Distributed Parameter Estimation in Sensor Networks: Nonlinear Observation Models and Imperfect Communication

Soummya Kar*, José M. F. Moura* and Kavita Ramanan†

Abstract

The paper studies the problem of distributed static parameter (vector) estimation in sensor networks with nonlinear observation models and imperfect inter-sensor communication. We introduce the concept of *separably estimable* observation models, which generalize the observability condition for linear centralized estimation to nonlinear distributed estimation. We study the algorithms $\mathcal{NU}$ (with its linear counterpart $\mathcal{LU}$) and $\mathcal{NLU}$ for distributed estimation in separably estimable models. We prove consistency (all sensors reach consensus almost sure and converge to the true parameter value), asymptotic unbiasedness and asymptotic normality of these algorithms. Both the algorithms are characterized by appropriately chosen decaying weight sequences in the estimate update rule. While the algorithm $\mathcal{NU}$ is analyzed in the framework of stochastic approximation theory, the algorithm $\mathcal{NLU}$ exhibits mixed time-scale behavior and biased perturbations and require a different approach, which we develop in the paper.

Keywords. Distributed parameter estimation, separable estimable, stochastic approximation, consistency, unbiasedness, asymptotic normality, spectral graph theory, Laplacian

I. Introduction

A. Background and Motivation

Wireless sensor network (WSN) applications generally consist of a large number of sensors which coordinate to perform a task in a distributed fashion. Unlike fusion-center based applications, there is no center and the task is performed locally at each sensor with intermittent inter-sensor message exchanges. In a coordinated environment monitoring or surveillance task, it translates to each sensor observing a part of the field of interest. With such local information, it is not possible for a particular sensor to get a reasonable estimate of the field. The sensors

Names appear in alphabetical order.

* Soummya Kar and José M. F. Moura are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA 15213 (e-mail: soummyak@andrew.cmu.edu, moura@ece.cmu.edu, ph: (412) 268-6341, fax: (412) 268-3890.)

† Kavita Ramanan is with the Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, USA 15213 (e-mail: kramanan@math.cmu.edu, ph: (412) 268-8485 , fax: (412) 268-6380.)

The work of Soummya Kar and José M. F. Moura was supported by the DARPA DSO Advanced Computing and Mathematics Program Integrated Sensing and Processing (ISP) Initiative under ARO grant # DAAD19-02-1-0180, by NSF under grants # ECS-0225449 and # CNS-0428404, and by an IBM Faculty Award. The work of Kavita Ramanan was supported by the NSF under grants DMS 0405343 and CMMI 0728064.

need to cooperate then and this is achieved by intermittent data exchanges among the sensors, whereby each sensor fuses its version of the estimate from time to time with those of other sensors with which it can communicate (in this context, see [1], [2], [3], [4], for a treatment of general distributed stochastic algorithms.) We consider the above problem in this paper in the context of distributed parameter estimation in WSNs. As an abstraction of the environment, we model it by a static vector parameter, whose dimension, M, can be arbitrarily large. We assume that each sensor receives noisy measurements (not necessarily additive) of only a part of the parameter vector. More specifically, if M_n is the dimension of the observation space of the n-th sensor, $M_n \ll M$. Assuming that the rate of receiving observations at each sensor is comparable to the data exchange rate among sensors, each sensor updates its estimate at time index i by fusing it appropriately with the observation (innovation) received at i and the estimates at i of those sensors with which it can communicate at i. We propose and study two generic recursive distributed iterative estimation algorithms in this paper, namely, $\mathcal{NU}$ and $\mathcal{NLU}$ for distributed parameter estimation with possibly nonlinear observation models at each sensor. As is required, even by centralized estimation schemes, for the estimate sequences generated by the $\mathcal{NU}$ and $\mathcal{NLU}$ algorithms at each sensor to have desirable statistical properties, we need to impose some observability condition. To this end, we introduce a generic observability condition, the *separably estimable* condition for distributed parameter estimation in nonlinear observation models, which generalize the observability condition of centralized parameter estimation.

The inter-sensor communication is quantized with random link (communication channel) failures. This is appropriate, for example, in *digital* communication WSN when the data exchanges among a sensor and its neighbors are quantized, and the communication channels (or links) among sensors may fail at random times, e.g., as when packet dropouts occur randomly. We consider a very generic model of temporally independent link failures, whereby it is assumed that the sequence of network Laplacians, $\{L(i)\}_{i\geq 0}$ are i.i.d. with mean $\overline{L}$ and satisfying $\lambda_2(\overline{L}) > 0$. We do not make any distributional assumptions on the link failure model. Although the link failures, and so the Laplacians, are independent at different times, during the same iteration, the link failures can be spatially dependent, i.e., correlated. This is more general and subsumes the erasure network model, where the link failures are independent over space *and* time. Wireless sensor networks motivate this model since interference among the wireless communication channels correlates the link failures over space, while, over time, it is still reasonable to assume that the channels are memoryless or independent. In particular, we do not require that the random instantiations of communication graph be connected; in fact, it is possible to have all these instantiations to be disconnected. We only require that the graph stays connected on *average*. This is captured by requiring that $\lambda_2\left(\overline{L}\right) > 0$, enabling us to capture a broad class of asynchronous communication models, as will be explained in the paper.

As is required by even centralized estimation schemes, for the estimate sequences generated by the $\mathcal{NU}$ and $\mathcal{NLU}$ algorithms to have desirable statistical properties, we need to impose some observability condition. To this end, we introduce a generic observability condition, the *separably estimable* condition for distributed parameter estimation in nonlinear observation models, which generalize the observability condition of centralized parameter estimation. To motivate the separably estimable condition for nonlinear problems, we start with the linear model for which it reduces to a rank condition on the overall observability Grammian. We propose the algorithm $\mathcal{LU}$ for the linear model and using stochastic approximation show that the estimate sequence generated at each sensor is consistent, asymptotically unbiased and asymptotically normal. We explicitly characterize the asymptotic variance and in certain cases, compare it with the asymptotic variance of a centralized scheme. The $\mathcal{LU}$ algorithm can be regarded as a generalization of consensus algorithms (see, for example, [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17]), the latter being a specific case of the $\mathcal{LU}$ with no innovations. The algorithm $\mathcal{NU}$ is the natural generalization of the $\mathcal{LU}$ to nonlinear separably estimably models. Under reasonably assumptions on the model, we prove consistency, asymptotic unbiasedness and asymptotic normality of the algorithm $\mathcal{NU}$. An important aspect of these algorithms is the time-varying weight sequences (decaying to zero as the iterations progress) associated with the consensus and innovation updates. The algorithm $\mathcal{NU}$ (and its linear counterpart $\mathcal{LU}$) is characterized by the same decay rate of the consensus and innovation weight sequences and hence, its analysis falls under the framework of stochastic approximation. The algorithm $\mathcal{NU}$, though provides desirable performance guarantees (consistency, asymptotic unbiasedness and asymptotic normality), requires further assumptions on the separably estimable observation models. We thus introduce the $\mathcal{NLU}$ algorithm, which leads to consistent and asymptotic unbiasedness estimators at each sensor for all separably estimable models. In the context of stochastic algorithms, $\mathcal{NLU}$ can be viewed as exhibiting mixed time-scale behavior (the weight sequences associated with the consensus and innovation updates decay at different rates) and consisting of unbiased perturbations (detailed explanation is provided in the paper.) The $\mathcal{NLU}$ algorithm does not fall under the purview of standard stochastic approximation theory, and its analysis requires an altogether different framework as developed in the paper. The algorithm $\mathcal{NLU}$ is thus more reliable than the $\mathcal{NU}$ algorithm, as the latter requires further assumptions on the separably estimable observation models. On the other hand, in cases where the $\mathcal{NU}$ algorithm is applicable, it provides convergence rate guarantees (for example, asymptotic normality) which follow from standard stochastic approximation theory, while $\mathcal{NLU}$ does not follow under the purview of standard stochastic approximation theory and hence does not inherit these convergence rate properties.

We comment on the relevant recent literature on distributed estimation in WSNs. The papers [18], [19], [20], [21] study the estimation problem in static networks, where either the sensors take a single snapshot of the field at the start and then initiate distributed consensus protocols (or more generally distributed optimization, as in [19]) to fuse the initial estimates, or the observation rate of the sensors is assumed to be much slower than the inter-sensor communicate rate, thus permitting a separation of the two time-scales. On the contrary, our work considers new observations at every and the consensus and observation (innovation) updates are incorporated in the same iteration. More relevant to our present work are [22], [23], [24], [25], which consider the linear estimation problem in non-random networks, where the observation and consensus protocols are incorporated in the same iteration. In [22], [24] the distributed linear estimation problems are treated in the context of distributed least-mean-square (LMS) filtering, where constant weight sequences are used to prove mean-square stability of the filter. The use of non-decaying combining weights in [22], [24], [25] lead to a residual error, however, under appropriate assumptions, these algorithms can be adapted for tracking certain time-varying parameters. The distributed LMS algorithm in [23] also considers decaying weight sequences, thereby establishing $\mathcal{L}_2$ convergence to the true parameter value. Apart from treating generic separably estimable nonlinear observation models, in the linear case our algorithm $\mathcal{LU}$ leads to asymptotic normality in addition to consistency and asymptotic unbiasedness in random time-varying networks with quantized inter-sensor communication.

We briefly comment on the organization of the rest of the paper. The rest of this section introduces notation and preliminaries, to be adopted throughout the paper. To motivate the generic nonlinear problem, we study the linear case (algorithm $\mathcal{LU}$) in Section II. Section III studies the generic separably estimable models and the algorithm $\mathcal{NU}$, whereas, algorithm $\mathcal{NLU}$ is presented in Section IV. Finally, Section V concludes the paper.

B. *Notation*

For completeness, this subsection sets notation and presents preliminaries on algebraic graph theory, matrices, and dithered quantization to be used in the sequel.

Preliminaries. We denote the k-dimensional Euclidean space by $\mathbb{R}^{k\times 1}$. The $k\times k$ identity matrix is denoted by I_k, while $\mathbf{1}_k, \mathbf{0}_k$ denote respectively the column vector of ones and zeros in $\mathbb{R}^{k\times 1}$. We also define the rank one $k\times k$ matrix P_k by

$$P_k = \frac{1}{k}\mathbf{1}_k\mathbf{1}_k^T \tag{1}$$

The only non-zero eigenvalue of P_k is one, and the corresponding normalized eigenvector is $\left(1/\sqrt{k}\right)\mathbf{1}_k$. The operator $\|\cdot\|$ applied to a vector denotes the standard Euclidean 2-norm, while applied to matrices denotes the induced 2-norm, which is equivalent to the matrix spectral radius for symmetric matrices.

We assume that the parameter to be estimated belongs to a subset $\mathcal{U}$ of the Euclidean space $\mathbb{R}^{M\times 1}$. Throughout the paper, the true (but unknown) value of the parameter is denoted by θ^*. We denote a canonical element of $\mathcal{U}$ by θ. The estimate of θ^* at time i at sensor n is denoted by $\mathbf{x}_n(i)\in\mathbb{R}^{M\times 1}$. Without loss of generality, we assume that the initial estimate, $\mathbf{x}_n(0)$, at time 0 at sensor n is a non-random quantity.

Throughout, we assume that all the random objects are defined on a common measurable space, $(\Omega, \mathcal{F})$. In case the true (but unknown) parameter value is θ^*, the probability and expectation operators are denoted by $\mathbb{P}_{\theta^*}[\cdot]$ and $\mathbb{E}_{\theta^*}[\cdot]$, respectively. When the context is clear, we abuse notation by dropping the subscript. Also, all inequalities involving random variables are to be interpreted a.s. (almost surely.)

Spectral graph theory. We review elementary concepts from spectral graph theory. For an *undirected* graph $G=(V,E)$, $V=[1\cdots N]$ is the set of nodes or vertices, $|V|=N$, and E is the set of edges, $|E|=M$, where $|\cdot|$ is the cardinality. The unordered pair $(n,l)\in E$ if there exists an edge between nodes n and l. We only consider simple graphs, i.e., graphs devoid of self-loops and multiple edges. A graph is connected if there exists a path[1], between each pair of nodes. The neighborhood of node n is

$$\Omega_n = \{l\in V \mid (n,l)\in E\} \tag{2}$$

[1]A path between nodes n and l of length m is a sequence $(n=i_0, i_1, \cdots, i_m=l)$ of vertices, such that, $(i_k, i_{k+1})\in E\,\forall\; 0\le k\le m-1$.

Node n has degree $d_n = |\Omega_n|$ (number of edges with n as one end point.) The structure of the graph can be described by the symmetric $N \times N$ adjacency matrix, $A = [A_{nl}]$, $A_{nl} = 1$, if $(n, l) \in E$, $A_{nl} = 0$, otherwise. Let the degree matrix be the diagonal matrix $D = \text{diag}\,(d_1 \cdots d_N)$. The graph Laplacian matrix, L, is

$$L = D - A \tag{3}$$

The Laplacian is a positive semidefinite matrix; hence, its eigenvalues can be ordered as

$$0 = \lambda_1(L) \leq \lambda_2(L) \leq \cdots \leq \lambda_N(L) \tag{4}$$

The smallest eigenvalue $\lambda_1(l)$ is always equal to zero, with $\left(1/\sqrt{N}\right)\mathbf{1}_N$ being the corresponding normalized eigenvector. The multiplicity of the zero eigenvalue equals the number of connected components of the network; for a connected graph, $\lambda_2(L) > 0$. This second eigenvalue is the algebraic connectivity or the Fiedler value of the network; see [26], [27], [28] for detailed treatment of graphs and their spectral theory.

Kronecker product. Since, we are dealing with vector parameters, most of the matrix manipulations will involve Kronecker products. For example, the Kronecker product of the $N \times N$ matrix L and I_M will be an $NM \times NM$ matrix, denoted by $L \otimes I_M$. We will deal often with matrices of the form $C = [I_{NM} - bL \otimes I_M - aI_{NM} - P_N \otimes I_M]$. It follows from the properties of Kronecker products and the matrices L, P, that the eigenvalues of this matrix C are $-a$ and $1 - b\lambda_i(L) - a,\ 2 \leq i \leq N$, each being repeated M times.

We now review results from statistical quantization theory.

Quantizer: We assume that all sensors are equipped with identical quantizers, which uniformly quantize each component of the M-dimensional estimates by the quantizing function, $\mathbf{q}(\cdot) : \mathbb{R}^{M\times 1} \to \mathcal{Q}^M$. For $\mathbf{y} \in \mathbb{R}^{M\times 1}$ the channel input,

$$\begin{aligned} \mathbf{q}(\mathbf{y}) &= [k_1\Delta, \cdots, k_M\Delta], \qquad (k_m - \frac{1}{2})\Delta \leq \ y_i \ < (k_m + \frac{1}{2})\Delta,\ 1 \leq m \leq M \quad (5) \\ &= \mathbf{y} + \mathbf{e}(\mathbf{y}), \qquad -\frac{\Delta}{2}\mathbf{1}_N \leq \mathbf{e}(\mathbf{y}) < \frac{\Delta}{2}\mathbf{1}_N, \quad \forall \mathbf{y} \quad (6) \end{aligned}$$

where $\mathbf{e}(\mathbf{y})$ is the quantization error and the inequalities are interpreted component-wise. The quantizer alphabet is

$$\mathcal{Q}^M = \left\{ [k_1\Delta, \cdots, k_M\Delta]^T \,\middle|\, k_i \in \mathbb{Z},\ \forall i \right\} \tag{7}$$

We take the quantizer alphabet to be countable because no *à priori* bound is assumed on the parameter.

Conditioned on the input, the quantization error $\mathbf{e}(\mathbf{y})$ is deterministic. This strong correlation of the error with the input creates unacceptable statistical properties. In particular, for iterative algorithms, it leads to error accumulation and divergence of the algorithm (see the discussion in [29].) To avoid this divergence, we consider dithered quantization, which makes the quantization error possess nice statistical properties. We review briefly basic results on dithered quantization, which are needed in the sequel.

Dithered Quantization: Schuchman Conditions Consider a uniform scalar quantizer $q(\cdot)$ of step-size Δ, where $y \in \mathbb{R}$ is the channel input. Let $\{y(i)\}_{i\geq 0}$ be a scalar input sequence to which we added a dither sequence $\{\nu(i)\}_{i\geq 0}$

of i.i.d. uniformly distributed random variables on $[-\Delta/2, \Delta/2)$, independent of the input sequence $\{y(i)\}_{i\geq 0}$. This is a sufficient condition for the dither to satisfy the Schuchman conditions (see [30], [31], [32], [33]). Under these conditions, the error sequence for subtractively dithered systems ([31]) $\{\varepsilon(i)\}_{i\geq 0}$

$$\varepsilon(i) = q(y(i) + \nu(i)) - (y(i) + \nu(i)) \tag{8}$$

is an i.i.d. sequence of uniformly distributed random variables on $[-\Delta/2, \Delta/2)$, which is independent of the input sequence $\{y(i)\}_{i\geq 0}$. To be more precise, this result is valid if the quantizer does not overload, which is trivially satisfied here as the dynamic range of the quantizer is the entire real line. Thus, by randomizing appropriately the input to a uniform quantizer, we can render the error to be independent of the input and uniformly distributed on $[-\Delta/2, \Delta/2)$. This leads to nice statistical properties of the error, which we will exploit in this paper.

Random Link Failure. In digital communications, packets may be lost at random times. To account for this, we let the links (or communication channels among sensors) to fail, so that the edge set and the connectivity graph of the sensor network are time varying. Accordingly, the sensor network at time i is modeled as an undirected graph, $G(i) = (V, E(i))$ and the graph Laplacians as a sequence of i.i.d. Laplacian matrices $\{L(i)\}_{i\geq 0}$. We write

$$L(i) = \overline{L} + \widetilde{L}(i), \ \forall i \geq 0 \tag{9}$$

where the mean $\overline{L} = \mathbb{E}[L(i)]$. We do not make any distributional assumptions on the link failure model. Although the link failures, and so the Laplacians, are independent at different times, during the same iteration, the link failures can be spatially dependent, i.e., correlated. This is more general and subsumes the erasure network model, where the link failures are independent over space *and* time. Wireless sensor networks motivate this model since interference among the wireless communication channels correlates the link failures over space, while, over time, it is still reasonable to assume that the channels are memoryless or independent.

Connectedness of the graph is an important issue. We do not require that the random instantiations $G(i)$ of the graph be connected; in fact, it is possible to have all these instantiations to be disconnected. We only require that the graph stays connected on *average*. This is captured by requiring that $\lambda_2\left(\overline{L}\right) > 0$, enabling us to capture a broad class of asynchronous communication models; for example, the random asynchronous gossip protocol analyzed in [34] satisfies $\lambda_2\left(\overline{L}\right) > 0$ and hence falls under this framework.

II. Distributed Linear Parameter Estimation: Algorithm $\mathcal{LU}$

In this section, we consider the algorithm $\mathcal{LU}$ for *distributed* parameter estimation when the observation model is linear. This problem motivates the generic *separably estimable* nonlinear observation models considered in Sections III and IV. Subsection II-A sets up the distributed linear estimation problem and presents the algorithm $\mathcal{LU}$. Subsection II-B establishes the consistency and asymptotic unbiasedness of the $\mathcal{LU}$ algorithm, where we show that, under the $\mathcal{LU}$ algorithm, all sensors converge a.s. to the true parameter value, θ^*. Convergence rate analysis (asymptotic normality) is carried out in Subsection II-C, while Subsection II-D illustrates $\mathcal{LU}$ with an example.

A. Problem Formulation: Algorithm $\mathcal{LU}$

Let $\theta^* \in \mathbb{R}^{M\times 1}$ be an M-dimensional parameter that is to be estimated by a network of N sensors. We refer to θ as a parameter, although it is a vector of M parameters. Each sensor makes i.i.d. observations of noise corrupted linear functions of the parameter. We assume the following observation model for the n-th sensor:

$$\mathbf{z}_n(i) = H_n(i)\theta^* + \zeta_n(i) \tag{10}$$

where: $\left\{\mathbf{z}_n(i) \in \mathbb{R}^{M_n \times 1}\right\}_{i\geq 0}$ is the i.i.d. observation sequence for the n-th sensor; $\{\zeta_n(i)\}_{i\geq 0}$ is a zero-mean i.i.d. noise sequence of bounded variance; and $\{H_n(i)\}_{i\geq 0}$ is an i.i.d. sequence of observation matrices with mean $\overline{H}_n$ and bounded second moment. For most practical sensor network applications, each sensor observes only a subset of M_n of the components of θ, with $M_n \ll M$. Under such a situation, in isolation, each sensor can estimate at most only a part of the parameter. However, if the sensor network is connected in the mean sense (see Section I-B), and under appropriate observability conditions, we will show that it is possible for each sensor to get a consistent estimate of the parameter θ^* by means of quantized local inter-sensor communication.

In this subsection, we present the algorithm $\mathcal{LU}$ for distributed parameter estimation in the linear observation model (10). Starting from some initial deterministic estimate of the parameters (the initial states may be random, we assume deterministic for notational simplicity), $\mathbf{x}_n(0) \in \mathbb{R}^{M\times 1}$, each sensor generates by a distributed iterative algorithm a sequence of estimates, $\{\mathbf{x}_n(i)\}_{i\geq 0}$. The parameter estimate $\mathbf{x}_n(i+1)$ at the n-th sensor at time $i+1$ is a function of: its previous estimate; the communicated quantized estimates at time i of its neighboring sensors; and the new observation $\mathbf{z}_n(i)$. As described in Section I-B, the data is subtractively dithered quantized, i.e., there exists a vector quantizer $\mathbf{q}(.)$ and a family, $\{\nu_{nl}^m(i)\}$, of i.i.d. uniformly distributed random variables on $[-\Delta/2, \Delta/2)$ such that the quantized data received by the n-th sensor from the l-th sensor at time i is $\mathbf{q}(\mathbf{x}_l(i) + \nu_{nl}(i))$, where $\nu_{nl}(i) = [\nu_{nl}^1(i), \cdots, \nu_{nl}^M(i)]^T$. It then follows from the discussion in Section I-B that the quantization error, $\varepsilon_{nl}(i) \in \mathbb{R}^{M\times 1}$ given by (8), is a random vector, whose components are i.i.d. uniform on $[-\Delta/2, \Delta/2)$ and independent of $\mathbf{x}_l(i)$.

Algorithm $\mathcal{LU}$ Based on the current state $\mathbf{x}_n(i)$, the quantized exchanged data $\{\mathbf{q}(\mathbf{x}_l(i) + \nu_{nl}(i))\}_{l\in\Omega_n(i)}$, and the observation $\mathbf{z}_n(i)$, we update the estimate at the n-th sensor by the following distributed iterative algorithm:

$$\mathbf{x}_n(i+1) = \mathbf{x}_n(i) - \alpha(i)\left[b\sum_{l\in\Omega_n(i)}\left(\mathbf{x}_n(i) - \mathbf{q}(\mathbf{x}_l(i) + \nu_{nl}(i))\right) - \overline{H}_n^T\left(\mathbf{z}_n(i) - \overline{H}_n\mathbf{x}_n(i)\right)\right] \tag{11}$$

In (11), $b > 0$ is a constant and $\{\alpha(i)\}_{i\geq 0}$ is a sequence of weights with properties to be defined below. Algorithm (11) is distributed because for sensor n it involves only the data from the sensors in its neighborhood $\Omega_n(i)$. Using eqn. (8), the state update can be written as

$$\mathbf{x}_n(i+1) = \mathbf{x}_n(i) - \alpha(i)\left[b\sum_{l\in\Omega_n(i)}\left(\mathbf{x}_n(i) - \mathbf{x}_l(i)\right) - \overline{H}_n^T\left(\mathbf{z}_n(i) - \overline{H}_n\mathbf{x}_n(i)\right) + b\nu_{nl}(i) + b\varepsilon_{nl}(i)\right] \tag{12}$$

We rewrite (12) in compact form. Define the random vectors, $\mathbf{\Upsilon}(i)$ and $\mathbf{\Psi}(i) \in \mathbb{R}^{NM\times 1}$ with vector components

$$\mathbf{\Upsilon}_n(i) = -\sum_{l\in\Omega_n(i)} \nu_{nl}(i) \tag{13}$$

$$\mathbf{\Psi}_n(i) = -\sum_{l\in\Omega_n(i)} \varepsilon_{nl}(i) \tag{14}$$

It follows from the Schuchman conditions on the dither, see Section I-B, that

$$\mathbb{E}\left[\mathbf{\Upsilon}(i)\right] = \mathbb{E}\left[\mathbf{\Psi}(i)\right] = \mathbf{0},\ \forall i \tag{15}$$

$$\sup_i \mathbb{E}\left[\|\mathbf{\Upsilon}(i)\|^2\right] = \sup_i \mathbb{E}\left[\|\mathbf{\Psi}(i)\|^2\right] \leq \frac{N(N-1)M\Delta^2}{12} \tag{16}$$

from which we then have

$$\begin{aligned}\sup_i \mathbb{E}\left[\|\mathbf{\Upsilon}(i)+\mathbf{\Psi}(i)\|^2\right] &\leq 2\sup_i \mathbb{E}\left[\|\mathbf{\Upsilon}(i)\|^2\right] + 2\sup_i \mathbb{E}\left[\|\mathbf{\Psi}(i)\|^2\right] \\ &\leq \frac{N(N-1)M\Delta^2}{3} \\ &= \eta_q \end{aligned}\tag{17}$$

Also, define the noise covariance matrix S_q as

$$S_q = \mathbb{E}\left[\left(\mathbf{\Upsilon}(i)+\mathbf{\Psi}(i)\right)\left(\mathbf{\Upsilon}(i)+\mathbf{\Psi}(i)\right)^T\right] \tag{18}$$

The iterations in (11) can be written in compact form as:

$$\mathbf{x}(i+1) = \mathbf{x}(i) - \alpha(i)\left[b(L(i)\otimes I_M)\mathbf{x}(i) - \overline{D}_{\overline{H}}\left(\mathbf{z}(i) - \overline{D}_{\overline{H}}^T\mathbf{x}(i)\right) + b\mathbf{\Upsilon}(i) + b\mathbf{\Psi}(i)\right] \tag{19}$$

Here, $\mathbf{x}(i) = \left[\mathbf{x}_1^T(i)\cdots\mathbf{x}_N^T(i)\right]^T$ is the vector of sensor states (estimates.) The sequence of Laplacian matrices $\{L(i)\}_{i\geq 0}$ captures the topology of the sensor network . They are random, see Section I-B, to accommodate link failures, which occur in packet communications. We also define the matrices $\overline{D}_{\overline{H}}$ and $D_{\overline{H}}$ as

$$\overline{D}_{\overline{H}} = \text{diag}\left[\overline{H}_1^T\cdots\overline{H}_N^T\right] \text{ and } D_{\overline{H}} = \overline{D}_{\overline{H}}\overline{D}_{\overline{H}}^T = \text{diag}\left[\overline{H}_1^T\overline{H}_1\cdots\overline{H}_N^T\overline{H}_N\right] \tag{20}$$

We refer to the recursive estimation algorithm in eqn. (19) as $\mathcal{LU}$. We now summarize formally the assumptions on the $\mathcal{LU}$ algorithm and their implications.

A.1)Observation Noise. Recall the observation model in eqn. (10). We assume that the observation noise process, $\left\{\zeta(i) = \left[\zeta_1^T(i),\cdots,\zeta_N^T(i)\right]^T\right\}_{i\geq 0}$ is an i.i.d. zero mean process, with finite second moment. In particular, the observation noise covariance is independent of i

$$\mathbb{E}\left[\zeta(i)\zeta^T(j)\right] = S_\zeta\delta_{ij},\ \forall i,j\geq 0 \tag{21}$$

where the Kronecker symbol $\delta_{ij} = 1$ if $i=j$ and zero otherwise. Note that the observation noises at different sensors may be correlated during a particular iteration. Eqn. (21) states only temporal independence. The spatial correlation of the observation noise makes our model applicable to practical sensor network problems, for instance, for distributed target localization, where the observation noise is generally correlated across sensors.

A.2)Observability. We assume that the observation matrices, $\{[H_1(i), \cdots, H_N(i)]\}_{i\geq 0}$, form an i.i.d. sequence with mean $\left[\overline{H}_1, \cdots, \overline{H}_N\right]$ and finite second moment. In particular, we have

$$H_n(i) = \overline{H}_n + \widetilde{H}_n(i), \; \forall i, n \tag{22}$$

where, $\overline{H}_n = \mathbb{E}\left[H_n(i)\right], \; \forall i, n$ and $\left\{\left[\widetilde{H}_1(i), \cdots, \widetilde{H}_N(i)\right]\right\}_{i\geq 0}$ is a zero mean i.i.d. sequence with finite second moment. Here, also, we require only temporal independence of the observation matrices, but allow them to be spatially correlated. We require the following global observability condition. The matrix G

$$G = \sum_{n=1}^{N} \overline{H}_n^T \overline{H}_n \tag{23}$$

is full-rank. This distributed observability extends the observability condition for a centralized estimator to get a consistent estimate of the parameter θ^*. We note that the information available to the n-th sensor at any time i about the corresponding observation matrix is just the mean $\overline{H}_n$, and *not* the random $H_n(i)$. Hence, the state update equation uses only the $\overline{H}_n$'s, as given in eqn. (11).

A.3)Persistence Condition. The weight sequence $\{\alpha(i)\}_{i\geq 0}$ satisfies

$$\alpha(i) > 0, \; \sum_{i\geq 0} \alpha(i) = \infty, \quad \sum_{i\geq 0} \alpha^2(i) < \infty \tag{24}$$

This condition is commonly assumed in adaptive control and signal processing and implies, in particular, that, $\alpha(i) \to 0$. Examples include

$$\alpha(i) = \frac{1}{i^\beta}, \;\; .5 < \beta \leq 1 \tag{25}$$

A.4)Independence Assumptions. The sequences $\{L(i)\}_{i\geq 0}$,$\{\zeta_n(i)\}_{1\leq n\leq N, \; i\geq 0}$,$\{H_n(i)\}_{1\leq n\leq N, i\geq 0}$,$\{\nu_{nl}^m(i)\}$ are mutually independent.

Markov. Consider the filtration, $\{\mathcal{F}_i^{\mathbf{x}}\}_{i\geq 0}$, given by

$$\mathcal{F}_i^{\mathbf{x}} = \sigma\left(\mathbf{x}(0), \{L(j), \mathbf{z}(j), \mathbf{\Upsilon}(j), \mathbf{\Psi}(j)\}_{0\leq j<i}\right) \tag{26}$$

It then follows that the random objects $L(i), \mathbf{z}(i), \mathbf{\Upsilon}(i), \mathbf{\Psi}(i)$ are independent of $\mathcal{F}_i^{\mathbf{x}}$, rendering $\{\mathbf{x}(i), \mathcal{F}_i^{\mathbf{x}}\}_{i\geq 0}$ a Markov process.

B. Consistency of $\mathcal{LU}$

We recall standard definitions from sequential estimation theory (see, for example, [35]).

Definition 1 (Consistency) : A sequence of estimates $\{\mathbf{x}^{\bullet}(i)\}_{i\geq 0}$ is called consistent if

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\mathbf{x}^{\bullet}(i)=\theta^*\right]=1,\ \ \forall\theta^*\in\mathcal{U} \tag{27}$$

or, in other words, if the estimate sequence converges a.s. to the true parameter value. The above definition of consistency is also called strong consistency. When the convergence is in probability, we get weak consistency. In this paper, we use the term consistency to mean strong consistency, which implies weak consistency.

Definition 2 (Asymptotic Unbiasedness) :

A sequence of estimates $\{\mathbf{x}^{\bullet}(i)\}_{i\geq 0}$ is called asymptotically unbiased if

$$\lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\mathbf{x}^{\bullet}(i)\right]=\theta^*,\ \ \ \forall\theta^*\in\mathcal{U} \tag{28}$$

The main result of this subsection concerns the consistency and asymptotic unbiasedness of the $\mathcal{LU}$ algorithm. Before proceeding further we state the following result.

Lemma 3 Consider the $\mathcal{LU}$ algorithm under Assumptions **A.1-4**. Then, the matrix $\left[b\overline{L}\otimes I_M+D_{\overline{H}}\right]$ is symmetric positive definite.

Proof: Symmetricity is obvious. It also follows from the properties of Laplacian matrices and the structure of $D_{\overline{H}}$ that these matrices are positive semidefinite. Then the matrix $\left[b\overline{L}\otimes I_M+D_{\overline{H}}\right]$ is positive semidefinite, being the sum of two positive semidefinite matrices. To prove positive definiteness, assume, on the contrary, that the matrix $\left[b\overline{L}\otimes I_M+D_{\overline{H}}\right]$ is not positive definite. Then, there exists, $\mathbf{x}\in\mathbb{R}^{NM\times 1}$, such that $\mathbf{x}\neq\mathbf{0}$ and

$$\mathbf{x}^T\left[b\overline{L}\otimes I_M+D_{\overline{H}}\right]\mathbf{x}=\mathbf{0} \tag{29}$$

From the positive semidefiniteness of $\overline{L}\otimes I_M$ and $D_{\overline{H}}$, and the fact that $b>0$, it follows

$$\mathbf{x}^T\left[\overline{L}\otimes I_M\right]\mathbf{x}=0,\ \ \mathbf{x}^TD_{\overline{H}}\mathbf{x}=0 \tag{30}$$

Write $\mathbf{x}$ in the partitioned form,

$$\mathbf{x}=\left[\mathbf{x}_1^T\cdots\mathbf{x}_N^T\right]^T,\ \ \mathbf{x}_n\in\mathbb{R}^{M\times 1},\ \forall 1\leq n\leq N \tag{31}$$

It follows from the properties of Laplacian matrices and the fact that $\lambda_2(\overline{L})>0$, that eqn. (30) holds *iff*

$$\mathbf{x}_n=\mathbf{a},\ \forall n \tag{32}$$

where $\mathbf{a}\in\mathbb{R}^{M\times 1}$, and $\mathbf{a}\neq\mathbf{0}$. Also, eqn. (30) implies

$$\sum_{n=1}^{N}\mathbf{x}_n^T\overline{H}_n^T\overline{H}_n\mathbf{x}_n=0 \tag{33}$$

This together with eqn. (32) implies

$$\mathbf{a}^T G \mathbf{a} = 0 \tag{34}$$

where G is defined in eqn. (23). This is clearly a contradiction, because, G is positive definite by Assumption **A.2** and $\mathbf{a} \neq \mathbf{0}$. Thus, we conclude that the matrix $\left[b\overline{L} \otimes I_M + D_{\overline{H}}\right]$ is positive definite. ∎

We now present the following result regarding the asymptotic unbiasedness of the estimate sequence.

Theorem 4 ($\mathcal{LU}$: Asymptotic unbiasedness) Consider the $\mathcal{LU}$ algorithm under Assumptions **A.1-4** and let $\{\mathbf{x}(i)\}_{i \geq 0}$ be the state sequence generated. Then we have

$$\lim_{i \to \infty} \mathbb{E}\left[\mathbf{x}_n(i)\right] = \theta^*, \;\; 1 \leq n \leq N \tag{35}$$

In other words, the estimate sequence, $\{\mathbf{x}_n(i)\}_{i \geq 0}$, generated at a sensor n is asymptotically unbiased.

Proof: Taking expectations on both sides of eqn. (19) and using the independence assumptions (Assumption **A.4**), we have

$$\mathbb{E}\left[\mathbf{x}(i+1)\right] = \mathbb{E}\left[\mathbf{x}(i)\right] - \alpha(i)\left[b\left(\overline{L} \otimes I_M\right)\mathbb{E}\left[\mathbf{x}(i)\right] + D_{\overline{H}}\mathbb{E}\left[\mathbf{x}(i)\right] - \overline{D}_{\overline{H}}\mathbb{E}\left[\mathbf{z}(i)\right]\right] \tag{36}$$

Subtracting $\mathbf{1}_N \otimes \theta^*$ from both sides of eqn. (36) and noting that

$$\left(\overline{L} \otimes I_M\right)\left(\mathbf{1}_N \otimes \theta^*\right) = \mathbf{0}, \;\; \overline{D}_{\overline{H}}\mathbb{E}\left[\mathbf{z}(i)\right] = D_{\overline{H}}\left(\mathbf{1}_N \otimes \theta^*\right) \tag{37}$$

we have

$$\mathbb{E}\left[\mathbf{x}(i+1)\right] - \mathbf{1}_N \otimes \theta^* = \left[I_{NM} - \alpha(i)\left(b\overline{L} \otimes I_M + D_{\overline{H}}\right)\right]\left[\mathbb{E}\left[\mathbf{x}(i)\right] - \mathbf{1}_N \otimes \theta^*\right] \tag{38}$$

Define, $\lambda_{\min}\left(b\overline{L} \otimes I_M + D_{\overline{H}}\right)$ and $\lambda_{\max}\left(b\overline{L} \otimes I_M + D_{\overline{H}}\right)$ to be the smallest and largest eigenvalues of the positive definite matrix $\left[b\overline{L} \otimes I_M + D_{\overline{H}}\right]$ (see Lemma 3.) Since, $\alpha(i) \to 0$ (Assumption **A.3**), there exists i_0, such that,

$$\alpha(i_0) \leq \frac{1}{\lambda_{\max}\left(b\overline{L} \otimes I_M + D_{\overline{H}}\right)}, \;\; \forall i \geq i_0 \tag{39}$$

Continuing the recursion in eqn. (38), we have, for $i > i_0$,

$$\mathbb{E}\left[\mathbf{x}(i)\right] - \mathbf{1}_N \otimes \theta^* = \left(\prod_{j=i_0}^{i-1}\left[I_{NM} - \alpha(j)\left(b\overline{L} \otimes I_M + D_{\overline{H}}\right)\right]\right)\left[\mathbb{E}\left[\mathbf{x}(i_0)\right] - \mathbf{1}_N \otimes \theta^*\right] \tag{40}$$

Eqn. (40) implies

$$\left\|\mathbb{E}\left[\mathbf{x}(i)\right] - \mathbf{1}_N \otimes \theta^*\right\| \leq \left(\prod_{j=i_0}^{i-1}\left\|I_{NM} - \alpha(j)\left(b\overline{L} \otimes I_M + D_{\overline{H}}\right)\right\|\right)\left\|\mathbb{E}\left[\mathbf{x}(i_0)\right] - \mathbf{1}_N \otimes \theta^*\right\|, \;\; i > i_0 \tag{41}$$

It follows from eqn. (39)

$$\left\|I_{NM} - \alpha(j)\left(b\overline{L} \otimes I_M + D_{\overline{H}}\right)\right\| = 1 - \alpha(j)\lambda_{\min}\left(b\overline{L} \otimes I_M + D_{\overline{H}}\right), \;\; j \geq i_0 \tag{42}$$

Eqns. (41,42) now give

$$\|\mathbb{E}\left[\mathbf{x}(i)\right]-\mathbf{1}_N\otimes\theta^*\|\leq\left(\prod_{j=i_0}^{i-1}\left(1-\alpha(j)\lambda_{\min}\left(b\overline{L}\otimes I_M+D_{\overline{H}}\right)\right)\right)\|\mathbb{E}\left[\mathbf{x}(i_0)\right]-\mathbf{1}_N\otimes\theta^*\|\,,\quad i>i_0 \tag{43}$$

Using the inequality, $1-a\leq e^{-a}$, for $0\leq a\leq 1$, we finally get

$$\|\mathbb{E}\left[\mathbf{x}(i)\right]-\mathbf{1}_N\otimes\theta^*\|\leq e^{-\lambda_{\min}\left(b\overline{L}\otimes I_M+D_{\overline{H}}\right)\sum_{j=i_0}^{i-1}\alpha(j)}\|\mathbb{E}\left[\mathbf{x}\left(i_0\right)\right]-\mathbf{1}_N\otimes\theta^*\|\,,\quad i>i_0 \tag{44}$$

Since, $\lambda_{\min}\left(b\overline{L}\otimes I_M+D_{\overline{H}}\right)>0$ and the weight sequence sums to infinity, we have

$$\lim_{i\to\infty}\|\mathbb{E}\left[\mathbf{x}(i)\right]-\mathbf{1}_N\otimes\theta^*\|=0 \tag{45}$$

and the theorem follows. ■

We prove that, under the assumptions of the $\mathcal{LU}$ algorithm (see Subsection II-A), the state sequence, $\{\mathbf{x}(i)\}_{i\geq 0}$, satisfies

$$\mathbb{P}\left[\lim_{i\to\infty}\mathbf{x}_n(i)=\theta^*,\ \forall n\right]=1 \tag{46}$$

In other words, the sensor states reach consensus asymptotically and converge a.s. to the true parameter value, θ^*, thus yielding a consistent estimate at each sensor.

In the following, we present some classical results on stochastic approximation from [36] regarding the convergence properties of generic stochastic recursive procedures, which will be used to characterize the convergence properties (consistency, convergence rate) of the $\mathcal{LU}$ algorithm.

Theorem 5 Let $\left\{\mathbf{x}(i)\in\mathbb{R}^{l\times 1}\right\}_{i\geq 0}$ be a random vector sequence in $\mathbb{R}^{l\times 1}$, which evolves according to:

$$\mathbf{x}(i+1)=\mathbf{x}(i)+\alpha(i)\left[R(\mathbf{x}(i))+\Gamma\left(i+1,\mathbf{x}(i),\omega\right)\right] \tag{47}$$

where, $R(\cdot):\mathbb{R}^{l\times 1}\longmapsto\mathbb{R}^{l\times 1}$ is Borel measurable and $\{\Gamma(i,\mathbf{x},\omega)\}_{i\geq 0,\ \mathbf{x}\in\mathbb{R}^{l\times 1}}$ is a family of random vectors in $\mathbb{R}^{l\times 1}$, defined on some probability space $(\Omega,\mathcal{F},\mathcal{P})$, and $\omega\in\Omega$ is a canonical element of Ω. Consider the following sets of assumptions:

B.1): The function $\Gamma(i,\cdot,\cdot):\mathbb{R}^{l\times 1}\times\Omega\longrightarrow\mathbb{R}^{l\times 1}$ is $\mathcal{B}^l\otimes\mathcal{F}$ measurable[2] for every i.

B.2): There exists a filtration $\{\mathcal{F}_i\}_{i\geq 0}$ of $\mathcal{F}$, such that, for each i, the family of random vectors $\{\Gamma\left(i,\mathbf{x},\omega\right)\}_{\mathbf{x}\in\mathbb{R}^{l\times 1}}$ is $\mathcal{F}_i$ measurable, zero-mean and independent of $\mathcal{F}_{i-1}$.

(Note that, if Assumptions **B.1, B.2** are satisfied, the process, $\{\mathbf{x}(i)\}_{i\geq 0}$, is Markov.)

B.3): There exists a function $V\left(\mathbf{x}\right)\in\mathbb{C}_2$ with bounded second order partial derivatives and a point $\mathbf{x}^*\in\mathbb{R}^{l\times 1}$

[2] $\mathcal{B}^l$ denotes the Borel algebra of $\mathbb{R}^{l\times 1}$.

satisfying:

$$V\left(\mathbf{x}^{*}\right)=0,\;\; V\left(\mathbf{x}\right)>0,\, \mathbf{x}\neq\mathbf{x}^{*},\;\; \lim_{\|\mathbf{x}\|\rightarrow\infty}V\left(\mathbf{x}\right)=\infty \tag{48}$$

$$\sup_{\epsilon<\|\mathbf{x}-\mathbf{x}^{*}\|<\frac{1}{\epsilon}}\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right)<0,\,\forall\epsilon>0 \tag{49}$$

B.4): There exist constants $k_1, k_2 > 0$, such that,

$$\|R\left(\mathbf{x}\right)\|^{2}+\mathbb{E}\left[\|\Gamma\left(i+1,\mathbf{x},\omega\right)\|^{2}\right]\leq k_{1}\left(1+V\left(\mathbf{x}\right)\right)-k_{2}\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right) \tag{50}$$

B.5): The weight sequence $\{\alpha(i)\}_{i\geq 0}$ satisfies

$$\alpha(i)>0,\,\sum_{i\geq 0}\alpha_{i}=\infty,\,\sum_{i\geq 0}\alpha^{2}(i)<\infty \tag{51}$$

C.1): The function $R\left(\mathbf{x}\right)$ admits the representation

$$R\left(\mathbf{x}\right)=B\left(\mathbf{x}-\mathbf{x}^{*}\right)+\delta\left(\mathbf{x}\right) \tag{52}$$

where

$$\lim_{\mathbf{x}\rightarrow\mathbf{x}^{*}}\frac{\|\delta\left(\mathbf{x}\right)\|}{\|\mathbf{x}-\mathbf{x}^{*}\|}=0 \tag{53}$$

(Note, in particular, if $\delta\left(\mathbf{x}\right)\equiv 0$, then eqn. (53) is satisfied.)

C.2): The weight sequence, $\{\alpha(i)\}_{i\geq 0}$ is of the form,

$$\alpha(i)=\frac{a}{i+1},\quad\forall i\geq 0 \tag{54}$$

where $a>0$ is a constant. (Note that **C.2** implies **B.5**.)

C.3): The matrix Σ, given by

$$\Sigma=aB+\frac{1}{2}I \tag{55}$$

is stable. Here I is the $l\times l$ identity matrix and a, B are given in eqns. (54,52), respectively.

C.4): The entries of the matrices

$$A\left(i,\mathbf{x}\right)=\mathbb{E}\left[\Gamma\left(i+1,\mathbf{x},\omega\right)\Gamma^{T}\left(i+1,\mathbf{x},\omega\right)\right],\;\;\forall i\geq 0,\;\; x\in\mathbf{R}^{l\times 1} \tag{56}$$

are finite and the following limit exists:

$$\lim_{i\rightarrow\infty,\;\mathbf{x}\rightarrow\mathbf{x}^{*}}A\left(i,\mathbf{x}\right)=S_{0} \tag{57}$$

C.5): There exists $\epsilon>0$, such that

$$\lim_{R\rightarrow\infty}\sup_{\|\mathbf{x}-\mathbf{x}^{*}\|<\epsilon}\sup_{i\geq 0}\int_{\|\Gamma(i+1,\mathbf{x},\omega)\|>R}\|\Gamma\left(i+1,\mathbf{x},\omega\right)\|^{2}dP=0 \tag{58}$$

Then we have the following:

Let the Assumptions **B.1-B.5** hold for the process, $\{\mathbf{x}(i)\}_{i\geq 0}$, given by eqn. (47). Then, starting from an arbitrary initial state, the Markov process, $\{\mathbf{x}(i)\}_{i\geq 0}$, converges a.s. to $\mathbf{x}^*$. In other words,

$$\mathbb{P}\left[\lim_{i\to\infty}\mathbf{x}(i)=\mathbf{x}^*\right]=1 \tag{59}$$

The normalized process, $\left\{\sqrt{i}\left(\mathbf{x}(i)-\mathbf{x}^*\right)\right\}_{i\geq 0}$, is asymptotically normal if, in addition to Assumptions **B.1-B.5**, Assumptions **C.1-C.5** are also satisfied. In particular, as $i\to\infty$

$$\sqrt{i}\left(\mathbf{x}(i)-\mathbf{x}^*\right)\Longrightarrow\mathcal{N}(\mathbf{0},S) \tag{60}$$

where $\Longrightarrow$ denotes convergence in distribution or weak convergence. Also, the asymptotic variance, S, in eqn. (60) is given by,

$$S=a^2\int_0^\infty e^{\Sigma v}S_0e^{\Sigma^T v}dv \tag{61}$$

Proof: For a proof see [36] (c.f. Theorems 4.4.4, 6.6.1). ■

In the sequel, we will use Theorem 5 to establish the consistency and asymptotic normality of the $\mathcal{LU}$ algorithm.

We now give the main result regarding the a.s. convergence of the iterate sequence.

Theorem 6 ($\mathcal{LU}$: Consistency) Consider the $\mathcal{LU}$ algorithm with the assumptions stated in Subsection II-A. Then,

$$\mathbb{P}\left[\lim_{i\to\infty}\mathbf{x}_n(i)=\theta^*,\ \forall n\right]=1 \tag{62}$$

In other words, the estimate sequence $\{\mathbf{x}_n(i)\}_{i\geq 0}$ at a sensor n, is a consistent estimate of the parameter θ.

Proof: The proof follows by showing that the process $\{\mathbf{x}(i)\}_{i\geq 0}$, generated by the $\mathcal{LU}$ algorithm, satisfies the Assumptions **B.1-B.5** of Theorem 5. Recall the filtration, $\{\mathcal{F}_i^{\mathbf{x}}\}_{i\geq 0}$, given in eqn. (26). By adding and subtracting the vector $\mathbf{1}_N\otimes\theta^*$ and noting that

$$\left(\overline{L}\otimes I_M\right)\left(\mathbf{1}_N\otimes\theta^*\right)=\mathbf{0} \tag{63}$$

eqn. (19) can be written as

$$\begin{aligned}\mathbf{x}(i+1) \quad = \quad & \mathbf{x}(i)-\alpha(i)\left[b\left(\overline{L}\otimes I_M\right)\left(\mathbf{x}(i)-\mathbf{1}_N\otimes\theta^*\right)+b\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{x}(i)+D_{\overline{H}}\left(\mathbf{x}(i)-\mathbf{1}_N\otimes\theta^*\right)\right.\\ & \left.-\overline{D}_{\overline{H}}\left(\mathbf{z}(i)-\overline{D}_{\overline{H}}^T\mathbf{1}_N\otimes\theta^*\right)+b\mathbf{\Upsilon}(i)+b\mathbf{\Psi}(i)\right]\end{aligned} \tag{64}$$

In the notation of Theorem 5, eqn. (64) can be written as

$$\mathbf{x}(i+1)=\mathbf{x}(i)+\alpha(i)\left[R(\mathbf{x}(i))+\Gamma\left(i+1,\mathbf{x}(i),\omega\right)\right] \tag{65}$$

where

$$R\left(\mathbf{x}\right) \quad = \quad -\left[b\overline{L}\otimes I_M+D_H\right]\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right) \tag{66}$$

$$\Gamma\left(i+1,\mathbf{x},\omega\right) \quad = \quad -\left[b\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{x}-\overline{D}_{\overline{H}}\left(\mathbf{z}(i)-\overline{D}_{\overline{H}}^T\mathbf{1}_N\otimes\theta^*\right)+b\mathbf{\Upsilon}(i)+b\mathbf{\Psi}(i)\right] \tag{67}$$

Under the Assumptions **A.1-A.4**, for fixed $i+1$, the random family, $\{\Gamma(i+1,\mathbf{x},\omega)\}_{\mathbf{x}\in\mathbb{R}^{NM\times 1}}$, is $\mathcal{F}_{i+1}^{\mathbf{x}}$ measurable, zero-mean and independent of $\mathcal{F}_{i}^{\mathbf{x}}$. Hence, the assumptions **B.1, B.2** of Theorem 5 are satisfied.

We now show the existence of a stochastic potential function $V(\cdot)$ satisfying the remaining Assumptions **B.3-B.4** of Theorem 5. To this end, define

$$V(\mathbf{x}) = (\mathbf{x}-\mathbf{1}_N\otimes\theta^*)^T\left[b\overline{L}\otimes I_M + D_{\overline{H}}\right](\mathbf{x}-\mathbf{1}_N\otimes\theta^*) \tag{68}$$

Clearly, $V(\mathbf{x})\in\mathbb{C}_2$ with bounded second order partial derivatives. It follows from the positive definiteness of $\left[b\overline{L}\otimes I_M + D_{\overline{H}}\right]$ (Lemma 3), that

$$V(\mathbf{1}_N\otimes\theta^*) = 0,\quad V(\mathbf{x})>0,\ \mathbf{x}\neq\mathbf{1}_N\otimes\theta^* \tag{69}$$

Since the matrix $\left[b\overline{L}\otimes I_M + D_{\overline{H}}\right]$ is positive definite, the matrix $\left[b\overline{L}\otimes I_M + D_{\overline{H}}\right]^2$ is also positive definite and hence, there exists a constant $c_1>0$, such that

$$(\mathbf{x}-\mathbf{1}_N\otimes\theta^*)^T\left[b\overline{L}\otimes I_M + D_{\overline{H}}\right]^2(\mathbf{x}-\mathbf{1}_N\otimes\theta^*) \geq c_1\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|^2,\ \forall\mathbf{x}\in\mathbb{R}^{NM\times 1} \tag{70}$$

It then follows that

$$\begin{aligned}\sup_{\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|>\epsilon}(R(\mathbf{x}),V_{\mathbf{x}}(\mathbf{x})) &= -2\inf_{\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|>\epsilon}(\mathbf{x}-\mathbf{1}_N\otimes\theta^*)^T\left[b\overline{L}\otimes I_M + D_{\overline{H}}\right]^2(\mathbf{x}-\mathbf{1}_N\otimes\theta^*)\\ &\leq -2\inf_{\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|>\epsilon} c_1\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|^2\\ &\leq -2c_1\epsilon^2\\ &< 0\end{aligned} \tag{71}$$

Thus, Assumption **B.3** is satisfied. From eqn. (66)

$$\begin{aligned}\|R(\mathbf{x})\|^2 &= (\mathbf{x}-\mathbf{1}_N\otimes\theta^*)^T\left[b\overline{L}\otimes I_M + D_{\overline{H}}\right]^2(\mathbf{x}-\mathbf{1}_N\otimes\theta^*)\\ &= -\frac{1}{2}(R(\mathbf{x}),V_{\mathbf{x}}(\mathbf{x}))\end{aligned} \tag{72}$$

From eqn. (67) and the independence assumptions (Assumption **A.4**)

$$\begin{aligned}\mathbb{E}\left[\|\Gamma(i+1,\mathbf{x},\omega)\|^2\right] &= \mathbb{E}\left[(\mathbf{x}-\mathbf{1}_N\otimes\theta^*)^T\left(b\widetilde{L}(i)\otimes I_M\right)^2(\mathbf{x}-\mathbf{1}_N\otimes\theta^*)\right]\\ &+\mathbb{E}\left[\left\|\overline{D}_{\overline{H}}\left(\mathbf{z}(i)-\overline{D}_{\overline{H}}^T\mathbf{1}_N\otimes\theta^*\right)\right\|^2\right]+b^2\mathbb{E}\left[\|\boldsymbol{\Upsilon}(i)+\boldsymbol{\Psi}(i)\|^2\right]\end{aligned}$$

Since the random matrix $\widetilde{L}(i)$ takes values in a finite set, there exists a constant $c_2>0$, such that

$$(\mathbf{x}-\mathbf{1}_N\otimes\theta^*)^T\left(b\widetilde{L}(i)\otimes I_M\right)^2(\mathbf{x}-\mathbf{1}_N\otimes\theta^*) \leq c_2\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|^2\quad\forall\mathbf{x}\in\mathbb{R}^{NM\times 1} \tag{73}$$

Again, since $\left(b\overline{L}\otimes I_M+D_{\overline{H}}\right)$ is positive definite, there exists a constant $c_3>0$, such that

$$\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)^T\left[b\overline{L}\otimes I_M+D_{\overline{H}}\right]\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)\geq c_3\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|^2\quad\forall\mathbf{x}\in\mathbb{R}^{NM\times 1}\tag{74}$$

We then have from eqns. (73,74)

$$\begin{aligned}\mathbb{E}\left[\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)^T\left(b\widetilde{L}(i)\otimes I_M\right)^2\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)\right] &\leq \frac{c_2}{c_3}\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)^T\left[b\overline{L}\otimes I_M+D_{\overline{H}}\right]\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)\\ &= c_4V\left(\mathbf{x}\right)\end{aligned}\tag{75}$$

for some constant $c_4=\frac{c_2}{c_3}>0$. The term $\mathbb{E}\left[\left\|\overline{D}_{\overline{H}}\mathbf{z}(i)-D_{\overline{H}}\mathbf{1}_N\otimes\theta^*\right\|^2\right]+b^2\mathbb{E}\left[\left\|\mathbf{\Upsilon}(i)+\mathbf{\Psi}(i)\right\|^2\right]$ is bounded by a finite constant $c_5>0$, as it follows from Assumptions **A.1-A.4**. We then have from eqns. (72,73)

$$\begin{aligned}\|R\left(\mathbf{x}\right)\|^2+\mathbb{E}\left[\left\|\Gamma\left(i+1,\mathbf{x},\omega\right)\right\|^2\right] &\leq -\frac{1}{2}\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right)+c_4V\left(\mathbf{x}\right)+c_5\\ &\leq c_6\left(1+V\left(\mathbf{x}\right)\right)-\frac{1}{2}\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right)\end{aligned}\tag{76}$$

where $c_6=\max\left(c_4,c_5\right)>0$. This verifies Assumption **B.4** of Theorem 5. Also, Assumption **B.5** is satisfied by the choice of $\{\alpha(i)\}_{i\geq 0}$ (Assumption **A.3**.) It then follows that the process $\{\mathbf{x}(i)\}_{i\geq 0}$ converges a.s. to $\mathbf{1}_N\otimes\theta^*$. In other words,

$$\mathbb{P}[\lim_{i\to\infty}\mathbf{x}_n(i)=\theta^*,\ \forall n]=1\tag{77}$$

which establishes the consistency of the $\mathcal{LU}$ algorithm. ■

C. Asymptotic Variance: $\mathcal{LU}$

In this subsection, we carry out a convergence rate analysis of the $\mathcal{LU}$ algorithm by studying its moderate deviation characteristics. We summarize here some definitions and terminology from the statistical literature, used to characterize the performance of sequential estimation procedures (see [35]).

Definition 7 (Asymptotic Normality) A sequence of estimates $\{\mathbf{x}^\bullet(i)\}_{i\geq 0}$ is asymptotically normal if for every $\theta^*\in\mathcal{U}$, there exists a positive semidefinite matrix $S(\theta^*)\in\mathbb{R}^{M\times M}$, such that,

$$\lim_{i\to\infty}\sqrt{i}\left(\mathbf{x}^\bullet(i)-\theta^*\right)\Longrightarrow\mathcal{N}\left(\mathbf{0}_M,S(\theta^*)\right)\tag{78}$$

The matrix $S(\theta^*)$ is called the asymptotic variance of the estimate sequence $\{\mathbf{x}^\bullet(i)\}_{i\geq 0}$.

In the following we prove the asymptotic normality of the $\mathcal{LU}$ algorithm and explicitly characterize the resulting asymptotic variance. To this end, define

$$S_H=\mathbb{E}\left[\left(\overline{D}_{\overline{H}}\begin{bmatrix}\widetilde{H}_1(i) & & \\ \ddots & \ddots & \ddots\\ & & \widetilde{H}_N(i)\end{bmatrix}\mathbf{1}_N\theta^*\right)\left(\overline{D}_{\overline{H}}\begin{bmatrix}\widetilde{H}_1(i) & & \\ \ddots & \ddots & \ddots\\ & & \widetilde{H}_N(i)\end{bmatrix}\mathbf{1}_N\theta^*\right)^T\right]\tag{79}$$

Let $\lambda_{\min}\left(b\overline{L}\otimes I_M + D_{\overline{H}}\right)$, be the smallest eigenvalue of $\left[b\overline{L}\otimes I_M + D_{\overline{H}}\right]$ and recall the definitions of S_ζ, S_q (eqns. (21,18)).

We now state the main result of this subsection, establishing the asymptotic normality of the $\mathcal{LU}$ algorithm.

Theorem 8 ($\mathcal{LU}$: Asymptotic normality and asymptotic efficiency) Consider the $\mathcal{LU}$ algorithm under **A.1-A.4** with link weight sequence, $\{\alpha(i)\}_{i\geq 0}$ given by:

$$\alpha(i) = \frac{a}{i+1},\ \forall i \tag{80}$$

for some constant $a > 0$. Let $\{\mathbf{x}(i)\}_{i\geq 0}$ be the state sequence generated. Then, if $a > \frac{1}{2\lambda_{\min}\left(b\overline{L}\otimes I_M + D_{\overline{H}}\right)}$, we have

$$\sqrt{(}i)\left(\mathbf{x}(i) - \mathbf{1}_N\otimes\theta^*\right) \Longrightarrow \mathcal{N}(\mathbf{0}, S(\theta^*)) \tag{81}$$

where

$$S(\theta^*) = a^2\int_0^\infty e^{\Sigma v}S_0 e^{\Sigma v}dv \tag{82}$$

$$\Sigma = -a\left[b\overline{L}\otimes I_M + D_{\overline{H}}\right] + \frac{1}{2}I \tag{83}$$

$$S_0 = S_H + \overline{D}_{\overline{H}}S_\zeta\overline{D}_{\overline{H}}^T + b^2 S_q \tag{84}$$

In particular, at any sensor n, the estimate sequence, $\{\mathbf{x}_n(i)\}_{i\geq 0}$ is asymptotically normal:

$$\sqrt{(}i)\left(\mathbf{x}_n(i) - \theta^*\right) \Longrightarrow \mathcal{N}(\mathbf{0}, S_{nn}(\theta^*)) \tag{85}$$

where, $S_{nn}(\theta^*) \in \mathbb{R}^{M\times M}$ denotes the n-th principal block of $S(\theta^*)$.

Proof: The proof involves a step-by-step verification of Assumptions **C.1-C.5** of Theorem 5, since the Assumptions **B.1-B.5** are already shown to be satisfied (see, Theorem 6.) We recall the definitions of $R(\mathbf{x})$ and $\Gamma(i+1, \mathbf{x}, \omega)$ from Theorem 6 (eqns. (66,67)) and reproduce here for convenience:

$$R(\mathbf{x}) = -\left[b\overline{L}\otimes I_M + D_H\right]\left(\mathbf{x} - \mathbf{1}_N\otimes\theta^*\right) \tag{86}$$

$$\Gamma(i+1, \mathbf{x}, \omega) = -\left[b\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{x} - \left(\overline{D}_H\mathbf{z}(i) - D_H\mathbf{1}_N\otimes\theta^*\right) + b\mathbf{\Upsilon}(i) + b\mathbf{\Psi}(i)\right] \tag{87}$$

From eqn. (86), Assumption **C.1** of Theorem 5 is satisfied with

$$B = -\left[b\overline{L}\otimes I_M + D_H\right] \tag{88}$$

and $\delta(\mathbf{x}) \equiv 0$. Assumption **C.2** is satisfied by hypothesis, while the condition $a > \frac{1}{2\lambda_{\min}\left(b\overline{L}\otimes I_M + D_{\overline{H}}\right)}$ implies

$$\Sigma = -a\left[b\overline{L}\otimes I_M + D_H\right] + \frac{1}{2}I_{NM} = aB + \frac{1}{2}I_{NM} \tag{89}$$

is stable, and hence Assumption **C.3**. To verify Assumption **C.4**, we have from Assumption **A.4**

$$\begin{aligned} A(i,\mathbf{x}) &= \mathbb{E}\left[\Gamma(i+1,\mathbf{x},\omega)\Gamma^T(i+1,\mathbf{x},\omega)\right] \\ &= b^2\mathbb{E}\left[\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{x}\mathbf{x}^T\left(\widetilde{L}(i)\otimes I_M\right)^T\right]+\mathbb{E}\left[\left(\overline{D}_{\overline{H}}\mathbf{z}(i)-D_{\overline{H}}\mathbf{1}_N\otimes\theta^*\right)\left(\overline{D}_{\overline{H}}\mathbf{z}(i)-D_{\overline{H}}\mathbf{1}_N\otimes\theta^*\right)^T\right] \\ &\quad +b^2\mathbb{E}\left[\left(\boldsymbol{\Upsilon}(i)+\boldsymbol{\Psi}(i)\right)\left(\boldsymbol{\Upsilon}(i)+\boldsymbol{\Psi}(i)\right)^T\right] \end{aligned} \tag{90}$$

From the i.i.d. assumptions, we note that all the three terms on the R.H.S. of eqn. (90) are independent of i, and, in particular, the last two terms are constants. For the first term, we note that

$$\lim_{\mathbf{x}\to\mathbf{1}_N\otimes\theta^*}\mathbb{E}\left[\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{x}\mathbf{x}^T\left(\widetilde{L}(i)\otimes I_M\right)^T\right]=\mathbf{0} \tag{91}$$

from the bounded convergence theorem, as the entries of $\left\{\widetilde{L}(i)\right\}_{i\geq 0}$ are bounded and

$$\left(\widetilde{L}(i)\otimes I_M\right)\left(\mathbf{1}_N\otimes\theta^*\right)=\mathbf{0} \tag{92}$$

For the second term on the R.H.S. of eqn. (90), we have

$$\begin{aligned} \mathbb{E}\left[\left(\overline{D}_{\overline{H}}\mathbf{z}(i)-D_{\overline{H}}\mathbf{1}_N\otimes\theta^*\right)\left(\overline{D}_{\overline{H}}\mathbf{z}(i)-D_{\overline{H}}\mathbf{1}_N\otimes\theta^*\right)^T\right] &= \mathbb{E}\left[\left(\overline{D}_{\overline{H}}\begin{bmatrix}\widetilde{H}_1(i) & & \\ \ddots & \ddots & \ddots \\ & & \widetilde{H}_N(i)\end{bmatrix}\mathbf{1}_N\theta^*\right)\right. \\ &\quad \left.\left(\overline{D}_{\overline{H}}\begin{bmatrix}\widetilde{H}_1(i) & & \\ \ddots & \ddots & \ddots \\ & & \widetilde{H}_N(i)\end{bmatrix}\mathbf{1}_N\theta^*\right)^T\right] + \mathbb{E}\left[\overline{D}_{\overline{H}}\zeta\zeta^T\overline{D}_{\overline{H}}^T\right] \\ &= S_H+\overline{D}_{\overline{H}}S_\zeta\overline{D}_{\overline{H}}^T \end{aligned} \tag{93}$$

where the last step follows from eqns. (79,21). Finally, we note the third term on the R.H.S. of eqn. (90) is b^2S_q (see eqn. (18).) We thus have from eqns. (90,91,93)

$$\begin{aligned} \lim_{i\to\infty,\ \mathbf{x}\to\mathbf{x}^*}A(i,\mathbf{x}) &= S_H+\overline{D}_{\overline{H}}S_\zeta\overline{D}_{\overline{H}}^T+b^2S_q \\ &= S_0 \end{aligned} \tag{94}$$

We now verify Assumption **C.5**. Consider a fixed $\epsilon>0$. We note that eqn. (58) is a restatement of the uniform integrability of the random family, $\left\{\|\Gamma(i+1,\mathbf{x},\omega)\|^2\right\}_{i\geq 0,\ \|\mathbf{x}-\theta^*\|<\epsilon}$. From eqn. (87) we have

$$\begin{aligned} \|\Gamma(i+1,\mathbf{x},\omega)\|^2 &= \left\|b\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{x}-\left(\overline{D}_H\mathbf{z}(i)-D_H\mathbf{1}_N\otimes\theta^*\right)+b\boldsymbol{\Upsilon}(i)+b\boldsymbol{\Psi}(i)\right\|^2 \\ &= \left\|b\left(\widetilde{L}(i)\otimes I_M\right)(\mathbf{x}-\theta^*)-\left(\overline{D}_H\mathbf{z}(i)-D_H\mathbf{1}_N\otimes\theta^*\right)+b\boldsymbol{\Upsilon}(i)+b\boldsymbol{\Psi}(i)\right\|^2 \qquad (95) \\ &\leq 9\left[\left\|\left(b\widetilde{L}(i)\otimes I_M\right)(\mathbf{x}-\theta^*)\right\|^2+\left\|\overline{D}_{\overline{H}}\mathbf{z}(i)-D_{\overline{H}}\mathbf{1}_N\otimes\theta^*\right\|^2+b^2\left\|\boldsymbol{\Upsilon}(i)+\boldsymbol{\Psi}(i)\right\|^2\right] \end{aligned}$$

where we used the inequality, $\|\mathbf{y}_1+\mathbf{y}_2+\mathbf{y}_3\|^2 \leq 9\left[\|\mathbf{y}_1\|^2+\|\mathbf{y}_2\|^2+\|\mathbf{y}_3\|^2\right]$, for vectors $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$. From eqn. (73) we note that, if $\|\mathbf{x}-\theta^*\| < \epsilon$,

$$\left\|\left(b\widetilde{L}(i)\otimes I_M\right)(\mathbf{x}-\theta^*)\right\|^2 \leq c_2\epsilon^2 \tag{96}$$

From (95), the family $\left\{\widetilde{\Gamma}\left(i+1,\mathbf{x},\omega\right)\right\}_{i\geq 0,\ \|\mathbf{x}-\theta^*\|<\epsilon}$ dominates the family $\left\{\|\Gamma\left(i+1,\mathbf{x},\omega\right)\|^2\right\}_{i\geq 0,\ \|\mathbf{x}-\theta^*\|<\epsilon}$, where

$$\widetilde{\Gamma}\left(i+1,\mathbf{x},\omega\right) = 9\left[c_2\epsilon^2 + \left\|\overline{D}_{\overline{H}}\mathbf{z}(i) - D_{\overline{H}}\mathbf{1}_N\otimes\theta^*\right\|^2 + b^2\left\|\mathbf{\Upsilon}(i)+\mathbf{\Psi}(i)\right\|^2\right] \tag{97}$$

It is clear that the family $\left\{\widetilde{\Gamma}\left(i+1,\mathbf{x},\omega\right)\right\}_{i\geq 0,\ \|\mathbf{x}-\theta^*\|<\epsilon}$ is i.i.d. and hence uniformly integrable (see [37]). Then the family $\left\{\|\Gamma\left(i+1,\mathbf{x},\omega\right)\|^2\right\}_{i\geq 0,\ \|\mathbf{x}-\theta^*\|<\epsilon}$ is also uniformly integrable since it is dominated by the uniformly integrable family $\left\{\widetilde{\Gamma}\left(i+1,\mathbf{x},\omega\right)\right\}_{i\geq 0,\ \|\mathbf{x}-\theta^*\|<\epsilon}$ (see [37]). Thus the Assumptions **C.1-C.5** are verified and the theorem follows. ■

D. An Example

From Theorem 8 and eqn. (79), we note that the asymptotic variance is independent of θ^*, if the observation matrices are non-random. In that case, it is possible to optimize (minimize) the asymptotic variance over the weights a and b. In the following, we study a special case permitting explicit computations and that leads to interesting results. Consider a scalar parameter $(M=1)$ and let each sensor n have the same i.i.d. observation model,

$$z_n(i) = h\theta^* + \zeta_n(i) \tag{98}$$

where $h\neq 0$ and $\{\zeta_n(i)\}_{i\geq 0,\ 1\leq n\leq N}$ is a family of independent zero mean Gaussian random variables with variance σ^2. In addition, assume unquantized inter-sensor exchanges. We define the average asymptotic variance per sensor attained by the algorithm $\mathcal{LU}$ as

$$S_{\mathcal{LU}} = \frac{1}{N}\mathrm{Tr}\left(S\right) \tag{99}$$

where S is given by eqn. (82) in Theorem 8. From Theorem 8 we have $S_0 = \sigma^2 h^2 I_N$ and hence from eqn. (82)

$$\begin{aligned} S_{\mathcal{LU}} &= \frac{a^2\sigma^2h^2}{N}\mathrm{Tr}\left(\int_0^\infty e^{2\Sigma v}dv\right) \\ &= \frac{a^2\sigma^2h^2}{N}\int_0^\infty \mathrm{Tr}\left(e^{2\Sigma v}\right)dv \end{aligned} \tag{100}$$

From eqn. (83) the eigenvalues of $2\Sigma v$ are $\left[-2ab\lambda_n(\overline{L}) - \left(2ah^2-1\right)\right]v$ for $1\leq n\leq N$ and we have

$$\begin{aligned} S_{\mathcal{LU}} &= \frac{a^2\sigma^2h^2}{N}\sum_{n=1}^{N}\int_0^\infty e^{\left[-2ab\lambda_n(\overline{L})-\left(2ah^2-1\right)\right]v}dv \\ &= \frac{a^2\sigma^2h^2}{N}\sum_{n=1}^{N}\frac{1}{2ab\lambda_n(\overline{L})+(2ah^2-1)} \\ &= \frac{a^2\sigma^2h^2}{N\left(2ah^2-1\right)} + \frac{a^2\sigma^2h^2}{N}\sum_{n=2}^{N}\frac{1}{2ab\lambda_n(\overline{L})+(2ah^2-1)} \end{aligned} \tag{101}$$

In this case, the constraint $a > \frac{1}{2\lambda_{\min}(b\overline{L}\otimes I_M + D_{\overline{H}})}$ in Theorem 8 reduces to $a > \frac{1}{2h^2}$, and hence the problem of optimum a, b design to minimize $S_{\mathcal{LU}}$ is given by

$$S_{\mathcal{LU}}^{*} = \inf_{a > \frac{1}{2h^2},\ b>0} S_{\mathcal{LU}} \tag{102}$$

It is to be noted, that the first term on the last step of eqn. (101) is minimized at $a = \frac{1}{h^2}$ and the second term (always non-negative under the constraint) goes to zero as $b \to \infty$ for any fixed $a > 0$. Hence, we have

$$S_{\mathcal{LU}}^{*} = \frac{\sigma^2}{Nh^2} \tag{103}$$

The above shows that by setting $a = \frac{1}{h^2}$ and b sufficiently large in the $\mathcal{LU}$ algorithm, one can make $S_{\mathcal{LU}}$ arbitrarily close to $S_{\mathcal{LU}}^{*}$.

We compare this optimum achievable asymptotic variance per sensor, $S_{\mathcal{LU}}^{*}$, attained by the distributed $\mathcal{LU}$ algorithm to that attained by a centralized scheme. In the centralized scheme, there is a central estimator, which receives measurements from all the sensors and computes an estimate based on all measurements. In this case, the sample mean estimator is an efficient estimator (in the sense of Cramér-Rao) and the estimate sequence $\{x_c(i)\}_{i\geq 0}$ is given by

$$x_c(i) = \frac{1}{Nih}\sum_{n,i} z_n(i) \tag{104}$$

and we have

$$\sqrt{i}\,(x_c(i) - \theta^{*}) \;\sim\; (0, \mathcal{S}_c) \tag{105}$$

where, S_c is the variance (which is also the one-step Fisher information in this case, see, [35]) and is given by

$$S_c = \frac{\sigma^2}{Nh^2} \tag{106}$$

From eqn. (103) we note that,

$$S_{\mathcal{LU}}^{*} = S_c \tag{107}$$

Thus the average asymptotic variance attainable by the distributed algorithm $\mathcal{LU}$ is the same as that of the optimum (in the sense of Cramér-Rao) centralized estimator having access to all information simultaneously. This is an interesting result, as it holds irrespective of the network topology. In particular, however sparse the inter-sensor communication graph is, the optimum achievable asymptotic variance is the same as that of the centralized efficient estimator. Note that weak convergence itself is a limiting result, and, hence, the rate of convergence in eqn. (81) in Theorem 8 will, in general, depend on the network topology.

III. Nonlinear Observation Models: Agorithm $\mathcal{NU}$

The previous section developed the algorithm $\mathcal{LU}$ for distributed parameter estimation when the observation model is linear. In this section, we extend the previous development to accommodate more general classes of nonlinear observation models. We comment briefly on the organization of this section. In Subsection III-A, we introduce notation and setup the problem, and in Subsection III-B we present the $\mathcal{NU}$ algorithm for distributed parameter estimation for nonlinear observation model and establish conditions for its consistency.

A. Problem Formulation-Nonlinear Case

We start by formally stating the observation and communication assumptions for the generic case.

D.1)Nonlinear Observation Model: Similar to Section II, let $\theta^{*} \in \mathcal{U} \subset \mathbb{R}^{M \times 1}$ be the true but unknown parameter value. In the general case, we assume that the observation model at each sensor n consists of an i.i.d. sequence $\{\mathbf{z}_n(i)\}_{i \geq 0}$ in $\mathbb{R}^{M_N \times 1}$ with

$$\mathbb{P}_{\theta^{*}}[\mathbf{z}_n(i) \in \mathcal{D}] = \int_{\mathcal{D}} dF_{\theta^{*}}, \quad \forall\ \mathcal{D} \in \mathbb{B}^{M_N \times 1} \tag{108}$$

where $F_{\theta^{*}}$ denotes the distribution function of the random vector $\mathbf{z}_n(i)$. We assume that the distributed observation model is *separably estimable*, a notion which we introduce now.

Definition 9 (Separably Estimable) Let $\{\mathbf{z}_n(i)\}_{i \geq 0}$ be the i.i.d. observation sequence at sensor n, where $1 \leq n \leq N$. We call the parameter estimation problem to be separably estimable, if there exist functions $g_n(\cdot) : \mathbb{R}^{M_N \times 1} \longmapsto \mathbb{R}^{M \times 1}$, $\forall 1 \leq n \leq N$, such that the function $h(\cdot) : \mathbb{R}^{M \times 1} \longmapsto \mathbb{R}^{M \times 1}$ given by

$$h(\theta) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}_{\theta}\left[g_n(\mathbf{z}_n(i))\right] \tag{109}$$

is invertible[3]

We will see that this condition is, in fact, necessary and sufficient to guarantee the existence of consistent distributed estimation procedures. This condition is a natural generalization of the observability constraint of Assumption **A.2** in the linear model. Indeed, if, assuming the linear model, we define $g_n(\theta) = \overline{H}_n^T \theta$, $\forall 1 \leq n \leq N$ in eqn. (109), we have $h(\theta) = G\theta$, where G is defined in eqn. (23). Then, invertibility of (109) is equivalent to Assumption **A.2**, i.e., to invertibility of G; hence, the linear model is an example of a separably estimable problem. Note that, if an observation model is separably estimable, then the choice of functions $g_n(\cdot)$ is not unique. Indeed, given a separably estimable model, it is important to figure out an appropriate decomposition, as in eqn. (109), because the convergence properties of the algorithms to be studied are intimately related to the behavior of these functions. At a particular iteration i, we do not require the observations across different sensors to be independent. In other words, we allow spatial correlation, but require temporal independence.

D.2)Random Link Failure, Quantized Communication. The random link failure model is the model given in Section I-B; similarly, we assume quantized inter-sensor communication with subtractive dithering.

D.3)Independence and Moment Assumptions. The sequences $\{L(i)\}_{i \geq 0}$,$\{\mathbf{z}_n(i)\}_{1 \leq n \leq N,\ i \geq 0}$,$\{\nu_{nl}^{m}(i)\}$ (the dither sequence, as in eqn. II-A) are mutually independent. Define the functions, $h_n(\cdot) : \mathbb{R}^{M \times 1} \longmapsto \mathbb{R}^{M \times 1}$, by

$$h_n(\theta) = \mathbb{E}_{\theta}\left[g_n(\mathbf{z}_n(i))\right], \ \forall 1 \leq n \leq N \tag{110}$$

[3]The factor $\frac{1}{N}$ in eqn. (109) is just for notational convenience, as will be seen later.

We make the assumption:

$$\mathbb{E}_{\theta}\left[\left\|\frac{1}{N}\sum_{n=1}^{N}g_n(\mathbf{z}_n(i))-h(\theta)\right\|^2\right]=\eta(\theta)<\infty,\ \forall\theta\in\mathcal{U} \tag{111}$$

In Subsection III-B and Section IV, we give two algorithms, $\mathcal{NU}$ and $\mathcal{NUI}$, respectively, for the distributed estimation problem **D1-D3** and provide conditions for consistency and other properties of the estimates.

B. Algorithm $\mathcal{NU}$

In this subsection, we present the algorithm $\mathcal{NU}$ for distributed parameter estimation in separably estimable models under Assumptions **D.1-D.3**.

Algorithm $\mathcal{NU}$. Each sensor n performs the following estimate update:

$$\mathbf{x}_n(i+1)=\mathbf{x}_n(i)-\alpha(i)\left[\sum_{l\in\Omega_n(i)}\beta\left(\mathbf{x}_n(i)-\mathbf{q}(\mathbf{x}_l(i)+\nu_{nl}(i))\right)+h_n(\mathbf{x}_n(i))-g_n(\mathbf{z}_n(i))\right] \tag{112}$$

based on $\mathbf{x}_n(i)$, $\{\mathbf{q}(\mathbf{x}_l(i)+\nu_{nl}(i))\}_{l\in\Omega_n(i)}$, and $\mathbf{z}_n(i)$, which are all available to it at time i. The sequence, $\left\{\mathbf{x}_n(i)\in\mathbb{R}^{M\times 1}\right\}_{i\geq 0}$, is the estimate (state) sequence generated at sensor n. The weight sequence $\{\alpha(i)\}_{i\geq 0}$ satisfies the persistence condition of Assumption **A.3** and $\beta>0$ is chosen to be an appropriate constant. Similar to eqn. (12) the above update can be written in compact form as

$$\mathbf{x}(i+1)=\mathbf{x}(i)-\alpha(i)\left[\beta(L(i)\otimes I_M)\mathbf{x}(i)+M(\mathbf{x}(i))-J(\mathbf{z}(i))+\mathbf{\Upsilon}(i)+\mathbf{\Psi}(i)\right] \tag{113}$$

where $\mathbf{\Upsilon}(i),\mathbf{\Psi}(i)$ are as in eqns. (13-16) and $\mathbf{x}(i)=[\mathbf{x}_1^T(i)\cdots\mathbf{x}_N^T(i)]^T$ is the vector of sensor states (estimates.) The functions $M(\mathbf{x}(i))$ and $J(\mathbf{z}(i))$ are given by

$$M(\mathbf{x}(i))=[h_1^T(\mathbf{x}_1(i))\cdots h_N^T(\mathbf{x}_N(i))]^T,\quad J(\mathbf{x}(i))=[g_1^T(\mathbf{z}_1(i))\cdots g_N^T(\mathbf{z}_N(i))]^T \tag{114}$$

We note that the update scheme in eqn. (113) is nonlinear and hence convergence properties can only be characterized, in general, through the existence of appropriate stochastic Lyapunov functions. In particular, if we can show that the iterative scheme in eqn. (113) falls under the purview of a general result like Theorem 5, we can establish properties like consistency, normality etc. To this end, we note, that eqn. (113) can be written as

$$\begin{aligned}\mathbf{x}(i+1)\ &=\ \mathbf{x}(i)-\alpha(i)\Big[\beta\left(\overline{L}\otimes I_M\right)(\mathbf{x}(i)-\mathbf{1}_N\otimes\theta^*)+\beta\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{x}(i)+\left(M(\mathbf{x}(i))-M(\mathbf{1}_N\otimes\theta^*)\right)\\ &\quad-\left(J(\mathbf{z}(i))-M(\mathbf{1}_N\otimes\theta^*)\right)+\mathbf{\Upsilon}(i)+\mathbf{\Psi}(i)\Big]\end{aligned} \tag{115}$$

which becomes in the notation of Theorem 5

$$\mathbf{x}(i+1)=\mathbf{x}(i)+\alpha(i)\left[R(\mathbf{x}(i))+\Gamma\left(i+1,\mathbf{x}(i),\omega\right)\right] \tag{116}$$

where

$$R\left(\mathbf{x}\right)=-\left[\beta\left(\overline{L}\otimes I_M\right)\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)+\left(M\left(\mathbf{x}\right)-M(\mathbf{1}_N\otimes\theta^*)\right)\right] \tag{117}$$

and

$$\Gamma\left(i+1,\mathbf{x},\omega\right)=-\left[\beta\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{x}-\left(J(\mathbf{z}(i))-M(\mathbf{1}_N\otimes\theta^*)\right)+\mathbf{\Upsilon}(i)+\mathbf{\Psi}(i)\right] \tag{118}$$

Consider the filtration, $\{\mathcal{F}_i\}_{i\geq 0}$,

$$\mathcal{F}_i=\sigma\left(\mathbf{x}(0),\left\{L(j),\{\mathbf{z}_n(j)\}_{1\leq N}\ ,\ \mathbf{\Upsilon}(j),\mathbf{\Psi}(j)\right\}_{0\leq j<i}\right) \tag{119}$$

Clearly, under Assumptions **D.1-D.3**, the state sequence, $\{\mathbf{x}(i)\}_{i\geq 0}$ generated by algorithm $\mathcal{NU}$ is Markov w.r.t. $\{\mathcal{F}_i\}_{i\geq 0}$, and the definition in eqn. (118) renders the random family, $\{\Gamma\left(i+1,\mathbf{x},\omega\right)\}_{\mathbf{x}\in\mathbb{R}^{NM\times 1}}$, $\mathcal{F}_{i+1}$ measurable, zero-mean, and independent of $\mathcal{F}_i$ for fixed $i+1$. Thus Assumptions **B.1, B.2** of Theorem 5 are satisfied, and we have the following immediately.

Proposition 10 ($\mathcal{NU}$: Consistency and asymptotic normality) Consider the state sequence $\{\mathbf{x}(i)\}_{i\geq 0}$ generated by the $\mathcal{NU}$ algorithm. Let $R\left(\mathbf{x}\right),\Gamma\left(i+1,\mathbf{x},\omega\right),\mathcal{F}_i$ be defined as in eqns. (117,118,119), respectively. Then, if there exists a function $V\left(\mathbf{x}\right)$ satisfying Assumptions **B.3, B.4** at $\mathbf{x}^*=\mathbf{1}_N\otimes\theta^*$, the estimate sequence $\{\mathbf{x}_n(i)\}_{i\geq 0}$ at any sensor n is consistent. In other words,

$$\mathbb{P}_{\theta^*}[\lim_{i\to\infty}\mathbf{x}_n(i)=\theta^*,\ \forall n]=1 \tag{120}$$

If, in addition, Assumptions **C.1-C.4** are satisfied, the estimate sequence $\{\mathbf{x}_n(i)\}_{i\geq 0}$ at any sensor n is asymptotically normal.

Proposition 10 states that, a.s. asymptotically, the network reaches consensus, and the estimates at each sensor converge to the true value of the parameter vector $\theta^\star$. The Proposition relates these convergence properties of $\mathcal{NU}$ to the existence of suitable Lyapunov functions. For a particular observation model characterized by the corresponding functions $h_n(\cdot),g_n(\cdot)$, if one can come up with an appropriate Lyapunov function satisfying the assumptions of Proposition 10, then consistency (asymptotic normality) is guaranteed. Existence of a suitable Lyapunov condition is sufficient for consistency, but may not be necessary. In particular, there may be observation models for which the $\mathcal{NU}$ algorithm is consistent, but there exists no Lyapunov function satisfying the assumptions of Proposition 10.[4] Also, even if a suitable Lyapunov function exists, it may be difficult to guess its form, because there is no systematic (constructive) way of coming up with Lyapunov functions for generic models.

However, for our problem of interest, some additional weak assumptions on the observation model, for example, Lipschitz continuity of the functions $h_n(\cdot)$, will guarantee the existence of suitable Lyapunov functions, thus establishing convergence properties of the $\mathcal{NU}$ algorithm. The rest of this subsection studies this issue and presents different sufficient conditions on the observation model, which guarantee that the assumptions of Proposition 10 are satisfied, leading to the a.s. convergence of the $\mathcal{NU}$ algorithm. We start with a definition.

[4]This is because converse theorems in stability theory do not always hold (see, [38].)

Definition 11 (Consensus Subspace) We define the consensus subspace, $\mathcal{C} \subset \mathbb{R}^{MN\times 1}$ as

$$\mathcal{C} = \left\{ \mathbf{y} \in \mathbb{R}^{NM\times 1} \,\middle|\, \mathbf{y} = \mathbf{1}_N \otimes \widetilde{\mathbf{y}},\, \widetilde{\mathbf{y}} \in \mathbb{R}^{M\times 1} \right\} \tag{121}$$

For $\mathbf{y} \in \mathbb{R}^{NM\times 1}$, we denote its component in $\mathcal{C}$ by $\mathbf{y}_{\mathcal{C}}$ and its orthogonal component by $\mathbf{y}_{\mathcal{C}}^{\perp}$.

Theorem 12 ($\mathcal{NU}$: Consistency under Lipschitz on h_n) Let $\{\mathbf{x}(i)\}_{i\geq 0}$ be the state sequence generated by the $\mathcal{NU}$ algorithm (Assumptions **D.1-D.3**.) Let the functions $h_n(\cdot), 1 \leq n \leq N$, be Lipschitz continuous with constants $k_n > 0,\ 1 \leq n \leq N$, respectively, i.e.,

$$\|h_n(\theta) - h_n(\widetilde{\theta})\| \leq k_n \|\theta - \widetilde{\theta}\|, \quad \forall\, \theta, \widetilde{\theta} \in \mathbb{R}^{M\times 1},\ 1 \leq n \leq N \tag{122}$$

and satisfy

$$\left(\theta - \widetilde{\theta}\right)^T \left(h_n(\theta) - h_n(\widetilde{\theta})\right) \geq 0, \quad \forall\, \theta \neq \widetilde{\theta} \in \mathbb{R}^{M\times 1},\ 1 \leq n \leq N \tag{123}$$

Define K as

$$K = \max(k_1, \cdots, k_N) \tag{124}$$

Then, for every $\beta > 0$, the estimate sequence is consistent. In other words,

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty} \mathbf{x}_n(i) = \theta^*,\ \forall n\right] = 1 \tag{125}$$

Before proceeding with the proof, we note that the conditions in eqns. (122,123) are much easier to verify than the general problem of guessing the form of the Lyapunov function. Also, as will be shown in the proof, the conditions in Theorem 12 determine a Lyapunov function explicitly, which may be used to analyze properties like convergence rate. The Lipschitz assumption is quite common in the stochastic approximation literature, while the assumption in eqn. (123) holds for a large class of functions. As a matter of fact, in the one-dimensional case ($M = 1$), it is satisfied if the functions $h_n(\cdot)$ are non-decreasing.

Proof: As noted earlier, the Assumptions **B.1, B.2** of Theorem 5 are always satisfied for the recursive scheme in eqn. (113.) To prove consistency, we need to verify Assumptions **B.3, B.4** only. To this end, consider the following Lyapunov function

$$V(\mathbf{x}) = \|\mathbf{x} - \mathbf{1}_N \otimes \theta^*\|^2 \tag{126}$$

Clearly,

$$V(\mathbf{1}_N \otimes \theta^*) = 0, \quad V(\mathbf{x}) > 0,\ \mathbf{x} \neq \mathbf{1}_N \otimes \theta^*, \quad \lim_{\|\mathbf{x}\|\to\infty} V(\mathbf{x}) = \infty \tag{127}$$

The assumptions in eqns. (122,123) imply that $h(\cdot)$ is Lipschitz continuous and

$$\left(\theta - \widetilde{\theta}\right)^T \left(h(\theta) - h(\widetilde{\theta})\right) > 0, \quad \forall\, \theta \neq \widetilde{\theta} \in \mathbb{R}^{M\times 1} \tag{128}$$

where eqn. (128) follows from the invertibility of $h(\cdot)$ and the fact that,

$$h\left(\theta\right)=\frac{1}{N}h_{n}\left(\theta\right),\quad\forall\ \theta\in\mathbb{R}^{M\times 1} \tag{129}$$

Recall the definitions of $R\left(\mathbf{x}\right),\Gamma\left(i+1,\mathbf{x},\omega\right)$ in eqns. (117,118) respectively. We then have

$$\begin{aligned}\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right) &= -2\beta\left(\mathbf{x}-\mathbf{1}_{N}\otimes\theta^{*}\right)^{T}\left(\overline{L}\otimes I_{M}\right)\left(\mathbf{x}-\mathbf{1}_{N}\otimes\theta^{*}\right)-2\left(\mathbf{x}-\mathbf{1}_{N}\otimes\theta^{*}\right)^{T}\left[M\left(\mathbf{x}\right)-M(\mathbf{1}_{N}\otimes\theta^{*})\right]\\ &= -2\beta\left(\mathbf{x}-\mathbf{1}_{N}\otimes\theta^{*}\right)^{T}\left(\overline{L}\otimes I_{M}\right)\left(\mathbf{x}-\mathbf{1}_{N}\otimes\theta^{*}\right)-2\sum_{n=1}^{N}\left[\left(\mathbf{x}_{n}-\theta^{*}\right)^{T}\left(h_{n}(\mathbf{x}_{n})-h_{n}(\theta^{*})\right)\right]\\ &\leq 0 \end{aligned} \tag{130}$$

where the last step follows from the positive-semidefiniteness of $\overline{L}\otimes I_{M}$ and eqn. (123). To verify Assumption **B.3**, we need to show

$$\sup_{\epsilon<\|\mathbf{x}-\mathbf{1}_{N}\theta^{*}\|<\frac{1}{\epsilon}}\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right)<0,\quad\forall\epsilon>0 \tag{131}$$

Let us assume on the contrary that eqn. (131) is not satisfied. Then from eqn. (130) we must have

$$\sup_{\epsilon<\|\mathbf{x}-\mathbf{1}_{N}\theta^{*}\|<\frac{1}{\epsilon}}\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right)=0,\quad\forall\epsilon>0 \tag{132}$$

Then, there exists a sequence, $\left\{\mathbf{x}^{k}\right\}_{k\geq 0}$ in $\left\{\mathbf{x}\in\mathbb{R}^{NM\times 1}\ \middle|\ \epsilon<\|\mathbf{x}-\mathbf{1}_{N}\theta^{*}\|<\frac{1}{\epsilon}\right\}$, such that

$$\lim_{k\to\infty}\left(R(\mathbf{x}^{k}),V_{\mathbf{x}}(\mathbf{x}^{k})\right)=0 \tag{133}$$

Since the set $\left\{\mathbf{x}\in\mathbb{R}^{NM\times 1}\mid\epsilon<\|\mathbf{x}-\mathbf{1}_{N}\theta^{*}\|<\frac{1}{\epsilon}\right\}$ is relatively compact, the sequence $\left\{\mathbf{x}^{k}\right\}_{k\geq 0}$ has a limit point, $\widehat{\mathbf{x}}$, such that, $\epsilon\leq\|\widetilde{\mathbf{x}}-\mathbf{1}_{N}\theta^{*}\|\leq\frac{1}{\epsilon}$, and from the continuity of $\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right)$, we must have

$$\left(R(\widetilde{\mathbf{x}}),V_{\mathbf{x}}(\widetilde{\mathbf{x}})\right)=0 \tag{134}$$

From eqns. (123,130), we then have

$$\left(\widetilde{\mathbf{x}}-\mathbf{1}_{N}\otimes\theta^{*}\right)^{T}\left(\overline{L}\otimes I_{M}\right)\left(\widetilde{\mathbf{x}}-\mathbf{1}_{N}\otimes\theta^{*}\right)=0,\ \ \left(\widetilde{\mathbf{x}}_{n}-\theta^{*}\right)^{T}\left(h_{n}(\widetilde{\mathbf{x}}_{n})-h_{n}(\theta^{*})\right)=0,\ \forall n \tag{135}$$

The first equality in eqn. (135) and the properties of the Laplacian imply that $\widetilde{\mathbf{x}}\in\mathcal{C}$ and hence there exists $\mathbf{a}\in\mathbb{R}^{M\times 1}$, such that,

$$\widetilde{\mathbf{x}}_{n}=\mathbf{a},\quad\forall n \tag{136}$$

The second set of inequalities in eqn. (135) then imply

$$\left(\mathbf{a}-\theta^{*}\right)^{T}\left(h(\mathbf{a})-h(\theta^{*})\right)=0 \tag{137}$$

which is a contradiction by eqn. (128) since $\mathbf{a}\neq\theta^{*}$. Thus, we have eqn. (131) that verifies Assumption **B.3**. Finally, we note that,

$$\begin{aligned}\|R(\mathbf{x})\|^2 &= \left\|\beta\left(\overline{L}\otimes I_M\right)(\mathbf{x}-\mathbf{1}_N\otimes\theta^*)+(M(\mathbf{x})-M(\mathbf{1}_N\otimes\theta^*))\right\|^2\\ &\leq 4\beta^2\left\|\left(\overline{L}\otimes I_M\right)(\mathbf{x}-\mathbf{1}_N\otimes\theta^*)\right\|^2+4\left\|M(\mathbf{x})-M(\mathbf{1}_N\otimes\theta^*)\right\|^2\\ &\leq 4\beta^2\lambda_N(\overline{L})\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|^2+4K^2\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|^2 \qquad (138)\end{aligned}$$

where the second step follows from the Lipschitz continuity of $h_n(\cdot)$ and K is defined in eqn. (124). To verify Assumption **B.4**, we have then along similar lines as in Theorem 6

$$\begin{aligned}\|R(\mathbf{x})\|^2+\mathbb{E}\left[\|\Gamma(i+1,\mathbf{x},\omega)\|^2\right] &\leq k_1(1+V(\mathbf{x}))\\ &\leq k_1(1+V(\mathbf{x}))-(R(\mathbf{x}),V_{\mathbf{x}}(\mathbf{x})) \qquad (139)\end{aligned}$$

for some constant $k_1>0$ (the last step follows from eqn. (130).) Hence, the required assumptions are satisfied and the claim follows.

■

It follows from the proof, that the Lipschitz continuity assumption in Theorem 12 can be replaced by continuity of the functions $h_n(\cdot),\ 1\leq n\leq N$, and linear growth conditions, i.e.,

$$\|h_n(\theta)\|^2\leq c_{n,1}+c_{n,2}\|\theta\|^2,\quad \forall\theta\in\mathbb{R}^{M\times 1},\ 1\leq n\leq N \qquad (140)$$

for constants $c_{n,1},c_{n,2}>0$.

We now present another set of sufficient conditions that guarantee consistency of the algorithm $\mathcal{NU}$. If the observation model is separably estimable, in some cases even if the underlying model is nonlinear, it may be possible to choose the functions, $g_n(\cdot)$, such that the function $h(\cdot)$ possesses nice properties. This is the subject of the next result.

Theorem 13 ($\mathcal{NU}$: Consistency under strict monotonicity on h) Consider the $\mathcal{NU}$ algorithm (Assumptions **D.1-D.3**.) Suppose that the functions $g_n(\cdot)$ can be chosen, such that the functions $h_n(\cdot)$ are Lipschitz continuous with constants $k_n>0$ and the function $h(\cdot)$ satisfies

$$\left(\theta-\widetilde{\theta}\right)^T\left(h(\theta)-h(\widetilde{\theta})\right)\geq\gamma\|\theta-\widetilde{\theta}\|^2,\ \forall\theta,\widetilde{\theta}\in\mathbb{R}^{M\times 1} \qquad (141)$$

for some constant $\gamma>0$. Then, if $\beta>\frac{K^2+K\gamma}{\gamma\lambda_2\overline{L}}$, the algorithm $\mathcal{NU}$ is consistent, i.e.,

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\mathbf{x}_n(i)=\theta^*,\ \forall n\right]=1 \qquad (142)$$

where, $K=\max(k_1,\cdots,k_N)$.

Before proceeding to the proof, we comment that, in comparison to Theorem 12, strengthening the assumptions on $h(\cdot)$, see eqn. (141), considerably weakens the assumptions on the functions $h_n(\cdot)$. Eqn. (141) is an analog of strict monotonicity. For example, if $h(\cdot)$ is linear, the left hand side of eqn. (141) becomes a quadratic and the condition says that this quadratic is strictly away from zero, i.e., monotonically increasing with rate γ.

Proof: As noted earlier, the Assumptions **B.1, B.2** of Theorem 5 are always satisfied by the recursive scheme in eqn. (113.) To prove consistency, we need to verify Assumptions **B.3, B.4** only. To this end, consider the following Lyapunov function

$$V\left(\mathbf{x}\right)=\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|^2 \tag{143}$$

Clearly,

$$V(\mathbf{1}_N\otimes\theta^*)=0,\quad V\left(\mathbf{x}\right)>0,\ \mathbf{x}\neq\mathbf{1}_N\otimes\theta^*,\quad \lim_{\|\mathbf{x}\|\rightarrow\infty}V\left(\mathbf{x}\right)=\infty \tag{144}$$

Recall the definitions of $R\left(\mathbf{x}\right),\Gamma\left(i+1,\mathbf{x},\omega\right)$ in eqns. (117,118), respectively, and the consensus subspace in eqn. (121). We then have

$$\begin{aligned}
\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right) &= -2\beta\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)^T\left(\overline{L}\otimes I_M\right)\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)-2\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)^T\left[M\left(\mathbf{x}\right)-M(\mathbf{1}_N\otimes\theta^*)\right]\\
&\leq -2\beta\lambda_2(\overline{L})\|\mathbf{x}_{\mathcal{C}^\perp}\|^2-2\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)^T\left[M\left(\mathbf{x}\right)-M(\mathbf{x}_{\mathcal{C}})\right]\\
&\qquad -2\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)^T\left[M(\mathbf{x}_{\mathcal{C}})-M(\mathbf{1}_N\otimes\theta^*)\right]\\
&\leq -2\beta\lambda_2(\overline{L})\|\mathbf{x}_{\mathcal{C}^\perp}\|^2+2\left\|\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)^T\left[M\left(\mathbf{x}\right)-M(\mathbf{x}_{\mathcal{C}})\right]\right\|\\
&\qquad -2\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)^T\left[M(\mathbf{x}_{\mathcal{C}})-M(\mathbf{1}_N\otimes\theta^*)\right]\\
&\leq -2\beta\lambda_2(\overline{L})\|\mathbf{x}_{\mathcal{C}^\perp}\|^2+2K\|\mathbf{x}_{\mathcal{C}^\perp}\|\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|\\
&\qquad -2\left(\mathbf{x}-\mathbf{1}_N\otimes\theta^*\right)^T\left[M(\mathbf{x}_{\mathcal{C}})-M(\mathbf{1}_N\otimes\theta^*)\right]\\
&= -2\beta\lambda_2(\overline{L})\|\mathbf{x}_{\mathcal{C}^\perp}\|^2+2K\|\mathbf{x}_{\mathcal{C}^\perp}\|\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|-2\mathbf{x}_{\mathcal{C}^\perp}^T\left[M(\mathbf{x}_{\mathcal{C}})-M(\mathbf{1}_N\otimes\theta^*)\right]\\
&\quad -2\left(\mathbf{x}_{\mathcal{C}}-\mathbf{1}_N\otimes\theta^*\right)^T\left[M(\mathbf{x}_{\mathcal{C}})-M(\mathbf{1}_N\otimes\theta^*)\right]\\
&\leq -2\beta\lambda_2(\overline{L})\|\mathbf{x}_{\mathcal{C}^\perp}\|^2+2K\|\mathbf{x}_{\mathcal{C}^\perp}\|\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|+2\left\|\mathbf{x}_{\mathcal{C}^\perp}^T\left[M(\mathbf{x}_{\mathcal{C}})-M(\mathbf{1}_N\otimes\theta^*)\right]\right\|\\
&\quad -2\left(\mathbf{x}_{\mathcal{C}}-\mathbf{1}_N\otimes\theta^*\right)^T\left[M(\mathbf{x}_{\mathcal{C}})-M(\mathbf{1}_N\otimes\theta^*)\right]\\
&\leq -2\beta\lambda_2(\overline{L})\|\mathbf{x}_{\mathcal{C}^\perp}\|^2+2K\|\mathbf{x}_{\mathcal{C}^\perp}\|\|\mathbf{x}-\mathbf{1}_N\otimes\theta^*\|+2K\|\mathbf{x}_{\mathcal{C}^\perp}\|\|\mathbf{x}_{\mathcal{C}}-\mathbf{1}_N\otimes\theta^*\|\\
&\qquad -2\gamma\left\|\mathbf{x}_{\mathcal{C}}-\mathbf{1}_N\otimes\theta^*\right\|^2\\
&= \left(-2\beta\lambda_2(\overline{L})+2K\right)\|\mathbf{x}_{\mathcal{C}^\perp}\|^2+4K\|\mathbf{x}_{\mathcal{C}^\perp}\|\|\mathbf{x}_{\mathcal{C}}-\mathbf{1}_N\otimes\theta^*\|-2\gamma\left\|\mathbf{x}_{\mathcal{C}}-\mathbf{1}_N\otimes\theta^*\right\|^2
\end{aligned} \tag{145}$$

where the second to last step is justified because $\mathbf{x}_{\mathcal{C}}=\mathbf{1}_N\otimes\widetilde{\mathbf{y}}$ for some $\widetilde{\mathbf{y}}\in\mathbb{R}^{M\times 1}$ and

$$\begin{aligned}\left(\mathbf{x}_{\mathcal{C}}-\mathbf{1}_N\otimes\theta^{*}\right)^{T}\left[M(\mathbf{x}_{\mathcal{C}})-M(\mathbf{1}_N\otimes\theta^{*})\right] &= \sum_{n=1}^{N}\left(\widetilde{\mathbf{y}}-\theta^{*}\right)^{T}\left[h_n(\widetilde{\mathbf{y}})-h_n(\theta^{*})\right]\\ &= \left(\widetilde{\mathbf{y}}-\theta^{*}\right)^{T}\sum_{n=1}^{N}\left[h_n(\widetilde{\mathbf{y}})-h_n(\theta^{*})\right]\\ &= N\left(\widetilde{\mathbf{y}}-\theta^{*}\right)^{T}\left[h(\widetilde{\mathbf{y}})-h(\theta^{*})\right]\\ &\geq N\gamma\left\|\widetilde{\mathbf{y}}-\theta^{*}\right\|^{2}\\ &= \gamma\left\|\mathbf{x}_{\mathcal{C}}-\mathbf{1}_N\otimes\theta^{*}\right\|^{2}\end{aligned}\tag{146}$$

It can be shown that, if $\beta>\frac{K^2+K\gamma}{\gamma\lambda_2\overline{L}}$, the term on the R.H.S. of eqn. (145) is always non-positive. We thus have

$$\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right)\leq 0,\quad\forall\mathbf{x}\in\mathbb{R}^{MN\times 1}\tag{147}$$

By the continuity of $\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right)$ and the relative compactness of $\left\{\mathbf{x}\in\mathbb{R}^{NM\times 1}\;\middle|\;\epsilon<\left\|\mathbf{x}-\mathbf{1}_N\theta^{*}\right\|<\frac{1}{\epsilon}\right\}$, we can show along similar lines as in Theorem 12 that

$$\sup_{\epsilon<\left\|\mathbf{x}-\mathbf{1}_N\theta^{*}\right\|<\frac{1}{\epsilon}}\left(R\left(\mathbf{x}\right),V_{\mathbf{x}}\left(\mathbf{x}\right)\right)<0,\quad\forall\epsilon>0\tag{148}$$

verifying Assumption **B.3**. Assumption **B.4** can be verified in an exactly similar manner as in Theorem 12 and the result follows. ■

IV. Nonlinear Observation Models: Algorithm $\mathcal{NLU}$

In this Section, we present the algorithm $\mathcal{NLU}$ for distributed estimation in separably estimable observation models. As will be explained later, this is a mixed time-scale algorithm, where the consensus time-scale dominates the observation update time-scale as time progresses. The $\mathcal{NLU}$ algorithm is based on the fact that, for separably estimable models, it suffices to know $h(\theta^{*})$, because θ^{*} can be unambiguously determined from the invertible function $h(\theta^{*})$. To be precise, if the function $h(\cdot)$ has a continuous inverse, then any iterative scheme converging to $h(\theta^{*})$ will lead to consistent estimates, obtained by inverting the sequence of iterates. The algorithm $\mathcal{NLU}$ is shown to yield consistent and unbiased estimators at each sensor for any separably observable model, under the assumption that the function $h(\cdot)$ has a continuous inverse. Thus, the algorithm $\mathcal{NLU}$ presents a more reliable alternative than the algorithm $\mathcal{NU}$, because, as shown in Subsection III-B, the convergence properties of the latter can be guaranteed only under certain assumptions on the observation model. We briefly comment on the organization of this section. The $\mathcal{NLU}$ algorithm for separably estimable observation models is presented in Subsection IV-A. Subsection IV-B offers interpretations of the $\mathcal{NLU}$ algorithm and presents the main results regarding consistency, mean-square convergence, asymptotic unbiasedness proved in the paper. In Subsection IV-C we prove the main results about the $\mathcal{NLU}$ algorithm and provide insights behind the analysis (in particular, why standard stochastic approximation results cannot be used directly to give its convergence properties.) Finally, Subsection V presents discussions on the $\mathcal{NLU}$ algorithm and suggests future research directions.

A. Algorithm $\mathcal{NLU}$

Algorithm $\mathcal{NLU}$: Let $\mathbf{x}(0) = [\mathbf{x}_1^T \cdots \mathbf{x}_N^T]^T$ be the initial set of states (estimates) at the sensors. The $\mathcal{NLU}$ generates the state sequence $\{\mathbf{x}_n(i)\}_{i\geq 0}$ at the n-th sensor according to the following distributed recursive scheme:

$$\mathbf{x}_n(i+1) = h^{-1}\left(h(\mathbf{x}_n(i)) - \beta(i)\left(\sum_{l\in\Omega_n(i)}\left(h(\mathbf{x}_n(i)) - \mathbf{q}\left(h(\mathbf{x}_l(i)) + \nu_{nl}(i)\right)\right)\right) - \alpha(i)\left(h(\mathbf{x}_n(i)) - g_n(\mathbf{z}_n(i))\right)\right) \tag{149}$$

based on the information, $\mathbf{x}_n(i), \{\mathbf{q}\left(h(\mathbf{x}_l(i)) + \nu_{nl}(i)\right)\}_{l\in\Omega_n(i)}, \mathbf{z}_n(i)$, available to it at time i (we assume that at time i sensor l sends a quantized version of $h(\mathbf{x}_l(i)) + \nu_{nl}(i)$ to sensor n.) Here $h^{-1}(\cdot)$ denotes the inverse of the function $h(\cdot)$ and $\{\beta(i)\}_{i\geq 0}, \{\alpha(i)\}_{i\geq 0}$ are appropriately chosen weight sequences. In the sequel, we analyze the $\mathcal{NLU}$ algorithm under the model Assumptions **D.1-D.3**, and in addition we assume:

D.4): There exists $\epsilon_1 > 0$, such that the following moment exists:

$$\mathbb{E}_\theta\left[\left\|J(\mathbf{z}(i)) - \frac{1}{N}\left(\mathbf{1}_N \otimes I_M\right)^T J(\mathbf{z}(i))\right\|^{2+\epsilon_1}\right] = \kappa(\theta) < \infty,\ \ \forall\theta\in\mathcal{U} \tag{150}$$

The above moment condition is stronger than the moment assumption required by the $\mathcal{NU}$ algorithm in eqn. (111), where only existence of the quadratic moment was assumed.

We also define

$$\mathbb{E}_\theta\left[\left\|J(\mathbf{z}(i)) - \frac{1}{N}\left(\mathbf{1}_N \otimes I_M\right)^T J(\mathbf{z}(i))\right\|\right] = \kappa_1(\theta) < \infty,\quad \forall\theta\in\mathcal{U} \tag{151}$$

$$\mathbb{E}_\theta\left[\left\|J(\mathbf{z}(i)) - \frac{1}{N}\left(\mathbf{1}_N \otimes I_M\right)^T J(\mathbf{z}(i))\right\|^{2}\right] = \kappa_2(\theta) < \infty,\quad \forall\theta\in\mathcal{U} \tag{152}$$

D.5): The weight sequences $\{\beta(i)\}_{i\geq 0}$,$\{\beta(i)\}_{i\geq 0}$ are given by

$$\alpha(i) = \frac{a}{(i+1)^{\tau_1}},\quad \beta(i) = \frac{b}{(i+1)^{\tau_2}} \tag{153}$$

where $a, b > 0$ are constants. We assume the following:

$$.5 < \tau_1, \tau_2 \leq 1,\quad \tau_1 > \frac{1}{2+\epsilon_1} + \tau_2,\quad 2\tau_2 > \tau_1 \tag{154}$$

We note that under Assumption **D.4** that $\epsilon_1 > 0$, such weight sequences always exist. As an example, if $\frac{1}{2+\epsilon_1} = .49$, then the choice $\tau_1 = 1$ and $\tau_2 = .505$ satisfies the inequalities in eqn. (154).

D.6): The function $h(\cdot)$ has a continuous inverse, denoted by $h^{-1}(\cdot)$ in the sequel.

To write the $\mathcal{NLU}$ in a more compact form, we introduce the *transformed* state sequence, $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$, where $\widetilde{\mathbf{x}}(i) = [\widetilde{\mathbf{x}}_1^T(i) \cdots \widetilde{\mathbf{x}}_N^T(i)]^T \in \mathbb{R}^{NM\times 1}$ and the iterations are given by

$$\widetilde{\mathbf{x}}(i+1) = \widetilde{\mathbf{x}}(i) - \beta(i)\left(L(i)\otimes I_M\right)\widetilde{\mathbf{x}}(i) - \alpha(i)\left[\widetilde{\mathbf{x}}(i) - J(\mathbf{z}(i))\right] - \beta(i)\left(\boldsymbol{\Upsilon}(i) + \boldsymbol{\Psi}(i)\right) \tag{155}$$

$$\mathbf{x}(i) = \left[\left(h^{-1}(\widetilde{\mathbf{x}}_1(i))\right)^T \cdots \left(h^{-1}(\widetilde{\mathbf{x}}_N(i))\right)^T\right]^T \tag{156}$$

Here $\mathbf{\Upsilon}(i), \mathbf{\Psi}(i)$ model the dithered quantization error effects as in algorithm $\mathcal{NU}$. The update model in eqn. (155) is a mixed time-scale procedure, where the consensus time-scale is determined by the weight sequence $\{\beta(i)\}_{i\geq 0}$. On the other hand, the observation update time-scale is governed by the weight sequence $\{\alpha(i)\}_{i\geq 0}$. It follows from Assumption **D.5** that $\tau_1 > \tau_2$, which in turn implies, $\frac{\beta(i)}{\alpha(i)} \to \infty$ as $i \to \infty$. Thus, the consensus time-scale dominates the observation update time-scale as the algorithm progresses making it a mixed time-scale algorithm that does not directly fall under the purview of stochastic approximation results like Theorem 5. Also, the presence of the random link failures and quantization noise (which operate at the same time-scale as the consensus update) precludes standard approaches like time-scale separation for the limiting system.

B. Algorithm $\mathcal{NLU}$: Discussions and Main Results

We comment on the $\mathcal{NLU}$ algorithm. As is clear from eqns. (155,156), the $\mathcal{NLU}$ algorithm operates in a *transformed* domain. As a matter of fact, the function $h(\cdot)$ (c.f. definition 9) can be viewed as an invertible transformation on the parameter space $\mathcal{U}$. The transformed state sequence, $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$, is then a transformation of the estimate sequence $\{\mathbf{x}(i)\}_{i\geq 0}$, and, as seen from eqn. (155), the evolution of the sequence $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$ is linear. This is an important feature of the $\mathcal{NLU}$ algorithm, which is linear in the transformed domain, although the underlying observation model is nonlinear. Intuitively, this approach can be thought of as a distributed stochastic version of homomorphic filtering (see [39]), where, by suitably transforming the state space, linear filtering is performed on a certain non-linear problem of filtering. In our case, for models of the separably estimable type, the function $h(\cdot)$ then plays the role of the analogous transformation in homomorphic filtering, and in this transformed space, one can design linear estimation algorithms with desirable properties. This makes the $\mathcal{NLU}$ algorithm significantly different from algorithm $\mathcal{NU}$, with the latter operating on the untransformed space and is non-linear. This linear property of the $\mathcal{NLU}$ algorithm in the transformed domain leads to nice statistical properties (for example, consistency asymptotic unbiasedness) under much weaker assumptions on the observation model as required by the nonlinear $\mathcal{NLU}$ algorithm.

We now state the main results about the $\mathcal{NLU}$ algorithm, to be developed in the paper. We show that, if the observation model is separably estimable, then, in the transformed domain, the $\mathcal{NLU}$ algorithm is consistent. More specifically, if θ^* is the true (but unknown) parameter value, then the transformed sequence $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$ converges a.s. and in mean-squared sense to $h(\theta^*)$. We note that, unlike the $\mathcal{NU}$ algorithm, this only requires the observation model to be separably estimable and no other conditions on the functions $h_n(\cdot), h(\cdot)$. We summarize these in the following theorem.

Theorem 14 Consider the $\mathcal{NLU}$ algorithm under the Assumptions **D.1-D.5**, and the sequence $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$ generated according to eqn. (155). We then have

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty} \widetilde{\mathbf{x}}_n(i) = h(\theta^*),\ \forall 1 \leq n \leq N\right] = 1 \tag{157}$$

$$\lim_{i\to\infty} \mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}_n(i) - h(\theta^*)\right\|^2\right] = 0,\ \forall 1 \leq n \leq N \tag{158}$$

In particular,

$$\lim_{i\to\infty} \mathbb{E}_{\theta^*}\left[\mathbf{x}_n(i)\right] = h(\theta^*), \quad \forall 1 \leq n \leq N \tag{159}$$

In other words, in the transformed domain, the estimate sequence $\{\widetilde{\mathbf{x}}_n(i)\}_{i\geq 0}$ at sensor n, is consistent, asymptotically unbiased and converges in mean-squared sense to $h(\theta^*)$.

As an immediate consequence of Theorem 14, we have the following result, which characterizes the statistical properties of the untransformed state sequence $\{\mathbf{x}(i)\}_{i\geq 0}$.

Theorem 15 Consider the $\mathcal{NLU}$ algorithm under the Assumptions **D.1-D.6**. Let $\{\mathbf{x}(i)\}_{i\geq 0}$ be the state sequence generated, as given by eqns. (155,156). We then have

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty} \mathbf{x}_n(i) = \theta^*, \ \forall\ 1 \leq n \leq N\right] = 1 \tag{160}$$

In other words, the $\mathcal{NLU}$ algorithm is consistent.

If in addition, the function $h^{-1}(\cdot)$ is Lipschitz continuous, the $\mathcal{NLU}$ algorithm is asymptotically unbiased, i.e.,

$$\lim_{i\to\infty} \mathbb{E}_{\theta^*}\left[\mathbf{x}_n(i)\right] = \theta^*, \ \forall\ 1 \leq n \leq N \tag{161}$$

The next subsection is concerned with the proofs of Theorems 14, 15.

C. Consistency and Asymptotic Unbiasedness of $\mathcal{NLU}$: Proofs of Theorems 14,15

The present subsection is devoted to proving the consistency and unbiasedness of the $\mathcal{NLU}$ algorithm under the stated Assumptions. The proof is lengthy and we start by explaining why standard stochastic approximation results like Theorem 5 do not apply directly. A careful inspection shows that there are essentially two different time-scales embedded in eqn. (155). The consensus time-scale is determined by the weight sequence $\{\beta(i)\}_{i\geq 0}$, whereas the observation update time-scale is governed by the weight sequence $\{\alpha(i)\}_{i\geq 0}$. It follows from Assumption **D.5** that $\tau_1 > \tau_2$, which, in turn, implies $\frac{\beta(i)}{\alpha(i)} \to \infty$ as $i \to \infty$. Thus, the consensus time-scale dominates the observation update time-scale as the algorithm progresses making it a mixed time-scale algorithm that does not directly fall under the purview of stochastic approximation results like Theorem 5. Also, the presence of the random link failures and quantization noise (which operate at the same time-scale as the consensus update) precludes standard approaches like time-scale separation for the limiting system.

Finally, we note that standard stochastic approximation assume that the state evolution follows a stable deterministic system perturbed by *zero-mean* stochastic noise. More specifically, if $\{\mathbf{y}(i)\}_{i\geq 0}$ is the sequence of interest, Theorem 5 assumes that $\{\mathbf{y}(i)\}_{i\geq 0}$ evolves as

$$\mathbf{y}(i+1) = \mathbf{y}(i) + \gamma(i)\left[R(\mathbf{y}(i)) + \Gamma(i+1,\omega,\mathbf{y}(i))\right] \tag{162}$$

where $\{\gamma(i)\}_{i\geq 0}$ is the weight sequence, $\Gamma(i+1,\omega,\mathbf{y}(i))$ is the *zero-mean* noise. If the sequence $\{\mathbf{y}(i)\}_{i\geq 0}$ is supposed to converge to $\mathbf{y}_0$, it further assumes that $R(\mathbf{y}_0) = \mathbf{0}$ and $\mathbf{y}_0$ is a stable equilibrium of the deterministic system

$$\mathbf{y}_d(i+1) = \mathbf{y}_d(i) + \gamma(i)R(\mathbf{y}_d(i)) \tag{163}$$

The $\mathcal{NU}$ algorithm (and its linear version, $\mathcal{LU}$) falls under the purview of this, and we can establish convergence properties using standard stochastic approximation (see Sections II,III-A.) However, the $\mathcal{NLU}$ algorithm cannot be represented in the form of eqn. (162), even ignoring the presence of multiple time-scales. Indeed, as established by Theorem 14, the sequence $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$ is supposed to converge to $\mathbf{1}_N \otimes h(\theta^*)$ a.s. and hence writing eqn. (155) as a stochastically perturbed system around $\mathbf{1}_N \otimes h(\theta^*)$ we have

$$\widetilde{\mathbf{x}}(i+1) = \widetilde{\mathbf{x}}(i) + \gamma(i)\left[R(\widetilde{\mathbf{x}}(i)) + \Gamma(i+1,\omega,\widetilde{\mathbf{x}}(i))\right] \tag{164}$$

where,

$$R(\widetilde{\mathbf{x}}(i)) = -\beta(i)\left(\overline{L}\otimes I_M\right)\left(\widetilde{\mathbf{x}}(i) - \mathbf{1}_N \otimes h(\theta^*)\right) - \alpha(i)\left(\widetilde{\mathbf{x}}(i) - \mathbf{1}_N \otimes h(\theta^*)\right) \tag{165}$$

and

$$\Gamma(i+1,\omega,\widetilde{\mathbf{x}}(i)) = -\beta(i)\left(\widetilde{L}(i)\otimes I_M\right)\left(\widetilde{\mathbf{x}}(i) - \mathbf{1}_N \otimes h(\theta^*)\right) - \beta(i)\left(\mathbf{\Upsilon}(i) + \mathbf{\Psi}(i)\right) + \alpha(i)\left(J(\mathbf{z}(i)) - \mathbf{1}_N \otimes h(\theta^*)\right) \tag{166}$$

Although, $R(\mathbf{1}_N \otimes h(\theta^*)) = \mathbf{0}$ in the above decomposition, the noise $\Gamma(i+1,\omega,\widetilde{\mathbf{x}}(i))$ is not unbiased as the term $(J(\mathbf{z}(i)) - \mathbf{1}_N \otimes h(\theta^*))$ is *not* zero-mean.

With the above discussion in mind, we proceed to the proof of Theorems 14,15, which we develop in stages. The detailed proofs of the intermediate results are provided in the Appendix.

In parallel to the evolution of the state sequence $\{\mathbf{x}(i)\}_{i\geq 0}$, we consider the following update of the auxiliary sequence, $\{\widetilde{\mathbf{x}}^{\circ}(i)\}_{i\geq 0}$:

$$\widetilde{\mathbf{x}}^{\circ}(i+1) = \widetilde{\mathbf{x}}^{\circ}(i) - \beta(i)\left(\overline{L}\otimes I_M\right)\widetilde{\mathbf{x}}^{\circ}(i) - \alpha(i)\left[\widetilde{\mathbf{x}}^{\circ}(i) - J(\mathbf{z}(i))\right] \tag{167}$$

with $\widetilde{\mathbf{x}}^{\circ}(0) = \widetilde{\mathbf{x}}(0)$. Note that in (167) the random Laplacian L is replaced by the average Laplacian $\overline{L}$ and the quantization noises $\mathbf{\Upsilon}(i)$ and $\mathbf{\Psi}(i)$ are not included. In other words, in the absence of link failures and quantization, the recursion (155) reduces to (167), i.e., the sequences $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$ and $\{\widetilde{\mathbf{x}}^{\circ}(i)\}_{i\geq 0}$ are the same.

Now consider the sequence whose recursion adds as input to the recursion in (167) the quantization noises $\mathbf{\Upsilon}(i)$ and $\mathbf{\Psi}(i)$. In other words, in the absence of link failures, but with quantization included, define similarly the sequence $\{\widehat{\mathbf{x}}(i)\}_{i\geq 0}$ given by

$$\widehat{\mathbf{x}}(i+1) = \widehat{\mathbf{x}}(i) - \beta(i)\left(\overline{L}\otimes I_M\right)\widehat{\mathbf{x}}(i) - \alpha(i)\left[\widehat{\mathbf{x}}(i) - J(\mathbf{z}(i))\right] - \beta(i)\left(\mathbf{\Upsilon}(i) + \mathbf{\Psi}(i)\right) \tag{168}$$

with $\widehat{\mathbf{x}}(0) = \widetilde{\mathbf{x}}(0)$. Like before, the recursions (155,156) will reduce to (168) when there are no link failures. However, notice that in (168) the quantization noise sequences $\mathbf{\Upsilon}(i)$ and $\mathbf{\Psi}(i)$ are the sequences resulting from quantizing $\widetilde{\mathbf{x}}(i)$ in (155) and not from quantizing $\widehat{\mathbf{x}}(i)$ in (168).

Define the instantaneous averages over the network as

$$\begin{aligned} \mathbf{x}_{\text{avg}}(i) &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n(i) = \frac{1}{N}\left(\mathbf{1}_N \otimes I_M\right)^T \mathbf{x}(i) \\ \widetilde{\mathbf{x}}_{\text{avg}}(i) &= \frac{1}{N}\sum_{n=1}^{N}\widetilde{\mathbf{x}}_n(i) = \frac{1}{N}\left(\mathbf{1}_N \otimes I_M\right)^T \widetilde{\mathbf{x}}(i) \end{aligned} \tag{169}$$

$$\begin{aligned} \mathbf{x}^{\circ}_{\text{avg}}(i) &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}^{\circ}_n(i) = \frac{1}{N}\left(\mathbf{1}_N \otimes I_M\right)^T \mathbf{x}^{\circ}(i) \\ \widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) &= \frac{1}{N}\sum_{n=1}^{N}\widetilde{\mathbf{x}}^{\circ}_n(i) = \frac{1}{N}\left(\mathbf{1}_N \otimes I_M\right)^T \widetilde{\mathbf{x}}^{\circ}(i) \end{aligned} \tag{170}$$

We sketch the main steps of the proof here. While proving consistency and mean-squared sense convergence, we first show that the average sequence, $\left\{\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\right\}_{i\geq 0}$, converges a.s. to $h(\theta^*)$. This can be done by invoking standard stochastic approximation arguments. Then we show that the sequence $\{\widetilde{\mathbf{x}}^{\circ}(i)\}_{i\geq 0}$ reaches consensus a.s., and clearly the limiting consensus value must be $h(\theta^*)$. Intuitively, the a.s. consensus comes from the fact that, after a sufficiently large number of iterations, the consensus effect dominates over the observation update effect, thus asymptotically leading to consensus. The final step in the proof uses a series of comparison arguments to show that the sequence $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$ also reaches consensus a.s. with $h(\theta^*)$ as the limiting consensus value.

We now detail the proofs of Theorems 14,15 in the following steps.

I: The first step consists of studying the convergence properties of the sequence $\left\{\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\right\}_{i\geq 0}$ (see eqn. (167)), for which we establish the following result.

Lemma 16 Consider the sequence, $\{\widetilde{\mathbf{x}}^{\circ}(i)\}_{i\geq 0}$, given by eqn. (167), under the Assumptions **D.1-D.5**. Then,

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\widetilde{\mathbf{x}}^{\circ}(i) = \mathbf{1}_N \otimes h(\theta^*)\right] = 1 \tag{171}$$

$$\lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^{\circ}(i) - \mathbf{1}_N \otimes h(\theta^*)\right\|^2\right] = 0 \tag{172}$$

Lemma 16 says that the sequence $\{\widetilde{\mathbf{x}}^{\circ}(i)\}_{i\geq 0}$ converges a.s. and in $\mathcal{L}_2$ to $\mathbf{1}_N \otimes h(\theta^*)$. For proving Lemma 16 we first consider the corresponding average sequence $\{\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\}_{i\geq 0}$ (see eqn. (170)). For the sequence $\{\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\}_{i\geq 0}$, we can invoke stochastic approximation algorithms to prove that it converges a.s. and in $\mathcal{L}_2$ to $h(\theta^*)$. This is carried out in Lemma 17, which we state now.

Lemma 17 Consider the sequence, $\left\{\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\right\}_{i\geq 0}$, given by eqn. (170), under the Assumptions **D.1-D.5**. Then,

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) = h(\theta^*)\right] = 1 \tag{173}$$

$$\lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) - h(\theta^*)\right\|^2\right] = 0 \tag{174}$$

In Lemma 16 we show that the sequence $\{\widetilde{\mathbf{x}}^{\circ}(i)\}_{i\geq 0}$ reaches consensus a.s. and in $\mathcal{L}_2$, which together with Lemma 17 establishes the claim in Lemma 16 (see Appendix II for detailed proofs of Lemmas 17,16.)

The arguments in Lemmas 17,16 and subsequent results require the following property of real number sequences, which we state here (see Appendix I for proof.)

Lemma 18 Let the sequences $\{r_1(i)\}_{i\geq 0}$ and $\{r_2(i)\}_{i\geq 0}$ be given by

$$r_1(i) = \frac{a_1}{(i+1)^{\delta_1}}, \quad r_2(i) = \frac{a_2}{(i+1)^{\delta_2}} \tag{175}$$

where $a_1, a_2, \delta_2 \geq 0$ and $0 \leq \delta_1 \leq 1$. Then, if $\delta_1 = \delta_2$, there exists $B > 0$, such that, for sufficiently large non-negative integers, $j < i$,

$$0 \leq \sum_{k=j}^{i-1} \left[\left(\prod_{l=k+1}^{i-1} (1 - r_1(l)) \right) r_2(k) \right] \leq B \tag{176}$$

Moreover, the constant B can be chosen independently of i, j. Also, if $\delta_1 < \delta_2$, then, for arbitrary fixed j,

$$\lim_{i\to\infty} \sum_{k=j}^{i-1} \left[\left(\prod_{l=k+1}^{i-1} (1 - r_1(l)) \right) r_2(k) \right] = 0 \tag{177}$$

(We use the convention that, $\prod_{l=k+1}^{i-1} (1 - r_l) = 1$, for $k = i-1$.)

We note that Lemma 18 essentially studies stability of time-varying deterministic scalar recursions of the form:

$$y(i+1) = r_1(i)y(i) + r_2(i) \tag{178}$$

where $\{y(i)\}_{i\geq 0}$ is a scalar sequence evolving according to eqn. (178) with $y(0) = 0$, and the sequences $\{r_1(i)\}_{i\geq 0}$ and $\{r_2(i)\}_{i\geq 0}$ are given by eqn. (175).

II: In this step, we study the convergence properties of the sequence $\{\widehat{\mathbf{x}}(i)\}_{i\geq 0}$ (see eqn. (168)), for which we establish the following result.

Lemma 19 Consider the sequence $\{\widehat{\mathbf{x}}(i)\}_{i\geq 0}$ given by eqn. (168) under the Assumptions **D.1-D.5**. We have

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty} \widehat{\mathbf{x}}(i) = \mathbf{1}_N \otimes h(\theta^*)\right] = 1 \tag{179}$$

$$\lim_{i\to\infty} \mathbb{E}_{\theta^*}\left[\left\|\widehat{\mathbf{x}}(i) - \mathbf{1}_N \otimes h(\theta^*)\right\|^2\right] = 0 \tag{180}$$

The proof of Lemma 19 is given in Appendix III, and mainly consists of a comparison argument involving the sequences $\left\{\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\right\}_{i\geq 0}$ and $\{\widehat{\mathbf{x}}(i)\}_{i\geq 0}$.

III: This is the final step in the proofs of Theorems 14,15. The proof of Theorem 14 consists of a comparison argument between the sequences $\{\widehat{\mathbf{x}}(i)\}_{i\geq 0}$ and $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$, which is detailed in Appendix IV. The proof of Theorem 15, also detailed in Appendix IV, is a consequence of Theorem 14 and the Assumptions.

V. Conclusion

This paper studies linear and nonlinear *distributed* (vector) parameter estimation problems as may arise in constrained sensor networks. Our problem statement is quite general, including communication among sensors that is quantized, noisy, and with channels that fail at randonm times. These are characteristic of packet communication in wireless sensor networks. We introduce a generic observability condition, the separable estimability condition, that generalizes to distributed estimation the general observability condition of centralized parameter estimation. We study three recursive distributed estimators, $\mathcal{ALU}$, $\mathcal{NU}$, and $\mathcal{NLU}$. We study their asymptotic properties, namely: consistency, asymptotic unbiasedness, and for the $\mathcal{ALU}$ and $\mathcal{NU}$ algorithms their asymptotic normality. The $\mathcal{NLU}$ works in a transformed domain where the recursion is actually linear, and a final nonlinear transformation, justified by the separable estimability condition, recovers the parameter estimate (a stochastic generalization of homeomorphic filtering.) For example, Theorem 14 shows that, in the transformed domain, the $\mathcal{NLU}$ leads to consistent and asymptotically unbiased estimators at every sensor for all separably estimable observation models. Since, the function $h(\cdot)$ is invertible, for practical purposes, a knowledge of $h(\theta^*)$ is sufficient for knowing θ^*. In that respect, the algorithm $\mathcal{NLU}$ is much more applicable than the algorithm $\mathcal{NU}$, which requires further assumptions on the observation model for the existence of consistent and asymptotically unbiased estimators. However, in case, the algorithm $\mathcal{NU}$ is applicable, it provides convergence rate guarantees (for example, asymptotic normality) which follow from standard stochastic approximation theory. On the other hand, the algorithm $\mathcal{NLU}$ does not follow under the purview of standard stochastic approximation theory (see Subsection IV-C) and hence does not inherit these convergence rate properties. In this paper, we presented a convergence theory (a.s. and $\mathcal{L}_2$) of the three algorithms under broad conditions. An interesting future research direction is to establish a convergence rate theory for the $\mathcal{NLU}$ algorithm (and in general, distributed stochastic algorithms of this form, which involve mixed time-scale behavior and biased perturbations.)

Appendix I
Proof of Lemma 18

Proof: [Proof of Lemma 18] We prove for the case $\delta_1 < 1$ first. Consider j sufficiently large, such that,

$$r_1(i) \leq 1, \quad \forall i \geq j \tag{181}$$

Then, for $k \geq j$, using the inequality, $1 - a \leq e^{-a}$, for $0 \leq a \leq 1$, we have

$$\prod_{l=k+1}^{i-1} (1 - r_1(l)) \leq e^{-\sum_{l=k+1}^{i-1} r_1(l)} \tag{182}$$

It follows from the properties of the Riemann integral that

$$\begin{aligned}\sum_{l=k+1}^{i-1} r_1(l) &= \sum_{l=k+1}^{i-1} \frac{a_1}{(i+1)^{\delta_1}} \\ &\geq a_1 \int_{k+2}^{i+1} \frac{1}{t^{\delta_1}} dt \\ &= \frac{a_1}{1-\delta_1}\left[(i+1)^{1-\delta_1} - (k+2)^{1-\delta_1}\right] \qquad (183)\end{aligned}$$

We thus have from eqns. (182,183)

$$\begin{aligned}\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}(1-r_1(l))\right) r_2(l)\right] &\leq \sum_{k=j}^{i-1}\left[e^{-\frac{a_1}{1-\delta_1}(i+1)^{1-\delta_1}} e^{\frac{a_1}{1-\delta_1}(k+2)^{1-\delta_1}}\right]\frac{a_2}{(k+1)^{\delta_2}} \\ &= a_2 e^{-\frac{a_1}{1-\delta_1}(i+1)^{1-\delta_1}} \sum_{k=j}^{i-1}\left[e^{\frac{a_1}{1-\delta_1}(k+2)^{1-\delta_1}} \frac{1}{(k+1)^{\delta_2}}\right] \qquad (184)\end{aligned}$$

Using the properties of Riemann integration again, for sufficiently large j, we have

$$\begin{aligned}\sum_{k=j}^{i-1}\left[e^{\frac{a_1}{1-\delta_1}(k+2)^{1-\delta_1}} \frac{1}{(k+1)^{\delta_2}}\right] &\leq \sum_{k=j}^{i-1}\left[e^{\frac{a_1}{1-\delta_1}(k+2)^{1-\delta_1}} \frac{1}{(\frac{k}{2}+1)^{\delta_2}}\right] \\ &= 2^{\delta_2}\sum_{k=j}^{i-1}\left[e^{\frac{a_1}{1-\delta_1}(k+2)^{1-\delta_1}} \frac{1}{(k+2)^{\delta_2}}\right] \\ &= 2^{\delta_2}\sum_{k=j+2}^{i+1}\left[e^{\frac{a_1}{1-\delta_1}k^{1-\delta_1}} \frac{1}{k^{\delta_2}}\right] \\ &= 2^{\delta_2} e^{\frac{a_1}{1-\delta_1}(i+1)^{1-\delta_1}} \frac{1}{(i+1)^{\delta_2}} + 2^{\delta_2}\sum_{k=j+2}^{i}\left[e^{\frac{a_1}{1-\delta_1}k^{1-\delta_1}} \frac{1}{k^{\delta_2}}\right] \\ &\leq 2^{\delta_2} e^{\frac{a_1}{1-\delta_1}(i+1)^{1-\delta_1}} \frac{1}{(i+1)^{\delta_2}} + 2^{\delta_2}\int_{j+2}^{i+1}\left[e^{\frac{a_1}{1-\delta_1}t^{1-\delta_1}} \frac{1}{t^{\delta_2}}\right] dt \qquad (185)\end{aligned}$$

Again by the fundamental theorem of calculus,

$$\begin{aligned}e^{\frac{a_1}{1-\delta_1}(i+1)^{1-\delta_1}} &= a_1\int_{j+2}^{i+1}\left[e^{\frac{a_1}{1-\delta_1}t^{1-\delta_1}} \frac{1}{t^{\delta_1}}\right] dt + C_1 \\ &= a_1\int_{j+2}^{i+1}\left[e^{\frac{a_1}{1-\delta_1}t^{1-\delta_1}} \frac{1}{t^{\delta_2}} t^{\delta_2-\delta_1}\right] dt + C_1 \qquad (186)\end{aligned}$$

where $C_1 = C_1(j) > 0$ for sufficiently large j. From eqns. (185,186) we have

$$\begin{aligned}
\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}(1-r_1(l))\right)r_2(i)\right] &= a_2 e^{-\frac{a_1}{1-\delta_1}(i+1)^{1-\delta_1}}\sum_{k=j}^{i-1}\left[e^{\frac{a_1}{1-\delta_1}(k+2)^{1-\delta_1}}\frac{1}{(k+1)^{\delta_2}}\right] \\
&\leq \frac{2^{\delta_2}a_2 e^{\frac{a_1}{1-\delta_1}(i+1)^{1-\delta_1}}\frac{1}{(i+1)^{\delta_2}} + 2^{\delta_2}a_2\int_{j+2}^{i+1}\left[e^{\frac{a_1}{1-\delta_1}t^{1-\delta_1}}\frac{1}{t^{\delta_2}}\right]dt}{e^{\frac{a_1}{1-\delta_1}(i+1)^{1-\delta_1}}} \\
&= \frac{2^{\delta_2}a_2}{(i+1)^{\delta_2}} + \frac{2^{\delta_2}a_2\int_{j+2}^{i+1}\left[e^{\frac{a_1}{1-\delta_1}t^{1-\delta_1}}\frac{1}{t^{\delta_2}}\right]dt}{e^{\frac{a_1}{1-\delta_1}(i+1)^{1-\delta_1}}} \\
&\leq \frac{2^{\delta_2}a_2}{(i+1)^{\delta_2}} + \frac{2^{\delta_2}a_2\int_{j+2}^{i+1}\left[e^{\frac{a_1}{1-\delta_1}t^{1-\delta_1}}\frac{1}{t^{\delta_2}}\right]dt}{a_1\int_{j+2}^{i+1}\left[e^{\frac{a_1}{1-\delta_1}t^{1-\delta_1}}\frac{1}{t^{\delta_2}}t^{\delta_2-\delta_1}\right]dt + C_1}
\end{aligned} \tag{187}$$

It is clear that the second term stays bounded if $\delta_1 = \delta_2$ and goes to zero as $i \to \infty$ if $\delta_1 < \delta_2$, thus establishing the Lemma for the case $\delta_1 < 1$. Also, in the case $\delta_1 = \delta_2$, we have from eqn. (187)

$$\begin{aligned}
\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}(1-r_1(l))\right)r_2(i)\right] &\leq \frac{2^{\delta_2}a_2}{(i+1)^{\delta_2}} + \frac{2^{\delta_2}a_2}{a_1 + C_1\left[\int_{j+2}^{i+1}\left[e^{\frac{a_1}{1-\delta_1}t^{1-\delta_1}}\frac{1}{t^{\delta_2}}\right]dt\right]^{-1}} \\
&\leq 2^{\delta_2}a_2 + \frac{2^{\delta_2}a_2}{a_1}
\end{aligned} \tag{188}$$

thus making the choice of B in eqn. (176) independent of i, j.

Now consider the case $\delta_1 = 1$. Consider j sufficiently large, such that,

$$r_1(i) \leq 1, \quad \forall i \geq j \tag{189}$$

Using a similar set of manipulations for $k \geq j$, we have

$$\begin{aligned}
\prod_{l=k+1}^{i-1}(1-r_1(l)) &\leq e^{-a_1\sum_{l=k+1}^{i-1}\frac{1}{l+1}} \\
&\leq e^{-a_1\int_{k+2}^{i+1}\frac{1}{t}dt} \\
&= e^{-a_1\ln\left(\frac{i+1}{k+2}\right)} \\
&= \frac{(k+2)^{a_1}}{(i+1)^{a_1}}
\end{aligned} \tag{190}$$

We thus have

$$\begin{aligned}
\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}(1-r_1(l))\right)r_2(i)\right] &\leq \frac{a_2}{(i+1)^{a_1}}\sum_{k=j}^{i-1}\frac{(k+2)^{a_1}}{(k+1)^{\delta_2}} \\
&\leq \frac{2^{\delta_2}a_2}{(i+1)^{a_1}}\sum_{k=j}^{i-1}\frac{(k+2)^{a_1}}{(k+2)^{\delta_2}} \\
&= \frac{2^{\delta_2}a_2}{(i+1)^{a_1}}\sum_{k=j+2}^{i+1}\frac{k^{a_1}}{k^{\delta_2}}
\end{aligned} \tag{191}$$

Now, if $a_1 \geq \delta_2$, then

$$\begin{aligned}\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}(1-r_1(l))\right)r_2(i)\right] &\leq \frac{2^{\delta_2}a_2}{(i+1)^{a_1}}\sum_{k=j+2}^{i+1}k^{a_1-\delta_2} \\ &= \frac{2^{\delta_2}a_2}{(i+1)^{a_1}}\left[(i+1)^{a_1-\delta_2}+\sum_{k=j+2}^{i}k^{a_1-\delta_2}\right] \\ &\leq \frac{2^{\delta_2}a_2}{(i+1)^{a_1}}\left[(i+1)^{a_1-\delta_2}+int_{j+2}^{i+1}t^{a_1-\delta_2}dt\right] \\ &= \frac{2^{\delta_2}a_2}{(i+1)^{\delta_2}}+\frac{2^{\delta_2}a_2}{a-\delta_2+1}\frac{(i+1)^{a-\delta_2+1}-(j+2)^{a-\delta_2+1}}{(i+1)^{a_1}} \end{aligned} \quad (192)$$

It is clear that the second term remains bounded if $\delta_2 = 1$ and goes to zero if $\delta_2 > 1$. The case $a_1 < \delta_2$ can be resolved similarly, which completes the proof.

■

APPENDIX II
PROOFS OF LEMMAS 17,16

Proof: [Proof of Lemma 17] It follows from eqns. (167,170) and the fact that

$$\left(\mathbf{1}_N \otimes I_M\right)^T \left(\overline{L} \otimes I_M\right) = \mathbf{0} \quad (193)$$

that the evolution of the sequence, $\left\{\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\right\}_{i\geq 0}$ is given by

$$\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i+1) = \widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) - \alpha(i)\left[\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) - \frac{1}{N}\sum_{n=1}^{N}g_n(\mathbf{z}_n(i))\right] \quad (194)$$

We note that eqn. (194) can be written as

$$\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i+1) = \widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) + \alpha(i)\left[R(\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)) + \Gamma(i+1, \widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i), \omega)\right] \quad (195)$$

where

$$R(\mathbf{y}) = -\left(\mathbf{y} - h(\theta^*)\right), \quad \Gamma(i+1, \mathbf{y}, \omega) = \frac{1}{N}\sum_{n=1}^{N}g_n(\mathbf{z}_n(i) - h(\theta^*), \quad \mathbf{y} \in \mathbb{R}^{M\times 1} \quad (196)$$

Such a definition of $R(\cdot), \Gamma(\cdot)$ clearly satisfies Assumptions **B.1,B.2** of Theorem 5. Now, defining

$$V(\mathbf{y}) = \|\mathbf{y} - h(\theta^*)\|^2 \quad (197)$$

we have

$$V(h(\theta^*) = 0, \quad V(\mathbf{y}) > 0, \ \mathbf{y} \neq h(\theta^*), \quad \lim_{\|\mathbf{y}\|\to\infty} V(\mathbf{y}) = \infty \quad (198)$$

Also, we have for $\epsilon > 0$

$$
\begin{aligned}
\sup_{\epsilon<\|\mathbf{y}-h(\theta^*)\|<\frac{1}{\epsilon}} (R(\mathbf{y}), V_{\mathbf{y}}(\mathbf{y})) &= \sup_{\epsilon<\|\mathbf{y}-h(\theta^*)\|<\frac{1}{\epsilon}} \left(-2\|\mathbf{y}-h(\theta^*)\|^2\right) \\
&\leq -2\epsilon^2 \\
&< 0
\end{aligned} \tag{199}
$$

thus verifying Assumption **B.3**. Finally from eqns. (111,196) we have

$$
\begin{aligned}
\|R(\mathbf{y})\|^2 + \mathbb{E}_{\theta^*}\left[\|\Gamma(i+1,\mathbf{y},\omega)\|^2\right] &= \|\mathbf{y}-h(\theta^*)\|^2 + \eta(\theta^*) \\
&\leq k_1(1+V(\mathbf{y})) \\
&\leq k_1(1+V(\mathbf{y})) - (R(\mathbf{y}), V_{\mathbf{y}}(\mathbf{y}))
\end{aligned} \tag{200}
$$

for $k_1 = \max(1, \eta(\theta^*))$. Thus the Assumptions **B.1-B.4** are satisfied, and we have the claim in eqn. (173).

To establish eqn. (174), we note that, for sufficiently large i,

$$
\begin{aligned}
\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) - h(\theta^*)\right\|^2\right] &= (1-\alpha(i-1))^2 \mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i-1) - h(\theta^*)\right\|^2\right] + \alpha^2(i-1)\eta(\theta^*) \\
&\leq (1-\alpha(i-1)) \mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i-1) - h(\theta^*)\right\|^2\right] + \alpha^2(i-1)\eta(\theta^*)
\end{aligned} \tag{201}
$$

where the last step follows from the fact that $0 \leq (1-\alpha(i)) \leq 1$ for sufficiently large i. Continuing the recursion in eqn. (201), we have for sufficiently large $j \leq i$

$$
\begin{aligned}
\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) - h(\theta^*)\right\|^2\right] &\leq \left(\prod_{k=j}^{i-1}(1-\alpha(k))\right)\left\|\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(0) - h(\theta^*)\right\|^2 + \eta(\theta^*)\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}(1-\alpha(l))\right)\alpha^2(k)\right] \\
&\leq \left(e^{-\sum_{k=j}^{i-1}\alpha(k)}\right)\left\|\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(0) - h(\theta^*)\right\|^2 + \eta(\theta^*)\sum_{k=0}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}(1-\alpha(l))\right)\alpha^2(k)\right]
\end{aligned} \tag{202}
$$

From Assumption **D.5**, we note that $\sum_{k=j}^{i-1}\alpha(k) \to \infty$ as $i \to \infty$ because $0.5 < \tau_1 \leq 1$. Thus, the first term in eqn. (202) goes to zero as $i \to \infty$. The second term in eqn. (202) falls under the purview of Lemma 18 with $\delta_1 = \tau_1$ and $\delta_2 = 2\tau_1$ and hence goes to zero as $i \to \infty$. We thus have

$$
\lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) - h(\theta^*)\right\|^2\right] = 0 \tag{203}
$$

■

Proof: [Proof of Lemma 16] Recall from eqns. (167,194) that the evolution of the sequences $\{\widetilde{\mathbf{x}}^{\circ}(i)\}_{i\geq 0}$ and $\{\widetilde{\mathbf{x}}^{\circ}(i)\}_{i\geq 0}$ are given by

$$
\widetilde{\mathbf{x}}^{\circ}(i+1) = \widetilde{\mathbf{x}}^{\circ}(i) - \beta(i)\left(\overline{L}\otimes I_M\right)\widetilde{\mathbf{x}}^{\circ}(i) - \alpha(i)\left[\widetilde{\mathbf{x}}(i) - J(\mathbf{z}(i))\right] \tag{204}
$$

$$
\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i+1) = \widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) - \alpha(i)\left[\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i) - \frac{1}{N}\sum_{n=1}^{N} g_n(\mathbf{z}_n(i))\right] \tag{205}
$$

To establish the claim in eqn. (171), from Lemma 17, it suffices to prove

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\left\|\widetilde{\mathbf{x}}^{\circ}(i)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\right)\right\|=0\right]=1 \tag{206}$$

To this end define the matrix

$$P=\frac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)\left(\mathbf{1}_N\otimes I_M\right)^T \tag{207}$$

and note that

$$P\widetilde{\mathbf{x}}^{\circ}(i)=\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i),\quad P\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)=\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i),\quad\forall i \tag{208}$$

From eqns. (204,205), we then have

$$\begin{aligned}\widetilde{\mathbf{x}}^{\circ}(i+1)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i+1)\right) &= \left[I_{NM}-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right]\left[\widetilde{\mathbf{x}}^{\circ}(i)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\right)\right]\\ &\quad+\alpha(i)\left[J(\mathbf{z}(i))-\frac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)^T J(\mathbf{z}(i))\right]\end{aligned} \tag{209}$$

Choose δ satisfying

$$0<\delta<\tau_1-\frac{1}{2+\epsilon_1}-\tau_2 \tag{210}$$

and note that such a choice exists by Assumption **D.5**. We now claim that

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\frac{1}{(i+1)^{\frac{1}{2+\epsilon_1}+\delta}}\left\|J(\mathbf{z}(i))-\frac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)^T J(\mathbf{z}(i))\right\|=0\right]=1 \tag{211}$$

Indeed, consider any $\epsilon>0$. We then have from Assumption **D.4** and Chebyshev's inequality

$$\begin{aligned}\sum_{i\geq 0}\mathbb{P}_{\theta^*}\left[\frac{1}{(i+1)^{\frac{1}{2+\epsilon_1}+\delta}}\left\|J(\mathbf{z}(i))-\frac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)^T J(\mathbf{z}(i))\right\|>\epsilon\right] &\leq \sum_{i\geq 0}\frac{1}{(i+1)^{1+\delta(2+\epsilon_1)}\epsilon^{2+\epsilon_1}}\\ &\quad\mathbb{E}_{\theta}\left[\left\|J(\mathbf{z}(i))-\frac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)^T J(\mathbf{z}(i))\right\|^{2+\epsilon_1}\right]\\ &= \frac{\kappa(\theta^*)}{\epsilon^{2+\epsilon_1}}\sum_{i\geq 0}\frac{1}{(i+1)^{1+\delta(2+\epsilon_1)}}\\ &< \infty\end{aligned}$$

It then follows from the Borel-Cantelli Lemma (see [37]) that for arbitrary $\epsilon>0$

$$\mathbb{P}_{\theta^*}\left[\frac{1}{(i+1)^{\frac{1}{2+\epsilon_1}+\delta}}\left\|J(\mathbf{z}(i))-\frac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)^T J(\mathbf{z}(i))\right\|>\epsilon\text{ i.o.}\right]=0 \tag{212}$$

where i.o. stands for infinitely often. Since the above holds for ϵ arbitrarily small, we have (see [37]) the a.s. claim in eqn. (211).

Consider the set $\Omega_1\subset\Omega$ with $\mathbb{P}_{\theta^*}\left[\Omega_1\right]=1$, where the a.s. property in eqn. (211) holds. Also, consider the set $\Omega_2\subset\Omega$ with $\mathbb{P}_{\theta^*}\left[\Omega_2\right]=1$, where the sequence $\left\{\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\right\}_{i\geq 0}$ converges to $h(\theta^*)$. Let $\Omega_3=\Omega_1\cap\Omega_2$. It is clear that $\mathbb{P}_{\theta^*}\left[\Omega_3\right]=1$. We will now show that, on Ω_3, the sample paths of the sequence $\{\widetilde{\mathbf{x}}^{\circ}(i)\}_{i\geq 0}$ converge to $(\mathbf{1}_N\otimes h(\theta^*))$, thus proving the Lemma. In the following we index the sample paths by ω to emphasize the fact that we are establishing properties pathwise.

From eqn. (209), we have on $\omega \in \Omega_3$

$$\begin{aligned}\left\|\widetilde{\mathbf{x}}^{\circ}(i+1,\omega)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i+1,\omega)\right)\right\| &\leq \left\|I-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right\|\left\|\widetilde{\mathbf{x}}^{\circ}(i,\omega)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i,\omega)\right)\right\| \\ &+\frac{a}{(i+1)^{\tau_1-\frac{1}{2+\epsilon_1}-\delta}}\left\|\frac{1}{(i+1)^{\frac{1}{2+\epsilon_1}+\delta}}\left[J(\mathbf{z}(i,\omega))-\frac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)^T J(\mathbf{z}(i,\omega))\right]\right\|\end{aligned}$$

For sufficiently large i, we have

$$\left\|I-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right\|\leq 1-\beta(i)\lambda_2(\overline{L}) \tag{213}$$

From eqn. (212) for $\omega\in\Omega_3$ we can choose $\epsilon>0$ and $j(\omega)$ such that

$$\left\|\frac{1}{(i+1)^{\frac{1}{2+\epsilon_1}+\delta}}\left[J(\mathbf{z}(i,\omega))-\frac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)^T J(\mathbf{z}(i,\omega))\right]\right\|\leq\epsilon,\quad\forall i\geq j(\omega) \tag{214}$$

Let $j(\omega)$ be sufficiently large such that eqn. (213) is also satisfied in addition to eqn. (214). We then have for $\omega\in\Omega_3,\ i\geq j(\omega)$

$$\begin{aligned}\left\|\widetilde{\mathbf{x}}^{\circ}(i,\omega)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i,\omega)\right)\right\| &\leq \left(\prod_{k=j(\omega)}^{i-1}\left(1-\beta(k)\lambda_2(\overline{L})\right)\right)\left\|\widetilde{\mathbf{x}}^{\circ}(j(\omega),\omega)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(j(\omega),\omega)\right)\right\| \\ &+a\epsilon\sum_{k=j(\omega)}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}\left(1-\beta(l)\lambda_2(\overline{L})\right)\right)\frac{1}{(k+1)^{\tau_1-\frac{1}{2+\epsilon_1}-\delta}}\right]\end{aligned}$$

For the first term on the R.H.S. of eqn. (215) we note that

$$\begin{aligned}\prod_{k=j(\omega)}^{i-1}\left(1-\beta(k)\lambda_2(\overline{L})\right) &\leq e^{-\lambda_2(\overline{L})\sum_{k=j(\omega)}^{i-1}\beta(k)} \\ &= e^{-b\lambda_2(\overline{L})\sum_{k=j(\omega)}^{i-1}\frac{1}{(k+1)^{\tau_2}}}\end{aligned} \tag{215}$$

which goes to zero as $i\to\infty$ since $\tau_2<1$ by Assumption **D.5**. Hence the first term on the R.H.S. of eqn. (215) goes to zero as $i\to\infty$. The summation in the second term on the R.H.S. of eqn. (215) falls under the purview of Lemma 18 with $\delta_1=\tau_2$ and $\delta_2=\tau_1-\frac{1}{2+\epsilon_1}-\delta$. It follows from the choice of δ in eqn. (210) and Assumption **D.5** that $\delta_1<\delta_2$ and hence the term $\sum_{k=j(\omega)}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}\left(1-\beta(l)\lambda_2(\overline{L})\right)\right)\frac{1}{(k+1)^{\tau_1-\frac{1}{2+\epsilon_1}-\delta}}\right]\to 0$ as $i\to\infty$. We then conclude from eqn. (215) that, for $\omega\in\Omega_3$

$$\lim_{i\to\infty}\left\|\widetilde{\mathbf{x}}^{\circ}(i,\omega)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i,\omega)\right)\right\|=0 \tag{216}$$

The Lemma then follows from the fact that $\mathbb{P}_{\theta^*}\left[\Omega_3\right]=1$.

To establish eqn. (172), we have from eqn. (209)

$$\begin{aligned}&\left\|\widetilde{\mathbf{x}}^{\circ}(i+1)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i+1)\right)\right\|^2\leq\left\|I-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right\|^2\left\|\widetilde{\mathbf{x}}^{\circ}(i)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\right)\right\|^2 \\ &+2\alpha(i)\left\|I-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right\|\left\|\widetilde{\mathbf{x}}^{\circ}(i)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^{\circ}_{\text{avg}}(i)\right)\right\|\left\|J(\mathbf{z}(i))-\tfrac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)^T J(\mathbf{z}(i))\right\| \\ &+\alpha^2(i)\left\|J(\mathbf{z}(i))-\tfrac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)^T J(\mathbf{z}(i))\right\|^2\end{aligned} \tag{217}$$

Taking expectations on both sides and from eqn. (151)

$$\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^\circ(i+1)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^\circ_{\text{avg}}(i+1)\right)\right\|^2\right]\leq\left\|I-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right\|^2\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^\circ(i)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^\circ_{\text{avg}}(i)\right)\right\|^2\right]$$
$$+2\alpha(i)\left\|I-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right\|\kappa_1\left(\theta^*\right)\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^\circ(i)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^\circ_{\text{avg}}(i)\right)\right\|^2\right]$$
$$+2\alpha(i)\left\|I-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right\|\kappa_1\left(\theta^*\right)+\alpha^2(i)\kappa_2(\theta^*)$$

where we used the inequality

$$\left\|\widetilde{\mathbf{x}}^\circ(i)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^\circ_{\text{avg}}(i)\right)\right\|\leq\left\|\widetilde{\mathbf{x}}^\circ(i)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^\circ_{\text{avg}}(i)\right)\right\|^2+1,\quad\forall i \tag{218}$$

Choose j sufficiently large such that

$$\left\|I-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right\|1-\beta(i)\lambda_2(\overline{L}),\quad\forall i\geq j \tag{219}$$

For $i\geq j$, it can then be shown that

$$\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^\circ(i+1)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^\circ_{\text{avg}}(i+1)\right)\right\|^2\right]\leq\left[1-\beta(i)\lambda_2(\overline{L})+2\alpha(i)\kappa_1(\theta^*)\right]\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^\circ(i)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^\circ_{\text{avg}}(i)\right)\right\|^2\right]+\alpha(i)c_1 \tag{220}$$

where $c_1>0$ is a constant. Now choose $j_1\geq j$ and $0<c_2<\lambda_2(\overline{L})$[5] such that,

$$1-\beta(i)\lambda_2(\overline{L})+2\alpha(i)\kappa_1(\theta^*)\leq 1-\beta(i)c_2,\quad\forall i\geq j_1 \tag{221}$$

Then for $i\geq j_1$

$$\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^\circ(i)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^\circ_{\text{avg}}(i)\right)\right\|^2\right]\leq\left(\prod_{k=j_1}^{i-1}\left(1-\beta(k)c_2\right)\right)\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^\circ(j_1)-\left(\mathbf{1}_N\otimes\widetilde{\mathbf{x}}^\circ_{\text{avg}}(j)\right)\right\|^2\right]+c_1\sum_{k=j_1}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}\left(1-\beta(l)c_2\right)\right)\alpha(k)\right] \tag{222}$$

The first term on the R.H.S. of eqn. (220) goes to zero as $i\to\infty$ by the argument given in eqn. (215), while the second term falls under the purview of Lemma 18 and also goes to zero as $i\to\infty$. We thus have the claim in eqn. (172). ■

APPENDIX III
PROOF OF LEMMA 19

Proof: [Proof of Lemma 19] From eqns. (167,168) we have

$$\widehat{\mathbf{x}}(i+1)-\widetilde{\mathbf{x}}^\circ(i+1)=\left[I_{NM}-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}\right]\left[\widehat{\mathbf{x}}(i)-\widetilde{\mathbf{x}}^\circ(i)\right]-\beta(i)\left(\boldsymbol{\Upsilon}(i)+\boldsymbol{\Psi}(i)\right) \tag{223}$$

[5] Such a choice exists because $\tau_1>\tau_2$.

For sufficiently large j, we have

$$\left\|I-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}\right\|\leq 1-\alpha(i),\quad \forall i\geq j \tag{224}$$

We then have from eqn. (223), for $i\geq j$,

$$\begin{aligned}\mathbb{E}_{\theta^*}\left[\left\|\widehat{\mathbf{x}}(i+1)-\widetilde{\mathbf{x}}^{\circ}(i+1)\right\|^2\right] &\leq (1-\alpha(i))^2\,\mathbb{E}_{\theta^*}\left[\left\|\widehat{\mathbf{x}}(i)-\widetilde{\mathbf{x}}^{\circ}(i)\right\|^2\right]+\beta^2(i)\mathbb{E}_{\theta^*}\left[\left\|\mathbf{\Upsilon}(i)+\mathbf{\Psi}(i)\right\|^2\right]\\ &\leq (1-\alpha(i))\,\mathbb{E}_{\theta^*}\left[\left\|\widehat{\mathbf{x}}(i)-\widetilde{\mathbf{x}}^{\circ}(i)\right\|^2\right]+\eta_q\beta^2(i)\end{aligned} \tag{225}$$

where the last step follows from the fact that $0\leq(1-\alpha(i))\leq 1$ for $i\geq j$ and eqn. (17). Continuing the recursion, we have

$$\mathbb{E}_{\theta^*}\left[\left\|\widehat{\mathbf{x}}(i)-\widetilde{\mathbf{x}}^{\circ}(i)\right\|^2\right]\leq\left(\prod_{k=j}^{i-1}(1-\alpha(k))\right)\left\|\widehat{\mathbf{x}}(j)-\widetilde{\mathbf{x}}^{\circ}(j)\right\|^2+\eta_q\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}(1-\alpha(l))\right)\beta^2(k)\right] \tag{226}$$

By a similar argument as in the proof of Lemma 17, we note that the first term on the R.H.S. of eqn. (226) goes to zero as $i\rightarrow\infty$. The second term falls under the purview of Lemma 18 with $\delta_1=\tau_1$ and $\delta_2=2\tau_2$ and goes to zero as $i\rightarrow\infty$ since by Assumption **D.5**, $2\tau_2>\tau_1$. We thus have

$$\lim_{i\rightarrow\infty}\mathbb{E}_{\theta^*}\left[\left\|\widehat{\mathbf{x}}(i)-\widetilde{\mathbf{x}}^{\circ}(i)\right\|^2\right]=0 \tag{227}$$

which shows that the sequence $\{\|\widehat{\mathbf{x}}(i)-\widetilde{\mathbf{x}}^{\circ}(i)\|\}_{i\geq 0}$ converges to 0 in $\mathcal{L}_2$ (mean-squared sense). We then have from Lemma 16

$$\begin{aligned}\lim_{i\rightarrow\infty}\mathbb{E}_{\theta^*}\left[\left\|\widehat{\mathbf{x}}(i)-\mathbf{1}_N\otimes h(\theta^*)\right\|^2\right] &\leq 2\lim_{i\rightarrow\infty}\mathbb{E}_{\theta^*}\left[\left\|\widehat{\mathbf{x}}(i)-\widetilde{\mathbf{x}}^{\circ}(i)\right\|^2\right]+2\lim_{i\rightarrow\infty}\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}^{\circ}(i)-\mathbf{1}_N\otimes h(\theta^*)\right\|^2\right]\\ &= 0\end{aligned} \tag{228}$$

thus establishing the claim in eqn. (180).

We now show that the sequence $\{\|\widehat{\mathbf{x}}(i)-\widetilde{\mathbf{x}}^{\circ}(i)\|\}_{i\geq 0}$ also converges a.s. to a finite random variable. Choose j sufficiently large as in eqn. (224). We then have from eqn. (223)

$$\begin{aligned}\widehat{\mathbf{x}}(i)-\widetilde{\mathbf{x}}^{\circ}(i) &= \left(\prod_{k=j}^{i-1}\left(I_{NM}-\beta(k)\left(\overline{L}\otimes I_M\right)-\alpha(k)I\right)\right)\left(\widehat{\mathbf{x}}(j)-\widetilde{\mathbf{x}}^{\circ}(j)\right)\\ &\quad-\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}\left(I_{NM}-\beta(l)\left(\overline{L}\otimes I_M\right)-\alpha(l)I\right)\right)\beta(k)\mathbf{\Upsilon}(k)\right]\\ &\quad-\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}\left(I_{NM}-\beta(l)\left(\overline{L}\otimes I_M\right)-\alpha(l)I\right)\right)\beta(k)\mathbf{\Psi}(k)\right]\end{aligned} \tag{229}$$

The first term on the R.H.S. of eqn. (229) converges a.s. to zero as $i\rightarrow\infty$ by a similar argument as in the proof of Lemma 17. Since the sequence $\{\mathbf{\Upsilon}(i)\}_{i\geq 0}$ is i.i.d., the second term is a weighted summation of independent random vectors. Define the triangular array of weight matrices, $\{A_{i,k},\ j \leq k \leq i-1\}_{i>j}$, by

$$A_{i,k} = \prod_{l=k+1}^{i-1} \left(I_{NM} - \beta(l)\left(\overline{L} \otimes I_M\right) - \alpha(l)I\right)\beta(k) \tag{230}$$

We then have

$$\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}\left(I_{NM} - \beta(l)\left(\overline{L} \otimes I_M\right) - \alpha(l)I\right)\right)\beta(k)\mathbf{\Upsilon}(k)\right] = \sum_{k=j}^{i-1} A_{i,k}\mathbf{\Upsilon}(k) \tag{231}$$

By Lemma 18 and Assumption **D.5** we note that

$$\begin{aligned}\limsup_{i\to\infty}\sum_{k=j}^{i-1}\|A_{i,k}\|^2 &\leq \limsup_{i\to\infty}\sum_{k=j}^{i-1}\left[\left(\prod_{l=k+1}^{i-1}(1-\alpha(l))\right)\beta^2(k)\right] \\ &= 0\end{aligned} \tag{232}$$

It then follows that

$$\sup_{i>j}\sum_{k=j}^{i-1}\|A_{i,k}\|^2 = C_3 < \infty \tag{233}$$

The sequence $\left\{\sum_{k=j}^{i-1} A_{i,k}\mathbf{\Upsilon}(k)\right\}_{i>j}$ then converges a.s. to a finite random vector by standard results from the limit theory of weighted summations of independent random vectors (see [40], [41], [42]).

In a similar way, the last term on the R.H.S of eqn. (229) converges a.s. to a finite random vector since by the properties of dither the sequence $\{\mathbf{\Psi}(i)\}_{i\geq 0}$ is i.i.d. It then follows from eqn. (229) that the sequence $\{\widehat{\mathbf{x}}(i) - \widetilde{\mathbf{x}}^{\circ}(i)\}_{i\geq 0}$ converges a.s. to a finite random vector, which in turn implies that the sequence $\{\|\widehat{\mathbf{x}}(i) - \widetilde{\mathbf{x}}^{\circ}(i)\|\}_{i\geq 0}$ converges a.s. to a finite random variable. However, we have already shown that the sequence $\{\|\widehat{\mathbf{x}}(i) - \widetilde{\mathbf{x}}^{\circ}(i)\|\}_{i\geq 0}$ converges in mean-squared sense to 0. It then follows from the uniqueness of the mean-squared and a.s. limit, that the sequence $\{\|\widehat{\mathbf{x}}(i) - \widetilde{\mathbf{x}}^{\circ}(i)\|\}_{i\geq 0}$ converges a.s. to 0. In other words,

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\|\widehat{\mathbf{x}}(i) - \widetilde{\mathbf{x}}^{\circ}(i)\| = 0\right] = 1 \tag{234}$$

The claim in eqn. (179) then follows from eqn. (234) and Lemma 16.

■

APPENDIX IV
PROOFS OF THEOREMS 14,15

Proof: [Proof of Theorem 14] Recall the evolution of the sequences $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$, $\{\widehat{\mathbf{x}}(i)\}_{i\geq 0}$ in eqns. (155,168). Then writing $L(i) = \overline{L} + \widetilde{L}(i)$ and using the fact that

$$\left(\widetilde{L}(i) \otimes I_M\right)\widehat{\mathbf{x}}(i) = \left(\widetilde{L}(i) \otimes I_M\right)\widehat{\mathbf{x}}_{\mathcal{C}^{\perp}}(i),\ \forall i \tag{235}$$

we have from eqns. (155,168)

$$\widetilde{\mathbf{x}}(i+1) - \widehat{\mathbf{x}}(i+1) = \left[I_{NM} - \beta(i)\left(L(i) \otimes I_M\right) - \alpha(i)I_{NM}\right]\left(\widetilde{\mathbf{x}}(i) - \widehat{\mathbf{x}}(i)\right) - \beta(i)\left(\widetilde{L}(i) \otimes I_M\right)\widehat{\mathbf{x}}_{\mathcal{C}^{\perp}}(i) \tag{236}$$

For ease of notation, introduce the sequence $\{\mathbf{y}(i)\}_{i\geq 0}$, given by

$$\mathbf{y}(i)=\widetilde{\mathbf{x}}(i)-\widehat{\mathbf{x}}(i) \tag{237}$$

To prove eqn. (157), it clearly suffices (from Lemma 19) to prove

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\mathbf{y}(i)=\mathbf{0}\right]=1 \tag{238}$$

From eqn. (236) we note that the evolution of the sequence $\{\mathbf{y}(i)\}_{i\geq 0}$ is given by

$$\mathbf{y}(i+1)=\left[I_{NM}-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}\right]\mathbf{y}(i)-\beta(i)\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{y}(i)-\beta(i)\left(\widetilde{L}(i)\otimes I_M\right)\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i) \tag{239}$$

The sequence $\{\mathbf{y}(i)\}_{i\geq 0}$ is not Markov, in general, because of the presence of the term $\beta(i)\left(\widetilde{L}(i)\otimes I_M\right)\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)$ on the R.H.S. However, it follows from Lemma 19 that

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)=\mathbf{0}\right]=1 \tag{240}$$

and, hence, asymptotically its effect diminishes. However, the sequence $\{\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)\}_{i\geq 0}$ is not uniformly bounded over sample paths and, hence, we use truncation arguments (see, for example, [36]). For a scalar a, define its truncation $(a)^R$ at level $R>0$ by

$$(a)^R=\begin{cases}\frac{a}{|a|}\min(|a|,R) & \text{if } a\neq 0\\ 0 & \text{if } a=0\end{cases} \tag{241}$$

For a vector, the truncation operation applies component-wise. For $R>0$, we also consider the sequences, $\{\mathbf{y}_R(i)\}_{i\geq 0}$, given by

$$\mathbf{y}_R(i+1)=\left[I_{NM}-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}\right]\mathbf{y}_R(i)-\beta(i)\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{y}_R(i)-\beta(i)\left(\widetilde{L}(i)\otimes I_M\right)\left(\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)\right)^R \tag{242}$$

We will show that for every $R>0$

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\mathbf{y}_R(i)=\mathbf{0}\right]=1 \tag{243}$$

Now, the sequence $\{\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)\}_{i\geq 0}$ converges a.s. to zero, and, hence, for every $\epsilon>0$, there exists $R(\epsilon)>0$ (see [37]), such that

$$\mathbb{P}_{\theta^*}\left[\sup_{i\geq 0}\left\|\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)-\left(\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)\right)^{R(\epsilon)}\right\|=0\right]>1-\epsilon \tag{244}$$

and, hence, from eqns. (239,242)

$$\mathbb{P}_{\theta^*}\left[\sup_{i\geq 0}\left\|\mathbf{y}(i)-\mathbf{y}^{R(\epsilon)}(i)\right\|=0\right]>1-\epsilon \tag{245}$$

This, together with eqn. (243), will then imply

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\mathbf{y}(i)=\mathbf{0}\right]>1-\epsilon \tag{246}$$

Since $\epsilon > 0$ is arbitrary in eqn. (246), we will be able to conclude eqn. (157). Thus, the proof reduces to establishing eqn. (243) for every $R > 0$, which is carried out in the following.

For a given $R > 0$ consider the recursion given in eqn. (242). Choose $\varepsilon_1 > 0$ and $\varepsilon_2 < 0$ such that

$$1 - \varepsilon_2 < 2\tau_2 - \varepsilon_1 \tag{247}$$

and note that the fact, that $\tau_2 > .5$ in Assumption **D.5** permits such choice of $\varepsilon_1, \varepsilon_2$. Define the function $V : \mathbb{N} \times \mathbb{R}^{NM \times 1} \longmapsto \mathbb{R}^+$ by

$$V(i, \mathbf{x}) = i^{\varepsilon_1} \mathbf{x}^T \left(\overline{L} \otimes I_M\right) \mathbf{x} + \rho i^{\varepsilon_2} \tag{248}$$

where $\rho > 0$ is a constant. Recall the filtration $\{\mathcal{F}_i\}_{i \geq 0}$ given in eqn. (119)

$$\mathcal{F}_i = \sigma\left(\mathbf{x}(0), \left\{L(j), \{\mathbf{z}_n(j)\}_{1 \leq N},\ \mathbf{\Upsilon}(j), \mathbf{\Psi}(j)\right\}_{0 \leq j < i}\right) \tag{249}$$

to which all the processes of interest are adapted. We now show that there exists an integer $i_R > 0$ sufficiently large, such that the process $\{V(i, \mathbf{y}_R(i))\}_{i \geq i_R}$ is a non-negative supermartingale w.r.t. the filtration $\{\mathcal{F}_i\}_{i \geq i_R}$. To this end, we note that

$$\begin{aligned}
\mathbb{E}_{\theta^*}\left[V(i+1, \mathbf{y}_R(i+1)) \,|\, \mathcal{F}_i\right] - V(i, \mathbf{y}_R(i)) \quad = \quad & (i+1)^{\varepsilon_1} \mathbf{y}_R^T(i+1) \left(\overline{L} \otimes I_M\right) \mathbf{y}_R(i+1) + \rho (i+1)^{\varepsilon_2} \\
& - i^{\varepsilon_1} \mathbf{y}_R^T(i) \left(\overline{L} \otimes I_M\right) \mathbf{y}_R(i) - \rho i^{\varepsilon_2} \\
= (i+1)^{\varepsilon_1} \Big[& \mathbf{y}_{R,\mathcal{C}^\perp}^T(i) \left(\overline{L} \otimes I_M\right) \mathbf{y}_{R,\mathcal{C}^\perp}(i) - 2\beta(i) \mathbf{y}_{R,\mathcal{C}^\perp}^T(i) \left(\overline{L} \otimes I_M\right)^2 \mathbf{y}_{R,\mathcal{C}^\perp}(i) \\
& - 2\alpha(i) \mathbf{y}_{R,\mathcal{C}^\perp}^T(i) \left(\overline{L} \otimes I_M\right) \mathbf{y}_{R,\mathcal{C}^\perp}(i) + 2\beta(i)\alpha(i) \mathbf{y}_{R,\mathcal{C}^\perp}^T(i) \left(\overline{L} \otimes I_M\right)^2 \mathbf{y}_{R,\mathcal{C}^\perp}(i) \\
& + \beta^2(i) \mathbf{y}_{R,\mathcal{C}^\perp}^T(i) \left(\overline{L} \otimes I_M\right)^3 \mathbf{y}_{R,\mathcal{C}^\perp}(i) + \alpha^2(i) \mathbf{y}_{R,\mathcal{C}^\perp}^T(i) \left(\overline{L} \otimes I_M\right) \mathbf{y}_{R,\mathcal{C}^\perp}(i) \\
& + \beta^2(i) \mathbb{E}_{\theta^*}\left[\mathbf{y}_{R,\mathcal{C}^\perp}^T(i) \left(\widetilde{L}(i) \otimes I_M\right) \left(\overline{L} \otimes I_M\right) \left(\widetilde{L}(i) \otimes I_M\right) \mathbf{y}_{R,\mathcal{C}^\perp}(i) \,\Big|\, \mathcal{F}_i\right] \\
& + 2\beta^2(i) \mathbb{E}_{\theta^*}\left[\mathbf{y}_{R,\mathcal{C}^\perp}^T(i) \left(\widetilde{L}(i) \otimes I_M\right) \left(\overline{L} \otimes I_M\right) \left(\widetilde{L}(i) \otimes I_M\right) \left(\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)\right)^R \,\Big|\, \mathcal{F}_i\right] \\
& + \beta^2(i) \mathbb{E}_{\theta^*}\left[\left(\widehat{\mathbf{x}}_{\mathcal{C}^\perp}^T(i)\right)^R \left(\widetilde{L}(i) \otimes I_M\right) \left(\overline{L} \otimes I_M\right) \left(\widetilde{L}(i) \otimes I_M\right) \left(\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)\right)^R \,|\, \mathcal{F}_i\right]\Big] \\
& + (i+1)^{\varepsilon_2} - i^{\varepsilon_1} \mathbf{y}_{R,\mathcal{C}^\perp}^T(i) \left(\overline{L} \otimes I_M\right) \mathbf{y}_{R,\mathcal{C}^\perp}(i) - \rho i^{\varepsilon_2}
\end{aligned}$$

where we repeatedly used the fact that

$$\left(\overline{L} \otimes I_M\right) \mathbf{y}_R(i) = \left(\overline{L} \otimes I_M\right) \mathbf{y}_{R,\mathcal{C}^\perp}(i), \quad \left(\widetilde{L}(i) \otimes I_M\right) \mathbf{y}_R(i) = \left(\widetilde{L}(i) \otimes I_M\right) \mathbf{y}_{R,\mathcal{C}^\perp}(i) \tag{250}$$

and $\widetilde{L}(i)$ is independent of $\mathcal{F}_i$.

In going to the next step we use the following inequalities, where $c_1 > 0$ is a constant:

$$\begin{aligned}
\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\left(\overline{L}\otimes I_M\right)^2\mathbf{y}_{R,\mathcal{C}^\perp}(i) &\geq \lambda_2^2(\overline{L})\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2 \\
&= \frac{\lambda_2^2(\overline{L})}{\lambda_N(\overline{L})}\lambda_N(\overline{L})\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2 \\
&\geq \frac{\lambda_2^2(\overline{L})}{\lambda_N(\overline{L})}\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\left(\overline{L}\otimes I_M\right)\mathbf{y}_{R,\mathcal{C}^\perp}(i) \qquad (251)
\end{aligned}$$

$$\begin{aligned}
\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\left(\overline{L}\otimes I_M\right)^2\mathbf{y}_{R,\mathcal{C}^\perp}(i) &\leq \lambda_N^2(\overline{L})\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2 \\
&= \frac{\lambda_N^2(\overline{L})}{\lambda_2(\overline{L})}\lambda_2(\overline{L})\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2 \\
&\leq \frac{\lambda_N^2(\overline{L})}{\lambda_2(\overline{L})}\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\left(\overline{L}\otimes I_M\right)\mathbf{y}_{R,\mathcal{C}^\perp}(i) \qquad (252)
\end{aligned}$$

$$\begin{aligned}
\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\left(\overline{L}\otimes I_M\right)^3\mathbf{y}_{R,\mathcal{C}^\perp}(i) &\leq \lambda_N^3(\overline{L})\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2 \\
&= \frac{\lambda_N^3(\overline{L})}{\lambda_2(\overline{L})}\lambda_2(\overline{L})\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2 \\
&\leq \frac{\lambda_N^3(\overline{L})}{\lambda_2(\overline{L})}\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\left(\overline{L}\otimes I_M\right)\mathbf{y}_{R,\mathcal{C}^\perp}(i) \qquad (253)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{\theta^*}\left[\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\left(\widetilde{L}(i)\otimes I_M\right)\left(\overline{L}\otimes I_M\right)\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{y}_{R,\mathcal{C}^\perp}(i)\,\middle|\,\mathcal{F}_i\right] &\leq \lambda_N(\overline{L})\mathbb{E}_{\theta^*}\left[\left\|\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2 \mid\mathcal{F}_i\right] \\
&\leq c_1\lambda_N(\overline{L})\mathbb{E}_{\theta^*}\left[\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2\,\middle|\,\mathcal{F}_i\right] \\
&= c_1\lambda_N(\overline{L})\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2 \\
&\leq \frac{c_1\lambda_N(\overline{L})}{\lambda_2}\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\left(\overline{L}\otimes I_M\right)\mathbf{y}_{R,\mathcal{C}^\perp}(i) \qquad (254)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{\theta^*}\left[\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\left(\widetilde{L}(i)\otimes I_M\right)\left(\overline{L}\otimes I_M\right)\left(\widetilde{L}(i)\otimes I_M\right)\left(\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)\right)^R\,\middle|\,\mathcal{F}_i\right] &\leq \mathbb{E}_{\theta^*}\Big[\left\|\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\right\|\left\|\left(\widetilde{L}(i)\otimes I_M\right)\right\| \\
&\quad\left\|\left(\overline{L}\otimes I_M\right)\right\|\left\|\left(\widetilde{L}(i)\otimes I_M\right)\right\|\left\|\left(\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)\right)^R\right\|\,\Big|\,\mathcal{F}_i\Big] \qquad (255) \\
&\leq Rc_1\lambda_N(\overline{L})\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\| \qquad (256) \\
&\leq Rc_1\lambda_N(\overline{L}) + Rc_1\lambda_N(\overline{L})\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2 \\
&\leq Rc_1\lambda_N(\overline{L}) \qquad (257) \\
&\quad + \frac{Rc_1\lambda_N(\overline{L})}{\lambda_2(\overline{L})}\mathbf{y}^T_{R,\mathcal{C}^\perp}(i)\left(\overline{L}\otimes I_M\right)\mathbf{y}_{R,\mathcal{C}^\perp}(i) \qquad (258)
\end{aligned}$$

$$\mathbb{E}_{\theta^*}\left[\left(\widehat{\mathbf{x}}^T_{\mathcal{C}^\perp}(i)\right)^R\left(\widetilde{L}(i)\otimes I_M\right)\left(\overline{L}\otimes I_M\right)\left(\widetilde{L}(i)\otimes I_M\right)\left(\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)\right)^R\,\middle|\,\mathcal{F}_i\right] \leq R^2c_1\lambda_N(\overline{L}) \qquad (259)$$

$$(i+1)^{\varepsilon_1} - i^{\varepsilon_1} \leq \varepsilon_1(i+1)^{\varepsilon_1-1} \qquad (260)$$

$$\rho(i+1)^{\varepsilon_2} - \rho i^{\varepsilon_2} \leq \rho\varepsilon_2 i^{\varepsilon_2-1} \qquad (261)$$

where going from eqn. (255) to eqn. (256) we use the fact that $\left\|\left(\widehat{\mathbf{x}}_{\mathcal{C}^\perp}(i)\right)^R\right\| \leq R$. Using inequalities (251-261), we have from eqn. (250)

$$\mathbb{E}_{\theta^*}\left[V(i+1,\mathbf{y}_R(i+1))\,\middle|\,\mathcal{F}_i\right]-V(i,\mathbf{y}_R(i)) \;\leq\; (i+1)^{\varepsilon_1}\left[\frac{\varepsilon_1}{(i+1)^1}-2\beta(i)\frac{\lambda_2^2(\overline{L})}{\lambda_N(\overline{L})}-2\alpha(i)\right.$$
$$+2\beta(i)\alpha(i)\frac{\lambda_N^2(\overline{L})}{\lambda_2(\overline{L})}+\beta^2(i)\frac{\lambda_N^3(\overline{L})}{\lambda_2(\overline{L})}$$
$$\left.+\alpha^2(i)+\beta^2(i)\frac{c_1\lambda_N(\overline{L})}{\lambda_2}+2\beta^2(i)\frac{Rc_1\lambda_N(\overline{L})}{\lambda_2(\overline{L})}\right]\mathbf{y}_{R,\mathcal{C}^\perp}^T(i)\left(\overline{L}\otimes I_M\right)\mathbf{y}_{R,\mathcal{C}^\perp}(i)$$
$$+\left[\frac{1}{2\tau_2-\varepsilon_1}\left(2Rc_1\lambda_N(\overline{L})+R^2c_1\lambda_N(\overline{L})\right)+\rho\varepsilon_2 i^{\varepsilon_2-1}\right] \tag{262}$$

For the first term on the R.H.S. of eqn. (262) involving $\mathbf{y}_{R,\mathcal{C}^\perp}^T(i)\left(\overline{L}\otimes I_M\right)\mathbf{y}_{R,\mathcal{C}^\perp}(i)$, the coefficient $-2\beta(i)(i+1)^{\varepsilon_1}$ dominates all other coefficients eventually ($\tau_2<1$ by Assumption **D.5**) and hence the first term on the R.H.S. of eqn. (262) becomes negative eventually (for sufficiently large i). The second term on the R.H.S. of eqn. (262) also becomes negative eventually because $\rho\varepsilon_2<0$ and $1-\varepsilon_2<2\tau_2-\varepsilon_1$ by assumption. Hence there exists sufficiently large i, say i_R, such that,

$$\mathbb{E}_{\theta^*}\left[V(i+1,\mathbf{y}_R(i+1))\,\middle|\,\mathcal{F}_i\right]-V(i,\mathbf{y}_R(i))\leq 0,\quad \forall i\geq i_R \tag{263}$$

which shows that the sequence $\{V(i,\mathbf{y}_R(i))\}_{i\geq i_R}$ is a non-negative supermartingale w.r.t. the filtration $\{\mathcal{F}_i\}_{i\geq i_R}$. Thus, $\{V(i,\mathbf{y}_R(i))\}_{i\geq i_R}$ converges a.s. to a finite random variable (see [37]). It is clear that the sequence ρi^{ε_2} goes to zero as $\varepsilon_2<0$. We then have

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty} i^{\varepsilon_1}\mathbf{y}_R^T(i)\left(\overline{L}\otimes I_M\right)\mathbf{y}_R(i)\text{ exists and is finite}\right]=1 \tag{264}$$

Since $i^{\varepsilon_1}\to\infty$ as $i\to\infty$, it follows

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\mathbf{y}_R^T(i)\left(\overline{L}\otimes I_M\right)\mathbf{y}_R(i)=0\right]=1 \tag{265}$$

Since $\mathbf{y}_R^T(i)\left(\overline{L}\otimes I_M\right)\mathbf{y}_R(i)\geq\lambda_2(\overline{L})\left\|\mathbf{y}_{R,\mathcal{C}^\perp}(i)\right\|^2$, from eqn. (265) we have

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\mathbf{y}_{R,\mathcal{C}^\perp}(i)=0\right]=1 \tag{266}$$

To establish eqn. (243) we note that

$$\mathbf{y}_{R,\mathcal{C}}(i)=\mathbf{1}_N\otimes\mathbf{y}_{R,\text{avg}}(i) \tag{267}$$

where

$$\mathbf{y}_{R,\text{avg}}(i+1)=(1-\alpha(i))\,\mathbf{y}_{R,\text{avg}}(i) \tag{268}$$

Since $\sum_{i\geq 0}\alpha(i)=\infty$, it follows from standard arguments that $\mathbf{y}_{R,\text{avg}}(i)\to 0$ as $i\to\infty$. We then have from eqn. (267)

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\mathbf{y}_{R,\mathcal{C}}(i)=0\right]=1 \tag{269}$$

which together with eqn. (266) establishes eqn. (243). The claim in eqn. (157) then follows from the arguments above.

We now prove the claim in eqn. (158). Recall the matrix P in eqn. (207). Using the fact,

$$P\left(L(i)\otimes I_M\right)=P\left(\overline{L}\otimes I_M\right)=\mathbf{0},\quad \forall i \tag{270}$$

we have

$$P\widetilde{\mathbf{x}}(i+1)=P\widetilde{\mathbf{x}}(i)-\alpha(i)\left[P\widetilde{\mathbf{x}}(i)-PJ(\mathbf{z}(i))\right]-\beta(i)P\left(\boldsymbol{\Upsilon}(i)+\boldsymbol{\Psi}(i)\right) \tag{271}$$

and similarly

$$P\widehat{\mathbf{x}}(i+1)=P\widehat{\mathbf{x}}(i)-\alpha(i)\left[P\widehat{\mathbf{x}}(i)-PJ(\mathbf{z}(i))\right]-\beta(i)P\left(\boldsymbol{\Upsilon}(i)+\boldsymbol{\Psi}(i)\right) \tag{272}$$

Since the sequences $\{P\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$ and $\{P\widehat{\mathbf{x}}(i)\}_{i\geq 0}$ follow the same recursion and start with the same initial state $P\widetilde{\mathbf{x}}(0)$, they are equal, and we have $\forall i$

$$\begin{aligned} P\mathbf{y}(i) &= P\left(\widetilde{\mathbf{x}}(i)-\widehat{\mathbf{x}}(i)\right) \\ &= 0 \end{aligned} \tag{273}$$

From eqn. (239) we then have

$$\mathbf{y}(i+1)=\left[I_{NM}-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right]\mathbf{y}(i)-\beta(i)\left(\widetilde{L}(i)\otimes I_M\right)\mathbf{y}(i)-\beta(i)\left(\widetilde{L}(i)\otimes I_M\right)\widehat{\mathbf{x}}(i) \tag{274}$$

By Lemma 19, to prove the claim in eqn. (157), it suffices to prove

$$\lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\|\mathbf{y}(i)\|^2\right]=0 \tag{275}$$

From Lemma 19, we note that the sequence $\{\widehat{\mathbf{x}}(i)\}_{i\geq 0}$ converges in $\mathcal{L}_2$ to $\mathbf{1}_N\otimes h(\theta^*)$ and hence $\mathcal{L}_2$ bounded, i.e., there exists constant $c_3>0$, such that,

$$\sup_{i\geq 0}\mathbb{E}_{\theta^*}\left[\|\widehat{\mathbf{x}}(i)\|^2\right]\leq c_3<\infty \tag{276}$$

Choose j large enough, such that, for $i\geq j$

$$\left\|I_{NM}-\beta(i)\left(\overline{L}\otimes I_M\right)-\alpha(i)I_{NM}-P\right\|\leq 1-\beta(i)\lambda_2(\overline{L}) \tag{277}$$

Noting that $\widetilde{L}(i)$ is independent of $\mathcal{F}_i$ and $\left\|\widetilde{L}(i)\right\| \leq c_2$ for some constant $c_2 > 0$, we have for $i \geq j$,

$$\begin{aligned}
\mathbb{E}_{\theta^*}\left[\|\mathbf{y}(i+1)\|^2\right] &= \mathbb{E}_{\theta^*}\left[\mathbf{y}^T(i)\left(I_{NM} - \beta(i)\left(\overline{L}\otimes I_M\right) - \alpha(i)I_{NM} - P\right)^2\mathbf{y}(i)\right. \\
&\quad +\beta^2(i)\mathbf{y}^T(i)\left(\widetilde{L}(i)\right)^2\mathbf{y}(i) + \beta^2(i)\widehat{\mathbf{x}}^T(i)\left(\widetilde{L}(i)\right)^2\widehat{\mathbf{x}}(i) \\
&\quad \left. +\beta^2(i)\mathbf{y}^T(i)\left(\widetilde{L}(i)\right)^2\widehat{\mathbf{x}}(i)\right] \\
&\leq \left(1 - \beta(i)\lambda_2(\overline{L})\right)\mathbb{E}_{\theta^*}\left[\|\mathbf{y}(i)\|^2\right] + c_2^2\beta^2(i)\mathbb{E}_{\theta^*}\left[\|\mathbf{y}(i)\|^2\right] \\
&\quad +c_2^2c_3\beta^2(i) + \left(2\beta^2(i)c_2^2c_3^{\frac{1}{2}}\right)\mathbb{E}_{\theta^*}^{\frac{1}{2}}\left[\|\mathbf{y}(i)\|^2\right] \\
&\leq \left(1 - \beta(i)\lambda_2(\overline{L}) + c_2^2\beta^2(i) + 2\beta^2(i)c_2^2c_3^{\frac{1}{2}}\right)\mathbb{E}_{\theta^*}\left[\|\mathbf{y}(i)\|^2\right] \\
&\quad +\beta^2(i)\left(c_2^2c_3 + 2c_2^2c_3^{\frac{1}{2}}\right)
\end{aligned} \tag{278}$$

where in the last step we used the inequality

$$\mathbb{E}_{\theta^*}^{\frac{1}{2}}\left[\|\mathbf{y}(i)\|^2\right] \leq \mathbb{E}_{\theta^*}\left[\|\mathbf{y}(i)\|^2\right] + 1 \tag{279}$$

Now similar to Lemma 16, choose $j_1 \geq j$ and $0 < c_4 < \lambda_2(\overline{L})$, such that,

$$1 - \beta(i)\lambda_2(\overline{L}) + c_2^2\beta^2(i) + 2\beta^2(i)c_2^2c_3^{\frac{1}{2}} \leq 1 - \beta(i)c_4, \quad \forall i \geq j_1 \tag{280}$$

Then for $i \geq j_1$, from eqn. (278)

$$\mathbb{E}_{\theta^*}\left[\|\mathbf{y}(i+1)\|^2\right] \leq \left(1 - \beta(i)c_4\right)\mathbb{E}_{\theta^*}\left[\|\mathbf{y}(i)\|^2\right] + \beta^2(i)\left(c_2^2c_3 + 2c_2^2c_3^{\frac{1}{2}}\right) \tag{281}$$

from which we conclude that $\lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\|\mathbf{y}(i)\|^2\right] = 0$ by Lemma 18 (see also Lemma 16.)

■

Proof: [Proof of Theorem 15] Consistency follows from the fact that by Theorem 14 the sequence $\{\widetilde{\mathbf{x}}(i)\}_{i\geq 0}$ converges a.s. to $\mathbf{1}_N \otimes \theta^*$, and the function $h^{-1}(\cdot)$ is continuous.

To establish the second claim, we note that, if $h^{-1}(\cdot)$ is Lipschitz continuous, there exists constant $k > 0$, such that

$$\left\|h^{-1}(\widetilde{\mathbf{y}}_1) - h^{-1}(\widetilde{\mathbf{y}}_2)\right\| \leq k\left\|\widetilde{\mathbf{y}}_1 - \widetilde{\mathbf{y}}_2\right\|, \quad \forall\ \widetilde{\mathbf{y}}_1, \widetilde{\mathbf{y}}_2 \in \mathbb{R}^{M\times 1} \tag{282}$$

Since $\mathcal{L}_2$ convergence implies $\mathcal{L}_1$, we then have from Theorem 14 for $1 \leq n \leq N$

$$\begin{aligned}
\lim_{i\to\infty}\left\|\mathbb{E}_{\theta^*}\left[\mathbf{x}_n(i) - \theta^*\right]\right\| &\leq \lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\left\|\mathbf{x}_n(i) - \theta^*\right\|\right] \\
&= \lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\left\|h^{-1}\left(\widetilde{\mathbf{x}}_n(i)\right) - h^{-1}\left(h(\theta^*)\right)\right\|\right] \\
&\leq k\lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\left\|\widetilde{\mathbf{x}}_n(i) - h(\theta^*)\right\|\right] \\
&= 0
\end{aligned} \tag{283}$$

which establishes the theorem. ■

References

[1] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D., Massachusetts Institute of Technology, Cambridge, MA, 1984.

[2] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-31, no. 9, pp. 803–812, 1986.

[3] D. Bertsekas, J. Tsitsiklis, and M. Athans, "Convergence theories of distributed iterative processes: A survey," *Technical Report for Information and Decision Systems, Massachusetts Inst. of Technology, Cambridge, MA*, 1984.

[4] H. Kushner and G. Yin, "Asymptotic properties of distributed and communicating stochastic approximation algorithms," *Siam J. Control and Optimization*, vol. 25, no. 5, pp. 1266–1290, Sept. 1987.

[5] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Automat. Contr.*, vol. 49, no. 9, pp. 1520–1533, Sept. 2004.

[6] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Automat. Contr.*, vol. AC-48, no. 6, pp. 988–1001, June 2003.

[7] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Contr. Lett.*, vol. 53, pp. 65–78, 2004.

[8] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3315–3326, July 2008.

[9] ——, "Distributed consensus algorithms in sensor networks with communication channel noise and random link failures," in *41st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2007.

[10] ——, "Distributed average consensus in sensor networks with quantized inter-sensor communication," in *Proceedings of the 33rd International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, Nevada, USA, April 1-4 2008.

[11] Y. Hatano and M. Mesbahi, "Agreement over random networks," in *43rd IEEE Conference on Decision and Control*, vol. 2, Dec. 2004, pp. 2010–2015.

[12] T. C. Aysal, M. Coates, and M. Rabbat, "Distributed average consensus using probabilistic quantization," in *IEEE/SP 14th Workshop on Statistical Signal Processing Workshop*, Maddison, Wisconsin, USA, August 2007, pp. 640–644.

[13] M. E. Yildiz and A. Scaglione, "Differential nested lattice encoding for consensus problems," in *ACM/IEEE Information Processing in Sensor Networks*, Cambridge, MA, April 2007.

[14] A. Kashyap, T. Basar, and R. Srikant, "Quantized consensus," *Automatica*, vol. 43, pp. 1192–1203, July 2007.

[15] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri, "Average consensus on networks with quantized communication," *Submitted to the Int. J. Robust and Nonlinear Control*, 2008.

[16] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *Technical Report 2778, LIDS-MIT*, Nov. 2007.

[17] M. Huang and J. Manton, "Stochastic approximation for consensus seeking: mean square and almost sure convergence," in *Proceedings of the 46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, Dec. 12-14 2007.

[18] A. Das and M. Mesbahi.

[19] A. R. I. D. Schizas and G. B. Giannakis, "Consensus in ad hoc wsns with noisy links - part i: Distributed estimation of deterministic signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, January 2008.

[20] S. Kar, S. A. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2609–2613, June 2008.

[21] U. A. Khan and J. M. F. Moura, "Distributing the kalman filter for large-scale systems," *Accepted for publication, IEEE Transactions on Signal Processing*, 2008.

[22] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, July 2008.

[23] S. Stankovic, M. Stankovic, and D. Stipanovic, "Decentralized parameter estimation by consensus based stochastic approximation," in *46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, 12-14 Dec. 2007, pp. 1535–1540.

[24] I. Schizas, G. Mateos, and G. Giannakis, "Stability analysis of the consensus-based distributed lms algorithm," in *Proceedings of the 33rd International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, Nevada, USA, April 1-4 2008, pp. 3289–3292.

[25] S. Ram, V. Veeravalli, and A. Nedic, “Distributed and recursive parameter estimation in parametrized linear state-space models,” *Submitted for publication*, April 2008.

[26] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI : American Mathematical Society, 1997.

[27] B. Mohar, “The Laplacian spectrum of graphs,” in *Graph Theory, Combinatorics, and Applications*, Y. Alavi, G. Chartrand, O. R. Oellermann, and A. J. Schwenk, Eds. New York: J. Wiley & Sons, 1991, vol. 2, pp. 871–898.

[28] B. Bollobás, *Modern Graph Theory*. New York, NY: Springer Verlag, 1998.

[29] S. Kar and J. Moura, “Distributed consensus algorithms in sensor networks: Quantized data,” November 2007, submitted for publication, 30 pages. [Online]. Available: http://arxiv.org/abs/0712.1609

[30] L. Schuchman, “Dither signals and their effect on quantization noise,” *IEEE Trans. Commun. Technol.*, vol. COMM-12, pp. 162–165, December 1964.

[31] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, “Quantization and dither: A theoretical survey,” *J. Audio Eng. Soc.*, vol. 40, pp. 355–375, May 1992.

[32] A. B. Sripad and D. L. Snyder, “A necessary and sufficient condition for quantization errors to be uniform and white,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 442–448, October 1977.

[33] R. M. Gray and T. G. Stockham, “Dithered quantizers,” *IEEE Trans. Information Theory*, vol. 39, pp. 805–811, May 1993.

[34] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE/ACM Trans. Netw.*, vol. 14, no. SI, pp. 2508–2530, 2006.

[35] E. Lehmann, *Theory of point estimation*. John Wiley and Sons, Inc., 1983.

[36] M. Nevel’son and R. Has’minskii, *Stochastic Approximation and Recursive Estimation*. Providence, Rhode Island: American Mathematical Society, 1973.

[37] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed. Springer Series in Statistics., 2002.

[38] N. Krasovskii, *Stability of motion*. Stanford University Press, 1963.

[39] A. V. Oppenheim and R. W. Schafer, *Digital signal processing*. Prentice-Hall, 1975.

[40] Y. Chow, “Some convergence theorems for independent random variables,” *Ann. Math. Statist.*, vol. 37, pp. 1482–1493, 1966.

[41] Y. Chow and T. Lai, “Limiting behavior of weighted sums of independent random variables,” *Ann. Prob.*, vol. 1, pp. 810–824, 1973.

[42] W. Stout, “Some results on the complete and almost sure convergence of linear combinations of independent random variables and martingale differences,” *Ann. Math. Statist.*, vol. 39, pp. 1549–1562, 1968.