

Team-Optimal Solution of Finite Number of Mean-Field Coupled LQG Subsystems

Jalal Arabneydi and Aditya Mahajan

Abstract—A decentralized control system with linear dynamics, quadratic cost, and Gaussian disturbances is considered. The system consists of a finite number of subsystems whose dynamics and per-step cost function are coupled through their mean-field (empirical average). The system has mean-field sharing information structure, i.e., each controller observes the state of its local subsystem (either perfectly or with noise) and the mean-field. It is shown that the optimal control law is unique, linear, and identical across all subsystems. Moreover, the optimal gains are computed by solving two decoupled Riccati equations in the full observation model and by solving an additional filter Riccati equation in the noisy observation model. These Riccati equations do not depend on the number of subsystems. It is also shown that the optimal decentralized performance is the same as the optimal centralized performance. An example, motivated by smart grids, is presented to illustrate the result.

I. INTRODUCTION

A. Motivation and literature overview

Team theory investigates multi-agent decision problems in which all agents share a common objective. Such problems arise in various applications including cyber-physical systems, networked control systems, surveillance and sensor networks, communication networks, smart grids, robotics, and organizational economics.

Not much is known regarding the optimal control of general team problems. Most of the results in the literature are for specific information structures such as delayed state sharing [1], delaying observation sharing [2], periodic sharing [3], belief sharing [4], mean-field sharing [5] and partial-history sharing [6]. We refer the reader to [7] for an overview.

There are two main challenges in solving team problems:

- 1) Team problems are *conceptually* difficult due to the decentralized nature of the information available to the controllers. In particular, controllers need to cooperate and coordinate to minimize a common cost but they have different information about the state of the environment. Due to this discrepancy of information, dynamic programming, which is one of the main solution techniques for the optimal design of stochastic systems, does not work directly for team problems.
- 2) Team problems are *computationally* difficult. Even when a dynamic programming decomposition is obtained, the solution complexity increases exponentially

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

Jalal Arabneydi and Aditya Mahajan are with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada. Email: {jalal.arabneydi@mail.mcgill.ca, aditya.mahajan@mcgill.ca}

or double exponentially with the number of controllers. So, it is difficult to use the dynamic programming solution for systems with large number of controllers.

Similar challenges exist in dynamic games; however, they have been successfully resolved for a class of models known as *mean-field games* [8]–[14]. The salient feature of these models is that the agents/players/controllers are coupled in their dynamics and per-step cost functions only through the mean-field (i.e., the empirical average or the empirical distribution). If the number of players are large, then an approximately optimal solution for these models can be obtained by solving the infinite population limit. The infinite population limit is easier to solve than the finite population model because, when the population is asymptotically large, the action of a single controller does not affect the mean-field. Therefore, the optimal solution can be obtained by solving two coupled equations: a backward dynamic programming equation that determines the best-response of a controller given the trajectory of the mean-field; and a forward Fokker-Planck equation that determines the evolution of the mean-field given the strategies of the individual controllers. A consistent solution of these equations determine Nash equilibrium strategies. A desirable feature of the solution is that the resultant control laws and the complexity of the solution do not depend on the number of controllers.

Motivated by these results, we investigate *team-optimal* (rather than Nash equilibrium) solution of mean-field models. In this paper, we focus on linear-quadratic models. In particular, the system consists of a *finite* number of subsystems; the dynamics of the subsystems are coupled through the mean-field of the states. The per-step cost is also coupled through the mean-field. We assume that the controllers have *mean-field sharing* information structure introduced in [5]; that is, each controller observes its local state (either completely or with noise) and the mean-field of the states of all subsystems. In some applications such as communication networks, the sharing of the mean-field happens naturally. In other applications, such as robotic teams, it is possible to share the mean-field using distributed consensus algorithms.

The main difference between our setup and mean-field games is the following: (i) We assume a finite number of controllers and obtain team-optimal control strategies. In mean-field games, one assumes asymptotically large number of controllers and obtains Nash equilibrium strategies. (ii) We assume that each controller observes the mean-field. In mean-field games, this assumption is not made. However, when the number of controllers goes to infinity, both the information

structures are equivalent because the mean-field becomes a deterministic process.

B. Contributions and salient features of the results

The linear quadratic mean-field model is a decentralized system with non-classical information structure. **Our main result is to show that linear control strategies are optimal.** The model is neither partially nested nor quadratic invariant. In contrast, almost all positive results for linear quadratic systems with non-classical information structures are for models that are either partially nested or quadratic invariant.

We also show that the optimal control laws at each subsystem are identical. In general, identical control laws are *not* optimal for systems with exchangeable subsystems (see [5] for one example). So, a natural question is the following: Given a decentralized system with exchangeable subsystems, under what conditions are optimal control laws identical across subsystems. This question warrants further investigation.

Our solution, and the solution complexity, do not depend on the number of subsystems. Irrespective of the number of subsystems, the optimal control law is given by the solution of *two* (backward) Riccati equations in the full observation model, and an additional filter Riccati equation in the noisy observation model. The parameters of these Riccati equations do not depend on the number of subsystems. Thus, **the optimal control laws can be computed without any knowledge of the number of subsystems.**

Since the optimal control laws do not depend on the number of subsystems, our results remain valid in the infinite population limit. In the infinite population limit, the mean-field becomes a deterministic process that can be pre-computed at all controllers. Thus, mean-field sharing is equivalent to the completely decentralized information structure (where each controller only observes its local state) considered in [8], [9]. Thus, **our results may be viewed as an alternative derivation of the infinite population mean-field games solution.**

As an intermediate step of the proof, **we show that the decentralized control laws—and, hence, the decentralized performance—are same as centralized** (i.e., when all control actions are chosen by a single controller that observes all the information available to all the decentralized controllers). A natural follow up question that warrants further investigation is the following: In general decentralized control problems, when is decentralized performance the same as the centralized performance? Or more generally, when is the performance under one information structure the same as the performance under a finer information structure?

From an implementation point of view, the above feature has an interesting consequence. If we have the freedom to design the information structure, then in the linear-quadratic mean-field systems **there is no advantage of sharing anything beyond the mean-field.** Since, it is easy to share the mean-field using distributed consensus algorithms, it is relatively easy to implement the optimal centralized solution in a distributed manner.

Furthermore, as we already argued, in the infinite population limit, mean-field sharing is equivalent to the completely decentralized setup (i.e., each controller only observes its own local state). Since, there is no advantage of sharing any information beyond the mean-field—which is already computable at each controller—the optimal completely decentralized solution is also the optimal centralized solution. An immediate corollary is that **the control laws obtained using the approach of mean-field games are also the optimal centralized solution.**

II. PROBLEM FORMULATION AND THE MAIN RESULTS

A. Notation

Uppercase letters denote random variables/vectors and lowercase letters denote their realization. $\mathbb{P}(\cdot)$ denotes the probability of an event and $\mathbb{E}[\cdot]$ denotes the expectation of a random variable. \mathbb{R} denotes the set of real numbers.

For a sequence of (column) vectors X, Y, Z, \dots , the notation $\text{vec}(X, Y, Z, \dots)$ denotes the vector $[X^\top, Y^\top, Z^\top]^\top$. The vector $\text{vec}(X_1, \dots, X_t)$ is also denoted by $X_{1:t}$.

Superscripts index subsystems and subscripts index time. Thus, X_t^i denotes a variable at subsystem i at time t . Bold letters denote collection of variables at all subsystems. For example, \mathbf{X} denotes (X^1, \dots, X^n) . $\langle \mathbf{X} \rangle$ denotes the mean-field of \mathbf{X} , i.e., $\langle \mathbf{X} \rangle = \frac{1}{n} \sum_{i=1}^n X^i$.

B. Model and problem formulation

Consider a decentralized control system with n homogeneous subsystems that operate for a finite horizon T . A controller is co-located with each subsystem. Let $X_t^i \in \mathbb{R}^{d_x}$ denote the state of subsystem i and $U_t^i \in \mathbb{R}^{d_u}$ denote the control action taken by controller i , $i \in \{1, \dots, n\}$, at time t . Let $Z_t = \langle \mathbf{X}_t \rangle$ denote the mean-field of the states at time t .

1) *System dynamics:* The initial states $\{X_1^i\}_{i=1}^n$ are independent Gaussian random variables with zero mean and covariance Σ_X . They evolve in time as follows:

$$X_{t+1}^i = A_t X_t^i + B_t U_t^i + D_t Z_t + W_t^i, \quad i \in \{1, \dots, n\}, \quad (1)$$

where $A_t, B_t,$ and D_t are matrices of appropriate dimensions and $\{W_t^i\}_{t=1}^T$ is an i.i.d. noise process where W_t^i is Gaussian with zero mean and covariance Σ_W . We assume that the primitive random variables $\{X_1^i, W_1^i, \dots, W_T^i\}_{i=1}^n$ are independent.

Thus, the dynamics of the subsystems is coupled to the others through the mean-field.

2) *Per-step cost:* At time t , the system incurs a cost that depends on the local state of the subsystems and the mean-field as follows:

$$c_t(\mathbf{X}_t, \mathbf{U}_t, Z_t) = \frac{1}{n} \sum_{i=1}^n \left[X_t^{i\top} Q_t X_t^i + U_t^{i\top} R_t U_t^i \right] + Z_t^\top P_t Z_t \quad (2)$$

where $Q_t, P_t, S_t,$ and R_t are matrices of appropriate dimension; Q_t and P_t are symmetric and non-negative definite and R_t is symmetric and positive definite.

Remark 1 The results derived in this paper also apply to a more general per-step cost of the form

$$\frac{1}{n} \sum_{i=1}^n \left[X_t^{i\top} Q_t X_t^i + X_t^{i\top} S_t Z_t + U_t^{i\top} R_t U_t^i \right] + Z_t^\top P_t Z_t$$

because it can be rewritten in the form (2).

3) *Observation model and information structure:* We consider two observation models that differ in the observation of the local state X_t^i at controller i . In the first model, called *full observation model*, controller i perfectly observes the local state X_t^i ; in the second model, called *noisy observation model*, controller i observes a noisy version $Y_t^i \in \mathbb{R}^{d_y}$ of the local state X_t^i that is given by

$$Y_t^i = C_t^x X_t^i + C_t^z Z_t + V_t^i, \quad i \in \{1, \dots, n\}, \quad (3)$$

where the observation noise $\{V_t^i\}_{t=1}^T$ is i.i.d. Gaussian process with zero mean and covariance Σ_V . In addition, we assume that the noise processes are independent across subsystems and also independent of $\{X_1^i, W_1^i, \dots, W_T^i\}_{i=1}^n$.

In both models, in addition to the local measurement of the state of its subsystem, each controller perfectly observes the mean-field Z_t . Following [5], we call this model *mean-field sharing* information structure. Controllers perfectly recall all the data they observe. Thus, in the full observation model, controller i chooses control action according to

$$U_t^i = g_t^i(X_{1:t}^i, U_{1:t-1}^i, Z_{1:t}), \quad (4)$$

while in the noisy observation model, controller i chooses control action according to

$$U_t^i = g_t^i(Y_{1:t}^i, U_{1:t-1}^i, Z_{1:t}). \quad (5)$$

The function g_t^i is called the *control law* of controller i . The collection $\mathbf{g}^i = (g_1^i, g_2^i, \dots, g_T^i)$ is called the *control strategy* of controller i . The collection $\mathbf{g} = (\mathbf{g}^1, \dots, \mathbf{g}^n)$ of control strategies of all controllers is called the *control strategy of the system*.

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

Problem 1 *In the model described above, find a joint strategy $\mathbf{g} = (\mathbf{g}^1, \dots, \mathbf{g}^n)$ that minimizes the following cost:*

$$J(\mathbf{g}) = \mathbb{E}^{\mathbf{g}} \left[\sum_{t=1}^T c_t(\mathbf{X}_t, \mathbf{U}_t, Z_t) \right], \quad (6)$$

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

C. The main results

1) *Full observation model:* For the full observation model, our main result is the following:

Theorem 1 *The optimal control strategy of Problem 1 is unique and given by*

$$U_t^i = K_t^x X_t^i + (K_t^z - K_t^x) Z_t, \quad i \in \{1, \dots, n\}. \quad (7)$$

The optimal gains are given as follows. For the terminal time-step $K_T^x = K_T^z = 0_{d_u \times d_x}$ and for $t \leq T-1$,

$$K_t^x = - (B_t^\top M_{t+1}^x B_t + R_t)^{-1} B_t^\top M_{t+1}^x A_t, \quad (8)$$

and

$$K_t^z = - (B_t^\top M_{t+1}^z B_t + R_t)^{-1} B_t^\top M_{t+1}^z \bar{A}_t \quad (9)$$

where $\bar{A}_t := A_t + D_t$ and $\{M_t^x\}_{t=1}^T$ and $\{M_t^z\}_{t=1}^T$ are solutions of the following two decoupled Riccati equations:

$$M_T^x = Q_T, \quad M_T^z = Q_T + P_T,$$

and for $t = T-1, \dots, 1$,

$$M_t^x = -A_t^\top M_{t+1}^x B_t (B_t^\top M_{t+1}^x B_t + R_t)^{-1} B_t^\top M_{t+1}^x A_t + A_t^\top M_{t+1}^x A_t + Q_t, \quad (10)$$

and

$$M_t^z = -\bar{A}_t^\top M_{t+1}^z B_t (B_t^\top M_{t+1}^z B_t + R_t)^{-1} B_t^\top M_{t+1}^z \bar{A}_t + \bar{A}_t^\top M_{t+1}^z \bar{A}_t + Q_t + P_t. \quad (11)$$

The above result may also be derived under the weaker assumption that the initial state $\mathbf{X}_1 = (X_1^1, \dots, X_1^n)$ as well as the noise $\mathbf{W}_t = (W_t^1, \dots, W_t^n)$ are correlated across the subsystems.

2) *Noisy observation model:* For the noisy observation model, our main result is the following:

Theorem 2 *For the noisy observation model, the optimal control strategy for Problem 1 is unique and is given by*

$$U_t^i = K_t^x \hat{X}_t^i + (K_t^z - K_t^x) Z_t, \quad i \in \{1, \dots, n\}, \quad (12)$$

where optimal gains K_t^x and K_t^z are the same as in Theorem 1 and $\hat{X}_t^i = \mathbb{E}[X_t^i \mid Y_{1:t}^i, Z_{1:t}, U_{1:t-1}^i]$ which is generated by the following Kalman filtering equations: $\hat{X}_1^i = 0$ and for $t > 1$,

$$\hat{X}_{t+1}^i = A_t \hat{X}_t^i + B_t U_t^i + K_t^F (Y_t^i - C_t^x \hat{X}_t^i - C_t^z Z_t) \quad (13)$$

where the Kalman filter gains are given by

$$K_t^F = A_t S_t C_t^{x\top} (C_t^x S_t C_t^{x\top} + \Sigma_V)^{-1}, \quad (14)$$

and the state estimation error covariances satisfy the (filter) Riccati equation: $S_1 = \Sigma_X$ and for $t > 1$,

$$S_{t+1} = -A_t S_t C_t^{x\top} (C_t^x S_t C_t^{x\top} + \Sigma_V)^{-1} C_t^x S_t A_t^\top + A_t S_t A_t^\top + \Sigma_W. \quad (15)$$

III. PROOF OF THE MAIN RESULTS

The main idea of the proof is as follows. We construct an auxiliary system whose state, control actions, and per-step cost are equivalent to \mathbf{X}_t , \mathbf{U}_t , and $c_t(\cdot)$, respectively (modulo a change of variables to be described later). However, this auxiliary system is centrally controlled by a single control that has access to all the information available to the n decentralized controllers in the original system. We show that the optimal centralized solution of this auxiliary system can be implemented in the original decentralized system, and is therefore also optimal for the decentralized system.

A. The auxiliary system

Define $\bar{X}_t^i = X_t^i - Z_t$ and $\bar{U}_t^i = U_t^i - U_t^z$, where $U_t^z = \langle \mathbf{U}_t \rangle$. The auxiliary system is a centralized system with state $\tilde{X}_t = \text{vec}(\bar{X}_t^1, \dots, \bar{X}_t^n, Z_t)$ and actions $\tilde{U}_t = \text{vec}(\bar{U}_t^1, \dots, \bar{U}_t^n, U_t^z)$. Note that \bar{X}_t^i is equivalent to \mathbf{X}_t and \bar{U}_t^i is equivalent to \mathbf{U}_t .

The dynamics are the same as the model in Sec. II-B. In particular,

$$\bar{X}_{t+1}^i = A_t \bar{X}_t^i + B_t \bar{U}_t^i + W_t^i - \langle \mathbf{W}_t \rangle \quad (16)$$

and

$$Z_{t+1} = (A_t + D_t)Z_t + B_t U_t^z + \langle \mathbf{W}_t \rangle \quad (17)$$

The per-step cost of the auxiliary model is given by $c_t(\mathbf{X}_t, \mathbf{U}_t, Z_t)$ defined in (2).

As in Sec. II-B, we consider two observation models for the auxiliary system: full observation and noisy observation. In both cases, there is a *single centralized controller* that chooses \tilde{U}_t based on the observations.

In the full observation model, the centralized controller observes \tilde{X}_t and chooses \tilde{U}_t according to

$$\tilde{U}_t = \bar{g}_t(\tilde{X}_{1:t}, \tilde{U}_{1:t-1}).$$

In the noisy observation model, the centralized controller observes $\mathbf{Y}_t = \text{vec}(Y_t^1, \dots, Y_t^N)$ where Y_t^i is given by (3) and chooses \tilde{U}_t according to

$$\tilde{U}_t = \bar{g}_t(\mathbf{Y}_{1:t}, \tilde{U}_{1:t-1}, Z_{1:t}).$$

In both models, we are interested in the following optimization problem:

Problem 2 *In the model described above, find a strategy $\bar{\mathbf{g}} = (\bar{g}_1, \dots, \bar{g}_T)$ that minimizes the following cost:*

$$\bar{J}(\bar{\mathbf{g}}) = \mathbb{E}^{\bar{\mathbf{g}}} \left[\sum_{t=1}^T c_t(\mathbf{X}_t, \mathbf{U}_t, Z_t) \right], \quad (18)$$

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

Let J^* and \bar{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 $J^* \geq \bar{J}^*$. 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.

A critical step in the proof is to rewrite the cost $c_t(\mathbf{X}_t, \mathbf{U}_t, Z_t)$ in terms of \tilde{X}_t and \tilde{U}_t . For that matter, we need the following key result.

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

$$\frac{1}{n} \sum_{i=1}^n x^{i\top} Q x^i = \frac{1}{n} \sum_{i=1}^n \bar{x}^{i\top} Q \bar{x}^i + z^\top Q z \quad (19)$$

PROOF The result follows from elementary algebra and the observation that $\sum_{i=1}^n \bar{x}^i = 0$.

An immediate consequence of Lemma 1 is the following:

Corollary 1 *For any time t , $c_t(\mathbf{X}_t, \mathbf{U}_t, Z_t) = \bar{c}_t(\tilde{X}_t, \tilde{U}_t)$, where*

$$\bar{c}_t(\tilde{X}_t, \tilde{U}_t) = \frac{1}{n} \sum_{i=1}^n \left[\bar{X}_t^{i\top} Q_t \bar{X}_t^i + \bar{U}_t^{i\top} R_t \bar{U}_t^i \right] + Z_t^\top (Q_t + P_t) Z_t + U_t^{z\top} R_t U_t^z. \quad (20)$$

Note that the auxiliary model has linear dynamics and in Corollary 1 we have shown that the cost is quadratic in the state and the control actions. Thus, the optimal control actions are linear in the state and the corresponding optimal gains can be obtained by solving an appropriate Riccati equation. However, the state \tilde{X}_t of the auxiliary system belongs to $\mathbb{R}^{(n+1)d_x}$, thus, a naive attempt to obtain an optimal solution will involve solving for $(n+1)d_x \times (n+1)d_x$ dimensional Riccati equations. We present an alternative approach in the next section that involves solving two $d_x \times d_x$ dimensional Riccati equations (independent of n).

B. Full observation model

The auxiliary system is a stochastic linear quadratic system. From the certainty equivalence principle [15], we know that the optimal control law is unique and identical to the control law in the corresponding deterministic problem, whose dynamics are given by

$$\bar{X}_{t+1}^i = A_t \bar{X}_t^i + B_t \bar{U}_t^i, \quad i \in \{1, \dots, n\}, \quad (21)$$

$$Z_{t+1} = (A_t + D_t)Z_t + B_t U_t^z. \quad (22)$$

and the per-step cost is $\bar{c}(\tilde{X}_t, \tilde{U}_t)$ given by (20).

Note that this system consists on $(n+1)$ components: n components with state \bar{X}_t^i and control \bar{U}_t^i , $i \in \{1, \dots, n\}$, and one component with state Z_t and control U_t^z . These components have decoupled dynamics and decoupled cost; and n of these are identical. Thus, the optimal control law of each component may be identified separately. Therefore, we have the following:

Theorem 3 *The optimal control strategy of auxiliary model is unique and given by*

$$\bar{U}_t^i = K_t^x \bar{X}_t^i, \quad U_t^z = K_t^z Z_t, \quad i \in \{1, \dots, N\}, \quad (23)$$

where the gains K_t^x and K_t^z are given as in Theorem 1.

To complete the proof of Theorem 1, note that

$$U_t^i = \bar{U}_t^i + U_t^z = K^x (X_t^i - Z_t) + K^z Z_t.$$

Thus, the control laws specified in Theorem 1 are the optimal *centralized* control laws, and, a fortiori, the optimal *decentralized* control laws.

C. Noisy observation model

Define

$$\hat{X}_t^i = \mathbb{E}[\bar{X}_t^i \mid \mathbf{Y}_{1:t}, \tilde{U}_{1:t-1}, Z_{1:t}],$$

$$\check{X}_t^i = \mathbb{E}[X_t^i \mid \mathbf{Y}_{1:t}, \tilde{U}_{1:t-1}, Z_{1:t}].$$

Since $X_t^i = \bar{X}_t^i + Z_t$, we have that $\check{X}_t^i = \hat{X}_t^i + Z_t$.

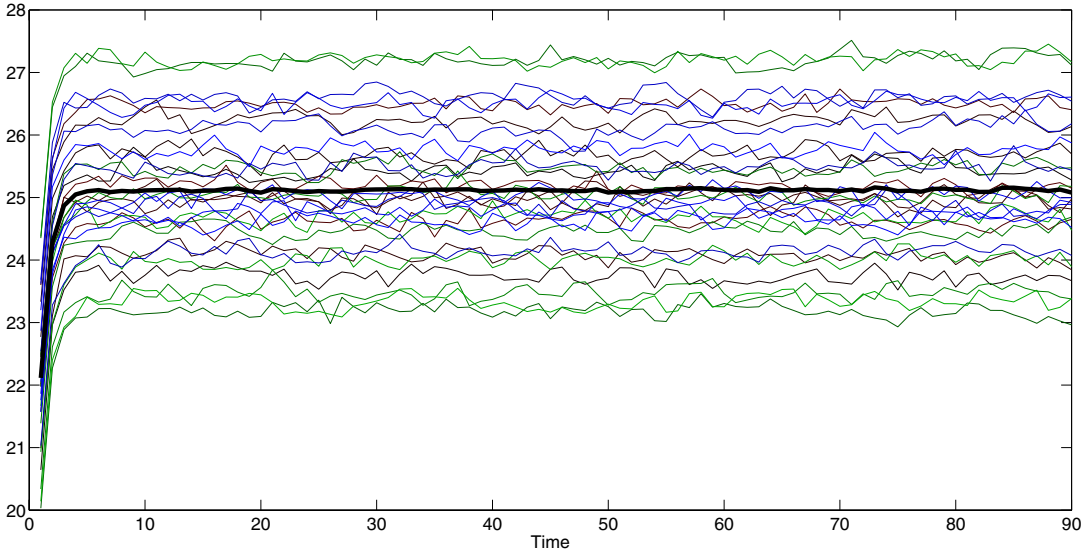


Fig. 1. An example of mean-field control of space heaters. The light lines show the temperature of individual heaters and the thick line depicts the mean-field of the trajectories. Note that the individual trajectories adjust around their initial values so that the mean-field tracks the reference trajectory $z_{ref} = 25$. The tracking is not perfect because of the friction between tracking the initial states locally and tracking the mean-field globally.

From the Separation Theorem for linear quadratic Gaussian systems [15], we know that the optimal (centralized) control law for the auxiliary system is given by

$$\bar{U}_t^i = K_t^x \bar{X}_t^i, \quad U_t^z = K_t^z Z_t, \quad i \in \{1, \dots, N\}.$$

Or equivalently,

$$U_t^i = K_t^x \check{X}_t^i + (K_t^z - K_t^x) Z_t, \quad i \in \{1, \dots, n\}.$$

Thus, to prove Theorem 2, we need to show that $\check{X}_t^i = \hat{X}_t^i$ (where \hat{X}_t^i is defined in Theorem 2). This is proved below.

Lemma 2 *In the auxiliary system with noisy observation, $\check{X}_t^i = \hat{X}_t^i$.*

PROOF Recall that X_t^i evolves according to (1) and Y_t^i is given by (3). Therefore, for any $i, j \in \{1, \dots, n\}$ such that $i \neq j$, $(X_{1:t}^i, Y_{1:t}^i)$ is conditionally independent of $(X_{1:t}^j, Y_{1:t}^j)$. Consequently,

$$\begin{aligned} \check{X}_t^i &= \mathbb{E}[X_t^i \mid \mathbf{Y}_{1:t}, \tilde{U}_{1:t-1}, Z_{1:t}] \\ &= \mathbb{E}[X_t^i \mid Y_{1:t}^i, \tilde{U}_{1:t-1}, Z_{1:t}] \\ &= \mathbb{E}[X_t^i \mid Y_{1:t}^i, \mathbf{U}_{1:t-1}, Z_{1:t}] \\ &\stackrel{(a)}{=} \mathbb{E}[X_t^i \mid Y_{1:t}^i, U_{1:t-1}^i, Z_{1:t}] \\ &= \hat{X}_t^i, \end{aligned}$$

where (a) follows because X_t^i depends on $\mathbf{U}_{1:t-1}$ only through $U_{1:t-1}^i$.

IV. NUMERICAL EXAMPLE

We simulate an example of inspired by the *collective target tracking mean-field model* of [16]. We consider a population of space heaters. The state X_t^i denotes the room temperature at heater i at time t . We assume that the initial mean temperature is z_1 degrees, with individual temperatures distributed according to Gaussian distribution with unit variance around

z_1 degrees. We assume that the temperature dynamics are linearized around the operating point and are given by

$$X_{t+1}^i = a(X_t^i - x_0) + bU_t^i + W_t^i$$

where x_0 is the ambient temperature, W_t^i is the randomness due to the environment and U_t^i is the control action of a local controller.

It is desired that the mean temperature increases to 21 degrees (denoted by z_{ref}) in T time steps (we assume $T = 90$ minutes). The per-step cost function is

$$\frac{1}{n} \sum_{i=1}^n \left[q(X_t^i - X_1^i)^2 + rU_t^i{}^2 \right] + p(Z_t - z_{ref})^2$$

The rationale of this cost function is as follows. It is assumed that the initial temperature is the comfort level of user i ; so we penalize local deviations from the initial temperature. The second term corresponds to the local control energy. The last term is the penalty for the mean-field deviating from the reference mean-field. The objective is to minimize the expected total cost over a finite horizon.

The above model is a tracking problem. The results of Theorem 1 extend to tracking problems in a natural manner. We solve the Riccati equations for the following values of the parameters:

$$\begin{aligned} n &= 30, & a &= 0.8, & b &= 1, \\ q &= 0.5, & p &= 1, & r &= 1, \\ x_0 &= 22, & z_1 &= 22, & z_{ref} &= 25, \\ T &= 90, & W_t^i &\sim \mathcal{N}(0, 1), & X_1^i &\sim \mathcal{N}(22, 2). \end{aligned}$$

The resultant trajectories are shown in Fig. 1.

V. CONCLUSION

We consider team-optimal control of finite number of mean-field coupled LQG subsystems with mean-field sharing

information structure. The optimal control law is unique, identical for all controllers, and linear in the current state and the current mean-field; the optimal gains are obtained by solving two decoupled Riccati equations for the full observation case and an additional filter Riccati equation for the noisy observation case. These Riccati equations do not depend on the number of subsystems.

To prove the main results, we consider an auxiliary centralized system whose dynamics and cost are equivalent to the original decentralized system. It is shown that the optimal centralized control law are implementable with mean-field sharing. Thus, the optimal decentralized control laws are identical to the optimal centralized control laws.

Although we only presented the results for the finite horizon optimal problem, the results extend to infinite horizon and to optimal tracking problems in a natural manner. Moreover, it is possible to extend the results to non-homogeneous population consisting of multiple types.

REFERENCES

- [1] M. Aicardi, F. Davoli, and R. Minciardi, "Decentralized optimal control of Markov chains with a common past information set," *IEEE Trans. Autom. Control*, vol. 32, no. 11, pp. 1028–1031, 1987.
- [2] A. Nayyar, A. Mahajan, and D. Teneketzis, "Optimal control strategies in delayed sharing information structures," *IEEE Trans. Autom. Control*, vol. 56, no. 7, pp. 1606–1620, Jul. 2011.
- [3] J. M. Ooi, S. M. Verbout, J. T. Ludwig, and G. W. Wornell, "A separation theorem for periodic sharing information patterns in decentralized control," *IEEE Trans. Autom. Control*, vol. 42, no. 11, pp. 1546–1550, Nov. 1997.
- [4] S. Yüksel, "Stochastic nestedness and the belief sharing information pattern," *IEEE Trans. Autom. Control*, pp. 2773–2786, Dec. 2009.
- [5] J. Arabneydi and A. Mahajan, "Team optimal control of coupled subsystems with mean field sharing," in *Proc. 53rd IEEE Conf. Decision and Control*, Los Angeles, CA, Dec. 2014, pp. 1669–1674.
- [6] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *IEEE Trans. Autom. Control*, vol. 58, no. 7, pp. 1644–1658, Jul. 2013.
- [7] A. Mahajan, N. Martins, M. Rotkowitz, and S. Yüksel, "Information structures in optimal decentralized control," in *Proc. 51st IEEE Conf. Decision and Control*, Maui, Hawaii, Dec. 2012, pp. 1291 – 1306.
- [8] M. Huang, P. Caines, and R. Malhame, "Large population cost-coupled LQG problems with non-uniform agents: individual mass behavior and decentralized epsilon-nash equilibrium," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1560–1571, Sep. 2007.
- [9] —, "An invariance principle in large population stochastic dynamic games," *Journal of Systems Science and Complexity*, vol. 20, pp. 162–172, 2007.
- [10] P. Lions and J. Largsy, "Mean-field games," *Japanese Journal of Mathematics*, vol. 2, no. 1, pp. 229–260, Mar. 2007.
- [11] 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.
- [12] O. Guéant, J.-M. Lasry, and P.-L. Lions, *Paris-Princeton Lecture on Mathematical Finance*. Springer, 2011, ch. Mean-field games and applications, pp. 205–266.
- [13] 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.
- [14] 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.
- [15] P. E. Caines, *Linear stochastic systems*. John Wiley & Sons, Inc., 1987.
- [16] A. Kizilkale and R. Malhame, "Mean field based control of power system dispersed energy storage devices for peak load relief," in *IEEE Conference on Decision and Control (CDC)*, Dec. 2013, pp. 4971–4976.