

Mean Field LQG Control in Leader-Follower Stochastic Multi-Agent Systems: Likelihood Ratio Based Adaptation

Mojtaba Nourian, *Student Member, IEEE*, Peter E. Caines, *Life Fellow, IEEE*, Roland P. Malhamé, *Member, IEEE*, and Minyi Huang, *Member, IEEE*

Abstract—We study large population leader-follower stochastic multi-agent systems where the agents have linear stochastic dynamics and are coupled via their quadratic cost functions. The cost of each leader is based on a trade-off between moving toward a certain reference trajectory which is unknown to the followers and staying near their own centroid. On the other hand, followers react by tracking a convex combination of their own centroid and the centroid of the leaders. We approach this large population dynamic game problem by use of so-called Mean Field (MF) linear-quadratic-Gaussian (LQG) stochastic control theory. In this model, followers are adaptive in the sense that they use a likelihood ratio estimator (on a sample population of the leaders' trajectories) to identify the member of a given finite class of models which is generating the reference trajectory of the leaders. Under appropriate conditions, it is shown that the true reference trajectory model is identified by each follower in finite time with probability one as the leaders' population goes to infinity. Furthermore, we show that the resulting sets of mean field control laws for both leaders and adaptive followers possess an almost sure ε_N -Nash equilibrium property for a system with population N where ε_N goes to zero as N goes to infinity. Numerical experiments are presented illustrating the results.

Index Terms—Adaptive control, leader-follower collective behavior, likelihood ratio based adaptation, mean field (MF) stochastic control, Nash equilibria, stochastic optimal control.

I. INTRODUCTION

MULTI-AGENT control and coordination problems arise in fields such as mobile robotics [1], vehicle formations [2], synchronization of coupled oscillators [3], flocking [4], and micro-economics [5] among others. Consequently, there has recently been a considerable amount of research devoted to the theory of multi-agent systems.

Manuscript received December 13, 2010; revised July 30, 2011; accepted March 15, 2012. Date of publication April 26, 2012; date of current version October 24, 2012. Recommended by Associate Editor H. S. Chang.

M. Nourian and P. E. Caines are with the Centre for Intelligent Machines (CIM) and the Department of Electrical & Computer Engineering, McGill University, Montreal, QC H3A 2A7 Canada. They are also with the Group for Research in Decision Analysis (GERAD), Montreal, QC H3T 2A7, Canada (e-mail: mnourian@cim.mcgill.ca; peterc@cim.mcgill.ca).

R. P. Malhamé is with the Department of Electrical Engineering, École Polytechnique de Montréal, Montreal, QC H3C 3A7, Canada. He is also with the Group for Research in Decision Analysis (GERAD), Montreal, QC H3T 2A7, Canada (e-mail: roland.malhamé@polymtl.ca).

M. Huang is with the School of Mathematics and Statistics, Carleton University, Ottawa, ON K1S 5B6, Canada. (e-mail: mhuang@math.carleton.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TAC.2012.2195797

For complex systems with many agents, it is not feasible for each individual agent to collect detailed state information on all the other agents' states in the way the single agent in a centralized control system does on all the components of its system's state. Therefore an important issue is the development of decentralized solutions so that each individual agent may implement a strategy based on its local information together with statistical information on the population of agents. For large population systems with mean field couplings, the Mean Field (MF) (or Nash Certainty Equivalence (NCE)) framework has been developed as a decentralized methodology in stochastic noncooperative dynamic games in a series of papers by Huang *et al.* (see [6], [7] for the MF (NCE) linear-quadratic-Gaussian (MF LQG) framework, and [8]–[10] for a general formulation of nonlinear McKean-Vlasov type MF (NCE) stochastic control problems). For this class of game problems a closely related approach has been independently developed by Lasry and Lions [11] (see also [12], [13]), while for models of many firm industry dynamics, Weintraub *et al.* proposed the notion of oblivious equilibrium by use of mean field approximations [14]. In [15], the MF LQG framework is extended to systems of agents with long run average costs.

The central idea of the MF control methodology is to specify a certain equilibrium relationship between the individual strategies and the mass effect (i.e., the overall effect of the population on a given agent) as the population size goes to infinity [7]. Specifically, in the equilibrium: (i) the individual strategy of each agent is a best response to the infinite population mass effect in the sense of a so-called ε -Nash equilibrium, and (ii) the set of strategies collectively replicates the mass effect, this being a dynamical game theoretic fixed point property. The defining property of the MF ε -Nash equilibrium (u_1^*, u_2^*, \dots) requires that for any given $\varepsilon > 0$, there exists $N(\varepsilon)$ such that for any population size $N \geq N(\varepsilon)$, when any agent j , $1 \leq j \leq N$, distinct from i employs u_j^* , then agent i can benefit at most ε by unilaterally deviating from his strategy u_i^* , and this holds for all $1 \leq i \leq N$.

The MF feedback strategies display the possibly counterintuitive nature of MF control which is that in the infinite population limit, except for some statistical information on the parameter distribution and the initial mean state distribution of the population of agents, no observations of other agents' states are necessary to achieve Nash equilibrium behavior and this property persists with negligible incremental cost for sufficiently large finite populations.

Decision making and collective behavior often involve some form of leader-follower behavior. This behavior is observed in humans [16] and many other species in nature [17], [18], and is studied in a variety of disciplines such as game theory [19], distributed networks [20], crowd flow dynamics [21] and biology [17], among others. Such behavior in nature is often attributed to the fact that there exist some individuals in the group which have more information than others, for instance the location of resources or migratory routes [17].

In this paper we study large population Leader-Follower (L-F) stochastic multi-agent systems where the agents have linear stochastic dynamics and are coupled via their quadratic cost functions. The cost of each leader is based on a trade-off between moving toward a certain reference trajectory which is unknown to the followers and staying near their own centroid. On the other hand, followers react by tracking a convex combination of their own centroid and the centroid of the leaders. Here, as in most practical leader-follower modeling of multi-agent systems, the leaders ignore the followers, but the followers' behaviors are influenced by the leaders. The model in this paper is a generalization of that in [15] to the case of collective dynamics which include leaders, followers and an unknown (to the followers) reference trajectory for the leaders.

We approach the large population L-F model of this paper by use of MF LQG stochastic control theory. In this framework the computation of the followers' control laws requires knowledge of the complete reference trajectory of the leaders which is in general not known to the followers, hence a likelihood ratio based adaptation scheme is proposed. The main contributions of the paper are as follows: (i) A likelihood ratio based adaptation algorithm (on a sample population of the leaders' trajectories) is employed by the adaptive followers to identify the member of a given finite class of models which is generating the reference trajectory of the leaders. Under appropriate conditions, it is shown that the true reference trajectory model is identified in finite time with probability one by each follower as the leaders' population goes to infinity. (ii) A demonstration that the use of the resulting MF control laws yields a set of leaders and adaptive followers' control laws possessing an almost sure (*a.s.*) ε_N -Nash equilibrium property, where ε_N goes to zero as the population size N goes to infinity.

The implementation of the overall MF control laws for the leaders and followers has the following form: (i) Each leader enacts an MF control law which consists of the feedback of its own local stochastic state and the precomputed leaders' deterministic mass effect. (ii) Each follower enacts an adaptive MF control law which consists of the feedback of its own local stochastic state and the estimation based mass effects of the leaders and followers.

Previous presentations of the methodology considered in this paper include [22], [23] which contained no proof details. In [22] we first developed a non-adaptive but general model with weighted couplings in the leaders and followers' cost functions (which depended on the locality parameters of the agents). [22] also presents the main adaptation result of the uniform cost coupling model in the case that the followers "only" track the centroid of the leaders. Subsequently, in [23] the optimality property of the (tracking like) adaptive followers' MF control laws is

studied. In this paper we present a complete analysis of a more general (and realistic) scenario where the followers are tracking a convex combination of their own centroid and the centroid of the leaders. Hence, we have an ε -Nash equilibrium property for the adaptive followers' MF control laws.

In the standard consensus literature the agents have little a priori information but communicate over possibly time varying graphs, then under connectivity assumptions (e.g., the union of the interaction graphs for the system is connected frequently enough as the system evolves) the agents reach consensus (see [24], among many other papers). By contrast, the leader-follower agents in this paper possess a priori data on the overall system; the leaders observe no one and the followers have limited observations on the leaders and a priori data on the possible trajectory scenarios of the leaders. This permits the computation by each follower agent of different tracking scenarios amongst which it chooses one at each instant, depending upon the observations received. Therefore, the agents in the model considered in this paper do not require communication except the limited observations on the leaders by the followers.

It is to be noted that the formulation in this paper does not include any collision avoidance or formation control between the agents beyond the optimal tracking property since the states do not necessarily correspond to positions in space. We further note that if the states are given a spatial interpretation the inherently stochastic volatility of the dynamics prevents any state from converging onto another state.

The organization of the paper is as follows. Section II is dedicated to the problem formulation, terminology and some applications of the model in multi-vehicle coordination control and economics (finance). The MF LQG equation systems of the L-F problem are derived and analyzed in Section III. In Section IV we present the estimation procedure for the followers. The stability analysis of the MF control laws and the adaptive MF algorithm for the followers are presented in Section V. The optimality properties of the MF control laws for both the leaders and the adaptive followers are given in Sections VI and VII presents sample numerical simulations of the model. Concluding remarks are stated in Section VIII.

II. PROBLEM FORMULATION, TERMINOLOGY AND SOME APPLICATIONS

The following notation will be used throughout the paper. We use the integer valued subscript as the label for a certain agent of the population and superscripts L and F for a leader and follower agent, respectively. In addition, an overline denotes the expected value of a random variable, i.e., $\bar{z}(t) := E z(t)$. $\|\cdot\|$ denotes the 2-norm of vectors and $\|\cdot\|_\infty$ denotes the infinity or sup norm. $\|z\|_Q := (z^T Q z)^{1/2}$ for any appropriate dimension vector z and matrix $Q \geq 0$. A^T denotes the transpose of a vector or matrix A and $\text{tr}(A)$ denotes the trace of a square matrix A . Let C_n be the family of all n -dimensional continuous functions on \mathbb{R}_+ and $C_n^b := \{f \in C_n : \|f\|_\infty := \sup_{t \geq 0} \|f(t)\| < \infty\}$. Note that C_n^b is a Banach space under the norm $\|\cdot\|_\infty$.

Let L denote the countably infinite set of leaders, L^{N_L} denotes the subset $\{L_1, \dots, L_{N_L}\} \subset L$ of cardinality N_L , and similarly for the set of countably infinite followers F and the

subset $F^{N_F} := \{F_1, \dots, F_{N_F}\} \subset F$ of cardinality N_F . We assume that $L \cap F = \emptyset$.

A. Leaders Stochastic LQG Dynamic Game Model

The dynamics for the N_L leaders are given by

$$dz_i^L = (A_i z_i^L + B_i u_i^L) dt + C_i dw_i^L, \quad t \geq 0, \quad 1 \leq i \leq N_L \quad (1)$$

where $z_i^L \in \mathbb{R}^n$ is the state, $u_i^L \in \mathbb{R}^m$ is the control input, and $\{w_i^L : 1 \leq i \leq N_L\}$ denotes a set of independent (i.e., mutually independent) p -dimensional standard Wiener processes. The matrices A_i , B_i and C_i have compatible dimensions.

Let $\theta_i := [A_i, B_i, C_i]$ be defined as the dynamical parameter associated with leader i , $1 \leq i \leq N_L$, where we assume that θ_i , for all $1 \leq i \leq N_L$, are in the compact set Θ_L . The initial states $\{z_i^L(0) : 1 \leq i \leq N_L\}$ are assumed to be independent and also independent of $\{w_i^L : 1 \leq i \leq N_L\}$. In addition we assume that $\sup_{1 \leq i \leq N_L} E \|z_i^L(0)\|^2 < \infty$.

The admissible control set for leader i , $1 \leq i \leq N_L$, is given by $\mathcal{U}_i^L := \{u_i^L : u_i^L(t) \text{ is adapted to sigma-field } \sigma(z_j^L(s), s \leq t, 1 \leq j \leq N_L), \|z_i^L(T)\| = o(\sqrt{T}), \int_0^T \|z_i^L(t)\|^2 dt = O(T) \text{ a.s.}\}$. Intuitively, the admissible controls are the controls that do not use any information about future increments of the driving noise processes, which is essentially a causality requirement in stochastic systems.

The cost function of the leaders is based on a trade-off between moving towards a common reference trajectory and keeping cohesion of the flock of leaders by also tracking their centroid. We let

$$\phi^L(z^{L, N_L})(\cdot) := \lambda h(\cdot) + (1 - \lambda) z^{L, N_L}(\cdot) \quad (2)$$

where λ is a scalar in $(0, 1)$, $h \in C_n^b$ is a reference trajectory known to all the leaders, and $z^{L, N_L}(\cdot) := 1/N_L \sum_{i=1}^{N_L} z_i^L(\cdot)$ is the centroid of the leaders. The objective of each individual leader i ($1 \leq i \leq N_L$) is to minimize its Long Run Average (LRA) cost function given by

$$J_i^{L, N_L}(u_i^L; u_{-i}^L) := \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\|z_i^L - \phi^L(z^{L, N_L})\|_Q^2 + \|u_i^L\|_R^2 \right) dt \quad (3)$$

where the matrices Q and R are symmetric positive semi-definite and symmetric positive definite, respectively, with compatible dimensions, and $u_{-i}^L := (u_1^L, \dots, u_{i-1}^L, u_{i+1}^L, \dots, u_{N_L}^L)$. To indicate the dependence of J_i^L on $u_i^L(\cdot)$, $u_{-i}^L(\cdot)$ and the leaders' population size N_L , we write it as $J_i^{L, N_L}(u_i^L; u_{-i}^L)$. Note that the leaders' mean field cost coupling (2) is the same as mean field couplings in the basic models considered in [7], but with time-varying offset term $h(\cdot)$. If $\lambda = 1$ then the leaders become independent such that each leader is interested in optimally tracking $h(\cdot)$.

B. Followers Stochastic LQG Dynamic Game Model

Similarly, the dynamics for the N_F followers are given by

$$dz_i^F = (A_i z_i^F + B_i u_i^F) dt + C_i dw_i^F, \quad t \geq 0, \quad 1 \leq i \leq N_F \quad (4)$$

where $z_i^F \in \mathbb{R}^n$ is the state, $u_i^F \in \mathbb{R}^m$ is the control input, and $\{w_i^F : 1 \leq i \leq N_F\}$ denotes a set of independent p -dimensional standard Wiener processes independent of both $\{w_i^L : 1 \leq i \leq N_L\}$ and $\{z_i^L(0) : 1 \leq i \leq N_L\}$. The matrices A_i , B_i and C_i have compatible dimensions.

Let $\theta_i := [A_i, B_i, C_i]$ be defined as the dynamical parameter associated with follower i , $1 \leq i \leq N_F$, where we assume that θ_i , $1 \leq i \leq N_F$, are in the compact set Θ_F . The initial states $\{z_i^F(0) : 1 \leq i \leq N_F\}$ are assumed to be independent and also independent of $\{w_i^F : 1 \leq i \leq N_F\}$, $\{w_i^L : 1 \leq i \leq N_L\}$, and $\{z_i^L(0) : 1 \leq i \leq N_L\}$. In addition we assume that $\sup_{1 \leq i \leq N_F} E \|z_i^F(0)\|^2 < \infty$.

The admissible control set for follower i , $1 \leq i \leq N_F$, is given by $\mathcal{U}_i^F := \{u_i^F : u_i^F(t) \text{ is adapted to } \sigma(z_j^F(s), z_k^L(s), s \leq t, 1 \leq j \leq N_F, 1 \leq k \leq N_L), \|z_i^F(T)\| = o(\sqrt{T}), \int_0^T \|z_i^F(t)\|^2 dt = O(T) \text{ a.s.}\}$.

The followers react by tracking a convex combination of their own centroid and the centroid of the leaders. We let

$$\phi^F(z^{L, N_L}, z^{F, N_F})(\cdot) := \eta z^{L, N_L}(\cdot) + (1 - \eta) z^{F, N_F}(\cdot) \quad (5)$$

where η is a scalar in $(0, 1)$, $z^{F, N_F}(\cdot) := 1/N_F \sum_{i=1}^{N_F} z_i^F(\cdot)$ is the centroid of the followers, and z^{L, N_L} is the centroid of the leaders defined in (2). The LRA cost function for an individual follower i ($1 \leq i \leq N_F$) is given by

$$J_i^{F, N}(u_i^F; u_{-i}^F, u^L) := \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\|z_i^F - \phi^F(z^{L, N_L}, z^{F, N_F})\|_Q^2 + \|u_i^F\|_R^2 \right) dt \quad (6)$$

where $u_{-i}^F := (u_1^F, \dots, u_{i-1}^F, u_{i+1}^F, \dots, u_{N_F}^F)$, $u^L := (u_1^L, \dots, u_{N_L}^L)$, $N = N_L + N_F$ is the population size of the system, and Q and R are defined in (3). To indicate the dependence of J_i^F on $u_i^F(\cdot)$, $u_{-i}^F(\cdot)$, $u^L(\cdot)$ and the population size of the system N , we write it as $J_i^{F, N}(u_i^F; u_{-i}^F, u^L)$.

We note that in this model: (i) the leaders are coupled to each other through their cost functions and respond to each other and their reference trajectory, and (ii) the followers attempt to track the convex combination of both their own centroid and the centroid of the leaders. These are captured in the two types of MF equation systems in Section III.

C. Followers Observation Processes

We assume that all adaptive followers observe a random fraction of the leaders' trajectories with some added noise through their individual observation processes. More precisely, we assume each adaptive follower i , $1 \leq i \leq N_F < \infty$, observes a non-empty random subset $O_i \subset L$ of size $|O_i| := M_i \leq N_L < \infty$ of the leaders' trajectories through the process $y_i^{M_i}(\cdot)$ which is described in terms of the stochastic differential equations

$$dy_i^{M_i} = \left(\frac{1}{M_i} \sum_{j \in O_i} z_j^L \right) dt + \frac{1}{M_i} \sum_{j=1}^{M_i} D_j^i dv_j^i, \quad t \geq 0 \quad (7)$$

where $y_i^{M_i} \in \mathbb{R}^n$ and $\{v_j^i : 1 \leq i \leq N_F, 1 \leq j \leq M_i\}$ is a set of independent standard Wiener processes independent of

$\{w_i^L : 1 \leq i \leq N_L\}$, $\{w_i^F : 1 \leq i \leq N_F\}$, $\{z_i^L(0) : 1 \leq i \leq N_L\}$ and $\{z_i^F(0) : 1 \leq i \leq N_F\}$. The set of constant matrices $\{D_j^i : 1 \leq i \leq N_F, 1 \leq j \leq M\}$ has compatible dimensions.

The sets $O_i \subset L$, $1 \leq i \leq N_F < \infty$, of cardinality M_i are chosen a priori with $M_i = M(N_L) = \lfloor \sqrt{N_L} \rfloor < N_L < \infty$ (where $\lfloor r \rfloor$ denotes greatest integer less than or equal to r) for $1 \leq i \leq N_F < \infty$, where $N_F \rightarrow \infty$ and $N_L \rightarrow \infty$ as the number of followers and leaders respectively goes to infinity. The sets $O_i \subset L$, $1 \leq i < \infty$, are chosen independently, and independent of all initial states and Wiener processes by uniformly distributed selections on the set of the leaders, L . We underline that the magnitude $\lfloor \sqrt{N_L} \rfloor$ is chosen simply for definition of modeling and exposition; any other integer value function $M(N_L)$ satisfying $M(N_L) \rightarrow \infty$ and $M(N_L)/N_L \rightarrow 0$, as $N_L \rightarrow \infty$, may be used for the theory in the paper.

We also assume that, as the prior information of the followers, the reference trajectory of the leaders $h(\cdot)$ is parameterized by $\delta \in \Delta$ where Δ is a finite set and $h_\delta \in C_n^b$ for any $\delta \in \Delta$. This is only for the followers, and $h(\cdot)$ is fully known by the leaders.

D. Applications

The leader-follower modeling of this paper is motivated by many practical problems in which some agents in a group have more information than the others.

A typical application is in multi-vehicle coordination control (see [25]) where the aim is that the states (e.g., the velocities) of all vehicles approach a reference signal (which could be an exogenous signal or which could evolve according to a dynamic model). In many realistic situations some vehicles (the leaders) have complete access to the reference trajectory, and the other vehicles (followers) do not have access to this trajectory and need to estimate it. We can formulate this multi-agent model as a leader-follower mean field LQG problem considered in this paper where the followers need to identify the member of a given finite class of models which is generating the reference trajectory.

Another application of the model is leader-follower dynamic version of Keynes' beauty contest games in economics (finance). Keynes proposed beauty contest games where a newspaper would print some photographs and people would vote for the prettiest faces. Everyone who picked the most popular face automatically entered a lottery to win a prize. Keynes remarked that the stock market is similar to beauty contest games where each investor would like to guess the other investors' guesses (see Example 1 in [26]). A similar approach to MF stochastic control is considered in [26] to study large population static aggregative games such as Keynes' beauty contest games. Now we formulate a leader-follower LQG dynamic version of Keynes' beauty contest games. Here we consider a large population of players divided into two groups: (i) the leaders as large well-informed players (e.g., institutional investors in the stock market), and (ii) the followers (e.g., retail investors in the stock market). The state of each player is its publicly announced prediction of the prettiest face where $z_i^L(\cdot)$ denotes the state of the i -th leader ($1 \leq i \leq N_L$) and $z_i^F(\cdot)$ denotes the state of the i -th follower ($1 \leq i \leq N_F$). The leaders and followers have linear stochastic dynamics given in (1) and (4) with different classes of parameters Θ_L and Θ_F . The average prediction of the leaders and followers are given by their centroids $z^{L,N_L}(\cdot) := (1/N_L) \sum_{i=1}^{N_L} z_i^L(\cdot)$ and $z^{F,N_F}(\cdot) := (1/N_F) \sum_{i=1}^{N_F} z_i^F(\cdot)$, respectively. Based on the quadratic payoff functions considered in [26], we formulate

cost functions of the agents as follows. The leaders would like to minimize their cost functions (3) based on a trade-off between making guesses close to the exogenous private informative signal h (which is unknown to the followers) and guessing close to their own average prediction $z^{L,N_L}(\cdot)$. On other hand, the followers would like to guess close to some convex combination of their own average prediction $z^{F,N_F}(\cdot)$ and the average prediction of the leaders $z^{L,N_L}(\cdot)$ (see (6)).

There are many other similar applications of the model considered in this paper in flocking [27], formation control [28], dynamic industry models [14], and social opinion models with a very large number of leaders (e.g., important members of a party) and followers [29].

III. MEAN FIELD LINEAR-QUADRATIC-GAUSSIAN (MF LQG) STOCHASTIC CONTROL THEORY

A. Preliminary LQG Optimal Control of a Single Agent

In this section first we consider a single agent with linear stochastic dynamics

$$dz = (Az + Bu)dt + Cdw, \quad t \geq 0 \quad (8)$$

where $z \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the control input, w denotes a p -dimensional standard Wiener process, and $z(0)$ is given. The matrices A , B and C have compatible dimensions. The initial condition $z(0)$ is independent of the process w .

Denote the admissible control set $\mathcal{U} := \{u : u(\cdot) \text{ is adapted to } \sigma(z(0), w(s), s \leq t), \|z(T)\| = o(\sqrt{T}), \int_0^T \|z(t)\|^2 dt = O(T) \text{ a.s.}\}$ (see [15]). For $u(\cdot) \in \mathcal{U}$, let the LRA cost function be given by

$$J(u(\cdot)) := \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\|z - \phi\|_Q^2 + \|u\|_R^2 \right) dt \quad (9)$$

where $\phi(\cdot) \in C_n^b$ is a known function, Q and R are, respectively, symmetric positive semi-definite and symmetric positive definite matrices with compatible dimensions.

Theorem 3.1: (Special case of [15]) For the LQG optimal control problem (8)–(9), assume (i) $[A, B]$ is stabilizable, (ii) $[A, Q^{1/2}]$ is detectable, and (iii) $\phi(\cdot) \in C_n^b$. Then we have:

- The algebraic Riccati equation $\Pi A + A^T \Pi - \Pi B R^{-1} B^T \Pi + Q = 0$ has a unique positive semi-definite solution Π .
- $\Gamma := A - B R^{-1} B^T \Pi$ is asymptotically stable.
- The differential equation $ds/dt = -(A - B R^{-1} B^T \Pi)^T s + Q \phi$ has a unique solution in C_n^b :

$$s(t) = - \int_t^\infty e^{-\Gamma(t-\tau)} Q \phi(\tau) d\tau, \quad t \geq 0.$$

- The optimal control law: $u^o(\cdot) := \arg \inf_{u(\cdot) \in \mathcal{U}} J(u(\cdot)) = -R^{-1} B^T (\Pi z(\cdot) + s(\cdot))$.

Proof: See Theorem 3.1 in [15]. \square

B. MF LQG Systems of Equations

Considering the L-F model (1)–(6) we need to solve a set of tracking optimal control problems where the tracking trajectories are $\phi^L(z^{L,N_L})(\cdot)$ and $\phi^F(z^{L,N_L}, z^{F,N_F})(\cdot)$ for leaders and followers, respectively. However, these tracking trajectories cannot be known *a priori* and so cannot be used for

constructing the control laws. Therefore, we take an MF stochastic control approach to approximate the coupling trajectory terms by purely deterministic processes called mass behaviors of leaders and followers.

In the MF methodology, each agent assumes that, in the large population limit, its individual action has no impact on the mass behavior. In turn, the impact of the mass on the agent is captured in the limit via a posited deterministic but unknown trajectory which through consistency requirements is then shown to satisfy a fixed point equation system. Hence, each individual agent's control law may be viewed as a result of that agent having solved an optimal tracking problem for which the resulting control law is a combination of a local state feedback and a pre-computable mass dependent open loop deterministic component.

For any leader or follower representative agent with dynamical parameter $\theta = [A, B, C]$ (where $\theta \in \Theta_L$ for leaders and $\theta \in \Theta_F$ for followers) let Π_θ be the solution to the algebraic Riccati equation

$$\Pi_\theta A + A^T \Pi_\theta - \Pi_\theta B R^{-1} B^T \Pi_\theta + Q = 0 \quad (10)$$

and let $\Gamma_\theta := A - B R^{-1} B^T \Pi_\theta$.

We have the following assumptions for the model under consideration:

(A1) For each $\theta = [A, B, C]$ from the compact sets Θ_L or Θ_F , we assume that the pair $[A, B]$ is stabilizable and the pair $[A, Q^{1/2}]$ is detectable. \square

Remark 3.1: Let Assumption **(A1)** hold. Then for any $\theta = [A, B, C] \in \Theta_L \cup \Theta_F$, the algebraic Riccati (10) has a unique positive semi-definite solution, Γ_θ is asymptotically stable and there exist positive γ, ρ such that $\|e^{\Gamma_\theta t}\| \leq \gamma e^{-\rho t}$ for all $t \geq 0$ [30].

For the sequence of the leaders and followers' dynamical parameters, $\{\theta_i \in \Theta_L : 1 \leq i \leq N_L\}$ and $\{\theta_i \in \Theta_F : 1 \leq i \leq N_F\}$, respectively, define the empirical distributions

$$F_N^L(\theta) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{\theta_i \in \Theta_L : \theta_i \leq \theta\}},$$

$$F_N^F(\theta) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{\theta_i \in \Theta_F : \theta_i \leq \theta\}}$$

where $\theta_i \leq \theta$ means the componentwise inequality for the two vectors θ_i and θ , and $\mathbf{1}_{\{\theta_i \in \Theta_L : \theta_i \leq \theta\}} = 1$ if $\Theta_L \ni \theta_i \leq \theta$ holds, and $\mathbf{1}_{\{\theta_i \in \Theta_L : \theta_i \leq \theta\}} = 0$ otherwise. Similarly, $\mathbf{1}_{\{\theta_i \in \Theta_F : \theta_i \leq \theta\}} = 1$ if $\Theta_F \ni \theta_i \leq \theta$ holds, and $\mathbf{1}_{\{\theta_i \in \Theta_F : \theta_i \leq \theta\}} = 0$ otherwise.

(A2) There exist two probability distributions $F^L(\cdot)$ and $F^F(\cdot)$ such that F_N^L converges to F^L and F_N^F converges to F^F weakly, i.e., for any bounded and continuous function $\phi(\theta)$,

$$\lim_{N \rightarrow \infty} \int \phi(\theta) dF_N^L(\theta) = \int \phi(\theta) dF^L(\theta),$$

$$\lim_{N \rightarrow \infty} \int \phi(\theta) dF_N^F(\theta) = \int \phi(\theta) dF^F(\theta)$$

where dF denotes the measure induced by the distribution functions F . \square

Remark 3.2: It is important to note that if the sequences $\{\theta_i \in \Theta_L : 1 \leq i \leq N_L\}$ and $\{\theta_i \in \Theta_F : 1 \leq i \leq N_F\}$ are generated by independent randomized observations on the distributions F^L and F^F , respectively, then **(A2)** holds with probability one by the Glivenko-Cantelli theorem [7].

By the probability distributions $F^L(\cdot)$ and $F^F(\cdot)$ we define

$$\psi^{L,\infty}(\cdot) := \int_{\Theta_L} \bar{z}_\theta^L(\cdot) dF^L(\theta),$$

$$\psi^{F,\infty}(\cdot) := \int_{\Theta_F} \bar{z}_\theta^F(\cdot) dF^F(\theta)$$

as the centroid of the leaders and followers, respectively, in the infinite population limit. The functions $\psi^{L,\infty}(\cdot)$ and $\psi^{F,\infty}(\cdot)$ are intended to respectively approximate $z^{L,N_L}(\cdot)$ and $z^{F,N_F}(\cdot)$, defined in (2) and (5), in the infinite population limit.

Denote Γ_{θ_i} by Γ_i . Applying the MF LQG approach [7], [15] to the leaders' dynamic game model (1)–(3), we obtain the leaders' MF system of equations in the infinite leaders' population limit

$$\frac{ds_i^L}{dt} = - (A_i - B_i R^{-1} B_i^T \Pi_{\theta_i})^T s_i^L + Q \phi^{L,\infty} \quad (11)$$

$$\frac{d\bar{z}_i^L}{dt} = (A_i - B_i R^{-1} B_i^T \Pi_{\theta_i}) \bar{z}_i^L - B_i R^{-1} B_i^T s_i^L \quad (12)$$

$$\psi^{L,\infty}(\cdot) = \int_{\Theta_L} \bar{z}_{\theta_i}^L(\cdot) dF^L(\theta_i) \quad (13)$$

$$\phi^{L,\infty}(\cdot) = \lambda h(\cdot) + (1 - \lambda) \psi^{L,\infty}(\cdot) \quad (14)$$

where $\bar{z}_i^L(\cdot) := \bar{z}_{\theta_i}^L(\cdot)$, $s_i^L(\cdot) := s_{\theta_i}^L(\cdot)$, $s_i^L(0) = - \int_0^\infty e^{\Gamma_i \tau} Q \phi^{L,\infty}(\tau) d\tau$ and $\bar{z}_i^L(0)$ is given for $\theta_i \in \Theta_L$.

The equation system (11)–(14) is similar to the MF LQG equation system in [15] but with a time-varying reference trajectory $h(\cdot)$, and is constructed such that each leader carries out optimal tracking of the leaders' mass behavior $\phi^{L,\infty}(\cdot)$. More precisely, (i) (11) is the mass offset tracking equation, (ii) (12) is obtained by taking expectations of the closed-loop dynamics of a leader with dynamical parameter $\theta_i \in \Theta_L$ while the best response is $u_i^{L,\infty}(\cdot) := -R^{-1} B_i^T (\Pi_{\theta_i} z_i^L(\cdot) + s_i^L(\cdot))$, (iii) $\psi^{L,\infty}(\cdot)$ in (13) is the centroid of the leaders in the infinite population limit, and (iv) $\phi^{L,\infty}(\cdot)$ in (14) is approximating $\phi^L(z^{L,N_L})(\cdot)$, defined in (2), in the infinite population limit.

Next, we obtain the followers' MF system of equations associated with the followers' dynamic game model (4)–(6) in the infinite population limit

$$\frac{ds_i^F}{dt} = - (A_i - B_i R^{-1} B_i^T \Pi_{\theta_i})^T s_i^F + Q \phi^{F,\infty} \quad (15)$$

$$\frac{d\bar{z}_i^F}{dt} = (A_i - B_i R^{-1} B_i^T \Pi_{\theta_i}) \bar{z}_i^F - B_i R^{-1} B_i^T s_i^F \quad (16)$$

$$\psi^{L,\infty}(\cdot) = \int_{\Theta_L} \bar{z}_{\theta_i}^L(\cdot) dF^L(\theta_i) \quad (17)$$

$$\psi^{F,\infty}(\cdot) = \int_{\Theta_F} \bar{z}_{\theta_i}^F(\cdot) dF^F(\theta_i) \quad (18)$$

$$\phi^{F,\infty}(\cdot) = \eta \psi^{L,\infty}(\cdot) + (1 - \eta) \psi^{F,\infty}(\cdot) \quad (19)$$

where $\bar{z}_i^F(\cdot) := \bar{z}_{\theta_i}^F(\cdot)$, $s_i^F(\cdot) := s_{\theta_i}^F(\cdot)$, $s_i^F(0) = - \int_0^\infty e^{\Gamma_i \tau} Q \phi^{F,\infty}(\tau) d\tau$ and $\bar{z}_i^F(0)$ is given for $\theta_i \in \Theta_L$.

The equation system (15)–(19) is constructed such that each follower carries out optimal tracking of the followers' mass behavior $\phi^{F,\infty}(\cdot)$. More precisely, (i) (15) is the mass offset tracking equation, (ii) (16) is obtained by taking expectations of the closed-loop dynamics of a follower with dynamical parameter θ_i while the best response is

$u_i^{F,\infty}(\cdot) := -R^{-1}B_i^T(\Pi_{\theta_i}z_i^F(\cdot) + s_i^F(\cdot))$, (iii) $\psi^{L,\infty}(\cdot)$ in (17) is the centroid of the leaders, in the infinite population limit, computed by the followers from the leaders' MF equation system (11)–(14), (iv) $\psi^{F,\infty}(\cdot)$ in (18) is the centroid of the followers in the infinite population limit, and (v) $\phi^{F,\infty}(\cdot)$ in (19) is approximating $\phi^F(z^{L,N_L}, z^{F,N_F})(\cdot)$, defined in (5), in the infinite population limit.

It is important to note that the followers' MF system of equations is coupled to the leaders' MF system of equations due to appearance of leaders' centroid $\psi^{L,\infty}(\cdot)$ in the mass behavior of the followers $\phi^{F,\infty}(\cdot)$. We make the following assumption:

(A3) We assume that for all possible reference trajectories $h_\delta(\cdot)$ parameterized by $\delta \in \Delta$:

- The leaders' MF system of (11)–(14) has a unique solution such that $\psi^{L,\infty} \in C_n^b$.
- The followers' MF system of (15)–(19) (implicitly depending upon the solution of (11)–(14)) has a unique solution such that $\psi^{F,\infty} \in C_n^b$.

□

Generally, it seems difficult to verify this assumption. However, in Section III-C we provide sufficient conditions for the existence and uniqueness of solutions to these MF equation systems by using a contractive mapping argument (see [7], [15]).

C. Analysis of MF Systems of Equations for Leaders and Followers

First, we consider the leaders' MF system of (11)–(14). For given $h(\cdot)$, $\psi^{L,\infty}(\cdot)$ and hence $\phi^{L,\infty}(\cdot)$, the unique solution of the tracking offset (11) for a “generic” leader agent with dynamical parameter $\theta_i = [A_i, B_i, C_i] \in \Theta_L$ is

$$\begin{aligned} s_i^L(t) &= - \int_t^\infty e^{-\Gamma_i^T(t-\tau)} Q \phi^{L,\infty}(\tau) d\tau \\ &= - \int_t^\infty e^{-\Gamma_i^T(t-\tau)} Q \\ &\quad \times (\lambda h(\tau) + (1-\lambda)\psi^{L,\infty}(\tau)) d\tau, \quad t \geq 0. \end{aligned} \quad (20)$$

Next, by solving (12) we have

$$\bar{z}_i^L(t) = e^{\Gamma_i t} \bar{z}_i^L(0) - \int_0^t e^{\Gamma_i(t-s)} B_i R^{-1} B_i^T s_i^L(s) ds, \quad t \geq 0 \quad (21)$$

which by substituting $s_i^L(\cdot)$ from (20) we get

$$\begin{aligned} \bar{z}_i^L(t) &= e^{\Gamma_i t} \bar{z}_i^L(0) + \int_0^t e^{\Gamma_i(t-s)} B_i R^{-1} B_i^T \\ &\quad \times \left(\int_s^\infty e^{-\Gamma_i^T(s-\tau)} Q (\lambda h(\tau) \right. \\ &\quad \left. + (1-\lambda)\psi^{L,\infty}(\tau)) d\tau \right) ds \\ &:= (\Upsilon_i^L(\psi^{L,\infty}, h))(t), \quad t \geq 0 \end{aligned} \quad (22)$$

where Υ_i^L is an operator acting on bounded continuous functions. This and (13) result in

$$\begin{aligned} \psi^{L,\infty}(t) &\equiv \int_{\Theta_L} \bar{z}_i^L(t) dF^L(\theta_i) \\ &= \int_{\Theta_L} \Upsilon_i^L(\psi^{L,\infty}, h)(t) dF^L(\theta_i) \\ &:= \Upsilon^L(\psi^{L,\infty}, h)(t), \quad t \geq 0 \end{aligned}$$

where we note $\psi^{L,\infty}(\cdot)$ is independent of i , $\theta_i \in \Theta_L$. Now for a given $h(\cdot)$ if the equation

$$\psi^{L,\infty} = \Upsilon^L(\psi^{L,\infty}, h) \quad (23)$$

has a unique solution $\psi^{L,\infty}(\cdot)$, then it can be used in (20) and (22) to compute the unique solution of the leaders' MF system of equations. In the following theorem we employ a contractive mapping argument to provide sufficient conditions under which (23) has a unique solution. We omit the proofs of the two following theorems which closely resemble that of Theorem 3.2 in [15].

Theorem 3.2: Assume **(A1)** and **(A2)** hold. For a given $h(\cdot) \in C_n^b$,

- Υ^L defined in (23) is an operator from C_n^b to C_n^b .
- If

$$(1-\lambda) \|R^{-1}\| \|Q\| \int_{\Theta_L} \|B_{\theta_i}\|^2 \left(\int_0^\infty \|e^{\Gamma_{\theta_i} t}\|^2 dt \right) dF^L(\theta_i) < 1 \quad (24)$$

then (23) has the unique solution $\psi^{L,\infty} \in C_n^b$, and so the leaders' MF system of equations, (11)–(14), has a unique solution which for a continuum of agents consists of the θ_i parameterized quadruple $(s_i^L(\cdot), \bar{z}_i^L(\cdot), \psi^{L,\infty}(\cdot), \phi^{L,\infty}(\cdot))$, $\theta_i \in \Theta_L$. □

Second, we consider the followers' MF system of equations, (15)–(19). For given $\psi^{L,\infty}(\cdot)$, $\psi^{F,\infty}(\cdot)$ and hence $\phi^{F,\infty}(\cdot)$, we define the operator Υ_i^F for a “generic” follower agent with dynamical parameter $\theta_i = [A_i, B_i, C_i] \in \Theta_F$ as

$$\begin{aligned} \bar{z}_{\theta_i}^F(t) &= e^{\Gamma_{\theta_i} t} \bar{z}_{\theta_i}^F(0) + \int_0^t e^{\Gamma_{\theta_i}(t-s)} B_i R^{-1} B_i^T \\ &\quad \times \left(\int_s^\infty e^{-\Gamma_{\theta_i}^T(s-\tau)} Q (\eta \psi^{L,\infty}(\tau) \right. \\ &\quad \left. + (1-\eta)\psi^{F,\infty}(\tau)) d\tau \right) ds \\ &:= \Upsilon_i^F(\psi^{F,\infty}, \psi^{L,\infty})(t), \quad t \geq 0 \end{aligned}$$

by using the solutions of (15) and (16) (similar to (22) for the i -th leader). This and (18) result in

$$\begin{aligned} \psi^{F,\infty}(t) &\equiv \int_{\Theta_F} \bar{z}_{\theta_i}^F(t) dF^F(\theta_i) \\ &= \int_{\Theta_F} \Upsilon_i^F(\psi^{F,\infty}, \psi^{L,\infty})(t) dF^F(\theta_i) \\ &:= \Upsilon^F(\psi^{F,\infty}, \psi^{L,\infty})(t), \quad t \geq 0 \end{aligned}$$

where we note $\psi^{F,\infty}(\cdot)$ is independent of $\theta_i \in \Theta_F$. Now for a given $\psi^{L,\infty}(\cdot)$ (the solution of (23)) if the equation

$$\psi^{F,\infty} = \Upsilon^F(\psi^{F,\infty}, \psi^{L,\infty}) \quad (25)$$

has a unique solution $\psi^{F,\infty}(\cdot)$, then it can be used in (15) and (16) to compute the unique solution of the followers' MF system of equations. In the following theorem we employ the contractive mapping argument to provide sufficient conditions under which (25) has a unique solution.

Theorem 3.3: Assume **(A1)** and **(A2)** hold. For a given $\psi^{L,\infty}(\cdot) \in C_n^b$,

- Υ^F defined in (25) is an operator from C_n^b to C_n^b .

(b) If

$$(1-\eta)\|R^{-1}\| \|Q\| \int_{\Theta_F} \|B_{\theta_i}\|^2 \left(\int_0^\infty \|e^{\Gamma_{\theta_i} t}\|^2 dt \right) dF^F(\theta_i) < 1 \quad (26)$$

then (25) has the unique solution $\psi^{F,\infty} \in C_n^b$, and so the followers' MF system of equations, (15)–(19), has a unique solution which for a continuum of agents consists of the θ_i parameterized quadruple $(s_i^F(\cdot), \bar{z}_i^F(\cdot), \psi^{F,\infty}(\cdot), \phi^{F,\infty}(\cdot))$, $\theta_i \in \Theta_F$. \square

Intuitively, (24) and (26) mean that λ and η should be reasonably close to 1, which means that the leaders should give sufficient attention to the reference trajectory, and on the other hand the followers should give sufficient attention to the group of leaders, so that at the end a desirable mean field behavior (fixed point) can set in.

Remark 3.3: It is important to note that the conditions (24) and (26) do not depend on the reference trajectory of the leaders $h(\cdot)$.

IV. ESTIMATION PROCEDURE FOR THE ADAPTIVE FOLLOWERS

The computation of the followers' MF system of equations (and hence followers' control laws) requires knowledge of the complete reference trajectory of the leaders $h(\cdot)$ which is in general not known to the followers. In this section we construct an adaptation procedure for a generic follower using a likelihood ratio based estimator (on a sample population of the leaders' trajectories) to identify the member generating the reference trajectory $h(\cdot)$ from a given finite set of possible parameters each identifying a single trajectory. Hence, followers are adaptive in the sense that they use an estimator to identify $h(\cdot)$. As stated earlier, we assume that the reference trajectory $h(\cdot)$ is parameterized with δ from a finite set Δ such that $h_\delta \in C_n^b$ for every $\delta \in \Delta$.

Likelihood ratio based estimation is a well-known method for generating estimates of an unknown stochastic model parameter (see, e.g., [31] and [32]). In this paper we use the general result on the convergence of likelihood ratio estimators for stochastic processes parameterized by a finite set of alternative values which was established in [33] (see [31]).

A. The Likelihood Function

For a generic adaptive follower, we define the *likelihood function* (see [34], [35]) on (a subset of) the leaders' trajectories by ($0 < t < \infty$)

$$L_t^M(\delta) := \exp \left(\int_0^t (z_{\delta,s}^{L,M})^T dy_s^M - \frac{1}{2} \int_0^t \|z_{\delta,s}^{L,M}\|^2 ds \right) \quad (27)$$

where $z_{\delta,t}^{L,M} := 1/M \sum_{i=1}^M z_{i,\delta}^L(t)$ is the centroid of the leaders' states when the defining parameter of $h(\cdot)$ is assumed to be $\delta \in \Delta$, and $y^M(\cdot)$ is the observation process of the generic follower (which observes a non-empty random subset $O \subset L$ of cardinality $M \equiv M(N_L) = \lfloor \sqrt{N_L} \rfloor$ of the leaders' trajectories) as defined in (7).

Without loss of generality assume $\delta_1 \in \Delta$ is the true parameter of the reference trajectory $h(\cdot)$ in the rest of the paper, that is

to say, the parameter of the leaders generating the data. Therefore, the observation process of the generic follower is of the form

$$dy^M = \left(\frac{1}{M} \sum_{i \in O} z_{i,\delta_1}^L \right) dt + \frac{1}{M} \sum_{i=1}^M D_i dv_i, \quad t \geq 0 \quad (28)$$

where $\{v_i : 1 \leq i \leq N_L\}$ is a set of independent standard Wiener processes.

We define the *asymptotic (in population) likelihood function* of the generic adaptive follower to be the deterministic function ($0 < t < \infty$)

$$L_t^\infty(\delta) := \exp \left(\int_0^t (\psi_{\delta,s}^{L,\infty})^T \psi_{\delta_1,s}^{L,\infty} ds - \frac{1}{2} \int_0^t \|\psi_{\delta,s}^{L,\infty}\|^2 ds \right) \quad (29)$$

where $\psi_{\delta,t}^{L,\infty} := \psi_\delta^{L,\infty}(t)$ such that $\psi_\delta^{L,\infty}(\cdot)$ is the deterministic infinite population leaders' centroid computed from the leaders' MF system of (11)–(14) when the defining parameter of $h(\cdot)$ is assumed to be $\delta \in \Delta$.

Since (i) the processes $z_\delta^{L,M}(\cdot)$, $\delta \in \Delta$, in (27) are not computable or observable for the adaptive followers, and (ii) the true infinite population centroid of the leaders $\psi_{\delta_1}^{L,\infty}(\cdot)$ in (29) is not known to the followers, we introduce the following *hybrid likelihood function* for a generic adaptive follower with observation process $y^M(\cdot)$ ($0 < t < \infty$)

$$H_t^M(\delta) := \exp \left(\int_0^t (\psi_{\delta,s}^{L,\infty})^T dy_s^M - \frac{1}{2} \int_0^t \|\psi_{\delta,s}^{L,\infty}\|^2 ds \right). \quad (30)$$

It is important to note that the hybrid likelihood function (30) is computable for adaptive followers and includes: (i) the infinite population centroid of the leaders $\psi_\delta^{L,\infty}(\cdot)$ which can be computed by each adaptive follower from the leaders' MF system of (11)–(14) when the defining parameter of the reference trajectory $h(\cdot)$ is assumed to be $\delta \in \Delta$, and (ii) the observation process $y^M(\cdot)$ which is given in (28) for a generic adaptive follower.

Proposition 4.1: Assume (A1)–(A3-a) hold. Then

(a) For each $\delta \in \Delta$ and $t > 0$

$$\lim_{M \rightarrow \infty} |L_t^M(\delta) - L_t^\infty(\delta)| = 0 \quad a.s.$$

i.e., for each $\delta \in \Delta$, $t > 0$ and $\varepsilon > 0$ there exists a random $M_{\delta,\varepsilon,t}$, $0 < M_{\delta,\varepsilon,t}(\omega) < \infty$, such that $|L_t^m(\delta) - L_t^\infty(\delta)| < \varepsilon$ a.s. for all $m > M_{\delta,\varepsilon,t}$.

(b) For each $\delta \in \Delta$ and $t > 0$,

$$\lim_{M \rightarrow \infty} |H_t^M(\delta) - L_t^\infty(\delta)| = 0 \quad (a.s.).$$

Proof: See Appendix. \square

B. The Likelihood Ratio

The likelihood ratio test provides the means for comparing the likelihood of the observations under one hypothesis about the unknown parameters of the model against the likelihood of the observations under alternative hypotheses.

At each instant $0 < t < \infty$ the set of the *likelihood ratios* is

$$LR^M(t) := \left\{ \frac{L_t^M(\delta_i)}{L_t^M(\delta_j)} : \delta_i, \delta_j \in \Delta, \delta_i \neq \delta_j \right\}$$

in which each ratio $L_t^M(\delta_i)/L_t^M(\delta_j)$ depends explicitly upon the hypotheses δ_i and δ_j .

It is important to note that for any fixed δ_i in Δ the process $(L_t^M(\delta_i)/L_t^M(\delta_1))_{t \geq 0}$ is a positive martingale (with respect to the filtration $\{\mathcal{F}_t^{z^L, y^M}\}_{t \geq 0}$ where $\mathcal{F}_t^{z^L, y^M}$ is defined as the σ -field $\sigma(z_i^L(\tau), y^M(\tau) : 1 \leq i \leq M, \tau < t)$) [33], [34]. Therefore, for any fixed $\delta_i \in \Delta$, $(L_t^M(\delta_i)/L_t^M(\delta_1))_{t \geq 0}$ a.s. converges to a limiting random variable by the Martingale Convergence Theorem [36]. Hence, for any fixed $\delta_i, \delta_j \in \Delta$, $\delta_i \neq \delta_j$, the process $(L_t^M(\delta_i)/L_t^M(\delta_j))_{t \geq 0}$ converges a.s. to a positive limiting random variable.

C. Main Estimation Theorem

We define the *Likelihood Ratio Estimator* (LRE) for a generic adaptive follower i , $1 \leq i \leq N_F$, with observation size m as

$$\hat{\delta}_i^m(t) := \left\{ \delta \in \Delta : \frac{H_{t_k}^m(\delta)}{H_{t_k}^m(\delta')} \geq 1, \forall \delta' \in \Delta, \delta' \neq \delta \right\} \quad (31)$$

where $t \in [t_k, t_k + \tau_i)$, τ_i is a pre-specific positive number and t_0, t_1, \dots is an infinite switching time sequence such that $t_{k+1} - t_k = \tau_i$, $k \geq 0$. $\hat{\delta}_i^m(\cdot)$ is a parameter selector between finite alternatives which is made at each instant. At each time t if $\hat{\delta}_i^m(t)$ has more than one member, a tie-breaking rule measurable with respect to the σ -field $\sigma(y_i^m(\tau) : \tau < t)$ is employed.

We now enunciate the following verifiable identifiability condition.

(A4) (Identifiability Condition) there exist a deterministic real number $\alpha > 0$, and deterministic time T_α , $0 < T_\alpha < \infty$, such that

$$\int_0^t \|\psi_{\delta_i, s}^{L, \infty} - \psi_{\delta_j, s}^{L, \infty}\|^2 ds > \alpha, \quad \forall \delta_i, \delta_j \in \Delta, \delta_i \neq \delta_j, t > T_\alpha$$

where for any fixed $\delta \in \Delta$, $\psi_\delta^{L, \infty}(\cdot)$ is the deterministic infinite population leaders' centroid computed from (11)–(14), when the defining parameter of $h(\cdot)$ is assumed to be δ . \square

Remark 4.1: This identifiability assumption implies that the corresponding centroid of the leaders for any two distinct parameters of the set Δ (which characterizes the reference trajectory $h(\cdot)$) is distinguishable, after some finite deterministic time.

Lemma 4.1: Assume **(A1)**–**(A3-a)** and **(A4)** hold. Then there exist a deterministic ε , $0 < \varepsilon < 1$, and a deterministic time T_ε , $0 < T_\varepsilon < \infty$, such that

$$\frac{L_t^\infty(\delta)}{L_t^\infty(\delta_1)} < 1 - \varepsilon, \quad \forall \delta \in \Delta, \delta \neq \delta_1, t > T_\varepsilon.$$

Proof: For any fixed $\delta \in \Delta$, $\delta \neq \delta_1$, and t , $0 < t < \infty$, by (29) we have

$$\begin{aligned} \frac{L_t^\infty(\delta)}{L_t^\infty(\delta_1)} &= \exp \left[\int_0^t \left(\psi_{\delta, s}^{L, \infty} - \psi_{\delta_1, s}^{L, \infty} \right)^T \psi_{\delta_1, s}^{L, \infty} ds \right. \\ &\quad \left. - \frac{1}{2} \int_0^t \left(\|\psi_{\delta, s}^{L, \infty}\|^2 - \|\psi_{\delta_1, s}^{L, \infty}\|^2 \right) ds \right] \\ &= \exp \left[-\frac{1}{2} \left(\int_0^t \|\psi_{\delta, s}^{L, \infty} - \psi_{\delta_1, s}^{L, \infty}\|^2 ds \right) \right]. \end{aligned}$$

But, by the Identifiability Condition **(A4)**, there exist a deterministic real number $\alpha > 0$, and deterministic time T_α , $0 < T_\alpha < \infty$, such that for all $t > T_\alpha$

$$\begin{aligned} &\exp \left[-\frac{1}{2} \left(\int_0^t \|\psi_{\delta, s}^{L, \infty} - \psi_{\delta_1, s}^{L, \infty}\|^2 ds \right) \right] \\ &< \exp \left(-\frac{1}{2} \alpha \right) \quad \delta \in \Delta, \delta \neq \delta_1 \end{aligned}$$

and setting $1 - \varepsilon = \exp(-\frac{1}{2}\alpha)$ yields the result with $0 < \varepsilon < 1$. \square

Lemma 4.2: Assume **(A1)**–**(A3-a)** and **(A4)** hold. Then there exist a deterministic η , $0 < \eta < 1$, a deterministic time T_η , $0 < T_\eta < \infty$, and, with probability one, a random M_η , $0 < M_\eta(\omega) < \infty$, such that for all $t > T_\eta$ and $m > M_\eta(\omega)$

$$\frac{H_t^m(\delta)}{H_t^m(\delta_1)} < 1 - \eta, \quad \delta \in \Delta, \delta \neq \delta_1 \quad (a.s.).$$

Proof: By Lemma 4.1 there exist a deterministic ε , $0 < \varepsilon < 1$, and a deterministic time T_ε , $0 < T_\varepsilon < \infty$, such that for any $\delta \in \Delta$, $\delta \neq \delta_1$, and $t > T_\varepsilon$

$$\frac{L_t^\infty(\delta)}{L_t^\infty(\delta_1)} < 1 - \varepsilon. \quad (32)$$

Now choose $\eta := \eta_\varepsilon$ such that $0 < \eta < \varepsilon < 1$. Then at each instant t , $t > T_\eta := T_{\eta_\varepsilon}$, by Proposition 4.1 with $\zeta := ((\varepsilon - \eta)L_t^\infty(\delta_1)/2 - \eta) > 0$, there exists, with probability one, $M_\eta(\omega) := M_{\zeta, T_\varepsilon(\omega)}(\omega)$, $0 < M_\eta(\omega) < \infty$, where $M_\eta(\omega) = \max \{M_{\delta_1, \eta}(\omega), M_{\delta_2, \eta}(\omega), \dots, M_{\delta_{|\Delta|}, \eta}(\omega)\}$ such that for all $m > M_\eta$ and $\delta \in \Delta$

$$|H_t^m(\delta) - L_t^\infty(\delta)| < \zeta \quad (a.s.). \quad (33)$$

Hence, by (32) and (33), for all $t > T_\eta$ and $m > M_\eta(\omega)$ we have

$$\begin{aligned} \frac{H_t^m(\delta)}{H_t^m(\delta_1)} &< \frac{L_t^\infty(\delta) + \zeta}{L_t^\infty(\delta_1) - \zeta} = \frac{\frac{L_t^\infty(\delta)}{L_t^\infty(\delta_1)} + \frac{\varepsilon - \eta}{2 - \eta}}{1 - \frac{\varepsilon - \eta}{2 - \eta}} \\ &< \frac{1 - \varepsilon + \frac{\varepsilon - \eta}{2 - \eta}}{1 - \frac{\varepsilon - \eta}{2 - \eta}} = 1 - \eta, \quad \delta \in \Delta, \delta \neq \delta_1 \quad (a.s.). \end{aligned}$$

\square

Theorem 4.1: Assume **(A1)**–**(A3-a)** and **(A4)** hold. Then for each generic adaptive follower i , $1 \leq i \leq N_F$, there exist a deterministic T_i , $0 < T_i < \infty$, and, with probability one, a random M_i , $0 < T_i, M_i(\omega) < \infty$, such that $\hat{\delta}_i^m(t) = \delta_1$ for all $t > T_i$ and $m > M_i(\omega)$.

Proof: By Lemma 4.2 for any adaptive follower i , $1 \leq i \leq N_F$, there exist a deterministic η_i , $0 < \eta_i < 1$, a deterministic time $T_i := T_{\eta_i}$, $0 < T_i < \infty$, and a random $M_i := M_{\eta_i}$, $0 < M_i(\omega) < \infty$, such that for all $t > T_i$, $m > M_i(\omega)$ and $\delta \in \Delta$ where $\delta \neq \delta_1$

$$\frac{H_t^m(\delta_1)}{H_t^m(\delta)} = \left(\frac{H_t^m(\delta)}{H_t^m(\delta_1)} \right)^{-1} \geq (1 - \eta_i)^{-1} > 1 \quad a.s.$$

which implies that $\hat{\delta}_i^m(t) = \delta_1$ for all $t > T_i$ and $m > M_i(\omega)$, based on the definition of the LRE in (31). \square

V. STABILITY ANALYSIS OF THE MEAN FIELD CONTROL LAWS

Definition 5.1: In a large but finite population, the decentralized MF control laws for the leaders and followers, respectively, are as follows.

- *Leaders' MF Control Laws:* The control strategy of each generic leader i , $1 \leq i \leq N_L$, with dynamical parameter $\theta_i = [A_i, B_i, C_i] \in \Theta_L$ is defined as

$$u_i^{L,\infty}(\cdot) := -R^{-1}B_i^T(\Pi_{\theta_i}z_i^L(\cdot) + s_i^L(\cdot)) \quad (34)$$

where $s_i^L(\cdot)$ is the solution of (11) presented in (20).

- *Followers' MF Control Laws:* The control strategy of each generic adaptive follower i , $1 \leq i \leq N_F$, with dynamical parameter $\theta_i = [A_i, B_i, C_i] \in \Theta_F$, and observation size m is defined as

$$\hat{u}_i^{F,\infty}(\cdot) := -R^{-1}B_i^T(\Pi_{\theta_i}z_i^F(\cdot) + s_{\delta_i^m}^F(\cdot)) \quad (35)$$

where $s_{\delta_i^m}^F(\cdot)$ is the solution of (15), when the defining parameter of the reference trajectory $h(\cdot)$ is assumed to be $\hat{\delta}_i^m(\cdot)$ defined in (31). \square

In the construction of individual strategies, (i) each leader needs to know λ , $h(\cdot)$, the leader's population initial mean and the distribution of the leaders' dynamical parameters ($F^L(\cdot)$), where the term know denotes that the control law of the agent in question may be an explicit function of the indicated information; and (ii) each follower needs to know λ , η , the population initial means and the distributions of both leaders and followers' dynamical parameters ($F^L(\cdot)$ and $F^F(\cdot)$), but each adaptive follower does not know the reference trajectory of the leaders $h(\cdot)$ and estimates it by likelihood estimation from a finite set of pre-defined signals based upon its observation process. Note that it is not required for any leader or follower agent to know specific information (such as the dynamical parameter) of any other particular agent.

A. Follower's Adaptive Mean Field Algorithm

The algorithm has the following two phases:

- 1) *Estimation Phase:* By observing a sample population of the leaders each follower computes the set of likelihood ratios (based on the hybrid likelihood functions defined in (30)) at each instant for alternative values of its hypothesis parameter $\delta \in \Delta$. Each follower also computes control laws by using the parameters in the finite set Δ . Therefore, each follower has a set of control strategies with respect to alternative defining parameters of the reference trajectory $h(\cdot)$, and at any instant uses the maximum likelihood ratio estimate (MLRE) without a guarantee that the MLRE has taken the true parameter value.
- 2) *Lock-on Phase:* As the observation size of each adaptive follower i , $1 \leq i \leq N_F$, goes to infinity, its estimate converges to the true parameter of the unknown reference trajectory at a deterministic time T_i (Theorem 4.1). In this phase, the control law of each adaptive follower i , $1 \leq i \leq N_F$, will necessarily be computed with the true parameter of the reference trajectory for all time $t > T_i$ and any sufficiently large random observation size.

Needless to say, an adaptive follower cannot deduce at which population size the lock-on phase has commenced since this occurs at some random size $M(\omega)$.

B. Stability Analysis

For the i -th leader, $1 \leq i \leq N_L$, with dynamical parameter $\theta_i = [A_i, B_i, C_i] \in \Theta_L$, denote $z_i^{L,\infty}(\cdot)$ as the closed-loop solution of its dynamics (1) while using the MF control law $u_i^{L,\infty}(\cdot)$ as defined in (34). In an analogous way let $\hat{z}_i^{F,\infty}(\cdot)$ be the closed-loop solution of the i -th adaptive follower's dynamics, $1 \leq i \leq N_F$, (4) where its dynamical parameter is $\theta_i = [A_i, B_i, C_i] \in \Theta_F$ and its adaptive control law is the MF control law $\hat{u}_i^{F,\infty}(\cdot)$ defined in (35).

Theorem 5.1: (Stability of the MF control laws in the sense of time average)

- (a) Assume **(A1)**-**(A3-a)** hold. Then (*a.s.*)

$$\sup_{N_L \geq 1} \max_{1 \leq i \leq N_L} \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T (\|z_i^{L,\infty}(s)\|^2 + \|u_i^{L,\infty}(s)\|^2) ds < \infty. \quad (36)$$

- (b) Assume **(A1)**-**(A4)** hold. Then (*a.s.*)

$$\sup_{N_F \geq 1} \max_{1 \leq i \leq N_F} \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T (\|\hat{z}_i^{F,\infty}(s)\|^2 + \|\hat{u}_i^{F,\infty}(s)\|^2) ds < \infty. \quad (37)$$

Proof: See Appendix. \square

It is important to note that these stability results depend on the boundedness of the infinite population centroids $\psi^{L,\infty}(\cdot)$ and $\psi^{F,\infty}(\cdot)$ computed from the leaders and followers' MF system of equations, respectively.

VI. ε -NASH EQUILIBRIUM PROPERTY OF THE MEAN FIELD BASED CONTROL LAWS

In a system of population size N , let the admissible control set of agent i , $1 \leq i \leq N$, be $\mathcal{U}_i := \{u_i(\cdot) : u_i(t) \text{ is adapted to } \sigma(z_j(s), s \leq t, 1 \leq j \leq N), \|z_i(T)\| = o(\sqrt{T}), \int_0^T \|z_i(t)\|^2 dt = O(T) \text{ a.s.}\}$. Note that the strategies in \mathcal{U}_i may use full state information (i.e., $u_i(t)$ is a function of both time t and the system state at this time, $(z_1(t), \dots, z_N(t))$). Denote $u_{-i} = (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_N)$. To indicate the dependence of the i -th agent's cost function J_i on u_i , u_{-i} and population size N , we write it as $J_i^N(u_i, u_{-i})$.

Definition 6.1: Given $\varepsilon > 0$, a set of controls $u_k^o \in \mathcal{U}_k$, $1 \leq k \leq N$, for N agents generates an *a.s.* ε -Nash equilibrium with respect to the costs J_k , $1 \leq k \leq N$, if for any i , $1 \leq i \leq N$

$$\begin{aligned} J_i^N(u_i^o, u_{-i}^o) - \varepsilon &\leq \inf_{u_i \in \mathcal{U}_i} J_i^N(u_i, u_{-i}^o) \\ &\leq J_i^N(u_i^o, u_{-i}^o) \quad (\text{a.s.}). \end{aligned}$$

\square

Let $z_i^{L,\infty}$ be the closed-loop solution of the i -th leader's dynamics (1) with the MF control input defined in (34), and $\psi^{L,\infty}(\cdot)$ be the infinite population centroid of the leaders (13), we denote

$$(\varepsilon_{N_L})^2 := \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left\| \psi^{L,\infty}(t) - \frac{1}{N_L} \sum_{i=1}^{N_L} z_i^{L,\infty}(t) \right\|^2 dt. \quad (38)$$

The proof of the following theorem is similar to the one of Theorem 6.1 in [15].

Theorem 6.1: Assume **(A1)**–**(A3-a)** hold. Then the set of the leaders' MF control laws is an *a.s.* $O(\varepsilon_{N_L})$ -Nash equilibrium, i.e., for any i , $1 \leq i \leq N_L$, we have

$$\begin{aligned} J_i^{L, N_L}(u_i^{L, \infty}, u_{-i}^{L, \infty}) - O(\varepsilon_{N_L}) &\leq \inf_{u_i \in \mathcal{U}_i^L} J_i^{L, N_L}(u_i, u_{-i}^{L, \infty}) \\ &\leq J_i^{L, N_L}(u_i^{L, \infty}, u_{-i}^{L, \infty}) \quad a.s. \end{aligned}$$

where $\lim_{N_L \rightarrow \infty} \varepsilon_{N_L} = 0$ (*a.s.*). \square

For any adaptive follower i , $1 \leq i \leq N_F$, let $u_{i, \delta_1}^{F, \infty}(\cdot)$ be the MF control law when the defining parameter of the reference trajectory is assumed to be the true parameter $\delta_1 \in \Delta$ and let $\hat{z}_i^{F, \infty}(\cdot)$ and $z_{i, \delta_1}^{F, \infty}(\cdot)$ be the closed-loop solutions of dynamics (4) with control input $\hat{u}_i^{F, \infty}(\cdot)$ and $u_{i, \delta_1}^{F, \infty}(\cdot)$, respectively (the explicit form of $\hat{z}_i^{F, \infty}(\cdot)$ is presented in **(A25)**). Let $u^{L, \infty} := (u_1^{L, \infty}, \dots, u_{N_L}^{L, \infty})$, $\hat{u}^{F, \infty} := (\hat{u}_1^{F, \infty}, \dots, \hat{u}_{N_F}^{F, \infty})$ and $u_{\delta_1}^{F, \infty} := (u_{1, \delta_1}^{F, \infty}, \dots, u_{N_F, \delta_1}^{F, \infty})$.

Lemma 6.1: Assume **(A1)**–**(A4)** hold. Then there exists a random M , $0 < M(\omega) < \infty$, such that for $N_L \geq M(\omega)$ and observation sizes $m_i \geq M(\omega)$, $1 \leq i \leq N_F$, we have

$$J_i^{F, N}(\hat{u}_i^{F, \infty}; \hat{u}_{-i}^{F, \infty}, u^{L, \infty}) \leq J_i^{F, N}(u_{i, \delta_1}^{F, \infty}; u_{-i, \delta_1}^{F, \infty}, u^{L, \infty}) \quad a.s.$$

for $1 \leq i \leq N_F$.

Proof: By Theorem 4.1 for each adaptive follower i , $1 \leq i \leq N_F$, there exist a deterministic time T_i and a random M_i , $0 < T_i$, $M_i(\omega) < \infty$, such that $\hat{\delta}_i^m(t) = \delta_1$ for all $t > T_i$ and $m > M_i(\omega)$. Let $T_F := \max\{T_1, \dots, T_{N_F}\} < \infty$ and $M := \max\{M_1, \dots, M_{N_F}\} < \infty$, then for $N_L \geq M(\omega)$ and observation sizes $m_i \geq M(\omega)$, $1 \leq i \leq N_F$, we have

$$\begin{aligned} &J_i^{F, N}(\hat{u}_i^{F, \infty}; \hat{u}_{-i}^{F, \infty}, u^{L, \infty}) \\ &\equiv \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\left\| \hat{z}_i^{F, \infty} - \left(\frac{\eta}{N_L} \sum_{j=1}^{N_L} z_j^{L, \infty} + \frac{(1-\eta)}{N_F} \sum_{j=1}^{N_F} \hat{z}_j^{F, \infty} \right) \right\|_Q^2 \right. \\ &\quad \left. + \|\hat{u}_i^{F, \infty}\|_R^2 \right) dt \\ &\leq \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^{T_F} \left(\left\| \hat{z}_i^{F, \infty} - \left(\frac{\eta}{N_L} \sum_{j=1}^{N_L} z_j^{L, \infty} + \frac{(1-\eta)}{N_F} \sum_{j=1}^{N_F} \hat{z}_j^{F, \infty} \right) \right\|_Q^2 \right. \\ &\quad \left. + \|\hat{u}_i^{F, \infty}\|_R^2 \right) dt \\ &\quad + \limsup_{T \rightarrow \infty} \frac{1}{T} \int_{T_F}^T \left(\left\| \hat{z}_i^{F, \infty} - \left(\frac{\eta}{N_L} \sum_{j=1}^{N_L} z_j^{L, \infty} + \frac{(1-\eta)}{N_F} \sum_{j=1}^{N_F} z_{j, \delta_1}^{F, \infty} \right) \right\|_Q^2 \right. \\ &\quad \left. + \|\hat{u}_i^{F, \infty}\|_R^2 \right) dt := I_1 + I_2. \end{aligned} \quad (39)$$

But, by the long run average stability of the MF control laws (37) there exists, with probability one, a real number k , $0 < k < \infty$, and independent of T such that

$$|I_1| \leq \limsup_{T \rightarrow \infty} \frac{kT_F}{T} = 0 \quad (a.s.). \quad (40)$$

Next consider $T' := T - T_F$, then

$$\begin{aligned} &J_i^{F, N}(\hat{u}_i^{F, \infty}; \hat{u}_{-i}^{F, \infty}, u^{L, \infty}) \\ &\leq \limsup_{T \rightarrow \infty} \frac{1}{T} \int_{T_F}^T \left(\left\| \hat{z}_i^{F, \infty} - \left(\frac{\eta}{N_L} \sum_{j=1}^{N_L} z_j^{L, \infty} + \frac{(1-\eta)}{N_F} \sum_{j=1}^{N_F} z_{j, \delta_1}^{F, \infty} \right) \right\|_Q^2 \right. \\ &\quad \left. + \|\hat{u}_i^{F, \infty}\|_R^2 \right) dt \end{aligned}$$

$$\begin{aligned} &= \limsup_{T' \rightarrow \infty} \frac{1}{T'} \int_0^{T'} \left(\left\| \hat{z}_i^{F, \infty} - \left(\frac{\eta}{N_L} \sum_{j=1}^{N_L} z_j^{L, \infty} + \frac{(1-\eta)}{N_F} \sum_{j=1}^{N_F} z_{j, \delta_1}^{F, \infty} \right) \right\|_Q^2 \right. \\ &\quad \left. + \|\hat{u}_i^{F, \infty}\|_R^2 \right) dt \equiv J_i^{F, N}(u_{i, \delta_1}^{F, \infty}; u_{-i, \delta_1}^{F, \infty}, u^{L, \infty}). \end{aligned}$$

by (39) and (40), which gives the result. \square

Denote

$$(\varepsilon_{N_F})^2 := \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left\| \psi_{\delta_1}^{F, \infty} - \frac{1}{N_F} \sum_{i=1}^{N_F} z_{i, \delta_1}^{F, \infty} \right\|^2 dt \quad (41)$$

where $z_{i, \delta_1}^{F, \infty}$ is the closed-loop solution of the i -th follower's dynamics (4) with the MF control law (35), and $\psi_{\delta_1}^{F, \infty}(\cdot)$ is the infinite population centroid of the followers (18) when the defining parameter of the leaders reference trajectory $h(\cdot)$ is assumed to be the true one, $\delta_1 \in \Delta$.

Lemma 6.2: Assume **(A1)**–**(A4)** hold. Then we have $\lim_{N_F \rightarrow \infty} \varepsilon_{N_F} = 0$ (*a.s.*).

Proof: We have

$$\begin{aligned} (\varepsilon_{N_F})^2 &\equiv \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left\| \psi_{\delta_1}^{F, \infty} - \frac{1}{N_F} \sum_{i=1}^{N_F} z_{i, \delta_1}^{F, \infty} \right\|^2 dt \\ &\leq 2 \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left\| \frac{1}{N_F} \sum_{i=1}^{N_F} \bar{z}_{i, \delta_1}^{F, \infty} - \frac{1}{N_F} \sum_{i=1}^{N_F} z_{i, \delta_1}^{F, \infty} \right\|^2 dt \\ &\quad + 2 \sup_{t \geq 0} \left\| \psi_{\delta_1}^{F, \infty}(t) - \frac{1}{N_F} \sum_{i=1}^{N_F} \bar{z}_{i, \delta_1}^{F, \infty}(t) \right\|^2. \end{aligned} \quad (42)$$

But, by Lemma 5.3 in [15] we have (*a.s.*)

$$\lim_{N_F \rightarrow \infty} \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left\| \frac{1}{N_F} \sum_{i=1}^{N_F} \left(\bar{z}_{i, \delta_1}^{F, \infty} - z_{i, \delta_1}^{F, \infty} \right) \right\|^2 dt = 0. \quad (43)$$

In addition

$$\begin{aligned} &\psi_{\delta_1}^{F, \infty}(\cdot) - \frac{1}{N_F} \sum_{i=1}^{N_F} \bar{z}_{i, \delta_1}^{F, \infty}(\cdot) \\ &= \int_{\Theta_F} \bar{z}_{\theta, \delta_1}^{F, \infty}(\cdot) dF^F(\theta) - \int_{\Theta_F} \bar{z}_{\theta, \delta_1}^{F, \infty}(\cdot) dF_{N_F}^F(\theta). \end{aligned}$$

So by the weak convergence of $F_{N_F}^F$ to F^F [Assumption **(A2)**] we have *a.s.* dF^F (see (33) in [15])

$$\lim_{N_F \rightarrow \infty} \sup_{t \geq 0} \left\| \psi_{\delta_1}^{F, \infty}(t) - \frac{1}{N_F} \sum_{i=1}^{N_F} \bar{z}_{i, \delta_1}^{F, \infty}(t) \right\| = 0. \quad (44)$$

Hence, by (42)–(44) we get $\lim_{N_F \rightarrow \infty} \varepsilon_{N_F} = 0$ (*a.s.*). \square

Theorem 6.2: Assume **(A1)**–**(A4)** hold. Then there exists a random M , $0 < M(\omega) < \infty$, such that for $N_L \geq M(\omega)$ and observation sizes $m_i \geq M(\omega)$, $1 \leq i \leq N_F$, the set of the followers' MF control laws is an *a.s.* ε_N -Nash equilibrium, i.e., for any i , $1 \leq i \leq N_F$, we have (*a.s.*)

$$\begin{aligned} &J_i^{F, N}(\hat{u}_i^{F, \infty}; \hat{u}_{-i}^{F, \infty}, u^{L, \infty}) - \varepsilon_N \\ &\leq \inf_{u_i \in \mathcal{U}_i^F} J_i^{F, N}(u_i; \hat{u}_{-i}^{F, \infty}, u^{L, \infty}) \\ &\leq J_i^{F, N}(\hat{u}_i^{F, \infty}; \hat{u}_{-i}^{F, \infty}, u^{L, \infty}) \end{aligned} \quad (45)$$

where $\varepsilon_N = O(\varepsilon_{N_L} + \varepsilon_{N_F})$ (*a.s.*).

Proof: The second inequality in (45) is trivial. Here, we shall prove the first inequality.

For any adaptive follower i , $1 \leq i \leq N$, by Lemma 6.1 there exists a random M , $0 < M(\omega) < \infty$, such that for $N_L \geq M(\omega)$ and observation sizes $m_i \geq M(\omega)$ we have

$$J_i^{F,N}(\hat{u}_i^{F,\infty}; \hat{u}_{-i}^{F,\infty}, u^{L,\infty}) \leq J_i^{F,N}(u_{i,\delta_1}^{F,\infty}; u_{-i,\delta_1}^{F,\infty}, u^{L,\infty}) \quad a.s. \quad (46)$$

But,

$$\begin{aligned} & J_i^{F,N}(u_{i,\delta_1}^{F,\infty}; u_{-i,\delta_1}^{F,\infty}, u^{L,\infty}) \\ & \equiv \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\left\| z_{i,\delta_1}^{F,\infty} - \left(\frac{\eta}{N_L} \sum_{j=1}^{N_L} z_j^{L,\infty} + \frac{(1-\eta)}{N_F} \sum_{j=1}^{N_F} z_{j,\delta_1}^{F,\infty} \right) \right\|_Q^2 \right. \\ & \quad \left. + \|u_{i,\delta_1}^{F,\infty}\|_R^2 \right) dt \\ & \leq \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\left\| z_{i,\delta_1}^{F,\infty} - \left(\eta \psi^{L,\infty} + (1-\eta) \psi_{\delta_1}^{L,\infty} \right) \right\|_Q^2 \right. \\ & \quad \left. + \|u_{i,\delta_1}^{F,\infty}\|_R^2 \right) dt + I_N \end{aligned} \quad (47)$$

where I_N denotes the rest of the terms. We now show that $I_N = O(\varepsilon_{N_L} + \varepsilon_{N_F})$. We have

$$\begin{aligned} I_N & := \limsup_{T \rightarrow \infty} \frac{\eta^2}{T} \int_0^T \left(\left\| \psi^{L,\infty} - \frac{1}{N_L} \sum_{i=1}^{N_L} z_i^{L,\infty} \right\|_Q^2 \right) dt \\ & + \limsup_{T \rightarrow \infty} \frac{(1-\eta)^2}{T} \int_0^T \left(\left\| \psi_{\delta_1}^{F,\infty} - \frac{1}{N_F} \sum_{i=1}^{N_F} z_{i,\delta_1}^{F,\infty} \right\|_Q^2 \right) dt \\ & + 2\eta(1-\eta) \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\psi^{L,\infty} - \frac{1}{N_L} \sum_{i=1}^{N_L} z_i^{L,\infty} \right)^T \\ & \quad \times Q \left(\psi_{\delta_1}^{F,\infty} - \frac{1}{N_F} \sum_{i=1}^{N_F} z_{i,\delta_1}^{F,\infty} \right) dt \\ & + 2\eta \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(z_{i,\delta_1}^{F,\infty} - \left(\eta \psi^{L,\infty} + (1-\eta) \psi_{\delta_1}^{L,\infty} \right) \right)^T \\ & \quad \times Q \left(\psi^{L,\infty} - \frac{1}{N_L} \sum_{i=1}^{N_L} z_i^{L,\infty} \right) dt \\ & + 2(1-\eta) \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(z_{i,\delta_1}^{F,\infty} - \left(\eta \psi^{L,\infty} + (1-\eta) \psi_{\delta_1}^{L,\infty} \right) \right)^T \\ & \quad \times Q \left(\psi_{\delta_1}^{F,\infty} - \frac{1}{N_F} \sum_{i=1}^{N_F} z_{i,\delta_1}^{F,\infty} \right) dt \\ & := I_{N_L}^1 + I_{N_F}^2 + I_{N_L, N_F}^3 + I_{N_L}^4 + I_{N_F}^5. \end{aligned} \quad (48)$$

But,

$$I_{N_L}^1 \leq \eta^2 \|Q\| \varepsilon_{N_L}^2, \quad I_{N_F}^2 \leq (1-\eta)^2 \|Q\| \varepsilon_{N_F}^2 \quad (49)$$

where ε_{N_L} and ε_{N_F} are defined in (38) and (41), respectively. By the Cauchy-Schwarz inequality we have

$$I_{N_L, N_F}^3 \leq 2\eta(1-\eta) \|Q\| \varepsilon_{N_L} \varepsilon_{N_F}. \quad (50)$$

By the stability property of the MF control laws (36) there exists a real number k , $0 < k < \infty$, independent of both N_L and N_F such that (*a.s.*)

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left\| z_{i,\delta_1}^{F,\infty} - \left(\eta \psi^{L,\infty} + (1-\eta) \psi_{\delta_1}^{L,\infty} \right) \right\|_Q^2 dt \leq k.$$

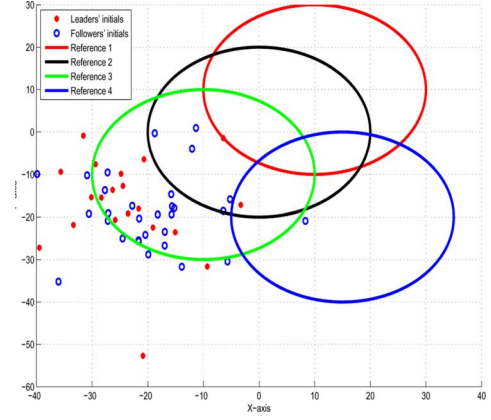


Fig. 1. Four possible reference trajectories of the leaders (reference one is the true reference trajectory) corresponding to parameters $\delta_1, \delta_2, \delta_3, \delta_4$ in the set Δ , respectively, and the initial states of leaders and followers.

This and the Cauchy-Schwarz inequality result in (*a.s.*)

$$I_{N_L}^4 \leq 2\eta \|Q\| \sqrt{k} \varepsilon_{N_L}, \quad I_{N_F}^5 \leq 2(1-\eta) \|Q\| \sqrt{k} \varepsilon_{N_F}. \quad (51)$$

Hence, (49)–(51) imply that $I_N = I_{N_L}^1 + I_{N_F}^2 + I_{N_L, N_F}^3 + I_{N_L}^4 + I_{N_F}^5 = O(\varepsilon_{N_L} + \varepsilon_{N_F})$ (*a.s.*).

But, by the construction of the MF system of equations for the followers (15)–(19), $u_{i,\delta_1}^{F,\infty}(\cdot)$ is the optimal tracking control with respect to $u^{L,\infty}(\cdot)$ and $u_{-i,\delta_1}^{F,\infty}(\cdot)$ which collectively generate $\psi^{L,\infty}(\cdot)$ and $\psi_{\delta_1}^{F,\infty}(\cdot)$. Therefore

$$\begin{aligned} & \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\left\| z_{i,\delta_1}^{F,\infty} - \left(\eta \psi^{L,\infty} + (1-\eta) \psi_{\delta_1}^{L,\infty} \right) \right\|_Q^2 + \|u_{i,\delta_1}^{F,\infty}\|_R^2 \right) dt \\ & \equiv J_i^{F,N}(u_{i,\delta_1}^{F,\infty}; \psi^{F,\infty}, \psi^{L,\infty}) \\ & \leq \inf_{u_i \in \mathcal{U}_i^F} J_i^{F,N}(u_i; \hat{u}_{-i}^{F,\infty}, u^{L,\infty}) + O(\varepsilon_{N_L} + \varepsilon_{N_F}) \quad a.s. \end{aligned} \quad (52)$$

where the last inequality follows from a similar argument as in (47)–(51). Hence, (46) and (52) imply the first inequality in (45). \square

VII. NUMERICAL EXAMPLE

Consider a system of 50 agents with 20 leaders and 30 followers. For simplicity and clarity of the simulation, we assume that one follower, called the *adaptive follower*, needs to estimate the true reference trajectory but this reference trajectory is fully known to all other followers and all leaders. It is important to note that the followers are estimating the true reference trajectory independently. If all followers are adaptive the overall computational load increases linearly with the population size of the followers which is a manageable complexity. For the adaptive follower, the possible set of reference trajectories have the general form $[a_1 + b_1 \cos(\omega t) \ a_2 + b_2 \sin(\omega t)]$, $t \in [0, 10]$, where $\delta = (a_1, b_1, a_2, b_2, \omega) \in \Delta$. We assume this set Δ has four parameters including the true parameter of the reference trajectory which is reference one in Fig. 1. All the four possible reference trajectories generated by each parameter of the set Δ and the initial states of all the leaders and followers are shown in Fig. 1.

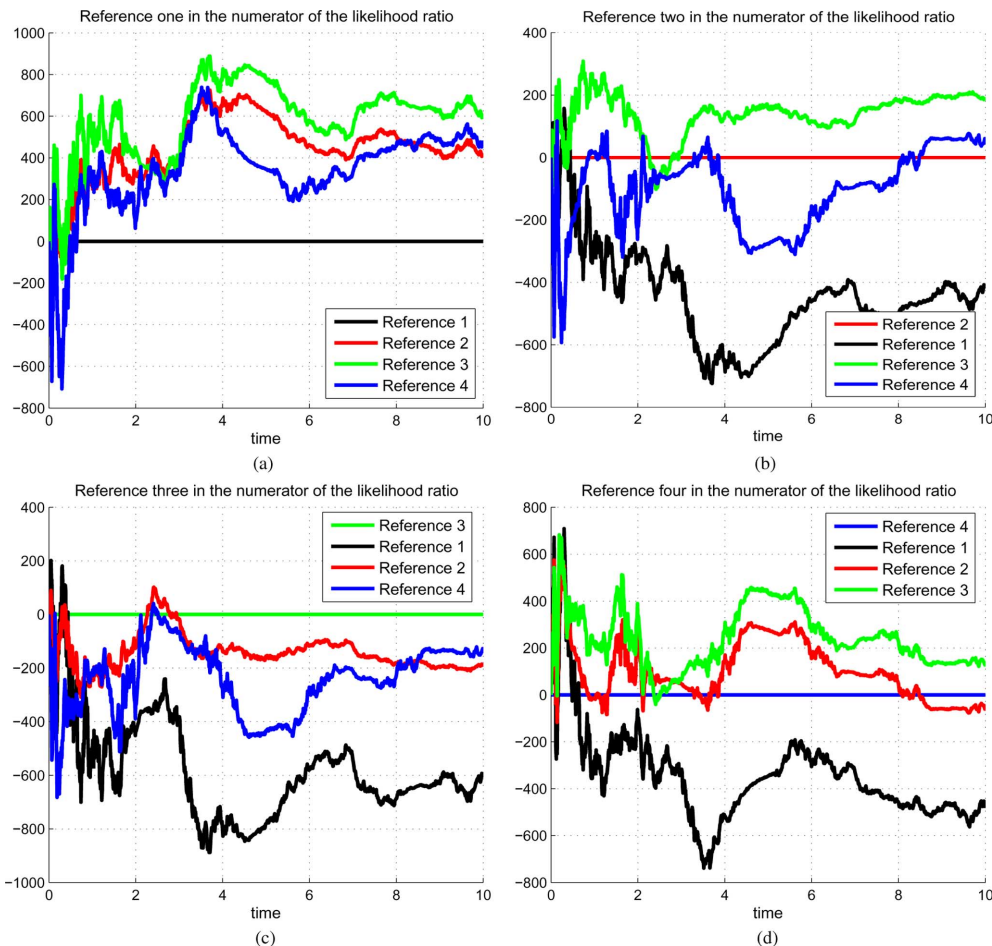


Fig. 2. $(1/t) \log$ of likelihood ratios $((1/t) \log(H_t^m(\delta_i)/H_t^m(\delta_j)))$, $\delta_i, \delta_j \in \Delta$ such that in each figure there is a fixed parameter of set Δ in the numerator of the ratio and the parameter in the denominator changes in the set Δ (a) $(1/t) \log(H_t^m(\delta_1)/H_t^m(\delta))$, $\delta \in \Delta$ (b) $(1/t) \log(H_t^m(\delta_2)/H_t^m(\delta))$, $\delta \in \Delta$ (c) $(1/t) \log(H_t^m(\delta_3)/H_t^m(\delta))$, $\delta \in \Delta$ (d) $(1/t) \log(H_t^m(\delta_4)/H_t^m(\delta))$, $\delta \in \Delta$.

The dynamics of leaders and followers are given in (1) and (4), respectively, where the leaders and followers' MF control laws are given in (34) and (35). In this simulation we have the following parameters: (i) In (1) and (4) matrices A_i of the leaders and followers are chosen randomly from a normal probability distribution around matrix $((0.2, -0.3)^T, (-0.4, 0.2)^T)$ with identity covariance, while matrices B_i of both leaders and followers are identity matrices, and the noise intensity matrices of both leaders and followers is $C = 4I$; (ii) In (3) and (6) let $Q = I$ and $R = 0.01I$; (iii) In (2) and (5) let $\lambda = 0.7$, $\eta = 0.6$ and the reference trajectory $h(\cdot)$ be reference one in Fig. 1; (iv) In (28) let the observation size of the adaptive follower be 15, and $D = 5I$; (v) In (31) let $t_0 = 0$, step-size $\tau = 1$ and the adaptive follower observes a non-empty subset of the leaders' trajectories of size 4.

It can be shown that this system satisfies the identifiability condition (A4). Fig. 2(a)–(d) correspond to $(1/t) \log(H_t^m(\delta_i)/H_t^m(\delta_j))$ for $\delta_i, \delta_j \in \Delta$ such that, for instance, in Fig. 2(a) the parameter $\delta_1 \in \Delta$ corresponding to reference trajectory one is in the numerator and the plots in this figure display $(1/t) \log(H_t^m(\delta_1)/H_t^m(\delta))$ for $\delta \in \Delta$. Based on the LRE defined in (31), Fig. 2(a)–(d) show that for the adaptive follower $\hat{\delta}^m(t) = \delta_1$ for all $t \geq 1$ since $H_t^m(\delta_1)/H_t^m(\delta) > 1$ (i.e., $\log(H_t^m(\delta_1)/H_t^m(\delta)) > 0$) for all $\delta \in \Delta$, $\delta \neq \delta_1$.

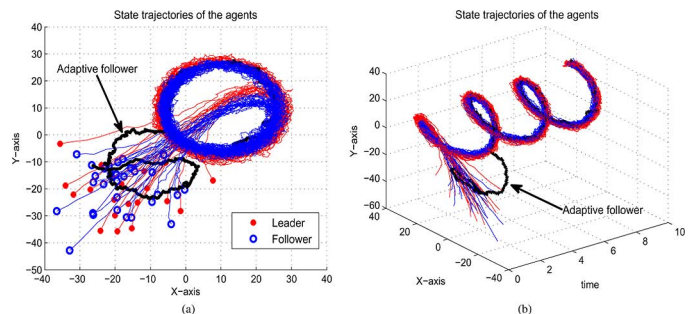


Fig. 3. State trajectories of leaders (red), non-adaptive followers (blue), and the adaptive follower (black) (a) 2-D and (b) 2-D in time.

In Fig. 3(a) and (b) the state trajectories of leaders, followers and the adaptive follower are shown. Based on $\hat{\delta}^m(\cdot)$ in (31) the adaptive follower initially considers the wrong reference trajectory signals until $t = 1$, reference 3 (generated by δ_3) from time zero to one, and then locks on the true reference trajectory (reference one generated by δ_1) as shown in Fig. 3(a) and (b).

VIII. CONCLUSION

In this paper, we have developed a linear-quadratic-Gaussian (LQG) dynamic game based model of collective dynamics which include leaders, followers and an unknown (to the followers) reference trajectory for the leaders. The Mean Field

(MF) (or Nash Certainty Equivalence (NCE)) LQG (MF LQG) equations characterizing the Nash equilibrium for infinite population systems were derived, and under appropriate conditions, they have a unique solution leading to decentralized control laws. Furthermore, for large but finite population systems, such controls were shown to correspond to so-called ε -Nash equilibria.

The computation of the followers' control laws requires knowledge of the complete reference trajectory which is in general not known to the followers but is estimated by a likelihood ratio based adaptation scheme based on noisy observations by followers of a random sample of leaders. Under appropriate identifiability conditions, it is established that this identification scheme is able to select the exact reference trajectory model within a finite class of candidates in a finite deterministic time almost surely as the number of samples goes to infinity. As a result, the estimation based adaptive mean field control laws of the followers together with the mean field control laws of the leaders give rise to a dynamic stochastic Nash equilibrium for the overall leader-follower system.

APPENDIX

Proof of Proposition 4.1: Here, we shall prove part (a). Part (b) follows directly from part (a)

The terms dy^M and $z_{\delta}^{L,M}$: The observation process (28) of a generic adaptive follower with the observation subset $O := \{L_1, \dots, L_M\} \subset L$ of cardinality M is given by

$$\begin{aligned} dy^M &= \left(\frac{1}{M} \sum_{i=1}^M z_{i,\delta_1}^L \right) dt + \frac{1}{M} \sum_{i=1}^M D_i dv_i \\ &\equiv \left(\frac{1}{M} \sum_{i=1}^M \bar{z}_{i,\delta_1}^L + \frac{1}{M} \sum_{i=1}^M \tilde{z}_{i,\delta_1}^L \right) dt + \frac{1}{M} \sum_{i=1}^M D_i dv_i \end{aligned} \quad (\text{A1})$$

where $\tilde{z}_{i,\delta_1}^L(\cdot) := z_{i,\delta_1}^L(\cdot) - \bar{z}_{i,\delta_1}^L(\cdot)$ in which $z_{i,\delta_1}^L(\cdot)$ and $\bar{z}_{i,\delta_1}^L(\cdot)$ are respectively the state trajectory and its expected value of the L_i -th leader, $1 \leq i \leq M$, where the defining parameter of $h(\cdot)$ is assumed to be $\delta_1 \in \Delta$. For each parameter $\delta \in \Delta$, the closed-loop solution of the L_i -th leader's dynamics, $1 \leq i \leq M$, in (1) with dynamical parameter $\theta_i = [A_i, B_i, C_i]$ is

$$\begin{aligned} z_{i,\delta}^L(t) &= e^{\Gamma_i t} z_i^L(0) - \int_0^t e^{\Gamma_i(t-\tau)} B_i R^{-1} B_i^T s_{i,\delta}^L(\tau) d\tau \\ &\quad + \int_0^t e^{\Gamma_i(t-\tau)} C_i dw_i^L \end{aligned}$$

where $s_{i,\delta}^L(\cdot)$ is the solution of the leaders' offset tracking (11) given in (20) in which the defining parameter of $h(\cdot)$ is assumed to be $\delta \in \Delta$. Furthermore, the expected value of the corresponding closed-loop solution of the L_i -th leader, $1 \leq i \leq M$, as given in (21) is

$$\bar{z}_{i,\delta}^L(t) = e^{\Gamma_i t} \bar{z}^L(0) - \int_0^t e^{\Gamma_i(t-\tau)} B_i R^{-1} B_i^T s_{i,\delta}^L(\tau) d\tau.$$

Therefore, we have $\tilde{z}_{i,\delta_1}^L = z_{i,\delta_1}^L - \bar{z}_{i,\delta_1}^L = e^{\Gamma_i t} (z_i^L(0) - \bar{z}^L(0)) + \int_0^t e^{\Gamma_i(t-\tau)} C_i dw_i^L$. Hence, for any $\delta \in \Delta$ we may write

$$z_{\delta}^{L,M}(t) := \frac{1}{M} \sum_{i=1}^M z_{i,\delta}^L(t)$$

$$\begin{aligned} &= \psi_{\delta}^{L,\infty}(t) + (\bar{z}_{\delta}^{L,M}(t) - \psi_{\delta}^{L,\infty}(t)) \\ &\quad + \frac{1}{M} \sum_{i=1}^M e^{\Gamma_i t} (z_i^L(0) - \bar{z}^L(0)) \\ &\quad + \frac{1}{M} \sum_{i=1}^M \int_0^t e^{\Gamma_i(t-\tau)} C_i dw_i^L \end{aligned} \quad (\text{A2})$$

where $\psi_{\delta}^{L,\infty}(\cdot)$ is the deterministic infinite population leaders' centroid computed from the leaders' MF system of equations, (11)–(14), when the defining parameter of $h(\cdot)$ is assumed to be $\delta \in \Delta$. In a similar way, we may write (A1) as

$$\begin{aligned} dy^M &= \psi_{\delta_1}^{L,\infty} dt + (\bar{z}_{\delta_1}^{L,M} - \psi_{\delta_1}^{L,\infty}) dt \\ &\quad + \left(\frac{1}{M} \sum_{i=1}^M e^{\Gamma_i t} (z_i^L(0) - \bar{z}^L(0)) \right. \\ &\quad \left. + \frac{1}{M} \sum_{i=1}^M \int_0^t e^{\Gamma_i(t-\tau)} C_i dw_i^L \right) dt + \frac{1}{M} \sum_{i=1}^M D_i dv_i \end{aligned} \quad (\text{A3})$$

where $\bar{z}_{\delta_1}^{L,M}(\cdot) := 1/M \sum_{i=1}^M \bar{z}_{i,\delta_1}^L(\cdot)$. The term $\int_0^t z_{\delta,s}^{L,M})^T dy_s^M$: By (A2) and (A3) we have

$$\begin{aligned} &\int_0^t (z_{\delta,s}^{L,M})^T dy_s^M \\ &\equiv \int_0^t \left(\psi_{\delta,s}^{L,\infty} + (\bar{z}_{\delta,s}^{L,M} - \psi_{\delta,s}^{L,\infty}) \right. \\ &\quad \left. + \frac{1}{M} \sum_{i=1}^M e^{\Gamma_i s} (z_i^L(0) - \bar{z}^L(0)) \right. \\ &\quad \left. + \frac{1}{M} \sum_{i=1}^M \int_0^s e^{\Gamma_i(s-\tau)} C_i dw_i^L \right)^T \\ &\quad \times \left(\psi_{\delta_1,s}^{L,\infty} ds + (\bar{z}_{\delta_1,s}^{L,M} - \psi_{\delta_1,s}^{L,\infty}) ds \right. \\ &\quad \left. + \left(\frac{1}{M} \sum_{i=1}^M e^{\Gamma_i s} (z_i^L(0) - \bar{z}^L(0)) \right. \right. \\ &\quad \left. \left. + \frac{1}{M} \sum_{i=1}^M \int_0^s e^{\Gamma_i(s-\tau)} C_i dw_i^L \right) ds \right. \\ &\quad \left. + \frac{1}{M} \sum_{i=1}^M D_i dv_i \right) \end{aligned} \quad (\text{A4})$$

where $z_{\delta,t}^{L,M} := z_{\delta}^{L,M}(t)$, $\bar{z}_{\delta,t}^{L,M} := \bar{z}_{\delta}^{L,M}(t)$, $y_t^M := y^M(t)$ and $\psi_{\delta,t}^{L,\infty} := \psi_{\delta}^{L,\infty}(t)$.

In the rest of the proof take fixed $\delta \in \Delta$ and $0 < T < \infty$. Let

$$\begin{aligned} \varepsilon_{M,\delta} &:= \sup_{t \geq 0} \|\bar{z}_{\delta}^{L,M}(t) - \psi_{\delta}^{L,\infty}(t)\| \\ &= \sup_{t \geq 0} \left\| \int_{\Theta_L} \bar{z}_{\delta,\theta}^L(t) dF_M^L(\theta) - \int_{\Theta_L} \bar{z}_{\delta,\theta}^L(t) dF^L(\theta) \right\| \end{aligned}$$

where $F_M^L(x)$ is the empirical distribution associated with the leaders. By (A2) and similar to (33) in [15] we have $\lim_{M \rightarrow \infty} \varepsilon_{M,\delta} = 0$ (a.s. dF^L). Therefore,

$$\int_0^T \|\bar{z}_{\delta,s}^{L,M} - \psi_{\delta,s}^{L,\infty}\|^2 ds \leq T \varepsilon_{M,\delta}^2 \rightarrow 0 \quad \text{a.s. } dF^L \quad (\text{A5})$$

as M goes to infinity. Hence, (A5) and the Cauchy-Schwarz inequality imply that

$$\begin{aligned} & \lim_{M \rightarrow \infty} \left| \int_0^T (\bar{z}_{\delta,s}^{L,M} - \psi_{\delta,s}^{L,\infty})^T (\bar{z}_{\delta_1,s}^{L,M} - \psi_{\delta_1,s}^{L,\infty}) ds \right| \\ & \leq \lim_{M \rightarrow \infty} \left(\int_0^T \|\bar{z}_{\delta,s}^{L,M} - \psi_{\delta,s}^{L,\infty}\|^2 ds \right)^{1/2} \times \\ & \left(\int_0^T \|\bar{z}_{\delta_1,s}^{L,M} - \psi_{\delta_1,s}^{L,\infty}\|^2 ds \right)^{1/2} = \lim_{M \rightarrow \infty} T \varepsilon_{M,\delta} \varepsilon_{M,\delta_1} = 0. \end{aligned} \quad (\text{A6})$$

Since for any $\delta \in \Delta$, $\psi_{\delta}^{L,\infty}(\cdot) \in C_n^b$, there exists a real number k , $0 < k < \infty$ independent of M such that

$$\int_0^T \|\psi_{\delta,s}^{L,\infty}\|^2 \leq kT. \quad (\text{A7})$$

By (A5), (A7) and the Cauchy-Schwarz inequality we have both (a.s.)

$$\begin{aligned} & \lim_{M \rightarrow \infty} \left| \int_0^T (\psi_{\delta,s}^{L,\infty})^T (\bar{z}_{\delta_1,s}^{L,M} - \psi_{\delta_1,s}^{L,\infty}) ds \right| \\ & \leq \lim_{M \rightarrow \infty} T \sqrt{k} \varepsilon_{M,\delta_1} = 0 \end{aligned} \quad (\text{A8})$$

and

$$\lim_{M \rightarrow \infty} \left| \int_0^T (\bar{z}_{\delta,s}^{L,M} - \psi_{\delta,s}^{L,\infty})^T \psi_{\delta_1,s}^{L,\infty} ds \right| \leq \lim_{M \rightarrow \infty} T \sqrt{k} \varepsilon_{M,\delta} = 0 \quad (\text{A9})$$

where k is a fixed real number independent of M given in (A7).

Analysis of disturbance terms via SLLN: By the Strong Law of Large Numbers (SLLN) [36] we have

$$\begin{aligned} & \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M \int_0^T e^{\Gamma_i(T-\tau)} C_i dw_i^L = 0 \quad a.s. \\ & = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M \int_0^T D_i dv_i = 0 \quad a.s. \\ & = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M e^{\Gamma_i T} (z_i^L(0) - \bar{z}^L(0)) = 0 \quad a.s. \end{aligned} \quad (\text{A10})$$

and

$$\lim_{M \rightarrow \infty} \int_0^T (\psi_{\delta,s}^{L,\infty})^T \left(\frac{1}{M} \sum_{i=1}^M D_i dv_i \right) = 0 \quad a.s. \quad (\text{A11})$$

since for any $\delta \in \Delta$, $\psi_{\delta}^{L,\infty}(\cdot) \in C_n^b$. Similarly, we have (a.s.)

$$\lim_{M \rightarrow \infty} \int_0^T (\bar{z}_{\delta,s}^{L,M} - \psi_{\delta,s}^{L,\infty})^T \left(\frac{1}{M} \sum_{i=1}^M D_i dv_i \right) = 0 \quad (\text{A12})$$

$$\lim_{M \rightarrow \infty} \int_0^T \left(\frac{1}{M} \sum_{i=1}^M e^{\Gamma_i s} (z_i^L(0) - \bar{z}^L(0)) \right)^T \left(\frac{1}{M} \sum_{i=1}^M D_i dv_i \right) = 0 \quad (\text{A13})$$

$$\lim_{M \rightarrow \infty} \int_0^T \left(\frac{1}{M} \sum_{i=1}^M \int_0^s e^{\Gamma_i(s-\tau)} C_i dw_i^L \right)^T \left(\frac{1}{M} \sum_{i=1}^M D_i dv_i \right) = 0. \quad (\text{A14})$$

Analysis of influence of initial conditions: Since for $t \geq 0$

$$\begin{aligned} & \left\| \frac{1}{M} \sum_{i=1}^M e^{\Gamma_i t} (z_i^L(0) - \bar{z}^L(0)) \right\| \\ & \leq \gamma e^{-\rho t} \sup_{\theta_i \in \Theta_L} \|z_i^L(0) - \bar{z}^L(0)\| < \infty, \end{aligned} \quad (\text{A15})$$

with the right hand side independent of M , and where positive γ and ρ are given in Remark 3.1, Lebesgue dominated convergence theorem [36] implies that

$$\begin{aligned} & \lim_{M \rightarrow \infty} \left(\int_0^T \left\| \frac{1}{M} \sum_{i=1}^M e^{\Gamma_i s} (z_i^L(0) - \bar{z}^L(0)) \right\|^2 ds \right) \\ & = \int_0^T \lim_{M \rightarrow \infty} \left\| \frac{1}{M} \sum_{i=1}^M e^{\Gamma_i s} (z_i^L(0) - \bar{z}^L(0)) \right\|^2 ds = 0 \quad a.s. \end{aligned} \quad (\text{A16})$$

where the last equality is obtained by (A10). Again, by the Cauchy-Schwarz inequality we have (a.s.)

$$\lim_{M \rightarrow \infty} \int_0^T \left(\frac{1}{M} \sum_{i=1}^M e^{\Gamma_i s} (z_i^L(0) - \bar{z}^L(0)) \right)^T \psi_{\delta,s}^{L,\infty} ds = 0 \quad (\text{A17})$$

by (A7) and (A16); and (a.s.)

$$\begin{aligned} & \lim_{M \rightarrow \infty} \int_0^T \left(\frac{1}{M} \sum_{i=1}^M e^{\Gamma_i s} (z_i^L(0) - \bar{z}^L(0)) \right)^T \\ & \times \left(\bar{z}_{\delta,s}^{L,M} - \psi_{\delta,s}^{L,\infty} \right) ds = 0 \end{aligned} \quad (\text{A18})$$

by (A5) and (A16). By the same argument as in proving Theorem 5.1 in [15] (see (31) in [15]) we get

$$\lim_{M \rightarrow \infty} \int_0^T \left\| \frac{1}{M} \sum_{i=1}^M \int_0^s e^{\Gamma_i(s-\tau)} C_i dw_i^L \right\|^2 ds = 0 \quad (a.s.). \quad (\text{A19})$$

Again by the Cauchy-Schwarz inequality we have (a.s.)

$$\lim_{M \rightarrow \infty} \int_0^T (\psi_{\delta,s}^{L,\infty})^T \left(\frac{1}{M} \sum_{i=1}^M \int_0^s e^{\Gamma_i(s-\tau)} C_i dw_i^L \right) ds = 0 \quad (\text{A20})$$

by (A7) and (A19),

$$\begin{aligned} & \lim_{M \rightarrow \infty} \int_0^T (\bar{z}_{\delta,s}^{L,M} - \psi_{\delta,s}^{L,\infty})^T \\ & \times \left(\frac{1}{M} \sum_{i=1}^M \int_0^s e^{\Gamma_i(s-\tau)} C_i dw_i^L \right) ds = 0 \end{aligned} \quad (\text{A21})$$

by (A5) and (A19), and

$$\begin{aligned} & \lim_{M \rightarrow \infty} \int_0^T \left(\frac{1}{M} \sum_{i=1}^M e^{\Gamma_i s} (z_i^L(0) - \bar{z}^L(0)) \right)^T \\ & \times \left(\frac{1}{M} \sum_{i=1}^M \int_0^s e^{\Gamma_i(s-\tau)} C_i dw_i^L \right) ds = 0 \end{aligned} \quad (\text{A22})$$

by (A16) and (A19).

Conclusion of the asymptotic analysis. By (A6), (A8)–(A9), (A11)–(A14), (A16)–(A18) and (A19)–(A22) we obtain (*a.s.*)

$$\begin{aligned} \lim_{M \rightarrow \infty} \exp \left(\int_0^T (z_{\delta,s}^{L,M})^T dy_s^M \right) \\ = \exp \left(\int_0^T (\psi_{\delta,s}^{L,\infty})^T \psi_{\delta_1,s}^{L,\infty} ds \right) \end{aligned} \quad (\text{A23})$$

from (A4) for any fixed $\delta \in \Delta$ and $T, 0 < T < \infty$, and in an analogous way one can show that (*a.s.*)

$$\lim_{M \rightarrow \infty} \exp \left(\int_0^T \|z_{\delta,s}^{L,M}\|^2 ds \right) = \exp \left(\int_0^T \|\psi_{\delta,s}^{L,\infty}\|^2 ds \right). \quad (\text{A24})$$

Finally, (A23) and (A24) imply that for any fixed $\delta \in \Delta$, and $t, 0 < t < \infty$, we have (*a.s.*)

$$\begin{aligned} \lim_{M \rightarrow \infty} L_t^M(\delta) \\ \equiv \lim_{M \rightarrow \infty} \exp \left(\int_0^t (z_{\delta,s}^{L,M})^T dy_s^M - \frac{1}{2} \int_0^t \|z_{\delta,s}^{L,M}\|^2 ds \right) \\ = L_t^\infty(\delta) \equiv \exp \left(\int_0^t (\psi_{\delta,s}^{L,\infty})^T \psi_{\delta_1,s}^{L,\infty} ds - \frac{1}{2} \int_0^t \|\psi_{\delta,s}^{L,\infty}\|^2 ds \right). \end{aligned}$$

□

Proof of Theorem 5.1: Part (a) is a special case of Theorem 4.1 in [15]. Here, we broadly follow the same approach to prove Part (b). For an adaptive follower $i, 1 \leq i \leq N_F$, with dynamical parameter $\theta_i = [A_i, B_i, C_i] \in \Theta_L$ by application of the adaptive MF control law (35) we have the closed-loop solution

$$\begin{aligned} \hat{z}_i^{F,\infty}(t) = e^{\Gamma_i t} z_i^F(0) - \int_0^t e^{\Gamma_i(t-\tau)} B_i R^{-1} B_i^T \hat{s}_i^{F,\infty}(\tau) d\tau \\ + \int_0^t e^{\Gamma_i(t-\tau)} C_i dw_i^F(\tau), \quad t \geq 0 \end{aligned} \quad (\text{A25})$$

where $\hat{s}_i^{F,\infty}(\cdot)$ is the solution of the tracking offset (15) given by

$$\begin{aligned} \hat{s}_i^{F,\infty}(t) \\ = - \int_t^\infty e^{-\Gamma_i^T(t-\tau)} Q \left(\eta \psi_{\delta}^{L,\infty}(\tau) + (1 - \eta) \psi_{\delta}^{F,\infty}(\tau) \right) d\tau. \end{aligned} \quad (\text{A26})$$

Denote $k_L := \sup_{\delta \in \Delta} \|\psi_{\delta}^{L,\infty}\|_\infty$, $k_F := \sup_{\delta \in \Delta} \|\psi_{\delta}^{F,\infty}\|_\infty$, and $k' = \max(k_L, k_F)$ then by Assumption (A3) we have $k' < \infty$. Subsequently, from (A26), we get $\|\hat{s}_i^{F,\infty}\|_\infty \leq \gamma \|Q\| k' / \rho =: k_s < \infty$, where positive γ and ρ are given in Remark 3.1, and hence, $\|\int_0^t e^{\Gamma_i(t-\tau)} B_i R^{-1} B_i^T \hat{s}_i^{F,\infty}(\tau) d\tau\| \leq \gamma k_s \|R^{-1}\| M_B^2 / \rho =: k_1 < \infty$, where $M_B := \sup_{\theta_i \in \Theta_F} \|B_i\| < \infty$ since Θ_F is compact. Therefore, we have (*a.s.*)

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left\| \int_0^t e^{\Gamma_i(t-\tau)} B_i R^{-1} B_i^T \hat{s}_i^{F,\infty}(\tau) d\tau \right\| dt \leq k_1 \quad (\text{A27})$$

and

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left\| \int_0^t e^{\Gamma_i(t-\tau)} B_i R^{-1} B_i^T \hat{s}_i^{F,\infty}(\tau) d\tau \right\|^2 dt \leq k_1^2. \quad (\text{A28})$$

Since Γ_i is asymptotically stable (Remark 3.1) we have

$$\begin{aligned} \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \|e^{\Gamma_i t} z_i^F(0)\| dt \\ = \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \|e^{\Gamma_i t} z_i^F(0)\|^2 dt = 0 \quad (\text{a.s.}) \end{aligned} \quad (\text{A29})$$

By the same argument as in proving Theorem 4.1 in [15] we get (*a.s.*)

$$\begin{aligned} \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left\| \int_0^t e^{\Gamma_i(t-\tau)} C_i dw_i^F \right\|^2 dt \\ = \int_0^\infty \text{tr}(e^{\Gamma_i t} C_i C_i^T e^{\Gamma_i^T t}) dt \\ \leq \frac{\gamma^2}{2\rho} \sup_{\theta_i \in \Theta_F} \|C_i\|^2 =: k_2. \end{aligned} \quad (\text{A30})$$

Thus, it follows from (A27)–(A30) that (*a.s.*)

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \|\hat{z}_i^{F,\infty}(t)\|^2 dt \leq k_1^2 + k_2 =: k < \infty. \quad (\text{A31})$$

Since $\hat{u}_i^{F,\infty}(\cdot) = -R^{-1} B_i^T \left(\Pi_i \hat{z}_i^{F,\infty}(\cdot) + \hat{s}_i^{F,\infty}(\cdot) \right)$ we have

$$\begin{aligned} \limsup_{T \rightarrow \infty} \frac{1}{T} \int_0^T \|\hat{u}_i^{F,\infty}(t)\|^2 dt \\ \leq \|R^{-1}\|^2 M_B^2 (M_P^2 k + k_s^2 + 2M_P k k_s) < \infty \quad \text{a.s.} \end{aligned} \quad (\text{A32})$$

where $M_P := \sup_{\theta_i \in \Theta_F} \|\Pi_i\| < \infty$ since Θ_F is compact. Since k, k_s, M_B and M_P are independent of i and N_F , by (A31) and (A32) we obtain (37). □

REFERENCES

- [1] Z. Lin, M. Broucke, and B. Francis, "Local control strategies for groups of mobile autonomous agents," *IEEE Trans. Autom. Control*, vol. 49, no. 4, pp. 622–629, Apr. 2004.
- [2] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1465–1476, Sep. 2004.
- [3] H. Yin, P. G. Mehta, S. P. Meyn, and U. V. Shanbhag, "Synchronization of coupled oscillators is a game," *IEEE Trans. Autom. Control*, vol. 57, no. 4, pp. 920–935, Apr. 2012.
- [4] L. Perea, G. Gómez, and P. Elosegui, "Extension of the Cucker-Smale control law to space flight formations," *J. Guid., Control, Dyn.*, vol. 32, no. 2, pp. 527–537, 2009.
- [5] V. Lambson, "Self-enforcing collusion in large dynamic markets," *J. Econ. Theory*, vol. 34, no. 2, pp. 282–291, Dec. 1984.
- [6] M. Huang, P. E. Caines, and R. P. Malhamé, "Individual and mass behaviour in large population stochastic wireless power control problems: Centralized and Nash equilibrium solutions," in *Proc. 42nd IEEE CDC*, Maui, HI, Dec. 2003, pp. 98–103.
- [7] M. Huang, P. E. Caines, and R. P. Malhamé, "Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ϵ -Nash equilibria," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1560–1571, Sep. 2007.
- [8] 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," *Comm. Inf. Syst.*, vol. 6, no. 3, pp. 221–252, 2006.

- [9] M. Huang, P. E. Caines, and R. P. Malhamé, "An invariance principle in large population stochastic dynamic games," *J. Syst. Sci. Complex.*, vol. 20, no. 2, pp. 162–172, 2007.
- [10] P. E. Caines, "Bode lecture: Mean field stochastic control," in *Proc. 48th IEEE CDC*, Shanghai, China, Dec. 2009.
- [11] J.-M. Lasry and P.-L. Lions, "Mean field games," *Jpn. J. Math.*, vol. 2, no. 1, pp. 229–260, 2007.
- [12] J.-M. Lasry and P.-L. Lions, "Jeux à champ moyen. I—Le cas stationnaire," *C. Rendus Math.*, vol. 343, no. 9, pp. 619–625, 2006.
- [13] J.-M. Lasry and P.-L. Lions, "Jeux à champ moyen. II—Horizon fini et contrôle optimal," *C. Rendus Math.*, vol. 343, no. 10, pp. 679–684, 2006.
- [14] G. Y. Weintraub, C. L. Benkard, and B. Van Roy, "Oblivious equilibrium: A mean field approximation for large-scale dynamic games," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2005.
- [15] T. Li and J.-F. Zhang, "Asymptotically optimal decentralized control for large population stochastic multiagent systems," *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1643–1660, Jul. 2008.
- [16] J. R. G. Dyer, A. Johansson, D. Helbing, I. D. Couzin, and J. Krause, "Leadership, consensus decision making and collective behaviour in humans," *Phil. Tran. Royal Soc. B: Bio. Sci.*, vol. 364, no. 1518, pp. 781–789, 2009.
- [17] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin, "Effective leadership and decision-making in animal groups on the move," *Nature*, vol. 433, no. 7025, pp. 513–516, Feb. 2005.
- [18] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, "Hierarchical group dynamics in pigeon flocks," *Nature*, vol. 464, no. 7290, pp. 890–893, 2010.
- [19] M. Simaan and J. B. Cruz, "On the Stackelberg strategy in nonzero-sum games," *J. Opt. Theory Appl.*, vol. 11, no. 5, pp. 533–555, 1973.
- [20] W. Wang and J. E. Slotine, "A theoretical study of different leader roles in networks," *IEEE Trans. Autom. Control*, vol. 51, no. 7, pp. 1156–1161, Jul. 2006.
- [21] F. Aubé and R. Shield, "Modeling the effect of leadership on crowd flow dynamics," *Cel. Aut.*, vol. 3305, pp. 601–611, 2004.
- [22] M. Nourian, R. P. Malhamé, M. Huang, and P. E. Caines, "Mean field (NCE) formulation of estimation based leader-follower collective dynamics," *Int. J. Robot. Automat.*, vol. 26, no. 1, pp. 120–129, 2011.
- [23] M. Nourian, R. P. Malhamé, M. Huang, and P. E. Caines, "Optimality of adaptation based mean field control laws in leader-follower stochastic collective dynamics," in *Proc. 49th IEEE CDC*, Atlanta, GA, Dec. 2010, pp. 2270–2275.
- [24] W. Ren, R. W. Beard, and E. M. Atkins, "A survey of consensus problems in multi-agent coordination," in *Proc. ACC*, Portland, OR, Jun. 2005, pp. 1859–1864.
- [25] W. Ren, "Multi-vehicle consensus with a time-varying reference state," *Syst. Control Lett.*, vol. 56, no. 7–8, pp. 474–483, 2007.
- [26] D. Acemoglu and M. Jensen, "Robust comparative statics in large static games," in *Proc. 49th IEEE CDC*, Atlanta, GA, Dec. 2010, pp. 3133–3139.
- [27] D. Gu and Z. Wang, "Leader-follower flocking: Algorithms and experiments," *IEEE Trans. Control Syst. Technol.*, vol. 17, no. 5, pp. 1211–1219, 2009.
- [28] H. Liu, J. Shan, and D. Sun, "Adaptive synchronization control of multiple spacecraft formation flying," *J. Dyn. Syst., Meas., Control*, vol. 129, pp. 337–337, 2007.
- [29] B. Düring, P. Markowich, J. Pietschmann, and M. Wolfram, "Boltzmann and Fokker-Planck equations modelling opinion formation in the presence of strong leaders," *Proc. Royal Soc. A*, vol. 465, no. 2112, pp. 3687–3707, 2009.
- [30] D. Delchamps, "Analytic feedback control and the algebraic Riccati equation," *IEEE Trans. Autom. Control*, vol. 29, no. 11, pp. 1031–1033, Nov. 1984.
- [31] P. E. Caines, *Linear Stochastic Systems*. New York: Wiley, 1987.
- [32] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [33] P. E. Caines, "A note on the consistency of maximum likelihood estimates for finite families of stochastic processes," *Annal. Stat.*, vol. 3, no. 2, pp. 539–546, 1975.
- [34] T. Duncan, "Evaluation of likelihood functions," *Inform. Control*, vol. 13, no. 1, pp. 62–74, Jul. 1968.
- [35] T. Duncan, "Likelihood functions for stochastic signals in white noise," *Inform. Control*, vol. 16, no. 4, pp. 303–310, 1970.
- [36] K. L. Chung, *A Course in Probability Theory*. San Diego, CA: Harcourt, Brace and World, Inc., 1968.



Mojtaba Nourian (S'06) received dual B.Sc. degrees in applied mathematics and in electrical engineering and the M.Sc. degree in applied mathematics from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2005 and 2007, respectively, and is currently pursuing the Ph.D. degree in electrical engineering in the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada.

His current research interests include mean field stochastic control theory and its applications.



Peter E. Caines (M'74–SM'83–F'86–LF'11) received the B.A. degree in mathematics from Oxford University, Oxford, U.K., in 1967 and the Ph.D. degree in systems and control theory from Imperial College, University of London, London, U.K., in 1970.

He was a Postdoctoral Researcher and Faculty Member at the University of Manchester Institute of Science and Technology, Manchester, U.K., Stanford University, Stanford, CA, University of California, Berkeley, University of Toronto, Toronto,

ON Canada, and Harvard University, Cambridge, MA. He joined McGill University, Montreal, QC Canada, in 1980, where he is the James McGill Professor and Macdonald Chair. He is the author of *Linear Stochastic Systems* (Wiley: 1988) and his research interests include stochastic, mean field and hybrid systems theory.

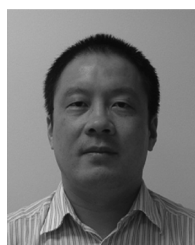
Dr. Caines is a Fellow of SIAM, IMA and the CIAR, and was elected to the Royal Society of Canada in 2003. In 2009, he received the IEEE Control Systems Society Bode Lecture Prize. In 2000, the adaptive control paper he coauthored in the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, 1980, was recognized by the IEEE Control Systems Society as one of the 25 seminal control theory papers of the 20th century.



Roland P. Malhamé (S'82–M'92) received the B.S. degree from the American University of Beirut, Beirut, Lebanon, in 1976, the M.S. degree from the University of Houston, Houston, TX, in 1978, and the Ph.D. degree from the Georgia Institute of Technology, Atlanta, in 1983, all in electrical engineering.

After single-year stays at the University of Quebec, Chicoutimi, Canada, and CAE Electronics Ltd, Montreal, Canada, he joined the École Polytechnique de Montréal, Montreal, in 1985, where he is Professor of Electrical Engineering. From June 2005 to April

2011, he was Director of the Groupe d'Études et de Recherche en Analyse des Décisions, GERAD, Montreal, QC, Canada. His current research interests are in the analysis of complex systems and the development of biology inspired schemes for their decentralized control, with applications in manufacturing, communications, smart grids, and collective dynamics. Past contributions include a statistical mechanics inspired approach to electric load modeling.



Minyi Huang (S'01–M'04) received the B.Sc. degree from Shandong University, Jinan, Shandong, China, in 1995, the M.Sc. degree from the Institute of Systems Science, Chinese Academy of Sciences, Beijing, China, in 1998, and the Ph.D. degree from the Department of Electrical and Computer Engineering, McGill University, Montreal, QC Canada, in 2003, all in systems and control.

He was a Research Fellow first in the Department of Electrical and Electronic Engineering, the University of Melbourne, Melbourne, Australia from February 2004 to March 2006, and then in the Department of Information Engineering, Research School of Information Sciences and Engineering, the Australian National University, Canberra, Australia from April 2006 to June 2007. He joined the School of Mathematics and Statistics, Carleton University, Ottawa, ON Canada as an Assistant Professor in July 2007, where he has been an Associate Professor since July 2011. His research interests include mean field stochastic control and dynamic games, multi-agent control and computation in distributed networks with applications.