Distributed Consensus With Quantized Data via Sequence Averaging

Jun Fang and Hongbin Li

Abstract—**The problem of distributed average consensus with quantized data is considered in this correspondence. Conventional consensus algorithms suffer from divergence when quantization errors are present. To address this issue, we introduce a modified quantization-based consensus protocol and exploit the temporal information collected from the iterative process, based on which we develop an efficient consensus algorithm. The proposed consensus algorithm is proved to converge to the true mean, i.e., the average of the initial state, in a mean square sense. It also presents an advantage of speeding up the convergence over the algorithm [P. Frasca, R. Carli, F. Fagnani, and S. Zampieri, "Average Consensus on Networks With Quantized Communication," *Int. J. Robust Non-Linear Control*, 2008, to be published] without exploitation of temporal information. Numerical results are presented to illustrate the effectiveness of the proposed algorithm.**

Index Terms—**Distributed average consensus, quantization, wireless sensor network (WSN).**

I. INTRODUCTION

Distributed average consensus has attracted much attention over the past few years. It is a fundamental problem arising from various wireless sensor network (WSN) processing tasks, such as distributed parameter estimation and distributed function computation. A multitude of studies on distributed average consensus have appeared recently. Among them, a major research direction focuses on accelerating the convergence rate of the distributed consensus algorithms by selecting the optimal weights [2], [3] or resorting to other more sophisticated algorithms [4]–[7]. In these works, they usually assume that the data are exchanged among neighboring sensors without distortion. This assumption, however, may not be true in practice due to the link noise and data quantization, in which case it has been shown [8] that the conventional consensus algorithms [2]–[7] diverge and may have an unbounded asymptotic mean square error. To address this issue, many schemes [9]–[14] have been proposed. Specifically, when only quantization is considered, [11] introduced a quantized consensus algorithm by imposing an integer constraint on the value of each node and preserving the network average at each iteration. However, the quantized consensus [11] is a relaxed consensus and the nodes do not have the same value. Strict consensus can be achieved by other algorithms by using predictive coding [12], employing a sequence of link weights satisfying a persistence condition [13], or resorting to dithered quantization [14]. These algorithms [12]–[14] converge to a random variable that usually deviates from the average of the initial state. Nevertheless, this random variable reveals some information of the initial state. For example, in [14], the consensus is achieved on one of the quantization levels and the deviation from the average of the initial state is

bounded by a quantity dependent on the quantization resolution and graph topology. Meanwhile, in [13], the mean square deviation between the random variable and the average of the initial state can be made arbitrarily small by rescaling its link weight sequence at a cost of slowing down the convergence rate.

In this correspondence, we study the problem of distributed average consensus with quantized data by exploiting the temporal information collected during the iterative process. Temporal information has been successfully utilized in [5]–[7] to accelerate the convergence rate. Nevertheless, its application to distributed average consensus with quantized data has received less attention. In this work, we firstly introduce a modified quantization-based consensus protocol. The quantized data are obtained by employing a bounded probabilistic quantization scheme [13], [14]. By exploiting the temporal independence property of the quantization errors, we develop a new consensus scheme which utilizes the temporal information collected from the iterative process. The consensus algorithm, unlike prior works [12], [14] which converge to a random variable, achieves the desired consensus at the average of the initial state. Theoretical analyses are conducted to prove the convergence and provide an upper bound for the mean square deviation.

II. CONSENSUS PROTOCOLS WITH QUANTIZED DATA

We model the WSN as an undirected graph $G = (V, E)$ whose vertices $V = \{1, 2, \ldots, N\}$ correspond to the sensors and whose edges $E = \{(i, j) | i, j \in V\}$ represent available communication links among sensors. An edge between i and j exists if sensor i can communicate directly with sensor j. We focus our study on the connected graph, i.e., there exists a multihop communication path connecting every pair of vertices. The structure of the graph can be described by an $N \times N$ symmetric affinity matrix $\mathbf{A}$

$$a_{i,j} = \begin{cases} 1, & \text{if } (i, j) \in E \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $a_{i,j}$ denotes the (i, j)th entry of $\mathbf{A}$. The Laplacian matrix of the graph G is defined as

$$\mathbf{L} \triangleq \mathbf{D} - \mathbf{A} \tag{2}$$

where $\mathbf{D} \triangleq \text{diag}(\mathbf{A1})$ is the degree matrix and $\mathbf{1}$ denotes a column vector with all unity elements. $\mathbf{L}$ is a positive semidefinite matrix with only one null eigenvalue associated with the eigenvector $(1/\sqrt{N})\mathbf{1}$ [15]. Assuming ideal links and no quantization, the conventional distributed average consensus algorithm [2] updates its state as

$$\mathbf{x}(t + 1) = (\mathbf{I} - \alpha \mathbf{L})\mathbf{x}(t) \triangleq \mathbf{W}\mathbf{x}(t) \tag{3}$$

where $\mathbf{x}(t) \triangleq [x_1(t) \; x_2(t) \; \ldots \; x_N(t)]^T$, $x_n(t)$ denotes the values of sensor n at iteration t. It can be easily verified that for $0 < \alpha < 2/\lambda_{\max}(\mathbf{L})$, the above system converges to the average of the initial state, i.e., $(1/N)\mathbf{1}\mathbf{1}^T\mathbf{x}(0)$, where $\lambda_{\max}(\mathbf{L})$ denotes the largest eigenvalue of $\mathbf{L}$.

Now consider practical scenarios where data are quantized before transmission. In this case, different variations of the conventional protocol (3) have been introduced and they differ in how to utilize the received quantized data and local unquantized data. Specifically, in [12] and [13], (3) is decomposed as

$$\mathbf{x}(t + 1) = (\mathbf{I} - \alpha \mathbf{D})\mathbf{x}(t) + \alpha \mathbf{A}\mathbf{x}(t) \tag{4}$$

where the first term on the right-hand side requires only local values and the second term uses data from neighboring sensors. Naturally the update equation in the presence of quantization becomes

$$\mathbf{x}(t+1) = (\mathbf{I} - \alpha\mathbf{D})\mathbf{x}(t) + \alpha\mathbf{A}Q(\mathbf{x}(t)) \qquad (5)$$

where $Q(\mathbf{x})$ denotes the quantized vector by carrying out quantization for each entry of $\mathbf{x}$ using a certain quantization scheme. Another variation of (3) uses only quantized data for the update:

$$\mathbf{x}(t+1) = \mathbf{W}Q(\mathbf{x}(t)). \qquad (6)$$

This consensus protocol (6) was investigated in [14]. It was found that both (5) and (6) can achieve a consensus in a mean square sense by utilizing predictive coding schemes [12] or by using dithered quantization schemes [14]. However, these algorithms converge to a random value which is usually not equal to the average of the initial state. An important reason for this deviation is that for both protocols (5) and (6), the quantization errors incurred at each iteration are preserved throughout the process [16].

To address this issue, we introduce the following consensus protocol (a similar protocol appeared in [1], which was brought to our attention by one reviewer):

$$\mathbf{x}(t+1) = \mathbf{x}(t) - \alpha\mathbf{L}Q(\mathbf{x}(t)) = \mathbf{W}\mathbf{x}(t) - \alpha\mathbf{L}\mathbf{v}(t) \qquad (7)$$

where $\mathbf{v}(t) \triangleq Q(\mathbf{x}(t)) - \mathbf{x}(t)$ denotes the quantization noise vector. Due to the fact $\mathbf{1}^T\mathbf{L} = \mathbf{0}$, this protocol preserves the network summation throughout the iterations. Also, it can suppress the noise propagation to a certain extent. To see this, we expand $\mathbf{x}(t)$ as

$$\mathbf{x}(t) = \mathbf{W}^t\mathbf{x}(0) - \alpha\sum_{i=0}^{t-1}\mathbf{W}^{t-1-i}\mathbf{L}\mathbf{v}(i). \qquad (8)$$

For a specific i, the noise components $\mathbf{v}(i)$ will eventually vanish as the system evolves over time because we have

$$\lim_{t\to\infty}\mathbf{W}^{t-1-i}\mathbf{L}\mathbf{v}(i) = \frac{1}{N}\mathbf{1}\mathbf{1}^T\mathbf{L}\mathbf{v}(i) = \mathbf{0}. \qquad (9)$$

As shown in [1], due to the noise propagation suppression, this protocol is able to drive the system near the desired consensus value, i.e., the average of the initial state. Nevertheless, we note that the noise propagation is not completely removed and there is still a nonvanishing noise term contributed by the recent noise components. The final state $\mathbf{x}(t)$, therefore, diverges (the divergence of the final state from the desired consensus was thoroughly examined in [1] by a worst case analysis and by a probabilistic analysis). In the next section, by utilizing the temporal information collected from the iterative process, we develop an efficient algorithm that is guaranteed to converge to the average of the initial state in a mean square sense.

III. PROPOSED CONSENSUS ALGORITHM

A. Main Result

Our algorithm is based on the following result, which constitutes the main contribution of this work.

Theorem 1: Considering the consensus protocol (7), suppose that the noise sequence $\{\mathbf{v}(t)\}$ is independent with zero mean and finite auto-covariance matrix $E[\mathbf{v}(t)\mathbf{v}(t)^T] = \mathbf{C}_{v,t}$. Then a *temporal* average of the sequence $\{\mathbf{x}(t)\}$ converges to the initial state's average

value in the mean square sense as the sample size M (for averaging) tends to infinity, i.e.,

$$\frac{1}{M}\sum_{t=t_0}^{t_0+M-1}\mathbf{x}(t) \to \frac{1}{N}\mathbf{1}\mathbf{1}^T\mathbf{x}(0) \quad \text{as } M \to \infty \qquad (10)$$

where t_0 is the starting point for sequence averaging and can be any positive integer. Specifically, when the sample size M is large, the mean square deviation is approximately upper bounded by

$$E[\|\mathbf{x}_{\mathrm{ave}}(M) - \bar{\mathbf{x}}_0\|_2^2] < \frac{\alpha^2 N \lambda_v \lambda_{\max}^2(\mathbf{L})}{M(1-\rho)^2} \qquad (11)$$

where

$$\mathbf{x}_{\mathrm{ave}}(M) \triangleq \frac{1}{M}\sum_{t=t_0}^{t_0+M-1}\mathbf{x}(t)$$

$$\bar{\mathbf{x}}_0 \triangleq \frac{1}{N}\mathbf{1}\mathbf{1}^T\mathbf{x}(0),$$

ρ denotes the spectral radius of $(\mathbf{W} - (1/N)\mathbf{1}\mathbf{1}^T)$, $\lambda_{\max}(\mathbf{L})$ denotes the largest eigenvalue of $\mathbf{L}$, and $\lambda_v \triangleq \max_i \lambda_{\max}(\mathbf{C}_{v,i})$ for $i \in \{t_0, \ldots, t_0 + M - 1\}$.

Proof: See Appendix A. ∎

The intuition behind this theorem is to recognize that the temporal average of the sequence, despite the fact that the successive states of the sequence are correlated, can be reformulated as an average of a set of independent random vectors with expected value $(1/N)\mathbf{1}\mathbf{1}^T\mathbf{x}(0)$ and finite covariance matrices. We emphasize that the sequence averaging algorithm is only applicable to the consensus protocol (7) and its application to other consensus protocols (5) and (6) does not lead to a convergence to the average of the initial state.

B. Discussions

Some remarks regarding Theorem 1 are as follows.

Remark 1: Note that convergence in the standard literature is measured by the mean square deviation between the final state of the sequence and the average of the initial state, i.e., $E[\|\mathbf{x}(t) - \bar{\mathbf{x}}_0\|_2^2]$. It has been shown by our analysis that such a convergence cannot be achieved by the consensus protocol (7). Here we construct a new estimate and define the convergence as the mean square deviation between the new estimate and the average of the initial state. This is meaningful in distributed estimation scenarios where we do not have to choose the final state of the sequence as the estimate of the unknown parameter. Any estimator that can yield a more accurate estimate is certainly desirable.

Remark 2: The upper bound of the mean square deviation given in (11) is a function of the parameters $\alpha\lambda_{\max}(\mathbf{L})$, ρ, and λ_v, in which $\alpha\lambda_{\max}(\mathbf{L})$ is within the region $(0,2)$ to assure the system convergence (see the discussion after (3)), ρ is a parameter dependent on the graph topology and controlling the convergence rate in conventional consensus algorithms (the smaller the value, the faster the convergence rate), and λ_v is a parameter closely related to the quantization resolution. We also see that the mean square deviation is inversely proportional to the sample size M. To obtain a more accurate estimate, we need to increase the sample size M, and consequently, the number of iterations. This, to some extent, resembles the consensus algorithm [13] which provides a trade-off between convergence rate and mean square deviation. From (11), we also notice that for a large M, the starting point t_0 for sequence averaging makes little difference to the mean square deviation. This is because the term dependent on t_0 is of order $\mathcal{O}(1/M^2)$ (see (16)). Therefore, when M is sufficiently large, this term can be neglected as compared with the term of order $\mathcal{O}(1/M)$.

Remark 3: The condition imposed on the noise sequence $\{\mathbf{v}(t)\}$ in Theorem 1 is not restrictive and can be satisfied by adopting a probabilistic quantization scheme [13], [14]. The quantization scheme is summarized as follows. We uniformly divide the sensor's dynamic range $[-\eta, \eta]$ into intervals of length $\Delta = 2\eta/(2^b - 1)$ and round the message x to the neighboring endpoints of these intervals in a probabilistic manner, where b denotes the number of bits for quantization. Suppose $-\eta + i\Delta \le x \le -\eta + (i+1)\Delta$, then x is quantized to $Q(x)$ according to

$$P(Q(x) = -\eta + i\Delta) = 1 - r$$
$$P(Q(x) = -\eta + (i+1)\Delta) = r \tag{12}$$

where $r = (x + \eta - i\Delta)/\Delta$. It can be easily verified that the quantized message $Q(x)$ is an unbiased estimator of x with variance $E[(Q(x) - x)^2] \le \Delta^2/4$. Hence, the quantization error induced in this way is a random variable with zero mean and a finite variance. Also, the quantization errors from different iterations are independent. We would like to emphasize that the strong correlation between successive states does not affect the independence between $\mathbf{v}(t)$ and $\mathbf{v}(t+1)$. Considering the extreme case where $\mathbf{x}(t) = \mathbf{x}(t+1)$, by adopting the probabilistic quantization scheme, both $v_i(t)$ and $v_i(t+1)$ ($v_i(t)$ denotes the quantization error of sensor i at iteration t) are random variables following the same Binomial distribution with two possible outcomes, and $v_i(t)$ and $v_i(t+1)$ are independent because the probabilistic quantization processes are independent. This is analogous to flipping a coin twice independently, the outcomes of these two flips are independent.

C. Summary of Algorithm

We now enumerate the steps for our proposed consensus algorithm.
1) Input: η, b, t_0 and M.
2) Generate the sequence $\{\mathbf{x}(t)\}$ using the consensus protocol (7), i.e., $\mathbf{x}(t+1) = \mathbf{x}(t) - \alpha \mathbf{L} Q(\mathbf{x}(t))$, where $Q(\mathbf{x}(t))$ are the quantized data by using the probabilistic quantization scheme.
3) Let the final estimate be the temporal average of the sequence $\{\mathbf{x}(t)\}$, i.e., $(1/M) \sum_{t=t_0}^{t_0+M-1} \mathbf{x}(t)$.

In this algorithm, only the second step requires data exchange among neighboring sensors. The third step involves a simple averaging computation that can be easily implemented. There is no buffer needed to store the previous states. With a specified t_0 and M, each sensor only needs to carry out the accumulation throughout the iterative process and then normalize by M, where the starting point t_0 can be any positive integer of user choice and the sample size M can be determined from (11) given a specified mean square deviation convergence performance.

The choice of η and b should take into account the convergence rate, energy/bandwidth budgets, and signal saturation. Due to the random nature of the quantization, saturation may occur during the iterative process. In this case, data are rounded to the end points. This truncation may have an unpredictable effect on our algorithm. Nevertheless, in [17] (a longer version of [13]), a sample path analysis was carried out and it is shown that the state sequence generated by the quantized consensus protocol is uniformly bounded with high probability, which means that we can choose proper parameters η and b to minimize the chance of data saturation/truncation. Also, our simulation results show that our proposed algorithm is robust to moderate saturation when η and b are set in reasonable ranges.

IV. SIMULATION RESULTS

We present simulation results to illustrate the performance of our proposed algorithm. The sensor network is constructed using a random geographic graph model [18], in which $N = 25$ sensors are placed uniformly at random on a two-dimensional unit area and communicate with their neighbors within a radius r. The transmission radius r is set
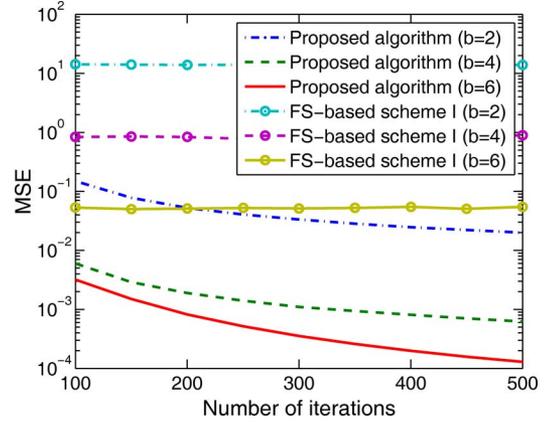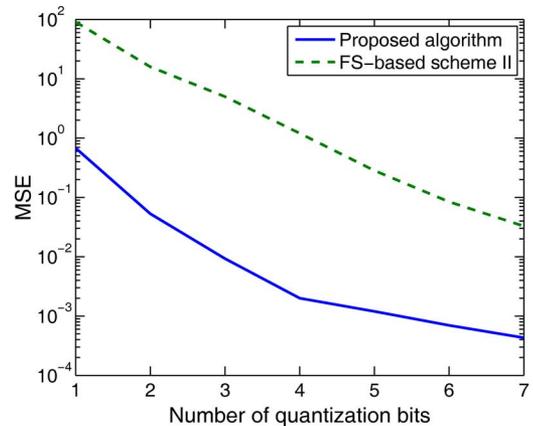

Fig. 1. MSEs versus the number of iterations.

Fig. 2. MSEs versus the number of quantization bits, b. The number of iterations is 200.

to be $\sqrt{(\log N)/N}$ to ensure that the graph is connected with a high probability [18]. The weights assigned to the edges connecting two neighboring sensors are equal to one (see (1)). The initial values of the sensors, $\{x_i(0)\}$, are generated according to a Gaussian distribution with zero mean and unit variance. The performance is measured by an empirical mean square error $\|\mathbf{x}_{\text{ave}}(M) - \bar{\mathbf{x}}_0\|_2^2$ (or $\|\mathbf{x}(t) - \bar{\mathbf{x}}_0\|_2^2$ for other schemes). Results are averaged over 1000 Monte Carlo runs, with the graph and the initial state independently generated for each run (only connected graphs are counted in). To show the effectiveness of our algorithm, we compare our method with two schemes [1] and [14] which use the final states of the sequences $\{\mathbf{x}(t)\}$ generated by protocols (7) and (6) as their estimates, and are denoted by FS-based scheme I and FS-based scheme II, respectively.

Fig. 1 shows the mean square errors (MSEs) of our proposed algorithm and the FS-based scheme I as a function of the number of iterations N_{itr}, where we set $\eta = 2$, $t_0 = 50$ and $M = N_{\text{itr}} - t_0$ for our proposed algorithm, and the number of quantization bits, b, varies from 2 to 6. From Fig. 1, we see that our proposed algorithm presents a clear performance advantage over the FS-based scheme I. Also, our method achieves consistent performance improvement with an increasing N_{itr}, i.e., the sample size M. More specifically, the MSE of our algorithm demonstrates a behavior that is roughly inversely proportional to the sample size M. This observation corroborates our theoretical analysis that defines the relationship between the MSE and the sample size M (see (11)). In contrast, the FS-based scheme I yield little performance gain as the iteration evolves. In Fig. 2, we plot the MSEs of our algorithm and the FS-based scheme II versus the number of quantization

bits, where we set $\eta = 2$, $t_0 = 50$, $N_{\text{itr}} = 200$ and $M = N_{\text{itr}} - t_0$. Again, we see that, even with a moderate number of iterations, a considerable performance improvement is achieved by our algorithm.

V. CONCLUSION

The problem of distributed average consensus with quantized data was studied. A new algorithm was developed by employing a probabilistic quantization scheme and exploiting the temporal information collected from the iterative process. Theoretical analysis proved that the proposed algorithm converges to the average of the initial state in a mean square sense. Simulate results were presented to corroborate our theoretical results.

APPENDIX A
PROOF OF THEOREM 1

Our objective is to show that

$$E[(\mathbf{x}_{\text{ave}}(M) - \bar{\mathbf{x}}_0)^T (\mathbf{x}_{\text{ave}}(M) - \bar{\mathbf{x}}_0)] \to 0 \quad \text{as } M \to \infty. \quad (13)$$

By using the expansion expression of $\mathbf{x}(t)$ (see (8)), we can express $\mathbf{x}_{\text{ave}}$ (for simplicity, we drop the explicit dependence on M) as the summation of following two terms:

$$\mathbf{x}_{\text{ave}} = \frac{1}{M} \sum_{t=t_0}^{t_0+M-1} \mathbf{W}^t \mathbf{x}(0) - \frac{\alpha}{M} \sum_{t=t_0}^{t_0+M-1} \sum_{i=0}^{t-1} \mathbf{W}^{t-1-i} \mathbf{L} \mathbf{v}(i)$$
$$\triangleq \mathbf{f}_1 - \mathbf{f}_2. \quad (14)$$

Since $\mathbf{f}_1$ is deterministic and $\mathbf{f}_2$ is a matrix-weighted combination of the zero-mean random vectors $\{\mathbf{v}(t)\}$, we can decompose the mean-square error as

$$E[(\mathbf{x}_{\text{ave}} - \bar{\mathbf{x}}_0)^T (\mathbf{x}_{\text{ave}} - \bar{\mathbf{x}}_0)]$$
$$= (\mathbf{f}_1 - \bar{\mathbf{x}}_0)^T (\mathbf{f}_1 - \bar{\mathbf{x}}_0) + E\left[\mathbf{f}_2^T \mathbf{f}_2\right] \triangleq \epsilon_1 + \epsilon_2. \quad (15)$$

Considering ϵ_1, we have

$$\epsilon_1 = \mathbf{x}^T(0) \left(\frac{1}{M} \sum_{t=t_0}^{t_0+M-1} \mathbf{W}^t - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right)^T$$
$$\cdot \left(\frac{1}{M} \sum_{t=t_0}^{t_0+M-1} \mathbf{W}^t - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) \mathbf{x}(0)$$
$$\stackrel{(a)}{<} \frac{\rho^{2t_0} \|\mathbf{x}(0)\|_2^2}{M^2 (1-\rho)^2} \triangleq \kappa_1 \quad (16)$$

where ρ denotes the spectral radius of $(\mathbf{W} - (1/N)\mathbf{1}\mathbf{1}^T)$, (a) comes by noting that

$$\mathbf{P} \triangleq \left(\frac{1}{M} \sum_{t=t_0}^{t_0+M-1} \mathbf{W}^t - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) = \frac{1}{M} \sum_{t=t_0}^{t_0+M-1} \left(\mathbf{W} - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right)^t \quad (17)$$

and the largest eigenvalue of $\mathbf{P}$ is upper bounded by $\rho^{t_0}/(M(1-\rho))$ (the derivation is straightforward and thus omitted here). We see that this term vanishes as t_0 or M increases.

Now turning to ϵ_2, we firstly express $\mathbf{f}_2$ as

$$\mathbf{f}_2 = \frac{\alpha}{M} \sum_{t=t_0}^{t_0+M-1} \sum_{i=0}^{t-1} \mathbf{W}^{t-1-i} \mathbf{L} \mathbf{v}(i)$$

$$= \frac{\alpha}{M} \sum_{i=0}^{t_0-1} \left(\sum_{j=0}^{M-1} \mathbf{W}^{t_0-1-i+j} \right) \mathbf{L} \mathbf{v}(i)$$
$$+ \frac{\alpha}{M} \sum_{i=t_0}^{t_0+M-2} \left(\sum_{j=0}^{M-2-i+t_0} \mathbf{W}^j \right) \mathbf{L} \mathbf{v}(i)$$
$$\triangleq \frac{\alpha}{M} \sum_{i=0}^{t_0-1} \mathbf{G}_i \mathbf{L} \mathbf{v}(i) + \frac{\alpha}{M} \sum_{i=t_0}^{t_0+M-2} \mathbf{H}_i \mathbf{L} \mathbf{v}(i). \quad (18)$$

By using (18), we have

$$\epsilon_2 = E[\mathbf{f}_2^T \mathbf{f}_2] = \text{trace}\left(E[\mathbf{f}_2 \mathbf{f}_2^T] \right)$$
$$\stackrel{(a)}{=} \frac{\alpha^2}{M^2} \left(\sum_{i=0}^{t_0-1} \text{trace}\left(\mathbf{G}_i \mathbf{L} \mathbf{C}_{v,i} \mathbf{L}^T \mathbf{G}_i^T \right) \right.$$
$$\left. + \sum_{i=t_0}^{t_0+M-2} \text{trace}\left(\mathbf{H}_i \mathbf{L} \mathbf{C}_{v,i} \mathbf{L}^T \mathbf{H}_i^T \right) \right)$$
$$= \frac{\alpha^2}{M^2} \left(\sum_{i=0}^{t_0-1} \|\mathbf{G}_i \mathbf{L} \mathbf{C}_{v,i}^{1/2}\|_F^2 + \sum_{i=t_0}^{t_0+M-2} \|\mathbf{H}_i \mathbf{L} \mathbf{C}_{v,i}^{1/2}\|_F^2 \right)$$
$$\stackrel{(b)}{\leq} \frac{\alpha^2 N}{M^2} \left(\sum_{i=0}^{t_0-1} \|\mathbf{G}_i \mathbf{L} \mathbf{C}_{v,i}^{1/2}\|_2^2 + \sum_{i=t_0}^{t_0+M-2} \|\mathbf{H}_i \mathbf{L} \mathbf{C}_{v,i}^{1/2}\|_2^2 \right)$$
$$\stackrel{(c)}{<} \frac{\alpha^2 N \lambda_v \lambda_{\max}^2(\mathbf{L})}{M^2} \left(\sum_{i=0}^{t_0-1} \frac{\rho^{2t_0-2-2i}}{(1-\rho)^2} + \frac{M}{(1-\rho)^2} \right)$$
$$\triangleq \kappa_2 + \kappa_3 \quad (19)$$

where (a) follows from the fact that the sequence $\{\mathbf{v}(t)\}$ is independent with finite auto-covariance matrix $E[\mathbf{v}(t)\mathbf{v}(t)^T] = \mathbf{C}_{v,t}$; (b) follows by recalling the matrix norm property [19, p. 56]: $\|\mathbf{X}\|_F \leq \sqrt{N}\|\mathbf{X}\|_2$ for $\mathbf{X} \in \mathbb{R}^{N \times N}$, where $\|\cdot\|_F$ and $\|\cdot\|_2$ denote the Frobenius norm and matrix 2-norm respectively; (c) is obtained by using the results (23)–(24) derived in Appendix B, and defining $\lambda_v \triangleq \max_i \lambda_{\max}(\mathbf{C}_{v,i})$ for $i \in \{t_0, \ldots, t_0 + M - 1\}$.

By combining (16) and (19), and noting that κ_1 and κ_2 are of order $\mathcal{O}(1/M^2)$ while κ_3 is of order $\mathcal{O}(1/M)$, when M is sufficiently large, the mean-square deviation therefore is approximately upper-bounded by κ_3, i.e.,

$$E[\|\mathbf{x}_{\text{ave}} - \bar{\mathbf{x}}_0\|_2^2] < \frac{\alpha^2 N \lambda_v \lambda_{\max}^2(\mathbf{L})}{M(1-\rho)^2}. \quad (20)$$

Since λ_v, and N are finite, $\rho < 1$ and $\lambda_{\max}(\mathbf{L})$ are parameters dependent on the network topology, the mean-square error goes to zero as the sample size, M, tends to infinity.

APPENDIX B
DERIVATION OF (19)

Since $\mathbf{1}^T \mathbf{L} = \mathbf{0}$, the following holds for any integer $t \geq 0$:

$$\mathbf{W}^t \mathbf{L} = \left(\mathbf{W}^t - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right) \mathbf{L} = \left(\mathbf{W} - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right)^t \mathbf{L}. \quad (21)$$

Therefore, we have

$$\mathbf{G}_i \mathbf{L} \mathbf{C}_{v,i}^{1/2} = \sum_{j=0}^{M-1} \left(\mathbf{W} - \frac{1}{N} \mathbf{1}\mathbf{1}^T \right)^{t_0-1-i+j} \mathbf{L} \mathbf{C}_{v,i}^{1/2} \triangleq \tilde{\mathbf{G}}_i \mathbf{L} \mathbf{C}_{v,i}^{1/2}. \quad (22)$$

Hence, $\|\mathbf{G}_i\mathbf{L}\mathbf{C}_{v,i}^{1/2}\|_2^2$ is bounded by

$$
\|\mathbf{G}_i\mathbf{L}\mathbf{C}_{v,i}^{1/2}\|_2^2 = \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^T\mathbf{C}_{v,i}^{1/2}\mathbf{L}^T\tilde{\mathbf{G}}_i^T\tilde{\mathbf{G}}_i\mathbf{L}\mathbf{C}_{v,i}^{1/2}\mathbf{x}
$$

$$
= \max_{\mathbf{y}=\mathbf{L}\mathbf{C}_{v,i}^{1/2}\mathbf{x}} \mathbf{y}^T\tilde{\mathbf{G}}_i^T\tilde{\mathbf{G}}_i\mathbf{y} < \|\mathbf{y}\|_2^2\left(\frac{\rho^{t_0-1-i}}{1-\rho}\right)^2
$$

$$
\overset{(a)}{\le} \lambda_{\max}(\mathbf{C}_{v,i})\lambda_{\max}^2(\mathbf{L})\left(\frac{\rho^{t_0-1-i}}{1-\rho}\right)^2 \tag{23}
$$

where in (a), $\lambda_{\max}(\mathbf{L})$ denotes the largest eigenvalue of $\mathbf{L}$ and $\lambda_{\max}(\mathbf{C}_{v,i})$ denotes the largest eigenvalue of $\mathbf{C}_{v,i}$. By following a similar derivation, $\|\mathbf{H}_i\mathbf{L}\mathbf{C}_{v,i}^{1/2}\|_2^2$ is bounded by

$$
\|\mathbf{H}_i\mathbf{L}\mathbf{C}_{v,i}^{1/2}\|_2^2 < \lambda_{\max}(\mathbf{C}_{v,i})\lambda_{\max}^2(\mathbf{L})\left(\frac{1}{1-\rho}\right)^2. \tag{24}
$$

REFERENCES

[1] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri, "Average consensus on networks with quantized communication," *Int. J. Robust Non-Linear Control*, 2008, to be published.

[2] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, no. 53, pp. 65–78, Feb. 2004.

[3] D. S. Scherber and H. C. Papadopoulos, "Distributed computation of averages over ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 776–787, Apr. 2005.

[4] C. C. Moallemi and B. V. Roy, "Consensus propagation," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 4753–4766, Nov. 2006.

[5] E. Kokiopoulou and P. Frossard, "Accelerating distributed consensus using extrapolation," *IEEE Signal Process. Lett.*, vol. 14, pp. 665–668, Oct. 2007.

[6] E. Kokiopoulou and P. Frossard, "Polynomial filtering for fast convergence in distributed consensus," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 342–354, Jan. 2009.

[7] T. C. Aysal, B. N. Oreshkin, and M. J. Coates, "Accelerated distributed average consensus via localized node state prediction," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1563–1576, Apr. 2009.

[8] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *J. Parallel Distrib. Comput.*, vol. 67, pp. 33–46, 2007.

[9] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links – Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.

[10] S. Barbarossa and G. Scutari, "Decentralized maximum-likelihood estimation for sensor networks composed of nonlinearly coupled dynamical systems," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3456–3470, Jul. 2007.

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

[12] 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.

[13] S. Kar and J. M. F. Moura, "Distributed average consensus in sensor networks with quantized inter-sensor communication," presented at the IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Las Vegas, NV, Mar. 30–Apr. 4, 2008.

[14] T. C. Aysal, M. Coats, and M. Rabbat, "Distributed average consensus with dithered quantization," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4905–4918, Oct. 2008.

[15] B. Mohar, "Some applications of Laplace eigenvalues of graphs," in *Graph Symmetry: Algebraic Methods and Applications*, ser. NATO ASI Series C 497.   Dordrecht, MA: Kluwer, 1997, pp. 225–275.

[16] A. Fasano and G. Scutari, "The effect of additive noise on consensus achievement in wireless sensor networks," presented at the IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Las Vegas, NV, Mar. 30–Apr. 4, 2008.

[17] S. Kar and J. M. F. Moura, *Distributed Consensus Algorithms in Sensor Networks: Quantized Data*, 2007, submitted for publication.

[18] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, pp. 388–404, Mar. 2000.

[19] G. Golub and C. van Loan, *Matrix Computations*, 3rd ed.   Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

A Hidden Markov Model With Binned Duration Algorithm

Stephen Winters-Hilt and Zuliang Jiang

Abstract—The hidden Markov model with duration (HMMD) is critically important when the distributions on state intervals deviate significantly from the geometric distribution, such as for multimodal distributions and heavy-tailed distributions. Heavy-tailed distributions, in particular, are widespread in describing phenomena across the sciences, where the log-normal, student's-T, and Pareto distributions are heavy-tailed distributions that are almost as common as the normal and geometric distributions in descriptions of physical phenomena or man-made phenomena. The standard hidden Markov model (HMM) constrains state occupancy durations to be geometrically distributed, while HMMD avoids this limitation, but at significant computational expense. We propose a new algorithm, hidden Markov model with binned duration, whose result shows no loss of accuracy compared to the HMMD decoding performance and a computational expense that only differs from the much simpler and faster HMM decoding by a constant factor.

Index Terms—Algorithms, artificial intelligence, hidden Markov models, signal processing.

I. INTRODUCTION

The Viterbi and Baum–Welch algorithms are the underlying communication, error-coding, and structure-identification algorithms used in cell-phone communications, deep-space satellite communications, voice recognition, and gene-structure identification [5], [6], with growing applications in areas such as image processing and channel current cheminformatics [3]. The hidden Markov model with duration (HMMD) generalization is important because the standard hidden Markov model (HMM)-based, Viterbi, and Baum—Welch algorithms, are critically constrained in their modeling ability to distributions on state intervals that are geometric. This can lead to a significant decoding failure in noisy environments when the state-interval distributions are not geometric (or approximately geometric), as will be shown in what follows. The starkest contrast occurs for multimodal distributions and heavy-tailed distributions.

Heavy-tailed distributions are widespread in describing phenomena across the sciences [7]. The log-normal and Pareto distributions are heavy-tailed distributions that are almost as common as the normal and geometric distributions in descriptions of physical phenomena or man-made phenomena (such as internet packet size). The length distribution for introns, in particular, has very strong support in an extended "heavy-tail" region; likewise for the length distribution on open reading frames (ORFs) in genomic DNA [3]. In fact, the anomalously long-tailed aspect of the ORF distributions is the key distinguishing feature of this distribution and has been the key attribute used by biologists to identify likely protein-coding regions in genomic DNA since the early days of (manual) gene structure identification. The length distributions on blockade states in channel