A Mixed Time-Scale Algorithm for Distributed Parameter Estimation : NonLinear Observation Models and Imperfect Communication

Soummya Kar
Department of Electrical and
Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213 USA
Email: soummyak@andrew.cmu.edu

José M. F. Moura
Department of Electrical and
Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213 USA
Email: moura@ece.cmu.edu

Abstract—The paper considers the algorithm $\mathcal{NLU}$ for distributed (vector) parameter estimation in sensor networks, where, the local observation models are nonlinear, and inter-sensor communication is imperfect, in the sense, that the network links fail randomly and inter-sensor transmission is quantized. The paper introduces the class of *separably estimable* observation models, which generalizes the notion of observability in centralized linear estimation to distributed nonlinear estimation. We show that the $\mathcal{NLU}$ algorithm leads to consistent and asymptotically unbiased estimates of the parameter at each sensor for separably estimable observation models. In other words, the sensors reach consensus almost sure (a.s.) to the true parameter value. The algorithm $\mathcal{NLU}$ is a mixed time scale stochastic algorithm, characterized by two different decreasing weight sequences associated with the consensus and innovation updates. The analysis of the $\mathcal{NLU}$ algorithm, thus, does not follow under the purview of standard stochastic approximation, making the analysis developed in the paper of independent theoretical interest.

Index Terms—Distributed parameter estimation, separably estimable, stochastic approximation, Laplacian, consenus.

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. Then, the sensors need to cooperate, 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. 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. In this context, we note that the linear distributed estimation problem has been studied in the literature in non-random networks with unquantized communication, where the observation and consensus protocols are incorporated in the same iteration (see [1], [2], [3], [4].) In this paper we propose and study a generic recursive distributed iterative estimation algorithm, namely, $\mathcal{NLU}$ for distributed parameter estimation with possibly nonlinear observation models at each sensor (see also [5].) As is required, even by centralized estimation schemes, for the estimate sequences generated by the $\mathcal{NLU}$ algorithm at each sensor to have desirable statistical properties, we impose an observability condition. To this end, we introduce a generic observability condition, the *separably estimable* condition for distributed parameter estimation in nonlinear observation models, which generalizes the observability condition of centralized parameter estimation.

The inter-sensor communication is quantized and the communication links among sensors are subject to random failures. This is appropriate, for example, in *digital* communication in WSN when the data exchanges among a sensor and its neighbors are quantized, and the communication channels may fail, 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 $\lambda_2(\overline{L}) > 0$ (see paper for details.) We show that under these conditions, the $\mathcal{NLU}$ algorithm leads to a consistent and asymptotically unbiased estimate of the parameter at each sensor for separably estimable observation models. In other words, the sensors reach consensus almost sure (a.s.) to the true parameter value. 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 (see [5] for a detailed explanation.) The $\mathcal{NLU}$ algorithm does not fall under the purview of

standard stochastic approximation theory, and its analysis requires an altogether different framework as presented in the paper.

We comment briefly on the organization of the rest of the paper. Subsection I-B sets basic notation and preliminaries about spectral graph theory and statistical quantization theory. Section II formulates the problem with the required assumptions and presents the algorithm $\mathcal{NLU}$. The main results regarding the convergence properties of the $\mathcal{NLU}$ algorithm are stated in Section III, whose proofs can be found in [5]. Finally, Section IV concludes the paper.

B. Notation and Preliminaries

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

$$P_k = \frac{1}{k} \mathbf{1}_k \mathbf{1}_k^T \qquad (1)$$

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.

Spectral graph theory. We model the inter-sensor communication graph at time index i by a simple undirected graph $G(i) = (V, E(i))$, where V is the set of sensors with $|V| = N$ and $E(i)$ denotes the (random) set of edges (communication links) active at i. The neighborhood of node n at i is

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

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

$$L(i) = D(i) - A(i) \qquad (3)$$

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

$$0 = \lambda_1(L(i)) \leq \lambda_2(L(i)) \leq \cdots \leq \lambda_N(L(i)) \qquad (4)$$

The smallest eigenvalue $\lambda_1(L(i))$ 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, $\lambda_2(L(i)) > 0$. This second eigenvalue is the algebraic connectivity or the Fiedler value of the network; see [6] for detailed treatment of graphs and their spectral theory.

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 $\mathbf{y} \in \mathbb{R}^{M \times 1}$ we have, $\mathbf{q}(\mathbf{y}) = [k_1 \Delta, \cdots, k_M \Delta] = \mathbf{y} + \mathbf{e}(\mathbf{y})$, where

$$(k_m - \frac{1}{2})\Delta \leq y_i < (k_m + \frac{1}{2})\Delta, \quad 1 \leq m \leq M \qquad (5)$$

$$\frac{\Delta}{2} \mathbf{1}_N \leq \mathbf{e}(\mathbf{y}) < \frac{\Delta}{2} \mathbf{1}_N \qquad (6)$$

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

$$\mathcal{Q}^M = \left\{ [k_1 \Delta, \cdots, k_M \Delta]^T \mid k_i \in \mathbb{Z}, \ \forall i \right\} \qquad (7)$$

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 [7].) 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 [8]). Under these conditions, the error sequence for subtractively dithered systems ([9]) $\{\varepsilon(i)\}_{i \geq 0}$

$$\varepsilon(i) = q(y(i) + \nu(i)) - (y(i) + \nu(i)) \qquad (8)$$

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.

II. PROBLEM FORMULATION AND ALGORITHM $\mathcal{NLU}$

In this section we present the $\mathcal{NLU}$ algorithm for distributed parameter estimation in separably estimable observation models. We start by formally listing the assumptions on the observation model (we introduce the notion of separably estimable) and the inter-sensor communication process involving random link failures and dithered quantized transmission.

A.1)Nonlinear Observation Model: 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

$$\mathbb{P}_{\theta^*}[\mathbf{z}_n(i) \in \mathcal{D}] = \int_{\mathcal{D}} dF_{\theta^*}, \quad \forall \, \mathcal{D} \in \mathbb{B}^{M_n \times 1} \qquad (9)$$

where F_{θ^*} denotes the distribution function of the random vector $\mathbf{z}_n(i)$ and $\mathcal{B}^{M_n \times 1}$ is the Borel sigma algebra of $\mathbb{R}^{M_n \times 1}$. We assume that the distributed observation model is *separably estimable*, a notion which we introduce now.

Definition 1 (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

$$h(\theta) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}_{\theta}\left[g_n(\mathbf{z}_n(i))\right] \qquad (10)$$

is invertible[1]

Before proceeding to interpret the notion of separably estimable models, we illustrate with a motivating example:
Example: Linear Model: Let $\theta \in \mathbb{R}^{M \times 1}$ be an M-dimensional parameter that is to be estimated by a network of N sensors. 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:

$$\mathbf{z}_n(i) = H_n \theta^* + \zeta_n(i) \qquad (11)$$

where $\{\mathbf{z}_n(i) \in \mathbb{R}^{M_n \times 1}\}_{i \geq 0}$ 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 denotes the observation matrix at the n-th sensor. 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. On the other extreme, for a centralized scheme (a fusion center architecture, where all sensors dump their observations) to get a consistent estimate of the parameter, certain observability conditions need to be satisfied by the matrices H_n. To this end, define the matrix

$$G = \sum_{n=1}^{N} H_n^T H_n \qquad (12)$$

It is well known from the theory of linear estimation, that a necessary and sufficient condition for a centralized estimator to yield consistent estimates of the parameter, requires the matrix G to be full-rank (invertible). We now show that, if G is full-rank, then the linear observation model in eqn. (11) becomes separably estimable (see defn. (1)) and hence the $\mathcal{NLU}$ algorithm will lead to asymptotically consistent estimates at each sensor. Indeed, if, assuming the linear model, we define $g_n(\theta) = \overline{H}_n^T \theta, \forall 1 \leq n \leq N$ in eqn. (10), we have $h(\theta) = G\theta$, where G is defined in eqn. (12). Then, separably

---

[1]The factor $\frac{1}{N}$ in eqn. (10) is just for notational convenience, as will be seen later. Also, the R.H.S. of eqn. (10) may not be defined for $\theta \notin \mathcal{U}$. In that case, we assume that $h(\cdot)$ has been properly extended to $\mathbb{R}^{M \times 1}$.

estimable reduces to the invertibility of G. In fact, it also illustrates the generality of the notion of separably estimable and the algorithm $\mathcal{NLU}$, in the sense, that the sufficient condition required by the distributed $\mathcal{NLU}$ algorithm is the necessary condition required by the centralized estimator.
In this regard, the notion of separably estimable can be viewed as an extension of the notion of observability in linear centralized models to nonlinear distributed models. Note that, if an observation model is separably estimable, then the choice of functions $g_n(\cdot)$ is not unique.
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.
A.2)Random Link Failure: The graph Laplacians are

$$L(i) = \overline{L} + \widetilde{L}(i), \ \forall i \geq 0 \qquad (13)$$

where $\{L(i)\}_{i \geq 0}$ is a sequence of i.i.d. Laplacian matrices with mean $\overline{L} = \mathbb{E}[L(i)]$, such that $\lambda_2(\overline{L}) > 0$ (we just require the network to be connected on the average.) We do not make any distributional assumptions on the link failure model. During the same iteration, the link failures can be spatially dependent, i.e., correlated across different edges of the network. This model subsumes the erasure network model, where the link failures are independent both over space and time. Wireless sensor networks motivate this model since interference among the sensors communication correlates the link failures over space, while over time, it is still reasonable to assume that the channels are memoryless or independent. We also note that the above assumption $\lambda_2(\overline{L}) > 0$ does not require the individual random instantiations of $L(i)$ to be connected; in fact, it is possible to have all the instantiations to be disconnected. This enables us to capture a broad class of asynchronous communication models, for example, the random asynchronous gossip protocol analyzed in [10] satisfies $\lambda_2(\overline{L}) > 0$ and hence falls under this framework. More generally, in the asynchronous set up, if the sensors nodes are equipped with independent clocks whose ticks follow a regular random point process (the ticking instants do not have an accumulation point, which is true for all renewal processes, in particular, the Poisson clock in [10]), and at each tick a random network is realized with $\lambda_2(\overline{L}) > 0$ independent of the the networks realized in previous ticks (this is the case with the link formation process assumed in [10]) our model applies.
A.3)Dithered Quantized Communication The sensors exchange quantized messages, where dither is added before quantization (see Subsection I-B.)

We now present the algorithm $\mathcal{NLU}$.

Algorithm $\mathcal{NLU}$: Let $\mathbf{x}(0) = [\mathbf{x}_1^T(0) \cdots \mathbf{x}_N^T(0)]^T$ 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:

$$
\begin{aligned}
\mathbf{x}_n(i+1) = h^{-1}\Bigg(h(\mathbf{x}_n(i)) - \beta(i)\Bigg(\sum_{l \in \Omega_n(i)} (h(\mathbf{x}_n(i)) \\
-\mathbf{q}\left(h(\mathbf{x}_l(i)) + \nu_{nl}(i)\right)) - \alpha(i)\left(h(\mathbf{x}_n(i))\right. \\
\left.-g_n(\mathbf{z}_n(i))\right)\Bigg)
\end{aligned}
$$
$$\qquad (14)$$

based on the information, $\mathbf{x}_n(i), \{\mathbf{q}(h(\mathbf{x}_l(i)) + \nu_{nl}(i))\}_{l \in \Omega_n(i)}, \mathbf{z}_n(i)$, 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, where $\{\nu_{nl}(i)\}$ is the dither sequence.) Here $h^{-1}(\cdot)$ denotes the inverse of the function $h(\cdot)$ and $\{\beta(i)\}_{i\geq0}$, $\{\alpha(i)\}_{i\geq0}$ are appropriately chosen weight sequences. In the sequel, we analyze the $\mathcal{NLU}$ algorithm under the model Assumptions **A.1-A.3**, and in addition we assume:

A.4)Independence and Moment Assumptions. The sequences $\{L(i)\}_{i\geq0}$, $\{\mathbf{z}_n(i)\}_{1\leq n\leq N,\ i\geq0}$, $\{\nu_{nl}^m(i)\}$ (the dither sequence) are mutually independent. We also make the following moment assumption: There exists $\epsilon_1 > 0$, such that the following moment exists $\forall\theta\in\mathcal{U}$:

$$\mathbb{E}_\theta\left[\left\|J(\mathbf{z}(i)) - \frac{1}{N}\left(\mathbf{1}_N\otimes I_M\right)^T J(\mathbf{z}(i))\right\|^{2+\epsilon_1}\right] = \kappa(\theta) < \infty \quad (15)$$

where $J(\mathbf{z}(i)) = [g_1^T(\mathbf{z}_1(i))\cdots g_N^T(\mathbf{z}_N(i))]^T$ and $\bigotimes$ denotes Kronecker product.

A.5): The weight sequences $\{\beta(i)\}_{i\geq0}$, $\{\beta(i)\}_{i\geq0}$ are given by

$$\alpha(i) = \frac{a}{(i+1)^{\tau_1}}, \quad \beta(i) = \frac{b}{(i+1)^{\tau_2}} \quad (16)$$

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

$$.5 < \tau_1, \tau_2 \leq 1, \quad \tau_1 > \frac{1}{2+\epsilon_1} + \tau_2, \quad 2\tau_2 > \tau_1 \quad (17)$$

We note that under Assumption **A.4)** that $\epsilon_1 > 0$, such weight sequences always exist. As an example, if $1/(2+\epsilon_1) = .49$, then the choice $\tau_1 = 1$ and $\tau_2 = .505$ satisfies the inequalities in eqn. (17). Also, note that, under this assumption, $\lim_{i\to\infty}\beta(i)/\alpha(i) = \infty$, giving $\mathcal{NLU}$ a mixed time-scale behavior.

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

III. STATEMENT OF MAIN RESULTS

In this section, we state the main results regarding consistency and asymptotic unbiasedness of the $\mathcal{NLU}$ algorithm. We start by introducing some terminology from the sequential estimation literature (see, for example, [11]).

Definition 2 (Consistency) : A sequence of estimates $\{\mathbf{x}^\bullet(i)\}_{i\geq0}$ is called consistent if

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\mathbf{x}^\bullet(i) = \theta^*\right] = 1, \quad \forall\theta^*\in\mathcal{U} \quad (18)$$

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 3 (Asymptotic Unbiasedness) :

A sequence of estimates $\{\mathbf{x}^\bullet(i)\}_{i\geq0}$ is called asymptotically unbiased if

$$\lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\mathbf{x}^\bullet(i)\right] = \theta^*, \quad \forall\theta^*\in\mathcal{U} \quad (19)$$

We now state the main result characterizing the convergence properties of the $\mathcal{NLU}$ algorithm.

Theorem 4 Consider the $\mathcal{NLU}$ algorithm under the Assumptions **A.1)-A.6)**. Let $\{\mathbf{x}(i)\}_{i\geq0}$ be the state sequence generated, as given by eqns. (14). We then have

$$\mathbb{P}_{\theta^*}\left[\lim_{i\to\infty}\mathbf{x}_n(i) = \theta^*, \ \forall\ 1\leq n\leq N\right] = 1 \quad (20)$$

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.,

$$\lim_{i\to\infty}\mathbb{E}_{\theta^*}\left[\mathbf{x}_n(i)\right] = \theta^*, \ \forall\ 1\leq n\leq N \quad (21)$$

Proof: For a proof, see [5]. ∎

Theorem 4 thus states that the sensors reach consensus asymptotically and the limiting consensus state is the true value of the parameter vector, θ^*. Also, under the assumption of Lipschitz continuity of $h^{-1}(\cdot)$ (which, for example, holds for the linear model example, given in Assumption **A.1)**), we have asymptotic unbiasedness.

IV. CONCLUSION

The paper presents the $\mathcal{NLU}$ algorithm for distributed parameter estimation in sensor networks, where the observation model at each sensor is possibly nonlinear and inter-sensor communication is imperfect because of random link failures and quantized inter-sensor transmission. We introduce the notion of separably estimable observation models, which generalize the notion of observability in linear centralized estimation to nonlinear decentralized estimation. We show that, for such models, the $\mathcal{NLU}$ algorithm yields consistent and asymptotically unbiased estimates of the parameter value at each sensor. The algorithm $\mathcal{NLU}$ is quite general, in this regard, as it does not require any distributional assumptions on the observation process, only some moment conditions suffice.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

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

[5] S. Kar, J. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," August 2008, submitted for publication, 51 pages. [Online]. Available: http://arxiv.org/abs/0809.0009

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

[7] 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

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

[9] 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.

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

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