Mean-field games among teams

Jayakumar Subramanian, Balaji Krishnamurthy, Akshat Kumar, and Aditya Mahajan

***Abstract*— In this paper, we consider a game among teams of agents. Agents within each team cooperate with each other while they compete with the agents in all other teams. All agents are coupled in the dynamics and the cost through the empirical distribution (or the mean-field) of the states of agents in each team. The system has a mean-field sharing information structure, i.e., each agent observes its local state and the mean-field. Thus, this system is a stochastic game with asymmetric information. For such games, a solution concept called common-information based Markov perfect equilibrium has been presented in literature. This refinement of Nash equilibrium has been shown to exist whenever a technical condition equivalent to the absence of signaling is satisfied. We show that for our model the current mean-field can be used as an information state and it satisfies the technical conditions for the absence of signaling. We identify coupled dynamic programs that can used to identify the common-information based Markov perfect equilibrium for our model. We then present an efficient method to sample from the mean-field dynamics to efficiently solve the coupled dynamic programs using sampling based methods. We then consider infinite population mean-field limit of the model and show that the Markov perfect equilibrium obtained using the mean-field limit is an ϵ-Markov perfect equilibrium.**

I. Introduction

Traditionally, agents in a multi-agent system are modeled either as cooperative agents who optimize a common system-wide objective or as strategic agents who optimize an individual objective. This difference in the behavior of agents leads to different conceptual difficulties and different solution concepts. As a result, the two settings are studied as separate sub-disciplines of decision theory: models with cooperative agents are studied under the heading of team theory [1] and models with strategic agents are studied under the heading of game theory [2]. For the most part, these two research sub-disciplines have evolved independently.

However, in recent years, many applications have emerged which may be considered as games among teams. Examples include:

- Multiple demand aggregators competing in the same electricity markets.
- Multiple ride-sharing companies competing in the same city.
- Multiple firms competing in the same industry with different levels of competitive advantages [3].
- The DARPA Spectrum Sharing Challenge, where teams of multiple radio units compete with other such teams in the same geographic area [4].

In such applications, teams of agents compete with other teams of agents. These models are different from traditional team theory models because agents belonging to different teams have separate objectives and are, therefore, not cooperative. These models are also different from traditional game theory models because agents belonging to the same team can enter into pre-game agreements; therefore the notion of equilibrium in games among teams must account for the possibility of simultaneous and coordinated deviations by all agents belonging to the same team. In some of the applications described above, agents have the option of switching teams, so certain ideas from coalition formation are relevant. However, we do not explore that direction in this paper. So, our discussion is distinct from the concerns in cooperative games [5].

There has been some recent interest in modeling and analyzing games among teams. A dynamic game among teams with delayed intra-agent information sharing is considered in [6], where common-information based refinements for Team-Nash equilibrium are presented. Mean-field games among teams have been considered in [7] in the context of transportation networks and in [8] in the context of large firms aiming at new product or technology development.

We also consider mean-field games among teams but our motivation, model and results are different from [7] and [8]. The focus in [7] is on designing incentive mechanisms to mitigate congestion, while [8] restricts attention to rank based rewards. In contrast, we consider a general cost for each team which depends on the state of the agents in the team and the mean-fields of all the teams. Moreover, [7] considers specific dynamics on traffic graphs and [8] considers a Poisson process with controlled intensities. In contrast, we consider finite-state controlled Markov dynamics. In addition, our assumptions on the behavior of the agents in a team and the solution concepts for the game are different from [7] and [8].

Our model may be viewed as a generalization of mean-field models of decentralized multi-agent systems. Broadly speaking, two classes of such models have been considered int he literature. (i) Mean-field games, which considers large-scale systems with strategic agents with mean-field coupling in the dynamics and the cost; see [3], [9], [10] and follow-up work. (ii) Mean-field teams, which considers cooperative control of teams with exchangeable agents; see [11], [12] and follow-up work. In this paper, we consider games among teams where the agents in a team are exchangeable and

J. Subramanian and B. Krishnamurthy are with the Media and Data Science Research Lab, Digital Experience Cloud, Adobe Inc., Noida, Uttar Pradesh, India. Emails: jasubram@adobe.com, kbalaji@adobe.com

A. Kumar is with the School of Computing and Information Systems at the Singapore Management University, Singapore. Email: akshatkumar@smu.edu.sg

A. Mahajan is with the Department of Electrical and Computer Engineering, McGill University, Montreal, Canada. Email: aditya.mahajan@mcgill.ca

the agents across teams have mean-field coupling in the dynamics and the cost.

The main contributions of this paper are as follows:

- We present a solution concept for mean-field games among teams and present an approach to determine the same.
- We present a mean-field limit model where the population of each team is assumed to be infinite. We show that the mean-field limit solution is ϵ optimal for the agents in any single team and derive the bound for ϵ.

II. Model and problem formulation

A. Model of mean-field games among teams (MFGT)

1) State and action spaces: Consider a multi-agent system with K teams of homogeneous agents. Team $k \in \mathcal{K} = \{1, \dots, K\}$ consists of $N^{(k)}$ agents denoted by the set $\mathcal{N}^{(k)}$, with state space $\mathcal{X}^{(k)}$ and action space $\mathcal{U}^{(k)}$. We assume that $\mathcal{X}^{(k)}$ and $\mathcal{U}^{(k)}$ are finite sets. At time t, the state and control action of a generic agent i in team k is denoted by $X_t^i \in \mathcal{X}^{(k)}$ and $U_t^i \in \mathcal{U}^{(k)}$, respectively. Moreover,

$$X_t^{(k)} = (X_t^i)_{i \in \mathcal{N}^{(k)}} \quad \text{and} \quad U_t^{(k)} = (U_t^i)_{i \in \mathcal{N}^{(k)}}$$

denote the states and control actions of all agents in team k and

$$X_t = (X_t^{(k)})_{k \in \mathcal{K}} \quad \text{and} \quad U_t = (U_t^{(k)})_{k \in \mathcal{K}}$$

denote the global state and control actions of the entire system.

Given a vector $x^{(k)} = (x^i)_{i \in \mathcal{N}^{(k)}}$, $x^i \in \mathcal{X}^{(k)}$, we use $\xi(x^{(k)})$ to denote the mean-field (or empirical distribution) of $x^{(k)}$, i.e.,

$$\xi(x^{(k)}) = \frac{1}{N^{(k)}} \sum_{i \in \mathcal{N}^{(k)}} \delta_{x^i},$$

where δ_x is a dirac delta measure centered at x. We use $Z_t^{(k)} = \xi(X_t^{(k)})$ to denote the mean-field of team k, $\mathcal{Z}^{(k)}$ to denote the space of realizations of $Z_t^{(k)}$, $Z_t = (Z_t^{(k)})_{k \in \mathcal{K}}$ to denote the mean-field of the entire system, and $\mathcal{Z}^*$ to denote the space of realizations of Z_t. Note that $\mathcal{Z}^{(k)}$ has at most $(N^{(k)} + 1)^{|\mathcal{X}^{(k)}|}$ elements.

2) System dynamics: Let $(x_{1:T}, u_{1:T})$ denote a realization of $(X_{1:T}, U_{1:T})$ and let $z_t^{(k)} = \xi(x_t^{(k)})$ denote the mean-field at time t. The initial states of all agents are independent, i.e.,

$$\begin{aligned}\mathbb{P}(X_1 = x_1) &= \prod_{k \in \mathcal{K}} \prod_{i \in \mathcal{N}^{(k)}} \mathbb{P}(X_1^i = x_1^i) \\ &=: \prod_{k \in \mathcal{K}} \prod_{i \in \mathcal{N}^{(k)}} P_0^{(k)}(x_1^i),\end{aligned}$$

where $P_0^{(k)}$ denotes the initial state distribution of agents in team k. The global state of the system evolves in a controlled Markovian manner and independently across agents, i.e.,

$$\begin{aligned}&\mathbb{P}(X_{t+1} = x_{t+1} \mid X_{1:t} = x_{1:t}, U_{1:t} = u_{1:t}) \\ &\quad = \prod_{k \in \mathcal{K}} \prod_{i \in \mathcal{N}^{(k)}} \mathbb{P}(X_{t+1}^i = x_{t+1}^i \mid X_t = x_t, U_t = u_t).\end{aligned}$$

And finally, all agents within a team are partially exchangeable. So the state evolution of a generic agent depends on the states and actions of other agents only through the mean-fields of the states of the teams, i.e., for agent i in team k,

$$\begin{aligned}&\mathbb{P}(X_{t+1}^i = x_{t+1}^i \mid X_t = x_t, U_t = u_t) \\ &\qquad = \mathbb{P}(X_t^i = x_t^i, U_t^i = u_t^i, Z_t = z_t) \\ &\qquad =: P^{(k)}(x_{t+1}^i \mid x_t^i, u_t^i, z_t),\end{aligned}$$

where $P^{(k)}$ denotes the controlled transition matrix for team k.

Combining all of the above, we have

$$\begin{aligned}&\mathbb{P}(X_{t+1} = x_{t+1} \mid X_{1:t} = x_{1:t}, U_{1:t} = u_{1:t}) \\ &\qquad = \prod_{k \in \mathcal{K}} \prod_{i \in \mathcal{N}^{(k)}} P^{(k)}(x_{t+1}^i \mid x_t^i, u_t^i, z_t). \quad (1)\end{aligned}$$

3) Cost function: There is a cost function: $c_t^{(k)} : \mathcal{X}^{(k)} \times \mathcal{U}^{(k)} \times \mathcal{Z}^* \to \mathbb{R}$ associated with each agent in team k. The per-step cost incurred by team k is the average of the cost associated with all agents in the team, i.e,

$$C_t^{(k)} = \frac{1}{N^{(k)}} \sum_{i \in \mathcal{N}^{(k)}} c_t^{(k)}(X_t^i, U_t^i, Z_t). \quad (2)$$

4) Information structure and control laws: We assume that the system has mean-field sharing information-structure [11], i.e., each agent has access to its local state X_t^i and the history of mean-field $Z_{1:t}$ of all teams. Thus, the information available to agent i is given by:

$$I_t^i = \{X_t^i, Z_{1:t}\}. \quad (3)$$

In addition, we assume that all agents in team k use identical (stochastic) control laws: $g_t^{(k)} \colon \mathcal{X}^{(k)} \times \mathcal{Z}^{*t} \to \Delta(\mathcal{U}^{(k)})$ to choose the control action at time t, i.e.,

$$U_t^i \sim g_t^{(k)}(X_t^i, Z_{1:t}), \quad (4)$$

where each agent randomizes independently.

In general, restricting attention to identical strategies for all agents in a team results in a loss of optimality for that team. However, as argued in [11], such a restriction is often justifiable due to other considerations such as simplicity, robustness and fairness.

Let $g^{(k)} := (g_1^{(k)}, \dots, g_T^{(k)})$ denote the strategy of team k. We use $g^{(-k)}$ to denote the strategy of all teams other than k. Given a strategy $g = (g^{(k)}, g^{(-k)})$ for all teams, the expected total cost incurred by team k is given by:

$$J^{(k)}(g^{(k)}, g^{(-k)}) = \mathbb{E}^{(g^{(k)}, g^{(-k)})}\left[\sum_{t=1}^{T} C_t^{(k)}\right]. \quad (5)$$

It is assumed that the system dynamics $(P_0^{(k)})_{k \in \mathcal{K}}$, $(P^{(k)})_{k \in \mathcal{K}}$, the cost functions $(c^{(k)})_{k \in \mathcal{K}}$ and the information-structure are common knowledge to all agents. Each team is interested in minimizing the total expected cost incurred over a finite horizon. Following [6], we say that a strategy profile $g^* = (g^{*(k)})_{k \in \mathcal{K}}$ is a **Team-Nash equilibrium** if for

all teams $k \in \mathcal{K}$ and all other strategy profiles $g^{(k)}$ for team k, we have:

$$J^{(k)}(g^{*(k)}, g^{*(-k)}) \leqslant J^{(k)}(g^{(k)}, g^{*(-k)}). \tag{6}$$

In the sequel, we refer to the model defined above as mean-field game among teams (MFGT). We are interested in the following:

Game 1: Identify a Team-Nash equilibrium of the game among teams (MFGT) model defined above.

B. Salient features of the model

We now discuss some salient features of the MFGT model described above:

1) All agents in a team are homogeneous: We have assumed that all agents in a team have homogeneous dynamics and cost functions. This assumption is made only for ease of exposition. It is conceptually straightforward to extend the discussion to models with heterogeneous agents where the agents have multiple types. In fact, such a model can be converted into a model with homogeneous agents by augmenting the state space and considering the type of each agent to be a (static) component of its state.

2) Independent randomization: We have assumed that all agents randomize independently. In principle at each time, agents in the same team could randomize in a correlated manner, but we do not consider that setup in this paper.

3) Simultaneous deviations by all agents in a team: In a general Team-Nash equilibrium [6], all agents in a team are allowed to deviate together and in a correlated manner. However, we have imposed an additional assumption that all agents in a team play identical strategies. Under this assumption, when agents in a team consider deviations, they only consider deviations in which all agents of that team are playing identical strategies.

4) Dynamic Game with asymmetric information: MFGT is a dynamic game (also called stochastic game), where there is a state space model which describes the evolution of the state of the environment. The agents have imperfect and asymmetric information about the state of the environment.

C. Refinements of Nash equilibrium

For dynamic games with perfect information, a refinement of Nash equilibrium known as Markov perfect equilibrium can be identified using dynamic programming [13], [14]. However, MFGT is a dynamic game with *asymmetric* information, where the notion of Markov perfect equilibrium is not applicable. Recently, Nayyar et al. [15] proposed a refinement of Nash equilibrium called *common information based Markov perfect equilibrium* (CIB-MPE) for dynamic games with asymmetric information. This refinement is applicable when the game satisfies a technical condition (strategy independence of common information based beliefs). It is argued in [16] that the technical condition of [15] is equivalent to absence of signaling [17], [18]. Loosely speaking, absence of signaling means that each player's private information at time t is payoff irrelevant to all players from time $(t+1)$ onward.

The characterization in [15] is based on the common-information-based belief on the local information. For MFGT this belief will correspond to $\mathbb{P}(X_t|Z_{1:t})$ which takes values in $\Delta(\prod_{k\in\mathcal{K}} \mathcal{X}^{(k)})$. Following an argument similar to [11], we show that it is possible to work with a simpler information state for mean-field models. In particular, we show that the current mean-field Z_t is an information state for the common information $Z_{1:t}$. We combine this result with the argument of [15] to present a CIB-MPE which uses Z_t as an information state.

III. Common information based Markov perfect equilibrium (CIB-MFE)

A. Some preliminary results

Given any strategy $g = (g^{(1)}, \dots, g^{(K)})$ for the system and any realization $z_{1:T}$ of the mean-field $Z_{1:T}$, we define the following partial functions, which we call *prescriptions*:

$$\gamma_t^{(k)} = g_t^{(k)}(\cdot, z_{1:t}), \quad \forall k \in \mathcal{K}. \tag{7}$$

When the realization $z_{1:t}$ is given, $\gamma_t^{(k)}$ is a function from $\mathcal{X}^{(k)}$ to $\mathcal{U}^{(k)}$. When $Z_{1:t}$ is a random variable, $g^{(k)}(\cdot, Z_{1:t})$ is a random function from $\mathcal{X}^{(k)}$ to $\mathcal{U}^{(k)}$ and we denote this random function by $\Gamma_t^{(k)}$. We use γ_t to denote $(\gamma_t^{(1)}, \dots, \gamma_t^{(K)})$ and Γ_t to denote $(\Gamma_t^{(1)}, \dots \Gamma_t^{(K)})$.

For any $z^{(k)} \in \mathcal{Z}^{(k)}$, let $\Xi^{(k)}(z^{(k)}) = \{x^{(k)} \in (\mathcal{X}^{(k)})^{N^{(k)}} : \xi(x^{(k)}) = z^{(k)}\}$ denote all possible states for agents in team k such that the mean-field of team k is $z^{(k)}$. Then we have the following results. The proofs are omitted due to space constraints. The main ideas for the proof are based on the arguments presented in [11] for establishing similar results for mean-field teams.

Lemma 1: For any strategy $g = (g^{(1)}, \dots, g^{(K)})$ and any realization $(z_{1:T}, \gamma_{1:T})$ of $(Z_{1:T}, \Gamma_{1:T})$, we have that for any $x = (x^{(1)}, \dots, x^{(K)})$, $x^{(k)} \in (\mathcal{X}^{(k)})^{N^{(k)}}$,

$$\begin{aligned}&\mathbb{P}(X_t = x \mid Z_{1:t} = z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}) \\ &= \prod_{k\in\mathcal{K}} \mathbb{P}(X_t^{(k)} = x_t^{(k)} \mid Z_{1:t} = z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}).\end{aligned} \tag{8}$$

Moreover for any $k \in \mathcal{K}$,

$$\begin{aligned}&\mathbb{P}(X_t^{(k)} = x^{(k)} \mid Z_{1:t} = z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}) \\ &\qquad = \mathbb{P}(X_t^{(k)} = x^{(k)} \mid Z_t^{(k)} = z_t^{(k)}) \\ &\qquad = \frac{1}{|\Xi^{(k)}(Z^{(k)})|} \mathbb{1}\{\xi(x^{(k)}) = z^{(k)}\}.\end{aligned} \tag{9}$$

Lemma 2: For any realization $(z_t, z_{t+1}^{(k)}, \gamma_t^{(k)})$ of $(Z_t, Z_{t+1}^{(k)}, \Gamma_t^{(k)})$, we have that for any $x_t^{(k)} \in \Xi^{(k)}(Z_t^{(k)})$,

$$\sum_{x_{t+1}^{(k)} \in \Xi^{(k)}(Z_{t+1}^{(k)})} \mathbb{P}(X_{t+1}^{(k)} = x_{t+1}^{(k)} \mid X_t^{(k)} = x_t^{(k)}, U_t^{(k)} = \gamma_t^{(k)}(x_t^{(k)}), Z_t = z_t)$$

is a constant that depends on $x_t^{(k)}$ only through $z_t^{(k)}$. We denote this constant by $Q_t^{(k)}(z_{t+1}^{(k)} \mid z_t, \gamma_t^{(k)})$.

Lemma 3: For any strategy $g = (g^{(1)}, \dots, g^{(K)})$ and any realization $(z_{1:T}, \gamma_{1:T})$ of $(Z_{1:T}, \Gamma_{1:T})$, we have that for any

$z = (z^{(1)}, \dots z^{(K)}) \in \mathcal{Z}^*$,

$$\mathbb{P}(Z_{t+1} = z \mid Z_{1:t} = z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}) = \prod_{k \in \mathcal{K}} Q^{(k)}(z^{(k)} \mid z_t, \gamma_t^{(k)}). \quad (10)$$

Lemma 4: For any realization $(z_{1:t}, \gamma_{1:t})$ of $(Z_{1:t}, \Gamma_{1:t})$, we have:

$$\begin{aligned}\mathbb{E}[C_t^{(k)} \mid Z_{1:t} = z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}] &= \mathbb{E}[C_t^{(k)} \mid Z_t = z_t, \Gamma_t^{(k)} = \gamma_t^{(k)}] \\ &=: \ell_t^{(k)}(z_t, \gamma_t^{(k)}) \end{aligned} \quad (11)$$

B. A game between virtual players

Following [15], we consider a stochastic game between K virtual players. At time t, the state is $Z_t = (Z_t^{(1)}, \dots, Z_t^{(K)}) \in \mathcal{Z}^*$; virtual player $k \in \mathcal{K}$ observes Z_t, chooses the prescription $\gamma_t^{(k)} \in \Gamma^{(k)}$, and incurs a per-step cost $\ell^{(k)}(Z_t, \gamma_t^{(k)})$ given by (11). The initial state Z_1 has a probability mass function given by:

$$\begin{aligned}\mathbb{P}(Z_1 = z) &= \prod_{k \in \mathcal{K}} \mathbb{P}(Z_1^{(k)} = z^{(k)}) \\ &= \prod_{k \in \mathcal{K}} \sum_{x^{(k)} \in (\mathcal{X}^{(k)})^{N^{(k)}}} \prod_{i \in \mathcal{N}^{(k)}} P_0^{(k)}(x_1^i). \end{aligned} \quad (12)$$

The state Z_t evolves in a controlled Markov manner according to (10).

In information available to the virtual player at time t is the history of mean-fields $Z_{1:t}$. Virtual player k selects the prescription according to a strategy $\varphi^{(k)}$, i.e.,

$$\Gamma_t^{(k)} = \varphi^{(k)}(Z_{1:t}).$$

Let $\varphi^{(k)} = (\varphi_1^{(k)}, \dots, \varphi_T^{(k)})$ denote the strategy of virtual player k. Then, the total cost incurred by virtual player k is given by:

$$L^{(k)}(\varphi^{(k)}, \varphi^{(-k)}) = \mathbb{E}\Big[\sum_{t=1}^{T} \ell_t^{(k)}(Z_t, \Gamma^{(k)})\Big], \quad (13)$$

We are interested in the following:

Game 2: Given the system model described above, identify a Nash equilibrium strategy $\varphi^* = (\varphi^{*(k)})_{k \in \mathcal{K}}$, i.e., $\varphi_t^{*(k)} \colon Z_t \mapsto \Gamma^{(k)}$, i.e., for any other strategy $\varphi = (\varphi^{(k)})_{k \in \mathcal{K}}$, we have

$$L^{(k)}(\varphi^{*(k)}, \varphi^{*(-k)}) \leqslant L^{(k)}(\varphi^{(k)}, \varphi^{*(-k)}), \quad \forall k \in \mathcal{K}. \quad (14)$$

C. Relationship between Games 1 and 2

We have the following result that connects the solutions of Game 1 and Game 2.

Theorem 1: Let $g = (g^{(1)}, \dots, g^{(K)})$ be a Team-Nash equilibrium of Game 1. Define a strategy $\varphi = (\varphi^{(1)}, \dots, \varphi^{(K)})$ for Game 2 as follows: for any $z_{1:t} \in \mathcal{Z}^{*t}$:

$$\varphi_t^{(k)}(z_{1:t}) = g_t^{(k)}(\cdot, z_{1:t}). \quad (15)$$

Then φ is a Nash equilibrium for Game 2.

Conversely, let $\varphi = (\varphi^{(1)}, \dots, \varphi^{(K)})$ by any Nash equilibrium for Game 2. Define a strategy $g = (g^{(1)}, \dots, g^{(k)})$ for Game 1 as follows: for any $x \in \mathcal{X}^{(k)}$ and $z \in \mathcal{Z}^*$,

$$g_t^{(k)}(x, z_{1:t}) = \varphi_t^{(k)}(z_{1:t})(x). \quad (16)$$

Then g is a Team-Nash equilibrium of Game 1.

Proof: The proof follows from a sample path equivalence argument between the two games. Details are omitted due to space limitations. ∎

D. Markov perfect equilibrium for Game 2

Game 2 among virtual players is a game with perfect information since all players choose prescriptions based on the history $Z_{1:t}$ of mean-field which is common knowledge between the players. Lemmas 2 and 3 imply that we can view the current mean-field Z_t as the "state" of system. Following [13], [14], we restrict attention to Markov perfect equilibrium for Game 2, which can be thought of as a subgame perfect equilibrium of Game 2 where all virtual players are playing Markov strategies which map current state to prescriptions. Such Markov perfect equilibrium can be obtained using dynamic programming as follows.

Theorem 2: Consider a strategy profile $\psi = (\psi^{(k)})_{k \in \mathcal{K}}$, where each virtual player is playing a Markov strategy, i.e., $\psi_t^{(k)} \colon Z_t \mapsto \Gamma_t^{(k)}$.

A necessary and sufficient condition for ψ to be a Markov perfect equilibrium for Game 2 is that it satisfy the following conditions:

1) For each possible realization z_T of Z_T, define the value function for virtual player k:
$$V_T^k(z_T) = \min_{\gamma_T^{(k)}} \ell_T^{(k)}(z_T, \gamma_T^{(k)}). \quad (17)$$
Then, $\psi_T^{(k)}(z_T)$ must be a minimizing $\gamma_T^{(k)}$ in (17).
2) For $t \in \{T-1, \dots, 1\}$ and for each possible realization z_t of Z_t, recursively define the value function for virtual player k:
$$\begin{aligned} V_t^k(z_t) = \min_{\gamma_t^{(k)}} \mathbb{E}\big[&\ell_t^{(k)}(z_t, \gamma_t^{(k)}) + \\ &V_{t+1}^k(Z_{t+1}) \mid Z_t = z_t, \Gamma_t^{(k)} = \gamma_t^{(k)}, \\ &\Gamma_t^{(-k)} = \psi_t^{(-k)}(z_t)\big]\big], \end{aligned} \quad (18)$$
where the expectation is with respect to the distribution (10). Then, $\psi_t^{(k)}(z_t)$ must be a minimizing $\gamma_t^{(k)}$ in (18).

Theorem 2 states that the Markov perfect equilibrium for the virtual players can be obtained by dynamic programming. Let ψ be such a Markov perfect equilibrium. Let g be the policy obtained by (16). Then, by Theorem 1, g is a Team-Nash equilibrium of Game 1 and is called the *common information based Markov perfect equilibrium*(CIB-MPE).

IV. Efficient algorithm to sample the mean-field dynamics

Theorem 2 provides a dynamic program to identify a CIB-MPE but numerically solving it suffers from the curse

of dimensionality. One alternative is to use sampling based methods [19]–[21] to approximately solve the dynamic program.

These sampling methods assume that given the current state and action, one has the ability to sample the next state. For the dynamic program of Game 2, this translates to being able to sample Z_{t+1} give (Z_t, Γ_t), i.e., sample from the transition kernel $Q^{(k)}$ defined in Lemma 2.

It is not immediately obvious how this sampling should be done. Explicitly constructing $Q^{(k)}$ as a transition matrix is infeasible due to exponentially large size of the state space $\mathcal{Z}^*$. Trying to construct a whole system simulator is also difficult when each team has a large number of agents. In this section, we present an efficient method to sample from the transition matrix $Q^{(k)}$.

To describe this method, we need to work with state counts rather than the mean-field. Given a realization $(x_t^{(k)}, u_t^{(k)}, x_{t+1}^{(k+1)})$ of $(X_t^{(k)}, U_t^{(k)}, X_{t+1}^{(k+1)})$ and any $x, x' \in \mathcal{X}^{(k)}$ and $u \in \mathcal{U}^{(k)}$, define the following:

- *State counts*, denoted by $m_t^{(k)}(x)$, which count the number of agents of team k in state x and is given by:
$$m_t^{(k)}(x) = \sum_{i \in N^{(k)}} \mathbb{1}\{X_t^i = x\}.$$
- *State-action counts*, deonted by $\bar{m}_t^{(k)}(x,u)$, count the number of agents of team k in state x that take action u and is given as:
$$\bar{m}_t^{(k)}(x,u) = \sum_{i \in N^{(k)}} \mathbb{1}\{X_t^i = x, U_t^i = u\}.$$
- *State-action-next state counts*, denoted by $\widehat{m}_t^{(k)}(x,u,x')$, count the number of agents of team k in state x that take action u and reach next state x' and is given by:
$$\widehat{m}_t^{(k)}(x,u,x') = \sum_{i \in N^{(k)}} \mathbb{1}\{X_t^i = x, U_t^i = u, X_{t+1}^u = x'\}.$$

We also note that:

$$z_t^{(k)} = \frac{1}{N^{(k)}} m_t^{(k)}. \tag{19}$$

Now observe that for any global states x_t, x_{t+1} and global actions u_t, we have

$$\begin{aligned}
&\mathbb{P}(X_{t+1} = x_{t+1} \mid x_t, u_t) \\
&= \prod_{k \in \mathcal{K}} \prod_{i \in \mathcal{N}^{(k)}} P^{(k)}(x_{t+1}^i \mid x_t^i, u_t^i, \xi(x_t)) \\
&= \prod_{k \in \mathcal{K}} \prod_{\substack{x \in \mathcal{X}^{(k)} \\ u \in \mathcal{U}^{(k)}}} P^{(k)}(x_{t+1}^i \mid x, u, \xi(x_t))^{\widehat{m}_t^{(k)}(x,u,x_{t+1}^i)}
\end{aligned} \tag{20}$$

Notice that computing the right-hand-side expression in (20) requires only aggregate information, namely counts $\widehat{m}^k(x,u,x_{t+1}^i)$, for each team k. Such symmetries can be exploited to directly sample each mean field $Z_{t+1}^{(k)}$ given $(z_t^{(k)}, \gamma_t)$. This process is going to be sample efficient as we no longer need to simulate the trajectory of each agent; mean field can be updated directly by mapping the dynamics of the mean field evolution to a collective graphical model [22]. Such a process has been explored earlier in collective decentralized POMDPs [23]. However, this earlier method only had a single team type, we show the sampling step when there are teams of different types.

Next we show how to use this method to sample $Z_{t+1}^{(k)}$ given $(z^{(k)}, \gamma_t^{(k)})$.

a) Sampling state-action counts: Let $m_t^{(k)}$ and $\bar{m}_t^{(k)}$ be consistent values of state-counts and state-action counts, i.e.,

$$m_t^{(k)}(x) = \sum_{u \in \mathcal{U}^{(k)}} \bar{m}_t^{(k)}(x,u), \quad \forall x \in \mathcal{X}^{(k)}.$$

Then, from a basic combinatorial counting argument, we get

$$\begin{aligned}
&P(\bar{M}_t^{(k)} = \bar{m}_t^{(k)} | M_t = m_t, \Gamma_t^{(k)} = \gamma_t^{(k)}) \\
&= \prod_{x \in \mathcal{X}^{(k)}} \Big[\frac{m_t^{(k)}(x)!}{\prod_u \bar{m}_t^{(k)}(x,u)} \times \prod_u \gamma_t^{(k)}(u|x)^{\bar{m}_t^{(k)}(x,u)} \Big],
\end{aligned} \tag{21}$$

which is a multinomial distribution. Thus, we can efficiently sample the state-action counts $\bar{M}_t^{(k)}$ given $(z_t, \gamma_t^{(k)})$ by sampling from a multinomial distribution

b) Sampling state-action-state counts: Let $m_t^{(k)}$, $\bar{m}_t^{(k)}$ and $\widehat{m}_t^{(k)}$ be consistent values of state counts, state-action counts and state-action-next state counts, i.e.,

$$\bar{m}_t^k(x,u) = \sum_{x' \in \mathcal{X}^{(k)}} \widehat{m}_t^{(k)}(x,u,x'), \quad \forall x,u \in \mathcal{X}^{(k)} \times U^{(k)}.$$

Let z_t be the mean-field corresponding to $(m_t^{(1)}, \dots, m_t^{(K)})$. Then, from a basic combinatorial counting argument, we get

$$\begin{aligned}
&\bar{P}(\widehat{M}_t^{(k)} = \widehat{m}_t^{(k)} | \bar{M}_t^{(k)} = \bar{m}_t^{(k)}, M_t = m_t) \\
&= \prod_{x,u} \Big[\frac{\bar{m}_t^{(k)}(x,u)!}{\prod_{x'} \widehat{m}_t^{(k)}(x,u,x')!} \times \\
&\qquad \prod_{x'} P^{(k)}\big(x' \mid x, u, z_t\big)^{\widehat{m}_t^{(k)}(x,u,x')} \Big],
\end{aligned} \tag{22}$$

which is also a multinomial distribution. Thus, we can efficiently sample $\widehat{M}_t^{(k)}$ given $(\bar{m}_t^{(k)}, z_t)$ by sampling from a multinomial distribution.

Finally observe that

$$M_{t+1}^{(k)}(x') = \sum_{x \in \mathcal{X}^{(k)}, u \in \mathcal{U}^{(k)}} \widehat{M}_t^{(k)}(x,u,x').$$

Thus, if we "marginalize" the sampled state-action-next state count $\widehat{M}_t^{(k)}$, we will obtain the state count $M_{t+1}^{(k)}$. Normalizing $M_{t+1}^{(k)}$ gives us the mean-field $Z_{t+1}^{(k)}$. In particular

$$Z_{t+1}^{(k)} = \frac{1}{N^{(k)}} M_{t+1}^{(k)}. \tag{23}$$

Thus, combining (20)–(23) provides an efficient method to sample $Z_{t+1}^{(k)}$ given $(z_t, \gamma^{(k)})$. This sampling procedure can be used in the inner loop of any sampling based algoritm to solve the coupling dynamic programs of Theorem 2.

V. Mean-field approximation

We note that solving the dynamic program (17), (18), to yield a solution to Game 2 requires knowing the $Q^{(k)}$ function given in (10), which is difficult to compute. Though, one option is to sample from this function as explained in Sec. IV, this is still computationally involved, especially in cases where $\mathcal{Z}^*$ is high-dimensional and/or Z_t has complex dynamics. In these situations, it is desirable to consider a mean-field approximation of Game 2, in which one assumes that each team has an infinite number of agents. In this limit, the mean field Z_t evolves in a deterministic manner, which drastically simplifies the dynamic program of Theorem 2. For such a mean-field approximation to be meaningful, the MPE equilibrium of the infinite population system must be an approximate equilibrium of the finite population system. We establish such an approximation result in this section.

A. Mean-field limit

In the limiting case, where each team has an infinite population, each of the mean-fields evolves in a deterministic manner. Let $\bar{z}^{(k)} \in \bar{\mathcal{Z}}^{(k)}$ denote the mean-field in such the infinite population system with $\bar{\mathcal{Z}}^{(k)} = \Delta(\mathcal{X}^{(k)})$ for each $k \in \mathcal{K}$. Furthermore, we extend the domain of the cost functions and transition functions to $\Delta(\mathcal{X}^{(k)})$. In such a system, the mean-field evolution given prescription γ and current mean-field $\bar{z}_t$ changes from the finite population case (10) to:

$$\begin{aligned}\bar{z}^{(k)}_{t+1}(x') &= \sum_{x\in\mathcal{X}^{(k)}} \bar{z}^{(k)}_t(x) P^{(k)}(x'|x, \gamma^{(k)}_t(x), \bar{z}_t) \\ &=: q^{(k)}(\bar{z}_t, \gamma^{(k)}_t),\end{aligned} \tag{24}$$

for each $k \in \mathcal{K}$. Furthermore, the per-step total team cost given by (2), (11) is modified as:

$$\bar{\ell}^{(k)}_t(\bar{z}, \gamma^{(k)}) = \sum_{x\in\mathcal{X}^{(k)}} \bar{z}(x) c^{(k)}_t(x, \gamma^{(k)}(x), \bar{z}). \tag{25}$$

The Markov perfect equilibrium of this mean-field limit is characterized as follow.

Theorem 3: Consider a strategy profile $\bar{\psi} = (\bar{\psi}^{(k)})_{k\in\mathcal{K}}$, where each virtual player is playing a Markov strategy.

A necessary and sufficient condition for $\bar{\psi}$ to be a Markov perfect equilibrium for the mean-field limit of Game 2 is that it satisfy the following conditions:

1) For each possible realization $\bar{z}_T$ of $\bar{Z}_T$, define the value function for virtual player k:
$$\bar{V}^k_T(z_T) = \min_{\gamma^{(k)}_T} \bar{\ell}^{(k)}_T(\bar{z}_T, \gamma^{(k)}_T). \tag{26}$$
Then, $\psi^{(k)}_T(\bar{z}_T)$ must be a minimizing $\gamma^{(k)}_T$ in (26).
2) For $t \in \{T-1, \dots, 1\}$ and for each possible realization $\bar{z}_t$ of $\bar{Z}_t$, recursively define the value function for virtual player k:
$$\begin{aligned}\bar{V}^k_t(\bar{z}_t) = \min_{\gamma^{(k)}_t} \Big\{ &\bar{\ell}^{(k)}_t(\bar{z}_t, \gamma^{(k)}_t) \\ &+ V^k_{t+1}\big((q^{(k)}(\bar{z}_t, \gamma^{(k)}_t))_{k\in\mathcal{K}}\big)\Big\}\end{aligned} \tag{27}$$
where the deterministic evolution of $\bar{z}_{t+1}$ is as given in (24). Then, $\bar{\psi}^{(k)}_t(\bar{z}_t)$ must be a minimizing $\gamma^{(k)}_t$ in (27).

B. Lipschitz continuity

We begin by defining some preliminary quantities. Let d_x be a metric on the state space $\mathcal{X}^{(k)}$ for all $k \in \mathcal{K}$. Then, based on this metric, let $d_{\mathfrak{w}}$ be the Kantorovich metric (also called Wasserstein metric) on the space of mean-fields for each team $k \in \mathcal{K}$. Define a metric on the set of mean-fields for all teams $\mathcal{Z}^*$ as:

$$d_{\mathcal{K}}(Z, \widehat{Z}) = \sum_{k\in\mathcal{K}} \Big(d_{\mathfrak{w}}(Z^{(k)}, \widehat{Z^{(k)}}) \Big), \tag{28}$$

We now make the following assumptions:

Assumption 1: For each $k \in \mathcal{K}$, the per step cost function $c^{(k)}_t(x, u, z)$, is Lipschitz continuous with respect to the local state and the mean-field with Lipschitz constants $\mathcal{L}_x$ and $\mathcal{L}_z$, respectively, i.e., for any $x, x_+ \in \mathcal{X}^{(k)}$, $u \in \mathcal{U}^{(k)}$ and $\bar{z}, \bar{z}_+ \in \bar{\mathcal{Z}}^*$:

$$|c^{(k)}_t(x, u, \bar{z}) - c^{(k)}_t(x_+, u, \bar{z})| \leqslant \mathcal{L}_x d_x(x, x_+) \tag{29}$$

$$|c^{(k)}_t(x, u, \bar{z}) - c^{(k)}_t(x, u, \bar{z}_+)| \leqslant \mathcal{L}_z d_{\mathcal{K}}(\bar{z}, \bar{z}_+). \tag{30}$$

Assumption 2: The transition function $P^{(k)}(X \mid x, u, \bar{z})$ is Lipschitz continuous with respect to the Wasserstein metric $d_{\mathfrak{w}}$, with respect to the local state and the mean-field, with Lipschitz constants $\mathcal{L}_{P,x}$ and $\mathcal{L}_{P,z}$ respectively, i.e., for any $x, x_+ \in \mathcal{X}^{(k)}$, $u \in \mathcal{U}^{(k)}$ and $\bar{z}, \bar{z}_+ \in \bar{\mathcal{Z}}^*$:

$$\begin{aligned}d_{\mathfrak{w}}(P^{(k)}(X \mid x, u, \bar{z}), &P^{(k)}(X \mid x_+, u, \bar{z})) \\ &\leqslant \mathcal{L}_{P,x} d_x(x, x_+)\end{aligned} \tag{31}$$

$$\begin{aligned}d_{\mathfrak{w}}(P^{(k)}(X \mid x, u, \bar{z}), &P^{(k)}(X \mid x, u, \bar{z}_+)) \\ &\leqslant \mathcal{L}_{P,z} d_{\mathcal{K}}(\bar{z}, \bar{z}_+).\end{aligned} \tag{32}$$

These assumptions lead to the following results. The proofs are provided in the appendix.

Lemma 5: The per-step cost function for virtual player $k \in \mathcal{K}$ in the mean-field limit, $\bar{\ell}^{(k)}_t(\bar{z}, \gamma^{(k)})$ given by (25) is Lipschitz continuous with respect to its first argument $\bar{z}$, with Lipschitz constant $\mathcal{L}_C = \mathcal{L}_z + \mathcal{L}_x$, i.e.,

$$|\bar{\ell}^{(k)}_t(\bar{z}, \gamma^{(k)}) - \bar{\ell}^{(k)}_t(\bar{z}_+, \gamma^{(k)})| \leqslant \mathcal{L}_C d_{\mathcal{K}}(\bar{z}, \bar{z}_+). \tag{33}$$

Lemma 6: The transition function given by (24) is Lipschitz continuous with respect to the mean-field $\bar{z}$ in terms of the Wasserstein metric $d_{\mathfrak{w}}$, with Lipschitz constant $\mathcal{L}_P$, i.e,:

$$d_{\mathfrak{w}}\Big(q^{(k)}(\bar{z}, \gamma^{(k)}), q^{(k)}(\bar{z}_+, \gamma^{(k)})\Big) \leqslant \mathcal{L}_P d_{\mathcal{K}}(\bar{z}, \bar{z}_+),$$

where $\mathcal{L}_P = \max_{k\in\mathcal{K}} \Big[\frac{\mathtt{diam}(\mathcal{X}^{(k)}) |\mathcal{X}^{(k)}|}{2} (\mathcal{L}_{P,z} + \mathcal{L}_{P,x}) \Big]$.

Lemma 7: The transition function for all the mean-fields, $q = (q^{(1)}, \dots, q^{(K)})$ where each $q^{(k)}$ us given by (24) is Lipschitz continuous with respect to the mean-field $\bar{z}$ in terms of the metric $d_{\mathcal{K}}$ defined in (28), with Lipschitz constant $\mathcal{L}_q = K\mathcal{L}_P$, i.e,:

$$d_{\mathcal{K}}\Big(q(\bar{z}, \gamma), q(\bar{z}_+, \gamma)\Big) \leqslant \mathcal{L}_q d_{\mathcal{K}}(\bar{z}, \bar{z}_+),$$

Lemma 8: For any Lipschitz continuous policy $\psi^{(k)}_t : Z_t \mapsto \gamma^{(k)}_t$, the value function $\bar{V}^{(k)}_t$ is Lipschitz continuous with Lipschitz constant $\mathcal{L}_V = \mathcal{L}_C \sum_{s=0}^{T-t} (\mathcal{L}_q)^s$.

Proof: The proof follows a similar approach to the one in [24], [25] for the finite horizon case, as given in [26], as both the per-step cost function and the transition function are Lipschitz continuous. The detailed proof is omitted due to space constraints. ■

C. ϵ-Team-Nash equilibrium

Given Assumptions 1 and 2, we have shown that the mean-field limit of Game 2 is such that from the perspective of each virtual player, the resulting Markov decision process for that virtual player has Lipschitz continuous cost function, transition function and value function. We will now present an approximation bound on the difference in the value functions for any virtual player k, when that virtual player unilaterally deviates from using the mean-field limit approximation and instead uses the exact dynamics. We use the approximate information state theorem from [27] to bound the performance difference between the mean-field limit case and the finite population case for virtual player k as given in the following theorem.

Theorem 4: Let $\bar{\psi} = (\bar{\psi}^{(k)})_{k\in\mathcal{K}}$ be a Markov perfect equilibrium for the infinite population mean-field described in Thm. 3. Then, for any virtual player $k \in \mathcal{K}$ and any other Markov strategy $\psi^{(k)}$,

$$L^{(k)}(\bar{\psi}^{(k)}, \bar{\psi}^{(-k)}) \leqslant L^{(k)}(\psi^{(k)}, \bar{\psi}^{(-k)}) + \epsilon, \qquad (34)$$

where $L^{(k)}$ is defined in (14) and $\epsilon = 2(T-t)\mathcal{L}_V \kappa \sum_{k\in\mathcal{K}} \frac{1}{\sqrt{N^{(k)}}}$

Proof: Fix a virtual player k and the strategy profile $\bar{\psi}^{(-k)}$ for the other virtual players and consider teh best response dynamics at virtual player k given by the dynamic program in Thm. 2. The idea of proof is to show that the history compression function $\nu_t(z_{1:t}, \gamma_{1:t}) = z_t$ dynamics $(q_t^{(k)})_{k\in\mathcal{K}}$ and the per-step cost $\bar{\ell}_t^{(k)}$ is an approximate information state (AIS) as defined in [27]. In particular, we observe that:

$$\mathbb{E}[\ell_t^{(k)}(z, \gamma^{(k)}) - \bar{\ell}_t^{(k)}(\nu_t(z_{1:t}, \gamma_{1:t}), \gamma^{(k)})] = 0 \coloneqq \varepsilon_t, \quad (35)$$

which follows from the definitions of $\ell^{(k)}$ and $\bar{\ell}^{(k)}$ and the fact that $\nu_t(z_{1:t}, \gamma_{1:t}) = z_t$. Furthermore, we have:

$$\begin{aligned} &d_{\mathcal{K}}(\mathbb{P}(Z_{t+1}|Z_t = z_t, \Gamma_t = \gamma_t), q_t(z_t, \gamma_t)) \\ &\quad \overset{(a)}{=} \sum_{k\in\mathcal{K}} d_{\mathfrak{w}}(\mathbb{P}(Z_{t+1}^{(k)}|Z_t = z_t, \Gamma_t^{(k)} = \gamma_t^{(k)}), q_t^{(k)}(z_t, \gamma_t^{(k)})) \\ &\quad \overset{(b)}{\leqslant} \sum_{k\in\mathcal{K}} \frac{\kappa}{\sqrt{N^{(k)}}} \coloneqq \delta_t, \end{aligned} \qquad (36)$$

where (a) follows from the definition of $d_{\mathcal{K}}$ and (b) follows from the concentration of empirical measure to statistical measure with respect to the Wasserstein distance [28] where κ is a constant that depends on the state spaces $\mathcal{X}^{(k)}$ and the metric d_x. Equations (35), (36) show that $(_t, (q_t^{(k)})_{k\in\mathcal{K}}, \bar{\ell}_t^{(k)})$ is an AIS. Then, the reuslt follows from [27, Theorem 9]. ■

From Thm 4, we find that the mean-field limit strategy is an ϵ-Team-Nash equilibrium for virtual player k, where $\epsilon = 2(T-t)\mathcal{L}_V \frac{\kappa}{K} \sum_{k\in\mathcal{K}} \frac{1}{\sqrt{N^{(k)}}}$.

VI. Conclusion

In this paper, we presented a model for mean-field games among teams and presented a common-information based refinement of the Team-Nash equilibrium for this game. This common-information based Markov perfect equilibrium can be obtained by solving coupled dynamic programs. These dynamic programs use the mean-field of all teams as a state. In general, solving such dynamic programs suffers from the curse of dimensionality.

We present two methods to efficiently solve these dynamic programs. The first is an efficient method to sample the mean-field dynamics by using the state counts. The second is using a mean-field limit to approximate finite population teams by an infinite population. We show that a Markov perfect equilibrium obtained using the mean-field approximations posses an ϵ-Nash property.

Appendix

A. Proof of Lemma 5

Consider

$$\begin{aligned} &\big|\bar{\ell}_t^{(k)}(\bar{z}, \gamma^{(k)}) - \bar{\ell}_t^{(k)}(\bar{z}_+, \gamma^{(k)})\big| \\ &\overset{(a)}{=} \Big|\sum_{x\in\mathcal{X}^{(k)}} \bar{z}(x) c_t^{(k)}(x, \gamma^{(k)}(x), \bar{z})\Big| \\ &\qquad - \Big|\sum_{x\in\mathcal{X}^{(k)}} \bar{z}_+(x) c_t^{(k)}(x, \gamma^{(k)}(x), \bar{z}_+)\Big| \\ &\overset{(b)}{\leqslant} \Big|\sum_{x\in\mathcal{X}^{(k)}} \bar{z}(x)(c_t^{(k)}(x, \gamma^{(k)}(x), \bar{z}) - c_t^{(k)}(x, \gamma^{(k)}(x), \bar{z}_+))\Big| \\ &\qquad + \Big|\sum_{x\in\mathcal{X}^{(k)}} c_t^{(k)}(x, \gamma^{(k)}(x), \bar{z}_+)(\bar{z}(x) - \bar{z}_+(x))\Big| \\ &\overset{(c)}{\leqslant} \mathcal{L}_z d_{\mathcal{K}}(\bar{z}, \bar{z}_+) + \mathcal{L}_x d_{\mathcal{K}}(\bar{z}, \bar{z}_+) \\ &= (\mathcal{L}_z + \mathcal{L}_x) d_{\mathcal{K}}(\bar{z}, \bar{z}_+), \end{aligned} \qquad (37)$$

where (a) follows from (25), (b) follows from adding and subtracting $\sum_{x\in\mathcal{X}^{(k)}} \bar{z}(x)(\bar{c}_t^{(k)}(x, \gamma^{(k)}(x), \bar{z}_+)$ and then using the triangle inequality and regrouping terms, the first term of (c) follows from (29) and the second term follows from (30) and the Kantorovich-Rubinstein duality.

B. Proof of Lemma 6

Let us consider the total divergence distance between the distributions which are the outputs of the transition function

$q^{(k)}$ in (24):

$$
\begin{aligned}
&d_{TV}\Big(q^{(k)}(\bar{z},\gamma), q^{(k)}(\bar{z}_+,\gamma)\Big) \\
&\quad \overset{(a)}{=} \frac{1}{2} \sum_{x' \in \mathcal{X}^{(k)}} \Big[\Big| \sum_{x \in \mathcal{X}^{(k)}} \bar{z}(x) P^{(k)}(x'|x, \gamma^{(k)}(x), \bar{z}) - \\
&\qquad \sum_{x \in \mathcal{X}^{(k)}} \bar{z}_+(x) P^{(k)}(x'|x, \gamma^{(k)}(x), \bar{z}_+) \Big| \Big] \\
&\quad \overset{(b)}{\leqslant} \frac{1}{2} \sum_{x' \in \mathcal{X}^{(k)}} \Big[\Big| \sum_{x \in \mathcal{X}^{(k)}} \bar{z}(x) P^{(k)}(x'|x, \gamma^{(k)}(x), \bar{z}) - \\
&\qquad \sum_{x \in \mathcal{X}^{(k)}} \bar{z}(x) P^{(k)}(x'|x, \gamma^{(k)}(x), \bar{z}_+) \Big| \Big] \\
&\qquad + \frac{1}{2} \sum_{x' \in \mathcal{X}^{(k)}} \Big[\Big| \sum_{x \in \mathcal{X}^{(k)}} \bar{z}(x) P^{(k)}(x'|x, \gamma^{(k)}(x), \bar{z}_+) - \\
&\qquad \sum_{x \in \mathcal{X}^{(k)}} \bar{z}_+(x) P^{(k)}(x'|x, \gamma^{(k)}(x), \bar{z}_+) \Big| \Big] \\
&\quad \overset{(c)}{=} \frac{|\mathcal{X}^{(k)}|}{2} (\mathcal{L}_z d_{\mathcal{K}}(\bar{z}, \bar{z}_+) + \mathcal{L}_x d_{\mathcal{K}}(\bar{z}, \bar{z}_+)), \qquad (38)
\end{aligned}
$$

where (a) follows from the definition of $q^{(k)}$, (b) from addition and subtraction of the same term and using the triangle inequality, the first term in (c) follows from Assumption 2 and the second term from the Kantorovich-Rubinstein duality. Then, we have

$$
\begin{aligned}
&d_{\mathfrak{w}}\Big(q^{(k)}(\bar{z},\gamma), q^{(k)}(\bar{z}_+,\gamma)\Big) \\
&\overset{(a)}{\leqslant} \mathtt{diam}(\mathcal{X}^{(k)}) d_{TV}\Big(q^{(k)}(\bar{z},\gamma), q^{(k)}(\bar{z}_+,\gamma)\Big) \\
&\overset{(b)}{\leqslant} \frac{\mathtt{diam}(\mathcal{X}^{(k)}) |\mathcal{X}^{(k)}|}{2} (\mathcal{L}_z + \mathcal{L}_x) d_{\mathcal{K}}(\bar{z}, \bar{z}_+), \qquad (39)
\end{aligned}
$$

where (a) follows from [29] and (b) from (38).

C. Proof of Lemma 7

Consider,

$$
\begin{aligned}
&d_{\mathcal{K}}\Big(q(\bar{z},\gamma), q(\bar{z}_+,\gamma)\Big) \\
&\overset{(a)}{\leqslant} \sum_{k \in \mathcal{K}} d_{\mathfrak{w}}\Big(q^{(k)}(\bar{z},\gamma), q^{(k)}(\bar{z}_+,\gamma)\Big) \\
&\overset{(b)}{\leqslant} K \mathcal{L}_P d_{\mathcal{K}}(\bar{z}, \bar{z}_+) \qquad (40)
\end{aligned}
$$

where (a) follows from (28) and the triangle inequality and (b) follows from Lemma 6.

REFERENCES

[1] J. Marschak and R. Radner, *Economic theory of teams.* Yale University Press, 1972.

[2] J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*. Princeton University Press, 1944.

[3] G. Y. Weintraub, C. L. Benkard, and B. Van Roy, "Markov perfect industry dynamics with many firms," *Econometrica*, vol. 76, no. 6, pp. 1375–1411, 2008.

[4] T. Harbert. (2014) Radio wrestlers fight it out at the DARPA spectrum challenge. [Online]. Available: https://spectrum.ieee.org/telecom/wireless/radio-wrestlers-fight-it-out-at-the-darpa-spectrum-challenge

[5] L. S. Shapley, "A value for n-person games," *Contributions to the Theory of Games*, vol. 2, no. 28, pp. 307–317, 1953.

[6] D. Tang, H. Tavafoghi, V. Subramanian, A. Nayyar, and D. Teneketzis, "Dynamic games among teams with delayed intra-team information sharing," *arXiv preprint arXiv:2102.11920*, 2021.

[7] A. R. Pedram and T. Tanaka, "Linearly-solvable mean-field approximation for multi-team road traffic games," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 1243–1248.

[8] X. Yu, Y. Zhang, and Z. Zhou, "Teamwise mean field competitions," *Available at SSRN 3638969*, 2020.

[9] M. Huang, P. E. Caines, and R. P. Malhamé, "Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized epsilon-Nash equilibria," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1560–1571, 2007.

[10] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japanese Journal of Mathematics*, vol. 2, no. 1, pp. 229–260, 2007.

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

[12] ——, "Linear quadratic mean field teams: Optimal and approximately optimal decentralized solutions," 2016, arXiv:1609.00056v2.

[13] E. Maskin and J. Tirole, "A theory of dynamic oligopoly, I: Overview and quantity competition with large fixed costs," *Econometrica: Journal of the Econometric Society*, pp. 549–569, 1988.

[14] ——, "A theory of dynamic oligopoly, II: Price competition, kinked demand curves, and edgeworth cycles," *Econometrica: Journal of the Econometric Society*, pp. 571–599, 1988.

[15] A. Nayyar, A. Gupta, C. Langbort, and T. Başar, "Common information based Markov perfect equilibria for stochastic games with asymmetric information: Finite games," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 555–570, 2013.

[16] Y. Ouyang, H. Tavafoghi, and D. Teneketzis, "Dynamic Games With Asymmetric Information: Common Information Based Perfect Bayesian Equilibria and Sequential Decomposition," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 222–237, 2017.

[17] D. M. Kreps and J. Sobel, "Signalling," *Handbook of game theory with economic applications*, vol. 2, pp. 849–867, 1994.

[18] Y.-C. Ho, "Team decision theory and information structures," *Proceedings of the IEEE*, vol. 68, no. 6, pp. 644–654, 1980.

[19] W. B. Haskell, R. Jain, and D. Kalathil, "Empirical dynamic programming," *Mathematics of Operations Research*, vol. 41, no. 2, pp. 402–429, 2016.

[20] W. B. Haskell, R. Jain, H. Sharma, and P. Yu, "A universal empirical dynamic programming algorithm for continuous state MDPs," *IEEE Transactions on Automatic Control*, vol. 65, no. 1, pp. 115–129, 2019.

[21] H. Chang, M. Fu, J. Hu, and S. Marcus, *Simulation-based Algorithms for Markov Decision Processes*, ser. Communications and Control Engineering. Springer, 2007.

[22] D. R. Sheldon and T. G. Dietterich, "Collective graphical models," in *Neural Information Processing Systems*, 2011, pp. 1161–1169.

[23] D. T. Nguyen, A. Kumar, and H. C. Lau, "Collective multiagent sequential decision making under uncertainty," in *AAAI Conference on Artificial Intelligence*, 2017, pp. 3036–3043.

[24] K. Hinderer, "Lipschitz continuity of value functions in Markovian decision processes," *Mathematical Methods of Operations Research*, vol. 62, no. 1, pp. 3–22, 2005.

[25] E. Rachelson and M. G. Lagoudakis, "On the locality of action domination in sequential decision making," in *International Symposium on Artificial Intelligence and Mathematics*, Fort Lauderdale, US, Jan. 2010.

[26] J. Subramanian, "Reinforcement learning in partially observed and multi-agent systems," Ph.D. dissertation, McGill University, 2020.

[27] J. Subramanian, A. Sinha, R. Seraj, and A. Mahajan, "Approximate information state for approximate planning and reinforcement learning in partially observed systems," 2020, arXiv:2010.08843.

[28] M. Sommerfeld, J. Schrieber, Y. Zemel, and A. Munk, "Optimal transport: Fast probabilistic approximation with exact solvers," 2018.

[29] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International statistical review*, vol. 70, no. 3, pp. 419–435, 2002.