Uncoupled and Convergent Learning in Monotone Games under Bandit Feedback

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

We study the problem of no-regret learning algorithms for general monotone and smooth games and their last-iterate convergence properties. Specifically, we investigate the problem under bandit feedback and strongly uncoupled dynamics, which allows modular development of the multi-player system that applies to a wide range of real applications. We propose a mirror-descent-based algorithm, which converges in $O(T^{-1/4})$ and is also no-regret. The result is achieved by a dedicated use of two regularizations and the analysis of the fixed point thereof. The convergence rate is further improved to $O(T^{-1/2})$ in the case of strongly monotone games. Motivated by practical tasks where the game evolves over time, the algorithm is extended to time-varying monotone games. We provide the first non-asymptotic result in converging monotone games and give improved results for equilibrium tracking games.

1 Introduction

We consider multi-player online learning in games. In this problem, the cost function for each player is unknown to the player, and they need to learn to play the game through repeated interaction with other players. We focus on a class of monotone and smooth games, which was first introduced by Rosen (1965). This encapsulates a wide array of common games, such as two-player zero-sum games, convex-concave games, and zero-sum polymatrix games (Bregman and Fokin, 1987). Our goal is to find algorithms that solve the problem under bandit feedback and strongly uncoupled dynamics. Within this context, each player can only access information regarding the cost function associated with their chosen actions without prior insight into their counterparts. This allows modular development of the multi-player system in real applications and leverages existing single-agent learning algorithms for reuse.

Many works have focused on the time-average convergence to Nash equilibrium on learning in monotone games (Even-Dar et al., 2009; Syrgkanis et al., 2015; Farina et al., 2022). However, these works only guarantee the convergence of the time average of the joint action profile. Such convergence properties are less appealing, because while the trajectories of the players converge in the time-average sense, it may still exhibit cycling (Mertikopoulos et al., 2018). This jeopardizes the practical use of such algorithms.

Popular no-regret algorithms such as mirror descent have demonstrated convergence in the last iterate within specific scenarios, such as two-player zero-sum games (Cai et al., 2023) and strongly monotone games (Bravo et al., 2018; Drusvyatskiy et al., 2022; Lin et al., 2021). Yet convergence to Nash equilibrium in monotone and smooth games is not available unless one assumes exact gradient feedback and coordination of players (Cai et al., 2022; Cai and Zheng, 2023). It remains open as to whether a no-regret algorithm can efficiently converge to a Nash equilibrium in monotone games with bandit feedback and strongly uncoupled dynamics. In this paper, we investigate the pivotal question:

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

How fast can no-regret algorithms converge (in the last iterate) to a Nash equilibrium in general monotone and smooth games with bandit feedback and strongly uncoupled dynamics?

In this work, we present a mirror-descent-based algorithm designed to converge to the Nash equilibrium in monotone and smooth games. Our algorithm is uncoupled and convergent and is applicable to the general monotone and smooth game setting. Motivated by real applications, where many games are also time-varying, we extend our study to encompass time-varying monotone games. This justifies that our algorithm could be deployed in both stationary and non-stationary tasks.

We achieve state-of-the-art results in both monotone games and time-varying monotone games.

- In monotone and smooth games:
  - Under bandit feedback and strongly uncoupled dynamics, we show our algorithm achieves a last-iterate convergence rate of $O(T^{-1/4})$.
  - In cases where the game exhibits strong monotonicity, our result improves to $O(T^{-1/2})$, matching the current best available convergence rates for strongly monotone games (Drusvyatskiy et al., 2022; Lin et al., 2021).
  - Our algorithm is no regret albeit players may be self-interested. The individual regret is at most $O(T^{3/4})$ in monotone games and at most $O(T^{1/2})$ in strongly monotone games.
- In time-varying monotone and smooth games:
  - If the game eventually converges to a static state within a time frame of $O(T^{\alpha})$, our algorithm achieves convergence in $O(T^{-1/4+\alpha})$.
  - If the game does not converge but experiences gradual changes in the Nash equilibrium that evolves in $O(T^{\varphi})$, our algorithm exhibits convergence rates of $O\left(\max\left\{T^{2\varphi/3-2/3}, T^{(4\varphi+5)^2/72-9/8}\right\}\right)$. The algorithm outperforms best available results of $T^{\varphi/5-1/5}$ by Duvocelle et al. (2023) and $T^{\varphi/3-2/3}$ by Yan et al. (2023).

Table 1 and Table 2 summarize our results and the results of previous works.

2 Related Works

Monotone games The convergence of monotone games has been studied in a significant line of research. For a strongly monotone game under exact gradient feedback, the linear last-iterate convergence rate is known (Tseng, 1995; Liang and Stokes, 2019; Zhou et al., 2020). Under noisy gradient feedback, Jordan et al. (2023) showed a last-iterate convergence rate of $O(T^{-1})$. Under bandit feedback, Bervoets et al. (2020) proposed an algorithm that asymptotically converges to the equilibrium if it is unique. Bravo et al. (2018) subsequently introduced an algorithm with a last-iterate convergence rate of $O(T^{-1/3})$, while also ensuring the no-regret property. Later works (Lin et al., 2021) further improved the last-iterate convergence rate to $O(T^{-1/2})$ under bandit feedback using the self-concordant barrier function. Jordan et al. (2023) gave a result of the same rate, but with the additional assumption that the Jacobian of each player's gradient is Lipschitz continuous. In the case of bandit but noisy feedback (with a zero-mean noise), Lin et al. (2021) showed that the convergence rate is still $O(T^{-1/2})$.

For monotone but not strongly monotone games, Mertikopoulos and Zhou (2019) leveraged the dual averaging algorithm to demonstrate an asymptotic convergence rate under noisy gradient feedback. With access to the exact gradient information, Cai and Zheng (2023) gave a last-iterate convergence rate of $O(T^{-1})$. In the context of bandit feedback, Tatarenko and Kamgarpour (2019) proposed an algorithm that asymptotically converges to the Nash equilibrium. Table 1 provides a summary of the recent results.

Time-varying monotone games Motivated by real-world applications such as Cournot competition, where multiple firms supply goods to the market and pricing is subject to fluctuations due to factors like weather, holidays, and politics. Duvocelle et al. (2023) studied the strongly monotone game under a time-varying cost function. When the game converges to a static state, they propose an algorithm that achieves asymptotic convergence under bandit feedback. Assuming the cost function varies

Table 1: Summary of results for monotone games. "E" stands for the result in expectation and "P" stands for the result held in high probability. Strongly monotone games are abbreviated to "StroM", while monotone games are abbreviated to "M". We use "linear*" to denote the two-player zero-sum game, which is a special case of the linear game. We use "(N)" to remark that the results can also be obtained with noisy feedback.

	Class of games	Feedback	Results
Bravo et al. (2018)	StroM	bandit	$O(T^{-1/3})$ (E)
Drusvyatskiy et al. (2022)	StroM	bandit	$O(T^{-1/2})$ (E)
Lin et al. (2021)	StroM	bandit (N)	$O(T^{-1/2})$ (E)
Jordan et al. (2023)	StroM	noisy gradient	$O(T^{-1})$
Ours	StroM	bandit (N)	$O(T^{-1/2})$ (E & P)
Mertikopoulos and Zhou (2019)	M	noisy gradient	asymptotic
Cai and Zheng (2023)	M	exact gradient	$O(T^{-1})$
Tatarenko and Kamgarpour (2019)	M	bandit	asymptotic
Ours	M	bandit (N)	$O(T^{-1/4})$ (E)
Cai et al. (2023)	linear*	bandit	$O(T^{-1/6})$ (E)
Ours	linear	bandit	$O(T^{-1/6})$ (E)

Table 2: Summary of last-iterate convergence results for time-varying games. All results here are in expectation results. Strongly monotone games are abbreviated to "StroM", and monotone games are abbreviated to "M".

	Class of games	Time-varying property	Feedback	Results
Duvocelle et al. (2023)	StroM	converging in $O(T^{\alpha})$	bandit	asymptotic
Ours	M	converging in $O(T^{\alpha})$	bandit	$O(T^{-1/4+\alpha})$
Duvocelle et al. (2023)	StroM	$O(T^{\varphi})$ variation path	bandit	$O(T^{\varphi/5-1/5})$
Yan et al. (2023)	StroM	$O(T^{\varphi})$ variation path	exact gradient	$O(T^{\varphi/3-2/3})$
Ours	M	$O(T^{\varphi})$ variation path	bandit	$O\left(\max\{T^{2\varphi/3-2/3}, T^{(4\varphi+5)^2/72-9/8}\}\right)$

$O(T^{\phi})$ across a horizon T, Duvocelle et al. (2023) provided an algorithm that attains a convergence rate of $O(T^{\phi/5-1/5})$ under bandit feedback. Subsequent work of Yan et al. (2023) further improved this rate to $O(T^{\phi/3-2/3})$ under exact gradient feedback.

3 Preliminaries

We consider a multi-player game with n players, with the set of players denoted as $\mathcal{N}$. Each player i takes action on a compact and convex set $\mathcal{X}_i \subseteq \mathbb{R}^d$ of d dimensions, and has cost function $c_i(x_i, x_{-i})$, where $x_i \in \mathcal{X}_i$ is the action of the i-th player and $x_{-i} \in \prod_{j\in[n], j\neq i} \mathcal{X}_j$ is the action of all other players. We assume the radius of $\mathcal{X}_i$ is bounded, i.e., $\|x - x'\| \leq B, \forall x, x' \in \mathcal{X}_i$. Without loss of generality, we further assume $c_i(x) \in [0, 1]$.

In this work, we study a class of monotone continuous games, where the gradient of the cost functions is monotone and the cost functions continuous (Assumption 3.1). Games that satisfy this assumption include convex-concave games, convex potential games, extensive form games, Cournot competition, and splittable routing games. A discussion of these games is available in Section 3.1. Note that the class of monotone continuous games is commonly studied in the literature (Lin et al., 2021; Farina et al., 2022).

Assumption 3.1. *For all player $i \in \mathcal{N}$, the cost function $c_i(x_i, x_{-i})$ is continuous, differentiable, convex, and ℓ_i-smooth in x_i. Further, c_i has bounded gradient $|\nabla_i c_i(x)| \leq G$ and the gradient $F(x) = [\nabla_i c_i(x)]_{i\in\mathcal{N}}$ is a monotone operator, i.e., $(F(x) - F(y))^\top (x - y) \geq 0$, $\forall x, y$.*

For notational convenience, we denote $L = \sum_{i\in\mathcal{N}} \ell_i$.

A common solution concept in the game is Nash equilibrium, which is a state of dynamic where no player can reduce its cost by unilaterally changing its action. Our aim is to learn a Nash equilibrium $x^* \in \prod_i \mathcal{X}_i$ of the game. Formally, the Nash equilibrium is defined as follows.

Definition 3.1 (Nash equilibrium)**.** *An action $x^* \in \prod_i \mathcal{X}_i$ is a Nash equilibrium if $c_i(x^*) \leq c_i(x_i, x^*_{-i})$, $\forall x_i \in \mathcal{X}_i, x_i \neq x_i^*, i \in \mathcal{N}$.*

When the game satisfies Assumption 3.1, and is with a compact action set, it is known that it must admit at least one Nash equilibrium (Debreu, 1952).

3.1 Examples of Monotone Continuous Games

A wide range of monotone games are captured by Assumption 3.1, and we now present a few classic examples. We include more examples in the appendix.

Example 3.1 (convex-concave game)**.** *Consider a two-player convex-concave game, where the objective function is $c_1(x_1, x_2) = f(x_1, x_2)$, $c_2(x_1, x_2) = -f(x_1, x_2)$. It is immediate that if f is continuous, differentiable, smooth, convex in x_1, concave in x_2, then the game satisfies Assumption 3.1. Examples are rock paper scissors and chicken games.*

Example 3.2 (Cournot competition)**.** *In the Cournot oligopoly model, there is a finite set of N firms, where firm i supplies the market with a quantity $x_i \in [0, C_i]$ of some good and C_i is the firm's production capacity. The good is priced as a decreasing function $P(x_{\mathrm{tot}}) = a - bx_{\mathrm{tot}}$, where $x_{\mathrm{tot}} = \sum_{i=1}^{N} x_i$ is the total number of goods supplied to the market, and $a, b > 0$ are positive constants. The cost of firm i is then given by $c_i(x_i, x_{-i}) = d_i x_i - x_i P(x_{\mathrm{tot}})$, where d_i is the cost of producing one unit of good. This is the associated production cost minus the total revenue from producing x_i units of goods. It is clear that c_i is continuous and differentiable, and Bravo et al. (2018) showed c_i has positive definite and bounded hessian (is convex and smooth).*

Example 3.3 (Splittable routing game)**.** *In a splittable routing game, each player directs a flow, denoted as f_i, from a source to a destination within an undirected graph $G = (V, E)$. Each edge $e \in E$ is linked to a latency function, represented as $\ell_e(f)$, which denotes the latency cost of the flow passing through the edge. The strategies available to player i are the various ways of dividing or "splitting" the flow f_i into distinct paths connecting the source and the destination. With some restrictions on the latency function, the game satisfies Assumption 3.1 (Roughgarden and Schoppmann, 2015).*

3.2 Bandit Feedback and Strongly Uncoupled Dynamic

In this work, we focus on learning under bandit feedback and strongly uncoupled dynamics. The bandit feedback setting restricts each player to only observe the cost function $c_i(x_i, x_{-i})$ with respect to the action taken x_i. The strongly uncoupled learning dynamic (Daskalakis et al., 2011) means players do not have prior knowledge of cost function or the action space of other players and can only keep track of a constant amount of historical information. As the bandit feedback and strongly uncoupled dynamic only require each player to access information of its own, this allows for modular development of the multi-player system, by reusing existing single-agent learning algorithms.

4 Algorithm

Our algorithm builds upon the renowned mirror-descent algorithm. The efficacy of online mirror-descent in solving Nash equilibrium has been demonstrated under full information, and in both linear or strongly monotone games, with extensive investigations into its last-iterate convergence investigated in Cen et al. (2021); Lin et al. (2021); Cai et al. (2023); Duvocelle et al. (2023).

Our algorithm differs from classic online mirror descent approaches by making use of two regularizers: A self-concordant barrier regularizer h to build an efficient Ellipsoidal gradient estimator and contest the bandit feedback; and a regularizer p to accommodate monotone (and not strongly monotone) games. Similar use of two regularizers has also been investigated (Lin et al., 2021). However, their method used the Euclidean norm regularization, which cannot be extended to our setting.

Regularizers Let h be a ν-self-concordant barrier function (Definition 4.1), p be a convex function with $\mu I \preceq \nabla^2 p(x) \preceq \zeta I$, $\zeta > 0, \mu \geq 0$. Let D_p denote the Bregman divergence induced by

p. We choose p such that for any $x_i, x_i' \in \mathcal{X}_i$, $D_p(x_i, x_i') \le C_p < \infty$, and for some $\kappa > 0$, $c_i(x_i, x_{-i}) - \kappa p(x_i)$ to be convex. Notice that when c_i is convex but not linear, we can always find such p when the action set is bounded. Intuitively, this is to interpolate a function p that possesses less curvature than all c_i. We will discuss the modification to the algorithm needed when c_i is linear in Section 5.3.

Definition 4.1. *A function $h : int(\mathcal{X}) \mapsto \mathbb{R}$ is a ν-self concordant barrier for a closed convex set $\mathcal{X} \subseteq \mathbb{R}^n$, where $int(\mathcal{X})$ is an interior of $\mathcal{X}$, if 1) h is three times continuously differentiable; 2) $h(x) \to \infty$ if $x \to \partial\mathcal{X}$, where $\partial\mathcal{X}$ is a boundary of $\mathcal{X}$; 3) for $\forall x \in \mathrm{int}(\mathcal{X})$ and $\forall \lambda \in \mathbb{R}^n$, we have $\left|\nabla^3 h(x)[\lambda,\lambda,\lambda]\right| \le 2\left(\lambda^\top \nabla^2 h(x)\lambda\right)^{3/2}$ and $\left|\nabla h(x)^\top \lambda\right| \le \sqrt{\nu}\left(\lambda^\top \nabla^2 h(x)\lambda\right)^{1/2}$ where $\nabla^3 h(x)\left[\lambda_1,\lambda_2,\lambda_3\right] = \frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} h\left(x + t_1\lambda_1 + t_2\lambda_2 + t_3\lambda_3\right)\Big|_{t_1=t_2=t_3=0}$.*

1. *h is three times continuously differentiable;*
2. *$h(x) \to \infty$ if $x \to \partial\mathcal{X}$, where $\partial\mathcal{X}$ is a boundary of $\mathcal{X}$;*
3. *for $\forall x \in \mathrm{int}(\mathcal{X})$ and $\forall \lambda \in \mathbb{R}^n$, we have $\left|\nabla^3 h(x)[\lambda,\lambda,\lambda]\right| \le 2\left(\lambda^\top \nabla^2 h(x)\lambda\right)^{3/2}$ and $\left|\nabla h(x)^\top \lambda\right| \le \sqrt{\nu}\left(\lambda^\top \nabla^2 h(x)\lambda\right)^{1/2}$ where $\nabla^3 h(x)\left[\lambda_1,\lambda_2,\lambda_3\right] = \frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} h\left(x + t_1\lambda_1 + t_2\lambda_2 + t_3\lambda_3\right)\Big|_{t_1=t_2=t_3=0}$.*

It is shown that any closed convex domain of $\mathbb{R}^d$ has a self-concordant barrier (Lee and Yue, 2021).

Ellipsoidal gradient estimator As our algorithm operates under bandit feedback and strongly uncoupled dynamics, we would need to design a gradient estimator while only using costs for the individual player.

Let $\mathbb{S}^d$, $\mathbb{B}^d$ be the d-dimensional unit sphere and the d-dimensional unit ball, respectively. Our algorithm estimates the gradient using the following ellipsoidal estimator:

$$\hat{g}_i^t = \frac{d}{\delta_t} c_i(\hat{x}^t)(A_i^t)^{-1} z_i^t\,, \quad A_i^t = (\nabla^2 h(x_i^t) + \eta_t(t+1)\nabla^2 p(x_i^t))^{-1/2}\,, \quad \hat{x}_i^t = x_i^t + \delta_t A_i^t z_i^t\,,$$

where z_i^t is uniformly independently sampled from $\mathbb{S}^d$ and $\delta_t, \eta_t \in [0,1]$ are tunable parameters.

One can show that $\hat{g}_i^t$ is an unbiased estimate of the gradient of a smoothed cost function $\hat{c}_i(x^t) = \mathbb{E}_{w_i^t \sim \mathbb{B}^d} \mathbb{E}_{\mathbf{z}_{-i}^t \sim \Pi_{j\neq i}\mathbb{S}^d}\left[c_i\left(x_i^t + A_i^t w_i^t, \hat{x}_{-i}^t\right)\right]$. When p is strongly convex, one can upper bound $\|\nabla_i \hat{c}_i(x) - \nabla_i c_i(x)\|$ by the maximum eigenvalue of A_i^t and it suffices to take $\delta_t = 1$, which recovers the results in Lin et al. (2021). However, when p is convex and not strongly convex, one would need to carefully tune δ_t to control the bias from estimating the smoothed cost function. This ellipsoidal gradient estimator was first introduced by Abernethy et al. (2008) for the case of c_i being linear, and was then extended by Hazan and Levy (2014) to the case of strongly convex costs. In learning for games, the ellipsoidal estimator was used in the case of strongly monotone games (Bravo et al., 2018; Lin et al., 2021).

Based on the ellipsoidal gradient estimator, we present our uncoupled and convergent algorithm for monotone games under bandit feedback.

Implementation Notice that solving Equation (1) is equivalent to solving a convex but potentially non-smooth optimization problem. Certain sets $\mathcal{X} \subseteq \mathbb{R}^d$, including the cases when $\mathcal{X}$ is the strategy space of a normal-form game or an extensive-form game, can be solved by proximal Newton algorithm provably in $O(\log^2(1/\epsilon))$ iterations (Farina et al., 2022). When such guarantees are not required, one could accommodate other optimization methods in solving (1). Our experiment section provides more details.

The choice of p and h is game-dependent. For example, when $c_i(x) = x^2$ and the action set is on the positive half line, we can use the negative log function as our self-concordant barrier function h and take $p = x$.

Algorithm 1: Algorithm

Input: Learning rate η_t, parameter δ_t, regularizer $h(\cdot), p(\cdot)$, constant κ

$x_i^1 = \operatorname{argmin}_{x_i \in \mathcal{X}_i} h(x_i)$

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

    Set $A_i^t = (\nabla^2 h(x_i^t) + \eta_t (t+1) \nabla^2 p(x_i^t))^{-1/2}$

    Play $\hat{x}_i^t = x_i^t + \delta_t A_i^t z_i^t$, receive bandit feedback $c_i(\hat{x}_i, \hat{x}_{-i})$, sample $z_i^t \sim \mathbb{S}^d$

    Update gradient estimator $\hat{g}_i^t = \frac{d}{\delta_t} c_i(\hat{x}^t)(A_i^t)^{-1} z_i^t$

    Update the strategy

$$x_i^{t+1} = \operatorname*{argmin}_{x_i \in \mathcal{X}_i} \left\{ \eta_t \left\langle x_i, \hat{g}_i^t \right\rangle + \eta_t \kappa (t+1) D_p(x_i, x_i^t) + D_h(x_i, x_i^t) \right\} \tag{1}$$

5 No-regret Convergence to Nash Equilibrium

In this section, we present our main results on the last-iterate convergence to the Nash equilibrium. We show that Algorithm 1 converges to the Nash equilibrium in monotone, strongly monotone, and linear games. Such convergence is no-regret, meaning that the individual regret of each player is sublinear.

For notational simplicity, we present the results in a perfect bandit feedback model, where player i observes exactly $c_i(x^t)$. The discussion of noisy bandit feedback, where player i observes $c_i(x^t) + \epsilon_i^t$, with ϵ_i^t be a zero-mean noise, is deferred to the appendix (Theorem D.1).

5.1 Perfect Bandit Feedback

The following theorem describes the last-iterate convergence rate (in expectation) for convex and strongly convex loss under perfect bandit feedback.

Theorem 5.1. *Take* $\eta_t = \begin{cases} \frac{1}{2dt^{3/4}} & \mu = 0 \\ \frac{1}{2dt^{1/2}} & \mu > 0 \end{cases}$, $\delta_t = \begin{cases} \frac{1}{t^{1/4}} & \mu = 0 \\ 1 & \mu > 0 \end{cases}$. *With Algorithm 1, we have*

$$\mathbb{E}\left[\sum_{i \in \mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\right] \leq \begin{cases} O\left(\frac{nd\nu \log(T)}{\kappa T^{1/4}} + \frac{n\zeta dB}{T^{3/4}} + \frac{nBL}{\kappa\sqrt{T}} + \frac{ndC_p}{T^{1/4}} + \frac{nd\log(T)}{\kappa T^{1/4}} + \frac{\sqrt{n}B^2 L \log(T)}{\kappa T^{1/4}}\right) & \mu = 0 \\ O\left(\frac{nd\nu \log(T)}{\kappa\sqrt{T}} + \frac{nd\zeta B}{T} + \frac{nBL}{\kappa\sqrt{T}} + \frac{ndC_p}{\sqrt{T}} + \frac{nd\log(T)}{\kappa\sqrt{T}} + \frac{BL\log(T)}{\mu\kappa\sqrt{T}}\right) & \mu > 0\,, \end{cases}.$$

In the case of the monotone games, Bravo et al. (2018) showed an asymptotic convergence to Nash equilibrium. To the best of our knowledge, Theorem 5.1 is the first result on the last-iterate convergence rate for monotone games. For strongly monotone games, Bravo et al. (2018) first gave a $O(T^{-1/3})$ last-iterate convergence rate, which was later improved to $O(T^{-1/2})$ by Lin et al. (2021).

While we defer the proof to the appendix, we discuss the main ideas for deriving the results. By the update rule, we can obtain the inequality

$$\begin{aligned} &D_h\left(\omega_i, x_i^{t+1}\right) + \eta_t \kappa (t+1) D_p\left(\omega_i, x_i^{t+1}\right) \\ &\leq D_h\left(\omega_i, x_i^t\right) + \eta_t \kappa (t+1) D_p\left(\omega_i, x_i^t\right) + \eta_t \left\langle \nabla_i c_i\left(x^t\right), \omega_i - x_i^t \right\rangle + \eta_t \cdot \text{residual terms}\,, \end{aligned} \tag{2}$$

where ω_i is a fixed point given.

When the game is strongly monotone, we can directly use strongly monotonicity and take p to be the Euclidean norm to obtain a recursive relation similar to $\left\|\omega_i - x_i^{t+1}\right\|_2^2 \leq (1 - \eta_t^2)\left\|\omega_i - x_i^{t+1}\right\|_2^2 +$ residual terms. This amounts to applying this recursion and upper-binding the residual terms individually to obtain a last-iterate convergence. However, when the game is monotone but not strongly monotone, we will need a different approach. Notice that $G = \nabla c_i - \nabla p$ is a monotone operator. Using the property of Bregman divergence, we have $\langle G(x) - G(x'), x' - x \rangle \leq -\sum_{i \in \mathcal{N}} \left(D_p\left(x_i, x_i'\right) + D_p\left(x_i', x_i\right)\right)$.

We then sum the recursive inequality and leverage the combination of two regularizations, which obtains $\eta_T \kappa(T+1) \sum_{i \in \mathcal{N}} D_p\left(\omega_i, x_i^{T+1}\right) \leq \sum_{i \in \mathcal{N}} D_h\left(\omega_i, x_i^1\right) + \kappa \sum_{i \in \mathcal{N}} D_p\left(\omega_i, x_i^1\right) + \sum_{t=1}^T \sum_{i \in \mathcal{N}} \eta_t \left\langle \nabla_i c_i(\omega), \omega_i - x_i^t \right\rangle + \sum_{t=1}^T \sum_{i \in \mathcal{N}} \eta_t \left\langle \hat{g}_i^t - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t \right\rangle + \sum_{t=1}^T \sum_{i \in \mathcal{N}} \eta_t \left\langle \hat{g}_i^t, x_i^t - x_i^{t+1} \right\rangle$ Now it suffices to properly choose a fixed point ω_i such that both the first term $\sum_{i \in \mathcal{N}} D_h\left(\omega_i, x_i^1\right)$ and the third term $\sum_{t=1}^T \sum_{i \in \mathcal{N}} \eta_t \left\langle \nabla_i c_i(\omega), \omega_i - x_i^t \right\rangle$ are bounded. When ω_i is the Nash equilibrium x_i^*, the third term can be upper bounded trivially using the monotonicity of c_i, while it does not imply a bounded first term. Therefore, we set $\omega_i = x_i^*$ when the first term can be bounded. Otherwise, we set it to a close enough point to x_i^*, such that the first term can be bounded and the third term is bounded through a more careful calculation.

High probability result In the case of a strongly monotone game, our results show that the $O(T^{-1/4})$ last-iterate convergence rate holds a high probability. This is the first high-probability result for last-iterate convergence in strongly monotone games.

Theorem 5.2. *With a probability of at least* $1 - \log(T)\delta$, $\delta \leq e^{-1}$, *and with Algorithm 1, we have* $\sum_{i \in \mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right) \leq O\left(\frac{nd\nu \log(T)}{\sqrt{T}} + \frac{nd\zeta B}{T} + \frac{nBL}{\sqrt{T}} + \frac{ndC_p}{\sqrt{T}} + \frac{nd\log(T)}{\sqrt{T}} + \frac{dBL\log(T)}{\mu\sqrt{T}} + \frac{nBd^2\log^2(1/\delta)\log(T)}{\min\{\sqrt{\mu},\mu\}\sqrt{T}}\right)$.

5.2 Individual Low Regret

Beyond the fast convergence to Nash equilibrium, our algorithm also ensures each player with a sublinear regret when playing against other players. The sublinear regret convergence is a desirable property as the players could be self-interested in general, and want to ensure their return even when other players are not adhering to the protocol. The low regret property remains true for players that are potentially adversarial, despite the convergence to Nash equilibrium no longer holds in that case.

For player i, and a sequence of actions $\{\hat{x}_i^t\}_{t=1}^T$, define the individual regret as the cumulative expected difference between the costs received and the cost of playing the hindsight optimal action. That is, $\sum_{t=1}^T \mathbb{E}\left[c_i\left(\hat{x}_i^t, x_{-i}^t\right) - c_i\left(\omega_i, x_{-i}^t\right)\right]$, where $\{x_{-i}^t\}_{t=1}^T$ is a fixed sequence of actions of other players. The following theorem shows a guarantee of the individual regret of each player.

Theorem 5.3. *Take* $\eta_t = \begin{cases} \frac{1}{2dt^{3/4}} & \mu = 0 \\ \frac{1}{2dt^{1/2}} & \mu > 0 \end{cases}$, $\delta_t = \begin{cases} \frac{1}{t^{1/4}} & \mu = 0 \\ 1 & \mu > 0 \end{cases}$. *For a fixed* $\omega_i \in \mathcal{X}_i$, *a fixed sequence of* $\{x_{-i}^t\}_{t=1}^T$, *and with Algorithm 1, we have*

$$\sum_{t=1}^T \mathbb{E}\left[c_i\left(\hat{x}_i^t, x_{-i}^t\right) - c_i\left(\omega_i, x_{-i}^t\right)\right] = \begin{cases} O\left(\nu d T^{3/4}\log(T) + G\sqrt{T} + \ell_i\sqrt{n}BT^{3/4}\right) & \mu = 0 \\ O\left(\nu d\sqrt{T}\log(T) + G\sqrt{T} + \frac{nB\ell_i\sqrt{T}}{\mu}\right) & \mu > 0 \end{cases}.$$

Our result matches the $\sqrt{T}$ regret bound for strongly monotone games (Lin et al., 2021), but applies to monotone games as well.

Implication on social welfare By designing the algorithm to be no-regret, we can also show that the social welfare attained by the algorithm also converges to the optimal value.

The social welfare for a joint action x is defined as $\mathrm{SW}(x) = \sum_{i \in \mathcal{N}} c_i(x)$. We let $\mathrm{OPT} = \min_x \mathrm{SW}(x)$ to denote the optimal social welfare.

Definition 5.1 (Roughgarden 2015; Syrgkanis et al. 2015)**.** *A game is* (C_1, C_2)*-smooth,* $C_1 > 0$, $C_2 < 1$, *if there exists a strategy* x', *such that for any* $x \in \mathcal{N}$, $\sum_{i \in \mathcal{N}} c_i(x_i', x_{-i}) \leq C_1\mathrm{OPT} + C_2\mathrm{SW}(x)$.

We have the following proposition which shows that the social welfare converges to optimal welfare on average.

Proposition 5.1. *With* $\eta_t = \frac{1}{2dt^{3/4}}, \delta_t = \frac{1}{t^{1/4}}$, *and suppose every player employ Algorithm 1, we have* $\frac{1}{T}\sum_{t=1}^T \mathbb{E}\left[\mathrm{SW}(\hat{x})\right] = O\left(\frac{C_1\mathrm{OPT}}{(1-C_2)} + \frac{n\nu d\log(T)}{(1-C_2)T^{1/4}} + \frac{\sqrt{n}B\sum_{i \in \mathcal{N}}\ell_i}{(1-C_2)T^{1/4}}\right)$.

5.3 Special Case: Linear Cost Function

When c_i is linear, there does not exist a p that is convex while making $c_i - \kappa p$ convex. Algorithm 1 therefore does not apply to the linear case. This coincides with our intuition that the landscape c_i does not provide enough curvature information for the algorithm to utilize.

To extend the algorithm to the linear case, we modify line 6 of Algorithm 1 as $x_i^{t+1} = \operatorname{argmin}_{x_i \in \mathcal{X}_i} \{\eta_t \langle x_i, \hat{g}_i^t \rangle + \eta_t \tau (t+1) D_p(x_i, x_i^t) + D_h(x_i, x_i^t)\}$. The idea is to first show the convergence of x^T to a game with the cost $c_i(x) + \tau p(x)$. With this regularized game, we choose p to be a strongly convex function and measure the convergence in terms of the gap function $\langle c_i(x), x_i - x^* \rangle$. By carefully controlling τ, we obtain the following result.

Theorem 5.4. *With* $\eta_t = \frac{1}{2d\sqrt{t}}$, $\tau = \frac{1}{T^{1/6}}$, $G_p = \sup_x \|\nabla p(x)\|$ *and Algorithm 1, we have* $\mathbb{E}\left[\sum_{i \in \mathcal{N}} \left\langle \nabla_i c_i\left(x^T\right), x_i^T - x_i^* \right\rangle\right] \leq \tilde{O}\left(\frac{BG_p + \sqrt{d(BL+G)(n\nu + nBL + nd^2)}}{T^{1/6}} + \frac{\sqrt{dBL(BL+G)}}{\sqrt{\mu}T^{1/6}} + \frac{\sqrt{dnC_p(BL+G)}}{\sqrt{\mu}T^{1/4}}\right)$.

Similar regularization techniques have been used in the analysis of the zero-sum game (Cen et al., 2021; Cai et al., 2023). Our result matches the last-iterate convergence for zero-sum matrix game (Cai et al., 2023), which is a class of games with linear cost functions. However, our result is more general as it applies to multi-player linear games with convex and compact action sets (while previous works only apply to a simplex action set). It remains open to how games with linear cost functions could be effectively learned and whether the convergence rate could be improved.

6 Application to Time-varying Game

In this section, we further apply Algorithm 1 to games that evolve over time. A time-varying game $\mathcal{G}_t$ is a game where the cost function $c_i^t(\cdot)$, $i \in \mathcal{N}$ depends on t. The game $\mathcal{G}_t$ is not revealed to the players before choosing their actions x_t. We assume that $\mathcal{G}_t$ satisfies Assumption 3.1 for every t.

Such evolving games have applications in Kelly's auction and power control, where the cost function may change as time-dependent values change, such as channel gains. While the changes of $\mathcal{G}_t$ can be random, we discuss two cases here, 1) when $\mathcal{G}_t$ converges to a static game $\mathcal{G}$ in $o(T)$ time, and 2) when the variation path of the Nash equilibrium, $\sum_{t=1}^T \|x_i^{t+1,*} - x_i^{t,*}\|$ is bounded in $o(T)$.

Converging monotone game Let $\mathcal{G}_t$ denote the game formed by the costs $\{c_i^t(\cdot)\}_{i \in \mathcal{N}}$, and $\mathcal{G}$ be the game formed by the costs $\{c_i(\cdot)\}_{i \in \mathcal{N}}$. Suppose $\mathcal{G}_t$ converges to $\mathcal{G}$, and let x^* be the set of Nash equilibrium of the game $\mathcal{G}$. The cost function c_i^t converges to some cost function c_i in $o(T)$ time. The following theorem shows the last iterate convergence to x^*.

Theorem 6.1. *With* $\sum_{t=1}^T \sum_{i \in \mathcal{N}} \max_x \|\nabla_i c_i(x) - \nabla_i c_i^t(x)\|_2 = T^\alpha$, *take* $\eta_t = \frac{1}{2dt^{3/4}}$, $\delta_t = \frac{1}{t^{1/4}}$, *and under Algorithm 1, we have* $\mathbb{E}\left[\sum_{i \in \mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\right] \leq O\left(\frac{nd\nu \log(T)}{\kappa T^{1/4}} + \frac{n\zeta dB}{T^{3/4}} + \frac{nBL}{\kappa\sqrt{T}} + \frac{ndC_p}{T^{1/4}} + \frac{nd\log(T)}{\kappa T^{1/4}} + \frac{\sqrt{n}B^2 L \log(T)}{\kappa T^{1/4}} + \frac{B}{T^{1/4-\alpha}}\right)$.

For monotone games, Duvocelle et al. (2023) showed an asymptotic last-iterate convergence rate. To the best of our knowledge, Theorem 6.1 is the first last-iterate convergence rate for the class of converging monotone game.

Evolving game and equilibrium tracking We now discuss the case where $\mathcal{G}_t$ does not necessarily converge to a game $\mathcal{G}$, but the cumulative changes of the equilibrium are bounded. We use the variation path $V_i(T) = \sum_{t \in [T]} \left\|x_i^{t+1,*} - x_i^{t,*}\right\|$ to track the cumulative changes of equilibrium. In this setting, the last-iterate convergence is not applicable, and the convergence is measured in terms of the average gap. Because of this, the algorithm is slightly modified and updates with $x_i^{t+1} = \operatorname{argmin}_{x_i \in \mathcal{X}_i} \{\eta_t \langle x_i, \hat{g}_i^t \rangle + D_h(x_i, x_i^t)\}$.

Theorem 6.2. *Assume* $V_i(T) \leq T^\varphi$, $\varphi \in [0,1]$. *Take* $\eta_t = \frac{1}{2dt^{\frac{(1-\varphi)}{3}}}$, $\delta_t = \frac{1}{t^{1/2}}$, *and under Algorithm 1, we have* $\frac{1}{T}\sum_{t=1}^T \sum_{i \in \mathcal{N}} \left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - x_i^{t,*} \right\rangle = \tilde{O}\left(\frac{n\nu d + Ln^{3/2}B^2 + nG}{T^{\frac{2(1-\varphi)}{3}}} + \frac{n}{T^{\frac{9}{8} - \frac{(4\varphi+5)^2}{72}}}\right)$.

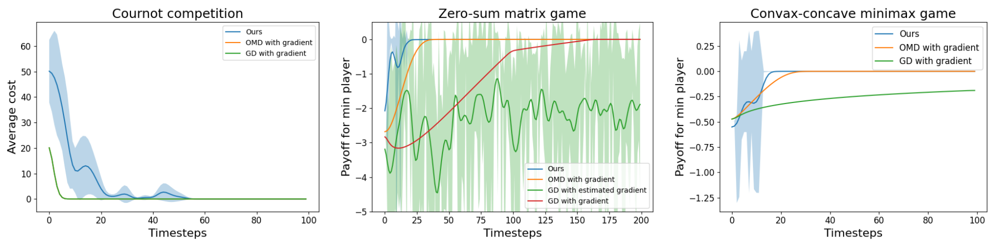

Figure 1: Experiment on Cournot competition, zero-sum two-player minimax game, and convex-concave game. In Cournot competition, the curves of OMD and GD overlap with each other.

In the case of a strongly monotone game, Duvocelle et al. (2023) gave a result of $T^{\varphi/5-1/5}$ and Yan et al. (2023) gave a result of $T^{\varphi/3-2/3}$. In comparison, Theorem 6.2 extends the study to monotone games, and improves the result to $O\left(\max\left\{T^{2\varphi/3-2/3}, T^{(4\varphi+5)^2/72-9/8}\right\}\right)$.

7 Experiment

In this section, we provide a numerical evaluation of our proposed algorithm in three static games. We repeat each experiment with 5 different random seeds. We ran all experiments with a 10-core CPU, with 32 GB memory. We set $\eta_t = \frac{1}{\sqrt{t+1}}$, and $\delta_t = 0.001$.

We present the results of the following example games described below. More results with other parameters can be found in the Appendix K.

Cournot competition In this Cournot duopoly model, n players compete with constant marginal costs, each having individual constant price intercepts and slopes. We model the game with 5 players, where the margin cost is 40, price intercept is $[30, 50, 30, 50, 30]$, and the price slope is $[50, 30, 50, 30, 50]$.

Zero-sum matrix game In this zero-sum matrix game, the two players aim to solve the bilinear problem $\min_x \max_y x^\top Ay$. We set this matrix A to be $[[1, 2], [3, 4]]$.

monotone zero-sum matrix game In this monotone version of the zero-sum matrix game, we regularize the game by the regularizer $x^2 + y^2$.

Algorithm 1 is evaluated against two baseline methods: online mirror descent and gradient descent, with exact gradient, or estimated gradient (bandit feedback). We set the learning rate η to be 0.01 in both zero-sum matrix games and monotone zero-sum matrix games and 0.09 in Cournot competition.

Figure 1 summarizes our experimental findings, where our algorithm attains comparable performance to online mirror descent and gradient descent with full information. This demonstrates the efficacy of our algorithm. We also compare our algorithm to gradient descent with an estimated gradient, using the same ellipsoidal gradient estimator, for a more fair comparison. However, apart from the zero-sum matrix game, we find the baseline algorithm performs too poorly to be compared.

8 Conclusion

In this work, we present a mirror-descent-based algorithm that converges in $O(T^{-1/4})$ in general monotone and smooth games under bandit feedback and strongly uncoupled dynamics. Our algorithm is no-regret, and the result can be improved to $O(T^{-1/2})$ in the case of strongly-monotone games. To our best knowledge, this is the first uncoupled and convergent algorithm in general monotone games under bandit feedback. We then extend our results to time-varying monotone games and present the first result of $O(T^{-1/4})$ for converging games and the improved result of $O\left(\max\{T^{2\varphi/3-2/3}, T^{(4\varphi+5)^2/72-9/8}\}\right)$ for equilibrium tracking. We further verify the effectiveness of our algorithm with empirical evaluations.

References

Abernethy, J., Hazan, E., and Rakhlin, A. (2008). Competing in the dark: An efficient algorithm for bandit linear optimization. In *Conference on Learning Theory*.

Bartlett, P., Dani, V., Hayes, T., Kakade, S., Rakhlin, A., and Tewari, A. (2008). High-probability regret bounds for bandit online linear optimization. In *Conference on Learning Theory*.

Bauschke, H. H., Bolte, J., and Teboulle, M. (2017). A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications. *Mathematics of Operations Research*, 42(2):330–348.

Bervoets, S., Bravo, M., and Faure, M. (2020). Learning with minimal information in continuous games. *Theoretical Economics*, 15(4):1471–1508.

Bravo, M., Leslie, D., and Mertikopoulos, P. (2018). Bandit learning in concave n-person games. *Advances in Neural Information Processing Systems*.

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

Cai, Y., Luo, H., Wei, C.-Y., and Zheng, W. (2023). Uncoupled and convergent learning in two-player zero-sum markov games. *arXiv preprint arXiv:2303.02738*.

Cai, Y., Oikonomou, A., and Zheng, W. (2022). Finite-time last-iterate convergence for learning in multi-player games. *Advances in Neural Information Processing Systems*.

Cai, Y. and Zheng, W. (2023). Doubly optimal no-regret learning in monotone games. In *International Conference on Machine Learning*.

Cen, S., Wei, Y., and Chi, Y. (2021). Fast policy extragradient methods for competitive games with entropy regularization. *Advances in Neural Information Processing Systems*.

Chen, P.-A. and Lu, C.-J. (2016). Generalized mirror descents in congestion games. *Artificial Intelligence*, 241:217–243.

Daskalakis, C., Deckelbaum, A., and Kim, A. (2011). Near-optimal no-regret algorithms for zero-sum games. In *Symposium on Discrete Algorithms*.

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

Drusvyatskiy, D., Fazel, M., and Ratliff, L. J. (2022). Improved rates for derivative free gradient play in strongly monotone games. In *Conference on Decision and Control*. IEEE.

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

Even-Dar, E., Mansour, Y., and Nadav, U. (2009). On the convergence of regret minimization dynamics in concave games. In *Symposium on Theory of computing*.

Farina, G., Anagnostides, I., Luo, H., Lee, C.-W., Kroer, C., and Sandholm, T. (2022). Near-optimal no-regret learning dynamics for general convex games. *Advances in Neural Information Processing Systems*.

Hazan, E. and Levy, K. (2014). Bandit convex optimization: Towards tight bounds. *Advances in Neural Information Processing Systems*.

Jordan, M. I., Lin, T., and Zhou, Z. (2023). Adaptive, doubly optimal no-regret learning in strongly monotone and exp-concave games with gradient feedback. arXiv:2310.14085.

Koller, D., Megiddo, N., and Von Stengel, B. (1996). Efficient computation of equilibria for extensive two-person games. *Games and Economic Behavior*, 14(2):247–259.

Lee, Y. T. and Yue, M.-C. (2021). Universal barrier is n-self-concordant. *Mathematics of Operations Research*, 46(3):1129–1148.

Liang, T. and Stokes, J. (2019). Interaction matters: A note on non-asymptotic local convergence of generative adversarial networks. In *International Conference on Artificial Intelligence and Statistics*, pages 907–915.

Lin, T., Zhou, Z., Ba, W., and Zhang, J. (2021). Doubly optimal no-regret online learning in strongly monotone games with bandit feedback. *arXiv preprint arXiv:2112.02856*.

Mertikopoulos, P., Papadimitriou, C., and Piliouras, G. (2018). Cycles in adversarial regularized learning. In *Proceedings of the twenty-ninth annual ACM-SIAM symposium on discrete algorithms*.

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

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

Roughgarden, T. (2015). Intrinsic robustness of the price of anarchy. *Journal of the ACM (JACM)*, 62(5):1–42.

Roughgarden, T. and Schoppmann, F. (2015). Local smoothness and the price of anarchy in splittable congestion games. *Journal of Economic Theory*, 156:317–342.

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

Tatarenko, T. and Kamgarpour, M. (2019). Learning nash equilibria in monotone games. In *IEEE 58th Conference on Decision and Control (CDC)*. IEEE.

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

Yan, Y.-H., Zhao, P., and Zhou, Z.-H. (2023). Fast rates in time-varying strongly monotone games. In *International Conference on Machine Learning*. PMLR.

Zhou, Z., Mertikopoulos, P., Bambos, N., Boyd, S. P., and Glynn, P. W. (2020). On the convergence of mirror descent beyond stochastic convex programming. *SIAM Journal on Optimization*, 30(1):687–716.

A More Example Games

Example A.1 (Extensive form game (EFG)). *EFGs are games on a directed tree. At terminal nodes denoted as $z \in \mathcal{Z}$, each player $i \in \mathcal{N}$ incurs a cost $c_i(z)$ based on a function $c_i : \mathcal{Z} \to \mathbb{R}$. The action set of each player, $\mathcal{X}_i$, is represented through a sequence-form polytope known as $\mathcal{X}_i$ Koller et al. (1996). Considering the probability $p(z)$ of reaching a terminal node $z \in \mathcal{Z}$, the cost for player i is expressed as $c_i(x) := \sum_{z \in \mathcal{Z}} p(z) c_i(z) \prod_{j \in \mathcal{N}} x_j \left[\sigma_{j,z}\right]$. Here, $x = (x_1, \ldots, x_n) \in \prod_{j \in \mathcal{N}} \mathcal{X}_j$ signifies the joint strategy profile, and $x_j \left[\sigma j, z\right]$ denotes the probability mass assigned to the last sequence $\sigma_{j,z}$ encountered by player j before reaching z. The smoothness and concavity of utilities directly arise from multilinearity.*

Example A.2 (monotone potential game). *A game is called a potential game if there exists a potential function $\Phi : \mathcal{X} \to \mathbb{R}$, such that, $c_i(x_i, x_{-i}) - c_i(x_i', x_{-i}) = \Phi(x_i, x_{-i}) - \Phi(x_i', x_{-i})$, for all $i \in \mathcal{N}$. If Φ is continuous, differentiable, smooth, and monotone in x_i, then the game satisfies Assumption 3.1. For example, a non-atomic congestion game satisfies Assumption 3.1, as shown in Proposition 1 and 2 of Chen and Lu (2016).*

B Proof of Theorem 5.1

Theorem 5.1. *Take* $\eta_t = \begin{cases} \frac{1}{2dt^{3/4}} & \mu = 0 \\ \frac{1}{2dt^{1/2}} & \mu > 0 \end{cases}$, $\delta_t = \begin{cases} \frac{1}{t^{1/4}} & \mu = 0 \\ 1 & \mu > 0 \end{cases}$. *With Algorithm 1, we have*

$$\mathbb{E}\left[\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\right]$$
$$\leq \begin{cases} O\left(\frac{nd\nu\log(T)}{\kappa T^{1/4}} + \frac{n\zeta dB}{T^{3/4}} + \frac{nBL}{\kappa\sqrt{T}} + \frac{ndC_p}{T^{1/4}} + \frac{nd\log(T)}{\kappa T^{1/4}} + \frac{\sqrt{n}B^2L\log(T)}{\kappa T^{1/4}}\right) & \mu = 0 \\ O\left(\frac{nd\nu\log(T)}{\kappa\sqrt{T}} + \frac{nd\zeta B}{T} + \frac{nBL}{\kappa\sqrt{T}} + \frac{ndC_p}{\sqrt{T}} + \frac{nd\log(T)}{\kappa\sqrt{T}} + \frac{BL\log(T)}{\mu\kappa\sqrt{T}}\right) & \mu > 0\,, \end{cases}.$$

Proof. We now upper bound the terms in Lemma J.1.

When $\mu = 0$, taking expectation conditioned on x^t, we have $\mathbb{E}\left[\left\|A_i^t\hat{g}_i^t\right\|^2 \mid x^t\right] = \frac{d^2}{\delta_t^2}\mathbb{E}\left[c_i(\hat{x}^t)^2\|z_i^t\|^2 \mid x^t\right] \leq \frac{d^2}{\delta_t^2}$. By Lemma J.2, and the choice $\eta_t = \frac{1}{2d\sqrt{t}}$, we have

$$\sum_{t=1}^{T}\eta_t\sum_{i\in\mathcal{N}}\mathbb{E}\left[\left\langle\hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle\right] \leq \sum_{t=1}^{T}\eta_t^2\sum_{i\in\mathcal{N}}\mathbb{E}\left[\left\|A_i^t\hat{g}_i^t\right\|^2\right] \leq nd^2\sum_{t=1}^{T}\frac{\eta_t^2}{\delta_t^2}.$$

By the definition of $\hat{c}_i$,

$$\begin{aligned}
&\sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\left\langle\hat{g}_i^t - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right] \\
&= \sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\left\langle\nabla_i\hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right] \\
&= \sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left\langle\nabla_i c_i\left(x_i^t + \delta_t A_i^t w_i, \hat{x}_{-i}^t\right) - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right] \\
&\leq B\sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left\|\nabla_i c_i\left(x_i^t + \delta_t A_i^t w_i, \hat{x}_{-i}^t\right) - \nabla_i c_i\left(x^t\right)\right\| \mid x^t\right]
\end{aligned}$$

By the smoothness of c_i,

$$\begin{aligned}
&\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left[\left\|\nabla_i c_i\left(x_i^t + \delta_t A_i^t w_i, \hat{x}_{-i}^t\right) - \nabla_i c_i\left(x^t\right)\right\|\right] \\
&\leq \ell_i\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left[\sqrt{\delta_t^2\left\|A_i w_i\right\|^2 + \delta_t^2\sum_{j\neq i}\left\|A_j z_j\right\|^2}\right].
\end{aligned}$$

Since p is monotone, $\nabla^2 p(x)$ is positive semi-definite, and $A_i^t \preceq (\nabla^2 h(x_i))^{-1/2}$. For $\bar{x}_i^t = x_i^t + A_i^t w_i^t$. Define $\|v\|_x = \sqrt{v^\top\nabla^2 h(x)v}$, we have $\|\bar{x}_i^t - x_i^t\|_{x_i} \leq \|\omega_i^t\| \leq 1$, and $\bar{x}_i^t \in W(x_i^t)$, where $W(x_i) = \{x_i' \in \mathbb{R}^d, \|x_i' - x_i\|_{x_i} \leq 1\}$ is the Dikin ellipsoid. Since $W(x_i) \subseteq \mathcal{X}_i, \forall x_i \in int(\mathcal{X}_i)$, we can upper bound $\|A_i w_i\|^2$ by B^2, the diameter of the set $\mathcal{X}_i$. Hence $\|\nabla_i\hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right)\| \leq \ell_i\delta_t\sqrt{n}B$. By Lemma J.5

$$\begin{aligned}
\sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\left\langle\hat{g}_i^t - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right] &= \sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\left\langle\nabla_i\hat{c}_i\left(x^t\right) - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right] \\
&\leq \sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\left\|\nabla_i\hat{c}_i\left(x^t\right) - \nabla_i c_i\left(x^t\right)\right\|\left\|\omega_i - x_i^t\right\| \mid x^t\right] \\
&\leq \sqrt{n}B^2\sum_{i\in\mathcal{N}}\ell_i\sum_{t=1}^{T}\eta_t\delta_t.
\end{aligned}$$

When $\mu > 0$, we set $\delta = 1$. Then, taking expectation conditioned on x^t, we have $\mathbb{E}\left[\|A_i^t \hat{g}_i^t\|^2 \mid x^t\right] = d^2 \mathbb{E}\left[c_i(\hat{x}^t)^2 \|z_i^t\|^2 \mid x^t\right] \leq d^2$. By Lemma J.2, and the choice $\eta_t = \frac{1}{2d\sqrt{t}}$, we have

$$\sum_{t=1}^T \eta_t \sum_{i \in \mathcal{N}} \mathbb{E}\left[\langle \hat{g}_i^t, x_i^t - x_i^{t+1} \rangle\right] \leq \sum_{t=1}^T \eta_t^2 \sum_{i \in \mathcal{N}} \mathbb{E}\left[\|A_i^t \hat{g}_i^t\|^2\right] \leq nd^2 \sum_{t=1}^T \eta_t^2 .$$

By Lemma J.5, for any $\omega_i \in \mathcal{X}_i$, we have

$$\begin{aligned}
\sum_{i \in \mathcal{N}} \sum_{t=1}^T \eta_t \mathbb{E}\left[\langle \hat{g}_i^t - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t \rangle \mid x^t\right] &= \sum_{i \in \mathcal{N}} \sum_{t=1}^T \eta_t \mathbb{E}\left[\langle \nabla_i \hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t \rangle \mid x^t\right] \\
&\leq \sum_{i \in \mathcal{N}} B\ell_i \sum_{t=1}^T \eta_t \mathbb{E}\left[\sum_{j \in \mathcal{N}} \left(\sigma_{\max}\left(A_j^t\right)^2\right) \mid x^t\right] \\
&\leq \sum_{i \in \mathcal{N}} B\ell_i \sum_{t=1}^T \frac{1}{\mu(t+1)} \\
&\leq \frac{B \sum_{i \in \mathcal{N}} \ell_i}{\mu} \sum_{t=1}^T \frac{1}{(t+1)} .
\end{aligned}$$

where the third inequality is by $\nabla^2 h(x)$ being positive definite, and $\nabla^2 p(x) \geq \mu I$.

Let $L = \sum_{i \in \mathcal{N}} \ell_i$. When $\mu = 0$, combing and rearranging the terms, we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i \in \mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\right] \\
&\leq O\left(\frac{n\nu \log(T)}{\kappa \eta_T T} + \frac{n\zeta B}{\eta_T T^{3/2}} + \frac{nBL}{\kappa\sqrt{T}} + \frac{n}{\kappa\sqrt{T}} + \frac{nC_p}{\eta_T T} + \frac{nd^2}{\kappa \eta_T T} \sum_{t=1}^T \frac{\eta_t^2}{\delta_t^2} + \frac{\sqrt{n} B^2 L \sum_{t=1}^T \eta_t \delta_t}{\kappa \eta_T T}\right) .
\end{aligned}$$

Take $\eta_t = \frac{1}{2dt^{3/4}}$, $\delta_t = \frac{1}{t^{1/4}}$, then $\sum_{t=1}^T \frac{\eta_t^2}{\delta_t^2} = O\left(\sum_{t=1}^T \frac{1}{t}\right) = O(\log(T))$, and $\sum_{t=1}^T \eta_t \delta_t = O\left(\sum_{t=1}^T \frac{1}{t}\right) = O(\log(T))$. Hence, we have

$$\mathbb{E}\left[\sum_{i \in \mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\right] \leq O\left(\frac{nd\nu \log(T)}{\kappa T^{1/4}} + \frac{n\zeta dB}{T^{3/4}} + \frac{nBL}{\kappa\sqrt{T}} + \frac{ndC_p}{T^{1/4}} + \frac{nd\log(T)}{\kappa T^{1/4}} + \frac{\sqrt{n} B^2 L \log(T)}{\kappa T^{1/4}}\right) .$$

When $\mu > 0$, combing and rearranging the terms, we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i \in \mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\right] \\
&\leq O\left(\frac{n\nu \log(T)}{\kappa \eta_T T} + \frac{n\zeta B}{\eta_T T^{3/2}} + \frac{nBL}{\kappa\sqrt{T}} + \frac{n}{\sqrt{T}} + \frac{nC_p}{\eta_T T} + \frac{nd^2}{\kappa \eta_T T} \sum_{t=1}^T \eta_t^2 + \frac{BL\log(T)}{\mu\kappa\eta_T T}\right) .
\end{aligned}$$

Take $\eta_t = \frac{1}{2dt^{1/2}}$, we have

$$\mathbb{E}\left[\sum_{i \in \mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\right] \leq O\left(\frac{nd\nu \log(T)}{\kappa\sqrt{T}} + \frac{nd\zeta B}{T} + \frac{nBL}{\kappa\sqrt{T}} + \frac{ndC_p}{\sqrt{T}} + \frac{nd\log(T)}{\kappa\sqrt{T}} + \frac{BL\log(T)}{\mu\kappa\sqrt{T}}\right) .$$

$\square$

C Proof of Theorem 5.3

Theorem 5.3. *Take* $\eta_t = \begin{cases} \frac{1}{2dt^{3/4}} & \mu = 0 \\ \frac{1}{2dt^{1/2}} & \mu > 0 \end{cases}$, $\delta_t = \begin{cases} \frac{1}{t^{1/4}} & \mu = 0 \\ 1 & \mu > 0 \end{cases}$. *For a fixed* $\omega_i \in \mathcal{X}_i$*, a fixed sequence of* $\{x_{-i}^t\}_{t=1}^T$*, and with Algorithm 1, we have*

$$\sum_{t=1}^T \mathbb{E}\left[c_i\left(\hat{x}_i^t, x_{-i}^t\right) - c_i\left(\omega_i, x_{-i}^t\right)\right] = \begin{cases} O\left(\nu d T^{3/4}\log(T) + G\sqrt{T} + \ell_i\sqrt{n}BT^{3/4}\right) & \mu = 0 \\ O\left(\nu d\sqrt{T}\log(T) + G\sqrt{T} + \frac{nB\ell_i\sqrt{T}}{\mu}\right) & \mu > 0 \end{cases}.$$

Proof. Define the smoothed version of c_i as $\hat{c}_i(x) = \mathbb{E}_{w_i \sim \mathbb{B}^d}\left[c_i\left(x_i + \delta A_i w_i, x_{-i}\right)\right]$. Then, we decompose as

$$\begin{aligned}\sum_{t=1}^T c_i\left(\hat{x}_i^t, x_{-i}^t\right) - c_i\left(\omega_i, x_{-i}^t\right) = &\sum_{t=1}^T\left(\hat{c}_i\left(x_i^t, x_{-i}^t\right) - \hat{c}_i\left(\omega_i, x_{-i}^t\right)\right) + \sum_{t=1}^T\left(c_i\left(x_i^t, x_{-i}^t\right) - \hat{c}_i\left(x_i^t, x_{-i}^t\right)\right) \\ &+ \sum_{t=1}^T\left(\hat{c}_i\left(\omega_i, x_{-i}^t\right) - c_i\left(\omega_i, x_{-i}^t\right)\right) + \sum_{t=1}^T\left(c_i\left(\hat{x}_i^t, x_{-i}^t\right) - c_i\left(x_i^t, x_{-i}^t\right)\right).\end{aligned}$$

For the first term, recall that by the update rule, we have,

$$\begin{aligned}&D_h\left(\omega_i, x_i^{t+1}\right) + \eta_t\kappa(t+1)D_p\left(\omega_i, x_i^{t+1}\right) \\ &= D_h\left(\omega_i, x_i^t\right) + \eta_t\kappa(t+1)D_p\left(\omega_i, x_i^t\right) + \eta_t\left\langle\nabla\hat{c}_i\left(x^t\right), \omega_i - x_i^t\right\rangle + \eta_t\left\langle\hat{g}_i^t - \nabla\hat{c}_i\left(x^t\right), \omega_i - x_i^t\right\rangle \\ &\quad + \eta_t\left\langle\hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle \\ &= D_h\left(\omega_i, x_i^t\right) + \eta_t\kappa(t+1)D_p\left(\omega_i, x_i^t\right) + \eta_t\left\langle\nabla\hat{c}_i\left(x^t\right) - \kappa\nabla p(x_i^t), \omega_i - x_i^t\right\rangle + \eta_t\left\langle\hat{g}_i^t - \nabla\hat{c}_i\left(x^t\right) + \kappa\nabla p(x_i^t), \omega_i - x_i^t\right\rangle \\ &\quad + \eta_t\left\langle\hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle.\end{aligned}$$

By Lemma J.5, for any $\omega_i \in \mathcal{X}_i$, we have

$$\begin{aligned}\mathbb{E}\left[\left\langle\hat{g}_i^t - \nabla\hat{c}_i\left(x^t\right) + \kappa\nabla p(x_i^t), \omega_i - x_i^t\right\rangle \mid x^t\right] &= \mathbb{E}\left[\left\langle\nabla_i\hat{c}_i(x^t) - \nabla_i\hat{c}_i\left(x^t\right) + \kappa\nabla p(x_i^t), \omega_i - x_i^t\right\rangle \mid x^t\right] \\ &= \mathbb{E}\left[\kappa\left\langle\nabla p(x_i^t), \omega_i - x_i^t\right\rangle \mid x^t\right] \\ &= \mathbb{E}\left[\kappa p(\omega_i) - \kappa p(x_i^t) - \kappa D_p(\omega_i, x_i^t) \mid x^t\right],\end{aligned}$$

where the last equality follows from the definition of Bregman divergence.

Therefore,

$$\begin{aligned}&\mathbb{E}\left[D_h\left(\omega_i, x_i^{t+1}\right) + \eta_t\kappa(t+1)D_p\left(\omega_i, x_i^{t+1}\right)\right] \\ &= \mathbb{E}\left[D_h\left(\omega_i, x_i^t\right) + \eta_t\kappa t D_p\left(\omega_i, x_i^t\right) + \eta_t\left\langle\nabla\hat{c}_i\left(x^t\right) - \kappa\nabla p(x_i^t), \omega_i - x_i^t\right\rangle\right] + \eta_t\mathbb{E}\left[\kappa p(\omega_i) - \kappa p(x_i^t)\right] \\ &\quad + \mathbb{E}\left[\eta_t\left\langle\hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle\right].\end{aligned}$$

By the monotoneity of $\hat{c}_i\left(x^t\right) - \kappa p(x_i^t)$, we have

$$\left\langle\nabla\hat{c}_i\left(x^t\right) - \kappa\nabla p(x_i^t), \omega_i - x_i^t\right\rangle \leq \left(\hat{c}_i\left(\omega_i, x_{-i}^t\right) - \kappa p(\omega_i)\right) - \left(\hat{c}_i\left(x_i^t, x_{-i}^t\right) - \kappa p(x_i^t)\right).$$

Hence

$$\begin{aligned}&\mathbb{E}\left[\hat{c}_i\left(x_i^t, x_{-i}^t\right) - \hat{c}_i\left(\omega_i, x_{-i}^t\right)\right] \\ &\leq \mathbb{E}\left[\frac{\left(D_h\left(\omega_i, x_i^t\right) - D_h\left(\omega_i, x_i^{t+1}\right)\right)}{\eta_t} + \kappa\left(tD_p\left(\omega_i, x_i^t\right) - (t+1)D_p\left(\omega_i, x_i^{t+1}\right)\right) + \left\langle\hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle\right].\end{aligned}$$

When $\mu = 0$, by Lemma J.2, we have $\mathbb{E}\left[\left\langle\hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle\right] \leq \eta_t\mathbb{E}\left[\left\|A_i^t\hat{g}_i^t\right\|^2\right]$. Taking expectation conditioned on x^t, we have $\mathbb{E}\left[\left\|A_i^t\hat{g}_i^t\right\|^2 \mid x^t\right] = \frac{d^2}{\delta_t^2}\mathbb{E}\left[\tilde{c}_i(\hat{x}^t)^2\|z_i^t\|^2 \mid x^t\right] \leq \frac{d^2}{\delta_t^2}$, and therefore $\mathbb{E}\left[\left\langle\hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle\right] \leq \frac{\eta_t d^2}{\delta_t^2}$.

Taking summation over T, and take $\eta_t = \frac{1}{2dt^{3/4}}$, $\delta_t = \frac{1}{t^{1/4}}$ we have

$$\sum_{t=1}^{T} \mathbb{E}\left[\hat{c}_i\left(x_i^t, x_{-i}^t\right) - \hat{c}_i\left(\omega_i, x_{-i}^t\right)\right] \leq dT^{3/4}\mathbb{E}\left[D_h\left(\omega_i, x_i^1\right)\right] + \kappa\mathbb{E}\left[D_p\left(\omega_i, x_i^1\right)\right] + \sum_{t=1}^{T}\frac{\eta_t d^2}{\delta^2}$$
$$\leq O\left(dT^{3/4}\mathbb{E}\left[D_h\left(\omega_i, x_i^1\right)\right] + \kappa C_p + T^{3/4}\right),$$

as we assumed $D_p(x_i, x_i')$ is bounded for any x_i, x_i'.

When $\mu > 0$, taking expectation conditioned on x^t, we have $\mathbb{E}\left[\|A_i^t\hat{g}_i^t\|^2 \mid x^t\right] = d^2\mathbb{E}\left[c_i(\hat{x}^t)^2\|z_i^t\|^2 \mid x^t\right] \leq d^2$. By Lemma J.2, and the choice $\eta_t = \frac{1}{2d\sqrt{t}}$, we have

$$\sum_{t=1}^{T}\sum_{i\in\mathcal{N}} \mathbb{E}\left[\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\rangle\right] \leq \sum_{t=1}^{T}\eta_t\sum_{i\in\mathcal{N}}\mathbb{E}\left[\left\|A_i^t\hat{g}_i^t\right\|^2\right] \leq nd^2\sum_{t=1}^{T}\eta_t = nd^2\sqrt{T}.$$

Taking summation over T, and take $\eta_t = \frac{1}{2dt^{1/2}}$, we have

$$\sum_{t=1}^{T} \mathbb{E}\left[\hat{c}_i\left(x_i^t, x_{-i}^t\right) - \hat{c}_i\left(\omega_i, x_{-i}^t\right)\right] \leq dT^{1/2}\mathbb{E}\left[D_h\left(\omega_i, x_i^1\right)\right] + \kappa\mathbb{E}\left[D_p\left(\omega_i, x_i^1\right)\right] + nd^2\sqrt{T},$$

as we assumed $D_p(x_i, x_i')$ is bounded for any x_i, x_i'.

Define $\pi_x(y) = \inf\left\{t \geq 0 : x + \frac{1}{t}(y - x) \in \mathcal{X}_i\right\}$. Notice that $x_i^1(x) = \operatorname{argmin}_{x_i\in\mathcal{X}_i} h(x_i)$, so $D_h(\omega_i, x_i^1) = h(\omega_i) - h(x_i^1)$.

- If $\pi_{x_i^1}(\omega_i) \leq 1 - \frac{1}{\sqrt{T}}$, then by Lemma J.6, $D_h(\omega_i, x_i^1) = \nu\log(T)$, and $\sum_{t=1}^{T}\mathbb{E}\left[\hat{c}_i\left(x_i^t, x_{-i}^t\right) - \hat{c}_i\left(\omega_i, x_{-i}^t\right)\right] = O\left(\nu dT^{3/4}\log(T)\right)$.
- Otherwise, we find a point ω_i' such that $\|\omega_i' - \omega_i\| = O(1/\sqrt{T})$ and $\pi_{x_i^1}(\omega_i') \leq 1 - \frac{1}{\sqrt{T}}$. Then $D_h(\omega_i', x_i^1) = \nu\log(T)$,

  $$\hat{c}_i\left(\omega_i', x_{-i}^t\right) - \hat{c}_i\left(\omega_i, x_{-i}^t\right) \leq \left\langle \nabla_i\hat{c}_i\left(\omega_i', x_{-i}^t\right), \omega_i' - \omega_i\right\rangle \leq \|\nabla_i\hat{c}_i\left(\omega_i', x_{-i}^t\right)\|\|\omega_i' - \omega_i\| \leq \frac{\max_x\|\nabla_i c_i\left(x\right)\|}{\sqrt{T}}.$$

  Therefore, $\sum_{t=1}^{T}\mathbb{E}\left[\hat{c}_i\left(x_i^t, x_{-i}^t\right) - \hat{c}_i\left(\omega_i, x_{-i}^t\right)\right] = O\left(\nu dT^{3/4}\log(T) + \max_x\|\nabla_i c_i\left(x\right)\|\sqrt{T}\right)$.

For the second term, by Jensen's inequality, we have

$$\hat{c}_i\left(x_i^t, x_{-i}^t\right)\mathbb{E}_{w_i^t\sim\mathbb{B}^d}\left[c_i\left(x_i^t + \delta_t A_i^t w_i^t, x_{-i}^t\right)\right] \geq c_i\left(\mathbb{E}_{w_i^t\sim\mathbb{B}^d}x_i^t + \delta_t A_i^t w_i^t, x_{-i}^t\right) = c_i\left(x_i^t, x_{-i}^t\right).$$

Therefore, we have $\sum_{t=1}^{T}\left(c_i\left(x_i^t, x_{-i}^t\right) - \hat{c}_i\left(x_i^t, x_{-i}^t\right)\right) = 0$.

When $\mu = 0$, by the definition of $\hat{c}_i$ and the smoothness of c_i,

$$\|\nabla_i\hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right)\| = \left\|\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left[\nabla_i c_i\left(x_i^t + \delta_t A_i^t w_i, \hat{x}_{-i}^t\right) - \nabla_i c_i\left(x^t\right)\right]\right\|$$
$$\leq \ell_i\sqrt{\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left[\delta_t^2\left\|\delta_t A_i w_i\right\|^2 + \delta_t^2\sum_{j\neq i}\left\|A_j z_j\right\|^2\right]}.$$

Since p is monotone, $\nabla^2 p(x)$ is positive semi-definite, and $A_i^t \preceq (\nabla^2 h(x_i))^{-1/2}$. For $\bar{x}_i^t = x_i^t + A_i^t w_i^t$. Define $\|v\|_x = \sqrt{v^\top\nabla^2 h(x)v}$, we have $\|\bar{x}_i^t - x_i^t\|_{x_i} \leq \|\omega_i^t\| \leq 1$, and $\bar{x}_i^t \in W(x_i^t)$, where $W(x) = \{x_i' \in \mathbb{R}^d, \|x_i' - x_i\|_{x_i} \leq 1\}$ is the Dikin ellipsoid. Since $W(x_i) \subseteq \mathcal{X}_i, \forall x_i \in int(\mathcal{X}_i)$, we can upper bound $\|A_i w_i\|^2$ by B^2, the diameter of the set $\mathcal{X}_i$. Hence $\|\nabla_i\hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right)\| \leq \ell_i\delta_t\sqrt{n}B$.

Therefore, for the third term, we have

$$\sum_{t=1}^{T} \mathbb{E}\left[\hat{c}_i\left(\omega_i, x_{-i}^t\right) - c_i\left(\omega_i, x_{-i}^t\right)\right] \leq O\left(\sum_{t=1}^{T} \ell_i \delta_t \sqrt{n} B\right).$$

Similarly, for the fourth term, we have $\sum_{t=1}^{T} \mathbb{E}\left[c_i\left(\hat{x}_i^t, x_{-i}^t\right) - c_i\left(x_i^t, x_{-i}^t\right)\right] \leq O\left(\sum_{t=1}^{T} \ell_i \delta_t \sqrt{n} B\right)$.

When $\mu > 0$, by Lemma J.5, for any $\omega_i \in \mathcal{X}_i$, we have

$$\left\|\nabla_i \hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right)\right\| \leq \ell_i \sqrt{\sum_{j \in \mathcal{N}} \left(\sigma_{\max}\left(A_j^t\right)^2\right)} \leq \frac{n \ell_i}{\sqrt{\mu(t+1)}}.$$

where the second inequality is by $\nabla^2 h(x)$ being positive definite, and $\nabla^2 p(x) \geq \mu I$.

Therefore, for the third term, we have

$$\sum_{t=1}^{T} \mathbb{E}\left[\hat{c}_i\left(\omega_i, x_{-i}^t\right) - c_i\left(\omega_i, x_{-i}^t\right)\right] \leq O\left(\frac{n B \ell_i \sqrt{T}}{\mu}\right).$$

Similarly, for the fourth term, we have $\sum_{t=1}^{T} \mathbb{E}\left[c_i\left(\hat{x}_i^t, x_{-i}^t\right) - c_i\left(x_i^t, x_{-i}^t\right)\right] \leq O\left(\frac{n B \ell_i \sqrt{T}}{\mu}\right)$.

When $\mu = 0$, with $\delta_t = \frac{1}{t^{1/4}}$, we have the regret as

$$\sum_{t=1}^{T} \mathbb{E}\left[c_i\left(\hat{x}_i^t, x_{-i}^t\right) - c_i\left(\omega_i, x_{-i}^t\right)\right] = O\left(\nu d T^{3/4} \log(T) + \max_x \|\nabla_i c_i\left(x\right)\| \sqrt{T} + \ell_i \sqrt{n} B T^{3/4}\right).$$

When $\mu > 0$, we have the regret as

$$\sum_{t=1}^{T} \mathbb{E}\left[c_i\left(\hat{x}_i^t, x_{-i}^t\right) - c_i\left(\omega_i, x_{-i}^t\right)\right] = O\left(\nu d T^{1/2} \log(T) + \max_x \|\nabla_i c_i\left(x\right)\| \sqrt{T} + \frac{n B \ell_i \sqrt{T}}{\mu}\right).$$

Combining the terms yields the final result. $\square$

D Proof of Theorem D.1

We now consider the case where every player receive $\tilde{c}_i(x^t) = c_i(x^t) + \epsilon_i^t$, where $\mathbb{E}[\epsilon_i^t \mid \hat{x}^t] = 0$, and $\|\epsilon_i^t\|^2 \le \sigma$. The following theorem describes the last-iterate convergence rate (in expectation) for monotone and strongly monotone loss under noisy bandit feedback.

Theorem D.1. *With* $\eta_t = \frac{1}{4d^2(1+\sigma)t^{3/4}}$, $\delta_t = \frac{1}{t^{1/4}}$

$$\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right) \le O\left(\frac{n\nu d^2(1+\sigma)\log(T)}{\kappa T^{1/4}} + \frac{n\zeta d^2(1+\sigma)B}{T^{3/4}} + \frac{nd^2(1+\sigma)C_p}{T^{1/4}} + \frac{\sqrt{n}B^2 L\log(T)}{\kappa T^{1/4}} + \frac{nd\log(T)}{\kappa(1+\sigma)^2 T^{1/4}}\right).$$

Proof. Similar to Theorem 5.1, with Lemma J.1, we have

$$\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right) \le O\left(\frac{n\nu\log(T)}{\kappa\eta_T T} + \frac{n\zeta B}{\eta_T T^{3/2}}\right) + O\left(\frac{nB\sum_{i\in\mathcal{N}}\ell_i}{\kappa T^{3/2}} + \frac{n}{\kappa T^{3/2}}\right)\frac{\sum_{t=1}^T \eta_t}{\eta_T} + O\left(\frac{nC_p}{\eta_T T}\right) + \frac{\sqrt{n}B^2 L\sum_{t=1}^T \eta_t\delta_t}{\eta_T\kappa(T+1)} + \frac{1}{\eta_T\kappa(T+1)}\sum_{i\in\mathcal{N}}\sum_{t=1}^T \eta_t\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle.$$

Taking expectation conditioned on x^t, we have $\mathbb{E}\left[\left\|A_i^t\hat{g}_i^t\right\|^2 \mid x^t\right] = \frac{d^2}{\delta_t^2}\mathbb{E}\left[\tilde{c}_i(\hat{x}^t)^2\|z_i^t\|^2 \mid x^t\right] \le \frac{d^2}{\delta_t^2}(2+2\sigma)$. By Lemma J.2, and the choice $\eta_t = \frac{1}{4d^2(1+\sigma)t^{3/4}}$, we have

$$\sum_{t=1}^T \eta_t \sum_{i\in\mathcal{N}}\mathbb{E}\left[\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle\right] \le \sum_{t=1}^T \eta_t^2\sum_{i\in\mathcal{N}}\mathbb{E}\left[\left\|A_i^t\hat{g}_i^t\right\|^2\right] \le nd^2\sum_{t=1}^T\frac{\eta_t^2}{\delta_t^2} = \frac{n\log(T)}{16(1+\sigma)^2}.$$

Combining everything, we have

$$\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right) \le O\left(\frac{n\nu d^2(1+\sigma)\log(T)}{\kappa T^{1/4}} + \frac{n\zeta d^2(1+\sigma)B}{T^{3/4}} + \frac{nd^2(1+\sigma)C_p}{T^{1/4}} + \frac{\sqrt{n}B^2 L\log(T)}{\kappa T^{1/4}} + \frac{nd\log(T)}{\kappa(1+\sigma)^2 T^{1/4}}\right).$$

□

E Proof of Theorem 5.2

Theorem 5.2. *With a probability of at least* $1 - \log(T)\delta$, $\delta \leq e^{-1}$, *and with Algorithm 1, we have* $\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right) \leq O\left(\frac{nd\nu\log(T)}{\sqrt{T}} + \frac{nd\zeta B}{T} + \frac{nBL}{\sqrt{T}} + \frac{ndC_p}{\sqrt{T}} + \frac{nd\log(T)}{\sqrt{T}} + \frac{dBL\log(T)}{\mu\sqrt{T}} + \frac{nBd^2\log^2(1/\delta)\log(T)}{\min\{\sqrt{\mu},\mu\}\sqrt{T}}\right)$.

Proof. Lemma J.1, we have

$$\begin{aligned}
&\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right) \\
&\leq O\left(\frac{n\nu\log(T)}{\kappa\eta_T T} + \frac{n\zeta B}{\eta_T T^{3/2}}\right) + O\left(\frac{nB\sum_{i\in\mathcal{N}}\ell_i}{\kappa T^{3/2}} + \frac{n}{\kappa T^{3/2}}\right)\frac{\sum_{t=1}^T \eta_t}{\eta_T} + O\left(\frac{nC_p}{\eta_T T}\right) \\
&\quad + \frac{1}{\kappa\eta_T(T+1)}\sum_{i\in\mathcal{N}}\sum_{t=1}^T \eta_t\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle + \frac{1}{\kappa\eta_T(T+1)}\sum_{t=1}^T\eta_t\sum_{i\in\mathcal{N}}\left\langle \hat{g}_i^t - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle .
\end{aligned}$$

By Lemma J.2, we have

$$\sum_{t=1}^T \eta_t \sum_{i\in\mathcal{N}}\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle \leq \sum_{t=1}^T \eta_t^2\sum_{i\in\mathcal{N}}\left\|A_i^t\hat{g}_i^t\right\|^2 \leq nd^2\sum_{t=1}^T\eta_t^2 .$$

We then decompose the last term as

$$\sum_{t=1}^T\eta_t\sum_{i\in\mathcal{N}}\left\langle \hat{g}_i^t - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle = \sum_{t=1}^T\eta_t\sum_{i\in\mathcal{N}}\left\langle g_i^t - \hat{c}_i^t(x_i^t), \omega_i - x_i^t\right\rangle + \sum_{i\in\mathcal{N}}\sum_{t=1}^T\eta_t\left\langle \nabla_i\hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle .$$

By Lemma E.1, we have

$$\sum_{t=1}^T \eta_t\left\langle g_i^t - \hat{c}_i^t(x_i^t), \omega_i - x_i^t\right\rangle \leq O\left(\frac{Bd\log^2(1/\delta)\log(T)}{\min\{\sqrt{\mu},\mu\}}\right),$$

with a probability of at least $1-\log(T)\delta$, $\delta \leq e^{-1}$.

By Lemma J.5, for any $\omega_i \in \mathcal{X}_i$, we have

$$\begin{aligned}
\sum_{i\in\mathcal{N}}\sum_{t=1}^T\eta_t\left\langle \nabla_i\hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle &\leq \sum_{i\in\mathcal{N}} B\ell_i\sum_{t=1}^T\eta_t\sum_{j\in\mathcal{N}}\left(\sigma_{\max}\left(A_j^t\right)^2\right) \mid x^t \\
&\leq \sum_{i\in\mathcal{N}} B\ell_i\sum_{t=1}^T\frac{1}{\mu(t+1)} \\
&\leq \frac{B\sum_{i\in\mathcal{N}}\ell_i}{\mu}\sum_{t=1}^T\frac{1}{(t+1)} \\
&\leq \frac{BL\log(T)}{\mu}
\end{aligned}$$

where the third inequality is by $\nabla^2 h(x)$ being positive definite, and $\nabla^2 p(x) \geq \mu I$.

Therefore,

$$\sum_{t=1}^T\eta_t\sum_{i\in\mathcal{N}}\left\langle \hat{g}_i^t - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle \leq O\left(\frac{BL\log(T)}{\mu} + \frac{nBd\log^2(1/\delta)\log(T)}{\min\{\sqrt{\mu},\mu\}}\right) .$$

Combining the terms, and with $\eta_t = \frac{1}{2d\sqrt{t}}$, we have

$$\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)$$

$$\leq O\left(\frac{nd\nu\log(T)}{\kappa\sqrt{T}}+\frac{nd\zeta B}{T}+\frac{nBL}{\kappa\sqrt{T}}+\frac{ndC_p}{\sqrt{T}}+\frac{nd\log(T)}{\kappa\sqrt{T}}+\frac{dBL\log(T)}{\kappa\mu\sqrt{T}}+\frac{nBd^2\log^2(1/\delta)\log(T)}{\kappa\min\{\sqrt{\mu},\mu\}\sqrt{T}}\right).$$

□

Lemma E.1. *With a probability of at least* $1-\log(T)\delta$, $\delta\leq e^{-1}$, *we have*

$$\sum_{t=1}^{T}\eta_t\left\langle g_i^t-\hat{c}_i^t(x_i^t),\omega_i-x_i^t\right\rangle\leq O\left(\frac{Bd\log^2(1/\delta)\log(T)}{\min\{\sqrt{\mu},\mu\}}\right).$$

Proof. Define $Z_t=\eta_t\left\langle g_i^t-\hat{c}_i^t(x_i^t),\omega_i-x_i^t\right\rangle$. $\mathrm{Var}[Z_t]\leq\eta^2(\omega_i-x_i^t)^\top\mathbb{E}[g_i^t(g_i^t)^\top](\omega_i-x_i^t)$. Then, with $\eta_t=\frac{1}{2d\sqrt{t}}$,

$$\max_t|Z_t|\leq\max_t\left\|\eta_t\left(g_i^t-\hat{c}_i^t(x_i^t)\right)\right\|\left\|\omega_i-x_i^t\right\|\leq O\left(Bd\max_t\|\eta_t(A_i^t)^{-1}z_i^t\|\right)\leq O\left(\max_t\frac{Bd}{\mu(t+1)}\right)\leq O\left(\frac{Bd}{\mu}\right),$$

where the third inequality is by the definition of A_i^t.

By the definition of gradient estimator, we have

$$(g_i^t)^\top g_i^t\leq d^2\left((A_i^t)^{-1}z_i^t\right)^\top\left((A_i^t)^{-1}z_i^t\right)\leq\frac{d^2}{\mu\eta_t(t+1)}.$$

Therefore, with $\eta_t=\frac{1}{2d\sqrt{t}}$

$$(\omega_i-x_i^t)^\top\mathbb{E}[g_i^t(g_i^t)^\top](\omega_i-x_i^t)\leq\frac{d^2\|\omega_i-x_i^t\|^2}{\mu\eta_t(t+1)}\leq\frac{d^2B^2}{\mu\eta_t(t+1)}\leq\frac{dB^2}{\mu\sqrt{t}}.$$

We have

$$\sqrt{\sum_{t=1}^{T}\eta_t^2(\omega_i-x_i^t)^\top\mathbb{E}[g_i^t(g_i^t)^\top](\omega_i-x_i^t)}\leq\sqrt{\sum_{t=1}^{T}\frac{B^2}{d\mu t^{3/2}}}\leq O\left(\frac{B\sqrt{\log(T)}}{\sqrt{d\mu}}\right).$$

Then, by Lemma 2 of Bartlett et al. (2008), with a probability of at least $1-\log(T)\delta$, $\delta\leq e^{-1}$,

$$\begin{aligned}\sum_{t=1}^{T}\eta_t\left\langle g_i^t-\hat{c}_i^t(x_i^t),\omega_i-x_i^t\right\rangle&\leq 2\max\left\{2\sqrt{\sum_{t=1}^{T}\mathrm{Var}[Z_t]},\max_t|Z_t|\log(1/\delta)\right\}\\&\leq\max\left\{O\left(\frac{B\sqrt{\log(T)}}{\sqrt{d\mu}}\right),O\left(\frac{Bd\log(1/\delta)}{\mu}\right)\right\}\cdot\log(1/\delta)\\&\leq O\left(\frac{Bd\log^2(1/\delta)\log(T)}{\min\{\sqrt{\mu},\mu\}}\right).\end{aligned}$$

□

F Proof of Theorem 5.4

Theorem 5.4. *With* $\eta_t = \frac{1}{2d\sqrt{t}}$, $\tau = \frac{1}{T^{1/6}}$, $G_p = \sup_x \|\nabla p(x)\|$ *and Algorithm 1, we have* $\mathbb{E}\left[\sum_{i\in\mathcal{N}} \left\langle \nabla_i c_i\left(x^T\right), x_i^T - x_i^*\right\rangle\right] \leq \tilde{O}\left(\frac{BG_p+\sqrt{d(BL+G)(n\nu+nBL+nd^2)}}{T^{1/6}} + \frac{\sqrt{dBL(BL+G)}}{\sqrt{\mu}T^{1/6}} + \frac{\sqrt{dnC_p(BL+G)}}{\sqrt{\mu}T^{1/4}}\right)$.

Proof. We consider a regularized game with operator $\tilde{F}(x) = [\tilde{F}_i(x)]_{i\in\mathcal{N}}$, where $\tilde{F}_i(x) = \nabla c_i(x) + \tau\nabla p(x_i)$, $\nabla p(x) = [\nabla_i p(x_i)]_{i\in\mathcal{N}}$.

Similar to Lemma J.1, we have

$$
\begin{aligned}
&\sum_{i\in\mathcal{N}} D_p\left(x_i^\tau, x_i^{T+1}\right) \\
&\leq O\left(\frac{n\nu\log(T)}{\eta_T\tau T} + \frac{n\mu B}{\eta_T\tau T^{3/2}}\right) + O\left(\frac{nB\sum_{i\in\mathcal{N}}\ell_i}{\tau T^{3/2}} + \frac{n}{\tau T^{3/2}}\right)\frac{\sum_{t=1}^T \eta_t}{\eta_T} + O\left(\frac{nC_p}{\eta_T T}\right) \\
&\quad + \frac{1}{\eta_T\tau(T+1)}\sum_{i\in\mathcal{N}}\sum_{t=1}^T \eta_t\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle + \frac{1}{\eta_T\tau(T+1)}\sum_{t=1}^T \eta_t \sum_{i\in\mathcal{M}}\left\langle \hat{g}_i^t - \tilde{F}_i\left(x^t\right), x_i^\tau - x_i^t\right\rangle \\
&\quad + \frac{1}{\eta_T\tau(T+1)}\sum_{t=1}^T \eta_t \sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle \hat{g}_i^t - \tilde{F}_i\left(x^t\right), \bar{x}_i - x_i^t\right\rangle .
\end{aligned}
$$

Taking expectation conditioned on x^t, we have $\mathbb{E}\left[\|A_i^t\hat{g}_i^t\|^2 \mid x^t\right] = d^2\mathbb{E}\left[c_i(\hat{x}^t)^2\|z_i^t\|^2 \mid x^t\right] \leq d^2$.
By Lemma J.2, and the choice $\eta_t = \frac{1}{2d\sqrt{t}}$, we have

$$
\sum_{t=1}^T \eta_t \sum_{i\in\mathcal{N}} \mathbb{E}\left[\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle\right] \leq \sum_{t=1}^T \eta_t^2 \sum_{i\in\mathcal{N}} \mathbb{E}\left[\left\|A_i^t\hat{g}_i^t\right\|^2\right] \leq nd^2\sum_{t=1}^T \eta_t^2 .
$$

By Lemma J.5, for any $\omega_i \in \mathcal{X}_i$, we have

$$
\begin{aligned}
\sum_{i\in\mathcal{N}}\sum_{t=1}^T \eta_t\mathbb{E}\left[\left\langle \hat{g}_i^t - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right] &= \sum_{i\in\mathcal{N}}\sum_{t=1}^T \eta_t\mathbb{E}\left[\left\langle \nabla_i\hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right] \\
&\leq \sum_{i\in\mathcal{N}}\sum_{t=1}^T \eta_t\mathbb{E}\left[\left\|\nabla_i\hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right)\right\|\left\|\omega_i - x_i^t\right\| \mid x^t\right] \\
&\leq \sum_{i\in\mathcal{N}} B\ell_i \sum_{t=1}^T \eta_t\mathbb{E}\left[\sum_{j\in\mathcal{N}}\left(\sigma_{\max}\left(A_j^t\right)^2\right) \mid x^t\right] \\
&\leq \sum_{i\in\mathcal{N}} B\ell_i \sum_{t=1}^T \frac{1}{\mu(t+1)} \\
&\leq \frac{B\sum_{i\in\mathcal{N}}\ell_i}{\mu}\sum_{t=1}^T \frac{1}{(t+1)} .
\end{aligned}
$$

where the third inequality is by $\nabla^2 h(x)$ being positive definite, and $\nabla^2 p(x) \geq \mu I$.

Combing and rearranging the terms, we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i\in\mathcal{N}} D_p\left(x_i^\tau, x_i^{T+1}\right)\right] \\
&\leq O\left(\frac{n\nu\log(T)}{\eta_T\tau T} + \frac{n\zeta B}{\eta_T\tau T^{3/2}}\right) + O\left(\frac{nB\sum_{i\in\mathcal{N}}\ell_i}{\tau\sqrt{T}} + \frac{n}{\tau\sqrt{T}}\right) + O\left(\frac{nC_p}{\eta_T T}\right) + O\left(\frac{nd^2}{\tau\eta_T T}\sum_{t=1}^T \eta_t^2 + \frac{B\sum_{i\in\mathcal{N}}\ell_i}{\tau\mu\eta_T T}\sum_{t=1}^T \frac{1}{t}\right) .
\end{aligned}
$$

Take $\eta_t = \frac{1}{2d\sqrt{t}}$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i\in\mathcal{N}} D_p\left(x_i^\tau, x_i^{T+1}\right)\right] \\
&\leq O\left(\frac{nd\nu\log(T)}{\tau\sqrt{T}} + \frac{nd\zeta B}{\tau T} + \frac{nB\sum_{i\in\mathcal{N}}\ell_i}{\tau\sqrt{T}} + \frac{n}{\tau\sqrt{T}} + \frac{ndC_p}{\sqrt{T}} + \frac{nd\log(T)}{\tau\sqrt{T}} + \frac{dB\log(T)\sum_{i\in\mathcal{N}}\ell_i}{\tau\mu\sqrt{T}}\right) .
\end{aligned}
$$

We can decompose as

$$
\begin{aligned}
&\left\langle F\left(x^T\right), x^T - x^*\right\rangle \\
&= \left\langle F\left(x^T\right), x^T - x^\tau\right\rangle + \left\langle F\left(x^T\right), x^\tau - x^*\right\rangle \\
&\leq G\left\|x^T - x^\tau\right\| + \left\langle F\left(x^\tau\right) + \tau\nabla p(x^\tau), x^\tau - x^*\right\rangle + \left\langle F\left(x^T\right) - F\left(x^\tau\right), x^\tau - x^*\right\rangle + \tau B\left\|\nabla p(x^\tau)\right\| \\
&\leq \sum_{i\in\mathcal{N}}(B\ell_i + G)\left\|x_i^T - x^\tau\right\| + \tau B\left\|\nabla p(x^\tau)\right\| .
\end{aligned}
$$

Since $\nabla^2 p(x) \succeq \mu I$, we have $\|x_i^\tau - x_i^T\| \leq \sqrt{D_p(x_i^\tau, x_i^T)}$. Let $G_p = \sup_x\|\nabla p(x)\|$, $L = \sum_{i\in\mathcal{N}}\ell_i$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i\in\mathcal{N}}\left\langle\nabla_i c_i\left(x^T\right), x_i^T - x_i^*\right\rangle\right] \\
&\leq O\left(\tau B G_p\right) + \tilde{O}\left(\frac{\sqrt{d(BL+G)(n\nu + nBL + nd^2)}}{\sqrt{\tau}T^{1/4}}\right) + \tilde{O}\left(\frac{\sqrt{dBL(BL+G)}}{\sqrt{\tau\mu}T^{1/4}}\right) + O\left(\frac{\sqrt{dnC_p(BL+G)}}{\sqrt{\mu}T^{1/4}}\right) \\
&\leq \tilde{O}\left(\frac{BG_p + \sqrt{d(BL+G)(n\nu + nBL + nd^2)}}{T^{1/6}}\right) + \tilde{O}\left(\frac{\sqrt{dBL(BL+G)}}{\sqrt{\mu}T^{1/6}}\right) + O\left(\frac{\sqrt{dnC_p(BL+G)}}{\sqrt{\mu}T^{1/4}}\right) ,
\end{aligned}
$$

where the last inequality is by taking $\tau = \frac{1}{T^{1/6}}$. $\square$

G Proof of Proposition 5.1

Proposition 5.1. *With* $\eta_t = \frac{1}{2dt^{3/4}}, \delta_t = \frac{1}{t^{1/4}}$*, and suppose every player employ Algorithm 1, we have* $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\mathrm{SW}(\hat{x})\right] = O\left(\frac{C_1\mathrm{OPT}}{(1-C_2)} + \frac{n\nu d\log(T)}{(1-C_2)T^{1/4}} + \frac{\sqrt{n}B\sum_{i\in\mathcal{N}}\ell_i}{(1-C_2)T^{1/4}}\right)$.

Proof. By Theorem 5.3, we have

$$\sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\mathbb{E}\left[c_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right)\right] \le \sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\mathbb{E}\left[c_i\left(\omega_i, \hat{x}_{-i}^t\right)\right] + O\left(n\nu dT^{3/4}\log(T) + \sqrt{n}BT^{3/4}\sum_{i\in\mathcal{N}}\ell_i\right)$$
$$\le C_1\mathrm{OPT}\cdot T + C_2\sum_{t=1}^{T}\mathbb{E}\left[\mathrm{SW}(\hat{x})\right] + O\left(n\nu dT^{3/4}\log(T) + \sqrt{n}BT^{3/4}\sum_{i\in\mathcal{N}}\ell_i\right).$$

As $\sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\mathbb{E}\left[c_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right)\right] = \mathbb{E}\left[\mathrm{SW}(\hat{x})\right]$, we solve for $\mathbb{E}\left[\mathrm{SW}(\hat{x})\right]$ and obtain

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\mathrm{SW}(\hat{x})\right] = O\left(\frac{C_1\mathrm{OPT}}{(1-C_2)} + \frac{n\nu d\log(T)}{(1-C_2)T^{1/4}} + \frac{\sqrt{n}B\sum_{i\in\mathcal{N}}\ell_i}{(1-C_2)T^{1/4}}\right).$$

□

H Proof of Theorem 6.1

Theorem 6.1. *With* $\sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\max_x\|\nabla_i c_i(x)-\nabla_i c_i^t(x)\|_2 = T^\alpha$*, take* $\eta_t = \frac{1}{2dt^{3/4}}$*,* $\delta_t = \frac{1}{t^{1/4}}$*, and under Algorithm 1, we have* $\mathbb{E}\left[\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\right] \leq O\left(\frac{nd\nu\log(T)}{\kappa T^{1/4}} + \frac{n\zeta dB}{T^{3/4}} + \frac{nBL}{\kappa\sqrt{T}} + \frac{ndC_p}{T^{1/4}} + \frac{nd\log(T)}{\kappa T^{1/4}} + \frac{\sqrt{n}B^2L\log(T)}{\kappa T^{1/4}} + \frac{B}{T^{1/4-\alpha}}\right)$.

Proof. Similar to Theorem 5.1, we have

$$
\begin{aligned}
&\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\\
&\leq O\left(\frac{n\nu\log(T)}{\eta_T\kappa T} + \frac{n\zeta B}{\eta_T T^{3/2}}\right) + O\left(\frac{nB\sum_{i\in\mathcal{N}}\ell_i}{\kappa T^{3/2}} + \frac{n}{\kappa T^{3/2}}\right)\frac{\sum_{t=1}^T\eta_t}{\eta_T} + O\left(\frac{nC_p}{\eta_T T}\right)\\
&\quad + \frac{1}{\kappa\eta_T(T+1)}\sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\left\langle\hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle + \frac{1}{\kappa\eta_T(T+1)}\sum_{t=1}^{T}\eta_t\sum_{i\in\mathcal{M}}\left\langle\hat{g}_i^t - \nabla_i c_i^t\left(x^t\right), x_i^* - x_i^t\right\rangle\\
&\quad + \frac{1}{\kappa\eta_T(T+1)}\sum_{t=1}^{T}\eta_t\sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle\hat{g}_i^t - \nabla_i c_i^t\left(x^t\right), \bar{x}_i - x_i^t\right\rangle + B\sum_{t=1}^{T}\Delta^t\,,
\end{aligned}
$$

where $\Delta^t = \sum_{i\in\mathcal{N}}\max_x\|\nabla_i c_i(x) - \nabla_i c_i^t(x)\|_2$.

We now upper bound the remaining terms by discussing them by cases.

When $\mu = 0$, taking expectation conditioned on x^t, we have $\mathbb{E}\left[\|A_i^t\hat{g}_i^t\|^2 \mid x^t\right] = \frac{d^2}{\delta_t^2}\mathbb{E}\left[c_i^t(\hat{x}^t)^2\|z_i^t\|^2 \mid x^t\right] \leq \frac{d^2}{\delta_t^2}$. By Lemma J.2, and the choice $\eta_t = \frac{1}{2d\sqrt{t}}$, we have

$$
\sum_{t=1}^{T}\eta_t\sum_{i\in\mathcal{N}}\mathbb{E}\left[\left\langle\hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle\right] \leq \sum_{t=1}^{T}\eta_t^2\sum_{i\in\mathcal{N}}\mathbb{E}\left[\left\|A_i^t\hat{g}_i^t\right\|^2\right] \leq nd^2\sum_{t=1}^{T}\frac{\eta_t^2}{\delta_t^2}\,.
$$

By the definition of $\hat{c}_i$,

$$
\begin{aligned}
&\sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\left\langle\hat{g}_i^t - \nabla_i c_i^t\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right]\\
&= \sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\left\langle\nabla_i\hat{c}_i^t(x^t) - \nabla_i c_i^t\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right]\\
&= \sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left\langle\nabla_i c_i^t\left(x_i^t + \delta_t A_i^t w_i, \hat{x}_{-i}^t\right) - \nabla_i c_i^t\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right]\\
&\leq B\sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\mathbb{E}\left[\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left\|\nabla_i c_i^t\left(x_i^t + \delta_t A_i^t w_i, \hat{x}_{-i}^t\right) - \nabla_i c_i^t\left(x^t\right)\right\| \mid x^t\right]
\end{aligned}
$$

By the smoothness of c_i^t,

$$
\begin{aligned}
&\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left[\left\|\nabla_i c_i^t\left(x_i^t + \delta_t A_i^t w_i, \hat{x}_{-i}^t\right) - \nabla_i c_i^t\left(x^t\right)\right\|\right]\\
&\leq \ell_i\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left[\sqrt{\delta_t^2\left\|A_i w_i\right\|^2 + \delta_t^2\sum_{j\neq i}\left\|A_j z_j\right\|^2}\right]\,.
\end{aligned}
$$

Since p is monotone, $\nabla^2 p(x)$ is positive semi-definite, and $A_i^t \preceq (\nabla^2 h(x_i))^{-1/2}$. For $\bar{x}_i^t = x_i^t + A_i^t w_i^t$. Define $\|v\|_x = \sqrt{v^\top\nabla^2 h(x)v}$, we have $\|\bar{x}_i^t - x_i^t\|_{x_i} \leq \|\omega_i^t\| \leq 1$, and $\bar{x}_i^t \in W(x_i^t)$, where $W(x_i) = \{x_i' \in \mathbb{R}^d, \|x_i' - x_i\|_{x_i} \leq 1\}$ is the Dikin ellipsoid. Since $W(x_i) \subseteq \mathcal{X}_i, \forall x_i \in int(\mathcal{X}_i)$, we can upper bound $\|A_i w_i\|^2$ by B^2, the diameter of the set $\mathcal{X}_i$. Hence $\|\nabla_i\hat{c}_i(x^t) - \nabla_i c_i\left(x^t\right)\| \leq$

$\ell_i \delta_t \sqrt{n} B$. By Lemma J.5

$$\begin{aligned}
\sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t \mathbb{E}\left[\left\langle \hat{g}_i^t - \nabla_i c_i^t\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right] &= \sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t \mathbb{E}\left[\left\langle \nabla_i \hat{c}_i^t\left(x^t\right) - \nabla_i c_i^t\left(x^t\right), \omega_i - x_i^t\right\rangle \mid x^t\right] \\
&\leq \sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t \mathbb{E}\left[\left\|\nabla_i \hat{c}_i^t\left(x^t\right) - \nabla_i c_i^t\left(x^t\right)\right\| \left\|\omega_i - x_i^t\right\| \mid x^t\right] \\
&\leq \sqrt{n} B^2 \sum_{i\in\mathcal{N}} \ell_i \sum_{t=1}^{T} \eta_t \delta_t \,.
\end{aligned}$$

Let $L = \sum_{i\in\mathcal{N}} \ell_i$. When $\mu = 0$, combing and rearranging the terms, we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\right] \\
&\leq O\left(\frac{n\nu\log(T)}{\kappa\eta_T T} + \frac{n\zeta B}{\eta_T T^{3/2}} + \frac{nBL}{\kappa\sqrt{T}} + \frac{n}{\kappa\sqrt{T}} + \frac{nC_p}{\eta_T T} + \frac{nd^2}{\kappa\eta_T T}\sum_{t=1}^{T}\frac{\eta_t^2}{\delta_t^2} + \frac{\sqrt{n}B^2 L\sum_{t=1}^{T}\eta_t\delta_t}{\kappa\eta_T T} + \frac{B\sum_{t=1}^{T}\Delta^t}{\eta_T T}\right) .
\end{aligned}$$

Take $\eta_t = \frac{1}{2dt^{3/4}}$, $\delta_t = \frac{1}{t^{1/4}}$, then $\sum_{t=1}^{T}\frac{\eta_t^2}{\delta_t^2} = O\left(\sum_{t=1}^{T}\frac{1}{t}\right) = O(\log(T))$, and $\sum_{t=1}^{T}\eta_t\delta_t = O\left(\sum_{t=1}^{T}\frac{1}{t}\right) = O(\log(T))$. Hence, we have

$$\mathbb{E}\left[\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right)\right] \leq O\left(\frac{nd\nu\log(T)}{\kappa T^{1/4}} + \frac{n\zeta dB}{T^{3/4}} + \frac{nBL}{\kappa\sqrt{T}} + \frac{ndC_p}{T^{1/4}} + \frac{nd\log(T)}{\kappa T^{1/4}} + \frac{\sqrt{n}B^2 L\log(T)}{\kappa T^{1/4}} + \frac{B\Delta}{T^{1/4}}\right) ,$$

where $\Delta = \sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\max_x \|\nabla_i c_i(x) - \nabla_i c_i^t(x)\|_2$. $\square$

I Proof of Theorem 6.2

Theorem 6.2. *Assume $V_i(T) \le T^{\varphi}$, $\varphi \in [0,1]$. Take $\eta_t = \frac{1}{2dt^{\frac{(1-\varphi)}{3}}}$, $\delta_t = \frac{1}{t^{1/2}}$, and under Algorithm 1, we have $\frac{1}{T}\sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - x_i^{t,*}\right\rangle = \tilde{O}\Big(\frac{n\nu d + Ln^{3/2}B^2 + nG}{T^{\frac{2(1-\varphi)}{3}}} + \frac{n}{T^{\frac{9}{8}-\frac{(4\varphi+5)^2}{72}}}\Big)$.*

Proof. We first fix a player i decomposes

$$\sum_{t=1}^{T}\left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - x_i^{t,*}\right\rangle = \sum_{t=1}^{T}\left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - \omega_i\right\rangle + \sum_{t=1}^{T}\left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \omega_i - x_i^{t,*}\right\rangle .$$

For the second term, we partition the horizon of play T into m batches T_k, $k \in [m]$, each of length $|T_k| = T^q$, $q \in [0,1]$. We will determine q later. Note that the number of batches is thus $m = T^{1-q}$. For the batch T_k, we pick ω_i to be the Nash equilibrium of the first game. Then

$$\begin{aligned}
\sum_{t\in[T_k]}\left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \omega_i - x_i^{t,*}\right\rangle &\le \sum_{t\in[T_k]}\left\|\nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right)\right\|\left\|\omega_i - x_i^{t,*}\right\| \\
&\le GT^q \max_{t\in[T_k]}\left\|\omega_i - x_i^{t,*}\right\| \\
&\le GT^q \sum_{t\in[T_k]}\left\|x_i^{t+1,*} - x_i^{t,*}\right\| \\
&\le GT^q V_i(T_k),
\end{aligned}$$

where the third inequality is by the definition of ω_i.

Therefore, we have

$$\sum_{t=1}^{T}\left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - x_i^{t,*}\right\rangle = \sum_{k=1}^{m}\sum_{t\in[T_k]}\left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - \omega_i\right\rangle + GT^q V_i(T) .$$

Define the smoothed version of c_i as $\hat{c}_i^t(x) = \mathbb{E}_{w_i\sim\mathbb{B}^d}\left[c_i^t\left(x_i + \delta A_i w_i, x_{-i}\right)\right]$. Then, for batch T_k, we decompose $\sum_{t=1}^{T}\left\langle \nabla_i c_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - \omega_i\right\rangle$ as

$$\begin{aligned}
&\sum_{t\in[T_k]}\left\langle \nabla_i c_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - \omega_i\right\rangle \\
&= \sum_{t\in[T_k]}\left\langle \nabla_i \hat{c}_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - \omega_i\right\rangle + \sum_{t\in[T_k]}\left\langle \nabla_i c_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right) - \nabla_i \hat{c}_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - \omega_i\right\rangle \\
&\le \sum_{t\in[T_k]}\left\langle \nabla_i \hat{c}_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - \omega_i\right\rangle + B\sum_{t\in[T_k]}\left\|\nabla_i c_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right) - \nabla_i \hat{c}_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right)\right\|_2 .
\end{aligned}$$

For the first term, recall that by the update rule, we have,

$$\begin{aligned}
D_h\left(\omega_i, \hat{x}_i^{t+1}\right) = D_h\left(\omega_i, \hat{x}_i^t\right) &+ \eta_t\left\langle \nabla\hat{c}_i^t\left(\hat{x}^t\right), \omega_i - \hat{x}_i^t\right\rangle + \eta_t\left\langle \hat{g}_i^t - \nabla\hat{c}_i^t\left(\hat{x}^t\right), \omega_i - \hat{x}_i^t\right\rangle \\
&+ \eta_t\left\langle \hat{g}_i^t, \hat{x}_i^t - \hat{x}_i^{t+1}\right\rangle .
\end{aligned}$$

By Lemma J.5, for any $\omega_i \in \mathcal{X}_i$, we have

$$\mathbb{E}\left[\left\langle \hat{g}_i^t - \nabla\hat{c}_i^t\left(\hat{x}^t\right), \omega_i - \hat{x}_i^t\right\rangle \mid \hat{x}^t\right] = \mathbb{E}\left[\left\langle \nabla_i\hat{c}_i^t(\hat{x}^t) - \nabla_i\hat{c}_i^t\left(\hat{x}^t\right), \omega_i - \hat{x}_i^t\right\rangle \mid \hat{x}^t\right] = 0 .$$

Therefore,

$$\mathbb{E}\left[D_h\left(\omega_i, \hat{x}_i^{t+1}\right)\right] = \mathbb{E}\left[D_h\left(\omega_i, \hat{x}_i^t\right) + \eta_t\left\langle \nabla\hat{c}_i^t\left(\hat{x}^t\right), \omega_i - \hat{x}_i^t\right\rangle\right] + \eta_t\mathbb{E}\left[\left\langle \hat{g}_i^t, \hat{x}_i^t - \hat{x}_i^{t+1}\right\rangle\right] .$$

Rearranging the terms yields

$$\mathbb{E}\left[\left\langle \nabla\hat{c}_i^t\left(\hat{x}^t\right), \hat{x}_i^t - \omega_i\right\rangle\right] \le \mathbb{E}\left[\frac{\left(D_h\left(\omega_i, \hat{x}_i^t\right) - D_h\left(\omega_i, \hat{x}_i^{t+1}\right)\right)}{\eta_t} + \eta_t\left\langle \hat{g}_i^t, \hat{x}_i^t - \hat{x}_i^{t+1}\right\rangle\right] .$$

By Lemma J.2, we have $\mathbb{E}\left[\langle \hat{g}_i^t, \hat{x}_i^t - \hat{x}_i^{t+1}\rangle\right] \leq \eta_t \mathbb{E}\left[\|A_i^t \hat{g}_i^t\|^2\right]$. Taking expectation conditioned on $\hat{x}^t$, we have $\mathbb{E}\left[\|A_i^t \hat{g}_i^t\|^2 \mid \hat{x}^t\right] = \frac{d^2}{\delta_t^2}\mathbb{E}\left[\tilde{c}_i^t(\hat{x}^t)^2 \|z_i^t\|^2 \mid \hat{x}^t\right] \leq \frac{d^2}{\delta_t^2}$, and therefore $\mathbb{E}\left[\langle \hat{g}_i^t, \hat{x}_i^t - \hat{x}_i^{t+1}\rangle\right] \leq \frac{\eta_t d^2}{\delta_t^2}$.

Taking summation over T, and take $\eta_t = \frac{1}{2dt^p}$, $\delta_t = \frac{1}{t^r}$ we have

$$\sum_{t\in[T_k]} \mathbb{E}\left[\left\langle \nabla \hat{c}_i^t\left(\hat{x}^t\right), \hat{x}_i^t - \omega_i\right\rangle\right] \leq dT^p \mathbb{E}\left[D_h\left(\omega_i, x_i^1\right)\right] + \sum_{t\in[T_k]} \frac{\eta_t d^2}{\delta^2}$$
$$\leq O\left(dT^p \mathbb{E}\left[D_h\left(\omega_i, x_i^1\right)\right] + T^{q(p-2r)}\right),$$

as we assumed $D_p(x_i, x_i')$ is bounded for any x_i, x_i'.

Define $\pi_x(y) = \inf\left\{t \geq 0 : x + \frac{1}{t}(y - x) \in \mathcal{X}_i\right\}$. Notice that $x_i^1(x) = \operatorname{argmin}_{x_i\in\mathcal{X}_i} h(x_i)$, so $D_h(\omega_i, x_i^1) = h(\omega_i) - h(x_i^1)$.

- If $\pi_{x_i^1}(\omega_i) \leq 1 - \frac{1}{\sqrt{T^q}}$, then by Lemma J.6, $D_h(\omega_i, x_i^1) = \nu\log(T^q)$, and $\sum_{t=1}^T \mathbb{E}\left[\hat{c}_i\left(\hat{x}_i^t, x_{-i}^t\right) - \hat{c}_i\left(\omega_i, x_{-i}^t\right)\right] = O\left(\nu dT^{1-p}\log(T^q)\right)$.
- Otherwise, we find a point ω_i' such that $\|\omega_i' - \omega_i\| = O(1/\sqrt{T^q})$ and $\pi_{x_i^1}(\omega_i') \leq 1 - \frac{1}{\sqrt{T^q}}$. Then $D_h(\omega_i', x_i^1) = \nu\log(T^q)$,
$$\left\langle \nabla_i \hat{c}_i^t\left(\omega_i', x_{-i}^t\right), \omega_i' - \omega_i\right\rangle \leq \|\nabla_i \hat{c}_i^t\left(\omega_i', x_{-i}^t\right)\|\|\omega_i' - \omega_i\| \leq \frac{G}{\sqrt{T^q}}.$$
Therefore, $\sum_{t\in[T_k]} \mathbb{E}\left[\hat{c}_i\left(\hat{x}_i^t, x_{-i}^t\right) - \hat{c}_i\left(\omega_i, x_{-i}^t\right)\right] = O\left(\nu dT^p\log(T^q) + GT^{q/2} + T^{q(p-2r)}\right)$.

By the definition of $\hat{c}_i$ and the smoothness of c_i,

$$\|\nabla_i \hat{c}_i(\hat{x}^t) - \nabla_i c_i\left(\hat{x}^t\right)\| = \left\|\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left[\nabla_i c_i\left(\hat{x}_i^t + \delta_t A_i^t w_i, \hat{x}_{-i}^t\right) - \nabla_i c_i\left(\hat{x}^t\right)\right]\right\|$$
$$\leq \ell_i \sqrt{\mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left[\delta_t^2\left\|\delta_t A_i w_i\right\|^2 + \delta_t^2\sum_{j\neq i}\left\|A_j z_j\right\|^2\right]}.$$

Since p is monotone, $\nabla^2 p(x)$ is positive semi-definite, and $A_i^t \preceq (\nabla^2 h(x_i))^{-1/2}$. For $\bar{x}_i^t = \hat{x}_i^t + A_i^t w_i^t$. Define $\|v\|_x = \sqrt{v^\top \nabla^2 h(x) v}$, we have $\|\bar{x}_i^t - \hat{x}_i^t\|_{x_i} \leq \|\omega_i^t\| \leq 1$, and $\bar{x}_i^t \in W(\hat{x}_i^t)$, where $W(x) = \{x_i' \in \mathbb{R}^d, \|x_i' - x_i\|_{x_i} \leq 1\}$ is the Dikin ellipsoid. Since $W(x_i) \subseteq \mathcal{X}_i, \forall x_i \in int(\mathcal{X}_i)$, we can upper bound $\|A_i w_i\|^2$ by B^2, the diameter of the set $\mathcal{X}_i$. Hence $\|\nabla_i \hat{c}_i(\hat{x}^t) - \nabla_i c_i\left(\hat{x}^t\right)\| \leq \ell_i \delta_t \sqrt{n} B$.

With $\delta_t = \frac{1}{t^r}$, we have

$$\sum_{t\in[T_k]} \mathbb{E}\left[\left\langle \nabla_i c_i\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - \omega_i\right\rangle\right] = O\left(\nu dT^p\log(T^q) + GT^{q/2} + T^{q(p-2r)} + \ell_i\sqrt{n}B^2 T^{q(1-r)}\right).$$

Combining, as $m = T^{1-q}$ we have

$$\sum_{t=1}^T \mathbb{E}\left[\left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - x_i^{t,*}\right\rangle\right]$$
$$= O\left(GT^q V_i(T)\right) + \sum_{j\in[m]} \tilde{O}\left(\nu dT^{1-p} + GT^{q/2} + T^{q(p-2r)} + \ell_i\sqrt{n}B^2 T^{q(1-r)}\right)$$
$$= \tilde{O}\left(\nu dT^{(1-q)+p} + GT^{(1-q)+q/2} + T^{(1-q)+q(p-2r)} + \ell_i\sqrt{n}B^2 T^{(1-q)+q(1-r)} + GT^q V_i(T)\right).$$

When $V_i(T) = T^{\varphi}$, $\varphi \in [0,1]$, we set $q = \frac{2(1-\varphi)}{3}$, $p = \frac{(1-\varphi)}{3}$, $r = \frac{1}{2}$, we have

$$\sum_{t=1}^{T} \mathbb{E}\left[\left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - x_i^{t,*}\right\rangle\right] = \tilde{O}\left(\left(\nu d + G + \ell_i\sqrt{n}B^2\right) T^{\frac{1+2\varphi}{3}} + T^{\frac{(2\varphi+1)(\varphi+2)}{9}}\right).$$

Divided by T, we have

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\left\langle \nabla_i c_i^t\left(\hat{x}_i^t, \hat{x}_{-i}^t\right), \hat{x}_i^t - x_i^{t,*}\right\rangle\right] = \tilde{O}\left(\frac{\nu d + G + \ell_i\sqrt{n}B^2}{T^{\frac{2(1-\varphi)}{3}}} + \frac{1}{T^{\frac{9}{8} - \frac{(4\varphi+5)^2}{72}}}\right).$$

Sum over $i \in \mathcal{N}$ and we have the claimed result. $\square$

J Auxiliary Lemmas

Lemma J.1. *With the update rule equation 1,*

$$
\begin{aligned}
&\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right) \\
&\leq O\left(\frac{n\nu\log(T)}{\eta_T\kappa T}+\frac{n\zeta B}{\eta_T T^{3/2}}\right)+O\left(\frac{nB\sum_{i\in\mathcal{N}}\ell_i}{\kappa T^{3/2}}+\frac{n}{\kappa T^{3/2}}\right)\frac{\sum_{t=1}^T\eta_t}{\eta_T}+O\left(\frac{nC_p}{\eta_T T}\right) \\
&\quad+\frac{1}{\kappa\eta_T(T+1)}\sum_{i\in\mathcal{N}}\sum_{t=1}^T\eta_t\left\langle\hat{g}_i^t, x_i^t-x_i^{t+1}\right\rangle+\frac{1}{\kappa\eta_T(T+1)}\sum_{t=1}^T\eta_t\sum_{i\in\mathcal{M}}\left\langle\hat{g}_i^t-\nabla_i c_i\left(x^t\right), x_i^*-x_i^t\right\rangle \\
&\quad+\frac{1}{\kappa\eta_T(T+1)}\sum_{t=1}^T\eta_t\sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle\hat{g}_i^t-\nabla_i c_i\left(x^t\right), \bar{x}_i-x_i^t\right\rangle\,,
\end{aligned}
$$

where $\bar{x}_i$ is a point such that $\|\bar{x}_i-x_i^*\|=O(1/\sqrt{T})$ and $\inf\left\{t\geq 0: x_i^1+\frac{1}{t}(\bar{x}_i-x_i^1)\in\mathcal{X}_i\right\}\leq 1-1/\sqrt{T}$.

Proof. By the update rule equation 1, we have

$$
\eta_t\hat{g}_i^t+\eta_t\kappa(t+1)\left(\nabla p\left(x_i^{t+1}\right)-\nabla p\left(x_i^t\right)\right)+\left(\nabla h\left(x_i^{t+1}\right)-\nabla h\left(x_i^t\right)\right)=0\,.
$$

For a fixed point ω_i, by the three-point equality of Bregman divergence, we have

$$
\begin{aligned}
&D_h\left(\omega_i, x_i^{t+1}\right) \\
&=D_h\left(\omega_i, x_i^t\right)-D_h\left(x_i^{t+1}, x_i^t\right)+\left\langle\nabla h\left(x_i^t\right)-\nabla h\left(x_i^{t+1}\right), \omega_i-x_i^{t+1}\right\rangle \\
&=D_h\left(\omega_i, x_i^t\right)-D_h\left(x_i^{t+1}, x_i^t\right)+\eta_t\left\langle\hat{g}_i^t, \omega_i-x_i^{t+1}\right\rangle+\eta_t\kappa(t+1)\left\langle\nabla p\left(x_i^{t+1}\right)-\nabla p\left(x_i^t\right), \omega_i-x_i^{t+1}\right\rangle \\
&=D_h\left(\omega_i, x_i^t\right)-D_h\left(x_i^{t+1}, x_i^t\right)+\eta_t\left\langle\hat{g}_i^t, \omega_i-x_i^{t+1}\right\rangle+\eta_t\kappa(t+1)\left(D_p\left(\omega_i, x_i^t\right)-D_p\left(\omega_i, x_i^{t+1}\right)-D_p\left(x_i^{t+1}, x_i^t\right)\right)\,.
\end{aligned}
$$

Rearranging and by the non-negativity of Bregman divergence, we have,

$$
\begin{aligned}
&D_h\left(\omega_i, x_i^{t+1}\right)+\eta_t\kappa(t+1)D_p\left(\omega_i, x_i^{t+1}\right) \\
&\leq D_h\left(\omega_i, x_i^t\right)+\eta_t\kappa(t+1)D_p\left(\omega_i, x_i^t\right)+\eta_t\left\langle\hat{g}_i^t, \omega_i-x_i^t\right\rangle+\eta_t\left\langle\hat{g}_i^t, x_i^t-x_i^{t+1}\right\rangle \\
&=D_h\left(\omega_i, x_i^t\right)+\eta_t\kappa(t+1)D_p\left(\omega_i, x_i^t\right)+\eta_t\left\langle\nabla_i c_i\left(x^t\right), \omega_i-x_i^t\right\rangle+\eta_t\left\langle\hat{g}_i^t-\nabla_i c_i\left(x^t\right), \omega_i-x_i^t\right\rangle+\eta_t\left\langle\hat{g}_i^t, x_i^t-x_i^{t+1}\right\rangle\,.
\end{aligned}
$$

By Lemma J.3 and the assumption that $c_i(x)-\kappa p(x_i)$ is monotone, we have

$$
\eta_t\sum_{i\in\mathcal{N}}\left\langle\nabla_i c_i\left(x^t\right), \omega_i-x_i^t\right\rangle\leq-\eta_t\kappa\sum_{i\in\mathcal{N}}\left(D_p\left(x_i^t, \omega_i\right)+D_p\left(\omega_i, x_i^t\right)\right)+\eta_t\sum_{i\in\mathcal{N}}\left\langle\nabla_i c_i\left(\omega\right), \omega_i-x_i^t\right\rangle\,.
$$

Therefore,

$$
\begin{aligned}
&\sum_{i\in\mathcal{N}} D_h\left(\omega_i, x_i^{t+1}\right)+\eta_t\kappa(t+1)\sum_{i\in\mathcal{N}} D_p\left(\omega_i, x_i^{t+1}\right) \\
&\leq\sum_{i\in\mathcal{N}} D_h\left(\omega_i, x_i^t\right)+\eta_t\kappa t\sum_{i\in\mathcal{N}} D_p\left(\omega_i, x_i^t\right)+\eta_t\sum_{i\in\mathcal{N}}\left\langle\nabla_i c_i(\omega), \omega_i-x_i^t\right\rangle+\eta_t\sum_{i\in\mathcal{N}}\left\langle\hat{g}_i^t-\nabla_i c_i\left(x^t\right), \omega_i-x_i^t\right\rangle \\
&\quad+\eta_t\sum_{i\in\mathcal{N}}\left\langle\hat{g}_i^t, x_i^t-x_i^{t+1}\right\rangle\,.
\end{aligned}
$$

Summing over T, by the non-negativity of Bregman divergence, we have

$$
\begin{aligned}
&\eta_T\kappa(T+1)\sum_{i\in\mathcal{N}} D_p\left(\omega_i, x_i^{T+1}\right) \\
&\leq\sum_{i\in\mathcal{N}} D_h\left(\omega_i, x_i^1\right)+\kappa\sum_{i\in\mathcal{N}} D_p\left(\omega_i, x_i^1\right)+\sum_{t=1}^T\sum_{i\in\mathcal{N}}\eta_t\left\langle\nabla_i c_i(\omega), \omega_i-x_i^t\right\rangle+\sum_{t=1}^T\sum_{i\in\mathcal{N}}\eta_t\left\langle\hat{g}_i^t-\nabla_i c_i\left(x^t\right), \omega_i-x_i^t\right\rangle
\end{aligned}
$$

$$
+\sum_{t=1}^{T}\sum_{i\in\mathcal{N}}\eta_t\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle \ .
$$

Define $\pi_x(y) = \inf\left\{t \geq 0 : x + \frac{1}{t}(y-x) \in \mathcal{X}_i\right\}$, let us consider x_i^*, the equilibrium of the game.

- If $\pi_{x_i^1}(x_i^*) \leq 1 - 1/\sqrt{T}$, we set $\omega_i = x_i^*$. Let this set of player be $\mathcal{M}$
- Otherwise, we find $\bar{x}_i \in \mathcal{X}_i$ such that $\|\bar{x}_i - x_i^*\| = O(1/\sqrt{T})$ and $\pi_{x_i^1}(\bar{x}_i) \leq 1 - 1/\sqrt{T}$. We set $\omega_i = \bar{x}_i$.

By Lemma J.6, and initializing x_i^1 to minimize h, thus $D_h(\omega_i, x_i^1) = h(\omega_i) - h(x_i^1) \leq \nu\log(T)$.

Therefore, we have

$$
\begin{aligned}
&\eta_T\kappa(T+1)\left(\sum_{i\in\mathcal{M}} D_p\left(x_i^*, x_i^{T+1}\right) + \sum_{i\in\mathcal{N}\setminus\mathcal{M}} D_p\left(\bar{x}_i, x_i^{T+1}\right)\right)\\
&\leq n\nu\log(T) + \kappa\sum_{i\in\mathcal{M}} D_p\left(x_i^*, x_i^1\right) + \kappa\sum_{i\in\mathcal{N}\setminus\mathcal{M}} D_p\left(\bar{x}_i, x_i^1\right) + \sum_{t=1}^{T}\eta_t\sum_{i\in\mathcal{M}}\left\langle \nabla_i c_i(x^*_{\mathcal{M}}, \bar{x}_{\mathcal{N}\setminus\mathcal{M}}), x_i^* - x_i^t\right\rangle\\
&\quad + \sum_{t=1}^{T}\eta_t\sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle \nabla_i c_i(x^*_{\mathcal{M}}, \bar{x}_{\mathcal{N}\setminus\mathcal{M}}), \bar{x}_i - x_i^t\right\rangle + \eta_t\sum_{t=1}^{T}\sum_{i\in\mathcal{M}}\left\langle \hat{g}_i^t - \nabla_i c_i\left(x^t\right), x_i^* - x_i^t\right\rangle\\
&\quad + \eta_t\sum_{t=1}^{T}\sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle \hat{g}_i^t - \nabla_i c_i\left(x^t\right), \bar{x}_i - x_i^t\right\rangle + \sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\eta_t\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle \ .
\end{aligned}
$$

By the three-point inequality and the non-negativity of Bregman divergence, we have

$$
\begin{aligned}
\sum_{i\in\mathcal{N}\setminus\mathcal{M}} D_p\left(\bar{x}_i, x_i^{T+1}\right) &= \sum_{i\in\mathcal{N}\setminus\mathcal{M}} D_p\left(\bar{x}_i, x_i^*\right) + \sum_{i\in\mathcal{N}\setminus\mathcal{M}} D_p\left(x_i^*, x_i^{T+1}\right) - \sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle \bar{x}_i - x_i^*, \nabla p\left(x_i^{T+1}\right) - \nabla p\left(\bar{x}_i\right)\right\rangle\\
&\geq \sum_{i\in\mathcal{N}\setminus\mathcal{M}} D_p\left(x_i^*, x_i^{T+1}\right) - \sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle \bar{x}_i - x_i^*, \nabla p\left(x_i^{T+1}\right) - \nabla p\left(\bar{x}_i\right)\right\rangle \ .
\end{aligned}
$$

By Cauchy-Schwarz and the smoothness of p, we have

$$
\begin{aligned}
\sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle \bar{x}_i - x_i^*, \nabla p\left(x_i^{T+1}\right) - \nabla p\left(\bar{x}_i\right)\right\rangle &\leq \sum_{i\in\mathcal{N}\setminus\mathcal{M}}\|\bar{x}_i - x_i^*\|\left\|\nabla p\left(x_i^{T+1}\right) - \nabla p\left(\bar{x}_i\right)\right\|\\
&\leq \zeta\sum_{i\in\mathcal{N}\setminus\mathcal{M}}\|\bar{x}_i - x_i^*\|\left\|x_i^{T+1} - \bar{x}_i\right\|\\
&\leq O\left(\frac{n\zeta B}{\sqrt{T}}\right)
\end{aligned}
$$

As x_i^* is a Nash equilibrium, we have $\sum_{i\in\mathcal{N}}\left\langle \nabla_i c_i(x^*), x_i^* - x_i^t\right\rangle = 0$, therefore,

$$
\begin{aligned}
&\eta_t\sum_{i\in\mathcal{M}}\left\langle \nabla_i c_i(x^*_{\mathcal{M}}, \bar{x}_{\mathcal{N}\setminus\mathcal{M}}), x_i^* - x_i^t\right\rangle + \eta_t\sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle \nabla_i c_i(x^*_{\mathcal{M}}, \bar{x}_{\mathcal{N}\setminus\mathcal{M}}), \bar{x}_i - x_i^t\right\rangle\\
&= \eta_t\sum_{i\in\mathcal{N}}\left\langle \nabla_i c_i(x^*), x_i^* - x_i^t\right\rangle + \eta_t\sum_{i\in\mathcal{N}}\left\langle \nabla_i c_i(x^*_{\mathcal{M}}, \bar{x}_{\mathcal{N}\setminus\mathcal{M}}) - \nabla_i c_i(x^*), x_i^* - x_i^t\right\rangle\\
&\quad + \eta_t\sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle \nabla_i c_i(x^*_{\mathcal{M}}, \bar{x}_{\mathcal{N}\setminus\mathcal{M}}), \bar{x}_i - x_i^*\right\rangle\\
&\leq \eta_t\sum_{i\in\mathcal{N}}\ell_i\left\|x_i^* - x_i^t\right\|\left(\sum_{i\in\mathcal{N}\setminus\mathcal{M}}\|x_i^* - \bar{x}_i\|\right) + \eta_t\sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\|\nabla_i c_i(x^*_{\mathcal{M}}, \bar{x}_{\mathcal{N}\setminus\mathcal{M}})\right\|\|\bar{x}_i - x_i^*\|
\end{aligned}
$$

$$\leq O\left(\frac{\eta_t n B \sum_{i\in\mathcal{N}} \ell_i}{\sqrt{T}} + \frac{\eta_t n}{\sqrt{T}}\right) .$$

Hence, as $D_p(x_i, x_i') \leq C_p, \forall x_i, x_i'$,

$$\begin{aligned}
&\sum_{i\in\mathcal{N}} D_p\left(x_i^*, x_i^{T+1}\right) \\
&\leq O\left(\frac{n\nu\log(T)}{\eta_T\kappa T} + \frac{n\zeta B}{\eta_T T^{3/2}}\right) + O\left(\frac{nB\sum_{i\in\mathcal{N}}\ell_i}{\kappa T^{3/2}} + \frac{n}{\kappa T^{3/2}}\right)\frac{\sum_{t=1}^T \eta_t}{\eta_T} + O\left(\frac{nC_p}{\eta_T T}\right) \\
&\quad + \frac{1}{\kappa\eta_T(T+1)}\sum_{i\in\mathcal{N}}\sum_{t=1}^T \eta_t\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle + \frac{1}{\kappa\eta_T(T+1)}\sum_{t=1}^T \eta_t \sum_{i\in\mathcal{M}}\left\langle \hat{g}_i^t - \nabla_i c_i\left(x^t\right), x_i^* - x_i^t\right\rangle \\
&\quad + \frac{1}{\kappa\eta_T(T+1)}\sum_{t=1}^T \eta_t \sum_{i\in\mathcal{N}\setminus\mathcal{M}}\left\langle \hat{g}_i^t - \nabla_i c_i\left(x^t\right), \bar{x}_i - x_i^t\right\rangle .
\end{aligned}$$

□

Lemma J.2. *Take $\eta_t \leq \frac{1}{2d}$, we have*

$$\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle = \eta_t \left\|A_i^t \hat{g}_i^t\right\|^2 .$$

Proof. Define

$$f(x_i) = \eta_t\left\langle x_i, \hat{g}_i^t\right\rangle + \eta_t(t+1)D_p(x_i, x_i^t) + D_h(x_i, x_i^t) .$$

As adding the linear term $\langle x_i, \hat{g}_i^t\rangle$ does not affect the self-concordant barrier property, and p is strongly monotone, $f(x)$ is a self-concordant barrier.

Define the local norm $\|h\|_x := \sqrt{h^\top \nabla^2 f(x) h}$, by Holder's inequality, we have

$$\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle = \left\|\hat{g}_i^t\right\|_{x_i^t,*}\left\|x_i^t - x_i^{t+1}\right\|_{x_i^t} .$$

Notice that

$$\nabla f(x_i^t) = \eta_t \hat{g}_i^t \, , \nabla^2 f(x_i^t) = \eta_t(t+1)\nabla^2 p(x_i^t) + \nabla^2 h(x_i^t) .$$

Therefore, by our assumption that $c_i(x) \in [0, 1]$,

$$\begin{aligned}
\left\|\left(\nabla^2 f(x_i^t)\right)^{-1}\nabla f(x_i^t)\right\|_{x_i^t} &= \eta_t\left\|A_i^t \hat{g}_i^t\right\| \\
&\leq \eta_t d|c_i(\hat{x}^t)| \leq \eta_t d .
\end{aligned}$$

By Lemma J.4, take $\eta_t \leq \frac{1}{2d}$, we have

$$\left\|x_i^t - x_i^{t+1}\right\|_{x_i^t} = \left\|x_i^t - \arg\min_x f(x_i^t)\right\|_{x_i^t} \leq 2\left\|\left(\nabla^2 f(x_i^t)\right)^{-1}\nabla f(x_i^t)\right\|_{x_i^t} \leq \eta_t\left\|A_i^t \hat{g}_i^t\right\| .$$

Therefore, we have

$$\left\langle \hat{g}_i^t, x_i^t - x_i^{t+1}\right\rangle = \eta_t \left\|A_i^t \hat{g}_i^t\right\|^2 .$$

□

Lemma J.3. *[Proposition 1 Bauschke et al. (2017)] For an operator G that $G - \nabla p(x)$ is monotone,*

$$\langle G(x) - G(x'), x' - x\rangle \leq -\sum_{i\in\mathcal{N}}\left(D_p\left(x_i, x_i'\right) + D_p\left(x_i', x_i\right)\right) .$$

Proof. By the monotonicity of $G - \nabla p(x)$, we have

$$\begin{aligned}\langle G(x) - G(x'), x' - x\rangle &\leq \langle \nabla p(x) - \nabla p(x'), x' - x\rangle \\ &\leq -\sum_{i\in\mathcal{N}} \left(D_p\left(x_i, x_i'\right) + D_p\left(x_i', x_i\right)\right),\end{aligned}$$

where the second inequality is due to the definition of Bregman divergence. $\square$

Lemma J.4 (Lemma 3 Lin et al. (2021))**.** *For any self-concordant function g and let $\lambda(x,g) \leq \frac{1}{2}$, $\lambda(x,g) := \|\nabla g(x)\|_{x,\star} = \left\|\left(\nabla^2 g(x)\right)^{-1} \nabla g(x)\right\|_x$, we have $\|x - \arg\min_{x'\in\mathcal{X}} g\left(x'\right)\|_x \leq 2\lambda(x,g)$, where $\|\cdot\|_x$ is the local norm given by $\|h\|_x := \sqrt{h^\top \nabla^2 g(x) h}$.*

Lemma J.5 (Lemma 7 of Lin et al. (2021))**.** *Suppose that c_i is a monotone function and $A_i \in \mathbb{R}^{d\times d}$ is an invertible matrix for each $i \in \mathcal{N}$, we define the smoothed version of c_i with respect to A_i by $\hat{c}_i(x) = \mathbb{E}_{w_i\sim\mathbb{B}^d}\mathbb{E}_{\mathbf{z}_{-i}\sim\Pi_{j\neq i}\mathbb{S}^d}\left[c_i\left(x_i + A_i w_i, \hat{x}_{-i}\right)\right]$ where $\mathbb{S}^d$ is a d-dimensional unit sphere, $\mathbb{B}^d$ is a d-dimensional unit ball and $\hat{x}_i = x_i + A_i z_i$ for all $i \in \mathcal{N}$. Then, the following statements hold true:*

- $\nabla_i \hat{c}_i(x) = \mathbb{E}\left[d \cdot c_i\left(\hat{x}_i, \hat{x}_{-i}\right)\left(A_i\right)^{-1} z_i \mid x_1, x_2, \ldots, x_N\right]$.
- *If ∇c_i is ℓ_i-Lipschitz continuous and we let $\sigma_{\max}(A)$ be the largest eigenvalue of A, we have $\|\nabla_i \hat{c}_i(x) - \nabla_i c_i(x)\| \leq \ell_i \sqrt{\sum_{j\in\mathcal{N}}\left(\sigma_{\max}\left(A_j\right)\right)^2}$.*

Lemma J.6 (Lemma 2 Lin et al. (2021))**.** *Suppose that $\mathcal{X}$ is a closed, monotone and compact set, R is a ν-self-concordant barrier function for $\mathcal{X}$ and $\bar{x} = \arg\min_{x\in\mathcal{X}} R(x)$ is a center. Then, we have $R(x) - R(\bar{x}) \leq \nu \log\left(\frac{1}{1-\pi_{\bar{x}}(x)}\right)$. For any $\epsilon \in (0,1]$ and $x \in \mathcal{X}_\epsilon$, we have $\pi_{\bar{x}}(x) \leq \frac{1}{1+\epsilon}$ and $R(x) - R(\bar{x}) \leq \nu \log\left(1 + \frac{1}{\epsilon}\right)$.*

K More Experimental Results

In Figure 2 and 3 we supplement more experiment results for zero-sum matrix games and Cournot competition. Note that in Figure 3, the curve of OMD with gradient coincides exactly with the curve GD with gradient. We found similar observations that our algorithm attains comparable performance to OMD and GD with full information gradient.

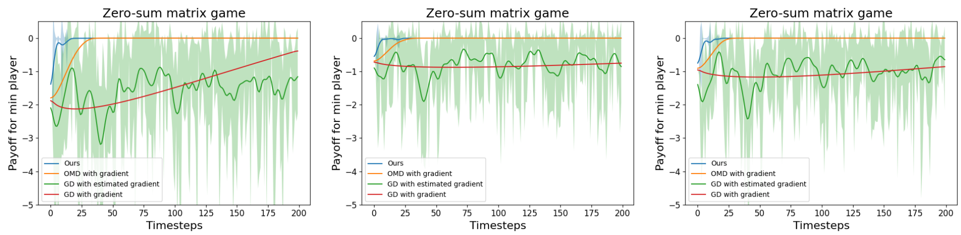

Figure 2: More examples on the zero-sum matrix game, with A being $[2, 1], [1, 3]$, $[3, 0], [0, 1]$, and $[1, 2], [2, 0]$.

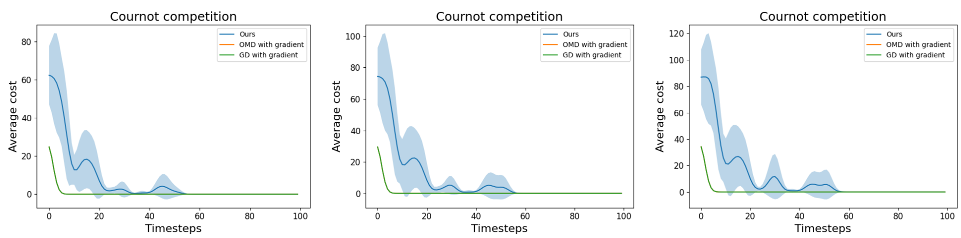

Figure 3: More examples on the Cournot competition, with the marginal cost being $50, 60, 70$.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately and clearly state the main theoretical claims made and discuss the main contributions and scope.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper clearly states the assumption made and gives examples of when the assumptions are satisfied. The paper also states clearly of the experimental settings, the computational resources needed, etc.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All assumptions are clearly stated, and all proofs are included in the appendix. All theorems and lemmas are properly referenced.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: All implementation details are given in the experiment section. Code will be released upon acceptance of this paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: We will provide open access to the code and data upon acceptance of this paper.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The experimental settings are described in the experiment settings in detail to reproduce the paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Yes, all figures for the experiments are shown with shading the one standard deviation error.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All computational resources used have been stated in the experiment section.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: I have reviewed the code of ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: The paper is foundational and theoretical research without particular application or deployment.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper is theoretical with no such risk.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: The experimental environment used in this paper is synthetic.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The experiment environments have been described in detail in this paper.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: the paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.