Adaptively Perturbed Mirror Descent for Learning in Games

Kenshi Abe [1 2] **Kaito Ariu** [1] **Mitsuki Sakamoto** [1] **Atsushi Iwasaki** [2]

Abstract

This paper proposes a payoff perturbation technique for the Mirror Descent (MD) algorithm in games where the gradient of the payoff functions is monotone in the strategy profile space, potentially containing additive noise. The optimistic family of learning algorithms, exemplified by optimistic MD, successfully achieves *last-iterate* convergence in scenarios devoid of noise, leading the dynamics to a Nash equilibrium. A recent re-emerging trend underscores the promise of the perturbation approach, where payoff functions are perturbed based on the distance from an anchoring, or *slingshot*, strategy. In response, we propose *Adaptively Perturbed MD* (APMD), which adjusts the magnitude of the perturbation by repeatedly updating the slingshot strategy at a predefined interval. This innovation empowers us to find a Nash equilibrium of the underlying game with guaranteed rates. Empirical demonstrations affirm that our algorithm exhibits significantly accelerated convergence.

1. Introduction

This study delves into a variant of Mirror Descent (MD) (Nemirovskij & Yudin, 1983; Beck & Teboulle, 2003) in the context of monotone games whose gradient of the payoff functions exhibits monotonicity concerning the strategy profile space. This encompasses diverse games, including Cournot competition (Bravo et al., 2018), λ-cocoercive games (Lin et al., 2020), concave-convex games, and zero-sum polymatrix games (Cai & Daskalakis, 2011; Cai et al., 2016). Due to their extensive applicability, various learning algorithms have been developed and scrutinized to compute a Nash equilibrium efficiently.

Agents, which are prescribed to play according to MD or its variant, choose strategies with higher expected payoffs more likely but do not move far away from current strategies via regularization. The dynamics is known to converge to an equilibrium in an average sense, which is referred to as *average-iterate convergence*. In other words, the averaged strategy profile over iterations converges to an equilibrium. Nevertheless, research has shown that the actual trajectory of the updated strategy profiles fails to converge even in two-player zero-sum games and a specific class within monotone games (Mertikopoulos et al., 2018; Bailey & Piliouras, 2018). On the contrary, optimistic learning algorithms, incorporating recency bias, have shown success. The updated strategy profile itself converges to a Nash equilibrium (Daskalakis et al., 2018; Daskalakis & Panageas, 2019; Mertikopoulos et al., 2019; Wei et al., 2021), termed *last-iterate convergence*.

However, the optimistic approach faces challenges with feedback contaminated by some noise. Typically, each agent updates his or her strategy according to the perfect gradient feedback of the payoff function at each iteration, denoted as *full feedback*. In a more realistic scenario, noise might distort this feedback. With *noisy feedback*, optimistic learning algorithms perform suboptimally. For instance, Abe et al. (2023) empirically demonstrated that optimistic Multiplicative Weights Update (OMWU) fails to converge to an equilibrium, orbiting around it.

Alternatively, perturbation of payoffs has emerged again as a pivotal concept for achieving last-iterate convergence, even under noise (Abe et al., 2023). Payoff perturbation is a classical technique, as seen in Facchinei & Pang (2003), and introduces strongly convex penalties to the players' payoff functions to stabilize learning. This leads to convergence to approximate equilibria, not only in the full feedback setting but also in the noisy feedback setting. However, to ensure convergence toward a Nash equilibrium of the underlying game, the magnitude of perturbation requires careful adjustment, which is calculated as the product of a strongly convex penalty function and a perturbation strength parameter. In fact, Liu et al. (2023) shrink the perturbation strength based on the current strategy profile's proximity to an underlying equilibrium. Similarly, iterative Tikhonov regularization methods (Koshal et al., 2010; Tatarenko & Kamgarpour, 2019) adjust the magnitude of perturbation by using a sequence of perturbation strengths that satisfy

[1]CyberAgent, Tokyo, Japan [2]University of Electro-Communications, Tokyo, Japan. Correspondence to: Kenshi Abe <abekenshi1224@gmail.com>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

certain conditions, such as diminishing as the iteration increases. Although these algorithms admit last-iterate convergence, it becomes challenging to choose an appropriate learning rate for the shrinking perturbation strength, which often leads to a failure in achieving rapid convergence for these algorithms and their variants.

In response to this, we adaptively determine the amount of the penalty from the distance $G(\cdot,\cdot)$ between the current strategy π and an anchoring, or *slingshot* strategy σ, while maintaining the perturbation strength parameter μ constant. Instead of carefully decaying the perturbation strength, the slingshot strategies σ are re-initialized at a predefined interval T_σ by the current strategies, and thus the magnitude of the perturbation $\mu G(\cdot,\cdot)$ is adjusted. To the best of our knowledge, Perolat et al. (2021) were the first to employ this idea and enabled Abe et al. (2023) to modify MWU to achieve last-iterate convergence. However, they have established the convergence only in an asymptotic manner. The significance of our work, in part, lies in extending these two studies and establishing non-asymptotic convergence results.

Our contributions are manifold. First, we identify our algorithm as *Adaptively Perturbed MD*[1] (APMD). Second, we analyze the case where both the perturbation function and the proximal regularizer are assumed to be the squared ℓ^2-distance and provide the last-iterate convergence rates to a Nash equilibrium, $\mathcal{O}(\ln T/\sqrt{T})$ and $\mathcal{O}(\ln T/T^{\frac{1}{10}})$ with full and noisy feedback, respectively. We also discuss the case where different distances from the squared ℓ^2-distance are utilized. Finally, we empirically reveal that our proposed APMD significantly outperforms MWU and OMWU in two zero-sum polymatrix games, regardless of the feedback type.

2. Preliminaries

Monotone games. This paper focuses on a continuous game, which is denoted by $([N],(\mathcal{X}_i)_{i\in[N]},(v_i)_{i\in[N]})$. $[N]=\{1,2,\cdots,N\}$ represents the set of N players, $\mathcal{X}_i\subseteq\mathbb{R}^{d_i}$ represents the d_i-dimensional compact convex strategy space for player $i\in[N]$, and we write $\mathcal{X}=\prod_{i\in[N]}\mathcal{X}_i$. Each player i chooses a *strategy* π_i from $\mathcal{X}_i$ and aims to maximize her differentiable payoff function $v_i:\mathcal{X}\to\mathbb{R}$. We write $\pi_{-i}\in\prod_{j\neq i}\mathcal{X}_i$ as the strategies of all players except player i, and denote the *strategy profile* by $\pi=(\pi_i)_{i\in[N]}\in\mathcal{X}$. This study particularly considers a *smooth monotone game*, where the gradient $(\nabla_{\pi_i}v_i)_{i\in[N]}$ of the payoff functions is monotone: $\forall\pi,\pi'\in\mathcal{X}$,

$$\sum_{i=1}^{N}\langle\nabla_{\pi_i}v_i(\pi_i,\pi_{-i})-\nabla_{\pi_i}v_i(\pi'_i,\pi'_{-i}),\pi_i-\pi'_i\rangle\leq 0,\quad(1)$$

and L-Lipschitz:

$$\sum_{i=1}^{N}\|\nabla_{\pi_i}v_i(\pi_i,\pi_{-i})-\nabla_{\pi_i}v_i(\pi'_i,\pi'_{-i})\|^2\leq L^2\|\pi-\pi'\|^2,\quad(2)$$

where $\|\cdot\|$ is the ℓ^2-norm.

Monotone games include many common and well-studied classes of games, such as concave-convex games, zero-sum polymatrix games, and Cournot competition.

Example 2.1 (Concave-Convex Games)**.** Let us consider a max-min game $(\{1,2\},(\mathcal{X}_1,\mathcal{X}_2),(v,-v))$, where $v:\mathcal{X}_1\times\mathcal{X}_2\to\mathbb{R}$. Player 1 aims to maximize v, while player 2 aims to minimize v. If v is concave in $x_1\in\mathcal{X}_1$ and convex in $x_2\in\mathcal{X}_2$, the game is called a concave-convex game or minimax optimization problem, and it is easy to confirm that the game is monotone.

Example 2.2 (Zero-Sum Polymatrix Games)**.** In a zero-sum polymatrix game, each player's payoff function can be decomposed as $v_i(\pi)=\sum_{j\neq i}u_i(\pi_i,\pi_j)$, where $u_i:\mathcal{X}_i\times\mathcal{X}_j\to\mathbb{R}$ is represented by $u_i(\pi_i,\pi_j)=\pi_i^\top\mathrm{M}^{(i,j)}\pi_j$ with some matrix $\mathrm{M}^{(i,j)}\in\mathbb{R}^{d_i\times d_j}$, and satisfies $u_i(\pi_i,\pi_j)=-u_j(\pi_j,\pi_i)$. In this game, each player i can be interpreted as playing a two-player zero-sum game with each other player $j\neq i$. From the linearity and zero-sum property of u_i, we can easily show that $\sum_{i=1}^{N}\langle\nabla_{\pi_i}v_i(\pi_i,\pi_{-i})-\nabla_{\pi_i}v_i(\pi'_i,\pi'_{-i}),\pi_i-\pi'_i\rangle=0$. Thus, the zero-sum polymatrix game is a special case of monotone games.

Nash equilibrium and gap function. A *Nash equilibrium* (Nash, 1951) is a common solution concept of a game, which is a strategy profile where no player can improve her payoff by deviating from her specified strategy. Formally, a Nash equilibrium $\pi^*\in\mathcal{X}$ satisfies the following condition:

$$\forall i\in[N],\forall\pi_i\in\mathcal{X}_i,\ v_i(\pi_i^*,\pi_{-i}^*)\geq v_i(\pi_i,\pi_{-i}^*).$$

We denote the set of Nash equilibria by Π^*. Note that a Nash equilibrium always exists for any smooth monotone game (Debreu, 1952). Furthermore, we measure the proximity to Nash equilibrium for a given strategy profile π by its *gap function*, which is defined as:

$$\mathrm{GAP}(\pi):=\max_{\tilde{\pi}\in\mathcal{X}}\sum_{i=1}^{N}\langle\nabla_{\pi_i}v_i(\pi_i,\pi_{-i}),\tilde{\pi}_i-\pi_i\rangle.$$

The gap function is a standard metric of proximity to Nash equilibrium for a given strategy profile π (Cai & Zheng, 2023). From the definition, $\mathrm{GAP}(\pi)\geq 0$ for any $\pi\in\mathcal{X}$, and the equality holds if and only if π is a Nash equilibrium.

Problem setting. In this study, we consider the online learning setting where the following process is repeated for T iterations: 1) At each iteration $t\geq 0$, each player

[1] An implementation of our method is available at https://github.com/CyberAgentAILab/adaptively-perturbed-md

Algorithm 1 APMD for player i.

Require: Learning rate sequence $\{\eta_t\}_{t\geq 0}$, divergence function for perturbation G, perturbation strength μ, update interval T_σ, initial strategy π_i^0

1: $k \leftarrow 0,\ \tau \leftarrow 0$
2: $\sigma_i^0 \leftarrow \pi_i^0$
3: **for** $t = 0, 1, 2, \cdots, T$ **do**
4: Receive the gradient feedback $\widehat{\nabla}_{\pi_i} v_i(\pi^t)$
5: Update the strategy by

$$\pi_i^{t+1} = \arg\max_{x\in\mathcal{X}_i} \left\{ \eta_t \left\langle \widehat{\nabla}_{\pi_i} v_i(\pi^t) - \mu \nabla_{\pi_i} G(\pi_i^t, \sigma_i^k), x \right\rangle - D_\psi(x, \pi_i^t) \right\}$$

6: $\tau \leftarrow \tau + 1$
7: **if** $\tau = T_\sigma$ **then**
8: $k \leftarrow k + 1,\ \tau \leftarrow 0$
9: $\sigma_i^k \leftarrow \pi_i^{t+1}$
10: **end if**
11: **end for**

$i \in [N]$ chooses her strategy $\pi_i^t \in \mathcal{X}_i$ based on the previously observed feedback; 2) Each player i receives the gradient feedback $\widehat{\nabla}_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t)$ as feedback. This study considers two feedback models: *full feedback* and *noisy feedback*. In the full feedback setting, each player receives the perfect gradient vector as feedback, i.e., $\widehat{\nabla}_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) = \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t)$. In the noisy feedback setting, each player's feedback is given by $\widehat{\nabla}_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) = \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) + \xi_i^t$, where $\xi_i^t \in \mathbb{R}^{d_i}$ is a noise vector. Specifically, we focus on the zero-mean and bounded-variance noise vectors.

Mirror Descent. Mirror Descent (MD) is a widely used algorithm for learning equilibria in games. Let us define $\psi : \mathbb{R}^{d_i} \to \mathbb{R} \cup \{\infty\}$ as the strictly convex *regularization function* and $D_\psi(\pi_i, \pi_i') = \psi(\pi_i) - \psi(\pi_i') - \langle \nabla\psi(\pi_i'), \pi - \pi_i' \rangle$ as the associated Bregman divergence. Then, MD updates each player i's strategy π_i^t at iteration t as follows:

$$\pi_i^{t+1} = \arg\max_{x\in\mathcal{X}_i} \left\{ \eta_t \left\langle \widehat{\nabla}_{\pi_i} v_i(\pi^t), x \right\rangle - D_\psi(x, \pi_i^t) \right\},$$

where $\eta_t \in (0, \infty)$ is the learning rate at iteration t.

Other notations. We denote a d-dimensional probability simplex by $\Delta^d = \{p \in [0,1]^d \mid \sum_{j=1}^d p_j = 1\}$. We define $\mathrm{diam}(\mathcal{X}) := \sup_{\pi,\pi'\in\mathcal{X}} \|\pi - \pi'\|$ as the diameter of $\mathcal{X}$. The *Kullback-Leibler* (KL) divergence is defined by $\mathrm{KL}(\pi_i, \pi_i') = \sum_{j=1}^{d_i} \pi_{ij} \ln \frac{\pi_{ij}}{\pi_{ij}'}$. Besides, with a slight abuse of notation, we represent the sum of Bregman divergences and the sum of KL divergences by $D_\psi(\pi, \pi') = \sum_{i=1}^N D_\psi(\pi_i, \pi_i')$, and $\mathrm{KL}(\pi, \pi') = \sum_{i=1}^N \mathrm{KL}(\pi_i, \pi_i')$, respectively. We finally denote the domain of ψ by $\mathrm{dom}\ \psi := \{x : \psi(x) < \infty\}$, and corresponding interior by $\mathrm{int}(\mathrm{dom}\ \psi)$.

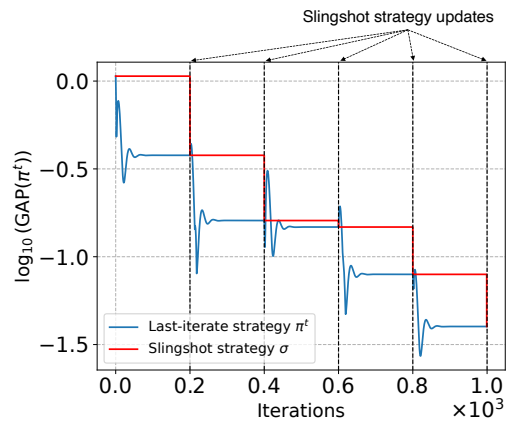


Figure 1: Illustration of the impact of the slingshot strategy updates on the gap function for π^t updated by APMD.

3. Adaptively Perturbed Mirror Descent

In this section, we present Adaptively Perturbed Mirror Descent (APMD), which is an extension of the standard MD algorithms. Algorithm 1 describes the pseudo-code. APMD employs two pivotal techniques: *slingshot payoff perturbation* and *slingshot strategy update*. Each of them corresponds to line 5 and line 9 in Algorithm 1, respectively.

3.1. Slingshot Payoff Perturbation

Letting us define the differentiable divergence function $G(\cdot,\cdot) : \mathbb{R}^{d_i} \times \mathbb{R}^{d_i} \to \mathbb{R} \cup \{\infty\}$ and the *slingshot* strategy $\sigma_i \in \mathcal{X}_i$, APMD perturbs each player's payoff by the divergence between the current strategy π_i^t and the slingshot strategy σ_i, i.e., $G(\pi_i^t, \sigma_i)$. Specifically, APMD updates each player's strategy according to

$$\pi_i^{t+1} = \arg\max_{x\in\mathcal{X}_i} \left\{ \eta_t \left\langle \widehat{\nabla}_{\pi_i} v_i(\pi^t) - \mu \nabla_{\pi_i} G(\pi_i^t, \sigma_i), x \right\rangle - D_\psi(x, \pi_i^t) \right\}. \tag{3}$$

where $\mu \in (0, \infty)$ is the *perturbation strength* and $\nabla_{\pi_i} G$ denotes differentiation with respect to first argument. We assume that $G(\cdot, \sigma_i)$ is strictly convex for every $\sigma_i \in \mathcal{X}_i$, and takes a minimum value of 0 at σ_i. Furthermore, we assume that ψ is differentiable and ρ-strongly convex on $\mathcal{X}_i$ with $\rho \in (0, \infty)$.

The conventional MD updates its strategy based on the gra-

dient feedback of the payoff function and the regularization term. The regularization term adjusts the next strategy so that it does not deviate significantly from the current strategy. APMD perturbs the gradient payoff vector by the divergence between the current strategy π_i^t and a predefined slingshot strategy σ_i. If the current strategy is far away from the slingshot strategy, the magnitude of the perturbation increases. Note that if both strategies are equivalent, no perturbation occurs, and APMD just seeks a strategy with a higher expected payoff. As Figure 1 illustrates, the divergence first fluctuates, and then the current strategy profile comes close to a stationary point where the gradient of the expected payoff vector is equal to the gradient of the magnitude of the perturbation so that the perturbation via slingshot stabilizes the learning dynamics. Indeed, Mutant Follow the Regularized Leader (Mutant FTRL) instantiated in Example 5.3 encompasses replicator-mutator dynamics, which is guaranteed to converge to an approximate equilibrium in two-player zero-sum games (Abe et al., 2022). We can argue that APMD inherits this nice feature.

3.2. Slingshot Strategy Update

The perturbation via slingshot enables π^t to converge quickly to a stationary point (Lemmas 5.6 and 5.7). Different slingshot strategy profiles induce different stationary points. Of course, when the slingshot strategy profile is set to be an equilibrium, the corresponding stationary point also becomes an equilibrium. However, it is virtually impossible to identify such an ideal slingshot strategy profile beforehand. To this end, APMD adjusts a slingshot strategy profile σ by replacing it with the (nearly) stationary point that is reached after predefined iterations T_σ.

Figure 1 illustrates how our slingshot strategy update brings the corresponding stationary points closer to an equilibrium. x- and y-axis indicate the number of iterations and the logarithm of the gap function of the last-iterate strategy profile. We here assume that the learning rate and the perturbation strength are $\eta = 0.1$ and $\mu = 1$, respectively. The initial slingshot strategy σ_i is given as a uniform distribution on the action space. After the first interval of $T_\sigma = 200$ iterations, APMD finds a stationary point. The slingshot strategy profile for the second interval is replaced with the stationary point. Figure 1 clearly shows the stationary point, i.e., the last-iterate strategy profile comes close to an equilibrium every time the slingshot strategy profile is updated. We also theoretically justify our slingshot strategy update in Theorem D.1, i.e., when the slingshot strategy profile is close to an equilibrium $\pi^* \in \Pi^*$, the stationary point is close to π^*.

4. Last-Iterate Convergence Rates

In this section, we establish the last-iterate convergence rates of APMD. More specifically, we examine a setting where both D_ψ and G is set to the squared ℓ^2-distance, i.e., $D_\psi(\pi_i, \pi_i') = G(\pi_i, \pi_i') = \frac{1}{2}\|\pi_i - \pi_i'\|^2$. This instance can be considered as an extended version of Gradient Descent, which incorporates our techniques of payoff perturbation and slingshot strategy update. We also assume that the gradient vector of v_i is bounded. We emphasize that we have obtained the overall last-iterative convergence rates of APMD for the entire T iterations in both full and noisy feedback settings.

4.1. Full Feedback Setting

First, we demonstrate the last-iterate convergence rate of APMD with *full feedback* where each player receives the perfect gradient vector $\widehat{\nabla}_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) = \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t)$, at each iteration t. Theorem 4.1 provides the APMD's convergence rate of $\mathcal{O}(\ln T/\sqrt{T})$ in terms of the gap function. Note that the learning rate is constant, and its upper bound is specified by perturbation strength μ and smoothness parameter L.

Theorem 4.1. *If we use the constant learning rate $\eta_t = \eta \in (0, \frac{2\mu}{3\mu^2+8L^2})$, and set D_ψ and G as the squared ℓ^2-distance $D_\psi(\pi_i, \pi_i') = G(\pi_i, \pi_i') = \|\pi_i - \pi_i'\|^2/2$, and set $T_\sigma = \Theta(\ln T)$, then the strategy profile π^T updated by APMD satisfies:*

$$\mathrm{GAP}(\pi^T) = \mathcal{O}\left(\frac{\ln T}{\sqrt{T}}\right).$$

The obtained rate herein is competitive with optimistic gradient and extra-gradient methods (Cai et al., 2022a) whose rates are $\mathcal{O}(1/\sqrt{T})$. Although it is open whether our convergence rate matches its lower bound, it closely aligns with the lower bound for the class of algorithms that includes 1-SCLI algorithms (Golowich et al., 2020b), which is different from our APMD. To the best of our knowledge, the fastest rate of $\mathcal{O}(1/T)$ is achieved by *Accelerated Optimistic Gradient* (AOG) (Cai & Zheng, 2023), which is an optimistic variant of the *Halpern iteration* (Halpern, 1967). At first sight, the update rule of AOG looks as if the perturbation strength was linearly decayed. However, it does not perturb the payoff functions and instead adjusts the regularization term by using the convex combination of the current strategy and the anchoring strategy as the proximal point in MD. Unlike our APMD, the anchoring strategy is never updated through the iterations.[2]

4.2. Proof Sketch of Theorem 4.1

This section sketches the proof for Theorem 4.1. We present the complete proofs for the theorem and associated lemmas in Appendix E.

[2]Regarding the rate of $\mathcal{O}(1/T)$, a companion paper is in preparation.

(1) Convergence rates to a stationary point with k-th slingshot strategy profile. We denote σ^k as the slingshot strategy profile after k updates. Since the slingshot strategy profile is overwritten by the current strategy profile π^t every T_σ iterations, we can write $\sigma^k = \pi^{kT_\sigma}$. We first prove that, as T_σ increases, the next $k+1$-th slingshot strategy profile σ^{k+1} approaches to the stationary point π^{μ,σ^k} under the slingshot strategy profile σ^k, which satisfies the following condition: $\forall i \in [N]$,

$$\pi_i^{\mu,\sigma^k} = \arg\max_{\pi_i \in \mathcal{X}_i} \left\{ v_i(\pi_i, \pi_{-i}^{\mu,\sigma^k}) - \mu G(\pi_i, \sigma_i^k) \right\}. \quad (4)$$

Note that π^{μ,σ^k} always exists since the perturbed game is still monotone. Using the strong convexity of $G(\pi_i, \sigma_i^k) = \frac{1}{2}\|\pi_i - \sigma_i^k\|^2$, we show that $\sigma^{k+1} \to \pi^{\mu,\sigma^k}$ as $T_\sigma \to \infty$:

Lemma 4.2. *Assume that D_ψ and G are set as the squared ℓ^2-distance. If we use the constant learning rate $\eta_t = \eta \in (0, \frac{2\mu}{3\mu^2+8L^2})$, the $k+1$-th slingshot strategy profile σ^{k+1} of APMD under the full feedback setting satisfies that:*

$$\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|^2 = \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \cdot \exp\left(-\mathcal{O}(T_\sigma)\right).$$

(2) Upper bound on the gap function. Next, we derive the upper bound on the gap function for σ^{k+1}. From the bounding technique of the gap function using the tangent residuals by Cai et al. (2022a) and the first-order optimality condition for π^{μ,σ^k}, the gap function for σ^{k+1} can be upper bounded by the distance between the slingshot strategy profile and the stationary point π^{μ,σ^k}:

Lemma 4.3. *If G is set as the squared ℓ^2-distance, the gap function for σ^{k+1} of APMD satisfies for $k \geq 0$:*

$$\mathrm{GAP}(\sigma^{k+1}) = \mathcal{O}\left(\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\| + \|\pi^{\mu,\sigma^k} - \sigma^k\| \right).$$

By Lemma 4.2, if we set D_ψ as the squared ℓ^2-distance, the first term in this lemma can be bounded as:

$$\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|^2 = \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \exp\left(-\mathcal{O}(T_\sigma)\right). \quad (5)$$

Therefore, it is enough to derive the convergence rate on the ℓ^2-distance between π^{μ,σ^k} and σ^k.

(3) Last-iterate convergence results for the slingshot strategy profile. Let us denote $K := \lfloor T/T_\sigma \rfloor$ as the total number of the slingshot strategy updates over the entire T iterations. Then, by adjusting $T_\sigma = \Omega(\ln T)$, we show that the ℓ^2-distance between the $K-1$-th slingshot strategy profile σ^{K-1} and the corresponding stationary point $\pi^{\mu,\sigma^{K-1}}$ decreases as K increases:

Lemma 4.4. *In the same setup of Theorem 4.1, the $K-1$-th slingshot strategy profile σ^{K-1} of APMD satisfies:*

$$\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\| = \mathcal{O}(1/\sqrt{K}).$$

Lemma 4.4 implies that as K increases, the variation of σ^K becomes negligible, signifying convergence in its behavior.

By combining (5) and Lemmas 4.3-4.4, we can derive the last-iterate convergence rate of the slingshot strategy profile σ^K:

$$\mathrm{GAP}(\sigma^K) = \mathcal{O}(1/\sqrt{K}).$$

Thus, since $\pi^T = \sigma^K$ and $K = \lfloor T/T_\sigma \rfloor = \Theta(T/\ln T)$, the statement of the theorem is concluded. □

4.3. Noisy Feedback Setting

Next, we consider the *noisy feedback* setting, where each player i receives a gradient vector with additive noise: $\nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) + \xi_i^t$. Define the sigma-algebra generated by the history of the observations: $\mathcal{F}_t = \sigma\left((\widehat{\nabla}_{\pi_i} v_i(\pi_i^0, \pi_{-i}^0))_{i\in[N]}, \ldots, (\widehat{\nabla}_{\pi_i} v_i(\pi_i^{t-1}, \pi_{-i}^{t-1}))_{i\in[N]} \right)$, $\forall t \geq 1$. We assume that the noise vectors $(\xi_i^t)_{t\geq 1}$ are with zero-mean and bounded variances. We also suppose that the noise vectors $(\xi_i^t)_{t\geq 1}$ are independent over t. In this setting, the last-iterate convergence rate is achieved by APMD using a decreasing learning rate sequence η_t. The convergence rate obtained by APMD is $\mathcal{O}(\ln T/T^{\frac{1}{10}})$:

Theorem 4.5. *Let $\theta = \frac{3\mu^2+8L^2}{2\mu}$ and $\kappa = \frac{\mu}{2}$. Assume that D_ψ and G are set as the squared ℓ^2-distance $D_\psi(\pi_i, \pi_i') = G(\pi_i, \pi_i') = \frac{1}{2}\|\pi_i - \pi_i'\|^2$, and $T_\sigma = \Theta(T^{4/5})$. If we choose the learning rate sequence of the form $\eta_t = 1/(\kappa(t - T_\sigma \cdot \lfloor t/T_\sigma \rfloor) + 2\theta)$, then the strategy profile π^T updated by APMD satisfies:*

$$\mathbb{E}\left[\mathrm{GAP}(\pi^T)\right] = \mathcal{O}\left(\frac{\ln T}{T^{\frac{1}{10}}}\right).$$

It should be noted that Theorem 4.5 provides a non-asymptotic convergence guarantee with a rate. This is a significant departure from the existing convergence results (Koshal et al., 2010; 2013; Tatarenko & Kamgarpour, 2019), which focus on the asymptotic convergence of iterative Tikhonov regularization methods in the noisy or bandit feedback settings.

4.4. Proof Sketch of Theorem 4.5

As in Theorem 4.1, we first derive the convergence rate of σ^{k+1} for the noisy feedback setting:

Lemma 4.6. *Let $\theta = \frac{3\mu^2+8L^2}{2\mu}$ and $\kappa = \frac{\mu}{2}$. Suppose that both D_ψ and G are defined as the squared ℓ^2-distance. Under the noisy feedback setting, if we use the learning rate sequence of the form $\eta_t = 1/(\kappa(t - T_\sigma \cdot \lfloor t/T_\sigma \rfloor) + 2\theta)$, the $k+1$-th slingshot strategy profile σ^{k+1} of APMD for each $k \geq 0$ satisfies that:*

$$\mathbb{E}\left[\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|^2\right] = \mathcal{O}\left(\frac{\ln T_\sigma}{T_\sigma}\right).$$

The proof is given in Appendix E.6 and is based on the standard argument of stochastic optimization, e.g., Nedić & Lee (2014). However, the proof is made possible by taking into account the monotonicity of the game and the relative (strong and smooth) convexity of the divergence function.

Next, in a similar manner for Lemma 4.4, we show the upper bound on the expected distance between $\pi^{\mu,\sigma^{K-1}}$ and σ^{K-1} under the noisy feedback setting:

Lemma 4.7. *In the same setup of Theorem 4.5, the $K-1$-th slingshot strategy profile σ^{K-1} satisfies:*

$$\mathbb{E}\left[\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|\right] = \mathcal{O}\left(\frac{\ln K}{\sqrt{K}}\right).$$

We note that Lemma 4.3 holds for any combination of σ^k and σ^{k+1}, regardless of the presence of noise. By combining Lemmas 4.3, 4.6, and 4.7, we can derive the following last-iterate convergence rate of σ^K in terms of the gap function:

$$\mathbb{E}\left[\mathrm{GAP}(\sigma^K)\right] = \mathcal{O}\left(\frac{\ln K}{\sqrt{K}}\right).$$

By setting $K = \lfloor T/T_\sigma \rfloor = \Theta(T^{1/5})$, we conclude the statement of the theorem. □

5. Beyond Squared ℓ^2-Payoff Perturbation

Section 4 assumes that both G and D_ψ are the squared ℓ^2-distance. This section considers more general choices of G and D_ψ.

5.1. Instantiation of Payoff-Perturbed Algorithms

First, we would like to emphasize that choosing appropriate combinations of G and D_ψ enables APMD to reproduce some existing learning algorithms that incorporate payoff perturbation. For example, the following learning algorithms can be viewed as instantiations of APMD.

Example 5.1 (Boltzmann Q-Learning (Tuyls et al., 2006))**.** Assume that $\mathcal{X}_i = \Delta^{d_i}$, the regularize is entropy: $\psi(\pi_i) = \sum_{j=1}^{d_i} \pi_{ij} \ln \pi_{ij}$. Let us set G as the KL divergence and the slingshot strategy σ_i as a uniform distribution, i.e., $G(\pi_i, \sigma_i) = \mathrm{KL}(\pi_i, \sigma_i)$ and $\sigma_i = (1/d_i)_{j\in[d_i]}$. Then, the corresponding continuous-time APMD dynamics can be expressed as:

$$\frac{d}{dt}\pi_{ij}^t = \pi_{ij}^t\left(q_{ij}^{\pi^t} - \sum_{k=1}^{d_i}\pi_{ik}^t q_{ik}^{\pi^t}\right) - \mu\pi_{ij}^t\left(\ln\pi_{ij}^t - \sum_{k=1}^{d_i}\pi_{ik}^t\ln\pi_{ik}^t\right),$$

which is equivalent to Boltzman Q-learning (Tuyls et al., 2006; Bloembergen et al., 2015).

Example 5.2 (Reward transformed FTRL (Perolat et al., 2021))**.** Consider the continuous-time APMD dynamics where $N = 2$, $\mathcal{X}_i = \Delta^{d_i}$, ψ is Legendre (Rockafellar, 1997; Lattimore & Szepesvári, 2020) on Δ^{d_i}, and $G(\pi_i, \sigma_i) = \mathrm{KL}(\pi_i, \sigma_i)$. Then, APMD dynamics can be described as follows:

$$\pi_{ij}^t = \arg\max_{x\in\Delta^{d_i}}\left\{\int_0^t\left(\sum_{k=1}^{d_i}x_k q_{ik}^{\pi^s} - \mu\sum_{k=1}^{d_i}x_k\ln\frac{\pi_{ik}^s}{\sigma_{ik}} + \mu\sum_{k=1}^{d_{-i}}\pi_{-ik}^s\ln\frac{\pi_{-ik}^s}{\sigma_{-ik}}\right)ds - \psi(x)\right\}.$$

This algorithm is equivalent to FTRL with reward transformation (Perolat et al., 2021).

Example 5.3 (Mutant FTRL (Abe et al., 2022))**.** Let us define $\mathcal{X}_i = \Delta^{d_i}$, and assume that the regularizer ψ is Legendre on Δ^{d_i}. If we set G as the reverse KL divergence, i.e., $G(\pi_i, \sigma_i) = \mathrm{KL}(\sigma_i, \pi_i) = \sum_{j=1}^{d_i}\sigma_{ij}\ln\frac{\sigma_{ij}}{\pi_{ij}}$, we can rewrite (3) as:

$$\pi_i^{t+1} = \arg\max_{x\in\Delta^{d_i}}\left\{\sum_{s=0}^{t}\eta\sum_{j=1}^{d_i}x_j\left(q_{ij}^{\pi^s} + \frac{\mu}{\pi_{ij}^s}\left(\sigma_{ij} - \pi_{ij}^s\right)\right) - \psi(x)\right\},$$

where $q_{ij}^{\pi^s} = (\widehat{\nabla}_{\pi_i}v_i(\pi_i^s, \pi_{-i}^s))_j$. This algorithm is equivalent to Mutant FTRL (Abe et al., 2022).

5.2. Convergence Results with General G and D_ψ

Next, we establish the convergence results for APMD with general combinations of G and D_ψ. For theoretical analysis, we assume a specific condition on G:

Assumption 5.4. *$G(\cdot, \sigma_i)$ is differentiable over $\mathrm{int}(\mathrm{dom}\,\psi)$. Moreover, $G(\cdot, \sigma_i)$ is β-smooth and γ-strongly convex relative to ψ, i.e., for any $\pi_i, \pi_i' \in \mathrm{int}(\mathrm{dom}\,\psi)$, $\gamma D_\psi(\pi_i', \pi_i) \le G(\pi_i', \sigma_i) - G(\pi_i, \sigma_i) - \langle\nabla_{\pi_i}G(\pi_i, \sigma_i), \pi_i' - \pi_i\rangle \le \beta D_\psi(\pi_i', \pi_i)$ holds.*

Note that these assumptions are always satisfied with $\beta = \gamma = 1$ whenever G is identical to D_ψ; thus, these are not strong assumptions. We also assume that π^t is well-defined over iterations:

Assumption 5.5. *π^t updated by APMD satisfies $\pi^t \in \mathrm{int}(\mathrm{dom}\,\psi)$ for any $t \in \{0, 1, \cdots, T\}$.*

Using Assumptions 5.4 and 5.5, we derive the convergence rate to π^{μ,σ^k} in (4):

Lemma 5.6. *Suppose that Assumptions 5.4 with $\beta, \gamma \in (0, \infty)$ and 5.5 hold. If we use the constant learning rate $\eta_t = \eta \in (0, \frac{2\mu\gamma\rho^2}{\mu^2\gamma\rho^2(\gamma+2\beta)+8L^2})$, the $k+1$-th slingshot strategy profile σ^{k+1} of APMD under the full feedback setting satisfies that:*

$$D_\psi(\pi^{\mu,\sigma^k}, \sigma^{k+1}) = D_\psi(\pi^{\mu,\sigma^k}, \sigma^k)\exp\left(-\mathcal{O}(T_\sigma)\right).$$

Lemma 5.7. *Let $\theta = \frac{\mu^2\gamma\rho^2(\gamma+2\beta)+8L^2}{2\mu\gamma\rho^2}$ and $\kappa = \frac{\mu\gamma}{2}$. Suppose that Assumptions 5.4 with $\beta, \gamma \in (0, \infty)$ and 5.5 hold, and the learning rate sequence of the form $\eta_t = 1/(\kappa(t - T_\sigma \cdot \lfloor t/T_\sigma \rfloor) + 2\theta)$ is used. Then, the $k+1$-th slingshot strategy profile σ^{k+1} of APMD under the noisy feedback setting satisfies that:*

$$\mathbb{E}\left[D_\psi(\pi^{\mu,\sigma^k}, \sigma^{k+1})\right] = \mathcal{O}\left(\frac{\ln T_\sigma}{T_\sigma}\right).$$

These lemmas imply that $\sigma^{k+1} \to \pi^{\mu,\sigma^k}$ as $T_\sigma \to \infty$. Therefore, when T_σ is sufficiently large, $k+1$-th slingshot strategy profile becomes almost equivalent to the stationary point π^{μ,σ^k}. From this, we anticipate that $\mathbb{E}[\mathrm{GAP}(\sigma^k)] \to 0$ as $k \to \infty$ even in the noisy feedback setting. The subsequent theorems provide asymptotic last-iterate convergence results for this ideal scenario. In particular, we show that the slingshot strategy profile σ^{k+1} converges to equilibrium when using the following divergence functions as G: 1) Bregman divergence; 2) α-divergence; 3) Rényi-divergence; 4) reverse KL divergence.

Theorem 5.8. *Assume that G is a Bregman divergence $D_{\psi'}$ for some strongly convex function ψ', and $\sigma^{k+1} = \pi^{\mu,\sigma^k}$ for $k \geq 0$. Then, there exists $\pi^* \in \Pi^*$ such that $\sigma^k \to \pi^*$ as $k \to \infty$.*

Theorem 5.9. *Let us define $\mathcal{X}_i = \Delta^{d_i}$. Assume that $\sigma^{k+1} = \pi^{\mu,\sigma^k}$ for $k \geq 0$, and G is one of the following divergence: 1) α-divergence with $\alpha \in (0,1)$; 2) Rényi-divergence with $\alpha \in (0,1)$; 3) reverse KL divergence. If the initial slingshot strategy profile σ^0 is in the interior of $\mathcal{X}$, the sequence $\{\sigma^k\}_{k\geq 1}$ converges to the set of Nash equilibria Π^* of the underlying game.*

We remark that these results cover the algorithms in Example 5.1, 5.2, and 5.3. Furthermore, we can incorporate our payoff perturbation techniques into FTRL, detailed in Appendix C.

6. Experiments

This section empirically compares the representative instance of MD, namely Multiplicative Weight Update (MWU) and its Optimistic version (OMWU), with our framework. Specifically, we consider the following three instances of APMD: (i) both are the squared ℓ^2-distance; (ii) both G and D_ψ are the KL divergence, which is also an instance of Reward transformed FTRL in Example 5.2. Note that if the slingshot strategy is fixed to a uniformly random strategy, this algorithm corresponds to Boltzmann Q-Learning in Example 5.1; (iii) the divergence function G is the reverse KL divergence, and the Bregman divergence D_ψ is the KL divergence, which matches Mutant FTRL in Example 5.3.

We focus on two zero-sum polymatrix games: Three-Player Biased Rock-Paper-Scissors (3BRPS) and three-player random payoff games with 10 and 50 actions. For the 3BRPS game, each player participates in two instances of the game in Table 2 in Appendix I simultaneously with two other players. For the random payoff games, each player i participates in two instances of the game with two other players j simultaneously. The payoff matrix for each instance is denoted as $M^{(i,j)}$. Each entry of $M^{(i,j)}$ is drawn independently from a uniform distribution on the interval $[-1, 1]$.

Figures 2 and 3 illustrate the logarithm of the gap function averaged over 100 instances with different random seeds. We assume that the initial slingshot strategy profile π^0 is chosen uniformly at random in the interior of the strategy space $\mathcal{X} = \prod_{i=1}^3 \Delta^{d_i}$ in each instance for 3BRPS, while π^0 is chosen as $(1/d_i)_{j\in[d_i]}$ for $i \in [3]$ in every instances for the random payoff games.

First, Figure 2 depicts the case of full feedback. Unless otherwise specified, we use a constant learning rate $\eta = 0.1$ and a perturbation strength $\mu = 0.1$ for APMD. Further details and additional experiments can be found in Appendix I. Figure 2 shows that APMD outperforms MWU and OMWU in all three games. Notably, APMD exhibits the fastest convergence in terms of the gap function when using the squared ℓ^2-distance as both G and D_ψ. Next, Figure 3 depicts the case of noisy feedback. We assume that the noise vector ξ_i^t is generated from the multivariate Gaussian distribution $\mathcal{N}(0,\ 0.1^2\mathbf{I})$ in an i.i.d. manner. To account for the noise, we use a lower learning rate $\eta = 0.01$ than the full feedback case. In OMWU, we use the noisy gradient vector $\widehat{\nabla}_{\pi_i} v_i(\pi_i^{t-1}, \pi_{-i}^{t-1})$ at the previous step $t-1$ as the prediction vector for the current iteration t. We observe the same trends as with full feedback. While MWU and OMWU exhibit worse performance, APMD maintains its fast convergence, as predicted by the theoretical analysis.

7. Related Literature

Recent progress in achieving no-regret learning with full feedback has been driven by optimistic learning (Rakhlin & Sridharan, 2013a;b). Optimistic versions of well-known algorithms like Follow the Regularized Leader (Shalev-Shwartz & Singer, 2006) and Mirror Descent (Zhou et al., 2017; Hsieh et al., 2021) have been proposed to admit last-iterate convergence in a wide range of game settings. These optimistic algorithms have been successfully applied to various classes of games, including bilinear games (Daskalakis et al., 2018; Daskalakis & Panageas, 2019; Liang & Stokes, 2019; de Montbrun & Renault, 2022), cocoercive games (Lin et al., 2020), and saddle point problems (Daskalakis & Panageas, 2018; Mertikopoulos et al., 2019; Golowich et al., 2020b; Wei et al., 2021; Lei et al., 2021; Yoon & Ryu, 2021; Lee & Kim, 2021; Cevher et al., 2023). The advance-

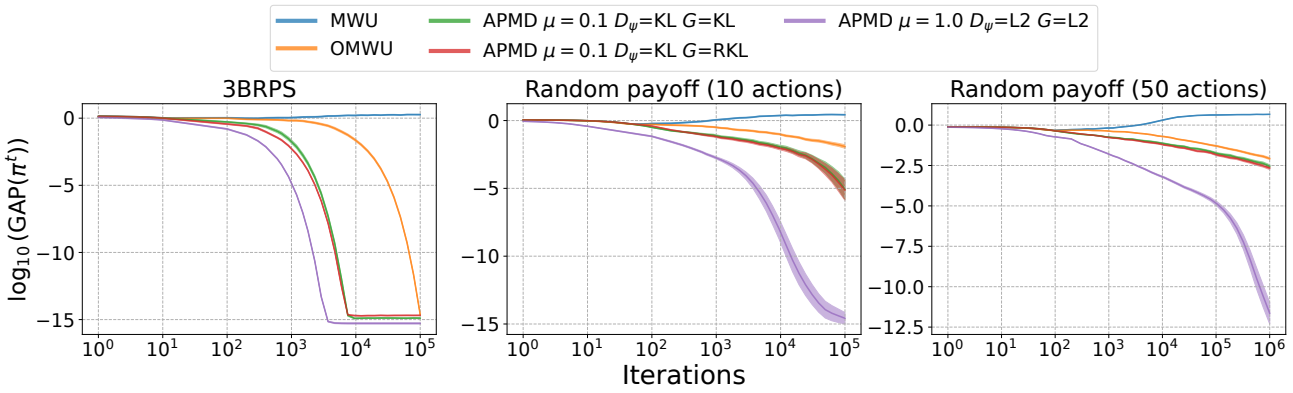


Figure 2: The gap function for π^t for APMD, MWU, and OMWU with full feedback. The shaded area represents the standard errors. Note that the KL divergence, reverse KL divergence, and squared ℓ^2-distance are abbreviated to KL, RKL, and L2, respectively.

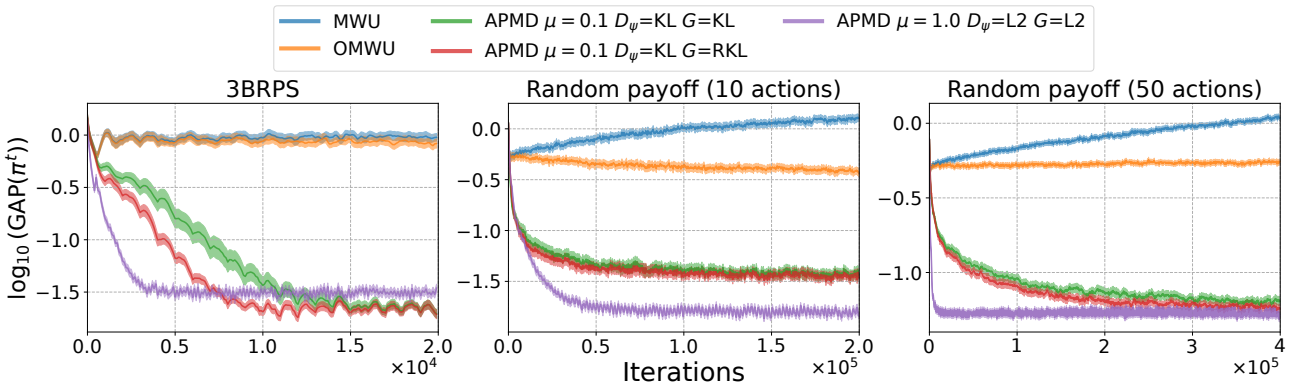


Figure 3: The gap function for π^t for APMD, MWU, and OMWU with noisy feedback.

ments have provided solutions to monotone games and have established convergence rates (Golowich et al., 2020a; Cai et al., 2022a;b; Gorbunov et al., 2022; Cai & Zheng, 2023).

The exploration of literature with noisy feedback poses significant challenges, in contrast to full feedback. In situations where feedback is imprecise or limited, algorithms must estimate action values at each iteration. There have been two trends in achieving last-iterate convergence: restricting the class of games and perturbing the payoff functions. On one hand, particularly noticeable works lie in potential games (Cohen et al., 2017), normal-form games with strict Nash equilibria (Giannou et al., 2021b;a), and two-player zero-sum games (Abe et al., 2023). Also, noisy feedback is handled with games whose payoff functions are assumed to be strictly (or strongly) monotone (Bravo et al., 2018; Kannan & Shanbhag, 2019; Hsieh et al., 2019; Anagnostides & Panageas, 2022), while to be strictly variational stable (Mertikopoulos & Zhou, 2019; Mertikopoulos et al., 2019; 2022; Azizian et al., 2021). Note that variationally stable games, often referred to in control theory, are a slightly broader class of monotone games. These studies requires the payoff functions to be strictly or strongly convex, while we do not require such an assumption.

On the other hand, payoff-perturbed algorithms have recently regained attention for their ability to demonstrate convergence in unrestricted games when noise is present. As described in Section 1, payoff perturbation is a textbook technique (Facchinei & Pang, 2003) that has been extensively studied (Koshal et al., 2010; 2013; Yousefian et al., 2017; Tatarenko & Kamgarpour, 2019; Hsieh et al., 2020; 2022; Abe et al., 2023; Cen et al., 2021; 2023; Cai et al., 2023; Pattathil et al., 2023). It is known that carefully adjusting the magnitude of perturbation ensures convergence to a Nash equilibrium. This magnitude is computed as the product of a strongly convex penalty and a perturbation strength parameter. Liu et al. (2023) shrink the perturbation strength based on a predefined hyper-parameter and the gap function of the current strategy. Likewise, Koshal et al. (2010) and Tatarenko & Kamgarpour (2019) have identified somewhat complex conditions that the sequence

of the perturbation strength parameters and learning rates should satisfy. Roughly speaking, as we have implied in Lemma 4.2, a smaller strength would require a lower learning rate. This potentially decelerates the convergence rate and complicates the task of finding an appropriate learning rate. For practicality, we have opted to keep the perturbation strength constant, independent of the iteration in APMD. Moreover, it must be emphasized that the existing literature has primarily provided asymptotic convergence results, while we have successfully provided the non-asymptotic convergence rate.

Finally, the idea of the slingshot strategy update was initiated by Perolat et al. (2021) and later extended by Abe et al. (2023). Our contribution partly owes to the significance of quantifying the convergence rate for the first time. We must also mention that Sokota et al. (2023) have proposed a very similar, but essentially different update rule to ours. It just adds an additional regularization term based on an anchoring strategy, which they call a *magnetic* one, and this means that it directly perturbs the (expected) payoff functions. In contrast, our APMD indirectly perturbs the payoff functions, i.e., perturbs the gradient vector. Furthermore, we have established non-asymptotic convergence results toward a Nash equilibrium, while Sokota et al. (2023) have only shown convergence toward a quantal response equilibrium (McKelvey & Palfrey, 1995; 1998), which is just equivalent to an approximate equilibrium. Similar results to them have been obtained with the Boltzmann Q-learning dynamics (Tuyls et al., 2006) and penalty-regularized dynamics (Coucheney et al., 2015) in continuous-time settings (Leslie & Collins, 2005; Hussain et al., 2023).

8. Conclusion

This paper proposes a novel variant of MD that achieves last-iterate convergence even when the noise is present, by adaptively adjusting the magnitude of the perturbation. This research could lead to several intriguing future studies, such as finding the best perturbation strength for the optimal convergence rate and achieving convergence with more limited feedback, for example, using bandit feedback (Bravo et al., 2018; Drusvyatskiy et al., 2022).

Acknowledgments

Kaito Ariu is supported by JSPS KAKENHI Grant Number 23K19986. Atsushi Iwasaki is supported by JSPS KAKENHI Grant Numbers 21H04890 and 23K17547.

Impact Statement

This paper focuses on discussing the problem of computing equilibria in games. There are potential social implications associated with our work, but none that we believe need to be particularly emphasized here.

References

Abe, K., Sakamoto, M., and Iwasaki, A. Mutation-driven follow the regularized leader for last-iterate convergence in zero-sum games. In *UAI*, pp. 1–10, 2022.

Abe, K., Ariu, K., Sakamoto, M., Toyoshima, K., and Iwasaki, A. Last-iterate convergence with full and noisy feedback in two-player zero-sum games. In *AISTATS*, pp. 7999–8028, 2023.

Anagnostides, I. and Panageas, I. Frequency-domain representation of first-order methods: A simple and robust framework of analysis. In *SOSA*, pp. 131–160, 2022.

Azizian, W., Iutzeler, F., Malick, J., and Mertikopoulos, P. The last-iterate convergence rate of optimistic mirror descent in stochastic variational inequalities. In *COLT*, pp. 326–358, 2021.

Bailey, J. P. and Piliouras, G. Multiplicative weights update in zero-sum games. In *Economics and Computation*, pp. 321–338, 2018.

Beck, A. and Teboulle, M. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

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

Bravo, M., Leslie, D., and Mertikopoulos, P. Bandit learning in concave N-person games. In *NeurIPS*, pp. 5666–5676, 2018.

Cai, Y. and Daskalakis, C. On minmax theorems for multiplayer games. In *SODA*, pp. 217–234, 2011.

Cai, Y. and Zheng, W. Doubly optimal no-regret learning in monotone games. *arXiv preprint arXiv:2301.13120*, 2023.

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

Cai, Y., Oikonomou, A., and Zheng, W. Finite-time last-iterate convergence for learning in multi-player games. In *NeurIPS*, pp. 33904–33919, 2022a.

Cai, Y., Oikonomou, A., and Zheng, W. Tight last-iterate convergence of the extragradient method for constrained monotone variational inequalities. *arXiv preprint arXiv:2204.09228*, 2022b.

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

Cen, S., Wei, Y., and Chi, Y. Fast policy extragradient methods for competitive games with entropy regularization. In *NeurIPS*, pp. 27952–27964, 2021.

Cen, S., Chi, Y., Du, S. S., and Xiao, L. Faster last-iterate convergence of policy optimization in zero-sum Markov games. In *ICLR*, 2023.

Cevher, V., Piliouras, G., Sim, R., and Skoulakis, S. Min-max optimization made simple: Approximating the proximal point method via contraction maps. In *Symposium on Simplicity in Algorithms (SOSA)*, pp. 192–206, 2023.

Cohen, J., Héliou, A., and Mertikopoulos, P. Learning with bandit feedback in potential games. In *NeurIPS*, pp. 6372–6381, 2017.

Coucheney, P., Gaujal, B., and Mertikopoulos, P. Penalty-regulated dynamics and robust learning procedures in games. *Mathematics of Operations Research*, 40(3):611–633, 2015.

Daskalakis, C. and Panageas, I. The limit points of (optimistic) gradient descent in min-max optimization. In *NeurIPS*, pp. 9256–9266, 2018.

Daskalakis, C. and Panageas, I. Last-iterate convergence: Zero-sum games and constrained min-max optimization. In *ITCS*, pp. 27:1–27:18, 2019.

Daskalakis, C., Ilyas, A., Syrgkanis, V., and Zeng, H. Training GANs with optimism. In *ICLR*, 2018.

de Montbrun, É. and Renault, J. Convergence of optimistic gradient descent ascent in bilinear games. *arXiv preprint arXiv:2208.03085*, 2022.

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

Drusvyatskiy, D., Fazel, M., and Ratliff, L. J. Improved rates for derivative free gradient play in strongly monotone games. In *CDC*, pp. 3403–3408. IEEE, 2022.

Facchinei, F. and Pang, J.-S. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

Giannou, A., Vlatakis-Gkaragkounis, E. V., and Mertikopoulos, P. Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information. In *COLT*, pp. 2147–2148, 2021a.

Giannou, A., Vlatakis-Gkaragkounis, E.-V., and Mertikopoulos, P. On the rate of convergence of regularized learning in games: From bandits and uncertainty to optimism and beyond. In *NeurIPS*, pp. 22655–22666, 2021b.

Golowich, N., Pattathil, S., and Daskalakis, C. Tight last-iterate convergence rates for no-regret learning in multi-player games. In *NeurIPS*, pp. 20766–20778, 2020a.

Golowich, N., Pattathil, S., Daskalakis, C., and Ozdaglar, A. Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems. In *COLT*, pp. 1758–1784, 2020b.

Gorbunov, E., Taylor, A., and Gidel, G. Last-iterate convergence of optimistic gradient method for monotone variational inequalities. In *NeurIPS*, pp. 21858–21870, 2022.

Halpern, B. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73(6):957 – 961, 1967.

Hart, S. and Mas-Colell, A. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5): 1127–1150, 2000.

Hsieh, Y.-G., Iutzeler, F., Malick, J., and Mertikopoulos, P. On the convergence of single-call stochastic extra-gradient methods. In *NeurIPS*, pp. 6938–6948, 2019.

Hsieh, Y.-G., Iutzeler, F., Malick, J., and Mertikopoulos, P. Explore aggressively, update conservatively: Stochastic extragradient methods with variable stepsize scaling. *Advances in Neural Information Processing Systems*, 33: 16223–16234, 2020.

Hsieh, Y.-G., Antonakopoulos, K., and Mertikopoulos, P. Adaptive learning in continuous games: Optimal regret bounds and convergence to Nash equilibrium. In *COLT*, pp. 2388–2422, 2021.

Hsieh, Y.-G., Antonakopoulos, K., Cevher, V., and Mertikopoulos, P. No-regret learning in games with noisy feedback: Faster rates and adaptivity via learning rate separation. In *NeurIPS*, pp. 6544–6556, 2022.

Hussain, A. A., Belardinelli, F., and Piliouras, G. Asymptotic convergence and performance of multi-agent Q-learning dynamics. *arXiv preprint arXiv:2301.09619*, 2023.

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

Koshal, J., Nedić, A., and Shanbhag, U. V. Single timescale regularized stochastic approximation schemes for monotone nash games under uncertainty. In *CDC*, pp. 231–236. IEEE, 2010.

Koshal, J., Nedic, A., and Shanbhag, U. V. Regularized iterative stochastic approximation methods for stochastic variational inequality problems. *IEEE Transactions on Automatic Control*, 58(3):594–609, 2013.

Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020.

Lee, S. and Kim, D. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. In *NeurIPS*, pp. 22588–22600, 2021.

Lei, Q., Nagarajan, S. G., Panageas, I., et al. Last iterate convergence in no-regret learning: constrained min-max optimization for convex-concave landscapes. In *AISTATS*, pp. 1441–1449, 2021.

Leslie, D. S. and Collins, E. J. Individual q-learning in normal form games. *SIAM Journal on Control and Optimization*, 44(2):495–514, 2005.

Liang, T. and Stokes, J. Interaction matters: A note on non-asymptotic local convergence of generative adversarial networks. In *AISTATS*, pp. 907–915, 2019.

Lin, T., Zhou, Z., Mertikopoulos, P., and Jordan, M. Finite-time last-iterate convergence for multi-agent learning in games. In *ICML*, pp. 6161–6171, 2020.

Liu, M., Ozdaglar, A., Yu, T., and Zhang, K. The power of regularization in solving extensive-form games. In *ICLR*, 2023.

McKelvey, R. D. and Palfrey, T. R. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.

McKelvey, R. D. and Palfrey, T. R. Quantal response equilibria for extensive form games. *Experimental economics*, 1:9–41, 1998.

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

Mertikopoulos, P., Papadimitriou, C., and Piliouras, G. Cycles in adversarial regularized learning. In *SODA*, pp. 2703–2717, 2018.

Mertikopoulos, P., Lecouat, B., Zenati, H., Foo, C.-S., Chandrasekhar, V., and Piliouras, G. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *ICLR*, 2019.

Mertikopoulos, P., Hsieh, Y.-P., and Cevher, V. Learning in games from a stochastic approximation viewpoint. *arXiv preprint arXiv:2206.03922*, 2022.

Nash, J. Non-cooperative games. *Annals of mathematics*, pp. 286–295, 1951.

Nedić, A. and Lee, S. On stochastic subgradient mirror-descent algorithm with weighted averaging. *SIAM Journal on Optimization*, 24(1):84–107, 2014.

Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

Nemirovskij, A. S. and Yudin, D. B. *Problem complexity and method efficiency in optimization*. Wiley, 1983.

Pattathil, S., Zhang, K., and Ozdaglar, A. Symmetric (optimistic) natural policy gradient for multi-agent learning with parameter convergence. In *AISTATS*, pp. 5641–5685, 2023.

Perolat, J., Munos, R., Lespiau, J.-B., Omidshafiei, S., Rowland, M., Ortega, P., Burch, N., Anthony, T., Balduzzi, D., De Vylder, B., et al. From Poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization. In *ICML*, pp. 8525–8535, 2021.

Rakhlin, A. and Sridharan, K. Online learning with predictable sequences. In *COLT*, pp. 993–1019, 2013a.

Rakhlin, S. and Sridharan, K. Optimization, learning, and games with predictable sequences. In *NeurIPS*, pp. 3066–3074, 2013b.

Rockafellar, R. T. *Convex analysis*, volume 11. Princeton university press, 1997.

Shalev-Shwartz, S. and Singer, Y. Convex repeated games and fenchel duality. *Advances in neural information processing systems*, 19, 2006.

Sokota, S., D'Orazio, R., Kolter, J. Z., Loizou, N., Lanctot, M., Mitliagkas, I., Brown, N., and Kroer, C. A unified approach to reinforcement learning, quantal response equilibria, and two-player zero-sum games. In *ICLR*, 2023.

Tammelin, O. Solving large imperfect information games using CFR+. *arXiv preprint arXiv:1407.5042*, 2014.

Tatarenko, T. and Kamgarpour, M. Learning Nash equilibria in monotone games. In *CDC*, pp. 3104–3109. IEEE, 2019.

Tuyls, K., Hoen, P. J., and Vanschoenwinkel, B. An evolutionary dynamical analysis of multi-agent learning in iterated games. *Autonomous Agents and Multi-Agent Systems*, 12(1):115–153, 2006.

Wei, C.-Y., Lee, C.-W., Zhang, M., and Luo, H. Linear last-iterate convergence in constrained saddle-point optimization. In *ICLR*, 2021.

Yoon, T. and Ryu, E. K. Accelerated algorithms for smooth convex-concave minimax problems with $\mathcal{O}(1/k^2)$ rate on squared gradient norm. In *ICML*, pp. 12098–12109, 2021.

Yousefian, F., Nedić, A., and Shanbhag, U. V. On smoothing, regularization, and averaging in stochastic approximation methods for stochastic variational inequality problems. *Mathematical Programming*, 165:391–431, 2017.

Zhou, Z., Mertikopoulos, P., Moustakas, A. L., Bambos, N., and Glynn, P. Mirror descent learning in continuous games. In *CDC*, pp. 5776–5783. IEEE, 2017.

Contents

1 Introduction **1**

2 Preliminaries **2**

3 Adaptively Perturbed Mirror Descent **3**

3.1 Slingshot Payoff Perturbation . . . 3

3.2 Slingshot Strategy Update . . . 4

4 Last-Iterate Convergence Rates **4**

4.1 Full Feedback Setting . . . 4

4.2 Proof Sketch of Theorem 4.1 . . . 4

4.3 Noisy Feedback Setting . . . 5

4.4 Proof Sketch of Theorem 4.5 . . . 5

5 Beyond Squared ℓ^2-Payoff Perturbation **6**

5.1 Instantiation of Payoff-Perturbed Algorithms . . . 6

5.2 Convergence Results with General G and D_ψ . . . 6

6 Experiments **7**

7 Related Literature **7**

8 Conclusion **9**

A Notations **15**

B Formal Theorems and Lemmas **16**

B.1 Full Feedback Setting . . . 16

B.2 Noisy Feedback Setting . . . 16

B.3 Convergence Results with General G and D_ψ . . . 17

C Extension to Follow the Regularized Leader **17**

D Additional Theorems **18**

E Proofs for Section 4 **18**

E.1 Proof of Theorem B.1 (Formal Version of Theoremf 4.1) . . . 18

E.2 Proof of Lemma B.2 (Formal Version of Lemma 4.2) . . . 19

E.3 Proof of Lemma B.3 (Formal Version of Lemma 4.3) . . . 19

E.4 Proof of Lemma B.4 (Formal Version of Lemma 4.4) . . . 21

E.5 Proof of Theorem B.6 (Formal Version of Theorem 4.5) . . . 22
E.6 Proof of Lemma B.7 (Formal Version of Lemma 4.6) . . . 24
E.7 Proof of Lemma B.8 (Formal Version of Lemma 4.7) . . . 24

F Proofs for Section 5 **26**
F.1 Proof of Lemma B.9 (Formal Version of Lemma 5.6) . . . 26
F.2 Proof of Lemma B.10 (Formal Version of Lemma 5.7) . . . 26
F.3 Proof of Theorem 5.8 . . . 27
F.4 Proof of Theorem 5.9 . . . 28
F.5 Proof of Lemma C.1 . . . 29
F.6 Proof of Lemma C.2 . . . 29

G Proofs for Section D **30**
G.1 Proof of Theorem D.1 . . . 30

H Proofs for Additional Lemmas **31**
H.1 Proof of Lemma E.2 . . . 31
H.2 Proof of Lemma E.3 . . . 33
H.3 Proof of Lemma E.4 . . . 34
H.4 Proof of Lemma F.1 . . . 36
H.5 Proof of Lemma F.2 . . . 37
H.6 Proof of Lemma F.3 . . . 38
H.7 Proof of Lemma F.4 . . . 39
H.8 Proof of Lemma F.5 . . . 40
H.9 Proof of Lemma F.6 . . . 41
H.10 Proof of Lemma F.7 . . . 41
H.11 Proof of Lemma F.8 . . . 44
H.12 Proof of Lemma H.1 . . . 44
H.13 Proof of Lemma H.2 . . . 44
H.14 Proof of Lemma H.3 . . . 45

I Additional Experimental Results and Details **48**
I.1 Payoff Matrix in Three-Player Biased RPS Game . . . 48
I.2 Experimental Setting for Section 6 . . . 48
I.3 Additional Experiments . . . 48
I.4 Comparison with the Averaged Strategies of No-Regret Learning Algorithms . . . 48
I.5 Sensitivity Analysis of Update Interval for the Slingshot Strategy . . . 49

A. Notations

In this section, we summarize the notations we use in Table 1.

Table 1: Notations

Symbol	Description
N	Number of players
$\mathcal{X}_i$	Strategy space for player i
$\mathcal{X}$	Joint strategy space: $\mathcal{X} = \prod_{i=1}^N \mathcal{X}_i$
v_i	Payoff function for player i
π_i	Strategy for player i
π	Strategy profile: $\pi = (\pi_i)_{i\in[N]}$
ξ_i^t	Noise vector for player i at iteration t
π^*	Nash equilibrium
Π^*	Set of Nash equilibria
$\mathrm{GAP}(\pi)$	Gap function of π: $\mathrm{GAP}(\pi) = \max_{\tilde{\pi}\in\mathcal{X}} \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi), \tilde{\pi}_i - \pi_i \rangle$
Δ^d	d-dimensional probability simplex: $\Delta^d = \{p \in [0,1]^d \mid \sum_{j=1}^d p_j = 1\}$
$\mathrm{diam}(\mathcal{X})$	Diameter of $\mathcal{X}$: $\mathrm{diam}(\mathcal{X}) = \sup_{\pi,\pi'\in\mathcal{X}} \lVert \pi - \pi' \rVert$
$\mathrm{KL}(\cdot,\cdot)$	Kullback-Leibler divergence
$D_\psi(\cdot,\cdot)$	Bregman divergence associated with ψ
$\nabla_{\pi_i} v_i$	Gradient of v_i with respect to π_i
η_t	Learning rate at iteration t
μ	Perturbation strength
σ	Slingshot strategy profile
$G(\cdot,\cdot)$	Divergence function for payoff perturbation
$\nabla_{\pi_i} G$	Gradient of G with respect to first argument
T_σ	Update interval for the slingshot strategy
K	Total number of the slingshot strategy updates
$\pi^{\mu,\sigma}$	Stationary point satisfies (4) for given μ and σ
π^t	Strategy profile at iteration t
σ^k	Slingshot strategy profile after k updates
L	Smoothness parameter of $(v_i)_{i\in[N]}$
ρ	Strongly convex parameter of ψ
β	Smoothness parameter of $G(\cdot,\sigma_i)$ relative to ψ
γ	Strongly convex parameter of $G(\cdot,\sigma_i)$ relative to ψ

B. Formal Theorems and Lemmas

B.1. Full Feedback Setting

Theorem B.1 (Formal version of Theorem 4.1)**.** *Assume that* $\sqrt{\sum_{i=1}^N \|\nabla_{\pi_i} v_i(\pi)\|^2} \le \zeta$ *for any* $\pi \in \mathcal{X}$*. If we use the constant learning rate* $\eta_t = \eta \in (0, \frac{2\mu\rho^2}{3\mu^2\rho^2+8L^2})$*, and set* D_ψ *and* G *as squared* ℓ^2*-distance* $D_\psi(\pi_i, \pi_i') = G(\pi_i, \pi_i') = \frac{1}{2}\|\pi_i - \pi_i'\|^2$*, and set* $T_\sigma = c \cdot \max(\frac{6}{\ln 2 - \ln(2-\eta\mu)} \ln T + \frac{2\ln 64}{\ln 2 - \ln(2-\eta\mu)}, 1)$ *for some constant* $c \ge 1$*, then the strategy profile* π^T *updated by APMD satisfies:*

$$\begin{aligned}&\mathrm{GAP}(\pi^T)\\ &\le \frac{2\sqrt{2}c\left((\mu+L)\cdot \mathrm{diam}(\mathcal{X}) + \zeta\right)\cdot\left(\frac{6}{\ln 2-\ln(2-\eta\mu)}\ln T + \frac{2\ln 64}{\ln 2-\ln(2-\eta\mu)} + 1\right)}{\sqrt{T}}\sqrt{\mathrm{diam}(\mathcal{X})\left(8\cdot\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)}.\end{aligned}$$

Lemma B.2 (Formal version of Lemma 4.2)**.** *Assume that* D_ψ *and* G *are set as the squared* ℓ^2*-distance. If we use the constant learning rate* $\eta_t = \eta \in (0, \frac{2\mu\rho^2}{3\mu^2\rho^2+8L^2})$*, the* $k+1$*-th slingshot strategy profile* σ^{k+1} *of APMD under the full feedback setting satisfies that:*

$$\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|^2 \le \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \left(1 - \frac{\eta\mu}{2}\right)^{T_\sigma}.$$

Lemma B.3 (Formal version of Lemma 4.3)**.** *Assume that* $\sqrt{\sum_{i=1}^N \|\nabla_{\pi_i} v_i(\pi)\|^2} \le \zeta$ *for any* $\pi \in \mathcal{X}$*. If* G *is set as the squared* ℓ^2*-distance, the gap function for* σ^{k+1} *of APMD satisfies for* $k \ge 0$*:*

$$\mathrm{GAP}(\sigma^{k+1}) \le \mu\cdot\mathrm{diam}(\mathcal{X})\cdot\|\pi^{\mu,\sigma^k} - \sigma^k\| + (L\cdot\mathrm{diam}(\mathcal{X}) + \zeta)\cdot\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|.$$

Lemma B.4 (Formal version of Lemma 4.4)**.** *Assume that* $T_\sigma \ge \max(\frac{6}{\ln 2-\ln(2-\eta\mu)}\ln T + \frac{2\ln 64}{\ln 2-\ln(2-\eta\mu)}, 1)$*. In the same setup of Theorem 4.1, the* $K-1$*-th slingshot strategy profile* σ^{K-1} *of APMD satisfies:*

$$\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\| \le \frac{2\sqrt{2}}{\sqrt{K}}\sqrt{\mathrm{diam}(\mathcal{X})\left(8\cdot\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)}.$$

B.2. Noisy Feedback Setting

For the noisy feedback setting, we assume that $\xi_i^t \in \mathbb{R}^{d_i}$ is a zero-mean independent random vector with bounded variance.

Assumption B.5. *For all* $t \ge 1$ *and* $i \in [N]$*, the noise vector* ξ_i^t *satisfies the following properties: (a) Zero-mean:* $\mathbb{E}[\xi_i^t|\mathcal{F}_t] = (0, \cdots, 0)^\top$*; (b) Bounded variance:* $\mathbb{E}[\|\xi_i^t\|^2|\mathcal{F}_t] \le C^2$ *with some constant* $C > 0$*.*

This is a standard assumption in learning in games with noisy feedback (Mertikopoulos & Zhou, 2019; Hsieh et al., 2019) and stochastic optimization (Nemirovski et al., 2009; Nedić & Lee, 2014). Under Assumption B.5, we can obtain the following convergence results for ADMP under the noisy feedback setting.

Theorem B.6 (Formal version of Theorem 4.5)**.** *Let* $\theta = \frac{3\mu^2\rho^2+8L^2}{2\mu\rho^2}$ *and* $\kappa = \frac{\mu}{2}$*. Suppose that Assumption B.5 holds and* $\sqrt{\sum_{i=1}^N \|\nabla_{\pi_i} v_i(\pi)\|^2} \le \zeta$ *for any* $\pi \in \mathcal{X}$*. We also assume that* D_ψ *and* G *are set as squared* ℓ^2*-distance* $D_\psi(\pi_i, \pi_i') = G(\pi_i, \pi_i') = \frac{1}{2}\|\pi_i - \pi_i'\|^2$*, and* $T_\sigma = c\cdot\max(T^{4/5}+2, 3)$ *for some constant* $c \ge 1$*. If we choose the learning rate sequence of the form* $\eta_t = 1/(\kappa(t - T_\sigma\cdot\lfloor t/T_\sigma\rfloor) + 2\theta)$*, then the strategy profile* π^T *updated by APMD satisfies:*

$$\begin{aligned}\mathbb{E}\left[\mathrm{GAP}(\pi^T)\right] &\le \frac{\sqrt{6c}\mu\cdot\mathrm{diam}(\mathcal{X})^2}{T^{1/10}}\\ &+ \frac{L\cdot\mathrm{diam}(\mathcal{X}) + \zeta + \mu\cdot\mathrm{diam}(\mathcal{X})\sqrt{18c\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu} + 1\right)}}{T^{1/10}}\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T + 1\right) + \frac{1}{2\theta}\right)}{\rho\kappa}}.\end{aligned}$$

Lemma B.7 (Formal version of Lemma 4.6)**.** *Let $\theta = \frac{3\mu^2\rho^2+8L^2}{2\mu\rho^2}$ and $\kappa = \frac{\mu}{2}$. Suppose that Assumption B.5 holds, and both D_ψ and G are defined as the squared ℓ^2-distance. Under the noisy feedback setting, if we use the learning rate sequence of the form $\eta_t = 1/(\kappa(t - T_\sigma \cdot \lfloor t/T_\sigma \rfloor) + 2\theta)$, the $k+1$-th slingshot strategy profile σ^{k+1} of APMD for each $k \geq 0$ satisfies that:*

$$\mathbb{E}\left[\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|^2\right] \leq \frac{2\theta - \kappa}{\kappa(T_\sigma - 1) + 2\theta}\mathbb{E}\left[\|\pi^{\mu,\sigma^k} - \sigma^{k}\|^2\right] + \frac{NC^2}{\rho(\kappa(T_\sigma - 1) + 2\theta)}\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T_\sigma - 1) + 1\right) + \frac{1}{2\theta}\right).$$

Lemma B.8 (Formal version of Lemma 4.7)**.** *Assume that $T_\sigma \geq \max(T^{4/5} + 2, 3)$. In the same setup of Theorem 4.5, the $K-1$-th slingshot strategy profile σ^{K-1} satisfies:*

$$\mathbb{E}\left[\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|\right] \leq \sqrt{\frac{\|\pi^* - \sigma^0\|^2 + \frac{3}{\rho\kappa}\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu} + 1\right)\left(\rho(2\theta - \kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T + 1\right) + \frac{1}{2\theta}\right)\right)}{K}}.$$

B.3. Convergence Results with General G and D_ψ

Lemma B.9 (Formal version of Lemma 5.6)**.** *Suppose that Assumptions 5.4 with $\beta, \gamma \in (0, \infty)$ and 5.5 hold. If we use the constant learning rate $\eta_t = \eta \in (0, \frac{2\mu\gamma\rho^2}{\mu^2\gamma\rho^2(\gamma+2\beta)+8L^2})$, the $k+1$-th slingshot strategy profile σ^{k+1} of APMD under the full feedback setting satisfies that:*

$$D_\psi(\pi^{\mu,\sigma^k}, \sigma^{k+1}) \leq D_\psi(\pi^{\mu,\sigma^k}, \sigma^{k})\left(1 - \frac{\eta\mu\gamma}{2}\right)^{T_\sigma}.$$

Lemma B.10 (Formal version of Lemma 5.7)**.** *Let $\theta = \frac{\mu^2\gamma\rho^2(\gamma+2\beta)+8L^2}{2\mu\gamma\rho^2}$ and $\kappa = \frac{\mu\gamma}{2}$. Suppose that Assumptions 5.4, 5.5, and B.5 hold, and the learning rate sequence of the form $\eta_t = 1/(\kappa(t - T_\sigma \cdot \lfloor t/T_\sigma \rfloor) + 2\theta)$ is used. Then, the $k+1$-th slingshot strategy profile σ^{k+1} of APMD under the noisy feedback setting satisfies that:*

$$\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \sigma^{k+1})] \leq \frac{2\theta - \kappa}{\kappa(T_\sigma - 1) + 2\theta}\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \sigma^{k})] + \frac{NC^2}{\rho(\kappa(T_\sigma - 1) + 2\theta)}\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T_\sigma - 1) + 1\right) + \frac{1}{2\theta}\right).$$

C. Extension to Follow the Regularized Leader

In Sections 4 and 5, we introduced and analyzed APMD, which extends the standard MD approach. Similarly, it is possible to extend the FTRL approach as well. In this section, we present Adaptively Perturbed Follow the Regularized Leader (APFTRL), which incorporates the perturbation term $\mu G(\cdot, \sigma_i)$ into the conventional FTRL algorithm:

$$\pi_i^{t+1} = \arg\max_{x \in \mathcal{X}_i}\left\{\sum_{s=0}^{t}\eta_s\left\langle \widehat{\nabla}_{\pi_i} v_i(\pi^s) - \mu\nabla_{\pi_i} G(\pi_i^s, \sigma_i), x\right\rangle - \psi(x)\right\}.$$

Throughout this section, we assume that $\mathcal{X}_i$ is an affine subset, i.e., there exists a matrix $A \in \mathbb{R}^{k_i \times d_i}$, a vector $b \in \mathbb{R}^{k_i}$ such that $A\pi_i = b$ for all $\pi_i \in \mathcal{X}_i$. Furthermore, we assume that ψ is Legendre (Rockafellar, 1997; Lattimore & Szepesvári, 2020). Then, APFTRL enjoys the last-iterate convergence of π^T in full and noisy feedback settings by proving the following lemmas:

Lemma C.1. *Suppose that Assumptions 5.4 with $\beta, \gamma \in (0, \infty)$ and 5.5 hold. If we use the constant learning rate $\eta_t = \eta \in (0, \frac{2\mu\gamma\rho^2}{\mu^2\gamma\rho^2(\gamma+2\beta)+8L^2})$, the $k+1$-th slingshot strategy profile σ^{k+1} of APFTRL under the full feedback setting satisfies that:*

$$D_\psi(\pi^{\mu,\sigma^k}, \sigma^{k+1}) \leq D_\psi(\pi^{\mu,\sigma^k}, \sigma^{k})\left(1 - \frac{\eta\mu\gamma}{2}\right)^{T_\sigma}.$$

Lemma C.2. *Let $\theta = \frac{\mu^2\gamma\rho^2(\gamma+2\beta)+8L^2}{2\mu\gamma\rho^2}$ and $\kappa = \frac{\mu\gamma}{2}$. Suppose that Assumptions 5.4, 5.5, and B.5 hold, and the learning rate sequence of the form $\eta_t = 1/(\kappa(t - T_\sigma \cdot \lfloor t/T_\sigma \rfloor) + 2\theta)$ is used. Then, the $k+1$-th slingshot strategy profile σ^{k+1} of APFTRL under the noisy feedback setting satisfies that:*

$$\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \sigma^{k+1})] \leq \frac{2\theta - \kappa}{\kappa(T_\sigma - 1) + 2\theta}\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \sigma^{k})] + \frac{NC^2}{\rho(\kappa(T_\sigma - 1) + 2\theta)}\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T_\sigma - 1) + 1\right) + \frac{1}{2\theta}\right).$$

The proofs of these theorems can be found in Appendix F.5, F.6.

D. Additional Theorems

Based on this theorem, we can show that the gap function for π^t converges to the value of $\mathcal{O}(\mu)$.

Theorem D.1. *In the same setup of Lemma 5.6, the gap function for APMD is bounded as:*

$$\mathrm{GAP}(\pi^t) \leq \mu \cdot \mathrm{diam}(\mathcal{X}) \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i)\|^2} + \mathcal{O}\left(\left(1 - \frac{\eta\mu\gamma}{2}\right)^{\frac{t}{2}}\right).$$

We note that lower μ reduces the gap function for $\pi^{\mu,\sigma}$ (as in the first term of Theorem D.1), whereas higher μ makes π^t converge faster (as in the second term of Theorem D.1). That is, μ controls a trade-off between the speed of convergence and the gap function.

E. Proofs for Section 4

E.1. Proof of Theorem B.1 (Formal Version of Theoremf 4.1)

Proof of Theorem B.1. First, from Lemma B.3, we have for any $k \geq 0$:

$$\mathrm{GAP}(\sigma^{k+1}) \leq \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \|\pi^{\mu,\sigma^k} - \sigma^k\| + (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|.$$

Using Lemma B.2, we can upper bound the term of $\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|$ as follows:

$$\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|^2 \leq \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \left(1 - \frac{\eta\mu}{2}\right)^{T_\sigma}.$$

Combining these inequalities, we have for any $k \geq 0$:

$$\begin{aligned}\mathrm{GAP}(\sigma^{k+1}) &\leq \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \|\pi^{\mu,\sigma^k} - \sigma^k\| + \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} \cdot (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \|\pi^{\mu,\sigma^k} - \sigma^k\| \\ &\leq ((\mu + L) \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \|\pi^{\mu,\sigma^k} - \sigma^k\|,\end{aligned}$$

where the second inequality follows from $T_\sigma \geq 1$. Let us denote $K := \lfloor T/T_\sigma \rfloor$ as the total number of the slingshot strategy updates over the entire T iterations. By letting $k = K - 1$ in the above inequality, we get:

$$\mathrm{GAP}(\sigma^K) \leq ((\mu + L) \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|. \tag{6}$$

Next, we derive the following upper bound on $\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|$ from Lemma B.4:

$$\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\| \leq \frac{2\sqrt{2}}{\sqrt{K}} \sqrt{\mathrm{diam}(\mathcal{X}) \left(8 \cdot \mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)}. \tag{7}$$

By combining (6) and (7), we get:

$$\mathrm{GAP}(\sigma^K) \leq \frac{2\sqrt{2}\,((\mu + L) \cdot \mathrm{diam}(\mathcal{X}) + \zeta)}{\sqrt{K}} \sqrt{\mathrm{diam}(\mathcal{X}) \left(8 \cdot \mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)}.$$

Finally, since $\pi^T = \sigma^K$, $K = \lfloor T/T_\sigma \rfloor$, and $T_\sigma = c \cdot \max(\frac{6}{\ln 2 - \ln(2-\eta\mu)} \ln T + \frac{2 \ln 64}{\ln 2 - \ln(2-\eta\mu)}, 1)$, we have:

$$\begin{aligned}
&\mathrm{GAP}(\pi^T) \\
&\leq \frac{2\sqrt{2}\left((\mu + L) \cdot \mathrm{diam}(\mathcal{X}) + \zeta\right)}{\sqrt{T/T_\sigma}} \sqrt{\mathrm{diam}(\mathcal{X}) \left(8 \cdot \mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)} \\
&\leq \frac{2\sqrt{2}c\left((\mu + L) \cdot \mathrm{diam}(\mathcal{X}) + \zeta\right) \cdot \left(\frac{6}{\ln 2 - \ln(2-\eta\mu)} \ln T + \frac{2 \ln 64}{\ln 2 - \ln(2-\eta\mu)} + 1\right)}{\sqrt{T}} \sqrt{\mathrm{diam}(\mathcal{X}) \left(8 \cdot \mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)}.
\end{aligned}$$

This concludes the statement of the theorem. □

E.2. Proof of Lemma B.2 (Formal Version of Lemma 4.2)

Proof of Lemma B.2. From the definition of the Bregman divergence, we have for all $\pi_i, \pi_i' \in \mathcal{X}_i$:

$$\begin{aligned}
&D_\psi(\pi_i', \sigma_i) - D_\psi(\pi_i, \sigma_i) - \langle \nabla_{\pi_i} D_\psi(\pi_i, \sigma_i), \pi_i' - \pi_i \rangle \\
&= \psi(\pi_i') - \psi(\sigma_i) - \langle \nabla\psi(\sigma_i), \pi_i' - \sigma_i \rangle - \psi(\pi_i) + \psi(\sigma_i) + \langle \nabla\psi(\sigma_i), \pi_i - \sigma_i \rangle - \langle \nabla\psi(\pi_i) - \nabla\psi(\sigma_i), \pi_i' - \pi_i \rangle \\
&= \psi(\pi_i') - \psi(\pi_i) - \langle \nabla\psi(\pi_i), \pi_i' - \pi_i \rangle \\
&= D_\psi(\pi_i', \pi_i).
\end{aligned}$$

Hence, assuming that G is identical to D_ψ, Assumption 5.4 is satisfied with $\beta = \gamma = 1$. Furthermore, since $\psi(x) = \frac{1}{2}\|x\|^2$, both $\rho = 1$ and $\mathrm{int}(\mathrm{dom}\ \psi) = \mathbb{R}^{d_i}$ hold. Therefore, Assumption 5.5 is also satisfied. Consequently, we can obtain the following convergence result from Lemma B.9 with $\beta = \gamma = \rho = 1$:

$$\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|^2 \leq \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \left(1 - \frac{\eta\mu}{2}\right)^{T_\sigma}.$$

□

E.3. Proof of Lemma B.3 (Formal Version of Lemma 4.3)

Proof of Lemma B.3. First, we have for any $\pi \in \mathcal{X}$:

$$\begin{aligned}
\mathrm{GAP}(\pi) &= \max_{\tilde{\pi}_i \in \mathcal{X}_i} \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi), \tilde{\pi}_i - \pi_i \rangle \\
&= \max_{\tilde{\pi} \in \mathcal{X}} \sum_{i=1}^N \left(\langle \nabla_{\pi_i} v_i(\pi'), \tilde{\pi}_i - \pi_i' \rangle - \langle \nabla_{\pi_i} v_i(\pi'), \pi_i - \pi_i' \rangle + \langle \nabla_{\pi_i} v_i(\pi) - \nabla_{\pi_i} v_i(\pi'), \tilde{\pi}_i - \pi_i \rangle \right). \qquad (8)
\end{aligned}$$

Here, we introduce the following lemma from Cai et al. (2022a):

Lemma E.1 (Lemma 2 of Cai et al. (2022a))**.** *For any $\pi \in \mathcal{X}$, we have:*

$$\max_{\tilde{\pi} \in \mathcal{X}} \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi), \tilde{\pi}_i - \pi_i \rangle \leq \mathrm{diam}(\mathcal{X}) \cdot \min_{(a_i) \in N_{\mathcal{X}}(\pi)} \sqrt{\sum_{i=1}^N \| -\nabla_{\pi_i} v_i(\pi) + a_i \|^2},$$

where $N_{\mathcal{X}}(\pi) = \{(a_i)_{i \in [N]} \in \prod_{i=1}^N \mathbb{R}^{d_i} \mid \sum_{i=1}^N \langle a_i, \pi_i' - \pi_i \rangle \leq 0,\ \forall \pi' \in \mathcal{X}\}$.

From Lemma E.1, the first term of (8) can be upper bounded as:

$$\max_{\tilde{\pi} \in \mathcal{X}} \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi'), \tilde{\pi}_i - \pi_i' \rangle \leq \mathrm{diam}(\mathcal{X}) \cdot \min_{(a_i) \in N_{\mathcal{X}}(\pi')} \sqrt{\sum_{i=1}^N \| -\nabla_{\pi_i} v_i(\pi') + a_i \|^2} \qquad (9)$$

Next, from Cauchy-Schwarz inequality, the second term of (8) can be upper bounded as:

$$\begin{aligned} -\sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\pi'), \pi_i - \pi_i' \rangle &\leq \|\pi - \pi'\| \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} v_i(\pi')\|^2} \\ &\leq \zeta \|\pi - \pi'\|. \end{aligned} \tag{10}$$

Again from Cauchy-Schwarz inequality, the third term of (8) can be upper bounded as:

$$\begin{aligned} \sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\pi) - \nabla_{\pi_i} v_i(\pi'), \tilde{\pi}_i - \pi_i \rangle &\leq \|\tilde{\pi} - \pi\| \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} v_i(\pi) - \nabla_{\pi_i} v_i(\pi')\|^2} \\ &\leq \mathrm{diam}(\mathcal{X}) \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} v_i(\pi) - \nabla_{\pi_i} v_i(\pi')\|^2} \\ &\leq L \cdot \mathrm{diam}(\mathcal{X}) \cdot \|\pi - \pi'\|. \end{aligned} \tag{11}$$

By combining (8), (9), (10), and (11), we get for any $\pi, \pi' \in \mathcal{X}$:

$$\mathrm{GAP}(\pi) \leq \mathrm{diam}(\mathcal{X}) \cdot \min_{(a_i) \in N_{\mathcal{X}}(\pi')} \sqrt{\sum_{i=1}^{N} \| - \nabla_{\pi_i} v_i(\pi') + a_i\|^2} + (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \|\pi - \pi'\|.$$

Thus, letting $\pi = \sigma^{k+1}$ and $\pi' = \pi^{\mu,\sigma^k}$, we have:

$$\mathrm{GAP}(\sigma^{k+1}) \leq \mathrm{diam}(\mathcal{X}) \cdot \min_{(a_i) \in N_{\mathcal{X}}(\pi^{\mu,\sigma^k})} \sqrt{\sum_{i=1}^{N} \| - \nabla_{\pi_i} v_i(\pi^{\mu,\sigma^k}) + a_i\|^2} + (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|. \tag{12}$$

On the other hand, from the first-order optimality condition for π^{μ,σ^k}, we have for any $\pi \in \mathcal{X}$:

$$\sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\pi^{\mu,\sigma^k}) - \mu \left(\pi_i^{\mu,\sigma^k} - \sigma_i^k\right), \pi_i - \pi_i^{\mu,\sigma^k} \rangle \leq 0,$$

and then $\left(\nabla_{\pi_i} v_i(\pi^{\mu,\sigma^k}) - \mu \left(\pi_i^{\mu,\sigma^k} - \sigma_i^k\right)\right)_{i \in [N]} \in N_{\mathcal{X}}(\pi^{\mu,\sigma^k})$. Thus, the first term of (12) can be bounded as:

$$\begin{aligned} &\min_{(a_i) \in N_{\mathcal{X}}(\pi^{\mu,\sigma^k})} \sqrt{\sum_{i=1}^{N} \| - \nabla_{\pi_i} v_i(\pi^{\mu,\sigma^k}) + a_i\|^2} \\ &\leq \sqrt{\sum_{i=1}^{N} \| - \nabla_{\pi_i} v_i(\pi^{\mu,\sigma^k}) + \nabla_{\pi_i} v_i(\pi^{\mu,\sigma^k}) - \mu \left(\pi_i^{\mu,\sigma^k} - \sigma_i^k\right) \|^2} \\ &= \mu \|\pi^{\mu,\sigma^k} - \sigma^k\|. \end{aligned} \tag{13}$$

Combining (12) and (13), we have:

$$\mathrm{GAP}(\sigma^{k+1}) \leq \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \|\pi^{\mu,\sigma^k} - \sigma^k\| + (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|$$

□

E.4. Proof of Lemma B.4 (Formal Version of Lemma 4.4)

Proof of Lemma B.4. First, we prove the following lemma:

Lemma E.2. *Assume that* $\sqrt{\sum_{i=1}^N \|\nabla_{\pi_i} v_i(\pi)\|^2} \le \zeta$ *for any* $\pi \in \mathcal{X}$*. If* G *is set as the squared* ℓ^2*-distance, we have for any* $k \in [K]$*:*

$$\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \le \|\sigma^k - \sigma^{k-1}\|^2 + \frac{2\zeta}{\mu}\|\pi^{\mu,\sigma^{k-1}} - \sigma^k\|.$$

From Lemma E.2, we can bound $\|\pi^{\mu,\sigma^k} - \sigma^k\|^2$ as:

$$\begin{aligned}
&\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \\
&\le \|\sigma^k - \sigma^{k-1}\|^2 + \frac{2\zeta}{\mu}\|\pi^{\mu,\sigma^{k-1}} - \sigma^k\| \\
&\le \|\pi^{\mu,\sigma^{k-1}} - \sigma^{k-1}\|^2 + \|\sigma^k - \pi^{\mu,\sigma^{k-1}}\|^2 + 2\|\sigma^k - \pi^{\mu,\sigma^{k-1}}\|\|\pi^{\mu,\sigma^{k-1}} - \sigma^{k-1}\| + \frac{2\zeta}{\mu}\|\pi^{\mu,\sigma^{k-1}} - \sigma^k\|. \qquad (14)
\end{aligned}$$

Next, we upper bound $\|\pi^{\mu,\sigma^k} - \sigma^k\|^2$ using the following lemma:

Lemma E.3. *Assume that* $T_\sigma \ge \max(\frac{6}{\ln 2 - \ln(2-\eta\mu)} \ln T + \frac{2\ln 64}{\ln 2 - \ln(2-\eta\mu)}, 1)$*. In the same setup of Theorem 4.1, we have for any Nash equilibrium* $\pi^* \in \Pi^*$*:*

$$\sum_{k=0}^{K-1} \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \le 16\|\pi^* - \sigma^0\|^2.$$

Using Lemma E.3, we have:

$$\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \le \sum_{k=0}^{K-1} \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \le 16\|\pi^* - \sigma^0\|^2,$$

and then from Lemma B.2, we get:

$$\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\| \le \|\pi^{\mu,\sigma^k} - \sigma^k\| \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} \le 4\|\pi^* - \sigma^0\| \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}}. \qquad (15)$$

By combining (14) and (15), we have:

$$\begin{aligned}
&\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \\
&\le \|\pi^{\mu,\sigma^{k-1}} - \sigma^{k-1}\|^2 + 16\|\pi^* - \sigma^0\| \left(1 - \frac{\eta\mu}{2}\right)^{T_\sigma} + 32\|\pi^* - \sigma^0\|^2 \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} + \frac{8\zeta}{\mu}\|\pi^* - \sigma^0\| \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} \\
&\le \|\pi^{\mu,\sigma^{k-1}} - \sigma^{k-1}\|^2 + 48\|\pi^* - \sigma^0\|^2 \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} + \frac{8\zeta}{\mu}\|\pi^* - \sigma^0\| \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} \\
&\le \|\pi^{\mu,\sigma^{k-1}} - \sigma^{k-1}\|^2 + 8\|\pi^* - \sigma^0\| \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} \left(6\|\pi^* - \sigma^0\| + \frac{\zeta}{\mu}\right).
\end{aligned}$$

Therefore, we get:

$$\begin{aligned}
\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|^2 &\le \|\pi^{\mu,\sigma^{K-2}} - \sigma^{K-2}\|^2 + 8\|\pi^* - \sigma^0\| \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} \left(6\|\pi^* - \sigma^0\| + \frac{\zeta}{\mu}\right) \\
&\le \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 + 8K\|\pi^* - \sigma^0\| \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} \left(6\|\pi^* - \sigma^0\| + \frac{\zeta}{\mu}\right). \qquad (16)
\end{aligned}$$

By combining (16) and Lemma E.3, we have:

$$\begin{aligned}
K\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|^2 &\leq \sum_{k=0}^{K-1} \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 + 8K^2\|\pi^* - \sigma^0\| \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} \left(6\|\pi^* - \sigma^0\| + \frac{\zeta}{\mu}\right) \\
&\leq 16\|\pi^* - \sigma^0\|^2 + 8K^2\|\pi^* - \sigma^0\| \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} \left(6\|\pi^* - \sigma^0\| + \frac{\zeta}{\mu}\right) \\
&\leq 16\|\pi^* - \sigma^0\|^2 + 8T^2\|\pi^* - \sigma^0\| \left(1 - \frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}} \left(6\|\pi^* - \sigma^0\| + \frac{\zeta}{\mu}\right).
\end{aligned}$$

Under the assumption that $T_\sigma \geq \max(\frac{6}{\ln 2 - \ln(2-\eta\mu)} \ln T + \frac{2\ln 64}{\ln 2 - \ln(2-\eta\mu)}, 1)$, we get:

$$\left(1 - \frac{\eta\mu}{2}\right)^{-\frac{T_\sigma}{2}} \geq \left(1 - \frac{\eta\mu}{2}\right)^{-\frac{3}{\ln 2 - \ln(2-\eta\mu)} \ln T} \left(1 - \frac{\eta\mu}{2}\right)^{\frac{\ln 64}{\ln\left(1 - \frac{\eta\mu}{2}\right)}} = 64\left(1 - \frac{\eta\mu}{2}\right)^{\frac{\ln T^3}{\ln(2-\eta\mu) - \ln 2}} = 64T^3.$$

Thus, we have:

$$\begin{aligned}
K\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|^2 &\leq 16\|\pi^* - \sigma^0\|^2 + 8\|\pi^* - \sigma^0\| \left(6\|\pi^* - \sigma^0\| + \frac{\zeta}{\mu}\right) \\
&= 8\|\pi^* - \sigma^0\| \left(8\|\pi^* - \sigma^0\| + \frac{\zeta}{\mu}\right) \\
&\leq 8 \cdot \mathrm{diam}(\mathcal{X}) \left(8 \cdot \mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right).
\end{aligned}$$

□

E.5. Proof of Theorem B.6 (Formal Version of Theorem 4.5)

Proof of Theorem B.6. First, from Lemma B.3, we have for any $k \geq 0$:

$$\mathbb{E}\left[\mathrm{GAP}(\sigma^{k+1})\right] \leq \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \mathbb{E}\left[\|\pi^{\mu,\sigma^k} - \sigma^k\|\right] + (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \mathbb{E}\left[\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|\right].$$

Using Lemma B.7, we can upper bound the term of $\mathbb{E}\left[\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|^2\right]$ as follows:

$$\mathbb{E}\left[\|\pi^{\mu,\sigma^k} - \sigma^{k+1}\|^2\right] \leq \frac{2\theta - \kappa}{\kappa(T_\sigma - 1) + 2\theta} \mathbb{E}\left[\|\pi^{\mu,\sigma^k} - \sigma^k\|^2\right] + \frac{NC^2}{\rho(\kappa(T_\sigma - 1) + 2\theta)} \left(\frac{1}{\kappa} \ln\left(\frac{\kappa}{2\theta}(T_\sigma - 1) + 1\right) + \frac{1}{2\theta}\right).$$

Combining these inequalities, we have for any $k \geq 0$:

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{GAP}(\sigma^{k+1})\right] \\
&\leq \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \mathbb{E}\left[\|\pi^{\mu,\sigma^k} - \sigma^k\|\right] + (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \sqrt{\frac{\rho(2\theta - \kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T + 1\right) + \frac{1}{2\theta}\right)}{\rho\kappa T^{4/5}}},
\end{aligned}$$

where we use $T_\sigma \geq \max(T^{4/5} + 2, 3)$. Let us denote $K := \lfloor T/T_\sigma \rfloor$ as the total number of the slingshot strategy updates over the entire T iterations. By letting $k = K - 1$ in the above inequality, we get:

$$\begin{aligned}
\mathbb{E}\left[\mathrm{GAP}(\sigma^K)\right] &\leq \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \mathbb{E}\left[\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|\right] \\
&\quad + (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \sqrt{\frac{\rho(2\theta - \kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T + 1\right) + \frac{1}{2\theta}\right)}{\rho\kappa T^{4/5}}},
\end{aligned} \tag{17}$$

Next, we derive the following upper bound on $\mathbb{E}\left[\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|\right]$ from Lemma B.8:

$$\begin{aligned}
&\mathbb{E}\left[\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|\right] \\
&\leq \sqrt{\frac{\|\pi^* - \sigma^0\|^2 + \frac{3}{\rho\kappa}\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu} + 1\right)\left(\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)\right)}{K}}.
\end{aligned} \tag{18}$$

By combining (17) and (18), we get:

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{GAP}(\sigma^K)\right] \\
&\leq \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \sqrt{\frac{\|\pi^* - \sigma^0\|^2 + \frac{3}{\rho\kappa}\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu} + 1\right)\left(\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)\right)}{K}} \\
&+ (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)}{\rho\kappa T^{4/5}}}.
\end{aligned}$$

Finally, since $\pi^T = \sigma^K$, $K = \lfloor T/T_\sigma \rfloor$, and $T_\sigma = c \cdot \max(T^{4/5} + 2, 3)$, we have:

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{GAP}(\pi^T)\right] \\
&\leq \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \sqrt{\frac{\|\pi^* - \sigma^0\|^2 + \frac{3}{\rho\kappa}\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu} + 1\right)\left(\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)\right)}{T/T_\sigma}} \\
&+ (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)}{\rho\kappa T^{4/5}}} \\
&\leq \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \sqrt{c(T^{4/5}+5)} \\
&\cdot \sqrt{\frac{\|\pi^* - \sigma^0\|^2 + \frac{3}{\rho\kappa}\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu} + 1\right)\left(\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)\right)}{T}} \\
&+ (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)}{\rho\kappa T^{4/5}}} \\
&\leq \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \sqrt{\frac{6c\|\pi^* - \sigma^0\|^2 + \frac{18c}{\rho\kappa}\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu} + 1\right)\left(\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)\right)}{T^{1/5}}} \\
&+ (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)}{\rho\kappa T^{4/5}}} \\
&\leq \mu \cdot \mathrm{diam}(\mathcal{X})^2 \cdot \sqrt{\frac{6c}{T^{1/5}}} \\
&+ \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \sqrt{\frac{18c\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu} + 1\right)\left(\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)\right)}{\rho\kappa T^{1/5}}} \\
&+ (L \cdot \mathrm{diam}(\mathcal{X}) + \zeta) \cdot \sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right) + \frac{1}{2\theta}\right)}{\rho\kappa T^{1/5}}} \\
&\leq \frac{\sqrt{6c}\mu \cdot \mathrm{diam}(\mathcal{X})^2}{T^{1/10}}
\end{aligned}$$

$$+\frac{L\cdot\mathrm{diam}(\mathcal{X})+\zeta+\mu\cdot\mathrm{diam}(\mathcal{X})\sqrt{18c\left(\mathrm{diam}(\mathcal{X})+\frac{\zeta}{\mu}+1\right)}}{T^{1/10}}\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2+NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho\kappa}}.$$

□

This concludes the statement of the theorem.

E.6. Proof of Lemma B.7 (Formal Version of Lemma 4.6)

Proof of Lemma B.7. Assuming that G is identical to D_ψ, Assumption 5.4 is satisfied with $\beta=\gamma=1$. Furthermore, since $\psi(x)=\frac{1}{2}\|x\|^2$, both $\rho=1$ and $\mathrm{int}(\mathrm{dom}\ \psi)=\mathbb{R}^{d_i}$ hold. Therefore, Assumption 5.5 is also satisfied. Consequently, we can apply Lemma B.10 with $\beta=\gamma=\rho=1$ and obtain:

$$\mathbb{E}\left[\|\pi^{\mu,\sigma^k}-\sigma^{k+1}\|^2\right]\leq\frac{2\theta-\kappa}{\kappa(T_\sigma-1)+2\theta}\mathbb{E}\left[\|\pi^{\mu,\sigma^k}-\sigma^k\|^2\right]+\frac{NC^2}{\rho(\kappa(T_\sigma-1)+2\theta)}\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T_\sigma-1)+1\right)+\frac{1}{2\theta}\right).$$

□

E.7. Proof of Lemma B.8 (Formal Version of Lemma 4.7)

Proof of Lemma B.8. From Lemmas E.2 and B.7, we have:

$$\begin{aligned}
&\mathbb{E}\left[\|\pi^{\mu,\sigma^k}-\sigma^k\|^2\right]\\
&\leq\mathbb{E}\left[\|\sigma^k-\sigma^{k-1}\|^2\right]+\frac{2\zeta}{\mu}\mathbb{E}\left[\|\pi^{\mu,\sigma^{k-1}}-\sigma^k\|\right]\\
&\leq\mathbb{E}\left[\|\sigma^k-\pi^{\mu,\sigma^{k-1}}\|^2\right]+\mathbb{E}\left[\|\pi^{\mu,\sigma^{k-1}}-\sigma^{k-1}\|^2\right]+\mathbb{E}\left[2\|\sigma^k-\pi^{\mu,\sigma^{k-1}}\|\|\pi^{\mu,\sigma^{k-1}}-\sigma^{k-1}\|+\frac{2\zeta}{\mu}\|\pi^{\mu,\sigma^{k-1}}-\sigma^k\|\right]\\
&\leq\frac{2\theta-\kappa}{\kappa(T_\sigma-2)+2\theta}\mathbb{E}\left[\|\pi^{\mu,\sigma^k}-\sigma^k\|^2\right]+\frac{NC^2}{\rho(\kappa(T_\sigma-2)+2\theta)}\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T_\sigma-2)+1\right)+\frac{1}{2\theta}\right)\\
&+2\left(\mathrm{diam}(\mathcal{X})+\frac{\zeta}{\mu}\right)\sqrt{\frac{2\theta-\kappa}{\kappa(T_\sigma-2)+2\theta}\mathbb{E}\left[\|\pi^{\mu,\sigma^{k-1}}-\sigma^{k-1}\|^2\right]+\frac{NC^2}{\rho(\kappa(T_\sigma-2)+2\theta)}\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T_\sigma-2)+1\right)+\frac{1}{2\theta}\right)}\\
&+\mathbb{E}\left[\|\pi^{\mu,\sigma^{k-1}}-\sigma^{k-1}\|^2\right]\\
&\leq\mathbb{E}\left[\|\pi^{\mu,\sigma^{k-1}}-\sigma^{k-1}\|^2\right]+\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2+NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T_\sigma-2)+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa(T_\sigma-2)+2\theta)}\\
&+2\left(\mathrm{diam}(\mathcal{X})+\frac{\zeta}{\mu}\right)\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2+NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T_\sigma-2)+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa(T_\sigma-2)+2\theta)}}.
\end{aligned}$$

Under the assumption that $T_\sigma\geq\max(T^{4/5}+2,3)$, we get:

$$\begin{aligned}
\mathbb{E}\left[\|\pi^{\mu,\sigma^k}-\sigma^k\|^2\right]\leq\mathbb{E}\left[\|\pi^{\mu,\sigma^{k-1}}-\sigma^{k-1}\|^2\right]+\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2+NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)}\\
+2\left(\mathrm{diam}(\mathcal{X})+\frac{\zeta}{\mu}\right)\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2+NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)}}.
\end{aligned}$$

Therefore, we get:

$$\begin{aligned}\mathbb{E}[\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|^2] &\leq \mathbb{E}[\|\pi^{\mu,\sigma^{K-2}} - \sigma^{K-2}\|^2] + \frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)} \\ &\quad + 2\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)}} \\ &\leq \mathbb{E}[\|\pi^{\mu,\sigma^{k}} - \sigma^{k}\|^2] + K\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)} \\ &\quad + 2K\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)}}. \end{aligned} \tag{19}$$

Here, we derive the following upper bound in terms of $\mathbb{E}[\|\pi^{\mu,\sigma^k} - \sigma^k\|^2]$:

Lemma E.4. *Assume that $T_\sigma \geq \max(T^{4/5}+2, 3)$. In the same setup of Theorem 4.5, we have for any Nash equilibrium $\pi^* \in \Pi^*$:*

$$\mathbb{E}\left[\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2\right] \leq \|\pi^* - \sigma^0\|^2 + K\cdot\mathrm{diam}(\mathcal{X})\cdot\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)}}.$$

By combining (19), Lemma E.4, and the assumption that $T_\sigma \geq \max(T^{4/5}+2, 3)$, we have:

$$\begin{aligned}&K\mathbb{E}[\|\pi^{\mu,\sigma^{K-1}} - \sigma^{K-1}\|^2] \\ &\leq \mathbb{E}\left[\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2\right] + K^2\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)} \\ &+ 2K^2\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)}} \\ &\leq \|\pi^* - \sigma^0\|^2 + K\cdot\mathrm{diam}(\mathcal{X})\cdot\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T-1)+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)}} \\ &+ K^2\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)} \\ &+ 2K^2\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)}} \\ &\leq \|\pi^* - \sigma^0\|^2 + T^{1/5}\cdot\mathrm{diam}(\mathcal{X})\cdot\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T-1)+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)}} \\ &+ T^{2/5}\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)} \\ &+ 2T^{2/5}\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu}\right)\sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5}+2\theta)}} \\ &\leq \|\pi^* - \sigma^0\|^2 + 3\left(\mathrm{diam}(\mathcal{X}) + \frac{\zeta}{\mu} + 1\right)\left(\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T+1\right)+\frac{1}{2\theta}\right)}{\rho\kappa}\right).\end{aligned}$$

□

F. Proofs for Section 5

F.1. Proof of Lemma B.9 (Formal Version of Lemma 5.6)

Proof of Lemma B.9. From the definition of the Bregman divergence, we have for any $t \in \{kT_\sigma, kT_\sigma+1, \cdots, (k+1)T_\sigma - 1\}$:

$$\begin{aligned}
&D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^{t+1}) - D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t) + D_\psi(\pi_i^{t+1}, \pi_i^t) \\
&= \psi(\pi_i^{\mu,\sigma^k}) - \psi(\pi_i^{t+1}) - \langle \nabla\psi(\pi_i^{t+1}), \pi_i^{\mu,\sigma^k} - \pi_i^{t+1}\rangle \\
&- \psi(\pi_i^{\mu,\sigma^k}) + \psi(\pi_i^t) + \langle \nabla\psi(\pi_i^t), \pi_i^{\mu,\sigma^k} - \pi_i^t\rangle \\
&+ \psi(\pi_i^{t+1}) - \psi(\pi_i^t) - \langle \nabla\psi(\pi_i^t), \pi_i^{t+1} - \pi_i^t\rangle \\
&= \langle \nabla\psi(\pi_i^t) - \nabla\psi(\pi_i^{t+1}), \pi_i^{\mu,\sigma^k} - \pi_i^{t+1}\rangle. \qquad (20)
\end{aligned}$$

From the first-order optimality condition for π_i^{t+1}, we get for any $t \in \{kT_\sigma, kT_\sigma+1, \cdots, (k+1)T_\sigma - 1\}$:

$$\langle \eta(\nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu\nabla_{\pi_i} G(\pi_i^t, \sigma_i^k)) - \nabla\psi(\pi_i^{t+1}) + \nabla\psi(\pi_i^t), \pi_i^{\mu,\sigma^k} - \pi_i^{t+1}\rangle \le 0. \qquad (21)$$

Note that $\nabla_{\pi_i} G(\pi_i^t, \sigma_i^k)$ and $\nabla_{\pi_i}\psi(\pi_i^t)$ are well-defined because of Assumptions 5.4 and 5.5. By combining (20) and (21), we have:

$$D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^{t+1}) - D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t) + D_\psi(\pi_i^{t+1}, \pi_i^t) \le \eta\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu\nabla_{\pi_i} G(\pi_i^t, \sigma_i^k), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle. \qquad (22)$$

Next, we derive the following convergence result for π^t:

Lemma F.1. *Suppose that Assumption 5.4 holds with $\beta, \gamma \in (0, \infty)$, and the updated strategy profile π^t satisfies the following condition: for any $t \in \{kT_\sigma, kT_\sigma+1, \cdots, (k+1)T_\sigma - 1\}$,*

$$D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) - D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + D_\psi(\pi^{t+1}, \pi^t) \le \eta\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu\nabla_{\pi_i} G(\pi_i^t, \sigma_i^k), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle.$$

Then, for any $t \in \{kT_\sigma, kT_\sigma+1, \cdots, (k+1)T_\sigma - 1\}$:

$$D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) \le D_\psi(\pi^{\mu,\sigma^k}, \pi^{kT_\sigma})\left(1 - \frac{\eta\mu\gamma}{2}\right)^{t-kT_\sigma+1}.$$

It is easy to confirm that (22) satisfies the assumption in Lemma F.1. Thus, taking $t = (k+1)T_\sigma - 1$, we have:

$$D_\psi(\pi^{\mu,\sigma^k}, \pi^{(k+1)T_\sigma}) \le D_\psi(\pi^{\mu,\sigma^k}, \pi^{kT_\sigma})\left(1 - \frac{\eta\mu\gamma}{2}\right)^{T_\sigma}.$$

Since $\sigma^k = \pi^{kT_\sigma}$ and $\sigma^{k+1} = \pi^{(k+1)T_\sigma}$, we conclude the statement. □

F.2. Proof of Lemma B.10 (Formal Version of Lemma 5.7)

Proof of Lemma B.10. Writing $g_i^t = \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu\nabla_{\pi_i} G(\pi_i^t, \sigma_i^k)$, from the first-order optimality condition for π_i^{t+1}, we get for any $t \in \{kT_\sigma, kT_\sigma+1, \cdots, (k+1)T_\sigma - 1\}$:

$$\langle \eta_t(g_i^t + \xi_i^t) - \nabla\psi(\pi_i^{t+1}) + \nabla\psi(\pi_i^t), \pi_i^{\mu,\sigma^k} - \pi_i^{t+1}\rangle \le 0. \qquad (23)$$

Note that $\nabla_{\pi_i} G(\pi_i^t, \sigma_i^k)$ and $\nabla_{\pi_i}\psi(\pi_i^t)$ are well-defined because of Assumptions 5.4 and 5.5. By combining (20) and (23), we have for any $t \in \{kT_\sigma, kT_\sigma+1, \cdots, (k+1)T_\sigma - 1\}$:

$$D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^{t+1}) - D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t) + D_\psi(\pi_i^{t+1}, \pi_i^t) \le \eta_t\langle g_i^t + \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle. \qquad (24)$$

We have the following Lemma that replaces the gradient with Bregman divergences:

Lemma F.2. *Under the noisy feedback setting, suppose that Assumption 5.4 holds with $\beta, \gamma \in (0, \infty)$, and the updated strategy profile π^t satisfies the following condition: for any $t \in \{kT_\sigma, kT_\sigma + 1, \cdots, (k+1)T_\sigma - 1\}$,*

$$\begin{aligned} &D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) - D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + D_\psi(\pi^{t+1}, \pi^t) \\ &\quad \leq \eta_t \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu \nabla_{\pi_i} G(\pi_i^t, \sigma_i^k) + \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle. \end{aligned}$$

Then, for any $t \in \{kT_\sigma, kT_\sigma + 1, \cdots, (k+1)T_\sigma - 1\}$:

$$\begin{aligned} &D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) - D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + D_\psi(\pi^{t+1}, \pi^t) \\ &\leq \eta_t \left(\left(\frac{\mu^2 \gamma \rho^2 (\gamma + 2\beta) + 8L^2}{2\mu\gamma\rho^2} \right) D_\psi(\pi^{t+1}, \pi^t) - \frac{\mu\gamma}{2} D_\psi(\pi^{\mu,\sigma^k}, \pi^t) \right) + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle. \end{aligned}$$

It is easy to confirm that (24) satisfies the assumption in Lemma F.2. Thus, for any $t \in \{kT_\sigma, kT_\sigma + 1, \cdots, (k+1)T_\sigma - 1\}$:

$$\begin{aligned} &D_\psi(\pi^{\mu,r}, \pi^{t+1}) - D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + D_\psi(\pi^{t+1}, \pi^t) \\ &\leq \eta_t \left(\left(\frac{\mu^2 \gamma \rho^2 (\gamma + 2\beta) + 8L^2}{2\mu\gamma\rho^2} \right) D_\psi(\pi^{t+1}, \pi^t) - \frac{\mu\gamma}{2} D_\psi(\pi^{\mu,\sigma^k}, \pi^t) \right) + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle \\ &= \eta_t (\theta D_\psi(\pi^{t+1}, \pi^t) - \kappa D_\psi(\pi^{\mu,\sigma^k}, \pi^t)) + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle. \end{aligned}$$

Then, using this inequality, we can bound the expected value of $D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1})$ as follows:

Lemma F.3. *Suppose that with some constants $\theta > \kappa > 0$, for all $t \in \{kT_\sigma, kT_\sigma + 1, \cdots, (k+1)T_\sigma - 1\}$, the following inequality holds:*

$$\begin{aligned} &D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) - D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + D_\psi(\pi^{t+1}, \pi^t) \\ &\quad \leq \eta_t (\theta D_\psi(\pi^{t+1}, \pi^t) - \kappa D_\psi(\pi^{\mu,\sigma^k}, \pi^t)) + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle. \end{aligned}$$

Then, under Assumption B.5, for any $t \in \{kT_\sigma, kT_\sigma + 1, \cdots, (k+1)T_\sigma - 1\}$,

$$\begin{aligned} &\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1})] \\ &\leq \frac{2\theta - \kappa}{\kappa(t - kT_\sigma) + 2\theta} \mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{kT_\sigma})] + \frac{NC^2}{\rho(\kappa(t - kT_\sigma) + 2\theta)} \left(\frac{1}{\kappa} \ln \left(\frac{\kappa}{2\theta}(t - kT_\sigma) + 1 \right) + \frac{1}{2\theta} \right). \end{aligned}$$

Taking $t = (k+1)T_\sigma - 1$, we have:

$$\begin{aligned} &\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{(k+1)T_\sigma})] \\ &\leq \frac{2\theta - \kappa}{\kappa(T_\sigma - 1) + 2\theta} \mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{kT_\sigma})] + \frac{NC^2}{\rho(\kappa(T_\sigma - 1) + 2\theta)} \left(\frac{1}{\kappa} \ln \left(\frac{\kappa}{2\theta}(T_\sigma - 1) + 1 \right) + \frac{1}{2\theta} \right). \end{aligned}$$

Since $\sigma^k = \pi^{kT_\sigma}$ and $\sigma^{k+1} = \pi^{(k+1)T_\sigma}$, we conclude the statement. □

F.3. Proof of Theorem 5.8

Proof of Theorem 5.8. When $G(\pi_i, \pi_i') = D_{\psi'}(\pi_i, \pi_i')$ for all $i \in [N]$ and $\pi, \pi' \in \mathcal{X}$, we can show that the Bregman divergence from a Nash equilibrium $\pi^* \in \Pi^*$ to σ^{k+1} monotonically decreases:

Lemma F.4. *Assume that G is a Bregman divergence $D_{\psi'}$ for some strongly convex function ψ'. Then, for any Nash equilibrium $\pi^* \in \Pi^*$ of the underlying game, we have for any $k \geq 0$:*

$$D_{\psi'}(\pi^*, \sigma^{k+1}) - D_{\psi'}(\pi^*, \sigma^k) \leq -D_{\psi'}(\sigma^{k+1}, \sigma^k).$$

By summing the inequality in Lemma F.4 from $k = 0$ to K, we have:

$$D_{\psi'}(\pi^*, \sigma^0) \geq \sum_{k=0}^{K} D_{\psi'}(\sigma^{k+1}, \sigma^k) \geq \frac{\rho}{2} \sum_{k=0}^{K} \|\sigma^{k+1} - \sigma^k\|^2,$$

where the second inequality follows from the strong convexity of ψ'. Therefore, $\sum_{k=0}^{\infty} \|\sigma^{k+1} - \sigma^k\|^2 < \infty$, which implies that $\|\sigma^{k+1} - \sigma^k\| \to 0$ as $k \to \infty$.

By the compactness of $\mathcal{X}$ and Bolzano–Weierstrass theorem, there exists a subsequence k_n and a limit point $\hat{\sigma} \in \mathcal{X}$ such that $\sigma^{k_n} \to \hat{\sigma}$ as $n \to \infty$. Since $\|\sigma^{k_n+1} - \sigma^{k_n}\| \to 0$ as $n \to \infty$, we have $\sigma^{k_n+1} \to \hat{\sigma}$ as $n \to \infty$. Thus, the limit point $\hat{\sigma}$ is the fixed point of the updating rule. From the following lemma, we show that the fixed point $\hat{\sigma}$ is a Nash equilibrium of the underlying game:

Lemma F.5. *Assume that G is a Bregman divergence $D_{\psi'}$ for some strongly convex function ψ', and $\sigma^{k+1} = \pi^{\mu,\sigma^k}$ for $k \geq 0$. If $\sigma^{k+1} = \sigma^k$, then σ^k is a Nash equilibrium of the underlying game.*

On the other hand, by summing the inequality in Lemma F.4 from $k = k_n$ to $k = K - 1$ for $K \geq k_n + 1$, we have:

$$0 \leq D_{\psi'}(\hat{\sigma}, \sigma^K) \leq D_{\psi'}(\hat{\sigma}, \sigma^{k_n}).$$

Since $\sigma^{k_n} \to \hat{\sigma}$ as $n \to \infty$, we have $\sigma^K \to \hat{\sigma}$ as $K \to \infty$. Since $\hat{\sigma}$ is a Nash equilibrium of the underlying game, we conclude the first statement of the theorem.

□

F.4. Proof of Theorem 5.9

Proof of Theorem 5.9. We first show that the divergence between Π^* and σ^k decreases monotonically as k increases:

Lemma F.6. *Suppose that the same assumptions in Theorem 5.9 hold. For any $k \geq 0$, if $\sigma^k \in \mathcal{X} \setminus \Pi^*$, then:*

$$\min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^{k+1}) < \min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^k).$$

Otherwise, if $\sigma^k \in \Pi^*$, then $\sigma^{k+1} = \sigma^k \in \Pi^*$.

From Lemma F.6, the sequence $\{\min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^k)\}_{k \geq 0}$ is a monotonically decreasing sequence and is bounded from below by zero. Thus, $\{\min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^k)\}_{k \geq 0}$ converges to some constant $b \geq 0$. We show that $b = 0$ by a contradiction argument.

Suppose $b > 0$ and let us define $B = \min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^0)$. Since $\min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^k)$ monotonically decreases, σ^k is in the set $\Omega_{b,B} = \{\sigma \in \mathcal{X} \mid b \leq \min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma) \leq B\}$ for all $k \geq 0$. Since $\min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \cdot)$ is a continuous function on $\mathcal{X}$, the preimage $\Omega_{b,B}$ of the closed set $[b, B]$ is also closed. Furthermore, since $\mathcal{X}$ is compact and then bounded, $\Omega_{b,B}$ is a bounded set. Thus, $\Omega_{b,B}$ is a compact set.

Next, we show that the function which maps the slingshot strategies σ to the associated stationary point $\pi^{\mu,\sigma}$ is continuous:

Lemma F.7. *Let $F(\sigma) : \mathcal{X} \to \mathcal{X}$ be a function that maps the slingshot strategies σ to the stationary point $\pi^{\mu,\sigma}$ defined by (4). In the same setup of Theorem 5.9, $F(\cdot)$ is a continuous function on $\mathcal{X}$.*

From Lemma F.7, $\min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, F(\sigma)) - \min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma)$ is also a continuous function. Since a continuous function has a maximum over a compact set $\Omega_{b,B}$, the maximum $\delta =$

$\max_{\sigma\in\Omega_{b,B}}\{\min_{\pi^*\in\Pi^*}\mathrm{KL}(\pi^*,F(\sigma))-\min_{\pi^*\in\Pi^*}\mathrm{KL}(\pi^*,\sigma)\}$ exists. From Lemma F.6 and the assumption that $b>0$, we have $\delta<0$. It follows that:

$$\min_{\pi^*\in\Pi^*}\mathrm{KL}(\pi^*,\sigma^k)=\min_{\pi^*\in\Pi^*}\mathrm{KL}(\pi^*,\sigma^0)+\sum_{l=0}^{k-1}\left(\min_{\pi^*\in\Pi^*}\mathrm{KL}(\pi^*,\sigma^{l+1})-\min_{\pi^*\in\Pi^*}\mathrm{KL}(\pi^*,\sigma^l)\right)$$
$$\leq B+\sum_{l=0}^{k-1}\delta=B+k\delta.$$

This implies that $\min_{\pi^*\in\Pi^*}\mathrm{KL}(\pi^*,\sigma^k)<0$ for $k>\frac{B}{-\delta}$, which contradicts $\min_{\pi^*\in\Pi^*}\mathrm{KL}(\pi^*,\sigma)\geq 0$. Therefore, the sequence $\{\min_{\pi^*\in\Pi^*}\mathrm{KL}(\pi^*,\sigma^k)\}_{k\geq 0}$ converges to 0, and σ^k converges to Π^*. □

F.5. Proof of Lemma C.1

Proof of Lemma C.1. First, we introduce the following lemma:

Lemma F.8. *Let us define $T(y_i)=\arg\max_{x\in\mathcal{X}_i}\{\langle y_i,x\rangle-\psi(x)\}$. Assuming $\psi:\mathcal{X}_i\to\mathbb{R}$ be a convex function of the Legendre type, we have for any $\pi_i\in\mathcal{X}_i$:*

$$D_\psi(\pi_i,T(y_i))=\psi(\pi_i)-\psi(T(y_i))-\langle y_i,\pi_i-T(y_i)\rangle.$$

Defining $y_i^t=\eta\sum_{s=0}^t\left(\nabla_{\pi_i}v_i(\pi_i^s,\pi_{-i}^s)-\mu\nabla_{\pi_i}G(\pi_i^s,\sigma_i^k)\right)$ and letting $\pi_i=\pi_i^{\mu,\sigma^k}$, $y_i=y_i^t$ in Lemma F.8, we have:

$$D_\psi(\pi_i^{\mu,\sigma^k},\pi_i^{t+1})=\psi(\pi_i^{\mu,\sigma^k})-\psi(\pi_i^{t+1})-\langle y_i^t,\pi_i^{\mu,\sigma^k}-\pi_i^{t+1}\rangle.$$

Note that $\nabla_{\pi_i}G(\pi_i^s,\sigma_i^k)$ is well-defined because of Assumptions 5.4 and 5.5. Using this equation, we get for any $t\in\{kT_\sigma,kT_\sigma+1,\cdots,(k+1)T_\sigma-1\}$:

$$\begin{aligned}
&D_\psi(\pi_i^{\mu,\sigma^k},\pi_i^{t+1})-D_\psi(\pi_i^{\mu,\sigma^k},\pi_i^t)+D_\psi(\pi_i^{t+1},\pi_i^t)\\
&=\psi(\pi_i^{\mu,\sigma^k})-\psi(\pi_i^{t+1})-\langle y_i^t,\pi_i^{\mu,\sigma^k}-\pi_i^{t+1}\rangle-\psi(\pi_i^{\mu,\sigma^k})+\psi(\pi_i^t)+\langle y_i^{t-1},\pi_i^{\mu,\sigma^k}-\pi_i^t\rangle\\
&+\psi(\pi_i^{t+1})-\psi(\pi_i^t)-\langle y_i^{t-1},\pi_i^{t+1}-\pi_i^t\rangle\\
&=-\langle y_i^t,\pi_i^{\mu,\sigma^k}-\pi_i^{t+1}\rangle+\langle y_i^{t-1},\pi_i^{\mu,\sigma^k}-\pi_i^{t+1}\rangle\\
&=\langle y_i^t-y_i^{t-1},\pi_i^{t+1}-\pi_i^{\mu,\sigma^k}\rangle\\
&=\eta\langle\nabla_{\pi_i}v_i(\pi_i^t,\pi_{-i}^t)-\mu\nabla_{\pi_i}G(\pi_i^t,\sigma_i^k),\pi_i^{t+1}-\pi_i^{\mu,\sigma^k}\rangle.
\end{aligned}\tag{25}$$

It is easy to confirm that (25) satisfies the assumption in Lemma F.1. Thus, taking $t=(k+1)T_\sigma-1$ in Lemma F.1, we have:

$$D_\psi(\pi^{\mu,\sigma^k},\pi^{(k+1)T_\sigma})\leq D_\psi(\pi^{\mu,\sigma^k},\pi^{kT_\sigma})\left(1-\frac{\eta\mu\gamma}{2}\right)^{T_\sigma}.$$

Since $\sigma^k=\pi^{kT_\sigma}$ and $\sigma^{k+1}=\pi^{(k+1)T_\sigma}$, we conclude the statement. □

F.6. Proof of Lemma C.2

Proof of Lemma C.2. Writing $y_i^t=\sum_{s=0}^t\eta_s(\nabla_{\pi_i}v_i(\pi_i^s,\pi_{-i}^s)+\xi_i^s-\mu\nabla_{\pi_i}G(\pi_i^s,r_i))$ and using Lemma F.8 in Appendix F.5, we have:

$$\begin{aligned}
&D_\psi(\pi_i^{\mu,\sigma^k},\pi_i^{t+1})-D_\psi(\pi_i^{\mu,\sigma^k},\pi_i^t)+D_\psi(\pi_i^{t+1},\pi_i^t)\\
&=\langle y_i^t-y_i^{t-1},\pi_i^{t+1}-\pi_i^{\mu,\sigma^k}\rangle\\
&=\eta_t\langle\nabla_{\pi_i}v_i(\pi_i^t,\pi_{-i}^t)-\mu\nabla_{\pi_i}G(\pi_i^t,\sigma_i^k)+\xi_i^t,\pi_i^{t+1}-\pi_i^{\mu,\sigma^k}\rangle.
\end{aligned}$$

Note that $\nabla_{\pi_i} G(\pi_i^t, \sigma_i^k)$ is well-defined because of Assumptions 5.4 and 5.5. Thus, we can apply Lemma F.2, and we have for any $t \in \{kT_\sigma, kT_\sigma + 1, \cdots, (k+1)T_\sigma - 1\}$:

$$\begin{aligned}
&D_\psi(\pi^{\mu,r}, \pi^{t+1}) - D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + D_\psi(\pi^{t+1}, \pi^t) \\
&\leq \eta_t \left(\left(\frac{\mu^2\gamma\rho^2(\gamma + 2\beta) + 8L^2}{2\mu\gamma\rho^2} \right) D_\psi(\pi^{t+1}, \pi^t) - \frac{\mu\gamma}{2} D_\psi(\pi^{\mu,\sigma^k}, \pi^t) \right) + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle \\
&= \eta_t(\theta D_\psi(\pi^{t+1}, \pi^t) - \kappa D_\psi(\pi^{\mu,\sigma^k}, \pi^t)) + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle.
\end{aligned}$$

Therefore, the assumption in Lemma F.3 is satisfied, and we get:

$$\begin{aligned}
&\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1})] \\
&\leq \frac{2\theta - \kappa}{\kappa(t - kT_\sigma) + 2\theta} \mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{kT_\sigma})] + \frac{NC^2}{\rho(\kappa(t - kT_\sigma) + 2\theta)} \left(\frac{1}{\kappa} \ln \left(\frac{\kappa}{2\theta}(t - kT_\sigma) + 1 \right) + \frac{1}{2\theta} \right).
\end{aligned}$$

Taking $t = (k+1)T_\sigma - 1$, we have:

$$\begin{aligned}
&\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{(k+1)T_\sigma})] \\
&\leq \frac{2\theta - \kappa}{\kappa(T_\sigma - 1) + 2\theta} \mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{kT_\sigma})] + \frac{NC^2}{\rho(\kappa(T_\sigma - 1) + 2\theta)} \left(\frac{1}{\kappa} \ln \left(\frac{\kappa}{2\theta}(T_\sigma - 1) + 1 \right) + \frac{1}{2\theta} \right).
\end{aligned}$$

Since $\sigma^k = \pi^{kT_\sigma}$ and $\sigma^{k+1} = \pi^{(k+1)T_\sigma}$, we conclude the statement. □

G. Proofs for Section D

G.1. Proof of Theorem D.1

Proof of Theorem D.1. Since $v_i(\cdot, \pi_{-i}^t)$ is concave, we can upper bound the gap function for π^t as:

$$\begin{aligned}
&\mathrm{GAP}(\pi^t) \\
&= \sum_{i=1}^N \max_{\tilde{\pi}_i \in \mathcal{X}_i} \langle \nabla_{\pi_i} v_i(\pi^t), \tilde{\pi}_i - \pi_i^t \rangle \\
&= \max_{\tilde{\pi} \in \mathcal{X}} \sum_{i=1}^N \left(\langle \nabla_{\pi_i} v_i(\pi^{\mu,\sigma}), \tilde{\pi}_i - \pi_i^{\mu,\sigma} \rangle - \langle \nabla_{\pi_i} v_i(\pi^{\mu,\sigma}), \pi_i^t - \pi_i^{\mu,\sigma} \rangle + \langle \nabla_{\pi_i} v_i(\pi^t) - \nabla_{\pi_i} v_i(\pi^{\mu,\sigma}), \tilde{\pi}_i - \pi_i^t \rangle \right).
\end{aligned} \tag{26}$$

From Lemma E.1, the first term of (26) can be upper bounded as:

$$\max_{\tilde{\pi} \in \mathcal{X}} \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi^{\mu,\sigma}), \tilde{\pi}_i - \pi_i^{\mu,\sigma} \rangle \leq \mathrm{diam}(\mathcal{X}) \cdot \min_{(a_i) \in N_{\mathcal{X}}(\pi^{\mu,\sigma})} \sqrt{\sum_{i=1}^N \| - \nabla_{\pi_i} v_i(\pi^{\mu,\sigma}) + a_i \|^2}.$$

From the first-order optimality condition for $\pi^{\mu,\sigma}$, we have for any $\pi \in \mathcal{X}$:

$$\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi^{\mu,\sigma}) - \mu \nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i), \pi_i - \pi_i^{\mu,\sigma} \rangle \leq 0,$$

and then $(\nabla_{\pi_i} v_i(\pi^{\mu,\sigma}) - \mu \nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i))_{i\in[N]} \in N_{\mathcal{X}}(\pi^{\mu,\sigma})$. Thus,

$$\begin{aligned}
&\max_{\tilde{\pi}\in\mathcal{X}} \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi^{\mu,\sigma}), \tilde{\pi}_i - \pi_i^{\mu,\sigma} \rangle \\
&\leq \mathrm{diam}(\mathcal{X}) \sqrt{\sum_{i=1}^{N} \| -\nabla_{\pi_i} v_i(\pi^{\mu,\sigma}) + \nabla_{\pi_i} v_i(\pi^{\mu,\sigma}) - \mu \nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i)\|^2} \\
&= \mu \cdot \mathrm{diam}(\mathcal{X}) \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i)\|^2}. \qquad (27)
\end{aligned}$$

Next, from Cauchy–Schwarz inequality, the second term of (26) can be bounded as:

$$-\sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi^{\mu,\sigma}), \pi_i^t - \pi_i^{\mu,\sigma} \rangle \leq \|\pi^t - \pi^{\mu,\sigma}\| \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} v_i(\pi^{\mu,\sigma})\|^2}. \qquad (28)$$

Again from Cauchy–Schwarz inequality, the third term of (26) is bounded by:

$$\begin{aligned}
\sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi^t) - \nabla_{\pi_i} v_i(\pi^{\mu,\sigma}), \tilde{\pi}_i - \pi_i^t \rangle &\leq \|\tilde{\pi} - \pi^t\| \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} v_i(\pi^t) - \nabla_{\pi_i} v_i(\pi^{\mu,\sigma})\|^2} \\
&\leq \mathrm{diam}(\mathcal{X}) \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} v_i(\pi^t) - \nabla_{\pi_i} v_i(\pi^{\mu,\sigma})\|^2} \\
&\leq L \cdot \mathrm{diam}(\mathcal{X}) \|\pi^t - \pi^{\mu,\sigma}\|, \qquad (29)
\end{aligned}$$

where the third inequality follows from (2).

By combining (26), (27), (28), and (29), we get:

$$\mathrm{GAP}(\pi^t) \leq \mu \cdot \mathrm{diam}(\mathcal{X}) \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i)\|^2} + \left(L \cdot \mathrm{diam}(\mathcal{X}) + \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} v_i(\pi^{\mu,\sigma})\|^2} \right) \|\pi^t - \pi^{\mu,\sigma}\|.$$

Thus, from Lemma 5.6 and the strong convexity of ψ, we have:

$$\begin{aligned}
\mathrm{GAP}(\pi^t) \leq{}& \mu \cdot \mathrm{diam}(\mathcal{X}) \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i)\|^2} \\
&+ \left(L \cdot \mathrm{diam}(\mathcal{X}) + \sqrt{\sum_{i=1}^{N} \|\nabla_{\pi_i} v_i(\pi^{\mu,\sigma})\|^2} \right) \sqrt{\frac{2D_\psi(\pi^{\mu,\sigma}, \pi^0)}{\rho} \left(1 - \frac{\eta\mu\gamma}{2}\right)^t}.
\end{aligned}$$

□

H. Proofs for Additional Lemmas

H.1. Proof of Lemma E.2

Proof of Lemma E.2. From the first-order optimality condition for π^{μ,σ^k} and $\pi^{\mu,\sigma^{k-1}}$, we have for any $k \geq 1$:

$$\begin{aligned}
&\langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}) - \mu \left(\pi_i^{\mu,\sigma^k} - \sigma_i^k \right), \sigma_i^k - \pi_i^{\mu,\sigma^k} \rangle \leq 0, \\
&\langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^{k-1}}, \pi_{-i}^{\mu,\sigma^{k-1}}) - \mu \left(\pi_i^{\mu,\sigma^{k-1}} - \sigma_i^{k-1} \right), \pi_i^{\mu,\sigma^k} - \pi_i^{\mu,\sigma^{k-1}} \rangle \leq 0.
\end{aligned}$$

Summing up these inequalities yields:

$$
\begin{aligned}
0 &\geq \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}) - \mu \left(\pi_i^{\mu,\sigma^k} - \sigma_i^k\right), \sigma_i^k - \pi_i^{\mu,\sigma^k} \rangle \\
&+ \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^{k-1}}, \pi_{-i}^{\mu,\sigma^{k-1}}) - \mu \left(\pi_i^{\mu,\sigma^{k-1}} - \sigma_i^{k-1}\right), \pi_i^{\mu,\sigma^k} - \pi_i^{\mu,\sigma^{k-1}} \rangle \\
&= \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \pi_i^{\mu,\sigma^{k-1}} - \pi_i^{\mu,\sigma^k} \rangle + \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \sigma_i^k - \pi_i^{\mu,\sigma^{k-1}} \rangle + \mu \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \\
&+ \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^{k-1}}, \pi_{-i}^{\mu,\sigma^{k-1}}), \pi_i^{\mu,\sigma^k} - \pi_i^{\mu,\sigma^{k-1}} \rangle + \mu \sum_{i=1}^{N} \langle \sigma_i^{k-1} - \pi_i^{\mu,\sigma^{k-1}}, \pi_i^{\mu,\sigma^k} - \pi_i^{\mu,\sigma^{k-1}} \rangle \\
&= \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}) - \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^{k-1}}, \pi_{-i}^{\mu,\sigma^{k-1}}), \pi_i^{\mu,\sigma^{k-1}} - \pi_i^{\mu,\sigma^k} \rangle \\
&+ \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \sigma_i^k - \pi_i^{\mu,\sigma^{k-1}} \rangle + \mu \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 + \mu \sum_{i=1}^{N} \langle \sigma_i^{k-1} - \pi_i^{\mu,\sigma^{k-1}}, \pi_i^{\mu,\sigma^k} - \pi_i^{\mu,\sigma^{k-1}} \rangle \\
&\geq \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \sigma_i^k - \pi_i^{\mu,\sigma^{k-1}} \rangle + \mu \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 + \mu \sum_{i=1}^{N} \langle \sigma_i^{k-1} - \pi_i^{\mu,\sigma^{k-1}}, \pi_i^{\mu,\sigma^k} - \pi_i^{\mu,\sigma^{k-1}} \rangle,
\end{aligned}
$$

where the last inequality follows from (1). Then, since

$$
\langle \sigma_i^{k-1} - \pi_i^{\mu,\sigma^{k-1}}, \pi_i^{\mu,\sigma^{k-1}} - \pi_i^{\mu,\sigma^k} \rangle = \frac{\|\pi_i^{\mu,\sigma^k} - \sigma_i^{k-1}\|^2}{2} - \frac{\|\pi_i^{\mu,\sigma^k} - \pi_i^{\mu,\sigma^{k-1}}\|^2}{2} - \frac{\|\pi_i^{\mu,\sigma^{k-1}} - \sigma_i^{k-1}\|^2}{2},
$$

we have:

$$
\begin{aligned}
0 &\geq \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \sigma_i^k - \pi_i^{\mu,\sigma^{k-1}} \rangle + \mu \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \\
&+ \frac{\mu}{2} \sum_{i=1}^{N} \langle \sigma_i^{k-1} - \pi_i^{\mu,\sigma^{k-1}}, \pi_i^{\mu,\sigma^k} - \pi_i^{\mu,\sigma^{k-1}} \rangle + \frac{\mu}{2} \sum_{i=1}^{N} \langle \sigma_i^{k-1} - \pi_i^{\mu,\sigma^{k-1}}, \pi_i^{\mu,\sigma^k} - \pi_i^{\mu,\sigma^{k-1}} \rangle \\
&\geq \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \sigma_i^k - \pi_i^{\mu,\sigma^{k-1}} \rangle + \mu \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \\
&- \frac{\mu}{4} \left(\|\sigma^{k-1} - \pi^{\mu,\sigma^{k-1}}\|^2 + \|\pi^{\mu,\sigma^k} - \pi^{\mu,\sigma^{k-1}}\|^2 - \|\pi^{\mu,\sigma^k} - \pi^{\mu,\sigma^{k-1}} + \sigma^{k-1} - \pi^{\mu,\sigma^{k-1}}\|^2 \right) \\
&- \frac{\mu}{4} \left(\|\pi^{\mu,\sigma^k} - \sigma^{k-1}\|^2 - \|\pi^{\mu,\sigma^k} - \pi^{\mu,\sigma^{k-1}}\|^2 - \|\pi^{\mu,\sigma^{k-1}} - \sigma^{k-1}\|^2 \right) \\
&\geq \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \sigma_i^k - \pi_i^{\mu,\sigma^{k-1}} \rangle + \mu \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{4} \|\pi^{\mu,\sigma^k} - \sigma^{k-1}\|^2 \\
&\geq \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \sigma_i^k - \pi_i^{\mu,\sigma^{k-1}} \rangle + \mu \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2} \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2} \|\sigma^k - \sigma^{k-1}\|^2 \\
&= \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \sigma_i^k - \pi_i^{\mu,\sigma^{k-1}} \rangle + \frac{\mu}{2} \|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2} \|\sigma^k - \sigma^{k-1}\|^2,
\end{aligned}
$$

where the third inequality follows from $(a+b)^2 \leq 2(a^2+b^2)$ for $a, b \in \mathbb{R}$. Thus,

$$
\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \leq \|\sigma^k - \sigma^{k-1}\|^2 + \frac{2}{\mu} \sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \pi_i^{\mu,\sigma^{k-1}} - \sigma_i^k \rangle
$$

$$\leq \|\sigma^k - \sigma^{k-1}\|^2 + \frac{2}{\mu}\|\pi^{\mu,\sigma^{k-1}} - \sigma^k\| \sqrt{\sum_{i=1}^N \|\nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k})\|^2}$$
$$\leq \|\sigma^k - \sigma^{k-1}\|^2 + \frac{2\zeta}{\mu}\|\pi^{\mu,\sigma^{k-1}} - \sigma^k\|.$$

□

H.2. Proof of Lemma E.3

Proof of Lemma E.3. From the first-order optimality condition for p^{k+1}, we have:

$$\langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}) - \mu\left(\pi_i^{\mu,\sigma^k} - \sigma_i^k\right), \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle \leq 0.$$

Thus, from the three-point identity $2\langle x-y, z-x\rangle = \|y-z\|^2 - \|x-y\|^2 - \|x-z\|^2$ and Young's inequality:

$$\begin{aligned}
&\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle \leq \mu \sum_{i=1}^N \langle \pi_i^{\mu,\sigma^k} - \sigma_i^k, \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle \\
&= \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \pi^{\mu,\sigma^k}\|^2 \\
&= \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1} + \sigma^{k+1} - \pi^{\mu,\sigma^k}\|^2 \\
&= \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2 - \frac{\mu}{2}\|\sigma^{k+1} - \pi^{\mu,\sigma^k}\|^2 - \mu\langle \pi^* - \sigma^{k+1}, \sigma^{k+1} - \pi^{\mu,\sigma^k}\rangle \\
&\leq \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2 + \mu\|\pi^* - \sigma^{k+1}\|\|\sigma^{k+1} - \pi^{\mu,\sigma^k}\| \\
&\leq \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2 + \frac{\mu}{64T^3}\|\pi^* - \sigma^{k+1}\|^2 + \frac{32\mu T^3}{2}\|\sigma^{k+1} - \pi^{\mu,\sigma^k}\|^2.
\end{aligned}$$

Here, since $T_\sigma \geq \max(\frac{6}{\ln 2 - \ln(2-\eta\mu)}\ln T + \frac{2\ln 64}{\ln 2 - \ln(2-\eta\mu)}, 1)$, we have:

$$\left(1 - \frac{\eta\mu}{2}\right)^{-T_\sigma} \geq \left(1 - \frac{\eta\mu}{2}\right)^{-\frac{3}{\ln 2 - \ln(2-\eta\mu)}\ln T} \left(1 - \frac{\eta\mu}{2}\right)^{\frac{\ln 64}{\ln\left(1 - \frac{\eta\mu}{2}\right)}} = 64\left(1 - \frac{\eta\mu}{2}\right)^{\frac{\ln T^3}{\ln(2-\eta\mu) - \ln 2}} = 64T^3.$$

Therefore, we get from Lemma B.2:

$$\begin{aligned}
&\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle \\
&\leq \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2 + \frac{\mu}{64T^3}\|\pi^* - \sigma^{k+1}\|^2 + \frac{32\mu T^3}{2}\|\sigma^k - \pi^{\mu,\sigma^k}\|^2\left(1 - \frac{\eta\mu}{2}\right)^{T_\sigma} \\
&\leq \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2 + \frac{\mu}{64T^3}\|\pi^* - \sigma^{k+1}\|^2 + \frac{\mu}{4}\|\sigma^k - \pi^{\mu,\sigma^k}\|^2 \\
&= \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2 - \frac{\mu}{4}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 + \frac{\mu}{64T^3}\|\pi^* - \sigma^{k+1}\|^2.
\end{aligned}$$

Summing up this inequality from $k = 0$ to $K - 1$ yields:

$$\begin{aligned}
&\frac{\mu}{2}\|\pi^* - \sigma^0\|^2 + \frac{\mu}{64T^3}\sum_{k=0}^{K-1}\|\pi^* - \sigma^{k+1}\|^2 - \frac{\mu}{4}\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 \\
&\geq \sum_{k=0}^{K-1}\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle \\
&\geq \sum_{k=0}^{K-1}\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^*, \pi_{-i}^*), \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle \\
&\geq 0.
\end{aligned}$$

Then, from Cauchy–Schwarz inequality, we have:

$$\begin{aligned}
\frac{\mu}{4}\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k}-\sigma^k\|^2 &\le \frac{\mu}{2}\|\pi^*-\sigma^0\|^2+\frac{\mu}{64T^3}\sum_{k=0}^{K-1}\|\pi^*-\sigma^{k+1}\|^2\\
&\le \frac{\mu}{2}\|\pi^*-\sigma^0\|^2+\frac{\mu}{64T^3}\sum_{k=0}^{K-1}\left(\sum_{\tau=0}^{k}\|\sigma^{\tau+1}-\sigma^\tau\|+\|\pi^*-\sigma^0\|\right)^2\\
&\le \frac{\mu}{2}\|\pi^*-\sigma^0\|^2+\frac{\mu}{32T^3}\sum_{k=0}^{K-1}\left(\left(\sum_{\tau=0}^{k}\|\sigma^{\tau+1}-\sigma^\tau\|\right)^2+\|\pi^*-\sigma^0\|^2\right)\\
&\le \frac{\mu}{2}\|\pi^*-\sigma^0\|^2+\frac{\mu}{32T^3}\sum_{k=0}^{K-1}\left(K\sum_{\tau=0}^{k}\|\sigma^{\tau+1}-\sigma^\tau\|^2+\|\pi^*-\sigma^0\|^2\right)\\
&\le \frac{\mu}{2}\|\pi^*-\sigma^0\|^2+\frac{\mu}{32T^3}\left(K^2\sum_{k=0}^{K-1}\|\sigma^{k+1}-\sigma^k\|^2+K\|\pi^*-\sigma^0\|^2\right)\\
&\le \frac{\mu}{2}\|\pi^*-\sigma^0\|^2+\frac{\mu}{32T^3}\left(K^2\sum_{k=0}^{K-1}\left(\|\sigma^{k+1}-\pi^{\mu,\sigma^k}\|+\|\pi^{\mu,\sigma^k}-\sigma^k\|\right)^2+K\|\pi^*-\sigma^0\|^2\right).
\end{aligned}$$

By applying Lemma B.2 to the above inequality, we get:

$$\begin{aligned}
&\frac{\mu}{4}\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k}-\sigma^k\|^2\\
&\le \frac{\mu}{2}\|\pi^*-\sigma^0\|^2+\frac{\mu}{32T^3}\left(K^2\sum_{k=0}^{K-1}\left(\|\sigma^k-\pi^{\mu,\sigma^k}\|^2\left(1-\frac{\eta\mu}{2}\right)^{\frac{T_\sigma}{2}}+\|\pi^{\mu,\sigma^k}-\sigma^k\|\right)^2+K\|\pi^*-\sigma^0\|^2\right)\\
&\le \frac{\mu}{2}\|\pi^*-\sigma^0\|^2+\frac{\mu}{32T^3}\left(K^2\sum_{k=0}^{K-1}\left(2\|\pi^{\mu,\sigma^k}-\sigma^k\|\right)^2+K\|\pi^*-\sigma^0\|^2\right)\\
&= \frac{\mu}{2}\|\pi^*-\sigma^0\|^2+\frac{\mu}{32T^3}\left(4K^2\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k}-\sigma^k\|^2+K\|\pi^*-\sigma^0\|^2\right)\\
&\le \frac{\mu}{2}\|\pi^*-\sigma^0\|^2+\frac{\mu}{32T^3}\left(4T^2\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k}-\sigma^k\|^2+T\|\pi^*-\sigma^0\|^2\right)\\
&= \mu\left(\frac{1}{2}+\frac{1}{T^2}\right)\|\pi^*-\sigma^0\|^2+\frac{\mu}{8T}\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k}-\sigma^k\|^2\\
&\le 2\mu\|\pi^*-\sigma^0\|^2+\frac{\mu}{8}\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k}-\sigma^k\|^2.
\end{aligned}$$

Therefore, for $K \ge 1$, we get:

$$\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k}-\sigma^k\|^2 \le 16\|\pi^*-\sigma^0\|^2.$$

□

H.3. Proof of Lemma E.4

Proof of Lemma E.4. From the first-order optimality condition for p^{k+1}, we have:

$$\langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k},\pi_{-i}^{\mu,\sigma^k})-\mu\left(\pi_i^{\mu,\sigma^k}-\sigma_i^k\right),\pi_i^*-\pi_i^{\mu,\sigma^k}\rangle \le 0.$$

Thus, from the three-point identity $2\langle a-b, c-a\rangle = \|b-c\|^2 - \|a-b\|^2 - \|a-c\|^2$ and Young's inequality:

$$\begin{aligned}
&\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle \leq \mu \sum_{i=1}^N \langle \pi_i^{\mu,\sigma^k} - \sigma_i^k, \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle \\
&= \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \pi^{\mu,\sigma^k}\|^2 \\
&= \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1} + \sigma^{k+1} - \pi^{\mu,\sigma^k}\|^2 \\
&= \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2 - \frac{\mu}{2}\|\sigma^{k+1} - \pi^{\mu,\sigma^k}\|^2 - \mu\langle \pi^* - \sigma^{k+1}, \sigma^{k+1} - \pi^{\mu,\sigma^k}\rangle \\
&\leq \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2 + \mu\|\pi^* - \sigma^{k+1}\|\|\sigma^{k+1} - \pi^{\mu,\sigma^k}\| \\
&\leq \frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2 + \mu \cdot \mathrm{diam}(\mathcal{X})\|\sigma^{k+1} - \pi^{\mu,\sigma^k}\|
\end{aligned}$$

Since $T_\sigma \geq \max(T^{4/5}+2, 3)$, we get from Lemma B.7:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle\right] \\
&\leq \mathbb{E}\left[\frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2\right] \\
&+ \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(T_\sigma - 1) + 1\right) + \frac{1}{2\theta}\right)}{\rho(\kappa(T_\sigma - 1) + 2\theta)}} \\
&\leq \mathbb{E}\left[\frac{\mu}{2}\|\pi^* - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2 - \frac{\mu}{2}\|\pi^* - \sigma^{k+1}\|^2\right] \\
&+ \mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T + 1\right) + \frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5} + 2\theta)}}.
\end{aligned}$$

Summing up this inequality from $k=0$ to $K-1$ and taking its expectation yields:

$$\begin{aligned}
&\frac{\mu}{2}\|\pi^* - \sigma^0\|^2 - \mathbb{E}\left[\frac{\mu}{2}\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2\right] \\
&+ K\mu \cdot \mathrm{diam}(\mathcal{X}) \cdot \sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T + 1\right) + \frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5} + 2\theta)}} \\
&\geq \mathbb{E}\left[\sum_{k=0}^{K-1}\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}), \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle\right] \\
&\geq \mathbb{E}\left[\sum_{k=0}^{K-1}\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^*, \pi_{-i}^*), \pi_i^* - \pi_i^{\mu,\sigma^k}\rangle\right] \\
&\geq 0.
\end{aligned}$$

Therefore, for $K \geq 1$, we get:

$$\mathbb{E}\left[\sum_{k=0}^{K-1}\|\pi^{\mu,\sigma^k} - \sigma^k\|^2\right] \leq \|\pi^* - \sigma^0\|^2 + K \cdot \mathrm{diam}(\mathcal{X}) \cdot \sqrt{\frac{\rho(2\theta-\kappa)\mathrm{diam}(\mathcal{X})^2 + NC^2\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}T + 1\right) + \frac{1}{2\theta}\right)}{\rho(\kappa T^{4/5} + 2\theta)}}.$$

□

H.4. Proof of Lemma F.1

Proof of Lemma F.1. We first decompose the inequality in the assumption as follows:

$$\begin{aligned}
&D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^{t+1}) - D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t) + D_\psi(\pi_i^{t+1}, \pi_i^t) \\
&\leq \eta\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu\nabla_{\pi_i} G(\pi_i^t, \sigma_i^k), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle \\
&= \eta\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle + \eta\mu\langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i^k), \pi_i^t - \pi_i^{t+1}\rangle + \eta\mu\langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i^k), \pi_i^{\mu,\sigma^k} - \pi_i^t\rangle. \qquad (30)
\end{aligned}$$

From the relative smoothness in Assumption 5.4 and the convexity of $G(\cdot, \sigma_i^k)$:

$$\begin{aligned}
&\langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i^k), \pi_i^t - \pi_i^{t+1}\rangle \\
&\leq G(\pi_i^t, \sigma_i^k) - G(\pi_i^{t+1}, \sigma_i^k) + \beta D_\psi(\pi_i^{t+1}, \pi_i^t) \\
&\leq G(\pi_i^t, \sigma_i^k) - G(\pi_i^{\mu,\sigma^k}, \sigma_i^k) + \langle \nabla_{\pi_i} G(\pi_i^{\mu,\sigma^k}, \sigma_i^k), \pi_i^{\mu,\sigma^k} - \pi_i^{t+1}\rangle + \beta D_\psi(\pi_i^{t+1}, \pi_i^t). \qquad (31)
\end{aligned}$$

Also, from the relative strong convexity in Assumption 5.4:

$$G(\pi_i^t, \sigma_i^k) - G(\pi_i^{\mu,\sigma^k}, \sigma_i^k) \leq \langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i^k), \pi_i^t - \pi_i^{\mu,\sigma^k}\rangle - \gamma D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t). \qquad (32)$$

By combining (30), (31), and (32), we have:

$$\begin{aligned}
&D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^{t+1}) - D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t) + D_\psi(\pi_i^{t+1}, \pi_i^t) \\
&\leq \eta\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle + \eta\mu\langle \nabla_{\pi_i} G(\pi_i^{\mu,\sigma^k}, \sigma_i^k), \pi_i^{\mu,\sigma^k} - \pi_i^{t+1}\rangle \\
&- \eta\mu\gamma D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t) + \eta\mu\beta D_\psi(\pi_i^{t+1}, \pi_i^t),
\end{aligned}$$

and then:

$$\begin{aligned}
&D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^{t+1}) - (1-\eta\mu\gamma) D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t) + (1-\eta\mu\beta) D_\psi(\pi_i^{t+1}, \pi_i^t) \\
&\leq \eta\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle + \eta\mu\langle \nabla_{\pi_i} G(\pi_i^{\mu,\sigma^k}, \sigma_i^k), \pi_i^{\mu,\sigma^k} - \pi_i^{t+1}\rangle \\
&= \eta\langle \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle + \eta\mu\langle \nabla_{\pi_i} G(\pi_i^{\mu,\sigma^k}, \sigma_i^k), \pi_i^{\mu,\sigma^k} - \pi_i^{t+1}\rangle \\
&+ \eta\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle.
\end{aligned}$$

Summing this inequality from $i = 1$ to N implies that:

$$\begin{aligned}
&D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) - (1-\eta\mu\gamma) D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + (1-\eta\mu\beta) D_\psi(\pi^{t+1}, \pi^t) \\
&\leq \eta \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle + \eta\mu \sum_{i=1}^N \langle \nabla_{\pi_i} G(\pi_i^{\mu,\sigma^k}, \sigma_i^k), \pi_i^{\mu,\sigma^k} - \pi_i^{t+1}\rangle \\
&+ \eta \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle \\
&\leq \eta \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}) - \mu\nabla_{\pi_i} G(\pi_i^{\mu,\sigma^k}, \sigma_i^k), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle \\
&+ \eta \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle \\
&\leq \eta \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle, \qquad (33)
\end{aligned}$$

where the second inequality follows from (1), and the third inequality follows from the first-order optimality condition for π^{μ,σ^k}.

Here, from Young's inequality, we have for any $\lambda > 0$:

$$\begin{aligned}
&\sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle \\
&= \sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{t}\rangle + \sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t} - \pi_i^{\mu,\sigma^k}\rangle \\
&\le \lambda \sum_{i=1}^{N}\|\nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}) - \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t)\|^2 + \frac{1}{2\lambda}\|\pi^{t+1} - \pi^t\|^2 + \frac{1}{2\lambda}\|\pi^t - \pi^{\mu,\sigma^k}\|^2 \\
&\le \left(L^2\lambda + \frac{1}{2\lambda}\right)\|\pi^{t+1} - \pi^t\|^2 + \frac{1}{2\lambda}\|\pi^t - \pi^{\mu,\sigma^k}\|^2 \\
&\le \frac{1}{\rho}\left(2L^2\lambda + \frac{1}{\lambda}\right) D_\psi(\pi^{t+1}, \pi^t) + \frac{1}{\rho\lambda} D_\psi(\pi^{\mu,\sigma^k}, \pi^t). \qquad (34)
\end{aligned}$$

where the second inequality follows from (2), and the fourth inequality follows from the strong convexity of ψ.

By combining (33) and (34), we get:

$$D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) \le \left(1 - \eta\left(\mu\gamma - \frac{1}{\rho\lambda}\right)\right) D_\psi(\pi^{\mu,\sigma^k}, \pi^t) - \left(1 - \eta\left(\mu\beta + \frac{2L^2\lambda}{\rho} + \frac{1}{\rho\lambda}\right)\right) D_\psi(\pi^{t+1}, \pi^t).$$

By setting $\lambda = \frac{2}{\mu\gamma\rho}$,

$$\begin{aligned}
D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) &\le \left(1 - \frac{\eta\mu\gamma}{2}\right) D_\psi(\pi^{\mu,\sigma^k}, \pi^t) - \left(1 - \eta\left(\frac{\mu(\gamma + 2\beta)}{2} + \frac{4L^2}{\mu\gamma\rho^2}\right)\right) D_\psi(\pi^{t+1}, \pi^t) \\
&= \left(1 - \frac{\eta\mu\gamma}{2}\right) D_\psi(\pi^{\mu,\sigma^k}, \pi^t) - \left(1 - \eta\left(\frac{\mu^2\gamma\rho^2(\gamma + 2\beta) + 8L^2}{2\mu\gamma\rho^2}\right)\right) D_\psi(\pi^{t+1}, \pi^t).
\end{aligned}$$

Thus, when $\eta \le \frac{2\mu\gamma\rho^2}{\mu^2\gamma\rho^2(\gamma+2\beta)+8L^2} < \frac{2}{\mu\gamma}$, we have for any $t \in \{kT_\sigma, kT_\sigma + 1, \cdots, (k+1)T_\sigma - 1\}$:

$$D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) \le D_\psi(\pi^{\mu,\sigma^k}, \pi^t)\left(1 - \frac{\eta\mu\gamma}{2}\right) \le D_\psi(\pi^{\mu,\sigma^k}, \pi^{kT_\sigma})\left(1 - \frac{\eta\mu\gamma}{2}\right)^{t - kT_\sigma + 1}.$$

□

H.5. Proof of Lemma F.2

Proof of Lemma F.2. We first decompose the inequality in the assumption as follows:

$$\begin{aligned}
&D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^{t+1}) - D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t) + D_\psi(\pi_i^{t+1}, \pi_i^t) \\
&\le \eta_t\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu\nabla_{\pi_i} G(\pi_i^t, \sigma_i^k) + \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle \\
&= \eta_t\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle + \eta_t\mu\langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i^k), \pi_i^t - \pi_i^{t+1}\rangle \\
&\quad + \eta_t\mu\langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i^k), \pi_i^{\mu,\sigma^k} - \pi_i^t\rangle + \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle. \qquad (35)
\end{aligned}$$

By combining (31), (32) in Appendix H.4, and (35),

$$\begin{aligned}
&D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^{t+1}) - D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t) + D_\psi(\pi_i^{t+1}, \pi_i^t) \\
&\le \eta_t\langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle + \eta_t\mu\langle \nabla_{\pi_i} G(\pi_i^{\mu,\sigma^k}, \sigma_i^k), \pi_i^{\mu,\sigma^k} - \pi_i^{t+1}\rangle + \eta_t\mu\beta D_\psi(\pi_i^{t+1}, \pi_i^t) \\
&\quad - \eta_t\mu\gamma D_\psi(\pi_i^{\mu,\sigma^k}, \pi_i^t) + \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k}\rangle.
\end{aligned}$$

Summing up these inequalities with respect to the player index,

$$
\begin{aligned}
&D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) - D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + D_\psi(\pi^{t+1}, \pi^t) \\
&\leq \eta_t \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu G(\pi_i^{\mu,\sigma^k}, \sigma_i^k), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle + \eta_t \mu \beta D_\psi(\pi^{t+1}, \pi^t) \\
&\quad - \eta_t \mu \gamma D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle \\
&= \sum_{i=1}^N \eta_t \langle \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}) - \mu \nabla_{\pi_i} G(\pi_i^{\mu,\sigma^k}, \sigma_i^k), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle - \eta_t \mu \gamma D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + \eta_t \mu \beta D_\psi(\pi^{t+1}, \pi^t) \\
&+ \eta_t \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle \\
&\leq \sum_{i=1}^N \eta_t \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma^k}, \pi_{-i}^{\mu,\sigma^k}) - \mu \nabla_{\pi_i} G(\pi_i^{\mu,\sigma^k}, \sigma_i^k), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle - \eta_t \mu \gamma D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + \eta_t \mu \beta D_\psi(\pi^{t+1}, \pi^t) \\
&+ \eta_t \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle \\
&\leq -\eta_t \mu \gamma D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + \eta_t \mu \beta D_\psi(\pi^{t+1}, \pi^t) \\
&+ \eta_t \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \nabla_{\pi_i} v_i(\pi_i^{t+1}, \pi_{-i}^{t+1}), \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle, \qquad (36)
\end{aligned}
$$

where the second inequality follows (1), and the third inequality follows from the first-order optimality condition for π^{μ,σ^k}. By combining (34) in Appendix H.4 and (36), we have for any $\lambda > 0$:

$$
\begin{aligned}
&D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) - D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + D_\psi(\pi^{t+1}, \pi^t) \\
&\leq -\eta_t \mu \gamma D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + \eta_t \mu \beta D_\psi(\pi^{t+1}, \pi^t) \\
&+ \frac{\eta_t}{\rho} \left(2L^2 \lambda + \frac{1}{\lambda} \right) D_\psi(\pi^{t+1}, \pi^t) + \frac{\eta_t}{\rho \lambda} D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle.
\end{aligned}
$$

By setting $\lambda = \frac{2}{\mu \gamma \rho}$,

$$
\begin{aligned}
&D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) - D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + D_\psi(\pi^{t+1}, \pi^t) \\
&\leq -\frac{\eta_t \mu \gamma}{2} D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + \eta_t \left(\frac{\mu^2 \gamma \rho^2 (\gamma + 2\beta) + 8L^2}{2\mu \gamma \rho^2} \right) D_\psi(\pi^{t+1}, \pi^t) + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle.
\end{aligned}
$$

This concludes the proof. □

H.6. Proof of Lemma F.3

Proof of Lemma F.3. Reforming the inequality in the assumption,

$$
\begin{aligned}
D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1}) &\leq (1 - \kappa \eta_t) D_\psi(\pi^{\mu,\sigma^k}, \pi^t) - (1 - \eta_t \theta) D_\psi(\pi^{t+1}, \pi^t) + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^{\mu,\sigma^k} \rangle \\
&= (1 - \kappa \eta_t) D_\psi(\pi^{\mu,\sigma^k}, \pi^t) - (1 - \eta_t \theta) D_\psi(\pi^{t+1}, \pi^t) \\
&\quad + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^t - \pi_i^{\mu,\sigma^k} \rangle + \eta_t \sum_{i=1}^N \langle \xi_i^t, \pi_i^{t+1} - \pi_i^t \rangle.
\end{aligned}
$$

By taking the expectation conditioned on $\mathcal{F}_t$ for both sides and using Assumption B.5 (a),

$$\begin{aligned}
\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1})|\mathcal{F}_t] &\leq (1-\kappa\eta_t)D_\psi(\pi^{\mu,\sigma^k}, \pi^t) - (1-\eta_t\theta)\mathbb{E}[D_\psi(\pi^{t+1}, \pi^t)|\mathcal{F}_t] + \sum_{i=1}^N \mathbb{E}[\langle \eta_t \xi_i^t, \pi_i^{t+1} - \pi_i^t\rangle|\mathcal{F}_t] \\
&= (1-\kappa\eta_t)D_\psi(\pi^{\mu,\sigma^k}, \pi^t) - (1-\eta_t\theta)\mathbb{E}[D_\psi(\pi^{t+1}, \pi^t)|\mathcal{F}_t] \\
&\quad + \sum_{i=1}^N \mathbb{E}\left[\left\langle \frac{\eta_t \xi_i^t}{\sqrt{\rho(1-\eta_t\theta)}}, \sqrt{\rho(1-\eta_t\theta)}(\pi_i^{t+1} - \pi_i^t)\right\rangle |\mathcal{F}_t\right] \\
&\leq (1-\kappa\eta_t)D_\psi(\pi^{\mu,\sigma^k}, \pi^t) - (1-\eta_t\theta)\mathbb{E}[D_\psi(\pi^{t+1}, \pi^t)|\mathcal{F}_t] \\
&\quad + \frac{\eta_t^2}{2\rho(1-\eta_t\theta)}\sum_{i=1}^N \mathbb{E}[\|\xi_i^t\|^2|\mathcal{F}_t] + \frac{\rho(1-\eta_t\theta)}{2}\mathbb{E}[\|\pi^{t+1} - \pi^t\|^2|\mathcal{F}_t] \\
&\leq (1-\kappa\eta_t)D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + \frac{\eta_t^2}{2\rho(1-\eta_t\theta)}\sum_{i=1}^N \mathbb{E}[\|\xi_i^t\|^2|\mathcal{F}_t] \\
&\leq \left(1 - \frac{1}{t - kT_\sigma + 2\theta/\kappa}\right) D_\psi(\pi^{\mu,\sigma^k}, \pi^t) + \frac{\eta_t^2}{\rho}\sum_{i=1}^N \mathbb{E}[\|\xi_i^t\|^2|\mathcal{F}_t].
\end{aligned}$$

Therefore, rearranging and taking the expectations,

$$(t - kT_\sigma + 2\theta/\kappa)\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1})] \leq (t - kT_\sigma - 1 + 2\theta/\kappa)\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^t)] + \frac{NC^2}{\rho\kappa(\kappa(t-kT_\sigma)+2\theta)}.$$

Telescoping the sum,

$$\begin{aligned}
&(t - kT_\sigma + 2\theta/\kappa)\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1})] \leq (2\theta/\kappa - 1)\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{kT_\sigma})] + \frac{NC^2}{\rho\kappa}\sum_{s=kT_\sigma}^{t} \frac{1}{\kappa(s-kT_\sigma)+2\theta} \\
\iff \quad &\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1})] \leq \frac{2\theta-\kappa}{\kappa(t-kT_\sigma)+2\theta}\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{kT_\sigma})] + \frac{NC^2}{\rho(\kappa(t-kT_\sigma)+2\theta)}\sum_{s=0}^{t-kT_\sigma}\frac{1}{\kappa s + 2\theta}.
\end{aligned}$$

Here, we introduce the following lemma, whose proof is given in Appendix H.12, for the evaluation of the sum.

Lemma H.1. *For any $\kappa, \theta \geq 0$ and $t \geq 0$,*

$$\sum_{s=0}^{t}\frac{1}{\kappa s + 2\theta} \leq \frac{1}{2\theta} + \frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}t + 1\right).$$

In summary, we obtain the following inequality:

$$\begin{aligned}
&\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{t+1})] \\
&\leq \frac{2\theta-\kappa}{\kappa(t-kT_\sigma)+2\theta}\mathbb{E}[D_\psi(\pi^{\mu,\sigma^k}, \pi^{kT_\sigma})] + \frac{NC^2}{\rho(\kappa(t-kT_\sigma)+2\theta)}\left(\frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}(t-kT_\sigma)+1\right) + \frac{1}{2\theta}\right).
\end{aligned}$$

This concludes the proof. □

H.7. Proof of Lemma F.4

Proof of Lemma F.4. Recall that $G(\pi_i, \pi_i') = D_{\psi'}(\pi_i, \pi_i')$ for any $i \in [N]$ and $\pi_i, \pi_i' \in \mathcal{X}_i$. By the first-order optimality condition for σ^{k+1}, we have for all $\pi^* \in \Pi^*$:

$$\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1}) - \mu(\nabla\psi'(\sigma_i^{k+1}) - \nabla\psi'(\sigma_i^k)), \pi_i^* - \sigma_i^{k+1}\rangle \leq 0.$$

Then,

$$\sum_{i=1}^{N}\langle \nabla\psi'(\sigma_i^{k+1}) - \nabla\psi'(\sigma_i^k), \sigma_i^{k+1} - \pi_i^*\rangle \leq \frac{1}{\mu}\sum_{i=1}^{N}\langle \sigma_i^{k+1} - \pi_i^*, \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1})\rangle.$$

Moreover, we have for any $\pi^* \in \Pi^*$:

$$\begin{aligned}
&D_{\psi'}(\pi^*, \sigma^{k+1}) - D_{\psi'}(\pi^*, \sigma^k) \\
&= \sum_{i=1}^{N}\left(\psi'(\pi_i^*) - \psi'(\sigma_i^{k+1}) - \langle \nabla\psi'(\sigma_i^{k+1}), \pi_i^* - \sigma_i^{k+1}\rangle - \psi'(\pi_i^*) + \psi'(\sigma_i^k) + \langle \nabla\psi'(\sigma_i^k), \pi_i^* - \sigma_i^k\rangle\right) \\
&= \sum_{i=1}^{N}\left(-\psi'(\sigma_i^{k+1}) + \psi'(\sigma_i^k) + \langle \nabla\psi'(\sigma_i^k), \sigma_i^{k+1} - \sigma_i^k\rangle - \langle \nabla\psi'(\sigma_i^{k+1}) - \nabla\psi'(\sigma_i^k), \pi_i^* - \sigma_i^{k+1}\rangle\right) \\
&= -D_{\psi'}(\sigma^{k+1}, \sigma^k) + \sum_{i=1}^{N}\langle \nabla\psi'(\sigma_i^{k+1}) - \nabla\psi'(\sigma_i^k), \sigma_i^{k+1} - \pi_i^*\rangle.
\end{aligned}$$

By combining these inequalities, we get for any $\pi^* \in \Pi^*$:

$$\begin{aligned}
D_{\psi'}(\pi^*, \sigma^{k+1}) - D_{\psi'}(\pi^*, \sigma^k) &\leq -D_{\psi'}(\sigma^{k+1}, \sigma^k) + \frac{1}{\mu}\sum_{i=1}^{N}\langle \sigma_i^{k+1} - \pi_i^*, \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1})\rangle \\
&\leq -D_{\psi'}(\sigma^{k+1}, \sigma^k) + \frac{1}{\mu}\sum_{i=1}^{N}\langle \sigma_i^{k+1} - \pi_i^*, \nabla_{\pi_i} v_i(\pi_i^*, \pi_{-i}^*)\rangle,
\end{aligned}$$

where the second inequality follows from (1). Since π^* is the Nash equilibrium, from the first-order optimality condition, we get:

$$\sum_{i=1}^{N}\langle \sigma_i^{k+1} - \pi_i^*, \nabla_{\pi_i} v_i(\pi_i^*, \pi_{-i}^*)\rangle \leq 0.$$

Thus, we have for $\pi^* \in \Pi^*$:

$$D_{\psi'}(\pi^*, \sigma^{k+1}) - D_{\psi'}(\pi^*, \sigma^k) \leq -D_{\psi'}(\sigma^{k+1}, \sigma^k).$$

□

H.8. Proof of Lemma F.5

Proof of Lemma F.5. By using the first-order optimality condition for σ_i^{k+1}, we have for all $\pi \in \mathcal{X}$:

$$\sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1}) - \mu(\nabla_{\pi_i}\psi'(\sigma_i^{k+1}) - \nabla_{\pi_i}\psi'(\sigma_i^k)), \pi_i - \sigma_i^{k+1}\rangle \leq 0,$$

and then

$$\sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1}), \pi_i - \sigma_i^{k+1}\rangle \leq \mu\sum_{i=1}^{N}\langle \nabla_{\pi_i}\psi'(\sigma_i^{k+1}) - \nabla_{\pi_i}\psi'(\sigma_i^k), \pi_i - \sigma_i^{k+1}\rangle.$$

Under the assumption that $\sigma^{k+1} = \sigma^k$, we have for all $\pi \in \mathcal{X}$:

$$\sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1}), \pi_i - \sigma_i^{k+1}\rangle \leq 0.$$

This is equivalent to the first-order optimality condition for $\pi^* \in \Pi^*$. Therefore, $\sigma^{k+1} = \sigma^k$ is a Nash equilibrium of the underlying game. □

H.9. Proof of Lemma F.6

Proof of Lemma F.6. First, we prove the first statement of the lemma by using the following lemmas:

Lemma H.2. *Assume that $\sigma^{k+1} = \pi^{\mu,\sigma^k}$ for $k \geq 0$, and G is one of the following divergence: 1) α-divergence with $\alpha \in (0,1)$; 2) Rényi-divergence with $\alpha \in (0,1)$; 3) reverse KL divergence. If $\sigma^{k+1} = \sigma^k$, then σ^k is a Nash equilibrium of the underlying game.*

Lemma H.3. *Assume that $\sigma^{k+1} = \pi^{\mu,\sigma^k}$ for $k \geq 0$, and G is one of the following divergence: 1) α-divergence with $\alpha \in (0,1)$; 2) Rényi-divergence with $\alpha \in (0,1)$; 3) reverse KL divergence. Then, if $\sigma^{k+1} \neq \sigma^k$, we have for any $\pi^* \in \Pi^*$ and $k \geq 0$:*

$$\mathrm{KL}(\pi^*, \sigma^{k+1}) - \mathrm{KL}(\pi^*, \sigma^k) < 0.$$

From Lemma H.2, when $\sigma^k \in \mathcal{X} \setminus \Pi^*$, $\sigma^{k+1} \neq \sigma^k$ always holds. Let us define $\pi^\star = \arg\min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^k)$. Since $\sigma^{k+1} \neq \sigma^k$, from Lemma H.3, we have:

$$\min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^k) = \mathrm{KL}(\pi^\star, \sigma^k) > \mathrm{KL}(\pi^\star, \sigma^{k+1}) \geq \min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^{k+1}).$$

Therefore, if $\sigma^k \in \mathcal{X} \setminus \Pi^*$ then $\min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^{k+1}) < \min_{\pi^* \in \Pi^*} \mathrm{KL}(\pi^*, \sigma^k)$.

Next, we prove the second statement of the lemma. Assume that there exists $\sigma^k \in \Pi^*$ such that $\sigma^{k+1} \neq \sigma^k$. In this case, we can apply Lemma H.3, hence we have $\mathrm{KL}(\pi^*, \sigma^{k+1}) < \mathrm{KL}(\pi^*, \sigma^k)$ for all $\pi^* \in \Pi^*$. On the other hand, since $\sigma^k \in \Pi^*$, there exists a Nash equilibrium $\pi^\star \in \Pi^*$ such that $\mathrm{KL}(\pi^\star, \sigma^k) = 0$. Therefore, we have $\mathrm{KL}(\pi^\star, \sigma^{k+1}) < \mathrm{KL}(\pi^\star, \sigma^k) = 0$, which contradicts $\mathrm{KL}(\pi^\star, \sigma^{k+1}) \geq 0$. Thus, if $\sigma^k \in \Pi^*$ then $\sigma^{k+1} = \sigma^k$. □

H.10. Proof of Lemma F.7

Proof of Lemma F.7. For a given $\sigma \in \mathcal{X}$, let us consider that π^t follows the following continuous-time dynamics:

$$\begin{aligned} \pi_i^t &= \arg\max_{\pi_i \in \mathcal{X}_i} \left\{ \langle y_i^t, \pi_i \rangle - \psi(\pi_i) \right\}, \\ y_{ij}^t &= \int_0^t \left(\frac{\partial}{\pi_{ij}} v_i(\pi_i^s, \pi_{-i}^s) - \mu \frac{\partial}{\pi_{ij}} G(\pi_i^s, \sigma_i) \right). \end{aligned} \tag{37}$$

We assume that $\psi(\pi_i) = \sum_{j=1}^{d_i} \pi_{ij} \ln \pi_{ij}$. Note that this dynamics is the continuous-time version of APFTRL, so clearly $\pi^{\mu,\sigma}$ defined by (4) is the stationary point of (37). We have for a given $\sigma' \in \mathcal{X}$ and the associated stationary point $\pi^{\mu,\sigma'} = F(\sigma')$:

$$\begin{aligned} \frac{d}{dt}\mathrm{KL}(\pi^{\mu,\sigma'}, \pi^t) &= \frac{d}{dt} D_\psi(\pi^{\mu,\sigma'}, \pi^t) \\ &= \sum_{i=1}^N \frac{d}{dt} \left(\psi(\pi_i^{\mu,\sigma'}) - \psi(\pi_i^t) - \langle y_i^t, \pi_i^{\mu,\sigma'} - \pi_i^t \rangle \right) \\ &= \sum_{i=1}^N \frac{d}{dt} \left(\langle y_i^t, \pi_i^t \rangle - \psi(\pi_i^t) - \langle y_i^t, \pi_i^{\mu,\sigma'} \rangle + \psi(\pi_i^{\mu,\sigma'}) \right) \\ &= \sum_{i=1}^N \frac{d}{dt} \left(\psi^*(y_i^t) - \langle y_i^t, \pi_i^{\mu,\sigma'} \rangle \right) \\ &= \sum_{i=1}^N \left(\left\langle \frac{d}{dt} y_i^t, \nabla \psi^*(y_i^t) \right\rangle - \left\langle \frac{d}{dt} y_i^t, \pi_i^{\mu,\sigma'} \right\rangle \right) \\ &= \sum_{i=1}^N \left\langle \frac{d}{dt} y_i^t, \nabla \psi^*(y_i^t) - \pi_i^{\mu,\sigma'} \right\rangle, \end{aligned}$$

where $\psi^*(y_i^t) = \max_{\pi_i \in \mathcal{X}_i} \{\langle y_i^t, \pi_i \rangle - \psi(\pi_i)\}$. When $\psi(\pi_i) = \sum_{j=1}^{d_i} \pi_{ij} \ln \pi_{ij}$, we have

$$\begin{aligned}\psi^*(y_i^t) &= \sum_{j=1}^{d_i} y_{ij}^t \frac{\exp(y_{ij}^t)}{\sum_{j'=1}^{d_i} \exp(y_{ij'}^t)} - \sum_{j=1}^{d_i} \frac{\exp(y_{ij}^t)}{\sum_{j'=1}^{d_i} \exp(y_{ij'}^t)} \ln \frac{\exp(y_{ij}^t)}{\sum_{j'=1}^{d_i} \exp(y_{ij'}^t)} \\ &= \frac{\ln \sum_{j'=1}^{d_i} \exp(y_{ij'}^t)}{\sum_{j'=1}^{d_i} \exp(y_{ij'}^t)} \sum_{j=1}^{d_i} \exp(y_{ij}^t),\end{aligned}$$

and then,

$$\begin{aligned}\frac{\partial}{\partial y_{ij}} \psi^*(y_i^t) &= \frac{\exp(y_{ij}^t)}{(\sum_{j'=1}^{d_i} \exp(y_{ij'}^t))^2} \sum_{j'=1}^{d_i} \exp(y_{ij'}^t) - \frac{\ln \sum_{j'=1}^{d_i} \exp(y_{ij'}^t)}{(\sum_{j'=1}^{d_i} \exp(y_{ij'}^t))^2} \exp(y_{ij}^t) \sum_{j'=1}^{d_i} \exp(y_{ij'}^t) \\ &\quad + \frac{\ln \sum_{j'=1}^{d_i} \exp(y_{ij'}^t)}{\sum_{j'=1}^{d_i} \exp(y_{ij'}^t)} \exp(y_{ij}^t) \\ &= \frac{\exp(y_{ij}^t)}{\sum_{j'=1}^{d_i} \exp(y_{ij'}^t)} = \pi_{ij}^t.\end{aligned}$$

Therefore, we get $\nabla \psi^*(y_i^t) = \pi_i^t$. Hence,

$$\begin{aligned}\frac{d}{dt} \mathrm{KL}(\pi^{\mu,\sigma'}, \pi^t) &= \sum_{i=1}^N \left\langle \frac{d}{dt} y_i^t, \pi_i^t - \pi_i^{\mu,\sigma'} \right\rangle \\ &= \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu \nabla_{\pi_i} G(\pi_i^t, \sigma_i), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle \\ &= \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu \nabla_{\pi_i} G(\pi_i^t, \sigma_i'), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle \\ &\quad + \mu \sum_{i=1}^N \langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i') - \nabla_{\pi_i} G(\pi_i^t, \sigma_i), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle. \end{aligned} \tag{38}$$

The first term of (38) can be written as:

$$\begin{aligned}&\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu \nabla_{\pi_i} G(\pi_i^t, \sigma_i'), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle \\ &\le \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma'}, \pi_{-i}^{\mu,\sigma'}) - \mu \nabla_{\pi_i} G(\pi_i^t, \sigma_i'), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle \\ &= \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma'}, \pi_{-i}^{\mu,\sigma'}), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle - \mu \sum_{i=1}^N \langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i'), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle \\ &= \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^{\mu,\sigma'}, \pi_{-i}^{\mu,\sigma'}) - \mu \nabla_{\pi_i} G(\pi_i^{\mu,\sigma'}, \sigma_i'), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle \\ &- \mu \sum_{i=1}^N \langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i') - \nabla_{\pi_i} G(\pi_i^{\mu,\sigma'}, \sigma_i'), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle \\ &\le -\mu \sum_{i=1}^N \langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i') - \nabla_{\pi_i} G(\pi_i^{\mu,\sigma'}, \sigma_i'), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle.\end{aligned}$$

where the first inequality follows from (1), and the second inequality follows from the first-order optimality condition for $\pi^{\mu,\sigma}$. When G is α-divergence, G has a diagonal Hessian is given as:

$$\nabla^2 G(\pi_i, \sigma_i') = \begin{bmatrix} \frac{\sigma_{i1}'}{(\pi_{i1})^{\alpha-2}} & & \\ & \ddots & \\ & & \frac{\sigma_{id_i}'}{(\pi_{id_i})^{\alpha-2}}, \end{bmatrix}$$

and thus, its smallest eigenvalue is lower bounded by $\min_{j\in[d_i]} \sigma_{ij}'$. Therefore,

$$\begin{aligned} &\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^t, \pi_{-i}^t) - \mu \nabla_{\pi_i} G(\pi_i^t, \sigma_i'), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle \\ &\leq -\mu \sum_{i=1}^N \langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i') - \nabla_{\pi_i} G(\pi_i^{\mu,\sigma'}, \sigma_i'), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle \\ &\leq -\mu \left(\min_{i\in[N],\ j\in[d_i]} \sigma_{ij}' \right) \|\pi^t - \pi^{\mu,\sigma'}\|^2. \end{aligned} \tag{39}$$

On the other hand, by compactness of $\mathcal{X}_i$, the second term of (38) is written as:

$$\begin{aligned} &\mu \sum_{i=1}^N \langle \nabla_{\pi_i} G(\pi_i^t, \sigma_i') - \nabla_{\pi_i} G(\pi_i^t, \sigma_i), \pi_i^t - \pi_i^{\mu,\sigma'} \rangle \\ &\leq \mu \cdot \mathrm{diam}(\mathcal{X}) \sqrt{\sum_{i=1}^N \|\nabla_{\pi_i} G(\pi_i^t, \sigma_i') - \nabla_{\pi_i} G(\pi_i^t, \sigma_i)\|^2} \end{aligned} \tag{40}$$

By combining (38), (39), and (40), we get:

$$\begin{aligned} &\frac{d}{dt} \mathrm{KL}(\pi^{\mu,\sigma'}, \pi^t) \\ &\leq -\mu \left(\min_{i\in[N],\ j\in[d_i]} \sigma_{ij}' \right) \|\pi^t - \pi^{\mu,\sigma'}\|^2 + \mu \cdot \mathrm{diam}(\mathcal{X}) \sqrt{\sum_{i=1}^N \|\nabla_{\pi_i} G(\pi_i^t, \sigma_i') - \nabla_{\pi_i} G(\pi_i^t, \sigma_i)\|^2}. \end{aligned}$$

Recall that $\pi^{\mu,\sigma}$ is the stationary point of (37). Therefore, by setting the start point as $\pi^0 = \pi^{\mu,\sigma}$, we have for all $t \geq 0, \pi^t = \pi^{\mu,\sigma}$. In this case, for all $t \geq 0$, $\frac{d}{dt}\mathrm{KL}(\pi^{\mu,\sigma'}, \pi^t) = 0$ and then:

$$\left(\min_{i\in[N],\ j\in[d_i]} \sigma_{ij}' \right) \|\pi^{\mu,\sigma'} - \pi^{\mu,\sigma}\|^2 \leq \mathrm{diam}(\mathcal{X}) \sqrt{\sum_{i=1}^N \|\nabla_{\pi_i} G(\pi_i^t, \sigma_i') - \nabla_{\pi_i} G(\pi_i^t, \sigma_i)\|^2}.$$

For a given $\varepsilon > 0$, let us define $\varepsilon' = \frac{\left(\min_{i\in[N],\ j\in[d_i]} \sigma_{ij}'\right)}{\mathrm{diam}(\mathcal{X})} \varepsilon^2$. Since $\nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \cdot)$ is continuous on $\mathcal{X}_i$, for ε', there exists $\delta > 0$ such that $\|\sigma' - \sigma\| < \delta \Rightarrow \sqrt{\sum_{i=1}^N \|\nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i') - \nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i)\|^2} < \varepsilon' = \frac{\left(\min_{i\in[N],\ j\in[d_i]} \sigma_{ij}'\right)}{\mathrm{diam}(\mathcal{X})} \varepsilon^2$. Thus, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\begin{aligned} &\|\sigma' - \sigma\| < \delta \\ &\Rightarrow \|\pi^{\mu,\sigma'} - \pi^{\mu,\sigma}\| \leq \sqrt{\frac{\mathrm{diam}(\mathcal{X})}{\left(\min_{i\in[N],\ j\in[d_i]} \sigma_{ij}'\right)}} \left(\sum_{i=1}^N \|\nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i') - \nabla_{\pi_i} G(\pi_i^{\mu,\sigma}, \sigma_i)\|^2 \right)^{\frac{1}{4}} < \varepsilon. \end{aligned}$$

This implies that $F(\cdot)$ is a continuous function on $\mathcal{X}$ when G is α-divergence. A similar argument can be applied to Rényi-divergence and reverse KL divergence. □

H.11. Proof of Lemma F.8

Proof of Lemma F.8. First, from the definition of the Bregman divergence, for any $\pi_i \in \mathcal{X}_i$:

$$D_\psi(\pi, T(y_i)) = \psi(\pi) - \psi(T(y_i)) - \langle \nabla\psi(T(y_i)), \pi_i - T(y_i)\rangle. \tag{41}$$

Recall that $\mathcal{X}_i$ satisfies $A\pi_i = b$ for all $\pi_i \in \mathcal{X}_i$ for a matrix $A \in \mathbb{R}^{k_i \times d_i}$ and $b \in \mathbb{R}^{k_i}$. From the assumption for ψ and the first-order optimality condition for the optimization problem of $\arg\max_{x\in\mathcal{X}}\{\langle y_i, x\rangle - \psi(x)\}$, there exists $\nu \in \mathbb{R}^{k_i}$ such that

$$y_i - \nabla\psi(T(y_i)) = A^\top \nu.$$

Thus, we get:

$$\begin{aligned}
\langle y_i, \pi_i - T(y_i)\rangle &= \langle A^\top\nu + \nabla\psi(T(y_i)), \pi_i - T(y_i)\rangle \\
&= \langle \nabla\psi(T(y_i)), \pi_i - T(y_i)\rangle + \nu^\top A\pi_i - \nu^\top A T(y_i) \\
&= \langle \nabla\psi(T(y_i)), \pi_i - T(y_i)\rangle + \nu^\top b - \nu^\top b \\
&= \langle \nabla\psi(T(y_i)), \pi_i - T(y_i)\rangle.
\end{aligned} \tag{42}$$

By combining (41) and (42), we have:

$$D_\psi(\pi, T(y_i)) = \psi(\pi) - \psi(T(y_i)) - \langle y_i, \pi_i - T(y_i)\rangle.$$

□

H.12. Proof of Lemma H.1

Proof of Lemma H.1. Since $\frac{1}{\kappa s + 2\theta}$ is a decreasing function for $s \geq 0$, for all $s \geq 1$,

$$\frac{1}{\kappa s + 2\theta} \leq \int_{s-1}^{s} \frac{1}{\kappa x + 2\theta} dx.$$

Using this inequality, we can upper bound the sum as follows.

$$\begin{aligned}
\sum_{s=0}^{t} \frac{1}{\kappa s + 2\theta} &= \frac{1}{2\theta} + \sum_{s=1}^{t} \frac{1}{\kappa s + 2\theta} \\
&\leq \frac{1}{2\theta} + \sum_{s=1}^{t} \int_{s-1}^{s} \frac{1}{\kappa x + 2\theta} dx \\
&= \frac{1}{2\theta} + \int_0^t \frac{1}{\kappa x + 2\theta} dx \\
&= \frac{1}{2\theta} + \frac{1}{\kappa}\int_0^t \frac{1}{x + \frac{2\theta}{\kappa}} dx \\
&= \frac{1}{2\theta} + \frac{1}{\kappa}\int_{\frac{2\theta}{\kappa}}^{t+\frac{2\theta}{\kappa}} \frac{1}{u} du \qquad \left(u = x + \frac{2\theta}{\kappa}\right) \\
&= \frac{1}{2\theta} + \frac{1}{\kappa}\ln\left(\frac{\kappa}{2\theta}t + 1\right).
\end{aligned}$$

This concludes the proof. □

H.13. Proof of Lemma H.2

Proof of Lemma H.2. By using the first-order optimality condition for σ_i^{k+1}, we have for all $\pi \in \mathcal{X}$:

$$\sum_{i=1}^{N} \langle \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1}) - \mu\nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \pi_i - \sigma_i^{k+1}\rangle \leq 0,$$

and then

$$\sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1}), \pi_i - \sigma_i^{k+1}\rangle \leq \mu \sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^{k}), \pi_i - \sigma_i^{k+1}\rangle.$$

When G is α-divergence, we have for all $\pi \in \mathcal{X}$:

$$\begin{aligned}\sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^{k}), \pi_i - \sigma_i^{k+1}\rangle &= \frac{1}{1-\alpha}\sum_{i=1}^{N}\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1} - \pi_{ij})\left(\frac{\sigma_{ij}^{k}}{\sigma_{ij}^{k+1}}\right)^{1-\alpha} \\ &= \frac{1}{1-\alpha}\sum_{i=1}^{N}\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1} - \pi_{ij}) = 0,\end{aligned}$$

where we use the assumption that $\sigma^{k+1} = \sigma^k$ and $\mathcal{X}_i = \Delta^{d_i}$. Similarly, when G is Rényi-divergence, we have for all $\pi \in \mathcal{X}$:

$$\begin{aligned}\sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^{k}), \pi_i - \sigma_i^{k+1}\rangle &= \frac{\alpha}{1-\alpha}\sum_{i=1}^{N}\frac{1}{\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1})^{\alpha}(\sigma_{ij}^{k})^{1-\alpha}}\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1} - \pi_{ij})\left(\frac{\sigma_{ij}^{k}}{\sigma_{ij}^{k+1}}\right)^{1-\alpha} \\ &= \frac{\alpha}{1-\alpha}\sum_{i=1}^{N}\frac{1}{\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1})^{\alpha}(\sigma_{ij}^{k})^{1-\alpha}}\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1} - \pi_{ij}) = 0.\end{aligned}$$

Furthermore, if G is reverse KL divergence, we have for all $\pi \in \mathcal{X}$:

$$\begin{aligned}\sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^{k}), \pi_i - \sigma_i^{k+1}\rangle &= \sum_{i=1}^{N}\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1} - \pi_{ij})\frac{\sigma_{ij}^{k}}{\sigma_{ij}^{k+1}} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1} - \pi_{ij}) = 0,\end{aligned}$$

Thus, we have for all $\pi \in \mathcal{X}$:

$$\sum_{i=1}^{N}\langle \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1}), \pi_i - \sigma_i^{k+1}\rangle \leq 0.$$

This is equivalent to the first-order optimality condition for $\pi^* \in \Pi^*$. Therefore, $\sigma^{k+1} = \sigma^k$ is a Nash equilibrium of the underlying game. □

H.14. Proof of Lemma H.3

Proof of Lemma H.3. First, we prove the statement for α-divergence: $G(\sigma_i^{k+1}, \sigma_i^k) = \frac{1}{\alpha(1-\alpha)}\left(1 - \sum_{j=1}^{d_i}(\sigma_{ij}^{k+1})^{\alpha}(\sigma_{ij}^{k})^{1-\alpha}\right)$. From the definition of α-divergence, we have for all $\pi^* \in \Pi^*$:

$$\begin{aligned}\sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^{k}), \sigma_i^{k+1} - \pi_i^*\rangle &= \frac{1}{1-\alpha}\sum_{i=1}^{N}\sum_{j=1}^{d_i}(\pi_{ij}^* - \sigma_{ij}^{k+1})\left(\frac{\sigma_{ij}^{k}}{\sigma_{ij}^{k+1}}\right)^{1-\alpha} \\ &= \frac{1}{1-\alpha}\sum_{i=1}^{N}\sum_{j=1}^{d_i}\pi_{ij}^*\left(\frac{\sigma_{ij}^{k}}{\sigma_{ij}^{k+1}}\right)^{1-\alpha} - \frac{1}{1-\alpha}\sum_{i=1}^{N}\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1})^{\alpha}(\sigma_{ij}^{k})^{1-\alpha}.\end{aligned}$$

Here, when $\alpha \in (0,1)$, we get $\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1})^\alpha(\sigma_{ij}^k)^{1-\alpha} \le 1$. Thus,

$$\begin{aligned}\sum_{i=1}^N \langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle &\ge \frac{1}{1-\alpha}\sum_{i=1}^N\sum_{j=1}^{d_i} \pi_{ij}^* \left(\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}\right)^{1-\alpha} - \frac{N}{1-\alpha} \\ &= \frac{N}{1-\alpha}\exp\left(\ln\left(\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^{d_i}\pi_{ij}^*\left(\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}\right)^{1-\alpha}\right)\right) - \frac{N}{1-\alpha} \\ &\ge \frac{N}{1-\alpha}\exp\left(\frac{1-\alpha}{N}\sum_{i=1}^N\sum_{j=1}^{d_i}\pi_{ij}^*\ln\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}\right) - \frac{N}{1-\alpha} \\ &= \frac{N}{1-\alpha}\exp\left(\frac{1-\alpha}{N}\left(\mathrm{KL}(\pi^*, \sigma^{k+1}) - \mathrm{KL}(\pi^*, \sigma^k)\right)\right) - \frac{N}{1-\alpha},\end{aligned}$$

where the second inequality follows from the concavity of the $\ln(\cdot)$ function and Jensen's inequality for concave functions. Since $\ln(\cdot)$ is strictly concave, the equality holds if and only if $\sigma^{k+1} = \sigma^k$. Therefore, under the assumption that $\sigma^{k+1} \ne \sigma^k$, we get:

$$\begin{aligned}\mathrm{KL}(\pi^*, \sigma^{k+1}) - \mathrm{KL}(\pi^*, \sigma^k) &< \frac{N}{1-\alpha}\ln\left(1 + \frac{1-\alpha}{N}\sum_{i=1}^N \langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle\right) \\ &\le \sum_{i=1}^N \langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle, \end{aligned} \tag{43}$$

where the second inequality follows from $\ln(1+x) \le x$ for $x > -1$. From the first-order optimality condition for σ_i^{k+1}, we have for all $\pi^* \in \Pi^*$:

$$\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1}) - \mu \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \pi_i^* - \sigma_i^{k+1} \rangle \le 0.$$

Then,

$$\begin{aligned}\sum_{i=1}^N \langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle &\le \frac{1}{\mu}\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\sigma_i^{k+1}, \sigma_{-i}^{k+1}), \sigma_i^{k+1} - \pi_i^* \rangle \\ &\le \frac{1}{\mu}\sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi_i^*, \pi_{-i}^*), \sigma_i^{k+1} - \pi_i^* \rangle, \end{aligned} \tag{44}$$

where the second inequality follows from (1). Moreover, since π^* is the Nash equilibrium, from the first-order optimality condition, we get:

$$\sum_{i=1}^N \langle \sigma_i^{k+1} - \pi_i^*, \nabla_{\pi_i} v_i(\pi_i^*, \pi_{-i}^*) \rangle \le 0. \tag{45}$$

By combining (43), (44), and (45), if $\sigma^{k+1} \ne \sigma^k$, we have any $\pi^* \in \Pi^*$:

$$\mathrm{KL}(\pi^*, \sigma^{k+1}) - \mathrm{KL}(\pi^*, \sigma^k) < 0.$$

Next, we prove the statement for Rényi-divergence: $G(\sigma_i^{k+1}, \sigma_i^k) = \frac{1}{\alpha-1}\ln\left(\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1})^\alpha(\sigma_{ij}^k)^{1-\alpha}\right)$. We have for all $\pi^* \in \Pi^*$:

$$\begin{aligned}\sum_{i=1}^N \langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle &= \frac{\alpha}{1-\alpha}\sum_{i=1}^N \frac{1}{\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1})^\alpha(\sigma_{ij}^k)^{1-\alpha}}\sum_{j=1}^{d_i}(\pi_{ij}^* - \sigma_{ij}^{k+1})\left(\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}\right)^{1-\alpha} \\ &= \frac{\alpha}{1-\alpha}\sum_{i=1}^N \frac{1}{\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1})^\alpha(\sigma_{ij}^k)^{1-\alpha}}\sum_{j=1}^{d_i}\pi_{ij}^*\left(\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}\right)^{1-\alpha} - \frac{N\alpha}{1-\alpha}.\end{aligned}$$

Again, by using $\sum_{j=1}^{d_i}(\sigma_{ij}^{k+1})^{\alpha}(\sigma_{ij}^{k})^{1-\alpha} \leq 1$ when $\alpha \in (0,1)$, we get:

$$\begin{aligned}
\sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle &\geq \frac{\alpha}{1-\alpha}\sum_{i=1}^{N}\sum_{j=1}^{d_i}\pi_{ij}^*\left(\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}\right)^{1-\alpha} - \frac{N\alpha}{1-\alpha} \\
&= \frac{N\alpha}{1-\alpha}\exp\left(\ln\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{d_i}\pi_{ij}^*\left(\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}\right)^{1-\alpha}\right)\right) - \frac{N\alpha}{1-\alpha} \\
&\geq \frac{N\alpha}{1-\alpha}\exp\left(\frac{1-\alpha}{N}\sum_{i=1}^{N}\sum_{j=1}^{d_i}\pi_{ij}^*\ln\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}\right) - \frac{N\alpha}{1-\alpha} \\
&= \frac{N\alpha}{1-\alpha}\exp\left(\frac{1-\alpha}{N}\left(\mathrm{KL}(\pi^*, \sigma^{k+1}) - \mathrm{KL}(\pi^*, \sigma^k)\right)\right) - \frac{N\alpha}{1-\alpha},
\end{aligned}$$

where the second inequality follows from Jensen's inequality for $\ln(\cdot)$ function. Since $\ln(\cdot)$ is strictly concave, the equality holds if and only if $\sigma^{k+1} = \sigma^k$. Therefore, under the assumption that $\sigma^{k+1} \neq \sigma^k$, we get:

$$\begin{aligned}
\mathrm{KL}(\pi^*, \sigma^{k+1}) - \mathrm{KL}(\pi^*, \sigma^k) &< \frac{N\alpha}{1-\alpha}\ln\left(1 + \frac{1-\alpha}{N\alpha}\sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle\right) \\
&\leq \sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle, \qquad (46)
\end{aligned}$$

where the second inequality follows from $\ln(1+x) \leq x$ for $x > -1$. Thus, by combining (44), (45), and (46), if $\sigma^{k+1} \neq \sigma^k$, we have any $\pi^* \in \Pi^*$:

$$\mathrm{KL}(\pi^*, \sigma^{k+1}) - \mathrm{KL}(\pi^*, \sigma^k) < 0.$$

Finally, we prove the statement for reverse KL divergence: $G(\sigma_i^{k+1}, \sigma_i^k) = \sum_{j=1}^{d_i}\sigma_{ij}^k \ln\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}$. We have for all $\pi^* \in \Pi^*$:

$$\begin{aligned}
\sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle &= \sum_{i=1}^{N}\sum_{j=1}^{d_i}(\pi_{ij}^* - \sigma_{ij}^{k+1})\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{d_i}\pi_{ij}^*\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}} - N \\
&= N\exp\left(\ln\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{d_i}\pi_{ij}^*\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}\right)\right) - N \\
&\geq N\exp\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{d_i}\pi_{ij}^*\ln\frac{\sigma_{ij}^k}{\sigma_{ij}^{k+1}}\right) - N \\
&= N\exp\left(\frac{1}{N}\left(\mathrm{KL}(\pi^*, \sigma^{k+1}) - \mathrm{KL}(\pi^*, \sigma^k)\right)\right) - N,
\end{aligned}$$

where the inequality follows from Jensen's inequality for $\ln(\cdot)$ function. Thus, under the assumption that $\sigma^{k+1} \neq \sigma^k$, we get:

$$\begin{aligned}
\mathrm{KL}(\pi^*, \sigma^{k+1}) - \mathrm{KL}(\pi^*, \sigma^k) &< N\ln\left(1 + \frac{1}{N}\sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle\right) \\
&\leq \sum_{i=1}^{N}\langle \nabla_{\pi_i} G(\sigma_i^{k+1}, \sigma_i^k), \sigma_i^{k+1} - \pi_i^* \rangle, \qquad (47)
\end{aligned}$$

where the second inequality follows from $\ln(1+x) \leq x$ for $x > -1$. Thus, by combining (44), (45), and (47), if $\sigma^{k+1} \neq \sigma^k$, we have any $\pi^* \in \Pi^*$:

$$\mathrm{KL}(\pi^*, \sigma^{k+1}) - \mathrm{KL}(\pi^*, \sigma^k) < 0.$$

□

I. Additional Experimental Results and Details

I.1. Payoff Matrix in Three-Player Biased RPS Game

Table 2: Three-Player Biased RPS game matrix.

	R	P	S
R	0	$-1/3$	1
P	$1/3$	0	$-1/3$
S	-1	$1/3$	0

I.2. Experimental Setting for Section 6

The experiments in Section 6 are conducted in Ubuntu 20.04.2 LTS with Intel(R) Core(TM) i9-10850K CPU @ 3.60GHz and 64GB RAM.

In the full feedback setting, we use a constant learning rate $\eta = 0.1$ for MWU and OMWU, and APMD in all three games. For APMD, we set $\mu = 0.1$ and $T_\sigma = 100$ for KL and reverse KL divergence perturbation, and set $\mu = 0.1$ and $T_\sigma = 20$ for squared ℓ^2-distance perturbation. As an exception, $\eta = 0.01$, $\mu = 1.0$, and $T_\sigma = 200$ are used for APMD with squared ℓ^2-distance perturbation in the random payoff games with 50 actions.

For the noisy feedback setting, we use the lower learning rate $\eta = 0.01$ for all algorithms, except APMD with squared ℓ^2-distance perturbation for the random payoff games with 50 actions. We update the slingshot strategy profile σ^k every $T_\sigma = 1000$ iterations in APMD with KL and reverse KL divergence perturbation, and update it every $T_\sigma = 200$ iterations in APMD with squared ℓ^2-distance perturbation. For APMD with ℓ^2-distance perturbation in the random payoff games with 50 actions, we set $\eta = 0.001$ and $T_\sigma = 2000$.

I.3. Additional Experiments

In this section, we compare the performance of APMD and APFTRL to MWU, OMWU, and optimistic gradient descent (OGD) (Daskalakis et al., 2018; Wei et al., 2021) in the full/noisy feedback setting. The parameter settings for MWU, OMWU, and APMD are the same as Section 6. For APFTRL, we use the squared ℓ^2-distance and the parameter is the same as APMD with squared ℓ^2-distance perturbation. For OGD, we use the same learning rate as APMD with squared ℓ^2-distance perturbation.

Figure 4 shows the logarithm of the gap function for π^t averaged over 100 instances with full feedback. We observe that APMD and APFTRL with squared ℓ^2-distance perturbation exhibit competitive performance to OGD. The experimental results in the noisy feedback setting are presented in Figure 5. Surprisingly, in the noisy feedback setting, all APMD-based algorithms and the APFTRL-based algorithm exhibit overwhelmingly superior performance to OGD in all three games.

I.4. Comparison with the Averaged Strategies of No-Regret Learning Algorithms

This section compares the last-iterate strategies π^t of APMD and APFTRL with the average of strategies $\frac{1}{t}\sum_{\tau=1}^{t} \pi^\tau$ of MWU, regret matching (RM) (Hart & Mas-Colell, 2000), and regret matching plus (RM+) (Tammelin, 2014). The parameter settings for MWU, APMD, and APFTRL, as used in Section I.3, are maintained. Figure 6 illustrates the logarithm of the gap function averaged over 100 instances with full feedback. The results show that the last-iterate strategies of APMD and APFTRL squared ℓ^2-distance perturbation exhibit a lower gap than the averaged strategies of MWU, RM, and RM+.

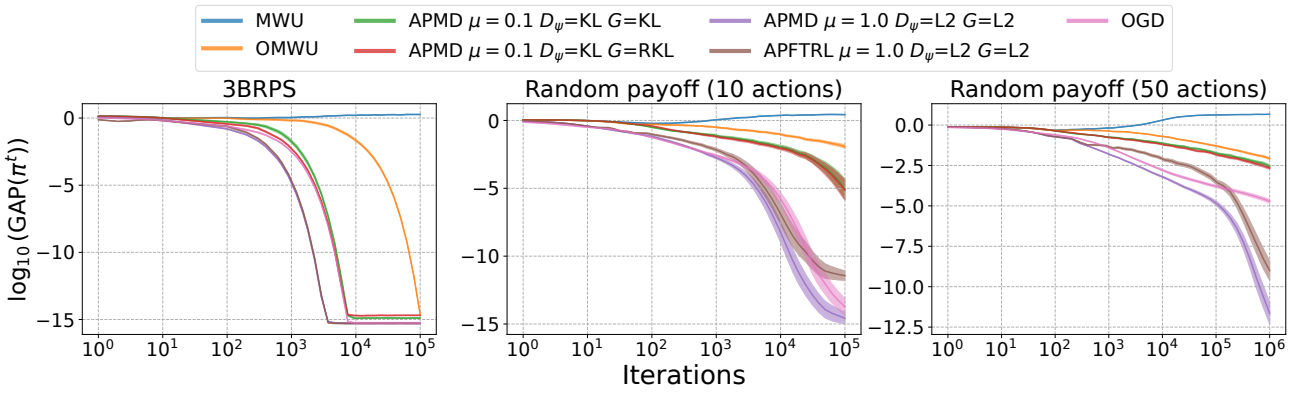


Figure 4: The gap function for π^t for APMD, APFTRL, MWU, OMWU, and OGD with full feedback. The shaded area represents the standard errors. Note that the KL divergence, reverse KL divergence, and squared ℓ^2-distance are abbreviated to KL, RKL, and L2, respectively.

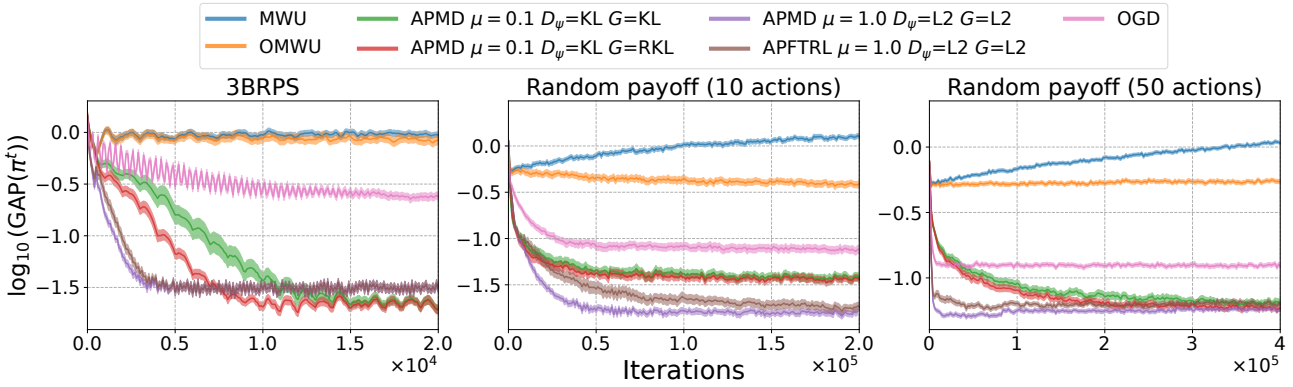


Figure 5: The gap function for π^t for APMD, APFTRL, MWU, OMWU, and OGD with noisy feedback. The shaded area represents the standard errors.

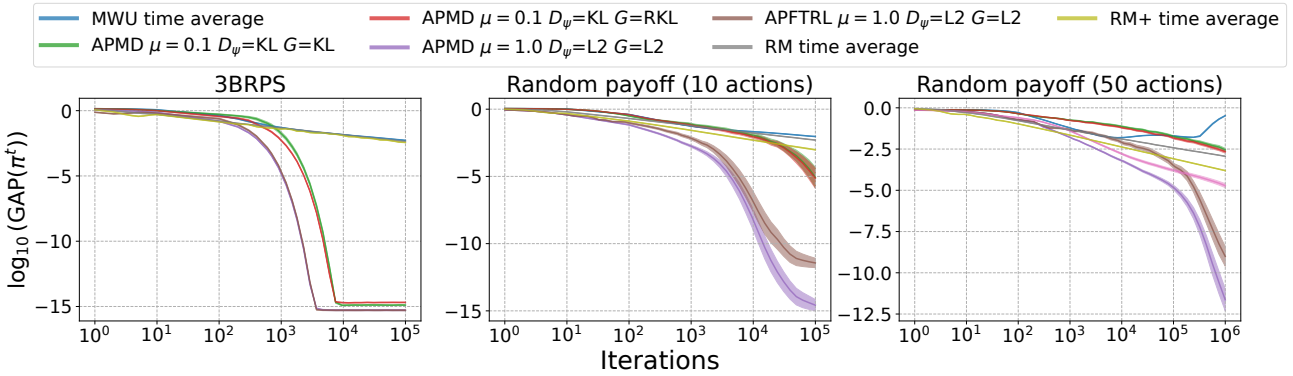
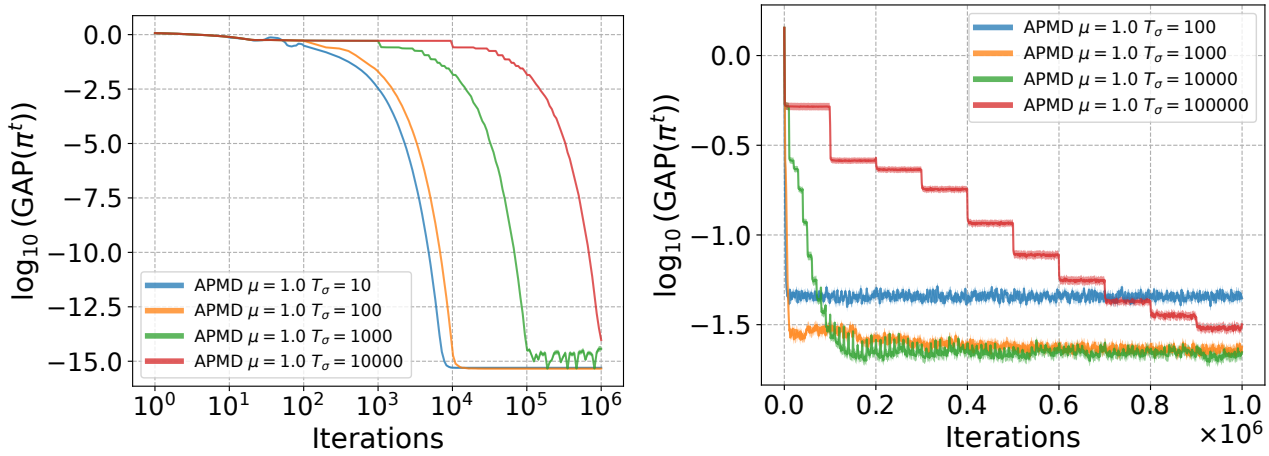


Figure 6: Comparison between the gap function of the last-iterate strategy profile of APMD, APFTRL, and the averaged strategy profile of MWU, RM, and RM+ with full feedback. The shaded area represents the standard errors.

I.5. Sensitivity Analysis of Update Interval for the Slingshot Strategy

In this section, we investigate the performance when changing the update interval of the slingshot strategy. We vary the T_σ of APMD with L2 perturbation in 3BRPS with full feedback to be $T_\sigma \in \{10, 100, 1000, 10000\}$, and with noisy feedback

to be $T_\sigma \in \{10, 100, 1000, 10000\}$. All other parameters are the same as in Section 6. Figure 7 shows the logarithm of the gap function for π^t averaged over 100 instances in 3BRPS with full/noisy feedback. We observe that the smaller the T_σ, the faster π^t converges. However, if T_σ is too small, π^t does not converge (See $T_\sigma = 10$ with full feedback, and $T_\sigma = 100$ with noisy feedback in Figure 7).



(a) Full feedback

(b) Noisy feedback

Figure 7: The gap function for π^t for APMD with varying T_σ in 3BRPS with full/noisy feedback. The shaded area represents the standard errors.