Linear Quadratic Mean Field Teams: Optimal and Approximately Optimal Decentralized Solutions

Jalal Arabneydi, *Student Member, IEEE,* and Aditya Mahajan, *Senior Member, IEEE*

***Abstract*—We consider team optimal control of decentralized systems with linear dynamics, quadratic costs, and arbitrary disturbance that consist of multiple sub-populations with exchangeable agents (i.e., exchanging two agents within the same sub-population does not affect the dynamics or the cost). Such a system is equivalent to one where the dynamics and costs are coupled across agents through the mean-field (or empirical mean) of the states and actions (even when the primitive random variables are non-exchangeable). Two information structures are investigated. In the first, all agents observe their local state and the mean-field of all sub-populations; in the second, all agents observe their local state but the mean-field of only a subset of the sub-populations. Both information structures are non-classical and not partially nested. Nonetheless, it is shown that linear control strategies are optimal for the first and approximately optimal for the second; the approximation error is inversely proportional to the size of the sub-populations whose mean-fields are not observed. The corresponding gains are determined by the solution of $K+1$ decoupled standard Riccati equations, where K is the number of sub-populations. The dimensions of the Riccati equations do not depend on the size of the sub-populations; thus the solution complexity is independent of the number of agents. Generalizations to major-minor agents, tracking cost, weighted mean-field, and infinite horizon are provided. The results are illustrated using an example of demand response in smart grids.**

***Index Terms*—Stochastic dynamic teams, multi-agent systems, decentralized control, non-classical information structures, linear quadratic systems, team theory, large-scale systems.**

I. Introduction

A. Motivation

Team optimal control of decentralized systems has been an important research topic since the mid 1960s. Many of the initial research results were negative and showed that even simple dynamical systems with two agents can be difficult to design—even in the celebrated linear quadratic Gaussian (LQG) framework. In particular, non-linear strategies can outperform the best linear strategy [2]; even if attention is restricted to linear strategies, the best linear strategy may not have a finite dimensional representation [3]. Since then, various solution methodologies for the optimal control of decentralized systems have been proposed and there has been considerable progress in understanding the nature of system dynamics and the information structure under which these methodologies work. See [4] and references therein for an overview.

In spite of this progress, there is a big gap between the theory and applications of optimal decentralized control. On the one hand, the envisioned applications—which include networked control systems, swarm robotics, and modern power systems—often consist of multiple interconnected dynamical systems and controllers. On the other hand, *explicit* optimal solutions are available only for systems with a few (often two or three) controllers [5]–[7]. The model and results presented in this paper attempt to reduce the gap between theory and applications.

In particular, we study decentralized control systems in which the dynamics and cost satisfy a property that we call *exchangeability*. In a dynamical system, we say agents i and j are *exchangeable* if exchanging (or interchanging) agents i and j does not affect the dynamics or the cost (the formal definition is given below). Or, equivalently, the dynamics and the cost do not depend on the index assigned to the two agents.

In many applications of decentralized systems, the system may be partitioned into sub-populations where all agents within a sub-population are exchangeable.[1] We call such systems as systems with *partially exchangeable agents*. In this paper, we develop a framework for the design of optimal decentralized control for such systems.

B. System with partially exchangeable agents

To formally define exchangeability, consider a multi-agent dynamical system where $\mathcal{N}$ denotes the set of agents. The state and action of agent $i, i \in \mathcal{N}$, at time t are denoted by x_t^i and u_t^i, where $x_t^i \in \mathcal{X}^i$ and $u_t^i \in \mathcal{U}^i$. Let $\mathbf{x}_t = (x_t^i)_{i\in\mathcal{N}}$ and $\mathbf{u}_t = (u_t^i)_{i\in\mathcal{N}}$ denote the state and action of the entire system. The dynamics are given by

$$\mathbf{x}_{t+1} = f_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t), \tag{1}$$

where f_t is system dynamics and $\{\mathbf{w}_t\}_{t\ge 1}$, where $\mathbf{w}_t = (w_t^i)_{i\in\mathcal{N}}$ and $w_t^i \in \mathcal{W}^i$, is the disturbance noise process. A per-step cost $c_t(\mathbf{x}_t, \mathbf{u}_t)$ is incurred at each time t.

For now, we do not specify the information structure as we want to identify the system properties that do not depend on the information structure.

J. Arabneydi and A. Mahajan are with the Department of Electrical and Computer Engineering, McGill University, Montreal, Quebec, Canada. Email: jalal.arabneydi@mail.mcgill.ca and aditya.mahajan@mcgill.ca.

This research was funded by the Natural Sciences and Engineering Research Council of Canada through Grant NSERC-RGPIN 402753-11.

Preliminary version of this paper [1] was presented at the 54th IEEE Conference on Decision and Control (CDC), Osaka, Japan, 2015.

[1]For example, consider an aggregator that provides demand response as a service by controlling the air conditioners in multiple neighborhoods in a city. The air-conditioners could be partitioned into sub-populations based on their tonnage and type (window, split, or packages AC). To the first-level of approximation, all air conditioners with the same tonnage and type have same dynamics and cost—and, therefore, are exchangeable. Similar situations arise in swarm robotics (where the subpopulations correspond to robots with different capabilities), and other engineering applications.

For any state $\mathbf{x}$ and agents $i, j \in \mathcal{N}$, let $\sigma_{i,j}\mathbf{x}$ denote the state when agents i and j are exchanged. For example, if $\mathbf{x} = (x^1, x^2, x^3, x^4, x^5)$, then $\sigma_{2,4}\mathbf{x} = (x^1, x^4, x^3, x^2, x^5)$. Similar interpretation holds for $\sigma_{i,j}\mathbf{u}$ and $\sigma_{i,j}\mathbf{w}$.

Definition 1 (Exchangeable agents) A pair (i, j) of agents is exchangeable if the following conditions hold:

1) $\mathcal{X}^i = \mathcal{X}^j$, $\mathcal{U}^i = \mathcal{U}^j$, and $\mathcal{W}^i = \mathcal{W}^j$, i.e., the states, actions, and disturbances of agents i and j have the same dimensions.
2) For any t, and any $\mathbf{x}_t$, $\mathbf{u}_t$, and $\mathbf{w}_t$,
$$f_t(\sigma_{i,j}\mathbf{x}_t, \sigma_{i,j}\mathbf{u}_t, \sigma_{i,j}\mathbf{w}_t) = \sigma_{i,j}\big(f_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t)\big),$$
i.e., exchanging agents i and j does not affect the system dynamics.
3) For any t, and any $\mathbf{x}_t$ and $\mathbf{u}_t$,
$$c_t(\sigma_{i,j}\mathbf{x}_t, \sigma_{i,j}\mathbf{u}_t) = c_t(\mathbf{x}_t, \mathbf{u}_t),$$
i.e., exchanging agents i and j does not affect the cost.

Definition 2 (Exchangeable set of agents) A set $\mathcal{S}$ of agents, $\mathcal{S} \subseteq \mathcal{N}$, is exchangeable if every pair of agents in $\mathcal{S}$ is exchangeable.

Definition 3 (System with partially exchangeable agents) The multi-agent system described above is called a *system with partially exchangeable agents* if the set $\mathcal{N}$ of agents can be partitioned into K disjoint subsets $\mathcal{N}^k$, $k \in \mathcal{K} := \{1, \dots, K\}$, such that for each $k \in \mathcal{K}$, the set $\mathcal{N}^k$ of agents is exchangeable.

In this paper, we investigate optimal decentralized control of linear quadratic system (i.e., a system where dynamics are linear and the per-step cost is quadratic) with partially exchangeable agents. In a subsequent paper, we will investigate systems with controlled Markovian dynamics.

C. Notation

For a set $\mathcal{N}$, $|\mathcal{N}|$ denotes its size. For a matrix A, $A^\intercal$ denotes its transpose, $\mathrm{Tr}(A)$ denotes its trace; if A is square, $A \geq 0$ (respectively $A > 0$) denotes that A is positive semi-definite (respectively positive definite). For matrices A and B of appropriate size, $A \leq B$ means $B - A \geq 0$, $\mathrm{diag}(A, B)$ denotes a block diagonal matrix with diagonal terms A and B, $\sqrt{A}$ denotes a matrix C such that[2] $C^\intercal C = A$, $A \circ B$ denotes Hadamard product, and $A \otimes B$ denotes Kronecker product. For matrices A, B, and C with the same number of columns, $\mathrm{rows}(A, B, C)$ denotes the matrix $[A^\intercal, B^\intercal, C^\intercal]^\intercal$. For vectors x, y, and z, $\mathrm{vec}(x, y, z)$ denotes the vector $[x^\intercal, y^\intercal, z^\intercal]^\intercal$.

Superscripts index agents (indexed by i) or sub-populations (indexed by k). Given a set $\mathcal{N}$ of agents and states $x^i, i \in \mathcal{N}$, bold $\mathbf{x}$ denotes $\mathrm{vec}(x^1, \dots, x^{|\mathcal{N}|})$; when all states are of the same dimension, $\langle (x^i)_{i \in \mathcal{N}} \rangle$ denotes the mean-field $\frac{1}{|\mathcal{N}|}\sum_{i=1}^{|\mathcal{N}|} x^i$ of $(x^i)_{i \in \mathcal{N}}$. For vectors and matrices, we use the short hand notation $x_{1:t}$ or $A_{1:t}$ to denote $(x_1, \dots, x_t)$ and $(A_1, \dots, A_t)$, respectively.

[2]Note that sometimes $\sqrt{A}$ is defined as a matrix C such that $CC = A$. We are not using that definition here.

$\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{R}_{>0}$ denote the sets of real, non-negative real, and positive real numbers, respectively. $\mathbb{1}_{n\times m}$ denotes $n \times m$ matrix of ones, $\mathbb{I}_n$ denotes $n \times n$ identity matrix. We omit the subscripts when the dimensions are clear from the context. For a random variable x, $\mathbb{E}[x]$ and $\mathrm{var}(x)$ denote its mean and variance, respectively.

Given horizon T and matrices $A_{1:T}$ and $Q_{1:T}$, the notation $M_{1:T} = \mathrm{DLE_T}(A_{1:T}, Q_{1:T})$ means that $M_{1:T}$ is the solution of the finite horizon discrete Lyapunov equation, i.e., $M_T = Q_T$, and for $t \in \{T-1, \dots, 1\}$, $M_t = A_t{}^\intercal M_{t+1} A_t + Q_t$.

Similarly, given a horizon T and matrices $A_{1:T}$, $B_{1:T}$, $Q_{1:T}$, and $R_{1:T}$, the notation $M_{1:T} = \mathrm{DRE_T}(A_{1:T}, B_{1:T}, Q_{1:T}, R_{1:T})$ means that $M_{1:T}$ is the solution of the finite horizon discrete Riccati equation, i.e., $M_T = Q_T$, and for $t \in \{T-1, \dots, 1\}$, $M_t = -A_t^\intercal M_{t+1} B_t \left(B_t^\intercal M_{t+1} B_t + R_t\right)^{-1} B_t^\intercal M_{t+1} A_t + A_t^\intercal M_{t+1} A_t + Q_t$.

Given a discount factor $\beta \in (0, 1]$ and matrices A, B, Q, and R, the notation $M = \mathrm{DALE}_\beta(A, Q)$ means that M is the solution of the discrete algebraic Lyapunov equation
$$M = \beta A^\intercal M A + Q.$$
and the notation $M = \mathrm{DARE}_\beta(A, B, Q, R)$ means that M is the solution of the discrete algebraic Riccati equation
$$M = -\beta A^\intercal M B \left(B^\intercal M B + \beta^{-1} R\right)^{-1} B^\intercal M A + \beta A^\intercal M A + Q.$$

II. Problem formulation and literature overview

A. Linear quadratic system with partially exchangeable agents

1) System Model: Suppose the dynamics (1) are linear, i.e.,
$$\mathbf{x}_{t+1} = A_t \mathbf{x}_t + B_t \mathbf{u}_t + \mathbf{w}_t, \tag{2}$$
where A_t and B_t are matrices of appropriate dimensions and $\{\mathbf{x}_1, \{\mathbf{w}_t\}_{t=1}^T\}$ are random variables defined on a common probability space. The cost is quadratic, i.e., for $t \in \{1, \dots, T-1\}$,
$$c_t(\mathbf{x}_t, \mathbf{u}_t) = \mathbf{x}_t{}^\intercal Q_t \mathbf{x}_t + \mathbf{u}_t{}^\intercal R_t \mathbf{u}_t, \tag{3}$$
and $t = T$,
$$c_T(\mathbf{x}_T) = \mathbf{x}_T{}^\intercal Q_T \mathbf{x}_T, \tag{4}$$
where Q_t and R_t are matrices of appropriate dimensions. Furthermore, assume that the above system is partially exchangeable, i.e., agents $\mathcal{N}$ can be partitioned into K disjoint sub-populations $\mathcal{N}^k$, $k \in \mathcal{K} := \{1, \dots, K\}$, such that for each $k \in \mathcal{K}$, the agents $\mathcal{N}^k$ are exchangeable. Moreover, for any sub-population $k \in \mathcal{K}$ and agent $i \in \mathcal{N}^k$, state x_t^i takes values in $\mathbb{R}^{d_x^k}$ and action u_t^i takes values in $\mathbb{R}^{d_u^k}$.

The *mean-field* of states[3] $\bar{x}_t^k$ of sub-population k, $k \in \mathcal{K}$, is defined as the empirical mean of the states of all agents in that sub-population, i.e.,
$$\bar{x}_t^k := \langle (x_t^i)_{i \in \mathcal{N}^k} \rangle = \frac{1}{|\mathcal{N}^k|} \sum_{i \in \mathcal{N}^k} x_t^i, \quad k \in \mathcal{K}.$$

[3]In the sequel, we refer to mean-field of the states simply as mean field.

Similarly, the mean-field of the actions $\bar{u}_t^k$ of sub-population k, $k \in \mathcal{K}$, is defined as the empirical mean of the actions of all agents in that sub-population, i.e.,

$$\bar{u}_t^k := \langle (u_t^i)_{i \in \mathcal{N}^k} \rangle = \frac{1}{|\mathcal{N}^k|} \sum_{i \in \mathcal{N}^k} u_t^i, \quad k \in \mathcal{K}.$$

The mean-field of states and actions of the entire population are denoted by $\bar{\mathbf{x}}_t$ and $\bar{\mathbf{u}}_t$ respectively, i.e.,

$$\bar{\mathbf{x}}_t = \text{vec}(\bar{x}_t^1, \dots, \bar{x}_t^K), \quad \bar{\mathbf{u}}_t = \text{vec}(\bar{u}_t^1, \dots, \bar{u}_t^K).$$

For ease of reference, the notation is summarized in Table I.

TABLE I
SUMMARY OF THE NOTATION USED IN THIS PAPER.

Notation used for agent $i \in \mathcal{N}^k$ belonging to sub-population $k \in \mathcal{K}$	
$x_t^i \in \mathbb{R}^{d_x^k}$	State of agent i
$u_t^i \in \mathbb{R}^{d_u^k}$	Action of agent i
Notation used for sup-population $k \in \mathcal{K} = \{1, \dots, K\}$	
$\mathcal{N}^k$	Entire sub-population k
$\bar{x}_t^k = \langle (x_t^i)_{i \in \mathcal{N}^k} \rangle$	Mean-field of states at time t
$\bar{u}_t^k = \langle (u_t^i)_{i \in \mathcal{N}^k} \rangle$	Mean-field of actions at time t
Notation used for entire population	
$\mathcal{N} = \bigcup_{k \in \mathcal{K}} \mathcal{N}^k$	Entire population
$\mathbf{x}_t = (x_t^i)_{i \in \mathcal{N}}$	Joint state of entire population at time t
$\mathbf{u}_t = (u_t^i)_{i \in \mathcal{N}}$	Joint action of entire population at time t
$\bar{\mathbf{x}}_t = \text{vec}(\bar{x}_t^1, \dots, \bar{x}_t^K)$	Mean-field of states of entire population at t
$\bar{\mathbf{u}}_t = \text{vec}(\bar{u}_t^1, \dots, \bar{u}_t^K)$	Mean-field of actions of entire population at t

Proposition 1 *In the linear quadratic system with partially exchangeable agents described above, there exist matrices $\{A_t^k, B_t^k, D_t^k, E_t^k, Q_t^k, R_t^k\}_{k \in \mathcal{K}}$ and P_t^x and P_t^u such that the dynamics of agent $i \in \mathcal{N}^k$ of sub-population $k, k \in \mathcal{K}$, may be written as*

$$x_{t+1}^i = A_t^k x_t^i + B_t^k u_t^i + D_t^k \bar{\mathbf{x}}_t + E_t^k \bar{\mathbf{u}}_t + w_t^i; \tag{5}$$

the per-step cost at time $t \in \{1, \dots, T-1\}$, may be written as

$$c_t(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) = \bar{\mathbf{x}}_t^\intercal P_t^x \bar{\mathbf{x}}_t + \bar{\mathbf{u}}_t^\intercal P_t^u \bar{\mathbf{u}}_t + \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}^k} \frac{1}{|\mathcal{N}^k|} \Big[(x_t^i)^\intercal Q_t^k x_t^i + (u_t^i)^\intercal R_t^k u_t^i \Big]; \tag{6}$$

and the per-step cost at time $t = T$, may be written as

$$c_T(\mathbf{x}_T, \bar{\mathbf{x}}_T) = \bar{\mathbf{x}}_T^\intercal P_T^x \bar{\mathbf{x}}_T + \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}^k} \frac{1}{|\mathcal{N}^k|} (x_T^i)^\intercal Q_T^k x_T^i. \tag{7}$$

The proof is presented in Appendix A.

Remark 1 In general, the matrices $(A_t^k, B_t^k, D_t^k, E_t^k, Q_t^k, R_t^k)$ and (P_t^x, P_t^u) may depend on the number $\{|\mathcal{N}^k|\}_{k \in \mathcal{K}}$ of agents in the sub-populations, but their dimensions do not.

Thus, any linear quadratic system with partial exchangeable agents—irrespective of the information structure—is equivalent to a mean-field coupled system with the same information structure. In the rest of this paper, we investigate the optimal control of such systems under the following two information structures.

2) Observation model and information structure: We consider two information structures; in both, agents perfectly recall all data that they observe. In the first information structure, which we call *mean field sharing* and denote by MFS-IS, every agent $i \in \mathcal{N}$ perfectly observes its local state x_t^i and the global mean-field $\bar{\mathbf{x}}_t$. Thus, the data I_t^i available to agent i at time t is given by

$$I_t^i = (x_{1:t}^i, u_{1:t-1}^i, \bar{\mathbf{x}}_{1:t}). \tag{MFS-IS}$$

In the second information structure, which we call *partial mean field sharing* and denote by PMFS-IS, there exists a subset $\mathcal{S}$ of the sub-populations $\mathcal{K}$ such that every agent $i \in \mathcal{N}$ perfectly observes its local state x_t^i and the mean-fields of sub-populations $\mathcal{S}$, i.e., $\{\bar{x}_t^k\}_{k \in \mathcal{S}}$. We use $\mathcal{S}^c$ to denote $\mathcal{K} \backslash \mathcal{S}$. The data I_t^i available to agent i at time t is given by

$$I_t^i = (x_{1:t}^i, u_{1:t-1}^i, (\bar{x}_{1:t}^k)_{k \in \mathcal{S}}). \tag{PMFS-IS}$$

Under both information structures, agent i chooses u_t^i as follows:

$$u_t^i = g_t^i(I_t^i). \tag{8}$$

The function g_t^i is called the *control law of agent i* at time t. The collection $\mathbf{g}^i = (g_1^i, g_2^i, \dots, g_T^i)$ is called the *control strategy of agent i*. The collection $\mathbf{g} = (\mathbf{g}^i)_{i \in \mathcal{N}}$ is called the *control strategy of the system*. The performance of strategy $\mathbf{g}$ is given by

$$J(\mathbf{g}) = \mathbb{E}^{\mathbf{g}} \Bigg[\sum_{t=1}^{T-1} c_t(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) + c_T(\mathbf{x}_T, \bar{\mathbf{x}}_T) \Bigg], \tag{9}$$

where the expectation is with respect to the measure induced on all the system variables by the choice of strategy $\mathbf{g}$.

3) The optimization problem: We are interested in the following optimization problem.

Problem 1 *In the model described above, find a strategy $\mathbf{g}^*$ that minimizes* (9), *i.e.,*

$$J^* := J(\mathbf{g}^*) = \inf_{\mathbf{g}} J(\mathbf{g}),$$

where the infimum is taken over all strategies of form (8).

B. Conceptual difficulties

There are several conceptual difficulties in solving Problem 1 because ***it has a non-classical information structure***. Information structure refers to the set of information known to all agents at all times. If every decision maker knows the observations and actions of all decision makers that acted before it, then the information structure is said to be *classical*; if every decision maker knows the observations and actions of all decision makers whose actions effect its observations, then the information structure is said to be *partially nested*; otherwise, the information structure is said to be *non-classical* [8], [9]. For linear quadratic systems with classical or partially nested information structures, when the primitive random variables are jointly Gaussian, the optimal control action is a linear (or affine) function of the observations[4]. This is not necessarily

[4]In classical information structure *with state feedback*, the optimal control action is linear function of the state and this result holds even when the primitive random variables are not Gaussian.

the case when the information structure is non-classical as is illustrated by the Witsenhausen counterexample [2], which presents a linear quadratic Gaussian model with non-classical information structure where non-linear strategies outperform the best linear strategy. The model presented in this paper is neither classical nor partially nested nor the primitive random variables are necessarily Gaussian, so it is not known a priori whether there is no loss of optimality in restricting attention to linear strategies.

Even when linear strategies are not optimal, sometimes attention is restricted to linear strategies because they are simple and easy to implement. For systems with non-classical information structure, the problem of finding the best linear strategy need not be convex; it is convex only for special sparsity pattern such as funnel causality [10] and quadratic invariance [11]. Furthermore, as is illustrated by the Whittle and Rudge counterexample [3], even when the problem of finding the best linear strategy is convex, the best linear strategy might not have a finite dimensional representation.

Finally, the usual curse of dimensionality is exasperated in systems with non-classical information structure. Even in systems with finite state and action spaces, the complexity of finding the optimal control strategy belongs to NEXP complexity class [12].

C. Contributions of the paper

1) We show that linear control laws are team optimal for MFS-IS (even when the noise processes are not Gaussian). As argued earlier, MFS-IS does not fall into the class of information structures for which linear strategies are known to be optimal. We show that the corresponding gains are computed by solving $K+1$ *decoupled* Riccati equations (where K is the number of sub-populations) (Theorem 1).
2) We propose a certainty equivalence linear strategy for PMFS-IS and show that the error satisfies a Lyapunov equation. The approximation error converges to zero at a rate that is inversely proportional to the number of agents in the sub-populations whose mean-fields are not observed (Theorem 2).
3) The salient feature of our main results is that the solution complexity does not depend on the number of agents in each sub-population; rather, it only depends on the number of sub-populations. Furthermore, the optimal gains can be computed in a decentralized manner such that each agent simply needs to solve at most two (rather than all) Riccati equations.
4) We show that our results generalize to variations of the basic model that are not partially exchangeable including: systems where the objective is to optimally track reference trajectories (Sec. IV-C) and systems where agents have individual weights (Sec. IV-D).
5) When the dynamics and the per-step cost are time-homogeneous, we show that our results extend to infinite horizon setups: both for the discounted cost setup with any discount factor in $(0,1)$ and for average-cost per unit time setup. For both setups, the optimal control strategy for MFS-IS and the approximately optimal control strategy for PMFS-IS are time-homogeneous and the corresponding gains are computed by solving $K+1$ *decoupled* algebraic Riccati equations.

D. Literature overview

Our model and results for MFS-IS are similar in spirit to those obtained in [13] under stronger modeling assumptions. In [13], the authors consider a *homogeneous* population of *dynamically decoupled* agents which are coupled in the cost through a weighted mean-field term. Two models are investigated: (a) *hard-constraint model* where the weighted mean-field of actions must equal a pre-specified linear function of the weighted mean-field of states; and (b) *soft-constraint model* where the above hard constraint is relaxed by penalizing it in the cost. For both models, the authors show that the optimal centralized control laws are linear in the local state and the mean field; the corresponding gains are computed by two decoupled Riccati equations. In section IV-D, we generalize our results to the case when a weighted empirical mean field is shared. In contrast to [13], we consider multiple sub-populations and allow agents to be coupled in dynamics. Note that approximation results similar to those for partial mean-field sharing were not considered in [13].

Our results have similar features to those obtained for *centralized* linear quadratic mean-field control [14], [15]. In these models, the dynamics and the cost depend on the *statistical* mean-field of the state and action. Such a model may be viewed as a special case of our model when we restrict to a single homogeneous sub-population and consider the limit of infinite number of agents (and therefore the empirical mean and the statistical mean are the same). Our proof technique, which relies on a simple change of variables, is conceptually simpler than that of [14], [15].[5] It is worth highlighting that the linear quadratic mean-field control model is a centralized control problem and the results of [14], [15] do not apply to the multi-agent models that we consider.

Recently, an iterative bidding strategy was proposed in [17] for the optimal control multi-agent systems with decoupled dynamics that are coupled through a constraint. For LQG agents, the scheme operates as follows: at each time, a coordinator sets a price profile for all future times; agents submit a bid profile for all future times; the coordinator updates the prices and the process continues until the bids have converged. Agents choose the first value of their bid as their action and the above process is repeated at the next time step. In this scheme, agents do not need to know the system dynamics of other agents. In contrast, we assume that the system dynamics are common knowledge to all agents. However, in our model, agents only need to share the mean-field of their states (which can be computed using a consensus algorithm) rather than iteratively sharing the bid profile for all future times.

A decomposition-coordination approach for optimal decentralized control of *deterministic* linear quadratic systems was

[5] In [14], first coupled forward and backward stochastic differential equations are derived and then they are decoupled into two Riccati equations using the four step technique of [16]. In [15], a matrix dynamical optimization method is used.

proposed in [18], [19]. This is an iterative approach. Each iteration consists of two steps: (i) a *decomposition step* in which each agent assumes decoupled dynamics and costs and computes its local control trajectory by solving an optimal tracking problem from pre-specified linear offsets for the dynamics and a reference trajectory for the cost; (ii) a *coordination* step in which the linear offsets for the dynamics and reference trajectories for the cost are computed for all agents from the pre-specified control trajectories. It is shown that this iterative process converges to the optimal centralized solution. In contrast to such decomposition-coordination methods, our proposed solution is not iterative. The optimal gains for all agents are computed in a single step by solving Riccati equations. Furthermore, our solution methodology works for deterministic as well as stochastic systems.

A related solution approach called mean-field games (MFG) was proposed in [20]–[28] to compute approximate Nash equilibrium for large population games. The main idea is to assume an infinite large size of each sub-population and solve a set of two coupled equations: a Hamilton-Jacobi-Bellman (HJB) equation to compute the best response of a generic agent playing against a "mass trajectory" and a Fokker-Planck-Kolmogorov (FPK) equation to compute the mass trajectory from the strategy of a generic agent. It is shown that a solution to these equations exists under appropriate conditions. The resulting strategies are ε-Nash when the sub-populations are finite, where the approximation error is $\mathcal{O}(1/\sqrt{n})$, where n denotes the size of the smallest sub-population. For linear quadratic systems, the coupled HJB-FPK equations simplify to K Riccati equations and two coupled forward and backward ODEs. In contrast, in our solution there is an additional Riccati equation instead of the coupled forward-backward equations. The coupled equations in MFG depend on the initial mean-field while the Riccati equations in our solution do not. The key difference between our results and the results in the MFG literature is that we obtain team optimal strategies of a decentralized control problem while in the MFG literature one typically obtains either Nash or Markov perfect equilibrium strategies of a large population dynamic game problem. These solution concepts are different.

The approach of mean-field games was used to obtain team optimal solution of linear quadratic systems with *decoupled* dynamics in [29]. It is shown that the MFG solution is ε-socially optimal (with $\varepsilon \in \mathcal{O}(1/\sqrt{n})$). We obtain a similar result for *dynamically coupled* agents with $\varepsilon \in \mathcal{O}(1/n)$.

It should be noted that identifying *team-optimal* control laws for systems with coupled dynamics is significantly more challenging than for systems with decoupled dynamics. This is because, when the agent dynamics are decoupled (and the primitive random variables are Gaussian), the information structure is partially nested, so one may restrict attention to linear strategies. Furthermore, for a finite horizon system, team-optimal strategies may be obtained by solving a set of linear equations.[6] In contrast, when the system dynamics are coupled, the information structure is non-classical and there is no general solution methodology to obtain a *team-optimal* solution.

III. Main Results

A. Exact solution for MFS-IS

We impose following standard assumptions on the model described in Proposition 1:

Assumption (A1) *The primitive random variables $\{\mathbf{x}_1, \{\mathbf{w}_t\}_{t=1}^T\}$ have zero mean, finite variance, and are mutually independent.*

Remark 2 Note that we do not require the primitive random variables to be Gaussian. Nor do we require the initial state $\mathbf{x}_1$ and the disturbance $\mathbf{w}_t$ to be independent or exchangeable across agents.

Assumption (A2) *For every t, P^x_t, P^u_t, Q^k_t, and R^k_t are symmetric matrices that satisfy*

$$Q^k_t \geq 0,\ \forall k \in \mathcal{K}, \qquad \operatorname{diag}(Q^1_t, \dots, Q^K_t) + P^x_t \geq 0, \tag{10}$$
$$R^k_t > 0,\ \forall k \in \mathcal{K}, \qquad \operatorname{diag}(R^1_t, \dots, R^K_t) + P^u_t > 0. \tag{11}$$

Note that matrices P^x_t and P^u_t are not required to be positive semi-definite as long as (10)–(11) hold.

Theorem 1 *Under (A1), (A2), and (MFS-IS), we have the following results for Problem 1.*

1) Structure of optimal strategy: *The optimal strategy for Problem 1 is unique and is linear in the local state and the mean-field of the system. In particular,*
$$u^i_t = \breve{L}^k_t (x^i_t - \bar{x}^k_t) + \bar{L}^k_t \bar{\mathbf{x}}_t, \tag{12}$$
where the gains $\{\breve{L}^k_t, \bar{L}^k_t\}_{t=1}^{T-1}$ are obtained by the solution of $K+1$ Riccati equations given below: one for computing each $\breve{L}^k_t$, $k \in \mathcal{K}$, and one for $\bar{L}_t := \operatorname{rows}(\bar{L}^1_t, \dots, \bar{L}^K_t)$.
2) Riccati equations: *Let*
$$\bar{A}_t := \operatorname{diag}(A^1_t, \dots, A^K_t) + \operatorname{rows}(D^1_t, \dots, D^K_t),$$
$$\bar{B}_t := \operatorname{diag}(B^1_t, \dots, B^K_t) + \operatorname{rows}(E^1_t, \dots, E^K_t),$$
$$\bar{Q}_t := \operatorname{diag}(Q^1_t, \dots, Q^K_t), \quad \bar{R}_t := \operatorname{diag}(R^1_t, \dots, R^K).$$
Then, for $t \in \{1, \dots, T-1\}$, define:
$$\breve{L}^k_t = -\left((B^k_t)^\intercal \breve{M}^k_{t+1} B^k_t + R^k_t\right)^{-1} (B^k_t)^\intercal \breve{M}^k_{t+1} A^k_t,$$
$$\bar{L}_t = -\left(\bar{B}^\intercal_t \bar{M}_{t+1} \bar{B}_t + \bar{R}_t + P^u_t\right)^{-1} \bar{B}^\intercal_t \bar{M}_{t+1} \bar{A}_t,$$
where $\{\breve{M}^k_t\}_{t=1}^T$ and $\{\bar{M}_t\}_{t=1}^T$ are the solutions of following Riccati equations:
$$\breve{M}^k_{1:T} = \operatorname{DRE_T}(A^k_{1:T}, B^k_{1:T}, Q^k_{1:T}, R^k_{1:T}), \tag{13}$$
$$\bar{M}_{1:T} = \operatorname{DRE_T}(\bar{A}_{1:T}, \bar{B}_{1:T}, \bar{Q}_{1:T} + P^x_{1:T}, \bar{R}_{1:T} + P^u_{1:T}). \tag{14}$$

[6] It is shown in [9] that a finite horizon system with partially nested information structure may be converted to a static team by an appropriate change of variables. The optimal control laws for such a static team may be obtained by solving a set of linear equations [30]. The key conceptual challenge in such problem is to identify sufficient statistics such that the optimal control laws can be computed efficiently and the results can generalize to infinite-horizon setup.

3) Optimal performance: *Let*

$$\breve{\Sigma}^k_t := \frac{1}{|\mathcal{N}^k|}\sum_{i\in\mathcal{N}^k}\operatorname{var}(w^i_t - \bar{w}^k_t), \quad \bar{\Sigma}_t := \operatorname{var}(\bar{\mathbf{w}}_t),$$
$$\breve{\Xi}^k := \frac{1}{|\mathcal{N}^k|}\sum_{i\in\mathcal{N}^k}\operatorname{var}(x^i_1 - \bar{x}^k_1), \quad \bar{\Xi} := \operatorname{var}(\bar{\mathbf{x}}_1).$$

Then, the optimal cost is given by

$$J^* = \sum_{k\in\mathcal{K}}\operatorname{Tr}(\breve{\Xi}^k\breve{M}^k_1) + \operatorname{Tr}(\bar{\Xi}\bar{M}_1) + \sum_{t=1}^{T-1}\Big[\sum_{k\in\mathcal{K}}\operatorname{Tr}(\breve{\Sigma}^k_t\breve{M}^k_{t+1}) + \operatorname{Tr}(\bar{\Sigma}_t\bar{M}_{t+1})\Big]. \quad (15)$$

The proof is presented in Section V. Note that the dimensions of Riccati equations (13) and (14) do not depend on the sizes of the sub-populations ($|\mathcal{N}^1|,\dots,|\mathcal{N}^K|$). Hence, the solution complexity depends only on the number K of sub-populations and it is independent of the number of agents in each sub-population. To implement the optimal control strategies:

- all agents must compute $\bar{L}_{1:T-1}$ by solving the Riccati equation (14),
- agents of sub-population k must compute $\breve{L}^k_{1:T-1}$ by solving the Riccati equation (13).

Then, an individual agent i of sub-population k, upon observing the local state x^i_t and the global mean-field $\bar{\mathbf{x}}_t$, chooses its local control action according to (12). Note that each agent needs to solve only two Riccati equations, although there are $K+1$ Riccati equations in Theorem 1.

Remark 3 An interesting feature of the solution is that all agents in a particular sub-population use identical control laws. This is a feature of the linear quadratic system and not of exchangeability.[7]

Remark 4 If the per-step cost has cross-terms involving $(x^i_t, \bar{\mathbf{x}}_t)$ and $(u^i_t, \bar{\mathbf{u}}_t)$, i.e.,

$$\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}^k}\frac{1}{|\mathcal{N}^k|}\Big[(x^i_t)^\intercal S^{x,k}_t\bar{\mathbf{x}}_t + (u^i_t)^\intercal S^{u,k}_t\bar{\mathbf{u}}_t\Big]$$

then, this cost can be re-written in the form of (6) and (7):

$$\bar{\mathbf{x}}_t^\intercal S^x_t\bar{\mathbf{x}}_t + \bar{\mathbf{u}}_t^\intercal S^u_t\bar{\mathbf{u}}_t$$

where

$$S^x_t := \operatorname{rows}(S^{x,1}_t,\dots,S^{x,K}_t), \quad S^u_t := \operatorname{rows}(S^{u,1}_t,\dots,S^{u,K}_t).$$

[7] The following example (which is based on an example presented in [31]) shows that asymmetric control strategies may outperform symmetric ones even in systems with exchangeable agents. Consider a system with 2 agents that runs for a horizon 2. Let $\mathcal{X} = \mathcal{U} = \{1,2\}$ and suppose the initial state (x^1_1, x^2_1) is uniformly distributed over all possible values. Suppose the dynamics are $x^i_2 = u^i_1$, $i\in\{1,2\}$ and the costs are $c_1(\mathbf{x}_1,\mathbf{u}_1) = 0$ and $c_2(\mathbf{x}_2,\mathbf{u}_2)$ is C when $\{x^1_2 = x^2_2\}$ (where C is a positive number) and 0 otherwise. The above system is exchangeable. Any symmetric control strategy puts positive probability on the event $\{u^1_1 = u^2_1\}$ (and, hence on the event $\{x^1_2 = x^2_2\}$) and, therefore, has a positive expected cost. On the other hand, the asymmetric strategy $u^i_1 = i$ has zero cost. Thus, symmetric control strategies are not optimal.

Remark 5 We assumed that there are no cross-terms of the form $x^\intercal Su$ in the per-step cost of (3) and (4). If such cross-terms are present, there will be cross-terms involving (x^i_t, u^i_t), $(x^i_t, \bar{u}_t)$, $(\bar{x}_t, u^i_t)$, and $(\bar{x}_t, \bar{u}_t)$ in the equivalent mean-field model presented in Proposition 1. These cross-terms can be treated in the standard manner as cross-terms are treated in centralized LQR.

Remark 6 Suppose in addition to (A1), we have that $\{x^i_1, \{w^i_t\}_{t\ge1}\}_{i\in\mathcal{N}}$ are independent and for any $k\in\mathcal{K}$, $(x^i_1)_{i\in\mathcal{N}^k}$ is i.i.d. with variance Ξ^k and $\{w^i_t\}_{i\in\mathcal{N}^k}$ is i.i.d. with variance Σ^k_t. Then, we have

$$\breve{\Sigma}^k_t = \frac{|\mathcal{N}^k|-1}{|\mathcal{N}^k|}\Sigma^k_t, \qquad \bar{\Sigma}_t = \operatorname{diag}(\Sigma^1_t,\dots,\Sigma^K_t),$$
$$\breve{\Xi}^k = \frac{|\mathcal{N}^k|-1}{|\mathcal{N}^k|}\Xi^k, \qquad \bar{\Xi} = \operatorname{diag}(\Xi^1,\dots,\Xi^K).$$

The expression of total cost (15) can be simplified accordingly.

B. Approximate solution for PMFS-IS

In this section, we consider Problem 1 under PMFS-IS. Based on the results of Theorem 1, we propose a certainty equivalence strategy for PMFS-IS and show that the performance of this strategy is close to the optimal performance under MFS-IS. We impose the following assumptions on the model.

Assumption (A1a) *In addition to (A1), for any $k\in\mathcal{S}$ and $k'\in\mathcal{S}^c$, initial states $(x^i_1)_{i\in\mathcal{N}^k}$ are independent of $(x^j_1)_{j\in\mathcal{N}^{k'}}$.*

Assumption (A1b) *The primitive random variables $\{x^i_1, \{w^i_t\}_{t=1}^T\}_{i\in\mathcal{N}}$ are independent. For any $k,k\in\mathcal{K}$, there exist finite matrices c^k_x and c^k_w such that*

$$\sup_{i\in\mathcal{N}^k}\operatorname{var}(x^i_1)\le c^k_x, \qquad \sup_{t\le T, i\in\mathcal{N}^k}\operatorname{var}(w^i_t)\le c^k_w.$$

Assumption (A3) *The dynamics $\{A^k_t, B^k_t, D^k_t, E^k_t\}_{k\in\mathcal{K}}$, cost $\{Q^k_t, R^k_t,\}_{k\in\mathcal{K}}$, P^x_t and P^u_t, and covariance bounds $\{c^k_x, c^k_w\}_{k\in\mathcal{K}}$ do not depend on the sizes $(|\mathcal{N}^1|,\dots,|\mathcal{N}^K|)$ of the sub-populations.*

Since we are comparing the system performance under two information structures, we use different notation for the two. Under MFS-IS, the state and action of agent i are denoted by x^i_t and u^i_t. Assume that u^i_t is generated as per Theorem 1. Under PMFS-IS, the state and action of agent i are denoted by s^i_t and v^i_t. The dynamics are same as (5). In particular for agent i of sub-population $k\in\mathcal{K}$, $s^i_1 = x^i_1$ and

$$s^i_{t+1} = A^k_t s^i_t + B^k_t v^i_t + D^k_t\bar{\mathbf{s}}_t + E^k_t\bar{\mathbf{v}}_t + w^i_t, \qquad (16)$$

where

$$\bar{\mathbf{s}}_t = \operatorname{vec}(\bar{s}^1_t,\dots,\bar{s}^K_t), \qquad \bar{s}^k_t = \langle(s^i_t)_{i\in\mathcal{N}^k}\rangle,$$
$$\bar{\mathbf{v}}_t = \operatorname{vec}(\bar{v}^1_t,\dots,\bar{v}^K_t), \qquad \bar{v}^k_t = \langle(v^i_t)_{i\in\mathcal{N}^k}\rangle.$$

Define a *(mean-field) approximation process* $\{\mathbf{z}_t\}_{t=1}^T$ as follows: $\mathbf{z}_t = \operatorname{vec}(z^1_t,\dots,z^K_t)$, where for any $k\in\mathcal{K}$,

$z_t^k \in \mathbb{R}^{d_x^k}$; the initial state $\mathbf{z}_1$ is given by z_1^k is $\bar{s}_1^k$ for $k \in \mathcal{S}$ and is 0 for $k \notin \mathcal{S}$.[8] The process evolves as:

$$z_{t+1}^k = \begin{cases} \bar{s}_{t+1}^k, & k \in \mathcal{S}, \\ A_t^k z_t^k + (B_t^k \bar{L}_t^k + D_t^k + E_t^k \bar{L}_t)\mathbf{z}_t, & k \in \mathcal{S}^c, \end{cases} \quad (17)$$

where $\bar{L}_t$ is as defined in Theorem 1. Note that the approximation process $\{\mathbf{z}_t\}_{t=1}^T$ is adapted to the filtration $\{\{\bar{s}_t^k\}_{k\in\mathcal{S}}\}_{t=1}^T$ which is known at all agents. Therefore, at time t, $\mathbf{z}_t$ can be computed at all agents.

Now, consider the following certainty equivalence strategy for PMFS-IS: for agent i of sub-population $k, k \in \mathcal{K}$,

$$v_t^i = \breve{L}_t^k (s_t^i - z_t^k) + \bar{L}_t^k \mathbf{z}_t. \quad (18)$$

The above strategy is similar to the optimal strategy for MFS-IS (given by (12) in Theorem 1) except that the mean field $\{\bar{s}_t^k\}_{k\in\mathcal{K}}$ has been replaced by its approximation $\mathbf{z}_t$.

For ease of exposition, let $d_x := \sum_{k\in\mathcal{K}} d_x^k$ and matrix $H = \text{rows}(H^1, \dots, H^K)$ be a binary matrix such that

$$H^k = \begin{cases} 0_{d_x^k \times d_x}, & k \in \mathcal{S}, \\ \mathbb{1}_{d_x^k \times d_x}, & k \in \mathcal{S}^c. \end{cases}$$

Let $\hat{J}$ denote the performance of strategy (18) and J^* denote the optimal performance under MFS-IS. Then, the difference in performance $\hat{J} - J^*$ is bounded. In particular, we have

Theorem 2 *Assume (A1a), (A2), and (PMFS-IS). Then,*

1) *The performance loss is given by*

$$\hat{J} - J^* = \text{Tr}(\tilde{X}_1 \tilde{M}_1) + \sum_{t=1}^{T-1} \text{Tr}(\tilde{W}_t \tilde{M}_{t+1}), \quad (19)$$

where $\tilde{X}_1 = \mathbb{1}_{2d_x \times 2d_x} \otimes [H \circ \text{var}(\bar{\mathbf{x}}_1)]$, $\tilde{W}_t = \mathbb{1}_{2d_x \times 2d_x} \otimes [H \circ \text{var}(\bar{\mathbf{w}}_t)]$, *and* $\tilde{M}_{1:T}$ *is the solution of following Lyapunov equation:*

$$\tilde{M}_{1:T} = \text{DLE}_\text{T}(\tilde{A}_{1:T}, \tilde{Q}_{1:T}), \quad (20)$$

where

$$\tilde{A} = \begin{bmatrix} \tilde{A}_t^1 & -(\mathbb{1}_{d_x\times d_x} - H) \circ \tilde{A}_t^2 \\ 0 & H \circ \tilde{A}_t^2 \end{bmatrix},$$

and $\tilde{Q}_t = \text{diag}(-\tilde{Q}_t^1, \tilde{Q}_t^2)$ *where* $\tilde{A}_t^1 = \bar{A}_t + \bar{B}_t \bar{L}_t$, $\tilde{A}_t^2 = \bar{A}_t + \bar{B}_t \breve{L}_t$, $\tilde{Q}_t^1 = P_t^x + \bar{Q}_t + \bar{L}_t^\intercal (P_t^u + \bar{R}_t)\bar{L}_t$, $\tilde{Q}_t^2 = P_t^x + \bar{Q}_t + \breve{L}_t^\intercal (P_t^u + \bar{R}_t)\breve{L}_t$, *and* $\breve{L}_t = \text{diag}(\breve{L}_t^1, \dots, \breve{L}_t^K)$.

2) *Let* $n = \min_{k\in\mathcal{S}^c}(|\mathcal{N}^k|)$. *Under (A1b) and (A3),*

$$\hat{J} - J^* \in \mathcal{O}\left(\frac{T}{n}\right).$$

The result is proved in Section VI.

Remark 7 As the number of agents in each sub-population $k \in \mathcal{S}^c$, becomes large, the approximation error $\hat{J} - J^*$ goes to zero; therefore, PMFS-IS is as informative as MFS-IS.

Note that when the mean-field of all sub-populations are shared, then $\mathcal{S} = \mathcal{K}$ and, therefore, H is zero. Consequently, the approximation error given by (19) is zero. Hence, the result of Theorem 2 is consistent with that of Theorem 1.

[8]If the initial states are non-zero mean, then $z_1^k = \mathbb{E}(\bar{x}_1^k)$ for $k \notin \mathcal{S}$.

Corollary 1 *When the mean-field is not shared, i.e.,* $\mathcal{S} = \emptyset$, *the approximation error* $\hat{J} - J^*$ *is*

$$\text{Tr}\left(\text{var}(\bar{\mathbf{x}}_1)(\tilde{M}_1^2 - \tilde{M}_1^1)\right) + \sum_{t=1}^{T-1} \text{Tr}(\text{var}\left(\bar{\mathbf{w}}_t)(\tilde{M}_{t+1}^2 - \tilde{M}_{t+1}^1)\right),$$

where $\tilde{M}_{1:T}^1$ *and* $\tilde{M}_{1:T}^2$ *are the solutions of following two decoupled Lyapunov equations:*

$$\tilde{M}_{1:T}^1 = \text{DLE}_\text{T}(\tilde{A}_{1:T}^1, \tilde{Q}_{1:T}^1), \quad \tilde{M}_{1:T}^2 = \text{DLE}_\text{T}(\tilde{A}_{1:T}^2, \tilde{Q}_{1:T}^2).$$

Proof: When $\mathcal{S} = \emptyset$, H is $\mathbb{1}_{d_x\times d_x}$; thus, $\tilde{A}_t$ is block diagonal. Consequently, the Lyapunov equation (20) decouples into the two smaller Lyapunov equations given above. ■

IV. Special cases and generalizations

In this section, we present two special cases and two generalizations of Problem 1. Due to space limitations, we only present the results for MFS-IS (i.e., the analogue of Theorem 1); the results for PMFS (i.e., the analogue of Theorem 2) may be derived in a similar manner.

A. Special case 1: major and minor agents

Suppose there exist $\mathcal{M} \subseteq \mathcal{K}$ sub-populations with only 1 agent, i.e., $|\mathcal{N}^k| = 1, k \in \mathcal{M}$. Then, for every $k \in \mathcal{M}$, $\bar{x}_1^k = x_t^k$. The rest of the dynamics and cost are the same as in Section II-A. Since the dynamics are coupled through the mean-field, the states of the agents of sub-populations $\mathcal{M}$ directly influence the dynamics of all other agents and the per-step cost. For this reason, such agents are called *major* agents. A variation of the above model with a single major agent was first introduced in [32] and other variations have been investigated in [33]–[35].

For above model, result of Theorem 1 simplifies as follows.

Corollary 2 *For any sub-population* $k \in \mathcal{K}\backslash\mathcal{M}$ *and minor agent* $i \in \mathcal{N}^k$, u_t^i *is given by* (12). *For any major agent* $i \in \mathcal{N}^k, k \in \mathcal{M}$, *the control law is given by* $u_t^k = \bar{L}_t^k \bar{\mathbf{x}}_t$.

Note that for $k \in \mathcal{M}$, $\breve{L}_t^k$ is not needed to compute u_t^k; so we do not need a Riccati equation to compute $\breve{M}_{1:T}^k$.

B. Special case 2: no local controls

Suppose that for all $k \in \mathcal{K}$, $B_t^k = 0$ and $R_t^k = 0$. Moreover assume that there exists a vector $\boldsymbol{\theta}_t = \text{rows}(\theta_t^1, \dots, \theta_t^K), \theta_t^k \in \mathbb{R}^{d_{\tilde{u}}} \times \mathbb{R}^{d_u}, k \in \mathcal{K}$, such that $E_t^k = \tilde{E}_t^k \boldsymbol{\theta}_t^\intercal$ for all $k \in \mathcal{K}$ and $P_t^u = \boldsymbol{\theta}_t^\intercal \tilde{P}_t^u \boldsymbol{\theta}_t$. In addition, let $\theta_t^{k^+}$ denote the right inverse of θ_t^k (i.e., $\theta_t^k \theta_t^{k^+} = \mathbb{I}_{\mathbb{R}^{d_{\tilde{u}}}}$), which is assumed to exist. This implies that the dynamics and cost are given as follows. Let

$$\tilde{u}_t := \boldsymbol{\theta}_t^\intercal \bar{\mathbf{u}}_t = \sum_{k\in\mathcal{K}} \theta_t^k \bar{u}_t^k.$$

Then, for agent $i \in \mathcal{N}^k$ of sub-population $k \in \mathcal{K}$, we have

$$x_{t+1}^i = A_t^k x_t^i + D_t^k \bar{\mathbf{x}}_t + \tilde{E}_t^k \tilde{u}_t + w_t^i.$$

At time $t \in \{1, \dots, T-1\}$, the per-step cost is given by,

$$c_t(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t, \tilde{u}_t) = \bar{\mathbf{x}}_t^\intercal P_t^x \bar{\mathbf{x}}_t + \tilde{u}_t^\intercal \tilde{P}_t^u \tilde{u}_t + \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}^k} \frac{1}{|\mathcal{N}^k|} (x_t^i)^\intercal Q_t x_t^i,$$

and $t = T$,

$$c_T(\mathbf{x}_T, \bar{\mathbf{x}}_T) = \bar{\mathbf{x}}_T^\intercal P_T^x \bar{\mathbf{x}}_T + \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}^k} \frac{1}{|\mathcal{N}^k|} (x_T^i)^\intercal Q_T x_T^i.$$

Corollary 3 *For the model described above, the optimal control law is given as follows. For all $k \in \mathcal{K}$ and $i \in \mathcal{N}^k$,*

$$u_t^i = \theta_t^{k+} \tilde{L}_t^k \bar{x}_t^k.$$

where $[\tilde{L}_t^1, \dots, \tilde{L}_t^K] =: \bar{L}_t$ is given as in Theorem 1 but with $\bar{B}_t$ replaced by $\tilde{B}_t = \operatorname{rows}(\tilde{E}_t^1, \dots, \tilde{E}_t^K)$ and P_t^u replaced by $\tilde{P}_t^u$.

The proof is presented in Appendix B.

Remark 8 Note that for the model defined above, each agent only needs to observe the mean-field of its sub-population (rather than the mean-field of entire population). Thus, this result is similar in spirit to [36, Theorem 1].

C. Generalization 1: tracking cost function

Consider a tracking problem in which we are given a tracking signal $\{s_t^k\}_{t=1}^T$, $s_t^k \in \mathbb{R}^{d_x^k}$ for the mean-field of sub-population $k \in \mathcal{K}$ and a tracking signal $\{r_t^i\}_{t=1}^T$, $r_t^i \in \mathbb{R}^{d_x^k}$, for each agent $i \in \mathcal{N}^k$.

Define $\bar{r}_t^k := \langle (r_t^i)_{i \in \mathcal{N}^k} \rangle, k \in \mathcal{K}$, $\bar{\mathbf{r}}_t := \operatorname{vec}(\bar{r}_t^1, \dots, \bar{r}_t^K)$, and $\mathbf{s}_t = \operatorname{vec}(s_t^1, \dots, s_t^K)$. The tracking cost is as follows. For $t \in \{1, \dots, T-1\}$,

$$c_t(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) = (\bar{\mathbf{x}}_t - \mathbf{s}_t)^\intercal P_t^x (\bar{\mathbf{x}}_t - \mathbf{s}_t) + \bar{\mathbf{u}}_t^\intercal P_t^u \bar{\mathbf{u}}_t + \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}^k} \frac{1}{|\mathcal{N}^k|} \left[(x_t^i - r_t^i)^\intercal Q_t^k (x_t^i - r_t^i) + (u_t^i)^\intercal R_t^k u_t^i \right],$$

and for $t = T$,

$$c_T(\mathbf{x}_T, \bar{\mathbf{x}}_T) = (\bar{\mathbf{x}}_T - \mathbf{s}_T)^\intercal P_T^x (\bar{\mathbf{x}}_T - \mathbf{s}_T) + \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}^k} \frac{1}{|\mathcal{N}^k|} (x_T^i - r_T^i)^\intercal Q_T^k (x_T^i - r_T^i).$$

We assume that, in addition to MFS-IS specified in Section II-A2, agent i also knows signals $\{r_t^i, \bar{\mathbf{r}}_t, \mathbf{s}_t\}_{t=1}^T$. The rest of the model is the same as in Section II-A.

Theorem 3 *Under (A1), (A2), and (MFS-IS), the optimal strategy is unique and given by*

$$u_t^i = \breve{L}_t^k (x_t^i - \bar{x}_t^k) + \bar{L}_t^k \bar{\mathbf{x}}_t + \breve{F}_t^k v_t^i + \bar{F}_t^k \bar{v}_t, \tag{21}$$

where the gains $\{\breve{L}_t^k, \bar{L}_t^k\}_{t=1}^{T-1}$ are obtained by the solution of $K+1$ Riccati equations defined in Theorem 1 and the gains $\{\breve{F}_t^k, \bar{F}_t^k\}_{t=1}^{T-1}$ and the correction signals $\{v_t^i, \bar{v}_t\}_{t=1}^T$ are given as follows. Let $\{\breve{M}_t^k\}_{t=1}^T$ and $\{\bar{M}_t\}_{t=1}^T$ be the solutions of $K+1$ Riccati equations defined in Theorem 1. For $t \in \{1, \dots, T-1\}$, the gains $\{\breve{F}_t^k, \bar{F}_t^k\}_{t=1}^T$ are given by

$$\breve{F}_t^k = \left((B_t^k)^\intercal \breve{M}_{t+1}^k B_t^k + R_t^k \right)^{-1} B_t^{k\intercal},$$

and $\operatorname{rows}(\bar{F}_t^1, \dots, \bar{F}_t^K) := \bar{F}_t$, where

$$\bar{F}_t = \left(\bar{B}_t^\intercal \bar{M}_{t+1} \bar{B}_t + \bar{R}_t + P_t^u \right)^{-1} \bar{B}_t^\intercal.$$

The correction signals $\{v_t^i, \bar{v}_t\}_{t=1}^T$ are given recursively as follows: for $t = T$,

$$v_T^i = Q_T^k r_T^i, \quad \bar{v}_T = \bar{Q}_T \bar{\mathbf{r}}_T + P_T^x \mathbf{s}_T, \tag{22}$$

and for $t \in \{T-1, \dots, 1\}$,

$$v_t^i = (A_t^k + B_t^k \breve{L}_t^k)^\intercal v_{t+1}^i + Q_t^k r_t^i, \tag{23}$$

$$\bar{v}_t = (\bar{A}_t + \bar{B}_t \bar{L}_t)^\intercal \bar{v}_{t+1} + \bar{Q}_t \bar{\mathbf{r}}_t + P_t^x \mathbf{s}_t. \tag{24}$$

The proof is presented in Appendix C. To implement the optimal control strategies:

- all agents must compute $\bar{L}_{1:T-1}$ and $\bar{F}_{1:T-1}$ by solving Riccati equation (14) and compute the global correction signal $\bar{v}_{1:T}$ by solving backward equations (22) and (24),
- agents of sub-population k must compute $\breve{L}_{1:T-1}^k$ and $\breve{F}_{1:T-1}^k$ by solving Riccati equation (13),
- an individual agent i of sub-population k must compute a local correction signal $v_{1:T}^i$ by solving backward equations (22) and (23).

Then, an individual agent i of sub-population k, upon observing the local state x_t^i and the global mean-field $\bar{\mathbf{x}}_t$, chooses its local control action according to (21).

D. Generalization 2: weighted mean-field

Suppose there are weights (a^i, λ^i, b^i) associated with each agent $i \in \mathcal{N}$ such that $a^i, \lambda^i \in \mathbb{R}$ and $b^i \in \mathbb{R}_{>0}$. For each sub-population $k \in \mathcal{K}$ define the weighted mean-field of states and actions as follows.

$$\bar{x}_t^{k,\lambda} = \frac{1}{|\mathcal{N}^k|} \sum_{i \in \mathcal{N}^k} \lambda^i x_t^i, \qquad \bar{u}_t^{k,\lambda} = \frac{1}{|\mathcal{N}^k|} \sum_{i \in \mathcal{N}^k} \lambda^i u_t^i,$$

$$\bar{\mathbf{x}}_t^\lambda = \operatorname{vec}(\bar{x}^{1,\lambda}, \dots, \bar{x}^{K,\lambda}), \qquad \bar{\mathbf{u}}_t^\lambda = \operatorname{vec}(\bar{u}^{1,\lambda}, \dots, \bar{u}^{K,\lambda}).$$

Also, define $\bar{a}^{k,\lambda} = \frac{1}{|\mathcal{N}^k|} \sum_{i \in \mathcal{N}^k} \lambda^i a^i$. For sub-population $k \in \mathcal{K}$, the state of agent $i \in \mathcal{N}^k$ evolves as follows.

$$x_{t+1}^i = A_t^k x_t^i + B_t^k u_t^i + a^i (D_t^k \bar{\mathbf{x}}_t^\lambda + E_t^k \bar{\mathbf{u}}_t^\lambda) + w_t^i.$$

The per-step cost is given by

$$c_t(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t^\lambda, \bar{\mathbf{u}}_t^\lambda) = (\bar{\mathbf{x}}_t^\lambda)^\intercal P_t^x \bar{\mathbf{x}}_t^\lambda + (\bar{\mathbf{u}}_t^\lambda)^\intercal P_t^u \bar{\mathbf{u}}_t^\lambda + \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}^k} \frac{b^i}{|\mathcal{N}^k|} \left[(x_t^i)^\intercal Q_t^k x_t^i + (u_t^i)^\intercal R_t^k u_t^i \right],$$

and the terminal cost is given by

$$c_T(\mathbf{x}_T, \bar{\mathbf{x}}_T^\lambda) = (\bar{\mathbf{x}}_T^\lambda)^\intercal P_T^x \bar{\mathbf{x}}_T^\lambda + \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}^k} \frac{b^i}{|\mathcal{N}^k|} \left[(x_T^i)^\intercal Q_T^k x_T^i \right].$$

Such models arise in applications where the interaction between two homogeneous agents is not symmetric but depends on their weights. For example, in wireless networks, the interference caused at the base-station depends on the distance of the agents from the base-station.

In the above model, agents are not partially exchangeable. Nonetheless, we are able to explicitly identify optimal control strategies under the following assumptions.

Assumption (A4) *For each sub-population $k \in \mathcal{K}$ and each agent $i \in \mathcal{N}^k$, $a^i b^i = \lambda^i \bar{a}^{k,\lambda}$.*

Given a sub-population $k \in \mathcal{K}$, examples of weights that satisfy (A4) are: for all $i \in \mathcal{N}^k$, (i) $a^i = 0$, (ii) $a^i = 1$ and $b^i = \lambda^i$, (iii) $a^i = \lambda^i$, $b^i = 1$, and $\frac{1}{|\mathcal{N}^k|}\sum_{i\in\mathcal{N}^k} \lambda^i = 1$. To simplify the exposition, define $\mu^k \coloneqq 2 - \frac{1}{|\mathcal{N}^k|}\sum_{i\in\mathcal{N}^k} \frac{(\lambda^i)^2}{b^i}$.

Assumption (A2a) *For every t, P^x_t, P^u_t, Q^k_t, and R^k_t are symmetric matrices that satisfy*

$$\begin{aligned} Q^k_t \geq 0,\ \forall k \in \mathcal{K}, \quad \operatorname{diag}(\mu^1 Q^1_t, \dots, \mu^K Q^K_t) + P^x_t \geq 0,\\ R^k_t > 0,\ \forall k \in \mathcal{K}, \quad \operatorname{diag}(\mu^1 R^1_t, \dots, \mu^K R^K_t) + P^u_t > 0. \end{aligned}$$

Note that if $\mu^k = 1$, (A2a) reduces to (A2). Each agent has mean-field sharing information structure, i.e., agent $i \in \mathcal{N}^k$ of sub-population $k \in \mathcal{K}$ observes the local state x^i_t and the weighted mean-field $\bar{\mathbf{x}}^\lambda_t$.

Theorem 4 *Under (A1), (A2a), (A4), and (MFS-IS), the optimal strategy is unique and given by*

$$u^i_t = \breve{L}^k_t \left(x^i_t - \frac{\lambda^i}{b^i} \bar{x}^{k,\lambda}_t \right) + \frac{\lambda^i}{b^i} \bar{L}^k_t \bar{\mathbf{x}}^\lambda_t,$$

where the gains $\{\breve{L}^k_t, \bar{L}^k_t\}_{t=1}^{T-1}$ are obtained by the solution of $K+1$ Riccati equations defined in Theorem 1 when $\bar{A}_t$, $\bar{B}_t$, $\bar{Q}_t$, and $\bar{R}_t$ are replaced by

$$\begin{aligned} \bar{A}_t &\coloneqq \operatorname{diag}(A^1_t, \dots, A^K_t) + \operatorname{rows}(\bar{a}^{1,\lambda} D^1_t, \dots, \bar{a}^{K,\lambda} D^K_t),\\ \bar{B}_t &\coloneqq \operatorname{diag}(B^1_t, \dots, B^K_t) + \operatorname{rows}(\bar{a}^{1,\lambda} E^1_t, \dots, \bar{a}^{K,\lambda} E^K_t),\\ \bar{Q}_t &\coloneqq \operatorname{diag}(\mu^1 Q^1_t, \dots, \mu^K Q^K_t),\\ \bar{R}_t &\coloneqq \operatorname{diag}(\mu^1 R^1_t, \dots, \mu^K R^K_t). \end{aligned}$$

The proof is presented in Appendix D.

Remark 9 The optimal strategy depends on the weights and, even within a sub-population, the gains of the mean-field terms are different for different agents.

Remark 10 If the dynamics of the agents are decoupled, i.e., $a^i = 0$ for all agents, then the results of Theorem 4 are similar to the model with soft constraints discussed in [13].

Note that if $a^i = b^i = \lambda^i = 1$ for all agents, then the weighted mean-field model reduces to the basic model described in Proposition 1 and the result of Theorem 4 reduces to that of Theorem 1.

V. PROOF OF THEOREM 1

We start with the model presented in Proposition 1. The proof proceeds in three steps.

- **Step 1**: We use a coordinate transformation to construct a system that is isomorphic to the original system.
- **Step 2**: We construct an auxiliary system which is system of Step 1 with classical information structure (i.e., all decisions are made a single agent).
- **Step 3**: We show that the optimal control laws of the auxiliary system can be implemented using MFS-IS. A fortiori, they are also optimal for MFS-IS.

A. Step 1: A coordinate transformation

Define $\breve{x}^i_t = x^i_t - \bar{x}^k_t$ and $\breve{u}^i_t = u^i_t - \bar{u}^k_t$ and consider the following coordinate transformation $\mathcal{T}$ of the state and action spaces: $\mathcal{T} \operatorname{vec}\big((x^i_t)_{i\in\mathcal{N}}\big) = \operatorname{vec}\big((\breve{x}^i_t)_{i\in\mathcal{N}}, \bar{\mathbf{x}}_t\big)$ and $\mathcal{T} \operatorname{vec}\big((u^i_t)_{i\in\mathcal{N}}\big) = \operatorname{vec}\big((\breve{u}^i_t)_{i\in\mathcal{N}}, \bar{\mathbf{u}}_t\big)$. Under this transformation, the dynamics (5) may be written as

$$\breve{x}^i_{t+1} = A^k_t \breve{x}^i_t + B^k_t \breve{u}^i_t + \breve{w}^i_t, \tag{25}$$

where $\breve{w}^i_t \coloneqq w^i_t - \bar{w}^k_t$ and $\bar{w}^k_t \coloneqq \langle (w^i_t)_{i\in\mathcal{N}^k} \rangle$ and

$$\bar{\mathbf{x}}_{t+1} = \bar{A}_t \bar{\mathbf{x}}_t + \bar{B}_t \bar{\mathbf{u}}_t + \bar{\mathbf{w}}_t, \tag{26}$$

where $\bar{\mathbf{w}}_t \coloneqq \operatorname{vec}(\bar{w}^1_t, \dots, \bar{w}^K_t)$ and $\bar{A}_t$ and $\bar{B}_t$ are defined as in Theorem 1.

The per-step cost $c_t(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)$ and terminal cost $c_T(\mathbf{x}_T, \bar{\mathbf{x}}_T)$ can also be written in terms of the transformed variables. For that matter, we need the following result that is similar to the Parallel-Axis Theorem (or Huygens-Steiner Theorem) in mechanics [37]:

Lemma 1 *For any $\mathbf{x} = \operatorname{vec}(x^1, \dots, x^N)$ and $\bar{x} = \langle \mathbf{x} \rangle$, let $\breve{x}^i = x^i - \bar{x}$, $i \in \{1, \dots, N\}$. Then, for any matrix Q of appropriate dimension,*

$$\frac{1}{N}\sum_{i=1}^N (x^i)^\intercal Q x^i = \frac{1}{N}\sum_{i=1}^N (\breve{x}^i)^\intercal Q \breve{x}^i + \bar{x}^\intercal Q \bar{x}.$$

Proof: The result follows from elementary algebra and the observation that $\sum_{i=1}^N \breve{x}^i = 0$. ■

An immediate consequence of Lemma 1 is the following:

Corollary 4 *For time t, $t \in \{1, \dots, T\}$, there exist functions $\{\breve{c}^k_t\}_{k\in\mathcal{K}}$ and $\bar{c}_t$ such that*

$$c_t(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) = \bar{c}_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) + \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}^k} \breve{c}^k_t(\breve{x}^i_t, \breve{u}^i_t), \tag{27}$$

where

$$\begin{aligned} \bar{c}_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) &= \bar{\mathbf{x}}_t^\intercal(\bar{Q}_t + P^x_t)\bar{\mathbf{x}}_t + \bar{\mathbf{u}}_t^\intercal(\bar{R}_t + P^u_t)\bar{\mathbf{u}}_t,\\ \breve{c}^k_t(\breve{x}^i_t, \breve{u}^i_t) &= \frac{1}{|\mathcal{N}^k|}\left[(\breve{x}^i_t)^\intercal Q^k_t \breve{x}^i_t + (\breve{u}^i_t)^\intercal R^k_t \breve{u}^i_t\right], \end{aligned}$$

and for $t = T$,

$$c_T(\mathbf{x}_T, \bar{\mathbf{x}}_T) = \bar{c}_T(\bar{\mathbf{x}}_T) + \sum_{i\in\mathcal{N}^k, k\in\mathcal{K}} \breve{c}^k_T(\breve{x}^i_T), \tag{28}$$

where

$$\bar{c}_T(\bar{\mathbf{x}}_T) = \bar{\mathbf{x}}_T^\intercal(\bar{Q}_T + P^x_T)\bar{\mathbf{x}}_T, \quad \breve{c}^k_T(\breve{x}^i_T) = \frac{1}{|\mathcal{N}^k|}(\breve{x}^i_T)^\intercal Q^k_T \breve{x}^i_T.$$

Since the transformation $\mathcal{T}$ is an isomorphism, the transformed model with dynamics (25) and (26) and the per-step cost (27) and (28) is equivalent to the original model in Proposition 1, *irrespective of the information structure.*

B. Step 2: An auxiliary system

Consider an auxiliary system with state $\mathring{\mathbf{x}}_t = \text{vec}((\breve{x}_t^i)_{i\in\mathcal{N}}, \bar{\mathbf{x}}_t)$ and action $\mathring{\mathbf{u}}_t = \text{vec}((\breve{u}_t^i)_{i\in\mathcal{N}}, \bar{\mathbf{u}}_t)$ (which is same as the transformed model of Step 1). There is a *single centralized agent* that chooses $\mathring{\mathbf{u}}_t$ based on the observations. In particular, the centralized agent observes $\mathring{\mathbf{x}}_t$ and chooses $\mathring{\mathbf{u}}_t$ according to

$$\mathring{\mathbf{u}}_t = \mathring{g}_t(\mathring{\mathbf{x}}_{1:t}, \mathring{\mathbf{u}}_{1:t-1}). \tag{29}$$

The performance of strategy $\mathring{\mathbf{g}} := (\mathring{g}_1, \dots, \mathring{g}_T)$ is given by

$$\mathring{J}(\mathring{\mathbf{g}}) = \mathbb{E}^{\mathring{\mathbf{g}}}\Bigg[\sum_{t=1}^{T-1} c_t(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) + c_T(\mathbf{x}_T, \bar{\mathbf{x}}_T)\Bigg], \tag{30}$$

where the expectation is with respect to the measure induced on all system variables by the choice of strategy $\mathring{\mathbf{g}}$. We are interested in the following optimization problem.

Problem 2 *In the auxiliary system, find strategy $\mathring{\mathbf{g}}^*$ that minimizes* (30), *i.e.,*

$$\mathring{J}^* := \mathring{J}(\mathring{\mathbf{g}}^*) = \inf_{\mathring{\mathbf{g}}} \mathring{J}(\mathring{\mathbf{g}}),$$

where the infimum is taken over all strategies of the form (29).

Let J^* and $\mathring{J}^*$ denote the optimal cost for Problem 1 and Problem 2, respectively. Since the per-step cost is the same in both cases, but Problem 2 is centralized, we have that

$$J^* \geq \mathring{J}^*.$$

We identify the optimal control laws for the auxiliary system and show that these laws can be implemented in, and therefore are optimal for, the original decentralized system.

C. Step 3: The Optimal Solution of the Auxiliary System

The auxiliary system is a stochastic linear quadratic system. So, the optimal control laws are linear and the optimal gains are given by the solution of an appropriate Riccati equation. However, the dimension of the state $\mathring{\mathbf{x}}_t$, and therefore the dimension of the Riccati equation, increases with the number of agents. To overcome this challenge, we present an alternative approach that involves solving $K+1$ Riccati equations that do not depend on the number of agents.

Since the auxiliary system is a stochastic linear quadratic system, the certainty equivalence principle [38, Theorem 6.1] holds. Therefore, the optimal control law is identical to the control law of the corresponding deterministic system, whose dynamics are given as follows: for $k \in \mathcal{K}$ and $i \in \mathcal{N}^k$

$$\breve{x}_{t+1}^i = A_t^k \breve{x}_t^i + B_t^k \breve{u}_t^i, \quad \bar{\mathbf{x}}_{t+1} = \bar{A}_t \bar{\mathbf{x}}_t + \bar{B}_t \bar{\mathbf{u}}_{\mathbf{t}},$$

and whose per-step cost is $\mathring{c}_t(\mathring{\mathbf{x}}_t, \mathring{\mathbf{u}}_t)$ given by Corollary 4. Under (A2), the deterministic centralized linear quadratic system is strictly convex; hence, the solution is unique [38, Theorem 4.1].

Note that this system consists of $(N+1)$ components: N components with state $\breve{x}_t^i$ and action $\breve{u}_t^i$, $i \in \mathcal{N}$, and one component with state $\bar{\mathbf{x}}_t$ and action $\bar{\mathbf{u}}_t$. The first N components are split into K classes of identical components—one for each sub-population. The components have decoupled dynamics and decoupled cost. Thus, the optimal control law of each class may be identified separately. In particular, from [38, Theorem 4.1], we have that

Theorem 5 *The optimal control strategy of the auxiliary system (i.e., Problem 2) is unique and given by*

$$\bar{\mathbf{u}}_t = \bar{L}_t \bar{\mathbf{x}}_t \quad \textit{and for } k \in \mathcal{K}, i \in \mathcal{N}^k, \quad \breve{u}_t^i = \breve{L}_t^k \breve{x}_t^i,$$

where the gains $\{\breve{L}_t^k, \bar{L}_t\}_{t=1}^{T-1}$ are given as in Theorem 1.

Now, we transform the optimal centralized solution, given by Theorem 5, back to the original model (by taking the inverse of coordinate transformation used in Step 1), to get

$$u_t^i = \breve{u}_t^i + \bar{u}_t^k = \breve{L}_t^k (x_t^i - \bar{x}_t^k) + \bar{L}_t^k \bar{\mathbf{x}}_t.$$

Note that the above control laws are implementable under MFS-IS. Therefore, the solution of Problem 2 coincides with the solution of Problem 1 with MFS-IS.

VI. Proof of Theorem 2

A. Preliminary results

We use the same transformation as Step 1 in Section V-A. In particular, for any $k \in \mathcal{K}$ and $i \in \mathcal{N}^k$, define $\breve{x}_t^i := x_t^i - \bar{x}_t^k$, $\breve{u}_t^i := u_t^i - \bar{u}_t^k$, $\breve{s}_t^i := s_t^i - \bar{s}_t^k$ and $\breve{v}_t^i := v_t^i - \bar{v}_t^k$. Then, we have

Lemma 2 *For all t, $\breve{s}_t^i = \breve{x}_t^i$ and $\breve{u}^i = \breve{v}_t^i$. Consequently,*

$$\hat{J} - J^* = \sum_{t=1}^{T} \left[\bar{c}_t(\bar{\mathbf{s}}_t, \bar{\mathbf{v}}_t) - \bar{c}_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)\right]. \tag{31}$$

Proof: We prove the first part by induction. Note that $\breve{x}_1^i = \breve{s}_1^i$ and $\breve{v}_1^i = \breve{L}_1^k \breve{s}_1^i = \breve{L}_1^k \breve{x}_1^i = \breve{u}_1^i$. This forms the basis of induction. Now assume that $\breve{s}_t^i = \breve{x}_t^i$ and $\breve{v}_t^i = \breve{u}_t^i$ and consider time $t+1$. Then,

$$\breve{s}_{t+1}^i = A_t^k \breve{s}_t^i + B_t^k \breve{v}_t^i + \breve{w}_t^i = A_t^k \breve{x}_t^i + B_t^k \breve{u}_t^i + \breve{w}_t^i = \breve{x}_{t+1}^i.$$

Moreover, $\breve{v}_{t+1}^i = \breve{L}_{t+1}^k \breve{s}_{t+1}^i = \breve{L}_{t+1}^k \breve{x}_{t+1}^i = \breve{u}_{t+1}^i$. Thus, the result is true by induction. Equation (31) immediately follows from the first part and Corollary 4. ∎

Next we simplify (31) in terms of the following relative errors: For any $k \in \mathcal{K}$, define

$$\zeta_t^k = \bar{x}_t^k - z_t^k \quad \text{and} \quad \xi_t^k = \bar{s}_t^k - z_t^k.$$

Let $\boldsymbol{\zeta}_t = \text{vec}(\zeta_t^1, \dots, \zeta_t^K)$ and $\boldsymbol{\xi}_t = \text{vec}(\xi_t^1, \dots, \xi_t^K)$. For ease of exposition, let vector $h = \text{vec}(h^1, \dots, h^K)$ be binary such that $h^k = 0_{d_x^k \times 1}$ if $k \in \mathcal{S}$ and $h^k = \mathbb{1}_{d_x^k \times 1}$ if $k \in \mathcal{S}^c$.

Lemma 3 *Let $\tilde{A}_t$ be defined as in Theorem 2. Then, $\boldsymbol{\zeta}_1 = h \circ \bar{\mathbf{x}}_1$ and $\boldsymbol{\xi}_1 = h \circ \bar{\mathbf{x}}_1$ and*

$$\begin{bmatrix} \boldsymbol{\zeta}_{t+1} \\ \boldsymbol{\xi}_{t+1} \end{bmatrix} = \tilde{A}_t \begin{bmatrix} \boldsymbol{\zeta}_t \\ \boldsymbol{\xi}_t \end{bmatrix} + \begin{bmatrix} h \circ \bar{\mathbf{w}}_t \\ h \circ \bar{\mathbf{w}}_t \end{bmatrix}.$$

Proof: From (16) and (18), we get

$$\begin{aligned} \bar{s}_{t+1}^k &= A_t^k \bar{s}_t^k + B_t^k \bar{v}_t^k + D_t^k \bar{\mathbf{s}}_t + E_t^k \bar{\mathbf{v}}_t + \bar{w}_t^k, \\ \bar{v}_t^k &= \breve{L}_t^k (\bar{s}_t^k - z_t^k) + \bar{L}_t^k \mathbf{z}_t, \end{aligned} \tag{32}$$

where $\bar{w}_t^k := \langle (w_t^i)_{i\in\mathcal{N}^k} \rangle$. Write (32) in a vectorized form,

$$\bar{\mathbf{s}}_{t+1} = \bar{A}_t \bar{\mathbf{s}}_t + \bar{B}_t \bar{\mathbf{v}}_t + \bar{\mathbf{w}}_t, \quad \bar{\mathbf{v}}_t = \breve{L}_t \boldsymbol{\xi}_t + \bar{L}_t \mathbf{z}_t,$$

where $\bar{\mathbf{w}}_t = \text{vec}(\bar{w}_k^1, \dots, \bar{w}_t^K)$. From Theorem 1, we can write the dynamics under the optimal strategy as follows

$$\bar{x}_{t+1}^k = A_t^k \bar{x}_t^k + (B_t^k \bar{L}_t^k + D_t^k + E_t^k \bar{L}_t)\bar{\mathbf{x}}_t + \bar{w}_t^k,$$
$$\bar{u}_t^k = \bar{L}_t^k \bar{\mathbf{x}}_t,$$

and in a vectorized form,

$$\bar{\mathbf{x}}_{t+1} = (\bar{A}_t + \bar{B}_t \bar{L}_t)\bar{\mathbf{x}}_t + \bar{\mathbf{w}}_t, \quad \bar{\mathbf{u}}_t = \bar{L}_t \bar{\mathbf{x}}_t.$$

Thus, the dynamics of the relative errors can be written as follows. If $k \in \mathcal{S}$,

$$\zeta_{t+1}^k = A_t^k \zeta_t^k + (B_t^k \bar{L}_t^k + D_t^k + E_t^k \bar{L}_t)\boldsymbol{\zeta}_t - (A_t^k + B_t^k \breve{L}_t^k)\xi_t^k - (D_t^k + E_t^k \breve{L}_t)\boldsymbol{\xi}_t,$$
$$\xi_{t+1}^k = 0,$$

and if $k \in \mathcal{S}^c$,

$$\zeta_{t+1}^k = A_t^k \zeta_t^k + (B_t^k \bar{L}_t^k + D_t^k + E_t^k \bar{L}_t)\boldsymbol{\zeta}_t + \bar{w}_t^k,$$
$$\xi_{t+1}^k = (A_t^k + B_t^k \breve{L}_t^k)\xi_t^k + (D_t^k + E_t^k \breve{L}_t)\boldsymbol{\xi}_t + \bar{w}_t^k.$$

Combining these, gives the result of the Lemma. ■

Let $\mathcal{F}_t = \{\bar{s}_{1:t}^k\}_{k \in \mathcal{S}}$ be the history of the mean-fields of sub-populations $\mathcal{S}$ that are observed.

Lemma 4 *For all t, $\mathbb{E}[\boldsymbol{\zeta}_t|\mathcal{F}_t] = \mathbb{E}[\boldsymbol{\xi}_t|\mathcal{F}_t] = 0$.*

Proof: If $k \in \mathcal{S}$, $\zeta_1^k = \xi_1^k = 0$ and if $k \in \mathcal{S}^c$, $\zeta_1^k = \xi_1^k = \bar{x}_1^k$, and from (A1a), $\mathbb{E}[\bar{x}_1^k|\mathcal{F}_1] = \mathbb{E}[\bar{x}_1^k] = 0$. Therefore, $\mathbb{E}[\boldsymbol{\zeta}_1|\mathcal{F}_1] = \mathbb{E}[\boldsymbol{\xi}_1|\mathcal{F}_1] = 0$. Thus, from Lemma 3 and $\mathbb{E}[\bar{\mathbf{w}}_t|\mathcal{F}_t] = 0$, we get that $\mathbb{E}[\boldsymbol{\zeta}_t|\mathcal{F}_t] = \mathbb{E}[\boldsymbol{\xi}_t|\mathcal{F}_t] = 0$. ■

Lemma 5 *$\mathbf{z}_t$ is measurable with respect to $\mathcal{F}_t$, therefore, $\mathbb{E}[\mathbf{z}_t|\mathcal{F}_t] = \mathbf{z}_t$.*

Proposition 2 *The relative loss is given*

$$\hat{J} - J^* = \mathbb{E}\left[\sum_{t=1}^{T} [\boldsymbol{\zeta}_t \quad \boldsymbol{\xi}_t]^\intercal \tilde{Q}_t [\boldsymbol{\zeta}_t \quad \boldsymbol{\xi}_t]\right].$$

Proof: Recall that $\bar{c}_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) = \bar{\mathbf{x}}_t^\intercal(\bar{Q}_t + P_t^x)\bar{\mathbf{x}}_t + \bar{\mathbf{u}}_t^\intercal(\bar{R}_t + P_t^u)\bar{\mathbf{u}}_t$. The proof follows immediately from (31) and the following observation:

Lemma 6 *Let $\hat{Q}_t \coloneqq \bar{Q}_t + P_t^x$ and $\hat{R}_t \coloneqq \bar{R}_t + P_t^u$. Then,*

$$\mathbb{E}[\bar{\mathbf{s}}_t^\intercal \hat{Q}_t \bar{\mathbf{s}}_t - \bar{\mathbf{x}}_t^\intercal \hat{Q}_t \bar{\mathbf{x}}_t|\mathcal{F}_t] = \mathbb{E}[\boldsymbol{\xi}_t^\intercal \hat{Q}_t \boldsymbol{\xi}_t - \boldsymbol{\zeta}_t^\intercal \hat{Q}_t \boldsymbol{\zeta}_t|\mathcal{F}_t],$$

and

$$\mathbb{E}[\bar{\mathbf{v}}_t^\intercal \hat{R}_t \bar{\mathbf{v}}_t - \bar{\mathbf{u}}_t^\intercal \hat{R}_t \bar{\mathbf{u}}_t|\mathcal{F}_t] = \mathbb{E}[\boldsymbol{\xi}_t^\intercal \breve{L}_t^\intercal \hat{R}_t \breve{L}_t \boldsymbol{\xi}_t|\mathcal{F}_t] - \mathbb{E}[\boldsymbol{\zeta}_t^\intercal \bar{L}_t^\intercal \hat{R}_t \bar{L}_t \boldsymbol{\zeta}_t|\mathcal{F}_t].$$

Therefore, the proof of Proposition 2 is complete. ■

Proof of Lemma 6:

1) Substituting $\bar{\mathbf{s}}_t = \boldsymbol{\xi}_t + \mathbf{z}_t$ and $\bar{\mathbf{x}}_t = \boldsymbol{\zeta}_t + \mathbf{z}_t$, we get

$$\begin{aligned}&\mathbb{E}[\bar{\mathbf{s}}_t^\intercal \hat{Q}_t \bar{\mathbf{s}}_t - \bar{\mathbf{x}}_t^\intercal \hat{Q}_t \bar{\mathbf{x}}_t|\mathcal{F}_t] \\ &\overset{(a)}{=} \mathbb{E}[\boldsymbol{\xi}_t^\intercal \hat{Q}_t \boldsymbol{\xi}_t - \boldsymbol{\zeta}_t^\intercal \hat{Q}_t \boldsymbol{\zeta}_t|\mathcal{F}_t] + 2\mathbb{E}[\boldsymbol{\xi}_t^\intercal \hat{Q}_t \mathbf{z}_t|\mathcal{F}_t] - 2\mathbb{E}[\boldsymbol{\zeta}_t^\intercal \hat{Q}_t \mathbf{z}_t|\mathcal{F}_t] \\ &= \mathbb{E}[\boldsymbol{\xi}_t^\intercal \hat{Q}_t \boldsymbol{\xi}_t - \boldsymbol{\zeta}_t^\intercal \hat{Q}_t \boldsymbol{\zeta}_t|\mathcal{F}_t],\end{aligned}$$

where the last two terms in (a) are zero by Lemmas 4 and 5.

2) Substituting $\bar{\mathbf{v}}_t = \breve{L}_t \boldsymbol{\xi}_t + \bar{L}_t \mathbf{z}_t$ and $\bar{\mathbf{u}}_t = \bar{L}_t \bar{\mathbf{x}}_t = \bar{L}_t(\boldsymbol{\zeta}_t + \mathbf{z}_t)$, we get

$$\begin{aligned}&\mathbb{E}[\bar{\mathbf{v}}_t^\intercal \hat{R}_t \bar{\mathbf{v}}_t - \bar{\mathbf{u}}_t^\intercal \hat{R}_t \bar{\mathbf{u}}_t|\mathcal{F}_t] \\ &\overset{(b)}{=} \mathbb{E}[\boldsymbol{\xi}_t^\intercal \breve{L}_t^\intercal \hat{R}_t \breve{L}_t \boldsymbol{\xi}_t - \boldsymbol{\zeta}_t^\intercal \bar{L}_t^\intercal \hat{R}_t \bar{L}_t \boldsymbol{\zeta}_t|\mathcal{F}_t] + 2\mathbb{E}[\boldsymbol{\xi}_t^\intercal \breve{L}_t^\intercal \hat{R}_t \bar{L}_t \mathbf{z}_t|\mathcal{F}_t] - 2\mathbb{E}[\boldsymbol{\zeta}_t^\intercal \bar{L}_t \hat{R}_t \bar{L}_t \mathbf{z}_t|\mathcal{F}_t] \\ &= \mathbb{E}[\boldsymbol{\xi}_t^\intercal \breve{L}_t^\intercal \hat{R}_t \breve{L}_t \boldsymbol{\xi}_t - \boldsymbol{\zeta}_t^\intercal \bar{L}_t^\intercal \hat{R}_t \bar{L}_t \boldsymbol{\zeta}_t|\mathcal{F}_t],\end{aligned}$$

where the last two terms in (b) are zero by Lemmas 4 and 5. ■

B. Proof of Theorem 2

To prove part 1, note that $\hat{J} - J^*$ is the expected total quadratic cost (given by Proposition 2) of a linear (uncontrolled) system (given by Lemma 3). Thus, $\hat{J} - J^*$ is given by (19) where $\tilde{M}_{1:T}$ is the solution of the Lyapunov equation (20). Note that the variance of the initial state and noises in Lemma 3 are given as follows:

$$\text{var}(h \circ \bar{\mathbf{x}}_1, h \circ \bar{\mathbf{x}}_1) = \mathbb{1}_{2d_x \times 2d_x} \otimes [H \circ \text{var}(\bar{\mathbf{x}}_1)] =: \tilde{X}_1,$$
$$\text{var}(h \circ \bar{\mathbf{w}}_t, h \circ \bar{\mathbf{w}}_t) = \mathbb{1}_{2d_x \times 2d_x} \otimes [H \circ \text{var}(\bar{\mathbf{w}}_t)] =: \tilde{W}_t.$$

To prove part 2 of Theorem 2, first observe that due to (A3), matrices $\tilde{A}_t$ and $\tilde{Q}_t$ do not depend on $(|\mathcal{N}^1|, \dots, |\mathcal{N}^K|)$; therefore, neither does $\tilde{M}_{1:T}$. Thus the only dependence on the size of the sub-population is due to $\tilde{X}_1$ and $\tilde{W}_t$. Under (A1b) and (A3), for any sub-population $k \in \mathcal{K}$,

$$\text{var}(\bar{x}_1^k) = \frac{1}{|\mathcal{N}^k|^2} \sum_{i \in \mathcal{N}^k} \text{var}(x_1^i) \le \frac{c_x^k}{n},$$
$$\text{var}(\bar{w}_t^k) = \frac{1}{|\mathcal{N}^k|^2} \sum_{i \in \mathcal{N}^k} \text{var}(w_t^i) \le \frac{c_w^k}{n}.$$

From (A1b), $\text{var}(\bar{\mathbf{x}}_1) = \text{diag}(\text{var}(\bar{x}_1^1), \dots, \text{var}(\bar{x}_1^K))$ and $\text{var}(\bar{\mathbf{w}}_t) = \text{diag}(\text{var}(\bar{w}_t^1), \dots, \text{var}(\bar{w}_t^K))$. Thus,

$$\tilde{X}_1 \le \frac{1}{n} \mathbb{1}_{2d_x \times 2d_x} \otimes \left[H \circ \text{diag}(c_x^1, \dots, c_x^K)\right],$$
$$\tilde{W}_t \le \frac{1}{n} \mathbb{1}_{2d_x \times 2d_x} \otimes \left[H \circ \text{diag}(c_w^1, \dots, c_w^K)\right].$$

Thus, $\tilde{X}_1$ and $\tilde{W}_t$ are $\mathcal{O}(\frac{1}{n})$. From (19), we have

$$|\hat{J} - J^*| \le \left|\text{Tr}\left(\tilde{X}_1 \tilde{M}_1\right)\right| + \sum_{t=1}^{T-1} \left|\text{Tr}\left(\tilde{W}_t \ \tilde{M}_{t+1}\right)\right|,$$

where each of above absolute values is $\mathcal{O}(\frac{1}{n})$. In particular, since $\tilde{X}_1$ and $\tilde{W}_t$ are $\mathcal{O}(\frac{1}{n})$ and $\tilde{M}_{1:T}$ do not depend on n, $|\text{Tr}(\tilde{X}_1 \tilde{M}_1|)$ and $|\text{Tr}(\tilde{W}_t \tilde{M}_{t+1})|$ are $\mathcal{O}(\frac{1}{n})$.

VII. Infinite Horizon

The results presented in Sections III and IV generalize to infinite horizon setup in a natural manner. Assume that the model is time-invariant, i.e., the matrices $\{A_t^k, B_t^k, D_t^k, E_t^k, Q_t^k, R_t^k, P_t^x, P_t^u\}$ and covariances $\{\breve{\Sigma}_t^k, \bar{\Sigma}_t, \breve{\Xi}_t^k, \bar{\Xi}_t\}$ (defined in Theorem 1) do not depend on time; hence, we remove the subscript t. The rest of the model is as same as that in Section II-A.

Consider the infinite horizon discounted cost and the infinite horizon long-term average setups as follows:

Problem 3 *Given discount factor $\beta \in (0,1)$, find a strategy* $\mathbf{g}$ *that minimizes the following cost:*

$$J_\beta(\mathbf{g}) = (1-\beta)\mathbb{E}^{\mathbf{g}}\left[\sum_{t=1}^{\infty} \beta^{t-1} c(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)\right],$$

where the expectation is with respect to the measure induced on all the system variables by the choice of strategy $\mathbf{g}$.

Problem 4 *Find a strategy* $\mathbf{g}$ *that minimizes the following cost:*

$$J_1(\mathbf{g}) = \lim_{T\to\infty} \mathbb{E}^{\mathbf{g}}\left[\frac{1}{T}\sum_{t=1}^{T} c(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)\right],$$

where the expectation is with respect to the measure induced on all the system variables by the choice of strategy $\mathbf{g}$.

Assumption (A5) *For each sub-population $k \in \mathcal{K}$, $(\sqrt{\beta}A^k, \sqrt{\beta}B^k)$ are stabilizable and $(\sqrt{\beta}A^k, \sqrt{Q^k})$ are detectable. In addition, for $\bar{A}_t$ and $\bar{B}_t$ defined in Theorem 1, $(\sqrt{\beta}\bar{A}, \sqrt{\beta}\bar{B})$ are stabilizable and $(\sqrt{\beta}\bar{A}, \sqrt{\bar{Q}+P^x})$ are detectable.*

A. Exact solution for MFS-IS

The optimal strategy under MFS-IS is as follows.

Theorem 6 *Under (A1), (A2), (A5), and (MFS-IS), the optimal strategy for Problems 3 and 4 are linear and time homogeneous and are given by*

$$u_t^i = \breve{L}^k(x_t^i - \bar{x}_t^k) + \bar{L}^k \bar{\mathbf{x}}_t, \tag{33}$$

where the gains $\{\breve{L}^k, \bar{L}^k\}$ are obtained by the solution of $K+1$ algebraic Riccati equations given below: one for computing each $\breve{L}^k, k \in \mathcal{K}$, and one for $\bar{L} := \operatorname{rows}(\bar{L}^1, \dots, \bar{L}^K)$. Let matrices $\bar{A}, \bar{B}, \bar{Q}$, and $\bar{R}$ be defined as in Theorem 1; then, given $\beta \in (0,1]$,

$$\breve{L}^k = -\left(B^{k\intercal}\breve{M}^k B^k + \beta^{-1}R^k\right)^{-1} B^{k\intercal}\breve{M}^k A^k,$$
$$\bar{L} = -\left(\bar{B}^\intercal \bar{M}\bar{B} + \beta^{-1}(\bar{R}+P^u)\right)^{-1} \bar{B}^\intercal \bar{M}\bar{A},$$

where $\breve{M}^k$ and $\bar{M}$ are the solutions of the following algebraic Riccati equations:

$$\breve{M}^k = \operatorname{DARE}_\beta(A^k, B^k, Q^k, R^k),$$
$$\bar{M} = \operatorname{DARE}_\beta(\bar{A}, \bar{B}, \bar{Q}+P^x, \bar{R}+P^u).$$

In addition, the optimal performance is given by

$$J_\beta^* = (1-\beta)\left[\sum_{k\in\mathcal{K}} \operatorname{Tr}\left(\breve{\Xi}^k \breve{M}^k\right) + \operatorname{Tr}(\bar{\Xi}\bar{M})\right] + \left[\sum_{k\in\mathcal{K}} \operatorname{Tr}\left(\breve{\Sigma}^k \breve{M}^k\right) + \operatorname{Tr}(\bar{\Sigma}\bar{M})\right],$$

where $\breve{\Sigma}^k$, $\bar{\Sigma}$, $\breve{\Xi}^k$, and $\bar{\Xi}$ are defined as in Theorem 1.

Proof: The proof follows along the same lines of the proof of Theorem 1. We construct an auxiliary system as in Section V, which consists of $|\mathcal{N}|+1$ components with decoupled cost and dynamics coupled only through the noise. Since the costs are infinite-horizon discounted and infinite-horizon long run average, the optimal solution is given by appropriate algebraic Riccati equations.[9] Under (A2) and (A5), these Riccati equations have a unique solution [38, Theorem 9.2]. ∎

B. Approximate solution for PMFS-IS

In this section, we propose an approximately optimal strategy for Problems 3 and 4 under PMFS-IS. Let $\breve{L} = \operatorname{diag}(\breve{L}^1, \dots, \breve{L}^K)$ denote a diagonal matrix with diagonal terms of $\breve{L}^k$ defined as in Theorem 6. We impose the following assumption.

Assumption (A6) *$\sqrt{\beta}(\bar{A}+\bar{B}\breve{L})$ is Hurwitz matrix.*

Let $\hat{J}_\beta$ denote the performance of strategy (33) where $\bar{\mathbf{x}}_t$ is replaced by $\mathbf{z}_t$ in (17) and J_β^* denote the optimal performance under MFS-IS. Then, the difference in performance $\hat{J}_\beta - J_\beta^*$ is bounded. In particular, we have the following

Theorem 7 *Assume (A1a), (A2), (A5), (A6) and (PMFS-IS). Then, for $\beta \in (0,1]$, we have*

1) *The performance loss is given by*

$$\hat{J}_\beta - J_\beta^* = (1-\beta)\operatorname{Tr}\left(\tilde{X}_1\tilde{M}\right) + \operatorname{Tr}\left(\tilde{W}\tilde{M}\right), \tag{34}$$

where $\tilde{X}_1$ and $\tilde{W}$ are time-homogeneous and defined as in Theorem 2 and $\tilde{M}$ is the solution of following algebraic Lyapunov equation:

$$\tilde{M} = \operatorname{DALE}_\beta(\tilde{A}, \tilde{Q}), \tag{35}$$

where $\tilde{A}$ and $\tilde{Q}$ are defined as in Theorem 2 and $\breve{L} = \operatorname{diag}(\breve{L}^1, \dots, \breve{L}^K)$ and $\bar{L}$ are computed as in Theorem 6.

2) *Let $n = \min_{k\in\mathcal{S}^c}(|\mathcal{N}^k|)$. Under (A1b) and (A3),*

$$\hat{J}_\beta - J_\beta^* \in \mathcal{O}\left(\frac{1}{n}\right).$$

Proof: The proof follows along the same lines of the proof of Theorem 2. In particular, under (A5) and (A6), $\sqrt{\beta}\tilde{A}$ of Proposition 2 is Hurwtiz; hence, the performance loss may be computed by the associated algebraic Lyapunov equation given by (35). Note that even though $\tilde{Q}$ is not positive semi-definite, the algebraic Lyapunov equation has a solution [39]. The proof of part 2 of Theorem 7 follows from (34) and observation that (i) $\tilde{M}$ given by (35) does not depend on n due to (A3); (ii) $(\tilde{X}_1, \tilde{W})$ are $\mathcal{O}(1/n)$ due to (A1b). ∎

Remark 11 Assumption (A6) is always satisfied if $D_t^k = 0$ and $E_t^k = 0$ for all $k \in \mathcal{K}$. In this case, $\sqrt{\beta}(\bar{A}+\bar{B}\breve{L})$ is

$$\operatorname{diag}(\sqrt{\beta}(A^1+B^1\breve{L}^1), \dots, \sqrt{\beta}(A^K+B^K\breve{L}^K)),$$

where each of the diagonal terms are Hurwitz by definition of $\breve{L}^k$ given in Theorem 6.

[9]Note that an infinite-horizon discounted problem with 4-tuple (A,B,Q,R) and discount factor β is equivalent to an undiscounted problem with 4-tuple $(\sqrt{\beta}A, \sqrt{\beta}B, Q, R)$.

VIII. Numerical Example

To illustrate our results, we consider an example that is motivated by demand response in power systems. In demand response, the volatility in renewable generation is compensated by making small changes in the demand of a large number of loads. We model the load dynamics according to a model proposed in [40], but consider a different per-step cost.

Consider a population $\mathcal{N}$ of space heaters that can be partitioned into K disjoint sub-populations $\mathcal{N}^k$, $k \in \mathcal{K} := \{1, \dots, K\}$. Each sub-population corresponds to a particular type of space heater that have similar physical characteristics such time response and nominal temperature. For space heater $i, i \in \mathcal{N}$, the state x^i_t denotes the room temperature at time t. Consider a nominal temperature x^k_{nom} for sub-population k, $k \in \mathcal{K}$, and let u^k_{nom} be the control input needed to maintain the room temperature at x^k_{nom}. Following [40], we linearize the dynamics of sub-population k around x^k_{nom}, i.e.,

$$x^i_{t+1} - x^k_{nom} = a^k(x^i_t - x^k_{nom}) + b^k u^i_t + w^i_t,$$

where u^i_t is the control input *in addition* to u^k_{nom} and w^i_t is a random disturbance. We assume u^k_{nom} is large enough such that $(u^i_t + u^k_{nom})$ is positive.

Let x^i_{des} denote the desired temperature of user i. It is assumed that the mean desired temperature $\bar{\mathbf{x}}_{des} = \text{vec}(\bar{x}^1_{des}, \dots, \bar{x}^K_{des})$ is known to everyone (e.g., independent system operator (ISO) could compute it and broadcast the mean value to everyone or it could be computed in a distributed manner using a consensus algorithm). For the purpose of demand response, time is divided into epochs of length T. At the beginning of each epoch, a central authority such as an ISO generates a reference mean temperature m_{ref} and broadcasts it to all users.

During an epoch, all users collectively minimize the total expected cost $\mathbb{E}[\sum_{t=1}^{T} c_t]$, where the per-step cost c_t is given by

$$\frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}}\Big[q(x^i_t - x^i_{des})^2 + r {u^i_t}^2\Big] + \frac{t}{T} p(m_t - m_{ref})^2,$$

where $m_t = \big(\sum_{i\in\mathcal{N}} x^i_t\big)/|\mathcal{N}|$. The rationale for the per-step cost is that we penalize deviations from the desired temperature (which corresponds to the user's comfort level), the control effort, and deviation of the mean temperature from the reference prescribed by the ISO. The weight $\frac{t}{T}$ is so that we linearly add more weight to meeting global preference.

The above problem is an optimal tracking problem and the optimal strategy is given by Theorem 3. As an example, we consider the following values of the parameters: $K = 2$, $p = 30$, $q = 2$, $r = 50$, $x^i_{des} = x^i_1 \sim \text{Normal}(20, 3)$, $w^i_t \sim \text{Normal}(0, 0.01)$, and

$$\begin{aligned} &|\mathcal{N}^1| = 40, && a^1 = 0.5, && b^1 = 1.5, && x^1_{nom} = 20,\\ &|\mathcal{N}^2| = 100, && a^2 = 0.8, && b^2 = 1.0, && x^2_{nom} = 20. \end{aligned}$$

and consider three epochs. In the first epoch, $1 \le t \le 50$, there is no reference signal and the space heaters are operating around their local set temperatures; in the second epoch, $50 < t \le 150$, $m_{ref} = 21$; in the third epoch, $150 < t \le 250$, $m_{ref} = 19$. The resultant trajectories of a subset of the users are shown in Fig. 1.

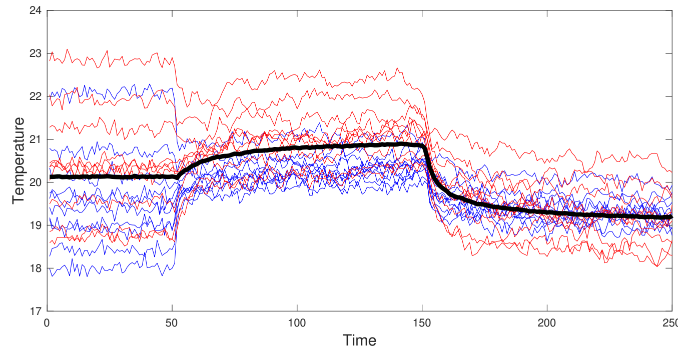

Fig. 1. Demand response with a population of 140 space heaters consisting of two sub-populations of size 40 and 100. In the initial phase, $1 \le t \le 50$, the system is uncontrolled. In the first epoch $50 < t \le 150$, the system tracks a mean reference temperature of $m_{ref} = 21$; in the second epoch $150 < t \le 250$, the system tracks a mean reference temperature of $m_{ref} = 19$. The thin lines show the local temperature of 20 out of the 140 space heaters, where blue lines corresponds to the first sub-population and red lines correspond to the second sub-population. The thick black line shows the mean-temperature achieved by the optimal strategy.

IX. Conclusion

We presented team optimal control of a decentralized system with partially exchangeable agents. Partial exchangeability implies that such a system is equivalent to one where the dynamics and the cost are coupled only through the mean-field. Our two main results are as follows. First, when the mean field is observed by all agents (the MFS information structure), the linear control laws are optimal and the corresponding gains are computed by solving $K + 1$ Riccati equations, where K is the number of sub-populations. The dimensions of these Riccati equations are independent of the size of sub-populations; consequently, the solution complexity depends only on the number K of sub-populations (rather than the size of the entire population). Second, when the mean-field of a (possibly empty) subset of sub-populations is observed by all agents (the PMFS information structure), a linear control law based on certainty equivalence is approximately optimal.

An important practical implication of these results is that they do not suffer from the curse of dimensionality. In fact, under assumption (A3), the solution does not even depend on the number of agents and the optimal gains can be computed without being aware of the size of each sub-population. Consequently, the solution methodology generalizes to the setup where the agents in a sub-population arrive and depart according to an exogenous process (e.g. number of electric vehicles plugged in for charging in smart grids).

The *raison d'etre* for investigating decentralized systems is that it is not possible—either physically or economically—to send all the state observation to a centralized controller. We show that when agents are partially exchangeable, we may circumvent the conceptual difficulties of decentralized control and achieve the centralized performance by sharing only the mean-field. Moreover, in view of the results of PMFS-IS, one may even decide not to share the mean-field of large sub-populations because there is only a small loss in performance in using the approximate value of the mean-field instead.

Throughout this paper, we assumed that when the mean-field is observed, it is observed without noise. In practice (especially if the mean-field is computed using a consensus algorithm), the mean-field will be observed with noise (and the noise will be different across agents). Our results show that if all sub-populations are large, such an observation noise will not matter. (In fact, the agents may completely ignore the mean-field observations and use the approximate values instead). However, if some of the sub-populations are small, the solution approach is not obvious. In particular, in the special case when all sub-populations have one agent, the problem reduces to the general decentralized control problem with non-classical information structure. Identifying a solution methodology for this general case remains a challenging research direction.

References

[1] J. Arabneydi and A. Mahajan, "Team-optimal solution of finite number of mean-field coupled LQG subsystems," *IEEE Conference on Decision and Control*, pp. 5308 – 5313, Dec. 2015.

[2] H. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal Of Cont. And Opt.*, vol. 6, pp. 131–147, Dec. 1968.

[3] P. Whittle and J. Rudge, "The optimal linear solution of a symmetric team control problem," *J. of Applied Probability*, pp. 377–381, 1974.

[4] A. Mahajan, N. C. Martins, M. C. Rotkowitz, and S. Yuksel, "Information structures in optimal decentralized control," *Proc. of Conf. on Decision and Control (CDC)*, pp. 1291–1306, Dec. 2012.

[5] Y. Ouyang and D. Teneketzis, "Signaling for decentralized routing in a queueing network," *Annals of Operations Research*, pp. 1–39, 2015.

[6] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order LTI systems," *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2013–2025, Sept 2011.

[7] L. Lessard and S. Lall, "Optimal control of two-player systems with output feedback," *IEEE Trans. Autom. Control*, vol. 60, no. 8, pp. 2129–2144, 2015.

[8] H. Witsenhausen, "Separation of estimation and control for discrete time systems," *Proc. of IEEE*, vol. 59, no. 11, pp. 1557–1566, Nov. 1971.

[9] Y. C. Ho and K. H. Chu, "Team decision theory and information structures in optimal control problems–part I," *IEEE Trans. Autom. Control*, vol. 17, no. 1, pp. 15–22, 1972.

[10] B. Bamieh and P. G. Voulgaris, "A convex characterization of distributed control problems in spatially invariant systems with communication constraints," *Sys. & Control Letters*, vol. 54, no. 6, pp. 575–583, 2005.

[11] M. Rotkowitz and S. Lall, "A characterization of convex problems in decentralized control," *IEEE Trans. Autom. Control*, vol. 51, no. 2, pp. 274–286, 2006.

[12] D. S. Bernstein, S. Zilberstein, and N. Immerman, "The complexity of decentralized control of markov decision processes," in *Conf. Uncert. in Artificial Intelligence (UAI)*, Stanford, CA, Jun. 2000, pp. 32–27.

[13] D. Madjidian and L. Mirkin, "Distributed control with low-rank coordination," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 53—63, 2014.

[14] J. Yong, "Linear-quadratic optimal control problems for mean-field stochastic differential equations," *SIAM on Control and Optimization*, vol. 51, no. 4, pp. 2809–2838, 2013.

[15] R. Elliott, X. Li, and Y.-H. Ni, "Discrete time mean-field stochastic linear-quadratic optimal control problems," *Automatica*, vol. 49, no. 11, pp. 3222–3233, 2013.

[16] J. Ma and J. Yong, "Forward-backward stochastic differential equations and their applications," *Springer Science & Business Media*, no. 1702, 1999.

[17] R. Singh, P. R. Kumar, and L. Xie, "The ISO problem: Decentralized stochastic control via bidding schemes," *CoRR*, vol. arXiv:1510.00983, 2015.

[18] Y. Takahara, "Multi-level approach to dynamic optimization," *Technical report, DTIC Document*, 1964.

[19] G. Cohen, "On an algorithm of decentralized optimal control," *Journal Math. Anal. and Appl.*, vol. 59, no. 2, pp. 242–259, 1977.

[20] 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 *IEEE Conference on Decision and Control*, 2003, pp. 98–103.

[21] J.-M. Lasry and P.-L. Lions, "Jeux á champ moyen. I – le cas stationnaire," *C. R. Acad. Sci. Paris, Ser. I*, vol. 343, pp. 619–625, 2006.

[22] ——, "Jeux á champ moyen. II – horizon fini et contrôle optimal," *C. R. Acad. Sci. Paris, Ser. I*, vol. 343, pp. 679–684, 2006.

[23] 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, 2007.

[24] 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, 2008.

[25] P. Caines, "Mean-field games," in *Encyclopedia of Systems and Control*, T. Samad and J. Baillieul, Eds. Springer-Verlag, Oct. 2013.

[26] D. A. Gomes and J. Saude, "Mean field games models: A brief survey," *Springer, Dynamic Games and Appl.*, vol. 4, pp. 1–45, Jun. 2014.

[27] J. Moon and T. Basar, "Linear quadratic risk-sensitive and robust mean field games," *IEEE Trans. Autom. Control*, vol. 62, no. 6, 2017.

[28] A. Bensoussan, K. C. J. Sung, S. C. P. Yam, and S. P. Yung, "Linear-quadratic mean field games," *Journal of Optimization Theory and Applications*, vol. 169, no. 2, pp. 496–529, 2016.

[29] M. Huang, P. E. Caines, and R. P. Malhamé, "Social optima in mean field LQG control: centralized and decentralized strategies," *IEEE Trans. Autom. Control*, vol. 57, no. 7, pp. 1736–1751, 2012.

[30] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, pp. 857–881, Sept. 1962.

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

[32] M. Huang, "Large-population LQG games involving a major player: the Nash certainty equivalence principle," *SIAM Journal on Control and Optimization*, vol. 48, no. 5, pp. 3318–3353, 2010.

[33] P. E. Caines and A. C. Kizilkale, "Mean field estimation for partially observed LQG systems with major and minor agents," in *IFAC World Congress*, 2014.

[34] J. Huang, S. Wang, and Z. Wu, "Mean field linear-quadratic-Gaussian LQG Games: Major and minor players," *arXiv preprint arXiv:1403.3999*, 2014.

[35] J. Arabneydi and A. Mahajan, "Team optimal control of coupled major-minor subsystems with mean-field sharing," in *Indian Control Conference (ICC)*, Jan. 2015.

[36] S. M. Asghari and A. Nayyar, "Decentralized control problems with substitutable actions," *arXiv preprint arXiv:1601.02250*, 2016.

[37] T. R. Kane and D. A. Levinson, "Dynamics, theory and applications," *McGraw Hill*, 1985.

[38] P. E. Caines, "Linear stochastic systems," *John Wiley & Sons, Inc.*, 1987.

[39] B. Hassibi, A. H. Sayed, and T. Kailath, "Indefinite-quadratic estimation and control: A unified approach to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ theories," *SIAM*, 1999.

[40] A. C. Kizilkale and R. P. Malhame, "Collective target tracking mean field control for Markovian jump-driven models of electric water heating loads," in *IFAC World Congress*, 2014.

Appendix

A. Proof of Proposition 1

Let $A_t^{i,j}$ denote the (i,j)-th block of matrix A_t. We use a similar notation for other matrices as well. Fix a sub-population $k, k \in \mathcal{K}$. If we exchange agents $i, j \in \mathcal{N}^k$, then property 2 of exchangeability implies that $A_t^{i,i} = A_t^{j,j}$ and for any other agent $n \in \mathcal{N}$, $A_t^{i,n} = A_t^{j,n}$ and $A_t^{n,i} = A_t^{n,j}$. (Similar relationships hold for B_t as well). Property 3 implies that $Q_t^{i,i} = Q_t^{j,j}$, $Q_t^{i,n} = Q_t^{j,n}$, and $Q_t^{n,i} = Q_t^{n,j}$. (Similar relationships hold for R_t as well). Define the following:

- For $i, j \in \mathcal{N}^k, A_t^{i,i} = A_t^{j,j}$ and $B_t^{i,i} = B_t^{j,j}$. Denote these by a_t^k and b_t^k, respectively.
- For $i, j \in \mathcal{N}^k$ and $n, m \in \mathcal{N}^l, l \neq k, A_t^{i,n} = A_t^{j,m}$ and $B_t^{i,n} = B_t^{j,m}$. Denote these by $d_t^{k,l}$ and $e_t^{k,l}$, respectively.
- For $i, j \in \mathcal{N}^k, Q_t^{i,i} = Q_t^{j,j}$ and $R_t^{i,i} = R_t^{j,j}$. Denote these by q_t^k and r_t^k, respectively.
- For $i, j \in \mathcal{N}^k$ and $n, m \in \mathcal{N}^l, l \neq k, Q_t^{i,n} = Q_t^{j,m}$ and $R_t^{i,n} = R_t^{j,m}$. Denote these by $p_t^{x,k,l}$ and $p_t^{u,k,l}$, respectively.

Now, consider the dynamics according to (2), the dynamics of agent i of sub-population k can be written as

$$x^i_{t+1} = A^{i\cdot}\mathbf{x}_t + B^{i\cdot}\mathbf{u}_t + w^i_t, \tag{36}$$

where $A^{i\cdot}$ and $B^{i\cdot}$ denote the rows corresponding to the ith block of A_t and B_t. Note that

$$\begin{aligned} A^{i\cdot}\mathbf{x}_t &= A^{i,i}_t x^i_t + \sum_{j\in\mathcal{N}^k, j\neq i} A^{i,j}_t x^j_t + \sum_{l\in\mathcal{K}, l\neq k}\sum_{n\in\mathcal{N}^l} A^{i,n} x^n_t \\ &= a^k_t x^i_t + d^{k,k}_t \sum_{j\in\mathcal{N}^k, j\neq i} x^j_t + \sum_{l\in\mathcal{K}, l\neq k} d^{k,l}_t \sum_{n\in\mathcal{N}^l} x^n_t \\ &= a^k_t x^i_t + d^{k,k}_t(|\mathcal{N}^k|\bar{x}^k_t - x^i_t) + \sum_{l\in\mathcal{K}, l\neq k} d^{k,l}_t |\mathcal{N}^l|\bar{x}^l_t \\ &=: A^k_t x^i_t + \sum_{l\in\mathcal{K}} D^{k,l}_t \bar{x}^l_t, \end{aligned} \tag{37}$$

where $A^k_t = a^k_t - d^{k,k}_t$ and $D^{k,l}_t = |\mathcal{N}^l| d^{k,l}_t$. By a similar algebra, we can define B^k_t and $E^{k,l}_t$ such that

$$B^{i\cdot}_t\mathbf{u}_t = B^k_t u^i_t + \sum_{l\in\mathcal{K}} E^{k,l}_t \bar{u}^l_t, \tag{38}$$

where $B^k_t = b^k_t - e^{k,k}_t$ and $E^{k,l}_t = |\mathcal{N}^l| e^{k,l}_t$. Substituting (37) and (38) in (36), we get (5). Now consider the per-step cost given by (3). Note that

$$\begin{aligned} \mathbf{x}_t^\intercal Q_t \mathbf{x}_t &= \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}}\sum_{i\in\mathcal{N}^k}\sum_{j\in\mathcal{N}^l} (x^i_t)^\intercal Q^{i,j}_t x^j_t \\ &= \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}, l\neq k}\sum_{i\in\mathcal{N}^k}\sum_{j\in\mathcal{N}^l} (x^i_t)^\intercal p^{x,k,l}_t x^j_t \\ &\quad + \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}^k}\sum_{j\in\mathcal{N}^k, j\neq i} (x^i_t)^\intercal p^{x,k,k}_t x^j_t \\ &\quad + \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}^k} (x^i_t)^\intercal q^k_t x^i_t \\ &= \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}, l\neq k} |\mathcal{N}^k||\mathcal{N}^l| (\bar{x}^k_t)^\intercal p^{x,k,l}_t \bar{x}^l_t \\ &\quad + \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}^k}\sum_{j\in\mathcal{N}^k} |\mathcal{N}^k|^2 (\bar{x}^k_t)^\intercal p^{x,k,k}_t \bar{x}^k_t \\ &\quad - \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}^k} (x^i_t)^\intercal p^{x,k,k}_t x^i_t + \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}^k} (x^i_t)^\intercal q^k_t x^i_t \\ &= \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{K}} |\mathcal{N}^k||\mathcal{N}^l| (\bar{x}^k_t)^\intercal p^{x,k,l}_t \bar{x}^l_t \\ &\quad + \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}^k} (x^i_t)^\intercal (q^k_t - p^{x,k,k}_t) x^i_t \\ &=: \mathbf{x}_t^\intercal P^x_t \mathbf{x}_t + \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}^k} \frac{1}{|\mathcal{N}^k|} (x^i_t)^\intercal Q^k_t x^i_t, \end{aligned} \tag{39}$$

where $P^{x\,k,l}_t = |\mathcal{N}^k||\mathcal{N}^l| p^{x,k,l}$ and $Q^k_t = |\mathcal{N}^k|(q^k_t - p^{x,k,k}_t)$. By similar algebraic manipulation, we can show

$$\mathbf{u}_t^\intercal R_t \mathbf{u}_t = \mathbf{u}_t^\intercal P^u_t \mathbf{u}_t + \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}^k} \frac{1}{|\mathcal{N}^k|} (u^i_t)^\intercal R^k_t u^i_t, \tag{40}$$

where $P^{u\,k,l}_t = |\mathcal{N}^k||\mathcal{N}^l| p^{u,k,l}$ and $R^k_t = |\mathcal{N}^k|(r^k_t - p^{u,k,k}_t)$. Substituting (39) and (40) in (3), we get (6).

B. Proof of Corollary 3

Under the assumptions on the model, the dynamics, given by (25) and (26), simplify to

$$\breve{x}^i_{t+1} = A^k_t \breve{x}^i_t + \breve{w}^i_t, \quad \bar{\mathbf{x}}_{t+1} = \bar{A}_t\bar{\mathbf{x}}_t + \tilde{B}_t\tilde{u}_t + \bar{\mathbf{w}}_t,$$

and $\bar{c}_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)$ of Corollary 4 simplifies to

$$\bar{c}_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) = \bar{\mathbf{x}}_t^\intercal(\bar{Q}_t + P^x_t)\bar{\mathbf{x}}_t + \tilde{u}_t^\intercal(\tilde{P}^u_t)\tilde{u}_t. \tag{41}$$

Thus, the N subsystems corresponding to $\breve{x}^i_t$ are uncontrolled and we need to identify $\tilde{u}_t$ to optimally control the dynamics of mean-field $\bar{\mathbf{x}}_t$ with per-step cost given by (41). Hence, the optimal solution is given by

$$\tilde{u}_t = \bar{L}_t\bar{\mathbf{x}}_t = \sum_{k\in\mathcal{K}} \tilde{L}^k_t \bar{x}^k_t,$$

where $\bar{L}_t$ is computed as explained in Corollary 3. To complete the proof, note that if agent $i\in\mathcal{N}^k$ of sub-population $k\in\mathcal{K}$ chooses action $u^i_t = \theta^{k+}_t \tilde{L}^k_t \bar{x}^k_t$, then we get $\theta^k_t \bar{u}^k_t = \tilde{L}^k_t \bar{x}^k_t$; consequently, $\tilde{u}_t = \sum_{k\in\mathcal{K}} \theta^k_t \bar{u}^k_t = \sum_{k\in\mathcal{K}} \tilde{L}^k_t \bar{x}^k_t$.

C. Proof of Theorem 3

As in the proof of Theorem 1 described in Section V, define $\breve{x}^i_t = x^i_t - \bar{x}^k_t$, $\breve{u}^i_t = u^i_t - \bar{u}^k_t$, $\mathring{\mathbf{x}}_t = \text{vec}((\breve{x}^i_t)_{i\in\mathcal{N}}, \bar{\mathbf{x}}_t)$, and $\mathring{\mathbf{u}}_t = \text{vec}((\breve{u}^i_t)_{i\in\mathcal{N}}, \bar{\mathbf{u}}_t)$. We identify a cost function $\{\breve{c}^k_t\}_{k\in\mathcal{K}}$ and $\bar{c}_t$ as in Corollary 4.

Corollary 5 *For time t, $t\in\{1,\dots,T\}$, there exist functions $\{\breve{c}^k_t\}_{k\in\mathcal{K}}$ and $\bar{c}_t$ such that*

$$\begin{aligned} c_t(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) = \bar{c}_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) &+ \sum_{i\in\mathcal{N}^k, k\in\mathcal{K}} \breve{c}^k_t(\breve{x}^i_t, \breve{u}^i_t) \\ &- \sum_{k\in\mathcal{K}} (\bar{r}^k_t)^\intercal Q^k_t \bar{r}^k_t, \end{aligned}$$

and for $t = T$,

$$c_T(\mathbf{x}_t, \bar{\mathbf{x}}_t) = \bar{c}_T(\bar{\mathbf{x}}_T) + \sum_{i\in\mathcal{N}^k, k\in\mathcal{K}} \breve{c}^k_T(\breve{x}^i_T) - \sum_{k\in\mathcal{K}} (\bar{r}^k_T)^\intercal Q^k_T \bar{r}^k_T.$$

To describe $\bar{c}_t(\cdot)$, define $\mathbf{y}_t := \left[\begin{smallmatrix}\bar{\mathbf{x}}_t - \bar{\mathbf{r}}_t \\ \bar{\mathbf{x}}_t - \mathbf{s}_t\end{smallmatrix}\right]$. Then,

$$\bar{c}_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) = \mathbf{y}_t^\intercal \begin{bmatrix}\bar{Q}_t & 0 \\ 0 & P^x_t\end{bmatrix}\mathbf{y}_t + \bar{\mathbf{u}}_t^\intercal(\bar{R}_t + P^u_t)\bar{\mathbf{u}}_t,$$

$$\bar{c}_T(\bar{\mathbf{x}}_T) = \mathbf{y}_T^\intercal \begin{bmatrix}\bar{Q}_T & 0 \\ 0 & P^x_T\end{bmatrix}\mathbf{y}_T.$$

Moreover,

$$\begin{aligned} \breve{c}^k_t(\breve{x}^i_t, \breve{u}^i_t) &= \frac{1}{|\mathcal{N}^k|}\left[(\breve{x}^i_t - r^i_t)^\intercal Q^k_t(\breve{x}^i_t - r^i_t) + (\breve{u}^i_t)^\intercal R^k_t \breve{u}^i_t\right], \\ \breve{c}^k_T(\breve{x}^i_T) &= \frac{1}{|\mathcal{N}^k|}\left[(\breve{x}^i_T - r^i_T)^\intercal Q^k_T(\breve{x}^i_T - r^i_T)\right]. \end{aligned}$$

Then, define a centralized auxiliary system where the state is $\mathring{\mathbf{x}}_t = \text{vec}((\breve{x}^i_t)_{i\in\mathcal{N}}, \bar{\mathbf{x}}_t)$, action is $\mathring{\mathbf{u}}_t = \text{vec}((\breve{u}^i_t)_{i\in\mathcal{N}}, \bar{\mathbf{u}}_t)$, and the per-step cost is given by Corollary 5. Note that the per-step cost is decomposed into terms that depend only on $(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)$ and terms that depend only on $(\breve{x}^i_t, \breve{u}^i_t)$ (and terms that do not depend on the control strategy). The rest of the proof follows along the same lines of the proof of Theorem 1. In particular, we consider a deterministic dynamical system and split it into $K+1$ classes. The agents in class $k, k \in \mathcal{K}$, are solving a tracking problem whose solution is given by

$$\breve{u}_t^i = \breve{L}_t^k \breve{x}_t^i + \breve{F}_t^k v_t^i.$$

The mean-field component is also solving a tracking problem whose solution is given by

$$\bar{\mathbf{u}}_t = \bar{L}_t \bar{\mathbf{x}}_t + \bar{F}_t \bar{v}_t.$$

The result of the Theorem follows from combining the above equations. Therefore, from standard results in LQR tracking problem, the optimal control law of agent $i \in \mathcal{N}^k$ of sub-population $k \in \mathcal{K}$ is given by

$$u_t^i = \breve{u}_t^i + \bar{u}_t^k = \left[\breve{L}_t^k (x_t^i - \bar{x}_t^k) + \breve{F}_t^k v_t^i\right] + \left[\bar{L}_t^k \bar{\mathbf{x}}_t + \bar{F}_t^k \bar{v}_t\right],$$

where gains $\{\breve{L}_t^k, \bar{L}_t^k, \breve{F}_t^k, \bar{F}_t^k\}_{t=1}^{T-1}$ are identical for all agents of sub-population k, $\bar{v}_t$ is identical for all agents of all sub-populations, and v_t^i may be different for each agent.

D. *Proof of Theorem 4*

The proof follows the same lines as the proof of Theorem 1 with the following differences. The mean-field is defined as $\bar{x}_t^{k,\lambda} = \frac{1}{|\mathcal{N}^k|}\sum_{i \in \mathcal{N}^k} \lambda^i x_t^i$ (similar interpretations hold for $\bar{u}_t^{k,\lambda}$ and $\bar{w}_t^{k,\lambda}$) and the breve variables are defined as $\breve{x}_t^i = x_t^i - \frac{\lambda^i}{b^i}\bar{x}_t^{k,\lambda}$ (similar interpretations hold for $\breve{u}_t^i$ and $\breve{w}_t^i$). Note that due to (A4), the dynamics of $\breve{x}_t^i$ and $\bar{\mathbf{x}}_t^\lambda$ are still given by (25) and (26), respectively, where $\bar{A}_t$ and $\bar{B}_t$ are defined as in Theorem 4.

The equivalent of Lemma 1 is the following:

Lemma 7 *Let $(\lambda^1, \dots, \lambda^N) \in \mathbb{R}^N$ and $(b^1, \dots, b^N) \in \mathbb{R}^N_{>0}$. In addition, for any $\mathbf{x} = \operatorname{vec}(x^1, \dots, x^N)$ and $\bar{x}^\lambda = \langle (\lambda^i x^i)_{i=1}^N \rangle$, let $\breve{x}^i = x^i - \frac{\lambda^i}{b^i}\bar{x}^\lambda$, $i \in \{1, \dots, N\}$. Then, for any matrix Q of appropriate dimension,*

$$\frac{1}{N}\sum_{i=1}^N b^i (x^i)^\intercal Q x^i = \frac{1}{N}\sum_{i=1}^N b^i (\breve{x}^i)^\intercal Q \breve{x}^i + (\bar{x}^\lambda)^\intercal \mu Q \bar{x}^\lambda,$$

where $\mu := 2 - \frac{1}{N}\sum_{i=1}^N \frac{(\lambda^i)^2}{b^i}$.

Consequently, the equivalent of Corollary 4 is the following

Corollary 6 *For time t, $t \in \{1, \dots, T\}$, there exist functions $\{\breve{c}^k\}_{k \in \mathcal{K}}$ and $\bar{c}_t$ such that*

$$c_t(\mathbf{x}_t, \mathbf{u}_t, \bar{\mathbf{x}}_t^\lambda, \bar{\mathbf{u}}_t^\lambda) = \bar{c}_t(\bar{\mathbf{x}}_t^\lambda, \bar{\mathbf{u}}_t^\lambda) + \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}^k} \breve{c}_t^i(\breve{x}_t^i, \breve{u}_t^i),$$

where

$$\bar{c}_t(\bar{\mathbf{x}}_t^\lambda, \bar{\mathbf{u}}_t^\lambda) = (\bar{\mathbf{x}}_t^\lambda)^\intercal (\bar{Q}_t + P_t^x)\bar{\mathbf{x}}_t^\lambda + (\bar{\mathbf{u}}_t^\lambda)^\intercal (\bar{R}_t + P_t^u)\bar{\mathbf{u}}_t^\lambda,$$
$$\breve{c}_t^i(\breve{x}_t^i, \breve{u}_t^i) = \frac{b^i}{|\mathcal{N}^k|}\left[(\breve{x}_t^i)^\intercal Q_t^k \breve{x}_t^i + (\breve{u}_t^i)^\intercal R_t^k \breve{u}_t^i\right],$$

and for $t = T$,

$$c_T(\mathbf{x}_t, \bar{\mathbf{x}}_t^\lambda) = \bar{c}_T(\bar{\mathbf{x}}_T^\lambda) + \sum_{i \in \mathcal{N}^k, k \in \mathcal{K}} \breve{c}_T^i(\breve{x}_T^i),$$

where

$$\bar{c}_T(\bar{\mathbf{x}}_T^\lambda) = (\bar{\mathbf{x}}_T^\lambda)^\intercal (\bar{Q}_T + P_T^x)\bar{\mathbf{x}}_T^\lambda,$$
$$\breve{c}_T^i(\breve{x}_T^i) = \frac{b^i}{|\mathcal{N}^k|}\left[(\breve{x}_T^i)^\intercal Q_T^k \breve{x}_T^i\right],$$

where $\bar{Q}_t$ and $\bar{R}_t$ are defined as in Theorem 4.

The rest of the proof is the same as in Section V-C. We can show that the optimal control strategy of auxiliary model is given by

$$\bar{\mathbf{u}}_t^\lambda = \bar{L}_t \bar{\mathbf{x}}_t^\lambda \quad \text{and for } k \in \mathcal{K}, i \in \mathcal{N}^k, \quad \breve{u}_t^i = \breve{L}_t^k \breve{x}_t^i,$$

where the gains $\{\breve{L}_t^k, \bar{L}_t\}_{t=1}^{T-1}$ are given as in Theorem 4. To complete the proof of Theorem 4, note that

$$u_t^i = \breve{u}_t^i + \frac{\lambda^i}{b^i}\bar{u}_t^{k,\lambda} = \breve{L}_t^k\left(x_t^i - \frac{\lambda^i}{b^i}\bar{x}_t^{k,\lambda}\right) + \frac{\lambda^i}{b^i}\bar{L}_t^k \bar{\mathbf{x}}_t^\lambda.$$

Thus, the control laws specified in Theorem 4 are the optimal *centralized* control laws, and, a fortiori, the optimal decentralized control laws.

PLACE PHOTO HERE

Jalal Arabneydi received Ph.D. degree in Electrical and Computer Engineering from McGill University, Montreal, Canada in 2016. He also received B.Sc. degree from Isfahan University of Technology, Isfahan, Iran, in 2006 and M.Sc. degree from University of Tehran, Tehran, Iran, in 2009 both in Electrical and Computer Engineering, majored in control systems. He is currently a postdoctoral fellow at Concordia University. He is the recipient of the best student paper award at the 53rd Conference on Decision and Control (CDC) in 2014. His research interests include decentralized stochastic control, mean field teams, game and team theory, large-scale system, multi-agent reinforcement learning with applications in complex networks including smart grids, swarm robotics, and finance.

PLACE PHOTO HERE

Aditya Mahajan (S'06–M'09–SM'14) received B.Tech degree in Electrical Engineering from the Indian Institute of Technology, Kanpur, India in 2003 and MS and PhD degrees in Electrical Engineering and Computer Science from the University of Michigan, Ann Arbor, USA in 2006 and 2008.

He is currently Associate Professor of Electrical and Computer Engineering at McGill University, Montreal. In 2017, he held a visiting position at the University of California, Berkeley. Prior to joining McGill University, he was postdoctoral researcher at Yale University, New Haven. He is senior member of the IEEE, member of SIAM, and member of Professional Engineers Ontario. He currently serves as an Associate Editor for Mathematics of Control, Signals, and Systems (MCSS). In the past, he has served as an Associate Editor on the IEEE Control Systems Society Conference Editorial Board.

His principal research interests include decentralized stochastic control, team theory, reinforcement learning, multi-armed bandits, and information theory.