

Distributed Parameter Estimation in Unreliable Sensor Networks via a Broadcast Gossip Algorithm

Huiwei Wang, Xiaofeng Liao, Zidong Wang, *IEEE Fellow*, Tingwen Huang, and Guo Chen

Abstract—In this paper, we present an asynchronous algorithm to estimate the unknown parameter under an unreliable network which allows new sensors to join and old sensors to leave, and can tolerate link failures. Each sensor has access to partially informative measurements when it is awakened. In addition, the proposed algorithm can avoid the interference among messages and effectively reduce the accumulated measurement and quantization errors. Based on the theory of stochastic approximation, we prove that our proposed algorithm almost surely converges to the unknown parameter. Finally, we present a numerical example to assess the performance and the communication cost of the algorithm.

Index Terms—Broadcast gossip algorithm, distributed parameter estimation, quantized communication, unreliable sensor network.

I. INTRODUCTION

Distributed parameter estimation is one of the fundamental problems in wireless sensor networks (WSNs), and the first step in a wider range of applications such as event detection and classification. In practice, WSNs may be unreliable [1]–[3] due to external malicious attacks, energy exhausting, incorrect sensing and other factors. A convenient way when dealing with agreement in unreliable networks is to exchange messages by an asynchronous framework. The best known one is gossip-based algorithms which have attracted considerable recent attention [5]–[10]. In addition, limitations on the sensor cost, bandwidth, and energy budget dictate that information transmitted between sensors has to be quantized in practice [11]–[14]. However, to the best of the authors’ knowledge, few authors have considered the distributed parameter estimation problem for unreliable WSNs by an asynchronous quantization communication framework so far.

To overcome the challenges from unreliable network environments, we modify the broadcast gossip algorithm [6]–[8] to estimate the unknown parameter associated with incompletely informative samples [4] with noise measured by local sensors and communicated through quantization channels. In a round, each randomly selected sensor broadcasts its quantized estimate to its one-hop neighbors, then each neighbor processes its

measurement and updates its estimate through local information exchange if it receives only one broadcasted message, and the remaining sensors sustain their estimates. It is shown that by iterating this procedure, the proposed algorithm is capable of estimating the unknown parameter with probability one.

In summary, our paper makes two main contributions. The first contribution is to provide a simple broadcast based gossiping algorithm for the distributed parameter estimation problem that is fully asynchronous in spite of noise measurements and estimate updates, which will effectively reduce the accumulated measurement and quantization errors raised by algorithms based on synchronous updates. The second one is that the broadcasting sensors of each round are usually probabilistic in nature and the iteration is updated by local exchange based on radio transmission, which guarantee that the algorithm is robust to sensor drifting in and out of the network, message interference and link failures.

II. RELATED WORK AND DISCUSSIONS

Recently, the paper [1] proposes a distributed EM algorithm to be implemented in a WSN with a decentralized architecture using diffusion-like strategies. In [2], the authors design a consensus-based algorithm for distributed estimation problem with linear or nonlinear measurements. Further improvement of this work is presented in [3] to handle the case: WSNs with randomly switching topologies. Of note is that all of them [1]–[3] are developed to deal with the unreliable environment including only link failures based on synchronous updates framework. Though synchronous algorithms have advantages in convergence time, they accumulate more and more quantization and measurement noise over time, which greatly affect the computations precision. Also, the algorithms based on synchronous updates are unstable or fail during periods of even modest disruption.

In comparison, the presented algorithm in this paper has overcome these drawbacks and has the following major advantages. First, our algorithm is fully asynchronous thus can effectively reduce the accumulated measurement and quantization noise, and further enhance the computation precision than algorithms based on synchronous updates [1]–[3]. In addition, our algorithm does not require error recovery mechanisms thus can solve gracefully distributed parameter estimation over networks with a huge number of nodes. Second, the broadcasting sensors of each round are usually probabilistic in nature and the iteration is updated by local exchange based on radio transmission, which will insure the algorithm to achieve high stability under stress and disruptions including not only

H. Wang and X. Liao are with College of Electronic and Information Engineering, Southwest University, Chongqing, 400715, P.R. China. (E-mail: huiwei.wang@gmail.com, xfliao@swu.edu.cn)

Z. Wang is with the Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex, UB8 3PH, U.K. (E-mail: zidong.wang@brunel.ac.uk)

T. Huang is with Texas A&M University at Qatar, Doha, P.O. Box 23874, Qatar. (E-mail: tingwen.huang@qatar.tamu.edu)

G. Chen is with School of Electrical and Information Engineering, The University of Sydney, Australia. (E-mail: guo.chen@sydney.edu.au)

common link failures but also sensor equipment failures due to the limited computational and energy resources. In particular, when new nodes are allowed to join in or leave from the network, by denoting a sensor number sequence $\{N(t)\}_{t \geq 0}$, $N_{\min} = \min_t \{N(t)\}$ and $N_{\max} = \max_t \{N(t)\}$, then we can assert that the estimate of each sensor converges to the unknown parameter with probability one as long as at least one sensor does not terminate finitely.

III. PRELIMINARY AND PRIMARY MOTIVATIONS

In the following, we briefly introduce the underlying graph of the WSNs and the framework of probabilistic quantization.

Network Model: We model an unreliable WSN as a time-varying undirected graph $\mathcal{G}(t) = \{\mathcal{V}(t), \mathcal{E}(t)\}$ of order $N(t)$, consisting of a set of nodes $\mathcal{V}(t) = \{1, 2, \dots, N(t)\}$ and a set of edges $\mathcal{E}(t) \subseteq \mathcal{V}(t) \times \mathcal{V}(t)$. An edge in graph $\mathcal{G}(t)$ is denoted by $e_{ij}(t) = (i, j)$. If there is an edge from node j to node i , then it is said that nodes i and j can communicate with each other reliably. The neighborhood set of node i at time t is denoted by $\mathcal{N}_i(t) = \{j | (i, j) \in \mathcal{E}(t)\}$, and $d_i(t) = |\mathcal{N}_i(t)|$. As usual, we assume that there is no self-loop in $\mathcal{G}(t)$, $\forall t$. Throughout the paper, we denote N_{\max} and N_{\min} by $\max_t \{N(t)\}$ and $\min_t \{N(t)\}$, respectively.

Quantization Scheme: Assume that all sensors are equipped with identical uniform probability quantizers $q(\cdot) : \mathbb{R}^d \rightarrow \mathcal{Q}^d$ applied componentwise. Denote $x_i = (x_{i1}, \dots, x_{id})^\top$. For $i \in \mathcal{V}(t)$, $j \in \{1, \dots, d\}$, $x_{ij} \in \mathbb{R}$ is identically uniformly bounded to a finite interval $[-U, U]$. Furthermore, we wish to obtain a quantized message $q[x_{ij}]$ with length ℓ bits, where ℓ represents the quantization precision. Therefore, we have $L = 2^\ell$ quantization points given by the set $\Pi = \{\pi_1, \pi_2, \dots, \pi_L\}$ where $\pi_1 = -U$ and $\pi_L = U$. The points are uniformly spaced such that $\Delta = \pi_{k+1} - \pi_k$ for $k \in \{1, 2, \dots, L-1\}$. It follows that $\Delta = 2U/(2^\ell - 1)$. Now suppose $x_{ij} \in [\pi_k, \pi_{k+1})$, then x_{ij} is quantized in a probabilistic manner

$$\mathbb{P}\{q[x_{ij}] = \pi_{k+1}\} = r \quad \text{and} \quad \mathbb{P}\{q[x_{ij}] = \pi_k\} = 1 - r$$

where $r = (x_{ij} - \pi_k)/\Delta$, we refer to [12] to make further relevant comments. It is easy to see that when the variable is exactly equal to a quantization centroid, there is zero probability of choosing another centroid. Therefore, it follows from [11] that the message $q[x_{ij}]$ is an unbiased representation of x_{ij} , i.e., $\mathbb{E}\{q[x_{ij}]\} = x_{ij}$, and $\mathbb{E}\{(q[x_{ij}] - x_{ij})^2\} \leq U^2/(2^\ell - 1)^2 = \Delta^2/4$.

We denote the quantization noisy by $\xi_i = (\xi_{i1}, \dots, \xi_{id})^\top$, where

$$\xi_{ij} = q[x_{ij}] - x_{ij},$$

it is then an i.i.d. sequence of uniformly distributed random variables on $[-\Delta/2, \Delta/2]$. As pointed out in [12], probabilistic quantization is equivalent to a ‘‘dithered quantization’’ method [2]. It has been shown by Schuchman that the subtractive dithering process yields error signal values that are statistically independent from each other and the input. We, therefore, conclude that $\{\xi_i(t)\}_{t \geq 0}$ is independent of the estimate sequence $\{x_i(t)\}_{t \geq 0}$, i.e., $\mathbb{E}\{\xi_i(t)\}_{t \geq 0} = 0$ and $\mathbb{E}\{\|\xi_i(t)\|_2^2\} \leq d\Delta^2/4$.

IV. DISTRIBUTED PARAMETER ESTIMATION VIA ASYNCHRONOUS ALGORITHMS

In this section, we will describe the problem in this paper. Let $\theta^* \in \mathbb{R}^d$ be an unknown d -dimensional parameter to be estimated by a sensor network. At each time, each sensor maintains an estimate of the unknown parameter, we denote the estimate of sensor i at time t by $x_i(t) \in \mathbb{R}^d$. In round t , the following events occur:

- Each sensor $i \in \mathcal{V}(t)$ is allowed to wake up, independent with probability p . Let $\mathcal{M}(t)$ be the set of awakened sensors, and $|\mathcal{M}(t)|$ be the cardinality of $\mathcal{M}(t)$;
- Each sensor $m \in \mathcal{M}(t)$ broadcasts wirelessly its estimate information $x_m(t)$, and then the broadcasted value is successfully received by the sensors that are within sensor m 's communication radius;
- To efficiently avoid the interference among messages, suppose $\mathcal{N}^*(t) \triangleq \{i \in \mathcal{V}(G) : |\mathcal{N}_i(t) \cap \mathcal{M}(t)| = 1\}$ to denote the set of sensors to receive the eligible broadcasted value. Let $m(i) \in \mathcal{M}(t)$ be the unique sensor index such that $i \in \mathcal{N}^*(t)$. Then, each sensor $i \in \mathcal{N}^*(t)$ has local access to partially informative measurements with noise about the unknown parameter

$$z_i(t) = H_i(t)\theta^* + \zeta_i(t) \in \mathbb{R}^{d_i}, \quad 0 < d_i \leq d$$

where $H_i(t) \in \mathbb{R}^{d_i \times d}$ is the measurement matrix with entries not all zero, which forms an i.i.d. sequence with mean \bar{H}_i and finite second moment, and $\zeta_i(t)$ is a zero mean noise and has variance σ^2 .

- For all $i \in \mathcal{N}^*(t)$, each sensor updates its estimate according to the following equation:

$$x_i(t+1) = (1 - \beta(t))x_i(t) + \beta(t)q[x_{m(i)}(t)] + \beta(t)\bar{H}_i^\top(z_i(t) - \bar{H}_i x_i(t)),$$

where $\beta(t) = 1/\gamma t^{1-\eta}$ is the stepsize with constants γ and η . The remaining sensors maintain their prior estimates as

$$x_i(t+1) = x_i(t), \quad i \notin \mathcal{N}^*(t).$$

- Let $t = t + 1$, and all sensors fall asleep.

This procedure takes place at every clock tick, and a flow chart is given in Fig. 1 to give a clear understanding of how it works. From the procedure, we construct the diffusion matrix $A(t) = [a_{ij}(t)]$ as follows

$$a_{ij}(t) = \begin{cases} \beta(t), & i \in \mathcal{N}^*(t), j = m(i) \in \mathcal{M}(t) \\ 1 - \beta(t), & i \in \mathcal{N}^*(t), i = j \\ 1, & i \notin \mathcal{N}^*(t), i = j \\ 0, & \text{elsewhere.} \end{cases}$$

Obviously, $A(t)$ is a row-stochastic matrix for any t .

It's worth pointing out here that, in the each round of the algorithm, the index of the each sensor is re-allocated, thus each sensor needs to keep their own parameter and estimate. To establish a uniform iterative protocol, we define $\mathbf{H}(t) =$

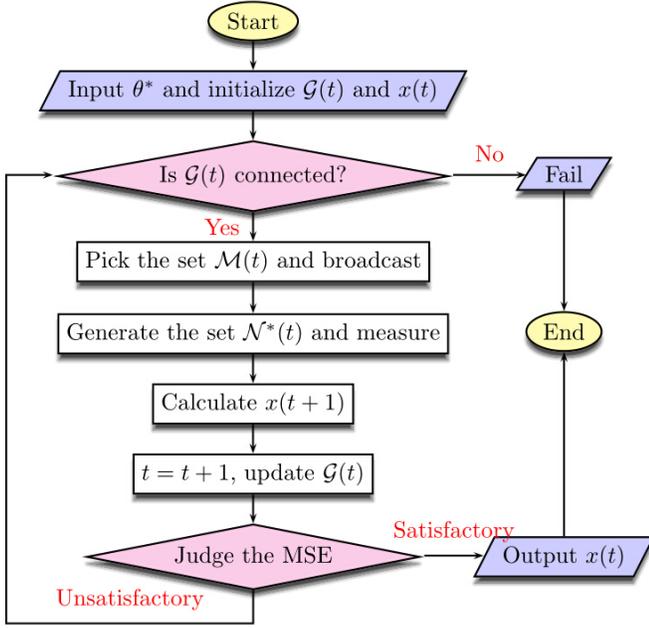


Fig. 1: The flow chart of the proposed algorithm.

$\text{diag}\{\bar{H}_1^\top, \dots, \bar{H}_{N(t)}^\top\}$ and $\Xi(t) = \text{diag}\{\delta_1(t), \dots, \delta_{N(t)}(t)\}$ with

$$\delta_i(t) = \begin{cases} 1, & i \in \mathcal{N}^*(t), \\ 0, & i \notin \mathcal{N}^*(t). \end{cases}$$

Next, we define two noise vectors: $\Phi(t) = [\Xi(t) \otimes \mathbf{I}_d] \mathbf{H}(t) \zeta(t) \in \mathbb{R}^{N(t)d}$, where \mathbf{I}_d is the identity matrix, and $\Psi(t) = (\psi_1^\top(t), \dots, \psi_{N(t)}^\top(t))^\top \in \mathbb{R}^{N(t)d}$ with vector components

$$\psi_i(t) = \begin{cases} \xi_{m(i)}(t), & i \in \mathcal{N}^*(t), \\ 0, & i \notin \mathcal{N}^*(t). \end{cases}$$

Suppose that the sequences $\{\xi(t)\}_{t \geq 0}$ and $\{\zeta(t)\}_{t \geq 0}$ are mutually independent, then $\mathbb{E}\{\Psi(t)\} = \mathbb{E}\{\Phi(t)\} = 0$. Also, it is easy to get

$$\sup_t \mathbb{E}\{\|\Psi(t)\|_2^2\} \leq N_{\max} d \Delta^2 / 4, \quad \text{and} \quad (1)$$

$$\sup_t \mathbb{E}\{\|\Phi(t)\|_2^2\} \leq N_{\max} d \sigma^2. \quad (2)$$

On the basis of the above, we now rewrite the proposed algorithm in a uniform compact form as

$$\begin{aligned} x(t+1) = & (A(t) \otimes \mathbf{I}_d - \beta(t)[\Xi(t) \otimes \mathbf{I}_d] \mathbf{H}(t) \mathbf{H}^\top(t)) x(t) \\ & + \beta(t)[\Xi(t) \otimes \mathbf{I}_d] \mathbf{H}(t) \mathbf{H}^\top(t) [\theta^* \otimes \mathbf{1}_{N(t)}] \\ & + \beta(t)[\Psi(t) + \Phi(t)]. \end{aligned} \quad (3)$$

Ideally, the estimate $x_i(t)$ converges to the unknown parameter θ^* for any $i \in \mathcal{V}(t)$, as t approaches infinity. The following proposition states our main result, and the corresponding proof is given in the Appendix.

Proposition 1: If the stepsize sequence is nonnegative, non-increasing, and it satisfies $\beta(t) = 1/\gamma t^{1-\eta}$ for $\eta \in (0, 1/2)$ and $\gamma \geq \max_t\{4, \max_k [\mathbf{H}(t) \mathbf{H}^\top(t)]_{kk}\}$, then the estimate

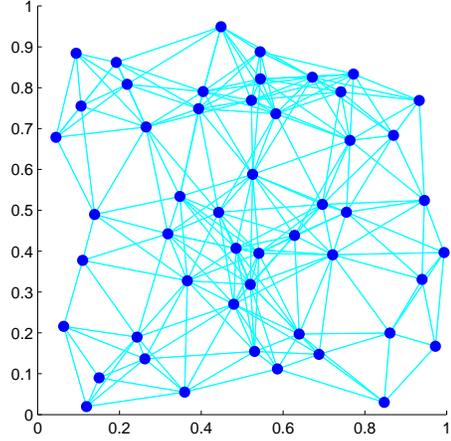


Fig. 2: The network topology without node or link failure.

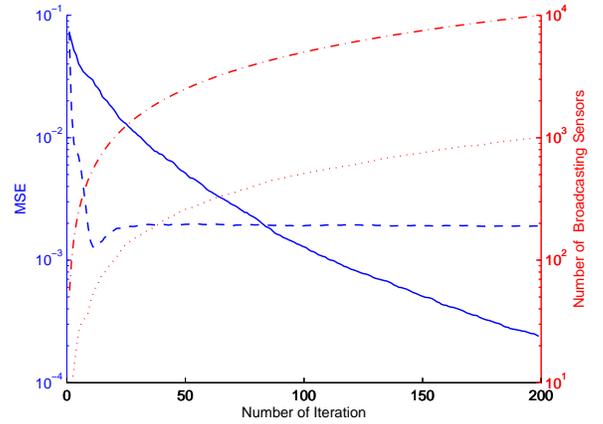


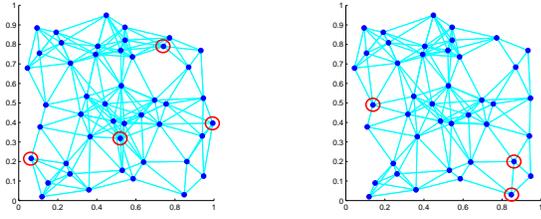
Fig. 3: The MSE performance and cost versus the iteration count. The parameters used here are: $\Delta = 0.01$, $\gamma = 4$ and $\eta = 0.495$. The blue solid and dash lines denote the MSE of DPE-AA and DPE-SA [2], respectively; the red dot and dash-dot lines denote the measurement costs of the DPE-AA and the DPE-SA [2], respectively.

sequence $\{x_i(t)\}_{t \geq 0}$ converges to the unknown parameter, almost surely, i.e.,

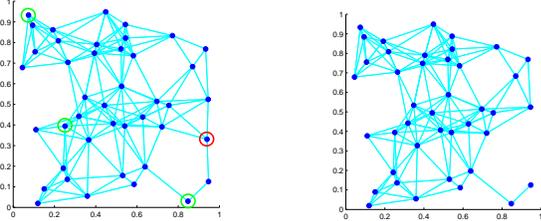
$$\mathbb{P}\left\{\lim_{t \rightarrow \infty} x_i(t) = \theta^*\right\} = 1, \quad \forall i \in \mathcal{V}(t).$$

V. NUMERICAL EXAMPLE

We simulate a sensor network with $N = 50$ nodes as a random geometric graph, as shown in Fig. 2, with a communication radius $R_c = \sqrt{\log(N)/N}$ on the unit square $[0, 1] \times [0, 1]$. Each node may wake up and broadcast its estimate independent with probability $p = 0.1$. Each node is initialized as uniformly distributed random values with support in the $[0, 1]$ interval and the mean $\bar{x}(0)$ is $0.5 \times \mathbf{1}_d$, where $\mathbf{1}_d$ is the d -dimensional vector of all 1. Let the unknown parameter θ^* be $0.5 \times \mathbf{1}_d$, since it is the optimal value in terms of convergence speed and provides a tradeoff for the MSE.



(a) The network topology $t = 1201$. (b) The topology when $t = 2401$.



(c) The topology when $t = 3601$. (d) The topology when $t = 4801$.

Fig. 4: The network topology with node and link failure.

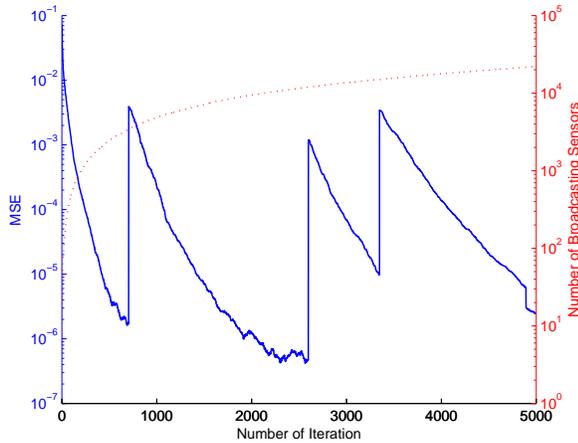


Fig. 5: The MSE performance and cost versus the iteration count. The parameters used here are: $\Delta = 0.01$, $\gamma = 4$ and $\eta = 0.495$. The blue solid and red dash lines denote the MSE and measurement costs of the DPE-AA, respectively.

We first take into account the MSE performance and costs of our algorithm (DPE-AA) under the reliable network topology without any failure, and the comparison with the algorithm proposed in [2] (DPE-SA). To simplify the analysis, we let $d = 1$, $d_i = 1$ and $H_i = 1$ for any i . As shown in Fig. 3, the DPE-SA [2] has absolute advantage in convergence rate for an equal iteration count but with enormous communication costs. As a result, it reaches a steady state MSE at a lower precision due to the accumulated quantization and measurement noise, which also verifies our theoretical analysis. By contrast, the MSE of the DPE-AA persistently decreases and will ultimately reach a higher precision with lower communication costs. In addition, the MSE precision of our algorithm is commensurate with the DPE-SA [2] for an equal communication costs.

Next, we conduct the performance studies for the DPE-

AA under the unreliable WSN which admits sensors join in and/or leave from the network. To provide an implementation independent result, we let each sensor leave from the network with a uniform probability of 0.005% and a new sensor join in the network with probability of 0.05%. We occasionally draw the network topology when the iteration satisfies $t \bmod 1200 = 1$, as shown in Fig. 4. Specifically, we add the red and green circles around the sensors' location to clearly show the changes in the network topology and scaling over time. It is still identified from the plot that the topology in Fig. 4(d) is evolved from the initial topology in Fig. 2. If we improve the above probability, the ultimate topology is entirely different from the initial topology, and is likely to further generate the isolated sub-topology.

Considered next are the performance and cost of the DPE-AA under the above unreliable WSNs environment. To simplify the analysis, we let $d = 2$, $d_i = 1$ and $H_i = [1 \ 1]$ for any i . As shown in Fig. 5, the MSE will persistently drop down and reach to the precision 10^{-6} or higher until the new sensor joins in the network. However, the MSE will undergo a pulse when the new sensor joins in the network. Note that the intensity of the pulse depends on the distance between the new node's initial estimate and the unknown parameter. The plot shows that it can maintain a high MSE precision for the whole network while physical topology is changing, then we can assert that the estimate of each sensor converges to the unknown parameter with probability one as long as at least one sensor does not terminate finitely.

VI. CONCLUSIONS

In this paper, a consensus-based asynchronous algorithm has been proposed for distributed parameter estimation problem in unreliable WSNs with possible sensor equipment and link failures. The novelty of the proposed scheme is that the measurement over the network is embedded in the iterative update of the selected gossiping broadcast nodes. By some theoretical analysis, the investigated algorithm achieves asymptotically to the unknown parameter with probability one. Finally, a numerical example is presented to evaluate the communication and measurement costs required to achieve a given MSE precision, and also verifies the theoretical analysis.

ACKNOWLEDGEMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 61273021 and Grant 61472331, in part by the Natural Science Foundation Project of Chongqing CSTC 2013jjB40008, and in part by Fundamental Research Funds for the Central Universities under Grant XDJK2015C008. This publication was made possible by NPRP grant #4-1162-1-181 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

- [1] S. S. Pereira, R. López-Valcarce, and A. Pagès-Zamora, "A diffusion-based EM algorithm for distributed estimation in unreliable sensor networks," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 595-598, Jun. 2013.

- [2] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: nonlinear observation models and imperfect communication," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575-3605, Jun. 2012.
- [3] Q. Zhang and J.-F. Zhang, "Distributed parameter estimation over unreliable networks with Markovian switching topologies," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2545-2560, Oct. 2012.
- [4] Z. Wang, H. Dong, B. Shen, and H. Gao, "Finite-horizon H-infinity filtering with missing measurements and quantization effects," *IEEE Trans. Autom. Control*, vol. 58, no. 7, pp. 1707-1718, Jul. 2013.
- [5] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, "Randomized gossip algorithms", *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508-2530, 2006.
- [6] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2748-2761, Oct. 2009.
- [7] F. Fagnani and P. Frasca, "Broadcast gossip averaging: interference and unbiasedness in large Abelian Cayley networks," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 4, pp. 866-875, 2011.
- [8] H. Wang, X. Liao, and T. Huang, "Average consensus in sensor networks via broadcast multi-gossip algorithms," *Neurocomputing*, vol. 117, no. 6, pp. 160-170, 2013.
- [9] N. Wang, D. Li, and Z. Yin, "Broadcast gossip algorithm with quantization," in *Proc. 9th IEEE Int. Conf. on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2012, pp. 2143-2147.
- [10] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847-1864, Nov. 2010.
- [11] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564-3575, Oct. 2005.
- [12] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Distributed average consensus with dithered quantization," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4905-4918, Oct. 2008.
- [13] K. Cai, and H. Ishii, "Quantized consensus and averaging on gossip digraphs," *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2087-2100, Sep. 2011.
- [14] J. Lavaei, and R. M. Murray, "Quantized consensus by means of gossip algorithm," *IEEE Trans. Autom. Control*, vol. 57, no. 1, pp. 19-32, Jan. 2012.
- [15] N. E. Leonard and A. Olshevsky, "Cooperative learning in multi-agent systems from intermittent measurements," 2012 [Online]. Available: <http://arxiv.org/abs/1209.2194>
- [16] B. T. Polyak, *Introduction to Optimization*. New York: Optimization Software, 1987.

APPENDIX

In the following, we will investigate the convergence of the distributed parameter estimation algorithm. We first define $v(t)$ as the vector of deviations of the components of $x(t)$ from the unknown vector θ^* at time t , i.e., $v(t) = x(t) - \theta^* \otimes \mathbf{1}_{N(t)}$. Next, we denote the natural measure of how far from convergence by an Euclidean distance measurement $V(t) = \|v(t)\|_2^2$. For any matrix $A \in \mathbb{R}^{N \times N}$, $r_i[A]$ denotes the row sum of the i -th row of $A^T A$, and $R_{[A]} = \text{diag}\{r_1[A], \dots, r_N[A]\}$. The following definition is useful to study the convergence of the algorithm, and the similar statements can be found in [15].

Definition 1: For a nonnegative stochastic matrix $A \in [a_{ij}] \in \mathbb{R}^{N \times N}$, the sieve constant $\kappa[A]$ is defined as

$$\kappa[A] = \min_{m=1, \dots, N} \min_{\|x\|_2=1} x_m^2 + \sum_{k \neq l} a_{kl}(x_k - x_l)^2.$$

Lemma 1: For any asymmetric matrix A ,

$$A^T A = R_{[A]} - \sum_{k < \ell} [A^T A]_{k\ell} (\mathbf{e}_k - \mathbf{e}_\ell)(\mathbf{e}_k - \mathbf{e}_\ell)^T,$$

where \mathbf{e}_k means the k -th basis column vector.

Proof: Since both the matrices $A^T A$ and $(\mathbf{e}_k - \mathbf{e}_\ell)(\mathbf{e}_k - \mathbf{e}_\ell)^T$ are symmetric, hence both sides of the above equation

are symmetric. On one hand, the matrix $(\mathbf{e}_k - \mathbf{e}_\ell)(\mathbf{e}_k - \mathbf{e}_\ell)^T$ has row sums of row, and consequently both sides of the above equation have identical row sums. On the other hand, the entry $[R_{[A]}]_{ij} = 0$ when $i \neq j$, this implies that all the $\{i, j\}$ -entries of both sides with $i \neq j$ are $[A^T A]_{ij}$. Combining the above, it suffices to prove that all the $\{i, j\}$ -entries of both sides with $i = j$ are the same. This completes the proof. ■

Lemma 2: For any asymmetric matrix A ,

$$\|Ay\|_2^2 = \|y\|_2^2 - \sum_{j=1}^N (1 - r_j[A]) y_j^2 - \sum_{k < \ell} [A^T A]_{k\ell} (y_k - y_\ell)^2.$$

Proof: By Lemma 1,

$$\begin{aligned} \|Ay\|_2^2 &= y^T A^T Ay \\ &= y^T R_{[A]} y - \sum_{k < \ell} [A^T A]_{k\ell} y^T (\mathbf{e}_k - \mathbf{e}_\ell)(\mathbf{e}_k - \mathbf{e}_\ell)^T y \\ &= \sum_{j=1}^N r_j[A] y_j^2 - \sum_{k < \ell} [A^T A]_{k\ell} (y_k - y_\ell)^2. \end{aligned}$$

Note that $\|y\|_2^2 = \sum_{j=1}^N y_j^2$, this implies the results of the lemma and completes the proof. ■

Proof of Proposition 1: Subtracting $\theta^* \otimes \mathbf{1}_{N(t)}$ from both side of (3) and noting the fact $[A(t) \otimes \mathbf{I}_d][\theta^* \otimes \mathbf{1}_{N(t)}] = \theta^* \otimes \mathbf{1}_{N(t)}$, we have

$$\begin{aligned} x(t+1) - \theta^* \otimes \mathbf{1}_{N(t)} &= W(t)[x(t) - \theta^* \otimes \mathbf{1}_{N(t)}] + \beta(t)[\Psi(t) + \Phi(t)] \end{aligned}$$

where $W(t) = A(t) \otimes \mathbf{I}_d - \beta(t)[\Xi(t) \otimes \mathbf{I}_d] \mathbf{H}(t) \mathbf{H}^T(t)$. In order to apply Lemma 2, we need to derive the lower bound of $[W^T(t)W(t)]_{k\ell}$ and the upper bound of $r_k[W^T(t)W(t)]$.

If $i \in \mathcal{N}^*(t)$, $j = m(i) \in \mathcal{M}(t)$, then $[W(t)]_{k\ell} = a_{ij}(t) = \beta(t)$ for $k \in [(i-1)d+1, id]$, $\ell = k - (i-j)d$; and $[W(t)]_{k\ell} = 0$ for other $k \neq \ell$. Since $\gamma \geq \max\{4, \max_k [\mathbf{H}(t) \mathbf{H}^T(t)]_{kk}\}$, we know that $[W(t)]_{kk} \geq [1 - \beta(t)] - \beta(t)[\mathbf{H}(t) \mathbf{H}^T(t)]_{kk} \geq 1/2$, which means that $W(t)$ is diagonally dominant. Then, we have

$$\begin{aligned} [W^T(t)W(t)]_{k\ell} &= \sum_{p=1}^{N(t)d} [W(t)]_{pk} [W(t)]_{p\ell} \\ &\geq [W(t)]_{kk} [W(t)]_{k\ell} + [W(t)]_{\ell k} [W(t)]_{\ell\ell} \\ &\geq ([W(t)]_{kk} + [W(t)]_{\ell k})/2 \geq \beta(t)/2, \end{aligned}$$

where the last inequality follows from that $W(t)$ is a non-negative matrix, i.e., $[W(t)]_{\ell k} \geq 0$. Moreover, if k has a measurement of θ^* then the row sum of the k -th row of $W(t)$ is equal to $1 - \beta(t)[\mathbf{H}(t) \mathbf{H}^T(t)]_{kk} \geq 3/4$, which implies that the k -th row sum of $W^T(t)W(t)$ is at most $1 - \beta(t)[\mathbf{H}(t) \mathbf{H}^T(t)]_{kk} \geq 3/4$. We apply Lemma 2 to obtain

$$\begin{aligned} &\| [x(t+1) - \theta^* \otimes \mathbf{1}_{N(t)}] - \beta(t)[\Psi(t) + \Phi(t)] \|_2^2 \\ &\leq \|x(t) - \theta^* \otimes \mathbf{1}_{N(t)}\|_2^2 - \beta(t) \sum_{i \in \mathcal{M}(t)} \|x_i(t) - \theta^*\|_2^2 \\ &\quad - \frac{\beta(t)}{2} \sum_{i \in \mathcal{N}^*(t), j=m(i) \in \mathcal{M}(t), i < j} \|x_i(t) - x_j(t)\|_2^2 \\ &\leq [1 - \beta(t)\kappa[A(t)]/2] \|x(t) - \theta^* \otimes \mathbf{1}_{N(t)}\|_2^2. \end{aligned}$$

Expanding the left hand of the above inequality, after taking expectations and applying the fact $(a+b)^2 \leq 2a^2 + 2b^2$ yields immediately the following statement

$$\begin{aligned} \mathbb{E}\{V(t+1)|x(t)\} &\leq [1 - \beta(t)\kappa_{[A]}^*/2]V(t) \\ &\quad + [\beta(t)]^2[2N_{\max}d\sigma^2 + N_{\max}d\Delta^2/2], \end{aligned}$$

where $\kappa_{[A]}^* = \min_{t=1, \dots, \infty} \{\kappa[A(t)]\}$. For $\beta(t) \propto 1/t^{1-\eta}$, $\eta \in (0, 1/2)$, it is easy to know

$$\sum_{t=1}^{\infty} \beta(t) = \infty, \quad \sum_{t=1}^{\infty} [\beta(t)]^2 < \infty, \quad \text{and} \quad \sup_{t \geq 1} \frac{\beta(t)}{\beta(t+c)} < \infty$$

for any integer c . From Lemma 3 in Section 2.2.1 of [16], we derive that the sequence $\lim_{t \rightarrow \infty} \mathbb{E}\{V(t)\}$ asymptotically approaching zero as $t \rightarrow \infty$. By the Lebesgue dominated convergence theorem, one obtains $\mathbb{E}\{\lim_{t \rightarrow \infty} V(t)\} = 0$. We make use of Lemma 10 in Section 2.2.1 of [16] to complete our proof. ■