[4] D. L. Kleinman, "On an iterative technique for Riccati equation computation," *IEEE Trans. Autom. Control*, vol. AC-13, no. 1, pp. 114–115, 1968.
[5] Y.-W. Liang and D.-C. Liaw, "Common stabilizers for linear control systems in the presence of actuators outage," *Appl. Math. Comput.*, vol. 177, no. 2, pp. 635–643, 2006.
[6] F. Liao, J. L. Wang, and G. H. Yang, "Reliable robust flight tracking control: An LMI approach," *IEEE Trans. Control Syst. Technol.*, vol. 10, no. 1, pp. 76–89, 2002.
[7] J. Lunze and T. Steffen, "Control reconfiguration after actuator failures using disturbance decoupling methods," *IEEE Trans. Autom. Control*, vol. 51, no. 10, pp. 1590–1601, Oct. 2006.
[8] M. Maki, J. Jiang, and K. Hagino, "A stability guaranteed active fault-tolerant control system against actuator failures," *Int. J. Robust Nonlin. Control*, vol. 14, no. 12, pp. 1061–1077, 2004.
[9] G. Pujol, J. Rodellar, J. M. Rossell, and F. Pozo, "Decentralized reliable guaranteed cost control of uncertain systems: An LMI design," *IET Control Theory Appl.*, vol. 1, no. 3, pp. 779–785, 2007.
[10] M. Staroswiecki, H. Yang, and B. Jiang, "Active fault tolerant control based on progressive accommodation," *Automatica*, vol. 43, no. 12, pp. 2070–2076, 2007.
[11] M. Staroswiecki, D. Berdjag, B. Jiang, and K. Zhang, "PACT: A PAssive/ACTive approach to fault tolerant stability under actuator outages," in *Proc. 48th IEEE CDC and 28th CCC*, Shanghai, China, 2009, pp. 7819–7824.
[12] M. Staroswiecki and D. Berdjag, "Passive/active fault tolerant control for LTI systems with actuator outages," in *Proc. ECC'09*, Budapest, Hungary, 2009, pp. 2506–2511.
[13] M. Staroswiecki and D. Berdjag, "A general fault tolerant linear quadratic control strategy under actuator outages," *Int. J. Syst. Sci.*, vol. 41, no. 8, pp. 971–985, 2010.
[14] G. Tao, S. M. Joshi, and X. L. Ma, "Adaptive state feedback and tracking control of systems with actuator failures," *IEEE Trans. Autom. Control*, vol. 46, no. 1, pp. 78–95, Jan. 2001.
[15] R. J. Veillette *et al.*, "Design of Reliable Control systems," *IEEE Trans. Autom. Control*, vol. 37, no. 3, pp. 290–304, Mar. 1992.
[16] R. J. Veillette, "Reliable linear-quadratic state-feedback control," *Automatica*, vol. 32, no. 1, pp. 137–143, 1995.
[17] X. Yu, Y. Li, X. Wang, and K. Zhao, "An autonomous robust fault tolerant control system," in *Proc. IEEE Int. Conf. Inf. Acquis.* , Shandong, China, 2006, pp. 1191–1196.
[18] G. H. Yang, J. L. Wang, and Y. C. Soh, "Reliable H_∞ controller design for linear systems," *Automatica*, vol. 37, no. 5, pp. 717–725, 2003.
[19] Z. Yang, M. Blanke, and M. Verhagen, "Robust control mixer method for reconfigurable control design using model matching," *Proc. IET Control Theory Appl.*, vol. 1, no. 1, pp. 349–357, 2007.

Averaging Over General Random Networks

Kai Cai

***Abstract*—This technical note studies the distributed averaging problem over general random networks, by means of augmenting state space. A general iterative scheme (with a certain structure) is proposed that is discrete-time, linear, and stochastic; its generality compared to the literature lies in that the weight matrices corresponding to the networks need not be column-stochastic, and the random process generating the update matrices need not be ergodic or i.i.d. It is then justified that the scheme achieves average consensus in the mean-square sense, which, in a special case, also implies averaging with probability one. A key technique to the justification is a matrix perturbation result, which describes the behavior of eigenvalues perturbed simultaneously by multiple parameters.**

***Index Terms*—Distributed averaging, distributed consensus, matrix perturbation theory, mean-square analysis, random networks/graphs, stationary stochastic systems.**

I. Introduction

Distributed consensus and averaging problems have been extensively studied for their potential applications in motion coordination, information fusion, load balancing, to name a few [1]–[3]. Recently much attention has been paid to the stochastic, discrete-time, linear iterative scheme: $x(t+1) = W_t x(t)$, where $x(t) \in \mathbb{R}^n$ is the state vector at time $t \in \{0, 1, 2, \ldots\}$ and W_t are the row-stochastic weight matrices with positive diagonal entries generated by some random process. References [4]–[8] considered the i.i.d. case of W_t; in particular [8] proved that every entry of $x(t)$ asymptotically converges to a common value (i.e. consensus) almost surely if and only if the second largest eigenvalue λ_2 of the expected weight matrix $\mathcal{E}W_t$ satisfies $|\lambda_2| < 1$. This characterization was later extended to the ergodic stationary case of W_t in [9]. For almost sure averaging (i.e. consensus value $\mathbf{1}^T x(0)/n$, $\mathbf{1} = [1 \cdots 1]^T$) it is found that, in addition to the above condition for consensus, W_t need to be column-stochastic with probability one [8], [10]; the latter condition ensures that the state sum $\mathbf{1}^T x(t)$ be an invariant. It might sometimes be desirable, however, not to require almost sure column-stochasticity of W_t; in that case, new iterative schemes have been proposed to achieve averaging by means of augmenting state space with an extra vector $y(t) \in \mathbb{R}^n$. [11], [12] designed a protocol involving the element-wise division $x(t)/y(t)$; and when W_t are from an ergodic stationary process, [12] proved that $x(t)/y(t)$ asymptotically tends to the initial average almost surely. References [13], [14], and developed the linear update algorithm

$$\begin{bmatrix} x(t+1) \\ y(t+1) \end{bmatrix} = \Gamma_t \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}, \quad \Gamma_t := \begin{bmatrix} W_t & \epsilon U_t \\ I - W_t & V_t - \epsilon U_t \end{bmatrix} \tag{1}$$

Manuscript received November 20, 2011; revised January 17, 2012; accepted April 19, 2012. Date of publication May 14, 2012; date of current version November 21, 2012. This work was supported in part by the Leave a Nest Grant. Recommended by Associate Editor L. Schenato.

The author is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: kai.cai@scg.utoronto.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2012.2199181

here $U_t(\neq 0)$ are nonnegative,[1] V_t column-stochastic whose nonzero entries are at the same locations as W_t, and finally ϵ is a real scalar. It is well to note that Γ_t are *not* nonnegative, for (at least) $I - W_t$ are not. The matrices W_t, U_t, and V_t studied in [13], have special forms corresponding, respectively, to the broadcast gossip [7] and the pairwise gossip [10]; and when Γ_t are i.i.d. and $\epsilon > 0$ is sufficiently small, justified mean-square averaging of $x(t)$.

This technical note studies the averaging problem, the main purpose being to bring forward the linear iterative scheme (1) to greater (theoretic) generality. The generalizations can offer more flexibility in algorithm implementation, as we shall remark in detail in the next section (after the notion of distributed iteration over digraphs is introduced). First, more general forms of W_t and V_t than gossip (pairwise or broadcast) are considered, and the requirement that nonzero entries of these matrices be at the same locations is dropped. Second, the term ϵU_t is extended to $U_t(p)$ that depends linearly on a real vector $p := [p_1 \cdots p_m]^T$, rather than on a scalar. Third (and notably), we suppose the sequence $\{\Gamma_0, \Gamma_1, \Gamma_2, \ldots\}$ is generated by a general stationary random process, which does not require ergodicity and subsumes the i.i.d. case (as compared to [4]–[12]). It is under all these generalizations that we establish, again, that when p is positive and small, $x(t)$ asymptotically converges to the initial average in mean-square. As a direct consequence, our results contain those of [13], as special cases. Also, contrasted with [4]–[10], our averaging scheme (with a specific structure) does not impose the column-stochastic condition on any W_t, however at the price of doubling the size of state space. Choosing one scheme over the other may depend probably on specific applications at hand; nevertheless, with the tradeoff identified, our approach can be a helpful alternative to the standard $x(t+1) = W_t x(t)$. Finally, we employ matrix perturbation theory to establish convergence, as , but in a more general form to handle a vector p of parameters.

II. Iterative Scheme

Let us begin with the discrete-time linear iterative equation $x(t+1) = W_t x(t)$, where $x(t) \in \mathbb{R}^n$ is the state vector and W_t the weight matrix at time $t \in \{0, 1, 2, \ldots\}$. For a fixed W_t, there is a natural way to define the corresponding digraph $\mathcal{G}W_t$ on n nodes: an edge (j, i) from node j to node i exists in $\mathcal{G}W_t$ if and only if the (i, j)-th entry of W_t is nonzero. As so defined, the equation $x(t+1) = W_t x(t)$ represents a distributed update scheme over the nodes of $\mathcal{G}W_t$: the state $x_i(t+1)$ depends solely on those in the set $\{x_j(t) : (j, i) \text{ is an edge of } \mathcal{G}W_t\}$. We say W_t is *irreducible* if the digraph $\mathcal{G}W_t$ is *strongly connected*; that is, there exists a sequence of directed edges connecting every pair of nodes [3, Appendix C].

Now augment state space with an additional vector $y(t) \in \mathbb{R}^n$, and consider the following discrete-time, linear iterative scheme:

$$\begin{bmatrix} x(t+1) \\ y(t+1) \end{bmatrix} = \Gamma_t \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}, \quad \Gamma_t := \begin{bmatrix} W_t & U_t(p) \\ I - W_t & V_t - U_t(p) \end{bmatrix} \tag{2}$$

where $U_t(p)$ are linear in a real parameter vector $p = [p_1 \cdots p_m]^T$. From $x(t+1) = W_t x(t) + U_t(p) y(t)$, we see that the second vector $y(t)$ through $U_t(p)$ influences the update of $x(t)$. Also observe

$$\begin{aligned} y(t+1) &= V_t y(t) + ((I - W_t)x(t) - U_t(p)y(t)) \\ &= V_t y(t) + (x(t) - x(t+1)); \end{aligned}$$

thus $y(t)$ is updated through V_t on one hand, and records the change of $x(t)$ on the other. For the scheme (2), we shall make several assumptions, below.

[1]A matrix is nonnegative if every of its entry is a nonnegative real number; and throughout the technical note, by nonnegative matrices we exclude the 0 matrix, i.e. there exists at least one positive entry.

(A1) W_t are row-stochastic (i.e. nonnegative and $W_t \mathbf{1} = \mathbf{1}$), and with positive diagonal entries.

(A2) V_t are column-stochastic (i.e. nonnegative and $\mathbf{1}^T V_t = \mathbf{1}^T$), and with positive diagonal entries.

By **(A1)** the update of $x(t)$ through W_t is to achieve consensus; the sum $\mathbf{1}^T x(t)$, however, need not be invariant. On the other hand, by **(A2)** the update of $y(t)$ through V_t preserves the sum $\mathbf{1}^T y(t)$. Then, according to (2) one derives $[\mathbf{1}^T \ \mathbf{1}^T]\Gamma_t = [\mathbf{1}^T \ \mathbf{1}^T]$, and it follows that the total sum $\mathbf{1}^T(x(t) + y(t))$ is a constant—this is the essential idea of using the extra $y(t)$ to *backup* every shift of $\mathbf{1}^T x(t)$ for the purpose of averaging. It should now be stressed that **(A2)** requires each node to know its out-degree. In addition, it is not required that the nonzero entries of W_t and V_t be at the same locations (contrast with (1) studied in [13],). This is originally motivated by reducing the communication effort of $y(t)$ (which may frequently be a practical need) when the digraph $\mathcal{G}V_t$ has only a (proper) subset of the digraph $\mathcal{G}W_t$'s edges. It then turns out that, as long as **(A1)** and **(A2)** hold, $\mathcal{G}W_t$ and $\mathcal{G}V_t$ can have arbitrarily different topologies, i.e. the connection structures of the components of $x(t)$ and $y(t)$.

(A3) The parameter vector $p := \epsilon e$, where $\epsilon \geq 0$ is a nonnegative scalar and $e = [e_1 \cdots e_m] > 0$ a positive vector.

Thus $p > 0$ or $p = 0$; and $p > 0$ if and only if $\epsilon > 0$. Allowing a vector of parameter values may offer more flexibility in a distributed implementation of the scheme (2) over the digraphs nodes, inasmuch as individual nodes may independently choose different values for these local parameters. This may also provide some robustness of the scheme (2) with respect to parameter variations.

(A4) $U_t(0) = 0$; and the derivatives $\partial U_t(p)/\partial p_i$, $i \in [1, m]$, are constant (free of p) nonnegative matrices.

The term $U_t(p)$ is more general than ϵU_t in (1) from two aspects: for one, it depends on a vector of parameters rather than a scalar; for the other, $\partial U_t(p)/\partial p_i$ can be arbitrary nonnegative matrices whereas $U_t(= d(\epsilon U_t)/d\epsilon)$ are specific to the gossip type [13], .

(A5) $y(0) = 0$.

The justification for this is: as observed above $y(t)$ records the change $x(t-1) - x(t)$, but there is no change of $x(0)$ at the initial time 0.

Finally, let us consider the matrices Γ_t in (2), $t = 0, 1, 2, \ldots$, are generated by a random process. Let $(\Omega_0, \mathcal{B}_0)$ be a measurable space, where

$$\Omega_0 := \{\text{set of } \Gamma_t \text{ satisfying assumptions } (\mathbf{A1}) - (\mathbf{A4})\} \tag{3}$$

and $\mathcal{B}_0$ the Borel σ-algebra on Ω_0. Define a product space $(\Omega, \mathcal{B})$ by $\Omega := \{(\omega_1, \omega_2, \ldots) : \omega_t \in \Omega_0\}$ and $\mathcal{B} := \mathcal{B}_0 \times \mathcal{B}_0 \times \cdots$, and consider a probability measure $\mathbb{P}$ so that $(\Omega, \mathcal{B}, \mathbb{P})$ forms a probability space. Thus, a sequence of Γ_t can be defined by $\Gamma_t : \Omega \to \Omega_0 : (\omega_1, \omega_2, \ldots) \mapsto \omega_t$, $t = 0, 1, 2, \ldots$. We say a random sequence $\{\Gamma_0, \Gamma_1, \Gamma_2, \ldots\}$ is *stationary* if the families $\{\Gamma_{t_1}, \Gamma_{t_2}, \ldots, \Gamma_{t_h}\}$ and $\{\Gamma_{t_1+l}, \Gamma_{t_2+l}, \ldots, \Gamma_{t_h+l}\}$ have the same joint distribution for all $t_1, t_2, \ldots, t_h$ and $l > 0$.

(A6) The sequence $\{\Gamma_0, \Gamma_1, \Gamma_2, \ldots\}$ is a stationary random process.

Thus Γ_t have the same (finite-dimensional) distribution for all t. We stress that stationarity is the only statistical property we assume for the sequence $\{\Gamma_0, \Gamma_1, \Gamma_2, \ldots\}$, which is more general than [4]–[12] in that neither ergodicity nor i.i.d. is required. This implies that in real networks, we may deal with cases where strong correlations exist in the sample of network topologies, as well as long-term samples are not regular enough to deduce the statistical properties of networks.

(**A7**) The expected matrices $\mathcal{E}W_t$ and $\mathcal{E}V_t$ are irreducible.[2]

For $\mathcal{E}W_t$, besides irreducibility (**A7**), it is also row-stochastic with positive diagonal entries owing to (**A1**). Thus by the Perron-Frobenius and Gershgorin Theorems (e.g. [15]), its second largest eigenvalue $|\lambda_2(\mathcal{E}W_t)| < 1$. The same conclusion holds for $\mathcal{E}V_t$, i.e. $|\lambda_2(\mathcal{E}V_t)| < 1$.

III. Convergence Result

Our main result is that the iterative scheme (2) achieves average consensus of $x(t)$ in mean-square.

Theorem 1: Suppose that the assumptions (**A1**)–(**A7**) hold. Then there exists $p > 0$ such that the iterative scheme (2) achieves mean-square average consensus, i.e. for every $x(0)$ there holds $\mathcal{E}\|x(t) - \mathbf{1}\mathbf{1}^T x(0)/n\|_2^2 \to 0$ and $\mathcal{E}\|y(t)\|_2^2 \to 0$ as $t \to \infty$.

We postpone the proof to the next section, and discuss here the entailed comparisons and implications.

First, Theorem 1 contains the main results of [13], as special cases, inasmuch as the scheme (2) extends (1) in a number of respects. Concretely, in (**A1**)–(**A4**), the matrices W_t, V_t, and U_t take more general forms than the gossip type (pairwise or broadcast), and $U_t(p)$ depends on a vector of parameters rather than a single scalar; in (**A6**), the stationary sequence $\{\Gamma_0, \Gamma_1, \Gamma_2, \ldots\}$ is a much broader class of random processes than the i.i.d. Technically (as will be seen in the section below), our extension centers around a more general matrix perturbation result, which describes the behavior of certain eigenvalues perturbed simultaneously by multiple parameters.

Second, we comment on the assumption that $\mathcal{E}W_t$ is irreducible in (**A7**), which means that the corresponding digraph $\mathcal{G}(\mathcal{E}W_t)$ is strongly connected. From [4]–[9] that studied the iteration $x(t+1) = W_t x(t)$, this assumption is *not* a necessary condition for consensus; and in fact, a necessary condition is weaker: the digraph $\mathcal{G}(\mathcal{E}W_t)$ contains a spanning tree. On the other hand, for the averaging of $x(t)$ using the scheme (1), proved that the irreducibility of $\mathcal{E}W_t$ is, indeed, necessary. Now for our scheme (2) to achieve averaging, it is well to emphasize that a necessary condition is that the matrix $(\mathcal{E}W_t + \mathcal{E}U_t(p))$ is irreducible. This condition is weaker than , for in the scheme (2) more general forms of $U_t(p)$ are allowed. In addition, this condition permits even that the digraph $\mathcal{G}(\mathcal{E}W_t)$ has no spanning tree, insofar as $U_t(p)$ is chosen so that $\mathcal{G}(\mathcal{E}W_t + \mathcal{E}U_t(p))$ is strongly connected.

Corollary 1: If the iterative scheme (2) achieves (mean-square) average consensus, then the matrix $(\mathcal{E}W_t + \mathcal{E}U_t(p))$ is irreducible.

Proof: Suppose on the contrary that $(\mathcal{E}W_t + \mathcal{E}U_t(p))$ is not irreducible; or equivalently, the digraph $\mathcal{G}(\mathcal{E}W_t + \mathcal{E}U_t(p))$ is not strongly connected. From graph theory, the latter implies that there exists a proper subset $\mathcal{V}_1$ of $h(\in [1, n-1])$ nodes to which the rest $\mathcal{V}_2$ of $n-h$ nodes are not connected; namely, no information flow, neither x through $\mathcal{E}W_t$ nor y through $\mathcal{E}U_t(p)$, is from the nodes in $\mathcal{V}_2$ to the nodes in $\mathcal{V}_1$. Thus, if an initial $x(0)$ is such that the nodes in $\mathcal{V}_1$ all have the same value c_1 and the nodes in $\mathcal{V}_2$ all have $c_2(\neq c_1)$, then no update of x or y can ever occur in $\mathcal{V}_1$. But the initial average is $\mathbf{1}^T x(0)/n = (hc_1 + (n-h)c_2)/n \neq c_1$. Consequently, the iterative scheme (2) cannot achieve average consensus in the mean, let alone in mean-square. ∎

The third remark is on the types of convergence, in particular mean-square (Theorem 1, [10],) and with probability one ([4], [6]–[9], [12]). In general, there is no implication between these two convergence notions. However, in the special case where the set Ω_0 of matrices Γ_t in (3) is finite, as well as under our linear, stationary setting, mean-square convergence is stronger than with probability one [16, Corollary 3.46].

[2]Owing to (**A6**), these expectations $\mathcal{E}W_t$ and $\mathcal{E}V_t$ are invariant over time.

Corollary 2: Let the assumptions of Theorem 1 hold, and suppose the set Ω_0 in (3) is finite. Then the iterative scheme (2) achieves almost sure average consensus.

Lastly, we should point out two open questions: how large may the parameter p be, and how fast do $x(t)$ and $y(t)$ converge under the scheme (2). These questions induce two corresponding problems: one is to derive an upper bound for the parameter p which ensures average consensus; the other is to find an optimal value of p which maximizes convergence speed. Systematic inquiries into these problems, however, would seem to require a complete, *analytic* characterization of the perturbed eigenvalue trajectories in terms of the vector p, which is itself, yet, an unsolved problem in matrix perturbation theory, and thus beyond the scope of the present technical note. Nevertheless, we shall shed some light on these raised issues by *simulation* in Section V, below.

IV. Proof Technique

We present the proof of Theorem 1, which is organized into three steps.[3] Throughout this section we suppose that the assumptions (**A1**)–(**A7**) hold.

In the first step, we derive a matrix spectral condition that characterizes mean-square average consensus. Let $\otimes$ stand for the Kronecker product, and consider the $4n^2 \times 4n^2$ matrix $\mathcal{E}(\Gamma_t \otimes \Gamma_t)$. Since the matrices Γ_t, $t = 0, 1, 2, \ldots$, are stationary (**A6**), so are $\Gamma_t \otimes \Gamma_t$ and we simplify $\mathcal{E}(\Gamma_t \otimes \Gamma_t)$ by the notation $\mathcal{E}(\Gamma \otimes \Gamma)$. Below is a necessary and sufficient condition in terms of the spectrum of $\mathcal{E}(\Gamma \otimes \Gamma)$ for mean-square average consensus; the proof is standard, and can be found in, e.g., [10], .

Lemma 1: The scheme (2) achieves mean-square average consensus if and only if $\mathcal{E}(\Gamma \otimes \Gamma)$ has a simple eigenvalue 1, and all the other eigenvalues with moduli smaller than one.

Directly analyzing the spectrum of the (large) matrix $\mathcal{E}(\Gamma \otimes \Gamma)$ is, however, not a straightforward task. Nevertheless, owing to the block structure of Γ_t, we propose a similarity transformation on $\mathcal{E}(\Gamma \otimes \Gamma)$ that generates a manageable structure. Let us write $\Gamma_t = \Phi_t + \Psi_t$, where

$$\Phi_t := \begin{bmatrix} W_t & 0 \\ I - W_t & V_t \end{bmatrix} \quad \text{and} \quad \Psi_t := \begin{bmatrix} 0 & U_t(p) \\ 0 & -U_t(p) \end{bmatrix}. \tag{4}$$

Then $\mathcal{E}(\Gamma \otimes \Gamma) = \mathcal{E}(\Phi \otimes \Phi) + \mathcal{E}(\Phi \otimes \Psi + \Psi \otimes \Phi + \Psi \otimes \Psi)$. Now let $h \in [1, 4n]$ and $h\mathbf{n} := \{(h-1)n + 1, \ldots, hn\}$. Consider the following permutation , which is to permute the columns of $\mathcal{E}(\Gamma \otimes \Gamma)$:

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

Denote by P the corresponding permutation matrix (which is orthogonal); we obtain

$$\hat{\Gamma} := P^T \mathcal{E}(\Gamma \otimes \Gamma) P = \hat{\Phi} + \hat{\Psi} \tag{5}$$

where $\hat{\Phi}$, $\hat{\Psi}$ are defined, respectively, in (6) and (7), as shown at the bottom of the following page. Note that the parameter vector p is only in $\hat{\Psi}$, not in $\hat{\Phi}$. As the above is a similarity transformation, the spectrum of $\hat{\Gamma}$ is the same as $\mathcal{E}(\Gamma \otimes \Gamma)$. Therefore, we have proved the following result.

[3]The approach used here is essentially the same as , however with the generalization of employing a multidimensional matrix perturbation theory to deal with a vector of parameters. There are also other extensions along the proof, due to that the scheme (2) is more general than (1) in several respects. For a reasonably self-contained presentation, we shall include necessary derivations.

Proposition 1: The scheme (2) achieves mean-square average consensus if and only if $\hat{\Gamma}$ in (5) has a simple eigenvalue 1, and all the other eigenvalues with moduli smaller than one.

The remaining two steps in the proof will together establish the desired spectral property of $\hat{\Gamma}$ in (5), using an eigenvalue perturbation result with the vector p viewed as the perturbing parameter. Step 2 investigates the spectrum of $\hat{\Phi}$ in (6) (without p), thereby laying an important basis for applying the perturbation result.

Proposition 2: The spectrum of the matrix $\hat{\Phi}$ in (6) satisfies

$$1 = \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 > |\lambda_5| \geq \cdots \geq |\lambda_{4n^2}|. \tag{8}$$

Moreover, the eigenvalue 1 is semi-simple, i.e., its geometric multiplicity equals 4.

To justify Proposition 2 the fact below is needed. It is an easy corollary of the Perron-Frobenius Theorem (cf. [15, Ch. XIII]), but rarely used in the consensus literature.

Lemma 2: Let M be a nonnegative and irreducible matrix, and λ be an eigenvalue of M. If there is a positive vector ξ such that $M\xi = \lambda\xi$ or $\xi^T M = \lambda\xi^T$, then $\lambda = \rho(M)$, the spectral radius of M.

Proof of Proposition 2: First, we show (8). Since $\hat{\Phi}$ in (6) is (block) lower-triangular, it suffices to prove that each of the four diagonal block matrices

$$\mathcal{E}(W \otimes W), \quad \mathcal{E}(W \otimes V), \quad \mathcal{E}(V \otimes W), \quad \mathcal{E}(V \otimes V) \tag{9}$$

has a simple eigenvalue 1 and all the other eigenvalues with moduli smaller than one. From (**A1**) and (**A2**), we obtain that these four matrices are all nonnegative and with positive diagonal entries; also, $W_t\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T V_t = \mathbf{1}^T$ for every t. On the other hand, it follows from (**A7**) and the Perron-Frobenius Theorem (e.g. [15]) that there exist positive vectors ξ_1 and ξ_2 such that

$$\xi_1^T \mathcal{E}(W) = \xi_1^T \quad \text{and} \quad \xi_1^T \mathbf{1} = 1 \tag{10}$$

$$\mathcal{E}(V)\xi_2 = \xi_2 \quad \text{and} \quad \mathbf{1}^T \xi_2 = 1. \tag{11}$$

Then a short calculation yields

$$\begin{aligned}
\mathcal{E}(W \otimes W)(\mathbf{1} \otimes \mathbf{1}) &= (\mathbf{1} \otimes \mathbf{1}) \\
\mathcal{E}(W \otimes V)(\mathbf{1} \otimes \xi_2) &= (\mathbf{1} \otimes \xi_2) \\
\left(\mathbf{1}^T \otimes \xi_1^T\right)\mathcal{E}(V \otimes W) &= \left(\mathbf{1}^T \otimes \xi_1^T\right) \\
(\mathbf{1}^T \otimes \mathbf{1}^T)\mathcal{E}(V \otimes V) &= (\mathbf{1}^T \otimes \mathbf{1}^T).
\end{aligned}$$

These equations imply that 1 is an eigenvalue of each matrix in (9), and has a positive right or left eigenvector. Further we *claim* that all the four matrices are irreducible. Hence, by the Perron-Frobenius Theorem, as well as Lemma 2, the eigenvalue 1 is simple and the spectral radius of each matrix in (9). Lastly it follows from the Gershgorin Theorem (e.g. [15]) that all the other eigenvalues of each matrix in (9) have moduli smaller than one.

Now we prove the claim. We shall justify in detail that $\mathcal{E}(W \otimes V)$ is irreducible; and for the other three matrices in (9), the justification follows virtually the same. Denote by $\hat{\mathcal{G}} = (\hat{\mathcal{V}}, \hat{\mathcal{E}})$ the digraph corresponding to $\mathcal{E}(W \otimes V)$. It will be shown that $\hat{\mathcal{G}}$ is strongly connected. Let $\mathcal{V}, \mathcal{V}'$ be the node sets respectively of $\mathcal{G}W_t$ and $\mathcal{G}V_t$ (in general with different topologies). Then $\hat{\mathcal{V}} = \mathcal{V} \times \mathcal{V}' = \{(i, i') : i \in \mathcal{V}, i' \in \mathcal{V}'\}$, and make the arrangement $\hat{\mathcal{V}} = \mathcal{V}_1 \cup \cdots \cup \mathcal{V}_n$, where $\mathcal{V}_i = \{(i,1), \ldots, (i,n)\}$ for every $i \in [1,n]$. Now consider the edge set $\hat{\mathcal{E}}$. For a fixed $W_t \otimes V_t$, every nonzero (i,j)-th entry of W_t corresponds to an edge from a node $(j,*)$ in $\mathcal{V}_j$ to a node $(i,*)$ in $\mathcal{V}_i$, and every nonzero (k,l)-th entry of V_t corresponds to an edge from (i,l) to (i,k) in each $\mathcal{V}_i$, $i \in [1,n]$. Based on these two correspondences, the assumption that $\mathcal{E}W$ is irreducible implies that the n sets $\mathcal{V}_1, \ldots, \mathcal{V}_n$ are strongly connected; and that $\mathcal{E}V$ is irreducible implies that the n nodes in $\mathcal{V}_i$ are strongly connected for every $i \in [1,n]$. Therefore, there exists a sequence of directed edges from (j,l) to (i,k) for every $i,j,k,l \in [1,n]$, i.e., $\hat{\mathcal{G}}$ is strongly connected.

It is left to show that the eigenvalue 1 is semi-simple. By a straightforward calculation one derives rank $(\hat{\Phi} - I) = 4n^2 - 4$, which implies that the geometric multiplicity of the eigenvalue 1 equals four, the same as its algebraic multiplicity. ∎

The third and last step in the proof employs a perturbation result, viewing the matrix $\hat{\Psi}$ in (7) with the parameter p as a perturbing term to the spectrum of $\hat{\Phi}$ in (6). Below is the perturbation result, borrowed from [17, Ch. 2]. This result is the multidimensional generalization of the one used in , which handles parametric perturbation simultaneously from multiple directions.

Lemma 3: Consider an $n \times n$ matrix $M(p)$ which depends smoothly on a real parameter vector $p = [p_1 \cdots p_m]^T$. Suppose $p := \epsilon e$, where $\epsilon \geq 0$ and $e = [e_1 \cdots e_m] > 0$; so $p > 0$ if and only if $\epsilon > 0$, in which case $e = p/\epsilon$. Fix $h \in [1,n]$; let $\lambda_1 = \cdots = \lambda_h$ be a semi-simple eigenvalue of $M(0)$, with (linearly independent) right eigenvectors $v_1, \ldots, v_h$ and left eigenvectors $u_1, \ldots, u_h$ such that

$$\begin{bmatrix} u_1^T \\ \vdots \\ u_h^T \end{bmatrix} [v_1 \quad \cdots \quad v_l] = I_h. \tag{12}$$

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

$$\begin{bmatrix} u_1^T \partial M v_1 & \cdots & u_1^T \partial M v_h \\ \vdots & & \vdots \\ u_h^T \partial M v_1 & \cdots & u_h^T \partial M v_h \end{bmatrix}, \text{ where } \partial M := \sum_{i=1}^{m} \left.\frac{\partial M(p)}{\partial p_i}\right|_{p=0} e_i. \tag{13}$$

$$\hat{\Phi} := P^T \mathcal{E}(\Phi \otimes \Phi) P = \mathcal{E} \begin{bmatrix} W \otimes W & 0 & 0 & 0 \\ W \otimes (I - W) & W \otimes V & 0 & 0 \\ (I - W) \otimes W & 0 & V \otimes W & 0 \\ (I - W) \otimes (I - W) & (I - W) \otimes V & V \otimes (I - W) & V \otimes V \end{bmatrix} \tag{6}$$

$$\begin{aligned}
\hat{\Psi} &:= P^T \mathcal{E}(\Phi \otimes \Psi + \Psi \otimes \Phi + \Psi \otimes \Psi) P \\
&= \mathcal{E} \begin{bmatrix} 0 & W \otimes U(p) & U(p) \otimes W & U(p) \otimes U(p) \\ 0 & -W \otimes U(p) & U(p) \otimes (I - W) & U(p) \otimes V - U(p) \otimes U(p) \\ 0 & (I - W) \otimes U(p) & -U(p) \otimes W & V \otimes U(p) - U(p) \otimes U(p) \\ 0 & (W - I) \otimes U(p) & U(p) \otimes (W - I) & U(p) \otimes (U(p) - V) - V \otimes U(p) \end{bmatrix}
\end{aligned} \tag{7}$$

Here comes the promised desired spectral property of the matrix $\hat{\Gamma}$ in (5).

Proposition 3: [4] If the parameter vector $p > 0$ is small, then the matrix $\hat{\Gamma} = \hat{\Phi} + \hat{\Psi}$ in (5) has a simple eigenvalue 1, and all the other eigenvalues with moduli smaller than one.

Proof: The matrix $\hat{\Phi}$ in (6) is free of the parameter vector p; and according to Proposition 2, $\hat{\Phi}$ has a semi-simple eigenvalue 1 of multiplicity four. To apply Lemma 3, we find the corresponding (linearly independent) right and left eigenvectors in (14), as shown at the bottom of the page. Here, ξ_1, ξ_2 are from (10) and (11); and one checks that the normalization condition (12) is satisfied.

Now with the parameter p, the matrix $\hat{\Psi}$ in (7) is added to $\hat{\Phi}$ to form $\hat{\Gamma}$. Denote the resulting eigenvalues by $\lambda_i(p)$, $i \in [1, 4n^2]$, corresponding to those λ_i in (8); we study their (instantaneous) behavior. Since $p > 0$ is small, let $\epsilon > 0$ be small and by (**A3**) $e := p/\epsilon > 0$. Also, from (5)–(7) and that $U_t(p)$ are linear in p, it follows that $\hat{\Gamma}$ depends smoothly on p. Thus, the conditions of Lemma 3 are all satisfied, and therefore the derivatives $d\lambda_i(p)/d\epsilon\big|_{\epsilon=0}$, $i \in [1, 4]$, are the eigenvalues of the matrix in (13). By $U_t(0) = 0$ and $\partial U_t(p)/\partial p_i$ constant (**A4**) one computes the matrix (13) in (15), where $\partial U(p) := \sum_{i=1}^{m} (\partial U(p)/\partial p_i)\big|_{p=0} \cdot (p_i/\epsilon)$. Hence

$$\left.\frac{d\lambda_1(p)}{d\epsilon}\right|_{\epsilon=0} = 0, \quad \left.\frac{d\lambda_2(p)}{d\epsilon}\right|_{\epsilon=0} = -n\xi_1^T \mathcal{E}\left(\partial U(p)\right)\xi_2$$
$$\left.\frac{d\lambda_3(p)}{d\epsilon}\right|_{\epsilon=0} = -n\xi_1^T \mathcal{E}\left(\partial U(p)\right)\xi_2$$
$$\left.\frac{d\lambda_4(p)}{d\epsilon}\right|_{\epsilon=0} = -2n\xi_1^T \mathcal{E}\left(\partial U(p)\right)\xi_2.$$

Since every $\partial U_t(p)/\partial p_i$ has positive entries (**A4**), $p_i/\epsilon > 0$, and ξ_1, ξ_2 are positive vectors, we derive $d\lambda_j(p)/d\epsilon\big|_{\epsilon=0} < 0$, $j = 2, 3, 4$. The above implies that when $\epsilon > 0$ is small, equivalently $p > 0$ small (**A3**), $\lambda_1(p)$ stays put while $\lambda_j(p)$, $j = 2, 3, 4$, move to the left along the real axis. So by continuity, there exists a (small) positive vector p^1 such that $\lambda_1(p^1) = 1$ and $\lambda_j(p^1) < 1$, $j = 2, 3, 4$. On the other hand, since eigenvalues are continuous functions of matrix entries [17], thus continuous of p, there exists a (small) positive vector p^2 such that

[4]Compared to , here all derivatives involved in the proof are partial derivatives, with respect to each direction of parametric perturbation.

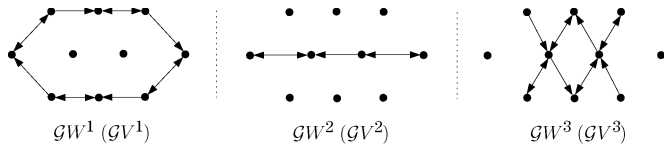

Fig. 1. Random networks of ten nodes that may stochastically switch among three different digraphs.

$|\lambda_l(p^2)| < 1$ for all $l \in [5, 4n^2]$, where the corresponding $|\lambda_l| < 1$ by (8). Take a small positive p such that $p < p^*$ (i.e., $p^* - p$ is positive), where $p_i^* := \min\{p_i^1, p_i^2\}$ for every $i \in [1, m]$; then $\lambda_1(p) = 1$ and $|\lambda_l(p)| < 1$ for all $l \in [2, 4n^2]$—the desired spectral property of $\hat{\Gamma}$. ■

Finally, our conclusion on mean-square average consensus in Theorem 1 follows immediately from Propositions 2 and 3.

V. Numerical Study on Parameter p

This section aims to illustrate, through a numerical study, the issue of how the values of the parameter vector p affect the convergence, as well as convergence speed, of $x(t)$ and $y(t)$ under the iterative scheme (2). By Lemma 1, we know that this amounts to studying the effect of p on the second largest eigenvalue of the matrix $\mathcal{E}(\Gamma \otimes \Gamma)$. Let us denote this special eigenvalue by $\lambda_2(p)$; and we will display the trajectories of $\lambda_2(p)$ with respect to the values of p.

Consider the three digraphs $\mathcal{G}W^l$, $l = 1, 2, 3$, in Fig. 1, each having 10 nodes but with different topologies. Construct the corresponding row-stochastic matrices W^l: the (i, j)-th off-diagonal entry equals 1/10 if and only if the edge (j, i) exists in $\mathcal{G}W^l$; and the values of the diagonals are determined to make each row sum 1. Likewise, construct the column-stochastic matrices V^l from the same digraphs. As a result, the positive entries of V^l and W^l appear at the same locations. Now consider the parameter vector $p = [p_1\ p_2]^T$, where $p_1, p_2 \geq 0$. Let

$$U(p) = \begin{bmatrix} p_1 I_5 & 0 \\ 0 & p_2 I_5 \end{bmatrix}, \quad \text{and} \quad \Gamma^l = \begin{bmatrix} W^l & U(p) \\ I - W^l & V^l - U(p) \end{bmatrix}$$

$l = 1, 2, 3$. Further, suppose the matrices Γ_t, $t = 0, 1, 2, \ldots$, are i.i.d., with the probability $\mathcal{P}(\Gamma_t = \Gamma^l) = 1/3$ for every $l \in [1, 3]$. All these simplifications (compared to the generality undertaken in the foregoing sections) are to provide ease of computing the eigenvalues of $\mathcal{E}(\Gamma_t \otimes \Gamma_t)$, especially the second largest one $\lambda_2(p)$; and considering

$$v_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ n\xi_2 \otimes n\xi_2 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 0 \\ n\xi_2 \otimes \mathbf{1} \\ -n\xi_2 \otimes n\xi_2 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 0 \\ \mathbf{1} \otimes n\xi_2 \\ 0 \\ -n\xi_2 \otimes n\xi_2 \end{bmatrix}, \quad v_4 = \begin{bmatrix} \mathbf{1} \otimes \mathbf{1} \\ -\mathbf{1} \otimes n\xi_2 \\ -n\xi_2 \otimes \mathbf{1} \\ n\xi_2 \otimes n\xi_2 \end{bmatrix}$$
$$u_1 = \begin{bmatrix} \frac{1}{n}\mathbf{1} \otimes \frac{1}{n}\mathbf{1} \\ \frac{1}{n}\mathbf{1} \otimes \frac{1}{n}\mathbf{1} \\ \frac{1}{n}\mathbf{1} \otimes \frac{1}{n}\mathbf{1} \\ \frac{1}{n}\mathbf{1} \otimes \frac{1}{n}\mathbf{1} \end{bmatrix}, \quad u_2 = \begin{bmatrix} \frac{1}{n}\mathbf{1} \otimes \xi_1 \\ 0 \\ \frac{1}{n}\mathbf{1} \otimes \xi_1 \\ 0 \end{bmatrix}, \quad u_3 = \begin{bmatrix} \xi_1 \otimes \frac{1}{n}\mathbf{1} \\ \xi_1 \otimes \frac{1}{n}\mathbf{1} \\ 0 \\ 0 \end{bmatrix}, \quad u_4 = \begin{bmatrix} \xi_1 \otimes \xi_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{14}$$

$$\begin{bmatrix} u_1^T \partial\hat{\Gamma} v_1 & u_1^T \partial\hat{\Gamma} v_2 & u_1^T \partial\hat{\Gamma} v_3 & u_1^T \partial\hat{\Gamma} v_4 \\ u_2^T \partial\hat{\Gamma} v_1 & u_2^T \partial\hat{\Gamma} v_2 & u_2^T \partial\hat{\Gamma} v_3 & u_2^T \partial\hat{\Gamma} v_4 \\ u_3^T \partial\hat{\Gamma} v_1 & u_3^T \partial\hat{\Gamma} v_2 & u_3^T \partial\hat{\Gamma} v_3 & u_3^T \partial\hat{\Gamma} v_4 \\ u_4^T \partial\hat{\Gamma} v_1 & u_4^T \partial\hat{\Gamma} v_2 & u_4^T \partial\hat{\Gamma} v_3 & u_4^T \partial\hat{\Gamma} v_4 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & 0 & 0 \\ n\xi_1^T \mathcal{E}\left(\partial U(p)\right)\xi_2 & -n\xi_1^T \mathcal{E}\left(\partial U(p)\right)\xi_2 & 0 & 0 \\ n\xi_1^T \mathcal{E}\left(\partial U(p)\right)\xi_2 & 0 & -n\xi_1^T \mathcal{E}\left(\partial U(p)\right)\xi_2 & 0 \\ 0 & n\xi_1^T \mathcal{E}\left(\partial U(p)\right)\xi_2 & n\xi_1^T \mathcal{E}\left(\partial U(p)\right)\xi_2 & -2n\xi_1^T \mathcal{E}\left(\partial U(p)\right)\xi_2 \end{bmatrix} \tag{15}$$

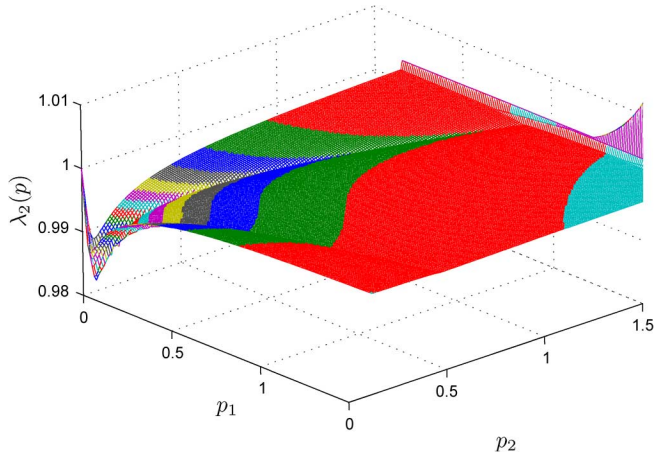

Fig. 2. Surface of $\lambda_2(p)$ with respect to parameter vector $p = [p_1\ p_2]^T$.

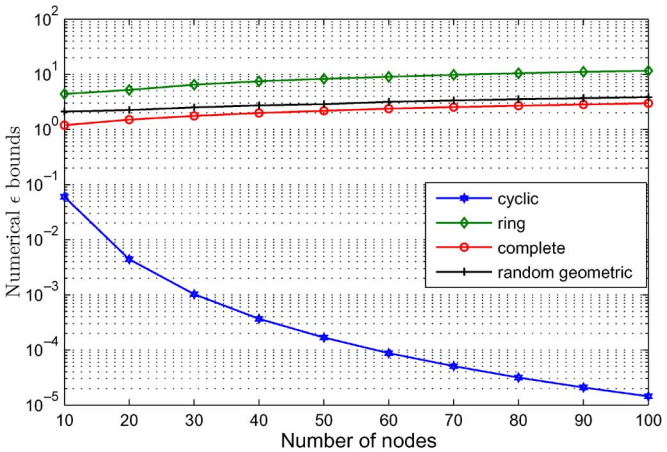

Fig. 3. Trends of ϵ bounds with respect to number of nodes for directed cyclic, bidirectional ring, complete, and random geometric (radius 0.5) graphs. Each plotted point is the average over 100 computations by the 3-step procedure.

just two parameters p_1, p_2 is to be able to visualize their influence on $\lambda_2(p)$. In Fig. 2 we display the trajectories of $\lambda_2(p)$ with respect to the values of p_1, p_2 for this example.

Observe that as p_1, p_2 both increase, the $\lambda_2(p)$ surface first sinks down (quickly), then rises up (slowly), and eventually grows over 1 where convergence would fail. Let us recall from the proof of Proposition 3 that (when p is small) certain eigenvalues move into the unit circle from 1, and some other eigenvalues move out from inside of the unit circle. When these two groups of eigenvalues *meet* (with equal moduli), that very moment is highly likely to produce the turning point from sink to rise of the $\lambda_2(p)$ surface; the turning point corresponds to the minimum value of $\lambda_2(p)$ (that is, the fastest convergence speed), and the corresponding vector $p^* = [p_1^*\ p_2^*]^T$ is in this sense optimal.

Now for a valid upper bound on the vector p which guarantees convergence, we would need to analyze the behavior of $\lambda_2(p)$ when its modulus is close to 1. To analytically characterize a tight bound for p is, however, very challenging, inasmuch as it seems to require the knowledge of detailed behavior of $\lambda_2(p)$ when p need not be small. Numerically, one may use the following procedure to derive a valid p value: Suppose an arbitrary positive vector e is chosen (locally by individual nodes), and $\epsilon(>0)$ to be designed such that $p = \epsilon e$. Also suppose Γ_t are given together with their stationary probability distribution. Then:

Step 1) Set $\epsilon = \epsilon_0 > 0$, an initial positive value.

Step 2) Compute $\lambda_2(p)$, the second largest eigenvalue of the matrix $\mathcal{E}(\Gamma \otimes \Gamma)$.

Step 3) If $\lambda_2(p) < 1$, terminate, and output ϵ. Otherwise, set $\epsilon = k\epsilon$, $k \in (0,1)$ constant, and go to Step 2.

We apply this (naive) procedure to study numerical ϵ bounds for different network topologies, by letting $e(>0)$ be an n-dimensional unit vector generated at random, $U(p) = \mathrm{diag}(\epsilon e)$, $\epsilon_0 = 12$, and $k = 0.99$. For simplicity, networks are assumed i.i.d. and their edges uniformly weighted. Computed results are displayed in Fig. 3, for directed cyclic, bidirectional ring, complete, and random geometric (radius 0.5) graphs. Observe that ϵ bounds decrease (quickly) in cyclic digraphs as the node number becomes larger, whereas they increase (slowly) in the other three types which are undirected graphs. Also, comparing the ϵ bounds among the latter three types might suggest that they could be related to the node degrees (it is seen here that the lower the degrees, the larger the ϵ bounds). Future research will aim to justify the above observations by analyzing ϵ bounds for these special topologies.

VI. Conclusions

We have justified that the iterative scheme (2) achieves average consensus under several generalizations compared to the literature: notably, the weight matrices W_t need not be column-stochastic, and the random process $\{\Gamma_0, \Gamma_1, \Gamma_2, \ldots\}$ need not be ergodic or i.i.d.. A key technique to the justification is the multi-parameter perturbation theory. Yet open problems include deriving an upper bound for the parameter vector p which ensures averaging, as well as an optimal value of p which maximizes convergence speed.

References

[1] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Upper Saddle River, NJ: Prentice-Hall, 1989.

[2] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[3] W. Ren and R. W. Beard, *Distributed Consensus Multi-vehicle Cooperative Control: Theory and Applications*. New York: Springer-Verlag, 2008.

[4] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Trans. Autom. Control*, vol. 50, no. 11, pp. 1867–1872, Nov. 2005.

[5] C. W. Wu, "Synchronization and convergence of linear dynamics random directed networks," *IEEE Trans. Autom. Control*, vol. 51, no. 7, pp. 1207–1210, Jul. 2006.

[6] M. Porfiri and D. J. Stilwell, "Consensus seeking over random weighted directed graphs," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1767–1773, Sep. 2007.

[7] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE J. Select. Areas Commun.*, vol. 26, no. 4, pp. 634–649, Apr. 2008.

[8] A. Tahbaz-Salehi and A. Jadbabaie, "A necessary and sufficient condition for consensus over random networks," *IEEE Trans. Autom. Control*, vol. 53, no. 3, pp. 791–795, Mar. 2008.

[9] A. Tahbaz-Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 225–230, Jan. 2010.

[10] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[11] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. 44th IEEE Symp. Foundations of Computer Science*, Washington, DC, 2003, pp. 482–491.

[12] F. Benezit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *Proc. IEEE Int. Symp. Information Theory*, Austin, TX, 2010, pp. 1753–1757.

[13] M. Franceschelli, A. Giua, and C. Seatzu, "Distributed averaging sensor networks based on broadcast gossip algorithms," *IEEE Sensors J.*, vol. 11, no. 3, pp. 808–817, 2011.

[14] K. Cai and H. Ishii, "Average consensus on arbitrary strongly connected digraphs with dynamic topologies," in *Proc. American Control Conf.*, Montreal, QC, Canada, 2012, pp. 14–19 [Online]. Available: http://arxiv.org/abs/1203.2563

[15] F. R. Gantmacher, *The Theory of Matrices*. New York: Chelsea, 1959, vol. 2.

[16] O. Costa, M. Fragoso, and R. Marques, *Discrete-Time Markov Jump Linear Systems*. New York: Springer-Verlag, 2004.

[17] A. Seyranian and A. Mailybaev, *Multiparameter Stability Theory With Mechanical Applications*. Singapore: World Scientific, 2004.