

# Convergence of Consensus Models With Stochastic Disturbances

Tuncer Can Aysal, *Member, IEEE*, and Kenneth E. Barner, *Senior Member, IEEE*

**Abstract**—We consider consensus algorithms in their most general setting and provide conditions under which such algorithms are guaranteed to converge, almost surely, to a consensus. Let  $\{A(t), B(t)\} \in \mathbb{R}^{N \times N}$  be (possibly) stochastic, nonstationary matrices and  $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  be state and perturbation vectors, respectively. For any consensus algorithm of the form  $x(t+1) = A(t)x(t) + B(t)m(t)$ , we provide conditions under which consensus is achieved almost surely, i.e.,  $\Pr\{\lim_{t \rightarrow \infty} x(t) = c\mathbf{1}\} = 1$  for some  $c \in \mathbb{R}$ . Moreover, we show that this general result subsumes recently reported results for specific consensus algorithm classes, including sum-preserving, nonsum-preserving, quantized, and noisy gossip algorithms. Also provided are the  $\epsilon$ -converging time for any such converging iterative algorithm, i.e., the earliest time at which the vector  $x(t)$  is  $\epsilon$  close to consensus, and sufficient conditions for convergence in expectation to the average of the initial node measurements. Finally, mean square error bounds of any consensus algorithm of the form discussed above are presented.

**Index Terms**—Convergence in expectation, convergence of random sequences, convergence to consensus, distributed average consensus, gossip algorithms, mean square error, noisy gossip algorithms, nonsum-preserving gossip algorithms, quantized gossip algorithms, sum-preserving gossip algorithms.

## I. INTRODUCTION

A FUNDAMENTAL problem in decentralized networked systems is that of having nodes reach a state of agreement [1]–[7]. Distributed agreement is a fundamental problem in *ad hoc* network applications, including distributed consensus and synchronization problems [4]–[6], [8], distributed coordination of mobile autonomous agents [2], [3], and distributed data fusion in sensor networks [1], [7], [9]. It is also a central topic for load balancing (with divisible tasks) in parallel computers [10].

Vicsek *et al.* provided a variety of simulation results which demonstrate that simple distributed algorithms allow all nodes to eventually agree on a parameter [4]. The work in [11] provided the theoretical explanation for behavior observed in these reported simulation studies. This paper focuses on a prototypical example of agreement in asynchronous networked systems,

Manuscript received July 25, 2008; revised December 02, 2009. Date of current version July 14, 2010. This work was supported in part by the National Science Foundation under award #0728904.

T. C. Aysal is with the Communications Research in Signal Processing Group, School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: tca27@cornell.edu).

K. E. Barner is with the Signal Processing and Communications Group, Electrical and Computer Engineering Department, University of Delaware, Newark, DE 19716 USA (e-mail: barner@udel.edu).

Communicated by H.-A. Loeliger, Associate Editor for Coding Techniques. Digital Object Identifier 10.1109/TIT.2010.2050940

namely, the randomized average consensus problem in a communication network.

### A. Average Consensus

Distributed averaging algorithms are extremely attractive for applications in networked systems because nodes maintain simple state information and exchange information with only their immediate neighbors. Consequently, there is no need to establish or maintain complicated routing structures. Also, there are no bottleneck links (as in tree or ring structures) where the result of in-network computations can be compromised, lost, or jammed by an adversary. Finally, consensus algorithms have the attractive property that, at termination, the computed value is available throughout the network, enabling a network user to query any node and immediately receive a response, rather than waiting for the query and response to propagate to and from a fusion center.

Gossip-based average consensus algorithms were initially introduced in 1984 by Tsitsiklis [12] to achieve consensus over a set of agents, with the approach receiving considerable recent attention from other researchers [1]–[3], [13]–[17]. The problems setup stipulates that, at time slot  $t \geq 0$ , each node  $i = 1, 2, \dots, N$  has an estimate  $x_i(t)$  of the global average, where  $x(t)$  denotes the  $N$ -vector of these estimates. The ultimate goal is to drive the estimate  $x(t)$  to, or as close as possible to, the average vector  $\bar{x}(0)\mathbf{1}$  using minimal amount of communications. In this notation,  $\mathbf{1}$  denotes the vector of ones and

$$\bar{x}(0) = \frac{1}{N} \sum_{i=1}^N x_i(0). \quad (1)$$

Notably, the quantity  $x(t)$  for  $t > 0$  is a random vector since the algorithms are stochastic in their behavior.

In the following, we discuss specific cases of related work that are subsumed by the general theoretical approach presented subsequently.

**Sum-Preserving Gossip:** Randomized average consensus gossiping uses an asynchronous time model wherein a node chosen at random (uniformly) wakes up, contacts a random neighbor within its *connectivity radius*, and exchanges values [2], [3], [13], [18]. The two nodes then update their values with the pairwise averages of their values. This operation preserves the nodal total sum, and, hence, also the mean. The algorithm converges to a consensus if the graph is, on the average, strongly connected. Because the transmitting node must send a packet to the chosen neighbor and then wait for the neighbor's packet, this scheme is vulnerable to packet collisions and yields a communication complexity (measured by number of radio transmissions required to drive the estimation error to within

$\Theta(N^{-\alpha})$ , for any  $\alpha > 0$ ) on the order of  $\Theta(N^2 \log N)$  for random geometric graphs [13]. In synchronous time model average consensus, nodes exchange values with all their neighbors simultaneously and update their state values accordingly. This model yields a faster convergence rate compared to the asynchronous case, but introduces congestion and collision issues [13], [18].

The recently proposed geographic gossip algorithm combines gossip with geographic routing [19]. Similarly to the standard gossip algorithm, a node randomly wakes up, but, in this case, chooses a random node within the *whole network*, rather than simply in its *neighborhood*, and performs pairwise averaging with the selected node. Geographic gossiping increases the diversity of every pairwise averaging operation. Moreover, the algorithm communication complexity is of the order  $O(N^{3/2} \sqrt{\log(N)})$ , which is an improvement with respect to the standard gossiping algorithm. More recently, a variety of the algorithm that “averages around the way” has been shown to converge in  $O(N \log N)$  transmissions [20].

**Nonsum-Preserving Gossip:** To overcome the drawbacks of the standard packet based sum-preserving gossip algorithms, broadcast gossip algorithms suitable for wireless sensor networks were recently proposed [17], [21]. Under this methodology, a node in the network wakes up randomly according to an asynchronous time model and broadcasts its value. This value is successfully received by the nodes in the connectivity radius of the broadcasting node. The nodes that received the broadcasted value update their state values, in a convex fashion, while the remaining nodes sustain their values.

By iterating the broadcast gossip protocol, the algorithm achieves consensus (with probability one) over the network, with the consensus valued neighboring the initial network sensor reading average. Moreover, although the convergence time of the algorithm is commensurate with the standard pairwise sum-preserving gossip algorithms, it is shown that, for modest network sizes, the broadcast gossip algorithm converges to consensus faster, and with fewer radio transmissions, than algorithms based on pairwise averages or routing [17], [21].

Also of note is that, when the number of nodes is large with respect to time, the convergence rate of algorithms that achieve consensus but do not necessarily converge to the initial average (e.g., asymmetric gossip, broadcast gossip, and packet-drop gossip), it is argued in [22], is better described by mean square analysis, while, when the number of nodes is small with respect to time, it is the Lyapunov exponent analysis that provides the appropriate convergence description [22].

**Quantized Gossip:** All algorithms discussed above assume internode communications are achieved with infinite precision. This assumption is clearly not satisfied in practice. Thus, recent interest in gossip algorithms has focused on node communications using quantized values [15], [16], [23]–[26]. While there are quantized algorithms that yield node state values as-close-as possible to each other, but that do not achieve consensus in the strict sense [26], we review only representative algorithms that strictly achieve consensus.

Aysal *et al.* consider a model in which the nodes utilize a probabilistic quantizer (i.e., dithering quantization) prior to transmission. The quantized consensus iterations, in this

approach, can be viewed as an absorbing Markov chain, with the absorbing states given by the quantization points [16], [23]. Similarly, Nedic *et al.* use a floor quantizer and show, using Lyapunov function analysis (as they approach the problem from a control theory perspective), that: (i) the state variances diminish to zero and (ii) all nodes converge to a (quantized value) consensus [15]. Finally, Yildiz and Scaglione use coding algorithms in order to reduce the quantization errors to zero, thereby achieving state consensus on a quantization value [25].

**Noisy Gossip:** Xiao, Boyd, and Kim extended the standard distributed consensus algorithm to admit noisy updates, deriving a method in which each node updates its local variable with a weighted average of neighboring values, where each new value is corrupted by zero mean, *fixed* variance, additive noise [27]. Accordingly, weight design procedures are proposed that lead to optimal steady-state behavior, based on the assumption that the noise terms are independent. The resulting algorithm processes through a random walk in which only the variation amongst the nodes converges to a steady state [27].

Hatano *et al.* [28], followed subsequently by Kar and Moura [29] and Rajagopal and Wainwright [30], consider a synchronous and nonrandom distributed agreement problem over realizations of a (Poisson) random geometric network with noisy interconnections. The noise is assumed independent, uncorrelated and Gaussian distributed, with zero mean and fixed variance. Sufficiency conditions for single parameter consensus are presented, albeit for the particular adaptive algorithm considered. That is, general conditions for almost sure convergence to consensus are not provided, nor are generic convergence rate and mean square error results [28]–[30].

The maximum likelihood (ML) estimate of initial observations, obtained through a decentralized convex optimization algorithm, is also considered in the literature [31]. Although the authors of this work do not specifically design their algorithm considering noisy links, they argue that their approach is robust to noise components for bounded noise covariance matrix cases. Unlike the previous method, however, the algorithm does not converge to a consensus when noisy links are considered.

## B. Main Contributions

The almost sure convergence results with regard to the algorithms discussed above are generally specific to the algorithm utilized in particular works. Thus, current almost sure convergence results reported in the literature fall short in terms of explaining the underlying true mechanics of the consensus framework. There are no “universal” almost sure convergence results that would allow researchers and engineers to assess the characteristics of a newly designed consensus system. This lack of knowledge forces researchers to investigate algorithm or problem-specific conditions. In some cases, extensive simulations alone are utilized to support the convergence of an algorithm. Also problematic is the fact that the standard system stationarity assumption excludes cases such as addition/extraction of nodes in the system that are crucial to current networking applications. Future consensus systems and algorithms, possibly not characterized by existing models, may be devised in the future, thereby requiring powerful analysis tools to identify their convergence properties.

Accordingly, this work studies consensus algorithms in their most general setting and provides conditions under which such algorithms are guaranteed to converge, almost surely, to a consensus, and addresses the discussed drawbacks of the algorithms, algorithm class and their explicit analysis. Specifically, we consider any consensus algorithm of the form

$$x(t+1) = A(t)x(t) + B(t)m(t) \quad (2)$$

where

- $\{A(t), B(t)\} \in \mathbb{R}^{N \times N}$  are (possibly) nonstationary random update and control matrices, respectively,
- $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  are state and perturbation vectors, respectively.

We provide conditions under which any such algorithm achieves consensus almost surely, i.e.,

$$\Pr \left\{ \lim_{t \rightarrow \infty} x(t) = c\mathbf{1} \right\} = 1 \quad (3)$$

for some  $c \in \mathbb{R}$ . We also make the following assumptions throughout the paper: 1)  $\{x(s) : s \leq t\}$  is independent from  $A(t)$ ,  $B(t)$  and  $m(t)$  for all  $t \geq 0$ , and 2)  $m(t)$  is independent from  $B(t)$  for all  $t \geq 0$ .

Moreover, we show that this general result subsumes recently reported results for specific consensus algorithm classes, including sum-preserving, nonsum-preserving, quantized, and noisy gossip algorithms. Also provided are the  $\epsilon$ -converging time for any such converging iterative algorithm, i.e., the earliest time at which the vector  $x(t)$  is  $\epsilon$  close to consensus, and sufficient conditions for convergence in expectation to the initial node measurements average. Finally, error bounds characterizing the mean square error performance of any consensus algorithm of the form discussed in this paper are given.

### C. Paper Organization

General consensus algorithm formulation is provided in Section II, along with the main result of the paper, i.e., sufficient conditions guaranteeing convergence to consensus for any consensus algorithm of the form considered in this paper. Moreover, the convergence rate to consensus of such algorithms are also detailed in this section. Section III details the conditions required to achieve consensus in expectation to the desired value for any consensus algorithm. Mean square error bounds of any consensus algorithm of the form discussed above are presented in Section IV. Finally, we conclude with Section V.

## II. CONSENSUS REVEALED: CONVERGENCE TO CONSENSUS

In this section, we consider consensus algorithms in their most general setting and provide conditions under which the algorithms are guaranteed to converge to a consensus almost surely. Moreover, we show how this result nicely subsumes the results corresponding to consensus algorithms reported in the literature, i.e., sum-preserving consensus algorithms (such as randomized and geographic gossip algorithms), nonsum-preserving algorithms (such as broadcast gossip algorithms), quantized gossip algorithms and noisy gossip algorithms.

### A. Convergence to Consensus

Let  $J = 1/N\mathbf{1}\mathbf{1}^T$ . We begin by stating one of the main results of this paper.

*Theorem 1: Let  $\{A(t), B(t)\} \in \mathbb{R}^{N \times N}$  be (possibly) nonstationary random matrices and  $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  be state and perturbation vectors, respectively. Consider any consensus algorithm of the form*

$$x(t+1) = A(t)x(t) + B(t)m(t). \quad (4)$$

Now suppose that

$$\begin{aligned} A(t)\mathbf{1} &= \mathbf{1}, \mathbb{E}\{m(t)\} = \mathbf{0} \\ \mathbb{E}\{m(t) | x(t), A(t), B(t)\} &= \mathbb{E}\{m(t)\}, \forall t \end{aligned} \quad (5)$$

and

$$\begin{aligned} 0 \leq \lambda_1[\mathbb{E}\{A(t)^T(I - J)A(t)\}] &\leq 1, \forall t, \\ \sum_{t=0}^{\infty} 1 - \lambda_1[\mathbb{E}\{A(t)^T(I - J)A(t)\}] &= \infty \quad (6) \\ \sum_{t=0}^{\infty} \lambda_1[\mathbb{E}\{B(t)^T(I - J)B(t)\}] \mathbb{E}\{\|m(t)\|_2^2\} &< \infty \\ \lim_{t \rightarrow \infty} \frac{\lambda_1[\mathbb{E}\{B(t)^T(I - J)B(t)\}] \mathbb{E}\{\|m(t)\|_2^2\}}{1 - \lambda_1[\mathbb{E}\{A(t)^T(I - J)A(t)\}]} &= 0 \quad (7) \end{aligned}$$

where  $\lambda_1[\cdot]$  denotes the largest eigenvalue of its argument. Then any algorithm of the form (4) converges to a consensus almost surely, i.e.,

$$\Pr \left\{ \lim_{t \rightarrow \infty} x(t) = c\mathbf{1} \right\} = 1 \quad (8)$$

for some  $c \in \mathbb{R}$ .

*Proof:* Consider the deviation vector  $v(t) = x(t) - Jx(t)$  and resulting recursion

$$v(t+1) = (A(t) - JA(t))v(t) + (B(t) - JB(t))m(t). \quad (9)$$

Using the fact that  $A(t)\mathbf{1} = \mathbf{1}$ ,  $\forall t$ , we have

$$\begin{aligned} (A(t) - JA(t))(-Jx(t)) &= -A(t)Jx(t) + JA(t)Jx(t) \\ &= -Jx(t) + Jx(t) = 0 \end{aligned} \quad (10)$$

indicating that

$$v(t+1) = (A(t) - JA(t))v(t) + (B(t) - JB(t))m(t). \quad (11)$$

Now, letting  $V(t) = \|v(t)\|_2^2$ , we obtain

$$\begin{aligned} V(t+1) &= \|(A(t) - JA(t))v(t)\|_2^2 \\ &\quad + \|(B(t) - JB(t))m(t)\|_2^2 \\ &\quad + 2v(t)^T(A(t) - JA(t))^T(B(t) - JB(t))m(t). \end{aligned} \quad (12)$$

Taking the expectation of the above, given  $\{V(s) : s \leq t\}$  and using the fact that  $m(t)$  is independent from  $A(t)$ ,  $B(t)$  and  $x(t)$ , yields

$$\begin{aligned} \mathbb{E}\{V(t+1) | V(s) : s \leq t\} &= \mathbb{E}\{\|(A(t) - JA(t))v(t)\|_2^2 | V(s) : s \leq t\} \\ &\quad + \mathbb{E}\{\|(B(t) - JB(t))m(t)\|_2^2\} \end{aligned} \quad (13)$$

since  $\mathbb{E}\{m(t) | v(t), A(t), B(t)\} = \mathbb{E}\{m(t)\}$  and  $\mathbb{E}\{m(t)\} = \mathbf{0}, \forall t$ . Moreover, expanding the norms and noting that  $m(t)$  and  $B(t)$  are independent with  $v(t)$  for all  $t \geq 0$ , we get

$$\begin{aligned} & \mathbb{E}\{V(t+1) | V(s) : s \leq t\} \\ &= \mathbb{E}\{v(t)^T(A(t) - JA(t))^T(A(t) \\ &\quad - JA(t))v(t) | V(s) : s \leq t\} \\ &\quad + \mathbb{E}\{m(t)^T \\ &\quad \times (B(t) - JB(t))^T(B(t) - JB(t))m(t)\} \quad (14) \end{aligned}$$

$$\begin{aligned} &=^{(a)} \mathbb{E}\{v(t)^T \mathbb{E}\{(A(t) - JA(t))^T(A(t) \\ &\quad - JA(t))\}v(t) | V(s) : s \leq t\} \\ &\quad + \mathbb{E}\{m(t)^T \mathbb{E}\{(B(t) - JB(t))^T(B(t) \\ &\quad - JB(t))\}m(t)\} \quad (15) \end{aligned}$$

$$\begin{aligned} &\leq^{(b)} \lambda_1 [\mathbb{E}\{(A(t) - JA(t))^T(A(t) \\ &\quad - JA(t))\}] \mathbb{E}\{V(t) | V(s) : s \leq t\} \\ &\quad + \lambda_1 [\mathbb{E}\{(B(t) - JB(t))^T(B(t) \\ &\quad - JB(t))\}] \mathbb{E}\{M(t)\} \quad (16) \end{aligned}$$

$$\begin{aligned} &=^{(c)} \lambda_1 [\mathbb{E}\{A(t)^T(I - J)A(t)\}]V(t) \\ &\quad + \lambda_1 [\mathbb{E}\{B(t)^T(I - J)B(t)\}] \mathbb{E}\{M(t)\} \quad (17) \end{aligned}$$

$$=^{(d)} \rho_A(t)V(t) + \rho_B(t)\mathbb{E}\{M(t)\} \quad (18)$$

where (a) follows by conditioning, (b) is due to Rayleigh-Ritz theorem, (c) is seen by noting that  $I - J$  is a projection matrix and denoting  $M(t) = \|m(t)\|_2^2$  and (d) follows by the fact that  $\mathbb{E}\{V(t) | V(s) : s \leq t\} = V(t)$  and the notation:  $\rho_A(t) \triangleq \lambda_1[\mathbb{E}\{A(t)^T(I - J)A(t)\}]$  and  $\rho_B(t) \triangleq \lambda_1[\mathbb{E}\{B(t)^T(I - J)B(t)\}]$ .

We will make use of the following Lemma to finish up our proof.

*Lemma 1:* [32] Consider a sequence of nonnegative random variables  $\{V(t)\}_{t \geq 0}$  with  $\mathbb{E}\{V(0)\} < \infty$ . Let

$$\mathbb{E}\{V(t+1) | V(t), \dots, V(1), V(0)\} \leq (1 - c_1(t))V(t) + c_2(t) \quad (19)$$

where

$$\begin{aligned} &0 \leq c_1(t) \leq 1, c_2(t) \geq 0, \forall t \\ &\sum_{t=0}^{\infty} c_2(t) < \infty, \sum_{t=0}^{\infty} c_1(t) = \infty \\ &\lim_{t \rightarrow \infty} \frac{c_2(t)}{c_1(t)} = 0. \quad (20) \end{aligned}$$

Then,  $V(t)$  almost surely converges to zero, i.e.,

$$\Pr \left\{ \lim_{t \rightarrow \infty} V(t) = 0 \right\} = 1. \quad (21)$$

In the following, we will fit our development to the above Lemma. Note the following items:

- $V(t) = \|v(t)\|_2^2 \geq 0$  is nonnegative for all  $t$  and  $\mathbb{E}\{V(0)\} = V(0) < \infty$ .
- Let  $c_2(t) = \rho_B(t)\mathbb{E}\{M(t)\}$  and note that  $c_2(t) \geq 0$  for all  $t$  since  $\rho_B(t) \geq 0$  and  $\mathbb{E}\{M(t)\} = \mathbb{E}\{\|m(t)\|_2^2\} \geq 0$  for all  $t$ .
- Let  $c_1(t) = 1 - \rho_A(t)$ .

Finally, using the next lemma from [21], [28] indicating the following.

*Lemma 2:* [21], [28] Let  $V(t) = \|x(t) - Jx(t)\|_2^2$ . Then

$$\Pr \left\{ \lim_{t \rightarrow \infty} V(t) = 0 \right\} = 1 \quad (22)$$

if and only if

$$\Pr \left\{ \lim_{t \rightarrow \infty} x(t) = c\mathbf{1} \right\} = 1 \quad (23)$$

for some  $c \in \mathbb{R}$ .

completes the proof of our theorem.  $\blacksquare$

The vast majority of recent literature in the randomized and deterministic gossip fields uses *stationary* update matrices. The following corollary gives the special case of Theorem 1 for this subclass of consensus algorithms.

*Corollary 1:* Let  $\{A(t), B(t)\} \in \mathbb{R}^{N \times N}$  be stationary random matrices and  $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  be state and perturbation vectors, respectively. Consider any consensus algorithm of the form

$$x(t+1) = A(t)x(t) + B(t)m(t). \quad (24)$$

Now suppose that

$$\begin{aligned} &A(t)\mathbf{1} = \mathbf{1}, \mathbb{E}\{m(t)\} = \mathbf{0}, \\ &\mathbb{E}\{m(t) | x(t), A(t), B(t)\} = \mathbb{E}\{m(t)\}, \forall t \quad (25) \end{aligned}$$

and

$$\begin{aligned} &0 \leq \lambda_1[\mathbb{E}\{A^T(I - J)A\}] < 1 \\ &\lambda_1[\mathbb{E}\{B^T(I - J)B\}] \sum_{t=0}^{\infty} \mathbb{E}\{\|m(t)\|_2^2\} < \infty \quad (26) \end{aligned}$$

$$\frac{\lambda_1[\mathbb{E}\{B^T(I - J)B\}]}{1 - \lambda_1[\mathbb{E}\{A^T(I - J)A\}]} \lim_{t \rightarrow \infty} \mathbb{E}\{\|m(t)\|_2^2\} = 0 \quad (27)$$

where  $\lambda_1[\cdot]$  denotes the largest eigenvalue of its argument and we denote  $\lambda_1[\mathbb{E}\{A^T(I - J)A\}] = \lambda_1[\mathbb{E}\{A(t)^T(I - J)A(t)\}]$  and  $\lambda_1[\mathbb{E}\{B^T(I - J)B\}] = \lambda_1[\mathbb{E}\{B(t)^T(I - J)B(t)\}]$  for all  $t \geq 0$ . Then, the algorithm converges to a consensus almost surely, i.e.,

$$\Pr \left\{ \lim_{t \rightarrow \infty} x(t) = c\mathbf{1} \right\} = 1 \quad (28)$$

for some  $c \in \mathbb{R}$ .

*Proof:* Recall the following two conditions of Theorem 1, considered for stationary matrices:

$$\begin{aligned} &0 \leq \lambda_1[\mathbb{E}\{A^T(I - J)A\}] \leq 1, \\ &\sum_{t=0}^{\infty} 1 - \lambda_1[\mathbb{E}\{A^T(I - J)A\}] = \infty. \quad (29) \end{aligned}$$

Note that the second condition is always satisfied as long as  $\lambda_1[\mathbb{E}\{A^T(I - J)A\}] < 1$ . Thus, we can fuse these two conditions into the one stated in the corollary:  $0 \leq \lambda_1[\mathbb{E}\{A^T(I - J)A\}] < 1$ . The remaining proof and conditions follow directly from Theorem 1.  $\blacksquare$

The above corollary reveals an important fact. Namely, that stationary consensus algorithms might not converge to consensus unless the perturbations is somehow driven to zero.

To see this, consider nonzero, finite  $\lambda_1[\mathbb{E}\{A^T(I - J)A\}]$  and  $\lambda_1[\mathbb{E}\{B^T(I - J)B\}]$ . The only way to satisfy (26) and (27) is to drive the perturbation norm to zero. This is in fact possible with clever algorithms, as we show further in the paper. In the quantized gossip algorithms, for instance, some authors use coding algorithms or probabilistic quantizers, thereby (unknowingly) achieving the task of driving quantization noise variance (perturbations) to zero.

1) *Sum-Preserving Gossip Algorithms:* In the following, we show that the theorem presented in this work reduces to that presented in sum-preserving gossip algorithms, such as randomized and geographic gossip algorithms [13], [18], [19], [27], [33]. The network-wide update, in this case, is given by

$$x(t + 1) = A(t)x(t) \quad (30)$$

where  $A(t)$  is the random and doubly stochastic (for all  $t \geq 0$ ), but *stationary*, weight matrix. Of note is that we consider the asynchronous case where  $A(t)$  is random. However, one can easily consider the synchronous case where  $A(t) = A$  for all  $t \geq 0$ , with the analysis following similarly to Theorem 1, which makes no assumptions on the time model.

The update equation in (30) is clearly subsumed by the model considered in the Corollary 1, which reduces to the former when  $B(t) = 0$  or  $m(t) = 0$  for all  $t \geq 0$  and is  $A(t)$  random but stationary. In this case, the Corollary conditions given in (27) are automatically satisfied since  $\lambda_1[\mathbb{E}\{B(t)^T(I - J)B(t)\}] = 0$  if  $B(t) = 0$ , or  $\mathbb{E}\{\|m(t)\|_2^2\} = 0$  if  $m(t) = 0$ , for all  $t \geq 0$ . The set of conditions in (25) is also satisfied since  $A(t)$  is doubly stochastic. Moreover, since the algorithm is sum-preserving, i.e.,  $\mathbf{1}^T A(t) = \mathbf{1}^T$  for all  $t \geq 0$ , the condition in (26) reduces to

$$\begin{aligned} \lambda_1[\mathbb{E}\{A^T(I - J)A\}] &= \lambda_1[\mathbb{E}\{A^T A - A^T J A\}] \\ &= \lambda_1[\mathbb{E}\{A - J\}] = \lambda_2(\mathbb{E}\{A\}) < 1 \end{aligned} \quad (31)$$

where the second equality follows from the facts that  $\mathbb{E}\{A^T A\} = \mathbb{E}\{A\}$  [13] and  $N^{-1}A^T \mathbf{1} \mathbf{1}^T A = N^{-1} \mathbf{1} \mathbf{1}^T$  since  $A(t)$  is doubly stochastic for all  $t \geq 0$  and taken to be arising from pairwise averaging algorithms. Thus, in addition to double stochasticity of all  $A(t)$ , we need to have that  $\lambda_2(\mathbb{E}\{A\}) < 1$ . This is indeed the convergence condition given in sum-preserving average consensus algorithms [13], [18], [19], [27], [33].

2) *Nonsum-Preserving Gossip Algorithms:* This section considers algorithms for which the network-wide sum is not preserved through iterations. A number of such gossip algorithms have recently been proposed, e.g., the broadcast gossip algorithm [17], [21], [22]. These algorithms provide faster convergence and require a smaller number of radio transmissions to achieve consensus, with the tradeoff being that they converge to a neighborhood of the average rather than strictly average.

The network-wide update, in this case, is also given as  $x(t + 1) = A(t)x(t)$  where, this time,  $A(t)$  is random and stationary, but not doubly stochastic for all  $t \geq 0$ . Through analysis similar to that above, it can be proven that Corollary 1 reduces to (with the same notation as above)

$$\lambda_1[\mathbb{E}\{A^T(I - J)A\}] < 1. \quad (32)$$

This is indeed the condition given in [17], [21], and [22] guaranteeing convergence of form (30), nonsum-preserving consensus algorithms.

3) *Quantized Gossip Algorithms:* All algorithms discussed above assume internode communications are carried out with infinite precision. This assumption is clearly violated in practice. Thus, recent efforts have focused on gossip algorithms that communicate using quantized values [15], [16], [23]–[26].

To consider the quantized case, let  $A(t)$  again be doubly stochastic for all  $t \geq 0$ . Then the network-wide update with quantized values is given by [16], [25]

$$\begin{aligned} x(t + 1) &= A(t)q(t) = A(t)Q[x(t)] \\ &= A(t)x(t) + A(t)m(t) \end{aligned} \quad (33)$$

where  $Q[\cdot]$ ,  $q(t)$  and  $m(t)$  denote any quantizer, quantized sample, and the quantization noise, respectively. Under mild conditions on the signal and quantization parameters, Schumann shows that the quantization noise samples are zero-mean and statistically independent amongst each other, and from the signal [34], [35]. Utilizing dithering also leads to these conditions and the quantized consensus model. Specifically, Schumann proves that the subtractive dithering process utilizing uniform random variable with support on  $[-\Delta/2, \Delta/2]$ -where  $\Delta$  is the quantization bin size-yields error signal values that are statistically independent from each other and the input [34], [35]. Of importance is that both infinite precision and quantized gossip algorithms reported in the literature employ stationary weight matrices.

Unfortunately, most convergence to consensus proofs in the quantized consensus field are algorithm and quantizer specific, and thus do not yield insight on the mechanics of quantized consensus systems [15], [16], [23]–[26]. In this section, we utilized Theorem 1 and Corollary 1 to give convergence conditions for such systems that generalize and subsume previous convergence proofs in this field. Of note is that the following corollaries, i.e., Corollary 2 and 3 consider quantization noise vector samples are independent from each other and the state vector, e.g., a quantized consensus system with dithered quantization.

*Corollary 2:* Let  $\{A(t)\} \in \mathbb{R}^{N \times N}$  be (possibly) nonstationary random matrix and  $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  be state and quantization noise vectors, respectively. Consider any quantized consensus algorithm of the form

$$x(t + 1) = A(t)x(t) + A(t)m(t). \quad (34)$$

Now suppose that  $A(t)\mathbf{1} = \mathbf{1}$  and

$$\begin{aligned} 0 \leq \lambda_1[\mathbb{E}\{A(t)^T(I - J)A(t)\}] &\leq 1, \forall t, \\ \sum_{t=0}^{\infty} 1 - \lambda_1[\mathbb{E}\{A(t)^T(I - J)A(t)\}] &= \infty \end{aligned} \quad (35)$$

$$\begin{aligned} \sum_{t=0}^{\infty} \lambda_1[\mathbb{E}\{A(t)^T(I - J)A(t)\}]\mathbb{E}\{\|m(t)\|_2^2\} &< \infty \\ \lim_{t \rightarrow \infty} \frac{\lambda_1[\mathbb{E}\{A(t)^T(I - J)A(t)\}]\mathbb{E}\{\|m(t)\|_2^2\}}{1 - \lambda_1[\mathbb{E}\{A(t)^T(I - J)A(t)\}]} &= 0 \end{aligned} \quad (36)$$

where  $\lambda_1[\cdot]$  denotes the largest eigenvalue of its argument. Then any quantized algorithm of the form (34) converges to a consensus almost surely, i.e.,

$$\Pr \left\{ \lim_{t \rightarrow \infty} x(t) = c\mathbf{1} \right\} = 1 \quad (37)$$

for some  $c \in \mathbb{R}$ .

*Proof:* The result simply follows by taking  $B(t) = A(t)$  in Theorem 1 and noting that the theorem conditions on  $m(t)$  are met by quantization procedures utilizing dither and satisfying Schumann's conditions. ■

The vast majority of recent literature in the randomized and deterministic quantized gossip fields uses *stationary* update matrices. The following corollary gives the special case of Theorem 1 for this subclass of quantized consensus algorithms.

*Corollary 3:* Let  $\{A(t)\} \in \mathbb{R}^{N \times N}$  be a nonstationary deterministic matrix and  $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  be state and quantization noise vectors, respectively. Consider any quantized consensus algorithm of the form

$$x(t+1) = Ax(t) + Am(t). \quad (38)$$

Now suppose that  $A\mathbf{1} = \mathbf{1}$  and

$$\begin{aligned} 0 \leq \lambda_1[\mathbb{E}\{A^T(I-J)A\}] < 1 \\ \sum_{t=0}^{\infty} \mathbb{E}\{\|m(t)\|_2^2\} < \infty \\ \lim_{t \rightarrow \infty} \mathbb{E}\{\|m(t)\|_2^2\} = 0 \end{aligned} \quad (39)$$

where  $\lambda_1[\cdot]$  denotes the largest eigenvalue of its argument. Then any quantized algorithm of the form (38) converges to a consensus almost surely, i.e.,

$$\Pr\left\{\lim_{t \rightarrow \infty} x(t) = c\mathbf{1}\right\} = 1 \quad (40)$$

for some  $c \in \mathbb{R}$ .

*Proof:* Clearly, the set of corollary conditions in (25) are met by quantization procedures utilizing dither and satisfying Schumann's conditions. Taking  $B(t) = A(t)$ , the remaining conditions of Corollary 1 reduce to

$$\begin{aligned} 0 \leq \lambda_1[\mathbb{E}\{A^T(I-J)A\}] < 1 \\ \lambda_1[\mathbb{E}\{A^T(I-J)A\}] \sum_{t=0}^{\infty} \mathbb{E}\{\|m(t)\|_2^2\} < \infty \end{aligned} \quad (41)$$

and

$$\frac{\lambda_1[\mathbb{E}\{A^T(I-J)A\}]}{1 - \lambda_1[\mathbb{E}\{A^T(I-J)A\}]} \lim_{t \rightarrow \infty} \mathbb{E}\{\|m(t)\|_2^2\} = 0. \quad (42)$$

Omitting the trivial case  $\lambda_1[\mathbb{E}\{A^T(I-J)A\}] = 0$  (which occurs when the graph is superconnected—all nodes are connected to all nodes) further reduces the above set of conditions to

$$\begin{aligned} 0 \leq \lambda_1[\mathbb{E}\{A^T(I-J)A\}] < 1, \\ \sum_{t=0}^{\infty} \mathbb{E}\{\|m(t)\|_2^2\} < \infty, \lim_{t \rightarrow \infty} \mathbb{E}\{\|m(t)\|_2^2\} = 0. \end{aligned} \quad (43)$$

This concludes the proof. ■

Interestingly, this result states that in addition to standard assumptions on the weight update matrix  $A(t)$ , convergent and bounded quantization noise variances are required. This corroborates the individual proofs provided in the quantized consensus literature where, for instance, Aysal *et al.* show that

the quantized consensus iterations can be viewed as an absorbing Markov chain and that the absorbing states are given by the quantization points [16], [23], [24]. The absorbing Markov chain requirements indeed give convergent quantization noise variances, as the above results require. Similarly, Nedic *et al.* use a floor quantizer and show, employing Lyapunov function analysis (approaching the problem from a control theory perspective), that the state variances diminishes to zero and all nodes converge to consensus on a quantization value [15]. This again yields quantization error series that converge to zero. Finally, Yildiz and Scaglione use coding algorithms in order to bring all the node state values closer to a quantization value, effectively trying to reduce the quantization noise variances to zero [25].

4) *Noisy Gossip Algorithms:* Xiao, Boyd, and Kim extended the distributed consensus algorithm to admit noisy updates, with each node updating its local variable as a weighted average of neighbor values and corrupting zero mean additive noise [27]

$$x(t+1) = Ax(t) + m(t) \quad (44)$$

where  $m(t)$  is the additive zero-mean noise with *fixed* variance. The algorithm yields a random walk with the variation among nodes converging to a steady state [27]. Hence, the authors pose and solve, under the assumption that the  $m(t)$  noise terms are independent, the problem of designing weights  $A$  that yield optimal steady-state behavior.

The generic model clearly subsumes the noisy gossip update, reducing to the latter for  $A$  time-invariant and deterministic, and  $B(t) = I$ . Also, the perturbation is taken to be zero-mean and independent of the node states. Recall that for stationary (or time-invariant in this case) update and control matrices cases, convergence to consensus is achieved by driving the perturbation to zero, as suggested by Corollary 1. But clearly, this condition is not met under the noisy gossip model. Thus, our findings corroborate those of Xiao *et al.* [27]; namely, the noisy gossip algorithm does not satisfy the sufficient conditions derived in this paper.

For fixed variance noise, it is clear that stationary  $B$  matrices are not able to drive the divergence cause by the perturbation to zero. Although more general noisy consensus algorithms are directly covered by Theorem 1, the following gives a corollary of Theorem 1 that considers the case where  $A(t)$  and  $B(t)$  are deterministic but time varying. This is indeed the case considered in noisy consensus algorithms [28]–[30].

*Corollary 4:* Let  $\{A(t), B(t)\} \in \mathbb{R}^{N \times N}$  be nonstationary and deterministic matrices and  $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  be state and perturbation vectors, respectively. Consider any consensus algorithm of the form

$$x(t+1) = A(t)x(t) + B(t)m(t). \quad (45)$$

Now suppose that

$$\begin{aligned} A(t)\mathbf{1} = \mathbf{1}, \mathbb{E}\{m(t)\} = \mathbf{0} \\ \mathbb{E}\{m(t) | x(t), A(t), B(t)\} = \mathbb{E}\{m(t)\}, \forall t \end{aligned} \quad (46)$$

and

$$\begin{aligned} 0 \leq \lambda_1[A(t)^T(I-J)A(t)] \leq 1, \forall t \\ \sum_{t=0}^{\infty} 1 - \lambda_1[A(t)^T(I-J)A(t)] = \infty \end{aligned} \quad (47)$$

$$\sum_{t=0}^{\infty} \lambda_1[B(t)^T(I-J)B(t)]\mathbb{E}\{\|m(t)\|_2^2\} < \infty$$

$$\lim_{t \rightarrow \infty} \frac{\lambda_1[B(t)^T(I-J)B(t)]\mathbb{E}\{\|m(t)\|_2^2\}}{1 - \lambda_1\{[A(t)^T(I-J)A(t)]\}} = 0 \quad (48)$$

where  $\lambda_1[\cdot]$  denotes the largest eigenvalue of its argument. Then any noisy consensus algorithm of the form (45) converges to a consensus almost surely, i.e.,

$$\Pr\left\{\lim_{t \rightarrow \infty} x(t) = c\mathbf{1}\right\} = 1 \quad (49)$$

for some  $c \in \mathbb{R}$ .

Moreover, it is easy to see that the conditions in (48) reduces to

$$\sum_{t=0}^{\infty} \lambda_1[B(t)^T(I-J)B(t)] < \infty,$$

$$\lim_{t \rightarrow \infty} \frac{\lambda_1[B(t)^T(I-J)B(t)]}{1 - \lambda_1\{[A(t)^T(I-J)A(t)]\}} = 0 \quad (50)$$

for  $\mathbb{E}\{\|m(t)\|_2^2\} < \infty$  which is the case for the consensus algorithms with noisy channels considered in [28]–[30].

In the following, we show how the above theorem [specifically, (47) and (50)] reduces to the results presented recently in the noisy consensus literature. Hatano *et al.* consider (also subsequently addressed in [29] and in [30]) the following nonrandom, synchronous model (rearranged for convenience) applied to agreement in independent zero-mean fixed variance  $\sigma^2$  Gaussian corrupted links [28]:

$$x(t+1) = [I - \gamma(t)L]x(t) + \gamma(t)m(t), \gamma(t) > 0 \in \mathbb{R} \quad (51)$$

where  $m_i(t) = \sum_{j=1}^N K_{ji}n_{ji}(t)$  (noise accumulated at the  $i$ th node after receiving all corrupted neighboring values), and  $L$  and  $K$  denote the Laplacian and adjacency matrices of a graph. This is indeed the model generated within the stochastic approximation theory dating back to 1970s [32], [36]–[38]. Notably, the generic model also subsumes this special case, with equivalence realized by taking a nonrandom synchronous  $A(t) = (I - \gamma(t)L)$  and nonrandom synchronous diagonal matrix  $B(t) = \gamma(t)I$ , where  $\gamma(t) \in \mathbb{R}$  and entries  $B_{ii}(t) = \gamma(t)$  for all  $i$  and  $t$ .

The set of conditions in (5) are clearly satisfied due to assumptions on the noise and the fact that  $L\mathbf{1} = \mathbf{0}$ , indicating that  $A(t)\mathbf{1} = \mathbf{1}$  for all  $t \geq 0$ . Moreover, this model greatly simplifies the convergence conditions through the elimination of expectations and the substitution of expressions for  $A(t)$

$$\lambda_1[A(t)^T(I-J)A(t)] = \lambda_1[A^2(t) - J] = \lambda_2[A^2(t)]$$

$$= \max\{\lambda_2^2[A(t)], \lambda_N^2[A(t)]\} \quad (52)$$

where  $\lambda_2[\cdot]$  denotes the second largest eigenvalue of its argument and we utilized the fact that  $L$  is symmetric and  $\mathbf{1}^T L = L\mathbf{1} = \mathbf{0}$ .

Recalling that  $A(t) = (I - \gamma(k)L)$  gives

$$\max\{\lambda_2^2(A(t)), \lambda_N^2[A(t)]\}$$

$$= (\max\{|\lambda_2[A(t)]|, |\lambda_N[A(t)]|\})^2 \quad (53)$$

$$= (\max\{|1 - \gamma(t)\lambda_F[L]|, |1 - \gamma(t)\lambda_1[L]|\})^2 \quad (54)$$

where we utilized the fact that  $\lambda_k[A(t)] = 1 - \gamma(k)\lambda_{N-k+1}[L]$  [2] and denote  $\lambda_F[L]$  as the Fiedler eigenvalue of the Laplacian. Note that  $(1 - \gamma(k)\lambda_F[L]) \geq (1 - \gamma(k)\lambda_1[L])$  since  $\lambda_F[L] \leq \lambda_1[L]$  and  $\gamma(t) > 0$ . Thus

$$0 \leq \lambda_1[A(t)^T(I-J)A(t)] \leq 1$$

$$\Rightarrow \max\{|1 - \gamma(t)\lambda_F[L]|, |1 - \gamma(t)\lambda_1[L]|\} \leq 1, \forall t. \quad (55)$$

Note that if  $|1 - \gamma(t)\lambda_i[L]| \leq 1$  for all  $i$ , then,  $\max\{|1 - \gamma(t)\lambda_F[L]|, |1 - \gamma(t)\lambda_1[L]|\} \leq 1$ . Now, observe the following set of inequalities:

$$|1 - \gamma(t)\lambda_i[L]| < 1 \Rightarrow -1 \leq 1 - \gamma(t)\lambda_i[L] \leq 1$$

$$\Rightarrow 0 \leq \gamma(t) \leq \frac{2}{\lambda_i[L]}. \quad (56)$$

Since  $2/\lambda_1[L] \leq 2/\lambda_i[L]$  for all  $i$ , we observe that

$$0 \leq \gamma(t) \leq \frac{2}{\lambda_1[L]}$$

$$\Rightarrow \max\{|1 - \gamma(t)\lambda_F[L]|, |1 - \gamma(t)\lambda_1[L]|\} \leq 1. \quad (57)$$

Moreover

$$\lambda_1[B(t)^T(I-J)B(t)] = \lambda_1[\gamma^2(t)(I-J)] = \gamma^2(t) \quad (58)$$

and  $\mathbb{E}\{\|m(t)\|_2^2\} \leq \sigma^2 \max_i\{L_{ii}\} < \infty$ . Thus, the remaining conditions of Theorem 1 reduce to

$$0 \leq (\max\{|1 - \gamma(t)\lambda_F[L]|, |1 - \gamma(t)\lambda_1[L]|\})^2 \leq 1,$$

$$\forall t \Leftarrow \gamma(t) \leq \frac{2}{\lambda_1[L]} \quad (59)$$

$$\sum_{t=0}^{\infty} 1 - (\max\{|1 - \gamma(t)\lambda_F[L]|, |1 - \gamma(t)\lambda_1[L]|\})^2$$

$$= \infty \Rightarrow \sum_{t=0}^{\infty} \gamma(t) = \infty \quad (60)$$

where the right-hand side (RHS) of the above comes from expanding the selective max operation, and

$$\sum_{t=0}^{\infty} \gamma^2(t) < \infty$$

$$\lim_{t \rightarrow \infty} \frac{\gamma^2(t)}{1 - (\max\{|1 - \gamma(t)\lambda_F[L]|, |1 - \gamma(t)\lambda_1[L]|\})^2}$$

$$= 0 \Rightarrow \lim_{t \rightarrow \infty} \gamma(t) = 0. \quad (61)$$

It should be noted that the  $\gamma(t)$  interval in (59) is a sufficient condition and that employing more sophisticated techniques to express the maximum eigenvalues of interest in (57) leads to a larger bound.

Although the researchers considering the limited model in (51) do not provide convergence rate and MSE results, they do, however, give conditions on the adaptive parameter  $\gamma(t)$  guaranteeing convergence to a consensus (addressing the problem from a control theory perspective and finding conditions under

which the *disagreement vector* goes to zero in the limit) [28], [29]

$$0 < \gamma(t) < \frac{2}{\lambda_1[L]}, \forall t, \lim_{t \rightarrow \infty} \gamma(t) = 0, \\ \sum_{t=0}^{\infty} \gamma(t) = \infty, \sum_{t=0}^{\infty} \gamma^2(t) < \infty. \quad (62)$$

It is clear that the first condition (above) provided in [28] is more restrictive than that derived here utilizing the more general result, (60). The remaining conditions are the same.

### B. Convergence Rate

In the following we generalize the concept of  $\epsilon$ -converging time, originally defined for standard sum preserving gossip-based averaging algorithms, to include non sum-preserving, e.g.,  $A(t)$  stochastic but not doubly stochastic for all  $t \geq 0$ , and perturbed gossip algorithms.

Of note is that the general  $\epsilon$ -converging time defined below is valid for sum-preserving and non sum-preserving, perturbed gossip algorithms, while the prior definition in the literature held only for sum-preserving algorithms.

*Definition 1:* Given  $\epsilon > 0$ , the  $\epsilon$ -converging time is

$$T(\epsilon) = \inf\{t \geq 0 : \Pr\{\|x(t) - Jx(t)\|_2 \geq \epsilon\} \leq \epsilon\} \quad (63)$$

where  $\|\cdot\|_2$  denotes the  $l_2$  norm of its argument.

In essence,  $\epsilon$ -convergence time,  $T(\epsilon)$ , is the earliest time at which the state vector  $x(t)$  is  $\epsilon$  close to consensus with probability greater than  $1 - \epsilon$ . Small  $\epsilon$  values give high probability bounds on the convergence time of the general consensus algorithms.

The following theorem gives the  $\epsilon$ -convergence time of any model of the form  $x(t+1) = A(t)x(t) + B(t)m(t)$  considered in this paper.

*Theorem 2:*  $\epsilon$ -converging time of any algorithm of the form  $x(t+1) = A(t)x(t) + B(t)m(t)$  is given by

$$T(\epsilon) \leq \inf \left\{ t : V(0) \prod_{k=0}^{t-1} \rho_A(k) \right. \\ \left. + \left[ \sum_{j=1}^{t-1} \prod_{k=j}^{t-1} \rho_A(k) \rho_B(j-1) \mathbb{E}\{M(j-1)\} \right. \right. \\ \left. \left. + \rho_B(t-1) \mathbb{E}\{M(t-1)\} \right] \leq \epsilon^3 \right\} \quad (64)$$

for any  $\epsilon > 0$ , where  $\rho_A(t) \triangleq \lambda_1[\mathbb{E}\{A(t)^T(I-J)A(t)\}]$ ,  $\rho_B(t) \triangleq \lambda_1[\mathbb{E}\{B(t)^T(I-J)B(t)\}]$ ,  $M(t) \triangleq \|m(t)\|_2^2$   $t \geq 0$ , and  $V(t) = \|v(t)\|_2^2$  with  $v(t) = x(t) - Jx(t)$ .

*Proof:* Given the definition of the  $\epsilon$  converging time, we have that

$$\Pr\{\|x(t) - Jx(t)\|_2 \geq \epsilon\} = \Pr\{\|x(t) - Jx(t)\|_2^2 \geq \epsilon^2\} \quad (65)$$

$$= \Pr\{V(t) \geq \epsilon^2\} \quad (66)$$

$$\leq \frac{\mathbb{E}\{V(t)\}}{\epsilon^2} \quad (67)$$

where the second line follows from the definition of  $V(t)$  and last line follows from the Markov inequality. Hence we need to characterize  $\mathbb{E}\{V(t)\}$  in terms of initial conditions which is considered in the following.

Note that from the proof of Theorem 1, we have that

$$\mathbb{E}\{V(t+1) | V(s) : s \leq t\} \\ \leq \rho_A(t)V(t) + \rho_B(t)\mathbb{E}\{M(t)\} \\ \Rightarrow \mathbb{E}\{V(t+1) | V(t)\} \\ \leq \rho_A(t)V(t) + \rho_B(t)\mathbb{E}\{M(t)\} \quad (68)$$

since the RHS of the first equation only depends on  $V(t)$ . Repeatedly conditioning and using the norm recursion given above yields

$$\mathbb{E}\{V(t)\} \leq V(0) \prod_{k=0}^{t-1} \rho_A(k) + \left[ \sum_{j=1}^{t-1} \prod_{k=j}^{t-1} \rho_A(k) \rho_B \right. \\ \left. \times (j-1) \mathbb{E}\{M(j-1)\} + \rho_B(t-1) \mathbb{E}\{M(t-1)\} \right]. \quad (69)$$

Substituting this into (67) gives

$$\Pr\{\|x(t) - Jx(t)\|_2 \geq \epsilon\} \\ \leq \epsilon^{-2} \left( V(0) \prod_{k=0}^{t-1} \rho_A(k) \right. \\ \left. + \left[ \sum_{j=1}^{t-1} \prod_{k=j}^{t-1} \rho_A(k) \rho_B(j-1) \mathbb{E}\{M(j-1)\} \right. \right. \\ \left. \left. + \rho_B(t-1) \mathbb{E}\{M(t-1)\} \right] \right). \quad (70)$$

Since we desire the RHS of the above to be less than  $\epsilon$ , the stated results is obtained.  $\blacksquare$

The theorem reveals that the convergence rate to consensus of any algorithm of the form considered in this work, i.e.,  $x(t+1) = A(t)x(t) + B(t)m(t)$ , is dependent on the *contraction* abilities of the update and control matrices  $A(t)$  and  $B(t)$ , i.e.,  $\rho_A(t)$  and  $\rho_B(t)$ , and the norm of the perturbation along iterations, i.e., *divergence* characteristics of the perturbations.

Similarly to the consensus case, the above theorem greatly simplifies if one only considers stationary update and control matrices.

*Corollary 5:*  $\epsilon$ -converging time of any algorithm of the form  $x(t+1) = A(t)x(t) + B(t)m(t)$  is given by

$$T(\epsilon) \leq \inf \left\{ t : V(0) \rho_A^t + \rho_B \right. \\ \left. \times \sum_{j=1}^t \rho_A^{t-j} \mathbb{E}\{M(j-1)\} \leq \epsilon^3 \right\} \quad (71)$$



for any  $\epsilon > 0$ , where  $\rho_A \triangleq \lambda_1[\mathbb{E}\{A^T(I - J)A\}]$ ,  $\rho_B \triangleq \lambda_1[\mathbb{E}\{B^T(I - J)B\}]$ ,  $M(t) \triangleq \|m(t)\|_2^2$ , for all  $t \geq 0$ , and  $V(t) = \|v(t)\|_2^2$  with  $v(t) = x(t) - Jx(t)$ .

The above corollary, as in the convergence to consensus case, reduces to previous sum-preserving and nonsum-preserving gossiping results reported in the literature [13], [21]. This is seen by taking  $\rho_B(t) = 0$  or  $\mathbb{E}\{M(t)\} = 0$  for all  $t \geq 0$ , i.e., no perturbation, no control matrix  $B(t)$ , and stationary statistics. Moreover, Corollary 5 directly applies to all the quantized consensus algorithms considered in the literature, where the vast majority of work utilizes synchronous and nonrandom update matrices  $A(t) = A$  and  $B(t) = A$  for all  $t \geq 0$ .

Of note is that  $T(\epsilon)$  might not be achievable for all  $\epsilon$  if  $\mathbb{E}\{M(t)\}$  does not form a series converging to zero. We omit this case for the above theorem and corollary since  $\epsilon$ -converging results are strictly for converging algorithms.

### III. CONSENSUS REVEALED: CONVERGENCE IN EXPECTATION

Although perturbation influenced consensus algorithm do not achieve consensus on the initial node measurements average, they do, as the following result indicates, achieve it in expectation (under mild conditions on the update matrices).

*Theorem 3:* Let  $\{A(t), B(t)\} \in \mathbb{R}^{N \times N}$  be (possibly) stochastic and nonstationary matrices and  $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  be state and perturbation vectors, respectively. Consider any consensus algorithm of the form

$$x(t + 1) = A(t)x(t) + B(t)m(t) \quad (72)$$

where the perturbation is zero-mean and the update matrix  $A(t)$  is independent of the state vector  $x(t)$  for all  $t \geq 0$ , and the control matrix  $B(t)$  is independent of the perturbation  $m(t)$  for all  $t \geq 0$ . Then

$$\lim_{t \rightarrow \infty} \mathbb{E}\{x(t)\} = Jx(0) \quad (73)$$

if  $\mathbb{E}\{A(t)\}\mathbf{1} = \mathbf{1}$ ,  $\mathbf{1}^T \mathbb{E}\{A(t)\} = \mathbf{1}^T$ , and

$$0 \leq \phi(\mathbb{E}\{A(t)\}^T \mathbb{E}\{A(t)\} - J) \leq 1, \forall t \geq 0$$

$$\sum_{t=0}^{\infty} 1 - \phi(\mathbb{E}\{A(t)\}^T \mathbb{E}\{A(t)\} - J) = \infty \quad (74)$$

where  $\phi$  denotes the spectral radius of its argument.

*Proof:* Since  $x(t + 1) = A(t)x(t) + B(t)m(t)$  and we assume zero-mean perturbation vectors and update matrix independence from the current state vector

$$\mathbb{E}\{x(t + 1)\} = \mathbb{E}\{A(t)\}\mathbb{E}\{x(t)\}$$

$$= \prod_{k=0}^{t-1} \mathbb{E}\{A(k)\}x(0). \quad (75)$$

Moreover, if we have  $\mathbb{E}\{A(t)\}\mathbf{1} = \mathbf{1}$ ,  $\mathbf{1}^T \mathbb{E}\{A(t)\} = \mathbf{1}^T$ , and

$$0 \leq \phi(\mathbb{E}\{A(t)\}^T \mathbb{E}\{A(t)\} - J) \leq 1, \forall t \geq 0$$

$$\sum_{t=0}^{\infty} 1 - \phi(\mathbb{E}\{A(t)\}^T \mathbb{E}\{A(t)\} - J) = \infty \quad (76)$$

since

$$\|\mathbb{E}\{x(t + 1)\} - Jx(0)\|_2^2 \leq \phi(\mathbb{E}\{A(t)\}^T \mathbb{E}\{A(t)\} - J) \|\mathbb{E}\{x(t) - Jx(0)\}\|_2^2 \quad (77)$$

then, it is easy to see that we have convergence in the expectation. ■

The stationary update and control matrices case is considered next. The proof of the results are omitted for brevity, as they follow similarly to the convergence to consensus case.

*Corollary 6:* Let  $\{A(t), B(t)\} \in \mathbb{R}^{N \times N}$  be (possibly) stochastic stationary matrices and  $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  be state and perturbation vectors, respectively. Consider any consensus algorithm of the form

$$x(t + 1) = A(t)x(t) + B(t)m(t) \quad (78)$$

where the perturbation is zero-mean and the update matrix  $A(t)$  is independent of the state vector  $x(t)$  for all  $t \geq 0$ , and the control matrix  $B(t)$  is independent of the perturbation  $m(t)$  for all  $t \geq 0$ . Then,

$$\lim_{t \rightarrow \infty} \mathbb{E}\{x(t)\} = Jx(0) \quad (79)$$

if  $\mathbb{E}\{A\}\mathbf{1} = \mathbf{1}$ ,  $\mathbf{1}^T \mathbb{E}\{A\} = \mathbf{1}^T$ , and  $\phi(\mathbb{E}\{A\}^T \mathbb{E}\{A\} - J) < 1$ .

This corollary reduces, if it is taken that  $\mathbb{E}\{A\}^T = \mathbb{E}\{A\}$  and  $(\mathbb{E}\{A\})^2 = \mathbb{E}\{A\}$  (as is done in [13], [16], [21], and [27]), to the conditions required for consensus in expectation in sum-preserving, nonsum-preserving, quantized and noisy gossip algorithms [13], [16], [21], [27].

### IV. MEAN SQUARE ERROR ANALYSIS

Note that consensus algorithms susceptible to perturbation are not able to converge to average in strict sense. It is, thus, natural to investigate the MSE performance of such algorithms. The following theorem gives an asymptotic MSE bound for the general consensus algorithm.

*Theorem 4:* Let  $\{A(t), B(t)\} \in \mathbb{R}^{N \times N}$  be (possibly) stochastic and nonstationary matrices and  $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  be state and perturbation vectors, respectively. Consider any consensus algorithm of the form

$$x(t + 1) = A(t)x(t) + B(t)m(t). \quad (80)$$

Now suppose that

$$A(t)\mathbf{1} = \mathbf{1}, \mathbb{E}\{m(t)\} = \mathbf{0},$$

$$\mathbb{E}\{m(t) | x(t), A(t), B(t)\} = \mathbb{E}\{m(t)\}, \forall t \quad (81)$$

and

$$0 \leq \lambda_2(\mathbb{E}\{A(t)^T A(t)\}) \leq 1, \forall t,$$

$$\sum_{t=0}^{\infty} 1 - \lambda_2(\mathbb{E}\{A(t)^T A(t)\}) = \infty \quad (82)$$

with

$$\begin{aligned} \infty &> \sum_{t=0}^{\infty} (1 - \lambda_2(\mathbb{E}\{A(t)^T A(t)\})) \lambda_{N-1} \\ &\times \left( \prod_{k=0}^{t-1} \mu_A(k)^T \prod_{k=0}^{t-1} \mu_A(t-1-k) - J \right) \geq C_1 \end{aligned} \quad (83)$$

$$-\infty < \sum_{t=0}^{\infty} \lambda_1(\mathbb{E}\{B(t)^T B(t)\}) \mathbb{E}\{M(t)\} \leq C_2 \quad (84)$$

where  $\mu_A(t) = \mathbb{E}\{A(t)\}$ ,  $\forall t$  and  $C_i$  for  $i \in \{1, 2\}$  denote finite constants. Then, the limiting MSE is asymptotically bounded by

$$\lim_{t \rightarrow \infty} \mathbb{E}\{S(t)\} \preceq S(0) \left[ 1 - C_1 + \frac{C_2}{S(0)} \right]. \quad (85)$$

Before proving the above theorem, we present a lemma giving a recursion the MSE follows. Of note is that this recursion is valid for all algorithm types considered in this paper.

*Lemma 3:* Let  $\delta(t) \triangleq x(t) - Jx(0)$  and  $S(t) = \|\delta(t)\|_2^2$ . Then

$$\begin{aligned} \mathbb{E}\{S(t+1)\} &\leq \mathbb{E}\{S(t)\} - (1 - \lambda_2(\mathbb{E}\{A(t)^T A(t)\})) \\ &\times \mathbb{E}\{V(t)\} + \lambda_1(B(t)^T B(t)) \mathbb{E}\{M(t)\}. \end{aligned} \quad (86)$$

where, as before,  $M(t) = \|m(t)\|_2^2$  and  $V(t) = \|v(t)\|_2^2$  with  $v(t) = x(t) - Jx(t)$ . ■

*Proof:* It is easy to check that

$$\delta(t+1) = A(t)\delta(t) + B(t)m(t). \quad (87)$$

Of note is that this recursion is slight different than that for the state deviation vector. Thus, taking the norm of the above, expanding the terms, and taking the statistical expectation yields

$$\begin{aligned} S(t+1) &= \|A(t)\delta(t)\|_2^2 + \|B(t)m(t)\|_2^2 \\ &\quad + 2\delta(t)^T A(t)^T B(t)m(t) \quad (88) \\ \Rightarrow \mathbb{E}\{S(t+1) | \delta(t)\} &= \delta(t)^T \mathbb{E}\{A(t)^T A(t)\} \delta(t) \\ &\quad + \mathbb{E}\{m(t)^T B(t)^T B(t)m(t)\} \end{aligned} \quad (89)$$

where the second line follows from the fact that the perturbation is zero-mean for all  $t \geq 0$ . Focus now on the first term on the RHS of the above. Note that  $A(t)\mathbf{1} = \mathbf{1}$  for all  $t \geq 0$ . Thus, unity is an eigenvalue of  $\mathbb{E}\{A(t)^T A(t)\}$  for all  $t \geq 0$ . Moreover, from the Peron-Frobenius theorem, the multiplicity is one. Thus, we have

$$\begin{aligned} \delta(t)^T \mathbb{E}\{A(t)^T A(t)\} \delta(t) \\ =^{(a)} \sum_{i=1}^N \lambda_i(\mathbb{E}\{A(t)^T A(t)\}) |y_i(t)|^2 \end{aligned} \quad (90)$$

$$= |y_1(t)|^2 + \sum_{i=2}^N \lambda_i(\mathbb{E}\{A(t)^T A(t)\}) |y_i(t)|^2 \quad (91)$$

$$\begin{aligned} =^{(b)} &|y_1(t)|^2 + \lambda_2(\mathbb{E}\{A(t)^T A(t)\}) |y_1(t)|^2 \\ &- \lambda_2(\mathbb{E}\{A(t)^T A(t)\}) |y_1(t)|^2 \\ &+ \sum_{i=2}^N \lambda_i(\mathbb{E}\{A(t)^T A(t)\}) |y_i(t)|^2 \end{aligned} \quad (92)$$

$$\begin{aligned} \leq &(1 - \lambda_2(\mathbb{E}\{A(t)^T A(t)\})) |y_1(t)|^2 \\ &+ \lambda_2(\mathbb{E}\{A(t)^T A(t)\}) \sum_{i=1}^N |y_i(t)|^2 \end{aligned} \quad (93)$$

$$\begin{aligned} =^{(c)} &(1 - \lambda_2(\mathbb{E}\{A(t)^T A(t)\})) \|J\delta(t)\|_2^2 \\ &+ \lambda_2(\mathbb{E}\{A(t)^T A(t)\}) \|\delta(t)\|_2^2 \end{aligned} \quad (94)$$

where (a) follows from the unitary eigendecomposition of  $\mathbb{E}\{A(t)^T A(t)\} = U(t)\Lambda(t)U(t)^T$  and (b) follows from adding and subtracting  $\lambda_2(\mathbb{E}\{A(t)^T A(t)\}) |y_1(t)|^2$  and (c) is due to the fact that  $y_1(t) = u_1(t)\delta(t) \Rightarrow |y_1(t)|^2 = \delta(t)^T J\delta(t) = \|J\delta(t)\|_2^2$  (with  $u_1(t) = N^{-1/2}\mathbf{1}^T$ ) and  $\sum_{i=1}^N |y_i(t)|^2 = y(t)^T y(t) = \delta(t)^T U(t)U(t)^T \delta(t) = \|\delta(t)\|_2^2$  (due to unitary decomposition). Finally, utilizing Rayleigh-Ritz theorem on the second term of the RHS of (89) [similarly to its application in (18)] gives

$$\begin{aligned} \mathbb{E}\{\|\delta(t+1)\|_2^2 | \delta(t)\} \\ \leq (1 - \lambda_2(\mathbb{E}\{A(t)^T A(t)\})) \|J\delta(t)\|_2^2 \\ + \lambda_2(\mathbb{E}\{A(t)^T A(t)\}) \|\delta(t)\|_2^2 \\ + \lambda_1(B(t)^T B(t)) \mathbb{E}\{M(t)\}. \end{aligned} \quad (95)$$

Now, to simplify the notational burden, let  $\rho'_A(t) = \lambda_2(\mathbb{E}\{A(t)^T A(t)\})$  and  $\rho'_B(t) = \lambda_1(B(t)^T B(t))$ . Moreover, using algebraic manipulations, it is easy to see that

$$\begin{aligned} \mathbb{E}\{\|J\delta(t)\|_2^2\} &= \mathbb{E}\{\|\delta(t)\|_2^2\} - \mathbb{E}\{\|v(t)\|_2^2\} \\ &= \mathbb{E}\{S(t)\} - \mathbb{E}\{V(t)\}. \end{aligned} \quad (96)$$

Substituting this into the recursion gives

$$\begin{aligned} \mathbb{E}\{S(t+1)\} &= \mathbb{E}\{S(t)\} - (1 - \rho'_A(t)) \\ &\times \mathbb{E}\{V(t)\} + \rho'_B(t) \mathbb{E}\{M(t)\}. \end{aligned} \quad (97)$$

Now, replacing the parameters with their equivalent expressions give the desired result. ■

The above Lemma reveals the mechanics of the MSE through iterations of the generic consensus algorithm and shows how each parameter affects these mechanics. Notably, the MSE shrinks proportionally to the contracting capabilities of the update and control matrices  $\lambda_2(\mathbb{E}\{A(t)^T A(t)\})$  and  $\lambda_1(B(t)^T B(t))$ , as well as decrease in the total variation  $\mathbb{E}\{V(t)\}$ . Conversely, the MSE expands proportionally to the perturbation vector norm  $\mathbb{E}\{M(t)\}$ .

As expected, if one takes  $A(t)$  stationary,  $\mathbf{1}^T A(t) = \mathbf{1}^T$ , and  $B(t) = 0$ , or  $m(t) = 0$ , for all  $t$ , then the recursion reduces to  $\mathbb{E}\{S(t+1)\} = \lambda_2(\mathbb{E}\{A(t)^T A(t)\}) \mathbb{E}\{S(t)\}$ , i.e., the MSE recursion for all sum preserving gossip algorithms

[13], [18]. If one only takes  $A(t)$  stationary and  $B(t) = 0$ , or  $m(t) = 0$ , for all  $t$ , then the recursion reduces to  $\mathbb{E}\{S(t+1)\} = \mathbb{E}\{S(t)\} - (1 - \lambda_2(\mathbb{E}\{A(t)^T A(t)\}))\mathbb{E}\{V(t)\}$ , i.e., the MSE recursion for any nonsum-preserving gossip algorithms [21]. Finally, if one takes  $A(t)$  stationary,  $\mathbf{1}^T A(t) = \mathbf{1}^T$ , and  $B(t) = A(t)$  for all  $t$ , then  $\mathbb{E}\{S(t+1)\} = \mathbb{E}\{S(t)\} - (1 - \lambda_2(\mathbb{E}\{A(t)^T A(t)\}))\mathbb{E}\{V(t)\} + \lambda_1(A(t)^T A(t))\mathbb{E}\{M(t)\}$ , i.e., a MSE recursion for any quantized gossip algorithms [16], [25].

We now begin the proof of Theorem 4.

*Proof:* To achieve the desired result we must lower bound  $\mathbb{E}\{V(t)\}$ . Let  $\mu_A(t) \triangleq \mathbb{E}\{A(t)\}$  and note that

$$\mathbb{E}\{\|v(t)\|_2^2\} \geq \|\mathbb{E}\{x(t) - Jx(t)\}\|_2^2 \quad (98)$$

$$= \left\| \left( \prod_{k=0}^{t-1} \mu_A(t-1-k) - J \right) x(0) \right\|_2^2 \quad (99)$$

$$= \left\| \left( \prod_{k=0}^{t-1} \mu_A(t-1-k) - J \right) v(0) \right\|_2^2 \quad (100)$$

where the second line follows from Theorem 3 and the last line from the fact that  $A(t)\mathbf{1} = \mathbf{1}$  for all  $t \geq 0$  with probability one. Recall that  $\mathbf{1}^T \mu_A(t) = \mathbf{1}^T$  for all  $t \geq 0$ . (This assumption is not crucial for the proof. It does, however, greatly simplify the notation burden.) Moreover, let  $\mu_A^\Pi(t-1) \triangleq \prod_{k=0}^{t-1} \mu_A(t-1-k)$  and  $\mu_{A'}^\Pi(t-1) \triangleq \mu_A^\Pi(t-1)^T \mu_A^\Pi(t-1)$ . Then

$$\mathbb{E}\{\|v(t)\|_2^2\} \geq v(0)^T (\mu_{A'}^\Pi(t-1) - J)^T \times (\mu_{A'}^\Pi(t-1) - J) v(0) \quad (101)$$

$$= v(0)^T (\mu_{A'}^\Pi(t-1) - J) v(0) \quad (102)$$

$$= \sum_{i=1}^N \lambda_i (\mu_{A'}^\Pi(t-1) - J) |z_i(t)|^2 \quad (103)$$

$$= \sum_{i=1}^{N-1} \lambda_i (\mu_{A'}^\Pi(t-1)) |z_i(t)|^2 \quad (104)$$

$$= \sum_{i=1}^{N-1} \lambda_i (\mu_{A'}^\Pi(t-1)) |z_i(t)|^2 - \lambda_{N-1} (\mu_{A'}^\Pi(t-1)) |z_N(t)|^2 + \lambda_{N-1} (\mu_{A'}^\Pi(t-1)) |z_N(t)|^2 \quad (105)$$

$$\geq \lambda_{N-1} (\mu_{A'}^\Pi(t-1)) \sum_{i=1}^N |z_i(t)|^2 - \lambda_{N-1} (\mu_{A'}^\Pi(t-1)) |z_N(t)|^2 \quad (106)$$

$$\geq \lambda_{N-1} (\mu_{A'}^\Pi(t-1)) S(0) \quad (107)$$

where the fourth line follows from the fact that  $\lambda_N(\mu_{A'}^\Pi(t-1)) = 0$  and the last line follows for large  $N$  since the summation has  $N$  terms and the fact that  $z(t)^T z(t) = v(0)^T U(t)U(t)^T v(0) = v(0)^T v(0) = \|v(0)\|_2^2 = V(0)$  where we used the properties of the unitary decomposition. Also of

note is that  $V(0) = S(0)$ . Substituting this information back into MSE recursion yields

$$\mathbb{E}\{S(t+1)\} \leq \mathbb{E}\{S(t)\} - (1 - \rho'_A(t))\gamma'_A(t)S(0) + \rho'_B(t)\mathbb{E}\{M(t)\} \quad (108)$$

where  $\gamma'_A(t) \triangleq \lambda_{N-1}(\mu_{A'}^\Pi(t-1))$ . Repeatedly using the above recursion gives:

$$\mathbb{E}\{S(t)\} \leq S(0) \left[ 1 - \sum_{k=0}^{t-1} (1 - \rho'_A(k))\gamma'_A(k) + S(0)^{-1} \sum_{k=0}^{t-1} \rho'_B(k)\mathbb{E}\{M(k)\} \right]. \quad (109)$$

Now, recall that by the definitions of the theorem

$$\sum_{t=0}^{\infty} (1 - \rho'_A(t))\gamma'_A(t) \geq C(\{A(t)\}_{t \geq 0}) \quad (110)$$

and

$$\sum_{t=0}^{\infty} \rho'_B(t)\mathbb{E}\{M(t)\} \leq C(\{B(t), M(t)\}_{t \geq 0}). \quad (111)$$

Taking the limit of MSE and noting that each limit exists, we obtain

$$\lim_{t \rightarrow \infty} \mathbb{E}\{S(t)\} \leq S(0) \left[ 1 - C(\{A(t)\}_{t \geq 0}) + \frac{C(\{B(t), M(t)\}_{t \geq 0})}{S(0)} \right] \quad (112)$$

concluding the proof.  $\blacksquare$

In the following, we consider the stationary version of the above theorem.

*Corollary 7:* Let  $\{A(t), B(t)\} \in \mathbb{R}^{N \times N}$  be (possibly) stochastic and stationary matrices and  $\{x(t), m(t)\} \in \mathbb{R}^{N \times 1}$  be state and perturbation vectors, respectively. Consider any consensus algorithm of the form

$$x(t+1) = A(t)x(t) + B(t)m(t). \quad (113)$$

Now suppose that

$$A(t)\mathbf{1} = \mathbf{1}, \mathbb{E}\{m(t)\} = \mathbf{0}, \mathbb{E}\{m(t) | x(t), A(t), B(t)\} = \mathbb{E}\{m(t)\}, \forall t \quad (114)$$

and

$$0 \leq \lambda_2(\mathbb{E}\{A^T A\}) < 1 \quad (115)$$

with

$$\begin{aligned} \infty &> \frac{1 - \lambda_2(\mathbb{E}\{A^T A\})}{1 - \lambda_{N-1}^2(\mathbb{E}\{A\} - J)} \geq C_1, \\ \infty &< \lambda_1(\mathbb{E}\{B^T B\}) \sum_{t=0}^{\infty} \mathbb{E}\{M(t)\} \leq C_2 \end{aligned} \quad (116)$$

where  $\mu_A = \mathbb{E}\{A(t)\}, \forall t$  and  $C_i$  for  $i \in \{1, 2\}$  denote finite constants. Then, the limiting MSE is asymptotically bounded by

$$\lim_{t \rightarrow \infty} \mathbb{E}\{S(t)\} \preceq S(0) \left[ 1 - C_1 + \frac{C_2}{S(0)} \right]. \quad (117)$$

The above corollary, for instance, directly applies to the synchronous and stationary quantized consensus algorithms adopted in [15], [16], and [25]. Moreover, it is easy to see that the MSE of any nonsum-preserving algorithm for (possibly) stochastic and nonstationary is

$$\lim_{t \rightarrow \infty} \mathbb{E}\{S(t)\} \preceq S(0)[1 - C_1], \quad (118)$$

with appropriate  $C_1$  value.

## V. CONCLUDING REMARKS

We consider general state update systems susceptible to perturbations approached from a consensus perspective. We derive conditions on the system parameters such as nonstationary, random update and control matrices, random perturbation vector, guaranteeing to achieve consensus. Given that these conditions are satisfied, we provide convergence rate to consensus expressions that depends on the system properties. Moreover, we provide conditions on the system parameters guaranteeing convergence in the mean. Finally, we derive an asymptotical upper bound on the mean square performance of these general update systems achieving consensus.

## REFERENCES

- [1] C. C. Moallemi and B. V. Roy, "Consensus propagation," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 4753–4766, Nov. 2006.
- [2] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.
- [3] W. Ren and R. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, no. 5, pp. 655–661, 2005.
- [4] 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, no. 6, pp. 1226–1229, 1995.
- [5] Y. Hatano and M. Mesbahi, "Agreement over random networks," in *Proc. IEEE Conf. Decision Control*, Paradise Island, Bahamas, Dec. 2004.
- [6] A. T. Salehi and A. Jadbabaie, "On consensus in random networks," in *Proc. Allerton Conf. Commun., Contr., Comput.*, IL, Sep. 2007, Allerton House.
- [7] 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, 2008.
- [8] N. Lynch, *Distributed Algorithms*. San Francisco, CA: Morgan Kaufmann, 1996.
- [9] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. IEEE/ACM Int. Symp. Inf. Process. Sens. Netw.*, Los Angeles, CA, Apr. 2005.
- [10] Y. Rabani, A. Sinclair, and R. Wanka, "Local divergence of Markov chains and the analysis of iterative load-balancing schemes," in *Proc. IEEE Symp. Found. Comp. Sci.*, Palo Alto, CA, Nov. 1998.
- [11] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988–1001, 2003.
- [12] J. Tsitsiklis, "Problems in Decentralized Decision Making and Computation," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., MIT, Boston, MA, 1984.
- [13] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.
- [14] D. Kempe, A. Dobra, and J. Gehrke, "Computing aggregate information using gossip," in *Proc. Found. Comput. Sci.*, Cambridge, MA, Oct. 2003.
- [15] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, On Distributed Averaging Algorithms and Quantization Effects M.I.T., Cambridge, LIDS Rep. 2274, Nov. 2007.
- [16] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Rates of convergence of distributed average consensus with probabilistic quantization," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Sep. 2007.
- [17] T. C. Aysal, M. E. Yildiz, A. Sarwate, and A. Scaglione, "Broadcast gossip algorithms: Design and analysis for consensus," in *Proc. IEEE Conf. Decision Contr.*, Cancun, Mexico, Dec. 2008.
- [18] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Contr. Lett.*, vol. 53, pp. 65–78, 2004.
- [19] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 3, Mar. 2008.
- [20] F. Benezit, A. G. Dimakis, P. Thiran, and M. Vetterli, "Gossip along the way: Order-optimal consensus through randomized path averaging," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Allerton, IL, Sep. 2007.
- [21] T. C. Aysal, M. E. Yildiz, and A. Scaglione, "Broadcast gossip algorithms," in *Proc. 2008 IEEE Inf. Theory Workshop*, Porto, Portugal, May 2008.
- [22] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 634–649, 2008.
- [23] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Distributed average consensus using probabilistic quantization," in *Proc. IEEE Statist. Signal Process. Workshop*, Madison, WI, Aug. 2007.
- [24] T. C. Aysal, M. J. Coates, and M. G. Rabbat, "Distributed average consensus using dithered quantization," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4905–4918, Oct. 2008.
- [25] M. E. Yildiz and A. Scaglione, "Differential nested lattice encoding for consensus problems," in *Proc. Inf. Process. Sens. Netw.*, Cambridge, MA, Apr. 2007.
- [26] A. Kashyap, T. Basar, and R. Srikant, "Quantized consensus," *Automatica*, vol. 43, pp. 1192–1203, Jul. 2007.
- [27] 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, Jan. 2007.
- [28] Y. Hatano, A. K. Das, and M. Mesbahi, "Agreement in presence of noise: Pseudogradients on random geometric networks," in *Proc. IEEE Conf. Decision and Contr., and Eur. Contr. Conf.*, Seville, Spain, Dec. 2005.
- [29] S. Kar and J. M. F. Moura, "Distributed average consensus in sensor networks with random link failures and communication channel noise," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2007.
- [30] R. Rajagopal and M. J. Wainwright, "Network-based consensus averaging with general noisy channels," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Allerton, IL, Sep. 2007.
- [31] I. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNS with noisy links-Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.
- [32] B. Poljak, *Introduction to Optimization*. New York: Optimization Software, 1987.
- [33] F. Benezit, A. Dimakis, P. Thiran, and M. Vetterli, "Gossip along the way: Order-optimal consensus through randomized path averaging," in *Proc. 45th Ann. Allerton Conf. Commun., Contr., Comput.*, Sep. 2007.
- [34] L. Schuchman, "A theory of nonsubtractive dither," *IEEE Trans. Commun. Technol.*, vol. COMM-12, pp. 162–165, Dec. 1964.
- [35] R. A. Wannamaker, S. P. Lipshitz, J. Vanderkooy, and J. N. Wright, "A theory of nonsubtractive dither," *IEEE Trans. Signal Process.*, vol. 8, no. 2, pp. 499–516, Feb. 2000.
- [36] B. T. Poljak and Y. Z. Tsympkin, "Pseudogradient adaptation and training algorithms," *Autom. Remote Contr.*, no. 3, pp. 45–68, 1973.
- [37] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Algorithms: Numerical Methods*. Upper Saddle River, NJ: Prentice-Hall, 1989.
- [38] M. Huang and J. H. Manton, "Stochastic approximation for consensus seeking: Mean square and almost sure convergence," in *Proc. 2007 (IEEE) Conf. Decision Contr.*, New Orleans, LA, Dec. 2007.

**Tuncer Can Aysal** (S'05–M'08) received the B.E. degree (high honors) from Istanbul Technical University, Istanbul, Turkey, in May 2003, and the Ph.D. degree from the University of Delaware, Newark, in February 2007, both in electrical and computer engineering.

After the Ph.D. degree, he held research positions with McGill University, Montreal, QC, Canada, and Cornell University, Ithaca, NY, from March 2007 to June 2008. His research interests include distributed signal processing, consensus algorithms, spectrum sensing, and robust, nonlinear, statistical signal, and image processing.

Dr. Aysal was the recipient of the Competitive Graduate Student Fellowship in 2005, a Signal Processing and Communications Graduate Faculty Award in 2006 (award is presented to an outstanding graduate student in this research area), and a University Dissertation Fellowship in 2007 from the University of Delaware. He was also a Best Student Paper finalist at the ICASSP 2007 and his dissertation was nominated to the prestigious Allan P. Colburn prize.

**Kenneth E. Barner** (S'84–M'92–SM'00) received the B.S.E.E. degree (*magna cum laude*) from Lehigh University, Bethlehem, PA, in 1987. He received the M.S.E.E. and Ph.D. degrees from the University of Delaware, Newark, in 1989 and 1992, respectively.

He was the duPont Teaching Fellow and a Visiting Lecturer with the University of Delaware in 1991 and 1992, respectively. From 1993 to 1997, he was an Assistant Research Professor with the Department of Electrical and Computer Engineering, University of Delaware, and a Research Engineer with the duPont Hospital for Children. He is currently Professor and Chairman with the Department of Electrical and Computer Engineering, University of Delaware. He is coeditor of the book *Nonlinear Signal and Image Processing: Theory, Methods, and Applications* (Boca Raton, FL: CRC), 2004. His research interests include signal and image processing, robust signal processing, nonlinear systems, sensor networks and consensus systems, compressive sensing, human-computer interaction, haptic and tactile methods, and universal access.

Dr. Barner is the recipient of a 1999 NSF CAREER award. He was the Co-Chair of the 2001 IEEE-EURASIP Nonlinear Signal and Image Processing (NSIP) Workshop and a Guest Editor for a Special Issue of the *EURASIP Journal of Applied Signal Processing on Nonlinear Signal and Image Processing*. He is a member of the Nonlinear Signal and Image Processing Board. He was the Technical Program Co-Chair for ICASSP 2005 and is currently serving on the IEEE Signal Processing Theory and Methods (SPTM) Technical Committee. He previously served on the IEEE Bio-Imaging and Signal Processing (BISP) Technical Committee, and is currently a member of the IEEE Delaware Bay Section Executive Committee. He has served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTION ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, and the IEEE *Signal Processing Magazine*. He is currently the Editor-in-Chief for the journal *Advances in Human-Computer Interaction*, an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS, a member of the Editorial Board of the *EURASIP Journal of Applied Signal Processing*, and served as a Guest Editor for that journal on the Super-Resolution Enhancement of Digital Video and Empirical Mode Decomposition and the Hilbert-Huang Transform special issues. For his dissertation "Permutation Filters: A Group Theoretic Class of Non-Linear Filters," he received the Allan P. Colburn Prize in Mathematical Sciences and Engineering for the Most Outstanding Doctoral Dissertation in the engineering and mathematical disciplines. He is a member of Tau Beta Pi, Eta Kappa Nu, and Phi Sigma Kappa.