2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Partial Information and Mean Field Games: The Case of a Linear Quadratic Stochastic Aggregative Games with Discrete Observations

Farid Rajabali, Roland Malhamé and Sadegh Bolouki

Abstract— **Mean Field Game equilibria are based on the assumption of instantaneous interactions within a population of interchangeable agents, where each agent's impact diminishes as the population size increases. However, in practical scenarios, agents may not continuously observe the overall population state. Instead, in some situations, agents observe the empirical mean state only at discrete time intervals. This observation structure likely influences the nature of Nash equilibria that agents can attain. This paper characterizes the best responses of agents under such discrete observation conditions. Sufficient conditions for the existence of a so-called Markov Nash equilibrium within a finite population of agents are presented. Additionally, the difference in cost due to discrete versus continuous mean observations is evaluated.**

I. INTRODUCTION

Information shapes in critical ways the decision-making process in multi-agent systems. In contexts such as Mean Field Games (MFGs) and aggregative games, based on access to aggregate information established to be sufficient, agents anticipate the statistical characteristics of the population to shape their control policies and navigate effectively. The assumption of sufficient aggregate information access underlies much of the existing MFG literature (See e.g. [1]–[5]). Several works do address situations of partial information within this framework. Thus [6] analyzes the impact on equilibrium of partial own state observability by agents, while [7]–[9] tackle various situations of partial observability within the so-called major-minor agent MFG framework. The papers address linear and nonlinear state models, partial observability by minor agents of their own state and that of the major agent state, as well as partial observability of the major agent state by itself. In addition, [8], [10] establish ε-Nash equilibria for a partially observed major agent. Paper [11] explores MFG with nonlinear dynamics cost functions and addresses a problem with partial state observations, leveraging nonlinear filtering theory and the separation principle. [12], [13] study major-minor agents MFGs with partial observability for all populations. Finally, [14] tackles the partial observability situation for MFGs in discrete time for a risk-sensitive cost structure.

The studies mentioned above generally assume that agents can observe a subset of the population, such as their neighbors, while those farther away remain unobservable. They then attempt to infer information about the unobservable portion based on the data from the visible subset.

In contrast, the objective of the current work is to investigate scenarios where agents only have access to the empirical mean state of the population at discrete time intervals. This setting is motivated by practical situations such as the movement of individuals within a crowd or the dynamics of vehicles in traffic, where continuous observation of the global state is not feasible. Under these conditions, the agents' control policies are determined by a dynamic programming analysis that accounts for the discrete observation structure. This paper characterizes the best response policies and identifies the conditions under which a Nash equilibrium (NE) may arise in a finite population setting. The analysis, conducted within a linear quadratic stochastic mean field framework over a finite time horizon, involves coupled dynamic programs that incorporate both continuous time dynamics and discrete time observations. The results provide insights into the impact of discrete observations on the expected cost incurred by agents due to the inability to continuously observe the empirical mean state.

The research contributions can be outlined as follows:

1) Establishing best response policies for agents under discrete, periodic information sharing, amidst continuous agent dynamics.
2) Quantifying performance degradation, termed as "regret," for periodic observation of empirical mean every Δt seconds and demonstrating a linear growth rate of regret.
3) Establishing the convergence of the specified game with partial observability to their counterparts with complete observation, when the population tends to infinity.

The rest of the paper is organized as follows: In Section 2, we discuss the formulation of the game. In Section 3, we use stochastic DP to find the best response policy for the problem. In Section 4, we calculate the loss of performance due to partial observability, referred to as regret, and show that the regret has a linear growth rate.

II. AN AGGREGATIVE GAME WITH SAMPLED EMPIRICAL MEAN OBSERVATIONS

Consider a non-cooperative game in a population of N agents that are uniform and have scalar dynamics. The dynamics equation for agent i is written in the following as a linear and stochastic differential equation.

$$dx_i(t) = \big(ax_i(t) + bu_i(t)\big)dt + \sigma dw_i(t), \quad t \geq 0 \qquad (1)$$

Research supported by NSERC-RGPIN-2022-0540.

F. Rajabali and R. Malhamé are with the Department of Electrical Engineering at Polytechnique Montréal and GERAD, Canada. Emails: farid.rajabali@polymtl.ca, roland.malhame@polymtl.ca and sadegh.bolouki@polymtl.ca

In (1), $x_i(t)$ is the state of agent i and $u_i(t)$ is the control input or action of agent i. Coefficients a, b are in $\mathbb{R}$ and σ is a non-negative finite value. Noises $w_i(t), i = 1, 2, \ldots, N$ are mutually-independent zero-mean Wiener processes that are also independent from initial agent states. The agents' initial conditions are assumed to be random with finite variance. The agents are assumed to have access to the global empirical mean state $\bar{x}^N = \frac{1}{N}\sum_{i=1}^{N} x_i(t)$ only at discrete-time instants $t_j = t_0 + j\Delta t, j = 0, \ldots, (T/\Delta t) = n$, where T is the time horizon and Δt is the inter-observation time interval, whereas they observe continuously their own state $x_i(t)$.

Agents wish to follow a target $\phi(\bar{x}^N(t)) = \Gamma\bar{x}^N(t) + \eta$, with $\Gamma \in \mathbb{R}^+$, $\eta \in \mathbb{R}$, while minimizing their control effort. This is captured by the following cost function, with $q, r > 0$, $h \geq 0$, $i = 1, \ldots, N$:

$$J_i = E\Bigg[\int_{t_0}^{T} \big[q(x_i(t) - \phi(\bar{x}^N(t)))^2 + ru_i^2(t)\big]\, dt + h\big(x_i(T) - \phi(\bar{x}^N(T))\big)^2\Bigg] \tag{2}$$

Cost function (2) can represent an energy function agents consume while attempting to follow the population mean. Note that an analysis of the game in equations (1), (2) is carried out in [1] under the assumption that $\bar{x}^N(t)$ is continuously observed, while sufficient Riccati equations-related conditions for existence of a NE are derived. Here we wish to identify a set of agent "Markov" control strategies (i.e. relying on the latest agent observations), leading to a potential NE under the described partial information structure. We denote as $X_i(t)$ the pair $(x_i(t), \bar{x}^N(t))$.

III. Predictor-Based Dynamic Programming Equations

A. Problem Formulation

Definition 1. We define the Markov control strategies u_i^* for $i = 1, \ldots, N$ as a Nash equilibrium of the game, if given u_{-i}^*, the vector of Markov strategies of agents other than i, agent i has no incentive to unilaterally change its strategy since doing so, cannot lead to a lower cost.

Recall that a Markov strategy for an agent has been defined as a feedback strategy that depends on time, the current state of the agent and the most recent empirical mean observation, i.e.:

$$u_i = f_i\left(x_i(t), \bar{x}^N(t_j), t\right), \quad t \in [t_j, t_{j+1}] \tag{3}$$

In order to determine this u_i, we first define the following value function for $i = 1, \ldots, N$, $j = 0, \ldots, n$ and $t \in [t_j, T]$:

$$V_{j,i}(t, X_i) = \inf_{u_i \in M_i} E\Big[\int_t^T \big(q(x_i(\tau) - \Gamma\bar{x}^N(\tau) - \eta)^2 + ru_i^2(\tau)\big)d\tau + h(x_i(T) - \Gamma\bar{x}^N(T) - \eta)^2 | \bar{x}^N(t_j)\Big] \tag{4}$$

And the predictor:

$$\hat{\bar{x}}_{j,i}^N(t) = E[\bar{x}^N(t) | \bar{x}^N(t_j), x_i(t)] \quad t \in [t_j, T] \tag{5}$$

where in (4), M_i is the admissible set of Markov control policies of agent i.

Assumption 1. To keep the analysis simple, we assume that N is large enough to neglect the impact of $u_i(t)$ on $\bar{x}^N(t)$. Thus when computing the best response in the context of the game, agent i treats $\bar{x}^N(t)$ as a known value and its predictor, $\hat{\bar{x}}_{j,i}^N(t)$, based on the most recent observations of empirical mean as deterministic although a priori unknown.

Remark 1. Note that since agent i is the only one observing its own state $x_i(t)$, its predictor of $\bar{x}^N(t)$ in (5) will be slightly different from that of other agents $k \neq i$. But based on Assumption 1, we neglect the local effects. Thus, herein, we shall assume that $\hat{\bar{x}}_{j,i}^N(t) \equiv \hat{\bar{x}}_j^N(t)$, $\forall i = 1, \ldots, N$ where:

$$\hat{\bar{x}}_j^N(t) = E[\bar{x}^N(t) | \bar{x}^N(t_j)], \quad t \in [t_j, T] \tag{6}$$

Also, for simplicity, and without loss of generality, we shall assume $\eta = 0$. We define the prediction error for $t \in [t_j, t_{j+1})$ as follows:

$$err_j(t) = \bar{x}^N(t) - \hat{\bar{x}}_j^N(t) \tag{7}$$

Since $\bar{x}^N(t)$ is not observable to agents continuously, we introduce a modified cost function $\widetilde{V}_{j,i}(t, X_i)$ for $t \in [t_j, T)$ that will be shown to result in the same control policy. This is stated in the lemma following definition (8) below.

$$\widetilde{V}_{j,i}(t, X_i) = \inf_{u_i \in M_i} E\Bigg[\int_t^T \big(q(x_i(\tau) - \Gamma\hat{\bar{x}}_j^N(\tau))^2 + ru_i^2(\tau)\big)d\tau + h(x_i(T) - \Gamma\bar{x}^N(T))^2 | \bar{x}^N(t_j)\Bigg] \tag{8}$$

Lemma 1. *The control policy that achieves the minimum cost to go $\widetilde{V}_{j,i}(t, X_i)$ given $\bar{x}^N(t_j)$ is identical to the one that achieves the minimum cost to go $V_{j,i}(t, X_i)$.*

Proof: By centering $\bar{x}^N(\tau)$ around $\hat{\bar{x}}_j^N(\tau)$, the optimal cost to go in (4) can be expressed as follows:

$$V_{j,i}(t, X_i) = \inf_{u_i \in M_i} E\Big[\int_t^T \big(q(x_i(\tau) - \Gamma\bar{x}^N(\tau) + \Gamma\hat{\bar{x}}_j^N(t) - \Gamma\hat{\bar{x}}_j^N(t))^2 + ru_i^2(\tau)\big)d\tau + h\big(x_i(T) - \Gamma\bar{x}^N(T)\big)^2 | \bar{x}^N(t_j)\Big] =$$

$$\inf_{u_i \in M_i} E\Big[\int_t^T \big(q(x_i(\tau) - \Gamma\hat{\bar{x}}_j^N(\tau))^2 + q\Gamma^2(\hat{\bar{x}}_j^N(\tau) - \bar{x}^N(\tau))^2 + ru_i^2(\tau)\big)d\tau + h\big(x_i(T) - \Gamma\bar{x}^N(T)\big)^2 | \bar{x}^N(t_j)\Big] \tag{9}$$

In (9), we have used the orthogonality of the prediction error, $err_j(\tau)$, and $\bar{x}^N(\tau)$. Furthermore, having neglected the influence of $x_i(\tau)$ on $\bar{x}^N(\tau)$ and thus $\hat{\bar{x}}_j^N(\tau)$, $E[x_i(\tau)err_j(\tau)|x^N(t_j)] = 0$. Note that normally the predictor $\hat{\bar{x}}_j^N(\tau)$ would have factored in all information available to agent i at time τ including $x_i(\tau)$ and orthogonality would hold anyway (See Ch. 3 of [15]). Now the optimization in (9) requires some interpretation. Indeed, if one considers (as in MFG types of arguments) that agents other than i have frozen their control policy, then, in accordance with Assumption 1, they would be responsible for producing the trajectory of the estimator $\hat{\bar{x}}_j^N(\tau)$, while the estimation error

variance, $E[err_j^2(\tau)|x^N(t_j)]$, would be only a function of time, independent of the specific control exerted by agent i. In other words, the trajectory of $\bar{x}^N(t)$ produced by the population and its predictor $\hat{\bar{x}}_j^N(t)$ based on the the latest empirical mean observation, are independent from agent i's control policy. In that respect, the best response policy, u_i^*, for agent i would be the same if one were to use the modified optimal cost to go function $\widetilde{V}_{j,i}(t, X_i)$. ■

Given the hybrid nature of observations, the construction of the best response policy for agent i requires two steps:

1. Discrete Component of the DP Equation: We shall write DP equation for $\widetilde{V}_{j,i}$ only at sampling times t_j for $t \in [t_j, t_{j+1}]$, with appropriate boundary conditions.

$$\widetilde{V}_{j,i}(t, X_i) = \inf_{u_i \in M_i} E\Bigg[\int_t^{t_{j+1}} \Big(q\left(x_i(\tau) - \Gamma \hat{\bar{x}}_j^N(\tau)\right)^2 + ru_i^2(\tau)\Big) d\tau + \widetilde{V}_{j+1,i}\left(t_{j+1}, X_i\right) \left| \bar{x}^N(t_j) \right. \Bigg] \tag{10}$$

Boundary conditions at $t = T$ for $\widetilde{V}_{n,i}$, and at t_{j+1} for $\widetilde{V}_{j,i}, j = 0, .., n-1$ can be written as follows:

$$\widetilde{V}_{n,i}(T, X_i) = h\left(x_i(T) - \Gamma \bar{x}^N(T)\right)^2 \tag{11}$$

$$\widetilde{V}_{j,i}\left(t_{j+1}, X_i\right) = E\left[\widetilde{V}_{j+1,i}\left(t_{j+1}, X_i\right) \middle| \bar{x}^N(t_j) \right] \tag{12}$$

2. Interval-Wise Continuous Component Analysis: We shall write the Hamilton-Jacobi-Bellman (HJB) equation between sampling intervals as a tracking problem since $\hat{\bar{x}}_j^N(t)$ is treated as a deterministic, albeit unknown trajectory. As written in (12) the values of $\widetilde{V}_{j,i}$ at the right-hand side of the intervals act as boundary conditions for the continuous interval-wise solution. By solving these equations, we derive the structure of the predictor-dependent best response policy for agent i. This analysis eventually leads to the dynamic equation for the predictor and the expression of the best response policy in terms of x_i and the empirical mean observation.

B. Interval-Wise Application of Dynamic Programming Equations

1) Finding Control Policy for $[t_{n-1}, T)$: The solution of the DP equation will proceed backwards, starting from the time interval $[t_{n-1}, T)$. A deterministic tracking trajectory $\hat{\bar{x}}_j^N(t)$, $t \in [t_{n-1}, T)$ is hypothesized, and by Lemma 1, to compute the best response policy, one needs to solve problem (10). At this point, one writes the following HJB equation for $\widetilde{V}_{n-1,i}(t, X_i)$ to find the optimal policy for $t \in [t_{n-1}, T)$.

$$0 = \frac{\partial \widetilde{V}_{n-1,i}}{\partial t} + \min_{u_i} \Bigg[\frac{\partial \widetilde{V}_{n-1,i}}{\partial x_i}\left(ax_i + bu_i\right) + \left[q\left(x_i - \Gamma \hat{\bar{x}}_{n-1}^N\right)^2 + ru_i^2 \right] + \frac{1}{2}\sigma^2 \frac{\partial^2 \widetilde{V}_{n-1,i}}{\partial x_i^2} \Bigg] \tag{13}$$

We use (11) to develop the boundary condition at time T for the HJB equation on $[t_{n-1}, T)$.

$$\begin{aligned} \widetilde{V}_{n-1,i}\left(T, X_i\right) &= E\left[\widetilde{V}_{n,i}\left(T, X_i\right) \middle| \bar{x}^N\left(t_{n-1}\right) \right] \\ &= E\left[h\left(x_i(T) - \Gamma \bar{x}^N(T)\right)^2 \middle| \bar{x}^N\left(t_{n-1}\right) \right] \\ &= E\left[h\left(x_i(T) - \Gamma(\bar{x}^N(T) + \hat{\bar{x}}_{n-1}^N(T) - \hat{\bar{x}}_{n-1}^N(T))\right)^2 \right] \\ &= h\left(x_i(T) - \Gamma \hat{\bar{x}}_{n-1}^N(T)\right)^2 + h\Gamma^2 E\left[err_{n-1}^2(T)\right] \end{aligned} \tag{14}$$

Variables p_{n-1}, s_{n-1}, and r_{n-1} are introduced to represent the solution of the HJB equation for $\widetilde{V}_{n-1,i}(t, X_i)$, where we assume the following quadratic form for it.

$$\widetilde{V}_{n-1,i}(t, X_i) = p_{n-1}(t)x_i^2(t) + 2s_{n-1}(t)x_i(t) + r_{n-1}(t) \tag{15}$$

The minimizer value of $\widetilde{V}_{n-1,i}(t, X_i)$ is u_i^* which can be found as follows:

$$\begin{aligned} u_i^*(t) &= -\frac{1}{2}\frac{b}{r}\left(\frac{\partial \widetilde{V}_{n-1,i}(t, X_i)}{\partial x_i} \right) = \\ &\quad -\frac{b}{r}(p_{n-1}(t)x_i(t) + s_{n-1}(t)) \end{aligned} \tag{16}$$

Substitution in (13) and identification of resulting polynomial coefficients yield a set of differential equations and boundary conditions for p_{n-1}, s_{n-1} and r_{n-1}.

$$\frac{dp_{n-1}}{dt} = -2p_{n-1}a + \frac{b^2}{r}p_{n-1}^2 - q, \qquad p_{n-1}(T) = h \tag{17}$$

$$\begin{aligned} \frac{ds_{n-1}}{dt} &= -\left(a - \frac{b^2}{r}p_{n-1}\right)s_{n-1} + q\Gamma \hat{\bar{x}}_{n-1}^N(t) \\ s_{n-1}(T) &= -h\Gamma \hat{\bar{x}}_{n-1}^N(T) \end{aligned} \tag{18}$$

$$\begin{aligned} \frac{dr_{n-1}}{dt} &= \frac{b^2}{r}s_{n-1}^2 - q\left(\Gamma \hat{\bar{x}}_{n-1}^N(t)\right)^2 - \sigma^2 p_{n-1} \\ r_{n-1}(T) &= h\Gamma^2\left(\hat{\bar{x}}_{n-1}^N(T)\right)^2 + h\Gamma^2 E\left[err_{n-1}^2(T)\right] \end{aligned} \tag{19}$$

In the following, we state the NE of the game for $[t_{n-1}, T)$, however, the proof is similar for $[t_j, T)$.

Proposition 1. *Suppose Assumption 1 holds and a unique solution exists for the following Riccati differential equation:*

$$\begin{aligned} \frac{d\alpha_{n-1}(t)}{dt} &= -2\left(a - \frac{b^2}{r}p_{n-1}(t)\right)\alpha_{n-1}(t) + \\ &\quad \frac{b^2}{r}\alpha_{n-1}^2(t) + q\Gamma, \qquad \alpha_{n-1}(T) = -h\Gamma \end{aligned} \tag{20}$$

Then for $t \in [t_{n-1}, T)$, $u_i^*(t) = -\frac{b}{r}(p_{n-1}(t)x_i(t) + \alpha_{n-1}(t)\hat{\bar{x}}^N_{n-1}(t))$, $i = 1, \ldots, N$ *is a set of Markov Nash equilibrium strategies of the game, where:*

$$\frac{d\hat{\bar{x}}^N_{n-1}}{dt} = \left(a - \frac{b^2}{r}p_{n-1}(t) - \frac{b^2}{r}\alpha_{n-1}(t)\right)\hat{\bar{x}}^N_{n-1} \quad (21)$$

Proof: The policy derived from DP is the NE of the game as stipulated in Definition 1. Under the assumptions of the proposition, we develop fixed-point calculations that characterize the Markov NE strategies on $[t_{n-1}, T)$. Indeed, the trajectory $\hat{\bar{x}}^N_{n-1}(t)$ must be a predictor of $\bar{x}^N(t)$ based on the solution of the problem of optimally tracking that predictor. As a result, $\hat{\bar{x}}^N_{n-1}(t)$ must be the solution of a fixed-point problem. To help compute that fixed-point, we assume, following [16], the following form for $s_{n-1}(t)$:

$$s_{n-1}(t) = \alpha_{n-1}(t)\hat{\bar{x}}^N_{n-1}(t) \quad (22)$$

If such a structure holds, it will allow a decoupling of the forward and backward propagating parts of the complete solution. Equation (22) yields after taking time derivatives:

$$\frac{ds_{n-1}(t)}{dt} = \frac{d\alpha_{n-1}(t)}{dt}\hat{\bar{x}}^N_{n-1}(t) + \alpha_{n-1}(t)\frac{d\hat{\bar{x}}^N_{n-1}}{dt} \quad (23)$$

Recalling the definition of $\hat{\bar{x}}^N_j(t)$ in (6), we substitute the closed loop control (16) in (1) for $i = 1, \ldots, N$, after recognizing that best response strategies must be identical for all agents, due to their assumed homogeneity. Taking expectations of the resulting empirical mean of the $x_i(t)$'s under closed loop dynamics, and using (22), one obtains the forward propagating dynamics of the fixed-point predictor trajectory $\hat{\bar{x}}^N_{n-1}(t)$ as in (21). One then uses (17), (21) and (23) to obtain the Riccati equation and boundary condition in (20) for $\alpha_{n-1}(t)$. ■

Remark 2. Equation (21) leads to the following solution for the predictor:

$$\hat{\bar{x}}^N_{n-1}(t) = \varphi_{\bar{x}}(t, t_{n-1})\bar{x}^N(t_{n-1}) \quad t \in [t_{n-1}, T) \quad (24)$$

$\varphi_{\bar{x}}(t, t_{n-1})$ denotes the state transition function for $\hat{\bar{x}}^N_{n-1}$.

Using (19), (22), and (24) to determine $r_{n-1}(t)$, one can write the solution of (19) as a function of $\bar{x}^N(t_{n-1})^2$ as follows.

$$r_{n-1}(t_{n-1}) = \psi_{n-1}(t_{n-1})\bar{x}^N(t_{n-1})^2 + \gamma_{n-1}(t_{n-1}) \quad (25)$$

where $\psi_{n-1}(t_{n-1})$ and $\gamma_{n-1}(t_{n-1})$ are found as:

$$\psi_{n-1}(t_{n-1}) = -\int_{t_{n-1}}^{T} \varphi_{\bar{x}}(\tau, t_{n-1})^2 \left(\frac{b^2}{r}\alpha^2_{n-1}(\tau) - q\Gamma^2\right) d\tau + h\left(\Gamma\varphi_{\bar{x}}(T, t_{n-1})\right)^2 \quad (26)$$

$$\gamma_{n-1}(t_{n-1}) = \sigma^2 \int_{t_{n-1}}^{T} p_{n-1}(\tau)d\tau + h\Gamma^2 E[err^2_{n-1}(T)]$$

2) Finding Best Response Policy for $[t_j, t_{j+1})$*:* We now move to determining agent best responses for the interval $[t_j, t_{j+1})$. A key distinction relative to the analysis on $[t_{n-1}, T)$ lies in the fact that we anticipate agents receiving new information about mean agent state at t_{j+1} which will impact subsequent policies. For $t \in [t_j, t_{j+1})$, we again test a quadratic ansatz, assuming the quadratic form of $\widetilde{V}_{j+1,i}(t, X_i)$ has already been validated:

$$\widetilde{V}_{j,i}(t, X_i) = p_j(t)x_i(t)^2 + 2s_j(t)x_i(t) + r_j(t) \quad (27)$$

The procedure for finding optimal control policy over $[t_j, T)$ mirrors the steps taken for $[t_{n-1}, T)$, and hence, we do not repeat the detailed process here. The key difference lies in determining the boundary condition for $\widetilde{V}_{j,i}$ which is obtained from (12) and computed as follows:

$$\widetilde{V}_{j,i}(t_{j+1}, X_i) = E[\widetilde{V}_{j+1,i}(t_{j+1}, X_i)\big|\bar{x}^N(t_j)] = p_{j+1}(t_{j+1})x_i^2(t_{j+1}) + 2\alpha_{j+1}(t_{j+1})x_i(t_{j+1})\hat{\bar{x}}^N_j(t_{j+1}) + \psi_{j+1}(t_{j+1})E[\bar{x}^N(t_{j+1})^2\big|\bar{x}^N(t_j)] + E[\gamma_{j+1}(t_{j+1})] \quad (28)$$

In (28), the computation of $E[\bar{x}^N(t_{j+1})^2|\bar{x}^N(t_j)]$ involves adding and subtracting $\hat{\bar{x}}^N_j(t_{j+1})$ as in (14). Furthermore, the definitions of $\psi_{j+1}(t_{j+1})$ and $\gamma_{j+1}(t_{j+1})$ are analogous to those of $\psi_{n-1}(t_{n-1})$ and $\gamma_{n-1}(t_{n-1})$ above.

$$E[\bar{x}^N(t_{j+1})^2\big|\bar{x}^N(t_j)] = E[err^2_j(t_{j+1})] + \hat{\bar{x}}^N_j(t_{j+1})^2 \quad (29)$$

Solving the HJB equation yields differential equations for p_j, s_j, and r_j, analogous to (17), (18), and (19), respectively. However, the boundary conditions for these functions differ and are derived directly from (28) as follows:

$$p_j(t_{j+1}) = p_{j+1}(t_{j+1}), \qquad s_j(t_{j+1}) = s_{j+1}(t_{j+1}) \quad (30)$$

Similar to (22), s_j can be expressed in terms of α_j.

$$\alpha_j(t_{j+1})\hat{\bar{x}}^N_j(t_{j+1}) = \alpha_{j+1}(t_{j+1})E[\bar{x}^N(t_{j+1})|\bar{x}^N(t_j)] \quad (31)$$

This leads to the boundary condition for α_j:

$$\alpha_j(t_{j+1}) = \alpha_{j+1}(t_{j+1}) \quad (32)$$

Remark 3. The boundary conditions in (30) and (32), along with quadratic forms of $\widetilde{V}_{n-1,i}$ and $\widetilde{V}_{j,i}$ in (15) and (27) imply that the differential equations for p_j and s_j, mirror those of p_{n-1} and s_{n-1}. Consequently, p_j and α_j can be treated as part of a continuous trajectory over $j = 0, 1, 2, \ldots, n-1$ with the same boundary conditions governing each segment:

$$\frac{dp(t)}{dt} = -2pa + \frac{b^2}{r}p^2 - q, \qquad p(T) = h \quad (33)$$

$$\frac{d\alpha(t)}{dt} = -2(a - \frac{b^2}{r}p(t))\alpha(t) + \frac{b^2}{r}\alpha^2(t) + q\Gamma, \quad \alpha(T) = -h\Gamma \quad (34)$$

Using (33), (34) and (28), we derive the differential equation and boundary condition for r_j:

$$\frac{dr_j}{dt} = \frac{b^2}{r}(\alpha\hat{\bar{x}}_j^N)^2 - q(\Gamma\hat{\bar{x}}_j^N)^2 - \sigma^2 p$$
$$r_j(t_{j+1}) = \psi_{j+1}(t_{j+1})(E[err_j^2(t_{j+1})] + \hat{\bar{x}}_j^N(t_{j+1})^2) + \gamma_{j+1}(t_{j+1}) \quad (35)$$

To solve r_j, we express it as a linear function of of $\bar{x}^N(t_j)^2$ using ψ_j and γ_j similar to (25):

$$\psi_j(t) = -\int_t^{t_{j+1}} (\frac{b^2}{r}\alpha(\tau)^2 - q\Gamma^2)\varphi_{\bar{x}}(\tau, t_j)^2 d\tau + \psi_{j+1}(t_{j+1})\varphi_{\bar{x}}(t_{j+1}, t)^2 \quad (36)$$

$$\gamma_j(t) = \gamma_{j+1}(t_{j+1}) + \int_t^{t_{j+1}} \sigma^2 p(\tau) d\tau + \psi_{j+1}(t_{j+1})E[err_j^2(t_{j+1})] \quad (37)$$

Given (26) and (36), we can also derive the differential equation for $\psi(t)$. For clarity, we omit the index j from ψ from this point forward:

$$\frac{d\psi}{dt} = -2\Big(a - \frac{b^2}{r}(p(t) + \alpha(t))\Big)\psi(t) + \left(\frac{b^2}{r}\alpha(t)^2 - q\Gamma^2\right)$$
$$\psi(T) = h\Gamma^2 \quad (38)$$

The above analysis and remarks lead us to the main result of the paper which is an interval-wise generalization of Proposition 1 and characterizes Markov Nash strategies.

Theorem 1. *Suppose Assumption 1 holds and a unique solution $\alpha(t)$ exists for* (34) *where $p(t)$ is the solution of* (33)*, then for $t \in [t_j, t_{j+1})$, $j = 0, \ldots, n-1$ and $i = 1, \ldots, N$, $u_i^*(t) = -\frac{b}{r}(p(t)x_i(t) + \alpha(t)\hat{\bar{x}}_j^N(t))$, is a set of Markov Nash equilibrium strategies of the game, where:*

$$\frac{d\hat{\bar{x}}_j^N}{dt} = \left(a - \frac{b^2}{r}p(t) - \frac{b^2}{r}\alpha(t)\right)\hat{\bar{x}}_j^N, \qquad \hat{\bar{x}}_j^N(t_j) = \bar{x}^N(t_j)$$

C. Error Calculation

In this section, we compute the error, err_j, based on observations of $\bar{x}^N(t_j)$ at time t_j. To do so, we first express $\bar{x}^N(t)$ over the interval $[t_j, t_{j+1})$ by solving the differential equation in (1). Using the results from [17], $\bar{x}^N(t)$ is the average of $x_i(t)$ under the closed-loop best response control law. The error is then derived from this expression. Here, $\varphi(t, t_0) = \exp\left(\int_{t_0}^t \left(a - \frac{b^2}{r}p(\tau)\right)d\tau\right)$ represents the state transition function for $x_i(t)$.

$$\bar{x}^N(t) = \varphi_{\bar{x}}(t, t_j)\bar{x}^N(t_j) + \sigma\int_{t_j}^t \varphi(t, s)d\bar{w}^N(s) = \hat{\bar{x}}_j^N(t) + \sigma\int_{t_j}^t \varphi(t, s)d\bar{w}^N(s), \quad \bar{w}^N(s) = \frac{1}{N}\sum_{i=1}^n dw_i(s)$$

$$err_j(t) = \sigma\int_{t_j}^t \varphi(t, s)d\bar{w}^N(s) \quad t \in [t_j, t_{j+1}) \quad (39)$$

Remark 4. Equation (39) further is consistent with Lemma 1 since it indicates that $err_j(t)$ is only a function of noises.

IV. Performance Evaluation

We now aim at calculating $V_{j,i}(t, X_i)$ based on the knowledge of $\widetilde{V}_{j,i}(t, X_i)$ that we have developed in the earlier sections. A DP equation for $V_{j,i}(t, X_i)$, analogous to that for $\widetilde{V}_{j,i}(t, X_i)$ in (10), is first written. We then compute the discrepancy between these two value functions using the knowledge from Lemma 1 that the associated best response policies are identical. Thus, we have:

$$V_{j,i}(t, X_i) = \inf_{u_i \in M_i} E\Bigg[\int_t^{t_{j+1}} \big(q(x_i(\tau) - \Gamma\bar{x}_j^N(\tau))^2 + ru_i^2(\tau)\big)d\tau + V_{j+1,i}(t_{j+1}, X_i)\big|\bar{x}_j^N(t_j)\Bigg] \quad (40)$$

Subtracting (40) from (10) yields interval wise:

$$\Delta V_j(t) := V_{j,i}(t, X_i) - \widetilde{V}_{j,i}(t, X_i) \quad (41)$$
$$= E\Bigg[\int_t^{t_{j+1}} \big(q(x_i(\tau) - \Gamma\bar{x}_j^N(\tau))^2 - q(x_i(\tau) - \Gamma\hat{\bar{x}}_j^N(\tau))^2\big)d\tau + \Delta V_{j+1}(t_{j+1})|\bar{x}_j^N(t_j)\Bigg] =$$
$$q\Gamma^2 E\left[\int_t^{t_{j+1}} err_j^2(\tau)d\tau\right] + E[\Delta V_{j+1}(t_{j+1})|\bar{x}_j^N(t_j)] \quad (42)$$

Note that $\Delta V_j(t)$ is 0 at T, and is *independent* of agent i. For calculation of $V_{0,i}$ from t_0 to T, we sum all ΔV_0 for $j = 0, 1, \ldots, n-1$:

$$V_{0,i}(t_0, X_i) = \widetilde{V}_{0,i}(t_0, X_i) + \Delta V_0(t_0) = \widetilde{V}_{0,i}(t_0, X_i) + q\Gamma^2\sum_{j=0}^{n-1} E\left[\int_{t_j}^{t_{j+1}} err_j^2(\tau)d\tau\right] = p(t_0)x_i^2(t_0) + 2\alpha(t_0)\bar{x}^N(t_0)x_i(t_0) + \psi(t_0)\bar{x}^N(t_0)^2 + \gamma_0(t_0) + q\Gamma^2\frac{\sigma^2}{N}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{t_j}^{\tau}\varphi^2(\tau, s)dsd\tau \quad (43)$$

A. Comparing Costs with Full Observation Game

In this section, we quantify the performance loss due to partial observability, referred to as regret. Since agents have limited discrete observations, their costs differ from the fully observable case. Therefore, we also consider the scenario where agents have continuous observations of the empirical

mean, leading to the control policy $u_i^{\text{Full}}(t) = -\frac{b}{r}(p(t)x_i(t) + \alpha(t)\bar{x}^N(t))$ derived through similar fixed-point calculations as presented in this paper. In the following, we calculate the cost for the continuous observation scenario:

$$V^{\text{Full}}(t_0) = V_{0,i}(t_0, X_i) + \sum_{j=0}^{n-1} E\Big[\int_{t_j}^{t_{j+1}} r(u_i^{\text{Full}}(\tau)^2 - u_i^2(\tau))d\tau \big| \bar{x}^N(t_0)\Big] = V_{0,i}(t_0, X_i) + \frac{b^2}{r}\sum_{j=0}^{n-1} E\Big[\int_{t_j}^{t_{j+1}} err_j^2(\tau)d\tau\Big] \quad (44)$$

The formula for regret is defined in the following:

$$Regret = E\left[V^{\text{Full}}(t_0) - \widetilde{V}_{0,i}(t_0, X_i)\big|\bar{x}^N(t_0)\right] = \left(q\Gamma^2 + \frac{b^2}{r}\right)\frac{\sigma^2}{N}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\int_{t_j}^{\tau}\varphi^2(\tau, s)dsd\tau + \sum_{j=0}^{n-1}\psi(t_j)E\left[err_j^2(t_j)\right] \quad (45)$$

Theorem 2. *The $Regret$ exhibits linear growth rate.*

Proof: Our goal is to demonstrate that $Regret$ exhibits linear growth. In analyzing the long-term behavior of $Regret$, we focus on the steady state solution of Riccati differential $p(t)$, and we know this steady state solution is $p_\infty := \frac{r}{b^2}\left(a \pm \sqrt{a^2 + \frac{b^2}{r}q}\right)$ [18]. To ensure a physically meaningful solution, we take the positive root, so $c := a - \frac{b^2}{r}p_\infty = -\sqrt{a^2 + \frac{b^2}{r}q} < 0$. This leads to the following expression in steady state:

$$\int_{t_j}^{t_{j+1}}\int_{t_j}^{\tau}\varphi^2(\tau, s)\,ds\,d\tau \approx \int_{t_j}^{t_{j+1}}\int_{t_j}^{\tau}\exp(2c(\tau - s))\,ds\,d\tau = -\frac{1}{2c}\Delta t + \frac{1}{4c^2}\exp(2c\Delta t) - \frac{1}{4c^2} \quad (46)$$

$$\lim_{T\to\infty}\frac{1}{T}Regret = -\frac{1}{2c}\left(q\Gamma^2 + \frac{b^2}{r}\right)\frac{\sigma^2}{N} \quad (47)$$

This shows that the first term of the $Regret$ grows linearly, while the second term decays exponentially [1]. Therefore, the overall $Regret$ exhibits a linear growth rate. ■

V. CONCLUSION

This paper introduces a multi-agent aggregative game characterized by continuous agent dynamics and discrete empirical mean observations over time. Leveraging dynamic programming principles, we identify the Markov Nash strategies of this game. In the subsequent section, we outline the cost function formula for the scenario of complete observability. By quantifying the disparity between the costs, termed as regret, we demonstrate that this regret exhibits a linear growth rate.

REFERENCES

[1] Minyi Huang and Mengjie Zhou. Linear quadratic mean field games: Asymptotic solvability and relation to the fixed point approach. *IEEE Transactions on Automatic Control*, 65(4):1397–1412, 2019.

[2] Daniel Lacker and Agathe Soret. A case study on stochastic games on large graphs in mean field and sparse regimes. *Mathematics of Operations Research*, 47(2):1530–1565, 2022.

[3] Minyi Huang, Peter E Caines, and Roland P Malhamé. 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, 2007.

[4] René Carmona, François Delarue, et al. *Probabilistic theory of mean field games with applications I-II*. Springer, 2018.

[5] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

[6] Minyi Huang, Peter E Caines, Roland P Malhamé, et al. Distributed multi-agent decision-making with partial observations: asymptotic nash equilibria. In *Proc. the 17th Internat. Symposium on Math. Theory on Networks and Systems (MTNS'06), Kyoto, Japan*, pages 2725–2730, 2006.

[7] Nevroz Şen and Peter E Caines. Mean field games with partially observed major player and stochastic mean field. In *53rd IEEE Conference on Decision and Control*, pages 2709–2715. IEEE, 2014.

[8] Nevroz Şen and Peter E Caines. ε-nash equilibria for a partially observed mean field game with major player. In *2015 American Control Conference (ACC)*, pages 4791–4797. IEEE, 2015.

[9] Nevroz Sen and Peter E Caines. Mean field game theory with a partially observed major agent. *SIAM Journal on Control and Optimization*, 54(6):3174–3224, 2016.

[10] Peter E Caines and Arman C Kizilkale. ε-nash equilibria for partially observed lqg mean field games with a major player. *IEEE Transactions on Automatic Control*, 62(7):3225–3234, 2016.

[11] Nevroz Sen and Peter E Caines. Mean field games with partial observation. *SIAM Journal on Control and Optimization*, 57(3):2064–2091, 2019.

[12] Dena Firoozi and Peter E Caines. ε-nash equilibria for partially observed lqg mean field games with major agent: Partial observations by all agents. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 4430–4437. IEEE, 2015.

[13] Dena Firoozi and Peter E Caines. ε-nash equilibria for major–minor lqg mean field games with partial observations of all agents. *IEEE Transactions on Automatic Control*, 66(6):2778–2786, 2020.

[14] Naci Saldi, Tamer Başar, and Maxim Raginsky. Partially observed discrete-time risk-sensitive mean field games. *Dynamic Games and Applications*, 13(3):929–960, 2023.

[15] David G Luenberger. *Optimization by vector space methods*. John Wiley & Sons, 1997.

[16] Roland P Malhamé and Christy Graves. Mean field games: A paradigm for individual-mass interactions. In *Proceedings of Symposia in Applied Mathematics*, volume 78, pages 3–32, 2020.

[17] Farid Rajabali and Roland Malhamé. Can mean field game equilibria amongst exchangeable agents survive under partial observability of their competitors' states? In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 8188–8193. IEEE, 2023.

[18] David H Jacobson and Jason L Speyer. Necessary and sufficient conditions for optimality for singular control problems: A limit approach. *Journal of Mathematical Analysis and Applications*, 34(2):239–266, 1971.