Distributed Parameter Estimation with Quantized Communication via Running Average
Shanying Zhu, Yeng Chai Soh, Senior Member, IEEE, and Lihua Xie, Fellow, IEEE

Abstract—In this paper, we consider the parameter estimation problem over sensor networks in the presence of quantized data and directed communication links. We propose a two-stage algorithm aiming at achieving the centralized sample mean estimate in a distributed manner. Different from the existing algorithms, a running average technique is utilized in the proposed algorithm to smoothen out the randomness caused by the probabilistic quantization scheme. With the running average technique, it is shown that the centralized sample mean estimate can be achieved both in the mean square and almost sure senses, which is not observed in the conventional consensus algorithms. In addition, the rates of convergence are given to quantify the mean square and almost sure performances. Finally, simulation results are presented to illustrate the effectiveness of the proposed algorithm and highlight the improvements by using running average technique.

Index Terms—Distributed estimation, probabilistic quantization, running average, directed topology

I. INTRODUCTION

Sensor networks, composed of a large number of signal processing devices (nodes), are massively distributed systems for sensing and processing of spatially dense data with wide applications both in military and civilian scenarios. A popular application of sensor networks is decentralized estimation of unknown parameters using samples collected from nodes [1]–[4]. Two prevailing topologies for such task are fusion center based networks and ad hoc networks [3]. Compared with fusion center based networks, ad hoc networks have several advantages including scalability and resilience of node failure. In a typical estimation problem in ad hoc networks, nodes make noisy measurements of variables of interest. The main concern is how to utilize the samples to produce a desired estimate by only exchanging data between neighboring nodes.

Distributed estimation in ad hoc networks is usually based on successive refinements of local estimates maintained at individual nodes. In most applications, nodes are powered by batteries with finite lifetime and thus have limited computing and communication capabilities. Another aspect is bandwidth constraint, which renders the transmission of large volume of real-valued data impractical. This means that the data exchanged between nodes needs to be quantized prior to transmission. However, this process introduces certain quantization errors which could have severe effects. The errors will be accumulated throughout the successive iterations, making the estimation process fluctuating or even divergent [5].

A number of distributed consensus algorithms have been proposed to address the problem of estimation with quantized communication. Most of them assume symmetric communication between nodes. Actually, in ad hoc networks, communication links between certain pairs of nodes may be directed, i.e., a node can receive information from another node but not vice versa. This could be caused by non-homogeneous interference, packet collision and so on. Motivated by this observation, in the paper, we consider the problem of distributed estimation over directed topologies and examine its convergence behavior under the effect of quantized communication.

A. Related work

Distributed consensus algorithms are powerful tools to solve the estimation problems in sensor networks, where the final states are mostly chosen as the estimates. Recently, much attention has been paid to the effect of quantization. For instance, deterministic quantization schemes are used in [6]–[9]. In particular, uniform and truncation quantizers were investigated in [6], [7], [9], where convergence can only be guaranteed up to a neighborhood of the target average and upper bounds characterizing the gaps were provided. Ref. [8] considered the logarithmic quantization scheme, which showed that the consensus error is upper bounded by a quantity depending on the quantization resolution and initial states. In [10], a quantized consensus algorithm was introduced with an additional constraint that the states of the nodes are integers. This constraint leads to an integer approximation of the target average. Extension to the directed topologies has been examined in [11].

Another thread is to adopt probabilistic quantization schemes. In [12], the dithered quantization scheme was used. It was shown that consensus to a random variable whose expectation is equal to the desired average can be reached almost surely. This kind of convergence was also observed for gossip algorithms [6]. In fact, even employing the decaying link weights satisfying a persistence condition cannot guarantee the convergence to the target average [13]. By exploring the temporal information of the successive states, the authors in [14] showed that the target average can be obtained in the mean square sense by introducing a new quantity as the estimate. Most of the above works assume that
the communication topology is symmetric. This assumption may not be realistic as discussed previously. Moreover, the symmetric requirement imposes much effort on the nodes to acquire necessary topology information to construct the weight matrices. Even the symmetric communication is assumed, the aforementioned results indicate that convergence to the target average is not possible in most cases using simple quantizers.

To further address the residual issue of quantization, dynamic encoding/decoding schemes were proposed in [15], [16] to ensure the convergence to the desired average value. Specifically, Ref. [16] showed that the number of quantization bits can be reduced merely to one by appropriately designing the scaling function and some control parameters. The result of [16] has been extended to directed graphs in [17], where the weighted average instead of the desired average was shown to be achievable. Although the dynamic quantizations perform quite well, some spectral properties of the Laplacian matrix of the underlying topology have to be known in advance based on which the encoder-decoder parameters are carefully chosen. A similar idea was adopted in [18] to design a progressive quantization schemes are possible to solve the parameter estimation problems over networks, provided that a suitable form of estimator is introduced.

The paper is organized as follows: In Section II we present the problem formulation and some preliminary results needed in the subsequent sections. In Section III we describe the proposed two-stage distributed algorithm along with some implementation considerations. Convergence analysis both in the mean square and almost sure senses is presented in Section IV. Section V presents the simulation results to illustrate the effectiveness of the proposed algorithm, followed by the conclusion in Section VI.

**Notation:** \( \mathbb{Z}_{ \geq a} \) stands for the subset of integers greater than \( a \). For two functions \( f(k) \) and \( g(k) \), \( f(k) = o(g(k)) \) means that \( \lim_{k \to \infty} f(k)/g(k) = 0 \). We use \( O(1) \) to denote a constant, which may vary at different places. \( \mathbb{R}^n \) is the \( n \)-dimensional Euclidean space endowed with \( l_2 \) norm \( \| \cdot \|_2 \). \( \mathbb{R}^{m \times n} \) denotes the set of all \( m \times n \) matrices with the Euclidean norm \( \| \cdot \|_2 \) and Frobenious norm \( \| \cdot \|_F \). We use bold uppercase and lowercase letters to denote matrices and vectors, respectively. In particular, \( \mathbf{I} \) is the identity matrix, \( \mathbf{1} \), \( \mathbf{0} \) are all-one and all-zero vectors, respectively. \( \lambda_{\text{max}}(\cdot) \) represents the largest eigenvalue of a symmetric matrix. For a random vector \( \mathbf{x} \), \( \mathbb{E}\{\mathbf{x}\} \) denotes its expectation and \( \text{Cov}(\mathbf{x}) \) its covariance.

**B. Summary of contributions**

In this paper, we consider the parameter estimation problem over directed communication topologies. Each node has real-valued states but can only exchange information with its neighbors utilizing quantized communication. Different from the conventional consensus algorithms, we introduce a new quantity rather than the final state as the estimate. The main contributions are summarized as follows:

Firstly, we propose a two-stage distributed estimation algorithm in which the nodes utilize probabilistic quantization. At the first stage, we estimate the left eigenvector with respect to the zero eigenvalue of the Laplacian matrix. This information is then used at the second stage to construct a correction term aiming at compensating the unidirectional effect of directed communication links. At both stages, the running average technique is utilized to limit the quantization effect on the estimation process such that the centralized estimate can be achieved in a distributed manner. Unlike [4–9], [12], [13], our algorithm does not require the weight matrix to be doubly stochastic. And it can be run over any strongly connected topologies without any knowledge of the out-neighbor information and the left eigenvector of the corresponding Laplacian matrix as those in [11], [17].

Secondly, a comprehensive convergence analysis of the proposed algorithm is given. With the running average technique, we show that the centralized sample mean estimate can be achieved exactly both in the mean square and almost sure senses. The results extend the one in [14] from undirected graphs to directed graphs. Moreover, the proposed algorithm does not depend on the complicated design of dynamic encoding/decoding schemes as in [15–18]. Our analysis relies on the theoretical tools of the laws of large numbers and the iterated logarithm. The theoretical results imply that simple quantization schemes are possible to solve the parameter estimation problems over networks, provided that a suitable form of estimator is introduced.

**II. Problem formulation**

Consider the estimation problem in a sensor network consisting of \( n \) homogeneous nodes, each making observations of an unknown parameter \( \theta \in \mathbb{R} \). The observations are corrupted by additive noises

\[
y_i = \theta + w_i, \quad i = 1, 2, \ldots, n,
\]

where \( w_i \) are zero mean, independent and identically distributed Gaussian noises. In the centralized case, it is shown that the linear minimum mean square error estimate can be computed using the sample mean estimator \( \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} y_i \), which is the best in the sense of Cramér-Rao lower bound [19]. This estimator is universal since it does not require any information of noises [20].

The distributed estimation problem is concerned with computing the centralized sample mean estimate \( \hat{\theta} \) iteratively at every node without requiring global knowledge of \( \{y_i\}_{i=1}^{n} \). We model the communication topology over which the nodes exchange information as a weighted directed graph \( \mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A}) \), where \( \mathcal{V} = \{1, 2, \ldots, n\} \) is the set of nodes, \( \mathcal{E} \subset \mathcal{V} \times \mathcal{V} \) denotes all the unidirectional communication links between nodes and \( \mathcal{A} = [a_{ij}]_{n \times n} \) is composed of weights \( a_{ij} > 0 \) associated with each directed edge \( (j, i) \in \mathcal{E} \). It is assumed that there are no self-loops in \( \mathcal{G} \), i.e., \( a_{ii} = 0 \) for all \( i \in \mathcal{V} \). The directed edge \( (j, i) \) means that node \( i \) can receive data from node \( j \). We denote \( \mathcal{N}_j \) as the set of neighbors of node \( i \). As for the graph, we assume that:

**Assumption 1:** Graph \( \mathcal{G} \) is strongly connected, i.e., for any two nodes \( i \) and \( j \), there exists a directed path from \( i \) to \( j \).

In the case of limited communication rate between nodes, each node will first quantize the data prior to its transmission to the neighbors. In this paper, we adopt the following estimation
algorithm at each node $i$,
\begin{equation}
    x_i(t+1) = \hat{x}_i(t) + \alpha \sum_{j \in \mathcal{N}_i} a_{ij} [Q(\hat{x}_j(t)) - Q(\hat{x}_i(t))],
\end{equation}
with initial guess $x_i(0) = y_i$, where $\alpha > 0$ is the weight, $Q(\cdot)$ denotes the quantization operation and $\hat{x}_i(t) \triangleq x_i(t) + \epsilon_i(t)$, in which $\epsilon_i(t)$ is a correction term to compensate for the unidirectional effects of communication links. The goal is to design an appropriate $\epsilon_i(t)$ such that all the nodes can acquire a reliable estimate of the centralized $\tilde{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i(0)$.

**Remark 1:** The algorithm (1) belongs to the compensating update rule [6], [14], [19], [21], where both the real-valued states and their quantized values are used to compute the states at next step. This strategy is meant to fully exploit the implicit channel feedback which comes from quantization.

**A. Probabilistic quantization**

In this section, we present a brief review of the quantization scheme used in the paper. Each node is equipped with a probabilistic quantizer $Q(\cdot) : \mathbb{R} \to \mathcal{S}_\Delta$ with the set of quantization levels $\mathcal{S}_\Delta = \{k\Delta : k \in \mathbb{Z}\}$, where $\Delta$ is the quantization step-size. For any $x \in \mathbb{R}$, it is quantized in a probabilistic manner:
\begin{equation}
    Q(x) = \begin{cases} 
    \lceil x \rceil \Delta, & \text{with probability } p, \\
    \lfloor x \rfloor \Delta, & \text{with probability } 1-p,
\end{cases}
\end{equation}
where $p = x/\Delta - \lfloor x/\Delta \rfloor$, $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ denote the floor and ceiling functions, respectively. We can prove that the quantized message $Q(x)$ is an unbiased estimator of $x$ with finite variance [6], [12], that is,
\begin{equation}
    \mathbb{E}\{Q(x)\} = x, \quad \mathbb{E}\{(Q(x) - x)^2\} \leq \frac{\Delta^2}{4}.
\end{equation}
Further, it is obvious that
\begin{equation}
    |Q(x) - x| \leq \Delta.
\end{equation}

Actually, the above quantization is equivalent to a subtrac- tively dithered method [12]. If the dither sequence satisfies the Schuchman conditions, then the quantization errors are statistically independent from each other and the input [22]. We make the following natural assumption of statistical independence:

**Assumption 2:** The quantization errors are independent from the data and are spatially and temporally independent.

**B. Averaging technique**

Distributed consensus with quantized transmission over undirected graphs has been studied recently. However, regardless of the type of quantizers (deterministic or probabilistic), the residue between the final state and the desired average will always occur. This means that the exact $\theta$ cannot be achieved even for undirected graphs in the presence of quantized data unless a dynamic quantizer is applied [15]. We need to find another form of estimator to tackle the quantization issue.

Statistics tells us that large samples have smoothing effects: The wild randomness that always exits in small samples will be smeared out [23]. This temporal information has been used in [14] to investigate the consensus seeking over undirected graphs. This observation motivates us to adopt the following running average to smooth the samples
\begin{equation}
    \bar{x}_i(K) \triangleq \frac{1}{K} \sum_{t=t_0}^{t_0+K-1} x_i(t), \quad \forall i = 1, 2, \ldots, n,
\end{equation}
where $K \in \mathbb{Z}_{\geq 1}$ is the window size and $t_0 \in \mathbb{Z}_{\geq 1}$ is the starting point of the averaging process. The new quantity $\bar{x}_i$ will be used as the estimate of $\theta$ at node $i$, which is different from those used for conventional consensus algorithms.

**C. Preliminaries**

One important concept for distributed algorithms is the Laplacian $L$ corresponding to graph $G$, which is defined as $L \triangleq D - A$, where $D \triangleq \text{diag}\{d_1, d_2, \ldots, d_n\}$ with $d_i = \sum_{j \in \mathcal{N}_i} a_{ij}$ denoting the in-degree of node $i$. It is clear that $L1 = 0$, that is, 0 is an eigenvalue of $L$.

**Lemma 1:** Let $\omega = [\omega_1, \omega_2, \ldots, \omega_n]^T$ be the left eigenvector corresponding to the zero eigenvalue of $L$ with $1^T\omega = 1$. Then under Assumption 1, $\omega$ is positive and the matrix $Q \triangleq P - \omega \omega^T$ with $P \triangleq I - \alpha L$ has the following properties:

i) **Spectrum:** Let $0 = \lambda_1(L), \lambda_2(L), \ldots, \lambda_c(L)$ be the eigenvalues of Laplacian $L$, then the spectrum of $Q$ is $\{0, 1 - \alpha \lambda_i(L), i = 2, 3, \ldots, n\}$.

ii) **Spectral radius:** The spectral radius $\rho(Q)$ is bounded by $\rho(Q) \leq \rho(Q^k) \leq c\omega \omega^T$.

iii) **Bounds on Frobenious norm:** The Frobenious norm of power $Q^k$, $\forall k \in \mathbb{Z}_{\geq 1}$, is bounded by
\begin{equation}
    \rho_k(Q) \leq \|Q^k\|_F \leq nc \omega \omega^T \rho_0(Q),
\end{equation}

where $c$ is a constant depending only on $Q$ and $q \triangleq \max_{\lambda_i(Q) \neq 0} \{q_i\}$, $q_i$ is the multiplicity of $\lambda_i(Q)$ in the minimal polynomial of $Q$.

**Proof:** See Appendix A.

The next lemma presents a way to choose the parameter $\alpha$ such that $Q$ has some desired properties as given in Lemma 1.

**Lemma 2:** Let $0 < \alpha < \frac{1}{\max_{d_i, d_j} d_i d_j}$, then under Assumption 1, $P$ is a nonnegative matrix and $\rho(Q) < 1$. Further, for all $k \in \mathbb{Z}_{\geq 1}$,
\begin{equation}
    \|I - Q^k\|_F \leq \sqrt{n} + 2 + n^2 c \omega \omega^T (1 - q)/(\epsilon \ln(\rho(Q)))^{2(q-1)},
\end{equation}
where $c \omega \omega^T \triangleq \sqrt{n} + 2 + n^2 c \omega \omega^T (1 - q)/(\epsilon \ln(\rho(Q)))^{2(q-1)}$.

**Proof:** See Appendix B.

**III. DISTRIBUTED ESTIMATION ALGORITHM OVER DIRECTED TOPOLOGIES VIA RUNNING AVERAGE**

In this section, the averaging technique proposed in the previous section is applied to the estimation problem with the aim of achieving the centralized sample mean estimate in a distributed manner over directed communication topologies.

The proposed algorithm is composed of two stages: At the first stage, we apply the averaging technique to estimate the left eigenvector $\omega$ associated with the zero eigenvalue of Laplacian $L$; At the second stage, we design the correction
term $\epsilon(t)$ in (1) by using the estimates of $\omega$ obtained at the first stage aiming at compensating the unidirectional effect of the directed communication topologies. A distributed estimation algorithm via interwinding these two stages is then proposed.

A. Distributed estimation of the left eigenvector $\omega$

At the first stage, each node $i$ maintains a vector $z_i = [z_{i1}, z_{i2}, \ldots, z_{in}]^T$ to store the estimate of $\omega$. At each iteration, the nodes update their variables as follows:

$$z_i(t + 1) = z_i(t) + \alpha \sum_{j \in \mathcal{N}_i} a_{ij} [Q(z_j(t)) - Q(z_i(t))],$$

with initial values $z_i(0) = 1$, $z_{ij}(0) = 0$, $\forall j \neq i$, where $0 < \alpha < \frac{1}{\max_i d_i}$ and $Q(\cdot)$ is componentwise for vectors. This algorithm is inspired by [24].

In order to ensure that all nodes can achieve reliable estimates of $\omega$, it suffices to guarantee that $Z(t) = [z_1(t), z_2(t), \ldots, z_n(t)]^T$ converges to $1\omega^T$. However, as discussed previously, this is impossible due to the incurred quantization error. We apply the averaging technique proposed in Section I to this eigenvector estimation problem, and adopt the running average

$$\bar{Z}(t) = \frac{1}{t} \sum_{k=1}^{t} Z(k)$$
as the estimate of $1\omega^T$.\footnote{Actually, we can start the running average from any integer $k_0 \in \mathbb{Z}_{\geq 1}$, i.e., $Z(t) = \frac{1}{t - k_0 + 1} \sum_{k=k_0}^{t} Z(k)$. For notational simplicity, we take $k_0 = 1$ in the sequel.}

Algorithm I shows the distributed estimation algorithm of $\omega$ at the $t$-th iteration run by node $i$. In the algorithm, we use the initial value $z_{ii}(0) = n^\kappa$ with $\kappa \geq 0$ instead of the original $z_{ii}(0) = 1$ in (5). One reason is that convergence of the original $\bar{z}_i$ to $\omega_i$ is equivalent to its convergence to $n^\kappa \omega_i$ in the new scale. Introducing $n^\kappa$ into the initial states does not affect the convergence. We also impose a starting point $k_0 \in \mathbb{Z}_{\geq 1}$ for the averaging process in Algorithm I.

Algorithm 1 Distributed estimation of $\omega$ at node $i$

**Input:** $\alpha$, $n$, $\kappa$, $a_{ij}$, $K_0$.

**Output:** $\bar{z}_i(n^\kappa)$.

1: **Initialization:** $z_{ii}(0) = n^\kappa$, $z_{ij}(0) = 0$, $\forall j \neq i$.
2: Receive data from neighbors: $Q(z_j(t))$, $j \in \mathcal{N}_i$.
3: Update the estimate of $\omega$ via (5).
4: if $t < k_0$ then
5: $\bar{z}_i(t) = z_i(t)$.
6: else
7: Update the average $\bar{z}_i(t)$:
$$\bar{z}_i(t + 1) = \frac{t - k_0 + 1}{t - k_0 + 2} \bar{z}_i(t) + \frac{1}{t - k_0 + 2} z_i(t + 1).$$
8: end if

B. Design of the correction term $\epsilon(t)$

The second stage is concerned with the design of an appropriate correction term $\epsilon(t)$ to compensate the unidirectional effect of the communication links.

Let $\tau \in \mathbb{Z}_{\geq 1}$ be the integer which triggers the estimation algorithm in Algorithm 1. Motivated by the consensus algorithm proposed in [25], we design the correction term $\epsilon_i(t)$ as follows

$$\epsilon_i(t) \triangleq \begin{cases} \left( \frac{1}{n z_{ii}(t) - 1} \right) x_i(t), & t = 0, \\ \left( \frac{1}{n z_{ii}(t + \tau) - 1} - \frac{1}{n z_{ii}(t + \tau - 1)} \right) x_i(t), & t \in \mathbb{Z}_{\geq 1}. \end{cases}$$

The main idea behind the form (6) is to inject a suitable exogenous input at each iteration $t$ aiming at regulating the weighted sum such that the centralized sample mean estimate can be achieved by the running average $\bar{x}$ asymptotically.\footnote{In fact, we can start the running average from any integer $k_0 \in \mathbb{Z}_{\geq 1}$, i.e., $Z(t) = \frac{1}{t - k_0 + 1} \sum_{k=k_0}^{t} Z(k)$. For notational simplicity, we take $k_0 = 1$ in the sequel.}

One issue remaining before the implementation is the well-definedness of $\epsilon_i(t)$, $\forall i$, that is, the denominators in $\epsilon_i(t)$ must be nonzero at least almost surely. This is much involved and we will elaborate on it in Section IV (see Theorem 3).

The proposed algorithm of the $t$-th iteration run by node $i$ at the second stage is shown in Algorithm 2. In the algorithm, we modify the definition of the correction in (6) (see lines 1 and 4) to accommodate the setup in Algorithm 1.

Algorithm 2 Distributed estimation algorithm with quantization via running average at node $i$

**Input:** $\alpha$, $n$, $\kappa$, $a_{ij}$, $\tau$, $t_0$, $x_i(0)$.

**Output:** $\bar{x}_i$.

1: **Initialization:** $\epsilon_i(t) = \left( \frac{n^{-1} - 1}{x_i(t)} \right) x_i(t)$.
2: Receive data from neighbors: $Q(x_j(t) + \epsilon_j(t))$, $j \in \mathcal{N}_i$.
3: Update the estimate of $\theta$ via (1).
4: Compute the correction:
$$\epsilon_i(t + 1) = n^\kappa x_i(t) \frac{\bar{x}_i(t + \tau) - \bar{x}_i(t + \tau + 1)}{z_{ii}(t + \tau) - z_{ii}(t + \tau + 1)}.$$
5: if $t \geq t_0$ then
6: Update the average $\bar{x}_i(K)$:
$$\bar{x}_i(K + 1) = \frac{1}{K+1} \bar{x}_i(K) + \frac{1}{K+1} x_i(t).$$
7: end if

C. Summary of the algorithm

At each iteration, the proposed distributed estimation algorithm with quantized data via running average is composed of Algorithm 1 and Algorithm 2. In the algorithm, we use an increasing window size $K = t - t_0$ for the averaging process (see line 6 of Algorithm 2). A fixed window size $K$ can also be used according to what level of the convergence performance is needed. This can be inferred from the theoretical results in Section IV where the rates of convergence are established.

We remark that the adjustment of the initial values in line 1 of Algorithm 1 has another consequence. It is known that $0 < \omega_i < 1$, $\forall i$ by Lemma 1 and for certain topologies, some $\omega_i$’s are rather close to 0. It is probable that zeros would occur in the denominators of $\epsilon_i$ during the quantization process. In this case, the original correction term in (6) will be meaningless for the first several iterations and we have to wait a long time before triggering Algorithm 2. Increasing the initial values from 1 to $n^\kappa$ is meant to tackle this concern. Our simulation results validate this consideration.

We also emphasize that no further buffer is needed to store the previous states $\bar{z}_i(t)$ and $\bar{x}_i(K)$ (see line 7 of Algorithm 1).
IV. CONVERGENCE ANALYSIS OF THE PROPOSED AVERAGING BASED ALGORITHM

In this section, we first present the convergence results for the estimation algorithm of the left eigenvector $\omega$, based on which the convergence analysis of the proposed averaging based algorithm is given.

A. Convergence results for the left eigenvector estimation algorithm

Write $Q(z_i(t)) = z_i(t) + u_i(t)$, where $u_i(t)$ is the quantization error with zero mean and $E\{\|u_i(t)\|^2\} \leq n\Delta^2/4$ in view of (2). Let $U(t) \triangleq [u_1(t), u_2(t), \ldots, u_n(t)]^T$, then we can write (5) in a compact form $Z(t+1) = PU(t) - \alpha LU(t)$, with initial value $Z(0) = I$. Hence it can be derived that

$$Z(t) = \frac{1}{\alpha} \sum_{k=1}^{t} \left( P^k - \alpha \sum_{s=0}^{k-1} P^{k-s-1}LU(s) \right).$$

(7)

Define the estimation error as $e_Z(t) \triangleq Z(t) - 1/\omega^T$. Recall that $L1 = \omega^T L = 0$, it is easy to verify that $P^k - \omega^T = Q^k$ and $P^k L = Q^k L$, $\forall k \in \mathbb{Z}_{\geq 1}$. This together with (7) implies

$$e_Z(t) = \frac{1}{\alpha} \sum_{k=1}^{t} \left( Q^k - \alpha \sum_{s=0}^{k-1} Q^{k-s-1}LU(s) \right).$$

(8)

By interchanging the order of summation, we can obtain

$$\sum_{k=0}^{t} \sum_{s=0}^{k} Q^{k-s-1}LU(s) = \sum_{k=0}^{t} \sum_{s=0}^{t-k} Q^s L U(k).$$

By Lemma 2 we know that $p(Q) < 1$, which implies that $I - Q$ is nonsingular. Combining the above two relations yields

$$e_Z(t) = \frac{1}{\alpha} \sum_{k=0}^{t} Q^k (I - Q^t) - \alpha \sum_{k=0}^{t} W_t(k)LU(k),$$

(9)

where $Q \triangleq (I - Q)^{-1}$, $L \triangleq (I - Q)^{-1} L$, and $W_t(k) \triangleq I - Q^{t-k}$, for $0 \leq k \leq t - 1$.

In the following, we will focus on the convergence of $e_Z(t)$ both in the mean square and almost sure senses.

1) Mean square performance: Let $D(t) \triangleq E\{U(t)U^T(t)\}$. Under Assumption 2, we can decompose it as $D(t) = F(t)^2$, where $F(t) \triangleq \text{diag} \{E\{\|u_1(t)\|^2\}, E\{\|u_2(t)\|^2\}, \ldots, E\{\|u_n(t)\|^2\}\}$. Invoking (9) and Assumption 2 on $\{U(t)\}_{t \geq 0}$ implies

$$E\{\|e_Z(t)\|^2_F\} \leq \frac{1}{t^2} \|Q(I - Q^t)\|^2_F + \frac{\alpha^2}{t} \sum_{k=0}^{t-1} \|W_t(k)LU(k)\|^2_F.$$

We have the following result regarding the mean square convergence of $e_Z(t)$.

**Theorem 1:** Under Assumptions 1 and 2, $Z(t)$ converges in mean square to $1/\omega^T$ as $t \to \infty$. Moreover, for large $t$, the mean square deviation is approximately given by

$$E\{\|e_Z(t)\|^2_F\} \leq \frac{n\Delta^2}{4t},$$

where $\nu \triangleq \alpha\sqrt{n + 2\Delta}\|L\|_2$.

**Proof:** Using the facts that $\|BC\|^2_F = \text{tr}(CC^T B^T B) \leq \lambda_{\text{max}}(CC^T)\|B\|^2_F$, for arbitrary $B, C \in \mathbb{R}^{n \times n}$, and $\max_i E\{\|u_i(t)\|^2\} \leq n\Delta^2/4$ in light of (2), we can obtain

$$E\{\|e_Z(t)\|^2_F\} \leq \frac{1}{t^2} \|Q\|^2_F \|I - Q^t\|^2_F + \frac{n\Delta^2\|L\|^2_F}{4t^2} \sum_{k=1}^{t} \|I - Q^k\|^2_F.$$

It thus follows from Lemma 2 that

$$E\{\|e_Z(t)\|^2_F\} \leq \frac{n\nu^2}{4t} + \frac{(n+2)\|Q\|^2_F}{t^2} + \frac{n\Delta^2\|L\|^2_F}{4t} \sum_{k=1}^{t} \|I - Q^k\|^2_F.$$

(10)

We will make use of the following lemma to finish up our proof, whose proof is given in Appendix C.

**Lemma 3:** Suppose that $p(Q) < 1$, then for any $q \in \mathbb{Z}_{\geq 1}$, we have $\sum_{k=1}^{q} \rho^{2(q-k)}(Q) \leq c'_q Q_k$. $\forall t \in \mathbb{Z}_{\geq 1}$, where $c'_q \triangleq \left(\frac{\rho^2(Q)}{1 - \rho(Q)}\right)^{2(q-1)} + \sum_{j=0}^{2q-2} \frac{(2q-2)!\rho^j(Q)}{((2q-2)-j)!\rho(Q)^j} \geq 1$. Recall that $\rho(Q) < 1$ in terms of Lemma 2 then by Lemma 3 one knows that $\sum_{k=1}^{t} \rho^{2(q-k)}(Q)$ is bounded for all $t \in \mathbb{Z}_{\geq 1}$. Moreover, $\rho^{2q}(Q) \to 0$ as $t \to \infty$ for any integer $k \in \mathbb{Z}_{\geq 0}$. Thus it follows from (10) that for all large $t$

$$E\{\|e_Z(t)\|^2_F\} \leq \frac{n\nu^2}{4t} + o\left(\frac{1}{t}\right),$$

from which the theorem follows.

**Remark 2:** Different from the standard consensus algorithm [26], the averaging based method has a universal convergence rate of $O(t^{-1})$, independent of the network topology. The possible effect of the network topology only lies in the rate coefficient $\lim_{t \to \infty} tE\{\|e_Z(t)\|^2_F\}$. In fact, the upper bound (10) gives a rough estimate of the rate coefficient of convergence, namely, $n(n+2)n\Delta^2/4\|I + \omega^T\|_2^2$ for any network topology. However, this depends on the parameter $\alpha$, the network topology through $n$, $\omega$, and the quantization scheme through $\Delta$. We note that a similar rate is established for undirected graphs in [14].
It is easy to see that the first two terms of the RHS of (12) tend to zero as $t \to \infty$. In the following, we will focus on the almost sure convergence of the third term and aim to provide some quantitative bounds of the rate at which convergence takes place. To this end, we define

$$r_t^U \triangleq \max_i \lambda_{\max} \left( \sum_{k=0}^{t-1} \text{Cov}(u_i(k)) \right).$$

The next theorem establishes such a bound with close relation to $r_t^U$, which is motivated by the law of the iterated logarithm for independent random variables [23].

**Theorem 2:** Under Assumptions 1 and 2, for all large $t$,

i) if $\sup_{t \geq 1} r_t^U < \infty$, then there exists a constant $c_U > 0$ such that $\max_i \| \sum_{k=0}^t u_i(k) \| \leq c_U \ a.s.$ and

$$\| e_Z(t) \|_F \leq \frac{\mu}{t} \ a.s.,$$

where $\mu \triangleq \sqrt{n + 2}\| \tilde{Q} \|_2^2 + \alpha(n \eta \| Q' \| \Delta + c_U \| L \|_2^2)$, and

$$c'_Q \triangleq \left\{ \begin{array}{ll}
\frac{\rho(Q)}{\ln(\ln t)} & , q = 1, \\
\frac{1 - \rho(Q)}{\ln(\ln t)} q^{-1} + \sum_{q = 0}^{q-1} \frac{q-1}{\ln(\ln t)} q^{-1}, & , q > 1,
\end{array} \right.$$}

ii) if $\lim_{t \to \infty} r_t^U = \infty$, then

$$\| e_Z(t) \|_F \leq \frac{\alpha n \| L \|_2^2 \sqrt{2r_t^U \ln \ln r_t^U}}{t} \ a.s.$$  

**Proof:** Note that $W_t(k)\tilde{L}U(k) = \tilde{L}U(k) - Q^{t-1} \tilde{L}U(k)$, for all $0 \leq k \leq t - 1$. In light of (3), we have $\| U(k) \|_F = \sqrt{\sum_{i=1}^n \| u_i(k) \|^2} \leq n\Delta$. These two relations imply

$$\left\| \sum_{k=0}^{t-1} W_t(k) \tilde{L}U(k) \right\|_F \leq \| \tilde{L}U \|_2 \left( n\Delta \sum_{k=1}^t \| Q^k \|_F \right) + \left\| \sum_{k=0}^{t-1} U(k) \right\|_F.$$  

Using a similar argument of Lemma 3 we can show that $\sum_{k=1}^{t-1} \rho^k(Q)$ is bounded by $c'_Q$. It thus follows from Lemma 1 that

$$\sum_{k=1}^t \| Q^k \|_F \leq nc_Q \sum_{k=1}^t k^{q-1} \rho^k(Q) \leq nc_Qc'_Q.$$  

This shows that the first term of the RHS of (16) is bounded.

It remains to provide the quantitative bound of the rate of convergence of $\| \sum_{k=0}^t U(k) \|_F$. First, by the definition of $U(t)$, it can be verified that $\| \sum_{k=0}^t U(k) \|^2_F = \sum_{i=1}^n \| \sum_{k=0}^{t-1} u_i(k) \|^2, \forall t \in \mathbb{Z}_1$, from which one has

$$\left\| \sum_{k=0}^{t-1} U(k) \right\|_F \leq \sum_{i=1}^n \left\| \sum_{k=0}^{t-1} u_i(k) \right\|.$$  

Now considering each $i = 1, 2, \ldots, n$, we have the next two cases:

**Case 1.** $\sup_{t \geq 1} r_t^U < \infty$. It is known that $\text{Cov}(u_i(k))$ is positive semidefinite and $\text{E}[\| u_i(k) \|^2] = \text{trace}(\text{Cov}(u_i(k))), \forall k \in \mathbb{Z}_2$. Consequently, in this case, we have

$$\sum_{k=0}^{t-1} \text{E}[\| u_i(k) \|^2] = \text{trace} \left( \sum_{k=0}^{t-1} \text{Cov}(u_i(k)) \right) \leq n \sup_{t \geq 1} r_t^U.$$  

Recall that $\{u_i(k)\}_{k \geq 0}$ is a sequence of independent bounded random vectors. By employing the Kolmogorov three series theorem [27], we know that $\sum_{k=0}^{t-1} u_i(k)$ converges almost surely as $t \to \infty$. Thus there exists a constant $c_U > 0$ so that

$$\max_i \left\| \sum_{k=0}^{t-1} u_i(k) \right\| \leq c_U, \forall t \in \mathbb{Z}_{\geq 1} \ a.s.$$  

Substituting (17), (18) and (19) into (16) implies that $\| \sum_{k=0}^{t-1} W_t(k) \tilde{L}U(k) \|_F \leq n(\eta c_Q c'_Q \Delta + c_U \| L \|_2^2)$. Moreover, $t^\rho 2^{\rho}(Q) \to 0$ as $t \to \infty$ for all $k \in \mathbb{Z}_{\geq 0}$. It then follows from (12) that for large $t$,

$$\| e_Z(t) \|_F \leq \frac{\mu}{t} + o \left( \frac{1}{t} \right) \ a.s.$$  

**Case II.** $\lim_{t \to \infty} r_t^U = \infty$. In this case, there is 1 at most $n$ satisfying $r_t^U \triangleq \lambda_{\max} \left( \sum_{k=0}^{t-1} \text{Cov}(u_i(k)) \right) \to \infty$ as $t \to \infty$. Hence $\ln t = o(r_t^U)$ as $t \to \infty$. On the other hand, one obtains $\| u_i(k) \|^2 \leq n\Delta^2, \forall k \in \mathbb{Z}_{\geq 0}$. It thus follows from Theorem 1.1 of [28] that

$$\limsup_{t \to \infty} \left( \sum_{k=0}^{t-1} u_i(k) \right) \leq 1 \ a.s.$$  

Now, invoking (18), (19) and noting that

$$\| \sum_{k=0}^{t-1} U(k) \|_F \leq 2r_t^U \ln \ln r_t^U,$$  

for all large $t$. Substituting (16), (17) and (21) into (12), we finally get for large $t$,

$$\| e_Z(t) \|_F \leq \frac{\alpha n \| L \|_2^2 \sqrt{2r_t^U \ln \ln r_t^U}}{t} + o \left( \sqrt{r_t^U \ln \ln r_t^U} \right) \ a.s.$$  

Combining the above two cases completes the proof. $\blacksquare$

**Theorem 3:** Under Assumptions 1 and 2, for any constant $0 < \eta < 1$, there exists $t_\eta \in \mathbb{Z}_2$ such that for all $t \in \mathbb{Z}_{\geq t_\eta}$,

$$\min_i \frac{\xi_i(t)}{\omega_i} \geq \eta \ a.s.$$  

**Proof:** Note that

$$\sum_{i=1}^n \| \xi_i(t) - \omega_i \|^2 \leq \| e_Z(t) \|_F^2, \forall t \in \mathbb{Z}_{\geq 20}.$$  

And it is easy to see that

$$r_t^U \leq \max_i \lambda_{\max} \left( \text{Cov}(u_i(k)) \right) \leq \max_i \sum_{k=0}^{t-1} \text{E}[\| u_i(k) \|^2] \leq \frac{n\Delta^2}{4},$$  

where the second step follows from the relation that $\lambda_{\max} \left( \text{Cov}(u_i(k)) \right) \leq \text{trace}(\text{Cov}(u_i(k))) = \text{E}[\| u_i(k) \|^2]$, and the last inequality is a direct consequence of (3). Thus it follows from Theorem 2 that we always have $\| e_Z(t) \| \to 0 \ a.s.$, as $t \to \infty$. This together with (23) implies $\lim_{t \to \infty} \xi_i(t) = \omega_i \ a.s.$ Moreover, by Lemma 1 we know that $\min_i \omega_i > 0.$
As a result, for all large $t \geq t_0$, one has $\tilde{z}_i(t) \geq \eta \omega_i$ a.s., $\forall i$. The proof is thus complete.

Remark 3: This result shows that $\epsilon(t)$ in (6) is well-defined for large $t$. To implement (6), we may choose $\tau = t_0$ to trigger the estimation algorithm (1) at the second stage. Actually, with the setup in Algorithm 1 it is possible to choose a much smaller $\tau < t_0$ (see the simulation results in Section VII). For clarity of presentation of the subsequent theoretical results, we always assume that $\tilde{z}_i(t + \tau) \geq \eta \omega_i$, $\forall t \in \mathbb{Z}_{\geq 0}$.

B. Convergence results for the distributed estimation algorithm

Write $Q(x_i(t)) = x_i(t) + \psi_i(t)$, $\forall i$, where $\psi_i(t)$ is the quantization error satisfying (2) and (3). Stack $x_i(t), \epsilon_i(t)$ and $v_i(t)$ into column vectors $x(t), \epsilon(t)$ and $v(t)$, respectively, then we can rewrite (1) more compactly into

$$x(t + 1) = P(x(t) + \epsilon(t)) - \alpha L v(t).$$

Hence the running average $\tilde{x}(K)$ of (1) can be expressed as

$$\tilde{x}(K) = \frac{1}{K} \sum_{t=0}^{t_0+K-1} P^t x(0) + \frac{1}{K} \sum_{t=t_0}^{t_0+K-1} \sum_{s=0}^{t-t_0-1} (P^s \epsilon(s) - \alpha L v(s)).$$

The next lemma provides the convergence properties of the correction term $\epsilon(t)$ of (6).

Lemma 4: Under Assumptions 1 and 2, $\epsilon(t)$ converges to zero both in mean square and almost sure senses. Moreover, for each $0 < \eta < 1$, we have for all large $t$,

$$\mathbb{E}\{\|\epsilon(t)\|^2\} \leq \frac{\nu^2 x_0^2}{2\eta^4(t + \tau - 1)},$$

and almost surely

$$\|\epsilon(t)\| \leq \left\{ \begin{array}{ll} \frac{2\sqrt{2x_0^2}}{4\omega_1^2 + \ln \ln r_{t+\tau}^U}, & \text{sup } r_{t+\tau}^U < \infty, \\ \frac{4\omega_1^2}{\eta^4(t + \tau - 1)}, & \text{otherwise,} \end{array} \right.$$

where $x_0^2 \equiv \max_i \omega_i^{-2} |x_i(0)|$.

Proof: See Appendix [D]

The compensation nature of $\epsilon(t)$ is demonstrated in the next lemma, which guarantees convergence of the weighted sum to the desired $\hat{\theta}$. To this end, denote $e_x(t) \triangleq \omega^T x(t) - \hat{\theta}$.

Lemma 5: Under Assumptions 1 and 2, $\omega^T x(t)$ converges to the centralized estimate $\hat{\theta}$ both in mean square and almost sure senses. Further, $e_x(t)$ is approximately bounded by

$$\mathbb{E}\{e_x^2(t + 1)\} \leq \frac{\nu^2 x_0^2}{4\eta^4(t + \tau)},$$

and almost surely

$$|e_x(t + 1)| \leq \left\{ \begin{array}{ll} \frac{x_0^2}{\eta^4 \ln(t + \tau + 1)}, & \text{sup } r_{t+\tau}^U < \infty, \\ \frac{2x_0^2}{\eta^4 \ln \ln r_{t+\tau}^U}, & \text{otherwise,} \end{array} \right.$$ 

where $x_0^2 \equiv \max_{1 \leq i \leq n} \omega_i^{-1} |x_i(0)|$.

Proof: See Appendix [E]

Based on Lemma 5, we can decompose the estimation error $e_x(K) \triangleq \tilde{x}(K) - \hat{\theta}$ into two parts: $e_x(K) = e_x(t_0 + K)1 + e_{x,x}(K)$, where $e_{x,x}(K) \triangleq \tilde{x}(K) - \omega^T x(t_0 + K)1$.

In the following, we will provide an upper bound of $\|e_{x,x}(K)\|$. Recall that $\omega^T L = 0$, we have $\omega^T x(t_0 + K) = \omega^T x(0) + \sum_{s=0}^{t_0+K-1} \epsilon(s))$. Similar to (9), one can obtain from (25) that

$$e_{x,x}(K) = \frac{1}{K} \left( Q Q_{t_0+1}^0 (I - Q^K) x(0) - \alpha \sum_{t=0}^{t_0+K-2} \tilde{W}_K(t) L v(t) \right) + \frac{Q}{n} \sum_{t=0}^{t_0+K-1} \tilde{W}_K(t) \epsilon(t) + \frac{1}{n} \sum_{t=0}^{t_0+K-1} x^T(0) \epsilon_K(t),$$

where $\tilde{W}_K(t)$ is defined by

$$\tilde{W}_K(t) \triangleq \left\{ \begin{array}{ll} (I - Q^K) Q_{t_0+1}^0 (t - t_0 - 1) \omega_{K} - \hat{\theta}, & 0 \leq t \leq t_0 - 1, \\ I - Q^t - K - 1 - t, & t_0 - 1 \leq t \leq t_0 + K - 2, \end{array} \right.$$ 

and $\epsilon_K(t) = [\epsilon_1(K), \cdots, \epsilon_n(K)]^T$ with the $i$-th entry being

$$\frac{1}{\tilde{z}_{ii}} - \frac{\omega_i}{\tilde{z}_{ii}(t_0 + K + x - 1)}.$$

1) Mean square performance: Let $t'_0 \triangleq t_0 + \tau - 2$, then we have the next result regarding the mean square convergence of $e_{x,x}(K)$.

Theorem 4: Under Assumptions 1 and 2, we have

$$\mathbb{E}\{\|e_{x,x}(K)\|^2\} \leq \frac{2n^2 \ln^2 x_0^2}{2n^2 \eta^4 K^2} \ln(t'_0 + K).$$

Proof: It follows from (20) that $\mathbb{E}\{\|e_{x,x}(K)\|^2\} \leq 3K^{-2} \mathbb{E}\{\|I(Z_1)\|^2\} + \mathbb{E}\{\|I(Z_2)\|^2\} + \mathbb{E}\{\|I(Z_3)\|^2\}$. Considering $\mathbb{E}\{\|I(Z_1)\|^2\}$, under Assumption 2, one has

$$\mathbb{E}\{\|I_1\|^2\} \leq \frac{\|Q Q_{t_0+1}^0 (I - Q^K)\|_2 x(0)^2}{2}.$$

Let $\text{Cov}^{1/2}(v(t))$ be the square root of $\text{Cov}(v(t))$, then one can obtain

$$\mathbb{E}\left\{ \|W_K(t)Lv(t)\|^2 \right\} \leq \frac{\|W_K(t)\|^2}{4} \|W_K(t)\|^2,$$

where in the last step use was made of the fact that $\|\text{Cov}^{1/2}(v(t))\|_2 \leq \max_i \sqrt{\mathbb{E}\{\epsilon_i^2\}}$ and (2). Moreover, by the definition of $W_K(t)$, we have

$$\sum_{t=0}^{t_0+K-1} \|W_K(t)\|_F^2 \leq \|I - Q^K\|_F^2 \sum_{t=0}^{t_0+K-1} \|Q^t\|_F^2 + \sum_{t=0}^{K-1} \|I - Q^t\|_F^2$$

$$\leq (n + 2)K + O(1),$$
where in the last step we make use of the relation \( \sum_{t=0}^{t_0-1} \|Q_t\|^2 = n^2 + \sum_{t=0}^{t_0-1} \|Q_t\|^2 \), and Lemmas 1, 2 and 3. Substituting (29) and (30) back into (28) yields
\[
\mathbb{E}\{\|I_2\|^2\} \leq \frac{\nu^2 K}{4} + O(1).
\] (31)

As for \( \mathbb{E}\{\|E_2\|^2\} \), by Lemma 4, we know that \( \sup_{t \geq t_0} \mathbb{E}\{\|e(t)\|^2\} \) is bounded. It thus follows from the \( c_r \) inequality (23), Lemmas 2 and 3 that for large \( K \),
\[
\mathbb{E}\{\|E_2\|^2\} \leq c_{Q,n}^2 \mathbb{E}\sum_{t=t_0}^{t_0+K-1} \|e(t)\|^2 + O(1)
\leq \frac{c_{Q,n}^2 \nu^2 n^2}{n^4} \ln(t_0 + K - 1) + o(\ln K).
\] (32)

where the last step follows from Lemma 4 and the relation
\[
\sum_{t=t_0}^{t_2} \frac{1}{t} \leq \int_{t_1}^{t_2} \frac{1}{t} dt + \frac{1}{t_1} = \ln(t_2/t_1) + \frac{1}{t_1},
\] (33)

for any two integers \( t_1, t_2 \in \mathbb{Z}_{\geq 1} \) with \( t_2 > t_1 \).

Let us turn to \( \mathbb{E}\{\|E_3\|^2\} \). It follows from (27) and Theorem 3 that
\[
\mathbb{E}\{\|E_3\|^2\} \leq \frac{1}{n} \sum_{t=t_0}^{t_0+K-1} \mathbb{E}\{x^T \mathbb{E}(\|e_2(t)\|^2)\}
\leq \frac{2\nu^2 K}{n^4} \mathbb{E}\{\|e_2(t_0 + K + 1)\|^2\}
+ \sum_{t=t_0+1}^{t_0+K} \mathbb{E}\{\|e_2(t)\|^2\}
\leq \frac{\nu^2 n^2}{2n^4} K \ln(t_0 + K) + o(\ln \ln K),
\] (34)

where in the last step was made of Theorem 1 and (33).

Combining the bounds (31), (32) and (34) all together, after some simplification, we finally complete the proof.

Recalling the fact that \((t_0 + K - 1)^{-1} = o(K^{-1} \ln K)\) for large \( K \), we immediately have the next result of Lemma 4 and Theorem 5.

**Theorem 5:** Under Assumptions 1 and 2, at each node \( i \), the running average \( \bar{x}_i(K) \) converges to the centralized estimate \( \bar{\theta} \) almost surely. Moreover, for large \( t \), the approximate upper bounds of \( \|e_2(K)\| \) are summarized in Table 1.

**Proof:** We only need to consider \( \|e_{2,X}(K)\| \). First, it is trivial that \( \|e_{2,X}(K)\| \leq K^{-1} (\|I_1\| + \|I_2\| + \|I_3\|) \).

For \( \|I_1\| \), by Lemma 2, we know that \( \|I - Q^k\|_F \leq \|I - Q^k\|_F \leq \|I - \sum_{t=t_0+K}^{t_0+K-1} \mathbb{E}\{e(t)\} + O(1),
\] (36)

where the first step follows from (35) and the last step follows from (17).

To establish the rate of convergence of \( \sum_{t=t_0}^{t_0+K-1} \mathbb{E}\{e(t)\} \), we use the similar arguments as that of the proof of Theorem 1. We consider two cases, separately.

**Case I.** \( \sup_{t \geq t_0} r_{2,K} < \infty \). Note that \( v(t) \) is uniformly bounded in light of (34). Under Assumption 2, the Kolmogorov three series theorem for random vectors (27) applies. It thus follows that \( \lim_{K \to \infty} \sum_{t=t_0}^{t_0+K-1} \mathbb{E}\{e(t)\} \) converges almost surely as \( K \) tends to \( \infty \). In particular, there exists a constant \( \alpha > 0 \) so that for all integers \( K \in \mathbb{Z}_{\geq 1} \), we have
\[
\mathbb{E}\{\sum_{t=t_0}^{t_0+K-1} \mathbb{E}\{e(t)\}\} \leq \alpha v a.s.
\]

**Case II.** \( \lim_{K \to \infty} r_{2,K} = \infty \). In this case, similar to (20), one has
\[
\limsup_{K \to \infty} \frac{\|\sum_{t=t_0}^{t_0+K-1} \mathbb{E}\{e(t)\}\}}{\sqrt{2K \ln K}} \leq 1 \text{ a.s.}
\]

Based on the above discussions, we obtain from (36) that \( \|I_2\| \) is approximately bounded by
\[
\mathbb{E}\{\|I_2\|^2\} \leq \frac{\alpha c_{Q,n} \|I_2\|^2 + O(1)}{\sqrt{2K \ln K}} \sup_{t \geq t_0} r_{2,K} < \infty, \quad \text{otherwise.}
\] (37)

Now turning to \( \|I_3\| \), similar to (32), we can get
\[
\mathbb{E}\|I_3\|^2 \leq \mathbb{E}\|Q_{n}E^2\|_2 \sum_{t=t_0}^{t_0+K-1} \mathbb{E}\{e(t)\} + O(1) \text{ a.s.}
\] (38)

As for \( \|I_3\| \), we can show that \( \|1_{x^T(0)}e_2(t)\|_F \leq \sqrt{2m} \mathbb{E}\{e_2(t + t_0 - 1)\}_F + \mathbb{E}\{e_2(t_0 + K + 1)\}_F \), which together with Theorem 2 implies
\[
\|I_3\| \leq \frac{\sqrt{2m} \mathbb{E}\{e_2(t_0 + K + 1)\}_F + \sum_{t=t_0+1}^{t_0+K} \mathbb{E}\{e_2(t)\}_F}{\sqrt{n}}.
\] (39)
Let $\varpi = 2c_{\text{UnifQ}, r_0^U} \|Q\|_2 + \sqrt{t_0''}$, then applying Theorem 2 and Lemma 4 to (38) and (39) yields
\[
\|I_2 + I_3\| \leq \frac{2\varpi}{n^2} \ln(t_0' + K) + o(\ln K), \text{ if } \sup_{t_1} r_{t_1}^U < \infty.
\]
For the case $\lim_{t_1 \to \infty} r_{t_1}^U = \infty$, noting that $r_{t_1}^U \ln r_{t_1}^U$ is monotone increasing of $t_1$, one has
\[
\|I_2 + I_3\| \leq \frac{2\varpi \|\tilde{L}\|_2}{n^2} \sqrt{r_{t_1}' + K} \ln r_{t_1}' + K \ln(t_0' + K) + o\left(\sqrt{r_{t_1}' \ln r_{t_1}' \ln K}\right).
\]

Based on the above analysis, we have the next four cases about the approximate upper bounds of $\|e_{k, x}(K)\|$

**Case I.** $\sup_{t_1, K \geq 1} \max\{r_{t_1}'', r_{t_1}''\} < \infty$.

\[
\|e_{k, x}(K)\| \leq \frac{2\varpi \|\tilde{L}\|_2}{n^2} \frac{r_{t_1}' + K}{K} \ln r_{t_1}' + K \ln(t_0' + K).
\]

**Case II.** $\sup_{K \geq 1} r_{t_1}' < \infty$ and $\lim_{t_1 \to \infty} r_{t_1}^U = \infty$.

\[
\|e_{k, x}(K)\| \leq \frac{2\varpi \|\tilde{L}\|_2}{n^2} \frac{r_{t_1}' + K}{K} \ln r_{t_1}' + K \ln(t_0' + K).
\]

**Case III.** $\lim_{K \to \infty} r_{t_1}' = \infty$ and $\sup_{t_1 \geq 1} r_{t_1}^U < \infty$.

\[
\|e_{k, x}(K)\| \leq \frac{2\varpi \|\tilde{L}\|_2}{n^2} \max\left\{\sqrt{r_{t_1}'' \ln r_{t_1}'}, \ln(t_0' + K)\right\}.
\]

**Case IV.** $\lim_{K \to \infty} r_{t_1}' = \infty$ and $\lim_{t_1 \to \infty} r_{t_1}^U = \infty$.

\[
\|e_{k, x}(K)\| \leq \frac{2\varpi}{n^2} \max\left\{\sqrt{r_{t_1}'' \ln r_{t_1}'}, \sqrt{r_{t_1}'' \ln r_{t_1}'}, \ln(t_0' + K)\right\}.
\]

Note that $t_0 - 1 \leq t_0'$ and $\sqrt{r_{t_1}'' \ln r_{t_1}'}, \sqrt{r_{t_1}'' \ln r_{t_1}'}$ are increasing functions of $t_1$, respectively. The above four cases together with Lemma 5 completes the proof.

**Remark 4:** From Theorems 5 and 6 we can see that the starting point $t_0$ contributes little to the rate of convergence of the proposed algorithm, since $\ln(t_0' + K) \approx \ln K$, for large $K$. This means that we can start the running average $x(K)$ at any time during the iteration. For example, at each stage, we can initiate the standard algorithm with the states as the estimates to achieve a better convergence rate, then start the running average to get higher accuracy. This is exactly what we have done in Algorithms 1 and 2 by introducing the starting points $k_0, t_0$ for the averaging processes.

**V. Simulation results**

In this section, we provide some simulation results to validate the theoretical analysis given in the previous sections.

Consider a sensor network with 12 nodes deployed to monitor an unknown parameter $\theta = 2$. The communication topology which is modeled as a directed graph, is shown in Fig. 1. Each node makes the measurement with $y_i = \theta + n_i$, where $n_i$ is the white Gaussian noise with zero mean and unit variance. As an illustration, we choose the Metropolis-type weight $a_{ij} = (1 + d_i)^{-1}$, if $j \in N_i$ and 0, otherwise. In this case, $\alpha = 1$ is sufficient for both Lemmas 1 and 2. For each implementation of the proposed algorithm, the initial state $x_i(0)$ is randomly chosen from the interval $[y_i - 1, y_i + 1]$, $\forall i$.

In the following simulations, both the deterministic uniform quantization (UnifQ) and probabilistic quantization (ProbQ)
Fig. 3. Comparison of the mean square errors of UnifQ, ProbQ and ProbQ-RA: $\text{MSE}_\omega$ ((a) and (b)), and $\text{MSE}_\theta$ ((c) and (d)) with respect to $\Delta \in \{0.2, 1\}$.

Fig. 4. Comparison of the mean square errors of ProbQ-RA-PQ, ProbQ-RA-TQ and ProbQ-RA: $\text{MSE}_\omega$ ((a) and (b)), and $\text{MSE}_\theta$ ((c) and (d)) with respect to $\Delta \in \{0.2, 1\}$. 
are considered and compared. The proposed averaging based algorithm is denoted as ProbQ-RA. Simulation results are presented by averaging over 100 independent runs.

A. Comparison of the deterministic and probabilistic quantization

First, we simulate the eigenvector estimation algorithm: Algorithm 1. Here, \( \kappa = 1.15 \) and the starting point for the running average of \( z \) is taken as \( k_0 = 25 \). Fig. 2 depicts the estimate of the left eigenvector \( \omega \) with respect to the zero eigenvalue of \( L \) at the first node for \( \Delta = 1 \), respectively. In Fig. 2 we just use \( z_1 \) of ProbQ as the estimates of \( \omega \) in the first 25 steps. From the results, we observe that steady residues occur for UnifQ, and there are fluctuations for ProbQ. While for the proposed ProbQ-RA, the running average has an obvious smoothing effect, where the randomness of ProbQ is smeared out. The performance of ProbQ-RA is satisfactory even for a low quantization resolution \( \Delta = 1 \) compared with the large residues observed in both UnifQ and ProbQ.

To quantify the performances, we use the average of the mean square error as an indicator, which is defined by

\[
\text{MSE}_Z = \frac{1}{n} \sum_{i=1}^{n} \| z_i(t) - \omega \|^2, \quad \text{MSE}_X = \frac{1}{n} \sum_{i=1}^{n} \| x_i(t) - \theta \|^2,
\]

for the estimate of the left eigenvector \( \omega \), and

\[
\text{MSE}_Z = \frac{1}{n} \sum_{i=1}^{n} (z_i(K) - \hat{\theta})^2, \quad \text{MSE}_X = \frac{1}{n} \sum_{i=1}^{n} (x_i(t) - \theta)^2
\]

for the sample mean. The starting points used in Algorithm 2 are set \( \tau = 1 \) and \( t_0 = 25 \), respectively. The results are shown in Fig. 3. It can be seen that the proposed ProbQ-RA outperforms UnifQ and ProbQ in both cases with the quantization resolutions \( \Delta = 0.2 \) and 1. The performances of UnifQ and ProbQ are acceptable for the estimates of the left eigenvector \( \omega \) in both cases (see Fig. 3(a) and (b)). However, with the errors accumulated from the first stage to the second stage, they degrade significantly for lower quantization resolution, e.g., \( \Delta = 1 \) (see Fig. 3(c) and (d)). Compared with UnifQ and ProbQ, the proposed ProbQ-RA degrades quite smoothly. There is only a modest increase of MSE with decreasing quantization resolution, i.e., increasing \( \Delta \) from 0.2 to 1. These results indicate that the averaging technique can improve the accuracy of the estimates especially for the case of low quantization resolutions, where its smoothing effect contributes much to the improvement.

B. Comparison with the partially quantized and totally quantized updating rules

In Fig. 4 we plot the results of the average mean square errors \( \text{MSE}_Z \) and \( \text{MSE}_X \) for the three updating rules using running average, where ProbQ-RA-PQ and ProbQ-RA-TQ denote the averaging based PQ and TQ rules. From the results, we can see that the averaging based PQ and TQ rules perform well for the left eigenvector estimation for both \( \Delta = 0.2 \) and 1. However, it is observed from Fig. 4(c) that the errors are quite large at the second stage even with a rather high quantization resolution \( \Delta = 0.2 \). Moreover, with the quantization resolution decreased from \( \Delta = 0.2 \) to \( \Delta = 1 \), both PQ and TQ rules will not produce acceptable results (see Fig. 4(d)). PQ rule diverges and TQ rule doesn’t provide any meaningful data for large \( \Delta \). Different from the PQ and TQ rules, the update rule used in (1) and (5) performs quite well for all the cases and the running average can further improve its accuracy. This is consistent with the aforementioned theoretical analysis.

Finally, we compare UnifQ, ProbQ, ProbQ-RA-PQ, ProbQ-RA-TQ and ProbQ-RA regarding the average mean square error for different quantization resolutions. The results are shown in Fig. 5 (as for ProbQ-RA-TQ, we only plot the results for \( \Delta \in \{0.05, 0.1, 0.2, 0.4, 0.8, 1\} \)). In order to avoid the transient periods, we take the average of the last 150 iterations of \( \text{MSE}_X \) in presenting the results. From the figure, we can see that the proposed ProbQ-RA works quite well even when \( \Delta = 1 \). There are significant improvements of the performance at lower quantization resolutions by using ProbQ-RA compared with other algorithms. The running average technique does improve the performance of PQ and TQ rules for smaller \( \Delta \). While for larger \( \Delta \), it seems that the running average does not have such effect on PQ and TQ rules. Although the running average has smoothing effects on random data, the above simulations indicate that only certain kinds of algorithms benefit from this consequence.

VI. Conclusion

We have studied the distributed parameter estimation problem over sensor networks in the presence of quantized data and directed communication links. We have proposed a two-stage algorithm such that the centralized sample mean estimate can be achieved in a distributed manner. In the algorithm, the running average technique is utilized to smear out the randomness caused by the probabilistic quantization scheme. We have shown that the proposed algorithm can achieve the centralized sample mean estimate both in the mean square and almost sure senses. Finally, we have presented simulation results illustrating the effectiveness of the proposed algorithm. Comparisons with other algorithms have also been provided to highlight the improvements of the proposed algorithms.
APPENDIX A
PROOF OF LEMMA 1

i) Since $G$ is strongly connected, $\lambda_1(L) = 0$ is a
principal eigenvalue and $\lambda_i(L) \neq 0$, for all
$i = 1, 2, \ldots, n$. In view of the principle of
biorthogonality [29], all the eigenvectors $\nu_i$
corresponding to $\lambda_i(L)$, $i = 1, 2, \ldots, n$, are
orthogonal to $\omega$, that is, $\omega^T \nu_i = 0$. This
implies that $Q \nu_i = (I - \alpha L) \nu_i = (1 - \alpha \lambda_i(L)) \nu_i$. Moreover, it can be verified that $Q \nu_i = 0$,
since $L \nu_i = 0$ and $1^T \omega = 1$.

ii) By i), one has $\rho(Q) = \max_{2 \leq i \leq n} |1 - \alpha \lambda_i(L)| = \max_{2 \leq i \leq n} \sqrt{\alpha^2 |\lambda_i(L)|^2 - 2 \alpha \text{Re}(\lambda_i(L)) + 1}$. Since $G$
is strongly connected, we know that $\text{Re}(\lambda_i(L)) > 0$, for
all $i = 2, 3, \ldots, n$. Hence, the equation
$\alpha^2 |\lambda_i(L)|^2 - 2 \alpha \text{Re}(\lambda_i(L)) + 1 = 1$ has two solutions
$0 = \alpha_1 < \alpha_2 = \frac{\text{Re}(\lambda_i(L))}{|\lambda_i(L)|}$
corresponding to each $\lambda_i(L)$, $i = 2, 3, \ldots, n$. This means that $\rho(Q) < 1$ if and only if
$0 < \alpha < \min_{2 \leq i \leq n} \frac{2 \text{Re}(\lambda_i(L))}{|\lambda_i(L)|}$.

iii) It follows from Theorem 1 of [31] that there is a constant
cdepending only on $Q$ so that
$$\sqrt{\sum_{i=1}^{n} |\lambda_i(Q)|^{2k}} \leq \|Q^k\|_F \leq c(k)k^{q-1} \sum_{i=1}^{n} |\lambda_i(Q)|^{k}.$$ 
It is immediate that $\rho^k(Q) \leq \|Q^k\|_F \leq n c(k)k^{q-1} \rho^k(Q)$.

APPENDIX B
PROOF OF LEMMA 2

First, it is easy to see that $P$ is nonnegative. Recall that
$\lambda_i(L) \neq 0$, for $i = 2, 3, \ldots, n$ for strongly
graphed connected graphs, we only need to show that
$$\max_{1 \leq i \leq n} \left\{ \frac{2 \text{Re}(\lambda_i(L))}{|\lambda_i(L)|^2} \right\} \leq \min_{2 \leq i \leq n} \left\{ \frac{2 \text{Re}(\lambda_i(L))}{|\lambda_i(L)|^2} \right\}$$
in view of Lemma 1. Indeed, by the Geršgorin disc theorem
[29], all the eigenvalues of $L$ are located in the union of
disks $\bigcup_{i=1}^{n} \{ |\lambda - d_i| \leq \sum_{j=1}^{n} |\lambda_{ij}| \}$. Consequently, for each $\lambda_i(L)$,
$i = 2, 3, \ldots, n$, we can find a $\lambda_i$ located on some circle such
that $\text{Re}(\lambda_i) = \text{Re}(\lambda_i(L))$ and $|\lambda_i(L)| \leq |\lambda_i|$ (see Fig. 6). This means that
$$\frac{2 \text{Re}(\lambda_i(L))}{|\lambda_i(L)|^2} \geq \frac{2 |\lambda_i| \cos(\beta)}{|\lambda_i|^2} = 2 \cos(\beta) \frac{1}{|\lambda_i|}. \quad (40)$$
On the other hand, we have $\cos(\beta) = |\lambda_i|/(2d_i)$, which along
with [40] gives $\min_{2 \leq i \leq n} \left\{ \frac{2 \text{Re}(\lambda_i(L))}{|\lambda_i(L)|^2} \right\} \geq \min_{2 \leq i \leq n} \left\{ \frac{1}{d_i} \right\}$.

Since $P$ is nonnegative and $1^T \omega = 1$, it is clear that
trace($Q^k$) = trace($P^k$) − 1, $\forall k \in \mathbb{Z}_{\geq 1}$. It thus follows
from Lemma 1 that $\|I - Q^k\|_F^2 = n - 2 \text{tr}(Q^k) + ||Q^k||_F^2 \leq n + 2 + 2 \rho^2(Q) \rho^{2k}(Q)$.

Consider the case of $q > 1$, let $T_0 \triangleq (1 - q)/\ln \rho(Q)$, it can be shown that $\rho^{2(q-1)} \rho^{2k}(Q)$ is monotone increasing
on the interval $(0, T_0]$ and decreasing on $[T_0, \infty)$. Therefore,
$\rho^{2(q-1)} \rho^{2k}(Q) \leq T_0^{2(q-1)} \rho^{2T_0}(Q) = T_0^{2(q-1)}$ since
$(\rho(Q))^{1/\ln \rho(Q)} = e$. This completes the proof.

APPENDIX C
PROOF OF LEMMA 3

The case of $q = 1$ is straightforward. We only need to
take the case of $q > 1$. From the proof of Lemma 2 we
know that $f(t) = t^{2(q-1)} \rho^{2t}(Q)$ is monotone increasing
on the interval $(0, T_0]$ and decreasing on $[T_0, \infty)$, where $T_0 \triangleq (1 - q)/\ln \rho(Q)$. We have the next three cases:

Case I. $[T_0] = 1$. In this case, we may have $t \leq 1$ and $f(t) \leq f(1) + \int_{1}^{t} f(s)ds$.

Case II. $1 < [T_0] < t$. In this case, we may have either $[T_0] = T_0$ or $[T_0] - T_0 = 1$. For the
former case, it is easy to get $t \leq f(T_0) + \int_{1}^{t} f(s)ds$.

The latter case can be proved similarly as $\sum_{k=1}^{t} f(k) \leq f(T_0) + \int_{1}^{t} f(s)ds$. Since
$[T_0] - T_0 = 1$, it follows that $f(T_0) \geq \min_{[T_0]} f(T_0)$.
Moreover, $f(T_0) = f(T_0) + \min_{[T_0]} f(T_0)$.

Combining the above three cases, we finally get
$$\sum_{k=1}^{t} f(k) \leq f(T_0) + \int_{1}^{t} f(s)ds. \quad (41)$$

On the other hand, by repeatedly using the integration by parts,
one can deduce that
$$\int_{1}^{t} f(s)ds = \sum_{k=0}^{t-2} (-1)^{k} k! \frac{2^{(q-1)k}}{(2(q-1))^{k}} \rho^{2t}(Q) - \rho^2(Q)\quad \quad (42)$$
where we adhere to the convention that $0! = 1$. Substituting
(42) into (41) completes the proof.

APPENDIX D
PROOF OF LEMMA 4

By the definition of $\epsilon_i(t)$ in (6), for each $i = 1, 2, \ldots, n$, we have
$$|\epsilon_i(t)| = \frac{|x_i(0)| - |x_i(t) - \bar{\epsilon}_i(t)|}{n} \leq \frac{|x_i(t) - \bar{\epsilon}_i(t)|}{n} \quad \forall t \in \mathbb{Z}_{\geq 1}. \quad \quad (43)$$
It thus follows from Theorem 4 that for all $t \in \mathbb{Z}_{\geq 1}$,
$$|\mathbb{E}[\epsilon_i(t)]| \leq \frac{|x_i(0)|}{n^{1/2}} \mathbb{E} \{ |x_i(t) - \bar{\epsilon}_i(t)| \} \quad \quad (44)$$
Moreover, applying the Lebesgue bounded convergence theorem
[23], we conclude that $\mathbb{E}[|x_i(t) - \bar{\epsilon}_i(t)|] \to 0$ as $t \to \infty$. This implies that $\mathbb{E} \{ |x_i(t) - \bar{\epsilon}_i(t)| \} \leq$
We can thus employ the Hölder inequality to obtain
\[ E \{ \| e(t) \|^2 \} \leq \frac{\nu^2}{n^2 \nu^2} \sum_{i=1}^{n} \frac{x_i^2(t)}{\nu^2} \| x_i(t) \|_F^2 + \| x_i(t) \|_F^2, \]

where the last inequality follows from \((24)\). Therefore, applying Theorem 2 to the previous relation completes the proof.

**APPENDIX E**

**PROOF OF LEMMA 5**

Note that \( \omega^T L = 0 \). By iterating \((24)\), one can get
\[ \omega^T x(t + 1) = \omega^T \left( x(0) + \sum_{s=0}^{t} \epsilon(s) \right) = \frac{1}{n} \left[ \frac{\omega_1}{z_{i1}(t + \tau)}, \ldots, \frac{\omega_n}{z_{in}(t + \tau)} \right] x(0). \]
We can thus employ the Hölder inequality to obtain
\[ c_x^2(t + 1) \leq \frac{1}{n^2} \sum_{i=1}^{n} \frac{x_i^2(t)}{z_{i}(t + \tau)^2} \sum_{i=1}^{n} \left( \omega_i - z_{ii}(t + \tau) \right)^2, \]
which together with \((22)\) and \((24)\) implies that
\[ \frac{c_x^2(t + 1)}{\nu^2} \leq \| e_x(t + 1) \|_F^2 \text{ and } E \{ c_x^2(t + 1) \} \leq \frac{\nu^2}{n^2} \| e_x(t + 1) \|_F^2. \]

The lemma thus follows from Theorems 1 and 2.

**REFERENCES**


