Gossip and Distributed Kalman Filtering: Weak Consensus Under Weak Detectability

Soummya Kar, Member, IEEE, and José M. F. Moura, Fellow, IEEE

Abstract—The paper presents the gossip interactive Kalman filter (GIKF) for distributed Kalman filtering for networked systems and sensor networks, where intersensor communication and observations occur at the same time-scale. The communication among sensors is random; each sensor occasionally exchanges its filtering state information with a neighbor depending on the availability of the appropriate network link. We show that under a weak distributed detectability condition: 1) the GIKF error process remains stochastically bounded, irrespective of the instability of the random process dynamics; and 2) the network achieves weak consensus, i.e., the conditional estimation error covariance at a (uniformly) randomly selected sensor converges in distribution to a unique invariant measure on the space of positive semidefinite matrices (independent of the initial state). To prove these results, we interpret the filtering states (estimates and error covariances) at each node in the GIKF as stochastic particles with local interactions. We analyze the asymptotic properties of the error process by studying as a random dynamical system the associated switched (random) Riccati equation, the switching being dictated by a nonstationary Markov chain on the network graph.

Index Terms—Consensus, gossip, Kalman filter, random algebraic Riccati equation, random dynamical systems.

I. INTRODUCTION

A. Background and Motivation

This paper presents the Gossip Interactive Kalman Filtering (GIKF). GIKF is a linear distributed estimator that filters noisy observations of a random process measured by a sparsely connected sensor network. Each sensor observes only a portion of the process; acting alone, no sensor can resolve the signal. GIKF is fundamentally different from other distributed implementations of the Kalman filter ([1]–[4]) that consensus on the sensor observations or estimates in between new observations; in contrast, in GIKF the consensus and observation steps are at the same time scale. GIKF runs at each sensor a local copy of the Kalman filter. At random times, a sensor randomly selects a neighbor and the two sensors swap their states (their local Kalman filter state estimate and conditional error covariance), before processing the current observation. In other words, when communication occurs, a sensor updates the state it receives from its neighbor with its present observation; otherwise, it updates its own previous state. Such collaboration through state swapping is asynchronous and occurs occasionally, dictated by the random network topology. Indeed, due to environmental randomness, the medium access control (MAC) protocol is randomized, not known at the local sensor level. We assume that the sensor network uses a generic random communication protocol, see Section II, that subsumes the widely used gos-piping protocol for real time embedded architectures, [5] and the graph matching based communication protocols for internet architectures [6].

Distributed estimation problems of time varying state-space models like in Kalman filtering occur in many applications, for example, in large-scale physical infrastructures and social networks. A typical example is the dynamic state estimation problem in smart grids, the state of interest there being the phase angles governing the network power flow (a static and nonlinear version of this problem is considered in [7]). The linear dynamical signal model has been used recently in [8] to explain dynamics of social networking phenomena like technology adoption in groups ([9], [10]) and collaborative belief formation about employment opportunities in job markets ([11], [12]). The GIKF algorithm offers a potential approach for belief dissemination in these collaborative processing networks.

The paper establishes GIKF and studies its error properties. We define a weak distributed detectability condition\(^1\) under which we show 1) the GIKF error process remains stochastically bounded, irrespective of the instability of the random process dynamics and 2) the network achieves weak consensus, i.e., the conditional estimation error covariance at a (uniformly) randomly selected sensor converges in distribution to a unique invariant measure on the space of positive semi-definite matrices (independent of the initial state.) To prove these results, we interpret as stochastic particles, which we refer to as tokens or particles, the filtering states (state estimate and error covariance) at each node as they are swapped in the GIKF. We focus on these traveling (swapped) states to prove the stochastic boundedness and analyze the asymptotic properties of the error process and the network weak consensus and not on the sequence of conditional error covariances at each sensor, which is not Markov. This particle point of view is reminiscent of the approach taken in fluid dynamics of

\(^1\)This condition is required even by a centralized estimator (having access to all sensor observations over all time) to yield an estimate with bounded error (for unstable systems).
studying the transport of a particle as it travels in the fluid (Lagrangian coordinates) rather than studying the transport at a fixed coordinate in space (Eulerian coordinates), [13]. We show that the sequence of traveling states or particles evolves according to a switched system of random Riccati operators, the switching being dictated by a nonstationary Markov chain on the graph. A key contribution is the analysis of the resulting random Riccati equation (RRE). In this context, we note that the RRE arises in the literature in several practical filtering and control formulations with nonclassical information. Prior work ([14]–[24]) mostly address qualitative properties of the RRE in terms of moment stability, whereas recent approaches focus on understanding the limit behavior in terms of weak convergence ([25]–[28], see also [29], [30]). In this paper, we utilize a random dynamical systems formulation of the RRE; however, in contrast with our work in [25], [26], the switching sequence is no longer stationary. Several approximation arguments of independent interest are developed to tackle this nonstationary behavior and to establish the asymptotic distributional properties of the RRE. Another paper that deserves attention in this regard is [30]. In [30] the random Riccati equation (centralized) was studied with nonstationary system parameters. Under general mixing conditions on the matrices, tightness (stochastic boundedness) results were studied. Tightness of the covariance sequence implies the existence of subsequential weak limits by Prohorov’s theorem. That these subsequential limits are unique were stated under some invertibility and stability assumptions (see [30, Assumptions (A1) and (E4)]) of the signal dynamics.

To summarize, the paper considers distributed estimation (consensus and observation at the same time scale) of dynamical models. It introduces distributed observability and the question of minimal observation pattern (what should be the minimal number of sensors and what should they observe,) so that there exists a successful filtering scheme. Weak detectability (see Section II-A) is in terms of a full rank network Grammian. Weak detectability leads to stochastic boundedness of the conditional filtering error at each sensor, irrespective of observability of individual sensors. Robust information flow addresses the minimal communication that maintains consistent (asymptotically) information dissemination in the network, weak connectedness formulated in Section II-A quantifies the rate of information flow (in random communication environments) as the mixing time of a particle on a random walk in the network with appropriate statistics. The positive recurrence of this Markov chain translates to information dissemination at a sufficient rate to cope with the (possible) instability in signal dynamics and leads to weak consensus of the filtering errors. The notion of weak consensus introduced in the paper is the best form of consensus possible in such a setup because, as opposed to familiar scenarios (average computation/static parameter estimation,) in a dynamic situation it is not possible to accomplish almost sure (pathwise) consensus of the estimate or error processes. On the contrary, the weak consensus we establish shows that the error processes at different sensors converge in distribution to the same invariant measure. We do not characterize here this invariant measure as a function of the communication and observation policies; instead, we resolve the minimal conditions for the existence of such an invariant measure and hence conditions for the stability of the filtering error processes.

We present the rest of the paper. Section I-B sets up notation and background material to be used in the paper. Section II sets-up the problem and introduces the GIKF algorithm together with the observability and connectivity assumptions in Section II-A. An interactive particle interpretation and important preliminary results are in Section II-B. The main results regarding the asymptotic properties of the GIKF are stated (without proof) and interpreted in Section III. To prove these results, we provide first in Section IV a random dynamic system (RDS) formulation of the switching iterates of the random Riccati equation arising in the GIKF. Appendix A recalls facts and results on random dynamical systems (RDS) needed in this Section. The main results of the paper are proved in Section VI. Two technical Lemmas are proven in Appendix B. Finally, Section VII concludes the paper.

Notation and Preliminaries

Let \( \mathbb{R} \) be the reals; \( \mathbb{R}^M \), the \( M \)-dimensional Euclidean space; \( \mathbb{T} \), the integers; \( \mathbb{T}_+ \), the nonnegative integers; \( \mathbb{N} \), the natural numbers and \( \mathcal{X} \), a generic space. For \( \mathcal{B} \subset \mathcal{X} \), \( \| \cdot \| : \mathcal{X} \mapsto \{0,1\} \) is the indicator function, i.e., 1 when its argument is in \( \mathcal{B} \) and zero otherwise and \( \text{id}_\mathcal{X} \) is the identity function on \( \mathcal{X} \).

1) Cones in Partially Ordered Banach Spaces: We summarize facts and definitions on the structure of cones in partially ordered Banach spaces. Let \( V \) be a Banach space (over the field of the reals) with a closed (w.r.t. the Banach space norm) convex cone \( V_+ \) and assume \( V_+ \cap (-V_+) = \{0\} \). The cone \( V_+ \) induces a partial order in \( V \), namely, for \( X, Y \in V_+ \), we write \( X \leq Y \), if \( Y - X \in V_+ \). In case \( X \leq Y \) and \( X \neq Y \), we write \( X < Y \). The cone \( V_+ \) is called solid, if it has a nonempty interior in \( V_+ \); in that case, \( V_+ \) defines a strong ordering in \( V \) and we write \( X \ll Y \), if \( Y - X \in \text{int} \, V_+ \). The cone \( V_+ \) is normal if the norm \( \| \cdot \| \) is semimonotone, i.e., \( \exists c > 0 \), s.t. \( 0 \leq X \leq Y \Rightarrow \|X\| \leq c\|Y\| \). There are various equivalent characterizations of normality, of which we note that the normality of \( V_+ \) ensures that the topology in \( V \) induced by the Banach space norm is compatible with the ordering induced by \( V_+ \), in the sense that any norm-bounded set \( B \subset V \) is contained in a conic interval of the form \( [X,Y] \), where \( X, Y \in V \). Finally, a cone is said to be minihedral, if every order-bounded (both upper and lower bounded) finite set \( B \subset V \) has a supremum (here bounds are w.r.t. the partial order.) We focus on the separable Banach space of symmetric \( n \times n \) matrices, \( \mathbb{S}^n \), with the induced 2-norm. The subset \( \mathbb{S}^n_+ \) of positive semidefinite matrices is a closed, convex, solid, normal, minihedral cone in \( \mathbb{S}^n \), with nonempty interior \( \mathbb{S}^n_+ \), the set of positive definite matrices. These denote the partial and strong ordering in \( \mathbb{S}^n \) induced by \( \mathbb{S}^n_+ \).

2) Probability Measures on Metric Spaces: Let: \( (\mathcal{X}, d_\mathcal{X}) \) a complete separable metric space \( \mathcal{X} \) with metric \( d_\mathcal{X} \); \( \mathbb{B}(\mathcal{X}) \) its Borel algebra; \( \mathbb{B}(\mathcal{X}) \) the Banach space of real-valued bounded functions on \( \mathcal{X} \), with the sup-norm,
\[
f \in B(\mathcal{X}), \quad \|f\| = \sup_{x \in \mathcal{X}} |f(x)|	ext{ and } C_{\mu}(\mathcal{X}) \text{ the subspace of } B(\mathcal{X}) \text{ of continuous functions. For } x \in \mathcal{X}, \text{ the open ball of radius } \varepsilon > 0 \text{ centered at } x \text{ is denoted by } B_{\varepsilon}(x), \text{ i.e., } B_{\varepsilon}(x) = \{y \in \mathcal{X} | \|x - y\| < \varepsilon\}. \text{ For any set } \Gamma \subset \mathcal{X}, \text{ the open } \varepsilon\text{-neighborhood of } \Gamma \text{ is given by } \Gamma_{\varepsilon} = \{y \in \mathcal{X} | \|x - y\| < \varepsilon\}. \Gamma_{\varepsilon} \text{ is an open set.}

Let } \mathcal{P}(\mathcal{X}) \text{ be the set of probability measures on } \mathcal{X}. \text{ A sequence } \{\mu_t\}_{t \in \mathbb{T}_+} \text{ of probability measures in } \mathcal{P}(\mathcal{X}) \text{ converges weakly to } \mu \in \mathcal{P}(\mathcal{X}) \text{ if } \lim_{t \to \infty} \langle f, \mu_t \rangle = \langle f, \mu \rangle, \forall f \in C_{\mu}(\mathcal{X}). \text{ By Portmanteau’s theorem, the above is equivalent to any one of the following:}

1) \text{ for all closed } F \in \mathcal{B}(\mathcal{X}), \quad \lim_{t \to \infty} \sup_{t \in \mathbb{T}_+} \mu_t(F) 
\leq \mu(F);

2) \text{ for all open } O \in \mathcal{B}(\mathcal{X}), \quad \lim_{t \to \infty} \inf_{t \in \mathbb{T}_+} \mu_t(O) \geq \mu(O).

Weak convergence, } \mu_t \Rightarrow \mu, \text{ is also referred to as convergence in distribution. The weak topology on } \mathcal{P}(\mathcal{X}) \text{ generated by weak convergence can be metrized; one has the Prohorov metric } d_\mathcal{P} \text{ on } \mathcal{P}(\mathcal{X}), \text{ [31], such that the metric space } (\mathcal{P}(\mathcal{X}), d_\mathcal{P}) \text{ is complete, separable and a sequence } \{\mu_t\}_{t \in \mathbb{T}_+} \text{ in } \mathcal{P}(\mathcal{X}) \text{ converges weakly to } \mu \text{ in } \mathcal{P}(\mathcal{X}) \text{ iff } \lim_{t \to \infty} d_\mathcal{P}(\mu_t, \mu) = 0. \text{ The distance between two probability measures } \mu_1, \mu_2 \text{ in } \mathcal{P}(\mathcal{X}) \text{ is computed as:}

\[
d_\mathcal{P}(\mu_1, \mu_2) = \inf \{\varepsilon > 0 | \|\mu_1(F) - \mu_2(F)\|_\varepsilon + \varepsilon, \forall \text{ closed set } F\}.
\]

II. Gossip Interactive Kalman Filter (GIKF)

A. Problem Setup

1) Signal/Observation Model: Let the signal model and observation model at the } n \text{th sensor be}

\[
x_{t+1} = F x_t + w_t \tag{1}
\]

\[
v_{t}^{n} = C_n x_t + v_{t}^{n} \tag{2}
\]

where } x_t \in \mathbb{R}^m \text{ is the signal (state) vector with initial state } x_0 \text{ distributed as a zero mean Gaussian vector with covariance } \Sigma; \text{ the system noise } \{w_t\} \text{ is an uncorrelated zero mean Gaussian sequence independent of } x_0 \text{ with covariance } \Sigma; \text{ the } n \text{th sensor observation at time } t; \text{ there are } N \text{ sensors, where } C_n \in \mathbb{R}^{m_n \times m} \text{ and } \{v_{t}^{n}\} \text{ is an uncorrelated zero mean Gaussian observation noise sequence with covariance } R_n \gg 0. \text{ Also, the noise sequences at different sensors are independent of each other, the system noise process and the initial system state. Because of the limited capability of the sensors, typically the dimension of } v_{t}^{n} \text{ is much smaller than that of the signal process and the observation process at each sensor is not sufficient to make the pair } \{x_t, v_{t}^{n}\} \text{ observable. We envision a totally distributed application where a reliable estimate of the signal process is required at each sensor. The sensors achieve collaboration with each other by means of occasional communication with their neighbors, whereby they exchange their filtering states (to be defined precisely.) We assume that time is slotted and intersensor communication and sensing (observation) take place at the same time-scale.}

2) Communication Model: Communication among sensors is constrained by proximity, transmit power, or receiving capabilities. We model the underlying communication network in terms of an undirected graph } (V, E) \text{ where } V \text{ denotes the set of } N \text{ sensors and } E \text{ is the set of edges or allowable communication links between the sensors. The notation } n \sim l \text{ indicates that sensors } n \text{ and } l \text{ can communicate, i.e., } E \text{ contains the undirected edge } (n, l). \text{ The graph can be represented in terms of its } N \times N \text{ symmetric adjacency matrix } A:

\[
A_{nl} = \begin{cases} 1, & \text{if } (n, l) \in E \\ 0, & \text{otherwise} \end{cases}
\]

The diagonal elements of } A \text{ are identically 1-a sensor } n \text{ can always communicate to itself. Note, that } E \text{ is the maximal allowable set of links in the network at any time, however, at a particular instant, each sensor may choose to communicate only to a fraction of its neighbors. The exact communication protocol is not so important for the analysis, as long as some weak connectivity assumptions are satisfied. For definiteness, we assume the following generic communication model, which subsumes the widely used gossiping protocol for real time embedded architectures ([5]) and the graph matching based communication protocols for internet architectures ([6]). We make this precise in the following, which we generalize later. Define the set } \mathcal{M} \text{ of symmetric 0 and } 1 N \times N \text{ matrices}

\[
\mathcal{M} = \{A | 1^T A = 1^T, \quad 1 A = 1, \quad A \leq \mathcal{E}\}. \tag{3}
\]

In other words, } M \text{ is the set of adjacency matrices, such that, every node is incident to exactly one edge (including self edges) and allowable edges are only those included in } \mathcal{E}. \text{ Let } D \text{ be a probability distribution on the space } \mathcal{M}. \text{ We make the following assumption of connectivity on the average.}

Assumption (C.1): The sequence of time-varying adjacency matrices, } \{A(t)\}_{t \in \mathbb{N}}, \text{ governing the intersensor communication, is an i.i.d. sequence in } \mathcal{M} \text{ with distribution } D \text{ and independent of the signal and observation processes.} \text{ Also define the assumed to be irreducible and aperiodic symmetric stochastic matrix } \bar{A} \text{ as}

\[
\bar{A} = E[A(t)] = \int_{\mathcal{M}} A dD(A). \tag{4}
\]

Remark 1: The stochasticity of } \bar{A} \text{ is inherited from the elements of } \mathcal{M}. \text{ We are not concerned with the properties of the distribution } D \text{ as long as the weak connectivity assumption is satisfied. Matrix } \bar{A} \text{ being irreducible depends on the allowable edges } E \text{ and the distribution } D. \text{ We do not pursue this in detail here. To show the applicability of assumption (C.1) and justify the notion of weak connectivity, we note that such a distribution } D \text{ is a}

\footnote{The inequality } A \leq \mathcal{E} \text{ is to be interpreted component-wise in (3).}

\footnote{The set } \mathcal{M} \text{ is always nonempty, in particular the } N \times N \text{ identity matrix } I_N \in \mathcal{M}. \text{ For convenience of presentation, we assume that } A(0) = I_N, \text{ although communication starts at slot } t = 1. \text{ (6) is a}

\footnote{For convenience of presentation, we assume that } A(0) = I_N, \text{ although communication starts at slot } t = 1. \text{ (6) is a}

\footnote{The set } \mathcal{M} \text{ is always nonempty, in particular the } N \times N \text{ identity matrix } I_N \in \mathcal{M}. \text{ For convenience of presentation, we assume that } A(0) = I_N, \text{ although communication starts at slot } t = 1. \text{ The}
always exists if the graph \((V, E)\) is connected. We give a Markov chain interpretation of the mean adjacency matrix \(\bar{A}\), which will be helpful for the analysis to follow. The matrix \(\bar{A}\) can be associated to the transition kernel of a time-homogeneous Markov chain on the state space \(V\). Since the state space \(V\) is finite, the irreducibility of \(\bar{A}\) suggests that the resulting Markov chain is positive recurrent. Due to symmetry, the Markov chain is reversible with unique invariant distribution \(\pi\) on \(V\), where \(\pi\) is the discrete uniform distribution on \(V\).

We make some observations on the nature of the random intersensor communication process. First, the randomness may be due to the wireless protocol itself. For example, a gossip type of protocol often comes from a carrier sense multiple access (CSMA) scheme, where randomness in the topology is due to the random back-off carrier sensing times ([5]). In this work we allow more general random protocols, where the link formation is spatially correlated but temporally independent. Spatial correlation captures communication network phenomena like interference avoidance, whereas temporal independence captures the class of memoryless channels, often assumed in communication networks. Secondly, wireless networks operate under scarce communication resources in unpredictable environments. Randomness in the communication channel is often due to infrastructure failures.

3) Observability Conditions: Weak Detectability: Even in the centralized setting (when the sensors forward their observations at all time to a fusion center), filtering requires some form of detectability and stabilizability. In the present distributed setting we impose the following weak assumptions on the signal/observation model.

4) Stabilizability: Assumption (S.1): The pair \((\mathcal{F}, Q^{1/2})\) is stabilizable. The nondegeneracy (positive definiteness) of \(Q\) ensures this.

5) Weak Detectability: Assumption (D.1): There exists a walk\(^8\) of length \(\ell \geq 1\), \((n_1, n_2, \ldots, n_\ell)\), covering the \(N\) nodes, such that, the matrix \(\sum_{\ell=1}^{\ell} (\mathcal{F}^{-1})^T C_{\ell}^T C_{n_\ell} \mathcal{F}^{-1}\) is invertible.

Remark 2: Note, as permitted by the general definition of a walk, the sequence \((n_1, n_2, \ldots, n_\ell)\) may consist of repeated vertices and in particular, self-loops (if permitted by \(\bar{A}\)).

Remark 3: When \(\mathcal{F}\) is invertible, (D.1) may be replaced by the full rank of

\[ G = \sum_{n=1}^{N} C_{n}^T C_{n}, \]

indeed, by the irreducibility of \(\bar{A}\) (equivalently, by the connectivity of the graph induced by \(\bar{A}\)) we can find a walk \((n_1, n_2, \ldots, n_\ell)\) of length \(\ell \geq N\), which covers the network, i.e., visits each node at least once. Hence, if \(\mathcal{F}\) is invertible and (5) holds, it follows that the matrix \(\sum_{\ell=1}^{\ell} (\mathcal{F}^{-1})^T C_{\ell}^T C_{n_\ell} \mathcal{F}^{-1}\) corresponding to this walk is invertible leading to Assumption (D.1).

Remark 4: From the positive definiteness of the measurement noise matrices \(R_{n_\ell}\), it follows that under (D.1), the matrix \(\sum_{\ell=1}^{\ell} (\mathcal{F}^{-1})^T C_{\ell}^T R_{n_\ell}^{-1} C_{n_\ell} \mathcal{F}^{-1}\) is invertible.

Remark 5: Assumption (D.1) is minimal, in the sense, that, even in a centralized setting (a center has access to all the sensor observations over all time,) it is required to ensure detectability for arbitrary choice of the matrix \(\mathcal{F}\) governing the signal dynamics. This justifies the nomenclature weak detectability.

Algorithm GIKF: We now present the algorithm GIKF (gossip based interacting Kalman filter) for distributed estimation of the signal process \(x_t\) over time. We start by introducing notation. Let the filter at sensor \(n\) be initialized with the pair \((\hat{x}_{0|0|n}, \hat{P}_{0}^n)\), where \(\hat{x}_{0|0|n}\) denotes the prior estimate of \(x_0\) (with no observation information) and \(\hat{P}_0^n\) the corresponding error covariance. Also, by \((\hat{x}_{t|t-1|n}, \hat{P}_t^n)\) denote the estimate at sensor \(n\) of \(x_t\) based on information\(^9\) till time \(t - 1\) and the corresponding conditional error covariance, respectively. The pair \((\hat{x}_{t|t-1|n}, \hat{P}_t^n)\) is also referred to as the state of sensor \(n\) at time \(t - 1\). To define the estimate update rule for the GIKF, let \(n_t\) be the neighbor of sensor \(n\) at time \(t\) w.r.t. the adjacency matrix\(^10\) \(A(t)\). We assume that all intersensor communication for time \(t\) occurs at the beginning of the slot, whereby communicating sensors swap their previous states, i.e., if at time \(t\), \(n_t\) is expanded, sensor \(n\) replaces its previous state \((\hat{x}_{t|t-1|n}, \hat{P}_t^n)\) by \((\hat{x}_{t|t-1|n}, \hat{P}_t^n)\) and sensor \(l\) replaces its previous state \((\hat{x}_{t|t-1|l}, \hat{P}_t^n)\) by \((\hat{x}_{t|t-1|n}, \hat{P}_t^n)\). The estimate update at sensor \(n\) at the end of the slot (after the communication and observation tasks have been completed) is shown in (6) and (7) at the bottom of the page. Due to conditional Gaussianity, the filtering steps above can be implemented through the time-varying Kalman filter recursions and it follows that

\[ \hat{x}_{t+1|t} = E \left[ x_{t+1} \Big| \hat{x}_{t+1|t}, \hat{P}_{t}^{n_t}; n_t; Y_t \right] \]

\[ \hat{P}_{t+1}^n = E \left[ (x_{t+1} - \hat{x}_{t+1|t}) (x_{t+1} - \hat{x}_{t+1|t})^T \right| \hat{x}_{t|t-1}; \hat{P}_t^{n_t}; n_t; Y_t \].

\[ \hat{x}_{t+1|t} = E \left[ x_{t+1} \Big| \hat{x}_{t+1|t}, \hat{P}_{t}^{n_t}; n_t; Y_t \right] \]

\[ \hat{P}_{t+1}^n = E \left[ (x_{t+1} - \hat{x}_{t+1|t}) (x_{t+1} - \hat{x}_{t+1|t})^T \right| \hat{x}_{t|t-1}; \hat{P}_t^{n_t}; n_t; Y_t \]
the sequence \( \{\tilde{P}_t^n\} \) of conditional predicted error covariance matrices at sensor \( n \) satisfies the Riccati recursion:

\[
\tilde{P}_{t+1}^n = \mathcal{F} \tilde{P}_t^n \mathcal{F}^T + Q - \mathcal{F} \tilde{P}_t^n C_n^T (C_n \tilde{P}_t^n C_n^T + R_n)^{-1} C_n \tilde{P}_t^n \mathcal{F}^T.
\]

(8)

Note that the sequence \( \{\tilde{P}_t^n\} \) is random, due to the random neighborhood selection function \( n^\rightarrow \). The goal of the paper is to study asymptotic properties of the sequence of random conditional error covariance matrices \( \{\tilde{P}_t^n\} \) at every sensor \( n \) and show in what sense they reach consensus, so that, in the limit of large time, every sensor provides an equally good (stable in the sense of estimation error) estimate of the signal process.

**B. An Interacting Particle Representation**

To compactify the notation in (8), we define the functions \( f_n : \mathcal{S}_n^\rightarrow \rightarrow \mathcal{S}_n^\rightarrow \) for \( n = 1, \ldots, N \) defining the respective Riccati operators:

\[
f_n(X) = \mathcal{F} X \mathcal{F}^T + Q - \mathcal{F} X C_n^T (C_n X C_n^T + R_n)^{-1} C_n X \mathcal{F}^T.
\]

(9)

Recall the sequence \( \{n^\rightarrow_t\} \) of neighborhoods of sensor \( n \). The sequence of conditional error covariance matrices \( \{\tilde{P}_t^n\} \) at sensor \( n \) then evolves according to

\[
\tilde{P}_{t+1}^n = f_n \left( \tilde{P}_t^n \right).
\]

(10)

The sequence \( \{\tilde{P}_t^n\} \) is non-Markov (not even semi-Markov given the random adjacency matrix sequence \( \{A(t)\} \)) as \( \tilde{P}_{t+1}^n \) at time \( t \) is a random functional of the conditional error covariance at time \( t - 1 \) of the sensor \( n^\rightarrow \), which, in general, is different from sensor \( n \). This makes the evolution of the sequence \( \{\tilde{P}_t^n\} \) difficult to track. To overcome this, we give the following interacting particle interpretation of the conditional error covariance evolution, which naturally leads us to track semi-Markov sequences of conditional error covariance matrices from which we can completely characterize the evolution of the desired covariance sequences \( \{\tilde{P}_t^n\} \) for \( n = 1, \ldots, N \).

To this end, we note that the link formation process given by the sequence \( \{A(t)\} \) can be represented in terms of \( N \) particles moving on the graph as identical Markov chains. The state of the \( n \)th particle is denoted by \( z_n(t) \) and the sequence \( \{z_n(t)\} \) takes values in \( \{1, \ldots, N\} \). The evolution of the \( n \)th particle is given as follows:

\[
z_n(t) = (z_n(t - 1))^\rightarrow, \quad z_n(0) = n.
\]

(11)

Recall the (random) neighborhood selection \( n^\rightarrow \). Thus, the \( n \)th particle can be viewed as originating from node \( n \) at time 0 and then traveling on the graph (possibly changing its location at each time) according to the link formation process \( \{A(t)\} \). The following proposition establishes important statistical properties of the sequence \( \{z_n(t)\} \) (see Appendix B for a proof).

**Proposition 6:**

1. For each \( n \), the process \( \{z_n(t)\} \) is a Markov chain on \( V = \{1, \ldots, N\} \) with transition probability matrix \( \Gamma \).
2. The Markov chain \( \{z_n(t)\} \) is ergodic with the uniform distribution on \( V \) being the attracting invariant measure.

For each of the Markov chains \( \{z_n(t)\} \), we define a sequence of switched Riccati iterates \( \{P_n(t)\} \):

\[
P_n(t + 1) = f_n(z_n(t))(P_n(t)).
\]

(12)

The sequence \( \{P_n(t)\} \) can be viewed as an iterated system of Riccati maps, the random switching sequence being governed by the Markov chain \( \{z_n(t)\} \). A more intuitive explanation comes from the particle interpretation, precisely the \( n \)th sequence may be viewed as a particle originating at node \( n \) and hopping around the network as a Markov chain with transition probability \( A \) whose instantaneous state \( P_n(t) \) evolves by the application of the Riccati operator corresponding to its current location. In particular, in contrary to the sequence of conditional error covariances at sensor \( n \), the sequence \( \{\tilde{P}_n(t)\} \) does not correspond to the error evolution at a particular sensor. The following proposition shows that the sequence \( \{P_n(t)\} \) is semi-Markov and establishes its relation to the sequence \( \{\tilde{P}_n(t)\} \) of interest.

**Proposition 7:**

1. The sequence \( \{P_n(t)\} \) is semi-Markov, given the Markov switching sequence, i.e.,

\[
E \left[ \Gamma (P_n(t + 1)) \right| \{P_n(s), z_n(s)\}_{0 \leq s \leq t} = E \left[ \Gamma (P_n(t + 1)) \right| \{P_n(s), z_n(s)\}, \forall t \in T_+, \Gamma \in \mathcal{B}(\mathcal{S}_N^\rightarrow) \]

(13)

2. Consider the sequence of random permutations \( \{\pi_t\} \) on \( V \), given by

\[
(\pi_{t+1}(1), \ldots, \pi_{t+1}(N)) = \left( (\pi_1(1))^\rightarrow, \ldots, (\pi_t(N))^\rightarrow \right)
\]

(14)

with initial condition

\[
(\pi_0(1), \ldots, \pi_0(N)) = (1, \ldots, N).
\]

(15)

(Note that \( \pi_t(n) := z_n(t) \) for every \( n \), where \( z_n(t) \) is defined in (11).) Then, for \( t \in T_+ \),

\[
(\tilde{P}_1(t + 1), \ldots, \tilde{P}_N(t + 1)) = \left( \tilde{P}_{\pi_1(t)}(t + 1), \ldots, \tilde{P}_{\pi_t(N)}(t + 1) \right).
\]

(16)

Part 2) of the above proposition suggests that the asymptotics of the desired sequence \( \{\tilde{P}_n(t)\} \) for every \( n \) can be obtained by studying the same for the sequences \( \{P_n(t)\} \). Also, part 1) of Proposition 7 demonstrates the nice structure of the sequence \( \{P_n(t)\} \). In the following, in particular, we will show that the sequences \( \{P_n(t)\} \) reach consensus in a weak sense, which by part 2) will establish weak consensus for the sequences \( \{\tilde{P}_n(t)\} \) of interest. Hence, in the subsequent sections, we will study the sequences \( \{P_n(t)\} \).
rather than working directly on the sequences \( \{ \hat{P}_n(t) \}_{t \in \mathbb{T}^+} \) of interest, which involve a much more complicated statistical dependence.

### III. Main Results

In this section, we present and discuss the main results of the paper under Assumptions (C.1), (S.1), (D.1). The first result does not directly concern the sequences \( \{ \hat{P}_n(t) \} \) for \( n = 1, \ldots, N \), but sets the stage for presenting the key result regarding the convergence of these sequences and is of independent interest.

**Theorem 8:** For a given \( \mathbb{A} \), let \( \{ \tilde{Z}(t) \}_{t \in \mathbb{T}^+} \) be a stationary Markov chain on \( V \) with transition probability matrix \( \mathbb{A} \), i.e., \( \tilde{Z}(0) \) is distributed uniformly on \( V \). Let \( \nu \) be a probability measure on \( \mathbb{P}_N^V \) and consider the random process \( \{ \tilde{P}(t) \}_{t \in \mathbb{T}^+} \) given by

\[
\tilde{P}(t+1) = f_{\tilde{Z}(t)}(\tilde{P}(t)), \quad t \in \mathbb{T}^+.
\]

(17)

where \( \tilde{P}(0) \) is distributed as \( \nu \) and independent of the Markov chain \( \{ \tilde{Z}(t) \} \). Then, there exists a probability measure (unique) \( \mu^\mathbb{A} \) (depending on \( \mathbb{A} \) only and supported on \( \mathbb{P}_N^V \)) such that, for every \( \nu \), the process \( \{ \tilde{P}(t) \} \) constructed above converges weakly to \( \mu^\mathbb{A} \). In other words, for any \( \nu \) if \( \tilde{P}(0) \sim \nu \) and independent of \( \{ \tilde{Z}(t) \} \), we have as \( t \to \infty \) that the composition of Riccati operators converges in distribution

\[
f_{\tilde{Z}(t)} \circ f_{\tilde{Z}(t-1)} \cdots \circ f_{\tilde{Z}(0)}(\tilde{P}(0)) \to \mu^\mathbb{A}.
\]

(18)

**Remark 9:** We stress here that the dependence of the invariant measure \( \mu^\mathbb{A} \) on the communication policy \( \mathbb{D} \) manifests only through the mean matrix \( \mathbb{A} \). We add some remarks on the weak convergence rate of the sequence \( \{ \tilde{P}(t) \} \) to its stationary measure \( \mu^\mathbb{A} \). Although we do not explicitly characterize in this work, the convergence rate of the iterates to the invariant measure \( \mu^\mathbb{A} \) will be geometric in the Prohorov metric. This follows from some generic results in RDS theory. As a matter of fact, any monotone, sublinear RDS (the Riccati RDS (Section IV) satisfies these properties, Lemma 11) can be shown to be a contractive system with respect to the Birkhoff or part metric (for details on this, please refer to the monograph [32]). In this sense, we establish exponential ergodicity of the sequence \( \{ \tilde{P}(t) \} \). However, explicit calculation of this geometric factor would require more analysis and due to space limitations and digression from the main focus we omit it in this paper.

We now state the key result characterizing the convergence properties of the sequences \( \{ \hat{P}_n(t) \} \).

**Theorem 10:**
1) Let \( q \) be a uniformly distributed random variable on \( V \) and independent of the sequence of adjacency matrices \( \{ A(t) \}_{t \in \mathbb{T}^+} \). Then, the sequence \( \{ \hat{P}_q(t) \}_{t \in \mathbb{T}^+} \) converges weakly to \( \mu^\mathbb{A} \) (the latter being defined in Theorem 8), i.e.,

\[
\hat{P}_q(t) \Rightarrow \mu^\mathbb{A}.
\]

(19)

In other words, the conditional error covariance \( \{ \hat{P}_q(t) \} \) of any randomly selected sensor (estimator) converges in distribution to \( \mu^\mathbb{A} \).

2) For every \( n \in \{ 1, \ldots, N \} \), the sequence \( \{ \hat{P}_n(t) \}_{t \in \mathbb{T}^+} \) (or the sequence \( \{ \hat{P}_{\pi(n)}(t) \}_{t \in \mathbb{T}^+} \)) is stochastically dominated by the distribution \( \mu^\mathbb{A} \) as \( t \to \infty \), i.e., for every \( \alpha > 0 \), we have

\[
\limsup_{t \to \infty} \mathbb{P}(\| \hat{P}_n(t) \| \geq \alpha) \leq \mu^\mathbb{A}(\{ X \in \mathbb{S}_+^N \| X \| \geq \alpha \})
\]

(20)

\[
\limsup_{t \to \infty} \mathbb{P}(\hat{P}_n(t) \geq \alpha I) \leq \mu^\mathbb{A}(\{ X \in \mathbb{S}_+^N | X \geq \alpha I \}).
\]

(21)

More generally, for a closed set \( F \) preserving monotonicity, i.e., \( X \in F \) implies \( Y \in F \) for all \( Y \geq X \), we have

\[
\limsup_{t \to \infty} \mathbb{P}(\hat{P}_n(t) \in F) \leq \mu^\mathbb{A}(F).
\]

(22)

In words, \( \forall n \), the pathwise error associated with \( \{ \hat{S}_{\pi(n)}(t) \} \) is stochastically dominated by \( \mu^\mathbb{A} \).

3) For each \( n \), the sequence of error covariances \( \{ \hat{P}_n(t) \}_{t \in \mathbb{T}^+} \) is stochastically bounded,

\[
\lim_{t \to \infty} \sup_{t \in \mathbb{T}^+} \mathbb{P}(\| \hat{P}_n(t) \| \geq J) = 0.
\]

(23)

Specifically, for all closed \( F \), we have

\[
\limsup_{t \to \infty} \mathbb{P}(\hat{P}_n(t) \in F) \leq N \mu^\mathbb{A}(F).
\]

(24)

**Remark 11:** We discuss the consequences of Theorem 10. The first part of the theorem reinforces the weak consensus achieved by the GIKF algorithm, i.e., the conditional error covariance at a randomly selected sensor converges in distribution to the invariant measure \( \mu^\mathbb{A} \). Reinterpreted, it provides an estimate \( \{ \tilde{S}_q(t) \} \) (in practice, obtained by uniformly selecting a sensor \( q \) independent of the random gossip protocol \( \{ A(t) \} \) and using its estimate \( \tilde{S}_q(t) \) for all time \( t \) with stochastically bounded conditional error covariance under the weak detectability and connectivity assumptions. Note that the results provided in this paper pertain to the limiting distribution of the conditional error covariance and hence, the pathwise filtering error. This is a much stronger result than providing moment estimates of the conditional error covariance, which does not provide much insight into the pathwise instantiations of the filter. In this paper, we do not provide analytic characterizations of the resulting invariant measure \( \mu^\mathbb{A} \). However, Theorem 8 also provides an efficient numerical characterization of \( \mu^\mathbb{A} \). In particular, the weak convergence in (18) shows that the empirical distribution obtained by plotting repeated instantiations of the process \( \{ \tilde{P}(t) \} \) (see (17)) would converge to \( \mu^\mathbb{A} \).

Another class of estimates obtained by the GIKF algorithm is demonstrated in the second part of Theorem 10. For each \( n \), the estimate \( \{ \tilde{S}_{\pi(n)}(t) \} \) is obtained in practice by starting at the node \( n \) and then performing a random walk, \( \pi(n) \), through
the graph and collecting the estimates on the way. Equations (17)–(22) show that, in the limit as $t \to \infty$, these estimates are at least as good as the estimate $\{\hat{X}(t)\}$ obtained by probing a randomly selected node and using its estimate throughout. For some $\pi$, whether the estimate $\{\hat{X}(\pi(t))\}$ is strictly better than the estimate $\{\hat{X}(t)\}$ asymptotically is an interesting technical question and not resolved in this paper. On the contrary, another possibility may be an extension of (22) to all closed $F$ leading to the weak convergence of $\{\hat{P}(\pi(n))\}$ to $P^{x}$ by Portmanteau’s theorem. However, the inequality in (22) cannot be strict for all $\pi$, as we have for all closed $F$ and $\epsilon > 0$ (see Section VI-B)

$$1/N \sum_{n=1}^{N} \lim\inf_{t \to \infty} P(\hat{P}(\pi(n))(l) \in F) \geq \mu x.$$

(25)

The last part of Theorem 10 shows that weak detectability (which is necessary for the error of a centralized estimator to be stochastic bounded) is sufficient in the distributed gossip setting to lead to sensor estimates with stochastically bounded errors. The upper bound presented in (24) is highly conservative and, in fact, we have for all closed $F$ (see Section VI-B)

$$\sum_{n=1}^{N} \lim\sup_{t \to \infty} P(\hat{P}(n)(l) \in F) \leq N \mu x.\n$$

(26)

We note a few important technical concerns in the proof of Theorem 10. The key technical difficulty in obtaining Theorem 10 from Theorem 8 is the correlation and nonstationarity induced among the various covariance and switching sequences. In particular, the random permutation (see (16)) leading to the sequences $\{\hat{P}(t)\}$, $n = 1, \leq N$ from the sequences $\{P(t)\}$, $n = 1, \leq N$ has both spatial and temporal correlation. This is due to the possible correlation among the switching sequences $\{z(t)\}$, $n = 1, \leq N$. More importantly, the switching sequences $\{z(t)\}$, $n = 1, \leq N$ are nonstationary (see (11)) and hence, the asymptotic properties of the switching sequences $\{P(t)\}$, $n = 1, \leq N$ do not follow from Theorem 8. This requires us to carry out an elaborate pathwise comparison analysis (Section VI-B) culminating to Theorem 10.

1) **Some General Discussions:** Before proceeding to the technical contents of the paper, we further clarify the major contributions of the paper, the relevance of the assumptions and fundamental differences from existing approaches. The paper addresses a fundamental qualitative question in distributed filtering of linear dynamical systems, in that, what are the minimal conditions on the network information flow and distributed observation (information) pattern, such that, there exists a distributed estimator leading to stable (and ergodic) error covariance at each sensor. We revisit the assumptions (C.1), (S.1), (D.1), and make the following observations justifying the minimality of these conditions to ensure successful filtering:

1) The system matrix $F$ is possibly unstable, no sensor has detectable observations by itself. However, the collection of all the sensor observations possesses a detectability property. Note, the absence of detectable observations at each sensor implies that the optimal local estimator at a sensor (with no cooperation) has unbounded error covariance. Hence, in order to achieve stable estimation error the sensors need to cooperate. The weak detectability assumption on the collective observation process is enforced with the hope that it may lead to a collaborative (distributed) estimation scheme with stable error. We add that the weak detectability condition is minimal, as, under reasonable situations (for example, $F$ is invertible, which is true for all discrete time systems obtained by direct sampling of a continuous evolution process), it is necessary even for the optimum centralized estimator to be stable.

2) We assume a very weak condition on the network connectivity, that restraints the cooperation or information exchange among the sensors. We assume that the inter-sensor communication graph (possibly random) is only connected in the mean. Note that, if the goal is to achieve stable estimation error through a distributed scheme, this assumption cannot be weakened in general. Indeed, if the mean network is not connected, the network separates into two disjoint components (at least), none of which might possess the weak detectability (or any form of detectability as a matter of fact) and any estimation scheme would lead to unstable error.

3) Further, we assume that the communication rate is equal to the system evolution rate, i.e., only one round of inter-sensor communication is possible per system evolution epoch. This truly enforces the distributed nature of all cooperation among sensors.

This work provides an affirmative answer to the above question of designing a successful estimator under the stated assumptions and shows that a simple distributed scheme, the GIKF, leads to error stability (and ergodicity) at each sensor. As our results show, under the communication and information constraints, the GIKF provides sufficient mixing, thus resolving the fundamental interplay between network information flow and spatially distributed information (observation). The qualitative result that weak connectivity and weak detectability leads to stable, ergodic error process at each sensor can be viewed as a distributed stochastic generalization of the fundamental qualitative result in classical Kalman filtering, i.e., detectability of the centralized system leads to finite asymptotic error covariance. At this point, we also briefly contrast our work with existing approaches for distributed Kalman filtering (please refer to the bibliography also.) Most of these schemes rely on an average consensus type step in between system evolution epochs for information mixing. To prove convergence, they mostly assume that a sufficiently large number of consensus rounds take place in each epoch of system evolution, thus violating the constraint that communication and system evolution occur at the same scale. Making the communication rate much faster than the signal dynamics weakens the distributed nature of the problem and in the extreme is equivalent to constructing a center at each sensor. Also, in general, relating the fundamental notions of network connectivity, distributed observability and error stability was not addressed in prior work in this generality.

Of course, like existing schemes, given the possibility of more communication (more than one communication round in every
epoch of system evolution), the GIKF can be modified trivially to obtain improved performance. To this end, think of the particle viewpoint of the GIKF. If there are say $t_0$ communication rounds in each signal evolution epoch, a GIKF particle can hop through $t_0$ sensors and accumulate all the sensor observations in its way. Finally, it updates the estimate (its own or after swapping from a sensor in the way) with all the $t_0$ acquired observations. The convergence analysis of the resulting approach will be a direct modification of the current approach, in that, it will lead to switched Riccati iterates, where each Riccati operator corresponds to the particular subset of sensors, the particle encounters in its way. The same weak detectability condition will lead us through the analysis establishing stability and ergodicity. In fact, a pathwise comparison would reveal that, as $t_0$ increases, the performance improves, i.e., the corresponding invariant distribution $\mu^{U_{t_0}}$ (now also indexed by the superscript $t_0$, that denotes the number of message exchanges in each epoch of signal evolution\(^\text{12}\)) of the conditional covariance becomes closer to $\delta_{P^*}$ (in the Prohorov metric), where $\delta_{P^*}$ denotes the Dirac mass at $P^*$. Here $P^*$ is the fixed point of the centralized Riccati operator corresponding to all observations at all times. In a forthcoming paper, we will show using large deviations theory that, as $t_0 \to \infty$, under generic assumptions the error measure $\mu^{U_{t_0}}$ converges in distribution to $\delta_{P^*}$ at an exponential rate (see also the techniques in [26] in this regard.) This will not only show that the (modified) GIKF will be reasonably close to the optimal centralized estimator (in terms of performance) by slightly increasing the communication rate, but will lead to a complete characterization of the invariant filtering measures whose existence and uniqueness are established here.

We make a few observations on the implementation complexity and the communication overhead of the GIKF scheme. The key computational effort is in the matrix inversion, when computing the Kalman gain. The dimension of this inverse depends on the dimension of the corresponding observation and not the signal dimension. Our problem is motivated by random field monitoring applications, where a network of low capability sensors are deployed to monitor a geographical area. Here the observation of each sensor is quite low dimensional (this also justifies the absence of local observability) and hence the GIKF requires modest computational effort. In fact, in these problems, as the signal dimension (the field area) increases it is customary to increase the number of sensors to guarantee global observability (weak detectability), the observation dimension at each sensor remaining low dimensional limited by its resolution capability. Hence, in this sense, the GIKF stays scalable as the signal size increases. In terms of the communication overhead, the GIKF requires the exchange of sensor estimates and covariances. In general, in wireless scenarios, although the communication channel is uncertain in terms of being active or not, the amount of data that a packet holds can be quite large. Examples are real time scenarios, that support live data streaming. However, if the signal dimension is very large, or if the allocated communication packet size is not large enough, we need to resort to some form of sophisticated encoding/decoding strategies. In fact, in some cases, the covariance matrices have nice enough structure (see, for example, [3]) leading to sufficient compression of the covariance data. In general, this operation may lead to some information loss. This would be an interesting future research question, in particular, the convergence analysis should capture the interplay between this induced distortion and the random Riccati iterates.

**IV. THE AUXILIARY SEQUENCE $\{\tilde{P}_t\}$: RDS FORMULATION**

The asymptotic analysis of the semi-Markov processes $\{P_n(t)\}$ for $n = 1, \ldots, N$ does not fall under the purview of standard approaches based on iterated random systems (IRS)\(^\text{13}\) or a random dynamical system (RDS)\(^\text{14}\) as the switching Markov chains $\{\gamma_n(t)\}$ are nonstationary. In this section, we consider an auxiliary process $\{\tilde{P}(t)\}$ whose evolution is governed by similar random Riccati iterates, the difference being that the switching Markov chain is stationary i.e., the switching Markov chain $\{\tilde{z}(t)\}$ is initialized with the uniform invariant measure on $V$. We analyze the asymptotic properties of the auxiliary sequence $\{\tilde{P}(t)\}$ by formulating it as a RDS on the space $\mathbb{S}_+$ and then in subsequent sections we derive the asymptotics of the sequences $\{P_n(t)\}$ for $n = 1, \ldots, N$ through comparison arguments. We start by formally defining the sequence $\{\tilde{P}(t)\}\(^\text{13}\)$.

Consider a Markov chain on the graph $V$, $\{\tilde{z}(t)\}_{t \in T}$, with transition matrix $\tilde{A}$ and uniform initial distribution, i.e.,

$$P[\tilde{z}(0) = n] = \frac{1}{N}, \quad n = 1, \ldots, N. \quad (27)$$

By Proposition 6, the Markov chain $\{\tilde{z}(t)\}$ is stationary.

We now define the auxiliary process $\{\tilde{P}(t)\}$ as follows:

$$\tilde{P}(t + 1) = f_{\tilde{z}(t)}(\tilde{P}(t)) \quad (28)$$

with (possibly random) initial condition $\tilde{P}(0)\(^\text{14}\)$. Before reading the next two sections, we refer the reader to Appendix A where we review preliminary facts and results from the theory of monotone, sublinear random dynamical systems (RDS)\(^\text{12}\) tailored to our needs. We then show in Section IV-A that the sequence $\{\tilde{P}(t)\}$ for each $n$, admits an ergodic RDS formulation evolving on $\mathbb{S}_+$ and establish some of its properties in Section IV-B.

**A. RDS Formulation of $\{\tilde{P}(t)\}$**

In this subsection, we construct a RDS $(\theta^R, \varphi^R)$ on $\mathbb{S}_+$, which is equivalent to the auxiliary sequence $\{\tilde{P}(t)\}$ in distribution. To this end, we construct the Markov chain $\{\tilde{z}(t)\}$ (in a distributional sense) on the canonical path space. Let $\Omega$.

\(^{12}\)In particular, $t_0 = 1$ corresponds to the current situation ($\mu^{A_1} = \mu^{A}$) of equal communication and signal evolution time scale.

\(^{13}\)Although the sequences $\{P_n(t)\}$ of interest have deterministic initial conditions, it is required for technical reasons (to be made precise later) to allow random initial states $\tilde{P}(0)$, when studying the auxiliary sequence $\{\tilde{P}(t)\}$.

\(^{14}\)We are interested in the distributional properties of the various processes of concern. The actual pathwise construction is not of importance as long as the required distributional equivalence holds. We assume that the measure space $(\Omega, \mathcal{F}, \mathbb{P})$ is rich enough (or suitably extended) to carry out constructions of the various auxiliary random variables.
denote the set \( \{1, \ldots, N\} \) with \( \mathcal{F} \) denoting the corresponding Borel algebra on \( \Omega \), which coincides with the power set of \( \{1, \ldots, N\} \). Denote by \( \Omega^R \) the two-sided infinite product of sets \( \Omega \). \( \Omega^R = \bigotimes_{\infty}^{\infty} \Omega \), i.e., \( \Omega^R \) is the space of double-sided sequences of entries in \( \{1, \ldots, N\} \), i.e.

\[
\Omega^R = \{ \omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots) | \omega_t \in \{1, \ldots, N\}, \forall t \in \mathbb{T} \},
\]

(29)

We equip \( \Omega^R \) with the corresponding product Borel algebra \( \mathcal{F}^R = \bigotimes_{\infty}^{\infty} \mathcal{F} \) generated by the cylinder sets. Note that \( \{\omega_t\}_{t \in \mathbb{T}} \) for all \( \omega \in \Omega^R \) denotes the canonical path space (trajectory) of the Markov chain \( \{\hat{z}(t)\}_{t \in \mathbb{T}} \). The reason for introducing two-sided sequences is a matter of technical convenience and will be evident soon. Consider the unique probability measure \( P^R \) on \( \mathcal{F}^R \), under which the stochastic process (two-sided) \( \{\omega_t\}_{t \in \mathbb{T}} \) is a stationary Markov chain on the finite state space \( \{1, \ldots, N\} \) with transition probability matrix \( \tilde{A} \). By the assumption of stationarity and Proposition 6, the distribution of \( \omega_t \) for each \( t \in \mathbb{T} \) is necessarily the uniform distribution on \( \{1, \ldots, N\} \). In particular, we note that the stochastic processes \( \{\hat{z}(t)\}_{t \in \mathbb{T}} \) and \( \{\omega_t\}_{t \in \mathbb{T}} \) are equivalent in terms of the distribution induced on path space. Define the family of transformations \( \{\theta^R_t\}_{t \in \mathbb{T}} \) on \( \Omega \) as the family of left-shifts, i.e.

\[
\theta^R_t \omega = \omega(t+\cdot) \quad \forall t \in \mathbb{T}.
\]

(30)

With this, the space \( (\Omega^R, \mathcal{F}^R, P^R, \{\theta^R_t, t \in \mathbb{T}\}) \) becomes the canonical path space of a two-sided stationary sequence equipped with the left-shift operator and hence (see, for example, [35]) satisfies the assumptions (A.1)–(A.3) in Definition 18 to be a metric dynamical system and in fact, is also ergodic.

We now set to define the cocycle \( \varphi^R \), see also Definition 18, over \( \mathbb{S}_N^R \), which gives the RDS of interest. We define \( \varphi^R : \mathbb{T}_+ \times \Omega^R \times \mathbb{S}_N^R \mapsto \mathbb{S}_N^R \) by

\[
\varphi^R(0, \omega, X) = X, \quad \forall \omega, X \tag{31}
\]

\[
\varphi^R(1, \omega, X) = f_\omega(\mathbf{X}), \quad \forall \omega, X \tag{32}
\]

\[
\varphi^R(t, \omega, X) = f_{\theta^R_{t-1}\omega}(\varphi^R(t-1, \omega, X)) = f_{\theta^R_{t-1}} \left( \varphi^R(t-1, \omega, X) \right), \quad \forall t > 1, \omega, X. \tag{33}
\]

(Note that, by property of the left shift \( \theta^R \), we have \( \theta^R_{t-1}\omega(0) = \omega(t) \), which explains the equality in (33).) The cocycle \( \varphi^R \) defined satisfies the assumptions of measurability jointly in its arguments and the continuity of the map \( \varphi^R(0, \omega, \cdot) : \mathbb{S}_N^R \mapsto \mathbb{S}_N^R \) w.r.t. the phase variable \( X \) for each fixed \( t \), \( \omega \) from the continuity of the corresponding Riccati operator. The pair \( (\theta^R, \varphi^R) \) thus forms a well-defined RDS on the phase space \( \mathbb{S}_N^R \). Now consider the sequence of random variables \( \{\varphi^R(\cdot, \omega, P_n(0))\}_{t \in \mathbb{T}_+} \) (as explained earlier, the randomness is induced by \( \omega \)) which can be viewed as successive (random) iterates of the RDS \( (\theta^R, \varphi^R) \) starting with the initial state \( P_n(0) \). By construction, it follows that the sequence \( \{\varphi^R(t, \omega, P_n(0))\}_{t \in \mathbb{T}_+} \) is distributionally equivalent to the sequence \( \{\tilde{P}(t)\}_{t \in \mathbb{T}_+} \). In particular,

\[
\varphi^R(t, \omega, P_n(0)) \overset{d}{=} \tilde{P}(t), \quad \forall t \in \mathbb{T}_+. \tag{34}
\]

Thus, analyzing the asymptotic distributional properties of the sequence \( \{\tilde{P}(t)\}_{t \in \mathbb{T}_+} \) is equivalent to studying the sequence \( \{\varphi^R(t, \omega, P_n(0))\}_{t \in \mathbb{T}_+} \), which we undertake in the next subsection.

B. Properties of the RDS \((\theta^R, \varphi^R)\)

We establish some basic properties of the RDS \((\theta^R, \varphi^R)\) representing the auxiliary sequence \( \{\tilde{P}(t)\} \).

**Lemma 12:**

1. The RDS \((\theta^R, \varphi^R)\) is conditionally compact.
2. The RDS \((\theta^R, \varphi^R)\) is order preserving.
3. If in addition \( Q \) is positive definite, i.e., \( Q \succ 0 \), then \( (\theta^R, \varphi^R) \) is strongly sublinear.

**Proof:** The claim in 1) (conditional compactness) is an immediate consequence of the finite dimensionality of the underlying vector space \( \mathbb{S}_N^R \).

The order preserving property 2) follows from the monotonicity of the individual Riccati operators \( f_n \) and hence finite compositions of them remain order-preserving.

The strong sublinearity uses the concavity of the Riccati operators and their monotone nature and is a routine extension to an arbitrary number \( N \) of Riccati operators, given the development in [25] (see [25, Lemma 21]) for the case of two Riccati operators.

V. ASYMMETRIC F. \( \{|\tilde{P}(t)\}_T\}

The main result here concerns the asymptotic properties of the auxiliary sequences \( \{\tilde{P}(t)\}_{t \in \mathbb{T}_+} \) for each \( n \in \{1, \ldots, N\} \). We have the following:

**Theorem 13:** Under the assumptions (C.1), (S.1), (D.1), see page 7, there exists a unique equilibrium probability measure \( \mu^A \) on the space of positive semidefinite matrices \( \mathbb{S}_N^R \), such that, for each \( n \in \{1, \ldots, N\} \), the sequence \( \{\tilde{P}(t)\}_{t \in \mathbb{T}_+} \) converges weakly (in distribution) to \( \mu^A \) from every initial condition \( P_n(0) \):

\[
\{\tilde{P}(t)\} \Rightarrow \mu^A, \quad \forall n \in \{1, \ldots, N\}. \tag{35}
\]

The rest of the subsection is devoted to the proof of the above result. But, before that, we highlight some consequences of Theorem 13.

**Remark 14:** It is important to note, as stated in Theorem 13, that the equilibrium measure \( \mu^A \) does not depend on the index \( n \) and the initial state \( \tilde{P}(0) \) of the sequence \( \{\tilde{P}(t)\}_T \), but is a functional of the network topology and the particular (randomized) communication protocol captured by the matrix \( A \). Theorem 13, thus concludes that the sequences \( \{\tilde{P}(t)\}_T \) reach consensus in
the weak sense to the same equilibrium measure irrespective of the initial states.

The proof of Theorem 13 is rather long and technical, which we accomplish in steps.

Lemma 15: Recall assumption (D.1), page 7 and let, in particular, \( u_0 = (n_1, \ldots, n_\ell) \) be a walk such that, the Grammian
\[
G_{u_0} = \sum_{i=1}^\ell (F_i^{i-1})^T C_n C_n F_i^{i-1} \tag{36}
\]
is invertible, where \( \ell \geq 1 \). Define the function \( g_{u_0} : S^N_+ \mapsto S^N_+ \) by
\[
g_{u_0}(X) = f_{n_\ell} \circ f_{n_{\ell-1}} \circ \cdots \circ f_{n_1}(X). \tag{37}
\]
Then, there exists a constant \( \alpha_0 > 0 \) such that the following uniformity condition holds:
\[
g_{u_0}(X) \leq \alpha_0 I, \quad \forall X \in S^N_+. \tag{38}
\]
In other words the iterate \( g_{u_0}(\cdot) \) is uniformly bounded irrespective of the value of the argument.

The proof is provided in Appendix B. Note that in (36) the observation matrix \( C_n \) is indexed by \( n_\ell \) the current site visited by the random walk \( u_0 \) introduced in Lemma 15. Also, note that the function \( g_{u_0}(X) \) defined in (37) is indexed by the walk \( u_0 \).

The following key lemma establishes asymptotic boundedness properties of \( \{\tilde{P}(t)\} \) and is proved in Appendix B.

Lemma 16: The sequence \( \{\tilde{P}(t)\} \) is stochastically bounded for each \( n_\ell \) under the Assumptions of Theorem 13, i.e.,
\[
\lim_{J \to \infty} \sup_{t \in T_+} \mathbb{P}\left(\|\tilde{P}(t)\| > J\right) = 0. \tag{39}
\]
We now complete the proof of Theorem 13.

From Lemma 12 we note that \( (\theta^R, \varphi^R) \) is strongly sublinear, conditionally compact and order-preserving. Also, the cone \( S^N_+ \) satisfies the conditions required in the hypothesis of Theorem 27. We note for \( t > 0 \)
\[
\varphi^R(t, \omega, 0) = f_{\omega(t-1)} (\varphi(t-1, \omega, 0)) \geq Q \geq 0. \tag{40}
\]
Thus the hypotheses of Theorem 27 are satisfied and precisely one of the assertions (a) and (b) holds. By an argument similar to Lemma 23 in [25], we can show that assertion (a) cannot hold in the face of stochastic boundedness of the sequence \( \{\tilde{P}(t)\}_{t \in T_+} \) (Lemma 16). Thus assertion (b) holds and as a direct consequence of Theorem 27, we establish the existence of a unique almost equilibrium \( u^\lambda(\omega) \geq 0 \) defined on a \( \theta^R \)-invariant set \( \Omega^* \in \mathcal{F}^R \) with \( \mathbb{P}(\Omega^*) = 1 \) such that, for any random variable \( \upsilon(\omega) \) possessing the property \( 0 \leq \upsilon(\omega) \leq \alpha u^\lambda(\omega) \) for all \( \omega \in \Omega^* \) and deterministic \( \alpha > 0 \), the following holds:
\[
\lim_{t \to \infty} \varphi(t, \theta^{-t} \omega, \upsilon(\theta^{-t} \omega)) = u^\lambda(\omega), \quad \omega \in \Omega^*. \tag{41}
\]
From the distributional equivalence of pull-back and forward orbits, Lemma 24 establishes the existence of a unique almost equilibrium \( u^\lambda \), i.e., a unique equilibrium measure for the process \( \{\tilde{P}(t)\} \) from the distributional equivalence of pull-back and forward orbits. However, to show that the measure induced by \( u^\lambda \) on \( S^N_+ \) is attracting for \( \{\tilde{P}(t)\} \), (41) must hold for all initial \( \upsilon \), whereas Lemma 24 establishes convergence for a restricted class of initial conditions \( \upsilon \). We need the following result to extend it to general initial conditions.

Lemma 17: Under the assumptions of Theorem 13, let \( u^\lambda \) be the unique almost equilibrium of the RDS \( (\theta^R, \varphi^R) \). Then
\[
\mathbb{P}\left(\omega : u^\lambda(\omega) \geq Q\right) = 1. \tag{42}
\]

Proof: The proof uses the fact that, for all \( n_\ell \), \( f_{n_\ell}(X) \geq Q \) and is routine given the corresponding development in Lemma 24 of [25].

We now complete the proof of Theorem 13.

Proof of Theorem 13: Let \( u^\lambda \) be the distribution of the unique almost equilibrium in (41). By Lemma 17 we have \( u^\lambda (S^N_+) = 1 \). Let \( P_0 \in S^N_+ \) be an arbitrary initial state. By construction of the RDS \( (\theta^R, \varphi^R) \), the sequences \( \{P_t\}_{t \in T_+} \) and \( \{\varphi^R(t, \omega, P_0)\}_{t \in T_+} \) are distributionally equivalent, i.e.,
\[
P_t \overset{d}{=} \varphi^R(t, \omega, P_0). \tag{43}
\]
Recall \( \Omega^* \) as the \( \theta^R \)-invariant set with \( \mathbb{P}^R(\Omega^*) = 1 \) in (41) on which the almost equilibrium \( u^\lambda \) is defined. By Lemma 17, there exists \( \Omega_1 \subset \Omega^* \) with \( \mathbb{P}^R(\Omega_1) = 1 \), such that
\[
u^\lambda(\omega) \geq Q, \quad \omega \in \Omega_1. \tag{44}
\]
Define the random variable \( \tilde{X} : \Omega \mapsto S^N_+ \) by
\[
\begin{cases}
P_0 & \text{if } \omega \in \Omega_1, \\
0 & \text{if } \omega \in \Omega_1^c. 
\end{cases}
\tag{45}
\]
Now choose \( \alpha > 0 \) sufficiently large such that
\[
0 \leq \tilde{X}(\omega) \leq \alpha u^\lambda(\omega), \quad \omega \in \Omega^*. \tag{46}
\]
Indeed, we have
\[
0 \leq P_0 = \tilde{X}(\omega) \leq \alpha Q \leq \alpha u^\lambda(\omega), \quad \omega \in \Omega_1. \tag{47}
\]
and
\[
0 = \tilde{X}(\omega) \leq \alpha u^\lambda(\omega), \quad \omega \in \Omega_1^c. \tag{48}
\]
We then have by the discussion preceding (41)
\[
\lim_{t \to \infty} \varphi^R(t, \theta^{-t} \omega, \tilde{X}(\theta^{-t} \omega)) = u^\lambda(\omega), \quad \omega \in \Omega^*. \tag{49}
\]
Since convergence \( \mathbb{P}^R \) a.s. implies convergence in distribution, we have
\[
\varphi^R(t, \theta^{-t} \omega, \tilde{X}(\theta^{-t} \omega)) \Rightarrow u^\lambda(\omega), \quad \omega \in \Omega^*. \tag{50}
\]
\[
\left\{ \varphi^R \left( t, \omega, \tilde{X}(\omega) \right) \right\}_{t \in \mathbb{T}_+} \text{ also converges in distribution to the unique stationary distribution } \mu^A, \text{ i.e., as } t \to \infty
\]
\[\varphi^R \left( t, \omega, \tilde{X}(\omega) \right) \Rightarrow \mu^A. \quad (51)\]

Now, since \( P^n (\Omega_1) = 1 \), by (44)
\[\varphi^R (t, \omega, P_0) = \varphi^R \left( t, \omega, \tilde{X}(\omega) \right), \quad P^n \text{ a.s., } t \in \mathbb{T}_+ \quad (52)\]
which implies
\[\varphi^R (t, \omega, P_0) \equiv \varphi^R \left( t, \omega, \tilde{X}(\omega) \right), \quad t \in \mathbb{T}_+. \quad (53)\]

From (51) and (53), we then have \( \varphi^R (t, \omega, P_0) \Rightarrow \mu^A \), which together with the distributional equivalence \( P_t \equiv \varphi^R (t, \omega, P_0) \) noted above implies, as \( t \to \infty \), \( P_t \Rightarrow \mu^A \).

VI. PROOFS OF MAIN RESULTS

A. Proof of Theorem 8

Proof: By Theorem 13 we know that such a sequence \( \{ \tilde{P}(t) \} \) converges weakly to \( \mu^A \) when started from a deterministic initial condition. In the case, \( \tilde{P}(0) \) is distributed as \( \nu \), we note that, by the independence of \( \tilde{P}(0) \) and the Markov chain \( \{ q(t) \} \),
\[E \left[ g \left( \tilde{P}(t) \right) \right] = \int_{\mathbb{S}_N} E \left[ \left( \tilde{P}(t) \right) \right] d\nu(X) \quad (54)\]
for any \( g \in C_b(\mathbb{S}_N^N) \). Now, the distribution of the sequence \( \{ \tilde{P}(t) \} \) conditioned on the event \( \tilde{P}(0) = X \) is the same as that when the sequence starts with the deterministic initial condition \( X \) (this is true because \( \tilde{P}(0) \) is independent of \( \{ q(t) \} \)). Hence, by Theorem 13
\[\lim_{t \to \infty} E \left[ \left( \tilde{P}(t) \right) \right] = \int_{\mathbb{S}_N^N} g(y) d\mu^A(Y) \quad (55)\]
for all \( X \). Since \( g \) is bounded, the dominated convergence theorem and (54) lead to
\[\lim_{t \to \infty} E \left[ \left( \tilde{P}(t) \right) \right] = \int_{\mathbb{S}_N^N} g(y) d\mu^A(Y) \quad (56)\]
for all \( g \in C_b(X) \) and hence the required weak convergence follows.

B. Proof of Theorem 10

Proof: We prove Theorem 10 in the order (1)–(3). Consider any \( \Gamma \in \mathcal{B}(\mathbb{S}_N^N) \). We estimate the probability \( P \left( \tilde{P}_q (t) \in \Gamma \right) \). To this end, we note that
\[P \left( \tilde{P}_q (t) \in \Gamma \right) = \sum_{n=1}^{N} P \left( \tilde{P}_n (t) \in \Gamma \right) P(q = n)\]
\[= \frac{1}{N} \sum_{n=1}^{N} P \left( \tilde{P}_n (t) \in \Gamma \right). \quad (57)\]
The first step holds because \( q \) is independent of the sequences \( \{ \tilde{P}_n (t) \} \) for all \( n \) and subsequently we use that \( q \) is uniformly distributed on \( V \). Denoting by \( \pi_t^{-1} \) the inverse of the permutation \( \pi_t \), we have
\[P \left( \tilde{P}_q (t) \in \Gamma \right) = P \left( \tilde{P}_q (t) \in \Gamma \right) \]
\[= \frac{1}{N} \sum_{l=1}^{N} P \left( \{ P_l (t) \in \Gamma \} \right) \]
\[= \frac{1}{N} \sum_{l=1}^{N} P \left( \{ P_l (t) \in \Gamma \} \right) \]
\[= \frac{1}{N} \sum_{l=1}^{N} P \left( \{ P_l (t) \in \Gamma \} \right). \quad (59)\]

Note the last step follows from the fact that
\[\sum_{n=1}^{N} P \left( \{ P_l (t) \in \Gamma \} \right) \]
\[= \frac{1}{N} \sum_{l=1}^{N} P \left( \{ P_l (t) \in \Gamma \} \right) \]
because the events \( \{ \pi_t^{-1}(n) = l \} \), \( n = 1, \ldots, N \) are mutually exclusive and exhaustive, \( \pi_t^{-1}(l) \) being a permutation.

Now consider a stationary Markov chain \( \{ \tilde{Z}(t) \} \) on \( V \) with transition probability \( A \) and let \( \{ \tilde{P}(t) \} \) be the sequence defined by
\[\tilde{P}(t+1) = f_{\tilde{Z}(t)} \left( \tilde{P}(t) \right), \quad t \in \mathbb{T}_+ \quad (61)\]
with initial condition \( \tilde{P}(0) = \tilde{P}(0) \). Then,
\[P \left( \tilde{P}(t) \in \Gamma \right) \]
\[= \frac{1}{N} \sum_{l=1}^{N} P \left( \tilde{P}(t) \in \Gamma \right) \]
\[= \frac{1}{N} \sum_{l=1}^{N} P \left( \tilde{P}(t) \in \Gamma \right) \]
\[= \frac{1}{N} \sum_{l=1}^{N} P \left( \tilde{P}(t) \in \Gamma \right) \]

By construction, the distribution of the sequence \( \{ \tilde{P}(t) \} \) conditioned on the event \( \{ \tilde{Z}(0) = l \} \) is equivalent to that of the sequence \( \{ P_l (t) \} \) and hence
\[P \left( \tilde{P}(t) \in \Gamma \right) \]
\[= \frac{1}{N} \sum_{l=1}^{N} P \left( \tilde{P}(t) \in \Gamma \right) \]
\[= \frac{1}{N} \sum_{l=1}^{N} P \left( \tilde{P}(t) \in \Gamma \right) \]

Hence by (59) and (63) we obtain
\[P \left( \tilde{P}_q (t) \in \Gamma \right) = P \left( \tilde{P}_q (t) \in \Gamma \right). \quad (64)\]
Thus, for all $t$, $P_q(t) \overset{d}{=} \hat{P}(t)$. By Theorem 13, we then have the weak convergence of the sequence $\{P_q(t)\}$ to $\mu^A$. For the third part, we note that for any $\Gamma \in B(S^n_+)$

$$\frac{1}{N} \sum_{n=1}^{N} P \left( \hat{P}_n(t) \in \Gamma \right) = P \left( \hat{P}_q(t) \in \Gamma \right)$$  \hspace{1cm} (65)

due to the independence of $q$ from $\{A(t)\}$. Taking the $\limsup$ and noting the nonnegativity of the terms, we have for closed $F$,

$$\limsup_{t \to \infty} \sum_{n=1}^{N} P \left( \hat{P}_n(t) \in F \right) \leq \sum_{n=1}^{N} P \left( \hat{P}_n(t) \in F \right)$$

$$= N \limsup_{t \to \infty} \sum_{n=1}^{N} \left[ \frac{1}{N} P \left( \hat{P}_n(t) \in F \right) \right]$$

$$= N \limsup_{t \to \infty} P \left( \hat{P}_q(t) \in F \right) \leq N \mu^A.$$  \hspace{1cm} (66)

The proof of the second part involves an auxiliary construction and approximation arguments to relate the limit properties of the sequences $\{P_n(t)\}$ to similar processes, where the underlying switching Markov chain is stationary. To this end consider any strictly positive $s \in \mathbb{T}_+$. Recall the Markov chains $\{z_n(t)\}$ for $n = 1, \ldots, N$ with transition probability matrix $A$ and initial state $z_n(0) = \pi$. The corresponding sequence of interacting particle processes $\{P_n(t)\}$ are constructed, for each $n$ as:

$$P_n(t+1) = f_{z_n(t)}(P_n(t))$$  \hspace{1cm} (67)

with initial condition $P_n(0) = \hat{P}$. Let $f_0 : S^n_+ \mapsto S^n_+$ denote the Lyapunov operator

$$f_0(X) = FXF^T + Q$$  \hspace{1cm} (68)

and note that the following ordering holds:

$$f_n(X) \leq f_0(X), \quad \forall n$$  \hspace{1cm} (69)

For a given $s > 0$ chosen above and for all $n$, define the processes $\{P^s_n(t)\}_{t \geq s}$ by

$$P^s_n(t+1) = f_{z_n(t)}(P^s_n(t))$$  \hspace{1cm} (70)

with deterministic initial value $P^s_n(s) = f_0 \left( \hat{P}(0) \right)$. By (68), for any $s$ tuple $(i_0, i_1, \ldots, i_{t-s})$ with $i_r \in [1, \ldots, N]$ for $r = 0, \ldots, s - 1$, we note that

$$f_{i_{t-s}} \circ f_{i_{t-2}} \circ \cdots \circ f_{i_0} \left( \hat{P}(0) \right) \leq f_0 \left( \hat{P}(0) \right)$$  \hspace{1cm} (71)

and hence, by the monotonicity of the Riccati operators, we conclude that for all $n$

$$P_n(t) \leq P^s_n(t), \quad t \geq s.$$  \hspace{1cm} (72)

Also consider a stationary Markov chain $\{q(t)\}_{t \geq s}$ with transition probability $A$, i.e., $q(0)$ is uniformly distributed on $V$ and define the process $\{Q^s(t)\}_{t \geq s}$ by

$$Q^s(t+1) = f_{q(t)}(Q^s(t))$$  \hspace{1cm} (73)

with deterministic initial value

$$Q^s(s) = f_0 \left( \hat{P}(0) \right).$$  \hspace{1cm} (74)

It is to be noted that by Theorem 13, the process $\{Q^s(t)\}$ converges weakly to $\mu^A$, i.e.,

$$\lim_{t \to \infty} d_{D^*} \left( Q^s(t), \mu^A \right) = 0$$  \hspace{1cm} (75)

where $d_{D^*}$ denotes the Prohorov metric. We now set to relate the limit properties of $\{P^s_n(t)\}$ to those of $\{Q^s(t)\}$. For $t \geq s$ define the total variation distance between $P^s_n(t)$ and $Q^s(t)$ by

$$d_{t-s} \left( P^s_n(t), Q^s(t) \right) = \sup_{\Gamma \in B(S^n_+)} \left[ |P \left( P^s_n(t) \in \Gamma \right) - P \left( Q^s(t) \in \Gamma \right)| \right].$$  \hspace{1cm} (76)

Since for any $t$, the two sequences considered above assume values in a finite set, we define a set of $(t-s)$ tuples $\Lambda(\Gamma)$ by (77) shown at the bottom of the page. It is clear that

$$\{P^s_n(t) \in \Gamma \} \iff \{(z_n(t), \ldots, z_n(t-1)) \in \Lambda(\Gamma)\} \hspace{1cm} (77)$$

$$\{Q^s(t) \in \Gamma \} \iff \{(q(t), \ldots, q(t-1)) \in \Lambda(\Gamma)\}.$$  \hspace{1cm} (78)

We then have

$$P \left( P^s_n(t) \in \Gamma \right) - P \left( Q^s(t) \in \Gamma \right)$$

$$= \sum_{i_1} P \left( z_n(t) = i_1 \right) \sum_{(i_1, \ldots, i_{t-s}) \in \Lambda(\Gamma)} \prod_{r=1}^{t-s-1} \lambda_{i_r,i_{r+1}}$$

$$- \sum_{i_1} P \left( q(t) = i_1 \right) \sum_{(i_1, \ldots, i_{t-s}) \in \Lambda(\Gamma)} \prod_{r=1}^{t-s-1} \lambda_{i_r,i_{r+1}}$$

$$= \sum_{i_1} \left[ P \left( z_n(t) = i_1 \right) - P \left( q(t) = i_1 \right) \right] \sum_{(i_1, \ldots, i_{t-s}) \in \Lambda(\Gamma)} \prod_{r=1}^{t-s-1} \lambda_{i_r,i_{r+1}}$$

\[ \Lambda(\Gamma) = \{(i_1, \ldots, i_{t-s}) | i_r \in [1, \ldots, N] \text{ for all } r \text{ and } f_{i_{t-s}} \circ \cdots \circ f_{i_1} \left( f_0 \left( \hat{P}(0) \right) \right) \}. \]  \hspace{1cm} (79)
and hence
\[
|P(P_n^\infty(t) \in \Gamma) - P(Q^\infty(t) \in \Gamma)|
\leq \sum_{i_1} \sum_{i_2} \cdots \sum_{i_{t-1}} \prod_{r=1}^{t-1} \text{A}_{i_r,i_{r+1}}
\leq \sum_{i_1} \sum_{i_2} \cdots \sum_{i_{t-1}} P(z_n(s) = i_1) - P(q(s) = i_1)
\leq \sum_{i_1} d_v(z_n(s), q(s)) \leq N d_v(z_n(s), q(s))
\]
where we have used the fact that
\[
\sum_{i_1} \sum_{i_2} \cdots \sum_{i_{t-1}} \prod_{r=1}^{t-1} \text{A}_{i_r,i_{r+1}}
= \prod_{i_1} \text{A} = P\left(z_n(s+1) = (i_2, \ldots, i_{t-s}) \mid z_n(s) = i_1\right)
\leq 1.
\]

We thus obtain
\[
d_v(P_n^\infty(t), Q^\infty(t)) \leq N d_v(z_n(s), q(s)), \quad \forall t \geq s. \tag{79}
\]
It is well known that the finite state Markov chain \(\{z_n(s)\}\) converges weakly at a geometric rate to the uniform measure, i.e., the measure induced by \(q(s)\) for each \(s\) and hence in variation. In other words
\[
\lim_{s \to \infty} d_v(z_n(s), q(s)) = 0. \tag{80}
\]
Thus, by (79), we have
\[
\lim_{s \to \infty} \sup_{t \geq s} d_v(P_n^\infty(t), Q^\infty(t)) = 0 \tag{81}
\]
and since convergence in total variation implies weak convergence ([36]), we have
\[
\lim_{s \to \infty} \sup_{t \geq s} d_P(P_n^\infty(t), Q^\infty(t)) = 0. \tag{82}
\]
Now consider \(\varepsilon > 0\). Then there exists \(s(\varepsilon)\), such that,
\[
d_P(P_n^\infty(t), Q^\infty(t)) \leq \frac{\varepsilon}{2}, \quad s \geq s(\varepsilon), \quad t \geq s. \tag{83}
\]
Since the sequence \(\{Q^\infty(t)\}\) converges weakly to \(\mu^A\) for all \(s\) (in particular for \(s = s(\varepsilon)\)), there exists \(t(\varepsilon) \geq s(\varepsilon)\) sufficiently large, such that,
\[
d_P(Q^\infty(\varepsilon)(t), \mu^A) \leq \frac{\varepsilon}{2}, \quad t \geq t(\varepsilon). \tag{84}
\]
Then, an application of the triangle inequality for the metric \(d_P\) leads to
\[
d_P(P_n^\infty(t), \mu^A)
\leq d_P(P_n^\infty(t), Q_n^\infty(t)) + d_P(Q_n^\infty(t), \mu^A) \leq \varepsilon \tag{85}
\]
for all \(t \geq t(\varepsilon)\). Now, by definition, see (86) shown at the bottom of the page where \(F^\varepsilon\) is defined as
\[
F^\varepsilon = \left\{ X \in \mathbb{S}^N_+ \mid \inf_{Y \in F} \|X - Y\| < \delta \right\}. \tag{87}
\]
Since, by (85), \(d_P(P_n^\infty(t), \mu^A) \leq \varepsilon\) for all \(t \geq t(\varepsilon)\), we have, for any closed set \(F\),
\[
P\left(P_n^\infty(t) \in F\right) \leq \mu^A(F^\varepsilon) + \varepsilon, \quad t \geq t(\varepsilon). \tag{88}
\]
In addition to \(F\) being closed, let us assume that \(F\) satisfies monotonicity, i.e., \(X \in F\) implies \(Y \in F\) for all \(Y \geq X\). By (71), we have
\[
P_n(t) \leq P_n^\infty(t), \quad t \geq t(\varepsilon) \tag{89}
\]
and hence
\[
P\left(P_n(t) \in F\right) \leq P\left(P_n^\infty(t) \in F\right), \quad t \geq t(\varepsilon). \tag{90}
\]
We then have from (88) for all \(t \geq t(\varepsilon)\)
\[
P\left(P_n(t) \in F\right) \leq \mu^A(F^\varepsilon) + \varepsilon, \tag{91}
\]
Taking the limit as \(t \to \infty\), we have
\[
\lim_{t \to \infty} P\left(P_n(t) \in F\right) \leq \mu^A(F^\varepsilon) + \varepsilon. \tag{92}
\]
The L.H.S. above is now independent of \(t\) and hence, \(\varepsilon\) through \(t(\varepsilon)\). Since the above holds for arbitrary \(\varepsilon > 0\), moving to the limit as \(\varepsilon \to 0\) yields
\[
\lim_{\varepsilon \to 0} \sup_{t \geq 0} P\left(P_n(t) \in F\right) \leq \lim_{\varepsilon \to 0} \mu^A(F^\varepsilon) = \mu^A(F). \tag{93}
\]
The last step follows from the continuity of the probability measure \(\mu^A\) and the fact that for closed \(F\)
\[
\bigcap_{\varepsilon > 0} F^\varepsilon = F, \tag{94}
\]
This establishes the result for general order preserving \(F\). The result for sets of the form \(\{X \in \mathbb{S}^N_+ \mid X \geq \alpha I\}\) or
\{ X \in \mathbb{F}_+^N \mid \|X\| \geq \alpha \} \text{ for } \alpha > 0 \text{ follow, as they satisfy the general hypothesis on } F. \]

\section{Concluding Remarks}

The paper develops the gossip interactive Kalman filter (GIKF) for distributed Kalman filtering in sensor networks, when observation sampling and intersensor communication occur at the same time scale. Inter-sensor collaboration is achieved by intermittent exchange of filtering states. A traveling particle interpretation of the filtering states leads to a random dynamical system (RDS) formulation of the sequence of conditional error covariances. Under a weak detectability assumption, the estimation error process at each sensor stays stochastically bounded (irrespective of the instability in signal dynamics), provided the network satisfies some weak connectivity conditions. Also, the network achieves weak consensus, i.e., the conditional error covariance (or the pathwise filtering error) at a randomly selected sensor converges in distribution to a unique invariant measure $\mu^A$. The invariant measure $\mu^A$ depends on the network connectivity process (the MAC protocol) through the mean $\bar{A}$ of the random adjacency matrix $A$.

The characterization of the invariant measure $\mu^A$ as a functional of the matrix $\bar{A}$ is interesting to study the sensitivity of the mapping, $\bar{A} \mapsto \mu^A$. This would lead to understanding the robustness of the above filtering approach to perturbations in the communication policy, i.e., whether a small change in the MAC protocol (a perturbation of $\bar{A}$) leads to a negligible change of $\mu^A$, or the filtering performance changes dramatically. Exploring such comparison principles for the mapping would lead to understanding the more complicated problem of characterizing the invariant measure $\mu^A$. Such a characterization, in general, is difficult as there seems to be no direct way of obtaining a functional mapping $\bar{A}$ to $\mu^A$. In fact, a much simpler situation (Kalman filtering with intermittent observations) involving a single sensor with observation packet losses demands the machinery of moderate deviations ([26]) and large random matrix theory ([28]) for a characterization of the invariant measures.

In conclusion, we note some natural extensions worthy of future research. Note, in the work we assume time-invariance in the system matrices $(F, Q, C_0, R_n)$. If these matrices are time-varying, the GIKF algorithm will be still applicable, as all the Riccati updates can be performed with time-varying matrices. However, the convergence properties would require further conditions, as is required even in the centralized setting. In a deterministic environment (the system matrices are not random), possibly some periodic assumptions on the time-varying matrices might lead to interesting results. In a random environment, if the sequence of time-varying system matrices form a stationary stochastic process, then the random dynamical system formulation will go through similarly by enlarging the probability space where the metric dynamical system is defined. The weak detectability assumption should then be replaced by positive definiteness of the Grammian (corresponding to a detectable path) in probability only. In the general nonstationary regime additional conditions on the stochastic dependence need to be imposed (the mixing conditions in [30], of course adapted to a distributed setting, might be useful in establishing tightness in this regard.) The assumption of temporally independent adjacency matrices (assumption (C.1)) might not be valid in networks with high degree of agent mobility and nonstationary channel characteristics. Incorporating nonstationary dynamics in the communication process could be another challenging research topic. Also, the current framework requires the knowledge of the system matrices. An interesting research question in this regard, would be the simultaneous analysis of the online system identification and the random Riccati recursions, that are fed by the output of the learning scheme. Even in the current framework, several modifications of the basic GIKF scheme may be possible with a view to improve the filter performance. As an example, suppose, instead of pairwise communication, we allow each sensor to talk to more than one neighbor at a time. Then, the sensor can obtain the observations from all its neighbors (including its own and the one it swaps estimate with) and then perform the Riccati update with all these observations. A pathwise coupling argument will reveal that the performance will improve. However, we emphasize here that one of the main concerns of the paper is to connect the fundamental notions of network connectivity, distributed observability and error stability. As in all control and signal processing questions, the first step consists of establishing qualitative properties like stability, ergodicity based on structural conditions, the second step being refining the theory for optimum performance. In this paper, we focus on the nontrivial first step of establishing stability and ergodicity under minimal network and observation conditions. This step is necessary for refining and understanding the subsequent steps of quantitative performance evolution in future research.

\section*{Appendix A}

\subsection*{Random Dynamical Systems: Facts and Results}

We start by defining a random dynamical system (RDS). In the sequel, we follow the notation in [32], [37].

\textbf{Definition 18 (RDS):} A RDS with (one-sided) time $\mathbb{T}_+$ and state space $X$ is a pair $(\theta, \varphi)$ with the following properties:

\begin{enumerate}
  \item[(A.1)] \( \theta_0 = \text{id}_X \), \( \theta_0 \circ \theta_s = \theta_{t+s}, \forall t, s \in \mathbb{T} \)
  \item[(A.2)] \( (t, \omega) \mapsto \theta_t(\omega) \) is measurable.
  \item[(A.3)] \( \theta_t \mathbb{P} \subseteq \mathbb{P} \) \( \forall t \in \mathbb{T} \), i.e., \( \mathbb{P}(\theta_t B) = \mathbb{P}(B) \) for all \( B \in \mathcal{F} \) and all \( t \in \mathbb{T} \).
\end{enumerate}

\textbf{A.} A cocycle $\varphi$ over $\theta$ of continuous mappings of $X$ with time $\mathbb{T}_+$, i.e., a measurable mapping $\varphi : \mathbb{T}_+ \times \Omega \times X \to X$, $(t, \omega, X) \mapsto \varphi(t, \omega, X)$ (95) such that

\begin{enumerate}
  \item[(B.1)] The mapping $X \mapsto \varphi(t, \omega, X) \equiv \varphi(t, \omega)_X$ is continuous in $X$ for every $t \in \mathbb{T}_+$ and $\omega \in \Omega$.
  \item[(B.2)] The mappings $\varphi(t, \omega)$ satisfy the cocycle property:
  \begin{equation}
  \varphi(0, \omega) = \text{id}_X, \quad \varphi(t+s, \omega) = \varphi(t, \theta_s \omega) \circ \varphi(s, \omega) \quad (96)
  \end{equation}
  \text{ for all } t, s \in \mathbb{T}_+ \text{ and } \omega \in \Omega.
\end{enumerate}

\footnote{The notation $\varphi(t, \omega)_X$ is to be interpreted as a function $\varphi(t, \omega)$ acting on the argument $X$ for fixed $t$, $\omega$.}
Although we consider in this paper discrete time RDS, the general notion of RDS, as defined in [32], applies equally well to dynamical systems with continuous time. In the above definition, the randomness is captured by the probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and iterates indexed by \(\omega\) indicates pathwise construction. For example, if \(X_0\) is the deterministic initial state of the system of interest at time \(t = 0\), the random state at time \(t \in T_+\) is given by
\[
X_t(\omega) = \varphi(t, \omega, X_0).
\]
(97)
The measurability assumptions in the definition above, guarantee that the random state \(X_t\) is a well-defined random variable. Also, note that the iterates are defined for nonnegative (one-sided) time, however, the family of transformations \(\{\theta_t\}\) is two-sided, which is purely for technical convenience, as will be seen later.

Some Results From RDS Theory: We summarize terminologies and notions used in the RDS literature (see [32], [34] for details.)

Consider a generic RDS \((\theta, \varphi)\) with state space \(X\) as in Definition 18. In the following we assume that \(X\) is a nonempty subset of a real Banach space \(V\) with a closed, convex, solid, normal (w.r.t. the Banach space norm,) minimal closed cone \(V_+\). We denote by \(\leq\) the partial order induced by \(V_+\) in \(X\) and \(\leq\) denotes the corresponding strong order. Although the development that follows may hold for arbitrary \(X \subseteq V\), in the sequel we assume \(X = V_+\) (which is true for the RDS \((\theta^R, \varphi^R)\) modeling the Random Algebraic Riccati Equation (RAE)).

Definition 19 (Order-Preserving RDS): An RDS \((\theta, \varphi)\) with state space \(V_+\) is called order-preserving if
\[
X \subseteq Y \implies \varphi(t, \omega, X) \subseteq \varphi(t, \omega, Y), \forall t \in T_+, \omega \in \Omega, X, Y \in V_+.
\]
(98)

Definition 20 (Sublinearity): An order-preserving RDS \((\theta, \varphi)\) with state space \(V_+\) is called sublinear if for every \(X \in V_+\) and \(\lambda \in (0, 1)\) we have
\[
\lambda \varphi(t, \omega, X) \leq \varphi(t, \omega, \lambda X), \forall t > 0, \omega \in \Omega.
\]
(99)
The RDS is said to be strictly sublinear if strict inequality in (99) holds for \(X \in \text{Int} V_+\), i.e., for \(X \in \text{Int} V_+\),
\[
\lambda \varphi(t, \omega, X) < \varphi(t, \omega, \lambda X), \forall t > 0, \omega \in \Omega.
\]
(100)
and strongly sublinear if in addition to (99), we have
\[
\lambda \varphi(t, \omega, X) \ll \varphi(t, \omega, X), \forall t > 0, \omega \in \Omega, X \in \text{Int} V_+.
\]
(101)

Definition 21 (Equilibrium): A random variable \(u : \Omega \mapsto V_+\) is called an equilibrium (fixed point, stationary solution) of the RDS \((\theta, \varphi)\) if it is invariant under \(\varphi\), i.e.,
\[
\varphi(t, \omega, u(\omega)) = u(\theta_t \omega), \forall t \in T_+, \omega \in \Omega.
\]
(102)
In case, (102) holds for all \(\omega \in \Omega\) except on a set of \(\mathbb{P}\) measure zero, we call \(u\) an almost equilibrium.

Since, the transformations \(\{\theta_t\}\) are measure-preserving, i.e., \(\theta_t \mathbb{P} = \mathbb{P}, \forall t,\) we have
\[
u(\theta_t \omega) = u(\omega), \forall t.
\]
(103)
Thus (102), in particular, implies that, for an almost equilibrium \(u\), the sequence of iterates \(\{\varphi(t, \omega, u(\omega))\}_{t \in T_+}\) have the same distribution, which is the distribution of \(u\).

Definition 22 (Part): The equivalence classes in \(V_+\) under the equivalence relation defined by \(X \sim Y\) if there exists \(\alpha > 0\) such that \(\alpha^{-1} X \subseteq Y \subseteq \alpha X\) are called parts of \(V_+\).

We call the part \(C_\omega\) generated by a random variable \(\varphi : \Omega \mapsto V_+\) as the collection of random variables \(u : \Omega \mapsto V_+\) such that there exists deterministic \(\alpha(u) \geq 1\) with
\[
\alpha(u)^{-1} u(\omega) \leq u(\omega) \leq \alpha(u) u(\omega), \forall \omega \in \Omega.
\]
(104)

Definition 23 ( Orbit): For a random variable \(u : \Omega \mapsto V_+\) we define the forward orbit \(\eta^L(u)\) emanating from \(u(\omega)\) as the random set \(\{\varphi(t, \omega, u(\omega))\}_{t \in T_+}\). The forward orbit gives the sequence of iterates of the RDS starting at \(u\).

Although \(\eta^L(u)\) is the object of practical interest, for technical convenience (will be seen later,) we also define the pull-back orbit \(\eta^R(u)\) emanating from \(u\) as the random set \(\{\varphi(t, \theta_{-t} \omega, u(\theta_{-t} \omega))\}_{t \in T_+}\). The reason for defining the pull-back orbit is that it is comparatively convenient to establish asymptotic properties for \(\eta^L(u)\). However, analyzing \(\eta^R(u)\) leads to understanding asymptotic distributional properties for \(\eta^L(u)\), because the random sequences \(\{\varphi(t, \omega, u(\omega))\}_{t \in T_+}\) and \(\{\varphi(t, \theta_{-t} \omega, u(\theta_{-t} \omega))\}_{t \in T_+}\) are equivalent in distribution. In other words
\[
\varphi(t, \omega, u(\omega)) \mathbb{P} \sim \varphi(t, \theta_{-t} \omega, u(\theta_{-t} \omega)), \forall t \in T_+.
\]
(105)
This follows from the fact that \(\theta_t \mathbb{P} = \mathbb{P}, \forall t \in T_\). Thus, in particular, we have the following assertion.

Lemma 24: Let the sequence \(\{\varphi(t, \omega, u(\omega))\}_{t \in T_+}\) converge in distribution to a measure \(\mu\) on \(V_+\), where \(u : \Omega \mapsto V_+\) is a random variable. Then the sequence \(\{\varphi(t, \omega, u(\omega))\}_{t \in T_+}\) also converges in distribution to the measure \(\mu\).

We now introduce notions of boundedness of RDS, which will be used in the sequel.

Definition 25 (Boundedness): Let \(a : \Omega \mapsto V_+\) be a random variable. The pull-back orbit \(\eta^R(u)\) emanating from \(a\) is said to be bounded on \(U \in \mathcal{F}\) if there exists a random variable \(C\) on \(U\) such that
\[
\|\varphi(t, \omega, a(\omega))\| \leq C(\omega), \forall t \in T_+, \omega \in U.
\]
(106)

Definition 26 (Conditionally Compact RDS): An RDS \((\theta, \varphi)\) in \(V_+\) is said to be conditionally compact if for any \(U \in \mathcal{F}\) and pull-back orbit \(\eta^R(u)\) which is bounded on \(U\) there exists a family of compact sets \(\{K(\omega)\}_{\omega \in U}\) such that
\[
\lim_{t \to \infty} \text{dist} (\varphi(t, \omega, a(\omega)), K(\omega)) = 0, \omega \in U.
\]
(107)
It is to be noted that conditionally compact is a topological property of the space \(V_+\).

We now state a limit set dichotomy result for a class of sublinear, order-preserving RDS.

Theorem 27 (Corollary 4.3.1. in [32]): Let \(V\) be a separable Banach space with a normal solid cone \(V_+\). Assume that \((\theta, \varphi)\) is a strongly sublinear conditionally compact order-preserving RDS over an ergodic metric dynamical system \(\theta\). Sup-
pose that $\varphi(t, \omega, 0) \geq 0$ for all $t > 0$ and $\omega \in \Omega$. Then precisely one of the following applies.

a) For any $X \in V_+$ we have

$$P \left( \lim_{t \to \infty} \| \varphi(t, \theta, \omega, X) \| = \infty \right) = 1.$$  

(108)

b) There exists a unique almost equilibrium $\pi(\omega) > 0$ defined on a $\theta$-invariant set $\mathcal{P} \in \mathcal{F}$ with $P(\Omega^*) = 1$ such that for any random variable $\psi(\omega)$ possessing the property $0 \leq \psi(\omega) \leq \alpha_0 \pi(\omega)$ for all $\omega \in \Omega^*$ and deterministic $\alpha_0 > 0$, the following holds:

$$\lim_{t \to \infty} \varphi(t, \theta, \omega, \psi(\theta, \omega)) = \pi(\omega), \quad \omega \in \Omega^*.$$  

(109)

Appendix B

Proofs in Sections II and V

Proof of Proposition 6: For part (i), we note that, by the independence of $\{A(t)\}$, for any $t \in T_+$ and $l_1, \ldots, l_\ell \in V$,

$$P[z_{n}(t) = l_1, \ldots, z_{n}(1) = l_1, z_0(0) = l_0] = P[z_{n}(t) = l_1] z_{n}(t-1) = l_1] = P[A_{l_1, 0} l_1 \mid \ldots \mid A\{t-1\} l_1 = l_0] = \tilde{A}_l l_{l-1} l_1]$$

(110)

where the last step follows from the fact, that the entries of $\{A(t-1)\}$ are binary. This establishes the desired Markovianity of the sequence $\{z_{n}(t)\}_{t \in T_+}$.

For part (ii), since the state space $V$ is finite, the irreducibility of $\tilde{A}$ implies its positive recurrence and hence the invariant measure (the uniform distribution on $V$) is unique. That this measure is attracting follows from the aperiodicity of $\tilde{A}$.

Proof of Lemma 15: The proof is obtained by constructing an approximate filter with suboptimal performance and then bounding its error by using the rank condition on the Grammian $C_{y_{n}}$. We detail such a construction now.

Consider $\ell$ steps $t = 1, \ldots, \ell$ of the linear time-varying signal/observation model given by 1 and 2, where in 2 we index the observation matrix as $C_{y_{n(t)}}$ where $y_{n(t)}$ indicates the current state of the walk $u_{n(t)}$. The signal vector $x_{h} \in \mathbb{R}^{M}$ with initial state $x_{1}$ is a Gaussian random variable with known mean $\mu_{x}$ and variance $X \in \mathbb{S}_{+}^{N}$. The system noise process $\{w_{t}\}$ is uncorrelated zero mean Gaussian with covariance $Q$. The observation noise process $\{v_{t}\}_{t=1}^{\ell}$ is uncorrelated zero mean Gaussian with time varying error covariance $R_{n(t)}$ and independent of the initial signal and the system noise process. By the above construction, the optimal estimate of the signal state $x_{n(t)}$ at time $t$, based on observations till that time, is given by the Kalman filter initialized with $X$ as the predicted conditional error covariance at time $t = 1$. In other words, the optimal m.m.s.e. state estimator (predictor form)

$$\hat{x}_{n(t)} = E[x_{n(t)} | y_{s}]_{1 \leq s \leq t}$$

(111)

of $x_{n(t)}$ based on observations $\{y_{s}\}_{1 \leq s \leq t}$ for $1 \leq t \leq \ell + 1$ can be recursively constructed through the Kalman filter and the corresponding predicted conditional error covariance sequence $\{P_{n(t)} \}_{1 \leq t \leq \ell + 1}$ satisfies the recursion:

$${P_{n(t+1)} = P_{n(t)} + K \ (F_{n(t)} P_{n(t)} F_{n(t)}^{T} + Q - K P_{n(t)} K^{T}) \ (F_{n(t)} P_{n(t)} F_{n(t)}^{T} + R_{n(t)})^{-1} C_{y_{n(t)}} P_{n(t)} F_{n(t)}^{T}}$$

(122)

with initial condition $P_{n(t)}(1) = X$. We then have

$$P_{n(t+1)} = f_{n(t)} \cdots f_{n(t)}(X)$$

(113)

the R.H.S. being the desired functional form in (37), i.e., $P_{n(t+1)} = g_{n(t)}(X)$. Since for a Kalman filter with deterministic system/observation matrices, the conditional error covariance is equal to the unconditional one and the fact that the Kalman filter minimizes any positive definite form of the estimation error, for a generic estimator $\hat{h}$ of $x_{n(t)}$ based on $\{y_{s}\}_{1 \leq s \leq t}$ we have

$$P_{n(t+1)} = E \left[ (x_{n(t)} - \hat{h}) (x_{n(t)} - \hat{h})^{T} \right]$$

(114)

where $\hat{h}$ refers to the partial order on $\mathbb{S}_{+}^{N}$. In order to upper bound the functional $g_{n(t)}$, we now construct a suboptimal state estimator with a guaranteed estimation performance. To this end, define the modified Grammian

$$\tilde{C}_{y_{n(t)}} = \sum_{t=1}^{\ell} (F_{n}^{T})^{t} C_{y_{n(t)}} R_{n(t)}^{-1} C_{y_{n(t)}} F_{n}^{T}$$

(115)

We note that $\tilde{C}_{y_{n(t)}}$ is invertible by the invertibility of $G_{y_{n(t)}}$ and the noise covariances $R_{n(t)}$. Define the suboptimal estimator of $x_{n(t)}$ by:

$$\tilde{x}_{n(t)}(t+1) = \phi_{n(t)} \sum_{t=1}^{\ell} (F_{n}^{T})^{t} C_{y_{n(t)}} R_{n(t)}^{-1} y_{t}$$

(116)

based on observations $\{y_{s}\}_{1 \leq s \leq t}$. Using the fact, that,

$$x_{t} = F_{n(t)} x_{1} + \sum_{s=1}^{t} (F_{n(t)}^{T})^{s} w_{s}, \quad 1 \leq t \leq \ell + 1$$

(117)

we have from (116)

$$\tilde{x}_{n(t)}(t+1)$$

$$\phi_{n(t)} \sum_{t=1}^{\ell} (F_{n(t)}^{T})^{t} C_{y_{n(t)}} R_{n(t)}^{-1} C_{y_{n(t)}} F_{n(t)}^{T}$$

(118)

$$+ \phi_{n(t)} \sum_{t=1}^{\ell} (F_{n(t)}^{T})^{t} C_{y_{n(t)}} R_{n(t)}^{-1} C_{y_{n(t)}} F_{n(t)}^{T}$$

(119)

$$= \phi_{n(t)} F_{n(t)} x_{1} + \phi_{n(t)} \sum_{t=1}^{\ell} (F_{n(t)}^{T})^{t} C_{y_{n(t)}} R_{n(t)}^{-1} C_{y_{n(t)}} F_{n(t)}^{T}$$

(120)

$$+ \phi_{n(t)} \sum_{t=1}^{\ell} (F_{n(t)}^{T})^{t} C_{y_{n(t)}} R_{n(t)}^{-1} C_{y_{n(t)}} F_{n(t)}^{T}$$

(121)

$$+ \phi_{n(t)} \sum_{t=1}^{\ell} (F_{n(t)}^{T})^{t} C_{y_{n(t)}} R_{n(t)}^{-1} C_{y_{n(t)}} F_{n(t)}^{T}$$

(122)

$$x_{t} = F_{n(t)} x_{1} + \sum_{s=1}^{t} (F_{n(t)}^{T})^{s} w_{s}, \quad 1 \leq t \leq \ell + 1$$

(123)

We have from (116)
The filtering error is then given by
\[
\mathbf{e}_{\mathbf{u}_0}(\ell + 1) = \mathbf{x}_{\ell+1} - \mathbf{z}_{\mathbf{u}_0}(\ell + 1) = - \mathcal{F}^{t-1} \mathcal{R}_{\mathbf{u}_0}^{-1} \sum_{i=1}^{\ell} (\mathcal{F}_{t-1})^i \mathbf{c}_{n_t} \mathcal{R}_{n_t}^{-1} \mathbf{c}_{n_t} \times \sum_{i=s}^{\ell-1} \mathcal{F}_{t-1}^{i-s} \mathbf{w}_s - \mathcal{F}^{t-1} \mathcal{R}_{\mathbf{u}_0}^{-1} \sum_{i=1}^{\ell} (\mathcal{F}_{t-1})^i \mathbf{c}_{n_t} \mathcal{R}_{n_t}^{-1} \mathbf{v}_{n_t}. 
\]
(118)

We note that the error above is independent of the initial state \( \mathbf{x}_1 \) (and hence the covariance \( \mathcal{X} \)) of the system and the mean square boundedness of the process noise \( \{ \mathbf{w}_t \} \) and observation noise \( \{ \mathbf{v}_t \} \) imply the existence of a constant \( \alpha_0 > 0 \), such that,
\[
\mathbb{E}[\mathbf{e}_{\mathbf{u}_0}(\ell + 1) \mathbf{e}_{\mathbf{u}_0}(\ell + 1)^T] \leq \alpha_0 I.
\]
(119)

The Lemma then follows by the optimality of the Kalman filter, as stated in (114).

**Proof of Lemma 16:** In case \( \mathcal{F} \) is stable, the claim is obvious, as the suboptimal estimate of 0 at each sensor for all time is stochastically bounded. So, in the sequel we assume \( \mathcal{F} \) is unstable.

The proof is somewhat technical and mainly uses the uniform boundedness of the composition of Riccati operators in Lemma 15 and the ergodicity of the underlying switching Markov chain \( \{ \mathcal{Z}(t) \}_{t \in \mathbb{T}_n} \). From Lemma 15 it follows that a successive application of \( \mathcal{F} \) Riccati maps (in the composition order \( f_{n_1} \circ \cdots \circ f_{n_1} \)) reduces the iterate in the cone interval \([0, \alpha_0 I] \) irrespective of its initial value. The approach is to relate the probability of large exceedance of \( \tilde{Z}_1 \) to the hitting time statistics of a modified Markov chain. We detail it below.

First, we note that the regularity of the distributions of \( \tilde{P}(t) \) for every \( t \), implies that it suffices to show
\[
\lim_{J \to \infty} \sup_{t \geq 0} \mathbb{P}(\| \tilde{P}(t) \| > J) = 0
\]
(120)
for some arbitrarily large \( t_0 \in \mathbb{T}_n \). For every \( n, \) the Riccati update is upper bounded by the Lyapunov operator, i.e.,
\[
f_n(\mathcal{X}) \leq \mathcal{F}_n \mathcal{F}^T + \mathcal{Q}, \quad \forall \mathcal{X} \in \mathcal{S}_n^+.
\]
(121)

For sufficiently large \( J > 0 \), define
\[
k(J) = \max_k \left\{ k \in \mathbb{T}_n \left| \alpha^{2k} \alpha_0 + \frac{\alpha^{2k} - 1}{\alpha^2 - 1} \| \mathcal{Q} \| \leq J \right. \right\}
\]
(122)
where \( \alpha = \| \mathcal{F} \|. \) Since \( \mathcal{F} \) is unstable (\( \alpha > 1 \)), we note that \( k(J) \to \infty \) as \( J \to \infty \).

We introduce additional notation here. For integers \( t_0, t_1 \geq \ell \), the phrase “there exists a \( (n_1, n_2, \ldots, n_\ell) \) cycle in the interval \([t_0, t_1] \)” indicates the existence of an integer \( t_0 \leq \ell \leq t_1 \), such that,
\[
\zeta(t - \ell + s) = n_s, \quad 1 \leq s \leq \ell
\]
(123)
where \( \{ \zeta(t) \}_{t \in \mathbb{T}_n} \) is the switching Markov chain.

We now make the following claim for relating the probabilities of interest for sufficiently large \( t \):
\[
\mathbb{P}(\| \tilde{P}(t) \| > J) \leq \mathbb{P}(\text{no } (n_1, n_2, \ldots, n_\ell) \text{ exists in } [t - k(J), t]).
\]
(124)
Indeed, assume on the contrary that a \( (n_1, \ldots, n_\ell) \) cycle exists in the interval \([t - k(J), t]\). Then there exists \( \ell \in [t - k(J), t] \), such that
\[
\zeta(t - \ell + s) = n_s, \quad 1 \leq s \leq \ell.
\]
(125)
This implies
\[
\tilde{P}(t) = f_{n_1} \circ \cdots \circ f_{n_\ell}(\tilde{P}(t - \ell + 1))
\]
(126)
and hence by Lemma 15
\[
\tilde{P}(t) \leq \alpha_0 I
\]
(127)
which holds irrespective of the value of \( \tilde{P}(t - \ell + 1) \). By (121), we note that
\[
\tilde{P}(s) \leq \mathcal{F} \tilde{P}(s - 1) \mathcal{F}^T + \mathcal{Q}, \quad \forall s.
\]
(128)
Continuing the recursion and noting \( \tilde{P}(t) \leq \alpha_0 I \), we have
\[
\| \tilde{P}(t) \| \leq \alpha^{2(t-s)} \| \tilde{P}(t) \| + \frac{\alpha^{2(t-s)} - 1}{\alpha^2 - 1} \| \mathcal{Q} \|
\]
\[
= \alpha^{2(t-s)} \alpha_0 + \frac{\alpha^{2(t-s)} - 1}{\alpha^2 - 1} \| \mathcal{Q} \|
\]
Since \((t - s) \leq k(J)\), it follows from the above
\[
\| \tilde{P}(t) \| \leq \alpha^{2(t-s)} \alpha_0 + \frac{\alpha^{2(t-s)} - 1}{\alpha^2 - 1} \| \mathcal{Q} \|
\]
\[
\leq \alpha^{2k(J)} \alpha_0 + \frac{\alpha^{2k(J)} - 1}{\alpha^2 - 1} \| \mathcal{Q} \| \leq J
\]
where the last step follows from the definition of \( k(J) \) (see (122)). We thus note that the existence of a \( (n_1, \ldots, n_\ell) \) cycle in \([t - k(J), t]\) implies \( \| \tilde{P}(t) \| \leq J \), i.e., we have the event inclusion
\[
\{ \text{there exists a } (n_1, \ldots, n_\ell) \text{ cycle in } [t - k(J), t] \} \subset \left\{ \| \tilde{P}(t) \| \leq J \right\}.
\]
(129)
The claim in (124) follows. Thus estimating the probability on the L.H.S. of (124) reduces to estimating the probability of a \( (n_1, \ldots, n_\ell) \) cycle in \([t - k(J), t]\). To this end we construct another Markov chain \( \{ z(t) \}_{t \geq \ell} \). The state space \( \mathcal{Z} \) is a subset of \( V^\ell \) given by:
\[
\mathcal{Z} = \{ z = (i_1, i_2, \ldots, i_\ell) \mid \mathcal{A}_{i_j, i_{j+1}} > 0, \quad 1 \leq j < \ell \}.
\]
(130)
The dynamics of the Markov chain \( \{ z(t) \}_{t \geq \ell} \) is given in terms of the Markov chain \( \{ \zeta(t) \}_{t \in \mathbb{T}_n} \) as follows:
\[
z(t) = (\zeta(t - \ell + 1), \zeta(t - \ell + 2), \ldots, \zeta(t)).
\]
(131)
\[ P \left( \text{no } (n_1, \ldots, n_L) \text{ exists in } [t - k(J), t] \right) = P \left( \tau^J_t > t \right) = \sum_{z \in \mathcal{Z}} \left[ P \left( z(t - k(J) - 1) = z \right) P \left( \tau^J_t > t \mid z(t - k(J) - 1) = z \right) \right] = \sum_{z \in \mathcal{Z}} \left[ P \left( z(t - k(J) - 1) = z \right) P_z \left( \tau_0 > k(J) + 1 \right) \right]. \] (137)

From the dynamics of \( \{\tilde{z}(t)\}_{t \in \mathbb{T}_+} \) it follows that \( \{z(t)\}_{t \geq \ell} \) is a Markov chain with transition probability \( \tilde{A}_t \) between allowable states \( (i_1, i_2, \ldots, i_{\ell-1}, n) \) and \( (i_2, \ldots, i_{\ell}, n, I) \). With state space \( \mathcal{Z} \), the Markov chain \( \{z(t)\} \) inherits irreducibility and aperiodicity from that of \( \{\tilde{z}(t)\} \). Also, \( \{z(t)\} \) is stationary from the stationarity of \( \{\tilde{z}(t)\} \) with invariant distribution:

\[ P \left( z(t) = (i_1, i_2, \ldots, i_{\ell}) \right) = \frac{1}{N} \prod_{j=1}^{\ell-1} \tilde{A}_{i_j, i_{j+1}}, (i_1, i_2, \ldots, i_{\ell}) \in \mathcal{Z}, \, t \geq \ell, \, t \in \mathbb{T}_+. \] (132)

Denote the hitting time \( \tau_0 \) of \( \{z(t)\} \) to the state \( (n_1, \ldots, n_L) \) by:

\[ \tau_0 = \min \left\{ t > \ell : z(t) = (n_1, \ldots, n_L) \right\} \] (133)

and for all \( z \in \mathcal{Z} \) define

\[ P_z \left( \tau_0 > s \right) = P \left( \tau_0 > s \mid z(t) = z \right). \] (134)

Also, for each \( t \geq \ell \) and \( J \) sufficiently large, define the stopping times

\[ \tau^J_t = \min \left\{ t \geq t - k(J) : \text{\#(z(t) = (n_1, \ldots, n_L))} \right\}. \] (135)

From the Markov property it then follows:

\[ P \left( \tau^J_t > t \mid z(t - k(J) - 1) = z \right) = P_z \left( \tau_0 > k(J) + 1 \right). \] (136)

It then follows successively as shown in (137) at the top of the page. Since the above development holds for all \( t \geq t_0 \) for some sufficiently large \( t_0 \), we conclude from (124)

\[ \sup_{t \geq t_0} P \left( \left\| \tilde{P}(t) \right\| > J \right) \leq \sum_{z \in \mathcal{Z}} P \left( z(t - k(J) - 1) = z \right) P_z \left( \tau_0 > k(J) + 1 \right). \] (138)

The recurrence (in fact positive recurrence) of the finite state Markov chain \( \{z(t)\} \) and the fact that \( k(J) \to \infty \) as \( J \to \infty \) imply, for all \( z \in \mathcal{Z} \),

\[ \lim_{J \to \infty} P_z \left( \tau_0 > k(J) + 1 \right) = 0. \] (139)

Since \( \mathcal{Z} \) is finite, letting \( J \to \infty \) in (138) leads to

\[ \lim_{J \to \infty} \sup_{t \geq t_0} P \left( \left\| \tilde{P}(t) \right\| > J \right) = 0 \] (140)

by the dominated convergence theorem and the Lemma follows.
[19] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, “On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage,” *Automatica*, vol. 42, no. 2, pp. 251–260, Feb. 2006.

[20] Y. Xu and J. Hespanha, “Estimation under controlled and uncontrolled communications in networked control systems,” in *Proc. IEEE Conf. Decision Contr.*, Sevilla, Spain, Dec. 2005, pp. 842–847.

[21] M. Huang and S. Dey, “Stability of Kalman filtering with Markovian packet losses,” *Automatica*, vol. 43, no. 4, pp. 598–607, Apr. 2007.

[22] K. Plarre and F. Bullo, “On Kalman filtering for detectable systems with Markovian binary switching,” *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 386–390, Feb. 2009.

[23] C. Robinson and P. R. Kumar, “Sending the most recent observation is not optimal in networked control: Linear temporal coding and towards the design of a control specific transport protocol,” in *Proc. IEEE CDC*, Dec. 2007, pp. 334–339.

[24] L. Xie and L. Xie, “Stability of a random Riccati equation with Markovian binary switching,” *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1759–1764, 2008.

[25] S. Kar, B. Sinopoli, and J. M. F. Moura, “Kalman filtering with intermittent observations: Weak convergence to a stationary distribution,” *IEEE Trans. Autom. Control*, Aug. 2009, submitted for publication.

[26] S. Kar and J. M. F. Moura, “Moderate deviations of the random Riccati equation,” Oct. 2009 [Online]. Available: http://arxiv.org/abs/0910.4686

[27] A. Censi, “On the performance of Kalman filtering with intermittent observations: A geometric approach with fractals,” in *Amer. Control Conf.*, 2009.

[28] A. Vakili and B. Hassibi, “A Stieltjes transform approach for studying the steady-state behavior of random Lyapunov and Riccati recursions,” in *Proc. IEEE CDC*, Dec. 2008, pp. 453–458.

[29] P. Bougerol, “Kalman filtering with random coefficients,” *SIAM J. Contr. Optim.*, vol. 31, no. 4, pp. 942–959, Jul. 1993.

[30] V. Solo, “Stability of the Kalman filter with stochastic time-varying parameters,” in *Proc. 35th IEEE Conf. Decision Contr.*, Dec. 1996, pp. 57–61.

[31] J. Jacod and A. Shiryaev, *Limit Theorems for Stochastic Processes*. Berlin, Germany: Springer-Verlag, 1987.

[32] I. Chueshov, *Monotone Random Systems—Theory and Applications*. New York: Springer, 2002.

[33] P. Diaconis and D. Freedman, “Iterated random functions,” *SIAM Rev.*, vol. 41, no. 1, pp. 45–76, 1999.

[34] L. Arnold, *Random Dynamical Systems*. New York: Springer, 1998.

[35] O. Kallenberg, *Foundations of Modern Probability*, ser. Springer Series in Statistics, 2nd ed. New York: Springer, 2002.

[36] S. Ethier and T. Kurtz, *Markov Processes: Characterization and Convergence*. New York: Wiley, 1986.

[37] L. Arnold and I. Chueshov, “Order-preserving random dynamical systems: Equilibria, attractors, applications,” *Dynam. Stability Syst.*, vol. 13, pp. 265–280, 1998.

Soummya Kar (S’05–M’10) received the B. Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in May 2005 and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in June 2010.

He is currently a Postdoctoral Researcher at Princeton University, Princeton, NJ.

His research interests include performance analysis and inference in large-scale networked systems, stochastic filtering, stochastic approximation and large deviations.

José M. F. Moura (S’71–M’75–SM’90–F’94) received the engenheiro electrotécnico degree from Instituto Superior Técnico (IST), Lisbon, Portugal and the M.Sc., E.E., and D.Sc. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge.

He is University Professor at Carnegie Mellon University (CMU). He was on the faculty at IST, held visiting faculty appointments at MIT and was a visiting research scholar at the University of Southern California. He is a founding co-director of the Center for Sensed Critical Infrastructures Research (CenSCIR) and manages a large education and research program between CMU and Portugal, www.icti.cmu.edu.

His research interests include statistical and algebraic signal processing, image, bioimaging and video processing and digital communications. He has published over 400 technical Journal and Conference papers, is the co-editor of two books, holds seven patents and has given numerous invited seminars at international conferences, United States and European Universities and industrial and government Laboratories.

Dr. Moura is the IEEE Division IX Director Elect and was the President (2008–2009) of the IEEE Signal Processing Society (SPS), Editor in Chief for the IEEE TRANSACTIONS IN SIGNAL PROCESSING, interim Editor in Chief for the IEEE SIGNAL PROCESSING LETTERS and was on the Editorial Board of several Journals, including the IEEE PROCEEDINGS, the IEEE SIGNAL PROCESSING MAGAZINE and the ACM Transactions on Sensor Networks. He was on the steering and technical committees of several Conferences. He is a Fellow of the IEEE, a Fellow of the American Association for the Advancement of Science (AAAS), and a corresponding member of the Academy of Sciences of Portugal (Section of Sciences). He was awarded the 2003 IEEE Signal Processing Society Meritorious Service Award and in 2000 the IEEE Millennium Medal. In 2007, he received the CMU’s College of Engineering Outstanding Research Award and in 2009 the 2008 Philip L. Dowd Fellowship Award for Contributions to Engineering Education. In 2010, he was elected University Professor at CMU.