THE INSTITUTE FOR SYSTEMS RESEARCH

ISR TECHNICAL REPORT 2010-11

Stochastic Average Consensus Filter for Distributed HMM Filtering: Almost Sure Convergence

Nader Ghasemi
Subhrakanti Dey
John S. Baras

Stochastic Average Consensus Filter for Distributed HMM Filtering: Almost Sure Convergence

Nader Ghasemi, *Student Member, IEEE,* Subhrakanti Dey, *Senior Member, IEEE,* and John S. Baras, *Fellow, IEEE*

***Abstract*—This paper studies 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 conditional probability 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 network topology 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), where each sensor node computes a conditional probability estimate of the state of the Markov chain based on its own observations and the messages received from its immediate neighbors. 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. This allows each node to track the time varying average of the logarithm of conditional observation probabilities available at the individual nodes in the network. 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 $\mathbb{P}$–a.s. to the true average of the logarithm of the conditional observation probability density functions of all the sensors. Convergence is proved by using a perturbed stochastic Lyapunov function technique. Numerical results suggest that the distributed Markov chain state estimates obtained at the individual sensor nodes based on this consensus algorithm track the centralized state estimate (computed on the basis of having access to observations of all the nodes) quite well, while formal results on convergence of the distributed HMM filter to the centralized one are currently under investigation.**

I. 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 where 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 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 [1] and [2]. Recently, there has been many studies on the use of consensus algorithms in distributed estimation, see, e.g., distributed Kalman filtering in [3], [4], [5], [6], approximate Kalman filter in [7], linear least square estimator in [8], and distributed information filtering in [9].

This paper will focus on analyzing asymptotic properties of a stochastic approximation type algorithm for dynamic average consensus introduced in [4]. Using the 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) technique[1] 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 average enters some compact set infinitely often $\mathbb{P}$–w.p.1. Then, using this result and stability of the mean ODE it is shown that the error process is bounded $\mathbb{P}$–w.p.1. This is then used towards proving almost sure convergence of the consensus algorithm.

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

N. Ghasemi and S. Dey are with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, Melbourne, VIC 3010, Australia. n.ghasemi@ee.unimelb.edu.au, sdey@unimelb.edu.au

J. S. Baras is with the Institute for Systems Research and the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA. baras@umd.edu

[1] see [10] and [11].

The outline of the paper is as follows. In Section II, we present the model for distributed HMM filtering and introduce the stochastic approximation algorithm for average consensus. Section III introduces required assumptions for convergence and provides convergence analysis of the consensus algorithm. Numerical results are presented in Section IV. Details of the proofs are given in the Appendix.

II. 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 $(\Omega, \mathcal{F})$ 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 $\mathbb{1}_n$, and $\mathbf{0}_n$ we denote n-dimensional[2] vectors with all elements equal to one, and zero respectively. Let $\mathbf{I}$ denote the identity matrix of proper dimension. For readability of the manuscript, matrix/vector symbols are in bold face with their elements presented within brackets [ ], uppercase letters denote random variables and lowercase is used for a realization of a random variable. Let $\|.\|_p$ denote the p-norm on a Euclidean space. In this paper, vector means a column vector, and $'$ denotes the transpose notation.

A. 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 $\mathcal{S} = \{1, \cdots, s\}$, $s \in \mathbb{N}$, where $x_{ij} = \mathbb{P}(X_k = j \mid X_{k-1} = i)$ 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 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 $\mathcal{N} = \{1, \cdots, n\}$, $n \in \mathbb{N}$ 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 $\mathcal{N}_j = \{i \in \mathcal{N} \mid (i, j) \in \mathcal{E}\}$. A k-regular graph is defined as a graph in which every vertex has k neighbors. A k-regular graph on $m = k + 1$ vertices is called a *complete* graph and is denoted by K_m. For convenience, in the following, the names, sensor and node will be used interchangeably. For brevity, an undirected graph will be simply referred to as a graph.

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

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 $f_\ell^m(.)$, where $\mathbb{P}(Y_k^m \in dy \mid X_k = \ell) = f_\ell^m(y)\varrho(dy)$, $\ell \in \mathcal{S}$. 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 $\mathcal{Y}_k = \sigma(Y_k^m, 1 \leq m \leq n)$ 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 $\mathcal{H} \triangleq (\mathbf{X}, \mathcal{S}, \pi, \mathbf{\Psi})$, where we define the matrix $\mathbf{\Psi}(\mathbf{y}) = \mathrm{diag}[\psi_i(\mathbf{y})]_{i \in \mathcal{S}}$, with i-th diagonal element $\psi_i(\mathbf{y})$ called state-to-observation probability density function for the Markov chain state $X_k = i$.

B. 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, $\bar{v}_k(j) \triangleq \mathbb{P}(X_k = j \mid \mathcal{Y}_k^1, \cdots, \mathcal{Y}_k^n)$ 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 K_n 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 2.1: Assume **A**-*1*. For a given sequence of the sensors' observations $\{\mathbf{y}_k\}$, where $\mathbf{y}_k = [y_k^1, \cdots, y_k^n]' \in \mathcal{Y}_k$ and for any $\ell \in \mathcal{S}$, the centralized filter $\bar{v}_k(\ell)$ satisfies the following recursion:

$$\bar{w}_k^\ell(\mathbf{y}_k) = n^{-1}\langle \mathbb{1}_n, \mathbf{z}_k^\ell(\mathbf{y}_k)\rangle, \ k \in \mathbb{Z}^+,$$
$$v_0(\ell) = e^{-n\bar{w}_0^\ell}\pi_\ell \ ,$$

$v_k(\ell) = e^{-n\bar{w}_k^\ell} \sum_{i=1}^s x_{i\ell} v_{k-1}(i),\ k \in \mathbb{N}$,
$\bar{v}_k(\ell) = \langle \mathbb{1}_s, \mathbf{v}_k \rangle^{-1} v_k(\ell),\ k \in \mathbb{Z}^+$,
where $\mathbf{v}_k = [v_k(\ell)]_{\ell \in \mathcal{S}}$ is the unnormalized centralized filter, $\mathbf{z}_k^\ell = [z_k^\ell(j)]_{j \in \mathcal{N}} \triangleq [-\log f_\ell^1(y_k^1), \cdots, -\log f_\ell^n(y_k^n)]'$ is the vector of sensors' contributions.

For any $\ell \in \mathcal{S}$, the random sequence $\{\bar{w}_k^\ell, k \in \mathbb{Z}^+\}$ is, in fact, the arithmetic mean of the individual sensor contributions. From Lemma 2.1, 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 $\bar{w}_k^\ell$ is known exactly at each node. It is clear, however, that in a distributed scenario, this average 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 $\bar{w}_k^\ell$ by exchanging appropriate messages only with its neighboring nodes. A possible approach to approximate $\bar{w}_k^\ell$ at each node is to run a dynamic average consensus filter for every $\ell \in \mathcal{S}$. In the following section, we introduce a stochastic approximation type algorithm for achieving consensus with respect to the average of time-varying (dynamical) inputs $\mathbf{z}_k^\ell$. Next, we focus on studying the asymptotic properties of the dynamic average consensus algorithm which is used in computing a distributed HMM filter as an approximation to the centralized filter. In particular, we study almost sure convergence of the average computed by using the consensus algorithm to the true average $\bar{w}_k^\ell$.

C. Stochastic Approximation Algorithm for Consensus Filter

In the following, we present a stochastic approximation algorithm for estimating centralized 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 2.1. 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, $j \in \mathcal{N}_i \cup \{i\}$. The state of each node $i \in \mathcal{N}$ is updated using the following algorithm (see [4]):

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

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 topology[3] of the network, and the parameter q_{ii} is defined by $q_{ii} \triangleq -(1 + 2\mathbf{A}_i \mathbb{1})$. Precise conditions on the step size ρ will be introduced later. For further details on the consensus algorithm (1) the reader is referred to [4].

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

Definition 1: Strong Average Consensus : Consider a stochastic process $\{\mathbf{Z}_k, k \in \mathbb{Z}^+\}$ with a given realization $\{\mathbf{z}_k = \mathbf{Z}_k(\omega), \omega \in \Omega\}$, where $\mathbf{z}_k = [z_k(i)]_{i \in \mathcal{N}}$ is the vector of 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 random variable $\bar{w}_k^* \triangleq n^{-1} \langle \mathbb{1}, \mathbf{z}_k \rangle$, the condition $\lim_{k \to \infty}\ (\hat{w}_k^i - \bar{w}_k^*) = 0$ $\mathbb{P}$–a.s. is satisfied uniformly in $i \in \mathcal{N}$.

We may write (1) in the form

$$\hat{\mathbf{w}}_k = \Pi_{\mathbb{H}}\ [\hat{\mathbf{w}}_{k-1} + \rho(\mathbf{\Lambda} \hat{\mathbf{w}}_{k-1} + \mathbf{\Gamma} \mathbf{z}_k)],\quad k \in \mathbb{Z}^+ \tag{2}$$

where $\Pi_{\mathbb{H}}$ is the projection onto a constraint set $\mathbb{H}$, the matrices $\mathbf{\Lambda}, \mathbf{\Gamma}$ are defined by $\mathbf{\Lambda} \triangleq \mathrm{diag}[q_{ii}]_{i \in \mathcal{N}} + \mathbf{A}$ and $\mathbf{\Gamma} \triangleq \mathbf{I} + \mathbf{A}$, and the initial condition $\hat{\mathbf{w}}_{-1}$ may be chosen as an arbitrary vector $\hat{\mathbf{w}}_{-1} \triangleq c\mathbb{1}$, for some $c \in \mathbb{R}^+$. It is noted that the iterates $\hat{\mathbf{w}}_k$ are confined to a proper subset $\mathbb{H}$ of the Euclidean space $\mathbb{R}^n$, such that if an iterate ever escapes the constraint set, it is projected back to the closest point in the constraint set. The constraint set $\mathbb{H}$ is assumed to be compact and its elements are admissible vectors satisfying the required constraints.

III. Convergence Analysis of the Consensus Algorithm

In this section, we study the convergence of the average consensus algorithm (2) introduced in the previous section. In what follows, we use the ordinary differential equation (ODE[4]) approach to prove $\mathbb{P}$ w.p.1 convergence of the consensus filter state $\hat{\mathbf{w}}_k$ to the centralized average quantity $\bar{\mathbf{w}}_k^* \triangleq n^{-1}\mathbb{1}\mathbb{1}'\mathbf{z}_k$. In the ODE method, the asymptotic behavior of the discrete time iterates $\hat{\mathbf{w}}_k$ is studied by analyzing asymptotic stability of a continuous time mean ODE, see [10] for further detail.

A. Preliminary Assumptions

We introduce the following 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 $\bar{\varrho}$ on appropriate Euclidean space, with $\bar{\varrho}$-a.e. positive density $\psi_\ell(.)$, where $\mathbb{P}(\mathbf{Y}_k \in d\mathbf{y} \mid X_k = \ell) = \psi_\ell(\mathbf{y})\bar{\varrho}(d\mathbf{y})$.

***A**-3:* The transition probability matrix $\mathbf{X} = [x_{ij}]$ of the Markov chain $\{X_k, k \in \mathbb{Z}^+\}$ is primitive[5] with index of primitivity r.

Remark 1: Under **A**-*2*, **A**-*3*, the extended Markov chain $\{(X_k, \mathbf{Y}_k),\ k \in \mathbb{Z}^+\}$ is geometrically ergodic (see [12]) with a unique invariant measure $\boldsymbol{\nu}_\circ = [\nu_\circ^\ell]_{\ell \in \mathcal{S}}$ on $\mathcal{S} \times \mathbb{R}^{nq}$, $\nu_\circ^\ell(d\mathbf{y}) = \gamma_\circ^\ell \psi_\ell(\mathbf{y})\bar{\varrho}(d\mathbf{y})$ for any $\ell \in \mathcal{S}$, where $\boldsymbol{\gamma}_\circ = [\gamma_\circ^\ell]_{\ell \in \mathcal{S}}$ defined on $\mathcal{S}$ is the unique stationary probability distribution of the Markov chain[6] $\{X_k, k \in \mathbb{Z}^+\}$.

Define the stochastic process $\{\boldsymbol{\eta}_k, k \in \mathbb{Z}^+\}$, where the error $\boldsymbol{\eta}_k = [\eta_k^i]_{i \in \mathcal{N}}$, defined as $\boldsymbol{\eta}_k \triangleq \hat{\mathbf{w}}_k - \bar{\mathbf{w}}_k^*$, is the error

[4]for relevant literature on ODE approach, the reader is referred to [11], [10], and references therein.

[5]equivalently, the Markov chain is irreducible and aperiodic.

[6]note that under **A**-*3*, the Markov chain $\{X_k\}$ is also geometrically ergodic.

between the consensus filter state and average of the nodes' data $\bar{\mathbf{w}}_k^* \triangleq \bar{w}_k^* \mathbb{1}$ at time k. For notational convenience, let $\boldsymbol{\xi}_k \triangleq (\mathbf{z}_k, \mathbf{z}_{k-1})$ adapted to $\mathcal{O}_k$ denote the extended data at time k, where $\mathcal{O}_k$ is the σ-algebra generated by $(\mathbf{Y}_k, \mathbf{Y}_{k-1})$ for $k \in \mathbb{Z}^+$.

Lemma 3.1: For a given sequence $\{\mathbf{z}_k(\mathbf{y}_k)\}$, where $\mathbf{y}_k \in \mathcal{Y}_k$, the error vector $\boldsymbol{\eta}_k$ evolves according to the following stochastic approximation algorithm

$$\boldsymbol{\eta}_{k+1} = \boldsymbol{\eta}_k + \rho \mathbb{Q}(\boldsymbol{\eta}_k, \boldsymbol{\xi}_{k+1}), \quad k \in \mathbb{Z}^+ \tag{3}$$

where $\mathbb{Q}(.)$ is a measurable function[7], which determines how the error is updated as a function of new input $\mathbf{z}_{k+1}$, defined by

$$\begin{aligned}\mathbb{Q}(\boldsymbol{\eta}_k, \boldsymbol{\xi}_{k+1}) \triangleq & \boldsymbol{\Lambda}\boldsymbol{\eta}_k + \boldsymbol{\Gamma}(\mathbf{z}_{k+1} - n^{-1}\mathbb{1}\mathbb{1}'\mathbf{z}_k) \\ & - (n\rho)^{-1}\mathbb{1}\mathbb{1}'(\mathbf{z}_{k+1} - \mathbf{z}_k)\end{aligned} \tag{4}$$

Remark 2: The argument may be verified by using the algorithm (2) and the equality[8] $\boldsymbol{\Lambda}\mathbb{1} = -\boldsymbol{\Gamma}\mathbb{1}$ for the undirected graph $\mathcal{G}$.

B. 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 $\bar{\mathbb{Q}}(\boldsymbol{\eta})$ as the limit average of the function $\mathbb{Q}(.)$ by

$$\bar{\mathbb{Q}}(\boldsymbol{\eta}) \triangleq \lim_{k\to\infty} \mathbb{E}_{\boldsymbol{\eta}} \ \mathbb{Q}(\boldsymbol{\eta}, \boldsymbol{\xi}_k) \tag{5}$$

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

$$\dot{\boldsymbol{\eta}}_\bullet = \bar{\mathbb{Q}}(\boldsymbol{\eta}_\bullet), \qquad \boldsymbol{\eta}_\bullet(0) = \boldsymbol{\eta}_0 \tag{6}$$

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

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

$$\begin{aligned}\chi_\iota(i) &\triangleq \max_{j\in\mathcal{S}} \int \big[\max_{\ell\in\mathcal{S}} \ |\log f_\ell^i(y^i)\,|\,\big]^\iota f_j^i(y^i)\varrho(dy^i) \\ \Delta_{(\iota)} &\triangleq \max_{j\in\mathcal{S}} \int \big[\max_{\ell\in\mathcal{S}} \ |\log \psi_\ell(\mathbf{y})\,|\,\big]^\iota \psi_j(\mathbf{y})\bar{\varrho}(d\mathbf{y})\end{aligned}$$

and the average

$$\bar{\mathbb{Q}}_k(\boldsymbol{\eta}) \triangleq (k+1)^{-1} \sum_{l=0}^{k} \mathbb{Q}(\boldsymbol{\eta}, \boldsymbol{\xi}_l) \tag{7}$$

[7]note that for each $(\mathbf{z}, \tilde{\mathbf{z}})$, $\mathbb{Q}(., \mathbf{z}, \tilde{\mathbf{z}})$ is a C^0-function in $\boldsymbol{\eta}$ on $\mathbb{R}^n$.

[8]note that for the undirected graph $\mathcal{G}$, the matrix $-(\boldsymbol{\Lambda} + \boldsymbol{\Gamma})$ is positive-semidefinite with $\mathbb{1}$ as an eigenvector corresponding to the trivial eigenvalue $\lambda_0 = 0$.

Proposition 3.2: Assume conditions **A-***2* and **A-***3*. If $\Delta_{(1)}$ is finite, then there exists a finite $\bar{\mathbb{Q}}(\boldsymbol{\eta})$ such that

$$\lim_{k\to\infty} \ \bar{\mathbb{Q}}_k(\boldsymbol{\eta}) = \bar{\mathbb{Q}}(\boldsymbol{\eta}) \ \ \mathbb{P}\text{–a.s.}$$

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

$$\bar{\mathbb{Q}}(\boldsymbol{\eta}) = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\Gamma}(\bar{\boldsymbol{z}} - n^{-1}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}}) \tag{8}$$

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

$$\bar{z}(i) = \int \ |\log f_\ell^i(y^i)\,|\mu_\circ^i(dy^i), \ \ell \in \mathcal{S} \tag{9}$$

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

In the following, we establish the global asymptotic ϵ-stability of the mean ODE (6) in sense of the following definition.

Definition 2: A set $\mathbb{E}^*$ is said to be asymptotically ϵ-stable for the ODE (6) if for each $\varepsilon_1 > 0$ there exists an $\varepsilon_2 > 0$ such that all trajectories $\boldsymbol{\eta}(t)$ of the ODE (6) 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.
We introduce the assumptions.

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

A-5: For any trajectory $\boldsymbol{\eta}_\bullet(.)$ solving the ODE (6) 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 $0 < c < \infty$, the derivative $\dot{V}(\boldsymbol{\eta}_\bullet(t))$ is strictly negative.

Proposition 3.3: Consider the ODE (6). 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 (6), with ϵ given by

$$\epsilon = 2\bar{\nu}\sqrt{n}(1 + d_{max})|\ \lambda_{max}(\boldsymbol{\Lambda})\ |^{-1} \tag{10}$$

where $\bar{\nu} \triangleq \max_{i\in\mathcal{N}} \ \bar{z}(i)$.

Proof: See Appendix A. ■

C. 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. Then, in the next section, using this result and the ODE method it is shown that $\{\boldsymbol{\eta}_k\}$ is bounded $\mathbb{P}$–w.p.1 and converges $\mathbb{P}$–w.p.1 to the largest bounded invariant set of the mean ODE (6) contained in $\Omega_{\bar{c}}$. 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 $\Omega_{\bar{c}}$. As the next step, using this result and the stability hypothesis on the mean ODE, it is shown that the error sequence $\{\boldsymbol{\eta}_k\}$ is bounded $\mathbb{P}$–w.p.1.

Define the filtration $\{\mathcal{F}_k, k \in \mathbb{Z}^+\}$ as a sequence of nondecreasing 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 $\sigma(\eta_0^i, \mathbf{Y}_k^j, j \in \mathcal{N}_i \cup \{i\})$. 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:

$$\delta\vartheta_k(\boldsymbol{\eta}) = \sum_{i=k}^{\infty} \rho\beta_{k+1}^i \mathbb{E}_k[\mathbb{Q}(\boldsymbol{\eta}, \boldsymbol{\xi}_{i+1}) - \bar{\mathbb{Q}}(\boldsymbol{\eta})] \tag{11}$$

In view of the fact that $\sup_k \sum_{i=k}^{\infty} \rho\beta_{k+1}^i < \infty$, the sum in the discounted perturbation (11) is well defined and we have[9]

$$\mathbb{E}_k\delta\vartheta_{k+1}(\boldsymbol{\eta}) = \sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_k[\mathbb{Q}(\boldsymbol{\eta}, \boldsymbol{\xi}_{i+1}) - \bar{\mathbb{Q}}(\boldsymbol{\eta})] \quad \mathbb{P}\text{–w.p.1} \tag{12}$$

Define the perturbed stochastic Lyapunov function

$$V_k(\boldsymbol{\eta}_k) \triangleq V(\boldsymbol{\eta}_k) + \nabla_{\boldsymbol{\eta}_k} V(\boldsymbol{\eta})\ \delta\vartheta_k(\boldsymbol{\eta}_k) \tag{13}$$

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

We introduce the assumptions.

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

$$\langle \boldsymbol{b}, \boldsymbol{\chi}_{(2)} \rangle - \langle \boldsymbol{b}, \boldsymbol{\chi}_{(1)}^2 \rangle \tag{14}$$

converge for sufficiently large n.

***A**-7:* The step size ρ is strictly positive[10] satisfying the condition $\rho < 2(1 + 3d_{max})^{-1}$.

The following theorem establishes a sufficient condition for recurrence of the error iterates $\boldsymbol{\eta}_k$.

Theorem 3.4: 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 $V(.)$ of the mean ODE (6) 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 $V_k(\boldsymbol{\eta}_k)$ is an $\mathcal{F}_k$–supermartingale for the stopped process $\boldsymbol{\eta}_k$ when $\boldsymbol{\eta}_k$ first visits some compact set $\Omega_{\bar{c}} \triangleq \{\boldsymbol{\eta} : V(\boldsymbol{\eta}(t)) \leq \bar{c}\}$, for $\bar{c} \in (0, \infty)$.

Proof: See Appendix B. ■

The following theorem establishes the recurrence of the error iterates $\boldsymbol{\eta}_k$.

Theorem 3.5: Consider the perturbed stochastic Lyapunov function $V_k(\boldsymbol{\eta}_k)$ defined in (13). Let $V_k(\boldsymbol{\eta}_k)$ be a real-valued supermartingale with respect to the filtration $\mathcal{F}_k$. Assume that $\mathbb{E}V(\boldsymbol{\eta}_0)$ is bounded. Then, for any $\delta \in (0, 1]$, there is a compact set $\mathbb{L}_\delta$ such that the iterates $\boldsymbol{\eta}_k$ enter $\mathbb{L}_\delta$ infinitely often with probability at least δ.

Proof: See Appendix C. ■

D. Almost Sure Convergence of the Consensus Algorithm

Recall the main result of the previous section, where a stochastic stability method based on a perturbed stochastic Lyapunov function is used to show that the error iterates $\boldsymbol{\eta}_k$ return to some compact set $\Omega_{\bar{c}}$ infinitely often $\mathbb{P}$–w.p.1. In this section, we use this recurrence result in combination with an ODE-type method to prove almost sure convergence of the error sequence $\{\boldsymbol{\eta}_k\}$ under rather weak conditions[11]. The ODE method shows that asymptotically the stochastic process $\{\boldsymbol{\eta}_k\}$, starting at the recurrence times when $\boldsymbol{\eta}_k$ enters the compact recurrence set $\Omega_{\bar{c}}$, converges to the largest bounded invariant set of the mean ODE (6) contained in $\Omega_{\bar{c}}$. Therefore, if the origin is globally asymptotically ϵ-stable for the mean ODE (6) with some invariant level set Ω_{c°, where $c^\circ < \bar{c}$, then $\{\boldsymbol{\eta}_k\}$ converges to an ϵ-neighborhood of the origin $\mathbb{P}$–w.p.1.

The following lemma establishes a nonuniform regularity condition on the function $\mathbb{Q}(., \boldsymbol{\xi})$ in $\boldsymbol{\eta}$ required for the proof of convergence.

Lemma 3.6: There exist nonnegative measurable functions $\mathbb{h}_1(.)$ and $\mathbb{h}_{k2}(.)$ of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, respectively, such that $\mathbb{h}_1(.)$ is bounded on each bounded $\boldsymbol{\eta}$-set and

$$\|\mathbb{Q}(\boldsymbol{\eta}, \boldsymbol{\xi}) - \mathbb{Q}(\tilde{\boldsymbol{\eta}}, \boldsymbol{\xi})\| \leq \mathbb{h}_1(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})\mathbb{h}_{k2}(\boldsymbol{\xi}) \tag{15}$$

where $\mathbb{h}_1(\boldsymbol{\eta}) \to 0$ as $\boldsymbol{\eta} \to \mathbf{0}$ and $\mathbb{h}_{k2}$ satisfies

$$\mathbb{P}\big[\limsup_l \sum_{k=l}^{\alpha(t_l+\bar{\tau})} \rho\mathbb{h}_{k2}(\boldsymbol{\xi}_k) < \infty\big] = 1, \tag{16}$$

for some $\bar{\tau} > 0$.

Proof: By applying Geršgorin theorem to the negative definite matrix $\boldsymbol{\Lambda}$, it is shown that its minimum eigenvalue satisfies $\lambda_{min}(\boldsymbol{\Lambda}) \geq -(1 + 3d_{max})$, where $d_{max} \triangleq \max_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}_i} a_{ij}$. Thus, from (4) we have

$$\|\boldsymbol{\Lambda}(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})\| \leq (1 + 3d_{max})\|\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}\|$$

where choosing $\mathbb{h}_{k2}$ as $\mathbb{h}_{k2}(\boldsymbol{\xi}) = (1 + 3d_{max})$ satisfies condition (16) for any finite $\bar{\tau} > 0$. Moreover, the function $\mathbb{h}_1(\boldsymbol{\eta}) = \|\boldsymbol{\eta}\|_p$ is bounded on each bounded $\boldsymbol{\eta}$-set and tends to 0 as $\boldsymbol{\eta} \to \mathbf{0}$. This completes the proof of the lemma. ■

We introduce the assumption.

***A**-8:* For each $\boldsymbol{\eta}$, let the rate of change of

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

[9] cf. [11, Chapter 6, Section 6.3.2]

[10] note that ρ must be kept strictly away from zero in order to allow $\hat{w}_k^i$ to track the time varying true average $\bar{w}_k$, see [11] for further detail.

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

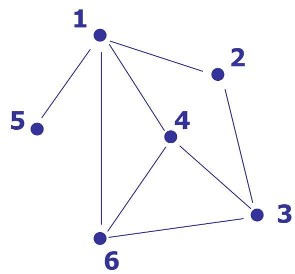

Fig. 1. Network topology $\mathcal{G}$

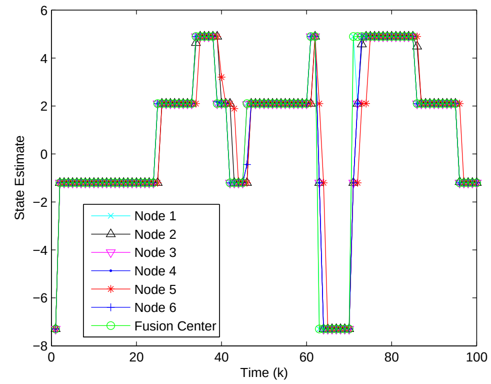

Fig. 2. Distributed and centralized state estimates

go to zero $\mathbb{P}$–w.p.1 as $t \to \infty$. This means the asymptotic rate of change condition[12]

$$\lim_{k} \sup_{j \geq k} \max_{0 \leq t \leq T} \mid \mathbb{Q}^{\circ}_{\boldsymbol{\eta}}(jT+t) - \mathbb{Q}^{\circ}_{\boldsymbol{\eta}}(jT) \mid = 0 \quad \mathbb{P}\text{–w.p.1} \tag{17}$$

is satisfied uniformly in $\boldsymbol{\eta}$ for every $T > 0$.

In the main theorem of this section, by assuming that some compact set $\Omega_{\bar{c}}$ is recurrent and the mean ODE (6) is stable, it is stated that the error process $\{\boldsymbol{\eta}_k\}$ is bounded $\mathbb{P}$–w.p.1 and converges to a bounded invariant set in $\Omega_{\bar{c}}$.

Theorem 3.7: Consider the unconstrained stochastic approximation algorithm (3). For any $\delta \in (0, 1]$, let there be a compact set $\mathbb{L}_\delta$ such that the iterates $\boldsymbol{\eta}_k$ return to $\mathbb{L}_\delta$ infinitely often with probability at least δ. Assume conditions **A**-*4* and **A**-*5*. Then, $\{\boldsymbol{\eta}_k\}$ is bounded $\mathbb{P}$–w.p.1 , that is,

$$\limsup_{k} \|\boldsymbol{\eta}_k\| < \infty \quad \mathbb{P}\text{–w.p.1}$$

Assume condition **A**-*8*. Also, assume that the function $\mathbb{Q}(., \boldsymbol{\xi})$ satisfies the nonuniform regularity condition in $\boldsymbol{\eta}$ established in Lemma 3.6. Then, there exists a null set $\mho$ such that for $\omega \notin \mho$, the set of functions $\{\boldsymbol{\eta}^k_\bullet(\omega, .), k < \infty\}$ is equicontinuous. Let $\boldsymbol{\eta}(\omega, .)$ denote the limit of some convergent subsequence $\{\boldsymbol{\eta}^{k'}_\bullet(\omega, .)\}$. Then, for $\mathbb{P}$–almost all $\omega \in \Omega$, the limits $\boldsymbol{\eta}(\omega, .)$ are trajectories of the mean ODE (6) in some bounded invariant set and the error iterates $\{\boldsymbol{\eta}_k\}$ converge to this invariant set. Moreover, let the origin be globally[13] asymptotically ϵ-stable[14] for the mean ODE (6) with some invariant level set Ω_{c°, where $\Omega_{c^\circ} \subset \mathbb{L}_1$. Then, $\{\boldsymbol{\eta}_k\}$ converges to the ϵ-neighborhood of the origin $\mathbb{P}$–w.p.1 as $k \to \infty$.

Proof: The proof follows from [11, Theorem 7.1 and Theorem 1.1, Chapter 6] and for brevity the details are omitted here. ∎

IV. 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 $Y^j_k = X_k + u^j_k$, where the measurement noises $\{\mathbf{u}_k = [u^j_k]_{j \in \mathcal{N}}, k \in \mathbb{Z}^+\}$ 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 $\hat{\mathbf{w}}_{-1} = c\mathbb{1}$, with $c = 3$.

Fig. 2 shows the distributed (or local) estimate $\hat{X}^j_k \triangleq \mathbb{E}^j[X_k \mid \mathcal{F}_k]$ 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 $\hat{\mathbf{v}}^j_k = [\hat{v}^j_k(\ell)]_{\ell \in \mathcal{S}}$ 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 $\hat{X}^j_k$ 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 $\bar{g}^j_k \triangleq (k+1)^{-1} \sum_{i=0}^{k} \mid \hat{X}^j_i - \hat{X}_i \mid$. This is done based on the fact that $\bar{g}^j_k$ converges $\mathbb{P}$–a.s. to the expectation[15] $\mathbb{E} \mid \hat{X}^j_k - \hat{X}_k \mid$. This is due to the geometric ergodicity of the extended Markov chain $\{(X_k, \mathbf{Y}_k),\ k \in \mathbb{Z}^+\}$. As it can

[12] see Section 5.3 and 6.1, [11] for further detail.

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

[14] this is shown in Proposition 3.3.

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

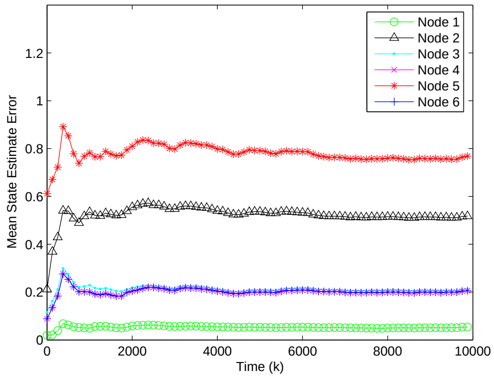

Fig. 3. Convergence in mean of the distributed state estimate at each node to the centralized one.

be seen for each node j, the average $\bar{g}_k^j$ converges to a $\bar{\delta}^j$-ball around the origin as $k \to \infty$. The radius $\|\bar{\delta}\|$, where $\bar{\delta} = [\bar{\delta}^j]_{j \in \mathcal{N}}$ 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.

V. Acknowledgments

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

References

[1] V. Borkar and P. P. Varaiya, "Asymptotic agreement in distributed estimation," *IEEE Transactions on Automatic Control*, vol. AC-27, no. 3, pp. 650–655, June 1982.
[2] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803–812, September 1986.
[3] R. Olfati-Saber, "Distributed Kalman Filter with Embedded Consensus Filters," in *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, Seville, Spain, December 12-15 2005, pp. 8179–8184.
[4] R. Olfati-Saber and J. S. Shamma, "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*, Seville, Spain, December 12-15 2005, pp. 6698–6703.
[5] P. Arliksson and A. Rantzer, "Distributed Kalman filtering using weighted averaging," in *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, Kyoto, Japan, July 2006.
[6] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 622–633, May 2008.
[7] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, "Approximate distributed Kalman filtering in sensors networks with qunatifiable performance," in *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks*, Los Angeles, CA, USA, April 2005, pp. 133–139.
[8] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proceedings of the fourth International Symposium on Information Processing in Sensor Networks*, April 2005, pp. 63–70.
[9] D. W. Casbeer and R. Beard, "Distributed information filtering using consensus filters," in *Proceedings of the 2009 American Control Conference (ACC 2009)*, Hyatt Regency Riverfront, St. Louis, MO, USA, June 2009, pp. 1882–1887.
[10] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, A. V. Balakrishnan, I. Karatzas, and M. Yor, Eds. vol. 22, in Applications of Mathematics, Berlin Heidelberg: Springer-Verlag, 1990.
[11] H. J. Kushner and G. G. Yin, *Stochastic approximation and recursive algorithms and applications*, 2nd ed., B. Rozovskii and M. Yor, Eds. vol. 35, in Applications of Mathematics, New York, NY, USA: Springer-Verlag New York, Inc., 2003.
[12] F. LeGland and L. Mevel, "Exponential Forgetting and Geometric Ergodicity in Hidden Markov Models," *Mathematics of Control, Signals, and Systems*, vol. 13, pp. 63–93, 2000.
[13] A. N. Shiryaev, *Probability*, 2nd ed., S. Axler, F. Gehring, and P. Halmos, Eds. vol. 95, in Graduate Texts in Mathematics, New York, NY, USA: Springer-Verlag New York, Inc., 1996.
[14] P. Billingsley, *Probability and Measure*, 3rd ed. New York, NY, USA: John Wiley & Sons, Inc., 1995.

Appendix A
Proof of Proposition 3.3

Proof: Consider the positive definite and radially unbounded Lyapunov function $V(.) : \mathbb{R}^n \mapsto \mathbb{R}$, where $V(.)$ is defined by $V(\boldsymbol{\eta}_\bullet) = \frac{1}{2}\boldsymbol{\eta}_\bullet'\boldsymbol{\eta}_\bullet$. The time derivative of $V(.)$ is computed as

$$\begin{aligned}
&\dot{V}(\boldsymbol{\eta}_\bullet) = \boldsymbol{\eta}_\bullet'\boldsymbol{\Lambda}\boldsymbol{\eta}_\bullet + \boldsymbol{\eta}_\bullet'\boldsymbol{\Gamma}(\bar{\boldsymbol{z}} - n^{-1}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}}) \\
&(\star) \leq \lambda_{max}(\boldsymbol{\Lambda})\|\boldsymbol{\eta}_\bullet\|^2 + \|\bar{\nu}\boldsymbol{\Lambda}\mathbb{1}\|\|\boldsymbol{\eta}_\bullet\| + \|\bar{\nu}n^{-1}\boldsymbol{\Lambda}\mathbb{1}\mathbb{1}'\mathbb{1}\|\|\boldsymbol{\eta}_\bullet\| \\
&\quad \leq \lambda_{max}(\boldsymbol{\Lambda})\|\boldsymbol{\eta}_\bullet\|^2 + 2\bar{\nu}\sqrt{n}(1+d_{max})\|\boldsymbol{\eta}_\bullet\| \\
&\quad = \left(\bar{\nu}\sqrt{n}(1+d_{max})|\ \lambda_{max}(\boldsymbol{\Lambda})\ |^{-\frac{1}{2}}\right)^2 - \\
&\quad \left(|\ \lambda_{max}(\boldsymbol{\Lambda})\ |^{\frac{1}{2}}\|\boldsymbol{\eta}_\bullet\| - \bar{\nu}\sqrt{n}(1+d_{max})|\ \lambda_{max}(\boldsymbol{\Lambda})\ |^{-\frac{1}{2}}\right)^2
\end{aligned} \tag{18}$$

where in $(\star)$, the Cauchy-Schwarz inequality and the equality $\boldsymbol{\Lambda}\mathbb{1} = -\boldsymbol{\Gamma}\mathbb{1}$ are used. Note that the matrix $\boldsymbol{\Lambda}$ is negative definite and $\boldsymbol{\Lambda}\mathbb{1} = [-(1+d_i)]_{i \in \mathcal{N}}$ with $d_i \stackrel{\triangle}{=} \mathbf{A}_i\mathbb{1}$ being the degree of the node i.

Consider a closed ϵ-ball with radius $\epsilon = 2\bar{\nu}\sqrt{n}(1 + d_{max})|\ \lambda_{max}(\boldsymbol{\Lambda})\ |^{-1}$ centered at the origin. The compact level set Ω_{c° with $c^\circ = \frac{1}{2}\epsilon^2$ contains this ϵ-ball. From (18), any solution $\boldsymbol{\eta}_\bullet(.)$ of the ODE (6), for which the initial condition $\boldsymbol{\eta}_\bullet(0)$ lies in $\mathbb{R}^n \setminus \Omega_{c^\circ}$, satisfies $\dot{V}(\boldsymbol{\eta}_\bullet) \leq -v,\ v > 0$. As such Ω_{c° is an invariant level set in that the trajectory $\boldsymbol{\eta}_\bullet(.)$ reaches the level set Ω_{c° in some finite time and stays in Ω_{c° afterward. Therefore, for the mean ODE (6), the origin is globally asymptotically ϵ-stable and the proof is concluded. ■

Appendix B
Proof of Theorem 3.4

Proof: For the proof of this theorem, we first present the following lemma which establishes a strong law of large numbers to compute the average of the likelihood of the beliefs over conditionally independent but not identically distributed nodes.

Lemma B.1: Assume **A-***1* and **A-***6*. If $\Delta_{(2)}$ is finite, then for sufficiently large n, the following

$$n^{-1}\mathbb{1}\mathbb{1}'(\mathbf{z}_k - \bar{\boldsymbol{z}}) \to \mathbf{0} \quad \mathbb{P}\text{–a.s.} \tag{19}$$

is satisfied uniformly in $k \in \mathbb{Z}^+$.

Remark 3: For the proof see [13, Theorem 2, §3, Chapter IV].

Proof of Theorem 3.4: From the definition (13) we have

$$\mathbb{E}_k\big[V_{k+1}(\boldsymbol{\eta}_{k+1}) - V_k(\boldsymbol{\eta}_k)\big] = \mathbb{E}_k\big[V(\boldsymbol{\eta}_{k+1}) - V(\boldsymbol{\eta}_k)\big] + \mathbb{E}_k\big[\boldsymbol{\eta}_{k+1}'\delta\vartheta_{k+1}(\boldsymbol{\eta}_{k+1}) - \boldsymbol{\eta}_k'\delta\vartheta_k(\boldsymbol{\eta}_k)\big] \quad (20)$$

Taylor series expansion of the Lyapunov function $V(\boldsymbol{\eta}) = \frac{1}{2}\|\boldsymbol{\eta}\|^2$ in a neighborhood of $\boldsymbol{\eta}_k$ yields

$$\begin{aligned}\mathbb{E}_k V(\boldsymbol{\eta}_{k+1}) - V(\boldsymbol{\eta}_k) = &\\ &\rho\boldsymbol{\eta}_k'\boldsymbol{\Lambda}\boldsymbol{\eta}_k + \rho\boldsymbol{\eta}_k'\mathbb{E}_k\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k) + \frac{1}{2}\rho^2\|\boldsymbol{\Lambda}\boldsymbol{\eta}_k\|^2 + \\ &\rho^2\boldsymbol{\eta}_k'\boldsymbol{\Lambda}\mathbb{E}_k\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k) + \frac{1}{2}\rho^2\mathbb{E}_k\|\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k)\|^2 \quad (21)\end{aligned}$$

where we define

$$\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k) \triangleq \boldsymbol{\Gamma}(\mathbf{z}_{k+1} - n^{-1}\mathbb{1}\mathbb{1}'\mathbf{z}_k) - (n\rho)^{-1}\mathbb{1}\mathbb{1}'(\mathbf{z}_{k+1} - \mathbf{z}_k)$$

Define also

$$\bar{\mathbb{G}} \triangleq \boldsymbol{\Gamma}(\bar{\boldsymbol{z}} - n^{-1}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}})$$

From (11), we write

$$\begin{aligned}\mathbb{E}_k\boldsymbol{\eta}_k'\delta\vartheta_k(\boldsymbol{\eta}_k) &= \boldsymbol{\eta}_k'\mathbb{E}_k\sum_{i=k}^{\infty}\rho\beta_{k+1}^i\mathbb{E}_k[\mathbb{Q}(\boldsymbol{\eta}_k,\boldsymbol{\xi}_{i+1}) - \bar{\mathbb{Q}}(\boldsymbol{\eta}_k)]\\ &= \rho\boldsymbol{\eta}_k'\mathbb{E}_k[\mathbb{Q}(\boldsymbol{\eta}_k,\boldsymbol{\xi}_{k+1}) - \bar{\mathbb{Q}}(\boldsymbol{\eta}_k)] +\\ &\quad (1-\rho)\boldsymbol{\eta}_k'\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_k[\mathbb{Q}(\boldsymbol{\eta}_k,\boldsymbol{\xi}_{i+1}) - \bar{\mathbb{Q}}(\boldsymbol{\eta}_k)]\\ &= \rho\boldsymbol{\eta}_k'\mathbb{E}_k\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k) - \rho\boldsymbol{\eta}_k'\bar{\mathbb{G}} +\\ &\quad (1-\rho)\boldsymbol{\eta}_k'\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_k\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i) - (1-\rho)\boldsymbol{\eta}_k'\bar{\mathbb{G}}\\ &= \rho\boldsymbol{\eta}_k'\mathbb{E}_k\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k) - \boldsymbol{\eta}_k'\bar{\mathbb{G}} +\\ &\quad (1-\rho)\boldsymbol{\eta}_k'\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_k\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i) \quad (22)\end{aligned}$$

Also, from (11) and (12), we write

$$\begin{aligned}&\mathbb{E}_k\boldsymbol{\eta}_{k+1}'\delta\vartheta_{k+1}(\boldsymbol{\eta}_{k+1})\\ &\quad = \mathbb{E}_k\boldsymbol{\eta}_{k+1}'\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_{k+1}[\mathbb{Q}(\boldsymbol{\eta}_{k+1},\boldsymbol{\xi}_{i+1}) - \bar{\mathbb{Q}}(\boldsymbol{\eta}_{k+1})]\\ &\quad = -\mathbb{E}_k\boldsymbol{\eta}_{k+1}'\bar{\mathbb{G}} + \mathbb{E}_k\boldsymbol{\eta}_{k+1}'\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_{k+1}\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i)\\ &\quad = -\mathbb{E}_k\big[\boldsymbol{\eta}_k' + \rho\boldsymbol{\eta}_k'\boldsymbol{\Lambda} + \rho\mathbb{G}'(\mathbf{z}_{k+1},\mathbf{z}_k)\big]\bar{\mathbb{G}}\\ &\quad + \mathbb{E}_k\Big[\big[\boldsymbol{\eta}_k' + \rho\boldsymbol{\eta}_k'\boldsymbol{\Lambda} + \rho\mathbb{G}'(\mathbf{z}_{k+1},\mathbf{z}_k)\big]\\ &\quad \sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_{k+1}\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i)\Big]\\ &\quad = -\boldsymbol{\eta}_k'\bar{\mathbb{G}} - \rho\boldsymbol{\eta}_k'\boldsymbol{\Lambda}\bar{\mathbb{G}} - \rho\mathbb{E}_k\mathbb{G}'(\mathbf{z}_{k+1},\mathbf{z}_k)\bar{\mathbb{G}} +\\ &\quad \boldsymbol{\eta}_k'\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_k\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i) +\\ &\quad \rho\boldsymbol{\eta}_k'\boldsymbol{\Lambda}\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_k\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i) +\\ &\quad \rho\mathbb{E}_k\big[\mathbb{G}'(\mathbf{z}_{k+1},\mathbf{z}_k)\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_{k+1}\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i)\big] \quad (23)\end{aligned}$$

For $k \in \mathbb{Z}^+$, define the *distributed* prediction filter at node $j \in \mathcal{N}$ by $\bar{\boldsymbol{p}}_k^j = [\bar{p}_k^j(\ell)]_{\ell\in\mathcal{S}}$, where

$$\bar{p}_k^j(\ell) \triangleq \mathbb{P}(X_k = \ell \mid \mathcal{F}_{k-1}^j)$$

for each $\ell \in \mathcal{S}$. The conditional expected value $\mathbb{E}_k\mathbf{z}_i$ for each $i \geq k+1$ may be written as

$$\mathbb{E}_k\mathbf{z}_i = [\bar{\boldsymbol{f}}^{j\,\prime}\mathbb{X}_{k+1}^{i\;\prime}\bar{\boldsymbol{p}}_{k+1}^j]_{j\in\mathcal{N}} \quad (24)$$

where we define for $\ell \in \mathcal{S}$

$$\bar{\boldsymbol{f}}^j = \big[\bar{f}_x^j\big]_{x\in\mathcal{S}} \triangleq \big[\int |\log f_\ell^j(y^j)|\, f_x^j(y^j)\varrho(dy^j)\big]_{x\in\mathcal{S}}$$

and $\mathbb{X}_k^{i\,\prime} \triangleq \prod_{\kappa=1}^{i-k}\mathbf{X}'$, where for $i = k$ the empty product is defined by $\mathbb{X}_k^{k\,\prime} \triangleq \mathbf{I}$. Define also

$$\bar{f}_{max}^{(\iota)} \triangleq \max_{j\in\mathcal{N}}\max_{x\in\mathcal{S}}\int |\log f_\ell^j(y^j)|^\iota f_x^j(y^j)\varrho(dy^j) \quad (25)$$

Since $\Delta_{(1)}$ and $\Delta_{(2)}$ are finite, $\bar{f}_{max}^{(\iota)}$ for both $\iota = 1, 2$ are finite. Substituting (21), (22), and (23) in (20) gives

$$\begin{aligned}&\mathbb{E}_k\big[V_{k+1}(\boldsymbol{\eta}_{k+1}) - V_k(\boldsymbol{\eta}_k)\big] =\\ &\quad \rho\boldsymbol{\eta}_k'(\mathbf{I} + \frac{1}{2}\rho\boldsymbol{\Lambda})\boldsymbol{\Lambda}\boldsymbol{\eta}_k + \rho^2\boldsymbol{\eta}_k'\boldsymbol{\Lambda}\mathbb{E}_k\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k) +\\ &\quad \frac{1}{2}\rho^2\mathbb{E}_k\|\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k)\|_2^2 + \rho\boldsymbol{\eta}_k'\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_k\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i)\\ &\quad + \rho\boldsymbol{\eta}_k'\boldsymbol{\Lambda}\big(\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_k\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i) - \bar{\mathbb{G}}\big)\\ &\quad + \rho\mathbb{E}_k\big[\mathbb{G}'(\mathbf{z}_{k+1},\mathbf{z}_k)\,\big(\sum_{i=k+1}^{\infty}\rho\beta_{k+2}^i\mathbb{E}_{k+1}\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i) - \bar{\mathbb{G}}\big)\big]\end{aligned} \quad (26)$$

Using Lemma B.1 and (24), we write

$$\begin{aligned}
&\rho^2 \boldsymbol{\eta}_k' \mathbf{\Lambda} \mathbb{E}_k \mathbb{G}(\mathbf{z}_{k+1}, \mathbf{z}_k) \\
&\quad = \rho^2 \boldsymbol{\eta}_k' \mathbf{\Lambda} \mathbf{\Gamma} (\mathbb{E}_k \mathbf{z}_{k+1} - n^{-1} \mathbb{1}\mathbb{1}' \bar{\boldsymbol{z}}) \\
&\quad \leq \big(\ \|\rho^2 \mathbf{\Lambda}\mathbf{\Gamma}[\bar{\boldsymbol{f}}^{j\prime} \bar{\boldsymbol{p}}_{k+1}^j]_{j\in\mathcal{N}}\| + \|n^{-1}\rho^2 \mathbf{\Lambda}\mathbf{\Gamma}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}}\| \ \big) \ \|\boldsymbol{\eta}_k\| \\
&\quad \leq \rho^2 \lambda_{max}(\mathbf{\Lambda}) \big(\ \|\bar{f}_{max}^{(1)} \mathbf{\Lambda}\mathbb{1}\| + \|n^{-1} \bar{f}_{max}^{(1)} \mathbf{\Lambda}\mathbb{1}\mathbb{1}'\mathbb{1}\| \ \big) \ \|\boldsymbol{\eta}_k\| \\
&\quad \leq 2\sqrt{n}\rho^2 \bar{f}_{max}^{(1)} \lambda_{max}(\mathbf{\Lambda})(1+d_{max}) \ \|\boldsymbol{\eta}_k\| \qquad (27)
\end{aligned}$$

Also, we have

$$\begin{aligned}
&\|\mathbb{G}(\mathbf{z}_{k+1}, \mathbf{z}_k)\|_2 \\
&\quad \leq \|\mathbf{\Gamma}\mathbf{z}_{k+1}\|_2 + \|n^{-1}\mathbf{\Gamma}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}}\|_2 \\
&\quad \leq \|\mathbf{\Gamma}\mathbf{z}_{k+1}\|_1 + \|n^{-1}\bar{f}_{max}^{(1)}\mathbf{\Gamma}\mathbb{1}\mathbb{1}'\mathbb{1}\|_2 \\
&\quad = (\mathbf{\Gamma}\mathbb{1})'\mathbf{z}_{k+1} + \bar{f}_{max}^{(1)}\|\mathbf{\Lambda}\mathbb{1}\|_2 \\
&\quad \leq \|\mathbf{\Lambda}\mathbb{1}\|_2 \|\mathbf{z}_{k+1}\|_2 + \sqrt{n}\bar{f}_{max}^{(1)}(1+d_{max}) \\
&\quad \leq \sqrt{n}(1+d_{max})(\|\mathbf{z}_{k+1}\|_2 + \bar{f}_{max}^{(1)})
\end{aligned}$$

and then under **A-***1* we may write

$$\begin{aligned}
&\frac{1}{2}\rho^2 \mathbb{E}_k \|\mathbb{G}(\mathbf{z}_{k+1}, \mathbf{z}_k)\|_2^2 \\
&\quad \leq \frac{1}{2} n\rho^2 (1+d_{max})^2 \big(\mathbb{E}_k \|\mathbf{z}_{k+1}\|_2^2 + 2\bar{f}_{max}^{(1)} \mathbb{E}_k \|\mathbf{z}_{k+1}\|_2 \\
&\quad + (\bar{f}_{max}^{(1)})^2 \big) \\
&\quad = \frac{1}{2} n\rho^2 (1+d_{max})^2 \big(\ \big\| \big[\mathbb{E}_{\mathcal{F}_k^j} z_{k+1}^2(j) \big]_{j\in\mathcal{N}} \big\|_1 \\
&\quad + 2\bar{f}_{max}^{(1)} \mathbb{E}_k \|\mathbf{z}_{k+1}\|_2 + (\bar{f}_{max}^{(1)})^2 \ \big) \\
&\quad \leq \frac{1}{2} n\rho^2 (1+d_{max})^2 \big(\big\| \bar{f}_{max}^{(2)} \mathbb{1} \big\|_1 + 2\bar{f}_{max}^{(1)} \mathbb{E}_k \|\mathbf{z}_{k+1}\|_1 \\
&\quad + (\bar{f}_{max}^{(1)})^2 \big) \\
&\quad = \frac{1}{2} n\rho^2 (1+d_{max})^2 \big(n\bar{f}_{max}^{(2)} + 2\bar{f}_{max}^{(1)} \|\mathbb{E}_k \mathbf{z}_{k+1}\|_1 \\
&\quad + (\bar{f}_{max}^{(1)})^2 \big) \\
&\quad \leq \frac{1}{2} n\rho^2 (1+d_{max})^2 \big(n\bar{f}_{max}^{(2)} + 2\bar{f}_{max}^{(1)} \|\bar{f}_{max}^{(1)} \mathbb{1}\|_1 \\
&\quad + (\bar{f}_{max}^{(1)})^2 \big) \\
&\quad = \frac{1}{2} n\rho^2 (1+d_{max})^2 \big(n\bar{f}_{max}^{(2)} + (2n+1)(\bar{f}_{max}^{(1)})^2 \big) \triangleq \rho\varphi_2^2 \\
&\qquad (28)
\end{aligned}$$

As $\sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i = 1$, using Lemma B.1 and (24) we write

$$\begin{aligned}
&\rho\boldsymbol{\eta}_k' \sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_k \mathbb{G}(\mathbf{z}_{i+1}, \mathbf{z}_i) \\
&\quad = \rho\boldsymbol{\eta}_k' \mathbf{\Gamma} \big(\sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_k \mathbf{z}_{i+1} - n^{-1}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}} \big) \\
&\quad \leq \rho \big(\big\| \mathbf{\Gamma}[\bar{\boldsymbol{f}}^{j\prime} \widetilde{\mathbf{X}}_{k,i}' \ \bar{\boldsymbol{p}}_{k+1}^j]_{j\in\mathcal{N}} \big\| + \|n^{-1}\mathbf{\Gamma}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}}\| \big) \ \|\boldsymbol{\eta}_k\|
\end{aligned}$$

where we define $\widetilde{\mathbf{X}}_{k,i}' \triangleq \sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{X}_k^{i\,\prime}$. It is clear that $\widetilde{\mathbf{X}}_{k,i}$ is a stochastic matrix and, thus, we have

$$\begin{aligned}
&\rho\boldsymbol{\eta}_k' \sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_k \mathbb{G}(\mathbf{z}_{i+1}, \mathbf{z}_i) \\
&\qquad \leq \rho \big(\|\bar{f}_{max}^{(1)} \mathbf{\Lambda}\mathbb{1}\| + \|n^{-1}\bar{f}_{max}^{(1)} \mathbf{\Lambda}\mathbb{1}\mathbb{1}'\mathbb{1}\| \big) \ \|\boldsymbol{\eta}_k\| \\
&\qquad \leq 2\sqrt{n}\rho \bar{f}_{max}^{(1)} (1+d_{max}) \|\boldsymbol{\eta}_k\| \qquad (29)
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\rho\boldsymbol{\eta}_k' \mathbf{\Lambda} \big(\sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_k \mathbb{G}(\mathbf{z}_{i+1}, \mathbf{z}_i) - \bar{\mathbb{G}} \big) \\
&\quad = \rho\boldsymbol{\eta}_k' \mathbf{\Lambda}\mathbf{\Gamma} \\
&\quad \big(\sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_k \mathbf{z}_{i+1} - n^{-1}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}} - \bar{\boldsymbol{z}} + n^{-1}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}} \big) \\
&\quad \leq \rho \big(\big\| \mathbf{\Lambda}\mathbf{\Gamma}[\bar{\boldsymbol{f}}^{j\prime} \widetilde{\mathbf{X}}_{k,i}' \ \bar{\boldsymbol{p}}_{k+1}^j]_{j\in\mathcal{N}} \big\| + \|\mathbf{\Lambda}\mathbf{\Gamma}\bar{\boldsymbol{z}}\| \big) \ \|\boldsymbol{\eta}_k\| \\
&\quad \leq \rho \big(\|\bar{f}_{max}^{(1)} \mathbf{\Lambda}\mathbf{\Gamma}\mathbb{1}\| + \|\bar{f}_{max}^{(1)} \mathbf{\Lambda}\mathbf{\Gamma}\mathbb{1}\| \big) \ \|\boldsymbol{\eta}_k\| \\
&\quad \leq 2\rho \bar{f}_{max}^{(1)} \lambda_{max}(\mathbf{\Lambda}) \|\mathbf{\Lambda}\mathbb{1}\| \ \|\boldsymbol{\eta}_k\| \\
&\quad \leq 2\sqrt{n}\rho \bar{f}_{max}^{(1)} \lambda_{max}(\mathbf{\Lambda})(1+d_{max}) \|\boldsymbol{\eta}_k\| \qquad (30)
\end{aligned}$$

Using Lemma B.1, we write

$$\begin{aligned}
&\sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_{k+1} \mathbb{G}(\mathbf{z}_{i+1}, \mathbf{z}_i) - \bar{\mathbb{G}} \\
&\qquad = \mathbf{\Gamma} \big(\sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_{k+1} \mathbf{z}_{i+1} - \bar{\boldsymbol{z}} \big)
\end{aligned}$$

and thus we have

$$\begin{aligned}
&\mathbb{G}'(\mathbf{z}_{k+1}, \mathbf{z}_k) \ \big(\sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_{k+1} \mathbb{G}(\mathbf{z}_{i+1}, \mathbf{z}_i) - \bar{\mathbb{G}} \big) \\
&\quad \leq \|\mathbf{\Gamma}(\mathbf{z}_{k+1} - n^{-1}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}})\|_2 \\
&\quad \|\mathbf{\Gamma}\big(\sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_{k+1} \mathbf{z}_{i+1} - \bar{\boldsymbol{z}} \big)\|_2 \\
&\quad \leq \big(\|\mathbf{\Gamma}\mathbf{z}_{k+1}\|_2 + \|n^{-1}\mathbf{\Gamma}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}}\|_2 \big) \ . \\
&\quad \big(\|\mathbf{\Gamma}[\bar{\boldsymbol{f}}^{j\prime} \widetilde{\mathbf{X}}_{k+1,i}' \ \bar{\boldsymbol{p}}_{k+1}^j]_{j\in\mathcal{N}}\|_2 + \|\mathbf{\Gamma}\bar{\boldsymbol{z}}\|_2 \big)
\end{aligned}$$

As before, the matrix $\widetilde{\mathbf{X}}_{k+1,i}' = \sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{X}_{k+1}^{i\,\prime}$ is left stochastic and we write

$$\begin{aligned}
&\mathbb{G}'(\mathbf{z}_{k+1}, \mathbf{z}_k) \ \big(\sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_{k+1} \mathbb{G}(\mathbf{z}_{i+1}, \mathbf{z}_i) - \bar{\mathbb{G}} \big) \\
&\quad \leq \big(\|\mathbf{\Gamma}\mathbf{z}_{k+1}\|_1 + \|n^{-1}\bar{f}_{max}^{(1)}\mathbf{\Gamma}\mathbb{1}\mathbb{1}'\mathbb{1}\|_2 \big) \ . \\
&\quad \big(\|\bar{f}_{max}^{(1)}\mathbf{\Gamma}\mathbb{1}\|_2 + \|\bar{f}_{max}^{(1)}\mathbf{\Gamma}\mathbb{1}\|_2 \big) \\
&\quad = 2\bar{f}_{max}^{(1)} \|\mathbf{\Lambda}\mathbb{1}\|_2 \big((\mathbf{\Gamma}\mathbb{1})'\mathbf{z}_{k+1} + \bar{f}_{max}^{(1)} \|\mathbf{\Lambda}\mathbb{1}\|_2 \big) \\
&\quad \leq 2\bar{f}_{max}^{(1)} \|\mathbf{\Lambda}\mathbb{1}\|_2 \big(\|\mathbf{\Lambda}\mathbb{1}\|_2 \|\mathbf{z}_{k+1}\|_2 + \bar{f}_{max}^{(1)} \|\mathbf{\Lambda}\mathbb{1}\|_2 \big) \\
&\quad \leq 2\bar{f}_{max}^{(1)} \|\mathbf{\Lambda}\mathbb{1}\|_2^2 \big(\|\mathbf{z}_{k+1}\|_1 + \bar{f}_{max}^{(1)} \big) \qquad (31)
\end{aligned}$$

thus we have

$$\begin{aligned}
&\rho\mathbb{E}_k \big[\mathbb{G}'(\mathbf{z}_{k+1}, \mathbf{z}_k) \ \big(\sum_{i=k+1}^{\infty} \rho\beta_{k+2}^i \mathbb{E}_{k+1} \mathbb{G}(\mathbf{z}_{i+1}, \mathbf{z}_i) - \bar{\mathbb{G}} \big) \big] \\
&\quad \leq 2n\rho \bar{f}_{max}^{(1)} (1+d_{max})^2 \big(\mathbb{E}_k \|\mathbf{z}_{k+1}\|_1 + \bar{f}_{max}^{(1)} \big) \\
&\quad \leq 2n\rho \bar{f}_{max}^{(1)} (1+d_{max})^2 \big(\|\mathbb{E}_k \mathbf{z}_{k+1}\|_1 + \bar{f}_{max}^{(1)} \big) \\
&\quad = 2n\rho \bar{f}_{max}^{(1)} (1+d_{max})^2 \big(\|[\bar{\boldsymbol{f}}^{j\prime} \ \bar{\boldsymbol{p}}_{k+1}^j]_{j\in\mathcal{N}}\|_1 + \bar{f}_{max}^{(1)} \big) \\
&\quad \leq 2n\rho \bar{f}_{max}^{(1)} (1+d_{max})^2 \big(\|\bar{f}_{max}^{(1)} \mathbb{1}\|_1 + \bar{f}_{max}^{(1)} \big) \\
&\quad = 2n(n+1)\rho(1+d_{max})^2 (\bar{f}_{max}^{(1)})^2 \\
&\qquad (32)
\end{aligned}$$

By applying Geršgorin theorem to the matrix $\mathbf{I}+\frac{1}{2}\rho\mathbf{\Lambda}$, it is shown that under Assumption **A-***7* all the eigenvalues are strictly positive and as such the matrix $\mathbf{Q}\stackrel{\triangle}{=}(\mathbf{I}+\frac{1}{2}\rho\mathbf{\Lambda})\mathbf{\Lambda}$ is negative definite. Substituting (27)- (32) in (26) yields

$$\begin{aligned}&\mathbb{E}_k\big[V_{k+1}(\boldsymbol{\eta}_{k+1})-V_k(\boldsymbol{\eta}_k)\big]\\&\quad\le\rho\lambda_{max}(\mathbf{Q})\|\boldsymbol{\eta}_k\|^2+\tilde{\varphi}\|\boldsymbol{\eta}_k\|+\varphi^2\\&\quad=\varphi^2+\frac{1}{4}\tilde{\varphi}^2\mid\rho\lambda_{max}(\mathbf{Q})\mid^{-1}\\&\quad-\big(\mid\rho\lambda_{max}(\mathbf{Q})\mid^{\frac{1}{2}}\|\boldsymbol{\eta}_k\|-\frac{1}{2}\tilde{\varphi}\mid\rho\lambda_{max}(\mathbf{Q})\mid^{-\frac{1}{2}}\big)^2\end{aligned}$$

where $\tilde{\varphi}\stackrel{\triangle}{=}\mid 2\sqrt{n}\rho\bar{f}^{(1)}_{max}(1+d_{max})(1+(1+\rho)\lambda_{max}(\mathbf{\Lambda}))\mid$ and $\varphi^2\stackrel{\triangle}{=}n\rho(1+d_{max})^2\big(\frac{1}{2}n\rho\bar{f}^{(2)}_{max}+(\frac{1}{2}\rho(2n+1)+2(n+1))(\bar{f}^{(1)}_{max})^2\big)$.

Now if the iterate $\boldsymbol{\eta}_k$ lies outside the interior $\Omega^{\circ}_{\bar{c}}$ of a compact level set $\Omega_{\bar{c}}$, where $\Omega_{\bar{c}}$ is defined as $\Omega_{\bar{c}}\stackrel{\triangle}{=}\{\boldsymbol{\eta}:V(\boldsymbol{\eta})\le\bar{c}\}$, with $\bar{c}=\frac{1}{2}\acute{c}^2$, where $\acute{c}>0$ is given by $\acute{c}=\mid\rho\lambda_{max}(\mathbf{Q})\mid^{-\frac{1}{2}}\mid\varphi\mid+\tilde{\varphi}\mid\rho\lambda_{max}(\mathbf{Q})\mid^{-1}$, then there exists an $\bar{\alpha}>0$ such that

$$\begin{aligned}&\|\boldsymbol{\eta}_k\|\ge\acute{c}:\ \boldsymbol{\eta}_k\in\mathbb{R}^n\setminus\Omega^{\circ}_c\ \Rightarrow\\&\mathbb{E}_k\big[V_{k+1}(\boldsymbol{\eta}_{k+1})-V_k(\boldsymbol{\eta}_k)\big]\\&\qquad\qquad\le-\tilde{\varphi}\mid\rho\lambda_{max}(\mathbf{Q})\mid^{-\frac{1}{2}}\mid\varphi\mid\stackrel{\triangle}{=}-\bar{\alpha}<0\end{aligned}$$

Thus, $\mathbb{E}_kV_{k+1}(\boldsymbol{\eta}_{k+1})<V_k(\boldsymbol{\eta}_k)$ $\mathbb{P}$–w.p.1 for $V(\boldsymbol{\eta}_k)\ge\bar{c}$. Define a random variable τ with values in $[0,\infty]$ as an $\mathcal{F}_k$–stopping time with respect to the error process $\{\boldsymbol{\eta}_k\}$ when $\boldsymbol{\eta}_k$ first enters $\Omega_{\bar{c}}$, that is, τ is finite $\mathbb{P}$–a.s. and the event $\{\tau<k\}$ is measurable with respect to $\mathcal{F}_k$ for each finite $k\in\mathbb{Z}^+$. Define $\tau\wedge k\stackrel{\triangle}{=}\min\{\tau,k\}$. Hence, $V_{\tau\wedge k}(\boldsymbol{\eta}_{\tau\wedge k})$ is an $\mathcal{F}_k$–supermartingale for the stopped process $\boldsymbol{\eta}_k$ with the $\mathcal{F}_k$–stopping time τ. This completes the proof of the theorem. ∎

APPENDIX C
PROOF OF THEOREM 3.5

Proof: In order to show that some compact set $\mathbb{L}_\delta$ is recurrent for the error process $\boldsymbol{\eta}_k$ with probability at least $\delta\in(0,1]$, we use the Doob's martingale convergence theorem. The sufficient condition for the Doob's theorem is for the $\mathcal{F}_k$–supermartingale $V_k(\boldsymbol{\eta}_k)$ to satisfy $\sup_k\ \mathbb{E}V^-_k(\boldsymbol{\eta}_k)<\infty$, where $V^-_k\stackrel{\triangle}{=}\max(-V_k,0)$ is defined as the negative part of the random variable $V_k(.)$. From (13), since $V_k(\boldsymbol{\eta}_k)$ is a summation of two terms (possibly with different signs), we need to show that $\mathbb{E}\mid V_k(\boldsymbol{\eta}_k)\mid$ is bounded above[16] for every $k\in\mathbb{Z}^+$. A sufficient condition for this is to show that

$$\begin{aligned}&\sup_k\ \mathbb{E}\ V(\boldsymbol{\eta}_k)<\infty\\&\sup_k\ \mathbb{E}\mid\boldsymbol{\eta}'_k\delta\vartheta_k(\boldsymbol{\eta}_k)\mid<\infty\end{aligned}$$

For the proof, we use induction on k. Assume $\mathbb{E}V(\boldsymbol{\eta}_0)$ is bounded. For the induction hypothesis, suppose that $\mathbb{E}V(\boldsymbol{\eta}_k)<\infty$ for some $k\in\mathbb{N}$. Then, we show that $\mathbb{E}V(\boldsymbol{\eta}_{k+1})<\infty$.

Using Lemma B.1 and (24), we write

$$\begin{aligned}&\rho\mid\boldsymbol{\eta}'_k\mathbb{E}_k\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k)\mid\\&\quad=\rho\mid\boldsymbol{\eta}'_k\mathbf{\Gamma}(\mathbb{E}_k\mathbf{z}_{k+1}-n^{-1}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}})\mid\\&\quad\le\rho\big(\ \|\mathbf{\Gamma}[\bar{\boldsymbol{f}}^{j'}\bar{\boldsymbol{p}}^j_{k+1}]_{j\in\mathcal{N}}\|+\|n^{-1}\mathbf{\Gamma}\mathbb{1}\mathbb{1}'\bar{\boldsymbol{z}}\|\ \big)\ \|\boldsymbol{\eta}_k\|\\&\quad\le\rho\big(\ \|\bar{f}^{(1)}_{max}\mathbf{\Lambda}\mathbb{1}\|+\|n^{-1}\bar{f}^{(1)}_{max}\mathbf{\Lambda}\mathbb{1}\mathbb{1}'\mathbb{1}\|\ \big)\ \|\boldsymbol{\eta}_k\|\\&\quad\le 2\sqrt{n}\rho\bar{f}^{(1)}_{max}(1+d_{max})\ \|\boldsymbol{\eta}_k\|\end{aligned}\tag{33}$$

By writing the Taylor series expansion of the Lyapunov function $V(\boldsymbol{\eta})$ in a neighborhood of $\boldsymbol{\eta}_k$, we have

$$\begin{aligned}&\mathbb{E}_kV(\boldsymbol{\eta}_{k+1})-V(\boldsymbol{\eta}_k)\\&\quad\le\rho\lambda_{max}(\mathbf{Q})\|\boldsymbol{\eta}_k\|^2+\rho\mid\boldsymbol{\eta}'_k\mathbb{E}_k\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k)\mid\\&\quad+\rho^2\mid\boldsymbol{\eta}'_k\mathbf{\Lambda}\mathbb{E}_k\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k)\mid+\frac{1}{2}\rho^2\mathbb{E}_k\|\mathbb{G}(\mathbf{z}_{k+1},\mathbf{z}_k)\|^2\end{aligned}\tag{34}$$

Substituting (27), (28), and (33) in (34) yields

$$\begin{aligned}&\mathbb{E}_kV(\boldsymbol{\eta}_{k+1})-V(\boldsymbol{\eta}_k)\\&\qquad\le\rho\lambda_{max}(\mathbf{Q})\|\boldsymbol{\eta}_k\|^2+\rho\varphi_1\|\boldsymbol{\eta}_k\|+\rho\varphi_2^2\end{aligned}$$

where $\varphi_1\stackrel{\triangle}{=}2\sqrt{n}\bar{f}^{(1)}_{max}(1+d_{max})(1+\rho\lambda_{max}(\mathbf{\Lambda}))$. If the error iterate $\boldsymbol{\eta}_k$ lies outside the unit sphere[17], then there exists a real $\varphi_3\stackrel{\triangle}{=}2(\varphi_1+\lambda_{max}(\mathbf{Q}))$ such that

$$\mathbb{E}_kV(\boldsymbol{\eta}_{k+1})\le(\rho\varphi_3+1)V(\boldsymbol{\eta}_k)+\rho\varphi_2^2\quad\mathbb{P}\text{–w.p.1}\tag{35}$$

where the marginal density of $(\boldsymbol{\eta}_0,\mathbf{Y}^j_k,j\in\mathcal{N}_i\cup\{i\})$ and the above inequality together with the induction hypothesis implies that $\mathbb{E}\ V(\boldsymbol{\eta}_{k+1})<\infty$.

Next, we show that under the induction hypothesis $\mathbb{E}V(\boldsymbol{\eta}_k)<\infty$, we have $\mathbb{E}\mid\boldsymbol{\eta}'_k\delta\vartheta_k(\boldsymbol{\eta}_k)\mid<\infty$ for some $k\in\mathbb{Z}^+$ and then it is shown that the following

$$\mathbb{E}\mid\boldsymbol{\eta}'_{k+1}\delta\vartheta_{k+1}(\boldsymbol{\eta}_{k+1})\mid<\infty$$

also holds.

Since $\sum_{i=k}^{\infty}\rho\beta^i_{k+1}=1$, from Lemma B.1, (11), and (24) we write

$$\begin{aligned}&\mid\boldsymbol{\eta}'_k\delta\vartheta_k(\boldsymbol{\eta}_k)\mid\\&\qquad\le\|\sum_{i=k}^{\infty}\rho\beta^i_{k+1}\mathbb{E}_k[\mathbb{Q}(\boldsymbol{\eta}_k,\boldsymbol{\xi}_{i+1})-\bar{\mathbb{Q}}(\boldsymbol{\eta}_k)]\|\ \|\boldsymbol{\eta}_k\|\\&\qquad\le\big(\|\mathbf{\Gamma}\sum_{i=k}^{\infty}\rho\beta^i_{k+1}\mathbb{E}_k\mathbf{z}_{i+1}\|+\|\mathbf{\Gamma}\bar{\boldsymbol{z}}\|\big)\|\boldsymbol{\eta}_k\|\\&\qquad\le\big(\|\mathbf{\Gamma}[\bar{\boldsymbol{f}}^{j'}\breve{\mathbf{X}}'_{k,i}\ \bar{\boldsymbol{p}}^j_{k+1}]_{j\in\mathcal{N}}\|+\|\bar{f}^{(1)}_{max}\mathbf{\Gamma}\mathbb{1}\|\big)\|\boldsymbol{\eta}_k\|\end{aligned}$$

where the matrix $\breve{\mathbf{X}}'_{k,i}\stackrel{\triangle}{=}\sum_{i=k}^{\infty}\rho\beta^i_{k+1}\mathbb{X}^{i\,\prime}_k$ is left stochastic and we have

$$\begin{aligned}\mid\boldsymbol{\eta}'_k\delta\vartheta_k(\boldsymbol{\eta}_k)\mid&\le 2\bar{f}^{(1)}_{max}\|\mathbf{\Lambda}\mathbb{1}\|\|\boldsymbol{\eta}_k\|\\&\le 2\sqrt{n}(1+d_{max})\bar{f}^{(1)}_{max}\|\boldsymbol{\eta}_k\|\end{aligned}$$

[16] note that $\mid V_k(\boldsymbol{\eta}_k)\mid=V^+_k(\boldsymbol{\eta}_k)+V^-_k(\boldsymbol{\eta}_k)$, where $V^+_k\stackrel{\triangle}{=}\max(V_k,0)$ is defined as the positive part of $V_k(.)$.

[17] note that in opposite case when $\|\boldsymbol{\eta}_k\|\le 1$, we get a uniform bound in (35) and the proof follows in a straightforward way.

and for $\boldsymbol{\eta}_k$ outside the unit sphere, there is a real $\varphi_4 \triangleq 4\sqrt{n}(1+d_{max})\bar{f}^{(1)}_{max}$ such that by the induction hypothesis we have

$$\mathbb{E} \mid \boldsymbol{\eta}'_k \delta\vartheta_k(\boldsymbol{\eta}_k) \mid \leq \ \varphi_4 \mathbb{E} V(\boldsymbol{\eta}_k) < \infty$$

Now, we show that $\mathbb{E} \mid \boldsymbol{\eta}'_{k+1}\delta\vartheta_{k+1}(\boldsymbol{\eta}_{k+1}) \mid < \infty$. From (11), we write

$$\begin{aligned}
&\mid \boldsymbol{\eta}'_{k+1}\delta\vartheta_{k+1}(\boldsymbol{\eta}_{k+1}) \mid \\
&\quad = \Big|\boldsymbol{\eta}'_{k+1} \sum_{i=k+1}^{\infty} \rho\beta^i_{k+2}\mathbb{E}_{k+1}[\mathbb{Q}(\boldsymbol{\eta}_{k+1},\boldsymbol{\xi}_{i+1}) - \bar{\mathbb{Q}}(\boldsymbol{\eta}_{k+1})]\Big| \\
&\quad \leq \Big|\boldsymbol{\eta}'_k \sum_{i=k+1}^{\infty} \rho\beta^i_{k+2}\mathbb{E}_{k+1}[\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i) - \bar{\mathbb{G}}]\Big| \\
&\quad (\star) + \rho\Big|\boldsymbol{\eta}'_k\boldsymbol{\Lambda} \sum_{i=k+1}^{\infty} \rho\beta^i_{k+2}\mathbb{E}_{k+1}[\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i) - \bar{\mathbb{G}}]\Big| \\
&\quad + \rho\Big|\mathbb{G}'(\mathbf{z}_{k+1},\mathbf{z}_k) \sum_{i=k+1}^{\infty} \rho\beta^i_{k+2}\mathbb{E}_{k+1}[\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i) - \bar{\mathbb{G}}]\Big|
\end{aligned} \tag{36}$$

We may compute

$$\sum_{i=k+1}^{\infty} \rho\beta^i_{k+2}\mathbb{E}_{k+1}\mathbf{z}_{i+1} = [\bar{\boldsymbol{f}}^{j'}\widetilde{\mathbf{X}}'_{k+1,i}\ \bar{\boldsymbol{p}}^j_{k+1}]_{j\in\mathcal{N}}$$

where $\widetilde{\mathbf{X}}'_{k+1,i} \triangleq \sum_{i=k+1}^{\infty}\rho\beta^i_{k+2}\mathbb{X}^{i\ \prime}_{k+1}$ is also a left stochastic matrix. As such, for the term $(\star)$ in (36) by replacing $\mathbb{E}_k$ with $\mathbb{E}_{k+1}$ in (30), the final upper bound will remain unchanged. Similar to (30), using Lemma B.1 we write

$$\begin{aligned}
&\Big|\boldsymbol{\eta}'_k \sum_{i=k+1}^{\infty} \rho\beta^i_{k+2}\mathbb{E}_{k+1}[\mathbb{G}(\mathbf{z}_{i+1},\mathbf{z}_i) - \bar{\mathbb{G}}]\Big| \\
&\quad = \Big|\boldsymbol{\eta}'_k\boldsymbol{\Gamma}\big(\sum_{i=k+1}^{\infty} \rho\beta^i_{k+2}\mathbb{E}_{k+1}\mathbf{z}_{i+1} - \bar{\boldsymbol{z}}\big)\Big| \\
&\quad \leq \big(\big\|\boldsymbol{\Gamma}[\bar{\boldsymbol{f}}^{j'}\widetilde{\mathbf{X}}'_{k+1,i}\ \bar{\boldsymbol{p}}^j_{k+1}]_{j\in\mathcal{N}}\big\| + \|\boldsymbol{\Gamma}\bar{\boldsymbol{z}}\|\big)\ \|\boldsymbol{\eta}_k\| \\
&\quad \leq \big(\|\bar{f}^{(1)}_{max}\boldsymbol{\Gamma}\mathbb{1}\| + \|\bar{f}^{(1)}_{max}\boldsymbol{\Gamma}\mathbb{1}\|\big)\ \|\boldsymbol{\eta}_k\| \\
&\quad \leq 2\bar{f}^{(1)}_{max}\|\boldsymbol{\Lambda}\mathbb{1}\|\ \|\boldsymbol{\eta}_k\| \\
&\quad \leq 2\sqrt{n}\bar{f}^{(1)}_{max}(1+d_{max})\|\boldsymbol{\eta}_k\|
\end{aligned} \tag{37}$$

Substituting (37), (30), and (31) in (36) we write for $\varphi_5 \triangleq 4\sqrt{n}\bar{f}^{(1)}_{max}(1+d_{max})(1+\rho\lambda_{max}(\boldsymbol{\Lambda}))$

$$\begin{aligned}
&\mid \boldsymbol{\eta}'_{k+1}\delta\vartheta_{k+1}(\boldsymbol{\eta}_{k+1}) \mid \\
&\quad \leq \ \frac{1}{2}\varphi_5\|\boldsymbol{\eta}_k\| + 2\rho\bar{f}^{(1)}_{max}\|\boldsymbol{\Lambda}\mathbb{1}\|^2_2\big(\|\mathbf{z}_{k+1}\|_1 + \bar{f}^{(1)}_{max}\big)
\end{aligned}$$

where again for $\boldsymbol{\eta}_k$ outside the unit sphere we have

$$\begin{aligned}
&\mathbb{E} \mid \boldsymbol{\eta}'_{k+1}\delta\vartheta_{k+1}(\boldsymbol{\eta}_{k+1}) \mid \\
&\quad \leq \ \varphi_5\mathbb{E}V(\boldsymbol{\eta}_k) + 2\rho\bar{f}^{(1)}_{max}\|\boldsymbol{\Lambda}\mathbb{1}\|^2_2\big(\mathbb{E}\|\mathbf{z}_{k+1}\|_1 + \bar{f}^{(1)}_{max}\big) \\
&\quad \leq \ \varphi_5\mathbb{E}V(\boldsymbol{\eta}_k) + 2n\rho\bar{f}^{(1)}_{max}(1+d_{max})^2\big(\|\mathbb{E}\mathbf{z}_{k+1}\|_1 + \bar{f}^{(1)}_{max}\big)
\end{aligned}$$

The expected value $\mathbb{E}\mathbf{z}_{k+1}$ can be computed as

$$\mathbb{E}\mathbf{z}_{k+1} = [\bar{\boldsymbol{f}}^{j'}\mathbb{X}^{k+1\ \prime}_0\pi]_{j\in\mathcal{N}} \tag{38}$$

Note that for every $k\in\mathbb{Z}^+$, $\mathbb{X}^{k\ \prime}_0 \triangleq \prod_{\kappa=1}^{k}\mathbf{X}'$ is a left stochastic matrix. Thus, by the induction hypothesis we have

$$\begin{aligned}
&\mathbb{E} \mid \boldsymbol{\eta}'_{k+1}\delta\vartheta_{k+1}(\boldsymbol{\eta}_{k+1}) \mid \\
&\leq \ \varphi_5\mathbb{E}V(\boldsymbol{\eta}_k) + 2n\rho\bar{f}^{(1)}_{max}(1+d_{max})^2\big(\|[\bar{\boldsymbol{f}}^{j'}\mathbb{X}^{k+1\ \prime}_0\pi]_{j\in\mathcal{N}}\|_1 + \bar{f}^{(1)}_{max}\big) \\
&\leq \ \varphi_5\mathbb{E}V(\boldsymbol{\eta}_k) + 2n\rho\bar{f}^{(1)}_{max}(1+d_{max})^2\big(\|\bar{f}^{(1)}_{max}\mathbb{1}\|_1 + \bar{f}^{(1)}_{max}\big) \\
&= \ \varphi_5\mathbb{E}V(\boldsymbol{\eta}_k) + 2n(n+1)\rho(1+d_{max})^2(\bar{f}^{(1)}_{max})^2 \ < \infty
\end{aligned}$$

Therefore by induction, we have shown that $\mathbb{E} \mid V_k(\boldsymbol{\eta}_k) \mid < \infty$ for all $k\in\mathbb{Z}^+$ and thus $\sup_k\ \mathbb{E}V^-_k(\boldsymbol{\eta}_k) \leq M < \infty$.

Now, from the martingale convergence theorem[18], due to Doob, there exists a random variable U satisfying $\mathbb{E} \mid U \mid \leq \ M$ such that the pointwise limit

$$\lim_{k\to\infty} V_k(\boldsymbol{\eta}_k(\omega)) = U(\omega) \tag{39}$$

exists for $\mathbb{P}$–*almost all* $\omega\in\Omega$. Hence, $V_k(\boldsymbol{\eta}_k) \to U$ $\mathbb{P}$–a.s. as $k\to\infty$. From this and Theorem 7.3 [11, Chapter 6], the compact set $\Omega_{\bar{c}}$ is again reached $\mathbb{P}$–w.p.1 after each time it is exited. This means that $\boldsymbol{\eta}_k$ returns to $\Omega_{\bar{c}}$ infinitely often $\mathbb{P}$–w.p.1. Also, from Theorem 7.3 [11, Chapter 6] we have that the sequence $\{\boldsymbol{\eta}_k\}$ is bounded in probability, that is,

$$\lim_{\mathcal{E}\to\infty}\sup_k\ \mathbb{P}[\|\boldsymbol{\eta}_k\| \geq \mathcal{E}] = 0$$

and thus using Theorem 7.2 [11, Chapter 6] given any $\delta\in(0,1]$, there is a compact set $\mathbb{L}_\delta$ such that the iterates $\boldsymbol{\eta}_k$ enter $\mathbb{L}_\delta$ infinitely often with probability at least δ. In particular, we may choose $\mathbb{L}_\delta = \Omega_{\bar{c}}$ for all $\delta\in(0,1]$. The proof of the theorem is concluded. ■

[18]the proof of the Doob's theorem can be found in most classic books on probability theory, see, e.g., [14, Theorem 35.5.].