Learning of Nash Equilibria in Risk-Averse Games

Zifan Wang, Yi Shen, Michael M. Zavlanos, and Karl H. Johansson

***Abstract*— This paper considers risk-averse learning in convex games involving multiple agents that aim to minimize their individual risk of incurring significantly high costs. Specifically, the agents adopt the conditional value at risk (CVaR) as a risk measure with possibly different risk levels. To solve this problem, we propose a first-order risk-averse leaning algorithm, in which the CVaR gradient estimate depends on an estimate of the Value at Risk (VaR) value combined with the gradient of the stochastic cost function. Although estimation of the CVaR gradients using finitely many samples is generally biased, we show that the accumulated error of the CVaR gradient estimates is bounded with high probability. Moreover, assuming that the risk-averse game is strongly monotone, we show that the proposed algorithm converges to the risk-averse Nash equilibrium. We present numerical experiments on a Cournot game example to illustrate the performance of the proposed method.**

I. Introduction

Convex games find applications in many domains ranging from online marketing [1] to transportation networks [2]. In these convex games, agents simultaneously take actions to minimize their loss functions, which are influenced by the actions taken by other agents. The concept of a Nash equilibrium is central in the analysis of such games and represents a stationary point from which no agent has an incentive to deviate, see, e.g., [3], [4].

An important consideration, especially in high-stakes applications, is sensitivity of the learned decisions to the possible risks due to the presence of uncertainty [5]–[8]. For example, in portfolio management [6], constructing a risk-averse portfolio rather than one that yields the highest expected return in a market with much uncertainty is preferred, since it reduces the risk of suffering from large losses. Additionally, in clinical trials [7], a drug that has high average performance and high probability of negative effects may not be desirable due to the safety critical nature of this application. The key idea in risk-averse learning is to replace the expectation in the objective function by a more general objective that employs measures of risk and considers the whole distribution of the stochastic cost. Popular risk measures include mean-deviation functionals [9], Value at risk (VaR) [10]–[12] and Conditional Value at Risk (CVaR) [13]–[15].

In this paper, we consider learning in risk-averse games with continuous action sets in which agents aim to minimize their risk-averse cost functions. Specifically, we employ CVaR as the risk measure and provide a sufficient condition that guarantees the strong monotonicity for the risk-averse game. We assume that the agents are capable of observing other agents actions and computing the gradient of their own stochastic cost function. Although computing the CVaR gradient is usually computationally difficult, we show that the CVaR gradient can be expressed as a function of the VaR value and the gradient of the stochastic cost function. Building on this insight, we develop VaR estimates using all historical samples and then use these VaR estimates to construct the CVaR gradient estimates. By constructing an upper confidence bound for the VaR estimate errors, we show that the proposed algorithm converges to the risk-averse Nash equilibrium with high probability under the established strong monotonicity condition. Finally, we present numerical experiments on a Cournot game to verify our results.

To the best of our knowledge, the analysis of risk-averse games is underexplored in the literature, with a few exceptions [16]–[19]. Specifically, [16] considers agents that aim to maximize the probability of receiving maximum reward instead of the expected payoff, and shows that risk-averse Nash equilibria always exist. However, specific algorithms that converge to these risk-averse equilibria are not proposed. The authors in [17] define a new concept of risk-averse equilibria in finite games using the mean-variance as the risk measure and prove that such equilibria exist. To find these equilibria, a fictitious play algorithm is proposed, but no theoretical convergence analysis is provided. The works in [18], [19] investigate risk-averse games with continuous action sets and propose several risk-averse learning algorithms, which all rely on one-point zeroth-order optimization and achieve no-regret learning with high probability. These works focus on the regret analysis but not on the Nash equilibrium convergence considered here.

The rest of the paper is organized as follows. Section II introduces preliminaries about convex games and the VaR and CVaR risk measures. In Section III, we formally define the risk-averse games and establish the strong monotonicity for such games. In Section IV, we propose a first-order risk-averse learning algorithm and analyze its convergence rate. The performance of the proposed method is illustrated in Section V via an online market problem. Finally, we conclude the paper in Section VI.

* This work was supported in part by Swedish Research Council Distinguished Professor Grant 2017-01078, Knut and Alice Wallenberg Foundation, Wallenberg Scholar Grant, the Swedish Strategic Research Foundation CLAS Grant RIT17-0046, AFOSR under award #FA9550-19-1-0169, and NSF under award CNS-1932011.

Zifan Wang and Karl H. Johansson are with Division of Decision and Control Systems, School of Electrical Enginnering and Computer Science, KTH Royal Institute of Technology, and also with Digital Futures, SE-10044 Stockholm, Sweden. Email: {zifanw,kallej}@kth.se.

Yi Shen and Michael M. Zavlanos are with the Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA. Email: {yi.shen478, michael.zavlanos}@duke.edu

II. PRELIMINARIES

A. Convex Games

Consider a repeated game $\mathcal{G}$ with N agents, whose goal is to learn their best individual actions that minimize their local loss functions. At each episode, each agent selects an action x_i from the convex set $\mathcal{X}_i \subseteq \mathbb{R}^d$ and receives a cost value of $J_i(x_i, x_{-i}) : \mathcal{X} \to \mathbb{R}$, where x_{-i} represents all agents' actions except for agent i, and $\mathcal{X} = \Pi_{i=1}^N \mathcal{X}_i$ is the joint action space. For ease of notation, we usually collect all agents' actions in a vector $x := (x_1, \dots, x_N)$. The game is formally defined as

$$\min_{x_i \in \mathcal{X}_i} J_i(x_i, x_{-i}). \tag{1}$$

We call the game (1) a convex game if $J_i(x_i, x_{-i})$ is convex in x_i for all $x_{-i} \in \mathcal{X}_{-i}$, where $\mathcal{X}_{-i} = \Pi_{j \neq i} \mathcal{X}_j$. As shown in [20], a convex game always has at least one Nash equilibrium. We denote by x^* a Nash equilibrium of the game (1) and this point satisfies $J_i(x^*) \leq J_i(x_i, x^*_{-i})$, for all $x_i \in \mathcal{X}_i$, $i = 1, \dots, N$. At this Nash equilibrium point, agents are strategically stable in the sense that each agent lacks incentives to change its action. Since the loss functions of the agents are convex, the Nash equilibrium can also be characterized by the first-order optimality condition, that is, $\langle \nabla_{x_i} J_i(x^*), x_i - x_i^* \rangle \geq 0$, $\forall x_i \in \mathcal{X}_i$, where $\nabla_{x_i} J_i(x)$ is the partial derivative of the loss function of agent i with respect to his own action. We write $\nabla_i J_i(x)$ instead of $\nabla_{x_i} J_i(x)$ whenever it is clear from the context. In general, it is not easy to show convergence to a fixed Nash equilibrium for games with multiple Nash Equilibria. For this reason, recent studies focus on games that are so-called strongly monotone and are well-known to have a unique Nash equilibrium [20]. The game (1) is said to be m-strongly monotone if for all $x, x' \in \mathcal{X}$, we have $\sum_{i=1}^N \langle \nabla_i J_i(x) - \nabla_i J_i(x'), x_i - x_i' \rangle \geq m \|x - x'\|^2$. Throughout this paper, we use $\|\cdot\|$ to denote the 2-norm for a vector.

B. CVaR and VaR

Conditional Value at Risk (CVaR) is a coherent risk measure that satisfies properties of monotonicity, sub-additivity, homogeneity, and translational invariance; see [9], [21]. Formally, for a random variable Z with the cumulative distribution function (CDF) denoted by F and a risk level $\alpha \in (0, 1]$, the CVaR value is defined as $\mathrm{CVaR}_\alpha[Z] = \mathbb{E}_F[Z | Z > \mathrm{VaR}_\alpha[Z]]$, where $\mathrm{VaR}_\alpha[Z] = \inf\{y : F_Z(y) \geq 1 - \alpha\}$ is the $1 - \alpha$ quantile of the distribution of the random variable Z, also known as the Value at Risk (VaR). Intuitively, CVaR represents the average of the worst $\alpha \times 100\%$ cost. If the risk level is selected as $\alpha = 1$, it becomes equivalent to the expected (risk-neutral) case. Thus, CVaR has a naturally distributional robust optimization formulation. The work [13] introduces a new approach to compute the CVaR value, i.e.,

$$\mathrm{CVaR}_\alpha[Z] = \min_{\nu \in \mathbb{R}} \left\{ \nu + \frac{1}{\alpha} \mathbb{E}_F[Z - \nu]_+ \right\}, \tag{2}$$

where $[x]_+ = \max\{x, 0\}$. Besides, [13] shows that the right hand side of (2) takes the minimal value when $\nu = \mathrm{VaR}_\alpha[Z]$.

III. RISK-AVERSE CONVEX GAMES

In this section, we formally define the risk-averse game under consideration, using the CVaR risk measure. Moreover, we establish a condition under which this game is strongly monotone, which lays the theoretical foundation for the subsequent analysis of the Nash equilibrium convergence.

A. Problem Formulation

We consider convex games with stochastic costs $J_i(x_i, x_{-i}, \xi_i) : \mathcal{X} \times \Xi_i \to \mathbb{R}$, where x_{-i} are the actions of all agents except for agent i and $\xi_i \in \Xi_i \subset \mathbb{R}^{n_\xi}$ characterizes the uncertainty associated with the cost function, following some static and unknown distribution. We sometimes instead write $J_i(x, \xi_i)$ for ease of notation, where $x = (x_i, x_{-i})$ is the concatenated vector of all agents' actions. We assume that the diameter of the convex set $\mathcal{X}_i$ is bounded by D for all $i = 1, \dots, N$.

The goal of the risk-averse agents is to minimize the CVaR value of the stochastic cost. We denote by $F_{i,x}(y) = \mathbb{P}\{J_i(x, \xi_i) \leq y\}$ the CDF of the random cost $J_i(x, \xi_i)$ and J^{α_i} the $1 - \alpha_i$ quantile of this distribution. Then, for a given risk level $\alpha_i \in (0, 1]$, the CVaR of the cost function $J_i(x, \xi_i)$ is defined as

$$\begin{aligned} C_i(x) &:= \mathrm{CVaR}_{\alpha_i}[J_i(x, \xi_i)] \\ &:= \mathbb{E}_{F_{i,x}}[J_i(x, \xi_i) | J_i(x, \xi_i) \geq J^{\alpha_i}]. \end{aligned} \tag{3}$$

Notice that the CVaR value is determined by the distribution function $F_{i,x}(y)$ for a given α_i, so we sometimes write CVaR as a function of the distribution function, i.e., $\mathrm{CVaR}_{\alpha_i}[F_{i,x}] := \mathrm{CVaR}_{\alpha_i}[J_i(x, \xi_i)]$. Given different risk levels α_i, $i = 1, \dots, N$, the goal of each agent is to minimize the individual risk-averse loss function, i.e.,

$$\min_{x_i \in \mathcal{X}_i} C_i(x_i, x_{-i}). \tag{4}$$

In this paper, we assume that the function $J_i(x_i, x_{-i}, \xi_i)$ is convex in x_i for every $x_{-i} \in \mathcal{X}_{-i}$ and $\xi_i \in \Xi_i$, for all $i = 1, \dots, N$. By virtue of (2), we have $C_i(x) = \nu_i^* + \frac{1}{\alpha} \mathbb{E}[J_i(x, \xi_i) - \nu_i^*]_+$, where $\nu_i^* = \mathrm{VaR}_{\alpha_i}[J_i(x, \xi_i)]$. Since both the operations of point-wise supremum over convex functions and expectation preserve convexity, we can obtain that $C_i(x)$ is convex in x_i for all $i = 1, \dots, N$. As a result, the risk-averse game (4) is a convex game and has at least one equilibrium point by Theorem 1 in [20].

We consider the game setting in which the agents compete in an open environment that discloses all agents' past actions publicly. Specifically, each agent selects an action at each episode and at the next episode the previous actions will be shared to all agents. Besides, we assume that the agents can sample the static uncertainty in the game and know the form of its own cost function J_i. Thus, each agent can obtain the gradient information of J_i. The goal of this paper is to analyze the CVaR gradient and then design a risk-averse learning algorithm for the risk-averse game (4) that achieves Nash equilibrium convergence.

B. Strong Monotonicity Analysis

In this section, we provide a sufficient condition that establishes strong monotonicity for the risk-averse game (4). We note that if the risk-neutral game is monotone, where agents aim to minimize the expected cost values, the risk-averse game may not necessarily be monotone. We provide an example in the following to illustrate this point. In this example, we show that when a risk-neutral game is strongly monotone and has a unique Nash equilibrium, the risk-averse game may have infinitely many Nash equilibria.

Example. Consider a game with two agents in which each agent has the cost function $J_i(x,\xi_i) = c + ax_i^2 + ax_ix_{-i} - abx_i + \frac{4a}{3d}x_ix_{-i}\xi_i$, $i = 1, 2$, where $a > 0$ and $\xi_i \sim U(0, d)$. In the risk-neutral game, each agent aims to minimize the expected cost values $\pi_i(x) = \mathbb{E}_{\xi_i}[J_i(x,\xi_i)]$, while in the risk-averse game, each agent aims to minimize $C_i(x) = \mathrm{CVaR}_{0.5}[J_i(x,\xi_i)]$. It is easily verified that the risk-neutral game satisfies the monotone condition

$$\sum_i \langle \nabla_i \pi_i(x) - \nabla_i \pi_i(x'), x_i - x_i' \rangle \geq m_0 \left\| x - x' \right\|^2,$$

with $m_0 = a$. However, in the risk-averse case, we have $C_i(x) = c + ax_i^2 + 2ax_ix_{-i} - abx_i$ and thus $\nabla_i C_i(x) = 2ax_i + 2ax_{-i} - ab$. Since Nash equilibria are the points that satisfy $\nabla_1 C_1(x) = \nabla_2 C_2(x) = 0$, which correspond to all the points located on the line $x_1 + x_2 = \frac{b}{2}$, it follows that the risk-averse game has an infinite number of Nash equilibria.

Note that we require that the risk-averse game is strongly monotone regardless of the choices of $\alpha_i \in (0, 1]$, thus, every quantile of the distribution of the stochastic cost is expected to satisfy the strong monotonicity condition. To establish the strong monotonicity condition of the risk-averse game (4) for all the choices of α_i, we make the following assumption on the stochastic cost.

Assumption 1. *For each agent i, the cost function can be decomposed as $J_i(x,\xi_i) = f_i(x_i,\xi_i) + g_i(x)$. Moreover, $f_i(x_i,\xi_i)$ is absolutely continuous and its VaR value is differentiable for any $x_i \in \mathcal{X}_i$. Besides, $f_i(x_i,\xi_i)$ is convex in x_i, for every $\xi_i \in \Xi_i$ and $g_i(x)$ satisfies $\sum_i \langle \nabla_i g_i(x) - \nabla_i g_i(x'), x_i - x_i' \rangle \geq m \left\| x - x' \right\|^2$, for all $x, x' \in \mathcal{X}$.*

Assumption 1 states that the uncertainty in the stochastic cost of each agent does not rely on other agents' actions. This assumption holds for many classes of games including Cournot games. For example, in a market consisting of multiple competitive companies that produce goods at different levels, the price of these goods consists of the deterministic term $P(x)$ that depends on the total amount of production and the random term ξ_i that captures the randomness in the market induced by, e.g., government policies or the climate changes. This randomness usually does not depend on the production levels. As a result, the reward of the i-th company can be given by $r_i(x) = -J_i(x) = (P(x) + \xi_i)x_i$.

Given Assumption 1, we show that the risk-averse game is strongly monotone in the following lemma.

Lemma 1. *Let Assumption 1 hold. Then, the risk-averse game* (4) *is m-strongly monotone.*

Proof. Given Assumption 1 and the translation invariance property of CVaR, we have that $C_i(x) = \mathrm{CVaR}_{\alpha_i}[J_i(x,\xi_i)] = \mathrm{CVaR}_{\alpha_i}[f_i(x_i,\xi_i)] + g_i(x)$. Define $\mathrm{CVaR}_{\alpha_i}[f_i(x_i,\xi_i)] = h_i(x_i)$. Since f_i is convex in x_i, by Lemma 3 in [7], $h_i(x_i)$ is convex in x_i. Then, it holds that

$$\begin{aligned}
&\sum_i \langle \nabla_i C_i(x) - \nabla_i C_i(x'), x_i - x_i' \rangle \\
=&\sum_i \langle \nabla_i h_i(x_i) - \nabla_i h_i(x_i'), x_i - x_i' \rangle \\
&+ \sum_i \langle \nabla_i g_i(x) - \nabla_i g_i(x'), x_i - x_i' \rangle \\
\geq&\sum_i \langle \nabla_i g_i(x) - \nabla_i g_i(x'), x_i - x_i' \rangle \geq m \left\| x - x' \right\|^2,
\end{aligned}$$

for all $x, x' \in \mathcal{X}$ and $\alpha_i \in (0, 1]$, $i = 1, \ldots, N$. The first inequality follows from the convexity of h_i. The proof is complete. □

Given that the risk-averse game is strongly monotone, the risk-avesrse Nash equilibrium x^* is unique [20]. In what follows, we require the following assumption on the cost function J_i.

Assumption 2. $\|\nabla_i J_i(x,\xi_i)\| \leq B$, *for all* $i = 1, \ldots, N$.

This assumption is common in the literature and holds in many applications, e.g., Cournot Games and Kelly auctions; see [1], [4].

IV. A Risk-Averse Learning Algorithm

In this section, we propose a method that enables each agent to use the gradient of the stochastic cost to minimize the risk-averse cost function.

A. Algorithm Description

Before presenting the designed algorithm, we analyze the expression of the CVaR gradient. To do so, we define the auxiliary function

$$L_i(x,\nu_i) := \nu_i + \frac{1}{\alpha_i}\mathbb{E}_{\xi_i}[J_i(x,\xi_i) - \nu_i]_+,$$

where $\nu_i \in \mathbb{R}$ is an auxiliary variable. We assume that $J_i(\cdot,\xi_i)$ is Lipschitz and diffentiable for every ξ_i. Then, it is shown in [8] that the (sub)gradient of L_i can be represented as

$$\nabla_i L_i(x,\nu_i) = \mathbb{E}_{\xi_i}\left[\begin{array}{c} \frac{1}{\alpha_i}\mathbf{1}\left\{J_i(x,\xi_i) \geq \nu_i\right\}\nabla_i J_i(x,\xi_i) \\ 1 - \frac{1}{\alpha_i}\mathbf{1}\left\{J_i(x,\xi_i) \geq \nu_i\right\} \end{array}\right],$$

where $\nabla_i J_i(x,\xi_i)$ represents the derivative of the function $J_i(x,\xi_i)$ with respect to x_i, $\nabla_i L_i(x,\nu_i)$ denotes the derivative of the function L_i with respect to (x_i,ν_i), and $\mathbf{1}\{\cdot\}$ is the indicator function. As shown in [13], the VaR value of the random cost $J_i(x,\xi_i)$, which we denote by $\nu_i^*(x)$, satisfies $C_i(x) = \mathrm{CVaR}_{\alpha_i}[J_i(x,\xi_i)] = L_i(x,\nu_i^*(x))$ and $\nu_i^*(x) = $ left endpoint of $\mathcal{A}_i^*(x)$, where the set $\mathcal{A}_i^*(x) := \mathrm{argmin}_{\nu_i} L_i(x,\nu_i)$.

Algorithm 1: First-order risk-averse learning

Require: Initial value x_0, step size η, the total number of episodes T, risk level α_i, $i = 1, \cdots, N$.

for *episode* $t = 1, \ldots, T$ **do**
  Each agent plays $x_{i,t}$, $i = 1, \ldots, N$
  for agent $i = 1, \ldots, N$ **do**
    Each agent samples ξ_i^t
    Each agent computes cost evaluations $J_i(x_t, \xi_i^k)$ and $\nabla_i J_i(x_t, \xi_i^k)$, $k = 1, \ldots, t$
    Build EDF by (5)
    Compute VaR estimate by (6)
    Construct gradient estimate by (7)
    Update action $x_{i,t+1}$ by (8)
  end for
end for

In the following lemma, we connect the CVaR gradient with the gradient of the function $L_i(x, \nu_i)$.

Lemma 2. *It holds that*

$$\begin{aligned}\nabla_i C_i(x) =& \nabla_{x_i} L_i(x, \nu_i)\Big|_{\nu_i = \nu_i^*(x)} \\ =& \mathbb{E}_{\xi_i}\left[\frac{1}{\alpha_i}\boldsymbol{1}\left\{J_i(x, \xi_i) \geq \nu_i^*(x)\right\}\nabla_i J_i(x, \xi_i)\right].\end{aligned}$$

Proof. Since $L_i(x, \nu_i)$ is convex in ν_i, $\nu_i^*(x)$ satisfies that $\nabla_{\nu_i} L_i(x, \nu_i^*(x))(\nu_i' - \nu_i^*(x)) \geq 0$, for all $\nu_i' \in \mathbb{R}$. Due to the fact that the value of $\nu_i^*(x)$ is bounded, we conclude that $\nabla_{\nu_i} L_i(x, \nu_i)\Big|_{\nu_i = \nu_i^*(x)} = 1 - \frac{1}{\alpha_i}\mathbf{1}\left\{J_i(x, \xi_i) \geq \nu_i^*(x)\right\} = 0$. Hence, we have

$$\begin{aligned}&\nabla_i C_i(x) = \nabla_{x_i} L_i(x, \nu_i^*(x)) \\ =& \nabla_{x_i} L_i(x, \nu_i)\Big|_{\nu_i = \nu_i^*(x)} + \nabla_{\nu_i} L_i(x, \nu_i)\nabla_{x_i}\nu_i^*(x)\Big|_{\nu_i = \nu_i^*(x)} \\ =& \nabla_{x_i} L_i(x, \nu_i)\Big|_{\nu_i = \nu_i^*(x)},\end{aligned}$$

which completes the proof. □

Lemma 2 provides an alternative way to compute the CVaR gradient by means of computing the gradient $\nabla_{x_i} L_i(x, \nu_i)$ and the VaR value $\nu_i^*(x)$. Similar arguments can also be found in [22]. Leveraging this result, we propose a first-order risk-averse learning algorithm to solve the problem (4), which is illustrated in Algorithm 1.

Specifically, at episode t, each agent plays the action $x_{i,t}$ and samples the random seed ξ_i^t. Then, each agent collects all the previous samples ξ_i^k, $k = 1, \ldots, t$, and computes the cost evaluations $J_i(x_t, \xi_i^k)$ and gradients $\nabla_i J_i(x_t, \xi_i^k)$, for all $k = 1, \ldots, t$. For agent i, we denote the CDF of the random cost $J_i(x_t, \xi_i)$ as $G_{i,t}(y) = \mathbb{P}\{J_i(x_t, \xi_i) \leq y\}$. With finitely many samples, the agents cannot obtain the accurate CDF but only construct the EDF $\hat{G}_{i,t}$ by

$$\hat{G}_{i,t}(y) = \frac{1}{t}\sum_{k=1}^{t}\mathbf{1}\{J_i(x_t, \xi_i^k) \leq y\}. \tag{5}$$

This distribution estimate is constructed by using all historical samples and thus each agent has t samples at episode t. Using the distribution estimate $\hat{G}_{i,t}$, each agent constructs the VaR estimate as

$$\nu_{i,t} = \mathrm{VaR}_{\alpha_i}[\hat{G}_{i,t}]. \tag{6}$$

Using the VaR estimate $\nu_{i,t}$, we design the CVaR gradient estimate as

$$g_{i,t} = \frac{1}{t\alpha_i}\sum_{k=1}^{t}\mathbf{1}\left\{J_i(x_t, \xi_i^k) \geq \nu_{i,t}\right\}\nabla_i J_i(x_t, \xi_i^k). \tag{7}$$

Then, each agent performs the following projected gradient-descent update

$$x_{i,t+1} \leftarrow \mathcal{P}_{\mathcal{X}_i}(x_{i,t} - \eta g_{i,t}). \tag{8}$$

B. Convergence Analysis

In this section, we provide the convergence analysis for Algorithm 1.

From (7), we have $\mathbb{E}[g_{i,t}] = \nabla_{x_i} L_i(x_t, \nu_{i,t})$, which indicates that the gradient estimate (7) is an unbiased estimate of $\nabla_{x_i} L_i(x_t, \nu_{i,t})$, but a biased estimate of $\nabla_i C_i(x_t)$. Hence, there exists a gradient estimation bias, which we define as

$$\varepsilon_{i,t} := \nabla_{x_i} L_i(x_t, \nu_{i,t}) - \nabla_i C_i(x_t).$$

We denote by $\nu_{i,t}^*$ the true VaR value of the random cost $J_i(x_t, \xi_i)$, i.e., $\nu_{i,t}^* = \mathrm{VaR}_{\alpha_i}[J_i(x_t, \xi_i)]$. By virtue of Lemma 2, we have $\varepsilon_{i,t} = \nabla_{x_i} L_i(x_t, \nu_{i,t}) - \nabla_{x_i} L_i(x_t, \nu_{i,t}^*)$, which indicates that the performance of CVaR gradient estimate is closely related to the VaR estimate error $|\nu_{i,t} - \nu_{i,t}^*|$.

It has been shown in [23] that the quantile estimation error using finitely many samples is inversely proportional to the probability density value at the VaR point. Therefore, we make the following assumption on the PDF of the random cost.

Assumption 3. *Let* $F_{i,x}(y) = \mathbb{P}\{J_i(x, \xi_i) \leq y\}$ *and* $\mathcal{Y}_{i,x} = \mathrm{Range}(J_i(x, \xi_i))$. *For every* $x \in \mathcal{X}$, *the distribution function* $F_{i,x}$ *is continuously differentiable and* L_0*-Lipschitz continuous. Moreover, there exists a lower bound* $\underline{p} > 0$ *on its probability density function, i.e.,* $F_{i,x}'(y) \geq \underline{p}$ *for all* $y \in \mathcal{Y}_{i,x}$.

Assumption 3 states that the PDF of the random cost $J_i(x, \xi_i)$ is both upper bounded and lower bounded. Note that this assumption is satisfied for some common random variables, e.g., uniform and weighted truncated random variables. In fact, we can construct the confidence bound for VaR estimates only when the value of the PDF at the VaR point is lower bounded, see [11], [12] for further discussions.

In what follows, we present a lemma that bounds the VaR estimation error.

Lemma 3. *Let Assumption 3 hold. Suppose that we have* t *samples at episode* t. *Then, we have* $\mathbb{P}\{|\nu_{i,t} - \nu_{i,t}^*| > \epsilon\} \leq 2e^{-2t\epsilon^2\underline{p}^2}$.

Proof. The proof is motivated by [24]. Let $\epsilon > 0$ be fixed. Note that, for any CDF $G(y)$, $G(t) \geq t$ if and only if $y \geq$

$G^{-1}(t)$. Hence, we have

$$\begin{aligned}
&\mathbb{P}\left\{\nu_{i,t} > \nu^*_{i,t} + \epsilon\right\} = \mathbb{P}\left\{\hat{G}_{i,t}(\nu_{i,t}) > \hat{G}_{i,t}(\nu^*_{i,t} + \epsilon)\right\} \\
=&\mathbb{P}\left\{G_{i,t}(\nu^*_{i,t} + \epsilon) - \hat{G}_{i,t}(\nu^*_{i,t} + \epsilon) > G_{i,t}(\nu^*_{i,t} + \epsilon) - \alpha_i\right\} \\
\leq&\mathbb{P}\left\{\sup_y |\hat{G}_{i,t}(y) - G_{i,t}(y)| > G_{i,t}(\nu^*_{i,t} + \epsilon) - \alpha_i\right\} \\
\leq&\mathbb{P}\Big\{\sup_y |\hat{G}_{i,t}(y) - G_{i,t}(y)| > G_{i,t}(\nu^*_{i,t} + \epsilon) - G_{i,t}(\nu^*_{i,t})\Big\},
\end{aligned}$$

where the second and last equalities follow from the facts $\hat{G}_{i,t}(\nu_{i,t}) = \alpha_i$ and $G_{i,t}(\nu^*_{i,t}) = \alpha_i$, respectively. By virtue of the Mean Value Theorem, there exists $\nu_1 \in (\nu^*_{i,t}, \nu^*_{i,t} + \epsilon)$ such that $G_{i,t}(\nu^*_{i,t} + \epsilon) - G_{i,t}(\nu^*_{i,t}) = G'_{i,t}(\nu_1)\epsilon$. Using Assumption 3, we have that

$$\begin{aligned}
&\mathbb{P}\left\{\nu_{i,t} > \nu^*_{i,t} + \epsilon\right\} \\
\leq&\mathbb{P}\left\{\sup_y |\hat{G}_{i,t}(y) - G_{i,t}(y)| > G'_{i,t}(\nu_1)\epsilon\right\} \\
\leq&\mathbb{P}\left\{\sup_y |\hat{G}_{i,t}(y) - G_{i,t}(y)| > \underline{p}\epsilon\right\}.
\end{aligned}$$

Using similar arguments yields the other side of the inequality. Hence, we have

$$\mathbb{P}\left\{|\nu_{i,t} - \nu^*_{i,t}| > \epsilon\right\} \leq \mathbb{P}\left\{\sup_y |\hat{G}_{i,t}(y) - G_{i,t}(y)| > \underline{p}\epsilon\right\}. \tag{9}$$

By virtue of the definition of $\hat{G}_{i,t}$ in (5), this EDF is constructed using t samples. Using the DvoretzkyKieferWolfowitz (DKW) inequality, we have

$$\mathbb{P}\left\{\sup_y |\hat{G}_{i,t}(y) - G_{i,t}(y)| > \underline{p}\epsilon\right\} \leq 2e^{-2t\underline{p}^2\epsilon^2}. \tag{10}$$

Substituting (10) into (9) completes the proof. □

Based on Lemma 3, the accumulated error of the CVaR gradient estimate is bounded, which is presented in the following lemma.

Lemma 4. *Given a confidence level $\bar{\gamma}$, we have that $\sum_{t=1}^T \|\varepsilon_{i,t}\| \leq \frac{\sqrt{2}BL_0}{\alpha_i \underline{p}} \sqrt{\ln \frac{2}{\bar{\gamma}}} \sqrt{T}$ with probability at least $1-\gamma$, where $\gamma = \bar{\gamma}T$.*

Proof. See Appendix A. □

For ease of notation, we define $S_1(\alpha) := \sum_i \frac{1}{\alpha_i}$ and $S_2(\alpha) := \sum_i \frac{1}{\alpha_i^2}$. Now we are ready to present the convergence result.

Theorem 1. *Let Assumptions 1, 2 and 3 hold, and select $\eta = \frac{D}{B}T^{-\frac{1}{2}}$. Then, Algorithm 1 achieves*

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T \mathbb{E}\|x_t - x^*\|^2 \\
=&\mathcal{O}\Big(T^{-\frac{1}{2}}\big(S_2(\alpha) + \sqrt{\ln(T/\gamma)}S_1(\alpha)\big)\Big),
\end{aligned} \tag{11}$$

with probability at least $1-\gamma$.

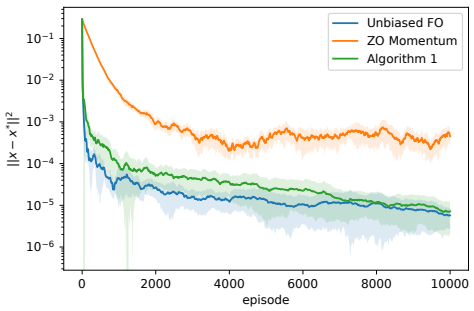

Fig. 1. Error to the Nash equilibrium among Algorithm 1, unbiased first-order algorithm (Unbiased FO), and the zeroth-order algorithm with momentum (ZO momentum) in [18]. The solid lines and shades are averages and standard deviations over 20 runs.

The proof can be found in the arxiv version. Theorem 1 shows that Algorithm 1 achieves a time-averaged convergence to the risk-averse Nash equilibrium with high probability. If the risk-averse game is convex and not necessarily strongly monotone, we can use similar techniques and show that Algorithm 1 achieves sub-linear regret.

V. Numerical Result

In this section, we illustrate the proposed algorithm on a Cournot game. Specifically, we consider two risk-averse agents $i = 1, 2$. Each agent determines the production level x_i of a homogeneous product and has an individual cost function $J_i(x) = 1 - (2 - \sum_j x_j)x_i + 0.2x_i + \xi_i x_i$, where $\xi_i \sim U(0,1)$ is a uniform random variable. Here we utilize the term $\xi_i x_i$ to represent the uncertainty occurred in the market, which is proportional to the production level x_i. It is easy to verify that the game satisfies Assumption 1 and is strongly monotone. The agents have their own risk levels α_i and they aim to minimize the CVaR value of their cost functions. We select $\alpha_1 = 0.4$ and $\alpha_2 = 0.8$. It can be verified that the Nash equilibrium point of the formulated risk-averse game is $x^* = (0.2667, 0.4667)$.

To compute the distribution function in Algorithm 1 in practice, we partition the interval $[0, U]$ into 1000 equal-width bins and approximate the expectation by the sum of finite terms. We compare Algorithm 1 with the zeroth-order algorithm in [19], which we term the zeroth-order algorithm with momentum. To explore the effect of bias in the estimates of the VaR values, we run another first-order algorithm with accurate VaR values at each iteration, which we term the unbiased first-order algorithm. Each algorithm is run for 20 trials and the parameters of these algorithms are separately optimally tuned. The convergence results are presented in Figure 1. We observe that our first-order method, i.e., Algorithm 1, outperforms the zeroth-order algorithm. Since Algorithm 1 cannot entirely eliminate the bias in the VaR value estimates, it performs worse than the unbiased first-order algorithm. However, as the number of samples increases and the VaR estimate error diminishes, the performance of these two algorithms becomes close.

VI. Conclusion

In this work, we proposed a first-order method to solve convex games with risk-averse agents. We showed that the VaR estimates play a key role in estimating the CVaR gradient. Assuming that the game is strongly monotone, we showed that the VaR estimates can be bounded and, as a result, our proposed algorithm converges to the Nash equilibrium with high probability. We provided numerical simulations to illustrate the performance of our algorithm.

Appendix

A. Proof of Lemma 4

By applying Lemma 3 and setting $\bar{\gamma} = 2e^{-2t\epsilon^2\underline{p}^2}$ with $\bar{\gamma} = \gamma/T$, we have

$$\mathbb{P}\left\{|\nu_{i,t} - \nu^*_{i,t}| > \frac{1}{\underline{p}\sqrt{2t}}\sqrt{\ln\frac{2C}{\bar{\gamma}}}\right\} \leq \bar{\gamma}. \tag{12}$$

Define the events in (12) as $\mathcal{B}_t$. Then, for all $t = 1, \dots, T$, we have $|\nu_{i,t} - \nu^*_{i,t}| \leq \frac{1}{\underline{p}\sqrt{2t}}\sqrt{\ln\frac{2}{\bar{\gamma}}},\ \forall t = 1, \dots, T$, with probability at least $1-\gamma$, since $1 - \mathbb{P}\{\bigcup_{t=1}^T \mathcal{B}_t\} \geq 1 - \sum_{t=1}^T \mathbb{P}\{\mathcal{B}_t\} \geq 1 - T\frac{\gamma}{T} \geq 1-\gamma$.

Next, we analyze the property of $\varepsilon_{i,t}$. Set $\nu_m = \min\{\nu_{i,t}, \nu^*_{i,t}\}$, $\nu_M = \max\{\nu_{i,t}, \nu^*_{i,t}\}$. We have

$$\begin{aligned}
&\varepsilon_{i,t} = \nabla_{x_i} L_i(x_t, \nu_{i,t}) - \nabla_i C_i(x_t)\\
=&\mathbb{E}\Big[\frac{1}{\alpha_i}\mathbf{1}\left\{J_i(x_t,\xi_i) \geq \nu_{i,t})\right\}\nabla_i J_i(x_t,\xi_i)\Big]\\
&-\mathbb{E}\Big[\frac{1}{\alpha_i}\mathbf{1}\left\{J_i(x_t,\xi_i) \geq \nu^*_{i,t})\right\}\nabla_i J_i(x_t,\xi_i)\Big]\\
=&\mathbb{E}\Big[\frac{1}{\alpha_i}\mathrm{sgn}\{\nu_{i,t}-\nu^*_{i,t}\}\mathbf{1}\left\{\nu_m \leq J_i(x,\xi_i) \leq \nu_M\right\}\nabla_i J_i(x_t,\xi_i)\Big].
\end{aligned}$$

Using the fact that $\|\mathbb{E}[XY]\| \leq \mathbb{E}[\|X\|\,\|Y\|]$ for any random variable X, Y, for all $t \geq 1$, we have

$$\begin{aligned}
\|\varepsilon_{i,t}\| \leq& \frac{1}{\alpha_i}\mathbb{E}[\mathbf{1}\left\{\nu_m \leq J_i(x,\xi_i) \leq \nu_M)\right\}]B\\
=&\frac{B}{\alpha_i}(G_{i,t}(\nu_M) - G_{i,t}(\nu_m))\\
\leq&\frac{BL_0}{\alpha_i}|\nu_{i,t} - \nu^*_{i,t}| \leq \frac{BL_0}{\alpha_i\underline{p}\sqrt{2t}}\sqrt{\ln\frac{2T}{\gamma}},
\end{aligned} \tag{13}$$

with probability at least $1-\gamma$. Summing up the inequality (13) over $t = 1, \dots, T$, we have

$$\begin{aligned}
&\sum_{t=1}^T \|\varepsilon_{i.t}\| \leq \sum_{t=1}^T \frac{BL_0}{\alpha_i\underline{p}\sqrt{2t}}\sqrt{\ln\frac{2T}{\gamma}}\\
\leq&\frac{BL_0}{\alpha_i\underline{p}\sqrt{2}}\sqrt{\ln\frac{2T}{\gamma}}\Big(1 + \int_1^T \frac{1}{\sqrt{t}}dt\Big)\\
\leq&\frac{\sqrt{2}BL_0}{\alpha_i\underline{p}}\sqrt{\ln\frac{2T}{\gamma}}\sqrt{T},
\end{aligned}$$

which completes the proof.

References

[1] Tianyi Lin, Zhengyuan Zhou, Wenjia Ba, and Jiawei Zhang. Doubly optimal no-regret online learning in strongly monotone games with bandit feedback. *arXiv preprint arXiv:2112.02856*, 2021.

[2] Pier Giuseppe Sessa, Ilija Bogunovic, Maryam Kamgarpour, and Andreas Krause. No-regret learning in unknown games with correlated payoffs. In *Advances in Neural Information Processing Systems*, volume 32, pages 13624–13633, 2019.

[3] Panayotis Mertikopoulos and Zhengyuan Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1):465–507, 2019.

[4] Mario Bravo, David Leslie, and Panayotis Mertikopoulos. Bandit learning in concave n-person games. In *Advances in Neural Information Processing Systems*, pages 5666–5676, 2018.

[5] Mohamadreza Ahmadi, Xiaobin Xiong, and Aaron D Ames. Risk-averse control via cvar barrier functions: Application to bipedal robot locomotion. *IEEE Control Systems Letters*, 6:878–883, 2021.

[6] Harry M Markowitz. Foundations of portfolio theory. *The Journal of Finance*, 46(2):469–477, 1991.

[7] Adrian Rivera Cardoso and Huan Xu. Risk-averse stochastic convex bandit. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 39–47. PMLR, 2019.

[8] Dionysios S Kalogerias. Fast and stable convergence of online sgd for cv@ r-based risk-aware learning. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6007–6011. IEEE, 2022.

[9] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczynski. *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.

[10] Balazs Szorenyi, Róbert Busa-Fekete, Paul Weng, and Eyke Hüllermeier. Qualitative multi-armed bandits: A quantile-based approach. In *International Conference on Machine Learning*, pages 1660–1668, 2015.

[11] Mengyan Zhang and Cheng Soon Ong. Quantile bandits for best arms identification. In *International Conference on Machine Learning*, pages 12513–12523. PMLR, 2021.

[12] Ravi Kumar Kolla, LA Prashanth, Sanjay P Bhat, and Krishna Jagannathan. Concentration bounds for empirical conditional value-at-risk: The unbounded case. *Operations Research Letters*, 47(1):16–20, 2019.

[13] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[14] Alex Tamkin, Ramtin Keramati, Christoph Dann, and Emma Brunskill. Distributionally-aware exploration for cvar bandits. In *NeurIPS 2019 Workshop on Safety and Robustness on Decision Making*, 2019.

[15] Lei You, Hui Ma, Tapan Kumar Saha, and Gang Liu. Gaussian mixture model based distributionally robust optimal power flow with cvar constraints. *arXiv preprint arXiv:2110.13336*, 2021.

[16] Ali Yekkehkhany, Timothy Murray, and Rakesh Nagi. Risk-averse equilibrium for games. *arXiv preprint arXiv:2002.08414*, 2020.

[17] Oliver Slumbers, David Henry Mguni, Stephen McAleer, Jun Wang, and Yaodong Yang. Learning risk-averse equilibria in multi-agent systems. *arXiv preprint arXiv:2205.15434*, 2022.

[18] Zifan Wang, Yi Shen, and Michael Zavlanos. Risk-averse no-regret learning in online convex games. In *International Conference on Machine Learning*, pages 22999–23017. PMLR, 2022.

[19] Zifan Wang, Yi Shen, Zachary I Bell, Scott Nivison, Michael M Zavlanos, and Karl H Johansson. A zeroth-order momentum method for risk-averse online convex games. In *2022 IEEE 61st Conference on Decision and Control*, pages 5179–5184. IEEE, 2022.

[20] J Ben Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, pages 520–534, 1965.

[21] Audrey Huang, Liu Leqi, Zachary Lipton, and Kamyar Azizzadenesheli. Off-policy risk assessment in contextual bandits. In *Advances in Neural Information Processing Systems*, pages 23714–23726, 2021.

[22] L Jeff Hong and Guangwu Liu. Simulating sensitivities of conditional value at risk. *Management Science*, 55(2):281–293, 2009.

[23] R Raj Bahadur. A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3):577–580, 1966.

[24] Jun Shao. *Mathematical statistics*. Springer Science & Business Media, 2003.