Convergence Rate for Consensus with Delays

Angelia Nedić* and Asuman Ozdaglar†

October 8, 2007

Abstract

We study the problem of reaching a consensus in the values of a distributed system of agents with time-varying connectivity in the presence of delays. We consider a widely studied consensus algorithm, in which at each time step, every agent forms a weighted average of its own value with values received from the neighboring agents. We study an asynchronous operation of this algorithm using delayed agent values. Our focus is on establishing convergence rate results for this algorithm. In particular, we first show convergence to consensus under a bounded delay condition and some connectivity and intercommunication conditions imposed on the multi-agent system. We then provide a bound on the time required to reach the consensus. Our bound is given as an explicit function of the system parameters including the delay bound and the bound on agents' intercommunication intervals.

*Department of Industrial and Enterprise Systems Engineering, University of Illinois, angelia@uiuc.edu

†Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, asuman@mit.edu

1 Introduction

There has been much recent interest in distributed cooperative control problems, in which several autonomous agents try to collectively accomplish a global objective. This is motivated mainly by the emerging large scale networks which are characterized by the lack of centralized access to information and control. Most recent literature in this area focused on the *consensus problem*, where the objective is to develop distributed algorithms under which agents can reach an agreement or consensus on a common decision (represented by a scalar or a vector). Consensus problem arises in a number of applications including coordination of UAVs, information processing in wireless sensor networks, and distributed multi-agent optimization.

A widely studied algorithm in the consensus literature involves, at each time step, every agent computing a weighted average of its own value with values received from some of the other agents. This algorithm has been proposed and analyzed in the seminal work by Tsitsiklis [18] (see also Tsitsiklis *et al.* [20]). The convergence properties of the consensus algorithm has been further studied under different assumptions on agent connectivity and information exchange by Jadbabaie *et al.* [9] and Blondel *et al.* [4]). Despite much work on the convergence of the consensus algorithm, there has not been a systematic study of the convergence rate of this algorithm in the presence of delays. The presence of delays is a good model for communication networks where there are delays associated with transmission of agent values. Establishing the rate properties of consensus algorithms in such systems is essential in understanding the robustness of the system against dynamic changes.

In this paper, we study convergence and convergence rate properties of the consensus algorithm in the presence of delays. Our analysis is based on reducing the consensus problem with delay to a problem with no delays using state augmentation, i.e., we enlarge the system by including a new agent for each delay element. The state augmentation allows us to represent the evolution of agent values using linear dynamics. The convergence and convergence rate analysis then translates to studying the properties of infinite products of stochastic matrices. Under a bounded delay assumption, we provide rate estimates for the convergence of products of stochastic matrices. Our estimates are per iteration and highlight the dependence on the system parameters including the delay bound.

Other than the papers cited above, our paper is also related to the literature on the consensus problem and average consensus problem (a special case, where the goal is to reach a consensus on the *average of the initial values of the agents*); see Olfati-Saber and Murray [15], Boyd *et al.* [5], Xiao and Boyd [21], Moallemi and Van Roy [12], Cao *et al.* [6], Olshevsky and Tsitsiklis [16], [17]). Recent work has studied the implications of noise and quantization effects on the limiting behavior of the consensus algorithm, see Kashyap *et al.* [10], Carli *et al.* [8], Carli *et al.* [7]. Consensus algorithm also plays a key role in the development of distributed optimization methods. The convergence properties of such methods have been investigated by Tsitsiklis and Athans [19], Li and Basar [11], Bertsekas and Tsitsiklis [2], and more recently in our work [13, 14].

There has also been some work on the convergence of consensus algorithms in the presence of delays. In particular, Bliman and Ferrari-Trecate [3] studied convergence of (average) consensus under *symmetric* delays for a continuous model of agent updates, i.e., a model that represents the evolution of agent values using partial differential equations (which is in contrast with the slotted update rule studied in this paper). Another related work is that of Angeli and Bliman [1], who consider consensus algorithms and the rate of convergence in the presence of delays assuming special topologies for agent connectivity, namely spanning-tree topologies. In contrast with this work, we establish convergence to consensus with delays without requiring any special topologies for agent connectivity. The main contribution of our work is the convergence rate result that *quantifies the algorithm's progress per iteration* and *provides a performance bound in terms of the system and algorithm parameters*.

The rest of this paper is organized as follows: In Section 2, we introduce our notation, formulate the consensus problem, and describe the assumptions imposed on the agent connectivity and information exchange. In Section 3, we introduce and analyze an equivalent consensus problem without a delay, but with an enlarged number of agents. This section also contains our main convergence and rate of convergence results. In Section 4, we provide concluding remarks.

2 Consensus Problem

In this section, we formulate a generic consensus problem and state our assumptions imposed on agent connectivity and local information exchange. To do this, we start by introducing the basic notation and notions that we use throughout the paper.

2.1 Basic Notation and Notions

A vector is viewed as a column, unless clearly stated otherwise. We denote by x_i or $[x]_i$ the i-th component of a vector x. When $x_i \geq 0$ for all components i of a vector x, we write $x \geq 0$. For a matrix A, we write A_i^j or $[A]_i^j$ to denote the matrix entry in the i-th row and j-th column. We write $[A]_i$ to denote the i-th row of the matrix A, and $[A]^j$ to denote the j-th column of A.

We write x' to denote the transpose of a vector x. The scalar product of two vectors $x, y \in \mathbb{R}^m$ is denoted by $x'y$. We use $\|x\|$ to denote the standard Euclidean norm, $\|x\| = \sqrt{x'x}$. We write $\|x\|_\infty$ to denote the max norm, $\|x\|_\infty = \max_{1 \leq i \leq m} |x_i|$.

A vector a is said to be a *stochastic vector* when $a_i \geq 0$ for all i and $\sum_i a_i = 1$. A square $m \times m$ matrix A is said to be a *stochastic matrix* when each row of A is a stochastic vector. A square $m \times m$ matrix A is said to be a *doubly stochastic* matrix when both A and A' are stochastic matrices.

2.2 Consensus Problem with Delay

We consider a network with m agents (or nodes). The neighbors of an agent i are the agents j communicating directly with agent i through a directed link (j, i). Each agent updates and sends its information to its neighbors at discrete times $t_0, t_1, t_2, \ldots$. We index agents' information states and any other information at time t_k by k. We use $x^i(k) \in \mathbb{R}^n$ to denote agent i information state (or estimates) at time t_k.

Each agent i updates its estimate $x^i(k)$ by combining it with the available delayed estimates $x^j(s)$ of its neighbors j. An agent combines the estimates by using nonnegative weights $a^i_j(k)$. These weights capture the information inflow to agent i at time k and the information delay. More specifically, suppose an agent j sends its estimate $x^j(s)$ to agent i. If agent i receives the estimate $x^j(s)$ at time k, then the delay is $t^i_j(k) = k - s$ and agent i assigns a weight $a^i_j(k) > 0$ to the estimate $x^j(s)$. Otherwise, agent i uses $a^i_j(k) = 0$. Formally, each agent i updates its estimate according to the following relation:

$$x^i(k+1) = \sum_{j=1}^{m} a^i_j(k) x^j(k - t^i_j(k)) \qquad \text{for } k = 0, 1, 2, \ldots, \tag{1}$$

where the vector $x^i(0) \in \mathbb{R}^n$ is initial state of agent i, the scalar $t^i_j(k)$ is nonnegative and it represents the delay of a message from agent j to agent i, while the scalar $a^i_j(k)$ is a nonnegative weight that agent i assigns to a delayed estimate $x^j(s)$ arriving from agent j at time k. We use the vector $a^i(k) = (a^i_1(k), \ldots, a^i_m(k))'$ to denote the set of nonnegative weights that agent i uses at time k.

The *consensus problem* involves determining conditions on the agents' connectivity and interactions (including conditions on the weights $a^i(k)$) that guarantee the convergence of the estimates $x^i(k)$, as $k \to \infty$, to a common vector i.e., a limit vector independent of i.

In the absence of a delay, we have $t^i_k(k) = 0$ and the update relation (1) reduces to an algorithm for the *consensus problem without a delay*. This algorithm has been proposed by Tsitsiklis [18]. Variations of this algorithm for various specialized choices of weights and including quantization effects have been recently studied (see [1], [4], [5], [9], [16], [17]), [10], [8], [7].

2.3 Assumptions

Here, we describe some rules that govern the information evolution of the agent system in time. Motivated by the model of Tsitsiklis [18] and the "consensus" setting of Blondel *et al.* [4], these rules include:

- A rule on the weights that an agent uses when combining its information with the information received from its neighbors.

- A connectivity rule ensuring that the information of each agent influences the information of any other agent infinitely often in time.

- A rule on the frequency at which an agent sends his information to the neighbors.

Specifically, we use the following assumption on the weights $a^i_j(k)$.

Assumption 1 *(Weights Rule)* We have:

(a) There exists a scalar η with $0 < \eta < 1$ such that for all $i \in \{1, \ldots, m\}$,

(i) $a_i^i(k) \geq \eta$ for all $k \geq 0$.

(ii) $a_j^i(k) \geq \eta$ for all $k \geq 0$, and all agents j whose (potentially delayed) information $x^j(s)$ reaches agent i in the time interval (t_k, t_{k+1}).

(iii) $a_j^i(k) = 0$ for all $k \geq 0$ and j otherwise.

(b) The vectors $a^i(k)$ are stochastic, i.e., $\sum_{j=1}^m a_j^i(k) = 1$ for all i and k.

Assumption 1(a) states that each agent gives significant weights to its own estimate $x^i(k)$ and the estimate $x^j(k)$ available from her neighboring agents j at the update time t_k. Note that, under Assumption 1, for the matrix $A(k)$ whose columns are $a^1(k), \ldots, a^m(k)$, the transpose $A'(k)$ is a stochastic matrix for all $k \geq 0$.

We now discuss the rules we impose on the information exchange among agents. Here, it is convenient to view the agents as a set of nodes $V = \{1, \ldots, m\}$. At each update time t_k, the information exchange among the agents may be represented by a directed graph (V, E_k) with the set E_k of directed edges given by

$$E_k = \{(j, i) \mid a_j^i(k) > 0\}.$$

Note that, by Assumption 1(a), we have $(i, i) \in E_k$ for each agent i and all k. Also, we have $(j, i) \in E_k$ if and only if agent i receives information x^j from agent j in the time interval (t_k, t_{k+1}).

We impose a connectivity assumption on the agent system, which can be stated as: following any time t_k, *the information of an agent j reaches each and every agent i directly or indirectly* (through a sequence of communications between the other agents). In other words, the information state of any agent i influences the information state of any other agent infinitely often in time. In formulating this, we use the set E_∞ consisting of edges (j, i) such that j is a neighbor of i who communicates with i infinitely often in time. The connectivity requirement is formally stated in the following assumption.

Assumption 2 *(Connectivity)* The graph (V, E_∞) is connected, where E_∞ is the set of edges (j, i) representing agent pairs communicating directly infinitely many times, i.e.,

$$E_\infty = \{(j, i) \mid (j, i) \in E_k \text{ for infinitely many indices } k\}.$$

To re-phrase, the assumption says that for any k and any two agents $u, v \in V$, there is a directed path from agent u to agent v with edges (j, i) in the set $\cup_{l \geq k} E_l$. Thus, Assumption 2 is equivalent to having the composite directed graph $(V, \cup_{l \geq k} E_l)$ connected for all k.

When analyzing the system state behavior, we use an additional assumption that the intercommunication intervals are bounded for those agents that communicate directly. In particular, we use the following.

Assumption 3 *(Bounded Intercommunication Interval)* There exists an integer $B \geq 1$ such that for every $(j, i) \in E_\infty$, agent j sends information to its neighbor i at least once every B consecutive time slots, i.e., at time t_k or at time t_{k+1} or ... or (at latest) at time t_{k+B-1} for any $k \geq 0$.

When there is no delay in the system, this assumption is equivalent to the requirement that there is $B \geq 1$ such that

$$(j, i) \in E_k \cup E_{k+1} \cup \cdots \cup E_{k+B-1} \qquad \text{for all } (j, i) \in E_\infty \text{ and } k \geq 0.$$

Thus, when there is no delay in the system, our Assumptions 1–3 coincide with those of Tsitsiklis [18] (see also the "consensus" setting of Blondel *et al.* [4] and our optimization model in [13]).

Finally, we assume that the delays $t_j^i(k)$ in delivering a message from an agent j to any neighboring agent i is uniformly bounded at all times.[1] Formally, this is imposed in the following assumption.

Assumption 4 *(Bounded Delays)* Let the following hold:

(a) We have $t_i^i(k) = 0$ for all agents i and all $k \geq 0$.

(b) We have $t_j^i(k) = 0$ for all agents j communicating with agent i directly and whose estimates x^j are not available to agent i at time t_{k+1}.

(c) There is an integer B_1 such that $0 \leq t_j^i(k) \leq B_1 - 1$ for all agents i, j, and all k.

Part (a) of the assumption states that each agent i has its own estimate $x^i(k)$ available (naturally) without any delay. Part (b) states that the delay is zero for those agents j whose (delayed) estimates $x^j(s)$ are not available to agent i at an update time. Under Weights Rule (a) [cf. Assumption 1 (a)], agent i assigns zero weight for the estimate x^j of such an agent j, i.e., $a_j^i(k) = 0$. Thus, under Weights Rule (a), the part (b) of the preceding assumption reduces to the following relation:

$$t_j^i(k) = 0 \qquad \text{for all agents } i \text{ and } j \text{ such that } a_j^i(k) = 0.$$

Finally, part (c) of Assumption 4 states that the delays are uniformly bounded at all times and for all neighboring agents i and j.

3 Convergence Analysis

In this section, we show that the agents updating their information according to Eq. (1) reach a consensus under the assumptions of Section 2.3. In particular, we establish the convergence of agent estimates to a consensus and provide a convergence rate estimate. Our analysis is based on reducing the consensus problem with a delay to a problem without a delay.

[1]The delay bound is used in our analysis. In the implementation of the algorithm, the bound need not be available to any agent.

3.1 Reduction to a Consensus Problem without Delay

Here, we reduce the original agent system with delays to a system without delays, under the Bounded Delays assumption [cf. Assumption 4]. In particular, we define an enlarged agent system that is obtained by adding "new" agents into the original system in order to deal with delays. With each agent i of the original system, we associate a new agent for each of the possible values of the delay that a message originating from agent i may experience. In view of the Bounded Delays assumption, it suffices to add $(m-1)B_1$ new agents handling the delays.[2]

To differentiate between the original agents in the system and the new agents, we introduce the notions of computing and noncomputing agents (or nodes). We refer to the original agents as **computing agents** since these agents maintain and update their information state (estimates). We refer to the new agents as **noncomputing agents** since these agents do not compute or update any information, but only pass the received information to their neighbors.

In the enlarged system, we enumerate the computing agents first and then the noncomputing agents. In particular, the computing agents are indexed by $1, \ldots, m$ and noncomputing agents are indexed by $m+1, \ldots, (B_1-1)m$. Furthermore, the noncomputing agents are indexed so that the first m of them model the delay of 1 for the computing agents, the next m of them model the delay of 2 for the computing agents, and so on. Formally, we have that for a computing agent i, the noncomputing agents $i+m, \ldots, i+(B_1-1)m$ model the nonzero delay values $t = 1, \ldots, (B_1-1)m$, respectively.

We now describe how the agents communicate in the enlarged system, i.e., we identify the neighbors of each agent. The computing agents are connected and communicate in the same way as in the original system. The noncomputing agents corresponding to the delays of different computing agents do not communicate among themselves. Specifically, for t with $1 \leq t < B_1 - 1$, a noncomputing agent $j + tm$ receives the information only from agent $j + (t-1)m$, and sends the same information to either agent $j + (t+1)m$ or to a computing agent i only if agent j communicates with agent i in the original system. A noncomputing agent $j + (B_1 - 1)m$ communicates only with computing agents and in particular, agent $j + (B_1-1)m$ communicates with agent i if and only if j communicates with i in the original system. The communication connections among the agents in the original system and the corresponding enlarged system is illustrated in Figure 1 for a system with 3 agents and a maximum delay of 3.

We let $\tilde{x}^i(k)$ denote the estimate of agent i of the enlarged system at time t_k. Then, the relation in Eq. (1) for the *evolution of estimates of computing agents* is given by: for all $i \in \{1, \ldots, m\}$,

$$\tilde{x}^i(k+1) = \sum_{h=1}^{mB_1} \tilde{a}^i_h(k)\tilde{x}^h(k) \qquad \text{for all } k \geq 0, \tag{2}$$

[2]This idea has also been used in the distributed computation model of Tsitsiklis [18], and it motivates our development here.

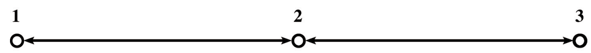

Figure 1 (a)

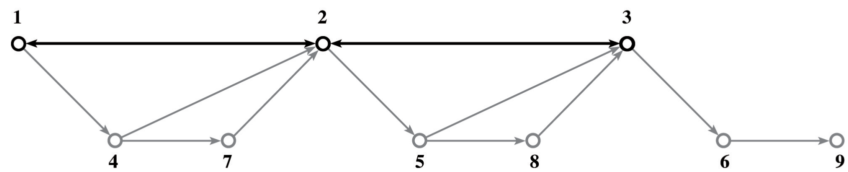

Figure 1 (b)

Figure 1: Figure 1 (a) illustrates an agent network with 3 agents, where agents 1 and 2, and agents 2 and 3 communicate directly. Figure 1(b) illustrates the enlarged network associated with the original network of part (a), when the delay bound is $B_1 = 3$. The noncomputing agents introduced in the system are $4, \ldots, 9$. Agents 4, 5, and 6 model the delay of 1 while agents 7, 8, and 9 model the delay of 2 for the computing nodes 1, 2 and 3, respectively.

where for all $h \in \{1, \ldots, mB_1\}$,

$$\tilde{a}_h^i(k) = \begin{cases} a_j^i(k) & \text{if } h = j + tm,\ t = k - t_j^i(k) \\ 0 & \text{otherwise} \end{cases} \qquad \text{for all } k \geq 0, \tag{3}$$

and $a_j^i(k)$ are the weights used by the agents in the original network. The *evolution of states for noncomputing agents* is given by: for all $i = m+1, \ldots, mB_1$,

$$\tilde{x}^i(k+1) = \tilde{x}^{i-m}(k) \qquad \text{for all } k \geq 0,$$

where the initial values are $\tilde{x}^i(0) = 0$. Therefore, for noncomputing agents i we have

$$\tilde{a}_h^i(k) = \begin{cases} 1 & \text{for } h = i - m \\ 0 & \text{otherwise} \end{cases} \qquad \text{for all } k \geq 0. \tag{4}$$

Using these definitions of weights, we can compactly write the **evolution of estimates $\tilde{x}^i(k)$ for all agents i in the enlarged system** as follows:

$$\tilde{x}^i(k+1) = \sum_{h=1}^{mB_1} \tilde{a}_h^i(k)\tilde{x}^h(k) \qquad \text{for all } i \in \{1, \ldots, mB_1\} \text{ and } k \geq 0, \tag{5}$$

where the initial vectors are given by

$$\begin{aligned} \tilde{x}^i(0) &= x^i(0) && \text{for } i \in \{1, \ldots, m\}, \\ \tilde{x}^i(0) &= 0 && \text{for } i \in \{m+1, \ldots, mB_1\}, \end{aligned} \tag{6}$$

and the weights $\tilde{a}_h^i(k)$ for **computing agents** $i \in \{1, \dots, m\}$ are given by Eq. (3), while the weights $\tilde{a}_h^i(k)$ for **noncomputing agents** $i \in \{m+1, \dots, mB_1\}$ are given by Eq. (4). For a noncomputing agent i, the sum of weights $\sum_{h=1}^{mB_1} \tilde{a}_h^i(k)$ is evidently equal to 1 for any k. For a computing agent $i \in \{1, \dots, m\}$, the sum of weights $\sum_{h=1}^{mB_1} \tilde{a}_h^i(k)$ is equal to 1 for any k *if and only if* the weights $a_j^i(k)$ of agent i in the original network sum to 1, i.e., $\sum_{j=1}^{m} a_j^i(k) = 1$ for all k.

In order to have a more compact representation of the evolution of the estimates $x^i(k)$ of Eq. (5), we rewrite this model using the matrices that govern the (linear) evolution. The resulting representation is also more suitable for our convergence analysis. In particular, we introduce matrices $\tilde{A}(s)$ whose i-th column is the vector $\tilde{a}^i(s)$. Using these matrices, we can relate estimate $\tilde{x}^i(k+1)$ to the estimates $\tilde{x}^1(s), \dots, \tilde{x}^m(s)$ for any $s \le k$. Specifically, it is straightforward to verify that for the iterates generated by Eq. (5), we have for any i, and any s and k with $k \ge s$,

$$\tilde{x}^i(k+1) = \sum_{h=1}^{mB_1} [\tilde{A}(s)\tilde{A}(s+1)\cdots\tilde{A}(k-1)\tilde{a}^i(k)]_h \tilde{x}^h(s). \tag{7}$$

As indicated by the preceding relation, to analyze the convergence of the iterates $\tilde{x}^i(k)$, we need to understand the behavior of the matrix product $\tilde{A}(s)\cdots\tilde{A}(k)$. Actually, this matrix product is the transition matrix for the agent system from time s to time k. We formally introduce these *transition matrices* as follows:

$$\tilde{\Phi}(k,s) = \tilde{A}(s)\tilde{A}(s+1)\cdots\tilde{A}(k-1)\tilde{A}(k) \qquad \text{for all } s \text{ and } k \text{ with } k \ge s, \tag{8}$$

where

$$\tilde{\Phi}(k,k) = \tilde{A}(k) \qquad \text{for all } k. \tag{9}$$

Note that the i-th column of $\tilde{\Phi}(k,s)$ is given by

$$[\tilde{\Phi}(k,s)]^i = \tilde{A}(s)\tilde{A}(s+1)\cdots\tilde{A}(k-1)\tilde{a}^i(k) \qquad \text{for all } i,\ s \text{ and } k \text{ with } k \ge s,$$

while the entry in i-th column and h-th row of $\tilde{\Phi}(k,s)$ is given by

$$[\tilde{\Phi}(k,s)]_h^i = [\tilde{A}(s)\tilde{A}(s+1)\cdots\tilde{A}(k-1)\tilde{a}^i(k)]_h \qquad \text{for all } i,\ h,\ s \text{ and } k \text{ with } k \ge s.$$

We can now rewrite relation (7) compactly in terms of the transition matrices $\tilde{\Phi}(k,s)$, as follows: for any $i \in \{1, \dots, mB_1\}$,

$$\tilde{x}^i(k+1) = \sum_{j=1}^{mB_1} [\tilde{\Phi}(k,s)]_j^i x^j(s) \qquad \text{for all } s \text{ and } k \text{ with } k \ge s \ge 0. \tag{10}$$

Under the Weights Rule [Assumption 1], from the definition of the weights $\tilde{a}_h^i(k)$ in Eqs. (3) and (4), it follows that each matrix $\tilde{A}(k)'$ is stochastic. Since the product of stochastic matrices is a stochastic matrix, it follows that **the transition matrices** $\tilde{\Phi}(k,s)'$ **are stochastic** for all $k \ge s \ge 0$. In what follows, we establish some additional properties of these matrices that will be important in our convergence analysis of the iterates generated by Eq. (10).

3.2 Properties of the Transition Matrices

Here, we explore some properties of the matrices $\tilde{\Phi}(k,s)$ under the assumptions imposed on agent interactions in Section 2.3. As defined in Section 2.3, we use E_∞ to denote the agent pairs communicating (directly) infinitely often in the original network (with delays). In particular, $(j,i) \in E_\infty$ means that agent j communicates its estimates to a neighbor i infinitely often in the presence of delays.

We start by considering the implications for the matrices $\tilde{\Phi}(k,s)$ under Weights Rule and Bounded Delays assumption.

Lemma 1 Let Weights Rule (a) hold for the weights $a_j^i(k)$ for $i,j \in \{1,\ldots,m\}$ and $k \geq 0$ [cf. Assumption 1(a)]. Let also Bounded Delay assumption hold [cf. Assumption 4]. Then:

(a) For a computing node j that sends its information at time s and the information has a nonzero delay t, we have

$$[\tilde{\Phi}(s+t-1,s)]_j^{j+tm} = 1.$$

(b) For any computing node $i \in \{1,\ldots,m\}$, we have

$$[\tilde{\Phi}(k,s)]_i^i \geq \eta^{k-s+1} \qquad \text{for all } k \text{ and } s \text{ with } k \geq s \geq 0.$$

(c) Under Bounded Intercommunication Intervals [cf. Assumption 3], for any two computing nodes $i,j \in \{1,\ldots,m\}$ such that $(j,i) \in E_\infty$, we have

$$[\tilde{\Phi}(s+B+B_1-1,s)]_j^i \geq \eta^{B+B_1}, \qquad \text{for all } s \geq 0,$$

where η is the lower bound on the nonzero weights of Assumption 1 (a), B is the bound on the intercommunication intervals of Assumption 3, and B_1 is the delay bound of Assumption 4(c).

Proof. (a) We prove that $[\tilde{\Phi}(s+t-1,s)]_j^{j+tm} = 1$ by induction on the delay value t. When the delay is $t=1$, we have $\tilde{\Phi}(s,s) = \tilde{A}(s)$. Since $\tilde{a}_{h-m}^h(s) = 1$ for all noncomputing nodes h [see Eq. (4)], it follows that $[\tilde{\Phi}(s,s)]_j^{j+m} = 1$. Suppose now that

$$[\tilde{\Phi}(s+t-1,s)]_j^{j+tm} = 1 \qquad \text{for a delay } t > 1. \tag{11}$$

Consider the case when the delay value is $t+1$. In this case, the path of the information sent from node j at time s is given by $j \to j+m \to j+2m \to \cdots \to j+tm \to j+(t+1)m$. Therefore,

$$[\tilde{\Phi}(s+t,s)]_j^{j+(t+1)m} = [\tilde{\Phi}(s+t-1,s)]_j^{j+tm}\, \tilde{a}(s+t)_{j+tm}^{j+(t+1)m}.$$

By using the inductive hypothesis and the relation $\tilde{a}_{h-m}^h(s) = 1$ for any noncomputing node [cf. Eqs. (11) and (4), respectively], it follows that

$$[\tilde{\Phi}(s+t,s)]_j^{j+(t+1)m} = 1.$$

(b) We let $s \geq 0$ be arbitrary and $i \in \{1, \ldots, m\}$ be an arbitrary computing node. We prove the relation

$$[\tilde{\Phi}(k,s)]_i^i \geq \eta^{k-s+1} \qquad \text{for all } k, s \text{ with } k \geq s \geq 0 \tag{12}$$

by induction on k. For $k = s$, from the definition of $\tilde{\Phi}(s,s)$ in Eq. (9) we see that

$$[\tilde{\Phi}(s,s)]_i^i = \tilde{a}_i^i(s) \qquad \text{for all } s \geq 0.$$

In view of Bounded Delays (a) [cf. Assumption 4 (a)], we have $t_i^i(s) = 0$ for all s. Thus, by the definition of $\tilde{a}_i^i(k)$ [cf. Eq. (3)], it follows that $\tilde{a}_i^i(s) = a_i^i(s)$. Furthermore, by Weights Rule (a) [cf. Assumption 1(a)], we have $a_i^i(s) \geq \eta$ for all s, and therefore,

$$\tilde{a}_i^i(s) = a_i^i(s) \geq \eta \qquad \text{for all } s \geq 0. \tag{13}$$

Hence, $[\tilde{\Phi}(s,s)]_i^i \geq \eta$ for $s \geq 0$, showing that the relation in Eq. (12) holds for $k = s$.

Now, assume that the relation in Eq. (12) holds for some k with $k > s$, and consider $[\tilde{\Phi}(k+1,s)]_i^i$. By the definition of the matrix $\tilde{\Phi}(k,s)$ [cf. Eq. (8)], we have

$$[\tilde{\Phi}(k+1,s)]_i^i = \sum_{h=1}^{mB_1} [\tilde{\Phi}(k,s)]_i^h \; \tilde{a}_h^i(k+1) \geq [\tilde{\Phi}(k,s)]_i^i \; \tilde{a}_i^i(k+1),$$

where the inequality in the preceding relation follows from the nonnegativity of the entries of $\tilde{\Phi}(k,s)$ for all k and s. By using the inductive hypothesis and the relation $\tilde{a}_i^i(k+1) = a_i^i(k+1) \geq \eta$ [cf. Eq. (13)], we obtain

$$[\tilde{\Phi}(k+1,s)]_i^i \geq \eta^{k-s+2}.$$

Hence, the relation in Eq. (12) holds for all $k \geq s$.

(c) Let $s \geq 0$ be arbitrary. Let i and j be two computing nodes with $(j,i) \in E_\infty$. Under Bounded Intercommunication Intervals [cf. Assumption 3], for any such nodes, node j sends its information to node i at time s or $s+1$ or ... or at time $s + B - 1$ at latest. Let the information be sent at time $s + \tau$ with $0 \leq \tau \leq B - 1$.

Suppose that there was no delay. Then

$$\begin{aligned}
[\tilde{\Phi}(s+B-1,s)]_j^i &= \sum_{h=1}^{mB_1} [\tilde{\Phi}(s+\tau,s)]_j^h \; [\tilde{\Phi}(s+B-1,s+\tau+1)]_h^i \\
&\geq [\tilde{\Phi}(s+\tau,s)]_j^i \; [\tilde{\Phi}(s+B-1,s+\tau+1)]_i^i \\
&\geq [\tilde{\Phi}(s+\tau,s)]_j^i \, \eta^{B-\tau-1},
\end{aligned} \tag{14}$$

where the first inequality follows from the nonnegativity of the entries $\tilde{\Phi}(k,s)$ and the last inequality follows from $[\tilde{\Phi}(k,s)]_i^i \geq \eta^{k-s+1}$ for all i, and all k and s with $k \geq s$ [cf. part (b)]. Similarly, by using the result in part (b), we have

$$[\tilde{\Phi}(s+\tau,s)]_j^i = \sum_{h=1}^{mB_1} [\tilde{\Phi}(s+\tau-1,s)]_j^h \, \tilde{a}(s+\tau)_h^i \geq [\tilde{\Phi}(s+\tau-1,s)]_j^j \; \tilde{a}(s+\tau)_j^i \geq \eta^\tau \, \tilde{a}(s+\tau)_j^i.$$

Since the information from j to i is sent at time $s+\tau$ and it arrives without a delay, we have $\tilde{a}(s+\tau)^i_j = a(s+\tau)^i_j$. By the Weights Rule (a), we also have $a(s+\tau)^i_j \ge \eta$, and therefore,

$$[\tilde{\Phi}(s+\tau,s)]^i_j \ge \eta^{\tau+1}.$$

Using the preceding relation and Eq. (14), we obtain $[\tilde{\Phi}(s+B-1,s)]^i_j \ge \eta^B$, which together with the relation $[\tilde{\Phi}(s+B+B_1-1,s+B)]^i_i \ge \eta^{B_1}$ [cf. part (b)], implies that

$$\begin{aligned}[\tilde{\Phi}(s+B+B_1-1,s)]^i_j &= \sum_{h=1}^{mB_1}[\tilde{\Phi}(s+B-1,s)]^h_j\,[\tilde{\Phi}(s+B+B_1-1,s+B)]^i_h \\ &\ge [\tilde{\Phi}(s+B-1,s)]^i_j\,[\tilde{\Phi}(s+B+B_1-1,s+B)]^i_i \\ &\ge \eta^{B+B_1}.\end{aligned}$$

Thus, the relation $[\tilde{\Phi}(s+B+B_1-1,s)]^i_j \ge \eta^{B+B_1}$ holds when the message from j to i is not delayed.

Consider now the case when the information from node j to i is delayed by t with $t \in \{1,\dots,B_1-1\}$. Thus, the information is sent from node j to the noncomputing node $j+m$ at time $s+\tau$ and reaches the computing node i at time $s+\tau+t$. In view of the information communication in the enlarged model, from time $s+\tau$ to $s+\tau+t$, the message path is $j \to j+m \to j+2m \to \cdots \to j+tm \to i$. In terms of the transition matrices, we formally have

$$\begin{aligned}[\tilde{\Phi}(s+\tau+t,s)]^i_j &= \sum_{h=1}^{mB_1}[\tilde{\Phi}(s+\tau-1,s)]^h_j\,[\tilde{\Phi}(s+\tau+t,s+\tau)]^i_h \\ &\ge [\tilde{\Phi}(s+\tau-1,s)]^j_j\,[\tilde{\Phi}(s+\tau+t,s+\tau)]^i_j \\ &\ge \eta^\tau\,[\tilde{\Phi}(s+\tau+t,s+\tau)]^i_j, \end{aligned} \tag{15}$$

where the first inequality follows from the nonnegativity of the entries in $\tilde{\Phi}(k,s)$ and the last inequality follows from $[\tilde{\Phi}(s+\tau-1,s)]^j_j \ge \eta^\tau$ [cf. part (b)]. We now consider the term $[\tilde{\Phi}(s+\tau+t,s+\tau)]$, for which we have

$$\begin{aligned}[\tilde{\Phi}(s+\tau+t,s+\tau)]^i_j &= \sum_{h=1}^{mB_1}[\tilde{\Phi}(s+\tau+t-1,s+\tau)]^h_j\,\tilde{a}(s+\tau+t)^i_h \\ &\ge [\tilde{\Phi}(s+\tau+t-1,s+\tau)]^{j+tm}_j\,\tilde{a}(s+\tau+t)^i_{j+tm} \\ &\ge [\tilde{\Phi}(s+\tau+t-1,s+\tau)]^{j+tm}_j\,\eta,\end{aligned}$$

where the last inequality follows from $\tilde{a}^i_h(k) = a^i_j(k)$ for $h = j+tm$ and all k [cf. Eq. (3)], the assumption that $a^i_j(k) \ge \eta$ for all k [cf. Assumption 1(a)], and the fact that the information arrives to node i from node $j+tm$ at time $k = s+\tau+t$. By part (a), we have $[\tilde{\Phi}(s+t-1,s)]^{j+tm}_j = 1$ for any computing node j that sends its information at time s and the information has a delay $t > 0$. Therefore, we obtain

$$[\tilde{\Phi}(s+\tau+t,s+\tau)]^i_j \ge \eta,$$

and in view of Eq. (15), it follows that

$$[\tilde{\Phi}(s+\tau+t,s)]_j^i \geq \eta^{\tau+1}.$$

By part (b), for computing node i, we have

$$[\tilde{\Phi}(s+B+B_1-1,s+\tau+t+1)]_i^i \geq \eta^{B+B_1-\tau-t-1}.$$

By using the preceding two relations, we have

$$\begin{aligned}
[\tilde{\Phi}(s+B+B_1-1,s)]_j^i &= \sum_{h=1}^{mB_1}[\tilde{\Phi}(s+\tau+t,s)]_j^h \, [\tilde{\Phi}(s+B+B_1-1,s+\tau+t+1)]_h^i \\
&\geq [\tilde{\Phi}(s+\tau+t,s)]_j^i \, [\tilde{\Phi}(s+B+B_1-1,s+\tau+t+1)]_i^i \\
&\geq \eta^{\tau+1}\,\eta^{B+B_1-\tau-t-1} \\
&= \eta^{B+B_1-t}.
\end{aligned}$$

Since $0<\eta<1$ [cf. Assumption 1] and $B+B_1-t<B+B_1$ for all nonzero delay values t, we have $\eta^{B+B_1-t}>\eta^{B+B_1}$, implying that

$$[\tilde{\Phi}(s+B+B_1-1,s)]_j^i \geq \eta^{B+B_1}.$$

■

Using Lemma 1, we next establish some additional properties of the transition matrices $\tilde{\Phi}(k,s)$. These properties hold when the agents are connected in the sense of Assumption 2. More specifically, we show that the entries of the row $[\tilde{\Phi}(s+(m-1)B+mB_1-1,s)]_j$ are uniformly bounded away from zero for all s and for all computing nodes $j\in\{1,\ldots,m\}$.

Lemma 2 Let Weights Rule (a), Connectivity, Bounded Intercommunication Interval, and Bounded Delay assumptions hold for the agents in the original network [cf. Assumptions 1(a), 2, 3, and 4]. Then, the following holds:

(a) For any computing nodes $i,j\in\{1,\ldots,m\}$, we have

$$[\tilde{\Phi}(k,s)]_j^i \geq \eta^{k-s+1} \qquad \text{for all } s\geq 0, \text{ and } k\geq s+(m-1)(B+B_1).$$

(b) For any computing node $j\in\{1,\ldots,m\}$, we have

$$[\tilde{\Phi}(s+(m-1)B+mB_1-1,s)]_j^i \geq \eta^{(m-1)B+mB_1} \qquad \text{for all nodes } i \text{ and all } s\geq 0.$$

Proof. For parts (a)-(b), we let $s\geq 0$ be arbitrary, but fixed.

(a) Let i and j be any two computing nodes. When $i=j$, by Lemma 1 (b), we have

$$[\tilde{\Phi}(k,s)]_i^i \geq \eta^{k-s+1} \qquad \text{for all } k\geq s.$$

Thus, assume that the nodes i and j are distinct.

Under the Connectivity [cf. Assumption 2], there is a directed path from node j to node i passing through some other computing nodes $i_1, \ldots, i_\kappa$ such that the nodes $j, i_1, \ldots, i_\kappa, i$ are distinct and the path edges $(j, i_1), (i_1, i_2), \ldots, (i_{\kappa-1}, i_\kappa), (i_\kappa, i)$ belong to the set E_∞. Let us re-label node i by $i_{\kappa+1}$, and let $B_2 = B + B_1$. By the definition of the transition matrix $\tilde{\Phi}$ [cf. Eq. (8)], we have

$$\begin{aligned}
&[\tilde{\Phi}(s+(\kappa+1)B_2-1,s)]_j^{i_{\kappa+1}} \\
&\quad = \sum_{h=1}^{mB_1}[\tilde{\Phi}(s+\kappa B_2-1,s)]_j^h\ [\tilde{\Phi}(s+(\kappa+1)B_2-1,s+\kappa B_2)]_h^{i_{\kappa+1}} \\
&\quad \geq [\tilde{\Phi}(s+\kappa B_2-1,s)]_j^{i_\kappa}\ [\tilde{\Phi}(s+(\kappa+1)B_2-1,s+\kappa B_2)]_{i_\kappa}^{i_{k+1}} \\
&\quad \geq [\tilde{\Phi}(s+\kappa B_2-1,s)]_j^{i_\kappa}\ \eta^{B_2},
\end{aligned}$$

where the first inequality follows from the nonnegativity of the entries of $\tilde{\Phi}$ and the last inequality follows from $[\tilde{\Phi}(s+(\kappa+1)B_2-1,s+\kappa B_2)]_{i_k}^{i_{k+1}} \geq \eta^{B_2}$ [cf. Lemma 1 part (c) and $B_2 = B + B_1$]. Hence, it follows that

$$[\tilde{\Phi}(s+(\kappa+1)B_2-1,s)]_j^{i_{\kappa+1}} \geq [\tilde{\Phi}(s+\kappa B_2-1,s)]_j^{i_\kappa}\ \eta^{B_2} \geq \cdots \geq [\tilde{\Phi}(s+B_2-1,s)]_j^{i_1}\ \eta^{\kappa B_2}.$$

By Lemma 1 (c) and $B_2 = B + B_1$, we have $[\tilde{\Phi}(s+B_2-1,s)]_j^{i_1} \geq \eta^{B_2}$, and therefore, $[\tilde{\Phi}(s+(\kappa+1)B_2-1,s)]_j^{i_{\kappa+1}} \geq \eta^{(\kappa+1)B_2}$. Since $i_{\kappa+1} = i$, we have

$$[\tilde{\Phi}(s+(\kappa+1)B_2-1,s)]_j^i \geq \eta^{(\kappa+1)B_2}.$$

Since there are m agents, and the nodes $j, i_1, \ldots, i_\kappa, i$ are distinct, it follows that $\kappa+2 \leq m$. By using the preceding relation and the definition of the transition matrix $\tilde{\Phi}$ [cf. Eq. (8)], we obtain for all $k \geq s + (m-1)B_2$,

$$\begin{aligned}
[\tilde{\Phi}(k,s)] &= \sum_{h=1}^{mB_1}[\tilde{\Phi}(s+(\kappa+1)B_2-1,s)]_j^h\ [\tilde{\Phi}(k,s+(\kappa+1)B_2)]_h^i \\
&\geq [\tilde{\Phi}(s+(\kappa+1)B_2-1,s)]_j^i\ [\tilde{\Phi}(k,s+(\kappa+1)B_2)]_i^i \\
&\geq \eta^{(\kappa+1)B_2}\ \eta^{k-s-(\kappa+1)B_2+1} \\
&= \eta^{k-s+1},
\end{aligned}$$

where in the last inequality we also use $[\tilde{\Phi}(k,s)]_i^i \geq \eta^{k-s+1}$ for all k, s with $k \geq s \geq 0$ [cf. Lemma 1 part (b)].

(b) For any computing nodes $i,\ j \in \{1, \ldots, m\}$, by part (a), we have

$$[\tilde{\Phi}(k,s)]_j^i \geq \eta^{k-s+1} \qquad \text{for all } s \geq 0 \text{ and } k \geq s + (m-1)(B+B_1).$$

Since $s + (m-1)B + mB_1 - 1 > s + (m-1)(B+B_1) - 1$, it follows that

$$[\tilde{\Phi}(s+(m-1)B+mB_1-1,s)]_j^i \geq \eta^{(m-1)B+mB_1}.$$

Thus, the desired relation holds for any computing nodes j and i.

Assume now that j is a computing agent and i is an arbitrary noncomputing agent, i.e., $i \in \{m+1, \ldots, mB_1\}$. We can express i as

$$i = l + mt \qquad \text{for some } l \in \{1, \ldots, m\} \text{ and } t \in \{1, \ldots, B_1 - 1\}. \tag{16}$$

By the definition of the matrix $\tilde{\Phi}$, we have

$$[\tilde{\Phi}(s + (m-1)B + mB_1 - 1, s)]_j^{l+mt} \geq$$
$$[\tilde{\Phi}(s + (m-1)B + mB_1 - 2, s)]_j^{l+m(t-1)} \tilde{a}_{l+m(t-1)}^{l+mt}(s + (m-1)B + mB_1 - 1).$$

Since $a_{l+m(t-1)}^{l+mt} = 1$ for a noncomputing agent $l + mt$ [cf. Eq. (4)], the preceding relation implies

$$\tilde{\Phi}(s + (m-1)B + mB_1 - 1, s)]_j^{l+mt} \geq [\tilde{\Phi}(s + (m-1)B + mB_1 - 2, s)]_j^{l+m(t-1)}.$$

If $t = 1$, we are done. Otherwise, repeating the same procedure recursively yields

$$[\tilde{\Phi}(s + (m-1)B + mB_1 - 1, s)]_j^i \geq [\tilde{\Phi}(s + (m-1)B + mB_1 - 1 - t, s)]_j^l. \tag{17}$$

Since $l \in \{1, \ldots, m\}$ and $s + (m-1)B + mB_1 - 1 - t > s + (m-1)(B + B_1) - 1$ [cf. Eq. (16)], it follows by part (a) that

$$[\tilde{\Phi}(s + (m-1)B + mB_1 - 1 - t, s)]_j^l \geq \eta^{(m-1)B + mB_1 - t} \geq \eta^{(m-1)B + mB_1},$$

where the last inequality follows since $t \geq 0$ and $\eta \in (0,1)$. ∎

We next present a result for stochastic matrices which will be used in establishing the convergence properties of the transition matrices $\tilde{\Phi}$.

Lemma 3 Let D be a stochastic matrix. Assume that D has a column with all entries bounded away from zero i.e., for some index j and some scalar $\nu > 0$, there holds $[D]_i^j \geq \nu$ all i. Let z be a nonnegative vector. Then, for any i^*, the following relation holds:

$$([D]_i - [D]_{i^*})'z \leq (1 - \nu)\|z\|_\infty \qquad \text{for all } i.$$

Proof. We assume without loss of generality that the entries in the first column of D are bounded away from zero by ν. For an arbitrary h, we define the index sets:

$$I^+ = \{i \mid [D]_h^i - [D]_{j^*}^i > 0\}, \qquad I^- = \{i \mid [D]_h^i - [D]_{j^*}^i < 0\}.$$

Assume first that $1 \in I^+$. Using the index sets I^+ and I^-, we can write

$$\begin{aligned}([D]_h - [D]_{j^*})'z = \sum_i ([D]_h^i - [D]_{j^*}^i) z_i &= \sum_{i \in I^+} ([D]_h^i - [D]_{j^*}^i) z_i + \sum_{i \in I^-} ([D]_h^i - [D]_{j^*}^i) z_i \\ &\leq \sum_{i \in I^+} ([D]_h^i - [D]_{j^*}^i) z_i,\end{aligned}$$

where the inequality follows from the nonnegativity of z_i. Since $[D]_h^i - [D]_{j^*}^i > 0$ for all $i \in I^+$, we have

$$\sum_{i\in I^+}([D]_h^i - [D]_{j^*}^i)z_i \le \sum_{i\in I^+}([D]_h^i - [D]_{j^*}^i)\|z\|_\infty = \left(\sum_{i\in I^+}[D]_h^i - \sum_{i\in I^+}[D]_{j^*}^i\right)\|z\|_\infty.$$

By the stochasticity of the vector $[D]_h$, we have that $\sum_{i\in I^+}[D]_h^i \le 1$. By the nonnegativity of the entries $D_{j^*}^i$ and the fact $i \in I^+$, we obtain $\sum_{i\in I^+}[D]_{j^*}^i \ge D_{j^*}^1$. Therefore,

$$\sum_{i\in I^+}([D]_h^i - [D]_{j^*}^i)z_i \le (1 - D_{j^*}^1)\|z\|_\infty \le (1-\nu)\|z\|_\infty,$$

where the last inequality follows from the assumption $D_{j^*}^1 > \nu$.

Assume now that $1 \notin I^+$. Using the nonnegativity of the vectors $[D]_h$, $[D]_{j^*}$, and z, we can write

$$([D]_h - [D]_{j^*})'z \le \sum_{i\in I^+}([D]_h^i - [D]_{j^*}^i)z_i \le \sum_{i\in I^+}[D]_h^i z_i \le \sum_{i\in I^+}[D]_h^i\|z\|_\infty.$$

Since D is stochastic and $1 \notin I^+$, we have

$$\sum_{i\in I^+}[D]_h^i \le \sum_{i\neq 1}[D]_h^i \le 1-\nu.$$

Combining the preceding two relations, we obtain

$$([D]_h - [D]_{j^*})'z \le (1-\nu)\|z\|_\infty.$$

■

We now give a lemma that plays a key role in establishing the convergence properties and assessing the convergence rate of the matrices $\tilde{\Phi}(k,s)$. In the lemma, we consider the products $\tilde{D}_k(s)\cdots\tilde{D}_1(s)$ of the matrices

$$\tilde{D}_k(s) = \tilde{\Phi}'\left(s + kB_2 - 1, s + (k-1)B_2\right),$$

where $B_2 = (m-1)B + mB_1$, and we show that these products converge as k increases to infinity. We use this later to establish the convergence of the composite weights $[\tilde{\Phi}(k,s)]^i$, as $k \to \infty$, for each computing agent i.

Lemma 4 Let Weights Rule, Connectivity, Bounded Intercommunication Interval, and Bounded Delay assumptions hold for the agents in the original network [cf. Assumptions 1, 2, 3, and 4].

$$\tilde{D}_k(s) = \tilde{\Phi}'\left(s + kB_2 - 1, s + (k-1)B_2\right) \qquad \text{for all } k \ge 1, \tag{18}$$

where $B_2 = (m-1)B + mB_1$. We then have:

(a) The limit $\tilde{D}(s) = \lim_{k\to\infty}\tilde{D}_k(s)\cdots\tilde{D}_1(s)$ exists.

(b) The limit $\tilde{D}(s)$ is an $mB_1 \times mB_1$ matrix whose rows are identical stochastic vectors (function of s) i.e.,

$$\tilde{D}(s) = e\tilde{\phi}'(s)$$

where $\tilde{\phi}(s) \in \mathbb{R}^{mB_1}$ is a stochastic vector.

(c) The convergence of $\tilde{D}_k(s) \cdots \tilde{D}_1(s)$ to $\tilde{D}(s)$ is geometric:

$$\left\| \tilde{D}_k(s) \cdots \tilde{D}_1(s)x - \tilde{D}(s)x \right\|_\infty \leq 2\left(1 + \eta^{-B_2}\right)\left(1 - \eta^{B_2}\right)^k \|x\|_\infty,$$

for every $x \in \mathbb{R}^{mB_1}$ and for all $k \geq 1$. In particular, for each $j \in \{1, \ldots, mB_1\}$, the entries $[\tilde{D}_k(s) \cdots \tilde{D}_1(s)]_i^j$ for $i \in \{1, \ldots, mB_1\}$, converge to the same limit $\tilde{\phi}_j(s)$ as $k \to \infty$ with a geometric rate: for each $j \in \{1, \ldots, mB_1\}$,

$$\left| [\tilde{D}_k(s) \cdots \tilde{D}_1(s)]_i^j - \tilde{\phi}_j(s) \right| \leq 2\left(1 + \eta^{-B_2}\right)\left(1 - \eta^{B_2}\right)^k,$$

for all $i \in \{1, \ldots, mB_1\}$, and all $k \geq 1$ and $s \geq 0$.

Proof. To simplify our notation in the proof, we suppress the explicit dependence of the matrices $\tilde{D}_k(s)$ on s.

(a) We prove that the limit of $\tilde{D}_k \cdots \tilde{D}_1$ exists by showing that the sequence $\{\tilde{D}_k \cdots \tilde{D}_1 x\}$ converges for every $x \in \mathbb{R}^{mB_1}$. For this, we let $x \in \mathbb{R}^{mB_1}$ be arbitrary, and we consider the vector sequence $\{x_k\} \subseteq \mathbb{R}^{mB_1}$ defined by

$$x_k = \tilde{D}_k \cdots \tilde{D}_1 x \qquad \text{for } k \geq 1.$$

We recursively decompose each vector x_k in the following form:

$$x_k = z_k + c_k e \quad \text{with} \quad z_k \geq 0 \qquad \text{for all } k \geq 0, \tag{19}$$

where $e \in \mathbb{R}^{mB_1}$ is the vector with all entries equal to 1 and $x_0 = x$. The recursion is initialized with

$$z_0 = x - \min_{1 \leq i \leq mB_1} [x]_i \qquad \text{and} \qquad c_0 = \min_{1 \leq i \leq mB_1} [x]_i. \tag{20}$$

Having the decomposition for x_k, we consider the vector $x_{k+1} = \tilde{D}_{k+1} x_k$. In view of relation (19) and the stochasticity of $\tilde{D}_{k+1}$, we have

$$x_{k+1} = \tilde{D}_{k+1} z_k + c_k e.$$

We define

$$z_{k+1} = \tilde{D}_{k+1} z_k - \left([\tilde{D}_{k+1}]'_{i^*} z_k \right) e, \tag{21}$$

$$c_{k+1} = [\tilde{D}_{k+1}]'_{i^*} z_k + c_k. \tag{22}$$

where i^* is the index of the row vector $[\tilde{D}_{k+1}]_i$ achieving the minimum of inner products $[\tilde{D}_{k+1}]'_i z_k$ over all $i \in \{1, \ldots, mB_1\}$. Clearly, we have $x_{k+1} = z_{k+1} + c_{k+1} e$ and $z_{k+1} \geq 0$.

By the definition of z_{k+1} in Eq. (21) it follows that for the components $[z_{k+1}]_i$ we have

$$[z_{k+1}]_i = [\tilde{D}_{k+1}]_i' z_k - [\tilde{D}_{k+1}]_{i^*}' z_k \qquad \text{for all } i \in \{1, \ldots, mB_1\}, \tag{23}$$

where $[\tilde{D}_{k+1}]_i$ is the i-th row vector of the matrix $\tilde{D}_{k+1}$. By Lemma 2 and the definition of the matrices $\tilde{D}_k$ [cf. Eq. (18)], we have that the first m columns of each matrix $\tilde{D}_k$ are bounded away from zero, i.e., for all $j \in \{1, \ldots, m\}$ and $i \in \{1, \ldots, mB_1\}$,

$$[\tilde{D}_{k+1}]_i^j \geq \eta^{B_2} \qquad \text{for all } k \geq 0.$$

Then, from relation (23) and Lemma 3, we have for all $i \in \{1, \ldots, mB_1\}$,

$$[z_{k+1}]_i = \left([\tilde{D}_{k+1}]_i - [\tilde{D}_{k+1}]_{i^*}\right)' z_k \leq (1 - \eta^{B_2}) \, \|z_k\|_\infty.$$

Because $z_{k+1} \geq 0$, it follows

$$\|z_{k+1}\|_\infty \leq \left(1 - \eta^{B_2}\right) \|z_k\|_\infty \qquad \text{for all } k \geq 0,$$

implying that

$$\|z_k\|_\infty \leq \left(1 - \eta^{B_2}\right)^k \|z_0\|_\infty \qquad \text{for all } k \geq 0. \tag{24}$$

Hence $z_k \to 0$ with a geometric rate.

Consider now the sequence $\{c_k\}$ satisfying Eq. (22), for which by the nonnegativity of the vector z_k and the stochasticity of $\tilde{D}_{k+1}$, we have

$$0 \leq c_{k+1} - c_k \leq [\tilde{D}_{k+1}]_{i^*}' z_k \leq \sum_{j=1}^{mB_1} [\tilde{D}_{k+1}]_{i^*}^j \, \|z_k\|_\infty = \|z_k\|_\infty \leq \left(1 - \eta^{B_2}\right)^k \|z_0\|_\infty,$$

where the last inequality in the preceding relation follows from the relation in Eq. (24). Therefore, for any $k \geq 1$ and $r \geq 1$,

$$c_{k+r} - c_k \leq c_{k+r} - c_{k+r-1} + \cdots + c_{k+1} - c_k \leq (q^{k+r-1} + \cdots + q^k) \|z_0\|_\infty = \frac{1 - q^r}{1 - q} \, q^k \|z_0\|_\infty,$$

where $q = 1 - \eta^{B_2}$. Hence, $\{c_k\}$ is a Cauchy sequence and therefore, it converges to some $\tilde{c} \in \mathbb{R}$. By letting $r \to \infty$ in the preceding relation, we obtain

$$\tilde{c} - c_k \leq \frac{q^k}{1 - q} \, \|z_0\|_\infty \qquad \text{for all } k \geq 0. \tag{25}$$

From the decomposition of x_k [cf. Eq. (19)], and the relations $z_k \to 0$ and $c_k \to \tilde{c}$, it follows that $(\tilde{D}_k \cdots \tilde{D}_1)x \to \tilde{c}e$ for any $x \in \mathbb{R}^{mB_1}$, with $\tilde{c}$ being a function of x. Therefore, the limit of $\tilde{D}_k \cdots \tilde{D}_1$ as $k \to \infty$ exists. We denote this limit by $\tilde{D}$, for which we have

$$\tilde{D}x = \tilde{c}(x)e \qquad \text{for all } x \in \mathbb{R}^{mB_1}. \tag{26}$$

(b) Since each $\tilde{D}_k$ is stochastic, each finite product matrix $\tilde{D}_k \cdots \tilde{D}_1$ is stochastic, and therefore the limit matrix $\tilde{D}$ is also stochastic, i.e., $\tilde{D}e = e$. Furthermore, the limit matrix $\tilde{D}$ has rank one [cf. Eq. (26)]. Thus, all its rows are collinear, and because each of its rows sum to 1, it follows that all rows of $\tilde{D}$ are identical. Therefore, for some stochastic vector $\tilde{\phi} \in \mathbb{R}^{mB_1}$, we have $\tilde{D} = e\tilde{\phi}'$.

(c) Let $x \in \mathbb{R}^{mB_1}$ be arbitrary and let $x_k = (\tilde{D}_k \cdots \tilde{D}_1)x$. By using the decomposition of x_k given in Eq. (19), we have

$$(\tilde{D}_k \cdots \tilde{D}_1)x - \tilde{D}x = z_k + (c_k - \tilde{c})e \qquad \text{for all } k \geq 1,$$

[cf. Eq. (26) where the explicit dependence on x in $\tilde{c}(x)$ is suppressed]. Using the estimates in Eqs. (24) and (25), we obtain for all $k \geq 1$,

$$\left\| (\tilde{D}_k \cdots \tilde{D}_1)x - \tilde{D}x \right\|_\infty \leq \|z_k\|_\infty + |c_k - \bar{c}| \leq \left(1 + \frac{1}{1-q}\right) q^k \, \|z_0\|_\infty.$$

Since $\|z_0\|_\infty \leq 2\|x\|_\infty$ [cf. Eq. (20)] and $q = 1 - \eta^{B_2}$, it follows

$$\left\| (\tilde{D}_k \cdots \tilde{D}_1)x - \tilde{D}x \right\|_\infty \leq 2\left(1 + \eta^{-B_2}\right)\left(1 - \eta^{B_2}\right)^k \|x\|_\infty \qquad \text{for all } k \geq 1, \tag{27}$$

establishing the first relation of part (c).

To show the second relation of part (c) of the lemma, let $j \in \{1, \ldots, mB_1\}$ be arbitrary. By letting $x = e_j$ in Eq. (27), and by using $\tilde{D} = e\tilde{\phi}'$ and $\|e_j\|_\infty = 1$, we obtain

$$\left\| [D_k \cdots D_1]^j - \tilde{\phi}_j e \right\|_\infty \leq 2\left(1 + \eta^{-B_2}\right)\left(1 - \eta^{B_2}\right)^k \qquad \text{for all } k \geq 1,$$

implying that for all $i \in \{1, \ldots, mB_1\}$,

$$\left| [D_k \cdots D_1]^j_i - \tilde{\phi}_j \right| \leq 2\left(1 + \eta^{-B_2}\right)\left(1 - \eta^{B_2}\right)^k \qquad \text{for all } k \geq 1.$$

■

We next use Lemma 4 to establish the convergence properties of the matrices $\tilde{\Phi}(k, s)$ for arbitrary s, as k goes to infinity. In particular, the following lemma states that the matrices $\tilde{\Phi}(k, s)$ have the same limit as the matrices $[\tilde{D}_k(s) \cdots D_k(s)]'$, when k increases to infinity. The proof is omitted since it is identical to the proof of a similar result for the case of no delay in Lemma 4 of our work [13].

Lemma 5 Let Weights Rule, Connectivity, Bounded Intercommunication Interval, and Bounded Delay assumptions hold [cf. Assumptions 1, 2, 3, and 4]. We then have:

(a) The limit $\tilde{\Phi}(s) = \lim_{k\to\infty} \tilde{\Phi}(k, s)$ exists for each s.

(b) The limit matrix $\tilde{\Phi}(s)$ has identical columns and the columns are stochastic, i.e.,

$$\tilde{\Phi}(s) = \tilde{\phi}(s)e',$$

where $\tilde{\phi}(s) \in \mathbb{R}^{mB_1}$ is a stochastic vector for each s.

(c) For each $j \in \{1, \ldots, mB_1\}$, the entries $[\tilde{\Phi}(k, s)]^i_j$, $i = 1, ..., mB_1$, converge to the same limit $\tilde{\phi}_j(s)$ as $k \to \infty$ with a geometric rate, i.e., for each $j \in \{1, \ldots, mB_1\}$ and all $s \geq 0$,

$$\left| [\tilde{\Phi}(k, s)]^i_j - \tilde{\phi}_j(s) \right| \leq 2\, \frac{1 + \eta^{-B_2}}{1 - \eta^{B_2}} \left(1 - \eta^{B_2}\right)^{\frac{k-s}{B_2}} \quad \text{for all } k \geq s \text{ and } i \in \{1, \ldots, mB_1\},$$

where $B_2 = (m-1)B + mB_1$.

3.3 Convergence Result

In this section, we prove the convergence of the iterates of Eq. (1) to a consensus and we provide a convergence rate estimate. Lemma 5 plays a key role in establishing these results.

Proposition 1 Let Weights Rule, Connectivity, Bounded Intercommunication Interval, and Bounded Delay assumptions hold [cf. Assumptions 1, 2, 3, and 4]. We then have:

(a) The sequences $\{x^i(k)\}, i = 1, \ldots, m$ generated by Eq. (1) converge to a consensus, i.e., there holds
$$\lim_{k\to\infty} x^i(k) = \bar{x} \qquad \text{for all } i.$$

(b) The consensus vector $\bar{x} \in \mathbb{R}^m$ is a nonnegative combination of the agent initial vectors $x^j(0), j = 1, \ldots, m$, i.e.,
$$\bar{x} = \sum_{j=1}^{m} w_j x^j(0),$$
with scalars $w_j \geq 0$ for all $j = 1, \ldots, m$, and such that $\sum_{j=1}^{m} w_j \leq 1$.

(c) The convergence rate to the consensus is geometric: for all agents $i \in \{1, \ldots, m\}$,
$$\|x^i(k+1) - \bar{x}\| \leq 2\, \frac{1+\eta^{-B_2}}{1-\eta^{B_2}} \left(1-\eta^{B_2}\right)^{\frac{k}{B_2}} \sum_{j=1}^{m} \|x^j(0)\| \qquad \text{for all } k \geq 0,$$
where $B_2 = (m-1)B + mB_1$.

Proof. In view of the equivalence between the evolution equations (1) for the original system and the evolution equations (5) for enlarged system, it suffices to show that the limit $\lim_{k\to\infty} \tilde{x}^i(k)$ exists for all computing nodes $i \in \{1, \ldots, m\}$, and that these limits are the same for all $i \in \{1, \ldots, m\}$.

To show this, we consider the compact representation of Eq. (5), which is given in Eq. (10). Letting $s = 0$ in relation (10), we obtain for any agent $i = \{1, \ldots, mB_1\}$ in the enlarged system,
$$\tilde{x}^i(k+1) = \sum_{j=1}^{mB_1} [\tilde{\Phi}(k,0)]^i_j x^j(0) \qquad \text{for all } k \geq 0.$$

Recall that the initial vectors $\tilde{x}^i(0)$ for computing agents $i \in \{1, \ldots, m\}$ in the enlarged system are the same as the initial vectors $x^i(0)$ of agents i in the original system, and that the initial vectors $\tilde{x}^i(0)$ are zero for noncomputing agents $i \in \{m+1, \ldots, mB_1\}$ [cf. Eq. (6)]. Thus, we have
$$\tilde{x}^i(k+1) = \sum_{j=1}^{m} [\tilde{\Phi}(k,0)]^i_j x^j(0) \qquad \text{for all } k \geq 0. \tag{28}$$

By Lemma 5, we have for all j,

$$\left|[\tilde{\Phi}(k,0)]^i_j - \tilde{\phi}_j(0)\right| \le 2\,\frac{1+\eta^{-B_2}}{1-\eta^{B_2}}\,\left(1-\eta^{B_2}\right)^{\frac{k}{B_2}} \quad \text{for all } k \ge 0 \text{ and } i \in \{1,\ldots,mB_1\},$$

where $\tilde{\phi}(0) \in \mathbb{R}^{mB_1}$ is a stochastic vector with components $\tilde{\phi}_j(0)$. Therefore,

$$\lim_{k\to\infty} \tilde{x}^i(k+1) = \sum_{j=1}^{m} \tilde{\phi}_j(0) x^j(0) \qquad \text{for all } i = 1,\ldots,m,$$

showing that

$$\bar{x} = \sum_{j=1}^{m} \tilde{\phi}_j(0) x^j(0). \tag{29}$$

This establishes the results in parts (a) and (b), where $w_j = \tilde{\phi}_j(0)$ and the properties of the weights w_j follow from the stochasticity of the vector $\tilde{\phi}(0)$.

From relations (28) and (29), we obtain for any $i \in \{1,\ldots,m\}$ and any $k \ge 0$,

$$\|\tilde{x}^i(k+1) - \bar{x}\| \le \sum_{j=1}^{m} \left|[\tilde{\Phi}(k,0)]^i_j - \tilde{\phi}_j(0)\right| \|x^j(0)\| \le \max_{1\le h\le m} \left|[\tilde{\Phi}(k,0)]^i_h - \tilde{\phi}_h(0)\right| \sum_{j=1}^{m} \|x^j(0)\|.$$

The rate estimate in part (c) follows immediately from the preceding relation and part (c) of Lemma 5. ■

As indicated in the proof of the preceding proposition [cf. Eqs. (28) and (29)], the consensus value $\bar{x}$ is a function of the vector $\tilde{\phi}(0)$ defining the limit of the transition matrices $\tilde{\Phi}(k,0)$ as $k \to \infty$, i.e., the vector $\tilde{\phi}(0)$ for which

$$\lim_{k\to\infty} \tilde{\Phi}(k,0) = \tilde{\phi}(0)e'.$$

The transition matrices depend on the maximum delay B_1 in the system, and therefore the vector $\tilde{\phi}(0)$ is also a function of B_1. Hence, the consensus value $\bar{x}$ implicitly depends on the delay bound B_1 through the vector $\tilde{\phi}(0)$. This dependence can be made more explicit by focusing on more structured weight choices $a^i_j(k)$ and topologies for the agent communication network.

4 Conclusions

We considered an algorithm for the consensus problem in the presence of delay in the agent values. Our analysis relies on reducing the problem to a consensus problem in an enlarged agent system without delays. We studied properties of the reduced model and through these properties, we established the convergence and rate of convergence properties for the consensus problem with delays. Our convergence rate estimate is explicitly given in terms of the system parameters. Furthermore, our rate result shows a geometric convergence to consensus. Future work includes incorporating the delayed consensus algorithm in the distributed optimization framework developed in [13] to account for delays in agent values.

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*, Athena Scientific, Belmont, MA, 1997.

[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] V.D. Blondel, J.M. Hendrickx, A. Olshevsky, and J.N. Tsitsiklis, *Convergence in multiagent coordination, consensus, and flocking*, Proceedings of IEEE CDC, 2005.

[5] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Gossip algorithms: Design, analysis, and applications*, Proceedings of IEEE INFOCOM, 2005.

[6] 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.

[7] R. Carli, F. Fagnani, T. Taylor P. Frasca, and S. Zampieri, *Average consensus on networks with transmission noise or quantization*, Proceedings of European Control Congerence, 2007.

[8] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, *Communication constraints in coordinated consensus problems*, Proceedings of IEEE ACC, 2006.

[9] 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.

[10] A. Kashyap, T. Basar, and R. Srikant, *Quantized consensus*, Proceedings of IEEE CDC, 2006.

[11] S. Li and T. Basar, *Asymptotic agreement and convergence of asynchronous stochastic algorithms*, IEEE Transactions on Automatic Control **32** (1987), no. 7, 612–618.

[12] C. Moallemi and B. Van Roy, *Consensus propagation*, Proceedings of Neural Information Processing Systems, 2005.

[13] A. Nedić and A. Ozdaglar, *Distributed subgradient methods for multi-agent optimization*, Tech. Report LIDS 2755, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 2007.

[14] ———, *On the rate of convergence of distributed subgradient methods for multi-agent optimization*, Proceedings of IEEE CDC, 2007.

[15] 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.

[16] A. Olshevsky and J.N. Tsitsiklis, *Convergence rates in distributed consensus and averaging*, Proceedings of IEEE CDC, 2006.

[17] ______, *Convergence speed in distributed consensus and averaging*, http://arxiv.org/abs/math.OC/0612682v1, 2006.

[18] J.N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, Dept. of Electrical Engineering and Computer Science Massachusetts Institute of Technology, 1984.

[19] 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.

[20] 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.

[21] L. Xiao and S. Boyd, *Fast linear iterations for distributed averaging*, Systems and Control Letters **53** (2004), 65–78.