Adaptive No-Regret Algorithms in Multi-Agent Games: Optimal Last-Iterate Convergence with Gradient Feedback

Chris Junchi Li$^\diamond$

Department of Electrical Engineering and Computer Sciences$^\diamond$
University of California, Berkeley

August 29, 2024

Abstract

In this work, we explore the development of adaptive no-regret algorithms for multi-agent learning in both strongly monotone and exp-concave game settings. We propose a unified framework that achieves optimal last-iterate convergence rates, extending the utility of online learning algorithms in complex, non-stationary environments. Our approach leverages gradient feedback mechanisms to adaptively optimize decision-making processes across diverse game-theoretic scenarios, ensuring double optimality in both single-agent and multi-agent contexts. Specifically, we introduce novel, feasible variants of online gradient descent (OGD) and online Newton step (ONS) algorithms that do not require prior knowledge of problem parameters. These algorithms demonstrate robust performance, achieving near-optimal regret bounds in single-agent settings and optimal convergence to Nash equilibria in multi-agent games. Our contributions provide theoretical insights and practical algorithms that bridge the gap between transient performance and long-term equilibrium outcomes in decentralized, adversarial, and stochastic environments.

Keywords: Online Learning with Gradient Feedback; No-Regret Algorithms; Adaptive Algorithms; Strongly Monotone Games; Exp-Concave Functions; Last-Iterate Convergence

1 Introduction

The problem of online learning with gradient feedback has garnered significant attention due to its wide applicability in adaptive decision-making processes across various fields, including machine learning, game theory, and optimization. At the core of this problem is an agent that interacts with an environment by selecting actions based on observed historical data and receiving feedback in the form of gradients from cost functions. This feedback mechanism allows the agent to iteratively refine its actions, aiming to minimize a cumulative cost over time. The generality of this framework is particularly attractive as it imposes no statistical regularity assumptions on the cost functions, allowing for the accommodation of non-stationary or adversarial environments.

A key metric for evaluating the performance of online learning algorithms is **regret**, defined as the difference between the cumulative cost incurred by the algorithm and that incurred by the best fixed action in hindsight. Algorithms that achieve sublinear regret are termed **no-regret** algorithms, as they guarantee that the average regret per period approaches zero as the time horizon increases. Among the most well-known no-regret algorithms are **Online Gradient Descent** (OGD) and **Online Newton Step** (ONS), both of which have been shown to achieve minimax optimal regret bounds in various settings. These algorithms are particularly valued for their simplicity, theoretical elegance, and broad applicability.

The robustness of no-regret learning algorithms in single-agent settings naturally leads to their consideration in multi-agent environments, where each agent's decision-making process affects and

is affected by the decisions of other agents. In such game-theoretic settings, the convergence of joint actions to Nash equilibria—where no agent has an incentive to unilaterally deviate from their strategy—is a critical question. Recent advances in the field have focused on understanding the last-iterate convergence properties of no-regret algorithms, a problem recognized as considerably more difficult than analyzing the convergence of time-averaged actions.

However, many existing theoretical guarantees for no-regret algorithms, particularly in multi-agent settings, require prior knowledge of problem parameters, such as the strong convexity or strong monotonicity of the cost functions. This dependency on specific parameter knowledge presents a significant obstacle to the practical implementation of these algorithms, especially in decentralized environments where such information may not be readily available.

To address this issue, we investigate the design of **adaptive no-regret learning algorithms** that are feasible, requiring no prior knowledge of problem parameters while still achieving optimal convergence guarantees. Our approach is to develop variants of OGD and ONS that are doubly optimal—achieving both optimal regret in single-agent settings and optimal last-iterate convergence in multi-agent settings.

1.1 An Online Learning Framework

Formally, the problem of online learning with gradient feedback [Blum(1998), Shalev-Shwartz(2012), Hazan(2016)] can be described in its essential form by the following adaptive decision-making process:

(i) An agent interfaces with the environment by choosing an **action** $x^t \in \mathcal{X}$ at period t where $\mathcal{X} \subseteq \mathbb{R}^d$ is a convex and compact set. For example, the action is a route in a traffic network or an output quantity in an oligopoly. The agent chooses x^t through an online learning algorithm, which makes its choice adaptively based on observable historical information.

(ii) The environment then returns a cost function $f_t(\cdot)$ so that the agent incurs cost $f_t(x^t)$ and receives $\nabla f_t(x^t)$ as feedback. The process then moves to the next period $t + 1$ and repeats.

One appealing feature of the online learning framework is that one need not impose any statistical regularity assumption: $f_1(\cdot), \ldots, f_T(\cdot)$ can be an arbitrary fixed sequence of cost functions, hence accommodating a non-stationary or even adversarial environment. Further, the cost function $f_t(\cdot)$ does need not to be known by the agent (and indeed in many applications it is not known); only the gradient feedback is needed. In this general framework, the standard metric for judging the performance of an online learning algorithm is **regret** [Blum and Mansour(2007)]—the difference between the total cost incurred by the algorithm up to T and the total cost incurred by the best fixed action in hindsight:

$$\text{Regret}(T) = \sum_{t=1}^{T} f_t(x^t) - \min_{x \in \mathcal{X}} \left\{ \sum_{t=1}^{T} f_t(x) \right\} \tag{1}$$

If the average regret (obtained by dividing by T) goes to zero, then the algorithm is referred to as a "no-regret learning algorithm."

A canonical example of a no-regret learning algorithm is online gradient descent (OGD), where the agent takes a gradient step (given the current action x_t) and performs a projection step onto $\mathcal{X}$ to obtain the next action x_{t+1}. Analyzing OGD in the standard setting where the cost function f_t is convex, [Zinkevich(2003)] proved that the algorithm with the stepsize rule $\eta_t = \frac{1}{\sqrt{T}}$ achieves

a regret bound of $\Theta(\sqrt{T})$, which is known to be minimax optimal [Hazan(2016)]. If f_t is further assumed to be μ-strongly convex, [Hazan et al.(2007)] proved that the algorithm with the stepsize rule $\eta_t = \frac{1}{\mu t}$ achieves a regret bound of $\Theta(\log T)$; again this rate is minimax optimal. In fact, the $\Theta(\log T)$ regret bound is achievable even for a class of cost functions that are more general than strongly convex cost functions: if f_t is exp-concave—a class of functions properly subsuming strongly convex functions that has found widespread applications (see Section 2 for a detailed discussion)—then the online Newton step (ONS) [Hazan et al.(2007)] achieves the minimax optimal regret bound of $\Theta(d \log T)$. In summary, OGD and ONS provide two of the most well-known optimal no-regret learning algorithms in the online learning/online convex optimization literature, with their algorithmic simplicity and theoretical elegance being matched by their broad applicability in practice.

Given their appealing theoretical and practical properties, the aforementioned no-regret learning algorithms such as OGD and ONS have also served as natural candidates for game-theoretic learning. In this setting, each agent makes online decisions in an environment consisting of other agents, each of whom are making adaptive decisions. Even if the game is fixed, the fact that all agents are adjusting their strategies simultaneously means that the stream of costs seen by any single agent is non-stationary and complex. It might be hoped that no-regret learning—given its robustness to assumptions—can cope with the complexity of the multi-agent setting. An extensive literature has shown that this hope is borne out—under no-regret dynamics, the time average of the joint actions converges to equilibria in various classes of games [Cesa-Bianchi and Lugosi(2006), Shoham and Leyton-Brown(2008), Viossat and Zapechelnyuk(2013), Bloembergen et al.(2015), Monnot and Piliou

1.2 First Question

Further progress in the online learning and optimization literature has yielded more refined statements regarding convergence; in particular, last-iterate rate guarantees have been established for many no-regret learning algorithms. This leads to an analogous question for the game-theoretic setting: ***If each agent employs a no-regret learning algorithm to minimize its own regret, can the joint action converge to a Nash equilibrium at an optimal last-iterate rate?***

An affirmative answer to this question[1] would establish a remarkable "double optimality" for an online algorithm: while the algorithm itself is only designed for maximizing the (transient) performance for a finite time horizon T, the resulting long-run performance would also be optimal for all agents, in the sense that any agent, by unilaterally deviating from the action suggested by the algorithm, could only incur higher cost (by the definition of a Nash equilibrium). Without this double optimality, an agent could incur "regret" in the long term, since it may do better by not following such an algorithm.

To address this question, it is necessary to study the last-iterate convergence of algorithms (i.e., the convergence of the ***actual*** joint action), a problem that has been recognized to be considerably more difficult than the characterization of convergence of time averages [Krichene et al.(2015), Balandat et al.(2016), Zhou et al.(2017), Zhou et al.(2018), Mertikopoulos et al.(2019), Mertikopoulos and Zhou(2 For instance, as pointed out by [Mertikopoulos et al.(2018)], there are situations where the time

---

[1]For instance, if each vehicle in a traffic network employs an optimal no-regret learning algorithm (such as OGD) to choose their route adaptively over a certain horizon, would the system converge to a stable traffic distribution or devolve to perpetual congestion as users ping-pong between different routes? If it does converge to a stable distribution, is it Nash? Because if not, each agent is being irrational—by ***not*** following the no-regret learning algorithm, agents can do individually do better.

average of the iterate converges to a Nash equilibrium, but the last iterate cycles around the equilibrium point. Progress has been made on this problem during the past five years, but much of it only provides qualitative or asymptotic results, with only a few quantitative (finite-time, last-iterate convergence guarantees) results obtained, for games having special structures or using metrics other than the distance to Nash equilibria (see Section A). In particular, [Zhou et al.(2021)] established the last-iterate convergence of multi-agent OGD to the unique Nash equilibrium in strongly monotone games[2] at an optimal rate of $\Theta(\frac{1}{T})$. This optimal convergence rate continues to hold even when the gradient feedback is corrupted by certain forms of noise, in which case we have $E[\|x_t - x^\star\|_2^2] = \Theta(\frac{1}{T})$, where x_t is the (random) joint action of all agents and $x^\star$ is the unique Nash equilibrium.[3] Thus, OGD is doubly optimal when a strong convexity structure is available (i.e., the game is strongly monotone or the cost functions are strongly convex from a single-agent perspective), giving a compelling argument for its adoption in single-agent and multi-agent settings.

1.3 Second Question

The above argument, however, suffers from a key, subtle weakness: the theoretical guarantees for OGD require choices of step size, in both single-agent [Hazan et al.(2007)] and multi-agent [Zhou et al.(2021)] settings, and these choices require prior knowledge of problem parameters. In particular, it is generally assumed that the strong convexity parameter of the cost functions (single-agent) or the strong monotonicity parameter of the game (multi-agent) are known. Thus, OGD's appealing guarantees are not feasible in practice if the choice of step sizes cannot be made ***fully adaptive*** to problem parameters. Further, the feasibility issue is more acute in the multi-agent setting: in addition to requiring prior knowledge of problem parameters for step-size designs, recent work on adaptive OGD has assumed that each agent determines their step size using global information from all agents [Lin et al.(2020), Antonakopoulos et al.(2021), Hsieh et al.(2021)]. This is a practical and theoretical conundrum—if the agents can achieve this level of coordination, learning would be unnecessary in the first place. The same issue occurs for ONS in the single-agent setting, where the exp-concave parameter is needed as an input to the algorithm.[4] Consequently, the feasibility considerations lead us to consider the following question: ***Can we design a feasible and doubly optimal variant of OGD under strong convexity and strong monotonicity? What about ONS?***

Our answer is a "yes" in a strong sense. We present a single feasible OGD algorithm—and hence a single parameter-adaptive scheme—that simultaneously (up to log factors) achieves optimal regret in the single-agent setting and optimal last-iterate convergence rate to the unique Nash equilibrium in the multi-agent setting. This analysis is different from and more challenging than that involved in the design of feasible OGD algorithms separately for single-agent and multi-agent settings. In particular, it could be that an effective adaptive scheme for the strong convexity parameter in single-agent setting is different from that for the strong monotonicity parameter in multi-agent setting, in which case one has ***at best*** either a feasible algorithm with optimal regret or a feasible algorithm that has an optimal convergence-to-Nash guarantee, but not both. Such results would still be of considerable value but our results in this paper show that such intermediate results can

---

[2]When the strongly monotone games have Lipschitz gradients—a condition that does not hold in many games of interest—a classic result from the variational inequality literature implies that multi-agent OGD converges to the unique Nash equilibrium at a geometric rate due to a contraction.

[3]This result is further generalized in [Loizou et al.(2021)], who obtain the same last-iterate convergence rate for strongly variationally stable games, under weaker noise assumptions.

[4]The multi-agent ONS has not yet been explored, and even the time-average convergence of ONS still remains to be established.

be bypassed; indeed, the best of both worlds can be achieved. We also develop a single feasible variant of ONS that (up to log factors) achieves optimal regret in the single-agent setting with exp-concave loss functions and optimal time-average convergence rate to the unique Nash equilibrium in the multi-agent setting. For the latter result we introduce and analyze a new class of exp-concave games.

Our results can also be cast in the framework of variational inequalities (VIs). Indeed, they can be viewed as contributing to the VI literature by presenting a decentralized, feasible optimization algorithm for finding a solution of a strongly monotone VI. We prefer to emphasize, however, the online learning perspective, and the design of no-regret algorithms, given the direct connection of those algorithms to game-theoretic settings. In multi-agent games, it is natural to focus on decentralized algorithms and on algorithms that make minimal assumptions about their environment, allowing that environment to consist of other agents that may be responding in complex ways to an agent's actions. Our double optimality contributions are best understood as a further weakening of these assumptions, allowing interacting agents to choose actions effectively in an unknown, possibly adversarial, environment.

1.4   Our Contributions

We present a feasible variant of OGD that we refer to as AdaOGD that does not require knowing any problem parameter. It is guaranteed to achieve a minimax optimal (up to a log factor) regret bound of $O(\log^2(T))$ in the single-agent setting with strongly convex cost functions. Further, in a strongly monotone game, if each agent employs AdaOGD, we show that the joint action converges to the unique Nash equilibrium in a **_last-iterate sense_** at a rate of $O(\frac{\log^3(T)}{T})$, again optimal up to log factors. In comparison, the existing single-agent OGD [Hazan(2016)] and multi-agent OGD [Zhou et al.(2021)] methods require prior knowledge of the strong convexity/monotonicity parameter (respectively) to perform the step-size design with theoretical guarantees. It is worth noting that if the game has only a single agent, the strong monotonicity parameter degenerates to the strong convexity parameter. However, when there are multiple agents, the strong monotonicity parameter depends on all agents' cost functions. As such, one would naturally think that these two settings would require two different adaptive schemes. Surprisingly, our AdaOGD algorithm, which is based on a single adaptive principle, works in both settings, achieving optimal regret in the single-agent setting and optimal last-iterate convergence in the multi-agent setting (up to log factors). A particularly important application of our results is the problem of learning to order in the classical newsvendor problem, where due to lost sales, only (noisy) gradient feedback can be observed. Our results immediately yield the first feasible near-optimal algorithm—both in the single-retailer setting [Huh and Rusmevichientong(2009)] and in the multi-retailer setting [Netessine and Rudi(2003)]. This is in contrast to previous work that requires problem parameters to be known. Indeed, the direct application of our results to the VI setting yields a decentralized, feasible optimization algorithm for finding a solution of a strongly monotone VI.

Additionally, we provide a feasible variant of ONS (that we refer to as AdaONS) that again does not require prior knowledge of any problem parameter. It is also guaranteed to achieve a minimax optimal regret bound of $O(d \log^2(T))$ (up to a log factor) in the single-agent setting with exp-concave cost functions. Further, we propose a new class of exp-concave (EC) games and show that if each agents employs AdaONS, an optimal time-average convergence rate of $O(\frac{d \log^2(T)}{T})$ is obtained. Much like strongly monotone games that provide a multi-agent generalization of strongly convex cost functions, the EC games that we introduce are a natural generalization of exp-concave

cost functions from the single-agent to a multi-agent setting. Again, in this case, a single adaptive scheme works for both of these two settings. One thing to note here is that we first establish an $O(\frac{d \log(T)}{T})$ time-average convergence rate for the multi-agent version of classical ONS that is non-adaptive (and hence not feasible). To the best of our knowledge, results of this kind have not appeared in the game-theoretic literature and they yield a decentralized, feasible optimization algorithm for finding a solution of a new class of VIs in that line of literature.

Perhaps the most surprising takeaway from our work is that both AdaOGD and AdaONS are based on a simple and unifying randomized strategy that selects the step size based on a set of independent and identically distributed geometric random variables.

2 Feasible Single-Agent Online Learning under Strongly Convex Costs (AdaOGD)

In this section, we present adaptive OGD (AdaOGD), a feasible single-agent online learning algorithm, and prove that AdaOGD achieves a near-optimal regret of $O(\log^2(T))$ for a class of strongly convex cost functions. We also show that our algorithm can be used to solve the problem of adaptive ordering in newsvendor problems with lost sales. To our knowledge, this is the first feasible no-regret learning algorithm for the newsvendor-with-lost-sales problem with strong regret guarantees.

2.1 Algorithmic Scheme

We continue with the setup in the introduction, focusing on strongly convex cost functions f_t:

Definition 1. *A function $f : \mathbb{R}^d \mapsto \mathbb{R}$ is β-strongly convex if $f(\cdot) - 0.5\beta\| \cdot \|^2$ is convex.*

We work with a more general (and relaxed) model of gradient feedback [Flaxman et al.(2005)]:

(i) At each round t, an unbiased and bounded gradient is observed. That is, the observed noisy gradient ξ^t satisfies $\mathbb{E}[\xi^t \mid x^t] = \nabla f_t(x^t)$ and $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq G^2$ for all $t \geq 1$.

(ii) The action set $\mathcal{X}$ is bounded by a diameter $D > 0$, i.e., $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$.

A lower bound of $\Omega(\log(T))$ has been established in [Hazan and Kale(2014), Theorem 18] under an assumption of perfect gradient feedback. However, even with noisy gradient feedback, OGD with a particular step size can achieve the minimax-optimal regret bound of $\Theta(\log(T))$ [Hazan et al.(2007)]. In particular, we write OGD as $x^{t+1} \leftarrow \mathcal{P}_{\mathcal{X}}(x^t - \frac{1}{\beta(t+1)}\nabla f_t(x^t))$, which is equivalent to

$$\eta^{t+1} \leftarrow \beta(t+1) \qquad x^{t+1} \leftarrow \underset{x \in \mathcal{X}}{\text{argmin}} \left\{ (x - x^t)^\top \nabla f_t(x^t) + \frac{\eta^{t+1}}{2}\|x - x^t\|^2 \right\}$$

The value of $\beta(t+1)$ comes from the key inequality for β-strongly convex functions:

$$f(x') \geq f(x) + (x' - x)^\top \nabla f(x) + \frac{\beta}{2}\|x' - x\|^2 \qquad (2)$$

Despite the elegance of OGD (and its optimal regret guarantee), however, it is inadequate since it requires knowledge of the problem parameter β. We can address this issue by a simple randomization strategy based on independent, identically distributed geometric random variables, $M^t \sim \text{Geometric}(p_0)$, for $p_0 = \frac{1}{\log(T+10)}$; i.e., $\mathbb{P}(M^t = k) = (1 - p_0)^{k-1}p_0$ for $k \in \{1, 2, \ldots\}$. See Algorithm 1.

6

Algorithm 1 AdaOGD(x^1, T)

1: **Input:** initial point $x^1 \in \mathcal{X}$ and the total number of rounds T
2: **Initialization:** $p_0 = \frac{1}{\log(T+10)}$
3: **for** $t = 1, 2, \ldots, T$ **do**
4:      sample $M^t \sim \mathsf{Geometric}(p_0)$
5:      set $\eta^{t+1} \leftarrow \frac{t+1}{\sqrt{1+\max\{M^1,\ldots,M^t\}}}$
6:      update $x^{t+1} \leftarrow \operatorname{argmin}_{x \in \mathcal{X}} \left\{ (x - x^t)^\top \xi^t + \frac{\eta^{t+1}}{2} \|x - x^t\|^2 \right\}$
7: **end for**

Remark (Compared with doubling trick)**.** In the context of online learning, the doubling trick [Shalev-Shwartz(2012)] is commonly used to make OGD adaptive to **specific** unknown parameters under **convex** costs. In particular, for any algorithm that enjoys a regret bound of $O(\sqrt{T})$ but requires the knowledge of T (such as OGD under convex costs), the doubling trick converts such an algorithm into an algorithm that does not require the knowledge of T. The idea is to divide the time into periods of increasing size and run the original algorithm on each period: for $m = 0, 1, 2, \ldots$, we run the original algorithm on the 2^m rounds $t = 2^m, \ldots, 2^{m+1} - 1$. The resulting algorithm still enjoys a regret bound of $O(\sqrt{T})$. However, it is nontrivial to apply the doubling trick to make OGD adaptive to the strongly convex parameter β under **strongly convex** costs. In particular, a natural adaptation of the doubling trick under strongly convex costs is as follows: for $m = 0, 1, 2, \ldots$, we run OGD with $\eta^t = \frac{t}{2^m}$ on the rounds $t = 2^m, \ldots, 2^{m+1} - 1$. The analysis contains two parts: (i) for $0 \le m \le \lfloor \log_2(1/\beta) \rfloor$, the regret for each round is $O(2^m)$. This leads to a total constant regret of $O(1/\beta)$; (ii) for $\lceil \log_2(1/\beta) \rceil \le m \le \lceil \log_2(T) \rceil$, the regret for each round is $O(2^m)$. This unfortunately leads to a linear regret of $O(T)$. This argument of course does not eliminate the possibility that some variant of doubling could lead to doubly optimal learning algorithms for strongly monotone games. Our results do suggest that a fruitful way to search for such a procedure would be via some form of randomization.

Remark (Comparison with online learning with adaptive regret)**.** The static regret defined in Eq. (1) can be unsuitable as the environments are non-stationary and the best action is drifting over time [Hazan(2016)]. To tackle this issue, there are two independently proposed notions: dynamic regret [Zinkevich(2003), Besbes et al.(2015)] and adaptive regret [Hazan and Seshadhri(2007), Daniely et al.(2015)]. Several recent studies have further investigated the adaptive/dynamic regret minimization by leveraging the curvature of loss functions, such as exponential concavity [Hazan and Seshadhri(2007), Baby and Wang(2021)] and strong convexity [Mokhtari et al.(2016), Baby and Wang(2022)]. However, the adaptivity in regret is **different** from that in learning rates and all of these algorithms require knowing the strong convexity parameter and the exp-concavity parameters.

Remark (Comparison with geometrization)**.** In the offline finite-sum optimization, the geometrization trick [Lei and Jordan(2017), Lei and Jordan(2020)]—which sets the length of each epoch as a geometric random variable—has been used to make the stochastic variance-reduced gradient (SVRG) algorithm of [Johnson and Zhang(2013)] adaptive to both strong convexity parameter and target accuracy. While similar in spirit, our approach is not a straightforward application of the geometrization trick. The key difference between OGD and SVRG is their dependence on the strongly convexity parameter. The former needs that parameter to set the stepsize while the latter algorithm needs it to set the length of each epoch. The intuition behind the geometrization trick is to randomly set the length of each epoch using geometric random variables, allowing terms

to telescope across the outer and inner loops; such telescoping does not happen in SVRG, a fact which leads to the loss of adaptivity for SVRG (see [Lei and Jordan(2020), Section 3.1]). In contrast, our technique is designed to randomly set the stepsize using geometric random variables, implementing a trade-off for bounding the regret and last-iterate convergence rate. The telescoping from [Lei and Jordan(2020)] occurs due to the specific nature of SVRG (or more broadly, the setting of offline finite-sum optimization), and does not appear in our analysis for AdaOGD and its multi-agent generalization.

2.2 Regret Guarantees

We present our result on the regret minimization property in the following theorem.

Theorem 1. *For an arbitrary fixed sequence of β-strongly convex functions $f_1, \ldots, f_T$, where each f_t satisfies $\mathbb{E}[\xi^t \mid x^t] = \nabla f_t(x^t)$ and $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq G^2$ for all $t \geq 1$, and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. If the agent employs Algorithm 1, we have*

$$\mathbb{E}[Regret(T)] \leq \frac{D^2}{2}(1 + e^{\frac{1}{\beta^2 \log(T+10)}}) + \frac{G^2 \log(T+1)}{2}\sqrt{1 + \log(T+10) + \log(T)\log(T+10)}$$

As a consequence, we have $\mathbb{E}[Regret(T)] = O(\log^2(T))$.

Remark. Theorem 1 demonstrates that Algorithm 1 achieves a near-optimal regret since the upper bound matches the lower bound up to a log factor; indeed, [Hazan and Kale(2014)] proved the lower bound of $\Omega(\log(T))$ for this setting. Further, Algorithm 1 dynamically adjusts η^{t+1} without any prior knowledge of problem parameters, only utilizing the noisy feedback $\{\xi^t\}_{t \geq 1}$.

3 Feasible Multi-Agent Online Learning in Strongly Monotone Games (MA-AdaOGD)

In this section, we consider feasible multi-agent learning in monotone games. Our main result is that if each agent applies AdaOGD in a strongly monotone game (the multi-agent generalization of strongly convex costs), the joint action of all agents converges in a last-iterate sense to the unique Nash equilibrium at a near-optimal rate. In contrast to previous work, our results provide the first feasible no-regret learning algorithm that is doubly optimal; in particular, in addition to not requiring any prior knowledge of the problem parameters, one does not need to adjust the step-size schedule based on whether an agent is in the single-agent setting or the multi-agent setting. It is important to note that these are two different merits, and our algorithm enjoys both of them.

3.1 Basic Definitions and Notations

We first review the definition of continuous games and consider a class of monotone games. In particular, we focus on continuous games played by a set of agents, $\mathcal{N} = \{1, 2, \ldots, N\}$. Each agent selects an ***action*** x_i from a convex and bounded $\mathcal{X}_i \subseteq \mathbb{R}^{d_i}$. The incurred cost for each agent is determined by the joint action $x = (x_i; x_{-i}) = (x_1, x_2, \ldots, x_N)$. We let $\|\cdot\|$ denote the Euclidean norm (Other norms can also be accommodated here and different $\mathcal{X}_i$'s can have different norms).

Definition 2. *A continuous game is a tuple $\mathcal{G} = \{\mathcal{N}, \mathcal{X} = \Pi_{i=1}^N \mathcal{X}_i, \{u_i\}_{i=1}^N\}$, where $\mathcal{N}$ is a set of N agents, $\mathcal{X}_i \subseteq \mathbb{R}^{d_i}$ is the ith agent's action set that is both convex and bounded, and $u_i : \mathcal{X} \to \mathbb{R}$ is the ith agent's cost function satisfying:*

(i) $u_i(x_i; x_{-i})$ is continuous in x and continuously differentiable in x_i

(ii) $v_i(x) = \nabla_{x_i} u_i(x_i; x_{-i})$ is continuous in x

For simplicity, we denote $v(\cdot) = (v_1(\cdot), v_2(\cdot), \ldots, v_N(\cdot))$ as the joint profile of all agents' individual gradients.

We work with an analogous model of gradient feedback:

(i) At each round t, an unbiased and bounded gradient is observed as a feedback signal. In particular, the observed noisy gradient ξ^t satisfies $\mathbb{E}[\xi^t \mid x^t] = v(x^t)$ and $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq G^2$ for all $t \geq 1$.

(ii) The action set $\mathcal{X}$ is bounded by a diameter $D > 0$, i.e., $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$.

The study of monotone games dates to [Rosen(1965)] who considered a class of games that satisfy the ***diagonal strict concavity*** (DSC) condition.[5] Further contributions appeared in [Sandholm(2015)] and [Sorin and Wan(2016)], where games that satisfy DSC are called "contractive" and "dissipative." In this context, a game $\mathcal{G}$ is ***monotone*** if $(x' - x)^\top(v(x') - v(x)) \geq 0$ for any $x, x' \in \mathcal{X}$. Intuitively, the notion of monotonicity generalize the notion of convexity; indeed, the gradient operator of a convex function f satisfies that $(x' - x)^\top(\nabla f(x') - \nabla f(x)) \geq 0$ for any $x, x' \in \mathcal{X}$.

We now define the class of strongly monotone games:

Definition 3. *A continuous game $\mathcal{G}$ is said to be β-**strongly monotone** if we have $\langle x' - x, v(x') - v(x) \rangle \geq \beta \|x' - x\|^2$ for all $x, x' \in \mathcal{X}$.*

A standard solution concept for non-cooperative games is the ***Nash equilibrium*** (NE), where no agent has an incentive to deviate from her strategy [Osborne and Rubinstein(1994)]. For the continuous games considered in this paper, we are interested in pure-strategy Nash equilibria since the randomness introduced by mixed strategies is unnecessary when each action lives in a continuum.

Definition 4. *An action profile $x^\star \in \mathcal{X}$ is called a **Nash equilibrium** of $\mathcal{G}$ if it is resilient to unilateral deviations; that is, $u_i(x_i^\star; x_{-i}^\star) \leq u_i(x_i; x_{-i}^\star)$ for all $x_i \in \mathcal{X}_i$ and $i \in \mathcal{N}$.*

[Debreu(1952)] proved that any continuous game admits at least one Nash equilibrium if all action sets are convex and bounded, and all cost functions are individually convex (i.e., $u_i(x_i; x_{-i})$ is convex in x_i for a fixed x_{-i}). Moreover, there is a variational characterization that forms the basis of equilibrium computation under an individual convexity condition [Facchinei and Pang(2007)]. We summarize this characterization in the following proposition.

Proposition 1. *If all cost functions are in a continuous game $\mathcal{G}$ are individually convex, the joint action $x^\star \in \mathcal{X}$ is a Nash equilibrium if and only if $(x - x^\star)^\top v(x^\star) \geq 0$ for all $x \in \mathcal{X}$.*

The notion of strong monotonicity arises in various application domains. Examples include strongly-convex-strongly-concave zero-sum games, atomic splittable congestion games in networks with parallel links [Orda et al.(1993), Sorin and Wan(2016), Mertikopoulos and Zhou(2019)], wireless network optimization [Weeraddana et al.(2012), Tan(2014), Zhou et al.(2021)] and classical online decision-making problems [Cesa-Bianchi and Lugosi(2006)].

---

[5]This condition is equivalent to the notion of strict monotonicity in convex analysis [Bauschke and Combettes(2011)]; see [Facchinei and Pang(2007)] for further discussion.

Strongly monotone games satisfy the individual convexity condition and hence the existence of at least one Nash equilibrium is ensured. Moreover, every strongly monotone game admits a unique Nash equilibrium [Zhou et al.(2021)]. Thus, one appealing feature of strongly monotone games is that finite-time convergence can be derived in terms of $\|\hat{x} - x^\star\|^2$ where $x^\star \in \mathcal{X}$ is a unique Nash equilibrium. Despite some recent progress on last-iterate convergence rates for non-strongly monotone games [Lin et al.(2020), Golowich et al.(2020), Cai et al.(2022)], last-iterate convergence rates in terms of $\|\hat{x} - x^\star\|^2$ are only available for strongly monotone games [Bravo et al.(2018), Zhou et al.(2021), Lin et al.(2021)]. An important gap in the literature is that there currently do not exist **doubly optimal** and **feasible** learning algorithms for strongly monotone games.

3.2 Algorithmic Scheme

We review the multi-agent OGD method that is a generalization of single-agent OGD. Letting $x_i^1 \in \mathcal{X}_i$ for all $i \in \mathcal{N}$, the multi-agent version of OGD (cf. Definition 3) performs the following step at each round:

$$\eta_i^{t+1} \leftarrow \beta(t+1) \qquad x_i^{t+1} \leftarrow \underset{x_i \in \mathcal{X}_i}{\operatorname{argmin}} \left\{ (x_i - x_i^t)^\top \xi_i^t + \frac{\eta_i^{t+1}}{2} \|x_i - x_i^t\|^2 \right\} \tag{3}$$

In the following theorem, we summarize the results from [Zhou et al.(2021)] on the optimal last-iterate convergence rate of multi-agent OGD in Eq. (3) using a squared Euclidean distance function.

Theorem 2 ([Zhou et al.(2021)])**.** *Suppose that a continuous game $\mathcal{G}$ is β-strongly monotone and let $G, D > 0$ be problem parameters satisfying $\mathbb{E}[\xi^t \mid x^t] = v(x^t)$ and let $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq G^2$ for all $t \geq 1$, and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. If all agents employ multi-agent OGD in Eq. (3), we have*

$$\mathbb{E}[\|x^T - x^\star\|^2] \leq \frac{4G^2}{\beta^2 T}$$

where $x^\star \in \mathcal{X}$ denotes the unique Nash equilibrium. As a consequence, we have $\mathbb{E}[\|x^T - x^\star\|^2] = O(\frac{1}{T})$.

Remark. Theorem 2 demonstrates that multi-agent OGD can achieve a near-optimal convergence rate in strongly monotone games; indeed, the convergence rate of multi-agent OGD matches the lower bound of $\Omega(\frac{1}{T})$ proved in [Nemirovski and Yudin(1983)] for strongly convex optimization.

A drawback of multi-agent OGD is that it requires knowledge of the strong monotonicity parameters in order to update η^{t+1} and is thus not feasible in practice. We are not aware of any research that addresses this key issue. This is possibly because existing adaptive techniques are specialized to **upper curvature information**, e.g., the Lipschitz constant of function values or gradients [Duchi et al.(2011), Kingma and Ba(2015), Mukkamala and Hein(2017), Levy(2017), Levy et al.(2018), Bach and Levy(2019), Antonakopoulos et al.(2021), Hsieh et al.(2021)], and are not suitable for estimating **lower curvature information** such as that encoded by the strong monotonicity parameter. The goal of this section is to extend Algorithm 1 to multi-agent learning in games, showing that our adaptive variant of OGD is both doubly optimal and feasible.

We again employ a randomization strategy such that the resulting algorithm is adaptive to the strong monotonicity parameter and other problem parameters. We again choose independently identical distributed geometric random variables, i.e., $M_i^t \sim \mathsf{Geometric}(p_0)$, for $p_0 = \frac{1}{\log(T+10)}$.[6]

---

[6]We can define agent-specific probabilities $p_0^i \in (0, 1)$ and prove the same finite-time convergence guarantee. For simplicity, we use an agent-independent probability $p_0 = \frac{1}{\log(T+10)}$.

Algorithm 2 MA-AdaOGD$(x_1^1, x_2^1, \ldots, x_N^1, T)$

1: **Input:** initial points $x_i^1 \in \mathcal{X}_i$ for all $i \in \mathcal{N}$ and the total number of rounds T
2: **Initialization:** $p_0 = \frac{1}{\log(T+10)}$
3: **for** $t = 1, 2, \ldots, T$ **do**
4:   **for** $i = 1, 2, \ldots, N$ **do**
5:     sample $M_i^t \sim \mathsf{Geometric}(p_0)$
6:     set $\eta_i^{t+1} \leftarrow \frac{t+1}{\sqrt{1+\max\{M_i^1, \ldots, M_i^t\}}}$
7:     update $x_i^{t+1} \leftarrow \operatorname{argmin}_{x_i \in \mathcal{X}_i} \left\{ (x_i - x_i^t)^\top \xi_i^t + \frac{\eta_i^{t+1}}{2} \|x_i - x_i^t\|^2 \right\}$
8:   **end for**
9: **end for**

Note that all the other updates are analogous to those of Algorithm 1 and the resulting algorithm is decentralized. This defines our adaptive multi-agent variant of OGD, as detailed in Algorithm 2.

3.3 Finite-Time Last-Iterate Convergence Guarantee

In the following theorem, we summarize our main results on the last-iterate convergence rate of Algorithm 2 using a distance function based on squared Euclidean norm.

Theorem 3. *Suppose that a continuous game $\mathcal{G}$ is β-strongly monotone and let $G, D > 0$ be problem parameters satisfying $\mathbb{E}[\xi^t \mid x^t] = v(x^t)$ and let $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq G^2$ for all $t \geq 1$, and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. If all agents employ Algorithm 2, we have*

$$
\begin{aligned}
\mathbb{E}[\|x^T - x^\star\|^2] &\leq \frac{D^2}{T}(1 + e^{\frac{1}{4\beta^2 \log(T+10)}})\sqrt{1 + \log(T + 10)} + \log(NT)\log(T+10) \\
&\quad + \frac{G^2}{T}\log(T+1)(1 + \log(T+10) + \log(NT)\log(T+10))
\end{aligned}
$$

where $x^\star \in \mathcal{X}$ is the unique Nash equilibrium. As a consequence, we have $\mathbb{E}[\|x^T - x^\star\|^2] = O(\frac{\log^3(T)}{T})$.

Remark. Theorem 3 demonstrates that Algorithm 2 achieves a near-optimal convergence rate since the upper bound matches the lower bound [Nemirovski and Yudin(1983)] up to a log factor. This result together with Theorem 1 shows that our adaptive variant of OGD is a doubly optimal and feasible learning algorithm for strongly monotone games.

Remark (Importance of doubly optimality)**.** To further appreciate the concept of double optimality note that our starting point is single-agent that is learning in an online manner in an arbitrarily non-stationary and possibly adversarial environment. It is well known that online gradient descent achieves the minimax optimal regret bounds $(\Theta(\log)T$ in this setting, for strongly convex cost functions and $\Theta(\sqrt{T}))$. As such, to an agent engaged in an online decision making process, it is natural to employ OGD in such an environment where statistical regularity may be lacking. In particular, the most common instantiation of such a non-stationary environment is one that comprises other agents who are simultaneously engaged in the online decision making process and whose actions impact all others' costs/rewards. In other words, each agent is acting in an environment whose cost/reward is determined by an ***opaque*** game: the cost/reward is determined by an unknown underlying game, where even the number of agents that comprise of the game may

be unknown. Consequently, it is natural to ask if each agent adopts OGD to maximize its finite-horizon cumulative reward, would the system jointly converge to a Nash equilibrium, a multi-agent optimal outcome where no agent has any incentive to unilaterally deviate? If not, then in the long run an agent would be able to do better by deviating from what is prescribed by the online learning algorithm, given what all the other agents are doing, thereby producing "regret." Consequently, an online learning algorithm that is doubly optimal—where a single agent adopting it attains optimal finite-time no-regret guarantees and when all agents adopt it they converge to a Nash equilibrium—effectively bridges optimal transient performance (i.e., finite-horizon performance) with optimal long-run performance (i.e., the equilibrium outcome).

4 Extensions to Exp-Concave Cost Functions and Games

In this section, we extend our results—for both single-agent and multi-agent learning—to a broader class of cost functions and game structures defined in terms of exp-concavity. For exp-concave cost functions, our adaptive variant of an online Newton step (AdaONS) can achieve a near-optimal regret of $O(d \log^2(T))$. For exp-concave games, we propose a multi-agent online Newton step with a near-optimal time-average convergence rate of $O(\frac{d \log(T)}{T})$. We also extend it to a multi-agent adaptive online Newton step (MA-AdaONS) that achieves a near-optimal rate of $O(\frac{d \log^2(T)}{T})$.

4.1 Single-Agent Learning with Exp-Concave Cost

We start by recalling the definition of exp-concave functions.

Definition 5. *A function $f : \mathbb{R}^d \mapsto \mathbb{R}$ is α-**exp-concave** if $e^{-\alpha f(\cdot)}$ is concave.*

Exp-Concave (EC) functions, as a strict subclass of convex functions and as a strict superclass of strongly convex functions, have found applications in many fields, including information theory [Cover(1999)], stochastic portfolio theory [Fernholz(2002)], optimal transport [Villani(2009)], optimization [Hazan et al.(2014), Mahdavi et al.(2015)], probability theory [Pal et al.(2018)] and statistics [Juditsky et al.(2008), Koren and Levy(2015), Yang et al.(2018)].

We work with the following model of perfect gradient feedback:[7]:

(i) At each round t, an exact gradient is observed. That is, we have $\nabla f_t(x^t)$ for all $t \geq 1$.

(ii) The action set $\mathcal{X}$ is bounded by a diameter $D > 0$, i.e., $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$.

For this class of cost functions, a lower bound on regret of $\Omega(d \log(T))$ is known (see Section 3 in [Abernethy et al.(2008)]). Two typical examples which admit EC cost functions are online linear regression [Kivinen and Warmuth(1999)] and universal portfolio management [Cover(1991)].

(i) ***In universal portfolio management***, we have $f_t(x) = -\log(a_t^\top x)$ where $x \in \Delta^d$ is a probability vector and $a^t \in \mathbb{R}_+^d$ is the d-dimensional vector of all d assets' relative prices between period t and period $t-1$. Here, f_t is EC and [Ordentlich and Cover(1998)] established a lower bound of $\Omega(d \log(T))$ for the cumulative regret.

---

[7]Can we generalize the aforementioned results to the setting with noisy gradient feedback? We leave the answer to this question to future work.

Algorithm 3 AdaONS(x^1, T)

1: **Input:** initial point $x^1 \in \mathcal{X}$ and the total number of rounds T
2: **Initialization:** $A^1 = I_d$ where $I_d \in \mathbb{R}^{d \times d}$ is an identity matrix and $p_0 = \frac{1}{\log(T+10)}$
3: **for** $t = 1, 2, \ldots, T$ **do**
4:     sample $M^t \sim$ Geometric(p_0)
5:     set $\eta^{t+1} \leftarrow \frac{1}{\sqrt{1 + \max\{M^1, \ldots, M^t\}}}$
6:     update $A^{t+1} \leftarrow A^t + \nabla f_t(x^t) \nabla f_t(x^t)^\top$
7:     update $x^{t+1} \leftarrow \text{argmin}_{x \in \mathcal{X}} \left\{ (x - x^t)^\top \nabla f_t(x^t) + \frac{\eta^{t+1}}{2}(x - x^t)^\top A^{t+1}(x - x^t) \right\}$
8: **end for**

(ii) *In online linear regression*, we have $f_t(x) = (a_t^\top x - b_t)^2$, which is in general not a strongly convex function, but is indeed EC for any a_t, b_t. Formally, [Vovk(1997)] proved that there exists a randomized strategy of the adversary for choosing the vectors a_t, b_t such that the expected cumulative regret scales as $\Omega(d \log(T))$.

This lower bound was established without making the connectio to the general EC function class. Note also that the $\Omega(d \log(T))$ lower bound is interesting given that $d \geq 1$ does not enter the lower bound for strongly convex functions [Hazan and Kale(2014)].

On the upper-bound side, there are many algorithms that achieve the minimax-optimal regret, such as the online Newton step (ONS) of [Hazan et al.(2007)]. Formally, let $\{f_t\}_{t \geq 1}$ be α-exp-concave cost functions and let $G > 0$ be problem parameters satisfying $\|\nabla f_t(x)\| \leq G$ for all $t \geq 1$. Then, the ONS algorithm is given by

$$
\eta \leftarrow \frac{1}{2} \min \left\{ \frac{1}{4GD}, \alpha \right\} \qquad A^{t+1} \leftarrow A^t + \nabla f_t(x^t) \nabla f_t(x^t)^\top
$$
$$
x^{t+1} \leftarrow \underset{x \in \mathcal{X}}{\text{argmin}} \left\{ (x - x^t)^\top \nabla f_t(x^t) + \frac{\eta}{2}(x - x^t)^\top A^{t+1}(x - x^t) \right\} \tag{4}
$$

The choice $\frac{1}{2} \min \left\{ \frac{1}{4GD}, \alpha \right\}$ comes from [Hazan et al.(2007), Lemma 3]: if $\|\nabla f(x)\| \leq G$ and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$, the α-exp-concave function f satisfies

$$
f(x') \geq f(x) + (x' - x)^\top \nabla f(x) + \frac{1}{4} \min \left\{ \frac{1}{4GD}, \alpha \right\} (x' - x)^\top (\nabla f(x) \nabla f(x)^\top)(x' - x) \tag{5}
$$

Intuitively, this equation implies that any α-exp-concave function can be approximated by a local quadratic function with a matrix $\eta \nabla f(x) \nabla f(x)^\top$ and moreover ONS can exploit such structure. Although ONS suffers from a quadratic dependence on the dimension in terms of per-iteration cost, there has been progress in alleviating this complexity. For example, [Luo et al.(2016)] proposed a variant of ONS that attains linear dependence on the dimension using sketching techniques.

4.2 Feasible Single-Agent Online Learning under Exp-Concave Cost

The scheme of ONS in Eq. (4) is infeasible in practice since it requires prior knowledge of the problem parameters α, G and D. To address this issue, we invoke the same randomization strategy as described in Algorithm 1. This results in the single-agent adaptive variant of ONS (cf. Algorithm 3).

We summarize our main results on the regret minimization properties of the algorithm in the following theorem.

13

Theorem 4. *Consider an arbitrary fixed sequence of α-exp-concave functions $f_1, \ldots, f_T$, where each f_t satisfies $\|\nabla f_t(x)\| \leq G$ for all $t \geq 1$ and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. If the agent employs Algorithm 3, we have*

$$\mathbb{E}[Regret(T)] \leq \frac{D^2}{2}(1 + G^2 e^{\frac{(\max\{8GD, 2\alpha^{-1}\})^2}{\log(T+10)}}) + \frac{d \log(TG^2 + 1)}{2}\sqrt{1 + \log(T+10) + \log(T)\log(T+10)}$$

Thus, we have $\mathbb{E}[Regret(T)] = O(d \log^2(T))$.

Remark. Theorem 4 demonstrates that Algorithm 3 achieves a near-optimal regret since the upper bound matches the lower bound up to a log factor; indeed, the lower bound of $\Omega(d \log(T))$ has been established for online linear regression and universal portfolio management. Further, Algorithm 3 dynamically adjusts η^{t+1} without any prior knowledge of problem parameters.

To prove Theorem 4, we again make use of a key descent inequality.

Lemma 1. *Consider an arbitrary fixed sequence of α-exp-concave functions $f_1, \ldots, f_T$, where each f_t satisfies $\|\nabla f_t(x)\| \leq G$ for all $t \geq 1$ and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. Letting the iterates $\{x^t\}_{t \geq 1}$ be generated by Algorithm 3, we have*

$$\sum_{t=1}^{T} f_t(x^t) - \sum_{t=1}^{T} f_t(x) \leq \frac{\eta^1}{2}(x^1 - x)^\top A^1(x^1 - x) + \frac{1}{2}\left(\sum_{t=1}^{T} \frac{1}{\eta^{t+1}} \nabla f_t(x^t)^\top (A^{t+1})^{-1} \nabla f_t(x^t)\right)$$

$$+ \sum_{t=1}^{T}(x^t - x)^\top \left(\frac{\eta^{t+1}}{2} A^{t+1} - \frac{\eta^t}{2} A^t - \frac{1}{4}\min\left\{\frac{1}{4GD}, \alpha\right\} \nabla f_t(x^t)\nabla f_t(x^t)^\top\right)(x^t - x), \quad \text{for all } x \in \mathcal{X}$$

The proofs of Lemma 1 and Theorem 4 are given in Section C.6.1 and C.3.

4.3 Exp-Concave (EC) Games

We define a class of exp-concave (EC) games via a game-theoretic generalization of the optimization framework in the previous section. Letting f be α-exp-concave with $\|\nabla f(x)\| \leq G$ and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$, [Hazan et al.(2007), Lemma 3] guarantees that

$$f(x') \geq f(x) + (x' - x)^\top \nabla f(x) + \frac{1}{4}\min\left\{\frac{1}{4GD}, \alpha\right\}(x' - x)^\top (\nabla f(x)\nabla f(x)^\top)(x' - x)$$

Equivalently, we have

$$(x' - x)^\top (\nabla f(x') - \nabla f(x)) \geq \frac{1}{4}\min\left\{\frac{1}{4GD}, \alpha\right\}(x' - x)^\top (\nabla f(x')\nabla f(x')^\top + \nabla f(x)\nabla f(x)^\top)(x' - x)$$

This leads to the following formal definition in which $v(\cdot)$ replaces $\nabla f(\cdot)$.

Definition 6. *A continuous game $\mathcal{G}$ is said to be α-**exp-concave** if we have*

$$\langle x' - x, v(x') - v(x) \rangle \geq \frac{1}{4}\min\left\{\frac{1}{4GD}, \alpha\right\} (\sum_{i=1}^{N}(x'_i - x_i)^\top (v_i(x')v_i(x')^\top + v_i(x)v_i(x)^\top)(x'_i - x_i))$$

for all $x, x' \in \mathcal{X}$.

Since EC games satisfy the individual convexity condition, we have the existence of at least one Nash equilibrium. However, multiple Nash equilibria might exist for EC games. For example, letting $d \geq 2$, we consider a single-agent game with the cost function $f(x) = \frac{1}{2}(a^\top x)^2$ and $\mathcal{X} = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$. It is easy to verify that this is a $(2/\|a\|^2)$-exp-concave game and any $x \in \mathcal{X}$ satisfying $a^\top x = 0$ will be a Nash equilibrium. It is thus natural to ask how to measure the optimality of a point $\hat{x} \in \mathcal{X}$. Proposition 1 inspires us to use a gap function, $\text{GAP}(\cdot) : \mathcal{X} \mapsto \mathbb{R}_+$, given by

$$\text{GAP}(\hat{x}) = \sup_{x \in \mathcal{X}} (\hat{x} - x)^\top v(x) \tag{6}$$

The above gap function is well defined and is nonnegative for all $\hat{x} \in \mathcal{X}$, given that at least one Nash equilibrium exists in EC games. Note that such a function has long been a standard example in the optimization literature [Facchinei and Pang(2007), Mertikopoulos and Zhou(2019)].

All strongly monotone games are clearly EC games if $\|v(x)\| \leq G$ for all $x \in \mathcal{X}$. Indeed, we have $\sum_{i=1}^N (x_i' - x_i)^\top (v_i(x')v_i(x')^\top + v_i(x)v_i(x)^\top)(x_i' - x_i) \leq 2G^2\|x' - x\|^2$ and

$$\langle x' - x, v(x') - v(x) \rangle \geq \beta\|x' - x\|^2 \geq \frac{\beta}{2G^2} \left(\sum_{i=1}^N (x_i' - x_i)^\top (v_i(x')v_i(x')^\top + v_i(x)v_i(x)^\top)(x_i' - x_i) \right)$$

This implies that a β-strongly monotone game is α-exp-concave if we have $\alpha \leq \frac{2\beta}{G^2}$. More generally, we remark that strongly monotone games are a rich class of games that arise in many real-world application problems [Bravo et al.(2018), Mertikopoulos and Zhou(2019), Lin et al.(2021)]. Moreover, there are also applications that can be cast as EC games rather than strongly monotone games, such as empirical risk minimization [Koren and Levy(2015), Yang et al.(2018)] and universal portfolio management [Cover(1991), Fernholz(2002)].

4.4 Multi-Agent Online Learning in EC Games

We start by extending the single-agent ONS algorithm in Eq. (4) to the multi-agent online learning in EC games (cf. Definition 6). We again work with a model of exact gradient feedback:

(i) At each round t, an exact gradient is observed. That is, we have $v(x^t)$ for all $t \geq 1$.

(ii) The action set $\mathcal{X} = \Pi_{i=1}^N \mathcal{X}_i$ is bounded by a diameter $D > 0$, i.e., $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$.

Letting $x_i^1 \in \mathcal{X}_i$ for all $i \in \mathcal{N}$, the multi-agent version of ONS performs the following step at each round ($A_i^1 = I_{d_i}$ for all $i \in \mathcal{N}$):

$$\begin{aligned} \eta_i &\leftarrow \frac{1}{2} \min\left\{ \frac{1}{4GD}, \alpha \right\} \qquad A_i^{t+1} \leftarrow A_i^t + v_i(x^t)v_i(x^t)^\top \\ x_i^{t+1} &\leftarrow \operatorname*{argmin}_{x_i \in \mathcal{X}_i} \left\{ (x_i - x_i^t)^\top v_i(x^t) + \frac{\eta_i}{2}(x_i - x_i^t)^\top A_i^{t+1}(x_i - x_i^t) \right\} \end{aligned} \tag{7}$$

We see from Eq. (7) that our multi-agent learning algorithm is a generalization of single-agent ONS in Eq. (4). Notably, it is a decentralized algorithm since each agent does not need to know any other agents' gradients. In the following theorem, we summarize our main results on the time-average convergence rate using the gap function in Eq. (6).

15

Algorithm 4 MA-AdaONS$(x_1^1, x_2^1, \ldots, x_N^1, T)$

1: **Input:** initial points $x_i^1 \in \mathcal{X}_i$ for all $i \in \mathcal{N}$ and the total number of rounds T

2: **Initialization:** $A_i^1 = I_{d_i}$ where $I_{d_i} \in \mathbb{R}^{d_i \times d_i}$ is an identity matrix and $p_0 = \frac{1}{\log(T+10)}$

3: **for** $t = 1, 2, \ldots, T$ **do**

4:      **for** $i = 1, 2, \ldots, N$ **do**

5:          sample $M_i^t \sim \mathsf{Geometric}(p_0)$

6:          set $\eta_i^{t+1} \leftarrow \frac{1}{\sqrt{1+\max\{M_i^1,\ldots,M_i^t\}}}$

7:          update $A_i^{t+1} \leftarrow A_i^t + v_i(x^t)v_i(x^t)^\top$

8:          update $x_i^{t+1} \leftarrow \operatorname{argmin}_{x_i \in \mathcal{X}_i} \left\{ (x_i - x_i^t)^\top v_i(x^t) + \frac{\eta_i^{t+1}}{2}(x_i - x_i^t)^\top A_i^{t+1}(x_i - x_i^t) \right\}$

9:      **end for**

10: **end for**

Theorem 5. *Suppose that a continuous game $\mathcal{G}$ is α-exp-concave and let $G, D > 0$ be problem parameters satisfying $\|v(x)\| \leq G$ and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. If all agents agree to employ multi-agent ONS in Eq. (7), we have*

$$\mathrm{GAP}(\bar{x}^T) \leq \frac{\alpha D^2}{4T} + \frac{d \log(TG^2+1)}{T} \max\left\{4GD, \frac{1}{\alpha}\right\}$$

where $\bar{x}^T = \frac{1}{T}\sum_{t=1}^{T} x^t$ denotes a time-average iterate. Thus, we have $\mathrm{GAP}(\bar{x}^T) = O(\frac{d\log(T)}{T})$.

Remark. Theorem 5 shows that the multi-agent ONS algorithm can achieve a near-optimal convergence rate in EC games; indeed, the convergence rate of multi-agent ONS matches up to a log factor the lower bound of $\Omega(\frac{d}{T})$ proved in [Mahdavi et al.(2015)] for exp-concave optimization.

4.5   Feasible Multi-Agent Online Learning in EC Games

We extend Algorithm 3 to multi-agent learning in EC games, showing that our adaptive variant of ONS is not only feasible but achieves a near-optimal convergence rate in terms of a gap function.

By employing a randomization strategy, we obtain an algorithm that is adaptive to exp-concavity parameter and other problem parameters. All the other updates are analogous to Algorithm 3 and the overall multi-agent framework, shown in Algorithm 4, is decentralized. we summarize our main results on the time-average convergence rate of Algorithm 4 using the gap function in Eq. (6).

Theorem 6. *Suppose that a continuous game $\mathcal{G}$ is α-exp-concave and let $G, D > 0$ be problem parameters satisfying $\|v(x)\| \leq G$ and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. If all agents agree to employ Algorithm 4, we have*

$$\mathbb{E}[\mathrm{GAP}(\bar{x}^T)] \leq \frac{D^2}{2T}\left(1 + G^2 e^{\frac{(\max\{8GD, 2\alpha^{-1}\})^2}{\log(T+10)}}\right) + \frac{d\log(TG^2+1)}{2T}\sqrt{1 + \log(T+10) + \log(T)\log(T+10)}$$

where $\bar{x}^T = \frac{1}{T}\sum_{t=1}^{T} x^t$ denotes a time-average iterate. Thus, we have $\mathbb{E}[\mathrm{GAP}(\bar{x}^T)] = O(\frac{d\log^2(T)}{T})$.

Directly extending Lemma 4 to multi-agent learning in EC games, we prove that the iterates

$\{x^t\}_{t\geq 1}$ generated by Algorithm 4 satisfy

$$
\begin{aligned}
\sum_{t=1}^{T}(x^t - x)^\top v(x) &\leq \sum_{i=1}^{N}\frac{\eta_i^1}{2}(x_i^1 - x_i)^\top A_i^1(x_i^1 - x_i) + \frac{1}{2}\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{1}{\eta_i^{t+1}}v_i(x^t)^\top (A_i^{t+1})^{-1}v_i(x^t)\right) \\
&+ \sum_{t=1}^{T}\sum_{i=1}^{N}(x_i^t - x_i)^\top \left(\frac{\eta_i^{t+1}}{2}A_i^{t+1} - \frac{\eta_i^t}{2}A_i^t - \frac{1}{4}\min\left\{\frac{1}{4GD},\alpha\right\}v_i(x^t)v_i(x^t)^\top\right)(x_i^t - x_i)
\end{aligned}
\tag{8}
$$

This inequality is crucial to the proof of Theorem 6, which can be found in Section C.5.

5 Discussion

Our results open up several directions for further research. First, we have assumed that the gradient feedback is always received at the end of each period. In practice, there may be delays. For instance, in the power-control example, the signal-to-noise ratio sent by the receiver to the transmitter through the feedback channel, from which the gradient can be recovered, is often received with a delay. As such, it is important to understand how delays impact the performance as well as to design the delay-adaptive algorithms that can operate robustly in such environments. Second, our paper has focused on (noisy) gradient feedback while another important yet more challenging type of feedback is bandit feedback, where we only observe (noisy) function values at the chosen action. This problem domain has been less explored than that of learning with (noisy) gradient feedback. For instance, it remains unknown what the minimax optimal regret bound for convex cost functions is; the optimal dependence on T is $\sqrt{T}$ [Bubeck et al.(2021)] but the optimal dependence on the dimension d is unknown (note that the OGD regret is dimension-independent if gradient feedback can be observed). Thus, it is desirable to develop the feasible and optimal bandit learning algorithms in both single-agent and multi-agent settings. Can our randomization technique be applicable in that setting and improve the existing algorithms [Bravo et al.(2018), Lin et al.(2021)]? This is a natural question for future work.

References

[Abernethy et al.(2008)] J. Abernethy, P. L. Bartlett, A. Rakhlin, and A. Tewari. Optimal strategies and minimax lower bounds for online convex games. In *COLT*, page 414–424. Omnipress, 2008.

[Alacaoglu and Malitsky(2022)] A. Alacaoglu and Y. Malitsky. Stochastic variance reduction for variational inequality methods. In *COLT*, pages 778–816. PMLR, 2022.

[Antonakopoulos et al.(2021)] K. Antonakopoulos, V. Belmega, and P. Mertikopoulos. Adaptive extra-gradient methods for min-max optimization and games. In *ICLR*, pages 1–12, 2021. URL https://openreview.net/forum?id=R0a0kFI3dJx.

[Baby and Wang(2021)] D. Baby and Y-X. Wang. Optimal dynamic regret in exp-concave online learning. In *COLT*, pages 359–409. PMLR, 2021.

[Baby and Wang(2022)] D. Baby and Y-X. Wang. Optimal dynamic regret in proper online learning with strongly convex losses and beyond. In *AISTATS*, pages 1805–1845. PMLR, 2022.

[Bach and Levy(2019)] F. Bach and K. Y. Levy. A universal algorithm for variational inequalities adaptive to smoothness and noise. In *COLT*, pages 164–194. PMLR, 2019.

[Balamurugan and Bach(2016)] P. Balamurugan and F. Bach. Stochastic variance reduction methods for saddle-point problems. In *NIPS*, pages 1416–1424, 2016.

[Balandat et al.(2016)] M. Balandat, W. Krichene, C. Tomlin, and A. Bayen. Minimizing regret on reflexive Banach spaces and Nash equilibria in continuous zero-sum games. In *NIPS*, pages 154–162, 2016.

[Bauschke and Combettes(2011)] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, volume 408. Springer, 2011.

[Besbes et al.(2015)] O. Besbes, Y. Gur, and A. Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015.

[Bloembergen et al.(2015)] D. Bloembergen, K. Tuyls, D. Hennes, and M. Kaisers. Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research*, 53:659–697, 2015.

[Blum and Mansour(2007)] A. Blum and Y. Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8(6):1307–1324, 2007.

[Blum(1998)] Avrim Blum. Online algorithms in machine learning. In *Online Algorithms*, pages 306–325. Springer, 1998.

[Bravo et al.(2018)] M. Bravo, D. Leslie, and P. Mertikopoulos. Bandit learning in concave N-person games. In *NeurIPS*, pages 5661–5671, 2018.

[Bubeck et al.(2021)] S. Bubeck, R. Eldan, and Y. T. Lee. Kernel-based methods for bandit convex optimization. *Journal of the ACM (JACM)*, 68(4):1–35, 2021.

[Cai et al.(2022)] Y. Cai, A. Oikonomou, and W. Zheng. Finite-time last-iterate convergence for learning in multi-player games. In *NeurIPS*, pages 33904–33919, 2022.

[Cesa-Bianchi and Lugosi(2006)] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[Chen et al.(2017)] Y. Chen, G. Lan, and Y. Ouyang. Accelerated schemes for a class of variational inequalities. *Mathematical Programming*, 165:113–149, 2017.

[Cover(1991)] T. M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

[Cover(1999)] T. M. Cover. *Elements of Information Theory*. John Wiley & Sons, 1999.

[Cutkosky(2020a)] A. Cutkosky. Better full-matrix regret via parameter-free online learning. In *NeurIPS*, pages 8836–8846, 2020a.

[Cutkosky(2020b)] A. Cutkosky. Parameter-free, dynamic, and strongly-adaptive online learning. In *ICML*, pages 2250–2259. PMLR, 2020b.

[Cutkosky and Orabona(2018)] A. Cutkosky and F. Orabona. Black-box reductions for parameter-free online learning in banach spaces. In *COLT*, pages 1493–1529. PMLR, 2018.

[Daniely et al.(2015)] Amit Daniely, Alon Gonen, and Shai Shalev-Shwartz. Strongly adaptive online learning. In *ICML*, pages 1405–1411, 2015.

[Debreu(1952)] G. Debreu. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences*, 38(10):886–893, 1952.

[Defazio et al.(2014)] A. Defazio, F. Bach, and S. Lacoste-Julien. SAGA: a fast incremental gradient method with support for non-strongly convex composite objectives. In *NIPS*, pages 1646–1654, 2014.

[Duchi et al.(2011)] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(7), 2011.

[Facchinei and Pang(2007)] F. Facchinei and J-S. Pang. *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer Science & Business Media, 2007.

[Fernholz(2002)] E. R. Fernholz. *Stochastic Portfolio Theory*, volume 48. Springer Science & Business Media, 2002.

[Flaxman et al.(2005)] A. D. Flaxman, A. T. Kalai, and H. B. McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *SODA*, pages 385–394, 2005.

[Foster et al.(2015)] D. J. Foster, A. Rakhlin, and K. Sridharan. Adaptive online learning. In *NeurIPS*, pages 3375–3383, 2015.

[Foster et al.(2017)] D. J. Foster, S. Kale, M. Mohri, and K. Sridharan. Parameter-free online learning via model selection. In *NeurIPS*, pages 6022–6032, 2017.

[Garber and Kretzu(2022)] D. Garber and B. Kretzu. New projection-free algorithms for online convex optimization with adaptive regret guarantees. In *COLT*, pages 2326–2359. PMLR, 2022.

[Goldsmith(2005)] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[Golowich et al.(2020)] N. Golowich, S. Pattathil, and C. Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. In *NeurIPS*, pages 20766–20778, 2020.

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

[Hazan and Kakade(2019)] E. Hazan and S. Kakade. Revisiting the Polyak step size. *ArXiv Preprint: 1905.00313*, 2019.

[Hazan and Kale(2014)] E. Hazan and S. Kale. Beyond the regret minimization barrier: Optimal algorithms for stochastic strongly-convex optimization. *Journal of Machine Learning Research*, 15(1):2489–2512, 2014.

[Hazan and Seshadhri(2007)] E. Hazan and C. Seshadhri. Adaptive algorithms for online decision problems. *ECCC*, 14(088), 2007.

[Hazan et al.(2007)] E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

[Hazan et al.(2014)] E. Hazan, T. Koren, and K. Y. Levy. Logistic regression: Tight bounds for stochastic and online optimization. In *COLT*, pages 197–209. PMLR, 2014.

[Hsieh et al.(2021)] Y-G. Hsieh, K. Antonakopoulos, and P. Mertikopoulos. Adaptive learning in continuous games: Optimal regret bounds and convergence to Nash equilibrium. In *COLT*, pages 2388–2422. PMLR, 2021.

[Huang and Zhang(2022)] K. Huang and S. Zhang. New first-order algorithms for stochastic variational inequalities. *SIAM Journal on Optimization*, 32(4):2745–2772, 2022.

[Huang et al.(2022)] K. Huang, N. Wang, and S. Zhang. An accelerated variance reduced extrapoint approach to finite-sum VI and optimization. *ArXiv Preprint: 2211.03269*, 2022.

[Huh and Rusmevichientong(2009)] W. T. Huh and P. Rusmevichientong. A nonparametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research*, 34(1): 103–123, 2009.

[Iusem et al.(2017)] A. N. Iusem, A. Jofré, R. I. Oliveira, and P. Thompson. Extragradient method with variance reduction for stochastic variational inequalities. *SIAM Journal on Optimization*, 27(2):686–724, 2017.

[Iusem et al.(2019)] A. N. Iusem, A. Jofré, R. I. Oliveira, and P. Thompson. Variance-based extragradient methods with line search for stochastic variational inequalities. *SIAM Journal on Optimization*, 29(1): 175–206, 2019.

[Jadbabaie et al.(2015)] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan. Online optimization: Competing with dynamic comparators. In *AISTATS*, pages 398–406. PMLR, 2015.

[Jalilzadeh and Shanbhag(2019)] A. Jalilzadeh and U. V. Shanbhag. A proximal-point algorithm with variable sample-sizes (PPAWSS) for monotone stochastic variational inequality problems. In *WSC*, pages 3551–3562. IEEE, 2019.

[Jiang and Xu(2008)] H. Jiang and H. Xu. Stochastic approximation approaches to the stochastic variational inequality problem. *IEEE Transactions on Automatic Control*, 53(6):1462–1475, 2008.

[Jin et al.(2022)] Y. Jin, A. Sidford, and K. Tian. Sharper rates for separable minimax and finite sum optimization via primal-dual extragradient methods. In *COLT*, pages 4362–4415. PMLR, 2022.

[Johnson and Zhang(2013)] R. Johnson and T. Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *NIPS*, pages 315–323, 2013.

[Juditsky et al.(2008)] A. Juditsky, P. Rigollet, A. B. Tsybakov, et al. Learning by mirror averaging. *The Annals of Statistics*, 36(5):2183–2206, 2008.

[Juditsky et al.(2011)] A. Juditsky, A. Nemirovski, and C. Tauvel. Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58, 2011.

[Jun and Orabona(2019)] K-S. Jun and F. Orabona. Parameter-free online convex optimization with sub-exponential noise. In *COLT*, pages 1802–1823. PMLR, 2019.

[Jun et al.(2017a)] K-S. Jun, F. Orabona, S. Wright, and R. Willett. Improved strongly adaptive online learning using coin betting. In *AISTATS*, pages 943–951. PMLR, 2017a.

[Jun et al.(2017b)] K-S. Jun, F. Orabona, S. Wright, and R. Willett. Online learning for changing environments using coin betting. *Electronic Journal of Statistics*, 11:5282–5310, 2017b.

[Kalhan et al.(2021)] D. S. Kalhan, A. S. Bedi, A. Koppel, K. Rajawat, H. Hassani, A. K. Gupta, and A. Banerjee. Dynamic online learning via Frank-Wolfe algorithm. *IEEE Transactions on Signal Processing*, 69:932–947, 2021.

[Kannan and Shanbhag(2019)] A. Kannan and U. V. Shanbhag. Optimal stochastic extragradient schemes for pseudomonotone stochastic variational inequality problems and their variants. *Computational Optimization and Applications*, 74(3):779–820, 2019.

[Kingma and Ba(2015)] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, pages 1–15, 2015. URL https://openreview.net/forum?id=8gmWwjFyLj.

[Kivinen and Warmuth(1999)] J. Kivinen and M. K. Warmuth. Averaging expert predictions. In *EuroCOLT*, pages 153–167. Springer, 1999.

[Koren and Levy(2015)] T. Koren and K. Levy. Fast rates for exp-concave empirical risk minimization. In *NeurIPS*, pages 1477–1485, 2015.

[Koshal et al.(2012)] J. Koshal, A. Nedić, and U. V. Shanbhag. Regularized iterative stochastic approximation methods for stochastic variational inequality problems. *IEEE Transactions on Automatic Control*, 58(3):594–609, 2012.

[Kotsalis et al.(2022)] G. Kotsalis, G. Lan, and T. Li. Simple and optimal methods for stochastic variational inequalities, I: Operator extrapolation. *SIAM Journal on Optimization*, 32(3):2041–2073, 2022.

[Krichene et al.(2015)] S. Krichene, W. Krichene, R. Dong, and A. Bayen. Convergence of heterogeneous distributed learning in stochastic routing games. In *Allerton*, pages 480–487. IEEE, 2015.

[Lei and Jordan(2017)] L. Lei and M. I. Jordan. Less than a single pass: Stochastically controlled stochastic gradient. In *AISTATS*, pages 148–156. PMLR, 2017.

[Lei and Jordan(2020)] L. Lei and M. I. Jordan. On the adaptivity of stochastic gradient-based optimization. *SIAM Journal on Optimization*, 30(2):1473–1500, 2020.

[Levy(2017)] K. Y. Levy. Online to offline conversions, universality and adaptive minibatch sizes. In *NeurIPS*, pages 1612–1621, 2017.

[Levy et al.(2018)] K. Y. Levy, A. Yurtsever, and V. Cevher. Online adaptive methods, universality and acceleration. In *NeurIPS*, pages 6501–6510, 2018.

[Li and Orabona(2019)] X. Li and F. Orabona. On the convergence of stochastic gradient descent with adaptive stepsizes. In *AISTATS*, pages 983–992. PMLR, 2019.

[Lin et al.(2020)] T. Lin, Z. Zhou, P. Mertikopoulos, and M. I. Jordan. Finite-time last-iterate convergence for multi-agent learning in games. In *ICML*, pages 6161–6171. PMLR, 2020.

[Lin et al.(2021)] T. Lin, Z. Zhou, W. Ba, and J. Zhang. Optimal no-regret learning in strongly monotone games with bandit feedback. *ArXiv Preprint: 2112.02856*, 2021.

[Loizou et al.(2021)] N. Loizou, H. Berard, G. Gidel, I. Mitliagkas, and S. Lacoste-Julien. Stochastic gradient descent-ascent and consensus optimization for smooth games: Convergence analysis under expected co-coercivity. In *NeurIPS*, pages 19095–19108, 2021.

[Lu et al.(2023)] Z. Lu, N. Brukhim, P. Gradu, and E. Hazan. Projection-free adaptive regret with membership oracles. In *ALT*, pages 1055–1073. PMLR, 2023.

[Luo et al.(2016)] H. Luo, A. Agarwal, N. Cesa-Bianchi, and J. Langford. Efficient second order online learning by sketching. In *NeurIPS*, pages 902–910, 2016.

[Mahdavi et al.(2015)] M. Mahdavi, L. Zhang, and R. Jin. Lower and upper bounds on the generalization of stochastic exponentially concave optimization. In *COLT*, pages 1305–1320. PMLR, 2015.

[Mertikopoulos and Zhou(2019)] P. Mertikopoulos and Z. Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1-2):465–507, 2019.

[Mertikopoulos et al.(2018)] P. Mertikopoulos, C. Papadimitriou, and G. Piliouras. Cycles in adversarial regularized learning. In *SODA*, pages 2703–2717. SIAM, 2018.

[Mertikopoulos et al.(2019)] P. Mertikopoulos, B. Lecouat, H. Zenati, C-S. Foo, V. Chandrasekhar, and G. Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra(-gradient) mile. In *ICLR*, pages 1–23, 2019. URL https://openreview.net/forum?id=Bkg8jjC9KQ.

[Mokhtari et al.(2016)] A. Mokhtari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro. Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *CDC*, pages 7195–7201. IEEE, 2016.

[Monnot and Piliouras(2017)] B. Monnot and G. Piliouras. Limits and limitations of no-regret learning in games. *The Knowledge Engineering Review*, 32, 2017.

[Mukkamala and Hein(2017)] M. C. Mukkamala and M. Hein. Variants of RMSProp and Adagrad with logarithmic regret bounds. In *ICML*, pages 2545–2553. PMLR, 2017.

[Nemirovski(2004)] A. Nemirovski. Prox-method with rate of convergence o(1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

[Nemirovski and Yudin(1983)] A. Nemirovski and D. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, 1983.

[Nesterov(2018)] Y. Nesterov. *Lectures on Convex Optimization*, volume 137. Springer, 2018.

[Netessine and Rudi(2003)] S. Netessine and N. Rudi. Centralized and competitive inventory models with demand substitution. *Operations Research*, 51(2):329–335, 2003.

[Nguyen et al.(2022)] L. M. Nguyen, M. van Dijk, D. T. Phan, P. H. Nguyen, T-W. Weng, and J. R. Kalagnanam. Finite-sum smooth optimization with SARAH. *Computational Optimization and Applications*, 82(3):561–593, 2022.

[Orabona and Pál(2016)] F. Orabona and D. Pál. Coin betting and parameter-free online learning. In *NeurIPS*, pages 577–585, 2016.

[Orda et al.(1993)] A. Orda, R. Rom, and N. Shimkin. Competitive routing in multiuser communication networks. *IEEE/ACM Transactions on Networking*, 1(5):510–521, 1993.

[Ordentlich and Cover(1998)] E. Ordentlich and T. M. Cover. The cost of achieving the best portfolio in hindsight. *Mathematics of Operations Research*, 23(4):960–982, 1998.

[Osborne and Rubinstein(1994)] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[Pal et al.(2018)] S. Pal, T-K. L. Wong, et al. Exponentially concave functions and a new information geometry. *The Annals of Probability*, 46(2):1070–1113, 2018.

[Polyak(1987)] B. Polyak. *Introduction to Optimization*. Optimization Software, 1987.

[Rappaport(2001)] T. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall PTR, 2001.

[Reddi et al.(2018)] S. J. Reddi, S. Kale, and S. Kumar. On the convergence of Adam and beyond. In *ICLR*, pages 1–23, 2018. URL `https://openreview.net/forum?id=ryQu7f-RZ`.

[Rockafellar and Sun(2019)] R. T. Rockafellar and J. Sun. Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging. *Mathematical Programming*, 174(1-2): 453–471, 2019.

[Rockafellar and Wets(2017)] R. T. Rockafellar and R. J. B. Wets. Stochastic variational inequalities: Single-stage to multistage. *Mathematical Programming*, 165(1):331–360, 2017.

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

[Roux et al.(2012)] N. L. Roux, M. Schmidt, and F. Bach. A stochastic gradient method with an exponential convergence rate for finite training sets. In *NIPS*, pages 2663–2671, 2012.

[Sandholm(2015)] W. H. Sandholm. Population games and deterministic evolutionary dynamics. In *Handbook of Game Theory with Economic Applications*, volume 4, pages 703–778. Elsevier, 2015.

[Shalev-Shwartz(2012)] S. Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

[Shanbhag(2013)] U. V. Shanbhag. Stochastic variational inequality problems: Applications, analysis, and algorithms. In *Theory Driven by Influential Applications*, pages 71–107. INFORMS, Catonsville, MD, 2013.

[Shoham and Leyton-Brown(2008)] Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2008.

[Sorin and Wan(2016)] S. Sorin and C. Wan. Finite composite games: Equilibria and dynamics. *Journal of Dynamics and Games*, 3(1):101–120, 2016.

[Tan(2014)] C. W. Tan. Wireless network optimization by Perron-Frobenius theory. In *CISS*, pages 1–6. IEEE, 2014.

[Vaswani et al.(2019)] S. Vaswani, A. Mishkin, I. Laradji, M. Schmidt, G. Gidel, and S. Lacoste-Julien. Painless stochastic gradient: interpolation, line-search, and convergence rates. In *NIPS*, pages 3732–3745, 2019.

[Villani(2009)] C. Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2009.

[Viossat and Zapechelnyuk(2013)] Y. Viossat and A. Zapechelnyuk. No-regret dynamics and fictitious play. *Journal of Economic Theory*, 148(2):825–842, 2013.

[Vovk(1997)] V. Vovk. Competitive online linear regression. In *NIPS*, pages 364–370, 1997.

[Wan et al.(2021)] Y. Wan, B. Xue, and L. Zhang. Projection-free online learning in dynamic environments. In *AAAI*, pages 10067–10075, 2021.

[Wan et al.(2023)] Y. Wan, L. Zhang, and M. Song. Improved dynamic regret for online Frank-Wolfe. *ArXiv Preprint: 2302.05620*, 2023.

[Wang et al.(2020)] G. Wang, S. Lu, Q. Cheng, W. Tu, and L. Zhang. SAdam: A variant of Adam for strongly convex functions. In *ICLR*, pages 1–21, 2020. URL https://openreview.net/forum?id=rye5YaEtPr.

[Ward et al.(2019)] R. Ward, X. Wu, and L. Bottou. AdaGrad stepsizes: Sharp convergence over nonconvex landscapes. In *ICML*, pages 6677–6686. PMLR, 2019.

[Weeraddana et al.(2012)] P. C. Weeraddana, M. Codreanu, M. Latva-aho, A. Ephremides, and C. Fischione. *Weighted Sum-Rate Maximization in Wireless Networks: A Review*. Now Foundations and Trends, 2012.

[Xu et al.(2017)] Y. Xu, Q. Lin, and T. Yang. Adaptive SVRG methods under error bound conditions with unknown growth parameter. In *NIPS*, pages 3279–3289, 2017.

[Yang et al.(2016)] T. Yang, L. Zhang, R. Jin, and J. Yi. Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient. In *ICML*, pages 449–457. PMLR, 2016.

[Yang et al.(2018)] T. Yang, Z. Li, and L. Zhang. A simple analysis for exp-concave empirical minimization with arbitrary convex regularizer. In *AISTATS*, pages 445–453, 2018.

[Yousefian et al.(2013)] F. Yousefian, A. Nedić, and U. V. Shanbhag. A regularized smoothing stochastic approximation (RSSA) algorithm for stochastic variational inequality problems. In *WSC*, pages 933–944. IEEE, 2013.

[Yousefian et al.(2014)] F. Yousefian, A. Nedić, and U. V. Shanbhag. Optimal robust smoothing extra-gradient algorithms for stochastic variational inequality problems. In *CDC*, pages 5831–5836. IEEE, 2014.

[Yu et al.(2022)] Y. Yu, T. Lin, E. Mazumdar, and M. I. Jordan. Fast distributionally robust learning with variance-reduced min-max optimization. In *AISTATS*, pages 1219–1250. PMLR, 2022.

[Zhang et al.(2018a)] L. Zhang, S. Lu, and Z-H. Zhou. Adaptive online learning in dynamic environments. In *NIPS*, pages 1330–1340, 2018a.

[Zhang et al.(2018b)] L. Zhang, T. Yang, and Z-H. Zhou. Dynamic regret of strongly adaptive methods. In *ICML*, pages 5882–5891. PMLR, 2018b.

[Zhang et al.(2019)] L. Zhang, T-Y. Liu, and Z-H. Zhou. Adaptive regret of convex and smooth functions. In *ICML*, pages 7414–7423. PMLR, 2019.

[Zhou et al.(2017)] Z. Zhou, P. Mertikopoulos, A. L. Moustakas, N. Bambos, and P. Glynn. Mirror descent learning in continuous games. In *CDC*, pages 5776–5783. IEEE, 2017.

[Zhou et al.(2018)] Z. Zhou, P. Mertikopoulos, S. Athey, N. Bambos, P. Glynn, and Y. Ye. Learning in games with lossy feedback. In *NIPS*, pages 1–11, 2018.

[Zhou et al.(2021)] Z. Zhou, P. Mertikopoulos, A. L. Moustakas, N. Bambos, and P. Glynn. Robust power management via learning and game design. *Operations Research*, 69(1):331–345, 2021.

[Zinkevich(2003)] M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *ICML*, pages 928–936, 2003.

[Zou et al.(2019)] F. Zou, L. Shen, Z. Jie, W. Zhang, and W. Liu. A sufficient condition for convergences of Adam and RMSProp. In *CVPR*, pages 11127–11135, 2019.

A Related Work

In both single-agent online learning and offline optimization, considerable attention has been paid to the development of **_adaptive gradient-based schemes_**. In particular, [Duchi et al.(2011)] presented an adaptive gradient algorithm (known as AdaGrad) for online learning with convex cost functions that updates the step sizes without needing to know the problem parameters. This algorithm is guaranteed to achieve a minimax-optimal regret of $O(\sqrt{T})$. Subsequently, Adam was proposed in the offline optimization setting to further exploit geometric aspects of iterate trajectories, exhibiting better empirical convergence performance [Kingma and Ba(2015)]. Theoretical guarantees have been obtained for Adam and other related adaptive algorithms in both offline optimization and online learning settings [Reddi et al.(2018), Zou et al.(2019)]. In parallel, the norm version of AdaGrad was developed and theoretical guarantees were established for related convex and/or nonconvex optimization problems [Levy(2017), Levy et al.(2018), Ward et al.(2019), Li and Orabona(2019)].

This adaptive family of algorithms has also been studied in the online learning literature under an assumption of strong convexity [Mukkamala and Hein(2017), Wang et al.(2020)]. The algorithms are guaranteed to achieve a minimax-optimal regret of $O(\log(T))$. However, unlike in the convex setting, these adaptive algorithms are not feasible in practice since they often require knowledge of the strong convexity parameter. In the offline optimization setting, the gradient-based methods can be made adaptive to the strong convexity parameter by exploiting the Polyak stepsize [Polyak(1987), Hazan and Kakade(2019)]. However, these algorithms do not extend readily to online learning since the sub-optimality gap is not well-defined. A recent line of research has shown that adaptive algorithms can be designed in the finite-sum setting, where they exhibit adaptivity to the strong convexity parameter [Roux et al.(2012), Defazio et al.(2014), Xu et al.(2017), Lei and Jordan(2017), Lei and Jordan(2020), Vaswani et al.(2019), Nguyen et al.(2022)]. In particular, [Lei and Jordan(2017), Lei and Jordan(2020)] showed that the use of random, geometrically-distributed epoch length yields full adaptivity in variance-controlled stochastic optimization under an assumption of strong convexity. However, these algorithm are not no-regret and thus their strategies do not extend readily to the online setting.

In terms of the last-iterate convergence to Nash equilibria, due to the difficulties mentioned earlier, much of the existing literature provides only qualitative convergence guarantees for non-adaptive (and hence infeasible) no-regret learning algorithms for various games [Krichene et al.(2015), Balandat et al.(2016), Zhou et al.(2017), Zhou et al.(2018), Mertikopoulos et al.(2019), Mertikopoulos and Zhou(2 More recently, finite-time last-iterate convergence rates have been obtained for specially structured games, such as strongly monotone games [Zhou et al.(2021), Loizou et al.(2021)], unconstrained cocoercive games [Lin et al.(2020)], unconstrained smooth games [Golowich et al.(2020)] and constrained smooth games [Cai et al.(2022)]. Except for a class of strongly monotone games, the last-iterate convergence rate is measured in metrics other than $\|x^t - x^\star\|_2^2$. Further, among these results, only [Lin et al.(2020)] provides an adaptive online learning algorithm that does not require knowing the cocoercivity parameter. However, their algorithm falls short in two respects: (i) it may not be no-regret; (ii) each agent needs to know all other agents' gradients, thus again rendering it infeasible in practice. Recently, [Antonakopoulos et al.(2021)] has developed an adaptive extragradient algorithm for strictly monotone games that converges asymptotically in a last-iterate sense to the unique Nash equilibrium. However, their algorithm also requires each agent to know all others' gradients and the no-regret property cannot be guaranteed; indeed, the original extra-gradient algorithm was shown to not be no-regret [Golowich et al.(2020)]. [Hsieh et al.(2021)] has proposed a no-regret adaptive online learning algorithm based on optimistic mirror descent and established a regret of $O(\sqrt{T})$ for convex cost functions. They also proved asymptotic last-iterate

convergence to the unique Nash equilibrium for strictly variationally stable games (a superset of strictly monotone games).

Another line of relevant literature focuses on stochastic approximation methods for solving strongly monotone VIs. An early proposal using such an approach was presented by [Jiang and Xu(2008)], who proposed a stochastic projection method for solving strongly monotone VIs with an almost-sure convergence guarantee. [Koshal et al.(2012)] and [Yousefian et al.(2013)] proposed various regularized iterative stochastic approximation methods for solving monotone VIs and also established almost-sure convergence. A survey of these methods, as well as applications and the theory behind stochastic VI, can be found in [Shanbhag(2013)]. [Juditsky et al.(2011)] was among the first to establish an iteration complexity bound for stochastic VI methods by extending the mirror-prox method [Nemirovski(2004)] to stochastic setting. [Yousefian et al.(2014)] further extended the stochastic mirror-prox method with a more general step size choice and proved the same iteration complexity. They also proved an improved complexity bound for the stochastic extragradient method for solving strongly monotone VIs. [Chen et al.(2017)] studied a specific class of VIs and proposed a method that combines the stochastic mirror-prox method with Nesterov's acceleration [Nesterov(2018)], resulting in an optimal iteration complexity for such problem class. [Kannan and Shanbhag(2019)] analyzed a general variant of an extragradient method (which uses general distance-generating functions) and proved an optimal iteration bound under a slightly weaker assumption than strong monotonicity. Several other stochastic methods have also been shown to yield an optimal iteration bound for solving strongly monotone VIs [Kotsalis et al.(2022), Huang and Zhang(2022)]. In recent years, there have been developments in variance-reduction-based methods [Balamurugan and Bach(2016), Iusem et al.(2017), Iusem et al.(2019), Jalilzadeh and Shanbhag(20 Yu et al.(2022), Alacaoglu and Malitsky(2022), Jin et al.(2022), Huang et al.(2022)]. In the line of research aiming to model multistage stochastic VI (as compared to the single-stage VI considered in the above literature), the dynamics between the actions and the arrival of future information plays a central role. For details regarding multistage stochastic VI, we refer to [Rockafellar and Wets(2017)] and [Rockafellar and Sun(2019)].

We are also aware of two related topics: (1) parameter-free online learning [Foster et al.(2015), Foster et al.(2017), Orabona and Pál(2016), Cutkosky and Orabona(2018), Jun and Orabona(2019), Cutkosky(2020a), Cutkosky(2020b)] and (2) online learning with adaptive/dynamic regret [Zinkevich(2003), Hazan and Seshadhri(2007), Daniely et al.(2015), Hazan(2016)]. The former line of works consider an alternative form of regret:

$$\text{P-Regret}_T(x) = \sum_{t=1}^{T} f_t(x^t) - \sum_{t=1}^{T} f_t(x)$$

where $x \in \mathcal{X}$ is an unknown **competitor**. The goal is to achieve expected regret bounds that have optimal dependency not only on T but also $\|x\|$. Notably, their framework provides a way to design algorithms that achieve minimax-optimal regret with respect to any competitor, without imposing a bounded set for the competitor nor any parameter to tune in online convex optimization. However, the strong convexity parameter and the exp-concavity parameter both characterize **the lower bound for curvature information**. Estimating these quantities will require new techniques. The second line of works consider two other form of regrets:

$$\text{D-Regret}_T(x_1, \ldots, x_T) = \sum_{t=1}^{T} f_t(x^t) - \sum_{t=1}^{T} f_t(x_t) \qquad \text{(dynamic)}$$

25

where $x_1, \ldots, x_T \in \mathcal{X}$ are any competitors, and

$$\text{A-Regret}_T(\tau) = \max_{1 \le s \le T+1-\tau} \left\{ \sum_{t=s}^{s+\tau-1} f_t(x^t) - \min_{x \in \mathcal{X}} \left\{ \sum_{t=s}^{s+\tau-1} f_t(x) \right\} \right\} \qquad \text{(adaptive)}$$

which is the maximum static regret over an interval with the length τ. In this context, it is impossible to obtain a sublinear dynamic regret with arbitrarily varying sequences and we often assume certain conditions, such as an upper bound for path-length $P_T = \sum_{t=2}^{T} \|x_{t-1} - x_t\|$. The seminar work of [Zinkevich(2003)] established the first general-case bound of $O(\sqrt{T}(1 + P_T))$ for OGD and [Zhang et al.(2018b), Zhang et al.(2018a)] improved it to $O(\sqrt{T(1 + P_T)})$. Several recent studies have further investigated the dynamic regret by leveraging the curvature of loss functions, such as exponential concavity [Baby and Wang(2021)] and strong convexity [Baby and Wang(2022)]. There is also the worst-case variant of dynamic regret: D-Regret$_T(x_1, \ldots, x_T) = \sum_{t=1}^{T} f_t(x^t) - \sum_{t=1}^{T} f_t(x_t^\star)$ where $x_t^\star = \operatorname{argmin}_{x \in \mathcal{X}} f_t(x_t)$ and has been extensively studied in operations research and computer science [Besbes et al.(2015), Jadbabaie et al.(2015), Mokhtari et al.(2016), Yang et al.(2016)]. Moreover, the previous works in adaptive regret minimization mainly focus on the setting of online convex optimization [Hazan and Seshadhri(2007), Daniely et al.(2015), Jun et al.(2017b), Jun et al.(2017a), Zhang et al.(2019)]. The seminar work of [Hazan and Seshadhri(2007)] first introduced the notion of adaptive regret in a weak form and proposed an algorithm with an $O(d \log^2(T))$ adaptive regret bound for exponentially concave losses. However, the weak form of adaptive regret could be dominated by long intervals and hence, cannot respect short intervals well. This issue was addressed by [Daniely et al.(2015)] who put forth the strongly adaptive regret A-Regret$_T(\cdot)$ and design a two-layer algorithm which combines OGD and the geometric covering intervals construction. In addition, a few projection-free online algorithms were developed to investigate adaptive/dynamic regret minimization [Kalhan et al.(2021), Wan et al.(2021), Wan et al.(2023), Garber and Kretzu(2022), Lu et al.(2023)]. However, all of these algorithms require knowing the strong convexity parameter and the exp-concavity parameters.

In summary, the possibility of designing an online algorithm that is both ***doubly optimal*** and ***feasible*** under strong convexity still remains open.

B Applications

B.1 Application of **AdaOGD**: Feasible Learning for Newsvendors with Lost Sales

The single-retailer version of the newsvendor problem is a well known model for perishable inventory control [Huh and Rusmevichientong(2009)]. The assumption in this model is that unsold inventory perishes at the end of each period. A retailer sells a product over a time horizon T and then makes inventory-ordering decisions $x^t \in [0, \bar{x}]$ at the beginning of each period t to maximize the profit. The unknown demand D^t is random and only realized with a value d^t after the retailer makes her decision. It is often assumed in the inventory control literature that the D^t are independent, corresponding to a stationary environment. Here, we do not need to make this assumption and we allow D^t to be arbitrary. Further, in the lost-sales setting, any unmet demand is lost and hence the retailer does *not* observe d^t; instead, she only observes the sales quantity $\min\{x^t, d^t\}$. The retailer's cost functions are defined by

$$f_t(x^t) = (p - c) \cdot \mathbb{E}[\max\{0, D^t - x^t\}] + c \cdot \mathbb{E}[\max\{0, x^t - D^t\}] \qquad (9)$$

Algorithm 5 Newsvendor-AdaOGD(x^1, T)

1: **Input:** initial point $x^1 \in \mathcal{X}$ and the total number of rounds T
2: **Initialization:** $p_0 = \frac{1}{\log(T+10)}$
3: **for** $t = 1, 2, \ldots, T$ **do**
4:     sample $M^t \sim \mathsf{Geometric}(p_0)$
5:     set $\eta^{t+1} \leftarrow \frac{t+1}{\sqrt{1+\max\{M^1, \ldots, M^t\}}}$
6:     observe the sales quantity of $\min\{x^t, d^t\}$
7:     update $x^{t+1} \leftarrow \begin{cases} \operatorname{argmin}_{x \in [0,\bar{x}]} \left\{ (x - x^t)c + \frac{\eta^{t+1}}{2}(x - x^t)^2 \right\} & \text{if } x^t \geq \min\{x^t, d^t\} \\ \operatorname{argmin}_{x \in [0,\bar{x}]} \left\{ (x - x^t)(c - p) + \frac{\eta^{t+1}}{2}(x - x^t)^2 \right\} & \text{otherwise} \end{cases}$
8: **end for**

---

where the unit purchase cost is $c > 0$ and the unit selling price is $p \geq c$. It is known that minimizing this cost is equivalent to maximizing the profit, $\mathbb{E}[p \cdot \min\{x^t, D^t\} - c \cdot x^t]$, where:

$$\mathbb{E}[p \cdot \min\{x^t, D^t\} - c \cdot x^t]$$
$$= \mathbb{E}\left[p \cdot \left(D^t - \max\{0, D^t - x^t\} \right) - c \cdot \left(D^t - \max\{0, D^t - x^t\} + \max\{0, x^t - D^t\} \right) \right]$$
$$= (p - c) \cdot \mathbb{E}[D^t] - (p - c) \cdot \mathbb{E}[\max\{0, D^t - x^t\}] - c \cdot \mathbb{E}[\max\{0, x^t - D^t\}]$$
$$= (p - c) \cdot \mathbb{E}[D^t] - f_t(x^t)$$

Note that the first term $(p - c) \cdot \mathbb{E}[D^t]$ is independent of x^t. Thus, the maximization of the profit $\mathbb{E}[p \cdot \min\{x^t, D^t\} - c \cdot x^t]$ is equivalent to minimizing the cost $f_t(\cdot)$ in Eq. (9).

In this context, [Huh and Rusmevichientong(2009)] have shown that the cost function $f_t(\cdot)$ is convex in general and αp-strongly convex if the demand is a random variable with a continuous density function q such that $\inf_{d \in [0,\bar{x}]} q(d) \geq \alpha > 0$. Moreover, the retailer only observes the sales quantity $\min\{x^t, d^t\}$ where d^t is a realization of D^t. Thus, the (noisy) bandit feedback is not observable. However, a noisy gradient feedback signal can be obtained:

$$\xi^t = \begin{cases} c & \text{if } x^t \geq \min\{x^t, d^t\} \\ c - p & \text{otherwise} \end{cases}$$

which is an unbiased and bounded gradient estimator. In this setting, the parameter $\alpha > 0$ is not available since the distribution of the demand is unknown. However, the retailer can apply our AdaOGD algorithm (cf. Algorithm 1) and obtain a near-optimal regret of $O(\log^2(T))$.

We specialize Algorithm 1 to the newsvendor problem in Algorithm 5 and we present the corresponding result on the regret minimization property in the following corollary.

Corollary 1. *In the single-retailer newsvendor problem, the retailer's cost functions are defined by Eq. (9) where the unit purchase cost is $c > 0$ and the unit selling price is $p \geq c$. Also, the demand is a random variable with a continuous density function q such that $\inf_{d \in [0,\bar{x}]} q(d) \geq \alpha > 0$. If the agent employs Algorithm 5, we have*

$$\mathbb{E}[Regret(T)] \leq \frac{\bar{x}^2}{2}\left(1 + e^{\frac{1}{(\alpha p)^2 \log(T+10)}}\right) + \frac{p^2 \log(T+1)}{2}\sqrt{1 + \log(T+10) + \log(T)\log(T+10)}$$

As a consequence, we have $\mathbb{E}[Regret(T)] = O(\log^2(T))$.

Proof. Recall that $f_t(x^t) = (p-c) \cdot \mathbb{E}[\max\{0, D^t - x^t\}] + c \cdot \mathbb{E}[\max\{0, x^t - D^t\}]$ and the noisy gradient feedback is given by

$$\xi^t = \begin{cases} c & \text{if } x^t \geq \min\{x^t, d^t\} \\ c - p & \text{otherwise} \end{cases}$$

So we have $\mathbb{E}[\xi^t \mid x^t] = \nabla f_t(x^t)$ and $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq p^2$ for all $t \geq 1$. Since the demand has a continuous density function q such that $\inf_{d \in [0, \bar{x}]} q(d) \geq \alpha > 0$, we have $f_t(\cdot)$ is αp-strongly convex. In addition, $\mathcal{X} = [0, \bar{x}]$ implies that $D = \bar{x}$. Thus, Theorem 1 can be applied and implies the desired result. $\square$

B.2 Applications of MA-AdaOGD: Feasible Multi-Agent Learning for Power Management and Newsvendors with Lost Sales

We present two typical examples that satisfy the conditions in Definition 3, where (noisy) gradient feedback is available even when (noisy) bandit feedback is not available.

Example 1 (Power Management in Wireless Networks)**.** Target-rate power management problems are well known in operations research and wireless communications [Rappaport(2001), Goldsmith(2005)]. We consider a wireless network of N communication links, each link consisting of a transmitter (e.g., phone, tablet and sensor) and an intended receiver (each transmitter consumes power to send signals to their intended receivers). Assume further that the transmitter in the ith link transmits with power $a_i \geq 0$ and let $a = (a_1, \ldots, a_N) \in \mathbb{R}^N$ denote the joint power profile of all transmitters in the network. In this context, the quality-of-service rate of the ith link depends not only on how much power its transmitter is employing but also on how much power all the other transmitters are concurrently employing. Formally, the ith link's quality-of-service rate is given by $r_i(a) = \frac{G_{ii}a_i}{\sum_{j \neq i} G_{ij}a_j + \eta_i}$ where η_i is the thermal noise associated with the receiver of the ith link and $G_{ij} \geq 0$ is the unit power gain between the transmitter in the jth link and the receiver in the ith link, which is determined by the network topology but is unknown in practice. Note that $\sum_{j \neq i} G_{ij}a_j$ is the interference caused by other links to link i—all else being equal, the larger the powers of transmitters in other links, the lower the service rate the ith link. Intuitively, the transmitter in the ith link aims to balance between two objectives: maintaining a target service rate $r_i^\star$ while consuming as little power as possible. This consideration leads to the following standard cost function [Rappaport(2001), Goldsmith(2005)]:

$$u_i(a) = \frac{a_i^2}{2}\left(1 - \frac{r_i^\star}{r_i(a)}\right)^2 = \frac{1}{2}\left(a_i - \frac{r_i^\star(\sum_{j \neq i} G_{ij}a_j + \eta_i)}{G_{ii}}\right)^2$$

Notably, $u_i(a) = 0$ if the realized service rate $r_i(a)$ is equal to the target rate $r_i^\star$. Otherwise, there will be a cost either due to not meeting $r_i^\star$ or due to consuming unnecessary power. Our prior work has shown that this is a β-strongly monotone game [Zhou et al.(2021)] with $\beta = \lambda_{\min}(I - \frac{1}{2}(W + W^\top)) > 0$, where $W_{ii} = 0$ for all $1 \leq i \leq N$ and $W_{ij} = \frac{r_i^\star G_{ij}}{G_{ii}}$ for $i \neq j$.

Finally, gradient feedback is available for the aforementioned class of strongly monotone games; indeed, the ith link's quality-of-service rate $r_i(\mathbf{a})$ is available to the transmitter in the ith link, who can use it to compute the gradient of $u_i(\mathbf{a})$ with respect to a_i. In this setting, the parameter $\beta > 0$ is not available in practice since finding the minimal eigenvalue is computationally prohibitive. However, the retailer can apply our MA-AdaOGD algorithm (cf. Algorithm 2) and obtain a near-optimal last-iterate convergence rate of $O(\frac{\log^3(T)}{T})$.

Example 2 (Newsvendor with Lost Sales). We consider the multi-retailer generalization of newsvendor problem [Netessine and Rudi(2003)]. For simplicity, we focus on a single product with same per-unit price $p > 0$ and perishable inventory control. Given the set of retailers $\mathcal{N} = \{1, 2, \ldots, N\}$, the ith retailer's action x_i is assumed to lie in the interval $[0, \bar{x}_i]$. For the ith retailer, the demand is random and depends on the inventory levels of other retailers, therefore we denote it as $D_i(x_{-i})$ and let $d_i \geq 0$ denote a realization of this random variable. The ith retailer only observes the sales quantity $\min\{x_i, d_i\}$. Using the previous argument, the ith retailer's cost function is defined by

$$u_i(x) = (p - c_i) \cdot \mathbb{E}[\max\{0, D_i(x_{-i}) - x_i\}] + c_i \cdot \mathbb{E}[\max\{0, x_i - D_i(x_{-i})\}] \tag{10}$$

and the noisy gradient feedback signal is given by

$$\xi_i = \begin{cases} c_i & \text{if } x_i \geq \min\{x_i, d_i\} \\ c_i - p & \text{otherwise} \end{cases}$$

It remains to extend the analysis from [Huh and Rusmevichientong(2009), Section 3.5] and provide a condition on the distribution of $D_i(x_{-i})$ that can guarantee that the multi-retailer newsvendor problem is β-strongly monotone for some constant $\beta > 0$. Indeed, after simple calculations, we have

$$v_i(x) = \nabla_{x_i} u_i(x) = p \cdot \mathbb{P}(D_i(x_{-i}) \leq x_i) - p + c_i = p \cdot F_i(x) - p + c_i$$

Letting $F = (F_1, F_2, \ldots, F_N)$ be an operator from $\prod_{i=1}^{N}[0, \bar{x}_i]$ to $[0, 1]^N$, if F is α-strongly monotone, we have

$$\langle x' - x, v(x') - v(x) \rangle = p \cdot \langle x' - x, F(x') - F(x) \rangle \geq \alpha p \|x' - x\|^2, \quad \text{for all } x, x' \in \prod_{i=1}^{N}[0, \bar{x}_i]$$

As a concrete example, we can let the distribution of a demand $D_i(x_{-i})$ be given by

$$\mathbb{P}(D_i(x_{-i}) \leq z) = 1 - \frac{1 + \sum_{j \neq i} x_j}{(1 + z + \sum_{j \neq i} x_j)^2}$$

It is clear that $\lim_{z \to +\infty} \mathbb{P}(D_i(x_{-i}) \leq z) = 1$ for all $x_{-i} \in \prod_{j \neq i}[0, \bar{x}_j]$. Also, we have

$$\mathbb{P}(D_i(x_{-i}) = 0) = 1 - \frac{1}{1 + \sum_{j \neq i} x_j}$$

which characterizes the dependence of the distribution of the demand for the ith retailer on other retailers' actions x_{-i}. Indeed, the demand $D_i(x_{-i})$ is likely to be small if the total inventory provided by other retailers $\sum_{j \neq i} x_j$ is large. Two extreme cases are (i) $\mathbb{P}(D_i(x_{-i}) = 0) \to 1$ as $\sum_{j \neq i} x_j \to +\infty$ and (ii) $\mathbb{P}(D_i(x_{-i}) = 0) = 0$ as $\sum_{j \neq i} x_j = 0$. Finally, we can prove that this is a β-strongly monotone game with $\beta = \frac{p}{(1 + \sum_{k=1}^{N} \bar{x}_k)^3} > 0$; see Section B.3 for the details.

Thus, if all the retailers agree to apply our MA-AdaOGD algorithm (cf. Algorithm 2), we can obtain a near-optimal last-iterate convergence rate of $O(\frac{\log^3(T)}{T})$.

We summarize Algorithm 2 specialized to the multi-retailer generalization of the newsvendor problem in Algorithm 6 and we present the corresponding result on the last-iterate convergence rate guarantee in the following corollary.

Algorithm 6 Newsvendor-MA-AdaOGD$(x_1^1, x_2^1, \ldots, x_N^1, T)$

1: **Input:** initial point $x_i^1 \in \mathcal{X}_i$ for all $i \in \mathcal{N}$ and the total number of rounds T
2: **Initialization:** $p_0 = \frac{1}{\log(T+10)}$
3: **for** $t = 1, 2, \ldots, T$ **do**
4:     **for** $i = 1, 2, \ldots, N$ **do**
5:         sample $M_i^t \sim \mathsf{Geometric}(p_0)$
6:         set $\eta_i^{t+1} \leftarrow \frac{t+1}{\sqrt{1+\max\{M_i^1, \ldots, M_i^t\}}}$
7:         observe the sales quantity of $\min\{x^t, d^t\}$
8:         update $x_i^{t+1} \leftarrow \begin{cases} \operatorname{argmin}_{x_i \in [0, \bar{x}_i]} \left\{ (x_i - x_i^t)c_i + \frac{\eta_i^{t+1}}{2}(x_i - x_i^t)^2 \right\} & \text{if } x_i^t \geq \min\{x_i^t, d_i^t\} \\ \operatorname{argmin}_{x_i \in [0, \bar{x}_i]} \left\{ (x_i - x_i^t)(c_i - p) + \frac{\eta_i^{t+1}}{2}(x_i - x_i^t)^2 \right\} & \text{otherwise} \end{cases}$
9:     **end for**
10: **end for**

Corollary 2. *In the multi-retailer generalization of the newsvendor problem, each retailer's cost functions are defined by Eq. (10) where the unit purchase cost is $c_i > 0$ and the unit selling price is $p \geq c$. Also, the demand is a random variable with a continuous density function defined in Example 2. If the agent employs Algorithm 6, we have*

$$\mathbb{E}[\|x^T - x^\star\|^2] \leq \frac{\sum_{i=1}^N \bar{x}_i^2}{T}(1 + e^{\frac{1}{4\beta^2 \log(T+10)}})\sqrt{1 + \log(T+10) + \log(NT)\log(T+10)}$$
$$+ \frac{Np^2}{T}\log(T+1)(1 + \log(T+10) + \log(NT)\log(T+10))$$

As a consequence, we have $\mathbb{E}[\|x^T - x^\star\|^2] = O(\frac{\log^3(T)}{T})$.

Proof. Recall that $u_i(x^t) = (p - c_i) \cdot \mathbb{E}[\max\{0, D_i(x_{-i}^t) - x_i^t\}] + c_i \cdot \mathbb{E}[\max\{0, x_i^t - D_i(x_{-i}^t)\}]$ and the noisy gradient feedback is given by

$$\xi_i^t = \begin{cases} c_i & \text{if } x_i^t \geq \min\{x_i^t, d_i^t\} \\ c_i - p & \text{otherwise} \end{cases}$$

So we have $\mathbb{E}[\xi^t \mid x^t] = v(x^t)$ and $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq Np^2$ for all $t \geq 1$. In Section B.3, we show that this is a β-strongly monotone game with $\beta = \frac{p}{(1+\sum_{k=1}^N \bar{x}_k)^3} > 0$. In addition, $\mathcal{X}_i = [0, \bar{x}_i]$ implies that $D^2 = \sum_{i=1}^N \bar{x}_i^2$. Thus, Theorem 3 can be applied and implies the desired result. □

B.3 Proofs for Example 1 and 2

We show that POWER MANAGEMENT IN WIRELESS NETWORKS and NEWSVENDOR WITH LOST SALES in Example 1 and 2 are β-strongly monotone games (cf. Definition 3) for some $\beta > 0$.

Power Management in Wireless Networks. Example 1 satisfies Definition 3 with $\beta = \lambda_{\min}(I - \frac{1}{2}(W + W^\top)) > 0$ where $W_{ii} = 0$ for all $1 \leq i \leq N$ and $W_{ij} = \frac{r_i^\star G_{ij}}{G_{ii}}$ for $i \neq j$. An analysis of this setting has been given in [Zhou et al.(2021)]; we provide the details for completeness. The cost function is given by

$$u_i(a) = \frac{1}{2}\left(a_i - \frac{r_i^\star(\sum_{j \neq i} G_{ij}a_j + \eta_i)}{G_{ii}}\right)^2$$

Taking the derivative of $u_i(a)$ with respect to a_i yields that $v_i(a) = a_i - \frac{r_i^\star(\sum_{j \neq i} G_{ij} a_j + \eta_i)}{G_{ii}}$. Consequently, by the definition of W_{ij}, we have

$$
\begin{aligned}
\langle a' - a, v(a') - v(a) \rangle &= \|a' - a\|^2 - \sum_{i=1}^{N} \frac{r_i^\star}{G_{ii}} \sum_{j \neq i} G_{ij} \langle a_i' - a_i, a_j' - a_j \rangle \\
&= \|a' - a\|^2 - \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} \langle a_i' - a_i, a_j' - a_j \rangle \\
&= \|a' - a\|^2 - \langle a' - a, W(a' - a) \rangle = \langle a' - a, (I - \frac{1}{2}(W + W^\top))(a' - a) \rangle \\
&\geq \lambda_{\min}(I - \frac{1}{2}(W + W^\top)) \|a' - a\|^2 = \beta \|a' - a\|^2
\end{aligned}
$$

This yields the desired result.

Newsvendor with Lost Sales. Example 2 satisfies Definition 3 with $\beta = \alpha p$. Indeed, each players' reward function is given by

$$
u_i(x) = (p - c_i) \cdot \mathbb{E}[\max\{0, D_i(x_{-i}) - x_i\}] + c_i \cdot \mathbb{E}[\max\{0, x_i - D_i(x_{-i})\}]
$$

Equivalently, we have

$$
u_i(x) = (p - c_i) \cdot \int_{x_i}^{+\infty} (k - x_i) \cdot d\mathbb{P}(D_i(x_{-i}) \leq k) + c_i \cdot \int_{-\infty}^{x_i} (x_i - k) \cdot d\mathbb{P}(D_i(x_{-i}) \leq k)
$$

Using the Leibniz integral rule, we have

$$
v_i(x) = \nabla_{x_i} u_i(x) = p \cdot \mathbb{P}(D_i(x_{-i}) \leq x_i) - p + c_i = p \cdot F_i(x) - p + c_i
$$

Let $F = (F_1, F_2, \ldots, F_N)$ be an operator from $\prod_{i=1}^{N}[0, \bar{x}_i]$ to $[0, 1]^N$. Then, if F is α-strongly monotone, we have

$$
\langle x' - x, v(x') - v(x) \rangle = p \cdot \langle x' - x, F(x') - F(x) \rangle \geq \alpha p \|x' - x\|^2, \quad \text{for all } x, x' \in \prod_{i=1}^{N}[0, \bar{x}_i]
$$

Considering the example where the distribution of a demand $D_i(x_{-i})$ is given by

$$
\mathbb{P}(D_i(x_{-i}) \leq z) = 1 - \frac{1 + \sum_{j \neq i} x_j}{(1 + z + \sum_{j \neq i} x_j)^2}
$$

Then, we have

$$
v_i(x) = p \cdot \left(1 - \frac{1 + \sum_{j \neq i} x_j}{(1 + \sum_{i=1}^{N} x_i)^2} \right) - p + c_i
$$

The following proposition, a modification of [Rosen(1965), Theorem 6], plays an important role in the subsequent analysis.

31

Proposition 2. *Given a continuous game $\mathcal{G} = (\mathcal{N}, \mathcal{X} = \prod_{i=1}^N \mathcal{X}_i, \{u_i\}_{i=1}^N)$, where each u_i is twice continuously differentiable. For each $x \in \mathcal{X}$, we define the Hessian matrix $H(x)$ as follows:*

$$H_{ij}(x) = \frac{1}{2}\nabla_j v_i(x) + \frac{1}{2}(\nabla_i v_j(x))^\top$$

If $H(x)$ is positive definite for every $x \in \mathcal{X}$, we have $\langle x' - x, v(x') - v(x) \rangle \geq 0$ for all $x, x' \in \mathcal{X}$ where the equality holds true if and only if $x = x'$.

As a consequence of Proposition 2, we have $\langle x' - x, v(x) - v(x) \rangle \geq \beta \|x' - x\|^2$ for all $x, x' \in \mathcal{X}$ if $H(x) \succeq \beta I_N$ for all $x \in \mathcal{X}$. For our example, it suffices to show that

$$H(x) \succeq \left(\frac{p}{(1 + \sum_{i=1}^N \bar{x}_i)^3} \right) I_N, \quad \text{for all } x \in \prod_{i=1}^N [0, \bar{x}_i] \tag{11}$$

Indeed, after a straightforward calculation, we have

$$\nabla_i v_i(x) = \frac{2p(1 + \sum_{k \neq i} x_k)}{(1 + \sum_{k=1}^N x_k)^3}, \qquad \nabla_j v_i(x) = \frac{p(1 + \sum_{k \neq i} x_k - x_i)}{(1 + \sum_{k=1}^N x_k)^3}, \qquad \nabla_i v_j(x) = \frac{p(1 + \sum_{k \neq j} x_k - x_j)}{(1 + \sum_{k=1}^N x_k)^3}$$

This implies that

$$H(x) = \frac{p}{(1 + \sum_{i=1}^N x_i)^3} \cdot \begin{pmatrix} 2 + 2\sum_{i \neq 1} x_i & 1 + \sum_{i \neq 1,2} x_i & 1 + \sum_{i \neq 1,3} x_i & \cdots & 1 + \sum_{i \neq 1,N} x_i \\ 1 + \sum_{i \neq 1,2} x_i & 2 + 2\sum_{i \neq 2} x_i & 1 + \sum_{i \neq 2,3} x_i & \cdots & 1 + \sum_{i \neq 2,N} x_i \\ 1 + \sum_{i \neq 1,3} x_i & 1 + \sum_{i \neq 2,3} x_i & 2 + 2\sum_{i \neq 3} x_i & \cdots & 1 + \sum_{i \neq 3,N} x_i \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 + \sum_{i \neq 1,N} x_i & 1 + \sum_{i \neq 2,N} x_i & 1 + \sum_{i \neq 3,N} x_i & \cdots & 2 + 2\sum_{i \neq N} x_i \end{pmatrix}$$

Equivalently, we have

$$
\begin{aligned}
H(x) &= \frac{p}{(1 + \sum_{i=1}^N x_i)^3} \cdot \left\{ \left(1 + \sum_{i=1}^N x_i \right) \cdot \begin{pmatrix} 2 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 1 & \cdots & 1 \\ 1 & 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 2 \end{pmatrix} \right. \\
&\qquad \left. - \begin{pmatrix} 2x_1 & x_1 + x_2 & x_1 + x_3 & \cdots & x_1 + x_N \\ x_1 + x_2 & 2x_2 & x_2 + x_3 & \cdots & x_2 + x_N \\ x_1 + x_3 & x_2 + x_3 & 2x_3 & \cdots & x_3 + x_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1 + x_N & x_2 + x_N & x_3 + x_N & \cdots & 2x_N \end{pmatrix} \right\} \\
&= \frac{p}{(1 + \sum_{i=1}^N x_i)^3} \cdot \left\{ \left(1 + \sum_{i=1}^N x_i \right) \cdot (I_N + \mathbf{1}_N \mathbf{1}_N^\top) - x\mathbf{1}_N^\top - \mathbf{1}_N x^\top \right\}
\end{aligned}
$$

Note that $x \in \prod_{i=1}^N [0, \bar{x}_i]$. If $x = \mathbf{0}_N$, we have $H(x) = p \cdot (I_N + \mathbf{1}_N \mathbf{1}_N^\top)$ and thus satisfy Eq. (11). Otherwise, we let $y = \frac{x}{\sum_{i=1}^N x_k}$ and obtain that $\|y\| \leq 1$. It is clear that $(y - \mathbf{1}_N)(y - \mathbf{1}_N)^\top$ is positive semidefinite. Thus, we have

$$yy^\top + \mathbf{1}_N \mathbf{1}_N^\top \succeq y\mathbf{1}_N^\top + \mathbf{1}_N y^\top$$

Using the definition of y, we have $yy^\top \preceq I_N$. This together with the above inequality implies that

$$\left(\sum_{i=1}^N x_i\right) \cdot (I_N + \mathbf{1}_N \mathbf{1}_N^\top) \succeq x \mathbf{1}_N^\top + \mathbf{1}_N x^\top$$

Putting these pieces together yields that

$$H(x) \succeq \frac{p}{(1 + \sum_{i=1}^N x_i)^3} \cdot (I_N + \mathbf{1}_N \mathbf{1}_N^\top) \succeq \left(\frac{p}{(1 + \sum_{i=1}^N \bar{x}_i)^3}\right) I_N, \quad \text{for all } x \in \prod_{i=1}^N [0, \bar{x}_i]$$

This completes the proof.

C Deferred Proofs

C.1 Proof of Theorem 1

To prove Theorem 1, we present a descent inequality for the iterates generated by Algorithm 1.

Lemma 2. *For an arbitrary fixed sequence of β-strongly convex functions $f_1, \ldots, f_T$, where each f_t satisfies $\mathbb{E}[\xi^t \mid x^t] = \nabla f_t(x^t)$ and $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq G^2$ for all $t \geq 1$, and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. Letting the iterates $\{x^t\}_{t \geq 1}$ be generated by Algorithm 1, we have for all $x \in \mathcal{X}$*

$$\sum_{t=1}^T \mathbb{E}[f_t(x^t) - f_t(x)] \leq \frac{\eta^1}{2}\|x^1 - x\|^2 + \sum_{t=1}^T \mathbb{E}\left[\left(\frac{\eta^{t+1} - \eta^t}{2} - \frac{\beta}{2}\right)\|x^t - x\|^2\right] + \frac{G^2}{2}\left(\sum_{t=1}^T \mathbb{E}\left[\frac{1}{\eta^{t+1}}\right]\right)$$

The proof of this lemma is deferred to Section C.6.2.

Proof of Theorem 1. Recall that $\mathcal{X}$ is convex and bounded with a diameter $D > 0$ and we have $\eta^{t+1} = \frac{t+1}{\sqrt{1 + \max\{M^1, \ldots, M^t\}}}$ in Algorithm 1, we have

$$\frac{\eta^1}{2}\|x^1 - x\|^2 \leq \frac{D^2}{2}, \qquad \eta^{t+1} - \eta^t \leq \frac{1}{\sqrt{1 + \max\{M^1, \ldots, M^t\}}}$$

By Lemma 2, we have

$$\sum_{t=1}^T \mathbb{E}[f_t(x^t) - f_t(x)] \leq \frac{D^2}{2} + \sum_{t=1}^T \mathbb{E}\left[\left(\frac{1}{2\sqrt{1 + \max\{M^1, \ldots, M^t\}}} - \frac{\beta}{2}\right)\|x^t - x\|^2\right] + \frac{G^2}{2}\left(\sum_{t=1}^T \mathbb{E}\left[\frac{1}{\eta^{t+1}}\right]\right) \tag{12}$$

Further, we have

$$\sum_{t=1}^T \frac{1}{\eta^{t+1}} \leq \sqrt{1 + \max\{M^1, \ldots, M^T\}}\left(\sum_{t=1}^T \frac{1}{t+1}\right) \leq \sqrt{1 + \max\{M^1, \ldots, M^T\}} \log(T+1) \tag{13}$$

Plugging Eq. (13) into Eq. (12) yields that

$$\sum_{t=1}^T \mathbb{E}[f_t(x^t) - f_t(x)] \leq \frac{D^2}{2} + \sum_{t=1}^T \mathbb{E}\left[\left(\frac{1}{2\sqrt{1 + \max\{M^1, \ldots, M^t\}}} - \frac{\beta}{2}\right)\|x^t - x\|^2\right]$$
$$+ \frac{G^2 \log(T+1)}{2} \mathbb{E}\left[\sqrt{1 + \max\{M^1, \ldots, M^T\}}\right]$$

Since $\mathcal{X}$ is convex and bounded with a diameter $D > 0$, we have

$$\sum_{t=1}^{T}\left(\frac{1}{2\sqrt{1+\max\{M^1,\ldots,M^t\}}}-\frac{\beta}{2}\right)\|x^t-x\|^2 \leq \frac{D^2}{2}\left(\sum_{t=1}^{T}\max\left\{0,\frac{1}{\sqrt{1+\max\{M^1,\ldots,M^t\}}}-\beta\right\}\right)$$

This implies that

$$\mathbb{E}[\mathrm{Regret}(T)] \leq \frac{D^2}{2}+\frac{D^2}{2}\underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\max\left\{0,\frac{1}{\sqrt{1+\max\{M^1,\ldots,M^t\}}}-\beta\right\}\right]}_{\mathbf{I}}$$
$$+\frac{G^2\log(T+1)}{2}\underbrace{\mathbb{E}\left[\sqrt{1+\max\{M^1,\ldots,M^T\}}\right]}_{\mathbf{II}}$$

It remains to bound the terms $\mathbf{I}$ and $\mathbf{II}$ using Proposition 3 (cf. Section D). Indeed, we have

$$\mathbf{I} = \sum_{t=1}^{T}\mathbb{E}\left[\max\left\{0,\frac{1}{\sqrt{1+\max\{M^1,\ldots,M^t\}}}-\beta\right\}\right] \leq \sum_{t=1}^{T}\mathbb{P}\left(\frac{1}{\sqrt{1+\max\{M^1,\ldots,M^t\}}}-\beta \geq 0\right)$$
$$= \sum_{t=1}^{T}\mathbb{P}\left(\sqrt{1+\max\{M^1,\ldots,M^t\}}\leq\frac{1}{\beta}\right) \leq \sum_{t=1}^{T}\mathbb{P}\left(\max\{M^1,\ldots,M^t\}\leq\frac{1}{\beta^2}\right)$$

Since $M^1,\ldots,M^t$ are i.i.d. geometric random variables with $p_0 = \frac{1}{\log(T+10)}$, Proposition 3 implies that

$$\sum_{t=1}^{T}\mathbb{P}\left(\max\{M^1,\ldots,M^t\}\leq\frac{1}{\beta^2}\right) \leq e^{\frac{p_0}{\beta^2}} = e^{\frac{1}{\beta^2\log(T+10)}}$$

Putting these pieces together yields that $\mathbf{I} \leq e^{\frac{1}{\beta^2\log(T+10)}}$.

By using Jensen's inequality and the concavity of $g(x) = \sqrt{x}$, we have

$$\mathbf{II} \leq \sqrt{1+\mathbb{E}\left[\max\{M^1,\ldots,M^T\}\right]}$$

Using Proposition 3 and $p_0 = \frac{1}{\log(T+10)}$, we have

$$\mathbb{E}\left[\max\{M^1,\ldots,M^T\}\right] \leq \frac{1+\log(T)}{p_0} = \log(T+10)+\log(T)\log(T+10)$$

Putting these pieces together yields that $\mathbf{II} \leq \sqrt{1+\log(T+10)+\log(T)\log(T+10)}$. Therefore, we conclude that

$$\mathbb{E}[\mathrm{Regret}(T)] \leq \frac{D^2}{2}(1+e^{\frac{1}{\beta^2\log(T+10)}})+\frac{G^2\log(T+1)}{2}\sqrt{1+\log(T+10)+\log(T)\log(T+10)}$$
$$= \frac{D^2}{2}(1+e^{\frac{1}{\beta^2\log(T+10)}})+\frac{G^2\log(T+1)}{2}\sqrt{1+\log(T+10)+\log(T)\log(T+10)}$$

This completes the proof.

C.2 Proof of Theorem 3

To prove Theorem 3, we provide another descent inequality for the iterates generated by Algorithm 2.

Lemma 3. *Suppose that a continuous game $\mathcal{G}$ is β-strongly monotone and let $G, D > 0$ be problem parameters satisfying $\mathbb{E}[\xi^t \mid x^t] = v(x^t)$ and $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq G^2$ for all $t \geq 1$, and let $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. Letting the iterates $\{x^t\}_{t\geq 1}$ be generated by Algorithm 2, we have*

$$\sum_{i=1}^{N} \eta_i^T \mathbb{E}\left[\|x_i^T - x_i^\star\|^2 \mid \{\eta_i^t\}_{1\leq i\leq N, 1\leq t\leq T}\right] \leq \sum_{i=1}^{N} \eta_i^1 \|x_i^1 - x_i^\star\|^2$$

$$+ D^2 \left(\sum_{t=1}^{T-1}\left(\max\left\{0, \max_{1\leq i\leq N}\{\eta_i^{t+1} - \eta_i^t\} - 2\beta\right\}\right)\right) + G^2 \left(\sum_{t=1}^{T-1}\left(\max_{1\leq i\leq N}\left\{\frac{1}{\eta_i^{t+1}}\right\}\right)\right)$$

where $x^\star \in \mathcal{X}$ is the unique Nash equilibrium.

We defer the proof of this lemma to Section C.6.3.

Proof of Theorem 3. Since $D > 0$ satisfies that $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$ and $\eta_i^{t+1} = \frac{t+1}{\sqrt{1+\max\{M_i^1,\ldots,M_i^t\}}}$ in Algorithm 2, we have

$$\sum_{i=1}^{N} \eta_i^1 \|x_i^1 - x_i^\star\|^2 \leq D^2, \qquad \eta_i^{t+1} - \eta_i^t \leq \frac{1}{\sqrt{1 + \max\{M_i^1,\ldots,M_i^t\}}}$$

By Lemma 3, we have

$$\sum_{i=1}^{N} \eta_i^T \mathbb{E}\left[\|x_i^T - x_i^\star\|^2 \mid \{\eta_i^t\}_{1\leq i\leq N, 1\leq t\leq T}\right] \leq D^2 \tag{14}$$

$$+ D^2 \left(\sum_{t=1}^{T-1}\left(\max\left\{0, \max_{1\leq i\leq N}\left\{\frac{1}{\sqrt{1 + \max\{M_i^1,\ldots,M_i^t\}}}\right\} - 2\beta\right\}\right)\right) + G^2 \left(\sum_{t=1}^{T-1}\left(\max_{1\leq i\leq N}\left\{\frac{1}{\eta_i^{t+1}}\right\}\right)\right)$$

Further, we have

$$\sum_{t=1}^{T-1}\left(\max_{1\leq i\leq N}\left\{\frac{1}{\eta_i^{t+1}}\right\}\right) \leq \sqrt{1 + \max_{1\leq i\leq N, 1\leq t\leq T}\{M_i^t\}}\left(\sum_{t=1}^{T-1}\frac{1}{t+1}\right) \leq \log(T+1)\sqrt{1 + \max_{1\leq i\leq N, 1\leq t\leq T}\{M_i^t\}} \tag{15}$$

Plugging Eq. (15) into Eq. (14) yields that

$$\sum_{i=1}^{N} \eta_i^T \mathbb{E}\left[\|x_i^T - x_i^\star\|^2 \mid \{\eta_i^t\}_{1\leq i\leq N, 1\leq t\leq T}\right]$$

$$\leq D^2 + D^2 \left(\sum_{t=1}^{T-1}\left(\max\left\{0, \max_{1\leq i\leq N}\left\{\frac{1}{\sqrt{1 + \max\{M_i^1,\ldots,M_i^t\}}}\right\} - 2\beta\right\}\right)\right)$$

$$+ G^2 \log(T+1)\sqrt{1 + \max_{1\leq i\leq N, 1\leq t\leq T}\{M_i^t\}}$$

35

By the definition of η_i^T, we have

$$\eta_i^T \geq \frac{T}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^T\}}} \geq \frac{T}{\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}}$$

which implies that

$$\sum_{i=1}^{N} \eta_i^T \mathbb{E}\left[\|x_i^T - x_i^\star\|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}\right] \geq \frac{T}{\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}} \cdot \mathbb{E}\left[\|x^T - x^\star\|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}\right]$$

Putting these pieces together yields that

$$\left(\frac{T}{\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}}\right) \mathbb{E}\left[\|x^T - x^\star\|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}\right]$$

$$\leq D^2 + D^2 \left(\sum_{t=1}^{T-1}\left(\max\left\{0, \max_{1 \leq i \leq N}\left\{\frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}\right\} - 2\beta\right\}\right)\right)$$

$$+ G^2 \log(T+1)\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}$$

Rearranging and taking the expectation of both sides, we have

$$T \cdot \mathbb{E}[\|x^T - x^\star\|^2] \leq D^2 \underbrace{\mathbb{E}\left[\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}\right]}_{\mathbf{I}} + G^2 \log(T+1) \underbrace{\mathbb{E}\left[1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}\right]}_{\mathbf{II}}$$

$$+ D^2 \underbrace{\mathbb{E}\left[\sum_{t=1}^{T-1} \max\left\{0, \max_{1 \leq i \leq N}\left\{\frac{\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}\right\} - 2\beta\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}\right\}\right]}_{\mathbf{III}} \quad (16)$$

By using the previous argument with Proposition 3 (cf. Section D) and $p = \frac{1}{\log(T+10)}$, we have

$$\mathbf{I} \leq \sqrt{1 + \frac{1 + \log(NT)}{p}} = \sqrt{1 + \log(T+10) + \log(NT)\log(T+10)} \quad (17)$$

and

$$\mathbf{II} \leq 1 + \frac{1 + \log(NT)}{p} = 1 + \log(T+10) + \log(NT)\log(T+10) \quad (18)$$

It remains to bound the term $\mathbf{III}$ using Proposition 3 and $p = \frac{1}{\log(T+10)}$. Indeed, we have

$$\mathbf{III} = \mathbb{E}\left[\sum_{t=1}^{T-1} \max\left\{0, \max_{1 \leq i \leq N}\left\{\frac{\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}\right\} - 2\beta\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T-1} \max_{1 \leq i \leq N}\left\{\frac{\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}\right\} \cdot \mathbb{I}\left(\max_{1 \leq i \leq N}\left\{\frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}\right\} - 2\beta \geq 0\right)\right]$$

Defining $i^t = \text{argmax}_{1 \leq i \leq N} \left\{ \frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}} \right\}$ as a random variable and recalling that $\{\max\{M_i^1, \ldots, M_i^t\}\}_{1 \leq i \leq N}$ are independent and identically distributed, we have that $i^t \in \{1, \ldots, N\}$ is uniformly distributed. This implies that

$$\mathbf{III} \leq \frac{1}{N} \sum_{t=1}^{T-1} \sum_{j=1}^{N} \mathbb{E} \left[\max_{1 \leq i \leq N} \left\{ \frac{\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}} \right\} \cdot \mathbb{I} \left(\max_{1 \leq i \leq N} \left\{ \frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}} \right\} - 2\beta \geq 0 \right) \mid i^t = j \right]$$

$$= \frac{1}{N} \sum_{t=1}^{T-1} \sum_{j=1}^{N} \mathbb{E} \left[\frac{\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}}{\sqrt{1 + \max\{M_j^1, \ldots, M_j^t\}}} \cdot \mathbb{I} \left(\frac{1}{\sqrt{1 + \max\{M_j^1, \ldots, M_j^t\}}} - 2\beta \geq 0 \right) \right]$$

$$\leq \frac{1}{N} \sum_{t=1}^{T-1} \sum_{j=1}^{N} \mathbb{E} \left[\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T, i \neq j}\{M_i^t\}} \cdot \mathbb{I} \left(\frac{1}{\sqrt{1 + \max\{M_j^1, \ldots, M_j^t\}}} - 2\beta \geq 0 \right) \right]$$

Since $\max_{1 \leq i \leq N, 1 \leq t \leq T, i \neq j}\{M_i^t\}$ is independent of $\max\{M_j^1, \ldots, M_j^t\}$, we have

$$\mathbf{III} \leq \frac{1}{N} \sum_{t=1}^{T-1} \sum_{j=1}^{N} \mathbb{E} \left[\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T, i \neq j}\{M_i^t\}} \right] \cdot \mathbb{P} \left(\frac{1}{\sqrt{1 + \max\{M_j^1, \ldots, M_j^t\}}} - 2\beta \geq 0 \right)$$

Since $\{M_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}$ is a sequence of independent and identically distributed geometric random variables with $p = \frac{1}{\log(T+10)}$, Proposition 3 (cf. Section D) implies that

$$\mathbb{E} \left[\max_{1 \leq i \leq N, 1 \leq t \leq T, i \neq j}\{M_i^t\} \right] \leq \frac{1 + \log(NT)}{p} = \log(T + 10) + \log(NT) \log(T + 10)$$

and

$$\sum_{t=1}^{T-1} \mathbb{P} \left(\frac{1}{\sqrt{1 + \max\{M_j^1, \ldots, M_j^t\}}} - 2\beta \geq 0 \right) \leq e^{\frac{p}{4\beta^2}} = e^{\frac{1}{4\beta^2 \log(T+10)}}$$

Putting these pieces together with Jensen's inequality yields that

$$\mathbf{III} \leq e^{\frac{1}{4\beta^2 \log(T+10)}} \sqrt{1 + \log(T + 10) + \log(NT) \log(T + 10)} \tag{19}$$

Plugging Eq. (17), Eq. (18) and Eq. (19) into Eq. (16) yields that

$$\begin{aligned}
\mathbb{E}[\|x^T - x^\star\|^2] &\leq \frac{D^2}{T} (1 + e^{\frac{1}{4\beta^2 \log(T+10)}}) \sqrt{1 + \log(T + 10) + \log(NT) \log(T + 10)} \\
&\quad + \frac{G^2}{T} \log(T + 1)(1 + \log(T + 10) + \log(NT) \log(T + 10))
\end{aligned}$$

This completes the proof.

It is worth remarking that our proof does not use the structure of $v(\cdot)$ (i.e., $v_i(x) = \nabla_{x_i} u_i(x_i; x_{-i})$). Thus, Theorem 3, when extended to the VI setting, yields a decentralized, feasible optimization algorithm for finding a solution of a strongly monotone VI, contributing to that line of literature.

Corollary 3. *Suppose that the variational inequality defined by $v(\cdot)$ (without the structure that $v_i(x) = \nabla_{x_i} u_i(x_i; x_{-i})$) and $\mathcal{X}$ (any convex set) is β-strongly monotone and let $G, D > 0$ be problem*

parameters satisfying $\mathbb{E}[\xi^t \mid x^t] = v(x^t)$ and $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq G^2$ for all $t \geq 1$, and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. Then, by employing Algorithm 2, we have

$$\mathbb{E}[\|x^T - x^\star\|^2] \leq \frac{D^2}{T}(1 + e^{\frac{1}{4\beta^2 \log(T+10)}})\sqrt{1 + \log(T+10) + \log(NT)\log(T+10)}$$

$$+ \frac{G^2}{T}\log(T+1)(1 + \log(T+10) + \log(NT)\log(T+10))$$

where $x^\star \in \mathcal{X}$ is the unique solution of he VI. As a consequence, we have $\mathbb{E}[\|x^T - x^\star\|^2] = O(\frac{\log^3(T)}{T})$.

C.3   Proof of Theorem 4

By the update formula of η^{t+1} in Algorithm 3, we have $\eta^{t+1} = \frac{1}{\sqrt{1+\max\{M^1,\ldots,M^t\}}}$. By the update formula of A^{t+1} in Algorithm 3, we have $A^1 = I_d$ where $I_d \in \mathbb{R}^{d \times d}$ is an identity matrix and $A^{t+1} = A^t + \nabla f_t(x^t)\nabla f_t(x^t)^\top$. Since $\mathcal{X}$ is convex and bounded with a diameter $D > 0$, we have

$$\frac{\eta^1}{2}(x^1 - x)^\top A^1(x^1 - x) \leq \frac{D^2}{2}, \qquad \eta^{t+1} A^{t+1} - \eta^t A^t \preceq \frac{1}{\sqrt{1 + \max\{M^1,\ldots,M^t\}}}\nabla f_t(x^t)\nabla f_t(x^t)^\top$$

By Lemma 1, we have

$$\sum_{t=1}^{T} f_t(x^t) - \sum_{t=1}^{T} f_t(x) \leq \frac{D^2}{2} + \frac{1}{2}\left(\sum_{t=1}^{T} \frac{1}{\eta^{t+1}}\nabla f_t(x^t)^\top (A^{t+1})^{-1}\nabla f_t(x^t)\right) \tag{20}$$

$$+ \sum_{t=1}^{T}\left(\frac{1}{2\sqrt{1+\max\{M^1,\ldots,M^t\}}} - \frac{1}{4}\min\left\{\frac{1}{4GD},\alpha\right\}\right)(x^t - x)^\top \nabla f_t(x^t)\nabla f_t(x^t)^\top (x^t - x)$$

Since $\|\nabla f_t(x)\| \leq G$, $A^1 = I_d$ where $I_d \in \mathbb{R}^{d \times d}$ is an identity matrix and $A^{t+1} = A^t + \nabla f_t(x^t)\nabla f_t(x^t)^\top$, we derive from [Hazan et al.(2007), Lemma 11] that

$$\sum_{t=1}^{T} \frac{1}{\eta^{t+1}}\nabla f_t(x^t)^\top (A^{t+1})^{-1}\nabla f_t(x^t) \leq \sqrt{1 + \max\{M^1,\ldots,M^T\}}\left(\sum_{t=1}^{T}\nabla f_t(x^t)^\top (A^{t+1})^{-1}\nabla f_t(x^t)\right)$$

$$\leq \sqrt{1 + \max\{M^1,\ldots,M^T\}}(d\log(TG^2 + 1)) \tag{21}$$

Plugging Eq. (21) into Eq. (20) yields that

$$\sum_{t=1}^{T} f_t(x^t) - \sum_{t=1}^{T} f_t(x) \leq \frac{D^2}{2} + \frac{d\sqrt{1+\max\{M^1,\ldots,M^T\}}}{2}\log(TG^2 + 1)$$

$$+ \sum_{t=1}^{T}\left(\frac{1}{2\sqrt{1+\max\{M^1,\ldots,M^t\}}} - \frac{1}{4}\min\left\{\frac{1}{4GD},\alpha\right\}\right)(x^t - x)^\top \nabla f_t(x^t)\nabla f_t(x^t)^\top (x^t - x)$$

Since $\|\nabla f_t(x)\| \leq G$ and $\mathcal{X}$ is convex and bounded with a diameter $D > 0$, we have

$$\sum_{t=1}^{T}\left(\frac{1}{2\sqrt{1+\max\{M^1,\ldots,M^t\}}} - \frac{1}{4}\min\left\{\frac{1}{4GD},\alpha\right\}\right)(x^t - x)^\top \nabla f_t(x^t)\nabla f_t(x^t)^\top (x^t - x)$$

$$\leq \frac{G^2 D^2}{2}\left(\sum_{t=1}^{T}\max\left\{0, \frac{1}{\sqrt{1+\max\{M^1,\ldots,M^t\}}} - \frac{1}{2}\min\left\{\frac{1}{4GD},\alpha\right\}\right\}\right)$$

Putting these pieces together yields that

$$\text{Regret}(T) \leq \frac{D^2}{2} + \frac{G^2 D^2}{2} \left(\sum_{t=1}^{T} \max \left\{ 0, \frac{1}{\sqrt{1 + \max\{M^1, \ldots, M^t\}}} - \frac{1}{2} \min \left\{ \frac{1}{4GD}, \alpha \right\} \right\} \right)$$
$$+ \frac{d\sqrt{1 + \max\{M^1, \ldots, M^T\}}}{2} \log(TG^2 + 1)$$

For simplicity, we let $\gamma = \frac{1}{2} \min \left\{ \frac{1}{4GD}, \alpha \right\}$. Taking the expectation of both sides, we have

$$\mathbb{E}[\text{Regret}(T)] \leq \frac{D^2}{2} + \frac{G^2 D^2}{2} \underbrace{\mathbb{E} \left[\sum_{t=1}^{T} \max \left\{ 0, \frac{1}{\sqrt{1 + \max\{M^1, \ldots, M^t\}}} - \gamma \right\} \right]}_{\mathbf{I}}$$
$$+ \frac{d \log(TG^2 + 1)}{2} \underbrace{\mathbb{E} \left[\sqrt{1 + \max\{M^1, \ldots, M^T\}} \right]}_{\mathbf{II}}$$

It remains to bound the terms **I** and **II** using Proposition 3. By using the same argument as applied in the proof of Theorem 1, we have

$$\mathbf{I} \leq e^{\frac{1}{\gamma^2 \log(T+10)}}$$

and

$$\mathbf{II} \leq \sqrt{1 + \log(T + 10) + \log(T) \log(T + 10)}$$

Therefore, we conclude that

$$\mathbb{E}[\text{Regret}(T)] \leq \frac{D^2}{2} (1 + e^{\frac{1}{\gamma^2 \log(T+10)}}) + \frac{d \log(TG^2 + 1)}{2} \sqrt{1 + \log(T + 10) + \log(T) \log(T + 10)}$$

This completes the proof.

C.4 Proof of Theorem 5

The proof of Theorem 5 is again based on a descent inequality.

Lemma 4. *Suppose that a continuous game $\mathcal{G}$ is α-exp-concave and let $G, D > 0$ be problem parameters satisfying $\|v(x)\| \leq G$ and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$. Letting the iterates $\{x^t\}_{t \geq 1}$ be generated by the multi-agent ONS in Eq. (7), we have*

$$\sum_{t=1}^{T} (x^t - x)^\top v(x) \leq \sum_{i=1}^{N} \frac{\eta_i}{2} (x_i^1 - x_i)^\top A_i^1 (x_i^1 - x_i) + \frac{1}{2} \left(\sum_{t=1}^{T} \sum_{i=1}^{N} \frac{1}{\eta_i} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t) \right)$$
$$+ \sum_{t=1}^{T} \sum_{i=1}^{N} (x_i^t - x_i)^\top \left(\frac{\eta_i}{2} A_i^{t+1} - \frac{\eta_i}{2} A_i^t - \frac{1}{4} \min \left\{ \frac{1}{4GD}, \alpha \right\} v_i(x^t) v_i(x^t)^\top \right) (x_i^t - x_i)$$

See Appendices C.6.4 for the proof of Lemma 4.

Proof of Theorem 5.   We can see from the multi-agent ONS algorithm that $\eta_i = \frac{1}{2}\min\left\{\frac{1}{4GD}, \alpha\right\}$, $A_i^1 = I_{d_i}$ where $I_{d_i} \in \mathbb{R}^{d_i \times d_i}$ is an identity matrix, and $A_i^{t+1} = A_i^t + v_i(x^t)v_i(x^t)^\top$. Since $\mathcal{X}$ is convex and bounded with a diameter $D > 0$, we have

$$\sum_{i=1}^N \frac{\eta_i}{2}(x_i^1 - x_i)^\top A_i^1 (x_i^1 - x_i) \le \frac{\alpha D^2}{4}, \qquad \eta_i A_i^{t+1} - \eta_i A_i = \frac{1}{2}\min\left\{\frac{1}{4GD}, \alpha\right\}v_i(x^t)v_i(x^t)^\top$$

By Lemma 4, we have

$$\sum_{t=1}^T (x^t - x)^\top v(x) \le \frac{\alpha D^2}{4} + \frac{1}{2}\left(\sum_{t=1}^T \sum_{i=1}^N \frac{1}{\eta_i} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t)\right) \tag{22}$$

Since $\|v(x)\| \le G$, $A_i^1 = I_{d_i}$ and $A_i^{t+1} = A_i^t + v_i(x^t)v_i(x^t)^\top$, [Hazan et al.(2007), Lemma 11] guarantees that

$$\sum_{t=1}^T v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t) \le d_i \log(TG^2 + 1)$$

which implies that

$$\sum_{t=1}^T \sum_{i=1}^N \frac{1}{\eta_i} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t) \le \max\left\{8GD, \frac{2}{\alpha}\right\}(d\log(TG^2 + 1)) \tag{23}$$

Plugging Eq. (23) into Eq. (22) yields that

$$\sum_{t=1}^T (x^t - x)^\top v(x) \le \frac{\alpha D^2}{4} + \max\left\{4GD, \frac{1}{\alpha}\right\}(d\log(TG^2 + 1))$$

By the definition of $\text{GAP}(\cdot)$ and $\bar{x}^T$ (i.e., $\bar{x}^T = \frac{1}{T}\sum_{t=1}^T x^t$), we have

$$\text{GAP}(\bar{x}^T) \le \frac{\alpha D^2}{4T} + \frac{d\log(TG^2 + 1)}{T}\max\left\{4GD, \frac{1}{\alpha}\right\}$$

This completes the proof.

C.5   Proof of Theorem 6

We can see from Algorithm 4 that $\eta_i^{t+1} = \frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}$, $A_i^1 = I_{d_i}$ where $I_{d_i} \in \mathbb{R}^{d_i \times d_i}$ is an identity matrix, and $A_i^{t+1} = A_i^t + v_i(x^t)v_i(x^t)^\top$. Since $\mathcal{X}$ is convex and bounded with a diameter $D > 0$, we have

$$\sum_{i=1}^N \frac{\eta_i^1}{2}(x_i^1 - x_i)^\top A_i^1 (x_i^1 - x_i) \le \frac{D^2}{2}, \qquad \eta_i^{t+1} A_i^{t+1} - \eta_i^t A_i \preceq \frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}v_i(x^t)v_i(x^t)^\top$$

We can see from Eq. (8) that

$$\sum_{t=1}^T (x^t - x)^\top v(x) \le \frac{D^2}{2} + \frac{1}{2}\left(\sum_{t=1}^T \sum_{i=1}^N \frac{1}{\eta_i^{t+1}} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t)\right) \tag{24}$$

$$+ \sum_{t=1}^T \sum_{i=1}^N \left(\frac{1}{2\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}} - \frac{1}{4}\min\left\{\frac{1}{4GD}, \alpha\right\}\right)(x_i^t - x_i)^\top v_i(x^t)v_i(x^t)^\top (x_i^t - x_i)$$

Since $\|v(x)\| \leq G$, $A_i^1 = I_{d_i}$ and $A_i^{t+1} = A_i^t + v_i(x^t)v_i(x^t)^\top$, [Hazan et al.(2007), Lemma 11] guarantees that

$$
\sum_{t=1}^{T} \frac{1}{\eta_i^{t+1}} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t) \;\leq\; \sqrt{1 + \max\{M_i^1, \ldots, M_i^T\}} \left(\sum_{t=1}^{T} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t) \right)
$$

$$
\leq \;\; \sqrt{1 + \max\{M_i^1, \ldots, M_i^T\}}(d_i \log(TG^2 + 1))
$$

which implies that

$$
\sum_{t=1}^{T} \sum_{i=1}^{N} \frac{1}{\eta_i^{t+1}} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t) \leq \sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}(d \log(TG^2 + 1)) \tag{25}
$$

Plugging Eq. (25) into Eq. (24) yields that

$$
\sum_{t=1}^{T}(x^t - x)^\top v(x) \leq \frac{D^2}{2} + \frac{d \log(TG^2 + 1)}{2} \sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}
$$

$$
+ \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\frac{1}{2\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}} - \frac{1}{4} \min\left\{ \frac{1}{4GD}, \alpha \right\} \right) (x_i^t - x_i)^\top v_i(x^t)v_i(x^t)^\top (x_i^t - x_i)
$$

Since $\|v(x)\| \leq G$ and $\mathcal{X}$ is convex and bounded with a diameter $D > 0$, we have

$$
\sum_{t=1}^{T} \sum_{i=1}^{N} \left(\frac{1}{2\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}} - \frac{1}{4} \min\left\{ \frac{1}{4GD}, \alpha \right\} \right) (x_i^t - x_i)^\top v_i(x^t)v_i(x^t)^\top (x_i^t - x_i)
$$

$$
\leq \;\; \frac{G^2 D^2}{2} \left(\sum_{t=1}^{T} \max\left\{ 0, \max_{1 \leq i \leq N} \left\{ \frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}} \right\} - \frac{1}{2} \min\left\{ \frac{1}{4GD}, \alpha \right\} \right\} \right)
$$

Putting these pieces together yields that

$$
\sum_{t=1}^{T}(x^t - x)^\top v(x) \leq \frac{D^2}{2} + \frac{d \log(TG^2 + 1)}{2} \sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}
$$

$$
+ \frac{G^2 D^2}{2} \left(\sum_{t=1}^{T} \max\left\{ 0, \max_{1 \leq i \leq N} \left\{ \frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}} \right\} - \frac{1}{2} \min\left\{ \frac{1}{4GD}, \alpha \right\} \right\} \right)
$$

For simplicity, we let $\gamma = \frac{1}{2} \min\left\{ \frac{1}{4GD}, \alpha \right\}$. By the definition of $\bar{x}^T$ (i.e., $\bar{x}^T = \frac{1}{T} \sum_{t=1}^{T} x^t$), we have

$$
\mathrm{GAP}(\bar{x}^T) \leq \frac{D^2}{2T} + \frac{G^2 D^2}{2T} \left(\sum_{t=1}^{T} \max\left\{ 0, \max_{1 \leq i \leq N} \left\{ \frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}} \right\} - \gamma \right\} \right)
$$

$$
+ \frac{d \log(TG^2 + 1)}{2T} \sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}
$$

Taking the expectation of both sides, we have

$$\mathbb{E}[\text{GAP}(\bar{x}^T)] \leq \frac{D^2}{2T} + \frac{G^2 D^2}{2T} \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \max\left\{0, \max_{1 \leq i \leq N}\left\{\frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}\right\} - \gamma\right\}\right]}_{\mathbf{I}} \quad (26)$$

$$+ \frac{d \log(TG^2 + 1)}{2T} \underbrace{\mathbb{E}\left[\sqrt{1 + \max_{1 \leq i \leq N, 1 \leq t \leq T}\{M_i^t\}}\right]}_{\mathbf{II}}$$

It remains to bound the terms **I** and **II** using Proposition 3. Indeed, we have

$$\mathbf{I} = \sum_{t=1}^{T} \mathbb{E}\left[\max\left\{0, \max_{1 \leq i \leq N}\left\{\frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}\right\} - \gamma\right\}\right]$$

$$\leq \sum_{t=1}^{T} \mathbb{P}\left(\max_{1 \leq i \leq N}\left\{\frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}\right\} - \gamma \geq 0\right)$$

Considering $i^t = \text{argmax}_{1 \leq i \leq N}\left\{\frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}\right\}$ that is a random variable and then recalling that $\{\max\{M_i^1, \ldots, M_i^t\}\}_{1 \leq i \leq N}$ are i.i.d., we have $i^t \in \{1, \ldots, N\}$ is uniformly distributed. This implies that

$$\mathbf{I} \leq \frac{1}{N} \sum_{t=1}^{T} \sum_{j=1}^{N} \mathbb{P}\left(\max_{1 \leq i \leq N}\left\{\frac{1}{\sqrt{1 + \max\{M_i^1, \ldots, M_i^t\}}}\right\} - \gamma \geq 0 \mid i^t = j\right)$$

$$= \frac{1}{N} \sum_{t=1}^{T} \sum_{j=1}^{N} \mathbb{P}\left(\frac{1}{\sqrt{1 + \max\{M_j^1, \ldots, M_j^t\}}} - \gamma \geq 0\right)$$

Since $\{M_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}$ are i.i.d. geometric random variables with $p_0 = \frac{1}{\log(T+10)}$, Proposition 3 implies that

$$\sum_{t=1}^{T} \mathbb{P}\left(\frac{1}{\sqrt{1 + \max\{M_j^1, \ldots, M_j^t\}}} - \gamma \geq 0\right) \leq e^{\frac{p}{\gamma^2}} = e^{\frac{1}{\gamma^2 \log(T+10)}}$$

Putting these pieces together yields that

$$\mathbf{I} \leq e^{\frac{1}{\gamma^2 \log(T+10)}} = e^{\frac{(\max\{8GD, 2\alpha^{-1}\})^2}{\log(T+10)}} \quad (27)$$

By using the similar argument with Proposition 3 and $p_0 = \frac{1}{\log(T+10)}$, we have

$$\mathbf{II} \leq \sqrt{1 + \frac{1 + \log(NT)}{p_0}} = \sqrt{1 + \log(T + 10) + \log(NT)\log(T + 10)} \quad (28)$$

Plugging Eq. (27) and Eq. (28) into Eq. (26) yields that

$$\mathbb{E}[\text{GAP}(\bar{x}^T)] \leq \frac{D^2}{2T}(1 + G^2 e^{\frac{(\max\{8GD, 2\alpha^{-1}\})^2}{\log(T+10)}}) + \frac{d\log(TG^2 + 1)}{2T}\sqrt{1 + \log(T + 10) + \log(T)\log(T + 10)}$$

This completes the proof.

C.6 Proofs of Auxiliary Lemmas

C.6.1 Proof of Lemma 1

Recall that the update formula of x^{t+1} in Algorithm 3 is

$$x^{t+1} \leftarrow \underset{x \in \mathcal{X}}{\operatorname{argmin}}\{(x - x^t)^\top \nabla f_t(x^t) + \frac{\eta^{t+1}}{2}(x - x^t)^\top A^{t+1}(x - x^t)\}$$

The first-order optimality condition implies that

$$(x - x^{t+1})^\top \nabla f_t(x^t) + \eta^{t+1}(x - x^{t+1})^\top A^{t+1}(x^{t+1} - x^t) \geq 0, \quad \text{for all } x \in \mathcal{X}$$

Equivalently, we have

$$\frac{\eta^{t+1}}{2}((x^{t+1} - x)^\top A^{t+1}(x^{t+1} - x) - (x^t - x)^\top A^{t+1}(x^t - x))$$

$$\leq \ (x - x^{t+1})^\top \nabla f_t(x^t) - \frac{\eta^{t+1}}{2}(x^{t+1} - x^t)^\top A^{t+1}(x^{t+1} - x^t)$$

$$= \ (x - x^t)^\top \nabla f_t(x^t) + (x^t - x^{t+1})^\top \nabla f_t(x^t) - \frac{\eta^{t+1}}{2}(x^{t+1} - x^t)^\top A^{t+1}(x^{t+1} - x^t)$$

$$\leq \ (x - x^t)^\top \nabla f_t(x^t) + \frac{1}{2\eta^{t+1}}\nabla f_t(x^t)^\top (A^{t+1})^{-1}\nabla f_t(x^t)$$

Since f_t is α-exp-concave and satisfies that $\|\nabla f_t(x)\| \leq G$ and $\|x - x'\| \leq D$ for all $x, x' \in \mathcal{X}$, we derive from [Hazan et al.(2007), Lemma 3] that

$$f_t(x) \geq f_t(x^t) + (x - x^t)^\top \nabla f_t(x^t) + \frac{1}{4}\min\left\{\frac{1}{4GD}, \alpha\right\}(x - x^t)^\top(\nabla f_t(x^t)\nabla f_t(x^t)^\top)(x - x^t)$$

For simplicity, we let $\gamma = \frac{1}{4}\min\left\{\frac{1}{4GD}, \alpha\right\}$. Putting these pieces together yields that

$$\frac{\eta^{t+1}}{2}((x^{t+1} - x)^\top A^{t+1}(x^{t+1} - x) - (x^t - x)^\top A^{t+1}(x^t - x)) \leq f_t(x) - f_t(x^t)$$

$$-\gamma(x - x^t)^\top(\nabla f_t(x^t)\nabla f_t(x^t)^\top)(x - x^t) + \frac{1}{2\eta^{t+1}}\nabla f_t(x^t)^\top(A^{t+1})^{-1}\nabla f_t(x^t)$$

Rearranging the above inequality yields that

$$f_t(x^t) - f_t(x) + \frac{\eta^{t+1}}{2}(x^{t+1} - x)^\top A^{t+1}(x^{t+1} - x) - \frac{\eta^t}{2}(x^t - x)^\top A^t(x^t - x)$$

$$\leq \ (x - x^t)^\top\left(\frac{\eta^{t+1}}{2}A^{t+1} - \frac{\eta^t}{2}A^t - \gamma\nabla f_t(x^t)\nabla f_t(x^t)^\top\right)(x - x^t) + \frac{1}{2\eta^{t+1}}\nabla f_t(x^t)^\top(A^{t+1})^{-1}\nabla f_t(x^t)$$

Summing over $t = 1, 2, \ldots, T$ yields the desired inequality.

C.6.2 Proof of Lemma 2

Recall that the update formula of x^{t+1} in Algorithm 1 is

$$x^{t+1} \leftarrow \underset{x \in \mathcal{X}}{\operatorname{argmin}}\{(x - x^t)^\top \xi^t + \frac{\eta^{t+1}}{2}\|x - x^t\|^2\}$$

The first-order optimality condition implies that

$$(x - x^{t+1})^\top \xi^t + \eta^{t+1}(x - x^{t+1})^\top (x^{t+1} - x^t) \geq 0, \quad \text{for all } x \in \mathcal{X}$$

Equivalently, we have

$$\frac{\eta^{t+1}}{2}(\|x^{t+1} - x\|^2 - \|x^t - x\|^2) \leq (x - x^{t+1})^\top \xi^t - \frac{\eta^{t+1}}{2}\|x^{t+1} - x^t\|^2$$

$$= (x - x^t)^\top \xi^t + (x^t - x^{t+1})^\top \xi^t - \frac{\eta^{t+1}}{2}\|x^{t+1} - x^t\|^2 \leq (x - x^t)^\top \xi^t + \frac{1}{2\eta^{t+1}}\|\xi^t\|^2$$

Since $\mathbb{E}[\xi^t \mid x^t] = \nabla f_t(x^t)$ and $\mathbb{E}[\|\xi^t\|^2 \mid x^t] \leq G^2$ for all $t \geq 1$, we have

$$\mathbb{E}\left[\frac{\eta^{t+1}}{2}\left(\|x^{t+1} - x\|^2 - \|x^t - x\|^2\right) \mid x^t\right] \leq (x - x^t)^\top \nabla f_t(x^t) + \mathbb{E}\left[\frac{G^2}{2\eta^{t+1}}\right]$$

Since f_t is β-strongly convex, we have

$$\mathbb{E}\left[\frac{\eta^{t+1}}{2}\left(\|x^{t+1} - x\|^2 - \|x^t - x\|^2\right) \mid x^t\right] \leq f_t(x) - f_t(x^t) - \frac{\beta}{2}\|x^t - x\|^2 + \mathbb{E}\left[\frac{G^2}{2\eta^{t+1}}\right]$$

Taking the expectation of both sides and rearranging the resulting inequality yields

$$\mathbb{E}\left[f_t(x^t) + \frac{\eta^{t+1}}{2}\|x^{t+1} - x\|^2 - \frac{\eta^t}{2}\|x^t - x\|^2\right] - f_t(x) \leq \mathbb{E}\left[\left(\frac{\eta^{t+1} - \eta^t}{2} - \frac{\beta}{2}\right)\|x^t - x\|^2 + \frac{G^2}{2\eta^{t+1}}\right]$$

Summing up the above inequality over $t = 1, 2, \ldots, T$ yields the desired inequality.

C.6.3 Proof of Lemma 3

Recall that the update formula of x_i^{t+1} in Algorithm 2 is

$$x_i^{t+1} \leftarrow \underset{x_i \in \mathcal{X}_i}{\arg\min}\{(x_i - x_i^t)^\top \xi_i^t + \frac{\eta_i^{t+1}}{2}\|x_i - x_i^t\|^2\}$$

The first-order optimality condition implies that

$$(x_i - x_i^{t+1})^\top \xi_i^t + \eta_i^{t+1}(x_i - x_i^{t+1})^\top (x_i^{t+1} - x_i^t) \geq 0, \quad \text{for all } x_i \in \mathcal{X}_i$$

Equivalently, we have

$$\frac{\eta_i^{t+1}}{2}(\|x_i^{t+1} - x_i\|^2 - \|x_i^t - x_i\|^2) \leq (x_i - x_i^{t+1})^\top \xi_i^t - \frac{\eta_i^{t+1}}{2}\|x_i^{t+1} - x_i^t\|^2$$

$$= (x_i - x_i^t)^\top \xi_i^t + (x_i^t - x_i^{t+1})^\top \xi_i^t - \frac{\eta_i^{t+1}}{2}\|x_i^{t+1} - x_i^t\|^2 \leq (x_i - x_i^t)^\top \xi_i^t + \frac{1}{2\eta_i^{t+1}}\|\xi_i^t\|^2$$

Letting $x = x^\star$ be a unique Nash equilibrium and rearranging the above inequality, we have

$$\frac{\eta_i^{t+1}}{2}\|x_i^{t+1} - x_i^\star\|^2 - \frac{\eta_i^t}{2}\|x_i^t - x_i^\star\|^2 \leq (x_i^\star - x_i^t)^\top \xi_i^t + \left(\frac{\eta_i^{t+1}}{2} - \frac{\eta_i^t}{2}\right)\|x_i^t - x_i^\star\|^2 + \frac{1}{2\eta_i^{t+1}}\|\xi_i^t\|^2$$

44

Summing up the above inequality over $i = 1, 2, \ldots, N$ and rearranging, we have

$$\sum_{i=1}^{N} \left(\eta_i^{t+1} \| x_i^{t+1} - x_i^\star \|^2 - \eta_i^t \| x_i^t - x_i^\star \|^2 \right)$$

$$\leq 2(x^\star - x^t)^\top \xi^t + \left(\sum_{i=1}^{N} (\eta_i^{t+1} - \eta_i^t) \| x_i^t - x_i^\star \|^2 \right) + \left(\max_{1 \leq i \leq N} \left\{ \frac{1}{\eta_i^{t+1}} \right\} \right) \| \xi^t \|^2$$

Note that $\{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}$ are generated independently of any noisy gradient feedback, we have

$$\mathbb{E}[(x^\star - x^t)^\top \xi^t \mid x^t, \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}] = \mathbb{E}[(x^\star - x^t)^\top v(x^t) \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}]$$

$$\mathbb{E}[\| \xi^t \|^2 \mid x^t, \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}] \leq G^2$$

Thus, we have

$$\sum_{i=1}^{N} \left(\eta_i^{t+1} \mathbb{E} \left[\| x_i^{t+1} - x_i^\star \|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T} \right] - \eta_i^t \mathbb{E} \left[\| x_i^t - x_i^\star \|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T} \right] \right)$$

$$\leq 2 \mathbb{E} \left[(x^\star - x^t)^\top v(x^t) \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T} \right]$$

$$+ \sum_{i=1}^{N} \left(\eta_i^{t+1} - \eta_i^t \right) \mathbb{E} \left[\| x_i^t - x_i^\star \|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T} \right] + G^2 \left(\max_{1 \leq i \leq N} \left\{ \frac{1}{\eta_i^{t+1}} \right\} \right)$$

Since a continuous game $\mathcal{G}$ is β-strongly monotone and $x^\star \in \mathcal{X}$ is a unique Nash equilibrium, we have

$$(x^\star - x^t)^\top v(x^t) \leq (x^\star - x^t)^\top v(x^\star) - \beta \| x^\star - x^t \|^2 \leq -\beta \| x^\star - x^t \|^2$$

Putting these pieces together yields that

$$\sum_{i=1}^{N} \left(\eta_i^{t+1} \mathbb{E} \left[\| x_i^{t+1} - x_i^\star \|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T} \right] - \eta_i^t \mathbb{E} \left[\| x_i^t - x_i^\star \|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T} \right] \right)$$

$$\leq \sum_{i=1}^{N} \left(\eta_i^{t+1} - \eta_i^t - 2\beta \right) \mathbb{E} \left[\| x_i^t - x_i^\star \|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T} \right] + G^2 \left(\max_{1 \leq i \leq N} \left\{ \frac{1}{\eta_i^{t+1}} \right\} \right)$$

Since $D > 0$ satisfies that $\| x - x' \| \leq D$ for all $x, x' \in \mathcal{X}$, we have

$$\sum_{i=1}^{N} \left(\eta_i^{t+1} \mathbb{E} \left[\| x_i^{t+1} - x_i^\star \|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T} \right] - \eta_i^t \mathbb{E} \left[\| x_i^t - x_i^\star \|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T} \right] \right)$$

$$\leq \left(\max_{1 \leq i \leq N} \{ \eta_i^{t+1} - \eta_i^t \} - 2\beta \right) \mathbb{E} \left[\| x^t - x^\star \|^2 \mid \{\eta_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T} \right] + G^2 \left(\max_{1 \leq i \leq N} \left\{ \frac{1}{\eta_i^{t+1}} \right\} \right)$$

$$\leq D^2 \left(\max \left\{ 0, \max_{1 \leq i \leq N} \{ \eta_i^{t+1} - \eta_i^t \} - 2\beta \right\} \right) + G^2 \left(\max_{1 \leq i \leq N} \left\{ \frac{1}{\eta_i^{t+1}} \right\} \right)$$

Summing over $t = 1, 2, \ldots, T - 1$ yields the desired inequality.

45

C.6.4 Proof of Lemma 4

Recall that the update formula of x_i^{t+1} in either the multi-agent ONS is

$$x_i^{t+1} \leftarrow \operatorname*{argmin}_{x_i \in \mathcal{X}_i} \{ (x_i - x_i^t)^\top v_i(x^t) + \frac{\eta_i}{2}(x_i - x_i^t)^\top A_i^{t+1}(x_i - x_i^t) \}$$

The first-order optimality condition implies that

$$(x_i - x_i^{t+1})^\top v_i(x^t) + \eta_i(x_i - x_i^{t+1})^\top A_i^{t+1}(x_i^{t+1} - x_i^t) \geq 0, \quad \text{for all } x_i \in \mathcal{X}_i$$

Equivalently, we have

$$\frac{\eta_i}{2}((x_i^{t+1} - x_i)^\top A_i^{t+1}(x_i^{t+1} - x_i) - (x_i^t - x_i)^\top A_i^{t+1}(x_i^t - x_i))$$

$$\leq \quad (x_i - x_i^{t+1})^\top v_i(x^t) - \frac{\eta_i}{2}(x_i^{t+1} - x_i^t)^\top A_i^{t+1}(x_i^{t+1} - x_i^t)$$

$$= \quad (x_i - x_i^t)^\top v_i(x^t) + (x_i^t - x_i^{t+1})^\top v_i(x^t) - \frac{\eta_i}{2}(x_i^{t+1} - x_i^t)^\top A_i^{t+1}(x_i^{t+1} - x_i^t)$$

$$\leq \quad (x_i - x_i^t)^\top v_i(x^t) + \frac{1}{2\eta_i} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t)$$

Rearranging this inequality, we have

$$\frac{\eta_i}{2}(x_i^{t+1} - x_i)^\top A_i^{t+1}(x_i^{t+1} - x_i) - \frac{\eta_i}{2}(x_i^t - x_i)^\top A_i^t(x_i^t - x_i)$$

$$\leq \quad (x_i - x_i^t)^\top v_i(x^t) + (x_i^t - x_i)^\top \left(\frac{\eta_i}{2} A_i^{t+1} - \frac{\eta_i}{2} A_i^t \right)(x_i^t - x_i) + \frac{1}{2\eta_i} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t)$$

Summing over $i = 1, 2, \ldots, N$, we have

$$\sum_{i=1}^{N} \left(\frac{\eta_i}{2}(x_i^{t+1} - x_i)^\top A_i^{t+1}(x_i^{t+1} - x_i) - \frac{\eta_i}{2}(x_i^t - x_i)^\top A_i^t(x_i^t - x_i) \right)$$

$$\leq \quad (x - x^t)^\top v(x^t) + \frac{1}{2} \left(\sum_{i=1}^{N} (x_i^t - x_i)^\top \left(\eta_i A_i^{t+1} - \eta_i A_i^t \right)(x_i^t - x_i) \right) + \frac{1}{2} \left(\sum_{i=1}^{N} \frac{1}{\eta_i} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t) \right)$$

Since $\mathcal{G}$ is α-exp-concave and satisfies that $\|v(x)\| \leq G_i$ and $\|x - x'\| \leq D_i$ for all $x_i, x_i' \in \mathcal{X}_i$, we have

$$\langle x - x^t, v(x) - v(x^t) \rangle \geq \frac{1}{4} \left(\sum_{i=1}^{N} \min\{ \frac{1}{4G_i D_i}, \alpha \}(x_i^t - x_i)^\top (v_i(x)v_i(x)^\top + v_i(x^t)v_i(x^t)^\top)(x_i^t - x_i) \right)$$

Putting these pieces together yields that

$$\sum_{i=1}^{N} \left(\frac{\eta_i}{2}(x_i^{t+1} - x_i)^\top A_i^{t+1}(x_i^{t+1} - x_i) - \frac{\eta_i}{2}(x_i^t - x_i)^\top A_i^t(x_i^t - x_i) \right)$$

$$\leq \quad (x - x^t)^\top v(x) + \sum_{i=1}^{N} (x_i^t - x_i)^\top (\frac{\eta_i}{2} A_i^{t+1} - \frac{\eta_i}{2} A_i^t - \frac{1}{4} \min \left\{ \frac{1}{4GD}, \alpha \right\} v_i(x^t)v_i(x^t)^\top)(x_i^t - x_i)$$

$$+ \frac{1}{2} \left(\sum_{i=1}^{N} \frac{1}{\eta_i} v_i(x^t)^\top (A_i^{t+1})^{-1} v_i(x^t) \right)$$

Summing over $t = 1, 2, \ldots, T$ yields the desired inequality.

D   A Toolbox Result

We provide a simple result on the maximum of independent identically distributed (i.i.d.) geometric random variables.

Proposition 3. *Let $X_1, \ldots, X_n$ be n i.i.d. geometric random variables: $X_i \sim$ Geometric(p_0) for all $i = 1, 2, \ldots, n$ and with $p_0 \in (0, 1)$. Defining $\bar{X}_n = \max_{1 \leq i \leq n} X_i$, we have*

$$\sum_{n=1}^{+\infty} \mathbb{P}(\bar{X}_n \leq x) \leq e^{xp_0} \tag{29}$$

and

$$1 \leq \mathbb{E}[\bar{X}_n] \leq \frac{1 + \log(n)}{p_0} \tag{30}$$

Proof of Proposition 3. Let $q_0 = 1 - p_0$. Fixing a constant $x \geq 1$, we have

$$\mathbb{P}(\bar{X}_n \leq x) = \mathbb{P}(\max\{X_1, X_2, \ldots, X_n\} \leq x) = \mathbb{P}(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x)$$

Since $X_1, X_2, \ldots, X_n$ are n independently distributed random variables, we have

$$\mathbb{P}(\bar{X}_n \leq x) = \mathbb{P}(X_1 \leq x)\mathbb{P}(X_2 \leq x) \cdots \mathbb{P}(X_n \leq x)$$

Since $X_i \sim$ Geometric(p_0) for all $i = 1, 2, \ldots, n$, we have

$$\mathbb{P}(X_i \leq x) = \sum_{k=1}^{\lfloor x \rfloor} \mathbb{P}(X_i = k) = \sum_{k=1}^{\lfloor x \rfloor} (1 - p_0)^{k-1} p_0 = p_0 \cdot \frac{1 - (1 - p_0)^{\lfloor x \rfloor}}{p_0} = 1 - q_0^{\lfloor x \rfloor}$$

where $\lfloor x \rfloor$ is the largest integer that is smaller than x. As such, we have $\mathbb{P}(\bar{X}_n \leq x) = (1 - q_0^{\lfloor x \rfloor})^n$. Given that $q_0 \in (0, 1)$, we have

$$\sum_{n=1}^{+\infty} \mathbb{P}(\bar{X}_n \leq x) = \sum_{n=1}^{+\infty} (1 - q_0^{\lfloor x \rfloor})^n = (1 - q_0^{\lfloor x \rfloor}) \cdot \frac{1}{q_0^{\lfloor x \rfloor}} = \frac{1}{q_0^{\lfloor x \rfloor}} - 1 \leq \frac{1}{q_0^x} = (1 - p_0)^{-x}$$

Since $1 + x \leq e^x$ for all $x \in \mathbb{R}$, we have $1 - p_0 \leq e^{-p_0}$. Putting these pieces together yields Eq. (29).

Moreover, it follows from the definition of $\bar{X}_n$ that $\bar{X}_n \geq 1$ and hence $\mathbb{E}[\bar{X}_n] \geq 1$. We also have

$$\mathbb{E}[\bar{X}_n] = \sum_{k=0}^{+\infty} \mathbb{P}(\bar{X}_n > k) = \sum_{k=0}^{+\infty} (1 - (1 - q_0^k)^n)$$

By viewing the infinite sum as an Riemann sum approximation of an integral, we obtain

$$\sum_{k=0}^{+\infty} (1 - (1 - q_0^k)^n) \leq 1 + \int_0^{+\infty} (1 - (1 - q_0^x)^n) \, dx$$

We perform the change of variables $u = 1 - q_0^x$ and obtain

$$
\begin{aligned}
\int_0^{+\infty} (1 - (1 - q_0^x)^n) dx &= -\frac{1}{\log(q_0)} \int_0^1 \frac{1 - u^n}{1 - u} \, du = -\frac{1}{\log(q_0)} \int_0^1 (1 + u + \ldots + u^{n-1}) \, du \\
&= -\frac{1}{\log(q_0)} (1 + \frac{1}{2} + \ldots + \frac{1}{n}) \leq -\frac{1}{\log(q_0)} \left(1 + \int_1^n \frac{1}{x} dx \right) \\
&= -\frac{1 + \log(n)}{\log(1 - p_0)}
\end{aligned}
$$

47

Recalling again that $1 + x \leq e^x$ for all $x \in \mathbb{R}$, we have $\log(1 - p_0) \leq -p_0$, which implies $-\frac{1}{\log(1-p_0)} \leq \frac{1}{p_0}$. Putting these pieces together yields Eq. (30). $\qquad\square$