Fast Rates in Time-Varying Strongly Monotone Games

Yu-Hu Yan [1] **Peng Zhao** [1] **Zhi-Hua Zhou** [1]

Abstract

Multi-player online games depict the interaction of multiple players with each other over time. Strongly monotone games are of particular interest since they have benign properties and also relate to many classic games that have applications in real life. Existing works mainly focus on the time-invariant case with provable guarantees established. However, the research of the more general *time-varying* games in changing environments is underexplored and the best-known result cannot match the guarantees in the time-invariant case. In this work, we present a new decentralized online algorithm for time-varying strongly monotone games, which greatly improves existing results and obtains fast rates, matching the best time-invariant guarantee without knowing the environmental non-stationarity. Furthermore, to achieve faster rates, we generalize the RVU property with smoothness and establish a series of problem-dependent bounds that also match the best time-invariant one. To realize all those results, we make a comprehensive use of the techniques in non-stationary and universal online learning.

1. Introduction

Multi-player online games (Daskalakis et al., 2011; Rakhlin & Sridharan, 2013b; Syrgkanis et al., 2015) is a versatile model that depicts the interaction of multiple players over time. At each round, each player makes a decision from a convex compact set, and meanwhile, the environment selects convex loss functions (also called *utility functions*) for them. Then each player suffers a loss decided by their own utility function and the joint decisions of all players. A fundamental task of game-theoretic learning is to find the *Nash equilibrium*, a stable set of decisions where no player has incentives to deviate (Nash Jr, 1950). However, solving an accurate Nash equilibrium is generally computation-hard or even PPAD-complete (Daskalakis et al., 2009). Fortunately, it has been revealed that when games exhibit certain benign properties, effective algorithms can be developed to achieve Nash equilibriums with provable guarantees. Particularly, recent advances show the great potential of the regret minimization framework (Freund & Schapire, 1999) for multi-player online games. Indeed, when each player minimizes their own regret, the joint decision eventually converges to a coarse correlated equilibrium. However, it may differ from the Nash equilibrium. Fortunately, with more structures in games, regret minimization algorithms can provably achieve Nash equilibriums.

Among those successes, *strongly monotone games* (Rosen, 1965) is a significant subclass that encompasses many real-world applications of interest and has been extensively studied due to its benign mathematical properties (Monderer & Shapley, 1996; Nemirovski et al., 2010). It is demonstrated that, when appropriate regret minimization algorithms are deployed to all players, the distance between their decisions and Nash equilibrium can provably converge to zero (Facchinei & Pang, 2003; Bravo et al., 2018). Existing results on strongly monotone games primarily concern the *time-invariant* case, i.e., with fixed utility functions. However, in real-world applications, many game-related scenarios are *time-varying*. For instance, an important game-theoretic application is the Cournot competition (Monderer & Shapley, 1996), where multiple firms provide goods to the market, and the goods are then priced as a function of the total supply. Due to various factors such as weather, holidays, politics, etc., the supply-market relationship is often subject to change, while existing studies for time-invariant scenarios cannot address this issue.

The only existing work for time-varying strongly monotone games is by Duvocelle et al. (2023), but the attained results are not favorable enough. Specifically, they investigated the *distance tracking error*, the distance between the decisions and the Nash equilibrium (a formal definition is introduced in (2.1)), and further proposed a restart-based algorithm to handle the non-stationarity, attaining an $\mathcal{O}(\sqrt{T}+T^{2/3}P_T^{1/3})$ guarantee, where T is the time horizon, and P_T measures the environmental non-stationarity. Notably, the result implies an $\mathcal{O}(\sqrt{T})$ tracking error in the time-invariant case (where $P_T = 0$), which unfortunately exhibits a large gap compared

[1]National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China. Correspondence to: Peng Zhao <zhaop@lamda.nju.edu.cn>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

Table 1: A summary of tracking error guarantees for time-varying strongly monotone games. The first column shows two setups about non-smooth and smooth games. The second column presents where the results are from. The second column provides the main results, where P_T, V_T, and W_T are different non-stationarity measures that are at most $\mathcal{O}(T)$ and become 0 in the time-invariant case. The third column shows implications to the time-invariant case, where our results match the best-known results, i.e., $\widetilde{\mathcal{O}}(1)$ for general utility functions (Bravo et al., 2018, Theorem 7) and $\mathcal{O}(1)$ for smooth functions (Facchinei & Pang, 2003, Section 12.3.2).

Setups	Works	Time-Varying Games	Time-Invariant Games
Non-Smooth Games	Duvocelle et al. (2023)	$\mathcal{O}(\sqrt{T}+T^{2/3}P_T^{1/3})$	$\mathcal{O}(\sqrt{T})$
	This Paper (Theorem 3)	$\widetilde{\mathcal{O}}(1+\min\{T^{1/3}P_T^{2/3}, W_T\})$	$\widetilde{\mathcal{O}}(1)$
Smooth Games	This Paper (Theorem 5)	$\mathcal{O}(\min\{\sqrt{(1+V_T+P_T)(1+P_T)}, 1+W_T\})$	$\mathcal{O}(1)$

to $\mathcal{O}(\log T)$, the best-known time-invariant result (Bravo et al., 2018). It is thus natural to ask for a more thorough investigation of time-varying strongly monotone games.

This paper presents a more comprehensive characterization of time-varying strongly monotone games. To optimize the distance tracking error, we propose a new *decentralized* online algorithm that achieves an $\widetilde{\mathcal{O}}(1+T^{1/3}P_T^{2/3})$ tracking error bound, where $\widetilde{\mathcal{O}}(\cdot)$ omits the poly-logarithmic dependence in T. The advantages of our result lie in three aspects: *(i)* Our result significantly improves upon the previous best-known $\mathcal{O}(\sqrt{T}+T^{2/3}P_T^{1/3})$ rate (Duvocelle et al., 2023), and importantly, it implies an $\widetilde{\mathcal{O}}(1)$ tracking error when specializing to the time-invariant scenario, hence matching the corresponding best-known $\widetilde{\mathcal{O}}(1)$ result (Bravo et al., 2018); *(ii)* Our algorithm does not require to know P_T, the variation of Nash equilibriums (a formal definition is deferred to Section 2.2), that is actually unknown in advance. In contrast, this quantity is required by Duvocelle et al. (2023); *(iii)* Our algorithm exhibits adaptivity to the strong monotonicity in the sense that the monotonicity coefficient is also not required. We further contribute an orthogonal improvement by showing that our algorithm additionally enjoys an $\widetilde{\mathcal{O}}(1+W_T)$ guarantee, where W_T quantifies the variance of the gradients of utility functions. Consequently, our algorithm can take advantage of both slow Nash equilibrium variation and small gradient variance *simultaneously*. The key to our improvement lies in a *novel* analysis to construct carefully designed strongly convex surrogate loss functions by the virtue of strong monotonicity.

To step further, we consider the possibility of obtaining even faster tracking error rates. Assuming the *smoothness* of the utility functions, we generalize the Regret bounded by Variation in Utilities (RVU) condition (Syrgkanis et al., 2015), a key property for fast-rate convergence in finite games, to continuous multi-player games under the time-varying scenarios inspired by the recent study of time-varying zero-sum games (Zhang et al., 2022c). Using the classic optimistic online gradient descent algorithm (Rakhlin & Sridharan, 2013a), we derive a series of *problem-dependent* bounds. Specifically, we obtain an $\mathcal{O}(\sqrt{(1+V_T+P_T)(1+P_T)})$ tracking error, where V_T is a problem-dependent quantity that is at most $\mathcal{O}(T)$ but can be much smaller in benign environments. In addition, our algorithm also benefits from small gradient variance with an $\mathcal{O}(1+W_T)$ bound. Note that our new results are faster than $\widetilde{\mathcal{O}}(1+T^{1/3}P_T^{2/3})$ without smoothness. For instance, in a game with S switches, the non-smooth guarantee ensures an $\widetilde{\mathcal{O}}(1+T^{1/3}S^{2/3})$ rate. In contrast, the results here give an optimal $\mathcal{O}(1+S)$ bound, matching the performance of the oracle learner who restarts once a switch happens and runs a time-invariant algorithm within each stationary period (thus suffering an $\mathcal{O}(S)$ bound in total, due to $\mathcal{O}(1)$ tracking error in each period and overall S periods). However, the aforementioned guarantees require different configurations (of the step sizes). To this end, we leverage a two-layer framework by properly hedging over many possibilities and finally obtain the same guarantees with a single algorithm. Table 1 summarizes the existing result and ours in non-smooth and smooth cases.

Notably, in *common interest games*, a setting usually encountered in distributed optimization problems (Gopal & Yang, 2013), where the utility functions remain the same across players, all our results hold for the newly proposed *utility tracking error*, which serves as an upper bound of the distance version and is thus more fundamental in this case.

Techniques. To achieve all those fast-rate guarantees, we make comprehensive use of recent online convex optimization techniques for non-stationarity (Zhao et al., 2021) and universality (van Erven & Koolen, 2016). In the non-smooth case, the key improvements over (Duvocelle et al., 2023) stem from two aspects. First, it is possible to directly handle the non-stationarity with *online ensemble* (Zhou, 2012; Zhao, 2021), which is realized by employing a group of online gradient descent (OGD) algorithms (Zinkevich, 2003) with different configurations as base learners and a meta learner to track the best one on the fly. As a result, the two-layer algorithm can track the moving clairvoyant and does not require P_T as input. The second improvement is carefully designed strongly convex surrogate loss functions utilizing the virtue of strong monotonicity, which allows us to leverage the recent progress of non-stationary online learning with strongly convex losses.

For the smooth case, our results draw inspiration from time-varying two-player zero-sum games (Zhang et al., 2022c) but with additional innovations. On the one hand, we generalize the key RVU condition to more complex scenarios with multiple players and general utility functions so that our bounds can safeguard the tracking error; alternatively, Zhang et al. (2022c) only bounded quantities related to the summation of two players' regret and thus cannot protect the tracking error. On the other hand, to obtain favorable results, we need a novel usage of correction terms for online ensemble, which hinges on the gradient-variation dynamic regret for online convex optimization (Zhao et al., 2021).

The rest is organized as follows. Section 2 formulates the problem. Section 3 proposes our fast-rate algorithm for time-varying strongly monotone games. Section 4 achieves faster rates with smoothness. Section 5 provides the empirical evaluations and finally Section 6 concludes the work.

2. Problem Setup

This section introduces the performance and non-stationarity measures in the time-varying strongly monotone games and a particular class called common interest games.

2.1. Time-Varying Strongly Monotone Games

A multi-player time-varying game contains T rounds and $N \geq 2$ players. In the t-th round, the i-th player ($i \in [N]$) chooses a decision $x_{t,i}$ from a compact convex set $\mathcal{X}_i \subseteq \mathbb{R}^d$. Simultaneously, the environments reveal a group of *time-varying* utility functions $u_{t,i} : \mathcal{X} \mapsto \mathbb{R}$ for each player, where $\mathcal{X} \triangleq \mathcal{X}_1 \times \ldots \times \mathcal{X}_N$. Afterwards, each player receives her local gradient feedback $v_{t,i}(\mathbf{x}_t)$, where $\mathbf{x}_t \triangleq (x_{t,1}, \ldots, x_{t,N})$ and $v_{t,i}(\mathbf{x}_t) \triangleq \nabla_{x_{t,i}} u_{t,i}(\mathbf{x}_t)$.

For the i-th player, a Nash equilibrium is a joint decision $\mathbf{x}^\star$ such that $u_{t,i}(x_i^\star; x_{-i}^\star) \leq u_{t,i}(x_i; x_{-i}^\star)$ for any $x_i \in \mathcal{X}_i$, where $x_{-i} \triangleq (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N)$. Variational inequality is also used to describe a Nash equilibrium: $\langle v_t(\mathbf{x}^\star), \mathbf{x} - \mathbf{x}^\star \rangle \geq 0$, where the global gradient $v_t(\mathbf{x}) \triangleq (v_{t,1}(\mathbf{x}), \ldots, v_{t,N}(\mathbf{x}))$. Since solving an accurate Nash equilibrium is computation-hard in general multi-player games, we focus on those with certain benign properties, especially strongly monotone games (Rosen, 1965).

Definition 1 (Strong Monotonicity). *A game with utility gradient v is μ-strongly monotone if $\langle v(\mathbf{x}) - v(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \mu \|\mathbf{x} - \mathbf{y}\|_2^2$ holds for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, where $\mu > 0$.*

Monotone games include many classic games close to real-world applications, including Cournot competition (Monderer & Shapley, 1996), Kelly auctions and Tullock competitions (Nemirovski et al., 2010), signal covariance and power control problems in wireless communications (d'Oro et al., 2015; Mertikopoulos & Moustakas, 2015), and so on. We refer readers to Facchinei & Kanzow (2010) for more applications. Besides, since a monotone game admits a unique Nash equilibrium (Rosen, 1965), we denote by $\mathbf{x}_t^\star$ the Nash equilibrium of the game of the t-th round.

In game theory and convex optimization, a well-studied performance measure is the *distance tracking error*, which reflects the algorithm's ability to chase some target measured by norm distance. In multi-player games, a natural target is the Nash equilibrium, and thus we investigate

$$\textsc{Dist-Err} \triangleq \sum_{t=1}^{T} \|\mathbf{x}_t - \mathbf{x}_t^\star\|^2 . \tag{2.1}$$

In time-invariant strongly monotone games, where $\mathbf{x}^\star$ is used to denote the unique Nash equilibrium, the distance tracking error $\|\mathbf{x}_t - \mathbf{x}^\star\|^2$ enjoys an $\mathcal{O}(t^{-1})$ last-iterate convergence (Bravo et al., 2018, Theorem 7) and can be improved to $\mathcal{O}(\rho^t)$ with smooth utility functions (Facchinei & Pang, 2003), where $\rho \in (0, 1)$. We end this part by listing the assumptions and notations used throughout the work.

Assumption 1. For any $i \in [N]$, the gradient satisfies $\|v_{t,i}(\cdot)\|, \|v_t(\cdot)\| \leq G$ and the domain satisfies $\|\mathbf{x} - \mathbf{y}\| \leq D$ for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $\|x - y\| \leq D$ for any $x, y \in \mathcal{X}_i$.

Assumption 2. All games with utility gradients $\{v_t\}_{t=1}^T$ are μ-strongly monotone (Definition 1).

We use Δ_d for a d-dimensional simplex, $\{a_t\}_{t=1}^T$ for the sequence $a_1, \ldots, a_T$, $\|\cdot\|$ for ℓ_2-norm in default. $a \lesssim b$ represents $a \leq \widetilde{\mathcal{O}}(b)$, where $\widetilde{\mathcal{O}}(\cdot)$-notation omits logarithmic factors in time horizon T.

2.2. Non-Stationarity Measures

We introduce the following non-stationarity measures to capture the varying intensity of the time-varying games:

- *Path length*: $P_T \triangleq \sum_{t=2}^T \|\mathbf{x}_t^\star - \mathbf{x}_{t-1}^\star\|$ measures the variation of the time-varying Nash equilibriums.
- *Gradient variation*: $V_T \triangleq \sum_{t=2}^T \sup_{\mathbf{x} \in \mathcal{X}} \|v_t(\mathbf{x}) - v_{t-1}(\mathbf{x})\|^2$ denotes the variation in gradients.
- *Gradient variance*: $W_T \triangleq \sum_{t=1}^T \sup_{\mathbf{x} \in \mathcal{X}} \|v_t(\mathbf{x}) - \bar{v}_T(\mathbf{x})\|$ reflects the variance of the gradients, where $\bar{v}_T(\cdot) = \sum_{t=1}^T v_t(\cdot)/T$ is the average gradient.

The three measures reflect different aspects of the games and are generally incomparable. We will give more detailed discussions in the rest of the paper.

2.3. Common Interest Games

Our setup in Section 2.1 exactly matches that of Duvocelle et al. (2023). Besides, we also investigate a particular class called *common interest games*, where the utility functions remain the same across players, i.e., $u_{t,i} = u_t$ for all $i \in [N]$.

This game is worth studying since it is widely encountered in distributed optimization problems, where the problem dimension is pretty large (Gopal & Yang, 2013).

In common interest games, we propose a more fundamental performance measure, which evaluates the function level. Specifically, we define *utility tracking error*:

$$\textsc{Util-Err} \triangleq \sum_{t=1}^{T} u_t(\mathbf{x}_t) - \sum_{t=1}^{T} u_t(\mathbf{x}_t^\star). \tag{2.2}$$

Note that (2.1) and (2.2) are incomparable in general, but we identify that, in common interest strongly monotone games, the utility tracking error serves as a natural upper bound of the distance tracking error, showing that (2.2) is more fundamental in this particular setup. See Proposition 1 for a formal statement with the proof in Appendix B.

Proposition 1. *The utility function u_t of a μ-strongly monotone game is μ-strongly convex, and the utility tracking error (2.2) can upper-bound the distance tracking error (2.1), specifically, $\mu\textsc{Dist-Err} \leq 2\textsc{Util-Err}$.*

3. Fast Rates for Strongly Monotone Games

This section presents our fast-rate results for time-varying strongly monotone games. First, Section 3.1 reviews the latest work and identifies the unsatisfactory guarantees. Then, Section 3.2 and Section 3.3 introduce our solution with a fast tracking error rate. Finally, Section 3.4 shows that our new method can also take advantage of small gradient variance.

3.1. Reviewing Latest Result

In this part, we review the algorithm and sketch the analysis of Duvocelle et al. (2023). They focused on the distance tracking error (2.1). By definition of strong monotonicity and property of Nash equilibriums, they upper-bounded the distance tracking error as $\mu\textsc{Dist-Err} \leq \sum_{t\in[T]}\langle v_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^\star\rangle$. Notably, the right-hand side is essentially the regret over linear losses $\{\langle v_t(\mathbf{x}_t), \cdot\rangle\}_{t=1}^T$ against a sequence of changing comparators $\{\mathbf{x}_t^\star\}_{t=1}^T$, which is unknown and thus in general hard to handle.

To overcome this issue, Duvocelle et al. (2023) leveraged a *restarting* strategy in analysis,[1] a common way to deal with non-stationarity (Besbes et al., 2015; Zhao et al., 2020a). Specifically, dividing the time horizon T into $K = \lceil T/\Delta \rceil$ periods of length Δ, $\sum_{t\in[T]}\langle v_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^\star\rangle$ (upper bound of distance tracking error) can be decomposed as

$$\sum_{k=1}^{K}\sum_{t\in\Delta_k}\langle v_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{v}_k\rangle + \sum_{k=1}^{K}\sum_{t\in\Delta_k}\langle v_t(\mathbf{x}_t), \mathbf{v}_k - \mathbf{x}_t^\star\rangle,$$

[1] We call it a method with restarts because it decomposes the regret into multiple periods, just like restart-based algorithms. Notably, the algorithm does not need to restart.

where Δ_k denotes the k-th period and $\{\mathbf{v}_k\}_{k=1}^K$ is a period-wise stationary comparator sequence. The first term is the summation of the static regret of all periods, and the second term represents the gap between two comparator sequences: $\{\mathbf{v}_k\}_{k=1}^K$ and $\{\mathbf{x}_t^\star\}_{t=1}^T$. For the first term, since the comparator is fixed as $\mathbf{v}_k$ inside the k-th period, an algorithm with static regret guarantees, e.g., OGD (Zinkevich, 2003), gives an $\mathcal{O}(\sqrt{\Delta})$ regret inside each period with step size $1/\sqrt{\Delta}$. The second term, using the period comparison technique from Besbes et al. (2015), can be bounded by $\mathcal{O}(\Delta P_T)$, where $P_T \triangleq \sum_{t=2}^T \|\mathbf{x}_t^\star - \mathbf{x}_{t-1}^\star\|$ is the path length of the Nash equilibriums. Summing together gives an upper bound of $\mathcal{O}(T/\Delta \cdot \sqrt{\Delta} + \Delta P_T)$. Choosing the period length optimally as $\Delta = \min\{(T/P_T)^{2/3}, T\}$, which is required in the step size of OGD, gives an $\mathcal{O}(\sqrt{T} + T^{2/3}P_T^{1/3})$ bound.

However, their result is not satisfactory enough. Consider the simplest time-invariant case, i.e., the Nash equilibrium is fixed $\mathbf{x}_1^\star = \ldots = \mathbf{x}_T^\star$ (so that $P_T = 0$). Their result implies an $\mathcal{O}(\sqrt{T})$ tracking error, which is exponentially worse than $\mathcal{O}(\log T)$, the best-known result for static strongly monotone games (Bravo et al., 2018, Theorem 7).

3.2. Tracking the Non-Stationarity Directly

The key idea of Duvocelle et al. (2023) is to divide the time horizon into multiple periods and treat the games inside each period as a time-invariant one in response to the non-stationarity of time-varying games. However, this may not be the best way. Indeed, in this online game setup, the feedback of each player is relatively adequate as they can observe the *gradient* information. Thus, instead of restarting a stationary algorithm, we manage to design online algorithms capable of *directly* tracking the non-stationarity and catching up with the changing nature.

With this high-level sense in mind, we present a new decomposition for the distance tracking error,

$$\sum_{t=1}^{T}\langle v_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^\star\rangle = \sum_{i=1}^{N}\sum_{t=1}^{T}\langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star\rangle,$$

which is the summation of all players' regret. The regret of the i-th player is defined on linear losses $\{\langle v_{t,i}(\mathbf{x}_t), \cdot\rangle\}_{t=1}^T$ against a sequence of *time-varying* comparators $\{x_{t,i}^\star\}_{t=1}^T$, and notably the latter is related to Nash equilibriums and is unknown in advance, or even after the game ends.

Benchmarking online performance against a sequence of changing comparators is the purpose of *dynamic regret* precisely, a central target of non-stationary online learning (Zinkevich, 2003). It is demonstrated that OGD has such tracking ability. Specifically, let us consider a general online convex optimization (OCO) setup, where the online learner submits decisions $\{\mathbf{w}_t\}_{t=1}^T$ inside a convex feasible domain $\mathcal{W} \subseteq \mathbb{R}^d$ and competes with changing comparators

$\{\mathbf{u}_t\}_{t=1}^T$ over convex losses $\{f_t(\cdot)\}_{t=1}^T$. A commonly used performance measure is dynamic regret:

$$\text{D-REG}(\{\mathbf{u}_t\}_{t=1}^T) \triangleq \sum_{t=1}^T f_t(\mathbf{w}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t). \qquad (3.1)$$

OGD, which updates via $\mathbf{w}_{t+1} = \Pi_{\mathcal{W}}[\mathbf{w}_t - \eta \nabla f_t(\mathbf{w}_t)]$, provably optimizes (3.1) as $\text{D-REG}(\{\mathbf{u}_t\}_{t=1}^T) \le \mathcal{O}(P_T/\eta + \eta T)$ with step size $\eta > 0$, where the path length of comparators $P_T \triangleq \sum_{t=2}^T \|\mathbf{u}_t - \mathbf{u}_{t-1}\|$ essentially measures the non-stationarity intensity and the regret bound of OGD scales with it. It is important that we now obtain an online learner that can directly track the changing comparators, *without any restarts* used in (Duvocelle et al., 2023).

In our problem, if each player knew her own path length $P_{T,i}$ defined on $\{x_{t,i}^\star\}_{t=1}^T$, her own regret would be bounded by at most $\mathcal{O}(\sqrt{T(1+P_{T,i})})$ by running OGD with step size $\eta = \mathcal{O}(\sqrt{P_{T,i}/T})$, leading to an $\mathcal{O}(\sqrt{T(1+P_T)})$ tracking error, which already improves the $\mathcal{O}(\sqrt{T} + T^{2/3}P_T^{1/3})$ bound of Duvocelle et al. (2023). However, the path-length quantities $\{P_{T,i}\}_{i=1}^N$ are in general unavailable. To address this issue, inspired by the recent progress in non-stationary OCO (Zhang et al., 2018), we deploy a two-layer online ensemble algorithm with totally M base learners and a meta learner. The i-th base learner maintains her own decision $\mathbf{w}_{t,i}$ and optimizes the linearized loss $\langle \nabla f_t(\mathbf{w}_t), \cdot \rangle$ by OGD with her own step size based on a guess of the path length. The meta learner maintains weights $\{p_{t,i}\}_{i=1}^M$ for each base learner and uses an expert-tracking algorithm, e.g., Hedge (Freund & Schapire, 1997), to track the best base learner on the fly. Interested readers can refer to Algorithm 3 for algorithmic details. In the following, we provide the theoretical guarantees of our proposed algorithm, whose proof is presented in Appendix C.1.

Theorem 1. *Under Assumptions 1 and 2, if each player runs the aforementioned algorithm (Algorithm 3 in Appendix C.1), the distance tracking error enjoys an upper bound of $\mathcal{O}(\sqrt{T(1+P_T)})$, where $P_T \triangleq \sum_{t=2}^T \|\mathbf{x}_t^\star - \mathbf{x}_{t-1}^\star\|$ is the path length. Notably, our algorithm does not require path-length quantities (P_T or $\{P_{T,i}\}_{i=1}^N$) as inputs.*

Remark 1. Theorem 1 improves the result of Duvocelle et al. (2023, Theorem 2) in two aspects: *(i)* Theorem 1 improves the tracking error bound from $\mathcal{O}(T^{2/3}P_T^{1/3})$ to $\mathcal{O}(\sqrt{TP_T})$ since $P_T \le \mathcal{O}(T)$; and *(ii)* Theorem 1 is agnostic about P_T because the ensemble structure can eliminate the environmental uncertainty. In contrast, the optimal tuning of the period length Δ in the method of Duvocelle et al. (2023) requires the path length P_T as an input. ¶

3.3. Further Exploiting Strong Monotonicity

Although Theorem 1 already improves the existing result (Duvocelle et al., 2023), we discover that a further improvement is possible due to the virtue of strong monotonicity.

The key to the improvement lies in a novel analysis that extracts *strong convexity* from the definition of strong monotonicity (Definition 1), constructing a strongly convex surrogate upper bound for the distance tracking error, as shown in Proposition 2 below. The proof is deferred to Appendix C.2.

Proposition 2. *Under Assumption 2, it is guaranteed that $\mu\text{DIST-ERR} \le 2\sum_{i=1}^N \sum_{t=1}^T (\ell_{t,i}(x_{t,i}) - \ell_{t,i}(x_{t,i}^\star))$, where*

$$\ell_{t,i}(x) \triangleq \langle v_{t,i}(\mathbf{x}_t), x \rangle + \frac{\mu}{2}\|x - x_{t,i}\|^2 \qquad (3.2)$$

is a μ-strongly convex surrogate loss function.

We first provide the definition of strong convexity for self-containedness. Specifically, for any $\mathbf{w}, \mathbf{u} \in \mathcal{W}$, a function f is λ-strongly convex if $f(\mathbf{w}) - f(\mathbf{u}) \le \langle \nabla f(\mathbf{w}), \mathbf{w} - \mathbf{u} \rangle - \frac{\lambda}{2}\|\mathbf{w} - \mathbf{u}\|^2$. Clearly, $\ell_{t,i}(\cdot)$ in (3.2) is μ-strongly convex. The strongly convex surrogate loss function design allows us to leverage the recent progress of non-stationary online learning with strongly convex losses (Baby & Wang, 2022). Specifically, assuming the strong monotonicity μ to be known for a moment, recent studies show that if the loss functions are strongly convex, algorithms optimizing the strongly adaptive regret (Daniely et al., 2015), i.e., $\sum_{t\in\mathcal{I}} f_t(\mathbf{w}_t) - \sum_{t\in\mathcal{I}} f_t(\mathbf{u})$ inside an arbitrary interval $\mathcal{I} \subseteq [1, T]$, can also guarantee the dynamic regret.[2] To conclude, Baby & Wang (2022, Theorem 8) proves that $\sum_{t\in[T]} \ell_{t,i}(x_{t,i}) - \sum_{t\in[T]} \ell_{t,i}(x_{t,i}^\star) \le \widetilde{\mathcal{O}}(1 + T^{1/3}P_{T,i}^{2/3})$. Summing all players' regret gives an $\widetilde{\mathcal{O}}(1 + T^{1/3}P_T^{2/3})$ tracking error bound, which matches the best-known time-invariant result of $\widetilde{\mathcal{O}}(1)$ (Bravo et al., 2018).

Notably, the aforementioned result requires the knowledge of strong monotonicity μ in both the construction of surrogate loss and the setting of algorithmic parameters. However, in real applications, this quantity can be hard to estimate at the beginning of the online games or even unknown, which asks for online algorithms with more adaptivity. To address so, we leverage the technique of *universal OCO* (van Erven & Koolen, 2016; Wang et al., 2019), which aims to achieve optimal regret bounds for multiple types of loss functions, including convex, exp-concave, and strongly convex ones, using only a single algorithm.

We illustrate the high-level idea of universal methods in the standard OCO setup (introduced in Section 3.2). In OCO, universal algorithms (e.g., Wang et al. (2019)) guarantee $\sum_{t\in[T]} \langle \nabla f_t(\mathbf{w}_t), \mathbf{w}_t - \mathbf{u} \rangle \le \widetilde{\mathcal{O}}(\sqrt{Q_T})$ for any $\mathbf{u} \in \mathcal{W}$, where $Q_T \triangleq \sum_{t\in[T]} \|\mathbf{w}_t - \mathbf{u}\|^2$. If the loss functions are μ-strongly convex, the $\widetilde{\mathcal{O}}(\sqrt{Q_T})$ bound can be then

[2] The strongly adaptive regret minimization algorithm is run on each dimension. Therefore the aggregated decision lies in a box domain and is then projected into the feasible domain. The detailed algorithm is deferred to Algorithm 6 in Appendix C.3.

Algorithm 1 TV-SMOG for the i-th player

Input: parameter pool $\mathcal{H}$, domain diameter D, gradient upper bound G
Initialize: M instances of Algorithm 6 $\mathcal{A}_1, \ldots, \mathcal{A}_M$
for $t = 1, \ldots, T$ **do**
  Receive $x_{t,i,j}$ from $\mathcal{A}_j$
  Submit $x_{t,i} = \sum_{j=1}^{M} p_{t,i,j} a_j x_{t,i,j} / \sum_{j=1}^{M} p_{t,i,j} a_j$
  Receive gradient feedback $v_{t,i}(\mathbf{x}_t)$
  Update $p_{t+1,i,j}$ via $p_{t+1,i,j} \propto p_{t,i,j} \exp(-s_{t,i,j}(x_{t,i,j}))$
  for $j = 1, \ldots, M$ **do**
    Construct surrogate loss $s_{t,i,j}$ with $a_j \in \mathcal{H}$
    $\mathcal{A}_j$ updates to $x_{t+1,i,j}$ with surrogate loss $s_{t,i,j}$
  end
end

canceled by the negative term $-\mu Q_T$ induced by strong convexity, without the need to know the value of μ, while attaining the virtue of strong convexity. We note the existing universal methods only guarantee the static regret (with a fixed comparator), and we need to adapt them to the dynamic regret setting (changing comparators) to fit our problem.

In the following, we introduce the algorithmic details of our solution. Without loss of generality, we consider the case of $1/T \le \mu \le 1$, because if $\mu < 1/T$, even the optimal static regret $\Theta(\log T/\mu)$ (Hazan et al., 2007) is linear in T. On the contrary, functions with $\mu > 1$ are also 1-strongly convex. Inspired by Wang et al. (2019), we design a sequence of strongly convex surrogate loss functions, defined as:

$$s_{t,i,j}(x) \triangleq a_j \langle v_{t,i}(\mathbf{x}_t), x - x_{t,i} \rangle + a_j^2 G^2 \|x - x_{t,i}\|^2,$$

where the hyper-parameter a_j is chosen from a pool $\mathcal{H} \triangleq \{a_j | a_j = 2^{-j}/(3DG), j \in [M]\}$ with $M = \lceil \frac{1}{2} \log_2 T \rceil + 1$. The j-th base learner runs the algorithm of Baby & Wang (2022) (restated in Algorithm 6) with surrogate losses $\{s_{t,i,j}(\cdot)\}_{t=1}^{T}$. Given that different base learners are running over *heterogeneous* loss functions, combining them with a small regret overhead to the best base learner becomes challenging. To this end, our meta learner employs Titled Exponentially Weighted Average (TEWA), which can combine base learners with different functions and enjoy a small meta regret (van Erven & Koolen, 2016). Algorithm 1 concludes our method for Time-Varying Strongly Monotone Online Games (abbreviated as TV-SMOG), which notably does not require strong monotonicity coefficient as an input. The guarantee is shown below, and the proof is in Appendix C.3.

Theorem 2. *Under Assumptions 1 and 2, Algorithm 1 enjoys the following distance tracking error bound*

$$\text{DIST-ERR} \le \widetilde{\mathcal{O}}\left(1 + T^{1/3} P_T^{2/3}\right),$$

without knowing strong monotonicity μ and path length P_T.

Theorem 2 implies an $\widetilde{\mathcal{O}}(1)$ tracking error in time-invariant games, which further improves Theorem 1 and matches the best-known result (Bravo et al., 2018, Theorem 7).

Remark 2. We remind that Algorithm 1 is *decentralized*, in the sense that each player conducts theirss own algorithm without cooperating with others, and thus valuable for real applications (Hsieh et al., 2020; Sentenac et al., 2021). ¶

3.4. Taking Advantage of Small Gradient Variance

Theorem 2 obtains an $\widetilde{\mathcal{O}}(1 + T^{1/3} P_T^{2/3})$ tracking error bound, which is now the state-of-the-art scaling with the path length P_T. In this part, we further demonstrate that without any modification, Algorithm 1 can take advantage of small gradient variance, defined as $W_T \triangleq \sum_{t \in [T]} \sup_{\mathbf{x} \in \mathcal{X}} \|v_t(\mathbf{x}) - \bar{v}_T(\mathbf{x})\|$, where $\bar{v}_T(\cdot) = \sum_{t \in [T]} v_t(\cdot)/T$ is the averaged gradient. To do so, we begin with a different decomposition:

$$\text{DIST-ERR} \le 2\sum_{t=1}^{T} \|\mathbf{x}_t - \bar{\mathbf{x}}_T^\star\|^2 + 2\sum_{t=1}^{T} \|\bar{\mathbf{x}}_T^\star - \mathbf{x}_t^\star\|^2,$$

where $\bar{\mathbf{x}}_T^\star$ is the Nash equilibrium of the averaged game of T rounds. Briefly, Algorithm 1 upper-bounds the first term by $\widetilde{\mathcal{O}}(1)$ via reusing Theorem 2 since $\bar{\mathbf{x}}_T^\star$ is fixed. Thus it remains to analyze the second term, which measures the gap between the averaged and time-varying Nash equilibriums, i.e., $\bar{\mathbf{x}}_T^\star$ and $\{\mathbf{x}_t^\star\}_{t=1}^{T}$. We prove the second term is algorithm-irrelevant and can be bounded by $\mathcal{O}(W_T)$.

To conclude, using Algorithm 1, we obtain the following variance-based, and thus *best-of-both-worlds*, fast tracking error rate. The proof can be found in Appendix C.4.

Theorem 3 (Main Result)**.** *Under Assumptions 1 and 2, Algorithm 1 enjoys the following tracking error bound*

$$\text{DIST-ERR} \le \widetilde{\mathcal{O}}\left(1 + W_T\right),$$

without knowing the strong monotonicity μ and gradient variance W_T. With Theorem 2, our algorithm thus achieves

$$\text{DIST-ERR} \le \widetilde{\mathcal{O}}\left(1 + \min\left\{T^{1/3} P_T^{2/3}, W_T\right\}\right),$$

without knowing strong monotonicity μ, gradient variance W_T, and path length P_T.

Note that path length P_T and gradient variance W_T are generally incomparable. Thus both measures have their own merit, and our algorithm achieves the best of both worlds. Interested readers can refer to Appendix C of Zhang et al. (2022c) for more detailed discussions about their relationship in the simpler two-player zero-sum games.

Finally, in Corollary 1 below, we show that in common interest strongly monotone games (defined in Section 2.3), our results in Theorem 3 also hold for the more fundamental utility tracking error. The proof is deferred to Appendix C.4.

Corollary 1. *In common interest strongly monotone games, under the same assumptions of Theorem 3, Algorithm 1 enjoys the same guarantees for the utility tracking error.*

4. Faster Rates in Smooth Games

In this section, we investigate the possibility of obtaining even faster rates in time-varying strongly monotone games.

4.1. Leveraging RVU Property

In multi-player finite games, an essential property for fast-rate convergences is the *Regret bounded by Variation in Utilities (RVU)* (Syrgkanis et al., 2015), restated as follows.

Definition 2 (RVU Property)**.** *An online algorithm satisfies the RVU property with parameters $\alpha > 0$ and $0 < \beta \leq \gamma$, if its regret $\sum_{t\in[T]}\langle \mathbf{g}_t, \mathbf{w}_t - \mathbf{u}\rangle$ on any loss sequence $\{\mathbf{g}_t\}_{t=1}^T$ with respect to any comparator $\mathbf{u}$ is bounded by $\alpha + \beta\sum_{t\in[T]}\|\mathbf{g}_t - \mathbf{g}_{t-1}\|^2 - \gamma\sum_{t\in[T]}\|\mathbf{w}_t - \mathbf{w}_{t-1}\|^2$.*

Recent progress of online learning shows that the RVU property is satisfied by the well-known optimistic online gradient descent (OOGD) (Rakhlin & Sridharan, 2013b), shown as follows in our notations:

$$x_{t,i} = \Pi_{\mathcal{X}_i}[\widehat{x}_{t,i} - \eta_i m_{t,i}],\ \widehat{x}_{t+1,i} = \Pi_{\mathcal{X}_i}[\widehat{x}_{t,i} - \eta_i v_{t,i}(\mathbf{x}_t)],$$

where the intermediate decision $\widehat{x}_{t,i}$ updates using the true gradient $v_{t,i}(\mathbf{x}_t)$ and the submitted decision $x_{t,i}$ is generated from $\widehat{x}_{t,i}$ using an optimism $m_{t,i}$, which can be viewed as an optimistic estimation of the current gradient $v_{t,i}(\mathbf{x}_t)$.

Due to the property of finite games (e.g., Theorem 4 of Syrgkanis et al. (2015)), the gradient variation of the i-th player (i.e., $\|\mathbf{g}_t - \mathbf{g}_{t-1}\|^2$) can be bounded by the summation of other players' switching costs (i.e., $\sum_{j\neq i}\|x_{t,j} - x_{t-1,j}\|^2$), with which the summation of all players' regret brings cancellations and can thus achieve faster rates.

Zhang et al. (2022c) generalize the RVU property to the time-varying setting, but only for zero-sum finite games. To achieve faster rates in our problem, we leverage RVU by exploiting the *smoothness* of the utility functions. [3]

Assumption 3 (Smoothness)**.** The utility gradient $v_t(\mathbf{x})$ is L-Lipschitz continuous: $\|v_t(\mathbf{x}) - v_t(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|$.

We remark that Assumption 3 is mild since similar assumptions naturally exist in matrix games as long as the matrix norm is bounded (Kangarshahi et al., 2018). In the following, we further generalize the RVU property to time-varying multi-player continuous games, as shown below in Lemma 1. And the proof is deferred to Appendix D.1.

Lemma 1. *Under Assumptions 1-3, if the i-th player runs* OOGD *with optimism $m_{t,i} = v_{t-1,i}(\mathbf{x}_{t-1})$ and step size η_i, the regret of the i-th player, $\sum_{t\in[T]}\langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x^\star_{t,i}\rangle$, can be non-asymptotically bounded by*

$$\frac{1 + P_{T,i}}{\eta_i} + \eta_i(1 + V_T) + \eta_i\sum_{j=1}^N S_j - \frac{1}{\eta_i}S_i, \tag{4.1}$$

where $S_j \triangleq \sum_{t=2}^T\|x_{t,j} - x_{t-1,j}\|^2$ is the cumulative switching cost of the j-th player, V_T denotes the gradient variation and $P_{T,i}$ represents the path length of the i-th player.

With specific choices of the step sizes $\{\eta_i\}_{i=1}^N$, the summation of the last two terms in (4.1) can be non-positive, with which the summation of all players' regret can achieve faster rates that only depend on the gradient variation V_T and path length P_T. The distance tracking error can be consequently guaranteed due to DIST-ERR $\leq \sum_{i\in[N]}\sum_{t\in[T]}\langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x^\star_{t,i}\rangle$. Besides, using the similar analysis in Section 3.4, we show that OOGD can also take advantage of small gradient variance W_T. Theorem 4 informally guarantees the tracking error in various problem-dependent quantities, which matches the best known $\mathcal{O}(1)$ bound in time-invariant games. The corresponding formal version and proof can be found in Appendix D.2.

Theorem 4 (informal)**.** *Under Assumptions 1-3, if the i-th player runs* OOGD *with $m_{t,i} = v_{t-1,i}(\mathbf{x}_{t-1})$, the distance tracking error enjoys $\mathcal{O}(\sqrt{(1 + V_T + P_T)(1 + P_T)})$ and $\mathcal{O}(1 + W_T)$ guarantees respectively with different step sizes. Note that the algorithm requires inputs of path length P_T, gradient variation V_T, and gradient variance W_T.*

Remark 3. Although the gradient variance W_T is an upper bound of the variation V_T, as shown in (D.6), our theoretical guarantees of $\mathcal{O}(\sqrt{(1 + V_T + P_T)(1 + P_T)})$ and $\mathcal{O}(1 + W_T)$ are in general incomparable. ¶

Remark 4. Theorem 4 obtains faster rates than Theorem 2 (without smoothness). For example, in time-invariant games, the $\mathcal{O}(1)$ tracking error induced by Theorem 4 improves the earlier $\widetilde{\mathcal{O}}(1)$ by $\log T$ factors. Another example is the S-switch games, where Theorem 4 implies an $\mathcal{O}(1+S)$ bound, faster than $\widetilde{\mathcal{O}}(1 + T^{1/3}S^{2/3})$ of Theorem 2 since $S \leq T$. ¶

4.2. Best-of-Both-Worlds Rates

Theorem 4 obtains faster tracking error guarantees than those without smoothness. However, different bounds require different step size setups, which even depend on the unknown problem-dependent quantities. In this part, we address this issue by designing a *single* algorithm that attains a best-of-both-worlds tracking error guarantee.

[3] In literature, smooth game (Roughgarden, 2009) also has another meaning: a game is (a, b)-smooth if there exists a joint decision $\mathbf{x}^\star$ such that $\sum_{i=1}^N u_i(x_i^\star; x_{-i}) \leq a\sum_{i=1}^N u_i(\mathbf{x}^\star) + b\sum_{i=1}^N u_i(\mathbf{x})$ for any $\mathbf{x}$. In words, any player using their optimal strategy continues to do well irrespective of other players.

Algorithm 2 TV-SMOG (smooth) for the i-th player

Input: step size pool $\mathcal{H}_i^\eta$ (D.18)
Initialize: M instances of OOGD $\mathcal{A}_1, \ldots, \mathcal{A}_M$ with step size $\eta_{i,j} \in \mathcal{H}_i^\eta$
for $t = 1, \ldots, T$ **do**
  Receive $x_{t,i,j}$ from $\mathcal{A}_j$
  Submit $x_{t,i} = \sum_{j=1}^M p_{t,i,j} x_{t,i,j}$
  Receive gradient feedback $v_{t,i}(\mathbf{x}_t)$
  Construct loss $\boldsymbol{\ell}_{t,i}$ and optimism $\boldsymbol{m}_{t,i}$ via (4.3)
  Update to $\boldsymbol{p}_{t+1,i}$ via (4.2) with learning rate $\varepsilon_{t,i}$ (D.7)
  for $j = 1, \ldots, M$ **do**
    $\mathcal{A}_j$ updates to $x_{t+1,i,j}$ with optimism $v_{t-1,i}(\mathbf{x}_{t-1})$
  end
end

To this end, we adopt the two-layer online ensemble framework to facilitate the algorithm with different kinds of adaptivity. Taking the i-th player as an example, the decision is given by $x_{t,i} = \sum_{j\in[M]} p_{t,i,j} x_{t,i,j}$, where $\boldsymbol{p}_{t,i} \triangleq (p_{t,i,1}, \ldots, p_{t,i,M}) \in \Delta_M$ denotes the weight returned by the meta learner and the j-th base learner updates her own decision $x_{t,i,j}$ via OOGD with step size $\eta_{i,j}$. However, one caveat is that the cancellations from Lemma 1 fail. To see this, for the j-th base learner, notice that the last two terms of (4.1) become $\eta_{i,j} \sum_{j=1}^N S_j - S_{i,j}/\eta_{i,j}$, where $S_j \triangleq \sum_{t=2}^T \|x_{t,j} - x_{t-1,j}\|^2$ denotes the cumulative switching cost of the final decision of the i-th player. However, the negative term relies on $S_{i,j} \triangleq \sum_{t=2}^T \|x_{t,i,j} - x_{t-1,i,j}\|^2$, the switching cost of the j-th base learner, and thus cannot be leveraged to cancel the positive switching cost.

To handle this, inspired by the work of Zhao et al. (2021) on the gradient-variation guarantees in non-stationary environments, we observe that the switching cost of the i-th player can be bounded as $\|x_{t,i} - x_{t-1,i}\|^2 \lesssim \|\boldsymbol{p}_{t,i} - \boldsymbol{p}_{t-1,i}\|_1^2 + \sum_{j\in[M]} p_{t,i,j} \|x_{t,i,j} - x_{t-1,i,j}\|^2$ (interested readers can refer to (D.10) for a detailed derivation). The first term $\|\boldsymbol{p}_{t,i} - \boldsymbol{p}_{t-1,i}\|_1^2$ can be canceled by adopting optimistic Hedge (Rakhlin & Sridharan, 2013a) as the meta algorithm (see Lemma 7 for the negative term in the analysis):

$$p_{t+1,i,j} \propto \exp\left(-\varepsilon_{t,i}\left(\sum_{s=1}^{t} \ell_{s,i,j} + m_{t+1,i,j}\right)\right), \quad (4.2)$$

where the learning rate $\varepsilon_{t,i}$, the optimism $\boldsymbol{m}_{t,i} \triangleq (m_{t,i,1}, \ldots, m_{t,i,M})$ and the loss $\boldsymbol{\ell}_{t,i} \triangleq (\ell_{t,i,1}, \ldots, \ell_{t,i,M})$ of the meta learner are to be specified later.

To illustrate the high-level solution to $\sum_{j\in[M]} p_{t,i,j} \|x_{t,i,j} - x_{t-1,i,j}\|^2$, we consider a simpler problem with regret $\sum_{t\in[T]} \langle \boldsymbol{\ell}_t, \boldsymbol{p}_t \rangle - \sum_{t\in[T]} \ell_{t,i^\star}$. If we instead optimize the *biased* loss vector $\boldsymbol{\ell}_t + \boldsymbol{b}_t$ and obtain a bound of R_T, the original regret is at most $R_T - \sum_{t\in[T]} \sum_{i\in[N]} p_{t,i} b_{t,i} + \sum_{t\in[T]} b_{t,i^\star}$, where the negative term of $-\sum_{t\in[T]} \sum_{i\in[N]} p_{t,i} b_{t,i}$ is useful. For our purpose, we set the bias term as $\lambda \|x_{t,i,j} - x_{t-1,i,j}\|^2$ (λ to be specified), the switching cost of the j-th base learner. The loss vector $\boldsymbol{\ell}_{t,i}$ and optimism $\boldsymbol{m}_{t,i}$ are set accordingly as

$$\begin{aligned} \ell_{t,i,j} &\triangleq \langle v_{t,i}(\mathbf{x}_t), x_{t,i,j} \rangle + \lambda \|x_{t,i,j} - x_{t-1,i,j}\|^2, \quad (4.3) \\ m_{t,i,j} &\triangleq \langle v_{t-1,i}(\mathbf{x}_{t-1}), x_{t,i,j} \rangle + \lambda \|x_{t,i,j} - x_{t-1,i,j}\|^2. \end{aligned}$$

Our method, TV-SMOG (smooth), is summarized in Algorithm 2. Theorem 5 gives the theoretical guarantees in terms of multiple problem-dependent quantities. Besides, the individual regret of each player (i.e., $\sum_{t\in[T]} \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star \rangle$), when all players agree to run the same algorithm distributedly, can also achieve faster rates. A formal version and the proof are deferred to Appendix D.3.

Theorem 5 (informal)**.** *Under Assumptions 1-3, Algorithm 2 guarantees a distance tracking error bound of*

$$\mathcal{O}(\min\{\sqrt{(1+V_T+P_T)(1+P_T)}, 1+W_T\}).$$

Simultaneously, the individual regret of each player is bounded by $\mathcal{O}(1)$ in the time-invariant case.

Although our solution draws inspiration from Zhang et al. (2022c), we additionally exploit the structure of strongly monotone games and can therefore guarantee the tracking error, while their method only enjoys implicit measures of the distance to Nash equilibriums.

Finally, Corollary 2 considers common interest strongly monotone games (defined in Section 2.3) and shows that the same rates hold for utility tracking error. Besides, our algorithm can take advantage of the *small loss* $F_T \triangleq \sum_{t=1}^T u_t(\mathbf{x}_t^\star)$, which is at most $\mathcal{O}(T)$ but can be much smaller in benign scenarios. The proof is in Appendix D.4.

Corollary 2. *In common interest strongly monotone games, under the assumptions of Theorem 5, Algorithm 2 enjoys an $\mathcal{O}(\min\{\sqrt{(1+\min\{V_T, F_T\}+P_T)(1+P_T)}, 1+W_T\})$ best-of-three-worlds utility tracking error guarantee.*

The small loss F_T is orthogonal to other problem-dependent measures, because it can be positive in the time-invariant case and also zero in time-varying games. Therefore, our algorithm can exploit the merit of different aspects of the environments and achieve a best-of-three-worlds guarantee.

5. Experiment

This section provides empirical evaluations of our proposed method in time-varying strongly monotone games.

Contenders. We compare our method, *TV-Smog* (Algorithm 1) and *TV-Smog (Smooth)* (Algorithm 2) with three contenders: *(i) OGD* does not consider the changing of games and runs simple online gradient descent; *(ii) Restart* is the method of Duvocelle et al. (2023), which knows the path length P_T in advance; and *(iii) Ader* is our Algorithm 3 proposed in Section 3.2, which is agnostic about the path length P_T, but does not exploit the strong monotonicity.

Game Setups. Our game setups are mostly inspired by Lin et al. (2022), and we modify them to be time-varying to adapt to our problem. We investigate three kinds of time-varying strongly monotone games, including distributed ℓ_2-regularized logistic regression, Cournot competition, and strongly convex-concave zero-sum games. In the following, we introduce the aforementioned setups respectively.

ℓ_2-regularized logistic regression is an important model in machine learning, where the performance is measured by the ℓ_2-regularized logistic loss $u_t(\mathbf{x}) \triangleq \log(1+\exp(-b_t \cdot \mathbf{a}_t^\top \mathbf{x})) + \mu\|\mathbf{x}\|_2^2$, where $\mathbf{a}_t \in \mathbb{R}^N$ and $b_t \in \{-1,+1\}$ are chosen from a dataset $\{\mathbf{a}_t, b_t\}_{t=1}^T$, μ is the parameter of the regularization term to prevent overfitting. In applications, the problem dimension (i.e., N) may be extremely large. Thus distributed computation is usually considered by modeling the problem as a common interest game. See Gopal & Yang (2013) for an example. Here we use four LIBSVM datasets to initialize the logistic loss. We set $\mu = 0.005$, i.e., u_t is 0.01-strongly monotone. The logistic loss is also smooth as long as $\|\mathbf{a}_t\|$ is bounded.

In the Cournot competition model, there are N firms, each supplying the market with a quantity $x_i \in [0, R]$ of some good up to the firm's production capacity. This good is then priced as a decreasing function $P(\mathbf{x})$ of the total supply, as determined by all firms' production. We focus on the standard linear model: $P(\mathbf{x}) \triangleq a - b\sum_{i=1}^N x_i$, where $a = 0.5$ and $b = 1/(NR)$. The reward of the i-th firm is given by $u_{t,i}(x) \triangleq (c_{t,i} - P(\mathbf{x}_t)) \cdot x$, where $c_{t,i} \in [0,1]$ is the time-varying marginal production cost of the i-th firm at the t-th round. This game is b-strongly monotone and $2b$-smooth.

We consider time-varying zero-sum strongly convex-concave games with utility $u_t(x,y) \triangleq \frac{1}{2}\|x\|^2 + x^\top A_t y - \frac{1}{2}\|y\|^2$, which is 1-strongly monotone and 1-smooth.

Results. We report average results with standard deviations of 5 independent runs. Only the randomness of the initial decisions is preserved. All hyper-parameters are set to be theoretically optimal. Figure 1 plots the tracking error of all methods. Smaller tracking error indicates better performance. The results show the supremacy of our *TV-Smog* and *TV-Smog (Smooth)*, supporting our theoretical results.

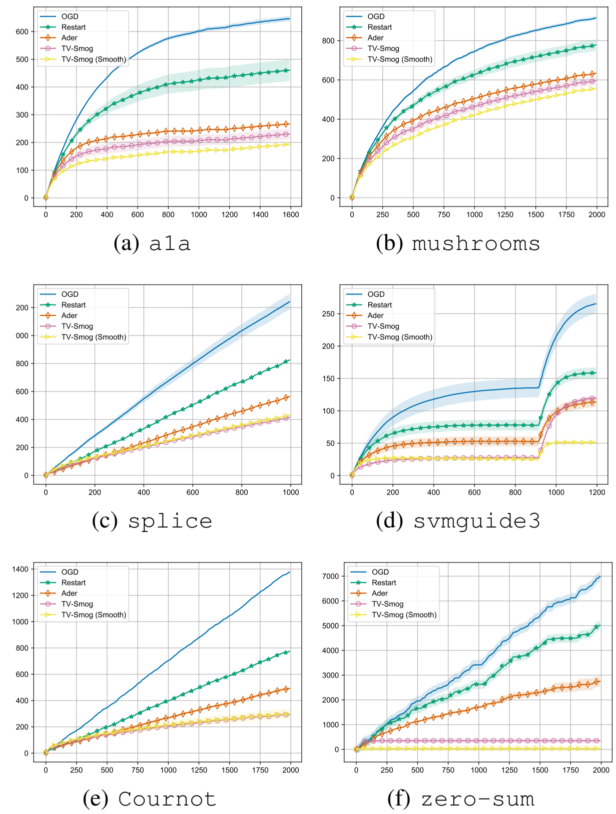

Figure 1: Tracking error of all methods in four time-varying distributed ℓ_2-regularized logistic regression tasks, named by `a1a`, `mushrooms`, `splice` and `svmguide3`, a time-varying Cournot competition game named `Cournot` and a time-varying strongly convex-concave zero-sum game named `zero-sum`.

6. Conclusion and Future Directions

This work presents an initial solution for the time-varying strongly monotone games. We develop novel decentralized online algorithms and establish a series of fast tracking error rates, in both non-smooth and smooth cases. Our results match the best-known time-invariant bounds and enjoy many favorable properties. We also investigate the more specialized common interest strongly monotone games, and show that our bounds also hold this setup regarding the more fundamental utility tracking error.

One major future direction is to explore the problem's lower bound. Another interesting question is to leverage the strong convexity induced by strong monotonicity to further improve current results in the smooth case.

Acknowledgements

This research was supported by National Key R&D Program of China (2022ZD0114800), NSFC (61921006, 62206125), and Fundamental Research Funds for the Central Universities (2023300246). The authors thank the ICML area chair for pointing out that the utility tracking error may only apply to common interest games rather than general games, which has been corrected in the camera-ready version. We are also grateful to Mengxiao Zhang for the helpful discussions.

References

Abernethy, J. D., Bartlett, P. L., Rakhlin, A., and Tewari, A. Optimal stragies and minimax lower bounds for online convex games. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, pp. 415–424, 2008a.

Abernethy, J. D., Hazan, E., and Rakhlin, A. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of the 21st Annual Conference Computational Learning Theory (COLT)*, pp. 263–274, 2008b.

Baby, D. and Wang, Y. Optimal dynamic regret in exp-concave online learning. In *Proceedings of the 34th Annual Conference Computational Learning Theory (COLT)*, pp. 359–409, 2021.

Baby, D. and Wang, Y. Optimal dynamic regret in proper online learning with strongly convex losses and beyond. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 1805–1845, 2022.

Beck, A. *First-order Methods in Optimization*. SIAM, 2017.

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

Bravo, M., Leslie, D. S., and Mertikopoulos, P. Bandit learning in concave N-person games. In *Advances in Neural Information Processing Systems 31 (NeurIPS)*, pp. 5666–5676, 2018.

Bregman, L. and Fokin, I. Methods of determining equilibrium situations in zero-sum polymatrix games. *Optimizatsia*, 40(57):70–82, 1987.

Cai, Y., Candogan, O., Daskalakis, C., and Papadimitriou, C. Zero-sum polymatrix games: A generalization of minmax. *Mathematics of Operations Research*, 41(2): 648–655, 2016.

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

Cutkosky, A. Parameter-free, dynamic, and strongly-adaptive online learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pp. 2250–2259, 2020.

Daniely, A., Gonen, A., and Shalev-Shwartz, S. Strongly adaptive online learning. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pp. 1405–1411, 2015.

Daskalakis, C. and Papadimitriou, C. H. On a network generalization of the minmax theorem. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming (ICALP)*, pp. 423–434, 2009.

Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

Daskalakis, C., Deckelbaum, A., and Kim, A. Near-optimal no-regret algorithms for zero-sum games. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 235–254, 2011.

d'Oro, S., Mertikopoulos, P., Moustakas, A. L., and Palazzo, S. Interference-based pricing for opportunistic multicarrier cognitive radio systems. *IEEE Transactions on Wireless Communications*, 14(12):6536–6549, 2015.

Drusvyatskiy, D., Fazel, M., and Ratliff, L. J. Improved rates for derivative free gradient play in strongly monotone games. In *Proceedings of the 61st IEEE Conference on Decision and Control (CDC)*, pp. 3403–3408, 2022.

Duvocelle, B., Mertikopoulos, P., Staudigl, M., and Vermeulen, D. Multi-agent online learning in time-varying games. *Mathematics of Operations Research*, 48(2):914–941, 2023.

Even-Dar, E., Mansour, Y., and Nadav, U. On the convergence of regret minimization dynamics in concave games. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pp. 523–532, 2009.

Facchinei, F. and Kanzow, C. Generalized Nash equilibrium problems. *Annals of Operations Research*, 175(1):177–211, 2010.

Facchinei, F. and Pang, J.-S. *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003.

Flaxman, A., Kalai, A. T., and McMahan, H. B. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 385–394, 2005.

Freund, Y. and Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

Freund, Y. and Schapire, R. E. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103, 1999.

Golowich, N., Pattathil, S., and Daskalakis, C. Tight last-iterate convergence rates for no-regret learning in multi-player games. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, 2020.

Gopal, S. and Yang, Y. Distributed training of large-scale logistic models. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, pp. 289–297, 2013.

Hazan, E. and Seshadhri, C. Adaptive algorithms for online decision problems. *Electronic Colloquium on Computational Complexity*, TR07-088, 2007.

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

Hsieh, K., Phanishayee, A., Mutlu, O., and Gibbons, P. B. The non-iid data quagmire of decentralized machine learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pp. 4387–4398, 2020.

Kangarshahi, E. A., Hsieh, Y., Sahin, M. F., and Cevher, V. Let's be honest: An optimal no-regret framework for zero-sum games. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pp. 2493–2501, 2018.

Kapralov, M. and Panigrahy, R. Prediction strategies without loss. In *Advances in Neural Information Processing Systems 24 (NIPS)*, pp. 828–836, 2011.

Lin, T., Zhou, Z., Mertikopoulos, P., and Jordan, M. I. Finite-time last-iterate convergence for multi-agent learning in games. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pp. 6161–6171, 2020.

Lin, T., Zhou, Z., Ba, W., and Zhang, J. Doubly optimal no-regret online learning in strongly monotone games with bandit feedback. *ArXiv preprint*, arXiv:2112.02856, 2022.

Loizou, N., Berard, H., Gidel, G., Mitliagkas, I., and Lacoste-Julien, S. Stochastic gradient descent-ascent and consensus optimization for smooth games: Convergence analysis under expected co-coercivity. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, pp. 19095–19108, 2021.

Lu, Z. and Hazan, E. On the computational efficiency of adaptive and dynamic regret minimization. *ArXiv preprint*, arXiv:2207.00646, 2022.

Mertikopoulos, P. and Moustakas, A. L. Learning in an uncertain world: Mimo covariance matrix optimization with imperfect feedback. *IEEE Transactions on Signal Processing*, 64(1):5–18, 2015.

Monderer, D. and Shapley, L. S. Potential games. *Games and Economic Behavior*, 14(1):124–143, 1996.

Nash Jr, J. F. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1): 48–49, 1950.

Nemirovski, A., Onn, S., and Rothblum, U. G. Accuracy certificates for computational problems with convex structure. *Mathematics of Operations Research*, pp. 52–78, 2010.

Pogodin, R. and Lattimore, T. On first-order bounds, variance and gap-dependent bounds for adversarial bandits. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 894–904, 2019.

Rakhlin, A. and Sridharan, K. Online learning with predictable sequences. In *Proceedings of the 26th Annual Conference Computational Learning Theory (COLT)*, pp. 993–1019, 2013a.

Rakhlin, A. and Sridharan, K. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems 26 (NIPS)*, pp. 3066–3074, 2013b.

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

Roughgarden, T. Intrinsic robustness of the price of anarchy. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pp. 513–522, 2009.

Sentenac, F., Boursier, E., and Perchet, V. Decentralized learning in online queuing systems. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, pp. 18501–18512, 2021.

Srebro, N., Sridharan, K., and Tewari, A. Smoothness, low noise and fast rates. In *Advances in Neural Information Processing Systems 23 (NIPS)*, pp. 2199–2207, 2010.

Syrgkanis, V., Agarwal, A., Luo, H., Luo, H., and Schapire, R. E. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems 28 (NIPS)*, pp. 2989–2997, 2015.

Tatarenko, T. and Kamgarpour, M. Learning Nash equilibria in monotone games. In *Proceedings of the 58th IEEE Conference on Decision and Control (CDC)*, pp. 3104–3109, 2019a.

Tatarenko, T. and Kamgarpour, M. Learning generalized Nash equilibria in a class of convex games. *IEEE Transactions on Automatic Control*, 64(4):1426–1439, 2019b.

Tatarenko, T. and Kamgarpour, M. On the rate of convergence of payoff-based algorithms to Nash equilibrium in strongly monotone games. *ArXiv preprint*, arXiv:2202.11147, 2022.

Tseng, P. On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, 60(1-2):237–252, 1995.

van Erven, T. and Koolen, W. M. Metagrad: Multiple learning rates in online learning. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pp. 3666–3674, 2016.

Wang, G., Lu, S., and Zhang, L. Adaptivity and optimality: A universal algorithm for online convex optimization. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 659–668, 2019.

Zhang, L., Lu, S., and Zhou, Z.-H. Adaptive online learning in dynamic environments. In *Advances in Neural Information Processing Systems 31 (NeurIPS)*, pp. 1330–1340, 2018.

Zhang, L., Jiang, W., Yi, J., and Yang, T. Smoothed online convex optimization based on discounted-normal-predictor. In *Advances in Neural Information Processing Systems 35 (NeurIPS)*, pp. 4928–4942, 2022a.

Zhang, L., Wang, G., Yi, J., and Yang, T. A simple yet universal strategy for online convex optimization. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pp. 26605–26623, 2022b.

Zhang, M., Zhao, P., Luo, H., and Zhou, Z.-H. No-regret learning in time-varying zero-sum games. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pp. 26772–26808, 2022c.

Zhang, Z., Cutkosky, A., and Paschalidis, I. C. Adversarial tracking control via strongly adaptive online learning with memory. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 8458–8492, 2022d.

Zhao, P. *Online Ensemble Theories and Methods for Robust Online Learning*. PhD thesis, Nanjing University, Nanjing, China, 2021. Advisor: Zhi-Hua Zhou.

Zhao, P., Zhang, L., Jiang, Y., and Zhou, Z.-H. A simple approach for non-stationary linear bandits. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 746–755, 2020a.

Zhao, P., Zhang, Y.-J., Zhang, L., and Zhou, Z.-H. Dynamic regret of convex and smooth functions. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, pp. 12510–12520, 2020b.

Zhao, P., Zhang, Y.-J., Zhang, L., and Zhou, Z.-H. Adaptivity and non-stationarity: Problem-dependent dynamic regret for online convex optimization. *ArXiv preprint*, arXiv:2112.14368, 2021.

Zhao, P., Xie, Y.-F., Zhang, L., and Zhou, Z.-H. Efficient methods for non-stationary online learning. In *Advances in Neural Information Processing Systems 35 (NeurIPS)*, pp. 11573–11585, 2022.

Zhou, Z.-H. *Ensemble Methods: Foundations and Algorithms*. Chapman & Hall/CRC Press, 2012.

Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, pp. 928–936, 2003.

A. Related Work

In this section, we briefly review related works of non-stationary (Appendix A.1) and universal (Appendix A.2) online learning, and monotone games (Appendix A.3), and compare our work with existing time-varying results (Appendix A.4).

A.1. Non-Stationary Online Convex Optimization

Online convex optimization (OCO), a versatile model which stems from the seminal work of Zinkevich (2003), depicts the learning of a player in adversarial environments, aiming to minimize the game theoretical performance measure called *regret* (Cesa-Bianchi & Lugosi, 2006):

$$\textsc{Reg}_T \triangleq \sum_{t=1}^{T} f_t(\mathbf{w}_t) - \min_{\mathbf{w}\in\mathcal{W}} \sum_{t=1}^{T} f_t(\mathbf{w}).$$

Zinkevich (2003) proposed the well-known OGD and proved an $\mathcal{O}(\sqrt{T})$ regret, which is minimax optimal (Abernethy et al., 2008a). In non-stationary environments, a more appropriate performance measure is *dynamic regret* (Zinkevich, 2003):

$$\text{D-}\textsc{Reg}(\{\mathbf{u}_t\}_{t=1}^T) \triangleq \sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t)$$

which aims to compete the performance of the online learner with changing comparators. For convex functions, the author proved that OGD enjoys a dynamic regret of $\mathcal{O}(\sqrt{T}(1+P_T))$, which is sub-optimal. Zhang et al. (2018) showed that the minimax lower bound is $\Omega(\sqrt{T(1+P_T)})$ and closed the gap by proposing an online ensemble algorithm with an $\mathcal{O}(\sqrt{T(1+P_T)})$ regret. When exploiting the smoothness, Zhao et al. (2020b) developed a novel online ensemble algorithm called Sword that enjoys a best-of-both-worlds problem-dependent dynamic regret bound of order $\mathcal{O}(\sqrt{(1+P_T+\min\{V_T,F_T\})(1+P_T)})$, where $V_T \triangleq \sum_{t=2}^{T} \sup_{\mathbf{w}\in\mathcal{W}} \|\nabla f_t(\mathbf{w}) - \nabla f_{t-1}(\mathbf{w})\|^2$ measures the variation of the gradients and $F_T \triangleq \sum_{t=1}^{T} f_t(\mathbf{u}_t)$ is the cumulative loss of comparators. In the subsequent extended version (Zhao et al., 2021), the authors further proposed the Sword++ algorithm that achieves the same dynamic regret guarantee while improving the per-round gradient complexity from $\mathcal{O}(\log T)$ to 1 via the collaborative online ensemble framework. For exp-concave functions, Baby & Wang (2021) proposed an improper learning method (the decisions can be outside the feasible domain) that enjoys the optimal $\widetilde{\mathcal{O}}(1+T^{1/3}P_T^{2/3})$ dynamic regret guarantee. Later, Baby & Wang (2022) showed that improper learning is not necessary for strongly convex functions and designed a proper learning method with the optimal dynamic regret bound in this case.

Another commonly used performance measure in non-stationary OCO is the *adaptive regret* (Hazan & Seshadhri, 2007; Daniely et al., 2015), which measures the regret inside an arbitrary interval $\mathcal{I} \subseteq [T]$:

$$\text{A-}\textsc{Reg}(\mathcal{I}) \triangleq \sum_{t\in\mathcal{I}} f_t(\mathbf{w}_t) - \min_{\mathbf{w}\in\mathcal{W}} \sum_{t\in\mathcal{I}} f_t(\mathbf{w}).$$

Hazan & Seshadhri (2007) first proposed the weakly adaptive regret, $\max_{\mathcal{I}\subseteq[T]} \text{A-}\textsc{Reg}(\mathcal{I})$, which measures the maximum regret over all possible intervals. They designed a meta algorithm called Follow the Leading History (FLH) and proved an $\widetilde{\mathcal{O}}(\sqrt{T})$ adaptive regret with OGD as base learners. Later, Daniely et al. (2015) optimized the adaptive regret $\text{A-}\textsc{Reg}(\mathcal{I})$ over arbitrary interval $\mathcal{I}$, which is called strongly adaptive regret, and proved an $\widetilde{\mathcal{O}}(\sqrt{|\mathcal{I}|})$ regret via a new meta algorithm called SAOL with OGD as base learners. For exp-concave and strongly convex functions, FLH-OGD (Hazan & Seshadhri, 2007) can achieve the optimal regret bound of $\mathcal{O}(\log T)$.

The aforementioned works all choose the weighted combination mechanism (i.e., $\mathbf{w}_t = \sum_i p_{t,i}\mathbf{w}_{t,i}$) in non-stationary online learning problems. We note that there are other methodologies for combining the multiple base learners, for example, using techniques from parameter-free online learning (Cutkosky, 2020; Zhang et al., 2022d) and techniques based on the discounted normal predictor (Kapralov & Panigrahy, 2011; Zhang et al., 2022a).

Finally, we mention that the computational efficiency of the non-stationary online learning, including dynamic and adaptive regret minimization, is recently considered (Zhao et al., 2022; Lu & Hazan, 2022). Typical algorithms for non-stationary online learning often utilize the online ensemble framework. The overall method requires to run multiple base learners (typically $\mathcal{O}(\log T)$ base learners) and a meta learner simultaneously, which may introduce some computational overheads compared to the static regret minimization. Zhao et al. (2022) investigated the projection complexity in non-stationary environments, an important and potentially time-consuming operation for common online learning algorithms such as OGD, and reduced the number of projection operations needed in each round to 1.

A.2. Universal Online Convex Optimization

When the loss functions are convex, using OGD with a fixed step size $\eta \approx 1/\sqrt{T}$ gives an optimal regret bound of $\mathcal{O}(\sqrt{T})$ (Zinkevich, 2003). When the loss functions are α-exp-concave, online Newton step enjoys a regret bound of $\mathcal{O}(d \log T)$ (Hazan et al., 2007). And when the loss functions are σ-strongly convex, using the same OGD algorithm but with a time-varying step size $\eta_t = 1/(\sigma t)$ gives an optimal $\mathcal{O}(\log T)$ guarantee (Hazan et al., 2007). Although the theories of OCO are rich, its application requires heavy domain knowledge: *(i)* the learner must know the type of functions in advance in order to select an appropriate algorithm; and *(ii)* for exp-concave and strongly-convex functions, the learner needs the strong convexity and the exp-concavity coefficients (i.e., values of α and σ) as prior knowledge. *Universal online learning* aims to remove the above barriers. A milestone is the MetaGrad proposed by van Erven & Koolen (2016). Briefly, MetaGrad is a two-layer algorithm, with multiple base learners running on some complicated surrogate losses to guess the curvature coefficients and a meta learner tracking the best one on the fly. Later, many improvements and extensions are proposed, that further broaden the application scope of universal methods. Specifically, Wang et al. (2019) improved the results of MetaGrad for strongly convex functions. Zhang et al. (2022b) proposed a simpler universal method with the same theoretical guarantees as previous works, but without the need to design handcrafted surrogate losses.

A.3. Monotone Games

Monotone game is a general class of games that satisfies the diagonal strict convexity condition of Rosen (1965). Many common and well-studied classes of games, such as zero-sum poly-matrix games (Bregman & Fokin, 1987; Daskalakis & Papadimitriou, 2009; Cai et al., 2016) and its generalization zero-sum socially-concave games (Even-Dar et al., 2009) are monotone. Solving the Nash equilibriums of a monotone game is equivalent to solving a variational inequality, where there is a vast literature, and we refer the readers to the work of Facchinei & Pang (2003) for further references. Besides, in the following, we only talk about the convergence of the last iterate, the decision of the last round of the game.

In generally monotone games, only asymptotic convergence can be guaranteed. With additional smoothness assumption, Golowich et al. (2020) obtained nearly optimal convergence rate in terms of the total function gap.

There are many works devoted to studying certain subclasses of monotone games. Specifically, Lin et al. (2020) considered co-coercive games, i.e., $\langle v(\mathbf{x}) - v(\mathbf{y}), \mathbf{x} - \mathbf{y}\rangle \geq \mu \|v(\mathbf{x}) - v(\mathbf{y})\|^2$, in the unconstrained setting (i.e., $v_t(\mathbf{x}^\star) = \mathbf{0}$) and chose $\|v_t(\mathbf{x}_t)\|^2$ as the performance measure. Although their setting is more general than ours, the two works cannot be compared directly with since their performance measure is also easier as $\|v_t(\mathbf{x}_t)\|^2$ is observable while ours $\|\mathbf{x}_t - \mathbf{x}^\star\|^2$ is not. Besides, Loizou et al. (2021) considered a more general setting with quasi-strong monotonicity and expected co-coercivity, and gave a linear tracking error convergence rate $\mathcal{O}(\rho^T)$ for some $\rho \in (0, 1)$ (see Corollary 4.2 therein). We refer readers to Loizou et al. (2021) for detailed relationship between monotonicity, co-coercivity and strong monotonicity.

For strongly monotone games, non-asymptotic convergence rates can be established. In the full information setting, an $\mathcal{O}(T^{-1})$ last-iterate convergence of the tracking error can be obtained (Bravo et al., 2018). With smoothness (i.e., Lipschitz continuous gradients), this rate can be improved to $\mathcal{O}(\rho^T)$ for some $\rho \in (0, 1)$ (Tseng, 1995; Facchinei & Pang, 2003). In the bandit feedback setting, existing works cannot obtain faster rates than the single-player results. Specifically, Bravo et al. (2018) extended the FKM algorithm (Flaxman et al., 2005) to the multi-player setting, and proposed an algorithm with an $\mathcal{O}(T^{-1/3})$ distance tracking error. With smoothness, Lin et al. (2022) leveraged the self-concordant barrier (Abernethy et al., 2008b), and obtained an improved result of $\mathcal{O}(T^{-1/2})$, matching the optimal result in the single-player setting. Drusvyatskiy et al. (2022) achieved the same guarantee with Lin et al. (2022), while their result depends on an additional assumption that the Jacobian of each player's gradient is Lipschitz continuous. There is also a different line of research (Tatarenko & Kamgarpour, 2019a;b; 2022) that can also obtain the optimal result using Tikhonov (ridge) regularization.

A.4. Time-Varying Games

The study of online time-varying games has just started in recent years and the corresponding related works are limited. In the following, we directly compare with two works that are mostly related to ours.

Comparison with (Zhang et al., 2022c). As for problem setup, while both their work and ours investigate time-varying games, theirs considers the two-player zero-sum games and ours considers the multi-player continuous strongly monotone games. As for techniques, in smooth games, the bias term injection technique is partially inspired by theirs, which originally stems from the work of Zhao et al. (2021). However, as mentioned in Section 4.2, the structure of $\|x_{t,i} - x_{t-1,i}\|^2$ naturally rises in our problem and the correction term injection is a well-studied solution to handle this term. We also note that the results of Zhang et al. (2022c) can be directly extended to the multi-player setting, but without tracking error guarantees. The key observation (Proposition 2) and the techniques used in the non-smooth games are different from theirs.

Comparison with (Duvocelle et al., 2023). Both works investigate time-varying strongly monotone games. Theirs obtained a dynamic regret of $\mathcal{O}(\sqrt{T} + T^{2/3}P_T^{1/3})$, where $P_T \triangleq \sum_{t=2}^T \|\mathbf{x}_t^\star - \mathbf{x}_{t-1}^\star\|$ denotes the variation of the Nash equilibriums. The authors argued that this bound is tight according to an $\Omega(T^{2/3}(V_T^f)^{1/3})$ lower bound established by Besbes et al. (2015). We point out this argument is flawed since the V_T^f quantity is in fact the *functional variation*: $V_T^f \triangleq \sum_{t=2}^T \sup_{\mathbf{x}} |f_t(\mathbf{x}) - f_{t-1}(\mathbf{x})|$. The dynamic regret bounds with respect to P_T and V_T^f are in general incomparable. Based on the fact that $\mathcal{O}(\sqrt{T} + T^{2/3}P_T^{1/3})$ is far from optimal, we obtain improved results in Section 3. Also, Duvocelle et al. (2023) mentioned that they cannot incorporate strongly convexity into this problem. In this work, we show that this is actually possible, as illustrated in Proposition 2.

B. Proof of Proposition 1

Proof. First we prove that strong monotonicity directly implies strong convexity. Specifically, since the game of the t-th round is μ-strongly monotone (see Assumption 2), strong monotonicity (see Definition 1) tells that for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, $\mu\|\mathbf{x} - \mathbf{y}\|^2 \le \langle v_t(\mathbf{x}) - v_t(\mathbf{y}), \mathbf{x} - \mathbf{y}\rangle$, where $v_t(\cdot) = \nabla u_t(\cdot)$. The following result shows that the definition of strong monotonicity is in fact an equivalent condition of strong convexity in common interest strongly monotone games, which means that the utility function u_t is μ-strongly convex.

Lemma 2 (Theorem 5.24 of Beck (2017))**.** *Let f be a proper closed and convex function. Then for a given $\sigma > 0$, the following three claims are equivalent: (i) f is σ-strongly convex; (ii) for any $\mathbf{x} \in \mathrm{dom}(\partial f), \mathbf{y} \in \mathrm{dom}(f)$ and $\mathbf{g} \in \partial f(\mathbf{x})$, $f(\mathbf{y}) \ge f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x}\rangle + \frac{\sigma}{2}\|\mathbf{y} - \mathbf{x}\|^2$; and (iii) for any $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(\partial f)$ and $\mathbf{g}_\mathbf{x} \in \partial f(\mathbf{x}), \mathbf{g}_\mathbf{y} \in \partial f(\mathbf{y})$, $\langle \mathbf{g}_\mathbf{x} - \mathbf{g}_\mathbf{y}, \mathbf{x} - \mathbf{y}\rangle \ge \sigma\|\mathbf{x} - \mathbf{y}\|^2$.*

Consequently, it holds that

$$\frac{\mu}{2} \cdot \textsc{Dist-Err} = \frac{\mu}{2}\sum_{t=1}^T \|\mathbf{x}_t - \mathbf{x}_t^\star\|^2 \le \sum_{t=1}^T \langle v_t(\mathbf{x}_t^\star), \mathbf{x}_t^\star - \mathbf{x}_t\rangle + \sum_{t=1}^T u_t(\mathbf{x}_t) - \sum_{t=1}^T u_t(\mathbf{x}_t^\star) \le \textsc{Util-Err},$$

where the first inequality is due to the second property in Lemma 2 and the last inequality is by the definition of the Nash equilibrium $\mathbf{x}_t^\star$, i.e., $\langle v_t(\mathbf{x}_t^\star), \mathbf{x}_t^\star - \mathbf{x}\rangle \le 0$ for any $\mathbf{x} \in \mathcal{X}$, which completes the proof. ∎

C. Analysis for Section 3

In this section, we provide the detailed analysis of Section 3, including the algorithm and proof of Theorem 1 in Appendix C.1, the proof of Proposition 2 in Appendix C.2, the algorithm and proof of Theorem 2 in Appendix C.3 and the proofs of Theorem 3 and Corollary 1 in Appendix C.4 and Appendix C.5. Finally, we list some useful lemmas in Appendix C.6.

C.1. Algorithm and Proof of Theorem 1

To begin with, we give a different upper bound for the distance tracking error. Specifically, it holds that

$$\mu\textsc{Dist-Err} \le \sum_{t=1}^T \langle v_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^\star\rangle = \sum_{t=1}^T\sum_{i=1}^N \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star\rangle, \tag{C.1}$$

where the first step uses the definition of strong monotonicity and the property of Nash equilibrium: $\langle v_t(\mathbf{x}_t^\star), \mathbf{x}_t^\star - \mathbf{x}\rangle \le 0$ for any $\mathbf{x} \in \mathcal{X}$. In the following, we deploy Ader (Zhang et al., 2018), restated in Algorithm 3 in our notations, to optimize $\sum_{t\in[T]} \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star\rangle$ for each player. Specifically, in the t-th round, the algorithm submits a weighted combination of base learners' decisions as $x_{t,i} = \sum_{j\in[M]} p_{t,i,j} x_{t,i,j}$ and simultaneously receives a gradient feedback $v_{t,i}(\mathbf{x}_t)$. Then

Algorithm 3 Deploying ADER (Zhang et al., 2018) for the i-th player

Input: learning rate of the meta learner ε_i

Initialize: learning rate of the meta learner $\varepsilon_i = 1/\sqrt{T}$, decisions of all base learners $x_{1,i,j} \in \mathcal{X}_i$, weights of the meta learner $p_{1,i,j} \propto 1/(j^2+j)$ for all $j \in [M]$, step size pool $\mathcal{H}_i^\eta$

$$\mathcal{H}_i^\eta \triangleq \left\{ \eta_{i,j} \;\middle|\; \eta_{i,j} = \frac{2^{j-1}D}{G}\sqrt{\frac{7}{2T}},\ j \in [M] \right\}$$

for $t = 1, \dots, T$ **do**

 Receive $x_{t,i,j}$ from the j-th base learner

 Submit $x_{t,i} = \sum_{j=1}^M p_{t,i,j} x_{t,i,j}$ and receive a gradient feedback $v_{t,i}(\mathbf{x}_t)$

 Meta update: Update the weights via $p_{t+1,i,j} \propto p_{t,i,j}\exp(-\varepsilon_i \langle v_{t,i}(\mathbf{x}_t), x_{t,i,j}\rangle)$

 for $j = 1, \dots, M$ **do**

  Base update: Update via $x_{t+1,i,j} = \Pi_{\mathcal{X}_i}[x_{t,i,j} - \eta_{i,j} v_{t,i}(\mathbf{x}_t)]$

 end

end

each base learner updates her own decision to $x_{t+1,i,j}$ using her own step size $\eta_{i,j}$ via OGD. Finally the meta learner updates her weights to $p_{t+1,i,j}$ using Hedge with learning rate ε_i and a linearized loss $\langle v_{t,i}(\mathbf{x}_t), x_{t,i,j}\rangle$. In the following, we provide the proof of Theorem 1.

Proof. The proof is direct by bounding the summation of all players' regret bounds and leveraging the theoretical guarantee of ADER (deferred to Lemma 3 in Appendix C.6),

$$\sum_{t=1}^T \sum_{i=1}^N \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star\rangle \lesssim \sum_{i=1}^N \sqrt{T(1+P_{T,i})} \le N\sqrt{T(1+P_T)}, \qquad \text{(by Lemma 3)}$$

where $a \lesssim b$ means $a \le \widetilde{\mathcal{O}}(b)$, $P_{T,i} \triangleq \sum_{t=2}^T \|x_{t,i}^\star - x_{t-1,i}^\star\|$ denotes the path length of the Nash equilibriums of the i-th player and the last step is true because

$$\sum_{i=1}^N P_{T,i} = \sum_{t=2}^T \sum_{i=1}^N \|x_{t,i}^\star - x_{t-1,i}^\star\| \le \sum_{t=2}^T \sqrt{N \sum_{i=1}^N \|x_{t,i}^\star - x_{t-1,i}^\star\|^2} = \sum_{t=2}^T \sqrt{N\|\mathbf{x}_t^\star - \mathbf{x}_{t-1}^\star\|^2} = \sqrt{N} P_T.$$

Leveraging (C.1) finishes the proof. ∎

C.2. Proof of Proposition 2

Proof. Through the definition of strong monotonicity (Definition 1) and moving $\frac{\mu}{2}\sum_{t=1}^T \|\mathbf{x}_t - \mathbf{x}_t^\star\|^2$ to the right-hand side,

$$\frac{\mu}{2}\sum_{t=1}^T \|\mathbf{x}_t - \mathbf{x}_t^\star\|^2 \le \sum_{t=1}^T \langle v_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^\star\rangle - \frac{\mu}{2}\sum_{t=1}^T \|\mathbf{x}_t - \mathbf{x}_t^\star\|^2.$$

Multiplying both sides by 2, it is easy to find that

$$\begin{aligned} \mu \sum_{t=1}^T \|\mathbf{x}_t - \mathbf{x}_t^\star\|^2 &\le 2\left(\sum_{t=1}^T \langle v_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^\star\rangle - \frac{\mu}{2}\sum_{t=1}^T \|\mathbf{x}_t - \mathbf{x}_t^\star\|^2\right) \\ &= 2\sum_{i=1}^N \left(\sum_{t=1}^T \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star\rangle - \frac{\mu}{2}\sum_{t=1}^T \|x_{t,i} - x_{t,i}^\star\|^2\right) \\ &= 2\sum_{i=1}^N \left(\sum_{t=1}^T \ell_{t,i}(x_{t,i}) - \sum_{t=1}^T \ell_{t,i}(x_{t,i}^\star)\right), \end{aligned}$$

Algorithm 4 FLH-OGD (Hazan & Seshadhri, 2007)

Input: feasible domain $\mathcal{W}$, learning rate of the meta learner ε, loss functions $\{h_t(\cdot)\}_{t=1}^T$
Initialize: T instances $\mathcal{A}_1, \ldots, \mathcal{A}_T$ of OGD
for $t = 1, \ldots, T$ **do**
  Receive $\mathbf{w}_{t,j}$ from $\mathcal{A}_j$ for all $j \in [t]$
  Submit decision $\mathbf{w}_t = \sum_{j=1}^t p_{t,j}\mathbf{w}_{t,j}$ and receive loss function h_t
  Meta update: Set $\widehat{p}_{t+1,t+1} = 0$ and update $\widehat{p}_{t+1,j} = \frac{p_{t,j}\exp(-\varepsilon h_t(\mathbf{w}_{t,j}))}{\sum_{j=1}^t p_{t,j}\exp(-\varepsilon h_t(\mathbf{w}_{t,j}))}$ for $j \in [t]$
  Meta update: Obtain $p_{t+1,t+1} = 1/(t+1)$ and $p_{t+1,j} = (1-(t+1)^{-1})\widehat{p}_{t+1,j}$ for $j \in [t]$
  for $j = 1, \ldots, t+1$ **do**
    Base update: Update $\mathbf{w}_{t,j}$ to $\mathbf{w}_{t+1,j}$ via $\mathcal{A}_j$ with loss function h_t
  end
end

where although the second step only holds for ℓ_2-norm, it can be extended to arbitrary norm up to constant factors. The last step holds by the definition of the strongly convex loss function (3.2), restated as follows: $\ell_{t,i}(x) \triangleq \langle v_{t,i}(\mathbf{x}_t), x\rangle + \frac{\mu}{2}\|x - x_{t,i}\|^2$, which finishes the proof. ■

C.3. Algorithm and Proof of Theorem 2

Before giving the proof of Theorem 2, we explain the complete algorithms of this part. Briefly, the overall algorithms are concluded in Algorithm 1, Algorithm 4, Algorithm 5 and Algorithm 6. Among them, Algorithm 1 is the top algorithm, which aims to deal with the unknown strong monotonicity parameter μ and takes Algorithm 6 as base learners. Algorithm 6 aims to handle general convex feasible domains and loss functions with known strong convexity parameter and obtains its decisions from Algorithm 5. Algorithm 5 works on box domains and runs Algorithm 4 on each dimension. Algorithm 4 runs FLH-OGD (Hazan & Seshadhri, 2007), an adaptive regret minimization algorithm which runs FLH as the meta learner and multiple OGDs as the base learners. Algorithm 1, Algorithm 5 and Algorithm 6 are illustrated in the language of online multi-player games while Algorithm 4 is summarized in the general online convex optimization notations.

In the following, we give the proof of Theorem 2.

Proof. To begin with, we decompose the regret on the surrogate loss function $s_{t,i,j}$, with respect to its j-th base learner:

$$\sum_{t=1}^T s_{t,i,j}(x_{t,i}) - \sum_{t=1}^T s_{t,i,j}(x^\star_{t,i}) \le \underbrace{\sum_{t=1}^T s_{t,i,j}(x_{t,i}) - \sum_{t=1}^T s_{t,i,j}(x_{t,i,j})}_{\textsc{Meta-Reg}} + \underbrace{\sum_{t=1}^T s_{t,i,j}(x_{t,i,j}) - \sum_{t=1}^T s_{t,i,j}(x^\star_{t,i})}_{\textsc{Base-Reg}}.$$

As for the meta regret, we can reuse the result of Wang et al. (2019), restated in Lemma 4, since we use the same meta algorithm therein. It remains to bound the base regret. Denoting by λ_j, G_j the strong convexity parameter and the gradient upper bound of the surrogate loss $s_{t,i,j}$, its gradient upper bound satisfies $\|\nabla s_{t,i,j}(\cdot)\| = \|a_j v_{t,i}(\mathbf{x}_t) + 2a_j^2 G^2(\cdot - x_{t,i})\| \le a_j G + 2a_j^2 G^2 D$ and its strong convexity parameter holds for $\|\nabla^2 s_{t,i,j}(\cdot)\| \le 2a_j^2 G^2$. For simplicity, we define $G_j \triangleq a_j G + 2a_j^2 G^2 D$ and $\lambda_j = 2a_j^2 G^2$. Following Lemma 5 (about the base regret with known strong convexity coefficient), denoting by $\mathrm{TV}_{T,i} \triangleq \sum_{t=2}^T \|x^\star_{t,i} - x^\star_{t-1,i}\|_1$ the total variation of the comparator sequence $\{x^\star_{t,i}\}_{t=1}^T$, it holds that

$$\textsc{Base-Reg} = \sum_{t=1}^T s_{t,i,j}(x_{t,i,j}) - \sum_{t=1}^T s_{t,i,j}(x^\star_{t,i}) \lesssim \frac{G_j^2}{\lambda_j}(1 + T^{1/3}\mathrm{TV}_{T,i}^{2/3}) \lesssim 1 + T^{1/3}P_{T,i}^{2/3},$$

where $a \lesssim b$ means $a \le \widetilde{\mathcal{O}}(b)$, the last step is due to the fact that $a_j \le \frac{1}{3DG}$, which leads to $G_j^2 \le \lambda_j$, and the relationship between ℓ_1-norm and ℓ_2-norm: $\|\mathbf{x}\|_1 \le \sqrt{d}\|\mathbf{x}\|_2$ for any $\mathbf{x} \in \mathbb{R}^d$. Denoting by $\textsc{Reg} \triangleq \textsc{Meta-Reg} + \textsc{Base-Reg}$, we obtain a problem-dependent bound for $\sum_{t\in[T]}\langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x^\star_{t,i}\rangle$:

$$\sum_{t=1}^T \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x^\star_{t,i}\rangle \le \frac{\textsc{Reg}}{a_j} + a_j G^2 \sum_{t=1}^T \|x_{t,i} - x^\star_{t,i}\|^2 \lesssim \textsc{Reg} + G\sqrt{\textsc{Reg}\sum_{t=1}^T \|x_{t,i} - x^\star_{t,i}\|^2},$$

Algorithm 5 Known strong monotonicity parameter μ and box domain (Baby & Wang, 2022)

Input: box domain $\mathcal{X}_i$ with diameter B, gradient upper bound G, strong monotonicity parameter μ, dimension d, loss functions $\{g_t(\cdot)\}_{t=1}^T$

Initialize: d FLH-OGD instances (Algorithm 4) $\mathcal{A}_1, \dots, \mathcal{A}_d$ with feasible domain $[-B, B]$ and learning rate $\varepsilon_k = \frac{1}{2(2B+G/\mu)^2}$

for $t = 1, \dots, T$ **do**

  Receive $x_{t,i}^{(k)}$ from $\mathcal{A}_k$ for $k \in [d]$

  Submit $x_{t,i} \triangleq (x_{t,i}^{(1)}, \dots, x_{t,i}^{(d)})$ and receive loss function $g_{t,i}$

  Construct surrogate loss $h_t^{(k)}(x) \triangleq (x - (x_{t,i}^{(k)} - \nabla g_t(x_{t,i})^{(k)}/\mu))^2$

  for $k = 1, \dots, d$ **do**

    Update $x_{t,i}^{(k)}$ to $x_{t+1,i}^{(k)}$ via $\mathcal{A}_k$ with loss function $h_{t,i}^{(k)}$

  end

end

Algorithm 6 Known strong monotonicity parameter μ (Baby & Wang, 2022)

Input: domain $\mathcal{X}_i$, gradient upper bound G, strong monotonicity parameter μ, loss functions $\{s_t(\cdot)\}_{t=1}^T$

Initialize: tightest box $\widehat{\mathcal{X}}_i$ that circumscribes $\mathcal{X}_i$, initialization $\mathcal{A}$ of Algorithm 5 with feasible set $\widehat{\mathcal{X}}_i$ and gradient upper bound $2G$

for $t = 1, \dots, T$ **do**

  Receive $\widehat{x}_{t,i}$ from $\mathcal{A}$

  Submit $x_{t,i} = \Pi_{\mathcal{X}_i}[\widehat{x}_{t,i}] = \arg\min_{x \in \mathcal{X}_i} \|x - \widehat{x}_{t,i}\|_1$ and receive gradient feedback $v_{t,i}(\mathbf{x}_t)$

  Construct surrogate loss $g_t(x) \triangleq s_t(x) + G \cdot \|x - \Pi_{\mathcal{X}_i}(x)\|_1$

  Update $\widehat{x}_{t,i}$ to $\widehat{x}_{t+1,i}$ via $\mathcal{A}$ with loss function g_t and strong monotonicity parameter μ

end

where the last step is by considering whether the optimal value of $a_j = \sqrt{\frac{\textsc{Reg}}{G^2 \sum_{t=1}^T \|x_{t,i} - x_{t,i}^\star\|^2}}$ is covered by the candidate pool $\mathcal{H} = \{a_j \mid a_j = 2^{-j}/(3DG), j \in [M]\}$ with $M = \lceil \frac{1}{2} \log_2 T \rceil + 1$. Finally, the i-th player's regret can be bounded by

$$\sum_{t=1}^T \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star \rangle - \frac{\mu}{2} \sum_{t=1}^T \|x_{t,i} - x_{t,i}^\star\|^2 \le \textsc{Reg} + G\sqrt{\textsc{Reg} \sum_{t=1}^T \|x_{t,i} - x_{t,i}^\star\|^2} - \frac{\mu}{2} \sum_{t=1}^T \|x_{t,i} - x_{t,i}^\star\|^2 \le \frac{1}{\mu}\textsc{Reg},$$

where the last step is due to $\sqrt{xy} \le \frac{ax}{2} + \frac{y}{2a}$ for any $x, y, a > 0$. Combining the regret bounds of all players, we obtain

$$\mu\textsc{Dist-Err} \le 2\sum_{i=1}^N \left(\sum_{t=1}^T \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star \rangle - \frac{\mu}{2} \sum_{t=1}^T \|x_{t,i} - x_{t,i}^\star\|^2 \right) \lesssim \frac{1}{\mu} \sum_{i=1}^N (1 + T^{1/3} P_{T,i}^{2/3}) \le \frac{N}{\mu}(1 + T^{1/3} P_T^{2/3}),$$

which finishes the proof. ∎

C.4. Proof of Theorem 3

Before the analysis of this part, we denote by $\bar{\mathbf{x}}_T^\star$ the Nash equilibrium of the averaged game from $t = 1$ to T, whose utility gradient is $\bar{v}_T(\cdot) \triangleq \sum_{t \in [T]} v_t(\cdot)/T$. Note that the averaged Nash equilibrium $\bar{\mathbf{x}}_T^\star$ must exist and is unique because: *(i)* the average of multiple strongly monotone games is still strongly monotone, due to Definition 1; and *(ii)* a monotone game admits a unique Nash equilibrium (Rosen, 1965).

In the following, we give the detailed proof of Theorem 3 .

Proof. The distance tracking error can be decomposed with the averaged Nash equilibrium $\bar{\mathbf{x}}_T^\star$ as an intermediate variable:

$$\mu\textsc{Dist-Err} = \mu\sum_{t=1}^T \|\mathbf{x}_t - \mathbf{x}_t^\star\|^2 \le 2\underbrace{\mu\sum_{t=1}^T \|\mathbf{x}_t - \bar{\mathbf{x}}_T^\star\|^2}_{\text{TERM (A)}} + 2\underbrace{\mu\sum_{t=1}^T \|\bar{\mathbf{x}}_T^\star - \mathbf{x}_t^\star\|^2}_{\text{TERM (B)}}.$$

To bound TERM (A), through the definition of strong monotonicity (Definition 1), it holds that

$$\frac{\mu}{2}\sum_{t=1}^T \|\mathbf{x}_t - \bar{\mathbf{x}}_T^\star\|^2 \le \sum_{t=1}^T \langle v_t(\mathbf{x}_t) - v_t(\bar{\mathbf{x}}_T^\star), \mathbf{x}_t - \bar{\mathbf{x}}_T^\star\rangle - \frac{\mu}{2}\sum_{t=1}^T \|\mathbf{x}_t - \bar{\mathbf{x}}_T^\star\|^2.$$

Multiplying both sides by 2, we obtain

$$\begin{aligned}
\text{TERM (A)} = \mu\sum_{t=1}^T \|\mathbf{x}_t - \bar{\mathbf{x}}_T^\star\|^2 &\le 2\underbrace{\left(\sum_{t=1}^T \langle v_t(\mathbf{x}_t), \mathbf{x}_t - \bar{\mathbf{x}}_T^\star\rangle - \frac{\mu}{2}\sum_{t=1}^T \|\mathbf{x}_t - \bar{\mathbf{x}}_T^\star\|^2\right)}_{\text{TERM (C)}} \\
&\quad + 2\sum_{t=1}^T \langle \bar{v}_T(\bar{\mathbf{x}}_T^\star), \bar{\mathbf{x}}_T^\star - \mathbf{x}_t\rangle + 2\sum_{t=1}^T \langle v_t(\bar{\mathbf{x}}_T^\star) - \bar{v}_T(\bar{\mathbf{x}}_T^\star), \bar{\mathbf{x}}_T^\star - \mathbf{x}_t\rangle \\
&\le 2\sum_{i=1}^N \left(\sum_{t=1}^T \ell_{t,i}(x_{t,i}) - \sum_{t=1}^T \ell_{t,i}(\bar{x}_{T,i}^\star)\right) + 2DW_T \\
&\lesssim \widetilde{\mathcal{O}}(1) + DW_T, \qquad \text{(by using Theorem 2 with } P_{T,i} = 0)
\end{aligned}$$

where the second inequality is due to the definition of the surrogate loss function $\ell_{t,i}$ (3.2), $\bar{\mathbf{x}}_T^\star$ is the Nash equilibrium of $\bar{v}_T$, namely, $\langle \bar{v}_T(\bar{\mathbf{x}}_T^\star), \bar{\mathbf{x}}_T^\star - \mathbf{x}\rangle \le 0$ for any $\mathbf{x} \in \mathcal{X}$ and

$$\sum_{t=1}^T \langle v_t(\bar{\mathbf{x}}_T^\star) - \bar{v}_T(\bar{\mathbf{x}}_T^\star), \bar{\mathbf{x}}_T^\star - \mathbf{x}_t^\star\rangle \le D\sum_{t=1}^T \|v_t(\bar{\mathbf{x}}_T^\star) - \bar{v}_T(\bar{\mathbf{x}}_T^\star)\| \le D\sum_{t=1}^T \sup_{\mathbf{x}\in\mathcal{X}} \|v_t(\mathbf{x}) - \bar{v}_T(\mathbf{x})\| = DW_T. \tag{C.2}$$

TERM (B) can be bounded as

$$\begin{aligned}
\text{TERM (B)} = \mu\sum_{t=1}^T \|\bar{\mathbf{x}}_T^\star - \mathbf{x}_t^\star\|^2 &\le \sum_{t=1}^T \langle v_t(\mathbf{x}_t^\star) - v_t(\bar{\mathbf{x}}_T^\star), \mathbf{x}_t^\star - \bar{\mathbf{x}}_T^\star\rangle \qquad \text{(by Definition 1)} \\
&= \sum_{t=1}^T \langle v_t(\mathbf{x}_t^\star), \mathbf{x}_t^\star - \bar{\mathbf{x}}_T^\star\rangle + \sum_{t=1}^T \langle \bar{v}_T(\bar{\mathbf{x}}_T^\star), \bar{\mathbf{x}}_T^\star - \mathbf{x}_t^\star\rangle + \sum_{t=1}^T \langle v_t(\bar{\mathbf{x}}_T^\star) - \bar{v}_T(\bar{\mathbf{x}}_T^\star), \bar{\mathbf{x}}_T^\star - \mathbf{x}_t\rangle \le DW_T
\end{aligned}$$

where the first two terms in the second last step is non-positive because $\mathbf{x}_t^\star$ and $\bar{\mathbf{x}}_T^\star$ are both Nash equilibriums of the t-th round and the averaged game, respectively, i.e., $\langle v_t(\mathbf{x}_t^\star), \mathbf{x}_t^\star - \mathbf{x}\rangle \le 0$ and $\langle \bar{v}_T(\bar{\mathbf{x}}_T^\star), \bar{\mathbf{x}}_T^\star - \mathbf{y}\rangle \le 0$ for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. The last term can be bounded using (C.2) again. Combining the bounds of TERM (A) and TERM (B) finishes the proof. ■

C.5. Proof of Corollary 1

Proof. The proof is direct by combining Proposition 1 and Theorem 3. ■

C.6. Useful Lemmas

Lemma 3 (Theorem 3 of Zhang et al. (2018))**.** *Assume the domain has a diameter D, i.e., $\|\mathbf{x} - \mathbf{y}\| \le D$ for any $\mathbf{x}, \mathbf{y} \in \mathcal{W}$ and the gradient of the loss functions f_t has an upper bound G, i.e., $\|\nabla f_t(\cdot)\| \le G$, then* Ader *enjoys*

$$\sum_{t=1}^T f_t(\mathbf{w}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \le \mathcal{O}\left(\sqrt{T(1 + P_T)}\right),$$

where $P_T \triangleq \sum_{t=2}^T \|\mathbf{u}_t - \mathbf{u}_{t-1}\|$ denotes the path length.

Lemma 4 (Lemma 1 of Wang et al. (2019)). *Due to the construction of the surrogate loss functions $s_{t,i,j}$, if the meta learner updates as in Algorithm 1, the meta regret on the surrogate loss can be bounded by*

$$\textsc{Meta-Reg} = \sum_{t=1}^{T} s_{t,i,j}(x_{t,i}) - \sum_{t=1}^{T} s_{t,i,j}(x_{t,i,j}) \leq \mathcal{O}(\log T).$$

Lemma 5 (Theorem 8 of Baby & Wang (2022)). *Let the feasible domain be $\mathcal{W} \subseteq \mathbb{R}^d$ and let the loss functions $\{f_t(\cdot)\}_{t=1}^T$ are H-strongly convex in ℓ_2-norm and satisfy $\|\nabla f_t(\cdot)\|_\infty \leq G$. Then the algorithm (Figure 3 therein, restated in Algorithm 6 for self-containedness) guarantees that*

$$\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \widetilde{\mathcal{O}}\left(\frac{G^2}{H}\max\{d, d^{1/3}T^{1/3}\mathrm{TV}_T^{2/3}\}\right),$$

for any comparator sequence $\{\mathbf{u}_t\}_{t=1}^T \in \mathcal{W}$ with $\mathrm{TV}_T \triangleq \sum_{t=2}^T \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_1$. $\widetilde{\mathcal{O}}(\cdot)$ hides the dependence of $\log T$.

Note that the original proof omits the dependence of the gradient upper bound G and the strong convexity parameter H. For self-containedness, we restate the proof to illuminate the concerned parameter dependence.

Proof of Lemma 5. The regret can be decomposed as the regret summation of the surrogate loss on each dimension:

$$\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \frac{H}{2}\sum_{k=1}^{d}\left(\sum_{t=1}^{T} h_t^{(k)}(\mathbf{w}_t^{(k)}) - \sum_{t=1}^{T} h_t^{(k)}(\mathbf{u}_t^{(k)})\right),$$

where $\mathbf{w}_t^{(k)}$ denotes the k-th dimension of $\mathbf{w}_t$ and so does $\mathbf{u}_t^{(k)}$. $h_t^{(k)}(x) \triangleq \left(x - \left(\mathbf{w}_t^{(k)} - \frac{\nabla f_t(\mathbf{w}_t)^{(k)}}{H}\right)\right)^2$ is a squared surrogate loss function. For general squared loss function $g_t(x) \triangleq (x-y)^2$, where $x \in [-B, B]$ and $y \in [-G, G]$, it is $1/(2(B+G)^2)$-exp-concave. It is known that for a α-exp-concave loss function, FLH-OGD enjoys $\mathcal{O}(\alpha^{-1}\log T)$ adaptive regret on each interval inside the whole time horizon (Hazan & Seshadhri, 2007, Lemma 8). As a result, the dynamic regret of the squared loss function $\{g_t(\cdot)\}_{t=1}^T$ is about $\mathcal{O}((B+G)^2(1+T^{2/3}P_T^{1/3}))$ (Baby & Wang, 2022, Lemma 18). Since the surrogate loss $h_t^{(k)}(\cdot)$ is α-exp-concave with parameter

$$\alpha = \frac{1}{2\left(2B + \frac{G}{H}\right)^2},$$

we obtain

$$\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \lesssim \frac{H}{2}\sum_{k=1}^{d} 2\left(2B + \frac{G}{H}\right)^2 (1 + T^{2/3}P_T^{1/3}) \lesssim \frac{G^2}{H}(1 + T^{2/3}P_T^{1/3}),$$

where $a \lesssim b$ means $a \leq \widetilde{\mathcal{O}}(b)$, and the last step is true since we consider $H \leq 1$ without loss of generality. ■

D. Analysis for Section 4

In this section, we provide the detailed analysis of Section 4, including the proof of Lemma 1 in Appendix D.1, the proof of Theorem 4 in Appendix D.2, the proof of Theorem 5 in Appendix D.3 and the proof of Corollary 2 in Appendix D.4. Finally, in Appendix D.5, we list some useful lemmas.

D.1. Proof of Lemma 1

Proof. First, by the standard dynamic regret analysis of OOGD (see Lemma 6), it holds that

$$\sum_{t=1}^{T}\langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star\rangle \leq \frac{D^2 + 2DP_{T,i}}{2\eta_i} + \eta_i\left(G^2 + \sum_{t=2}^{T}\|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2\right) - \frac{1}{4\eta_i}\sum_{t=2}^{T}\|x_{t,i} - x_{t-1,i}\|^2.$$

Diving into the second term (i.e., gradient variation) above, we have

$$\sum_{t=2}^{T} \|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2 = \sum_{t=2}^{T} \|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_t) + v_{t-1,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2$$
$$\leq 2\sum_{t=2}^{T} \sup_{\mathbf{x}\in\mathcal{X}} \|v_{t,i}(\mathbf{x}) - v_{t-1,i}(\mathbf{x})\|^2 + 2\sum_{t=2}^{T} \|v_{t-1,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2$$
$$\leq 2V_T + 2\sum_{t=2}^{T} \|v_{t-1}(\mathbf{x}_t) - v_{t-1}(\mathbf{x}_{t-1})\|^2 \leq 2V_T + 2L^2 \sum_{t=2}^{T}\sum_{j=1}^{N} \|x_{t,j} - x_{t-1,j}\|^2, \tag{D.1}$$

where the second inequality holds because if we denote by $h(\mathbf{x}) \triangleq \|v_{t,i}(\mathbf{x}) - v_{t-1,i}(\mathbf{x})\|^2$ and $g(\mathbf{x}) \triangleq \|v_t(\mathbf{x}) - v_{t-1}(\mathbf{x})\|^2$, then $h(\mathbf{x}) \leq g(\mathbf{x})$ is true obviously, which consequently implies $\sup_{\mathbf{x}\in\mathcal{X}} h(\mathbf{x}) \leq \sup_{\mathbf{x}\in\mathcal{X}} g(\mathbf{x})$. Thus we obtain

$$\sum_{t=1}^{T} \sup_{\mathbf{x}\in\mathcal{X}} \|v_{t,i}(\mathbf{x}) - v_{t-1,i}(\mathbf{x})\|^2 \leq \sum_{t=1}^{T} \sup_{\mathbf{x}\in\mathcal{X}} \|v_t(\mathbf{x}) - v_{t-1}(\mathbf{x})\|^2 = V_T.$$

The last inequality in (D.1) uses Assumption 3 and the fact that $\|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 = \sum_{j=1}^{N} \|\mathbf{x}_{t,j} - \mathbf{x}_{t-1,j}\|^2$. Finally, we have

$$\sum_{t=1}^{T} \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star \rangle \leq \frac{D^2 + 2DP_{T,i}}{2\eta_i} + \eta_i(G^2 + 2V_T) + 2\eta_i L^2 \sum_{j=1}^{N} S_j - \frac{1}{4\eta_i} S_i,$$

where $S_j \triangleq \sum_{t=2}^{T} \|x_{t,j} - x_{t-1,j}\|^2$ denotes the cumulative switching cost of the j-th player. ■

D.2. Proof of Theorem 4

In the following, we provide a formal version of Theorem 4 and the corresponding proof.

Theorem 6. *Under Assumptions 1-3, if all problem-dependent quantities are known a priori, the i-th player runs* OOGD*:*

$$x_{t,i} = \Pi_{\mathcal{X}_i}[\widehat{x}_{t,i} - \eta_i m_{t,i}],\ \widehat{x}_{t+1,i} = \Pi_{\mathcal{X}_i}[\widehat{x}_{t,i} - \eta_i v_{t,i}(\mathbf{x}_t)], \tag{D.2}$$

with optimism $m_{t,i} = v_{t-1,i}(\mathbf{x}_{t-1})$, then the distance tracking error can be upper-bounded by

- $\mathcal{O}(\sqrt{(1 + V_T + P_T)(1 + P_T)})$ *if choosing the step size as*

$$\eta_i = \min\left\{\sqrt{\frac{D^2 + 2DP_{T,i}}{2G^2 + 4V_T}}, \frac{1}{2L\sqrt{N}}\right\};$$

- $\mathcal{O}(1 + W_T)$ *if choosing the step size as*

$$\eta_i = \min\left\{\sqrt{\frac{D^2}{2G^2 + 4V_T}}, \frac{1}{2L\sqrt{N}}\right\}.$$

Proof of Theorem 6. We first prove the gradient-variation bound, then the gradient-variance bound.

Gradient-Variation Bound. Summing the regret bounds of all players using Lemma 1 gives

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star \rangle = \sum_{i=1}^{N} \left(\frac{D^2 + 2DP_{T,i}}{2\eta_i} + \eta_i(G^2 + 2V_T)\right) + 2L^2 \sum_{i=1}^{N} \eta_i \sum_{j=1}^{N} S_j - \sum_{i=1}^{N} \frac{1}{4\eta_i} S_i.$$

Our goal is to design the players' step sizes $\eta_1, \dots, \eta_N$ such that

$$2L^2 \sum_{i=1}^N \eta_i \sum_{j=1}^N S_j - \sum_{i=1}^N \frac{1}{4\eta_i} S_i = \sum_{i=1}^N \left(2L^2 \sum_{j=1}^N \eta_j - \frac{1}{4\eta_i} \right) S_i < 0,$$

which means the step sizes need to satisfy $2L^2 \sum_{j\in[N]} \eta_j \leq \frac{1}{4\eta_i}$. This requirement is easy to satisfy. Assume there is a constant X such that $2L^2 \sum_{j\in[N]} \eta_j \leq X \leq 1/(4\eta_i)$. Solving the second inequality gives $\eta_i \leq 1/(4X)$ for all $i \in [N]$. Plugging this into the first inequality, we have $L^2 N/(2X) \leq X$. We can set $X = L\sqrt{N/2}$, and it suffices to choose

$$\eta_i \leq \frac{1}{2L\sqrt{N}}.$$

Overall the distance tracking error is at most

$$\begin{aligned}
\mu\text{DIST-ERR} &\leq \sum_{i=1}^N \sum_{t=1}^T \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x^\star_{t,i} \rangle \leq \sum_{i=1}^N \left(\frac{D^2 + 2DP_{T,i}}{2\eta_i} + \eta_i (G^2 + 2V_T) \right) \qquad (D.3)\\
&\leq \sum_{i=1}^N \left(\sqrt{2(G^2 + 2V_T)(D^2 + 2DP_{T,i})} + L\sqrt{N}(D^2 + 2DP_{T,i}) \right) \lesssim \sqrt{(1 + V_T + P_T)(1 + P_T)},
\end{aligned}$$

where $a \lesssim b$ means $a \leq \widetilde{\mathcal{O}}(b)$ and the last inequality is by choosing the step size as

$$\eta_i = \min \left\{ \sqrt{\frac{D^2 + 2DP_{T,i}}{2G^2 + 4V_T}}, \frac{1}{2L\sqrt{N}} \right\}. \qquad (D.4)$$

Gradient-Variance Bound. Let $\bar{\mathbf{x}}^\star_T$ be the Nash equilibrium of the averaged game over the whole time horizon T. Following the similar analysis in Appendix C.4, we have

$$\begin{aligned}
\mu\text{DIST-ERR} &\leq 2 \sum_{i=1}^N \sum_{t=1}^T \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - \bar{x}^\star_{T,i} \rangle + 4DW_T \leq 2 \sum_{i=1}^N \left(\frac{D^2}{2\eta_i} + \eta_i (G^2 + 2V_T) \right) + 4DW_T\\
&\leq 2 \sum_{i=1}^N \left(D\sqrt{2(G^2 + 2V_T)} + LD^2\sqrt{N} \right) + 4DW_T \lesssim 1 + W_T,
\end{aligned}$$

where the second step reuse Lemma 6 with $P_{T,i} = 0$ and the third step is by setting the step size as

$$\eta = \min \left\{ \sqrt{\frac{D^2}{2G^2 + 4V_T}}, \frac{1}{2L\sqrt{N}} \right\}, \qquad (D.5)$$

and the last step is due to the relationship between gradient variation V_T and gradient variance W_T:

$$\begin{aligned}
V_T &= \sum_{t=1}^T \sup_{\mathbf{x}\in\mathcal{X}} \|v_t(\mathbf{x}) - v_{t-1}(\mathbf{x})\|^2 = 4G^2 \sum_{t=1}^T \sup_{\mathbf{x}\in\mathcal{X}} \left(\frac{\|v_t(\mathbf{x}) - v_{t-1}(\mathbf{x})\|}{2G} \right)^2\\
&\leq 2G \sum_{t=1}^T \sup_{\mathbf{x}\in\mathcal{X}} \|v_t(\mathbf{x}) - v_{t-1}(\mathbf{x})\| \qquad \text{(by Assumption 1)}\\
&\leq 4G \sum_{t=1}^T \sup_{\mathbf{x}\in\mathcal{X}} \|v_t(\mathbf{x}) - \bar{v}_T(\mathbf{x})\| + 4G \sum_{t=1}^T \sup_{\mathbf{x}\in\mathcal{X}} \|v_{t-1}(\mathbf{x}) - \bar{v}_T(\mathbf{x})\|\\
&\leq 8G \sum_{t=1}^T \sup_{\mathbf{x}\in\mathcal{X}} \|v_t(\mathbf{x}) - \bar{v}_T(\mathbf{x})\| = 8GW_T, \qquad (D.6)
\end{aligned}$$

which finishes the proof. ■

Remark 5. The tuning of the step sizes requires the players number N. We note that this requirement is mild, which also appears in other multi-player analyses, e.g., in the work of Syrgkanis et al. (2015). ¶

D.3. Proof of Theorem 5

To start with, we specify the step sizes of the meta learner:

$$\varepsilon_{0,i} = \varepsilon_{1,i} = \frac{1}{L}, \text{ and } \varepsilon_{t,i} = \frac{1}{\sqrt{L^2 + \sum_{s=2}^{t-1} \|v_{s,i}(\mathbf{x}_s) - v_{s-1,i}(\mathbf{x}_{s-1})\|^2}} \text{ for } 2 \le t \le T-1. \tag{D.7}$$

In the following, we provide a formal version of Theorem 5 and the corresponding proof.

Theorem 7. *Under Assumptions 1-3, Algorithm 2 enjoys the following guarantees simultaneously for the distance tracking error and regret of individual players when they agree to run the same algorithm (abbreviated as honest regret):*

$$\textsc{Dist-Err} \le \begin{cases} \mathcal{O}\left(\sqrt{(1+V_T+P_T)(1+P_T)}\right) \\ \mathcal{O}(1+W_T) \end{cases}, \quad \sum_{t=1}^{T} \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^{\star} \rangle \le \begin{cases} \mathcal{O}\left(\sqrt{(1+V_T+P_T)(1+P_{T,i})}\right) \\ \mathcal{O}\left(\sqrt{(1+W_T+P_{T,i})(1+P_{T,i})}\right) \end{cases},$$

where P_T, V_T, W_T denote the path length, gradient variation and gradient variance, respectively.

Proof of Theorem 7. We first prove the gradient-variation bound and then the gradient-variance bound. For simplicity, we assume all problem-dependent quantities are known for a moment and illuminate the setting of the step size pool later.

Gradient-Variation Bound. First we decompose the regret of the i-th player with respect to its j-th base learner:

$$\sum_{t=1}^{T} \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^{\star} \rangle = \underbrace{\sum_{t=1}^{T} \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i,j} \rangle}_{\textsc{Meta-Reg}} + \underbrace{\sum_{t=1}^{T} \langle v_{t,i}(\mathbf{x}_t), x_{t,i,j} - x_{t,i}^{\star} \rangle}_{\textsc{Base-Reg}}.$$

The regret of the meta learner, due to Lemma 7 (optimistic Hedge is a special case of optimistic FTRL by choosing negative entropy as the regularizer), can be bounded as

$$\sum_{t=1}^{T} \langle \boldsymbol{\ell}_{t,i}, \boldsymbol{p}_{t,i} - \boldsymbol{e}_j \rangle \le \frac{\mathrm{KL}(\boldsymbol{e}_j, \boldsymbol{p}_{1,i})}{\varepsilon_{T-1,i}} + \sum_{t=1}^{T} \varepsilon_{t-1,i} \|\boldsymbol{\ell}_{t,i} - \boldsymbol{m}_{t,i}\|_\infty^2 - \sum_{t=1}^{T-1} \frac{1}{8\varepsilon_{t-1,i}} \|\boldsymbol{p}_{t,i} - \boldsymbol{p}_{t+1,i}\|_1^2, \tag{D.8}$$

where the first term above can be bounded by plugging the setup of learning rate (D.7):

$$\frac{\mathrm{KL}(\boldsymbol{e}_j, \boldsymbol{p}_{1,i})}{\varepsilon_{T-1,i}} = \frac{\ln(1/p_{1,i,j})}{\varepsilon_{T-1,i}} = \ln M \sqrt{L^2 + \sum_{t=2}^{T} \|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2}.$$

The second term in (D.8) can be analyzed as

$$\begin{aligned}
\sum_{t=1}^{T} \varepsilon_{t-1,i} \|\boldsymbol{\ell}_{t,i} - \boldsymbol{m}_{t,i}\|_\infty^2 &= \varepsilon_{0,i} \|\boldsymbol{\ell}_{1,i}\|_\infty^2 + \sum_{t=2}^{T} \varepsilon_{t-1,i} \max_{j \in [M]} \langle v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1}), x_{t,i,j} \rangle^2 \\
&\le \frac{D^2(G+\lambda D)^2}{L} + D^2 \sum_{t=2}^{T} \varepsilon_{t-1,i} \left\|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\right\|^2 \\
&= \mathcal{O}(1) + D^2 \sum_{t=2}^{T} \frac{\left\|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\right\|^2}{\sqrt{L^2 + \sum_{s=2}^{t-1} \left\|v_{s,i}(\mathbf{x}_s) - v_{s-1,i}(\mathbf{x}_{s-1})\right\|^2}} \qquad \text{(by (D.7))} \\
&\le \mathcal{O}(1) + 4D^2 \sqrt{L^2 + \sum_{t=2}^{T} \left\|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\right\|^2},
\end{aligned}$$

where the first step is because of $\boldsymbol{m}_{1,i} = \mathbf{0}$, the second step is due to $\max_{t,i,j} \ell_{t,i,j} \leq GD + \lambda D^2$ (see (4.3) for the definition of the meta learner's loss) and the last step uses $\sum_{t=1}^T a_t / \sqrt{1 + \sum_{s=1}^{t-1} a_s} \leq 4\sqrt{1 + \sum_{t=1}^T a_t} + \max_{t \in [T]} a_t$ for any $a_1, a_2, \ldots, a_T > 0$ (Pogodin & Lattimore, 2019, Lemma 4.8). Overall, the regret of the meta learner is at most

$$\sum_{t=1}^T \langle \boldsymbol{\ell}_{t,i}, \boldsymbol{p}_{t,i} - \boldsymbol{e}_j \rangle \leq (4D^2 + \ln M) \sqrt{L^2 + \sum_{t=2}^T \|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2} - \frac{L}{8} \sum_{t=1}^{T-1} \|\boldsymbol{p}_{t,i} - \boldsymbol{p}_{t+1,i}\|_1^2 + \mathcal{O}(1).$$

Plugging in the definition of the surrogate loss function $\boldsymbol{\ell}_{t,i}$ (4.3), the meta regret can be bounded as

$$\begin{aligned} \textsc{Meta-Reg} \leq{}& (4D^2 + \ln M) \sqrt{L^2 + \sum_{t=2}^T \|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2} - \frac{L}{8} \sum_{t=1}^{T-1} \|\boldsymbol{p}_{t,i} - \boldsymbol{p}_{t+1,i}\|_1^2 \\ & - \lambda \sum_{t=2}^T \sum_{j=1}^M p_{t,i,j} \|x_{t,i,j} - x_{t-1,i,j}\|^2 + \lambda \sum_{t=2}^T \|x_{t,i,j} - x_{t-1,i,j}\|^2 + \mathcal{O}(1). \end{aligned}$$

As for the base regret, following the standard analysis of OOGD (deferred to Lemma 6), it holds that

$$\textsc{Base-Reg} \leq \frac{D^2 + 2DP_{T,i}}{2\eta_{i,j}} + \eta_{i,j} \left(G^2 + \sum_{t=2}^T \|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2 \right) - \frac{1}{4\eta_{i,j}} \sum_{t=2}^T \|x_{t,i,j} - x_{t-1,i,j}\|^2.$$

Suppose there is a uniform upper bound X for $\eta_{i,j}$, the step size of the j-th base learner of the i-th player, for all $i \in [N], j \in [M]$. Overall, the regret of the i-th player is at most

$$\begin{aligned} \sum_{t=1}^T \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star \rangle \leq{}& \mathcal{O}(1) - \frac{L}{8} \sum_{t=1}^{T-1} \|\boldsymbol{p}_{t,i} - \boldsymbol{p}_{t+1,i}\|_1^2 - \lambda \sum_{t=2}^T \sum_{j=1}^M p_{t,i,j} \|x_{t,i,j} - x_{t-1,i,j}\|^2 \\ & + \left(\lambda - \frac{1}{4X}\right) \sum_{t=2}^T \|x_{t,i,j} - x_{t-1,i,j}\|^2 + \frac{5D^2 + \ln M + 2DP_{T,i}}{2\eta_{i,j}} \\ & + \eta_{i,j} \left(G^2 + L^2 + 2 \sum_{t=2}^T \|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2 \right) \qquad \text{(D.9)} \\ \leq{}& \frac{c_1 + 2DP_{T,i}}{2\eta_{i,j}} + \eta_{i,j}(c_2 + 4V_T) + 8L^2 D^2 X \sum_{t=2}^T \sum_{i=1}^N \|\boldsymbol{p}_{t,i} - \boldsymbol{p}_{t-1,i}\|^2 \qquad \text{(by (D.1))} \\ & + 8L^2 X \sum_{i=1}^N \sum_{t=2}^T \sum_{j=1}^M p_{t,i,j} \|x_{t,i,j} - x_{t-1,i,j}\|^2 + \left(\lambda - \frac{1}{4X}\right) \sum_{t=2}^T \|x_{t,i,j} - x_{t-1,i,j}\|^2 \\ & - \lambda \sum_{t=2}^T \sum_{j=1}^M p_{t,i,j} \|x_{t,i,j} - x_{t-1,i,j}\|^2 - \frac{L}{8} \sum_{t=1}^{T-1} \|\boldsymbol{p}_{t,i} - \boldsymbol{p}_{t+1,i}\|_1^2 + \mathcal{O}(1) \end{aligned}$$

where $c_1 = 5D^2 + \ln M, c_2 = G^2 + L^2$. The first inequality holds since $\sqrt{xy} \leq \frac{ax}{2} + \frac{y}{2a}$ for any $x, y, a > 0$,

$$\sqrt{L^2 + \sum_{t=2}^T \|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2} \leq \frac{1}{2\eta_{i,j}} + \frac{\eta_{i,j}\left(L^2 + \sum_{t=2}^T \|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2\right)}{2},$$

and the second step is true since (D.1) and

$$\|x_{t,i} - x_{t-1,i}\|^2 = \left\| \sum_{j=1}^M p_{t,i,j} x_{t,i,j} - \sum_{j=1}^M p_{t-1,i,j} x_{t-1,i,j} \right\|^2$$

$$= \left\| \sum_{j=1}^{M} p_{t,i,j} x_{t,i,j} - \sum_{j=1}^{M} p_{t,i,j} x_{t-1,i,j} + \sum_{j=1}^{M} p_{t,i,j} x_{t-1,i,j} - \sum_{j=1}^{M} p_{t-1,i,j} x_{t-1,i,j} \right\|^2$$

$$\leq 2 \sum_{j=1}^{M} p_{t,i,j} \left\| x_{t,i,j} - x_{t-1,i,j} \right\|^2 + 2D^2 \left\| \boldsymbol{p}_{t,i} - \boldsymbol{p}_{t-1,i} \right\|_1^2. \tag{D.10}$$

Summing all players' regret, we have

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^{\star} \rangle \leq \mathcal{O}(1) + \sum_{i=1}^{N} \left(\frac{c_1 + 2DP_{T,i}}{2\eta_{i,j}} + \eta_{i,j}(c_2 + 4V_T) \right)$$
$$+ \sum_{i=1}^{N} \left(\lambda - \frac{1}{4X} \right) \sum_{t=2}^{T} \left\| x_{t,i,j} - x_{t-1,i,j} \right\|^2 + \left(8NL^2D^2X - \frac{L}{8} \right) \sum_{i=1}^{N} \sum_{t=2}^{T} \left\| \boldsymbol{p}_{t,i} - \boldsymbol{p}_{t-1,i} \right\|_1^2$$
$$+ (8NL^2X - \lambda) \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{j=1}^{M} p_{t,i,j} \left\| x_{t,i,j} - x_{t-1,i,j} \right\|^2. \tag{D.11}$$

Following the above analysis, for all $i \in [N]$ and $j \in [M]$, we choose

$$\lambda = 16NLD^2, \quad \eta_{i,j} \leq \frac{1}{128NLD^2} = X$$

such that

$$\lambda - \frac{1}{4X} = -16NLD^2 < 0, \quad 8NL^2X - \lambda \leq -15NLD^2 < 0, \quad 8NL^2D^2X - \frac{L}{8} = -\frac{1}{16} < 0.$$

The distance tracking error can be thus bounded as

$$\mu \textsc{Dist-Err} \leq \sum_{t=1}^{T} \langle v_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^{\star} \rangle = \sum_{i=1}^{N} \sum_{t=1}^{T} \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^{\star} \rangle$$
$$\leq \sum_{i=1}^{N} \left(\frac{c_1 + 2DP_{T,i}}{2\eta_{i,j}} + \eta_{i,j}(c_2 + 4V_T) \right) \lesssim \sqrt{(1 + V_T + P_T)(1 + P_T)}, \tag{D.12}$$

where $a \lesssim b$ means $a \leq \widetilde{\mathcal{O}}(b)$ and the last step is by choosing the step size optimally as $\eta_{i,j} = \eta_i^{\star}$, where

$$\eta_i^{\star} = \min \left\{ \sqrt{\frac{c_1 + 2DP_{T,i}}{2c_2 + 8V_T}}, \frac{1}{128NLD^2} \right\}. \tag{D.13}$$

As for the honest regret of individual players, to start with, since $\sum_{t \in [T]} \langle v_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^{\star} \rangle \geq \sum_{t \in [T]} \langle v_t(\mathbf{x}_t) - v_t(\mathbf{x}_t^{\star}), \mathbf{x}_t - \mathbf{x}_t^{\star} \rangle \geq \mu \sum_{t \in [T]} \|\mathbf{x}_t - \mathbf{x}_t^{\star}\|^2 \geq 0$, it means that the left-hand side of (D.11) is positive. As a result, we can move the terms that are canceled to the left hand side of (D.11) to upper-bound them. By doing so, we obtain

$$\sum_{t=2}^{T} \|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 \leq 2 \sum_{i=1}^{N} \sum_{t=2}^{T} \left(D^2 \left\| \boldsymbol{p}_{t,i} - \boldsymbol{p}_{t-1,i} \right\|^2 + \sum_{j=1}^{M} p_{t,i,j} \left\| x_{t,i,j} - x_{t-1,i,j} \right\|^2 \right) \lesssim \sqrt{(1 + V_T + P_T)(1 + P_T)}.$$

Thus the honest regret of the i-th player, via Lemma 1, can be bounded as

$$\sum_{t=1}^{T} \langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^{\star} \rangle \leq \frac{D^2 + 2DP_{T,i}}{2\eta_{i,j}} + \eta_{i,j}(G^2 + 2V_T) + 2\eta_{i,j} L^2 \sum_{t=2}^{T} \|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2$$
$$\lesssim \frac{D^2 + 2DP_{T,i}}{2\eta_{i,j}} + \eta_{i,j}(G^2 + 2V_T) + 2\eta_{i,j} L^2 \sqrt{(1 + V_T + P_T)(1 + P_T)}$$

$$\lesssim \frac{D^2 + 2DP_{T,i}}{2\eta_{i,j}} + \eta_{i,j}(1 + V_T + P_T) \lesssim \sqrt{(1 + P_{T,i})(1 + V_T + P_T)}, \tag{D.14}$$

where the last step is by setting the step size optimally as

$$\eta_i^\star = \sqrt{\frac{D^2 + 2DP_{T,i}}{N(1 + V_T + P_T)}}. \tag{D.15}$$

Gradient-Variance Bound. Following the same analysis as in Appendix C.4, the environmental non-stationarity can be extracted into the gradient-variance term:

$$\mu\text{DIST-ERR} \le 2\sum_{i=1}^{N}\sum_{t=1}^{T}\langle v_{t,i}(\mathbf{x}_t), x_{t,i} - \bar{x}_{T,i}^\star\rangle + 4DW_T \lesssim \sqrt{1 + V_T} + W_T \lesssim 1 + W_T,$$

where the last step holds due to (D.12) with $P_T = 0$ and the relationship between gradient variation V_T and gradient variance W_T (D.6). As for the honest regret of individual players, following the same argument of the last part, we have

$$\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 \le 2\sum_{i=1}^{N}\sum_{t=2}^{T}\left(D^2\|\boldsymbol{p}_{t,i} - \boldsymbol{p}_{t-1,i}\|^2 + \sum_{j=1}^{M} p_{t,i,j}\|x_{t,i,j} - x_{t-1,i,j}\|^2\right) \lesssim 1 + W_T.$$

Consequently, via Lemma 1, we have

$$\begin{aligned}
&\sum_{t=1}^{T}\langle v_{t,i}(\mathbf{x}_t), x_{t,i} - x_{t,i}^\star\rangle \le \frac{D^2 + 2DP_{T,i}}{2\eta_{i,j}} + \eta_{i,j}(G^2 + 2V_T) + 2\eta_{i,j}L^2\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 \\
&\lesssim \frac{D^2 + 2DP_{T,i}}{2\eta_{i,j}} + \eta_{i,j}(G^2 + 2V_T) + 2\eta_{i,j}L^2(1 + W_T) \\
&\lesssim \frac{D^2 + 2DP_{T,i}}{2\eta_{i,j}} + \eta_{i,j}(1 + W_T) \le \sqrt{(1 + P_{T,i})(1 + W_T + P_{T,i})},
\end{aligned} \tag{D.16}$$

where the last step is by setting the step size optimally as

$$\eta_i^\star = \sqrt{\frac{D^2 + 2DP_{T,i}}{N(1 + W_T)}}. \tag{D.17}$$

Step Size Pool Configuration. In the last part of the proof, we consider together different optimal tunings of the above guarantees, including (D.13), (D.15) and (D.17), and give the setting of step size pool that covers all of them. Specifically, we choose the step size pool as

$$\mathcal{H}_i^\eta \triangleq \left\{\eta_{i,j} \,\middle|\, \eta_{i,j} \approx \min\left\{\sqrt{\frac{1}{1+T}} \cdot 2^{j-1}, \frac{1}{NL}\right\}, j \in [M]\right\}, \tag{D.18}$$

where $M \approx \lceil \log_2(T+1) \rceil + 1 = \mathcal{O}(\log T)$, which finishes the proof. ■

D.4. Proof of Corollary 2

Proof. The proofs of gradient-variation and gradient-variance bound can be obtained directly by combining Proposition 1 and Theorem 5. As a result, in the following, we focus on the small-loss bound.

Small-Loss Bound. If we choose the step size of all players as η (to be specified later), following Lemma 6, it holds that

$$\sum_{t=1}^{T}\langle v_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_t^\star\rangle \le \frac{N(D^2 + 2DP_T)}{2\eta} + \eta NG^2 + \eta\sum_{i=1}^{N}\sum_{t=2}^{T}\|v_{t,i}(\mathbf{x}_t) - v_{t-1,i}(\mathbf{x}_{t-1})\|^2.$$

Next, we show that the second term above, i.e., the gradient-variation term, naturally implies small-loss term. Specifically,

$$\begin{aligned}\sum_{i=1}^{N}\sum_{t=2}^{T}\|v_{t,i}(\mathbf{x}_t)-v_{t-1,i}(\mathbf{x}_{t-1})\|^2 &\le 2\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\|v_{t,i}(\mathbf{x}_t)\|^2+\|v_{t-1,i}(\mathbf{x}_{t-1})\|^2\right)\\ \le 4\sum_{i=1}^{N}\sum_{t=1}^{T}\|v_{t,i}(\mathbf{x}_t)\|^2 &= 4\sum_{t=1}^{T}\|v_t(\mathbf{x}_t)\|^2 \le 16L\sum_{t=1}^{T}u_t(\mathbf{x}_t),\end{aligned} \tag{D.19}$$

where the equality uses the fact that $\sum_{i\in[N]}\|v_{t,i}(\mathbf{x}_t)\|^2 = \|v_t(\mathbf{x}_t)\|^2$ and the last step holds due to the property of nonnegative and smooth functions ($\|\nabla f(\cdot)\|_2^2 \le 4Lf(\cdot)$ for any L-smooth and non-negative function f) (Srebro et al., 2010, Lemma 3.1). Rearranging gives

$$\sum_{t=1}^{T}\langle v_t(\mathbf{x}_t),\mathbf{x}_t-\mathbf{x}_t^\star\rangle - 16\eta L\left(\sum_{t=1}^{T}u_t(\mathbf{x}_t)-u_t(\mathbf{x}_t^\star)\right) \le \frac{N(D^2+2DP_T)}{2\eta}+\eta NG^2+16\eta LF_T,$$

where $F_T \triangleq \sum_{t\in[T]}u_t(\mathbf{x}_t^\star)$ denotes the cumulative game value. Since $\sum_{t\in[T]}u_t(\mathbf{x}_t)-u_t(\mathbf{x}_t^\star) \le \sum_{t\in[T]}\langle v_t(\mathbf{x}_t),\mathbf{x}_t-\mathbf{x}_t^\star\rangle$,

$$(1-16\eta L)\left(\sum_{t=1}^{T}\langle v_t(\mathbf{x}_t),\mathbf{x}_t-\mathbf{x}_t^\star\rangle\right) \le \frac{N(D^2+2DP_T)}{2\eta}+\eta(NG^2+16LF_T).$$

Finally, we upper-bound the utility tracking error by

$$\textsc{Util-Err} \le \sum_{t=1}^{T}\langle v_t(\mathbf{x}_t),\mathbf{x}_t-\mathbf{x}_t^\star\rangle \le \frac{N(D^2+2DP_T)}{\eta}+2\eta(NG^2+16LF_T) \lesssim \sqrt{(1+F_T+P_T)(1+P_T)},$$

where the last step is by choosing the step sizes of each player equally as

$$\eta_1=\ldots=\eta_N=\eta=\min\left\{\sqrt{\frac{N(D^2+2DP_T)}{2NG^2+32LF_T}},\frac{1}{32L}\right\}. \tag{D.20}$$

Since (D.20) is covered by (D.18), the proof is finished. ■

D.5. Useful Lemmas

Lemma 6 (Optimistic OGD (Zhao et al., 2021, Theorem 1))**.** *If the domain diameter is bounded by D, i.e., $\|\mathbf{w}_1-\mathbf{w}_2\| \le D$ for any $\mathbf{w}_1,\mathbf{w}_2 \in \mathcal{W}$, the gradient norm is bounded by G, i.e., $\|\nabla f(\cdot)\| \le G$, the dynamic regret of* OOGD *is bounded by*

$$\sum_{t=1}^{T}f_t(\mathbf{w}_t)-\sum_{t=1}^{T}f_t(\mathbf{u}_t) \le \frac{D^2+2DP_T}{2\eta}+\eta\left(G^2+\sum_{t=2}^{T}\|\nabla f_t(\mathbf{w}_t)-\boldsymbol{m}_t\|^2\right)-\frac{1}{4\eta}\sum_{t=2}^{T}\|\mathbf{w}_t-\mathbf{w}_{t-1}\|^2,$$

where $P_T \triangleq \sum_{t=2}^{T}\|\mathbf{u}_t-\mathbf{u}_{t-1}\|$ denotes the path length.

Lemma 7 (Optimistic FTRL (Rakhlin & Sridharan, 2013a))**.** *Suppose the learning rates are non-increasing: $\eta_t \ge \eta_{t+1}$ for all $t \ge 1$. Optimistic FTRL with the following update rule:*

$$\mathbf{w}_{t+1}=\operatorname*{arg\,min}_{\mathbf{w}\in\mathcal{W}}\left\{\eta_t\left\langle\sum_{s=1}^{t}\nabla f_s(\mathbf{w}_s)+\boldsymbol{m}_{t+1},\mathbf{w}\right\rangle+\psi(\mathbf{w})\right\},$$

where $\psi(\mathbf{w}) \triangleq \sum_{i\in[d]}w_i\ln w_i$, guarantees that

$$\sum_{t=1}^{T}\langle\nabla f_t(\mathbf{w}_t),\mathbf{w}_t-\mathbf{u}\rangle \le \frac{\mathrm{KL}(\mathbf{u},\mathbf{w}_1)}{\eta_{T-1}}+\sum_{t=1}^{T}\eta_{t-1}\|\nabla f_t(\mathbf{w}_t)-\boldsymbol{m}_t\|_*^2-\sum_{t=1}^{T-1}\frac{1}{8\eta_{t-1}}\|\mathbf{w}_t-\mathbf{w}_{t+1}\|^2,$$

where $\mathrm{KL}(\mathbf{x},\mathbf{y}) \triangleq \sum_{i=1}^{d}x_i\ln(x_i/y_i)$ denotes the Kullback-Leibler divergence.