Split-Spectrum Based Distributed State Estimation for Linear Systems

Lili Wang\textsuperscript{a}, Ji Liu\textsuperscript{b}, Brian D. O. Anderson\textsuperscript{d} A. Stephen Morse\textsuperscript{c},

\textsuperscript{a}Samueli School of Engineering, University of California, Irvine
\textsuperscript{b}Department of Electrical and Computer Engineering, Stony Brook University
\textsuperscript{c}Department of Electrical Engineering, Yale University
\textsuperscript{d}School of Engineering, Australian National University

Abstract

This paper studies a distributed state estimation problem for both continuous- and discrete-time linear systems. A simply structured distributed estimator \{comprising interconnected local estimators\} is first described for estimating the state of a continuous and multi-channel linear system whose sensed outputs are distributed across a fixed multi-agent network. The estimator is then extended to non-stationary networks whose graphs switch according to a switching signal. The estimator is guaranteed to solve the problem, provided a network-wide shared high gain condition achieving a form of spectrum separation is satisfied. As an alternative to sharing a common gain across the network, a fully distributed version of the estimator is also studied in which each agent adaptively adjusts a local gain, though the practicality of this approach is subject to a robustness issue common to adaptive control. A discrete-time version of the distributed state estimation problem is also studied, and a corresponding estimator based again on spectrum separation, but not high gain, is proposed for time-varying networks. For each scenario, it is explained how to construct the estimator so that the state estimation errors in the local estimators all converge to zero exponentially fast at a fixed but arbitrarily chosen rate, provided the network’s graph is strongly connected for all time. The proposed estimators are inherently resilient to abrupt changes in the number of agents and communication links in the inter-agent communication graph upon which the algorithms depend, provided the network is redundantly strongly connected and redundantly jointly observable.

Key words: Distributed Estimation, Multi-Agent Systems, Cooperative Control, Linear Systems

1 Introduction

With the growing interest in sensor networks and multi-agent systems, the problem of estimating the state of a dynamical system whose measured outputs are distributed across a network has been under study in one form or another for a number of years [1,3–15].

Depending on the nature of the system to be estimated, the distributed estimation problem has continuous- and discrete-time versions. In its simplest form, the continuous-time version of the distributed state estimation problem starts with a network of \( m > 1 \) agents labeled \( 1, 2, \ldots, m \) which are able to receive information from their neighbors. Neighbor relations are characterized by a directed graph \( \mathbb{N} \), which may or may not depend on time, whose vertices correspond to agents and whose arcs depict neighbor relations. Each agent \( i \) senses a signal \( y_i \in \mathbb{R}^s, \ i \in \mathbb{m} \triangleq \{1, 2, \ldots, m\} \) generated by a continuous-time system of the form \( \dot{x} = Ax, \ y_i(t) = C_ix, \ i \in \mathbb{m} \) and \( x \in \mathbb{R}^n \). It is typically assumed that \( \mathbb{N} \) is strongly connected and that the system is jointly observable. It is invariably assumed that each agent receives certain real-time signals from its neighbors although what is received can vary from one problem formulation to the next. In all formulations, the goal is to devise local estimators, one for each agent, whose outputs are all asymptotically correct estimates of \( x \). The local estimator dynamics for agent \( i \) are typically
assumed to depend only on the pair \((C_i, A)\) and certain properties of \(N\). The problem is basically the same in discrete time, except that rather than the continuous-time model just described, the discrete-time model \(x(\tau + 1) = Ax(\tau), \ y_i(\tau) = C_i x(\tau), \ i \in m, \ x \in \mathbb{R}^n\) is considered instead. More precise problem formulations will be given later.

1.1 Background

The study of distributed state estimation for linear systems can be dated back to the so-called distributed Kalman filter problem [16], which involves system and measurement noise in the problem formulation and has been widely studied for years [3, 17]. Most available Kalman filter based approaches [3, 4, 6, 16, 17] require the agents to both share “signal information”, which can be measurements or local state estimates, and fuse certain “structural information”, which forms the covariance or information matrix of the nominal centralized Kalman filter for each agent. For the problem just described, the existing literature based on only “signal information” sharing can be classified into two categories, namely continuous- and discrete-time estimators/observers, except for our earlier work of [15, 18] in which a hybrid observer was proposed for a continuous-time linear system.

Continuous-time distributed estimators have recently appeared in [1, 10, 12–14, 19]. By recasting and then solving the distributed estimation problem as a classical decentralized control problem, the resulting estimator becomes capable of estimating the state at a pre-assigned exponentially fast rate, assuming \(N\) is a constant strongly connected graph [10]. The work of [12] seeks to propose a distributed estimator for a continuous-time system at the expense of certain design flexibility. This is done, in essence, by exploiting the \(A\)-invariance of the unobservable spaces of the pairs \((C_i, A)\); this in turn enables one to “split” the local estimators into two parts, one based on conventional spectrum assignment techniques applied to the observable part of the state at each local estimator and the other based on consensus among the unobserved parts of the state at each local estimator. The two parts are interacting but the use of a high gain serves to simplify the stability issue because of a split in the spectrum arising from the design of the estimator. The idea has been further developed in [13, 14, 20]. Specifically, these latter references start to move beyond a restriction in [12] permitting only constant, undirected, connected neighbor graphs to be addressed. The work of [13, 14] extends the result of [12] to the case when the neighbor graph is constant, directed, strongly connected, while requiring that one chooses gains to ensure that certain LMIs hold which are difficult to grasp intuitively. In [21], motivated by a distributed least squares problem, a modified algorithm is proposed to deal with measurement noise constant, undirected, connected neighbor graphs. A distributed adaptive algorithm has recently been proposed in [20] which allows agents to join or leave the network over time, provided the resulting agent network always remains jointly connected and joint detectable. An evident disadvantage of all these existing continuous-time distributed estimators is that they require a somewhat complicated gain computation procedure, and partially because of this, do not, at least not directly, admit discrete-time counterparts.

Discrete-time distributed estimators have been recently studied in [9, 11, 22–29]. Notable among them is the paper [9]. Published prior to the appearance of the early continuous time paper [10] applying to the same class of distributed systems, [9] solves the discrete-time distributed estimation problem for jointly observable, linear systems with constant, directed, strongly connected neighbor graphs. It builds on the idea of recasting the estimation problem as a classical decentralized control problem. Although the observer is limited to discrete-time systems, it has been proved possible to make use of the ideas in [9] to obtain, as noted earlier, a distributed observer for continuous-time systems [10], but still for constant neighbor graphs. There are however other discrete-time distributed observers/estimators which do not admit continuous-time extensions, illustrating that passage between discrete-time and continuous-time thinking may be harder than intuition initially suggests for distributed estimation problems. By expanding on earlier work in [11], the papers [26, 30] provide a procedure for constructing a centralized designed distributed observer for time-varying neighbor graphs. It requires the sharing of an index that records the age of the information across the network, and the agents are designed to act in a sequential manner to do state estimation. The resulting algorithm, which is tailored exclusively to discrete-time systems, requires a network-wide initialization step that serves to sort the agents in a specific order. Thereby it can deal with state estimation under assumptions that are weaker than strong connectivity.

Different approaches to the distributed state estimation problem are summarized in Table 1. It turns out that the current paper is the first paper that can deal with both continuous-time and discrete-time systems while ensuring exponential convergence under time-varying neighbor graphs.

The contribution of this paper rests on the following three distinguishing features, differentiating it and highlighting it as a development of earlier work:

- The paper describes a simply structured, unified approach to the distributed state estimation problem and to design and analyze the corresponding distributed estimators for both continuous- and discrete-time linear systems with possibly time-varying graphs. It is termed the “split-spectrum” approach because it “splits” the system spectrum into disjoint subsets corresponding to observable and unobservable subspaces. In continuous time, this is achieved by a high gain mechanism, but in discrete-time by a different
mechanism, viz. the adoption of two integrally related sampling rates. Though the mechanisms are instrumentally different, their purpose is fundamentally the same. It is termed ‘unified’ because the approach is shown to work for both continuous- and discrete-time linear systems over both constant and time-varying neighbor graphs.

- A fully distributed version of the estimator is separately studied where each agent can adaptively adjust a local gain, with simpler gain computation procedure and analysis compared with [20].

- Exponential convergence of the error dynamics is ensured with an arbitrarily assigned convergence rate. A great advantage of our methodology is that the designs and algorithms developed under a noiseless assumption are then necessarily tolerant of some level of noise, simply because we take care to ensure an exponential convergence.

It is assumed in this paper that the neighbor graph of the network is always strongly connected. From the perspective of the real world, requiring the underlying network to be strongly connected “at every time step” is an assumption that will occur in a great many (though obviously not all) cases, and as such, is deserving of a separate study in its own right. The extension to more general time-varying graphs is one future direction. It may not be conceptually difficult, however intricate the details may be.

The paper first describes the split-spectrum based distributed estimator for the case when the system dynamics are continuous with a stationary network in §2.1, and with associated background analysis is given in §2.2. The estimator is then extended to deal with non-stationary networks whose neighbor graphs switch according to a switching signal with a fixed dwell time or a variable dwell time with prescribed average, the ideas being detailed in §2.3. In the case when the interconnection among the agents can always be modelled using doubly stochastic matrices (e.g., undirected graphs with the Metropolis weights [5]), it is shown in §2.3 that the estimator functions correctly even if the neighbor graph switches arbitrarily, provided the graph is always strongly connected. The estimators mentioned above all rely on the existence of a sufficiently large, network-wide shared gain. A fully distributed version of the estimator is then studied in §2.4 where each agent can adaptively adjust a local gain. The adaptive estimator is subject to a robustness issue common to adaptive control. The proposed estimators, except for the adaptive one, are inherently resilient to abrupt changes in the number of agents and communication links in the inter-agent communication graph upon which the algorithms depend, an issue which is discussed in §2.5. Then the split-spectrum based estimator design is extended to the case when the system dynamics is discrete in §3 for both constant and time-varying neighbor graphs. Simulation validation is provided in §4.

The material in this paper was partially presented in [1, 2], but this paper presents a more comprehensive treatment of the work. Specifically, the paper crafts continuous-time distributed estimators for two types of non-stationary networks in §2.3 and a fully distributed adaptive estimator in §2.4, which were not included in [1, 2].

## 2 Continuous-Time Distributed Estimator

We are interested in a network of $m > 0$ (possibly mobile) autonomous agents labeled $1, 2, \ldots, m$ which are able to receive information from their “neighbors”, where by a neighbor of agent $i$ is meant any other agent within agent $i$’s reception range. We write $\mathcal{N}_i(t)$ for the labels of agent $i$’s neighbors at time $t \in [0, \infty)$ and always take agent $i$ to be a neighbor of itself. Neighbor relations at time $t$ are characterized by a directed graph $\mathbb{N}(t)$ with $m$ vertices and a set of arcs defined so that there is an arc in $\mathbb{N}(t)$ from vertex $j$ to vertex $i$ whenever agent $j$ is a neighbor of agent $i$ at time $t$. Since each agent $i$ is always a neighbor of itself, $\mathbb{N}(t)$ has a self-arc at each of its vertices. Each agent $i$ can sense a continuous-time signal $y_i \in \mathbb{R}^{n_i}$, $i \in \mathfrak{m} \triangleq \{1, 2, \ldots, m\}$, where

\begin{align}
    y_i &= C_i x, \quad i \in \mathfrak{m} \\
    \dot{x} &= Ax
\end{align}

\section*{Table 1. Comparison of Different Approaches to Design Distributed Estimators}

| Nature of Approach | Reference | Continuous-time Systems | Discrete-time Systems | Exponential Convergence | Time-varying Graphs |
|--------------------|-----------|-------------------------|-----------------------|------------------------|---------------------|
| Kalman Filter Based Approach | [3, 6, 16, 17] | × | ✓ | × | × |
| Observability Based Approach | [4] | ✓ | × | ✓ | × |
| Decomposition Based Approach | [11] | × | ✓ | ✓ | × |
| Decentralized Control Based Approach | [26, 30] | × | ✓ | ✓ | ✓ |
| Split-Spectrum Based Approach | [9, 24, 25] | × | ✓ | ✓ | × |
| This work | ✓ | ✓ | ✓ | ✓ | ✓ |
and $x \in \mathbb{R}^n$. We assume throughout that $C_i \neq 0$, $i \in \mathbf{m}$, and that the system defined by (1) and (2) is jointly observable; i.e., with $C = \left[ C'_1 \ C'_2 \ \cdots \ C'_{m} \right]'$, the matrix pair $(C, A)$ is observable. Joint observability is equivalent to the requirement that $\bigcap_{i \in \mathbf{m}} V_i = 0$, where $V_i$ is the unobservable space of $(C, A)$; i.e., $V_i = \ker \left[ C'_i \ (C_i A)' \ \cdots \ (C_i A^{n-1})' \right]'$. As is well known, $V_i$ is the largest $A$-invariant subspace contained in the kernel of $C_i$. Generalizing the results that follow to the case when $(C, A)$ is only detectable is quite straightforward and can be accomplished using well-known ideas. However, the commonly made assumption that each pair $(C_i, A)$, $i \in \mathbf{m}$, is observable, or even just detectable, is very restrictive, grossly simplifies the problem and is unnecessary. The assumption $C_i \neq 0$ is not necessary provided the more relaxed problem is properly formulated. The assumption is made for the sake of simplicity. The problem of interest is to construct a suitably defined family of linear estimators in such a way so that an arbitrarily chosen but fixed positive number.

This section proposes the estimator first, and then analyses the estimator beginning with the condition that the neighbor graph $N(t)$ is constant. A time-varying neighbor graph $\dot{N}(t)$ is then considered in which changes occur according to a switching signal. Later, a fully distributed algorithm based on use of multiplicative adaptive gain control is developed.

### 2.1 The Estimator

The estimator to be considered consists of $m$ local or private estimators of the form for each $i \in \mathbf{m}$,

$$\dot{x}_i = \left(A + K_i C_i \right)x_i - K_i y_i - gP_i \left(x_i - \frac{1}{m_i(t)} \sum_{j \in N_i(t)} x_j \right) \quad (3)$$

where $m_i(t)$ is the number of labels in $N_i(t)$, $g$ is a suitably defined positive gain common to all local estimators, each $K_i$ is a suitably defined matrix, and for each $i \in \mathbf{m}$, $P_i$ is the orthogonal projection on the unobservable space of $(C_i, A)$. The term $(A + K_i C_i)x_i - K_i y_i$ is designed to enable each agent $i$ to be able to recover the observable part of the state by itself, and the term $-gP_i \left(x_i - \frac{1}{m_i(t)} \sum_{j \in N_i(t)} x_j \right)$ is for the purpose of utilizing information from its neighbors to recover the unobservable part of the state. It will be shown that with this design, the spectrum of the system matrix can be split into two subsets. Details on how to design the parameters $K_i$ and $g$ will be provided in the following.

#### 2.1.1 The Error Dynamics

This subsection provides the error dynamics of the proposed estimator. It will be shown that the spectrum of the error system matrix can be split into two subsets based on the observability of each agent with $K_i$ defined properly.

First note from (3), that the state estimation error $e_i = x_i - x$ satisfies

$$\dot{e}_i = \left( A + K_i C_i \right)e_i - gP_i \left(e_i - \frac{1}{m_i(t)} \sum_{j \in N_i(t)} e_j \right) \quad (4)$$

Consequently the overall error vector $e = \left[ e'_1 \ \cdots \ e'_m \right]'$ satisfies

$$\dot{e} = (\bar{A} - gP(I_{mn} - \bar{S}(t)))e \quad (5)$$

where $\bar{A} = \text{block diag} \{ A + K_1 C_1, A + K_2 C_2, \ldots, A + K_m C_m \}$, $P = \text{block diag} \{ P_1, P_2, \ldots, P_m \}$, $\bar{S}(t) = S(t) \otimes I_n$ with $S(t) = D_{n(t)}^{-1} A'_{n(t)}$. Here $I_k$ is the $k \times k$ identity matrix, and $A_{n(t)}$ is the adjacency matrix of $N(t)$ and $D_{n(t)}$ is the diagonal matrix whose $i$th diagonal entry is the in-degree of $N(t)$’s $i$th vertex. Note that $N(t)$ is the graph of $S'(t)$ and that the diagonal entries of $S'(t)$ are all positive because each agent is a neighbor of itself. The matrix $S(t)$ is evidently a stochastic matrix.

#### Proposition 1

The spectrum of the error system matrix $\bar{A} - gP(I_{mn} - \bar{S})$ splits into two subsets, One subset contains the union of certain subsets of the eigenvalues associated with the $i$-th local estimator, $i \in \mathbf{m}$, these being able to be arbitrarily positioned by choice of the $K_i$ and independent of $g$. The second subset is independent of the choice of the $K_i$, and depends, though not to the extent of being able to be arbitrarily positioned, on $g$.

We remark that in the proof below, we will explain how to choose the $K_i$ to ensure that the associated set of eigenvalues has degree of stability $\lambda$ (ensuring an estimation error decay at least as fast as $\exp(-\lambda t)$), while subsequently we will explain how to choose $g$ to ensure stability of the remaining part of the spectrum with the same minimum exponential decay rate.

#### Proof of Proposition 1

The definitions of $K_i$ and $g$ begin with the specification of a desired convergence rate bound $\lambda > 0$. To begin with, each matrix $K_i$ is defined as follows. For each fixed $i \in \mathbf{m}$, write $Q_i$ for any full rank matrix whose kernel is the unobservable space of $(C_i, A)$ and let $C_i$ and $A_i$ be the unique solutions to $C_i Q_i = C_i$ and $Q_i A_i = A_i Q_i$, respectively. Then the matrix pair $(C_i, A_i)$ is observable. A matrix $\bar{K}_i$ can be chosen to ensure that the convergence of $\exp \{ (\bar{A}_i + \bar{K}_i C_i) t \}$ to zero is as fast as the convergence of $\exp(-\lambda t)$ to zero is. There are several well-documented ways to do this (e.g., spectrum assignment algorithms or Riccati equation solvers), since each pair $(C_i, A_i)$ is observable. Having chosen such $\bar{K}_i$, $K_i$ is then chosen to be $K_i = Q_i^{-1} \bar{K}_i$.

$\dagger$ The graph of an $n \times n$ matrix $M$ is that directed graph on $n$ vertices possessing a directed arc from vertex $i$ to vertex $j$ if $m_{i,j} \neq 0$ (p. 357, [31]).
where $Q_i^{-1}$ is a right inverse for $Q_i$. The definition implies that
\[
Q_i(A + K_i C_i) = (\tilde{A}_i + \tilde{K}_i \tilde{C}_i)Q_i
\]
and that $(A + K_i C_i) V_i \subset V_i$. The latter, in turn, implies that there is a unique matrix $A_i$ which satisfies
\[
(A + K_i C_i) V_i = V_i A_i
\]
where $V_i$ is a basis matrix\footnote{For simplicity, we assume that the columns of $V_i$ constitute an orthonormal basis for $V_i$, in which case $P_i = V_i V_i^T$.} for $V_i$.

Next we show what defining the $K_i$ in this way accomplishes. Note that the subspace $V = V_1 \oplus V_2 \oplus \cdots \oplus V_m$ is $A$-invariant because $(A + K_i C_i) V_i \subset V_i$, $i \in m$. Next, let $Q = \text{block diag} \{Q_1, Q_2, \ldots, Q_m\}$ and $V = \text{block diag} \{V_1, V_2, V_3, \ldots, V_m\}$ where $V_i$ is a matrix whose columns form an orthonormal basis for $V_i$. Then $Q$ is a full rank matrix whose kernel is $V$ and $V$ is a basis matrix for $V$ whose columns form an orthonormal set. It follows that $P = V V'$, that $Q P = 0$, and that
\[
\begin{align*}
Q \tilde{A} &= \tilde{A}_V Q \\
AV &= V A_V
\end{align*}
\]
where $\tilde{A} = \text{block diag} \{A_1, A_2, \ldots, A_m\}$ and
\[
A_V = \text{block diag} \{\tilde{A}_1 + \tilde{K}_1 \tilde{C}_1, \ldots, \tilde{A}_m + \tilde{K}_m \tilde{C}_m\}. \tag{10}
\]

Let $V^{-1}$ be any left inverse for $V$ and let $Q^{-1}$ be that right inverse for $Q$ for which $V^{-1} Q^{-1} = 0$. Then
\[
\tilde{A} - g P (I_{mn} - \bar{S}(t)) T \begin{bmatrix} \tilde{A}_V & 0 \\ \tilde{A}_V(t) & A_V(t) \end{bmatrix} T^{-1}
\]
where $\tilde{A}_V(t) = V^{-1} (\tilde{A} - g (I_{mn} - \bar{S}(t)) Q^{-1}$ and $A_V(t) = \tilde{A} - g V' (I_{mn} - \bar{S}(t)) V$. Here $T = \begin{bmatrix} Q^{-1} & V \end{bmatrix}$. It is easy to check that $T^{-1} = \begin{bmatrix} Q' & V \end{bmatrix}$. According to (11), the spectrum of $\tilde{A} - g P (I_{mn} - \bar{S})$ is equivalent to the union of the spectrum of $\tilde{A}_V$ and $A_V$. Recall that the $\tilde{K}_i$ have been already been chosen so that each matrix exponential $\exp(\tilde{A}_i + \tilde{K}_i \tilde{C}_i t)$ converges to zero as fast as $\exp(-\lambda t)$ does. Because of this and the fact that $\tilde{A}_V(t)$ is a bounded matrix, to ensure that for each fixed $\tau$, the state transition matrix $\Phi(t, \tau)$ converges to zero as fast as $\exp(-\lambda t)$ does, it is enough to choose $g$ so that the state transition matrix of $A_V(t)$ converges to zero at least as fast as $\exp(-\lambda t)$ does. The requisite condition on $g$ is provided below for three different neighbor graph connectivity assumptions.

### 2.2 Constant Neighbor Graph

This subsection focuses on the case when the neighbor graph $N(t)$ is a constant graph $N$. The following result can be obtained.

**Theorem 1** For any given positive number $\lambda$, if the neighbor graph $N$ is fixed and strongly connected, and the system defined by (1) and (2) is jointly observable, there are matrices $K_i$, $i \in m$ such that for $g$ sufficiently large, each estimation error $x_i(t) - \hat{x}_i(t)$ of the distributed estimator defined by (3), converges to zero as $t \to \infty$ as fast as $\exp(-\lambda t)$ converges to zero.

The proof of the theorem involves making use of the following result, with all proofs being contained in Appendix A. In particular, the proof of Proposition 2 makes use of the properties of strong connectivity of the neighbor graph and joint observability of the system.

**Proposition 2** $-V'(I_{nn} - \bar{S})V$ is a continuous-time stability matrix.

### 2.3 Switching Neighbor Graph

In the sequel the problem is studied under the assumption that $N(t)$ changes according to a switching signal with a fixed dwell time or a variable dwell time with fixed average. To characterize the assumed time dependence of $N(t)$, let $G = \{G_1, G_2, \ldots, G_n\}$ denote the set of all directed, strongly connected graphs on $m$ vertices which have self-arcs at all vertices; here $n_G$ is the number of graphs in $G$. In some situations, the switching signals always have consecutive discontinuities separated by a value which is no less than a fixed positive real number $\tau_D$. It is called a dwell time [32]. In certain situations, the switching signals may occasionally have consecutive discontinuities separated by less than $\tau_D$, but for which the average interval between consecutive discontinuities is no less than $\tau_D$. This leads to the concept of an average dwell-time. With $\tau_D$ and $\delta$ fixed define $S_{avg}$ for the set of all piecewise-constant switching signals $\sigma : [0, \infty) \to \{1, 2, \ldots, |G|\}$ satisfying $\delta_\sigma(t_0, t) \leq \delta_0 + \frac{t - t_0}{\tau_D}$. Here $\delta_\sigma(t_0, t)$ denotes the number of discontinuities of $\sigma$ in the open interval $(t_0, t)$. The constant $\tau_D$ is called the average dwell-time and $\delta_0$ the chatter bound [33]. By the set of all time-varying neighbor graphs with average dwell-time $\tau_D$ is meant the set $G_\sigma : \sigma \in S_{avg}$. Note that switching according to an average dwell time is a special case of switching according to an average dwell time. In the following, it is assumed that $N \in \{G_\sigma : \sigma \in S_{avg}\}$.

The problem to which this subsection is addressed is this. For fixed averaged dwell-time $\tau_D$ and the chatter bound $\delta_0$, devise a procedure for crafting $m$ local estimators, one for each agent, so that for each neighbor graph $N \in \{G_\sigma : \sigma \in S_{avg}\}$, all state estimation errors converge to zero exponentially fast at a prescribed rate.

The estimator to be considered is still the same as the
estimator described in (3), with the exception that \( g \) is
different. The following result can be derived.

**Theorem 2** For any fixed positive numbers \( \tau_D \) and \( \lambda \),
there exists a positive number \( g^* \) with the following property.
For any value of \( g \geq g^* \), any neighbor graph \( N \in \{ G_\sigma : \sigma \in \mathcal{S}_{avg} \} \), if the system defined by (1) and (2)
is jointly observable, there are matrices \( K_i, i \in \mathbf{m} \) such that
each state estimation error \( e_i = x_i - x^* \), \( i \in \mathbf{m} \)
of the distributed estimator defined by (3) converges to zero
as \( t \to \infty \) as fast as \( \exp(-\lambda t) \) does.

The proof of Theorem 2 depends on the following lemma.

**Lemma 1** Let \( M_1, M_2, \ldots, M_{|G|} \) be a set of \( n \times n \) expon-
entially stable real matrices associated with a set \( G = \{ G_1, G_2, \ldots, G_{|G|} \} \) of directed strongly
connected graphs with self-arcs at all vertices. Let \( \sigma \) denote the switching
signal with average dwell time \( \tau_D \) governing the selection of
a graph from \( G \). Then for any \( n \times n \) real matrix \( N \) and
positive number \( \lambda \) there is a positive number \( g^* \), depending
on \( \tau_D \) for which, for each \( \sigma \in \mathcal{S}_{avg} \) and \( g \geq g^* \), all solutions to
\[
\dot{x} = (N + gM_\sigma)x
\]
(12)
converge to zero as fast as \( \exp(-\lambda t) \) does.

The proofs of Lemma 1 and Theorem 2 can be found in
Appendix A.

In the sequel the problem is studied for a certain type
of switching neighbor graphs. It turns out that if the
stochastic matrices of undirected neighbor graphs are chosen
to be doubly stochastic, there exist estimators which can deal
with arbitrary switching signals, and the notion of dwell times ceases to be relevant.
The estimator to be considered is again the same as (3), with the
exception that \( g \) is chosen differently.

**Theorem 3** For any fixed positive number \( \lambda \), there exists
a positive number \( g^* \) with the following property. For any
value of \( g \geq g^* \), any time-varying neighbor graph \( N(t) \), if the system defined by (1) and (2) is jointly ob-
serveable, the neighbor graph \( N(t) \) is undirected and con-
cected, and the stochastic matrix \( S(t) \) of graph \( N(t) \) is
doubly stochastic, there are matrices \( K_i, i \in \mathbf{m} \) such that
each state estimation error \( x_i(t) - x^* \) of the distributed
observer defined by (3), converges to zero as \( t \to \infty \) as
fast as \( \exp(-\lambda t) \) converges to zero.

The proof of Theorem 3 can be found in Appendix A.

Given that \( \hat{A} \) has eigenvalues which are a subset of those
of \( A \), it is seen that the effect of large \( g \) is to force the in-
stantaneous value of the eigenvalues of \( A(t) \) well to the
left of those of \( \hat{A} \), and indeed the same for the Lyapunov
exponent. This is a spectral separation idea – consensus
dynamics within the estimator are faster than those of the
original system.

### 2.4 Distributed Estimator with Adaptive Gains

Obviously it may be disadvantageous to share a gain
across the whole network, and here we aim to design a
simple adaptive distributed estimator to get gains for
each estimators in a distributed way. The estimator for
each agent \( i \) still has the form (3) while each agent \( i \) has
its own gain \( g_i \) which is obtained by
\[
\dot{g}_i = |V_i| \sum_{j \in N_i} \frac{1}{m_i} |x_j - x_i|^2, \quad i \in \mathbf{m}
\]
(13)
where \( | \cdot |_2 \) denotes the two norm of a matrix and \( g_i(0) \)
is nonnegative but otherwise arbitrary. Key questions arising are whether the \( g_i \) are bounded, and whether the matrices \( K_i \) can be chosen in the same way as previously.
We have in fact with \( K_i \) chosen as in the proof of
Proposition 1:

**Theorem 4** For any fixed positive numbers \( \tau_D \) and \( \lambda \),
if the system defined by (1) and (2) is jointly ob-
serveable, and the gain is defined by (13), there are matrices \( K_i, i \in \mathbf{m} \) such that all the \( g_i \) are bounded, and
each state estimation error \( e_i = x_i - x^* \) of the distributed
observer defined by (3) asymptotically converges to zero
as \( t \to \infty \).

The proof of Theorem 4 can be found in Appendix A.

As with any adaptive control algorithm, we must recog-
nize that there are fundamental challenges that can arise
in practice and have the potential to undermine the ap-
proach [34]; these include the need to work with models
of plants that may be very accurate but are virtually
never exact; the inability to know, given an unknown
plant, whether a desired control objective is practical or
impractical, and the possibility of transient instability,
or extremely large signals occurring before convergence.
Thus, for this paper, our preference is to stick with a
given \( g \) instead of using an adaptive algorithm.

### 2.5 Resilience

The concept of a passively resilient algorithm is pro-
posed in [18]. By a passively resilient algorithm for a
distributed process is meant an algorithm which, by ex-
ploring built-in network and data redundancies, is able
to continue to function correctly in the face of abrupt
changes in the number of vertices and arcs in the inter-
agent communication graph upon which the algorithm
depends. All the proposed continuous-time distributed
estimators, except for the adaptive one, are inherently resil-
ient to these abrupt changes provided the network is
redundantly strongly connected and redundantly jointly
observable, with a careful gain picking before the algo-
rum starts. Details can be found in Section 5 of [18].
The same resilience property is also possessed by the
discrete-time distributed estimators to be developed in
the next section.

### 3 Discrete-Time Distributed Estimator

In this section, a discrete time version of the distributed
estimator problem is studied. The estimator which solves
this problem in discrete time is described. Of central
over for bors do not change between event times. In other words, each successive pair of event times; what this informa-
from its neighbors at a finite number of times between
consists of
With this assumption, the estimator to be considered
rect estimate of
i
3
x
Each agent
i
jointly ob-
and
x
τ
∈
IR
y
(τ)
is an “averaged state” computed recursively
steps during the real time interval \([τT, (τ + 1)T]\)
and \(x \in \mathbb{R}^n\). We assume throughout that \(\mathbb{N}(t)\) is strongly connected, and the system defined by \((14)\) is jointly ob-
Each agent \(i\) is to estimate \(x\) using a dynamical system whose output \(x_i(τ) \in \mathbb{R}^n\) is to be an asymptotically correct estimate of \(x(τ)\) in the sense that the estimation error \(x_i(τ) - x(τ)\) converges to zero as \(τ \to \infty\) as fast as \(λ^*\) does, where \(λ\) is an arbitrarily chosen but fixed positive number\(^3\) less than 1. To accomplish this it is assumed that the information agent \(i\) can receive from neighbor \(j\) at event time \(τT\) is \(x_j(τ)\). It is further assumed that agent \(i\) can also receive certain additional information from its neighbors at a finite number of times between each successive pair of event times; what this information is will be specified below.

3.1 The Estimator

In this section it will be assumed that each agent’s neighbors do not change between event times. In other words, for \(i \in \mathbb{m}\),

\[
\mathcal{N}_i(t) = \mathcal{N}_i(τT), \quad t \in [τT, (τ + 1)T), \quad τ = 0, 1, 2, \ldots
\]

With this assumption, the estimator to be considered consists of \(m\) private estimators, one for each agent. The estimator for agent \(i\) is of the form

\[
x_i(τ + 1) = (A + K_iC_i)\bar{x}_i(τ) - K_iy_i(τ)
\]

where \(\bar{x}_i(τ)\) is an “averaged state” computed recursively over \(q\) steps during the real time interval \([τT, (τ + 1)T]\) using the update equations

\[
z_i(0, τ) = x_i(τ)
\]

\[
z_i(k, τ) = (I - P_i)z_i(k - 1, τ) + \frac{1}{m_i(τ)} P_i \sum_{j \in \mathcal{N}_i(τT)} z_j(k - 1, τ), \quad k \in q
\]

\[
\bar{x}_i(t) = z_i(q, τ)
\]

Here \(m_i(τ)\) is the number of labels in \(\mathcal{N}_i(τT)\), \(q\) is a suit-
ably defined positive integer, further detail being given below, \(q = \{1, 2, \ldots, q\}\), and \(P_i\) is the orthogonal projection on the unobservable space of \((C_i, A)\). Each matrix \(K_i\) is defined as explained in the next paragraph. Meanwhile, we note that the estimators incorporate two time scales. An agent’s local estimator obtains data from its neighbors at a rate \(q\) times that at which it obtains measurement data from the underlying system.

As described in §2.1, for fixed \(i \in \mathbb{m}\), write \(Q_i\) for any full rank matrix whose kernel is the unobservable space of \((C_i, A)\), and let \(C_i\) and \(A_i\) be the unique solutions to \(C_iQ_i = C_i\) and \(Q_iA = A_iQ_i\) respectively. Let \(λ\) be a positive value bounded by 1. Then the matrix pair \((C_i, A_i)\) is observable. Thus by using a standard spectrum assign-
ment algorithm, a matrix \(K_i\) can be chosen to ensure that the convergence of \((A_i + K_iC_i)τ\) to zero as \(τ \to \infty\) is as fast as the convergence to zero of \(λ^*\). Having chosen such \(K_i\), \(K_i\) is then defined to be \(K_i = Q_i^-K_i\) where \(Q_i^-\) is a right inverse for \(Q_i\). To explain what needs to be considered in choosing \(q\), which is a rough analog of the gain \(g\) of the continuous-time solution, it is necessary to describe the structure of the “error model” of the overall estimator. This will be done next.

3.1.1 The Error Dynamics

For \(i \in \mathbb{m}\), write \(e_{i}(τ)\) for the state estimation error \(e_{i}(τ) = x_{i}(τ) - x(τ)\). In view of \((15)\),

\[
e_{i}(τ + 1) = (A + K_iC_i)e_{i}(τ)
\]

where \(e_{i}(τ) = \bar{x}_{i}(τ) - x(τ)\). Moreover if \(e_{i}(k, τ) \triangleq z_{i}(k, τ) - x(τ), \quad k \in \{0, 1, \ldots, q\}\) then for \(k \in q\,

\[
e_{i}(0, τ) = e_{i}(τ)
\]

\[
e_{i}(k, τ) = (I - P_i)e_{i}(k - 1, τ) + \frac{1}{m_i(τ)} P_i \sum_{j \in \mathcal{N}_i(τT)} e_j(k - 1, τ)
\]

\[
e_{i}(τ) = e_{i}(q, τ)
\]

because of \((16) - (18)\). It is possible to combine these \(m\) subsystems into a single system. Paralleling §2.1.1 let \(e = \text{col} \{e_1, e_2, \ldots, e_m\}\), define \(\bar{A} = \text{block diag} \{A + K_1C_1, A + K_2C_2, \ldots, A + K_mC_m\}, \quad \bar{P} = \text{block diag} \{P_1, P_2, \ldots, P_m\}\) and write \(S(τ)\) for the stochastic matrix \(S(τ) = D_{N(τT)}^{-1} A_i N(τT)\) where \(A_i N(τT)\) is the adjacency matrix of \(N(τT)\) and \(D_{N(τT)}\) is the diagonal matrix whose \(i\)th diagonal entry is the in-degree of \(N(τT)\)’s
ith vertex. Let \(\tilde{e}(\tau) = \text{col}\{e_1(\tau), e_2(\tau), \ldots, e_m(\tau)\}\) and \(e(k, \tau) = \text{col}\{e_1(k, \tau), e_2(k, \tau), \ldots, e_m(k, \tau)\}\). Then
\[
e(\tau + 1) = \tilde{A}\tilde{e}(\tau)
\]
and
\[
\epsilon(0, \tau) = \epsilon(\tau)
\]
\[
\epsilon(k, \tau) = (I_{mn} - P(I_{mn} - \bar{S}(\tau)))\epsilon(k - 1, \tau), \quad k \in \mathbb{Q}
\]
where \(\bar{S}(\tau) = S(\tau) \otimes I_n\). Clearly \(\epsilon(\tau) = (I_{mn} - P(I_{mn} - \bar{S}(\tau)))^q\epsilon(\tau)\), so
\[
\epsilon(\tau + 1) = \bar{A}(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^q\epsilon(\tau)
\]
(19)

Our immediate aim is now to explain why for \(q\) sufficiently large, the time-varying matrix \(\bar{A}(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^q\) appearing in (19) is a discrete-time stability matrix for which the product
\[
\Phi(\tau) = \prod_{s=1}^{\tau} \bar{A}(I_{mn} - P(I_{mn} - \bar{S}(s)))^q
\]
converges to zero as \(\tau \to \infty\) as fast as \(\lambda^*\) does.

To begin with, we explore the property of matrix \(\bar{A}(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^q\).

**Proposition 3** The spectrum of the error system matrix \(\bar{A}(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^q\) splits into two subsets. One subset contains the union of certain subsets of the eigenvalues associated with the \(i\)-th local estimator, \(i \in \mathbb{m}\), these being able to be arbitrarily positioned by choice of the \(K_i\) and independent of \(q\). The second subset is independent of the choice of the \(K_i\) and depends, though not to the extent of being able to be arbitrarily positioned, on \(q\).

**Proof of Proposition 3:** As described in §2.1, note that the subspace \(\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \cdots \oplus \mathcal{V}_m\) is \(A\)-invariant because \((A + K_iC_i)\mathcal{V}_i \subset \mathcal{V}_i\), \(i \in \mathbb{m}\). Next, let \(Q = \text{block diag}\{Q_1, Q_2, \ldots, Q_m\}\) and \(V = \text{block diag}\{V_1, V_2, \ldots, V_m\}\). It follows that \(P = VV'\), (8) and (9). Also as before, \((A + K_iC_i)V_i = V_iA_i\). Moreover
\[
Q(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^q = Q
\]
(21)
\[
(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^qV = V(V'\bar{S}(\tau)V)^q
\]
(22)

Note that (21) holds because \(QP = 0\). To understand why (22) is true, note first that \((I_{mn} - P(I_{mn} - \bar{S}(\tau)))V = V(I_{m} - V'(I_{mn} - \bar{S}(\tau)V)\text{ because }P = VV'\text{; here }\bar{m} = \text{dim}V\). But \(I_{m} - V'(I_{mn} - \bar{S}(\tau)V = V'S(V\bar{S}(\tau)V\text{ because }V'V = I_{m}\). Thus (22) holds for \(q = 1\); it follows by induction that (22) holds for any positive integer \(q\).

Using (8), (9), (21), and (22), one obtains the equations
\[
Q\bar{A}(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^q = \bar{A}VQ
\]
(23)
\[
\bar{A}(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^qV = VAV(\tau)
\]
(24)

where
\[
AV(\tau) = \tilde{A}(V'\tilde{S}(\tau)V)^q
\]
(25)

with \(\tilde{A} = \text{block diag}\{A_1, A_2, \ldots, A_m\}\). These equations imply that
\[
\bar{A}(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^q = T
\]
(26)

\[\begin{bmatrix}
\bar{A}V & 0 \\
\bar{A}V(\tau) & A\bar{V}(\tau)
\end{bmatrix}
\]

with \(T = \left[Q^{-1} V\right]\) and \(\bar{A}\bar{V}(\tau) = V^{-1}\bar{A}(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^qQ^{-1}\).

According to (26), the spectrum of \(\bar{A}(I_{mn} - P(I_{mn} - \bar{S}(\tau)))^q\) is equivalent to the union of the spectrum of \(AV\) and \(AV(\tau)\).

3.2 Time-varying Neighbor Graph

The following result can be obtained when the neighbor graph is time-varying but strongly connected.

**Theorem 5** For any given \(\lambda\) with \(|\lambda| < 1\), if the neighbor graph \(\mathbb{N}(\tau)\) is strongly connected, and the system defined by (14) is jointly observable, there are matrices \(K_i\), \(i \in \mathbb{m}\) such that for sufficiently large \(q\), each estimation error \(x_i(\tau) - x(\tau)\) of the distributed estimator defined by (15)-(18), converges to zero as \(\tau \to \infty\) as fast as \(\lambda^*\) converges to zero.

The proof of the theorem involves studying the transition matrix and making use of the following results, with all proofs being contained in Appendix B.

**Lemma 2** Let \(M\) be an \(m \times m\) row stochastic matrix which has a strongly connected graph. There exists a diagonal matrix \(\Pi_M\) whose diagonal entries are positive for which the matrix \(L_M = \Pi_M - M\Pi_M\) is positive semi-definite; moreover \(L_M1 = 0\) where \(1\) is the \(m\)-vector of 1s. If, in addition, the diagonal entries of \(M\) are all positive, then the kernel of \(L_M\) is one-dimensional.

**Proposition 4** For each fixed value of \(\tau\),
\[
(V'\tilde{S}(\tau)V)'R(\tau)(V'\tilde{S}(\tau)V) - R(\tau) < 0
\]
(27)

where \(R(\tau)\) is the positive definite matrix, \(R(\tau) = V'(\Pi S(\tau) \otimes I_n)V\).

**Remark 1** It should be noticed that the computation of certain gains (\(q\) for the continuous-time case, and \(q\) for the discrete-time case) requires a centralized design. Besides this, all other design steps are distributed. Even though the computation of certain gains requires going

\[\text{The notation } \bar{A}\bar{V} \text{ and } A\bar{V} \text{ are different from the two defined in §2.}\]
over all possible strongly connected directed graphs on \( n \) vertices, which is a computationally intensive step, a clear distinction needs to be drawn between the computations required for designing the algorithm, and those required to run it. In design, we can afford to do more computations.

4 Simulations

This section provides simulations to illustrate the state estimation performance for both continuous and discrete time systems. The neighbor graph in some simulations will switch back and forth between Fig. 1 (a) and Fig. 1 (b). On occasion, it can serve to model a connection failure happening between agent 1 and agent 3 randomly.

4.1 Continuous dynamics

Consider the three channel, four-dimensional, continuous-time system described by the equations \( \dot{x} = Ax, \quad y_i = C_i x, \ i \in \{1, 2, 3\} \), where

\[
A = \begin{bmatrix}
0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & -2 & 0
\end{bmatrix}
\]

and \( C_i \) is the \( i \)th unit row vector in \( \mathbb{R}^{1 \times 4} \). Note that \( A \) is a matrix with eigenvalues at \( \pm 1, \pm j \), and \( \pm 1.4142j \). While the system is jointly observable, no single pair \( (C_i, A) \) is observable. The local observer convergence rate is designed to be at least with rate \( \lambda = 1 \). The first step is to design \( K_i \) as stated in §2.1. This is to control the spectrum of the matrix \( \hat{A}_V \) as defined in (11).

For agent 1:

\[
A_1 = \begin{bmatrix}
0 & -1 \\
1 & 0
\end{bmatrix}, \quad Q_1 = \begin{bmatrix}
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0
\end{bmatrix}, \quad V_1 = \begin{bmatrix}
0 & 0 & 1 & 0
\end{bmatrix}, \quad K_1 = \begin{bmatrix}
-5 & -5 & 0 & 0
\end{bmatrix}
\]

For agent 2:

\[
A_2 = \begin{bmatrix}
0 & -1 \\
1 & 0
\end{bmatrix}, \quad Q_2 = \begin{bmatrix}
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}, \quad V_2 = \begin{bmatrix}
0 & 0 & 1 & 0
\end{bmatrix}, \quad K_2 = \begin{bmatrix}
5 & -5 & 0 & 0
\end{bmatrix}
\]

For agent 3:

\[
A_3 = \begin{bmatrix}
0 & -2 \\
1 & 0
\end{bmatrix}, \quad Q_3 = \begin{bmatrix}
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0
\end{bmatrix}, \quad V_3 = \begin{bmatrix}
1 & 0 & 0 & 0
\end{bmatrix}, \quad K_3 = \begin{bmatrix}
0 & 0 & -5 & -4
\end{bmatrix}
\]

Two cases are considered. First, suppose the neighbor graph \( N(t) \) is fixed as shown in Figure 1(a). The system considered includes input white noise with zero mean, that is \( \dot{x} = Ax + v \) where \( E[v(t)] = 0, \quad E[v(t)v'(s)] = 0.5^2 \delta(t - s) \). With \( g = 10 \) obtained using (30), the real part of the right most eigenvalue of \( A_V \) is less than \( -1 \). With randomly chosen initial state values, traces of this simulation are shown in Fig. 2a where \( x_1 \) and \( x_3 \) denote the first components of \( x_1 \) and \( x_3 \) respectively. Moreover, the norm of the estimation error is plotted in Fig. 2b from which we can see that it is exponentially convergent with the approximate rate \( \lambda = 1 \).

Second, suppose the neighbor graph \( N(t) \) is time-varying and switching back and forth between Figure 1(a) and Figure 1(b) according to the indicator function in Fig. 3. That is, when the function value is 1, the neighbor graph is Figure 1(a), and when the function value is 0, the neighbor graph is Figure 1(b). It is arranged that the average dwell time is \( \tau_D = 0.0369 \) for this simulation. With zero measurement noise the corresponding solution trajectories for \( x_1 \) and \( x_3 \) are shown in Fig. 4a and the norm of the estimation error is shown in Fig. 4b.
rate we can see that it is exponentially convergent with the
of the estimation error is plotted in Fig. 6a from which
a spectral radius of \( \tilde{\lambda} = 0 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed as stated in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
scribed in \( \S 4.1 \). The observer convergence rate is de-
Therefore $2I_m - L$ must be irreducible. Note that the row sums of $2I_m - L$ all equal 2. Since $2I_m - L$ is nonnegative, its infinity norm is 2 so its spectral radius is no greater than 2 \{Theorem 5.6.9, \cite{31}\}. Moreover 2 is an eigenvalue of $2I_m - L$. Thus by the Perron-Frobenius Theorem, the geometric multiplicity of this eigenvalue is one. It follows that the geometric multiplicity of the eigenvalue of $L$ at 0 is also one.

We claim that $L$ is positive semi-definite. To establish this claim, note that $L$ can also be written as $L = D - \hat{A}$ where $D$ is a diagonal matrix whose entries are the diagonal entries of $L$ and $\hat{A}$ is the nonnegative matrix $\hat{A} = D - L$. As such, $L$ is the generalized Laplacian \cite{35} of that simple undirected graph $\mathcal{G}$ whose adjacency matrix is the matrix which results when the nonzero entries $a_{ij}$ in $\hat{A}$ are replaced by ones. Since $L$ can also be written as

$$L = \sum_{(i,j) \in \mathcal{E}} a_{ij}(u_i - u_j)(u_i - u_j)'$$

where $u_i$ is the $i$th unit vector and $\mathcal{E}$ is the edge set of $\mathcal{G}$, $L$ is positive semi-definite as claimed.

To proceed, set

$$H = \text{block diag} \{ \pi_1I_{n_1}, \pi_2I_{n_2}, \ldots, \pi_mI_{n_m} \}$$

where $n_i = \dim \mathcal{V}_i$ and note that $H = \Pi \otimes I_n$. Since $S - I_m$ is positive definite, it is enough to show that $H(V'(I_{mn} - \bar{S})V)'$ converges to zero as fast as $\exp(-\lambda t)$ does.

Observe that this is a Lyapunov equation for the positive definite function $z'Hz$. Therefore to show that $-V'(I_{mn} - \bar{S})V$ is a stability matrix, it is enough to show that $V'(L \otimes I_n)V$ is positive definite.

Since $L$ is positive semi-definite, so must be $L \otimes I_n$. Therefore $V'(L \otimes I_n)V$ is at least positive semi-definite. Suppose $z'V'(L \otimes I_n)Vz = 0$ where $z = \col \{ z_1, z_2, \ldots, z_m \}$ and $z_i \in \mathbb{R}^{\dim(\mathcal{V}_i)}$. To show that $V'(L \otimes I_n)V$ is positive definite, it is enough to show that $z = 0$. Since $L$’s eigenvalue at 0 has multiplicity one, ker $L$ = span \{1\}; therefore ker $(L \otimes I_n) = \text{column span} \ 1 \otimes I_n$. The hypothesis $z'V'(L \otimes I_n)Vz = 0$ implies that $(L \otimes I_n)Vz = 0$ so $Vz \in \ker(L \otimes I_n)$. Therefore $V_i z_i = V_{jz} = 0$, $i, j \in \mathbf{m}$. But because of joint observability, $\bigcap_{i \in \mathbf{m}} \mathcal{V}_i = 0$ so $V_i z_i = 0$, $i \in \mathbf{m}$. Thus $z_i = 0$, $i \in \mathbf{m}$ so $z = 0$ implying that $V'(L \otimes I_n)V$ is positive definite. Therefore $-V'(I_{mn} - \bar{S})V$ is a continuous-time stability matrix as claimed.

Proof of Theorem 1: Recall that the state estimation error satisfied (4). The overall error dynamic is defined as (5). According to (11), the spectrum of $\bar{A} - gP(I_{mn} - \bar{S})$ is equivalent to the union of the spectrum of $A_V$ and $A_P$. Since the spectrum of $A_i + \hat{K}_i \hat{C}_i$, $i \in \mathbf{m}$, is assignable with $\hat{K}_i$, to show for $g$ sufficiently large that $A - gP(I_{mn} - \bar{S})$ is a continuous-time stability matrix with a prescribed convergence rate as large as $\lambda$, it is enough to show that for $g$ sufficiently large, the matrix $A_V = \bar{A} - gV'(I_{mn} - \bar{S})V$ is a continuous-time stability matrix with a prescribed convergence rate as large as $\lambda$.

To show that $\exp((\bar{A} - gV'(I_{mn} - \bar{S})V)\tau)$ can be made to converge to zero as fast as $\exp(-\lambda \tau)$ does by choosing $g$ sufficiently large, we exploit (29). Note in particular that

$$H(\lambda I + A_V) + (\lambda I + A_V)'H = H(\lambda I + \bar{A}) + (\lambda I + \bar{A})'H - gV'(L \otimes I_n)V$$

Since $V'(L \otimes I_n)V$ is positive definite, by picking $g$ sufficiently large, $H(\lambda I + \bar{A}) + (\lambda I + \bar{A})'H$ will be negative definite implying that $A + A_V$ is a stability matrix and thus that $\bar{A} - gV'(I_{mn} - \bar{S})V$ is a stability matrix for which $\exp((\bar{A} - gV'(I_{mn} - \bar{S})V)\tau)$ converges to zero as fast as $\exp(-\lambda \tau)$ does. In other words, any value of $g$ will have the desired property provided

$$g \geq \frac{\lambda_{\min}(V'(L \otimes I_n)V)}{\lambda_{\max}(H(\lambda I + \bar{A}) + (\lambda I + \bar{A})'H)}$$

where $\lambda_{\max}(.)$ and $\lambda_{\min}(.)$ are the largest eigenvalue and the smallest eigenvalue of a symmetric matrix respectively.

Proof of Lemma 1: By hypothesis, each $M_i$ is exponentially stable. Thus there are positive constants $c_i > 1$ and $\lambda_i$ such that

$$\|\exp(M_i t)\| \leq c_i \exp(-\lambda_i t)$$

for any $i \in \{1, 2, \ldots, |G|\}$. Here $\|\cdot\|$ is any given submultiplicative norm on $\mathbb{R}^{n \times n}$. Let

$$c = \max_{i \in \{1, 2, \ldots, |G|\}} c_i, \text{ and } \lambda^* = \min_{i \in \{1, 2, \ldots, |G|\}} \lambda_i.$$  

Fix $\lambda > 0$ and let $g$ be any gain satisfying

$$g \geq \frac{\tau D(\lambda + |N| c) + \ln c}{\tau D \lambda^*}$$

We claim that for any number $\tau$, and any switching signal $\sigma \in \mathcal{S}_{avg}$, the transition matrix of $gM_{\sigma}$, namely

\footnote{The symbols used in this proof such as $g$, $c$ and $\lambda^*$ are generic and do not have the same meanings as the same symbols do when used elsewhere in this paper.}
\( \Phi_{\sigma}(t, \tau) \), converges to zero as fast as \( \exp(-\alpha t) \) does where
\[
\alpha = \lambda + \|N\|c
\]
(33)
To understand why this is so, by (31),
\[
\|\Phi_{\sigma}(t, \tau)\| \leq e^{\delta_\sigma(t, \tau)} \exp(-g\lambda^*(t - \tau))
\]
where \( \delta_\sigma(t, \tau) \) is the number of switching between \( (\tau, t) \).

By (32), \( \exp(g\lambda^*) \geq e^{\frac{t}{\tau_D}} \exp(\alpha) \). From this and the fact that
\[
\delta_\sigma(t, \tau) \leq \delta_0 + \frac{t}{\tau_D},
\]
(34)
\[
\|\Phi_{\sigma}(t, \tau)\| \leq e^{\delta_0} e^{\frac{t}{\tau_D}} \exp(-g\lambda^*(t - \tau)) \leq e^{\delta_0} \exp(-\alpha(t - \tau)).
\]

Thus, the claim is true.

In view of (12) and the variation of constants formula,
\[
x(t) = \Phi_{\sigma}(t, 0)x(0) + \int_0^t \Phi_{\sigma}(t, \mu)N x(\mu)d\mu
\]
(35)
As \( \|\Phi_{\sigma}(t, \tau)\| \leq c \exp(-\alpha t) \) for all \( \tau \) and \( \sigma \in S_{\text{avg}} \),
\[
\|x(t)\| \leq c \exp(-\alpha t)\|x(0)\| + \int_0^t c \exp(-\alpha(t - \mu))\|N\|\|x(\mu)\|d\mu
\]
By multiplying by \( \exp(\alpha t) \) on both sides, one obtains
\[
\exp(\alpha t)\|x(t)\| \leq c\|x(0)\| + \int_0^t c \exp(\alpha(t - \mu))\|N\|\|x(\mu)\|d\mu.
\]
From this and the Bellman-Gronwall Lemma there follows
\[
\exp(\alpha t)\|x(t)\| \leq c\|x(0)\| \exp\left(\int_0^t \|N\|d\mu\right)
\]
Since \( \int_0^t \|N\|d\mu = \|N\|c t \), it follows that \( \exp(\alpha t)\|x(t)\| \leq c\|x(0)\| \exp(\|N\|c t) \) and thus that
\[
\|x(t)\| \leq c\|x(0)\| \exp(\|N\|c - \alpha t)
\]
From this and (33) it follows that
\[
\|x(t)\| \leq c\|x(0)\| \exp(-\lambda t)
\]
This completes the proof. ■

**Proof of Theorem 2:** Recall that \( A_V(t) = \tilde{A} - gV'(I_{mn} - S(t))V \). By Proposition 2, for any fixed time \( \tau \), \( -V'(I_{mn} - S(t))V \) is exponentially stable if the graph of \( S(t') \) is strongly connected. Note \( \tilde{A} \) is fixed and bounded. According to Lemma 1, for each \( \sigma \in S_{\text{avg}} \) there is a positive number \( g \), depending on \( \tau_D \) so that

\[
\frac{\ln c + (\lambda + \|A\|c)\tau_D}{\lambda^*\tau_D}
\]
(36)
where \( c \) and \( \lambda^* \) are two positive numbers chosen so that for any fixed \( \tau \), \( \|\exp(-V'(I_{mn} - \bar{S}(\tau))V t)\| \leq c \exp(-\lambda^* t) \), and \( c > 1 \). This completes the proof. ■

**Proof of Theorem 3:** Recall \( \Phi_V(t, \tau) \) is the transition matrix of \( A_V(t) \) for any \( t \geq \tau \geq 0 \). If we can show that there exist a constant \( c \) so that
\[
\|\Phi_V(t, \tau)\| \leq c \exp(-\lambda(t - \tau)) \quad \forall t \geq \tau \geq 0
\]
the remaining proof is exactly the same as the proof of Theorem 1 which is omitted here.

It is left to show that \( \|\Phi_V(t, \tau)\| \leq c \exp(-\lambda(t - \tau)) \) for all \( t \geq \tau \geq 0 \) by choosing \( g \) sufficiently large. We explore the matrix \( A_V(t) \). Recall that \( A_V(t) = A - gV'(I_{mn} - S(t)) \otimes I_n \). In particular,
\[
(\lambda I + A_V(t)) + (\lambda + A_V(t))^t
\]
\[
= (\lambda I + \tilde{A}) + (\lambda + \tilde{A})^t
\]
\[
- gV'(2I_m - S(t) - S'(t)) \otimes I_n V
\]
Since each \( S(t) \) is doubly stochastic, \( 2I_m - S(t) - S'(t) \) has row sum 0, all its off-diagonal entries are non-positive, and all its diagonal entries are positive. That is, this matrix can be seen as a generalization of the Laplacian matrix of a connected graph. By Proposition 2, for any \( t \), \( -V'(2I_m - S(t) - S'(t)) \otimes I_n V = -V'(I_{mn} - S(t))V - V'(I_{mn} - S'(t))V \) is negative definite. Thus by picking \( g \) sufficiently large, \( (\lambda I + A_V(t)) + (\lambda + A_V(t))^t \) will be negative definite for any time \( t \).

Consider system
\[
\dot{z} = A_V(t)\tilde{z}
\]
Let \( V = \tilde{z}'\tilde{z} \). Then
\[
\dot{V} = \tilde{z}'(A_V(t)^t + A_V(t))\tilde{z} \leq -2\lambda\tilde{z}'\tilde{z}
\]
Therefore, \( \Phi_V(t, \tau) \) converges to zero as fast as \( \exp(-\lambda(t - \tau)) \) does, i.e.,
\[
\|\Phi_V(t, \tau)\| \leq c \exp(-\lambda(t - \tau)) \quad \forall t \geq \tau \geq 0
\]
This completes the proof. ■

**Proof of Theorem 4:** Equation (13) can be rewritten as
\[
\dot{y}_i = |V_i'W_i e|_2^2, \quad i \in m
\]
(37)
where $W_i = [W_{i1} \ldots W_{in}] \in \mathbb{R}^{n \times m}$. Here $W_{ij} \in \mathbb{R}^{n \times n}$ is $\frac{1}{m_i} I_n$ if $j \neq i$ and $j \in \mathcal{N}_i$, and $W_{ij}$ is $-I_n$ if $j = i$. Otherwise $W_{ij}$ is a $0$ matrix. Let column $(W_1, W_2, \ldots, W_m) = W$.

Different from (5), the error model turns to

$$\dot{e} = (A - G(t)P)(I_{mn} - S(t))e$$

where $G(t) = \text{block diag} \{g_1(t)I_n, \ldots, g_m(t)I_n\}$. Let $[(Qe')^T] = T^{-1}e$ where $T = [Q^{-1} V]$ as defined earlier. Here $z = [z'_1, \ldots, z'_m]'$ with $z_i = V'_i e_i$.

Based on (11) and (38), the dynamic of $z_i$ can be written in the following form

$$\dot{z}_i = \tilde{A}_iz_i - g_i(t)V'_i M e + \tilde{A}_i Q e_i, \quad i \in m$$

where $\tilde{A}_i = V'_i (A + K_i C_i) V_i$, and $\tilde{A}_i = V'_i (A + K_i C_i) Q_i$.

First, we want to show that all $g_i(t)$ are bounded. We prove this by contradiction. Without generality, suppose that $g_i$ for $i \in \mathcal{V}_u = \{1, 2, \ldots, m_1\}$ are unbounded, and $g_i$ for $i \in \mathcal{V}_b = \{m_1 + 1, m_1 + 2, \ldots, m\}$ are bounded where $\mathcal{V}_u \cap \mathcal{V}_b = 0$ and $\mathcal{V}_u \cup \mathcal{V}_b = m$.

Let $R = R_1 + R_2 + R_3 + R_4$ where the individual $R_i$ involve new positive parameters $p$, $\alpha_0, \alpha_{m_1+1}, \ldots, \alpha_m$.

$$R_1 = \frac{1}{2} \sum_{i=1}^{m_1} \pi_i \frac{p}{g_i(t)} |z_i|^2, \quad R_2 = \frac{1}{2} \sum_{i=m_1+1}^{m} \pi_i \frac{g_i(0)}{g_i(t)} |z_i|^2$$

$$R_3 = - \sum_{i=m_1+1}^{m} \alpha_i g_i(t), \quad R_4 = - \alpha_0 \int_0^t |Qe|^2 dt$$

The way to pick positive parameters $p$, $\alpha_0$, and $\alpha_i$ for $i \in \mathcal{V}_b$ is specified as follows.

**Picking $p \geq 1$:**

Let $\mathcal{W}_1$ be a positive matrix matrix chosen such that

$$z' W_1 z = \sum_{i=1}^{m} \pi_i |\tilde{A}_i| \sum_{i=1}^{m} |z_i|^2$$

According to (29),

$$F = \frac{1}{2} (HV' (\tilde{S} - I_{mn})V + V' (\tilde{S} - I_{mn})' VH) > 0$$

Pick $p$ so that $W_2 = pF - W_1 > 0$

**Picking $\alpha_0$, and $\alpha_i$, $i \in \mathcal{V}_b$:**

Using the Cauchy-Schwarz inequality, the following three inequalities can be derived. For $\beta_1 > 0$, $\beta_2 > 0$ and $\lambda_i > 0$ for $i \in \mathcal{V}_b$, all for the moment otherwise arbitrary, write $\tilde{A} = \text{block diag} \{A_1, \ldots, A_m\}$

$$z'H \tilde{A} Q e \leq \frac{\beta_1}{2} |\tilde{A}' H z|^2 + \frac{1}{2\beta_2} |Qe|^2,$$

$$-pz'H(\tilde{S} - I_{mn})Qe \leq \frac{\beta_2}{2} Q(\tilde{S} - I_{mn})' VH z_i^2 + \frac{1}{2\beta_2} |Qe|^2$$

$$-\pi_i (g_i(0) - p) z_i^2 \leq \frac{\lambda_i}{2} |\pi_i (g_i(0) - p) z_i|^2 + \frac{1}{2\lambda_i} |V_i' W_i e|^2.$$
It can be observed that $z'HV'(\bar{S} - I_{mn})e = z'HV'(\bar{S} - I_{mn})(VVe' + Qe)e = z'HV'(\bar{S} - I_{mn})Vz + z'HV'(\bar{S} - I_{mn})Qe$. From this and (41), $z'HV'(\bar{S} - I_{m})e = z'Fz + z'HV'((\bar{S} - I_{m}) \otimes I_{n})e$. Thus

$$
\dot{R}_1 + \dot{R}_2 \leq z'(\mathcal{W}_1 - pF)e + z'HAQe - pz'HV'
$$

$$(\bar{S} - I_{mn})Qe - \sum_{i=m+1}^m \pi_i(g_i(0) - p)z_i'Ve_ie$$

$$=-z'\mathcal{W}_2z + z'H \dot{A}Qe - pz'HV'$$

$$(\bar{S} - I_{mn})Qe - \sum_{i=m+1}^m \pi_i(g_i(0) - p)z_i'Ve_ie$$

From this, (42), (43), and (44),

$$\dot{R}_1 + \dot{R}_2 \leq z'\mathcal{W}_2z + \frac{\beta_1}{2}|A'Hz|^2 + \frac{\beta_2}{2}|Q(\bar{S} - I_{mn})'VHz|^2$$

$$+ \sum_{i=m+1}^m \frac{\lambda_i}{2}|\pi_i(g_i(0) - p)z_i|^2 + \frac{1}{2\beta_1}|Qe|^2$$

$$+ \frac{1}{2\beta_2}|Qe|^2 + \sum_{i=m+1}^m \frac{1}{2\beta_1}|Ve_ie|^2$$

(49)

It is direct to get that

$$\dot{R}_3 + \dot{R}_4 = -\sum_{i=m+1}^m \alpha_i|Ve_ie|^2 - \alpha_0|Qe|^2$$

(50)

From (49), (50), and (46),

$$\dot{R} \leq -z'\mathcal{W}_3z < 0$$

(51)

Since $Qe$ is exponentially convergent, the limit of $R_4(t)$ as $t$ goes to infinity exists. Due to the assumption that for $i \in \mathcal{V}_q$, the $g_i$ are bounded. Thus $R_3$ is bounded. Therefore, $R$ is lower bounded. From this and (51) it follows that $z \in \mathcal{L}^2$. From this and the fact that $Qe \in \mathcal{L}^2$, we conclude $e \in \mathcal{L}^2$. This with the definition of $\hat{g}_i$ imply that all $g_i$ for $i \in m$ are bounded. Thus by contradiction all $g_i$ are bounded.

Next, we want to show that $e$ converges to zero. According to Theorem 1, let $G_1 = g_1I$ be a matrix chosen so that $\bar{A} - G_1P(I_{mn} - \bar{S}(t))$ is a stable matrix. From (38), the dynamic of $e$ can be rewritten as

$$\dot{e} = (\bar{A} - G_1P(I_{mn} - \bar{S}(t)))e + (G_1 - G)P(I_{mn} - \bar{S}(t))e$$

(52)

Since $(I_{mn} - \bar{S}(t))e \in \mathcal{L}^2$, $G_1 - G$ is bounded, and $\bar{A} - G_1P(I_{mn} - S(t))$ is stable, thus $\dot{e} \in \mathcal{L}^2$. Thus (52) is input-to-state stable which implies that $e$ must converge to zero asymptotically.

**Appendix B Proofs for Discrete-Time Distributed Estimator**

**Proof of Lemma 2:** Since $M$ is an $m \times m$ row stochastic matrix which also is a strongly connected graph, $M$ is irreducible {Theorem 6.2.24, [31]}. Thus by the Perron-Frobenius Theorem there must be a positive vector $\pi$ such that $M^t\pi = \pi$. Without loss of generality, assume $\pi$ is normalized so that the sum of its entries equals 1; i.e., $\pi$ is a probability vector. Let $\Pi_M$ be that diagonal matrix whose diagonal entries are the entries of $\pi$. Then $\Pi_M^M = \pi$. Since $M^t1 = 1, \Pi_M^M = \pi$, and $M^t\pi = \pi$, it must be true that $M^t\Pi_M^M = \pi$ and thus that $L_M^M1 = 0$. Thus $L_M$ can also be written as $L_M = D - A$ where $D$ is a diagonal matrix whose diagonal entries are the diagonal entries of $L_M$ and $A$ is the nonnegative matrix $A = D - L_M$. Arguing as in the proof of Proposition 2, it can be shown that $L_M$ is positive-semidefinite.

Now suppose that the diagonal entries of $M$ are all positive. Then the diagonal entries of $M^t\Pi_M$ must also all be positive. It follows that every arc in the graph of $M$ must be an arc in the graph of $M^t\Pi_M$ which goes to infinity exists. To the assumption that for $i \in \mathcal{V}_0$, the $g_i$ are bounded. Thus $R_3$ is bounded. Therefore, $R$ is lower bounded. From this and (51) it follows that $z \in \mathcal{L}^2$. From this and the fact that $Qe \in \mathcal{L}^2$, we conclude $e \in \mathcal{L}^2$. This with the definition of $\hat{g}_i$ imply that all $g_i$ for $i \in m$ are bounded. Thus by contradiction all $g_i$ are bounded.

Write $R$ for $R(t)$. To prove the proposition it is enough to show that the matrix

$$Q = R - (V'SV)R(V'SV)$$

(53)

is positive definite.
To proceed, set \( \bar{L} = L \otimes I_n \) in which case \( \bar{L} \) is positive semi-definite because \( L \) is. Moreover, \( \bar{L} = \Pi - S^T \Pi S \) where \( \Pi = \Pi_S \otimes I_n \). Note that \( V' R V' = \bar{P} \bar{R} \) where \( \bar{P} \) is the orthogonal projection matrix \( P = V V' \). Clearly \( V' R V' = \bar{P} \bar{R} \bar{P} \bar{R} \). Note that both \( \bar{P} \) and \( \bar{R} \bar{P} \) are block diagonal matrices with corresponding diagonal blocks of the same size. Because of this and the fact that each diagonal block in \( \bar{P} \pi \) is a scalar times the identity matrix, it must be true that \( \bar{P} \) and \( \bar{R} \bar{P} \) commute; thus \( \bar{P} \bar{R} \bar{P} = \bar{P} \bar{R} \bar{P} \). From this and the fact that \( \bar{P} \) is idempotent, it follows that \( V' R V' = \bar{P} \bar{R} \). Clearly \( \bar{P} \bar{R} \bar{P} \bar{R} = \bar{P} \bar{R} \bar{P} \) so \( V' R V' \leq \bar{P} \). It follows using (53) that

\[
Q \geq V' \bar{L} V
\]

(54)

In view of this, to complete the proof it is enough to show that \( V' \bar{L} V \) is positive definite. This can be shown by the same proof of Proposition 2.

Therefore \( Q \) is positive definite because of (54). From this and (53) it follows that (27) is true.

Proof of Theorem 5: First it will be assumed that each \( \bar{K}_i \) has been selected so that the the matrix \( \bar{A}_V \) defined by (10), is such that \( \bar{A}_V \tau \rightarrow \infty \) as fast as \( \lambda^\tau \) does. This can be done using standard spectrum assignment techniques to make the spectral radius of \( \bar{A}_V \) at least as small as \( \lambda \). In view of (26), it is clear that to assign the convergence rate of the state transition matrix of \( \bar{A}(I_{mn} - P(I_{mn} - S(\tau))^q) \) it is necessary and sufficient to control the convergence rate of the state transition matrix of \( \bar{A}_V(\tau) \). This, as we will now show, can be accomplished by choosing \( q \) sufficiently large. We will actually detail two different ways to do this, each utilizing a different matrix norm. Both approaches will be explained next using the abbreviated notation \( B(\tau) = V' S(\tau) V \); note that with this simplification, \( \bar{A}_V(\tau) = \bar{A} \bar{B}^q(\tau) \) because of (25).

Weighted Two-Norm: For each fixed \( \tau \) and each appropriately-approximated matrix \( M \), write \( \|M\|_{R(\tau)} \) for the matrix norm induced by the vector norm \( \|x\|_{R(\tau)} \triangleq \sqrt{\tau R(\tau) x^T} \). Note that \( \|M\|_{R(\tau)} \) is the largest singular value of \( R^\frac{1}{2}(\tau) M R^\frac{1}{2}(\tau) \). Note in addition that

\[
(R^\frac{1}{2}(\tau) B(\tau) R^\frac{1}{2}(\tau))^T (R^\frac{1}{2}(\tau) B(\tau) R^\frac{1}{2}(\tau)) < I
\]

because of (27). This shows that the largest singular value of \( R^\frac{1}{2}(\tau) B(\tau) R^\frac{1}{2}(\tau) \) is less than one. Therefore

\[
\|B(\tau)\|_{R(\tau)} < 1
\]

(55)

a) \( N \) is constant

In this case both \( B(\tau) \) and \( R(\tau) \) are constant, so it is sufficient so choose \( q \) so that \( \|\bar{A} \bar{B}^q(\tau)\|_{R(\tau)} \leq \lambda \). Since \( \|\cdot\|_{R(\tau)} \) is submultiplicative, this can be done by choosing \( q \) so that

\[
\|B(\tau)\|_{R(\tau)}^q \leq \frac{\lambda}{\|A\|_{R(\tau)}}
\]

(56)

This can always be accomplished because of (55).

b) \( N \) changes with time

In this case it is not possible to use the weighted two-norm \( \|\cdot\|_{R(\tau)} \) because it is time-dependent. A simple fix, but perhaps not the most efficient one, would be to use the standard two-norm \( \|\cdot\|_2 \) instead since it does not depend on time. Using this approach, the first step would be to first choose, for each fixed \( \tau \), an integer \( p_1(\tau) \) large enough so that \( |B^{p_1(\tau)}/\tau|_2 < 1 \). Such values of \( p_1(\tau) \) must exist because each \( B(\tau) \) is a discrete-time stability matrix or equivalently, a matrix with a spectral radius less than 1. Computing such a value amounts to looking at the largest singular value of \( B(p_1(\tau)\tau) \) for successively largest values of \( p_1(\tau) \) until that singular value is less than 1. Having accomplished this, a number \( p \) can easily be computed so that \( |B^p(\tau)|_2 < 1 \forall \tau \) since there are only a finite number of distinct strongly connected graphs on \( m \) vertices and consequently only a finite number of distinct matrices \( B(\tau) \) in the set \( B = \{B(\tau) : \tau \geq 0\} \). Choosing \( p \) to be the maximum of the \( p_1(\tau) \) with respect to \( \tau \) is thus a finite computation. The next step would be to compute an integer \( \bar{p} \) large enough so that each \( |\bar{A} B^p(\tau)|_2 \leq \lambda \). A value of \( q \) with the required property would then be \( q = \bar{p} \).

Mixed Matrix Norm: There is a different way to choose \( q \) which does not make use of either Lemma 2 or Proposition 4. The approach exploits the “mixed matrix norm” introduced in [36]. To define this norm requires several steps. To begin, let \( |\cdot|_\infty \) denote the standard induced infinity norm and write \( \mathbb{R}^{mn \times mn} \) for the vector space of all \( m \times m \) block matrices \( M = [M_{ij}] \) whose \( ij \)th entry is a matrix \( M_{ij} \in \mathbb{R}^{n \times n} \). With \( n_i = \dim V_i \), \( i \in m \), and \( n = n_1 + n_2+ \cdots + n_m \), write \( \mathbb{R}^{m \times n} \) for the vector space of all \( m \times m \) block matrices \( M = [M_{ij}] \) whose \( ij \)th entry is a matrix \( M_{ij} \in \mathbb{R}^{n_i \times n_j} \). Similarly write \( \mathbb{R}^{n \times n} \) for the vector space of all \( m \times m \) block matrices \( M = [M_{ij}] \) whose \( ij \)th entry is a matrix \( M_{ij} \in \mathbb{R}^{n \times n} \).

Note that \( B \in \mathbb{R}^{mn \times mn} \), \( \bar{A} \in \mathbb{R}^{n \times n} \), \( V \in \mathbb{R}^{mn \times n} \), and \( V' \in \mathbb{R}^{m \times mn} \). For \( M \) in any one of these four spaces, the mixed matrix norm [36] of \( M \), written \( \|\cdot\|_M \), is

\[
\|M\| = \|\langle M \rangle\|_\infty
\]

(57)

where \( \langle M \rangle \) is the matrix in \( \mathbb{R}^{n \times m} \) whose \( ij \)th entry is \( \|M_{ij}\|_2 \). It is very easy to verify that \( \|\cdot\|_M \) is in fact
a norm. It is even sub-multiplicative whenever matrix multiplication is defined. Note in addition that $\|V\| = 1$ and $\|V'\| = 1$ because the columns of each $V_i$ form an orthonormal set.

Recall that $P = VV'$ is an orthogonal projection matrix. Using this, the definition of $B(\tau)$ and the fact that $PV = V$, it is easy to see that for any integer $p > 0$, $B_p(\tau) = V'(PS(\tau)P)pV$. Thus $\|B_p(\tau)\| \leq \|PS(\tau)P\|^p$. Using this and the fact that the graph of $S'$ is strongly connected, one can conclude that for $p \geq (m - 1)^2$, $\|PS(\tau)P\|^p < 1$. This is a direct consequence of Proposition 2 of [36]. Thus

$$\|B_p(\tau)\| < 1, \quad p \geq (m - 1)^2$$  \hspace{1cm} (58)

a) $\mathbb{N}$ is constant

In this case $B(\tau)$ is constant so it is sufficient to choose $q$ so that $\|AB_p(\tau)\| \leq \lambda$. This can be done by choosing $q = \overline{p}p$ where $p \geq (m - 1)^2$ and $\bar{p}$ is such that

$$\|B_p(\tau)\|^{\bar{p}(\tau)} \leq \frac{\lambda}{\|A\|}$$  \hspace{1cm} (59)

This can always be accomplished because of (58).

b) $\mathbb{N}$ changes with time

Note that (58) holds for all $\tau$. Assuming $p$ is chosen so that $p \geq (m - 1)^2$ it is thus possible to find, for each $\tau$, a positive integer $\bar{p}(\tau)$, for which

$$\|B_p(\tau)\|^{\bar{p}(\tau)} \leq \frac{\lambda}{\|A\|}$$  \hspace{1cm} (60)

Having accomplished this, a number $\bar{p}$ can easily be computed so that

$$\|B_p(\tau)\|^{\bar{p}} \leq \frac{\lambda}{\|A\|}$$  \hspace{1cm} (61)

holds for all $\tau$, since there are only a finite number of distinct strongly connected graphs on $m$ vertices and consequently only a finite number of distinct matrices $B(\tau)$ in the set $\mathcal{B}$ defined earlier. Choosing $\bar{p}$ to be the maximum of $\bar{p}(\tau)$ with respect to $\tau$ is thus a finite computation. A value of $q$ with the required property would then be $q = \overline{p}p$.

References

[1] L. Wang, J. Liu, and A. S. Morse. A distributed observer for a continuous-time linear system. In Proceedings of 2019 American Control Conference, pages 86–89, Philadelphia, PA, USA, July 2019.

[2] L. Wang, J. Liu, A. S. Morse, and B. D. O. Anderson. A Distributed Observer for a Discrete-Time Linear System. In Control and Decision conference, 2019.

[3] U. A. Khan and Ali Jadbabaie. On the stability and optimality of distributed Kalman filters with finite-time data fusion. In Proceedings of the 2011 American Control Conference, pages 3405–3410, 2011.

[4] R. Olfati-Reza and J. S. Shamma. Consensus filters for sensor networks and distributed sensor fusion. In Proc IEEE CDC, pages 6698–6703, 2005.

[5] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In Proceedings of the 4th International Conference on Information Processing in Sensor Networks, pages 63–70, 2005.

[6] R. Olfati-Reza. Kalman-consensus filter: Optimality, stability, and performance. In Proc IEEE CDC, pages 7036–7042, 2009.

[7] F. Dorfler, F. Pasqualetti, and F. Bullo. Continuous-time distributed observers with discrete communication. IEEE Journal of Selected Topics in Signal Processing, 7(2):296–304, 2013.

[8] Y. Li, S. Phillips, and R. G. Sanfelice. Robust distributed estimation for linear systems under intermittent information. IEEE Transactions on Automatic Control, 63(4):973–988, 2018.

[9] S. Park and N. C. Martins. Design of distributed LTI observers for state omniscience. IEEE Transactions on Automatic Control, pages 561–576, 2017.

[10] L. Wang and A. S. Morse. A distributed observer for a time-invariant linear system. IEEE Transactions on Automatic Control, 63(7):2123–2130, 2018.

[11] A. Mitra and S. Sundaram. Distributed observers for LTI systems. IEEE Transactions on Automatic Control, 63(11):3689–3704, 2018.

[12] T. Kim, H. Shim, and D. D. Cho. Distributed Luenberger observer design. In 2016 IEEE 55th Conference on Decision and Control (CDC), pages 6928–6933, Dec 2016.

[13] W. Han, H. L. Trentelman, Z. Wang, and Y. Shen. A simple approach to distributed observer design for linear systems. IEEE Transactions on Automatic Control, 64(1):329–336, 2019.

[14] W. Han, H. L. Trentelman, Z. Wang, and Y. Shen. Towards a minimal order distributed observer for linear systems. Systems and Control Letters, pages 59–65, 2018.

[15] L. Wang, A. S. Morse, D. Fullmer, and J. Liu. A hybrid observer for a distributed linear system with a changing neighborhood graph. In Proceedings of the 2017 IEEE Conference on Decision and Control, pages 1024–1029, 2017.

[16] R. Olfati-Saber. Distributed Kalman filter with embedded consensus filters. Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference, CDC-ECC ’05, pages 8179–8184, 2005.

[17] R. Olfati-Saber. Distributed Kalman filtering for sensor networks. In Proceedings of the 46th IEEE Conference on Decision and Control, pages 5492–5498, 2007.

[18] L. Wang, A. S. Morse, and J. Liu. A hybrid observer for estimating the state of a distributed linear system. Automatica, 146:110633, 2022.

[19] L. Wang and A. S. Morse. A distributed observer for a time-invariant linear system. In Proceedings of the 2017 American Control Conference, pages 2020–2025, 2017.

[20] T. Kim, C. Lee, and H. Shim. Completely decentralized design of distributed observer for linear systems. IEEE Transactions on Automatic Control, 65(11):4664–4678, Nov 2020.
[21] J. G. Lee and H. Shim. A distributed algorithm that finds almost best possible estimate under non-vanishing and time-varying measurement noise. *IEEE Control Systems Letters*, 4(1):229–234, 2020.

[22] M. Doostmohammadian and U. A. Khan. On the genericity properties in distributed estimation: Topology design and sensor placement. *IEEE Journal of Selected Topics in Signal Processing*, 7(2):195–204, 2013.

[23] U. A. Khan and A. Jadbabaie. Collaborative scalar-gain estimators for potentially unstable social dynamics with limited communication. *Autom.*, 50:1909–1914, 2014.

[24] S. Park and N. C. Martins. An augmented observer for the distributed estimation problem for LTI systems. In *Proceedings the 2012 American Control Conference*, pages 6775–6780, 2012.

[25] S. Park and N. C. Martins. Necessary and sufficient conditions for the stabilizability of a class of LTI distributed observers. In *Proceedings of the 51st IEEE Conference on Decision and Control*, pages 7431–7436, 2012.

[26] A. Mitra, J. A. Richards, S. Bagchi, and S. Sundaram. Finite-time distributed state estimation over time-varying graphs: Exploiting the age-of-information. *2019 American Control Conference (ACC)*, pages 4006–4011, 2019.

[27] B. Açıkmese, M. Mandić, and J. L. Speyer. Decentralized observers with consensus filters for distributed discrete-time linear systems. *Automatica*, 50(4):1037–1052, 2014.

[28] V. A. Ugrinovskii. Distributed robust estimation over randomly switching networks using $h_{\infty}$ consensus. *Automatica*, 49:160–168, 2013.

[29] F. C. Rego, Y. Pu, A. Alessandretti, A. P. Aguiar, A. M. Pascoal, and C. N. Jones. A distributed Luenberger observer for linear state feedback systems with quantized and rate-limited communications. *IEEE Transactions on Automatic Control*, 66(9):3922–3937, 2021.

[30] A. Mitra, J. A. Richards, S. Bagchi, and S. Sundaram. Distributed state estimation over time-varying graphs: Exploiting the age-of-information. *IEEE Transactions on Automatic Control*, 67(12):6349–6365, 2022.

[31] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, England, 1985.

[32] A. S. Morse. Dwell-time switching. In *Proceedings of the 2nd European Control Conference*, pages 176–181, 1993.

[33] J. P. Hespahna and A. S. Morse. Switched systems with average dwell time. In *Proceedings of the IEEE Conference on Decision and Control*, pages 2655–2660, 1999.

[34] B. D. O. Anderson and M. Gevers. *Fundamental Problems in Adaptive Control*, pages 9–21. Springer London, London, 1998.

[35] C. Godsil and G. Royle. *Algebraic Graph Theory*. Springer, 2001.

[36] S. Mou, J. Liu, and A. S. Morse. A distributed algorithm for solving a linear algebraic equation. *IEEE Transactions on Automatic Control*, 60(11):2863–2878, Nov 2015.
This figure "discretesfixerror.jpg" is available in "jpg" format from:

http://arxiv.org/ps/2210.01232v2
This figure "discretevaringerror.jpeg" is available in "jpeg" format from:

http://arxiv.org/ps/2210.01232v2
This figure "discretevaringerror.jpg" is available in "jpg" format from:

http://arxiv.org/ps/2210.01232v2
This figure "example.jpg" is available in "jpg" format from:

http://arxiv.org/ps/2210.01232v2
This figure "fix1.jpg" is available in "jpg" format from:

http://arxiv.org/ps/2210.01232v2
This figure "fixnorm.jpg" is available in "jpg" format from:

http://arxiv.org/ps/2210.01232v2
