

Asymmetric Learning in Convex Games

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

Abstract—This paper considers convex games involving multiple agents that aim to minimize their own cost functions using locally available information. A common assumption in the study of such games is that the agents are symmetric, meaning that they have access to the same type of information. Here we lift this assumption, which is often violated in practice, and instead consider asymmetric agents; specifically, we assume some agents have access to first-order gradient information and others have access to the zeroth-order oracles (cost function evaluations). We propose an asymmetric learning algorithm that combines the agent information mechanisms. We analyze the regret and Nash equilibrium convergence of this algorithm for convex and strongly monotone games, respectively. Specifically, we show that our algorithm always performs between pure first- and zeroth-order methods, and can match the performance of these two extremes by adjusting the number of agents with access to zeroth-order oracles. Therefore, our algorithm incorporates the pure first- and zeroth-order methods as special cases. We provide numerical experiments on a market problem for both deterministic and risk-averse games to demonstrate the performance of the proposed algorithm.

Index Terms—Asymmetric learning, Nash equilibrium, convex games, regret analysis

I. INTRODUCTION

Convex optimization [1]–[4] is widely applicable in various fields such as economics, engineering, and machine learning. Recently, convex optimization has been employed in multi-agent games with applications in traffic routing [5] and market optimization [6]. In these applications, agents are usually assumed rational with the goal to minimize their own cost functions by leveraging limited information received from the environment, which falls into the category of convex games [7], [8].

The performance of optimization algorithms for convex games is typically evaluated using the notion of regret [3], which captures the difference between agents' sequential actions and the achievable best actions in hindsight. An algorithm is said to achieve no-regret learning if the regret of a sequence of actions generated by the algorithm is sub-linear in the total number of episodes T , meaning that agents are able to

* 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 SUCCESS Grant, AFOSR under award #FA9550-19-1-0169. (Corresponding author: Xinlei Yi.)

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

Xinlei Yi is with Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai, 201804, China, and is also with Lab for Information & Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Email: xinleiyi@tongji.edu.cn.

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

eventually learn the best actions. Another important measure of performance in convex games is that of a Nash equilibrium, which is defined as a point at which no agent has incentive to change its decision. Recently, there is a growing literature in game theory focusing on designing algorithms that achieve no-regret learning [5], [9]–[11] or Nash equilibrium convergence [8], [12]–[19]. Common in these works is the assumption that the agents have access to similar or symmetric type of information. However, in many real-world settings, agents are asymmetric, meaning that they have access to different types or amounts of information; for instance, in financial markets, investment banks have access to more information compared to individual investors. Moreover, in Cournot competition, where multiple companies aim to maximize their own profit, certain dominant companies have the capability to observe their competitors' strategies while others do not have this advantage. Information asymmetry is also evident in security systems [20], [21], where an attacker operates in obscurity and has the capability to observe the actions of the defender. Conversely, the defender operates with transparency, often without access to the attacker's information or actions. In these cases, the asymmetric information gives rise to asymmetric update strategies for the agents.

To the best of our knowledge, asymmetric learning in convex games has not been explored in the literature. Most closely related to our study is symmetric learning in convex games [8], [22]–[33], where authors have proposed methods for no-regret learning and/or Nash equilibrium convergence. Specifically, when gradient information is available, [22] develops optimistic gradient methods for continuous games with noisy gradient estimates that achieve constant regret under multiplicative noise. Similarly, [23] proposes a first-order gradient descent algorithm for λ -coercive games with unconstrained continuous action sets, which attains the last-iterate convergence to a Nash equilibrium. Assuming bandit feedback, i.e., agents only have access to zeroth-order oracles, [8] shows Nash equilibrium convergence for strongly monotone games, which is a special class of convex games. The convergence rate of the zeroth-order method in [8] is further improved in [30] relying on the additional assumption that the Jacobian of the gradient function is Lipschitz continuous. Common in these works is that the agents perform symmetric updates using the same kind of information. The methods cannot be directly extended to asymmetric agents.

There are a few works that analyze asymmetric learning in other classes of games with different types of asymmetry [21], [34]–[37]. For example, [34] considers information asymmetry in Stackelberg games, where one agent (leader) can observe the action of the other agent (follower), and proposes a learning method that utilizes the theory of Markov games. Subsequently, [35] proposes an asymmetric Q-learning algorithm

for two-agent Markov games and discusses the existence of Nash equilibria and convergence. In [36], Bayesian Stackelberg games are analyzed under double-sided information asymmetry where the leader hides its action from the follower and the follower holds information about its payoff private. It is shown that the leader can improve its payoff by strategically revealing part of the action to the follower. We note that all the works above focus on Stackelberg games or Markov games with two agents acting *consecutively*, which differ from the convex games considered by us.

In this paper, we consider asymmetry in the form of gradient available to the agents during the learning process. Specifically, we assume that some agents have access to zeroth-order oracles, while other agents additionally have access to first-order gradient information. This situation may arise when some agents can observe other agents' actions and thus can compute the first-order gradient while the remaining agents have only access to cost function evaluations at each episode. Asymmetry also arises in stochastic games involving both risk-neutral and risk-averse agents. When Conditional Value at Risk (CVaR) is used as a risk measure, the CVaR gradient can rarely be explicitly derived even if the form of the stochastic cost function is known [38]. The risk-averse agents cannot easily have access to gradient information unlike the risk-neutral agents.

Our main contributions are summarized as follows.

- 1) We develop a novel framework that relies on asymmetric information to learn optimal actions in convex games. In this framework, agents belong to two distinct groups: agents that only have access to zeroth-order oracles and update their actions using zeroth-order optimization techniques, and agents that have access to first-order gradient information and update their actions accordingly.
- 2) While the asymmetric setting complicates the system dynamics, we theoretically show that no-regret learning is achieved for every agent for convex games, and last-iterate Nash equilibrium convergence is guaranteed for strongly monotone games. We show that the performance of the proposed asymmetric learning algorithm lies between the pure first- and zeroth-order methods.
- 3) Experimentally, we validate our algorithm on online markets, specifically, a deterministic and a risk-averse Cournot game. In the latter case, the cost of each agent is stochastic and agents may be risk-neutral or risk-averse to avoid catastrophically high costs. We show that when there are risk-neutral agents, our asymmetric learning algorithm converges faster compared to the pure zeroth-order method in [6].

To the best of our knowledge, this is the first paper to address asymmetric agent updates in convex games with theoretical results on regret analysis and Nash equilibrium convergence. Perhaps closest to the method proposed here is the one in [39], which addresses a distributed consensus optimization problem where the agents perform either Newton- or gradient-type updates. Linear convergence is shown for strongly convex objective functions regardless of updates. We note that the games considered here are different than the

consensus optimization problems in [39], as is also the type of asymmetric information. As a result, the techniques in [39] cannot be applied to analyze the asymmetric games considered here.

The rest of this paper is organized as follows. In Section II, we formally define asymmetric games and provide some basic notation. The asymmetric learning algorithm is proposed in Section III. In Section IV, we analyze regret of the algorithm for convex games. Section V provides Nash equilibrium convergence analysis for strongly monotone games of the asymmetric algorithm. Section VI experimentally illustrates the algorithm in the application to deterministic and risk-averse Cournot games. Finally, we conclude the paper in Section VII. The detailed proofs of the results presented in this paper can be found in the arXiv paper [40].

II. PROBLEM DEFINITION

Consider a repeated game with N non-cooperative agents, whose goals are to learn the best actions that minimize their own cost functions. Let $\mathcal{N} = \{1, 2, \dots, N\}$ denote the set of agent indices. For each agent $i \in \mathcal{N}$, the cost function is $\mathcal{C}_i(x_i, x_{-i}) : \mathcal{X} \rightarrow \mathbb{R}$, where $x_i \in \mathcal{X}_i$ is the action of agent i and x_{-i} the actions of all agents except for agent i . The joint action space is defined as $\mathcal{X} = \prod_{i=1}^N \mathcal{X}_i$, where $\mathcal{X}_i \subset \mathbb{R}^d$, $d > 0$, is a convex set. We write $x := (x_1, \dots, x_N)$ as the collection of all agents' actions. Throughout this paper, we use the notation (x, y) to denote the concatenated vector $[x^T, y^T]^T$.

The goal of each agent i is to minimize its individual cost function, i.e.,

$$\underset{x_i \in \mathcal{X}_i}{\text{minimize}} \mathcal{C}_i(x_i, x_{-i}). \quad (1)$$

The game (1) is defined as a convex game when each agent's cost function is convex in its individual action. We consider the following class of convex games.

Assumption 1. For each agent $i \in \mathcal{N}$, $\mathcal{C}_i(x_i, x_{-i})$ is convex in x_i for all $x_{-i} \in \mathcal{X}_{-i}$, where $\mathcal{X}_{-i} = \prod_{j=1, j \neq i}^N \mathcal{X}_j$. Moreover, we assume that \mathcal{X}_i contains the ball with radius r centered at the origin and has a bounded diameter $D > 0$, for all $i \in \mathcal{N}$.

Assumption 1 is common in the literature, see, e.g., [30]. As shown in [41], there always exists at least one Nash equilibrium in a convex game (1). We denote by x^* such a Nash equilibrium. By definition, we have that

$$\mathcal{C}_i(x^*) \leq \mathcal{C}_i(x_i, x_{-i}^*), \text{ for all } x_i \in \mathcal{X}_i, i \in \mathcal{N}.$$

At a Nash equilibrium point, no agent can reduce its own cost by unilaterally changing its individual action. Since every agent's cost function is convex in its own action, the Nash equilibrium can be characterized using the first-order optimality condition

$$\langle \nabla_i \mathcal{C}_i(x^*), x_i - x_i^* \rangle \geq 0, \text{ for all } x_i \in \mathcal{X}_i, i \in \mathcal{N},$$

where we write $\nabla_i \mathcal{C}_i(x)$ instead of $\nabla_{x_i} \mathcal{C}_i(x)$ for the ease of notation and the symbol ∇_{x_i} means taking the gradient with respect to x_i .

We consider the case when agents have access to asymmetric information. Specifically, there are $N_z \geq 0$ agents that only have access to zeroth-order oracles, and $N - N_z$ agents that have access to first-order gradient information. We index the agents that only have access to zeroth-order oracles by i_z , where $i_z \in \mathcal{N}_z := \{1, \dots, N_z\}$, and the agents that have access to first-order gradient information by i_f , where $i_f \in \mathcal{N}_f := \{N_z + 1, \dots, N\}$. When $N_z = 0$, we denote $\mathcal{N}_z = \emptyset$, and when $N_z = N$, $\mathcal{N}_f = \emptyset$. Let the vectors $x_z := (x_1, \dots, x_{N_z})$ and $x_f := (x_{N_z+1}, \dots, x_N)$ represent the action profiles of agents that have access to zeroth-order oracles and first-order gradient information, respectively. Finally, we use the notation $-i_z$ to represent all agents that have access to zeroth-order oracles except for the agent i_z .

We make the following assumptions on the game (1).

Assumption 2. For each agent $i \in \mathcal{N}$ and $x \in \mathcal{X}$, there exists $U > 0$ such that $|\mathcal{C}_i(x)| \leq U$.

Assumption 3. For each agent $i \in \mathcal{N}$, there exists $L_0 > 0$ such that $\mathcal{C}_i(x)$ is L_0 -Lipschitz continuous in x , i.e.,

$$|\mathcal{C}_i(x) - \mathcal{C}_i(y)| \leq L_0 \|x - y\|, \quad \forall x, y \in \mathcal{X}.$$

Assumption 4. For each agent $i \in \mathcal{N}$, there exists $L_1 > 0$ such that $\nabla_i \mathcal{C}_i(x)$ is L_1 -Lipschitz continuous in x , i.e.,

$$\|\nabla_i \mathcal{C}_i(x) - \nabla_i \mathcal{C}_i(y)\| \leq L_1 \|x - y\|, \quad \forall x, y \in \mathcal{X}.$$

These assumptions are common in the literature and hold in many applications, e.g., Cournot games and Kelly Auctions [8], [29], [42].

Convergence analysis for games with multiple Nash equilibria is in general hard. For this reason, recent research has focused on the so-called strongly monotone games, which are shown in [41] to admit a unique Nash equilibrium. The game (1) is said to be m -strongly monotone with $m > 0$, if for all $x, x' \in \mathcal{X}$ it satisfies

$$\sum_{i=1}^N \langle \nabla_i \mathcal{C}_i(x) - \nabla_i \mathcal{C}_i(x'), x_i - x'_i \rangle \geq m \|x - x'\|^2. \quad (2)$$

For convex games, the ability of agents to efficiently learn their optimal actions can be quantified using the notion of regret, which captures the difference between the agents' sequential actions and the achievable best actions in hindsight. Given the sequences of actions $\{x_{i,t}\}_{t=1}^T, i = 1, \dots, N$, the regret of agent i is defined as

$$R_i(T) = \sum_{t=1}^T \mathcal{C}_i(x_{i,t}, x_{-i,t}) - \min_{x_i \in \mathcal{X}_i} \sum_{t=1}^T \mathcal{C}_i(x_i, x_{-i,t}). \quad (3)$$

An algorithm is said to be no-regret if the regret of each agent is sub-linear in T .

Our goal is to design an asymmetric learning algorithm to solve the game (1), when the zeroth-order agents $i_z \in \mathcal{N}_z$ and the first-order agents $i_f \in \mathcal{N}_f$ update their actions using the specific type information available to them. We aim to show that the proposed algorithm achieves no-regret learning and Nash equilibrium convergence for convex and strongly monotone games, respectively.

Algorithm 1: Asymmetric learning

```

1: Input: Initial value  $x_1$ , positive sequences  $\eta_{f,t}, \eta_{z,t}, \delta_t$ , parameters  $N_z, N, T$ .
2: for episode  $t = 1, \dots, T$  do
3:   Agents play their actions:
4:   for agent  $i = 1, \dots, N$  do
5:     Agent  $i_z \in \mathcal{N}_z$  samples  $u_{i_z,t} \in \mathbb{S}$  and plays
6:      $\hat{x}_{i_z,t} = x_{i_z,t} + \delta_t u_{i_z,t}$ 
7:     Agent  $i_f \in \mathcal{N}_f$  plays  $x_{i_f,t}$ 
8:   end for
9:   Agents receive information:
10:  for agent  $i = 1, \dots, N$  do
11:    Agent  $i_z \in \mathcal{N}_z$  obtains  $\mathcal{C}_{i_z}(\hat{x}_{z,t}, x_{f,t})$ 
12:    Agent  $i_f \in \mathcal{N}_f$  obtains  $\nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t})$ 
13:  end for
14:  Agents perform updates:
15:  for agent  $i = 1, \dots, N$  do
16:    Agent  $i_z \in \mathcal{N}_z$  updates according to (4)
17:    Agent  $i_f \in \mathcal{N}_f$  updates according to (6)
18:  end for
19: Output:  $\{x_t\}$ .

```

III. AN ASYMMETRIC LEARNING ALGORITHM

In this section, we propose an asymmetric learning algorithm that allows the agents to update their actions using first-order or zeroth-order gradient information, whichever is available to them.

By saying that agent i has access to the first-order gradient information, we mean that agent i can compute the gradient $\nabla_i \mathcal{C}_i$. If agent i only has access to zeroth-order oracle, then the available information for agent i is its function evaluation \mathcal{C}_i . We assume that agents are rational and always use first-order gradient information to update their actions when available, while those with only zeroth-order gradient information use this information for updating.

The detailed algorithm is presented as Algorithm 1. At each episode t , each agent $i_z \in \mathcal{N}_z$ that has access to a zeroth-order oracle perturbs its action by an amount of $\delta_t u_{i_z,t}$ and plays the perturbed action $\hat{x}_{i_z,t} = x_{i_z,t} + \delta_t u_{i_z,t}$, where $u_{i_z,t} \in \mathbb{S}$ is a random perturbation direction sampled from a unit sphere $\mathbb{S} \subset \mathbb{R}^d$ and δ_t is the size of this perturbation. In contrast, each agent $i_f \in \mathcal{N}_f$ with first-order gradient information plays the action $x_{i_f,t}$. Here we collect the actions of all agents that have access to zeroth-order oracles in a vector $\hat{x}_{z,t} := (\hat{x}_{1,t}, \dots, \hat{x}_{N_z,t})$, and the actions of all agents that have access to first-order gradient information in a vector $x_{f,t} := (x_{N_z+1,t}, \dots, x_{N,t})$. After all agents have played their actions, they receive their own asymmetric information. Specifically, given the played action profile $(\hat{x}_{z,t}, x_{f,t})$ at time step t , each agent i_z obtains the returned function evaluation $\mathcal{C}_{i_z}(\hat{x}_{z,t}, x_{f,t})$, while each agent i_f gets the first-order gradient information $\nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t})$. Then, each agent i_z performs the following projected update

$$x_{i_z,t+1} \leftarrow \mathcal{P}_{\mathcal{X}_{i_z}^{\delta_t}} \left(x_{i_z,t} - \eta_{z,t} g_{i_z,t} \right), \quad (4)$$

where the step size $\eta_{z,t}$ of agents with zeroth-order oracles will be designed later. The zeroth-order gradient estimate is constructed as

$$g_{i_z,t} = \frac{d}{\delta_t} \mathcal{C}_{i_z}(\hat{x}_{z,t}, x_{f,t}) u_{i_z,t}. \quad (5)$$

The projection set is defined as $\mathcal{X}_{i_z}^{\delta_t} = \{x_{i_z} \in \mathcal{X}_{i_z} \mid \frac{1}{1-\delta_t/r} x_{i_z} \in \mathcal{X}_{i_z}\}$. The projection step guarantees the feasibility of the sampled action $\hat{x}_{i_z,t}$, since

$$\begin{aligned} & \left(1 - \frac{\delta_t}{r}\right) \mathcal{X}_{i_z} \oplus \delta_t \mathbb{S} \\ &= \left(1 - \frac{\delta_t}{r}\right) \mathcal{X}_{i_z} \oplus \frac{\delta_t}{r} r \mathbb{S} \\ &\subseteq \left(1 - \frac{\delta_t}{r}\right) \mathcal{X}_{i_z} \oplus \frac{\delta_t}{r} \mathcal{X}_{i_z} = \mathcal{X}_{i_z}. \end{aligned}$$

Here, \oplus denotes the Minkowski sum of two sets. Each agent i_f performs the action update

$$x_{i_f,t+1} \leftarrow \mathcal{P}_{\mathcal{X}_{i_f}} \left(x_{i_f,t} - \eta_{f,t} \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}) \right), \quad (6)$$

where $\eta_{f,t}$ is the step size of agents with first-order gradient information and will be designed later.

In our framework, each agent $i_z \in \mathcal{N}_z$ with no access to first-order gradient information plays a randomly perturbed action $\hat{x}_{i_z,t}$, as in standard zeroth-order optimization [43]. Then, the agents with access to the first-order gradients obtain gradient information at the random point $(x_{f,t}, \hat{x}_{z,t})$ rather than $(x_{f,t}, x_{z,t})$. In other words, agents with access to zeroth-order oracles add randomness to the action updates of agents with first-order gradient information. This randomness does not exist in the pure first-order method where all agents obtain gradient information at a deterministic point x_t . Consequently, such asymmetric updates lead to more complex agent interactions.

Although Algorithm 1 may appear to be a combination of the pure first- and zeroth-order methods, the analysis of these two methods cannot be directly applied to our algorithm. On one hand, as stated above, the presence of agents with zeroth-order oracles introduces extra randomness to the action updates of agents with first-order gradient information, thereby increasing the whole system's complexity. On the other hand, agents with first-order oracles introduce discoordination for the whole system. Specifically, the dynamics of first-order agents directly affect the cost function values, thereby affecting the gradient estimates for agents with zeroth-order oracles. Consequently, the constructed gradient estimates become biased of the smoothed functions in pure zeroth-order case, rendering traditional zeroth-order techniques inapplicable here.

IV. CONVERGENCE RESULTS FOR CONVEX GAMES

In this section, we analyze the convergence of Algorithm 1 given that the game (1) is convex.

The standard smoothed functions commonly used in analyzing pure zeroth-order methods rely on the premise that all agents obtain gradient estimates based on zeroth-order optimization techniques. However, the presence of agents with first-order gradient information in our asymmetric game

renders the premise invalid, making these smooth functions unsuitable for our analysis. To solve this problem, we construct new smoothed functions tailored for such asymmetric games. Specifically, we define $\mathcal{C}_{i_z,t}(x_z) := \mathcal{C}_{i_z}(x_z, x_{f,t})$ and the smoothed function $\mathcal{C}_{i_z,t}^{\delta_t}(x_z) := \mathbb{E}_{w_{i_z} \sim \mathbb{B}, u_{-i_z} \sim \mathbb{S}_{-i_z}} \mathcal{C}_{i_z,t}(x_z + \delta_t w_{i_z}, x_{-i_z} + \delta_t u_{-i_z})$ which serves as an approximation of $\mathcal{C}_{i_z,t}(x_z)$. Here, \mathbb{B} , \mathbb{S} denote the unit ball and unit sphere in \mathbb{R}^d , respectively, and $\mathbb{S}_{-i_z} := \Pi_{j=1}^{N_z-1} \mathbb{S}$. The smoothed cost function in symmetric games [8] is smoothed over actions of all agents, while our smoothed cost function is only smoothed over actions of zeroth-order agents. This modification lays the foundation for our subsequent analysis. When all agents use zeroth-order optimization, the smoothed function reduces to the one in [8]. It can be shown that the function $\mathcal{C}_{i_z,t}^{\delta_t}(x_z)$ has the following properties [40].

Lemma 1. *Let Assumptions 1 and 3 hold. Then, we have that for all $t \geq 1$*

- 1) $\mathcal{C}_{i_z,t}^{\delta_t}(x_z)$ is convex in x_z ;
- 2) $\mathcal{C}_{i_z,t}^{\delta_t}(x_z)$ is L_0 -Lipschitz continuous in x_z ;
- 3) $|\mathcal{C}_{i_z,t}^{\delta_t}(x_z) - \mathcal{C}_{i_z,t}(x_z)| \leq \delta_t L_0 \sqrt{N_z}$;
- 4) $\mathbb{E} \left[\frac{d}{\delta_t} \mathcal{C}_{i_z,t}(\hat{x}_z) u_{i_z} \right] = \nabla_{i_z} \mathcal{C}_{i_z,t}^{\delta_t}(x_z)$.

The smoothed function $\mathcal{C}_{i_z,t}^{\delta_t}$ is defined as a time-varying cost function of the actions of all agents that have access to zeroth-order oracles. Despite its time-varying nature, Lemma 1 shows that some properties of the function $\mathcal{C}_{i_z,t}^{\delta_t}$ still hold for any $x_{f,t}$. Note that the last property in Lemma 1 shows that the term $\frac{d}{\delta_t} \mathcal{C}_{i_z,t}(\hat{x}_z) u_{i_z}$ is an unbiased estimate of the gradient of the smoothed function $\mathcal{C}_{i_z,t}^{\delta_t}(x_z)$.

We use the notion of regret to measure the performance of our algorithm. Given a sequence of agent actions $\{\hat{x}_{i_z,t}\}_{t=1}^T$, $i_z \in \mathcal{N}_z$, and $\{x_{i_f,t}\}_{t=1}^T$, $i_f \in \mathcal{N}_f$ generated by Algorithm 1, (3) yields the regret of agent $i_z \in \mathcal{N}_z$ as

$$\begin{aligned} R_{i_z}(T) &= \mathbb{E} \left[\sum_{t=1}^T \mathcal{C}_{i_z}(\hat{x}_{z,t}, x_{f,t}) \right. \\ &\quad \left. - \min_{x_{i_z} \in \mathcal{X}_{i_z}} \sum_{t=1}^T \mathcal{C}_{i_z}(x_{i_z}, \hat{x}_{-i_z,t}, x_{f,t}) \right], \end{aligned}$$

and agent $i_f \in \mathcal{N}_f$ as

$$\begin{aligned} R_{i_f}(T) &= \mathbb{E} \left[\sum_{t=1}^T \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}) \right. \\ &\quad \left. - \min_{x_{i_f} \in \mathcal{X}_{i_f}} \sum_{t=1}^T \mathcal{C}_{i_f}(x_{i_f}, x_{-i_f,t}, \hat{x}_{z,t}) \right]. \end{aligned}$$

We note that the definition of regret depends on actions of other agents, and the asymmetric updates of other agents make regret analysis more complex. By appropriately selecting the parameters $\eta_{f,t}$, $\eta_{z,t}$ and δ_t , we show that Algorithm 1 achieves sub-linear regret. The formal result is presented in the following theorem, in which the notion \mathcal{O} hides all constant factors except for N , N_z and T . Due to space limitations, the detailed proof is provided in [40].

Theorem 1. *Let Assumptions 1–3 hold, and select $\delta_t = \frac{1}{N_z^{1/4}t^{1/4}}$, $\eta_{z,t} = \frac{1}{N_z^{1/4}t^{3/4}}$, $\eta_{f,t} = \frac{1}{\sqrt{t}}$, for all $t \geq 1$. Then, for any $T \geq 1$, Algorithm 1 achieves regrets*

$$\begin{aligned} R_{i_z}(T) &= \mathcal{O}\left(N_z^{1/4}T^{3/4}\right), \quad \forall i_z \in \mathcal{N}_z; \\ R_{i_f}(T) &= \mathcal{O}\left(T^{1/2}\right), \quad \forall i_f \in \mathcal{N}_f. \end{aligned} \quad (7)$$

The design of the positive sequences $\eta_{z,t}$, $\eta_{f,t}$ and δ_t does not require the knowledge of the total number of episodes T . Using these sequences, Theorem 1 shows that Algorithm 1 achieves sub-linear regret for all agents, although agents that have access to first-order gradient information have a smaller regret bound in terms of T compared to agents accessing zeroth-order oracles.

Using similar techniques as in [3], one can show that the pure first-order method achieves a regret of $\mathcal{O}(T^{1/2})$ while the pure zeroth-order method achieves a regret of $\mathcal{O}(N^{1/4}T^{3/4})$. We observe that the regret for agents with first-order gradient information is not affected by the asymmetric setting. This is due to the facts that other agents' varying actions can be deemed as a source of varying cost functions and first-order gradient descent update is no-regret under varying cost functions. However, the agents with zeroth-order oracles are affected by this coordination induced by the asymmetric setting. Specifically, the dynamics of the first-order agents directly affect the cost function values, thereby affecting the zeroth-order gradient estimates. Despite this, we show that the constructed gradient estimate is an unbiased estimate of the gradient of the newly constructed smooth function, and the regret bound of zeroth-order agents achieved by Algorithm 1 is always better than that achieved by the pure zeroth-order method. Therefore, Algorithm 1 not only ensures that the agents accessing first-order gradient information inherit the lower regret bound $\mathcal{O}(T^{1/2})$ of the pure first-order method, but also improves the regret bound of the agents accessing zeroth-order oracles, compared to the regret achieved by the pure zeroth-order method. Let the total regret denote the sum of every agent's regret, which we denote by $R = \sum_{i_z \in \mathcal{N}_z} R_{i_z} + \sum_{i_f \in \mathcal{N}_f} R_{i_f}$. The total regret achieved by Algorithm 1 is lower bounded by that of the pure zeroth-order method and upper bounded by that of the pure first-order method.

V. CONVERGENCE RESULTS FOR STRONGLY MONOTONE GAMES

In this section, we analyze Nash equilibrium convergence given that the game (1) is strongly monotone. It is well-known that the Nash equilibrium is unique in strongly monotone games [41], which we denote by x^* . In what follows, we provide the last-iterate Nash equilibrium convergence result for Algorithm 1.

Theorem 2. *Suppose that the game (1) is m -strongly monotone. Let Assumptions 1–4 hold, and select $\delta_t = \frac{N_z^{1/6}}{N^{1/3}t^{1/3}}$, $\eta_{z,t} = \eta_{f,t} = \frac{1}{mt}$, for all $t = 1, \dots, T$. Then, Algorithm 1 satisfies*

$$\mathbb{E} \|x_T - x^*\|^2$$

$$= \mathcal{O}\left(N_z^{2/3}N^{2/3}T^{-1/3} + (N - N_z)T^{-1}\right). \quad (8)$$

Proof Sketch: We analyze the convergence to the Nash equilibrium by separately examining the error dynamics for first- and zeroth-order agents and then combining these analyses together. Specifically, for first-order agents, we study the evolution of the squared error $\|x_{i_f,t+1} - x_{i_f}^*\|^2$ leveraging the corresponding update rule (6). For zeroth-order agents, due to the varying parameters of δ_t in the smoothed function for zeroth-order agents, we first analyze the evolution of $\|x_{i_z,t+1} - x_{i_z,t}^*\|$ leveraging the update rule (4), where $x_{i_z,t}^* = (1 - \frac{\delta_t}{r})x_{i_z}^*$. Then, we analyze $\|x_{i_z,t+1} - x_{i_z}^*\|$ by establishing a bound on $\|x_{i_z}^* - x_{i_z,t}^*\|^2$.

Although the dynamics of the two agent types are interconnected, we can bound the influence of one agent type on the other by leveraging the properties established in Lemma 1. Next, we combine the dynamics of all agents to analyze the behavior at the group level, i.e., the convergence toward the Nash equilibrium. By applying the strong monotonicity condition, we characterize the evolution of the Nash equilibrium error, quantified by $\|x_t - x^*\|^2$. Finally, we prove convergence using induction. ■

Theorem 2 shows that Algorithm 1 achieves the last-iterate Nash equilibrium convergence for strongly monotone games with diminishing smoothing parameters δ_t and diminishing step sizes $\eta_{f,t}$ and $\eta_{z,t}$. Note that the design of diminishing parameters does not require the information of the total number of episodes T . Moreover, under Assumption 3, Nash equilibrium convergence implies sub-linear regret for each agent since the regret is upper bounded by the sum of differences between their actions and Nash equilibrium.

When $N_z = 0$, meaning that all agents use the first-order gradient-descent update, the convergence rate $O(NT^{-1})$ matches the result of learning in strongly monotone games with stochastic gradient information [44]. In the other extreme case $N_z = N$, i.e., all agents have only access to zeroth-order oracles, the term $(N - N_z)T^{-1}$ disappears and the convergence rate $O(N^{4/3}T^{-1/3})$ matches the result of zeroth-order learning in games [8]. We observe that the convergence rate of Algorithm 1 lies between those of first- and zeroth-order methods. It is worth noting that the convergence rate of $O(T^{-1/3})$ for zeroth-order optimization in games is sub-optimal, as indicated in [42]. Therefore, there is potential to improve the convergence rate in (8). However, achieving this in the asymmetric setting is challenging, and we leave it for future work.

To end this section, we would like to briefly explain the challenges in the proof of Theorem 2. Algorithm 1 is a combination of the pure first- and zeroth-order methods. However, the Nash equilibrium convergence analysis of Algorithm 1 cannot be obtained by directly combining the results of these two pure methods, since the dynamics of the two groups of agents, i.e., agents with first-order gradient information and zeroth-order oracles, are coupled together. On the one hand, agents with access to zeroth-order oracles spread their randomness to the dynamics of the agents with first-order gradient information. On the other hand, the presence of first-order

agents slightly disrupt the coordination among those agents with zeroth-order oracles aiming to minimize the smoothed cost function. This coupled dynamics, combined with the setting of diminishing parameters, complicates the analysis of Nash equilibrium convergence of the whole system.

VI. NUMERICAL EXPERIMENTS

In this section we use a Cournot game to illustrate the performance of Algorithm 1. Consider a market problem with N agents. Suppose that each agent i supplies the market with quantity x_i and the total supply of all agents decides the price of the goods in the market. Given the price, each agent i has the cost $C_i(x)$. Each agent aims to minimize its own cost through repeated learning in the game. In what follows, we consider two cases, a deterministic game and a stochastic game with risk-averse agents.

A. Deterministic Cournot Game

Consider a market with $N = 10$ agents. The cost function of each agent is given by

$$C_i(x) = x_i \left(\frac{a_i x_i}{2} + b_i \sum_{j \neq i} x_j - e_i \right) + 1,$$

where $a_i > 0$, b_i , e_i are constant parameters. It is easy to verify that

$$\nabla_i C_i(x) = a_i x_i + b_i x_{-i} - e_i.$$

The parameters are selected as

$$\begin{aligned} a &= [2, 2, 1.5, 1.8, 2, 1.8, 2, 1.4, 1.8, 2], \\ b &= [0.2, 0.3, 0.3, 0.2, 0.3, 0.2, 0.3, 0.2, 0.3, 0.3], \\ e &= [1.8, 1.9, 1.5, 1.6, 1.8, 1.3, 1.2, 1.5, 1.8, 1.6]. \end{aligned}$$

It can be verified that the game is m -strongly monotone with $m = 1.284$, and the Nash equilibrium is

$$x^* = [0.57, 0.44, 0.29, 0.52, 0.38, 0.33, 0.026, 0.61, 0.43, 0.26].$$

The projection set is defined as $\mathcal{X}_i = [0, 3]$.

We run Algorithm 1 with different values of N_z , as well as the pure first- and zeroth-order methods. We choose the step sizes $\eta_{f,t} = \eta_{z,t} = \frac{0.6}{t}$ and the parameter $\delta_t = \frac{0.5}{t}$. Fig. 1 illustrates the convergence rate of these algorithms. We observe that our asymmetric learning algorithm with different values of N_z always performs worse than the pure first-order method but better than the pure zeroth-order method, which agrees with Theorem 2. Besides, a larger value of N_z , which means that fewer agents have access to first-order gradient information, leads to a slower convergence rate; when N_z approaches N , the convergence rate is similar to that of the pure zeroth-order method.

Fig. 2 illustrates the error between the action x_i and the corresponding Nash equilibrium x_i^* for every agent i . We term the agents with access to first-order gradient information and zeroth-order oracles as first-order agents and zeroth-order agents, respectively. We select $N_z = 5$ and compute the average error to the Nash equilibrium for first- and zeroth-order agents. As shown in Fig. 2, first-order agents converge faster than the zeroth-order agents. Fig. 3 plots the average

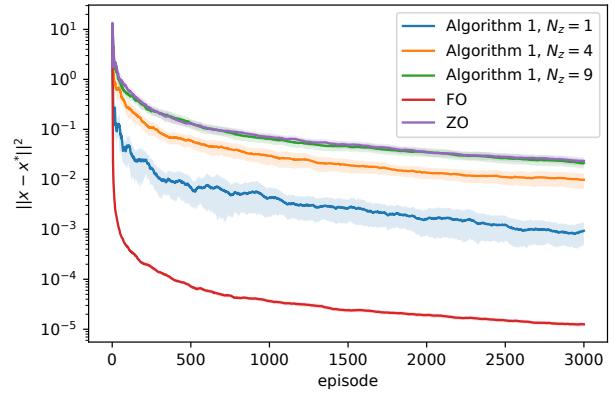


Fig. 1. Nash equilibrium convergence achieved by the pure first-order method (FO) and the pure zeroth-order method (ZO), and our asymmetric learning algorithm with different values of N_z . The solid lines and shades are averages and standard deviations over 50 runs.

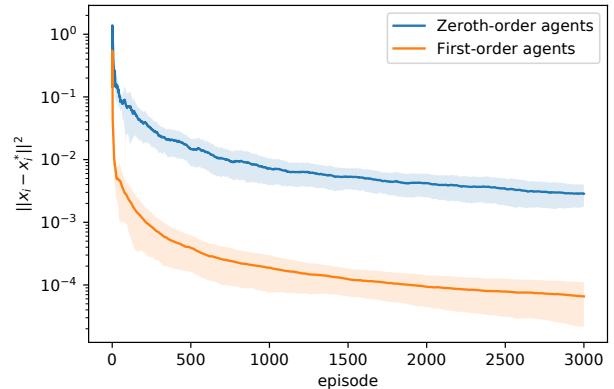


Fig. 2. Average error to Nash equilibrium point of first-order agents and zeroth-order agents achieved by Algorithm 1 when $N_z = 5$. The solid lines and shades are averages and standard deviations over 50 runs.

regret over 50 runs for first-order agents and zeroth-order agents. We observe that agents that use first-order gradient have much smaller regret than the agents with zeroth-order oracles, which aligns with our theoretical result.

B. Risk-averse Cournot Game

Consider a market with two agents, i.e., $N = 2$. Each agent decides the quantity x_i , $i = 1, 2$. The stochastic cost of each agent is defined by $J_i(x) = -(2 - \sum_{j=1}^2 x_j)x_i + 0.2x_i + \xi_i x_i + 1$, where ξ_i is a random variable used to represent the uncertainty in the market. Agents aim to minimize the risk of incurring high costs, i.e., agents are risk-averse. We use CVaR as the risk measure and denote the risk level of each agent i by α_i . It is well known that the CVaR value represents the average of the worst α_i percent of the stochastic cost, and when $\alpha_i = 1$, it is equivalent to the risk-neutral case. Let agent 1 select the risk level $\alpha_1 = 0.5$ while agent 2 is risk-neutral, i.e., $\alpha_2 = 1$. The objective functions of these

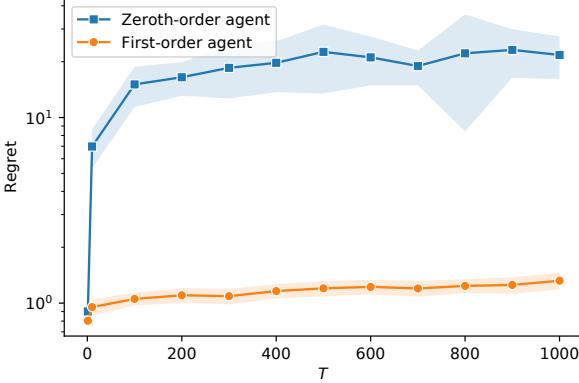


Fig. 3. Regret $\sum_{t=1}^T \mathcal{C}_i(x_t) - \min_{x_i \in \mathcal{X}_i} \sum_{t=1}^T \mathcal{C}_i(x_i, x_{-i,t})$ of first-order agents and zeroth-order agents achieved by Algorithm 1 when $N_z = 5$. The solid lines and shades are averages and standard deviations over 50 runs.

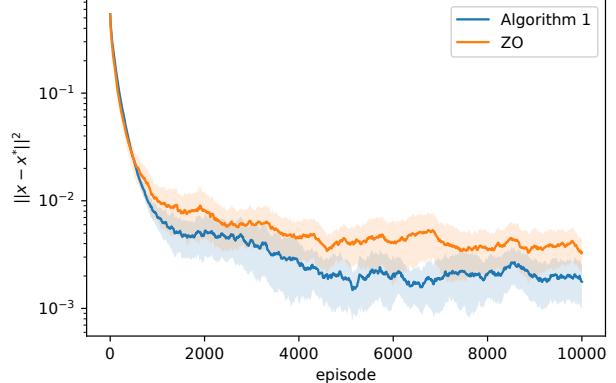


Fig. 5. Error to Nash equilibrium point achieved by Algorithm 1 and the pure zeroth-order method (ZO) in risk-averse games. The solid lines and shades are averages and standard deviations over 50 runs.

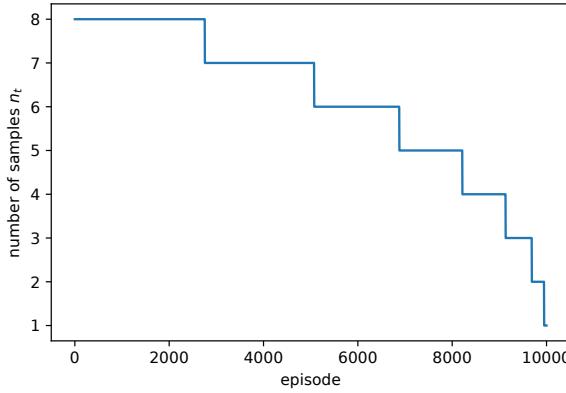


Fig. 4. The number of samples of Algorithm 1 and the pure zeroth-order method (Algorithm 1 in [6]).

two agents are $\mathcal{C}_1(x) := \text{CVaR}_{\alpha_1}[J_1(x, \xi_1)]$ and $\mathcal{C}_2(x) := \mathbb{E}[J_2(x, \xi_2)]$, respectively, where the definition of CVaR can be found in [6].

The gradient of the CVaR function is hard to compute in most cases even when the function $J_1(x, \xi_1)$ is completely known [41]. We assume that given the played action profile x , the risk-averse agent does not have access to gradient information but only to the cost evaluation $J_1(x, \xi_1)$. The risk-neutral agent has access to gradient information $\nabla_2 \mathcal{C}_2(x)$.

Since the CVaR value cannot be estimated from only one sample, we use the sampling strategy proposed in [6] that uses a decreasing number of samples with respect to the number of iterations. The number of samples n_t chosen here is shown in Fig. 4. With this sampling strategy, the agents have n_t samples at episode t to estimate the CVaR values.

Let $\xi_i \sim U(0, 0.4)$ be a uniform random variable. Then, we can obtain an explicit expression of the cost function $\mathcal{C}_1(x)$. Further, using the first-order optimality condition we can compute the Nash equilibrium to be $(0.4667, 0.5667)$. We run Algorithm 1 and the pure zeroth-order algorithm

(Algorithm 1 in [6]), and select the step size $\eta_t = \frac{0.005}{t}$ and the parameter $\delta = \frac{0.5}{t}$ for both algorithms. Fig. 5 shows that our asymmetric learning algorithm converges faster than the pure zeroth-order method.

VII. CONCLUSION

In this work, we proposed an asymmetric learning algorithm for convex games in which the agents update their actions using either first-order gradient information or zeroth-order oracles. We showed that our algorithm achieves sub-linear regret for convex games and last-iterate Nash equilibrium convergence for the class of strongly monotone games. Our theoretical analysis further established that the performance of the proposed algorithm consistently lies between those of the pure first-order and zeroth-order methods. This result highlights the algorithm's flexibility, as it can interpolate between these two extremes by varying the number of agents with access to gradient or oracle information. In particular, our algorithm recovers both the pure first-order and pure zeroth-order methods as special cases. We demonstrated the effectiveness of our algorithm through numerical experiments on deterministic and stochastic risk-averse Cournot games. These simulations illustrate the adaptability of the algorithm in handling heterogeneous agent information structures.

Several promising directions emerge for future research. One avenue is the development of a generalized theoretical framework to systematically study asymmetric learning algorithms that enable agents to employ multiple strategies simultaneously. Another intriguing direction involves exploring the fairness in such asymmetric settings. Besides, extending our analysis to more complex game structures, such as non-convex or hierarchical games, and studying the implications on equilibrium computation and convergence rates would also be interesting.

REFERENCES

[1] Stephen P Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[2] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of International Conference on Machine Learning*, pages 928–936, 2003.

[3] Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

[4] Yurii Nesterov. *Lectures on Convex Optimization*, volume 137. Springer, 2018.

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

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

[7] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of Reinforcement Learning and Control*, pages 321–384, 2021.

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

[9] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Pilopoulos. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2703–2717, 2018.

[10] 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 *Proceedings of IEEE Conference on Decision and Control*, pages 5179–5184, 2022.

[11] Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. Adaptive learning in continuous games: Optimal regret bounds and convergence to Nash equilibrium. In *Proceedings of Conference on Learning Theory*, pages 2388–2422, 2021.

[12] Giuseppe Belgioioso, Angelia Nedić, and Sergio Grammatico. Distributed generalized Nash equilibrium seeking in aggregative games on time-varying networks. *IEEE Transactions on Automatic Control*, 66(5):2061–2075, 2020.

[13] Chung-Wei Lee, Christian Kroer, and Haipeng Luo. Last-iterate convergence in extensive-form games. In *Proceedings of Advances in Neural Information Processing Systems*, pages 14293–14305, 2021.

[14] Sebastian Bervoets, Mario Bravo, and Mathieu Faure. Learning with minimal information in continuous games. *Theoretical Economics*, 15(4):1471–1508, 2020.

[15] Tatiana Tatarenko and Maryam Kamgarpour. Bandit online learning of Nash equilibria in monotone games. *arXiv preprint arXiv:2009.04258*, 2020.

[16] Tatiana Tatarenko and Maryam Kamgarpour. Learning generalized Nash equilibria in a class of convex games. *IEEE Transactions on Automatic Control*, 64(4):1426–1439, 2018.

[17] Dario Paccagnan, Basilio Gentile, Francesca Parise, Maryam Kamgarpour, and John Lygeros. Nash and Wardrop equilibria in aggregative games with coupling constraints. *IEEE Transactions on Automatic Control*, 64(4):1373–1388, 2018.

[18] Paul Frihauf, Miroslav Krstic, and Tamer Başar. Nash equilibrium seeking in noncooperative games. *IEEE Transactions on Automatic Control*, 57(5):1192–1207, 2011.

[19] Hamidou Tembine, Quanyan Zhu, and Tamer Başar. Risk-sensitive mean-field games. *IEEE Transactions on Automatic Control*, 59(4):835–850, 2013.

[20] S Rasoul Etesami and Tamer Başar. Dynamic games in cyber-physical security: An overview. *Dynamic Games and Applications*, 9(4):884–913, 2019.

[21] Haifeng Xu, Zinovi Rabinovich, Shaddin Dughmi, and Milind Tambe. Exploring information asymmetry in two-stage security games. In *Proceedings of AAAI Conference on Artificial Intelligence*, pages 1057–1063, 2015.

[22] Yu-Guan Hsieh, Kimon Antonakopoulos, Volkan Cevher, and Panayotis Mertikopoulos. No-regret learning in games with noisy feedback: Faster rates and adaptivity via learning rate separation. In *Proceedings of Advances in Neural Information Processing Systems*, pages 6544–6556, 2022.

[23] Tianyi Lin, Zhengyuan Zhou, Panayotis Mertikopoulos, and Michael Jordan. Finite-time last-iterate convergence for multi-agent learning in games. In *Proceedings of International Conference on Machine Learning*, pages 6161–6171, 2020.

[24] Noah Golowich, Sarah Pattathil, and Constantinos Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. In *Proceedings of Advances in neural information processing systems*, pages 20766–20778, 2020.

[25] Panayotis Mertikopoulos and Zhengyuan Zhou. Gradient-free online learning in games with delayed rewards. In *Proceedings of International Conference on Machine Learning*, pages 4172–4181, 2020.

[26] Panayotis Mertikopoulos and Mathias Staudigl. Equilibrium tracking and convergence in dynamic games. In *Proceedings of IEEE Conference on Decision and Control*, pages 930–935, 2021.

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

[28] Adhyayan Narang, Evan Faulkner, Dmitry Drusvyatskiy, Maryam Fazel, and Lillian Ratliff. Learning in stochastic monotone games with decision-dependent data. In *Proceedings of International Conference on Artificial Intelligence and Statistics*, pages 5891–5912, 2022.

[29] Benoit Duvocelle, Panayotis Mertikopoulos, Mathias Staudigl, and Dries Vermeulen. Multiagent online learning in time-varying games. *Mathematics of Operations Research*, 48(2):914–941, 2023.

[30] Dmitry Drusvyatskiy, Maryam Fazel, and Lillian J Ratliff. Improved rates for derivative free gradient play in strongly monotone games. In *Proceedings of IEEE Conference on Decision and Control*, pages 3403–3408, 2022.

[31] David Balduzzi, Sébastien Racanière, James Martens, Jakob Foerster, Karl Tuyls, and Thore Graepel. The mechanics of n-player differentiable games. In *Proceedings of International Conference on Machine Learning*, pages 354–363, 2018.

[32] Benjamin Chasnov, Lillian Ratliff, Eric Mazumdar, and Samuel Burden. Convergence analysis of gradient-based learning in continuous games. In *Proceedings of Uncertainty in Artificial Intelligence*, pages 935–944, 2020.

[33] Eric Mazumdar, Lillian J Ratliff, and S Shankar Sastry. On gradient-based learning in continuous games. *SIAM Journal on Mathematics of Data Science*, 2(1):103–131, 2020.

[34] Ville Könönen. Asymmetric multiagent reinforcement learning. *Web Intelligence and Agent Systems: An International Journal*, 2(2):105–121, 2004.

[35] Ezra Tampubolon, Haris Ceribasic, and Holger Boche. On information asymmetry in competitive multi-agent reinforcement learning: convergence and optimality. *arXiv preprint arXiv:2010.10901*, 2020.

[36] Tao Li and Quanyan Zhu. Commitment with signaling under double-sided information asymmetry. *arXiv preprint arXiv:2212.11446*, 2022.

[37] Yuksel Arslantaş, Ege Yuceel, Yigit Yalin, and Muhammed O Sayin. Convergence of heterogeneous learning dynamics in zero-sum stochastic games. *arXiv preprint arXiv:2311.00778*, 2023.

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

[39] Xiaochun Niu and Ermin Wei. Dish: A distributed hybrid primal-dual optimization framework to utilize system heterogeneity. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 6503–6510. IEEE, 2022.

[40] Zifan Wang, Xinlei Yi, Yi Shen, Michael M Zavlanos, and Karl H Johansson. Asymmetric learning in convex games. *arXiv preprint arXiv:2307.08812*, 2023.

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

[42] 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.

[43] Abraham D. Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, page 385–394, USA, 2005. Society for Industrial and Applied Mathematics.

[44] Michael Jordan, Tianyi Lin, and Zhengyuan Zhou. Adaptive, doubly optimal no-regret learning in strongly monotone and exp-concave games with gradient feedback. *Operations Research*, 2024.

[45] Halsey Lawrence Royden and Patrick Fitzpatrick. *Real Analysis*, volume 32. Macmillan New York, 1988.

APPENDIX

A. Proof of Lemma 1

1) Since the function $\mathcal{C}_i(x)$ is convex in x_i for any fixed x_{-i} , we have that $\mathcal{C}_{i_z,t}(x_z) = \mathcal{C}_{i_z}(x_z, x_{f,t})$ is convex in x_{i_z} . Therefore, for any $p_1, p_2 \in \mathcal{X}_{i_z}^{\delta_t}$ and $\theta \in [0, 1]$, we have that $\mathcal{C}_{i_z,t}(\theta p_1 + (1 - \theta)p_2, x_{-i_z}) \leq \theta \mathcal{C}_{i_z,t}(p_1, x_{-i_z}) + (1 - \theta) \mathcal{C}_{i_z,t}(p_2, x_{-i_z})$. Based on this, it gives that

$$\begin{aligned} & \mathcal{C}_{i_z,t}^{\delta_t}(\theta p_1 + (1 - \theta)p_2, x_{-i_z}) \\ &= \mathbb{E}[\mathcal{C}_{i_z,t}(\theta p_1 + (1 - \theta)p_2 + \delta_t w_{i_z}, x_{-i_z} + \delta_t u_{-i_z})] \\ &= \mathbb{E}[\mathcal{C}_{i_z,t}(\theta(p_1 + \delta_t w_{i_z}) + (1 - \theta)(p_2 + \delta_t w_{i_z}), \\ & \quad x_{-i_z} + \delta_t u_{-i_z})] \\ &\leq \mathbb{E}[\theta \mathcal{C}_{i_z,t}(p_1 + \delta_t w_{i_z}, x_{-i_z} + \delta_t u_{-i_z}) \\ & \quad + (1 - \theta) \mathcal{C}_{i_z,t}(p_2 + \delta_t w_{i_z}, x_{-i_z} + \delta_t u_{-i_z})] \\ &= \theta \mathcal{C}_{i_z,t}^{\delta_t}(p_1, x_{-i_z}) + (1 - \theta) \mathcal{C}_{i_z,t}^{\delta_t}(p_2, x_{-i_z}), \end{aligned}$$

where the expectations above are taken with respect to $w_{i_z} \sim \mathbb{B}$ and $u_{-i_z} \sim \mathbb{S}_{-i_z}$. The proof is complete.

2) Since the function $\mathcal{C}_i(x)$ is L_0 -Lipschitz continuous in x , we have that $\mathcal{C}_{i_z,t}(x_z)$ is L_0 -Lipschitz continuous in x_z . Thus, for arbitrary two points $x_z, x'_z \in \mathcal{X}_z$ with $\mathcal{X}_z = \prod_{i_z=1}^{N_z} \mathcal{X}_{i_z}$, we have that

$$\begin{aligned} & |\mathcal{C}_{i_z,t}^{\delta_t}(x_z) - \mathcal{C}_{i_z,t}^{\delta_t}(x'_z)| \\ &= \left| \mathbb{E}_{w_{i_z} \sim \mathbb{B}, u_{-i_z} \sim \mathbb{S}_{-i_z}} [\mathcal{C}_{i_z,t}(x_{i_z} + \delta_t w_{i_z}, x_{-i_z} + \delta_t u_{-i_z}) \right. \\ & \quad \left. - \mathcal{C}_{i_z,t}(x'_{i_z} + \delta_t w_{i_z}, x'_{-i_z} + \delta_t u_{-i_z})] \right| \\ &\leq \mathbb{E}_{w_{i_z} \sim \mathbb{B}, u_{-i_z} \sim \mathbb{S}_{-i_z}} [L_0 \|x_z - x'_z\|] \\ &\leq L_0 \|x_z - x'_z\|. \end{aligned}$$

The proof is complete.

3) From the definition of $\mathcal{C}_{i_z,t}$, it directly follows that

$$\begin{aligned} & |\mathcal{C}_{i_z,t}^{\delta_t}(x_z) - \mathcal{C}_{i_z,t}(x_z)| \\ &= \left| \mathbb{E}_{w_{i_z} \sim \mathbb{B}, u_{-i_z} \sim \mathbb{S}_{-i_z}} [\mathcal{C}_{i_z,t}(x_{i_z} + \delta_t w_{i_z}, x_{-i_z} + \delta_t u_{-i_z}) \right. \\ & \quad \left. - \mathcal{C}_{i_z,t}(x_z)] \right| \\ &\leq \mathbb{E}_{w_{i_z} \sim \mathbb{B}, u_{-i_z} \sim \mathbb{S}_{-i_z}} L_0 \|(\delta_t w_{i_z}, \delta_t u_{-i_z})\| \\ &\leq L_0 \delta_t \sqrt{N_z}. \end{aligned}$$

4) The proof of this claim can be directly adapted from Lemma C.1 in [8] and thus omitted.

B. Proof of Theorem 1

We first focus on the regret analysis of agents with first-order gradient information. As a reminder, we use i_f to index an agent with access to first-order gradient information and i_z to index an agent using zeroth-order oracles. Define $y_{i_f}^* := \operatorname{argmin}_{x_{i_f} \in \mathcal{X}_{i_f}} \sum_{t=1}^T \mathcal{C}_{i_f}(x_{i_f}, x_{-i_f,t}, \hat{x}_{z,t})$. From the update rule (6), we have that

$$\|x_{i_f,t+1} - y_{i_f}^*\|^2$$

$$\begin{aligned} &= \left\| \mathcal{P}_{\mathcal{X}_{i_f}} \left(x_{i_f,t} - \eta_{f,t} \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}) \right) - y_{i_f}^* \right\|^2 \\ &\leq \|x_{i_f,t} - \eta_{f,t} \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}) - y_{i_f}^*\|^2 \\ &\leq \|x_{i_f,t} - y_{i_f}^*\|^2 + \eta_{f,t}^2 \|\nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t})\|^2 \\ &\quad - 2\eta_{f,t} \left\langle \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}), x_{i_f,t} - y_{i_f}^* \right\rangle, \end{aligned} \quad (9)$$

where the first inequality is due to the facts that $\mathcal{P}_{\mathcal{X}_{i_f}}(y_{i_f}^*) = y_{i_f}^*$ and the projection operator is non-expansive. From the convexity of the function $\mathcal{C}_{i_f}(x)$ in x_{i_f} , we have that

$$\begin{aligned} R_{i_f}(T) &= \mathbb{E} \left[\sum_{t=1}^T \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}) - \sum_{t=1}^T \mathcal{C}_{i_f}(y_{i_f}^*, x_{-i_f,t}, \hat{x}_{z,t}) \right] \\ &\leq \mathbb{E} \left[\sum_{t=1}^T \left\langle \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}), x_{i_f,t} - y_{i_f}^* \right\rangle \right] \\ &\leq \sum_{t=1}^T \frac{1}{2\eta_{f,t}} \left(\mathbb{E} \|x_{i_f,t} - y_{i_f}^*\|^2 - \mathbb{E} \|x_{i_f,t+1} - y_{i_f}^*\|^2 \right. \\ &\quad \left. + \eta_{f,t}^2 \mathbb{E} \|\nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t})\|^2 \right) \\ &\leq \sum_{t=1}^T \mathbb{E} \|x_{i_f,t} - y_{i_f}^*\|^2 \left(\frac{1}{2\eta_{f,t}} - \frac{1}{2\eta_{f,t-1}} \right) + \sum_{t=1}^T \frac{\eta_{f,t} L_0^2}{2} \\ &\leq \frac{D^2}{2} \sum_{t=1}^T \left(\frac{1}{\eta_{f,t}} - \frac{1}{\eta_{f,t-1}} \right) + \sum_{t=1}^T \frac{\eta_{f,t} L_0^2}{2} \\ &\leq \frac{D^2}{2\eta_{f,T}} + \sum_{t=1}^T \frac{\eta_{f,t} L_0^2}{2} \\ &\leq \frac{\sqrt{T}}{2} (D^2 + 2L_0^2), \end{aligned}$$

where $\frac{1}{\eta_{f,0}} := 0$. The second inequality follows from (9) and the last inequality follows since $\eta_{f,t} = \frac{1}{\sqrt{t}}$ and $\sum_{t=1}^T \frac{1}{\sqrt{t}} \leq 2\sqrt{T}$. The third inequality holds since, for any two sequences $\{a_t\}_{t=1}^T, \{b_t\}_{t=1}^T$, with $\frac{1}{a_0} := 0, a_t > 0, b_t \geq 0, \forall t = 1, \dots, T$, we have

$$\begin{aligned} \sum_{t=1}^T \frac{1}{a_t} (b_t - b_{t+1}) &= \sum_{t=1}^T b_t \left(\frac{1}{a_t} - \frac{1}{a_{t-1}} \right) - \frac{b_{T+1}}{a_T} \\ &\leq \sum_{t=1}^T b_t \left(\frac{1}{a_t} - \frac{1}{a_{t-1}} \right). \end{aligned}$$

The proof for the agents with first-order gradient information is complete.

Next we turn to the regret analysis of the agents with zeroth-order oracles. Define $y_{i_z}^* := \operatorname{argmin}_{x_{i_z} \in \mathcal{X}_{i_z}} \sum_{t=1}^T \mathcal{C}_{i_z}(x_{i_z}, \hat{x}_{-i_z,t}, x_{f,t})$ and $y_{i_z}^{\delta_t} := (1 - \frac{\delta_t}{r}) y_{i_z}^*$. Since $y_{i_z}^* \in \mathcal{X}_{i_z}$, we have $y_{i_z}^{\delta_t} \in \mathcal{X}_{i_z}^{\delta_t}$. By virtue of the update rule (4), we have that

$$\begin{aligned} & \|x_{i_z,t+1} - y_{i_z}^{\delta_t}\|^2 \\ &= \left\| \mathcal{P}_{\mathcal{X}_{i_z}^{\delta_t}} \left(x_{i_z,t} - \eta_{z,t} g_{i_z,t} \right) - y_{i_z}^{\delta_t} \right\|^2 \\ &\leq \|x_{i_z,t} - \eta_{z,t} g_{i_z,t} - y_{i_z}^{\delta_t}\|^2 \end{aligned}$$

$$\leq \left\| x_{i_z,t} - y_{i_z}^{\delta_t} \right\|^2 + \eta_{z,t}^2 \| g_{i_z,t} \|^2 - 2\eta_{z,t} \left\langle x_{i_z,t} - y_{i_z}^{\delta_t}, g_{i_z,t} \right\rangle, \quad (10)$$

where the first inequality is due to the fact that $\mathcal{P}_{\mathcal{X}_{i_z}^{\delta_t}}(y_{i_z}^{\delta_t}) = y_{i_z}^{\delta_t}$. Taking expectations of both sides on the inequality (10), we obtain that

$$\mathbb{E} \left\| x_{i_z,t+1} - y_{i_z}^{\delta_t} \right\|^2 \leq \mathbb{E} \left\| x_{i_z,t} - y_{i_z}^{\delta_t} \right\|^2 + \frac{\eta_{z,t}^2 d^2 U^2}{\delta_t^2} - 2\eta_{z,t} \mathbb{E} \left\langle x_{i_z,t} - y_{i_z}^{\delta_t}, \nabla_{i_z} \mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) \right\rangle, \quad (11)$$

where the inequality follows from Assumption 2 and the fact that $\| g_{i_z,t} \| = \left\| \frac{d}{\delta_t} \mathcal{C}_{i_z}(\hat{x}_{z,t}, x_{f,t}) u_{i_z,t} \right\| \leq \frac{dU}{\delta_t}$. Then, we have that

$$\begin{aligned} R_{i_z}(T) &= \mathbb{E} \left[\sum_{t=1}^T \mathcal{C}_{i_z}(\hat{x}_{z,t}, x_{f,t}) - \sum_{t=1}^T \mathcal{C}_{i_z}(y_{i_z}^*, \hat{x}_{-i_z,t}, x_{f,t}) \right] \\ &= \mathbb{E} \left[\sum_{t=1}^T \left(\mathcal{C}_{i_z,t}(\hat{x}_{z,t}) - \mathcal{C}_{i_z,t}(y_{i_z}^*, \hat{x}_{-i_z,t}) - \mathcal{C}_{i_z,t}(x_{z,t}) \right. \right. \\ &\quad \left. \left. + \mathcal{C}_{i_z,t}(x_{z,t}) - \mathcal{C}_{i_z,t}(y_{i_z}^*, x_{-i_z,t}) + \mathcal{C}_{i_z,t}(y_{i_z}^*, x_{-i_z,t}) \right) \right] \\ &\leq \mathbb{E} \left[\sum_{t=1}^T \left(\mathcal{C}_{i_z,t}(x_{z,t}) - \mathcal{C}_{i_z,t}(y_{i_z}^*, x_{-i_z,t}) + 2L_0 \delta_t \sqrt{N_z} \right) \right] \\ &\leq \mathbb{E} \left[\sum_{t=1}^T \left(\mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) - \mathcal{C}_{i_z,t}^{\delta_t}(y_{i_z}^*, x_{-i_z,t}) + 4L_0 \delta_t \sqrt{N_z} \right) \right] \\ &= \mathbb{E} \left[\sum_{t=1}^T \left(\mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) - \mathcal{C}_{i_z,t}^{\delta_t}(y_{i_z}^*, x_{-i_z,t}) + \mathcal{C}_{i_z,t}^{\delta_t}(y_{i_z}^*, x_{-i_z,t}) \right. \right. \\ &\quad \left. \left. - \mathcal{C}_{i_z,t}^{\delta_t}(y_{i_z}^*, x_{-i_z,t}) + 4L_0 \delta_t \sqrt{N_z} \right) \right] \\ &\leq \mathbb{E} \left[\sum_{t=1}^T \left(\mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) - \mathcal{C}_{i_z,t}^{\delta_t}(y_{i_z}^*, x_{-i_z,t}) + \frac{L_0 \delta_t D}{r} \right. \right. \\ &\quad \left. \left. + 4L_0 \delta_t \sqrt{N_z} \right) \right] \\ &\leq \sum_{t=1}^T \mathbb{E} \left\langle x_{i_z,t} - y_{i_z}^{\delta_t}, \nabla_{i_z} \mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) \right\rangle \\ &\quad + L_0 \left(\frac{D}{r} + 4\sqrt{N_z} \right) \sum_{t=1}^T \delta_t. \end{aligned} \quad (12)$$

The first inequality follows from Assumption 3 and the second inequality follows from the third item in Lemma 1. The third inequality holds because

$$|\mathcal{C}_{i_z,t}^{\delta_t}(y_{i_z}^*, x_{-i_z,t}) - \mathcal{C}_{i_z,t}^{\delta_t}(y_{i_z}^*, x_{-i_z,t})| \leq L_0 \frac{\delta_t}{r} \| y_{i_z}^* \| \leq \frac{L_0 \delta_t D}{r}.$$

The last inequality in (12) follows from the first item in Lemma 1. Define $\frac{1}{\eta_{z,0}} = 0$. Substituting the inequality (11) into the inequality (12), we have that

$$\begin{aligned} R_{i_z}(T) &\leq \sum_{t=1}^T \frac{1}{2\eta_{z,t}} \left(\mathbb{E} \left\| x_{i_z,t} - y_{i_z}^{\delta_t} \right\|^2 - \mathbb{E} \left\| x_{i_z,t+1} - y_{i_z}^{\delta_t} \right\|^2 \right) \\ &\quad + \sum_{t=1}^T \frac{\eta_{z,t} d^2 U^2}{2\delta_t^2} + L_0 \left(\frac{D}{r} + 4\sqrt{N_z} \right) \sum_{t=1}^T \delta_t \end{aligned}$$

$$\begin{aligned} &= \sum_{t=1}^T \frac{1}{2\eta_{z,t}} \left(\mathbb{E} \left\| x_{i_z,t} - y_{i_z}^{\delta_t} \right\|^2 - \mathbb{E} \left\| x_{i_z,t+1} - y_{i_z}^{\delta_{t+1}} \right\|^2 \right. \\ &\quad \left. + \mathbb{E} \left\| x_{i_z,t+1} - y_{i_z}^{\delta_{t+1}} \right\|^2 - \mathbb{E} \left\| x_{i_z,t+1} - y_{i_z}^{\delta_t} \right\|^2 \right) \\ &\quad + \sum_{t=1}^T \frac{\eta_{z,t} d^2 U^2}{2\delta_t^2} + L_0 \left(\frac{D}{r} + 4\sqrt{N_z} \right) \sum_{t=1}^T \delta_t \\ &\leq \frac{1}{2} \sum_{t=1}^T \mathbb{E} \left\| x_{i_z,t} - y_{i_z}^{\delta_t} \right\|^2 \left(\frac{1}{\eta_{z,t}} - \frac{1}{\eta_{z,t-1}} \right) \\ &\quad + \sum_{t=1}^T \frac{1}{2\eta_{z,t}} \mathbb{E} \left[\langle y_{i_z}^{\delta_{t+1}} - y_{i_z}^{\delta_t}, y_{i_z}^{\delta_t} + y_{i_z}^{\delta_{t+1}} - 2x_{i_z,t+1} \rangle \right] \\ &\quad + \sum_{t=1}^T \frac{\eta_{z,t} d^2 U^2}{2\delta_t^2} + L_0 \left(\frac{D}{r} + 4\sqrt{N_z} \right) \sum_{t=1}^T \delta_t \\ &\leq \frac{D^2}{2\eta_{z,T}} + \sum_{t=1}^T \frac{\eta_{z,t} d^2 U^2}{2\delta_t^2} + L_0 \left(\frac{D}{r} + 4\sqrt{N_z} \right) \sum_{t=1}^T \delta_t \\ &\quad + \sum_{t=1}^T \frac{1}{2\eta_{z,t}} \mathbb{E} \left\| y_{i_z}^{\delta_{t+1}} - y_{i_z}^{\delta_t} \right\| 2D \\ &\leq \frac{D^2}{2\eta_{z,T}} + \sum_{t=1}^T \frac{\eta_{z,t} d^2 U^2}{2\delta_t^2} + L_0 \left(\frac{D}{r} + 4\sqrt{N_z} \right) \sum_{t=1}^T \delta_t \\ &\quad + \sum_{t=1}^T \frac{(\delta_t - \delta_{t+1}) D^2}{r \eta_{z,t}} \\ &\leq \frac{D^2}{2\eta_{z,T}} + \sum_{t=1}^T \frac{\eta_{z,t} d^2 U^2}{2\delta_t^2} + L_0 \left(\frac{D}{r} + 4\sqrt{N_z} \right) \sum_{t=1}^T \delta_t \\ &\quad + \frac{D^2}{r} \sum_{t=1}^T \delta_t \left(\frac{1}{\eta_{z,t}} - \frac{1}{\eta_{z,t-1}} \right). \end{aligned} \quad (13)$$

In the second and last inequalities, we use again the trick that $\sum_{t=1}^T \frac{1}{a_t} (b_t - b_{t+1}) \leq \sum_{t=1}^T b_t \left(\frac{1}{a_t} - \frac{1}{a_{t-1}} \right)$ for any two sequences $\{a_t\}_{t=1}^T$ and $\{b_t\}_{t=1}^T$ with $\frac{1}{a_0} := 0$, $a_t > 0$, $b_t \geq 0$, $\forall t = [1, T]$. The fourth inequality holds since $\| y_{i_z}^{\delta_{t+1}} - y_{i_z}^{\delta_t} \| = \left\| \frac{(\delta_{t+1} - \delta_t) y_{i_z}^*}{r} \right\| \leq \frac{(\delta_t - \delta_{t+1}) D}{r}$. Substituting $\eta_{z,t} = \frac{1}{N_z^{\frac{1}{4}} t^{\frac{3}{4}}}$ and $\delta_t = \frac{1}{N_z^{\frac{1}{4}} t^{\frac{1}{4}}}$ into the inequality (13), we have

$$\begin{aligned} R_{i_z}(T) &\leq \frac{D^2 N_z^{\frac{1}{4}} T^{\frac{3}{4}}}{2} + \frac{L_0 \left(\frac{D}{r} + 4\sqrt{N_z} \right)}{N_z^{\frac{1}{4}}} \sum_{t=1}^T \frac{1}{t^{\frac{1}{4}}} + \frac{d^2 U^2}{2} \sum_{t=1}^T \frac{1}{t} \\ &\quad + \frac{D^2}{r} \sum_{t=1}^T \frac{1}{t^{\frac{1}{4}}} (t^{\frac{3}{4}} - (t-1)^{\frac{3}{4}}) \\ &\leq \frac{D^2 N_z^{\frac{1}{4}} T^{\frac{3}{4}}}{2} + \frac{L_0 \left(\frac{D}{r} + 4\sqrt{N_z} \right)}{N_z^{\frac{1}{4}}} \sum_{t=1}^T \frac{1}{t^{\frac{1}{4}}} + \frac{d^2 U^2 \ln T}{2} \\ &\quad + \frac{D^2}{r} \sum_{t=1}^T \frac{1}{t^{\frac{1}{4}}} \\ &\leq \frac{D^2 N_z^{\frac{1}{4}} T^{\frac{3}{4}}}{2} + \frac{4L_0 \left(\frac{D}{r} + 4\sqrt{N_z} \right)}{3N_z^{\frac{1}{4}}} T^{\frac{3}{4}} + \frac{d^2 U^2 \ln T}{2} \end{aligned}$$

$$+ \frac{4D^2}{3} T^{\frac{3}{4}} \\ = \mathcal{O}\left(N_Z^{\frac{1}{4}} T^{\frac{3}{4}}\right),$$

where the last inequality follows from the fact that $\sum_{t=1}^T t^{-\frac{1}{4}} \leq 1 + \int_1^T t^{-\frac{1}{4}} dt \leq \frac{4}{3} T^{\frac{3}{4}}$. The second inequality holds since $f(t) := t^{\frac{3}{4}} - (t-1)^{\frac{3}{4}} \leq 1$ for all $t \geq 1$, which is easily verified through computing the derivative of $f(t)$. The proof is complete.

C. Proof of Theorem 2

To clarify the notations, we define the following:

$$\begin{aligned} x^* &= (x_z^*, x_f^*), \quad x_z^* = (x_1^*, x_2^*, \dots, x_{N_z}^*), \\ x_f^* &= (x_{N_z+1}^*, x_{N_z+2}^*, \dots, x_N^*), \quad x_t = (x_{z,t}, x_{f,t}), \\ x_{z,t} &= (x_{1,t}, \dots, x_{N_z,t}), \quad x_{f,t} = (x_{N_z+1,t}, \dots, x_{N,t}), \\ x_t^* &= (x_{z,t}^*, x_f^*), \quad x_{z,t}^* = (1 - \frac{\delta_t}{r})x_z^*, \\ A_{i_z,t} &:= \|x_{i_z,t} - x_{i_z}^*\|^2, \quad A_{i_f,t} := \|x_{i_f,t} - x_{i_f}^*\|^2, \\ \bar{A}_{i_z,t} &:= \|x_{i_z,t} - x_{i_z,t-1}^*\|^2, \quad \hat{A}_{i_z,t} := \|x_{i_z,t} - x_{i_z,t}^*\|^2, \\ A_t &:= \sum_{i_z=1}^{N_z} A_{i_z,t} + \sum_{i_f=N_z+1}^N A_{i_f,t} = \|x_t - x^*\|^2, \\ \bar{A}_t &:= \sum_{i_z=1}^{N_z} \bar{A}_{i_z,t} + \sum_{i_f=N_z+1}^N A_{i_f,t} = \|x_t - x_{t-1}^*\|^2, \\ \hat{A}_t &:= \sum_{i_z=1}^{N_z} \hat{A}_{i_z,t} + \sum_{i_f=N_z+1}^N A_{i_f,t} = \|x_t - x_t^*\|^2. \end{aligned}$$

Recall that we use i_f to index an agent with access to first-order gradient information and i_z to index an agent using zeroth-order oracles. We first focus on the evolution of $A_{i_f,t}$, i.e., the error dynamics of the agents with first-order gradient information. From the update rule (6), we have that

$$\begin{aligned} A_{i_f,t+1} &= \|x_{i_f,t+1} - x_{i_f}^*\|^2 \\ &= \left\| \mathcal{P}_{\mathcal{X}_{i_f}} \left(x_{i_f,t} - \eta_{f,t} \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}) \right) - x_{i_f}^* \right\|^2 \\ &\leq \|x_{i_f,t} - \eta_{f,t} \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}) - x_{i_f}^*\|^2 \\ &\leq \|x_{i_f,t} - x_{i_f}^*\|^2 + \eta_{f,t}^2 \|\nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t})\|^2 \\ &\quad - 2\eta_{f,t} \left\langle \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}), x_{i_f,t} - x_{i_f}^* \right\rangle \\ &\leq A_{i_f,t} - 2\eta_{f,t} \left\langle \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}), x_{i_f,t} - x_{i_f}^* \right\rangle \\ &\quad + \eta_{f,t}^2 L_0^2, \end{aligned} \tag{14}$$

where the first inequality follows since $\mathcal{P}_{\mathcal{X}_{i_f}}(x_{i_f}^*) = x_{i_f}^*$ and the projection operator is non-expansive. Since x^* is a Nash equilibrium of the convex game (1), we have that $\langle \nabla_{i_f} \mathcal{C}_{i_f}(x^*), x_{i_f,t} - x_{i_f}^* \rangle \geq 0$, $i = 1, \dots, N$. Combining this Nash equilibrium condition with the inequality (14), we have that

$$A_{i_f,t+1}$$

$$\begin{aligned} &\leq A_{i_f,t} + \eta_{f,t}^2 L_0^2 + 2\eta_{f,t} \left\langle \nabla_{i_f} \mathcal{C}_{i_f}(x^*), x_{i_f,t} - x_{i_f}^* \right\rangle \\ &\quad - 2\eta_{f,t} \left\langle \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}), x_{i_f,t} - x_{i_f}^* \right\rangle \\ &= A_{i_f,t} - 2\eta_{f,t} \left\langle \nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}) - \nabla_{i_f} \mathcal{C}_{i_f}(x_t), \right. \\ &\quad \left. x_{i_f,t} - x_{i_f}^* \right\rangle + \eta_{f,t}^2 L_0^2 \\ &\quad - 2\eta_{f,t} \left\langle \nabla_{i_f} \mathcal{C}_{i_f}(x_t) - \nabla_{i_f} \mathcal{C}_{i_f}(x_t^*), x_{i_f,t} - x_{i_f}^* \right\rangle \\ &\quad - 2\eta_{f,t} \left\langle \nabla_{i_f} \mathcal{C}_{i_f}(x_t^*) - \nabla_{i_f} \mathcal{C}_{i_f}(x^*), x_{i_f,t} - x_{i_f}^* \right\rangle \\ &\leq A_{i_f,t} + \eta_{f,t}^2 L_0^2 + 2\eta_{f,t} L_1 \delta_t \frac{2\eta_{f,t} L_1 \delta_t D^2 \sqrt{N_z}}{r} \\ &\quad - 2\eta_{f,t} \left\langle \nabla_{i_f} \mathcal{C}_{i_f}(x_t) - \nabla_{i_f} \mathcal{C}_{i_f}(x_t^*), x_{i_f,t} - x_{i_f}^* \right\rangle, \end{aligned} \tag{15}$$

where the last inequality follows from Assumptions 3, 4, and the facts that $\|\nabla_{i_f} \mathcal{C}_{i_f}(x_{f,t}, \hat{x}_{z,t}) - \nabla_{i_f} \mathcal{C}_{i_f}(x_t)\| \leq L_1 \|\hat{x}_{z,t} - x_{z,t}\| \leq L_1 \delta_t \sqrt{N_z}$, $\|\nabla_{i_f} \mathcal{C}_{i_f}(x_t^*) - \nabla_{i_f} \mathcal{C}_{i_f}(x^*)\| \leq L_1 \|x_t^* - x^*\| = L_1 \|x_{z,t}^* - x_z^*\| \leq \frac{L_1 \delta_t D \sqrt{N_z}}{r}$.

Next we turn our attention on the evolution of $\bar{A}_{i_z,t}$. Similarly, from the update equation (4), we obtain that

$$\begin{aligned} \bar{A}_{i_z,t+1} &= \|x_{i_z,t+1} - x_{i_z,t}^*\|^2 \\ &= \left\| \mathcal{P}_{\mathcal{X}_{i_z}^{\delta_t}} \left(x_{i_z,t} - \eta_{z,t} g_{i_z,t} \right) - x_{i_z,t}^* \right\|^2 \\ &\leq \|x_{i_z,t} - \eta_{z,t} g_{i_z,t} - x_{i_z,t}^*\|^2 \\ &= \hat{A}_{i_z,t} + \eta_{z,t}^2 \|g_{i_z,t}\|^2 - 2\eta_{z,t} \langle g_{i_z,t}, x_{i_z,t} - x_{i_z,t}^* \rangle \\ &\leq \hat{A}_{i_z,t} + \frac{\eta_{z,t}^2 d^2 U^2}{\delta_t^2} - 2\eta_{z,t} \langle g_{i_z,t}, x_{i_z,t} - x_{i_z,t}^* \rangle, \end{aligned} \tag{16}$$

where the first inequality follows from the fact that $\mathcal{P}_{\mathcal{X}_{i_z}^{\delta_t}}(x_{i_z,t}^*) = x_{i_z,t}^*$ and the projection operator is non-expansive. Taking expectation of both sides of the above inequality (16), we have that

$$\begin{aligned} &\mathbb{E}[\bar{A}_{i_z,t+1}] \\ &\leq \mathbb{E}[\hat{A}_{i_z,t}] - 2\eta_{z,t} \mathbb{E} \left\langle \nabla_{i_z} \mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}), x_{i_z,t} - x_{i_z,t}^* \right\rangle \\ &\quad + \frac{\eta_{z,t}^2 d^2 U^2}{\delta_t^2} \\ &\leq \mathbb{E}[\hat{A}_{i_z,t}] + \frac{\eta_{z,t}^2 d^2 U^2}{\delta_t^2} \\ &\quad - 2\eta_{z,t} \mathbb{E} \left\langle \nabla_{i_z} \mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) - \nabla_{i_z} \mathcal{C}_{i_z}(x_t), x_{i_z,t} - x_{i_z,t}^* \right\rangle \\ &\quad - 2\eta_{z,t} \mathbb{E} \left\langle \nabla_{i_z} \mathcal{C}_{i_z}(x_t) - \nabla_{i_z} \mathcal{C}_{i_z}(x_t^*), x_{i_z,t} - x_{i_z,t}^* \right\rangle \\ &\quad - 2\eta_{z,t} \mathbb{E} \left\langle \nabla_{i_z} \mathcal{C}_{i_z}(x_t^*) - \nabla_{i_z} \mathcal{C}_{i_z}(x^*), x_{i_z,t} - x_{i_z,t}^* \right\rangle \\ &\quad - 2\eta_{z,t} \mathbb{E} \left\langle \nabla_{i_z} \mathcal{C}_{i_z}(x^*), (x_{i_z,t} - x_{i_z,t}^*) - (x_{i_z,t} - x_{i_z}^*) \right\rangle \\ &\leq \mathbb{E}[\hat{A}_{i_z,t}] + \frac{\eta_{z,t}^2 d^2 U^2}{\delta_t^2} + \frac{2\eta_{z,t} L_1 \delta_t D^2}{r} \\ &\quad - 2\eta_{z,t} \mathbb{E} \left\langle \nabla_{i_z} \mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) - \nabla_{i_z} \mathcal{C}_{i_z}(x_t), x_{i_z,t} - x_{i_z,t}^* \right\rangle \\ &\quad - 2\eta_{z,t} \mathbb{E} \left\langle \nabla_{i_z} \mathcal{C}_{i_z}(x_t) - \nabla_{i_z} \mathcal{C}_{i_z}(x_t^*), x_{i_z,t} - x_{i_z,t}^* \right\rangle \\ &\quad - 2\eta_{z,t} \mathbb{E} \left\langle \nabla_{i_z} \mathcal{C}_{i_z}(x_t^*), x_{i_z,t}^* - x_{i_z,t}^* \right\rangle \\ &\leq \mathbb{E}[\hat{A}_{i_z,t}] + \frac{\eta_{z,t}^2 d^2 U^2}{\delta_t^2} + \frac{2\eta_{z,t} L_1 \delta_t D^2}{r} + \frac{2\eta_{z,t} L_0 \delta_t D}{r} \end{aligned}$$

$$\begin{aligned}
& -2\eta_{z,t}\mathbb{E}\left\langle \nabla_{i_z}\mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) - \nabla_{i_z}\mathcal{C}_{i_z}(x_t), x_{i_z,t} - x_{i_z,t}^* \right\rangle \\
& -2\eta_{z,t}\mathbb{E}\left\langle \nabla_{i_z}\mathcal{C}_{i_z}(x_t) - \nabla_{i_z}\mathcal{C}_{i_z}(x_t^*), x_{i_z,t} - x_{i_z,t}^* \right\rangle, \quad (17)
\end{aligned}$$

where the first inequality follows from Lemma 1, the second inequality holds due to the Nash equilibrium first-order optimality condition, and the third inequality follows since $\|\nabla_{i_z}\mathcal{C}_{i_z}(x_t^*) - \nabla_{i_z}\mathcal{C}_{i_z}(x^*)\| \leq L_1\|x_t^* - x^*\| \leq \frac{L_1\delta_t D}{r}$. The last inequality follows from Assumption 3 and the fact that $\|x_{i_z}^* - x_{i_z,t}^*\| \leq \frac{\delta_t D}{r}$.

Now we are in a position to bound the term $\|\nabla_{i_z}\mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) - \nabla_{i_z}\mathcal{C}_{i_z}(x_t)\|$. Since \mathcal{C}_{i_z} is bounded and continuous for all $i_z \in \mathcal{N}_z$, by Lebesgue's dominated convergence theorem (Chapter 4 in [45]), the order of integration and differentiation can be interchanged. Then, it follows that

$$\begin{aligned}
& \nabla_{i_z}\mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) \\
& = \nabla_{i_z}\mathbb{E}[C_{i_z,t}(x_{i_z,t} + \delta_t w_{i_z,t}, x_{-i_z,t} + \delta_t u_{-i_z,t})] \\
& = \mathbb{E}[\nabla_{i_z}C_{i_z,t}(x_{i_z,t} + \delta_t w_{i_z,t}, x_{-i_z,t} + \delta_t u_{-i_z,t})] \\
& = \mathbb{E}[\nabla_{i_z}C_{i_z}(x_{i_z,t} + \delta_t w_{i_z,t}, x_{-i_z,t} + \delta_t u_{-i_z,t}, x_{f,t})],
\end{aligned}$$

where the expectations in the above inequality are taken w.r.t $w_{i_z,t} \sim \mathbb{B}$ and $u_{-i_z,t} \sim \mathbb{S}_{-i_z}$. Then, we have that

$$\begin{aligned}
& \|\nabla_{i_z}\mathcal{C}_{i_z,t}^{\delta_t}(x_{z,t}) - \nabla_{i_z}\mathcal{C}_{i_z}(x_t)\| \\
& \leq L_1\delta_t\|(w_{i_z,t}, u_{-i_z,t})\| \leq L_1\delta_t\sqrt{N_z}. \quad (18)
\end{aligned}$$

Substituting (18) into (17), we obtain that

$$\begin{aligned}
& \mathbb{E}[\bar{A}_{i_z,t+1}] \\
& \leq \mathbb{E}[\hat{A}_{i_z,t}] + \frac{\eta_{z,t}^2 d^2 U^2}{\delta_t^2} + \frac{2\eta_{z,t} L_1 \delta_t D^2}{r} + \frac{2\eta_{z,t} L_0 \delta_t D}{r} \\
& \quad - 2\eta_{z,t}\mathbb{E}\left\langle \nabla_{i_z}\mathcal{C}_{i_z}(x_t) - \nabla_{i_z}\mathcal{C}_{i_z}(x_t^*), x_{i_z,t} - x_{i_z,t}^* \right\rangle \\
& \quad + 2\eta_{z,t} L_1 \delta_t \sqrt{N_z} D. \quad (19)
\end{aligned}$$

Define $\delta_0 = 0$ and $x_{i_z,0}^* = x_{i_z}^*$. From the definition of $\hat{A}_{i_z,t}$ and $\bar{A}_{i_z,t}$, we have

$$\begin{aligned}
\hat{A}_{i_z,t} & = \|x_{i_z,t} - x_{i_z,t-1}^* + x_{i_z,t-1}^* - x_{i_z,t}^*\|^2 \\
& = \bar{A}_{i_z,t} + \|x_{i_z,t-1}^* - x_{i_z,t}^*\|^2 \\
& \quad + 2\langle x_{i_z,t} - x_{i_z,t-1}^*, x_{i_z,t-1}^* - x_{i_z,t}^* \rangle \\
& \leq \bar{A}_{i_z,t} + \frac{|\delta_{t-1} - \delta_t|^2 D^2}{r^2} + \frac{2|\delta_{t-1} - \delta_t| D^2}{r}, \quad (20)
\end{aligned}$$

for all $t \geq 1$. The inequality (20) holds at $t = 1$ since $\|x_{i_z,0}^* - x_{i_z,1}^*\| = \frac{\delta_1}{r}x_{i_z}^* \leq \frac{|\delta_0 - \delta_1| D}{r}$. Substituting (20) into (19), we have

$$\begin{aligned}
& \mathbb{E}[\bar{A}_{i_z,t+1}] \\
& \leq \mathbb{E}[\bar{A}_{i_z,t}] + \frac{|\delta_{t-1} - \delta_t|^2 D^2}{r^2} + \frac{2|\delta_{t-1} - \delta_t| D^2}{r} \\
& \quad - 2\eta_{z,t}\mathbb{E}\left\langle \nabla_{i_z}\mathcal{C}_{i_z}(x_t) - \nabla_{i_z}\mathcal{C}_{i_z}(x_t^*), x_{i_z,t} - x_{i_z,t}^* \right\rangle \\
& \quad + \frac{\eta_{z,t}^2 d^2 U^2}{\delta_t^2} + \frac{2\eta_{z,t} L_1 \delta_t D^2}{r} + \frac{2\eta_{z,t} L_0 \delta_t D}{r} \\
& \quad + 2\eta_{z,t} L_1 \delta_t \sqrt{N_z} D. \quad (21)
\end{aligned}$$

Summing (15) over $i_f = N_z + 1, \dots, N$ and (21) over $i_z = 1, \dots, N_z$ altogether, and setting $\eta_{z,t} = \eta_{f,t} = \eta_t$ we have that

$$\begin{aligned}
\mathbb{E}[\bar{A}_{t+1}] & = \sum_{i_z=1}^{N_z} \mathbb{E}[\bar{A}_{i_z,t+1}] + \sum_{i_f=N_z+1}^N \mathbb{E}[A_{i_f,t+1}] \\
& \leq \sum_{i_z=1}^{N_z} \left(\mathbb{E}[\bar{A}_{i_z,t}] + \frac{|\delta_{t-1} - \delta_t|^2 D^2}{r^2} + \frac{2|\delta_{t-1} - \delta_t| D^2}{r} \right. \\
& \quad \left. - 2\eta_t\mathbb{E}\left\langle \nabla_{i_z}\mathcal{C}_{i_z}(x_t) - \nabla_{i_z}\mathcal{C}_{i_z}(x_t^*), x_{i_z,t} - x_{i_z,t}^* \right\rangle \right. \\
& \quad \left. + \frac{\eta_t^2 d^2 U^2}{\delta_t^2} + \frac{2\eta_t L_1 \delta_t D^2}{r} + \frac{2\eta_t L_0 \delta_t D}{r} \right. \\
& \quad \left. + 2\eta_t L_1 \delta_t \sqrt{N_z} D \right) \\
& \quad + \sum_{i_f=N_z+1}^N \left(\mathbb{E}[A_{i_f,t}] + \eta_t^2 L_0^2 + 2\eta_t L_1 \delta_t \sqrt{N_z} D \right. \\
& \quad \left. - 2\eta_t\mathbb{E}\left\langle \nabla_{i_f}\mathcal{C}_{i_f}(x_t) - \nabla_{i_f}\mathcal{C}_{i_f}(x_t^*), x_{i_f,t} - x_{i_f,t}^* \right\rangle \right. \\
& \quad \left. + \frac{2\eta_t L_1 \delta_t D^2 \sqrt{N_z}}{r} \right) \\
& = \mathbb{E}[\bar{A}_t] + N_z \frac{|\delta_{t-1} - \delta_t|^2 D^2}{r^2} + \frac{2N_z|\delta_{t-1} - \delta_t| D^2}{r} \\
& \quad - 2\eta_t \sum_{i=1}^N \mathbb{E}\left\langle \nabla_i\mathcal{C}_i(x_t) - \nabla_i\mathcal{C}_i(x_t^*), x_{i,t} - x_{i,t}^* \right\rangle \\
& \quad + N_z \left(\frac{\eta_t^2 d^2 U^2}{\delta_t^2} + \frac{2\eta_t L_1 \delta_t D^2}{r} + \frac{2\eta_t L_0 \delta_t D}{r} \right. \\
& \quad \left. + 2\eta_t L_1 \delta_t \sqrt{N_z} D \right) \\
& \quad + (N - N_z) \left(\eta_t^2 L_0^2 + 2\eta_t L_1 \delta_t \sqrt{N_z} D + \frac{2\eta_t L_1 \delta_t D^2 \sqrt{N_z}}{r} \right) \\
& \leq (1 - 2m\eta_t)\mathbb{E}[\bar{A}_t] + \frac{2N_z|\delta_{t-1} - \delta_t| D^2}{r} \\
& \quad + \frac{N_z|\delta_{t-1} - \delta_t|^2 D^2}{r^2} + \frac{\eta_t^2 N_z d^2 U^2}{\delta_t^2} + (N - N_z)\eta_t^2 L_0^2 \\
& \quad + 2\eta_t \delta_t N L_1 \left(\frac{D^2 \sqrt{N_z}}{r} + \sqrt{N_z} D \right) + 2\eta_t \delta_t N_z L_0 \frac{D}{r}, \quad (22)
\end{aligned}$$

where the last inequality follows from the strong monotonicity condition (2). Recalling $\delta_t = \frac{N_z^{\frac{1}{6}}}{N^{\frac{1}{3}} t^{\frac{1}{3}}}$, it is easy to verify that

$$|\delta_{t-1} - \delta_t| \leq \frac{N_z^{\frac{1}{6}}}{N^{\frac{1}{3}} t^{\frac{4}{3}}}, \quad (23)$$

for all $t \geq 1$. The inequality (23) holds at $t = 1$ since $|\delta_0 - \delta_1| = \delta_1$. Substituting $\eta_t = \frac{1}{mt}$, $\delta_t = \frac{N_z^{\frac{1}{6}}}{N^{\frac{1}{3}} t^{\frac{1}{3}}}$ and (23) into the inequality (22), we have that

$$\begin{aligned}
& \mathbb{E}[\bar{A}_{t+1}] \\
& \leq (1 - \frac{2}{t})\mathbb{E}[\bar{A}_t] + \frac{2N_z^{\frac{7}{6}} D^2}{r N^{\frac{1}{3}} t^{\frac{4}{3}}} + \frac{N_z^{\frac{4}{3}} D^2}{r^2 N^{\frac{2}{3}} t^{\frac{8}{3}}} + \frac{d^2 U^2 N_z^{\frac{2}{3}} N^{\frac{2}{3}}}{m^2 t^{\frac{4}{3}}} \\
& \quad + \frac{2N_z^{\frac{1}{6}} N^{\frac{2}{3}} L_1}{m t^{\frac{4}{3}}} \left(\frac{D^2 \sqrt{N_z}}{r} + \sqrt{N_z} D \right) + (N - N_z) \frac{L_0^2}{m^2 t^2} \\
& \quad + \frac{2N_z^{\frac{7}{6}} L_0 D}{m r N^{\frac{1}{3}} t^{\frac{4}{3}}}
\end{aligned}$$

$$\begin{aligned}
&\leq (1 - \frac{2}{t})\mathbb{E}[\bar{A}_t] + \frac{2N_z^{\frac{2}{3}}N^{\frac{1}{6}}D^2}{rt^{\frac{4}{3}}} + \frac{N_z^{\frac{2}{3}}D^2}{r^2t^{\frac{8}{3}}} + \frac{d^2U^2N_z^{\frac{2}{3}}N^{\frac{2}{3}}}{m^2t^{\frac{4}{3}}} \\
&\quad + \frac{2N_z^{\frac{1}{6}}N^{\frac{2}{3}}}{mt^{\frac{4}{3}}} \left(\frac{L_1D^2\sqrt{N_z}}{r} + \sqrt{N_z}L_1D + \frac{L_0D}{r} \right) \\
&\quad + (N - N_z)\frac{L_0^2}{m^2t^2} \\
&\leq (1 - \frac{2}{t})\mathbb{E}[\bar{A}_t] + \frac{N_z^{\frac{2}{3}}N^{\frac{2}{3}}}{t^{\frac{4}{3}}} \left(\frac{2D^2}{r} + \frac{D^2}{r^2} + \frac{d^2U^2}{m^2} \right) \\
&\quad + \frac{2N_z^{\frac{1}{6}}N^{\frac{2}{3}}}{mt^{\frac{4}{3}}} \left(\frac{L_1D^2\sqrt{N_z}}{r} + \sqrt{N_z}L_1D + \frac{L_0D}{r} \right) \\
&\quad + (N - N_z)\frac{L_0^2}{m^2t^2}, \tag{24}
\end{aligned}$$

where the second inequality follows since $N_z \leq N$.

In the following, we use the standard mathematical inductive method to prove that the following inequality holds for all $t \geq 1$:

$$\begin{aligned}
\mathbb{E}[\bar{A}_t] &\leq \frac{N_z^{\frac{2}{3}}N^{\frac{2}{3}}}{t^{\frac{1}{3}}} \left(\frac{2D^2}{r} + \frac{D^2}{r^2} + \frac{d^2U^2}{m^2} \right) + (N - N_z)\frac{L_0^2}{m^2t} \\
&\quad + \frac{2N_z^{\frac{1}{6}}N^{\frac{2}{3}}}{mt^{\frac{1}{3}}} \left(\frac{L_1D^2\sqrt{N_z}}{r} + \sqrt{N_z}L_1D + \frac{L_0D}{r} \right) + \frac{D^2}{t}. \tag{25}
\end{aligned}$$

Obviously, the inequality (25) holds at the initial time step $t = 1$ since $\bar{A}_1 \leq D^2$. In what follows, we show that, if the statement (25) holds at time step t , it must also hold for the next time step $t + 1$. To do so, we combine the inequalities (24) and (25), yielding that

$$\begin{aligned}
&\mathbb{E}[\bar{A}_{t+1}] \\
&\leq (1 - \frac{2}{t}) \left(\frac{N_z^{\frac{2}{3}}N^{\frac{2}{3}}}{t^{\frac{1}{3}}} \left(\frac{2D^2}{r} + \frac{D^2}{r^2} + \frac{d^2U^2}{m^2} \right) + (N - N_z)\frac{L_0^2}{m^2t} \right. \\
&\quad \left. + \frac{2N_z^{\frac{1}{6}}N^{\frac{2}{3}}}{mt^{\frac{1}{3}}} \left(\frac{L_1D^2\sqrt{N_z}}{r} + \sqrt{N_z}L_1D + \frac{L_0D}{r} \right) + \frac{D^2}{t} \right) \\
&\quad + \frac{N_z^{\frac{2}{3}}N^{\frac{2}{3}}}{t^{\frac{4}{3}}} \left(\frac{2D^2}{r} + \frac{D^2}{r^2} + \frac{d^2U^2}{m^2} \right) + (N - N_z)\frac{L_0^2}{m^2t^2} \\
&\quad + \frac{2N_z^{\frac{1}{6}}N^{\frac{2}{3}}L_1}{mt^{\frac{4}{3}}} \left(\frac{L_1D^2\sqrt{N_z}}{r} + \sqrt{N_z}L_1D + \frac{L_0D}{r} \right) \\
&= (1 - \frac{1}{t}) \left(\frac{N_z^{\frac{2}{3}}N^{\frac{2}{3}}}{t^{\frac{1}{3}}} \left(\frac{2D^2}{r} + \frac{D^2}{r^2} + \frac{d^2U^2}{m^2} \right) + (N - N_z)\frac{L_0^2}{m^2t} \right. \\
&\quad \left. + \frac{2N_z^{\frac{1}{6}}N^{\frac{2}{3}}}{mt^{\frac{1}{3}}} \left(\frac{L_1D^2\sqrt{N_z}}{r} + \sqrt{N_z}L_1D + \frac{L_0D}{r} \right) \right) \\
&\quad + (1 - \frac{2}{t})\frac{D^2}{t} \\
&\leq \frac{N_z^{\frac{2}{3}}N^{\frac{2}{3}}}{(t+1)^{\frac{1}{3}}} \left(\frac{2D^2}{r} + \frac{D^2}{r^2} + \frac{d^2U^2}{m^2} \right) + (N - N_z)\frac{L_0^2}{m^2(t+1)} \\
&\quad + \frac{2N_z^{\frac{1}{6}}N^{\frac{2}{3}}}{m(t+1)^{\frac{1}{3}}} \left(\frac{L_1D^2\sqrt{N_z}}{r} + \sqrt{N_z}L_1D + \frac{L_0D}{r} \right) + \frac{D^2}{t+1}, \tag{26}
\end{aligned}$$

where the last inequality follows since $(1 - \frac{1}{t})\frac{1}{t^{\frac{1}{3}}} \leq \frac{t}{t+1}\frac{1}{t^{\frac{1}{3}}} \leq \frac{1}{(t+1)^{\frac{1}{3}}}$ and $(1 - \frac{2}{t})\frac{1}{t} \leq (1 - \frac{1}{t})\frac{1}{t} \leq \frac{1}{t+1}$. The inequality (26)

shows that the statement (25) also holds at $t + 1$ given that it holds at t , and thus we can conclude that the statement (25) holds for all $t \geq 1$. Finally, we have

$$\begin{aligned}
&\mathbb{E}[A_T] = \mathbb{E}[\|x_T - x^*\|^2] \\
&= \mathbb{E}[\|x_T - x_{T-1}^* + x_{T-1}^* - x^*\|^2] \\
&\leq \mathbb{E}[2\bar{A}_T + 2\|x_{T-1}^* - x^*\|^2] \\
&\leq \frac{2N_z^{\frac{2}{3}}N^{\frac{2}{3}}}{T^{\frac{1}{3}}} \left(\frac{2D^2}{r} + \frac{D^2}{r^2} + \frac{d^2U^2}{m^2} \right) + (N - N_z)\frac{2L_0^2}{m^2T} \\
&\quad + \frac{4N_z^{\frac{1}{6}}N^{\frac{2}{3}}L_1}{mT^{\frac{1}{3}}} \left(\frac{L_1D^2\sqrt{N_z}}{r} + \sqrt{N_z}L_1D + \frac{L_0D}{r} \right) + \frac{2D^2}{T} \\
&\quad + \frac{2N_z^{\frac{1}{3}}D^2}{N^{\frac{2}{3}}(T-1)^{\frac{2}{3}}r^2} \\
&= \mathcal{O}\left(N_z^{\frac{2}{3}}N^{\frac{2}{3}}T^{-\frac{1}{3}} + (N - N_z)T^{-1}\right),
\end{aligned}$$

where the first inequality follows since $(a + b)^2 \leq 2a^2 + 2b^2$. The proof is complete.