Distributed Consensus Algorithms in Sensor Networks With Imperfect Communication: Link Failures and Channel Noise

Soummya Kar and José M. F. Moura, *Fellow, IEEE*

Abstract—The paper studies average consensus with random topologies (intermittent links) and noisy channels. Consensus with noise in the network links leads to the bias-variance dilemma—running consensus for long reduces the bias of the final average estimate but increases its variance. We present two different compromises to this tradeoff: the $\mathcal{A} - \mathcal{ND}$ algorithm modifies conventional consensus by forcing the weights to satisfy a *persistence* condition (slowly decaying to zero;) and the $\mathcal{A} - \mathcal{NC}$ algorithm where the weights are constant but consensus is run for a fixed number of iterations $\hat{\imath}$, then it is restarted and rerun for a total of $\hat{p}$ runs, and at the end averages the final states of the $\hat{p}$ runs (Monte Carlo averaging). We use controlled Markov processes and stochastic approximation arguments to prove almost sure convergence of $\mathcal{A} - \mathcal{ND}$ to a finite consensus limit and compute explicitly the mean square error (mse) (variance) of the consensus limit. We show that $\mathcal{A} - \mathcal{ND}$ represents the best of both worlds—zero bias and low variance—at the cost of a slow convergence rate; rescaling the weights balances the variance versus the rate of bias reduction (convergence rate). In contrast, $\mathcal{A} - \mathcal{NC}$, because of its constant weights, converges fast but presents a different bias-variance tradeoff. For the same number of iterations $\hat{\imath}\hat{p}$, shorter runs (smaller $\hat{\imath}$) lead to high bias but smaller variance (larger number $\hat{p}$ of runs to average over.) For a static nonrandom network with Gaussian noise, we compute the optimal gain for $\mathcal{A} - \mathcal{NC}$ to reach in the shortest number of iterations $\hat{\imath}\hat{p}$, with high probability $(1 - \delta)$, (ϵ, δ)-consensus (ϵ residual bias). Our results hold under fairly general assumptions on the random link failures and communication noise.

Index Terms—Additive noise, consensus, sensor networks, stochastic approximation, random topology.

I. INTRODUCTION

DISTRIBUTED computation in sensor networks is a well-studied field with an extensive body of literature (see, for example, [1] for early work.) Average consensus computes iteratively the global average of distributed data using local communications, see [2]–[5] that consider versions and extensions of basic consensus. A review of the consensus literature is in [6].

Reference [7] designs the optimal link weights that optimize the convergence rate of the consensus algorithm when the connectivity graph of the network is fixed (not random). Our previous work, [8]–[11], extends [7] by designing the topology, i.e., both the weights and the connectivity graph, under a variety of conditions, including random links and link communication costs, under a network communication budget constraint.

We consider distributed average consensus when *simultaneously* the network topology is random (link failures, like when packets are lost in data networks) *and* the communications among sensors is commonly noisy. A typical example is time division multiplexing, where in a particular user's time slot the channel may not be available, and, if available, we assume the communication is analog and noisy. Our approach can handle spatially correlated link failures and certain types of temporally Markovian sequences of Laplacians and Markovian noise, which go beyond independently, identically distributed (i.i.d.) Laplacian matrices and i.i.d. communication noise sequences. Noisy consensus leads to a tradeoff between bias and variance. Running consensus longer reduces bias, i.e., the mean of the error between the desired average and the consensus reached. But, due to noise, the variance of the limiting consensus grows with longer runs. To address this dilemma, we consider two versions of consensus with link failures *and* noise that represent two different bias-variance tradeoffs: the $\mathcal{A} - \mathcal{ND}$ and the $\mathcal{A} - \mathcal{NC}$ algorithms.

$\mathcal{A} - \mathcal{ND}$ updates each sensor state with a weighted fusion of its current neighbors' states (received distorted by noise). The fusion weights $\alpha(i)$ satisfy a *persistence* condition, decreasing to zero, but not too fast. $\mathcal{A} - \mathcal{ND}$ falls under the purview of controlled Markov processes, and we use stochastic approximation techniques to prove its almost sure (a.s.) consensus when the network is connected on the average: the sensor state vector sequence converges a.s. to the *consensus subspace*. A simple condition on the mean Laplacian, $\overline{L} = \mathbb{E}\{L\}$, for connectedness is on its second eigenvalue $\lambda_2(\overline{L}) > 0$. We establish that the sensor states converge asymptotically a.s. to a finite random variable θ and, in particular, the expected sensor states converge to the desired average r (asymptotic unbiasedness.) We determine the variance of θ, which is the mean square error (mse) between θ and the desired average. By properly tuning the weights sequence $\{\alpha(i)\}$, the variance of θ can be made arbitrarily small, though at a cost of slowing $\mathcal{A} - \mathcal{ND}$s convergence rate, i.e., the rate at which the bias goes to zero.

$\mathcal{A} - \mathcal{NC}$ is a repeated averaging algorithm that performs in-network Monte Carlo simulations: it runs consensus $\hat{p}$ times

with constant weight α, for a fixed number of iterations $\hat{\imath}$ each time, and then each sensor averages its $\hat{p}$ values of the state at the final iteration $\hat{\imath}$ of each run. $\mathcal{A} - \mathcal{N}\mathcal{C}$s constant weight α speeds its convergence rate relative to $\mathcal{A} - \mathcal{N}\mathcal{D}$s, whose weights $\alpha(i)$ decrease to zero. We determine the number of iterations $\widehat{\imath p}$ required to reach (ϵ, δ)-*consensus*, i.e., for the bias of the consensus limit at each sensor to be smaller than $K\epsilon$, with high probability $(1 - \delta)$. For nonrandom networks, we establish a tight upper bound on the minimizing $\widehat{\imath p}$ and compute the corresponding optimal constant weight α. We quantify the tradeoff between the number of iterations $\hat{\imath}$ per Monte Carlo run and the number of runs $\hat{p}$.

Finally, we compare the bias-variance tradeoffs between the two algorithms and the network parameters that determine their convergence rate and noise resilience. The fixed weight $\mathcal{A} - \mathcal{N}\mathcal{C}$ algorithm can converge faster but requires greater intersensor coordination than the $\mathcal{A} - \mathcal{N}\mathcal{D}$ algorithm.

A. Comparison With Existing Literature

Random link failures and additive channel noise have been considered *separately*. Random link failures, but noise*less* consensus, is in [11]–[16]. References [11]–[13] assume an erasure model: the network links fail independently in space (independently of each other) and in time (link failure events are temporally independent.) Papers [14] and [16] study directed topologies with only time i.i.d. link failures, but impose distributional assumptions on the link formation process. In [15], the link failures are i.i.d. Laplacian matrices, the graph is directed, and no distributional assumptions are made on the Laplacian matrices. The paper presents necessary and sufficient conditions for consensus using the ergodicity of products of stochastic matrices.

Similarly, [17]–[19] consider consensus with additive noise, but *fixed* or static, *nonrandom* topologies (no link failures.) They use a decreasing weight sequence to guarantee consensus. These references do not characterize the mse. For example, [18] and [19] rely on the existence of a unique solution to an algebraic Lyapunov equation. The more general problem of distributed estimation (of which average consensus is a special case) in the presence of additive noise is in [20], again with a fixed topology. Both [17] and [20] assume a temporally white noise sequence, while our approach can accommodate a more general Markovian noise sequence, in addition to white noise processes.

In summary, with respect to [11]–[20], our approach considers: i) random topologies *and* noisy communication links *simultaneously*; ii) spatially correlated (Markovian) dependent random link failures; iii) time Markovian noise sequences; iv) undirected topologies; v) no distributional assumptions; vi) consensus (estimation being considered elsewhere); and vii) two versions of consensus representing different compromises of bias versus variance.

Briefly, the paper is as follows. Sections II and III summarize relevant spectral graph and average consensus results. Sections IV and V treat the additive noise *with* random link failure communication analyzing the $\mathcal{A} - \mathcal{N}\mathcal{D}$ and $\mathcal{A} - \mathcal{N}\mathcal{C}$ algorithms, respectively. Finally, Section VI concludes the paper.

II. ELEMENTARY SPECTRAL GRAPH THEORY

We summarize briefly facts from spectral graph theory. For an undirected graph $G = (V, E)$, $V = [1 \cdots N]$ is the set of nodes or vertices $|V| = N$, and E is the set of edges $|E| = M$. The unordered pair $(n, l) \in E$ if there exists an edge between nodes n and l. We only consider simple graphs, i.e., graphs devoid of self-loops and multiple edges. A path between nodes n and l of length m is a sequence $(n = i_0, i_1, \ldots, i_m = l)$ of vertices, such that, $(i_k, i_{k+1}) \in E$, $0 \leq k \leq m - 1$. A graph is connected if there exists a path, between each pair of nodes. The neighborhood of node n is

$$\Omega_n = \{l \in V \mid (n, l) \in E\}. \tag{1}$$

Node n has degree $d_n = |\Omega_n|$ (number of edges with n as one end point.) The structure of the graph can be described by the symmetric $N \times N$ adjacency matrix, $A = [A_{nl}]$, $A_{nl} = 1$, if $(n, l) \in E$, 0 otherwise. Let the degree matrix be the diagonal matrix $D = \text{diag}(d_1 \cdots d_N)$. The graph Laplacian matrix L is

$$L = D - A. \tag{2}$$

The Laplacian is a positive semidefinite matrix; hence, its eigenvalues can be ordered as

$$0 = \lambda_1(L) \leq \lambda_2(L) \leq \cdots \leq \lambda_N(L). \tag{3}$$

The multiplicity of the zero eigenvalue equals the number of connected components of the network; for a connected graph $\lambda_2(L) > 0$. This second eigenvalue is the algebraic connectivity or the Fiedler value of the network; see [21]–[23] for detailed treatment of graphs and their spectral theory.

III. DISTRIBUTED AVERAGE CONSENSUS WITH IMPERFECT COMMUNICATION

In a simple form, distributed average consensus computes the average r of the initial node data

$$r = \frac{1}{N} \sum_{n=1}^{N} x_n(0) \tag{4}$$

by local data exchanges among neighbors. For noiseless and unquantized data exchanges across the network links, the state of each node is updated iteratively by

$$x_n(i+1) = w_{nn}(i) x_n(i) + \sum_{l \in \Omega_n(i)} w_{nl}(i) x_l(i), \quad 1 \leq n \leq N \tag{5}$$

where the link weights, w_{nl}s, may be constant or time varying. Similarly, the topology of a time-varying network is captured by making the neighborhoods Ω_ns, to be a function of time. Because noise causes consensus to diverge, [10], [24], we let

the link weights to be the same across different network links, but vary with time. Equation (5) becomes

$$x_n(i+1) = [1 - \alpha(i)d_n(i)] x_n(i) + \alpha(i) \sum_{l \in \Omega_n(i)} x_l(i), \quad 1 \leq n \leq N. \quad (6)$$

We address consensus with imperfect intersensor communication, where each sensor receives noise corrupted versions of its neighbors' states. We modify the state update (6) to

$$x_n(i+1) = [1 - \alpha(i)d_n(i)] x_n(i) + \alpha(i) \sum_{l \in \Omega_n(i)} f_{nl,i}[x_l(i)], \quad 1 \leq n \leq N \quad (7)$$

where $\{f_{nl,i}\}_{1 \leq n, l \leq N, i \geq 0}$ is a sequence of functions (possibly random) modeling the channel imperfections. In the following sections, we analyze the consensus problem given by (7), when the channel communication is corrupted by additive noise. In [25], we consider the effects of quantization (see also [26] for a treatment of consensus algorithms with quantized communication.) Here, we study two different algorithms. The first, $\mathcal{A} - \mathcal{ND}$, considers a decreasing weight sequence ($\alpha(i) \rightarrow 0$) and is analyzed in Section IV. The second, $\mathcal{A} - \mathcal{NC}$, uses repeated averaging with a constant link weight and is detailed in Section V.

IV. $\mathcal{A} - \mathcal{ND}$: CONSENSUS IN ADDITIVE NOISE AND RANDOM LINK FAILURES

We consider distributed consensus when the network links fail or become alive at random times, and data exchanges are corrupted by additive noise. The network topology varies randomly across iterations. We analyze the convergence properties of the $\mathcal{A} - \mathcal{ND}$ algorithm under this generic scenario. We start by formalizing the assumptions underlying $\mathcal{A} - \mathcal{ND}$ in the next Subsection.

A. Problem Formulation and Assumptions

We compute the average of the initial state $\mathbf{x}(0) = [x_1(0) \cdots x_N(0)]^T \in \mathbb{R}^{N \times 1}$ with the distributed consensus algorithm with communication channel imperfections given in (7). Let $\{v_{nl}(i)\}_{1 \leq n, l \leq N, i \geq 0}$ be a sequence of independent zero mean random variables. For additive noise,

$$f_{nl,i}(y) = y + v_{nl}(i). \quad (8)$$

Recall the Laplacian L defined in (2). Collecting the states $x_n(i)$ in the vector $\mathbf{x}(i)$, (7) is

$$\mathbf{x}(i+1) = \mathbf{x}(i) - \alpha(i)[L(i)\mathbf{x}(i) + \mathbf{n}(i)] \quad (9)$$
$$[\mathbf{n}(i)]_l = n_l(i)$$
$$= - \sum_{k \in \Omega_l(i)} v_{lk}(i), \ 1 \leq l \leq N, \ i \geq 0. \quad (10)$$

We now state the assumptions of the $\mathcal{A} - \mathcal{ND}$ algorithm.[1]
1) **Random Network Failure**: We propose two models; the second is more general than the first.

[1]See also [27], where parts of the results are presented.

1.1) Temporally i.i.d. Laplacian Matrices: The graph Laplacians are

$$L(i) = \overline{L} + \widetilde{L}(i), \ \forall i \geq 0 \quad (11)$$

where $\{L(i)\}_{i \geq 0}$ is a sequence of i.i.d. Laplacian matrices with mean $\overline{L} = \mathbb{E}[L(i)]$, such that $\lambda_2(\overline{L}) > 0$. We do not make any distributional assumptions on the link failure model, and, in fact, as long as the sequence $\{L(i)\}_{i \geq 0}$ is independent with constant mean $\overline{L}$, satisfying $\lambda_2(\overline{L}) > 0$, the i.i.d. assumption can be dropped. During the same iteration, the link failures can be spatially dependent, i.e., correlated across different edges of the network. This model subsumes the erasure network model, where the link failures are independent both over space and time. Wireless sensor networks motivate this model since interference among the sensors communication correlates the link failures over space, while over time, it is still reasonable to assume that the channels are memoryless or independent.

Connectedness of the graph is an important issue. We do not require that the random instantiations $G(i)$ of the graph be connected; in fact, it is possible to have all these instantiations to be disconnected. We only require that the graph stays connected on *average*. This is captured by requiring that $\lambda_2(\overline{L}) > 0$, enabling us to capture a broad class of asynchronous communication models; for example, the random asynchronous gossip protocol analyzed in [28] satisfies $\lambda_2(\overline{L}) > 0$ and hence falls under this framework.

1.2) Temporally Markovian Laplacian Matrices: Our results hold when the Laplacian matrix sequence $\{L(i, \mathbf{x}(i))\}_{i \geq 0}$ is state-dependent. More precisely, we assume that there exists a two-parameter random field, $\{L(i, \mathbf{x})\}_{i \geq 0, \mathbf{x} \in \mathbb{R}^{N \times 1}}$ of Laplacian matrices such that

$$\mathbb{E}[L(i, \mathbf{x})] = \overline{L}, \ \forall i, \mathbf{x} \quad (12)$$

and $\lambda_2(\overline{L}) > 0$. We also require that, for a fixed i, the random matrices, $\{L(i, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^{N \times 1}}$, are independent of the sigma algebra, $\sigma(\mathbf{x}(j), 0 \leq j \leq i)$.[2] It is clear then that the Laplacian matrix sequence, $\{L(i, \mathbf{x}(i))\}_{i \geq 0}$, is Markov. We will show that our convergence analysis holds also for this general link failure model. Such a model may be appropriate in stochastic formation control scenarios, see [29]–[31], where the network topology is state-dependent.

2) **Communication Noise Model**: We propose two models; the second is more general than the first.

2.1) Independent Noise Sequence: The additive noise $\{v_{nl}(i)\}_{1 \leq n, l \leq N, i \geq 0}$ is an independent sequence

$$\mathbb{E}[v_{nl}(i)] = 0, \ \forall 1 \leq n, l \leq N, \ i \geq 0,$$
$$\sup_{n,l,i} \mathbb{E}[v_{nl}^2(i)] = \mu < \infty. \quad (13)$$

[2]This guarantees that the Laplacian $L(i, \mathbf{x}(i))$ may depend on the past state history $\{\mathbf{x}(j), j \leq i\}$, only through the present state $\mathbf{x}(i)$.

The sequences, $\{L(i)\}_{i \geq 0}$ and $\{v_{nl}(i)\}_{1 \leq n, l \leq N, i \geq 0}$ are mutually independent. Hence, $L(i)$, $\{v_{nl}(i)\}$, $1 \leq n, l \leq N$, $i \geq 0$ are independent of $\sigma(\mathbf{x}(j), 0 \leq j \leq i)$, $\forall i$. Then, from (10)

$$\mathbb{E}[\mathbf{n}(i)] = \mathbf{0}, \ \forall i$$
$$\sup_i \mathbb{E}\left[\|\mathbf{n}(i)\|^2\right] = \eta \leq N(N-1)\mu < \infty. \tag{14}$$

No distributional assumptions are required on the noise sequence.

2.2) Markovian Noise Sequence: Our approach allows the noise sequence to be Markovian through state-dependence. Let the two-parameter random field, $\{\mathbf{n}(i, \mathbf{x})\}_{i \geq 0, \mathbf{x} \in \mathbb{R}^{N \times 1}}$ of random vectors

$$\mathbb{E}[\mathbf{n}(i, \mathbf{x})] = 0, \ \forall i, \mathbf{x}. \tag{15}$$

For fixed i, the random vectors $\{\mathbf{n}(i, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^{N \times 1}}$ are independent of the σ-algebra, $\sigma(\mathbf{x}(j), 0 \leq j \leq i)$ and the random families $\{L(i, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^{N \times 1}}$ and $\{\mathbf{n}(i, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^{N \times 1}}$ are independent. It is clear then that the noise vector sequence $\{\mathbf{n}(i, \mathbf{x}(i))\}_{i \geq 0}$ is Markov. Note, however, in this case the resulting Laplacian and noise sequences $\{L(i, \mathbf{x}(i))\}_{i \geq 0}$ and $\{\mathbf{n}(i, \mathbf{x}(i))\}_{i \geq 0}$ are no longer independent; they are coupled through the state $\mathbf{x}(i)$. In addition to (15), we require the variance of the noise component orthogonal to the consensus subspace [see (31)] to satisfy, for constants $c_1, c_2 \geq 0$

$$\mathbb{E}[\|\mathbf{n}_{\mathcal{C}^\perp}(i, \mathbf{x})\|^2] \leq c_1 + c_2 \|\mathbf{x}_{\mathcal{C}^\perp}\|^2. \tag{16}$$

We do not restrict the variance growth rate of the noise component in the consensus subspace. This clearly subsumes the bounded noise variance model. An example of such noise is

$$\mathbf{n}(i, \mathbf{x}(i)) = \vartheta(i)\left(\mathbf{x}(i) + \mathbf{w}(i)\right) \tag{17}$$

where $\{\vartheta(i)\}_{i \geq 0}$ and $\{\mathbf{w}(i)\}_{i \geq 0}$ are zero mean finite variance mutually i.i.d. sequences of scalars and vectors, respectively. It is then clear that the condition in (16) is satisfied, and the noise model **2.2)** applies. The model in (17) arises, for example, in multipath effects in MIMO systems, when the channel adds multiplicative noise whose amplitude is proportional to the transmitted data.

3) **Persistence Condition**: The weights decay to zero, but not too fast

$$\alpha(i) > 0, \ \sum_{i \geq 0} \alpha(i) = \infty, \ \sum_{i \geq 0} \alpha^2(i) < \infty. \tag{18}$$

This condition is commonly assumed in adaptive control and signal processing. Examples include

$$\alpha(i) = \frac{1}{i^\beta}, \ .5 < \beta \leq 1. \tag{19}$$

For clarity, in the main body of the paper, we prove the results for the $\mathcal{A} - \mathcal{ND}$ algorithm under Assumptions **1.1)**, **2.1)**, and **3)**. In the Appendix, we point out how to modify the proofs when the more general assumptions **1.2)** and **2.2)** hold.

We now prove the almost sure (a.s.) convergence of the $\mathcal{A} - \mathcal{ND}$ algorithm in (9) by using results from the theory of stochastic approximation algorithms [32].

B. A Result on Convergence of Markov Processes

A systematic and thorough treatment of stochastic approximation procedures has been given in [32]. In this section, we modify slightly a result from [32] and restate it as a theorem in a form relevant to our application. We follow the notation of [32], which we now introduce.

Let $\mathbf{X} = \{\mathbf{x}(i)\}_{i \geq 0}$ be a Markov process on $\mathbb{R}^{N \times 1}$. The generating operator $\mathcal{L}$ of $\mathbf{X}$ is

$$\mathcal{L}V(i, \mathbf{x}) = \mathbb{E}\left[V(i+1, \mathbf{x}(i+1)) \mid \mathbf{x}(i) = \mathbf{x}\right] - V(i, \mathbf{x}) \tag{20}$$

for functions $V(i, \mathbf{x})$, $i \geq 0$, $\mathbf{x} \in \mathbb{R}^{N \times 1}$, provided the conditional expectation exists. We say that $V(i, \mathbf{x}) \in D_{\mathcal{L}}$ in a domain A, if $\mathcal{L}V(i, \mathbf{x})$ is finite for all $(i, \mathbf{x}) \in A$.

Denote the Euclidean metric by $\rho(\cdot)$. For $B \subset \mathbb{R}^{N \times 1}$, the ϵ-neighborhood of B and its complement is

$$U_\epsilon(B) = \left\{ x \mid \inf_{y \in B} \rho(x, y) < \epsilon \right\} \tag{21}$$

$$V_\epsilon(B) = \mathbb{R}^{N \times 1} \backslash U_\epsilon(B). \tag{22}$$

We now state the desired theorem, whose proof we sketch in the Appendix.

Theorem 1: Let $\mathbf{X}$ be a Markov process with generating operator $\mathcal{L}$. Let there exist a nonnegative function $V(i, \mathbf{x}) \in D_{\mathcal{L}}$ in the domain $i \geq 0$, $\mathbf{x} \in \mathbb{R}^{N \times 1}$, and $B \subset \mathbb{R}^{N \times 1}$ with the following properties:

1)

$$\inf_{i \geq 0, \mathbf{x} \in V_\epsilon(B)} V(i, \mathbf{x}) > 0, \ \forall \epsilon > 0 \tag{23}$$

$$V(i, \mathbf{x}) \equiv 0, \ \mathbf{x} \in B \tag{24}$$

$$\lim_{\mathbf{x} \to B} \sup_{i \geq 0} V(i, \mathbf{x}) = 0 \tag{25}$$

2)

$$\mathcal{L}V(i, \mathbf{x}) \leq g(i)(1 + V(i, \mathbf{x})) - \alpha(i)\varphi(i, \mathbf{x}) \tag{26}$$

where $\varphi(i, \mathbf{x}), i \geq 0$, $\mathbf{x} \in \mathbb{R}^{N \times 1}$ is a nonnegative function such that

$$\inf_{i, \mathbf{x} \in V_\epsilon(B)} \varphi(i, \mathbf{x}) > 0, \ \forall \epsilon > 0 \tag{27}$$

3)

$$\alpha(i) > 0, \ \sum_{i \geq 0} \alpha(i) = \infty \tag{28}$$

$$g(i) > 0, \ \sum_{i \geq 0} g(i) < \infty. \tag{29}$$

Then, the Markov process $\mathbf{X} = \{\mathbf{x}(i)\}_{i \geq 0}$ with arbitrary initial distribution converges a.s. to B as $i \to \infty$. In other words,

$$\mathbb{P}\left(\lim_{i \to \infty} \rho(\mathbf{x}(i), B) = 0\right) = 1 \tag{30}$$

Proof of Convergence of the $\mathcal{A} - \mathcal{N}\mathcal{D}$ Algorithm

The $\mathcal{A} - \mathcal{N}\mathcal{D}$ distributed consensus algorithm is given by (9) in Section IV-A. To establish its a.s. convergence using Theorem 1, define the *consensus* subspace, $\mathcal{C}$, aligned with $\mathbf{1}$, the vector of 1s,

$$\mathcal{C} = \left\{\mathbf{x} \in \mathbb{R}^{N \times 1} : \mid \mathbf{x} = a\mathbf{1}, a \in \mathbb{R}\right\}. \tag{31}$$

We recall a result on distance properties in $\mathbb{R}^{N \times 1}$ to be used in the sequel. We omit the proof.

Lemma 2: Let $\mathcal{S}$ be a subspace of $\mathbb{R}^{N \times 1}$. For $\mathbf{x} \in \mathbb{R}^{N \times 1}$, consider the orthogonal decomposition $\mathbf{x} = \mathbf{x}_{\mathcal{S}} + \mathbf{x}_{\mathcal{S}^\perp}$. Then $\rho(\mathbf{x}, \mathcal{S}) = \|\mathbf{x}_{\mathcal{S}^\perp}\|$.

Theorem 3: ($\mathcal{A} - \mathcal{N}\mathcal{D}$ *a.s. convergence*) Let assumptions **1.1)**, **2.1)**, and **3)** hold. Consider the $\mathcal{A} - \mathcal{N}\mathcal{D}$ consensus algorithm in (9) in Section IV-A with initial state $\mathbf{x}(0) \in \mathbb{R}^{N \times 1}$. Then,

$$\mathbb{P}\left[\lim_{i \to \infty} \rho(\mathbf{x}(i), \mathcal{C}) = 0\right] = 1 \tag{32}$$

Proof: Under the assumptions, the process $\mathbf{X} = \{\mathbf{x}(i)\}_{i \geq 0}$ is Markov. Define

$$V(i, \mathbf{x}) = \mathbf{x}^T \overline{L} \mathbf{x}. \tag{33}$$

The potential function $V(i, \mathbf{x})$ is nonnegative. Since $\mathbf{x} \in \mathcal{C}$ is an eigenvector of $\overline{L}$ with zero eigenvalue

$$V(i, \mathbf{x}) \equiv 0, \; \mathbf{x} \in \mathcal{C}, \quad \lim_{\mathbf{x} \to \mathcal{C}} \sup_{i \geq 0} V(i, \mathbf{x}) = 0. \tag{34}$$

The second condition follows from the continuity of $V(i, \mathbf{x})$. By Lemma 2 and the definition in (22) of the complement of the ϵ-neighborhood of a set

$$\mathbf{x} \in V_\epsilon(\mathcal{C}) \implies \|\mathbf{x}_{\mathcal{C}^\perp}\| \geq \epsilon. \tag{35}$$

Hence, for $\mathbf{x} \in V_\epsilon(\mathcal{C})$,

$$\begin{aligned} V(i, \mathbf{x}) &= \mathbf{x}^T \overline{L} \mathbf{x} \\ &\geq \lambda_2\left(\overline{L}\right) \|\mathbf{x}_{\mathcal{C}^\perp}\|^2 \\ &\geq \lambda_2\left(\overline{L}\right) \epsilon^2. \end{aligned} \tag{36}$$

Then, since by assumption **1.1)** $\lambda_2(\overline{L}) > 0$ (note that the assumption $\lambda_2(\overline{L}) > 0$ comes into play here), we get

$$\inf_{i \geq 0, \mathbf{x} \in V_\epsilon(\mathcal{C})} V(i, \mathbf{x}) \geq \lambda_2(\overline{L}) \epsilon^2 > 0. \tag{37}$$

Now consider $\mathcal{L}V(i, \mathbf{x})$ in (20). Using (9) in (20), we obtain

$$\begin{aligned} \mathcal{L}V(i, \mathbf{x}) &= \mathbb{E}\left[\mathbf{x}(i+1)^T \overline{L}\mathbf{x}(i+1) \mid \mathbf{x}(i) = \mathbf{x}\right] - \mathbf{x}^T \overline{L}\mathbf{x} \\ &= \mathbb{E}\Big[\big[\mathbf{x} - \alpha(i)\overline{L}\mathbf{x} - \alpha(i)\widetilde{L}(i)\mathbf{x} - \alpha(i)\mathbf{n}(i)\big]^T \\ &\quad \cdot \overline{L}\big[\mathbf{x} - \alpha(i)\overline{L}\mathbf{x} - \alpha(i)\widetilde{L}(i)\mathbf{x} - \alpha(i)\mathbf{n}(i)\big]\Big] \\ &\quad - \mathbf{x}^T \overline{L}\mathbf{x}. \end{aligned} \tag{38}$$

Using the independence of $\widetilde{L}(i)$ and $\mathbf{n}(i)$ with respect to $\mathbf{x}(i)$, that $\widetilde{L}(i)$ and $\mathbf{n}(i)$ are zero-mean, and that the subspace $\mathcal{C}$ lies in the null space of $\overline{L}^3$ and $\widetilde{L}(i)$ (the latter because this is true for both $L(i)$ and $\overline{L}$), (38) leads successively to

$$\begin{aligned} \mathcal{L}V(i, \mathbf{x}) &= -2\alpha(i)\mathbf{x}^T \overline{L}^2 \mathbf{x} + \alpha^2(i)\mathbf{x}^T \overline{L}^3 \mathbf{x} \\ &\quad + \mathbb{E}\left[\alpha^2(i)\left(\widetilde{L}(i)\mathbf{x}\right)^T \overline{L}\left(\widetilde{L}(i)\mathbf{x}\right)\right] \\ &\quad + \mathbb{E}\left[\alpha^2(i)\mathbf{n}(i)^T \overline{L}\mathbf{n}(i)\right] \\ &\leq -2\alpha(i)\mathbf{x}^T \overline{L}^2 \mathbf{x} + \alpha^2(i)\lambda_N(\overline{L})^3 \|\mathbf{x}_{\mathcal{C}^\perp}\|^2 \\ &\quad + \alpha^2(i)\lambda_N(\overline{L})\mathbb{E}\left[\|\widetilde{L}(i)\mathbf{x}\|^2\right] \\ &\quad + \alpha^2(i)\lambda_N(\overline{L})\mathbb{E}\left[\|\mathbf{n}(i)\|^2\right] \\ &\leq -2\alpha(i)\mathbf{x}^T \overline{L}^2 \mathbf{x} + \alpha^2(i)\lambda_N(\overline{L})^3 \|\mathbf{x}_{\mathcal{C}^\perp}\|^2 \\ &\quad + \alpha^2(i)\lambda_N(\overline{L})\mathbb{E}\left[\lambda_{\max}^2\left(\widetilde{L}(i)\right)\right]\|\mathbf{x}_{\mathcal{C}^\perp}\|^2 \\ &\quad + \alpha^2(i)\lambda_N(\overline{L})\eta \\ &\leq -2\alpha(i)\mathbf{x}^T \overline{L}^2 \mathbf{x} + \alpha^2(i)\lambda_N(\overline{L})^3 \|\mathbf{x}_{\mathcal{C}^\perp}\|^2 \\ &\quad + 4\alpha^2(i)N^2 \lambda_N(\overline{L})\|\mathbf{x}_{\mathcal{C}^\perp}\|^2 \\ &\quad + \alpha^2(i)\lambda_N(\overline{L})\eta. \end{aligned} \tag{39}$$

The last step follows because all the eigenvalues of $\widetilde{L}(i)$ are less than $2N$ in absolute value, by the Gershgorin circle theorem. Now, by the fact $\mathbf{x}^T \overline{L}\mathbf{x} \geq \lambda_2(\overline{L})\|\mathbf{x}_{\mathcal{C}^\perp}\|^2$ and $\lambda_2(\overline{L}) > 0$, we have

$$\begin{aligned} &\mathcal{L}V(i, \mathbf{x}) \\ &\leq -2\alpha(i)\mathbf{x}^T \overline{L}^2 \mathbf{x} + \alpha^2(i)\left[\lambda_N(\overline{L})\eta + \frac{\lambda_N^3(\overline{L})}{\lambda_2(\overline{L})}\mathbf{x}^T \overline{L}\mathbf{x}\right. \\ &\quad \left. + \frac{4N^2 \lambda_N(\overline{L})}{\lambda_2(\overline{L})}\mathbf{x}^T \overline{L}\mathbf{x}\right] \\ &\leq -\alpha(i)\varphi(i, \mathbf{x}) + g(i)\left[1 + V(i, \mathbf{x})\right] \end{aligned} \tag{40}$$

where

$$\begin{aligned} \varphi(i, \mathbf{x}) &= 2\mathbf{x}^T \overline{L}^2 \mathbf{x}, \\ g(i) &= \alpha^2(i)\max\left(\lambda_N(\overline{L})\eta, \frac{\lambda_N^3(\overline{L})}{\lambda_2(\overline{L})} + \frac{4N^2 \lambda_N(\overline{L})}{\lambda_2(\overline{L})}\right). \end{aligned} \tag{41}$$

It is easy to see that $\varphi(i, \mathbf{x})$ and $g(i)$ defined above satisfy the conditions for Theorem 2. Hence,

$$\mathbb{P}\left[\lim_{i \to \infty} \rho(\mathbf{x}(i), \mathcal{C}) = 0\right] = 1. \tag{42}$$

∎

Theorem 3 assumes **1.1)**, **2.1)**, **3)**. For an equivalent statement under **1.2)**, **2.2)**, see Theorem 11 in the Appendix.

Theorem 3 shows that the sample paths approach the consensus subspace, $\mathcal{C}$, with probability 1 as $i \to \infty$. We now show that the sample paths, in fact, converge a.s. to a finite point in $\mathcal{C}$.

Theorem 4: (a.s. consensus: Limiting random variable) Let assumptions **1.1)**, **2.1)**, and **3)** hold. Consider the $\mathcal{A} - \mathcal{ND}$ consensus algorithm in (9) in Section IV-A with initial state $\mathbf{x}(0) \in \mathbb{R}^{N \times 1}$. Then, there exists an almost sure finite real random variable $\boldsymbol{\theta}$ such that

$$\mathbb{P}\left[\lim_{i \to \infty} \mathbf{x}(i) = \boldsymbol{\theta}\mathbf{1}\right] = 1 \qquad (43)$$

Proof: Denote the average of $\mathbf{x}(i)$ by

$$x_{\text{avg}}(i) = \frac{1}{N}\mathbf{1}^T\mathbf{x}(i). \qquad (44)$$

The conclusions of Theorem 3 imply that

$$\mathbb{P}\left[\lim_{i \to \infty} \|\mathbf{x}(i) - x_{\text{avg}}(i)\mathbf{1}\| = 0\right] = 1. \qquad (45)$$

Recall the distributed average consensus algorithm in (9). Premultiplying both sides of (9) by $1^T/N$, we get the stochastic difference equation for the average as

$$\begin{aligned} x_{\text{avg}}(i+1) &= x_{\text{avg}}(i) - \boldsymbol{\xi}(i) \\ &= x_{\text{avg}}(0) - \sum_{0 \le j \le i} \boldsymbol{\xi}(j) \end{aligned} \qquad (46)$$

where

$$\boldsymbol{\xi}(i) = \frac{\alpha(i)}{N}\mathbf{1}^T\mathbf{n}(i).$$

Given (14), in particular, the sequence $\{\mathbf{n}(i)\}$ is time independent, it follows that

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{\xi}(i)\right] &= 0, \ \forall i \\ \sum_{i \ge 0} \mathbb{E}[\boldsymbol{\xi}(i)]^2 &= \sum_{i \ge 0} \frac{\alpha^2(i)}{N^2}\mathbb{E}\left[\|\mathbf{n}(i)\|^2\right] \\ &\le \frac{\eta}{N^2}\sum_{i \ge 0} \alpha^2(i) \\ &< \infty \end{aligned} \qquad (47)$$

which implies

$$\mathbb{E}\left[(x_{\text{avg}}(i))^2\right] \le x_{\text{avg}}^2(0) + \frac{\eta}{N^2}\sum_{j \ge 0} \alpha^2(j), \ \forall i. \qquad (48)$$

Thus, the sequence $\{x_{\text{avg}}(i)\}_{i \ge 0}$ is an $\mathcal{L}_2$ bounded martingale and, hence, converges a.s. and in $\mathcal{L}_2$ to a finite random variable θ (see, [33]). The theorem then follows from (45). ∎

Again, we note that we obtain an equivalent statement under Assumptions **1.2)**, **2.1)** in Theorem 7 in the Appendix. Proving under Assumption **2.2)** requires more specific information about the mixing properties of the Markovian noise sequence, which due to space limitations is not addressed here.

C. Mean Square Error

By Theorem 3, the sensors reach consensus asymptotically and converge a.s. to a finite random variable θ. Viewed as an estimate of the initial average r (see (4)), θ should possess desirable properties like unbiasedness and small mse. The next Lemma characterizes the desirable statistical properties of θ.

Lemma 5: Let θ be as in Theorem 4 and r as in (4). Define the mse ζ as

$$\zeta = \mathbb{E}\left[(\boldsymbol{\theta} - r)^2\right]. \qquad (49)$$

Then, the consensus limit θ is unbiased and its mse bounded as shown

$$\mathbb{E}\left[\theta\right] = r \qquad (50)$$

$$\zeta \le \frac{\eta}{N^2}\sum_{i \ge 0} \alpha^2(i). \qquad (51)$$

Proof: From (46), we have

$$\mathbb{E}\left[x_{\text{avg}}(i)\right] = r, \ \forall i. \qquad (52)$$

Since $\{x_{\text{avg}}(i)\}_{i \ge 0}$ converges in $\mathcal{L}_2$ to θ (Theorem 4), it also converges in $\mathcal{L}_1$, and we have

$$\mathbb{E}\left[\theta\right] = \lim_{i \to \infty} \mathbb{E}\left[x_{\text{avg}}(i)\right] = r \qquad (53)$$

which proves (50). For (51), by Theorem 4, the sequence $\{(x_{\text{avg}}(i) - r)^2\}_{i \ge 0}$ converges in $\mathcal{L}_2$ to $(\theta - r)^2$. Hence

$$\begin{aligned} \zeta &= \mathbb{E}\left[(\boldsymbol{\theta} - r)^2\right] \\ &= \lim_{i \to \infty} \mathbb{E}\left[(x_{\text{avg}}(i) - r)^2\right] \\ &= \sum_{i \ge 0} \frac{\alpha^2(i)}{N^2}\mathbb{E}\left[\|\mathbf{n}(i)\|^2\right]. \end{aligned} \qquad (54)$$

Equation (51) follows from (54) and (48), with η the bound on the noise variance. ∎

Lemma 13 in the Appendix shows equivalent results under Assumptions **1.2)** and **2.1)**. Proving under Assumption **2.2)** requires more specific information about the mixing properties of the Markovian noise sequence, which we do not pursue here.

Equation (54) gives the exact representation of the mse. As a byproduct, we obtain the following corollary for an erasure network with identical link failure probabilities and i.i.d. channel noise.

Corollary 6: Consider an erasure network with M realizable links, identical link failure probability, p, for each link, and the noise sequence $\{v_{nl}(i)\}_{1 \le n,l \le N, \ i \ge 0}$ be of identical variance σ^2. Then the mse is

$$\zeta = \frac{2M\sigma^2(1-p)}{N^2}\sum_{j \ge 0} \alpha^2(j). \qquad (55)$$

Proof: Using the fact that the link failures and the channel noise are independent, we have

$$\mathbb{E}[\|\mathbf{n}(i)\|^2] = 2\sigma^2 M(1-p). \qquad (56)$$

The result then follows from (54). ∎

While interpreting the dependence of (55) on the number of nodes, N, it should be noted that the number of realizable links M must be $O(N)$ for the connectivity assumptions to hold.

Lemma 5 shows that, for a given bound on the noise variance, η, the mse ζ can be made arbitrarily small by properly scaling the weight sequence, $\{\alpha(j)\}_{j \geq 0}$. As an example, consider the weight sequence

$$\alpha(j) = \frac{1}{j+1}, \ \forall j.$$

Clearly, this choice of $\alpha(i)$ satisfies the persistence conditions of (28) and, in fact

$$\sum_{j \geq 0} \alpha^2(j) = \sum_{j \geq 1} \frac{1}{j^2} = \frac{\pi^2}{6}.$$

Then, for any $\epsilon > 0$, the scaled weight sequence $\{\tilde{\alpha}(j)\}_{j \geq 0}$

$$\tilde{\alpha}(j) = \frac{\sqrt{6\epsilon}N}{\sqrt{\eta}\pi(j+1)}$$

will guarantee that $\zeta \leq \epsilon$. However, reducing the mse by scaling the weights in this way will reduce the convergence rate of the algorithm; this tradeoff is considered in the next section.

D. Convergence Rate

The $\mathcal{A} - \mathcal{ND}$ algorithm falls under the framework of stochastic approximation and, hence, a detailed convergence rate analysis can be done through the ODE method (see, for example, [34].) For clarity of presentation, we skip this detailed analysis here; rather, we present a simpler convergence rate analysis, involving the mean state vector sequence only under assumptions **1.1), 2.1)** , and **3)** . From the asymptotic unbiasedness of θ, it follows

$$\lim_{i \to \infty} \mathbb{E}\left[\mathbf{x}(i)\right] = r\mathbf{1}. \tag{57}$$

Our goal is to determine the rate at which the sequence $\{\mathbb{E}\left[\mathbf{x}(i)\right]\}_{i \geq 0}$ converges to $r\mathbf{1}$. Since $L(i)$ and $\mathbf{x}(i)$ are independent, and $\mathbf{n}(i)$ is zero mean, we have

$$\mathbb{E}\left[\mathbf{x}(i+1)\right] = \left(I - \alpha(i)\overline{L}\right)\mathbb{E}\left[\mathbf{x}(i)\right], \ \forall i. \tag{58}$$

By the persistence condition (18) the sequence $\alpha(i) \to 0$. Without loss of generality, we can assume that

$$\alpha(i) \leq \frac{2}{\lambda_2(\overline{L}) + \lambda_N(\overline{L})}, \ \forall i. \tag{59}$$

Then, it can be shown that (see [11])

$$\left\| \mathbb{E}\left[\mathbf{x}(i)\right] - r\mathbf{1} \right\| \leq \left(\prod_{0 \leq j \leq i-1} \left(1 - \alpha(j)\lambda_2(\overline{L})\right) \right) \left\| \mathbb{E}\left[\mathbf{x}(0)\right] - r\mathbf{1} \right\|. \tag{60}$$

Now, since $1 - a \leq e^{-a}$, $0 \leq a \leq 1$, we have

$$\left\| \mathbb{E}\left[\mathbf{x}(i)\right] - r\mathbf{1} \right\| \leq \left(e^{-\lambda_2(\overline{L})\left(\sum_{0 \leq j \leq i-1} \alpha(j)\right)} \right) \left\| \mathbb{E}\left[\mathbf{x}(0)\right] - r\mathbf{1} \right\|. \tag{61}$$

Equation (61) shows that the rate of convergence of the mean consensus depends on the topology through the algebraic connectivity $\lambda_2(\overline{L})$ of the mean graph and through the weights $\alpha(i)$. It is interesting to note that the way the random link failures affect the convergence rate (at least for the mean state sequence) is through $\lambda_2(\overline{L})$, the algebraic connectivity of the *mean* Laplacian, in the exponent, whereas, for a static network, this reduces to the algebraic connectivity of the static Laplacian L, recovering the results in [17].

Equations (61) and (51) show a tradeoff between the mse and the rate of convergence at which the sequence $\{\mathbb{E}\left[\mathbf{x}(i)\right]\}_{i \geq 0}$ converges to $r\mathbf{1}$. Equation (61) shows that this rate of convergence is closely related to the rate at which the weight sequence $\alpha(i)$ sums to infinity. For a faster rate, we want the weights to sum up fast to infinity, i.e., the weights to be large. In contrast, (51) shows that, to achieve a small ζ, the weights should be small.

We studied the tradeoff between convergence rate and mse of the mean state vectors only. In general, more effective measures of convergence rate are appropriate; intuitively, the same tradeoffs will be exhibited, in the sense that the rate of convergence will be closely related to the rate at which the weight sequence, $\alpha(i)$, sums to infinity, as verified by the numerical studies presented next.

E. Numerical Studies—$\mathcal{A} - \mathcal{ND}$

We present numerical studies on the $\mathcal{A} - \mathcal{ND}$ algorithm that verify the analytical results. The first set of simulations confirms the a.s. consensus in Theorems 3 and 4. Consider an erasure network on $N = 100$ nodes and $M = 5N$ realizable links, with identical probability of link failure $p = .4$ and identical channel noise variance $\sigma^2 = 15$. We take $\alpha(i) = 1/4i$ and plot on Fig. 1 on the left, the sample paths $\{x_n(i)\}_{1 \leq n \leq N}$, of the sensors over an instantiation of the $\mathcal{A} - \mathcal{ND}$ algorithm. We note that the sensor states converge to consensus, thus verifying our analytical results.

The second set of simulations confirms the mse in Corollary 6. We consider the same erasure network, but take $\sigma^2 = 30$ and $\alpha(i) = 1/5i$. We simulate 50 runs of the $\mathcal{A} - \mathcal{ND}$ algorithm from the initial state. Fig. 1 on the center plots the propagation of the squared error $(x_n(i) - r)^2$ for a randomly chosen sensor n for each of the 50 runs. The cloud of (blue) lines denotes the 50 runs, whereas the extended dashed (red) line denotes the exact mse computed in Corollary 6. The paths are clustered around the exact mse, thus verifying our results.

The third set of simulations studies the tradeoff between mse and convergence rate. We consider the same erasure network, but take $\sigma^2 = 50$, and run the $\mathcal{A} - \mathcal{ND}$ algorithm from the same initial conditions, but for the weight sequences $\{\alpha_s(i) = s/i\}_{i \geq 0}$, $s = .33, .1$. Fig. 1 on the right depicts the propagation of the squared error averaged over all the sensors $(1/N)\sum_{n=1}^{N}(x_n(i) - r)^2$ for each case. We see that the solid (blue) line $(s = .33)$ decays much faster initially than the dotted (red) line $(s = .1)$ and reaches a steady state. The dotted line ultimately crosses the solid line and continues to decay at a very slow rate, thus verifying the mse versus convergence
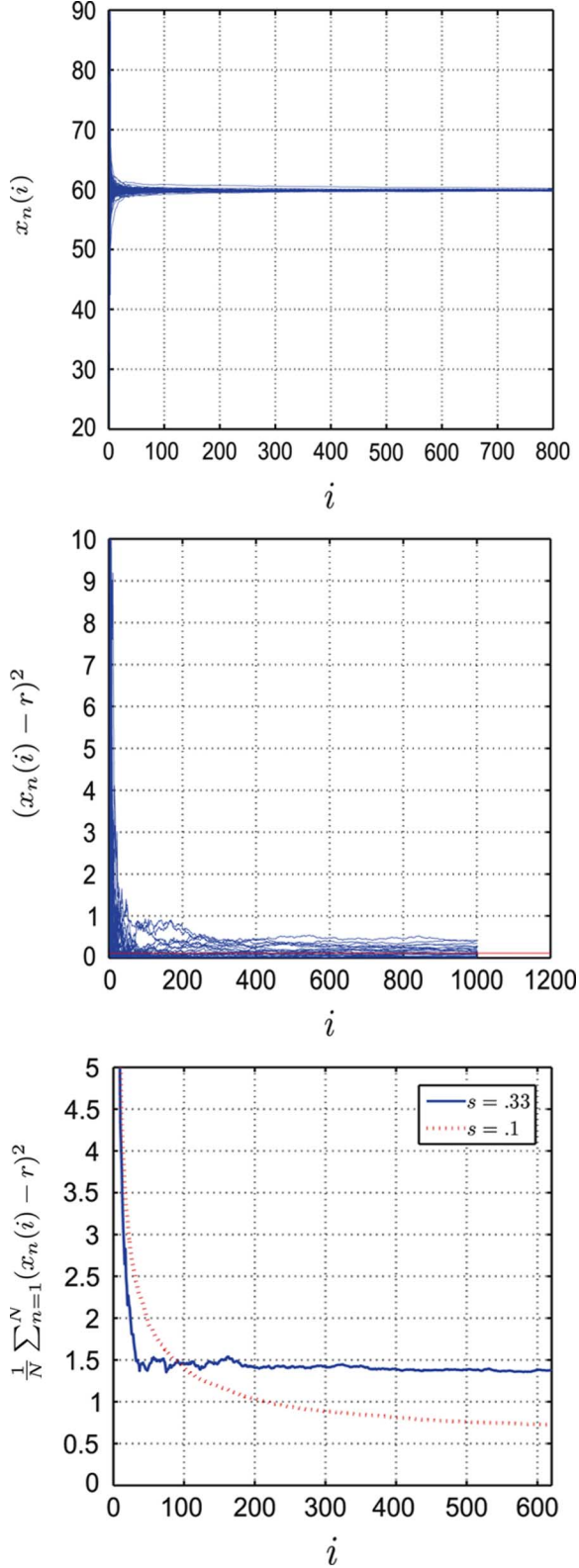
Fig. 1. Left: Sensors sample paths, $\{x_n(i)\}_{1 \leq n \leq N}$, of the $\mathcal{A} - \mathcal{ND}$ algorithm, verifying a.s. consensus. Center: Sample paths of the $\mathcal{A} - \mathcal{ND}$ algorithm, verifying the mse bound. Right: mse versus convergence rate tradeoff.

rate tradeoff, which we established rigorously by restricting attention to the mean state only.

V. $\mathcal{A} - \mathcal{NC}$: CONSENSUS WITH REPEATED AVERAGING

The stochastic approximation approach to the average consensus problem, discussed in Section IV, achieves arbitrarily small mse, see (51), possibly at the cost of a lower convergence rate, especially, when the desired mse is small. This is mainly because the weights $\alpha(i)$s decrease to zero, slowing the convergence as time progresses, as discussed in Section IV-E. In this section, we consider an alternative approach based on repeated averaging, which removes this difficulty. We use a constant link weight (or step size) and run the consensus iterations for a fixed number, $\hat{\imath}$, of iterations. This procedure is repeated $\hat{p}$ times, each time starting from the same initial state $\mathbf{x}(0)$. Since the $\hat{p}$ final states obtained at iteration $\hat{\imath}$ of each of the $\hat{p}$ runs are independent, we average them and get the law of large numbers to work for us. There is an interesting tradeoff between $\hat{\imath}$ and $\hat{p}$ for a constant total number of iterations $\hat{\imath}\hat{p}$. We describe and analyze this algorithm and consider its tradeoffs next. The section is organized as follows. Section V-A sets up the problem, states the assumptions, and gives the algorithm $\mathcal{A} - \mathcal{NC}$ for distributed average consensus with noisy communication links. We analyze the performance of the $\mathcal{A} - \mathcal{NC}$ algorithm in Section V-B. In Section V-C, we present numerical studies and suggest generalizations in Section V-D.

A. $\mathcal{A} - \mathcal{NC}$: Problem Formulation and Assumptions

Again, we consider distributed consensus with communication channel imperfections in (7) to average the initial state, $\mathbf{x}(0) \in \mathbb{R}^{N \times 1}$. The setup is the same as in (8) to (10).

$\mathcal{A} - \mathcal{NC}$ **Algorithm**: The $\mathcal{A} - \mathcal{NC}$ algorithm is the following Monte Carlo (MC) averaging procedure:

$$\mathbf{x}^p(i+1) = \mathbf{x}^p(i) - \alpha \left(L\mathbf{x}^p(i) + \mathbf{n}^p(i) \right)$$
$$0 \leq i \leq \hat{\imath} - 1, \ 1 \leq p \leq \hat{p}, \ \mathbf{x}^p(0) = \mathbf{x}(0) \quad (62)$$

where $x_n^p(i)$ is the state at sensor n at the ith iteration of the pth MC run. In particular, $x_n^p(\hat{\imath})$ is the state at sensor n at the end of the pth MC run. Each run of the $\mathcal{A} - \mathcal{NC}$ algorithm proceeds for $\hat{\imath}$ iterations and there are $\hat{p}$ MC runs. Finally, the estimate $\overline{x}_n^{\hat{p}}(\hat{\imath})$ of the average $x_{\text{avg}}(0)$ at sensor n is

$$\overline{x}_n^{\hat{p}}(\hat{\imath}) = \frac{1}{\hat{p}} \sum_{p=1}^{\hat{p}} x_n^p(\hat{\imath}). \quad (63)$$

We analyze the $\mathcal{A} - \mathcal{NC}$ algorithm under the following assumptions. These make the analysis tractable, but, are not necessary. They can be substantially relaxed, as shown in Section V-D.

1) **Static Network**: The Laplacian L is fixed (deterministic,) and the network is connected, $\lambda_2(L) > 0$. (In Section V-D we will allow random link failures.)

2) **Independent Gaussian Noise Sequence**: The additive noise $\{v_{nl}^p(i)\}_{1 \leq n, l \leq N, \ i, p \geq 0}$ is an independent Gaussian sequence with

$$\mathbb{E}\left[v_{nl}^p(i)\right] = 0, \ \forall 1 \leq n, l \leq N, \ i, p \geq 0,$$
$$\sup_{n, l, i, p} \mathbb{E}\left[\left(v_{nl}^p(i)\right)^2\right] = \mu < \infty. \quad (64)$$

From (10), it then follows that

$$\mathbb{E}\left[\mathbf{n}^p(i)\right] = \mathbf{0}, \ \forall i, p$$
$$\sup_{i,l,p} \mathbb{E}\left[|n_l^p(i)|^2\right] = \phi_{\max}^2 \leq (N-1)\mu < \infty \qquad (65)$$

3) **Constant link weight**: The link weight α is constant across iterations and satisfies

$$0 < \alpha < \frac{2}{\lambda_N(L)}. \qquad (66)$$

Let $r = 1^T \mathbf{x}(0)/N$ be the initial average. To define a uniform convergence metric, assume that the initial sensor observations $\mathbf{x}(0)$ belong to the following set (for some $K \in [0, \infty)$):

$$\mathcal{K} = \left\{\mathbf{x}(0) \in \mathbb{R}^{N \times 1} : | \ \|\mathbf{x}(0) - r\mathbf{1}\| \leq K\right\}. \qquad (67)$$

As performance metric for the $\mathcal{A} - \mathcal{NC}$ approach, we adopt the $\epsilon - \delta$ averaging time $T^\alpha(\epsilon, \delta)$ given by

$$T^\alpha(\epsilon, \delta) = i^* p^* \qquad (68)$$

where

$$(i^*, p^*)$$
$$= \arg \inf_{\widetilde{i,p}} \left\{ (\widehat{ip}) \mid \inf_{\mathbf{x}(0) \in \mathcal{K}} \inf_n \mathbb{P}\left(\frac{|\overline{x}_n^p(\widehat{i}) - r|}{K} \leq \epsilon\right) \geq 1 - \delta \right\} \qquad (69)$$

and the superscript α denotes explicitly the dependence on the link weight α.

We say that the $\mathcal{A} - \mathcal{NC}$ algorithm achieves (ϵ, δ)-*consensus* if $T^\alpha(\epsilon, \delta)$ is finite. A similar notion of averaging time has been used by others, see for example, [35] and [36]. The next section upper bounds the averaging time and analyzes the performance of $\mathcal{A} - \mathcal{NC}$.

B. Performance Analysis of $\mathcal{A} - \mathcal{NC}$

The $\mathcal{A} - \mathcal{NC}$ iterations can be rewritten as

$$\mathbf{x}^p(i+1) = W\mathbf{x}^p(i) + \boldsymbol{\chi}^p(i), \ 0 \leq i \leq \widehat{i} - 1$$
$$1 \leq p \leq \widehat{p}, \ \mathbf{x}^p(0) = \mathbf{x}(0) \qquad (70)$$

where

$$W = I - \alpha L$$
$$\boldsymbol{\chi}^p(i) = -\alpha \mathbf{n}^p(i), \ 0 \leq i \leq \widehat{i} - 1, \ 1 \leq p \leq \widehat{p}. \qquad (71)$$

Also, for the choice of α given in (66), the following can be shown (see [7]) for the spectral norm

$$\gamma_2 = \rho\left(W - \frac{1}{N}J\right)$$
$$= \rho\left(I - \alpha L - \frac{1}{N}J\right) < 1, \ J = \mathbf{1}\mathbf{1}^T \qquad (72)$$

where $\rho(\cdot)$ is the spectral radius, which is equal to the induced matrix 2 norm for symmetric matrices.

We next develop an upper bound on the averaging time $T^\alpha(\epsilon, \delta)$ given in (68). Actually, we derive a general bound

that holds for generic weight matrices W. This we do next, in Theorem 7. We come back in Theorem 10 to bounding the averaging time (68) for the model (70) when the weight matrix W is as in (71) and the spectral norm is given by (72).

Theorem 7: Averaging Time Consider the distributed iterative procedure (70). The weight matrix W is generic, i.e., is not necessarily of the form (71). It does satisfy the following assumptions:

1) **Symmetric Weights**:

$$W = W^T, \ W\mathbf{1} = \mathbf{1}, \ \gamma_2 = \rho\left(W - \frac{1}{N}J\right) < 1. \qquad (73)$$

2) **Noise Assumptions**: The sequence $\{\boldsymbol{\chi}^p(i)\}_{i \geq 0, p \geq 1}$ is a sequence of independent Gaussian noise vectors with uncorrelated components, such that

$$\mathbb{E}\left[\boldsymbol{\chi}^p(i)\right] = \mathbf{0}, \ \sup_{i \geq 0, p \geq 1} \sup_{1 \leq l \leq N} \mathbb{E}\left[(\chi_l^p(i))^2\right] \leq \phi_{\max}^2 < \infty. \qquad (74)$$

Then, we have the following upper bound on the averaging time $T^{\gamma_2}(\epsilon, \delta)$ given by (68)

$$T^{\gamma_2}(\epsilon, \delta) \leq \widehat{T}^{\gamma_2}(\epsilon, \delta) \qquad (75)$$
$$\widehat{T}^{\gamma_2}(\epsilon, \delta) = \left(\frac{\ln \frac{\epsilon}{2}}{\ln \gamma_2} + 1\right)\left[\left(\frac{4\phi_{\max}^2 \ln\left(\frac{2}{\delta}\right)}{K\epsilon}\right)\right.$$
$$\left.\cdot\left(\frac{\ln \frac{\epsilon}{2}}{N \ln \gamma_2} + \frac{1}{N} + \frac{1 - \frac{\gamma_2^2 \epsilon^2}{4}}{1 - \gamma_2^2}\left(1 - \frac{1}{N}\right)\right) + 1\right]. \qquad (76)$$

(Note we replace the superscript α by γ_2 because we prove results here for arbitrary W satisfying (73), not necessarily of the form $I - \alpha L$.)

For the proof we need a result from [10], which we state as a Lemma.

Lemma 8: Let the assumptions in Theorem 7 hold true. Define

$$m_l(\widehat{i}) = \mathbb{E}\left[x_l^p(\widehat{i})\right],$$
$$v_l(\widehat{i}) = \mathbb{E}\left[(x_l^p(\widehat{i}) - m_l(\widehat{i}))^2\right]$$
$$1 \leq p \leq \widehat{p}, \ 1 \leq l \leq N. \qquad (77)$$

(Note that these quantities do not depend on p.) Then we have the following:

1) **Bound on error mean**:

$$|m_l(\widehat{i}) - r| \leq \gamma_2^{\widehat{i}}\|\mathbf{x}(0) - r\mathbf{1}\|$$
$$\leq \gamma_2^{\widehat{i}}K, \ \forall \mathbf{x}(0) \in \mathcal{K}, \ 1 \leq l \leq N \qquad (78)$$

2) **Bound on error variance**:

$$v_l(\widehat{i}) \leq \phi_{\max}^2\left[\frac{\widehat{i}}{N} + \frac{1 - \gamma_2^{2\widehat{i}}}{1 - \gamma_2^2}\left(1 - \frac{1}{N}\right)\right]$$
$$\forall \mathbf{x}(0) \in \mathbb{R}^{N \times 1}, \ 1 \leq l \leq N \qquad (79)$$

3) $\{x_l^p(\widehat{i})\}_{p=1}^{\widehat{p}}$ is an i.i.d. sequence, with

$$x_l^p(\widehat{i}) \stackrel{\mathrm{d}}{=} \mathcal{N}(m_l(\widehat{i}), v_l(\widehat{i})), \ 1 \leq p \leq \widehat{p}. \qquad (80)$$

Proof: For proof, see [10]. ∎

We now return to the proof of Theorem 7.

Proof: [Theorem 7] The estimate $\overline{x}_l^{\widehat{p}}(\widehat{\imath})$ of the average at each sensor after $\widehat{p}$ runs is in (63). From Lemma 8 and a standard Chernoff type bound for Gaussian random variables (see [37]), then

$$\mathbb{P}\left(\frac{\left|\overline{x}_l^{\widehat{p}}(\widehat{\imath}) - m_l(\widehat{\imath})\right|}{K} \le \epsilon\right) \ge 1 - 2\exp\left\{\frac{-\widehat{p}K\epsilon}{2v_l(\widehat{\imath})}\right\}. \quad (81)$$

(For the present derivation we assume Gaussian noise; however, the analysis for other noise models may be done by using the corresponding large deviation rate functions.)

For arbitrary $\epsilon > 0$, define

$$\widehat{\imath}(\epsilon) = \left\lceil \frac{\ln \epsilon}{\ln \gamma_2} \right\rceil. \quad (82)$$

Also, for arbitrary $\epsilon, \delta > 0$, define

$$\widehat{p}(\epsilon, \delta) = \left\lceil \frac{2v_l(\widehat{\imath})}{K\epsilon} \ln \frac{2}{\delta} \right\rceil. \quad (83)$$

Then, we have from (78) in Lemma 8

$$\frac{|m_l(\widehat{\imath}) - r|}{K} \le \frac{\epsilon}{2}, \quad \forall \widehat{\imath} \ge \widehat{\imath}\left(\frac{\epsilon}{2}\right), \ 1 \le l \le N. \quad (84)$$

Also, we have from (81),

$$\mathbb{P}\left(\frac{\left|\overline{x}_l^{\widehat{p}}(\widehat{\imath}) - m_l(\widehat{\imath})\right|}{K} \le \frac{\epsilon}{2}\right) \ge 1 - \delta$$
$$\forall \widehat{p} \ge \widehat{p}\left(\frac{\epsilon}{2}, \delta\right), \ 1 \le l \le N. \quad (85)$$

From the triangle inequality, we get

$$\left|\overline{x}_l^{\widehat{p}}(\widehat{\imath}) - r\right| \le \left|\overline{x}_l^{\widehat{p}}(\widehat{\imath}) - m_l(\widehat{\imath})\right| + |m_l(\widehat{\imath}) - r|, \forall 1 \le l \le N. \quad (86)$$

It then follows from (84) that

$$\left|\overline{x}_l^{\widehat{p}}(\widehat{\imath}) - m_l(\widehat{\imath})\right| \le K\frac{\epsilon}{2} \implies \left|\overline{x}_l^{\widehat{p}}(\widehat{\imath}) - r\right| \le K\epsilon,$$
$$\forall \widehat{\imath} \ge \widehat{\imath}\left(\frac{\epsilon}{2}\right), \ 1 \le l \le N. \quad (87)$$

We, thus, have for $\widehat{\imath} \ge \widehat{\imath}(\epsilon/2)$ and $\widehat{p} \ge \widehat{p}(\epsilon/2, \delta)$

$$\mathbb{P}\left(\frac{\left|\overline{x}_l^{\widehat{p}}(\widehat{\imath}) - r\right|}{K} \le \epsilon\right) \ge \mathbb{P}\left(\frac{\left|\overline{x}_l^{\widehat{p}}(\widehat{\imath}) - m_l(\widehat{\imath})\right|}{K} \le \frac{\epsilon}{2}\right)$$
$$\ge 1 - \delta, \ 1 \le l \le N \quad (88)$$

where the last inequality follows from (85).

From the definition of $T^{\gamma_2}(\epsilon, \delta)$, see (68), we then have

$$T^{\gamma_2}(\epsilon, \delta) \le \widehat{\imath}\left(\frac{\epsilon}{2}\right)\widehat{p}\left(\frac{\epsilon}{2}, \delta\right) \quad (89)$$
$$\le \left(\frac{\ln\frac{\epsilon}{2}}{\ln \gamma_2} + 1\right)\left(\frac{4v_l(\widehat{\imath})}{K\epsilon}\ln\frac{2}{\delta} + 1\right)$$
$$\le \left(\frac{\ln\frac{\epsilon}{2}}{\ln \gamma_2} + 1\right)\left[\left(\frac{4\phi_{\max}^2 \ln\left(\frac{2}{\delta}\right)}{K\epsilon}\right)\right.$$
$$\left.\cdot\left(\frac{\ln\frac{\epsilon}{2}}{N\ln\gamma_2} + \frac{1}{N} + \frac{1 - \frac{\gamma_2^2\epsilon^2}{4}}{1 - \gamma_2^2}\left(1 - \frac{1}{N}\right)\right) + 1\right]$$
$$= \widehat{T}^{\gamma_2}(\epsilon, \delta) \quad (90)$$

where the third inequality follows from (79). ∎

We call the upper bound on $T^{\gamma_2}(\epsilon, \delta)$ in (76) given in Theorem 7 the *approximate averaging time*. We use $\widehat{T}^{\gamma_2}(\epsilon, \delta)$ to characterize the convergence rate of $\mathcal{A} - \mathcal{NC}$. We state a property of $\widehat{T}^{\gamma_2}(\epsilon, \delta)$.

Lemma 9: Recall the spectral radius, γ_2, defined in (73). Then, for $0 < \epsilon, \delta < 1$, $\widehat{T}^{\gamma_2}(\epsilon, \delta)$ is an increasing function of γ_2 in the interval $0 < \gamma_2 < 1$.

Proof: The lemma follows by differentiating $\widehat{T}^{\gamma_2}(\epsilon, \delta)$ with respect to γ_2. ∎

We now study the convergence properties of the $\mathcal{A} - \mathcal{NC}$ algorithm, (62), i.e., when the weight matrix W is of the form (71) and the spectral norm satisfies (72). Then, the averaging time becomes a function of the weight α. We make this explicit by denoting as $T^\alpha(\epsilon, \delta)$ and $\widehat{T}^\alpha(\epsilon, \delta)$ the averaging time and the approximate averaging time, respectively.

Theorem 10: ($\mathcal{A} - \mathcal{NC}$ *Averaging Time*) Consider the $\mathcal{A} - \mathcal{NC}$ algorithm for distributed averaging, under the assumptions given in Section V-A. Define

$$T^*(\epsilon, \delta) = \inf_{0 < \alpha < 2/\lambda_N(L)} T^\alpha(\epsilon, \delta)$$
$$\widehat{T}^*(\epsilon, \delta) = \inf_{0 < \alpha < 2/\lambda_N(L)} \widehat{T}^\alpha(\epsilon, \delta). \quad (91)$$

Then
1) $0 < \alpha < \frac{2}{\lambda_N(L)}$,

$$T^\alpha(\epsilon, \delta) < \infty, \quad \forall \epsilon > 0, \delta < 1. \quad (92)$$

This essentially means that the $\mathcal{A} - \mathcal{NC}$ algorithm is realizable for α in the interval $(0, 2/\lambda_N(L))$.
2) For $0 < \alpha < 2/\lambda_N(L)$, we have

$$T^\alpha(\epsilon, \delta) \le \widehat{T}^\alpha(\epsilon, \delta) \quad (93)$$
$$\widehat{T}^\alpha(\epsilon, \delta) = \left(\frac{\ln\frac{\epsilon}{2}}{\ln\gamma_2} + 1\right)\left[\left(\frac{4\alpha^2\phi_{\max}^2 \ln\left(\frac{2}{\delta}\right)}{K\epsilon}\right)\right.$$
$$\left.\cdot\left(\frac{\ln\frac{\epsilon}{2}}{N\ln\gamma_2} + \frac{1}{N} + \frac{1 - \frac{\gamma_2^2\epsilon^2}{4}}{1 - \gamma_2^2}\left(1 - \frac{1}{N}\right)\right) + 1\right] \quad (94)$$

and $\gamma_2 = \rho(I - \alpha L - 1/NJ)$.
3) For a given choice of the pair $0 < \epsilon, \delta < 1$, the best achievable averaging time $T^*(\epsilon, \delta)$ is bounded above by $\widehat{T}^*(\epsilon, \delta)$, given by

$$\widehat{T}^*(\epsilon, \delta) = \inf_{0 < \alpha \le (2/\lambda_2(L) + \lambda_N(L))}\left(\frac{\ln\frac{\epsilon}{2}}{\ln(1 - \alpha\lambda_2(L))} + 1\right)$$
$$\cdot\left[\left(\frac{4\alpha^2\phi_{\max}^2 \ln\left(\frac{2}{\delta}\right)}{K\epsilon}\right)\right.$$
$$\times\left(\frac{\ln\frac{\epsilon}{2}}{N\ln(1-\alpha\lambda_2(L))} + \frac{1}{N}\right.$$
$$\left.\left. + \frac{1 - \frac{(1-\alpha\lambda_2(L))^2\epsilon^2}{4}}{1 - (1 - \alpha\lambda_2(L))^2}\left(1 - \frac{1}{N}\right)\right) + 1\right]. \quad (95)$$

Note that in (95) the optimization is over a smaller range on α than in (91).

Proof: The iterations for the $\mathcal{A} - \mathcal{NC}$ algorithm are given by (70), where the weight matrix W is given by (71) and the spectral norm γ_2 by (72). Also, since $\boldsymbol{\chi}^p(i) = -\alpha \mathbf{n}^p(i)$, we get

$$\sup_{i \geq 0, p \geq 1} \sup_{1 \leq l \leq N} \mathbb{E}\left[(\chi_l^p(i))^2 \right] \leq \alpha^2 \phi_{\max}^2 < \infty. \quad (96)$$

Then, the assumptions (73), (74) in Theorem 7 are satisfied for α in the range (66) and the two items (92) and (94) follow. To prove item **3)**, we note that it follows from **2)** that $T^*(\epsilon, \delta) \leq \widehat{T}^*(\epsilon, \delta)$, where

$$\widehat{T}^*(\epsilon, \delta) = \inf_{0 < \alpha < 2/\lambda_N(L)} \widehat{T}^\alpha(\epsilon, \delta)$$
$$= \inf_{0 < \alpha < 2/\lambda_N(L)} \left(\frac{\ln \frac{\epsilon}{2}}{\ln \gamma_2} + 1 \right)$$
$$\times \left[\left(\frac{4\alpha^2 \phi_{\max}^2 \ln\left(\frac{2}{\delta}\right)}{K\epsilon} \right) \right.$$
$$\left. \left(\frac{\ln \frac{\epsilon}{2}}{N \ln \gamma_2} + \frac{1}{N} + \frac{1 - \frac{\gamma_2^2 \epsilon^2}{4}}{1 - \gamma_2^2} \left(1 - \frac{1}{N} \right) \right) + 1 \right]$$

and $\gamma_2 = \rho(I - \alpha L - 1/NJ)$. Now, consider the functions

$$g(\gamma_2) = \frac{4\phi_{\max}^2 \ln\left(\frac{2}{\delta}\right)}{K\epsilon} \left(\frac{\ln \frac{\epsilon}{2}}{\ln \gamma_2} + 1 \right)$$
$$\cdot \left[\frac{\ln \frac{\epsilon}{2}}{N \ln \gamma_2} + \frac{1}{N} + \frac{1 - \frac{\gamma_2^2 \epsilon^2}{4}}{1 - \gamma_2^2} \left(1 - \frac{1}{N} \right) \right] \quad (97)$$

and

$$h(\gamma_2) = \left(\frac{\ln \frac{\epsilon}{2}}{\ln \gamma_2} + 1 \right) \quad (98)$$

with γ_2 as before. Similar to Lemma 9, we can show that, $g(\gamma_2)$ and $h(\gamma_2)$ are nondecreasing functions of γ_2. It can be shown that $\gamma_2 = \rho(I - \alpha L - 1/NJ)$ attains its minimum value at $\alpha = \alpha^\bullet$ (see [7]), where

$$\alpha^\bullet = \frac{2}{\lambda_2(L) + \lambda_N(L)}. \quad (99)$$

We, thus, have

$$g\left(\rho\left(I - \alpha^\bullet L - \frac{1}{N}J \right) \right) \leq g\left(\rho\left(I - \alpha L - \frac{1}{N}J \right) \right)$$
$$\alpha^\bullet \leq \alpha < \frac{2}{\lambda_N(L)} \quad (100)$$

and

$$h\left(\rho\left(I - \alpha^\bullet L - \frac{1}{N}J \right) \right) \leq h\left(\rho\left(I - \alpha L - \frac{1}{N}J \right) \right)$$
$$\alpha^\bullet \leq \alpha < \frac{2}{\lambda_N(L)} \quad (101)$$

which implies, that, for $\alpha^\bullet \leq \alpha < 2/\lambda_N(L)$

$$\widehat{T}^\alpha(\epsilon, \delta) = \alpha^2 g\left(\rho\left(I - \alpha L - \frac{1}{N}J \right) \right)$$
$$+ h\left(\rho\left(I - \alpha L - \frac{1}{N}J \right) \right)$$
$$\geq (\alpha^\bullet)^2 g\left(\rho\left(I - \alpha^\bullet L - \frac{1}{N}J \right) \right)$$
$$+ h\left(\rho\left(I - \alpha^\bullet L - \frac{1}{N}J \right) \right)$$
$$= \widehat{T}^{\alpha^\bullet}(\epsilon, \delta). \quad (102)$$

So, there is no need to consider $\alpha > \alpha^\bullet$. This leads to

$$\widehat{T}^*(\epsilon, \delta) = \inf_{0 < \alpha < 2/\lambda_N(L)} \widehat{T}^\alpha(\epsilon, \delta)$$
$$= \inf_{0 < \alpha \leq 2/(\lambda_2(L) + \lambda_N(L))} \widehat{T}^\alpha(\epsilon, \delta). \quad (103)$$

Also, it can be shown that (see [7])

$$\gamma_2 = \rho\left(I - \alpha L - \frac{1}{N}J \right) = 1 - \alpha \lambda_2(L)$$
$$0 < \alpha \leq \frac{2}{\lambda_2(L) + \lambda_N(L)}. \quad (104)$$

This, together with (103) proves item **3)**. ∎

We comment on Theorem 10. For a given connected network, it gives explicitly the weight α for which the $\mathcal{A} - \mathcal{NC}$ algorithm is realizable (i.e., the averaging time is finite for any (ϵ, δ) pair.) For these choices of α, it provides an upper bound on the averaging time. The Theorem also addresses the problem of choosing the α that minimizes this upper bound.[3] There is, in general, no closed form solution for this problem and, as demonstrated by (95), even if a minimizer exists, its value depends on the actual choice of the pair (ϵ, δ). In other words, in general, there is no uniform minimizer of the averaging time.

C. $\mathcal{A} - \mathcal{NC}$: Numerical Studies

We present numerical studies on the $\mathcal{A} - \mathcal{NC}$ algorithm. We consider a sensor network of $N = 230$ nodes, with communication topology given by an LPS-II Ramanujan graph (see [10]), of degree 6.[4] For the first set of simulations, we take $\phi_{\max}^2 = 100$, $K = 50$, and fix δ at .05 (this guarantees that the estimate belongs to the ϵ-ball with probability at least 0.95.) We vary ϵ in steps, keeping the other parameters fixed, and compute the optimal (ϵ, δ) averaging time, $\widehat{T}^*(\epsilon, \delta)$, given by (95) and the corresponding optimum α^*. Fig. 2 on the top plots $\widehat{T}^*(\epsilon, \delta)$ as a function of ϵ, while Fig. 2 on the bottom plots α^* versus ϵ. As expected, $\widehat{T}^*(\epsilon, \delta)$ decreases with increasing ϵ. The behavior of α^* is interesting. It shows that, to improve accuracy (small ϵ), the link weight α should be chosen appropriately small, while, for lower accuracy, α can be increased, which speeds the algorithm. Also, as ϵ becomes larger, α^* increases to make the averaging time smaller, ultimately saturating at $\alpha^\bullet$, given by (99). This

---

[3]Note that the minimizer, α^* of the upper bound $\widehat{T}^\alpha(\epsilon, \delta)$ does not necessarily minimize the actual averaging time $T^\alpha(\epsilon, \delta)$. However, as will be demonstrated by simulation studies, the upper bound is tight and hence α^* obtained by minimizing $\widehat{T}^\alpha(\epsilon, \delta)$ is a good enough design criterion.

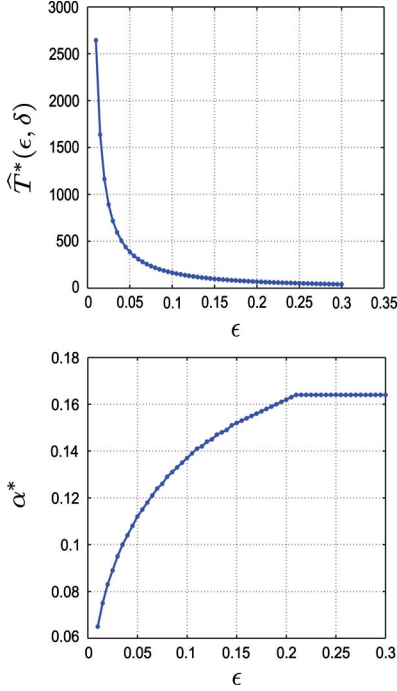[4]This is a 6-regular graph, i.e., all the nodes have degree 6.

Fig. 2. Top: Plot of $\widehat{T}^*(\epsilon, \delta)$ with varying ϵ, keeping δ fixed at .05. Bottom: Plot of α^* with varying ϵ, keeping δ fixed at .05.

behavior is similar to the $\mathcal{A} - \mathcal{ND}$ algorithm, where slower decreasing (smaller) weight sequences correspond to smaller asymptotic mse at a cost of lower convergence rate (increased accuracy.)

In the second set of simulations, we study the tradeoff between the number of iterations per Monte Carlo pass, $\widehat{\imath}$, and the total number of passes, $\widehat{p}$. Define the quantities, as suggested by (89)

$$\widehat{\imath}^* = \frac{\ln \frac{\epsilon}{2}}{\ln(1 - \alpha^* \lambda_2(L))} + 1 \tag{105}$$

$$\widehat{p}^* = \left(\frac{4\alpha^{*2} \phi_{\max}^2 \ln\left(\frac{2}{\delta}\right)}{K\epsilon} \right)$$
$$\cdot \left(\frac{\widehat{\imath}^*}{N} + \frac{1 - \frac{(1-\alpha^*\lambda_2(L))^2 \epsilon^2}{4}}{1 - (1-\alpha^*\lambda_2(L))^2} \left(1 - \frac{1}{N}\right) \right) + 1 \tag{106}$$

where α^* is the minimizer in (95). In the following, we vary ϵ and the channel noise variance $\phi_{\max}^2$, taking $K = 50$, $\delta = 0.05$, and using the same communication network. In particular, in Fig. 3 (top) we plot $(\widehat{\imath}^*, \widehat{p}^*)$ versus ϵ for $\phi_{\max}^2 = 10$, while in Fig. 3 (center) and Fig. 3 (bottom), we repeat the same for $\phi_{\max}^2 = 30$ and $\phi_{\max}^2 = 100$, respectively. The figures demonstrate an interesting tradeoff between $\widehat{\imath}^*$ and $\widehat{p}^*$, and show that for smaller values of the channel noise variance, the number of Monte Carlo passes, $\widehat{p}^*$ are much smaller than the number of iterations per pass, $\widehat{\imath}^*$, as expected. As the channel noise variance grows, i.e., as the channel becomes more unreliable, we note, that $\widehat{p}^*$ increases to combat the noise accumulated at each pass. Finally, we present a numerical study to verify the tightness of the bound $\widehat{T}^*(\epsilon, \delta)$. For this, we consider a network of $N = 100$ nodes with $M = 5N$ edges (generated according
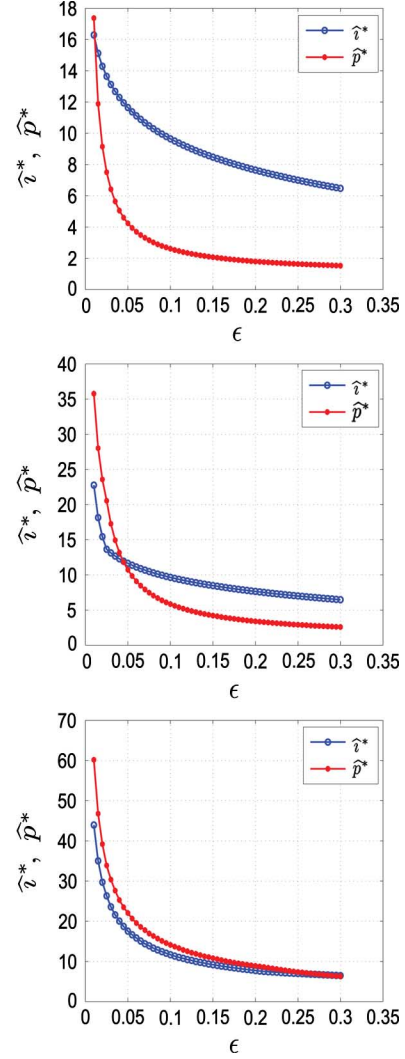

Fig. 3. Plot of $(\widehat{\imath}^*, \widehat{p}^*)$ with varying ϵ: Top: $\phi_{\max}^2 = 10$. Center: $\phi_{\max}^2 = 30$. Bottom: $\phi_{\max}^2 = 100$.

to an Erdös-Renýi graph, see [37].) We consider $\phi_{\max}^2 = 80$, $K = 50$, and $\delta = .05$. To obtain $T^*(\epsilon, \delta)$ for varying ϵ, we fix α, sample $\mathbf{x}(0) \in \mathcal{K}$ and for each such $\mathbf{x}(0)$, we generate 100 runs of the $\mathcal{A} - \mathcal{NC}$ algorithm. We check the condition in (69) and compute $T^\alpha(\epsilon, \delta)$. We repeat this experiment for varying α to obtain $T^*(\epsilon, \delta) = \inf_\alpha T^\alpha(\epsilon, \delta)$.[5] We also obtain $\widehat{T}^*(\epsilon, \delta)$ from (95). Fig. 4 on the top plots $T^*(\epsilon, \delta)$ (solid red line) and $\widehat{T}^*(\epsilon, \delta)$ (dotted blue line) with respect to ϵ, while Fig. 4 on the bottom plots the ratio $\widehat{T}^*(\epsilon, \delta)/T^*(\epsilon, \delta)$ with respect to ϵ. The plots show that the bound $\widehat{T}^*(\epsilon, \delta)$ is reasonably tight, especially at small and large values of ϵ, with fluctuations in between. Note that, because the numerical values obtained for $T^*(\epsilon, \delta)$ are a lower bound, the bound $\widehat{T}^*(\epsilon, \delta)$ is actually tighter than appears in the plots.

The bound $\widehat{T}^*(\epsilon, \delta)$ is significant since: i) it is easy to compute from (95); ii) it is a reasonable approximation to the exact $T^*(\epsilon, \delta)$; iii) it avoids the costly simulations involved in computing $T^*(\epsilon, \delta)$ by Monte Carlo; and iv) it gives the right

[5]The definition of $T^*(\epsilon, \delta)$ [see (69)] requires the infimum for all $\mathbf{x}(0) \in \mathcal{K}$, which is uncountable, so we verify (69) by sampling points from $\mathcal{K}$. The $T^*(\epsilon, \delta)$, thus obtained, is in fact, a lower bound for the actual $T^*(\epsilon, \delta)$.
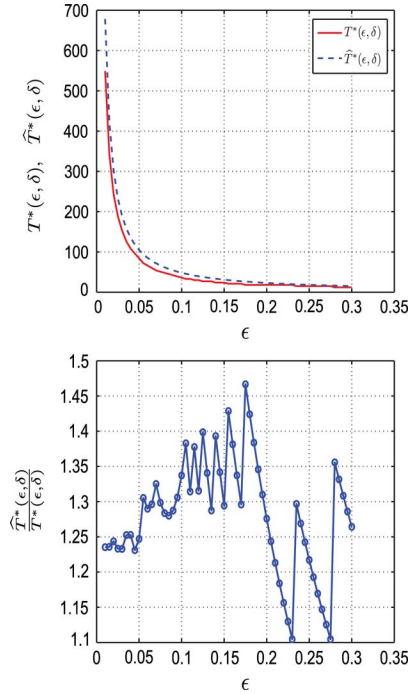
Fig. 4. Top: Plot of $T^*(\epsilon, \delta), \widehat{T}^*(\epsilon, \delta)$ with varying ϵ. Bottom: Plot of $\widehat{T}^*(\epsilon, \delta)/T^*(\epsilon, \delta)$ with varying ϵ.

tradeoff between $\widehat{\imath}^*$ and $\widehat{p}^*$ [see (105)–(106)], thus determining the stopping criterion of the $\mathcal{A} - \mathcal{NC}$ algorithm.

D. $\mathcal{A} - \mathcal{NC}$: Generalizations

In this Subsection, we suggest generalizations to the $\mathcal{A} - \mathcal{NC}$ algorithm. For convenience of analysis, we assumed before a static network and Gaussian noise. These assumptions can be considerably weakened. For instance, the static network assumption may be replaced by a random link failure model with $\lambda_2(\overline{L}) > 0$, where $\overline{L} = \mathbb{E}[L]$. Also, the independent noise sequence in (64) may be non-Gaussian. In this case, the $\mathcal{A} - \mathcal{NC}$ algorithm iterations will take the form

$$\mathbf{x}^p(i+1) = (\mathbf{x}^p(i) - \alpha L^p(i)\mathbf{x}^p(i)) - \alpha \mathbf{n}^p(i),$$
$$0 \le i \le \widehat{\imath} - 1, \, 1 \le p \le \widehat{p}, \, \mathbf{x}^p(0) = \mathbf{x}(0) \quad (107)$$

where $L^p(i)$ is an i.i.d. sequence of Laplacian matrices with $\lambda_2(\overline{L}) > 0$. We then have

$$\mathbf{x}^p(\widehat{\imath}) = \left(\prod_{j=0}^{\widehat{\imath}-1} (I - \alpha L^p(j)) \right) \mathbf{x}(0) + H^p(\widehat{\imath}), \, 1 \le p \le \widehat{p}. \quad (108)$$

It is clear that $\{H^p(\widehat{\imath})\}_{1 \le p \le \widehat{p}}$ is an i.i.d. sequence of zero mean random variables. Under the assumption $\lambda_2(\overline{L}) > 0$, there exists α (see [11]) such that

$$\mathbb{P}\left[\left(\prod_{j \ge 0} (I - \alpha L^p(j)) \right) \mathbf{x}(0) = r\mathbf{1} \right] = 1, \, 1 \le p \le \widehat{p}. \quad (109)$$

We, thus, choose $\widehat{\imath}$ so that $\left(\prod_{j=0}^{\widehat{\imath}-1} (I - \alpha L^p(j)) \right) \mathbf{x}(0)$ is sufficiently *close* to $r\mathbf{1}$. The final estimate is

$$\overline{\mathbf{x}^p}(\widehat{\imath}) = \frac{1}{\widehat{p}} \sum_{p=1}^{\widehat{p}} \left(\prod_{j=0}^{\widehat{\imath}-1} (I - \alpha L^p(j)) \right) \mathbf{x}(0) + \frac{1}{\widehat{p}} \sum_{p=1}^{\widehat{p}} H^p(\widehat{\imath}). \quad (110)$$

The first sum is close to $r\mathbf{1}$ by choice of $\widehat{\imath}$. We now choose $\widehat{p}$, large enough, so that the second term is *close* to zero. In this way, we can apply the $\mathcal{A} - \mathcal{NC}$ algorithm to more general scenarios.

The above argument guarantees that (ϵ, δ)-*consensus* is achievable under these generic conditions, in the sense that the corresponding averaging time, $T^\alpha(\epsilon, \delta)$, will be finite. A thorough analysis requires a reasonable computable upper bound like (94), followed by optimization over α to give the best achievable convergence rate. (Note a computable upper bound is required, because, as pointed earlier, it is very difficult to find stopping criterion using Monte Carlo simulations, and the resulting $T^*(\epsilon, \delta)$ will be a lower bound since the set $\mathcal{K}$ is uncountable.) One way to proceed is to identify the appropriate large deviation rate [as suggested in (81).] However, the results will depend on the specific nature of the link failures and noise, which we do not pursue in this paper due to lack of space.

VI. CONCLUSION

We consider distributed average consensus when the topology is random (links may fail at random times) *and* the communication in the channels is corrupted by additive noise. Noisy consensus leads to a bias-variance dilemma. We considered two versions of consensus that lead to two compromises to this problem: i) $\mathcal{A} - \mathcal{ND}$ fits the framework of stochastic approximation. It a.s. converges to the consensus subspace and to a consensus random variable θ–an unbiased estimate of the desired average, whose variance we compute and bound; and ii) $\mathcal{A} - \mathcal{NC}$ uses repeated averaging by Monte Carlo, achieving (ϵ, δ)-*consensus*. In $\mathcal{A} - \mathcal{ND}$ the bias can be made arbitrarily small, but the rate at which it decreases can be traded for variance—tradeoff between mse and convergence rate. $\mathcal{A} - \mathcal{NC}$ uses a constant weight α and hence outperforms $\mathcal{A} - \mathcal{ND}$ in terms of convergence rate. Computation-wise, $\mathcal{A} - \mathcal{ND}$ is superior since $\mathcal{A} - \mathcal{NC}$ requires more intersensor coordination to execute the independent passes. The estimate obtained by $\mathcal{A} - \mathcal{NC}$ does not possess the nice statistical properties, including unbiasedness, as the computation is terminated after a finite time in each pass.

Finally, these algorithms may be applied to other problems in sensor networks with random links and noise, e.g., distributed load balancing in parallel processing or distributed network flow.

APPENDIX
PROOF OF THEOREM 1 AND $\mathcal{A} - \mathcal{ND}$ GENERALIZATIONS UNDER ASSUMPTIONS **1.2)** AND **2.2)**

Proof: [Theorem 1] The proof follows from that of [32, Theorem 2.7.1]. Suffices to prove it for $\mathbf{x}(0) = \mathbf{x}_0$ a.s., $\mathbf{x}_0 \in \mathbb{R}^{N \times 1}$ is a deterministic starting state. Let the filtration

$\left\{\mathcal{F}_i^{\mathbf{X}} = \sigma\left\{\mathbf{x}(j) : 0 \le j \le i\right\}\right\}_{i \ge 0}$ w.r.t. which $\{\mathbf{x}(i)\}_{i \ge 0}$, $\mathbf{x} \in \mathbb{R}^{N \times 1}$ (and, hence, functionals of $\{\mathbf{x}(i)\}_{i \ge 0}$) are adapted.

Define the function $W(i, \mathbf{x}), i \ge 0$, $\mathbf{x} \in \mathbb{R}^{N \times 1}$ as

$$W(i, \mathbf{x}) = (1 + V(i, \mathbf{x})) \prod_{j \ge i} [1 + g(j)]. \tag{111}$$

It can be shown that

$$\mathcal{L}W(i, \mathbf{x}) \le -\alpha(i)\varphi(i, \mathbf{x}), \; i \ge 0, \; \mathbf{x} \in \mathbb{R}^{N \times 1} \tag{112}$$

and, hence, under the assumptions ([32, Theorem 2.5.1])

$$\mathbb{P}\left(\liminf_{i \to \infty} \rho(\mathbf{x}(i), B) = 0\right) = 1 \tag{113}$$

which, together with assumption (25), implies

$$\mathbb{P}\left(\liminf_{i \to \infty} V(i, \mathbf{x}(i)) = 0\right) = 1. \tag{114}$$

Also, it can be shown that the process $\left(W(i, \mathbf{x}(i)), \mathcal{F}_i^{\mathbf{X}}\right)$ is a nonnegative supermartingale ([32, Theorem 2.2.2]) and, hence, converges a.s. to a finite value. It then follows from (111) that $V(i, \mathbf{x}(i))$ also converges a.s. to a finite value. Together with (114), the a.s. convergence of $V(i, \mathbf{x}(i))$ implies

$$\mathbb{P}\left(\lim_{i \to \infty} V(i, \mathbf{x}(i)) = 0\right) = 1. \tag{115}$$

The theorem then follows from assumptions (23) and (24) (see also [32, Theorem 2.7.1].) ∎

Theorem 11: ($\mathcal{A} - \mathcal{ND}$: *Convergence*) Consider the $\mathcal{A} - \mathcal{ND}$ algorithm given in Section IV-A with arbitrary initial state $\mathbf{x}(0) \in \mathbb{R}^{N \times 1}$, under the Assumptions **1.2), 2.2), 3)**. Then

$$\mathbb{P}\left[\lim_{i \to \infty} \rho(\mathbf{x}(i), \mathcal{C}) = 0\right] = 1. \tag{116}$$

Proof: In the $\mathcal{A} - \mathcal{ND}$ (9), the Laplacian $L(i, \mathbf{x}(i))$ and the noise $\mathbf{n}(i, \mathbf{x}(i))$ are both state dependent. We follow Theorem 3 until (38) and modify (39) according to the new assumptions. The sequences $\{L(i, \mathbf{x})\}$ and $\{\mathbf{n}(i, \mathbf{x})\}$ are independent. By the Gershgorin circle theorem, the eigenvalues of $\left(L(i, \mathbf{x}) - \overline{L}\right)$ are less than $2N$ in magnitude. From the noise variance growth condition we have

$$\begin{aligned}
&\mathbb{E}\left[\alpha^2(i)\mathbf{n}^T(i, \mathbf{x})\overline{L}\mathbf{n}(i, \mathbf{x})\right] \\
&= \alpha^2(i)\mathbb{E}\left[\mathbf{n}_{\mathcal{C}^\perp}^T(i, \mathbf{x})\overline{L}\mathbf{n}_{\mathcal{C}^\perp}(i, \mathbf{x})\right] \\
&\le \alpha^2(i)\lambda_N(L)\mathbb{E}\left[\|\mathbf{n}_{\mathcal{C}^\perp}(i, \mathbf{x})\|^2\right] \\
&\le \alpha^2(i)\lambda_N(L)\left[c_1 + c_2\|\mathbf{x}_{\mathcal{C}^\perp}\|^2\right].
\end{aligned} \tag{117}$$

Using (117) and a sequence of steps similar to (39) we have

$$\begin{aligned}
&\mathcal{L}V(i, \mathbf{x}) \\
&= -2\alpha(i)\mathbf{x}^T\overline{L}^2\mathbf{x} + \alpha^2(i)\mathbf{x}^T\overline{L}^3\mathbf{x} \\
&\quad + \mathbb{E}\left[\alpha^2(i)\left(\left(L(i, \mathbf{x}) - \overline{L}\right)\mathbf{x}\right)^T \overline{L}\left(L(i, \mathbf{x}) - \overline{L}\right)\mathbf{x}\right] \\
&\quad + \mathbb{E}\left[\alpha^2(i)\mathbf{n}(i, \mathbf{x})^T\overline{L}\mathbf{n}(i, \mathbf{x})\right]
\end{aligned}$$

$$\begin{aligned}
&\le -2\alpha(i)\mathbf{x}^T\overline{L}^2\mathbf{x} + \alpha^2(i)\lambda_N(\overline{L})^3\|\mathbf{x}_{\mathcal{C}^\perp}\|^2 \\
&\quad + 4\alpha^2(i)N^2\lambda_N(\overline{L})\|\mathbf{x}_{\mathcal{C}^\perp}\|^2 \\
&\quad + \alpha^2(i)\lambda_N(\overline{L})\left[c_1 + c_2\|\mathbf{x}_{\mathcal{C}^\perp}\|^2\right].
\end{aligned}$$

Now, using the fact that $\mathbf{x}^T\overline{L}\mathbf{x} \ge \lambda_2(\overline{L})\|\mathbf{x}_{\mathcal{C}^\perp}\|^2$, we have

$$\begin{aligned}
\mathcal{L}V(i, \mathbf{x}) &\le -2\alpha(i)\mathbf{x}^T\overline{L}^2\mathbf{x} + \alpha^2(i) \\
&\quad \times \left[c_1\lambda_N(\overline{L}) + \frac{\lambda_N^3(\overline{L})}{\lambda_2(\overline{L})}\mathbf{x}^T\overline{L}\mathbf{x} \right. \\
&\qquad \left. + \frac{4N^2\lambda_N(\overline{L})}{\lambda_2(\overline{L})}\mathbf{x}^T\overline{L}\mathbf{x} + \frac{c_2\lambda_N(\overline{L})}{\lambda_2(\overline{L})}\mathbf{x}^T\overline{L}\mathbf{x}\right] \\
&\le -\alpha(i)\varphi(i, \mathbf{x}) + g(i)\left[1 + V(i, \mathbf{x})\right]
\end{aligned}$$

where $\varphi(i, \mathbf{x}) = 2\mathbf{x}^T\overline{L}^2\mathbf{x}, g(i) = \alpha^2(i)\max(c_1\lambda_N(\overline{L}), \lambda_N^3(\overline{L})/\lambda_2(\overline{L}) + 4N^2\lambda_N(\overline{L})/\lambda_2(\overline{L}) + c_2\lambda_N(\overline{L})/\lambda_2)$. It can be verified that $\varphi(i, \mathbf{x})$ and $g(i)$ satisfy the conditions for Theorem 1. Hence, (116). ∎

Theorem 12: Consider the $\mathcal{A} - \mathcal{ND}$ algorithm under the Assumptions **1.2)**, **2.1)**, and **3)**. Then, there exists a.s. a finite real random variable $\boldsymbol{\theta}$ such that

$$\mathbb{P}\left[\lim_{i \to \infty} \mathbf{x}(i) = \boldsymbol{\theta}\mathbf{1}\right] = 1 \tag{118}$$

Proof: Note that $\mathbf{1}^T L(i, \mathbf{x}(i)) = \mathbf{0}, \; \forall i$. The proof then follows from Theorem 4, since the noise assumptions are the same. ∎

Lemma 13: Let θ be as given in Theorem 12 and r, the initial average, as given in (4), under the Assumptions **1.2)**, **2.1)**, and **3)**. Let the mse $\zeta = \mathbb{E}\left[\boldsymbol{\theta} - r\right]^2$. Then we have

1) **Unbiasedness**:

$$\mathbb{E}[\theta] = r. \tag{119}$$

2) **M.S.E. Bound**

$$\zeta \le \frac{\eta}{N^2}\sum_{i \ge 0}\alpha^2(i). \tag{120}$$

Proof: Follows from Theorem 12 and Lemma 5. ∎

It is possible to have results similar to Theorem 12 and Lemma 13 under Assumption **2.2)** on the noise. In that case, we need exact mixing conditions on the sequence. Also, Assumption **2.2)** places no restriction on the growth rate of the variance of the noise component in the consensus subspace. By Theorem 6, we still get a.s. consensus, but the mse may become unbounded, if no growth restrictions are imposed.

REFERENCES

[1] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D., MIT, Cambridge, MA, 1984.

[2] R. O. Saber and R. M. Murray, "Consensus protocols for networks of dynamic agents," in *Proc. Amer. Control Conf.*, Jun. 2003, vol. 2, pp. 951–956.

[3] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. AC-48, pp. 988–1001, Jun. 2003.

[4] C. Reynolds, "Flocks, birds, and schools: A distributed behavioral model," *Comput. Graph.*, vol. 21, pp. 25–34, 1987.

[5] T. Vicsek, A. Czirok, E. B. Jacob, I. Cohen, and O. Schochet, "Novel type of phase transitions in a system of self-driven particles," *Phys. Rev. Lett.*, vol. 75, pp. 1226–1229, 1995.

[6] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *IEEE Proc.*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[7] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Contr. Lett.*, vol. 53, pp. 65–78, 2004.

[8] S. Kar and J. M. F. Moura, "Ramanujan topologies for decision making in sensor networks," in *Proc. 44th Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 2006.

[9] S. Kar and J. M. F. Moura, "Topology for global average consensus," in *Proc. 40th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 2006.

[10] S. Kar, S. A. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2609–2613, Jun. 2008.

[11] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3315–3326, Jul. 2008.

[12] Y. Hatano and M. Mesbahi, "Agreement over random networks," in *Proc. 43rd IEEE Conf. Decision Control*, Dec. 2004, vol. 2, pp. 2010–2015.

[13] M. G. Rabbat, R. D. Nowak, and J. A. Bucklew, "Generalized consensus computation in networked systems with erasure links," in *Proc. 6th Int. Wkshp. Signal Proc. Adv. in Wireless Commun.*, New York, 2005, pp. 1088–1092.

[14] C. Wu, "Synchronization and convergence of linear dynamics in random directed networks," *IEEE Trans. Autom. Control*, vol. 51, pp. 1207–1210, Jul. 2006.

[15] A. T. Salehi and A. Jadbabaie, "On consensus in random networks," in *Proc. Allerton Conf. Commun., Control, Comput.*, Allerton House, IL, Sep. 30, 2006.

[16] M. Porfiri and D. Stilwell, "Stochastic consensus over weighted directed networks," in *Proc. 2007 Amer. Control Conf.*, New York, Jul. 11–13, 2007.

[17] Y. Hatano, A. K. Das, and M. Mesbahi, "Agreement in presence of noise: pseudogradients on random geometric networks," in *Proc. 44th IEEE Conf. Decision Control, and Eur. Control Conf. (CDC-ECC)*, Seville, Spain, Dec. 2005.

[18] M. Huang and J. Manton, "Stochastic Lyapounov analysis for consensus algorithms with noisy measurements," in *Proc. 2007 Amer. Control Conf.*, New York City, Jul. 11–13, 2007.

[19] M. Huang and J. Manton, "Stochastic approximation for consensus seeking: mean square and almost sure convergence," in *IEEE 46th Conf. Decision Control*, New Orleans, LA, Dec. 12–14, 2007.

[20] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus-based distributed parameter estimation in ad hoc wireless sensor networks with noisy links," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Honolulu, HI, 2007, pp. 849–852.

[21] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI: Amer. Math. Soc., 1997.

[22] B. Mohar, "The Laplacian spectrum of graphs," in *Graph Theory, Combinatorics, and Applications*, Y. Alavi, G. Chartr, O. R. Oellermann, and A. J. Schwenk, Eds. New York: Wiley, 1991, vol. 2, pp. 871–898.

[23] B. Bollobás, *Modern Graph Theory*. New York: Springer-Verlag, 1998.

[24] 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.

[25] S. Kar and J. Moura, *Distributed Consensus Algorithms in Sensor Networks: Quantized Data*, Nov. 2007 [Online]. Available: http://arxiv.org/abs/0712.1609, submitted for publication

[26] T. C. Aysal, M. Coates, and M. Rabbat, "Distributed average consensus using probabilistic quantization," in *Proc. IEEE/SP 14th Wkshp. Statistical Signal Process.*, Madison, WI, Aug. 2007, pp. 640–644.

[27] S. Kar and J. M. F. Moura, "Distributed average consensus in sensor networks with random link failures and communication channel noise," in *Proc. 41st Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2007, pp. 676–680.

[28] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Trans. Netw.*, vol. 14, no. SI, pp. 2508–2530, 2006.

[29] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: Algorithms and theory," *IEEE Trans. Autom. Control*, vol. 51, pp. 401–420, 2006.

[30] H. Tanner, A. Jadbabaie, and G. J. Pappas, "Flocking in fixed and switching networks," *IEEE Trans. Autom. Control*, vol. 52, pp. 863–868, 2007.

[31] Y. Kim and M. Mesbahi, "On maximizing the second smallest eigenvalue of a state-dependent graph Laplacian," *IEEE Trans. Autom. Control*, vol. 51, pp. 116–120, Jan. 2006.

[32] M. Nevel'son and R. Has'minskii, *Stochastic Approximation and Recursive Estimation*. Providence, RI: Amer. Math. Soc., 1973.

[33] D. Williams, *Probability With Martingales*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[34] H. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer, 2003.

[35] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2508–2530, Jun. 2006.

[36] D. Mosk-Aoyama and D. Shah, "Computing separable functions via gossip," in *PODC '06: Proc. 25th Ann. ACM Symp. Principles of Distrib. Comput.*, New York, 2006, pp. 113–122.

[37] F. R. K. Chung and L. Lu, *Complex Graphs and Networks*. Boston, MA: Amer. Math. Soc., 2006.

Soummya Kar received the Bachelor of Technology (B.Tech.) degree from the Indian Institute of Technology, Kharagpur, in May, 2005.

He joined the Department of Electrical and Computer Engineering, Carnegie-Mellon University, Pittsburgh, PA, in 2005, where he is currently a third-year Ph.D. degree student. His research interests include distributed computations in sensor networks, stochastic approximations, stochastic networks, and stochastic filtering in the context of sensor networks.

José M. F. Moura (S'71–M'75–SM'90–F'94) received the electrical engineering degree from the Instituto Superior Téecnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and D.Sc. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge.

He is a Professor of Electrical and Computer Engineering and, by courtesy, of BioMedical Engineering, at Carnegie Mellon University (CMU), Pittsburgh, PA. He was on the faculty of IST and has held visiting faculty appointments at MIT and was a research scholar with the University of Southern California, Los Angeles. He is a founding codirector of the Center for Sensed Critical Infrastructures Research (CenSCIR) and of the Information and Communications Technologies Institute, a joint venture between CMU and Portugal that manages a large education and research program between CMU and Portugal (www.icti.cmu.edu). His research interests include statistical and algebraic signal processing, image, bioimaging and video processing, and digital communications. He has published more than 300 technical journal and conference papers, is the coeditor of two books, holds seven patents on image and video processing, and digital communications with the U.S. Patent Office, and has given numerous invited seminars at U.S. and European Universities and industrial and government Laboratories.

Dr. Moura is the President (2008–2009) of the IEEE Signal Processing Society (SPS). He was Editor-in-Chief for the IEEE TRANSACTIONS IN SIGNAL PROCESSING, interim Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS, and was on the Editorial Board of several journals, including the IEEE PROCEEDINGS, the IEEE *Signal Processing Magazine*, and the ACM TRANSACTIONS ON SENSOR NETWORKS. He was Vice-President for Publications for the IEEE Sensors Council. He chaired the IEEE TAB Transactions Committee, served on the IEEE TAB Periodicals Review Committee, and is currently Vice-Chair of the IEEE Publication Services and Products Board. He was on the steering and technical committees of several Conferences and on the IEEE Press Board. He is a Fellow of the American Association for the Advancement of Science (AAAS), and a corresponding member of the Academy of Sciences of Portugal (Section of Sciences). He was awarded the 2003 IEEE Signal Processing Society Meritorious Service Award and in 2000, the IEEE Millennium Medal. In 2007, he received the CMU's College of Engineering Outstanding Research Award. He is affiliated with several IEEE societies, Sigma Xi, AMS, IMS, and SIAM.