

---

# Team-Fictitious Play for Reaching Team-Nash Equilibrium in Multi-team Games

---

**Ahmed Said Donmez**  
Bilkent University  
said.donmez@bilkent.edu.tr

**Yuksel Arslantas**  
Bilkent University  
yuksel.arslantas@bilkent.edu.tr

**Muhammed O. Sayin**  
Bilkent University  
sayin@ee.bilkent.edu.tr

## Abstract

Multi-team games, prevalent in robotics and resource management, involve team members striving for a joint best response against other teams. Team-Nash equilibrium (TNE) predicts the outcomes of such coordinated interactions. However, can teams of self-interested agents reach TNE? We introduce Team-Fictitious Play (Team-FP), a new variant of fictitious play where agents respond to the last actions of team members and the beliefs formed about other teams with some inertia in action updates. This design is essential in team coordination beyond the classical fictitious play dynamics. We focus on zero-sum potential team games (ZSPTGs) where teams can interact pairwise while the team members do not necessarily have identical payoffs. We show that Team-FP reaches near TNE in ZSPTGs with a quantifiable error bound. We extend Team-FP dynamics to multi-team Markov games for model-based and model-free cases. The convergence analysis tackles the challenge of non-stationarity induced by evolving opponent strategies based on the optimal coupling lemma and stochastic differential inclusion approximation methods. Our work strengthens the foundation for using TNE to predict the behavior of decentralized teams and offers a practical rule for team learning in multi-team environments. We provide extensive simulations of Team-FP dynamics and compare its performance with other widely studied dynamics such as smooth fictitious play and multiplicative weights update. We further explore how different parameters impact the speed of convergence.

## 1 Introduction

Multi-team games are increasingly common, e.g., in robotics and resource management [Silva and Chaimowicz, 2017, Vinyals et al., 2019, Jaderberg et al., 2019]. Unlike non-cooperative multi-agent settings, team members strive for a joint best response against other teams as if the entire team is a single decision-maker. Team-Nash equilibrium (TNE), where team members coordinate in the best team response against other teams, can capture this to predict the outcome of coordinated team interactions [Farina et al., 2018, Zhang et al., 2021]. Game-theoretical equilibrium is often justified by its emergence from non-equilibrium adaptation of self-interested learners (e.g., see [Fudenberg and Levine, 2009]). However, the question of whether the teams of self-interested agents can reach TNE in multi-team games remains largely unexplored. This paper investigates this very question.

For example, TNE generally can arise if the team members can learn to correlate their actions in the best team response independent of the opponent. However, the widely studied fictitious play (FP) dynamics do not necessarily reach the best team outcome even when there are no opponents, e.g., in

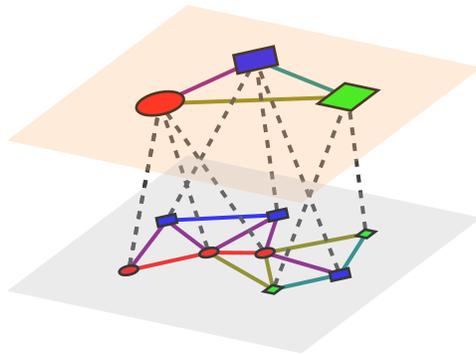


Figure 1: An illustration of networked interconnections from different teams. Nodes in bottom and top layers refer, resp., to team members and teams. Undirected edges represent the impact of actions on the payoff functions. We use different colors and shapes to represent agents from the same teams, and they are connected via dashed edges.

potential games. We present a slight adjustment of FP, called *Team-FP*, provably reaching TNE in multi-team competition even with networked interconnections, where agents' payoffs depend only on the neighbors' actions. Similar to the FP, here, agents respond greedily to the beliefs formed about the opponent teams' joint play based on past observations. Different from the FP, *Team-FP* incorporates the key features: (i) response to the last actions of team members, and (ii) inertia in action updates. These features, inspired by log-linear learning dynamics [Blume, 1993], play a crucial role in driving team coordination towards TNE. Notably, *Team-FP* reduces to the smoothed FP [Fudenberg and Kreps, 1993] (or log-linear learning) when each team has a single agent (or there is a single team).

Multi-team competition spans diverse domains. For example, robotics, resource management, online gaming, and financial markets [Kitano et al., 1997, Cardenas et al., 2009, Silva and Chaimowicz, 2017, Vinyals et al., 2019, Jaderberg et al., 2019] involve multi-team competition. To model such interactions, we consider multiple teams with possibly different number of team members. These team members have networked interconnections, as depicted in Figure 1. We focus on multi-team *zero-sum potential team games* where teams have *pairwise interactions* (ZSPTGs). For any opponent team strategy, team members effectively play an underlying potential game, as in distributed optimization applications, e.g., see [Arslan et al., 2007, Xu et al., 2012, Zheng et al., 2014]. Notably, ZSPTGs reduce to zero-sum polymatrix games [Cai et al., 2016] if each team has a single agent, and to potential games [Monderer and Shapley, 1996a] if there is a single team. Additionally, widely studied two-team zero-sum games, e.g., see [Farina et al., 2018, Zhang et al., 2021, Carminati et al., 2022, Kalogiannis et al., 2022], are a special case of ZSPTGs.

We show that the *Team-FP* dynamics reach near TNE in ZSPTGs. This means that the empirical average of team actions converge to the near best response each team can take against the average actions of other teams. We quantify the approximation error, showing it decreases with the level of exploration in the agents' responses. Similar to the FP dynamics, *Team-FP* is also *rational*: teams can learn (near) optimal strategies if opponent teams play stationary strategies. These results strengthen the applicability of TNE for predicting team behavior in multi-team competition and provide a practical rule for teams of self-interested agents to learn coordination in multi-team settings.

A key challenge in our analysis is handling the non-stationary nature of learning, as opponent teams' strategies change over time. We address this by leveraging the optimal coupling lemma (e.g. see [Levin and Peres, 2017, Chapter 4]) and stochastic differential inclusion approximation methods (e.g., see [Benaïm et al., 2005, Perkins and Leslie, 2013]) to the repeated play of games. Motivated from the recent interest in multi-agent reinforcement learning, we can extend *Team-FP* dynamics to finite horizon multi-team Markov games for both model-based and model-free cases. We discuss this extension and analyze its convergence numerically in Appendix C.

**Related works.** FP and its variants offer convergence guarantees in important classes of games [Fudenberg and Levine, 2009], yet not in every class of games [Hart and Mas-Colell, 2003]. For example, they reach equilibrium in potential games [Monderer and Shapley, 1996a], but not necessarily the most efficient one for the team. Log-linear learning can achieve efficient equilibrium for the team [Marden and Shamma, 2012, Tatarenko, 2017]. However, it is not clear whether such dynamics can track efficient equilibrium in dynamic environments (induced by the evolving strategies of opponent teams). Notably, Tatarenko [2018] and Donmez et al. [2024] addressed, resp., efficient learning under non-stationarity induced by the decaying exploration in agents' responses for the repeated play of potential games and non-stationarity induced by evolving stage games in Markov team problems

(also known as identical-interest Markov games). These approaches are orthogonal to our analysis to extend our results to the exact TNE convergence in repeated multi-team games or to learning in infinite horizon Markov games.

FP and its variants can reach equilibrium in two-agent zero-sum games [Hofbauer and Sandholm, 2002] yet not necessarily in multi-agent zero-sum games with more than two agents. We can transform any general-sum game to a multi-agent zero-sum game by introducing a non-effective auxiliary agent (with a single action). There have been several attempts to address zero-sum games beyond two-agent cases [Bergman and Fokin, 1998, Cai and Daskalakis, 2011, Cai et al., 2016]. For example, Ewerhart and Valkanova [2020] addressed the convergence of continuous and discrete-time FP in zero-sum network games, where each agent plays multiple two-agent games with separate actions, and the overall utilities sum to zero. Notably, Cai et al. [2016] introduced zero-sum polymatrix games where agents have network separable pairwise interactions with applications in security. Recently, FP has been shown to reach Nash equilibrium in zero-sum polymatrix games [Park et al., 2023]. Following the same trend, we focus on learning in ZSPTGs extending two-team zero-sum games to multi-team games with pairwise team interactions. However, we highlight that two-team zero-sum or ZSPTGs are not necessarily multi-agent zero-sum polymatrix games (as the agent payoffs do not necessarily sum to zero) and the classical FP dynamics do not necessarily converge TNE or Nash equilibrium in such multi-team games.

Recent studies on two-team zero-sum games and adversarial team games (e.g., see [Celli and Gatti, 2018, Farina et al., 2018, Zhang et al., 2022, Carminati et al., 2022]) have primarily focused on the efficient computation of team-minimax equilibrium. In particular, Celli and Gatti [2018] examines the efficiency of different communication types, highlighting the promising results of *ex ante* communication, referring to pre-play communication among team members. Consequently, the studies of Farina et al. [2018], Zhang et al. [2022], Carminati et al. [2022] often model teams as a single agent with imperfect recall, incorporating ex ante communication within the team. Notably, Farina et al. [2018] introduced the Fictitious Team Play (FTP) algorithm for extensive-form two-team zero-sum games with imperfect information, where team members communicate ex ante. In this approach, Farina et al. [2018] used fictitious play (FP) on a simplified version of the original game, embedded within the game tree. This method essentially applied FTP to the original adversarial team game. To find the best response, they used mixed-integer linear programming. Team-FP differs from such approaches by letting agents *learn* to team up while following their self-interest based on simple behavioral rules, as in the log-linear learning and FP dynamics.

The rest of the paper is organized as follows. We describe ZSPTGs in §2. We present the (independent) Team-FP dynamics in §3. We provide analytical and numerical results, resp., in §4 and §5. We conclude the paper with some remarks in §6. Appendices include the proofs of the technical results, some further numerical experiments and the extension to multi-team Markov games.

*Notation:* Given a finite set  $A$ , we let  $\Delta(A)$  denote the probability simplex over  $A$ . We let  $f(\mu) = E_{a \sim \mu}[f(a)]$  for any probability distribution  $\mu \in \Delta(A)$  and any functional  $f : A \rightarrow \mathbb{R}$ . Furthermore, we define the smoothed best response to any functional  $q : A \rightarrow \mathbb{R}$  by

$$\text{br}_\tau(q)(a) = \frac{\exp(q(a)/\tau)}{\sum_{\tilde{a} \in A} \exp(q(\tilde{a})/\tau)} \quad \forall a \in A \quad \Leftrightarrow \quad \text{br}_\tau(q) = \underset{\mu \in \Delta(A)}{\text{argmax}} \{q(\mu) + \tau \mathcal{H}(\mu)\}, \quad (1)$$

for some temperature parameter  $\tau > 0$ , where  $\mathcal{H}(\mu) := E_{a \sim \mu}[-\log(\mu(a))]$  for all  $\mu \in \Delta(A)$  is the entropy regularization.

## 2 Game Formulation

Consider a *multi-team game*, characterized by the tuple  $\langle \mathcal{T}, (A^i, u^i)_{i \in \mathcal{I}} \rangle$ , where  $\mathcal{T}$  and  $\mathcal{I}$  denote, resp., the teams' and agents' index sets, and  $A^i$  and  $u^i : A \rightarrow \mathbb{R}$  (with  $A := \prod_{j \in \mathcal{I}} A^j$ ) denote, resp., the agent  $i$ 's finite action set and payoff function. Agents take their actions to maximize their payoff functions.

**Definition 2.1** (Zero-sum Potential Team Game). Let  $\mathcal{I}^m$  denote the index set of agents in team  $m$  and  $A^m := \prod_{i \in \mathcal{I}^m} A^i$  denote the set of joint actions for team  $m$ . We say that a multi-team game is *zero-sum potential team game* (ZSPTG) if for each team  $m \in \mathcal{T}$ , there exists a potential function

$\phi^m : A \rightarrow \mathbb{R}$  such that

$$u^i(\hat{a}^i, a^{-i}, \underline{a}^{-m}) - u^i(a) = \phi^m(\hat{a}^i, a^{-i}, \underline{a}^{-m}) - \phi^m(a), \quad (2)$$

for all  $(\hat{a}^i, a) \in A^i \times A$  and  $i \in \mathcal{I}^m$ , where  $a^{-i} := \{a^j\}_{j \in \mathcal{I}^m \setminus \{i\}}$  are the actions of other team members,  $\underline{a}^{-m} := \{\underline{a}^\ell\}_{\ell \neq m}$  are the action profiles of other teams, where  $\underline{a}^\ell \in \underline{A}^\ell$  is team  $\ell$ 's action profile. The potential functions sum to zero, i.e., we have

$$\sum_{m \in \mathcal{T}} \phi^m(a) = 0 \quad \forall a \in A. \quad (3)$$

Furthermore, the actions have network separable interactions across teams such that we can separate the potential functions and correspondingly payoff functions as

$$\phi^m \equiv \sum_{\ell \neq m} \phi^{m\ell} \quad \text{and} \quad u^i \equiv \sum_{\ell \neq m} u^{i\ell} \quad \forall i \in \mathcal{I}^m, \quad (4)$$

for some  $\phi^{m\ell} : \underline{A}^m \times \underline{A}^\ell \rightarrow \mathbb{R}$  and  $u^{i\ell} : \underline{A}^m \times \underline{A}^\ell \rightarrow \mathbb{R}$ .

The following example generalizes the potential game formulation for distributed optimization (e.g., see [Arslan et al., 2007, Xu et al., 2012, Zheng et al., 2014]) to two-team zero-sum potential games.

*Example 2.2.* Consider two teams of agents interacting over a network. We can represent their interactions via a graph  $G = (V, E)$ , where the set of vertices  $V$  refers to the agents and the set of (undirected) edges refers to their interactions. Let agent  $i$  from team  $m$  receive a local payoff  $r^i : A^i \times \prod_{j:(i,j) \in E} A^j \rightarrow \mathbb{R}$  depending on the actions of the neighboring agents only. Agent  $i$  adds the neighboring team members' local payoffs whereas subtracts the other neighbors' local payoffs in his/her total payoff. In other words, his/her total payoff is given by

$$u^i \equiv \sum_{j:(i,j) \in E} \mathbb{I}_{\{j \in \mathcal{I}^m\}} r^j - \sum_{j:(i,j) \in E} \mathbb{I}_{\{j \notin \mathcal{I}^m\}} r^j. \quad (5)$$

This yields that the team  $m$  has the potential function

$$\phi^m \equiv \sum_{j \in \mathcal{I}^m} r^j - \sum_{j \notin \mathcal{I}^m} r^j, \quad (6)$$

and therefore, these potential functions sum to zero. However, the potential function is not generally the sum of the payoffs in potential games.

*Remark 2.3* (General-sum ZSPTGs). In ZSPTGs, the underlying game can be a *general-sum* game although the team-potentials sum to zero, as described in (3). For example, consider two competing teams whose team members have identical payoffs corresponding to their team potentials. If the teams have different number of members, then the agents' payoffs do not sum to zero or a constant while the team potentials do so.

In the following, we introduce TNE, generalizing team-minimax equilibrium for two-team zero-sum games, e.g., see [Von Stengel and Koller, 1997], to multi-team games. Particularly, at TNE, no team has an incentive to change their team strategy.

**Definition 2.4** (Team-Nash Gap). Given the strategy profile of teams  $\{\pi^m \in \Delta(\underline{A}^m)\}_{m \in \mathcal{T}}$ , we define the *team-Nash gap* for team  $m$  as

$$\text{TNG}^m(\pi) := \max_{\tilde{\pi} \in \Delta(\underline{A}^m)} \{\phi^m(\tilde{\pi}, \pi^{-m})\} - \phi^m(\pi), \quad (7)$$

and  $\text{TNG}(\pi) := \sum_{m \in \mathcal{T}} \text{TNG}^m(\pi)$ , where  $\pi^{-m} := \{\pi^\ell\}_{\ell \neq m}$ . Correspondingly, we say that the strategy profile of teams  $\{\pi^m\}_{m \in \mathcal{T}}$  is  $\epsilon$ -TNE if  $\text{TNG}(\pi) \leq \epsilon$ .

### 3 Team-FP Dynamics

We first present the Team-FP dynamics combining the log-linear learning and fictitious play for learning in multi-team games played repeatedly, and then extend Team-FP to multi-team MGs in Appendix C.

---

**Algorithm 1** (Independent) Team-FP

---

**initialize:**  $\{\pi_0^\ell\}_{\ell \neq m}$  and  $\{a_{-1}^j\}_{j \in \mathcal{I}^m \setminus \{i\}}$  arbitrarily  
**for** each stage  $k = 0, 1, \dots$  **do**  
    play  $a_k^i \sim \text{br}_\tau(u^i(\cdot, a_{k-1}^{-i}, \pi_k^{-m}))$  or  $a_k^i = a_{k-1}^i$  in a coordinated way (or independently)  
    update  $\pi_{k+1}^\ell = \pi_k^\ell + \alpha_k(\underline{a}_k^\ell - \pi_k^\ell)$  for all  $\ell \neq m$   
**end for**

---

Let  $a_k^i \in A^i$  denote the agent  $i$ 's action at the  $k$ th repetition in the repeated play of the underlying ZSPTG. Correspondingly,  $\underline{a}_k^m = (a_k^i)_{i \in \mathcal{I}^m}$  denote the team  $m$ 's action profile. Observing the joint actions of team  $m$ , agents  $j \notin \mathcal{I}^m$  can form a belief about the team  $m$ 's joint strategy. Let  $\pi_k^m \in \Delta(\underline{A}^m)$  denote the belief they formed. Consider actions as pure strategy where the associated action gets played with probability 1. Then, agents  $j \notin \mathcal{I}^m$  can update their beliefs about team  $m$ 's strategy according to

$$\pi_{k+1}^m = \pi_k^m + \alpha_k(\underline{a}_k^m - \pi_k^m) \quad \forall k = 0, 1, \dots \quad (8)$$

such that the belief  $\pi_{k+1}^m$  also corresponds to the (weighted) empirical average of the past action profiles  $\{\underline{a}_0^m, \dots, \underline{a}_k^m\}$ .

Agent  $i \in \mathcal{I}^m$  either takes the previous action  $a_{k-1}^i$  (i.e.,  $a_k^i = a_{k-1}^i$ ), or takes the action  $a_k^i \sim \text{br}_\tau(u^i(\cdot, a_{k-1}^{-i}, \pi_k^{-m}))$  according to the smoothed best response (as described in (1)) to the previous actions of team members  $a_{k-1}^{-i} := \{a_{k-1}^j\}_{j \in \mathcal{I}^m \setminus \{i\}}$  and the beliefs  $\pi_k^{-m} := \{\pi_k^\ell\}_{\ell \neq m}$  formed about other teams. It is instructive to note that the definition of potential function  $\phi^m(\cdot)$ , as described in (2), yields that

$$\text{br}_\tau(u^i(\cdot, a_{k-1}^{-i}, \pi_k^{-m})) \equiv \text{br}_\tau(\phi^m(\cdot, a_{k-1}^{-i}, \pi_k^{-m})) \quad \forall i \in \mathcal{I}^m. \quad (9)$$

We introduce Team-FP and independent Team-FP dynamics depending on how agents update their actions. In the former, a single agent can get chosen randomly, as in the classical log-linear learning. In the latter, each agent can update his/her action with probability  $\delta \in (0, 1)$  independent of others, as in the independent log-linear learning. The latter addresses the coordination burden in the update of actions within teams. Algorithm 1 provides descriptions of these dynamics for the typical agent  $i$  from team  $m$ .

*Remark 3.1* (Scalability). Agents can have networked interactions such that their payoff functions depend only on the actions of certain agents such as one/two-hop neighbors. For such cases, agents can form beliefs about these neighbors' strategies only, as if these neighbors play according to some stationary strategy. For example, assume that the payoff of agent  $i \notin \mathcal{I}^m$  (outside team  $m$ ) depends *only* on the actions of team- $m$  members from some neighborhood  $\mathcal{N}^i$ , i.e.,  $\{j : j \in \mathcal{N}^i \cap \mathcal{I}^m\}$ . Agent  $i$  can form a belief about these agents' strategies according to

$$\pi_{k+1}^{im} = \pi_k^{im} + \alpha_k(\underline{a}_k^{im} - \pi_k^{im}) \quad \forall k = 0, 1, \dots \quad (10)$$

where  $\underline{a}_k^{im} = \{a_k^j\}_{j \in \mathcal{N}^i \cap \mathcal{I}^m}$ . Then, the linearity of the update rule yields that the local empirical average  $\pi_k^{im}$  corresponds to the marginalization of  $\pi_k^m$  such that

$$\pi_k^{im}(\{a^j\}_{j \in \mathcal{N}^i \cap \mathcal{I}^m}) = \sum_{\{a^j\}_{j \in \mathcal{I}^m \setminus \mathcal{N}^i}} \pi_k^m(\{a^j\}_{j \in \mathcal{I}^m}) \quad \forall k. \quad (11)$$

Therefore, local observations (within neighborhoods) would still be sufficient to follow Algorithm 1. We demonstrate the scalability of Team-FP by a large-scale experiment in Appendix D, Figure 7.

We focus on the homogeneous cases where agents  $i \notin \mathcal{I}^m$  have a common belief about team  $m$ 's strategy. Homogeneous beliefs are possible if the agents have common initial beliefs and step sizes.

We moved the extension of (Independent) Team-FP to multi-team Markov games and its numerical analysis to Appendix C.

## 4 Convergence Results

Team-FP and Independent Team-FP dynamics reduce, resp., to the classical log-linear learning and independent log-linear learning dynamics if there is only one team. These log-linear learning

dynamics are known to reach team-optimal solution in potential games. For example, consider an  $n$ -agent potential game  $\langle (A^i, u^i)_{i \in [n]} \rangle$  with the potential function  $\phi(\cdot)$ . In both dynamics, the action profiles played form irreducible and aperiodic Markov chains. Let  $\hat{\mu}$  and  $\check{\mu}$  denote the unique stationary distributions, resp., for the classical and independent versions. For the former, we have  $\hat{\mu} = \text{br}_\tau(\phi(\cdot))$  [Marden and Shamma, 2012, Section 3] and (1) yields that

$$0 \leq \max_a \{\phi(a)\} - \phi(\hat{\mu}) \leq \tau \log |A|. \quad (12)$$

On the other hand, the smaller  $\delta > 0$  implies closer stationary distributions in the classical and independent versions. Particularly, we have

$$\|\hat{\mu} - \check{\mu}\|_1 \leq \Lambda(\delta, \epsilon_\phi), \quad (13)$$

for some function  $\Lambda(\cdot)$  decaying to zero as  $\delta \rightarrow 0^+$  for any  $\epsilon_\phi > 0$ , and  $0 < \epsilon_\phi \leq \text{br}_\tau(\phi(\cdot, a^{-i}))$  for any  $a^{-i}$  and  $i$  is a lower bound on local actions get played in the smoothed best response [Donmez et al., 2024, Lemma 5.6].

Team-FP dynamics have similar convergence guarantees for multi-team games under the following assumption on the step sizes used.

**Assumption 4.1.** The step size  $\alpha_k \in [0, 1]$  satisfies the stochastic approximation conditions:  $\alpha_k \rightarrow 0$  as  $k \rightarrow \infty$ ,  $\sum_{k=0}^{\infty} \alpha_k = \infty$ , and  $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$ . Additionally, we have  $\lim_{k \rightarrow \infty} \alpha_k / \alpha_{k+1} = 1$ , and  $\alpha_k - \alpha_{k+1} \geq \alpha_k \alpha_{k+1}$ .

The last condition in Assumption 4.1 ensures that recent observations have comparable weight in the beliefs formed. The classical choice  $\alpha_k = 1/(k+1)$ , leading to the empirical averages of the actions played, satisfies Assumption 4.1.

The following theorem shows the convergence of Algorithm 1 to near TNE almost surely with the approximation levels (similar to (12) and (13)) inherited from the (independent) log-linear learning dynamics.

**Theorem 4.2.** *Given a ZSPTG characterized by  $\langle \mathcal{T}, (A^i, u^i)_{i \in \mathcal{I}} \rangle$ , let every agent follow either Team-FP or Independent Team-FP, as described in Algorithm 1. If Assumption 4.1 holds, then the team-Nash gap for  $\pi_k := (\pi_k^m)_{m \in \mathcal{T}}$  satisfies*

$$\limsup_{k \rightarrow \infty} \text{TNG}(\pi_k) \leq \begin{cases} \tau \log |A| & \text{for Team-FP} \\ \tau \log |A| + |\mathcal{T}|^2 \bar{\phi} \cdot \Lambda(\delta, \epsilon) & \text{for Independent Team-FP} \end{cases} \quad (14)$$

almost surely, where  $\bar{\phi} := \max_{(m,l,a)} |\phi^{ml}(a)|$ .

The key challenge in our convergence analysis is the non-stationarity arising from opponent teams adapting their strategies while the team members learn to coordinate against them. We address this by constructing a *reference scenario* where the team members' beliefs about opponent strategies are *frozen* over finite-length epochs, allowing them to hypothetically "team up" under these fixed beliefs. By comparing the dynamics in the actual scenario and the reference scenario, and exploiting the averaging nature of belief formation, we can bound the approximation error between the two. The main proof concept, along with the Team-FP algorithm for a two-team scenario, is illustrated in Figure 2. This approach is similar to the one used in [Donmez et al., 2024] to handle non-stationarity in Markov team problems. However, unlike [Donmez et al., 2024], we cannot directly show the error bound decay due to the lack of a contraction property in our dynamics.

To overcome this limitation, we view Team-FP as smoothed fictitious play dynamics in zero-sum polymatrix games with an additive bounded error term. The additive error captures the fact that team members may not perfectly achieve team coordination. We then relax the problem by considering any approximation error within the formulated bounds, rather than the actual error. To address this relaxation, we leverage stochastic differential inclusion approximation methods [Benaïm et al., 2005, Perkins and Leslie, 2013]. Finally, by constructing a suitable Lyapunov function addressing arbitrary bounded errors in continuous-time smoothed best response dynamics, we establish the convergence of the discrete-time Team-FP updates.

The following corollary to the main result shows the rationality of the (independent) Team-FP dynamics.

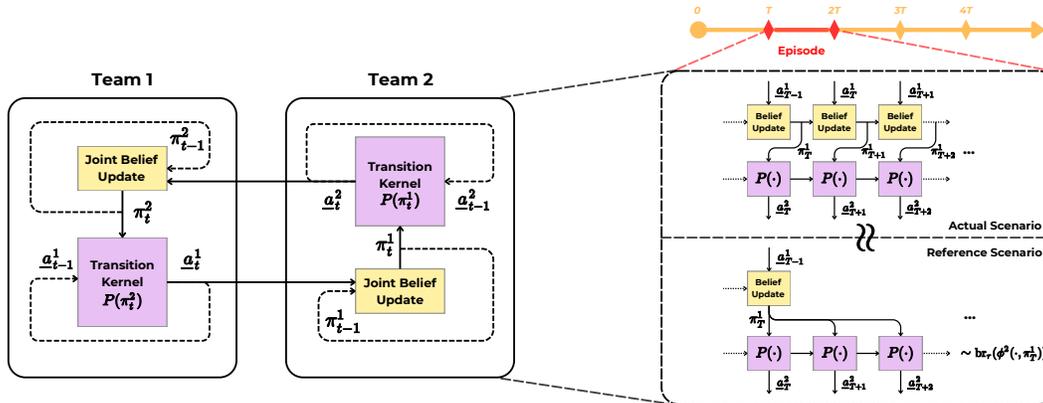


Figure 2: An illustration of Team-FP dynamics for two-team games on the left-hand side. Team actions change according to a transition kernel depending on the beliefs formed about the other teams. Dashed lines represent the time shift. On the right-hand side, we depict the key proof idea that we approximate the evolution of the team actions with a reference scenario where beliefs are stationary such that team actions form a homogeneous Markov chain whose unique stationary distribution corresponds to the best team response.

**Corollary 4.3.** *Given a ZSPTG characterized by  $\langle \mathcal{T}, (A^i, u^i)_{i \in \mathcal{I}} \rangle$ , let agents from team  $m$  follow either Team-FP or Independent Team-FP while other teams play according to some stationary strategy  $\pi^{-m}$ . If Assumption 4.1 holds, then empirical average of the action profiles played by team  $m$  satisfy*

$$\limsup_{k \rightarrow \infty} \text{TNG}^m(\pi_k^m, \pi^{-m}) \leq \begin{cases} \tau \log |A^m| & \text{for team-FP} \\ \tau \log |A^m| + \bar{\phi} \cdot \Lambda(\delta, \epsilon) & \text{for Independent Team-FP} \end{cases} \quad (15)$$

almost surely.

Theorem 4.2 can be generalized to the case where the rewards are random with bounded support, rather straightforwardly. Therefore, the proof of Corollary 4.3 follows from Theorem 4.2 if we view the underlying game as there is a single team receiving random rewards with bounded support.

## 5 Illustrative Examples

In this section, we present various simulation results demonstrating the coordination speed of Team-FP and compare it to pure FP, no-regret algorithms, and a stationary opponent. We also observe the effect of parameters on the convergence speed. In addition, we examine the long-run behavior of team-FP in games beyond ZSPTG games, where we intuitively expect it to converge. All simulations are averaged over 10 independent trials to reduce the randomness. In all figures, the mean is plotted with a thick colorful line, while one standard deviation below and above the mean is shown with a shaded area of the same color. Also, for all simulations, temperature parameter  $\tau$  is chosen to be 0.1 unless another option is mentioned. We conduct simulations for ZSPTG with two different setups: one with three teams, each consisting of three agents, and another with two teams, each consisting of four agents. Unless explicitly stated otherwise, the default setting consists of two teams, with four agents in each team. The step size is chosen to be  $\alpha_k = 1/(k + 1)$  for all simulations.

All the simulations are executed on a computer equipped with an Intel Xeon W7-3455 CPU and 128 GB RAM. Run-time for 10 independent trials over  $10^6$  iterations vary between 1-5 hours depending on the experiment.

**Achieving Implicit Coordination in Team-FP** In this section, we compare the performance of Team-FP to the explicit coordination of team members. We also compare Team-FP to algorithms such as Multiplicative Weights Update (MWU) and Smoothed FP (SFP), and show that these algorithms fail to achieve team coordination. We also show that Team-FP achieves near-optimum policy against a stationary opponent as stated in Corollary 4.3.

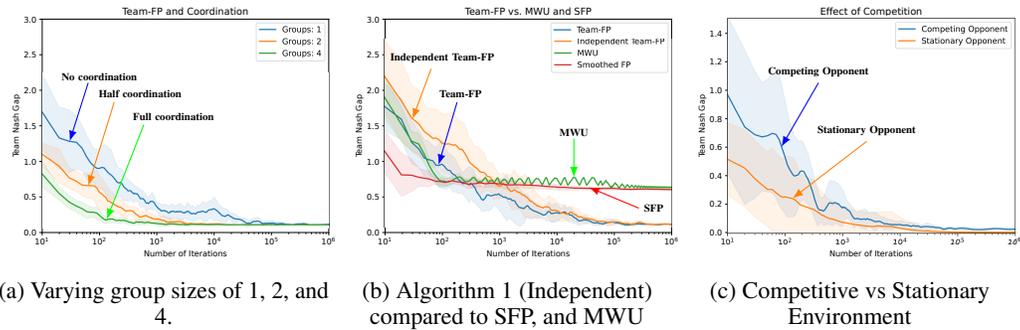


Figure 3: All the above figures show the variation of TNG over time. (a) Comparison of different levels of explicit coordination for Team-FP: independent agents (group size 1), pairs of cooperating agents (group size 2), and fully coordinated teams (group size 4). (b) Performance of Team-FP and Independent Team-FP compared to MWU and SFP algorithms in a 2-team ZSPTG. (c) Convergence of Team-FP against stationary and competitive opponents in a 3-team ZSPTG.

*Impact of Team Size in Team Coordination:* In Team-FP self-interested team members act together without explicit communication in a coordinated way to reach TNE. We can measure how this independent cooperation compares to the explicit coordination of team members. For that, we propose an example game with two teams and four agents in each team. For the first scenario, four agents act separately and use Team-FP. In the second scenario, the agents form groups of two and act in a coordinated way as if they are a single agent, using Team-FP. In the final scenario, all agents in a team behave as if they were a single agent, equivalent to the standard fictitious play (FP) dynamics in a two-person zero-sum game. This case mimics the ex-ante communication scheme from [Farina et al., 2018]. The scenarios are described by group sizes. Group sizes are one, two, and four respectively for these scenarios. The simulation results are presented in Figure 3a. As expected, the explicit coordination of all members in a team converges the fastest, followed by the explicit coordination of groups of two, and finally, Team-FP converges slowly as the agents do not coordinate explicitly.

*Team-FP Compared to MWU and SFP:* The strong side of Team-FP is, even though the agents do not communicate, the average of joint actions of teams can reach TNE, unlike other algorithms. We compare the equilibrium behavior of team-FP with a well-known no-regret learning algorithm Multiplicative Weights Update (MWU), in which the average strategies converge to NE in zero-sum polymatrix games [Cai et al., 2016]. We also compare Team-FP with the usual SFP, where each agent holds beliefs about other agents and uses the smoothed best response against them. In Figure 3b, we see that both Team-FP and Independent Team-FP dynamics converge to near TNE, while the other algorithms fail to do so.

*Rationality Against Stationary Opponent:* In this part, we use 3-team setting where each team has 3 agents. We let a team using Team-FP compete against two other stationary teams and compared it with the performance of the same team in the same game when the opponents are also using Team-FP competitively. In Figure 3c, we observe that both algorithms converge, while the convergence is much faster against stationary opponents.

**Team-FP in Application** In this part, we provide an example to demonstrate that Team-FP has applications in various contexts.

*Security Game Example:* We model an airport security scenario as a two-team game between defender and attacker teams. In our example, a security chief on the defender team faces three independent intruders on the attacker team, as shown in Figure 4a. The chief can defend a gate at a cost, while intruders decide whether to attack a gate or remain idle. Intruders gain or lose payoffs based on whether they attack undefended or defended gates, and the chief's payoffs mirror these outcomes. We conducted multiple trials and presented the evolution of the average Team-Nash Gap with standard deviations on the right side of Figure 4b. From a higher level, this example shows that team-minimax equilibrium can predict the outcome of games with different uncoordinated attackers. It also justifies the algorithms developed to compute team-minimax equilibrium efficiently.

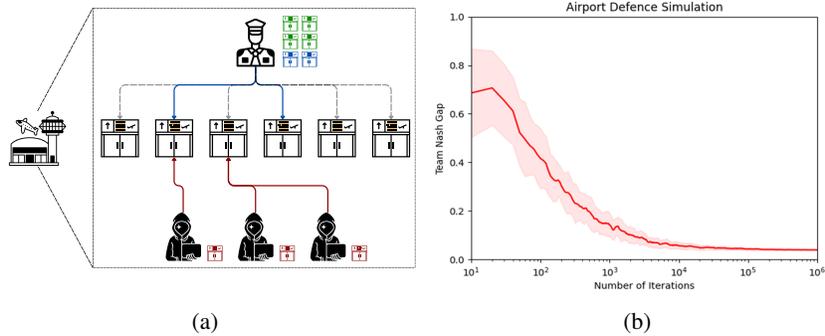


Figure 4: (a) The illustration of an airport security game: a security chief guarding the six gates of an airport against three different intruders making decisions autonomously. (b) The evolution of Team Nash Gap in airport security game, showing that Team-FP dynamics reach near team-minimax equilibrium.

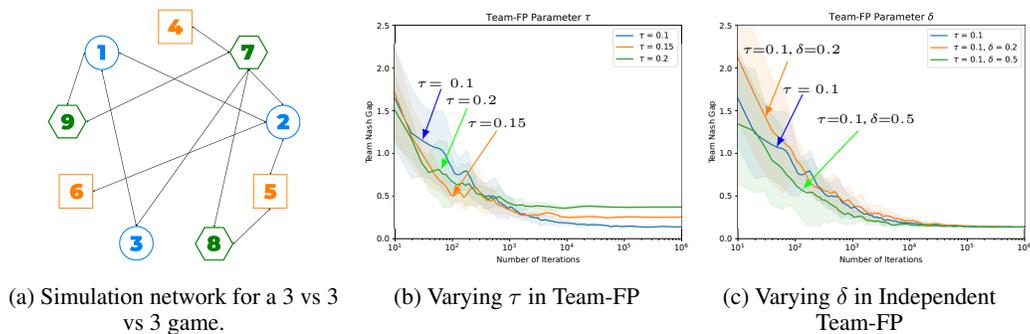


Figure 5: The 3-team experiments are tested on the randomly generated network structure (a). The other figures (b), and (c), shows the variation of TNG over iterations. (a) The simulation network for a multi-team ZSPTG, in which there are 3-teams with 3 agents in each team. (b) The impact of varying temperature parameter  $\tau$  (0.1, 0.15, 0.2) in Algorithm 1 on the closeness to TNE. (c) The effect of different  $\delta$  values (0.2, 0.5) in (Independent) Algorithm 1 on the convergence speed with  $\tau$  fixed at 0.1

**Effect of parameters in Team-FP** In this part, we examine how Team-FP performs for different  $\tau$  and  $\delta$  values. Given an example ZSPTG of three teams where each team has three agents in Figure 6a, we examine the evolution of TNG in the Team-FP dynamics for different values of  $\tau \in \{0.1, 0.15, 0.2\}$ . We also compare the evolution of NG in the Team-FP and Independent Team-FP dynamics for  $\tau = 0.1, \delta = 0.2, \delta = 0.5$ . All simulation results (see Figure 5) show convergence, and we observe lower final values of  $NG(\pi)$  for smaller  $\tau$  as we predicted (see. Figure 5b). The Independent Team-FP requires more iterations when  $\delta = 0.2$  for its convergence, while it is much faster when  $\delta = 0.5$  (see. Figure 5c). This is expected as updates are much more frequent as  $\delta$  increases. However, increasing  $\delta$  too much may harm the coordination behavior of the team.

**Beyond ZSPTG** In this part, we try several other games for Team-FP without proof of convergence. We expect Team-FP to converge in  $2 \times N$  and potential games other than zero-sum games as FP converges in these settings. For the first case, we try a game where one team has a single agent with only two actions with random rewards, while the other team has three agents with an underlying potential function. In this case, Team-FP converges (see. Figure 6b). In another setting, we try two teams of four agents. However, the potential functions for both teams are identical rather than summing to zero, resulting in a potential function that encompasses the individual potential functions. The team-FP algorithm converges to TNE in this case as well (see. Figure 6c). However, the equilibrium is not unique in this case. Finally, we create an extension of the team-FP for Markov games (see. Appendix C) in an RL framework and try simulations on this setting. In this case, there

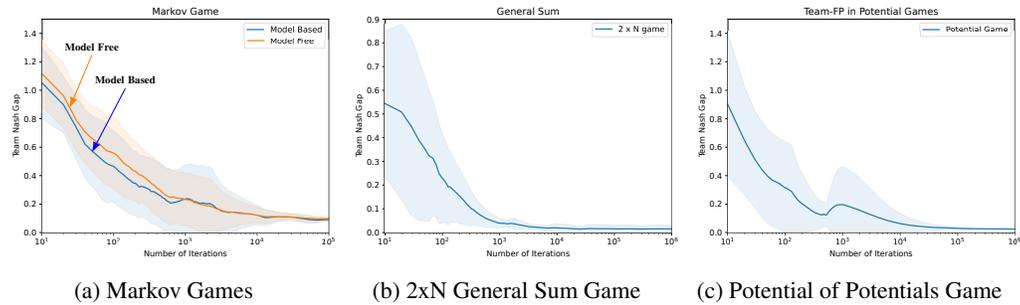


Figure 6: All the above figures describes the variation of TNG over iterations for Algorithms that are related to but outside the scope of ZSPTG. (a) The model-free and model-based games of Algorithm 2, and 3, for a game of 2-team each with 2 agents, with 2 states and 10 horizon length. (b) The behavior of Team-FP dynamics in a 2xN general sum game, where a team competes against a single agent with random rewards. (c) The behavior of Team-FP dynamics in a potential game over the underlying potential functions.

are two teams each having two agents, competing against each other in a finite horizon Markov game with a horizon length of ten. The number of states is two, and the state transition matrices are generated randomly. We observe that team-Nash Gap for MG's defined in (65), converges to near-zero.

## 6 Conclusion

In this paper, we introduced Team-FP, a novel fictitious play variant designed for multi-team games. We showed that Team-FP provably achieves near-TNE in ZSPTGs, with a quantifiable error bound. Our convergence analysis addressed the non-stationarity challenge arising from evolving opponent team strategies, by leveraging the optimal coupling lemma and stochastic differential inclusion approximation methods. We also extended Team-FP to multi-team Markov games, encompassing both model-based and model-free scenarios, with applications in multi-team reinforcement learning. Furthermore, we conducted detailed numerical analysis of the Team-FP dynamics, focusing on the trade-off in learning to team up and competition in comparison to the classical FP and no-regret learning dynamics. We further examined the convergence of Team-FP dynamics to TNE in multi-team games beyond ZSPTGs. These results strengthen the theoretical foundations for applying TNE to predict decentralized team behavior and provide a framework for team learning in multi-team settings.

**Limitations and Broader Impacts** Our work quantified the almost sure convergence of the Team-FP dynamics asymptotically yet did not provide guarantees for the convergence rate. We conducted detailed numerical examples and presented the evolution of the gap in TNE to exemplify this rate qualitatively. The challenge for the non-asymptotic analysis is inherited from the *discrete-time* FP dynamics for which there are only rough rate analysis [Karlin, 1959] and negative examples for some edge cases [Brandt et al., 2010, Daskalakis and Pan, 2014].

Our work introduces no new ethical concerns in multi-team systems and shares the assumption of stationary opponents with learning dynamics like Q-learning and fictitious play. This assumption does not disadvantage our approach. We argue that treating uncoordinated attackers as a single decision-maker is necessary, as they can learn to bypass security measures. Our paper provides a theoretical basis for this, ensuring more reliable AI-based solutions.

**Future Research Directions** This work paves the way to further explore the behavior of decentralized teams in multi-team interactions when the team members follow different types of dynamics for teaming up within teams (other than log-linear learning) and adapting to other teams' play (other than FP). Numerical examples we conducted for multi-team games beyond ZSPTGs are also promising to show the provable convergence of Team-FP dynamics in multi-team (Markov) games reducing to games with the *fictitious play property* (e.g., see [Monderer and Shapley, 1996b]) if the teams coordinate in acting as a single player.

## Acknowledgements

This work was supported by The Scientific and Technological Research Council of Türkiye (TUBITAK) BIDEB 2232-B International Fellowship for Early Stage Researchers under Grant Number 121C124.

## References

- G. Arslan, M. F. Demirkol, and Y. Song. Equilibrium efficiency improvement in MIMO interference systems: A decentralized stream control approach. *IEEE Transactions on Wireless Communications*, 6(8):2984–2993, 2007.
- M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. *SIAM J. Control Optim*, 44(1):328–348, 2005.
- L. M. Bergman and I. N. Fokin. On separable non-cooperative zero-sum games. *Optimization*, 44(1): 69–84, 1998.
- L. E. Blume. The statistical mechanics of strategic interaction. *Games Econom. Behav.*, 5(3):387–424, 1993.
- F. Brandt, F. Fischer, and P. Harrenstein. On the rate of convergence of fictitious play. In *Algorithmic Game Theory: Third International Symposium, SAGT 2010, Athens, Greece, October 18–20, 2010. Proceedings 3*, pages 102–113. Springer, 2010.
- Y. Cai and C. Daskalakis. On minmax theorems for multiplayer games. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 217–234. SIAM, 2011.
- Y. Cai, O. Candogan, C. Daskalakis, and C. Papadimitriou. Zero-sum polymatrix games: A generalization of minmax. *Math. Oper. Res.*, 41(2):648–655, 2016.
- A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, S. Sastry, et al. Challenges for securing cyber physical systems. In *Workshop on Future Directions in Cyber-physical Systems Security*, volume 5, page 7, 2009.
- L. Carminati, F. Cacciamani, M. Ciccone, and N. Gatti. A marriage between adversarial team games and 2-player games: Enabling abstractions, no-regret learning, and subgame solving. In *Int. Conf. Machine Learn. (ICML)*, pages 2638–2657. PMLR, 2022.
- A. Celli and N. Gatti. Computational results for extensive-form adversarial team games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- C. Daskalakis and Q. Pan. A counter-example to Karlin’s strong conjecture for fictitious play. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 11–20. IEEE, 2014.
- A. S. Donmez, O. Unlu, and M. O. Sayin. Logit-Q dynamics for efficient learning in stochastic teams. *arXiv preprint arXiv:2302.09806*, 2024.
- C. Ewerhart and K. Valkanova. Fictitious play in networks. *Games Econom. Behav.*, 123:182–206, 2020.
- G. Farina, A. Celli, N. Gatti, and T. Sandholm. Ex ante coordination and collusion in zero-sum multi-player extensive-form games. *Advances in Neural Inform. Process. (NeurIPS)*, 31, 2018.
- D. Fudenberg and D. M. Kreps. Learning mixed equilibria. *Games Econom. Behav.*, 5(3):320–367, 1993.
- D. Fudenberg and D. K. Levine. Learning and equilibrium. *Annual Rev. Econ.*, 1(1):385–420, 2009.
- S. Hart and A. Mas-Colell. Uncoupled dynamics do not lead to Nash equilibrium. *The American Econ. Rev.*, 83(5):1830–1836, 2003.

- J. Hofbauer and W. H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.
- J. Hu and M. P. Wellman. Nash Q-learning for general-sum stochastic games. *J. Machine Learn. Res.*, pages 1039–1069, 2003.
- M. Jaderberg, W. M. Czarnecki, I. Dunning, L. Marris, G. Lever, A. G. Castaneda, C. Beattie, N. C. Rabinowitz, A. S. Morcos, A. Ruderman, et al. Human-level performance in 3d multiplayer games with population-based reinforcement learning. *Science*, 364(6443):859–865, 2019.
- F. Kalogiannis, I. Panageas, and E-V. Vlatakis-Gkaragkounis. Towards convergence to nash equilibria in two-team zero-sum games. In *Internat. Conf. Learning Representations (ICLR)*, 2022.
- S. Karlin. Mathematical methods and theory in games. *Programming, and Economics*, 1,2, 1959.
- H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, E. Osawa, and H. Matsubara. Robocup: A challenge problem for AI. *AI magazine*, 18(1):73–73, 1997.
- D. A. Levin and Y. Peres. *Markov Chains and Mixing Times*. American Mathematical Society, 2nd edition, 2017.
- M. L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Int. Conf. Machine Learn. (ICML)*, 1994.
- J. R. Marden and J. S. Shamma. Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation. *Games Econom. Behav.*, 75(2):788–808, 2012.
- D. Monderer and L. S. Shapley. Potential games. *Games Econom. Behav.*, 14(1):124–143, 1996a.
- D. Monderer and L. S. Shapley. Fictitious play property for games with identical interests. *Journal of economic theory*, 68(1):258–265, 1996b.
- C. Park, K. Zhang, and A. Ozdaglar. Multi-player zero-sum Markov games with networked separable interactions. In *Advances in Neural Inform. Process. (NeurIPS)*, pages 37354–37369, 2023.
- S. Perkins and D. S. Leslie. Asynchronous stochastic approximation with differential inclusions. *Stochastic Systems*, 2(2):409–446, 2013.
- L. S. Shapley. Stochastic games. *Proc. Natl. Acad. Sci. USA*, 39(10):1095–1100, 1953.
- V. N. Silva and L. Chaimowicz. MOBA: A new arena for game AI. *arXiv preprint arXiv:1705.10443*, 2017.
- T. Tatarenko. *Game-theoretic Learning and Distributed Optimization in Memoryless Multi-agent Systems*. Springer, 2017.
- T. Tatarenko. Independent log-linear learning in potential games with continuous actions. *IEEE Trans. on Control of Network Systems*, 5(3):913–923, 2018.
- O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.
- B. Von Stengel and D. Koller. Team-maxmin equilibria. *Games Econom. Behav.*, 21(1-2):309–321, 1997.
- Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y. D. Yao. Opportunistic spectrum access in cognitive radio networks: Global optimization using local interaction games. *IEEE Journal of Selected Topics in Signal Processing*, 6(2):180–194, 2012.
- B. Zhang, L. Carminati, F. Cacciamani, G. Farina, P. Olivieri, N. Gatti, and T. Sandholm. Subgame solving in adversarial team games. *Advances in Neural Inform. Process. (NeurIPS)*, 35:26686–26697, 2022.

- K. Zhang, Z. Yang, H. Liu, T. Zhang, and T. Başar. Finite-sample analysis of decentralized batch multiagent reinforcement learning with networked agents. *IEEE Trans. on Automatic Control*, 66(12):5925–5940, 2021.
- J. Zheng, Y. Cai, Y. Liu, Y. Xu, B. Duan, and X. Shen. Optimal power allocation and user scheduling in multicell networks: Base station cooperation using a game-theoretic approach. *IEEE Transactions on Wireless Communications*, 13(12):6928–6942, 2014.

## A Proof of Theorem 4.2

We can separate the proof into two main steps: (i) showing that team members can learn to team up approximately by constructing a reference scenario where beliefs got frozen, and (ii) addressing the bounded approximation error by leveraging the stochastic differential inclusion approximations.

### A.1 Step (i) - Reference Scenario and Error Analysis

We divide the horizon into  $T$ -length epochs. Then, we can write the belief update (8) accumulated from  $k = nT$  to  $(n + 1)T$  as

$$\pi_{(n+1)T}^m = \left( \sum_{k=nT}^{(n+1)T-1} (1 - \alpha_k) \right) \cdot \pi_{nT}^m + \sum_{k=nT}^{(n+1)T-1} \alpha_k \left( \prod_{\ell=k+1}^{(n+1)T-1} (1 - \alpha_\ell) \right) \cdot \underline{a}_k^m \quad (16)$$

for any  $n = 0, 1, \dots$  denoting the epoch index. Let  $\pi_{(n)}^m := \pi_{nT}^m$  denote the belief about team  $m$  in epoch  $n$ . Furthermore, for notational simplicity, we define

$$\beta_k := \alpha_k \prod_{\ell=k+1}^{(n+1)T-1} (1 - \alpha_\ell) \quad \text{and} \quad \beta_{(n)} := \sum_{k=nT}^{(n+1)T-1} \beta_k. \quad (17)$$

Then, we can simplify (16) as

$$\pi_{(n+1)}^m = (1 - \beta_{(n)}) \cdot \pi_{(n)}^m + \beta_{(n)} \left( \sum_{k=nT}^{(n+1)T-1} \frac{\beta_k}{\beta_{(n)}} \cdot \underline{a}_k^m \right). \quad (18)$$

Due to 4.1, the step size  $\beta_{(n)}$  decays to zero monotonically at a certain rate such that [Donmez et al., 2024, Lemma 5.4]

$$\sum_{n=0}^{\infty} \beta_{(n)} = \infty \quad \text{and} \quad \sum_{n=0}^{\infty} \beta_{(n)}^2 < \infty. \quad (19)$$

Let  $\mathcal{F}_{(n)}$  be a filtration generated by the  $\sigma$ -algebra  $\sigma(A_0, \dots, A_{nT-1})$ , where  $A_t$  denotes the joint actions taken by each team at stage  $t$ , i.e.,  $A_t = (\underline{a}_t^1, \dots, \underline{a}_t^{|T|})$ . It is instructive to note that  $\pi_{(n)}^m$  is  $\mathcal{F}_{(n)}$ -measurable. We define the joint action distributions of team  $m$  at time  $k$  in epoch  $n$  by

$$\mu_{(n),k}^m := \mathbb{E}[\underline{a}_k^m \mid \mathcal{F}_{(n)}]. \quad (20)$$

Then, we can write (18) as in the form of stochastic approximation

$$\pi_{(n+1)}^m = (1 - \beta_{(n)})\pi_{(n)}^m + \beta_{(n)} \left( \sum_{k=nT}^{(n+1)T-1} \frac{\beta_k}{\beta_{(n)}} \cdot \mu_{(n),k}^m + \omega_{(n+1)}^m \right), \quad (21)$$

where  $\omega_{(n+1)}^m$  is a Martingale difference sequence defined by

$$\omega_{(n+1)}^m := \sum_{k=nT}^{(n+1)T-1} \frac{\beta_k}{\beta_{(n)}} \left( \underline{a}_k^m - \mu_{(n),k}^m \right). \quad (22)$$

Now, let's consider a *reference scenario* for the analysis in which the beliefs (denoted by  $\hat{\pi}_t^m$ ) are only updated at the ends of  $T$ -length epochs. In other words, we have  $\hat{\pi}_t^m = \pi_{(n)}^m$  for all  $nT \leq t \leq (n + 1)T - 1$  and  $m \in \mathcal{T}$ . Due to the fixed beliefs about the opponent plays, Team-FP dynamics reduce to the log-linear learning in the reference scenario. Let  $\hat{a}_{(n),k}^m$  denote the joint action of team  $m$  in the reference scenario. By the nature of the log-linear learning,  $\{\hat{a}_{(n),k}^m\}_{k=nT}^{\infty}$  form a homogeneous Markov chain (MC) even though the actual action profiles  $\{a_k^m\}_{k=nT}^{\infty}$  do not necessarily do so. Denote the stationary distribution of the MC in the reference scenario by  $\tilde{\mu}_{(n),*}^m$  and  $\hat{\mu}_{(n),*}^m$  for Team-FP and Independent Team-FP, respectively. The former is given by

$\tilde{\mu}_{(n),\star}^m = \text{br}_\tau(\phi^m(\cdot, \pi_{(n)}^{-m}))$  due to the log-linear learning update [Blume, 1993, Marden and Shamma, 2012]. Then, we can write the belief update (21) for team  $m$  as

$$\pi_{(n+1)}^m = (1 - \beta_{(n)})\pi_{(n)}^m + \beta_{(n)} \left( \text{br}_\tau(\phi^m(\cdot, \pi_{(n)}^{-m})) + \omega_{(n+1)}^m + e_{(n)}^m \right), \quad (23)$$

where we decompose the error  $e_{(n)}^m := \hat{e}_{(n)}^m + \check{e}_{(n)}^m$  as

$$\hat{e}_{(n)}^m := \sum_{k=nT}^{(n+1)T-1} \frac{\beta_k}{\beta_{(n)}} (\mu_{(n),k}^m - \hat{\mu}_{(n),\star}^m) \quad (24)$$

$$\check{e}_{(n)}^m := \hat{\mu}_{(n),\star}^m - \tilde{\mu}_{(n),\star}^m. \quad (25)$$

The update (23) corresponds to the SFP dynamics of teams acting as a single decision-maker with the additive error  $e_{(n)}^m$ .

The following lemma, based on designing the optimal coupling between the actual and reference scenarios as in [Donmez et al., 2024, Lemma 5.5], plays an important role in bounding  $\|\hat{e}_{(n)}^m\|$  from above.

**Lemma A.1.** *For some constants  $c, d, \rho \geq 0$ , we have*

$$\|\mu_{(n),k}^m - \hat{\mu}_{(n),\star}^m\|_1 \leq c \rho^{k-nT} + dT\alpha_{nT} \quad \forall m \in \mathcal{T}. \quad (26)$$

By (24), Lemma A.1 yields that we can bound the error  $\hat{e}_{(n)}^m$  from above by

$$\|\hat{e}_{(n)}^m\|_1 \leq c \sum_{k=nT}^{(n+1)T-1} \frac{\beta_k}{\beta_{(n)}} \rho^{k-nT} + dT\alpha_{nT}. \quad (27)$$

Due to the no-recency-bias condition in 4.1, we have  $\beta_{k+1}/\beta_k \leq 1$  by the definition (17). Then, we have monotonically decaying  $\beta_k$ , which yields

$$\frac{\beta_k}{\beta_{(n)}} \leq \frac{\beta_{nT}}{T\beta_{(n+1)T-1}} \leq \frac{\alpha_{nT}}{T\alpha_{(n+1)T}}. \quad (28)$$

By the assumption 4.1, we have

$$\lim_{n \rightarrow \infty} \frac{\alpha_{nT}}{T\alpha_{(n+1)T}} = \frac{1}{T} \lim_{n \rightarrow \infty} \prod_{k=nT}^{(n+1)T-1} \frac{\alpha_k}{\alpha_{k+1}} = \frac{1}{T}. \quad (29)$$

By (27), (28), (29), and the decaying property of  $\alpha_k$ , we obtain

$$\limsup_{n \rightarrow \infty} \|\hat{e}_{(n)}^m\|_1 \leq \frac{c}{T} \frac{1}{1-\rho} \quad \forall m \in \mathcal{T}. \quad (30)$$

On the other hand,  $\check{e}_{(n)}^m$  is non-zero only for Independent Team-FP and corresponds to the difference between the stationary distributions for the classical and independent log-linear learning. We can bound  $\check{e}_{(n)}^m \leq \Lambda(\delta, \epsilon)$  for some  $\epsilon > 0$  based on (13). Hence, given any  $\epsilon > 0$ , there exists  $N_\epsilon$  such that

$$\|e_{(n)}^m\|_1 < C(\epsilon, T) \quad \forall n \geq N_\epsilon, \quad (31)$$

where

$$C(\epsilon, T) := \begin{cases} \epsilon + \frac{c}{T} \frac{1}{1-\rho} & \text{for Team-FP} \\ \epsilon + \frac{c}{T} \frac{1}{1-\rho} + \Lambda(\delta, \epsilon) & \text{for Independent Team-FP} \end{cases} \quad (32)$$

Note that  $C(\epsilon, T)$  can become arbitrarily close to  $\Lambda(\delta, \epsilon)$  for sufficiently large  $T$  and small  $\epsilon$ , which are chosen arbitrarily just for the analysis.

## A.2 Step (ii) - Convergence Analysis with Relaxed Errors

We focus on the convergence analysis of (23) based on the bound (31). To this end, we define the set-valued mapping

$$F(\pi) := \left\{ \left( \text{br}_\tau(\phi^m(\cdot, \pi^{-m})) - \pi^m + e^m \right)_{m \in \mathcal{T}} : \right. \\ \left. \|e^m\|_1 \leq C(\varepsilon, T) \text{ and } \text{br}_\tau(\phi^m(\cdot, \pi^{-m})) + e^m \in \Delta(\underline{A}^m) \forall m \right\} \quad (33)$$

for all  $\pi \in \Pi := \prod_{m \in \mathcal{T}} \Delta(\underline{A}^m)$ . Then, the update (23) and (31) yield that for sufficiently large  $n$ , the empirical averages  $\{\pi_{(n)}\}_{n \geq 0}$  satisfy

$$\pi_{(n+1)} - \pi_{(n)} - \beta_{(n)} \cdot \omega_{(n+1)} \in \beta_{(n)} \cdot F(\pi_{(n)}), \quad (34)$$

where the Martingale difference sequence  $\{\omega_{(n+1)}\}$  is as described in (22). We have set inclusion in (34) different from the equality in (23) since we relax the update rule by considering any error satisfying the bound (31), rather than the actual error.

The following proposition shows that  $F(\cdot)$  is a peculiar set-valued mapping. Particularly, given the sets  $X, Y$  from some underlying Euclidean space, we say that a set-valued function  $\bar{F}(\cdot)$  mapping each point  $x \in X$  to a set  $\bar{F}(x) \subset Y$  is a *Marchaud map*, e.g., see [Perkins and Leslie, 2013, Definition 2.1] and [Benaïm et al., 2005, Hypothesis 1.1], if it satisfies the following conditions:

- (i)  $\bar{F}(\cdot)$  is upper semi-continuous, or equivalently,  $\text{Graph}(\bar{F}) = \{(x, y) : y \in \bar{F}(x)\}$  is a closed subset of  $X \times Y$ .
- (ii) For all  $x \in X$ , the set  $\bar{F}(x)$  is a non-empty, compact, and convex subset of  $Y$ .
- (iii) There exists a  $c > 0$  such that  $\sup_{y \in \bar{F}(x)} \|y\| \leq c(1 + \|x\|)$  for all  $x \in X$ .

**Proposition A.2.** *The set-valued function  $F(\cdot)$  is a Marchaud map.*

Next, we approximate the relaxation (34) by the differential inclusion

$$\dot{\pi} \in F(\pi). \quad (35)$$

Particularly, due to Proposition A.2 and (19), a linear interpolation of the iterative process  $\{\pi_{(n)}\}_{n \geq 0}$  is a perturbed solution to the differential inclusion (35) [Benaïm et al., 2005, Proposition 1.4] and its limit set is internally chain transitive [Benaïm et al., 2005, Theorem 3.6]. Furthermore, we can characterize internally chain transitive sets (and therefore, the limit set of linear interpolations) through a Lyapunov function. Particularly, let  $\Phi_t(\pi)$  be the set of solutions to (35) with initial point  $\pi$ . We say that a continuous function  $\bar{V} : \Pi \rightarrow \mathbb{R}$  is a *Lyapunov function* of (35) for a subset  $\Lambda \subset \Pi$  provided that

- $\bar{V}(\tilde{\pi}) < \bar{V}(\pi)$  for all  $\pi \in \Pi \setminus \Lambda$ ,  $\tilde{\pi} \in \Phi_t(\pi)$ ,  $t > 0$ ,
- $\bar{V}(\tilde{\pi}) \leq \bar{V}(\pi)$  for all  $\pi \in \Lambda$ ,  $\tilde{\pi} \in \Phi_t(\pi)$ ,  $t \geq 0$ .

Given such a Lyapunov function for  $\Lambda$ , every chain internally chain transitive set  $L$  is contained in  $\Lambda$  if the set  $\{\bar{V}(\pi) : \pi \in \Lambda\}$  has empty interior [Benaïm et al., 2005, Proposition 3.27]. Therefore, we propose the candidate Lyapunov function

$$V(\pi) = \max \{0, L(\pi) - \Xi(\lambda, \varepsilon, T)\}, \quad (36)$$

where the auxiliary terms are defined by

$$L(\pi) := \sum_{m \in \mathcal{T}} \max_{\mu \in \Delta(\underline{A}^m)} \{ \phi^m(\mu, \pi^{-m}) + \tau \mathcal{H}(\mu) \} \quad (37a)$$

$$\Xi(\lambda, \varepsilon, T) := \tau \log |A| + \lambda |\mathcal{T}|^2 \bar{\phi} \cdot C(\varepsilon, T) > 0 \quad (37b)$$

for some arbitrary  $\lambda > 1$  to characterize the long-run behavior of (34) based on (35).

**Proposition A.3.** *The continuous function  $V(\cdot)$  is a Lyapunov function of (35) for the set  $\Lambda = \{\pi : V(\pi) = 0\}$ .*

Proposition A.3 yields that there exists some sequence  $\{\zeta_k \in \mathbb{R}\}_{k \geq 0}$  such that  $|\zeta_k| \rightarrow 0$  as  $k \rightarrow \infty$  almost surely and

$$L(\pi_k) < \Xi(\lambda, \varepsilon, T) + \zeta_k \quad \forall k. \quad (38)$$

By (3), we can write  $L(\pi_k)$  as

$$\begin{aligned} L(\pi_k) &= \sum_{m \in \mathcal{T}} \left( \max_{\mu \in \Delta(\underline{A}^m)} \{ \phi^m(\mu, \pi_k^{-m}) + \tau \mathcal{H}(\mu) \} - \phi^m(\pi_k) \right), \\ &\geq \sum_{m \in \mathcal{T}} \left( \max_{\mu \in \Delta(\underline{A}^m)} \{ \phi^m(\mu, \pi_k^{-m}) \} - \phi^m(\pi_k) \right). \end{aligned} \quad (39)$$

Since  $\max_{\mu} \{ \phi^m(\mu, \pi^{-m}) \} - \phi^m(\pi) \geq 0$  for all  $\pi$ , the inequalities (38) and (39) yield that

$$0 \leq \sum_{m \in \mathcal{T}} \left( \max_{\mu \in \Delta(\underline{A}^m)} \{ \phi^m(\mu, \pi_k^{-m}) \} - \phi^m(\pi_k) \right) \leq \Xi(\lambda, \varepsilon, T) + \zeta_k \quad \forall m \text{ and } n. \quad (40)$$

Recall that we can select  $\lambda > 1$ ,  $\varepsilon > 0$  and  $T \in \mathbb{N}$  arbitrarily. The definitions (32) and (37b) yield that  $\Xi(\lambda, \varepsilon, T)$  can be arbitrarily close to  $\tau \log |A|$  for Team-FP and  $\tau \log |A| + |\mathcal{T}|^2 \bar{\phi} \cdot \Lambda(\delta, \epsilon)$  for Independent Team-FP. Furthermore,  $\zeta_k$  decays to zero. Therefore, we can obtain (14) based on (40) and Definition 2.4. This completes the proof.

## B Proofs of Technical Results

The followings are the proofs of Lemma A.1, and Propositions A.2 and A.3 used in Appendix A.

### B.1 Proof of Lemma A.1

We define two discrete-time processes from the beginning of epoch  $n$ . One of them is the joint actions of teams in the reference scenario, defined as  $\{\hat{\omega}_k\}_{k \geq nT} := \{(\hat{\underline{a}}_k^m)_{m \in \mathcal{T}} | \mathcal{F}_{(n)}\}_{k \geq nT}$ , and the other one is the joint actions of teams in the actual scenario, with  $\{\omega_k\}_{k \geq nT} := \{(\underline{a}_k^m)_{m \in \mathcal{T}} | \mathcal{F}_{(n)}\}_{k \geq nT}$ .

The transition probabilities between states depend only on the beliefs on the other teams. Therefore, for the reference scenario, during an epoch, this process is a homogeneous MC. Let,  $P_{(n),k}$  be the transition probabilities between the states, i.e., joint action profiles of all teams, of the original discrete-time process at step  $k$ , and let  $\hat{P}_{(n)}$  be the transition probabilities between the states for the reference scenario. Let  $\hat{\mu}_{(n),k}$  be the distribution of  $\hat{\omega}_k$ , and let  $\mu_{(n),k}$  be the distribution of  $\omega_k$ . In this case, the expected joint actions of a team at step  $k$  within an epoch in real and fictional scenarios can be expressed in terms of the distributions as:

$$\mu_{(n),k}^m(a) = \sum_{\omega: \{a^m=a\}} \mu_{(n),k}(\omega), \quad \hat{\mu}_{(n),k}^m(a) = \sum_{\hat{\omega}: \{a^m=a\}} \hat{\mu}_{(n),k}(\hat{\omega}). \quad (41)$$

**Step 1.** In classical log-linear learning, at each step, only one agent can change their action and due to the soft-max nature of this action change, any action has positive probability which is bounded from below. This bound only depends on the minimum and maximum values of any agents' reward, which are defined by the game and bounded by definition, and the temperature parameter. If we divide that bound by the number of agents in the team ( $|\mathcal{I}^m|$ ), we can obtain a lower bound on the probability of changing to any joint action which can be reached within a single step. Let's call that bound  $\xi$ . Then, for any state  $\omega$  or  $\hat{\omega}$ , there is a  $\xi > 0$  such that

$$P_{(n),k}(\omega | \omega_k) > \xi^{|\mathcal{T}|} \iff P_{(n),k}(\omega | \omega_k) > 0, \quad (42)$$

$$\hat{P}_{(n),k}(\hat{\omega} | \hat{\omega}_k) > \xi^{|\mathcal{T}|} \iff \hat{P}_{(n),k}(\hat{\omega} | \hat{\omega}_k) > 0. \quad (43)$$

Then, we can find a path with positive probability for both of these scenarios, where their state does not match until they reach  $\kappa = \max_m |\mathcal{I}^m|$ . In other words,  $\omega_{nT+\kappa} = \hat{\omega}_{nT+\kappa}$  and  $\omega_k \neq \hat{\omega}_k$  for all  $k = nT, \dots, nT + \kappa - 1$  with

$$\Pr \left( \bigwedge_{k=nT+1}^{nT+\kappa} \hat{\omega}_k \mid \hat{\omega}_{nT} \right) \geq \xi^\kappa \quad \text{and} \quad \Pr \left( \bigwedge_{k=nT+1}^{nT+\kappa} \omega_k \mid \omega_{nT} \right) \geq \xi^\kappa. \quad (44)$$

Hence, (44) satisfies the first condition for [Donmez et al., 2024, Lemma 2].

**Step 2.** For this part, we introduce a notation  $a^{jm} \in A^{jm}$ , where  $jm$  represents the  $j^{th}$  agent from team  $m$ . For example, if all teams have identical agent numbers and agent indexes are ordered for teams,  $jm = (m - 1)|T_m| + j$ . Also, let  $\pi_{(n),k} := (\pi_{(n),k}^m)_{m \in \mathcal{T}}$ , and  $\hat{\pi}_{(n),k} := (\hat{\pi}_{(n),k}^m)_{m \in \mathcal{T}}$ . Now, consider the total variation distance between two transition probabilities. Transition probabilities between state  $\omega = \{(a^{im}, a^{-im})\}_{m \in \mathcal{T}}$  and  $\tilde{\omega} = \{(\tilde{a}^{im}, a^{-im})\}_{m \in \mathcal{T}}$ , where  $im$  can be any random agent from team  $m$  with a slight abuse of notation, can be written as a function of the belief as

$$P_{\omega \rightarrow \tilde{\omega}}(\pi_{(n),k}) = \prod_{m \in \mathcal{T}} P_{\omega \rightarrow \tilde{\omega}}^m(\pi_{(n),k}), \quad (45)$$

where  $P_{\omega \rightarrow \tilde{\omega}}^m$  is defined to be

$$P_{\omega \rightarrow \tilde{\omega}}^m(\pi_{(n),k}) = \frac{1}{|\mathcal{I}^m|} \frac{\exp\left(\left(\phi^m\left(\tilde{a}^{im}, a^{-im}, \pi_{(n),k}^{-m}\right)\right) / \tau\right)}{\sum_{\tilde{a}' \in A^{im}} \exp\left(\phi^m\left(\tilde{a}', a^{-im}, \pi_{(n),k}^{-m}\right) / \tau\right)}, \quad (46)$$

and  $\pi_{(n),k}^{-m} := (\pi_{(n),k}^\ell)_{\ell \neq m}$ . Remember that due to the separable structure of the network in ZSPTG we have

$$\phi^m(a^{im}, (a^{-im}), \pi_{(n),k}^{-m}) = \sum_{\ell \neq m} \mathbb{E}_{\underline{a}^\ell \sim \pi_{(n),k}^\ell} \phi^{m\ell}(a^{im}, a^{-im}, \underline{a}^\ell) \quad (47a)$$

$$= \sum_{\ell \neq m} (\underline{a}^m)^T \Phi^{m\ell} \pi_{(n),k}^\ell, \quad (47b)$$

where  $\Phi^{m\ell}$  is the matrix form of the potential function whose rows are the joint actions of team  $m$  and columns are the joint actions of team  $\ell$ .

Now, for all  $\omega_{k+1}$ , which is reachable in one step from the state  $\omega_k$ , i.e., at most a single agent changes action in each team, the relation between state transition probabilities and  $P_{\omega \rightarrow \tilde{\omega}}^m(\cdot)$  can be expressed as

$$P_{(n),k}(\omega_{k+1} | \omega_k, \dots, \omega_{nT}) = \prod_{m \in \mathcal{T}} P_{\omega_k \rightarrow \omega_{k+1}}^m(\pi_{(n),k}), \quad (48a)$$

$$\hat{P}_{(n)}(\omega_{k+1} | \omega_k) = \prod_{m \in \mathcal{T}} P_{\omega_k \rightarrow \omega_{k+1}}^m(\hat{\pi}_{(n),k}). \quad (48b)$$

Note that  $\pi_{(n),k}$  is a function of history of actions and the initial  $\pi_{(n)}$ . Therefore, it can be computed given the past states  $(\omega_k, \dots, \omega_{nT})$ .

We know that  $P_{\omega \rightarrow \tilde{\omega}}^m(\pi_{(n),k})$  is bounded as it is a probability. Now, using the Lipschitz property of the soft-max function, and since  $\phi^m(\cdot, \pi_{(n),k}^{-m})$  is a linear function of  $\pi_{(n),k}$ , we can say that  $P_{\omega_k \rightarrow (\cdot)}^m(\pi)$  is a Lipschitz continuous function with respect to the input  $\pi$ . Also, using the fact that the product of bounded Lipschitz continuous functions is also Lipschitz continuous, and (48), we can say that there exists an  $\mathcal{L} < \infty$  such that the total variation distance between two transition probabilities is bounded as

$$\left\| P_{(n),k}(\cdot | \omega_k, \dots, \omega_{nT}) - \hat{P}_{(n)}(\cdot | \omega_k) \right\|_{\text{TV}} \leq \mathcal{L} \sum_{m \in \mathcal{T}} \left\| (\pi_{(n),k}^\ell - \hat{\pi}_{(n),k}^\ell) \right\|_1. \quad (49)$$

The distance between the belief of the original scenario and the reference scenario can also be bounded thanks to the small step size. If we bound  $\|a_k^\ell - \pi_k^\ell\|_1 < \|a_k^\ell\|_1 + \|\pi_k^\ell\|_1 = 2$ . Then using triangle inequality

$$\left\| (\pi_{(n),k}^\ell - \hat{\pi}_{(n),k}^\ell) \right\|_1 = \left\| (\pi_{(n),k}^\ell - \pi_{(n),nT}^\ell) \right\|_1 \leq \sum_{t=nT}^{k+NT-1} 2\alpha_t \quad (50a)$$

$$\leq \sum_{t=nT}^{(n+1)T-1} 2\alpha_t \quad (50b)$$

$$\leq 2T\alpha_{nT}. \quad (50c)$$

Let's consider late epochs where  $\alpha_{nT} < \frac{1}{2T\mathcal{L}|\mathcal{T}|}$ , and set  $2T\mathcal{L}|\mathcal{T}|\alpha_{nT} = 1 - \lambda_{nT}$  with  $0 < \lambda_{nT} \leq 1$ . Then,

$$\left\| P_{\omega \rightarrow \tilde{\omega}}(\pi_{(n),k}^{-m}) - P_{\omega \rightarrow \tilde{\omega}}(\hat{\pi}_{(n),k}^{-m}) \right\|_{\text{TV}} \leq 1 - \lambda_{nT}. \quad (51)$$

Hence, the second condition of [Donmez et al., 2024, Lemma 2] is met, and we can invoke the corresponding Lemma such that given the distributions of the original and reference scenarios, following inequality holds for all  $k \geq nT$ ,

$$\left\| \mu_{(n),k} - \hat{\mu}_{(n),k} \right\|_1 \leq 2(1 - \varepsilon)^{\frac{k-nT}{\kappa} - 1} + 2(1 - \lambda_{nT}^\kappa) \frac{1 + \varepsilon}{\varepsilon}, \quad (52)$$

where  $\varepsilon = \xi^{2\kappa}$ . If we define constants,  $c := 2(1 - \varepsilon)^{\frac{1}{\kappa} - 1}$ ,  $\rho := (1 - \varepsilon)^{\frac{1}{\kappa}}$ ,  $d := 4\frac{1 + \varepsilon}{\varepsilon}$ , and assume that the reference scenario initial distribution is the stationary distribution of the MC,  $\hat{\mu}_{(n),\star}$ , we can rewrite the inequality (52) as follows

$$\left\| \mu_{(n),k} - \hat{\mu}_{(n),\star} \right\|_1 \leq c \cdot \rho^{k-nT} + d \cdot T\alpha_{nT}, \quad (53)$$

for all  $k \geq nT$ . Then, by (41), and triangle inequality, we can conclude

$$\left\| \mu_{(n),k}^m - \hat{\mu}_{(n),\star}^m \right\|_1 \leq \left\| \mu_{(n),k} - \hat{\mu}_{(n),\star} \right\|_1 \leq c \cdot \rho^{k-nT} + d \cdot T\alpha_{nT}, \quad (54)$$

for all  $k \geq nT$ .

## B.2 Proof of Proposition A.2

The set  $\Pi$  is compact set by definition as it is a Cartesian product of probability simplexes. Let's consider a convergent sequence  $(\pi_n, \mu_n - \pi_n + e_n)_{n=1,2,\dots}$  in the set  $\{(\pi_n, y) : y \in F(\pi)\}$ , and let  $(\pi^*, \mu^* - \pi^* + e^*)$  be the point that the sequence converges to. Given  $\pi_n$ , any  $\mu_n$  is a fixed and unique value, and it is an element of the compact set that is generated by mapping probability simplex with the continuous soft-max function. Then, for any  $\pi^* \in \Pi$ ,  $\mu^* - \pi^*$  is a fixed value within another compact set. Furthermore, the error term must remain within the compact set  $e^* \in e$ . As a result,  $(\pi^*, \mu^* - \pi^* + e^*)$  is also within the set  $\{(\pi_n, y) : y \in F(\pi)\}$ , and  $F : \Pi \rightarrow A^{\sum_{m \in \mathcal{T}} \|\Delta^m\|}$  is a closed-set valued map. Therefore, the condition (i) is satisfied. Given a  $\pi \in \Pi$ ,  $\mu_\star \in \Pi$  is a fixed value corresponding to the smoothed best responses to  $\pi_{\{m \in \mathcal{T}\}}^{-m}$ . Hence,  $\mu_\star - \pi$  is a fixed value for a given  $\pi$ . Note that each  $e^m \in e$  is a non-empty, bounded, closed and convex subset of  $\mathbb{R}^{\sum_{m \in \mathcal{T}} \|\Delta^m\|}$ . Therefore, for any given  $\pi \in \Pi$ ,  $F(\pi) = \mu_\star - \pi + e$  is a non-empty, compact, convex subset of  $\mathbb{R}^{\sum_{m \in \mathcal{T}} \|\Delta^m\|}$ . As a result, (ii) is also satisfied. The function  $F$  is bounded such that

$$\sup_{y \in F(x)} \|y\|_1 \leq \sup_{\pi \in \Delta} \|\pi\|_1 + \sup_{\mu \in \Delta} \|\mu\|_1 + \sup \|e\| \leq 2M + M \left( \frac{1}{T} \frac{1}{1 - \rho} + K^m(\delta) \right). \quad (55)$$

Hence, it satisfies the condition (iii). Since all three conditions are satisfied,  $F$  is a Marchaud Map.

## B.3 Proof of Proposition A.3

The smoothness of the entropy regularization in (1) yields that we can invoke the envelope theorem to compute the time derivative of  $L(\pi)$  as

$$\frac{d}{dt} L(\pi) = \sum_{m \in \mathcal{T}} \phi^m(\mu_\star^m, \dot{\pi}^{-m}) \quad (56)$$

$$\stackrel{(a)}{=} \sum_{m \in \mathcal{T}} \sum_{\ell \neq m} \phi^{m\ell}(\mu_\star^m, \dot{\pi}^\ell), \quad (57)$$

where  $\mu_\star^m := \text{br}_\tau(\phi^m(\cdot, \pi^{-m}))$  and (a) follows from (4). By (35) and (33), we have  $\dot{\pi}^\ell = \mu_\star^\ell - \pi^\ell + e^\ell$ . Therefore, we can write (57) as

$$\frac{d}{dt} L(\pi) = \sum_{m \in \mathcal{T}} \sum_{\ell \neq m} \left( \phi^{m\ell}(\mu_\star^m, \mu_\star^\ell) - \phi^{m\ell}(\mu_\star^m, \pi^\ell) + \sum_{\underline{a}^\ell \in \underline{A}^\ell} e^\ell(\underline{a}^\ell) \phi^{m\ell}(\mu_\star^m, \underline{a}^\ell) \right) \quad (58)$$

For the first two terms on the right-hand side, we have

$$\sum_{m \in \mathcal{T}} \sum_{\ell \neq m} \left( \phi^{m\ell}(\mu_\star^m, \mu_\star^\ell) - \phi^{m\ell}(\mu_\star^m, \pi^\ell) \right) \stackrel{(a)}{=} - \sum_{m \in \mathcal{T}} \phi^m(\mu_\star^m, \pi^{-m}), \quad (59)$$

$$\stackrel{(b)}{\leq} -L(\pi) + \tau \log |A|, \quad (60)$$

where (a) is due to (3) and (4), and (b) follows from the definition (37a) as  $\mathcal{H}(\mu_\star^m) \leq \log |\underline{A}^m|$ . On the other hand, we have

$$\sum_{m \in \mathcal{T}} \sum_{\ell \neq m} \sum_{\underline{a}^\ell \in \underline{A}^\ell} e^\ell(\underline{a}^\ell) \phi^{m\ell}(\mu_\star^m, \underline{a}^\ell) \leq \sum_{m \in \mathcal{T}} \sum_{\ell \neq m} \|e^\ell\|_1 \cdot \bar{\phi} \leq |\mathcal{T}|^2 \bar{\phi} \cdot C(\varepsilon, T), \quad (61)$$

due to the bound on the errors. By the bounds (60) and (61), we can bound the time derivative of  $L(\pi)$  as

$$\frac{d}{dt} L(\pi) < -L(\pi) + \Xi(\lambda, \varepsilon, T) \Leftrightarrow \frac{d}{dt} (L(\pi) - \Xi(\lambda, \varepsilon, T)) < -L(\pi) + \Xi(\lambda, \varepsilon, T), \quad (62)$$

where the constant  $\Xi(\lambda, \varepsilon, T) > 0$  is as described in (37b). Since we have  $V(\pi) = \min(0, L(\pi) - \Xi(\lambda, \varepsilon, T))$ , the strict inequality in (62), yields that  $V(\cdot)$  is a Lyapunov function.

## C Extension to Multi-team Markov Games

Markov games (MGs), introduced by Shapley [1953], generalizes Markov decision processes (MDPs) to non-cooperative multi-agent environments. We can characterize a multi-team MG by the tuple  $\langle H, \mathcal{T}, S, (A^i, r^i)_{i \in \mathcal{I}}, p, p_o \rangle$ , where  $H$  is the horizon length,  $\mathcal{T}$  and  $\mathcal{I}$  again denote the index sets for the teams and agents,  $S$  denotes the *finite* set of states, and  $A^i$  and  $r^i : S \times A \rightarrow \mathbb{R}$  denote, resp., the agent  $i$ 's finite action set and *immediate reward* function, depending on current state and joint actions.<sup>1</sup> The state of the underlying game can change according to the transition kernel  $p(\cdot | s, a)$ , depending on the current state and joint actions, and the initial state is determined by the probability distribution  $p_o \in \Delta(S)$ .

Let each team  $m$  randomize their actions contingent on the current state  $s \in S$  and stage  $h \in [H] := \{1, \dots, H\}$  via a stationary strategy  $\pi^m : S \times [H] \rightarrow \Delta(\underline{A}^m)$ . Note that team members do not necessarily randomize their actions independently. Given the strategy profile of teams  $\underline{\pi} = (\pi^m)_{m \in \mathcal{T}}$ , agent  $i$ 's utility function is defined by

$$U^i(\underline{\pi}) := \mathbb{E} \left[ \sum_{h=1}^H r^i(s_h, a_h) \right], \quad (63)$$

where the pair  $(s_h, a_h)$  denotes the state and action profile at stage  $h$ , the expectation is taken with respect to the randomness on these pairs  $(s_h, a_h)$  induced by the strategy profile  $\underline{\pi}$  and the underlying transition kernel.

For each state  $s$ , let the reward functions  $\{r^i(s, \cdot)\}_{i \in \mathcal{I}}$  induce a ZSPTG where team  $m \in \mathcal{T}$  has the potential function  $\phi^m(s, \cdot) : A \rightarrow \mathbb{R}$  satisfying (2), (3), and (4). Correspondingly, team  $m$ 's utility function is given by

$$\underline{U}^m(\underline{\pi}) := \mathbb{E} \left[ \sum_{h=1}^H \phi^m(s_h, a_h) \right]. \quad (64)$$

Let  $\underline{\Pi}^m := \{\pi^m | \pi^m : S \times [H] \rightarrow \Delta(\underline{A}^m)\}$  denote the set of stationary strategy profiles for team  $m$ . Then, the following is the counterpart of Definition 2.4 for multi-team MGs.

**Definition C.1** (Team-Nash Gap for MGs). Given the strategy profile of teams  $\{\pi^m \in \underline{\Pi}^m\}_{m \in \mathcal{T}}$ , we define the *team-Nash gap* for team  $m$  as

$$\text{TNG}^m(\underline{\pi}) := \max_{\tilde{\pi} \in \underline{\Pi}^m} \{ \underline{U}^m(\tilde{\pi}, \pi^{-m}) \} - \underline{U}^m(\underline{\pi}), \quad (65)$$

and  $\text{TNG}(\underline{\pi}) := \sum_{m \in \mathcal{T}} \text{TNG}^m(\underline{\pi})$ , where  $\pi^{-m} := \{\pi^\ell\}_{\ell \neq m}$ . Correspondingly, we say that the strategy profile of teams  $\{\pi^m\}_{m \in \mathcal{T}}$  is  $\varepsilon$ -TNE if  $\text{TNG}(\underline{\pi}) \leq \varepsilon$ .

<sup>1</sup>The results can be generalized to state-variant action sets rather straightforwardly.

Next, we describe the stage-game framework, going back to the introduction of Markov games [Shapley, 1953], and also used in multi-agent reinforcement learning algorithms such as Minimax-Q [Littman, 1994] and Nash-Q [Hu and Wellman, 2003]. Particularly, at each stage, the agents' joint actions determine the immediate rewards they receive and the next state, and therefore, the future rewards to be received. Given the strategy profile of teams  $\underline{\pi} = (\underline{\pi}^m)_{m \in \mathcal{T}}$ , let  $v^i : S \times [H] \rightarrow \mathbb{R}$  denote agent  $i$ 's *value function* such that  $v^i(s)$  is the game value for the stage game associated with state  $s$ . Similarly, let  $Q^i : S \times [H] \times A \rightarrow \mathbb{R}$  be the *Q-function* such that  $Q^i(s, \cdot)$  corresponds to agent  $i$ 's payoff function for the stage game associated with state  $s$ . The definition of the utility (63) yields that

$$Q^i(s, h, a) = r^i(s, a) + \mathbb{I}_{\{h < H\}} \sum_{s_+ \in S} p(s_+ | s, a) \cdot v^i(s_+, h + 1), \quad (66a)$$

$$v^i(s, h) = Q^i(s, h, \underline{\pi}(s, h, \cdot)). \quad (66b)$$

The Q-function and value function implicitly depend on  $\underline{\pi}$ .

We can extend Team-FP dynamics to MGs played repeatedly. Based on the stage game framework, agents play some stage game associated with each state  $s \in S$  and stage  $h \in [H]$  pair repeatedly whenever the underlying MG visits state  $s$  at stage  $h$ . Correspondingly, agents form beliefs about the opponent teams as if the opponent teams play according to some stationary strategies across these repetitions.

Let  $\underline{\pi}_k^\ell : S \times [H] \rightarrow \Delta(A^\ell)$  denote the beliefs formed by agents  $i \notin \mathcal{I}^\ell$  about team  $\ell$ , and  $Q_k^i : S \times [H] \times A \rightarrow \mathbb{R}$  denote agent  $i$ 's Q-function estimate at the  $k$ th repetition. If the underlying MG visits state  $s$  at stage  $h$  at the  $k$ th repetition, then agent  $i$  follows Team-FP dynamics as if the payoff function is the Q-function estimate  $Q_k^i(s, h, \cdot)$ . Agents also recall the previous actions of their teams specific to each stage game. We denote agent  $i$ 's previous action for the pair of state  $s$  and stage  $h$  until and including the  $k$ th repetition by  $a_k^i(s, h) \in A^i$ , with a slight abuse of notation. Then, at stage  $k$ , agent  $i \in \mathcal{I}^m$  either takes the previous action for that stage game (i.e.,  $a_k^i(s, h) = a_{k-1}^i(s, h)$ ), or takes the action  $a_k^i(s, h) \sim \text{br}_\tau(Q_k^i(s, h, \cdot), a_{k-1}^{-i}(s, h), \underline{\pi}_k^{-m}(s, h))$  according to the smoothed best response to the previous actions of the team members  $a_{k-1}^{-i}(\cdot) := \{a_{k-1}^j(\cdot)\}_{j \in \mathcal{I}^m \setminus \{i\}}$  and the beliefs  $\underline{\pi}_k^{-m} := \{\underline{\pi}_k^\ell\}_{\ell \neq m}$  formed about other teams.

Let  $s_{h,k}$  and  $a_{h,k}$  denote, resp., the state and action profile at stage  $h$  at the  $k$ th repetition. Then, given some reference step size  $\{\alpha_c\}_{c \geq 0}$ , agents  $j \notin \mathcal{I}^m$  update their beliefs about team  $m$ 's strategy according to

$$\underline{\pi}_{k+1}^m(s, h) = \underline{\pi}_k^m(s, h) + \lambda_k(s, h) \cdot (\underline{a}_k^m(s, h) - \underline{\pi}_k^m(s, h)) \quad \forall k = 0, 1, \dots, \quad (67a)$$

$$\lambda_k(s, h) = \mathbb{I}_{\{s=s_{h,k}\}} \alpha_{c_k(s, h)}, \quad (67b)$$

where  $\underline{a}_k^m(\cdot) := \{a_k^j(\cdot)\}_{j \in \mathcal{I}^m}$  and  $c_k(s, h)$  is the number of times state  $s$  get visited at stage  $h$  until the  $k$ th repetition. Correspondingly, by (66), each agent  $i \in \mathcal{I}^m$  updates their Q-function estimates according to

$$Q_{k+1}^i(s, h, a) = Q_k^i(s, h, a) + \bar{\lambda}_k(s, h, a) \left( \widehat{Q}_k^i(s, h, a) - Q_k^i(s, h, a) \right), \quad (68)$$

where  $\bar{\lambda}_k(s, h, a) \in [0, 1]$  is also some step size. If the agent  $i$  knows the model of the underlying MG, then we have

$$\widehat{Q}_k^i(s, h, a) = r^i(s, a) + \mathbb{I}_{\{h < H\}} \sum_{s_+ \in S} p(s_+ | s, a) \cdot v_k^i(s_+, h + 1), \quad (69a)$$

$$\bar{\lambda}_k(s, h, a) = \mathbb{I}_{\{s=s_{h,k}\}} \alpha_{c_k(s, h)}, \quad (69b)$$

for all  $a \in A$  and

$$v_k^i(s, h) = Q_k^i(s, h, \underline{a}_k^m(s, h), \underline{\pi}_k^{-m}(s, h)) \quad \forall (s, h) \in S \times [H]. \quad (70)$$

If the agent does not know the model, then we have

$$\widehat{Q}_k^i(s, h, a) = r_{h,k}^i + \mathbb{I}_{\{h < H\}} v_k^i(s_{h+1,k}, h + 1), \quad (71a)$$

$$\bar{\lambda}_k(s, h, a) = \mathbb{I}_{\{(s,a)=(s_{h,k}, a_{h,k})\}} \alpha_{c_k(s, h, a)}, \quad (71b)$$

---

**Algorithm 2** Model-based (Independent) Team-FP for MGs

---

- 1: **initialize:**  $\{\underline{\pi}_0^\ell(\cdot)\}_{\ell \neq m}$ ,  $\{a_{-1}^j(\cdot)\}_{j \in \mathcal{I}^m \setminus \{i\}}$ , and  $Q_0^i(\cdot)$  arbitrarily for the typical agent  $i \in \mathcal{I}^m$
- 2: **for** each repetition  $k = 0, 1, \dots$  **do**
- 3:   **for** each stage  $h = 1, \dots, H$  **do**
- 4:     **require:**  $\{\underline{\pi}_k^\ell(\cdot)\}_{\ell \neq m}$ ,  $\{a_{k-1}^j(\cdot)\}_{j \in \mathcal{I}^m \setminus \{i\}}$ , and  $Q_k^i(\cdot)$
- 5:     observe current state  $s$
- 6:     set  $\bar{s} = (s, h)$
- 7:     play  $a_k^i(\bar{s}) \sim \text{br}_\tau(Q_k^i(\bar{s}, \cdot, a_{k-1}^{-i}(\bar{s}), \underline{\pi}_k^{-m}(\bar{s})))$  or  $a_k^i(\bar{s}) = a_{k-1}^i(\bar{s})$  in a coordinated way (or independently) simultaneously with other agents
- 8:     observe  $a_{h,k}^{-i} = a_k^{-i}(\bar{s})$
- 9:     receive  $r_{h,k}^i = r^i(s, a_{h,k})$
- 10:   **end for**
- 11:   **require:** trajectory  $\{s_{h,k}, a_{h,k}^{-i}\}_{h=1}^H$
- 12:   set

$$v_k^i(s, h) = Q_k^i(s, h, \underline{a}_k^m(s, h), \underline{\pi}_k^{-m}(s, h))$$
$$\widehat{Q}_k^i(s, h, \cdot) = r^i(s, \cdot) + \mathbb{I}_{\{h < H\}} \sum_{s_+ \in S} p(s_+ | s, \cdot) \cdot v_k^i(s_+, h + 1)$$

- 13:   update the beliefs and the Q-functions

$$\underline{\pi}_{k+1}^\ell(s, h) = \underline{\pi}_k^\ell(s, h) + \mathbb{I}_{\{s=s_{h,k}\}} \alpha_{c_k(s,h)} \cdot (\underline{a}_k^\ell(s, h) - \underline{\pi}_k^\ell(s, h)) \quad \forall \ell \neq m$$
$$Q_{k+1}^i(s, h, \cdot) = Q_k^i(s, h, \cdot) + \mathbb{I}_{\{s=s_{h,k}\}} \alpha_{c_k(s,h)} \left( \widehat{Q}_k^i(s, h, \cdot) - Q_k^i(s, h, \cdot) \right)$$

for all  $(s, h) \in S \times [H]$

- 14: **end for**
- 

where  $r_{h,k}^i$  denotes the reward received at stage  $h$  at the  $k$ th repetition and we approximate the expected continuation payoff by looking one stage ahead, as in the classical Q-learning algorithm, and  $c_k(s, h, a)$  is the number of times the pair  $(s, a)$  gets realized at stage  $h$  until the  $k$ th repetition. Algorithms 2 and 3 provide descriptions of the extensions, resp., for the model-based and model-free cases from the perspective of the typical agent  $i \in \mathcal{I}^m$  from the typical team  $m \in \mathcal{T}$ .

*Remark C.2.* Algorithm 2 reduces to Algorithm 1 if  $|S| = 1$  and  $H = 1$ .

## D Large-scale Numerical Examples

In this section, we give a large-scale experiment that show the scalability of Team-FP in a networked game where only 2-hop neighbors of agents affect their payoff function. We simulated a three-team game with nine agents per team, resulting in a large joint action space of size  $2^{27}$ . After ten independent trials, we plotted the evolution of the Team-Nash Gap in Figure 7. Despite the problem's scale, the empirical averages of team actions converge to the Team-Nash equilibrium at a similar rate, even with sparse network interconnections, as shown in the top right of Figure 7.

---

**Algorithm 3** Model-free (Independent) Team-FP for MGs
 

---

- 1: **initialize:**  $\{\underline{\pi}_0^\ell(\cdot)\}_{\ell \neq m}$ ,  $\{a_{-1}^j(\cdot)\}_{j \in \mathcal{I}^m \setminus \{i\}}$ , and  $Q_0^i(\cdot)$  arbitrarily for the typical agent  $i \in \mathcal{I}^m$
- 2: **for** each repetition  $k = 0, 1, \dots$  **do**
- 3:   **for** each stage  $h = 1, \dots, H$  **do**
- 4:     **require:**  $\{\underline{\pi}_k^\ell(\cdot)\}_{\ell \neq m}$ ,  $\{a_{k-1}^j(\cdot)\}_{j \in \mathcal{I}^m \setminus \{i\}}$ , and  $Q_k^i(\cdot)$
- 5:     observe current state  $s$
- 6:     set  $\bar{s} = (s, h)$
- 7:     play  $a_k^i(\bar{s}) \sim \text{br}_\tau(Q_k^i(\bar{s}, \cdot, a_{k-1}^{-i}(\bar{s}), \underline{\pi}_k^{-m}(\bar{s})))$  or  $a_k^i(\bar{s}) = a_{k-1}^i(\bar{s})$  in a coordinated way (or independently) simultaneously with other agents
- 8:     observe  $a_{h,k}^{-i} = a_k^{-i}(\bar{s})$
- 9:     receive  $r_{h,k}^i = r^i(s, a_{h,k})$
- 10:   **end for**
- 11:   **require:** trajectory  $\{s_{h,k}, a_{h,k}, r_{h,k}\}_{h=1}^H$
- 12:   set

$$v_k^i(s, h) = Q_k^i(s, h, \underline{a}_k^m(s, h), \underline{\pi}_k^{-m}(s, h))$$

$$\widehat{Q}_k^i(s, h, \cdot) = r_{h,k}^i + \mathbb{I}_{\{h < H\}} v_k^i(s_{h+1,k}, h+1)$$

- 13:   update the beliefs and the Q-functions

$$\underline{\pi}_{k+1}^\ell(s, h) = \underline{\pi}_k^\ell(s, h) + \mathbb{I}_{\{s=s_{h,k}\}} \alpha_{c_k}(s, h) \cdot (\underline{a}_k^\ell(s, h) - \underline{\pi}_k^\ell(s, h)) \quad \forall \ell \neq m$$

$$Q_{k+1}^i(s, h, a) = Q_k^i(s, h, a) + \mathbb{I}_{\{(s,a)=(s_{h,k}, a_{h,k})\}} \alpha_{c_k}(s, h, a) \left( \widehat{Q}_k^i(s, h, a) - Q_k^i(s, h, a) \right)$$

for all  $(s, h, a) \in S \times [H] \times A$

- 14: **end for**
- 

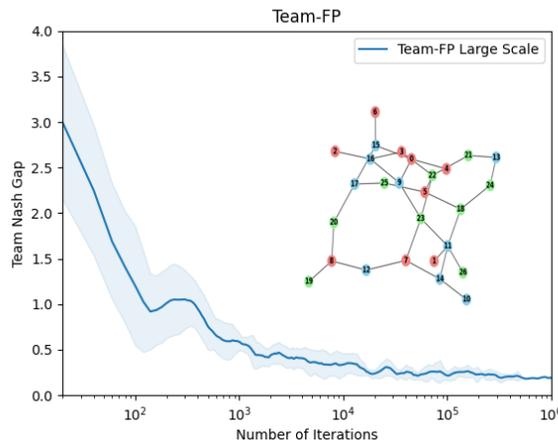


Figure 7: The evolution of Team Nash Gap in the large-scale example provided in the top right, showing that Team-FP dynamics reach near team-minimax equilibrium.

## NeurIPS Paper Checklist

### 1. Claims

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

Answer: [Yes]

Justification: The abstract and introduction is clear about the claims of the paper.

Guidelines:

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

## 2. Limitations

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

Answer: [Yes]

Justification: The limitations are discussed at the end before the conclusion.

Guidelines:

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

## 3. Theory Assumptions and Proofs

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

Answer: [Yes]

Justification: For all the theorems, we provide assumptions and proofs along with proper references.

Guidelines:

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

- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

#### 4. Experimental Result Reproducibility

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

Answer: [Yes]

Justification: Although there are some randomness in the generation, and we did not save the seeds, we share the code, and it should work as in the simulations for any other randomly generated games. Therefore, anyone can reproduce the results.

Guidelines:

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

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

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

Answer: [Yes]

Justification: We share code for experiments. However, the code may be too complex to understand as most of the documentation for the code is missing.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.

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

## 6. Experimental Setting/Details

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

Answer: [Yes]

Justification: We specify experimental settings of each experiment.

Guidelines:

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

## 7. Experiment Statistical Significance

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

Answer: [Yes]

Justification: We include mean and standard deviation of independent trials.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).

- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

## 8. Experiments Compute Resources

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

Answer: [Yes]

Justification: We give specifications of the computer and rough run-time of the algorithms. However, we do not provide exact run-times or required time for preliminary experiments.

Guidelines:

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

## 9. Code Of Ethics

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

Answer: [Yes]

Justification: We discuss how our work does not raise any ethical concerns in the conclusion section.

Guidelines:

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

## 10. Broader Impacts

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

Answer: [NA]

Justification: There is no social impact of the work.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

## 11. Safeguards

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

Answer: [NA]

Justification: The paper does not contain risks

Guidelines:

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

## 12. Licenses for existing assets

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

Answer: [NA]

Justification: We do not use existing assets

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

## 13. New Assets

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

Answer: [NA]

Justification: There are no new assets.

Guidelines:

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

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

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

Answer: [NA]

Justification: There are no human subjects or potential risks.

Guidelines:

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

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

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

Answer: [NA]

Justification: There are no human subjects or potential risks.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.