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

Ahmed Said Donmez [1] Yuksel Arslantas [1] M. Omer Sayin [1]

NeurIPS 2024

[1]Bilkent University

Motivation

Two-team games are common in many scenarios.

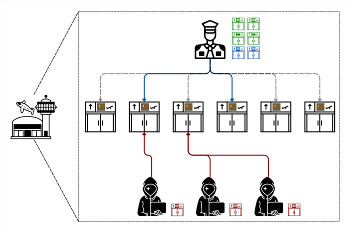


Figure 1: Example two team game (**Airport Security**): Consider a security chief guarding the six gates of an airport against three different autonomous intruders.

Key Question

In such multi-team games, can agents within teams learn to coordinate and act according to the best team strategy without explicit communication?

Difficulties of the Problem

- Common independent algortihms do not have any guarantees for joint team behavior.
- Prior works assume that agents can communicate beforehand and act as if they are a single agent during the game.

Main Difficulty

While a team tries to learn the best strategy, other teams will also learn and change strategies.

Team (Potential) Games

- All players have aligning objectives
- There exist a common potential function ϕ, with the following property: If any player changes their action while the rest keep it the same, change in the payoff = change in the potential function.

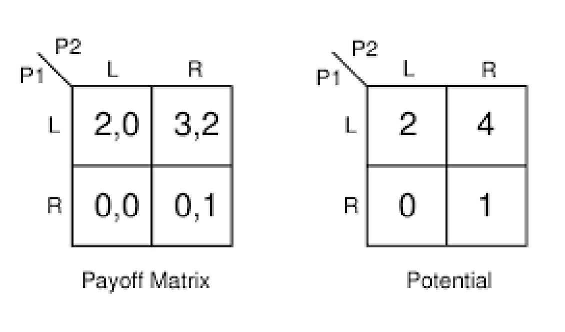


Figure 2: Example Potential Game

Zero-sum Potential Team Games (ZSPTG)

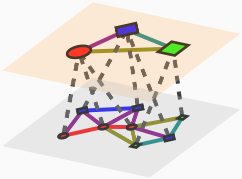


Figure 3: Multi-team Game

- All teams have different potential functions ϕ^m.

Zero-sum Potential Team Games (ZSPTG)

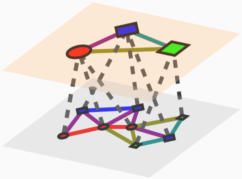


Figure 3: Multi-team Game

- All teams have different potential functions ϕ^m.
- Between the teams the relation might be competitive, cooperative or a mixture of them.

Zero-sum Potential Team Games (ZSPTG)

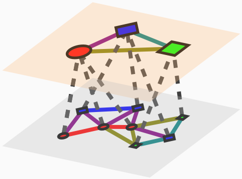


Figure 3: Multi-team Game

- All teams have different potential functions ϕ^m.
- Between the teams the relation might be competitive, cooperative or a mixture of them.

We focus on competitive teams:
Zero-sum Potential Team Games (ZSPTG)

Zero-sum Potential Team Games (ZSPTG)

A two-team ZSPTG consists of:

- Team 1 and Team 2, with agents $\mathcal{I}^1$, and $\mathcal{I}^2$.

Zero-sum Potential Team Games (ZSPTG)

A two-team ZSPTG consists of:

- Team 1 and Team 2, with agents $\mathcal{I}^1$, and $\mathcal{I}^2$.
- Action sets A^i for $i \in \mathcal{I} := \mathcal{I}^1 \cup \mathcal{I}^2$, and joint action sets $\underline{A}^m := \prod_{i \in \mathcal{I}^m} A^i$

Zero-sum Potential Team Games (ZSPTG)

A two-team ZSPTG consists of:

- Team 1 and Team 2, with agents $\mathcal{I}^1$, and $\mathcal{I}^2$.
- Action sets A^i for $i \in \mathcal{I} := \mathcal{I}^1 \cup \mathcal{I}^2$, and joint action sets $\underline{A}^m := \prod_{i \in \mathcal{I}^m} A^i$
- Zero-sum property: Potential functions ϕ^m for each team $m = 1, 2$ such that

$$\phi^1 + \phi^2 = 0.$$

Zero-sum Potential Team Games (ZSPTG)

A two-team ZSPTG consists of:

- Team 1 and Team 2, with agents $\mathcal{I}^1$, and $\mathcal{I}^2$.
- Action sets A^i for $i \in \mathcal{I} := \mathcal{I}^1 \cup \mathcal{I}^2$, and joint action sets $\underline{A}^m := \prod_{i \in \mathcal{I}^m} A^i$
- Zero-sum property: Potential functions ϕ^m for each team $m = 1, 2$ such that

$$\phi^1 + \phi^2 = 0.$$

- Possibly a graph structure.

Zero-sum Potential Team Games (ZSPTG)

Two-team ZSPTG can be generalized to multi-team ZSPTG with adding the condition of separable potential functions.

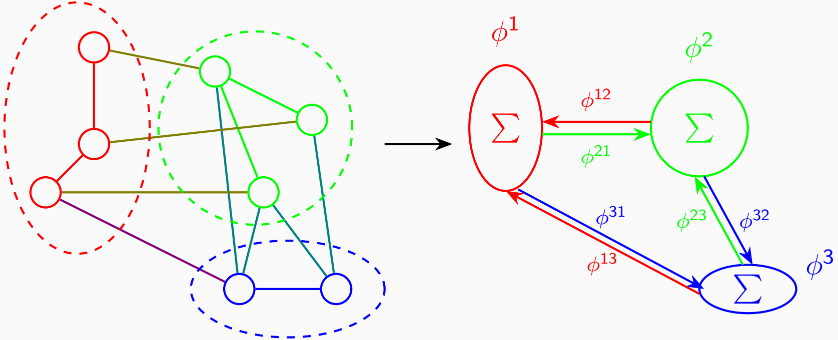


Figure 4: Example multi-team ZSPTG

Team Nash Equilibrium

Definition

Given the strategy profile of teams $\pi := \{\pi^m \in \Delta(\underline{A}^m)\}_{m\in\mathcal{T}}$, we define the **team-Nash gap** for team m as
$\mathsf{TNG}(\pi) := \sum_{m\in\mathcal{T}} \mathsf{TNG}^m(\pi)$, with $\mathsf{TNG}(\pi)$ is defined as

$$\mathsf{TNG}^m(\pi) := \max_{\tilde{\pi}^m \in \Delta(\underline{A}^m)} \left\{\phi^m(\tilde{\pi}^m, \pi^{-m})\right\} - \phi^m(\pi),$$

where $\pi^{-m} := \{\pi^\ell\}_{\ell\neq m}$.
Correspondingly, we say that the strategy profile of teams π is ϵ-TNE if

$$\mathsf{TNG}(\pi) \leq \epsilon.$$

For two-teams, this is also known as team-maxmin equilibrium!

Algorithm Idea

- **Log-linear learning** (Converges to near efficient equilibrium in potential games)
  - Only a single player can change action
  - Keep track of the last actions of others
  - Soft best response to the last actions of teammates
- **(Smoothed) Fictitious Play** (Converges to nash equilibrium in zero-sum polymatrix games)
  - Every player updates their action
  - Keep track of beliefs about every agent (average of actions they played)
  - (Soft) best response to the beliefs

Algorithm Idea

- **Log-linear learning** (Converges to near efficient equilibrium in potential games)
  - Only a single player can change action
  - Keep track of the last actions of others
  - Soft best response to the last actions of teammates
- **(Smoothed) Fictitious Play** (Converges to nash equilibrium in zero-sum polymatrix games)
  - Every player updates their action
  - Keep track of beliefs about every agent (average of actions they played)
  - (Soft) best response to the beliefs

Let's combine these two algorithms!

Our Algorithm (Team-FP)

- In each team, only a single player changes action (can be relaxed for independent case)
- Keep track of beliefs about joint actions of opponent teams and last actions of the teammates
- Soft best response to the beliefs and last actions

Main Result:

Theorem

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 Algorithm. If stepsize assumption holds, then the team-Nash gap for $\pi_k := (\pi_k^m)_{m\in\mathcal{T}}$ satisfies

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

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

Proof Sketch

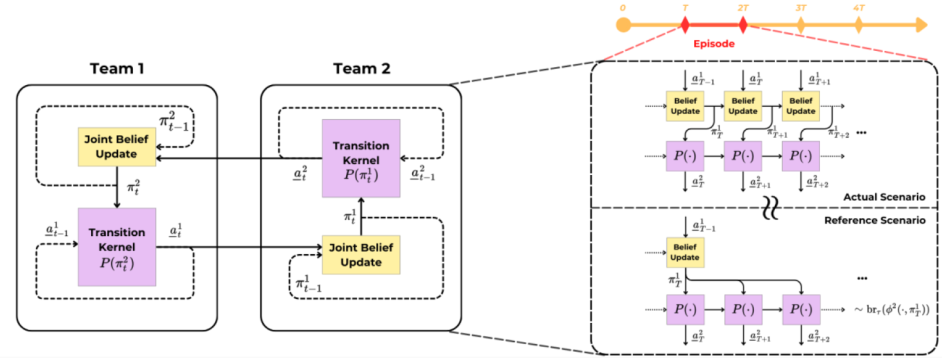
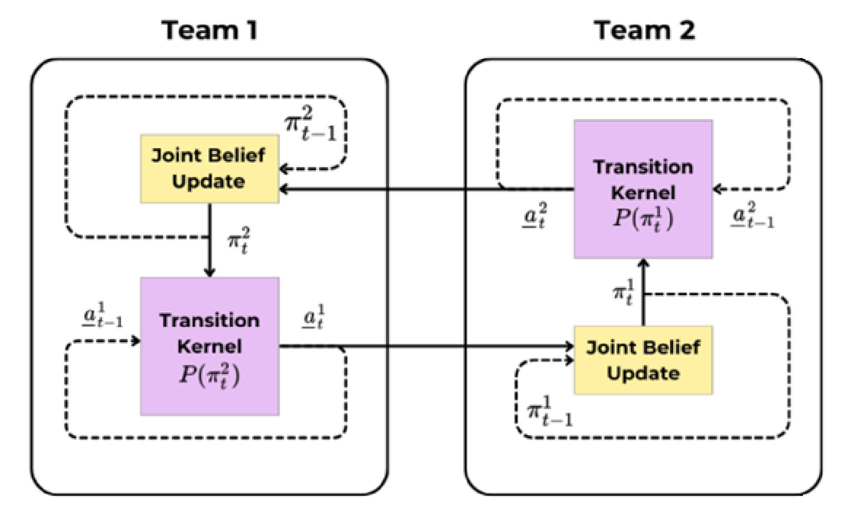
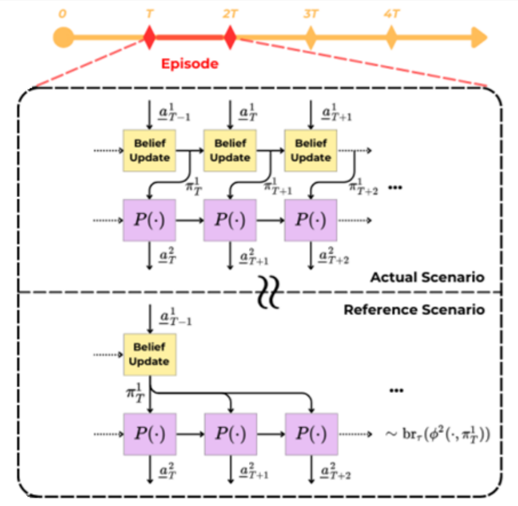


Figure 5: Team-FP Algorithm Dynamics for two teams (left), and the main proof idea (right). We show the Markov Chain of a reference scenario where beliefs π_t^1 do not change, behaves similar to the actual scenario.

Proof Sketch



Proof Sketch



Fictionally, we split the horizon into T-epoch lengths!

Proof Sketch

- The action profiles form a Markov chain within an epoch in the fictional scenario!
- The stationary distribution of the fictional scenario Markov chain: $\mu^m_{(n)}(\underline{a}^m)$
- The action distribution of the actual scenario: $\mu^m_{(n),k} := \mathrm{E}[\underline{a}^m_k \mid \mathcal{F}_{(n)}]$
- The difference $\|\mu^m_{(n),k} - \mu^m_{(n)}\|$ is bounded with arbitrarily small bounds.
- Using stochastic differential inclusion methods on the cumulative epoch update,

$$\pi^m_{(n+1)} = (1 - \beta_{(n)})\pi^m_{(n)} + \beta_{(n)}\left(\hat{\mu}^m_{(n),\star} + \omega^m_{(n+1)} + e^m_{(n)}\right),$$

we obtain convergence.

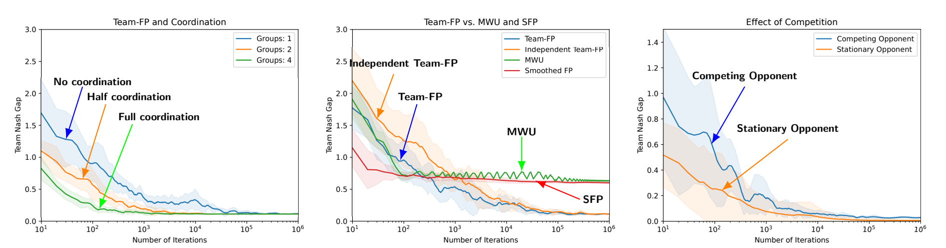
Beyond ZSPTG

We expect Team-FP to converge in other type of games where FP converges:

- Potential games
- 2xN games

We also provide a finite-horizon Markov Game generalization of Team-FP!

Numerical Results

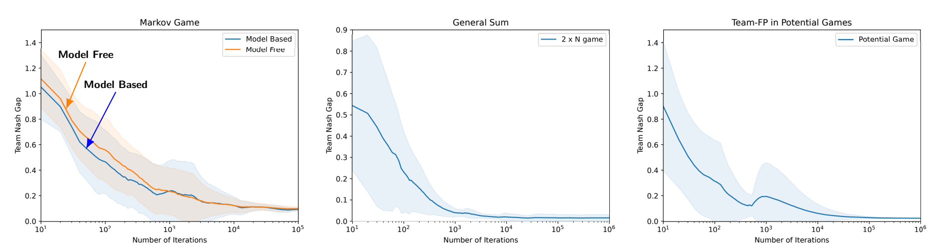


(a) Varying group sizes of 1, 2, and 4.

(b) Team-FP Algorithm (Independent) compared to Smoothed FP, and MWU

(c) Competitive vs Stationary Environment

Figure 6: 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 Multiplicative Weights Update (MWU) and Smoothed FP (SFP) algorithms in a 2-team ZSPTG. (c) Convergence of Team-FP against stationary and competitive opponents in a 3-team ZSPTG.



(a) Finite-horizon Markov Games

(b) 2xN General Sum Game

(c) Potential of Potentials Game

Figure 7: 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 finite-horizon Markov games for extension Algorithms of Team-FP, 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.

Yes! The initial example also show convergence of average joint actions of intruders to the Team Nash Equilibrium.

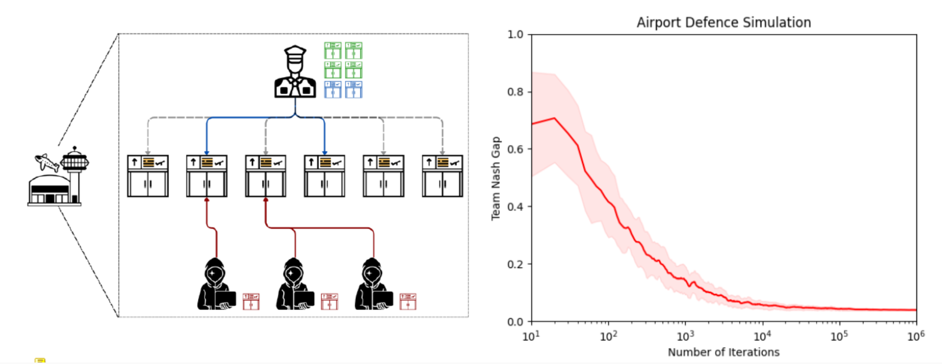


Figure 8: Airport Security Example