

# Distributed Consensus over Network with Noisy Links

**Behrouz Touri**

Department of Industrial  
and Enterprise Systems Engineering  
University of Illinois  
Urbana, IL 61801  
[touri1@illinois.edu](mailto:touri1@illinois.edu)

**Angelia Nedić\***

Department of Industrial  
and Enterprise Systems Engineering  
University of Illinois  
Urbana, IL 61801  
[angelia@illinois.edu](mailto:angelia@illinois.edu)

**Abstract** – *We consider a distributed consensus problem where a set of agents want to agree on a common value through local computations and communications. We assume that agents communicate over a network with time-varying topology and noisy communication links. We are interested in the case when the link noise is independent in time, and it has zero mean and bounded variance. We present and study an iterative algorithm with a diminishing stepsize. We show that the algorithm converges in expectation and almost surely to a “random” consensus, and we characterize the statistics of the consensus. In particular, we give the expected value of the consensus and provide an upper bound on its variance.*

**Keywords:** Consensus, time varying network, noisy links.

## 1 Introduction

One of the basic problems arising in distributed coordination and control is the consensus problem, also known as the agreement problem. Such a problem arises in a number of applications including decentralized coordination of UAV’s, resource allocation among heterogeneous agents in large-scale networks, and information processing and estimation in sensor networks. In a consensus problem, we have a set of agents each of which has some different initial value (a scalar or a vector). The objective is to develop a distributed and local algorithm that the agents can execute to align their values (agree on a common value). The algorithm needs to be local in the sense that each agent can perform local computations and communicate only with its immediate neighbors.

In this paper, we consider a consensus problem over a network with directed and noisy communication links. We study the case when the network topology is dynamically changing and each agent uses time-varying weights to combine its value with the incoming values from a (time-varying) set of its immediate neighbors. In addition, the agents use a diminishing stepsize to attenuate the effects of the noise in the incoming information. Under a frequent network connectivity assumption and some standard assumptions on the stepsize, we show that the agents reach a *random consensus* in expectation and almost surely. The convergence in expectation is primarily used in computing the expected value of the almost sure limit. We further characterize the statistics of the random consensus by providing a bound on its variance.

The work in this paper is closely related to [11, 12, 13, 16], where consensus over noisy links have been considered. Unlike [11, 12, 13], where a network with a static topology is considered, we consider a network with a dynamically changing topology. In [16], a consensus problem is considered over a network with noisy links and random link failures for a class of networks where the “underlying expected” connectivity model is symmetric and undirected. Furthermore, all agents use the same weights for the incoming information. In contrast with [16], we study exclusively the impact of link noise for a network with a general connectivity structure (not necessarily symmetric and directed) and more general agents’ weights (not necessarily uniform). In our analysis, we use a supermartingale convergence combined with the recent results on distributed averaging of [22], which allows for elegant proofs and insights.

Also related is the more general work on decentralized computational models [39, 41, 40, 18]

---

\*This research is supported by the National Science Foundation under CAREER grant CMMI 07-42538.

(see also, [2]); the literature on distributed coordination and control, and self-organized systems [42, 14, 6, 10, 28, 7, 29, 30, 5, 21, 38, 37, 1, 43, 20]; the effects of delay, quantization and link failure [17, 9, 8, 3, 4, 31]; and consensus-based distributed optimization with various effects [23, 26, 27, 15, 33, 34, 36, 19, 35].

The rest of this paper is organized as follows. In Section 2, we describe the model and an iterative algorithm for reaching consensus over a dynamically changing network with noisy communication links. In Section 3, we study convergence of the algorithm in expectation and almost surely. We also characterize the statistics of the resulting *random consensus*. Section 4 contains our concluding remarks and some future directions of this work.

**Notation and Basic Notions.** We view all vectors as columns. For a vector  $x$ , we write  $x_i$  or  $[x]_i$  to denote its  $i$ th entry. We write  $x^T$  to denote the transpose of a vector  $x$ , and we write  $x^T y$  to denote the inner product of vectors  $x$  and  $y$ . We use  $\|x\|$  to denote the standard Euclidean vector norm i.e.,  $\|x\| = \sqrt{\sum_i x_i^2}$ .

We use  $e_i$  to denote the vector with  $i$ th entry equal to 1 and all other entries equal to 0. We write  $e$  to denote a vector with all entries equal to 1. We use  $I$  to denote the identity matrix. We write  $A^T$  to denote the transpose of a matrix  $A$ . For a matrix  $A$ , we write  $a_{ij}$  or  $[A]_{ij}$  to denote the element in  $i$ th row and  $j$ th column.

A vector  $a$  is said to be *stochastic* when its components  $a_i$  are nonnegative and  $\sum_i a_i = 1$ . A square matrix  $A$  is said to be *stochastic* when each row of  $A$  is a stochastic vector, and it is said to be *doubly stochastic* when both  $A$  and  $A^T$  are stochastic.

## 2 Distributed Model

In this section, we describe our problem of interest and provide some preliminary results.

### 2.1 Problem Formulation

We consider an averaging consensus problem among a set of agents over a network with noisy communication links and time varying topology. We assume that there are  $m$  agents, indexed by  $1, \dots, m$ , and each agent has an initial<sup>1</sup> estimate  $x_i(0) \in \mathbb{R}$ .

Each agent updates its estimate at discrete times  $t_k, 0 = 1, 2, \dots$ . We denote by  $x_i(k)$  the estimate of

<sup>1</sup>Our model is applicable to a more general case where the agent estimates  $x_i(k)$  are vectors. In this case, our analysis can be applied component-wise.

agent  $i$  at time  $t_k$ . When updating, an agent  $i$  combines its current estimate with the noisy estimates received from its neighboring agents. Specifically, agent  $i$  updates its estimates by setting

$$x_i(k+1) = x_i(k) - \alpha_k \sum_{j \in N_i(k)} w_{ij}(k) (x_i(k) - x_j(k) - \zeta_{ij}(k)), \quad (1)$$

where  $N_i(k)$  is the set of neighbors communicating with agent  $i$  at time  $t_k$ . The scalar  $\alpha_k$  is a stepsize with  $\alpha_k > 0$  which is introduced to attenuate the effect of the noise, while the scalars  $w_{ij}(k)$  are positive weights. The term  $x_j(k) + \zeta_{ij}(k)$  represents a noisy estimate that agent  $i$  receives from its neighboring agent  $j$ , with  $\zeta_{ij}(k)$  being the random noise on directed link  $(j, i)$  at time  $t_k$ . The time varying network topology is captured through the use of time varying sets  $N_i(k)$  of agent neighbors. As seen from (1), the agents assign link-dependent time varying weights  $w_{ij}(k)$  to the incoming information.

For the weights  $w_{ij}(k)$ , we assume that

$$\begin{aligned} w_{ij}(k) &> 0 \quad \text{for all } j \in N_i(k), \text{ all } i \text{ and } k, \\ \sum_{j \in N_i(k)} w_{ij}(k) &< 1 \quad \text{for all } i \text{ and } k. \end{aligned} \quad (2)$$

For notational convenience, we define

$$\begin{aligned} w_{ii}(k) &= 1 - \sum_{j \in N_i(k)} w_{ij}(k) > 0, \\ w_{ij}(k) &= 0 \quad \text{iff } j \notin N_i(k) \text{ and } j \neq i, \end{aligned} \quad (3)$$

where  $w_{ii}(k)$  is positive in view of (2). We use these weights to give a more compact description of the evolution of the estimates  $x_i(k)$  generated by (1). In particular, we re-write Eq. (1) as follows

$$x_i(k+1) = x_i(k) - \alpha_k \sum_{j \neq i} w_{ij}(k) (x_i(k) - x_j(k)) + \alpha_k \xi_i(k), \quad (4)$$

where  $\xi_i(k)$  is the total noise that affects the information incoming to agent  $i$  at time  $k$ , given by

$$\xi_i(k) = \sum_{j \neq i} w_{ij}(k) \zeta_{ij}(k). \quad (5)$$

From relations (3) and (4), we have

$$\begin{aligned}
x_i(k+1) &= x_i(k) - \alpha_k \sum_{j \neq i} w_{ij}(k)x_i(k) \\
&\quad + \alpha_k \sum_{j \neq i} w_{ij}(k)x_j(k) + \alpha_k \xi_i(k) \\
&= x_i(k) - \alpha_k(1 - w_{ii}(k))x_i(k) \\
&\quad + \alpha_k \sum_{j \neq i} w_{ij}(k)x_j(k) + \alpha_k \xi_i(k) \\
&= (1 - \alpha_k)x_i(k) \\
&\quad + \alpha_k \sum_{j=1}^m w_{ij}(k)x_j(k) + \alpha_k \xi_i(k). \tag{6}
\end{aligned}$$

By introducing the matrix  $W(k)$  whose entries are  $w_{ij}(k)$  as given in (2)–(3), we can write the preceding relation compactly as follows:

$$x(k+1) = A(k)x(k) + \alpha_k \xi(k), \tag{7}$$

where  $A$  is the matrix given by

$$A(k) = (1 - \alpha_k)I + \alpha_k W(k), \tag{8}$$

the vector  $x(k) \in \mathbb{R}^m$  has entries  $x_i(k)$ , and the vector  $\xi(k) \in \mathbb{R}^m$  has entries  $\xi_i(k)$  defined in (5).

In our subsequent development, we use transition matrices to capture the evolution of estimates  $x(k)$  over a period of time. In particular, we define the *transition matrix*  $\Phi(k, s)$  for any  $s$  and  $k$  with  $k \geq s$ , as follows:

$$\Phi(k, s) = A(k)A(k-1) \cdots A(s+1)A(s). \tag{9}$$

Using these transition matrices, from relation (7) we can relate the estimate  $x(k+1)$  to the estimate  $x(s)$  for any  $s \leq k$ , as follows:

$$x(k+1) = \Phi(k, s)x(s) + \sum_{t=s}^k \alpha_t \Phi(k, t+1)\xi(t), \tag{10}$$

where  $\Phi(k, k+1) = \alpha_k I$ . As seen from the preceding relation, the asymptotic behavior of the estimates  $x(k)$  is closely related to the convergence behavior of the transition matrices  $\Phi(k, s)$ . In our analysis, we consider the case when the weight matrices  $W(k)$  are doubly stochastic and have a uniform lower bound on their positive entries.

**Assumption 1.** *The matrices  $W(k)$  are doubly stochastic. Furthermore, there is a scalar  $\eta > 0$  such that if  $w_{ij}(k) > 0$ , then  $w_{ij}(k) \geq \eta$ .*

In view of relation (8), when  $W(k)$  is doubly stochastic, the matrix  $A(k)$  is also doubly stochastic for any  $\alpha_k \in [0, 1]$ . Consequently, each transition matrix  $\Phi(k, s)$  is doubly stochastic. Furthermore, when the matrix  $W(k)$  has positive entries

bounded away from zero by  $\eta$ , the matrix  $A(k)$  has its positive entries bounded away from zero by  $\alpha_k \eta$ . In particular, by (8), we have

$$[A(k)]_{ij} \geq \alpha_k \eta \quad \text{whenever } [A(k)]_{ij} > 0. \tag{11}$$

In our analysis, we will use some results for the doubly stochastic matrices established<sup>2</sup> in [22]. For this, we need the agent network to be connected over time. At each time  $t_k$ , the agents' connectivity can be represented by a *directed graph*  $(V, \mathcal{E}(W(k)))$ , where  $V = \{1, \dots, m\}$  and  $\mathcal{E}(W)$  is the set of directed edges induced by the positive entries of a matrix  $W$ , i.e.,  $(j, i) \in \mathcal{E}(W)$  whenever  $w_{ij} > 0$ , including self-edges  $(i, i)$ . We assume that the agents are connected frequently enough to persistently influence each other.

**Assumption 2.** *There exists an integer  $B \geq 1$  such that the directed graph*

$$\begin{aligned}
&(V, \mathcal{E}(W(kB)) \cup \mathcal{E}(A(kB+1)) \cup \mathcal{E}(A(kB+2)) \cup \\
&\quad \cdots \cup \mathcal{E}(W((k+1)B-1)))
\end{aligned}$$

*is strongly connected for all  $k \geq 0$ .*

When the matrices  $W(k)$  satisfy Assumption 1, the matrix  $A(k) = (1 - \alpha_k)I + \alpha_k W(k)$  with  $\alpha_k \in (0, 1]$  induces the same directed graph as the matrix  $W(k)$ , i.e.,  $\mathcal{E}(A(k)) = \mathcal{E}(W(k))$  for all  $k$ . Thus, for any  $\alpha_k \in (0, 1]$ , when the matrices  $W(k)$  satisfy Assumptions 1–2, so do the matrices  $A(k)$ .

## 2.2 Preliminary Results

Consider a rule of the form

$$z(k+1) = A(k)z(k), \tag{12}$$

where  $z(0) \in \mathbb{R}^m$  is an arbitrary initial vector. For  $z \in \mathbb{R}^m$ , let us define:

$$\mathcal{V}(z) = (z - \bar{z}e)^T(z - \bar{z}e), \tag{13}$$

where  $\bar{z} = \frac{1}{m}z^T e$  is the mean of the vector  $z$ . The following result has been established in [22], Lemma 5 for the case when  $A(k)$  is given by (8) with<sup>3</sup>  $\alpha_k = 1$  for all  $k$ .

**Lemma 1.** *Let Assumptions 1 and 2 hold. Let the sequence  $\{z(k)\}$  be generated by (12) with  $A(k)$  given by (8), where  $\alpha_k = 1$  for any  $k$ . Then,  $\mathcal{V}(z(k))$  is non-increasing in  $k$ ,*

$$\mathcal{V}(z(k+1)) \leq \mathcal{V}(z(k)) \quad \text{for all } k \geq 0,$$

<sup>2</sup>See also a short conference paper [24].

<sup>3</sup>But under a weaker assumption than Assumption 2.

and we have for any initial vector  $z(0) \in \mathbb{R}^m$  and any  $k \geq 0$ , we have

$$\mathcal{V}(z((k+1)B)) \leq \left(1 - \frac{\eta}{2m^2}\right) \mathcal{V}(z(kB)).$$

Using Lemma 1, it can be seen that we have for all  $k$  and  $s$  with  $k \geq s \geq 0$ ,

$$\mathcal{V}(z(k)) \leq \left(1 - \frac{\eta}{2m^2}\right)^{\lceil \frac{k-s+1}{B} \rceil - 1} \mathcal{V}(z(s)). \quad (14)$$

The following result extends Lemma 1 to a general choice of  $\alpha_k$ 's.

**Lemma 2.** *Let Assumptions 1 and 2 hold. Let the sequence  $\{z(k)\}$  be generated by (12) with  $A(k)$  given by (8), where  $\alpha_k \in (0, 1]$  for all  $k$ . Then,  $\mathcal{V}(z(k))$  is non-increasing in  $k$ ,*

$$\mathcal{V}(z(k+1)) \leq \mathcal{V}(z(k)) \quad \text{for all } k \geq 0,$$

and we have<sup>4</sup> for any vector  $z(0) \in \mathbb{R}^m$  and  $k \geq 0$ ,

$$\mathcal{V}(z((k+1)B)) \leq \left(1 - \frac{\mu_k \eta}{2m^2}\right) \mathcal{V}(z(kB)),$$

where  $\mu_k = \min_{kB \leq t < (k+1)B} \{\alpha_t\}$ .

The preceding result can be seen by following the line of the proof of Lemmas 4 and 5 in [22], and by using relation (11).

We now introduce a condition on the stepsize  $\alpha_k$  in algorithm (7)–(8). In particular, let the stepsize satisfy the following.

**Assumption 3.** *The sequence  $\{\alpha_k\}$  is non-increasing with  $\alpha_k \in (0, 1]$  for all  $k$  and*

$$\sum_{k=0}^{\infty} \alpha_k = \infty.$$

Using Lemma 2, we have the following result.

**Lemma 3.** *Let Assumptions 1–3 hold. Let the sequence  $\{z(k)\}$  be generated by (12) with  $A(k)$  given by (8). Then, we have*

$$\lim_{k \rightarrow \infty} \mathcal{V}(z(kB)) = 0,$$

implying that  $\lim_{k \rightarrow \infty} z_i(k) = \bar{z}(0)$  for all  $i$ .

*Proof.* Based on Lemma 2, we have

$$\begin{aligned} \mathcal{V}(z(kB)) &= \prod_{h=0}^k \left(1 - \frac{\mu_h \eta}{2m^2}\right) \mathcal{V}(z(0)) \\ &\leq e^{-\frac{\eta}{2m^2} \sum_{h=0}^k \mu_h} \mathcal{V}(z(0)). \end{aligned}$$

<sup>4</sup>In fact, it can be seen that  $\mathcal{V}(z((k+1)B)) \leq \left(1 - \frac{\mu_k(1-\mu_k)\eta}{m^2}\right) \mathcal{V}(z(kB))$  when  $\alpha_k \leq \frac{1}{2}$  for all  $k$ .

Under Assumption 3, we have

$$\sum_{h=0}^k \mu(h) = \sum_{h=0}^k \alpha(hB).$$

Furthermore, we have  $\sum_{h=0}^{\infty} \alpha(hB) = \infty$  since  $\sum_{h=0}^{\infty} \alpha_h = \infty$ . Hence,  $\lim_{k \rightarrow \infty} \mathcal{V}(z(kB)) = 0$ . By Lemma 2, the sequence  $\{\mathcal{V}(z(k))\}$  is nonincreasing, and therefore,  $\lim_{k \rightarrow \infty} \mathcal{V}(z(k)) = 0$ . The rest follows by the definition of  $\mathcal{V}$  in (13) and observing that  $\bar{z}(k) = \bar{z}(0)$  in view of the doubly stochasticity of the matrices  $A(k)$ . **Q.E.D.**

Lemma 3 shows that the averaging algorithm (7)–(8), in the absence of noise (i.e.,  $\xi(k) \equiv 0$ ), converges to the average  $\bar{x}(0)$  of the agent initial values. Thus, the averaging algorithm converges even when the matrices  $A(k)$  may have positive entries decreasing to zero at a “slow” enough rate (i.e., at rate such that  $\sum_k \alpha_k = \infty$ ). This extends the known results for convergence of such algorithms which typically assume that all the entries of the weight matrix  $A(k)$  have to be bounded away from zero uniformly in time<sup>5</sup>.

## 2.3 Link Noise

In this section, we discuss the assumptions we impose on the noise. In particular, we use the following.

**Assumption 4.** *Assume that:*

- (a) *The link noise is zero mean, i.e., for all  $1 \leq i \leq m$  and  $k \geq 0$ ,*

$$E[\zeta_{ij}(k)] = 0 \quad \text{for all } j \in N_i(k).$$

- (b) *The noise across links and time has uniformly bounded variance,*

$$E[\zeta_{ij}^2(k)] \leq \sigma^2$$

for all  $j \in N_i(k)$ ,  $1 \leq i \leq m$  and  $k \geq 0$ .

- (c) *The noise is independent<sup>6</sup> across time: for all  $k$  and  $t$  with  $k > t$ ,*

$$E[\zeta_{ij}(k)\zeta_{\ell\kappa}(t)] = 0,$$

for all  $j \in N_i(k)$ ,  $\kappa \in N_{\ell}(t)$ , and all  $i$  and  $\ell$ .

- (d) *The initial vector  $x(0)$  is random with finite expectation and covariance.*

<sup>5</sup>See for example [39, 14, 25, 4, 27].

<sup>6</sup>The noise independence is not crucial; it can be replaced by uncorrelated noise, which amounts to almost the same line of analysis.

In view of the preceding assumption, the aggregate noise  $\xi_i(k)$  [cf. (5)] experienced at agent  $i$  at time  $k$  is also zero mean, and the random vectors  $\xi(k)$  and  $\xi(k')$  are independent for  $k \neq k'$ .

By the definition of  $\xi_i(k)$  and Assumption 4(b), we have

$$\begin{aligned} |E[\xi_i(k) \xi_j(k)]| &\leq \sum_{\ell \neq i} \sum_{\kappa \neq j} w_{i\ell}(k) w_{j\kappa}(k) |E[\zeta_{i\ell}(k) \zeta_{j\kappa}(k)]| \\ &\leq \sigma^2 \sum_{\ell \neq i} w_{i\ell}(k) \sum_{\kappa \neq j} w_{j\kappa}(k). \end{aligned}$$

Furthermore, by (2) and (3), it follows

$$|E[\xi_i(k) \xi_j(k)]| \leq \sigma^2 (1 - w_{ii}(k))(1 - w_{jj}(k)).$$

Thus, since the positive entries of the matrix  $W(k)$  are bounded away from zero by  $\eta$  [cf. Assumption 1], we have for all  $i, j$  and all  $k$ ,

$$|E[\xi_i(k) \xi_j(k)]| \leq \sigma^2 (1 - \eta)^2. \quad (15)$$

### 3 Convergence Results

We study the convergence of the method in expectation and almost surely.

#### 3.1 Convergence in Expectation

For a stepsize satisfying Assumption 3, we show that the algorithm converges in expectation to the average of the vector  $E[x(0)]$ .

**Lemma 4.** *Let Assumptions 1–4 hold. Then, for the iterates generated by the averaging algorithm in (7)–(8), we have for all  $i$ ,*

$$\lim_{k \rightarrow \infty} E[x_i(k)] = \frac{1}{m} \sum_{j=1}^m E[x_j(0)].$$

*Proof.* By taking the expectation in relation (7) and using the zero mean noise assumption, we obtain

$$E[x(k+1)] = A(k)E[x(k)] \quad \text{for all } k.$$

Define  $z(k) = E[x(k)]$  and note that  $z(k)$  satisfies

$$z(k+1) = A(k)z(k) \quad \text{for all } k.$$

The result now follows by applying Lemma 3. **Q.E.D.**

*Remark:* The result of Lemma 4 holds for a constant stepsize i.e.,  $\alpha_k = \alpha$  for all  $k$  and some  $\alpha \in (0, 1]$ , since such a stepsize satisfies the conditions of Assumption 3.

#### 3.2 Almost Sure Convergence

In this section, we investigate almost sure convergence of the method. This requires a typical assumption that the sum of the squared stepsize values,  $\sum_{k=0}^{\infty} \alpha_k^2$ , is finite, which accommodates the use of a standard supermartingale result. The following supermartingale result is due to Robbins and Siegmund, and can be found in [32], Chapter 2, Lemma 11.

**Theorem 1.** *Let  $v_k, u_k, a_k$  and  $b_k$  be nonnegative random variables such that almost surely*

$$E[v_{k+1} | F_k] \leq (1 + a_k)v_k - u_k + b_k \quad \text{for all } k,$$

$$\sum_{k=0}^{\infty} a_k < \infty, \quad \sum_{k=0}^{\infty} b_k < \infty,$$

where  $E[v_{k+1} | F_k]$  denotes the conditional expectation for the given filtration  $F_k$ . Then, we have almost surely

$$\lim_{k \rightarrow \infty} v_k = v, \quad \sum_{k=0}^{\infty} u_k < \infty,$$

where  $v \geq 0$  is some random variable with finite mean.

Throughout this paper, we consider the  $F_k$  to be the natural filtration of the process.

Using the preceding result, we first show that the averages of the iterates  $x(k+1)$  converge almost surely.

**Lemma 5.** *Let Assumptions 1–4 hold. Also, assume that the stepsize  $\alpha_k$  is such that  $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$ . Then, for the average vectors  $\bar{x}(k) = \frac{1}{m} \sum_{j=1}^m x_j(k)$  of the iterates  $x(k)$  generated by the algorithm in (7)–(8), we have*

$$\lim_{k \rightarrow \infty} \bar{x}(k) = \gamma,$$

where  $\gamma$  is a scalar random variable. The mean  $E[\gamma]$  is given by

$$E[\gamma] = \frac{1}{m} \sum_{i=1}^m E[x_i(0)],$$

and the variance of  $\gamma$  is bounded as follows

$$\text{Var}(\gamma) \leq \max_i \text{Var}(x_i(0)) + \sigma^2 (1 - \eta)^2 \sum_{k=0}^{\infty} \alpha_k^2.$$

*Proof.* By taking the average of both sides in relation (7) and by using the doubly stochasticity of  $A(k)$ , we obtain

$$\bar{x}(k+1) = \bar{x}(k) + \alpha_k \bar{\xi}(k) \quad \text{for all } k,$$

where  $\bar{\xi}(k) = \frac{1}{m} \sum_{j=1}^m \xi_j(k)$ . Hence, by subtracting an arbitrary scalar  $c$  from both sides in the preceding expression, and letting  $y(k) = \bar{x}(k) - c$ , we have almost surely for all  $k$ ,

$$y(k+1) = y(k) + \alpha_k \bar{\xi}(k).$$

Let  $F_k$  denote the history of the method (7)–(8) up to time  $k$ . i.e.,  $F_k = \sigma(x(0), \xi(0), \dots, \xi(k))$ . By taking squares and the conditional expectation on  $F_k$ , we obtain almost surely for all  $k$ ,

$$\begin{aligned} E[y^2(k+1) | F_k] &= y^2(k) + 2\alpha_k y(k) E[\bar{\xi}(k)] \\ &\quad + \alpha_k^2 E[\bar{\xi}^2(k)]. \end{aligned}$$

Using the zero mean noise assumption, we have  $E[\bar{\xi}(k)] = 0$ . By uniformly bounded noise variance assumption, [i.e., using (15)], we can see that  $\max_k E[\bar{\xi}^2(k)] \leq \sigma^2(1-\eta)^2$ . Therefore, almost surely we have for all  $k$ ,

$$E[y^2(k+1) | F_k] \leq y^2(k) + \alpha_k^2 \sigma^2(1-\eta)^2. \quad (16)$$

We now apply Theorem 1 with  $v_k = y^2(k)$ ,  $a_k = u_k = 0$  and  $b_k = \alpha_k^2 \sigma^2(1-\eta)^2$ . We note that since  $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$ , the conditions of the theorem are satisfied, and therefore it follows that the variable  $y^2(k) = (\bar{x}(k) - c)^2$  converges almost surely for any  $c \in \mathbb{R}$ . This, however, implies that  $\bar{x}(k)$  converges almost surely to a scalar random variable, say  $\gamma$ .

By Lemma 4, we have  $\lim_{k \rightarrow \infty} E[x_i(k)] = \frac{1}{m} \sum_{j=1}^m E[x_j(0)]$ , implying that

$$\lim_{k \rightarrow \infty} E[\bar{x}(k)] = \frac{1}{m} \sum_{j=1}^m E[x_j(0)].$$

Since  $\lim_{k \rightarrow \infty} \bar{x}(k) = \gamma$  almost surely, it follows that  $\lim_{k \rightarrow \infty} E[\bar{x}(k)] = E[\gamma]$ , and consequently,  $E[\gamma] = \frac{1}{m} \sum_{j=1}^m E[x_j(0)]$ .

By taking the expectation in (16), we can see that

$$E[y^2(k+1)] \leq E[y(0)^2] + \sigma^2(1-\eta)^2 \sum_{t=0}^k \alpha_t^2.$$

Since  $y(k) = \bar{x}(k) - c$  with arbitrary  $c$  and, as shown above,  $\bar{x}(k)$  converges almost surely to  $\gamma$ , by Fatou's lemma it follows that

$$\begin{aligned} E[(\gamma - c)^2] &\leq \liminf_{k \rightarrow \infty} E[y^2(k+1)] \\ &\leq E[(\bar{x}(0) - c)^2] + \sigma^2(1-\eta)^2 \sum_{t=0}^{\infty} \alpha_t^2, \end{aligned}$$

where the last term is finite by our assumption on the stepsize. Letting  $c = E[\gamma]$ , we obtain

$$\begin{aligned} \text{Var}(\gamma) &= E[(\gamma - E[\gamma])^2] \leq E[(\bar{x}(0) - E[\gamma])^2] \\ &\quad + \sigma^2(1-\eta)^2 \sum_{t=0}^{\infty} \alpha_t^2. \end{aligned}$$

Since  $E[\gamma] = \frac{1}{m} \sum_{j=1}^m E[x_j(0)]$ , it can be seen that

$$E[(\bar{x}(0) - E[\gamma])^2] \leq \max_i \text{Var}(x_i(0)).$$

By combining the preceding two relations, we obtain the given bound on variance of  $\gamma$ . **Q.E.D.**

We now show that actually the iterates  $x(k)$  converge almost surely to the random vector  $\gamma e$ .

**Proposition 2.** *Let Assumptions 1–4 hold. Also, let the stepsize  $\alpha_k$  be such that  $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$ . Then, the iterates  $x(k)$  generated by the algorithm in (7)–(8) converge almost surely to the random vector  $\gamma e$ , i.e., we have almost surely*

$$\lim_{k \rightarrow \infty} x_i(k) = \gamma \quad \text{for all } i,$$

where  $\gamma$  is the random variable established in Lemma 5.

*Proof.* We start by considering the subsequence  $\{x(kB)\}$  and showing that this subsequence has the desired convergence property. Then, we show that the whole sequence  $\{x(k)\}$  exhibits the same convergence behavior as the subsequence  $\{x(kB)\}$ .

From the relation in (10), we have

$$\begin{aligned} x((k+1)B) &= \Phi((k+1)B - 1, kB)x(kB) \\ &\quad + \sum_{t=kB}^{(k+1)B-1} \alpha_t \Phi((k+1)B - 1, t+1)\xi(t). \end{aligned}$$

Being a product of doubly stochastic matrices,  $\Phi(k, s)$  is a doubly stochastic matrix for any  $k$  and  $s$  [see (9)]. Using this, and taking the average of the vectors on both sides in the preceding relation, we obtain

$$\bar{x}((k+1)B) = \bar{x}(kB) + \sum_{t=kB}^{(k+1)B-1} \alpha_t \bar{\xi}(t).$$

By combining the preceding two relations, we obtain

$$\begin{aligned} x((k+1)B) - \bar{x}((k+1)B)e &= d((k+1)B) \\ &\quad + \sum_{t=kB}^{(k+1)B-1} \alpha_t (\Phi((k+1)B - 1, t+1)\xi(t) - \bar{\xi}(t)e), \end{aligned}$$

where

$$d((k+1)B) = \Phi((k+1)B-1, kB)x(kB) - \bar{x}(kB)e.$$

By taking squared norms, and then the conditional expectation on the past  $F_{kB}$  (up to time  $kB$ ), while using the zero mean and independence conditions on the noise, we further have almost surely for all  $k$ ,

$$\begin{aligned} E[\|x((k+1)B) - \bar{x}((k+1)B)e\|^2 | F_{kB}] \\ = \|d((k+1)B)\|^2 + \sum_{t=kB}^{(k+1)B-1} \alpha_t^2 E[\|S(k+1, t)\|^2]. \end{aligned}$$

with

$$S(k+1, t) = \Phi((k+1)B-1, t+1)\xi(t) - \bar{\xi}(t)e.$$

Note that, in view of the definition of the function  $\mathcal{V}$  in (13), the first term can be written as

$$\begin{aligned} E[\|x((k+1)B) - \bar{x}((k+1)B)e\|^2 | F_{kB}] = \\ E[\mathcal{V}(x((k+1)B)) | F_{kB}]. \end{aligned}$$

Note that in view of the doubly stochasticity of the matrix  $\Phi((k+1)B-1, kB)$  the term  $d((k+1)B)$  has the form  $d((k+1)B) = y - \bar{y}e$  with  $y = \Phi((k+1)B-1, kB)x(kB)$ . Hence,

$$\|d((k+1)B)\|^2 = \mathcal{V}(\Phi((k+1)B-1, kB)x(kB)).$$

with  $y = \Phi((k+1)B-1, kB)x(kB)$ . In view of Lemma 2, where  $z(k+1)B = \Phi((k+1)B-1, kB)x(kB)$ , we have

$$\begin{aligned} \mathcal{V}(\Phi((k+1)B-1, kB)x(kB)) \\ \leq \left(1 - \frac{\mu_k \eta}{2m^2}\right) \mathcal{V}(x(kB)), \end{aligned}$$

where  $\mu_k = \min_{kB \leq t < (k+1)B} \{\alpha_t\} = \alpha_{(k+1)B}$  by the nonincreasing property of the stepsize. By combining the preceding four relations, we see that almost surely for all  $k$ ,

$$\begin{aligned} E[\mathcal{V}(x((k+1)B)) | F_{kB}] \leq \left(1 - \frac{\alpha_{(k+1)B} \eta}{2m^2}\right) \mathcal{V}(x(kB)) \\ + \sum_{t=kB}^{(k+1)B-1} \alpha_t^2 E[\|S(k+1, t)\|^2]. \end{aligned}$$

with

$$S(k+1, t) = \Phi((k+1)B-1, t+1)\xi(t) - \bar{\xi}(t)e.$$

By the assumption on bounded noise variance and by stochasticity of the matrices  $\Phi(k, s)$ , we can see

that  $E[\|S(k+1, t)\|^2]$  is bounded by some scalar  $\rho_1$ , implying

$$\begin{aligned} E[\mathcal{V}(x((k+1)B)) | F_{kB}] \leq \left(1 - \frac{\alpha_{(k+1)B} \eta}{2m^2}\right) \mathcal{V}(x(kB)) \\ + \sum_{t=kB}^{(k+1)B-1} \rho_1 \alpha_t^2. \end{aligned}$$

Using the identification

$$\begin{aligned} v_k = \mathcal{V}(x(kB)), \quad u_k = \frac{\alpha_{(k+1)B} \eta}{2m^2} \mathcal{V}(x(kB)), \\ a_k = 0, \quad b_k = \sum_{t=kB}^{(k+1)B-1} \rho_1 \alpha_t^2, \end{aligned}$$

we can see that all conditions of Theorem 1 are satisfied. By this theorem, it follows that almost surely

$$\lim_{k \rightarrow \infty} \mathcal{V}(x(kB)) = v, \quad (17)$$

where  $v \geq 0$  is a scalar random variable, and

$$\sum_{k=0}^{\infty} \frac{\alpha_{(k+1)B} \eta}{2m^2} \mathcal{V}(x(kB)) < \infty. \quad (18)$$

Since  $\{\alpha_k\}$  is monotone and  $\sum_{k=0}^{\infty} \alpha_k = \infty$ , we have  $\sum_{k=0}^{\infty} \alpha_{(k+1)B} = \infty$ . Thus, from (18) it follows

$$\liminf_{k \rightarrow \infty} \mathcal{V}(x(kB)) = 0 \quad \text{almost surely,}$$

which together with relation (17) yields

$$\lim_{k \rightarrow \infty} \mathcal{V}(x(kB)) = 0 \quad \text{almost surely.} \quad (19)$$

We now observe that the preceding analysis applies to any time-shifted subsequence of the form  $\{x(kB+T)\}$ , where  $T \geq 0$  is a fixed time-shift. This is the case since  $x(0)$  and  $\xi(t)$  have finite variances. Therefore, using the same argument as in the preceding, we can see that for all  $T \in [0, B-1]$ ,

$$\lim_{k \rightarrow \infty} \mathcal{V}(x(kB+T)) = 0 \quad \text{almost surely.}$$

Hence,  $\lim_{k \rightarrow \infty} \mathcal{V}x(kB+T) = 0$  almost surely for any  $T \in [0, B-1]$ . This result along with the result of Lemma 5, implying that  $\lim_{k \rightarrow \infty} x_i(k) = \gamma$  for all  $1 \leq i \leq m$  almost surely. **Q.E.D.**

As shown in Proposition 2, the averaging method generates iterates  $x_i(k)$  converging almost surely to a random variable  $\gamma$ . Lemma 5 provides the expectation of  $\gamma$  and a bound on its variance.

As one may observe from the proofs of Lemma 5 and Proposition 2, the independence condition for the noise from the past noise [cf. Assumption 4(c)] is not crucial. These results will hold if the independence condition of Assumption 4(c) is replaced by a weaker condition that the noise is conditionally independent on the past.

## 4 Conclusion

We have considered an iterative algorithm for reaching a consensus over a network with noisy links. We have studied the case when the agents use time-varying weights and a diminishing step-size rule. We showed that for a dynamically changing topology but frequently connected, the agents reach a consensus on a common random value in expectation and almost surely. We have also characterized the statistics of the random consensus. While the proposed algorithm is distributed, it does require that agents update simultaneously and coordinate the stepsize value. Our future work includes the development of iterative algorithms that will allow the agents to update asynchronously and use uncoordinated step sizes.

## References

- [1] D. Angeli and P.-A. Bliman, *Convergence speed of unsteady distributed consensus: decay estimate along the settling spanning-trees*, <http://arxiv.org/abs/math.OC/0610854>, 2006.
- [2] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Prentice-Hall Inc., 1989.
- [3] P.-A. Bliman and G. Ferrari-Trecate, *Average consensus problems in networks of agents with delayed communications*, <http://arxiv.org/abs/math.OC/0503009v2>, 2005.
- [4] P.A. Bliman, A. Nedić, and A. Ozdaglar, *Rate of convergence for consensus with delays*, Proceedings of the 47th IEEE Conference on Decision and Control, 2008.
- [5] V.D. Blondel, J.M. Hendrickx, A. Olshevsky, and J.N. Tsitsiklis, *Convergence in multiagent coordination, consensus, and flocking*, Proceedings of IEEE CDC, 2005.
- [6] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Gossip algorithms: Design, analysis, and applications*, Proceedings of IEEE INFOCOM, 2005.
- [7] M. Cao, D.A. Spielman, and A.S. Morse, *A lower bound on convergence of a distributed network consensus algorithm*, Proceedings of IEEE CDC, 2005.
- [8] R. Carli, F. Fagnani, P. Frasca, T. Taylor, and S. Zampieri, *Average consensus on networks with transmission noise or quantization*, Proc. of European Control Conference, 2007.
- [9] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, *Communication constraints in coordinated consensus problems*, Proceedings of IEEE ACC, 2006.
- [10] A. Dimakis, A. Sarwate, and M. Wainwright, *Geographic gossip: Efficient averaging for sensor networks*, IEEE Transactions on Signal Processing **56** (2008), no. 3, 1205–1216.
- [11] M. Huang and J.H. Manton, *Stochastic approximation for consensus seeking: Mean square and almost sure convergence*, Proceedings of the 46th IEEE Conference on Decision and Control, 2007, pp. 306–311.
- [12] ———, *Stochastic lyapunov analysis for consensus algorithms with noisy measurements*, Proceedings of the 2007 American Control Conference, 2007, pp. 1419–1424.
- [13] ———, *Stochastic consensus seeking with measurement noise: Convergence and asymptotic normality*, Proceedings of the 2008 American Control Conference, 2008, pp. 1337–1342.
- [14] A. Jadbabaie, J. Lin, and S. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Transactions on Automatic Control **48** (2003), no. 6, 988–1001.
- [15] B. Johansson, *On distributed optimization in networked systems*, Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden, 2008.
- [16] S. Kar and J.M.F. Moura, *Distributed consensus algorithms in sensor networks with imperfect communications: Link failures and channel noise*, IEEE Transactions on Signal Processing **57** (2009), no. 1, 355–369.
- [17] A. Kashyap, T. Basar, and R. Srikant, *Quantized consensus*, Automatica **43** (2007), no. 7, 1192–1203.
- [18] S. Li and T. Basar, *Asymptotic agreement and convergence of asynchronous stochastic algorithms*, IEEE Transactions on Automatic Control **32** (1987), no. 7, 612–618.
- [19] I. Lobel and A. Ozdaglar, *Distributed subgradient methods over random networks*, Lab. for Information and Decision Systems, MIT, Report 2800, 2008.

- [20] C. Moallemi and B. Van Roy, *Consensus propagation*, Proceedings of Neural Information Processing Systems, 2005.
- [21] L. Moreau, *Stability of multi-agent systems with time-dependent communication links*, IEEE Transactions on Automatic Control **50** (2005), no. 2, 169–182.
- [22] A. Nedić, A. Olshevsky, A. Ozdaglar, and J.N. Tsitsiklis, *On distributed averaging algorithms and quantization effects*, Tech. Report LIDS 2778, MIT, 2007.
- [23] ———, *Distributed subgradient methods and quantization effects*, Proceedings of the 47th IEEE Conference on Decision and Control, 2008.
- [24] ———, *On distributed averaging algorithms and quantization effects*, Proceedings of the 47th IEEE Conference on Decision and Control, 2008.
- [25] A. Nedić and A. Ozdaglar, *Convergence rate for consensus with delays*, Tech. Report LIDS 2774, Laboratory for Information and Decision Systems, MIT, 2007.
- [26] ———, *On the rate of convergence of distributed subgradient methods for multi-agent optimization*, Proceedings of IEEE CDC, 2007.
- [27] ———, *Distributed subgradient methods for multi-agent optimization*, IEEE Transactions on Automatic Control, forthcoming.
- [28] R. Olfati-Saber and R.M. Murray, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Transactions on Automatic Control **49** (2004), no. 9, 1520–1533.
- [29] A. Olshevsky and J.N. Tsitsiklis, *Convergence rates in distributed consensus and averaging*, Proceedings of IEEE CDC, 2006.
- [30] ———, *Convergence speed in distributed consensus and averaging*, forthcoming in SIAM Journal on Control and Optimization, 2008.
- [31] S. Patterson, B. Bamieh, and A.E. Abbadi, *Distributed average consensus with stochastic communication failures*, Proceedings of 46th IEEE Conference on Decision and Control, 2007, pp. 4215–4220.
- [32] B.T. Polyak, *Introduction to optimisation*, Optimization Software, Inc., New York, 1987.
- [33] S. Sundhar Ram, A. Nedić, and V.V. Veeravalli, *Convergence of distributed algorithms: Case of gradient Lipschitz functions*, Working Paper, 2008.
- [34] ———, *Distributed stochastic subgradient algorithm for convex optimization*, Available at <http://arxiv.org/abs/0811.2595>, 2008.
- [35] ———, *Distributed subgradient projection algorithm for convex optimization*, Forthcoming in ICASSP Proceedings, 2009.
- [36] S. Sundhar Ram, V.V. Veeravalli, and A. Nedić, *Distributed and non-autonomous power control through distributed convex optimization*, IEEE INFOCOM, 2009.
- [37] W. Ren, *Consensus seeking in multi-vehicle systems with a time-varying reference state*, Proceedings of IEEE ACC, 2007.
- [38] W. Ren and R.W. Beard, *Consensus seeking in multi-agent systems under dynamically changing interaction topologies*, IEEE Transactions on Automatic Control **50** (2005), no. 5, 655–661.
- [39] J.N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, MIT, 1984.
- [40] J.N. Tsitsiklis and M. Athans, *Convergence and asymptotic agreement in distributed decision problems*, IEEE Transactions on Automatic Control **29** (1984), no. 1, 42–50.
- [41] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Transactions on Automatic Control **31** (1986), no. 9, 803–812.
- [42] T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and O. Schochet, *Novel type of phase transitions in a system of self-driven particles*, Physical Review Letters **75** (1995), no. 6, 1226–1229.
- [43] L. Xiao and S. Boyd, *Fast linear iterations for distributed averaging*, Systems and Control Letters **53** (2004), 65–78.