

Stochastic Average Consensus Filter for Distributed HMM Filtering: Almost Sure Convergence^{*}

Nader Ghasemi^{*} Subhrakanti Dey^{*} John S. Baras^{**}

^{*} *Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Melbourne, VIC 3010, Australia (e-mail: n.ghasemi@ee.unimelb.edu.au, sdey@unimelb.edu.au).*

^{**} *Institute for Systems Research and Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA, (e-mail: baras@umd.edu)*

Abstract: In this paper, we study almost sure convergence of a dynamic average consensus algorithm which allows distributed computation of the product of n time-varying conditional probability density functions. These density functions (often called as “belief functions”) correspond to the conditional probability of observations given the state of an underlying Markov chain, which is observed by n different nodes within a sensor network. The topology of the sensor network is modeled as an undirected graph. The average consensus algorithm is used to obtain a distributed state estimation scheme for a hidden Markov model (HMM). We use the ordinary differential equation (ODE) technique to analyze the convergence of a stochastic approximation type algorithm for achieving average consensus with a constant step size. It is shown that, for a connected graph, under mild assumptions on the first and second moments of the observation probability densities and a geometric ergodicity condition on an extended Markov chain, the consensus filter state of each individual sensor converges almost surely to the true average of the logarithm of the belief functions of all the sensors. Convergence is proved by using a perturbed stochastic Lyapunov function technique. Numerical results suggest that the distributed estimates of the Markov chain state obtained at the individual sensor nodes based on this consensus algorithm track the centralized state estimate (computed on the basis of having access to the observations of all the nodes) quite well, while more formal results on convergence of the distributed HMM filter to the centralized one are currently under investigation.

Keywords: Convergence analysis, stochastic approximation, state estimation, stochastic stability, asymptotic properties.

1. INTRODUCTION

The study of distributed estimation algorithms in a network of spatially distributed sensor nodes has been the subject of extensive research. A fundamental problem in distributed estimation is to design *scalable* estimation algorithms for multi-sensor networked systems such that the data of a sensor node is communicated only to its immediate neighbor nodes. This is in contrast to the centralized estimation where the data from all the sensors are transmitted to a central unit, known as the fusion center, where the task of data fusion is performed. The centralized scheme, clearly, is not energy-efficient in terms of message exchange. Also, this approach makes the estimation algorithms susceptible to single point failure. Moreover, for a large scale network, performing a centralized estimation algorithm at the fusion center may not be computationally

feasible. As such, the centralized approach is not robust and also not efficient in terms of both computation and communication.

Recently, designing scalable distributed estimation algorithms using consensus schemes has attracted significant surge of interest. For this, consensus filters are used to combine the individual node data in a way that every node can compute an approximation to a quantity, which is based on data from all the nodes, by using input data only from its nearest neighbors. Then, by decomposing the centralized algorithm into some subalgorithms where each subalgorithm can be implemented using a consensus algorithm, each node can run a distributed algorithm which relies only on the data from its neighboring nodes. The problem, then, is to study how close the distributed estimate is to the estimate obtained by the centralized algorithm.

Some pioneering works in distributed estimation were done by Borkar and Varaiya (1982) and Tsitsiklis et al. (1986). Recently, there has been many studies on the use of consensus algorithms in distributed estimation, see, e.g.,

^{*} This work was supported in part by the Australian Research Council (ARC) under Grant ARC DP 0985397 and the Robert Bage Memorial Scholarship. This research was performed while N. Ghasemi was a visiting scholar at the University of Maryland, College Park.

distributed Kalman filtering in Olfati-Saber (2005), Olfati-Saber and Shamma (2005), Carli et al. (2008), approximate Kalman filter in Spanos et al. (2005), linear least square estimator in Xiao et al. (2005), and distributed information filtering in Casbeer and Beard (2009).

This paper will focus on analyzing asymptotic properties of a stochastic approximation type algorithm for average consensus introduced in Olfati-Saber and Shamma (2005). In particular, we study almost sure convergence of the algorithm proposed in Olfati-Saber and Shamma (2005). Using the proposed dynamic average consensus algorithm, we compute the product of n time-varying conditional probability density functions, known as beliefs, corresponding to n different nodes within a sensor network. The stochastic approximation algorithm uses a constant step size to track the time-varying average of the logarithm of the belief functions. We use the ordinary differential equation (ODE) technique¹ in stochastic approximation to study almost sure convergence of the consensus algorithm. In order to prove convergence, we use a stochastic stability method where we introduce a perturbed stochastic Lyapunov function to show that the error between the consensus filter state at each node and the true (centralized) average enters some compact set infinitely often w.p.1. Then, using this result and stability of the mean ODE it is shown that the error is bounded w.p.1. This is then used towards proving almost sure convergence of the consensus algorithm.

2. PROBLEM STATEMENT

Notations: In this paper, \mathbb{R} denotes the set of real numbers and \mathbb{N} and \mathbb{Z}^+ represent the sets of positive and nonnegative integers, respectively. We denote by C^n the class of n -times continuously differentiable functions. Let (Ω, \mathcal{F}) be a measurable space consisting of a sample space Ω and the corresponding σ -algebra \mathcal{F} of subsets of Ω . The symbol ω denotes the canonical point in Ω . Let \mathbb{P} represent probability distribution with respect to some σ -finite measure and \mathbb{E} denote the expectation with respect to the probability measure \mathbb{P} . By $\mathbf{1}_n$, and $\mathbf{0}_n$ we denote n -dimensional² vectors with all elements equal to one, and zero respectively. Let \mathbf{I} denote the identity matrix of proper dimension. Let $\|\cdot\|_p$ denote the p -norm on a Euclidean space. In this paper, vector means a column vector, and $'$ denotes the transpose notation.

2.1 Distributed Filtering Model: Preliminaries & Notations

Let a stochastic process $\{X_k, k \in \mathbb{Z}^+\}$, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, represent a discrete time homogeneous Markov chain with transition probability matrix $\mathbf{X} = [x_{ij}]$ and finite state space $\mathcal{S} = \{1, \dots, s\}$, $s \in \mathbb{N}$, where $x_{ij} = \mathbb{P}(X_k = j | X_{k-1} = i)$ for $i, j \in \mathcal{S}$. Assume that $s > 1$ is fixed and known. Note that \mathbf{X} is a stochastic matrix, that is, $x_{ij} \geq 0$, $\sum_j x_{ij} = 1, \forall i \in \mathcal{S}$. The initial probability distribution of $\{X_k\}$ is denoted by $\pi = [\pi_i]_{i \in \mathcal{S}}$, where $\pi_i = \mathbb{P}(X_0 = i)$.

The Markov process $\{X_k\}$ is assumed to be hidden and observed indirectly through noisy measurements obtained

¹ see Benveniste et al. (1990) and Kushner and Yin (2003).

² for convenience, the dimension subscript n may be omitted when it is clear from the context.

by a set of sensor nodes. Consider a network of spatially distributed sensor nodes, observing the Markov process $\{X_k\}$, where the network topology is represented by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, with $\mathcal{N} = \{1, \dots, n\}$, $n \in \mathbb{N}$ denoting the set of vertices (nodes) and $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$ representing the set of edges. An edge between node i and j is denoted by an unordered pair $(i, j) \in \mathcal{E}$. In this paper, all graphs are assumed undirected and simple (with no self-loop), i.e., for every edge $(i, j) \in \mathcal{E}$, $i \neq j$. The set of neighbors of node j is denoted by $\mathcal{N}_j = \{i \in \mathcal{N} | (i, j) \in \mathcal{E}\}$.

For each node $m \in \mathcal{N}$, the sequence of observations is denoted by $\{Y_k^m, k \in \mathbb{Z}^+\}$, which is a sequence of conditionally independent random variables given a realization $\{x_k\}$ of $\{X_k\}$. The conditional probability distribution of the observed data Y_k^m , taking values in \mathbb{R}^q , given the Markov chain state $X_k = \ell, \ell \in \mathcal{S}$ is assumed to be absolutely continuous with respect to a nonnegative and σ -finite measure ϱ on \mathbb{R}^q , with the density function $f_\ell^m(\cdot)$, where $\mathbb{P}(Y_k^m \in dy | X_k = \ell) = f_\ell^m(y)\varrho(dy)$, $\ell \in \mathcal{S}$. Let \mathbf{Y}_k^m , adapted to \mathcal{Y}_k^m , denote the sequence of observed data at node $m \in \mathcal{N}$ up to time instant k , where $\mathcal{Y}_k^m = \sigma(Y_l^m, 0 \leq l \leq k)$ is the σ -algebra generated by the corresponding random observations. Define also \mathbf{Y}_k , measurable on \mathcal{Y}_k , as the random vector of the observations obtained by all n number of sensors at time k , where $\mathcal{Y}_k = \sigma(Y_l^m, 1 \leq m \leq n)$ is the corresponding σ -algebra. We introduce the following assumption:

A-1. The observations $\mathbf{Y}_k = [Y_k^m]_{m \in \mathcal{N}}$ are mutually conditionally independent with respect to the node index m given the Markov chain state $X_k = \ell, \ell \in \mathcal{S}$.

We specify an HMM corresponding to the observation sequence $\{\mathbf{Y}_k, k \in \mathbb{Z}^+\}$ by $\mathcal{H} \triangleq (\mathbf{X}, \mathcal{S}, \pi, \Psi)$, where we define the matrix $\Psi(\mathbf{y}) = \text{diag}[\psi_i(\mathbf{y})]_{i \in \mathcal{S}}$, with i -th diagonal element $\psi_i(\mathbf{y})$ called state-to-observation probability density function for the Markov chain state $X_k = i$.

2.2 Distributed Information State Equations

For $k \in \mathbb{Z}^+$, define the *centralized* information state vector or normalized filter $\bar{\mathbf{v}}_k = [\bar{v}_k(j)]_{j \in \mathcal{S}}$, as the conditional probability mass function of the Markov chain state X_k given the observed data from all n number of nodes up to time k , that is, $\bar{v}_k(j) \triangleq \mathbb{P}(X_k = j | \mathcal{Y}_k^1, \dots, \mathcal{Y}_k^n)$ for each $j \in \mathcal{S}$.

Clearly, in the centralized estimation scenario, where each node transmits its observations to a (remote) fusion center, $\bar{\mathbf{v}}_k$ can be computed at the fusion center using the received measurements from all the sensors. However, in the distributed scenario, in order to compute the centralized filter $\bar{\mathbf{v}}_k$ at each node, \mathcal{G} must be a complete graph which may not be a practical assumption for most (large scale) sensor networks. A practical approach is to express the filter equation in terms of summations of the individual node observations or some function of the observations, as shown in the following lemma. Each node, then, can approximate those summations using dynamic average consensus filters by exchanging appropriate messages only with its immediate neighbors. In this way, the communication costs for each sensor are largely reduced which leads to a longer life time of the overall network. It is clear, however, that without the knowledge of all the sensors' measurements

and distribution models, each node may only be able to find an *approximation* to the centralized filter $\bar{\mathbf{v}}_k$. The following lemma presents the equivalent *distributed* form of the centralized filter equations.

Lemma 1. Assume **A-1**. For a given sequence of the sensors' observations $\{\mathbf{y}_k\}$, where $\mathbf{y}_k = [y_k^1, \dots, y_k^n]'$ and for any $\ell \in \mathcal{S}$, the centralized filter $\bar{v}_k(\ell)$ satisfies the following recursion:

$$\bar{w}_k^\ell(\mathbf{y}_k) = n^{-1} \langle \mathbf{1}_n, \mathbf{z}_k^\ell(\mathbf{y}_k) \rangle, \quad k \in \mathbb{Z}^+,$$

$$v_0(\ell) = e^{-n\bar{w}_0^\ell \pi_\ell},$$

$$v_k(\ell) = e^{-n\bar{w}_k^\ell} \sum_{i=1}^s x_{i\ell} v_{k-1}(i), \quad k \in \mathbb{N},$$

$$\bar{v}_k(\ell) = \langle \mathbf{1}_s, \mathbf{v}_k \rangle^{-1} v_k(\ell), \quad k \in \mathbb{Z}^+,$$

where $\mathbf{v}_k = [v_k(\ell)]_{\ell \in \mathcal{S}}$ is the unnormalized centralized filter, $\mathbf{z}_k^\ell = [z_k^\ell(j)]_{j \in \mathcal{N}} \triangleq [-\log f_\ell^1(y_k^1), \dots, -\log f_\ell^n(y_k^n)]'$ is the vector of sensors' contributions.

From Lemma 1, assuming the knowledge of HMM parameters $(\mathbf{X}, \mathcal{S}, \pi)$ at each node, the centralized filter $\bar{v}_k(\ell)$ may be computed exactly with no error if the average quantity \bar{w}_k^ℓ is known exactly at each node. It is clear, however, that in a distributed scenario, this average quantity could be calculated with no error only for a complete graph with all-to-all communication topology. In practice, for other network topologies, each node may only be able to compute an approximation to \bar{w}_k^ℓ by exchanging appropriate messages only with its neighboring nodes. A possible approach to approximate \bar{w}_k^ℓ at each node is to run a dynamic average consensus filter for every $\ell \in \mathcal{S}$.

2.3 Stochastic Approximation Algorithm for Consensus Filter

In the following, we present a stochastic approximation algorithm for estimating the centralized (time-varying) quantity $\bar{w}_k^\ell \in \mathbb{R}^+$ as the average of the vector elements $z_k^\ell(j)$, $j \in \mathcal{N}$ defined in Lemma 1. Since the same algorithm is performed for every Markov chain state $\ell \in \mathcal{S}$, to simplify the notation, henceforth we omit the superscript dependence on the Markov chain state, e.g., $\bar{w}_k^\ell, \mathbf{z}_k^\ell = [z_k^\ell(j)]$ will be simply denoted by $\bar{w}_k, \mathbf{z}_k = [z_k(j)]$ respectively.

Let the consensus filter state for node $i \in \mathcal{N}$ at time $k \in \mathbb{Z}^+$ be denoted by \hat{w}_k^i which is, in fact, the node's estimate of the centralized (or true) average \bar{w}_k . Let $\hat{\mathbf{w}}_k = [\hat{w}_k^i]_{i \in \mathcal{N}}$ denote the vector of all the nodes' estimates. Each node i employs a stochastic approximation algorithm to estimate \bar{w}_k using the input messages $z_k(j)$ and consensus filter states \hat{w}_k^j only from its immediate neighbors, that is, $j \in \mathcal{N}_i \cup \{i\}$. For $k \in \mathbb{Z}^+$, the state of each node $i \in \mathcal{N}$ is updated using the following algorithm introduced in Olfati-Saber and Shamma (2005):

$$\hat{w}_k^i = (1 + \rho q_{ii}) \hat{w}_{k-1}^i + \rho (\mathbf{A}_i \hat{\mathbf{w}}_{k-1} + \mathbf{A}_i \mathbf{z}_k + z_k(i)) \quad (1)$$

where ρ is a fixed small scalar gain called step size, \mathbf{A}_i is i -th row of the matrix $\mathbf{A} = [a_{ij}]_{i,j \in \mathcal{N}}$ which specifies the interconnection topology³ of the network, and the parameter q_{ii} is defined by $q_{ii} \triangleq -(1 + 2\mathbf{A}_i \mathbf{1})$. Precise conditions on the step size ρ will be introduced later.

Definition 1. Strong Average Consensus : Consider a stochastic process $\{\mathbf{Z}_k, k \in \mathbb{Z}^+\}$ with a given realization

³ in this paper, it is assumed that $a_{ij} > 0$ for $j \in \mathcal{N}_i$ and is zero otherwise.

$\{\mathbf{z}_k = \mathbf{Z}_k(\omega), \omega \in \Omega\}$, where $\mathbf{z}_k = [z_k(i)]_{i \in \mathcal{N}}$ is the random data assigned to the set \mathcal{N} of nodes at time k . It is said that all the nodes have reached strong consensus with respect to the average of the input vector \mathbf{z}_k if for $\bar{w}_k^* \triangleq n^{-1} \langle \mathbf{1}, \mathbf{z}_k \rangle$, the condition $\lim_{k \rightarrow \infty} (\hat{w}_k^i - \bar{w}_k^*) = 0$ \mathbb{P} -a.s. is satisfied uniformly in $i \in \mathcal{N}$.

We may write (1) in the form

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + \rho (\mathbf{A} \hat{\mathbf{w}}_{k-1} + \mathbf{\Gamma} \mathbf{z}_k), \quad k \in \mathbb{Z}^+ \quad (2)$$

where the matrices $\mathbf{A}, \mathbf{\Gamma}$ are defined⁴ by $\mathbf{A} \triangleq \text{diag}[q_{ii}]_{i \in \mathcal{N}} + \mathbf{A}$ and $\mathbf{\Gamma} \triangleq \mathbf{I} + \mathbf{A}$, and the initial condition $\hat{\mathbf{w}}_{-1}$ may be chosen as an arbitrary vector $\hat{\mathbf{w}}_{-1} \triangleq c \mathbf{1}$, for some $c \in \mathbb{R}^+$.

3. CONVERGENCE ANALYSIS OF THE CONSENSUS ALGORITHM

3.1 Preliminary Assumptions

A-2. For any $\ell \in \mathcal{S}$, and $k \in \mathbb{Z}^+$, the conditional probability distribution of the observed data \mathbf{Y}_k given the Markov chain state $X_k = \ell$ is absolutely continuous with respect to a nonnegative and σ -finite measure $\bar{\rho}$ on appropriate Euclidean space, with $\bar{\rho}$ -a.e. positive density $\psi_\ell(\cdot)$, where $\mathbb{P}(\mathbf{Y}_k \in d\mathbf{y} \mid X_k = \ell) = \psi_\ell(\mathbf{y}) \bar{\rho}(d\mathbf{y})$.

A-3. The transition probability matrix $\mathbf{X} = [x_{ij}]$ of the Markov chain $\{X_k, k \in \mathbb{Z}^+\}$ is primitive⁵.

Remark 1. Under **A-2, A-3**, the extended Markov chain $\{(X_k, \mathbf{Y}_k), k \in \mathbb{Z}^+\}$ is geometrically ergodic⁶ with a unique invariant measure $\nu_\circ = [\nu_\circ^\ell]_{\ell \in \mathcal{S}}$ on $\mathcal{S} \times \mathbb{R}^{nq}$, $\nu_\circ^\ell(d\mathbf{y}) = \gamma_\circ^\ell \psi_\ell(\mathbf{y}) \bar{\rho}(d\mathbf{y})$ for any $\ell \in \mathcal{S}$, where $\gamma_\circ = [\gamma_\circ^\ell]_{\ell \in \mathcal{S}}$ defined on \mathcal{S} is the unique stationary probability distribution of the Markov chain $\{X_k, k \in \mathbb{Z}^+\}$.

Define the stochastic process $\{\eta_k, k \in \mathbb{Z}^+\}$, where the error $\eta_k = [\eta_k^i]_{i \in \mathcal{N}}$, defined as $\eta_k \triangleq \hat{\mathbf{w}}_k - \bar{\mathbf{w}}_k^*$, is the error between the consensus filter state and average of the nodes' data $\bar{\mathbf{w}}_k^* \triangleq \bar{w}_k^* \mathbf{1}$ at time k . For notational convenience, let $\xi_k \triangleq (\mathbf{z}_k, \mathbf{z}_{k-1})$ adapted to \mathcal{O}_k denote the extended data, where \mathcal{O}_k is the σ -algebra generated by $(\mathbf{Y}_k, \mathbf{Y}_{k-1})$.

Lemma 2. For a given sequence $\{\mathbf{z}_k(\mathbf{y}_k)\}$, where $\mathbf{y}_k \in \mathcal{Y}_k$, the error vector η_k evolves according to the following stochastic approximation algorithm:

$$\eta_{k+1} = \eta_k + \rho \mathbf{Q}(\eta_k, \xi_{k+1}), \quad k \in \mathbb{Z}^+ \quad (3)$$

where $\mathbf{Q}(\cdot)$ is a measurable function⁷, which determines how the error is updated as a function of new input \mathbf{z}_{k+1} , defined⁸ by

$$\begin{aligned} \mathbf{Q}(\eta_k, \xi_{k+1}) &\triangleq \mathbf{A} \eta_k + \mathbf{\Gamma} (\mathbf{z}_{k+1} - n^{-1} \mathbf{1} \mathbf{1}' \mathbf{z}_k) \\ &\quad - (n\rho)^{-1} \mathbf{1} \mathbf{1}' (\mathbf{z}_{k+1} - \mathbf{z}_k) \end{aligned} \quad (4)$$

⁴ Note that for the undirected graph \mathcal{G} , the matrix $-(\mathbf{A} + \mathbf{\Gamma})$ is positive-semidefinite with $\mathbf{1}$ as an eigenvector corresponding to the trivial eigenvalue $\lambda_0 = 0$.

⁵ equivalently, the Markov chain is irreducible and aperiodic.

⁶ see LeGland and Mevel (2000).

⁷ note that for each $(\mathbf{z}, \bar{\mathbf{z}})$, $\mathbf{Q}(\cdot, \mathbf{z}, \bar{\mathbf{z}})$ is a C^0 -function in η on \mathbb{R}^n .

⁸ The argument may be verified by using the algorithm (2) and the equality $\mathbf{A} \mathbf{1} = -\mathbf{\Gamma} \mathbf{1}$ for the undirected graph \mathcal{G} .

3.2 Mean ODE

In the following, we define, for $t \in \mathbb{R}$, a continuous time interpolation $\boldsymbol{\eta}_\bullet(t)$ of the sequence $\{\boldsymbol{\eta}_k\}$ in terms of the step size ρ . Let $t_0 = 0$ and $t_k = k\rho$. Define the map $\alpha(t) = k$, for $t \geq 0, t_k \leq t < t_{k+1}$, and $\alpha(t) = 0$ for $t < 0$. Define the piecewise constant interpolation $\boldsymbol{\eta}_\bullet(t)$ on $t \in (-\infty, \infty)$ with interpolation interval ρ as follows: $\boldsymbol{\eta}_\bullet(t) = \boldsymbol{\eta}_k$, for $t \geq 0, t_k \leq t < t_{k+1}$ and $\boldsymbol{\eta}_\bullet(t) = \boldsymbol{\eta}_0$ for $t \leq 0$. Define also the sequence of shifted processes $\boldsymbol{\eta}_\bullet^k(t) = \boldsymbol{\eta}_\bullet(t_k + t)$ for $t \in (-\infty, \infty)$.

Define mean vector field $\bar{\mathbf{Q}}(\boldsymbol{\eta})$ as the limit average of the function $\mathbf{Q}(\cdot)$ by $\bar{\mathbf{Q}}(\boldsymbol{\eta}) \triangleq \lim_{k \rightarrow \infty} \mathbb{E}_\boldsymbol{\eta} \mathbf{Q}(\boldsymbol{\eta}, \boldsymbol{\xi}_k)$ where $\mathbb{E}_\boldsymbol{\eta}$ denotes the expectation with respect to the distribution of $\boldsymbol{\xi}_k$ for a fixed $\boldsymbol{\eta}$. In order to analyze the asymptotic properties of the error iterates $\boldsymbol{\eta}_k$ in (3), we define the ODE determined by the mean dynamics as

$$\dot{\boldsymbol{\eta}}_\bullet = \bar{\mathbf{Q}}(\boldsymbol{\eta}_\bullet), \quad \boldsymbol{\eta}_\bullet(0) = \boldsymbol{\eta}_0 \quad (5)$$

where $\boldsymbol{\eta}_0$ is the initial condition. In Proposition 3, we present a strong law of large numbers to specify the mean vector field $\bar{\mathbf{Q}}(\cdot)$.

Define $\chi_{(\iota)} \triangleq [\chi_\iota(i)]_{i \in \mathcal{N}}$, where

$$\chi_\iota(i) \triangleq \max_{j \in \mathcal{S}} \int \left[\max_{\ell \in \mathcal{S}} | \log f_\ell^i(y^i) | \right]^\iota f_j^i(y^i) \varrho(dy^i)$$

$$\Delta_{(\iota)} \triangleq \max_{j \in \mathcal{S}} \int \left[\max_{\ell \in \mathcal{S}} | \log \psi_\ell(\mathbf{y}) | \right]^\iota \psi_j(\mathbf{y}) \bar{\varrho}(d\mathbf{y})$$

and the average $\bar{\mathbf{Q}}_k(\boldsymbol{\eta}) \triangleq (k+1)^{-1} \sum_{l=0}^k \mathbf{Q}(\boldsymbol{\eta}, \boldsymbol{\xi}_l)$.

Proposition 3. Assume conditions **A-2** and **A-3**. If $\Delta_{(1)}$ is finite, then there exists a finite $\bar{\mathbf{Q}}(\boldsymbol{\eta})$ such that

$$\lim_{k \rightarrow \infty} \bar{\mathbf{Q}}_k(\boldsymbol{\eta}) = \bar{\mathbf{Q}}(\boldsymbol{\eta}) \quad \mathbb{P}\text{-a.s.}$$

is satisfied uniformly in $\boldsymbol{\eta}$, where

$$\bar{\mathbf{Q}}(\boldsymbol{\eta}) = \boldsymbol{\Lambda} \boldsymbol{\eta} + \boldsymbol{\Gamma}(\bar{\mathbf{z}} - n^{-1} \mathbf{1} \mathbf{1}' \bar{\mathbf{z}}) \quad (6)$$

and $\bar{\mathbf{z}} = [\bar{z}(i)]_{i \in \mathcal{N}}$, in which we have

$$\bar{z}(i) = \int | \log f_\ell^i(y^i) | \mu_\circ^i(dy^i), \quad \ell \in \mathcal{S} \quad (7)$$

with μ_\circ^i denoting the marginal density of the invariant measure ν_\circ for node $i \in \mathcal{N}$ defined on \mathbb{R}^q .

We now establish the global asymptotic ϵ -stability of the ODE (5) in sense of the following definition.

Definition 2. A set \mathbb{E}^* is said to be asymptotically ϵ -stable for the ODE (5) if for each $\epsilon_1 > 0$ there exists an $\epsilon_2 > 0$ such that all trajectories $\boldsymbol{\eta}(t)$ of the ODE (5) with initial condition $\boldsymbol{\eta}_\bullet(0)$ in an ϵ_2 -neighborhood of \mathbb{E}^* will remain in an ϵ_1 -neighborhood of \mathbb{E}^* and ultimately converge to an ϵ -neighborhood of \mathbb{E}^* . If this holds for the set of all initial conditions, then \mathbb{E}^* is globally asymptotically ϵ -stable.

A-4. There exists a real-valued C^1 -function $V(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}$ of $\boldsymbol{\eta}_\bullet$ such that $V(\mathbf{0}) = 0$, $V(\boldsymbol{\eta}_\bullet) > 0$ for $\boldsymbol{\eta}_\bullet \neq \mathbf{0}$ and $V(\boldsymbol{\eta}_\bullet) \rightarrow \infty$ as $\|\boldsymbol{\eta}_\bullet\| \rightarrow \infty$.

A-5. For any trajectory $\boldsymbol{\eta}_\bullet(\cdot)$ solving the ODE (5) for which the initial condition $\boldsymbol{\eta}_\bullet(0)$ lies in $\mathbb{R}^n \setminus \Omega_c$, where Ω_c is a compact level set defined by $\Omega_c \triangleq \{\boldsymbol{\eta}_\bullet : V(\boldsymbol{\eta}_\bullet(t)) \leq c\}$, for some $0 < c < \infty$, $\dot{V}(\boldsymbol{\eta}_\bullet(t))$ is strictly negative.

Proposition 4. Consider the ODE (5). Assume **A-4**. In particular, consider the Lyapunov function $V(\boldsymbol{\eta}_\bullet) = \frac{1}{2} \boldsymbol{\eta}'_\bullet \boldsymbol{\eta}_\bullet$.

Also, assume **A-5** holds for some compact set Ω_{c° , where $c^\circ = \frac{1}{2} \epsilon^2$ for some $\epsilon > 0$. Then, the origin is globally asymptotically ϵ -stable for the mean ODE (5), with ϵ given by $\epsilon = 2\bar{\nu} \sqrt{n(1 + d_{max})} |\lambda_{max}(\boldsymbol{\Lambda})|^{-1}$, where $\bar{\nu} \triangleq \max_{i \in \mathcal{N}} \bar{z}(i)$.

Proof. See Appendix A, Ghasemi et al. (2010).

3.3 Stochastic Stability of the Consensus Error Iterates

Since the error iterates $\boldsymbol{\eta}_k$ in (3) are not known to be bounded a priori and not confined to a compact constraint set, in this section, we use a stochastic stability method to prove that the sequence $\{\boldsymbol{\eta}_k\}$ is *recurrent*, which means that the error process $\{\boldsymbol{\eta}_k\}$ visits some compact set $\Omega_{\bar{c}} \triangleq \{\boldsymbol{\eta} : V(\boldsymbol{\eta}(t)) \leq \bar{c}\}$, $0 < \bar{c} < \infty$ infinitely often \mathbb{P} -w.p.1. In order to prove that some compact set $\Omega_{\bar{c}}$ is recurrent, we introduce a perturbed stochastic Lyapunov function in which the Lyapunov function of the mean ODE is slightly perturbed in a way that the resulting stochastic Lyapunov function has the supermartingale property. The Doob's martingale convergence theorem is then used to show that the compact set $\Omega_{\bar{c}}$ is reached again \mathbb{P} -w.p.1 after each time the error process $\{\boldsymbol{\eta}_k\}$ exits $\Omega_{\bar{c}}$.

Define the filtration $\{\mathcal{F}_k, k \in \mathbb{Z}^+\}$ as a sequence of non-decreasing sub- σ -algebras of \mathcal{F} defined as $\mathcal{F}_k \triangleq [\mathcal{F}_k^i]_{i \in \mathcal{N}}$ such that for each $i \in \mathcal{N}$, $\mathcal{F}_k^i \subset \mathcal{F}_{k+1}^i$ is satisfied for all $k \in \mathbb{Z}^+$, and \mathcal{F}_k^i measures at least $\sigma(\boldsymbol{\eta}_0^i, \mathbf{Y}_k^j, j \in \mathcal{N}_i \cup \{i\})$. Let \mathbb{E}_k denote the conditional expectation given \mathcal{F}_k . For $i \geq k$, define the discount factor β_k^i by $\beta_k^i \triangleq (1 - \rho)^{i-k+1}$ and the empty product $\beta_k^i \triangleq 1$ for $i < k$.

Define the discounted perturbation $\delta\vartheta_k(\boldsymbol{\eta}) : \mathbb{R}^n \mapsto \mathbb{R}^n$ as follows:

$$\delta\vartheta_k(\boldsymbol{\eta}) = \sum_{i=k}^{\infty} \rho \beta_{k+1}^i \mathbb{E}_k[\mathbf{Q}(\boldsymbol{\eta}, \boldsymbol{\xi}_{i+1}) - \bar{\mathbf{Q}}(\boldsymbol{\eta})] \quad (8)$$

In view of the fact that $\sup_k \sum_{i=k}^{\infty} \rho \beta_{k+1}^i < \infty$, the sum in the discounted perturbation (8) is well defined and we have⁹

$$\mathbb{E}_k \delta\vartheta_{k+1}(\boldsymbol{\eta}) = \sum_{i=k+1}^{\infty} \rho \beta_{k+2}^i \mathbb{E}_k[\mathbf{Q}(\boldsymbol{\eta}, \boldsymbol{\xi}_{i+1}) - \bar{\mathbf{Q}}(\boldsymbol{\eta})] \quad \mathbb{P}\text{-w.p.1}$$

Define the perturbed stochastic Lyapunov function

$$V_k(\boldsymbol{\eta}_k) \triangleq V(\boldsymbol{\eta}_k) + \nabla_{\boldsymbol{\eta}_k} V(\boldsymbol{\eta}) \delta\vartheta_k(\boldsymbol{\eta}_k) \quad (9)$$

where $\nabla_{\boldsymbol{\eta}_k} V(\boldsymbol{\eta}) = \nabla V(\boldsymbol{\eta}) |_{\boldsymbol{\eta}=\boldsymbol{\eta}_k}$, with $\nabla V(\boldsymbol{\eta})$ denoting the gradient of $V(\cdot)$. Note that $V_k(\boldsymbol{\eta}_k)$ is \mathcal{F}_k -measurable.

A-6. Let there be positive numbers $\{b_i, i \in \mathcal{N}\}$ and define $\mathbf{b} \triangleq [b_i^{-2}]_{i \in \mathcal{N}}$ such that $b_n \rightarrow \infty$ for large n . In particular, let $b_n = n$. Let the following series

$$\langle \mathbf{b}, \chi_{(2)} \rangle - \langle \mathbf{b}, \chi_{(1)}^2 \rangle \quad (10)$$

converge for sufficiently large n .

A-7. The step size ρ is strictly positive¹⁰ satisfying the condition $\rho < 2(1 + 3d_{max})^{-1}$.

⁹ cf. (Kushner and Yin, 2003, Chapter 6, Section 6.3.2)

¹⁰ note that ρ must be kept strictly away from zero in order to allow \dot{w}_k^i to track the time varying true average \bar{w}_k , see Kushner and Yin (2003) for further detail.

The following theorem establishes a sufficient condition for recurrence of the error iterates η_k .

Theorem 5. Consider the unconstrained stochastic approximation algorithm (3). Assume conditions **A-1**, **A-2**, **A-3**, and **A-6** hold. Let the real-valued Lyapunov function $V(\cdot)$ of the mean ODE (5) have bounded second mixed partial derivatives and satisfy condition **A-4**. Also, assume $\Delta_{(1)}$ and $\Delta_{(2)}$ are finite and let the step size ρ satisfy condition **A-7**. Then, the perturbed stochastic Lyapunov function $V_k(\eta_k)$ is an \mathcal{F}_k -supermartingale for the stopped process η_k when η_k first visits some compact set $\Omega_{\bar{c}} \triangleq \{\eta : V(\eta(t)) \leq \bar{c}\}$, for $\bar{c} \in (0, \infty)$.

Proof. See Appendix B, Ghasemi et al. (2010).

The following theorem establishes the recurrence of the error iterates η_k .

Theorem 6. Consider the perturbed stochastic Lyapunov function $V_k(\eta_k)$ defined in (9). Let $V_k(\eta_k)$ be a real-valued supermartingale with respect to the filtration \mathcal{F}_k . Assume that $\mathbb{E}V(\eta_0)$ is bounded. Then, for any $\delta \in (0, 1]$, there is a compact set \mathbb{L}_δ such that the iterates η_k enter \mathbb{L}_δ infinitely often with probability at least δ .

Proof. See Appendix C, Ghasemi et al. (2010).

3.4 Almost Sure Convergence of the Consensus Algorithm

In this section, we use the recurrence result of the previous section in combination with an ODE-type method to prove almost sure convergence of the error sequence $\{\eta_k\}$ under rather weak conditions¹¹. The ODE method shows that asymptotically the stochastic process $\{\eta_k\}$, starting at the recurrence times when η_k enters the compact recurrence set $\Omega_{\bar{c}}$, converges to the largest bounded invariant set of the mean ODE (5) contained in $\Omega_{\bar{c}}$. Therefore, if the origin is globally asymptotically ϵ -stable for the mean ODE (5) with some invariant level set Ω_{c° , where $c^\circ < \bar{c}$, then $\{\eta_k\}$ converges to an ϵ -neighborhood of the origin \mathbb{P} -w.p.1.

The following lemma establishes a nonuniform regularity condition on the function $Q(\cdot, \xi)$ in η .

Lemma 7. There exist nonnegative measurable functions $\mathfrak{h}_1(\cdot)$ and $\mathfrak{h}_2(\cdot)$ of η and ξ , respectively, such that $\mathfrak{h}_1(\cdot)$ is bounded on each bounded η -set and

$$\|Q(\eta, \xi) - Q(\tilde{\eta}, \xi)\| \leq \mathfrak{h}_1(\eta - \tilde{\eta})\mathfrak{h}_2(\xi) \quad (11)$$

where $\mathfrak{h}_1(\eta) \rightarrow 0$ as $\eta \rightarrow \mathbf{0}$ and for some $\bar{\tau} > 0$, \mathfrak{h}_2 satisfies

$$\mathbb{P}\left[\limsup_l \sum_{k=l}^{\alpha(t_l + \bar{\tau})} \rho \mathfrak{h}_2(\xi_k) < \infty\right] = 1 \quad (12)$$

Proof. See Appendix D, Ghasemi et al. (2010).

A-8. For each η , let the rate of change of

$$Q_\eta^\circ(t) \triangleq \sum_{i=0}^{\alpha(t)-1} \rho [Q(\eta, \xi_{i+1}) - \bar{Q}(\eta)]$$

go to zero \mathbb{P} -w.p.1 as $t \rightarrow \infty$. This means the asymptotic rate of change condition¹²

¹¹ for example, the square summability condition on the step size ρ is not needed.

¹² see Section 5.3 and 6.1, Kushner and Yin (2003) for further detail.

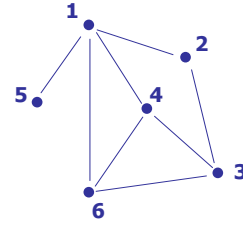


Fig. 1. Network topology \mathcal{G}

$$\limsup_k \max_{j \geq k} \max_{0 \leq t \leq T} |Q_\eta^\circ(jT + t) - Q_\eta^\circ(jT)| = 0 \quad \mathbb{P}\text{-w.p.1}$$

is satisfied uniformly in η for every $T > 0$.

Theorem 8. Consider the unconstrained stochastic approximation algorithm (3). For any $\delta \in (0, 1]$, let there be a compact set \mathbb{L}_δ such that the iterates η_k return to \mathbb{L}_δ infinitely often with probability at least δ . Assume conditions **A-4** and **A-5**. Then, $\{\eta_k\}$ is bounded \mathbb{P} -w.p.1, that is,

$$\limsup_k \|\eta_k\| < \infty \quad \mathbb{P}\text{-w.p.1}$$

Assume condition **A-8**. Also, assume that the function $Q(\cdot, \xi)$ satisfies the nonuniform regularity condition in η established in Lemma 7. Then, there exists a null set $\bar{\mathcal{U}}$ such that for $\omega \notin \bar{\mathcal{U}}$, the set of functions $\{\eta_k^\bullet(\omega, \cdot), k < \infty\}$ is equicontinuous. Let $\eta(\omega, \cdot)$ denote the limit of some convergent subsequence $\{\eta_k^{k'}(\omega, \cdot)\}$. Then, for \mathbb{P} -almost all $\omega \in \Omega$, the limits $\eta(\omega, \cdot)$ are trajectories of the mean ODE (5) in some bounded invariant set and the error iterates $\{\eta_k\}$ converge to this invariant set. Moreover, let the origin be globally¹³ asymptotically ϵ -stable¹⁴ for the mean ODE (5) with some invariant level set Ω_{c° , where $\Omega_{c^\circ} \subset \mathbb{L}_1$. Then, $\{\eta_k\}$ converges to the ϵ -neighborhood of the origin \mathbb{P} -w.p.1 as $k \rightarrow \infty$.

Proof. The proof follows from (Kushner and Yin, 2003, Theorem 7.1 and Theorem 1.1, Chapter 6) and for brevity the details are omitted here.

4. NUMERICAL RESULTS

In this section, we numerically evaluate the performance of the distributed HMM filter computed using the average consensus algorithm (1), and study its average behavior relative to the centralized filter. To this end, we present some numerical results for distributed estimation over a sensor network with the irregular topology \mathcal{G} depicted in Fig. 1. We consider a dynamical system whose state evolves according to a four-state Markov chain $\{X_k, k \in \mathbb{Z}^+\}$ with state space $\mathcal{S} = \{-7.3, -1.2, 2.1, 4.9\}$ and transition kernel

$$\mathbf{X} = \begin{bmatrix} 0.80 & 0.10 & 0 & 0.10 \\ 0.05 & 0.90 & 0.05 & 0 \\ 0 & 0.10 & 0.85 & 0.05 \\ 0.05 & 0 & 0.10 & 0.85 \end{bmatrix}$$

The initial distribution of $\{X_k\}$ is chosen as an arbitrary vector $\pi = [0.20, 0.15, 0.30, 0.35]$. The Markov process $\{X_k\}$ is observed by every node j according to

¹³ note that in case of local asymptotic stability, convergence result holds if \mathbb{L}_δ is in the domain of attraction of the ODE equilibrium.

¹⁴ this is shown in Proposition 4.

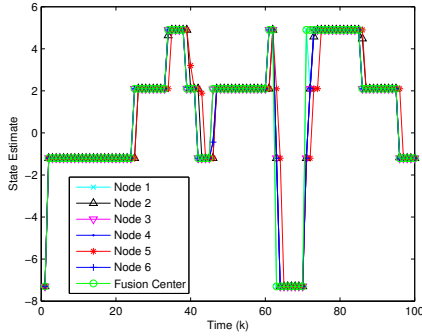


Fig. 2. Distributed and centralized state estimates

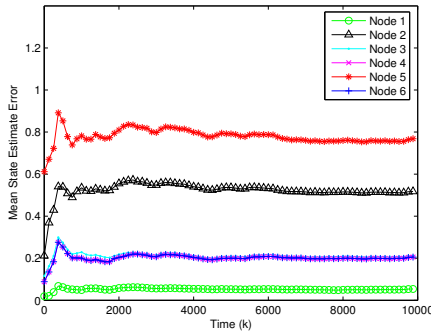


Fig. 3. Convergence in mean of the distributed state estimate at each node to the centralized one.

$Y_k^j = X_k + u_k^j$, where the measurement noises $\{\mathbf{u}_k = [u_k^j]_{j \in \mathcal{N}}, k \in \mathbb{Z}^+\}$ are assumed to be zero-mean white Gaussian noise processes with the noise variance vector $[0.29 + 0.01j]_{j \in \mathcal{N}}$. The initial condition $\hat{\mathbf{w}}_{-1}$ is chosen $\hat{\mathbf{w}}_{-1} = c\mathbf{1}$, with $c = 3$.

Fig. 2 shows the distributed (or local) estimate $\hat{X}_k^j \triangleq \mathbb{E}^j[X_k | \mathcal{F}_k^j]$ of the Markov chain state $\{X_k\}$ at each node $j \in \mathcal{N}$, where the expectation \mathbb{E}^j is with respect to distributed filter $\hat{\mathbf{v}}_k^j = [\hat{v}_k^j(\ell)]_{\ell \in \mathcal{S}}$ computed using the average consensus filter (1). Although node 5 and 2 have direct access to only one and two nodes' observations respectively, they maintain an estimate of $\{X_k\}$ but with some time delay. The reason is because these two nodes receive the observations of other nodes in the network indirectly through the consensus algorithm which incur some delay. Nevertheless, every node follows the state transition of the Markov process $\{X_k\}$ at each time k .

Fig. 3 shows the convergence in mean of the local state estimate \hat{X}_k^j for each node j to the centralized state estimate \hat{X}_k obtained by using the observations of all the nodes. The mean state estimate error is computed as the time average $\bar{g}_k^j \triangleq (k+1)^{-1} \sum_{i=0}^k |\hat{X}_i^j - \hat{X}_i|$. This is done based on the fact that \bar{g}_k^j converges \mathbb{P} -a.s. to the expectation¹⁵ $\mathbb{E} |\hat{X}_k^j - \hat{X}_k|$. This is due to the geometric ergodicity of the extended Markov chain $\{(X_k, \mathbf{Y}_k), k \in \mathbb{Z}^+\}$. As it can be seen for each node j , the average \bar{g}_k^j converges to a $\bar{\delta}^j$ -ball around the origin as $k \rightarrow \infty$. The radius $\|\bar{\delta}\|$, where

¹⁵here we have used the standard notion of convergence in mean.

$\bar{\delta} = [\bar{\delta}^j]_{j \in \mathcal{N}}$ and the rate of convergence, though, depends on how well connected the network is. Precise results on the exact nature of convergence of the distributed HMM filter to the centralized HMM filter and the corresponding proof of convergence are currently under investigation.

ACKNOWLEDGEMENTS

This work was supported in part by the Australian Research Council (ARC) under Grant ARC DP 0985397 and the Robert Bage Memorial Scholarship. This research was performed while N. Ghasemi was a visiting scholar at the University of Maryland, College Park.

REFERENCES

- Benveniste, A., Métivier, M., and Priouret, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, Berlin Heidelberg.
- Borkar, V. and Varaiya, P.P. (1982). Asymptotic agreement in distributed estimation. *IEEE Transactions on Automatic Control*, AC-27(3), 650–655.
- Carli, R., Chiuso, A., Schenato, L., and Zampieri, S. (2008). Distributed Kalman filtering based on consensus strategies. *IEEE Journal on Selected Areas in Communications*, 26(4), 622–633.
- Casbeer, D.W. and Beard, R. (2009). Distributed information filtering using consensus filters. In *Proceedings of the 2009 American Control Conference (ACC 2009)*, 1882–1887. St. Louis, MO, USA.
- Ghasemi, N., Dey, S., and Baras, J.S. (2010). Stochastic Average Consensus Filter for Distributed HMM Filtering: Almost Sure Convergence. Technical report, *Institute for Systems Research (ISR) Technical report*, TR-2010-11, <http://hdl.handle.net/1903/10069>.
- Kushner, H.J. and Yin, G.G. (2003). *Stochastic approximation and recursive algorithms and applications*. Springer-Verlag New York, Inc., vol. 35, in Applications of Mathematics, New York, NY, USA, 2nd edition.
- LeGland, F. and Mevel, L. (2000). Exponential Forgetting and Geometric Ergodicity in Hidden Markov Models. *Math. of Control, Signals, and Systems*, 13, 63–93.
- Olfati-Saber, R. (2005). Distributed Kalman Filter with Embedded Consensus Filters. In *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, 8179–8184. Seville, Spain.
- Olfati-Saber, R. and Shamma, J.S. (2005). Consensus filters for sensor networks and distributed sensor fusion. In *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, 6698–6703. Seville, Spain.
- Spanos, D.P., Olfati-Saber, R., and Murray, R.M. (2005). Approximate distributed Kalman filtering in sensors networks with quantifiable performance. In *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks*, 133–139. CA, USA.
- Tsitsiklis, J.N., Bertsekas, D.P., and Athans, M. (1986). Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9), 803–812.
- Xiao, L., Boyd, S., and Lall, S. (2005). A scheme for robust distributed sensor fusion based on average consensus. In *Proceedings of the fourth International Symposium on Information Processing in Sensor Networks*, 63–70.