from

$$P_i \in \arg\max_{Q_i \in \mathfrak{M}_i^N} \frac{1}{N} \sum_{j=1}^N q_i^j \phi_i(x_1,x_2,\xi^j), \tag{5.3}$$

where $\phi(x_1,x_2,\xi) = f_i(x_1,x_2,\xi) + g_i(x_1,x_2)$, $Q_i = (q_i^1, \cdots, q_i^N)$ and

$$\mathfrak{M}_i^N := \left\{ P \in \mathbb{R}_+^N : \sum_{j=1}^N (p^j - \frac{1}{N})^2 \le \frac{0.05}{N},\ \sum_{j=1}^N p_j = 1 \right\}.$$

We consider sample size $N = (50, 100, 300, 600, 1200)$. For each sample size, we generate 20 group of samples and solve the corresponding DRVI (5.1) - (5.3) by Algorithm 1 with $\tau = 0.2$ and randomly generated $z^0 \in [0,1]^5$ using the uniform distribution in Matlab.

Since the two players are symmetric, then $x_1 = x_2$ and we only show x_1 with $x_1 = (x_{11}, x_{12})^\top$ in Figures 1-2. From the two figures, we can observe the tendency of convergence as sample size increases, which is consistent with our convergence results.

Algorithm 1 Projection method for solving DRVI.

1: Choose a parameter $\tau \in (0,1)$ and an initial point $z^0 = ((x_1^0)^\top, (x_2^0)^\top, \mu^0)^\top$. Set $k \leftarrow 0$.

2: Solve

$$P_i^k = \arg \max_{Q_i \in \mathfrak{M}_i^N} \mathbb{E}_{Q_i}[\phi_i(x^k, \xi)], \quad \text{for} \quad i = 1, 2.$$

3: Set

$$F^k(z) = \begin{pmatrix} \sum_{j=1}^N (p_1^j)^k \nabla_{x_1} f_1(x, \xi^j) + \nabla_{x_1} g_1(x) + b_1 \mu \\ \sum_{j=1}^N (p_2^j)^k \nabla_{x_2} f_2(x, \xi^j) + \nabla_{x_2} g_2(x) + b_2 \mu \\ c - b_1 x_1 - b_2 x_2 \end{pmatrix}$$

where $((p_i^1)^k, \cdots, (p_i^N)^k) = P_i^k$, $i = 1, 2$.

4: If $\| \min(z^k, F^k(z^k)) \|_2 \leq 10^{-8}$, stop, otherwise find z^{k+1} such that

$$\|z^{k+1} - \mathrm{Proj}_{\mathbb{R}_+^5}(z^{k+1} - \tau F^k(z^{k+1}))\| \leq 10^{-8}.$$

5: $k \leftarrow k + 1$, go to Step 2.

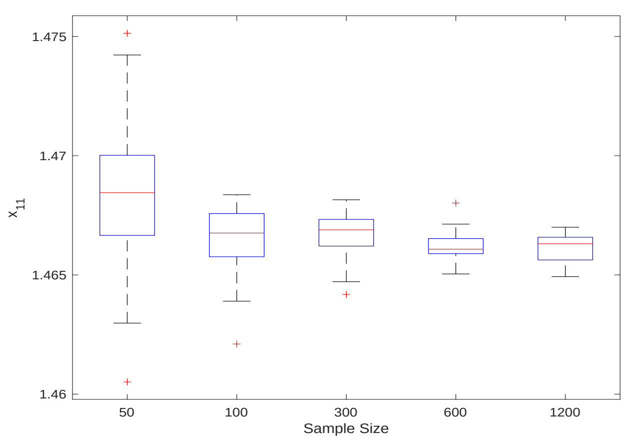

Figure 1: convergence of x_{11}

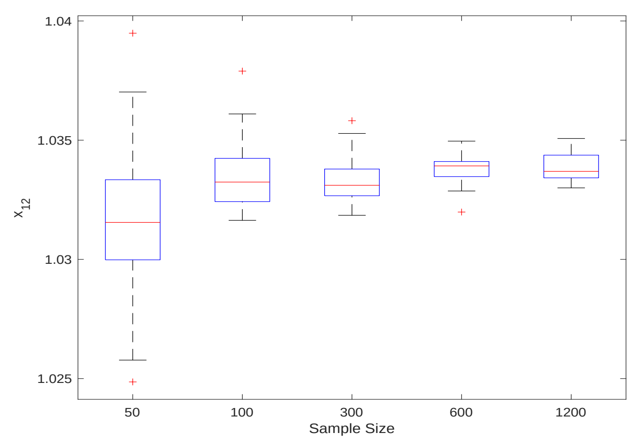

Figure 2: convergence of x_{12}

6 Conclusion remarks

To deal with uncertainties of probability distributions $\mathbb{P}$ in the SVI (1.1), we propose a formulation of the DRVI in Definition 2.1. This formulation provides a unified framework for the research of many important problems including the optimality conditions for distributionally robust optimization and distributionally robust games. We show the existence of solutions of the DRVI under the conditions that the set X of decision variables is convex and bounded or the operator in the DRVI is monotone and coercive. Moreover, under the condition that the set of densities associated with the ambiguity set $\mathfrak{M}$ is law invariant, we propose a SAA approach to the DRVI by using the corresponding law invariant risk measure $\mathcal{R}$ and establish its convergence properties as the sample size N goes to infinity. The formulation of the DRVI, solutions of the DRVI, the monotone condition, the SAA approach and the convergence properties of the SAA are illustrated by seven examples. Within this new DRVI framework, some new algorithms can be developed for finding robust solutions of optimization and equilibrium problems under uncertain environment.

References

[1] Bayraksan, G., Love, D. K.: Data-driven stochastic programming using phi-divergences, in Tutorials in Operations Research, INFORMS, Catonsville, MD (2015)

[2] Ben-Tal, A., Teboulle, M.: Penalty functions and duality in stochastic programming via phi-divergence functionals, Math. Oper. Res. 12, 224–240 (1987)

[3] Billingsley, P.: Convergence of Probability Measures, John Wiley & Sons, New York (1999).

[4] Bonnans J. F., Shapiro, A.: Perturbation Analysis of Optimization Problems, Wiley Series in Probability and Statistics, New York (2000)

[5] Chen, X., Fukushima, M.: Expected residual minimization method for stochastic linear complementarity problems. Math. Oper. Res. 30, 1022–1038 (2005)

[6] Chen, X., Pong, T. K., Wets, R.: Two-stage stochastic variational inequalities: an ERM-solution procedure. Math. Program. 165, 71-112 (2015)

[7] Chen, X., Wets, R., Zhang, Y.: Stochastic variational inequalities: Residual minimization smoothing. SIAM J. Optim. 22, 649–673 (2012)

[8] Chen, X., Sun, H., Xu, H.: Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems. Math. Program. 177, 255-289 (2019)

[9] 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)

[10] Chen, Y., Sun, H., Xu, H.: Decomposition and discrete approximation methods for solving two-stage distributionally robust optimization problems, Optimization-online, (2020)

[11] Chen, Y., Lan, G., Ouyang, Y.: Accelerated schemes for a class of variational inequalities. Math. Program. 165, 113149 (2017)

[12] Chieu, N. H., Trang, N. T. Q.: Coderivative and monotonicity of continuous mappings. Taiwanese J. Math. 16, 353–365 (2012)

[13] Csiszár, I.: Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von markoffschen ketten, Magyar. Tud. Akad. Mat. Kutato Int. Kozls, 8 (1963).

[14] Dommel, P., Pichler, A.: Convex risk measures based on divergence, Optimization-online (2020)

[15] Facchinei, F., Pang, J. S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, Springer-Verlag, New York (2003)

[16] Hadjisavvas, H., Komlósi, S., Schaible, S.: Handbook of Generalized Convexity and Generalized Monotonicity, Springer, New York (2005)

[17] Morimoto, T.: Markov processes and the h-theorem, J. Phys. Soc. Japan. 18, 328 - 333 (1963)

[18] Milz, J., Ulbrich, M.: An approximation scheme for distributionally robust nonlinear optimization. SAIM J. Optim. 30, 1996–2025 (2020)

[19] Morton, S.: Lagrange multipliers revisited, Cowles Commission Discussion Paper No. 403 (1950)

[20] Pardo, L.: Statistical Inference Based on Divergence Measures. Chapman and Hall/CRC, Boca Raton, FL (2005)

[21] Römisch, W.: Stability of Stochastic Programming Problems, in *Stochastic Programming*, Ruszczyński A., Shapiro, A. eds., Elsevier, Amsterdam (2003)

[22] Rockafellar, R. T., Wets, R. J.-B.: Stochastic variational inequalities: single-stage to multistage. Math. Program. 165, 331–360 (2017)

[23] Rockafellar R. T., Sun, J.: Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging. Math. Program. 174, 453-471 (2018)

[24] Shapiro, A.: Consistency of sample estimates of risk averse stochastic programs. J. Appl. Probab. 50, 533-541 (2013)

[25] Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory, second edition. SIAM, Philadelphia (2014)

[26] Shapiro, A.: Distributionally robust stochastic programming. SIAM J. Optim. 27, 2258–2275 (2017)

[27] Shanbhag U. V.: Stochastic variational inequality problems: applications, analysis, and algorithms. INFORMS Tutor. Oper. Res. 71–107 (2013)

[28] Sun, H., Xu, H.: Convergence analysis for distributionally robust optimization and equilibrium problems. Math. Oper. Res. 41, 377-401 (2016)

[29] Sun, H., Chen, X.: Two-stage stochastic variational inequalities: theory, algorithms and application. J. Oper. Res. Soc. China, (2019) https://doi.org/10.1007/s40305-019-00267-8

[30] Xu, H., Liu, Y., Sun, H.: Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. Math. Program. 169, 489529 (2018)

7 Appendix

In this appendix, we give some proofs and necessary results used in this paper.

Proposition 7.1 *Suppose (i) $\phi(\cdot,\xi)$ is Lipschitz continuous in $x\in X$ with a uniform Lipschitz modules k_ϕ, (ii) the CDF of $\phi^{\bar x}$ is strictly monotone, and (iii) $\{x_N\}$ is a sequence in X converging to $\bar x$, (iv) $|\kappa_\alpha^{x'}|$ is bounded by a constant number for all $x'\in\mathcal{B}(x)\cap X$, then for any bounded and continuous function g, $\int_\Xi g(s)dP_N^{x_N}(s)$ converges to $\int_{\phi^{\bar x}(\xi)>\kappa_\alpha^{\bar x}} g(s)d\mathbb{P}(s)$.*

Proof. For any continuous and bounded function $g(s)$,

$$
\begin{aligned}
&\left|\frac{1}{\alpha N}\sum_{\phi^{x_N}(\xi^j)>\kappa_{\alpha,N}^{x_N}} g(\xi^j)-\frac{1}{\alpha}\int_{\phi^{\bar x}(\xi)>\kappa_\alpha^{\bar x}} g(s)d\mathbb{P}(s)\right|\\
\leq\;&\left|\frac{1}{\alpha N}\sum_{\phi^{x_N}(\xi^j)>\kappa_{\alpha,N}^{x_N}} g(\xi^j)-\frac{1}{\alpha N}\sum_{\phi^{\bar x}(\xi^j)>\kappa_\alpha^{\bar x}} g(\xi^j)\right|+\left|\frac{1}{\alpha N}\sum_{\phi^{\bar x}(\xi^j)>\kappa_\alpha^{\bar x}} g(\xi^j)-\frac{1}{\alpha}\int_{\phi^{\bar x}(\xi)>\kappa_\alpha^{\bar x}} g(s)d\mathbb{P}(s)\right|.
\end{aligned}
\tag{7.1}
$$

We first prove that $\kappa_{\alpha,N}^{x_N}$ converges to $\kappa_\alpha^{\bar x}$ w.p.1. To see this, by condition (ii), we have $\mathbb{P}\{\phi^{\bar x}(\xi)=\kappa_\alpha^{\bar x}\}=0$, then

$$\kappa_\alpha^{\bar x}=\arg\min_\tau\ \tau+\frac{1}{1-\alpha}\mathbb{E}_\mathbb{P}[(\phi^{\bar x}-\tau)_+]$$

and

$$\kappa_{\alpha,N}^{x_N}\in\arg\min_\tau\ \tau+\frac{1}{(1-\alpha)N}\sum_{j=1}^N(\phi^{x_N}(\xi^j)-\tau)_+.$$

It is easy to observe that $(\phi^{\bar{x}}(\xi)-\tau)_+$ is continuous w.r.t. x and dominated by an integrable function, then by uniform law of large numbers [25, Theorem 7.48]

$$\left|\mathbb{E}_{\mathbb{P}}[(\phi^{\bar{x}}-\tau)_+] - \frac{1}{N}\sum_{j=1}^{N}(\phi^{x_N}(\xi^j)-\tau)_+\right| \to 0$$

as $N\to\infty$ w.p.1. Then by conditions (i)-(iv) and [4, Proposition 4.4], we have $\kappa_{\alpha,N}^{x_N}$ converges to $\kappa_\alpha^{\bar{x}}$ w.p.1.

Then we prove the convergence of first part in the right side of (7.1). Let

$$A_N^1 = \{\xi\in\Xi : \phi^{x_N}(\xi) > \kappa_{\alpha,N}^{x_N}, \phi^{\bar{x}}(\xi) \le \kappa_\alpha^{\bar{x}}\},\ \ A_N^2 = \{\xi\in\Xi : \phi^{x_N}(\xi) \le \kappa_{\alpha,N}^{x_N}, \phi^{\bar{x}}(\xi) > \kappa_\alpha^{\bar{x}}\}$$

$$A_N^3 = \{\xi\in\Xi : \phi^{\bar{x}}(\xi) \ge \kappa_{\alpha,N}^{x_N} - k_\phi\|\bar{x}-x_N\|, \phi^{\bar{x}}(\xi) \le \kappa_\alpha^{\bar{x}}\},$$

and

$$A_N^4 = \{\xi\in\Xi : \phi^{\bar{x}}(\xi) \le \kappa_{\alpha,N}^{x_N} + k_\phi\|\bar{x}-x_N\|, \phi^{\bar{x}}(\xi) \ge \kappa_\alpha^{\bar{x}}\}.$$

Then

$$\begin{aligned}\left|\frac{1}{\alpha N}\sum_{\phi^{x_N}(\xi^j)>\kappa_{\alpha,N}^{x_N}} g(\xi^j) - \frac{1}{\alpha N}\sum_{\phi^{\bar{x}}(\xi^j)>\kappa_\alpha^{\bar{x}}} g(\xi^j)\right| &\le \tfrac{1}{\alpha}\mathbb{P}_N(A_N^1\cup A_N^2)|\max_s g(s)| \\ &\le \tfrac{1}{\alpha}\mathbb{P}_N(A_N^3\cup A_N^4)|\max_s g(s)|,\end{aligned}$$

where $\mathbb{P}_N$ is an empirical estimation of $\mathbb{P}$. Note that $\kappa_{\alpha,N}^{x_N}\to\kappa_\alpha^{\bar{x}}$ and $x_N\to\bar{x}$, $A_N^1\subset A_N^3$ and $A_N^2\subset A_N^4$, and A_N^3 and A_N^4 converge to singleton sets. Then by condition (i) and (ii), $\mathbb{P}_N(A_N^1\cup A_N^2)\le\mathbb{P}_N(A_N^3\cup A_N^4)\to 0$ as $N\to\infty$ w.p.1, which implies

$$\left|\frac{1}{\alpha N}\sum_{\phi^{x_N}(\xi^j)>\kappa_{\alpha,N}^{x_N}} g(\xi^j) - \frac{1}{\alpha N}\sum_{\phi^{x_N}(\xi^j)>\kappa_\alpha^{\bar{x}}} g(\xi^j)\right| \to 0 \tag{7.2}$$

as $N\to\infty$ w.p.1.

Then we consider the second part in the right side of (7.1). Since g is continuous and bounded, by classical law of large numbers, as $N\to\infty$ w.p.1,

$$\left|\frac{1}{\alpha N}\sum_{\phi^{\bar{x}}(\xi^j)>\kappa_\alpha^{\bar{x}}} g(\xi^j) - \frac{1}{\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha^{\bar{x}}} g(s)d\mathbb{P}(s)\right| \to 0.$$

Combining discussion above, we have the conclusion. □

Then we derive a kind of uniform Glivenko-Cantelli theorem which we need in the proof of Lemma 4.1. Let $f(x,\xi)$ be a random function and $\{x_N\}\to x$ as $N\to\infty$. Moreover, suppose $f(x,\xi)$ is Lipschitz continuous w.r.t. x and ξ, and the Lipschitz modules $\kappa(\xi)$ of $f(\cdot,\xi)$ is integrable. We use $H_{x_N}(t)$ and $H_x(t)$ to denote the CDF of $f(x_N,\xi)$ and $f(x,\xi)$ w.r.t. $\mathbb{P}$ and $H_{x_N}^N(t)$ and $H_x^N(t)$ are used to denote the CDF of their empirical distributions i.i.d samples $\{\xi^1,\cdots,\xi^N\}$.

Lemma 7.1 *Suppose $f(x,\xi)$ is integrable and continuous w.r.t. x, and $\mathbb{P}$ is a continuous distribution. Then for each $\epsilon > 0$, there exists a finite partition of the real line of the form $-\infty = t_0 < t_1 < \cdots < t_k = \infty$ such that for $0 \le j \le k-1$, $H(x_N, t_{j+1}) - H(x_N, t_j) \le \epsilon$ for all N sufficiently large.*

Proof. Since $\mathbb{P}$ is a continuous distribution, $H_x(t)$ is a CDF of continuous distribution and then, for any $\epsilon > 0$ there exists $-\infty = t_0 < t_1 < \cdots < t_k = \infty$ such that for $0 \le j \le k-1$, $H_x(t_{j+1}) - H_x(t_j) \le \frac{\epsilon}{2}$. Moreover, since $f(x,\xi)$ is integrable and continuous w.r.t. x, by Lebesgue's dominated convergence theorem, for any continuous and bounded function h,

$$\lim_{N\to\infty} \mathbb{E}[h(f(x,\xi)) - h(f(x_N,\xi))] = \mathbb{E}\left[\lim_{N\to\infty}(h(f(x,\xi)) - h(f(x_N,\xi)))\right] = 0.$$

Then $f(x,\cdot)$ converges to $f(x_N,\cdot)$ weakly, which equivalents to $\lim_{n\to\infty} |H_{x_N}(t) - H_x(t)| = 0$ for any $t \in \mathbb{R}$. Then there exists sufficiently large n such that $\sup_{j\in\{0,\cdots,k\}} |H_{x_N}(t_j) - H_x(t_j)| \le \frac{\epsilon}{4}$. Then we have

$$\begin{aligned} |H_{x_N}(t_{j+1}) - H_{x_N}(t_j)| &\le |H_{x_N}(t_{j+1}) - H_x(t_{j+1})| + |H_x(t_{j+1}) - H_x(t_j)| + |H_x(t_j) - H_{x_N}(t_j)| \\ &\le \tfrac{\epsilon}{4} + \tfrac{\epsilon}{2} + \tfrac{\epsilon}{4} = \epsilon. \end{aligned}$$

□

Theorem 7.1 *Suppose $f(x,\xi)$ is Lipschitz continuous w.r.t. x and ξ, and the Lipschitz modules $\kappa(\xi)$ of $f(\cdot,\xi)$ is integrable, $f(x,\cdot) \in \mathcal{L}_P(\Xi,\mathcal{F},\mathbb{P})$ and $\mathbb{P}$ is a continuous distribution. Then w.p.1*

$$\lim_{N\to\infty} \sup_{t\in\mathbb{R}} |H^N_{x_N}(t) - H_x(t)| = 0, \tag{7.3}$$

and $(H^N_{x_N})^{-1}$ converges w.p.1 to H_x^{-1} in the norm topology of $\mathcal{L}_p$ as $N \to \infty$.

Proof. Note that

$$|H^N_{x_N}(t) - H_x(t)| \le |H^N_{x_N}(t) - H_{x_N}(t)| + |H_{x_N}(t) - H_x(t)|.$$

It is sufficiently to show that for any $\epsilon > 0$,

$$\limsup_{N\to\infty} \sup_t |H^N_{x_N}(t) - H_{x_N}(t)| \le \epsilon \tag{7.4}$$

and

$$\limsup_{N\to\infty} \sup_t |H_{x_N}(t) - H_x(t)| \le \epsilon. \tag{7.5}$$

We consider (7.4) firstly. By Lemma 7.1, there exists $-\infty = t_0 < t_1 < \cdots < t_k = \infty$ such that for $0 \le j \le k-1$, $H_{x_N}(t_{j+1}) - H_{x_N}(t_j) \le \frac{\epsilon}{2}$ for all n sufficiently large. For any t, there exists j such that $t_j \le t \le t_{j+1}$. For such j,

$$H^N_{x_N}(t_j) \le H^N_{x_N}(t) \le H^N_{x_N}(t_{j+1}) \text{ and } H_{x_N}(t_j) \le H_{x_N}(t) \le H_{x_N}(t_{j+1}),$$

which implies

$$H^N_{x_N}(t_j) - H_{x_N}(t_{j+1}) \le H^N_{x_N}(t) - H_{x_N}(t) \le H^N_{x_N}(t_{j+1}) - H_{x_N}(t_j).$$

Then we have

$$H^N_{x_N}(t_j) - H_{x_N}(t_j) + H_{x_N}(t_j) - H_{x_N}(t_{j+1}) \leq H^N_{x_N}(t) - H_{x_N}(t)$$

and

$$H^N_{x_N}(t_{j+1}) - H_{x_N}(t_{j+1}) + H_{x_N}(t_{j+1}) - H_{x_N}(t_j) \geq H^N_{x_N}(t) - H_{x_N}(t).$$

Note that by Lemma 7.1 and by uniform law of large numbers [25, Theorem 7.48], $H_{x_N}(t_{j+1}) - H_{x_N}(t_j) \leq \frac{\epsilon}{2}$ and $|H^N_{x_N}(t_{j+1}) - H_{x_N}(t_j)| \leq \frac{\epsilon}{4}$ for all N sufficiently large and $j = 0, \cdots, k$, then we have (7.4). Now we consider (7.5). Similar as the procedure above, For any t, there exists j such that $t_j \leq t \leq t_{j+1}$. For such j,

$$H_x(t_j) \leq H_x(t) \leq H_x(t_{j+1}) \text{ and } H_{x_N}(t_j) \leq H_{x_N}(t) \leq H(x_N, t_{j+1}).$$

Then by continuous distribution of $\mathbb{P}$, Lipschitz continuous of $f(x,\xi)$ w.r.t. x and Lemma 7.1, for any $t \in \mathbb{R}$,

$$\begin{aligned}
|H_{x_N}(t) - H_x(t)| &\leq |H_x(t) - H_{x_N}(t_j)| + |H_{x_N}(t_j) - H_x(t_j)| + |H_x(t_j) - H_x(t)| \\
&\leq |H_{x_N}(t_{j+1}) - H_{x_N}(t_j)| + |H_{x_N}(t_j) - H_x(t_j)| + |H_x(t_j) - H_x(t_{j+1})| \\
&\leq \epsilon.
\end{aligned}$$

Combine (7.4) and (7.5), we have (7.3).

Moreover, (7.3) implies $(H^N_{x_N})^{-1}$ pointwise converges to H_x^{-1} on the set $[0,1]$. Then, if the sequence $\{|(H^N_{x_N})^{-1}(s) - H_x^{-1}(s)|^p\}$ is uniformly integrable, $(H^N_{x_N})^{-1}$ converges w.p.1 to H_x^{-1} in the norm topology of $\mathcal{L}_p$ as $N \to \infty$, that is w.p.1

$$\lim_{N\to\infty} \int_0^1 |(H^N_{x_N})^{-1}(s) - H_x^{-1}(s)|^p ds = \int_0^1 \lim_{N\to\infty} |(H^N_{x_N})^{-1}(s) - H_x^{-1}(s)|^p ds = 0,$$

where the first equality comes from the Lebesgue's dominated convergence theorem.

Let us show that the uniform integrability indeed holds. By triangle inequality,

$$|(H^N_{x_N})^{-1}(s) - H_x^{-1}(s)|^p \leq |(H^N_{x_N})^{-1}(s)|^p + |H_x^{-1}(s)|^p.$$

Then we only need to show the uniform integrability of $|(H^N_{x_N})^{-1}(s)|^p$. Note that

$$\int_0^1 |(H^N_{x_N})^{-1}(s)|^p ds = \int_\Xi |f(x_N,\xi)|^p dH^N_{x_N} = \frac{1}{N}\sum_{i=1}^N |f(x_N,\xi^i)|^p.$$

Since the Lipschitz continuity of $f(x,\xi)$ with Lipschitz modules $\kappa(\xi)$,

$$\begin{aligned}
|\tfrac{1}{N}\textstyle\sum_{i=1}^N |f(x_N,\xi^i)|^p - \mathbb{E}_\mathbb{P}[|f(x,\xi)|^p]| &\leq |\tfrac{1}{N}\textstyle\sum_{i=1}^N |f(x_N,\xi^i)|^p - \tfrac{1}{N}\sum_{i=1}^N |f(x,\xi^i)|^p| \\
&+ |\tfrac{1}{N}\textstyle\sum_{i=1}^N |f(x,\xi^i)|^p - \mathbb{E}_\mathbb{P}[|f(x,\xi)|^p]| \\
&\leq |\tfrac{1}{N}\textstyle\sum_{i=1}^N \kappa(\xi)(x - x_N)|^p \\
&+ |\tfrac{1}{N}\textstyle\sum_{i=1}^N |f(x,\xi^i)|^p - \mathbb{E}_\mathbb{P}[|f(x,\xi)|^p]|.
\end{aligned}$$

Moreover, by the Law of Large Numbers and $x_N \to x$, $\frac{1}{N}\sum_{i=1}^N \kappa(\xi) \to \mathbb{E}_{\mathbb{P}}[\kappa(\xi)]$, $|\frac{1}{N}\sum_{i=1}^N \kappa(\xi)(x - x_N)|^p \to 0$ and $|\frac{1}{N}\sum_{i=1}^N |f(x,\xi^i)|^p - \mathbb{E}_{\mathbb{P}}[|f(x,\xi)|^p]| \to 0$ as $N \to \infty$ w.p.1. It follows that $|(H_{x_N}^N)^{-1}(s)|^p$ converges w.p.1 to a finite limit, which implies that w.p.1 $|(H_{x_N}^N)^{-1}(s)|^p$ is uniformly integrable. $\square$

Proof of Proposition 4.1. For any continuous and bounded function $g : \Xi \to \mathbb{R}$, we have that

$$\int_\Xi g(s)\bar{\zeta}^{\bar{x}}(s)d\mathbb{P}(s) = \int_{[0,1)}\int_\Xi g(s)\bar{\zeta}_\alpha^{\bar{x}}(s)d\mathbb{P}(s)d\bar{\mu}(\alpha) = \int_{[0,1)}\frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\bar{\mu}(\alpha),$$

where $\bar{\mu}$ is corresponding to $\bar{\sigma}$. Moreover,

$$\int_\Xi g(s)dP_N^{x_N}(s) = \int_{[0,1)}\frac{1}{N}\sum_{j=1}^N g(\xi^j)(\zeta_j^{x_N})_\alpha d\mu_N(\alpha) = \int_{[0,1)}\frac{1}{(1-\alpha)N}\sum_{\phi^{x_N}(\xi^j)>\kappa_{N,x_N}^\alpha} g(\xi^j)d\mu_N(\alpha).$$

Then

$$\begin{aligned}
&|\textstyle\int_\Xi g(s)dP_N^{x_N}(s) - \int_\Xi g(s)\bar{\zeta}^{\bar{x}}(s)d\mathbb{P}(s)| \\
&\le \left|\int_{[0,1)}\frac{1}{(1-\alpha)N}\sum_{\phi^{x_N}(\xi^j)>\kappa_{N,x_N}^\alpha} g(\xi^j)d\mu_N(\alpha) - \int_{[0,1)}\frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\mu_N(\alpha)\right| \\
&+ \left|\int_{[0,1)}\frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\mu_N(\alpha) - \int_{[0,1)}\frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\bar{\mu}(\alpha)\right|.
\end{aligned} \tag{7.6}$$

We first prove

$$\left|\int_{[0,1)}\frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\mu_N(\alpha) - \int_{[0,1)}\frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\bar{\mu}(\alpha)\right| \to 0 \tag{7.7}$$

as $N \to \infty$. From condition (iii), $g(\xi)$ is continuous and bounded and $\phi^{\bar{x}}(\xi)$ is continuous w.r.t. ξ, then for any $\alpha' \to \alpha$, $\alpha', \alpha \in [0,1)$, we have

$$\left|\int_{\phi^{\bar{x}}(\xi)>\kappa_{\alpha'}} g(s)d\mathbb{P}(s) - \int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)\right| \le \mathbb{P}((A_{\alpha'} - A_\alpha)\cup(A_\alpha - A_{\alpha'}))\max_s g(s),$$

where $A_\alpha = \{\xi : \phi^{\bar{x}}(\xi) > \kappa_\alpha\}$ and $A_{\alpha'} = \{\xi : \phi^{\bar{x}}(\xi) > \kappa_{\alpha'}\}$. Note that $a' \to a$ and the CDF of $\phi^{\bar{x}}$ is strictly monotone, $A_{\alpha'} \to A_\alpha$ and $\mathbb{P}((A_{\alpha'} - A_\alpha)\cup(A_\alpha - A_{\alpha'})) \to 0$. Then we have $\frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)$ is continuous and bounded w.r.t. α, and (7.7) is from the fact that μ_N weak* converges to μ. Indeed, by Lemma 4.1, σ_N weak* converges to $\bar{\sigma}$, then for any continuous and bounded function $g(t)$, $\int_{[0,1)} g(t)\sigma_N(t)dt \to \int_{[0,1)} g(t)\bar{\sigma}(t)dt$ as $N \to \infty$. Then

$$\begin{aligned}
|\textstyle\int_{[0,1)} g(\alpha)\mu_N(\alpha)d\alpha - \int_{[0,1)} g(\alpha)\bar{\mu}(\alpha)d\alpha| &= (1-\alpha)|\textstyle\int_{[0,1)} g(\alpha)\sigma_N(\alpha)d\alpha - \int_{[0,1)} g(\alpha)\bar{\sigma}(\alpha)d\alpha| \\
&+ \textstyle\int_{[0,1)}\int_0^\alpha g(\alpha)(\sigma_N(t) - \bar{\sigma}(t))dtd\alpha \to 0
\end{aligned}$$

as $N \to \infty$, which implies μ_N weak* converges to $\bar{\mu}$.

Then we prove

$$\left| \int_{[0,1)} \frac{1}{(1-\alpha)N} \sum_{\phi^{x_N}(\xi^j) > \kappa^{\alpha}_{N,x_N}} g(\xi^j) d\mu_N(\alpha) - \int_{[0,1)} \frac{1}{1-\alpha} \int_{\phi^{\bar{x}}(\xi) > \kappa_\alpha} g(s) d\mathbb{P}(s) d\mu_N(\alpha) \right| \to 0.$$

Note that $\phi^{\bar{x}}(\xi)$ is Lipschitz continuous w.r.t. x, the given g is continuous and bounded function w.r.t. ξ and $\{\xi^j\}_{j=1}^N$ is i.i.d. samples from $\mathbb{P}$, both $\frac{1}{(1-\alpha)N} \sum_{\phi^{x_N}(\xi^j) > \kappa^{\alpha}_{N,x_N}} g(\xi^j)$ and $\frac{1}{1-\alpha} \int_{\phi^{\bar{x}}(\xi) > \kappa_\alpha} g(s) d\mathbb{P}(s)$ are bounded by $\max_{s\in[0,1]} g(s)$ and by Proposition 7.1

$$\lim_{N\to\infty} \left| \frac{1}{(1-\alpha)N} \sum_{\phi^{x_N}(\xi^j) > \kappa^{\alpha}_{N,x_N}} g(\xi^j) - \frac{1}{1-\alpha} \int_{\phi^{\bar{x}}(\xi) > \kappa_\alpha} g(s) d\mathbb{P}(s) \right| = 0.$$

We then have

$$\begin{aligned}
&\lim_{N\to\infty} \left| \int_{[0,1)} \frac{1}{(1-\alpha)N} \sum_{\phi^{x_N}(\xi^j) > \kappa^{\alpha}_{N,x_N}} g(\xi^j) d\mu_N(\alpha) - \int_{[0,1)} \frac{1}{1-\alpha} \int_{\phi^{\bar{x}}(\xi) > \kappa_\alpha} g(s) d\mathbb{P}(s) d\mu_N(\alpha) \right| \\
&\le \lim_{N\to\infty} \int_{[0,1)} \left| \frac{1}{\alpha N} \sum_{\phi^{x_N}(\xi^j) > \kappa^{\alpha}_{N,x_N}} g(\xi^j) - \int_{[0,1)} \frac{1}{\alpha} \int_{\phi^{\bar{x}}(\xi) > \kappa_\alpha} g(s) d\mathbb{P}(s) \right| d\mu_N(\alpha) \\
&\le \lim_{N\to\infty} \int_{[0,1)} \left| \frac{1}{\alpha N} \sum_{\phi^{x_N}(\xi^j) > \kappa^{\alpha}_{N,x_N}} g(\xi^j) - \int_{[0,1)} \frac{1}{\alpha} \int_{\phi^{\bar{x}}(\xi) > \kappa_\alpha} g(s) d\mathbb{P}(s) \right| d\hat{\mu}(\alpha) \\
&= 0,
\end{aligned}$$

where the second inequality is from condition (iii) and the third equality is from Lebesgue's dominated convergence theorem.

Combining the above analysis, we have (7.6), that is then $P_N^{x_N}$ converges weakly to $P^{\bar{x}}$. $\square$

Proof of Theorem 4.1. By conditions (c) - (e),

$$P^{\bar{x}} \in \arg\max_{Q\in\mathfrak{M}} \mathbb{E}_Q[\phi(\bar{x}, \xi)].$$

Then we only need to prove $\bar{x}$ is a solution of (3.3), that is equivalent to

$$0 \in \mathbb{E}_{P^{\bar{x}}}[\Phi(\bar{x}, \xi)] + \mathcal{N}_X(\bar{x}). \tag{7.8}$$

Since $\hat{x}_N \to \bar{x}$,

$$\limsup_{N\to\infty} \mathcal{N}_X(\hat{x}_N) \subset \mathcal{N}_X(\bar{x}). \tag{7.9}$$

Moreover,

$$\begin{aligned}
\|\mathbb{E}_{P_N^{\hat{x}_N}}[\Phi(\hat{x}_N, \xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(\bar{x}, \xi)]\| &\le \|\mathbb{E}_{P_N^{\hat{x}_N}}[\Phi(\hat{x}_N, \xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(\hat{x}_N, \xi)]\| \\
&+ \|\mathbb{E}_{P^{\bar{x}}}[\Phi(\hat{x}_N, \xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(\bar{x}, \xi)]\|.
\end{aligned}$$

Note that since $P_N^{\hat{x}_N} \to P^{\bar{x}}$ weakly and by Assumption 4.2 (b), $P_N^{\hat{x}_N} \to P^{\bar{x}}$ under Wasserstein metric [21]. Then by condition (b), we have for any N, $\Phi(\hat{x}_N, \cdot)$ is Lipschitz continuous and

$$\lim_{N\to\infty} \sup_{z\in\bar{Z}} \|\mathbb{E}_{P_N^{\hat{x}_N}}[\Phi(z,\xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(z,\xi)]\| = 0, \tag{7.10}$$

where $\bar{Z} := \{\hat{x}_N, N = 1, 2, \cdots\}$. Moreover,

$$\begin{aligned}
\lim_{N\to\infty} \|\mathbb{E}_{P^{\bar{x}}}[\Phi(\bar{x},\xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(\hat{x}_N,\xi)]\| &\leq \lim_{N\to\infty} \mathbb{E}_{P^{\bar{x}}}[\kappa(\xi)]\|\bar{x} - \hat{x}_N\| \\
&\leq \lim_{N\to\infty} \sup_{P\in\hat{\mathfrak{M}}} \mathbb{E}_P[\kappa(\xi)]\|\bar{x} - \hat{x}_N\| \\
&= 0.
\end{aligned} \tag{7.11}$$

Combining (7.10)-(7.11), we have

$$\lim_{N\to\infty} \|\mathbb{E}_{P_N^{\hat{x}_N}}[\Phi(\hat{x}_N,\xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(\bar{x},\xi)]\| = 0, \tag{7.12}$$

which implies (7.8). $\square$