Team Optimal Decentralized Kalman Filtering with Coupled Cost

Mohammad Afshari and Aditya Mahajan

***Abstract*— We consider the problem of optimal decentralized estimation of a Gauss-Markov process with multiple agents. Each agent obtains noisy observations of the state of the process and shares its observations with its neighbors over a graph with communication delays. At each time, each agent generates an estimate of the state of the process based on all the information available to it. The objective is to minimize the estimation cost over a finite horizon where the per-step estimation cost depends on how close the estimate of each node is to the true state of the system and to the estimate of its neighbors (according to an estimation weight graph that may be different from the communication graph). We show the optimal estimates are linear in the local estimates (i.e., estimates which only minimize the instantaneous error in estimating the true state of the process but ignore the coupling with the neighbors). The corresponding gains depend on the cost coupling between the agents as well as on the covariance between the agents' local estimates. We show how to recursively compute the local estimates and the covariance between the local estimates for three representative graphs: a graph with no edges, complete graphs with d unit delay along each communication link, and strongly connected graphs with one unit delay along each communication link.**

I. Introduction

A. Motivation and literature overview

In many large scale systems, measurements of the state of the system are obtained from geographically distributed sensors. Examples include networked control systems, sensor networks, power systems, and transportation networks. In an ideal situation when communication cost and communication constraints are absent, the measurements from all of the sensors can be collected at a central location where an estimate of the state can be computed. However, in a real system, agents have limited communication capabilities, and communication constraints must be taken into account when generating state estimates.

Due to its importance, several variations of decentralized estimation has been investigated in the literature. Broadly speaking, the literatures on optimal decentralized estimation can be classified as follows. The first category consists of models where agents communicate a function of their observations (typically their local estimates and some correction terms) to a fusion center both for Gauss Markov processes [1]–[7] and general setup [8]. These papers identify conditions under which the fusion center can compute the centralized estimate but they do not take communication constraints such as link delays into account. A variation of these consists of models where, in addition to the fusion center, the agents can communicate directly with one another [9], [10]. In such models, conditions are identified under which the fusion center can compute the centralized estimates.

The second category is those models in which there is no fusion center, but agents compute individual estimates. These estimates are coupled according to a per-step cost function. Such models were considered in [11], [12] for a Gauss-Markov process. Recursive expressions for the optimal estimates are derived using Hilbert space projections.

The third category is those models in which the agents communicate a function of their observation to other agents over a pre-specified communication graph. Estimating of a static random variable is considered in [13], and it is shown that for tree networks, exchanging local estimates achieves the same estimation performance as exchanging all observations. In [14], distributed estimation under partially nested information structures is studied. The objective is to minimize a coupled cost function, subject to sparsity constraints which are reflected by the dynamics of the system. Distributed state estimation of linear Kalman filtering is considered in [15] where nodes are required to collaborate to estimate the state of a dynamic system while communicating with their direct neighbors. A diffusion Kalman filter is presented and the mean and mean square performance of the proposed algorithm is analyzed. Consensus based distributed Kalman filtering is considered in [16]–[18] and various asymptotically stable filtering algorithms are proposed.

In this paper, we consider a general model in the second category presented above. We are interested in team optimal decentralized estimation. We consider a setup where agents share their observation over a pre-specified communication graph and the estimation cost is coupled. We use ideas from team theory to identify optimal estimation strategy. Due to lack of space, we do not provide an overview of the literature on team theory but refer the reader to [19], [20] for an overview.

B. Notation

Given a matrix A, A_{ij} denotes its (i,j)-th element, $A^{\intercal}$ denotes its transpose, $\operatorname{vec}(A)$ denotes the column vector of A formed by vertically stacking the columns of A. Given matrices A and B, $\operatorname{diag}(A,B)$ denotes the matrix obtained by putting A and B in diagonal blocks. Given matrices A and B, with the same number of columns, $\operatorname{rows}(A,B)$ denotes the matrix obtained by stacking A on top of B. I_n denotes an $n \times n$ identity matrix. We simply use I when the dimension is clear for context. Given any vector valued process $\{y(t)\}_{t\geq 0}$, any any time instances t_1, t_2 such that $t_1 \leq t_2$, $y(t_1{:}t_2)$ is a

The authors are with the Department of Electrical and Computer Engineering, McGill University, Montreal, Canada. Emails: mohammad.afshari2@mail.mcgill.ca, aditya.mahajan@mcgill.ca.

short hand notation for $\text{vec}(y(t_1), y(t_1+1), \dots, y(t_2))$. Given random variables x and y, $\mathbb{E}[x]$ and $\text{var}(x)$ denote the mean and variance of x while $\text{cov}(x, y)$ denotes the covarinace between x and y.

C. Preliminaries on graph theory

A weighted graph $\mathcal{G}$ is an odered set (N, E, ρ) where N is the set of nodes and $E \subset N \times N$ is the set of edges, and $\rho\colon E \to \mathbb{R}^k$ is a weight function. In a directed graph, an edge (i, j) in E is considered directed from i to j; i is the *in-neighbor* of j; j is the *out-neighbor* of i; and i and j are neighbors. The set of in-neighbors of i, called the *in-neighborhood* of i is denoted by N_i^-; the set of out-neighbors of i, called the *out-neighborhood* is denoted by N_i^+.

In a directed graph, a *path* is a sequnce of nodes such that each node has a directed edge to the next node in the sequence. The *length* of a path is the number of edges in the path. The *geodesic distance* between two nodes i and j, denoted by $\ell_{i,j}$, is the length of the shortest path connecting the two nodes. The *diameter* of the graph is the largest geodesic distance between any two nodes. A graph is called *strongly connected* if it is possible to reach any node from any other node by traversing the edges in the direction in which they point.

II. Problem Statement

Consider a discrete-time Gauss-Markov process $\{x(t)\}_{t\geq 0}$, $x(t) \in \mathbb{R}^{d_x}$, where $x(0) \sim \mathcal{N}(0, \Sigma_x)$ and for $t \geq 0$,

$$x(t+1) = Ax(t) + w_0(t), \tag{1}$$

where A is a $d_x \times d_x$ matrix and $w_0(t) \in \mathbb{R}^{d_x}$ is the process noise with $w_0(t) \sim \mathcal{N}(0, Q_0)$. There are n agents that observe the process with noise. Let $N = \{1, \dots, n\}$ denote the set of agents. At time t, the observation $y_i(t) \in \mathbb{R}^{d_y^i}$ of agent i, $i \in N$, is given by:

$$y_i(t) = C_i x(t) + w_i(t), \tag{2}$$

where C_i is a $d_y^i \times d_x$ matrix and $w_i(t) \in \mathbb{R}^{d_y^i}$ is the observation noise with $w_i(t) \sim \mathcal{N}(0, Q_i)$.

We make the following assumptions on the model:

Assumption 1 *The primitive random variables* $(x(0), \{w_0(t)\}_{t\geq 0}, \{w_1(t)\}_{t\geq 0}, \dots, \{w_n(t)\}_{t\geq 0})$ *are independent.*

Assumption 2 *The noise covariance matrices* $Q_0, \{Q_i\}_{i\in N}$ *are positive definite.*

The agents are connected over a *communication graph* $\mathcal{G}^c$, which is a weighted directed graph with vertex set N. For any node i, let N_i^{c-} and N_i^{c+} denote the set of in- and out-neighbors of i. For every edge (i, j), the associated weight d_{ij} is a positive integer that denotes the communication delay from node i to node j.

Let $I_i(t)$ denote the information available to agent i at time t. We assume that agent i knows the history of all its observations and d_{ji} step delayed information of its in-neighbor j, $j \in N_i$, i.e.,

$$I_i(t) = \{y_i(0{:}t), \{I_j(t - d_{ji})\}_{j\in N_i^{c-}}\}. \tag{3}$$

In (3), we implicitly assume that $I_i(t) = \emptyset$ for any $t < 0$. Let $z_i(t) = I_i(t) \setminus I_i(t-1)$ denote the new information that becomes available to agent i at time t. Then, $z_i(0) = y_i(0)$ and for $t > 0$,

$$z_i(t) = \{y_i(t), \{z_j(t - d_{ji})\}_{j\in N_i^{c-}}\}.$$

It is assumed that at each time t, agent $j, j \in N$, communicates $z_j(t)$ to all its out-neighbors. This information reaches the out-neighbor i of agent j, at time $t + d_{ji}$.

At time t, agent $i, i \in N$, generates an estimate $\hat{x}_i(t) \in \mathbb{R}^{d_x}$ of $x(t)$ according to

$$\hat{x}_i(t) = g_{i,t}(I_i(t)),$$

where $g_{i,t}$ is called the *estimation rule* at time t. The collection $g_i \coloneqq (g_{i,0}, g_{i,1}, \dots)$ is called the *estimation strategy* of agent i and $g \coloneqq (g_1, \dots, g_n)$ is the *estimation strategy profile* of all agents.

There is another graph $\mathcal{G}^e$ associated with the agents, which we call the *estimation weight graph*, which is a weighted undirected graph with self-loops at each node. For any node i, let N_i^e denotes the set of estimation neighbors of i. For every edge (i, j), the associated weight M_{ij} is a $d_x \times d_x$ positive definite matrix.

For ease of notation, let $\hat{x}(t) = \text{vec}(\hat{x}_1(t), \dots, \hat{x}_n(t))$. The estimation error at time t is measured by

$$\begin{aligned} c(x(t), \hat{x}(t)) &= \sum_{i\in N} (x(t) - \hat{x}_i(t))^\intercal M_{ii} (x(t) - \hat{x}_i(t)) \\ &+ \frac{1}{2} \sum_{i\in N} \sum_{j\in N_i^e} (\hat{x}_i(t) - \hat{x}_j(t))^\intercal M_{ij} (\hat{x}_i(t) - \hat{x}_j(t)). \end{aligned} \tag{4}$$

Given a horizon T, the performance of an estimation strategy profile g is given by

$$J(g) = \mathbb{E}^g\Bigg[\sum_{t=0}^{T-1} c(x(t), \hat{x}(t))\Bigg]. \tag{5}$$

We are interested in the following optimization problem.

Problem 1 *Given matrices A, $\{C_i\}_{i\in N}$, Σ_x, Q_0, $\{Q_i\}_{i\in N}$, a* communication graph *$\mathcal{G}^c$ (and the corresponding weights d_{ij}), an* estimation weight graph *$\mathcal{G}^e$ (and the corresponding weights M_{ij}), and a horizon T, choose an estimation strategy profile g to minimize $J(g)$ given by* (5).

Remark 1 The per-step cost in (4) maybe interpreted as follows. Each agent wants to choose an estimate that is close to the true state $x(t)$ and is also close to the estimates chosen by its neighbors (according to the estimation weight graph). In the limiting case, when the coupling weights M_{ij}, $i \neq j$, between agents go to zero, the estimation problem is decoupled and the optimal estimates are given by $\mathbb{E}[x(t)|I_i(t)]$. On the other hand, when the coupling weights M_{ij}, $i \neq j$, between agents go to infinity (or equivalently, the self weights M_{ii} go to zero), the optimal estimation strategy is for all agents to pick a constant value, say 0. For moderate values of coupling weights M_{ij}, $i \neq j$, one expects the optimal estimates to be something in between these two extremes. This intuition is formalized in our main result.

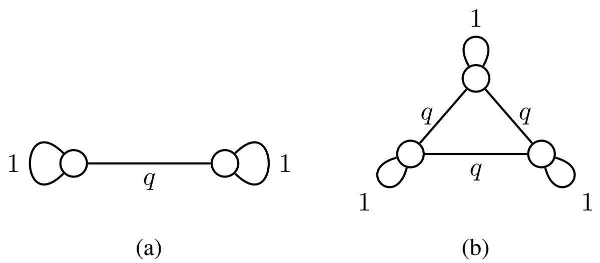

Fig. 1: Examples of estimation weight graphs with two and three nodes.

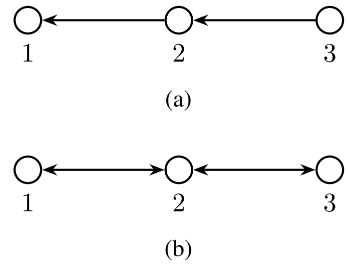

Fig. 2: Examples of neighborhood sharing information structure. In both graphs, the delay along each link is 1.

III. Some Illustrative Examples

To fix ideas, in this section we provide some illustrative examples of estimation weight graphs $\mathcal{G}^e$ and communication graphs $\mathcal{G}^c$.

A. Examples of Estimation weight Graphs

1) A two node estimation weight graph: Consider a system with two agents and an estimation weight graph as shown in Figure 1a. Suppose, $d_x = 1$, $M_{11} = M_{22} = 1$, and $M_{12} = q$. Thus, the cost function is

$$(x(t) - \hat{x}_1(t))^2 + (x(t) - \hat{x}_2(t))^2 + q(\hat{x}_1(t) - \hat{x}_2(t))^2.$$

Each agent wants to estimate $x(t)$ but also wants to be close to the estimates of the other agent. In the special case when $T = 1$, $C_i = 1$, and $Q_i = Q$ is a constant for $i \in N$, it can be shown that the optimal estimates are

$$\hat{x}_i(0) = \frac{\Sigma_x}{\Sigma_x + (1+q)Q} y_i(0).$$

In the sequel, we will show how to find similar estimates when $T > 1$.

2) A three node estimation weight graph: Consider a system with three agents and an estimation weight graph as shown in Figure 1b. Suppose $d_x = 1$, $M_{ii} = 1$, $M_{ij} = q$ if $|i - j| = 1 \pmod 3$ and 0 otherwise. In the special case when $T = 1$, $C_i = 1$, and $Q_i = Q$ is constant for all $i \in N$, it can be shown that the optimal estimates are

$$\hat{x}_i(0) = \frac{\Sigma_x}{\Sigma_x + (1+2q)Q} y_i(0).$$

In the sequel, we will show how to find similar estimates when $T > 1$.

In both cases, define

$$\hat{x}_i^{\mathrm{loc}}(0) = \mathbb{E}[x(0)|I_i(0)] = \frac{\Sigma_x}{\Sigma_x + Q} y_i(0)$$

Then the optimal estimates are of the form

$$\hat{x}_i(0) = F_i(0)\hat{x}_i^{\mathrm{loc}}(0).$$

Our main result is to show that such a structure holds in general; i.e., there exist gains $\{\{F_i(t)\}_{i\in N}\}_{t\geq 0}$ such that

$$\hat{x}_i(t) = F_i(t)\hat{x}_i^{\mathrm{loc}}(t).$$

The details are presented in Theorem 1.

B. Examples of Communication Graphs

The communication graph determines the information structure of problem. For that reason, we characterize a communication graph by the information structure induced by it.

1) No sharing information structure: Perhaps the simplest communication graph is one in which the agents don't share any information. We call such a graph as a graph with no information sharing.

2) Neighborhood sharing information structure: Consider a scenario in which agents share their information to their out-neighbors with one step delay. We refer to the resulting information structure as (one step delayed) neighborhood sharing information structure.

For example, consider the two communication graphs shown in Fig 2. In the communication graph of Fig 2a, the information structure is

$$\begin{aligned} I_1(t) &= \{y_1(0{:}t), y_2(0{:}t-1), y_3(0{:}t-2)\}, \\ I_2(t) &= \{y_2(0{:}t), y_3(0{:}t-1)\}, \\ I_3(t) &= \{y_3(0{:}t)\}. \end{aligned}$$

Similarly, in the communication graph of Fig 2b, the information structure is

$$\begin{aligned} I_1(t) &= \{y_1(0{:}t), y_2(0{:}t-1), y_3(0{:}t-2)\}, \\ I_2(t) &= \{y_1(0{:}t-1), y_2(0{:}t), y_3(0{:}t-1)\}, \\ I_3(t) &= \{y_1(0{:}t-2), y_2(0{:}t-1), y_3(0{:}t)\}. \end{aligned}$$

3) d-step delayed broadcast information structure: Consider a scenario in which agents broadcast their information to all other agents with d-step delay. This corresponds to a complete graph with a d-unit delay along each edge. We refer to the resulting information structure as d-step delayed broadcast information structure. The information available at each agent is

$$I_i(t) = \{y_i(t-d+1{:}t), y(1{:}t-d)\}.$$

This information structure is same as the d-step delayed sharing information structure introduced in [21]. Note that as $d \to \infty$, the d-step delayed broadcast information structure converges to the no sharing information structure.

IV. Main Results

A. Structure of optimal estimation strategies

Let L^e denote the weighted Laplacian matrix of the estimation weight graph $\mathcal{G}^e$, i.e., $L^e = [L^e_{ij}]_{i,j\in N}$, where

$$L^e_{ij} = \begin{cases} M_{ii} + \sum_{k\in N^e_i} M_{ik} & i = j \\ -M_{ij} & i \neq j. \end{cases}$$

Define the local estimates at agent $i, i \in N$, at time t as

$$\hat{x}^{\text{loc}}_i(t) = \mathbb{E}[x(t)|I_i(t)].$$

The local estimate $\hat{x}^{\text{loc}}_i(t)$ may be viewed as the agent's estimate when there is a positive weight on the estimate being close to the true state $x(t)$ but a zero weight on the estimate being close to its (estimation) neighbors (also see Remark 1 above and Corollary 1 below).

Let $\hat{P}_{ij}(t)$ denote $\operatorname{cov}\left(\hat{x}^{\text{loc}}_i(t), \hat{x}^{\text{loc}}_j(t)\right)$. Then, our main result is the following.

Theorem 1 *The optimal estimation strategy in Problem 1 is unique and linear in the local estimates, i.e.,*

$$\hat{x}_i(t) = F_i(t)\hat{x}^{\text{loc}}_i(t), \tag{6}$$

where the gains $\{F_i(t)\}_{i\in N}$ are given as follows. Define:

$$\begin{aligned} \Gamma_{ij}(t) &= \hat{P}_{ij}(t) \otimes L^e_{ij}, \\ \Gamma(t) &= [\Gamma_{ij}(t)]_{i,j\in N}, \\ F(t) &= \operatorname{vec}(F_1(t), \dots, F_n(t)), \\ \eta(t) &= \operatorname{vec}(M_{11}\hat{P}_{11}(t), \dots, M_{nn}\hat{P}_{nn}(t)). \end{aligned}$$

Then,

$$F(t) = \Gamma(t)^{-1}\eta(t). \tag{7}$$

Furthermore, the optimal performance is given by

$$J^* = \sum_{t=0}^{T-1} \left[\eta(t)^\intercal \Gamma(t)^{-1}\eta(t) + \operatorname{Tr}(MP_x(t))\right], \tag{8}$$

where $M = \sum_{i\in N} M_{ii}$, and $P_x(t) = \operatorname{var}(x(t))$ which can be recursively computed as follows: $P_x(0) = \Sigma_x$ and for $t > 0$,

$$P_x(t) = AP_x(t-1)A^\intercal + Q_0. \tag{9}$$

The proof is presented in Section V-A.

Remark 2 Theorem 1 shows there is a one way separation between the impact of the estimation weight graph and the communication graph on the optimal estimates. In particular, for any agent $i \in N$, the optimal estimate $\hat{x}_i(t)$ is a linear function of the local estimate $\hat{x}^{\text{loc}}_i(t)$. The local estimates $\{\hat{x}^{\text{loc}}_i(t)\}_{i\in N}$ depend only on the communication graph $\mathcal{G}^c$ but not on the estimation graph $\mathcal{G}^e$. This is the one way separation between the estimation and the communication graphs. On the other hand, the gains $\{F_i(t)\}_{i\in N}$ depend both on the communication graph $\mathcal{G}^c$ (through $\hat{P}_{ij}(t)$) and the estimation weight graph $\mathcal{G}^e$ (through L^e_{ij}).

Corollary 1

1) *If for all $i, j \in N$, $i \neq j$, we have that $M_{ij} = 0$, then $\hat{x}_i(t) = \hat{x}^{\text{loc}}_i(t)$.*
2) *If for all $i \in N$, we have that $M_{ii} = 0$, then $\hat{x}_i(t) = 0$.*

The proof is presented in Section V-B. See [22] foro details.

It is possible to recursively compute the local estimates $\{\hat{x}^{\text{loc}}_i(t)\}_{i\in N}$ and their covariances $\{\hat{P}_{ij}(t)\}_{i\in N}$. The exact recursive expressions depend on the communication graph $\mathcal{G}^c$. Below, we show how to derive such recursive expressions for the illustrative examples considered in Sec. III-B.

B. An Alternative Cost Structure

In certain applications, it is more natural to work with a per step estimation cost of the form

$$c(X(t), \hat{x}(t)) = (X(t)\hat{x}(t))^\intercal S(X(t) - \hat{x}(t)).$$

where $X(t) = \operatorname{vec}(x(t), \dots, x(t))$ and $\hat{x}(t) = \operatorname{vec}(\hat{x}_1(t), \dots, \hat{x}_n(t))$. Note that the above cost function can be written as

$$c(X(t), \hat{x}(t)) = x(t)^\intercal \bar{S}x(t) + \hat{x}(t)^\intercal S\hat{x}(t) + 2\hat{x}(t)^\intercal Dx(t),$$

where $\bar{S} = \sum_{i\in N}\sum_{j\in N^e_i} S_{ij}$, $D = \operatorname{rows}(D_1, \dots, D_n)$, and $D_i = -(S_{ii} + \sum_{k\in N^e_i} S_{ik})$.

The cost is a quadratic function in $x(t)$ and $\hat{x}(t)$. So, the result of Theorem 1 (which relies on Radner's result for static teams [23]) is also applicable for the above model with $L^e = S$ and $\eta = \operatorname{vec}(D_1\hat{P}_{11}, \dots, D_n\hat{P}_{nn})$.

C. No Sharing Information Structure

For the no sharing information structure, the information available at agent i is given by $I_i(t) = y_i(0{:}t)$. Thus, the local estimates $\hat{x}^{\text{loc}}_i(t) = \mathbb{E}[x(t)|I_i(t)]$ can be computed using standard Kalman filtering equations, one at each agent. In addition, we need to carry around some additional covariance matrices to recursively compute $\hat{P}_{ij}(t)$. To succiently write the results, for any $i \in N$, define $\hat{x}^{\text{loc}}_i(t^-) = \mathbb{E}[x(t)|y_i(0{:}t-1)]$, $\tilde{x}^{\text{loc}}_i(t) = x(t) - \hat{x}^{\text{loc}}_i(t)$, and $\tilde{x}^{\text{loc}}_i(t^-) = x(t) - \hat{x}^{\text{loc}}_i(t^-)$. Furthermore, for any $i, j \in N$, define $P_{ij}(t) = \operatorname{cov}(\tilde{x}_i(t), \tilde{x}_j(t))$ and $P_{ij}(t^-) = \operatorname{cov}(\tilde{x}_i(t^-), \tilde{x}_j(t^-))$. Then, we have the following.

Theorem 2 *In the communication graph with no sharing information structure, the local estimates $\hat{x}^{\text{loc}}_i(t)$ are given by standard Kalman filtering equations. In particular,*

$$\hat{x}^{\text{loc}}_i(t) = \hat{x}^{\text{loc}}_i(t^-) + K_i(t)(y_i(t) - C_i\hat{x}^{\text{loc}}_i(t^-)), \tag{10}$$

$$\hat{x}^{\text{loc}}_i(t^-) = A\hat{x}^{\text{loc}}_i(t-1), \tag{11}$$

where $K_i(t)$ is a $d_x \times d^i_y$ matrix given by

$$K_i(t) = P_{ii}(t^-)C_i^\intercal[C_iP_{ii}(t^-)C_i^\intercal + Q_i]^{-1}.$$

and $P_{ij}(t)$ and $P_{ij}(t^-)$ can be recursively computed as follows: $P_{ij}(0) = \Sigma_x$, and for $t \geq 0$,

$$P_{ij}(t) = \Delta_i(t)P_{ij}(t^-)\Delta_j(t)^\intercal + \delta_{i,j}K_i(t)Q_iK_j(t)^\intercal, \tag{12}$$

$$P_{ij}(t^-) = AP_{ij}(t-1)A^\intercal + Q_0, \tag{13}$$

where $\Delta_i(t) = I - K_i(t)C_i$ and δ_{ij} is the Kronecker delta function. Finally, the covariance between local estimates is

$$\hat{P}_{ij}(t) = P_x(t) - P_{ii}(t) - P_{jj}(t) + P_{ij}(t), \qquad (14)$$

where $P_x(t)$ is given by (9).

The proof is presented in Section V-C. We can show that (14) holds by using the Orthogonal Projection Theorem. The rest of the derivation is similar to standard calculations in Kalman filtering. See [22] for details.

Remark 3 The covariances $P_{ij}(t)$, $P_{ij}(t^-)$, $\hat{P}_{ij}(t)$, and $P_x(t)$ are data independent and can be pre-computed before the system starts running. Since the Kalman gains $K_i(t)$ only depend on these covariances and the system matrices A and $\{C_i\}_{i\in N}$, they can also be pre-computed before the system starts running. When the system is running, the agents only need to keep track of $\hat{x}_i^{\mathrm{loc}}(t)$ using (10) and (11).

The no sharing information structure with the cost function introduced in Section IV-B was investigated in [11], [12]. They used a different method and directly identified the estimates $\hat{x}_i(t)$ without computing the local estimates $\hat{x}_i^{\mathrm{loc}}(t)$. However, their method is tuned to the no sharing graph and does not immediately generalize to other information structures.

D. d-step delayed broadcast information structure

For the d-step delayed broadcast (as well as the neighborhood sharing) information structure, we follow the common information approach of [24], we write the information structure in (3) as

$$I_i(t) = \{I_i^{\mathrm{loc}}(t), I^{\mathrm{com}}(t)\},$$

where

$$I^{\mathrm{com}}(t) \coloneqq \bigcap_{i\in N} I_i(t) = y(1{:}t-d)$$

is the *common information* among all agents and

$$I_i^{\mathrm{loc}}(t) \coloneqq I_i(t) \setminus I^{\mathrm{com}}(t) = y_i(t-d+1{:}t)$$

is the *local information* at agent i.

With this split of the information set into common and local information, we may view the local estimate

$$\hat{x}_i^{\mathrm{loc}}(t) \coloneqq \mathbb{E}[x(t)|I_i(t)] = \mathbb{E}[x(t)|I^{\mathrm{com}}(t), I_i^{\mathrm{loc}}(t)]$$

as the update equation of a Kalman filter when the agent has information $I^{\mathrm{com}}(t)$ and observes new information $I_i^{\mathrm{loc}}(t)$. Then, using standard Kalman filtering equations, we can write

$$\begin{aligned}\hat{x}_i^{\mathrm{loc}}(t) &= \mathbb{E}[x(t)|I^{\mathrm{com}}(t)] \\ &\quad + K_i(t)\big[I_i^{\mathrm{loc}}(t) - \mathbb{E}[I_i^{\mathrm{loc}}(t)|I^{\mathrm{com}}(t)]\big], \qquad (15)\end{aligned}$$

for an appropriately defined Kalman filtering gain $K_i(t)$. In the remainder of this subsection, we show how to simplify $\mathbb{E}[x(t)|I^{\mathrm{com}}(t)]$ and $\mathbb{E}[I_i^{\mathrm{loc}}(t)|I^{\mathrm{com}}(t)]$ and identify the gain $K_i(t)$.

Following the idea proposed in [21], we define the common information based d-step delayed state estimate as follows:

$$\begin{aligned}\hat{x}^{\mathrm{com}}(t-d+1) &\coloneqq \mathbb{E}[x(t-d+1) \mid I^{\mathrm{com}}(t)] \\ &= \mathbb{E}[x(t-d+1) \mid y(0{:}t-d)].\end{aligned}$$

The recursive updates for $\hat{x}^{\mathrm{com}}(t-d+1)$ can be written using standard Kalman filtering equations. In order to do so, observe that the common observation $y(t)$ may be written as

$$y(t) = Cx(t) + w(t), \qquad (16)$$

where $C = \mathrm{rows}(C_1, \ldots, C_n)$ and $w(t) = \mathrm{vec}(w_1(t), \ldots, w_n(t))$. Therefore, we have $\hat{x}^{\mathrm{com}}(0) = 0$ and for $t > 0$,

$$\hat{x}^{\mathrm{com}}(t+1) = A\hat{x}^{\mathrm{com}}(t) + AK(t)\big[y(t) - C\hat{x}^{\mathrm{com}}(t)\big], \quad (17)$$

where

$$K(t) = P(t)C^{\intercal}\big[CP(t)C^{\intercal} + Q\big]^{-1},$$

$Q = \mathrm{diag}(Q_1, \ldots, Q_n)$, and $P(t) = \mathrm{var}(x(t) - \hat{x}^{\mathrm{com}}(t))$.

Since all the primitive random variables are Gaussian, the covariance $P(t)$ is data independent and can be pre-computed recursively as follows: $P(0) = \Sigma_x$ and for $t > 0$,

$$\begin{aligned}P(t+1) &= A\Delta(t)P(t)\Delta(t)^{\intercal}A^{\intercal} \\ &\quad + AK(t)QK(t)^{\intercal}A^{\intercal} + Q_0, \quad (18)\end{aligned}$$

where $\Delta(t) = I - K(t)C$.

To succinctly write $I_i^{\mathrm{loc}}(t)$ in terms of the delayed state $x(t-d+1)$, we define the following. For any $k \in \mathbb{Z}_{>0}$:

1) define the operator $\mathscr{C}^{(k)}$ that maps any $m \times d_x$ matrix $\tilde{C}$ (where $m \in \mathbb{Z}_{>0}$ is arbitrary) to a $km \times d_x$ matrix as follows:
$$\mathscr{C}^{(k)}\tilde{C} = \mathrm{rows}(\tilde{C}A^{k-1}, \cdots, \tilde{C}A, \tilde{C})$$
2) for any $t, \ell \in \mathbb{Z}_{>0}$, such that $\ell \le t$, define the $d_x \times 1$ random vector $w_0^{(k)}$ as follows:
$$\bar{w}_0^{(k)}(\ell, t) = \sum_{\tau=t-k+1}^{t-\ell-1} A^{t-\ell-\tau-1}w_0(\tau). \qquad (19)$$
Note that when $\ell \ge k-1$, then $w_0^{(k)}(\ell, t) = 0$.
3) for any $\ell \in \mathbb{Z}_{>0}$ such that $\ell < k$, define the $(k-\ell)d_x \times 1$ vectors $W_0^{(\ell,k)}(t)$ and $W_i^{(\ell,k)}(t)$ as follows:
$$W_0^{(\ell,k)}(t) = \mathrm{vec}(\bar{w}_0^{(k)}(\ell, t), \ldots, \bar{w}_0^{(k)}(k-1, t)) \quad (20)$$
$$W_i^{(\ell,k)}(t) = \mathrm{vec}(w_i(\ell), \ldots, w_i(k-1)). \qquad (21)$$

Then, from the dynamics (1) and observations (2), we can write

$$x(t) = A^{d-1}x(t-d+1) + \bar{w}_0(t), \qquad (22)$$
$$I_i^{\mathrm{loc}}(t) = \bar{C}_i x(t-d+1) + \bar{w}_i(t), \qquad (23)$$

where $\bar{w}_0(t) = w^{(d)}(0, t)$, and for $i \in N$, $\bar{C}_i = \mathscr{C}^{(d)}C_i$, and

$$\bar{w}_i(t) = C_i W_0^{(0,d)}(t) + W_i^{(0,d)}(t). \qquad (24)$$

From (16) and (24), observe that $I^{\mathrm{com}}(t)$ is a function of the primitive random variables up to time $t-d$, while $\bar{w}_i(t)$ is a function of the primitive random variables from time $t-d+1$ onwards. Thus, $\bar{w}_i(t)$ is independent of $I^{\mathrm{com}}(t)$. Hence, from (22) and (23), we get

$$\mathbb{E}[x(t) \mid I^{\mathrm{com}}(t)] = A^{d-1}\hat{x}^{\mathrm{com}}(t-d+1),$$
$$\mathbb{E}[I_i^{\mathrm{loc}}(t) \mid I^{\mathrm{com}}(t)] = \bar{C}_i\hat{x}^{\mathrm{com}}(t-d+1).$$

Substituting the above in (15), we get the following result.

Theorem 3 *In a communication graph with d-step delayed broadcast information structure, local estimates can be computed as follows:*

$$\hat{x}_i^{\text{loc}}(t) = A^{d-1}\hat{x}^{\text{com}}(t-d+1) + K_i(t)\big[I_i^{\text{loc}}(t) - \bar{C}_i\hat{x}^{\text{com}}(t-d+1)\big], \quad (25)$$

where $\hat{x}^{\text{com}}(t)$ is updated according to (17)–(18), $K_i(t)$ is a $d_x \times (d \cdot d_y^i)$ matrix given by

$$K_i(t) = [A^{d-1}P(t-d+1)\bar{C}_i^{\mathsf{T}} + \bar{P}_{0i}(t)] [\bar{C}_i P(t-d+1)\bar{C}_i^{\mathsf{T}} + \bar{P}_{ii}(t)]^{-1}, \quad (26)$$

$P(t)$ is given by (18), and $\bar{P}_{ij}(t) = \operatorname{cov}(\bar{w}_i(t), \bar{w}_j(t))$, $i \in \{0\} \cup N$, $j \in N$. The formulas for computing $\bar{P}_{ij}(t)$ are given later in (42) and (43).

The covariances between local estimates are:

$$\hat{P}_{ij}(t) = A^{d-1}(P_x(t-d+1) - P(t-d+1))(A^{d-1})^{\mathsf{T}} + K_i(t)[\bar{C}_i P(t-d+1)\bar{C}_j^{\mathsf{T}} + \bar{P}_{ij}(t)]K_j(t)^{\mathsf{T}},$$

where $P_x(t)$ is defined in (9).

The proof is presented in Section V-D. Comments similar to Remark 3 holds in this case as well.

Remark 4 As in the case of no sharing information structure, the covariances $\bar{P}_{ij}(t)$, $\hat{P}_{ij}(t)$, $P(t)$, and $P_x(t)$ are data independent and can be pre-computed before the system starts running. Since the Kalman gains $K_i(t)$ and $K(t)$ only depend on these covariances and the system matrices A and $\{C_i\}_{i\in N}$, they can also be pre-computed before the system starts running. When the system is running, the agents only need to keep track of $\hat{x}^{\text{com}}(t)$ and $\hat{x}_i^{\text{loc}}(t)$ using (17) and (25).

For the case of one step delayed broadcast information structure (i.e., $d = 1$), the covariance between the agent's estimates simplifies. In particular, we have the following

Corollary 2 *For $d = 1$,*

$$\hat{P}_{ij}(t) = P_x(t) - P(t) + K_i(t)[C_i P(t) C_j^{\mathsf{T}} + \delta_{i,j} Q_i]K_j(t)^{\mathsf{T}}.$$

Corollary 3 *When $d = 1$, $\mathcal{G}_1^c$ represents a one step delayed sharing information structure and the information available to agent i at time t is $I_i(t) = \{y_i(t), y(1{:}t-1)\}$. Then,*

$$\hat{x}_i^{\text{loc}}(t) = \hat{x}^{\text{com}}(t) + K_i(t)\big[y_i(t) - C_i\hat{x}^{\text{com}}(t)\big],$$

where $\hat{x}^{\text{com}}(t)$ is updated from (17)-(18),

$$K_i(t) = P(t)C_i^{\mathsf{T}}[C_i P(t) C_i^{\mathsf{T}} + Q_i]^{-1},$$

and $P(t)$ is given by (18). The covariance between local estimates is:

$$\hat{P}_{ij}(t) = P_x(t) - P(t) + K_i(t)[C_i P(t) C_j^{\mathsf{T}} + \delta_{i,j} Q_i]K_j(t)^{\mathsf{T}}, \quad (27)$$

where $\delta_{i,j}$ denotes the Kronecker delta function.

E. Strongly connected graph with neighborhood sharing information structure

Consider a *strongly connected* communication graph with neighborhood sharing information structure. Let d^* be the diameter of the graph, i.e.,

$$d^* = \max_{i,j\in N} \ell_{ij},$$

where ℓ_{ij} is the geodesic distance between nodes i and j. For any agent $i \in N$ and any positive integer $k \le d^*$, let $N_i^{c,k}$ denote the k-hop in-neighbors of node i, i.e.,

$$N_i^{c,k} = \{j \in N : \ell_{ji} = k\}.$$

For ease of notation, we define $N_i^{c,0} = \{i\}$. Then, the information available at agent $i \in N$ can be written as

$$I_i(t) = \bigcup_{k=0}^{d^*} \bigcup_{j\in N_i^{c,k}} \{y_j(1{:}t-k)\}.$$

As before, we split this information into common information $I^{\text{com}}(t)$ and local information $I_i^{\text{loc}}(t)$, where

$$\begin{aligned} I^{\text{com}}(t) &\coloneqq \bigcap_{i\in N} I_i(t) = \{y(1{:}t-d^*)\}, \\ I^{\text{loc}}(t) &\coloneqq I_i(t) \setminus I^{\text{com}}(t) \\ &= \bigcup_{k=0}^{d^*} \bigcup_{j\in N_i^{c,k}} \{y_j(t-d^*+1{:}t-k)\}. \end{aligned}$$

We follow the same approach as the d-step delayed broadcast information structure. In particular, we decompose $\hat{x}_i^*(t)$ as in (15).

As before, in order to simplify $\mathbb{E}[x(t) \mid I^{\text{com}}(t)]$ and $\mathbb{E}[I_i^{\text{loc}}(t) \mid I^{\text{com}}(t)]$, we define the common information based d^*-step delayed estimate as

$$\begin{aligned} \hat{x}^{\text{com}}(t-d^*+1) &\coloneqq \mathbb{E}[x(t-d^*+1) \mid I^{\text{com}}(t)] \\ &= \mathbb{E}[x(t-d^*+1) \mid y(1{:}t-d^*)]. \end{aligned}$$

We can write Kalman filtering equations for recursively updating $\hat{x}^{\text{com}}(t-d^*+1)$ that are similar to (17)–(18) with d replaced by d^*. Similar to (22) and (23), we can write

$$x(t) = A^{d^*-1}x(t-d^*+1) + \bar{w}_0^*(t), \quad (28)$$

$$I_i^{\text{loc}}(t) = \bar{C}_i^* x(t-d^*+1) + \bar{w}_i^*(t), \quad (29)$$

where $\bar{w}_0^*(t) = w^{(d^*)}(0,t)$,

$$\bar{C}_i^* = \operatorname{rows}(\mathscr{C}^{(d^*)}C_i^{(0)}, \mathscr{C}^{(d^*-1)}C_i^{(1)}, \dots, \mathscr{C}^{(1)}C_i^{(d^*-1)}),$$

where $C_i^{(k)} = \operatorname{rows}\big(\{C_j\}_{j\in N_i^{c,k}}\big)$,

$$W_0^*(t) = \operatorname{vec}(C_i^{(0)}W_0^{(0,d^*)}(t), C_i^{(1)}W_0^{(1,d^*)}(t), \dots, C_i^{(d^*-1)}W_0^{(d^*-1,d^*)}(t)) \quad (30)$$

and

$$W_i^*(t) = \operatorname{vec}(W_i^{(0,d^*)}(t), W_i^{(1,d^*)}(t), \dots, W_i^{(d^*-1,d^*)}(t)). \quad (31)$$

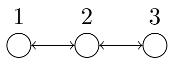

Fig. 3: A complete graph with $n = 3$ and $d = 1$

As before, we have

$$\mathbb{E}[x(t) \mid I^{\text{com}}(t)] = A^{d^*-1}\hat{x}^{\text{com}}(t - d^* + 1),$$
$$\mathbb{E}[I_i^{\text{loc}}(t) \mid I^{\text{com}}(t)] = \bar{C}_i^* \hat{x}^{\text{com}}(t - d^* + 1).$$

Substituting this in (15), we get the following which is similar to Theorem 3.

Theorem 4 *In the strongly connected communication graph, the local estimates can be computed as follows:*

$$\hat{x}_i^{\text{loc}}(t) = A^{d^*-1}\hat{x}^{\text{com}}(t - d^* + 1) + K_i(t)\big[I_i^{\text{loc}}(t) - \bar{C}_i^* \hat{x}^{\text{com}}(t - d^* + 1)\big], \quad (32)$$

where $K_i(t)$ is given by

$$K_i(t) = [A^{d^*-1} P(t - d^* + 1)\bar{C}_i^{*\intercal} + \bar{P}_{0i}^*(t)] [\bar{C}_i^* P(t - d^* + 1)\bar{C}_i^{*\intercal} + \bar{P}_{ii}^*(t)]^{-1},$$

where $\bar{P}_{ij}^*(t) = \operatorname{cov}(\bar{w}_i^*(t), \bar{w}_j^*(t))$, $i \in \{0\} \cup N$, $j \in N$ and the formulas for computing $\bar{P}_{ij}^*(t)$ can be written similar to (42) *and* (43).

The covariances between local estimates are:

$$\hat{P}_{ij}(t) = A^{d^*-1}(P_x(t - d^* + 1) - P(t - d^* + 1))(A^{d^*-1})^\intercal + K_i(t)[\bar{C}_i^* P(t - d^* + 1)\bar{C}_j^{*\intercal} + \bar{P}_{ij}^*(t)]K_j(t)^\intercal,$$

where $P_x(t)$ is defined in (9).

The proof is given in Section V-E. A remark similar to Remark 3 holds for this case as well.

As an illustrative example, reconsider the communication graph of Fig. 2b. The common information $I^{\text{com}}(t)$ is $\{y(1{:}t - 2)\}$. The local information available at the different agents is

$$I_1^{\text{loc}} = \big[y_1(t), y_1(t-1), y_2(t-1)\big],$$

$$I_2^{\text{loc}} = \big[y_1(t-1), y_2(t), y_2(t-1), y_3(t-1)\big],$$

$$I_1^{\text{loc}} = \big[y_3(t), y_3(t-1), y_2(t-1)\big],$$

Example 1 Consider the strongly connected graph shown in Figure 3 with $n = 3$ and $d^* = 1$. The information available to agent 1 at time t is

$$I_i(t) = \{y_1(t), y_1(t-1), y_2(t-1), y(1 : t-2)\}.$$

Therefore, the local information is

$$I_i^{\text{loc}}(t) = \begin{bmatrix} y_1(t) \\ y_1(t-1) \\ y_2(t-2) \end{bmatrix} = \begin{bmatrix} C_1 x(t) + w_1(t) \\ C_1 x(t-1) + w_1(t-1) \\ C_2 x(t-2) + w_2(t-2) \end{bmatrix}$$

Using (1) we can find $\bar{C}_1^*$, $\bar{w}_0^*(t)$ and $\bar{w}_1^*(t)$ as follows.

$$\bar{C}_1^* = \begin{bmatrix} C_1 A^2 \\ C_1 A \\ C_2 \end{bmatrix},$$

$$\bar{w}_1^*(t) = \begin{bmatrix} C_1 A w_0(t-2) + C_1 w_0(t-1) + w_1(t) \\ C_1 w_0(t-2) + w_1(t-1) \\ w_2(t-2) \end{bmatrix} = \begin{bmatrix} C_1 A w_0(t-2) + C_1 w_0(t-1) \\ C_1 w_0(t-2) \\ 0 \end{bmatrix} + \begin{bmatrix} w_1(t) \\ w_1(t-1) \\ w_2(t-2) \end{bmatrix},$$

and $\bar{w}_0^*(t) = C_1 A w_0(t-2) + C_1 w_0(t-1)$. We can write $\bar{C}_i^*$ and $\bar{w}_i^*(t)$ for all of the agents using the same approach. Hence,

$$\bar{P}_{0i}^* = [C_i A Q_0 A^\intercal C_i^\intercal + C_i Q_0 C_i^\intercal, C_i A Q_0 C_i^\intercal + C_i Q_0, 0]$$

and for $i = 1$ and $j = 2$

$$\bar{P}_{11} = \begin{bmatrix} C_1 A Q_0 A^\intercal C_1^\intercal + C_1 Q_0 C_1^\intercal & C_1 A Q_0 C_1^\intercal & 0 \\ C_1 Q_0 A^\intercal C_1^\intercal & C_1 Q_0 C_1^\intercal & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} Q_1 & 0 & 0 \\ 0 & Q_1 & 0 \\ 0 & 0 & Q_2 \end{bmatrix}$$

V. Proofs

A. Proof of Theorem 1

Following Witsenhausen's intrinsic model [25], we consider each agent at each time as a separate decision maker (DM). Thus, the system consists of nT DMs, indexed as $\{\{\mathrm{DM}_{i.t}\}_{i=1}^n\}_{t=0}^{T-1}$, where $\mathrm{DM}_{i.t}$ corresponds to agent i and time t. The actions of any DM do not effect the observation of any other DM. Thus, the team is a static team. Moreover, the primitive random variables $\{x(0), w_0(0{:}T-2), w_1(0{:}T-1), \dots, w_n(0{:}T-1)\}$ are jointly Gaussian and the cost function is quadratic in the primitive random variables and the actions of all DMs. Therefore, from [23, Theorem 5], we get that the optimal decision rule is of the form

$$\hat{x}_i(t) = F_i^\circ(t) I_i(0{:}t), \quad (33)$$

where the gains $\{\{F_i^\circ(t)\}_{i=1}^n\}_{t=0}^{T-1}$ are computed as in [23, Theorem 5]. Note that there is no additive offset term in (33) because all primitive random variables are zero mean.

One of the difficulties is that the dimension of the information set $I_i(t)$ is increasing with time, and, therefore, so is the dimension of $F_i^\circ(t)$. To avoid this dependence on time, observe that

$$\hat{x}_i^{\text{loc}}(t) = \mathbb{E}[x(t) | I_i(t)] = \operatorname{cov}(x(t), I_i(t)) \operatorname{var}(I_i(t))^{-1} I_i(t).$$

Thus, we can write

$$I_i(t) = \operatorname{var}(I_i(t)) \operatorname{cov}(x(t), I_i(t))^{-1} \hat{x}_i^{\text{loc}}(t). \quad (34)$$

Substituting (34) in (33), we get that the optimal decision rules can be written as

$$\hat{x}_i(t) = F_i(t) \hat{x}_i^{\text{loc}}(t),$$

with $F_i(t) = F_i^\circ(t)\operatorname{var}(I_i(t))\operatorname{cov}(x(t), I_i(t))^{-1}$. This proves the first part of Theorem 1.

The equations for computing the gains as well as the optimal performance can be obtained by following the same approach as in [23]. The details are given in the Appendix. in [22].

B. Proof of Corollary 1

We prove the two parts separately.

1) When $M_{ij} = 0$, $i \neq j$, then

$$\Gamma(t) = \operatorname{diag}(\hat{P}_{11}(t) \otimes M_{11}, \dots, \hat{P}_{nn}(t) \otimes M_{nn}).$$

Hence,

$$\begin{aligned}\operatorname{vec}(F_i(t)) &= (\hat{P}_{ii}(t) \otimes M_{ii})^{-1} \operatorname{vec}(M_{ii}\hat{P}_{ii}(t)) \\ &= (\hat{P}_{ii}(t)^{-1} \otimes M_{ii}^{-1}) \operatorname{vec}(M_{ii}\hat{P}_{ii}(t)) \\ &\overset{(a)}{=} \operatorname{vec}(M_{ii}^{-1}M_{ii}\hat{P}_{ii}(t)\hat{P}_{ii}(t)^{-1}) \\ &= \operatorname{vec}(I_{d_x \times d_x}),\end{aligned}$$

where (a) uses the following: for any matrices A, B, C of appropriate dimensions, $(C^\intercal \otimes A)\operatorname{vec}(B) = \operatorname{vec}(ABC)$. The only $d_x \times d_x$ matrix that satisfies the above equation is $F_i(t) = I_{d_x \times d_x}$. Therefore,

$$\hat{x}_i(t) = \hat{x}_i^{\text{loc}}(t).$$

2) When $M_{ii} = 0$, then $\eta(t) = 0$. Hence, $F(t) = 0$ and, therefore,

$$\hat{x}_i(t) = 0.$$

C. Proof of Theorem 2

The Kalman filtering equations for each agent are the standard ones. We need to compute $\hat{P}_{ij}(t)$ to complete the proof. We have,

$$\begin{aligned}\hat{P}_{ij}(t) &= \mathbb{E}[\hat{x}_i^{\text{loc}}(t)\hat{x}_j^{\text{loc}}(t)^\intercal] \\ &= \mathbb{E}[(x(t) - \tilde{x}_i^{\text{loc}}(t))(x(t) - \tilde{x}_j^{\text{loc}}(t))^\intercal] \\ &= \mathbb{E}[x(t)x(t)^\intercal] - \mathbb{E}[\tilde{x}_i^{\text{loc}}(t)x(t)^\intercal] \\ &\quad - \mathbb{E}[x(t)\tilde{x}_j^{\text{loc}}(t)^\intercal] + \mathbb{E}[\tilde{x}_i^{\text{loc}}(t)\tilde{x}_j^{\text{loc}}(t)^\intercal] \\ &= \mathbb{E}[x(t)x(t)^\intercal] - \mathbb{E}[\tilde{x}_i^{\text{loc}}(t)(\hat{x}_i^{\text{loc}}(t) + \tilde{x}_i^{\text{loc}}(t))^\intercal] \\ &\quad - \mathbb{E}[(\hat{x}_j^{\text{loc}}(t) + \tilde{x}_j^{\text{loc}}(t))\tilde{x}_j^{\text{loc}}(t)^\intercal] + \mathbb{E}[\tilde{x}_i^{\text{loc}}(t)\tilde{x}_j^{\text{loc}}(t)^\intercal] \\ &\overset{(a)}{=} \mathbb{E}[x(t)x(t)^\intercal] - \mathbb{E}[\tilde{x}_i^{\text{loc}}(t)\tilde{x}_i^{\text{loc}}(t)^\intercal] \\ &\quad - \mathbb{E}[\tilde{x}_j^{\text{loc}}(t)\tilde{x}_j^{\text{loc}}(t)^\intercal] + \mathbb{E}[\tilde{x}_i^{\text{loc}}(t)\tilde{x}_j^{\text{loc}}(t)^\intercal] \\ &= P_x(t) - P_{ii}(t) - P_{jj}(t) + P_{ij}(t),\end{aligned}$$

where (a) uses $\mathbb{E}[\hat{x}_i^{\text{loc}}(t)\tilde{x}_i^{\text{loc}}(t)^\intercal] = 0$. From (10) and (11), we have that

$$\begin{aligned}\tilde{x}_i^{\text{loc}}(t) &= \tilde{x}_i^{\text{loc}}(t^-) - K_i(t)(C_i\tilde{x}_i^{\text{loc}}(t^-) + w_i(t)) \\ &= \Delta_i(t)\tilde{x}_i^{\text{loc}}(t^-) - K_i(t)w_i(t), \\ \tilde{x}_i^{\text{loc}}(t^-) &= A\tilde{x}_i^{\text{loc}}(t-1) + w_0(t-1).\end{aligned}$$

Therefore,

$$\begin{aligned}P_{ij}(t) &= \mathbb{E}[(x_i(t) - \hat{x}_i^{\text{loc}}(t))(x_i(t) - \hat{x}_i^{\text{loc}}(t))^\intercal] \\ &= \Delta_i(t)P_{ij}(t^-)\Delta_j(t)^\intercal + \delta_{i,j}K_i(t)Q_iK_j(t)^\intercal, \\ P_{ij}(t^-) &= \mathbb{E}[(x_i(t) - \hat{x}_i^{\text{loc}}(t^-))(x_i(t) - \hat{x}_i^{\text{loc}}(t^-))^\intercal] \\ &= AP_{ij}(t-1)A^\intercal + Q_0,\end{aligned}$$

where $\Delta_i(t) = I - K_i(t)C_i$. Finally, combining (12), (13), and (14) gives $\hat{P}_{ij}(t)$.

D. Proof of Theorem 3

$$\tilde{x}(t) = x(t) - \mathbb{E}[x(t) \mid I^{\text{com}}(t)], \tag{35}$$

$$\tilde{x}^{\text{com}}(t) = x(t) - \hat{x}^{\text{com}}(t), \tag{36}$$

and

$$\tilde{I}_i^{\text{loc}}(t) = I_i^{\text{loc}}(t) - \mathbb{E}[I_i^{\text{loc}}(t) \mid I^{\text{com}}(t)]. \tag{37}$$

Using (22) and (23), we can write

$$\tilde{x}(t) = A^{d-1}\tilde{x}^{\text{com}}(t-d+1) + \bar{w}_0(t), \tag{38}$$

and

$$\tilde{I}_i^{\text{loc}}(t) = \bar{C}_i\tilde{x}^{\text{com}}(t-d+1) + \bar{w}_i(t). \tag{39}$$

Recall the Kalman filtering equation given in (15). The Kalman gain $K_i(t)$ is given by

$$K_i(t) = \operatorname{cov}(x(t), \tilde{I}_i^{\text{loc}}(t))\operatorname{var}(\tilde{I}_i^{\text{loc}}(t))^{-1}. \tag{40}$$

Using (35), we can write the first term in (40) as follows:

$$\operatorname{cov}(x(t), \tilde{I}_i^{\text{loc}}(t)) = \operatorname{cov}(\tilde{x}(t) + \mathbb{E}[x(t) \mid I^{\text{com}}(t)], \tilde{I}_i^{\text{loc}}(t))$$

$\mathbb{E}[\mathbb{E}[x(t) \mid I^{\text{com}}(t)], \tilde{I}_i^{\text{loc}}(t)] = 0$ as they are two orthogonal Gaussian random variables. Therefore,

$$\operatorname{cov}(x(t), \tilde{I}_i^{\text{loc}}(t)) = \operatorname{cov}(\tilde{x}(t), \tilde{I}_i^{\text{loc}}(t))$$

and

$$K_i(t) = \operatorname{cov}(\tilde{x}(t), \tilde{I}_i^{\text{loc}}(t))\operatorname{var}(\tilde{I}_i^{\text{loc}}(t))^{-1}. \tag{41}$$

In order to succinctly write the expressions for $\operatorname{cov}(\tilde{x}(t), \tilde{I}_i^{\text{loc}}(t))$ and $\operatorname{var}(\tilde{I}_i^{\text{loc}}(t))$, we define $\bar{P}_{00}(\ell_1, \ell_2, t) = \operatorname{cov}(w_0^{(d)}(\ell_1, t), w_0^{(d)}(\ell_2, t))$, and for $i \in \{0\} \cup N$ and $j \in N$, $\bar{P}_{ij}(t) = \operatorname{cov}(\bar{w}_i(t), \bar{w}_j(t))$.

Let $\delta_{i,j}$ denote the Kronecker delta function. Then from (19) we get,

$$\begin{aligned}&\bar{P}_{00}(\ell_1, \ell_2, t) = \\ &\sum_{l=\min\{0,t-d+1\}}^{t-\ell_1-1} \sum_{k=\min\{0,t-d+1\}}^{t-\ell_2-1} A^{t-\ell_1-l-1}Q_0(A^{t-\ell_2-k-1})^\intercal\delta_{l,k} \\ &= \sum_{l=\min\{0,t-d+1\}}^{t-\max\{\ell_1,\ell_2\}-1} A^{t-\ell_1-l-1}Q_0A^{t-\ell_2-l-1},\end{aligned}$$

which is zero when $|\ell_2 - \ell_1| \geq d$. Moreover, from (19) and (24),

$$\bar{P}_{0i}(t) = \begin{bmatrix}\bar{P}_{00}(0,0,t) & \cdots & \bar{P}_{00}(0,d-1,t)\end{bmatrix} C_i^\intercal \tag{42}$$

and from (24),

$$\bar{P}_{ij}(t) = \operatorname{diag}(\overbrace{Q_i, \ldots, Q_i}^{d \text{ times}})\delta_{i,j} + C_i \begin{bmatrix} \bar{P}_{00}(0,0,t) & \cdots & \bar{P}_{00}(0,d-1,t) \\ \vdots & \ddots & \vdots \\ \bar{P}_{00}(d-1,0,t) & \cdots & \bar{P}_{00}(d-1,d-1,t) \end{bmatrix} C_j^{\intercal}. \tag{43}$$

Note that the last block column of $\bar{P}_{0i}(t)$ and last block row and last block column of $\bar{P}_{ij}(t)$ are zero.

Now we come back to the expression of $K_i(t)$ given in (41). First consider

$$\begin{aligned} \operatorname{cov}(\tilde{x}(t), \tilde{I}_i^{\text{loc}}(t)) &= \mathbb{E}[\tilde{x}(t)\tilde{I}_i^{\text{loc}}(t)^{\intercal}] \\ &= A^{d-1}P(t-d+1)\bar{C}_i^{\intercal} + \bar{P}_{0i}(t), \end{aligned} \tag{44}$$

where the last equality follows from (38), (39), and the definition of $P(t-d+1)$ and $\bar{P}_{0i}(t)$. Now consider,

$$\begin{aligned} \operatorname{cov}(\tilde{I}_i^{\text{loc}}(t), \tilde{I}_j^{\text{loc}}(t)) &= \mathbb{E}[\tilde{I}_i^{\text{loc}}(t)\tilde{I}_i^{\text{loc}}(t)^{\intercal}] \\ &= \bar{C}_i P(t-d+1)\bar{C}_j^{\intercal} + \bar{P}_{ij}(t), \end{aligned} \tag{45}$$

where the last equality follows from (38), (39), and the definition of $P(t-d+1)$ and $\bar{P}_{ij}(t)$.

Substituting (44) and (45) in (41) gives the expression (26) for $K_i(t)$.

To derive the expression for $\hat{P}_{ij}(t)$, recall that $\hat{P}_{ij}(t) = \operatorname{cov}(\hat{x}_i^{\text{loc}}(t), \hat{x}_j^{\text{loc}}(t))$. Note that $\mathbb{E}[\hat{x}_i^{\text{loc}}(t)] = 0$. Therefore, using (25) and (38), we get

$$\begin{aligned} \hat{P}_{ij}(t) &= \mathbb{E}[\hat{x}_i^{\text{loc}}(t)\hat{x}_j^{\text{loc}}(t)^{\intercal}] \\ &= A^{d-1}\operatorname{var}(\hat{x}^{\text{com}}(t-d+1))\big(A^{d-1}\big)^{\intercal} \\ &\quad + K_i(t)\operatorname{cov}(\tilde{I}_i^{\text{loc}}(t), \tilde{I}_j^{\text{loc}}(t))K_j(t)^{\intercal}. \end{aligned} \tag{46}$$

Note that $\operatorname{var}(x(t)) = \operatorname{var}(\hat{x}^{\text{com}}(t)) + \operatorname{var}(\tilde{x}^{\text{com}}(t))$, which implies

$$\operatorname{var}(\hat{x}^{\text{com}}(t)) = \operatorname{var}(x(t)) - \operatorname{var}(\tilde{x}^{\text{com}}(t)) \tag{47}$$
$$= P_x(t) - P(t). \tag{48}$$

Substituting (45) and (48) in (46) gives the expression for $\hat{P}_{ij}(t)$.

E. Proof of Theorem 4

Define

$$\tilde{x}(t) = x(t) - \mathbb{E}[x(t) \mid I^{\text{com}}(t)],$$
$$\tilde{x}^{\text{com}}(t) = x(t) - \hat{x}^{\text{com}}(t),$$

and

$$\tilde{I}_i^{\text{loc}}(t) = I_i^{\text{loc}}(t) - \mathbb{E}[I_i^{\text{loc}}(t) \mid I^{\text{com}}(t)]$$

from (28) and (29), we can write

$$\tilde{x}(t) = A^{d^*-1}\tilde{x}^{\text{com}}(t-d^*+1) + \bar{w}_0^*(t), \tag{49}$$

and

$$\tilde{I}_i^{\text{loc}}(t) = \bar{C}_i^* \tilde{x}^{\text{com}}(t-d^*+1) + \bar{w}_i^*(t). \tag{50}$$

The rest of the proof is similar to the proof of Theorem 3. See [22] for details. The Kalman gain $K_i(t)$ is given by (41).

Now define, $\bar{P}_{00}^*(\ell_1, \ell_2, t) = \operatorname{cov}(w_0^*(\ell_1, t), w_0^*(\ell_2, t))$, and for $i \in \{0\} \cup N$ and $j \in N$, $\bar{P}_{ij}^*(t) = \operatorname{cov}(\bar{w}_i^*(t), \bar{w}_j^*(t))$. Then from (19) we get,

$$\begin{aligned} &\bar{P}_{00}^*(\ell_1, \ell_2, t) = \\ &\sum_{l=\min\{0,t-d^*+1\}}^{t-\ell_1-1} \sum_{k=\min\{0,t-d^*+1\}}^{t-\ell_2-1} A^{t-\ell_1-l-1}Q_0(A^{t-\ell_2-k-1})^{\intercal}\delta_{l,k} \\ &= \sum_{l=\min\{0,t-d^*+1\}}^{t-\max\{\ell_1,\ell_2\}-1} A^{t-\ell_1-l-1}Q_0 A^{t-\ell_2-l-1}, \end{aligned}$$

which is zero when $|\ell_2 - \ell_1| \geq d$. Moreover, from (19) and (30),

$$\bar{P}_{00}^{(*,\ell)}(t) = \big[(\bar{P}_{00}^*(0,\ell,t), \ldots, \bar{P}_{00}^*(0,d^*-1,t)C_i^{(\ell)\intercal}\big].$$

Then

$$\bar{P}_{0i}^*(t) = \big[\bar{P}_{00}^{(*,0)}(t), \bar{P}_{00}^{(*,1)}(t), \ldots, \bar{P}_{00}^{(*,d^*-1)}(t)\big]. \tag{51}$$

Let's define

$$\bar{P}_{ij}^*(\ell_1, \ell_2, t) = \mathbb{E}[C_i^{(\ell_1)}W_0^{(\ell_1,d^*)}(t)(C_j^{(\ell_2)}W_0^{(\ell_2,d^*)}(t))^{\intercal}].$$

Then from (20),

$$\bar{P}_{ij}^*(\ell_1, \ell_2, t) = C_i^{(k)} \begin{bmatrix} \bar{P}_{00}^{(*,0)}(\ell_1,\ell_2,t) & \cdots & \bar{P}_{00}(\ell_1,d^*-1,t) \\ \vdots & \ddots & \vdots \\ \bar{P}_{00}(d^*-1,\ell_2,t) & \cdots & \bar{P}_{00}(d^*-1,d^*-1,t) \end{bmatrix} C_j^{(l)\intercal}.$$

and from (31),

$$\bar{P}_{ij}^*(t) = \bar{Q}_i^* + \begin{bmatrix} \bar{P}_{ij}^*(0,0,t) & \cdots & \bar{P}_{ij}^*(0,d^*-1,t) \\ \vdots & \ddots & \vdots \\ \bar{P}_{ij}^*(d^*-1,0,t) & \cdots & \bar{P}_{ij}^*(d^*-1,d^*-1,t) \end{bmatrix}, \tag{52}$$

where $\bar{Q}_i^* = \operatorname{diag}\big(((Q_i)_{i \in N_i^{c,k}})_{k=0}^{d^*-1}\big)\delta_{i,j}$. Now consider

$$\begin{aligned} \operatorname{cov}(\tilde{x}(t), \tilde{I}_i^{\text{loc}}(t)) &= \mathbb{E}[\tilde{x}(t)\tilde{I}_i^{\text{loc}}(t)^{\intercal}] \\ &= A^{d-1}P(t-d+1)\bar{C}_i^{*\intercal} + \bar{P}_{0i}^*(t), \end{aligned} \tag{53}$$

where the last equality follows from (49), (50), and the definition of $P(t-d+1)$ and $\bar{P}_{0i}(t)$. Morever,

$$\begin{aligned} \operatorname{cov}(\tilde{I}_i^{\text{loc}}(t), \tilde{I}_j^{\text{loc}}(t)) &= \mathbb{E}[\tilde{I}_i^{\text{loc}}(t)\tilde{I}_i^{\text{loc}}(t)^{\intercal}] \\ &= \bar{C}_i^* P(t-d+1)\bar{C}_j^{*\intercal} + \bar{P}_{ij}^*(t), \end{aligned} \tag{54}$$

where the last equality follows from (49), (50), and the definition of $P(t-d+1)$ and $\bar{P}_{ij}^*(t)$.

Substituting (53) and (54) in (41) gives the expression for $K_i(t)$.

For the expression for $\hat{P}_{ij}(t)$, recall that $\hat{P}_{ij}(t)$ is $\operatorname{cov}(\hat{x}_i^{\text{loc}}(t), \hat{x}_j^{\text{loc}}(t))$. First note that $\operatorname{var}(x(t)) = \operatorname{var}(\hat{x}^{\text{com}}(t)) + \operatorname{var}(x(t) - \hat{x}^{\text{com}}(t))$, which implies

$$\operatorname{var}(\hat{x}^{\text{com}}(t)) = \operatorname{var}(x(t)) - \operatorname{var}(x(t) - \hat{x}^{\text{com}}(t)) \tag{55}$$
$$= P_x(t) - P(t) \tag{56}$$

Next note that $\mathbb{E}[\hat{x}_i^{\text{loc}}(t)] = 0$. Therefore, using (25) and (49), we get that

$$\begin{aligned}\hat{P}_{ij}(t) &= \mathbb{E}[\hat{x}_i^{\text{loc}}(t)\hat{x}_j^{\text{loc}}(t)^\intercal] \\ &= A^{d-1}\operatorname{var}(\hat{x}^{\text{com}}(t-d+1))\big(A^{d-1}\big)^\intercal \\ &\quad + K_i(t)\operatorname{cov}(\tilde{I}_i^{\text{loc}}(t), \tilde{I}_j^{\text{loc}}(t))K_j(t)^\intercal. \end{aligned} \tag{57}$$

Substituting (56) and (54) in (57) gives the expression for $\hat{P}_{ij}(t)$.

VI. Conclusion

We consider the problem of optimal decentralized estimation of a Gauss-Markov process with multiple agents. The per-step estimation cost is an estimation weight graph. In addition the agents communicate information to their neighbors over a directed communication graph with link delays.

Our main result is to show that the optimal estimates are linear in the local estimates at the agent (i.e., estimates determined by ignoring the cost coupling across the agents). The optimal gains can be obtained by solving a system of linear equations. We also derive expressions for recursively computing the local estimates for three representative communication graph.

In this paper, we restrict to finite horizon models. The results extend to infinite horizon long run average setup under standard stabilizability and observability conditions. We also restrict attention to the setup where all agents have a common (team) objectives. The results can be extended to setups where agents have individual (game) objectives.

Our results depend on the fact that there is no dynamical coupling between the agents. In general, it is difficult to say something about the general setup when the dynamics are controlled by the actions of all agents. It might be possible to generalize some of the results here to the case of controlled dynamics when the information structure is partially nested. We hope to address this case in the future.

Appendix

Expanding (4), we can write the following expression for the cost at time t as follows.

$$\begin{aligned}c(x(t), \hat{x}(t)) = &\sum_{k\in N}\sum_{j\in N}\hat{x}_k(t)^\intercal L^e_{kj}\hat{x}_j(t) \\ &- 2\sum_{k\in N}\hat{x}_k(t)^\intercal M_{kk}x(t) + x(t)^\intercal M x(t)\end{aligned}$$

The approach to find the gains as well as the optimal performance is the same as one in [23]. Consider agent i and arbitrarily fix the strategy $g_{-i,t}$ of all agents other than i at time t. For the strategy of agent i to be the best response to $g_{-i,t}$, it must satisfy

$$\begin{aligned}&\mathbb{E}^{g_{-i,t}}[c(x(t), g_{i,t}(I_i(t)), g_{-i,t}(I_{-i}(t))) \mid I_i(t)] \le \\ &\qquad \mathbb{E}^{g_{-i,t}}[c(x(t), \hat{x}_i(t), g_{-i,t}(I_{-i}(t))) \mid I_i(t)],\end{aligned}$$

where

$$I_{-i}(t) = \bigcup_{j\in N, j\neq i} I_j(t),$$

and $g_{-i,t}$ is the decision made by all of the players except player i. From [23, Theorem 1], a sufficient condition for the above to hold is

$$\frac{\partial}{\partial \hat{x}_i(t)}\mathbb{E}^{g_{-i,t}}[c(x(t), \hat{x}_i(t), g_{-i,t}(I_{-i}(t))) \mid I_i(t)] = 0. \tag{58}$$

Now, instead of using the local information we use the belief state $\hat{x}_i^{\text{loc}}(t)$ at each agent which is a sufficient information. Assuming we can interchange differentiation and expectation, the left hand side simplifies as follows

$$\begin{aligned}\text{LHS of (58)} &= \mathbb{E}^{g_{-i,t}}\Bigg[\frac{\partial}{\partial \hat{x}_i(t)}\Bigg[\sum_{k\in N}\sum_{j\in N}\hat{x}_k(t)^\intercal L^e_{kj}\hat{x}_j(t) \\ &\quad - 2\sum_{k\in N}\hat{x}_k(t)^\intercal M_{kk}x(t) + x(t)^\intercal M x(t)\Bigg] \,\Bigg|\, \hat{x}_i^{\text{loc}}(t)\Bigg] \\ &= 2\ \mathbb{E}^{g_{-i,t}}\Bigg[\sum_{j\in N} L^e_{ij}\hat{x}_j(t) - M_{ii}x(t) \,\Bigg|\, \hat{x}_i^{\text{loc}}(t)\Bigg].\end{aligned}$$

Thus, a sufficient condition for a strategy $g_{j,t}$ to be optimal is that

$$\sum_{j\in N} L^e_{ij}\mathbb{E}[g_{j,t}(\hat{x}_j^{\text{loc}}(t)) \mid \hat{x}_i^{\text{loc}}(t)] - M_{ii}\mathbb{E}[x(t) \mid \hat{x}_i^{\text{loc}}(t)] = 0. \tag{59}$$

The strategy (6) satisfies (59) if

$$\begin{aligned}&\sum_{j\in N} L^e_{ij}\big[F_j(t)\hat{P}_{ji}(t)\hat{P}_{ii}(t)^{-1}\hat{x}_i^{\text{loc}}(t)\big] \\ &\qquad - M_{ii}\big[\hat{P}_{ii}(t)\hat{P}_{ii}(t)^{-1}\hat{x}_i^{\text{loc}}(t)\big] = 0.\end{aligned}$$

Since the above equality must hold for all $\hat{x}_i^{\text{loc}}(t)$, the coefficients must be zero. Hence,

$$\sum_{j\in N} L^e_{ij}F_j(t)\hat{P}_{ji}(t) - M_{ii}\hat{P}_{ii}(t) = 0, \tag{60}$$

where we have cancelled out the positive definite matrix $\hat{P}_{ii}(t)^{-1}$. The above equation can be further simplified by vectorizing both sides and using the relation $\operatorname{vec}(ABC) = (C^\intercal \otimes A) \times \operatorname{vec}(B)$. Combining the equations for all $i \in N$, we get the linear equations described in the Theorem 1.

To compute the optimal performance, substituting (6) in (4), we get

$$\begin{aligned}J^*(t) &= \sum_{i\in N}\sum_{j\in N}\Big[\mathbb{E}[(\hat{x}_i^{\text{loc}}(t))^\intercal F_i(t)^\intercal L^e_{ij}F_j(t)\hat{x}_j^{\text{loc}}(t)]\Big] \\ &\quad - 2\sum_{i\in N}\Big[\mathbb{E}[(\hat{x}_i^{\text{loc}}(t))^\intercal F_i^\intercal(t)M_{ii}x(t)]\Big] \\ &\quad + \mathbb{E}[x^\intercal(t)Mx(t)] \\ &\overset{(a)}{=} \sum_{i\in N}\sum_{j\in N}\Big[\operatorname{Tr}(F_i(t)\hat{P}_{ij}(t)F_j(t)^\intercal L^e_{ji})\Big] \\ &\quad - 2\sum_{i\in N}\Big[\operatorname{Tr}(F_i(t)\hat{P}_{ii}(t)M_{ii}^\intercal)\Big] + \operatorname{Tr}(MP_x(t)) \\ &= \sum_{i\in N}\operatorname{Tr}\bigg(F_i(t)\Big(\sum_{j\in N}\hat{P}_{ij}(t)F_j(t)^\intercal L^e_{ji} - 2\hat{P}_{ii}(t)M_{ii}^\intercal\Big)\bigg) \\ &\quad + \operatorname{Tr}(MP_x(t))\end{aligned}$$

$$\overset{(b)}{=} -\sum_{i\in N} \mathrm{Tr}(F_i(t)\hat{P}_{ii}(t)M_{ii}^{\mathsf{T}}) + \mathrm{Tr}(MP_x(t))$$
$$\overset{(c)}{=} -\sum_{i\in N} \mathrm{vec}(F_i(t))^{\mathsf{T}}\,\mathrm{vec}(M_{ii}\hat{P}_{ii}(t)) + \mathrm{Tr}(MP_x(t))$$
$$= -F(t)^{\mathsf{T}}\eta(t) + \mathrm{Tr}(MP_x(t))$$
$$\overset{(d)}{=} -\eta(t)^{\mathsf{T}}\Gamma(t)^{-1}\eta(t) + \mathrm{Tr}(MP_x(t))$$

where $P_x(t)$ is defined in (9) and, the simplifications in (a) use the following: for any vectors x and y and matrices H, K, and R of appropriate dimensions, $\mathbb{E}[y^{\mathsf{T}}K^{\mathsf{T}}RHx] = \mathbb{E}[\mathrm{Tr}(y^{\mathsf{T}}K^{\mathsf{T}}RHx)] = \mathbb{E}[\mathrm{Tr}(Kyx^{\mathsf{T}}H^{\mathsf{T}}R^{\mathsf{T}})] = \mathrm{Tr}(K\mathbb{E}[yx^{\mathsf{T}}]H^{\mathsf{T}}R^{\mathsf{T}})$; those in (b) use (60); those in (c) use that for any matrices A and B of appropriate dimension $\mathrm{Tr}(AB^{\mathsf{T}}) = \mathrm{vec}(A)^{\mathsf{T}}\,\mathrm{vec}(B)$; and those in (d) use (6). Finally, summing the optimal cost from time 0 to $T-1$, gives (8).

REFERENCES

[1] J. Speyer, "Computation and transmission requirements for a decentralized linear-quadratic-Gaussian control problem," *IEEE Transactions on Automatic Control*, vol. 24, no. 2, pp. 266–269, April 1979.

[2] C.-Y. Chong, "Hierarchical estimation," in *Proc. MIT/ONR Workshop on C3*, 1979.

[3] M. Hassan, G. Salut, M. Singh, and A. Titli, "A decentralized computational algorithm for the global Kalman filter," *IEEE Transactions on Automatic Control*, vol. 23, no. 2, pp. 262–268, Apr 1978.

[4] A. Willsky, M. Bello, D. Castanon, B. Levy, and G. Verghese, "Combining and updating of local estimates and regional maps along sets of one-dimensional tracks," *IEEE Transactions on Automatic Control*, vol. 27, no. 4, pp. 799–813, Aug 1982.

[5] C. Sanders, E. Tacker, T. Linton, and R. Ling, "Specific structures for large-scale state estimation algorithms having information exchange," *IEEE Transactions on Automatic Control*, vol. 23, no. 2, pp. 255–261, Apr 1978.

[6] B. S. Rao and H. F. Durrant-Whyte, "Fully decentralised algorithm for multisensor kalman filtering," *IEE Proceedings D - Control Theory and Applications*, vol. 138, no. 5, pp. 413–420, Sept 1991.

[7] H. R. Hashemipour, S. Roy, and A. J. Laub, "Decentralized structures for parallel kalman filtering," *IEEE Transactions on Automatic Control*, vol. 33, no. 1, pp. 88–94, Jan 1988.

[8] D. Castanon and D. Teneketzis, "Distributed estimation algorithms for nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 418–425, May 1985.

[9] Z. Duda, "Fusion Kalman filtration for distributed multisensor systems," *Archives of Control Sciences*, vol. 24, no. 1, pp. 53–65, 2014.

[10] ——, "Fusion Kalman filtration with k-step delay sharing pattern," *Archives of Control Sciences*, vol. 25, no. 3, pp. 307–318, 2015.

[11] S. M. Barta, "On linear control of decentralized stochastic systems," Ph.D. dissertation, Massachusetts Institute of Technology, 1978.

[12] M. S. Andersland and D. Teneketzis, "Measurement scheduling for recursive team estimation," *Journal of Optimization Theory and Applications*, vol. 89, no. 3, pp. 615–636, Jun 1996.

[13] M. O. Sayin, S. S. Kozat, and T. Basar, "Team-optimal distributed estimation in general and tree networks," *Digital Signal Processing*, vol. 64, pp. 83 – 95, 2017.

[14] A. Gattami, "Distributed estimation and control under partially nested pattern," in *2009 American Control Conference*, June 2009, pp. 866–871.

[15] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2069–2084, Sept 2010.

[16] R. Olfati-Saber and P. Jalalkamali, "Coupled distributed estimation and control for mobile sensor networks," *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2609–2614, Oct 2012.

[17] G. Battistelli, L. Chisci, G. Mugnai, A. Farina, and A. Graziano, "Consensus-based linear and nonlinear filtering," *IEEE Transactions on Automatic Control*, vol. 60, no. 5, pp. 1410–1415, May 2015.

[18] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 90–104, Jan 2012.

[19] A. Mahajan, N. C. Martins, M. C. Rotkowitz, and S. Yksel, "Information structures in optimal decentralized control," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, Dec 2012, pp. 1291–1306.

[20] S. Yüksel and T. Başar, *Stochastic networked control systems: Stabilization and optimization under information constraints*. Springer Science & Business Media, 2013.

[21] H. S. Witsenhausen, "Separation of estimation and control for discrete time systems," *Proceedings of the IEEE*, vol. 59, no. 11, pp. 1557–1566, 1971.

[22] M. Afshari and A. Mahajan, "Team optimal decentralized Kalman filtering with coupled cost (extended version)," available at http://cim.mcgill.ca/~adityam/projects/info-structures/conference/2018-ecc-extended.pdf.

[23] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 857–881, 1962.

[24] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1644–1658, July 2013.

[25] H. S. Witsenhausen, "On information structures, feedback and causality," *SIAM Journal on Control*, vol. 9, no. 2, pp. 149–160, 1971.