Stochastic Consensus Algorithms over General Noisy Networks

Kenta Hanada *, Takayuki Wada **, Izumi Masubuchi ***, Toru Asai ****, and Yasumasa Fujisaki **

Abstract: Stochastic consensus algorithms are considered for multi-agent systems over noisy unbalanced directed networks. The graph which represents a communication network of the system is assumed to contain a directed spanning tree, that is, a given digraph is weakly connected. Then two types of stochastic consensus are investigated, where one is for the agent states themselves and the other is for the time averages of the agent states. The convergence of the algorithms is investigated, which gives a stopping rule, i.e., an explicit relation between the number of iterations and the closeness of the agreement.

Key Words: multi-agent system, stochastic consensus, time averaging algorithm, unbalanced network, stopping rule.

1. Introduction

A multi-agent based consensus is a distributed iterative algorithm, where each agent seeks to obtain an agreement of the state of the agents without any global communication. We can realize various kinds of agreement such as an optimal value of optimization problems [1],[2], a maximum (minimum) value of the initial state [3], and averaging value of the initial state [4],[5]. Stochastic averaging consensus algorithms have been developed and analyzed for systems which contain a certain disturbance [5], e.g., an observation or communication noise, because they have many applications like sensor networks and load balancing. Although stochastic consensus algorithms are iterative, we need to terminate the algorithm within a finite time in real situations. Stopping rules are, therefore, proposed in [6],[7], where they reveal a relation between the closeness of the agreement and the number of iterations explicitly with a probabilistic guarantee in advance of execution, while the existing stopping rules [8]–[10] are adaptive and thus stopping criteria are given during execution. The significance of the stopping rules [6],[7] is that they can evaluate the effect of the randomness to the current state, that is, they give how close to the expected value of the agreement the current state is. In [6],[7], the stopping rules are developed over unidirected and balanced directed networks, while an unbalanced directed graph should be considered.

In this paper, we consider two kinds of stochastic consensus algorithms of the multi-agent system over noisy unbalanced networks. One is a standard consensus algorithm with a diminishing communication gain, while the other is a time averaging consensus algorithm [11]–[13] with a static communication gain. We establish a rigorous stopping rule for each algorithm.

This result includes the existing ones [6],[13] as a special case. That is, if the communication network is defined by a balanced graph or an undirected one, the result of the present study coincides with those of [6],[13].

This paper is organized as follows. We argue problem statements to be considered in Section 2. We analyze the convergence of the consensus which gives the stopping rule for standard and time averaging consensus algorithms in Section 3. We show numerical examples in Section 4. Finally, we conclude this paper in Section 5. The preliminary versions of this study were presented at conferences [14],[15], where unweighted graphs are considered. On the other hand, the present paper deals with general weighted graphs, where it is shown that the stopping rules can be basically the same, and thus introducing the weights does not affect the computational complexity of the consensus algorithms.

2. Problem Statements

2.1 Multi-Agent Consensus

Let us consider $N$ agents having the same dynamics

$$x_i(k+1) = x_i(k) + u_i(k), \quad i \in \mathcal{V},$$

(1)

where $x_i(k) \in \mathbb{R}$ is the state of agent $i$, $u_i(k) \in \mathbb{R}$ is the input of agent $i$, $\mathcal{V} = \{1,2,\ldots,N\}$ is the set of agents, $N \in \mathbb{N}$ is the number of agents, and $k \in \mathbb{N}$ is the discrete time. For this set of the agents, we introduce an agent interaction

$$u_i(k) = r(k) \sum_{j \in \mathcal{N}} a_{ij}(y_{ij}(k) - x_i(k)), \quad i \in \mathcal{V}, j \in \mathcal{N}(k),

y_{ij}(k) = x_j(k) + w_{ij}(k),$$

(2)

where $r(k) \in \mathbb{R}$ is the communication gain to be determined later, $a_{ij}$ is a constant weight corresponding to agents $i$ and $j$ $(i \neq j)$ at time $k$, $y_{ij}(k) \in \mathbb{R}$ is the information which agent $i$ receives from agent $j$, $w_{ij}(k) \in \mathbb{R}$ is the communication noise.

Here we assume that the expectation of the noise $w_{ij}(k)$ is 0 and the variance has a finite upper bound $\var^2 \in \mathbb{R}$, i.e., $\mathbb{E}[w_{ij}(k)] = 0$ and $\mathbb{V}[w_{ij}(k)] \leq \var^2 < \infty$, where $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ denote the expectation and the variance of a random variable $\cdot$, respectively. We assume that the noise $w_{ij}(k)$ is independent and identically distributed with respect to $i$, $j$, and $k$ as well.
Furthermore, $N_i \subseteq \mathcal{V} \setminus \{i\}$ denotes the set of in-degree neighbors of agent $i$, which introduces a weighted digraph $G = (\mathcal{V}, E, A)$, where $E \subseteq \mathcal{V} \times \mathcal{V}$ is the set of directed edges and $A = [a_{ij}]$ is the adjacency matrix such that $a_{ij} > 0$ if $(j, i) \in E$, otherwise $a_{ij} = 0$. Throughout this paper, we assume that the weighted digraph $G$ has a directed spanning tree, which implies that the graph is weakly connected. However, we do not assume that $G$ is balanced, which means that in-degree and out-degree of a node can be different. This is a significant feature of this paper. In fact, in the existing work [6], [13], a weakly connected graph $G$ is considered under the assumption that $G$ is balanced, which always implies that $G$ is strongly connected. That is, the setting of the present paper is more general than that of the existing work.

In this paper, we consider two kinds of stochastic consensus. The multi-agent system (1) with noisy interaction (2) is said to achieve the standard consensus if

$$\lim_{k \to \infty} P \left( \exists i, j \in \mathcal{V} \text{ s.t. } \left|x_i(k) - x_j(k)\right| > \epsilon \right) = 0 \tag{3}$$

holds for any $\epsilon > 0$, where $P$ is the probability measure on the noise sequence. The multi-agent system (1) with noisy interaction (2) is said to achieve the time averaging consensus if

$$\lim_{k \to \infty} P \left( \exists i, j \in \mathcal{V} \text{ s.t. } \left|\hat{x}_i(k) - \hat{x}_j(k)\right| > \epsilon \right) = 0 \tag{4}$$

holds for any $\epsilon > 0$, where the time average is defined as

$$\hat{x}_i(k) = \frac{1}{k} \sum_{t=1}^{k} x_i(t), \quad i \in \mathcal{V}. \tag{5}$$

In the rest of the paper, we establish stopping rules for the consensus algorithms (1), (2), and (5) which achieve (3) or (4).

### 2.2 Compact Form and Some Preliminaries

We first derive a compact form of the multi-agent system (1) with noisy interaction (2), which is useful for further analysis.

To this end, we define the graph Laplacian $L$ of the graph $G$:

$$L = [l_{ij}], \quad l_{ij} = \begin{cases} -a_{ij} & \text{if } j \in N_i, \\ \sum_{j=1, j \neq i}^{N} a_{ij} & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

Equations (1) and (2) are then rewritten as

$$x(k + 1) = x(k) + u(k), \tag{6}$$

$$u(k) = -r(k) L x(k) + r(k) d(k), \tag{7}$$

where

$$x(k) = \begin{bmatrix} x_1(k) & x_2(k) & \cdots & x_N(k) \end{bmatrix}^\top \in \mathbb{R}^{N},$$

$$u(k) = \begin{bmatrix} u_1(k) & u_2(k) & \cdots & u_N(k) \end{bmatrix}^\top \in \mathbb{R}^{N},$$

$$d(k) = W(k) 1_N \in \mathbb{R}^{N}, \quad W(k) = [a_{ij} w_{ij}(k)] \in \mathbb{R}^{N \times N}.$$ 

Here $1_N \in \mathbb{R}^{N}$ denotes a column vector whose elements are all one. Substituting (7) into (6), we have

$$x(k + 1) = (I_N - r(k) L) x(k) + r(k) d(k). \tag{8}$$

The time averaging equation (5) can also be rewritten as

$$\hat{x}(k) = \frac{1}{k} \sum_{t=1}^{k} x(t), \tag{9}$$

where

$$\hat{x}(k) = \begin{bmatrix} \hat{x}_1(k) & \hat{x}_2(k) & \cdots & \hat{x}_N(k) \end{bmatrix}^\top \in \mathbb{R}^{N}.$$ 

Now let us summarize properties of the graph Laplacian $L$.

First of all, we see

$$L 1_N = 0,$$

which follows the definition of $L$. Moreover, since we assume that the graph $G$ contains a directed spanning tree, there always exists a vector

$$f = [f_1, f_2, \cdots, f_N]^\top \in \mathbb{R}^{N}$$

such that

$$f^\top L = 0, \quad f^\top 1_N = 1, \quad f_i \geq 0 \forall i \in \mathcal{V}. \tag{10}$$

Thus we have the following lemma.

**Lemma 1** There always exist positive scalars $\lambda, \sigma \in \mathbb{R}$ and matrices $S, T \in \mathbb{R}^{N \times (N-1)}$ such that

$$\begin{bmatrix} \frac{1}{\sigma}(S^\top LT + T^\top L^\top S) \geq \lambda I_{N-1}, \\ T^\top L^\top S^\top LT \leq \sigma^2 I_{N-1}, \end{bmatrix} \tag{11}$$

$$\begin{bmatrix} S^\top \\ f^\top \end{bmatrix} \begin{bmatrix} T & 1_N \end{bmatrix} = \begin{bmatrix} T & 1_N \end{bmatrix} \begin{bmatrix} S^\top \\ f^\top \end{bmatrix} = I_N. \tag{12}$$

**Proof:** We choose matrices $\hat{S} \in \mathbb{R}^{N \times (N-1)}$ and $\hat{T} \in \mathbb{R}^{N \times (N-1)}$ such that

$$\text{Im } \hat{S} = \ker 1_N^\top, \quad \text{Im } \hat{T} = \ker f^\top.$$ 

Then they satisfy

$$\hat{S}^\top 1_N = 0, \quad \text{rank } \hat{S} = N-1, \quad \frac{\text{rank } \hat{T}}{\text{rank } T} = N-1,$$

which means that

$$\begin{bmatrix} \hat{S}^\top \\ f^\top \end{bmatrix} \begin{bmatrix} T & 1_N \end{bmatrix} = \begin{bmatrix} \hat{S}^\top T & 0 \\ 0 & 1 \end{bmatrix} = I_N.$$

We first see that $\hat{S}^\top \hat{T}$ is invertible because $\begin{bmatrix} \hat{T} & 1_N \end{bmatrix}$ and $\begin{bmatrix} \hat{S} & f \end{bmatrix}$ are invertible. In fact, if $1_N \in \text{Im } \hat{T} = \ker f^\top$, it means $f^\top 1_N = 0$, which contradicts (10). That is, $1_N \notin \text{Im } \hat{T}$. Similarly, we see $f \notin \text{Im } \hat{S}$.

We now define $\hat{S} \in \mathbb{R}^{N \times (N-1)}$ and $\hat{T} \in \mathbb{R}^{N \times (N-1)}$ as

$$\hat{S} = \hat{S}, \quad \hat{T} = \hat{T} (S^\top T)^{-1}.$$ 

Since the graph Laplacian $L$ contains a directed spanning tree, we have

$$\begin{bmatrix} \hat{S}^\top \\ f^\top \end{bmatrix} \begin{bmatrix} L & \hat{T} & 1_N \end{bmatrix} = \begin{bmatrix} \hat{S}^\top L \hat{T} & 0 \\ 0 & 0 \end{bmatrix}.$$
where all eigenvalues of $\tilde{S}^T \tilde{L}^T$ should have positive real parts. Considering the Lyapunov inequality with $-\tilde{S}^T \tilde{L}^T$, we see that there always exist a positive definite matrix $P = P^T > 0$, and a positive scalar $\zeta$ such that

$$P (\tilde{S}^T \tilde{L}^T) + (\tilde{S}^T \tilde{L}^T)^T P \geq \zeta P > 0.$$ 

Since $P^{1/2} > 0$, we see that (11) and (13) hold with

$$S = \tilde{S} P^{1/2}, \quad T = \tilde{T} P^{-1/2}, \quad \Delta = \frac{1}{2} \zeta.$$

Moreover, there always exists a positive scalar $\tilde{\sigma}^2$ such that (12) holds, which establishes Lemma 1.

This lemma implies the following lemma, which is the key tool used in the proof of the main results.

**Lemma 2** The graph Laplacian $L$ satisfies

$$\left\| I_{N-1} - \frac{\Delta}{\tilde{\sigma}^2} \tilde{S}^T L T \right\| \leq 1 - \frac{\tilde{\sigma}^2}{\Delta},$$

where $\| \cdot \|$ denotes the spectral norm.

**Proof:** We see that

$$\left\| I_{N-1} - \frac{\Delta}{\tilde{\sigma}^2} \tilde{S}^T L T \right\|^2 = \lambda_{\text{max}} \left( I_{N-1} - \frac{\Delta}{\tilde{\sigma}^2} \left( \tilde{S}^T L T + T^T L^T \tilde{S} \right) + \frac{\tilde{\sigma}^2}{\Delta^2} T^T L^T \tilde{S} \tilde{S}^T L T \right)$$

is satisfied, where $\lambda_{\text{max}}(\cdot)$ is the maximum eigenvalue. From (11) in Lemma 1, we have

$$\frac{\Delta}{\tilde{\sigma}^2} \left( \tilde{S}^T L T + T^T L^T \tilde{S} \right) \geq \frac{2 \tilde{\sigma}^2}{\Delta^2} I_{N-1} \geq \left( \frac{3 \tilde{\sigma}^2}{\Delta^2} - \frac{\tilde{\sigma}^2}{4 \Delta^2} \right) I_{N-1}.$$ 

From (12), we also have

$$\frac{\tilde{\sigma}^2}{\Delta^2} T^T L^T \tilde{S} \tilde{S}^T L T \leq \frac{\tilde{\sigma}^2}{\Delta^2} I_{N-1}.$$ 

We, therefore, see that

$$I_{N-1} - \frac{\Delta}{\tilde{\sigma}^2} \left( \tilde{S}^T L T + T^T L^T \tilde{S} \right) + \frac{\tilde{\sigma}^2}{\Delta^2} T^T L^T \tilde{S} \tilde{S}^T L T \leq I_{N-1} - \left( \frac{2 \tilde{\sigma}^2}{\Delta^2} - \frac{\tilde{\sigma}^2}{4 \Delta^2} \right) I_{N-1} + \frac{\tilde{\sigma}^2}{\Delta^2} I_{N-1}$$

$$= \left( 1 - \frac{\tilde{\sigma}^2}{2 \Delta^2} \right) I_{N-1}$$

is satisfied, which establishes Lemma 2.

We next prepare the expectation and covariance of the noise $d(k)$. With the adjacency matrix $A$ of the graph $G$, let $A_{\text{cov}} \in \mathbb{R}^{N \times N}$ be the diagonal matrix such that

$$A_{\text{cov}} = \text{diag} \left( (A \circ A) I_N \right),$$

where $\circ$ denotes the Hadamard product (element-wise product) and $\text{diag}(q)$ denotes the diagonal matrix whose diagonal elements correspond to the column vector $q$. Then we have the following lemma.

**Lemma 3** The noise $d(k)$ satisfies

$$\mathbb{E}[d(k)] = 0,$$

$$\text{Cov}[d(k)] \leq \nu^2 A_{\text{cov}},$$

where $\text{Cov}[q]$ denotes the covariance of a random vector $q$.

**Proof:** We first rewrite the noise $d(k)$ as

$$d(k) = \left( I_N \otimes I_N \right) \text{diag} \left( \text{vec} (A) \right) \text{vec} (W(k)),$$

where $\otimes$ denotes the Kronecker product and $\text{vec} (A)$ denotes a vector $[ a_1^T a_2^T \cdots a_N^T ]^T \in \mathbb{R}^{N^2}$ for a matrix $A = [a_{ij}] \in \mathbb{R}^{N \times N}$ with $a_{ij} = [ a_{ij1} a_{ij2} \cdots a_{ijN} ]^T$.

From the assumption $\mathbb{E}[w_{ij}(k)] = 0$, we see that

$$\mathbb{E}[d(k)] = 0$$

holds. On the other hand, the covariance of the noise can be derived as

$$\text{Cov}[d(k)] = \mathbb{E} \left[ (d(k) - \mathbb{E}[d(k)]) (d(k) - \mathbb{E}[d(k)])^T \right]$$

$$= \mathbb{E} \left[ (d(k)) (d(k))^T \right]$$

$$= \left( I_N \otimes I_N \right) \text{diag} \left( \text{vec} (A) \right) \mathbb{E} \left[ \text{vec} (W(k)) (\text{vec} (W(k))^T \right]$$

$$= \nu^2 \left( I_N \otimes I_N \right) \text{diag} \left( \text{vec} (A) \right)^2 \left( I_N \otimes I_N \right)^T$$

$$= \nu^2 \text{diag} ((A \circ A) I_N).$$

We, therefore, see that Lemma 3 is established.

### 3. Convergence Analysis

#### 3.1 Standard Consensus

In this section, we investigate the convergence of the standard consensus. Employing a state coordinate transformation

$$\begin{bmatrix} \xi_1(k) \\ \xi_2(k) \end{bmatrix} = \begin{bmatrix} S^T \\ f^T \end{bmatrix} x(k), \quad x(k) = \begin{bmatrix} T & I_N \end{bmatrix} \begin{bmatrix} \xi_1(k) \\ \xi_2(k) \end{bmatrix},$$

where $S$ and $T$ satisfy (13) in Lemma 1, we obtain

$$\xi_1(k + 1) = \left( I_{N-1} - r(k) S^T L T \right) \xi_1(k) + r(k) d_1(k),$$

$$\xi_2(k + 1) = \xi_2(k) + r(k) d_2(k),$$

where

$$d_1(k) = S^T d(k) \in \mathbb{R}^{N-1}, \quad d_2(k) = f^T d(k) \in \mathbb{R}.$$ 

From Lemma 3, we have

$$\mathbb{E}[d_1(k)] = 0,$$

$$\text{Cov}[d_1(k)] \leq \nu^2 S^T A_{\text{cov}} S,$$

$$\mathbb{E}[d_2(k)] = 0,$$

$$\text{Var}[d_2(k)] \leq \nu^2 f^T A_{\text{cov}} f.$$

By using $\xi_1(k)$ and $\xi_2(k)$, the weighted average $\mu(k) \in \mathbb{R}$ of the agent states $x_i(k)$ at time $k$ and the deviation $d(k) \in \mathbb{R}^N$ from the weighted average $\mu(k)$ can be represented as

$$\mu(k) = f^T x(k) = \xi_2(k),$$

where $\mu(k)$ denotes the weighted average of the graph.
\[ \delta(k) = x(k) - \mathbf{1}_N \mu(k) = \left( I_N - \mathbf{1}_N f^T \right) x(k) = T \xi_1(k). \]

We further introduce positive constants \( \bar{\eta}, \eta \in \mathbb{R} \):

\[ \frac{\bar{\eta}^2}{2} I_{N-1} \leq T^T T \leq \frac{\eta^2}{2} I_{N-1}. \quad (19) \]

Then, the norm of \( \delta(k) \) can be bounded as

\[ \eta \| \delta_1(k) \| \leq \| \delta(k) \| \leq \eta \| \delta_1(k) \|, \quad (20) \]

where \( \| \cdot \| \) denotes the Euclidean norm. That is, we can consider the convergence of \( \delta(k) \) as that of \( \xi_1(k) \).

Now, let us select the communication gain as

\[ r(k) = \frac{2}{\lambda(k_0 + k)}, \quad k_0 \geq \frac{2\sigma^2}{I^2} - 1, \quad (21) \]

where \( k_0 \in \mathbb{N} \). Then we have the main result.

**Theorem 1** For given constants \( \alpha \in (0, \infty), \beta \in (0, \infty), \) and \( \gamma \in (0, 1) \), select \( k_f \in \mathbb{N} \) which satisfies

\[ k_f \geq \max(\tau_1, \tau_2), \]

\[ \tau_1 = \left( \frac{\bar{\eta}}{\gamma \eta} - 1 \right) k_0 + 1, \quad (22) \]

\[ \tau_2 = 4v^2 \bar{\eta}^2 \beta \gamma^2 \text{Tr} \left( S^T A_{\text{cov}} S \right) - k_0 + 1, \quad (23) \]

where \( \text{Tr}(Q) \) is the trace of a matrix \( Q \). Then, for any initial state \( \mu(1) \),

\[ \mathbb{P} \left( \| \delta(k_f) \| \leq \alpha \| \delta(1) \| + \beta \right) \geq 1 - \gamma \quad (24) \]

holds.

**Proof:** Following [6], we have

\[ \mathbb{P} \left( \| \xi_1(k) \| \leq \| \mathbb{E}[\xi_1(k)] \| + \sqrt{\frac{\text{Tr}(\text{Cov}[\xi_1(k)])}{\gamma}} \right) \geq 1 - \gamma. \]

From (20), the above relation is transformed into

\[ \mathbb{P} \left( \| \delta(k) \| \leq \bar{\eta} \| \mathbb{E}[\xi_1(k)] \| + \sqrt{\bar{\eta}^2 \text{Tr}(\text{Cov}[\xi_1(k)])} \right) \geq 1 - \gamma. \]

We, therefore, see that we have to evaluate \( \bar{\eta} \| \mathbb{E}[\xi_1(k)] \| \) and \( \bar{\eta}^2 \text{Tr}(\text{Cov}[\xi_1(k)]) \) to prove the theorem.

Let us introduce \( \Gamma(k) \) and \( \Phi(k, m) \) as

\[ \Gamma(k) = I_{N-1} - r(k) S^T LT, \quad (25) \]

\[ \Phi(k, m) = \begin{cases} \Gamma(k-1) \Gamma(k-2) \cdots \Gamma(m) & \text{if } k > m, \\ I_{N-1} & \text{otherwise}. \end{cases} \quad (26) \]

Then, the solution of \( \xi_1(k) \) with the time varying communication gain \( (21) \) can be expressed as

\[ \xi_1(k) = \Phi(k, 1) \xi_1(1) + \sum_{m=1}^{k-1} r(m) \Phi(k, m + 1) d_1(m). \]

From (17), its expectation and covariance are

\[ \mathbb{E}[\xi_1(k)] = \Phi(k, 1) \xi_1(1), \]

\[ \text{Cov}[\xi_1(k)] = \mathbb{E} \left[ \sum_{m=1}^{k-1} r^2(m) \Phi(k, m + 1) d_1(m) \right]. \]

Following the procedure in [6] again with Lemma 2 and \( r(k) \) defined by (21), we have \( \| \Gamma(k) \| \leq (k_0 + k - 1)/(k_0 + k) \), and thus

\[ \| \Phi(k, m + 1) \| \leq \frac{k_0 + m}{k_0 + k - 1} \]

holds. From (20) and the above inequality, we have

\[ \bar{\eta} \| \mathbb{E}[\xi_1(k)] \| \leq \frac{\bar{\eta} k_0}{k_0 + k - 1} \| \xi_1(1) \| \leq \frac{\bar{\eta} k_0}{k_0 + k - 1} \| \delta(1) \| \]

\[ \leq \frac{\bar{\eta} k_0}{\eta (k_0 + k - 1)} \| \delta(1) \| \]

\[ \leq \frac{\eta (k_0 + k - 1) \| \delta(1) \|}{\eta (k_0 + k - 1)} \| \delta(1) \| \]

\[ = \alpha \| \delta(1) \| \]

if \( k \geq \tau_1 \), where \( \tau_1 \) is denoted as (22). We also have

\[ \bar{\eta}^2 \text{Tr}(\text{Cov}[\xi_1(k)]) = \bar{\eta}^2 \text{Tr} \left( \frac{1}{\lambda(k_0 + k)} \right) \leq \bar{\eta}^2 \text{Tr} \left( S^T A_{\text{cov}} S \right) \]

\[ \leq \frac{k - 1}{(k_0 + k - 1)^2} \frac{4v^2 \bar{\eta}^2}{\beta} \text{Tr} \left( S^T A_{\text{cov}} S \right) \]

\[ \leq \frac{k_0 + k - 1}{(k_0 + k - 1)^2} \frac{4v^2 \bar{\eta}^2}{\beta} \text{Tr} \left( S^T A_{\text{cov}} S \right) \]

\[ \leq \frac{1}{k_0 + \tau_2 - 1} \frac{4v^2 \bar{\eta}^2}{\beta} \text{Tr} \left( S^T A_{\text{cov}} S \right) \]

\[ = \beta^2 \gamma \]

if \( k \geq \tau_2 \), where \( \tau_2 \) is denoted as (23). We, therefore, see that

\[ \mathbb{P} \left( \| \delta(k_f) \| \leq \alpha \| \delta(1) \| + \sqrt{\frac{\beta^2 \gamma}{\gamma}} \right) \]

\[ = \mathbb{P} \left( \| \delta(k_f) \| \leq \alpha \| \delta(1) \| + \beta \right) \]

\[ \geq 1 - \gamma \]

holds if we select \( k_f \) such that \( k_f \geq \max(\tau_1, \tau_2) \).

**Theorem 2** For any initial state \( \mu(1) \), the average \( \mu(k) \) satisfies

\[ \mathbb{E}[\mu(k)] = \mu(1) = f^T x(1), \]

\[ \text{Var}[\mu(k)] = \frac{4v^2}{k_0 d^2} f^T A_{\text{cov}} f. \]

**Proof:** By using (16), the solution of \( \mu(k) \) can be expressed as
\( \mu(k) = \xi_2(1) + \sum_{m=1}^{k-1} r(m)d_2(m) \).

From (17) and (18), we obtain
\[
\mathbb{E}[\mu(k)] = \xi_2(1) = \mu(1),
\]
\[
\text{Var}[\mu(k)] = \mathbb{E}\left[ \sum_{m=1}^{k-1} r(m)(d_2(m))^2 \right] 
\leq \nu^2 f^T A_{\text{cov}} f \sum_{m=1}^{k-1} \frac{r^2(m)}{k} 
\leq 4\nu^2 \frac{r^2}{k^2} f^T A_{\text{cov}} f.
\]

which concludes Theorem 2.

3.2 Time Averaging Consensus

In this section, we investigate the convergence of the time averaging consensus. We employ the same state coordinate transformation in the previous section, that is, we use (15) and (16). The time average \( \hat{\xi}_1(k) \) of \( \xi_1(k) \) and \( \hat{\xi}_2(k) \) of \( \xi_2(k) \) can be denoted as
\[
\hat{\xi}_1(k) = \frac{1}{k} \sum_{i=1}^{k} \xi_1(i), \quad \hat{\xi}_2(k) = \frac{1}{k} \sum_{i=1}^{k} \xi_2(i).
\]

Similarly, the time average \( \hat{\mu}(k) \) of \( \mu(k) \) and \( \hat{\delta}(k) \) of \( \delta(k) \) can be written as
\[
\hat{\mu}(k) = \frac{1}{k} \sum_{i=1}^{k} \mu(i) = \frac{1}{k} \sum_{i=1}^{k} \xi_2(i) = \hat{\xi}_2(k),
\]
\[
\hat{\delta}(k) = \frac{1}{k} \sum_{i=1}^{k} \delta(i) = \frac{1}{k} \sum_{i=1}^{k} T \xi_1(i) = T \hat{\xi}_1(k).
\]

As well as the previous section, the norm of \( \hat{\delta}(k) \) can also be bounded as
\[
\frac{\eta}{k} \|\hat{\delta}(k)\| \leq \|\hat{\delta}(k)\| \leq \frac{\eta}{k} \|\hat{\xi}_1(k)\|
\]
by the bound (20) of \( T^T T \).

Now, let us select the static communication gain as
\[
r(k) \equiv r = \frac{4}{\delta^2} \quad \forall k > 0.
\]

Then we have the main result.

**Theorem 3** For given constants \( \alpha \in (0, \infty) \), \( \beta \in (0, \infty) \), and \( \gamma \in (0, 1) \), select \( k_f \in \mathbb{N} \) which satisfies
\[
k_f \geq \max(\tau_1, \tau_2),
\]
\[
\tau_1 = \frac{2\nu^2 \eta}{\alpha \ell^2},
\]
\[
\tau_2 = \frac{4\nu^2 \ell^2}{\beta \gamma \lambda^2} \text{Tr}\left( S^T A_{\text{cov}} S \right).
\]

Then, for any initial state \( x(1) \),
\[
P(\|\hat{\delta}(k)\| \leq \alpha \|\hat{\xi}_1(1)\| + \sqrt{\frac{\eta^2 \text{Tr}(\text{Cov}(\hat{\xi}_1(k)))}{\gamma}}) \geq 1 - \gamma
\]
holds.

**Proof:** We can prove the theorem in a similar manner as in the previous section. In fact, since the following relation
\[
P\left( \frac{\|\hat{\delta}(k)\|}{\|\hat{\xi}_1(1)\|} \leq \frac{\eta}{k} \mathbb{E}[\hat{\xi}_1(k)] + \sqrt{\frac{\eta^2 \text{Tr}(\text{Cov}(\hat{\xi}_1(k)))}{\gamma}} \right)
\]
holds, we see that we have to evaluate \( \frac{\eta}{k} \mathbb{E}[\hat{\xi}_1(k)] \) and \( \mathbb{E}[\hat{\delta}(k)] \).

For a simple expression, we employ again \( \Gamma = I_{N-1} - r N^T L T \) of (25), where we write \( \Gamma \) instead of \( \Gamma(k) \) because the step size \( \gamma_i \) is now constant. Then, the solution of (15) with the static communication gain \( r \) can be denoted as
\[
\hat{\xi}_1(k) = \Gamma^{k-1} \xi_1(1) + r \sum_{m=1}^{k-1} \Gamma^{k-1-m} d_1(m).
\]

The solution of \( \hat{\xi}_1(k) \) can be expressed as
\[
\hat{\xi}_1(k) = \frac{1}{k} \sum_{i=1}^{k} \Gamma^{i-1} \xi_1(1)
\]
\[
+ \frac{r}{k} \sum_{i=1}^{k} \sum_{m=1}^{k-1} \Gamma^{k-1-m} d_1(m).
\]
if \( k \geq \tau_1 \), where \( \tau_1 \) can be denoted as (28). We also obtain
\[
\tilde{\eta}^2 \mathrm{Tr}(\text{Cov}[\hat{x}_1(k)])
\leq \tilde{\eta}^2 \left( \frac{r^2}{k^2} \left( \mathbb{E} \left[ \sum_{i=1}^{k} \left( \sum_{i=m}^{k} \sum_{j=m}^{\infty} d_i(m) \right)^2 \right] \right) \right)
\leq \frac{4 \tilde{\eta}^2 r^2}{k^2} \left( S^T A_{\infty} S \right)
\leq \frac{4 \tilde{\eta}^2 \nu^2}{r^2} \left( S^T A_{\infty} S \right)
= \beta^2 \gamma
\]
if \( k \geq \tau_2 \), where \( \tau_2 \) can be denoted as (29). We, therefore, see that
\[
P \left( \| \tilde{x}_k(k) \| \leq \alpha \| \hat{x}_1(1) \| + \beta \sqrt{\nu^2 \gamma} \right)
\geq 1 - \gamma
\]
holds if we select \( k \) such that \( k \geq \text{max}(\tau_1, \tau_2) \).

\[\Box\]

**Theorem 4** The time average \( \hat{\mu}(k) \) satisfies
\[
\mathbb{E}[\hat{\mu}(k)] = \mu(1),
\]
\[
\text{Var}[\hat{\mu}(k)] \leq \frac{1}{6k^3} \left( k^2 + (2k - 1) \right) \eta^2 f^T A_{\infty} f.
\]

**Proof:** By using (16), the solution of \( \hat{\mu}(k) \) can be expressed as
\[
\hat{\mu}(k) = \hat{x}_2(1) = \frac{1}{k} \sum_{i=1}^{k} \hat{x}_1(1) + \frac{r}{k} \sum_{i=m}^{\infty} d_i(m).
\]
From (17) and (18), we obtain
\[
\mathbb{E}[\hat{\mu}(k)] = \frac{1}{k} \sum_{i=1}^{k} \hat{x}_1(1) = \mu(1),
\]
\[
\text{Var}[\hat{\mu}(k)] = \mathbb{E} \left[ \left( \frac{r^2}{k^2} \sum_{i=m}^{\infty} \sum_{j=m}^{\infty} d_i(m) \right)^2 \right]
\leq \frac{r^2}{k^2} \left( \sum_{i=m}^{\infty} \sum_{j=m}^{\infty} d_i(m) \right)^2
\leq \frac{4 \tilde{\eta}^2 \nu^2}{r^2} \left( S^T A_{\infty} S \right)
\leq \frac{4 \tilde{\eta}^2 \nu^2}{r^2} \left( S^T A_{\infty} S \right)
= \beta^2 \gamma
\]
which concludes Theorem 4.

**4. Numerical Examples**

In this section, we consider numerical examples. We implemented a standard consensus algorithm which follows Theorem 1 and a time averaging consensus one which follows Theorem 3.

First, we describe common problem settings for algorithms in our numerical examples. Let us consider a multi-agent system with \( N = 4 \), where \( \mathcal{V} = \{1, 2, 3, 4\} \) and \( \mathcal{E} = \{(1, 3), (3, 2), (3, 4), (4, 1)\} \). See also Fig. 1. The adjacency matrix and its graph Laplacian are given by
\[
A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 \end{bmatrix}, \quad L = \begin{bmatrix} 2 & 0 & -2 \\ 0 & 3 & -3 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & -4 & 4 \end{bmatrix}.
\]

We selected \( f = [2/7 \ 0 \ 4/7 \ 1/7]^T \) of (10). Note that \( f \) is in fact the left eigenvector corresponding to the zero eigenvalue of the graph Laplacian \( L \). We set \( \text{Var}[w_j(k)] = 1 \) for any \( (i, j) \in \mathcal{E} \) and their upper bound \( \nu = 1 \).

We chose matrices \( S \) and \( T \) as
\[
S = \begin{bmatrix} 0 & -3/7 & 1/7 \\ 1/2 & 0 & 1/2 \\ -1/2 & 1/7 & -3/14 \\ 0 & 2/7 & 1/7 \end{bmatrix}, \quad T = \begin{bmatrix} 1 & -2 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & 1 \\ 2 & 0 & -2 \end{bmatrix}.
\]

Thus we have
\[
S^T LT = \begin{bmatrix} 4 & -3/2 & -1 \\ 4 & 1 & -4 \\ -2 & 1/2 & 5 \end{bmatrix},
\]

where \( a = 0.2120 \) from (11) and \( \sigma^2 = 69.2114 \) from (12). We also have \( \eta = 1.1531 \) and \( \tilde{\eta} = 3.7724 \) from (19). It should be noted that these constants \( \sigma, \eta, \tilde{\eta} \) are determined by \( L \) according to Lemma 1 and (19).

The probabilistic parameters related to Theorem 1 and Theorem 3 were chosen as \( \alpha = 0.3625, \beta = 1.5000 \) and \( \gamma = 0.2000 \). Notice that \( \alpha, \beta, \) and \( \gamma \) are tuning parameters which can be used for specifying the probabilistic guarantee of the solutions given by the consensus algorithms.

The initial state was \( x(1) = [4 \ 8 \ 3 \ 5]^T \).

**4.1 Standard Consensus Algorithm**

We consider a numerical example of the standard consensus algorithm in this section.

We selected \( k_0 = 3, 080 \) so that it is the minimum integer which satisfies (21) because larger gain \( r(k) \) is usually preferable for fast convergence. Then, the number of iterations \( k_f \) was selected as 24, 731. Since \( \| \tilde{x}_1(1) \| = 4.7078 \), Theorem 1 says
\[
P(\| \tilde{x}(24, 731) \| \leq 3.2066) \geq 0.8000.
\]

We executed 1,000 times with different noise sequences. Figures 2, 3, and 4 show that the behavior of the state \( x(k) \).

Fig. 1 The multi-agent network.
average $\mu(k)$, and the deviation $\delta(k)$ at a certain trial respectively. Then, the worst value of $\|\hat{\delta}(24, 731)\|$ in the trials was 0.0890, i.e., the inequality $\|\hat{\delta}(24, 731)\| \leq 3.2066$ is always satisfied for all trials. These results are consistent with Theorem 1. In Fig. 3, the dashed line denotes $E[\mu(k)]$. Here we note that the sample average among trials was nearly equal to this value for any number of iterations. Figure 5 shows the sample variance among 1,000 simulations and the upper bound of the variance in Theorem 2. This result did not invalidate the theorem. In Fig. 2, we see that we have obtained a solution with enough accuracy after around 1,000 iterations. That is, the proposed stopping rule could be conservative for a specific probabilistic distribution. This is because it is based on a distribution-free inequality (Chebyshev inequality) and employs only a part of the information on the probabilistic distribution. Notice that there exists a distribution such that the inequality becomes an equality.

4.2 Time Averaging Consensus Algorithm

We consider a numerical example of the time averaging consensus algorithm in this section. The number of iterations $k_f$ for (27) was selected as 27,810. Since $\|\hat{\delta}(1)\| = 4.7078$, the theorem says

$$P(\|\hat{\delta}(27, 810)\| \leq 3.2066) \geq 0.8000.$$  \hspace{1cm} (32)

We executed 1,000 times with different noise sequences. Figures 6, 7, 8, and 9 show the behaviors of the state $x(k)$, the time average $\hat{x}(k)$ of the state, the average $\hat{\mu}(k)$, and the deviation $\hat{\delta}(k)$ at a certain trial, respectively. As shown in Fig 7, all of the time averages $\hat{x}(k)$ follow the same behavior in sufficiently large $k \geq k_f = 27, 810$. Then, the worst value of $\|\hat{\delta}(27, 810)\|$ in the trials was 0.0345, i.e., the inequality $\|\hat{\delta}(27, 810)\| \leq 3.2066$ is always satisfied for all trials. These results are consistent with Theorem 3.

Theorem 4 mentions the expectation and the variance of $\hat{\mu}(k)$. According to the theorem, the expectation of $\hat{\mu}(k)$ is $\mu(1) = f^T x(1) = 3.5714$, which is depicted by a dashed line in Fig. 8. On the other hand, Fig. 5 shows the variance of the average $\hat{\mu}(k)$. A solid line indicates the sample variance of $\hat{\mu}(k)$ and a dashed line indicates the upper bound of $\text{Var}[\hat{\mu}(k)]$, which is given in Theorem 4. According to Figs. 8 and 10, $\hat{\mu}(k)$ moves around 3.5714 and the sample variance is always smaller than the dashed line. These results are consistent with Theorem 4.

5. Concluding Remarks

We have proposed the stopping rule for the standard and time averaging consensus algorithms in the multi-agent system over
noisy unbalanced networks. The theorems establish the relation between the closeness of the agreement and the number of iterations explicitly with a probabilistic guarantee. Numerical examples showed that the theorem is empirically consistent.

Acknowledgments

This research was supported by JST CREST Grant Number JPMJCR15K2, Japan.

References

[1] A. Nedić and A. Ozdaglar: Distributed subgradient methods for multi-agent optimization, *IEEE Transactions on Automatic Control*, Vol. 54, No. 1, pp. 48–61, 2009.

[2] I. Masubuchi, T. Wada, T. Asai, L.T.H. Nguyen, Y. Ohta, and Y. Fujisaki: Distributed multi-agent optimization based on an exact penalty method with equality and inequality constraints, *SICE Journal of Control, Measurement, and System Integration*, Vol. 9, No. 4, pp. 179–186, 2016.

[3] J. He, P. Cheng, L. Shi, J. Chen, and Y. Sun: Time synchronization in WSNs: A maximum-value-based consensus approach, *IEEE Transactions on Automatic Control*, Vol. 59, No. 3, pp. 660–675, 2014.

[4] F. Fagnani and P. Frasca: Introduction to Averaging Dynamics over Networks, Springer, 2018.

[5] L. Xiao, S. Boyd, and S-J. Kim: Distributed average consensus with least-mean-square deviation, *Journal of Parallel and Distributed Computing*, Vol. 67, No. 1, pp. 33–46, 2007.

[6] R. Morita, T. Wada, I. Masubuchi, T. Asai, and Y. Fujisaki: Multiagent consensus with noisy communication: Stopping rules based on network graphs, *IEEE Transactions on Control of Network Systems*, Vol. 3, No. 4, pp. 358–365, 2016.

[7] K. Hanada, T. Wada, I. Masubuchi, T. Asai, and Y. Fujisaki: Stochastic consensus over time-varying networks of linear symmetric agents, *Transactions of the Institute of Systems, Control and Information Engineers*, Vol. 31, No. 1, pp. 28–35, 2018.

[8] P. Xie, K. You, and C. Wu: How to stop consensus algorithms, locally?, *Proceedings of the 56th IEEE Conference on Decision and Control*, pp. 4544–4549, 2017.

[9] M. Prakash, S. Talukdar, S. Attree, V. Yadav, and M.V. Salapaka: Distributed stopping criterion for consensus in the presence of delays, *IEEE Transactions on Control of Network Systems*, Vol. 5, No. 1, pp. 85–95, 2020.

[10] N.E. Manita and C.N. Hadjicostis: Distributed stopping for average consensus in digraphs, *IEEE Transactions on Control of Network Systems*, Vol. 5, No. 3, pp. 957–967, 2018.

[11] D. Ruppert: Efficient estimations from a slowly convergent Robbins-Monro process, Technical Report 781, Cornell University Operations Research and Industrial Engineering, 1988.

[12] B.T. Polyak and A.B. Juditsky: Acceleration of stochastic approximation by averaging, *SIAM Journal on Control and Optimization*, Vol. 30, No. 4, pp. 838–855, 1992.

[13] R. Morita, T. Wada, I. Masubuchi, T. Asai, and Y. Fujisaki: Time averaging algorithms with stopping rules for multi-agent consensus with noisy measurements, *Asian Journal of Control*, Vol. 18, No. 6, pp. 1969–1982, 2016.

[14] K. Hanada, T. Wada, I. Masubuchi, T. Asai, and Y. Fujisaki: A stopping rule for multi-agent consensus over unbalanced noisy networks, *Proc. 50th ISIE International Symposium on Stochastic Systems Theory and Its Applications*, pp. 117–120, 2018.

[15] K. Hanada, T. Wada, I. Masubuchi, T. Asai, and Y. Fujisaki: A stopping rule for time averaging multi-agent consensus over unbalanced noisy networks, *SICE International Symposium on Control Systems 2019*, 211–1, 2019.

Kenta Hanada (Member)

He received his B.S. and M.S. degrees in Maritime Science and Technology, Ph.D. degree in engineering from Kobe University, Kobe, Japan, in 2011, 2013, and 2016, respectively. During 2014 and 2016, he was a Research Fellow of Japan Society for the Promotion of Science (JSPS). From 2016 to 2020, he joined Graduate School of Information Science and Technology in Osaka University. Since 2020, he has been an Assistant Professor in Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Japan. His research interests include distributed optimization.

Takayuki Wada (Member)

He received the B.E., M.E., and Ph.D. degrees in engineering from Kobe University, Kobe, Japan, in 2004, 2006, and 2009, respectively. During 2008 and 2010, he was a Research Fellow of JSPS. From 2010 to 2012, he was an Assistant Professor in Suzuki National College of Technology, Suzuka, Japan. Since 2012, he has been with Graduate School of Information Science and Technology, Osaka University, Suita, Japan where he has been an Associate Professor from 2017. His research interests include probabilistic approach to analysis and design of systems.

Izumi Masubuchi (Member)

He received the B.E., M.E. and D.E. degrees from Osaka University, Toyonaka, Japan, in 1991, 1993, and 1996, respectively. Since April 2016, he has been a Professor of the Department of Applied Mathematics, Graduate School of System Informatics, Kobe University, Kobe, Japan. His research interests include systems and control theory and its applications.

Toru Asai (Member)

He received his B.E., M.E. and Ph.D. degrees from the Tokyo Institute of Technology, Japan, in 1991, 1993, and 1996, respectively. He has worked as a Research Fellow of JSPS between 1996 and 1998. In 1999, he joined the faculty of Osaka University. Since 2015, he is an Associate Professor of the Sub-department of Mechatronics, Nago University. His research interests include robust control and switching control.

Yasumasa Fujisaki (Member)

He received his B.S., M.S., and Ph.D. degrees in engineering from Kobe University, Kobe, Japan, in 1986, 1988, and 1994, respectively. From 1988 to 1991, he was a Research Member of the Electronics Research Laboratory, Kobe Steel Ltd., Kobe, Japan. From 1991 to 2010, he was with Kobe University and served as a Research Associate and an Associate Professor. Since 2010, he has been a Professor of the Department of Information and Physical Sciences, Osaka University, Suita, Japan. His research interests include randomized algorithms in systems and control.