Mitigating Quantization Effects on Distributed Sensor Fusion: A Least Squares Approach

Shanying Zhu, *Member, IEEE*, Cailian Chen, *Member, IEEE*, Jinming Xu, Xinping Guan, *Fellow, IEEE*, Lihua Xie, *Fellow, IEEE*, and Karl Henrik Johansson, *Fellow, IEEE*

***Abstract*—In this paper, we consider the problem of sensor fusion over networks with asymmetric links, where the common goal is linear parameter estimation. For the scenario of bandwidth-constrained networks, existing literature shows that nonvanishing errors always occur, which depend on the quantization scheme. To tackle this challenging issue, we introduce the notion of virtual measurements and propose a distributed solution LS-DSFS, which is a combination of a quantized consensus algorithm and the least squares approach. We provide detailed analysis of the LS-DSFS on its performance in terms of unbiasedness and mean square property. Analytical results show that the LS-DSFS is effective in smearing out the quantization errors, and achieving the minimum mean square error (MSE) among the existing centralized and distributed algorithms. Moreover, we characterize its rate of convergence in the mean square sense and that of the mean sequence. More importantly, we find that the LS-DSFS outperforms the centralized approaches within a moderate number of iterations in terms of MSE, and will always consume less energy and achieve more balanced energy expenditure as the number of nodes in the network grows. Simulation results are presented to validate theoretical findings and highlight the improvements over existing algorithms.**

***Index Terms*—Distributed sensor fusion, bandwidth-constrained network, asymmetric links, least squares approach.**

I. Introduction

Recent advances in sensing and wireless technologies have enabled rapid development of distributed sensor networks and their wide applications both in military and civil areas. A key feature of such distributed networks is that high-level tasks, e.g., decentralized inference and tracking and in-network computing, can be accomplished collaboratively by nodes even though each of them is resource limited [1]–[4]. For instance, in a typical sensor fusion problem in a network with n homogeneous nodes, each node makes a snapshot observation of an unknown parameter $\theta \in \mathbb{R}$,[1] which is a noisy version given by

$$y_i = \theta + \omega_i,\ i = 1, 2, \ldots, n, \tag{1}$$

where ω_i are zero mean, i.i.d. random noises with variance σ^2. These data are either transmitted to a fusion center (FC) or shared among nodes for reliable and efficient processing. We know that the ideal sample mean estimate (ISME) $\hat{\theta} \triangleq (1/n)\sum_{i=1}^{n} y_i$ achieves the minimum mean square error (MSE) among all linear estimates, which is the best in the sense of Cramér-Rao lower bound (CRLB) if the noises are Gaussian [6]. It is worth mentioning that $\hat{\theta}$ can be ensured only if all the samples are collected at a FC ideally without any distortion.

One critical aspect of distributed sensor networks is that bandwidth is limited and nodes usually have bounded energy resources. To reduce the bandwidth requirement between nodes, quantization at the node side is one necessary and efficient way. In this case, estimator at the FC can only be formed based on the quantized versions of $\{y_i\}_{i=1}^{n}$. Its performance will thus be degraded with respect to $\hat{\theta}$. For instance, consider the scenario that nodes can transmit the quantized data directly to the FC once per data, i.e., single hop mode of communication. It was shown that the mean square error could be increased by a factor of $(1 + \Delta/(2\sigma))^2$ for the *uniform* quantizer [7], where Δ is the quantization step-size, and a factor of 2 for the *probabilistic* quantizer [8].

As a robust alternative, distributed information processing techniques have received much attention due to their scalability with respect to network size and robustness to node failure, where nodes have access to only local information and perform local computation to achieve the global goal collaboratively [9], [10]. Consensus strategy is one of these techniques that efficiently solve the problem of distributed sensor fusion in the presence of infinite bandwidth. However, when consensus meets quantization, the behavior of the system is too complex to analyze exactly. For instance, in [12], it showed that either a finite-time convergence or a cyclic behavior oscillating around $\hat{\theta}$ was identified for the *truncation* and *uniform* quantizers, which depends on the initial conditions. This phenomenon was also found in [11], where an ADMM strategy was developed to solve the consensus problem using finite-bit bounded

Manuscript received May 17, 2017; revised October 19, 2017 and February 6, 2018; accepted April 10, 2018. Date of publication April 27, 2018; date of current version June 1, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Joao Xavier. This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0901900 and in part by the NSF of China under the Grants 61521063, 61633017, 61603251, 61622307, and U1405251.
(Corresponding author: Shanying Zhu.)

S. Zhu, C. Chen, and X. Guan are with the Department of Automation, Shanghai Jiao Tong University and the Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China (e-mail: shyzhu@sjtu.edu.cn; cailianchen@sjtu.edu.cn; xpguan@sjtu.edu.cn).

J. Xu is with the Ira A. Fulton Schools of Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: xuji0016@e.ntu.edu.sg).

L. Xie is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: elhxie@ntu.edu.sg).

K. H. Johansson is with the ACCESS Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm 10044, Sweden (e-mail: kallej@kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2018.2830304

[1] The algorithm and analysis developed in this paper can be easily extended to deal with the vector case $\boldsymbol{\theta}$ with the linear measurement model $\mathbf{y}_i = \mathbf{C}_i\boldsymbol{\theta} + \boldsymbol{\omega}_i$. The tweak is to introduce a second consensus algorithm as in [5].

quantizer with possibly unbounded data. Ref. [13] showed that the states achieve consensus in finite time with the *truncation* quantizer, but a non-vanishing error between the ideal $\hat{\theta}$ and the final state exists, which is a function of the quantization resolution. For a general bounded quantizer, similar behaviors were also observed in [15], where a stochastic approximation method was used to ensure its asymptotic convergence. In [14], the *probabilistic* quantizer was adopted to ensure almost sure consensus at a random quantization level for fixed undirected graphs, whose expectation equals $\hat{\theta}$. However, the deviation of the consensus value from $\hat{\theta}$ is not tightly bounded. Carli *et al.* in [16] elaborated on the impact of updating rules on the performance of gossip algorithms for the *uniform* and *probabilistic* quantizers, revealing that convergence can only be guaranteed up to a neighborhood around the ideal $\hat{\theta}$. Similar phenomena were also observed for the *logarithmic* quantizer [17] and the additive quantization model [18], where the errors between the state and $\hat{\theta}$ are upper bounded by quantities depending on the quantization resolution and initial conditions. A workaround was proposed in [19], where it added dither to the nodes' data before quantization and adopted decaying link weights satisfying the persistence condition to suppress the quantization noises. The mean square error can be made arbitrarily small by tuning these weights, which, however, would significantly slow down the convergence of the algorithm. We emphasize that all the above works can produce estimates with acceptable accuracy only for rather high resolutions. Another way to solve the problem of distributed sensor fusion is to formulate it within the consensus+innovation framework as in [20], which mimics algorithms for continuous observations by combining two time scales in one step as in [21], [22]. Convergence to $\hat{\theta}$ can be established in the case of infinite bandwidth by using techniques from stochastic approximation. However, such formulation cannot handle the situation with quantized transmissions.

In summary, previous work (both centralized and distributed approaches) always assumed that the achievable precision would have been in the order of the quantization step. In [23], [24], a complicated *dynamic* quantizer was proposed to show that the ideal $\hat{\theta}$ can be achieved asymptotically. Of particular note is that each node needs n decoders at the worst case, which is obviously resource prohibited. Furthermore, they assume bidirectional communications between nodes for theoretical simplicity. This is quite rare in practice since nodes typically broadcast at different power levels and have different interference and noise patterns. We emphasize that all these algorithms will fail when applied to networks with asymmetric links. In particular, it was shown in [25] that if the states are integer-valued then nodes can only achieve an integer approximation of $\hat{\theta}$ for *uniform* quantizer over gossip asymmetric networks. Even if the *dynamic* quantizer [23] was adopted, there is no way to eliminate the bias introduced into $\hat{\theta}$, which is intrinsic imposed by the network topology [26]. In our previous work [27], we proposed a two-stage distributed algorithm to tackle the directed nature of communication links coupled with quantization residues, which can achieve $\hat{\theta}$ both in the mean square and almost sure senses. Extension to additive quantization model can be found in [28].

In this paper, we aim to address the above challenging issues. Specifically, we give a positive answer to the following question: given a certain quantization, how precise can the sensor fusion be? We gave a partial answer to this question in the previously published conference paper [29]. The main contributions of the paper are summarized as follows:

- Firstly, we characterize the quantization error as additive noise and exploit its temporal properties to propose a baseline distributed quantized algorithm to accommodate the constraints imposed by asymmetric links. We then introduce the notion of virtual measurements to describe the impact of quantization noise. By fusing all the virtual measurements, we propose the distributed solution, termed as LS-DSFS, which is a combination of the quantized consensus algorithm and the least squares approach. Numerically, LS-DSFS is shown to have the ability of mitigating the effect of quantization and achieve the minimum MSE among prevailing centralized and distributed algorithms, e.g., [7], [8], [12], [14], [16], [17], [24] (see Section III and Table I). More importantly, we find that LS-DSFS will always consume less energy and achieve more balanced energy expenditure than the centralized approaches as the number of nodes in the network grows.
- Secondly, we give a detailed analysis of the proposed LS-DSFS for generic conditions. Convergence results of the mean sequence as well as its mean square property are provided. Furthermore, we present the rate at which LS-DSFS converges to quantify its convergence time theoretically. In particular, for the special case of running average estimate, we can establish its rate of almost sure convergence by resorting to a law of iterated logarithm for independent random vectors. The involved arguments adopt decaying step-size as is common in the literature to suppress the propagation of noise over networks. Unlike those in [19], [21], [30] demanding the square-summable property, we only require them to be non-summable (see Sections III and IV).

The structure of the paper is as follows: Section II introduces the network model and our distributed quantized algorithm for asymmetric networks. In Section III, we describe the proposed LS-DSFS and summarize the main theoretical results. Section IV is devoted to the detailed performance analysis of LS-DSFS. Simulation results evaluating the performance of LS-DSFS and validating the theoretical findings are provided in Section VI, followed by the conclusions in Section VII.

II. Problem Formulation

In this section, we present the network model and the baseline distributed algorithm that are adopted in subsequent analysis.

Network model: Consider a sensor network where nodes are linked with their one-hop neighbors via asymmetric links. We model the communication network over which nodes exchange data as a *directed graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ is the set of nodes, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ denotes all the asymmetric links between nodes. Each edge $(j, i) \in \mathcal{E}$ represents an unidirectional link meaning that node i can receive data from node j. In order to avoid any isolated nodes, we assume that graph $\mathcal{G}$ is strongly connected, i.e., each node i can reach any other node via a directed path (possibly multi-hop).

We define $\mathbf{A} = [a_{ij}]_{n\times n}$ as the 0-1 adjacency matrix, where $a_{ij} = 1 \Leftrightarrow (j, i) \in \mathcal{E}$. Let $\mathcal{N}_i^+ \triangleq \{j : a_{ij} = 1\}$ and $\mathcal{N}_i^- \triangleq \{j : a_{ji} = 1\}$ as the *in-neighbors* and *out-neighbors* of node i, respectively. Accordingly, $d_i^+ = \sum_{j=1}^n a_{ij}$ and $d_i^- = \sum_{i=1}^n a_{ji}$ are called its *in-degree* and *out-degree*, respectively.

Quantization scheme: Due to bandwidth limitations, each node needs to quantize its data using a quantizer $q(\cdot): \mathbb{R} \to \mathcal{S} = \{k\Delta, k \in \mathbb{Z}\}$ before transmitting to its neighbors at each step. Widely used quantizers in the literature includes the uniform, truncation, rounding and probabilistic quantizers, etc [8], [12], [14], [16]. We represent the quantized message of data x_i at node i as

$$q(x_i) = x_i + w_i \text{ with } |w_i| \le \Delta, \forall i = 1, 2, \dots, n, \tag{2}$$

where w_i is the quantization error, and Δ is the quantization step-size. In this paper, we propose to characterize w_i as an additive quantization noise with zero mean and bounded variance. One example of the schemes possessing these properties is the probabilistic quantizer [8], [14], [16], [19], [27], [31]

$$q(x) = \begin{cases} \left\lceil \frac{x}{\Delta} \right\rceil \Delta, & \text{with probability } p_{x,\Delta}, \\ \left\lfloor \frac{x}{\Delta} \right\rfloor \Delta, & \text{with probability } 1 - p_{x,\Delta}, \end{cases} \tag{3}$$

where $p_{x,\Delta} = x/\Delta - \lfloor x/\Delta \rfloor$, $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ denote the floor and ceiling functions, respectively. In fact, it can be shown that $\mathbb{E}\{w_i\} = 0$ and $\mathbb{E}\left\{w_i^2\right\} \le \Delta^2/4$.

Distributed quantized algorithm: In order to tackle the effect of quantization coupled with asymmetric links, we propose to use the following algorithm as the basis of our distributed sensor fusion algorithm.

1) *Initialization:* Each node i randomly take an initial value $x_i(0)$, and set $s_i(0) = 0$, $\forall i = 1, 2, \dots, n$.

2) *Updates:* Each node i updates its states upon receiving data from its in-neighbors,

$$x_i(t+1) = x_i(t) - \alpha d_i^+ q(x_i(t)) + \alpha \sum_{j \in \mathcal{N}_i^+} q(x_j(t)) + \beta s_i(t), \tag{4}$$

$$s_i(t+1) = (1-\beta) s_i(t) - \alpha \left[d_i^- q(s_i(t)) - d_i^+ q(x_i(t)) \right] + \alpha \sum_{j \in \mathcal{N}_i^+} \left[q(s_j(t)) - q(x_j(t)) \right], \tag{5}$$

where $\alpha < 1/\max_i d_i^+$ and $\beta > 0$ is a tuning parameter.

In the algorithm, each node i keeps track of the state $x_i(t)$ along with $s_i(t)$ to locally record the state changes. Of special note is the companion variable $s_i(t)$ in (4), which is incorporated to compensate for the unidirectional effects of communication links. This is motivated by [32]. Besides the companion variable, we use both the exact and quantized information about their own states, namely, $x_i(t), s_i(t)$ and $q(x_i(t)), q(s_i(t))$, in (4) and (5). To give an intuitive illustration behind these endeavors, we note that quantized consensus algorithm preserve the initial average, i.e., $\sum_{i=1}^n (x_i(t) + s_i(t)) = \sum_{i=1}^n x_i(0)$, $\forall t > 0$, which is essential for standard average consensus algorithms. Further, we remark that the information of out-degree d_i^+ is necessary for distributed information processing over asymmetric networks [33].

It is worthy to note that such endeavors only work to some extent in that the quantization operation $q(\cdot)$, on the other hand, prevents the states of (4) and (5) from converging. We simulate the algorithm in a simple ring network with $n = 4$ nodes, see Fig. 1. For one specific realization, it can be seen that fluctuating behaviors of $x_i(t)$, $\forall i = 1, 2, \dots, n$ are observed with non-vanishing errors. This demonstrates that simply taking the states $\{x_i(t)\}_{i=1}^n$ of (4) as estimates of θ does not work in the presence of quantized transmissions.

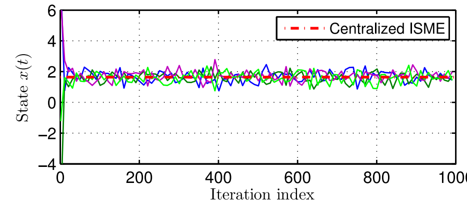


Fig. 1. States of (4) over a directed ring graph with $n = 4$ and $\Delta = 1$.

Our goal in this paper are the following: (i) *design* an appropriate form of estimator to mitigate the quantization effects in (4) and (5) such that the ideal centralized estimate $\hat{\theta}$ can be achieved at each node; and, (ii) *provide* rigorous convergence analysis and *characterize* its rate of convergence. We achieve this by adopting a least squares approach. With this approach, the randomness incurred by quantization errors will be mitigated gradually and finally smeared out.

To facilitate the ensuing analysis, we will be using the following assumptions:

(a-1) The initial values $x_i(0), i = 1, 2, \dots, n$ are random variables with bounded variance (not necessarily independent of each other).

(a-2) The quantization error is temporally uncorrelated with zero mean and auto-covariance matrix $\mathbb{E}\{\mathbf{w}(t)\mathbf{w}(t)^T\} = \mathbf{W}_t$, and is also uncorrelated with the input messages at each step and node.[2]

III. Mitigating the Quantization Effect: A Least Squares Formulation

In this section, we will propose a novel least squares approach to smear out the quantization errors collected from the iterative process (4) and (5), based on which our distributed sensor fusion scheme is proposed.

Our main idea is motivated by the following result.

Lemma 1: Consider the distributed quantized consensus algorithm defined by (4) and (5) under the quantization scheme (2), and assume that $\mathcal{G}$ is strongly connected. Then, under Assumption (a-2), $x_i(t)$ converges in expected value to the average $\bar{x}(0) \triangleq \mathbf{1}^T \mathbf{x}(0)/n$, $\forall i$, for sufficiently small $\beta > 0$.

Proof: For each node $i = 1, 2, \dots, n$, write the quantized message of $x_i(t)$ and $s_i(t)$ as $q(x_i(t)) = x_i(t) + u_i(t)$ and $q(s_i(t)) = s_i(t) + v_i(t)$, where $u_i(t), v_i(t)$ represent the quantization errors, respectively. Stack $x_i(t)$, $s_i(t)$, $u_i(t)$ and $v_i(t)$ into column vectors $\mathbf{x}(t)$, $\mathbf{s}(t)$, $\mathbf{v}(t)$, respectively, and let $\mathbf{z}(t) = [\mathbf{x}(t)^T, \mathbf{s}(t)^T]^T$, $\mathbf{w}(t) = [\mathbf{u}(t)^T, \mathbf{v}(t)^T]^T$, we can express (4) and (5) in a compact form

$$\mathbf{z}(t+1) = \mathbf{P}\mathbf{z}(t) + \alpha \mathbf{L}_{\text{aug}} \mathbf{w}(t) \tag{6}$$

where

$$\mathbf{P} \triangleq \begin{bmatrix} \mathbf{I} - \alpha \mathbf{L} & \beta \mathbf{I} \\ \alpha \mathbf{L} & (1-\beta)\mathbf{I} - \alpha \mathbf{L}^- \end{bmatrix}, \mathbf{L}_{\text{aug}} \triangleq \begin{bmatrix} -\mathbf{L} & \mathbf{0} \\ \mathbf{L} & -\mathbf{L}^- \end{bmatrix},$$

[2] In [34], Sripad *et al.* established sufficient conditions to enable such assumption for scalar quantization noise. Particularly, if the joint characteristic function of the input message is band-limited with the upper bound $2\pi/\Delta$, then Assumption (a-2) is satisfied. In practice, the characteristic functions are not exactly band-limited and the above assumption holds valid in an approximate sense. Under sufficiently large quantization rates, however, such an assumption closely reflects the actual system behavior. On the other hand, introducing dithering to the system can also lead to independent quantization errors, if the Schuchman conditions are satisfied [14], [19], [35].

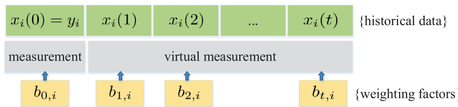


Fig. 2. A schematic view of the least squares formulation for the case $x_i(0) = y_i, \forall i = 1, 2, \ldots, n$.

$\mathbf{L} \triangleq \text{diag}\{d_1^+, \ldots, d_n^+\} - \mathbf{A}$ and $\mathbf{L}^- \triangleq \text{diag}\{d_1^-, \ldots, d_n^-\} - \mathbf{A}$. By Assumption (a-2), we know that $\mathbb{E}\{\mathbf{u}(t)\} = \mathbb{E}\{\mathbf{v}(t)\} = \mathbf{0}$. One thus obtains $\mathbb{E}\{\mathbf{z}(t+1)\} = \mathbf{P}\mathbb{E}\{\mathbf{z}(t)\}$, where the expectation $\mathbb{E}\{\mathbf{z}(t)\}$ is taken over the initial condition $\mathbf{x}(0)$ and quantization noises $\{\mathbf{w}(\tau)\}_{\tau=0}^{t-1}$. Applying [32, Theorem 4] to the above recursion shows that for small $\beta > 0$, $\lim_{t\to\infty} \mathbb{E}\{\mathbf{x}(t) - \bar{x}(0)\mathbf{1}\} = \mathbf{0}$, and $\lim_{t\to\infty} \mathbb{E}\{\mathbf{s}(t)\} = \mathbf{0}$, from which the lemma follows. ■

Lemma 1 states that $x_i(t)$ at each node i does converge to $\mathbb{E}\{\bar{x}(0)\}$ in mean, although its realizations have a fluctuating behavior as observed in Fig. 1. This motivates us to express the state $x_i(t)$ as

$$x_i(t) = \mathbb{E}\{\bar{x}(0)\} + noise_i(t), \ \forall i = 1, 2, \ldots, n, \quad (7)$$

where $noise_i(t)$ is the error capturing the fluctuation of $x_i(t)$ from $\mathbb{E}\{\bar{x}(0)\}$. In this way, we may regard $x_i(t)$ as a *virtual measurement* of $\mathbb{E}\{\bar{x}(0)\}$ corrupted by $noise_i(t)$ for node i.

Our objective is then to find an optimal estimator of $\mathbb{E}\{\bar{x}(0)\}$ for each node i by fusing the virtual measurement $\{x_i(t)\}_{i=1}^n$. It is noted, however, that the expression (7) makes the precise statistical characterization of $noise_i(t)$ hard to obtain except the fact that asymptotically $\mathbb{E}\{noise_i(t)\}$ tends to 0 as $t \to \infty$. A natural idea is to adopt the least squares approach, which is effective to handle the situation where the distributional knowledge of $noise_i(t)$ is unavailable [6]. Although no optimality can be ensured, the least squares approach is widely used in practice due to its ease of implementation.

A. The Algorithm: LS-DSFS

Each node i stores the virtual measurements $\{x_i(0), x_i(1), x_i(2), \ldots, x_i(t)\}$ up to time t, and does the batch least squares estimation by minimizing the cost function

$$J_i(t) = \min_{\hat{x}\in\mathbb{R}} \sum_{s=0}^{t} b_{s,i}(x_i(s) - \hat{x})^2, \quad (8)$$

where $\{b_{s,i}\}_{s=0}^t$ are weighting factors that emphasize the contributions of the historical data $\{x_i(s)\}_{s=0}^t$ (see Fig. 2). As a standard procedure, the above batch processing can be sequentially done in the following way

$$\hat{x}_i(t+1) = \hat{x}_i(t) + \gamma_{t,i}\big[x_i(t+1) - \hat{x}_i(t)\big], \quad (9)$$

where $\gamma_{t,i} > 0$ is the gain factor determined by $\{b_{s,i}\}_{s=0}^t$.

In the above framework of least squares formulation, the post-processed $\hat{x}_i(t)$ of (9), rather than the state $x_i(t)$ of (4), is taken as the estimate of $\mathbb{E}\{\bar{x}(0)\}$ at node i. We summarize the proposed Least Squares based Distributed Sensor Fusion Scheme (LS-DSFS) for asymmetric bandwidth-constrained sensor networks in Algorithm 1, where its t-th iteration run by node i is presented.

We discuss the communication cost of LS-DSFS. At each iteration t, each node i transmits $x_i(t)$ and $s_i(t)$ to its neighbors, resulting in $2|\mathcal{E}|$ total number of such transmissions across the entire network. As for the memory burden, each node only needs to store 3 quantities at each iteration, namely, $x_i(t)$, $s_i(t)$ and $\hat{x}_i(t)$. As such, LS-DSFS is scalable with respect to network size.

Algorithm 1: LS-DSFS at node i.

Input: α, β, d_i^-, $\gamma_{t,i}$.
Output: $\hat{x}_i$.
1: **Initialization:** arbitrary $x_i(0)$, $s_i(0) = 0$, and $\hat{x}_i(0) = x_i(0)$.
2: Quantize the data $\{x_i(t), s_i(t)\}$ using the probabilistic quantizer (3).
3: Receive data from the in-neighbors: $q(x_j(t)), q(s_j(t))$, $j \in \mathcal{N}_i^+$.
4: Run the quantized consensus algorithm (4) and (5) to update the intermediate states $\{x_i(t), s_i(t)\}$.
5: Update the least squares estimate $\hat{x}_i(t)$ by processing the virtual measurements $\{x_i(s)\}_{s=0}^t$ sequentially via (9).

We also remark that particularly by letting its measurement y_i in (1) as the initial guess of the unknown θ for each node i, i.e., $x_i(0) = y_i$, $\forall i = 1, 2, \ldots, n$, we then have $\hat{\theta} = \bar{x}(0)$. In this way, the aforementioned distributed sensor fusion problem can be exactly solved by LS-DSFS.

B. Main Results

Our first result demonstrates the correctness of the least squares method in smearing out the quantization noises for arbitrary gain factors $\{\gamma_{t,i}\}_{t\geq 0, 1\leq i\leq n}$ so long as they are non-summable and the relative maximum discrepancy is bounded.

Theorem 1: Consider LS-DSFS in Algorithm 1 and assume that $\mathcal{G}$ is strongly connected. Suppose the following conditions hold:

(c-1) The sequence of gain factors $\{\gamma_{t,i}\}_{t\geq 0, 1\leq i\leq n}$ are in the interval $(0, 1]$ satisfying $\sum_{t=0}^{\infty} \gamma_{t,i} = \infty, \forall i$;

(c-2) There are two positive sequences $\{\delta_t\}_{t\geq 0}$ and $\{\Gamma_t\}_{t\geq 0}$ such that $\delta_t \leq \gamma_{t,i} \leq \Gamma_t$ and $\Gamma_t \leq \kappa\delta_t, \forall i, \forall t$, for some constant $\kappa \geq 1$;

then under Assumptions (a-1) and (a-2), for sufficiently small $\beta > 0$,

$$\lim_{t\to\infty} \mathbb{E}\{\hat{\mathbf{x}}(t) - \bar{x}(0)\mathbf{1}\} = 0,$$

where the expectation is taken over the initial condition $\mathbf{x}(0)$ and quantization noises $\{\mathbf{w}(\tau)\}_{\tau=0}^{t-1}$. If further,

(c-3) $\lim_{t\to\infty} \gamma_{t,i} = 0, \forall i$, then LS-DSFS also has the following mean square property:

$$\lim_{t\to\infty} \mathbb{E}\{\|\hat{\mathbf{x}}(t) - \bar{x}(0)\mathbf{1}\|^2\} = 0.$$

It is worth noting that the requirements (c-1) and (c-2) imposed on the gain factors $\{\gamma_{t,i}\}$ embrace the constant case $\gamma_{t,i} \equiv c_i \in (0, 1], \forall t \geq 0$, as a special case. Moreover, we remark that both (c-1) and (c-3) are necessary to ensure the unbiasedness and mean square properties in Theorem 1.

- *Necessity of (c-1):* We use a contradiction argument. If $\sum_{t=0}^{\infty} \gamma_{t,i_*} < \infty$, for some i_*, let $e_{i_*}^x(t) \triangleq x_{i_*}(t) - \bar{x}(0)$ and $\hat{e}_{i_*}^x(t) \triangleq \hat{x}_{i_*}(t) - \bar{x}(0)$, it then follows from (9) that $|\mathbb{E}\{\hat{e}_{i_*}^x(t+1)\}| \leq |\mathbb{E}\{\hat{e}_{i_*}^x(0)\}| + \sum_{t=1}^{\infty} \gamma_{t-1,i_*} |\mathbb{E}\{e_{i_*}^x(t)\}|$, which together with Lemma 1 implies that both $|\mathbb{E}\{e_{i_*}^x(t)\}|$ and $|\mathbb{E}\{\hat{e}_{i_*}^x(t)\}|$ are bounded.

On the other hand, one can obtain

$$\begin{aligned}
&|\mathbb{E}\{\hat{e}^x_{i_*}(t+1)\}| \\
&\geq |\mathbb{E}\{\hat{e}^x_{i_*}(0)\}| - \sum_{t=0}^{\infty} \gamma_{t,i_*}\big[|\mathbb{E}\{\hat{e}^x_{i_*}(t)\}| + |\mathbb{E}\{e^x_{i_*}(t+1)\}|\big] \\
&= (1-\gamma_{0,i_*})|\mathbb{E}\{\hat{e}^x_{i_*}(0)\}| \\
&\quad - \sum_{t=1}^{\infty}\big[\gamma_{t,i_*}|\mathbb{E}\{\hat{e}^x_{i_*}(t)\}| + \gamma_{t-1,i_*}|\mathbb{E}\{e^x_{i_*}(t)\}|\big] > 0,
\end{aligned}$$

if the initial guess $\mathbb{E}\{\hat{x}_{i_*}(0)\}$ is far from reliable. Hence, $\mathbb{E}\{\hat{x}_{i_*}(t)\}$ will never converge to $\mathbb{E}\{\bar{x}(0)\}$ in this case. And so does $\mathbb{E}\{(\hat{x}_{i_*}(t)-\bar{x}(0))^2\}$ by recalling the fact that $|\mathbb{E}\{\hat{e}^x_i(t)\}|^2 \leq \mathbb{E}\{|\hat{e}^x_i(t)|^2\}$, $\forall t, i$, in view of the Hölder inequality [36, p.129]. This shows the necessity of (c-1).

- *Necessity of (c-3):* Suppose that $\liminf_{t\to\infty}\gamma_{t,i} = \gamma^* > 0$, then by the Hölder inequality, we derive from (9) that $\mathbb{E}\{\hat{e}^x_i(t+1)^2\} \geq \big(\gamma_{t,i}\sqrt{\mathbb{E}\{e^x_i(t+1)^2\}} - (1-\gamma_{t,i})\sqrt{\mathbb{E}\{\hat{e}^x_i(t)^2\}}\big)^2$. Using the properties of $\limsup$ and $\liminf$, in particular, $\limsup_{t\to\infty}(-\gamma_{t,i}) = -\liminf_{t\to\infty}\gamma_{t,i}$ and
$$\limsup_{t\to\infty}\sqrt{\mathbb{E}\{e^x_i(t)^2\}} = \sqrt{\limsup_{t\to\infty}\mathbb{E}\{e^x_i(t)^2\}},$$
the above relation yields
$$\gamma^*\sqrt{\limsup_{t\to\infty}\mathbb{E}\{e^x_i(t)^2\}} \leq (2-\gamma^*)\sqrt{\limsup_{t\to\infty}\mathbb{E}\{\hat{e}^x_i(t)^2\}}.$$
In general, we have $\limsup_{t\to\infty}\mathbb{E}\{e^x_i(t)^2\} > 0$ by taking Lemma 5 into consideration (see also Fig. 1 for an illustration). As a result, it is necessary to use decaying gain factors satisfying (c-3) in order to have the desired mean square performance of LS-DSFS.

The choice of the parameter β depends on the network structure and the number of nodes. Such a requirement when bounding some parameters in consensus algorithms can also be found in [10], [23]. A theoretical upper bound of β is established in [32] using matrix perturbation theory. This upper bound, however, is quite conservative, observed from the simulation results in Section VI.

Our second result presents the convergence rate result of LS-DSFS by establishing the rate at which $\hat{\mathbf{x}}(t)$ converges to $\bar{x}(0)\mathbf{1}$ in mean and mean squares senses. The theorem formally makes use of the matrix $\mathbf{Q} \triangleq \mathbf{P} - \mathbf{P}_\infty$ that captures the coefficient matrix of the error dynamics $[\mathbf{x}(t)^T, \mathbf{s}(t)^T]^T - [\bar{x}(0)\mathbf{1}^T, \mathbf{0}]^T$, where $\mathbf{P}_\infty \triangleq \left[\begin{smallmatrix}\mathbf{1}\mathbf{1}^T/n & \mathbf{1}\mathbf{1}^T/n \\ \mathbf{0} & \mathbf{0}\end{smallmatrix}\right]$.

Theorem 2: Suppose that all the assumptions of Theorem 1 hold. Then we have

$$\|\mathbb{E}\{\hat{\mathbf{x}}(t) - \bar{x}(0)\mathbf{1}\}\| \leq \begin{cases} \mathcal{O}\left(t^n \rho^t_{\mathbf{Q}}\right), & 1-\rho_{\mathbf{Q}} < \inf_t \delta_t \leq 1, \\ \mathcal{O}\left(e^{-\sum_{s=0}^{t-1}\delta_s}\right), & 0 < \sup_t \delta_t < 1-\rho_{\mathbf{Q}}, \\ \mathcal{O}\left(t^n e^{-\sum_{s=0}^{t-1}\delta_s}\right), & 0 < \inf_t \delta_t \leq \sup_t \delta_t = 1-\rho_{\mathbf{Q}}, \end{cases}$$

where $\rho_{\mathbf{Q}}$ is the spectral radius of $\mathbf{Q}$. Let $\gamma_{t,i} = a_i/(t+1)^{\tau_i}$, where $a_i > 0, 0 \leq \tau_i \leq 1$, $\forall t, i$, with $\max_i \tau_i < 2\min_i \tau_i < \min_i a_i + 1$, then for small $\beta > 0$,

$$\mathbb{E}\{\|\hat{\mathbf{x}}(t) - \bar{x}(0)\mathbf{1}\|^2\} \leq \mathcal{O}\left(1/t^{2\underline{\tau}-\bar{\tau}}\right),$$

where $\underline{\tau} = \min_i \tau_i$ and $\bar{\tau} = \max_i \tau_i$.

We remark that for the constant gain factor $\gamma_{t,i} \equiv c_i$, $\forall t$, Theorem 2 provides a complete characterization of the rate of convergence of $\mathbb{E}\{\hat{\mathbf{x}}(t)\}$. Noting, further, that $\lim_{t\to\infty} t^n\rho^t_{\mathbf{Q}}/r^t = 0$, for any constant $r > \rho_{\mathbf{Q}}$, it actually shows the exponential convergence of $\mathbb{E}\{\hat{\mathbf{x}}(t) - \bar{x}(0)\mathbf{1}\}$ in this case. As for the mean square analysis, we take the gain factor which decays as $\gamma_{t,i} \sim t^{-\tau_i}$, $\forall i$, in order to get a convergence rate result. Such choice is popular in stochastic approximation methods [30]. Theorem 2 then implies that for such kinds of gain factors LS-DSFS converges at a rate of $\mathcal{O}(1/t^{2\underline{\tau}-\bar{\tau}})$. Moreover, it is not surprising that the value of the rate coefficient depends on the network structure and the quantization scheme, whose explicit expression will be given in Section IV.

C. Comparison With Centralized Approaches

In this section, we will illustrate the efficiency of the least squares method in solving the sensor fusion problem by comparing LS-DSFS with centralized approaches. To this end, let $x_i(0)$ be the measurement y_i in (1) for each node, i.e., $x_i(0) = y_i$, $\forall i$. Then one has

$$\mathbb{E}\{\bar{x}(0)\} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{y_i\} = \theta,$$

since ω_i, $i = 1, 2, \ldots, n$, are zero mean. Then, under Assumption 2, we apply Theorem 1 to conclude that $\hat{\mathbf{x}}(t)$ of LS-DSFS is asymptotically unbiased with the rate of convergence

$$\|\mathbb{E}\{\hat{\mathbf{x}}(t)\} - \theta\mathbf{1}\| \leq \mathcal{O}\left(e^{-\sum_{s=0}^{t-1}\delta_s}\right).$$

For comparison, we choose the prevailing quantized sample mean estimate (QSME) [7], the decentralized estimation scheme (DES) [8] and the ISME as the centralized benchmarks. The performances in terms of mean square error MSE $= \mathbb{E}\left\{(\text{estimate} - \theta)^2\right\}$ are provided in Table I. As noted, there are always degradation in the performance of QSME and DES with respect to ISME $\hat{\theta}$ due to quantization, and the level of degradation generally depends on the magnitude of quantization step-size Δ and measurement noise variance σ^2.

Now consider the performance of the proposed LS-DSFS, for each i, denote $\text{MSE}_i(\theta) = \mathbb{E}\{(\hat{x}_i(t)-\theta)^2\}$, where the expectation is taken over the measurement noise $\omega_i, 1 \leq i \leq n$ and quantization noises $\{w_i(t), 1 \leq i \leq n\}^t_{\tau=0}$, and $\bar{\omega} = (1/n)\sum_{i=1}^n \omega_i$, respectively. We then obtain from (1) that

$$\begin{aligned}
\text{MSE}_i(\hat{\theta}) &= \text{MSE}_i(\theta) + \mathbb{E}\left\{\bar{\omega}^2\right\} - 2\mathbb{E}\left\{(\hat{x}_i(t)-\theta)\bar{\omega}\right\} \\
&\overset{(a)}{\geq} \text{MSE}_i(\theta) + \mathbb{E}\left\{\bar{\omega}^2\right\} - 2\sqrt{\text{MSE}_i(\theta)\mathbb{E}\left\{\bar{\omega}^2\right\}} \\
&\overset{(b)}{=} \text{MSE}_i(\theta) + \frac{\sigma^2}{n} - \frac{2\sigma}{\sqrt{n}}\sqrt{\text{MSE}_i(\theta)},
\end{aligned}$$

where (a) follows from the Hölder inequality [36, p.129], and (b) is due to the facts that $\{\omega_i\}_{i=1}^n$ are i.i.d. and $\mathbb{E}\left\{\bar{\omega}^2\right\} = 1/n^2\sum_{i=1}^n\mathbb{E}\{\omega_i^2\} = \sigma^2/n$. It thus follows that $\left(\sqrt{\text{MSE}_i(\theta)} - \sigma/\sqrt{n}\right)^2 \leq \text{MSE}_i(\hat{\theta})$. Applying Theorem 1 to the above relation yields $\lim_{t\to\infty}\text{MSE}_i(\theta) = \frac{\sigma^2}{n}$, $\forall i$. This means that the proposed LS-DSFS eventually achieves the minimum MSE, demonstrating the effectiveness of the least squares formulation (9) in mitigating the quantization effects.

TABLE I
PERFORMANCE COMPARISON BETWEEN LS-DSFS AND THE CENTRALIZED APPROACHES

Algorithm	QSME [7]	DES [8]	M-QSME	ISME	LS-DSFS
Unbiasedness	NO	YES	NO	YES	Asymptotically YES
MSE	$\leq (1+\frac{\Delta}{2\sigma})^2\frac{\sigma^2}{n}$	$\leq \frac{2\sigma^2}{n}$	$\to \frac{\sigma^2}{n}$	$\frac{\sigma^2}{n}$	$\to \frac{\sigma^2}{n}, \forall i$
Energy gap		$\mu r^m(K^m-1)$			$\leq (\max_i \lvert\mathcal{N}_i\rvert - \min_i \lvert\mathcal{N}_i\rvert)l + \mu r^m$
Avg. energy consumption		$> \mathcal{O}(n^{m/2})$			$< \mathcal{O}(n)$

Comparison results are summarized in Table I. The statement may seem puzzling at first, since one may think that distributed algorithms such as LS-DSFS relying solely on localized interactions could impossibly outperform the centralized counterpart which has a FC having access to the global information. In the case of constrained bandwidth, it is, however, reasonable. For the centralized approaches, namely, QSME and DES, the FC can only acquire the quantized data from the nodes once per data with a single hop mode. Performance degradation with respect to ISME $\hat{\theta}$ is clearly observed in Table I. The situation becomes even worse if the data has to travel through multiple hops until reaching the FC, since several rounds of quantization will be done on $\{y_i\}_{i=1}^n$, making it harder to recover at the FC side. Specifically, Ref. [37] derived an upper bound of MSE for the multi-hop mode, which tends to be constant independent of the number of nodes n. This is much different than QSME and DES in single hop mode. If multiple rounds of transmissions from nodes to the FC are allowed in the single hop mode (e.g., the multiple-round version of QSME, denoted as M-QSME), then it is possible for the FC to reach ISME $\hat{\theta}$ to arbitrary precision.[3] While for LS-DSFS, it is noted that each node keeps one perfect copy of y_i locally along with the quantized version $q(y_i)$. By exploiting this hidden information and the temporal properties of quantization noises appropriately, the nodes can collaboratively achieve better performance than QSME and DES. This is what we have done in LS-DSFS by introducing both quantized $\{q(x_i(t)), q(s_i(t))\}$ and unquantized $\{x_i(t), s_i(t)\}$ in line 4 and the least squares step in line 5 of Algorithm 1. One interesting property of the distributed LS-DSFS is that it can achieve the improved MSE performance level of the centralized M-QSME, by relying solely on interactions with neighboring nodes without the need for a FC as in M-QSME.

On the other hand, the centralized approaches, i.e., QSME, DES and M-QSME, are prone to the problem of uneven energy depletion, which can drastically reduce the network lifetime. In the case of a single hop network, nodes located farthest from the FC have to consume the maximum amount of energy. Whereas for the multiple hop mode, nodes closer to the FC need to relay more traffic and suffer much faster energy consumption, which is apt to cause the "energy hole" problem [38]. Such phenomenon has also been demonstrated previously in [39]. Consider a sensor network with n nodes uniformly distributed over a circular region and the FC is located at the center. For simplicity, we divide the circular region into K concentric rings of thickness r. Here r is the communication radius of all the nodes. Let E_i be the energy expenditure of a node in the i-th ring. The energy gap is thus $\max_i E_i - \min_i E_i$, which measures how uneven the energy depletion among all nodes is. It can be shown that the energy gaps are $EG_{single} = \mu r^m(K^m - 1)$ for the single hop mode and $EG_{multiple} = (2l + \mu r^m)(K^2 - 1)$ for the multi-hop mode, respectively, where l, μ, and m are energy-related parameters [40]. Note that, typically, $2 \leq m \leq 6$, and $K \gg 1$ for large scale networks. This means that EG_{single} and $EG_{multiple}$ grow at least with an order of $\mathcal{O}(K^2)$. While for the proposed LS-DSFS, only local communication with neighboring nodes is permitted, its energy gap is smaller than $(\max_i |\mathcal{N}_i| - \min_i |\mathcal{N}_i|)l + \mu r^m$, where $|\mathcal{N}_i|$ is the cardinality of $\mathcal{N}_i$. Hence, the energy expenditure is more balanced.

Let us have a closer look at the total average energy EA spent during one round of transmission for the centralized QSME, DES and M-QSME, and the distributed LS-DSFS. For the centralized approaches, letting n_i be the number of nodes in the i-th ring, and $d_{j0,i}$ be the distance between node j in the i-th ring and the FC, we can obtain

$$\begin{aligned} EA_{\text{centralized}} &= \frac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(l + \mu d_{j0,i}^m) \\ &\geq \frac{1}{n}\sum_{i=2}^{K} n_i(l + \mu((i-1)r)^m) \\ &\geq l\left(1 - \frac{n_1}{n}\right) + \mu r^m \sum_{i=1}^{K-1} i^m(2i+1) \\ &> \mathcal{O}(K^m) = \mathcal{O}(n^{m/2}), \end{aligned}$$

where in the above steps we use the facts that $n = \rho\pi(Kr)^2$, ρ is the density of nodes, $n_i = (i^2 - (i-1)^2)/K^2$, and $\sum_{i=1}^K i^m > K^{m+1}/(m+1)$. We emphasize that this is a quite optimistic analysis without considering the energy spent on the routing updates, etc. In general, much more energy will be required for the centralized approaches. On the other hand, for LS-DSFS, it is easy to show that

$$\begin{aligned} EA_{\text{LS-DSFS}} &= \frac{1}{n}\sum_{i=1}^{n}\left(l|\mathcal{N}_i| + l + \mu \max_j d_{ij}^m\right) \\ &\leq \frac{l}{n}\sum_{i=1}^{n}|\mathcal{N}_i| + l + \mu r^m \\ &< nl + \mu r^m = \mathcal{O}(n). \end{aligned}$$

Taking the above two relations into account, and recalling that $2 \leq m \leq 6$, we can clearly see that the distributed LS-DSFS scales better than the centralized approaches as the number of nodes grows.

The important implication of the above analysis regarding the performance and energy efficiency is that cooperate locally, we can do better!

[3] One possible strategy is brought to our attention by one of the anonymous referees. To be specific, taking $\Delta = 1$ as an example, each node can send the digits of its data one by one to the FC. Upon reception, the FC can recover these digits and thus the data from all the nodes to arbitrary precision.

IV. Performance Analysis of LS-DSFS

In the section, we focus on the proofs of our main results. Several intermediate results will be provided to facilitate the theoretical analysis. Proofs are relegated to the Appendix.

A. Preliminaries

We first establish two basic results regarding properties of the power matrix $\mathbf{Q}^k$ for arbitrary integer $k \geq 0$.

Lemma 2: Assume that $\mathcal{G}$ is strongly connected, for sufficiently small $\beta > 0$, the spectral radius of $\mathbf{Q}$ satisfies $0 < \rho_{\mathbf{Q}} < 1$. And for any integer $k \geq 0$, we have

$$\|\mathbf{Q}^k\|_2 \leq \begin{cases} (\eta+1)^{n-1}\rho_{\mathbf{Q}}^k, & k < n, \\ e^{\eta}k^{n-1}\rho_{\mathbf{Q}}^k, & k \geq n, \end{cases} \tag{10}$$

where $\eta \triangleq \|\mathbf{Q}-\mathbf{Q}^T\|_F/\big(\sqrt{2}\rho_{\mathbf{Q}}\big)$.

Lemma 3: Assume that $\mathcal{G}$ is strongly connected, and conditions (c-2) and (c-3) are satisfied. Then

i) we have

$$\lim_{t\to\infty}\sum_{s=0}^{t}\phi_{t,s}\|\mathbf{Q}^{t-s+1}\|_2 = 0, \tag{11}$$

where $\phi_{t,s} \triangleq \Gamma_{s-1}\prod_{h=s}^{t}\max_i|1-\gamma_{h,i}|$.

ii) for the special form of gain factors $\gamma_{t,i} = a_i/(t+1)^{\tau_i}$, $\forall t, i$ with $a_i > 0$, $0 \leq \tau_i \leq 1$, $\forall i$, one further obtains

$$\sum_{s=0}^{t}\phi_{t,s}\|\mathbf{Q}^{t-s+1}\|_2 = \mathcal{O}\left(1/(t+1)^{\min_i \tau_i}\right). \tag{12}$$

Our next tweak is to introduce an auxiliary system similar to $\hat{x}_i(t)$ in (9) for analysis only

$$\hat{s}_i(t+1) = \hat{s}_i(t) + \gamma_{t,i}[s_i(t+1) - \hat{s}_i(t)], \forall i = 1, 2, \ldots, n. \tag{13}$$

We will later find that the introduction of (13) greatly simplifies the theoretical derivations. Denote by $\mathbf{e}(t) \triangleq [\mathbf{e}^x(t)^T, \mathbf{e}^s(t)^T]^T = \mathbf{z}(t)^T - [\bar{x}(0)\mathbf{1}^T, \mathbf{0}]^T$ the estimation error, and $\hat{\mathbf{e}}(t) \triangleq [\hat{\mathbf{e}}^x(t)^T, \hat{\mathbf{e}}^s(t)^T]^T = [\hat{\mathbf{x}}(t)^T, \hat{\mathbf{s}}(t)^T]^T - [\bar{x}(0)\mathbf{1}^T, \mathbf{0}]^T$ the augmented estimation error, where $\hat{\mathbf{s}}(t) \triangleq [\hat{s}_1(t), \ldots, \hat{s}_n(t)]^T$. The next lemma summarizes two important results of cross relations between $\mathbf{e}(t)$ and $\{\mathbf{w}(t), \hat{\mathbf{e}}(t)\}$.

Lemma 4: Under Assumptions (a-1) and (a-2), we have

i) $\mathbb{E}\{\mathbf{e}(t_1)\mathbf{w}(t_2)^T\} = \mathbf{0}$, $\forall t_1 \leq t_2$.

ii) $\mathbb{E}\{\mathbf{e}(t+1)\hat{\mathbf{e}}(t)^T\} = \mathbf{Q}^{t+1}\mathbb{E}\{\mathbf{e}(0)\mathbf{e}(0)^T\}\mathbf{\Phi}_{t-1,0} + \sum_{s=1}^{t}\mathbf{Q}^{t-s+1}\mathbb{E}\{\mathbf{e}(s)\mathbf{e}(s)^T\}\mathbf{\Phi}_{t-1,s}$, for all $t \geq 0$, where $\mathbf{\Phi}_{t,s} \triangleq \mathbf{\Lambda}_{s-1}\prod_{h=s}^{t}(\mathbf{I}-\mathbf{\Lambda}_h)$, $\mathbf{\Lambda}_t \triangleq \mathrm{diag}\{\gamma_{t,1}, \ldots, \gamma_{t,n}\}$ and we denote $\mathbf{\Lambda}_{-1} = \prod_{h=s_2}^{s_1}(\mathbf{I}-\mathbf{\Lambda}_h) = \mathbf{I}$, $\forall s_2 > s_1$.

Lemma 4 states that the quantization error $\mathbf{w}(t)$ at t-th iteration is uncorrelated with all the historical intermediate errors $\mathbf{e}(s), s = 0, 2, \ldots, t$.

B. Boundedness Results

We next have the following result regarding boundedness of intermediate states $\mathbf{x}(\mathbf{t})$, $\mathbf{s}(t)$ of the distributed quantized algorithm (4) and (5).

Lemma 5: Assume that $\mathcal{G}$ is strongly connected, then for small $\beta > 0$, we have

i)

$$\|\mathbf{e}(t)\| \leq c_{\mathbf{Q}}\|\mathbf{x}(0)\| + \alpha\sqrt{n}\Delta c'_{\mathbf{Q}}\|\mathbf{L}_{\text{aug}}\|, \tag{14}$$

where $c_{\mathbf{Q}} = (\eta+1)^{n-1}$ for $t < n$, and $e^{\eta}[(1-n)/(e\log\rho_{\mathbf{Q}})]^{n-1}$ for $t \geq n$, and $c'_{\mathbf{Q}} \triangleq (\eta+1)^{n-1}(1-\rho_{\mathbf{Q}}^n)/(1-\rho_{\mathbf{Q}}) + e^{\eta}[(1-n)/(e\log\rho_{\mathbf{Q}})]^{n-1} + e^{\eta}\rho_{\mathbf{Q}}^n(n-1)!(1-\log\rho_{\mathbf{Q}})^{n-1}/(-\log\rho_{\mathbf{Q}})$.

ii) if, in addition, Assumptions (a-1) and (a-2) are satisfied, then

$$\limsup_{t\to\infty}\mathbb{E}\{\|\mathbf{e}(t)\|^2\} \leq \frac{n\alpha^2 c''_{\mathbf{Q}}\|\mathbf{L}_{\text{aug}}\|_2^2\sup_t \mathrm{tr}(\mathbf{W}_t)}{2}, \tag{15}$$

where the expectation is taken with respect to the initial condition $\mathbf{x}(0)$ and quantization noises $\{\mathbf{w}(\tau)\}_{\tau=0}^t$, $\mathbf{W}_t = \mathbb{E}\{\mathbf{w}(t)\mathbf{w}(t)^T\}$, $c''_{\mathbf{Q}} \triangleq (\eta+1)^{2(n-1)}(1-\rho_{\mathbf{Q}}^{2n})/(1-\rho_{\mathbf{Q}}^2) + e^{2\eta}[(1-n)/(e\log\rho_{\mathbf{Q}})]^{2(n-1)} + e^{2\eta}\rho_{\mathbf{Q}}^{2n}(2n-2)!(1-2\log\rho_{\mathbf{Q}})^{2(n-1)}(-2\log\rho_{\mathbf{Q}})^{1-2n}$.

One comment on the above result is in order. The explicit upper bound (14) together with (2) reveals that the quantized message $\{q(x_i(t)), q(s_i(t))\}$ is bounded at each node i.

Next, we estimate the average length of quantized message $\{q(x_i(t)), q(s_i(t))\}$ for LS-DSFS with $x_i(0) = y_i$, $\forall i$, where y_i is the measurement in (1). In fact, the quantized message $\{q(x_i(t)), q(s_i(t))\}$ has the length

$$b_{i,t} \leq 2 + \left\lceil \log_2\left\lceil\frac{|x_i(t)|}{\Delta}\right\rceil\right\rceil + \left\lceil \log_2\left\lceil\frac{|s_i(t)|}{\Delta}\right\rceil\right\rceil,$$

where the first term accounts for the sign bits of $q(x_i(t)), q(s_i(t))$, and the last two terms bound the binary lengths of $|q(x_i(t))|, |q(s_i(t))|$. Thus the average length of $\{q(x_i(t)), q(s_i(t))\}$ is bounded as follows:

$$\begin{aligned}\mathbb{E}\{b_{i,t}\} &\leq 4 + \mathbb{E}\left\{\log_2\left(1+\frac{|x_i(t)|}{\Delta}\right)\right\} \\ &\quad + \mathbb{E}\left\{\log_2\left(1+\frac{|s_i(t)|}{\Delta}\right)\right\} \\ &\leq 4 + 2\log_2\left(1+\frac{\mathbb{E}\{\|\mathbf{z}(t)\|\}}{\Delta}\right)\end{aligned}$$

where the last inequality follows from Jensen's inequality.

Further, one can obtain $\{|x_i(t)|, |s_i(t)|\} \leq \|\mathbf{z}(t)\| \leq (1/\sqrt{n})\|\mathbf{x}(0)\|_1 + c_{\mathbf{Q}}\|\mathbf{x}(0)\| + \alpha\sqrt{n}\Delta c'_{\mathbf{Q}}\|\mathbf{L}_{\text{aug}}\|$, where $\|\cdot\|_1$ denotes the 1-norm of vectors. It follows from (1) that $\mathbb{E}\{\|\mathbf{z}(t)\|\} \leq (c_{\mathbf{Q}}+1)\sqrt{n(\theta^2+\sigma^2)} + \alpha\sqrt{n}\Delta c'_{\mathbf{Q}}\|\mathbf{L}_{\text{aug}}\|$. The upper bound is clearly not available since θ is unknown. However, we note that θ^2/σ^2 actually represents the signal-to-noise ratio (SNR) of each node. It is realistic to assume that each node could estimate SNR by simply measuring the received signal power in the presence and absence of the incoming signal. A more aggressive way is just to replace θ with the dynamic range of each node $[-D, D]$, which can be regarded as part of the design specifications. We then obtain

$$\begin{aligned}\mathbb{E}\{\|\mathbf{z}(t)\|\} &\leq \sigma(c_{\mathbf{Q}}+1)\sqrt{n(\mathrm{SNR}+1)} + \alpha\sqrt{n}\Delta c'_{\mathbf{Q}}\|\mathbf{L}_{\text{aug}}\| \\ &\leq (c_{\mathbf{Q}}+1)\sqrt{n(D^2+\sigma^2)} + \alpha\sqrt{n}\Delta c'_{\mathbf{Q}}\|\mathbf{L}_{\text{aug}}\|.\end{aligned}$$

Putting the above relations together, we know that the average message length is decided by quantization step-size Δ, local SNR θ^2/σ^2 (or equivalently dynamic range of each node D), measurement noise variance σ^2, and network topology $\mathcal{G}$ through $\mathbf{L}$ and $\mathbf{L}^{-}$. Obviously, the bound of $\mathbb{E}\{b_{i,t}\}$ is conservative (see Section VI for some simulations). If more information is available, e.g., n is known by each node, then node i could quantize $x_i(t)/n, s_i(t)/n$ instead of $x_i(t), s_i(t)$. And the upper bound of $\mathbb{E}\{b_{i,t}\}$ can be reduced to $4 + 2\log_2(1+\xi/n)$, where $\xi \triangleq \sqrt{n}\sigma(c_{\mathbf{Q}}+1)\sqrt{\text{SNR (or } D) + 1}/\Delta + \alpha\sqrt{n}c'_{\mathbf{Q}}\|\mathbf{L}_{\text{aug}}\|$.

C. Convergence: Proof of Theorem 1

With the above lemmas at hand, we are now in the position to formally prove Theorem 1. For clarity of presentation, we divide the proof into several steps.

Asymptotic Unbiasedness Property: Putting (9) and the auxiliary system (13) together and express them in matrix-vector form, we obtain the augmented error dynamics

$$\hat{\mathbf{e}}(t+1) = (\mathbf{I} - \mathbf{\Lambda}_t)\hat{\mathbf{e}}(t) + \mathbf{\Lambda}_t \mathbf{e}(t+1), \tag{16}$$

where $\mathbf{e}(t)$ is the error of the original system (6).

Recalling that $0 < \delta_t \le \gamma_{t,i} \le 1$, $\forall t \ge 0$, this along with (c-2) and (16) implies

$$\|\mathbb{E}\{\hat{\mathbf{e}}(t+1)\}\| \le (1-\delta_t)\|\mathbb{E}\{\hat{\mathbf{e}}(t)\}\| + \kappa\delta_t\|\mathbb{E}\{\mathbf{e}(t+1)\}\|, \tag{17}$$

where the expectations $\mathbb{E}\{\hat{\mathbf{e}}(t)\}$ and $\mathbb{E}\{\mathbf{e}(t+1)\}$ are taken with respect to the initial condition $\mathbf{x}(0)$ and quantization noises $\{\mathbf{w}(\tau)\}_{\tau=0}^{t}$.

In view of (c-1) and (c-2), we know that $\sum_{t=0}^{\infty}\delta_t = \infty$. Moreover, Lemma 1 shows that $\lim_{t\to\infty}\|\mathbb{E}\{\mathbf{e}(t)\}\| = 0$. Hence, applying the Robbins-Siegmund theorem [41] to (17) yields $\lim_{t\to\infty}\mathbb{E}\{\hat{\mathbf{e}}(t)\} = 0$. This means that $\hat{x}_i(t)$ is asymptotically unbiased for each node i.

Mean Square Property: Now let us turn to the mean square property following the next three steps.

Step 1: (Error of the augmented system) It follows from the condition (c-2) and (16) that

$$\begin{aligned}
\mathbb{E}\{\|\hat{\mathbf{e}}(t+1)\|^2\} &= \mathbb{E}\{\hat{\mathbf{e}}(t)^T(\mathbf{I}-\mathbf{\Lambda}_t)^2\hat{\mathbf{e}}(t)\} \\
&\quad + \mathbb{E}\{\mathbf{e}(t+1)^T\mathbf{\Lambda}_t^2\mathbf{e}(t+1)\} \\
&\quad + 2\mathbb{E}\{\hat{\mathbf{e}}(t)^T\mathbf{\Lambda}_t(\mathbf{I}-\mathbf{\Lambda}_t)\mathbf{e}(t+1)\} \\
&\le (1-\delta_t)^2\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} + \underbrace{\Gamma_t^2\mathbb{E}\{\|\mathbf{e}(t+1)\|^2\}}_{\mathcal{T}_1} \\
&\quad + \underbrace{2\mathbb{E}\{\hat{\mathbf{e}}(t)^T\mathbf{\Lambda}_t(\mathbf{I}-\mathbf{\Lambda}_t)\mathbf{e}(t+1)\}}_{\mathcal{T}_2},
\end{aligned} \tag{18}$$

where we use condition (c-3) to obtain that for large t, say $t \ge t_1$, $0 < \max_i \gamma_{t,i} < 1$, $\forall t \ge t_1$.

Step 2: (Bounding the mean square error $\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\}$) We need to estimate the bounds of the above two terms $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively.

i) *Bound of $\mathcal{T}_1$:* By Lemma 5, we know that $\mathbb{E}\{\|\mathbf{e}(t)\|^2\}$ is bounded, provided that $\beta > 0$ is small enough. Hence,

$$\mathcal{T}_1 = \mathcal{O}(\Gamma_t^2). \tag{19}$$

ii) *Bound of $\mathcal{T}_2$:* Note that $\mathbb{E}\{\hat{\mathbf{e}}(t)^T\mathbf{\Lambda}_t(\mathbf{I}-\mathbf{\Lambda}_t)\mathbf{e}(t+1)\} = \text{trace}(\mathbf{\Lambda}_t(\mathbf{I}-\mathbf{\Lambda}_t)\mathbb{E}\{\mathbf{e}(t+1)\hat{\mathbf{e}}(t)^T\})$ and $\mathbf{\Lambda}_t$ is diagonal, by Lemma 4, one can obtain for all $t \ge 0$,

$$\begin{aligned}
\mathcal{T}_2 &= 2\text{trace}\left(\mathbf{\Lambda}_t\mathbf{Q}^{t+1}\mathbb{E}\{\mathbf{e}(0)\mathbf{e}(0)^T\}\mathbf{\Phi}_{t,0}\right) \\
&\quad + 2\sum_{s=1}^{t}\text{trace}\left(\mathbf{\Lambda}_t\mathbf{Q}^{t-s+1}\mathbb{E}\{\mathbf{e}(s)\mathbf{e}(s)^T\}\mathbf{\Phi}_{t,s}\right) \\
&\le 2\|\mathbf{\Phi}_{t,0}\mathbf{\Lambda}_t\mathbf{Q}^{t+1}\|_2\mathbb{E}\{\|\hat{\mathbf{e}}(0)^T\mathbf{e}(0)\|\} \\
&\quad + 2\sum_{s=1}^{t}\|\mathbf{\Phi}_{t,s}\mathbf{\Lambda}_t\mathbf{Q}^{t-s+1}\|_2\mathbb{E}\{\|\mathbf{e}(s)\|^2\} \\
&\le 2\Gamma_t\sum_{s=0}^{t}\phi_{t,s}\|\mathbf{Q}^{t-s+1}\|_2\psi_s,
\end{aligned}$$

where $\psi_s \triangleq \max\{\mathbb{E}\{\|\mathbf{e}(0)\|^2\}, \mathbb{E}\{\|\mathbf{e}(s)\|^2\}\}$ is bounded by Lemma 5. This along with Lemma 3 implies

$$\mathcal{T}_2 = \mathcal{O}\left(\Gamma_t\sum_{s=0}^{t}\phi_{t,s}\|\mathbf{Q}^{t-s+1}\|_2\right) = o(\Gamma_t), \text{ as } t\to\infty. \tag{20}$$

Step 3: (Convergence of $\mathbb{E}\{\|\hat{\mathbf{x}}(t) - \bar{x}(0)\mathbf{1}\|^2\}$) Combining (19) and (20), we can derive from (18) that for all $t \ge t_1$,

$$\begin{aligned}
\mathbb{E}\{\|\hat{\mathbf{e}}(t+1)\|^2\} &\le (1-\delta_t(2-\delta_t))\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} \\
&\quad + \mathcal{O}(\Gamma_t^2) + o(\Gamma_t).
\end{aligned}$$

Considering the above recurrent inequality, it is clear that $\sum_{t=t_1}^{\infty}\delta_t(2-\delta_t) \ge \sum_{t=t_1}^{\infty}\delta_t = \infty$ in view of the condition (c-1). Furthermore, we have $0 < \delta_t(2-\delta_t) \le 1$ for sufficiently large t by recalling the condition (c-3), and thus

$$\lim_{t\to\infty}\frac{\mathcal{O}(\Gamma_t^2) + o(\Gamma_t)}{\delta_t(2-\delta_t)} \le \lim_{t\to\infty}\mathcal{O}(\kappa^2\delta_t) = 0.$$

By the Robbins-Siegmund theorem [41], we can conclude that $\lim_{t\to\infty}\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} = 0$. Since $\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} = \mathbb{E}\{\|\hat{\mathbf{x}}(t) - \bar{x}(0)\mathbf{1}\|^2\} + \mathbb{E}\{\|\hat{\mathbf{s}}(t)\|^2\}$, it follows that $\lim_{t\to\infty}\mathbb{E}\{\|\hat{\mathbf{x}}(t) - \bar{x}(0)\mathbf{1}\|^2\} \le \lim_{t\to\infty}\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} = 0$, from which the theorem is obtained.

D. Rate of Convergence: Proof of Theorem 2

We now turn to the proof of Theorem 2 on convergence rate analysis for the estimate sequence $\{\hat{\mathbf{x}}(t)\}_{t\ge0}$.

Convergence of $\{\mathbb{E}\{\hat{\mathbf{x}}(t)\}\}_{t\ge0}$: The proof for the case $\inf_{t\ge0}\delta_t = 1$ is trivial. In the following, it suffices to consider the case $\inf_{t\ge0}\delta_t < 1$.

Continuing the recursion of (17) gives

$$\begin{aligned}
\|\mathbb{E}\{\hat{\mathbf{e}}(t)\}\| &\le \prod_{s=0}^{t-1}(1-\delta_s)\|\mathbb{E}\{\hat{\mathbf{e}}(0)\}\| \\
&\quad + \sum_{s=0}^{t-1}\prod_{h=s+1}^{t-1}(1-\delta_h)\Gamma_s\|\mathbb{E}\{\mathbf{e}(s+1)\}\| \\
&\triangleq \Xi_1 + \Xi_2.
\end{aligned} \tag{21}$$

Considering the first term Ξ_1, we have $\Xi_1 \le e^{-\sum_{s=0}^{t-1}\delta_s}\|\mathbb{E}\{\hat{\mathbf{e}}(0)\}\|$, where use was made of the fact that $1 - x \le e^{-x}$, for any scalar $0 \le x \le 1$.

As for the second term Ξ_2, taking expectation on both sides of (33) and applying Lemma 2 yields $\|\mathbb{E}\{\mathbf{e}(t)\}\| \le e^{\eta}t^{n-1}\rho_{\mathbf{Q}}^{t}\|\mathbb{E}\{\mathbf{e}(0)\}\|$, $\forall t \ge n$. Moreover, under the conditions (c-1) and (c-2), we have $\Gamma_t \le \kappa, \forall t$. Hence, $\Xi_2 \le \kappa \prod_{s=0}^{t-1}(1-\delta_s)\sum_{l=0}^{t-1}\prod_{h=0}^{l}(1-\delta_h)^{-1}\|\mathbb{E}\{\mathbf{e}(l+1)\}\| \le \kappa \prod_{s=0}^{t-1}(1-\delta_s)$ $(C_1 + e^{\eta}\|\mathbb{E}\{\mathbf{e}(0)\}\|\sum_{l=n-1}^{t-1} b_l)$, for certain constant $C_1 > 0$, where $b_s \triangleq \prod_{h=0}^{s}(1-\delta_h)^{-1}(s+1)^{n-1}\rho_{\mathbf{Q}}^{s+1}$.

To finish the proof, we claim that

$$\sum_{s=n-1}^{t-1} b_s < \begin{cases} \frac{\rho_{\mathbf{Q}} b_{t-1}}{\rho_{\mathbf{Q}}-1+\inf \delta_t}, & 1-\rho_{\mathbf{Q}} < \inf \delta_t < 1, \\ \infty, & 0 < \sup \delta_t < 1-\rho_{\mathbf{Q}}, \\ \frac{t^n-(n-1)^n}{n}, & 0 < \inf \delta_t \le \sup \delta_t = 1-\rho_{\mathbf{Q}}. \end{cases} \tag{22}$$

We refer the readers to the Appendix for its proof. Accordingly, let us consider the next three cases:

Case I: $1-\rho_{\mathbf{Q}} < \inf_{t\ge 0}\delta_t < 1$. In this case, we obtain $\Xi_2 \le \frac{\kappa e^{\eta}\|\mathbb{E}\{\mathbf{e}(0)\}\|\rho_{\mathbf{Q}} t^{n-1}\rho_{\mathbf{Q}}^t}{\rho_{\mathbf{Q}}-1+\inf_t \delta_t} + o(t^n \rho_{\mathbf{Q}}^t)$, where we use the facts that $\prod_{s=0}^{t-1}(1-\delta_s) \le \left(1-\inf_t \delta_t\right)^t$ and $1-\inf_t \delta_t < \rho_{\mathbf{Q}}$.

Case II: $0 < \sup_{t\ge 0}\delta_t < 1-\rho_{\mathbf{Q}}$. Note that $1-x \le e^{-x}$, for any scalar $0 \le x \le 1$. Substituting this relation shows that $\Xi_2 \le \kappa C_2 e^{-\sum_{s=0}^{t-1}\delta_s}$, for some constant $C_2 > C_1$.

Case III: $0 < \inf_{t\ge 0}\delta_t \le \sup_{t\ge 0}\delta_t = 1-\rho_{\mathbf{Q}}$. Since $\inf_{t\ge 0}\delta_t > 0$, one has $e^{-\lim\inf_{t\to\infty}\delta_t} < 1$, from which we know that $\sum_{t=1}^{\infty} t^n e^{-\sum_{s=0}^{t-1}\delta_s}$ is convergent. Thus, $\lim_{t\to\infty} t^n e^{-\sum_{s=0}^{t-1}\delta_s} = 0$. This together with (22) implies $\Xi_2 \le (1/n)\kappa e^{\eta}$ $\|\mathbb{E}\{\mathbf{e}(0)\}\| t^n e^{-\sum_{s=0}^{t-1}\delta_s} + o\left(t^n e^{-\sum_{s=0}^{t-1}\delta_s}\right)$.

We substitute all the above relations to (21) to establish the rate of convergence of $\|\mathbb{E}\{\hat{\mathbf{x}}(t)\} - \bar{x}(0)\mathbf{1}\|$.

Mean Square Convergence: Considering $\gamma_{t,i} = a_i/(t+1)^{\tau_i}$, by definition, we have $\delta_t = \underline{a}/(t+1)^{\bar{\tau}}$, $\Gamma_t = \bar{a}/(t+1)^{\underline{\tau}}$, where $\underline{a} \triangleq \min_i a_i$, $\bar{a} \triangleq \max_i a_i$, $\underline{\tau} \triangleq \min_i \tau_i$ and $\bar{\tau} \triangleq \max_i \tau_i$.

Under the condition (c-3), one can show that $\delta_t \le \Gamma_t < 1$, for all $t \ge t_1 \triangleq (\bar{a}+1)^{\frac{1}{\underline{\tau}}} - 1$. It thus follows from (18) that for all $t \ge t_1$,

$$\mathbb{E}\{\|\hat{\mathbf{e}}(t+1)\|^2\} \le (1-\delta_t)\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} + \left(\Gamma_t^2 + 2\Gamma_t \sum_{s=0}^{t}\phi_{t,s}\|\mathbf{Q}^{t-s+1}\|_2\right)c_{\mathbf{e}}, \tag{23}$$

where $\phi_{t,s} \triangleq \Gamma_{s-1}\prod_{h=s}^{t}\max_i |1-\gamma_{h,i}|$ and $c_{\mathbf{e}} \triangleq \sup_{t\ge 0}$ $\mathbb{E}\{\|\mathbf{e}(t)\|^2\} < \infty$ in light of Lemma 5.

We use definitions of δ_t, Γ_t and Lemma 3 to rewrite (23) as follows

$$\mathbb{E}\{\|\hat{\mathbf{e}}(t+1)\|^2\} \le \left(1 - \frac{\underline{a}}{(t+1)^{\bar{\tau}}}\right)\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} + \frac{\bar{a}^2(2c'_{\mathbf{Q}}+1)c_{\mathbf{e}}}{(t+1)^{2\underline{\tau}}} + o\left(\frac{1}{(t+1)^{2\underline{\tau}}}\right),$$

where $c'_{\mathbf{Q}}$ is given in Lemma 5. Remember that $\bar{\tau} < 2\underline{\tau} < \underline{a}+1$, applying Chung's lemma [42] to the above relation yields

$$\limsup_{t\to\infty}(t+1)^{2\underline{\tau}-\bar{\tau}}\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} \le \begin{cases} \frac{\bar{a}^2(2c'_{\mathbf{Q}}+1)c_{\mathbf{e}}}{\underline{a}}, & \bar{\tau} < 1, \\ \frac{\bar{a}^2(2c'_{\mathbf{Q}}+1)c_{\mathbf{e}}}{\underline{a}+1-2\underline{\tau}}, & \bar{\tau} = 1. \end{cases}$$

This completes the proof.

V. Running Average Estimate: Special Form of LS-DSFS

In the previous sections, we have provided the performance analysis of LS-DSFS and given some comparison results with the prevailing centralized approaches. In this section, we will focus on a special form of LS-DSFS, namely, the running average estimation.

Let $x_i(0) = \hat{x}_i(0) = y_i$ and

$$\gamma_{t,i} = \frac{1}{t+2}, \ \forall i = 1, 2, \ldots, n. \tag{24}$$

Substituting (24) into (9) yields

$$\hat{x}_i^{\text{RA}}(t) = \frac{t}{t+1}\hat{x}_i^{\text{RA}}(t-1) + \frac{1}{t+1}x_i(t) = \frac{1}{t+1}\sum_{k=0}^{t} x_i(k), \ \forall i. \tag{25}$$

Such $\hat{x}_i^{\text{RA}}(t)$ is hereinafter referred to as *running average estimate*. This technique is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends [27], [31]. It is easy to verify that $\gamma_{t,i}$ in (24) satisfies all the conditions (c-1), (c-2) and (c-3), thus Theorem 1 still holds. As for the rate of convergence, actually, we have an enhanced version of Theorem 2 shown below.

Theorem 3: Consider LS-DSFS in Algorithm 1 with $\{\gamma_{t,i}\}_{t,1\le i\le n}$ given in (24) and assume that $\mathcal{G}$ is strongly connected. Then under Assumptions (a-1) and (a-2), for sufficiently small $\beta > 0$, we have $\mathbb{E}\{\hat{\mathbf{x}}^{\text{RA}}(t)\} = \theta\mathbf{1}$ and

$$\limsup_{t\to\infty} t\|\mathbb{E}\{\|\hat{\mathbf{x}}^{\text{RA}}(t) - \hat{\theta}\mathbf{1}\|^2\}\| \le \alpha^2\|(\mathbf{I}-\mathbf{Q})^{-1}\|_2^2\|\mathbf{L}_{\text{aug}}\|_2^2 \sup_t \text{tr}(\mathbf{W}_t),$$

where the expectation is over the initial condition $\mathbf{x}(0)$ and quantization noises $\{\mathbf{w}(\tau)\}_{\tau=0}^t$.

More important, the running average method has another appealing feature that enables a dedicate rate of convergence result in the almost sure sense.

Theorem 4: Suppose that all the assumptions of Theorem 3 hold and the quantization noises $\{\mathbf{w}(t)\}_{t\ge 0}$ are temporally independent. Moreover, there is a positive constant $d > 2$ such that $\sup_{t\ge 0}\mathbb{E}\{\|\mathbf{w}(t)\|^d\} < \infty$. Then for small $\beta > 0$,

$$\limsup_{t\to\infty}\frac{\|\hat{\mathbf{x}}^{\text{RA}}(t) - \hat{\theta}\mathbf{1}\|}{t^{-1}\sqrt{r_t \log\log r_t}} \le \alpha\sqrt{2n\sigma_{\mathbf{w}}}, \ a.s.$$

where $r_t = \sum_{k=0}^{t-1}\|(\mathbf{I}-\mathbf{Q}^{t-k})(\mathbf{I}-\mathbf{Q})^{-1}\|_2^2$ and $\sigma_{\mathbf{w}} = \sup_{t\ge 0}$ $\|\mathbf{W}(t)\|_2$.

Proof: Let $\mathbf{Q}_{t,k} \triangleq (\mathbf{I}-\mathbf{Q}^{t-k})(\mathbf{I}-\mathbf{Q})^{-1}$. Then from the proof of Theorem 3, one can obtain

$$\|\hat{\mathbf{e}}(t)\| \le \frac{1}{t+1}\|(\mathbf{I}-\mathbf{Q}^{t+1})(\mathbf{I}-\mathbf{Q})^{-1}\hat{\mathbf{e}}(0)\| + \frac{\alpha}{t+1}\left\|\sum_{k=0}^{t-1}\mathbf{Q}_{t,k}\mathbf{L}_{\text{aug}}\mathbf{w}(k)\right\| \triangleq \Phi_1 + \Phi_2. \tag{26}$$

By Lemma 2, the first term Φ_1 is of the order $\mathcal{O}(1/t)$. It remains to consider the second term Φ_2. Actually, we find that Φ_2 is in the form of the sum of random vectors $\mathbf{L}_{\text{aug}}\mathbf{w}(k)$ weighted by the matrix $\mathbf{Q}_{t,k}$.

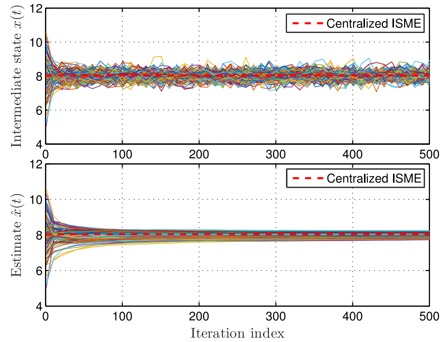


Fig. 3. Intermediate state $\mathbf{x}(t)$ of (4) and the estimate $\hat{\mathbf{x}}(t)$ generated by LS-DSFS for $\Delta = 1$.

- Firstly, we claim that the weight matrix is bounded. To see this, we can verify that $\mathbf{Q}[\bar{x}(0)\mathbf{1}^T, \mathbf{0}]^T = \mathbf{0}$. Hence, $(\mathbf{I} - \mathbf{Q}^{t-k})[\bar{x}(0)\mathbf{1}^T, \mathbf{0}]^T = (\mathbf{I} - \mathbf{Q})[\bar{x}(0)\mathbf{1}^T, \mathbf{0}]^T$, $\forall t > k$, which shows that 1 is an eigenvalue of $\mathbf{Q}_{t,k}$. This means that $\|\mathbf{Q}_{t,k}\|_2 \geq 1$, $\forall t > k$. On the other hand, it is obvious that $\|\mathbf{Q}_{t,k}\|_2 \leq (1 + \|\mathbf{Q}^{t-k}\|_2)\|(\mathbf{I} - \mathbf{Q})^{-1}\|_2$, which by Lemma 2 is also upper bounded for all $t \geq k$.
- Secondly, we can show that $\sum_{k=1}^{\infty} \|\mathbf{Q}^k\|_2^2$ is a convergent series in view of the fact that $\rho_{\mathbf{Q}} < 1$. As a consequence, for any $t_2 \geq t_1$, we can find a constant $B > 0$ such that $\sum_{k=0}^{t_1-1} \|\mathbf{Q}^{t_2-k} - \mathbf{Q}^{t_1-k}\|_2^2 \leq B$.

It thus follows from the law of the iterated logarithm for weighted sums of independent random vectors [28] that $\limsup_{t\to\infty} \frac{\|\sum_{k=0}^{t-1} \mathbf{Q}_{t,k} \mathbf{L}_{\text{aug}} \mathbf{w}(k)\|}{\sqrt{r_t \log\log r_t}} < \sqrt{2n\sigma_{\mathbf{w}}}$ $a.s.$ As a result, the second term Φ_2 is in the order of $\mathcal{O}(\sqrt{r_t \log\log r_t}/t)$.

Substituting the above results into (26) yields $\|\hat{\mathbf{e}}(t)\| \leq \mathcal{O}(\sqrt{r_t \log\log r_t}/t)$. This completes the proof. ∎

VI. Simulation Results

In this section, simulation results are presented to demonstrate the efficiency and effectiveness of the proposed LS-DSFS for solving the problem of distributed sensor fusion.

Consider a random network of $n = 100$ nodes to monitor an unknown parameter $\theta = 8$ with each observation given by $y_i = \theta + w_i$, where w_i is the Gaussian noise with unit variance. To simulate certain pairs of asymmetric links between nodes, we first generate an undirected network using the random geometric graph model, i.e., nodes are placed uniformly at random over $[0, 1] \times [0, 1]$ and a pair of nodes is connected by two unidirectional links if the distance is less than $\sqrt{\log n/n}$. With this setting, the network topology is guaranteed to be connected with high probability. Afterwards, 60% of the unidirectional links are randomly removed to generate an asymmetric network. In the following simulations, we adopt the probabilistic quantization scheme (3). The parameters are set as follows: $\alpha = 1/(0.1 + \max_i d_i^+)$ and $\beta = 0.1$ for (4), and $\gamma_t = \sqrt{0.5 \log\log(t + 10)}/t$ for (9). It is easy to check that such γ_t satisfies the conditions of Theorem 1.

Fig. 3 depicts the intermediate state $\mathbf{x}(t)$ of (4), and the estimate $\hat{\mathbf{x}}(t)$ generated by LS-DSFS for $\Delta = 1$. It is clear that LS-DSFS can smooth out fluctuations exhibited in the intermediate state $\mathbf{x}(t)$. On the other hand, the centralized ISME $\hat{\theta}$ is asymptotically approached, which is only possible for estimation with infinite bandwidth, showing that LS-DSFS has the ability of mitigating the quantization effect on distributed sensor fusion. This corroborates the theoretical result given in Theorem 1.

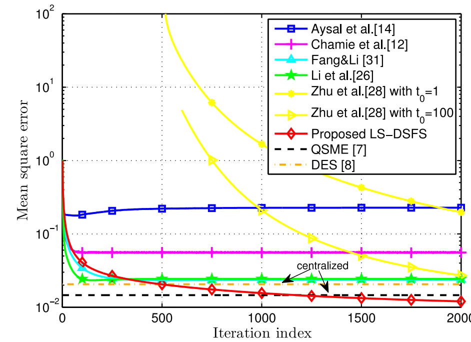


Fig. 4. Comparison results of e_{mse} between different algorithms for $\Delta = 1$.

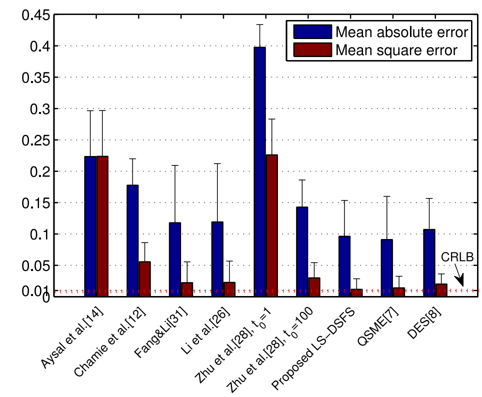


Fig. 5. Comparison results of ensemble average of e_{mae} and e_{mse} between different algorithms for $\Delta = 1$. An average of the last 200 iterations is taken for each bar to avoid transient periods.

Performances of the algorithms are measured by comparing with the true parameter θ in terms of mean absolute error $e_{\text{mae}}(t) = (1/n)\sum_{i=1}^{n} |\text{estimate}_i(t) - \theta|$ and mean square error $e_{\text{mse}}(t) = (1/n)\sum_{i=1}^{n} (\text{estimate}_i(t) - \theta)^2$, where $e_{\text{mae}}(t)$ is introduced to measure unbiasedness of the estimate. The following simulation results are for 200 realizations of initial values $\mathbf{x}(0)$, averaged over 200 runs of the algorithms. Fig. 4 depicts $e_{\text{mse}}(t)$'s of LS-DSFS, from which we find that there is a continuous decrease in the mean square error down to a low level within a few steps, approaching the CRLB $= \sigma^2/n = 0.01$. Also plotted are the centralized approaches QSME [7] and DES [8], and five distributed algorithms [12], [14], [26], [28], [31], for comparison purpose. We can easily identify that our LS-DSFS not only achieves a superior estimation performance to other distributed algorithms, but also outperforms the centralized QSME and DES within a moderate number of iterations. This is a direct implication of Theorem 1 and is consistent with the theoretical comparison results given in Table I. More comparison results can be found in Fig. 5, where 2000 iterations are performed for each algorithm and the averages over the last 200 iterations of $e_{\text{mae}}(t)$ and $e_{\text{mse}}(t)$ are presented. From the figure, we can see that LS-DSFS attains both the smallest mean absolute error and mean square error. This corroborates the convergence results given in Theorems 1 and 2. It should also be noted that the distributed algorithms proposed for symmetric networks [12], [14], [31] fail in the network with asymmetric links, where $e_{\text{mse}}(t)$'s remain almost unchanged after several steps. Moreover, the

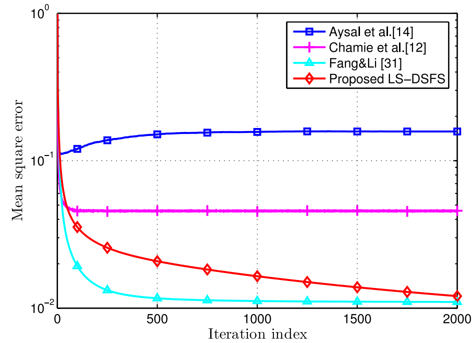


Fig. 6. Comparison results of e_{mse} between different algorithms for a symmetric network with $n = 100$ and $\Delta = 1$.

dynamic encoder-decoder scheme designed for asymmetric networks [26] can not solve the problem of sensor fusion either as can be noticed in Figs. 4 and 5. We also note that the two-stage algorithm [27], [28] can solve the sensor fusion problem over asymmetric networks as well, but its performance heavily depends on convergence of the first stage. In general, it should wait for several iterations before triggering the second stage. In the simulation, we let the first stage run for 500 iterations and then start the second stage. The averaging operation of the second stage starts at $500 + t_0$. It is observed that the convergence is still quite slow when $t_0 = 100$. In practice, these parameters should be carefully tuned for better performance. As a comparison, the proposed LS-DSFS adopts the static probabilistic quantizer and does not need to choose such parameters while achieving the CRLB asymptotically. These simulation results validate the effectiveness of the least squares approach of our LS-DSFS in solving the distributed sensor fusion problem in asymmetric networks with bandwidth constraints.

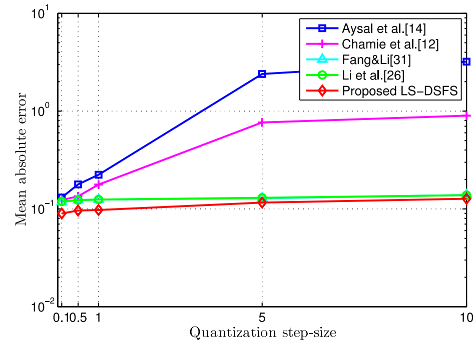
Fig. 6 depicts the comparison results between LS-DSFS and those in [12], [14], [31] for symmetric networks, from which we can see that both LS-DSFS and the algorithm in [31] approach the CRLB 0.01, although LS-DSFS converges with a little bit slower rate. The reason is due to the introduction of a second variable $s_i(t)$ in (5) of LS-DSFS, and the parameter $\beta > 0$ in (4) which should be small to ensure the convergence. Actually, $s_i(t)$ is introduced to handle the case of asymmetric networks. For symmetric networks, we find that $s_i(t)$ is not needed, and (4) and (5) become

$$x_i(t+1) = x_i(t) - \alpha d_i^+ q(x_i(t)) + \alpha \sum_{j \in \mathcal{N}_i^+} q(x_j(t)). \quad (4')$$

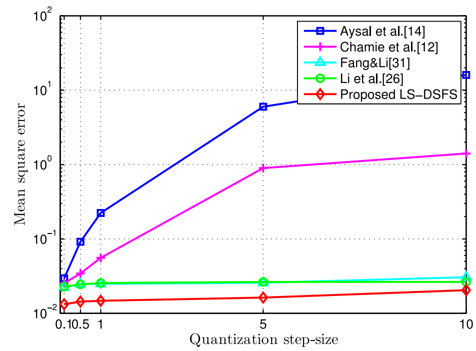
Theoretical analysis can still be enforced in this case, and LS-DSFS reduces to a weighted version of the algorithm in [37].

Next, we compare LS-DSFS with other algorithms with respect to $e_{mae}(t)$ and $e_{\text{mse}}(t)$ for different quantization step-sizes $\Delta \in \{0.1, 0.5, 1, 5, 10\}$. We simulate the LS-DSFS with different initial guesses, all the results show averages of 200 runs. And the results are averaged over the last 200 iterations of total 2000 iterations to avoid transient periods, respectively. From Fig. 7, we observe that the proposed LS-DSFS works quite well for all quantization step-sizes even for the relatively coarse quantization $\Delta = 10$. For smaller quantization step-sizes, the improvements of the performance by using LS-DSFS compared with other distributed algorithms is obvious. To look further into the effect of quantization step-size on the convergence, we run LS-DSFS for different $\Delta \in \{0.1, 0.5, 1, 5, 10\}$ and collect the number of iterations such that $e_{\text{mse}}(t)$ gets to 2*CRLB $= 0.02$. Note that 2*CRLB is the upper bound for MSE of the centralized DES [8]. Averaged over 200 independent runs, we obtain the respective average number of iterations as follows: $\{285.18, 304.87, 456.45, 781.67, 1452.60\}$. It is noted that the average number of iterations increases with Δ, and more iterations are needed to achieve the same level of performance for larger quantization step-sizes. This is intuitively true, since larger quantization step-sizes introduce bigger volume of quantization noises into the system, and thus more iterations are needed to smear out these noises.

We also investigate the effects of network topology on the convergence rate of LS-DSFS, which is characterized by the average number of iterations to get to specific accuracy. To this end, several asymmetric networks are randomly generated using the random geometric graph model mentioned



(a)



(b)

Fig. 7. Average e_{mae}'s and e_{mse}'s of five distributed algorithms for $\Delta \in \{0.1, 0.5, 1, 5, 10\}$. Results are for 200 realizations of initial guesses $x_i(0), \forall i$, averaged over 200 independent runs of LS-DSFS.

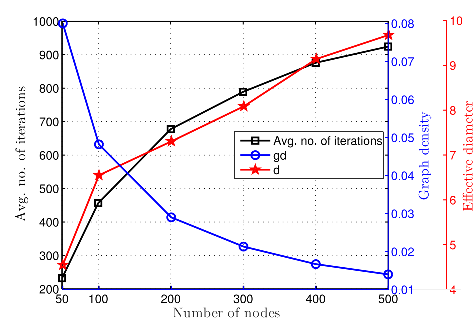


Fig. 8. Average number of iterations to get to $e_{\text{mse}}(t)$ of 2*CRLB along with the graph density and effective diameter with the number of nodes n varying from 50 to 500.

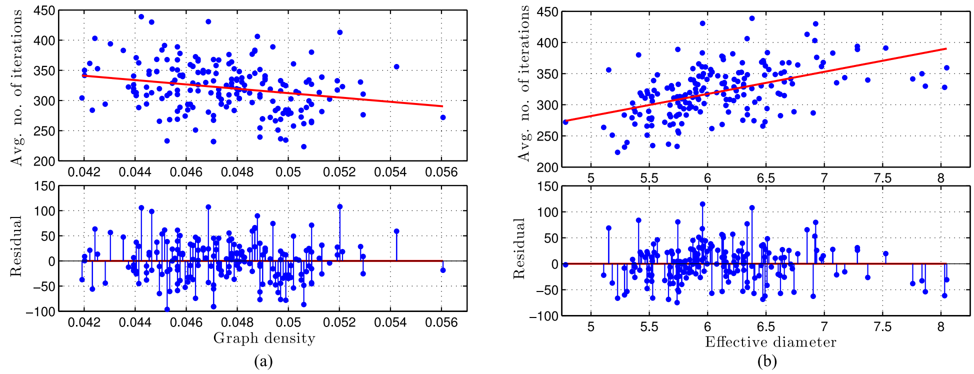


Fig. 9. Scatter plots and residuals of average number of iterations to get to $e_{\mathrm{mse}}(t)$ of 2*CRLB versus graph density and effective diameter, respectively, where 200 asymmetric networks with $n = 100$ are generated randomly.

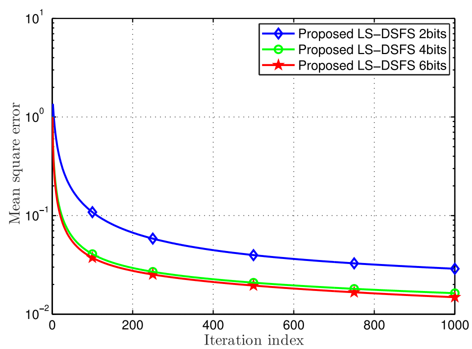


Fig. 10. Mean square error e_{mse} of the proposed LS-DSFS for a truncated probabilistic quantizer with 2 bits, 4 bits, and 6 bits.

earlier. Fig. 8 depicts the average number of iterations to get to $e_{\mathrm{mse}}(t)$ of 2*CRLB against the number of nodes varying from $n = 50$ to 500. The impacts of network topology are measured by two widely used indexes, namely, the graph densities $gd = \#links/[n(n-1)/2]$ and the effective diameters d. In the simulations, we take the effective diameter as the minimum value d such that at least 80% of the connected node pairs are at distance at most d. The plots suggest that as the network tends to larger and sparser, the number of iterations taken by LS-DSFS to achieve the same level of performance increases, and that the number of iterations is directly related with the effective diameter d. To further investigate their impacts, we fix the number of nodes $n = 100$ and randomly generate 200 asymmetric graphs. The scatter plots and corresponding residuals of average number of iterations to get to $e_{\mathrm{mse}}(t)$ of 2*CRLB versus graph density and effective diameter are presented in Fig. 9. The results illustrate that the average number of iterations is related with graph density roughly in a negative manner, while is positively related with effective diameter. This is consistent with the results shown in Fig. 8. All these simulation results show that gd and d have a coupled influence on the convergence time of LS-DSFS. We emphasize that the impact of the network topology on the performance of LS-DSFS has been recognized theoretically in Theorem 2 and the remarks followed.

To see the performance of the proposed LS-DSFS under finite-bit quantizers, we implement a truncated version of the probabilistic quantizer (3), i.e., projecting the transmitted message into $[-\max_i |x_i(0)|, \max_i |x_i(0)|]$. This range is divided into intervals of length $\Delta = 2\max_i |x_i(0)|/(2^b - 1)$, where b denotes the number of bits for quantization. Fig. 10 depicts simulation results for 200 realizations of initial values x(0), each being averaged over 200 runs. Of note that the proposed LS-DSFS also performs quite well for the above finite-bit quantizer. Moreover, it is observed that a few number of bits suffice to produce acceptable mean square errors, i.e., LS-DSFS is resilient to truncation errors of the probabilistic quantizer (3). This in turn reveals that the boundedness results of Lemma 5 and the discussions followed are quite conservative. A rigorous analysis of the proposed LS-DSFS with finite-bit quantizer, however, deserves further investigation.

VII. Conclusion

We proposed a least squares approach for distributed consensus algorithm in bandwidth-constrained and asymmetric sensor networks with applications to distributed sensor fusion, termed LS-DSFS. In LS-DSFS, we first introduced a surplus variable to keep track of local states in order to account for asymmetric links, and then adopted a compensating updating rule, resulting in a quantized consensus algorithm. A least squares reformulation of these local states was then performed to generate local estimates. It was shown that all such estimates converge to the ideal centralized estimate in mean and mean square senses, outperforming prevailing centralized and distributed algorithms. Furthermore, we found that the LS-DSFS will always consume less energy and achieve more balanced energy expenditure as the number of nodes in the network grows. Rate of convergence of LS-DSFS was also established. Finally, we provided some simulation studies to validate the theoretical results.

Appendix

A. Proof of Lemma 2

Firstly, for sufficiently small $\beta > 0$, we can show that $\lim_{k\to\infty} \mathbf{P}^k = \mathbf{P}_\infty$, i.e., $\lim_{k\to\infty} \mathbf{Q}^k = \mathbf{0}$, and thus $\rho_{\mathbf{Q}} < 1$.

To prove the second assertion, for each integer $k \geq 0$, we apply [43, Theorem 2] to the matrix $\mathbf{Q}^k$ so that

$$\begin{aligned}\|\mathbf{Q}^k\|_2 &\leq \sum_{i=0}^{\min\{k,n-1\}} \binom{k}{i} \rho_{\mathbf{Q}}^{k-i} \mathrm{dn}_2(\mathbf{Q})^i \\ &\leq \rho_{\mathbf{Q}}^k \sum_{i=0}^{\min\{k,n-1\}} \binom{k}{i} \left(\frac{\|\mathbf{Q} - \mathbf{Q}^T\|_F}{\sqrt{2}\rho_{\mathbf{Q}}} \right)^i, \end{aligned} \quad (27)$$

where $\mathrm{dn}_2(\mathbf{Q})$ is the departure from normality in spectral norm, and the last step follows from an inequality in [44] that $\mathrm{dn}_2(\mathbf{Q}) \le \|\mathbf{Q}-\mathbf{Q}^T\|_F/\sqrt{2}$.

Let $\eta \triangleq \|\mathbf{Q}-\mathbf{Q}^T\|_F/(\sqrt{2}\rho_{\mathbf{Q}})$. We have two cases. (i) For $k<n$, one has $\sum_{i=0}^{k}\binom{k}{i}\eta^i = (\eta+1)^k \le (\eta+1)^{n-1}$; (ii) While for $k \ge n$, it follows from Taylor expansion that $\sum_{i=0}^{n-1}\binom{k}{i}\eta^i \le k^{n-1}\sum_{i=0}^{n-1}\eta^i/i! \le e^{\eta}k^{n-1}$. Substituting the previous two inequalities into (27) completes the proof.

B. Proof of Lemma 3

i) By Lemma 2, we have $\lim_{t\to\infty}\|\mathbf{Q}^t\|_2 = 0$. Moreover, $\lim_{t\to\infty}\gamma_{t,i}=0, \forall i$ by condition (c-3). In consequence, for arbitrary $\epsilon>0$, there are $t_2>t_1$ such that $0<\max_i \gamma_{t,i}<1$, $\forall t\ge t_1$ and $\|\mathbf{Q}^t\|_2\le\epsilon$, $\forall t \ge t_2$, and thus $\sup_{t\ge 0}\|\mathbf{Q}^t\|_2 \le c_*$ for some constant $c_*>0$. This yields

$$\sum_{s=0}^{t}\phi_{t,s}\|\mathbf{Q}^{t-s+1}\|_2 \le c_*\sum_{s=t-t_2+2}^{t}\phi_{t,s} + \epsilon\sum_{s=0}^{t-t_2+1}\phi_{t,s}. \quad (28)$$

For the first term of the right-hand side of (28), we use the fact that $0<\max_i\gamma_{t,i}<1, \forall t\ge t_1$ to obtain

$$\lim_{t\to\infty}\sum_{s=t-t_2+2}^{t}\phi_{t,s} \le \lim_{t\to\infty}\sum_{s=1}^{t_2-1}\Gamma_{t-s} = 0. \quad (29)$$

As for the second term of the right-hand side, one has

$$\begin{aligned}\sum_{s=0}^{t-t_2+1}\phi_{t,s} &\overset{(a)}{\le} \sum_{s=0}^{t_1}\phi_{t_1,s} + \sum_{s=t_1}^{t-t_2}\Gamma_s\prod_{h=s+1}^{t}(1-\delta_h)\\ &\overset{(b)}{\le} \sum_{s=0}^{t_1}\phi_{t_1,s} + \kappa\sum_{s=t_1}^{t}\delta_s\prod_{h=s+1}^{t}(1-\delta_h)\\ &\overset{(c)}{=} \sum_{s=0}^{t_1}\phi_{t_1,s} + \kappa\left(1-\prod_{h=t_1}^{t}(1-\delta_h)\right) < \infty, \end{aligned} \quad (30)$$

where (a) is obtained by the fact that $0<\max_i\gamma_{t,i}<1$, $\forall t \ge t_1$; (b) is due to (c-2) and (c-3); (c) follows from the identity $\prod_{h=t_1}^{t}(1-\delta_h)+\sum_{s=t_1}^{t}\delta_s\prod_{h=s+1}^{t}(1-\delta_h)=1$, $\forall t\ge t_1$, which can be easily checked by induction on t. Substituting (29) and (30) into (28) shows that $\sum_{s=0}^{t}\phi_{t,s}\|\mathbf{Q}^{t-s+1}\|_2$ can be arbitrarily small for large t. This proves (11).

ii) For $\gamma_{t,i}=a_i/(t+1)^{\tau_i}$, $\forall t,i$, we can simply choose $\delta_t = \underline{a}/(t+1)^{\bar{\tau}}$, $\Gamma_t = \bar{a}/(t+1)^{\underline{\tau}}$, $\forall t\ge 0$, where $\underline{a}\triangleq\min_i a_i$, $\bar{a}\triangleq\max_i a_i$, $\underline{\tau}\triangleq\min_i\tau_i$ and $\bar{\tau}\triangleq\max_i\tau_i$. Hence, similar to (28), one can show that

$$\begin{aligned}\sum_{s=0}^{t}\phi_{t,s}\|\mathbf{Q}^{t-s+1}\|_2 &\le \max_{0\le k\le t_1}\phi_{t_1,k}\sum_{s=t-t_1+1}^{t+1}\|\mathbf{Q}^s\|_2\\ &\quad + \sum_{s=t_1}^{t-1}\Gamma_s\prod_{h=s+1}^{t}(1-\delta_h)\|\mathbf{Q}^{t-s}\|_2\\ &\triangleq \Phi_1+\Phi_2. \end{aligned} \quad (31)$$

For the first term Φ_1, taking $t_2 \triangleq t_1 - 1 + \max\{n, (1-n)/\log\rho_{\mathbf{Q}}\}$, and noting that $s^{n-1}\rho_{\mathbf{Q}}^s$ is an increasing function of s in the interval $[(1-n)/\log\rho_{\mathbf{Q}}, \infty)$, we obtain from Lemma 2 that

$$\begin{aligned}\Phi_1 &\le e^{\eta}\max_{0\le k\le t_1}\phi_{t_1,k}\sum_{s=t-t_1+1}^{t+1}s^{n-1}\rho_{\mathbf{Q}}^s\\ &\le (t_1+1)e^{\eta}\max_{0\le k\le t_1}\phi_{t_1,k}(t+1)^{n-1}\rho_{\mathbf{Q}}^{t-t_1+1}, \forall t \ge t_2. \end{aligned} \quad (32)$$

Regarding the second term Φ_2, we have $\prod_{h=s+1}^{t}(1-\delta_h) \le e^{-\sum_{h=s+1}^{t}\delta_h} \le e^{-\int_{s+1}^{t+1}\frac{\underline{a}}{(h+1)^{\bar{\tau}}}dh} = e^{-\frac{\underline{a}}{1-\bar{\tau}}\left[(t+2)^{1-\bar{\tau}}-(s+2)^{1-\bar{\tau}}\right]}$, where use was made of the fact that $1-x\le e^{-x}$, for any scalar $0\le x\le 1$. Moreover, it is straightforward to verify that $f(s)\triangleq(s+1)^{-\underline{\tau}}e^{\frac{\underline{a}}{1-\bar{\tau}}(s+2)^{1-\bar{\tau}}}$ is an increasing function of s in the interval $[(2\underline{\tau}/\underline{a})^{\frac{1}{1-\bar{\tau}}}-2, \infty)$. It then follows from Lemma 2 that for all $t_1 \ge (2\underline{\tau}/\underline{a})^{\frac{1}{1-\bar{\tau}}}-2$,

$$\begin{aligned}\Phi_2 &\le \bar{a}e^{-\frac{\underline{a}}{1-\bar{\tau}}(t+2)^{1-\bar{\tau}}}\sum_{s=t_1}^{t-1}f(s)\|\mathbf{Q}^{t-s}\|_2\\ &\le \frac{\bar{a}}{(t+1)^{\underline{\tau}}}\left((\eta+1)^{n-1}\sum_{s=1}^{n-1}\rho_{\mathbf{Q}}^s + e^{\eta}\sum_{s=n}^{t-t_1}s^{n-1}\rho_{\mathbf{Q}}^s\right),\\ &= \mathcal{O}\left(\frac{1}{(t+1)^{\underline{\tau}}}\right), \end{aligned}$$

where in the last step we use [27, Lemma D.1] to show that the series $\sum_{s=n}^{\infty}s^{n-1}\rho_{\mathbf{Q}}^s$ is bounded.

Since $\rho_{\mathbf{Q}}<1$, we know that $\lim_{t\to\infty}(t+1)^{n+\underline{\tau}}\rho_{\mathbf{Q}}^t = 0$. This reveals that Φ_1 will be dominated by Φ_2 for large t. Combining this with (31) completes the proof.

C. Proof of Lemma 4

i) First, we derive from Assumption (a-2) that $\mathbb{E}\{\mathbf{e}(0)\mathbf{w}(0)^T\} = (\mathbf{I}-\mathbf{P}_\infty)\mathbb{E}\{\mathbf{z}(0)\mathbf{w}(0)^T\} = \mathbf{0}$. Moreover, noting that $\mathbf{1}^T(\mathbf{x}(t)+\mathbf{s}(t)) = \mathbf{1}^T\mathbf{x}(0)$, $\forall t\ge 0$, we have $\mathbf{e}(t+1) = \mathbf{z}(t+1) - \mathbf{P}_\infty\mathbf{z}(t)$. This, coupled with (6) and the fact that $\mathbf{Q}[\bar{x}(0)\mathbf{1}^T, \mathbf{0}]^T = \mathbf{0}$, yields

$$\mathbf{e}(t+1) = \mathbf{Q}\mathbf{e}(t) + \alpha\mathbf{L}_{\text{aug}}\mathbf{w}(t), \quad (33)$$

where we use the facts that $\mathbf{L}\mathbf{1} = \mathbf{0} = \mathbf{1}^T\mathbf{L}^-$. Hence, $\mathbf{e}(t) = \mathbf{Q}^t\mathbf{z}(0) + \alpha\sum_{k=0}^{t-1}\mathbf{Q}^{t-1-k}\mathbf{L}_{\text{aug}}\mathbf{w}(k)$, $\forall t>0$. By Assumption (a-2) again, we then can obtain $\mathbb{E}\{\mathbf{e}(t)\mathbf{w}(t)^T\} = \mathbf{0}$. Using the induction argument on t, we can complete the proof.

ii) We use induction on t to prove this lemma. First, considering $t=0$, we use (33) and Assumption (a-2) to derive that $\mathbb{E}\{\mathbf{e}(1)\hat{\mathbf{e}}(0)^T\} = \mathbf{Q}\mathbb{E}\{\mathbf{e}(0)\mathbf{e}(0)^T\}$, since $\mathbf{e}(0)=\hat{\mathbf{e}}(0)$. That is, statement ii) is true for $t=0$. Next, assume that statement ii) holds for some $t>0$, and consider the case for $t+1$. Now applying statement i) to (16) and (33) yields $\mathbb{E}\{\mathbf{e}(t+2)\hat{\mathbf{e}}(t+1)^T\} = \mathbf{Q}\mathbb{E}\{\mathbf{e}(t+1)\hat{\mathbf{e}}(t+1)^T\} = \mathbf{Q}\mathbb{E}\{\mathbf{e}(t+1)\hat{\mathbf{e}}(t)^T\}(\mathbf{I}-\mathbf{\Lambda}_t) + \mathbf{Q}\mathbb{E}\{\mathbf{e}(t+1)\mathbf{e}(t+1)^T\}\mathbf{\Lambda}_t = \mathbf{Q}^{t+2}\mathbb{E}\{\mathbf{e}(0)\mathbf{e}(0)^T\}\mathbf{\Phi}_{t,0} + \sum_{s=1}^{t+1}\mathbf{Q}^{t-s+2}\mathbb{E}\{\mathbf{e}(s)\mathbf{e}(s)^T\}\mathbf{\Phi}_{t-1,s}$, where the last equality follows by the inductive hypothesis.

D. Proof of Lemma 5

i) It follows from (2) and (33) that for all $t > 0$,

$$\|\mathbf{e}(t)\| \leq \|\mathbf{Q}^t\|\|\mathbf{z}(0)\| + \alpha\sqrt{n}\Delta\|\mathbf{L}_{\text{aug}}\| \sum_{k=0}^{t-1} \|\mathbf{Q}^k\|. \quad (34)$$

By Lemma 2, we can show that $\|\mathbf{Q}^t\| \leq (\eta+1)^{n-1}$ for $t < n$, and $\leq e^{\eta}[(1-n)/(e\log\rho_{\mathbf{Q}})]^{n-1}$, for $t \geq n$. Moreover, following a similar argument as in the proof of [27, Lemma D.1], one can show that $\sum_{k=n}^{\infty} k^{n-1}\rho_{\mathbf{Q}}^k \leq (\frac{1-n}{e\log\rho_{\mathbf{Q}}})^{n-1} + \sum_{k=0}^{n-1}\frac{(n-1)!n^k\rho_{\mathbf{Q}}^n}{k!(-\log\rho_{\mathbf{Q}})^{n-k-1}}$. Noting that $\sum_{k=0}^{n-1}\|\mathbf{Q}^k\| \leq (\eta+1)^{n-1}\frac{1-\rho_{\mathbf{Q}}^n}{1-\rho_{\mathbf{Q}}}$, we then use Lemma 2 to derive (14).

ii) Let us consider (15). By Lemma 4-i), and recalling that $\mathbf{z}(0) = [\mathbf{x}(0)^T, \mathbf{0}]^T$, we get

$$\begin{aligned}\mathbb{E}\{\|\mathbf{e}(t)\|^2\} &= \mathbb{E}\left\{\left\|\mathbf{Q}^t\mathbf{z}(0)\right\|^2\right\} \\ &\quad + \alpha^2\sum_{k=0}^{t-1}\mathbb{E}\left\{\left\|\mathbf{Q}^{t-1-k}\mathbf{L}_{\text{aug}}\mathbf{w}(k)\right\|^2\right\} \\ &\leq \left\|\mathbf{Q}^t\right\|_2^2\mathbb{E}\{\|\mathbf{x}(0)\|^2\} \\ &\quad + \frac{n\alpha^2\|\mathbf{L}_{\text{aug}}\|_2^2\sup_t \text{tr}(\mathbf{W}_t)}{2}\sum_{k=0}^{t-1}\|\mathbf{Q}^k\|_2^2.\end{aligned}$$

Since $\rho_{\mathbf{Q}} < 1$ by Lemma 2, by taking limits on both sides of the previous relation, we can show that (15) holds.

E. Proof of Claim (22)

We consider three cases:

Case I: $1-\rho_{\mathbf{Q}} < \inf_{t\geq 0}\delta_t < 1$. In this case, we have $b_s = (\rho_{\mathbf{Q}}/(1-\delta_s))(1+1/s)^{n-1}b_{s-1}, \quad \forall s \geq 1$. Obviously, we have $\lim_{s\to\infty} b_s = \infty$. Hence, one can obtain $\sum_{s=n-1}^{t-1} b_s \geq \sum_{s=n-1}^{t-1}\frac{\rho_{\mathbf{Q}}}{1-\delta_s}b_{s-1} + b_{n-1} \geq \frac{\rho_{\mathbf{Q}}}{1-\inf_t\delta_t}(\sum_{s=n-1}^{t-1} b_s - b_{t-1}) + b_{n-1}$. After rearranging the terms, we have $\sum_{s=n-1}^{t-1} b_s \leq \frac{\rho_{\mathbf{Q}}b_{t-1}-(1-\inf_t\delta_t)b_{n-1}}{\rho_{\mathbf{Q}}-1+\inf_t\delta_t} \leq \frac{\rho_{\mathbf{Q}}b_{t-1}}{\rho_{\mathbf{Q}}-1+\inf_t\delta_t}$, since $1 < \rho_{\mathbf{Q}}/(1-\inf_t\delta_t) < \infty$.

Case II: $0 < \sup_{t\geq 0}\delta_t < 1-\rho_{\mathbf{Q}}$. It is easy to check that $\limsup_{s\to\infty}\frac{b_{s+1}}{b_s} = \rho_{\mathbf{Q}}\limsup_{s\to\infty}\frac{1}{1-\delta_s} == \frac{\rho_{\mathbf{Q}}}{1-\limsup_{s\to\infty}\delta_s} < 1$. The above relation shows that the series $\sum_{s=n-1}^{t-1} b_s$ is convergent, i.e., $\sum_{s=n-1}^{\infty} b_s < \infty$.

Case III: $0 < \inf_{t\geq 0}\delta_t \leq \sup_{t\geq 0}\delta_t = 1-\rho_{\mathbf{Q}}$. In this case, we use the monotone property of s^n over the interval $(0,\infty)$ to get $\sum_{s=n-1}^{t-1} b_s \leq \sum_{s=n-1}^{t}(s+1)^{n-1} \leq \int_{n-1}^{t} s^{n-1}ds \leq (1/n)\big(t^n - (n-1)^n\big)$.

Combining the above three cases, we complete the proof.

F. Proof of Theorem 3

We start the relation (4) recursively to obtain

$$\hat{\mathbf{e}}(t) = \frac{1}{t+1}\sum_{k=0}^{t}\mathbf{Q}^k\mathbf{e}(0) + \frac{\alpha}{t+1}\sum_{k=0}^{t}\sum_{h=0}^{k-1}\mathbf{Q}^{k-1-h}\mathbf{L}_{\text{aug}}\mathbf{w}(k). \quad (35)$$

By Lemma 2, we know that $\mathbf{I}-\mathbf{Q}$ is nonsingular. Taking expectation on both sides of (35), and noting that $\mathbb{E}\{\mathbf{w}(t)\} = 0$ by Assumption 2, one obtains $\|\mathbb{E}\{\hat{\mathbf{e}}(t)\}\| = (1/(t+1))\|(\mathbf{I}-\mathbf{Q}^{t+1})(\mathbf{I}-\mathbf{Q})^{-1}\mathbb{E}\{\hat{\mathbf{e}}(0)\}\| = \mathcal{O}(1/t)$, for arbitrary $\hat{\mathbf{e}}(0)$. Specifically, for $x_i(0) = \hat{x}_i(0) = y_i, \forall i$, one has $\bar{x}(0) = \hat{\theta}$. The fact that $\mathbb{E}\{\hat{\theta}\} = \theta$ then implies $\mathbb{E}\{\hat{\mathbf{e}}(0)\} = \mathbf{0}$. This show that $\hat{x}_i^{\text{RA}}$ is unbiased for each node i.

To show the mean square performance, we use similar arguments as in [27] to obtain that $\limsup_{t\to\infty} t\|\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\}\| \leq \alpha^2\|(\mathbf{I}-\mathbf{Q})^{-1}\|_2^2\|\mathbf{L}_{\text{aug}}\|_2^2\sup_t \text{tr}(\mathbf{W}_t)$. Note that $\mathbb{E}\{\|\hat{\mathbf{e}}(t)\|^2\} = \mathbb{E}\{\|\hat{\mathbf{x}}(t)-\hat{\theta}\mathbf{1}\|^2\} + \mathbb{E}\{\|\hat{\mathbf{s}}(t)\|^2\}$, the theorem thus follows.

References

[1] J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 27–41, Jul. 2006.

[2] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "Convergence of asynchronous distributed gradient methods over stochastic networks," *IEEE Trans. Automat. Control*, vol. 63, no. 2, pp. 434–448, Feb. 2017.

[3] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "A Bregman splitting scheme for distributed optimization over networks," *IEEE Trans. Automat. Control*, 2018, to be published, doi: 10.1109/TAC.2018.2805260.

[4] A. Ribeiro, I. D. Schizas, S. I. Roumeliotis, and G. B. Giannakis, "Kalman filtering in wireless sensor networks," *IEEE Control Syst.*, vol. 30, no. 2, pp. 66–86, Apr. 2010.

[5] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. Int. Conf. Inf. Process. Sensor Netw.*, Los Angeles, CA, USA, Apr. 2005, pp. 63–70.

[6] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. New York, NY, USA: Wiley, 2001.

[7] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-Part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[8] 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.

[9] 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.

[10] W. Ren and R. W. Beard, *Distributed Consensus in Multi-vehicle Cooperative Control: Theory and Applications*. London, U.K.: Springer, 2008.

[11] S. Zhu and B. Chen, "Distributed average consensus with bounded quantizer and unbounded input," arXiv preprint arXiv:1602.04193, 2017.

[12] M. E. Chamie, J. Liu, and T. Başar, "Design and analysis of distributed averaging with quantized communication," *IEEE Trans. Automat. Control*, vol. 61, no. 12, pp. 3870–3884, Dec. 2016.

[13] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Trans. Automat. Control*, vol. 54, no. 11, pp. 2506–2517, Nov. 2009.

[14] 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.

[15] S. Dasarathan, C. Tepedelenliolu, M. K. Banavar, and A. Spanias, "Nonlinear distributed average consensus using bounded transmissions," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 6000–6009, Dec. 2013.

[16] R. Carli, F. Fagnani, P. Frasca, and S. Zampieri, "Gossip consensus algorithms via quantized communication," *Automatica*, vol. 46, no. 1, pp. 70–80, 2010.

[17] S. Liu, T. Li, L. Xie, M. Fu, and J. Zhang, "Continuous-time and sampled-data based average consensus with logarithmic quantizers," *Automatica*, vol. 49, no. 11, pp. 3329–3336, 2013.

[18] M. E. Yildiz and A. Scaglione, "Coding with side information for rate-constrained consensus," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3753–3764, Aug. 2008.

[19] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1383–1400, Mar. 2010.

[20] L. Xie, D.-H. Choi, S. Kar, and H. V. Poor, “Fully distributed state estimation for wide-area monitoring systems,” *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1154–1169, Sep. 2012.

[21] 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.

[22] S. Zhu, C. Chen, X. Ma, B. Yang, and X. Guan, “Consensus based estimation over relay assisted sensor networks for situation monitoring,” *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 2, pp. 278–291, Mar. 2015.

[23] T. Li, M. Fu, L. Xie, and J.-F. Zhang, “Distributed consensus with limited communication data rate,” *IEEE Trans. Automat. Control*, vol. 56, no. 2, pp. 279–292, Feb. 2011.

[24] T. Li and L. Xie, “Distributed consensus over digital networks with limited bandwidth and time-varying topologies,” *Automatica*, vol. 47, no. 9, pp. 2006–2015, 2011.

[25] K. Cai and H. Ishii, “Quantized consensus and averaging on gossip digraphs,” *IEEE Trans. Automat. Control*, vol. 56, no. 9, pp. 2087–2100, Sep. 2011.

[26] D. Li, Q. Liu, X. Wang, and Z. Lin, “Consensus seeking over directed networks with limited information communication,” *Automatica*, vol. 49, no. 2, pp. 610–618, 2013.

[27] S. Zhu, Y. C. Soh, and L. Xie, “Distributed parameter estimation with quantized communication via running average,” *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4634–4646, Sep. 2015.

[28] S. Zhu, S. Liu, Y. C. Soh, and L. Xie, “Performance analysis of averaging based distributed estimation algorithm with additive quantization model,” *Automatica*, vol. 80, pp. 95–101, 2017.

[29] S. Zhu, J. Xu, C. Chen, and X. Guan, “A least square approach for distributed sensor fusion in bandwidth-constrained sensor networks,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, Mar. 2016, pp. 33 751–3755.

[30] M. Huang, “Stochastic approximation for consensus: A new approach via ergodic backward products,” *IEEE Trans. Automat. Control*, vol. 57, no. 12, pp. 2994–3008, Dec. 2012.

[31] J. Fang and H. Li, “Distributed consensus with quantized data via sequence averaging,” *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 944–948, Feb. 2010.

[32] K. Cai and H. Ishii, “Average consensus on general strongly connected digraphs,” *Automatica*, vol. 48, no. 11, pp. 2750–2761, 2012.

[33] J. M. Hendrickx and J. N. Tsitsiklis, “Fundamental limitations for anonymous distributed systems with broadcast communications,” in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep./Oct. 2015, pp. 9–16.

[34] 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 Process.*, vol. 25, no. 5, pp. 442–448, Oct. 1977.

[35] R. M. Gray and T. G. Stocham, “Dithered quantizers,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, May 1993.

[36] A. Gut, *Probability: A Graduate Course*, 2nd ed. New York, NY, USA: Springer, 2013.

[37] Y. Huang and Y. Hua, “Multihop progressive decentralized estimation in wireless sensor networks,” *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 1004–1007, Dec. 2007.

[38] S. Olariu and I. Stojmenović, “Design guidelines for maximizing lifetime and avoiding energy holes in sensor networks with uniform distribution and uniform reporting,” in *Proc. IEEE Int. Conf. Comput. Commun.*, Barcelona, Spain, Apr. 2006, pp. 1–12.

[39] Z. Xu, L. Chen, C. Chen, and X. Guan, “Joint clustering and routing design for reliable and efficient data collection in large-scale wireless sensor networks,” *IEEE Internet Things J.*, vol. 3, no. 4, pp. 520–532, Aug. 2016.

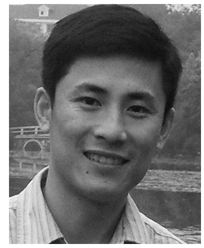
[40] V. Mhatre and C. Rosenberg, “Design guidelines for wireless sensor networks: communication, clustering and aggregation,” *Ad Hoc Netw.*, vol. 2, no. 1, pp. 45–63, 2004.

[41] B. T. Polyak, *Introduction to Optimization*. New York, NY, USA: Optimization Software, 1987.

[42] V. Fabian, “Approximation of minima with improved asymptotic speed,” *Ann. Math. Stat.*, vol. 38, no. 1, pp. 191–200, 1967.

[43] P. Henrici, “Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices,” *Numer. Math.*, vol. 4, no. 1, pp. 24–40, 1962.

[44] M. Gil, “Estimate for the norm of matrix-valued functions,” *Linear Multilinear Algebra*, vol. 35, no. 1, pp. 65–73, 1993.



Shanying Zhu (M’15) received the B.S. degree in information and computing science from the North China University of Water Resources and Electric Power, Zhengzhou, China, in 2006, the M.S. degree in applied mathematics from the Huazhong University of Science and Technology, Wuhan, China, in 2008, and the Ph.D. degree in control theory and control engineering from Shanghai Jiao Tong University, Shanghai, China, in 2013. From 2013 to 2015, he was a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, and also with the Berkeley Education Alliance for Research in Singapore, Singapore. In 2015, he joined Shanghai Jiao Tong University, China, where he is currently an Associate Professor with the Department of Automation, School of Electronic Information and Electrical Engineering. His research interests include multi-agent systems and wireless sensor networks, particularly in coordination control of mobile robots and distributed detection and estimation in sensor networks and their applications in industrial networks.

He has served as the Local Arrangement Co-chair of WiOpt’2018, session chair of CDC’2016, ICCA’2018, etc. He is the member of IEEE IES Technical Committee on Industrial Informatics, a member of Youth Committee of Chinese Association of Automation.

Cailian Chen (M’06) received the B.Eng. and M.Eng. degrees in automatic control from Yanshan University, Qinhuangdao, China, in 2000 and 2002, respectively, and the Ph.D. degree in control and systems from the City University of Hong Kong, Hong Kong, in 2006. She was a Senior Research Associate with the City University of Hong Kong, Hong Kong, in 2006. From 2006 to 2008, she was a Postdoctoral Research Associate with the University of Manchester, U.K. In 2008, she joined the Department of Automation, Shanghai Jiao Tong University, Shanghai, China as an Associate Professor, where she is currently a Full Professor. She was a Visiting Professor with the University of Waterloo, Waterloo, ON, Canada, from 2013 to 2014. Her research interests include wireless sensor networks and industrial applications, computational intelligence and distributed situation awareness, cognitive radio networks and system design, internet of vehicles and applications in intelligent transportation, and distributed optimization. She has authored and/or coauthored 2 research monographs and more than 100 referred international journal and conference papers. She is the inventor of more than 20 patents.

She is the recipient of the prestigious “IEEE Transactions on Fuzzy Systems Outstanding Paper Award” in 2008, and “Best Paper Award of the Ninth International Conference on Wireless Communications and Signal Processing” in 2017. She won the First Prize of Natural Science Award twice from The Ministry of Education of China in 2006 and 2016, respectively. She was honored “Changjiang Young Scholar” by the Ministry of Education of China, in 2015 and “Excellent Young Researcher” by NSF of China in 2016. She has been actively involved in various professional services. She serves as an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Peer-to-peer Networking and Applications* (Springer), *The World Scientific Journal: Computer Science*, and *ISRN Sensor Networks*. She also served as a Guest Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, Symposium TPC Co-chair of IEEE Globecom 2016 and VTC2016-fall, Workshop Co-chair of WiOpt’18, and TPC member of many flagship conferences including IEEE Globecom, IEEE ICC, IEEE VTC, ICCVE, and IEEE WCCI.

Jinming Xu received the B.S. degree in mechanical engineering from Shandong University, Jinan, China, in 2009 and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2016. He was a Research Fellow with the Internet of Things Lab, NTU, from fall 2016 to fall 2017. He is currently a Postdoctoral Scholar with the Ira A. Fulton Schools of Engineering, Arizona State University, Tempe, AZ, USA. His research interests include distributed optimization and its applications into large-scale machine learning and signal processing, with a recent focus on cybersecurity of networked systems.

Xinping Guan (F'18) received the Ph.D. degree in control and systems from Harbin Institute of Technology, Harbin, China in 1999. He is currently a Chair Professor of Shanghai Jiao Tong University, Shanghai, China, where he is the Deputy Director of University Research Management Office, and the Director of the Key Laboratory of Systems Control and Information Processing, Ministry of Education of China. From 1998 to 2008, he was a Professor and the Dean of School of Electrical Engineering, Yanshan University, China.

His current research interests include industrial cyberphysical systems, wireless networking and applications in smart city and smart factory, and underwater sensor networks. He has authored and/or coauthored 4 research monographs, more than 270 papers in IEEE Transactions and other peer-reviewed journals, and numerous conference papers. As a Principal Investigator, he has finished/been working on many national key projects. He is the Leader of the prestigious Innovative Research Team of the National Natural Science Foundation of China. He is an Executive Committee Member of Chinese Automation Association Council and the Chinese Artificial Intelligence Association Council. He received the First Prize of Natural Science Award from the Ministry of Education of China, in 2006 and 2016, and the Second Prize of the National Natural Science Award of China, in 2008. He was the recipient of the "IEEE Transactions on Fuzzy Systems Outstanding Paper Award" in 2008. He is a "National Outstanding Youth" honored by NSF of China, "Changjiang Scholar" by the Ministry of Education of China, and "State-level Scholar" of "New Century Bai Qianwan Talent Program" of China.

Lihua Xie (F'07) received the B.E. and M.E. degrees in electrical engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1983 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Newcastle, Callaghan, NSW, Australia, in 1992. Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a Professor and the Director with the Delta-NTU Corporate Laboratory for cyberphysical systems. From 1986 to 1989, he held teaching appointments with the Department of Automatic Control, Nanjing University of Science and Technology. From July 2011 to June 2014, he was the Head of the Division of Control and Instrumentation. His research interests include robust control and estimation, networked control systems, multi-agent networks, and localization and unmanned systems.

He is an Editor-in-Chief for Unmanned Systems and an Associate Editor for the IEEE TRANSACTIONS ON NETWORK CONTROL SYSTEMS. He was an Editor of IET Book Series in Control and an Associate Editor of a number of journals including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Automatica*, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II. He is an Elected Member of the Board of Governors, IEEE Control System Society (January 2016–December 2018). He is a Fellow of IFAC.

Karl Henrik Johansson (F'13) received the M.Sc. and Ph.D. degrees from Lund University, Lund, Sweden. He is the Director with the Stockholm Strategic Research Area ICT-The Next Generation and a Professor with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests include networked control systems, cyberphysical systems, and applications in transportation, energy, and automation. He is a member of the IEEE Control Systems Society Board of Governors, the IFAC Executive Board, and the European Control Association Council. He has received several best paper awards and other distinctions. He is a Distinguished Professor with the Swedish Research Council and a Wallenberg Scholar and he was the recipient of the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He is a Fellow of the Royal Swedish Academy of Engineering Sciences, and an IEEE Distinguished Lecturer.