Automatica 48 (2012) 2750–2761

Contents lists available at SciVerse ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Average consensus on general strongly connected digraphs☆

Kai Cai [a,1], Hideaki Ishii [b]

[a] Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ontario M5S 3G4, Canada

[b] Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, 4259-J2-54, Nagatsuta-cho, Midori-ku, Yokohama 226-8502, Japan

ARTICLE INFO

Article history:
Received 21 December 2010
Received in revised form
14 March 2012
Accepted 29 April 2012
Available online 10 September 2012

Keywords:
Multi-agent consensus
Surplus-based averaging
Gossip algorithm
Directed graph
Eigenvalue perturbation

ABSTRACT

We study the average consensus problem of multi-agent systems for general network topologies with unidirectional information flow. We propose two linear distributed algorithms, deterministic and gossip, respectively for the cases where the inter-agent communication is synchronous and asynchronous. In both cases, the developed algorithms guarantee state averaging on arbitrary strongly connected digraphs; in particular, this graphical condition does not require that the network be balanced or symmetric, thereby extending previous results in the literature. The key novelty of our approach is to augment an additional variable for each agent, called "surplus", whose function is to locally record individual state updates. For convergence analysis, we employ graph-theoretic and nonnegative matrix tools, plus the eigenvalue perturbation theory playing a crucial role.

1. Introduction

This paper presents a new approach to the design of distributed algorithms for *average consensus*: that is, a system of networked agents reaches an agreement on the average value of their initial states, through merely local interaction among peers. The approach enables multi-agent systems to achieve average consensus on arbitrary strongly connected network topologies with unidirectional information flow, where the state sum of the agents need not stay put as time evolves.

There has been an extensive literature addressing multi-agent consensus problems. Many fundamental distributed algorithms (developed in, e.g., Bertsekas & Tsitsiklis, 1989; Jadbabaie, Lin, & Morse, 2003; Olfati-Saber & Murray, 2004; Ren & Beard, 2008; Xiao & Boyd, 2004) are of the *synchronous* type: At an arbitrary time, individual agents are assumed to sense and/or communicate with all the neighbors, and then simultaneously execute their local updating protocols. In particular, Olfati-Saber and Murray (2004) studied algorithms of such type to achieve average consensus on static (i.e., time-invariant) digraphs, and justified that a *balanced* and *strongly connected* topology is necessary and sufficient to guarantee convergence. More recently Boyd, Ghosh, Prabhakar, and Shah (2006) proposed a compelling "gossip" algorithm, which provides an asynchronous approach to treat average consensus. Specifically, the algorithm assumes that at each time instant, exactly one agent wakes up, contacts only one of its neighbors selected at random, and then both agents average out their states. The graph model that the algorithm is based on is undirected (or symmetric), and average consensus is ensured as long as the topology is *connected*. Since then, the gossip approach has been widely adopted (Carli, Fagnani, Frasca, & Zampieri, 2010; Kashyap, Başar, & Srikant, 2007; Lavaei & Murray, 2012) in tackling average consensus on undirected graphs, with additional constraints on quantized information flow; see also Ishii and Tempo (2010) for related distributed computation problems in search engines.

In this paper, and its conference precursor (Cai & Ishii, 2011b), we study the average consensus problem under both synchronous and asynchronous setups, as in Boyd et al. (2006) and Olfati-Saber and Murray (2004). In each case, we propose a novel type of linear distributed algorithms, which can be seen as extensions of the corresponding algorithms in Boyd et al. (2006) and Olfati-Saber and Murray (2004); and we prove that these new algorithms guarantee state averaging on arbitrary strongly connected digraphs, therefore generalizing the graphical conditions derived in Boyd et al. (2006) and Olfati-Saber and Murray (2004). We note that digraph models have been studied extensively in the consensus literature (Olfati-Saber & Murray,

☆ This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology in Japan under Grants-in-Aid for Scientific Research, No. 21760323 and 23760385. The material in this paper was partially presented at the 50th IEEE Conference on Decision and Control (CDC 2011) and European Control Conference, December 12–15, 2011, Orlando, Florida, USA. This paper was recommended for publication in revised form by Associate Editor Valery Ugrinovskii under the direction of Editor Ian R. Petersen.

E-mail addresses: kai.cai@scg.utoronto.ca (K. Cai), ishii@dis.titech.ac.jp (H. Ishii).

[1] Tel.: +1 416 978 6299; fax: +1 416 978 0804.

0005-1098/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2012.08.003

2004; Ren & Beard, 2008; Xiao & Boyd, 2004), and considered to be generally more economical for information exchange than their undirected counterpart (refer to Olfati-Saber & Murray, 2004 for a more detailed motivation of using digraphs). Our underlying theoretic interest in this paper is to generalize the connectivity conditions on digraphs for average consensus.

The primary challenge of average consensus on arbitrary strongly connected digraphs lies in that the state sum of agents need not be preserved, thereby causing shifts in the average. We note that there are a few efforts in the literature having addressed this issue. In Franceschelli, Giua, and Seatzu (2009), an auxiliary variable is associated to each agent and a linear "broadcast gossip" algorithm is proposed; however, the convergence of that algorithm is not proved, and remarked to be difficult. Benezit, Blondel, Thiran, Tsitsiklis, and Vetterli (2010) and Kempe, Dobra, and Gehrke (2003) also use extra variables, and a nonlinear (division involved) algorithm is designed and proved to achieve state averaging on non-balanced digraphs. The idea is based on computing the stationary distribution for the Markov chain characterized by the agent network, and is thus different from consensus-type algorithms (Boyd et al., 2006; Olfati-Saber & Murray, 2004). In Bertsekas and Tsitsiklis (1989, Section 7.4), the load balancing problem is tackled in which interprocessor communication is asynchronous and with bounded delay. The underlying topology is assumed undirected; owing to asynchronism and delay, however, the total amount of loads at processors is not invariant. A switched linear algorithm is proposed to achieve load balancing in this scenario, the rules of which rely on, however, bidirectional communication. In addition, a different and interesting approach is presented in Gharesifard and Cortés (2011): Given a general strongly connected digraph, find a corresponding doubly stochastic matrix (which, when used as a distributed updating scheme, guarantees state averaging (Xiao & Boyd, 2004)). An algorithm is designed to achieve this goal by adding selfloop edges with proper weights to balance flow-in and -out information. Finally, time-varying state sum caused by packet loss or communication failure is considered in Fagnani and Zampieri (2009) and Patterson, Bamieh, and El Abbadi (2007), and the deviation from the initial average is analyzed.

We develop a new approach to handle the problem that the state sum of agents need not be preserved. Similar to Franceschelli et al. (2009), we also augment an additional variable for each agent, which we call "surplus"; different from Benezit et al. (2010) and Kempe et al. (2003), the function of surplus variables is to record the state change of the associated agent at each time. Thus, in effect, these variables collectively maintain the information of the average shift amount.[2] Using this novel idea, our main contribution is the design of linear algorithms (without switching) to achieve average consensus on general strongly connected digraphs, in contrast with the types of algorithms designed in Benezit et al. (2010), Bertsekas and Tsitsiklis (1989, Section 7.4) and Kempe et al. (2003). Also, linearity allows us to employ certain matrix tools in analysis, which are different from the proof methods used in Benezit et al. (2010), Bertsekas and Tsitsiklis (1989, Section 7.4) and Kempe et al. (2003). Moreover, our technical contribution in this paper is the demonstration of matrix perturbation tools (including eigenvalue perturbation, optimal matching distance, and the Bauer–Fike Theorem (Bhatia, 1996; Seyranian & Mailybaev, 2004; Stewart & Sun, 1990)) in analyzing convergence properties, which seems unexplored in the consensus literature.

Our idea of adding surplus variables is indeed a continuation of our own previous work in Cai and Ishii (2011a), where the original surplus-based approach is proposed to tackle *quantized* average consensus on general digraphs. There we developed a quantized (thus nonlinear) averaging algorithm, and the convergence analysis is based on finite Markov chains. By contrast, the algorithms designed in the present paper are linear, and hence the convergence can be characterized by the spectral properties of the associated matrices. On the other hand, our averaging algorithms differ also from those standard ones (Boyd et al., 2006; Olfati-Saber & Murray, 2004) in that the associated matrices of our algorithms contain negative entries. Consequently for our analysis tools, besides the usual nonnegative matrix theory and algebraic graph theory, the matrix perturbation theory is found instrumental.

The paper is organized as follows. Section 2 formulates distributed average consensus problems in both synchronous and asynchronous setups. Sections 3 and 4 present the respective solution algorithms, which are rigorously proved to guarantee state averaging on general strongly connected digraphs. Further, Section 5 explores certain special topologies that lead us to specialized results, and Section 6 provides a set of numerical examples for demonstration. Finally, Section 7 states our conclusions.

Notation. Let $\mathbf{1} := [1 \cdots 1]^T \in \mathbb{R}^n$ be the vector of all ones. For a complex number λ, denote its real part by $\mathrm{Re}(\lambda)$, imaginary part by $\mathrm{Im}(\lambda)$, conjugate by $\bar{\lambda}$, and modulus by $|\lambda|$. For a set $\mathcal{S}$, denote its cardinality by $\mathrm{card}(\mathcal{S})$. Given a matrix M, $|M|$ denotes its determinant; the spectrum $\sigma(M)$ is the set of its eigenvalues; the spectral radius $\rho(M)$ is the maximum modulus of its eigenvalues. In addition, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ denote the 2-norm and infinity norm of a vector/matrix.

2. Problem formulation

Given a network of n (>1) agents, we model its interconnection structure by a digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: Each *node* in $\mathcal{V} = \{1, \ldots, n\}$ stands for an agent, and each directed *edge* (j, i) in $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denotes that agent j communicates to agent i (namely, the information flow is from j to i). Selfloop edges are not allowed, i.e., $(i, i) \notin \mathcal{E}$. In $\mathcal{G}$ a node i is *reachable* from a node j if there exists a path from j to i which respects the direction of the edges. We say $\mathcal{G}$ is *strongly connected* if every node is reachable from every other node. A *closed strong component* of $\mathcal{G}$ is a maximal set of nodes whose corresponding subdigraph is strongly connected and closed (i.e., no node inside the subdigraph is reachable from any node outside). Also a node i is called *globally reachable* if every other node is reachable from i.

At time $k \in \mathbb{Z}_+$ (nonnegative integers) each node $i \in \mathcal{V}$ has a scalar state $x_i(k) \in \mathbb{R}$; the aggregate state is denoted by $x(k) = [x_1(k) \cdots x_n(k)]^T \in \mathbb{R}^n$. The *average consensus problem* aims at designing distributed algorithms, where individual nodes update their states using only the local information of their neighboring nodes in the digraph $\mathcal{G}$ such that all $x_i(k)$ eventually converge to the initial average $x_a := \mathbf{1}^T x(0)/n$. To achieve state averaging on general digraphs, the main difficulty is that the state sum $\mathbf{1}^T x$ need not remain invariant, which can result in losing track of the initial average x_a. To deal with this problem, we propose associating to each node i an additional variable $s_i(k) \in \mathbb{R}$, called *surplus*; write $s(k) = [s_1(k) \cdots s_n(k)]^T \in \mathbb{R}^n$ and set $s(0) = 0$. The function of surplus is to locally record the state changes of individual nodes such that $\mathbf{1}^T(x(k) + s(k)) = \mathbf{1}^T x(0)$ for all time k; in other words, surplus keeps the quantity $\mathbf{1}^T(x + s)$ constant over time.

In the first part of this paper, we consider synchronous networks as in Olfati-Saber and Murray (2004): At each time, every node communicates with all of its neighbors simultaneously, and then makes a corresponding update.

[2] The method of augmenting auxiliary variables is also found in Aysal, Oreshkin, and Coates (2009) and Liu, Anderson, Cao, and Morse (2009), as *predictors* estimating future states and *shift registers* storing past states respectively, in order to accelerate consensus speed. How the predictors or registers are used in these references is, however, different from our usage of surpluses.

Definition 1. A network of agents is said to achieve *average consensus* if for every initial condition $(x(0), s(0) = 0)$, it holds that $(x(k), s(k)) \to (x_a\mathbf{1}, 0)$ as $k \to \infty$.

Problem 1. Design a distributed algorithm such that agents achieve average consensus on general digraphs.

To solve this problem, we will propose in Section 3 a surplus-based distributed algorithm, under which we justify that average consensus is achieved for general strongly connected digraphs.

In the second part, we consider the setup of asynchronous networks as in Boyd et al. (2006). Specifically, communication among nodes is by means of *gossip*: At each time, exactly one edge $(j, i) \in \mathcal{E}$ is activated at random, independently from all earlier instants and with a time-invariant, strictly positive probability $p_{ij} \in (0, 1)$ such that $\sum_{(j,i)\in\mathcal{E}} p_{ij} = 1$. Along this activated edge, node j sends its state and surplus to node i, while i receives the information and makes a corresponding update.

Definition 2. A network of agents is said to achieve

(i) *mean-square average consensus* if for every initial condition $(x(0), s(0) = 0)$, it holds that $E\left[\|x(k) - x_a\mathbf{1}\|_2^2\right] \to 0$ and $E\left[\|s(k)\|_2^2\right] \to 0$ as $k \to \infty$;
(ii) *almost sure average consensus* if for every initial condition $(x(0), s(0) = 0)$, it holds that $(x(k), s(k)) \to (x_a\mathbf{1}, 0)$ as $k \to \infty$ with probability one.

As defined, the mean-square convergence is concerned with the second moments of the state and surplus processes, whereas the almost sure convergence is with respect to the corresponding sample paths. It should be noted that in general there is no implication between these two convergence notions (e.g., Grimmett & Stirzaker, 2001, Section 7.2).

Problem 2. Design a distributed algorithm such that agents achieve mean-square and/or almost sure average consensus on general digraphs.

For this problem, we will propose in Section 4 a surplus-based gossip algorithm, under which we justify that both mean-square and almost sure average consensus can be achieved for general strongly connected digraphs.

3. Averaging in synchronous networks

This section solves Problem 1. First we present a (discrete-time) distributed algorithm based on surplus, which may be seen as an extension of the standard consensus algorithms in the literature (Bertsekas & Tsitsiklis, 1989; Jadbabaie et al., 2003; Olfati-Saber & Murray, 2004; Ren & Beard, 2008; Xiao & Boyd, 2004). Then we prove convergence to average consensus for general strongly connected digraphs.

3.1. Algorithm description

Consider a system of n agents represented by a digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. For each node $i \in \mathcal{V}$, let $\mathcal{N}_i^+ := \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}$ denote the set of its "in-neighbors", and $\mathcal{N}_i^- := \{h \in \mathcal{V} : (i, h) \in \mathcal{E}\}$ the set of its "out-neighbors". Note that $\mathcal{N}_i^+ \neq \mathcal{N}_i^-$ in general; and $i \notin \mathcal{N}_i^+$ or $\mathcal{N}_i^-$, for selfloop edges do not exist. There are three operations that every node i performs at time $k \in \mathbb{Z}_+$. First, node i sends its state information $x_i(k)$ and weighted surplus $b_{ih}s_i(k)$ to each out-neighbor $h \in \mathcal{N}_i^-$; here the *sending weight* b_{ih} is assumed to satisfy that $b_{ih} \in (0, 1)$ if $h \in \mathcal{N}_i^-$, $b_{ih} = 0$ if $h \in \mathcal{V} - \mathcal{N}_i^-$, and $\sum_{h\in\mathcal{N}_i^-} b_{ih} < 1$. Second, node i receives[3] state information $x_j(k)$ and weighted surplus $b_{ji}s_j(k)$ from each in-neighbor $j \in \mathcal{N}_i^+$. Third, node i updates its own state $x_i(k)$ and surplus $s_i(k)$ as follows:

$$x_i(k+1) = x_i(k) + \sum_{j\in\mathcal{N}_i^+} a_{ij}(x_j(k) - x_i(k)) + \epsilon s_i(k), \tag{1}$$

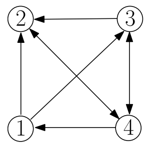
$$\begin{aligned} s_i(k+1) = {} & \left(1 - \sum_{h\in\mathcal{N}_i^-} b_{ih}\right) s_i(k) + \sum_{j\in\mathcal{N}_i^+} b_{ji}s_j(k) \\ & - \Big(x_i(k+1) - x_i(k)\Big), \end{aligned} \tag{2}$$

where the *updating weight* a_{ij} is assumed to satisfy that $a_{ij} \in (0, 1)$ if $j \in \mathcal{N}_i^+$, $a_{ij} = 0$ if $j \in \mathcal{V} - \mathcal{N}_i^+$, and $\sum_{j\in\mathcal{N}_i^+} a_{ij} < 1$; in addition, the parameter ϵ is a positive number which specifies the amount of surplus used to update the state.

We discuss the implementation of the above protocol in applications of sensor networks. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represent a network of sensor nodes. Our protocol deals particularly with scenarios where (i) sensors have different communication ranges owing possibly to distinct types or power supplies; (ii) communication is by means of broadcasting (e.g., Franceschelli et al., 2009) which again might have different ranges; and (iii) strategy of random geographic routing is used for efficient and robust node value aggregation in one direction (Benezit et al., 2010; Kempe et al., 2003). In these scenarios, information flow among sensors is typically directed. A concrete example is using sensor networks for monitoring geological areas (e.g., volcanic activities), where sensors are fixed at certain locations. At the time of setting them up, the sensors may be given different transmission power for saving energy (such sensors must run for a long time) or owing to geological reasons. Once the power is fixed, the neighbors (and their IDs) can be known to each sensor. Thus, digraphs can arise in static sensor networks where the neighbors can be fixed and known. To implement states and surpluses, we see from (1), (2) that they are ordinary variables locally stored, updated, and exchanged; thus they may be implemented by allocating memories in sensors. For the parameter ϵ, we will see that it plays a crucial role in the convergence of our algorithm; however, ϵ must be chosen sufficiently small, and a valid bound for its value involves non-local information of the digraph. The latter constraint (in bounding a parameter) is often found in consensus algorithms involving more than one variable (Li, Duan, Chen, & Huang, 2010; Li, Fu, Xie, & Zhang, 2011; Ren & Beard, 2008). One may overcome this by computing a valid bound offline, and notifying that ϵ value to every node.

Now let the *adjacency matrix* A of the digraph $\mathcal{G}$ be given by $A := [a_{ij}] \in \mathbb{R}^{n\times n}$, where the entries are the updating weights. Then the *Laplacian matrix* L is defined as $L := D - A$, where $D = \mathrm{diag}(d_1, \ldots, d_n)$ with $d_i = \sum_{j=1}^n a_{ij}$. Thus L has nonnegative diagonal entries, nonpositive off-diagonal entries, and zero row sums. Then the matrix $I - L$ is nonnegative $\left(\text{by } \sum_{j\in\mathcal{N}_i^+} a_{ij} < 1\right)$, and every row sums up to one; namely $I - L$ is *row stochastic*. Also let $B := [b_{ih}]^T \in \mathbb{R}^{n\times n}$, where the entries are the sending weights (note that the transpose in the notation is needed because $h \in \mathcal{N}_i^-$ for b_{ih}). Define the matrix $S := (I - \tilde{D}) + B$, where $\tilde{D} = \mathrm{diag}(\tilde{d}_1, \ldots, \tilde{d}_n)$ with $\tilde{d}_i = \sum_{h=1}^n b_{ih}$. Then S is nonnegative $\left(\text{by } \sum_{h\in\mathcal{N}_i^-} b_{ih} < 1\right)$, and every column sums up to one; i.e., S is *column stochastic*. As can be observed from (2), the matrix S

[3] We assume throughout the paper that the state and surplus information can be transmitted without loss.



$$\mathcal{N}_1^+ = \{4\} \qquad \mathcal{N}_1^- = \{2,3\}$$
$$\mathcal{N}_2^+ = \{1,3,4\} \qquad \mathcal{N}_2^- = \{4\}$$
$$\mathcal{N}_3^+ = \{1,4\} \qquad \mathcal{N}_3^- = \{2,4\}$$
$$\mathcal{N}_4^+ = \{2,3\} \qquad \mathcal{N}_4^- = \{1,2,3\}$$

Fig. 1. Illustrating example of four agents: communication topology and neighbor sets.

captures the part of update induced by sending and receiving surplus.

With the above matrices, the iterations (1) and (2) can be written in a matrix form as

$$\begin{bmatrix} x(k+1) \\ s(k+1) \end{bmatrix} = M \begin{bmatrix} x(k) \\ s(k) \end{bmatrix}, \quad \text{where } M := \begin{bmatrix} I-L & \epsilon I \\ L & S-\epsilon I \end{bmatrix}. \tag{3}$$

Notice that (i) the matrix M has negative entries due to the presence of the Laplacian matrix L in the (2, 1)-block; (ii) the column sums of M are equal to one, which implies that the quantity $\mathbf{1}^T(x(k) + s(k))$ is a constant for all $k \geq 0$; and (iii) the state evolution specified by the (1, 1)-block of M, i.e.,

$$x(k+1) = (I - L)x(k), \tag{4}$$

is that of the *standard* consensus algorithm studied in the literature (e.g., Bertsekas & Tsitsiklis, 1989; Olfati-Saber & Murray, 2004; Xiao & Boyd, 2004). We will henceforth refer to (3) as the deterministic algorithm, and analyze its convergence properties in the next subsection.

Example 3. For an illustration of the algorithm (3), consider a network of four nodes with neighbor sets shown in Fig. 1. Fixing $i \in [1, 4]$, let $a_{ij} = 1/\big(\mathrm{card}(\mathcal{N}_i^+) + 1\big)$ for every $j \in \mathcal{N}_i^+$ and $b_{ih} = 1/\big(\mathrm{card}(\mathcal{N}_i^-) + 1\big)$ for every $h \in \mathcal{N}_i^-$. Then the matrix M of this example is given by

$$M = \left[\begin{array}{cccc|cccc} \frac{1}{2} & 0 & 0 & \frac{1}{2} & \epsilon & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \epsilon & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \epsilon & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & \epsilon \\ \hline \frac{1}{2} & 0 & 0 & -\frac{1}{2} & \frac{1}{3}-\epsilon & 0 & 0 & \frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{3} & \frac{1}{2}-\epsilon & \frac{1}{3} & \frac{1}{4} \\ -\frac{1}{3} & 0 & \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3}-\epsilon & \frac{1}{4} \\ 0 & -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4}-\epsilon \end{array}\right].$$

We see that M has negative entries, and every column sums up to one.

3.2. Convergence result

The following is a graphical characterization for the deterministic algorithm (3) to achieve average consensus. The proof is deferred to Section 3.3.

Theorem 4. *Using the deterministic algorithm (3) with the parameter $\epsilon > 0$ sufficiently small, the agents achieve average consensus if and only if the digraph $\mathcal{G}$ is strongly connected.*

Without augmenting surplus variables, it is well known (Olfati-Saber & Murray, 2004) that a necessary and sufficient graphical condition for state averaging is that the digraph $\mathcal{G}$ is both strongly connected and *balanced*.[4] A balanced structure can be restrictive because when all the weights a_{ij} are identical, it requires the number of incoming and outgoing edges at each node in the digraph to be the same. By contrast, our algorithm (3) ensures average consensus for arbitrary strongly connected digraphs (including those non-balanced).

A surplus-based averaging algorithm was initially proposed in Cai and Ishii (2011a) for a quantized consensus problem. It guarantees that the integer-valued states converge to either $\lfloor x_a \rfloor$ (the largest integer smaller than or equal to x_a) or $\lceil x_a \rceil$ (the smallest integer larger than or equal to x_a). There, the steady-state surpluses are nonzero in general; in addition, the set of states and surpluses is finite, and thus arguments of finite Markov chain type are employed in the proof. Distinct from Cai and Ishii (2011a), with the algorithm (3) the states converge to the exact average x_a and the steady-state surpluses are zero. Moreover, since the algorithm (3) is linear, its convergence can be analyzed using tools from matrix theory, as detailed below. This last linearity point is also in contrast with the division involved algorithm designed in Benezit et al. (2010) and Kempe et al. (2003).

The choice of the parameter ϵ depends on the graph structure and the number of agents. In the following, we present an upper bound on ϵ for general networks.

Proposition 5. *Suppose that the digraph $\mathcal{G}$ is strongly connected. The deterministic algorithm (3) achieves average consensus if the parameter ϵ satisfies $\epsilon \in (0, \bar{\epsilon}^{(d)})$, where*

$$\bar{\epsilon}^{(d)} := \frac{1}{(20 + 8n)^n} (1 - |\lambda_3|)^n, \tag{5}$$

where λ_3 is the third largest eigenvalue of M in (3) by setting $\epsilon = 0$.

The proof of Proposition 5 is presented in Section 3.4, which employs a fact from matrix perturbation theory relating ϵ to the distance between perturbed and unperturbed eigenvalues (e.g., Bhatia, 1996; Stewart & Sun, 1990). Also, we will stress that this proof is based on that of Theorem 4. The above bound $\bar{\epsilon}^{(d)}$ ensures average consensus for arbitrary strongly connected topologies. Due to the power n, however, the bound is rather conservative. This power is indeed unavoidable for any perturbation bound result with respect to general matrices, as is well known in matrix perturbation literature (Bhatia, 1996; Stewart & Sun, 1990). In Section 5, we will exploit structures of some special topologies, which yield less conservative bounds on ϵ. Also, we see that the bound in (5) involves λ_3, the second largest eigenvalue of either $I - L$ or S (matrix M is block-diagonal when $\epsilon = 0$). This infers that, in order to bound ϵ, we need to know the structure of the agent network. Such a requirement when bounding some parameters in consensus algorithms, unfortunately, does not seem to be unusual (Li et al., 2010, 2011; Ren & Beard, 2008).

3.3. Proof of Theorem 4

We present the proof of Theorem 4. First, we state a necessary and sufficient condition for average consensus in terms of the spectrum of the matrix M.

Proposition 6. *The deterministic algorithm (3) achieves average consensus if and only if 1 is a simple eigenvalue of M, and all other eigenvalues have moduli smaller than one.*

Proof. The sufficiency part is standard, and we refer it to Cai and Ishii (2011b). For necessity, first we claim that the eigenvalue 1 of M is always simple. Suppose on the contrary that the algebraic multiplicity of 1 equals two. The corresponding geometric multiplicity, however, equals one; this is checked by verifying $\mathrm{rank}(M - I) = 2n - 1$. Thus there exists a generalized

[4] A digraph $\mathcal{G}$ with its adjacency matrix $A = [a_{ij}]$ is balanced if $\sum_{j=1}^n a_{ij} = \sum_{j=1}^n a_{ji}$ for all i. Equivalently, the system matrix $I - L$ of the standard consensus algorithm (4) is both row and column stochastic (Olfati-Saber & Murray, 2004; Xiao & Boyd, 2004).

right eigenvector $u = [u_1^T\ u_2^T]^T \in \mathbb{R}^{2n}$ such that $(M - I)^2 u = 0$, and $(M - I)u$ is a right eigenvector with respect to the eigenvalue 1. Since $[\mathbf{1}^T\ 0]^T$ is also a right eigenvector corresponding to the eigenvalue 1, it must hold:

$$(M - I)u = c[\mathbf{1}^T\ 0]^T, \quad \text{for some scalar } c \neq 0$$
$$\Rightarrow \begin{bmatrix} -L & \epsilon I \\ L & S - I - \epsilon I \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = c \begin{bmatrix} \mathbf{1} \\ 0 \end{bmatrix}$$
$$\Rightarrow \begin{cases} -Lu_1 + \epsilon u_2 = c\mathbf{1} \\ Lu_1 + (S - I)u_2 - \epsilon u_2 = 0 \end{cases}$$
$$\Rightarrow (S - I)u_2 = c\mathbf{1}.$$

One may verify that $\operatorname{rank}(S - I) = n - 1$ but $\operatorname{rank}([S - I\ \ c\mathbf{1}]) = n$. Hence there is no solution for u_2, which in turn implies that the generalized right eigenvector u cannot exist. This proves our claim.

Now suppose that there is an eigenvalue λ of M such that $\lambda \neq 1$ and $|\lambda| \geq 1$. But this immediately implies that $\lim_{k\to\infty} M^k$ does not exist (Xiao & Boyd, 2004). Therefore, average consensus cannot be achieved. □

Next, we introduce an important result from matrix perturbation theory (e.g., Seyranian & Mailybaev, 2004, Chapter 2), which is found crucial in analyzing the spectral properties of the matrix M in (3). The proof of this result can be found in Seyranian and Mailybaev (2004, Sections 2.8 and 2.10). An eigenvalue of a matrix is said to be *semi-simple* if its algebraic multiplicity is equal to its geometric multiplicity.

Lemma 7. *Consider an $n \times n$ matrix $W(\epsilon)$ which depends smoothly on a real parameter $\epsilon \geq 0$. Fix $l \in [1, n]$; let $\lambda_1 = \cdots = \lambda_l$ be a semi-simple eigenvalue of $W(0)$, with (linearly independent) right eigenvectors $y_1, \ldots, y_l$ and (linearly independent) left eigenvectors $z_1, \ldots, z_l$ such that*

$$\begin{bmatrix} z_1^T \\ \vdots \\ z_l^T \end{bmatrix} \begin{bmatrix} y_1 & \cdots & y_l \end{bmatrix} = I.$$

Consider a small $\epsilon > 0$, and denote by $\lambda_i(\epsilon)$ the eigenvalues of $W(\epsilon)$ corresponding to λ_i, $i \in [1, l]$. Then the derivatives $d\lambda_i(\epsilon)/d\epsilon|_{\epsilon=0}$ exist, and they are the eigenvalues of the following $l \times l$ matrix:

$$\begin{bmatrix} z_1^T \dot{W} y_1 & \cdots & z_1^T \dot{W} y_l \\ \vdots & & \vdots \\ z_l^T \dot{W} y_1 & \cdots & z_l^T \dot{W} y_l \end{bmatrix}, \quad \text{where } \dot{W} := dW(\epsilon)/d\epsilon|_{\epsilon=0}. \tag{6}$$

Now we are ready to prove Theorem 4. The necessity argument follows from the one for Cai and Ishii (2011a, Theorem 2); indeed, the class of strongly connected digraphs characterizes the existence of a distributed algorithm that can solve average consensus. For the sufficiency part, let

$$M_0 := \begin{bmatrix} I - L & 0 \\ L & S \end{bmatrix} \quad \text{and} \quad F := \begin{bmatrix} 0 & I \\ 0 & -I \end{bmatrix}. \tag{7}$$

Then $M = M_0 + \epsilon F$, and we view M as being obtained by "perturbing" M_0 via the term ϵF. Also, it is clear that M depends smoothly on ϵ. Concretely, we will first show that the eigenvalues λ_i of the unperturbed matrix M_0 satisfy

$$1 = \lambda_1 = \lambda_2 > |\lambda_3| \geq \cdots \geq |\lambda_{2n}|. \tag{8}$$

Then using Lemma 7 we will establish that after a small perturbation ϵF, the obtained matrix M has only a simple eigenvalue 1 and all other eigenvalues have moduli smaller than one. This is the characteristic part of our proof. Finally, it follows from Proposition 6 that average consensus is achieved. It should be pointed out that, unlike the standard consensus algorithm (4), the tools in nonnegative matrix theory cannot be used to analyze the spectrum of M directly due to the existence of negative entries.

Proof of Theorem 4. (*Necessity*) Suppose that $\mathcal{G}$ is not strongly connected. Then at least one node of $\mathcal{G}$ is not globally reachable. Let $\mathcal{V}_g^*$ denote the set of non-globally reachable nodes, and write its cardinality $\operatorname{card}(\mathcal{V}_g^*) = r$, $r \in [1, n]$. If $r = n$, i.e. $\mathcal{G}$ does not have a globally reachable node, then $\mathcal{G}$ has at least two distinct closed strong components (Lin, 2008, Theorem 2.1). In this case, if the nodes in different components have different initial states, then average consensus cannot be achieved. It is left to consider $r < n$. Let $\mathcal{V}_g := \mathcal{V} - \mathcal{V}_g^*$ denote the set of all globally reachable nodes; thus $\mathcal{V}_g$ is the unique closed strong component in $\mathcal{G}$ (Lin, 2008, Theorem 2.1). Consider an initial condition $(x(0), 0)$ such that all nodes in $\mathcal{V}_g$ have the same state $c \in \mathbb{R}$, and not all the states of the nodes in $\mathcal{V}_g^*$ equal c. Hence $x_a \neq c$. But no state or surplus update is possible for the nodes in $\mathcal{V}_g$ because it is closed, and therefore average consensus cannot be achieved.

(Sufficiency) First, we prove the assertion (8). Since M_0 is block (lower) triangular, its spectrum is $\sigma(M_0) = \sigma(I - L) \cup \sigma(S)$. Recall that the matrices $I - L$ and S are row and column stochastic, respectively; so their spectral radii satisfy $\rho(I - L) = \rho(S) = 1$. Now owing to that $\mathcal{G}$ is strongly connected, $I - L$ and S are both *irreducible*; thus by the Perron–Frobenius Theorem (see, e.g., Horn & Johnson, 1990, Chapter 8) $\rho(I - L)$ (resp. $\rho(S)$) is a simple eigenvalue of $I - L$ (resp. S). This implies (8). Moreover, for $\lambda_1 = \lambda_2 = 1$, one derives that the corresponding geometric multiplicity equals two by verifying $\operatorname{rank}(M_0 - I) = 2n - 2$. Hence the eigenvalue 1 is semi-simple.

Next, we will qualify the changes of the semi-simple eigenvalue $\lambda_1 = \lambda_2 = 1$ of M_0 under a small perturbation ϵF. We do this by computing the derivatives $d\lambda_1(\epsilon)/d\epsilon$ and $d\lambda_2(\epsilon)/d\epsilon$ using Lemma 7; here $\lambda_1(\epsilon)$ and $\lambda_2(\epsilon)$ are the eigenvalues of M corresponding respectively to λ_1 and λ_2. To that end, choose the right eigenvectors y_1, y_2 and left eigenvectors z_1, z_2 of the semi-simple eigenvalue 1 as follows:

$$Y := \begin{bmatrix} y_1 & y_2 \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{1} \\ v_2 & -nv_2 \end{bmatrix}, \qquad Z := \begin{bmatrix} z_1^T \\ z_2^T \end{bmatrix} = \begin{bmatrix} \mathbf{1}^T & \mathbf{1}^T \\ v_1^T & 0 \end{bmatrix}.$$

Here $v_1 \in \mathbb{R}^n$ is a left eigenvector of $I - L$ with respect to $\rho(I - L)$ such that it is positive and scaled to satisfy $v_1^T\mathbf{1} = 1$; and $v_2 \in \mathbb{R}^n$ is a right eigenvector of S corresponding to $\rho(S)$ such that it is positive and scaled to satisfy $\mathbf{1}^T v_2 = 1$. The fact that positive eigenvectors v_1 and v_2 exist follows again from the Perron–Frobenius Theorem. With this choice, one readily checks $ZY = I$. Now since $dM/d\epsilon|_{\epsilon=0} = F$, the matrix (6) in the present case is

$$\begin{bmatrix} z_1^T F y_1 & z_1^T F y_2 \\ z_2^T F y_1 & z_2^T F y_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ v_1^T v_2 & -nv_1^T v_2 \end{bmatrix}.$$

It follows from Lemma 7 that for small $\epsilon > 0$, the derivatives $d\lambda_1(\epsilon)/d\epsilon$, $d\lambda_2(\epsilon)/d\epsilon$ exist and are the eigenvalues of the above matrix. Hence $d\lambda_1(\epsilon)/d\epsilon = 0$, and $d\lambda_2(\epsilon)/d\epsilon = -nv_1^T v_2 < 0$. This implies that when ϵ is small, $\lambda_1(\epsilon)$ stays put while $\lambda_2(\epsilon)$ moves to the left along the real axis. Then by continuity, there must exist a positive δ_1 such that $\lambda_1(\delta_1) = 1$ and $\lambda_2(\delta_1) < 1$. On the other hand, since eigenvalues are continuous functions of matrix entries (e.g., Bhatia, 1996; Stewart & Sun, 1990), there must exist a positive δ_2 such that $|\lambda_i(\delta_2)| < 1$ for all $i \in [3, 2n]$. Thus for any sufficiently small $\epsilon \in (0, \min\{\delta_1, \delta_2\})$, the matrix M has a simple eigenvalue 1 and all other eigenvalues have moduli smaller than one. Therefore, from Proposition 6, the conclusion that average consensus is achieved follows. □

Remark 8. Assuming that the deterministic algorithm (3) converges to the average, the speed of its convergence is governed by the second largest (in modulus) eigenvalue of the matrix M. We denote this particular eigenvalue by $\lambda_2^{(d)}$, and refer to it as the *convergence factor* of algorithm (3). Note that $\lambda_2^{(d)}$ depends not only on the graph topology but also on the parameter ϵ, and $\lambda_2^{(d)} < 1$ is equivalent to average consensus (by Proposition 6).

Remark 9. Because of adding surpluses, the matrix M in (3) is double in size and is not nonnegative. Hence standard nonnegative matrix tools cannot be directly applied; this point was also discussed in Franceschelli et al. (2009). In Liu et al. (2009) a system matrix containing negative entries was analyzed, which depends however on symmetry of network structures. By contrast, we deal with general network topologies and have demonstrated that certain matrix perturbation tools are useful in proving convergence.

3.4. Proof of Proposition 5

Some preliminaries will be presented first, based on which Proposition 5 follows. Henceforth in this subsection, the digraph $\mathcal{G}$ is assumed to be strongly connected. We begin by introducing a metric for the distance between the spectrums of M_0 and M; here $M = M_0 + \epsilon F$, with M_0 and F in (7). Let $\sigma(M_0) := \{\lambda_1, \dots, \lambda_{2n}\}$ (where the numbering is the same as that in (8)) and $\sigma(M) := \{\lambda_1(\epsilon), \dots, \lambda_{2n}(\epsilon)\}$. The *optimal matching distance* $d(\sigma(M_0), \sigma(M))$ (Bhatia, 1996; Stewart & Sun, 1990) is defined by

$$d(\sigma(M_0), \sigma(M)) := \min_{\pi} \max_{i \in [1,2n]} |\lambda_i - \lambda_{\pi(i)}(\epsilon)|, \tag{9}$$

where π is taken over all permutations of $\{1, \dots, 2n\}$. Thus if we draw $2n$ identical circles centered respectively at $\lambda_1, \dots, \lambda_{2n}$, then $d(\sigma(M_0), \sigma(M))$ is the smallest radius such that these circles include all $\lambda_1(\epsilon), \dots, \lambda_{2n}(\epsilon)$. Here is an upper bound result for the optimal matching distance (Bhatia, 1996, Theorem VIII.1.5).

Lemma 10. *An upper bound on* $d(\sigma(M_0), \sigma(M))$ *is*

$$d(\sigma(M_0), \sigma(M)) \le 4\,(\|M_0\|_\infty + \|M\|_\infty)^{1-\frac{1}{n}}\ \|\epsilon F\|_\infty^{\frac{1}{n}}.$$

Next, we are concerned with the eigenvalues $\lambda_3(\epsilon), \dots, \lambda_{2n}(\epsilon)$ of M, whose corresponding unperturbed counterparts $\lambda_3, \dots, \lambda_{2n}$ of M_0 lie strictly inside the unit circle (see the proof of Theorem 4).

Lemma 11. *If the parameter* $\epsilon \in (0, \bar{\epsilon}^{(d)})$ *with* $\bar{\epsilon}^{(d)}$ *in* (5), *then* $|\lambda_3(\epsilon)|, \dots, |\lambda_{2n}(\epsilon)| < 1$.

Proof. Since $L = D - A$ and $S = (I - \tilde{D}) + B$, one can compute $\|L\|_\infty = 2\max_{i\in[1,n]} d_i < 2$ and $\|S\|_\infty < n$. Then $\|M_0\|_\infty \le \|L\|_\infty + \|S\|_\infty < 2 + n$ and $\|F\|_\infty \le 1$. By Lemma 10,

$$\begin{aligned} d(\sigma(M_0), \sigma(M)) &\le 4\,(2\|M_0\|_\infty + \epsilon\|F\|_\infty)^{1-\frac{1}{n}}\ (\epsilon\|F\|_\infty)^{\frac{1}{n}} \\ &< 4\,(4 + 2n + \epsilon)^{1-\frac{1}{n}}\ \epsilon^{\frac{1}{n}} \\ &< 4\,(4 + 2n + \epsilon)\,\epsilon^{\frac{1}{n}} < 1 - |\lambda_3|. \end{aligned}$$

The last inequality is due to $\epsilon < \bar{\epsilon}^{(d)}$. Now recall from the proof of Theorem 4 that the unperturbed eigenvalues $\lambda_3, \dots, \lambda_{2n}$ of M_0 lie strictly inside the unit circle; in particular, (8) holds. Therefore, perturbing the eigenvalues $\lambda_3, \dots, \lambda_{2n}$ by an amount less than $\bar{\epsilon}$, the resulting eigenvalues $\lambda_3(\epsilon), \dots, \lambda_{2n}(\epsilon)$ will remain inside the unit circle. □

It is left to consider the eigenvalues $\lambda_1(\epsilon)$ and $\lambda_2(\epsilon)$ of M. Since every column sum of M equals one for an arbitrary ϵ, we obtain that 1 is always an eigenvalue of M. Hence $\lambda_1(\epsilon)$ must be equal to 1 for any ϵ. On the other hand, for $\lambda_2(\epsilon)$ the following is true.

Lemma 12. *If the parameter* $\epsilon \in (0, \bar{\epsilon}^{(d)})$ *with* $\bar{\epsilon}^{(d)}$ *in* (5), *then* $|\lambda_2(\epsilon)| < 1$.

Proof. First recall from the proof of Theorem 4 that $\lambda_2 = 1$ and $d\lambda_2(\epsilon)/d\epsilon < 0$; so for sufficiently small $\epsilon > 0$, it holds that $|\lambda_2(\epsilon)| < 1$. Now suppose that there exists $\delta \in (0, \bar{\epsilon}^{(d)})$ such that $|\lambda_2(\delta)| \ge 1$. Owing to the continuity of eigenvalues, it suffices to consider $|\lambda_2(\delta)| = 1$. There are three such possibilities, for each of which we derive a contradiction.

Case 1: $\lambda_2(\delta)$ is a complex number with nonzero imaginary part and $|\lambda_2(\delta)| = 1$. Since M is a real matrix, there must exist another eigenvalue $\lambda_i(\delta)$, for some $i \in [3, 2n]$, such that $\lambda_i(\delta)$ is a complex conjugate of $\lambda_2(\delta)$. Then $|\lambda_i(\delta)| = |\lambda_2(\delta)| = 1$, which contradicts that all the eigenvalues $\lambda_3(\delta), \dots, \lambda_{2n}(\delta)$ stay inside the unit circle as $\delta \in (0, \bar{\epsilon}^{(d)})$ by Lemma 11.

Case 2: $\lambda_2(\delta) = -1$. This implies at least $d(\sigma(M_0), \sigma(M)) = 2$, which contradicts $d(\sigma(M_0), \sigma(M)) < 1 - |\lambda_3| < 1$ when $\delta < \bar{\epsilon}^{(d)}$.

Case 3: $\lambda_2(\delta) = 1$. This case is impossible because the eigenvalue 1 of M is always simple, as we have justified in the necessity proof of Proposition 6. □

Summarizing Lemmas 11 and 12, we obtain that if the parameter $\epsilon \in (0, \bar{\epsilon}^{(d)})$ with $\bar{\epsilon}^{(d)}$ in (5), then $\lambda_1(\epsilon) = 1$ and $|\lambda_2(\epsilon)|, |\lambda_3(\epsilon)|, \dots, |\lambda_{2n}(\epsilon)| < 1$. Therefore, by Proposition 6 the deterministic algorithm (3) achieves average consensus; this establishes Proposition 5.

4. Averaging in asynchronous networks

We move on to solve Problem 2. First, a surplus-based gossip algorithm is designed for digraphs, which extends those algorithms (Boyd et al., 2006; Carli et al., 2010; Kashyap et al., 2007; Lavaei & Murray, 2012) only for undirected graphs. Then, mean-square and almost sure convergence to average consensus is justified for arbitrary strongly connected topologies.

4.1. Algorithm description

Consider again a network of n agents modeled by a digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Suppose that at each time, exactly one edge in $\mathcal{E}$ is activated at random, independently from all earlier instants. Say edge (j, i) is activated at time $k \in \mathbb{Z}_+$, with a constant probability $p_{ij} \in (0, 1)$. Along the edge, the state information $x_j(k)$ and surplus $s_j(k)$ are transmitted from node j to i. The induced updates are described as follows:

(i) Let $w_{ij} \in (0, 1)$ be the updating weight, and $\epsilon > 0$ be a parameter. For node i:

$$x_i(k+1) = x_i(k) + w_{ij}(x_j(k) - x_i(k)) + \epsilon w_{ij} s_i(k), \tag{10}$$

$$s_i(k+1) = s_i(k) + s_j(k) - (x_i(k+1) - x_i(k)). \tag{11}$$

(ii) For node j: $x_j(k+1) = x_j(k)$ and $s_j(k+1) = 0$.

(iii) For other nodes $l \in \mathcal{V} - \{i, j\}$: $x_l(k+1) = x_l(k)$ and $s_l(k+1) = s_l(k)$.

We discuss potential applications of this protocol in sensor networks. Our focus is again on the situations of directed information flow, like asynchronous communication with variable ranges or unidirectional geographic routing (Benezit et al., 2010; Kempe et al., 2003). First, the states and surpluses can be implemented as ordinary variables in sensors, since their exchange and updating rules are fairly simple and purely local. Also, we will see that the parameter ϵ, as in the algorithm (3), affects the convergence of the algorithm, and must be chosen to be sufficiently small. A valid upper bound for ϵ again involves non-local information of the network; thus computing a bound offline and then notifying that value to every node is one possible way to deal with this restriction.

Now let A_{ji} be the adjacency matrix of the digraph $\mathcal{G}_{ji} = (\mathcal{V}, \{(j,i)\})$ given by $A_{ji} = w_{ij} f_i f_j^T$, where f_i, f_j are unit vectors of the standard basis of $\mathbb{R}^n$. Then the Laplacian matrix L_{ji} is given by $L_{ji} := D_{ji} - A_{ji}$, where $D_{ji} = w_{ij} f_i f_i^T$. Thus L_{ji} has zero row sums, and the matrix $I - L_{ji}$ is row stochastic. Also define $S_{ji} := I - (f_j - f_i) f_j^T$; it is clear that S_{ji} is column stochastic. With these matrices, the iteration of states and surpluses when edge (j, i) is activated at time k can be written in a matrix form as

$$\begin{bmatrix} x(k+1) \\ s(k+1) \end{bmatrix} = M(k) \begin{bmatrix} x(k) \\ s(k) \end{bmatrix},$$

$$\text{where } M(k) = M_{ji} := \begin{bmatrix} I - L_{ji} & \epsilon D_{ji} \\ L_{ji} & S_{ji} - \epsilon D_{ji} \end{bmatrix}. \tag{12}$$

We have several remarks regarding this algorithm. (i) The matrix $M(k)$ has negative entries due to the Laplacian matrix L_{ji} in the (2, 1)-block. (ii) The column sums of $M(k)$ are equal to one, which implies that the quantity $\mathbf{1}^T(x(k) + s(k))$ is constant for all k. (iii) By the assumption on the probability distribution of activating edges, the sequence $M(k), k = 0, 1, \ldots,$ is independent and identically distributed (i.i.d.). Henceforth we refer to (12) as the gossip algorithm, and establish its mean-square and almost sure convergence in the sequel.

Example 13. Consider again the network of four nodes in Fig. 1. We give one instance of the matrix $M(k)$ when the edge (3, 2) is activated, with the weight $w_{23} = 1/2$.

$$M_{32} = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & \epsilon/2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & -1/2 & 0 & 0 & 1-\epsilon/2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

We see that $M(k)$ has negative entries, and every column sums up to one.

4.2. Convergence result

We present our main result in this section.

Theorem 14. *Using the gossip algorithm (12) with the parameter $\epsilon > 0$ sufficiently small, the agents achieve mean-square average consensus if and only if the digraph $\mathcal{G}$ is strongly connected.*

We remark that Theorem 14 generalizes the convergence result in Boyd et al. (2006) from undirected to directed graphs. The problem of achieving average consensus on gossip digraphs is, however, more difficult in that the state sum of the nodes need not be invariant at each iteration. The key in our extension is to augment surplus variables which keep track of individual state updates, thereby ensuring average consensus for general strongly connected digraphs. This approach was previously exploited in Franceschelli et al. (2009) for a "broadcast gossip" algorithm, however without a convergence proof. We remark that our technique to prove Theorem 14, based on matrix perturbation theory, can be applied to Franceschelli et al. (2009) and justify the algorithm convergence.

We note that in the literature, many works for agents with non-scalar dynamics deal only with static networks (e.g., Cao & Ren, 2010; Li et al., 2010). Some exceptions include Liu et al. (2009) which relies heavily on graph symmetry and Ren and Beard (2008) which is based on dwell-time switching. By contrast, we study general digraphs that switch at every discrete time and each resulting update matrix is not nonnegative. The corresponding analysis is difficult, and we will demonstrate again that matrix perturbation tools are instrumental in proving convergence.

To prove Theorem 14, three preliminary results are to be established in order. The first is a necessary and sufficient condition for mean-square average consensus characterized by the spectrum of the matrix $E[M(k) \otimes M(k)]$, where $\otimes$ stands for the Kronecker product. This condition will be used in the sufficiency proof of Theorem 14. Since the matrices $M(k)$ are i.i.d. we denote $E[M(k) \otimes M(k)]$ by $E[M \otimes M]$. This result corresponds to Proposition 6 for the deterministic algorithm in Section 3. The proof is standard, and can be found in e.g. Boyd et al. (2006) and Cai and Ishii (2011b).

Proposition 15. *The gossip algorithm (12) achieves mean-square average consensus if and only if 1 is a simple eigenvalue of $E[M \otimes M]$, and all the other eigenvalues have moduli smaller than one.*

The second preliminary is an easy corollary of the Perron–Frobenius Theorem.

Lemma 16 (*cf.* *Gantmacher, 1959, Chapter XIII*)**.** *Let W be a nonnegative and irreducible matrix, and λ be an eigenvalue of W. If there is a positive vector v such that $Wv = \lambda v$, then $\lambda = \rho(W)$.*

Proof. Since W is nonnegative and irreducible, the Perron–Frobenius Theorem implies that $\rho(W)$ is a simple eigenvalue of W and there is a positive left eigenvector w corresponding to $\rho(W)$, i.e., $w^T W = w^T \rho(W)$. Then

$$\begin{aligned} \rho(W)(v^T w) &= v^T(\rho(W)w) = v^T(W^T w) \\ &= (Wv)^T w = (\lambda v)^T w = \lambda(v^T w), \end{aligned}$$

which yields $(\lambda - \rho(W))(v^T w) = 0$. Since both v and w are positive, we conclude that $\lambda = \rho(W)$. □

The final preliminary is on the spectral properties of the following four matrices: $E[(I - L) \otimes (I - L)]$, $E[(I - L) \otimes S]$, $E\big[S \otimes (I - L)\big]$, and $E[S \otimes S]$. For the proof, see Cai and Ishii (2011b).

Lemma 17. *Suppose that the digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is strongly connected. Then each of the four matrices $E[(I - L) \otimes (I - L)]$, $E[(I - L) \otimes S]$, $E[S \otimes (I - L)]$, and $E[S \otimes S]$ has a simple eigenvalue 1 and all other eigenvalues with moduli smaller than one.*

We are now ready to provide the proof of Theorem 14. The necessity argument is the same as Theorem 4. Below is the sufficiency part.

Proof of Theorem 14. By Proposition 15 it suffices to show that $E[M \otimes M]$ has a simple eigenvalue 1, and all other eigenvalues with moduli smaller than one. Let

$$M_0(k) := \begin{bmatrix} I - L(k) & 0 \\ L(k) & S(k) \end{bmatrix} \quad \text{and} \quad F(k) := \begin{bmatrix} 0 & D(k) \\ 0 & -D(k) \end{bmatrix};$$

from (12) we have $M(k) = M_0(k) + \epsilon F(k)$. Then write

$$\begin{aligned} E[M \otimes M] &= E[(M_0 + \epsilon F) \otimes (M_0 + \epsilon F)] \\ &= E[M_0 \otimes M_0] + \epsilon E[M_0 \otimes F + F \otimes M_0 + F \otimes \epsilon F]. \end{aligned}$$

Let $p \in [1, 4n]$, and $p\mathbf{n} := \{(p-1)n + 1, \ldots, pn\}$. Consider the following permutation:

$$\begin{aligned} \{&\mathbf{n}, 3\mathbf{n}, \ldots, (2n-1)\mathbf{n};\ 2\mathbf{n}, 4\mathbf{n}, \ldots, 2n\mathbf{n}; \\ &(2n+1)\mathbf{n}, (2n+3)\mathbf{n}, \ldots, (4n-1)\mathbf{n}; \\ &(2n+2)\mathbf{n}, (2n+4)\mathbf{n}, \ldots, 4n\mathbf{n}\}. \end{aligned}$$

Denoting by P the corresponding permutation matrix (which is orthogonal), one derives that

$$\begin{aligned} P^T E\left[M \otimes M\right] P &= P^T E\left[M_0 \otimes M_0\right] P \\ &\quad + \epsilon P^T E\left[M_0 \otimes F + F \otimes M_0 + F \otimes \epsilon F\right] P \\ &=: \hat{M}_0 + \epsilon \hat{F}, \end{aligned} \tag{13}$$

where

$$\hat{M}_0 := E\begin{bmatrix} (I-L)\otimes(I-L) & 0 & 0 & 0 \\ (I-L)\otimes L & (I-L)\otimes S & 0 & 0 \\ L\otimes(I-L) & 0 & S\otimes(I-L) & 0 \\ L\otimes L & L\otimes S & S\otimes L & S\otimes S \end{bmatrix},$$

$$\hat{F} := E\begin{bmatrix} 0 & (I-L)\otimes D & D\otimes(I-L) & D\otimes \epsilon D \\ 0 & (L-I)\otimes D & D\otimes L & D\otimes(S-\epsilon D) \\ 0 & L\otimes D & D\otimes(L-I) & (S-\epsilon D)\otimes D \\ 0 & -L\otimes D & -D\otimes L & D\otimes(\epsilon D - S) - S\otimes D \end{bmatrix}.$$

Based on the above similarity transformation, we henceforth analyze the spectral properties of the matrix $\hat{M}_0 + \epsilon\hat{F}$. For this, we resort again to a perturbation argument, which proceeds similarly to the one for Theorem 4. First, since the digraph $\mathcal{G}$ is strongly connected, it follows from Lemma 17 that the eigenvalues of the matrix $\hat{M}_0$ satisfy

$$1 = \hat{\lambda}_1 = \hat{\lambda}_2 = \hat{\lambda}_3 = \hat{\lambda}_4 > |\hat{\lambda}_5| \geq \cdots \geq |\hat{\lambda}_{4n^2}|. \tag{14}$$

For the eigenvalue 1, one derives that its geometric multiplicity equals four by verifying $\mathrm{rank}(\hat{M}_0 - I) = 4n^2 - 4$. Thus 1 is a semi-simple eigenvalue.

Next, we will qualify the changes of the semi-simple eigenvalue $\hat{\lambda}_1 = \hat{\lambda}_2 = \hat{\lambda}_3 = \hat{\lambda}_4 = 1$ of $\hat{M}_0$ under a small perturbation $\epsilon\hat{F}$. We do this by computing the derivatives $d\hat{\lambda}_i(\epsilon)/d\epsilon$, $i \in [1, 4]$, using Lemma 7; here $\hat{\lambda}_i(\epsilon)$ are the eigenvalues of $\hat{M}_0 + \epsilon\hat{F}$ corresponding to $\hat{\lambda}_i$. To that end, choose the right and left eigenvectors of the semi-simple eigenvalue 1 as follows:

$$\begin{aligned} Y &:= \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 & \mathbf{1}\otimes\mathbf{1} \\ 0 & 0 & \mathbf{1}\otimes nv_2 & -\mathbf{1}\otimes nv_2 \\ 0 & nv_2\otimes\mathbf{1} & 0 & -nv_2\otimes\mathbf{1} \\ nv_2\otimes nv_2 & -nv_2\otimes nv_2 & -nv_2\otimes nv_2 & nv_2\otimes nv_2 \end{bmatrix}, \end{aligned}$$

$$\begin{aligned} Z &:= \begin{bmatrix} z_1^T \\ z_2^T \\ z_3^T \\ z_4^T \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n}\mathbf{1}^T\otimes\frac{1}{n}\mathbf{1}^T & \frac{1}{n}\mathbf{1}^T\otimes\frac{1}{n}\mathbf{1}^T & \frac{1}{n}\mathbf{1}^T\otimes\frac{1}{n}\mathbf{1}^T & \frac{1}{n}\mathbf{1}^T\otimes\frac{1}{n}\mathbf{1}^T \\ \frac{1}{n}\mathbf{1}^T\otimes v_1^T & 0 & \frac{1}{n}\mathbf{1}^T\otimes v_1^T & 0 \\ v_1^T\otimes\frac{1}{n}\mathbf{1}^T & v_1^T\otimes\frac{1}{n}\mathbf{1}^T & 0 & 0 \\ v_1^T\otimes v_1^T & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Here v_1 is positive such that $v_1^T E\left[I - L\right] = v_1^T$ and $v_1^T\mathbf{1} = 1$, and v_2 is positive such that $E\left[S\right] v_2 = v_2$ and $\mathbf{1}^T v_2 = 1$. With this choice, it is readily checked that $ZY = I$. Now the matrix $\hat{M}_0 + \epsilon\hat{F}$ depends smoothly on ϵ, and the derivative $d(\hat{M}_0 + \epsilon\hat{F})/d\epsilon|_{\epsilon=0}$ is

$$\begin{aligned} \hat{F}_0 &:= \left.\frac{d(\hat{M}_0 + \epsilon\hat{F})}{d\epsilon}\right|_{\epsilon=0} = \left.\left(\hat{F} + \epsilon\frac{d\hat{F}}{d\epsilon}\right)\right|_{\epsilon=0} \\ &= E\begin{bmatrix} 0 & (I-L)\otimes D & D\otimes(I-L) & 0 \\ 0 & -(I-L)\otimes D & D\otimes L & D\otimes S \\ 0 & L\otimes D & -D\otimes(I-L) & S\otimes D \\ 0 & -L\otimes D & -D\otimes L & -D\otimes S - S\otimes D \end{bmatrix}. \end{aligned}$$

Hence the matrix (6) in the present case is

$$\begin{aligned} &\begin{bmatrix} z_1^T\hat{F}_0 y_1 & z_1^T\hat{F}_0 y_2 & z_1^T\hat{F}_0 y_3 & z_1^T\hat{F}_0 y_4 \\ z_2^T\hat{F}_0 y_1 & z_2^T\hat{F}_0 y_2 & z_2^T\hat{F}_0 y_3 & z_2^T\hat{F}_0 y_4 \\ z_3^T\hat{F}_0 y_1 & z_3^T\hat{F}_0 y_2 & z_3^T\hat{F}_0 y_3 & z_3^T\hat{F}_0 y_4 \\ z_4^T\hat{F}_0 y_1 & z_4^T\hat{F}_0 y_2 & z_4^T\hat{F}_0 y_3 & z_4^T\hat{F}_0 y_4 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ nv_1^T E[D]v_2 & -nv_1^T E[D]v_2 & 0 & 0 \\ nv_1^T E[D]v_2 & 0 & -nv_1^T E[D]v_2 & 0 \\ 0 & nv_1^T E[D]v_2 & nv_1^T E[D]v_2 & -2nv_1^T E[D]v_2 \end{bmatrix}. \end{aligned}$$

It follows from Lemma 7 that for small $\epsilon > 0$, the derivatives $d\hat{\lambda}_i(\epsilon)/d\epsilon$, $i \in [1, 4]$, exist and are the eigenvalues of the above matrix. Hence $d\hat{\lambda}_1(\epsilon)/d\epsilon = 0$, $d\hat{\lambda}_2(\epsilon)/d\epsilon = d\hat{\lambda}_3(\epsilon)/d\epsilon = -nv_1^T E[D]v_2 < 0$, and $d\hat{\lambda}_4(\epsilon)/d\epsilon = -2nv_1^T E[D]v_2 < 0$. This implies that when ϵ is small, $\hat{\lambda}_1(\epsilon)$ stays put, while $\hat{\lambda}_2(\epsilon)$, $\hat{\lambda}_3(\epsilon)$, and $\hat{\lambda}_4(\epsilon)$ move to the left along the real axis. So by continuity, there exists a positive δ_1 such that $\lambda_1(\delta_1) = 1$ and $\lambda_2(\delta_1), \lambda_3(\delta_1), \lambda_4(\delta_1) < 1$. On the other hand, by the eigenvalue continuity there exists a positive δ_2 such that $|\lambda_i(\delta_2)| < 1$ for all $i \in [5, 4n^2]$. Therefore for any sufficiently small $\epsilon \in (0, \min\{\delta_1, \delta_2\})$, the matrix $\hat{M}_0 + \epsilon\hat{F}$ has a simple eigenvalue 1 and all other eigenvalues with moduli smaller than one. □

Remark 18. Assuming that the gossip algorithm (12) converges to the average in mean square, the speed of its convergence is determined by the second largest (in modulus) eigenvalue of the matrix $E\left[M \otimes M\right]$. We denote this particular eigenvalue by $\lambda_2^{(g)}$, and refer to it as the *convergence factor* of algorithm (12). Notice that $\lambda_2^{(g)}$ depends not only on the graph topology but also on the parameter ϵ, and $\lambda_2^{(g)} < 1$ is equivalent to mean-square average consensus (by Proposition 15).

Remark 19. We have established that for small enough ϵ, the gossip algorithm (12) achieves mean-square average consensus. Using the same notion of optimal matching distance and following the procedures as in Section 3.4, it may be possible to derive a general bound for ϵ by solving the inequality $4\,(\|\hat{M}_0\|_\infty + \|\hat{M}_0 + \epsilon\hat{F}\|_\infty)^{1-1/n}\,\|\epsilon\hat{F}\|_\infty^{1/n} < 1 - |\hat{\lambda}_5|$, where $\hat{M}_0, \hat{F}$ are from (13) and $\hat{\lambda}_5$ from (14). The corresponding computation is however rather long, since the involved matrices are of much larger sizes. Such a general bound unavoidably again involves n, the number of agents in the network, and $\hat{\lambda}_5$, the second largest eigenvalue of one of the four matrices in Lemma 17. Consequently, the bound for ϵ is conservative and requires knowing the structure of the network.

Finally, we consider almost sure average consensus. Note that the gossip algorithm (12) can be viewed as a *jump linear system*, with i.i.d. system matrices $M(k)$, $k \in \mathbb{Z}_+$. For such systems, it is known (e.g., Costa, Fragoso, & Marques, 2004, Corollary 3.46) that almost sure convergence can be implied from mean-square convergence. Therefore, the result on almost sure convergence is immediate.

Corollary 20. *Using the gossip algorithm (12) with the parameter $\epsilon > 0$ sufficiently small, the agents achieve almost sure average consensus if and only if the digraph $\mathcal{G}$ is strongly connected.*

5. Special topologies

We turn now to a special class of topologies — strongly connected and balanced digraphs — and investigate the required upper bound on the parameter ϵ for the deterministic algorithm (3). Furthermore, when these digraphs are restricted to symmetric or cyclic respectively, we derive less conservative ϵ bounds compared to the general one in (5).

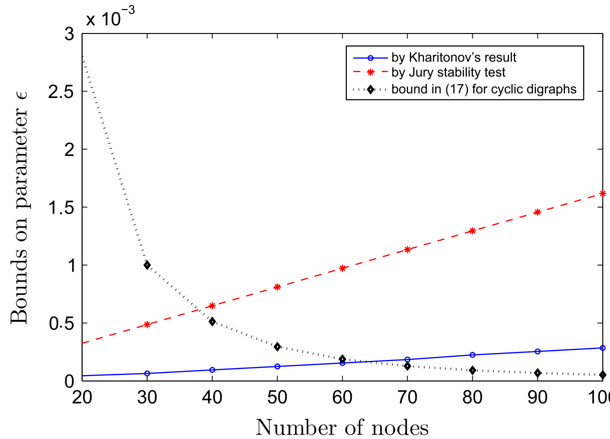


Fig. 2. Upper bounds on parameter ϵ such that deterministic algorithm (3) achieves average consensus on general strongly connected balanced digraphs (solid and dashed curves) and cyclic digraphs (dotted curve).

Given a digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, its *degree d* is defined by $d := \max_{i\in\mathcal{V}} \mathrm{card}(\mathcal{N}_i^+)$. In the deterministic algorithm (3) choose the updating and sending weights to be respectively $a_{ij} = 1/(2dn)$ and $b_{ij} = 1/(dn)$, for every $(j, i) \in \mathcal{E}$. This choice renders the two matrices $I - 2L$ and S identical, when the digraph $\mathcal{G}$ is balanced. We will see that the equality $I - 2L = S$ supports a similarity transformation in dealing with the cyclic case below.

Proposition 21. *Suppose that the parameter ϵ satisfies $\epsilon \in (0, 2)$, and the zeros of the following polynomial for every $\mu \neq 0$ with $|\mu - 1/(2n)| \le 1/(2n)$ lie strictly inside the unit circle:*

$$p(\lambda) := \lambda^2 + \alpha_1\lambda + \alpha_0, \tag{15}$$

where $\alpha_0 := 2\mu^2 - 3\mu - \epsilon + 1, \alpha_1 := 3\mu + \epsilon - 2$. Then the deterministic algorithm (3) achieves average consensus on strongly connected and balanced digraphs.

We refer to Cai and Ishii (2011b) for the proof. Now we investigate the values of ϵ that ensure the zeros of the polynomial $p(\lambda)$ in (15) inside the unit circle, which in turn guarantee average consensus on strongly connected and balanced digraphs by Proposition 21. For this, we view the polynomial $p(\lambda)$ as *interval polynomials* (Barmish, 1994) by letting μ take any value in the square: $0 \le \mathrm{Re}(\mu) \le 1/n, -1/(2n) \le \mathrm{Im}(\mu) \le 1/(2n)$. Applying the bilinear transformation we obtain a new family of interval polynomials:

$$\begin{aligned}\tilde{p}(\gamma) &:= (\gamma - 1)^2 p\left(\frac{\gamma+1}{\gamma-1}\right) \\ &= (1 + \alpha_0 + \alpha_1)\gamma^2 + (2 - 2\alpha_0)\gamma + (1 + \alpha_0 - \alpha_1).\end{aligned}$$

Then by Kharitonov's result for the complex-coefficient case, the stability of $\tilde{p}(\gamma)$ (its zeros have negative real parts) is equivalent to the stability of eight *extreme polynomials* (Barmish, 1994, Section 6.9), which in turn suffices to guarantee that the zeros of $p(\lambda)$ lie strictly inside the unit circle. Checking the stability of eight extreme polynomials results in upper bounds on ϵ in terms of n. This is displayed in Fig. 2 as the solid curve. We see that the bounds grow linearly, which is in contrast with the general bound $\bar{\epsilon}^{(d)}$ in Proposition 5 that decays exponentially and is known to be conservative. This is due to that, from the robust control viewpoint, the uncertainty of μ in the polynomial coefficients becomes smaller as n increases.

Alternatively, we employ the Jury stability test (Jury, 1988) to derive that the zeros of the polynomial $p(\lambda)$ are strictly inside the unit circle if and only if

$$\begin{aligned}\beta_0 &:= \begin{vmatrix} 1 & \alpha_0 \\ \bar{\alpha}_0 & 1 \end{vmatrix} > 0, \\ \beta_1 &:= \begin{vmatrix} \begin{vmatrix} 1 & \alpha_0 \\ \bar{\alpha}_0 & 1 \end{vmatrix} & \begin{vmatrix} 1 & \alpha_1 \\ \bar{\alpha}_0 & \bar{\alpha}_1 \end{vmatrix} \\ \begin{vmatrix} 1 & \bar{\alpha}_1 \\ \alpha_0 & \alpha_1 \end{vmatrix} & \begin{vmatrix} 1 & \bar{\alpha}_0 \\ \alpha_0 & 1 \end{vmatrix} \end{vmatrix} > 0.\end{aligned} \tag{16}$$

Here β_0 and β_1 turn out to be polynomials in ϵ of second and fourth order, respectively; the corresponding coefficients are functions of μ and n. Thus selecting μ such that $\mu \neq 0$ and $|\mu - 1/(2n)| \le 1/(2n)$, we can solve the inequalities in (16) for ϵ in terms of n. Thereby we obtain the dashed curve in Fig. 2, each plotted point being the minimum value of ϵ over 1000 random samples such that the inequalities in (16) hold. This simulation confirms that the true bound on ϵ for the zeros of $p(\lambda)$ to be inside the unit circle is between the solid and the dashed curves.

Here ends our discussion on ϵ bounds for arbitrary balanced (and strongly connected) digraphs. In the sequel, we will further specialize the balanced digraph $\mathcal{G}$ to be symmetric or cyclic, respectively, and provide analytic ϵ bounds less conservative than (5) for the general case. In particular, the exponent n is not involved.

5.1. Connected undirected graphs

A digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is *symmetric* if $(j, i) \in \mathcal{E}$ implies $(i, j) \in \mathcal{E}$. That is, $\mathcal{G}$ is undirected.

Proposition 22. *Consider a general connected undirected graph $\mathcal{G}$. Then the deterministic algorithm (3) achieves average consensus if the parameter ϵ satisfies $\epsilon \in (0, (1 - (1/n))(2 - (1/n)))$.*

Refer to Cai and Ishii (2011b) for the proof. It is noted that for connected undirected graphs, the upper bound on ϵ ensuring average consensus grows as n increases. This characteristic is in agreement with that of the bounds for the more general class of balanced digraphs as we observed in Fig. 2.

5.2. Cyclic digraphs

A digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is *cyclic* if $\mathcal{V} = \{1, \dots, n\}$ and $\mathcal{E} = \{(1, 2), (2, 3), \dots, (n-1, n), (n, 1)\}$. So a cyclic digraph is strongly connected.

Proposition 23. *Suppose that the digraph $\mathcal{G}$ is cyclic. Then the deterministic algorithm (3) achieves average consensus if the parameter ϵ satisfies*

$$\epsilon \in \left(0, \frac{\sqrt{2}}{3+\sqrt{5}}(1 - |\lambda_3|)\right), \quad \text{with } \lambda_3 \text{ as in (8)}. \tag{17}$$

Further, in this case

$$|\lambda_3| = \sqrt{1 - (1/n) + (1/(2n^2)) + (1/n)(1 - 1/(2n))\cos 2\pi/n}.$$

The proof can be found in Cai and Ishii (2011b), which relies on a perturbation result, the Bauer–Fike Theorem, for diagonalizable matrices (e.g., Horn & Johnson, 1990, Section 6.3). In Fig. 2 we plot the upper bound on ϵ in (17) for the class of cyclic digraphs. We see that this bound decays as the number n of nodes increases, which contrasts with the bound characteristic of the more general class of balanced digraphs. This may indicate the conservativeness of our current approach based on perturbation theory. Nevertheless, since the perturbation result used here is specific only to diagonalizable matrices, the derived upper bound in (17) is less conservative than the general one in (5).

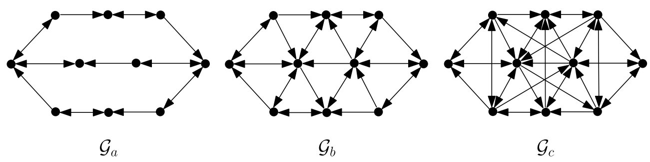


Fig. 3. Three examples of strongly connected but non-balanced digraphs.

Table 1
Convergence factors $\lambda_2^{(d)}$ and $\lambda_2^{(g)}$ with respect to three different values of parameter ϵ.

	$\epsilon = 0.2$		$\epsilon = 0.7$		$\epsilon = 2.15$	
	$\lambda_2^{(d)}$	$\lambda_2^{(g)}$	$\lambda_2^{(d)}$	$\lambda_2^{(g)}$	$\lambda_2^{(d)}$	$\lambda_2^{(g)}$
$\mathcal{G}_a$	0.9963	0.9963	0.9993	1.0003	1.0003	1.0020
$\mathcal{G}_b$	0.9951	0.9951	0.9969	0.9969	0.9985	1.0000
$\mathcal{G}_c$	0.9883	0.9883	0.9930	0.9930	0.9966	0.9993

6. Numerical examples

6.1. Convergence paths

Consider the three digraphs displayed in Fig. 3, with 10 nodes and respectively 17, 29, and 38 edges. Note that all the digraphs are strongly connected, and in the case of uniform weights they are non-balanced (indeed, no single node is balanced). We apply both the deterministic algorithm (3), with uniform weights $a = 1/(2\text{card}(\mathcal{E}))$ and $b = 1/\text{card}(\mathcal{E})$, and the gossip algorithm (12), with uniform weight $w = 1/2$ and probability $p = 1/\text{card}(\mathcal{E})$.

The convergence factors $\lambda_2^{(d)}$ and $\lambda_2^{(g)}$ (see Remarks 8 and 18) for three different values of the parameter ϵ are summarized in Table 1. We see that small ϵ ensures convergence of both algorithms (the gossip algorithm (12) requires smaller values of ϵ for mean-square convergence), whereas large ϵ can lead to instability. Moreover, in those converging cases the factors $\lambda_2^{(d)}$ and $\lambda_2^{(g)}$ decrease as the number of edges increases from $\mathcal{G}_a$ to $\mathcal{G}_c$, which indicates faster convergence when there are more communication channels available for information exchange. We also see that the algorithms are more robust on digraphs with more edges, in the sense that the range of allowed ϵ values expands from $\mathcal{G}_a$ to $\mathcal{G}_c$.

For a random initial state $x(0)$ with the average $x_a = 0$ and the initial surplus $s(0) = 0$, we display in Fig. 4 the trajectories of both states and surpluses when the deterministic algorithm (3) is applied on digraph $\mathcal{G}_a$ with parameter $\epsilon = 0.7$. Observe that asymptotically, state averaging is achieved and surplus vanishes. Under the same conditions, the gossip algorithm (12), however, fails to converge, as shown in Fig. 5. Applying algorithm (12) instead on the digraphs $\mathcal{G}_b$ and $\mathcal{G}_c$ which have more edges, average consensus is again reached, and faster convergence occurs in $\mathcal{G}_c$ (see Fig. 5).

6.2. Convergence speed versus parameter ϵ

We have seen that a sufficiently small parameter ϵ ensures convergence of both algorithms (3) and (12). Now we investigate the influence of ϵ on the speed of convergence, specifically the convergence factors $\lambda_2^{(d)}$ and $\lambda_2^{(g)}$. To reduce the effect of network topology in this investigation, we employ a type of random digraphs where an edge between every pair of nodes can exist with probability 1/2, independent across network and invariant over time; we take only those that are strongly connected.

For the deterministic algorithm (3), consider random digraphs of 50 nodes and uniform weights $a = b = 1/50$. Fig. 6 displays the curve of convergence factor $\lambda_2^{(d)}$ with respect to the parameter ϵ, each plotted point being the mean value of $\lambda_2^{(d)}$ over 100 random digraphs. To account for the trend of this curve, first

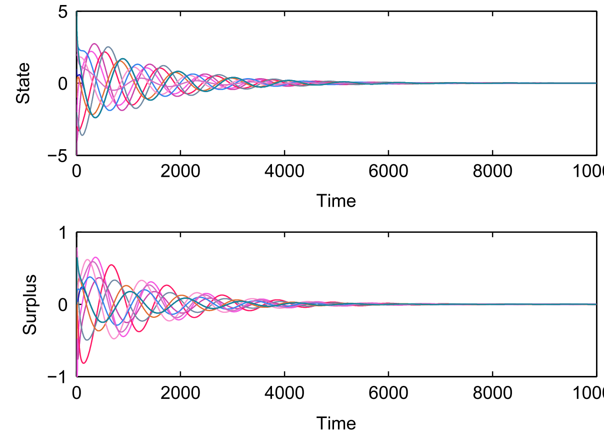


Fig. 4. Convergence paths of states and surpluses: Obtained by applying the deterministic algorithm (3) with parameter $\epsilon = 0.7$ on digraph $\mathcal{G}_a$.

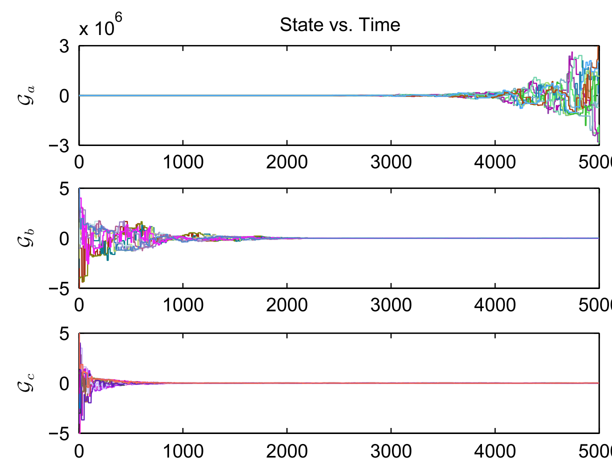


Fig. 5. Sample paths of states: Obtained by applying the gossip algorithm (3) with parameter $\epsilon = 0.7$ on digraphs $\mathcal{G}_a$, $\mathcal{G}_b$, and $\mathcal{G}_c$.

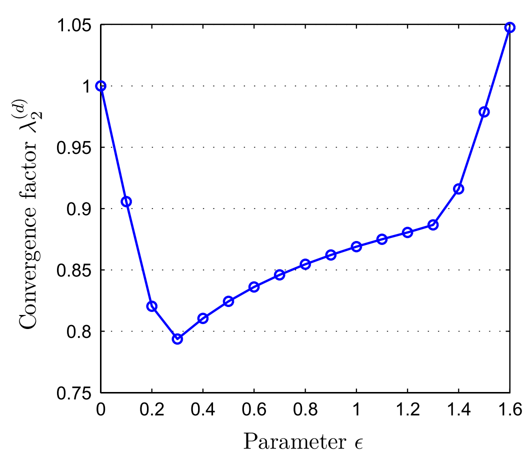


Fig. 6. Convergence factor $\lambda_2^{(d)}$ of the deterministic algorithm (3) with respect to parameter ϵ.

recall from the perturbation argument for Theorem 4 that the matrix M in (3) has two (maximum) eigenvalues 1 when $\epsilon = 0$, and small ϵ causes that one of them (denote its modulus by λ_{in}) moves into the unit circle. Meanwhile, some other eigenvalues of M inside the unit circle move outward; denote the maximum modulus among these by λ_{out}. In our simulations it is observed that when ϵ is small, $\lambda_2^{(d)} = \lambda_{in}$ ($>\lambda_{out}$) and λ_{in} moves further inside as perturbation becomes larger; so $\lambda_2^{(d)}$ decreases (faster convergence) as ϵ increases in the beginning.

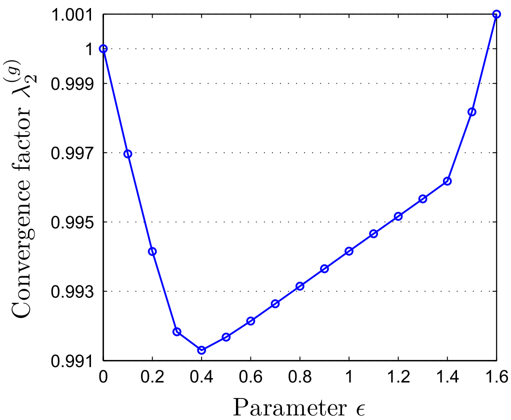


Fig. 7. Convergence factor $\lambda_2^{(g)}$ of the gossip algorithm (12) with respect to parameter ϵ.

Since the eigenvalues move continuously, there exists some ϵ such that $\lambda_{in} = \lambda_{out}$, corresponding to the fastest convergence speed. After that, $\lambda_2^{(d)} = \lambda_{out}$ ($>\lambda_{in}$) and λ_{out} moves further outside as ϵ increases; hence $\lambda_2^{(d)}$ increases and convergence becomes slower, and eventually divergence occurs (i.e., $\lambda_2^{(d)} \geq 1$).

An analogous experiment is conducted for the gossip algorithm (12), with random digraphs of 30 nodes, uniform probability $p = 1/\text{card}(\mathcal{E})$, and uniform weight $w_{ij} = 1/2$. We see in Fig. 7 a similar trend of $\lambda_2^{(g)}$ as the parameter ϵ increases, though it should be noted that the changes in $\lambda_2^{(g)}$ are smaller than those in $\lambda_2^{(d)}$. From these observations, it would be of ample interest to exploit the values of ϵ when the convergence factors achieve their minima, corresponding to the fastest convergence speed.

7. Conclusions

We have proposed surplus-based linear distributed algorithms which enable networks of agents to achieve average consensus on arbitrary strongly connected digraphs. Specifically, in synchronous networks a deterministic algorithm ensures asymptotic state averaging, and in asynchronous networks a gossip algorithm guarantees average consensus in the mean-square sense and with probability one. To emphasize, our derived graphical condition is more general than those previously reported in the literature, in the sense that it does not require a balanced network structure; also, the matrix perturbation theory plays an important role in the convergence analysis. Moreover, special regular digraphs are investigated to give less conservative bounds on the parameter ϵ; and numerical examples are provided to illustrate the convergence results, with emphasis on convergence speed.

For future research, one direction of interest would be to extend the deterministic algorithm (3) to the more realistic scenario of *switching* digraphs $\mathcal{G}(k) = (\mathcal{V}, \mathcal{E}(k))$; namely, the network topology is time-varying. If every $\mathcal{G}(k)$, $k \geq 0$, is strongly connected, then it is possible to ensure convergence by introducing *slow switching* (e.g., dwell time) as in Casbeer, Beard, and Swindlehurst (2008) and Ren and Beard (2008). Under the weaker graphical condition that digraphs $\mathcal{G}(k)$ are *jointly* strongly connected (Jadbabaie et al., 2003; Lin, 2008), to verify if average consensus can be achieved seems to be more challenging and requires further investigation.

On the other hand, in the literature on gossip algorithms (Boyd et al., 2006; Frasca & Fagnani, 2010; Kashyap et al., 2007), a variety of practical communication issues have been discussed such as link failure, message collision, broadcast protocol, and synchronized node selection (i.e., multiple nodes are selected at the same time). We thus aim at addressing these issues by making suitable extensions of our gossip algorithm (12).

Acknowledgments

The authors would like to thank Sandro Zampieri for the helpful discussion, and the anonymous reviewers for the valuable comments.

References

Aysal, T. C., Oreshkin, B. N., & Coates, M. J. (2009). Accelerated distributed average consensus via localized node state prediction. *IEEE Transactions on Signal Processing*, *57*(4), 1563–1576.

Barmish, B. R. (1994). *New tools for robustness of linear systems*. Macmillan.

Benezit, F., Blondel, V., Thiran, P., Tsitsiklis, J., & Vetterli, M. (2010). Weighted gossip: distributed averaging using non-doubly stochastic matrices. In *Proc. IEEE int. symposium on information theory* (pp. 1753–1757). Austin, TX.

Bertsekas, D. P., & Tsitsiklis, J. N. (1989). *Parallel and distributed computation: Numerical methods*. Prentice-Hall.

Bhatia, R. (1996). *Matrix analysis*. Springer-Verlag.

Boyd, S., Ghosh, A., Prabhakar, B., & Shah, D. (2006). Randomized gossip algorithms. *IEEE Transactions on Information Theory*, *52*(6), 2508–2530.

Cai, K., & Ishii, H. (2011a). Quantized consensus and averaging on gossip digraphs. *IEEE Transactions on Automatic Control*, *56*(9), 2087–2100.

Cai, K., & Ishii, H. (2011b). Average consensus on general digraphs. In *Proc. 50th IEEE conf. on decision and control and euro. control conf.* (pp. 1956–1961). Orlando, FL. Also Technical Report (available at http://arxiv.org/abs/1203.2563).

Cao, Y., & Ren, W. (2010). Multi-vehicle coordination for double-integrator dynamics under fixed undirected/directed interaction in a sampled-data setting. *International Journal of Robust and Nonlinear Control*, *20*(9), 987–1000.

Carli, R., Fagnani, F., Frasca, P., & Zampieri, S. (2010). Gossip consensus algorithms via quantized communication. *Automatica*, *46*(1), 70–80.

Casbeer, D.W., Beard, R., & Swindlehurst, A.L. (2008). Discrete double integrator consensus. In *Proc. 47th IEEE conf. on decision and control* (pp. 2264–2269). Cancun, Mexico.

Costa, O. L. V., Fragoso, M. D., & Marques, R. P. (2004). *Discrete-time Markov jump linear systems*. Springer-Verlag.

Fagnani, F., & Zampieri, S. (2009). Average consensus with packet drop communication. *SIAM Journal on Control and Optimization*, *48*(1), 102–133.

Franceschelli, M., Giua, A., & Seatzu, C. (2009). Consensus on the average on arbitrary strongly connected digraphs based on broadcast gossip algorithms. In *Proc. 1st IFAC workshop on estimation and control of networked systems* (pp. 66–71). Venice, Italy.

Frasca, P., & Fagnani, F. (2010). Broadcast gossip averaging algorithms: interference and asymptotical error in large networks. arXiv: 1005.1292.

Gantmacher, F. R. (1959). *The theory of matrices, vol. 2*. Chelsea.

Gharesifard, B., & Cortés, J. (2011). Distributed strategies for generating weight-balanced and doubly stochastic digraphs. arXiv:0911.0232.

Grimmett, G., & Stirzaker, D. (2001). *Probability and random processes*. Oxford University Press.

Horn, R. A., & Johnson, C. R. (1990). *Matrix analysis*. Cambridge University Press.

Ishii, H., & Tempo, R. (2010). Distributed randomized algorithms for the PageRank computation. *IEEE Transactions on Automatic Control*, *55*(9), 1987–2002.

Jadbabaie, A., Lin, J., & Morse, A. S. (2003). Coordination of groups of mobile autonomous agents sing nearest neighbor rules. *IEEE Transactions on Automatic Control*, *48*(6), 988–1001.

Jury, E. I. (1988). Modified stability table for 2-d digital filters. *IEEE Transactions on Circuits and Systems*, *35*(1), 116–119.

Kashyap, A., Başar, T., & Srikant, R. (2007). Quantized consensus. *Automatica*, *43*(7), 1192–1203.

Kempe, D., Dobra, A., & Gehrke, J. (2003). Gossip-based computation of aggregate information. In *Proc. 44th annual IEEE symposium on foundations of computer science* (pp. 482–491). Washington DC.

Lavaei, J., & Murray, R. M. (2012). Quantized consensus by means of gossip algorithm. *IEEE Transactions on Automatic Control*, *57*(1), 19–32.

Li, Z., Duan, Z., Chen, G., & Huang, L. (2010). Consensus of multiagent systems and synchronization of complex networks: a unified viewpoint. *IEEE Transactions on Circuits and Systems*, *57*(1), 213–224.

Li, T., Fu, M., Xie, L., & Zhang, J. (2011). Distributed consensus with limited communication data rate. *IEEE Transactions on Automatic Control*, *56*(2), 279–292.

Lin, Z. (2008). *Distributed control and analysis of coupled cell systems*. VDM Verlag.

Liu, J., Anderson, B.D.O., Cao, M., & Morse, A.S. (2009). Analysis of accelerated gossip algorithms. In *Proc. 48th IEEE conf. on decision and control* (pp. 871–876). Shanghai, China.

Olfati-Saber, R., & Murray, R. M. (2004). Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, *49*(9), 1520–1533.

Patterson, S., Bamieh, B., & El Abbadi, A. (2007). Distributed average consensus with stochastic communication failures. In *Proc. 46th IEEE conf. on decision and control* (pp. 4215–4220). New Orleans, LA.

Ren, W., & Beard, R. W. (2008). *Distributed consensus in multi-vehicle cooperative control: Theory and applications*. Springer-Verlag.

Seyranian, A. P., & Mailybaev, A. A. (2004). *Multiparameter stability theory with mechanical applications*. World Scientific.

Stewart, G. W., & Sun, J. (1990). *Matrix perturbation theory*. Academic Press.
Xiao, L., & Boyd, S. (2004). Fast linear iterations for distributed averaging. *Systems & Control Letters*, *53*(1), 65–78.

Kai Cai received his B. Eng. degree in Electrical Engineering from Zhejiang University (China) in 2006, his M.A.Sc. degree in Electrical and Computer Engineering from the University of Toronto (Canada) in 2008, and his Ph.D. degree in Systems Science from Tokyo Institute of Technology (Japan) in 2011.

He is currently a Postdoctoral Fellow in Systems Control, with the Department of Electrical and Computer Engineering of the University of Toronto. His research interests are distributed control of multi-agent systems, distributed control of discrete-event systems, and theory of control architecture.

Hideaki Ishii received his M.Eng. degree in applied systems science from Kyoto University in 1998, and his Ph.D. degree in electrical and computer engineering from the University of Toronto in 2002.

He was a Postdoctoral Research Associate of the Coordinated Science Laboratory, the University of Illinois at Urbana-Champaign from 2001 to 2004, and a Research Associate of the Department of Information Physics and Computing, The University of Tokyo from 2004 to 2007. He is currently an Associate Professor of the Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology. He is a co-author of the book *Limited Data Rate in Control Systems with Networks*, Lecture Notes in Control and Information Sciences, vol. 275, Springer, Berlin, 2002. He is the Chair of the IFAC Technical Committee on Networked Systems. He is on the editorial board of *Automatica* and previously of the *IEEE Transactions on Automatic Control*. His research interests are in networked control systems, multi-agent systems, hybrid systems, and probabilistic algorithms.