Zero-Sum Games between Mean-Field Teams: Reachability-Based Analysis under Mean-Field Sharing

Yue Guan, Mohammad Afshari, Panagiotis Tsiotras

Georgia Institute of Technology
{yguan44, mafshari, tsiotras}@gatech.edu

Abstract

This work studies the behaviors of two large-population teams competing in a discrete environment. The team-level interactions are modeled as a zero-sum game while the agent dynamics within each team is formulated as a collaborative mean-field team problem. Drawing inspiration from the mean-field literature, we first approximate the large-population team game with its infinite-population limit. Subsequently, we construct a fictitious centralized system and transform the infinite-population game to an equivalent zero-sum game between two coordinators. Via a novel reachability analysis, we study the optimality of coordination strategies, which induce decentralized strategies under the original information structure. The ϵ-optimality of the resulting strategies is established in the original finite-population game, and the theoretical guarantees are verified by numerical examples.

Introduction

Multi-agent decision-making arises in many applications, ranging from warehouse robots (Li et al. 2021) to organizational economics (Gibbons, Roberts et al. 2013). While the majority of the literature formulates the problems within either the cooperative or competitive setting, results on mixed collaborative-competitive team behaviors are relatively sparse. In this work, we consider a competitive team game, where two teams, each comprising a large number of intelligent agents, compete at the team level, while agents collaborate within each team. Such hierarchical interactions hold significant relevance in domains such as military operations (Tyler et al. 2020) and other multi-agent systems operating in adversarial environments (Shishika et al. 2022).

There are two major challenges when trying to solve such competitive team problems:

1. Large-population team problems are *computationally* challenging since the solution complexity increases exponentially with the number of agents, and, in general, the team optimal control problems belong to the NEXP complexity class (Bernstein et al. 2002).
2. Competitive team problems are *conceptually* challenging due to the elusive nature of the opponent team. In particular, one may want to impose assumptions on the opponent team to obtain tractable solutions, but these assumptions may not be valid. It is thus unclear whether one can deploy techniques that are readily available in the large-population game literature.

The scalability challenge in large-population multi-agent systems has been addressed for a specific class of games known as the mean-field games (Huang, Malhamé, and Caines 2006; Lasry and Lions 2007). The salient feature of a mean-field game is that agents are weakly-coupled in their dynamics and rewards through their state distribution (the so-called mean-field). The interactions among agents can then be approximated as the interaction between a typical agent and the "mass" of infinitely many other agents. This approximation technique has been extended to *single*-team settings known as the mean-field team problem (Arabneydi and Mahajan 2014). A dynamic programming decomposition is developed for this problem, where all agents within the team deploy the same strategy prescribed by a fictitious coordinator. However, in competitive team setting, although one may restrict the strategies used by her/his team to be identical, extending the same assumption to the opponent team may lead to a substantial underestimation of the opponent's capabilities and thus requires further justification.

Main Contributions

We address the two aforementioned challenges for the class of discrete zero-sum mean-field team games (ZS-MFTGs), which is an extension to the mean-field (single) team problems. Importantly, ZS-MFTG models competitive team behaviors and draws focus to the justifiable approximation of the opponent team strategies.

We develop a dynamic program that constructs ϵ-optimal strategies to the proposed large-population competitive team problem. Notably, our approach finds an optimal solution at the infinite-population limit and considers only *identical* team strategies. This avoids both the so-called "curse of dimensionality" issue in multi-agent systems and the book-keeping of individual strategies. Our main results provide a sub-optimality bound on the exploitability for our proposed solution in the original finite-population game, even when the opponent team is allowed to use non-identical strategies. Specifically, we show that the sub-optimality decreases at the rate of $\mathcal{O}(\underline{N}^{-0.5})$, where $\underline{N}$ is the size of the smaller team.

Our results stem from a novel reachability-based analysis of the mean-field approximation. In particular, we show that any finite-population team behavior can be effectively approximated by an infinite-population team that uses identical team strategies. This result allows us to approximate the original problem with two competing infinite-population teams and transform the resulting problem into a zero-sum game between two *fictitious* coordinators. Such transformation leads to a simple dynamic program based on the common-information approach (Nayyar, Mahajan, and Teneketzis 2013).

Related Literature

Mean-Field Games The mean-field game (MFG) model was introduced in (Huang, Caines, and Malhamé 2007; Lasry and Lions 2007) to address scalability issues in large-population games. The salient feature of MFG is that *selfish* agents are weakly-coupled in their dynamics and rewards through the mean-field (state distribution). If the population is sufficiently large, then an approximately optimal solution can be obtained by solving the infinite-population limit which is known as the mean-field equilibrium (MFE). See (Laurière et al. 2022) for an overview of the results in the MFG literature. The main differences between our setup and the MFG are the following: (a) we seek team optimal strategies while MFG seeks a Nash equilibrium. In particular, we provide performance guarantees when the entire opponent team deviates, while MFG only considers single-agent deviations; (b) The MFE assumes that all agents apply the same strategy and solves the mean-field flow offline. Hence, the MFE strategy is open-loop to the MF. However, under the ZS-MFTG setting, different opponent team strategies lead to different mean-field trajectories. Consequently, we require feedback on the MFs to respond to the strategies deployed by the opponent team.

Mean-Field Teams The single-team problem was explored in (Arabneydi and Mahajan 2014), where agents share a common team reward, resulting in a collaborative problem. The work of (Arabneydi and Mahajan 2015) assumes that all agents within the team apply the same strategy and the optimality for the finite-population game is *only* assured in the LQG setting (Mahajan and Nayyar 2015). Our work encompasses a more intricate two-team zero-sum scenario and justifies the identical team strategy assumptions.

The concurrent work of (Sanjari, Saldi, and Yüksel 2023) studies a similar team-against-team problem but in a continuous state and action setting. The authors analyze the existence of equilibria by modeling randomized strategies as Borel probability measures. Our work differs in the following aspects: (a) The work of Sanjari, Saldi, and Yüksel relies on the Kakutani fixed point theorem to establish the existence of a Nash equilibrium. In contrast, the best-response correspondence is nonconvex given the discrete nature of our formulation (see Numerical Example 1). Therefore, our approach focuses on the single-sided optimality based on the lower and upper game values; (b) our approach transforms the team-against-team problem into a zero-sum game between two coordinators, which allows the deployment of dynamic programming; (c) our results, which incorporate reachability-analysis and additional Lipschitz assumptions, provide the convergence rate of the finite-population team performance to its infinite-population limit.

Notations

We use $[n]$ to denote $\{1, 2, \dots, n\}$. The indicator function is denoted as $\mathbb{1}_{\cdot}(\cdot)$, such that $\mathbb{1}_a(b) = 1$ if $a = b$ and 0 otherwise. We use uppercase letters to denote random variables (e.g., X and $\mathcal{M}$) and lowercase letters to denote their realizations (e.g., x and μ). For a finite set E, we denote the space of all probability measures over E as $\mathcal{P}(E)$.

Problem Formulation

Finite-Population Team Games

Consider a discrete-time system with two large teams of agents that operates over a finite horizon T. The Blue team consists of N_1 *homogeneous* agents, and the Red team consists of N_2 homogeneous agents. The total system size is denoted as $N = N_1 + N_2$, and $\rho = N_1/N$ reflects the size ratio between the two teams. Let $X^{N_1}_{i,t} \in \mathcal{X}$ and $U^{N_1}_{i,t} \in \mathcal{U}$ denote the random variables representing the state and action taken by Blue agent $i \in [N_1]$ at time t. Here, $\mathcal{X}$ and $\mathcal{U}$ are the *finite* individual state and action spaces for each Blue agent, independent of i and t. Similarly, we use $Y^{N_2}_{j,t} \in \mathcal{Y}$ and $V^{N_2}_{j,t} \in \mathcal{V}$ to denote the individual state and action of Red agent $j \in [N_2]$. The joint state and action of the Blue team and the Red team are denoted as $(\mathbf{X}^{N_1}_t, \mathbf{U}^{N_1}_t)$ and $(\mathbf{Y}^{N_2}_t, \mathbf{V}^{N_2}_t)$, respectively.

Definition 1. The *empirical distribution* (ED) for the Blue and Red teams are defined as

$$\mathcal{M}^{N_1}_t(x) = \frac{1}{N_1}\sum_{i=1}^{N_1} \mathbb{1}_x(X^{N_1}_{i,t}), \quad x \in \mathcal{X}, \tag{1a}$$

$$\mathcal{N}^{N_2}_t(y) = \frac{1}{N_2}\sum_{j=1}^{N_2} \mathbb{1}_y(Y^{N_2}_{j,t}), \quad y \in \mathcal{Y}. \tag{1b}$$

Notice that $\mathcal{M}^{N_1}_t(x)$ gives the fraction of Blue agents at state x. Hence, the random vector $\mathcal{M}^{N_1}_t = [\mathcal{M}^{N_1}(x)]_{x\in\mathcal{X}}$ is a probability measure, i.e., $\mathcal{M}^{N_1}_t \in \mathcal{P}(\mathcal{X})$. Similarly, we have $\mathcal{N}^{N_2}_t \in \mathcal{P}(\mathcal{Y})$. We use the following two operators to denote the operation in (1) that relates joint states to their corresponding EDs:

$$\mathcal{M}^{N_1}_t = \mathrm{Emp}_\mu(\mathbf{X}^{N_1}_t), \qquad \mathcal{N}^{N_2}_t = \mathrm{Emp}_\nu(\mathbf{Y}^{N_2}_t).$$

Note that the Emp operators remove agent index information and thus one *cannot* tell the state of a specific Blue agent i given only the Blue ED.

We use total variation to measure the distance between distributions. Formally, for a finite set E, the total variation between two probability measures $\mu, \mu' \in \mathcal{P}(E)$ is given by

$$\mathrm{d_{TV}}(\mu, \mu') = \frac{1}{2}\sum_{e\in E} |\mu(e) - \mu'(e)| = \frac{1}{2}\left\|\mu - \mu'\right\|_1.$$

Dynamics We consider weakly-coupled dynamics, where the dynamics of each individual agent is coupled with other agents through the EDs. For Blue agent i, its stochastic transition is governed by the transition kernel f_t and satisfies

$$\mathbb{P}(X_{i,t+1}^{N_1} = x_{i,t+1}^{N_1} | U_{i,t}^{N_1} = u_{i,t}^{N_1}, \mathbf{X}_t^{N_1} = \mathbf{x}_t^{N_1}, \mathbf{Y}_t^{N_2} = \mathbf{y}_t^{N_2}) \\ = f_t(x_{i,t+1}^{N_1} | x_{i,t}^{N_1}, u_{i,t}^{N_1}, \mu_t^{N_1}, \nu_t^{N_2}),$$

where $\mu_t^{N_1} = \mathrm{Emp}_\mu(\mathbf{x}_t^{N_1})$ and $\nu_t^{N_2} = \mathrm{Emp}_\nu(\mathbf{y}_t^{N_2})$. Similarly, the dynamics of Red agent j is governed by the transition kernel $g_t(y_{j,t+1}^{N_2} | y_{j,t}^{N_2}, v_{j,t}^{N_2}, \mu_t^{N_1}, \nu_t^{N_2})$.

Assumption 1 (Lipschitz Dynamics)**.** For all $x \in \mathcal{X}$, $\mu, \mu' \in \mathcal{P}(\mathcal{X})$, $\nu, \nu' \in \mathcal{P}(\mathcal{Y})$ and $t \in \{0, \dots, T-1\}$, there exist a constant $L_f \geq 0$ such that

$$\sum_{x' \in \mathcal{X}} |f_t(x'|x, u, \mu, \nu) - f_t(x'|x, u, \mu', \nu')| \\ \leq L_f \Big(\mathrm{d_{TV}}(\mu, \mu') + \mathrm{d_{TV}}(\nu, \nu') \Big).$$

We assume that g_t is L_g-Lipschitz as well.

Reward Structure Under the team-game framework, agents in the same team receive the same reward. Similar to the dynamics, we consider a weakly-coupled team reward

$$r_t : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to [-R_{\max}, R_{\max}].$$

Assumption 2 (Lipschitz Rewards)**.** For all $\mu, \mu' \in \mathcal{P}(\mathcal{X})$, $\nu, \nu' \in \mathcal{P}(\mathcal{Y})$ and $t \in \{0, \dots, T\}$, there exists $L_r \geq 0$ such that

$$|r_t(\mu, \nu) - r_t(\mu', \nu')| \leq L_r \big(\mathrm{d_{TV}}(\mu, \mu') + \mathrm{d_{TV}}(\nu, \nu') \big).$$

Under the zero-sum structure, we let the Blue team maximize the reward while the Red team minimizes it.

Information Structure We assume a mean-field sharing information structure (Arabneydi and Mahajan 2015). Specifically, at each time step t, Blue agent i observes its own state $X_{i,t}^{N_1}$ and the EDs $\mathcal{M}_t^{N_1}$ and $\mathcal{N}_t^{N_2}$. Similarly, Red agent j observes $Y_{j,t}^{N_2}$, $\mathcal{M}_t^{N_1}$ and $\mathcal{N}_t^{N_2}$. We consider the following mixed Markov policies:

$$\begin{aligned} \phi_{i,t} &: \mathcal{U} \times \mathcal{X} \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to [0, 1], \\ \psi_{j,t} &: \mathcal{V} \times \mathcal{Y} \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to [0, 1], \end{aligned} \tag{2}$$

where $\phi_{i,t}(u | X_{i,t}^{N_1}, \mathcal{M}_t^{N_1}, \mathcal{N}_t^{N_2})$ is the probability that Blue agent i selects action u given its state $X_{i,t}^{N_1}$ and the team EDs $\mathcal{M}_t^{N_1}$ and $\mathcal{N}_t^{N_2}$. Note that agent's individual state is the private information, while the team EDs are the common information shared among all agents.

An individual strategy is defined as a time sequence $\phi_i = \{\phi_{i,t}\}_{t=0}^T$. A Blue team strategy $\phi^{N_1} = \{\phi_i\}_{i=1}^{N_1}$ is the collection of individual strategies used by each Blue agent. We use Φ_t and Φ to denote, respectively, the set of individual policies and strategies available to each Blue agent. The set of Blue team strategies is then defined as the Cartesian product $\Phi^{N_1} = \times_{i=1}^{N_1} \Phi$. The notations extend naturally to the Red team.

In summary, an instance of a finite-population zero-sum mean-field team game is defined as the tuple $\texttt{ZS-MFTG} = \langle \mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{V}, f_t, g_t, r_t, N_1, N_2, T \rangle$.

Optimization Problem The performance of the team strategy pair (ϕ^{N_1}, ψ^{N_2}) is given by the expected cumulative reward

$$J^{N, \phi^{N_1}, \psi^{N_2}}(\mathbf{x}_0^{N_1}, \mathbf{y}_0^{N_2}) \\ = \mathbb{E}_{\phi^{N_1}, \psi^{N_2}} \left[\sum_{t=0}^{T} r_t(\mathcal{M}_t^{N_1}, \mathcal{N}_t^{N_2}) \Big| \mathbf{X}_0^{N_1} = \mathbf{x}_0^{N_1}, \mathbf{Y}_0^{N_2} = \mathbf{y}_0^{N_2} \right],$$

where $\mathcal{M}_t^{N_1} = \mathrm{Emp}_\mu(\mathbf{X}_t^{N_1})$ and $\mathcal{N}_t^{N_2} = \mathrm{Emp}_\nu(\mathbf{Y}_t^{N_2})$, and the expectation is with respect to the distribution of all system variables induced by ϕ^{N_1} and ψ^{N_2}.

When the Blue team considers its worst-case performance, we have the following max-min optimization:

$$\underline{J}^{N*}(\mathbf{x}_0^{N_1}; \mathbf{y}_0^{N_2}) = \max_{\phi^{N_1} \in \Phi^{N_1}} \min_{\psi^{N_2} \in \Psi^{N_2}} J^{N, \phi^{N_1}, \psi^{N_2}}(\mathbf{x}_0^{N_1}; \mathbf{y}_0^{N_2}), \tag{3}$$

where $\underline{J}^{N*}$ is the lower game value for the finite-population game. Note that the game value may not always exist, i.e., max-min value may differ from the min-max value (Elliott and Kalton 1972). Consequently, we consider the following optimality condition for the Blue team strategy.

Definition 2. A Blue team strategy $\phi^{N_1 *}$ is ϵ-optimal if

$$\underline{J}^{N*} \geq \min_{\psi^{N_2} \in \Psi^{N_2}} J^{N, \phi^{N_1 *}, \psi^{N_2}} \geq \underline{J}^{N*} - \epsilon.$$

The strategy $\phi^{N_1 *}$ is optimal if $\epsilon = 0$.

Similarly, the minimizing Red team considers a min-max optimization problem, which leads to the upper game value

$$\bar{J}^{N*} = \min_{\psi^{N_2} \in \Psi^{N_2}} \max_{\phi^{N_1} \in \Phi^{N_1}} J^{N, \phi^{N_1}, \psi^{N_2}}.$$

The ϵ-optimality of Red team strategies is defined similarly.

A ZS-MFTG Example

Consider a simple team game on a two-node graph in Figure 1. The state spaces are given by $\mathcal{X} = \{x^1, x^2\}$ and $\mathcal{Y} = \{y^1, y^2\}$, and the action spaces are $\mathcal{U} = \{u^1, u^2\}$ and $\mathcal{V} = \{v^1, v^2\}$. The Blue action u^1 corresponds to staying on the current node and u^2 represents moving to the other node. The same connotations apply to Red actions v^1 and v^2.

The maximizing Blue team's objective is to maximize its presence at node 2 (state x^2), and hence $r_t(\mu, \nu) = \mu(x^2)$. The Blue transition kernel at x^1 under u^2 is defined as

$$f_t(x^1 | x^1, u^2, \mu, \nu) = 0.5 \big(1 - \big(\rho \mu(x^1) - (1 - \rho) \nu(y^1) \big) \big), \\ f_t(x^2 | x^1, u^2, \mu, \nu) = 0.5 \big(1 + \big(\rho \mu(x^1) - (1 - \rho) \nu(y^1) \big) \big).$$

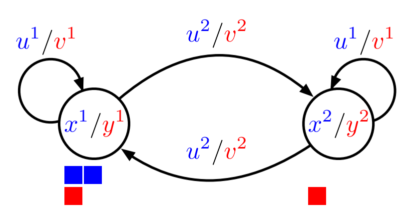


Figure 1: An example of ZS-MFTG over a two-node graph, where $N_1 = 2$, $N_2 = 2$ and $\rho = 0.5$.

Under this transition kernel, the probability of a Blue agent transitioning from node 1 to node 2 depends on the Blue team's numerical advantage over the Red team at node 1.

The initial joint states depicted in Figure 1 are given by $\mathbf{x}_0^2 = [x^1, x^1]$ and $\mathbf{y}_0^2 = [y^1, y^2]$. The corresponding EDs are $\mu_0^2 = [1, 0]$, $\nu_0^2 = [0.5, 0.5]$, and the running reward is $r_0 = \mu_0^2(x^2) = 0$. Suppose the Blue team applies a team strategy such that $\phi_0^i(u^2|x^1, \mu_0^2, \nu_0^2) = 1$ for both $i \in [2]$ (under which both Blue agents apply u^2). The probability of an individual Blue agent transitioning to node 2 is 0.625. Thus, the next Blue ED is a random vector with three possible realizations: (i) $\mathcal{M}_1^2 = [1, 0]$ with probability 0.14 (both Blue agents remain on node 1); (ii) $\mathcal{M}_1^2 = [0.5, 0.5]$ with probability 0.47 (one moves and one remains); and (iii) $\mathcal{M}_1^2 = [0, 1]$ with probability 0.39 (both move). Suppose the game terminates at $T = 1$, then the value under ϕ^2 is given by $J^{4,\phi^2,\psi^2}(\mathbf{x}_0^2, \mathbf{y}_0^2) = 0 + (0.14 \cdot 0 + 0.47 \cdot 0.5 + 0.39 \cdot 1) = 0.63$.

Infinite-Population Team Game

The preceding max-min optimization in (3) is intractable for large-population systems since the dimension of the joint policy spaces Φ^{N_1} and Ψ^{N_2} grows exponentially with the number of the agents. To address this scalability issue, we consider the infinite-population limit of the ZS-MFTGs, and further assume that agents in the same infinite-population team deploy the same strategy. As a result, we can model the behavior of an entire team as the distribution of a *typical agent*, i.e., the mean-field (Lasry and Lions 2007).

We first introduce the class of identical team strategies.

Definition 3 (Identical Blue Team Strategy)**.** The Blue team strategy $\phi^{N_1} = \{\phi_1, \dots, \phi_{N_1}\}$ is an identical team strategy, if $\phi_{i_1,t} = \phi_{i_2,t}$ for all $i_1, i_2 \in [N_1]$ and $t \in \{0, 1, \dots, T{-}1\}$.

We slightly abuse the notation and use ϕ to denote the identical Blue team strategy, under which all Blue agents apply the same individual strategy ϕ. Consequently, Φ is used to denote both the set of Blue individual strategies and the set of identical Blue team strategies. The definitions and notations extend to the identical Red team strategies.

We define the mean-field (MF) as the state distribution of a typical agent in an infinite-population team game.

Definition 4. Given identical team strategies $\phi \in \Phi$ and $\psi \in \Psi$, the MFs are defined as the sequence of vectors that adhere to the following deterministic dynamics with (μ_0^ρ, ν_0^ρ) as initial conditions:

$$\mu_{t+1}^\rho(x') = \sum_{x \in \mathcal{X}} \Big[\sum_{u \in \mathcal{U}} f_t(x'|x, u, \mu_t^\rho, \nu_t^\rho) \phi_t(u|x, \mu_t^\rho, \nu_t^\rho) \Big] \mu_t^\rho(x),$$
$$\nu_{t+1}^\rho(y') = \sum_{y \in \mathcal{Y}} \Big[\sum_{v \in \mathcal{V}} g_t(y'|y, v, \mu_t^\rho, \nu_t^\rho) \psi_t(v|y, \mu_t^\rho, \nu_t^\rho) \Big] \nu_t^\rho(y).$$

Later, in Theorem 1 we will show that the *deterministic* MF above is an approximation of the (stochastic) finite-population ED, and the approximation error goes to zero when $N_1, N_2 \to \infty$. Thus, we can regard the mean-field as the empirical distribution of an infinite-population team.

For simplicity, we express the MF dynamics in a compact matrix form as

$$\begin{aligned} \mu_{t+1}^\rho &= \mu_t^\rho F_t(\mu_t^\rho, \nu_t^\rho, \phi_t), \\ \nu_{t+1}^\rho &= \nu_t^\rho G_t(\mu_t^\rho, \nu_t^\rho, \psi_t), \end{aligned} \tag{4}$$

where $F_t \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$ is the transition matrix for a typical Blue agent under ϕ_t, which can be computed based on the transition kernel f_t. The matrix G_t is defined similarly.

Consider the infinite-population limit of the example in Figure 1 with $\mu_0^{0.5} = [1, 0]$, $\nu_0^{0.5} = [0.5, 0.5]$ and $\rho = 0.5$. If the Blue team applies the identical team strategy $\phi_0(u^2|x^1, \mu_0^{0.5}, \nu_0^{0.5}) = 1$, then the next Blue MF is deterministically given by $\mu_1^{0.5} = [0.375, 0.625]$.

For the infinite-population game, the performance of the identical team strategies $(\phi, \psi) \in \Phi \times \Psi$ is given by

$$J^{\rho,\phi,\psi}(\mu_0^\rho, \nu_0^\rho) = \sum_{t=0}^{T} r_t(\mu_t^\rho, \nu_t^\rho), \tag{5}$$

where the propagation of μ_t^ρ and ν_t^ρ is subject to (4).

The worst-case performance of the maximizing Blue team is then given by the lower game value

$$\underline{J}^{\rho*}(\mu_0^\rho, \nu_0^\rho) = \max_{\phi \in \Phi} \min_{\psi \in \Psi} J^{\rho,\phi,\psi}(\mu_0^\rho, \nu_0^\rho). \tag{6}$$

Remark 1. Different from the infinite-population game value (6), the finite-population value (3) takes joint states as arguments rather than EDs. The difference comes from the non-identical strategies considered in the finite-population game, which require each agent's state and index information to sample actions and predict the game's evolution.

Zero-Sum Game Between Coordinators

The mean-field sharing structure in (2) allows us to reformulate the *infinite*-population competitive team problem (6) as an equivalent two-player game from the perspective of two fictitious[1] coordinators. The coordinators know the common information (MFs) and selects a local policy that maps each agent's local information (individual state) to its actions. Through this common-information approach (Nayyar, Mahajan, and Teneketzis 2013), we provide a dynamic program that constructs optimal strategies for all agents under the original mean-field sharing information structure.

Equivalent Centralized Problem

We use $\pi_t : \mathcal{U} \times \mathcal{X} \to [0, 1]$ to denote a local Blue policy, which is *open-loop* with respect to the MFs. Specifically, $\pi_t(u|x)$ is the probability that a Blue agent selects action u at state x regardless of the current MFs. The set of open-loop Blue local policies is denoted as Π_t. Similarly, $\sigma_t : \mathcal{V} \times \mathcal{Y} \to [0, 1]$ and Σ_t denote a Red local policy and its admissible set. Under the local policy π_t, the Blue MF propagates as

$$\mu_{t+1}^\rho(x') = \sum_{x \in \mathcal{X}} \Big[\sum_{u \in \mathcal{U}} f_t(x'|x, u, \mu_t^\rho, \nu_t^\rho) \pi_t(u|x) \Big] \mu_t^\rho(x), \tag{7}$$

and the Red team MF dynamics under Red local policies is defined similarly.

At each time t, a Blue coordinator observes the MFs of both teams (common information) and prescribes a local

[1] The coordinators are fictitious since they are introduced as an auxiliary concept and are not required for the actual implementation of the obtained strategies.

policy $\pi_t \in \Pi_t$ to all Blue agents within its team. The local policy is selected based on:

$$\pi_t = \alpha_t\big(\mu_t^\rho, \nu_t^\rho\big),$$

where $\alpha_t : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \Pi_t$ is a deterministic Blue *coordination policy*. Similarly, the Red coordinator observes the MFs and selects a local policy $\sigma_t \in \Sigma_t$ according to $\sigma_t = \beta_t\big(\mu_t^\rho, \nu_t^\rho\big)$. We refer to the time sequence $\alpha = \big(\alpha_1, \ldots, \alpha_{T-1}\big)$ as the Blue team *coordination strategy* and $\beta = \big(\beta_1, \ldots, \beta_{T-1}\big)$ as the Red team coordination strategy. The sets of admissible coordination strategies are denoted as $\mathcal{A}$ and $\mathcal{B}$.

Remark 2. There is a one-to-one correspondence between Blue (Red) coordination strategies and *identical* Blue (Red) team strategies such that

$$\phi_t(u|x, \mu, \nu) = \underbrace{\big[\alpha_t(\mu, \nu)\big]}_{\pi_t}(u|x).$$

The original competitive team problem in (6) can now be viewed as an equivalent zero-sum game played between the two coordinators, where the game state is the joint MF (μ_t^ρ, ν_t^ρ) and the actions are the local policies π_t and σ_t selected by the coordinators. The cumulative reward is computed based on the mean-field trajectory induced by the coordination strategy pair (α, β). The objective of the Blue coordinator is to maximize the cumulative rewards, while the Red coordinator minimizes.

In summary, the zero-sum coordinator game is defined as the tuple $\texttt{ZS-CG} = \langle \mathcal{P}(\mathcal{X}), \mathcal{P}(Y), \Pi_t, \Sigma_t, F_t, G_t, r_t, \rho, T\rangle$, where the state and action spaces are all continuous.

We use the following dynamic programming recursion scheme to find the lower value of the coordinator game:

$$\underline{J}_{\mathrm{cor},T}^{\rho*}(\mu_T^\rho, \nu_T^\rho) = r_T(\mu_T^\rho, \nu_T^\rho) \tag{8}$$

$$\underline{J}_{\mathrm{cor},t}^{\rho*}(\mu_t^\rho, \nu_t^\rho) = r_t(\mu_t^\rho, \nu_t^\rho) \qquad \text{for } t = 0, \ldots, T-1 \tag{9}$$
$$+ \max_{\pi_t \in \Pi_t} \min_{\sigma_t \in \Sigma_t} \underline{J}_{\mathrm{cor},t+1}^{\rho*}\big(\mu_t^\rho F_t(\mu_t^\rho, \nu_t^\rho, \pi_t), \nu_t^\rho G_t(\mu_t^\rho, \nu_t^\rho, \sigma_t)\big).$$

With the optimal value function, the optimal Blue team coordination strategy can be constructed via

$$\alpha_t^*(\mu_t^\rho, \nu_t^\rho) \in \tag{10}$$
$$\operatorname*{argmax}_{\pi_t \in \Pi_t} \min_{\sigma_t \in \Sigma_t} \underline{J}_{\mathrm{cor},t+1}^{\rho*}\big(\mu_t^\rho F_t(\mu_t^\rho, \nu_t^\rho, \pi_t), \nu_t^\rho G_t(\mu_t^\rho, \nu_t^\rho, \sigma_t)\big).$$

Note that the optimal Blue team coordination strategy induces an identical Blue team strategy that satisfies the mean-field sharing information structure and can be implemented in the finite-population game (Remark 2).

Reachable Sets

At the infinite-population limit, the MF dynamics is deterministic, and thus selecting the local policies π_t and σ_t at time t is equivalent to selecting the desirable MFs at the next time step. Consequently, we examine the set of MFs that can be reached from the current MFs.

Definition 5. The Blue and Red team reachable sets, starting from μ_t^ρ and ν_t^ρ, are defined as

$$\mathcal{R}_{\mu,t}(\mu_t^\rho, \nu_t^\rho) \triangleq \{\mu_{t+1}^\rho | \exists \pi_t \in \Pi_t \text{ s.t.} \, \mu_{t+1}^\rho = \mu_t^\rho F_t(\mu_t^\rho, \nu_t^\rho, \pi_t)\}.$$
$$\mathcal{R}_{\nu,t}(\mu_t^\rho, \nu_t^\rho) \triangleq \{\nu_{t+1}^\rho | \exists \sigma_t \in \Sigma_t \text{ s.t.} \, \nu_{t+1}^\rho = \nu_t^\rho G_t(\mu_t^\rho, \nu_t^\rho, \sigma_t)\}.$$

In the sequel, we regard the reachable sets as correspondences, i.e., set-valued functions (Freeman and Kokotovic 2008).

Remark 3. Note that the reachable sets are constructed based on identical team strategies, since under the coordinator game formulation, all agents in the same team follow the same local policies prescribed by their coordinator.

Now, we can change the optimization domains in (9) from the local policy spaces to the reachable sets as follows

$$\underline{J}_{\mathrm{cor},t}^{\rho*}(\mu_t^\rho, \nu_t^\rho) = r_t(\mu_t^\rho, \nu_t^\rho) \tag{11}$$
$$+ \max_{\mu_{t+1}^\rho \in \mathcal{R}_{\mu,t}(\mu_t^\rho, \nu_t^\rho)} \min_{\nu_{t+1}^\rho \in \mathcal{R}_{\nu,t}(\mu_t^\rho, \nu_t^\rho)} \underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mu_{t+1}^\rho, \nu_{t+1}^\rho).$$

In the sequel, we primarily work with the reachability-based optimization in (11). There are two advantages to this approach: First, the reachable sets generally have a lower dimension than the coordinator action spaces, which is desirable for numerical algorithms; Second, the reachability-based optimization allows us to compare the "reachability" induced by non-identical and identical team strategies (Theorem 1) and then study the performance loss due to the identical strategy assumption.

Main Results

Recall that the optimal Blue team coordination strategy α^* is constructed for the infinite-population game assuming that both teams employ identical team strategies. This section establishes the performance guarantees for α^* in the finite-population games where both teams are allowed to deploy non-identical strategies.

Approximation Error

As α^* is solved at the infinite-population limit, it is essential to understand how well the infinite-population game approximates the original finite-population problem. The following theorem states that the reachable set constructed using identical strategies is rich enough to approximate any empirical distributions induced by *non-identical* team strategies in finite-population games.

Theorem 1. *Let $\mathbf{X}_t^{N_1}$, $\mathbf{Y}_t^{N_2}$, $\mathcal{M}_t^{N_1}$, and $\mathcal{N}_t^{N_2}$ be the joint states and the corresponding EDs of a finite-population game. Denote the next Blue team ED induced by a (potentially non-identical) Blue team policy $\phi_t^{N_1} \in \Phi_t^{N_1}$ as $\mathcal{M}_{t+1}^{N_1}$. Then, there exists $\mu_{t+1} \in \mathcal{R}_{\mu,t}(\mathcal{M}_t^{N_1}, \mathcal{N}_t^{N_2})$ such that*

$$\mathbb{E}_{\phi_t^{N_1}}\Big[\mathrm{d_{TV}}\big(\mathcal{M}_{t+1}^{N_1}, \mu_{t+1}\big)\big|\mathbf{X}_t^{N_1}, \mathbf{Y}_t^{N_2}\Big] \leq \frac{|\mathcal{X}|}{2}\sqrt{\frac{1}{N_1}}. \tag{12}$$

Proof. The key idea is to construct an identical local policy $\pi_{\mathrm{apprx},t}$ that has its action distribution matching the average of the policies used by the Blue agents. One can then leverage $\pi_{\mathrm{apprx},t}$ to mimic the population behavior and use a modified law of large numbers to show that the MF μ_{t+1} induced by $\pi_{\mathrm{apprx},t}$ satisfies the error bound in (12). This idea is visualized in Figure 2. See the full version (Guan, Afshari, and Tsiotras 2023) for a detailed proof. ☐

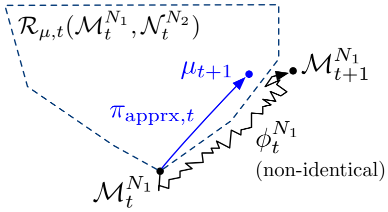


Figure 2: An illustration of the key idea behind Theorem 1.

Corollary 1. *Let $\mathbf{X}_t^{N_1}$, $\mathbf{Y}_t^{N_2}$, $\mathcal{M}_t^{N_1}$, and $\mathcal{N}_t^{N_2}$ be the joint states and the corresponding EDs of a finite-population game. Denote the next Blue ED induced by an identical Blue team policy $\phi_t \in \Phi_t$ as $\mathcal{M}_{t+1}^{N_1}$. Then, the following holds:*

$$\mathbb{E}_{\phi_t}\left[\mathrm{d_{TV}}\big(\mathcal{M}_{t+1}^{N_1}, \mu_{t+1}\big)\,\big|\, \mathbf{X}_t^{N_1}, \mathbf{Y}_t^{N_2}\right] \leq \frac{|\mathcal{X}|}{2}\sqrt{\frac{1}{N_1}},$$

where $\mu_{t+1} = \mathcal{M}_t^{N_1} F_t(\mathcal{M}_t^{N_1}, \mathcal{N}_t^{N_2}, \phi_t)$.

Corollary 1 implies that the MF induced by an identical team policy is a good approximation to the ED induced by the same identical team policy in a finite-population system.

Lipschitz Continuity of the Value Functions

Next, we examine the continuity of the coordinator game values, which is essential for the performance guarantees. We start with the continuity of the reachability correspondences under the Hausdorff distance $\mathrm{dist_H}$.[2]

Lemma 1. *For all $\mu_t, \mu_t' \in \mathcal{P}(\mathcal{X})$ and $\nu_t, \nu_t' \in \mathcal{P}(\mathcal{Y})$, the reachability correspondence $\mathcal{R}_{\mu,t}$ satisfies*

$$\begin{aligned}&\mathrm{dist_H}(\mathcal{R}_{\mu,t}(\mu_t,\nu_t), \mathcal{R}_{\mu,t}(\mu_t',\nu_t')) \\ &\qquad \leq L_{R_\mu}\big(\mathrm{d_{TV}}\big(\mu_t,\mu_t'\big) + \mathrm{d_{TV}}\big(\nu_t,\nu_t'\big)\big),\end{aligned} \tag{13}$$

where the Lipschitz constant is given by $L_{R_\mu} = 1 + \frac{1}{2}L_f$. The Red reachability correspondence satisfies a similar inequality with a Lipschitz constant $L_{R_\nu} = 1 + \frac{1}{2}L_g$.

Leveraging the continuity of the reachability correspondences, the following theorem establishes the Lipschitz continuity of the optimal coordinator game value.

Theorem 2. *For all $\mu_t^\rho, \mu_t^{\rho'} \in \mathcal{P}(\mathcal{X})$ and $\nu_t^\rho, \nu_t^{\rho'} \in \mathcal{P}(\mathcal{Y})$, the lower coordinator game value satisfies*

$$\begin{aligned}&\big|\underline{J}_{\mathrm{cor},t}^{\rho *}(\mu_t^\rho,\nu_t^\rho) - \underline{J}_{\mathrm{cor},t}^{\rho *}(\mu_t^{\rho'},\nu_t^{\rho'})\big| \\ &\qquad \leq L_{J,t}\big(\mathrm{d_{TV}}\big(\mu_t^\rho,\mu_t^{\rho'}\big) + \mathrm{d_{TV}}\big(\mu_t^\rho,\nu_t^{\rho'}\big)\big),\end{aligned} \tag{14}$$

where the Lipschitz constant is given by $L_{J,t} = L_r\big(1 + L_{\mathcal{R}}(1-L_{\mathcal{R}}^{T-t})/(1-L_{\mathcal{R}})\big)$ and $L_{\mathcal{R}} = L_{R_\mu} + L_{R_\nu}$.

Proof. Observe that the lower value in (11) takes the form: $f(x,y) = c(x,y) + \max_{p\in\Gamma(x,y)} \min_{q\in\Theta(x,y)} g(p,q)$, which is an extension of the maximization marginal function (Freeman and Kokotovic 2008) to the max-min case. We present a continuity result for this type of marginal function in the extended version of this paper (Guan, Afshari, and Tsiotras 2023), based on which we can prove the above theorem through an inductive argument. □

[2] The Hausdorff distance between sets $A, B \subseteq \mathcal{X}$ is defined as $\mathrm{dist_H}(A,B) = \max\big\{\sup_{a\in A}\inf_{b\in B}\|a-b\|, \sup_{b\in B}\inf_{a\in A}\|a-b\|\big\}$.

Performance Guarantees

The following main theorem compares the worst-case performance of the identical Blue team strategy induced by α^* (Remark 2) to the original max-min optimization in (3), where non-identical strategies are allowed.

Theorem 3. *The optimal Blue coordination strategy α^* in* (10) *induces an ϵ-optimal Blue team strategy. Formally, for all $\mathbf{x}^{N_1} \in \mathcal{X}^{N_1}$ and $\mathbf{y}^{N_2} \in \mathcal{Y}^{N_2}$,*

$$\begin{aligned}\underline{J}^{N*}(\mathbf{x}^{N_1}, \mathbf{y}^{N_2}) &\geq \min_{\psi^{N_2}\in\Psi^{N_2}} J^{N,\alpha^*,\psi^{N_2}}(\mathbf{x}^{N_1}, \mathbf{y}^{N_2}) \\ &\geq \underline{J}^{N*}(\mathbf{x}^{N_1}, \mathbf{y}^{N_2}) - \mathcal{O}\Big(\frac{1}{\sqrt{\underline{N}}}\Big),\end{aligned} \tag{15}$$

where $\underline{N} = \min\{N_1, N_2\}$.

Proof. The first inequality is straightforward. We break the second inequality into two steps: (i) $\min_{\psi^{N_2}} J^{N,\alpha^*,\psi^{N_2}} \geq \underline{J}_{\mathrm{cor}}^{\rho *} - \mathcal{O}(1/\sqrt{\underline{N}})$; and (ii) $\underline{J}_{\mathrm{cor}}^{\rho *} \geq \underline{J}^{N*} - \mathcal{O}(1/\sqrt{\underline{N}})$. The proofs for both steps are constructed based on induction, and we only present the proof for step (i) in the appendix of this paper. A detailed proof of Theorem 3 is presented in the extended version (Guan, Afshari, and Tsiotras 2023). □

Remark 4. Recall that α^* is solved at the *infinite-population* limit under the restriction that both teams apply *identical* team strategies. Theorem 3 states that the *identical* Blue team strategy induced by α^* is still ϵ-optimal, even if (i) it is deployed in a *finite-population* game and (ii) the opponent team employs *non-identical* strategies to exploit.

Remark 5. Continuity Assumptions 1 and 2 are necessary to translate the infinite-population performance back to the finite-population game. See the appendix of (Guan, Afshari, and Tsiotras 2023) for a discontinuous example where the infinite-population game value is different from that of the finite-population problem.

Numerical Examples

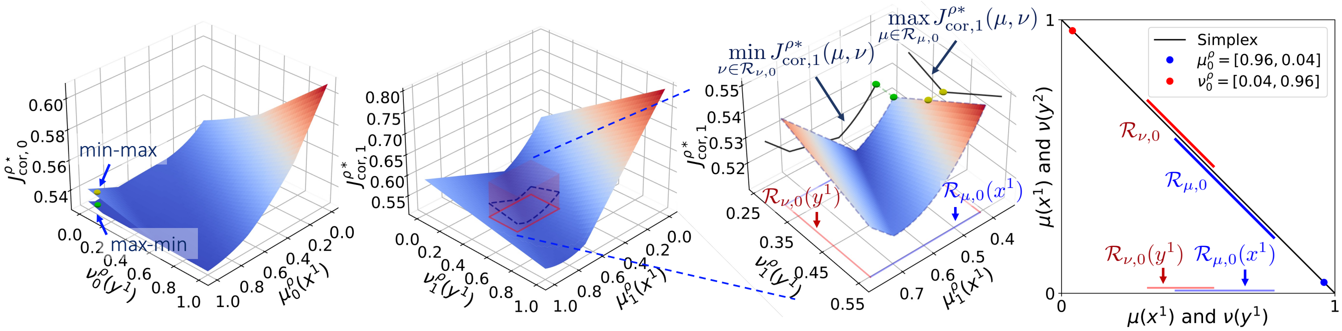
In this section, we provide two numerical examples. For both examples, the state spaces are $\mathcal{X} = \{x^1, x^2\}$ and $\mathcal{Y} = \{y^1, y^2\}$, and the action spaces are $\mathcal{U} = \{u^1, u^2\}$ and $\mathcal{V} = \{v^1, v^2\}$. The two-state state spaces allow the *joint* MFs to be characterized solely by $\mu_t(x^1)$ and $\nu_t(y^1)$.

Numerical Example 1

This example serves to illustrate the reachability-based optimization in equation (11) and to demonstrate that the coordinator game value may not exist, contrary to the continuous setting as discussed in (Sanjari, Saldi, and Yüksel 2023). For a comprehensive description of the dynamics and reward setup, see (Guan, Afshari, and Tsiotras 2023).

The coordinator game values in Figure 3 are computed through discretization, where the two-dimensional simplexes $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ are meshed into 1,000 bins[3]. While the game value $J_{\mathrm{cor},1}^{\rho *}$ exists at time $t = 1$, it is not convex-concave. Hence, the upper (max-min) and lower (min-max)

[3] An error bound on the difference between the discretized value and the true optimal value can be readily derived using the Lipschitz constants of the coordinator game values.



(a) $t = 0$ (b) $t = 1$ (c) $t = 1$ (d)

Figure 3: Subplots (a)-(c) present the game values computed via discretization. The x- and y-axes correspond to $\mu_t^\rho(x^1)$ and $\nu_t^\rho(y^1)$, respectively. Subplot (d) illustrates the reachable sets starting from $\mu_0 = [0.96, 0.04]$ and $\nu_0 = [0.04, 0.96]$.

game values at the previous step $t = 0$ differs, as observed in subplot (a). Specifically, at $\mu_0^\rho = [0.96, 0.04]$ and $\nu_0^\rho = [0.04, 0.96]$, we have the lower value $\underline{J}^{\rho*}_{\text{cor},0} = 0.5298$ and the upper value $\bar{J}^{\rho*}_{\text{cor},0} = 0.5384$, which are visualized as the green and yellow points. This discrepancy in the game values implies the absence of a Nash equilibrium in this coordinator game.

Based on (11), the optimization domains of $J^{\rho*}_{\text{cor},0}$ are $\mathcal{R}_{\mu,0}(\mu_0^\rho, \nu_0^\rho)$ for the maximization and $\mathcal{R}_{\nu,0}(\mu_0^\rho, \nu_0^\rho)$ for the minimization, both of which are plotted in (d). Subplot (c) presents a zoom-in for the optimization $\max_{\mathcal{R}_{\mu,0}} \min_{\mathcal{R}_{\nu,0}} J^{\rho*}_{\text{cor},1}$ and its min-max counterpart. The marginal functions are also plotted, from which the max-min and min-max values at $t = 0$ can be directly obtained.

Remark 6. In the extended version (Guan, Afshari, and Tsiotras 2023), we show that a Nash equilibrium exists when agents' dynamics are completely decoupled from each other.

Numerical Example 2

It is generally challenging to verify the suboptimality bound in Theorem 3, since computing the true optimal performance of a finite-population team game is intractable. However, for the following designed example, we have the optimal team strategies for large finite-population teams.

Consider a ZS-MFTG with $T=2$. The game setup is similar to the one in Figure 1 but with different dynamics and rewards. The (minimizing) Red team's objective is to maximize its presence at state y^1 at $t = 2$, which translates to

$$r_0(\mu, \nu) = r_1(\mu, \nu) = 0, \qquad r_2(\mu, \nu) = -\nu(y^1).$$

The Blue transition is time-invariant, deterministic, and independent of the MFs. Formally, for all μ, ν and $t \in \{0, 1\}$,

$$\begin{aligned} &f_t(x^1|x^1, u^1, \mu, \nu) = 1, \;\; f_t(x^2|x^1, u^2, \mu, \nu) = 1, \\ &f_t(x^2|x^2, u^1, \mu, \nu) = 1, \;\; f_t(x^1|x^2, u^2, \mu, \nu) = 1. \end{aligned} \tag{16}$$

Under the above transition kernel, a Blue agent can freely move between the two nodes (action u^1 to "stay" on the current node and u^2 to "switch" node).

At $t = 0$, all Red agents are frozen at both states, i.e., no action can change a Red agent's state. At $t = 1$, Red agents at y^1 are still frozen, but Red agents at y^2 can use v^2 to transition to y^1 and the transition probability is given by

$$\begin{aligned} &g_1^\rho(y^1|y^2, v^2, \mu_1, \nu_1) \\ &= \min\Big\{5\Big((\mu_1(x^1) - \frac{1}{\sqrt{2}})^2 + (\mu_1(x^2) - (1 - \frac{1}{\sqrt{2}}))^2\Big), 1\Big\}. \end{aligned} \tag{17}$$

Note that, under the above dynamics, if the Blue team achieves the target distribution $\mu_1 = [1/\sqrt{2}, 1 - 1/\sqrt{2}]$ at time $t = 1$, no Red agent can transition from y^2 to y^1.

Infinite-population case. For the Red team, only the decisions of Red agents at y^2 at time $t = 1$ have an impact on the game outcome. As a result, the above setup leads to a dominant optimal Red team strategy: all Red team agents at y^2 use v^2 at $t = 1$ and attempt to transit to state y^1. On the other hand, the Blue team should try to achieve the distribution $\mu_1 = [1/\sqrt{2}, 1 - 1/\sqrt{2}]$ to minimize the probability of Red team agents transitioning from y^2 to y^1 at $t=1$. The Blue dynamics in (16) ensures that the Blue team reachable set covers the entire simplex $\mathcal{P}(\mathcal{X})$ regardless of the initial distributions. Hence, the target distribution can always be achieved at $t=1$ with an *infinite* population.

Under the optimal strategies discussed above, the Blue team completely blocks the Red team agents' migration from y^2 to y^1, and thus only the Red agents spawn on y^1 will count towards the terminal rewards. Consequently, the infinite-population game value is given by $J^{\rho*} = -\nu_0(y^1)$.

Finite-population case. The Red team's optimal strategy remains the same as the infinite-population case. Note that this Red team strategy is optimal against all Blue team strategies. The Blue team, on the other hand, cannot achieve the *irrational* target distribution with a finite number of agents. While the Blue team can still match the target distribution *in expectation* using a (stochastic) identical team strategy, the following analysis shows that a non-identical deterministic Blue team strategy achieves a better performance.

Consider a Blue team with three agents and all Blue agents are on node 1, i.e., $\mu_0^3 = [1, 0]$. The optimal Blue coordination strategy prescribes that all Blue agents pick u^1 ("stay") with probability $1/\sqrt{2}$ and u^2 ("move to x^2") with probability $(1 - 1/\sqrt{2})$ to reach the target distribution in expectation. Such action selection leads to the following four possible outcomes of the next Blue team ED μ_1^3: $\mathbb{P}([1,0]) = 0.354$, $\mathbb{P}([2/3,1/3]) = 0.439$, $\mathbb{P}([1/3,2/3]) = 0.182$, and $\mathbb{P}([0,1]) = 0.025$. In expectation, these empirical distributions lead to a transition probability of 0.518 for a Red team agent moving from y^2 to y^1. Consequently, we have the worst-case performance of the optimal Blue coordinator strategy as $\min_{\psi^{N_2}} J^{3,\alpha^*,\psi^{N_2}} = -\nu_0(y^1) - 0.518\nu_0(y^2)$.

Next, consider the non-identical deterministic Blue team strategy, under which Blue agents 1 and 2 apply action u^1 and Blue agent 3 applies u^2. This Blue team strategy deterministically leads to $\mathcal{M}_1^3 = [2/3, 1/3]$ at $t = 1$, and the resultant Red team transition probability from y^2 to y^1 is 0.016. Clearly, the non-identical Blue team strategy significantly outperforms the identical mixed team strategy in this *three-agent* case. Furthermore, this Blue team strategy is optimal over the entire non-identical Blue team strategy set, resulting in a finite-population optimal game value $J^{3*} = -\nu_0(y^1) - 0.016\nu_0(y^2)$.

We repeat the above computation for multiple Blue team size N_1 and plot the suboptimality gap as the blue line in Figure 4, which verifies the $\mathcal{O}(1/\sqrt{\underline{N}})$ decrease rate predicted by Theorem 3.

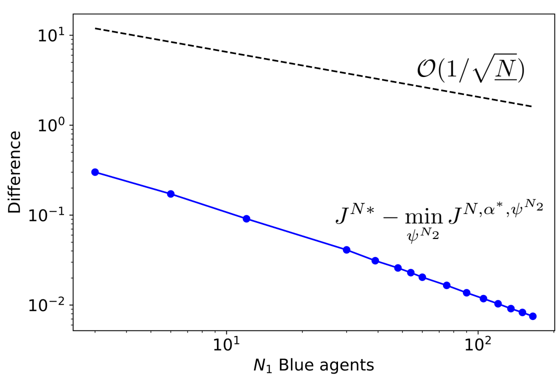


Figure 4: Performance loss of the optimal Blue coordination strategy with $\mu_0 = [1, 0]$ and $\nu_0 = [0.4, 0.6]$.

Conclusion

In this work, we introduced a discrete zero-sum mean-field team game to model the behaviors of competing large-population teams. We developed a dynamic programming approach that approximately solves this large-population game at its infinite-population limit where only identical team strategies are considered. Our analysis demonstrated that the identical strategies constructed are ϵ-optimal within the general class of non-identical team strategies when deployed in the original finite-population game. The derived performance guarantees are verified through numerical examples. Future work will investigate the LQG setup of this problem and explore machine-learning techniques to address more complex zero-sum mean-field team problems. Additionally, we aim to generalize our results to the infinite-horizon discounted case and problems with heterogeneous sub-populations.

Appendix

Proof of Theorem 3. We only provide an inductive proof for step (i). Fix an arbitrary Red team strategy $\psi^{N_2} \in \Psi^{N_2}$.
Base case: At time T, we have for all $\mathbf{x}_T^{N_1}$ and $\mathbf{y}_T^{N_2}$ that

$$J_T^{N,\alpha^*,\psi^{N_2}}(\mathbf{x}_T^{N_1}, \mathbf{y}_T^{N_2}) = \underline{J}_{\mathrm{cor},T}^{\rho*}(\mu_T^{N_1}, \nu_T^{N_1}) = r_T(\mu_T^{N_1}, \nu_T^{N_2}),$$

where $\mu_T^{N_1} = \mathrm{Emp}_\mu(\mathbf{x}_T^{N_1})$ and $\nu_T^{N_2} = \mathrm{Emp}_\nu(\mathbf{y}_T^{N_2})$.
Inductive hypothesis: Assume that for all $\mathbf{x}_{t+1}^{N_1}$ and $\mathbf{y}_{t+1}^{N_2}$,

$$J_{t+1}^{N,\alpha^*,\psi^{N_2}}(\mathbf{x}_{t+1}^{N_1}, \mathbf{y}_{t+1}^{N_2}) \geq \underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mu_{t+1}^{N_1}, \nu_{t+1}^{N_2}) - \mathcal{O}\Big(\frac{1}{\sqrt{\underline{N}}}\Big).$$

Induction: At timestep t, consider arbitrary joint states $(\mathbf{x}_t^{N_1}, \mathbf{y}_t^{N_2})$ and the corresponding EDs $(\mu_t^{N_1}, \nu_t^{N_2})$. Define

$$\mu_{t+1}^* = \mu_t^{N_1} F_t(\mu_t^{N_1}, \nu_t^{N_2}, \alpha_t^*).$$

Note that, from the optimality of α_t^* in (10), we have

$$\mu_{t+1}^* \in \operatorname*{argmax}_{\mu_{t+1} \in \mathcal{R}_{\mu,t}(\mu_t^{N_1}, \nu_t^{N_2})} \min_{\nu_{t+1} \in \mathcal{R}_{\nu,t}(\mu_t^{N_1}, \nu_t^{N_2})} \underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mu_{t+1}, \nu_{t+1}). \tag{18}$$

Furthermore, from Theorem 1, there exists a $\nu_{\mathrm{apprx},t+1} \in \mathcal{R}_{\nu,t}(\mu_t^{N_1}, \nu_t^{N_2})$ for the fixed Red policy $\psi_t^{N_2}$ such that

$$\mathbb{E}_{\psi_t^{N_2}}\Big[\mathrm{d_{TV}}\big(\mathcal{N}_{t+1}^{N_2}, \nu_{\mathrm{apprx},t+1}\big)\Big] \leq \frac{|\mathcal{Y}|}{2}\sqrt{\frac{1}{N_2}}. \tag{19}$$

Then, for all $\mathbf{x}_t^{N_1} \in \mathcal{X}^{N_1}$ and $\mathbf{y}_t^{N_2} \in \mathcal{Y}^{N_2}$, we have

$$\begin{aligned}
&J_t^{N,\alpha^*,\psi^{N_2}}(\mathbf{x}_t^{N_1}, \mathbf{y}_t^{N_2}) \\
&= r_t(\mu_t^{N_1}, \nu_t^{N_2}) + \mathbb{E}_{\alpha^*,\psi^{N_2}}\Big[J_{t+1}^{N,\alpha^*,\psi^{N_2}}(\mathbf{X}_{t+1}^{N_1}, \mathbf{Y}_{t+1}^{N_2})\Big] \\
&\overset{\text{(i)}}{\geq} r_t(\mu_t^{N_1}, \nu_t^{N_2}) + \mathbb{E}_{\alpha^*,\psi^{N_2}}\Big[\underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mathcal{M}_{t+1}^{N_1}, \mathcal{N}_{t+1}^{N_2})\Big] - \mathcal{O}\Big(\frac{1}{\sqrt{\underline{N}}}\Big) \\
&= r_t(\mu_t^{N_1}, \nu_t^{N_2}) - \mathcal{O}\Big(\frac{1}{\sqrt{\underline{N}}}\Big) + \mathbb{E}_{\alpha^*,\psi^{N_2}}\Big[\underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mathcal{M}_{t+1}^{N_1}, \mathcal{N}_{t+1}^{N_2}) \\
&\quad - \underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mu_{t+1}^*, \nu_{\mathrm{apprx},t+1}) + \underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mu_{t+1}^*, \nu_{\mathrm{apprx},t+1})\Big] \\
&\overset{\text{(ii)}}{\geq} r_t(\mu_t^{N_1}, \nu_t^{N_2}) + \underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mu_{t+1}^*, \nu_{\mathrm{apprx},t+1}) - \mathcal{O}\Big(\frac{1}{\sqrt{\underline{N}}}\Big) \\
&\quad - L_{J,t+1}\Big(\underbrace{\mathbb{E}_{\alpha^*}[\mathrm{d_{TV}}(\mathcal{M}_{t+1}^{N_1}, \mu_{t+1}^*)]}_{\mathcal{O}(1/\sqrt{N_1}) \text{ due to Corollary 1}} + \underbrace{\mathbb{E}_{\psi^{N_2}}[\mathrm{d_{TV}}(\mathcal{N}_{t+1}^{N_2}, \nu_{\mathrm{apprx},t+1})]}_{\mathcal{O}(1/\sqrt{N_2}) \text{ due to (19)}}\Big) \\
&\overset{\text{(iii)}}{\geq} r_t(\mu_t^{N_1}, \nu_t^{N_2}) + \underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mu_{t+1}^*, \nu_{\mathrm{apprx},t+1}) - \mathcal{O}\Big(\frac{1}{\sqrt{\underline{N}}}\Big) \\
&\overset{\text{(iv)}}{\geq} r_t(\mu_t^{N_1}, \nu_t^{N_2}) + \min_{\nu_{t+1} \in \mathcal{R}_{\nu,t}(\mu_t^{N_1}, \nu_t^{N_2})} \underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mu_{t+1}^*, \nu_{t+1}) - \mathcal{O}\Big(\frac{1}{\sqrt{\underline{N}}}\Big) \\
&\overset{\text{(v)}}{=} \underline{J}_{\mathrm{cor},t}^{\rho*}(\mu_t^{N_1}, \nu_t^{N_2}) - \mathcal{O}\Big(\frac{1}{\sqrt{\underline{N}}}\Big).
\end{aligned}$$

Inequality (i) is due to the inductive hypothesis; inequality (ii) leverages the Lipschitz continuity of the coordinator value function (Theorem 2); inequality (iii) bounds the error terms using Theorem 1 and Corollary 1; inequality (iv) is due to the fact that $\nu_{\mathrm{apprx},t+1}$ is in the reachable set; equality (v) comes from the optimality of μ_{t+1}^* in (18).

Finally, since $\psi^{N_2} \in \Psi^{N_2}$ is arbitrary, we have

$$\min_{\psi^{N_2} \in \psi^{N_2}} J_0^{N,\alpha^*,\psi^{N_2}}(\mathbf{x}^{N_1}, \mathbf{y}^{N_2}) \geq \underline{J}_{\mathrm{cor}}^{\rho*}(\mu^{N_1}, \nu^{N_2}) - \mathcal{O}\Big(\frac{1}{\sqrt{\underline{N}}}\Big),$$

thus completing the proof. ☐

References

Arabneydi, J.; and Mahajan, A. 2014. Team optimal control of coupled subsystems with mean-field sharing. In *53rd IEEE Conference on Decision and Control*, 1669–1674. Los Angeles, CA.

Arabneydi, J.; and Mahajan, A. 2015. Team-optimal solution of finite number of mean-field coupled LQG subsystems. In *54th IEEE Conference on Decision and Control*, 5308–5313. Osaka, Japan, Dec. 15–18, 2015.

Bernstein, D. S.; Givan, R.; Immerman, N.; and Zilberstein, S. 2002. The complexity of decentralized control of Markov decision processes. *Mathematics of Operations Research*, 27(4): 819–840.

Elliott, R. J.; and Kalton, N. J. 1972. *The Existence of Value in Differential Games*, volume 126. American Mathematical Soc.

Freeman, R.; and Kokotovic, P. V. 2008. *Robust Nonlinear Control Design: State-space and Lyapunov Techniques*. Springer Science & Business Media.

Gibbons, R.; Roberts, J.; et al. 2013. *The Handbook of Organizational Economics*. Princeton University Press Princeton, NJ.

Guan, Y.; Afshari, M.; and Tsiotras, P. 2023. Zero-Sum Games between Mean-Field Teams: A Common Information and Reachability based Analysis. *arXiv preprint arXiv:2303.12243*.

Huang, M.; Caines, P. E.; and Malhamé, R. P. 2007. Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ϵ-Nash equilibria. *IEEE Transactions on Automatic Control*, 52(9): 1560–1571.

Huang, M.; Malhamé, R. P.; and Caines, P. E. 2006. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information & Systems*, 6(3): 221–252.

Lasry, J.-M.; and Lions, P.-L. 2007. Mean field games. *Japanese Journal of Mathematics*, 2(1): 229–260.

Laurière, M.; Perrin, S.; Geist, M.; and Pietquin, O. 2022. Learning Mean Field Games: A Survey. *arXiv preprint arXiv:2205.12944*.

Li, J.; Tinka, A.; Kiesel, S.; Durham, J. W.; Kumar, T. S.; and Koenig, S. 2021. Lifelong multi-agent path finding in large-scale warehouses. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 11272–11281.

Mahajan, A.; and Nayyar, A. 2015. Sufficient Statistics for Linear Control Strategies in Decentralized Systems With Partial History Sharing. *IEEE Transactions on Automatic Control*, 60(8): 2046–2056.

Nayyar, A.; Mahajan, A.; and Teneketzis, D. 2013. Decentralized stochastic control with partial history sharing: A common information approach. *IEEE Transactions on Automatic Control*, 58(7): 1644–1658.

Sanjari, S.; Saldi, N.; and Yüksel, S. 2023. Nash equilibria for exchangeable team against team games and their mean field limit. In *American Control Conference*, 1104–1109. San Diego, CA.

Shishika, D.; Guan, Y.; Dorothy, M.; and Kumar, V. 2022. Dynamic defender-attacker Blotto game. In *American Control Conference*, 4422–4428. Atlanta, GA.

Tyler, J.; Arnold, R.; Abruzzo, B.; and Korpela, C. 2020. Autonomous Teammates for Squad Tactics. In *International Conference on Unmanned Aircraft Systems*, 1667–1672. Athens, Greece, Sept. 1–4, 2020.