Zero-Sum Games Between Large-Population Heterogeneous Teams: A Reachability-based Analysis under Mean-Field Sharing

Yue Guan[1] Mohammad Afshari[2] Panagiotis Tsiotras[3]

***Abstract*— This work studies the behaviors of two large-population teams competing in a discrete environment, where agents within each team are of different types. The team-level interactions are modeled as a zero-sum game, while the dynamics within each team is formulated as a collaborative mean-field team problem. Following the mean-field literature, we first approximate the large-population team game with its infinite-population limit. We then introduce two fictitious coordinators and transform the infinite-population game to an equivalent zero-sum coordinator game. We study the optimal strategies for each team via a novel reachability analysis. We show that the obtained team strategies are decentralized and are ϵ-optimal for the original finite-population game. The theoretical guarantees are verified by numerical examples.**

INTRODUCTION

Multi-agent decision-making arises in many applications, ranging from warehouse robots [1] to organizational economics [2]. 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 comprised of a large number of intelligent agents, compete at the team level, while the agents within the same team collaborate. Such hierarchical interactions are of particular interest to military operations and other multi-agent systems that operate in adversarial environments.

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.
2) Competitive team problems are *conceptually* challenging due to the unknown nature of the opponent strategies. In particular, one may want to impose additional assumptions on the opponent team (e.g., all adversarial agents apply the same strategy) to obtain tractable solutions, but these assumptions may not hold in practice.

The scalability challenge has been addressed in a specific class of *single*-team problems known as the mean-field team game [3]. The salient feature of a mean-field team is that a group of *homogeneous* agents are weakly-coupled in their dynamics and rewards through their state distribution (the so-called mean-field). Under the assumption that all agents apply the same strategy, the intractable interactions among agents can be reduced to the interaction between a typical agent and the 'mass' of infinitely many other agents. A dynamic programming decomposition is then developed leveraging the common-information approach [4] so that all agents within the team deploy the same strategy prescribed by a fictitious coordinator. In the competitive team setting, although one may restrict the strategies used by the agents within his/her team to be identical, making the same assumption on the opponent team may result in a significant underestimation of the opponent's capabilities, and thus such assumption needs further justification.

Our contribution: We begin by formulating a zero-sum mean-field team game (ZS-MFTG) under the discrete-time, finite state and action spaces setting, where the agents' dynamics and the team rewards are coupled through the state distributions of both teams. Different from the single-team setting in [3], agents in our formulation collaborate within each team while compete at the team level.

Leveraging the common-information approach [5], we show that the ZS-MFTG at its infinite-population limit is equivalent to a zero-sum game between two *fictitious* coordinators, and the optimal team strategy can be easily obtained via dynamic programming. Through a reachability-based analysis, we prove that the solution obtained under the identical team strategy assumption is still ϵ-optimal in the original finite-population game, even if the opponent team is allowed to deploy a non-identical team strategy.

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 probability measures over E as $\mathcal{P}(E)$.

PROBLEM FORMULATION

Consider a discrete-time system with two large-population teams that operate over a finite horizon T. The Blue team has L sub-populations, with the ℓ-th sub-population consisting of N_1^ℓ homogeneous agents. On the other hand, the Red team has M sub-populations, and there are N_2^m homogeneous agents in the m-th sub-population. The Blue and Red team sizes are given by $N_1 = \sum_{\ell=1}^{L} N_1^\ell$ and $N_2 = \sum_{m=1}^{M} N_m^2$, and the total system size is denoted as $N = N_1 + N_2$. We use the vector $\rho = (\frac{N_1^1}{N}, \dots, \frac{N_1^L}{N}, \frac{N_2^1}{N}, \dots, \frac{N_2^M}{N})$ to characterize

[1]Yue Guan is a PhD student with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. yguan44@gatech.edu

[2]Mohammad Afshari is a Postdoctoral Fellow with the Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA, USA. mafshari@gatech.edu

[3]Panagiotis Tsiotras is the David & Andrew Lewis Chair Professor with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. tsiotras@gatech.edu

the size ratio among the sub-populations. Let $X_{i,t}^{\ell,N_1} \in \mathcal{X}^\ell$ and $U_{i,t}^{\ell,N_1} \in \mathcal{U}^\ell$ denote the random variables representing the state and action taken by a type-ℓ Blue agent $i \in [N_1^\ell]$ at time t. Here, $\mathcal{X}^\ell$ and $\mathcal{U}^\ell$ represent the *finite* individual state and action spaces for each type-ℓ Blue agent, independent of i and t. Similarly, we use $Y_{j,t}^{m,N_2} \in \mathcal{Y}^m$ and $V_{j,t}^{m,N_2} \in \mathcal{V}^m$ to denote the individual state and action of a type-m Red agent j. The joint state of Blue sub-population ℓ is denoted as $\mathbf{X}_t^{\ell,N_1}$ and the joint state of the Blue team is denoted as $\mathbf{X}_t^{N_1} = (\mathbf{X}_t^{\ell,N_1})_{\ell\in[L]}$. The joint state $\mathbf{Y}_t^{N_2}$ for the Red team is constructed similarly.

Definition 1: The *empirical distribution* (ED) for the Blue and Red sub-populations are defined as

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

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

Note that $\mathcal{M}_t^{\ell,N_1} \in \mathcal{P}(\mathcal{X}^\ell)$ and $\mathcal{N}_t^{m,N_2} \in \mathcal{P}(\mathcal{Y}^m)$. The team EDs are denoted as $\mathcal{M}_t^{N_1} = (\mathcal{M}_t^{\ell,N_1})_{\ell\in[L]}$ and $\mathcal{N}_t^{N_2} = (\mathcal{N}_t^{m,N_2})_{m\in[M]}$. With a slight abuse of notations, we define the spaces for team distributions as $\mathcal{P}(\mathcal{X}) = \times_\ell \mathcal{P}(\mathcal{X}^\ell)$ and $\mathcal{P}(\mathcal{Y}) = \times_m \mathcal{P}(\mathcal{X}^m)$, and we use the following operator to relate the team joint states and the corresponding team EDs:

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

Definition 2: We define the distance between two Blue team distributions $\mu, \mu' \in \mathcal{P}(\mathcal{X})$ as

$$\mathrm{d_{TV}}(\mu,\mu') = \max_{\ell\in[L]} \frac{1}{2}\sum_{x\in\mathcal{X}^\ell} \left|\mu^\ell(x) - \mu^{\ell'}(x)\right|.$$

The distance between Red team distributions are similarly defined.

Dynamics: We consider weakly-coupled dynamics, where the dynamics of each individual agent is coupled with other agents through the EDs. The stochastic transition of type-ℓ Blue agent i is governed by the kernel $f_t^{\ell,\rho}$ such that

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

where $\mu_t^{N_1} = \mathrm{Emp}(\mathbf{x}_t^{N_1})$ and $\nu_t^{N_2} = \mathrm{Emp}(\mathbf{y}_t^{N_2})$ are the *team* EDs. Similarly, the dynamics of type-m Red agent j of is governed by $g_t^{m,\rho}(y_{j,t+1}^{m,N_2} | y_{j,t}^{m,N_2}, v_{j,t}^{m,N_2}, \mu_t^{N_1}, \nu_t^{N_2})$.

Assumption 1: For all $x \in \mathcal{X}$, $u \in \mathcal{U}$, and $\ell \in [L]$, there exist a positive constant L_{f_t} such that, for all $\mu, \mu' \in \mathcal{P}(\mathcal{X})$ and $\nu, \nu' \in \mathcal{P}(\mathcal{Y})$,

$$\begin{aligned}\sum_{x'\in\mathcal{X}} \left|f_t^{\ell,\rho}(x'|x,u,\mu,\nu) - f_t^{\ell,\rho}(x'|x,u,\mu',\nu')\right| \\ \leqslant L_{f_t}\Big(\mathrm{d_{TV}}(\mu,\mu') + \mathrm{d_{TV}}(\nu,\nu')\Big).\end{aligned}$$

We also assume that g_t^ρ is L_{g_t}-Lipschitz.

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

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

Assumption 2: For all $\mu, \mu' \in \mathcal{P}(\mathcal{X})$, $\nu, \nu' \in \mathcal{P}(\mathcal{Y})$ and $t \in \{0, \ldots, T\}$, there exist $L_r \geqslant 0$ such that

$$\left|r_t^\rho(\mu,\nu) - r_t^\rho(\mu',\nu')\right| \leqslant L_r\big(\mathrm{d_{TV}}(\mu,\mu') + \mathrm{d_{TV}}(\nu,\nu')\big).$$

Under the zero-sum reward 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 [3]. Specifically, at each time step t, Blue agent i observes its state $X_{i,t}^{\ell,N_1}$ and the team EDs $\mathcal{M}_t^{N_1}$ and $\mathcal{N}_t^{N_2}$. Similarly, Red agent j observes $Y_{j,t}^{m,N_2}$ and the team EDs. We consider the following mixed Markov policies:

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

where $\phi_{i,t}^\ell(u | X_{i,t}^{\ell,N_1}, \mathcal{M}_t^{N_1}, \mathcal{N}_t^{N_2})$ is the probability that the type-ℓ Blue agent i selects action u given its state $X_{i,t}^{\ell,N_1}$ and the team EDs $\mathcal{M}_t^{N_1}$ and $\mathcal{N}_t^{N_2}$. An individual strategy for the Blue agent i is defined as a time sequence $\phi_i^\ell = \{\phi_{i,t}^\ell\}_{t=0}^T$. Since Blue agents of the same type have the same state and action spaces, they share the same policy space. Hence, we denote Φ_t^ℓ and Φ^ℓ to be the set of individual policies and strategies available to each type-ℓ Blue agent. The Blue team strategy $\phi^{N_1} = \{\phi_i^\ell\}_{\ell\in[L], i\in[N_1^\ell]}$ is the collection of individual strategies used by each Blue agent. We use $\Phi^{N_1} = \times_{\ell\in[L], i\in[N^\ell]} \Phi^\ell$ to denote the set of Blue team strategies. Note that Φ^{N_1} contains team strategies where agents of the same type apply different strategies. The notation extends naturally to the Red team.

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

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

where $\mathcal{M}_t^{N_1} = \mathrm{Emp}(\mathbf{X}_t^{N_1})$ and $\mathcal{N}_t^{N_2} = \mathrm{Emp}(\mathbf{Y}_t^{N_2})$, and the expectation is with respect to the distribution of all system variables induced by $(\phi^{N_1}, \psi^{N_2}) \in \Phi^{N_1} \times \Psi^{N_2}$.

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

$$\underline{J}^{N*} = \max_{\phi^{N_1}\in\Phi^{N_1}} \min_{\psi^{N_2}\in\Psi^{N_2}} J^{N,\phi^{N_1},\psi^{N_2}}, \tag{2}$$

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 [6]. Consequently, we consider the following optimality condition for the Blue team strategy.

Definition 3: A Blue team strategy ϕ^{N_1*} is ϵ-optimal if

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

Note that ϵ measures the exploitability of a Blue team strategy, and 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.

The rest of the paper focuses on the performance analysis from the Blue team's perspective (max-min optimization), but the techniques developed are applicable to the Red team's side due to the symmetry of the problem formulation.

Mean-Field Approximation

The preceding max-min optimization 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 first consider the infinite-population limit of the large-population team games. We further assume that agents of the same type employ the same strategy. As a result, the behavioral of the entire sub-population can be represented by a *typical agent* [7]. As we will show later, due to the law of large numbers, the ED of an infinite-size sub-population converges to the state distribution of its typical agent. This limiting distribution is known as the *mean-field* (MF). In the sequel, we formulate the mean-field team game at its infinite-population limit and introduce additional concepts essential to the performance analysis in later sections.

Mean-field dynamics

We first introduce the class of identical team strategies.

Definition 4: The Blue team strategy $\phi^{N_1} = \{\phi_i^\ell\}_{\ell\in[L],i\in[N_1^\ell]}$ is an identical team strategy, if for all sub-population $\ell \in [L]$, $\phi_{i_1}^\ell = \phi_{i_2}^\ell$ for all $i_1, i_2 \in [N_1^\ell]$.

When all type-ℓ Blue agents apply the same individual strategy ϕ^ℓ, we slightly abuse the notation and use ϕ^ℓ to denote the identical sub-population strategy. Furthermore, we use $\Phi = \times_{\ell=1}^{L} \Phi^\ell$ to denote the set of Blue team strategies, where each Blue sub-population applies an identical strategy. Note that, under an identical team strategy, agents of different types may still apply different strategies. The definitions and notations extend to the identical Red team strategies.

Under identical team strategies, we define the mean-field (MF) as the state distribution of a typical agent.

Definition 5: Under identical team strategies $\phi = \{\phi^\ell\}_\ell \in \Phi$ and $\psi = \{\psi^m\}_m \in \Psi$, the sub-population MFs propagate according to the following *deterministic* dynamics:

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

The above deterministic mean-field dynamics can be expressed in a compact matrix form as

$$\begin{aligned}\mu_{t+1}^{\ell,\rho} &= \mu_t^{\ell,\rho} F_t^{\ell,\rho}(\mu_t^\rho,\nu_t^\rho,\phi_t^\ell),\\ \nu_{t+1}^{m,\rho} &= \nu_t^{m,\rho} G_t^{m,\rho}(\mu_t^\rho,\nu_t^\rho,\psi_t^m),\end{aligned} \tag{3}$$

where $F_t^{\ell,\rho} \in \mathbb{R}^{|\mathcal{X}^\ell|\times|\mathcal{X}^\ell|}$ is the transition matrix for a typical type-ℓ Blue agent under ϕ_t^ℓ, and $G_t^{m,\rho}$ is defined similarly.

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

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

where the propagation of the team mean-fields $\mu_t^\rho = \{\mu_t^{\ell,\rho}\}_\ell$ and $\nu_t^\rho = \{\nu_t^{m,\rho}\}_m$ is subject to the dynamics in (3).

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

$$\underline{J}^{\rho}*(\mu_0^\rho,\nu_0^\rho) = \max_{\phi\in\Phi}\min_{\psi\in\Psi} J^{\rho,\phi,\psi}(\mu_0^\rho,\nu_0^\rho). \tag{4}$$

Reachable Sets

Due to the deterministic dynamics in (3), designing the identical team policies ϕ_t and ψ_t at time t is equivalent to selecting the desirable next MFs for each sub-population. Consequently, we examine the set of MFs that can be achieved at the next time step. We use $\pi_t^\ell : \mathcal{U}^\ell\times\mathcal{X}^\ell \to [0,1]$ to denote an identical local policies of the type-ℓ Blue agents, which is *open-loop* with respect to the team MFs. Specifically, $\pi_t^\ell(u^\ell|x^\ell)$ is the probability that a Blue agent selects action u^ℓ at state x^ℓ regardless of the current MFs. The set of Blue local policies is denoted as Π_t^ℓ. Similarly, $\sigma_t^m : \mathcal{V}^m\times\mathcal{Y}^m \to [0,1]$ and Σ_t^m denote the Red local policy and its admissible set. Under local policy π_t^ℓ, the type-ℓ sub-population MF propagates according to

$$\mu_{t+1}^{\ell,\rho}(x') = \sum_{x\in\mathcal{X}^\ell}\Big[\sum_{u\in\mathcal{U}^\ell} f_t^{\ell,\rho}(x'|x,u,\mu_t^\rho,\nu_t^\rho)\pi_t^\ell(u|x)\Big]\mu_t^{\ell,\rho}(x), \tag{5}$$

and the dynamics of Red sub-population MF under local policies are defined similarly.

To facilitate later analysis, we provide the following definition of the reachable sets for the Blue and Red team MFs.

Definition 6: The Blue reachable set, starting from the team MFs $\mu_t^\rho = \{\mu_t^{\ell,\rho}\}_\ell$ and $\nu_t^\rho = \{\nu_t^{m,\rho}\}_m$ is defined as

$$\begin{aligned}\mathcal{R}_{\mu,t}^\rho(\mu_t^\rho,\nu_t^\rho) \triangleq \{\mu_{t+1}^\rho = \{\mu_{t+1}^{\ell,\rho}\}_\ell | \forall \ell\in[L],\ \exists \pi_t^\ell\in\Pi_t^\ell \text{ s.t.}\\ \mu_{t+1}^{\ell,\rho} = \mu_t^{\ell,\rho} F_t^{\ell,\rho}(\mu_t^\rho,\nu_t^\rho,\pi_t^\ell)\}.\end{aligned}$$

In the sequel, we regard the reachable sets as correspondences (set-valued functions) [8], i.e., $\mathcal{R}_{\mu,t}^\rho(\mu_t^\rho,\nu_t^\rho) : \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y}) \rightsquigarrow \mathcal{P}(\mathcal{X})$. The following lemma justifies the use of the reachable sets constructed based on the local policies to analyze the reachability of identical team policies.

Lemma 1: For all $\mu_t^\rho \in \mathcal{P}(\mathcal{X})$, $\nu_t^\rho \in \mathcal{P}(\mathcal{Y})$, we have that

$$\begin{aligned}\mathcal{R}_{\mu,t}^\rho(\mu_t^\rho,\nu_t^\rho) = \big\{\mu_{t+1}^\rho = \{\mu_{t+1}^{\ell,\rho}\}_\ell | \forall\ell\in[L],\ \exists \phi_t^\ell\in\Phi_t^\ell\\ \text{s.t. } \mu_{t+1}^{\ell,\rho} = \mu_t^{\ell,\rho} F_t^{\ell,\rho}(\mu_t^\rho,\nu_t^\rho,\phi_t^\ell)\big\}.\end{aligned}$$

Approximation Error

The following theorem states that the reachable set constructed under the identical policy assumption at the infinite-population limit is actually rich enough to approximate the empirical distributions induced by any non-identical team policy in the finite-population games.

Theorem 1: Consider a finite-population game and 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 $\mathcal{M}_{t+1}^{N_1}$. There exists $\mu_{t+1} \in \mathcal{R}_{\mu,t}^{\rho}(\mathcal{M}_t^{N_1}, \mathcal{N}_t^{N_2})$ such that

$$\mathbb{E}_{\phi_t^{N_1}}\Big[\mathrm{d_{TV}}\big(\mathcal{M}_{t+1}^{N_1}, \mu_{t+1}\big)\big|\mathbf{X}_t^{N_1}, \mathbf{Y}_t^{N_2}\Big] \leqslant \frac{\overline{|\mathcal{X}|}}{2}\sqrt{\frac{1}{\underline{N}_1}}, \quad (6)$$

where $\overline{|\mathcal{X}|} = \max_{\ell\in[L]} |\mathcal{X}^\ell|$ and $\underline{N}_1 = \min_{\ell\in[L]} N_1^\ell$.

Proof: The key step in the proof is to construct an identical local policy $\pi_{\mathrm{apprx},t}^\ell$ for each sub-population that has its action distribution matching the average of the policies used by the type-ℓ agents at each state. One can then leverage $\pi_{\mathrm{apprx},t}^\ell$ to mimic the population behavior and use a modified l_2 weak law of large numbers to show that the sub-population MF induced by $\pi_{\mathrm{apprx},t}^\ell$ satisfies the error bound in (6) with error $|\mathcal{X}^\ell|/2\sqrt{N_1^\ell}$ and is within the reachable set. See [9] for a detailed proof. ∎

Zero-Sum Game Between Coordinators

To efficiently solve the infinite-population game, we construct a fictitious centralized coordinated system by introducing the Blue and Red coordinators for the two teams respectively. At time t, the Blue coordinator observes the MFs of both teams and chooses a local policy $\pi_t^\ell \in \Pi_t^\ell$ for each of its sub-population according to:

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

where $\alpha_t^\ell : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \Pi_t^\ell$ is a *deterministic* Blue coordination policy for the type-ℓ population, and $\pi_t^\ell(u_t^\ell|x_t^\ell) \triangleq \alpha_t^\ell(\mu_t^\rho, \nu_t^\rho)(u_t^\ell|x_t^\ell)$ gives the probability that a type-ℓ Blue agent selects action u_t^ℓ at state x_t^ℓ. Similarly, the Red coordinator observes the team MFs and selects a local policy for its type-m sub-population according to $\sigma_t^m = \beta_t^m\big(\mu_t^\rho, \nu_t^\rho\big)$.

We refer to the collection $\pi_t = (\pi_t^\ell)_\ell$ as the Blue team local policy and denote its admissible set as Π_t. Similarly, the Red team local policy is denoted as $\sigma_t = (\sigma_t^m)_m \in \Sigma_t$. The coordination policy for the whole Blue team can then be defined as $\alpha_t : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \Pi_t$, which selects local policies for each sub-population. Similarly, the Red coordination strategy is defined as $\beta_t : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \Sigma_t$. We refer to $\alpha = \{\alpha_t\}_t$ as the *coordination strategy* for the Blue team and $\beta = \{\beta_t\}_t$ as the Red coordination strategy, and the admissible sets are denoted as $\mathcal{A}$ and $\mathcal{B}$.

Remark 1: There is a one-to-one correspondence between the Blue (Red) coordination strategies and the identical Blue (Red) team strategies.

The equivalent centralized system can be viewed as a zero-sum game played between the two coordinators, where the game state is the joint team MF (μ_t^ρ, ν_t^ρ), and the actions are the team local policies $\pi_t = (\pi_t^\ell)_\ell$ and $\sigma_t = (\sigma_t^m)_m$. Note that both the state and action spaces of the coordinator game are continuous.

Similar to the standard two-player zero-sum games, we use a backward induction scheme to find the lower value of the coordinator game. The lower value at the terminal time T is given by $\underline{J}_{\mathrm{cor},T}^{\rho*}(\mu_T^\rho, \nu_T^\rho) = r_T^\rho(\mu_T^\rho, \nu_T^\rho)$. For all previous time steps, the two coordinators optimize their cumulative reward by choosing their actions (i.e., local team policies) $\pi_t = (\pi_t^\ell)_\ell$ and $\sigma_t = (\sigma_t^m)_m$. Consequently, we have

$$\underline{J}_{\mathrm{cor},t}^{\rho*}(\mu_t^\rho, \nu_t^\rho) = r_t^\rho(\mu_t^\rho, \nu_t^\rho) + \max_{\pi_t\in\Pi_t}\min_{\sigma_t\in\Sigma_t} \underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mu_{t+1}^\rho, \nu_{t+1}^\rho), \quad (7)$$

where the Blue team MF $\mu_{t+1}^\rho = (\mu_{t+1}^{\ell,\rho})_\ell$ is propagated for each sub-population using (5) under the local policy $(\pi_t^\ell)_\ell$, and similarly for ν_{t+1}^ρ.

With the optimal value function, the optimal Blue team coordination policy can then be easily constructed via

$$\alpha_t^*(\mu_t^\rho, \nu_t^\rho) \in \quad (8)$$
$$\operatorname*{argmax}_{\pi_t\in\Pi_t}\min_{\sigma_t\in\Sigma_t} \underline{J}_{\mathrm{cor},t+1}^{\rho*}\big(\mu_t^\rho F_t^\rho(\mu_t^\rho, \nu_t^\rho, \pi_t), \nu_t^\rho G_t^\rho(\mu_t^\rho, \nu_t^\rho, \sigma_t)\big).$$

Exploiting the deterministic mean-field dynamics, we can change the optimization domains in (7) from the policy spaces to the corresponding reachable sets.

$$\underline{J}_{\mathrm{cor},t}^{\rho*}(\mu_t^\rho, \nu_t^\rho) = r_t^\rho(\mu_t^\rho, \nu_t^\rho) \quad (9)$$
$$+ \max_{\mu_{t+1}^\rho\in\mathcal{R}_{\mu,t}^\rho(\mu_t^\rho,\nu_t^\rho)}\ \min_{\nu_{t+1}^\rho\in\mathcal{R}_{\nu,t}^\rho(\mu_t^\rho,\nu_t^\rho)} \underline{J}_{\mathrm{cor},t+1}^{\rho*}(\mu_{t+1}^\rho, \nu_{t+1}^\rho).$$

For the rest of the paper, we will work with the above reachability-based optimization problem. There are two advantages for such an approach: (i) the reachable sets generally have a lower dimension than the coordinator action spaces[1], which is desirable for numerical algorithms, and (ii) the reachability-based optimization allows us to apply Theorem 1 and study the performance loss due to the identical-strategy assumption introduced by the mean-field approximation.

Lipschitz Continuity of the Value Functions

To obtain performance guarantees for a finite-population system, we need to first examine the continuity of the coordinator game value. If the value function is not continuous, a small disturbance in the EDs due to stochasticity may lead to a performance that is drastically different from the mean-field prediction.

Before analyzing the value function, we first study the continuity of the two reachability correspondences under the Hausdorff distance $\mathrm{dist_H}$.[2]

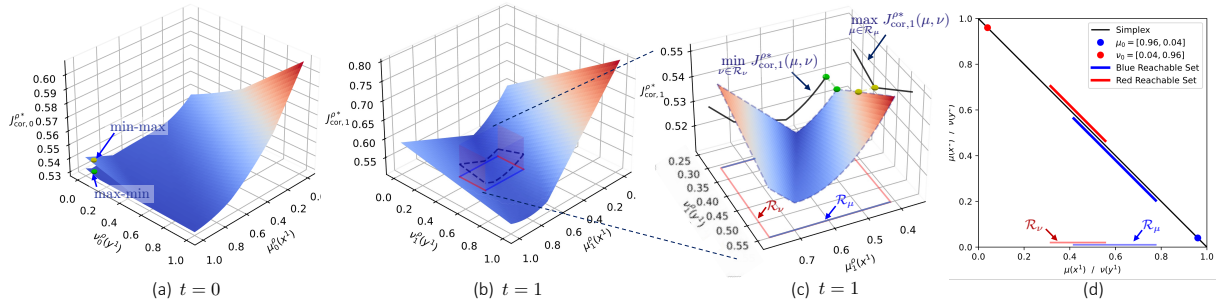
Lemma 2: For all $\mu_t, \mu_t' \in \mathcal{P}(\mathcal{X})$ and $\nu_t, \nu_t' \in \mathcal{P}(\mathcal{Y})$, the reachability correspondence $\mathcal{R}_{\mu,t}$ satisfies

$$\mathrm{dist_H}(\mathcal{R}_{\mu,t}^\rho(\mu_t, \nu_t), \mathcal{R}_{\mu,t}^\rho(\mu_t', \nu_t')) \quad (10)$$
$$\leqslant L_{R_{\mu,t}}\big(\mathrm{d_{TV}}\big(\mu_t, \mu_t'\big) + \mathrm{d_{TV}}\big(\nu_t, \nu_t'\big)\big).$$

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

[1]The Blue reachable set is a subset of $\times_\ell \mathcal{P}(\mathcal{X}^\ell)$, while the Blue coordinator action space is given by $\times_\ell(\mathcal{P}(\mathcal{U}^\ell))^{|\mathcal{X}^\ell|}$.

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



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

Fig. 1. 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 $\mu_0^\rho = [0.96, 0.04]$ and $\nu_0^\rho = [0.04, 0.96]$.

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

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

$$\begin{aligned}\left|\underline{J}^{\rho*}_{\mathrm{cor},t}(\mu_t^\rho, \nu_t^\rho) - \underline{J}^{\rho*}_{\mathrm{cor},t}(\mu_t^{\rho\prime}, \nu_t^{\rho\prime})\right| \\ \leqslant L_{J,t}\big(\mathrm{d_{TV}}(\mu_t^\rho, \mu_t^{\rho\prime}) + \mathrm{d_{TV}}(\mu_t^\rho, \nu_t^{\rho\prime})\big),\end{aligned} \tag{11}$$

where the Lipschitz constant is given by $L_{J,t} = L_r\big(1 + \sum_{k=t}^{T-1}\prod_{\tau=t}^{k}(L_{\mathcal{R}^\rho_{\mu,\tau}} + L_{\mathcal{R}^\rho_{\nu,\tau}})\big)$.

Proof: Observe that the lower value in (9) takes the form: $f(x,y) = \max_{p\in\Gamma(x,y)}\min_{q\in\Theta(x,y)} g(p,q)$, which is an extension of the maximization marginal function [8] to the max-min case. Based on the continuity result for the max-min marginal function, we can prove the theorem through an inductive argument. See [9] for a detailed proof. ∎

Performance Guarantees

Recall that the optimal Blue coordination strategy α^* is constructed for the infinite-population game where both teams apply identical team strategies. The following main theorem compares the worst-case performance of the identical Blue team strategy induced by α^* (Remark 1) to the original max-min optimization in (2), where non-identical strategies are allowed.

Theorem 3: The optimal Blue coordination strategy α^* in (8) induces an ϵ-optimal Blue team strategy. Formally, for all joint states $\mathbf{x}^{N_1} \in \mathcal{X}^{N_1}$, $\mathbf{y}^{N_2} \in \mathcal{Y}^{N_2}$,

$$\begin{aligned}\underline{J}^{N*}(\mathbf{x}^{N_1}, \mathbf{y}^{N_2}) &\geqslant \min_{\psi^{N_2}\in\Psi^{N_2}} J^{N,\alpha^*,\psi^{N_2}}(\mathbf{x}^{N_1}, \mathbf{y}^{N_2}) \\ &\geqslant \underline{J}^{N*}(\mathbf{x}^{N_1}, \mathbf{y}^{N_2}) - \mathcal{O}\Big(\frac{1}{\sqrt{\underline{N}}}\Big),\end{aligned} \tag{12}$$

where $\underline{N} = \min\{\underline{N}_1, \underline{N}_2\}$.

Proof: The first inequality in (12) is straightforward, since α^* is restricted to identical team strategy space. The second inequality can be split into two lemmas: (i) $\min_{\psi^{N_2}\in\Psi^{N_2}} \underline{J}^{N,\alpha,\psi^{N_2}} \geqslant \underline{J}^{\rho*}_{\mathrm{cor}} - \epsilon_1$, and (ii) $\underline{J}^{\rho*}_{\mathrm{cor}} \geqslant \underline{J}^{N*} - \epsilon_2$. Finally, one can show that both error terms are of order $\mathcal{O}(1/\sqrt{\underline{N}})$. The proofs of both lemmas make uses of the Lipschitz results of $\underline{J}^{\rho*}_{\mathrm{cor}}$ (Theorem 2) and the approximation result (Theorem 1). See the full version for more details [9]. ∎

In other words, the above theorem states that the Red team can at most gain an ϵ performance increase through using a non-identical team strategy, even if the Blue team assumes that both teams are restricted to identical strategies. As a result, the above theorem significantly reduces the search space for ϵ-optimal strategies under a ZS-MFTG formulation.

Numerical Examples

Numerical Example 1

This example is used to demonstrate the reachability-based optimization scheme and show that the coordinator game value may not always exist, contrary to the continuous setting [10]. For simplicity and visualization purpose, we consider two homogeneous teams, each only has a single sub-population, and the individual state spaces only consist two states (i.e., $\mathcal{X}^1 = \{x^1, x^2\}$ and $\mathcal{Y}^1 = \{y^1, y^2\}$). We drop the superscript ℓ and m for the rest of this example. See the full version in [9] for the detailed setups.

The coordinator game values in Fig. 1 are computed through discretization, where we uniformly mesh the two-dimensional simplexes $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ into 1000 bins.

Since the value function $J^{\rho*}_{\mathrm{cor},1}$ is not convex-concave, the Nash equilibrium does not exist and the upper and lower game values differ at $t=0$ as shown in Fig. 1(a). Specifically, at $\mu_0^\rho = [0.96, 0.04]$ and $\nu_0^\rho = [0.04, 0.96]$, we have the lower value $\underline{J}^{\rho*}_{\mathrm{cor},0} = 0.5298$ and the upper value $\bar{J}^{\rho*}_{\mathrm{cor},0} = 0.5384$, which are visualized as the green and yellow points. The reachable sets from μ_0^ρ and ν_0^ρ are plotted in Fig. 1(d) and also visualized as the box in Fig. 1(b), which serve as the optimization domain in (9) at $t=0$. Fig. 1(c) presents a zoom-in for the optimization and the marginalized functions.

The red line in Fig. 2 shows the performance loss of the proposed identical Blue team strategy in a finite-population game, when the opponent team uses a non-identical team strategy to exploit. It verifies the claim of Theorem 3.

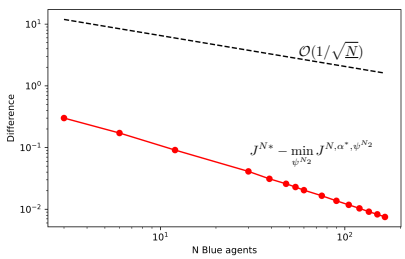


Fig. 2. Performance loss of the optimal Blue coordination strategy.

Numerical Example 2

We consider a simple defense scenario on a graph with two nodes (N1 and N2) as in Fig. 3, where Node 2 (N2) is a high-value target to be guarded. The Blue team is the defending team and consists of two sub-populations: type-1 (blue) and type-2 (cyan). The Red team is the attacking team and is homogeneous. The running reward is the difference between the numbers of Blue agents (both types) and Red agents on N2, so that the Blue team tries to maintain a numerical advantage over the Red team at N2. While the type-2 Blue agents can freely move on the graph, the type-1 agents can only move from N2 to N1, meaning that once a type-1 agent is "committed" to N1, it cannot come back to N2. The Red agents' transition probability is influenced by the distributions of the Blue agents: the more Blue agents are presented on N1, the less likely a Red agent can transit to N2 (the blocking effect); similarly, the more Blue agents on N2, the less likely a Red agent can stay on N2 (the fending effect). The blocking and fending effects are saturated when the Blue team's numerical advantage is over a certain threshold. We assume that the type-1 Blue agents are more effective in both blocking and fending off the Red agents than their type-2 peers. For example, the transition probability of a Red agent from N1 to N2 under the "move-to-node2" action v is given by

$$g_t^{1,\rho}(2|1,v,\mu_t^{1,N_1},\mu_t^{2,N_1},\nu_t^{1,N_2}) = \max\big\{0.1,$$
$$\rho^{1,N_2}\nu_t^{1,N_2}(1) - 0.6\rho^{1,N_1}\mu_t^{1,N_1}(1) - 0.4\rho^{2,N_1}\mu_t^{2,N_1}(1)\big\},$$

where 0.1 is the transition probability after the Blue team saturates the blocking effect.

As a result of the setup, a type-1 Blue agent needs to decide whether it should migrate to N1 to impede the Red team's transition, or to stay on N2 for running reward and fending off Red agents on N2.

In Fig. 3, we present the mean-field trajectories under two different team size ratios. In the first scenario, we have four times more type-2 Blue agents than type-1. Note that all cyan agents migrate to N1 to block the Red team's transition from N1 to N2 at $t=1$. The cyan agents then use their mobility and move back to N2 at $t=2$ to maximize the numerical advantage at the terminal time step. To collaborate with the type-2 agents and saturate the blocking effect on N1 at $t=1$, 20% of the type-1 Blue agents commit to N1, at the price of less type-1 agents on N2 at $t=2$.

The second scenario has more type-1 Blue agents than type-2. As the type-1 Blue agents are now the major force in deciding the reward, keeping them on N2 for the running reward out-benefits committing them to N1 for blocking, and hence all type-1 Blue agents stay on N2. With no type-1 agents going to N1, type-2 agents alone have limited influence on the Red team's transition. Consequently, the type-2 Blue agents prioritize N2 at $t=1$ and saturate the fending effect with 90% of the type-2 population. The rest of the type-2 agents (10%) moves to N1 at $t=1$ to impede the Red team's transition. At the terminal time step, that 10% of the type-2 agents return to N2 to maximize the numerical advantage over the Red team.

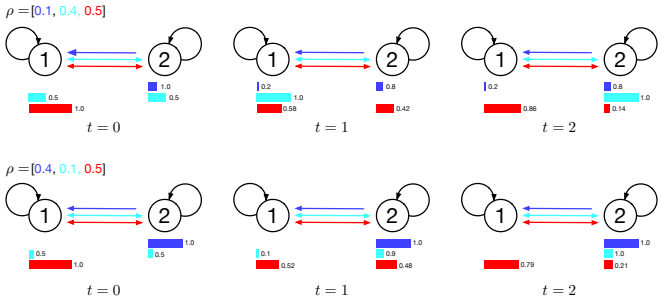


Fig. 3. A graph-based defense scenario. The sub-population mean-fields are visualized as bars with bar length normalized according to ρ.

Conclusion

In this work, we have formulated a zero-sum game between two heterogeneous teams under the mean-field sharing information structure. We approximated the team game with an infinite-population game and further transformed it to an equivalent zero-sum game between two coordinators. We showed that even though the optimal strategies are solved assuming that both teams employ identical team strategies, the strategies constructed are still ϵ-optimal for the original finite-population game and the general class of non-identical team strategies. The derived performance guarantees are verified through numerical examples. Future work will investigate the special case of the LQG setup of this problem and deploy machine learning techniques to solve zero-sum mean-field team problems in more complex environments.

References

[1] J. Li, A. Tinka, S. Kiesel, J. W. Durham, T. S. Kumar, and S. Koenig, "Lifelong multi-agent path finding in large-scale warehouses," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 11 272–11 281.

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

[3] J. Arabneydi and A. Mahajan, "Team optimal control of coupled subsystems with mean-field sharing," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 1669–1674.

[4] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1644–1658, 2013.

[5] D. Kartik, A. Nayyar, and U. Mitra, "Common information belief based dynamic programs for stochastic zero-sum games with competing teams," *arXiv preprint arXiv:2102.05838*, 2 2021. [Online]. Available: http://arxiv.org/abs/2102.05838

[6] R. J. Elliott and N. J. Kalton, *The Existence of Value in Differential Games*. American Mathematical Soc., 1972, vol. 126.

[7] M. Huang, R. P. Malhamé, and P. E. Caines, "Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the nash certainty equivalence principle," *Communications in Information & Systems*, vol. 6, no. 3, pp. 221–252, 2006.

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

[9] Y. Guan, M. Afshari, and P. Tsiotras, "Zero-sum games between mean-field teams: A common information and reachability based analysis," *arXiv preprint arXiv:2303.12243*, 2023.

[10] S. Sanjari, N. Saldi, and S. Yüksel, "Nash equilibria for exchangeable team against team games and their mean field limit," in *2023 American Control Conference*. IEEE, 2023, pp. 1104–1109.