Team Optimal Decentralized State Estimation

Mohammad Afshari and Aditya Mahajan

Abstract— **We consider the problem of optimal decentralized estimation of a linear stochastic process by multiple agents. Each agent receives a noisy observation of the state of the process and delayed observations of its neighbors (according to a pre-specified, strongly connected, communication graph). Based on their observations, all agents generate a sequence of estimates of the state of the process. The objective is to minimize the total expected weighted mean square error between the state and the agents' estimates over a finite horizon. In centralized estimation with weighted mean square error criteria, the optimal estimator does not depend on the weight matrix in the cost function. We show that this is not the case when the information is decentralized. The optimal decentralized estimates depend on the weight matrix in the cost function. In particular, we show that the optimal estimate consists of two parts: a common estimate which is the conditional mean of the state given the common information and a correction term which is a linear function of the offset of the local information from the conditional expectation of the local information given the common information. The corresponding gain depends on the weight matrix as well as on the covariance between the offset of agents' local information from the conditional expectation of the local information given the common information. We show that the common estimate can be computed from single Kalman filter and derive recursive expressions for computing the offset covariances and the estimation gains.**

I. INTRODUCTION

In his seminal counterexample [1], Witsenhausen showed that non-linear strategies may outperform the best linear (or affine) strategy in decentralized system with non-classical information structure, even if the dynamics are linear, the cost is quadratic, and the disturbances are Gaussian. Broadly speaking, three directions have been pursued in the subsequent literature: identifying conditions under which linear strategies are optimal; identifying conditions under which the domain of the control strategies may be restricted to a sufficient statistic or an information state; and identifying conditions under which a dynamic programming decomposition may be obtained. Due to lack of space, we provide a brief overview of only the first direction and refer the reader to [2], [3] for a detailed overview.

The simplest form of decentralized LQG (linear quadratic Gaussian) model is the static team problem in which all agents take a single action to minimize a common cost. Static team problems were first investigated in [4], [5], who identified necessary and sufficient conditions to determine team optimal strategies. In the special case when all primitive random variables are Gaussian and the cost is quadratic, it was shown that the optimal strategies are linear in the observation; the corresponding optimal gains are given by the solution of a system of simultaneous matrix equations.[1]

Based on the results for static teams, Witsenhausen [6] asserted that linear strategies are optimal for LQG dynamic teams with delayed sharing information structure. This assertion was shown to be true for one-step delayed sharing in [7], [8]. One-step delayed sharing is a special case of partially nested information structure. Ho and Chu [9] showed that linear strategies are optimal for general partially nested teams. Their proof was based on showing that a linear transformation reduces a partially nested team to a static team. The linear strategies obtained via such a transformation depend on the entire history of observations rather than a sufficient statistic. Subsequently, various models with partially nested information structure have been investigated where sufficient statistics are identified [10]–[12]. However, in many of these results, the equations for updating the sufficient statistics are coupled with those for computing the controller gain. This is in contrast to the celebrated two way separation in centralized LQG systems where the Kalman filtering equations are decoupled from the Riccati equations.

In view of this, the motivation of the current work is two-fold. Our first motivation is to understand the role of sufficient statistics in static team problems. To do so, we pose and solve a decentralized state estimation problem. Although the problem is a static team, directly using the results of [4], [5] gives estimation strategies that depend on the entire history of observations. We modify the proof of [4], [5] to obtain optimal strategies in terms of the common sufficient statistics and the innovation. We believe that the structure of optimal estimation strategies that is identified in this paper might be useful for general partially nested team problems as well.

Our second motivation is to study the decentralized state estimation problem in its own right. Decentralized estimation is a key component of many large scale systems including wireless sensor networks, power systems, target tracking, vehicle platooning, and networked control systems. Broadly speaking, the literature on decentralized state estimation can be classified into three categories. 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 and the objective is to compute the centralized estimate (see [13]–[18] and subsequent work). The second category consists of models where agents communicate a function of their observations to

The authors are with the Department of Electrical and Computer Engineering, McGill University, 3480 Rue University, Montreal, QC, H3A-0E9, Canada. Emails: mohammad.afshari2@mail.mcgill.ca, aditya.mahajan@mcgill.ca

[1] This system of matrix equations can be transformed into a system of simultaneous linear equations by vectorization.

other agents over a pre-specified communication graph and the objective is the asymptotic stability of the mean-squared error (see [19]–[21] and subsequent work). The third category consists of models where there is no fusion center and no inter-agent communication and the objective is to minimize a coupled cost function [22], [23]. The model that we consider is similar in spirit to [22], [23] but we allow inter-agent communication.

Our results highlight a feature of decentralized state estimation problem that makes it fundamentally different from centralized state estimation. To explain this difference, we consider the one-step version of both problems below.

A. A remark on centralized vs decentralized state estimation

First consider a centralized (one-step) state estimation problem. Let $x \in \mathbb{R}^{d_x}$, $x \sim \mathcal{N}(0, \Sigma_x)$, denote the state of a system. An agent observes $y \in \mathbb{R}^{d_y}$, where $y = Cx + v$, where C is a $d_y \times d_x$ matrix and $v \in \mathbb{R}^{d_y}$, $v \sim \mathcal{N}(0, R)$, is independent of x. The objective is to choose an estimate $\hat{z} \in \mathbb{R}^{d_z}$ of the state according to $\hat{z} = g(y)$ (where g can be any measurable function) to minimize

$$\mathbb{E}[(Lx - \hat{z})^\intercal S(Lx - \hat{z})],$$

where S is a $d_z \times d_z$ dimensional positive definite matrix and L is a $d_z \times d_x$ matrix. It is well known that the optimal estimate is given by L times the conditional mean $\hat{x}$ of the state given the observation, i.e.,

$$\hat{z} = L\hat{x}, \quad \text{where } \hat{x} := \mathbb{E}[x|y].$$

Alternatively, the optimal estimate may be written as a linear function of the observation y, i.e.,

$$\hat{z} = LKy, \quad \text{where } K = \Sigma_x C^\intercal (C\Sigma_x C^\intercal + R)^{-1}$$

It is worth highlighting the fact that *the optimal estimate does not depend on the weight matrix* S. It is perhaps for this reason that most standard texts on state estimation assume that the weight matrix $S = I$. However, when it comes to decentralized state estimation, the weight matrix S plays an important role.

To see this, consider a two-agent (one-step) decentralized state estimation problem. Let $x \in \mathbb{R}^{d_x}$, $x \sim \mathcal{N}(0, \Sigma_x)$, denote the state of a system. There are two agents indexed by $i \in \{1, 2\}$. Agent i, $i \in \{1, 2\}$, observes $y_i = C_i x + v_i$, $y \in \mathbb{R}^{d^i_y}$, where C_i is a $d^i_y \times d_x$ matrix and $v_i \in \mathbb{R}^{d^i_y}$, $v_i \sim \mathcal{N}(0, R_i)$. Assume that (x, v_1, v_2) are independent. The objective is for each agent to choose an estimate $\hat{z}_i \in \mathbb{R}^{d^i_z}$ according to $\hat{z}_i = g_i(y_i)$ (where g_i is a measurable function) to minimize

$$\mathbb{E}\left[\begin{bmatrix} L_1 x - \hat{z}_1 \\ L_2 x - \hat{z}_2 \end{bmatrix}^\intercal S \begin{bmatrix} L_1 x - \hat{z}_1 \\ L_2 x - \hat{z}_2 \end{bmatrix}\right],$$

where L_i and S are matrices of appropriate dimensions and S is positive definite.

Theorem 1 shows that the optimal estimates are given by

$$\hat{z}_i = F_i y_i,$$

where F_i is given by the solution of the following system of matrix equations:

$$\sum_{j \in \{1,2\}} \Big[S_{ij} F_j \Sigma_{ji} - S_{ij} L_j \Theta_i \Big] = 0, \quad \forall i \in \{1, 2\},$$

where $\Sigma_{ij} = \operatorname{cov}(y_i, y_j) = C_i \Sigma_x C_j^\intercal + \delta_{ij} R_i$, $\Theta_i = \operatorname{cov}(x, y_i) = \Sigma_x C_i^\intercal$, and δ_{ij} is the Dirac function.

In contrast to the centralized case, the gains F_i depend on the weight matrix S. Thus, in decentralized state estimation, the weight matrix S plays an important role, which makes decentralized state estimation fundamentally different from centralized state estimation.

B. Notations

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. Given a square matrix A, $\operatorname{Tr}(A)$ denotes the sum of its diagonal elements. Given a positive symmetric matrix A, the notation $A > 0$ and $A \geq 0$ mean that A is positive definite and semi-definite, respectively. Given a vector x, $\|x\|^2$ denotes $x^\intercal x$. $\mathbf{I}_n$ is the $n \times n$ identity matrix. We omit the subscript when the dimensions are clear from context.

Given any vector valued process $\{y(t)\}_{t \geq 1}$ and any time instances t_1, t_2 such that $t_1 \leq t_2$, $y(t_1{:}t_2)$ is a short hand notation for $\operatorname{vec}(y(t_1), y(t_1 + 1), \dots, y(t_2))$. Given matrices $\{A(i)\}_{i=1}^n$ with the same number of rows and vectors $\{w(i)\}_{i=1}^n$, $\operatorname{rows}(\bigodot_{i=1}^n A_i)$ and $\operatorname{vec}(\bigodot_{i=1}^n w(i))$ denote $\operatorname{rows}(A(1), \dots, A(n))$ and $\operatorname{vec}(w(1), \dots, w(n))$, respectively.

Given random vectors x and y, $\mathbb{E}[x]$ and $\operatorname{var}(x)$ denote the mean and variance of x while $\operatorname{cov}(x, y)$ denotes the covariance between x and y.

C. Preliminaries on graphs

A directed weighted graph $\mathcal{G}$ is an ordered set (N, E, d) where N is the set of nodes, $E \subset N \times N$ is the set of ordered edges, and $d \colon E \to \mathbb{R}^k$ is a weight function. 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 *directed path* $(v_1, v_2, \dots, v_k)$ is a sequence of distinct nodes such that $(v_i, v_{i+1}) \in E$. 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 directed graph is called *strongly connected* if for every pair of nodes $i, j \in N$, there is a directed path from i to j and from j to i. A directed graph is called *complete* if for every pair of nodes $i, j \in N$, there is a directed edge from i to j and from j to i.

II. Problem Formulation and Main Results

A. Observation Model

Consider a linear stochastic process $\{x(t)\}_{t \geq 1}$, $x(t) \in \mathbb{R}^{d_x}$, where $x(1) \sim \mathcal{N}(0, \Sigma_x)$ and for $t \geq 1$,

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

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

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

where C_i is a $d_y^i \times d_x$ matrix and $v_i(t) \in \mathbb{R}^{d_y^i}$, $v_i(t) \sim \mathcal{N}(0, R_i)$, is the observation noise. Eq. (2) may be written in vector form as

$$y(t) = Cx(t) + v(t),$$

where $C = \text{rows}(C_1, \dots, C_n)$, $y(t) = \text{vec}(y_1(t), \dots, y_n(t))$, and $v(t) = \text{vec}(v_1(t), \dots, v_n(t))$.

The agents are connected over a ***communication graph*** $\mathcal{G}$, which is a *strongly connected* weighted directed graph with vertex set N. 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.,

$$I_i(t) = \{y_i(1{:}t)\} \cup \Big(\bigcup_{j \in N_i^-} \{I_j(t - d_{ji})\} \Big). \tag{3}$$

In (3), we implicitly assume that $I_i(t) = \emptyset$ for any $t \leq 0$.

Let $\zeta_i(t) = I_i(t) \setminus I_i(t-1)$ denote the new information that becomes available to agent i at time t. Then, $\zeta_i(1) = y_i(1)$ and for $t > 1$,

$$I_i(t) = \text{vec}(y_i(t), \{\zeta_j(t - d_{ji})\}_{j \in N_i^-}).$$

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

Some examples of the communication graph are as follows.

Example 1 Consider a complete graph with d-step delay along each edge. The resulting information structure is

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

which is the *d-step delayed sharing information strucutre* [6]. □

Example 2 Consider a strongly connected graph with unit delay along each edge. Recall that ℓ_{ij} denotes the geodesic distance between nodes i and j. Let $d^* = \max_{i,j \in N} \ell_{ij}$, denote the diameter of the graph and $N_i^k = \{j \in N : \ell_{ji} = k\}$, denote the k-hop in-neighbors of i with $N_i^0 = \{i\}$. The resulting information structure is

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

which we call the *neighborhood sharing information structure*. □

At time t, agent $i \in N$ generates an estimate $\hat{z}_i(t) \in \mathbb{R}^{d_z^i}$ of $L_i x(t)$ (where L_i is a $\mathbb{R}^{d_z^i \times d_x}$ matrix) according to

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

where $g_{i,t}$ is a measurable function called the *estimation rule* at time t. The collection $g_i \coloneqq (g_{i,1}, g_{i,2}, \dots)$ is called the *estimation strategy* of agent i and $g \coloneqq (g_1, \dots, g_n)$ is the *team estimation strategy profile* of all agents.

B. Estimation Cost

Let $\hat{z}(t) = \text{vec}(\hat{z}_1(t), \dots, \hat{z}_n(t))$ denote the estimate of all agents. Then the estimation error $c(x(t), \hat{z}(t))$ is a weighted quadratic function of $(Lx(t) - \hat{z}(t))$. In particular,

$$c(x(t), \hat{z}(t)) = (Lx(t) - \hat{z}(t))^\intercal S (Lx(t) - \hat{z}(t)), \tag{4}$$

where S and L are defined as follows:

$$S = \begin{bmatrix} S_{11} & \cdots & S_{1n} \\ \vdots & \ddots & \vdots \\ S_{n1} & \cdots & S_{nn} \end{bmatrix} \quad \text{and} \quad L = \begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix}. \tag{5}$$

As an example of the cost function of the form (4), consider the following scenario. Suppose $x(t) = \text{vec}(x_1(t), \dots, x_n(t))$, where we may think of $x_i(t)$ as the local state of agent $i \in N$. Suppose the agents want to estimate their own local state, but at the same time, want to make sure that the average $\bar{z}(t) \coloneqq \frac{1}{n} \sum_{i \in N} \hat{z}_i(t)$ of their estimates is close to the average $\bar{x}(t) \coloneqq \frac{1}{n} \sum_{i \in N} x_i(t)$ of their local states. In this case, the cost function is

$$c(x(t), \hat{z}(t)) = \sum_{i \in N} \|x_i(t) - \hat{z}_i(t)\|^2 + \lambda \|\bar{x}(t) - \bar{z}(t)\|^2, \tag{6}$$

where $\lambda \in \mathbb{R}_{>0}$. This can be written in the form (4) with $L = \mathbf{I}$, and

$$S = \begin{bmatrix} (1 + \frac{\lambda}{n^2})\mathbf{I} & \frac{\lambda}{n^2}\mathbf{I} & \cdots & \frac{\lambda}{n^2}\mathbf{I} \\ \frac{\lambda}{n^2}\mathbf{I} & (1 + \frac{\lambda}{n^2})\mathbf{I} & \cdots & \frac{\lambda}{n^2}\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\lambda}{n^2}\mathbf{I} & \frac{\lambda}{n^2}\mathbf{I} & \cdots & (1 + \frac{\lambda}{n^2})\mathbf{I} \end{bmatrix},$$

As an other example, suppose the agents are moving in a line (e.g.. a vehicular platoon) and want to estimate their local state but, at the same time, want to ensure that the difference $\hat{d}_i(t) \coloneqq \hat{z}_i(t) - \hat{z}_{i+1}(t)$ between their estimates is close to the difference $d_i(t) \coloneqq x_i(t) - x_{i+1}(t)$ of their local states.

In this case, the cost function is

$$c(x(t), \hat{z}(t)) = \sum_{i \in N} \|x_i(t) - \hat{z}_i(t)\|^2 + \lambda \sum_{i=1}^{n-1} \|d_i(t) - \hat{d}_i(t)\|^2, \tag{7}$$

where $\lambda \in \mathbb{R}_{>0}$. This can be written in the form (4) with $L = \mathbf{I}$ and

$$S = \begin{bmatrix} (1+\lambda)\mathbf{I} & -\lambda\mathbf{I} & 0 & \cdots & 0 \\ -\lambda\mathbf{I} & (1+2\lambda)\mathbf{I} & -\lambda\mathbf{I} & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \cdots \\ \vdots & \vdots & -\lambda\mathbf{I} & (1+2\lambda)\mathbf{I} & -\lambda\mathbf{I} \\ 0 & \cdots & \cdots & -\lambda\mathbf{I} & (1+\lambda)\mathbf{I} \end{bmatrix},$$

C. Problem Formulation

We consider the following assumptions on the model.

(A1) The cost matrix S is positive definite.

(A2) The noise covariance matrices $\{R_i\}_{i\in N}$ are positive definite and Q and Σ_x are positive semi-definite.

(A3) The primitive random variables $(x(1), \{w(t)\}_{t\geq 1}, \{v_1(t)\}_{t\geq 1}, \dots, \{v_n(t)\}_{t\geq 1})$ are independent.

We are interested in the following optimization problem.

Problem 1 *Given matrices A, $\{C_i\}_{i\in N}$, Σ_x, Q, $\{R_i\}_{i\in N}$, L, S, a* communication graph $\mathcal{G}$ *(and the corresponding weights d_{ij}), and a horizon T, choose a team estimation strategy profile g to minimize $J_T(g)$ given by*

$$J_T(g) = \mathbb{E}^g\Bigg[\sum_{t=1}^{T} c(x(t), \hat{z}(t))\Bigg]. \tag{8}$$

III. Main Results

A. Preliminaries on centralized Kalman filtering

Consider a centralized agent that observes $y(1{:}t)$ and wants to generate an estimate $\hat{z}^*(t)$ to minimize

$$\mathbb{E}[(Lx(t) - \hat{z}^*(t))^\intercal(Lx(t) - \hat{z}^*(t))].$$

It is well known from Kalman filtering theory [24] that

$$\hat{z}^*(t) = L\hat{x}^*(t)$$

where $\hat{x}^*(t) = \mathbb{E}[x(t)|y(1{:}t)]$. We have that $\hat{x}^*(0) = 0$ and for $t \geq 0$, $\hat{x}^*(t)$ can be recursively updated as

$$\hat{x}^*(t+1) = A\hat{x}^*(t) + K^*(t)[y(t+1) - CA\hat{x}^*(t)], \tag{9}$$

where

$$K^*(t) = [AP^*(t)A^\intercal C^\intercal + QC^\intercal] \\ [CAP^*(t)A^\intercal C^\intercal + CQC^\intercal + R]^{-1}, \tag{10}$$

$R = \operatorname{diag}(R_1, \dots, R_n)$, and $P^*(t) = \operatorname{var}(x(t) - \hat{x}^*(t))$ is the covariance of the error $\tilde{x}^*(t) \coloneqq x(t) - \hat{x}^*(t)$. $P^*(t)$ can be pre-computed recursively using the forward Riccati equation: $P^*(0) = 0$ and for $t \geq 0$,

$$P^*(t+1) = AP^*(t)A^\intercal - K^*(t)[AP^*(t)A^\intercal C^\intercal + QC^\intercal] + Q. \tag{11}$$

In a second scenario, consider the centralized agent that observes $y(1{:}t-1)$ to generate an estimate $\hat{z}^{\text{cen}}(t)$ to minimize

$$\mathbb{E}[(Lx(t) - \hat{z}^{\text{cen}}(t))^\intercal(Lx(t) - \hat{z}^{\text{cen}}(t))].$$

Again from [24],

$$\hat{z}^{\text{cen}}(t) = L\hat{x}(t),$$

where $\hat{x}(t) = \mathbb{E}[x(t)|y(1{:}t-1)]$ is the delayed centralized estimate of the state. We have that $\hat{x}(1) = 0$ and for $t \geq 1$,

$$\hat{x}(t+1) = A\hat{x}(t) + AK(t)[y(t) - C\hat{x}(t)], \tag{12}$$

where

$$K(t) = P(t)C^\intercal[CP(t)C^\intercal + R]^{-1}, \tag{13}$$

and $P(t) = \operatorname{var}(x(t) - \hat{x}(t))$ is the covariance of the error $\tilde{x}(t) \coloneqq x(t) - \hat{x}(t)$. $P(t)$ can be pre-computed recursively using the forward Riccati equation: $P(1) = \Sigma_x$ and for $t \geq 1$,

$$P(t+1) = A\Delta(t)P(t)\Delta(t)^\intercal A^\intercal + AK(t)RK(t)^\intercal A^\intercal + Q, \tag{14}$$

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

B. Structure of optimal estimation strategy

Following [25], we define

$$I^{\text{com}}(t) = \bigcap_{i\in N} I_i(t)$$

as the *common information* among all agents[2]. Since the information is shared over a strongly connected graph, the common information is

$$I^{\text{com}}(t) = y(1{:}t - d^*),$$

where d^* is the diameter of the graph.

We define the local information at agent i as

$$I_i^{\text{loc}}(t) = I_i(t) \setminus I^{\text{com}}(t).$$

Then, $I_i(t) = I^{\text{com}}(t) \cup I_i^{\text{loc}}(t)$.

Furthermore, we define

$$\hat{x}^{\text{com}}(t) = \mathbb{E}[x(t)|I^{\text{com}}(t)]$$

as the *common estimate* of the state and

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

as the common estimate of local information of agent i. Here we assume that $I_i^{\text{loc}}(t)$ (ande hence $\hat{I}_i^{\text{loc}}(t)$) is a vector. Following [26], we define

$$\tilde{I}_i^{\text{loc}}(t) = I_i^{\text{loc}}(t) - \mathbb{E}[I_i^{\text{loc}}(t)|I^{\text{com}}(t)] \tag{15}$$

as the innovation in the local information at agent i.

To find a convenient expression for the *innovation* $\tilde{I}_i^{\text{loc}}(t)$, we follow [6] and express $I_i^{\text{loc}}(t)$ in terms of the *delayed state* $x(t - d^* + 1)$. For that matter, for any $t, \ell \in \mathbb{Z}_{>0}$, define the $d_x \times 1$ random vector $w^{(k)}(\ell, t)$ as follows:

$$w^{(k)}(\ell, t) = \sum_{\tau=\max\{1,t-k\}}^{t-\ell-1} A^{t-\ell-\tau-1}w(\tau). \tag{16}$$

Note that $w^{(k)}(\ell, t) = 0$ if $t \leq \min\{k, \ell+1\}$ or $\ell \geq k$. For any $t \geq k$, we may write

$$x(t) = A^k x(t-k) + w^{(k)}(0, t), \tag{17}$$

$$y_i(t) = C_i A^k x(t-k) + C_i w^{(k)}(0, t) + v_i(t). \tag{18}$$

[2]Our mothodology relies on the split of the total information into common and local information as proposed in [25]. However, the specific details on how the common information is used is different from [25].

By definition $I_i^{\text{loc}}(t) \subseteq y(t-d^*+1{:}t)$. Thus, for any $i \in N$, we can identify matrix C_i^{loc} and random vectors $w_i^{\text{loc}}(t)$ and $v_i^{\text{loc}}(t)$ (which are linear functions of $w(t-d^*+1{:}t-1)$ and $v_i(t-d^*+1{:}t)$) such that

$$I_i^{\text{loc}}(t) = C_i^{\text{loc}} x(t-d^*+1) + w_i^{\text{loc}}(t) + v_i^{\text{loc}}(t). \tag{19}$$

To write the expressions for $(C_i^{\text{loc}}, w_i^{\text{loc}}(t), v_i^{\text{loc}}(t))$ for the delayed sharing and neighborhood sharing information structures below, we define for any $\ell \le d^*$,

$$\mathcal{W}_i(\ell,t) = \begin{bmatrix} C_i w^{(d^*-1)}(d^*-1,t) \\ C_i w^{(d^*-1)}(d^*-2,t) \\ \vdots \\ C_i w^{(d^*-1)}(\ell,t) \end{bmatrix},$$

$$\mathcal{C}_i(\ell) = \begin{bmatrix} C_i \\ C_i A \\ \vdots \\ C_i A^{d^*-\ell-1} \end{bmatrix}, \quad \mathcal{V}_i(\ell,t) = \begin{bmatrix} v_i(t-d^*+1) \\ v_i(t-d^*+2) \\ \vdots \\ v_i(t-\ell) \end{bmatrix}.$$

□

Example 1 (cont.) For the d-step delayed sharing information structure $I_i^{\text{loc}}(t) = y_i(t-d^*+1{:}t)$. Thus,

$$C_i^{\text{loc}} = \mathcal{C}_i(0), \quad w_i^{\text{loc}}(t) = \mathcal{W}_i(0,t), \quad v_i^{\text{loc}}(t) = \mathcal{V}_i(0,t).$$

Example 2 (cont.) For the neighborhood sharing information structure, $I_i(t) = \bigcup_{k=0}^{d^*} \bigcup_{j \in N_i^k} \{y_j(1{:}t-k)\}$. Thus,

$$C_i^{\text{loc}} = \text{rows}\left(\bigodot_{\ell=0}^{d^*-1} \bigodot_{j \in N_i^\ell} \mathcal{C}_j(\ell) \right),$$

$$w_i^{\text{loc}}(t) = \text{vec}\left(\bigodot_{\ell=0}^{d^*-1} \bigodot_{j \in N_i^\ell} \mathcal{W}_j(\ell,t) \right),$$

$$v_i^{\text{loc}}(t) = \text{vec}\left(\bigodot_{\ell=0}^{d^*-1} \bigodot_{j \in N_i^\ell} \mathcal{V}_j(\ell,t) \right).$$

□

Now define,

$$\begin{aligned} \hat{x}(t-d^*+1) &= \mathbb{E}[x(t-d^*+1) \mid I^{\text{com}}(t)] \\ &= \mathbb{E}[x(t-d^*+1) \mid y(1{:}t-d^*)] \end{aligned} \tag{20}$$

as the *delayed centralized* estimate of the state and

$$\tilde{x}(t-d^*+1) = x(t-d^*+1) - \hat{x}(t-d^*+1)$$

as the error of the delayed centralized estimate. Note that this notation is consistent with the notation for the delayed centralized Kalman filtering used in Section III-A. Thus, $\hat{x}(t-d^*+1)$ can be updated recursively using (12).

Lemma 1 *$w_i^{\text{loc}}(t)$, $v_i^{\text{loc}}(t)$, $\tilde{x}(t-d^*+1)$, and $I^{\text{com}}(t)$ are independent.* □

Proof: The proof is omitted due to space limitations. ■

From Lemma 1 and from (19), we get

$$\hat{I}_i^{\text{loc}}(t) = C_i^{\text{loc}}(t) \hat{x}(t-d^*+1), \tag{21}$$

Our main result is as follows:

Theorem 1 *Under* (A1)–(A3)*, we have the following:*

1) *Optimal decentralized estimates are*

$$\hat{z}_i(t) = L_i \hat{x}^{\text{com}}(t) + F_i(t) \tilde{I}_i^{\text{loc}}(t), \tag{22}$$

where

$$\hat{x}^{\text{com}}(t) = A^{d^*-1} \hat{x}(t-d^*+1), \tag{23}$$

$\hat{x}(t-d^*+1)$ is computed according to the delayed centralized Kalman filter (12)–(14)*, and*

$$\begin{aligned} \tilde{I}_i^{\text{loc}}(t) &= I_i^{\text{loc}}(t) - C_i^{\text{loc}} \hat{x}(t-d^*+1) \\ &= C_i^{\text{loc}} \tilde{x}(t-d^*+1) + w_i^{\text{loc}}(t) + v_i^{\text{loc}}(t), \end{aligned} \tag{24}$$

2) *The optimal gains $\{F_i(t)\}_{i \in N}$ are given by the (unique) solution of the following system of matrix equations.*

$$\sum_{j \in N} \left[S_{ij} F_j(t) \hat{\Sigma}_{ji}(t) - S_{ij} L_j \hat{\Theta}_i(t) \right] = 0, \quad \forall i \in N, \tag{25}$$

where $\hat{\Sigma}_{ij}(t) = \text{cov}\left(\tilde{I}_i(t), \tilde{I}_j(t)\right)$ and is given by

$$\hat{\Sigma}_{ij}(t) = C_i^{\text{loc}} P(t-d^*+1) C_j^{\text{loc}\,\intercal} + P_{ij}^w(t) + P_{ij}^v(t), \tag{26}$$

where $P_{ij}^w(t) = \text{cov}(w_i^{\text{loc}}(t), w_j^{\text{loc}}(t))$, $P_{ij}^v(t) = \text{cov}(v_i^{\text{loc}}(t), v_j^{\text{loc}}(t))$ and $\hat{\Theta}_i(t) = \text{cov}(x, \tilde{I}_i(t))$ and is given by

$$\hat{\Theta}_i(t) = [A^{d^*-1} P(t-d^*+1) C_i^{\text{loc}\,\intercal} + P_i^\sigma(t)] \tag{27}$$

where $P_i^\sigma(t) = \text{cov}(w^{(d^*-1)}(0,t), w_i^{\text{loc}}(t))$.

3) *Finally, the optimal performance is given by*

$$J_T^* = \sum_{t=1}^{T} \Bigg[\text{Tr}(L^\intercal S L P_0(t)) - \sum_{i \in N} \text{Tr}\left(F_i^\intercal \sum_{j \in N} S_{ij} L_j \hat{\Theta}_i(t) \right) \Bigg]. \tag{28}$$

where $P_0(t) = \text{var}(x(t) - \hat{x}^{\text{com}}(t))$ and is given by

$$P_0(t) = A^{d^*-1} P(t-d+1) (A^{d^*-1})^\intercal + \Sigma^w(t), \tag{29}$$

and $\Sigma^w(t) = \text{var}(w^{(d^*-1)}(0,t))$. □

Proof: Since the choice of the estimates does not affect the evolution of the system, choosing an estimation profile $g = (g_1, \dots, g_n)$ to minimize $J_T(g)$ is equivalent to solving the following T separate optimization problems.

$$\min_{(g_{1,t}, \dots, g_{n,t})} \mathbb{E}[c(x(t), \hat{z}(t))], \quad \forall t \in \{1, \dots, t\}. \tag{30}$$

From [27, Theorem 1], we get that the strategy given by (22) is optimal for (30). We defer the proof of existence and uniqueness of the solution of (25) to Theorem 2.

The expression (23) for $\hat{x}^{\text{com}}(t)$ follows from (17). The expression (24) for $\tilde{I}_i^{\text{loc}}(t)$ follows from (19) and (21). Substituting (19) in (24), we get

$$\tilde{I}_i^{\text{loc}}(t) = C_i^{\text{loc}} \tilde{x}(t-d^*+1) + w_i^{\text{loc}}(t) + v_i^{\text{loc}}(t). \tag{31}$$

Thus, we get the expression (26) for $\hat{\Sigma}_{ij}(t)$ from Lemma 1.

From (17) and (31), and Lemma 1, we get the expression (27) for $\hat{\Theta}_i(t)$. Finally the expression for $P_0(t)$ follows from (17) and (23) and the performance of the strategy is given by (28). ∎

Theorem 2 *Equation* (25) *has a unique solution and can be written more compactly as*

$$F(t) = \Gamma(t)^{-1}\eta(t), \tag{32}$$

where

$$\begin{aligned} F(t) &= \operatorname{vec}(F_1(t), \dots, F_n(t)), \\ \eta(t) &= \operatorname{vec}(S_1 L\hat{\Theta}_1(t), \dots, S_n L\hat{\Theta}_n(t)), \\ S_i &= [S_{i1}, \dots, S_{in}], \\ \Gamma(t) &= [\Gamma_{ij}(t)]_{i,j\in N}, \quad \textit{where } \Gamma_{ij}(t) = \hat{\Sigma}_{ij}(t) \otimes S_{ij}. \end{aligned}$$

Furthermore, the optimal performance can be written as

$$J_T^* = \sum_{t=1}^{T} \big[\operatorname{Tr}(L^\intercal S L P_0(t)) - \eta(t)^\intercal \Gamma(t)^{-1}\eta(t) \big]. \tag{33}$$

□

Proof: First, we start by observing that $\hat{\Sigma}_{ii} > 0$. This follows from the fact that $\hat{\Sigma}_{ii}$ is the variance of the innovation in the standard Kalman filtering equation. Thus, the positive definiteness of R_i in assumption (A2) ensures that $\hat{\Sigma}_{ii}$ is positive definite [24, Section 3.4]. The result then follows from [7, Lemma 1] ∎

Remark 1 Let us contrast the results of Theorem 2 with the Kalman estimator

$$Z_i^{\text{Kal}}(t) = L_i \hat{x}_i(t)$$

where $\hat{x}_i(t) = \mathbb{E}[x(t)|I_i(t)]$. We can view $\hat{x}_i(t)$ as a Kalman filter update when the agent has information $I^{\text{com}}(t)$ and receives new information $I_i^{\text{loc}}(t)$. Hence, from Kalman filtering update, we have

$$\hat{x}_i(t) = \hat{x}^{\text{com}}(t) + K_i(t)\tilde{I}_i^{\text{loc}}(t),$$

where $K_i(t) = \hat{\Theta}_i(t)\hat{\Sigma}_{ii}^{-1}$. □

Therefore, we have

$$Z_i^{\text{Kal}}(t) = L_i\hat{x}^{\text{com}}(t) + L_i K_i(t)\tilde{I}_i^{\text{loc}}(t). \tag{34}$$

Thus, the structure of the optimal estimator in Theorem 1 is the same as the Kalman estimator (34). The difference is that, in the optimal estimator, the gains $\{F_i(t)\}_{i\in N}$ are obtained by solving a system of simultaneous matrix equations that depend on the weight matrix S where in the Kalman estimator the gains $\{L_i K_i(t)\}_{i\in N}$ do not depend on the cost matrix S.

When S is block diagonal, the optimal estimator is same as the Kalman estimator, as shown in the following.

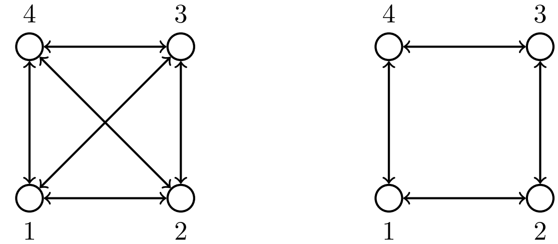
Corollary 1 *If* $S_{ij} = 0$ *for all* $i, j \in N$, $i \neq j$, *then*

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

□

Proof: For a block diagonal S, Eq. (25) reduces to

$$S_{ii}F_i(t)\hat{\Sigma}_{ii}(t) = S_{ii}L_i\hat{\Theta}_i(t). \tag{35}$$



(a) A system with 2-step delay sharing information structure.

(b) A system with neighborhood sharing information structure.

Fig. 1: Two systems with different information structures.

Note that when S is block diagonal, (A3) implies that each S_{ii} is postive-definite, and hence invertible. Moreover, $\hat{\Sigma}_{ii}(t)$ is positive definite and invertible [24, Section 3.4]. Thus, Eq. (35) simplifies to $F_i(t) = L_i\hat{\Theta}_i(t)\hat{\Sigma}_{ii}^{-1}(t)$. Substituting this in (22) gives

$$\begin{aligned} \hat{z}_i(t) &= L_i\hat{x}^{\text{com}}(t) + L_i\hat{\Theta}_i(t)\hat{\Sigma}_{ii}^{-1}(t)\tilde{I}_i^{\text{loc}}(t) \\ &\overset{(a)}{=} L_i\big[\mathbb{E}[x(t)|I^{\text{com}}(t)] + \mathbb{E}[x(t)|\tilde{I}_i^{\text{loc}}(t)]\big] \\ &\overset{(b)}{=} L_i\big[\mathbb{E}[x(t)|I_i(t)], \end{aligned}$$

where (a) uses $\hat{x}^{\text{com}}(t) = \mathbb{E}[x(t)|I^{\text{com}}(t)]$ and the following equation for Gaussian zero mean random variables a and b:

$$\mathbb{E}[a|b] = \operatorname{cov}(a,b)\operatorname{var}(b)b,$$

and (b) uses the orthogonal projection because $\tilde{I}_i^{\text{loc}}(t)$ is orthogonal to $I^{\text{com}}(t)$ [26]. ∎

IV. Some examples

To illustrate our main results, we consider two examples of four node networks with different information structures shown in Fig. 1 and show the computations $C_i^{\text{loc}}, w_i^{\text{loc}}(t), v_i^{\text{loc}}(t)$, $P_i^{\sigma}(t)$, $P_{ij}^{w}(t)$, and $P_{ij}^{v}(t)$.

A. Four node network with 2*-step delayed sharing*

Consider the complete graph with 2-step delay information structure shown in Fig. 1(a). The information structure is given by

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

Thus, $I^{\text{com}}(t) = y(1{:}t-2)$, $I_i^{\text{loc}}(t) = y_i(t-1{:}t)$, $C_i^{\text{loc}} = \operatorname{rows}(C_i, C_iA)$, $w_i^{\text{loc}}(t) = \operatorname{vec}(0, C_i w(t-1))$, and $v_i^{\text{loc}}(t) = \operatorname{vec}(v_i(t-1), v_i(t))$. Using these, we get that

- For $t = 1$,

$$\begin{aligned} \Sigma^w(1) &= 0, & P_i^{\sigma}(t) &= \begin{bmatrix} 0 & 0 \end{bmatrix}, \\ P_{ii}^{w}(1) &= \operatorname{diag}(0,0), & P_{ij}^{w}(1) &= \operatorname{diag}(0,0), \\ P_{ii}^{v}(1) &= \operatorname{diag}(0,R_i), & P_{ij}^{v}(1) &= \operatorname{diag}(0,0). \end{aligned}$$

- For $t \geq 2$,

$$\begin{aligned} \Sigma^w(t) &= Q, & P_i^{\sigma}(t) &= \begin{bmatrix} 0 & QC_i^\intercal \end{bmatrix}, \\ P_{ii}^{w}(t) &= \operatorname{diag}(0, C_iQC_i^\intercal), & P_{ij}^{w}(t) &= \operatorname{diag}(0, C_iQC_j^\intercal), \\ P_{ii}^{v}(t) &= \operatorname{diag}(R_i,R_i), & P_{ij}^{v}(t) &= \operatorname{diag}(0,0). \end{aligned}$$

Substituting these, we get that $\hat{\Sigma}_{ij}(1) = \delta_{ij}\text{diag}(0, R_i)$, and for $t \geq 2$,

$$\hat{\Sigma}_{ij}(t) = \begin{bmatrix} C_i \\ C_i A \end{bmatrix} P(t-1) \begin{bmatrix} C_j \\ C_j A \end{bmatrix}^\intercal + \begin{bmatrix} \delta_{ij} R_i & 0 \\ 0 & C_i Q C_i^\intercal + \delta_{ij} R_i \end{bmatrix}.$$

Finally, substituting $\hat{\Sigma}_{ij}(t)$ in (25) or (32) gives us the optimal gains.

B. Four node network with neighborhood sharing

Consider the graph with neighborhood sharing information structure shown in Fig 1(b). The information structure is given by

$$I_i(t) = \{y(1{:}t-2), y_{i-1}(t-1), y_i(t-1{:}t), y_{i+1}(t-1)\},$$

where we have assumed that the subscripts $i+1$ and $i-1$ are evaluated modulo 4 over the residue system $\{1, 2, 3, 4\}$. Thus, $I^{\text{com}}(t) = y(1{:}t-2)$ and

$$I_i^{\text{loc}}(t) = \{y_{i-1}(t-1), y_i(t-1{:}t), y_{i+1}(t-1)\}.$$

Thus, $C_i^{\text{loc}} = \text{rows}(C_{i-1}, C_i, C_i A, C_{i+1})$,

$$w_i^{\text{loc}}(t) = \begin{bmatrix} 0 \\ 0 \\ C_i w(t-1) \\ 0 \end{bmatrix}, \quad v_i^{\text{loc}}(t) = \begin{bmatrix} v_{i-1}(t-1) \\ v_i(t-1) \\ v_i(t) \\ v_{i+1}(t-1) \end{bmatrix}.$$

Using these, we get that

- For $t = 1$,

$$\begin{aligned} \Sigma^w(1) &= 0, & P_i^\sigma(t) &= \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}, \\ P_{ii}^w(1) &= \text{diag}(0,0,0,0), & P_{ij}^w(1) &= \text{diag}(0,0,0,0), \\ P_{ii}^v(1) &= \text{diag}(0,0,R_i,0), & P_{ij}^v(1) &= \text{diag}(0,0,0,0). \end{aligned}$$

- for $t \geq 2$,

$$\begin{aligned} \Sigma^w(t) &= Q, \\ P_{ij}^w(t) &= \text{diag}(0, 0, C_i Q C_j^\intercal, 0), \\ P_i^\sigma(t) &= [0, 0, QC_i^\intercal, 0], \\ P_{ii}^v(t) &= \text{diag}(R_{i-1}, R_i, R_i, R_{i+1}) \end{aligned}$$

$$P_{i,i+1}^v(t) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ R_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & R_{i+1} & 0 & 0 \end{bmatrix}, P_{i,i+2}^v(t) = \begin{bmatrix} 0 & 0 & 0 & R_{i-1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ R_{i+1} & 0 & 0 & 0 \end{bmatrix},$$

$P_{i+1,i}^v(t) = P_{i,i+1}^v(t)^\intercal$, and $P_{i+2,i}^v(t) = P_{i,i+2}^v(t)^\intercal$.

V. Conclusions

We investigated the problem of decentralized state estimation and identified the structure of optimal estimation strategies. The optimal estimates are linear function of the common estimate and the offset of the local observation from its conditional expectation given the common information. The gain of the offset term depends on the weight matrix of the cost function. This feature makes the decentralized state estimation problem fundamentally different from centralized state estimation. We restricted attention to finite horizon models. It can be shown that the results generalize to infinite horizon long run average setup under standard conditions on stabilizability and observability.

We believe that decentralized estimation plays a fundamental role in decentralized control problems and plan to investigate this relationship in the future.

References

[1] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, vol. 6, no. 1, pp. 131–147, 1968.

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

[3] A. Mahajan and M. Mannan, "Decentralized stochastic control," *Annals of Operations Research*, vol. 241, no. 1, pp. 109–126, Jun 2016.

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

[5] J. Marschak and R. R, *Economic theory of teams*. Yale University Press, New Haven, 1972.

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

[7] N. Sandell and M. Athans, "Solution of some nonclassical LQG stochastic decision problems," *IEEE Trans. Autom. Control*, vol. 19, no. 2, pp. 108–116, Apr 1974.

[8] T. Yoshikawa, "Dynamic programming approach to decentralized stochastic control problems," *IEEE Trans. Autom. Control*, vol. 20, no. 6, pp. 796–797, Dec 1975.

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

[10] H. R. Feyzmahdavian, A. Alam, and A. Gattami, "Optimal distributed controller design with communication delays: Application to vehicle formations," in *IEEE Conf. on Decision and Control (CDC)*, 2012, pp. 2232–2237.

[11] L. Lessard and A. Nayyar, "Structural results and explicit solution for two-player LQG systems on a finite time horizon," in *IEEE Conf. on Decision and Control (CDC)*, 2013, pp. 6542–6549.

[12] N. Nayyar, D. Kalathil, and R. Jain, "Optimal decentralized control with asymmetric one-step delayed information sharing," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 1, pp. 653–663, March 2018.

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

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

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

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

[17] 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 Trans. Autom. Control*, vol. 27, no. 4, pp. 799–813, Aug 1982.

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

[19] R. Olfati-Saber, "Distributed kalman filter with embedded consensus filters," in *Proceedings of the 44th IEEE Conf. on Decision and Control*, Dec 2005, pp. 8179–8184.

[20] ——, "Distributed Kalman filtering for sensor networks," in *IEEE Conf. on Decision and Control (CDC)*, Dec 2007, pp. 5492–5498.

[21] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in communications*, vol. 26, no. 4, 2008.

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

[23] 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.

[24] P. E. Caines, *Linear Stochastic Systems*. New York, NY, USA: John Wiley & Sons, Inc., 1987.

[25] 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, July 2013.

[26] T. Kailath, "An innovations approach to least-squares estimation–part I: Linear filtering in additive white noise," *IEEE Trans. Autom. Control*, vol. 13, no. 6, pp. 646–655, December 1968.

[27] M. Afshari and A. Mahajan, "Static teams with common information," *20th IFAC World Congress*, vol. 50, no. 1, pp. 11 926 – 11 931, 2017.