

Network-Based Consensus Averaging With General Noisy Channels

Ram Rajagopal and Martin J. Wainwright, *Senior Member, IEEE*

Abstract—This paper focuses on the consensus averaging problem on graphs under general imperfect communications. We study a particular class of distributed consensus algorithms based on damped updates, and using the ordinary differential equation method, we prove that the updates converge almost surely to the consensus average for various models of perturbation of data exchanged between nodes. The convergence is not asymptotic in the size of the network. Our analysis applies to various types of stochastic disturbances to the updates, including errors in update calculations, dithered quantization and imperfect data exchange among nodes. Under a suitable stability condition, we prove that the error between estimated and true averages is asymptotically Gaussian, and we show how the asymptotic covariance is specified by the graph Laplacian. For additive perturbations, we show how the scaling of the asymptotic MSE is controlled by the spectral gap of the Laplacian.

Index Terms—Consensus protocols, distributed averaging, gossip algorithms, graph Laplacian, message-passing, sensor networks, stochastic approximation.

I. INTRODUCTION

CONSENSUS problems, in which a group of nodes want to arrive at a common decision in a distributed manner, have a long history, dating back to seminal work from over 20 years ago [7], [11], [29]. A particular type of consensus estimation is the distributed averaging problem, in which a group of nodes want to compute the average (or more generally, a linear function) of a set of values. Due to its applications in sensor and wireless networking, this distributed averaging problem has been the focus of substantial recent research. The distributed averaging problem can be studied either in continuous-time [25], or in the discrete-time setting (e.g., [8], [13], [22], [30]). In both cases, there is now a fairly good understanding of the conditions under which various distributed averaging algorithms converge, as well as the rates of convergence for different graph structures.

Manuscript received June 09, 2009; accepted August 15, 2010. Date of publication September 16, 2010; date of current version December 17, 2010. This work was supported by NSF Grant DMS-0605165, and an NSF CCF-0545862 CAREER grant. The material in this paper was presented at the Allerton Conference on Control, Computing and Communication, September 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gerald Matz.

R. Rajagopal is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: ramr@eecs.berkeley.edu).

M. J. Wainwright is with the Department of Statistics and the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: wainwrig@eecs.berkeley.edu).

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

Digital Object Identifier 10.1109/TSP.2010.2077282

The bulk of early work on consensus has focused on the case of perfect communication between nodes [12]. Given that noiseless communication may be an unrealistic assumption for sensor networks, a more recent line of work has addressed the issue of noisy communication links. With imperfect observations, many of the standard consensus protocols might fail to reach an agreement. Xiao *et al.* [31] observed this phenomenon, and opted to instead redefine the notion of agreement, obtaining a protocol that allows nodes to obtain a steady-state agreement, whereby all nodes are able to track but need not obtain consensus agreement. Schizas *et al.* [28] study distributed algorithms for optimization, including the consensus averaging problem, and establish stability under noisy updates, in that the iterates are guaranteed to remain within a ball of the correct consensus, but do not necessarily achieve exact consensus. Kashyap *et al.* [21] study consensus updates with the additional constraint that the value stored at each node must be integral, and establish convergence to quantized consensus. Fagnani and Zampieri [14] study the case of packet-dropping channels, and propose various updates that are guaranteed to achieve consensus. Picci *et al.* [27] show almost sure convergence of consensus updates when the communication is noiseless, but the connectivity is random, such as in packet dropping channel models. Aysal *et al.* [3] use probabilistic forms of quantization to develop algorithms that achieve quantized average consensus. Aysal *et al.* [4] propose a broadcasting-based consensus algorithm that converges faster than standard consensus algorithms. Yildiz and Scaglione [32] suggest coding strategies to deal with quantization noise and rate constrained channels, and establish MSE convergence as the number of nodes in the network goes to infinity. In related work, Carli *et al.* [9] propose a sum-preserving difference update to handle quantization and communication noise.

In the current paper, we address the discrete-time average consensus problem in general fixed networks, modeling the communication between neighboring nodes as a general stochastic channel. Our main contributions are to propose and analyze simple distributed protocols that are guaranteed to converge almost surely (a.s.) to the exact consensus mean, and under suitable stability conditions, whose \sqrt{n} -rescaled error (where n is the number of iterations) is asymptotically normal with covariance controlled by the graph structure. These exactness guarantees are obtained using protocols with decreasing step sizes, which smooths out the noise factors. The framework described here is based on the classic ordinary differential equation method [24], which allows us to explicitly identify the deterministic limit, and moreover to establish asymptotic normality. This framework allows for the analysis of several different and important scenarios, namely:

- Noisy storage: stored values at each node are corrupted by noise, with known covariance structure.
- Noisy transmission: messages across each edge are corrupted by noise, with known covariance structure.
- Bit constrained channels: dithered quantization is applied to messages prior to transmission.

In this paper, we analyze protocols that can achieve arbitrarily small mean-squared error (MSE) for distributed averaging with noise. Closely related to part of our results are past work by Hatano *et al.* [16], who analyze averaging under additive noise and use a Lyapunov function approach to prove almost sure convergence to the so-called consensus subspace (i.e., $C = \text{span}(\vec{1})$). Related work by Kar and Moura [18] allows for more general models (including Markovian dynamics), and also uses the Lyapunov function approach to prove almost sure convergence of the updates to a random element of the consensus subspace. This paper studies a more restricted class of models with independent noise, and as a preliminary result, for several important noise models we establish almost sure convergence of the updates to either a deterministic or random element of the consensus subspace, depending on the noise model under consideration. We use the ordinary differential equation or ODE method [24] to approximate the long term behavior of the mean evolution, another technique from stochastic approximation theory. The two techniques (Lyapunov and ODE methods) provide some complementary insight into convergence. For instance, the Lyapunov approach provides insight into transient behavior not afforded by the ODE approach, whereas the ODE approach provides a deterministic path that captures the behavior of the stochastic iteration.

The main contribution of our paper is to show that under appropriate stability conditions, the error in the estimate (after the usual \sqrt{n} -rescaling) is *asymptotically normal*, and to demonstrate explicitly how its asymptotic covariance matrix is controlled by the graph topology. This analysis reveals how different graph structures—ranging from ring graphs at one extreme to expander graphs at the other—lead to different variance scaling behaviors, as determined by the eigenspectrum of the graph Laplacian [10]. Our theory predicts that the number of iterations required to achieve δ -error will scale very differently in the network size for different graph topologies, ranging from quadratic scaling in network size for ring graphs to constant scaling (independence of network size) for expander graphs. In addition, our simulation results show excellent agreement with the theoretical predictions. In subsequent work, other authors have used related techniques to establish asymptotic normality for more general problems of distributed parameter estimation [20].

The remainder of this paper is organized as follows. We begin in Section II by describing the distributed averaging problem in detail, and defining the class of stochastic algorithms studied in this paper. In Section III, we state our main results on the almost-sure convergence and asymptotic normality of our protocols, and illustrate some of their consequences for particular classes of graphs. In particular, we illustrate the sharpness of our theoretical predictions by comparing them to simulation results, on various classes of graphs. Sections IV and V are devoted to

the proofs of our main results, and we conclude the paper with discussion in Section VI.

1) *Comment on Notation:* Throughout this paper, we use the following standard asymptotic notation: for functions f and g , the notation $f(n) = \mathcal{O}(g(n))$ means that $f(n) \leq Cg(n)$ for some constant $C < \infty$; the notation $f(n) = \Omega(g(n))$ means that $f(n) \geq C'g(n)$ for some constant $C' > 0$, and $f(n) = \Theta(g(n))$ means that $f(n) = \mathcal{O}(g(n))$ and $f(n) = \Omega(g(n))$. We use the index n to refer to discrete iterations of the gossip algorithm and the variable t to refer to continuous time. The $|\cdot|$ operator, when applied to a set, corresponds to its cardinality. The notation $L \succeq 0$ means matrix L is positive semidefinite.

II. PROBLEM SETUP

In this section, we describe the distributed averaging problem, and specify the class of stochastic algorithms studied in this paper.

A. Continuous Consensus Matrices and Stochastic Updates

Consider a set of $m = |V|$ nodes, each representing a particular sensing and processing device. We model this system as an undirected graph $G = (V, E)$, with processors associated with nodes of the graph, and the edge set $E \subset V \times V$ representing pairs of processors that can communicate directly. For each node $i \in V$, we let $N(i) := \{j \in V \mid (i, j) \in E\}$ be its neighborhood set.

Suppose that each vertex i makes a real-valued measurement $x(i)$, and consider the goal of computing the average $\bar{x} = 1/m \sum_{i \in V} x(i)$. We assume that $|x(i)| \leq x_{\max}$ for all $i \in V$, as dictated by physical constraints of sensing. For iterations $n = 0, 1, 2, \dots$, let $\theta^n = \{\theta^n(i), i \in V\}$ represent an m -dimensional vector of estimates. Solving the distributed averaging problem amounts to having θ^n converge to $\theta^* := \bar{x} \vec{1}$, where $\vec{1} \in \mathbb{R}^m$ is the vector of all ones. Various algorithms for distributed averaging [8], [25] are based on symmetric consensus matrices $L \in \mathbb{R}^{m \times m}$ with the properties

$$L(i, j) \neq 0 \quad \text{only if } (i, j) \in E \quad (1a)$$

$$L\vec{1} = \vec{0}, \quad \text{and} \quad (1b)$$

$$L \succeq 0. \quad (1c)$$

The simplest example of such a matrix is the *graph Laplacian*, defined as follows. Let $A \in \mathbb{R}^{m \times m}$ be the adjacency matrix of the graph G , i.e., the symmetric matrix with entries

$$A_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

and let $D = \text{diag}\{d_1, d_2, \dots, d_m\}$ where $d_i := |N(i)|$ is the degree of node i . Assuming that the graph is connected (so that $d_i \geq 1$ for all i), the graph Laplacian is given by

$$L(G) = D - A. \quad (3)$$

Our analysis applies to the graph Laplacian, as well as to various weighted forms of graph Laplacian matrices [10], as long as they satisfy the properties (1). For a d -regular graph, it can be shown directly that the normalized graph Laplacian satisfies these properties.

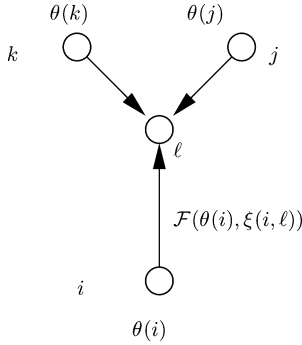


Fig. 1. Illustration of the distributed protocol. Each node $j \in V$ maintains an estimate $\theta(j)$. At each round, for a fixed reference node $\ell \in V$, each neighbor $i \in N(\ell)$ sends the message $\mathcal{F}(\theta(i), \xi(i, \ell))$ along the edge $i \rightarrow \ell$.

The updates are designed to respect the neighborhood structure of the graph G , in the sense that at each iteration, the estimate $\theta^{n+1}(v_r)$ at a receiving node $v_r \in V$ is a function of only¹ the estimates $\{\theta^n(v_t), v_t \in N(v_r)\}$ associated with *transmitting nodes* v_t in the neighborhood of node v_r . In order to model noise and uncertainty in the storage and communication process, we introduce random variables $\xi(v_t, v_r)$ associated with the transmission link from v_t to v_r ; we allow for the possibility that $\xi(v_t, v_r) \neq \xi(v_r, v_t)$, since the noise structure might be asymmetric.

With this setup, we consider algorithms that generate a stochastic sequence $\{\theta^n, n = 0, 1, 2, \dots\}$ in the following manner:

1. At time step $n = 0$, initialize $\theta^0(v) = x(v)$ for all $v \in V$.
2. For time steps $n = 0, 1, 2, \dots$, each node $v_t \in V$ computes the random variables

$$Y^{n+1}(v_r, v_t) = \begin{cases} \mathcal{F}(\theta^n(v_t), \xi^{n+1}(v_t, v_r)) & \text{if } (v_t, v_r) \in E \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where \mathcal{F} is the *communication-noise function* defining the model.

3. Generate estimate $\theta^{n+1} \in \mathbb{R}^m$ as

$$\theta^{n+1} = \theta^n + \epsilon_n \left[- (L \odot Y^{n+1}) \vec{1} \right] \quad (5)$$

where \odot denotes the Hadamard (elementwise) product between matrices, and $\epsilon_n > 0$ is a decaying step size parameter.

See Fig. 1 for an illustration of the message-passing update of this protocol. In this paper, we focus on step size parameters ϵ_n that scale as $\epsilon_n = \Theta(1/n)$. On an elementwise basis, the update (5) takes the form

$$\theta^{n+1}(v_r) = \theta^n(v_r) - \epsilon_n \left[L(v_r, v_r) \mathcal{F}(\theta^n(v_r), \xi^{n+1}(v_r, v_r)) + \sum_{v_t \in N(v_r)} L(v_r, v_t) \mathcal{F}(\theta^n(v_t), \xi^{n+1}(v_t, v_r)) \right].$$

¹In fact, our analysis is easily generalized to the case where $\theta^{n+1}(v_r)$ depends only on vertices $v_t \in N'(v_r)$, where $N'(v_r)$ is a (possibly random) subset of the full neighborhood set $N(v)$. However, to bring our results into sharp focus, we restrict attention to the case $N'(v_r) = N(v_r)$.

B. Communication and Noise Models

It remains to specify the form of the function \mathcal{F} that controls the communication and noise model in the local computation step in (4).

1) *Noiseless Real Number Model*: The simplest model, as considered by the bulk of past work on distributed averaging, assumes noiseless communication of real numbers. This model is a special case of the update (4) with $\xi^n(v_t, v_r) = 0$, and

$$\mathcal{F}(\theta^n(v_t), \xi^{n+1}(v_t, v_r)) = \theta^n(v_t). \quad (6)$$

2) *Additive Edge-Based Noise Model (AEN)*: In this model, the values stored in a node can be observed without noise, so that $\mathcal{F}(\theta^n(v_r), \xi^{n+1}(v_r, v_r)) = \theta^n(v_r)$. The term $\xi^n(v_t, v_r)$ is zero-mean additive random noise variable that is associated with the transmission $v_t \rightarrow v_r$, and the communication function takes the form

$$\mathcal{F}(\theta^n(v_t), \xi^{n+1}(v_t, v_r)) = \theta^n(v_t) + \xi^{n+1}(v_t, v_r). \quad (7)$$

We assume that the random variables $\xi^{n+1}(v_t, v_r)$ and $\xi^{n+1}(v_t', v_r)$ are independent for distinct edges (v_t', v_r) and (v_t, v_r) , and identically distributed with zero-mean and variance $\sigma^2 = \text{Var}(\xi^{n+1}(v_t, v_r))$.

3) *Additive Node-Based Noise Model (ANN)*: In this model, the function \mathcal{F} takes the same form (7) as the edge-based noise model for $v_t \neq v_r$, but the values stored at the node itself are noisy, so that $\mathcal{F}(\theta^n(v_r), \xi^{n+1}(v_r, v_r)) = \theta^n(v_r) + \xi^{n+1}(v_r, v_r)$. Moreover, we assume for each $v_t \in V$, the noise takes the form

$$\xi^{n+1}(v_t, v_r) = \xi^{n+1}(v_t) \quad \text{for all } v_r \in N(v_t) \quad (8)$$

where $\xi^{n+1}(v_t)$ is a single noise variable associated with node v_t , with zero mean and variance $\sigma^2 = \text{Var}(\xi^n(v_t))$. Thus, the random variables $\xi^{n+1}(v_t, v_r)$ and $\xi^{n+1}(v_t, v_r')$ are all *identical* for all edges outgoing from the transmitting node v_t .

4) *Bit-Constrained Communication (BC)*: Suppose that the channel from node v_t to v_r is bit-constrained, so that one can transmit at most B bits, which is then subjected to random dithering. Under these assumptions, the communication function \mathcal{F} takes the form

$$\mathcal{F}(\theta(v_t), \xi(v_t, v_r)) = Q_B(\theta(v_t) + \xi(v_t, v_r)) \quad (9)$$

where $Q_B(\cdot)$ represents the B -bit quantization function with maximum value M and $\xi(v_t, v_r)$ is random dithering. We assume that the random dithering is applied prior to transmission across the channel outgoing from vertex v_t , so that $\xi(v_t, v_r) = \xi(v_t)$ is the same random variable across all neighbors $v_r \in N(v_t)$. Note that the specification of the iteration uses $\mathcal{F}(\theta(v_r), \xi(v_r, v_r)) = Q_B(\theta(v_r) + \xi(v_r))$, so the node value is quantized in its self update as well. We assume that the dithering variables $\xi^n(v_t)$ are independent and bounded by Δ with variance at most σ^2 , and that the initial values satisfy the bound $|\theta^0(v_t)| \leq M - \Delta$ or alternatively $x_{\max} \leq M - \Delta$.

III. STATEMENT OF RESULTS AND CONSEQUENCES

In this section, we state our main results, concerning the stochastic behavior of sequence $\{\theta^n\}$ generated by the updates (5).

We then illustrate its consequences for the specific communication and noise models described in Section II-B, and conclude with a discussion of behavior for specific graph structures.

Consider the factor $L \odot Y$ that drives the updates (5). An important element of our analysis is the conditional covariance of this update factor, denoted by $\Sigma = \Sigma_\theta$ and given by

$$\Sigma_\theta := \mathbb{E} \left[(L \odot Y) \tilde{\mathbf{1}} \tilde{\mathbf{1}}^T (L \odot Y)^T \mid \theta \right] - L\theta(L\theta)^T \quad (10)$$

where

$$Y(v_r, v_t) = \begin{cases} \mathcal{F}(\theta(v_t), \xi(v_t, v_r)) & \text{if } (v_t, v_r) \in E \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

A little calculation shows that the (i, j) th element of this matrix is given by

$$\Sigma_\theta(i, j) = \sum_{k, \ell=1}^m L(i, k)L(j, \ell) \times \mathbb{E} [Y(i, k)Y(j, \ell) - \theta(k)\theta(\ell) \mid \theta]. \quad (12)$$

Moreover, the eigenstructure of the consensus matrix L plays an important role in our analysis. Since it is symmetric and positive semidefinite, we can write

$$L = UJU^T \quad (13)$$

where U be an $m \times m$ orthogonal matrix with columns defined by unit-norm eigenvectors of L , and $J := \text{diag}\{\lambda_1(L), \dots, \lambda_m(L)\}$ is a diagonal matrix of eigenvalues, with

$$0 = \lambda_1(L) < \lambda_2(L) \leq \dots < \lambda_m(L). \quad (14)$$

It is convenient to let \tilde{U} denote the $m \times (m-1)$ matrix with columns defined by eigenvectors associated with positive eigenvalues of L —that is, excluding column $U_1 = \tilde{\mathbf{1}}/\|\tilde{\mathbf{1}}\|_2$, associated with the zero-eigenvalue $\lambda_1(L) = 0$. With this notation, we have

$$\tilde{J} = \text{diag}\{\lambda_2(L), \dots, \lambda_m(L)\} = \tilde{U}^T L \tilde{U}. \quad (15)$$

We demonstrate two theorems: strong consistency and asymptotic normality. For the BC model, we make use of the assumption that the initial conditions are inside the unsaturated region of the quantizer, so that the iteration avoids saturated equilibria and has a unique global equilibrium. Our main result is the asymptotic normality of the procedure. In particular, we analyze the asymptotic error behavior of the procedure for various graphs, and relate the performance to specific graph properties. We are able to identify the error scaling behavior and compare it to the noiseless case.

A. Consistency

We begin by considering the asymptotic behavior of the (unrescaled) sequence $\{\theta^n\}$, and its relation to the average $\theta^* = \bar{x}\tilde{\mathbf{1}} \in \mathbb{R}^m$. Theorem 1 asserts that the sequence $\{\theta^n\}$ converges a.s. to the deterministic quantity θ^* for both the ANN and BC models. As opposed to weak consistency, this result guarantees that for almost any realization of the algorithm, the associated

sample path converges to the exact consensus solution. In contrast, for the AEN model, Theorem 1 shows that the solution converges a random quantity belonging to the consensus subspace, and centered at the average θ^* .

Theorem 1: Consider the random sequence $\{\theta^n\}$ generated by the update (5) with consensus matrix L , and step size parameter $\epsilon_n = \Theta(1/n)$. Then for all initial node value vectors θ^0 .

- (a) For the ANN noise model (8) or the BC noise model (9), we have $\theta^n \rightarrow \theta^*$ a.s.
- (b) Under the AEN model (7), the sequence $\{\theta^n\}$ converges a.s. to $\theta^* + \eta\tilde{\mathbf{1}}$, where the random variable η is zero-mean with variance $\pi^2/6 \sigma^2/m^2 [\sum_{i,j=1}^m [L(i, j)]^2]$.

Remarks: The essential condition for the communication function used in the AEN and ANN models is that

$$\mathbb{E}[\mathcal{F}(\theta(v_t), \xi(v_t, v_r)) \mid \theta(v_t)] = \theta(v_t). \quad (16)$$

A more general condition for the communication function is that

$$\mathbb{E}[\mathcal{F}(\theta(v_t), \xi(v_t, v_r)) \mid \theta(v_t)] = r(\theta(v_t)) \quad (17)$$

where r is a function such that the ODE $\dot{\theta} = -L r(\theta)$ has $\theta = \theta^* \mathbf{1}$ as the unique asymptotically stable equilibrium. One simple condition is if the function r is a monotonic nondecreasing function. The assertions of the theorem continue to hold, except that the matrix $J = r'(\theta^*) L$, so the asymptotic variance is scaled by $1/r'(\theta^*)$. One can perhaps choose functions r that accelerate the mean behavior convergence and reduce the asymptotic variance for the method. Notice that a large $r'(\theta^*)$ factor is helpful in such a situation, but this also implies that more power is being used by the system. If power is constrained (e.g., $r'(\theta^*) = 1$), the only gain is in mean behavior. Note that the convergence behavior depends on the noise model: for the ANN or BC models, we have almost sure (a.s.) convergence to a constant (namely, θ^*), whereas the AEN model yields only a.s. convergence to a subspace, as in the papers [16]–[18].

B. Asymptotic Normality

Theorem 2 establishes that for appropriate choices of consensus matrices, the rate of MSE convergence is of order $1/n$ for the ANN and BC models, since the \sqrt{n} -rescaled error converges to a nondegenerate Gaussian limit. For the AEN model, the asymptotic normality is observed in a $m-1$ dimensional subspace, since the estimation is biased. Such a rate is to be expected in the presence of sufficient noise, since the number of observations received by any given node (and hence the inverse variance of estimate) scales as n . The solution of the Lyapunov (20) specifies the precise form of this asymptotic covariance, which (as we will see) depends on the graph structure.

Theorem 2: Consider the random sequence $\{\theta^n\}$ generated by the update (5) with step size parameter $\epsilon_n = \Theta(1/n)$, and consensus matrix L with second smallest eigenvalue $\lambda_2(L) > 1/2$. Then the following properties hold:

- (a) For the ANN model (8) and BC model (9), we have

$$\sqrt{n}(\theta_n - \theta^*) \xrightarrow{d} N \left(0, U \begin{bmatrix} 0 & 0 \\ 0 & \tilde{P} \end{bmatrix} U^T \right). \quad (18)$$

(b) For the AEN model (7), we have

$$\sqrt{n} U \tilde{U}^T (\theta_n - \theta^*) \xrightarrow{d} N \left(0, U \begin{bmatrix} 0 & 0 \\ 0 & \tilde{P} \end{bmatrix} U^T \right). \quad (19)$$

In both statements, the $(m-1) \times (m-1)$ matrix \tilde{P} is the solution of the continuous time Lyapunov equation

$$\left(\tilde{J} - \frac{I}{2} \right) \tilde{P} + \tilde{P} \left(\tilde{J} - \frac{I}{2} \right)^T = \tilde{\Sigma}_{\theta^*} \quad (20)$$

where \tilde{J} is the diagonal matrix (15), and $\tilde{\Sigma}_{\theta^*} = \tilde{U}^T \Sigma_{\theta^*} \tilde{U}$ is the transformed version of the conditional covariance (10) evaluated at $\theta = \theta^* = \bar{\theta} \mathbf{1}$.

Theorem 2 makes some more specific predictions for different communication models, as we describe here. Under the same conditions as Theorem 2, we define the average mean-squared error (AMSE) as

$$\text{AMSE}(L; \theta^*) := \frac{1}{m} \text{trace}(\tilde{P}(\theta^*)) \quad (21)$$

corresponding to asymptotic error variance, averaged over nodes of the graph. For the AEN model it captures the asymptotic distance to the consensus subspace.

Corollary 1 (Asymptotic MSE for Different Communication Models): Given a consensus matrix L with second-smallest eigenvalue $\lambda_2(L) > 1/2$, the sequence $\{\theta^n\}$ has asymptotic MSE relative to θ^* , given by:

(a) For the additive edge-based noise (AEN) model (7)

$$\text{AMSE}(L; \theta^*) \leq \frac{\sigma^2}{m} \sum_{i=2}^m \left[\frac{\max_{j=1, \dots, m, k \neq j} \sum [L(j, k)]^2}{2\lambda_i(L) - 1} \right] + \text{Var}(n). \quad (22)$$

(b) For the ANN model (8) and the BC model (9)

$$\text{AMSE}(L; \theta^*) = \frac{\sigma^2}{m} \sum_{i=2}^m \left[\frac{[\lambda_i(L)]^2}{2\lambda_i(L) - 1} \right] \quad (23)$$

where the variance term σ^2 is given by the quantization noise $\mathbb{E} [Q_B(\theta + \xi)^2 - \theta^2 \mid \theta]$ for the BC model, and the noise variance $\text{Var}(\xi(i))$ for the ANN model.

Proof: The essential ingredient controlling the asymptotic MSE is the conditional covariance matrix Σ_{θ^*} , which specifies \tilde{P} via the Lyapunov (20). For analyzing model AEN, it is useful to establish first the following auxiliary result. For each $i = 1, \dots, m-1$, we have

$$\tilde{P}_{ii} \leq \frac{\|\Sigma_{\theta^*}\|_2}{2\lambda_{i+1}(L) - 1} \quad (24)$$

where $\|\Sigma_{\theta^*}\|_2 = \|\Sigma\|_2$ is the spectral norm (maximum eigenvalue for a positive semidefinite symmetric matrix). To see this fact, note that

$$\tilde{U}^T \Sigma \tilde{U} \preceq \tilde{U}^T \|\Sigma\|_2 I \tilde{U} = \|\Sigma\|_2 I.$$

Since \tilde{P} satisfies the Lyapunov equation, we have

$$\left(\tilde{J} - \frac{I}{2} \right) \tilde{P} + \tilde{P} \left(\tilde{J} - \frac{I}{2} \right)^T \preceq \|\Sigma\|_2 I.$$

Note that the diagonal entries of the matrix $\left(\tilde{J} - I/2 \right) \tilde{P} + \tilde{P} \left(\tilde{J} - I/2 \right)^T$ are of the form $(2\lambda_{i+1} - 1) \tilde{P}_{ii}$. The difference between the RHS and LHS matrices constitute a positive semidefinite matrix, which must have a nonnegative diagonal, implying the claimed inequality (24).

In order to use the bound (24), it remains to compute or upper bound the spectral norm $\|\Sigma\|_2$, which is most easily done using the elementwise representation (12). In the following analysis, it is important to note that variances will be evaluated at $\theta = \theta^*$.

(a) For the AEN model (7), we have

$$\mathbb{E} [Y(i, k)Y(j, \ell) - \theta_k \theta_\ell \mid \theta] = \mathbb{E} [\xi(i, k)\xi(j, \ell)]. \quad (25)$$

Since we have assumed that the random variables $\xi(i, k)$ on each edge (i, k) are i.i.d., with zero-mean and variance σ^2 , we have

$$\begin{aligned} \mathbb{E} [Y(i, k)Y(j, \ell) - \theta(k)\theta(\ell) \mid \theta] \\ = \begin{cases} \sigma^2 & \text{if } (i, k) = (j, \ell) \text{ and } i \neq k \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Consequently, from the elementwise expression (12), we conclude that Σ is diagonal, with entries

$$\Sigma(j, j) = \sigma^2 \sum_{k \neq j} [L(j, k)]^2$$

so that $\|\Sigma\|_2 = \sigma^2 \max_{j=1, \dots, m} \sum_{k \neq j} [L(j, k)]^2$, which establishes the claim (22).

(b) For the BC model (9), we have

$$\begin{aligned} \mathbb{E} [Y(i, k)Y(j, \ell) - \theta(k)\theta(\ell) \mid \theta] \\ = \begin{cases} \sigma_{\text{qnt}}^2 & \text{if } i = j \text{ and } k = \ell \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

where $\sigma_{\text{qnt}}^2 := \mathbb{E} [Q_B(\theta + \xi)^2 - \theta^2 \mid \theta]$ is the quantization noise variance. Notice this equation is valid for $\theta = \theta^*$, where the expectation of quantization is in the linear regime by assumption of the BC model. Therefore, we have $\Sigma(\theta^*) = \sigma_{\text{qnt}}^2 L^2$, and using the fact that \tilde{U} consists of eigenvectors of L (and hence also L^2 , the Lyapunov (20) takes the form

$$\left(\tilde{J} - \frac{I}{2} \right) \tilde{P} + \tilde{P} \left(\tilde{J} - \frac{I}{2} \right)^T = \sigma_{\text{qnt}}^2 (\tilde{J})^2,$$

which has the explicit diagonal solution \tilde{P} with entries $\tilde{P}_{ii} = \sigma_{\text{qnt}}^2 \lambda_{i+1}^2(L) / (2\lambda_{i+1}(L) - 1)$. Computing the asymptotic MSE $1/m \sum_{i=1}^{m-1} \tilde{P}_{ii}$ yields the claim (23). The proof of the same claim for the ANN model is analogous. \blacksquare

C. Scaling and Graph Topology

We can obtain further insight by considering how Corollary 1 as a function of the graph topology and consensus matrix L . For a fixed graph G , consider the graph Laplacian $L(G)$ defined in (3). It is easy to see that $L(G)$ is always positive semidefinite, with minimal eigenvalue $\lambda_1(L(G)) = 0$, corresponding to the constant vector. For a connected graph, the second smallest eigenvalue $L(G)$ is strictly positive [10]. Therefore, given an undirected graph G that is connected, the most straightforward manner in which to obtain a consensus matrix L satisfying the conditions of Corollary 1 is to rescale the graph Laplacian $L(G)$, as defined in (3), by its second smallest eigenvalue $\lambda_2(L(G))$, thereby forming the rescaled consensus matrix

$$R(G) := \frac{1}{\lambda_2(L(G))} L(G). \quad (26)$$

with $\lambda_2(R(G)) = 1 > 1/2$.

With this choice of consensus matrix, let us consider the implications of Corollary 1(b), in application to the additive node-based noise (ANN) model, for various graphs. Define the normalized trace of the Laplacian

$$\bar{\alpha}(L(G)) := \frac{\text{trace}(L(G))}{m}. \quad (27)$$

For a normalized graph Laplacian, we have $\bar{\alpha}(L(G)) = 1$. Otherwise, for any (unnormalized) graph Laplacian, we have $\bar{\alpha}(L(G)) \leq (m-1)$ and $\bar{\alpha}(L(G)) \geq \lambda_2(L(G))(m-1)/m$. Finally, for graphs with bounded degree d , we have $\bar{\alpha}(L(G)) \leq d$. (See Appendix A for proofs of these properties.) With this definition, we have the following simple corollary, showing that, up to constants, the scaling behavior of the asymptotic MSE is controlled by the second smallest eigenvalue $\lambda_2(L(G))$.

Corollary 2: For any connected graph G , using the rescaled Laplacian consensus matrix (26), the asymptotic MSE for the ANN model (8) satisfies the bounds

$$\frac{\sigma^2 \bar{\alpha}(L(G))}{2\lambda_2(L(G))} \leq \text{AMSE}(R(G); \theta^*) \leq \frac{\sigma^2 \bar{\alpha}(L(G))}{\lambda_2(L(G))} \quad (28)$$

where $\lambda_2(L(G))$ is the second smallest eigenvalue of the graph.

We provide the proof of this claim in Appendix A. Combined with known results from spectral graph theory [10], Corollary 2 allows us to make specific predictions about the number of iterations required, for a given graph topology of a given size m , to reduce the asymptotic MSE to any $\delta > 0$. Recall Theorem 2 guarantees that the asymptotic MSE per node scales as $\text{AMSE}(R(G); \theta^*)n$. Using this fact and Corollary 2, we have

$$n = \Theta\left(\frac{\sigma^2 d}{\lambda_2(L(G)) \delta}\right) \quad (29)$$

for a graph with maximum degree d . If a normalized Laplacian is used in the update, we have the same scaling of n , but with the term d rescaled to 1. It is interesting to note that this scaling (29) is similar but different from the scaling of noiseless updates [8], [13], where the MSE is (with high probability) upper bounded by δ for $n = \Theta(\log(1/\delta)/-\log(1 - \lambda_2(L(G))))$. When the

spectral gap $\lambda_2(L(G))$ shrinks as the graph size grows, then this scaling can be expressed as

$$n = \Theta\left(\frac{\log\left(\frac{1}{\delta}\right)}{\lambda_2(L(G))}\right). \quad (30)$$

Therefore, we pay a price for noise in the updates (the factor $\log(1/\delta)$ versus $1/\delta$), but the graph topology enters the bounds in the same way—namely, in the form of the spectral gap $\lambda_2(L(G))$.

D. Illustrative Simulations

We illustrate the predicted scaling (29) and the role of the Laplacian eigenspectrum via simulations on different classes of graphs. For all experiments reported here, we set the step size parameter $\epsilon_n = 1/(n+100)$. The additive offset serves to ensure stability of the updates in very early rounds, due to the possibly large gain specified by the rescaled Laplacian (26). We performed experiments for a range of graph sizes, for the ANN model (8), with noise variance $\sigma^2 = 0.1$ in all cases. For each graph size m , we measured the number of iterations n required to reach a fixed level δ of mean-squared error.

1) *Cycle Graph:* Consider the ring graph C_m on m vertices, as illustrated in Fig. 2(a). For this example, we use the normalized Laplacian, since all nodes have the same degree, and consequently, we have $\bar{\alpha}(L(G)) = 1$. Panel (b) provides a log-log plot of the MSE versus the iteration number n ; each trace corresponds to a particular sample path. Notice how the MSE over each sample converges to zero. Moreover, since Theorem 2 predicts that the MSE should drop off as $1/n$, the linear rate shown in this log-log plot is consistent with the theory. Fig. 2(c) plots the number of iterations (vertical axis) required to achieve a given constant MSE versus the size of the ring graph (horizontal axis). For the ring graph, it can be shown (see [10]) that the second smallest eigenvalue scales as $\lambda_2(L(C_m)) = \Theta(1/m^2)$, which implies that the number of iterations to achieve a fixed MSE for a ring graph with m vertices should scale as $n = \Theta(m^2)$. Consistent with this prediction, the plot in Fig. 2(c) shows a quadratic scaling; in particular, note the excellent agreement between the theoretical prediction and the data.

2) *Lattice Model:* Fig. 3(a) shows the two-dimensional four nearest-neighbor lattice graph with m vertices, denoted F_m . For this example, we also use the normalized Laplacian Graph, so $\bar{\alpha}(L(G)) = 1$. Again, panel (b) corresponds to a log-log plot of the MSE versus the iteration number n , with each trace corresponding to a particular sample path, again showing a linear rate of convergence to zero. Panel (c) shows the number of iterations required to achieve a constant MSE as a function of the graph size. For the lattice, it is known [10] that $\lambda_2(L(F_m)) = \Theta(1/m)$, which implies that the critical number of iterations should scale as $n = \Theta(m)$. Note that panel (c) shows linear scaling, again consistent with the theory.

3) *Expander Graphs:* Consider a bipartite graph $G = (V_1, V_2, E)$, with $m = |V_1| + |V_2|$ vertices and edges joining only vertices in V_1 to those in V_2 , and constant degree d ; see Fig. 4(a) for an illustration with $d = 3$. A bipartite graph of this form is an expander [1], [2], [10] with parameters

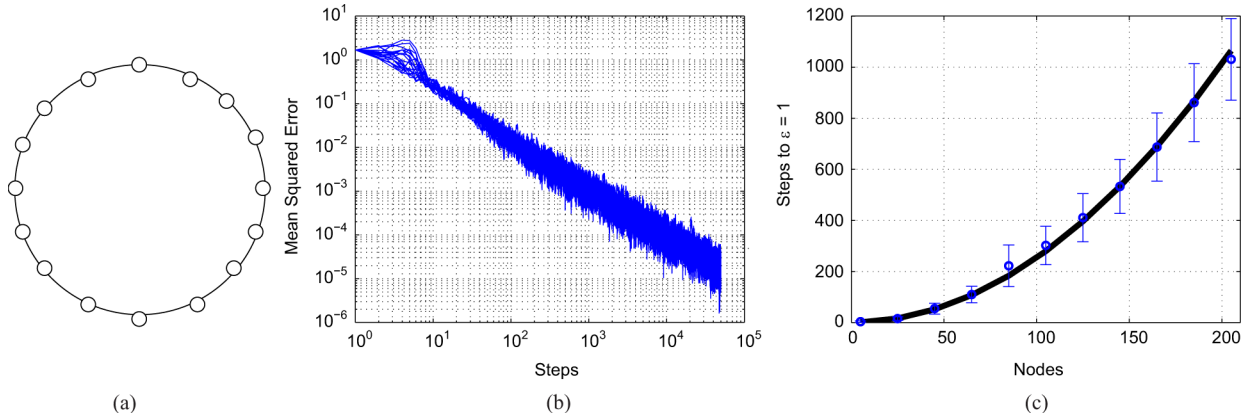


Fig. 2. Simulation of ANN model. Comparison of empirical simulations to theoretical predictions for the ring graph in panel (a). (b) Sample path plots of log MSE versus log iteration number: as predicted by the theory, the log MSE scales linearly with log iterations. (c) Plot of number of iterations (vertical axis) required to reach a fixed level of MSE versus the graph size (horizontal axis). For the ring graph, this quantity scales quadratically in the graph size, consistent with Corollaries 1 and 2. Solid line shows theoretical prediction.

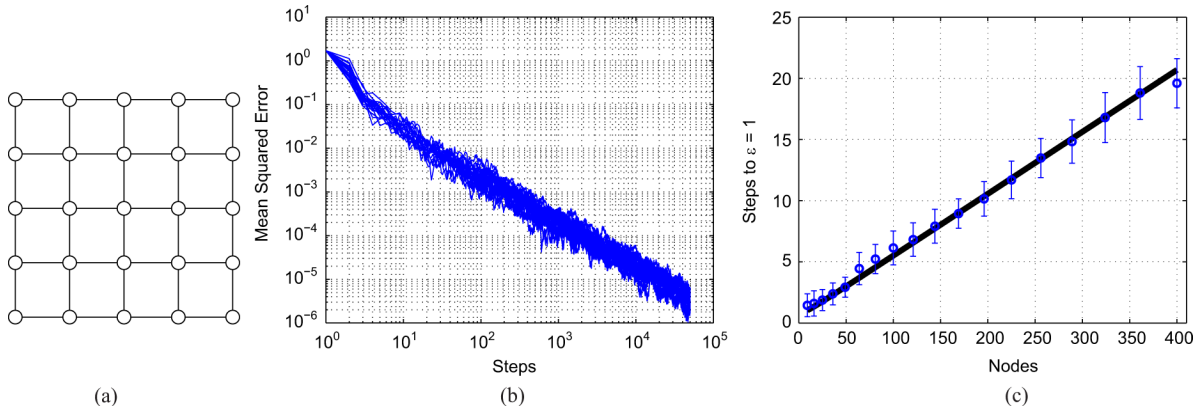


Fig. 3. Simulation of ANN model. Comparison of empirical simulations to theoretical predictions for the four nearest-neighbor lattice [panel (a)]. (b) Sample path plots of log MSE versus log iteration number: as predicted by the theory, the log MSE scales linearly with log iterations. (c) Plot of number of iterations (vertical axis) required to reach a fixed level of MSE versus the graph size (horizontal axis). For the lattice, graph, this quantity scales linearly in the graph size, consistent with Corollaries 1 and 2. Solid line shows theoretical prediction.

$\alpha, \delta \in (0, 1)$, if for all subsets $S \subset V_1$ of size $|S| \leq \alpha|V_1|$, the neighborhood set of S —namely, the subset

$$N(S) := \{t \in V_2 \mid (s, t) \text{ for some } s \in S\}$$

has cardinality $|N(S)| \geq \delta d|S|$. Intuitively, this property guarantees that each subset of V_1 , up to some critical size, “expands” to a relatively large number of neighbors in V_2 . (Note that the maximum size of $|N(S)|$ is $d|S|$, so that δ close to 1 guarantees that the neighborhood size is close to its maximum, for all possible subsets S .) Expander graphs have a number of interesting theoretical properties, including the property that $\lambda_2(L(K_m)) = \Theta(1)$ —that is, a bounded spectral gap [1], [10].

In order to investigate the behavior of our algorithm for expanders, we construct a random bi-partite graph as follows: for an even number of nodes m , we split them into two subsets $V_i, i = 1, 2$, each of size $m/2$. We then fix a degree d , construct a random matching on $dm/2$ nodes, and use it to connect the vertices in V_1 to those in V_2 . This procedure forms a random bipartite d -regular graph; using the probabilistic method, it can be shown to be an edge-expander with probability $1 - o(1)$, as

the graph size tends to infinity [1], [15]. Since the graph is d -regular, the normalized Laplacian can be used for consensus, and $\bar{\alpha}(L(G)) = 1$.

Given the constant spectral gap $\lambda_2(L(K_m)) = \Theta(1)$, the scaling in number of iterations to achieve constant MSE is $n = \Theta(1)$. This theoretical prediction is compared to simulation results in Fig. 4; note how the number of iterations soon settles down to a constant, as predicted by the theory.

4) *Random Geometric Graphs:* Random geometric graphs (RGG) are a class of graphs constructed by distributing m nodes uniformly at random in a unit square, and connecting nodes within distance of r from each other (see [26] for background, and details of the properties that we describe here). It is well known that the RGG will be connected with high probability as long as $r_m = \Omega(\sqrt{\log m/m})$, and Fig. 5(a) shows an example of a random geometric graph with this scaling of the radius. One important observation is that a typical RGG is not degree-regular, so the nonnormalized Laplacian has to be used. Using the fact the graph is connected with high probability, we have the lower bound $\bar{\alpha}(L(G)) \geq 1$. It is also known that the second smallest eigenvalue scales as

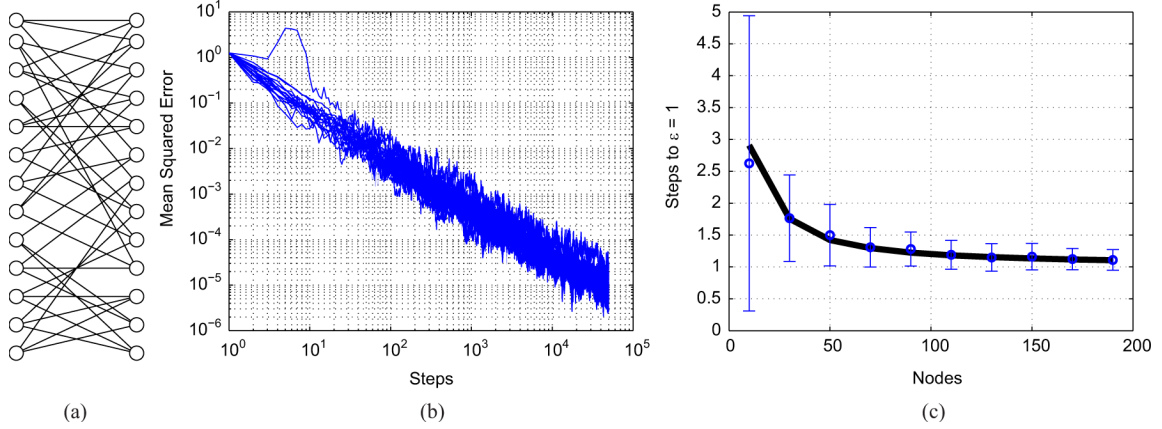


Fig. 4. Simulation of ANN model. Comparison of empirical simulations to theoretical predictions for the bipartite expander graph in panel (a). (b) Sample path plots of log MSE versus log iteration number: as predicted by the theory, the log MSE scales linearly with log iterations. (c) Plot of number of iterations (vertical axis) required to reach a fixed level of MSE versus the graph size (horizontal axis). For an expander, this quantity remains essentially constant with the graph size, consistent with Corollaries 1 and 2. Solid line shows theoretical prediction.

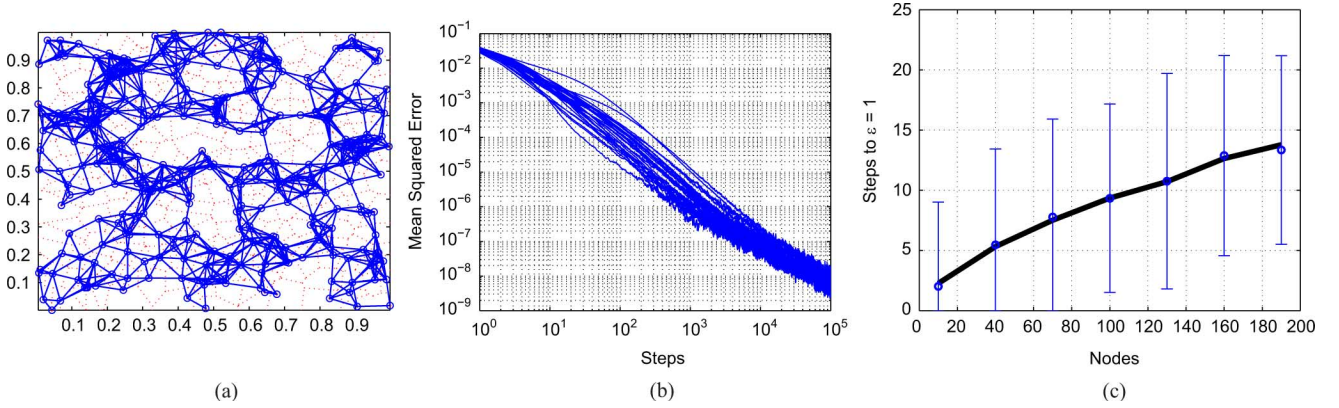


Fig. 5. Simulation of ANN model. Comparison of empirical simulations to theoretical predictions for the random geometric graph in panel (a). (b) Sample path plots of log MSE versus log iteration number: as predicted by the theory, the log MSE scales linearly with log iterations. (c) Plot of number of iterations (vertical axis) required to reach a fixed level of MSE versus the graph size (horizontal axis). For an RGG, this quantity grows linearly with the graph size, consistent with Corollaries 1 and 2. Solid line shows theoretical prediction.

$\lambda_2(L(K_m)) = \mathcal{O}(\log^4 m/m)$. Putting together the pieces, we conclude that $\text{AMSE} = \Omega(\sigma^2 m / \log^4 m)$.

With the radius r_m chosen as above, the expected degree of each vertex scales as $\bar{d} = \mathcal{O}(\log m)$, so that we have $\bar{\alpha}(L(G)) = \mathcal{O}(\log m)$. For the graph Laplacian, it can be shown [5] that Poincaré coefficient κ is $\mathcal{O}(m^2 / \log m)$ with high probability, where a single node is updated per round, chosen uniformly at random. The second smallest eigenvalue of the graph Laplacian is related to this coefficient by $\lambda_2(\tilde{L}(K_m)) \geq 1/\kappa$, where $\tilde{L}(K_m) = 1/m L(K_m)$ is the Laplacian due to the uniformly chosen single node update. Therefore $\lambda_2(L(K_m)) \geq m/\kappa$, and using Corollary 2 we can conclude that $\text{AMSE} = \mathcal{O}(\sigma^2 m)$.

Fig. 5(c) shows the empirical scaling obtained by applying our algorithm, in comparison to the theoretical prediction. Here the theoretical bound is also computed as the average of the theoretical AMSE of the realized graphs, since this quantity is a function of the whole spectrum of each realized graph instance. It is a well known fact that the average behavior of RGGs can be approximated by a grid graph, so it is not a surprise that we observe a similar AMSE scaling behavior. Fig. 5 confirms this

intuition; note however that the variability is larger, since (in contrast to a fixed grid) the graph spectrum is now also random.

Proof of Theorem 1: We now turn to the proof of Theorem 1. The basic idea is to relate the behavior of the stochastic recursion (5) to an ordinary differential equation (ODE), and then use the ODE method [24] to analyze its properties. The ODE involves a function $t \mapsto \theta_t \in \mathbb{R}^m$, with its specific structure depending on the communication and noise model under consideration. For the ANN model, the relevant ODE is given by

$$\frac{d\theta_t}{dt} = -L\theta_t. \quad (31)$$

For the BC model, the approximating ODE is given by

$$\frac{d\theta_t}{dt} = -L C_M(\theta_t) \quad \text{with} \quad C_M(u) := \begin{cases} u & \text{if } |u| < M \\ -M & \text{if } u \leq -M \\ +M & \text{if } u \geq +M. \end{cases} \quad (32)$$

In both cases, the ODE must satisfy the initial condition $\theta_0(v) = x(v)$.

For the AEN model, an ODE in a projected subspace is satisfied. In particular, given the projected variable $\tilde{\theta}_t = \tilde{U}^T \theta_t \in \mathbb{R}^{m-1}$, the approximating ODE is given by

$$\frac{d\tilde{\theta}_t}{dt} = -\tilde{J}\tilde{\theta}_t, \quad \text{with initial condition } \tilde{\theta}_0 = \tilde{U}^T x. \quad (33)$$

The following lemma summarizes a key technical result, connecting the discrete-time stochastic process $\{\theta^n\}$ to the deterministic ODE solution, and allows us to prove Theorem 1.

Lemma 1:

- (a) The ODEs (31) and (32) with initial condition $\theta_0(v) = x(v)$ each have $\theta^* = \bar{x}\bar{1}$ as their unique stable fixed point, and for all $\delta > 0$, we have
- $$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \|\theta^n - \theta^*\| > \delta\right) = 0, \quad \text{for } t_n = \sum_{k=1}^n \frac{1}{k} \quad (34)$$

which implies that $\theta^n \rightarrow \theta^*$ a.s.

- (b) In the AEN model, the reduced ODE (33) with the given initial condition $\tilde{\theta}_0$ has $\tilde{\theta}^* = 0$ as its unique stable fixed point, and $\tilde{\theta}^n \rightarrow \tilde{\theta}^*$ a.s.

The proof of this claim for each noise model relies on three steps: 1) first, we show that the corresponding ODE has a unique fixed point; 2) second, we prove the stochastic iteration is bounded; and 3) third, we conclude by establishing that the stochastic iteration follows the ODE closely. One issue that needs careful attention is the zero eigenvalue of the Laplacian matrix. Our strategy in following these three steps is to verify that [23, Ch. 5, Th. 2.1] can be applied, which allows us to link the stochastic iteration to the ODE.

E. Establishing Uniqueness of Fixed Points

We begin with step (1), establishing uniqueness and global asymptotic stability of fixed points.

1) *Analysis for ANN and AEN Models:* Using the eigendecomposition $L = UJU^T$, we define the rotated vector $\gamma_t = U^T \theta_t \in \mathbb{R}^m$. Note that we have $J_{11} = 0$, corresponding to the zero eigenvalue for the subspace $U_1 = (\bar{1}^T/\sqrt{m})$. Consequently, we have $d\gamma_t(1)/dt = 0$, and hence $\gamma_t(1) = c$ where $c = U_1^T x$. If we define $\tilde{\gamma}_t(k-1) = \gamma_t(k)$ for $k = 2, \dots, m$, then the reduced ODE for remaining subspace takes the form

$$\frac{d\tilde{\gamma}_t}{dt} = -\tilde{J}\tilde{\gamma}_t. \quad (35)$$

We now show that $\tilde{\gamma} = 0$ is globally asymptotically via LaSalle's principle. In order to do so, we denote $h(\tilde{\gamma}) = -\tilde{J}\tilde{\gamma}$, and then define the function $V(\tilde{\gamma}) = \|\tilde{\gamma}\|_2^2$. We note that $V(\tilde{\gamma}) > 0$ for all $\tilde{\gamma} \neq 0$, and moreover that $\nabla V(\tilde{\gamma})^T h(\tilde{\gamma}) < 0$ for all $\tilde{\gamma} \neq 0$, so that LaSalle's principle implies that $\tilde{\gamma} = 0 \in \mathbb{R}^{m-1}$ is globally asymptotically stable in the reduced subspace. The same argument shows that $\tilde{\gamma} = 0$ is also globally asymptotically for the AEN model.

2) *Analysis for the BC Model:* For the BC model, the analysis is somewhat more involved, since the quantization function saturates the output at $\pm M$. For the dithered quantization model (9), we have

$$\mathbb{E}_\xi[-(L \odot Y(\theta_t, \xi)) \mid \theta_t] = -L C_M(\theta_t)$$

where $C_M(\cdot)$ is the saturation function (32). We now claim that θ^* is also the unique asymptotically stable fixed point of the ODE $d\theta_t/dt = -L C_M(\theta_t)$ subject to the initial condition $\theta_0(v) = x(v)$. Consider the eigendecomposition $L = UJU^T$, where $J = \text{diag}\{0, \lambda_2(L), \dots, \lambda_m(L)\}$. Define the rotated variable $\gamma_t := U^T \theta_t$, so that the ODE (32) can be rewritten as

$$\frac{d\gamma_t(1)}{dt} = 0 \quad (36a)$$

$$\frac{d\gamma_t(k)}{dt} = -\lambda_k(L)U_k^T C_M(U\gamma_t), \quad \text{for } k = 2, \dots, m \quad (36b)$$

where U_k denotes the k^{th} column of U . Note that $U_1 = \bar{1}/\|\bar{1}\|_2$, since it is associated with the eigenvalue $\lambda_1(L) = 0$. Consequently, the solution to (36a) takes the form

$$\gamma_t(1) = U_1^T \theta_0 = \sqrt{m} \bar{x} \quad (37)$$

with unique fixed point $\gamma^*(1) = \sqrt{m} \bar{x}$, where $\bar{x} := 1/m \sum_{i=1}^m x(i)$ is the average value. Define $\tilde{\gamma}$ as before. Substituting $U\gamma_t = \bar{x}\bar{1} + \tilde{U}\tilde{\gamma}$, we obtain the reduced ODE $d\tilde{\gamma}_t/dt = -\tilde{J}\tilde{U}^T C_M(\bar{x}\bar{1} + \tilde{U}\tilde{\gamma}_t)$. A fixed point $\tilde{\gamma}^* \in \mathbb{R}^{m-1}$ for this ODE requires $\tilde{U}^T C_M(\bar{x}\bar{1} + \tilde{U}\tilde{\gamma}_t) = 0$. Given that the columns of U form an orthogonal basis, this implies $C_M(\bar{x}\bar{1} + \tilde{U}\tilde{\gamma}^*) = \alpha\bar{1}$ for some constant $\alpha \in \mathbb{R}$. Given the relation $\theta^* = \bar{x}\bar{1} + \tilde{U}\tilde{\gamma}^*$, we conclude that $C_M(\theta^*) = \alpha\bar{1}$.

Therefore, given the piecewise linear nature of the saturation function, one of three following options must hold: (a) in the case $\alpha = M$, we have the elementwise inequality $\theta^* > M$; (b) in the case $\alpha = -M$, we have the elementwise inequality $\theta^* < -M$; or, as the final option, (c) in the case $\alpha \in (-M, +M)$, we have $\theta^* = \alpha$. In case (a), some algebra yields $U_1^T \theta^* > \sqrt{m}M$. Using $\theta^* = \bar{x}\bar{1} + \tilde{U}\tilde{\gamma}^*$ and the properties of U , this leads to $\bar{x} > M$, which is a contradiction. Similarly, case (b) leads to the contradiction $\bar{x} < -M$. Turning to the only remaining possibility (namely, case (c)), we conclude $\theta^* = \alpha\bar{1}$ for some constant $\alpha \in (-M, +M)$. Using the relation $\theta^* = \bar{x}\bar{1} + \tilde{U}\tilde{\gamma}^*$ together with the orthogonality of $\bar{1}$ and columns of \tilde{U} , we conclude that the the unique solution is $\alpha = \bar{x}$ and $\tilde{\gamma} = 0$. Finally, we verify global asymptotic stability via LaSalle's principle. Defining the function $V(\tilde{\gamma}) = \tilde{\gamma}^T \tilde{J}^{-1} \tilde{\gamma}$, we note that $V(\tilde{\gamma}) > 0$ for all $\tilde{\gamma} \neq 0$, and moreover

$$\begin{aligned} \nabla V(\tilde{\gamma})^T [-\tilde{J}\tilde{U}^T C_M(\bar{x}\bar{1} + \tilde{U}\tilde{\gamma})] \\ &= -2\tilde{\gamma}^T \tilde{U}^T C_M(\bar{x}\bar{1} + \tilde{U}\tilde{\gamma}) \\ &= -2[U^T(\theta - \bar{x}\bar{1})]^T [U^T(C_M(\theta) - \bar{x}\bar{1})] \\ &= -2(\theta - \bar{x}\bar{1})^T (C_M(\theta) - \bar{x}\bar{1}). \end{aligned}$$

This quantity is strictly negative as long as $\theta \neq \bar{x}\bar{1}$, or equivalently as long as $\tilde{\gamma} \neq 0$. Consequently, by La Salle's stability theorem, the fixed point $\tilde{\gamma} = 0$ is globally asymptotically stable.

In general quantization scenarios requiring less assumptions, the quantized iterates have to be shown to return infinitely often to the original quantization range (i.e., all values avoid saturating the quantizer). Such treatment is developed in the paper [19] using the martingale property of the updates.

F. Establishing Boundedness and Equivalence to ODE

We now turn to steps (2) and (3). In order to establish boundedness of sample paths and equivalence to the ODE, we make use of the following result.

Lemma 2: Consider the iteration $\theta^{n+1} = \theta^n + \epsilon_n H^n$ with deterministic initial condition. Let $\epsilon_n = 1/n$ and $h(\theta) = \mathbb{E}[H^n | \theta_0, \dots, \theta^n]$. Suppose that there exists a function $V : \mathbb{R}^m \rightarrow \mathbb{R}$ with bounded mixed derivatives such that $V(\theta) > 0$ for all $\theta \neq 0$, and $|V(\theta)| \rightarrow \infty$ if $\|\theta\|_2 \rightarrow \infty$. Suppose moreover that

$$\nabla V(\theta)^T h(\theta) < 0 \quad \text{for } \|\theta\|_2 > \epsilon, \quad \text{and} \quad (38a)$$

$$\mathbb{E}\|H^n\|_2^2 + |\nabla V(\theta)^T h(\theta)| \leq K_1 V(\theta) + K_2. \quad (38b)$$

Then we are guaranteed that (i) $\sup_n \mathbb{E}\|H^n\|_2^2 < \infty$ and (ii) $\sup_n |\theta^n| < \infty$ with probability one.

As shown in Appendix B, this result follows by adapting results from [23].

1) *Analysis of ANN Model:* In the ANN case, we have $(\theta(v), \xi(v, v_r)) = \theta(v) + \xi$, which implies that $\theta^{n+1} = \theta^n - \epsilon_n L(\theta^n + \xi)$. Let us transform from θ^n to the new variables $\gamma^n = U^T \theta^n$, and note that as in the ODE, the first component takes the form $\gamma^n(1) = U_1^T x$. Disregarding this first component, the reduced iteration in \mathbb{R}^{m-1} takes the form

$$\tilde{\gamma}^{n+1} = \tilde{\gamma}^n - \epsilon_n \tilde{J}(\tilde{\gamma}^n + \tilde{U}^T \xi). \quad (39)$$

Based on our earlier discussion, the function $V(\tilde{\gamma}) = \|\tilde{\gamma}\|_2^2$ that we considered earlier is a valid candidate for applying Lemma 2 as long as (38a) and (38b) hold. Note that we have $h(\tilde{\gamma}) = -\tilde{J}\tilde{\gamma}$ and $\nabla V(\tilde{\gamma}) = 2\tilde{\gamma}$, as well as $\mathbb{E}\|H^n\|_2^2 = \tilde{\gamma}^T \tilde{J}^2 \tilde{\gamma} + m\sigma^2$. As verified earlier, (38a) holds; moreover, (38b) holds with $K_2 = m\sigma^2$ and $K_1 = 2\lambda_m(L)$. Therefore, we conclude that the iteration is bounded, and moreover that $\sup_n \mathbb{E}\|H^n\|_2^2 < \infty$. We can now apply [23, Ch. 5, p. 127, Th. 2.1], since we have verified (A2.1)–(A2.5) required for this result. We conclude that the iterates (39) strongly approximates the ODE in (35), and hence that $\tilde{\gamma}^n$ converges a.s. to 0. Since $\gamma^n(1) = U_1^T x$ for the first-coordinate, we conclude that γ^n converges a.s. to the vector $[U_1^T x \ 0 \ \dots \ 0]$. Using the inverse transformation back to the original coordinates, we conclude that the sequence $\theta^n = U\gamma^n$ converges a.s. to $\bar{x}\bar{1}$, as claimed.

2) *Analysis of AEN Model:* For the AEN model, the argument is analogous except for a few important details. The iteration now takes the form $\theta^{n+1} = \theta^n - \epsilon_n [L\theta^n + (L \odot \xi^{n+1}) \bar{1}]$. In this case, the noise vector is still zero-mean with independent components; the variance of component i is $\sigma^2(i) = \sum_{j=1}^m L^2(i, j)$. After applying the transformation $\theta \mapsto \gamma = U^T \theta$, we need to consider two terms in the transformed noise. The first term $\tilde{U}^T (L \odot \xi^{n+1}) \bar{1}$ has bounded variance, and can be controlled with an argument analogous to the ANN case. The other new term is the quantity $U_1(L \odot \xi^{n+1}) \bar{1}$ for the first component of γ , which is no longer equal to zero. We instead obtain the univariate update

$\gamma^{n+1}(1) = \gamma^n(1) + \epsilon_n \tilde{\xi}^{n+1}(1)$, where $\tilde{\xi}^{n+1}(1)$ has variance $\sigma^2/m \sum_{i,j=1}^m L^2(i, j)$. Unwrapping this recursion, we obtain

$$\gamma^n(1) = U_1^T x + \sum_{k=1}^n \epsilon_k \tilde{\xi}^{k+1}(1).$$

We conclude the proof by noting the inverse relation $\theta = U\gamma$. After this transformation, the second term introduces the vector noise term $\eta = \lim_{n \rightarrow +\infty} \gamma^n(1) \bar{1} / \sqrt{m}$, which has variance as stated in Theorem 1. Here we have recalled that $U_1 = \bar{1} / \sqrt{m}$ by definition.

3) *Analysis of BC Model:* Establishing $V(\tilde{\gamma})$ as a valid test function in Lemma 2 requires verifying that (38b) holds. Using our choice of V and h for this model, we obtain

$$\begin{aligned} |\nabla V(\tilde{\gamma})^T h(\tilde{\gamma})| &= |2\tilde{\gamma}^T \tilde{U}^T C_M(\bar{x}\bar{1} + \tilde{U}\tilde{\gamma})| \\ &= |2(\theta - \bar{x}\bar{1})^T (C_M(\theta) - \bar{x}\bar{1})| \\ &\leq 2(\|\theta\|_2^2 + mM\bar{x}) \end{aligned}$$

where we have used $\bar{1}^T \theta = m\bar{x}$. Moreover, we have

$$\begin{aligned} \mathbb{E}\|H^n\|_2^2 &= \mathbb{E}[-\tilde{J}\tilde{U}^T Q_B(\bar{x}\bar{1} + \tilde{U}\tilde{\gamma} + \xi)]^T \\ &\quad \times [-\tilde{J}\tilde{U}^T Q_B(\bar{x}\bar{1} + \tilde{U}\tilde{\gamma} + \xi)] \\ &= \mathbb{E}[Q_B(\theta + \xi)^T L^2 Q_B(\theta + \xi)] \\ &\leq \|L\|_2^2 \mathbb{E}\|Q_B(\theta + \xi)\|_2^2 \\ &\leq \lambda_m^2(L) \mathbb{E}\|\theta + \xi + \Delta\bar{1}\|_2^2 \\ &= \lambda_m^2(L) (\|\theta\|_2^2 + 2m\Delta\bar{x} + m\Delta^2 + m\sigma^2). \end{aligned}$$

Noting $\|\theta\|_2^2 = \|\tilde{\gamma}\|_2^2 + m\bar{x}^2$, some algebra shows yields that the growth condition (38b) holds with $K_1 = (2 + \lambda_m^2(L))\lambda_m(L)$ and $K_2 = \lambda_m^2(L) (2m\Delta\bar{x} + m\Delta^2 + m\sigma^2) + (2 + \lambda_m^2(L)) m\bar{x}^2 + 2mM\bar{x}$. As before, we have verified [23, conditions A.2.1 to A.2.5 of Th. 2.1], and proved that the iteration remains bounded with probability 1, and the equilibrium is globally asymptotic stable, so that the claim follows.

Proof of Theorem 2: We analyze the update (5) using results from [6]. In particular, given the stochastic iteration $\theta^{n+1} = \theta^n + \epsilon_n H(\theta^n, Y^{n+1})$, define the expectation $h(\theta) = \mathbb{E}[H(\theta, X)]$, its Jacobian matrix $\nabla h(\theta)$, and the covariance matrix $\Sigma(\theta) = \mathbb{E}[(H(\theta, X) - h(\theta))(H(\theta, X) - h(\theta))^T]$. Then [6, Th. 3, p. 110] asserts that as long as the eigenvalues $\lambda(\nabla h(\theta))$ are strictly below $-1/2$, we have

$$\sqrt{n}(\theta_n - \theta^*) \xrightarrow{d} N(0, Q) \quad (40)$$

where the covariance matrix Q is the unique solution to the Lyapunov equation

$$\left(\frac{I}{2} + \nabla h(\theta^*)\right) Q + Q \left(\frac{I}{2} + \nabla h(\theta^*)\right)^T + \Sigma_{\theta^*} = 0. \quad (41)$$

We begin by proving part (a) of the claim. For the model ANN, the conditional expectation $h(\theta)$ takes the form

$$h(\theta) = -L\theta \quad (42)$$

since the conditional expectation of the random matrix Y is given by $\mathbb{E}[Y \mid \theta] = \theta \bar{\mathbf{1}}$. For the BC model, since the quantization is finite with maximum value M , the expectation is given by

$$h(\theta) = -L C_M(\theta) \quad (43)$$

where the saturation function was defined previously in (32). In addition, we computed form of the covariance matrix Σ_{θ^*} previously in (10). Finally, we note that

$$\nabla h(\theta^*) = -L \quad (44)$$

for both models. (This fact is immediate for model ANN; for the BC model, note that Theorem 1 guarantees that θ^* falls in the middle linear portion of the saturation function.)

We cannot immediately conclude that asymptotic normality (40) holds, because the matrix L has a zero eigenvalue ($\lambda_1(L) = 0$). However, let us decompose $L = UJU^T$ where U is the matrix with unit norm columns as eigenvectors, and $J = \text{diag}\{0, \lambda_2(L), \dots, \lambda_m(L)\}$. Let \tilde{U} denote the $m \times (m-1)$ matrix obtained by deleting the first column of U . Defining the $(m-1)$ vector $\beta^n = \tilde{U}^T \theta^n$, we can rewrite the update in $(m-1)$ -dimensional space as

$$\beta^{n+1} = \beta^n + \frac{1}{n} \left[-\tilde{U}^T (L \odot Y^{n+1}(\theta^n)) \bar{\mathbf{1}} \right] \quad (45)$$

for which the new effective h function is given by $\tilde{h}(\beta) = -\tilde{J}\beta$, with $\tilde{J} = \text{diag}\{\lambda_2(L), \dots, \lambda_m(L)\}$. Since $\lambda_2(L) > 1/2$ by assumption, the asymptotic normality (40) applies to this reduced iteration, so that we can conclude that

$$\sqrt{n} (\beta^n - \beta^*) \xrightarrow{d} N(0, \tilde{P})$$

where \tilde{P} solves the Lyapunov equation

$$\left(\tilde{J} - \frac{I}{2} \right) \tilde{P} + \tilde{P} \left(\tilde{J} - \frac{I}{2} \right)^T = \tilde{U}^T \Sigma_{\theta^*} \tilde{U}.$$

We conclude by noting that the asymptotic covariance of θ^n is related to that of β^n by

$$P = U \begin{bmatrix} 0 & 0 \\ 0 & \tilde{P} \end{bmatrix} U^T \quad (46)$$

from which Theorem 2(a) follows.

For Theorem 2(b), it suffices to note that the reduced iteration in (45) holds, so asymptotic normality holds in the $m-1$ projected dimensions, as claimed.

IV. DISCUSSION

In this paper, we analyzed the convergence and asymptotic behavior of distributed averaging algorithms on graphs with general noise models. Using suitably damped updates, we showed that it is possible to obtain exact consensus, as opposed to approximate or near consensus, even in the presence of noise.

We guaranteed almost sure convergence of our algorithms under fairly general conditions, and moreover, under suitable stability conditions, we showed that the error is asymptotically normal, with a covariance matrix that can be predicted from the structure of the underlying graph. We provided a number of simulations that illustrate the sharpness of these theoretical predictions. One interesting consequence is that in the presence of noise, the number of iterations required to achieve an error of δ is $\Theta(\sigma^2 d / \lambda_2(L(G)) 1/\delta)$. This rate should be contrasted with the rate $\Theta(\log(1/\delta) / \lambda_2(L(G)))$ achievable by standard gossip algorithms under perfect (noiseless) communication. This comparison shows that there is some loss in convergence rates due to noisiness, but the influence of the graph structure—namely, via the spectral gap $\lambda_2(L(G))$ —is similar. Finally, although the current paper has focused exclusively on the averaging problem, the methods of analysis in this paper are applicable to other types of distributed inference problems, such as computing quantiles or order statistics, as well as computing various types of M -estimators. Obtaining analogous results for more general problems of distributed statistical inference is an interesting direction for future research.

APPENDIX

Proof of Corollary 2: We observe that $\text{trace}(L(G)) = \sum_{i \in V} d_i$, where d_i are the positive diagonal elements of $L(G)$. For the normalized Laplacian, we have $\text{trace}(L(G)) = m$. A direct computation shows that the normalized Laplacian is a valid consensus matrix only when all nodes have the same degree. Since the smallest eigenvalue is 0, and the others are positive, we have the bound $(m-1)\lambda_2(L(G)) \leq \sum_{i \in V} d_i$. Define

$$\bar{\alpha}(L(G)) := \frac{\text{trace}(L(G))}{m}$$

and notice that for normalized graphs, $\bar{\alpha}(L(G)) = 1$. Furthermore, for any graph we have the lower bound $\bar{\alpha}(L(G)) \geq \lambda_2(L(G)) (m-1)/m$, and in addition, we also have the upper bound $\bar{\alpha}(L(G)) \leq (m-1)$.

Using these facts, we establish Corollary 2 as follows. Recall that by construction, we have $R(G) = L(G)/\lambda_2(L(G))$, so that the second smallest eigenvalue of $R(G)$ is $\lambda_2(R(G)) = 1$, and the remaining eigenvalues are greater than or equal to one. Applying Corollary 1 to the ANN model, we have

$$\begin{aligned} \text{AMSE}(L; \theta^*) &= \frac{\sigma^2}{m} \sum_{i=2}^m \left[\frac{[\lambda_i(R(G))]^2}{[2\lambda_i(R(G)) - 1]} \right] \\ &= \frac{\sigma^2}{m \lambda_2(L(G))} \\ &\quad \times \sum_{i=2}^m \left[\frac{[\lambda_i(L(G))]^2}{[2\lambda_i(L(G)) - \lambda_2(L(G))]} \right] \\ &\geq \frac{\sigma^2}{2\lambda_2(L(G)) m} \text{trace}(L(G)) \\ &= \frac{\sigma^2 \bar{\alpha}(L(G))}{2\lambda_2(L(G))}. \end{aligned}$$

In the other direction, using the fact that $\lambda_2(R(G)) \geq 1$ and the bound $x^2/2x - 1 \leq x$ for $x \geq 1$, we have

$$\begin{aligned} \text{AMSE}(L; \theta^*) &= \frac{\sigma^2}{m} \sum_{i=2}^m \left[\frac{[\lambda_i(R(G))]^2}{2\lambda_i(R(G)) - 1} \right], \\ &\leq \frac{\sigma^2}{m} \text{trace}(R(G)) \\ &= \frac{\sigma^2}{\lambda_2(L(G))m} \text{trace}(L(G)) \\ &= \frac{\sigma^2 \bar{\alpha}(L(G))}{\lambda_2(L(G))} \end{aligned}$$

as claimed.

Proof of Lemma 2: Part (a) follows directly from [23 Lemma 4.1, pp. 146–147]. In order to establish part (b), we define the set $\mathcal{V} = \{\theta : V(\theta) \leq V(\theta^0)\}$ and its complement \mathcal{V}^c . Clearly, $\theta^0 \in \mathcal{V}$. We next adapt an argument in [23, p. 148] to show that the event $\theta^n \in \mathcal{V}^c$ cannot happen infinitely often. When this holds, the claimed boundedness follows as a consequence. Indeed, we let τ be the last exit time such that $\theta^\tau \in \mathcal{V}^c$. By definition, we then have $\theta^n \in \mathcal{V}$ for all $n > \tau$, so that the sequence is bounded. Moreover, since $\tau < \infty$, the terms $\theta^k = \sum_{r=0}^{k-1} \epsilon_r H^r$ are all finite for each $k = 1, \dots, \tau$. The last statement holds for each sample path ω , so $\theta_n(\omega, \cdot)$ is bounded for almost all ω , and thus with probability one.

We now establish that the event $\theta^n \in \mathcal{V}^c$ cannot happen infinitely often. Using the definition $\theta^{n+1} = \theta^n + \epsilon_n H^n$, a truncated Taylor series expansion, and boundedness of mixed partial derivatives of V , we obtain

$$\begin{aligned} V(\theta^{n+1}) &= V(\theta^n) + \epsilon_n \nabla V(\theta^n)^T h(\theta^n) \\ &\quad + \epsilon_n \nabla V(\theta^n)^T \delta M^n + \mathcal{O}(\epsilon_n^2 \|H^n\|_2^2) \end{aligned}$$

where $\delta M^n = H^n - h(\theta^n)$ is a martingale difference sequence. From part (a), we are guaranteed that $\sup_n \mathbb{E} \|H^n\|_2^2 < K < \infty$, so that

$$\begin{aligned} \mathbb{P} \left[\sum_{k=1}^{\infty} \mathcal{O}(\epsilon_k^2) \|H^k\|_2^2 \geq \mu \right] &\leq \sum_{k=1}^{\infty} \mathcal{O} \left(\frac{\epsilon_k^2}{\mu} \right) \\ &\leq \frac{\beta}{\mu} \rightarrow 0 \quad \text{as } \mu \rightarrow \infty \end{aligned}$$

so it is finite with probability 1. Moreover, (38a) (namely, that $\nabla V(\theta^k)^T h(\theta^k) < -\delta$) ensures the second term cannot drive θ^n from \mathcal{V} , but also pulls it back in. The only remaining term to consider is the martingale difference term δM^n . If we define $M^n := \sum_{k=1}^n \epsilon_k \delta M^k$, then using the martingale inequality and the condition $\sup_n \mathbb{E} \|H^n\|_2^2 < K$, we have for each $\mu > 0$

$$\lim_m \mathbb{P} \left[\sup_{k \geq m} |M^j - M^m| \geq \mu \right] = 0$$

so M^n converges a.s. to some finite value. (See [23, pp. 97, 128] for details.) Consequently, using the facts $\nabla V(\theta^k)^T h(\theta^k) < -\delta$ and $\sum_{k=1}^{\infty} \epsilon_k = \infty$, we conclude that the term $\sum_{k=1}^n \epsilon_k h(\theta^k)$ can thus compensate the effect of M^n for any finite amount. Therefore, the term M^n cannot drive θ^n outside of \mathcal{V} infinitely often, and we have established the claim.

ACKNOWLEDGMENT

The authors thank P. Varaiya and A. Willsky for helpful comments, and the reviewers for pointing out references and helping to improve the original manuscript. They thank N. Noorshams and one reviewer for pointing out a technical error in our original version of Theorem 1.

REFERENCES

- [1] N. Alon, "Eigenvalues and expanders," *Combinatorica*, vol. 6, no. 2, pp. 83–96, 1986.
- [2] N. Alon and J. Spencer, *The Probabilistic Method*. New York: Wiley Intersci., 2000.
- [3] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Distributed average consensus with dithered quantization," *IEEE Trans. Signal Process.*, vol. 56, pp. 4905–4918, Oct. 2008.
- [4] T. C. Aysal, M. E. Yildiz, and A. Scaglione, "Broadcast gossip algorithms," in *Proc. IEEE Inf. Theory Workshop*, Porto, Portugal, 2008.
- [5] F. Benezit, A. G. Dimakis, P. Thiran, and M. Vetterli, "Gossip along the way: Order-optimal consensus through randomized path averaging," in *Proc. 45th Allerton Conf.*, 2007.
- [6] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. New York: Springer-Verlag, 1990.
- [7] V. Borkar and P. Varaiya, "Asymptotic agreement in distributed estimation," *IEEE Trans. Autom. Control*, vol. 27, no. 3, pp. 650–655, 1982.
- [8] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2508–2530, 2006.
- [9] R. Carli, F. Fagnani, P. Frasca, T. Taylor, and S. Zampieri, "Average consensus on networks with transmission noise or quantization," in *Proc. Eur. Contr. Conf.*, 2007.
- [10] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI: Amer. Math. Soc., 1991.
- [11] M. H. deGroot, "Reaching a consensus," *J. Amer. Statist. Assoc.*, vol. 69, no. 345, pp. 118–121, Mar. 1974.
- [12] P. Denantes, F. Benezit, and P. T. Vetterli, "Which distributed averaging algorithm should i choose for my sensor network?," in *Proc. 27th IEEE Conf. Comput. Commun. (INFOCOM 2008)*, 2008, pp. 986–994.
- [13] A. G. Dimakis, A. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE Trans. Signal Process.*, vol. 53, pp. 1205–1216, Mar. 2008.
- [14] F. Fagnani and S. Zampieri, "Average consensus with packet drop communication," *SIAM J. Contr. Optimiz.*, 2007.
- [15] J. Feldman, T. Malkin, R. A. Servedio, C. Stein, and M. J. Wainwright, "LP decoding corrects a constant fraction of errors," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 82–89, Jan. 2007.
- [16] Y. Hatano, A. K. Das, and M. Mesbahi, "Agreement in presence of noise: Pseudogradients on random geometric networks," in *Proc. 44th IEEE Conf. Decision and Contr., and Eur. Contr. Conf.*, Dec. 2005.
- [17] S. Kar and J. M. F. Moura, "Distributed average consensus in sensor networks with quantized inter-sensor communication," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2008)*, 2008, pp. 2281–2284.
- [18] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, pp. 355–369, Jan. 2009.
- [19] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Trans. Signal Process.*, vol. 58, pp. 1383–1400, 2010.
- [20] S. Kar and J. M. F. Moura, "A mixed time-scale algorithm for distributed parameter estimation: Nonlinear observation models and imperfect communication," *ICASSP*, 2009.
- [21] A. Kashyap, T. Basar, and R. Srikant, "Quantized consensus," *Automatica*, vol. 43, pp. 1192–1203, 2007.
- [22] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," *Proc. 44th Ann. IEEE FOCS*, pp. 482–491, 2003.
- [23] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*, 2nd ed. New York: Springer-Verlag, 2003.
- [24] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Autom. Control*, vol. 22, pp. 551–575, 1977.
- [25] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, pp. 215–233, 2007.
- [26] M. Penrose, *Random Geometric Graphs. Oxford Studies in Probability*. Oxford, U.K.: Oxford Univ. Press, 2003.

- [27] G. Picci and T. Taylor, "Almost sure convergence of random gossip algorithms," in *Proc. 46th IEEE Conf. Decision and Contr.*, 2007, pp. 282–287.
- [28] 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, pp. 350–364, 2008.
- [29] J. Tsitsiklis, "Problems in decentralized decision-making and computation," Ph.D. dissertation, Dep. EECS, Mass. Inst. Technol. (MIT), Cambridge, 1984.
- [30] L. X. Boyd, "Fast linear iterations for distributed averaging," *Syst. Contr. Lett.*, vol. 52, pp. 65–78, 2004.
- [31] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *J. Parallel Distrib. Comput.*, vol. 67, no. 1, pp. 33–46, 2007.
- [32] M. E. Yildiz and A. Scaglione, "Coding with side information for rate constrained consensus," *IEEE Trans. Signal Process.*, 2008.



Ram Rajagopal received the B.S. degree in electrical engineering from the Federal University of Rio de Janeiro, Brazil, in 1998 and the M.S. degree in electrical and computer engineering from the University of Texas, Austin. He received the Ph.D. degree in electrical engineering and computer sciences and the M.A. degree in statistics from the University of California, Berkeley, both in 2009.

He is currently a postdoctoral researcher with the Electrical Engineering and Computer Sciences Department, University of California, Berkeley. His research interests include sequential statistics, stochastic signal processing, and applications of sensor networks and statistics in the design and operations of transportation and energy networks.



Martin J. Wainwright (M'03–SM'10) received the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge.

He is currently an Associate Professor with the University of California, Berkeley, with a joint appointment between the Department of Statistics and the Department of Electrical Engineering and Computer Sciences (EECS). His research interests include statistical signal processing, coding and information theory, statistical machine learning, and

high-dimensional statistics.

Prof. Wainwright has been awarded an Alfred P. Sloan Foundation Fellowship, an NSF CAREER Award, the George M. Sprowls Prize for his dissertation research (EECS Department, MIT), a Natural Sciences and Engineering Research Council of Canada 1967 Fellowship, an IEEE Signal Processing Society Best Paper Award in 2008, and several outstanding conference paper awards.